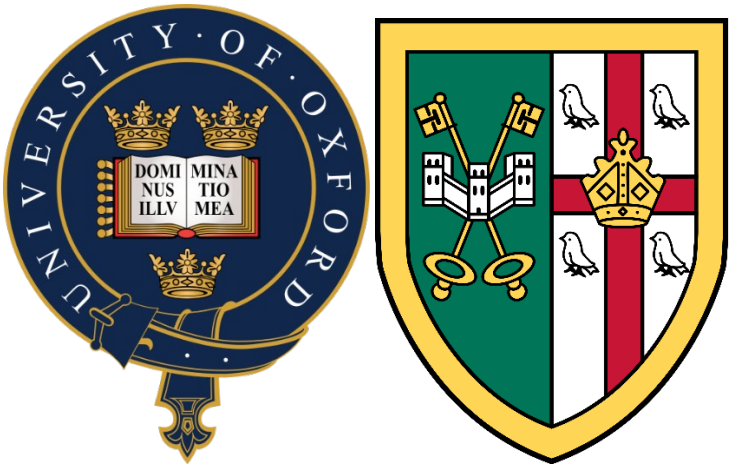
Uncovering Epithelial Lineage Complexity and Cell Fate Trajectories in Human Intestinal Organoids through advanced Single-cell Genomics

Nuffield Department of Medicine

Ana Sousa Gerós, MSc

St Peter's College, University of Oxford



Thesis submitted for the degree of

Doctor or Philosophy

2025

“I have a friend who's an artist and has sometimes taken a view which I don't agree with very well. He'll hold up a flower and say "look how beautiful it is," and I'll agree. Then he says "I as an artist can see how beautiful this is but you as a scientist take this all apart and it becomes a dull thing," and I think that he's kind of nutty (…) I see much more about the flower than he sees. I could imagine the cells in there, the complicated actions inside, which also have a beauty. I mean it's not just beauty at this dimension, at one centimeter; there's also beauty at smaller dimensions, the inner structure, also the processes. The fact that the colors in the flower evolved in order to attract insects to pollinate it is interesting; it means that insects can see the color. It adds a question: does this aesthetic sense also exist in the lower forms? Why is it aesthetic? All kinds of interesting questions which the science knowledge only adds to the excitement, the mystery and the awe of a flower. It only adds. I don't understand how it subtracts.”

— Richard P. Feynman

From the stirrings of quiet and disorder, a flower came to be.

To my husband and son — the gravity and the light that held me whole. To my parents (all four of them), family and friends —the roots, water and warmth that nourished every step. To my mentors and colleagues — the breeze and pollinators that shook my perspectives and helped me grow stronger. This work is so much more than what blooms on the surface; it carries the unseen beauty of all that made it possible.

Acknowledgements and funding

Tissue samples were collected by research nurses led by S. Fourie; I thank them, the patients, and the endoscopists in the Department of Gastroenterology, Oxford University Hospitals. I acknowledge the support of the Oxford University Hospitals NHS Foundation Trust Colorectal Surgical team and the TGU Investigators. I acknowledge support from Paul Sopp, Craig Waugh, and Sally Clark from the WIMM FACS Facility, Neil Ashley and Maria Greco from the WIMM Single Cell Facility, and Tim Rostron from the WIMM Sequencing Facility. Data from Smillie et al. (2019) was obtained from the Broad Data Use Oversight System (DUOS-000110) following institutional approval, and I thank and acknowledge the original authors and funders.

Funding for this study was generously provided by the Medical Research Council, the Wellcome Investigator Award (219523/Z/19/Z) and the MRC (A.S.). Diagrams shown in this work were generated using BioRender.com.

Abstract

The human intestinal epithelium is a rapidly renewing, functionally diverse tissue responsible for nutrient absorption, barrier defence, and immunological regulation. Its complexity — marked by continuous turnover and the presence of rare, specialised lineages — presents substantial challenges for experimental modelling. While human intestinal epithelial organoids (hIEOs) have emerged as physiologically relevant *ex vivo* systems, limitations persist in reliably inducing full lineage complexity, resolving rare or transitional cell types, and mapping lineage dynamics under perturbation. This thesis addresses these methodological gaps by developing, validating, and applying high-resolution, scalable single-cell and lineage tracing strategies to interrogate epithelial differentiation, plasticity, and rare cell emergence in hIEOs - with a particular focus on Best4 Cells, a recently characterised cell set within the absorptive lineage, potentially implicated in coordinating responses to luminal pH, osmolarity, or microbial metabolites.

To enable systematic and high-throughput analysis of epithelial identity across multiple experimental conditions, a multiplexed single-cell RNA sequencing (scRNA-seq) pipeline was developed. This platform integrated cell hashing with droplet-based scRNA-seq and supported the simultaneous profiling of over 40 conditions in a single experiment, including donor samples from colon and terminal ileum (TI), fetal and adult stages, and healthy and ulcerative colitis (UC) states. This methodological advance improved resolution, reduced batch effects, and significantly increased experimental scalability.

Applying this approach, three candidate organoid differentiation protocols — House, Fujii, and Holloway — were benchmarked side-by-side across multiple donor-derived organoid lines. Distinct media-specific biases in epithelial output were observed, with House media promoting balanced absorptive differentiation, Fujii enhancing proliferation and secretory lineages such as

Enteroendocrine (EEC) and Tuft cells, and Holloway supporting absorptive maturation and M-Cell priming. However, across all conditions, rare epithelial subtypes (particularly Best4$^+$ cells) remained poorly represented or transcriptionally ambiguous.

To investigate whether this reflected biological absence or technical under-detection, Smart-seq2 was applied to FACS-enriched EPCAM$^+$/BEST4$^+$ cells from both organoid and primary intestinal tissue. While primary-derived cells showed robust co-expression of *BEST4, CA7, SPIB*, and *MEIS1* — defining a coherent Best4$^+$ signature — organoid-derived cells failed to replicate this profile. This confirmed that current differentiation protocols do not support full commitment toward the Best4 lineage, despite expression of some surface markers. These findings validated the use of full-length transcriptomics for rare cell confirmation and established the *SPIB–CA7–BEST4–MEIS1* axis as a reliable transcriptional signature for Best4 cell identity.

To better capture transitional or epigenetically primed states that might not manifest at the transcriptomic level, a single-cell multiomic workflow (scMultiome-seq) was implemented. Following protocol optimisation — including the use of digitonin-based permeabilization and sample hashing compatibility — paired transcriptomic and chromatin accessibility data were obtained from eight donor-derived hIEO lines. While ATAC-based clustering alone lacked resolution due to sparse signal, integration with RNA profiles markedly improved identification of transitional and rare lineages. Chromatin accessibility at key loci associated with Best4 differentiation (e.g., *BEST4*, *MEIS1*, *SPIB)* was detected in absorptive progenitors, even in the absence of full transcriptional activation. This suggested that these lineages may be epigenetically primed but fail to commit due to insufficient niche cues or differentiation stimuli. Multiome data also improved the resolution of Tuft and EEC subtypes and revealed chromatin-level signatures consistent with known lineage-specific regulatory pathways, further validating the value of combining transcriptomic and epigenetic modalities.

To trace how lineage decisions unfold over time and under defined environmental cues, the CellTag lentiviral barcoding system was adapted for use in hIEOs. A triple-tagging scheme was developed to deliver three sequential CellTag libraries (V1–V3) during transitions through conditioning medium (CM), differentiation medium (DM), and cytokine stimulation (DM + RANKL). To facilitate efficient transduction, hIEOs were transitioned into 2D monolayers, increasing viral access to the apical surface. Despite some technical challenges — including cell loss from high viral load and reduced hashing efficiency — sufficient tag recovery enabled transcriptome-linked clonal reconstruction. Network-based visualisation of clonal relationships revealed lineage bifurcations and convergence patterns, including M-Cell and Goblet Cell emergence from shared progenitor pools. Cells tagged early (V1) contributed broadly to multiple lineages, while later-tagged cells (V3) were more fate-restricted, demonstrating temporal resolution of differentiation potential. Although mature Best4^{+} cells were not resolved in this dataset, the platform provided a scalable and modular framework for future fate-mapping studies in hIEOs.

Biological validation of these methods was carried out through a series of targeted case studies spanning anatomical regions (TI vs colon), developmental stages (fetal vs adult), and disease states (healthy vs inflamed UC). In all contexts, hIEOs retained key transcriptional features of their tissue of origin. For example, colon-derived organoids expressed *SPINK5*, *CFTR*, and *LEFTY1*, while ileal organoids expressed region-specific markers such as *CCL25*, *SLC10A2*, and *LYZ*. Fetal organoids exhibited elevated *OLFM4* and *CLDN10* expression, consistent with immature, progenitor-enriched states. Organoids derived from inflamed UC tissue retained transcriptional signatures of inflammation, including upregulation of *OLFM4* and *IFITM2*, and downregulation of absorptive markers such as *KRT20* and *FABP1*. These features were preserved across conditions and donors, affirming the biological fidelity of the model system. Furthermore, stimulation with RANKL led to reproducible induction of canonical M-cell markers *(SPIB, GP2, TNFAIP2)*, while IL13/IL4 promoted

secretory skewing and Tuft Cell expansion *(POU2F3, TRPM5, IL13RA2)*, demonstrating the responsiveness of hIEOs to immune-mediated differentiation cues.

The findings of this thesis establish that Best4$^+$ Cells are inconsistently recovered across current hIEO differentiation protocols and, when detected, remain transcriptionally immature, lacking co-expression of canonical markers such as *CA7, SPIB*, and *MEIS1*. However, integrative single-cell and multiomic analyses revealed chromatin accessibility at key Best4-associated loci, suggesting that this lineage may be epigenetically primed but transcriptionally inactive under standard culture conditions. These results indicate that the absence of fully specified Best4$^+$ Cells likely reflects an unmet requirement for additional inductive cues rather than a fundamental limitation of the hIEO platform itself. Recent work in the field has identified interferon signalling — particularly IFN-γ — as a potent driver of BEST4$^+$/CA7$^+$ lineage emergence, linking this absorptive subtype to immune regulation and host-pathogen defence. Together, these insights provide a compelling framework for future research by combining the methodological platforms developed here — such as scMultiome-seq and temporal CellTag tracing — with targeted interferon stimulation, which could illuminate the developmental origin, regulatory logic, and functional roles of this rare but biologically significant epithelial population.

In conclusion, this thesis presents a comprehensive and methodologically rigorous framework for studying human intestinal epithelial biology. By integrating multiplexed scRNA-seq, multiomic profiling, and clonal lineage tracing, it enables high-resolution dissection of differentiation pathways, rare lineage emergence, and the influence of niche signals. These methodological innovations offer validated solutions to longstanding technical challenges and provide a rich resource of datasets, protocols, and conceptual tools. Together, they establish hIEOs as powerful, physiologically relevant models for probing epithelial plasticity, and lay the groundwork for future

translational applications in regenerative medicine, infection biology, and personalised therapeutics.

Statement of contribution to work

I am aware of and understand the University's policy on plagiarism. I certify that the work presented in this thesis is entirely my own, unless otherwise stated or explicitly indicated by references. The work presented herein has not been submitted in support of another degree or qualification at this or any other university or institute of learning.

This work was conducted under the supervision of Professor Alison Simmons, with additional guidance provided by Dr Agne Antanaviciute and Dr Kaushal Parikh. I planned and carried out all experiments, either independently or with assistance - during large-scale experiments - from Dr Paulina Siejka-Zielińska or Dr Zoe Christoforidou. Experimental data were analysed and interpreted by myself. Computational analyses and data visualisation methods were performed by me, either developed independently or, where appropriate, by building upon prior algorithmic optimisations established by Dr Agne Antanaviciute, Dr Paulina Siejka-Zielińska, and Dr Rosana Babu. The interpretation of results and the writing of this thesis were undertaken entirely on my own.

Table of Contents

Acknowledgements and funding.....*iii*

Abstract.....*iv*

Statement of contribution to work.....*ix*

Table of Contents.....*x*

List of Abbreviations and Acronyms.....*xvi*

List of Tables.....*xxii*

List of Figures.....*xxiv*

1 Introduction*.....*1

1.1 Understanding Intestinal Homeostasis.....1

1.1.1 Cellular Diversity and Functional Organisation of the Intestinal Epithelium.....3

1.1.2 Epithelial Junctional Architecture and Barrier Function.....8

1.1.3 Molecular Regulation and Adaptive Capacity of the Intestinal Epithelium.....10

1.1.4 Intercellular Communication in the Intestinal Microenvironment.....13

1.1.5 Molecular Mechanisms Underpinning Barrier Dysfunction.....17

1.2 Simulating Intestinal Epithelial Dynamics using Organoid Cultures.....23

1.2.1 Overview of Organoid Culture Systems.....23

1.2.2 Organoids as Biomimetic Models of Health and Disease.....25

1.3 Analytical Methods for Characterising Human Intestinal Epithelial Organoids. .28

1.3.1 Functional Single-Cell Technologies: From Primary Tissue to Organoids.....28

1.3.2 Integrating Multi-Dimensional Functional Single-Cell Technologies to Advance Epithelial Biology Exploration.....32

1.3.3 Single-Cell Lineage Tracing of Epithelial Plasticity in Human Intestinal Organoids.....36

1.4 Overall Project Aims.....41

2 Materials and Methods.....44

2.1 Materials.....44

2.1.1 Consumables and Commercial Assays.....44

2.1.2 Chemicals, Peptides and Recombinant Proteins.....45

2.1.3 Solutions, Buffers and Media.....49

2.1.4 Cell Lines.....52

2.1.5 Recombinant DNA: CellTagging Libraries and Plasmids.....52

2.2 Methods.....53

2.2.1 Processing of Intestinal Samples.....53

2.2.2 Intestinal Epithelial Organoid Cultures.....54

2.2.3 Antibody Conjugation.....56

2.2.4 Antibody Staining.....56

2.2.5 Flow Cytometry and Cell Sorting.....57

2.2.6 DNA and RNA techniques.....57

2.2.7 Cell Tagging.....57

2.2.8 Single-Cell Sequencing Pipelines.....62

2.2.9 Computational Processing and Analyses.....65

3 Standardising Organoid Cultures to Model the Human Intestinal Epithelium.....75

3.1 Introduction.....75

3.2 Development of a Multiplexed Single-Cell RNA-seq Assay for High-Throughput Profiling of hIEO Cultures.....79

3.2.1 Mapping Epithelial Heterogeneity and Rare Lineages in hIEOs at Single-Cell Level.....79
3.2.2 Experimental Design: High-Throughput Evaluation of Differentiation Efficiency in hIEOs Across Multiple Conditions Using Multiplexed scRNA-seq.....80
3.2.3 Transcriptomic Profiling and Quality Control.....83
3.2.4 CM vs DM: Media-Specific Effects on Epithelial Composition.....85
3.2.5 Identification of Partial Best4 Cell States.....88
3.2.6 Methodological Benchmarking and Resolution Validation.....90
3.2.7 Interpretation of Findings.....92
3.3 Comparative Meta-Analysis of Published Organoid scRNA-seq Datasets.....94
3.3.1 Selection and Profiling of External Organoid Datasets.....94
3.3.2 Reprocessing, Clustering, and Identification of BEST4+ cells.....95
3.3.3 Integrated Transcriptomic Landscape of hIEO-derived Datasets.....97
3.3.4 Validation of Best4 Marker Expression Across Datasets.....99
3.3.5 Quantitative Benchmarking of Epithelial Lineage Representation.....101
3.3.6 Downsampled Integration for Dataset Size Normalisation.....102
3.3.7 Evaluation of Best4 Identity in Equalised Cell Populations.....106
3.3.8 Cross-Dataset Fidelity and Marker Expression in Best4 Cells.....108
3.3.9 Interpretation of Findings.....111
3.4 Technical Evaluation of hIEO Differentiation Protocols Using Single-Cell RNA Phenotyping: A Best4 Case Study.....113
3.4.1 Assessing House Protocol Performance via Targeted Sorting and Full-Length scRNA-Seq 113
3.4.2 Benchmarking SmartSeq2: Capture of BEST4^{+} Cells from hIEOs vs. Primary Epithelium. 116
3.4.3 Comparative Transcriptomics: Evaluating Lineage Resolution in hIEOs vs Primary Tissue 117

3.4.4 Surface Marker vs. Transcriptome Fidelity: Technical Implications for Lineage Annotation 120
3.4.5 Controlled Protocol Benchmarking via Multiplexed Droplet-Based scRNA-Seq.....121
3.4.6 Data Quality and Experimental Scope.....121
3.4.7 Protocol-Specific Performance in Supporting Rare Lineage Emergence.....122
3.4.8 Quantifying BEST4+ Resolution Across Protocols: A Technical Perspective.....127
3.4.9 Interpretation of Findings.....130
3.5 Chapter Summary.....133
4 Dissecting rare epithelial lineages with multiomic and lineage tracing frameworks 134
4.1 Introduction.....134
4.2 Integrative Multi-Omic Profiling to Overcome Transcriptomic Resolution Limitations.....136
4.2.1 Objective and Rationale: Enhancing Lineage Detection via Chromatin Accessibility and Transcriptomics.....136
4.2.2 Protocol Optimisation: Whole-Cell scMultiome via Digitonin vs. PFA.....137
4.2.3 Proof-of-Concept: Multiplexed scMultiome-seq Across Diverse hIEO Samples.....140
4.2.4 Benchmarking Modality Performance: RNA vs. ATAC vs. Integrated UMAP Clustering.....142
4.2.5 Evaluating Chromatin Cluster Fidelity Through Accessibility Metrics.....144
4.2.6 Chromatin Resolution Limits: Interpreting Sparse Accessibility in Transitional and Rare Epithelial States.....147
4.2.7 Functional Inference from Chromatin Accessibility: GO Enrichment Across Lineages....151
4.2.8 Validating Lineage Identity via Joint Transcriptomic and Epigenetic Signatures.....152
4.2.9 Validating Lineage Resolution Through Pseudotime Mapping of Integrated Multiomic Clusters.....154
4.2.10 Multiomic Insights into the Epigenetic and Transcriptomic Landscape of Best4 Cells. 156

4.2.11 Interpretation of Findings.....159

4.3 Engineering a Lineage Tracing Platform Using CellTagging in Epithelial Organoid Monolayers.....162

4.3.1 Rationale for Technological Approach.....162

4.3.2 Rationale For Adopting a 2D hIEO Conformation.....163

4.3.3 Design of a Preliminary Validation Strategy to Evaluate Biological Fidelity Between 2D and 3D Organoid Cultures.....164

4.3.4 Experimental Design: High-Throughput Single-Cell Characterisation of Different Culture Conditions in 2D vs 3D conformations.....164

4.3.5 Evaluating the Impact of 2D vs 3D Conformation in the Transcriptomic Signatures of hIEOs 167

4.3.6 Evaluating the Impact of Passage Number on the Transcriptomic Fidelity of hIEO Monolayers.....175

4.3.7 Generating CellTag Libraries.....181

4.3.8 Generating CellTag Lentiviral Preps and Optimising Infectivity in hIEOs.....183

4.3.9 Pilot Experiment 1: Assessing The Impact of Sample Storage Conditions on Cell Quality 185

4.3.10 Pilot Experiment 2: Assessing the Impact of Hashing Conditions and High Virus Titer on scRNA-seq Data Quality.....190

4.3.11 Proof-of-Concept Experiment: Multiplexed Lineage Tracing Platform in hIEOs.....196

4.3.12 Distribution of Tagged Cells Across CellTag Generations.....200

4.3.13 Clonal Complexity Across Epithelial Cell Types in hIEOs.....201

4.3.14 Network-Based Lineage Reconstruction Reveals Temporal Clonal Relationships.....203

4.3.15 Exploring the Feasibility of a Multiome Single Cell Sequencing Lineage Tracing Approach 206

4.3.16 Interpretation of Findings.....206

4.5 Chapter Summary.....210

5 Mapping Context-Specific Signatures in hIEO Models 212

5.1 Introduction 212

5.2 Contextual Influences on Epithelial Organoid Identity Across Multiplexed Experiments 214

5.2.1 Location, Inflammation, and Developmental Signatures in House-Protocol Cultured hIEOs 214

5.2.2 Regional Identity Retention in House-Protocol Cultured hIEOs 214

5.2.3 Preservation of Inflammatory Gene Expression Profiles in House-Protocol Cultured hIEOs 220

5.2.4 Maintenance of Developmental Stage-Specific Signatures in House-Protocol Cultured hIEOs 225

5.2.5 Differential Responses to Stimulation in House-Protocol Cultured hIEOs: NaB and RANKL as Divergent Modulators 231

5.2.6 Media-Dependent Variation in hIEO Responses to Location and Stimulation 244

5.2.7 Impact of IL13/IL4 Stimulation on the Emergence of Secretory Cell Populations 258

5.3 Integrated Transcriptomic and Epigenomic Profiling (Multiome) 261

5.4 Comparing Organoid Cell States with Primary Epithelial Phenotypes 271

5.5 Chapter Summary 276

6 Discussion 280

6.1 Rationale for Research 280

6.2 Summary of Findings 281

6.2.1 Multiplexed Single-cell RNA Sequencing Workflow 281

6.2.2 Comparative Meta-Analysis Of Organoid Single-Cell RNA Sequencing Datasets 283

6.2.3 Benchmarking Of Candidate Differentiation Protocols For Rare Cell Types 286

6.2.4 Integrating Transcriptomic And Epigenetic Modalities To Resolve Rare And Transitional Epithelial Cell States.....288
6.2.5 Development Of A Lineage Tracing Platform In Human Epithelial Organoids.....291
6.2.6 Biological Validation Of Human Epithelial Organoids As Models Of Intestinal Biology.....292
6.3 Relevance And Contributions Of The Current Research.....295
6.4 Late-Breaking Insights on BEST4/CA7^{+} Cells.....301
6.5 Implications For Future Research.....306
6.6 Final Summary.....310
7 References.....311

List of Abbreviations and Acronyms

2D/3D – Two/Three-Dimensional

ADTs - Antibody-Derived Tags

AJs - Adherens Junctions

APCs - Antigen-Presenting Cells

ARI- Adjusted Rand Index

ASAP-seq - Assay for Transposase-Accessible Chromatin with Select Antigen Profiling by sequencing

ASCs - Adult Stem Cells

Best4 -Bestrophin-4

Best4 Cells -Bestrophin-4-expressing Cells

BMP - Bone Morphogenic Protein

B2M - Beta-2 Microglobulin

cAMP - Cyclic Adenosine Monophosphate

CD – Crohn's Disease

cDNA - Complementary DNA

CFUs - Colony-Forming Units

CITE-seq - Cellular Indexing of Transcriptomes and Epitopes by sequencing

CLR - Centred Log-Ratio (transformation)

CM – Conditioning Medium

CRC - Colorectal Cancer

DM – Differentiation Medium

ECM - Extracellular Matrix

EECs - Enteroendocrine Cells

EGF(R) - Epidermal Growth Factor (Receptors)

EMP - Epithelial–Mesenchymal Plasticity

EMT - Epithelial-Mesenchymal Transition

ESCs - Embryonic Stem Cells

FACS - Fluorescence-Activated Cell Sorting

FAK - Focal Adhesion Kinase

FGF2 - Fibroblast Growth Factor 2

FMO - Fluorescence Minus One

FMT - Faecal Microbiota Transplantation

FST - Follistatin

GALT – Gut-Associated Lymphoid Tissue

GLP-1 - Glucagon-Like Peptide-1

GPs - Gap Junctions

GPCRs - G-Protein Coupled Receptors

Grem1 - Gremlin1

GO – Gene Ontology

HAP1 - Huntingtin-Associated Protein 1

HC – healthy colon

HDAC - Histone Deacetylase

Hh – Hedgehog (pathway)

hIEOs - Human Intestinal Epithelial Organoids

HTOs - Hashtag Nucleotides

IBD - Inflammatory Bowel Disease

IBS - Irritable Bowel Syndrome

IFN-γ - Interferon Gamma

IGF-1 - Insulin-like Growth Factor 1

IL - Interleukin

ILC2s - Innate Lymphoid Cells

INFLAREs – Inflammatory Epithelial Cells

iPSCs - Induced Pluripotent Stem Cells

ISCs – Intestinal Stem Cells

JAMs - Junctional Adhesion Molecules

Jag – Jagged (ligands)

kNN - k-nearest neighbour

LBPs - Live Biotherapeutics Products

lncRNAs - long non-coding RNAs

logFC - log2 Fold Changes

LR - Ligand-Receptor

M-Cells - Microfold Cells

MLCK - Myosin Light-Chain Kinase

MOI - Multiplicity Of Infection

mtDNA – Mitochondrial DNA

NaB – Sodium Butyrate

NASH - Non-Alcoholic Steatohepatitis

NEAA - Non-Essential Amino Acids

NRG1 - Neuregulin 1

Nog - Noggin

PCA - Principal Component Analysis

pcw – post-conceptional weeks

PDZ - post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (zo-1) (domain)

PFA - paraformaldehyde

PPs - Peyer's Patches

PRRs - Pattern Recognition Receptors

PSCs - Pluripotent Stem Cells

PDOs - Patient-Derived Organoids

RANKL – Receptor Activator of Nuclear Factor κB Ligand

RNI – RNase Inhibitor

ROCK - Rho-associated protein kinase

ROS – Reactive oxygen species

Rspo-1 - R-Spondin 1

scATAC-seq - Single-Cell Assay for Transposase-Accessible Chromatin Sequencing

scMultiome-seq - Single-Cell Multiome Sequencing

scRNA-seq - Single-Cell RNA Sequencing

SCFAs - Short-Chain Fatty Acids

smLiveFISH - single-molecule live-cell fluorescence in situ hybridization

SNN - Shared Nearest Neighbour

SPIB – Spi-B Tanscription Factor

SOST- Sclerostin

TAs - Transit-Amplifying Cells

TAZ - Transcriptional Co-Activator with PDZ-Binding Motif

TE – Transformation Efficiency

TEER - Transepithelial Electrical Resistance (measurements)

TF – Transcription Factor

TGF-β - Transforming Growth Factor Beta

TI – Terminal Ileum

TJs - Tight Junctions

TLRs - Toll-Like Receptors

TNF(-α) - Tumour Necrosis Factor(-Alpha)

TSS – Transcriptional Starting Site

YAP - Yes-Associated Protein

UC – Ulcerative Colitis

UMAP - Uniform Manifold Approximation and Projection for Dimension Reduction

UMI – Unique Molecule Identifier

YAP – Yes-associated protein

WNN - Weighted Nearest Neighbours

Wnt - Wingless-related integration site (pathway)

ZA - *Zonula Adhaerens*

ZO - *Zonula Occludens*

List of Tables

Table 1- Consumables and critical commercially available assays used in this study.....44

Table 2 - Reagents used in this study.....45

Table 3 - Table of antibodies used in this study.....47

Table 4 - Table of oligos used in this study.....48

Table 5 - List of reagents for the preparation of Transport Media.....49

Table 6 - List of reagents for the preparation of HPGA Buffer.....49

Table 7 - List of reagents for the preparation of Chelation Media.....49

Table 8 - List of reagents for the preparation of Staining Buffer.....49

Table 9 - List of reagents for the preparation of “Fujii” basal media, as per Fujii M. et. al (2018).....49

Table 10 - List of reagents for the preparation of “Fujii” conditioning media, as per Fujii M. et. al (2018).....50

Table 11 - List of reagents for the preparation of “Fujii” differentiation media, as per Fujii M. et. al (2018).....50

Table 12 - Preparation of “House” Conditioning Media, adapted from Sato T et. al (2011).....50

Table 13 - List of reagents for the preparation of “House” Differentiation Media, adapted from Sato T et. al (2011).....51

Table 14 - List of reagents for the preparation of “Holloway” conditioning media, as per Holloway E. et al (2021).....51

Table 15 - List of reagents for the preparation of “Holloway” differentiation media, adapted from Holloway E. et al (2021).....51

Table 16 - List of reagents for the preparation of Lysis Buffer (with Digitonin).....51

Table 17 - List of reagents for the preparation of Washing Buffer (after Digitonin Lysis).....52

Table 18 - List of reagents for the preparation of Virus Production Medium.....52
Table 19 - List of cell lines used in this study.....52
Table 20 - List of recombinant DNA sources used in this study.....52
Table 21 - List of reagents for the preparation of “Fujii” vs “Want” conditioning media, as per Fujii et. al (2018) and Wang et. al (2025).....303
Table 22 - List of reagents for the preparation of “Fujii” vs “Want” differentiation media, as per Fujii et. al (2018) and Wang et. al (2025).....304

List of Figures

Figure 1 - Multiplexed scRNA-seq workflow for high-throughput evaluation of differentiation efficiency and system perturbation effects in hIEOs..... 81

Figure 2 - Experimental design for Experiment 1: High-throughput screening of organoid differentiation and response to RANKL and Sodium Butyrate stimulation across distinct developmental stages and patient-derived samples..... 82

Figure 3 – Overview of relevant experimental metrics: Uniform Manifold Approximation and Projection (UMAP) as a function of distinct experimental variables, multiplexing efficiency and other key quality control metrics from the pilot single-cell multiplexed sequencing run of hIEOs…84

Figure 4 – Overview of the transcriptional landscape of differentiated versus undifferentiated hIEOs cultured in House medium..... 87

Figure 5 - Expression profile of BEST4, SPIB, and a combined Best4 Cell Score across the whole CM+DM dataset..... 89

Figure 6 - Impact of dataset downsampling on cell clustering and annotation in an intestinal epithelial datasets (Elmentaite et al., 2020)..... 91

Figure 7 - Meta-analysis of scRNA-seq data from independent, publicly available organoid datasets 96

Figure 8 - Integration and annotation of epithelial cell types across harmonised intestinal datasets 98

Figure 9 - Molecular and spatial characterisation of Best4 Cells within the intestinal epithelium...100

Figure 10 - Relative proportions of epithelial cell types across datasets, after integration.....102

Figure 11 – Number of cells contributed per dataset after integration.....103

Figure 12 – Positioning of the Best4 Cells cluster within the integrated dataset after size normalisation..... 104
Figure 13 – Cell clustering and prevalence after integration and downsampling, by dataset of origin 105
Figure 14 – Detailed transcriptomic profiling of Best4 cells in the downsampled, integrated dataset 107
Figure 15 – Best4 cell abundance and transcriptomic fidelity across datasets after integration and dataset balancing..... 109
Figure 16 - Comparative expression of Best4+ cell markers across datasets.....111
Figure 17 – Experimental design for Experiment 2: isolation of epithelial cells from either hIEOs cultured under the House protocol, or matched primary tissue-derived cells, for full-length transcriptomic profiling..... 115
Figure 18 – Best4 Cell distribution across both primary tissue and hIEO cultures.....117
Figure 19 – Transcriptional signature of $BEST4^{+}$and $BEST4^{-}$ sorted cells across primary intestinal cells and hIEOs..... 119
Figure 20 - Experimental design for Experiment 3: multiplexed scRNA-seq assay across different growth protocols, anatomical locations, and disease contexts.....122
Figure 21 - UMAP visualisation of epithelial cell types across the three different hIEO protocols included in this dataset..... 123
Figure 22 - Comparison of epithelial cell type distributions across the three hIEO protocols.....126
Figure 23 - Distribution and transcriptional fidelity of BEST4+ cells across culture conditions.....128
Figure 24 - Comparison of RNA, ATAC, and HTO signal quality between Digitonin and PFA fixation protocols in the scMultiome-seq pilot experiment on whole hIEO cells.....139

Figure 25 – Experimental design for Experiment 5: scMultiome-seq profiling of multiple hIEO conditions using the Fujii protocol..... 141
Figure 26 - Demultiplexing of scMultiome-seq samples using antibody-derived hashtags (HTOs). 142
Figure 27 - UMAP visualisations of RNA, chromatin accessibility (ATAC), and integrated modalities from scMultiome-seq intestinal epithelial dataset..... 144
Figure 28 - ATAC-seq quality control metrics across epithelial cell clusters..... 146
Figure 29 – Differentially accessible genomic loci identified across chromatin-defined clusters.... 148
Figure 30 - Chromatin priming at the HAP1 locus in Tuft Cells and TAs..... 150
Figure 31 – GO analysis of differentially accessible genes in chromatin-based clusters..... 152
Figure 32 - Expression of selected population-specific signature genes across intestinal epithelial cell types..... 153
Figure 33 – Pseudotime trajectories of absorptive and secretory epithelial lineages..... 156
Figure 34 - Gene expression and chromatin accessibility of Best4 markers in epithelial cell populations..... 158
Figure 35 – Experimental layout for Experiment 6: stimulation assays across multiple culture conditions and sample types..... 166
Figure 36 – Integrated cell clustering across experimental conditions and samples..... 167
Figure 37 - PCA of transcriptomic variation as a function of hIEO culture conformation..... 168
Figure 38 - UMAP visualisations of epithelial cell type annotation and culture format distribution 170
Figure 39 – Volcano plot showing differential gene expression between 3D and 2D epithelial organoid cultures..... 172
Figure 40 – Gene-wise logFC correlation of organoid culture conformation vs primary epithelium (3D: primary cells vs 2D: primary cells)..... 173
Figure 41 – GO enrichment of biological processes in 2D vs 3D conformations..... 174

Figure 42 - PCA of transcriptomic variation as a function of the passage number.....176
Figure 43 - Cell type composition and lineage distribution in epithelial cells across the passage subset.....178
Figure 44 - UMAP of the passage subset grouped by passage number.....179
Figure 45 - Diversity/dissimilarity metrics for epithelial cell type composition across organoid passages.....180
Figure 46 - CellTag library generation and benchmarking across V1, V2, and V3 constructs.....182
Figure 47 - Comparison of transduction efficiency before and after viral ultracentrifugation across CellTag libraries.....184
Figure 48 - Experimental setup for Experiment 7: assessment of impact of storage conditions on CellTag libraries, cell hashing, and RNA content.....187
Figure 49 - Demultiplexing and quality control of a 5-sample pooled scRNA-seq experiment using HTO hashing.....189
Figure 50 - Experimental setup for Experiment 8: assessing suitability of sequential CellTag barcoding and condition-specific stimulation across multiplexed reactions.....192
Figure 51 - Comparison of lineage tracing outcomes between multiplexed and single-sample CellTag experiments.....195
Figure 52 - Experimental design for Experiment 9: sequential CellTag barcoding and stimulation conditions in control and test monolayer samples.....197
Figure 53 - Multidimensional analysis of epithelial identity, culture condition, and CellTag history in human intestinal epithelial organoids.....199
Figure 54 - Distribution of tagged cells across CellTag generations V1–V3.....200
Figure 55 - Clonal complexity across epithelial cell types in hIEOs.....202

Figure 56 - Network visualisation of lineage relationships across temporal CellTag barcodes in hIEOs..... 205
Figure 57 - Subsetting strategy and overall transcriptional profiles for regional identity analysis. .216
Figure 58 - Regional gene expression differences between terminal ileum and colon in unstimulated hIEOs..... 218
Figure 59 - GO term enrichment analysis of region-specific gene expression in adult intestinal epithelial organoids..... 220
Figure 60 - Transcriptional divergence between healthy and ulcerative colitis–inflamed adult colon epithelial organoids..... 221
Figure 61 - Transcriptomic changes in inflamed ulcerative colitis (UCI) epithelium compared to healthy colon (HC)..... 223
Figure 62 - GO enrichment analysis of differentially expressed genes in healthy control (HC) and ulcerative colitis inflamed (UCI) epithelial samples..... 224
Figure 63 - Transcriptional differences between fetal and adult intestinal epithelial organoids.....226
Figure 64 - Transcriptomic comparison between fetal and adult human colonic epithelium.....228
Figure 65 - GO enrichment analysis of differentially expressed genes in fetal versus adult human colonic epithelium..... 230
Figure 66 - UMAP projections showing transcriptional responses to sodium butyrate (NaB) and RANKL stimulation in hIEOs..... 231
Figure 67 - Transcriptomic effects of RANKL stimulation on hIEOs..... 233
Figure 68 - GO enrichment analysis of genes upregulated in RANKL-stimulated versus unstimulated intestinal epithelial organoids..... 235
Figure 69 - UMAP clustering and gene expression profiles of RANKL-stimulated epithelial cells showing marker expression across transcriptional clusters..... 237

Figure 70 - UMAP projection and differential gene expression analysis of SPIB⁻ cells within RANKL-stimulated epithelial monolayers..... 239
Figure 71 - Network-based visualisation of enriched Gene Ontology terms and associated genes in RANKL-stimulated epithelial cells..... 241
Figure 72 - Enriched GO terms among upregulated genes in RANKL-stimulated epithelial cells.....243
Figure 73 - Compositional differences in epithelial cell types across media conditions, intestinal regions, and disease states..... 245
Figure 74 - Comparative analysis of gene expression in hIEOs across different media.....253
Figure 75 - Effect of medium, intestinal region, and inflammation on epithelial lineage composition in colon organoids subject to RANKL stimulation..... 256
Figure 76 - Epithelial cell states following IL13/IL4 stimulation versus unstimulated controls.....258
Figure 77 - Transcriptomic and pathway-level responses of epithelial organoids to IL13/IL4 stimulation..... 260
Figure 78 - Differential gene expression in inflamed UC versus healthy epithelial samples.....262
Figure 79 - Gene ontology enrichment of biological processes altered in ulcerative colitis inflammation..... 264
Figure 80 - Comparative transcriptomic and chromatin accessibility changes in inflamed versus non-inflamed UC epithelium relative to healthy controls.....267
Figure 81 - Multiomic analysis of TNFSF15 gene expression and regulatory accessibility in UC-derived epithelial organoids..... 269
Figure 82 - Epithelial cell type annotation and culture format distribution.....273
Figure 83 - Expression of Best4-Cell associated markers in integrated hIEO and primary tissue dataset..... 275

1 Introduction

1.1 Understanding Intestinal Homeostasis

The intestine is a complex and multifunctional organ, uniquely adapted to manage a diverse set of physiological tasks critical to survival. Extending from the stomach to the rectum, it forms a highly dynamic interface between the external environment and the internal milieu. As the principal site for nutrient absorption, waste excretion, and immune sampling, the intestine is continuously exposed to dietary antigens, microbial communities, and environmental signals[1–5]. Thus, to carry out its functions while maintaining systemic equilibrium, it must coordinate structural compartmentalisation with functional flexibility.

At the core of this capacity is the intestine's intricate histological organisation. The intestinal wall is composed of multiple, highly specialised layers that operate in concert to integrate digestion, motility, immunity, and barrier function. The innermost layer — the mucosa — hosts the intestinal epithelium, a monolayer of cells that forms the primary physical and immunological interface with the luminal contents[6–10]. Just beneath it lies the lamina propria, a connective tissue matrix enriched with immune and stromal cells, capillaries, and lymphatics, facilitating bidirectional signalling between the epithelium and underlying immune networks[6,11–13]. Deeper still, the muscularis mucosae modulates mucosal motion, while the submucosa contains neural and vascular elements critical for secretory and circulatory regulation[9]. The muscularis externa and the outer serosa provide motility and mechanical support, completing a multilayered structure finely tuned for spatial and functional specialisation[9].

Finally, the enteric nervous system (ENS) forms a highly specialised and semi-autonomous neural network embedded throughout the gastrointestinal tract, comprising myenteric and

submucosal plexuses that collectively regulate motility, secretion, vascular tone, and epithelial homeostasis.[8, 41, 47, 66] Often referred to as the “second brain”, the ENS contains an extensive repertoire of neurons and glial cells that interact intimately with the intestinal epithelium, immune system, and resident microbiota to coordinate local reflexes independently of central nervous system input. ENS-derived signals are increasingly recognised as key regulators of epithelial proliferation, differentiation, and barrier function, mediated through neuropeptides, neurotransmitters, and direct neuro–epithelial contact. Moreover, developmental and functional defects in the ENS contribute to diverse intestinal pathologies, including Hirschsprung disease, inflammatory bowel disease, and dysmotility syndromes, underscoring its central role in maintaining intestinal physiology. Recent single-cell and spatial transcriptomic studies have further revealed unexpected cellular diversity and dynamic neuromodulatory circuits within the ENS, highlighting its integration into broader epithelial and stromal regulatory networks relevant to organoid biology and intestinal regeneration.

Overall, this layered architecture supports not only structural integrity but also the maintenance of intestinal homeostasis — a dynamic, balanced state in which physiological function is preserved despite continuous exposure to perturbing agents. Homeostasis arises from coordinated interactions among epithelial, immune, and stromal cells, as well as with the resident gut microbiota[4,14–16]. These interactions span the luminal, mucosal, and submucosal compartments, enabling the intestine to sustain barrier function, facilitate immune tolerance, and mount rapid responses to injury or infection[3]. Disruptions to this delicate equilibrium can drive inflammation, infection, or chronic disorders such as inflammatory bowel disease (IBD) and colorectal cancer (CRC), underlining the necessity of understanding the cellular and molecular mechanisms that underpin intestinal homeostasis[12,17–21].

1.1.1 Cellular Diversity and Functional Organisation of the Intestinal Epithelium

The intestinal epithelium is a dynamic, highly specialised single-cell layer tissue that forms the primary interface between the host and the external environment across the gastrointestinal tract. It performs essential functions, including nutrient absorption, immune surveillance, and barrier maintenance, while responding adaptively to a constantly fluctuating luminal landscape[3,22,23]. To achieve this, epithelial cells engage in continuous crosstalk with immune and stromal cells of the lamina propria, as well as with resident microbial communities[17,18,24,25]. Collectively, this triad modulates epithelial integrity and function, shaping tissue homeostasis and defence.

Functionally, the epithelium acts as a selective barrier that facilitates nutrient and electrolyte uptake while restraining pathogenic microorganisms and antigens[5,26]. Tight junctions between epithelial cells, together with antimicrobial peptides and mucus secretion, contribute to this containment[27–32]. The gut-associated lymphoid tissue (GALT) — including Peyer's Patches (PPs), isolated lymphoid follicles, and lamina propria lymphocytes — integrates local antigen sampling with systemic immune response[33–35]. In parallel, the gut microbiota exerts profound effects on epithelial maturation and function, producing metabolites such as short-chain fatty acids (SCFAs) that reinforce barrier integrity and modulate epithelial gene expression[33,35–37].

1.1.1.1 Crypt-Villus Architecture and Stem Cell-Driven Renewal

The epithelium is structurally compartmentalised into crypt-villus units in the small intestine and crypt-surface domains in the colon[38–40]. These units host multiple functionally distinct epithelial lineages whose distribution and composition vary along the proximal-distal and crypt–luminal axes[10,41,42]. At the base of each crypt resides a pool of multipotent, actively cycling LGR5^{+} Intestinal Stem Cells (ISCs), which sustain rapid epithelial turnover — approximately every 4–5 days — through tightly regulated programmes of proliferation, differentiation, and

apoptosis[38,39,43,44]. Following ISC division, transit-amplifying (TA) progenitors undergo lineage specification into absorptive or secretory trajectories, driven by graded transcriptional and epigenetic cues[38,41,45].

1.1.1.2 Absorptive Lineages and Specialised Variants: Enterocytes vs Best4 Cells and M-Cells

Absorptive progenitors differentiate primarily into Enterocytes, specialised in nutrient absorption, but can also give rise to rarer epithelial cell types, such as Microfold Cells (M-Cells) and Bestrophin-4–expressing cells (Best4 Cells, or Best4)[19,38,46,47]. Enterocytes constitute the most abundant epithelial cell population in the intestinal lining. While their primary function is nutrient absorption, they also contribute critically to barrier integrity, forming the first line of defence against luminal pathogens and toxins[48]. These cells are characterised by an apical brush border composed of densely packed microvilli, which dramatically increase the epithelial surface area, thereby enhancing both absorptive efficiency and microbial defence[39,49]. The microvilli are ensheathed by a glycocalyx rich in digestive enzymes and transporter proteins, facilitating nutrient breakdown and uptake[50]. Enterocyte maturation follows a well-defined differentiation gradient along the crypt–villus (small intestine) or along the crypt-surface (colon) axes, evolving from immature to fully competent enterocytes in a terminally differentiated state[51].

M-Cells are a specialised epithelial lineage primarily located in the follicle-associated epithelium overlying Peyer's patches in the small intestine[52–54]. These cells play an essential role in mucosal immunity by transcytosing luminal antigens, particulates, and microbes to underlying Antigen-Presenting Cells (APCs) such as dendritic cells and macrophages[52,55]. By contrast with Enterocytes, M-Cells possess a reduced brush border and modified apical surface, optimised for sampling and transport rather than absorption. Transcriptional profiling has established that M-Cells arise from LGR5$^+$ ISCs, akin to absorptive and secretory lineages[52–

[54,56,57]. However, their differentiation is primarily governed by stromal cues from the subepithelial dome — most notably the cytokine RANKL (Receptor Activator of Nuclear Factor κB Ligand), which activates NF-κB signalling and induces expression of the transcription factor Spi-B *(SPIB)*[58]. Although morphologically and functionally distinct, M-Cells share elements of the Enterocyte gene programme, suggesting they represent a specialised adaptation of this lineage, repurposed to support immune surveillance rather than nutrient uptake[54].

Best4 Cells represent a recently discovered epithelial population, identified through high-resolution single-cell transcriptomic profiling, particularly in the human colon[41,47]. These cells are characterised by a distinct genetic signature, including co-expression of *BEST4*, which encodes a calcium-activated chloride channel, and *SPIB*, traditionally associated with immune-modulatory epithelial programmes[38,41,47]. They also frequently express CA7 - a zinc metalloenzyme that catalyses the reversible hydration of carbon dioxide - and OTOP2 - a proton channel implicated in luminal pH sensing - collectively suggesting a specialised role in ionic regulation and chemosensory homeostasis at the luminal surface of colonic crypts[38,41,47]. A recent publication - based on transcriptomic studies in rat models - has identified *MEIS1* as a selectively enriched transcription factor in these populations, implicated in the regulation of *CFTR* and other sensory machinery[59]. This supports a model in which *MEIS1* contributes to the transcriptional programming that underpins the pH-sensitive and sensory features of Best4 Cells. Its consistent co-expression with *SPIB* and *BEST4* across both primary and *in vitro* datasets underscores its utility as both a marker and a potential regulator of this lineage. However, the physiological function of Best4 remains incompletely defined. Unlike established secretory or absorptive populations, they appear to operate at the interface of epithelial sensing and local microenvironmental modulation, potentially coordinating responses to luminal pH, osmolarity, or microbial metabolites[59]. Preliminary studies suggest their numbers

and transcriptional states may be altered in pathological conditions such as IBD and CRC, implicating them in disease-associated epithelial remodelling. Notably, Best4 partially express an absorptive gene programme, consistent with a lineage origin from $LGR5^+$ ISCs and shared developmental pathways with Enterocytes. However, their unique gene signature, functional markers, and spatial restriction supports their classification as a distinct lineage with potential region-specific roles. $BEST4^+$ cell development appears to be governed by canonical signalling pathways, though the precise niche-derived factors required for their maintenance remain to be elucidated. Like M-Cells, they exemplify the functional plasticity of the intestinal epithelium: both arise from common ISC progenitors, share partial absorptive gene expression, and have diverged to perform specialised, non-absorptive roles within the intestinal milieu[59].

1.1.1.3 Secretory Lineages: Goblet, Paneth, Enteroendocrine, and Tuft Cells

In parallel, secretory progenitors give rise to lineages involved in highly specialised functions, such as mucus production, antimicrobial defence, and hormonal regulation[38,60,61]. Goblet Cells, interspersed among Enterocytes, secrete mucins that form a protective mucus layer, shielding the epithelium from microbial insults and mechanical abrasion. This physical barrier traps pathogens and particulate matter, preventing them from reaching epithelial cells and underlying tissues. Paneth cells, located at the base of small intestinal crypts, secrete antimicrobial peptides such as lysozyme and α-defensins, thereby shaping the intestinal microbiota and neutralising pathogens[30,62,63]. They are also critical for ISC maintenance, by secreting a range of signalling molecules involved in their proliferation and fate decisions[56,57]. Enteroendocrine Cells (EECs), although very sparse, play a vital role in modulating gut physiology by secreting hormones like Glucagon-Like Peptide-1 (GLP-1) and serotonin, which regulate various aspects of digestion, appetite, and gut motility[26,64–66]. They are also key in luminal sensing activities - G-protein-coupled receptors (GPCRs) are an example of EEC

receptor proteins that detect nutrients and microbial metabolites, leading to the release of hormones that regulate digestion, appetite, and insulin secretion[67]. Tuft cells, another rare epithelial lineage regulated by SPIB, function as chemosensory sentinels that detect parasitic infections via taste receptor signalling[68–72]. Upon activation, they secrete cytokines such as IL-25, which initiate downstream type 2 immune responses through the activation of group 2 innate lymphoid cells (ILC2s) and associated signalling cascades[70].

1.1.1.4 Epithelial Cell Specialisation and Rare Lineage Phenotypes

Epithelial populations all originate from a common pool of $LGR5^+$ intestinal stem cells, undergoing tightly regulated differentiation programmes in response to spatial, temporal, and environmental cues[44,56,73]. Through a finely orchestrated balance of lineage specification and transcriptional regulation, the intestinal epithelium generates a functionally diverse cellular landscape that supports nutrient absorption, immune defence, hormonal signalling, and barrier maintenance, all of which are essential for intestinal homeostasis[43]. While much is known about the major lineages, increasing attention is now being directed towards rarer epithelial populations such as M-Cells, Best4 Cells, EECs, and Tuft cells. In the human intestine, however, these rare cell types remain under-characterised relative to animal models, primarily due to their low abundance and the technical challenges of isolating and studying them. This has required the use of advanced *in vivo* systems, lineage tracing, and emerging single-cell technologies. Notably, despite their distinct morphological and functional roles, Best4 Cells, M-Cells, and Tuft cells all share expression of the transcription factor SPIB, suggesting a convergent regulatory signature that may aid in their identification and manipulation[58,59,71]. Their divergent developmental trajectories, nonetheless, underscore the complex specialisation pathways of the intestinal epithelium and reinforce the need for robust, human-relevant models to accurately recapitulate their differentiation and function. Advancing our

knowledge of these rare lineages will require the integration of robust *in vitro* systems and high-resolution technologies such as single-cell transcriptomics, which are indispensable for decoding their developmental trajectories and functional roles in both physiological and pathological contexts.

Having delineated the diverse cellular landscape of the intestinal epithelium, the subsequent section will explore how these cells collectively maintain epithelial integrity and ensure effective barrier function. This structural and functional cohesion is essential for protecting the host from luminal antigens, pathogens, and toxins, and relies heavily on specialised intercellular junctions that regulate paracellular permeability and mechanical stability.

1.1.2 Epithelial Junctional Architecture and Barrier Function

The integrity of the intestinal epithelium is maintained by cell junctions - namely tight junctions (TJs), adherens junctions (AJs), gap junctions (GPs) and desmosomes - which regulate paracellular permeability and ensure a robust barrier against pathogens and toxins[3,28,74].

TJs, or *zonulae occludentes*, as the most apical junctions, are paramount in regulating paracellular permeability[28,75]. TJs are characterised structurally as semipermeable mesh-like networks surrounding epithelial cells, composed of different transmembrane and cytosolic building blocks which interact with each other to form a complex architecture[27,29]. These include the transmembrane proteins claudins, occludin, and junctional adhesion molecules (JAMs), which are intracellularly linked to scaffold proteins known as *zonula occludens* (ZO) proteins that anchor the complex to the actin cytoskeleton[27]. TJs form a seal at the apical region between adjacent epithelial cells and are therefore critical in regulating selective paracellular permeability — that is, the passage of ions, solutes, and water between cells - to

ensure adequate nutrient absorption while preventing uncontrolled passage of harmful substances across the epithelial sheet[75].

AJs are typically located just basal to TJs in the apical lateral membrane and are primarily composed of cadherin proteins - like E-cadherin - which mediate calcium-dependent, homophilic interactions between adjacent cells[74]. Intracellularly, E-cadherin binds to catenins which link the adhesion complex to the actin cytoskeleton, facilitating both mechanical stability and dynamic remodelling of cell–cell contacts[76,77]. AJs maintain mechanical intercellular adhesion and are critical for establishing tissue architecture and morphogenesis, maintaining apico-basal polarity, and transmitting mechanical signals.

GJs provide direct cytoplasmic continuity between neighbouring cells, allowing passage of small molecules (typically <1 kDa), ions, second messengers (e.g., Ca^{2+}, cyclic adenosine monophosphate, “cAMP”), and small metabolites[78]. They consist of connexin proteins, which assemble into hexameric hemichannels (connexons) in each cell membrane. When connexons from adjacent cells dock, they form an intercellular channel[78,79].

Desmosomes are specialised intercellular junctions that provide robust mechanical adhesion between epithelial cells, particularly in tissues subject to significant shear or tensile stress, such as the intestinal epithelium[80–82]. Unlike TJs and AJs, which are primarily associated with the actin cytoskeleton, desmosomes are anchored to intermediate filaments — predominantly keratins — via intracellular proteins[80]. This arrangement forms a dense adhesive structure that confers tensile strength and ensures structural cohesion of the epithelial sheet during cellular turnover, peristaltic motion, and inflammatory challenge[83,84].

In summary, beyond cellular composition, the epithelial barrier relies on highly organised junctional complexes that preserve structural cohesion and regulate paracellular permeability. TJs, AJs, GJs, and desmosomes coordinate to maintain epithelial integrity in the face of

constant renewal, luminal stress, and microbial exposure. Nonetheless, having explored the cellular diversity and architectural organisation of the intestinal epithelium, it becomes clear that its ability to maintain integrity and perform complex functions hinges on more than structural components alone. The epithelium is inherently responsive and adaptable, constantly integrating cues from its environment to regulate cell fate, coordinate repair, and preserve homeostasis. This capacity is orchestrated by a network of evolutionarily conserved signalling pathways. The next section will examine the molecular regulators and adaptive programs that underpin epithelial plasticity, renewal, and functional modulation in response to physiological and pathological stimuli.

1.1.3 Molecular Regulation and Adaptive Capacity of the Intestinal Epithelium

The intestinal epithelium sustains extraordinary regenerative and adaptive capacity, driven by a rapid 4–5-day cell turnover and guided by a layered signalling network integrated with its niche[85,86]. Orchestration of cell fate and renewal involves several key regulatory mechanisms, as well as interactions with the surrounding environment - including mesenchymal cells, immune cells, and the extracellular matrix (ECM)[56,86,87]. At the core of epithelial renewal are three evolutionarily conserved pathways: Wnt (Wingless-related integration site)/β-catenin, Notch, and Bone Morphogenetic Protein (BMP)[61,88–90]. However, the integrity and adaptability of the epithelial landscape are further shaped by additional signalling axes and intrinsic cellular properties - such as epithelial plasticity - which confer the capacity to respond to environmental perturbations and injury[91,92].

1.1.3.1 Canonical Signalling (Wnt, Notch, BMP)

The key signalling pathways involved in coordinating maintenance, differentiation, apoptosis, and regeneration of the intestinal epithelium at a molecular level include Wnt/β-catenin,

Notch and BMP)[61,88–90]. The Wnt/β-catenin pathway, modulated by various extracellular and intracellular factors, maintains ISC proliferation and self-renewal at the crypt base, ensuring continuous and controlled epithelial turnover[88,93,94]. Activation of Wnt signalling promotes expression of target genes that drive proliferation and prevent differentiation[39]. In contrast to Wnt signalling, the Notch signalling pathway primarily governs cell fate decisions within the intestinal epithelium[95,96]. Notch signalling interacts with Wnt signalling to coordinate the complex dynamics of epithelial renewal and is essential for determining how progenitor cells differentiate: high Notch activity typically favours the generation of absorptive enterocytes, while inhibition of Notch signalling leads to an increase in secretory cell types[89,90]. Balancing proliferative signals from Wnt and differentiation cues from Notch is the BMP signalling pathway[61,63]. BMP signalling acts as a negative regulator of ISC proliferation, primarily restricting expansion of the stem cell pool and promoting differentiation as cells migrate upwards from the crypt base[61]. BMP signals are modulated by an increasing activity gradient established from the crypt base to the crypt-top (colon) or villus tip (small intestine), which encourages cell differentiation and limits the potential for uncontrolled cell growth, thus ensuring proper spatial organization of the epithelium[63,97]. Together, Wnt-Notch-BMP balance ensures crypt spatial organisation and homeostatic proliferation.

1.1.3.2 Auxiliary Signalling (Hedgehog, TGF-β, Hippo/YAP)

Beyond canonical regulators, other pathways also play an important role in maintaining intestinal homeostasis[98–100]. The Hedgehog (Hh) pathway contributes to the regulation of the crypt-villus architecture by modulating fibroblast activity and migration, while also modulating immune responses within the lamina propria[101]. Transforming Growth Factor Beta (TGF-β) signalling plays a complex role in most cellular processes, namely cell survival, metabolism, growth, proliferation, differentiation, adhesion, migration, and death[102]. In particular, under

normal conditions TGF-β plays a pivotal tumour suppressor role by inducing apoptosis and inhibiting cell proliferation[100,103]. The Hippo pathway controls epithelial cell proliferation and tissue homeostasis by modulating the activity of the Yes-associated protein (YAP) with PDZ (post synaptic density protein [PSD95], Drosophila disc large tumor suppressor [Dlg1], and zonula occludens-1 protein [zo-1])-binding motif transcriptional coactivators (TAZ - Transcriptional Co-Activator with PDZ-binding motif), thus responding to mechanical cues and cellular density[63,98,104,105]. Together, these pathways form an integrated network that governs intestinal epithelial homeostasis, with their intricate crosstalk ensuring a tightly regulated environment conducive to maintenance of intestinal integrity and regeneration. These non-canonical pathways integrate mechanical, spatial, and inflammatory inputs to fine-tune epithelial responses, especially during injury, remodelling, or disease progression.

1.1.3.3 Plasticity and Regeneration

A further dimension of epithelial regulation is cellular plasticity - the ability of epithelial cells to reversibly change their identity in response to injury or environmental perturbations[4,91,92,106–109]. Under normal conditions, ISCs and their progeny follow a well-defined path of proliferation and differentiation[110]. However, the epithelium can exhibit remarkable dynamism in response to injury or inflammation[109,111–113]. Several studies have highlighted the ability of differentiated cells to acquire a hybrid, stem-like state, contributing to tissue repair and regeneration[43,71,114,115]. For example, *Zhou et al.* claim that, upon tissue injury, YAP/TAZ become activated and translocate to the nucleus, where they reprogram cellular identity by inducing the expression of genes that promote cell proliferation and inhibit apoptosis[116]. This activation leads to dedifferentiation of mature epithelial cells into stem-like cells, replenishing the ISC pool and facilitating tissue repair[117]. Haerinck *et al.* have expanded the concept of epithelial plasticity beyond stem cell dynamics to encompass broader epithelial–mesenchymal plasticity (EMP), a

reversible spectrum of states between epithelial and mesenchymal phenotypes[118]. EMP enables cells to transiently acquire hybrid phenotypes, balancing epithelial integrity with mesenchymal features - such as motility - which are critical during development, tissue repair, and inflammation[119]. Mechanistically, EMP is not governed solely by deterministic transcriptional programmes but is also shaped by lineage identity, microenvironmental signals, and stochastic fluctuations, which confer memory of previous cell states[92,118,119]. Understanding the molecular triggers and limits of epithelial plasticity remains a critical frontier for regenerative biology and disease modelling.

In summary, the intestinal epithelium is governed by a sophisticated network of signalling pathways and cellular plasticity programs that ensure its rapid renewal, structural integrity, and adaptive capacity in the face of physiological stress or injury. Yet, these epithelial responses do not occur in isolation. They are deeply influenced by the surrounding cellular and acellular environment — including mesenchymal cells, immune populations, and microbial metabolites — through continuous bidirectional communication. The next section will explore these intricate intercellular interactions that coordinate homeostasis and immune surveillance within the intestinal microenvironment.

1.1.4 Intercellular Communication in the Intestinal Microenvironment

Cell-cell communication within the intestinal microenvironment is essential for maintaining homeostasis, coordinating cellular behaviour, and responding to environmental stimuli[120]. This complex interplay involves several sophisticated mechanisms, including paracrine signalling, direct cell-cell contact, as well as dynamic interactions with the ECM and the gut microbiota[17,121–124]. Most major signalling pathways within the intestinal epithelium - such as

Wnt, Notch, BMP, Hh, and TGF-β - are canonical examples of cell-cell communication, underpinning tissue organisation and stem cell regulation[51,95,101,102,104,114].

1.1.4.1 Paracrine Signalling in the Niche

Paracrine signalling plays a pivotal role in the intestinal niche primarily through ligand-receptor interactions, wherein cells secrete diffusible factors that act locally on neighbouring cells to modulate their behaviour. In the small intestine, Paneth Cells, strategically located adjacent to ISCs at the crypt base, are key contributors to the epithelial niche[56]. They secrete a range of critical signalling molecules, including Wnt and Notch ligands, and epidermal growth factor (EGF), which collectively support ISC maintenance, proliferation, and fate decisions[57,125]. Beyond the epithelium itself, mesenchymal stromal cells contribute essential paracrine signals by producing both BMP (which promotes epithelial differentiation) and its antagonists, such as Gremlin1 (Grem1), Noggin (Nog), follistatin (FST), and sclerostin (SOST)[38,42,126,127]. The balance between these opposing cues fine-tunes the equilibrium between proliferation and differentiation, underpinning epithelial tissue renewal and integrity[97,123,128]. This carefully orchestrated paracrine signalling landscape establishes a supportive niche that ensures ISC function and epithelial renewal, setting the stage for direct cell-cell interactions that further contribute to epithelial integrity.

1.1.4.2 Direct Cell–Cell Contact & Junctional Signalling

Direct cell-cell interactions refer to the transmission of signals through physical contact between neighbouring cells, rather than through diffusible signalling molecules[129–131]. Cell-cell interactions also play a crucial role in regulating epithelial structure and function, and are largely mediated by specialised junctional complexes, including AJs, TJs, and GJs[29,74,77,78]. Beyond barrier function, TJs also participate in signalling cascades that influence gene expression, proliferation, and differentiation[29]. Disruption of TJ integrity is an early event in

many pathological states, including inflammation and cancer. AJs' cadherins, intracellularly linked to the actin cytoskeleton via adaptor proteins as β-catenin and α-catenin, provide a connection which not only strengthens mechanical cohesion across the epithelial sheet but also acts as a mechanosensor, allowing cells to respond to mechanical stress and adjust adhesion dynamically[76,132]. GJ communication is crucial for synchronising cellular activities, such as coordinated contraction in smooth muscle layers, calcium wave propagation, and metabolic coupling[133]. In the intestinal epithelium, GJs contribute to the synchronised renewal and response to injury of epithelial cells[78,79,133]. Beyond their mechanical role, emerging evidence suggests that desmosomes also participate in signal transduction pathways regulating cell proliferation, differentiation, and response to injury, positioning them as integral components of epithelial integrity and homeostasis[82,134]. Together, epithelial intercellular junctions not only preserve tissue integrity but also act as active communication hubs, transmitting mechanical and biochemical cues essential for coordinated epithelial responses. Epithelial cells can also sense the tension generated at cell–cell interfaces (and transmitted through intercellular junctions, as discussed above) to then modulate signal transduction pathways such as Hippo, Wnt, and β-catenin signalling[135]. As such, these junctional complexes not only contribute to the epithelial barrier integrity but also serve as mechanosensitive platforms that relay physical cues into intracellular responses — linking the structural environment with broader tissue regulation via the extracellular matrix. Furthermore, contact-dependent signalling through membrane-bound ligands and receptors also constitutes a form of direct communication[136]. For example, in the Notch pathway, the interaction between Delta-like or Jagged (Jag) ligands on one cell and Notch receptors on an adjacent cell leads to receptor cleavage and intracellular signalling, regulating fate decisions in a highly spatially restricted manner[137].

1.1.4.3 ECM-Mediated Signalling & Mechanotransduction

The ECM plays an important role in intercellular communication by functioning as a dynamic signalling platform that facilitates indirect interactions between cells within the intestinal microenvironment. Beyond its structural role, the ECM transmits biochemical and mechanical cues that shape cell fate and function[138,139]. Epithelial cells engage with the ECM primarily through integrins — transmembrane receptors that link the ECM to the intracellular actin cytoskeleton[140]. ECM-cell interactions initiate "outside-in" signalling cascades, activating intracellular kinases such as focal adhesion kinase (FAK) and Src family kinases[141–144]. These pathways regulate critical cellular processes, including proliferation, survival, migration, and differentiation. Notably, the biochemical composition and mechanical properties of the ECM can profoundly modulate cell behaviour[145,146]. For instance, increased matrix stiffness has been associated with enhanced epithelial proliferation and altered differentiation, highlighting how this dynamic and continuous crosstalk between cells and the ECM can be crucial for tissue morphogenesis, repair, and long-term homeostasis[147]. By integrating biochemical and biomechanical signals, ECM-mediated pathways extend the scope of cellular communication, bridging local microenvironments with tissue-wide coordination — an interplay influenced by microbial and stromal components.

1.1.4.4 Microbial and Stromal Interactions

The gut microbiota plays a significant role in modulating epithelial homeostasis too, through all three cell-cell communication mechanisms described. For example, commensal bacteria interact with epithelial cells through pattern recognition receptors (PRRs), such as Toll-like receptors (TLRs), influencing immune responses and epithelial renewal[14]. Dysbiosis, or microbial imbalance, can disrupt these interactions, leading to IBD and other intestinal disorders[148]. Similarly, fibroblasts and myofibroblasts can also respond to environmental and

inflammatory signals by producing ECM components, contributing to tissue remodelling and fibrotic processes[12]. These microbial and stromal inputs exemplify how extrinsic factors co-opt paracrine, contact-mediated, and ECM-based signalling to shape epithelial dynamics, reinforcing the need to consider intercellular communication as a multi-tiered regulatory network essential for intestinal homeostasis.

Having explored the intricate modes of intercellular and microenvironmental communication that sustain intestinal homeostasis, the next section examines how disruptions in these interactions compromise epithelial barrier function and contribute to the pathogenesis of intestinal diseases.

1.1.5 Molecular Mechanisms Underpinning Barrier Dysfunction

Disruption of intestinal barrier integrity is implicated in both acute systemic insults — such as sepsis following acute pancreatitis — and chronic conditions like irritable bowel syndrome (IBS) and non-alcoholic steatohepatitis (NASH), where increased permeability fuels immune activation and disease progression[149]. Barrier dysfunction arises from diverse molecular mechanisms — including oxidative stress, signalling imbalances, ECM remodelling, and microbial perturbations — which will be each considered in turn[17,139,149–152].

1.1.5.1 Oxidative Stress

Disruption in the function of intercellular junctions can lead to a compromised barrier, contributing to various gastrointestinal disorders[3,23,153]. Oxidative stress, characterized by an overproduction of reactive oxygen species (ROS), can damage cellular components, including lipids, proteins, and DNA[154]. In the intestinal epithelium, oxidative stress can lead to the disruption of TJs by modifying their proteins, or by activating signalling pathways that result in their disassembly[154,155]. This mechanism is particularly relevant in conditions like IBD, where

chronic inflammation leads to sustained ROS production, perpetuating barrier dysfunction[156]. Alterations in TJ composition or function can lead to increased intestinal permeability, often referred to as "leaky gut"[157]. A leaky gut is unable to efficiently separate the intestinal flora and luminal contents from the intestinal immune system, allowing the translocation of luminal antigens, toxins, and microorganisms into the lamina propria[157]. This permeability may trigger an immune response which can result in chronic inflammation and tissue damage, characteristic of conditions such as IBD, where barrier dysfunction is both the cause and consequence of inflammation[157–159]. Understanding how oxidative damage weakens junctional integrity helps explain why antioxidants and redox modulators are being investigated as protective strategies in IBD[157,158].

1.1.5.2 Cytokine-Mediated Damage

Key inflammatory cytokines including tumour necrosis factor-alpha (TNF-α), interferon-gamma (IFN-γ), interleukin-1β (IL-1β), and IL-13, secreted upon immune system activation, destabilise tight junctions, fuelling a self-amplifying permeability-inflammation cycle that correlates with IBD relapse and colorectal cancer risk [3,160–163]. These insights have spurred the development of therapeutic strategies aimed at restoring epithelial barrier function by targeting the cytokine–tight junction axis. Anti-inflammatory agents, including anti-TNF-α antibodies, IL-13 and IL-6 inhibitors, and small-molecule myosin light-chain kinase (MLCK) blockers, have shown promise in reinforcing epithelial integrity and disrupting the feedback loop between inflammation and permeability implicated in both IBD and inflammation-associated colorectal cancer[164–166]. Notably, both patients with ulcerative colitis (UC) — a form of IBD characterised by continuous mucosal inflammation of the colon — and Crohn's disease (CD) — which features patchy, transmural inflammation that can affect any part of the gastrointestinal tract — exhibit reduced expression of occludin, likely driven by TNF-α and IFN-

γ signalling[162,165]. This is accompanied by downregulation of sealing claudins (e.g., claudin-5 and -8 in CD; claudin-4, -7, and -15 in UC) and a concurrent upregulation of the pore-forming claudin-2[165,167]. Together, these alterations reinforce a cycle of barrier disruption, immune dysregulation, and tissue pathology, highlighting key targets for therapeutic intervention.

1.1.5.3 ECM Remodelling and Epithelial-Mesenchymal Transition

The ECM can also play a role in amplifying barrier dysfunction in pathological states. For example, under chronic inflammatory conditions, such as those observed in IBD, dysregulated fibroblast activation leads to excessive ECM deposition and pathological intestinal fibrosis[168,169]. This pathological remodelling not only disrupts normal tissue architecture but also exacerbates disease progression, emphasising the critical importance of understanding stromal-epithelial crosstalk in both health and disease[127,170,171]. Similarly, inflammatory signals can also induce epithelial-mesenchymal transition (EMT), a process where epithelial cells acquire mesenchymal traits, enhancing their migratory and invasive capabilities[172]. EMT is implicated in fibrosis and cancer metastasis, where chronic inflammation drives the continuous activation of EMT programs, leading to pathological tissue remodelling and cancer progression[173]. Finally, chronic inflammation associated with barrier disruption can also lead to genomic instability and promote colorectal carcinogenesis[174].

1.1.5.4 Signalling Pathway Dysregulation

Dysregulated signalling networks undermine barrier integrity and contribute to disease progression through three principal mechanisms: perturbation of the stem cell and growth axis (via Wnt/β-catenin and BMP pathways)[94,97,111,175]; disruption of cell fate and differentiation programmes (through Notch and Hh signalling)[90,96,98,99]; and, in genetically altered epithelium, co-option of tumour-suppressor pathways (such as TGF-β). Dysregulation of Wnt/β-catenin activation can trigger unchecked intestinal epithelial proliferation and tumourigenesis; for

example, mutations in *APC*, a negative regulator, are characteristic of colorectal cancer, underscoring the pathway's centrality to intestinal health and disease[111,114]. Dysregulation of BMP signalling, documented in both IBD and CRC, leads to aberrant crypt architecture and impaired epithelial maturation gradients[128]. Similarly, aberrant Notch signalling disrupts the delicate balance between absorptive and secretory lineages, with sustained Notch activation linked to reduced goblet cell differentiation and impaired barrier function — phenomena observed in both chronic colitis and colorectal neoplasia[98,176,177]. Impairment of the Hh pathway is similarly implicated in IBD pathogenesis, as reduced epithelial Hh activity correlates with increased pro-inflammatory cytokine expression and immune cell infiltration[101,178]. Lastly, epithelial TGF-β signalling, typically tumour-suppressive, can be disrupted in the presence of oncogenic mutations such as *APC* or *KRAS*, prompting a functional switch towards tumour promotion and progression in colorectal cancer[103]. Together, these pathways are therefore central not only to intestinal renewal but also to pathogenesis, making them key therapeutic targets in IBD and CRC.

1.1.5.5 Microbiota-Induced Barrier Dysregulation

Finally, the gut microbiota also plays a pivotal role in modulating barrier integrity. It is estimated that microbial cells within the human body outnumber human cells by more than an order of magnitude. The gut microbiota alone comprises over a thousand distinct microbial species, with a collective metagenome that encodes approximately one hundred times more genes than the human genome[179–182]. The microbiome is crucial for diverse processes including nutrition, energy metabolism, host defence, and the development of the immune system[183]. Commensal microbes safeguard the barrier by producing SCFAs, enhancing mucus secretion, and promoting TJ protein expression[25,184]. However, a shift in the regular composition of the gut microbiome – gut dysbiosis - can result in pathogenic species overgrowth, compromising

barrier function and leading to excessive immune activation and systemic inflammation, which further compromises the epithelial barrier[29,185]. Dysbiosis, featuring expansions in *Proteobacteria* and depletion of *Firmicutes*, is a hallmark of IBD and CRC[186–189]. Several microbiome-targeted therapies — including faecal microbiota transplantation (FMT), multi-strain probiotics like VSL#3, and bacteria-based live biotherapeutics products (LBPs) like *Akkermansia muciniphila* — are currently under evaluation[190]. Preclinical studies also demonstrate strain-specific benefits: *Lactobacillus rhamnosus GG* (activating EGFR signaling), *Roseburia intestinalis* (reducing immune infiltration), *Faecalibacterium prausnitzii* and *Bacteroides thetaiotaomicron* (promoting regulatory T-cell responses, suppressing Th17 cells, increasing mucosal IL-10 production, and modulating butyrate)[148,191] . Altogether, these findings highlight how gut microbiota is both a guardian and a potential disruptor of barrier integrity, positioning microbial modulation as a promising therapeutic avenue for restoring homeostasis in inflammatory and neoplastic intestinal diseases.

In sum, epithelial barrier integrity depends on the orchestration of structural junctions, redox balance, finely tuned signalling pathways, and immune-microbial interactions. Disruption across any of these axes can increase permeability, impair epithelial plasticity, and drive intestinal disorders. However, key molecular insights are now guiding targeted therapies — ranging from cytokine blockers and MLCK inhibitors, to anti-Wnt agents, anti-fibrotics, and microbiome-based interventions[192,193]. Despite these advances, barrier biology remains incompletely understood, highlighting the need for robust human-relevant models to translate these findings into durable clinical strategies.

To better dissect the complexity of intestinal barrier regulation and accelerate therapeutic development, researchers increasingly turn to organoid cultures — three-dimensional, stem cell–derived systems that recapitulate key features of epithelial architecture and function.

These models offer unprecedented access to human-specific epithelial biology under both physiological and pathological conditions, overcoming the limitations of traditional monolayer cultures and animal models. The following section explores the emergence, utility, and evolving sophistication of intestinal organoid platforms as dynamic tools to simulate and interrogate epithelial homeostasis and disease.

1.2 Simulating Intestinal Epithelial Dynamics using Organoid Cultures

The remarkable complexity of the intestinal epithelium — its rapid turnover, intricate cell–cell communication, and dependence on environmental cues — has historically posed a challenge for *in vitro* modelling. Recent advances in three-dimensional (3D) culture systems have led to the development of intestinal organoids, which offer a powerful and physiologically relevant platform for simulating epithelial dynamics in health and disease. These self-organising structures, derived from either pluripotent stem cells or adult intestinal stem cells, faithfully recapitulate key aspects of native intestinal architecture, cellular heterogeneity, and function. By bridging the gap between reductionist 2D monolayers and non-human animal models, organoids enable precise investigation of epithelial behaviour in a controlled, human-specific context. The following sections explore the foundational principles, myriad applications, and emerging frontiers of intestinal organoid technologies, positioning them as indispensable tools for dissecting epithelial biology, modelling disease, and testing targeted interventions.

1.2.1 Overview of Organoid Culture Systems

The conceptual foundation of organoid biology rests on the discovery that, when provided with the appropriate ECM and niche-specific cues, stem cells can self-organise into 3D structures that recapitulate their tissue of origin. A pivotal breakthrough in the field was the generation of the first intestinal organoids from $LGR5^+$ crypt base columnar stem cells isolated from murine small intestine. These cells autonomously formed crypt–villus-like domains *in vitro*, preserving both region-specific differentiation and long-term self-renewal[44]. This finding laid the groundwork for patient-derived organoid systems and catalysed widespread adoption of organoids in developmental and disease modelling.

Organoid technology has since expanded to encompass a broad range of human tissues, including stomach, colon, liver, pancreas, brain, kidney, and lung. Depending on the biological question, organoids may be derived from embryonic stem cells (ESCs), induced pluripotent stem cells (iPSCs), adult stem cells (ASCs), or more recently, transdifferentiated somatic cells. The source of origin profoundly influences the organoid's developmental trajectory and utility: ESC/iPSC-derived organoids are particularly suited for studying early development and congenital disorders, whereas ASC-derived organoids more faithfully retain adult tissue identity and are highly relevant for disease modelling and regenerative applications[194].

Human intestinal epithelial organoids (hIEOs) stand out for their ability to retain key physiological features of the native intestine, including cellular polarity, barrier integrity, nutrient transport, and lineage specification[44,87,195,196]. Unlike conventional cell lines that often lose tissue-specific characteristics, hIEOs preserve the complex multicellular architecture and functionality of the intestinal epithelium[197]. This makes them especially valuable for studying epithelial–microbial interactions, dietary and pharmaceutical responses, and toxicological effects in a human-relevant context.

The successful generation and maintenance of organoids requires a reconstitution of the *in vivo* niche, which is typically achieved by embedding stem cells in a 3D ECM scaffold — most commonly Matrigel[198,199]. This basement membrane extract, rich in laminin, collagen IV, and entactin, supports organoid growth but suffers from batch variability and undefined composition. To improve reproducibility and enable clinical translation, synthetic hydrogels with tuneable mechanical and biochemical properties are under active development[199].

In addition to ECM components, organoid media must be supplemented with critical niche-specific growth factors and signalling molecules. For intestinal organoids, these typically include Wnt agonists, R-spondin, EGF, and Noggin, which together mimic the crypt

microenvironment and sustain ISC self-renewal and differentiation[10,57,87,123,200]. Recent innovations also emphasise the importance of physical and mechanical cues — such as oxygen gradients, ECM stiffness, and spatial constraints — in shaping epithelial behaviour[201,202]. To better simulate physiological conditions, techniques like microfluidic platforms and rotating bioreactors are increasingly used to enhance nutrient exchange, enable growth of larger organoids, and promote tissue maturation[203,204]. Together, these advancements underscore how organoids have evolved into versatile and high-fidelity systems that enable the simulation of epithelial behaviour with unprecedented accuracy, positioning them as central tools in translational research, therapeutic development, and personalised medicine.

1.2.2 Organoids as Biomimetic Models of Health and Disease

As *in vitro* proxies for human development, organoids facilitate mechanistic studies that are otherwise inaccessible *in vivo.* In developmental biology, organoids derived from pluripotent stem cells (PSCs) have provided novel insights into morphogenesis and lineage specification. Cerebral organoids, for instance, have been instrumental in studying cortical development, neurogenesis, and the aetiology of neurodevelopmental disorders such as microcephaly[205,206]. Similarly, intestinal organoids have enabled dissection of crypt–villus axis formation, Wnt-dependent stem cell maintenance, and regional identity along the gut tube[41,47,87,125].

Clinically, patient-derived organoids (PDOs) have redefined disease modelling. Generated from primary tissues or iPSCs, PDOs preserve the genetic, epigenetic, and phenotypic signatures of their source, including complex traits like cellular heterogeneity and spatial organisation[39,204,207]. These models have been established across a range of conditions — cystic fibrosis, IBD, infectious diseases, and various cancers — enabling both mechanistic studies and

translational applications[208–211]. Notably, organoids from IBD patients often retain disease-specific transcriptomic and epigenetic signatures *ex vivo*[212,213]. A 2024 single-cell RNA-sequencing study of rectal organoids from Crohn's disease patients revealed patient-specific transcriptional profiles across 10 epithelial subtypes, with altered gene expression in progenitor and stem-like cells, particularly in pathways related to Wnt signalling and chromatin remodelling[214]. Crucially, these differences were maintained in the absence of immune stimulation, suggesting stable, epithelial-intrinsic alterations, —potentially linked to genetic predispositions or chronic inflammatory priming. Organoid-based infection platforms have also been developed for pathogens including norovirus, rotavirus, *Salmonella*, *Clostridium difficile*, and SARS-CoV-2[215–218]. These models provide human-specific insights into host-pathogen interactions, immune responses, and barrier disruption, which are often difficult to capture in animal models[219]. In oncology, tumour-derived organoids retain mutational profiles, histoarchitecture, and drug response patterns of the primary tumour, allowing for high-throughput drug screening and predictive testing[220]. Importantly, emerging clinical studies suggest that PDO responses to chemotherapy often correlate with patient outcomes, opening avenues for personalised therapeutic strategies[221].

Beyond modelling, organoids are also advancing regenerative medicine. Preclinical studies have shown that intestinal and hepatic organoids can engraft into injured tissue, integrate functionally, and restore physiological activity[222,223]. Although clinical translation remains at an early stage, these findings demonstrate the feasibility of organoids as bioengineered grafts or therapeutic reservoirs[224]. Despite their advantages, organoids face important limitations. The absence of vasculature restricts their size, limits nutrient diffusion, and hampers full tissue maturation[225,226]. Similarly, most current organoid systems lack immune, neural, and stromal

components, which are critical for recapitulating the full physiological context[207,227]. To address these gaps, co-culture systems incorporating endothelial cells, immune cells, and fibroblasts — as well as organ-on-a-chip and microfluidic platforms — are under active development[204,228]. Standardisation is another challenge: variability in donor tissues, matrix compositions (e.g., Matrigel), and media formulations can compromise reproducibility. Defined synthetic matrices and automated culture protocols are emerging to improve experimental robustness and enable clinical translation[229,230]. Ethical considerations are also gaining prominence, particularly for complex organoids such as cerebral organoids, which may exhibit rudimentary neural activity[231]. Questions surrounding the moral status of organoids, patient consent, and data privacy will require ongoing ethical oversight as the technology matures.

In summary, organoids offer a high-fidelity, modular, and human-specific platform to investigate a broad array of biological processes. Their growing sophistication, combined with advances in bioengineering, multi-omics, and computational modelling, may soon contribute to significant transformations in biomedical research and therapeutic development.

As organoids increasingly serve as indispensable models for studying human tissue physiology and pathology, there is a parallel need for robust analytical methods to characterise their structural, functional, and molecular properties. The following section outlines the key technologies and strategies currently employed to evaluate organoid cultures, providing insights into their development, differentiation, and relevance to disease.

1.3 Analytical Methods for Characterising Human Intestinal Epithelial Organoids

In complex organisms, billions of cells — spanning diverse specialised types and dynamic states — are intricately organised into spatial compartments that coordinate physiological functions. Reconstructing this cellular complexity *in vitro* has long been a central goal of developmental and regenerative biology. Organoid cultures have emerged as powerful biomimetic models that recapitulate the structural, cellular, and functional hallmarks of native tissues, offering an accessible and manipulable platform to study development, disease, and regeneration under controlled conditions. Yet, the full potential of organoids as investigative tools has been unlocked only through their integration with high-resolution molecular profiling techniques. Among these, single-cell technologies have proven particularly transformative.

1.3.1 Functional Single-Cell Technologies: From Primary Tissue to Organoids

As discussed in previous sections, the intestinal epithelium is a dynamic and highly adaptable tissue, marked by rapid turnover and functional diversity. Traditional bulk transcriptomic methods, while informative, often obscure rare cell types, transitional progenitor states, and plastic lineage behaviours that are central to both tissue homeostasis and repair[232]. The emergence of single-cell profiling approaches has fundamentally transformed our ability to resolve this complexity at high resolution in primary human tissues.

The most widely adopted technologies bridging single-cell biology and molecular profiling are single-cell RNA sequencing (scRNA-seq) scRNA-seq and single-cell Assay for Transposase-Accessible Chromatin using sequencing (scATAC-seq)[233–236]. These tools enable high-resolution

interrogation of transcriptional states and regulatory architecture within organoid systems, transforming them into dynamic platforms for studying development, disease, and environmental response.

scRNA-seq profiles gene expression at single-cell resolution by isolating individual cells, capturing their RNA transcripts, converting them into complementary DNA (cDNA), and sequencing to reconstruct the transcriptomic landscape[237]. This enables the identification of distinct cell types, transcriptional states, and developmental trajectories[41,238,239]. Leveraging scRNA-seq, several large-scale atlases have mapped the cellular composition of the human gut across developmental stages and anatomical regions, revealing extensive epithelial and immune heterogeneity[22,46,47,213].

scRNA-seq has confirmed the presence of canonical, broadly enriched epithelial lineages such as Enterocytes, Goblet Cells, and TA Cells. More significantly, it has enabled high-resolution characterisation of epithelial subsets that were previously poorly defined or entirely uncharacterised. This includes SPIB-regulated lineages such as M-Cells, distinct tuft cell subtypes (Tuft-1 and Tuft-2), diverse EEC subtypes, and Best4 Cells[22,38,64,71]. These rare or niche populations are often absent or underrepresented in bulk transcriptomic datasets, underscoring the superior resolution of single cell approaches in capturing the full cellular heterogeneity of the intestinal epithelium. These discoveries have culminated in the generation of comprehensive single-cell atlases that map epithelial diversity across intestinal regions and developmental stages. Such studies have uncovered not only the full spectrum of mature cell types but also transitional progenitor states and unexpected differentiation paths, including proliferative epithelial precursors other than canonical $LGR5^+$ stem cells[10,41,240]. These insights have challenged classical models of stem cell identity and epithelial plasticity.

Crucially, these lineage-specific transcriptional signatures are largely preserved in human intestinal hIEOs. Comparative analyses have shown that hIEOs maintain the molecular identity of their tissue of origin, including both stem and differentiated cell types[241]. This fidelity enables organoids to serve not just as culture models, but as *bona fide* avatars of native intestinal tissue for functional studies, especially when profiling rare epithelial populations.

Beyond developmental lineage mapping, scRNA-seq has revealed how environmental factors — including microbial metabolites, dietary inputs, and inflammatory cues — dynamically influence epithelial gene expression[38,46,242,243]. For instance, recent studies demonstrate how microbial stimuli modulate immune-epithelial crosstalk and transcriptional programs, thereby altering barrier integrity and cell fate[244,245]. In hIEOs, scRNA-seq has captured rare epithelial subtypes, disease-specific transcriptional signatures, and altered differentiation trajectories, offering high-resolution insight into both healthy and pathological states[214,246,247].

Beyond basic characterisation, scRNA-seq has also revealed how epithelial composition is reshaped in disease. IBD, for instance, is associated with expanded secretory lineages and the emergence of inflammation-responsive or metaplastic states[176,213,215,249,250]. In UC, scRNA-seq has identified increased frequencies of SPIB$^+$ M-like cells and IL-13–responsive epithelial subsets[251]. Multi-cohort analyses have also uncovered the presence of MUC5AC$^+$/MUC6$^+$ goblet-like inflammatory epithelial (INFLARE) cells, markers of a pyloric-type metaplasia that emerges during chronic inflammation[251]. Importantly, many of these transcriptional alterations persist in hIEOs derived from patient biopsies. For example, rectal organoids from Crohn's disease patients retain disease-associated gene expression programs — particularly in Stem Cells and TA compartments — even after extended culture *ex vivo*[215]. This stability underscores the capacity of organoids to preserve intrinsic, cell-specific transcriptional "memory" from their tissue of origin.

By contrast, scATAC-seq, interrogates chromatin accessibility by employing a hyperactive transposase to insert sequencing adapters into open chromatin regions, thereby identifying active promoters, enhancers, and transcription factor binding sites[252,253]. This technique is particularly relevant in the intestinal epithelium, where precise regulatory control is essential for maintaining cell identity amid rapid turnover and environmental fluctuation[253,254]. In hIEOs, scATAC-seq has elucidated how epigenetic landscapes influence lineage specification and has revealed context-specific regulatory programs disrupted by inflammation or injury[255]. When integrated with scRNA-seq, these data enable the construction of cell-type-resolved gene regulatory networks that link chromatin accessibility to transcriptional output[255,256].

Even when used independently, these technologies provide an unprecedented window into the cellular heterogeneity and regulatory complexity of organoid cultures. They facilitate the reconstruction of lineage trajectories, the identification of rare or transitional populations, and the mapping of cell-specific regulatory circuits—extending the capacity of organoids to model dynamic tissue states. In this way, organoids and single-cell technologies form a synergistic toolkit: organoids offer a structured, physiologically relevant, and experimentally tractable model system, while single-cell methods deconvolute the molecular logic embedded within complex multicellular structures.

More recently, multiomic platforms have emerged that enable the simultaneous profiling of transcriptomes, chromatin accessibility, and protein abundance within the same cell[257–261]. Such integrative strategies are especially powerful in studying the intestinal epithelium—a tissue characterised by continuous renewal, high functional diversity, and spatial specialisation. When coupled with organoid models, single-cell and multiomic approaches allow fine-grained analysis of how environmental cues — including microbial metabolites, inflammatory cytokines, and dietary factors — shape epithelial cell states and fate decisions

across physiological and pathological contexts. The next sections will explore how these technologies extend beyond transcriptomics to dissect regulatory landscapes, spatial architecture, and functional phenotypes at single-cell resolution.

1.3.2 Integrating Multi-Dimensional Functional Single-Cell Technologies to Advance Epithelial Biology Exploration

Advancements in multimodal single-cell technologies now allow the simultaneous capture of multiple molecular layers — transcriptomes, epigenomes, and proteomes — within the same cell. These approaches are transforming our ability to investigate the regulatory complexity and cell-state diversity of epithelial organoids.

1.3.2.1 Multimodal Platforms Bridging Transcriptomic, Epigenetic, and Proteomic Layers

Recent advances in single-cell profiling have enabled dual-modality platforms that capture two complementary layers of molecular information within individual cells. These technologies bridge transcriptional output, chromatin accessibility, and surface protein expression, offering deeper insight into cellular identity, regulatory state, and phenotypic plasticity — particularly valuable in complex, dynamic systems like hIEOs.

Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITE-seq) represents one such platform, quantifying both mRNA and surface protein abundance in single cells by including oligonucleotide-labelled antibodies[262]. This integrated readout improves cell-type resolution and reveals discrepancies between gene expression and protein presence — important in post-transcriptionally regulated tissues[263]. Importantly, CITE-seq can be readily adapted for sample multiplexing through hashing strategies, enabling simultaneous cell barcoding and antibody-based profiling across multiple conditions or donors[264]. In hIEOs, cell

hashing has enabled the high-throughput screening of epithelial cell composition modulators[265].

Complementing this, the canonical Single-cell Multiome platform (scMultiome-seq) simultaneously captures gene expression and chromatin accessibility (scRNA-seq + scATAC-seq) from the same cell[260,266]. This strategy maps cis-regulatory elements to their transcriptional consequences, enabling the reconstruction of lineage hierarchies and gene regulatory networks. In organoid systems — where niche architecture and cell fate trajectories may diverge from *in vivo* counterparts — scMultiome profiling has proven crucial for resolving the regulatory logic underpinning differentiation, injury responses, and developmental reprogramming[267].

Extending this dual framework further, Assay for Transposase-Accessible Chromatin with Select Antigen Profiling by sequencing (ASAP-seq) integrates scATAC-seq with CITE-seq[257]. This modality allows researchers to link chromatin accessibility landscapes to protein phenotypes, facilitating high-resolution annotation of rare, transient, or activation-specific states.

As multimodal technologies evolve, DOGMA-seq represents the cutting edge in single-cell profiling[257]. This tri-modal platform simultaneously measures transcriptomic (scRNA-seq), epigenetic (chromatin accessibility, via scATAC-seq) and surface proteomic (via oligo-tagged antibodies) profiles within the same cell, enabling the construction of highly resolved regulatory maps that link DNA-level regulatory logic to both RNA and protein expression. In organoid studies — particularly those exploring epithelial plasticity and disease states or scrutinising differentiation trajectories — this level of molecular resolution is transformative, as it allows the dissection of how transcriptional programs are executed and stabilised at the protein level within individual cells, offering deep insights into functional phenotypes that span multiple molecular layers.

Together with the dual-modality platforms introduced earlier, DOGMA-seq pushes the frontier of cellular deconvolution. It offers unparalleled clarity into the molecular mechanisms guiding cell identity, lineage commitment, and adaptive responses under experimental or pathological conditions. The next section explores complementary modalities — including full-length SMART-seq technologies and spatial profiling techniques — that further extend our ability to characterise organs and organoids[268,269]. These approaches add depth and context, capturing rare transcripts, preserving tissue architecture, and linking molecular signatures back to their spatial origins.

1.3.2.2 Expanding the Single-Cell Toolkit: Additional Modalities

Beyond the established multimodal platforms, a range of complementary single-cell technologies continues to broaden the scope and depth of cellular analysis within organoid systems. Among these, SMART-seq platforms (including SMART-seq2 and SMART-seq3) offer full-length transcript coverage and superior sensitivity for low-abundance transcripts[268–270]. This makes them especially valuable for resolving rare epithelial populations such as Tuft Cells, M-Cells, or Best4 Cells — populations often underrepresented in droplet-based methods. Another powerful addition is single-cell DNA methylation sequencing, which enables the profiling of stable epigenetic modifications, such as CpG methylation[271–273]. Unlike chromatin accessibility assays that capture transient regulatory states, this method provides a longer-term view of cell lineage commitment, developmental history, and epigenomic memory[274].

Yet, while these approaches excel at resolving transcriptional and regulatory heterogeneity, they all share a critical limitation: the loss of spatial context due to tissue dissociation. This becomes increasingly problematic in increasingly sophisticated organoid systems (such as those involving developmental trajectories or programmable gene perturbations) where spatial cues — such as radial organisation — play decisive roles in cell fate and function[275,276].

To overcome this, spatial transcriptomics has emerged as a pivotal innovation[48,277–279]. By preserving the positional identity of gene expression within intact organoid sections, this approach bridges molecular signatures with tissue architecture[280]. It enables the localisation of distinct cell states, lineage boundaries, and niche-dependent expression programs, offering an integrated view of form and function.

In sum, these additional modalities enhance the dimensionality of single-cell analysis, filling key gaps in resolution and functional insight. However, their true potential is realised when integrated with physiologically relevant *in vitro* systems that recapitulate the spatial and mechanical complexity of native tissues— such as gut-on-a-chip systems and microengineered organoid cultures.

1.3.2.3 Integrating Single-Cell Modalities with Advanced *In Vitro* Models

The integration of single-cell technologies with advanced *in vitro* models has significantly extended their utility. Organoid cultures and gut-on-a-chip platforms now provide tractable systems for studying intestinal epithelial biology under near-physiological conditions[281]. A gut-on-a-chip is a microfluidic device that mimics the structure and function of the human gastrointestinal tract, enabling the study of intestinal physiology, disease, and drug responses in a controlled, dynamic environment[229,282]. When paired with scRNA-seq, these models reveal how specific environmental or mechanical cues influence cell fate and function[283]. For example, comparisons between traditional transwell systems and gut-on-a-chip platforms using scRNA-seq have shown that the latter more closely recapitulate *in vivo*-like gene expression profiles and cellular heterogeneity, highlighting the importance of incorporating biomechanical forces and luminal flow into *in vitro* systems to better mimic native epithelial environments[284]. Similarly, integration of single-cell data with live-cell imaging (e.g. smLiveFISH), and perturbation-based screening (e.g. Perturb-seq) is further advancing our

ability to reconstruct gene regulatory networks and dissect causal relationships between transcriptional programmes and functional phenotypes[285–287].

In summary, single-cell technologies offer a powerful lens through which to study cellular heterogeneity, regulatory dynamics, and environmental interactions within the intestinal epithelium. When integrated with organoid models and emerging multiomic platforms, these tools enable a systems-level understanding of tissue organisation and function. As the field continues to refine and converge these tools, the prospect of reconstructing tissue-level organisation and function from molecular first principles become increasingly tangible — paving the way toward a systems-level understanding of epithelial biology in health and disease.

1.3.3 Single-Cell Lineage Tracing of Epithelial Plasticity in Human Intestinal Organoids

While single-cell transcriptomics and multiomic platforms have greatly advanced our understanding of cell identity and regulatory states within organoids, they primarily offer static snapshots. To fully capture the dynamic nature of epithelial behaviour — particularly processes like regeneration, metaplasia, or lineage bifurcation — time-resolved approaches are essential[288]. Single-cell lineage tracing fills this gap by revealing how individual cells transition between states over time and under specific perturbations[289]. When applied to hIEOs, lineage tracing strategies allow researchers to dissect how stem cells and progenitors generate diverse lineages, how cells adopt new fates under stress, and how plasticity contributes to both tissue repair and disease progression[290,291]. The integration of cell barcoding, CRISPR-based recorders, and multi-timepoint sampling with single-cell resolution enables reconstruction of clonal dynamics and fate maps *in vitro*. This section explores how

these emerging tools are redefining our understanding of epithelial plasticity in human-relevant systems.

1.3.3.1 Overview of Major Lineage Tracing Techniques

Traditional lineage tracing — such as inducible Cre–lox recombination with fluorescent reporters — has been instrumental in mapping epithelial hierarchies *in vivo*[292]. In the context of intestinal biology, these tools have elucidated stem cell dynamics, revealed cellular plasticity, and traced differentiation trajectories across homeostasis, regeneration, and disease[293,294]. However, while powerful, such approaches are largely confined to animal systems and lack the scalability and resolution required for transcriptome-integrated or high-throughput analysis[295].

The need for greater cell-type-specific granularity has driven the development of lineage tracing strategies compatible with single-cell sequencing[289,290,296–298]. One highly robust approach to lineage inference involves the use of mitochondrial DNA (mtDNA) mutations[298–300]. However, because these mutations arise passively and stochastically, this method lacks experimental control and requires ultra-deep sequencing alongside computationally intensive algorithms to reliably detect and interpret lineage relationships. By contrast, CellTagging employs a high-diversity lentiviral barcode library to stably integrate synthetic DNA sequences — termed *CellTags* — into the 3′ untranslated region of transcripts[296,297]. These heritable barcodes are passed from progenitor to descendant cells and are co-transcribed with mRNA, enabling concurrent recovery of lineage identity and transcriptional state via scRNA-seq. This dual readout provides a uniquely integrated perspective on cell fate and molecular phenotype. Other genetic barcoding approaches, such as polyclonal lentiviral labelling or transposon-based insertions, also support high-resolution fate mapping, but often lack direct transcriptomic integration[301,302]. Meanwhile, CRISPR-based recorders — such as LINNAEUS, scGESTALT and iTracer — introduce cumulative, heritable edits at synthetic target arrays to

generate lineage “scars” readable through sequencing[290,303,304]. These systems are particularly well-suited for reconstructing deep lineage hierarchies but often entail complex engineering and computational reconstruction.

Collectively, these techniques form a powerful toolkit for dissecting epithelial cell dynamics. They are especially valuable in disease modelling: in murine colitis, for example, multicolour lineage reporters combined with scRNA-seq have revealed that crypt regeneration post-injury arises from selective clonal expansion rather than uniform progenitor activation[305]. This mosaic architecture mirrors observations in human IBD, where regenerating epithelial patches often display clonality and harbour shared somatic mutations — a phenomenon known as *clonal mosaicism* that may increase cancer risk if dominant clones accumulate oncogenic hits[306,307].

1.3.3.2 Harnessing CellTagging in hIEOs to Uncover Rare Cell Dynamics

Integrating lineage recording technologies with single-cell transcriptomics offers a powerful approach to not only trace cellular trajectories but also capture the gene expression states that accompany each fate decision[297]. While foundational studies — such as Kok *et al.* (2020), which combined AI-driven live imaging with lineage reconstruction in murine organoids — have demonstrated the feasibility of this approach in intestinal organoids, its application to hIEOs remains in its early stages[308].

The potential for lineage tracing in hIEOs is especially compelling given the unique features of the intestinal epithelium: rapid turnover, pronounced plasticity, and multiple progenitor pools that contribute to both steady-state maintenance and injury-induced regeneration. Within organoid systems, where experimental variables can be precisely controlled, lineage tracing allows researchers to dissect how intrinsic programmes and extrinsic cues shape clonal behaviour, fate bias, and tissue remodelling. This approach could be particularly valuable in identifying rare or transitional epithelial states — such as regenerative progenitors,

metaplastic intermediates, or inflammation-associated subsets. These populations often play outsized roles in disease progression or repair but are difficult to capture using static, cross-sectional techniques[54,72]. Lineage tracing would enable dynamic tracking of their emergence, expansion, and fate commitment under homeostatic or perturbed conditions. Furthermore, organoid-based lineage tracing opens the door to comparative analyses across anatomical sites (e.g., small intestine vs. colon), developmental stages, or disease contexts including IBD. hIEOs derived from distinct regions or clinical backgrounds can be subjected to unified tracing paradigms, revealing how regional identity, inflammation, or epigenetic memory influence epithelial plasticity and clonal architecture.

In this context, CellTagging presents a particularly attractive strategy. By introducing high-diversity DNA barcodes into hIEOs via lentiviral transduction, this method could promote scalable, transcriptome-integrated clonal tracking without requiring genetic modification of the donor tissue, which is ideal for patient-derived systems. Applying CellTagging to hIEOs would enable the reconstruction of lineage trees alongside single-cell transcriptomic profiles, providing a dual-resolution framework to study how individual progenitors contribute to lineage hierarchies over time.

This method holds promise for uncovering the behaviour of rare epithelial cell types that are both transcriptionally subtle and often obscured within broader lineage contexts — such as Best4 Cells and M-Cells. These populations, while functionally distinct, share a strong transcriptomic resemblance to Enterocytes, making them difficult to resolve using standard scRNA-seq approaches alone. Their rarity is further compounded by the overwhelming abundance of canonical Enterocytes, which can dilute their signatures in bulk or unsorted single-cell datasets. By enabling longitudinal clonal tracking and integration with transcriptomic profiles, CellTagging could illuminate the dynamics of their emergence,

persistence, and plasticity. This would offer critical insights into their contributions to intestinal homeostasis, immune surveillance, and disease-associated remodelling, laying the groundwork for more targeted functional studies and potential therapeutic exploitation.

In conclusion, lineage tracing provides a powerful framework for decoding how cellular identities emerge, diverge, and adapt over time. Single-cell lineage tracing elevates this approach by coupling clonal history with molecular phenotype, revealing dynamic processes otherwise obscured in static snapshots. Applied to hIEOs, it holds exceptional promise for charting stem cell behaviour, capturing rare lineage trajectories, and uncovering context-specific plasticity — transforming hIEOs into both a model and a map of epithelial regeneration and disease.

1.4 Overall Project Aims

As discussed above, the human intestinal epithelium is a highly dynamic tissue that continuously regenerates through tightly regulated coordination of stem cell renewal, lineage commitment, and microenvironmental input. While scRNA-seq has unveiled previously unrecognised epithelial heterogeneity and revealed the transcriptional basis of many rare or disease-associated states, it remains limited in its capacity to trace lineage relationships or resolve how dynamic transitions unfold over time. Likewise, while hIEOs offer an experimentally tractable system to model epithelial biology *ex vivo*, the full potential of integrating these systems with high-resolution, multiomic, and lineage-tracing strategies remains largely untapped.

This thesis aims to bridge this gap by integrating high-resolution single cell approaches with hIEO models to uncover the regulatory logic of epithelial plasticity, with a focus on rare cell lineage representation, which display the most variable representation across experiments and datasets. The overarching goal was to develop and apply a multimodal framework — spanning scRNA-seq, multiomic profiling, and single-cell lineage tracing — to investigate how context-dependent factors shape cell fate decisions, and to evaluate whether a unifying strategy could serve as a standardised framework for generating reproducible results.

These aims are pursued through four sequential and interconnected objectives. The first was to develop and benchmark a high-throughput, multiplexed scRNA-seq screening pipeline applied to hIEOs cultured under different physiological and/or culture conditions. This method was shown to provide fast, affordable profiling of epithelial heterogeneity across different experimental settings.

The second objective was to establish a standardised analytical approach and conduct a meta-analysis of publicly available organoid datasets. This framework provided efficient and robust validation of lineage fidelity, reproducibility of protocols, and facilitated the identification of rare epithelial cell populations across multiple conditions.

Thirdly, methodological validation was performed using orthogonal single-cell platforms, including full-length RNA sequencing. Guided by insights from the meta-analysis, optimal culture conditions were identified and experimentally compared to evaluate their ability to enhance lineage resolution, detection sensitivity, and transcriptional fidelity, particularly for rare epithelial subsets. The results indicated that even minor protocol variations significantly influence transcriptomic profiles. Additionally, the analysis highlighted that precise replication of protocols does not consistently yield identical outcomes or conclusions, likely due to a combination of technical, analytical, and biological variability.

The fourth objective leveraged integrative single-cell multiomic profiling—combining RNA expression and chromatin accessibility data—to explore whether regulatory information could enhance the resolution of transitional or ambiguous epithelial states identified in previous analyses. This approach demonstrated the potential of using both modalities to validate the technical and biological limitations of current models in characterising complex lineage dynamics. However, extracting meaningful insights from chromatin accessibility data requires sequencing at substantially higher depth, presenting a trade-off between rapid and cost-effective initial screening and more comprehensive in-depth characterisation.

The fifth objective examined the feasibility of implementing CellTagging — a scalable, transcriptome-compatible lineage tracing method — to reconstruct clonal dynamics within hIEO cultures. To enable this, a custom-engineered 2D culture system was developed to facilitate efficient barcode delivery and high-resolution clonal tracking. While further

optimisation is required, this proof-of-principle demonstrates promising potential for quantifying lineage bias, identifying rare or plastic progenitor populations, and assessing how environmental perturbations influence epithelial fate trajectories.

Finally, the sixth objective evaluated the biological fidelity and practical applicability of the developed methodologies by highlighting key insights that can be derived from these experimental frameworks. This included evaluating how well transcriptomic signatures associated with intestinal region, developmental stage, disease state, and external stimuli were retained in hIEOs compared to primary epithelial tissue.

Collectively, these strategies offer a comprehensive, multidimensional approach to understanding the emergence, stability, and functional significance of epithelial lineages in health and disease contexts, positioning hIEOs as a next-generation tool for mechanistic and translational studies.

Following from this conceptual framework, this thesis is organised as follows:

- The current Chapter 1 provides a detailed introduction to human intestinal biology in health and disease, the principles and applications of organoid culture technology, and advanced single-cell methodologies.
- Chapter 2 presents a comprehensive account of the materials and experimental methods employed throughout this study.
- Chapters 3 to 5 report the results of the experimental and analytical work undertaken, outlining methodological advancements developed to achieve the stated scientific objectives, concluding with the synthesis of this work and discussion of key findings and their implications for future research directions.
- Chapter 6 concludes this thesis with a discussion of its findings, its relevance and implications for future research, and overall conclusions.

1 2 Materials and Methods

2 2.1 Materials

3 2.1.1 Consumables and Commercial Assays

4 *Table 1- Consumables and critical commercially available assays used in this study*

Consumables and Kits	**Brand**	**Catalog Number**
Costar 6.5 mm transwell 0.4 µm pore polyester membrane inserts	STEMCELL	38024
Costar 12 mm transwell 0.4 µm pore polyester membrane inserts	STEMCELL	38023
Cell Strainer 70 µm	Falcon	352350
Plasmid Plus Mega Kit	QIAGEN	12981
Cell Strainer 40 µm strainer	Falcon	10737821
Costar 24-well Plates	Corning	3524
Costar 48-well Plates	Corning	3548
TC-Treated Culture Dishes	Corning	431110
Large TC-Treated Culture Dishes	Corning	10117320
KAPA library qnt kit (Illumina) universal qPCR mix	Kapa biosystems	KK4824
RNeasy Mini Kit	QIAGEN	74106
High-Capacity RNA-to-cDNA™	Thermo Fisher Scientific	4387406
Kapa HiFi Hotstart Readymix (2x)	Roche	KK2601
QuBit dsDNA HS Assay Kit (used with QuBit 3.0)	Invitrogen	Q32851
High Sensitivity DNA reagents (Used with Agilent 2100 Bioanalyser system)	Agilent Technologies	5067-4626
HS D5000 Reagents	Agilent Technologies	5067-5593
HS D5000 Screen Tape	Agilent Technologies	5067-5592
Nextera XT kit	Illumina	FC-131-1096
Nextseq 500/550 Hi Output kit v2.5	Illumina	20024907
Novaseq 6000 S4 150bp PE reads	Illumina	20012866
Chromium Next GEM Chip M Single Cell Kit	Chromium	1000371
Chromium Next GEM Chip G Single Cell Kit	Chromium	1000127
Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle	10X Genomics	PN-1000283
Chromium Next GEM Chip J Single Cell Kit	10X Genomics	1000234
Chromium i7 Multiplex Kit	10X Genomics	PN-120262
Single Index Kit N Set A	10X Genomics	PN-1000212
Dual Index Kit TT Set A	10X Genomics	PN-1000215
DepleteX™ Single Cell RNA Boost Kit - (24 Samples)	Jumpcode	KIT1037
SMARTScribe reverse transcriptase	Takara Bio	639537
Oligonucleotide Conjugation Kit	Abcam	ab218260
Jumpcode Deplete X	JUMPCODE Genomics	KIT1018

5

1 2.1.2 Chemicals, Peptides and Recombinant Proteins

2 *Table 2 - Reagents used in this study*

Reagents	**Brand**	**Catalog Number**
Pen-Strep	Sigma-Aldrich	P0781-100ML
Hepes	Thermo Fisher Scientific	15630080
EDTA	Invitrogen	AM9260G
Pierce DTT	Thermo Fisher Scientific	A39255
FBS	Gibco	A5670701
FBS F7524	Sigma-Aldrich	F7524
Zeocin	Invitrogen	R25001
Geneticin	Gibco	10131027
HBSS	Lonza	10-543F
GlutaMAX	Gibco	35050-061
TrypLE™ Express Enzyme	Thermo Fisher Scientific	12605010
DNAse	Thermo Fisher Scientific	18047019
BSA 10%	Sigma-Aldrich	A1595
BSA 30%	Sigma-Aldrich	A9576-50ML
Tween-20 10%	Sigma-Aldrich	11332465001
DPBS	STEMCELL	37350
Buffer EB	QIAGEN	19086
SSC Buffer	Sigma	S6639-1L
Levitation Agent	LevitasBio	1003001
Matrigel growth factor reduced	Corning	356231
N-Acetyl-L-cysteine	Sigma-Aldrich	A9165
[Leu^{15}]-Gastrin I human	Sigma-Aldrich	G9145
Recombinant human R-spondin1	R&D Systems	4645-RS
Recombinant mouse Noggin	Peprotech	250-38
Recombinant mouse EGF	Thermo Fisher Scientific	PMG8043
Recombinant human IGF-1	BioLegend	590904
Recombinant human FGF-2	Peprotech	100-17A
Recombinant human NRG1	R&D Systems	5898-NR-050
Advanced Dulbecco's Modified Eagle's Medium/F12	Thermo Fisher Scientific	12634010
DMEM High Glucose	Thermo Fisher Scientific	31966
A83-01	Tocris	2939/10
Y27632 Rho-associated kinase inhibitor	STEMCELL	72307
N-2 Supplement	Thermo Fisher Scientific	17502001
B-27 Supplement	Thermo Fisher Scientific	35050061
SB202190 p38 inhibitor	Sigma-Aldrich	J60950-MB
Nicotinamide	Sigma-Aldrich	N0636-100G
CryostorCS10	Sigma-Aldrich	2874-100ML
RNaseZap RNase Decontamination Solution	Thermo Fisher Scientific	AM9780
Protector RNase Inhibitor (RNI)	Roche	3335402001
RNase Inhibitor	NEB	M0314L
Proteinase K, Molecular Biology Grade	NEB	P8107S
MgCl2	Invitrogen	AM9530G
Tris-HCl, 1M, pH=8	Thermo Fisher Scientific	15893661
Digitonin [5%]	Invitrogen	BN2006
5M NaCl	Invitrogen	AM9759
SPRIselect Bead-Based Reagent	Beckman Coulter	B23318
Ampure XP Beads	Beckman Coulter	A63880
UltraPure DNase/RNase-Free Distilled Water	Invitrogen	10977
Nuclease-free Water (not DEPC-Treated)	Thermo Fisher Scientific	AM9937

Reagents	**Brand**	**Catalog Number**
Stellar chemically competent *E.coli*	Takara Bio	636766
X-tremeGENE 9 DNA Transfection Reagent	Sigma-Aldrich	6365779001
Protamine Sulfate	Sigma-Aldrich	P3369
Gelatin Solution 2%	Sigma-Aldrich	G1393
Methanol	Thermo Fisher Scientific	A4521
DpnI	NEB	R0176S
Phusion High-Fidelity DNA Polymerase	NEB	M0530S
dNTP mix	Clontech	639125
SOC Medium	Thermo Fisher Scientific	15544034
LB Broth Tablets	Sigma-Aldrich	L7275
LB Agar Tablets	Sigma-Aldrich	L7025
Glycerol	Sigma-Aldrich	49767
Ampicillin 100 mg/mL	Sigma-Aldrich	A5354
BSA powder	Sigma-Aldrich	A8806
SSC buffer 20x	Sigma-Aldrich	S6639
NxGen RNAse inhibitor	Lucigen	30281-1
Non-Essential Amino Acids (NEAA) 100X	Thermo Fisher Scientific	11140050
SMARTScribe	Takara Bio	639538

1
2

Table 3 - Table of antibodies used in this study

Antibodies	**Source**	**Catalog Number**
TotalSeq™-A0251 anti-human Hashtag 1 Antibody	BioLegend	394601
TotalSeq™-A0252 anti-human Hashtag 2 Antibody	BioLegend	394603
TotalSeq™-A0253 anti-human Hashtag 3 Antibody	BioLegend	394605
TotalSeq™-A0254 anti-human Hashtag 4 Antibody	BioLegend	394607
TotalSeq™-A0255 anti-human Hashtag 5 Antibody	BioLegend	394609
TotalSeq™-A0256 anti-human Hashtag 6 Antibody	BioLegend	394611
TotalSeq™-A0257 anti-human Hashtag 7 Antibody	BioLegend	394613
TotalSeq™-A0258 anti-human Hashtag 8 Antibody	BioLegend	394615
APC anti-human CD326 (EpCAM)	Miltenyi Biotec Ltd	130-113-260
Brilliant Violet 421(TM) anti-human CD16	BioLegend	360723
Brilliant Violet 421(TM) anti-human CD15 (SSEA-1)	BioLegend	323039
PE anti-human CD45	BioLegend	368510
Human Trustain FcX (Fc Block)	BioLegend	422302
IgG Rabbit anti-Goat AF488	InVitrogen	Ab150141
Anti-Best4 Antibody	Atlas Antibodies	HPA058564
Purified anti-human CD326 (EpCAM) clone C9	BioLegend	369802
DAPI (-4′,6-diamidino-2-phenylindole)	BioLegend	422801
ZOMBIE Aqua	BioLegend	423102

Table 4 - Table of oligos used in this study

Custom Oligos	**Sequence**	**Source**
SI-PCR_primer Sequence	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC*T* C	IDT
HTO_additive_primer_v2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C* T	IDT
D701_LONG	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	IDT
D702_LONG	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	IDT
D703_LONG	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	IDT
D704_LONG	CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T	IDT
D705_LONG	CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGA T*C*T	IDT
D706_LONG	CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGA T*C*T	IDT
D707_LONG	CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGA T*C*T	IDT
D708_LONG	CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACGTGTGCTCTTCCGA T*C*T	IDT
ADT1 oligo	/5AmMC12/CCT TGG CAC CCG AGA ATT CCA CAT GAT TGG CTC AAA AAA AAA AAA AAA AAA AAA AAA AAA A*A*A	IDT
ADT2 oligo	/5AmMC12/CCT TGG CAC CCG AGA ATT CCA GAG GCG ATT GAT AAA AAA AAA AAA AAA AAA AAA AAA AAA A*A*A	IDT
ADT3 oligo	/5AmMC12/CCT TGG CAC CCG AGA ATT CCA TGT CCG GCA ATA AAA AAA AAA AAA AAA AAA AAA AAA AAA A*A*A	IDT
ADT4 oligo	/5AmMC12/CCT TGG CAC CCG AGA ATT CCA TGG TGA ACC TGG AAA AAA AAA AAA AAA AAA AAA AAA AAA A*A*A	IDT
ADT5 oligo	/5AmMC12/CCT TGG CAC CCG AGA ATT CCA GAT CGT AAT ACC AAA AAA AAA AAA AAA AAA AAA AAA AAA A*A*A	IDT
ADT6 oligo	/5AmMC12/CCT TGG CAC CCG AGA ATT CCA AAG CGC TTG GCA AAA AAA AAA AAA AAA AAA AAA AAA AAA A*A*A	IDT
ADT7 oligo	/5AmMC12/CCT TGG CAC CCG AGA ATT CCA CAT CGG TGT ACA AAA AAA AAA AAA AAA AAA AAA AAA AAA A*A*A	IDT
ADT8 oligo	/5AmMC12/CCT TGG CAC CCG AGA ATT CCA GTC TAG ACT TCG AAA AAA AAA AAA AAA AAA AAA AAA AAA A*A*A	IDT
Forward CellTag library sequencing primer	5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT CATGGACGAGCTGTACAAGTAA-3′	IDT
Reverse CellTag library sequencing primer	5′-CAAGCAGAAGACGGCATACGAGATACAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGT GCAGGGGAAAGAATAGTAGAC-3′	IDT

Legend: The asterisk (*) indicates a phosphorothioate (PS) bond.

2.1.3 Solutions, Buffers and Media

Table 5 - List of reagents for the preparation of Transport Media

Reagent	**Final Concentration**
HEPES	10 mM
Penicillin - Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin
In HBSS	-

Table 6 - List of reagents for the preparation of HPGA Buffer

Reagent	**Final Concentration**
Hepes	10 mM
Penicillin - Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin
In HBSS	-

Table 7 - List of reagents for the preparation of Chelation Media

Reagent	**Final Concentration**
EDTA	1 mM
DTT	1 mM
FBS	1%
In HPGA Buffer	-

Table 8 - List of reagents for the preparation of Staining Buffer

Reagent	**Final Concentration**
Tween-20 10%	0.01%
BSA	2%
In PBS	-

Table 9 - List of reagents for the preparation of "Fujii" basal media, as per Fujii M. et. al (2018)

Reagent	**Final Concentration**
Hepes	10 mM
GlutaMAX	2 mM
B27 Supplement	1x
Penicillin - Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin
In Advanced Dulbecco's Modified Eagle's Medium/F12	-

Table 10 - List of reagents for the preparation of "Fujii" conditioning media, as per Fujii M. et. al (2018)

Reagent	**Final Concentration**
Recombinant mouse EGF	50 ng/mL
Recombinant mouse Noggin	100 ng/mL
Recombinant human IGF-1	100 ng/mL
Recombinant human FGF-2	50 ng/mL
Recombinant human R-spondin1	1 µg/mL
A83-01	500 nM
N-Acetyl-L-cysteine	1 mM
[Leu^{15}]-Gastrin I human	10 nM
WREN medium	50%
In IGF-FGF Basal Media	-

Table 11 - List of reagents for the preparation of "Fujii" differentiation media, as per Fujii M. et. al (2018)

Reagent	**Final Concentration**
Recombinant mouse Noggin	100 ng/mL
Recombinant human IGF-1	100 ng/mL
Recombinant human FGF-2	50 ng/mL
Recombinant human R-spondin1	1 µg/mL
A83-01	500 nM
N-Acetyl-L-cysteine	1 mM
[Leu^{15}]-Gastrin I human	10 nM
WREN medium	50%
In IGF-FGF Basal Media	-

Table 12 - Preparation of "House" Conditioning Media, adapted from Sato T et. al (2011)

Reagent	**Final Concentration**
GlutaMAX	2 mM
HEPES	10 mM
Penicillin - Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin
N-2 supplement	1x
B27 Supplement	1x
Nicotinamide	10 mM
A83-01	0.1 µM
N-Acetyl-L-cysteine	1 mM
[Leu^{15}]-Gastrin I human	1 µg/mL
SB202190	10 µM
Y27632*	10 µM
R-Spondin1 conditioned medium	10%
Noggin conditioned medium	10%
WREN medium	50%
In Advanced Dulbecco's Modified Eagle's Medium/F12	-

*Legend: * Added only right after plating organoids to enhance cell recovery and proliferation.*

Table 13 - List of reagents for the preparation of "House" Differentiation Media, adapted from Sato T et. al (2011)

Reagent	**Final Concentration**
GlutaMAX	2 mM
HEPES	10 mM
N-2 supplement	1x
B27 Supplement	1x
Nicotinamide	10 mM
A83-01	0.1 µM
N-Acetyl-L-cysteine	1 mM
[Leu^{15}]-Gastrin I human	1 µg/mL
R-Spondin1 conditioned medium	10%
Noggin conditioned medium	10%
WNT3A conditioned medium	5%
Advanced Dulbecco's Modified Eagle's Medium/F12	-

Table 14 - List of reagents for the preparation of "Holloway" conditioning media, as per Holloway E. et al (2021)

Reagent	**Final Concentration**
GlutaMAX	4 mM
HEPES	20 mM
Penicillin - Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin
N-2 supplement	2x
B27 Supplement	2x
Nicotinamide	20 mM
N-Acetyl-L-cysteine	2 mM
[Leu^{15}]-Gastrin I human	1 µg/mL
Advanced Dulbecco's Modified Eagle's Medium/F12	-

Table 15 - List of reagents for the preparation of "Holloway" differentiation media, adapted from Holloway E. et al (2021)

Reagent	**Final Concentration**
Recombinant human NRG1	100 ng/mL
Wnt	15%
Nog	5%
rsp	5%
NRG1 Basal Media	75%

Table 16 - List of reagents for the preparation of Lysis Buffer (with Digitonin)

Reagent	**Final Concentration**
Digitonin	0.01%
Tris-HCl @ pH 7.4	20 mM
NaCl	150 mM
$MgCl_2$	3 mM
RNase Inhibitor	2 U/µL
In UltraPure RNAse/DNAse Free Water	-

Table 17 - List of reagents for the preparation of Washing Buffer (after Digitonin Lysis)

Reagent	**Final Concentration**
Tris-HCl @ pH 7.4	20 mM
NaCl	150 mM
$MgCl_2$	3 mM
RNase Inhibitor	2 U/µL
In UltraPure RNAse/DNAse Free Water	-

Table 18 - List of reagents for the preparation of Virus Production Medium

Reagent	**Final Concentration**
FBS	10%
Sodium Pyruvate	1 mM
NEAA	1 mM
In DMEM	-

2.1.4 Cell Lines

Table 19 - List of cell lines used in this study

Cell line	**Notes**
Caco2	Used passages P30-P50
HT29	Used passages P30-P50
HEK293T – Recombinant Noggin	Used passages P1-P5 HEK293T stably transfected with pcDNA3.Neo insert mouse Noggin. C-terminal tag is human IgG1 –Fc.
L cells - Recombinant Wnt3A	Used passages P1-P5 L-cells stably transfected with pCDNA3.1Zeo-mouse Wnt3a MvdW.
Cultrex® HA-R-Spondin1-Fc 293T Cells – Recombinant R-Spondin1	Used passages P1-P5 HA-R-Spondin1-Fc 293T Cells. N-terminal HA epitope tag and fused to a C-terminal murine IgG2a Fc fragment.

2.1.5 Recombinant DNA: CellTagging Libraries and Plasmids

Table 20 - List of recombinant DNA sources used in this study

Plasmids and libraries	**Catalog number**
pCMV-VSV-G	Addgene Plasmid ID #8454
pCMV-dR8.2 dvpr	Addgene Plasmid ID #8455
Pooled CellTag V1 library	Addgene #115643
Pooled CellTag V2 library	Addgene #115644
Pooled CellTag V3 library	Addgene #115645

2.2 Methods

2.2.1 Processing of Intestinal Samples

Tissue samples from adult participants were collected under two conditions: from patients undergoing routine colonoscopic screening without an inflammatory diagnosis (referred to as Healthy Controls) and from patients receiving clinical care for IBD who exhibited active inflammation. In patients with a confirmed histological diagnosis of UC, biopsies were obtained from both inflamed and adjacent non-inflamed regions. Detailed information about the adult tissue samples is provided in the Appendix (Supplementary Methods, Section 1.6).

For experiments involving both scRNA-seq and organoid cultures, 3-4 pairs of biopsies from the cohort were collected. These samples were immediately placed in Transport Media (Table 5) and kept on ice during transport to the laboratory. Upon arrival, each tissue sample was transferred to 10 mL of chelation medium pre-warmed to 37°C (Table 6 and Table 7) and incubated in a water bath at 37°C for 15 minutes. The samples were agitated briefly (~10 seconds) every 5 minutes using a vortex. After incubation, the supernatant containing the initial fraction of epithelial crypts was collected, centrifuged at 300g for 4 minutes, and the pellet was resuspended in 2 mL of ice-cold Transport Media. This tube was then placed on ice until all collected epithelial fractions were combined.

The remaining tissue was further processed by adding 5 mL of fresh chelation medium, also pre-warmed to 37°C, and repeating the incubation and agitation process two additional times. The collected epithelial fractions from these steps were then combined and centrifuged at 300g for 4 minutes. The final pellet was resuspended in TrypLE Express Enzyme solution containing 50 µg/mL DNAse and incubated at 37°C with continuous agitation (MACSMix, Miltenyi Biotec). This suspension was then filtered through a 70 µm cell strainer. If incomplete

dissociation was observed, indicated by the presence of cell clumps, an additional filtration step was performed using either another 70 µm strainer or a 40 µm strainer. Cell concentration and viability were assessed using both a Countess II automated cell counter (Thermo Fisher) and/or manually with a haemocytometer as needed. Detailed methodologies and key resources have been documented in a previously published protocols[1]. The resulting epithelial cell-enriched suspension was immediately processed further for scRNA-seq, fluorescence-activated cell sorting (FACS), organoid culture, or other required techniques.

2.2.2 Intestinal Epithelial Organoid Cultures

Throughout this study, organoid media were prepared in-house using several key formulations. These included media prepared entirely from basic components following adaptations on the methodology described by *Sato T. et al. (2011)*, *Fujii M. et al. (2018)* and *Holloway E. et al. (2021)*[88,126,309]. For clarity and ease of reference within this dissertation, these growth formulations are hereafter referred to as "House," "Fujii" and "Holloway" media, respectively.

2.2.2.1 Preparation of Conditioning Media

Conditioned media used in this study was produced in-house from genetically modified cell lines, as fully described elsewhere[38,41,47]. Briefly, production involved three distinct protocols: Wnt3A was produced from L-cells transfected with pCDNA3.1Zeo-mouse Wnt3a. These cells were cultured in DMEM supplemented with FBS and Penicillin/Streptomycin (Pen/Strep) and maintained under Zeocin selection. Noggin was obtained from HEK-293 cells transfected with pcDNA3 NEO, which contains a mouse Noggin gene fused to a human IgG1-Fc tag. These cells were also cultured in DMEM enriched with FBS and pen/strep and selected using Geneticin. R-Spondin 1 was generated from 293T cells expressing murine Rspo1 with a C-terminal murine

IgG2a Fc fragment. Cultivation was in DMEM supplemented with necessary additives and maintained under Zeocin selection for stringent control. The protocol used to produce R-Spondin1 conditioned media (Production of R-Spondin 1-conditioned Medium using Cultrex® HA-R-Spondin1-Fc 293T Cells) is a publicly available protocol which can be found elsewhere[310]. The remaining protocols can be found in the Appendix (Supplementary Methods, Sections 1.1 and 1.2).

2.2.2.2 3D Organoid Cultures

Single-cell epithelial suspensions were embedded in Matrigel and plated onto 24-well and 48-well plates as 50 µL and 25 µL droplets, respectively. Following the polymerization of the Matrigel, the crypts were covered with the specified culture medium. House media were prepared according to Table 12 and Table 13 for conditioning and differentiation purposes, respectively. Fujii's media involved preparing expansion media as detailed in Table 9 and Table 10, while the differentiation media were adjusted by omitting EGF as outlined in Table 9 and Table 11. Holloway media formulations were followed as described in Table 14 and Table 15. Organoids were passaged weekly using a mix of mechanical and enzymatic dissociation methods, involving cycles of incubation in TrypLE™ Express Enzyme at 37°C and vortexing. Depending on the proliferation rate, organoids were split at ratios of 1:1, 1:3, or 1:5. The culture medium was enhanced with 10 µM Y-27632 during the initial two days post-passage.

2.2.2.3 Organoid Monolayers

Organoid monolayers were prepared by modifying the protocol provided by STEMCELL Technologies[311]. Briefly, organoids were harvested and dissociated using the combined mechanical and enzymatic methods outlined in the previous section (2.2.2.2). The seeding density was optimised to ensure the formation of a perfect monolayer. Prior to seeding, the culture surfaces of the transwell inserts were coated with Matrigel. The dissociated cells were

then seeded onto these prepared surfaces and cultured for a period that had also been optimised to promote monolayer formation. The integrity and confluency of the monolayers were assessed through transepithelial electrical resistance (TEER) measurements and visual inspection.

2.2.2.4 Organoid Stimulations

Stimulants were added to the organoid cultures as required, depending on the specific experimental design. RANKL was supplemented at a concentration of 200 ng/mL prior to organoid harvesting, in accordance with previous research findings. Sodium butyrate was applied at a concentration of 5 mM based on prior unpublished work. IL4 and IL13 were supplemented at a concentration of 400 ng/mL and 40 ng/mL, respectively, as per previous publications.

2.2.3 Antibody Conjugation

Custom oligo-antibody conjugation was performed to obtain antibody-derived tags (ADTs) following The Oligonucleotide Conjugation Kit (Abcam, ab218260). Briefly, a custom barcoded hashing antibody was generated by combining an EpCAM antibody with custom design oligos (Table 4).

2.2.4 Antibody Staining

Following resuspension in staining buffer (Table 8), primary antibody staining (Best4 - 1:100, EpCAM – 1:100) was performed for 30 minutes at 4°C, followed by blocking with Trustain FcX (Biolegend) for 10 minutes. When required, a secondary antibody staining was performed similarly to the primary (Rabbit anti-Goat AF488 - 1:400). Finally, a 10-minute incubation of the

cell suspension with DAPI or Zombie Aqua (1:1000) was performed. Cells were washed thrice in staining buffer between each step to remove unbounded antibodies.

2.2.5 Flow Cytometry and Cell Sorting

Cell quantification and/or cell sorting (FACS) was used in specific experiments for sorting of populations of interest (Best4^{+} and GFP^{+} cells). Primary cells or organoid-derived cells were processed and stained as described above (Sections 2.2.1 and 2.2.2, respectively). Flow cytometry quantification analysis was conducted in a BD LSR Fortessa X20, while cell sorting required a SonySH800 or a BD FACSARIA III. To accurately delineate population gates, compensations were regularly performed with singularly stained samples and fluorescence minus one (FMO) samples. Data generated was analysed with FlowJo (FlowJo v10.5.3).

2.2.6 DNA and RNA techniques

For quantitative RT–PCR experiments with larger cell numbers (more than 25,000 cells), total RNA was isolated using the RNeasy microkit, according to manufacturer's instructions. cDNA was then synthesized using the high-capacity RNA-to-cDNA kit, with RT–PCR then performed using applicable Taqman gene expression assays on the QuantStudio 7-Flex system (ThermoFisher).

2.2.7 Cell Tagging

A detailed, step-by-step summary of the CellTagging protocol, including intermediate results, is provided in the Appendix (Supplementary Methods, Section 1.5). Key aspects of the procedure are summarised below.

2.2.7.1 Processing Packaging Plasmids

Bacterial stabs containing the pCMV-dR8.2 and pCMV-VSV-G plasmids were streaked onto LB agar plates containing 100 μg/mL ampicillin to isolate bacterial colonies. Individual colonies were used to inoculate overnight LB cultures for plasmid amplification and glycerol stock preparation, following Addgene's standard protocols [313,314]. Glycerol stocks were prepared from these cultures and stored at −80 °C for long-term preservation of the plasmid constructs. Plasmid DNA for the lentiviral packaging components pCMV-VSV-G and pCMV-dR8.2 dvpr was purified using the Qiagen Plasmid Plus Maxiprep Kit following the manufacturer's protocol. These plasmids were subsequently used for lentivirus production.

2.2.7.2 Generating Pooled Cell Tag Libraries

Pooled CellTag libraries (V1, V2 and V3, Table 20) were amplified using Stellar chemically competent cells. For each library, 100 μL aliquots of competent cells were combined with 10 ng of pooled CellTag plasmid DNA. The mixtures were gently mixed by pipetting and incubated on ice for 30 minutes. Heat shock was performed at 42 °C for 60 seconds, followed by immediate cooling on ice for 1 minute. Subsequently, 1 mL of pre-warmed SOC medium was added to each transformation tube, and samples were incubated at 37 °C with shaking at 250 rpm for 1 hour. Following recovery, serial dilutions (1:10, 1:100, and 1:1000) of the transformation mixtures were prepared in LB broth, and 50 μL of each dilution was plated onto LB agar plates supplemented with 50 μg/mL ampicillin. Concurrently, the remaining transformation mixtures were transferred into 500 mL of LB broth containing 100 μg/mL ampicillin in 1 L Erlenmeyer flasks and incubated overnight at 37 °C with shaking. This approach enabled both colony quantification and the expansion of successfully transformed bacteria for downstream plasmid extraction. Following transformation and plating, colony-forming units (CFUs) were quantified

to ensure sufficient library complexity. Colony counts from dilution plates were used to calculate CFU/mL using the equation:

$$\frac{CFU}{mL} = \frac{Number\ of\ Colonies \times Dilution\ Factor}{Volume\ Plated}$$

The transformation efficiency (TE) was subsequently calculated using:

$$TE = \frac{Number\ of\ Colonies \div Amount\ of\ DNA\,(\mu g)}{Dilution},$$

where:

$$Dilution = \frac{Initial\ Volume\ Sampled}{Initial\ Volume} \times \frac{Diluted\ Volume\ Plated}{Total\ Diluted\ Volume}$$

Overnight liquid cultures from the transformed colonies were harvested, and plasmid DNA was purified using the QIAGEN Plasmid Plus Mega Kit (QIAGEN), following the manufacturer's protocol. Libraries were eluted in 1 mL of Elution Buffer. Purified DNA was quantified using a NanoDrop spectrophotometer. Pooled plasmid libraries were stored at −20 °C for long-term use.

2.2.7.3 Amplifying Cell Tag Libraries

To amplify each pooled CellTag library (V1, V2, V3), PCR reactions were set up using a 3x master mix formulation with custom sequencing primers. First, a master mix was prepared (2× KAPA HiFi HotStart ReadyMix PCR mix, 10 µM of Forward Sequencing Primer and 10 µM of Reverse Sequencing Primer) and mixed by pipetting. For each CellTag variant (V1, V2, V3), individual PCR reactions were assembled by combining 500 ng of the corresponding CellTag plasmid library (calculated from the concentration determined in the previous step), 28 µL of the

previously prepared master mix, and nuclease-free water to a final reaction volume of 50 µL. The mixtures were pipette-mixed for subsequent thermal cycling. Thermal cycling was performed by setting: 1) initial denaturation at 95 °C for 3 minutes; 2) 20 amplification cycles of 98 °C for 20 seconds, 65 °C for 15 seconds and 72 °C for 20 seconds; 3) final extension at 72 °C for 1 minute; and 4) hold at 4 °C indefinitely. Purified PCR products were cleaned using AMPure XP beads (Beckman Coulter) in a 0.6X ratio (as per manufacturer's instructions) to remove primer dimers and small fragments. Library quality and size distribution were assessed on the Agilent Tapestation with the High Sensitivity D5000 ScreenTape, following manufacturer's instructions.

2.2.7.4 Assessment of Cell Tag Library Complexity

Final libraries were first validated by Sanger sequencing to confirm successful amplification and correct insert identity. Subsequently, pooled CellTag libraries were sequenced on an Illumina MiSeq platform to assess library complexity and barcode diversity. This sequencing approach provided sufficient read depth for evaluating the representation and quality of all three pooled CellTag libraries. Following sequencing, a custom whitelist of around 60,000 unique, high-confidence CellTag barcodes was generated from the raw read data. This whitelist was then used for downstream demultiplexing and lineage tracing analyses.

2.2.7.5 Production of Cell Tag Lentivirus

On day 1, HEK 293T cells were seeded into 100-mm tissue culture dishes at an initial confluency of approximately 50–60%, corresponding to ~5 million cells per plate. Cells were cultured in Virus Production Medium supplemented with pen/strep (1:1000). On day 2, approximately two hours prior to transfection, the culture medium was replaced with fresh Virus Production Medium (without antibiotics). For each CellTag library, two transfection mixtures were

prepared. In the first, 15 µL of X-tremeGENE 9 DNA transfection reagent was diluted in 185 µL of pure DMEM. In the second, 2 and 4 µg of the respective CellTag plasmid, 2 µg of pCMV-dR8.2 dvpr, and 200 ng of pCMV-VSV-G were combined and adjusted to 200 µL with DMEM. The DNA mixture was added dropwise to the transfection reagent tube, gently mixed by pipetting, and incubated at room temperature for 15 minutes. The transfection mixture was then added dropwise to the plated 293T cells, which were gently agitated to ensure even distribution, and incubated overnight at 37°C. On day 3, media were aspirated and replaced with fresh virus production medium (without antibiotics). Viral supernatant was collected on day 4 and day 5 post-transfection. Each collection was filtered through a 0.45 µm low-protein-binding syringe filter to remove cellular debris. Collected virus was at −80°C for long-term storage. Titration was performed using flow cytometry following a standard lentivirus titration protocol (Merck)[315].

2.2.7.6 Transduction of Organoid Epithelial Cells with CellTag Libraries

hIEOs were dissociated into single-cell suspensions using gentle enzymatic dissociation (e.g. TrypLE Express) and filtered through a 40 µm strainer. Cells were counted and resuspended in organoid culture medium supplemented with Rho-associated protein kinase (ROCK) inhibitor (10 µM Y-27632) to enhance post-transduction viability. Lentiviral CellTag libraries, previously titrated on HEK293T cells to determine infectious titre, were added at a low multiplicity of infection (MOI ~0.1) to favour single-vector integration events per cell. Polybrene was added at a final concentration of 8 µg/mL to enhance infection efficiency. The cell-virus mixture was gently agitated and incubated for 6–8 h at 37°C and 5% CO_2 before pelleting and washing twice with fresh organoid culture medium. Transduced cells were then embedded in Matrigel domes and cultured in standard organoid medium. Expression of CellTags was allowed to establish for 48 h prior to downstream assays or single-cell library preparation.

2.2.7.7 Concentrating CellTag Lentiviral Preparations

Lentiviral particles containing CellTag libraries were generated by transient transfection of HEK293T cells with the CellTag plasmid pool and standard packaging vectors (psPAX2 and pMD2.G) in Opti-MEM using Protamine Sulphate. Supernatants were harvested at 48 h and 72 h post-transfection, clarified by centrifugation at 500 × g for 10 min, and passed through 0.45 µm PVDF filters to remove cellular debris. Viral supernatants were then concentrated 6x by ultracentrifugation at 100,000 × g for 2h at 4°C in a swinging-bucket rotor according to the manufacturer's instructions. Pelleted virus was carefully resuspended in sterile PBS and stored at −80°C in single-use aliquots. Viral titres were estimated by functional transduction of HEK293T and hIEOs lines.

2.2.8 Single-Cell Sequencing Pipelines

2.2.8.1 Chromium scRNA-seq (3' Gene Expression)

Following antibody-based hashtag staining, cells were centrifuged and re-suspended in PBS supplemented with 0.04% BSA. For each reaction, an input of approximately 40,000 single cells per pool was loaded in each channel of the 10X Chromium Controller or Chromium X single-cell platforms (10X Genomics), with a recovery rate of approximately 8,000-10,000 cells per reaction. Gene Expression and Protein library preparation were carried out as per manufacturer's instructions (*10X Genomics, CG000388 - Single Cell 3' Reagent Kits User Guide (v3.1 Chemistry Dual Index) with Feature Barcoding technology for Cell Multiplexing*, and *10X Genomics, CG000419 - Chromium Next GEM Single Cell 3' HT Reagent Kits v3.1 (Dual Index) with Feature Barcode technology for Cell Multiplexing*). Library quality and quantification were assessed in Bioanalyzer 2100 (Agilent) using High Sensitivity DNA chips, and final library concentrations measured in Qubit 4.0 using Qubit dsDNA High Sensitivity Assays (ThermoFisher).

2.2.8.2 Chromium scMultiome-seq – Pooling and permeabilization

Following antibody-based hashtag staining, cells from 2-8 distinct biological samples were pooled and processed according to an adaptation of the DOGMA-seq and ASAP-seq methods and the 10X Genomics protocol (10X Genomics, CG000338, Rev F)[257]. Briefly, cells were resuspended in a permeabilization buffer (Lysis Buffer, Table 16). Incubation was performed on ice for 7 minutes. Reactions were quenched by adding 1 mL of chilled wash buffer (Table 17), followed by gentle mixing. Cells were pelleted by centrifugation at 500g for 5 minutes at 4 °C, resuspended in wash buffer, and counted using Trypan Blue exclusion with a Countess II FL Automated Cell Counter. For each reaction, an input of approximately 40,000 single cells per pool was added to each channel of the Chromium X single-cell platform (10X Genomics) with a recovery rate of approximately 10,000 cells.

2.2.8.3 Chromium scMultiome-seq – Downstream Processing

Cell suspensions were processed using the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression platform (10x Genomics, User Guide CG000338 Rev A), with several modifications based on the ASAP-seq protocol. 1) Tagmentation Input: Around 40,000 cells were used per tagmentation reaction. When sufficient material was available, two independent reactions were set up per pool to increase total cell recovery; 2) ADT Primer Spike-In: During pre-amplification (Step 4.1), 1 µL of a 0.2 µM ADT/HTO additive primer (sequence: CCTTGGCACCCGAGAATTCC) was added to enable parallel capture of protein tag information; 3) Sample Split for Library Prep: Following SPRI cleanup (Step 4.3.k from the 10x Genomics Protocol), beads were eluted in 100 µL EB buffer. The eluted sample was split for downstream library preparation: 25% was allocated to ATAC indexing, 35% to cDNA amplification, and 35% to ADT library preparation. 4) ADT Library Amplification: Protein tag libraries were amplified from the allocated 35% sample using SI-PCR and RPI-x indexing primers (TruSeq Small RNA

handle-based primers with sample-specific barcodes). PCR conditions were: 95 °C for 3 minutes; 10 cycles of 95 °C for 20 seconds, 60 °C for 30 seconds, and 72 °C for 20 seconds; final extension at 72 °C for 5 minutes. Final ATAC, gene expression, and ADT/HTO libraries were quantified using a Qubit dsDNA High Sensitivity Assay (ThermoFisher) and quality-checked on a Bioanalyzer 2100 (Agilent) using High Sensitivity DNA chips. Specific protocol modifications and all primers used for surface library construction can be found in Table 4.

2.2.8.4 Smart-seq2

For Smart-seq2, single cells were sorted as described in Section 2.2.5, and plate-based scRNA-seq was performed as per the original protocol, with minor adaptations[269]. Briefly, for RNA amplification, reverse transcription was performed using 0.75U of SMARTScribe reverse transcriptase per reaction, followed by PCR pre-amplification with 5′-biotinylated IS PCR primers (Biomers) for 25 cycles. Amplicons were purified using Ampure XP beads at a 0.8:1 bead-to-cDNA ratio. Eluted cDNA was resuspended in elution buffer and assessed for concentration and fragment size using a High Sensitivity DNA kit. Indexed libraries were prepared using the Nextera XT kit on the Biomek FXp automation platform and pooled for sequencing on an Illumina NextSeq system.

2.2.8.5 Mitochondrial and Ribosomal Depletion (Chromium scRNA-seq and Chromium scMultiome-seq)

To further enhance sequencing quality, the Jumpcode Deplete X kit was used according to the manufacturer's instructions, for both 3’ Gene Expression and Multiome assays. By leveraging Cas9 + guide RNA nucleic acid depletion, this method allowed precise removal of ribosomal, mitochondrial mRNA, and non-transcriptomic reads.

2.2.8.6 Sequencing

Final scRNA-seq or scMultiome-seq gene expression libraries were prepared at a concentration of 4nM and subsequently denatured and diluted for analysis on an Illumina Nextseq 550 using 150 base-pair paired end reads with sequencing depth determined on expected cell recovery. These were subsequently sequenced together with HTO libraries, to a depth of 40,000 reads per cell and 5% of gene expression depth, respectively, and addition of 1-5% phiX.

V1, V2 and V3 CellTag amplicon libraries were prepared at a concentration of 4nM, denatured and diluted for analysis on an Illumina Miseq using 150 base-pair single end reads and 25% phiX.

Sequencing of the ATAC libraries obtained in the last CellTagging Multiome experiment were diluted for analysis on a NovaSeq platform, with a target of ~155,000 total reads.

2.2.9 Computational Processing and Analyses

2.2.9.1 Preprocessing of Raw Data (scRNA-seq and scMultiome-seq)

Raw sequencing reads were initially assessed for quality using *FastQC*[316]. Alignment and quantification of unique molecular identifiers (UMIs) were performed with Cell Ranger (v7.2) or Cell Ranger ARC (v2.0.2) from 10x Genomics, using the GRCh38 human reference genome (10x reference: *refdata-gex-GRCh38-2020-A*), retrieved from the official 10x Genomics software portal[317]. Cell calling was conducted using the emptyDrops function from the *DropletUtils* package (v1.26.0), which distinguishes true cells from empty droplets based on ambient RNA signatures within the 10x barcode whitelist[318]. During quality control, cells were first filtered based on statistical significance and RNA content to remove low-confidence transcriptomes. Specifically, cells were retained only if they exhibited an adjusted false discovery rate (FDR) below 0.05 and contained more than 200 detected RNA molecules, ensuring exclusion of low-

quality or poorly sequenced events. Subsequent thresholds for identifying low-quality barcodes (e.g. droplets with low total RNA or macs2 UMIs) and potential multiplets (high total UMIs) were defined empirically for each pool based on observed UMI count distributions and manually applied.

2.2.9.2 Demultiplexing of hashed samples and multiplet removal (scRNA-seq and scMultiome-seq)

For multiplexed experiments, hashing antibody libraries were processed as feature barcoding data alongside scRNA-seq or scMultiome-seq libraries. ADTs and/or HTOs were counted using a custom feature reference containing the corresponding TotalSeq antibody barcode sequences[319]. Unfiltered UMI count matrices were then imported into R (v4.4.1) for downstream processing. Hashtag antibody UMI count matrices were filtered to retain only cell barcodes that passed quality control based on mRNA expression data. Briefly, hashtag demultiplexing was performed using the HTODemux function (Seurat, assay = “ADT” or "HTO”) with a stringent positive quantile threshold of 0.99. This high quantile cutoff was applied to maximise confidence in sample assignment by restricting positive hashtag calls to cells with the strongest antibody-derived tag (ADT) or hashtag oligonucleotides (HTO) signals. Together, these steps ensured robust quality control by removing low-quality transcriptomes and enabling accurate assignment of cells to their corresponding experimental conditions. For each individual pool, antibody tags not included in the specific multiplexing design and non-hashtag antibody counts were excluded. UMI counts for each hashtag were normalised using a centred log-ratio (CLR) transformation[320]. An initial clustering was then performed using CLARA KMeans or KMedoids clustering (where the number of clusters k was set to one plus the number of expected samples in the pool, i.e., k = N_samples + 1)[321]. For each antibody tag, a tag-specific threshold was defined as the 99th percentile of the normalised counts under a

fitted negative binomial distribution. Cells with expression above this threshold were considered positive for the corresponding tag. Each cell was assigned to a single sample-of-origin based on its individual tag expression profile. Cells positive for multiple tags were classified as doublets or multiplets and removed from further analysis. Given random mixing of samples in the pool, most doublets are expected to arise from inter-sample collisions, which can be effectively identified and excluded using this barcoding strategy. Additionally, a minor subset of cells did not exceed threshold for any hashtag and were classified as unassigned. These untagged cells typically exhibited lower total mRNA content and did not correspond to specific transcriptional clusters, suggesting they represented dying cells, damaged nuclei, or unstained contaminants, and were also excluded from downstream analyses. Demultiplexed cells were visualised using t-SNE on Euclidean distance matrices of CLR-normalised hashtag expression[322]. When relevant, further validation of demultiplexing accuracy was conducted to verify that sex-specific gene expression (e.g. *XIST*) segregated as expected by sample-of-origin.

2.2.9.3 Preprocessing Gene Expression Libraries (plate-based scRNA-seq)

Smart-seq2 libraries were sequenced on an Illumina platform to generate paired-end reads. Raw sequencing files were first assessed for read quality using FastQC. Illumina Nextera sequencing adapters, Smart-seq2 oligonucleotide sequences, poly-d(T) and poor-quality (fewer than 20-base-pair) sequences were trimmed using Cutadapt software. Trimmed reads were then aligned to the human genome reference (GRCh38) using STAR software. Gene-level raw read quantification was performed with the featureCounts tool, generating a raw count matrix. Data-quality metrics for each well were aggregated using multiQC tool. Downstream analyses were carried out in R (v4.4.1). The R package *scater* was used to process raw count data. Poor-quality wells were filtered on the basis of the following criteria: less than 60% of reads uniquely mapped, less than 500 genes detected, more than 20% of reads mapping to

mitochondrial features. Library normalization size factors were computed using the R package *SCNorm*.

2.2.9.4 Preprocessing Gene Expression Libraries (droplet-based scRNA-seq)

Distributions of RNA and HTO UMI counts per cell were visualised using the *ggplot2* package (v3.5.1). Data normalisation, dimensionality reduction, and clustering were performed using Seurat (v5.1.0). Normalisation was applied per cell by scaling gene expression counts to total UMIs. Highly variable genes were identified by modelling the mean-variance relationship. Principal component analysis (PCA) was used for dimensionality reduction, and clustering was performed via the Louvain algorithm on a k-nearest neighbour (kNN) graph. UMAP was used for visualisation (parameters: n.neighbors = 30, spread = 1, min.dist = 0.1). For integrated analysis, cells from different pools or reactions were merged into a single Seurat object. Batch effects—primarily associated with 10x reactions—were corrected using the *Harmony* algorithm (v1.2.3), allowing adjustment for technical variation without compromising biological heterogeneity. UMAP and clustering analyses were then repeated using Harmony embeddings in place of PCA. Consistency in cell type composition was confirmed by comparing merged clusters with those derived from individual pools. In analyses where focus was on a specific subset of cells or a biological variable, the reverse approach was employed: large, multiplexed datasets were stratified into smaller subsets for focused re-analysis.

2.2.9.5 Preprocessing Multiome Libraries (scMultiome-seq)

Multiomic datasets were processed using both the Cell Ranger ARC (ATAC-seq data) and Cell Ranger (Gene Expression and hashing data) pipelines and subsequently integrated on the basis of shared cellular barcodes and UMI-derived features. Raw chromatin accessibility data sequencing reads were processed with Cell Ranger ARC (version 2.0.2) and reads were aligned to the GRCh38 human reference genome using the multiome-compatible reference package

(refdata-cellranger-arc-GRCh38-2020-A). The ATAC component was then filtered, and peak-counted using *cellranger-arc count*, generating high-quality chromatin-accessibility matrices and associated metadata for each GEM well. In parallel, gene-expression and antibody-derived tag (ADT) libraries were processed with *cellranger count* to obtain transcriptomic and surface-protein UMI matrices. All outputs were imported into R using the *Seurat* (v5.1.0) and *Signac* (v1.13.0) packages. After independent quality control and normalisation, chromatin-accessibility and GEX+ADT datasets were merged by exact barcode matching to ensure that only droplets detected across modalities were retained. A multi-assay Seurat object was created for each sample, containing the following assays: 1) "RNA": gene expression counts; 2) "ATAC": chromatin accessibility fragments via *CreateChromatinAssay*; 3) "ADT" or “HTO”: hashing antibody counts for demultiplexing.

This UMI-based integration strategy allows each cell to be represented by its corresponding RNA, protein, and ATAC profiles, enabling joint downstream analyses such as multiomic clustering, regulatory inference, and peak–gene linkage. By anchoring the integration on shared barcodes rather than post-hoc matching, the merged dataset preserves true single-cell correspondence across modalities and avoids artefacts arising from imputed or probabilistic pairing.

Quality control of multiome data was performed using a combination of RNA, ATAC, and hashing-derived metrics to ensure retention of high-quality cells. First, sample demultiplexing was carried out using a stringent 99% positive quantile threshold on ADT signals, enabling confident assignment of cells to their corresponding hashed samples. RNA-based filtering then excluded cells with fewer than 200 total UMI counts (nCount_RNA), fewer than 200 detected genes (nFeature_RNA), or with mitochondrial transcript proportions exceeding 30%, thereby

removing low-complexity or damaged transcriptomes. ATAC-derived quality metrics were applied in parallel, retaining only nuclei with more than 5,000 total ATAC fragments (nATAC) and at least 1,000 fragments mapping to called peaks, ensuring sufficient chromatin accessibility signal for downstream peak calling and integration. Cells were kept only if they passed all RNA, ATAC, and hashing thresholds, generating a high-confidence multiome dataset suitable for integrative analysis.

2.2.9.6 Processing Chromatin Accessibility (scMultiome-seq)

Chromatin fragments were processed using Signac's *TSSEnrichment()*, *NucleosomeSignal()*, and fragment histogram visualisations. Only high-quality cells were retained (TSS [Transcription Starting Site] enrichment > 2, nucleosome signal < 4, > 300 fragments in peaks). Peaks were called with MACS2 (v2.2.7.1) in pseudobulk mode using pooled fragments from each condition. These consensus peaks were used to construct a peak × cell matrix via *FeatureMatrix()*. Accessibility data were TF-IDF normalised, and LSI dimensionality reduction (*RunSVD()*) was applied. Clustering was performed using the Louvain algorithm on LSI components and visualised with UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) (*RunUMAP()*).

2.2.9.7 Gene activity and multimodal Integration (scMultiome-seq)

Gene activity scores were computed from chromatin data with *GeneActivity()* and log-normalised. Multi-modal integration of RNA + ATAC was carried out using *FindMultiModalNeighbors()* and *RunUMAP()* on the joint graph. Where needed, Harmony (*RunHarmony()*) was used to correct for pool batch effects across modalities, preserving shared biological variance while removing technical artefacts. The final integrated object contained paired RNA, ATAC, and sample-of-origin metadata, enabling downstream analyses

such as gene regulation inference, transcription factor (TF) motif accessibility with chromVAR (*RunChromVAR()*), and lineage tracing overlays.

2.2.9.8 Clustering and cell type annotation

Cell clustering and annotation were performed using an iterative, multi-step strategy designed to robustly identify epithelial subtypes and lineage relationships, while accounting for batch complexity and multimodal signal integration. Identification of cell types followed a combination of known marker genes and validation against several location, stage, or disease specific reference datasets (see Supplementary Methods in the Appendix, Section 1.3). After dimensionality reduction using PCA, a shared nearest neighbour (SNN) graph was constructed, followed by community detection using the Louvain algorithm implemented in Seurat's *FindClusters()* function. UMAP was used to visualise clustering outputs in two dimensions, using the top 30 principal components (or Harmony embeddings, when batch correction was applied). Cell type annotation was achieved through a combination of approaches. First, canonical epithelial lineage markers were overlaid onto clustering solutions to guide initial annotations. For example, *LGR5* and *OLFM4* marked stem/progenitor populations; *MUC2* and *TFF3* indicated goblet cells; *CHGA* and *NEUROD1* marked enteroendocrine cells; *BEST4*, *SPIB*, and *CA7* denoted Best4 Cells; and *SPIB* and *SOX8* identified M-cell–like populations. Second, annotations were refined through automated label transfer using several reference atlases of human intestinal tissue. These included datasets spanning multiple anatomical locations, developmental stages, and pathological contexts[22,38,41,42,47]. Annotation consistency was verified across references, and label uncertainty was manually resolved by examining marker gene expression, expression gradients, and cluster co-localisation with known lineage markers in UMAP space. In multi-modal analyses (e.g., DOGMA-seq), congruent cell-type assignments across RNA and ATAC modalities were further confirmed using multi-modal weighted nearest

neighbours (WNN) analysis (*FindMultiModalNeighbors()*) and joint UMAPs. Gene activity scores derived from chromatin accessibility data (via *GeneActivity()*) were used to validate transcriptional cluster identities at the epigenomic level. This integrated strategy enabled high-resolution annotation of intestinal epithelial cell types, including rare or previously under-characterised populations. Final annotations were stored as Seurat object metadata and used in all downstream comparisons.

2.2.9.9 Pseudotime and Lineage Trajectory Analyses

Cell-differentiation trajectories were reconstructed using the R package *monocle3* (version 1.4.26) on a harmonised multiomic dataset. This approach was applied to reconstruct continuous differentiation gradients across absorptive and secretory lineages in hIEO cultures. Chromatin accessibility data (MACS2-derived peaks) were used as the primary assay to define the joint embedding. Cell identities were assigned based on known transcriptomic signatures from RNA+ATAC integration.

The *cell_data_set* (CDS) object was constructed using *as.cell_data_set()* and clustered in UMAP space. Trajectory graphs were learned using *learn_graph()*, with or without partitioning depending on lineage structure. The root cell for pseudotime ordering was defined as the cell closest to the centroid of the Stem Cell cluster in UMAP space. Pseudotime was inferred using *order_cells()*, and visualisation was performed with *plot_cells()* coloured by pseudotime, cluster identity, and lineage label. To validate the results obtained from Monocle3 and assess trajectory robustness, Slingshot was applied using the *slingshot* R package. The Seurat object was converted to a *SingleCellExperiment* object, and UMAP coordinates were assigned to the *reducedDims* slot. Slingshot was run using cluster identities as input labels, with "Stem Cells"

set as the root cluster. Pseudotime values were extracted and used for independent visualisation and gene expression trend analysis across lineages.

2.2.9.10 Overview of CellTagging Analyses in scRNA-seq

Prior to the multiome CellTag experiment, the diversity and quality of the lentiviral CellTag library were assessed by transducing HEK293T cells with a low multiplicity of infection and sequencing the resulting CellTag amplicons on a MiSeq platform as per the original CellTag protocol. MiSeq reads were processed to identify the full repertoire of CellTags present in the library and to estimate their relative abundances. Only tags supported by ≥10 reads were considered bona fide and retained in a high confidence "whitelist." This MiSeq-derived whitelist was subsequently supplied to the demultiplexing pipeline for the multiome datasets to filter the raw UMI matrix. By restricting analysis to this vetted list of tags, we reduced noise introduced by sequencing errors and ensured that the detected tags originated from the validated HEK293T test library, thereby improving the accuracy of clonal tracking across all multiplexed pools. CellTag lineage tracing analysis was performed using a modified version of the original *CellTagR* pipeline in R, with adaptations for compatibility with multiplexed single-cell multiome data. In contrast to traditional workflows, where CellTag libraries are generated as part of the gene expression library, here the CellTag-specific amplicons were also amplified directly from final gene expression libraries prior to sequencing. This approach was designed to enhance recovery of low-abundance CellTag transcripts while retaining full transcriptomic and chromatin accessibility profiles from the same cells. Raw sequencing reads corresponding to the CellTag-enriched libraries were processed using a custom Cell Ranger run with a feature barcoding reference containing known CellTag sequences, appended as synthetic genes to the gene expression reference. The resulting UMI count matrix was then imported into R and

processed with the *CellTagR* framework, using *Seurat* (v5.1.0), *DropletUtils* (v1.26.0), and *dplyr* for cell filtering and matrix manipulation.

2.2.9.11 Barcode Filtering and Matrix Preparation (scRNA-seq+CellTagging and scMultiome-seq+CellTagging)

The CellTag UMI matrix was filtered to retain only barcodes that passed RNA-based QC (as described in Section *2.2.9.10*). Hashtag-negative cells or barcodes below RNA quality thresholds were removed. The filtered UMI matrix was then binarised using *binarizeMatrix()* to indicate presence or absence of individual CellTags in each cell.

2.2.9.12 Tag Filtering and Abundance Thresholding (scRNA-seq+CellTagging and scMultiome-seq+CellTagging)

CellTags detected in fewer than a minimum number of cells (<10) were removed to eliminate sequencing noise and spurious tags. Tags with excessive frequency were also excluded to prevent inflation from common ambient tags or artefacts.

2.2.9.13 Cell Lineage Group Assignment (scRNA-seq+CellTagging and scMultiome-seq+CellTagging)

A Jaccard similarity matrix was computed based on the overlap of CellTag presence between cells. This matrix was then used to perform hierarchical clustering (hclust) and define clonal groups using a dynamic cut height selected empirically to balance clone resolution and stringency.

2.2.9.14 Clone Calling and Visualisation (scRNA-seq+CellTagging and scMultiome-seq+CellTagging)

Clonal groups (cells sharing similar CellTag combinations) were assigned unique clone IDs and stored as metadata. Clonal structure was overlaid onto RNA UMAPs, and clone sizes and

frequencies were summarised using *ggplot2* and *Seurat* plotting functions. Clones were coloured by cell type or pool origin to investigate lineage restriction or expansion.

2.2.9.15 Clone Composition and Heterogeneity Assessment (scRNA-seq+CellTagging and scMultiome-seq+CellTagging)

For each clone, the proportion of constituent cell types was computed to evaluate lineage bias. Briefly, clonal reconstruction from CellTag data was performed using the CellTagR pipeline with default parameters, following the criteria established in the original methodology. Briefly, CellTag UMI counts were first collapsed and error-corrected, after which tags were binarised such that a CellTag was considered present in a cell only if supported by ≥ 2 UMIs. A percentile-based whitelist was then generated using the default abundance threshold to retain high-confidence CellTags. Cells carrying fewer than two or more than twenty distinct CellTags were excluded to remove, respectively, poorly labelled cells and putative multiplets or hyperinfected artefacts. Clonal groups were subsequently defined using the default Jaccard similarity cutoff of 0.7, which identifies sets of cells sharing sufficiently overlapping CellTag signatures to represent inheritance from a common progenitor. As per the original implementation, only groups composed of three or more related cells were retained as bona fide clones. These default criteria ensured robust and reproducible identification of clonal structures across the dataset, consistent with the validated parameters of the CellTagR framework. A Shannon entropy score was also calculated for each clone to quantify transcriptional diversity and plasticity. All scripts were implemented in R (version 4.4.1) using custom functions and *CellTagR* utilities. Figures and statistics presented in Chapters 3 and 4 were generated using this reproducible pipeline. All clonal assignments, clone UMAPs, and composition metrics were validated against randomised background controls to ensure specificity of tagging and avoid over-clustering.

1

3 Standardising Organoid Cultures to Model the Human Intestinal Epithelium

3.1 Introduction

The human intestinal epithelium comprises a diverse array of functionally specialised cell types, arranged along the crypt–villus axis in the small intestine and the crypt–surface axis in the colon[44,51]. *In vitro* recapitulation of this complexity has been enabled by the development of hIEOs derived from adult or fetal intestinal stem cells. Cultured within 3D extracellular matrix environments, hIEOs self-renew and self-organise, offering powerful platforms to investigate epithelial development, differentiation, and disease mechanisms[39,323,324]. Nonetheless, reproducing the full spectrum of *in vivo* epithelial phenotypes *in vitro* remains a significant challenge[325].

Standard culture protocols tend to favour the expansion of enriched, well-characterised populations, such as intestinal stem cells and transit-amplifying progenitors (“Stem Cells” and “TAs”, respectively) and lineage-committed progenitors (“Absorptive Progenitors” and “Secretory Progenitors”). By contrast, rare or terminally differentiated lineages — including enteroendocrine cells (“EECs”), microfold cells (“M-Cells”), Best4 cells (“Best4 Cells” or “Best4”), Paneth cells (“Paneth Cells”), and tuft cells (“Tuft Cells”) — are frequently underrepresented or entirely absent. This bias likely stems from the proliferative advantage of intermediary populations under conventional *in vitro* conditions, limiting the recovery and analysis of rarer or more specialised cell types.

Emerging evidence suggests that epithelial diversity is highly sensitive to even minor changes in culture protocols[208,227,325]. Subtle modifications can lead to marked shifts in lineage representation, complicating the selection of conditions that most faithfully recapitulate native epithelial composition. This variability raises important concerns regarding the reproducibility, comparability, and physiological relevance of hIEO models, particularly when applied to disease modelling or therapeutic evaluation.

To address these limitations, this work presents a systematic framework for optimising and benchmarking hIEO culture protocols using rare epithelial lineages — such as Best4 Cells — as a readout of epithelial diversity and differentiation fidelity. Although not the central focus of this study, Best4 Cells provide an ideal case study due to their well-defined rarity, transcriptomic specificity, and absence in most standard culture settings. The focus on this cell type here is therefore mainly methodological, enabling evaluation of the extent to which refined protocols support the emergence of lineages typically lost under standard conditions.

Best4 Cells, recently characterised as a distinct epithelial subtype implicated in luminal pH regulation, are poorly represented *in vitro* and are absent from most conventional animal models. Despite their classification as terminally differentiated within the adult intestinal epithelium, Best4 are enriched in fetal tissues from as early as six post-conceptional weeks (pcw), suggesting a role in intestinal morphogenesis[47]. They are most abundant in the colonic surface epithelium, though lower-level representation is also observed throughout the small intestine[41,47,60]. The heterogeneity of their transcriptional signature across regions and developmental stages remains unresolved, likely reflecting a combination of biological and technical variability. Furthermore, while consistently present in the healthy intestinal landscape, Best4 Cells are frequently diminished or absent in the inflamed mucosa of individuals with UC. The mechanistic basis of this depletion remains unclear and may involve

environmental disruption post-maturation — such as barrier erosion and microenvironmental hostility — or earlier molecular perturbations, including epigenetic dysregulation and impaired lineage specification[38].

Notably, while BEST4-expressing (*BEST4* +) cells may sporadically appear in hIEOs, expression of *BEST4* alone does not reliably indicate full lineage identity. High-resolution single-cell datasets have shown that expression of putative markers such as *OTOP2*, *HES4*, and *CA4* is frequently low or non-specific in hIEO-derived Best4 Cells, limiting their utility for robust classification. By contrast, *BEST4* and *SPIB* emerge as the most consistently co-expressed and lineage-defining markers across diverse experimental conditions, with *CA7* and *MEIS1* serving as secondary markers in larger datasets where Best4 representation is enriched. Accordingly, this study employed a stringent definition of Best4 Cells, restricting classification to cells co-expressing quantifiable levels of both *BEST4* and *SPIB*, with supplementary *CA7* and *MEIS1* expression used to support ambiguous cases. This approach prioritised specificity and enabled rigorous benchmarking of protocol performance.

To accommodate the spatial and temporal dynamics of enterocyte maturation along the crypt–villus and crypt–surface axes, the enterocyte lineage was stratified along a transcriptional continuum. Given the comparatively increased heterogeneity of *in vitro* hIEO-derived enterocytes relative to *in vivo* primary tissue[326], a conservative classification framework was adopted: (1) “Early Enterocytes”, representing transitional intermediates emerging from Absorptive Progenitors with retained plasticity; (2) “Absorptive Enterocytes”, reflecting progressive lineage commitment and increased functional maturation; and (3) “Crypt-Top Enterocytes”, corresponding to the most transcriptionally mature absorptive cells, aligned with the luminal surface compartment *in vivo*. This classification was guided by a curated gene signature list, provided in the Appendix (Supplementary Methods, Section 1.3).

The experimental framework outlined in this chapter focuses on the development of scalable and cost-effective methods for high-throughput screening of hIEO cultures. It supports the overarching aim of systematically comparing candidate protocols based on their ability to support epithelial lineage diversification, using rare cell types — such as Best4 cells — as sensitive indicators of differentiation potential. A key goal is to identify commonalities across protocols that could form the basis of a canonical framework, capable of producing consistent and representative outcomes — including the full spectrum of epithelial lineages — across laboratories and experimental contexts. Beginning with a slightly modified version of a gold-standard organoid culture protocol, this study addresses the reproducibility and maturation dynamics of hIEO cultures across multiple donor-derived lines.

This chapter is organised into three sections:

3.2 Development of a scalable, multiplexed scRNA-seq platform for benchmarking epithelial composition across standardised hIEO cultures.

3.3 Meta-analysis of publicly available single cell hIEO datasets to contextualise findings and prioritise candidate protocols.

3.4 Application of Smart-seq2 to interrogate limitations of the canonical protocol, followed by systematic benchmarking of alternative conditions using the multiplexed platform.

Together, these efforts aim to expand the methodological toolkit for interrogating rare intestinal epithelial populations *in vitro*, and to support the development of refined protocols that more faithfully reflect human epithelial diversity. Although experimental settings included a variety of health statuses, anatomical regions, and organoid stimulation conditions, the analyses presented throughout this and the following chapters are restricted to technical performance across metrics such as cell-type recovery, differentiation potential, and

transcriptomic resolution. Biological comparisons, such as disease-specific variation or regional identity, are deferred to Chapter 6.

3.2 Development of a Multiplexed Single-Cell RNA-seq Assay for High-Throughput Profiling of hIEO Cultures

3.2.1 Mapping Epithelial Heterogeneity and Rare Lineages in hIEOs at Single-Cell Level

The multiplexed scRNA-seq framework presented in this section was developed to evaluate the differentiation potential of a modified version of the Sato *et al.* (2011) protocol (hereafter referred to as the "House Protocol"). While quantitative PCR (qPCR) remains a standard tool for rapid transcript-level validation, its bulk nature precludes single-cell resolution and limits the detection of transcriptionally heterogeneous or low-abundance populations, such as Best4 Cells. High inter-replicate variability and inconsistent expression estimates further underscored its inadequacy for comprehensive epithelial profiling (Supplementary Figure S1, included in the Appendix to this thesis).

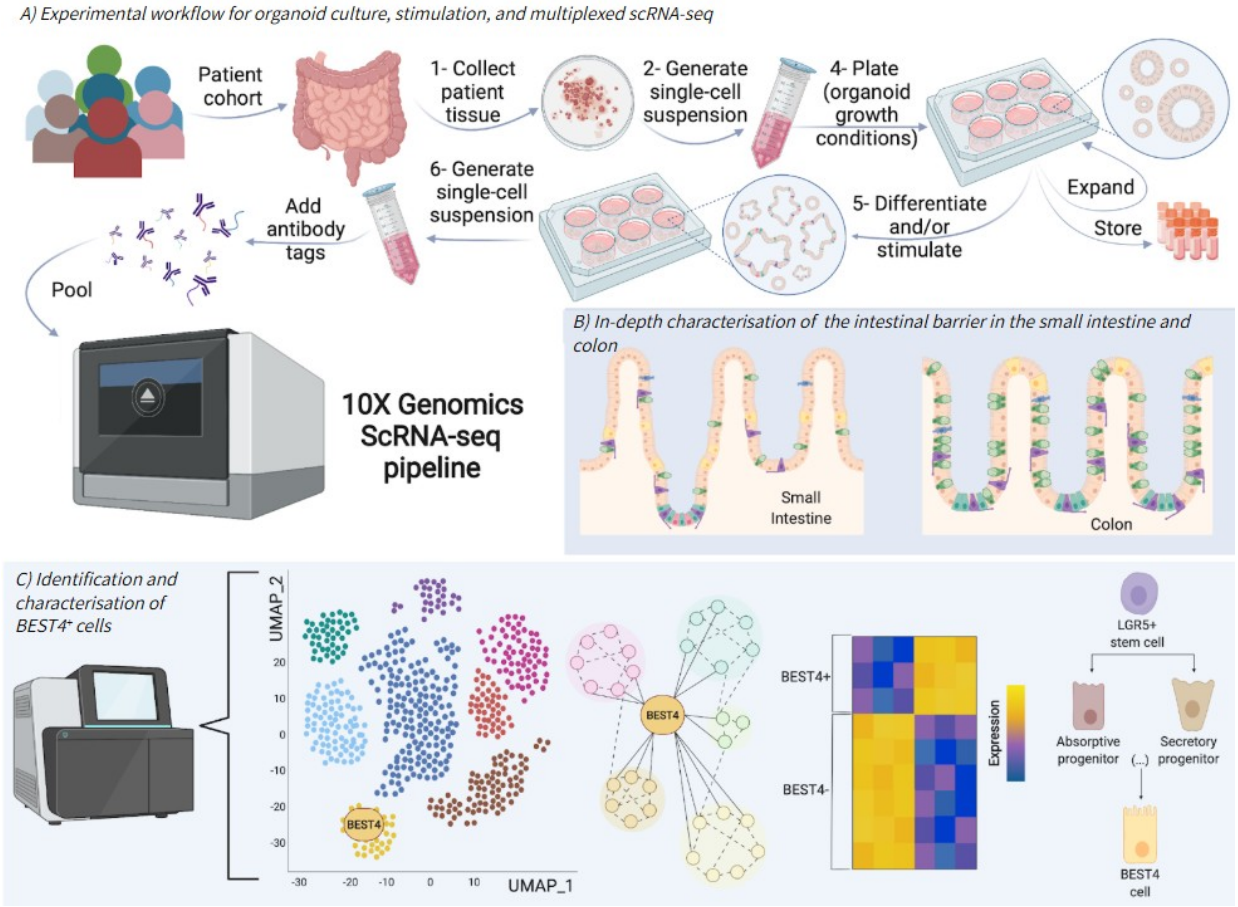
The House Protocol involves two sequential culture stages defined by distinct niche factor compositions (see Methods, Section 2.2.2.1)[126]. In the initial phase, cells are expanded in Wnt-enriched media - Conditioning Media (CM) - which maintains a proliferative, stem-like state reminiscent of the *in vivo* crypt-base niche inhabited by *LGR5+* intestinal stem cells. Upon reaching optimal organoid density, CM is replaced by Wnt-depleted Differentiation Media (DM), which promotes epithelial maturation by mimicking the attenuated Wnt signalling experienced by cells migrating along the crypt–villus or crypt–surface axis.

This platform was designed to systematically benchmark the emergence of both canonical and rare epithelial lineages *in vitro* and across multiple experimental conditions, using the reproducible detection of Best4 Cells as a proxy for differentiation fidelity.

3.2.2 Experimental Design: High-Throughput Evaluation of Differentiation Efficiency in hIEOs Across Multiple Conditions Using Multiplexed scRNA-seq

A multiplexed scRNA-seq workflow, previously optimised for native human intestinal tissue, was adapted for high-throughput profiling of hIEO cultures[41]. By incorporating cell hashing through oligo-conjugated antibodies - hashtag nucleotides ("HTOs") and, where applicable, antibody-derived tags ("ADTs") - this method enabled simultaneous profiling of multiple samples while reducing batch effects and enhancing analytical throughput (Figure 1). The assay was evaluated for its capacity to resolve epithelial heterogeneity and to detect rare lineages, including Best4 Cells, under diverse experimental conditions.

Figure 1 - Multiplexed scRNA-seq workflow for high-throughput evaluation of differentiation efficiency and system perturbation effects in hIEOs



Legend: Experimental design overview. (A) Patient-derived intestinal tissue is dissociated into single-cell suspensions, from which epithelial fractions are isolated. These are then plated across multiple experimental conditions, such as different organoid culture protocols and cytokine stimulations. Samples are labelled individually with oligo-conjugated antibody tags and pooled into a single library prior to scRNA-seq (10X Genomics) processing. (B) High-resolution characterisation of epithelial heterogeneity across small intestine and colon is enabled through this workflow, allowing detailed regional comparisons of lineage composition. (C) By screening multiple conditions, the goal is to achieve an accurate representation of the full epithelial landscape - including rare Best4 Cells - with trajectory inference linking Best4 cells to secretory progenitors. Then, expression heatmaps can highlight lineage-specific gene signatures distinguishing Best4 populations. Together, this multiplexing strategy provides a robust, high-throughput platform to profile epithelial differentiation efficiencies across multiple hIEO experimental conditions.

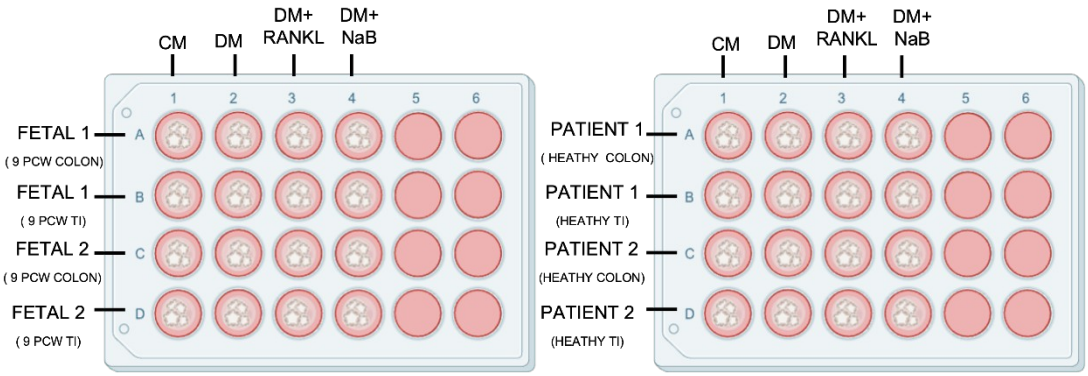
For this experiment (EXP01), hIEOs were established from both adult and fetal tissue samples, derived from two anatomically distinct sites: the terminal ileum (TI) and the colon. To assess potential pathological modulation, the adult cohort included organoids from healthy donors and patients diagnosed with UC - a condition previously associated with diminished Best4 abundance. To further interrogate the influence of microenvironmental cues on Best4 Cell

emergence, external stimuli were incorporated based on prior evidence implicating them in modulating *SPIB* and *BEST4* expression: sodium butyrate (NaB), a microbial metabolite known to promote epithelial maturation, and RANKL, a cytokine associated with *SPIB*-dependent lineage specification[37,58,327–329].

Each organoid line was exposed to four experimental conditions: CM only (control); CM+DM ("DM"), and CM+DM supplemented with either RANKL or NaB ("DM+RANKL" and "DM+NaB", respectively). Following initial expansion in CM, test samples (DM, DM+RANKL and DM+NaB) were transitioned to DM-based conditions to initiate differentiation. For the DM+RANKL and DM+NaB groups, stimulations were applied 48 hours prior to cell harvesting (Figure 2; see Methods, Section 2.2.2.4). Control samples remained in CM throughout.

For the scope of this section, comparative analyses are restricted to the CM and DM conditions to establish a baseline differentiation benchmark. The effects of RANKL and NaB treatments are examined separately in Chapter 5.

Figure 2 - Experimental design for Experiment 1: High-throughput screening of organoid differentiation and response to RANKL and Sodium Butyrate stimulation across distinct developmental stages and patient-derived samples



Legend: Left: Fetal-derived intestinal organoids from 9 and 20 post-conception weeks (pcw), obtained from both terminal ileum (TI) and colon, cultured were under four different conditions: CM (conditioning medium), DM (differentiation medium), RANKL supplementation, and NaB supplementation (both

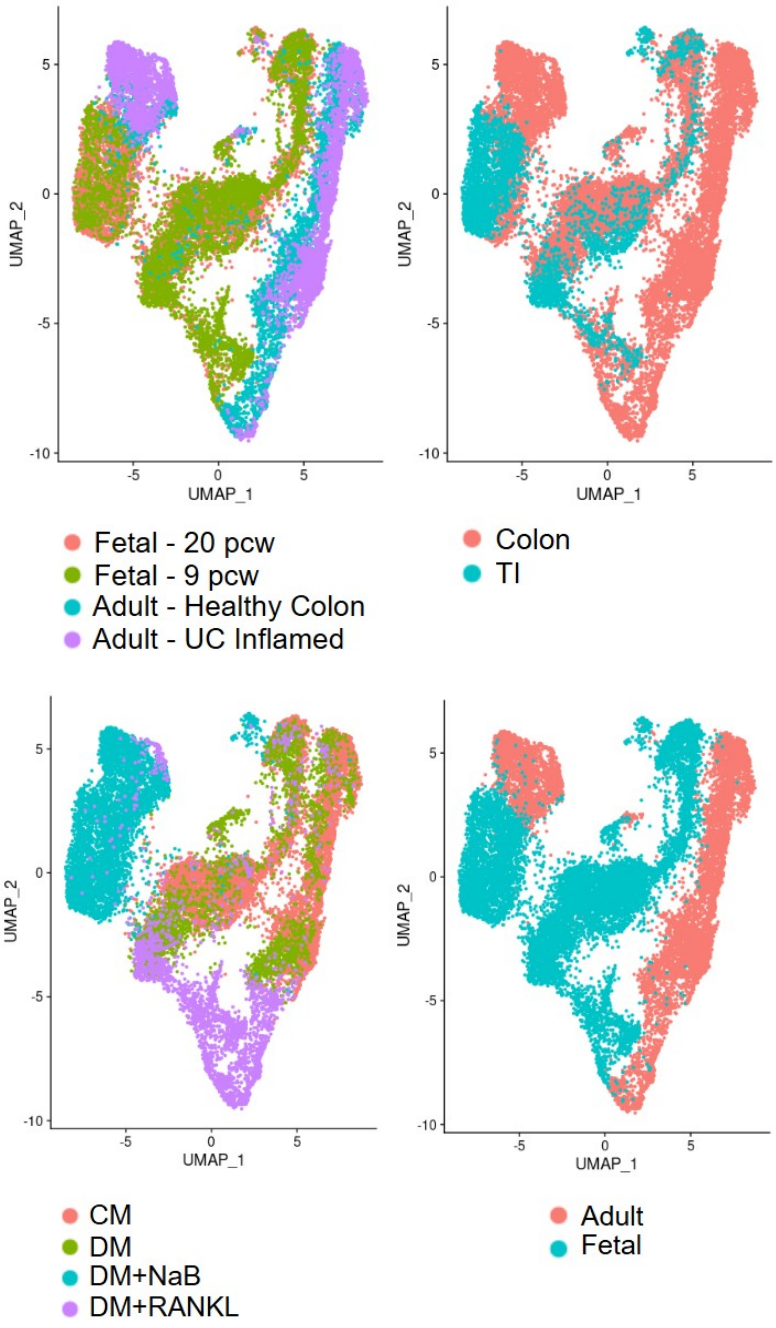
added to DM for 48h before the end of the experiment). Right: Adult patient-derived organoids from healthy and IBD samples exposed to the same conditions to assess differential responses.

3.2.3 Transcriptomic Profiling and Quality Control

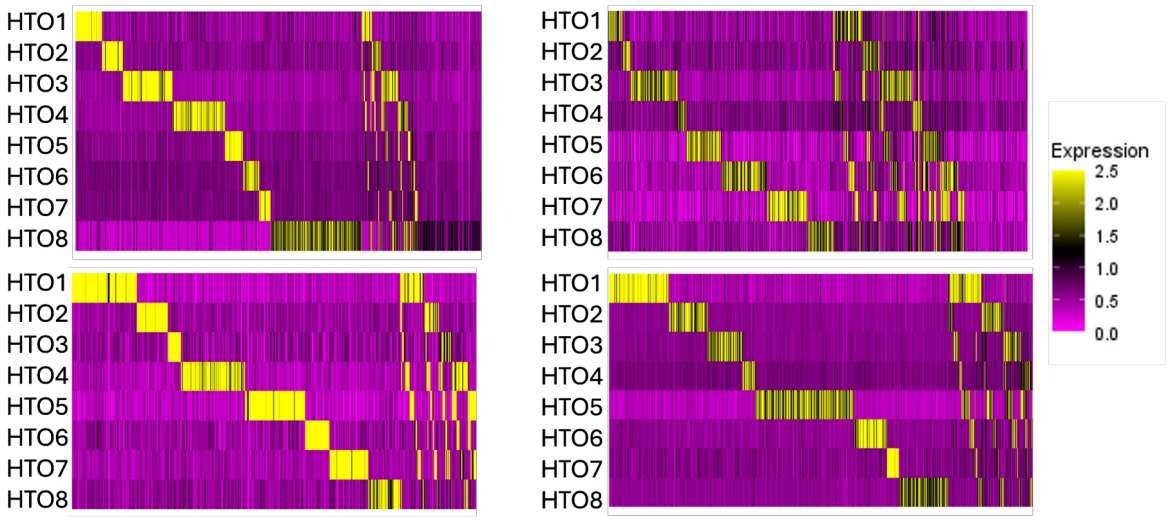
Single-cell transcriptomic profiling was performed using the 10x Genomics 3´ Gene Expression platform in combination with Illumina's NextSeq platform — a rapid and cost-effective sequencing option compared to deeper platforms such as Illumina's NovaSeq platform[330]. A total of 29,013 cells were profiled across four pooled reactions, each containing eight uniquely antibody-labelled samples. Sample multiplexing was achieved through oligo-conjugated antibody cell hashing, enabling accurate assignment of cells to their original experimental condition while minimising batch effects (Figure 3A).

From the initial dataset, 21,481 cells were confidently demultiplexed and attributed to specific conditions. The remaining 7,532 cells — identified as either doublets or hashing-negative — were associated with poor RNA quality metrics, likely corresponding to dead or dying cells. These observations underscore the effectiveness of cell hashing in filtering suboptimal transcriptomes from downstream analysis (Figure 3B; Supplementary Figure S2A). Following stringent quality control, 17,418 high-confidence epithelial transcriptomes were retained for downstream analyses (Figure 3C; Supplementary Figure S2B), providing a robust dataset for evaluating differentiation dynamics and rare lineage emergence across conditions.

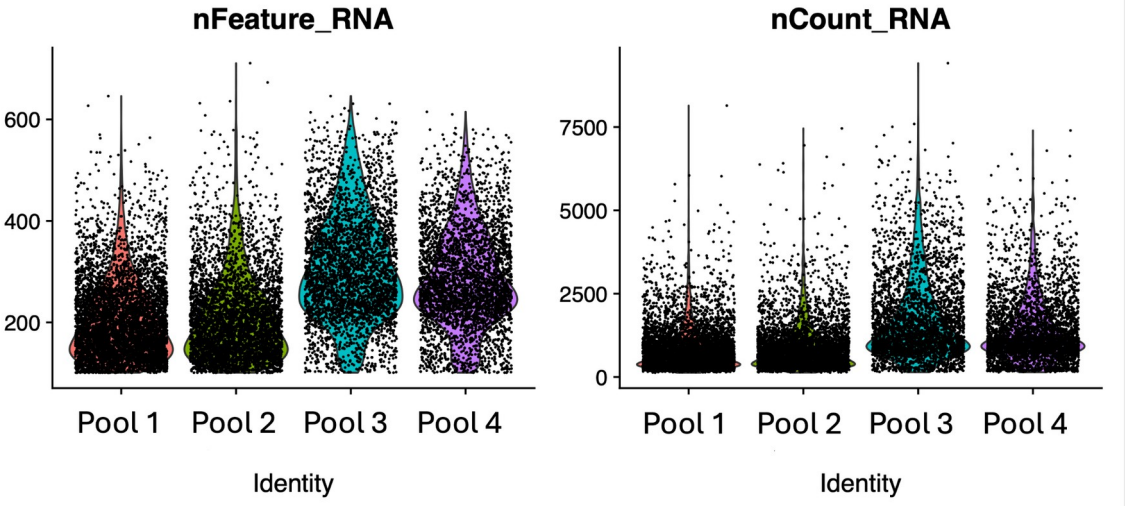
Figure 3 – Overview of relevant experimental metrics: Uniform Manifold Approximation and Projection (UMAP) as a function of distinct experimental variables, multiplexing efficiency and other key quality control metrics from the pilot single-cell multiplexed sequencing run of hIEOs
A) UMAP projections of integrated hIEO scRNA-seq data



B) Heatmap of HTO expression across individual cells according to their assigned tag



C) RNA quality control metrics



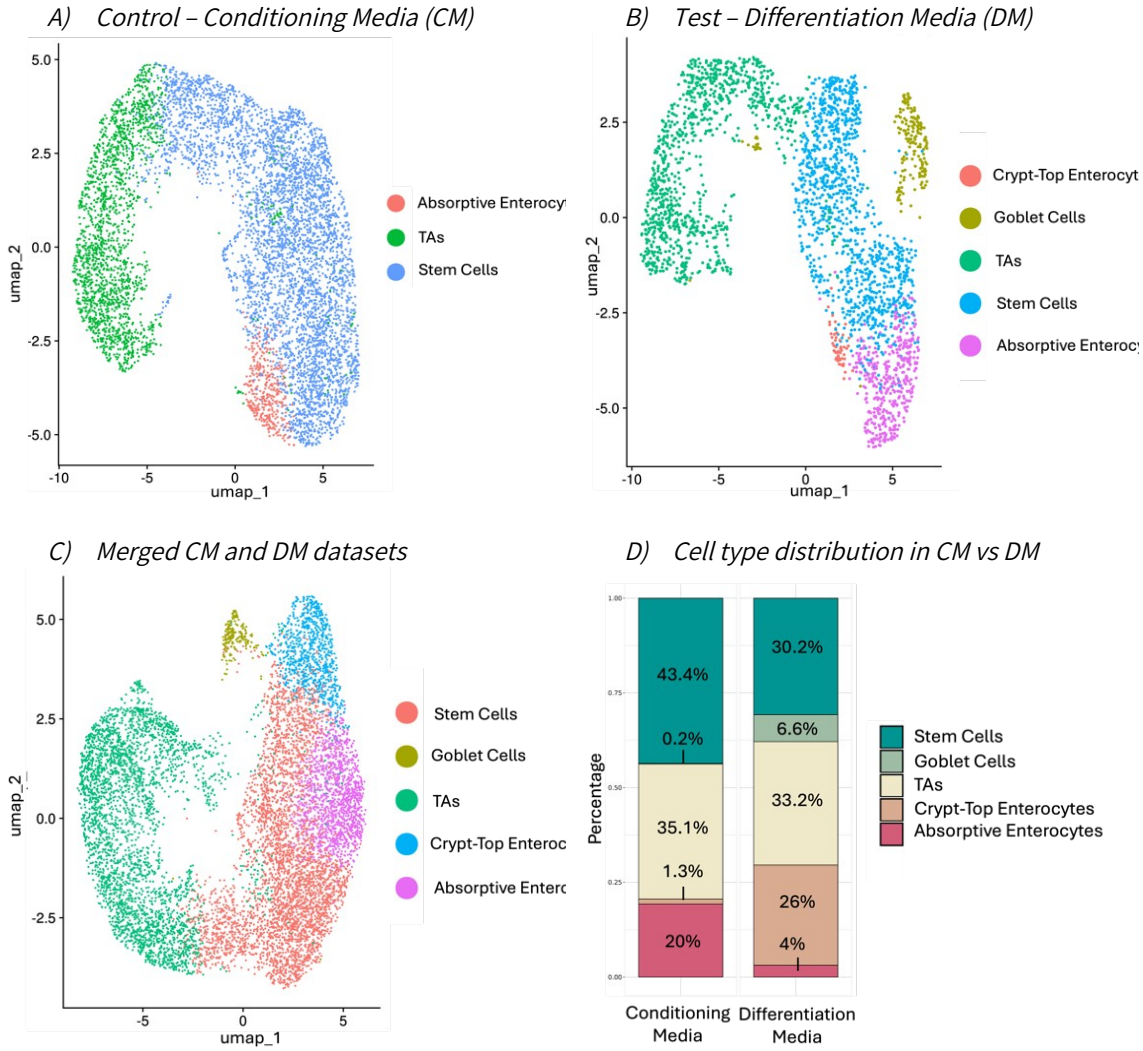
Legend: A - Final UMAP of integrated hIEO scRNA-seq data, coloured by sample source (top left), tissue region (top right), in vitro culture condition (bottom left), and developmental stage (bottom right). These results highlight the influence of both developmental origin and experimental context on epithelial transcriptional states. B - Heatmaps showing expression of HTOs across individual cells according to their assigned tag. Clear diagonal enrichment patterns confirm successful and specific sample labelling, with minimal multiplet rates or untagged cell fractions. C - Distribution of RNA quality control metrics across the four multiplexed pools, where nFeature_RNA indicates the number of unique genes detected per cell, and nCount_RNA reflects the total number of RNA molecules (i.e. UMI counts) detected per cell. Consistent distributions across pools suggest comparable library quality and cell recovery across multiplexed conditions.

3.2.4 CM vs DM: Media-Specific Effects on Epithelial Composition

To investigate the impact of niche modulation on epithelial differentiation, a subset of 9,692 cells cultured under either CM or DM conditions (“CM+DM House Differentiation subset”, Figure

4A-C) was selected for detailed analysis. Dimensionality reduction and unsupervised clustering using Uniform Manifold Approximation and Projection (UMAP) revealed distinct transcriptional landscapes between the two conditions (Figure 4A-C). Cells cultured in CM exhibited strong enrichment for intestinal stem cell signatures *(LGR5, ASCL2, SMOC2)* and highly proliferative TAs *(MKI67, TOP2A, UBE2C)*, indicative of an undifferentiated state. By contrast, DM-treated cultures showed increased transcriptional complexity, characterised by the emergence of Goblet Cells *(MUC2, SPDEF, WFDC2)* and Absorptive Enterocytes *(FABP1, SLC15A1, ALPI)*, suggesting partial progression along differentiation trajectories. However, despite these gains in lineage diversity, no transcriptionally discrete clusters corresponding to rare cell types - such as EECs, Best4, Tuft Cells, Paneth Cells, or M-Cells - were identified under either condition (Figure 4B–D). These findings highlight the limited capacity of this baseline differentiation protocol to support full epithelial lineage specification *in vitro*, particularly for low-abundance subtypes.

Figure 4 – Overview of the transcriptional landscape of differentiated versus undifferentiated hIEOs cultured in House medium



Legend: (A)-(C) - UMAP plots showing unsupervised clustering of scRNA-seq data from organoids cultured in Conditioning Media (CM) (A), Differentiation Media (DM) (B), and their joint representation within the overall (CM+DM) differentiation subset (C). Cells are annotated by major epithelial cell types based on their transcriptomic profiles. (D) Bar plot showing the proportional distribution of each epithelial cell type across CM and DM conditions, highlighting differences in lineage output between conditions.

3.2.5 Identification of Partial Best4 Cell States

Unsupervised clustering did not yield a transcriptionally distinct cluster corresponding to *BEST4+* cells. However, targeted analysis of cells expressing *BEST4* identified 369 cells with medium to high transcript levels (Figure 5A-B). While low-level background expression can result from technical noise — such as ambient RNA contamination or artefacts introduced during data processing — the number of *BEST4*-expressing cells observed here is substantially higher than the baseline typically seen for transcripts not genuinely expressed, suggesting that the detected signal reflects true biological expression rather than artefactual background. Among these, only 19 cells co-expressed *SPIB*, and just 3 also expressed *CA7* - a transcriptional profile suggestive of partial or transitional identity, rather than full Best4 lineage commitment (Supplementary Figure S3). These sparse *BEST4+* cells exhibited low-to-moderate expression levels of a combined Best4 Cell score and remained transcriptionally dispersed within the broader absorptive compartment (Figure 5C), lacking the coherence required for independent cluster formation.

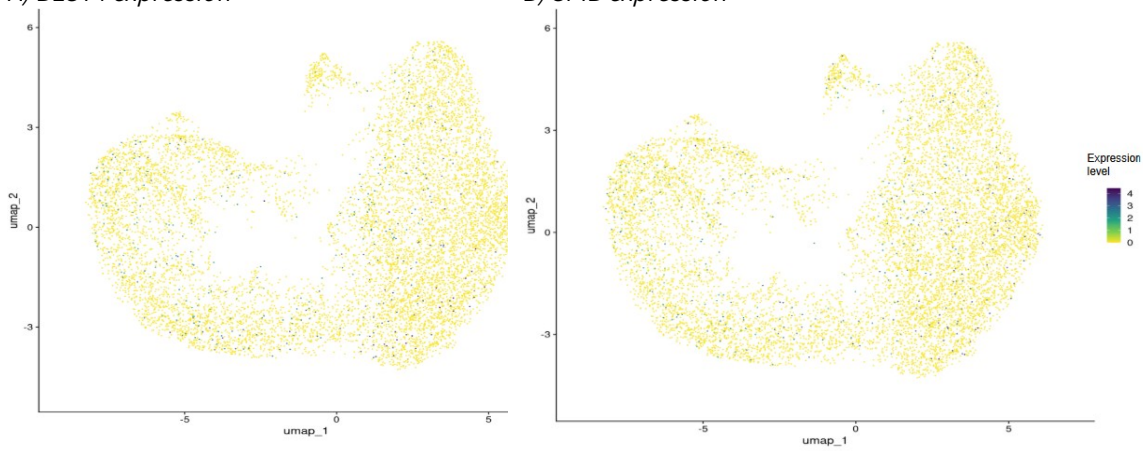
This sparse distribution and lack of spatial concentration or coherent clustering likely reflects early or incomplete differentiation trajectories. This is consistent with cells representing either undifferentiated or partially specified states that retain low-level expression of multiple lineage-associated markers, or absorptive intermediates undergoing transcriptional priming toward the Best4 fate. The observed heterogeneity supports the presence of lineage-skewed progenitors, yet also highlights the absence of sufficient *in vitro* cues — such as niche-derived signals or prolonged differentiation windows — required for full maturation into a stable Best4 identity. These findings suggest that the protocols used here may lack key environmental or temporal niche signals necessary to drive terminal differentiation into fully committed Best4 Cells.

1

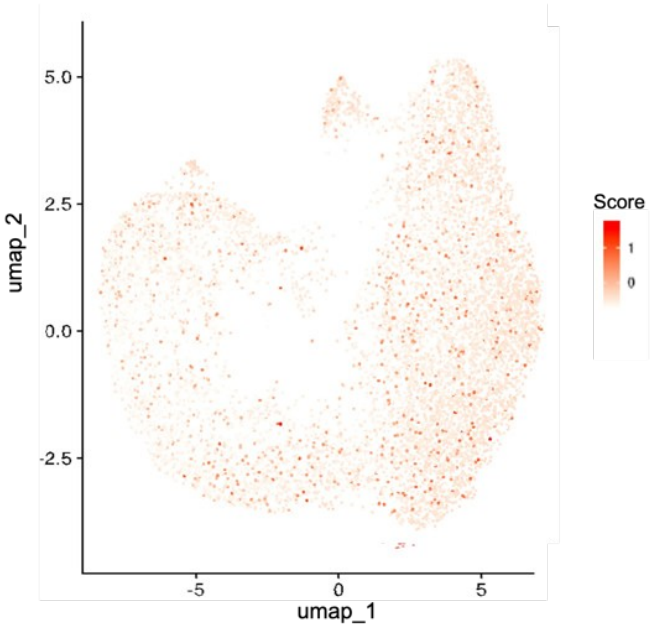
Figure 5 - Expression profile of BEST4, SPIB, and a combined Best4 Cell Score across the whole CM+DM dataset

A) BEST4 expression

B) SPIB expression



C) Best4 Cell Score



Legend: (A) UMAP projection showing the expression of BEST4, a canonical marker of Best4+ Cells. (B) Expression of SPIB, a transcription factor associated with Best4 Cells' fate. (C) Best4 Cell Score calculated based on the enrichment of a curated Best4+ gene signature consisting of BEST4, SPIB, CA7 and MEIS1, highlighting sparse and discrete cell populations across the UMAP embedding.

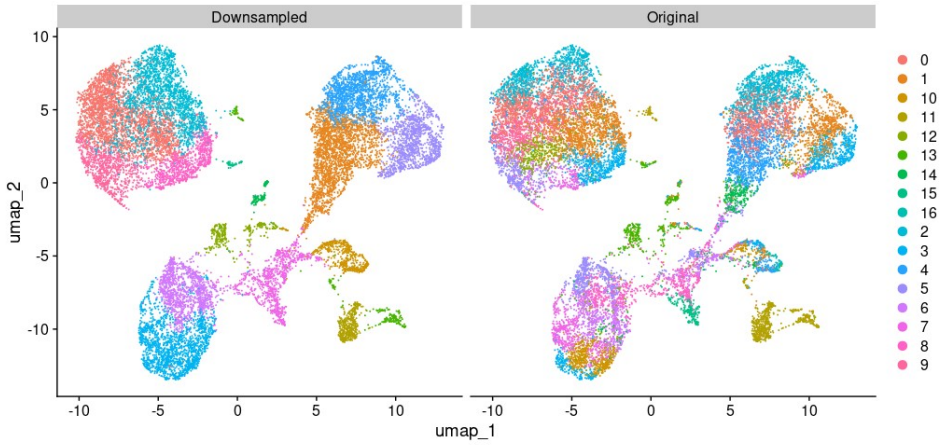
3.2.6 Methodological Benchmarking and Resolution Validation

To evaluate whether the lower per-condition cell yield — an inherent consequence of sample multiplexing — and the use of the Illumina NextSeq platform - as opposed to greater sequencing coverage approaches that improve detection sensitivity - compromised transcriptional resolution, a benchmarking simulation was conducted using a publicly available hIEO scRNA-seq dataset from Elmentaite *et al.* (2020). This dataset was selected for its high-resolution delineation of epithelial subtypes, reported presence of a Best4 cell population, and comprehensive transcriptomic coverage, as well as its quality and relevance and at the time of analysis.

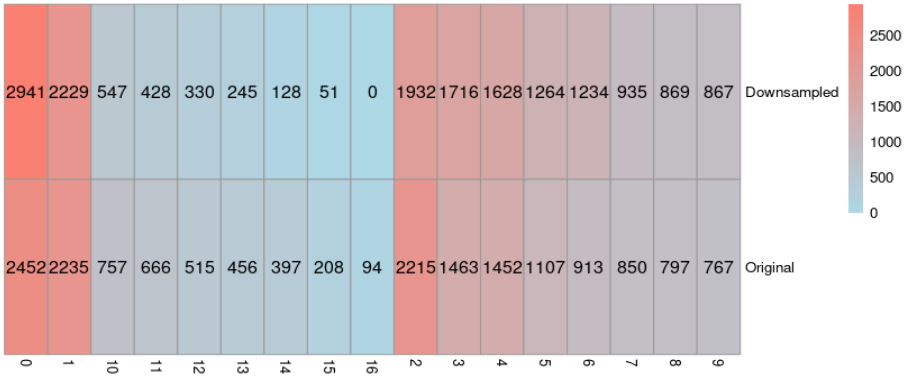
Downsampling was applied to approximate the sequencing depth observed in the present study, using Total UMI count as a proxy[331,332]. Clustering concordance was evaluated using the Adjusted Rand Index (ARI), which quantifies similarity between clustering solutions (with 1.0 indicating perfect agreement and 0.0 indicating random correspondence). The downsampled dataset yielded an ARI of 0.48, indicating moderate preservation of cluster structure. This discrepancy is illustrated in Figure 6A-B. However, independent annotation of both the original and downsampled datasets, followed by dataset integration, revealed that final cell-type-based cluster annotations remained highly similar across both scenarios (Figure 6B-C). These findings suggested that, while some subtle transcriptional attributes within cell types may be lost, the core features that define major cell identities are preserved at the current sequencing depth. These results validate the methodological robustness of the multiplexing strategy for assessing overall population dynamics across experimental conditions and confirm that biologically meaningful transcriptional heterogeneity can be reliably recovered, even under reduced sequencing depth and limited per-condition cell yield.

Figure 6 - Impact of dataset downsampling on cell clustering and annotation in an intestinal epithelial datasets (Elmentaite et al., 2020)

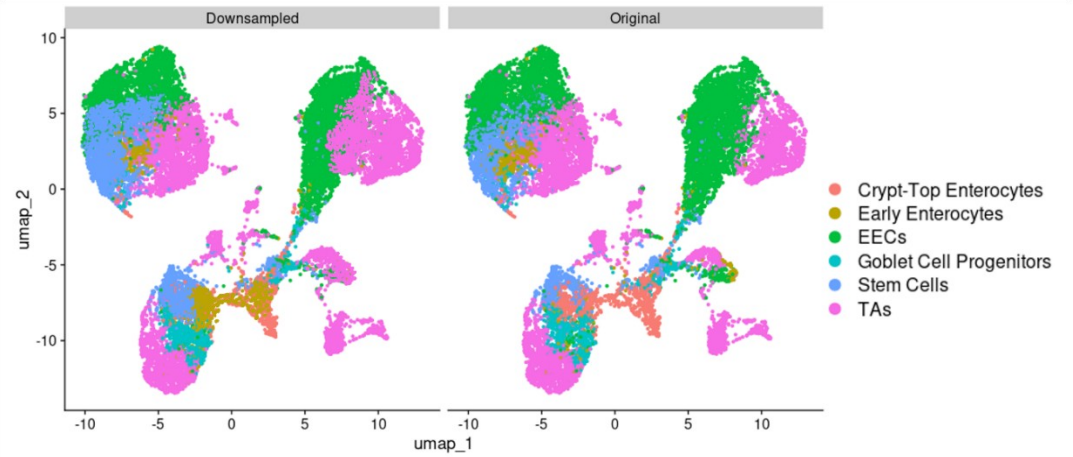
A) UMAP projections of unsupervised clustering results after downsampling (left) versus in the original dataset (right)



B) Cluster overlap before and after downsampling



C) Clustering overlap between the downsampled (left) and original datasets (right) after independent annotation and integration



Legend: (A) UMAP visualisation of Elmentaite et al., 2020 unsupervised clustering results before (right) and after (left) downsampling, coloured by original cluster identity (clusters 0–16). (B) Heatmap showing overlap between clusters in the original and downsampled datasets, with cell counts represented by colour intensity, indicating strong concordance in cluster identity across conditions. (C) UMAP showing clustering annotation results before and after downsampling. Key populations, including Stem Cells, Transit-Amplifying cells (TAs),

Goblet Cels, and Enterocytes, are retained after data reduction, supporting the stability of clustering and cell type annotation after downsampling and validating its use in subsequent comparisons.

3.2.7 Interpretation of Findings

The multiplexed scRNA-seq approach demonstrated strong technical performance, producing reproducible and cost-effective transcriptional profiles across 32 hIEO cultures. These findings underscore the potential utility of this assay in evaluating hIEOs as biomimetic models for the human intestinal epithelium across spatial, temporal, and pathological contexts. Moreover, by minimising batch effects and maximising throughput, this strategy enhances the interpretability, reproducibility, and analytical robustness of downstream single-cell analyses.

Cell-type compositions broadly recapitulated those observed in primary tissue reference datasets (as further elaborated in Chapter 5), affirming both the accuracy and biological relevance of the multiplexed dataset. However, under the differentiation conditions specified by the House Protocol, maturation was incomplete - stem-like and early progenitor states predominated, while differentiated secretory and rare epithelial lineages remained minimally represented. Although Enterocytes constitute the bulk of the *in vivo* intestinal epithelium, secretory and specialised subtypes — including Goblet Cells, EECs, Tuft Cells, Paneth Cells, M-Cells, and Best4 Cells — typically account for 10–20% of the tissue and were either absent or detected at very low frequencies *in vitro*.

Multiple factors may underlie these limitations, including inadequate differentiation duration, suboptimal tuning of niche signals (e.g., Notch, BMP, Wnt), absence of microbial, immune, or metabolic cues, and reduced statistical power to detect low-frequency populations. The sporadic co-expression of *BEST4*, *SPIB*, *CA7* and *MEIS1* — below the resolution threshold for discrete clustering — points to the possible presence of transitional or incompletely specified states. Similar patterns were noted for other rare epithelial lineages, such as M-Cells and EECs

(data not shown), reinforcing the need for refined protocols to support the emergence of the full spectrum of *in vivo* epithelial diversity. These partially committed profiles initiated under proliferative conditions could be lacking the requisite cues to complete maturation, ultimately leading to reversion to a less committed state or cell death. Altogether, these findings underscore the nuanced, non-linear nature of epithelial fate specification *in vitro*, highlighting that differentiation is not driven solely by Wnt withdrawal, but instead required the coordinated integration of multiple temporally dynamic, niche-derived signals.

This work successfully established a high-throughput, multiplexed scRNA-seq screening pipeline for profiling hIEO cultures under varying physiological and experimental contexts. While effective for surveying epithelial diversity, this approach failed to robustly support the emergence of rare cell types such as Best4 Cells. Building on these findings, the next section introduces a comparative analytical framework aimed at benchmarking selected protocols that demonstrated more complete epithelial representation, with the goal of identifying culture conditions that enhance differentiation fidelity and support the emergence of rare lineages.

3.3 Comparative Meta-Analysis of Published Organoid scRNA-seq Datasets

3.3.1 Selection and Profiling of External Organoid Datasets

To investigate whether the incomplete differentiation observed in the previous experiment was due to protocol-specific limitations or broader challenges inherent to hIEO systems, three publicly available organoid single-cell RNA sequencing (scRNA-seq) datasets were re-analysed:

1. Elmentaite *et al.* (2020), profiling fetal TI-derived cells cultured in a standard epidermal growth factor (EGF)-based medium (hereafter referred to as "Elmentaite" dataset)[10] ;
2. Holloway *et al.* (2021), focusing on fetal duodenal organoids, grown in a culture medium supplemented with neuregulin 1 (NRG1) instead of EGF ("Holloway")[309];
3. Fujii *et al.* (2018), comprised of organoids derived from both adult TI and colon, maintained in a refined culture condition with a modified EGF stimulation (versus Sato *et al., 2011*) and incorporating insulin-like growth factor 1 (IGF-1) and fibroblast growth factor 2 (FGF-2) ("Fujii")[88].

The Elmentaite and Holloway datasets were generated using 10x Genomics' 3′ Gene Expression assays, similarly to the methodology described in Section 3.2.2. By contrast, the Fujii dataset was generated using Smart-seq2, a plate-based scRNA-seq platform designed to increase transcriptomic depth by enabling full-length transcript coverage with high sensitivity[269]. This technique is particularly advantageous for detecting low-abundance or transcriptionally subtle epithelial populations that are often underrepresented in droplet-based platforms due to limited coverage or capture efficiency. UMAP plots from published analyses are presented in Supplementary Figure S4.

Despite differences in donor age, tissue origin, culture conditions, and protocol duration across these studies, all three reported the emergence of *BEST4+* cells and other rare epithelial

subtypes (Fujii reported *BEST4+* cells as M-Cells as these had not yet been defined a separate group). Their presence across diverse datasets supported the hypothesis that, under specific conditions, multiple hIEO culture systems can generate these rarer lineages, underscoring the relevance of methodological variables in epithelial differentiation outcomes[10,88,309].

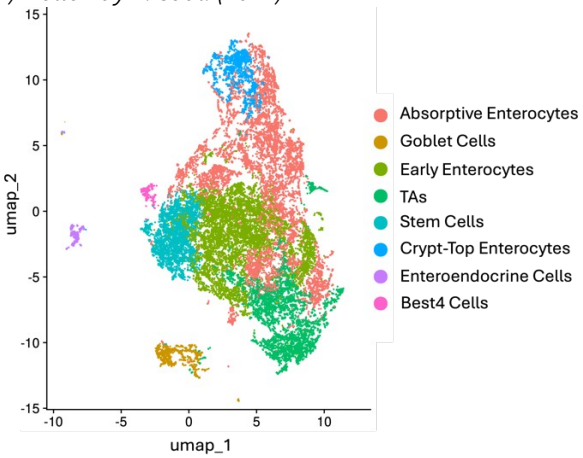
3.3.2 Reprocessing, Clustering, and Identification of *BEST4+* cells

Reprocessing and harmonised annotation following unsupervised clustering demonstrated notable variations in transcriptional diversity and cell-type representation among the 3 datasets. The Fujii dataset exhibited the broadest range of epithelial lineages, although specialised populations such as Tuft Cells and M-Cells remained rare or absent across all datasets (Figure 7 A,C,E). Both Fujii and Holloway datasets contained identifiable clusters of Best4 Cells — 37 in Fujii and 211 in Holloway. In Holloway, an additional 261 *BEST4+* cells were distributed across absorptive and transitional clusters, indicating partial or heterogeneous identities. By contrast, the Elmentaite dataset displayed only sparse, dispersed expression of Best4-associated markers: a total of 472 *BEST4+* cells were identified across multiple absorptive clusters, yet these failed to consolidate into a discrete transcriptional cluster.

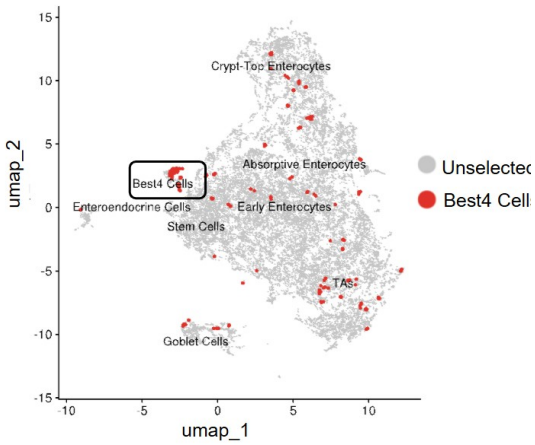
These findings suggest that the emergence of a coherent Best4 cluster is highly dependent on culture conditions, with only Fujii and Holloway achieving the transcriptional fidelity required to resolve the full lineage identity.

Figure 7 - Meta-analysis of scRNA-seq data from independent, publicly available organoid datasets

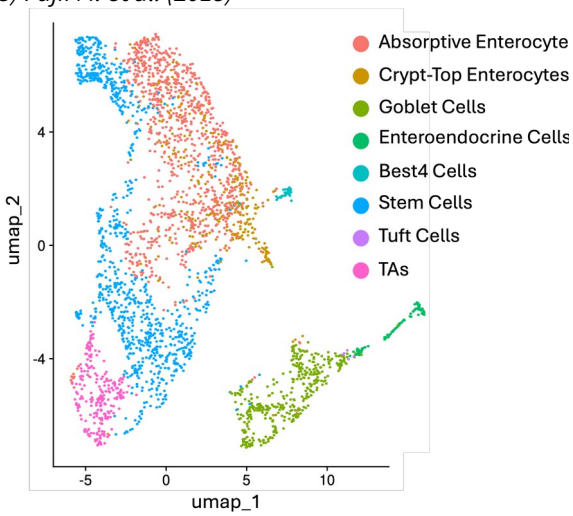
A) Holloway E. et al. (2021)



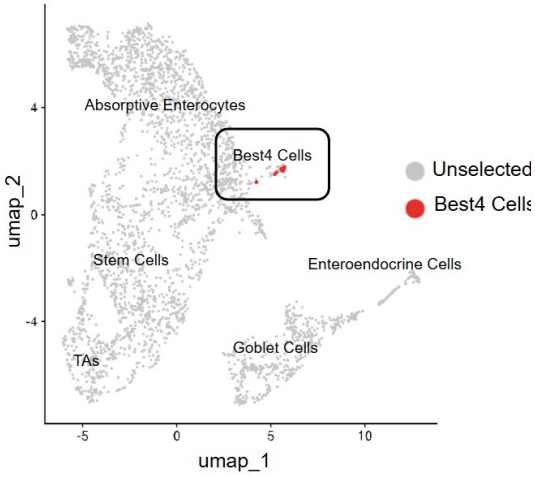
B) Best4 Cells in Holloway E. et al. (2021)



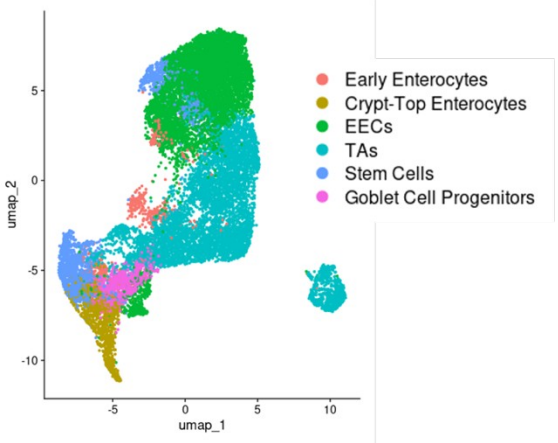
C) Fujii M. et al. (2018)



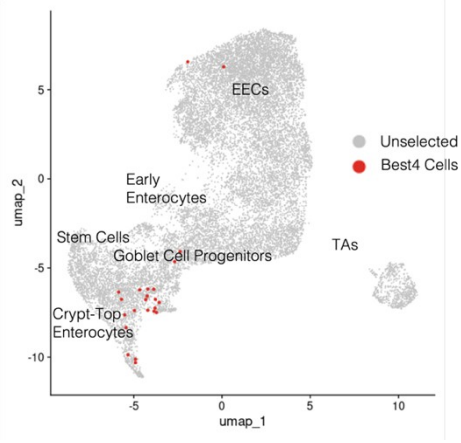
D) Best4 Cells in Fujii M. et al. (2018)



E) Elmentaite R. et al. (2021)



F) Best4 Cells in Elmentaite R. et al. (2021)



Legend: Each dataset is annotated by cell type (left panel, figures A, C, E) and by expression of Best4⁺ cells in the right panels (B, D, F). The presence or absence of a transcriptionally distinct Best4 Cell cluster is indicated, with any identifiable clusters circled in black.

3.3.3 Integrated Transcriptomic Landscape of hIEO-derived Datasets

To further contextualise the observed differences in epithelial lineage representation, an integrative analysis was performed by combining reprocessed single-cell datasets from Elmentaite, Fujii, and Holloway with the subset of the House dataset analysed in Section 3.2.5. All four datasets were subjected to batch correction, normalisation, and scaling within a harmonised computational pipeline to mitigate technical variability and ensure consistent cell-type annotation.

This integrative framework enabled both visual and quantitative cross-study comparisons of epithelial composition, particularly enhancing resolution for rare populations such as Best4 Cells by increasing the cumulative cell count and statistical power. The resulting unified UMAP projection revealed a coherent transcriptional landscape of the intestinal epithelium across divergent protocols and sources, highlighting both conserved lineages and protocol-specific divergences in differentiation outcomes (Figure 8).

Following integration and unsupervised clustering of 55,258 cells across datasets, epithelial cells from all sources coalesced into transcriptionally coherent clusters defined by canonical lineage markers (Figure 8 A,C). Effective batch correction was further confirmed by dataset origin overlays, which demonstrated strong alignment across independent datasets, indicating successful integration without artificial segregation by source (Figure 8B). The integrated UMAP resolved major epithelial lineages, including Stem Cells, TAs, Early Enterocytes, Absorptive Enterocytes, Crypt-Top Enterocytes, Goblet Cells, EECs, and a distinct cluster of Best4 Cells (Figure 8 A,C).

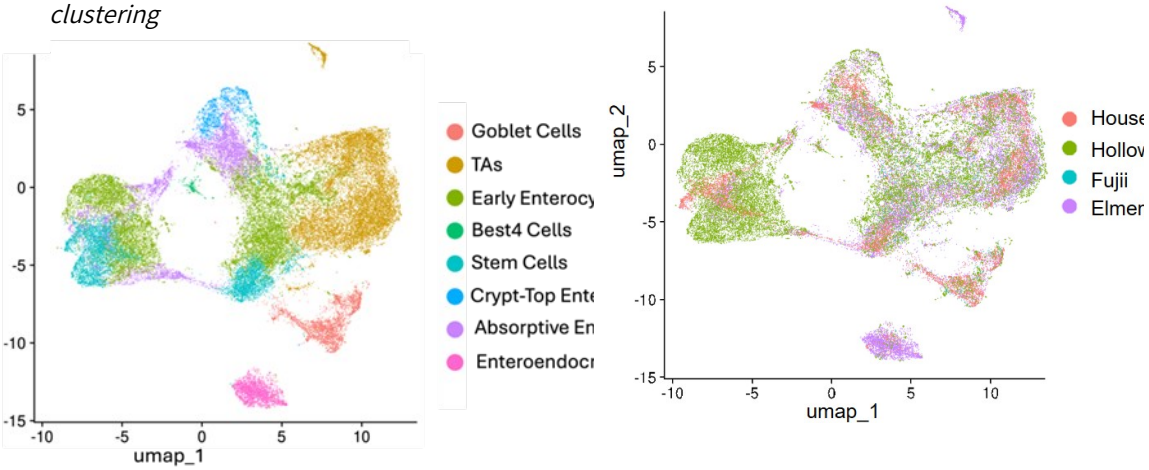
Despite improved resolution, integration highlighted limitations in detecting certain rare populations. For example, the Tuft Cell cluster originally identified in Fujii — comprising fewer than ten cells — dissipated following integration, likely due to statistical dilution within the

harmonised dataset (Figure 8A; Figure 7C). Conversely, a discrete Best4 Cell cluster emerged robustly, with contributions from all four datasets — including House and Elmentaite, which failed to resolve this population independently (Figure 8A).

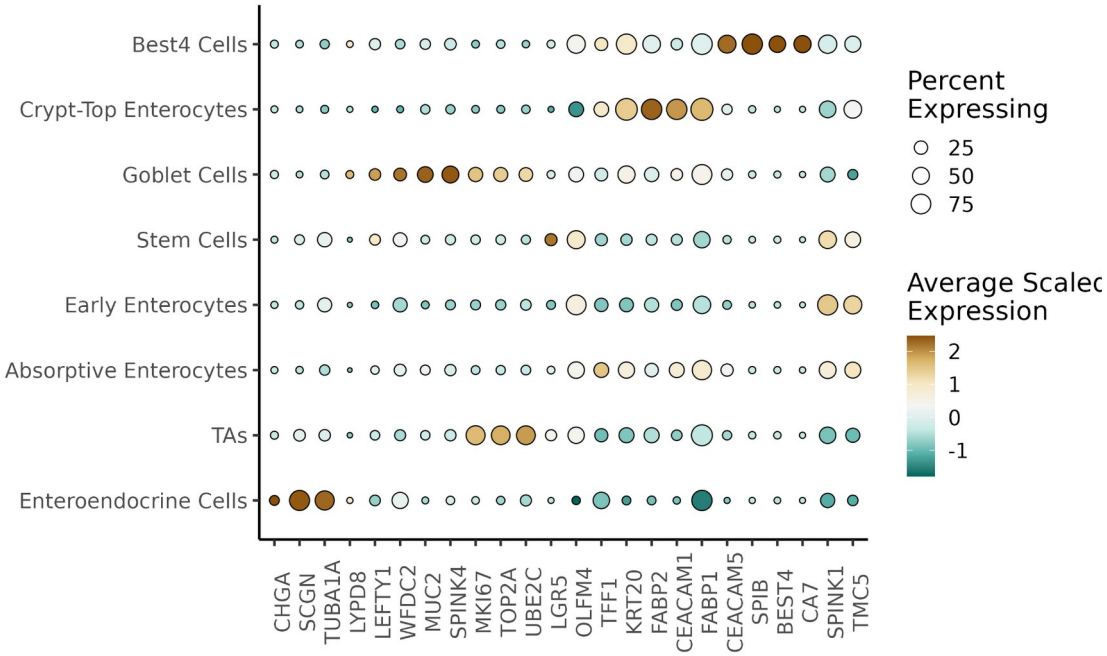
Figure 8 - Integration and annotation of epithelial cell types across harmonised intestinal datasets

A) Labelled cell types following harmonised clustering

B) UMAP showcasing dataset origin correction



C) Transcriptomic profiles across cell types



Legend: (A) UMAP visualisation of harmonised clusters annotated into eight epithelial cell types: Goblet Cells, Transit-Amplifying Cells (TAs), Early and Absorptive Enterocytes, Best4 Cells, Stem Cells, Crypt-Top Enterocytes, and Enteroendocrine Cells (EECs). (B) UMAP coloured by dataset origin (House, Holloway, Fujii, Elmentaite), showing effective integration with minimal batch segregation. (C) Expression of marker genes across annotated cell types.

3.3.4 Validation of Best4 Marker Expression Across Datasets

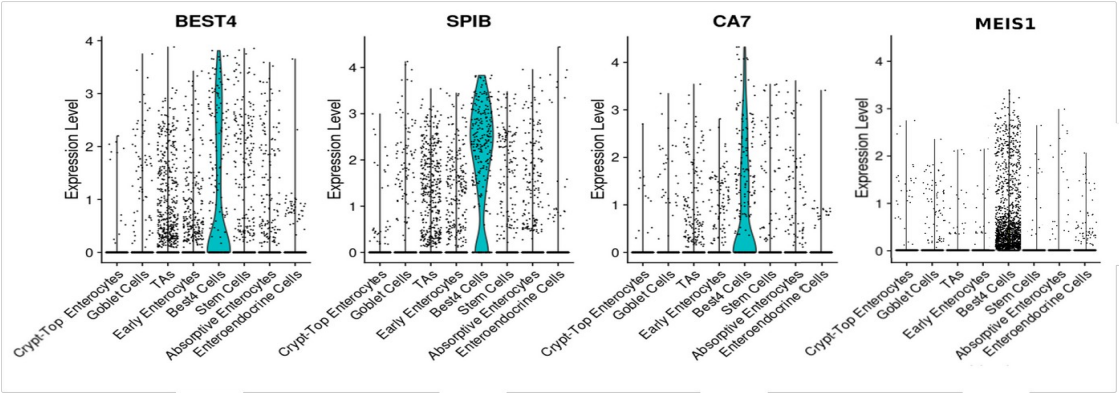
To confirm the identity of the Best4 cluster within the integrated dataset, transcriptomic profiles were examined for lineage-specific markers. Cells within this cluster demonstrated robust, spatially enriched co-expression of *BEST4, SPIB, CA7,* and *MEIS1*, reinforcing the transcriptional fidelity of this population (Figure 9A). Notably, *SPIB* exhibited more consistent expression than *BEST4*, supporting its role as an earlier lineage marker. Both *CA7* and *MEIS1* overlapped strongly with the cluster, providing reliable secondary validation.

By contrast, *OTOP2* displayed weak and diffuse expression across multiple epithelial subsets, limiting its specificity (Figure 9B). Additional candidate markers — *GUCA2A/B*, *CA4*, *GFI1B*, and *HES4* — showed low, inconsistent, or broad expression patterns. *HES4* in particular exhibited widespread transcriptional activity beyond the Best4 cluster, reducing its discriminatory value.

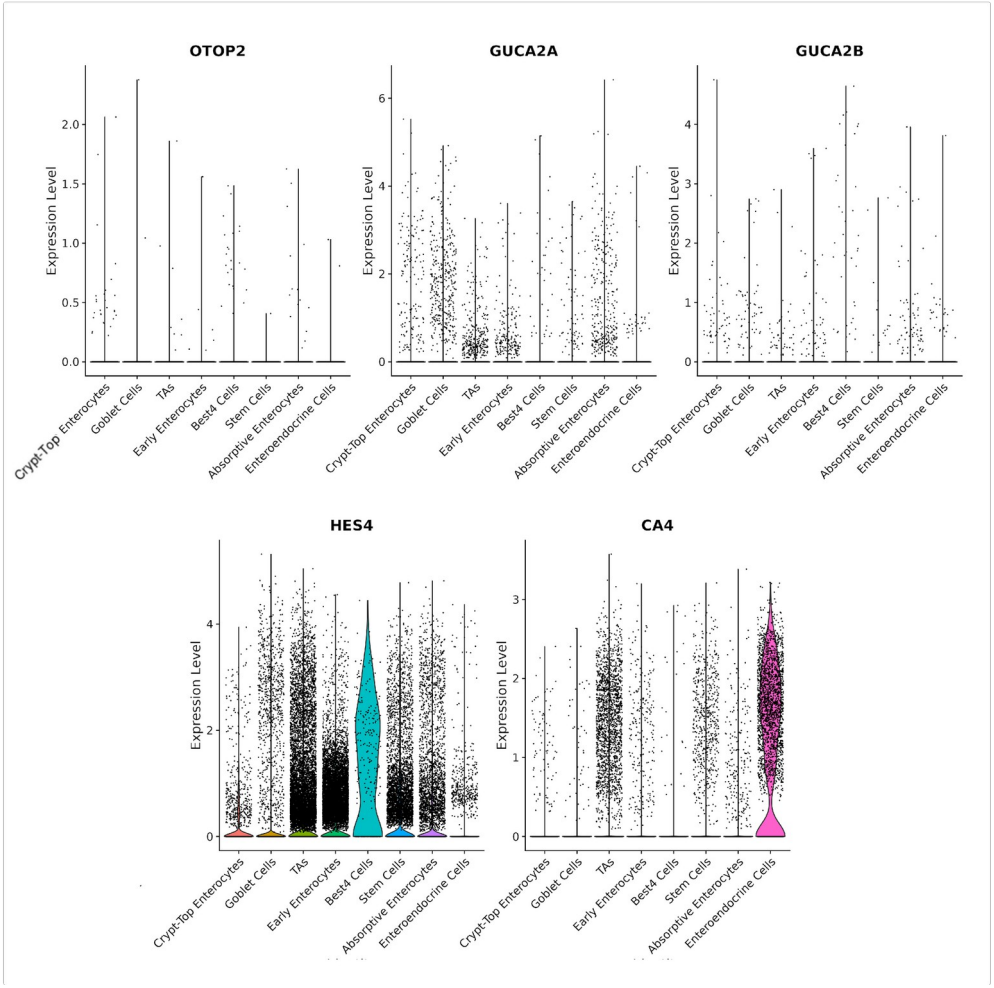
Cells expressing low levels of *BEST4* and *SPIB* were also enriched among TAs and Early Enterocyte populations, suggesting the presence of transcriptionally primed progenitors initiating — but not completing — Best4 lineage commitment (Figure 9A). Accordingly, the UMAP in Figure 9C revealed that, although a transcriptionally coherent Best4 cluster had formed, a subset of *BEST4+* cells remained scattered across adjacent absorptive states, indicative of transitional or incompletely specified progenitors. These observations support a gradient model of differentiation whereby *SPIB* — a transcription factor essential for Best4 lineage specification — is upregulated prior to full induction of *BEST4* and its associated markers, suggesting a progressive model of lineage commitment.

Figure 9 - Molecular and spatial characterisation of Best4 Cells within the intestinal epithelium

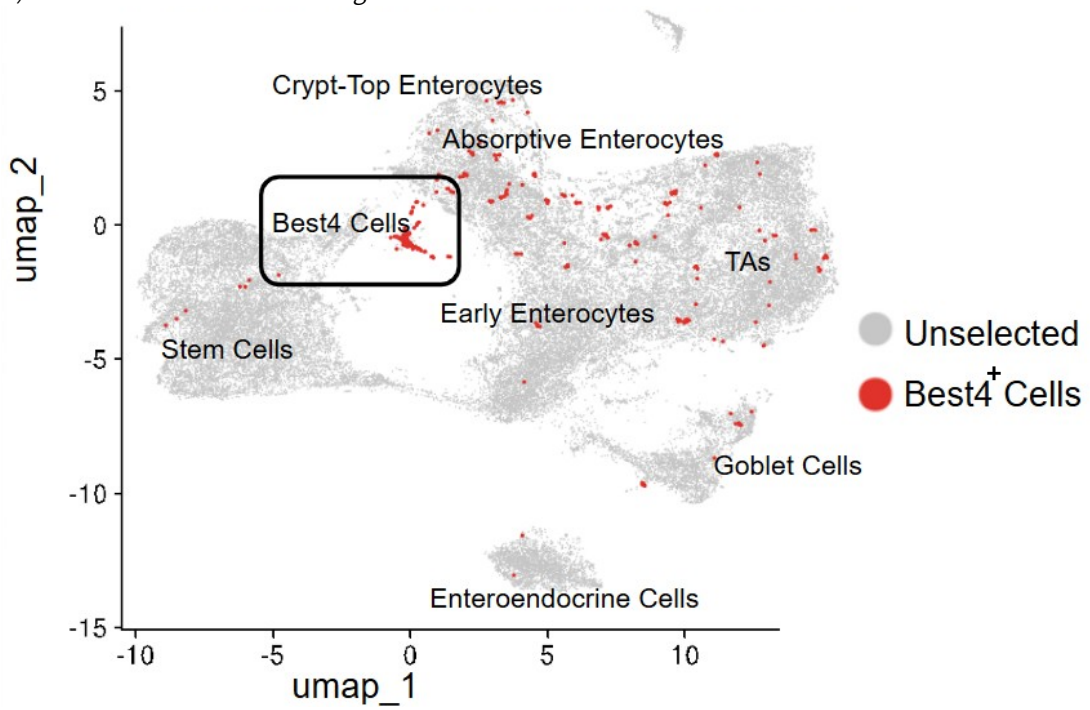
A) Best4 cluster signature markers defined in this study



B) Other canonical markers typically expressed in primary Best4 Cells



C) Best4 Cells' cluster in the integrated dataset



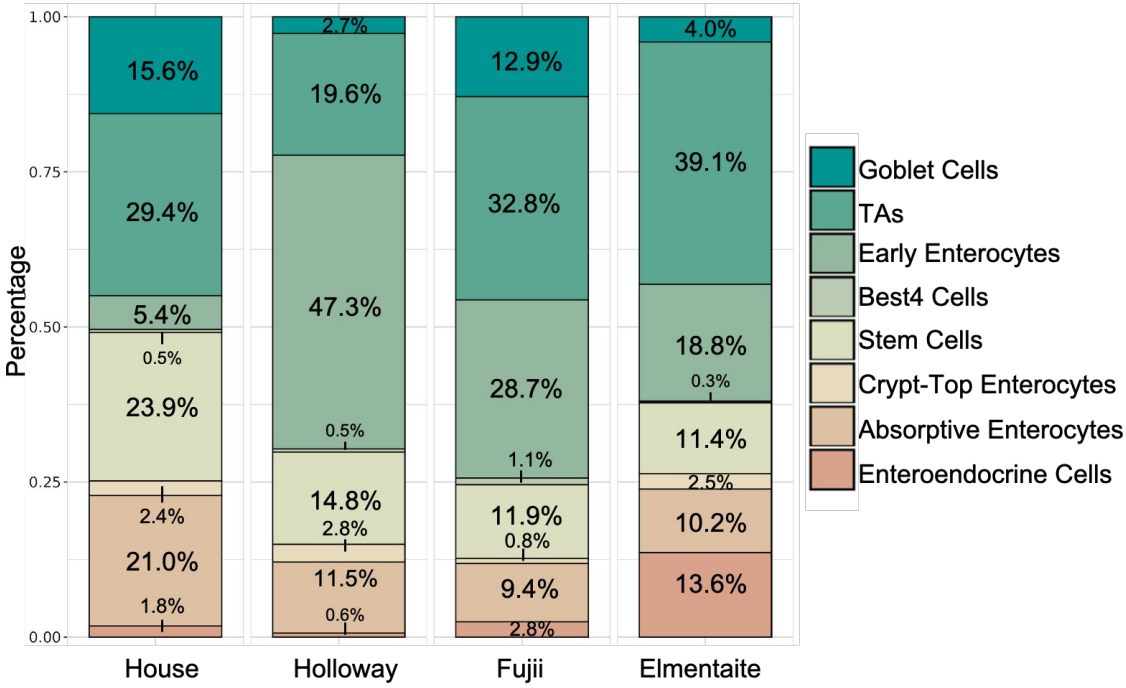
Legend: (A–B) Expression of the canonical Best4 Cell markers used in this study (BEST4, SPIB, CA7, and MEIS1) and other primary Best4 Cell associated genes (OTOP2, GUCA2A/B, HES4, CA4) across annotated epithelial cell types. (C) Spatial localisation of Best4+ cells within the epithelial landscape. A distinct transcriptional cluster of Best4 Cells (highlighted within the black circle) formed, adjacent to Absorptive/ Crypt-Top Enterocytes and Stem Cells. Together, these data reinforce the molecular distinctiveness and defined localisation of Best4+ cells within the differentiated epithelium.

3.3.5 Quantitative Benchmarking of Epithelial Lineage Representation

To evaluate the impact of culture conditions on epithelial differentiation, the relative abundance of major lineages was quantified across datasets (Figure 10). Holloway displayed the strongest absorptive bias, with Enterocytes spanning various maturation stages comprising 61.6% of the epithelial landscape, and a modest Best4 Cells fraction of 0.5%. Elmentaite showed the highest secretory bias, with 13.6% EECs and 4% Goblet Cells, totalling 17.6%. House featured the largest undifferentiated cell fraction (53.3%), but also the highest Goblet Cell proportion (15.6%). By contrast, Fujii demonstrated the most balanced epithelial architecture, comprising 40% absorptive, 15.7% secretory, and 44.7% undifferentiated cells. It

also yielded the highest proportion of Best4 Cells (1.1%), suggesting that this protocol may have a slightly enhanced capacity to generate broader epithelial diversity.

Figure 10 - Relative proportions of epithelial cell types across datasets, after integration



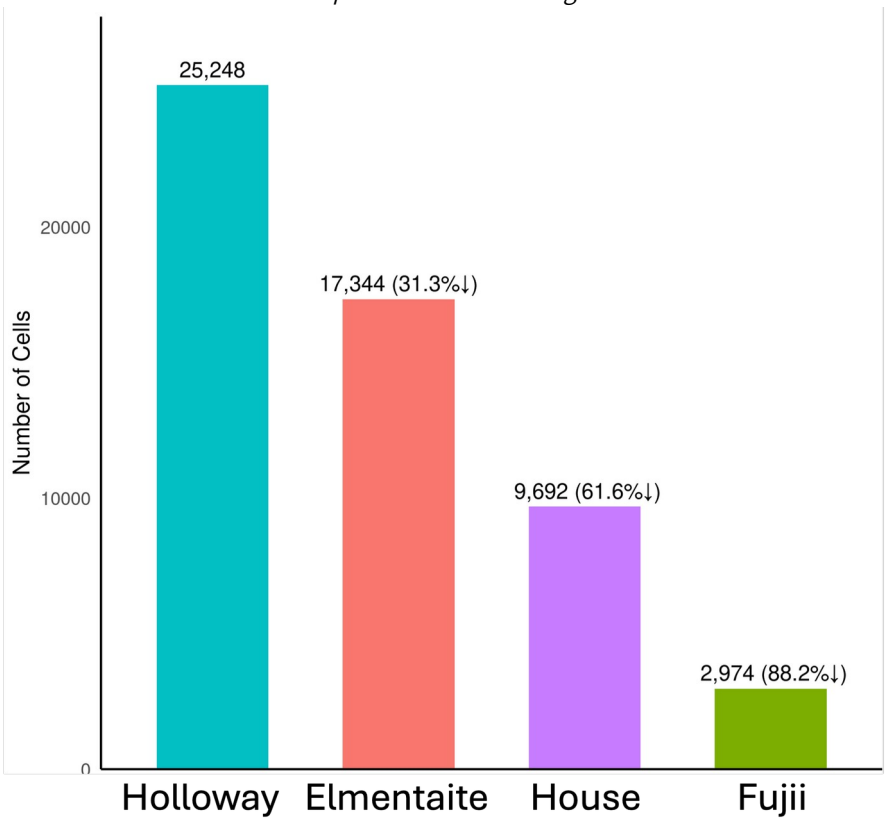
Legend: Relative abundance of annotated epithelial cell types within each dataset (House, Holloway, Fujii, and Elmentaite). While most major cell types are present across datasets, their proportions vary significantly. Best4 Cells are consistently low across all datasets.

3.3.6 Downsampled Integration for Dataset Size Normalisation

As illustrated in Figure 11, the number of cells contributed by each dataset varied considerably — from approximately 3,000 cells in the Fujii dataset to over 25,000 in the Holloway dataset. This discrepancy reflects methodological differences: the Fujii dataset was generated using Smart-seq2, a plate-based scRNA-seq method with enhanced transcriptomic resolution but inherently lower throughput, whereas the other datasets employed the higher-throughput 10x Genomics droplet-based platform. These disparities have critical implications for downstream analyses; for instance, although a discrete Tuft Cell cluster was detectable in the original Fujii dataset (Figure 7C), this population was lost during integration due to insufficient statistical

representation in the harmonised dataset. This pattern highlights the potential for skewed lineage assignment due to uneven sampling depths and/or cell number representation and underscores the importance of appropriate dataset balancing when comparing protocols or conducting cross-study benchmarking.

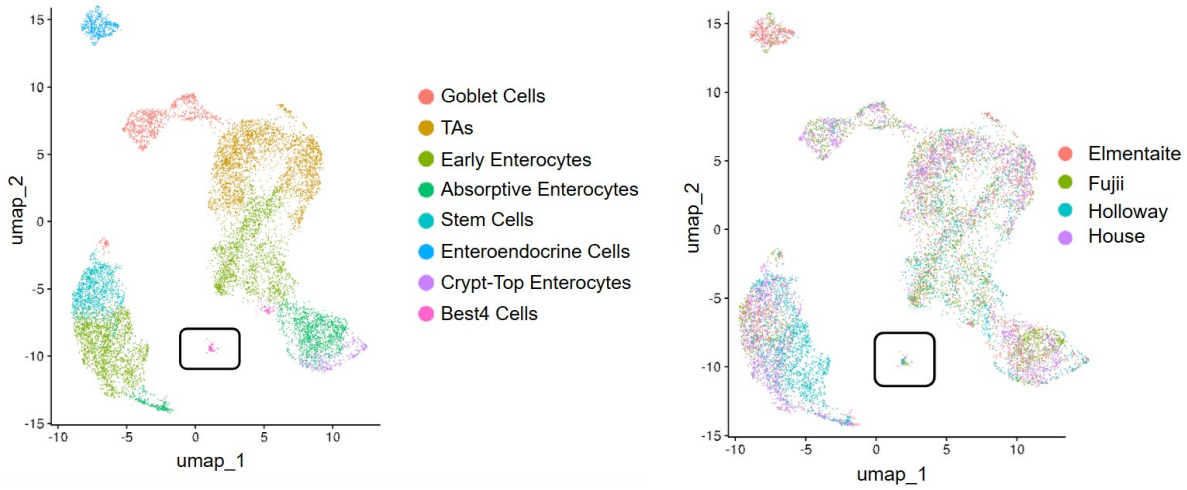
Figure 11 – Number of cells contributed per dataset after integration



To address this issue, the integrated dataset was downsampled to match the smallest dataset — Fujii, with 2,974 cells (Figure 11). This reprocessing allowed for unbiased comparisons of lineage representation across culture conditions. Following reprocessing and UMAP re-embedding, major epithelial lineages remained well-resolved, and rare populations — including EECs and Best4 Cells — were still detectable (Figure 12A). Figure 12B confirms that cell contributions from each dataset were balanced.

Figure 12 – Positioning of the Best4 Cells cluster within the integrated dataset after size normalisation

A) UMAP coloured by cell cluster *B) UMAP coloured by dataset origin*

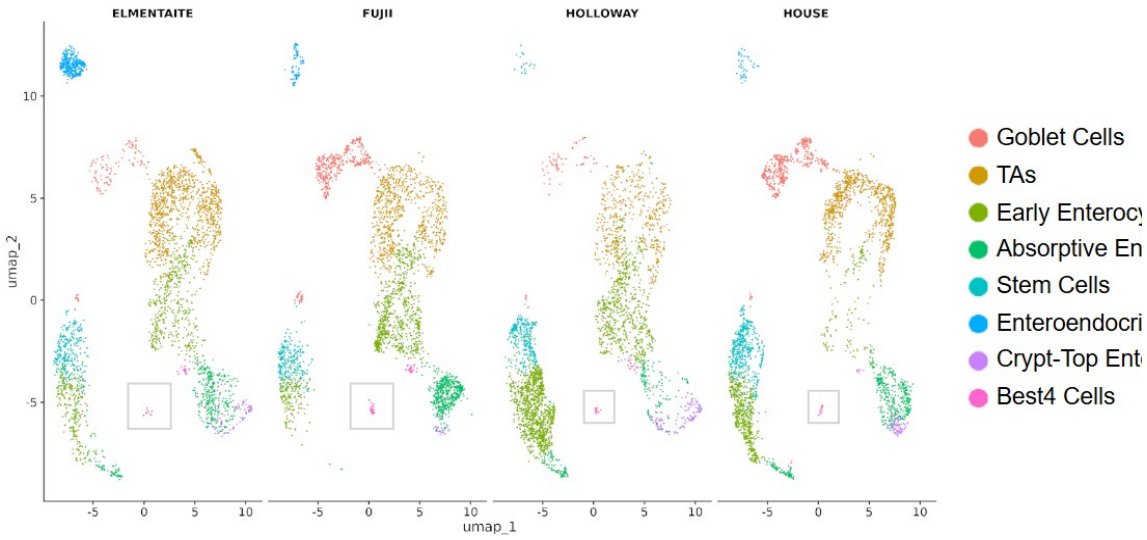


Legend: (A) UMAP coloured by annotated cell type, highlighting the distinct localisation of Best4 Cells (boxed) within the epithelial landscape. (B) UMAP coloured by dataset of origin, showing that Best4 Cells (boxed) are still present across multiple datasets.

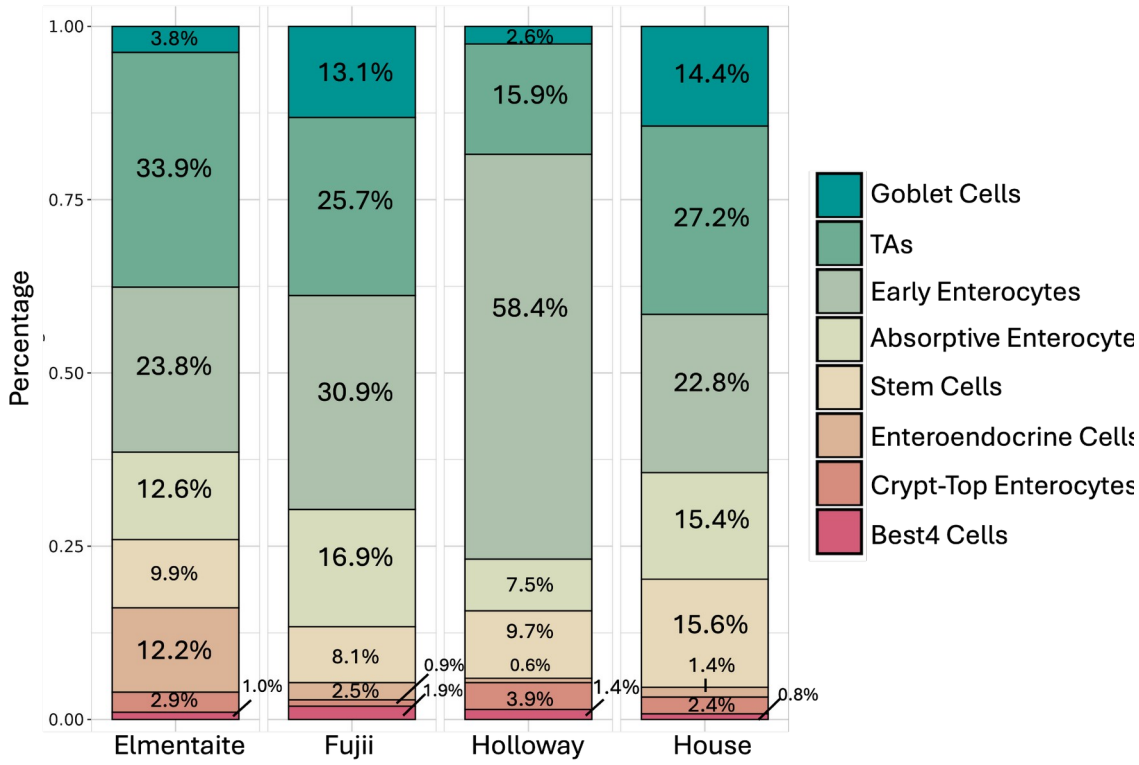
Subsequently, condition-specific UMAPs revealed discernible protocol-dependent lineage biases (Figure 13A): Goblet Cells were enriched in the House and Fujii datasets, whereas Elmentaite maintained the highest EEC representation, in line with previous observations (Figure 10). Holloway featured the highest proportion of Early Enterocytes and TAs, along with a robust Crypt-Top Enterocyte presence. Importantly, after downsizing, the relative abundance of Best4 Cells was still the highest in the Fujii dataset (1.9%), followed by Holloway (1.4%), Elmentaite (1.0%), and House (0.8%) (Figure 13B). These results reinforce the importance of methodical benchmarking to capture rare lineage output and protocol-specific differentiation capacity within organoid systems.

Figure 13 – Cell clustering and prevalence after integration and downsampling, by dataset of origin

A) Unsupervised cell cluster annotation split across datasets of origin



B) Cell type distribution per dataset



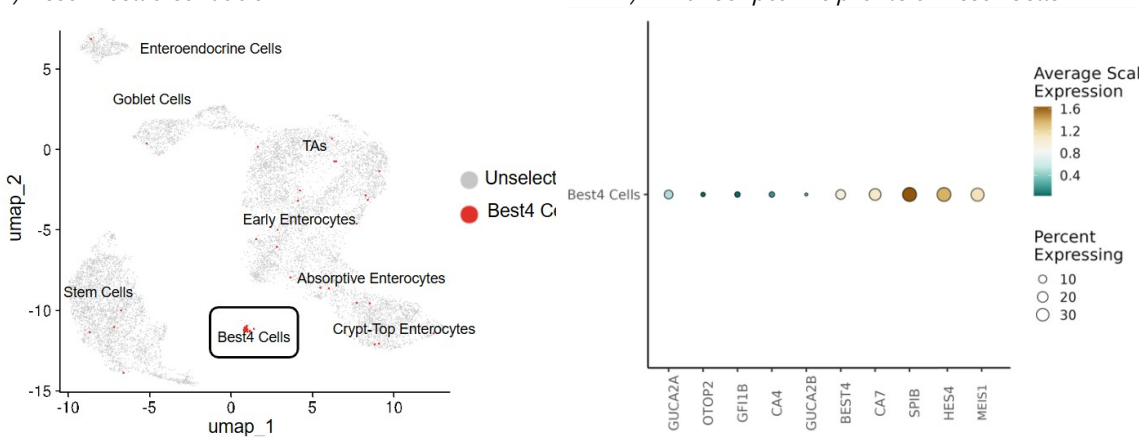
Legend: (A) Cell type annotation within each dataset after integration and downsampling. Best4 Cells (pink) consistently localise to a distinct region (boxed) in each dataset, confirming their reproducible detection. (B) Proportional distribution of annotated cell types per dataset. Despite compositional variability across studies, Best4 Cells remain detectable in all samples.

3.3.7 Evaluation of Best4 Identity in Equalised Cell Populations

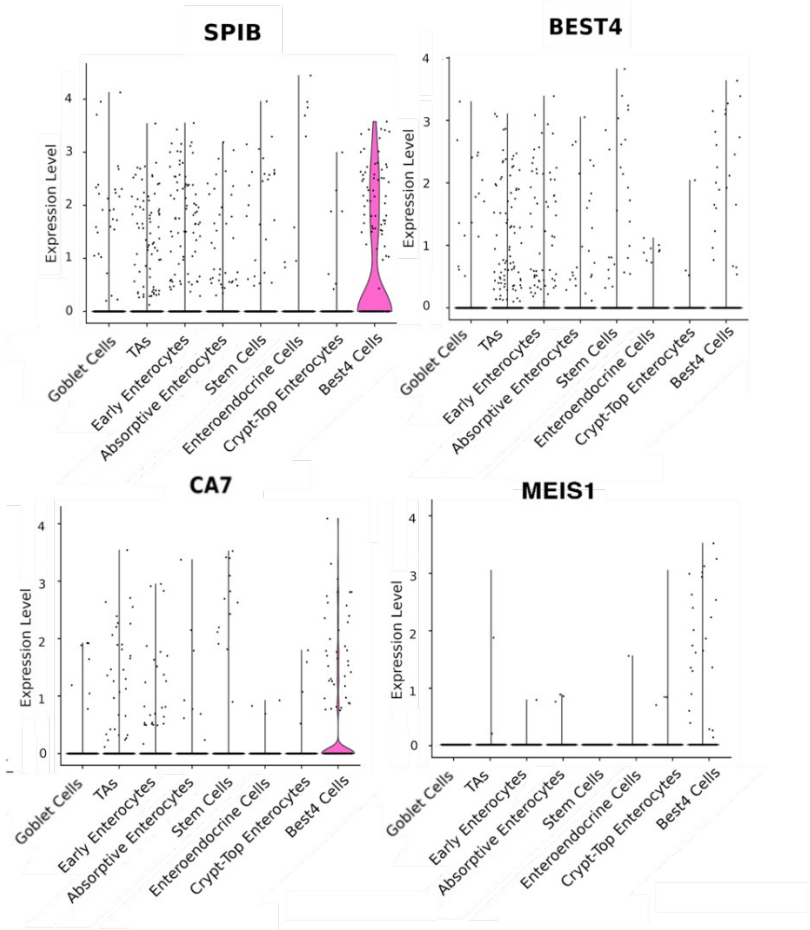
Analysis of the downsampled UMAP projection revealed that most *BEST4+* cells continued to localise within a discrete cluster situated between Stem Cells and absorptive lineages (Figure 14). As in the full dataset, a significant fraction of *BEST4+* cells remained scattered across adjacent clusters, consistent with transitional or partially differentiated states (Figure 14A). Marker gene expression within the downsampled Best4 cluster confirmed enriched co-expression of *BEST4*, *SPIB*, *CA7* and *MEIS1,* reinforcing the specificity of this signature (Figure 14; Supplementary Figure S5). While *SPIB* and *CA7* retained clear cluster enrichment, *BEST4* and *MEIS1* expression were proportionally lower and more broadly distributed, particularly among undifferentiated or early committed populations and absorptive lineages (Figure 14). This pattern aligns with a model of progressive differentiation, where transcriptional commitment precedes terminal marker upregulation.

Figure 14 – Detailed transcriptomic profiling of Best4 cells in the downsampled, integrated dataset

A) Best4+ cell distribution *B) Transcriptomic profile of Best4 Cells*



C) Expression of canonical Best4 Cell genes across cell types



Legend: (A) UMAP highlighting Best4+ cells (red) within the epithelial landscape compared to all cells (grey). (B) Average scaled expression and percentage of expressing cells for key marker genes within the Best4 Cell cluster. Canonical Best4 Cell markers (BEST4, SPIB, CA7, MEIS1) are specifically enriched. (C) Expression of SPIB, BEST4, CA7, and MEIS1 across cell clusters, illustrating selective enrichment of these genes within Best4 Cells relative to other epithelial cell types.

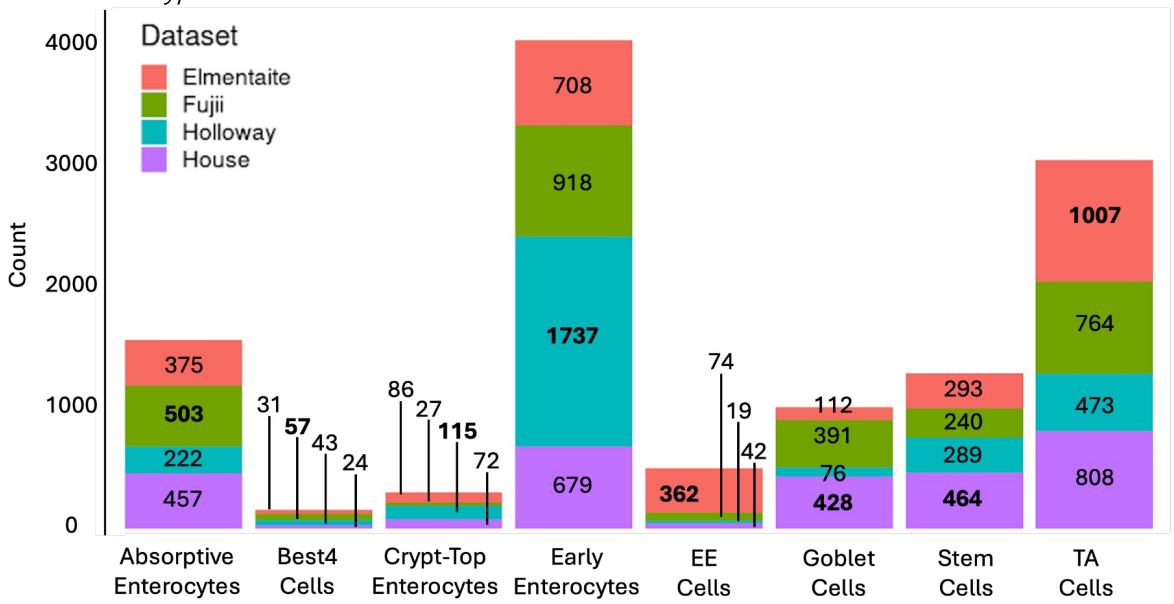
3.3.8 Cross-Dataset Fidelity and Marker Expression in Best4 Cells

The absolute contribution of cells to the Best4 cluster varied across datasets: Fujii contributed the greatest number (n = 57), followed by Holloway (n = 43), Elmentaite (n = 31), and House (n = 24) (Figure 15A). To evaluate lineage fidelity, cells were classified using progressively stringent marker criteria: (1) detection of the *BEST4* gene alone; (2) assignment to the Best4 cluster via unsupervised clustering; (3) co-expression of *BEST4* and *SPIB*; and (4) co-expression of *BEST4*, *SPIB*, and *CA7* (Figure 15; Supplementary Figure S6). Across all datasets, only a minority of *BEST4+* cells satisfied this signature threshold, and many were located outside the core cluster — suggesting incomplete or partial lineage commitment.

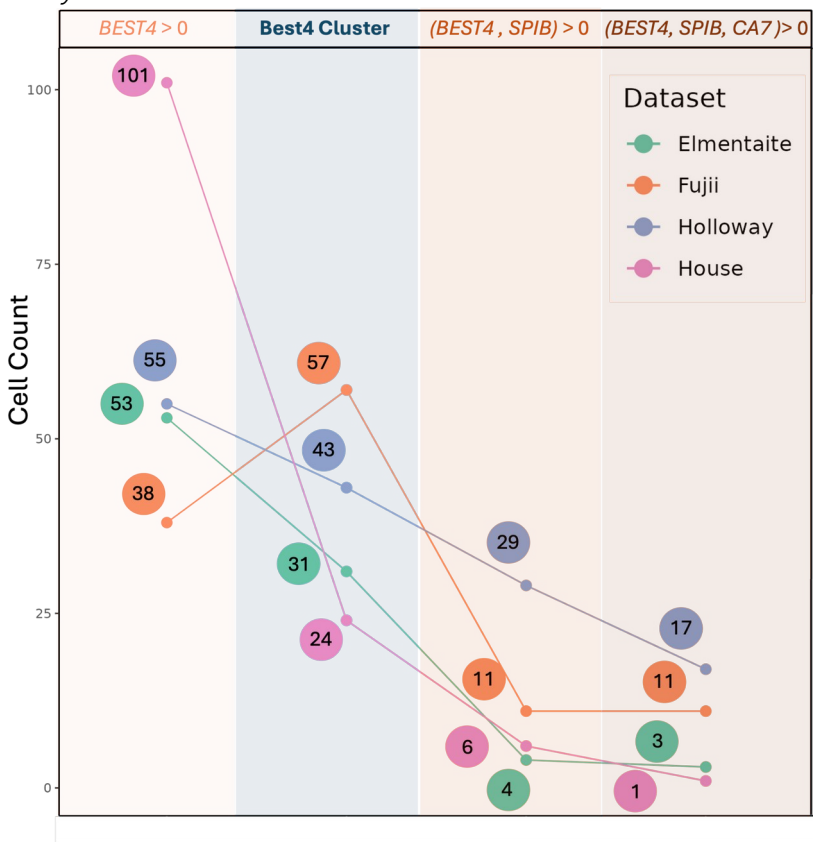
Intriguingly, in the Fujii dataset, the Best4 cluster included more cells than were identified as *BEST4*-expressing cells. Transcriptomic profiling revealed that these *BEST4*- (negative) cells exhibited elevated expression of *SPIB, FABP1, FABP2, HDAC1,* and *SOD2* — markers associated with early absorptive progenitor states — and no evidence of M-cell specific markers, suggesting that this cluster encompasses cells in a pre-Best4 stage (Supplementary Figure S7). By contrast, other datasets struggled to consolidate actual *BEST4+* cells into a coherent cluster, likely due to lower transcriptomic resolution, which resulted in these cells appearing more transcriptionally similar to other closely related epithelial cell types.

Figure 15 – Best4 cell abundance and transcriptomic fidelity across datasets after integration and dataset balancing

A) Absolute cell type abundance across datasets



B) Transcriptomic fidelity of Best4+ cells and cells within the Best4 Cells' cluster across datasets



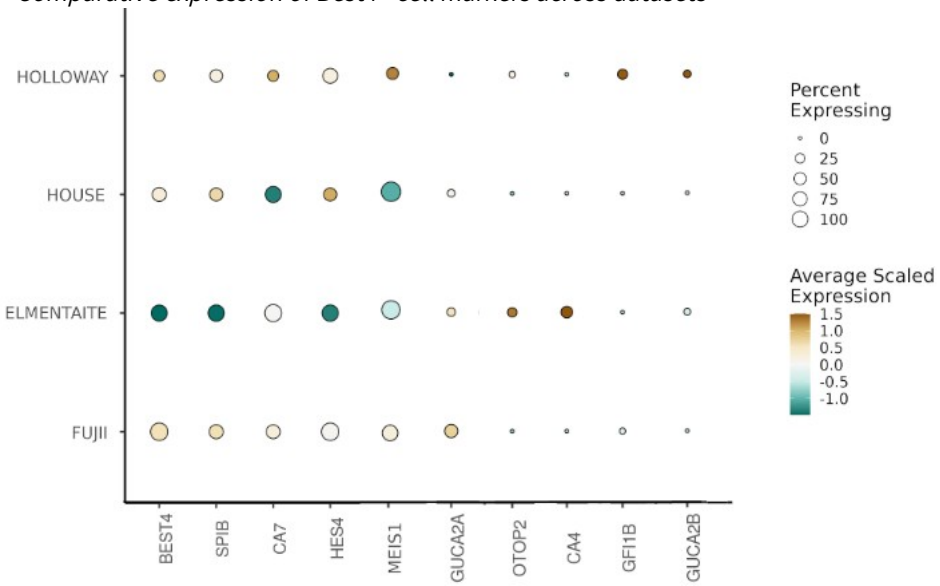
Legend: (A) Absolute number of cells per annotated epithelial cell type, stratified by dataset. Best4 Cells are rare across all datasets but consistently identified. (B) Progressive filtering of Best4+ cells using increasingly stringent transcriptional criteria: (1) BEST4 expression > 0, (2) Best4 cluster assignment resulting from unsupervised clustering, (3) BEST4 and SPIB > 0, (4) BEST4, SPIB, and CA7 > 0. Overall, cell

counts decrease accordingly, with lines coloured by dataset, reflecting variation in transcript detection across studies.

A comparative analysis of Best4 Cells representation across datasets (Figure 16) revealed substantial differences in both cluster contribution and marker fidelity, in line with the previous analysis (Figure 15B). While the Fujii dataset contributed the highest number of cells to the Best4 cluster, these cells exhibited a more immature transcriptional profile — characterised by lower average expression of canonical Best4 markers — compared to those in the Holloway dataset. By contrast, Holloway-derived Best4 Cells demonstrated more robust co-expression of *BEST4*, *SPIB*, *CA7* and *MEIS1*, suggesting a more mature and transcriptionally consolidated identity. Interestingly, the House dataset showed distinguishable expression of *BEST4* and *SPIB* but had notably low levels of *CA7* and *MEIS1*. Similarly, despite diverging the most from the canonical Best4 signature, Elmentaite-derived Best4 Cells showed the highest expression of *OTOP2* and *CA4*.

Overall, these findings highlight the protocol-dependent variability in Best4 differentiation and underscore the importance of applying both quantitative (e.g., cluster size, expression intensity) and qualitative (e.g., marker coherence, transcriptional maturity) metrics when evaluating rare and underrepresented epithelial lineages in *in vitro* models.

Figure 16 - Comparative expression of Best4+ cell markers across datasets



Legend: Expression of canonical and Best4 marker genes (BEST4, SPIB, CA7, MEIS1, and other primary Best4 Cells' markers such as GUCA2A/B, OTOP2, HES4, CA4, GIF1B) across datasets (Holloway, House, Elmentaite, Fujii). Dot size reflects the percentage of Best4 Cells expressing each gene, and colour indicates average scaled expression.

3.3.9 Interpretation of Findings

The work presented in this section aimed to establish a standardised analytical framework for the evaluation of single-cell hIEO datasets. Publicly available datasets were selected based on their reported ability to capture a broad and diverse epithelial landscape, with Best4 Cells used as a case study for rare lineage analysis. Furthermore, the dataset generated in this study (section 3.2) was incorporated into the meta-analysis to allow for direct, side-by-side comparisons. By systematically assessing the impact of protocol design on epithelial lineage representation, this framework enabled robust validation of lineage fidelity and protocol reproducibility across diverse experimental conditions. Notably, it proved effective in detecting and comparing rare epithelial populations, such as Best4 Cells, and revealing both shared and context-dependent features.

The comparative meta-analysis demonstrated that epithelial lineage specification in hIEO cultures is strongly influenced by experimental design. These findings underscore the critical roles of growth factor context, sample selection, and platform sensitivity in shaping *in vitro* epithelial outcomes. Among the evaluated protocols, those developed by Fujii and Holloway emerged as particularly promising for promoting epithelial complexity and supporting rare lineage fidelity. The Fujii protocol yielded the most balanced epithelial composition, encompassing absorptive, secretory, and undifferentiated populations — supporting its potential suitability for modelling physiologically relevant epithelial architecture. In parallel, the Holloway dataset exhibited the most clearly defined Best4 Cluster, indicating that its culture conditions may better support commitment to this rare fate. Collectively, these findings demonstrate that while Best4 Cell specification appears achievable across multiple hIEO platforms, its fidelity and prevalence remain suboptimal and are likely to be highly context-dependent. This underscores the need for refined protocols and robust validation strategies when assessing rare epithelial identities *in vitro*.

Finally, this analysis further emphasises the importance of rigorous benchmarking and cross-dataset validation when evaluating differentiation efficiency in hIEO systems. Integrated and downsampled analyses reinforced the impact of protocol-specific variables on rare lineage output, underlining the value of harmonised frameworks for distinguishing biological variation from technical noise.

3.4 Technical Evaluation of hIEO Differentiation Protocols Using Single-Cell RNA Phenotyping: A Best4 Case Study

3.4.1 Assessing House Protocol Performance via Targeted Sorting and Full-Length scRNA-Seq

In the previous section, the number of *BEST4+* cells detected under the House protocol was relatively high compared to other protocols analysed, whilst the total number of cells in the Best4 Cell cluster was lower (see Figure 15B). This raised the question of whether Best4 differentiation was genuinely suppressed under the standard culture conditions used in this study, or simply remained unresolved due to technical constraints (such as shallow sequencing depth, or insufficient transcriptomic capture).

To determine whether the absence of a transcriptomically distinct Best4 population in the House protocol was a true biological outcome, or instead a limitation of droplet-based scRNA-seq combined with suboptimal sequencing depth, a targeted, high-sensitivity strategy was deployed. Two full-length, plate-based scRNA-seq protocols were selected to benchmark transcriptomic sensitivity: SmartSeq2, a well-established high-resolution method, and SmartSeq-Total, a newer approach designed to capture both polyadenylated and non-polyadenylated RNAs[333]. SmartSeq-Total employs enzymatic polyadenylation and Cas9-mediated rRNA depletion to broaden transcriptome coverage, enabling detection of long non-coding RNAs (lncRNAs), histone transcripts, and other regulatory RNA species typically missed by poly(A)-biased protocols.

Given the low abundance and transcriptional ambiguity of Best4 Cells within hIEO cultures, fluorescence-activated cell sorting (FACS) was employed to selectively enrich $EPCAM^+/BEST4^+$ epithelial subsets. A validated anti-BEST4 antibody, benchmarked through immunohistochemistry and flow cytometry on human intestinal tissue, reliably labelled a

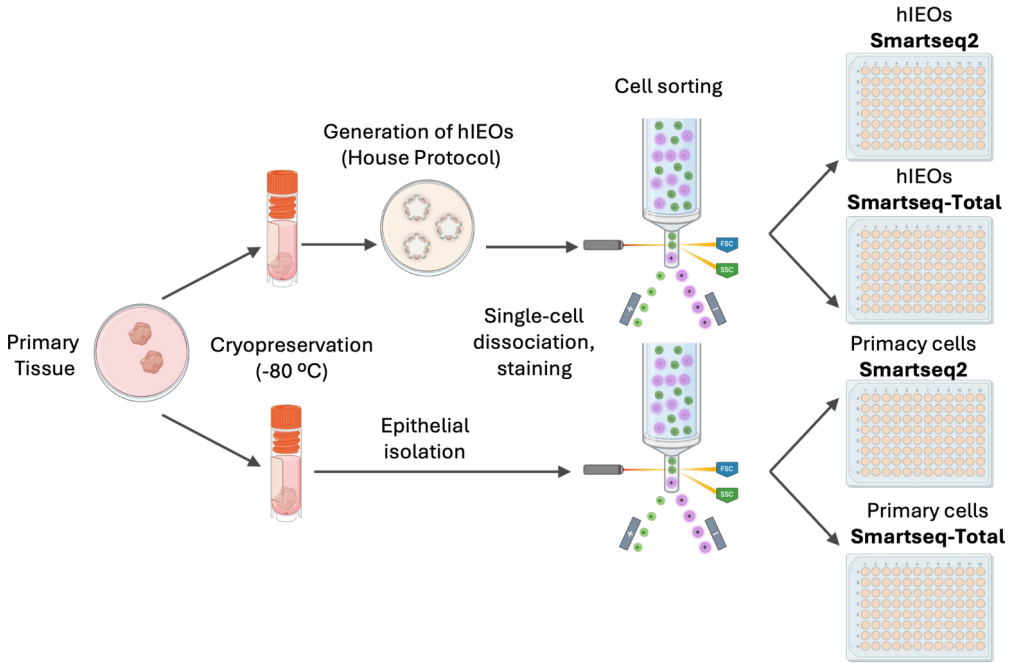
distinct epithelial population (Supplementary Figure S8, Supplementary Figure S9). For this experiment (EXP02), following antibody optimization, live $EPCAM^{+}/BEST4^{-}$ and $EPCAM^{+}/BEST4^{+}$ epithelial cells were isolated from both hIEOs cultured in House Medium and matched primary colonic tissue (Figure 17A). Sorted cells were captured in 96-well plates to preserve transcriptomic integrity and minimise technical dropout.

This sorting and capture strategy was designed to test the biological efficacy of the House protocol, while simultaneously enabling a technical benchmarking comparison between two full-length scRNA-seq platforms. By isolating $EPCAM^{+}/BEST4^{+}$ and $EPCAM^{+}/BEST4^{-}$ epithelial subsets from both hIEOs and matched primary tissue and processing them under identical sequencing and library preparation conditions (Figure 17B), the experiment provided a controlled framework for evaluating transcriptomic resolution across biological and technical contexts.

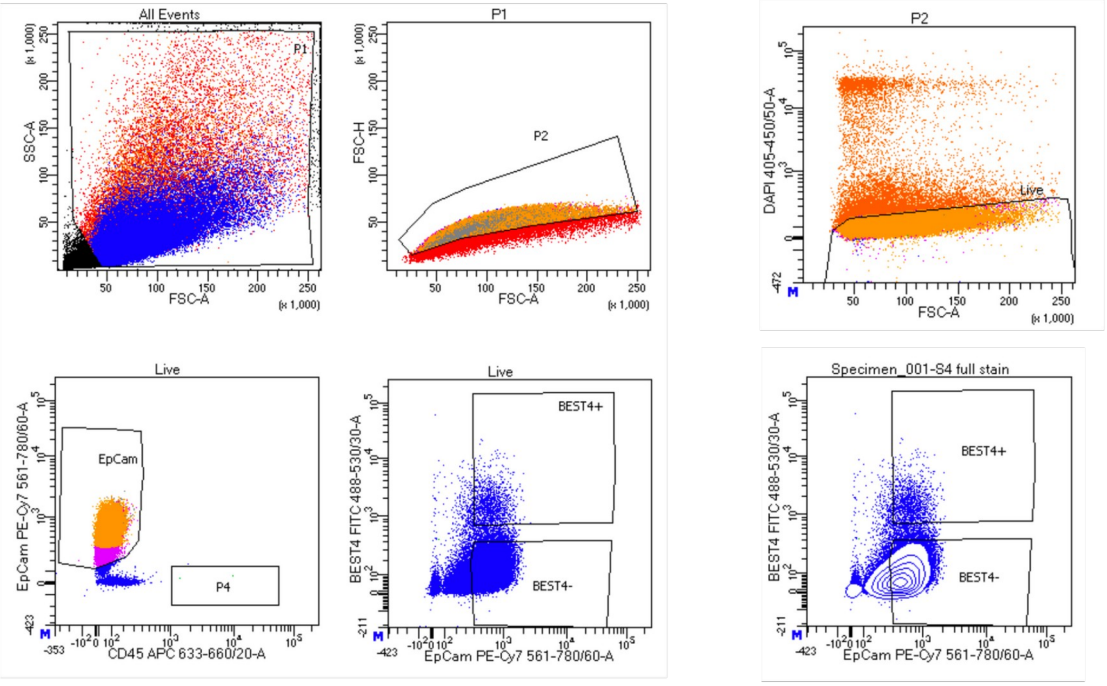
Despite its theoretical advantages, SmartSeq-Total consistently failed to generate high-quality libraries across multiple attempts. Sequencing data revealed low complexity, with read pools dominated by degraded RNA fragments and adapter-dimer contamination. Although the published protocol was revised and modified over the course of this study, applying the suggested adjustments to input quality, reaction volumes, and enzymatic steps was unsuccessful nonetheless (Supplementary Figure S10)[334]. Consequently, SmartSeq-Total was excluded from downstream analyses.

Figure 17 – Experimental design for Experiment 2: isolation of epithelial cells from either hIEOs cultured under the House protocol, or matched primary tissue-derived cells, for full-length transcriptomic profiling

A) Workflow for epithelial isolation and Smart-seq profiling



B) Flow cytometry gating strategy for the isolation of EPCAM⁺/BEST4- and EPCAM⁺/BEST4⁺ epithelial cells



Legend: (A) Schematic of the experimental workflow. Primary intestinal tissue was cryopreserved, followed by epithelial isolation and in vitro expansion into hIEOs using the House protocol. Both primary and hIEO-derived epithelial cells underwent single-cell dissociation, staining, and sorting for downstream Smart-seq2 or SmartSeq-Total profiling. (B) Representative FACS gating strategy for isolating live epithelial cells. Initial gating excluded debris and dead cells (DAPI+), followed by

enrichment for EpCAM+ epithelial cells and subsequent sorting of Best4+ (high expression) and Best4− populations.

3.4.2 Benchmarking SmartSeq2: Capture of *BEST4*$^+$ Cells from hIEOs vs. Primary Epithelium

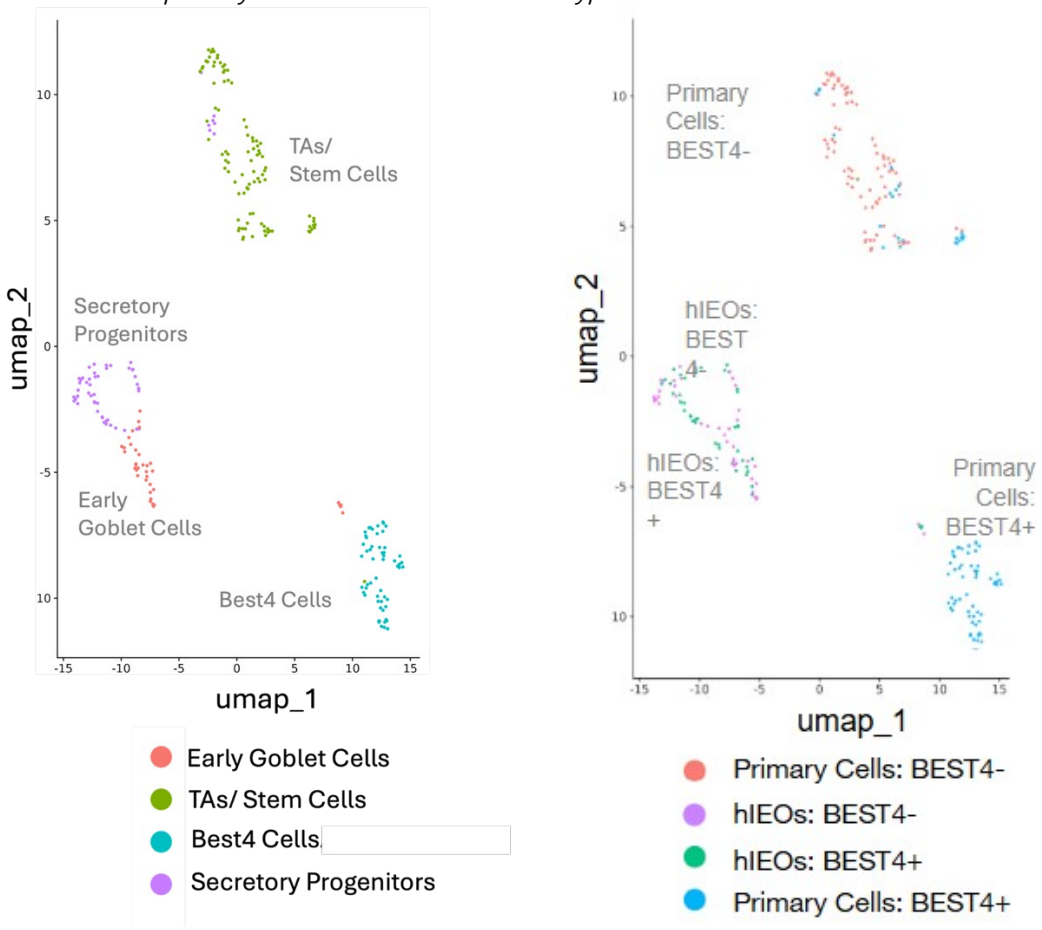
In contrast to SmartSeq-Total, SmartSeq2 delivered robust and reproducible performance across all tested conditions, generating high-quality, full-length transcriptomes from 243 epithelial cells derived from both hIEOs and matched primary colonic tissue. Despite the limited cell number, major epithelial lineages — including Stem Cells, TAs, Goblet Cells, Secretory Progenitors, and Best4 Cells — were reliably recovered from both sample types (Figure 18A; Supplementary Figure S11).

Critically, a pronounced divergence emerged in the detection of Best4 Cells. In primary intestinal tissue, EPCAM$^+$/BEST4$^+$ cells sorted by FACS clustered into a clearly defined, transcriptionally coherent population, enriched for hallmark Best4 markers. By contrast, EPCAM$^+$/BEST4$^+$ cells isolated from hIEOs cultured in House Medium failed to form a distinct cluster and instead appeared transcriptionally diffuse and indistinct (Figure 18B). This discrepancy suggests a biological failure of the House protocol to support mature Best4 differentiation, rather than a technical detection issue — especially given the sensitivity and resolution of SmartSeq2.

Figure 18 – Best4 Cell distribution across both primary tissue and hIEO cultures

A) Annotated cell clustering distribution across both hIEOs and primary cells

B) Distribution of BEST4 expression by sample type



Legend: (A) UMAP projection of sorted cells using SmartSeq2, coloured by inferred cell states. (B) UMAP coloured by dataset and Best4 status.

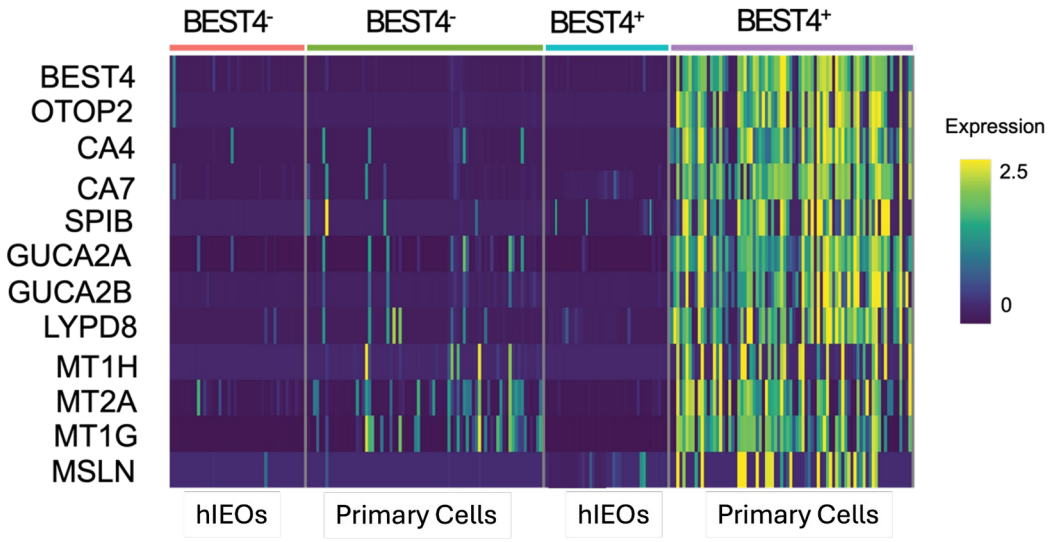
3.4.3 Comparative Transcriptomics: Evaluating Lineage Resolution in hIEOs vs Primary Tissue

Transcriptomic profiling of SmartSeq2 data revealed a clear dichotomy between hIEO-derived and tissue-derived EPCAM$^+$/BEST4$^+$ cells. Marker gene analysis confirmed the absence of canonical Best4 signatures in hIEO-derived cells, which instead mapped predominantly to proliferative and early secretory progenitor compartments (Figure 19; Supplementary Figure S11). This finding held despite the high transcriptomic sensitivity and full-length coverage of Smart-seq2, suggesting that the isolated BEST4$^+$ hIEO cells were likely products of nonspecific antibody binding or misclassification, rather than true members of the Best4 lineage.

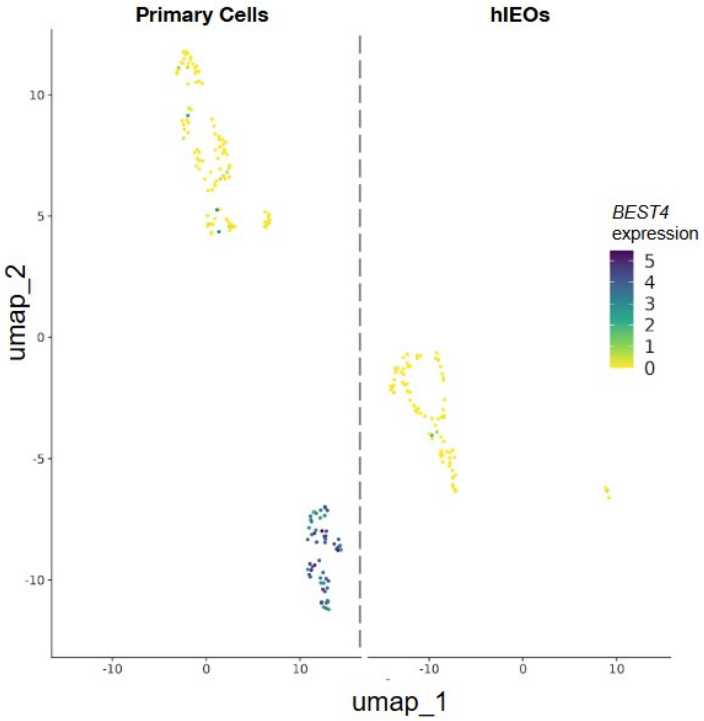
By contrast, BEST4^{+} cells from matched primary tissue displayed transcriptional profiles consistent with those of Best4 Cells identified within *in vivo* reference atlases. These cells robustly co-expressed *BEST4, SPIB*, and *CA7*, and also showed high expression of supporting markers such as *OTOP2, CA4*, and *GUCA2A/B*, thus confirming their identity as mature, lineage-committed Best4 Cells (Figure 19A). Approximately 90% of sorted BEST4^{+} primary cells met the full expression market signature, validating both the antibody's specificity and the sorting protocol in native epithelium. A comparative UMAP embedding clearly illustrates this contrast: primary-derived cells formed a discrete, high-expressing *BEST4* cluster, while hIEO-derived BEST4^{+} cells exhibited diffuse and transcriptionally ambiguous expression patterns (Figure 19B). Together, these findings demonstrate that the House protocol fails to support maturation into transcriptionally stable Best4 Cells, establishing that the lack of Best4 identity in prior experiments (see Section 3.2) was primarily due to a biological limitation rather than insufficient detection sensitivity.

Figure 19 – Transcriptional signature of BEST4⁺ and BEST4⁻ sorted cells across primary intestinal cells and hIEOs

A) Differentially expressed Best4 Cell genes



B) Expression of the BEST4 gene across the primary and hIEO cell datasets



Legend: (A) Heatmap of selected Best4 Cell marker genes showing consistent expression patterns across BEST4+ cells derived from primary epithelial cells, with low to null expression in matched BEST4− cells and both BEST4+ and BEST4− hIEO cells. (B) UMAP highlighting BEST4 expression across primary cells and hIEOs, showing negligible expression patterns in hIEOs.

3.4.4 Surface Marker vs. Transcriptome Fidelity: Technical Implications for Lineage Annotation

This discrepancy between transcriptomic profiles and surface cell markers in hIEO-derived cells underscores a broader challenge in sorting-based models: surface protein detection can be misleading and uncoupled from full transcriptional programming, particularly in immature or unstable cellular states. While the most plausible explanation is nonspecific antibody binding or technical artefact, alternative possibilities include altered membrane trafficking and epitope mislocalisation — phenomena commonly observed in stem cell–derived systems[335]. These findings highlight the risk of false-positive identification when relying solely on surface-marker-based enrichment, especially in the absence of orthogonal transcriptomic validation.

Crucially, the apparent failure of hIEOs cultured under the House protocol to progress beyond early absorptive intermediates underscores the epigenetic and microenvironmental barriers to full lineage commitment. Although early markers such as *SPIB* may be sporadically expressed, downstream targets like *CA7* and *BEST4* remained largely silent, suggesting that the differentiation trajectory is incompletely executed or stalled. Without permissive chromatin states or reinforcing niche signals, cells may stall in transitional states or revert, undermining stable lineage specification.

These findings emphasise the need for integrated validation approaches — combining surface marker profiling, full-length transcriptomic analysis, and careful marker scoring — particularly when evaluating rare or under-characterised epithelial subsets in *in vitro* systems. They also reinforce that high-fidelity modelling of epithelial diversity requires not just optimised culture conditions, but also robust technical and analytical frameworks to distinguish true lineage commitment from transient or misleading phenotypes.

3.4.5 Controlled Protocol Benchmarking via Multiplexed Droplet-Based scRNA-Seq

Having established that the House Protocol could not sustainably promote differentiation of Best4 Cell populations, a controlled benchmarking experiment was developed to systematically evaluate the influence of hIEO culture protocols on epithelial differentiation, with a particular focus on the emergence of rare lineages such as Best4 Cells. Matched patient-derived hIEO lines were divided into parallel culture arms using three basal protocols:

1. House Protocol: A modified version of the established Sato *et al.* (2011) formulation, serving as the reference standard (as used in previous experiments).
2. Holloway Protocol: Employing NRG1 in place of EGF.
3. Fujii Protocol: Incorporating a modified EGF stimulation approach (versus Sato *et al.*, 2011) and niche-promoting growth factors (IGF-1, FGF-2).

Although RANKL-stimulated arms were included for both the Fujii and Holloway protocols — to complement the House protocol RANKL data from the first experiment, as detailed in Section 3.2.2 — the present section focuses exclusively on the basal, unstimulated conditions. The impact of RANKL stimulation, alongside disease- and region-specific differences, will be explored in Section 5.2.3. NAB stimulation was not included here as it did not yield biologically-plausible results in the first experiment (as further detailed in Section 5.2.2.2).

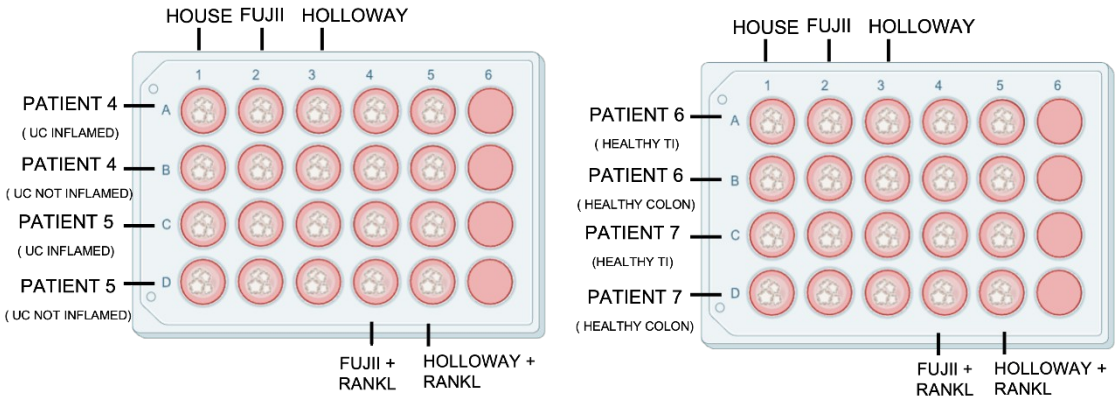
3.4.6 Data Quality and Experimental Scope

For this experiment (EXP03), organoids from eight donor samples — comprising single TI and colonic biopsies from four healthy controls, and paired inflamed and non-inflamed colonic biopsies from two UC patients — were seeded in parallel across the three culture protocols (see Supplementary Methods, Appendix Section 1.6). A total of 40 conditions were processed simultaneously using a single multiplexed 10x Genomics 3′ scRNA-seq run, employing the cell

hashing workflow validated above (Section 3.2.1) to generate 4 pooled reactions of 10 conditions each, thereby minimising technical batch variability (Figure 20).

Following stringent quality control, 12,477 high-quality cells were retained across the 40 conditions. For the comparative analysis presented in this chapter, 10,112 cells cultured under basal, unstimulated conditions (House, Fujii, Holloway) were selected. These donor-matched, uniformly processed cells provide a robust framework for side-by-side evaluation of protocol-specific differentiation potential, focusing on the emergence of rare Best4 epithelial populations.

Figure 20 - Experimental design for Experiment 3: multiplexed scRNA-seq assay across different growth protocols, anatomical locations, and disease contexts

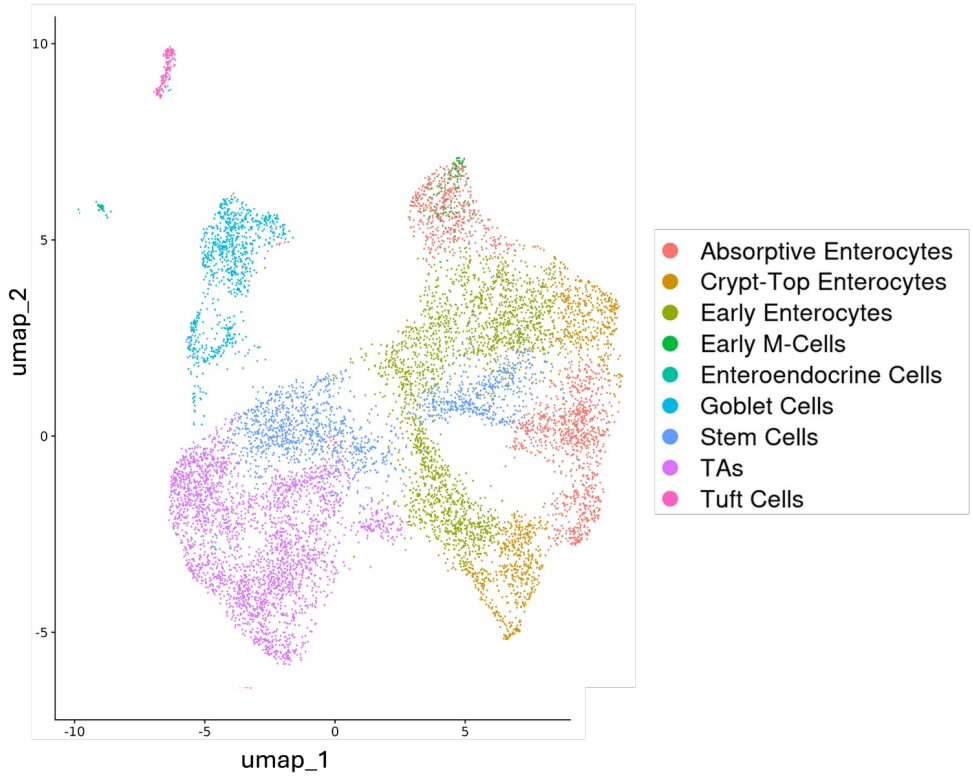


3.4.7 Protocol-Specific Performance in Supporting Rare Lineage Emergence

The Fujii protocol yielded the highest number of high-quality epithelial cells (6,843), substantially outperforming both Holloway (1,877) and House (1,752). Given that all samples were processed, pooled, and sequenced under identical conditions, this disparity likely reflects intrinsic biological differences between protocols, such as impacts on epithelial integrity, proliferation dynamics, or cell survival. The greater cellular recovery achieved with Fujii may have contributed to the improved detection of rare populations in downstream transcriptomic analyses seen earlier (Figure 7C, Figure 10, Figure 12).

UMAP dimensionality reduction and unsupervised clustering — using the consistent annotation strategy employed throughout this study — resolved major epithelial lineages across conditions, including Stem Cells, TAs, Enterocytes across all maturation stages, Goblet Cells, Tuft Cells, and EECs (Figure 21).

Figure 21 - UMAP visualisation of epithelial cell types across the three different hIEO protocols included in this dataset



Protocol-specific stratification revealed distinct differences in epithelial composition across conditions (Figure 22A), reflecting both the differentiation potential and overall recovery efficiency of each protocol. As anticipated, House was predominantly skewed toward undifferentiated (35.3%) and immature absorptive states (Early and Absorptive Enterocytes, 44.4%). While it yielded a modest fraction of Goblet Cells (7.9%), the emergence of rarer secretory terminally differentiated lineages — such as Tuft Cells and EECs — was minimal. This

outcome supports earlier observations that EGF-based formulations are limited in their ability to drive complete secretory differentiation. The resulting epithelial profile suggests that although secretory lineage pathways may be partially activated, the absence of balanced or instructive niche cues likely redirects differentiation predominantly toward the Goblet Cell lineage. This pattern may reflect a compensatory or default response under suboptimal secretory conditions, where precursor cells adopt the most accessible secretory fate in the absence of more specialised signals.

By contrast, both Fujii and Holloway supported more diverse and mature epithelial profiles. These formulations generated greater absolute representation of rare and specialised subsets, indicating improved engagement of lineage-specifying signals. Quantitative analysis confirmed these observations (Figure 22B). While TAs and Early Enterocytes remained prevalent, Fujii was distinguished by the highest frequency of Goblet Cells (9.6%), Tuft Cells (1.9%), and EECs (0.5%). Notably, Fujii also generated a small but transcriptionally distinct population of early M-Cells (expressing *SPIB*, *TNFAIP2*, *SOX8*), reflecting the protocol's broad and enhanced rare lineage specification potential.

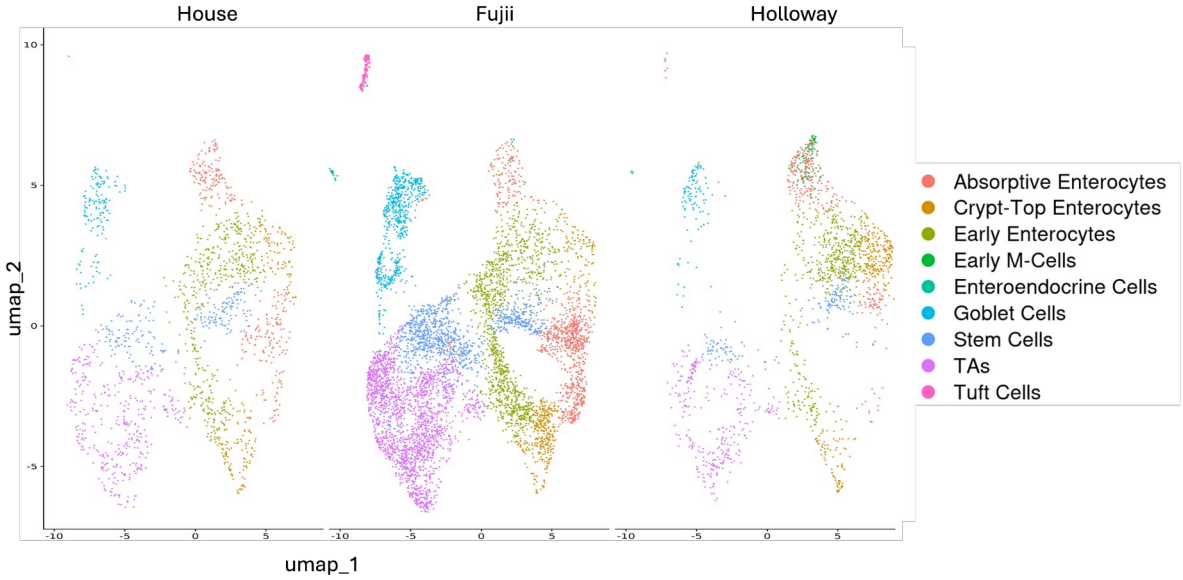
The Holloway protocol, while less efficient than Fujii at promoting secretory lineage commitment, induced robust absorptive differentiation. This was reflected in the high proportion of Enterocytes (62.9%) - including the largest fraction of Crypt-Top Enterocytes (16.8%) - alongside the highest representation of early M-Cells (5.2%). These findings align with trends observed in the earlier meta-analysis and suggest that Holloway effectively engages *SPIB*-regulated transcriptional programmes, underscoring the protocol's distinctive capacity to drive advanced absorptive maturation and specialisation. Of note, analysis of RANKL-stimulation (a cytokine previously shown to promote *SPIB* expression) across the 3 experimental conditions tested showed no discernible impact on the emergence of Best4 cells,

and was therefore not considered further. Nonetheless, this stimulation resulted in marked enrichment of M-cell populations, as further detailed in Section 5.2.3.2.

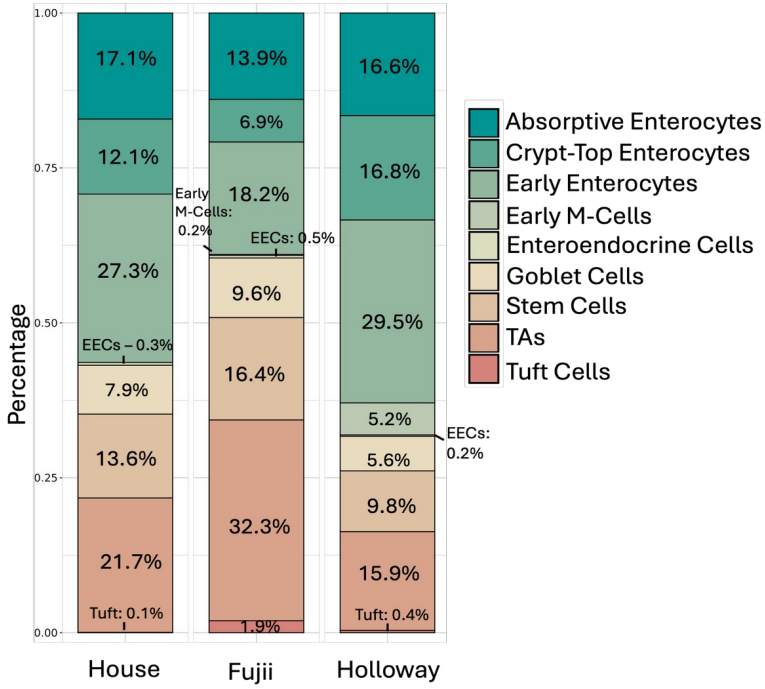
In summary, both Fujii and Holloway protocols substantially improved epithelial complexity relative to the House condition, though they promoted differentiation in distinct directions: Fujii favoured rarer secretory lineages, while Holloway supported advanced absorptive specialisation. These findings reinforce the need for goal-oriented media optimisation in hIEO systems, tailored to the specific epithelial lineages of interest.

Figure 22 - Comparison of epithelial cell type distributions across the three hIEO protocols

A) UMAP visualisation of epithelial cell types across House, Fujii, and Holloway protocols



B) Proportional abundance of epithelial lineages under each protocol



Legend: (A) UMAP projections of integrated organoid datasets cultured under the House, Fujii, or Holloway protocols, annotated by epithelial cell type. (B) Stacked bar plot showing the relative abundance of each cell type per protocol. Variability in the proportions of most cell types is evident, suggesting protocol-specific effects on lineage bias and differentiation outcomes.

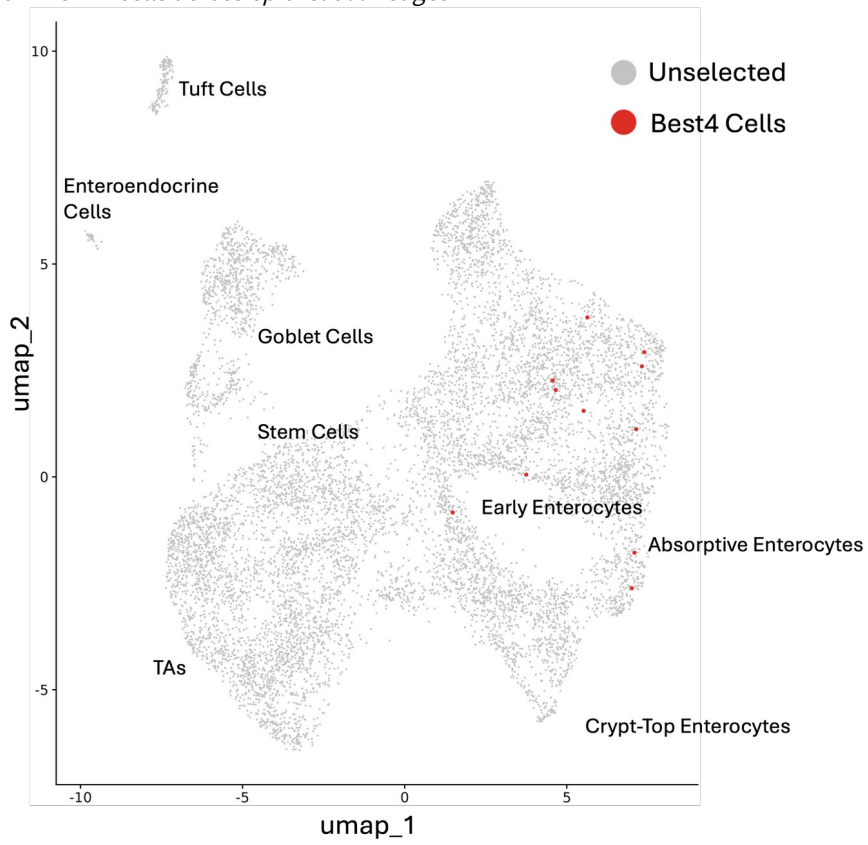
3.4.8 Quantifying *BEST4+* Resolution Across Protocols: A Technical Perspective

Contrary to earlier expectations and prior meta-analytical findings, no discrete cluster of Best4 Cells emerged within the integrated UMAP of unstimulated samples (Figure 23). To determine whether early or transitional Best4-like cells might still be present but transcriptionally diffuse, canonical marker co-expression was re-examined across all baseline conditions. Cells were screened for expression of *BEST4* and *SPIB* (Figure 23A), followed by progressive classification across three increasingly stringent thresholds (similarly to that employed earlier in Section 3.3.8): (1) expression of *BEST4* alone, (2) co-expression of *BEST4* and *SPIB*, and (3) co-expression of *BEST4*, *SPIB*, and *CA7* (Figure 23B–C). This systematic analysis revealed that while *BEST4+* cells were detectable across all protocols, they were scarce and transcriptionally heterogeneous, lacking the coherence necessary to consolidate into a dedicated cluster. The Fujii protocol yielded the highest number of *BEST4+* cells (n=31), but only two co-expressed *SPIB*, and none met the three-marker signature. By contrast, the Holloway and House protocols contributed fewer initial *BEST4+* cells overall, yet maintained similar counts across the co-expression thresholds, although still very low.

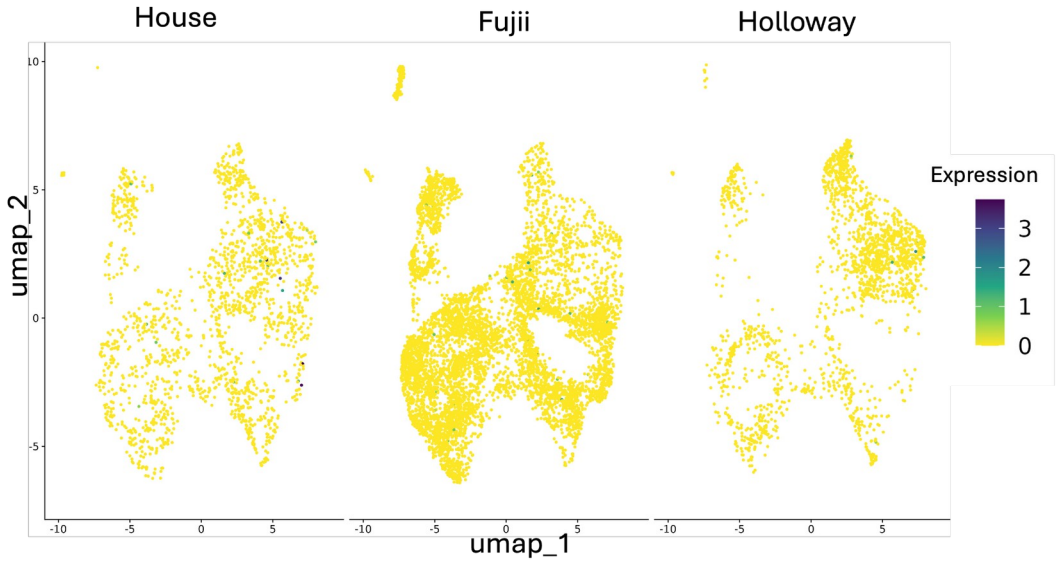
These findings reinforce the technical and biological complexity of inducing stable Best4 differentiation *in vitro*. The sparse and partial expression profiles suggest that many *BEST4+* cells may represent primed or transitional progenitors, rather than fully committed lineage identities. Moreover, the discrepancy between these findings and previous observations of Best4 enrichment in the Fujii and Holloway protocols may reflect donor-specific variability or protocol-context interactions that modulate lineage output. Overall, these results highlight the need for cautious interpretation of marker expression in underrepresented cell types, while emphasising the importance of rigorous multi-parameter validation when evaluating rare epithelial fates in organoid systems.

Figure 23 - Distribution and transcriptional fidelity of BEST4+ cells across culture conditions

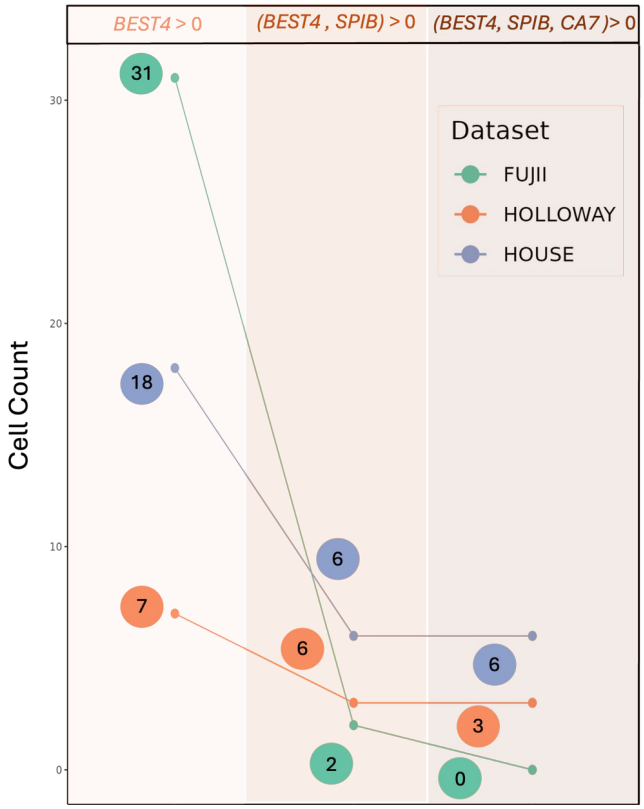
A) Distribution of BEST4+ cells across epithelial lineages



B) BEST4 expression across organoid cultures under three protocols



C) Quantification of Best4 candidate cells by transcriptional stringency across culture protocols



Legend: (A) Localisation of transcriptionally defined BEST4+ cells (red) within the epithelial landscape, compared to all other cells (grey), annotated by lineage. (B) BEST4 expression across the House, Fujii, and Holloway datasets, confirming near absence of BEST4+ cells under all conditions. (C) Number of cells retained per dataset after applying sequential transcriptional filters: BEST4 > 0, (BEST4, SPIB) > 0, and (BEST4, SPIB, CA7) > 0, illustrating the inability to generate Best4 Cells.

3.4.9 Interpretation of Findings

This section systematically benchmarked culture protocols and sequencing platforms using Best4 Cells as a representative case study for evaluating rare epithelial lineages. To validate whether the House protocol was genuinely incapable of supporting Best4 differentiation, SmartSeq2 — one of the gold-standard methods for transcriptomic profiling — was employed. Despite its high sensitivity, SmartSeq2 failed to recover a mature Best4 cluster from hIEO-derived $EPCAM^+/BEST4^+$ cells. Instead, these cells transcriptionally resembled progenitor-like states rather than committed Best4 Cells, providing strong evidence that the absence of a Best4 signature in the first House protocol experiment reflects a genuine biological deficit rather than a detection artefact. Nonetheless, SmartSeq2's ability to detect rare, transcriptionally distinct clusters in primary epithelium confirmed its utility for benchmarking lineage identity in low-abundance cell types.

Despite negative results, the technical framework employed in this study proved highly effective. The experimental workflow — consisting of initial droplet-based multiplexed scRNA-seq, followed by targeted sorting and full-length Smart-seq2 validation — provided a robust framework for assessing rare lineage detection across varying experimental conditions. This approach enabled rigorous co-expression validation, ensured high cell and transcript capture rates, and facilitated direct, cross-protocol benchmarking while minimising technical bias. The use of cell hashing for multiplexed single-cell analysis further maximised efficiency by increasing analytical throughput while simultaneously controlling for donor variability and processing artefacts. This was achieved by strategically pooling samples based on their characteristics and intended downstream comparisons.

The use of stringent marker criteria proved essential for defining true Best4 lineage commitment, with co-expression of *SPIB*, *BEST4*, *CA7*, and *MEIS1* representing a reproducible

and high-fidelity transcriptional signature for Best4 Cells across both tissue and hIEO contexts. These markers provided a rigorous filter for identifying true lineage commitment, avoiding false-positive assignment in transitional or immature cell states. However, the House protocol exhibited consistent limitations in supporting Best4 differentiation. While *SPIB*, *BEST4* or *CA7* expression occurred occasionally, a distinct Best4 population was never induced. This pattern points to incomplete lineage resolution and suggests that EGF-based formulations are insufficient to drive full commitment without additional niche-derived cues. While differentiation time could be suboptimal, extended differentiation attempts led to high apoptosis and reduced marker expression (Supplementary Figure S12), further highlighting that barriers to differentiation were the lack of specific biological cues. Similarly, discrepancies in epithelial maturity were also observed. In primary tissue, BEST4 protein expression reliably marked mature epithelial cells in primary tissue and aligned with a robust transcriptomic profile. By contrast, hIEO-derived cells in the House condition often remained immature or transcriptionally unstable, undermining surface sorting strategies and increasing susceptibility to nonspecific antibody binding.

Direct comparisons across protocols demonstrated that epithelial diversification is strongly governed by media formulation. IGF-1/FGF2 (Fujii) and NRG1 (Holloway) promoted broader epithelial differentiation, enhancing recovery of secretory and absorptive rare cell types. These effects were absent in the House protocol, which remained skewed toward undifferentiated or early absorptive states. Notably, the original Fujii dataset, generated using SmartSeq2 without $BEST4^{+}$ enrichment, successfully captured a distinct Best4 cluster, thus emphasising the relevance of media design. However, this outcome may also reflect several confounding factors, including the use of full-length scRNA-seq with tenfold greater cell numbers, enhanced transcript coverage, potential unpublished nuances in media handling, and donor-specific

biological predispositions that favour Best4 lineage emergence. Despite improvements in culture conditions, resolving a mature Best4 population remains a persistent challenge. Candidate cells were transcriptionally scattered and failed to co-express all defining markers. These findings suggest that full specification of Best4 cells in hIEOs may require epigenetic priming, enhanced transcriptomic resolution, the ability to capture subtle and temporally dynamic phenotypes, or additional extrinsic signalling cues.

3.5 Chapter Summary

This chapter established a systematic framework for benchmarking intestinal organoid culture protocols, with a particular emphasis on their capacity to generate rare epithelial lineages such as Best4^{+} cells. By combining multiplexed droplet-based scRNA-seq, integrative meta-analysis of published datasets, and targeted SmartSeq2 validation, it was possible to identify both the strengths and limitations of existing approaches. The analyses demonstrated that while major epithelial lineages are broadly represented across protocols, the emergence of rare subsets is inconsistent, highly protocol-dependent, and often transcriptionally incomplete. For example, Best4^{+} cells were detected sporadically but typically exhibited partial or diffuse signatures, underscoring the difficulty of driving full lineage commitment *in vitro*.

Cross-protocol comparisons revealed distinct biases: EGF-based formulations such as the House protocol favoured stem and early absorptive states, but failed to promote advanced maturation, while the Fujii and Holloway protocols supported more diverse epithelial outputs, including rare secretory and absorptive specialisations. Nonetheless, even under these optimised conditions, transcriptionally coherent Best4 populations remained scarce, highlighting the persistence of biological and technical barriers to resolving rare epithelial fates. Importantly, SmartSeq2 benchmarking confirmed that these deficits reflect genuine biological limitations rather than insufficient detection sensitivity, emphasising the need for approaches that capture subtle and transitional states more effectively. Together, these findings underscore that while hIEOs provide a powerful model of intestinal epithelium, conventional transcriptomic profiling and static culture systems fall short in resolving rare lineages and capturing dynamic cell fate decisions.

4 Dissecting rare epithelial lineages with multiomic and lineage tracing frameworks

4.1 Introduction

The preceding chapters highlighted both the promise and the limitations of human intestinal epithelial organoids (hIEOs) as models for epithelial differentiation and lineage specification. Multiplexed scRNA-seq experiments established the feasibility of systematically benchmarking media conditions, donor sources, and culture conformations, and confirmed that hIEOs broadly recapitulate the cellular diversity of the intestinal epithelium. However, these experiments also underscored persistent technical and biological barriers to achieving full resolution of epithelial states. Most notably, rare and transitional lineages — such as Best4$^+$ enterocytes, M-cells, and certain tuft or enteroendocrine intermediates — remained inconsistently detected. Even under optimised conditions, these subsets were either sparsely represented or transcriptionally diffuse, limiting confidence in their classification and preventing robust reconstruction of their differentiation trajectories.

Two recurring issues emerged. First, the reliance on scRNA-seq alone constrained detection to actively expressed transcripts. This presented a fundamental challenge for resolving partially committed or primed states, in which regulatory programmes may be engaged epigenetically long before measurable transcription occurs (i.e. where relevant chromatin areas may be unfolded and suitable for transcription to occur, but without active gene transcription taking place yet). For example, earlier analyses showed chromatin remodelling and signalling

responses consistent with niche priming, yet these were not reflected in transcript abundance, making rare populations difficult to distinguish from their more abundant neighbours. Second, the static nature of scRNA-seq offered only snapshots of cellular composition. While these data captured diversity at given timepoints, they could not reveal how progenitor pools diversified, how rare fates emerged, or how specific perturbations influenced lineage bifurcations over time. These gaps limited the ability to move from descriptive cataloguing of states to mechanistic models of epithelial fate decisions.

To overcome these challenges, this chapter introduces two complementary strategies: integrative single-cell multiomic profiling, which couples RNA expression with chromatin accessibility to identify primed or transitional states that precede overt transcriptional commitment, and transcriptome-compatible lineage tracing using sequential CellTag barcoding, which links clonal history to transcriptional identity and enables temporal reconstruction of differentiation trajectories.

Together, these approaches directly address the limitations of RNA-only datasets by providing both regulatory context and temporal resolution, offering a more mechanistic view of epithelial fate dynamics and setting the stage for biological validation of rare lineages in the following chapter.

4.2 Integrative Multi-Omic Profiling to Overcome Transcriptomic Resolution Limitations

4.2.1 Objective and Rationale: Enhancing Lineage Detection via Chromatin Accessibility and Transcriptomics

As evidenced above, resolving rare or transient epithelial lineages within the intestinal epithelium presents an ongoing technical challenge, particularly in organoid systems where differentiation states often remain immature or unstable. This is especially true for cell types which are low in abundance like Best4 Cells, frequently eluding detection in conventional scRNA-seq datasets. Such difficulties are compounded by the continuous nature of epithelial differentiation along the crypt–villus and crypt–surface axes, where overlapping transcriptional programs blur the distinction between closely related cell states.

While scRNA-seq has become a cornerstone in epithelial profiling, its resolution is fundamentally limited to expressed transcripts, which may not fully capture early lineage commitment or epigenetic priming events. This limitation is made particularly clear when attempting to detect rare or partially committed populations, where gene expression may be low, transient, or completely absent despite underlying regulatory activity.

To address these constraints, single-cell Multiome sequencing (scMultiome-seq) — which simultaneously captures transcriptomic (RNA) and epigenomic (chromatin accessibility) data from the same cell — was employed[257]. This dual-modality profiling approach offers several technical advantages. First, it enables the identification of epigenetically primed cell states where chromatin at lineage-defining loci is accessible, even if transcription has not yet been initiated. Second, it provides a more robust framework for resolving transitional or immature cell types by integrating orthogonal layers of regulatory information. Third, by linking

chromatin states to gene expression within the same single cell, it offers higher-confidence annotation and improved detection fidelity for rare epithelial subsets, such as Best4 Cells.

In this study, scMultiome-seq was strategically implemented to overcome prior limitations observed in RNA-only datasets. Specifically, the hypothesis was that epigenetic priming at loci from signature genes of rare cell types (such as *BEST4* and *CA7)* and transcription factors (such as *MEIS1* and *SPIB)* could precede transcriptional commitment, and that integrated RNA+ATAC profiles would enhance the classification of elusive or transitional cell states. This approach represents a technical extension of the earlier SmartSeq2 and 10x-based profiling, aimed at refining cell state resolution, reducing false negatives, and better defining the epigenetic underpinnings of lineage fidelity in hIEO cultures.

4.2.2 Protocol Optimisation: Whole-Cell scMultiome via Digitonin vs. PFA

Conventional scMultiome-seq workflows rely on nuclear isolation to achieve efficient DNA tagmentation for the chromatin accessibility assay (ATAC-seq). However, this strategy inherently sacrifices cytoplasmic RNA, yielding a truncated, nuclear-only transcriptome. As a result, such gene expression profiles do not fully capture the complexity of cellular transcription and are not directly comparable to standard whole-cell scRNA-seq data, which include both nuclear and cytoplasmic transcripts. Furthermore, nuclear isolation disrupts cell membrane integrity, limiting compatibility with surface-based multiplexing methods such as antibody-based cell hashing (HTO).

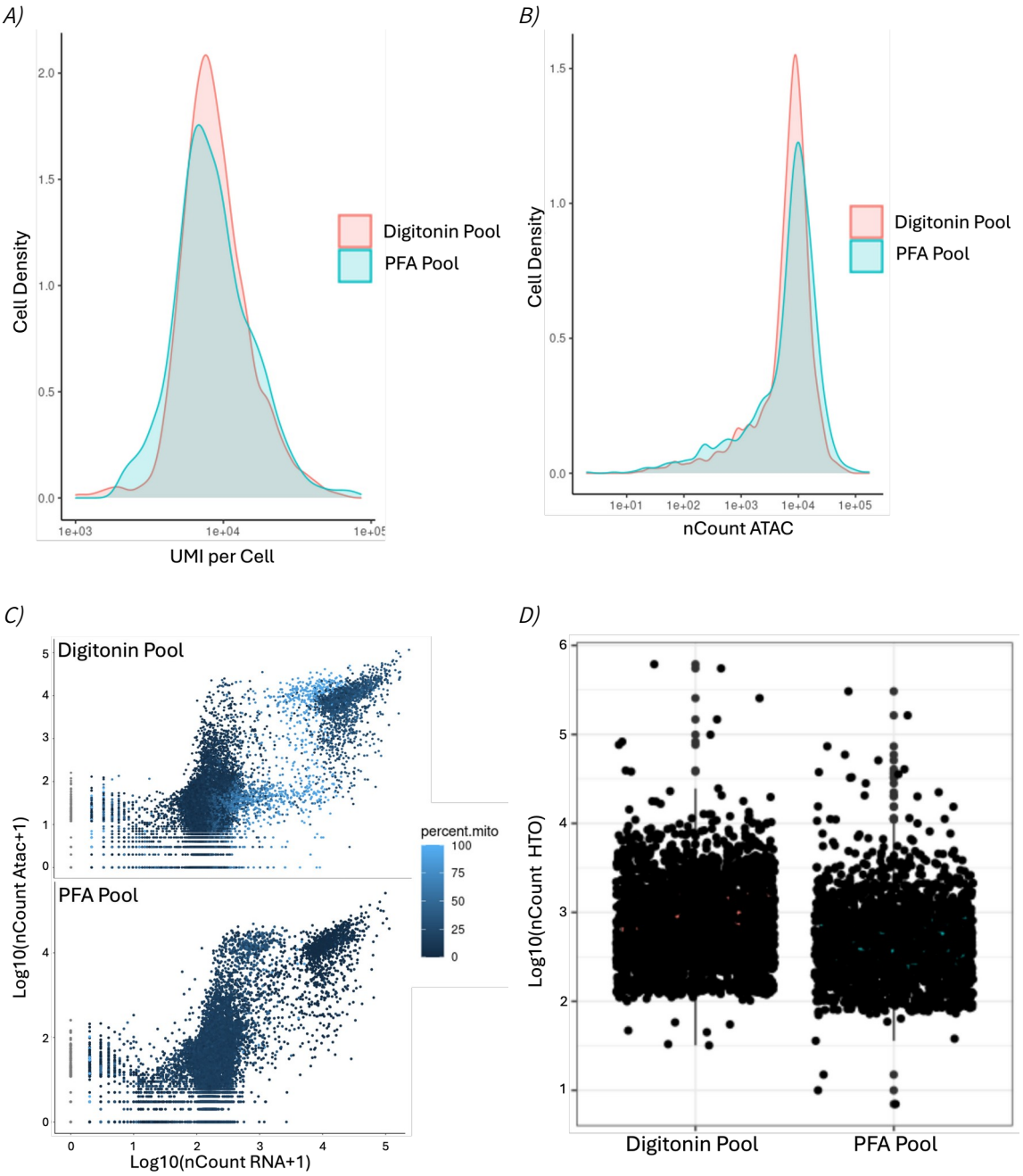
To overcome these limitations, the standard scMultiome-seq workflow from 10X Genomics was strategically modified to enable whole-cell profiling, preserving both cytoplasmic RNA and chromatin architecture (see Methods, Section 2.2.8.2). This adaptation aimed to achieve high-

quality, multimodal profiling across heterogeneous epithelial populations, while maintaining compatibility with HTO-based sample multiplexing (Supplementary Figure S13).

For this experiment (EXP04), two permeabilization strategies were evaluated, based on previous literature: paraformaldehyde (PFA)-based fixation and digitonin-based membrane permeabilisation[336]. While both protocols supported multiomic library construction, digitonin consistently outperformed PFA across all major technical metrics. Digitonin-treated cells exhibited higher transcriptomic complexity, reflected in increased mRNA capture and greater diversity of detected genes (Figure 24A). Chromatin profiling also benefited, with better-defined ATAC signal density and more coherent accessibility profiles (Figure 24B). Importantly, PFA-treated samples included a large subset of cells exhibiting strong ATAC signals but diminished RNA output — likely a consequence of cytoplasmic RNA leakage during permeabilization. By contrast, digitonin-treated samples showed a tighter correlation between ATAC and RNA signals, indicating superior preservation of cellular integrity and more balanced multimodal readouts (Figure 24C). Cell hashing performance mirrored these trends. While both strategies enabled successful multiplexing with low doublet rates, digitonin-treated cells exhibited slightly more uniform HTO signal distributions, suggesting enhanced membrane integrity during antibody labelling, and more consistent barcode incorporation (Figure 24D, Supplementary Figure S14).

Taken together, these results demonstrate that digitonin-based permeabilization provides a reproducible technical advantage in scMultiome-seq workflows. It enables deeper transcriptomic coverage, more accurate chromatin accessibility profiles, and improved compatibility with HTO multiplexing, supporting its use in all subsequent experiments.

Figure 24 - Comparison of RNA, ATAC, and HTO signal quality between Digitonin and PFA fixation protocols in the scMultiome-seq pilot experiment on whole hIEO cells



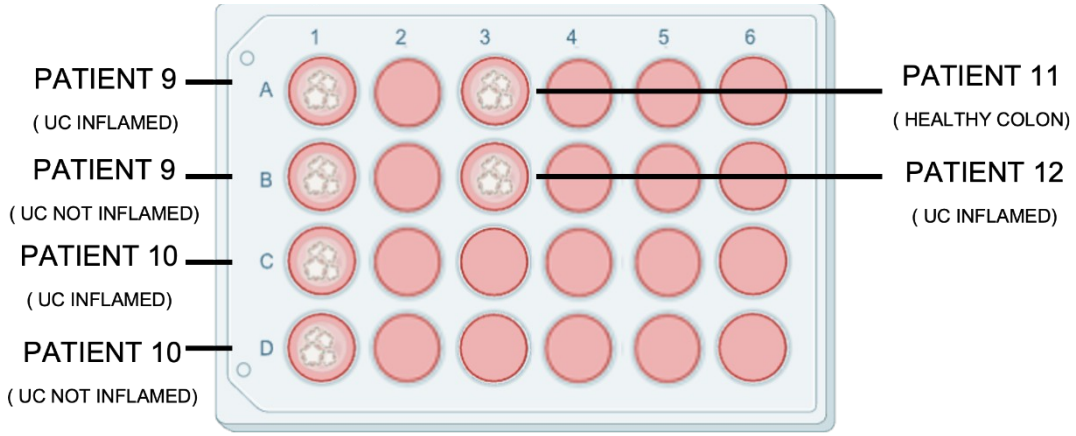
Legend: (A) Distribution of RNA UMI counts per cell in Digitonin- and PFA-fixed pools. Both methods yielded comparable recovery, with a slight advantage observed for the Digitonin protocol. (B) Distribution of ATAC fragment counts per cell across the two fixation methods, showing a slight enrichment in Digitonin-treated samples. (C) Comparison of RNA vs ATAC counts per cell, coloured by mitochondrial read percentage, for each fixation method. Digitonin pools show slightly higher joint signal with lower mitochondrial content. (D) HTO barcode count distributions are comparable between fixation methods, indicating preserved hashing efficiency across conditions and a slightly better performance for the Digitonin protocol.

4.2.3 Proof-of-Concept: Multiplexed scMultiome-seq Across Diverse hIEO Samples

To evaluate the utility of integrated RNA and chromatin accessibility profiling for resolving rare intestinal epithelial lineages, a multiplexed scMultiome-seq experiment was conducted across a biologically heterogeneous panel of hIEOs. For this experiment (EXP05), eight hIEO lines were derived from four colonic biopsies, including paired inflamed and non-inflamed regions from two UC donors, a single inflamed sample from a third UC donor, and one non-inflamed control sample from a healthy individual (see Figure 25 and Supplementary Methods, Section 1.6). This design intentionally captured a spectrum of epithelial states — ranging from homeostatic to chronically inflamed — enabling an assessment of how disease state and donor variability influence differentiation capacity and the emergence of rare epithelial subsets.

All lines were cultured under the Fujii differentiation protocol (Figure 25), chosen due to the superiority demonstrated earlier in supporting epithelial diversity and high cell yield (Figure 22). Although prior analyses (Section 3.4.5) did not reveal distinct Best4 cluster formation under this protocol, it remained the most suitable choice due to its robust representation of the broader epithelial landscape. This was particularly important given the technical demands of the scMultiome-seq assay, which benefits from enhanced lineage complexity and organoid viability — critical factors in mitigating the extended handling time and cellular stress associated with multimodal single-cell workflows.

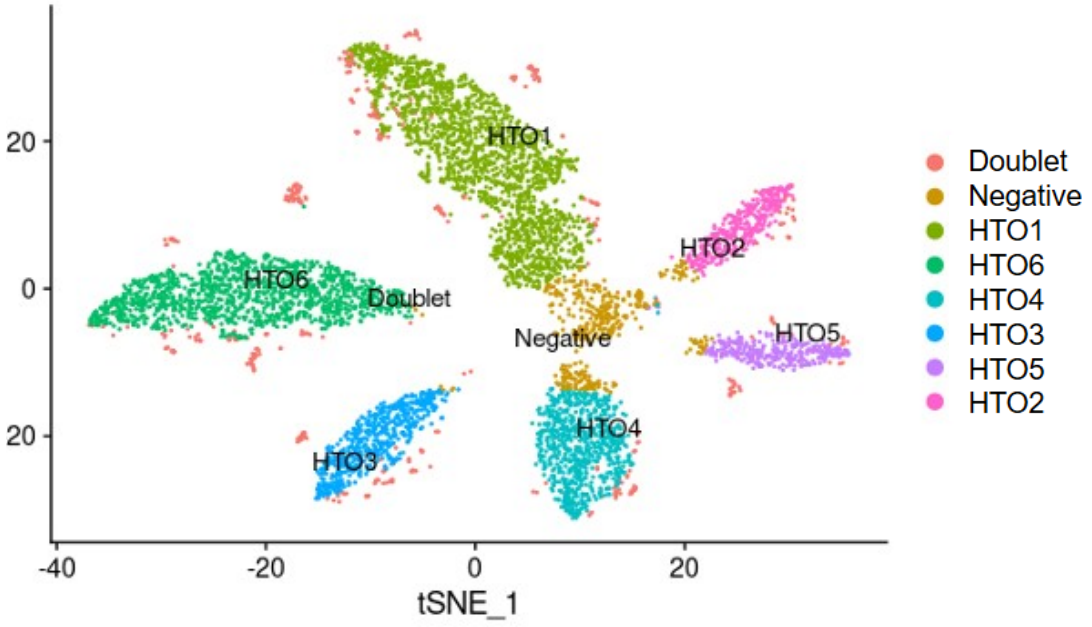
Figure 25 – Experimental design for Experiment 5: scMultiome-seq profiling of multiple hIEO conditions using the Fujii protocol



Legend: Experimental design for profiling intestinal organoids derived from ulcerative colitis (UC) inflamed (INF), non-inflamed (N.INF), and healthy control (HC) biopsies.

Following the specified differentiation period, all samples were processed using the digitonin-optimised scMultiome-seq protocol, enabling simultaneous profiling of chromatin accessibility and transcriptomic signatures from whole cells. Oligo-conjugated antibody labelling facilitated sample multiplexing within a single 10x Genomics reaction. RNA and HTO libraries were pooled and separated from ATAC libraries, sequencing each pool in an independent flow cell using the Illumina NextSeq platform. This approach produced 3,985 high-quality, paired RNA+ATAC profiles, with successful deconvolution of HTO signals confirming accurate sample demultiplexing (Figure 26, Supplementary Figure S15). This proof-of-concept establishes the feasibility of applying whole-cell Multiome-seq to multiplexed hIEO cultures, generating integrated molecular profiles suitable for evaluating both technical performance and biological lineage resolution across heterogeneous donor conditions.

Figure 26 - Demultiplexing of scMultiome-seq samples using antibody-derived hashtags (HTOs)



Legend: tSNE visualisation of the same dataset after demultiplexing. Cells are coloured by HTO classification, identifying distinct clusters corresponding to each individual sample (HTO1–HTO6), with separate annotations for doublets (red) and unclassified/negative cells (grey). This demonstrates efficient and specific sample multiplexing, enabling robust identification of pooled samples in downstream single-cell workflows.

4.2.4 Benchmarking Modality Performance: RNA vs. ATAC vs. Integrated UMAP Clustering

To evaluate the technical performance of each molecular modality in resolving epithelial differentiation states, unsupervised clustering was performed separately on the RNA, ATAC, and integrated RNA+ATAC components of the multiomic data. The resulting UMAP embeddings are shown in Figure 27. The RNA-based UMAP (Figure 27A) exhibited well-resolved major epithelial lineages, including Stem Cells, TAs, secretory subsets (Goblet Cells, Tuft Cells, and EECs), and absorptive lineages (Early-, Absorptive-, and Crypt-Top Enterocytes), confirming high transcriptomic definition and robust cluster segmentation.

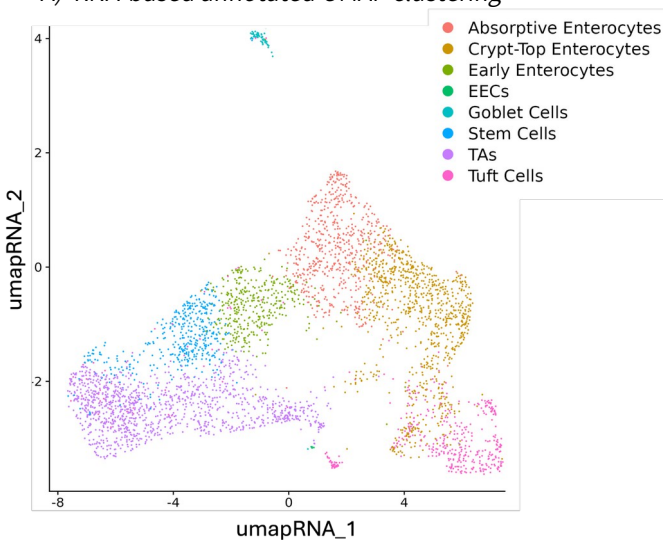
By contrast, the ATAC-based UMAP (Figure 27B) showed markedly lower resolution, with poor separation between both intermediate progenitor populations and rare lineages. Secretory subsets — especially Goblet and Tuft Cells — were weakly defined or collapsed into

neighbouring clusters. This likely reflects the sparse and noisy nature of single-cell ATAC-seq data, particularly at the sequencing depth achieved (below 20% saturation). Unlike scRNA-seq, chromatin accessibility requires substantially greater coverage to resolve genome-wide regulatory signals and delineate fine-grained lineage distinctions[337].

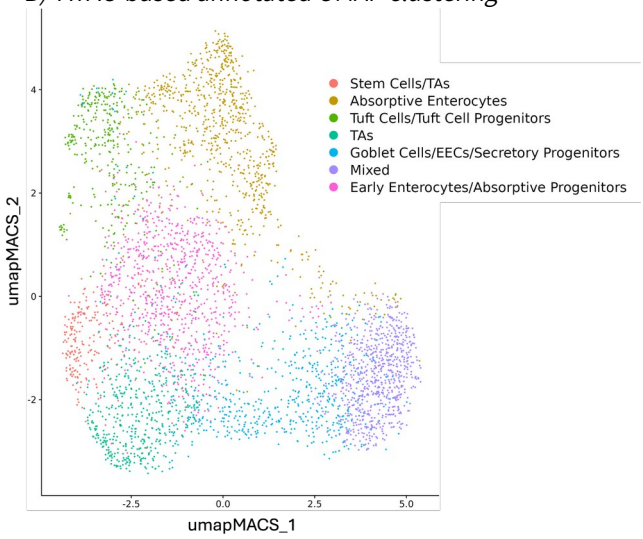
Nevertheless, the integrated RNA+ATAC UMAP (Figure 27C), leveraging the complementary strengths of both modalities, restored clear stratification of key lineages. Notably, rare populations such as Tuft Cells, Tuft Cell Progenitors, and EECs were resolved with improved fidelity, while maintaining the expected transcriptional hierarchies along the absorptive trajectory — from Stem Cells and TAs, through Early Enterocytes, to Crypt-Top Enterocytes. These findings highlight the technical advantages of multimodal integration in refining lineage classification, especially when characterising rare or transcriptionally intermediate epithelial states.

Figure 27 - UMAP visualisations of RNA, chromatin accessibility (ATAC), and integrated modalities from scMultiome-seq intestinal epithelial dataset

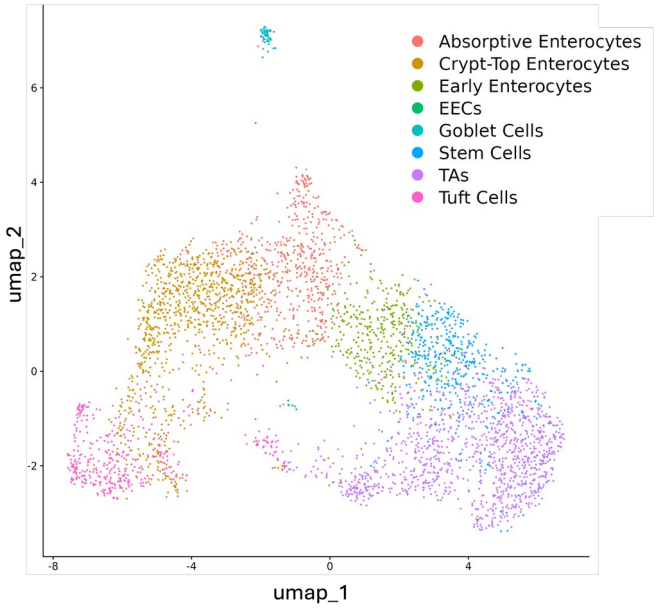
A) RNA-based annotated UMAP clustering



B) ATAC-based annotated UMAP clustering



C) Integrated RNA and ATAC UMAP with harmonised annotation



Legend: (A) UMAP projection of gene expression (scRNA-seq) from scMultiome-seq data coloured by epithelial cell type annotation. (B) UMAP projection of chromatin accessibility (scATAC-seq) from the same dataset, showing very poor lineage separation. (C) Joint UMAP visualisation after RNA-ATAC integration confirmed consistent resolution of key epithelial lineages.

4.2.5 Evaluating Chromatin Cluster Fidelity Through Accessibility Metrics

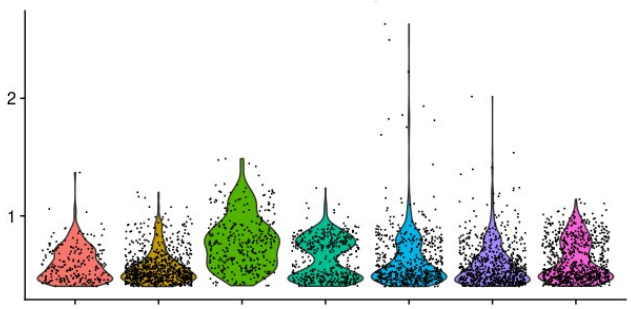
Analysis of overall chromatin accessibility metrics revealed critical insights into the poor resolution of ATAC-derived subpopulations (Figure 28). Clusters such as Goblet

Cells/EECs/Secretory Progenitors, and Early Enterocytes/Absorptive Progenitors consistently exhibited lower accessibility, as evidenced by reduced peak counts and elevated nucleosome signals (Figure 28A–B). These patterns likely reflect a combination of biological and technical factors. Biologically, quiescence and partial differentiation are often associated with global epigenetic repression and reduced chromatin accessibility, as observed in various stem and progenitor cell systems[338]. Technically, the inherently sparse nature of single-cell ATAC-seq presents limitations in resolving low-abundance or transitional populations, where signal dropout and under-sampling can obscure subtle regulatory features.

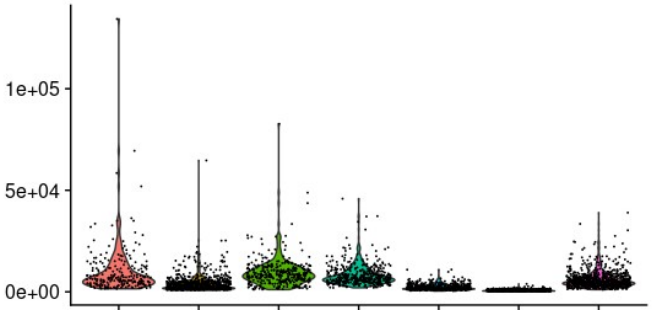
The “Mixed” cluster was particularly inconsistent, displaying conflicting trends between transcription start site (TSS) enrichment and nucleosome occupancy patterns — hallmarks of unresolved identity and epigenetic heterogeneity (Figure 28A-C). Even clusters with stronger RNA-based resolution, such as Stem Cells/TAs, TAs, and Tuft Cells/Tuft Cell Progenitors, displayed moderate epigenetic variability. Notably, TAs showed an accentuated bimodal nucleosome distribution, pointing to coexisting subpopulations with distinct chromatin states. Similarly, the Tuft-like cluster showed mild nucleosome elevation, suggesting subtle intra-cluster differences in chromatin compaction and regulatory activity.

Overall, these challenges underscore the limitations of relying on chromatin accessibility under conditions of low sequencing depth per cell, suggesting that significantly higher sequencing coverage or targeted epigenomic assays may be necessary to resolve fine-grained regulatory features with higher fidelity.

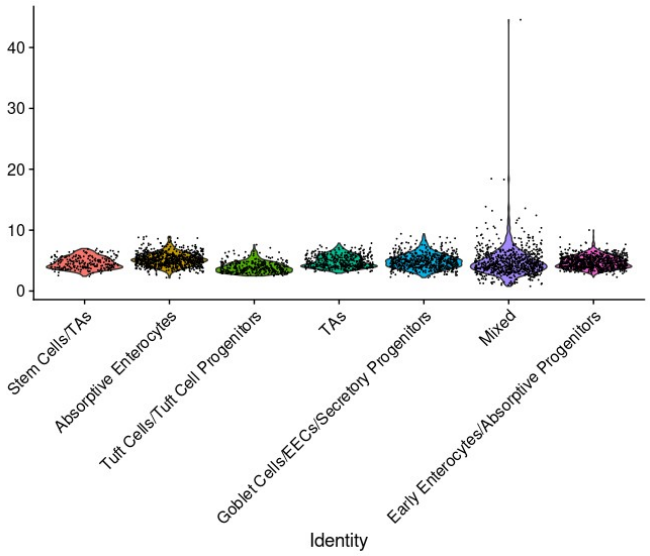
Figure 28 - ATAC-seq quality control metrics across epithelial cell clusters.
A) Nucleosome signal distribution per cluster



B) ATAC peak counts per cell across clusters



C) *Transcription start site (TSS) enrichment per cluster*



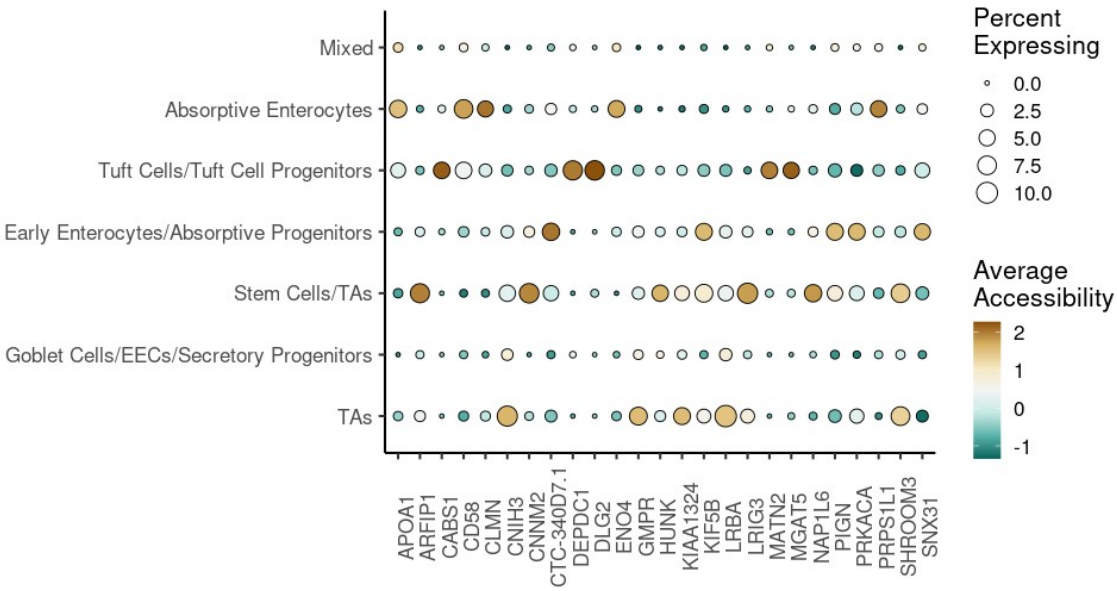
Legend: Labels for the common horizontal axis are shown in panel C. (A) Nucleosome signal per cluster; higher values suggest increased chromatin compaction or reduced accessibility. (B) Peak counts per cell across annotated epithelial clusters, reflecting the number of accessible chromatin regions detected per cell; measure of data richness and signal quality. (C) TSS enrichment scores per cluster, quantifying the enrichment of transposase accessibility at transcription start sites; indicator of chromatin integrity and transcriptional activity.

4.2.6 Chromatin Resolution Limits: Interpreting Sparse Accessibility in Transitional and Rare Epithelial States

Identification of cluster-level accessible loci supported the notion of poorly resolved chromatin profiles (Figure 29). Nevertheless, while chromatin accessibility data presented limitations in resolution and interpretability, certain biologically meaningful patterns could still be discerned. Stem and progenitor-rich populations exhibited broad accessibility at loci such as *ARFIP1*, *CNNM2*, *DEPDC1*, and *KIF5B*, consistent with their epigenetic plasticity and developmental potential. TAs also showed focused accessibility near *GMPR*, *CNIH3*, and *KIAA1324*, reinforcing their proliferative activity. By contrast, Absorptive Enterocytes displayed sharply defined accessibility at lineage-specific genes *(APOA1, CD58, CLMN)*, indicative of epigenetic commitment. As anticipated, Goblet Cells/EECs/Secretory Progenitors exhibited notably low chromatin accessibility across almost all tested loci.

These observations reinforce the need for high sequencing coverage in ATAC-seq assays, especially when resolving rare or transitional epithelial populations. In these contexts, chromatin-based clustering alone lacks the resolution to confidently delineate cell types. Integrated RNA+ATAC clustering mitigates this issue by leveraging complementary transcriptional data to refine lineage identities. However, achieving high-resolution epigenetic mapping — particularly for low-abundance or plastic cell states — requires highly sensitive profiling strategies, specifically designed to capture subtle, time-constrained variations across a vast and diverse set of genomic loci.

Figure 29 – Differentially accessible genomic loci identified across chromatin-defined clusters



Legend: Percentage of cells with detectable chromatin accessibility (dot size) and the average accessibility score (colour scale) across a panel of differentially accessible genomic loci. Notably, absorptive lineages show enriched accessibility at loci such as APOA1, CD58, and CLMN, while stem-like clusters exhibit broader but less lineage-specific accessibility patterns. These patterns suggest differential epigenetic priming across epithelial subsets.

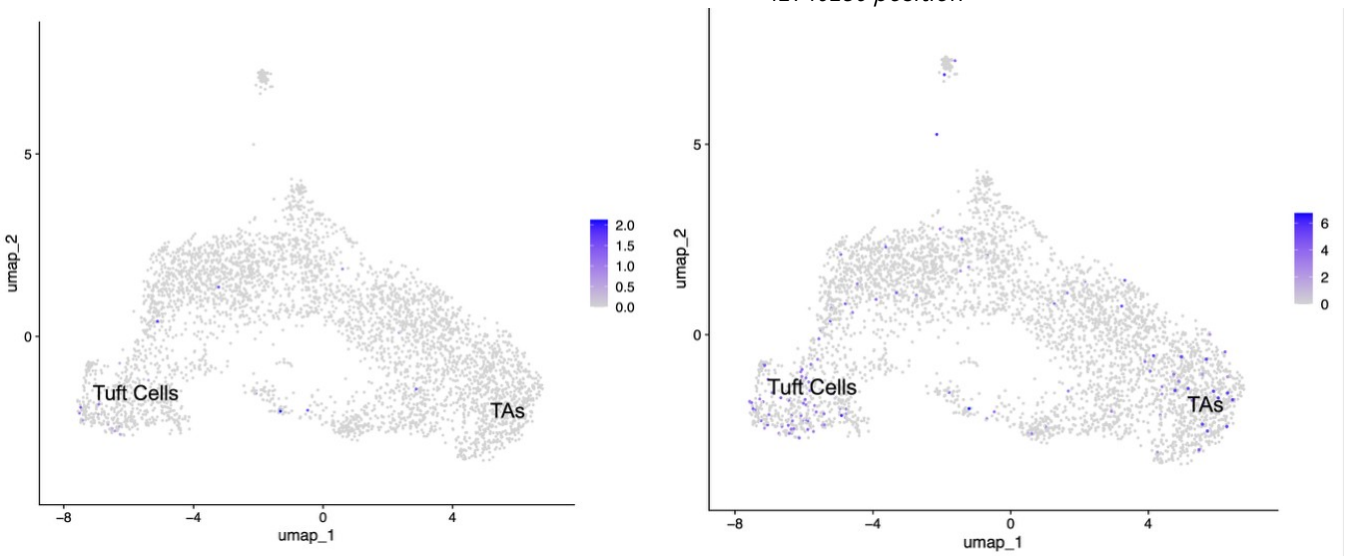
Of particular interest, despite showing overall poorer performance when compared with scRNA-seq, chromatin accessibility data still provided some complementary insights. Notably, some genes appeared to be epigenetically primed for expression in specific cell types before corresponding transcripts became detectable. One such example was observed at the *HAP1* locus. The *HAP1* (Huntingtin-associated protein 1) gene plays a key role in intracellular trafficking and vesicle-mediated transport of proteins[339]. It has been specifically implicated in the internalisation and recycling of epidermal growth factor receptors (EGFR), suggesting a potential role in modulating signalling dynamics in epithelial cells[339,340]. HAP1 transcript levels were low or undetectable across most cells in the RNA modality (Figure 30A), yet chromatin accessibility at the corresponding locus (Figure 30B) revealed a distinct enrichment in both Tuft Cells and TAs. This was further supported by the aggregated chromatin accessibility signal (Figure 30C), where both cell types showed strong accessibility peaks upstream of *HAP1*. This

suggests that *HAP1* is epigenetically primed in Tuft Cells, with chromatin already accessible before transcription is fully initiated. Additionally, the enrichment observed in a subset of TAs may reflect early stages of lineage commitment towards a differentiated cell state. Together, these results highlight the added value of multiomic profiling in capturing early, lineage-specific regulatory states that precede detectable gene expression.

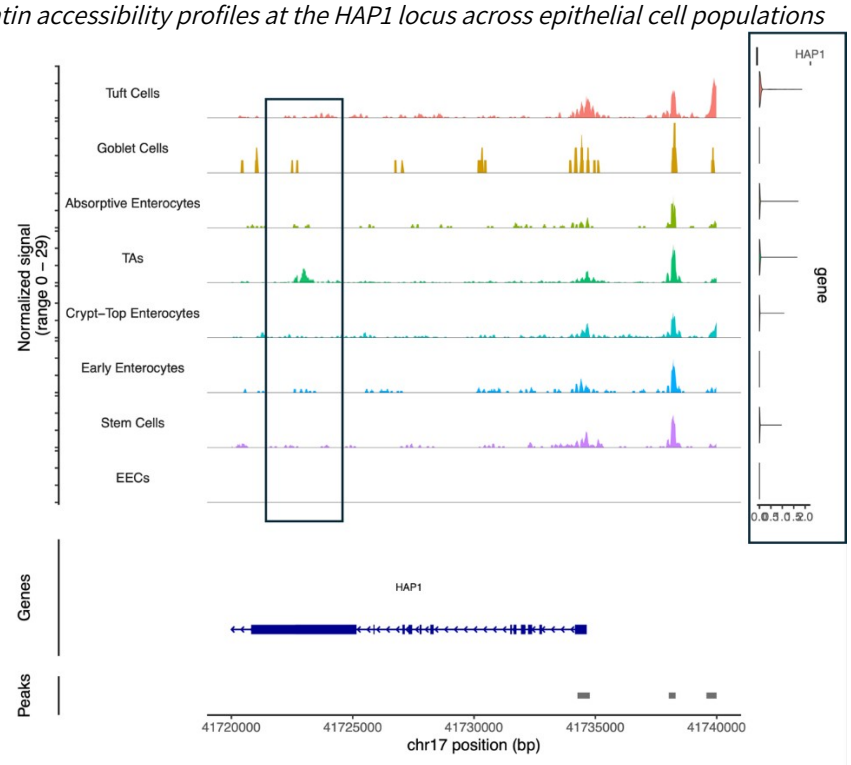
Figure 30 - Chromatin priming at the HAP1 locus in Tuft Cells and TAs

A) Gene expression levels of HAP1

B) Acessibility signal at the HAP1 locus - chromossome 17-41739574-41740180 position



C) Normalised chromatin accessibility profiles at the HAP1 locus across epithelial cell populations



Legend: (A) RNA expression of HAP1 across cell types shows low or undetectable transcript levels in most cells. (B) Chromatin accessibility signal at the HAP1 locus reveals clear enrichment in Tuft Cells and TAs, suggesting regulatory priming. (C) Aggregated chromatin accessibility signal demonstrates strong accessibility peaks upstream of HAP1 specifically in Tuft Cells and TAs. These findings indicate that HAP1 is epigenetically primed prior to transcriptional activation.

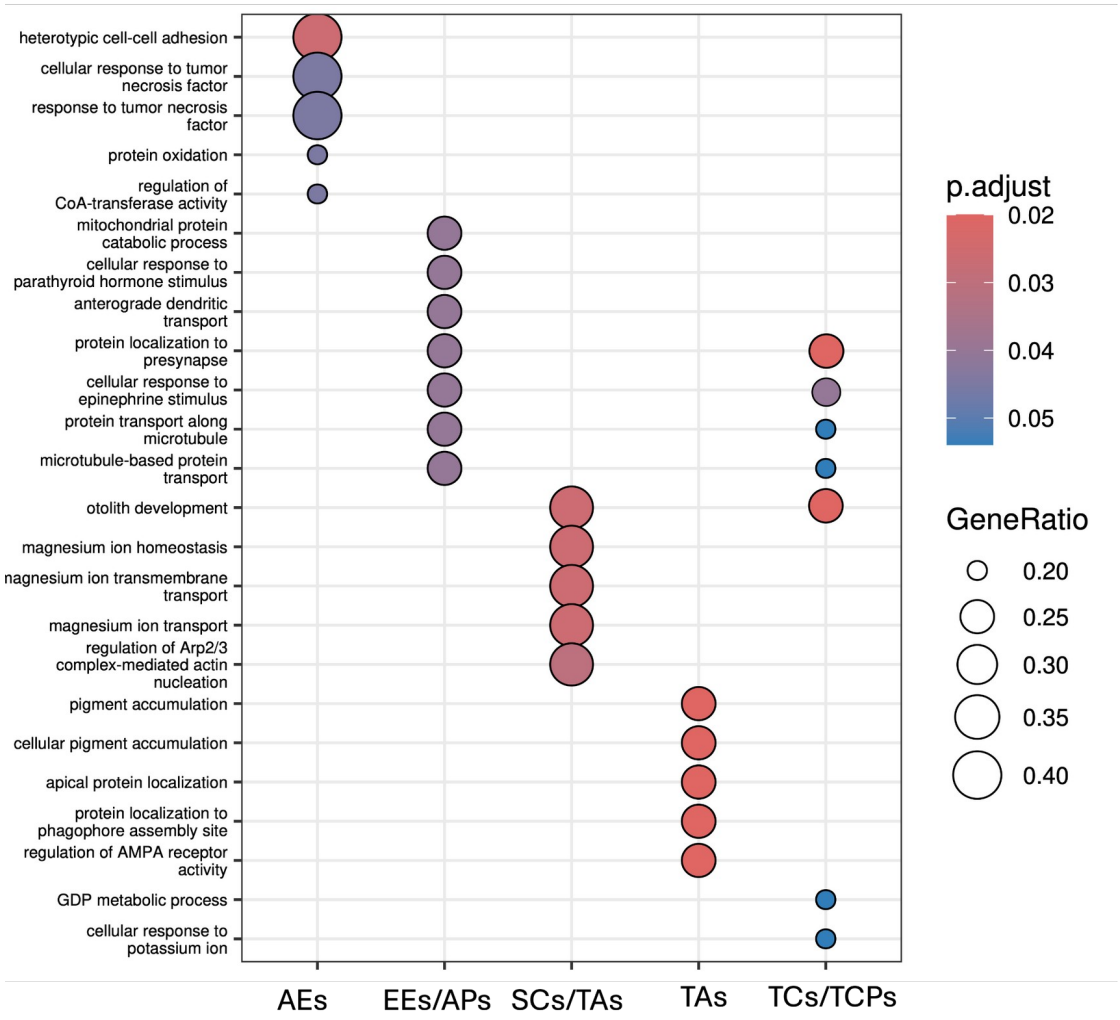
4.2.7 Functional Inference from Chromatin Accessibility: GO Enrichment Across Lineages

To functionally validate chromatin-based annotations, Gene Ontology (GO) enrichment analysis was performed on accessibility peaks within five major epithelial clusters: Absorptive Enterocytes, Early Enterocytes/Absorptive Progenitors, Stem Cells/TAs, TAs, and Tuft Cells/Tuft Cell Progenitors (Figure 31). This analysis provided biological context to the chromatin accessibility profiles and reinforced the lineage assignments derived from integrated RNA+ATAC clustering.

As expected, clusters with low accessibility and reduced representation — namely “Goblet Cells/EECs/Secretory Progenitors”, and the “Mixed” clusters — did not yield statistically significant enrichment, likely due to limited peak counts and insufficient cell numbers, underscoring the impact of sparse data on functional interpretation. Among the best resolved clusters, “Absorptive Enterocytes” showed strong enrichment for pathways related to cell-cell adhesion, TNF response, and membrane protein localisation, consistent with their role in epithelial barrier maintenance and immune interface. “Early Enterocytes/Absorptive Progenitors” were enriched for mitochondrial protein localisation, CoA-transferase activity, and microtubule-based transport, reflecting early metabolic and structural remodelling during differentiation. “Stem Cells/TAs” exhibited GO terms associated with magnesium ion transport, pigmentation, and actin filament nucleation — features indicative of proliferative plasticity and cytoskeletal reorganization during early lineage specification. “TAs” further showed enrichment for protein localisation and apical trafficking, suggesting emerging absorptive polarity and function. Finally, “Tuft Cells/Tuft Cell Progenitors” displayed specific enrichment for “cellular response to epinephrine stimulus” and “adrenergic stimulus,” capturing known chemosensory and neuroimmune roles of this rare epithelial lineage. Taken together, these results highlight the capacity of chromatin accessibility-based GO analysis to provide

functional insights into cell identity and reinforce the accuracy of cluster annotations. Supplementary Figures S16-S17 present extended enrichment analyses of accessibility gains and losses across clusters, further supporting the robustness of the multiomic clustering framework.

Figure 31 – GO analysis of differentially accessible genes in chromatin-based clusters.

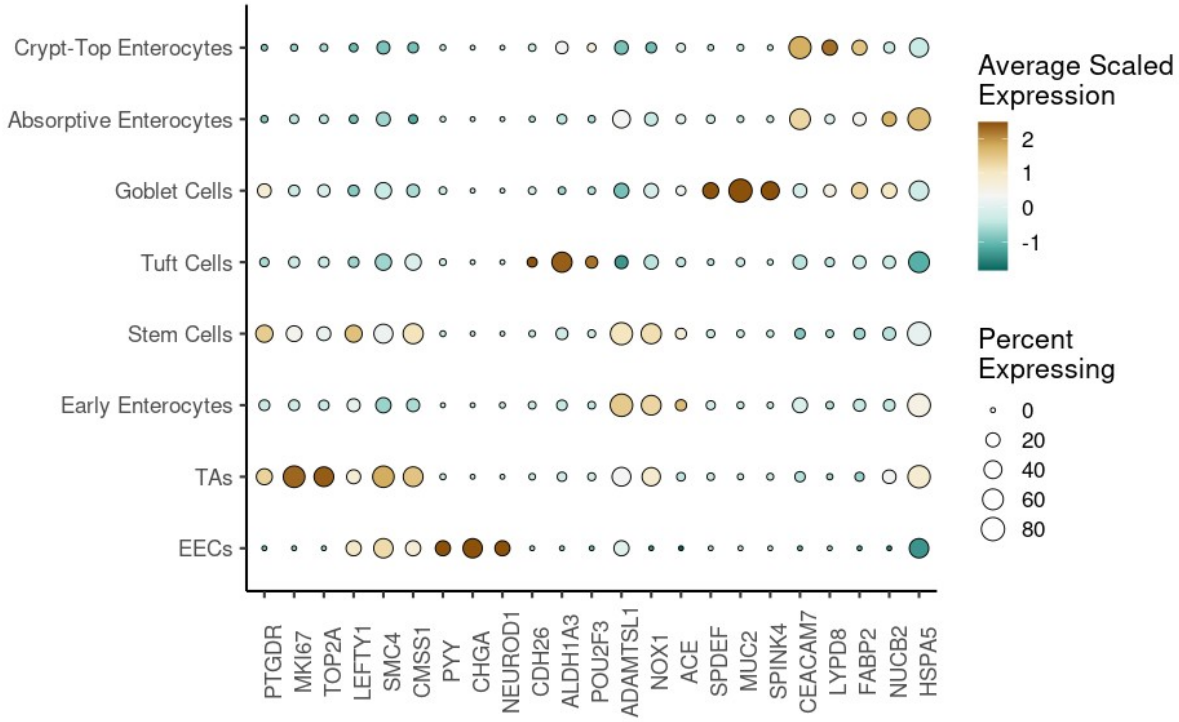


Legend: Enriched biological processes across epithelial cell types defined by chromatin accessibility. Each dot represents a GO term enriched in the respective cell type: absorptive enterocytes (AEs), early enterocytes/absorptive progenitors (EEs/APs), stem cells/transit-amplifying cells (SCs/TAs), transit-amplifying cells (TAs), and tuft cells/tuft cell progenitors (TCs/TCPs). Dot size indicates the gene ratio (proportion of genes in the GO category among input genes), and colour indicates adjusted p-value.

4.2.8 Validating Lineage Identity via Joint Transcriptomic and Epigenetic Signatures

To assess the fidelity of joint RNA+ATAC clustering, gene expression profiles were evaluated across all annotated epithelial cell types (Figure 32). Transcriptomic analysis confirmed robust lineage segregation, with canonical markers displaying expected cell-type specificity. Proliferative compartments — including TAs and Stem Cells — showed strong expression of *MKI67*, validating their proliferative identity. *FABP2*, a hallmark of Absorptive Enterocytes, was selectively enriched in absorptive clusters, confirming successful differentiation along this axis. Secretory populations were similarly well defined: *CHGA* marked EECs, *SPDEF* and *MUC2* were enriched in Goblet Cells, and *ALDH1A3* and *POU2F3* underpinned the Tuft Cells lineage. EECs also displayed strong expression of neuroendocrine markers such as *PYY* and *NEUROG1*, further validating their identity. Together, these transcriptional signatures corroborate the accuracy of integrated cluster assignments and demonstrate the capacity of the joint profiling approach to resolve diverse epithelial fates with high fidelity.

Figure 32 - Expression of selected population-specific signature genes across intestinal epithelial cell types.



Legend: (A) Average scaled expression (colour) and percentage of expressing cells (dot size) for selected genes across annotated epithelial cell types including Crypt-Top Enterocytes, Absorptive Enterocytes, Goblet Cells, Tuft Cells, Stem Cells, Early Enterocytes, TAs, and Enteroendocrine Cells (EECs).

4.2.9 Validating Lineage Resolution Through Pseudotime Mapping of Integrated Multiomic Clusters

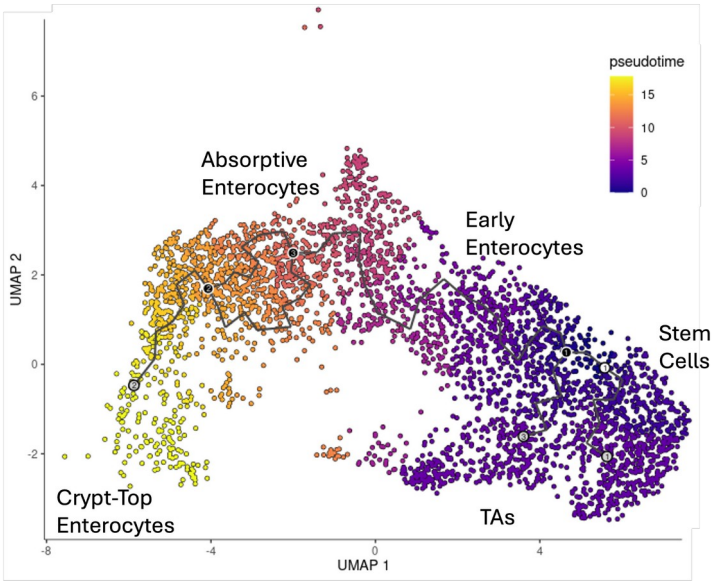
To evaluate the biological coherence and maturation gradient plausibility of the integrated RNA+ATAC clusters, pseudotime analysis was applied to the joint embedding. Pseudotime inference offers a powerful framework to model continuous cellular differentiation processes, a particularly valuable approach in epithelial systems where lineage specification occurs along gradual trajectories rather than discrete steps. Integrating RNA and ATAC signals enhances the robustness of these trajectories, as transcriptional and epigenetic markers together capture both gene expression states and upstream regulatory potential. Finally, given the limitations in chromatin-only clustering (as shown in Sections 4.2.5-4.2.6), pseudotime projections provided an orthogonal strategy to assess whether the integrated clusters recapitulate expected developmental hierarchies and lineage bifurcations. This analysis therefore served both as a validation tool for cluster fidelity based on multiomic data and to dissect differentiation pathways in hIEO cultures.

A comprehensive pseudotime analysis pipeline was implemented, combining droplet-based multimodal data with lineage-focused reconstruction (see Methods section 2.2.9.9). Integrated multiomic data were used to define absorptive and secretory lineages, and stem-cell-derived root populations were identified in UMAP space to initialise pseudotime ordering. Pseudotime was inferred independently for each lineage, and gene expression dynamics were tracked to evaluate trajectory fidelity. These lineage scores were mapped back to the full joint embedding, enabling fine-grained characterisation of lineage commitment. Finally, the use of integrated chromatin and transcriptomic information enhanced cell state resolution and helped reduce technical noise in trajectory inference.

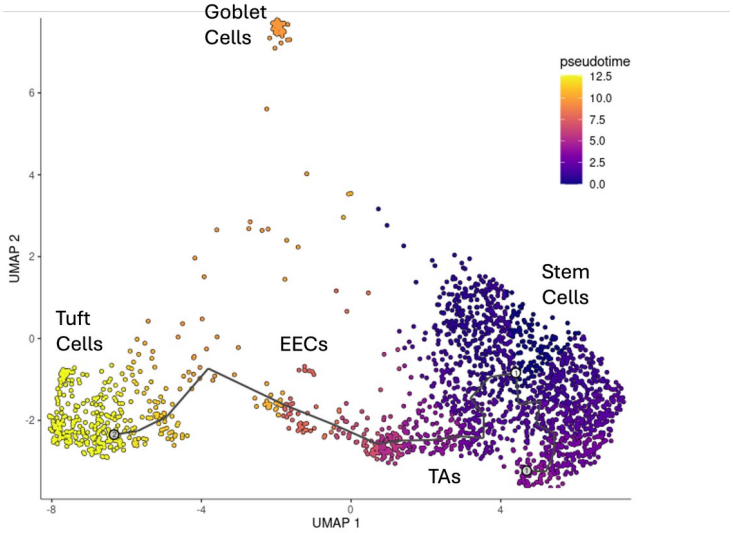
Pseudotime analysis of the integrated multiomic dataset confirmed the biological plausibility of joint RNA+ATAC cluster annotations, faithfully recapitulating known epithelial lineage hierarchies within hIEO-derived cultures (Figure 33; Supplementary Figure S18). The absorptive pathway revealed a clear continuum maturation process from Stem Cells and TAs, through Early Enterocytes and Absorptive Enterocytes, and culminating in Crypt-Top subtypes (Figure 33A; Supplementary Figure S18 A,C). This trajectory aligns with established absorptive differentiation pathways and reflects progressive transcriptional and chromatin-based maturation. By contrast, the secretory trajectory exhibited a more branched structure (Figure 33B; Supplementary Figure S18 B,D). From a shared progenitor region, trajectories diverged toward distinct secretory fates — Goblet Cells, EECs, and Tuft Cells. However, these branches exhibited comparatively diffuse resolution. This ambiguity may reflect overlapping or plastic intermediate states, characteristic of secretory lineage bifurcation in the intestinal epithelium. Alternatively, it may result from biological immaturity, suboptimal differentiation cues, or limited sequencing saturation — underscoring the need for refined protocol optimisation and deeper profiling to enhance fate separation and resolve lineage commitment in rare secretory subsets.

Figure 33 – Pseudotime trajectories of absorptive and secretory epithelial lineages

A) Absorptive Pathway



B) Secretory Pathway



Legend: (A) UMAP coloured by pseudotime values along the absorptive lineage, showing progression from Stem Cells, through Transit-Amplifying (TA) cells, to Early and fully differentiated Absorptive Enterocytes. (B) UMAP coloured by pseudotime along the secretory lineage, mapping differentiation from Stem and TA cells to enteroendocrine cells (EECs), Tuft Cells, and Goblet Cells. Pseudotime was inferred using lineage-specific trajectories.

4.2.10 Multiomic Insights into the Epigenetic and Transcriptomic Landscape of Best4 Cells

Despite the enhanced resolution offered by scMultiome-seq profiling, the reliable identification of *BEST4+* cells remained elusive. Detailed transcriptomic analysis revealed only eight cells

with detectable *BEST4* expression — of which only one co-expressed *SPIB*, and none co-expressed *BEST4, SPIB,* and *CA7* (Figure 34A). UMAP projections confirmed that expression of canonical Best4-associated genes was sparse and diffusely distributed, lacking the discrete clustering characteristic of mature, transcriptionally resolved cell types.

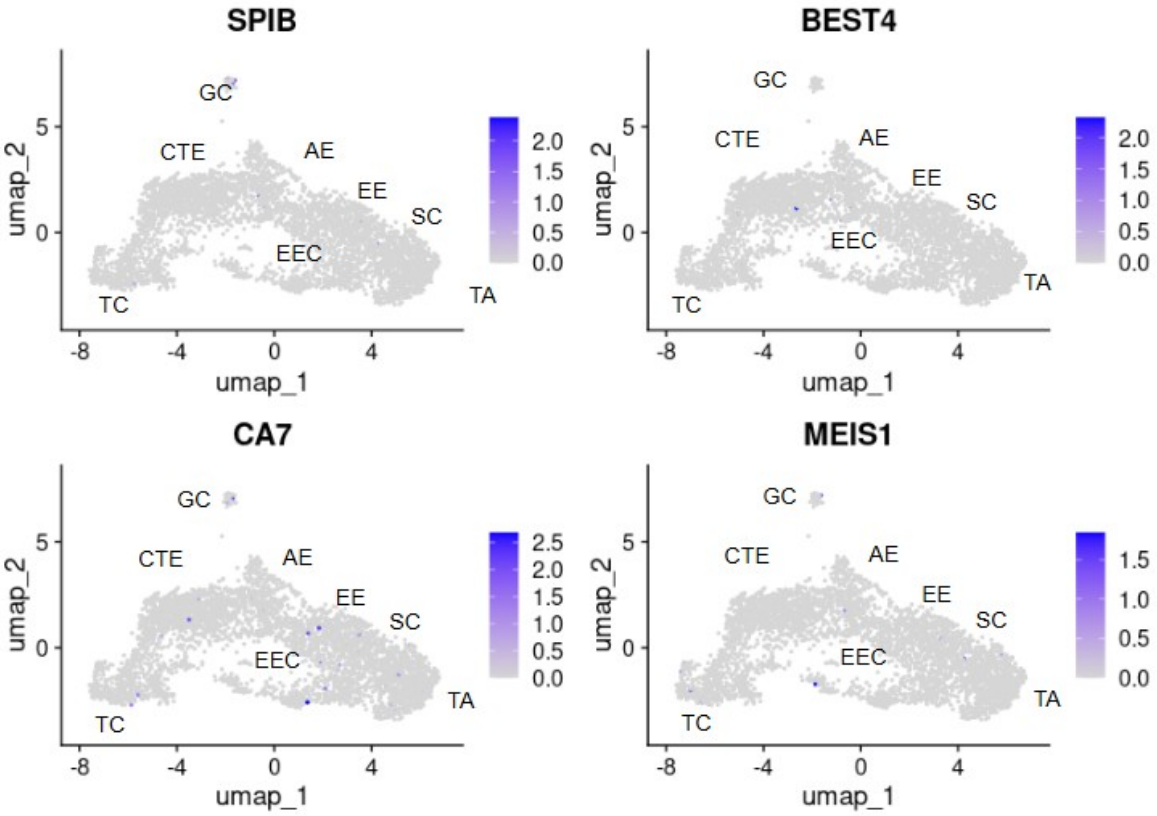
Interestingly, chromatin accessibility data provided additional, if partial, support for potential lineage priming. Several cell types — notably undifferentiated and absorptive populations — exhibited elevated accessibility at key Best4 regulatory loci, including *SPIB* (chr19:50431531–50432133), *BEST4* (chr1:44784405–44784622), and *MEIS1* (chr2:66431315–66431547) (Figure 34B; Supplementary Table S1). These findings suggest an epigenetically permissive landscape for Best4 activation, even in the absence of robust transcript expression.

Closer inspection of the *BEST4* locus further supports this view. Chromatin accessibility spanned both promoter and intragenic regions in Enterocytes, Absorptive Enterocytes, TAs, and Stem Cells (Figure 34C), highlighting latent regulatory activity within absorptive compartments. By contrast, Goblet Cells, and EECs exhibited minimal accessibility, consistent with transcriptional repression and lineage exclusion. Notably, Tuft Cells displayed partial accessibility - mostly identified in intragenic regions - alongside low-level *BEST4* expression, raising the possibility that some Tuft-like progenitors may represent transitional or plastic states with latent potential for Best4 differentiation. Co-accessibility analysis further reinforced this interpretation by identifying high-confidence cis-regulatory interactions linking distal enhancer regions to the *BEST4* promoter, indicative of poised or active transcriptional regulation (data not shown).

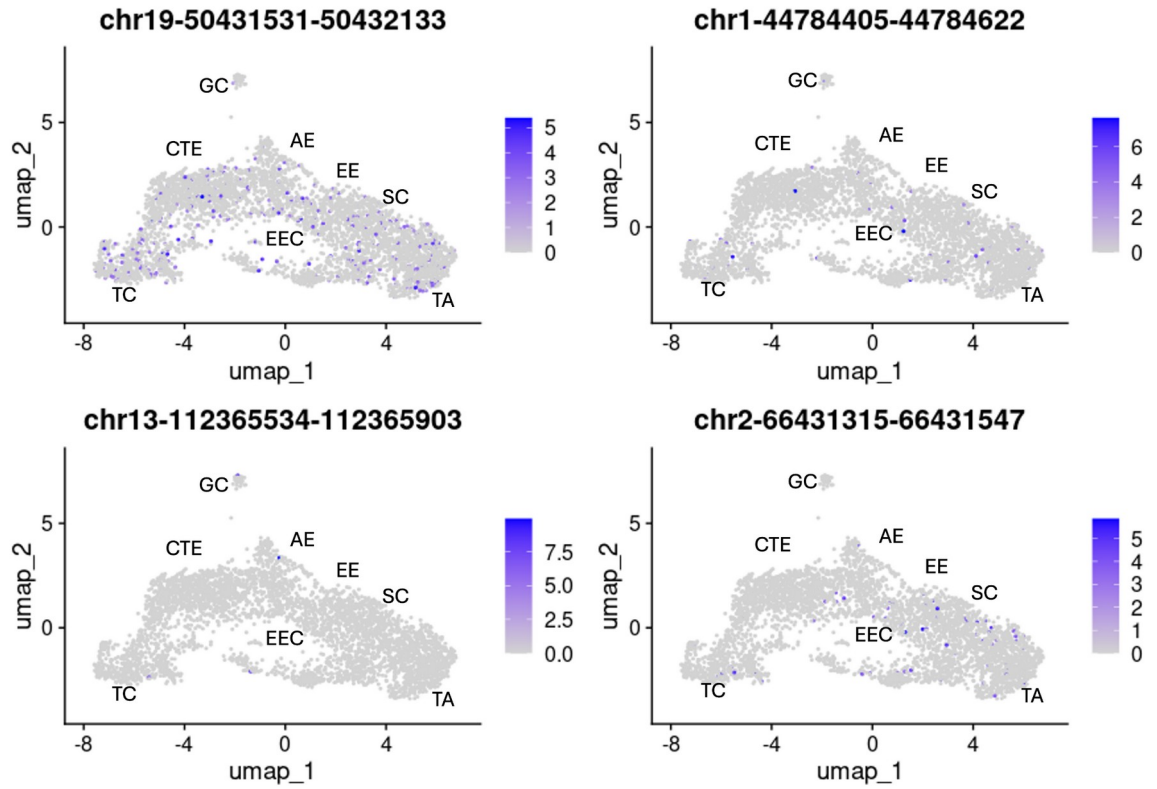
Together, these data suggest that while the epigenetic machinery necessary for Best4 identity is at least partially engaged, transcriptional commitment remains highly constrained — likely

due to incomplete differentiation cues, or insufficient niche specification under current *in vitro* conditions.

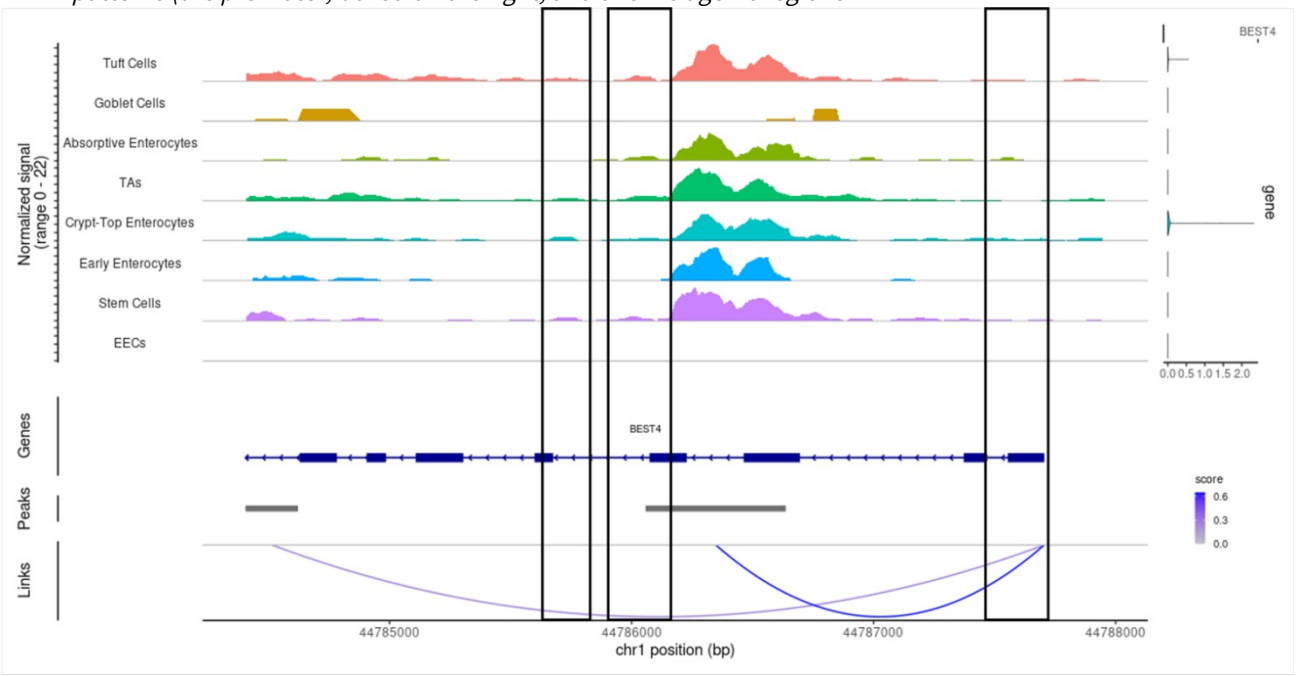
Figure 34 - Gene expression and chromatin accessibility of Best4 markers in epithelial cell populations.
A) Expression of canonical and associated Best4 cell genes



B) Chromatin accessibility at genomic loci linked to Best4 cell markers



C) Chromatin accessibility tracks for the BEST4 loci, highlight three regions with distinct accessibility patterns (the promoter, boxed on the right, and two intragenic regions



Legend: (A) UMAP plots showing expression of SPIB, BEST4, CA7, and MEIS1 across annotated epithelial cell types. TC- Tuft Cells, CTE- Crypt-Top Enterocytes, GC- Goblet Cells, AE- Absorptive Enterocytes, EE- Early Enterocytes, EEC- Enteroendocrine Cells, SC-Stem Cells, TA- TAs. (B) Chromatin accessibility UMAPs for selected genomic regions associated with Best4+ markers (displayed in the same order as the corresponding transcript expression plots shown above): SPIB (chr19:50431531–50432133), BEST4 (chr1:44784405–44784622), CA7 (chr13:112365534-112365903), and MEIS1 (chr2:66431315–66431547). (C) Genomic browser view showing chromatin accessibility peaks across samples at the BEST4 locus, highlighting accessibility differences among epithelial subtypes.

4.2.11 Interpretation of Findings

Despite the successful implementation of multiomic profiling, this section highlighted several critical technical and biological limitations that continue to constrain the detection and resolution of rare epithelial lineages such as Best4 Cells. One of the primary challenges lies in sequencing depth. While RNA-seq components reached moderate saturation (just below 20%), this level proved insufficient for chromatin accessibility profiling of hIEOs. ATAC-seq requires deeper coverage to resolve sparse, genome-wide signals and capture the full spectrum of open chromatin regions. The resulting ATAC datasets were notably unresolved, especially in low-frequency or transitional cell populations, reducing peak resolution and limiting the fidelity of

chromatin-based clustering. Even transcriptionally well-defined clusters exhibited heterogeneous chromatin accessibility patterns. Metrics such as nucleosome signal, TSS enrichment, and peak count revealed broad variability within clusters, suggesting that limited sequencing depth may obscure finer epigenetic distinctions. In some cases, entire clusters — particularly those comprising rare or transitional states like Goblet Cells, EECs, or progenitors — lacked sufficient chromatin signal to yield meaningful insights, further emphasizing the need for higher-resolution epigenomic datasets.

Biological constraints also emerged. Although the Fujii protocol was selected for its superior epithelial complexity and recovery rate, it failed to consistently induce full transcriptional commitment to the Best4 lineage. While chromatin regions near lineage-defining loci (e.g., *BEST4*, *SPIB*, and *MEIS1*) showed signs of accessibility — suggesting epigenetic priming — actual expression of these genes remained sparse and diffuse. This disconnect indicates that the *in vitro* conditions tested here were likely to lack critical niche-derived signals or environmental cues necessary for terminal differentiation. Extended culture duration or alternative stimulation strategies (e.g., exposure to microbial metabolites, cytokine stimulation, or BMP modulation) may be necessary to bridge this gap between chromatin readiness and transcriptional activation.

Additionally, isolated instances of chromatin accessibility at Best4-associated loci were observed in unrelated lineages, such as Tuft Cells. These cases likely represent plastic or transitional states rather than true lineage commitment, highlighting the complexity of lineage bifurcation in the intestinal epithelium. Finally, inter-donor variability presented another confounding factor. Despite profiling hIEOs from healthy, inflamed, and non-inflamed UC-derived samples, no consistent pattern of Best4 emergence was observed, pointing to intrinsic biological differences that further complicate rare lineage recovery.

In summary, both technical and biological factors limit the current capacity to resolve rare epithelial identities using scMultiome-seq approaches. Overcoming these challenges will require significantly greater sequencing depth, optimized protocols tailored to capture of epigenetic and transcriptional information, and refined *in vitro* differentiation strategies designed to mimic *in vivo* niche cues with greater fidelity. Nevertheless, with further optimisation and the use of higher-throughput sequencing platforms, this framework holds substantial potential for providing meaningful insights into epithelial lineage dynamics in organoid systems.

4.3 Engineering a Lineage Tracing Platform Using CellTagging in Epithelial Organoid Monolayers

4.3.1 Rationale for Technological Approach

Given the technical and biological constraints identified throughout this work — particularly the challenges in resolving the differentiation trajectories of intermediate cell states, which may hold the key to priming rare epithelial lineages — an alternative strategy was proposed to more effectively investigate epithelial cell identity and fate decisions. To investigate epithelial cell dynamics and lineage commitment over time, a lineage tracing framework was incorporated into the experimental design. This technique enables high-resolution, transcriptome-linked clonal analysis, allowing for the reconstruction of lineage trees alongside gene expression profiling. By capturing how individual clones expand, differentiate, or diverge in response to environmental cues, this strategy could provide critical insights into fate decisions and lineage plasticity. This would be particularly valuable for understanding the dynamics of rare cell populations and identifying the specific cues that drive their emergence or suppression in organoid cultures. This section addresses the development and validation of this approach.

Among the available methodologies for lineage tracing, lentiviral transduction coupled with the CellTagging protocol was selected due to its flexibility in barcode design, compatibility with single-cell technologies, and suitability for transcriptome-wide lineage reconstruction[289,290,298,341–343]. CellTagging is a transcriptome-compatible lineage tracing technique that enables the reconstruction of clonal relationships at single-cell resolution. It relies on the lentiviral delivery of synthetic DNA barcodes—termed CellTags—which are stably integrated into the genome and transcribed alongside endogenous genes. These barcodes can

be detected during single-cell RNA sequencing, allowing individual cells to be linked to their clonal origin. By introducing distinct CellTag libraries at multiple timepoints during an experiment, it is possible to track how progenitor cells give rise to differentiated descendants and how clonal trajectories evolve over time. Cells that share early barcodes but diverge in later tags can be interpreted as arising from a common ancestor, enabling the inference of lineage bifurcations and developmental hierarchies. This temporally resolved barcoding approach is especially powerful for studying dynamic cellular processes, such as differentiation, regeneration, and response to perturbations, while simultaneously capturing gene expression profiles.

4.3.2 Rationale For Adopting a 2D hIEO Conformation

Despite its advantages, applying the CellTagging approach within conventional 3D organoid cultures presents technical limitations. The cystic and enclosed architecture of 3D organoids embedded in Matrigel restricts viral access to the luminal surface, limiting the efficiency of lentiviral delivery[344]. Since successful viral transduction relies on direct interaction between viral particles and the target cell membrane, the spatial inaccessibility of many epithelial populations in 3D structures limits the uniformity and reproducibility of labelling across the epithelial population.

To circumvent these challenges, hIEOs were transitioned into a two-dimensional (2D) monolayer format. In this configuration, cells grow as a flat epithelial sheet, with both apical and basolateral surfaces exposed to the culture medium (see Supplementary Figure S19). This significantly enhances the accessibility of the entire epithelial surface to viral particles, improving transduction efficiency and enabling more uniform barcode delivery. As a result, the probability of achieving sufficient clonal complexity and coverage is greatly increased.

4.3.3 Design of a Preliminary Validation Strategy to Evaluate Biological Fidelity Between 2D and 3D Organoid Cultures

Prior to implementation of the lineage tracing strategy in this new 2D format, it was essential to evaluate whether organoid monolayers maintained the biological relevance and transcriptional fidelity of their 3D counterparts. A central concern was the potential for transcriptional drift or phenotypic loss due to adaptation to planar culture, or to the higher passage numbers required to generate sufficient cell quantities for downstream applications.

To address these considerations, an initial comparative transcriptomic profiling experiment was performed. Matched samples from 3D and 2D cultures were analysed to determine whether the monolayers could faithfully recapitulate the lineage complexity, differentiation potential, and functional states characteristic of conventional 3D organoids. The results of this analysis (shown below, Sections 4.3.5-4.3.6) provided foundational evidence supporting the biological validity of the 2D system as a suitable platform for lineage tracing and cell fate mapping.

While this experiment enabled a broad assessment of organoid culture fidelity across multiple dimensions, this section will focus on two primary objectives:

1. To evaluate the transcriptional and functional equivalency between organoids cultured in 3D and their 2D monolayer counterparts; and
2. To investigate the effects of extended passaging on transcriptomic stability and lineage representation.

4.3.4 Experimental Design: High-Throughput Single-Cell Characterisation of Different Culture Conditions in 2D vs 3D conformations

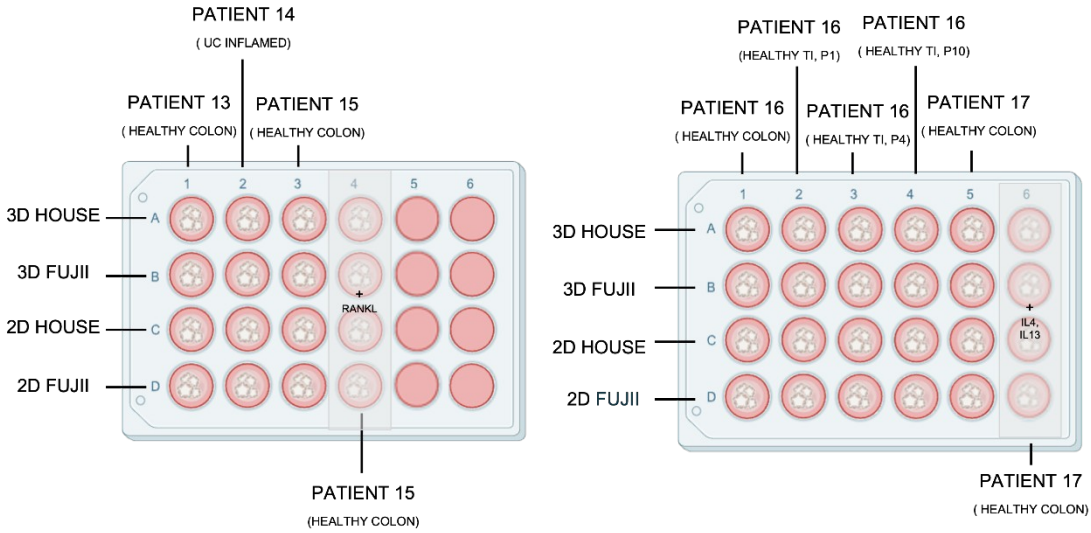
Organoids cultured as monolayers exhibit increased cellular packing and adhesion compared to their 3D counterparts, making them more challenging to dissociate into single-cell suspensions[345,346]. This is primarily due to the formation of a continuous, polarised epithelial sheet with reinforced cell–cell junctions, including TJs, AJs, and desmosomes, which are more uniformly developed on rigid two-dimensional substrates. The increased surface contact with the culture matrix also promotes stronger cell–matrix adhesion through integrin-mediated interactions. These physical and molecular features result in greater structural cohesion, requiring more extensive enzymatic digestion or mechanical disruption for efficient single-cell recovery. However, overly aggressive dissociation may compromise cell viability and surface marker integrity, presenting a key technical challenge when preparing monolayer cultures for downstream applications.

The successful establishment of this platform required the development of a robust and reproducible dissociation protocol capable of producing uniform single-cell suspensions from 2D organoid cultures. Such protocol should minimize cell clumping while maintaining high cell viability, thereby ensuring compatibility with the 10X Genomics platform. Achieving a high-quality single-cell suspension is essential to prevent chip clogging and avoid compromising data resolution. However, dissociating 2D monolayers proved considerably more challenging than 3D organoids, as monolayer cultures develop strong intercellular junctions and densely packed mucus layers that hinder enzymatic penetration. Through iterative refinement of publicly available methods, an optimized enzymatic-mechanical dissociation workflow was established (see Supplementary Methods, Section 1.4). Next, a comprehensive, multiplexed single-cell RNA sequencing experiment was designed to interrogate the transcriptional

integrity, lineage fidelity, and responsiveness of organoid monolayers under various conditions. In addition to assessing basal culture states, this experiment incorporated stimulation with selected factors — such as RANKL — to enhance differentiation into rare cell types.

For this experiment (EXP06), organoids were derived from five donor samples, including four healthy controls (three from colon only, one paired colon-TI) and one UC patient (inflamed region only). These were cultured across House and Fujii media, with 2D and 3D conformations, and subject to either no stimulation, stimulation with RANKL, or stimulation with IL13/IL4, yielding a total of different 40 conditions (Figure 35).

Figure 35 – Experimental layout for Experiment 6: stimulation assays across multiple culture conditions and sample types



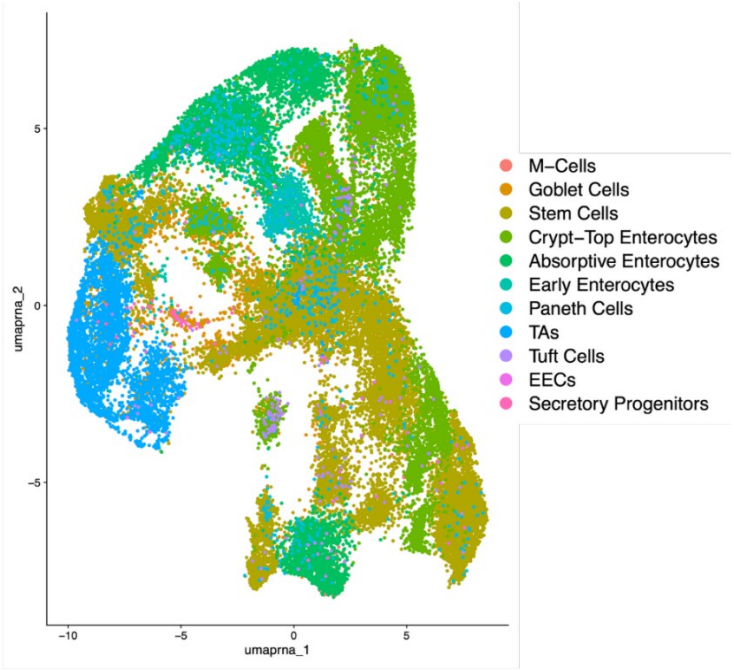
Legend: For patient 16, the different samples labelled P1, P4, and P10 represent organoids derived from cells subject to different passage numbers.

Fujii medium was prioritised here versus Holloway as it had shown improved cell recovery, better representation of rare cell types, and retention of a broader epithelial landscape in previous experiments, with House medium kept as a baseline control condition with balanced cell type representation (Figure 22, Figure 23, and Section 5.2.3.1). Although RANKL stimulation

was not able to induce sustained differentiation of Best4 Cells in EXP03 (Sections 3.4.7, 5.2.2.3, and 5.2.3.2), stimulations with RANKL and IL13/IL4 were used here for their ability to engage SPIB signalling pathways (and therefore potentially support Best4-Cell emergence) and drive shifts towards rare or specialised cell types (M-Cells for RANKL, and Tuft Cells for IL13/IL4) - which could serve as valuable use-cases to validate the proposed lineage tracing approach. To maximise biological insight while accommodating practical limitations, stimulation conditions were selectively applied only to two hIEO lines derived from healthy colonic biopsies, rather than uniformly distributed across all donor samples. This strategic variation was driven by the high cell input requirements of the experiment, which restricted the number of conditions that could be implemented per sample. Only the two most proliferative lines yielded sufficient cells to support additional stimulation arms in both monolayer and 3D formats, enabling a broader evaluation of media effects and lineage responses under varied environmental cues.

All samples were processed concurrently in a single Chromium Next GEM Single-Cell 3' High Throughput + Feature Barcoding run, using the validated multiplexed single-cell RNA-sequencing workflow outlined in Section 3.2.2. This approach ensured uniform sample handling and minimized technical variability, thereby enhancing the reliability of downstream analyses. After stringent quality control, 37,779 high-quality cells were retained across the dataset (Figure 36).

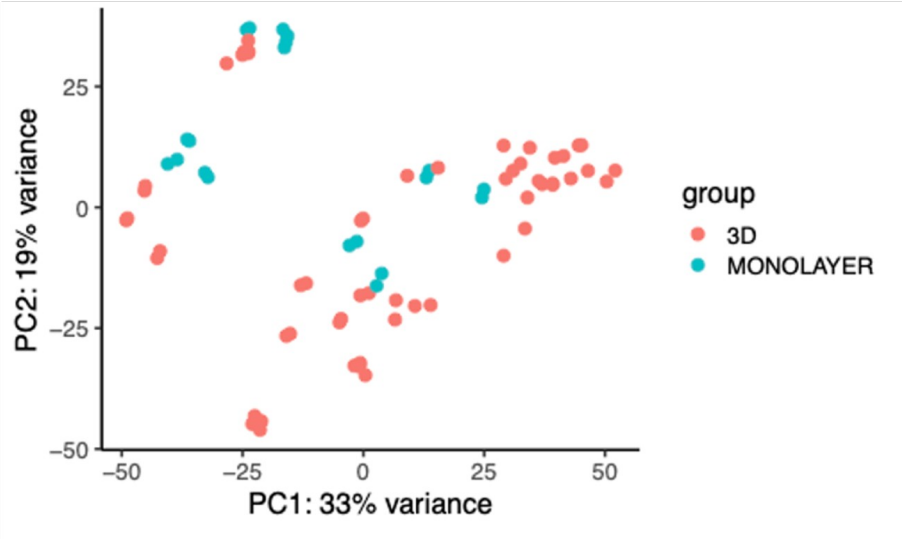
Figure 36 – Integrated cell clustering across experimental conditions and samples



4.3.5 Evaluating the Impact of 2D vs 3D Conformation in the Transcriptomic Signatures of hIEOs

To evaluate the impact of 2D culture configurations on the transcriptomic activity of hIEOs, the dataset was first subjected to pseudobulk analysis to identify the principal sources of variability across all conditions, with a particular focus on differences between 2D and 3D formats. Figure 37 presents a PCA analysis of transcriptomic variation grouped by culture conformation. This projection reveals partial separation between 2D monolayers and 3D cultures along the first two principal components, which together explain 52% of the total variance. These results indicate that culture geometry might contribute fairly to transcriptional divergence, supporting the relevance of physical context in shaping epithelial cell states.

Figure 37 - PCA of transcriptomic variation as a function of hIEO culture conformation

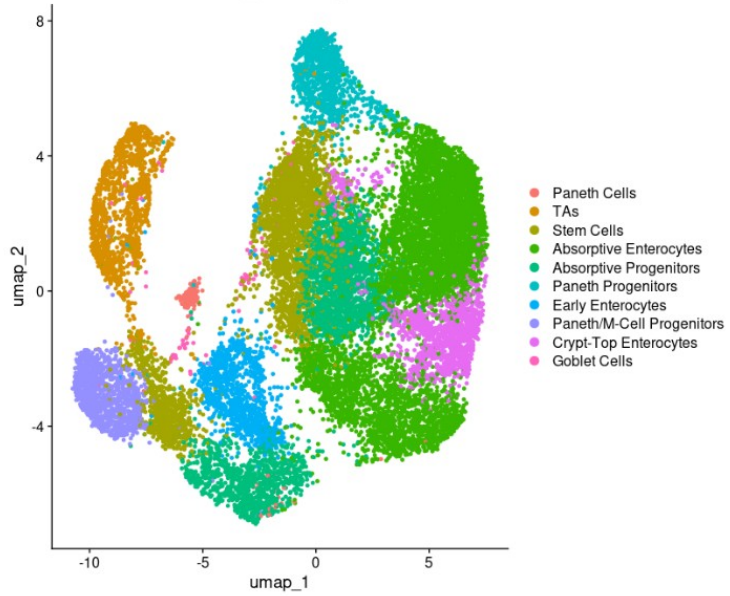


Legend: PCA of global gene expression profiles in matched hIEO cultures grown in either conventional 3D conformation (red) or 2D monolayer format (blue). Each point represents an individual sample. PC1 and PC2 explain 33% and 19% of total variance, respectively. Although monolayer samples exhibit a slight shift along PC1, the overall clustering pattern indicates broadly similar transcriptional states across both formats. These results suggest that hIEO monolayers maintain transcriptomic fidelity relative to their 3D counterparts and can serve as a biologically relevant model for downstream applications.

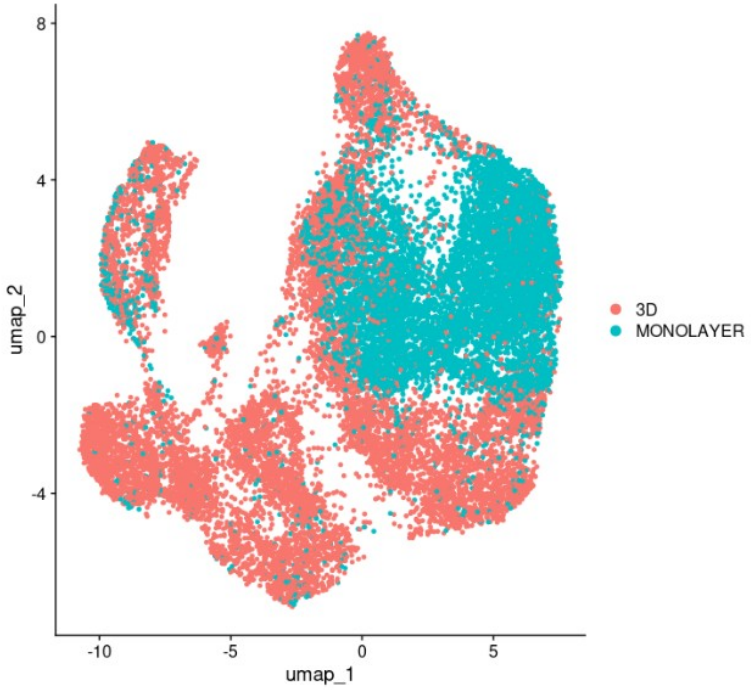
As the central objective of this study was to investigate epithelial differentiation trajectories and the specification of rare cell types, it was critical to determine whether the observed transcriptomic differences between 2D and 3D culture configurations reflected superficial global shifts or fundamentally altered cell fate decisions - while it is logical and expectable that overall gene expression profiles differ between conformation types, the key question was whether these differences affected the composition and identity of specific epithelial subtypes. If differentiation into major lineages remained consistent across formats, then 2D monolayer cultures could be confidently used as a valid model for studying cell type specification and rare lineage emergence. To address this, a subset of donor-matched, unstimulated TI and colon samples was isolated (“conformation subset”), resulting in 19,033 high-quality cell transcriptomes (Figure 38). By isolating conformation as the only experimental source of

variation, this subset enabled a focused assessment of whether structural changes significantly impact epithelial identity at the lineage level.

Figure 38 - UMAP visualisations of epithelial cell type annotation and culture format distribution
A) UMAP of the conformation subset grouped by annotated cluster



B) UMAP of the conformation subset grouped by culture conformation

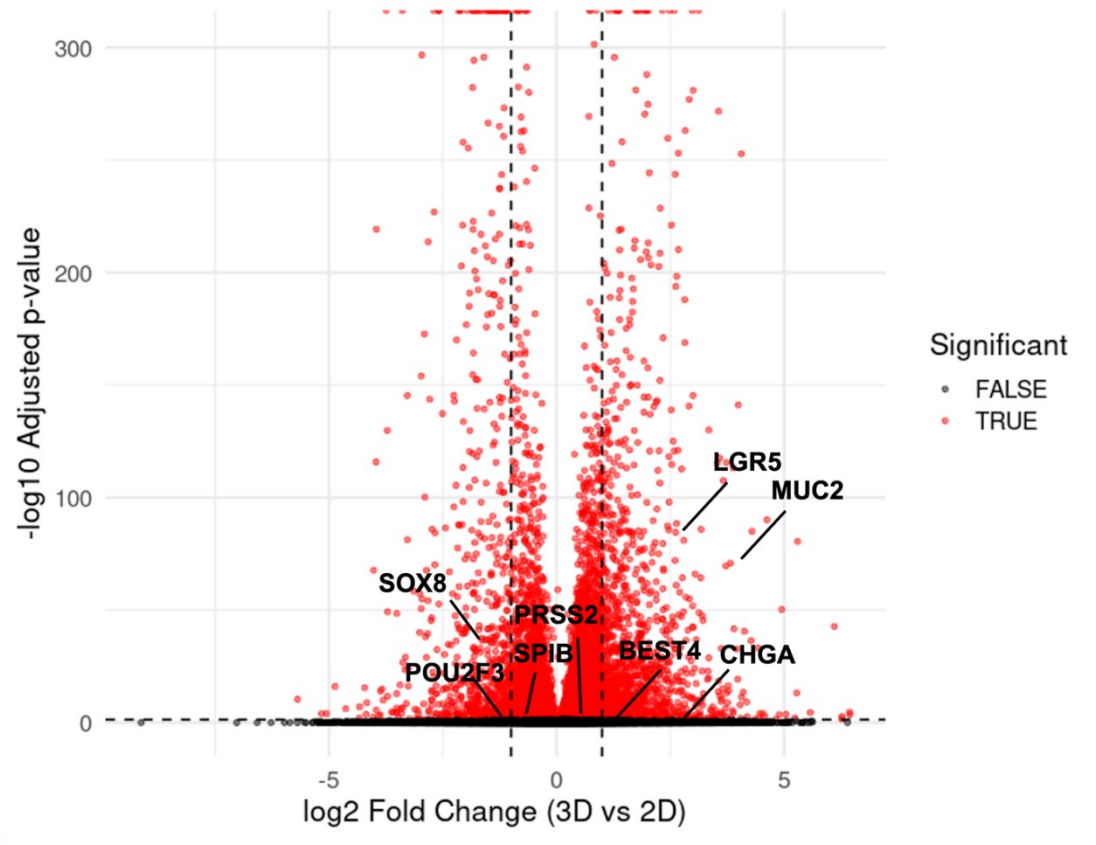


1

2 The first step was to evaluate whether 2D organoids recapitulate the transcriptomic features of
3 their 3D counterparts. Given that 3D organoids have been shown to closely resemble primary
4 intestinal epithelial cells — as supported by published literature, and further demonstrated in
5 Section 5.3 — establishing a similar degree of fidelity in 2D cultures would support their use as

a valid model for studying epithelial differentiation. To assess this, a comprehensive differential expression analysis was performed, comparing 2D and 3D cultures both at the global level and within individual epithelial clusters (Figure 39 and Supplementary Figure S20). Despite broad transcriptional differences between the two formats (**Figure 39**), core epithelial lineage markers – including *BEST4* - were consistently expressed in both 2D and 3D conditions, indicating preservation of major cell identities. While 3D cultures exhibited slightly elevated expression of genes associated with actively cycling Stem Cells *(LGR5)* and Goblet Cells *(MUC2)*, 2D cultures retained the essential transcriptional programmes defining epithelial subtypes. These findings suggest that, although certain lineage outputs may be enhanced in 3D environments, 2D cultures reliably maintain the molecular identity of key epithelial lineages. Similar findings were observed when comparing expression profiles between 2D and 3D conformations for samples stimulated with either RANKL or IL13/IL4 (data not shown). Thus, 2D organoid monolayers systems represent a valid and tractable model for studying epithelial cell identity, fate decisions, and the emergence of rare populations such as Best4 Cells.

Figure 39 – Volcano plot showing differential gene expression between 3D and 2D epithelial organoid cultures.

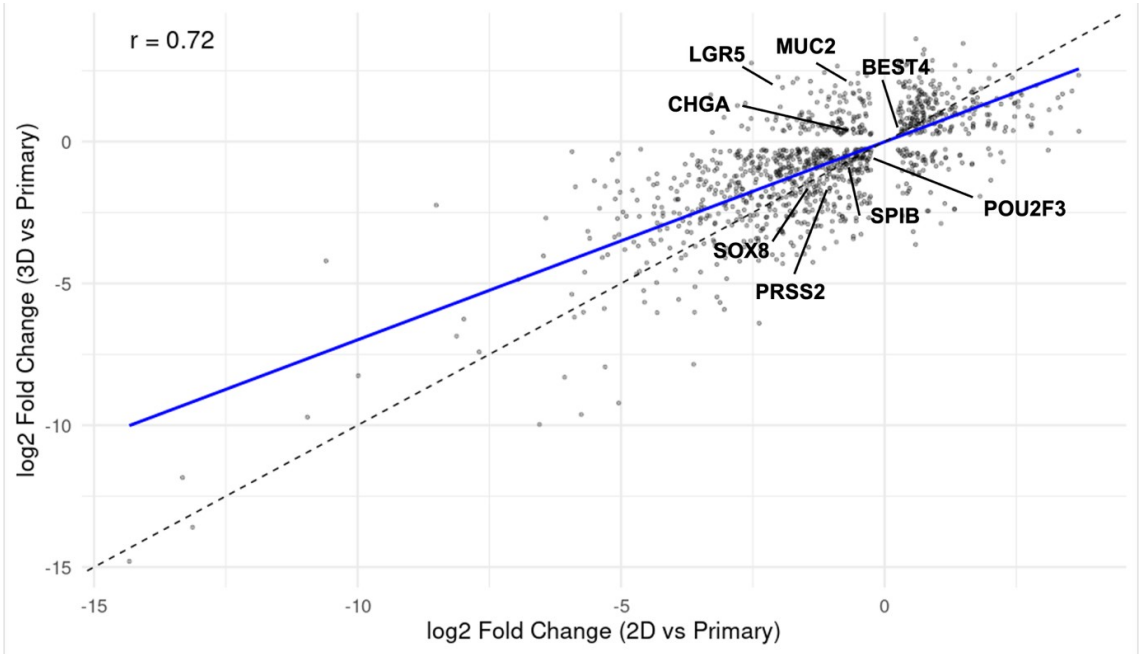


Legend: Volcano plot showing log2 fold change (x-axis) versus – log10 adjusted p-value (y-axis) for all genes comparing 3D to monolayer (2D) culture formats. Genes significantly differentially expressed (adjusted p-value < 0.01) are highlighted in red. Selected marker genes with known roles in intestinal identity and differentiation (e.g. LGR5, MUC2, CHGA, BEST4, SPIB) are annotated.

Next, to rigorously evaluate the fidelity of 2D organoid cultures as a model for intestinal epithelial biology, a differential expression analysis was performed comparing both 2D and 3D cultured cells to primary epithelial cells. As a gold-standard reference, the large, well-annotated epithelial single-cell dataset from Smillie *et al.* (2019) was used as the benchmark. This comparative approach enabled a systematic assessment of how closely each *in vitro* format recapitulates the transcriptional landscape of native tissue, providing insight into the physiological relevance and limitations of each model system[42]. For each gene, log2 fold changes (logFC) relative to the primary reference dataset were calculated independently in 2D

and 3D conditions, and the resulting values were plotted against each another (Figure 40). This gene-wise comparison revealed a strong positive correlation (r = 0.72) between the two systems, indicating that the transcriptional changes induced by culture conditions are broadly consistent. The regression trend closely follows the identity line (y = x), particularly for moderately expressed and lineage-specific genes, supporting the notion that 2D and 3D cultures largely preserve the same expression programs. For example, key epithelial lineage markers — including *BEST4*, *MUC2*, *CHGA*, and *SPIB* — were found to maintain consistent expression patterns across both systems relative to the primary tissue, further validating the biological relevance of the 2D culture approach. While 3D cultures showed slightly stronger preservation of some transcriptomic features, the overall directionality and identity of expression changes were retained in 2D, supporting the use of this culture conformation as a valid experimental model.

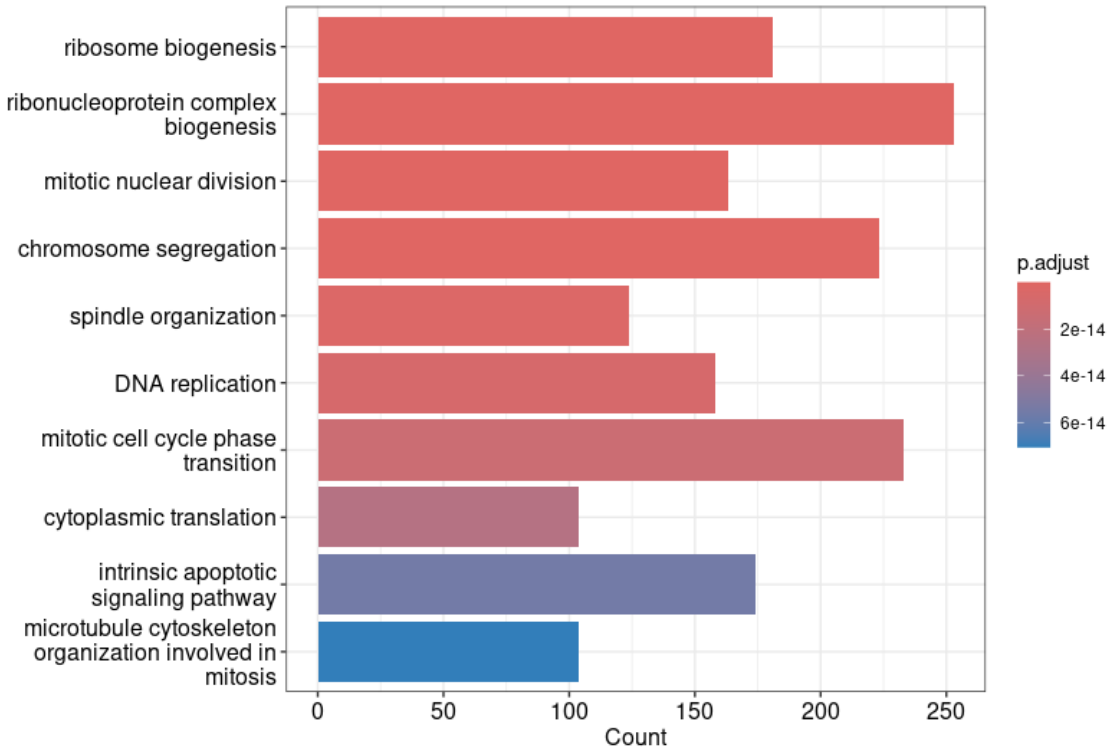
Figure 40 – Gene-wise logFC correlation of organoid culture conformation vs primary epithelium (3D: primary cells vs 2D: primary cells).



Legend: Comparison of log2 fold changes in gene expression between 2D and primary tissue (x-axis) and between 3D and primary tissue (y-axis). The positive correlation suggests that both 2D and 3D organoids deviate similarly from the primary tissue transcriptome, with 2D cultures showing a slightly stronger divergence for some lineage markers.

To further evaluate the biological relevance of differentially expressed genes between 2D and 3D cultures, GO enrichment analysis was performed. The results revealed that differentially upregulated genes in 2D conformations were significantly enriched for terms related to ribosome biogenesis, mitotic division, and cell cycle regulation (Figure 41), processes commonly associated with altered growth dynamics or substrate exposure *in vitro*. Importantly, no enrichment was observed for pathways defining epithelial lineage specification, supporting the conclusion that transcriptional differences between 2D and 3D organoids primarily reflect culture-induced metabolic and proliferative variation, rather than fundamental shifts in cell identity.

Figure 41 – GO enrichment of biological processes in 2D vs 3D conformations



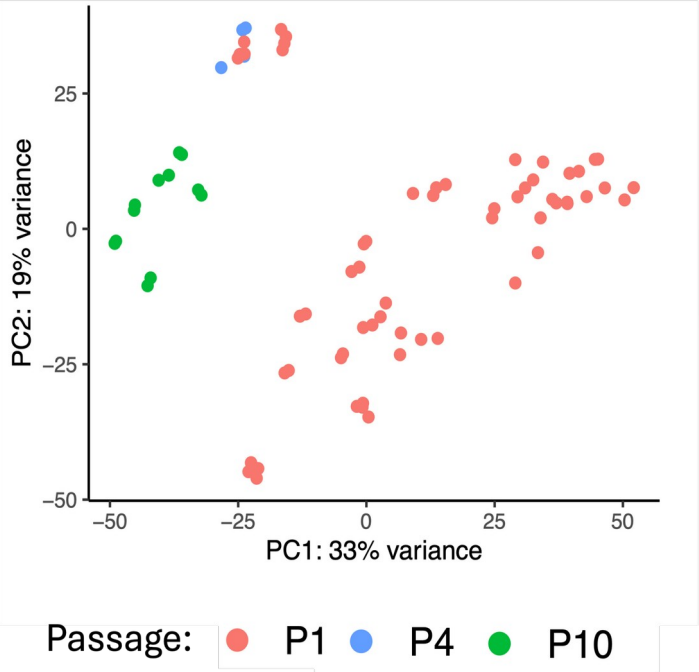
Legend: Bar plot showing significantly enriched GO biological processes among differentially accessible or expressed genes in selected epithelial cell populations. Bars represent the number of genes per GO term, coloured by adjusted p-value. Highly enriched terms include ribosome biogenesis, mitotic division, and chromosome segregation, indicating strong proliferative and biosynthetic activity.

4.3.6 Evaluating the Impact of Passage Number on the Transcriptomic Fidelity of hIEO Monolayers

Given the complexity of this experiment — involving multiple rounds of viral infection and extensive processing steps — it was necessary to expand organoids to a significant cell number. While all previous 3D experiments were consistently performed at P1-P3, maintaining this early passage number during development of monolayer cultures would be more challenging. Therefore, the second objective of these preliminary analyses was to determine whether passage number influences the transcriptomic landscape of hIEO-derived monolayers.

To address this, organoid monolayers that had been cultured at early (P1), intermediate (P4), and late (P10) passages were profiled and subjected to pseudobulk PCA. The resulting projection, grouped by passage number, revealed some similarity between P1 and P4, but clear separation from these to P10, indicating that passaging induces progressive transcriptomic divergence (Figure 42). These shifts likely reflect cumulative changes in cellular composition, epigenetic regulation, or adaptation to *in vitro* conditions.

Figure 42 - PCA of transcriptomic variation as a function of the passage number.

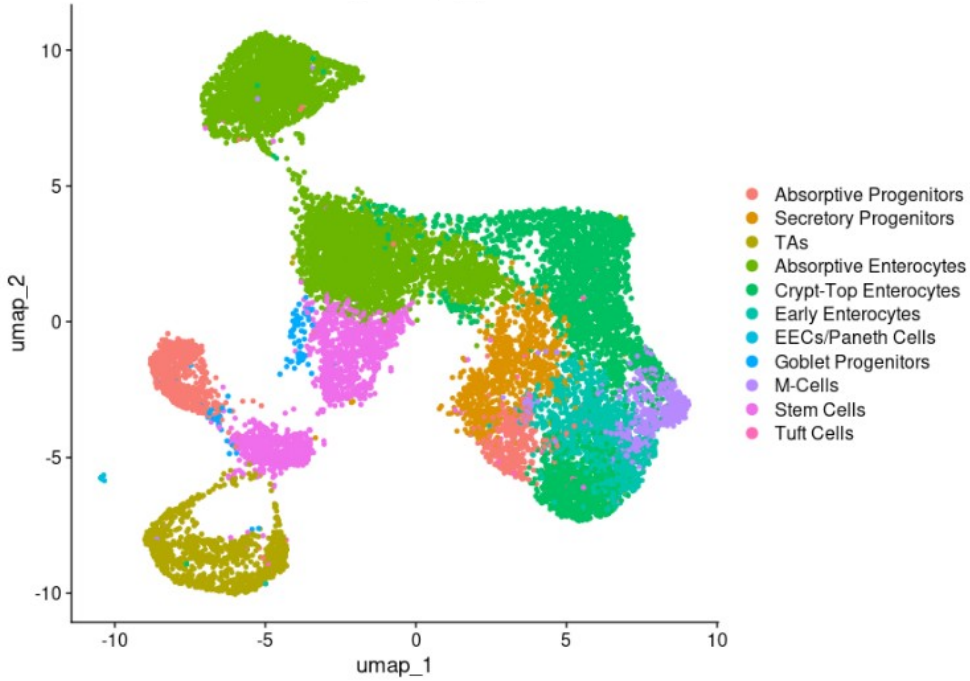


Legend: PCA plot showing transcriptomic variance among epithelial organoid samples at passage 1 (P1), passage 4 (P4), and passage 10 (P10). Each point represents a sample, coloured by passage number. The separation by passage suggests that transcriptomic profiles diverge with extended culture, indicating passage-dependent transcriptional drift.

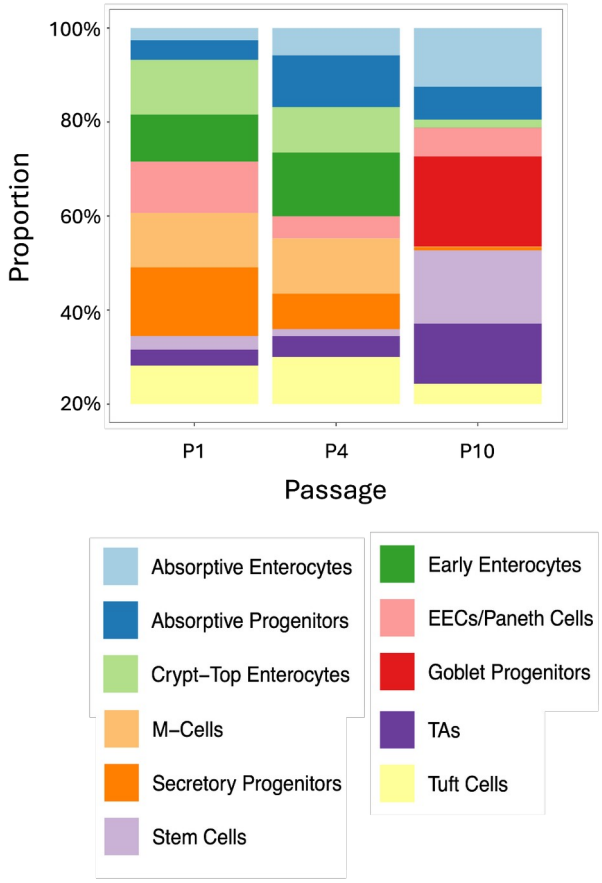
A subset of donor-matched, unstimulated TI samples was selected to compare transcriptomic profiles across identical passage numbers, yielding a total of 14,941 high-quality cells ("Passage subset"). The choice of TI samples was opportunistic, as these were the only samples for which organoids derived from such high passage numbers were available at the time. These were grouped by passage number into three datasets — P1, P4, and P10 — and processed independently for clustering and annotation before integration (Figure 43A). As anticipated, cell type annotation of each dataset revealed marked shifts in epithelial subtype representation with increasing passage number (Figure 43B). Early (P1) and intermediate (P4) passages exhibit a balanced distribution of cell types, including a substantial proportion of absorptive progenitors, TAs, secretory progenitors, and Stem Cells. By contrast, late passage cultures (P10) showed a marked shift towards undifferentiated states, accompanied by a

reduction in absorptive progenitors and Crypt-Top Enterocytes. Notably, secretory progenitors appeared increasingly biased toward Goblet Cell fates, and M-Cells were nearly absent by P10, underscoring the depletion of rarer and more specialised epithelial subtypes with continued passaging. These findings suggest that extended culture expansion disrupts normal differentiation trajectories, potentially due to epigenetic drift, exhaustion of progenitor pools, or altered responsiveness to niche cues.

1 *Figure 43 - Cell type composition and lineage distribution in epithelial cells across the passage subset.*
A) UMAP of epithelial cell types across all passage conditions



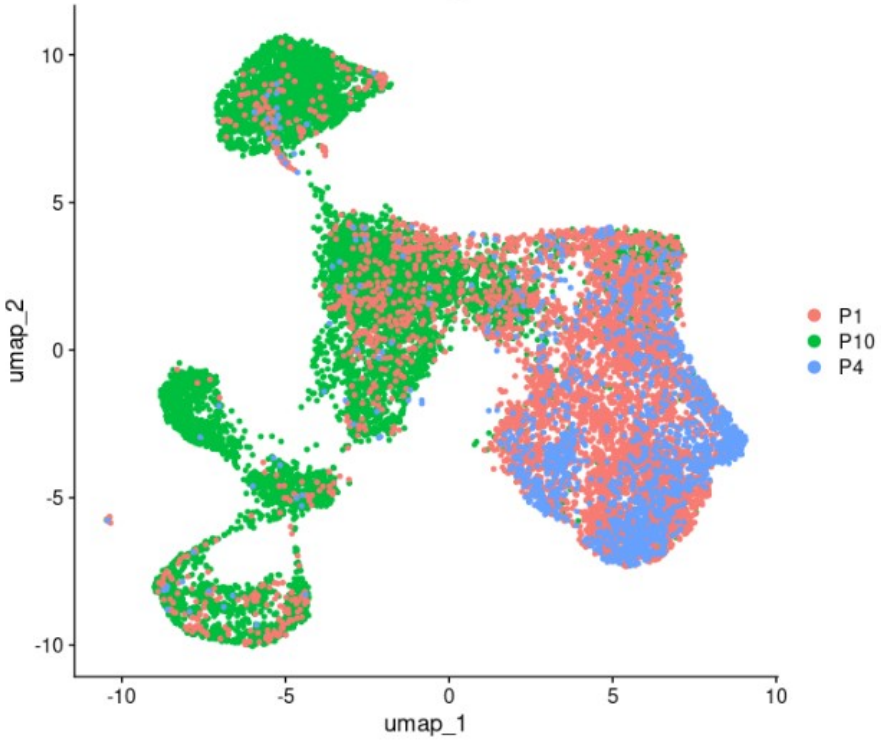
B) Proportional composition of cell clusters by passage number



Legend: (A) UMAP projection of integrated single-cell transcriptomes from organoids at passages P1, P4, and P10, coloured by epithelial cell type annotation. (B) Stacked bar plot showing the relative abundance of each cell type at each passage, indicating shifts in lineage distribution over time.

Accordingly, the UMAP projection of the integrated dataset grouped by passage number revealed clear passage-dependent transcriptional dynamics (Figure 43C). Cells from early (P1, red) and intermediate (P4, blue) passages were broadly mixed across the UMAP space, indicating high transcriptional similarity and minimal divergence between these two conditions. By contrast, cells from late passage (P10, green) show a distinct distribution, with prominent accumulation in regions of the UMAP that are underrepresented or nearly absent in P1 and P4, further reinforcing that prolonged passaging drives a shift in cellular composition and/or transcriptional states.

Figure 44 - UMAP of the passage subset grouped by passage number

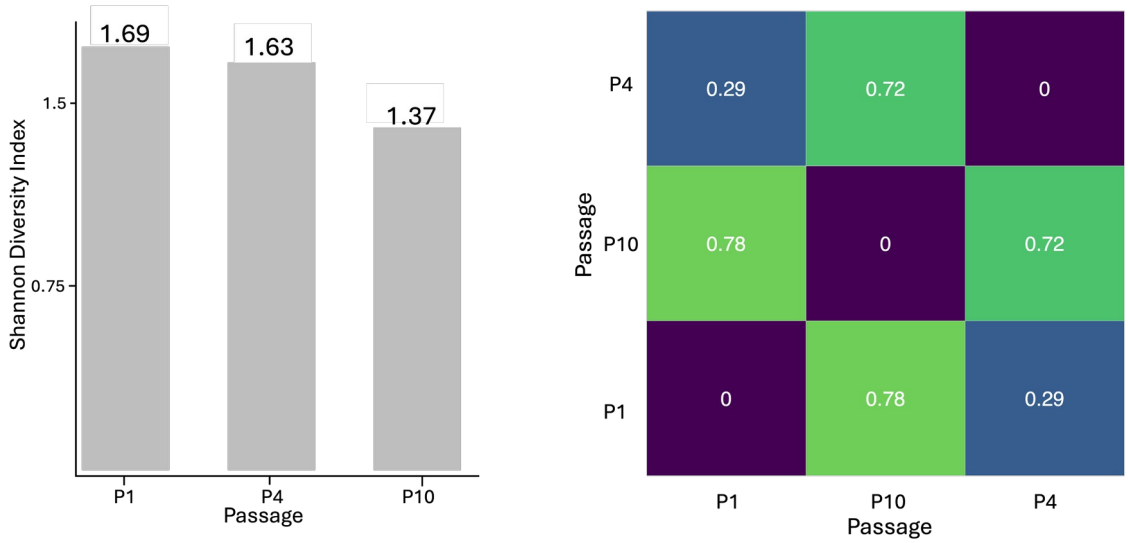


Legend: UMAP plot showing cells coloured by passage number, illustrating spatial shifts in cell distribution corresponding to transcriptomic changes.

To quantitatively assess the impact of passaging on epithelial diversity, two complementary metrics were calculated: the Shannon Diversity Index and Bray–Curtis Dissimilarity (Figure 45). The Shannon Diversity Index – a measurement of the overall richness and evenness of cell type distribution within each condition – showed that P1 and P4 displayed similar diversity values (1.69 and 1.63, respectively), indicating that a broad range of epithelial lineages was maintained at these stages. By contrast, the diversity score dropped more steeply at P10, (1.37), suggesting a reduction in cellular heterogeneity and a skew toward fewer dominant populations (Figure 45A). The Bray–Curtis dissimilarity matrix - a quantification of how different the cell type compositions are between passage pairs – showed lowest dissimilarity between P1 and P4 (0.29), consistent with their overlapping lineage profiles and shared transcriptional signatures (Figure 45B). However, comparisons involving P10 yielded substantially higher dissimilarity values (0.72–0.78), reflecting a moderate shift in population structure at later passages.

Figure 45 - Diversity/dissimilarity metrics for epithelial cell type composition across organoid passages.

A) Shannon Diversity Index by passage number *B) Bray-Curtis Dissimilarity by passage number*



Legend: (A) Shannon diversity index of epithelial cell type composition at passages P1, P4, and P10, indicating a decrease in transcriptional heterogeneity over time. (B) Bray–Curtis dissimilarity scores between cell type distributions at different passages, with higher values indicating greater dissimilarity. Together, these metrics suggest a progressive loss of cellular diversity and a shift in composition with extended organoid passaging.

Together, these results emphasise the need to account for passage effects when interpreting differentiation outcomes, particularly in studies focused on rare cell types or subtle lineage transitions. While early and intermediate passage numbers (P1 and P4) preserved a diverse and physiologically relevant epithelial composition, prolonged passaging (P10) seemed to result in progressive lineage imbalance, marked transcriptomic divergence, and reduced diversity. This divergence manifested not only as shifts in lineage representation, but also as broader alterations in global transcriptional identity, suggesting that late-passage cultures may progressively deviate from physiologically relevant epithelial states. Collectively, these metrics underscored the importance of prioritising low-passage cultures in studies aimed at modelling differentiation or resolving rare epithelial lineages, and informed the decision to use the earliest possible passages throughout this study to ensure fidelity and minimise artefactual divergence.

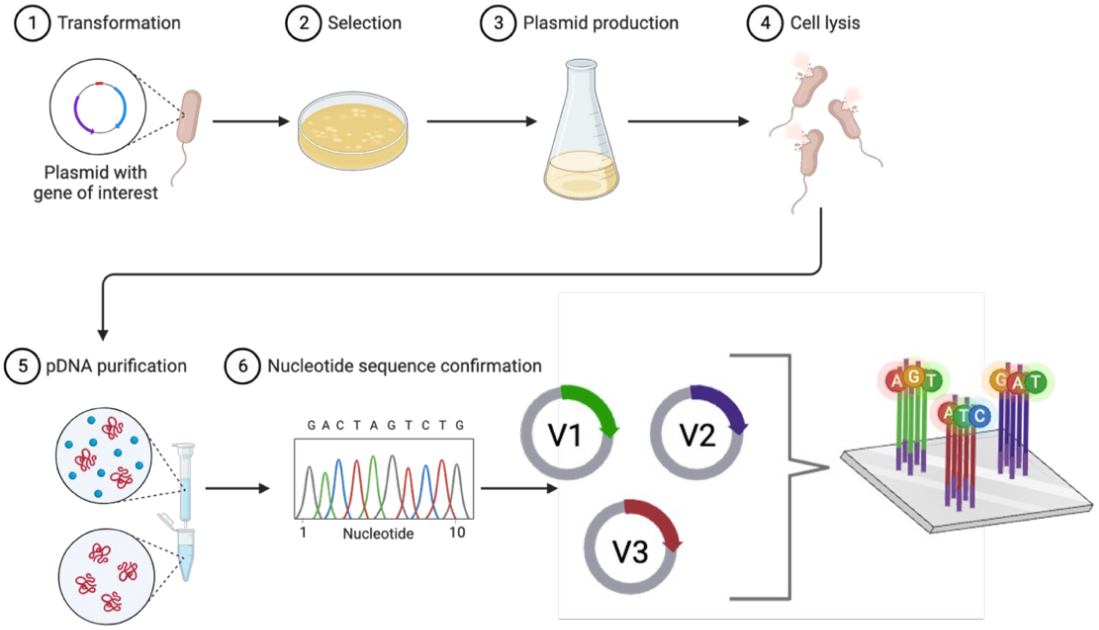
4.3.7 Generating CellTag Libraries

An overview of the initial library preparation steps is illustrated in Figure 46A. Quality control analysis revealed that all three libraries (V1, V2, and V3) displayed comparable levels of diversity, with a slight enrichment in V1 tags (Figure 46B). The total number of unique CellTags identified in each library ranged between ~50,000 and >60,000, exceeding the diversity obtained for V2 in the original CellTag publication (labelled as “V2*”; Figure 46BC). To account for differences in sequencing depth between experiments, tag counts were downsampled to match the sequencing saturation of the original dataset. Despite normalisation, the libraries

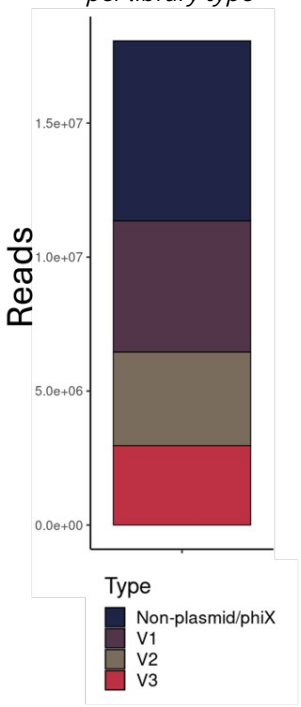
generated in this study retained a greater number of unique tags (Figure 46D), validating their quality and complexity for downstream lineage-tracing applications.

Figure 46 - CellTag library generation and benchmarking across V1, V2, and V3 constructs.

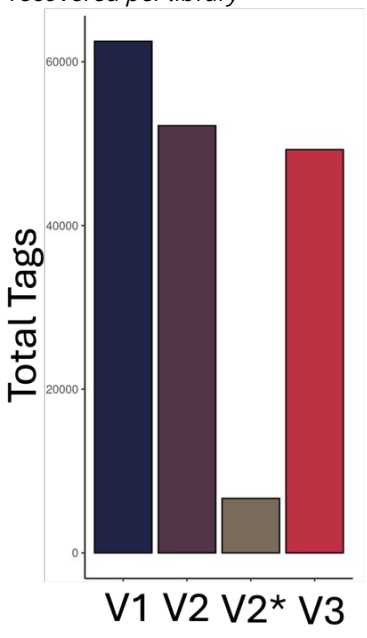
A) Schematic of CellTag lentiviral library generation and validation



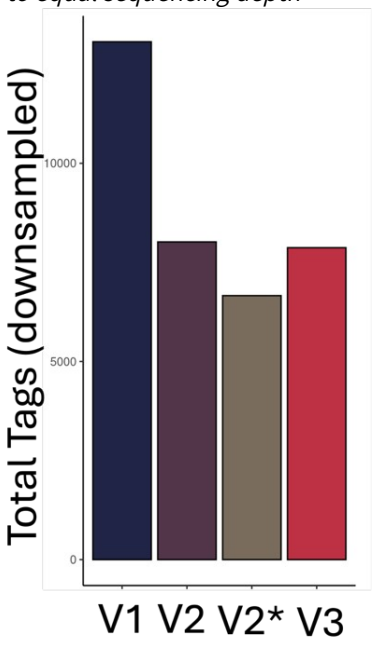
B) Sequencing reads per library type



C) Total number of unique tags recovered per library



D) Total tags after downsampling to equal sequencing depth



Legend: (A) Following the original CellTagging workflow, plasmids containing CellTag barcode cassettes were transformed into competent E. coli (Step 1), selected on antibiotic plates (Step 2), expanded in culture (Step 3), and harvested for plasmid extraction (Steps 4–5). Prior to lentiviral packaging, sequence confirmation and barcode diversity assessment were conducted. Sanger sequencing was first used for targeted validation (Step 6), followed by Illumina MiSeq high-throughput sequencing to

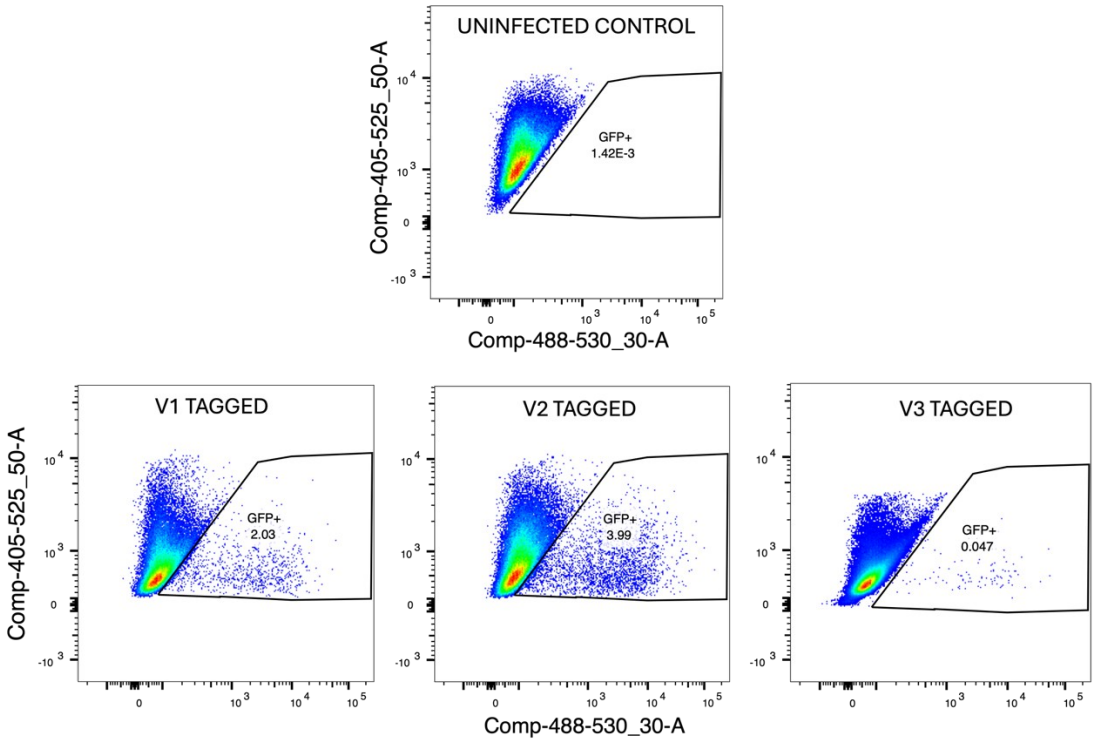
comprehensively evaluate library complexity and confirm sufficient diversity for downstream lineage tracing applications. V2: total tag count in the original protocol publication (Kong et al., 2020)[296].*

4.3.8 Generating CellTag Lentiviral Preps and Optimising Infectivity in hIEOs

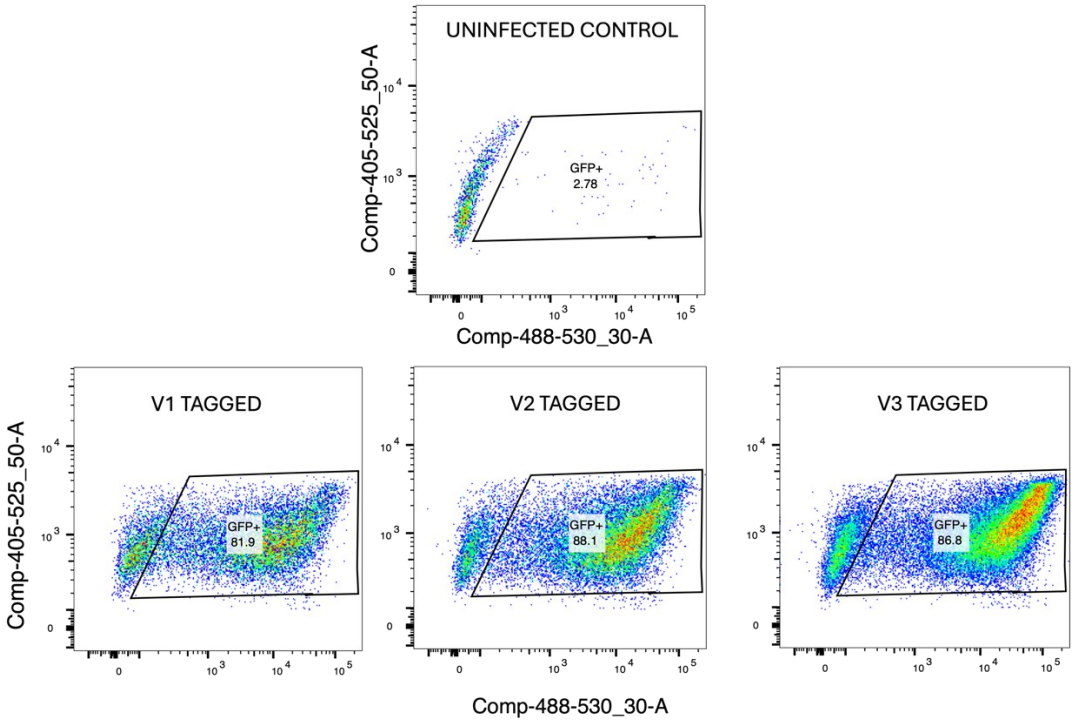
After initial viral preparations and preliminary infectivity assays, hIEOs consistently exhibited poor infection efficiency, as indicated by the very low percentage of GFP-positive cells (Supplementary Methods, Section 1.5; Supplementary Figures S21-S23). To overcome this limitation, several technical optimisation steps were implemented (Supplementary Figures S24-S25), including the addition of an ultracentrifugation step (as detailed in Supplementary Methods, Section 1.5). This modification aimed not only to concentrate the viral particles but also to enable buffer exchange, thereby ensuring a more organoid-compatible medium. Initial validation of the ultracentrifugation step generated very high viral concentrations and yielded highly favourable infectivity results (Supplementary Methods, Section 1.5; Figure 47; Supplementary Figure S26), justifying the use of this approach in subsequent experiments.

Figure 47 - Comparison of transduction efficiency before and after viral ultracentrifugation across CellTag libraries.

A) GFP+ cell yields prior to viral ultracentrifugation



B) GFP+ cell yields after viral ultracentrifugation



Legend: Flow cytometry plots depicting the percentage of GFP+ cells following transduction with V1, V2, or V3 CellTag lentiviral libraries, compared to uninfected controls. The top panel shows infection yields using crude lentiviral supernatant, while the bottom panel shows improved yields after ultracentrifugation-based viral concentration. Percentage values represent the fraction of live cells within the GFP+ gate, illustrating enhanced infection efficiency post-concentration.

1

4.3.9 Pilot Experiment 1: Assessing The Impact of Sample Storage Conditions on Cell Quality

Ideally, according to the original CellTagging protocol, cells would be split into two fractions at each timepoint: one processed immediately for single-cell sequencing (or stored to be run alongside the final fraction at the experiment's endpoint — thereby reducing batch effects and sequencing costs) and another fraction that continues to be sequentially tagged through the experiment. This strategy would allow for parallel temporal snapshots and lineage tracing from a common origin.

However, implementing this approach in organoid systems presents significant challenges. Unlike suspension cultures, or robust and phenotypically homogeneous cell lines, organoid-derived cells are highly sensitive to dissociation and re-seeding - especially in monolayer formats. Processing and fractionating cells at each timepoint would likely lead to the loss of rare or highly differentiated populations, skewing the representation of cell types. More critically, the structural integrity of organoid monolayers is essential for proper differentiation and cell–cell signalling. Disrupting this architecture, particularly the tight junctions crucial for maintaining monolayer identity, would not only reset the system's state but also render it biologically unrepresentative and technically infeasible.

A further refinement of this method involves generating biological replicates of the organoid system at each tagging point. In this design, two identical cultures would be infected simultaneously with the same CellTag library: after the infection period, one would be fixed and immediately stored, and the other allowed to proceed to the next tagging step. While the exact same cell progeny cannot be traced across biological replicates, the identical initial conditions and experimental procedure provide a controlled framework to examine progression across stages. This approach maintains biological integrity while allowing

approximate reconstruction of lineage trajectories and temporal states, circumventing the limitations imposed by organoid architecture and differentiation dynamics.

A pilot experiment (EXP07) was therefore developed, following the overall design described in Figure 48. The primary objectives were to investigate whether CellTags could be successfully retrieved, and to assess how different fixation and storage conditions (−80 °C freezing versus 4 °C cold storage) affect the preservation of hashing labels and high-quality RNA for single-cell analysis. For this run, hIEOs derived from a single healthy colon biopsy (see Supplementary Methods, Section 1.6) were cultured as monolayers under the two experimental conditions described below:

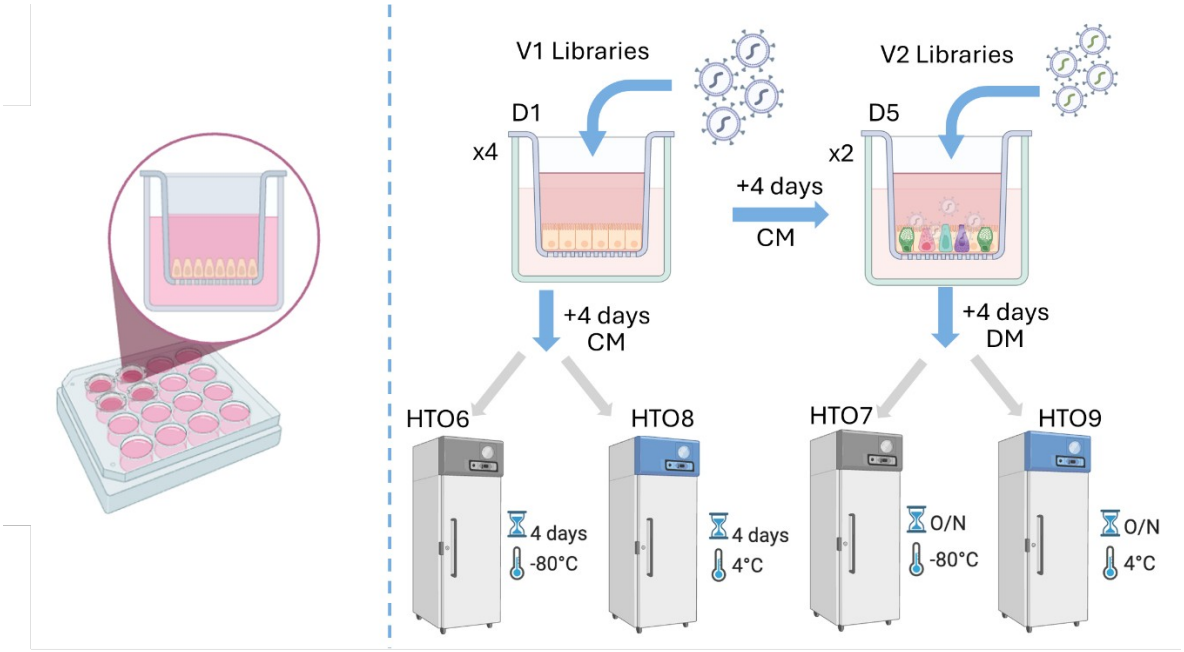
(1) fixation, antibody labelling, and freezing at −80 °C of two matched samples prior to the single-cell run – the first subjected to V1 tagging only and stored at −80 °C for 4 days (V1), and the other subjected to V1 tagging, then maintained in culture for an additional 3 days during which it underwent differentiation and V2 tagging, before being stored overnight at −80 °C (V1+V2; V3 tagging was not performed as part of this preliminary experiment);

(2) the same samples and procedure as in (1), but with storage carried out at 4 °C instead of −80 °C.

Of note, only the Fujii protocol was used for this and subsequent experiments, as it provided the most favourable outcomes in terms of rare cell type emergence, and overall cell recovery. Its use helped streamline the experimental design in these proof-of-concept studies.

Figure 48 - Experimental setup for Experiment 7: assessment of impact of storage conditions on CellTag libraries, cell hashing, and RNA content

A) Conformation of monolayer cultures within transwells *B) Experimental procedure*

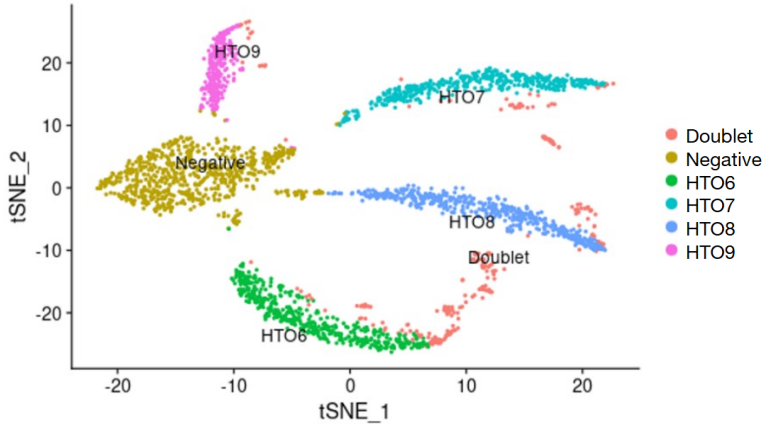


Legend: (A) Conformation of epithelial monolayers cultured on transwell inserts, used to enable polarised growth and apical accessibility. (B) Experimental setup used to evaluate the impact of storage conditions on cell quality in the context of sequential CellTag barcoding. On Day 1 (D1), 4 monolayer replicates were transduced with V1 CellTag libraries in conditioning medium (CM), followed by incubation for 4 days. On Day 5 (D5), two replicates were methanol-fixed and stored (either at 4 °C or –80 °C, for 4 days) and the other two introduced in Fujii differentiation medium (DM) and transduced with V2 CellTag libraries. These cells were incubated for an additional 4 days, after which they were methanol-fixed and stored either at 4 °C or –80 °C overnight (O/N). Each sample was then hashed (HTO6-9) to allow multiplexed scRNA-seq. This setup allowed comparison of storage temperature and timing effects on barcode recovery and transcriptomic integrity.

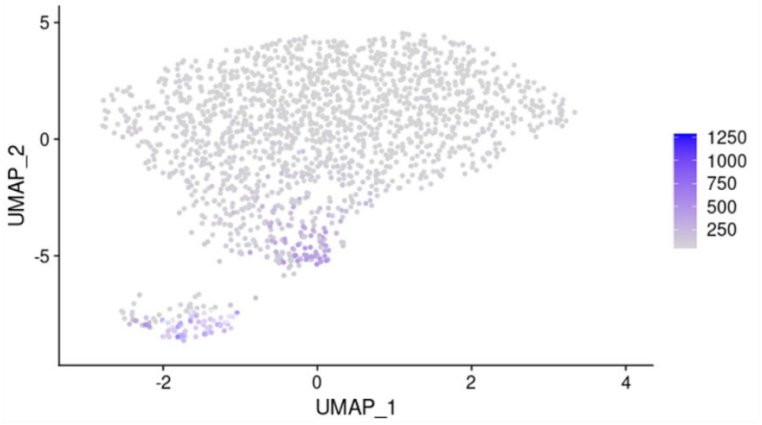
Cell hashing remained detectable in both storage conditions, as shown in Figure 49A. However, neither method preserved RNA integrity sufficiently for downstream library construction, rendering these samples unsuitable for further transcriptomic analysis (Figure 49B). Notably, there was no meaningful difference in RNA quality between the −80 °C and 4 °C conditions (data not shown). Despite suboptimal RNA preservation, lentiviral transduction was successfully achieved (Figure 49C–D), with high transduction efficiencies ranging from 50% to 90% across unsupervised cell clusters. Notably, this result shows that the presence of CellTags did not lead to the formation of distinct tag-exclusive clusters, indicating that tagged cells remained transcriptionally similar to their untagged counterparts. This suggests that the tagging process did not introduce substantial transcriptional drifts that could compromise cell identity, thereby confirming the feasibility of lentiviral delivery in hIEO cultures and supporting the use of this approach in subsequent experiments. However, these findings also emphasise the need to optimise viral load, multiplexing strategies, and sample handling protocols to enhance RNA quality and maintain cell viability.

Figure 49 - Demultiplexing and quality control of a 5-sample pooled scRNA-seq experiment using HTO hashing.

A) tSNE plot of HTO demultiplexed samples

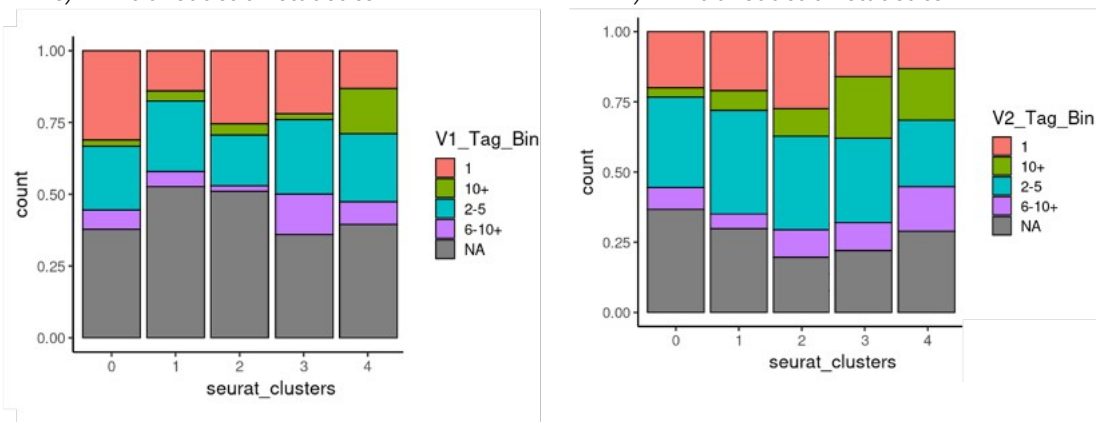


B) UMAP plot showing distribution of detected gene features per cell



C) V1 transduction statistics

D) V2 transduction statistics



Legend: (A) Cell hashing-based demultiplexing was successful, as evidenced by the distinct assignment of hash identities. HTO6 and HTO7 correspond to V1- and V1+V2-tagged cells stored at −80°C, respectively, and HTO8 and HTO9 to the same conditions but stored at 4°C. (B) UMAP projection showing the distribution of unique detected genes per cell (nFeature_RNA) and clustering before quality control filtering. The data exhibit poor transcriptomic quality, with low gene recovery and minimal separation between clusters — likely reflecting cytotoxic effects or transcriptional suppression due to the high viral titer used during transduction. (C) Proportion of cells in each cluster containing 1, 2–5, 6–10, or more than 10 CellTagV1 sequences. (D) Proportion of cells in each cluster containing 1, 2–5, 6–10, or more than 10 CellTagV2 sequences. These distributions reflect the clonal complexity and transduction efficiency

per cluster, offering insight into how distinct lineages were marked at different timepoints during organoid differentiation.

4.3.10 Pilot Experiment 2: Assessing the Impact of Hashing Conditions and High Virus Titer on scRNA-seq Data Quality

Despite the inability to fix or store samples for subsequent bulk analysis, the design of this experiment was deliberately structured to ensure that meaningful insights could still be obtained even if only tagged cells at the final timepoint could be analysed. For instance, if a given cell type consistently emerges only following a specific perturbation or exclusively within cells marked by a particular CellTag combination, this supports the notion that it arose from temporally earlier populations exposed to that condition. This approach poses a different challenge, which requires evaluating cell transduction efficiency without relying on storage of preliminary cell fractions. It was essential to ensure that all tagged populations would remain viable and retain sufficient RNA integrity throughout the full experimental duration (10 days). Additionally, the persistence of tag detectability was critical — for example, verifying whether a tag introduced on day 1 (CellTag V1) would still be reliably detectable on day 10. This required not only robust cell viability and transcriptional stability but also consistent tag retention across temporal windows.

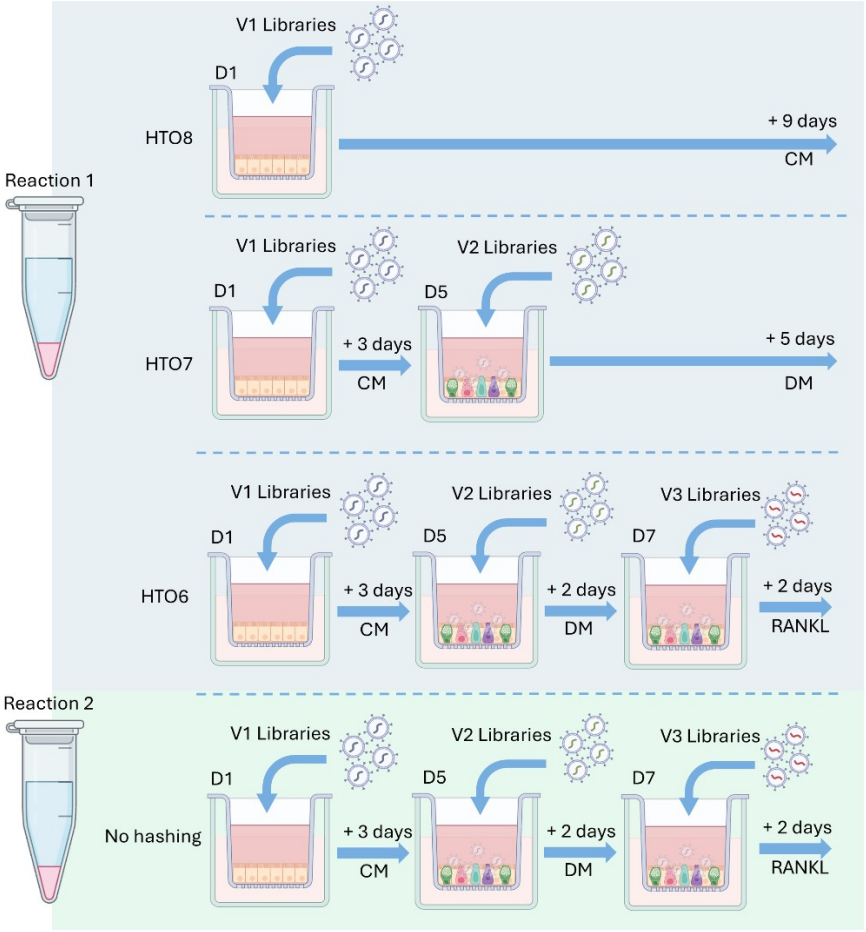
To address this, two lineage tracing test conditions were established using parallel experimental groups to assess temporal tagging under differentiation-induced perturbation (EXP08, as depicted in Figure 50). In the first group (Reaction 1), three identical hIEO samples were initially exposed to CM and spinfected with CellTag V1 lentiviruses. The first sample remained in CM throughout the 10-day experiment. On day 5, the second and third samples were transitioned to differentiation medium (DM) and spinfected with CellTag V2 lentiviral suspension. The second sample was then maintained in DM for the remainder of the experiment. On day 8, the third sample was transferred to freshly prepared DM supplemented

with RANKL and spinfected with CellTag V3 lentiviral suspension, remaining in DM+RANKL for the final two days. At the conclusion of the experiment, all three samples were antibody-hashed, pooled, and subjected to single-cell analysis via the 10x Genomics platform.

In the second group (Reaction 2), an identical hIEO sample that sequentially underwent exposure to CM, DM, and DM+RANKL — with corresponding spinfection using CellTag libraries V1, V2, and V3 (i.e. a technical replicate of sample 3 in Reaction 1) – was sequenced independently. This served as a contingency in the event that multiplexing, combined with reduced cell recovery per condition, resulted in insufficient CellTag representation.

Of note, only RANKL stimulation was used in this experiment, as IL13/IL4 treatment led to less cell type-specific priming—characterised by a broad expansion of secretory lineages—and failed to induce the emergence of Best4 Cells (as further detailed in 5.2.7).

Figure 50 - Experimental setup for Experiment 8: assessing suitability of sequential CellTag barcoding and condition-specific stimulation across multiplexed reactions



Legend: Schematic representation of the experimental timeline for sequential CellTag barcoding under different stimulation conditions. Four monolayer samples were included across two pooled single-cell RNA-seq reactions. In Reaction 1, hashed samples HTO8, HTO7, and HTO6 received V1 CellTags on Day 1 (D1) under conditioning medium (CM). HTO8 remained unstimulated for 9 days in CM. HTO7 received V2 tags on Day 5 and was switched to differentiation medium (DM) for 4 days. HTO6 received V2 tags on Day 5, followed by DM for 2 days and then V3 tags with RANKL for a final 2 days. In Reaction 2, an un-hashed sample followed the same triple-tag protocol as HTO6 (D1: V1 + CM; D5: V2 + DM; D7: V3 + RANKL) and served as a validation for the impact of stimulation and tagging without multiplexing. This design enabled lineage reconstruction and condition-specific fate mapping across both pooled and unpooled conditions.

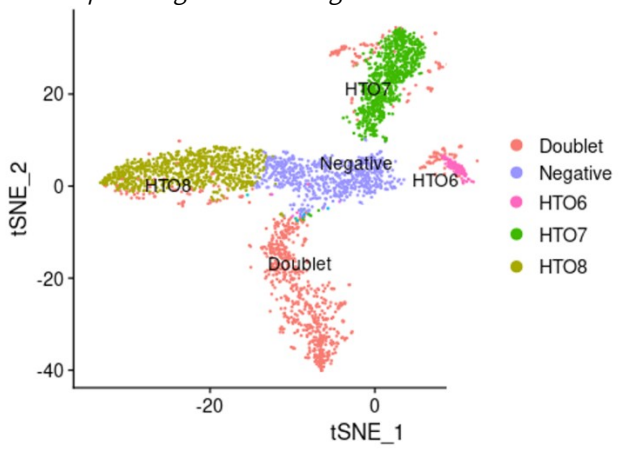
As anticipated, the high viral load led to substantial cell death and low overall cell recovery (Figure 51A), likely due to cellular toxicity and stress responses triggered by excessive lentiviral particle exposure, including activation of innate immune pathways and cellular toxicity, as previously described in lentiviral transduction models[347,348]. Nonetheless, transfection efficiency was notably high, reaching >75% fraction of tagged cells per cluster (Figure 51B-C). Comparisons between Reaction 1 (multiplexed samples) and Reaction 2 (individually run and sequenced V1+V2+V3 sample) revealed that the number of CellTags recovered per cell or cluster was highly similar across both groups, suggesting that the multiplexing approach did not negatively impact CellTag detectability. Importantly, multiplexing provided clear advantages, yielding higher overall cell recovery and more comprehensive information across conditions, while substantially reducing the cost and resource burden associated with separate sequencing runs. These findings support the feasibility of multiplexed single-cell lineage tracing in hIEO systems, demonstrating that it is both an efficient and cost-effective strategy without compromising tag resolution or data quality.

Although the majority of cells contained at least one detectable CellTag, relatively few were clonally tagged - defined as the presence of identical tag combinations across multiple cells, thus indicative of a shared lineage origin. This distinction is critical: the presence of a single tag confirms successful viral transduction, but clonal tagging reflects the expansion of a progenitor cell that has passed its unique combination of tags to its progeny over time. In this experiment, the low frequency of clonal tags suggests that while transduction efficiency was high, only a limited number of transduced cells underwent sufficient proliferation to generate detectable clonal populations. This outcome likely reflects limitations in cell recovery and potentially suboptimal sensitivity in CellTag detection, underscoring the importance of optimising both cell infections and tag amplification to enhance lineage resolution in future studies.

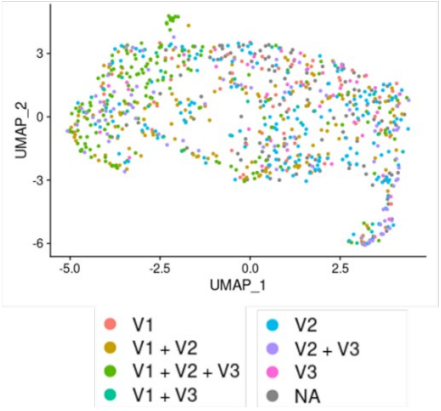
Nevertheless, despite reduced cell counts, several clones could still be reconstructed, with some exhibiting multi-stage branching patterns across tagging timepoints, indicative of functional clonal behaviour. Overall, while the tagging strategy was effective, the quality of the single-cell transcriptomic data remained suboptimal. Importantly, as tag-specific enrichment was not performed during library preparation, these results suggest that incorporating a targeted PCR amplification step to generate a separate tag-enriched library could substantially enhance tag recovery in future experiments, a strategy that was subsequently implemented.

Figure 51 - Comparison of lineage tracing outcomes between multiplexed and single-sample CellTag experiments

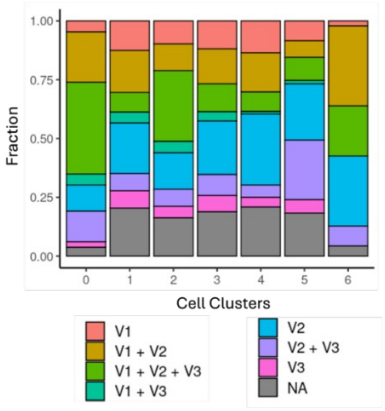
A) Demultiplexing of pooled sample using HTO labelling



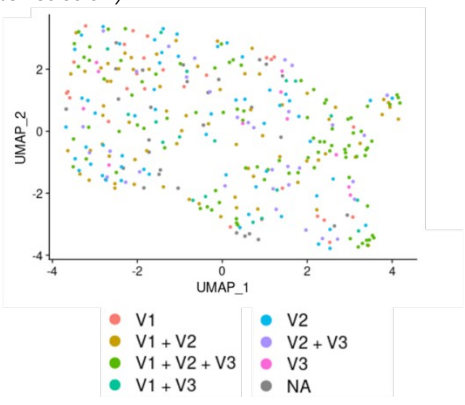
B) CellTag history across timepoints (multiplexed reaction)



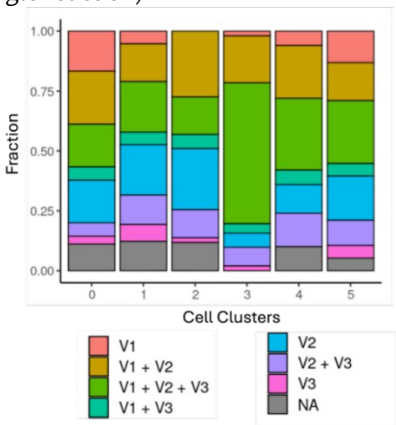
C) CellTag composition across cell types (multiplexed reaction)



D) CellTag history across timepoints (single reaction)



E) CellTag composition across cell types (single reaction)



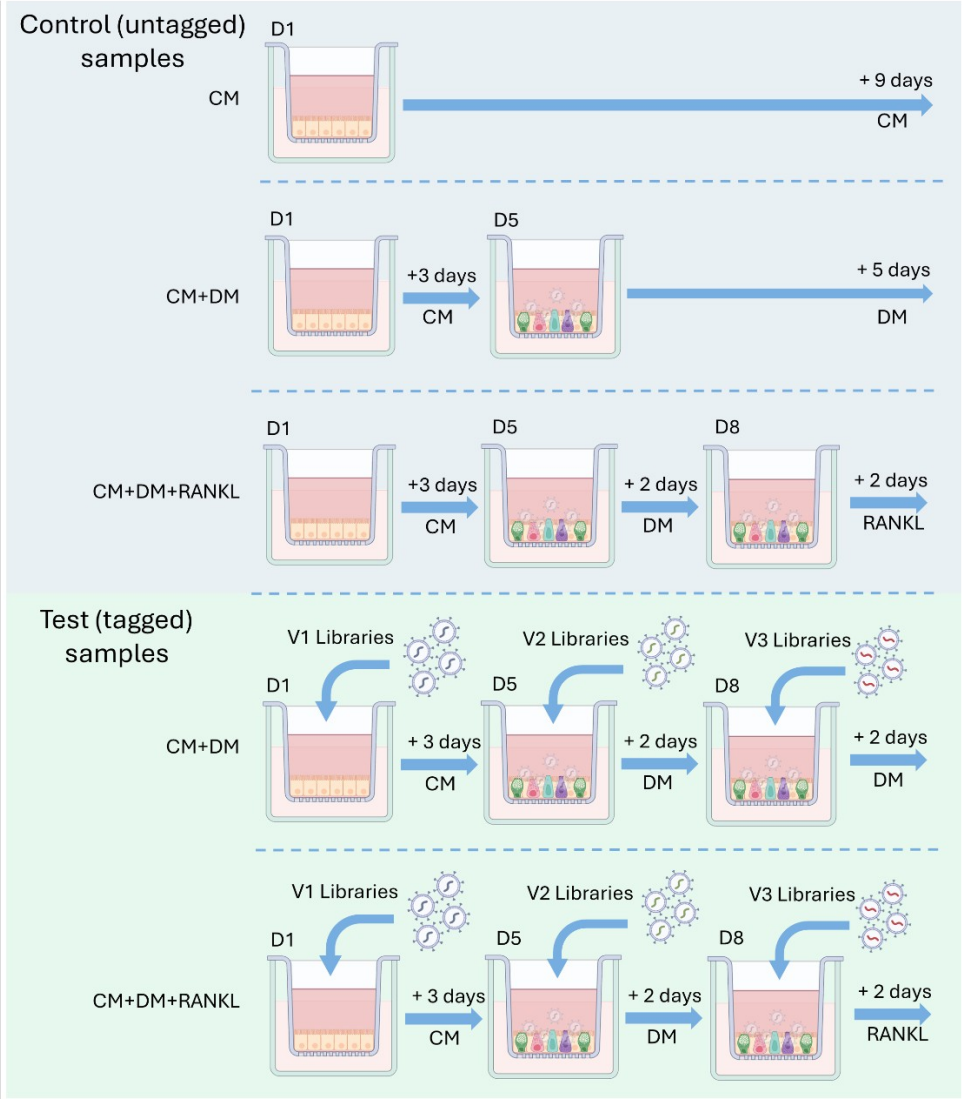
4.3.11 Proof-of-Concept Experiment: Multiplexed Lineage Tracing Platform in hIEOs

Based on the results from the previous experiments, a final proof-of-concept experiment was developed (with adjustments to viral titers; Supplementary figure S26) to enable temporal characterization of organoid differentiation, allowing lineage trajectories to be inferred based on CellTag inheritance patterns. The experimental design is shown in Figure 52. For this experiment (EXP08), five experimental conditions were developed for 1 healthy colon sample (Patient 20; see Supplementary Methods, Section 1.6), comprising three control and two test conditions:

1. CM only (control 1): cells maintained in CM for 10 days.
2. CM + DM (control 2): CM for 4 days, followed by 6 days in DM.
3. CM + DM + RANKL (control 3): CM for 4 days, DM for 3 days, followed by RANKL stimulation in DM for 3 days.
4. CM + DM (test 1): sequential V1, V2, and V3 library delivery at Day 1, Day 5, and Day 8, respectively, with media transition from CM to DM at Day 5.
5. CM + DM + RANKL (test 2): same barcoding schedule as above, but with RANKL added to DM at Day 8 (coinciding with the final 2-day V3 window).

All stepwise instructions and modifications of the CellTagging protocol are outlined in the Methods (Section 2.2.7) and in Supplementary Methods (Section 1.5).

Figure 52 - Experimental design for Experiment 9: sequential CellTag barcoding and stimulation conditions in control and test monolayer samples



Legend: Schematic representation of the experimental design for control (untagged) and test (CellTag-tagged) samples used in the temporal lineage tracing study. Monolayer hIEOs were cultured under three differentiation conditions: CM only, CM→DM, and CM→DM→RANKL. Control samples (top panel) underwent identical media transitions without CellTag lentiviral exposure, serving as unperturbed references for transcriptomic analysis. Test samples (bottom panel) were sequentially transduced with V1, V2, and V3 CellTag libraries at defined timepoints (D1, D5, D8), corresponding to transitions between media conditions. This approach enabled temporal barcoding of epithelial cells at key differentiation stages, allowing reconstruction of clonal dynamics and fate trajectories in response to environmental cues.

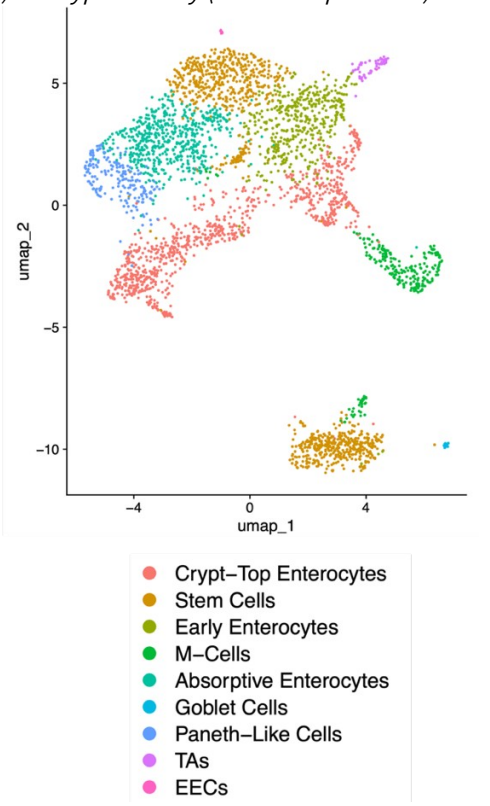
Following filtering, a total of 5,448 high-quality cells were retained. However, a substantial fraction of negative-hashed cells (2,128 cells) was observed, 91% of which carried valid CellTag barcodes (Supplementary Table S2, Supplementary Figure S29). These cells were neglected as

they could not be traced back to their original samples despite high RNA quality, with only 3,320 cells considered going forward. This pattern indicates that high viral titer or cumulative tagging may reduce hashing efficiency. Nevertheless, tagging coverage and barcode recovery were sufficient for clonal reconstruction and downstream analyses (Figure 53). The platform was shown to be highly scalable and potentially compatible with further multiplexing. For example, simultaneous CellTag tracing across three media conditions (CM, DM, and RANKL stimulation) revealed condition-specific clonal contributions to distinct cell types, notably M-cells and Paneth-like cells (Figure 53A-B), underscoring the method's capacity to detect rare or transient lineage transitions within the organoid model.

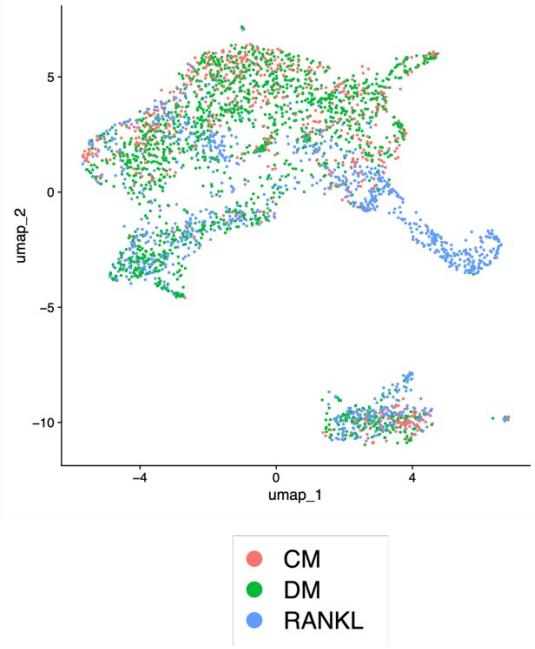
Well-resolved clusters and lineage trajectories were obtained across the combined dataset. Overlaying clonal tag identity onto UMAP space demonstrated the efficiency and specificity of the lineage tracing strategy. Quantitative analysis confirmed successful barcode recovery, clonal expansion, and preservation of key epithelial cell states, including absorptive, secretory, and progenitor subtypes (Figure 53C-D).

Figure 53 - Multidimensional analysis of epithelial identity, culture condition, and CellTag history in human intestinal epithelial organoids

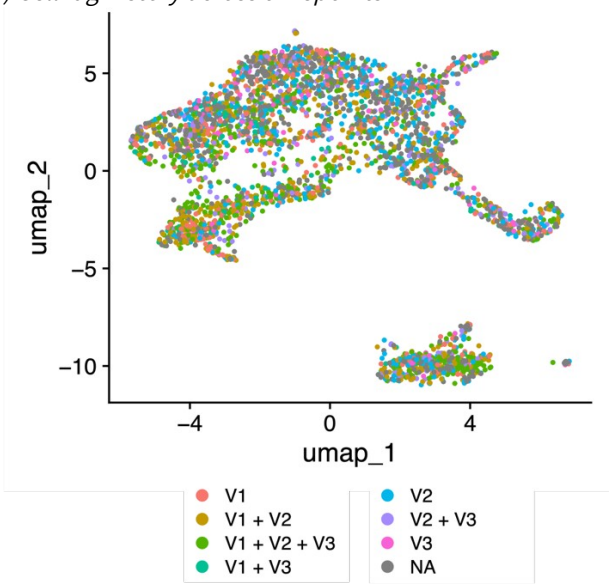
A) Cell type identity (scRNA-seq clusters)



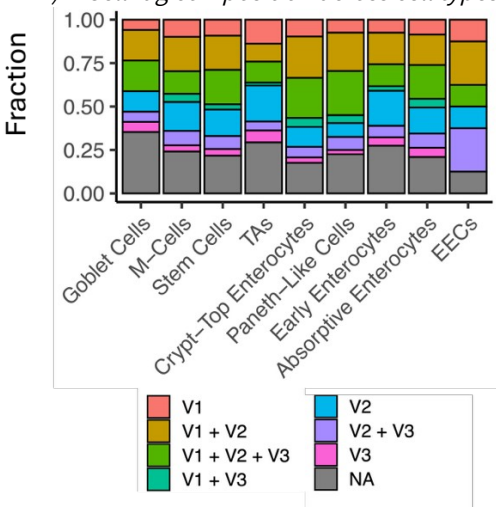
B) Culture condition (CM, DM, RANKL)



C) CellTag history across timepoints



E) CellTag composition across cell types

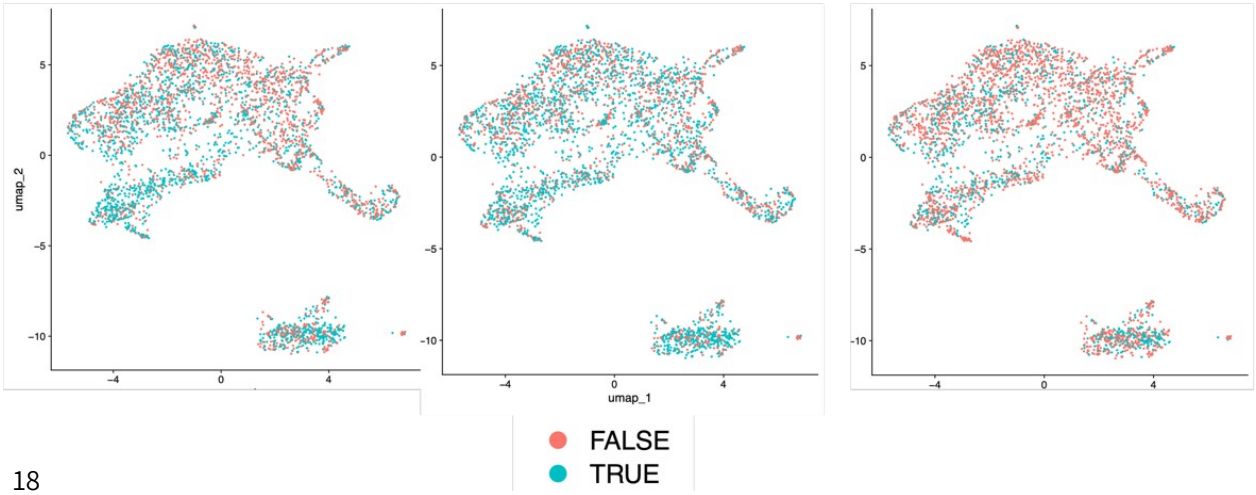


4.3.12 Distribution of Tagged Cells Across CellTag Generations

To assess the efficiency and distribution of lentiviral tagging over time, UMAP projections were used to visualise cells labelled with each generation of CellTags (V1–V3), as shown in Figure 54. V1 labelling was broadly distributed across the epithelial landscape, reflecting early accessibility and lentiviral uptake across multiple progenitor and probably highly plastic populations, since these tags were introduced during the initial culture in CM. V2 tagging was applied at the transition to DM, and while tagging was somehow broadly successful, differences in spatial distribution relative to V1 suggest a shift in accessibility and viral integration efficiency as cells underwent differentiation. Finally, V3 labelling, introduced at the final timepoint (RANKL stimulation), displayed more restricted coverage overall, consistent with reduced infection efficiency or altered viral accessibility in terminally differentiated cells. Overall, these results confirm that each CellTag generation successfully labelled a representative subset of cells at each time point, with tagging density and spatial distribution varying according to the timing of lentiviral exposure and the underlying cell state at each stage.

Figure 54 - Distribution of tagged cells across CellTag generations V1–V3

A) V1 tagged cells *B) V2 tagged cells* *C) V3 tagged cells*



Legend: UMAP visualisation showing the distribution of tagged (TRUE, blue) and untagged (FALSE, red) cells for each CellTag generation (V1–V3). Each panel highlights the spatial distribution of successfully

labelled cells across the epithelial landscape. Differences in tagging density reflect the timing of lentiviral exposure and cellular accessibility at each stage.

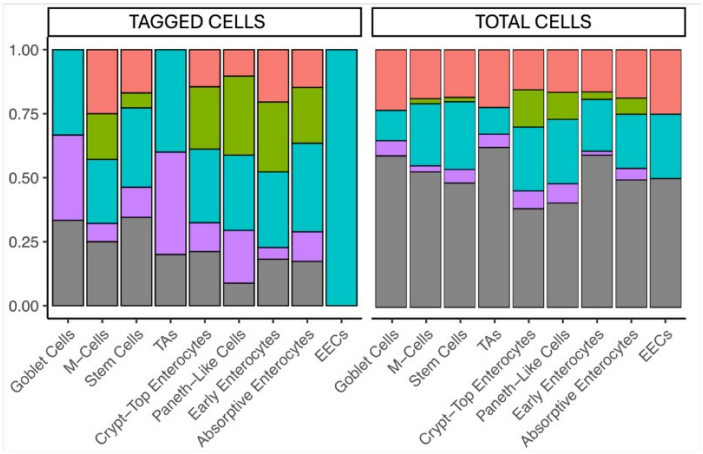
4.3.13 Clonal Complexity Across Epithelial Cell Types in hIEOs

To evaluate the clonal diversity achieved through CellTag lineage tracing, the number of unique clonal identities per cell type was quantified across the three tagging timepoints (V1–V3). Figure 55 depicts the distribution of clone counts within each epithelial subtype, separated into the subset of tagged cells (test samples 1 and 2 - left panels) and total cells (test samples together with control samples 1, 2 and 3 - right panels), for each tagging generation. Most clusters contained cells with 2–10+ clonal tags, indicating both high infection efficiency and active proliferation and clonal expansion from early progenitors. The overall ratio of cells taking up CellTags is shown to decrease from V1 to V3, as previously mentioned. This pattern may be due to 1) reduced proliferation time between cell tagging and the end of the experiment, 2) reduced proliferation resulting from media conditions leading to increasing maturation and differentiation, and 3) increased cell stress and toxicity, and a probably larger number of tagged cells become apoptotic. All cell clusters showed permeability to infection, with differences in the proportions of tagged vs untagged cells likely reflecting cluster abundance – for example, as EECs are less abundant, the resulting tagged fraction is higher but with lower variation in terms of clonal complexity.

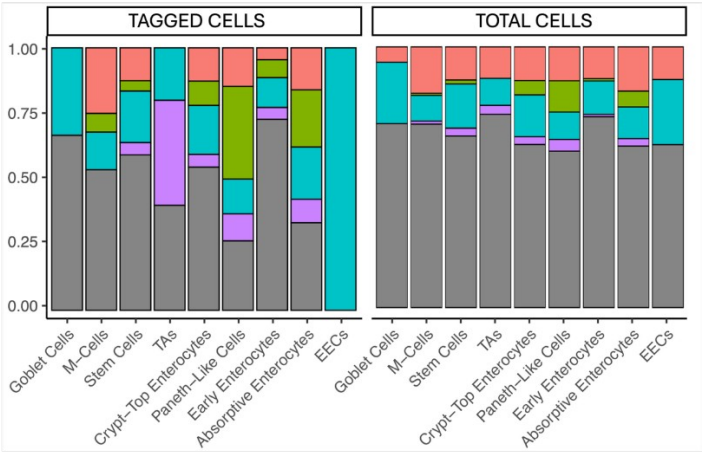
Collectively, these results demonstrate that clonal diversity is highest when tags are introduced early, and that different epithelial subtypes exhibit distinct patterns of clonal representation depending on the timing of tagging. This supports the temporal resolution of lineage dynamics and highlights the capacity of this system to capture clonal complexity across heterogeneous cell states.

Figure 55 - Clonal complexity across epithelial cell types in hIEOs

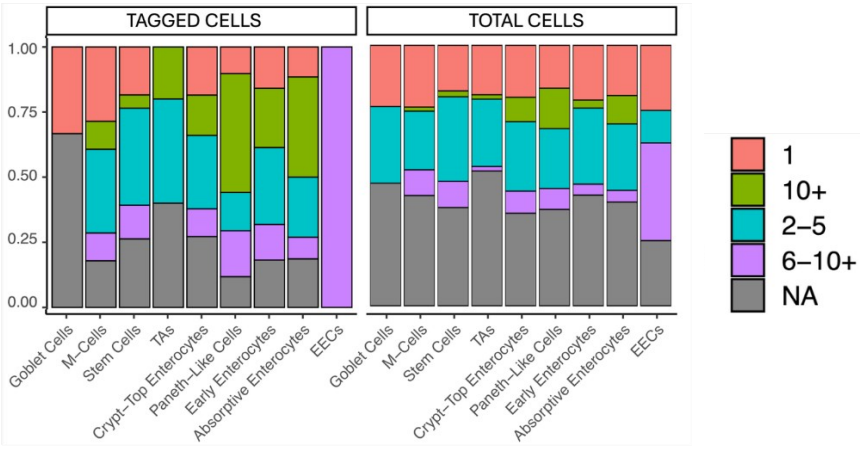
A) V1 tags



B) V2 tags



C) V3 tags



Legend: The color coding shows the number of distinct clones detected for each cell cluster at each time point (V1-V3).

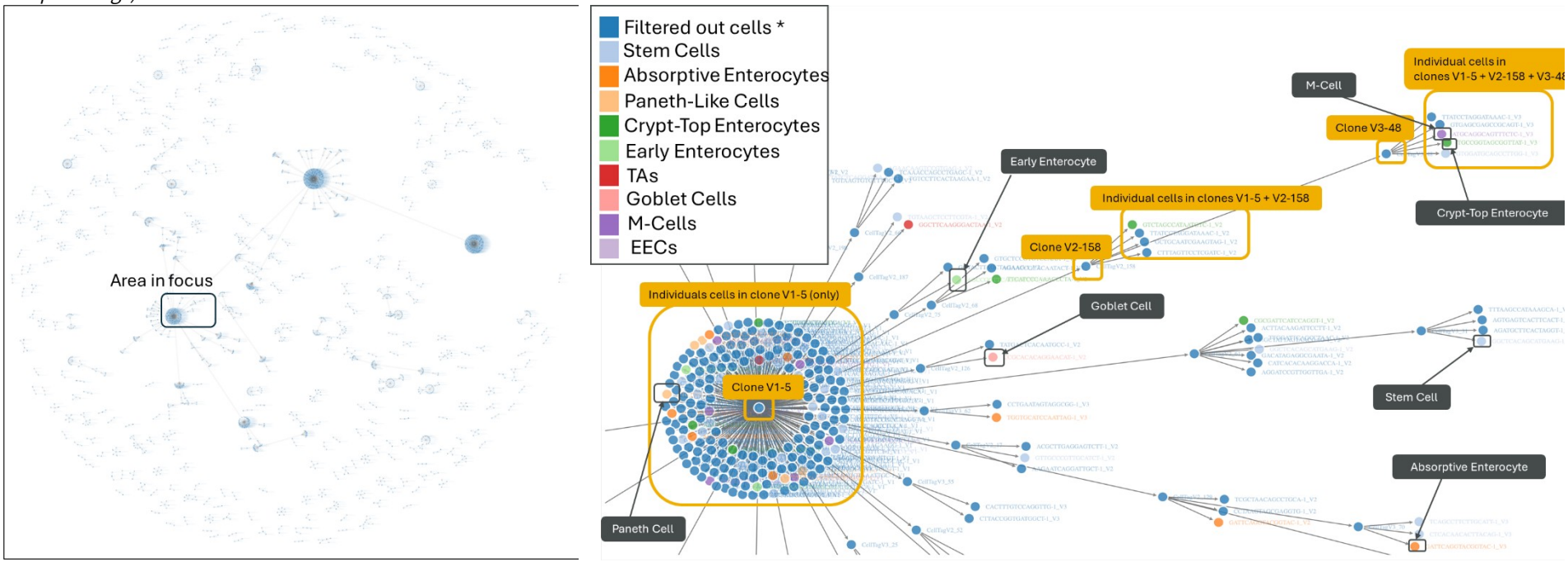
4.3.14 Network-Based Lineage Reconstruction Reveals Temporal Clonal Relationships

To resolve clonal trajectories across time, a network-based approach was used to visualise lineage relationships between individual cells and their inferred clone memberships across the three temporal CellTag generations (V1, V2, V3). As shown in Figure 56, the complete network (panel A) depicts all cells and their respective connections, with edges linking each cell to its assigned clone, and each clone to the respective inferred “parent” clone at the previous timepoint. Cell types are color-coded based on transcriptomic identity; cells lacking clear lineage assignments or with incomplete tag information were excluded and are labelled as “Filtered out cells”. An area of interest within the network is highlighted and shown in greater detail in panel B. The zoomed-in region illustrates how distinct cell types can arise from a shared ancestral clone. For example, the purple-coloured M-cell located at the top right (tag ATGGCAGGATTTCTC) belongs to clone V3-48, which is derived from clone V2-158, and originally from V1-5. The same is true for all other cells in clone V3-48. This progressive clonal structure demonstrates how lineage tracing across multiple tagging events can capture dynamic differentiation outcomes. Here, a diverse set of M-Cells, Goblet Cells, Crypt-Top Enterocytes, Paneth Cells, as well as Stem Cells and TAs, harvested simultaneously at the same terminal timepoint, all share the same lineage origin, diverging at later stages and in different timepoints in response to specific environmental cues.

This network analysis provides a visual and analytical framework for tracing clonal branching and fate diversification in human organoid systems. The method highlights the utility of temporal CellTagging in uncovering lineage hierarchies and differentiation patterns across complex epithelial landscapes. While the foundational framework is robust, further refinement is warranted. Optimisations in viral delivery, tag enrichment strategies, and computational tools for lineage resolution would enhance accuracy, sensitivity, and interpretability. These

improvements are beyond the scope of the current thesis, but present clear opportunities for future methodological advancement.

Figure 56 - Network visualisation of lineage relationships across temporal CellTag barcodes in hIEOs
A) Network graph of all individual cells and clones (across all temporal tags) *B) Network graph, zoomed in on area in focus*



*Legend: Network diagrams representing clonal relationships between individual cells inferred from CellTag lineage tracing across three temporal barcoding stages (V1, V2, V3). (A) The full network shows all cells and inferred clone assignments, with an area of interest indicated for further inspection in panel B. (B) Zoomed-in region from panel A, coloured by cell type (as inferred from final transcriptomic state). Arrows connect each cell to the clone it maps to in the previous timepoint. Representative cells are labelled for clarity. In this example, the M-cell at the top right corner (tag ATGCAGCAGTTTCTC) is part of clone V3-48, which itself is derived from clone V2-158, and originally from clone V1-5. All cells in V3-48 are also in V2-158, and all cells in V2-158 are part of V1-5 (even if not shown as members of V2-158, due to either belonging to other V2 clones or not forming part of any V2 clone). This progressive clonal relationship illustrates how different fates (e.g., M-cells, Goblet Cells, Crypt-Top Enterocytes) can emerge from a shared ancestor, depending on time and condition. * Cells with low-confidence lineage assignments or incomplete tag information were excluded ("Filtered out cells").*

4.3.15 Exploring the Feasibility of a Multiome Single Cell Sequencing Lineage Tracing Approach

The established lineage tracing pipeline using CellTag technology in hIEOs was implemented within a multiplexed, multiome sequencing framework, building from the insights garnered in Section 4.2 (namely employing higher depth scRNAseq). This approach would enable the simultaneous capture of transcriptomic profiles (RNA), chromatin accessibility (ATAC), and clonal lineage information at single-cell resolution across multiple experimental conditions assayed simultaneously. Integration of CellTagging with multiome data could facilitate the reconstruction of clonal relationships, mapping of lineage progression, and characterisation of transcriptional states within a complex, physiologically relevant human epithelial system.

Although high-quality ATAC-seq data could be generated as part of the multiome workflow, the total number of reads attributed to these samples reached only approximately 50% of the estimated depth required for optimal resolution (even when using NovaSeq). As a result, the chromatin accessibility signal was insufficient to produce informative clustering. Instead of enhancing resolution, the ATAC component obscured cell-type distinctions and produced diffuse clusters, often merging biologically distinct populations into a single chromatin-based identity (see Supplementary Figure S30). As a result, no further integrative analysis could be undertaken, as further optimisation of scMultiome-seq combined with CellTagging could not be pursued within the timelines of this DPhil degree. Future experiments incorporating ATAC modality at higher sequencing depth would be necessary to fully leverage chromatin accessibility in resolving lineage dynamics and regulatory states.

4.3.16 Interpretation of Findings

This section has detailed the conceptualisation, optimisation, and preliminary implementation of a scalable lineage tracing framework in hIEOs, leveraging CellTag lentiviral barcoding in

monolayer cultures to resolve differentiation dynamics at single-cell resolution. The experimental rationale stemmed from the need to capture rare or transient epithelial states — particularly those elusive to conventional snapshot approaches — and to reconstruct cell fate trajectories over time in a physiologically relevant system.

Adapting the CellTagging protocol to hIEOs required extensive methodological development, including (1) transitioning from 3D to 2D culture formats to overcome transduction inefficiencies inherent to Matrigel-embedded organoids; (2) generating and validating high-diversity V1–V3 CellTag libraries; (3) optimising lentiviral infection through ultracentrifugation-based viral concentration; and (4) implementing antibody-based sample multiplexing and a tailored bioinformatics pipeline for tag recovery, clone assignment, and lineage network reconstruction.

Proof-of-concept experiments confirmed that sequential tagging allowed the recovery of clonal lineages with temporally structured barcodes, and demonstrated that the resulting CellTag identities could be effectively overlaid onto single-cell transcriptomic data to infer lineage bifurcations. Key epithelial lineages — including absorptive and secretory progenitors, as well as Goblet Cell, M-Cell, and Paneth-Cell populations — were successfully reconstructed across timepoints, underscoring the utility of this approach in resolving epithelial fate transitions. Network visualisation revealed biologically plausible trajectories, such as the derivation of M-cells from absorptive-biased intermediates, and the shared ancestry of crypt-top and secretory lineages.

Nevertheless, several critical technical limitations constrained the overall resolution and scale of the platform. Most notably, the majority of tagged cells were lost during QC filtering due to low-quality RNA and/or hashing read recovery, resulting in a mismatch between high-quality cells from a transcriptomic perspective vs clonally informative cells. This discrepancy likely reflects the cytotoxic burden imposed by high viral loads, leading to a trade-off between transduction

efficiency and transcriptomic integrity. Tagged cells were enriched for stem and progenitor states, suggesting that heavily infected cells may have impaired differentiation capacity or increased susceptibility to stress-induced apoptosis.

From a computational standpoint, existing lineage tracing pipelines required extensive adaptation to accommodate the specificities of organoid cultures, the unique structure of CellTag data, and the multiplexed experimental design. Nonetheless, further adaptations are likely required to fully accommodate integration of the CellTag framework within hIEOs. For example, cells harbouring repeated tags across multiple timepoints were often excluded from clone assignment due to rigid thresholds (e.g. requiring ≥2 unique tags), despite evidence of clonal propagation. Likewise, network-based lineage reconstructions were challenged by sparse tagging or partial tag dropout, and further improvements in clone calling, barcode enrichment, and visualisation tools will be necessary to scale this approach for future studies.

Despite these constraints, this represents the first demonstration of a fully multiplexed, single-cell lineage tracing framework applied to hIEOs. The approach enabled simultaneous capture of culture condition, transcriptomic state, and clonal identity across thousands of epithelial cells, opening the door to high-resolution modelling of lineage behaviour in human systems. The work presented here lays the foundation for more robust future applications — for example, combining improved tag enrichment protocols (e.g. targeted tag capture) with optimised viral titration strategies and higher-throughput designs (e.g. condition-per-channel or patient-per-channel layouts) to minimise sample dropout and maximise clonal diversity.

Importantly, although a pilot scMultiome-seq experiment was included to explore the integration of chromatin accessibility with lineage information, the limited sequencing depth achieved — particularly for ATAC — precluded full resolution of regulatory trajectories. As such,

the potential of this dual-modality approach remains untapped and will require further investment in sequencing depth and protocol integration.

In conclusion, while the biological scalability of this approach has been successfully demonstrated, technical and temporal constraints precluded a more comprehensive exploration of epithelial differentiation and fate decisions using the current lineage tracing framework. This would require further optimisation cycles and follow-up experiments which could not be pursued within the temporal constraints underpinning this DPhil degree. Nonetheless, this developmental work offers critical insights into how rare cell populations — such as M-cells and Best4^{+} enterocytes — can be traced and modelled within human-derived organoid systems using temporal barcoding strategies. It provides a clear blueprint for the design, implementation, and refinement of lineage tracing experiments in physiologically relevant systems. The next section of this thesis will therefore shift from technological development to the biological exploration and validation of the experimental models developed throughout this work, using the optimised experimental conditions identified.

4.4

4.5 Chapter Summary

In this chapter, two complementary methodological frameworks were developed and applied to address the key limitations identified in earlier transcriptomic analyses of hIEOs. First, integrative multiomic profiling (RNA + ATAC) was implemented to capture both transcriptional output and epigenetic accessibility within the same cell. While chromatin accessibility data were constrained by sequencing depth, this approach nevertheless demonstrated the value of combining modalities to enhance resolution of transitional populations and to identify loci that are epigenetically primed even in the absence of detectable transcript expression. These findings provided initial evidence that elusive lineages, including Best4^{+} Cells, may be poised within absorptive compartments, highlighting the potential of multiomics to uncover hidden regulatory states.

Second, a transcriptome-compatible lineage tracing framework was established using sequential CellTag barcoding. Proof-of-concept experiments confirmed that temporal barcode inheritance could be reliably recovered in hIEO monolayers and linked to cell-type annotations at single-cell resolution. This enabled the reconstruction of clonal trajectories across multiple conditions, including RANKL stimulation, and revealed plausible lineage bifurcations from shared progenitor pools into absorptive, secretory, and rare epithelial fates. Despite technical constraints, including reduced cell viability under high viral load and suboptimal tag recovery, this work represents the first demonstration of scalable lineage tracing in hIEOs, providing a blueprint for future methodological refinement.

Together, these methodological advances extend the scope of hIEO-based single-cell analysis beyond descriptive profiling, offering mechanistic and temporal perspectives on epithelial fate decisions. They establish that rare or underrepresented populations can be probed not only

through transcriptional signatures but also through their regulatory context and clonal origin. The insights gained here set the stage for the next chapter, which shifts from methodological development to biological validation, assessing how well these experimental models recapitulate primary intestinal epithelium and how rare lineages, such as Best4^{+} cells, can be robustly defined and validated in a physiological context.

5 Mapping Context-Specific Signatures in hIEO Models

5.1 Introduction

The intestinal epithelium is a highly dynamic tissue that integrates spatial, developmental, and functional cues across a diverse range of cell types. Its architecture is shaped by intrinsic differentiation programmes and environmental stimuli, making it an ideal system for probing context-dependent cell fate decisions. Having established and benchmarked a robust experimental framework in previous chapters, this section transitions from protocol development to biological validation — using the optimised culture systems and multiplexed workflows to interrogate how epithelial identity is modulated across diverse conditions. These refined hIEO platforms are deployed to investigate whether transcriptional signatures observed *in vivo* are preserved or altered *in vitro*. Particular attention is given to retention of transcriptional signatures across different anatomical regions (TI vs colon), disease status (healthy vs UC inflamed) and developmental stages (fetal vs adult). Additionally, the effects of culture conformation (monolayer vs 3D) and exposure to extrinsic stimuli, including key intestinal cytokines such as RANKL, IL13 and IL4 are also evaluated.

Rather than focusing on mechanistic dissection, the analyses presented here offer broad validation of model fidelity and responsiveness. In earlier sections, rare epithelial subtypes were leveraged as sensitive indicators of differentiation potential. Here, global differential gene expression and ontology enrichment analyses are employed to assess how specific stimuli influence epithelial plasticity and phenotypic outcomes. Representative effects—such as RANKL-induced M-cell differentiation or IL13/IL4-driven shifts towards secretory lineage phenotypes—

are highlighted, while subsequent sections focus on broader patterns in lineage distribution and the preservation of regional identity.

To enhance interpretability, the multiplexed datasets were systematically subset to isolate the effect of individual variables. For example, when comparing tissue region (TI vs colon), analyses were stratified by media or stimulation condition to avoid confounding effects. Datasets were pre-processed independently and annotated following unsupervised clustering, using curated gene panels as detailed in the Appendix (Supplementary Methods, Section 1.3). Where no substantial differences were observed — such as comparable cellular compositions between fetal and adult unstimulated organoids — datasets were merged and batch-corrected to increase statistical power and streamline downstream analyses. By contrast, when a condition induced clear transcriptional divergence — as seen in RANKL-stimulated samples — these were analysed separately and excluded from unrelated comparisons. This selective analysis strategy — merging when appropriate, isolating when necessary — was applied consistently and is explicitly stated at the start of each section. It allowed biological comparisons to remain both robust and interpretable, despite the inherent complexity of multiplexed designs. Together, these analyses demonstrate that optimised hIEO culture systems — when combined with scalable multiplexing — can partially recapitulate key features of intestinal epithelial identity, including rare, region-specific, and immune-modulated lineages, across a wide range of experimental conditions.

5.2 Contextual Influences on Epithelial Organoid Identity Across Multiplexed Experiments

5.2.1 Location, Inflammation, and Developmental Signatures in House-Protocol Cultured hIEOs

The baseline condition for assessment of hIEO fidelity is the experimental setup described for EXP01 (Figure 2) – i.e. culture under the House protocol, a modified version of the standard organoid culture protocol described in Sato *et al.* (2011). The UMAP projections shown in Figure 3 (Section 3.2.3) depict the transcriptional landscape of hIEOs, with cells coloured by key experimental variables including developmental stage, anatomical region and inflammatory status, and media/stimulation condition. Since the comparison between CM and DM was examined in detail in Section 3.2.4, DM will be considered as the baseline control condition for downstream analyses.

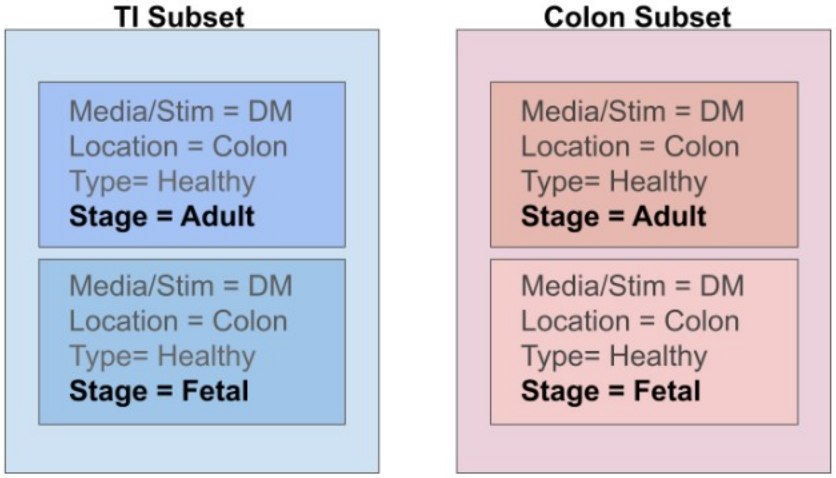
In Figure 3, clear clustering was observed by developmental origin (fetal vs adult), anatomical location (colon vs TI), and disease status (healthy vs inflamed), indicating that major *in vivo* signatures are at least partially preserved under standard culture conditions (i.e. House protocol). Furthermore, stimulation with NaB or RANKL introduced distinct transcriptional shifts, supporting the responsiveness of hIEOs to environmental perturbations. These initial patterns motivate a structured analysis of how regional identity, inflammatory signatures, and developmental cues are retained or modulated *in vitro*. The following sections examine each of these axes in turn.

5.2.2 Regional Identity Retention in House-Protocol Cultured hIEOs

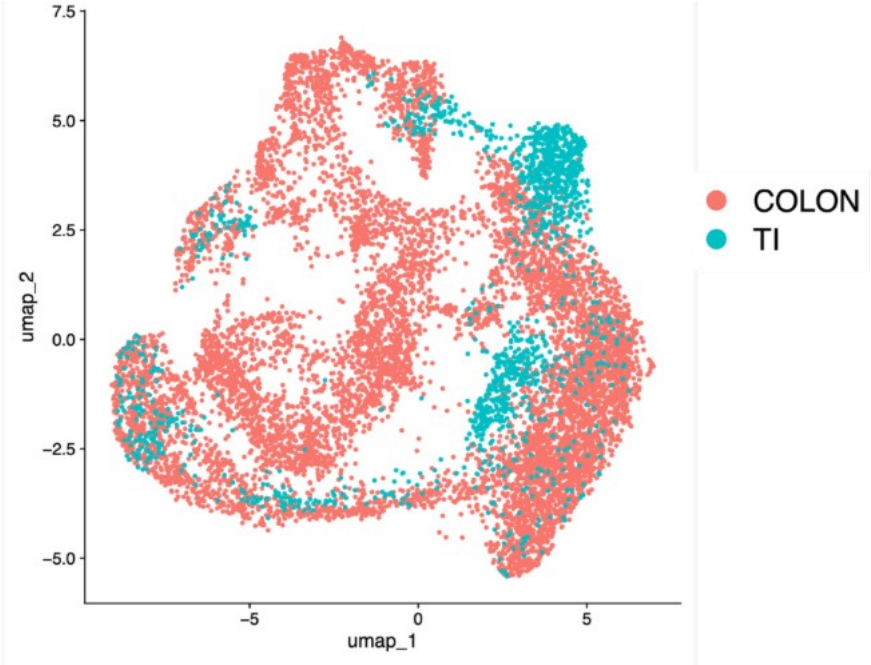
To assess the preservation of regional identity, TI and colon datasets were generated by combining healthy, unstimulated adult and fetal hIEOs for each anatomical site. Batch correction was applied to account for donor, pool, and developmental stage effects. Stimulation

conditions were excluded from this analysis, as they introduced strong and distinct cell identity signatures — particularly following RANKL exposure — that could not be effectively corrected without obscuring subtler, region-specific differences. Diseased states were also removed as they were only available in adult samples. This approach enabled a more accurate interrogation of transcriptional variation attributable specifically to anatomical location (Figure 57).

Figure 57 - Subsetting strategy and overall transcriptional profiles for regional identity analysis
A) Metadata structure for the TI and colon subsets



B) UMAP of epithelial cells by anatomical region

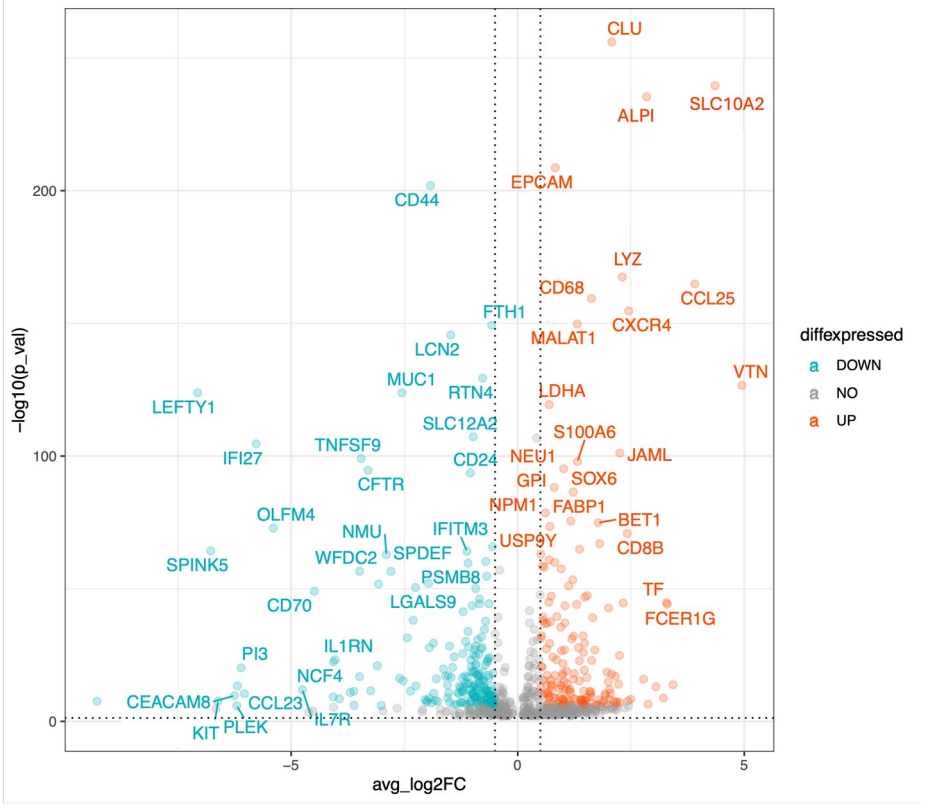


Legend: Overview of the experimental subset (top) and UMAP projection (bottom) of intestinal epithelial cells derived from healthy fetal and adult tissue cultured in differentiation medium (DM). The schematic outlines the sample metadata used to generate the UMAP, comprising matched colon and terminal ileum (TI) samples across both fetal and adult developmental stages. The UMAP visualisation (bottom) shows cells coloured by anatomical origin: colon (red) and TI (blue). Despite uniform culture conditions, cells cluster partially by region, indicating that regional identity is retained and contributes to transcriptional divergence in vitro. These findings highlight the importance of considering tissue-of-origin when interpreting epithelial differentiation and gene expression patterns in human intestinal organoid models.

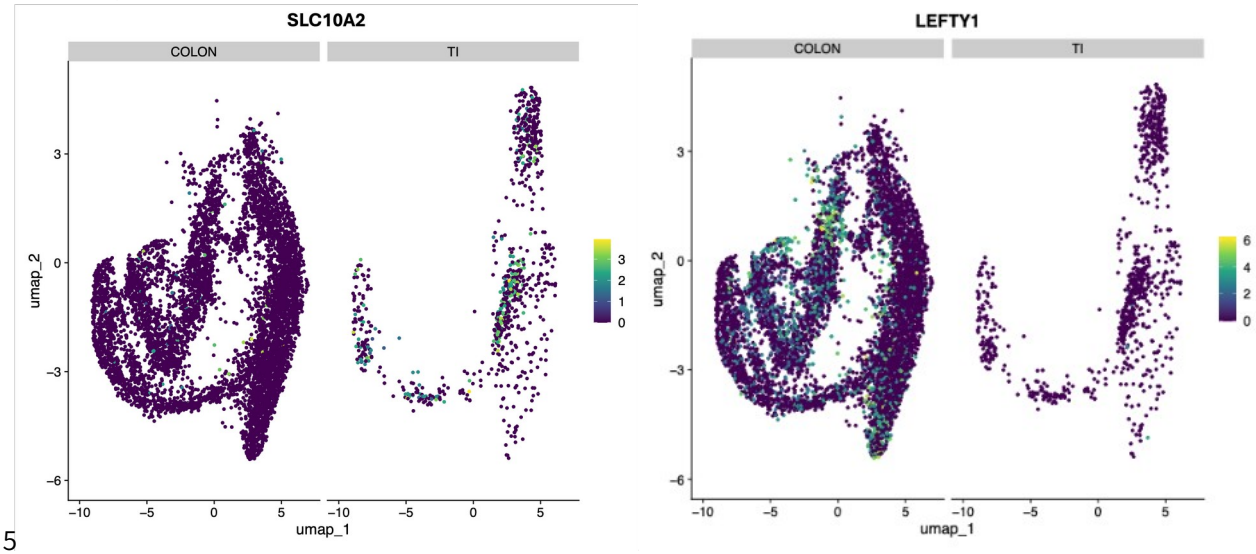
To assess regional identity *in vitro*, differential gene expression analysis was performed between the TI and colon subsets. Figure 58A illustrates transcriptomic differences, with genes upregulated in TI shown in orange, and those upregulated in the colon (i.e. downregulated in TI) shown in blue. Several well-established regional markers were recovered, validating the retention of location-specific signatures. For instance, *CCL25*, *SLC10A2*, and *LYZ* were enriched in TI-derived organoids, while *SPINK5*, *LEFTY1*, and *CFTR* were elevated in colon samples — consistent with well-established findings from multiple studies and single-cell datasets of primary colon and terminal ileum epithelium [22,38,41,42,241]. These findings were confirmed when assessing regional transcriptomic signatures for location-specific genes, such as *SLC10A2* and *LEFTY1* (Figure 58B). The recovery of known markers alongside a broader set of differentially expressed genes suggests that hIEOs preserve key aspects of their tissue-of-origin identity under standard culture conditions, supporting the use of unstimulated hIEOs for modelling region-specific intestinal epithelial biology.

Figure 58 - Regional gene expression differences between terminal ileum and colon in unstimulated hIEOs

A) Volcano plot of differentially expressed genes



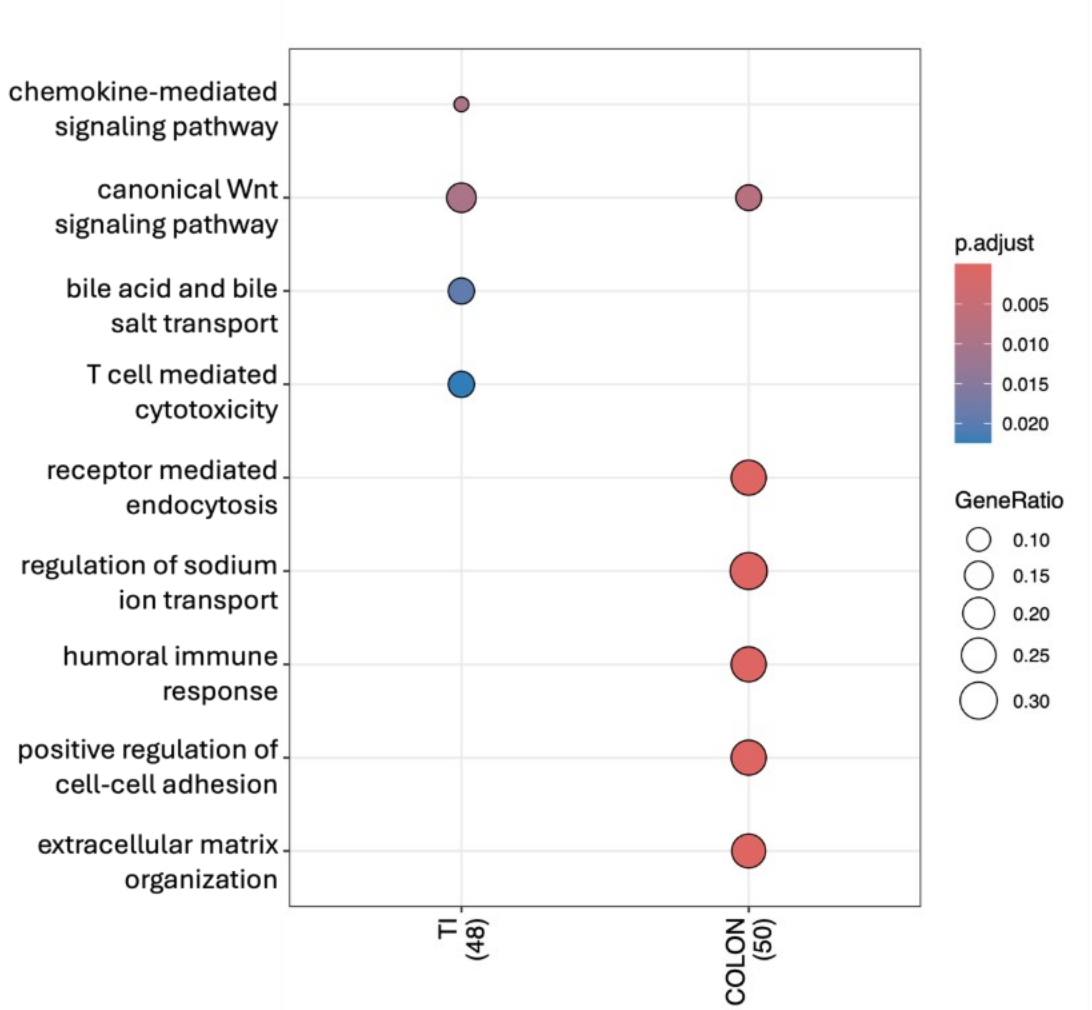
B) Regional expression patterns of SLC10A2 and LEFTY1



Legend: (A) Differential gene expression between terminal ileum (TI) and colon samples, with significantly upregulated genes in the colon highlighted in orange and those upregulated in the TI highlighted in blue. Genes not meeting the differential expression threshold are shown in grey. Notable regional markers include FOXD1 and SERPINA1 enriched in TI-derived samples, and LEFTY1, CCL25, and CEACAM8 enriched in colon-derived samples. (B) UMAP projections showing the expression of selected regional markers across cells from colon (left) and TI (right) samples. FOXD1 is predominantly expressed in colon cells, whereas LEFTY1 is enriched in the TI compartment.

GO analyses analysis further highlighted region-specific transcriptional programmes in hIEOs (Figure 59). While both TI and colon samples showed enrichment in canonical Wnt signalling, distinct functional pathways were uniquely associated with each region. Colon-derived hIEOs demonstrated increased expression of genes involved in epithelial maintenance and structural organisation, including pathways related to extracellular matrix organisation, cell–cell adhesion, and receptor-mediated endocytosis. By contrast, TI-derived organoids showed significant enrichment for pathways related to immune signalling and nutrient transport, including bile acid transport, T cell-mediated cytotoxicity, and chemokine signalling. These findings reinforce the notion that regionally derived hIEOs retain transcriptional signatures reflective of their tissue of origin, independent of developmental stage, supporting their application in modelling location-specific aspects of intestinal epithelial biology *in vitro*.

Figure 59 - GO term enrichment analysis of region-specific gene expression in adult intestinal epithelial organoids



Legend: Significantly enriched GO biological processes among genes differentially expressed between terminal ileum (TI) and colon-derived epithelial organoids (adult, healthy donors). Distinct regional enrichments are observed: TI-enriched genes are associated with developmental and signalling processes such as bile acid transport, T cell-mediated cytotoxicity, and canonical Wnt signalling, whereas colon-enriched genes are associated with functions like immune response, epithelial adhesion, endocytosis, and extracellular matrix organisation.

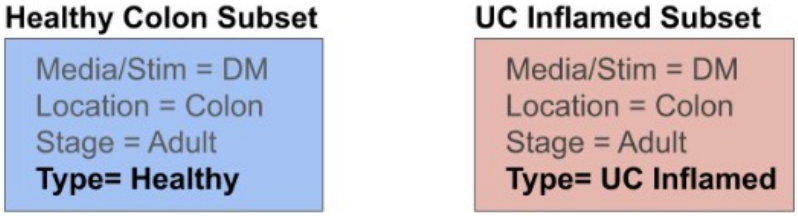
5.2.3 Preservation of Inflammatory Gene Expression Profiles in House-Protocol Cultured hIEOs

To investigate whether hIEOs retain inflammation-associated transcriptional programs, differential expression analysis was performed between hIEOs derived from HC and inflamed UC tissue, cultured under unstimulated differentiation conditions (DM). As with regional identity comparisons, stimulation was excluded from this analysis to avoid introducing secondary

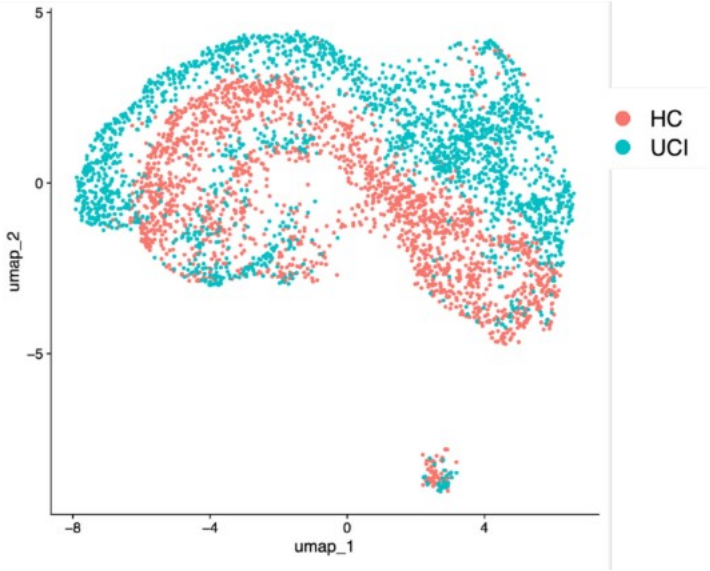
effects that could mask or distort disease-related transcriptional differences. Instead, the analysis focused on capturing baseline, donor-intrinsic inflammatory signatures. Similarly, fetal stage was removed since UC samples were only available for adult specimens (Figure 60).

Figure 60 - Transcriptional divergence between healthy and ulcerative colitis–inflamed adult colon epithelial organoids

A) Metadata for healthy colon and UC inflamed subsets



B) UMAP of epithelial cells by disease state



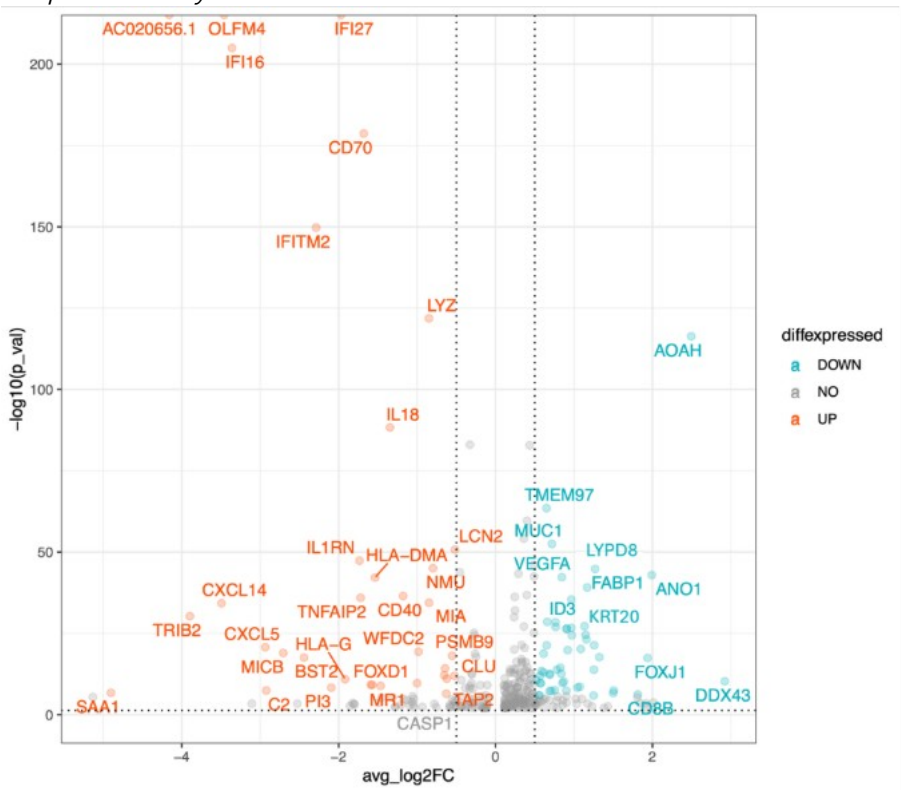
Legend: Metadata summary (top) and UMAP projection (bottom) of epithelial cells derived from adult colon organoids cultured in differentiation medium (DM). The samples represent either healthy controls (HC, red) or ulcerative colitis–inflamed (UCI, blue) tissue. The UMAP reveals that although HC and UCI cells partially overlap, inflamed cells occupy distinct transcriptional territories, suggesting inflammation-induced shifts in epithelial composition or cell state. This spatial divergence indicates transcriptional remodelling in response to chronic inflammatory cues in UC.

The volcano plot in Figure 61A reveals that hIEOs derived from inflamed UC tissue retain a characteristic inflammatory gene expression profile, marked by upregulation of genes such as *IL18*, *IFITM2*, and *OLFM4*, alongside downregulation of epithelial differentiation and absorptive

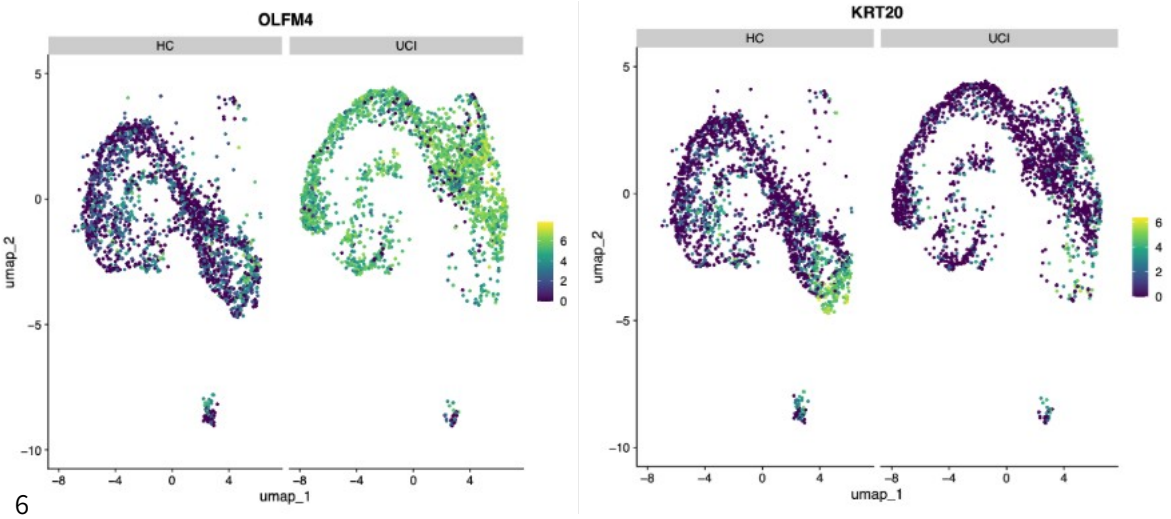
markers including *FABP1*, *LYPD8*, and *KRT20*. Examination of representative genes — *KRT20*, a hallmark of mature crypt-top enterocytes, and *OLFM4*, a progenitor marker typically expanded during inflammatory crypt hyperplasia — further underscores the transcriptional divergence between healthy and inflamed states (Figure 61B). These changes closely mirror patterns observed in primary inflamed colonic tissue, supporting the notion that epithelial organoids can robustly maintain disease-associated transcriptional programmes *in vitro*, even in the absence of direct immune cues.

Figure 61 - Transcriptomic changes in inflamed ulcerative colitis (UCI) epithelium compared to healthy colon (HC)

A) Differential expression analysis of UCI versus HC



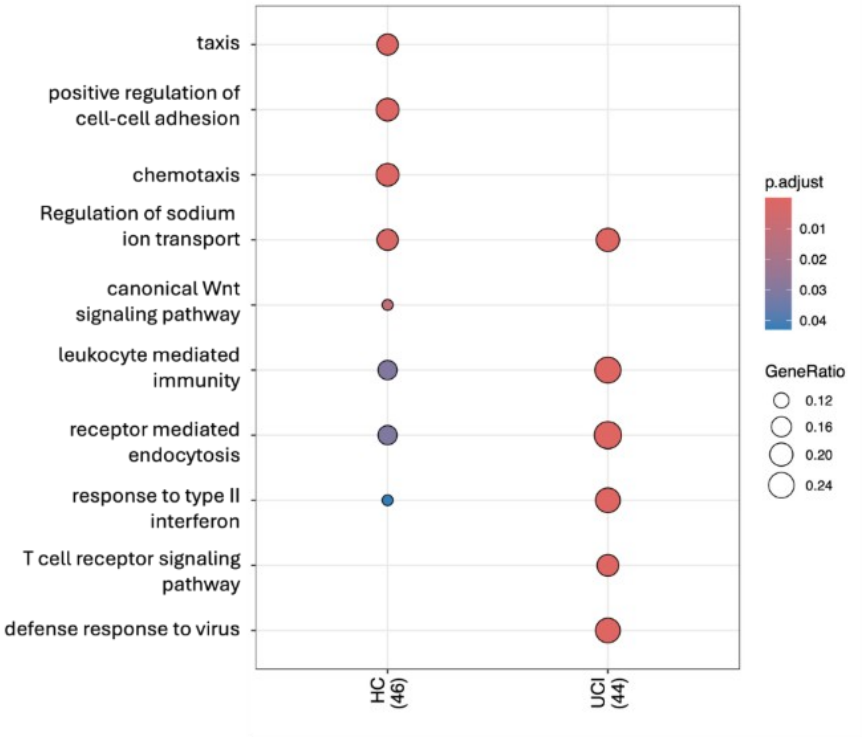
B) Expression of OLFM4 and KRT20 in HC and UCI samples



Legend: (A) Differentially expressed genes in inflamed ulcerative colitis (UCI) compared to healthy colon (HC). Genes upregulated in UCI are shown in orange, with prominent inflammatory and interferon response markers such as OLFM4, IFI16, IFI27, and IL1RN. Downregulated genes, shown in blue, include epithelial differentiation markers such as KRT20, FABP1, and FOXJ1. Grey points denote genes not meeting the differential expression threshold. (B) UMAP projections of epithelial cells coloured by the expression of OLFM4 (left) and KRT20 (right), across HC and UCI conditions. OLFM4 is expanded in UCI samples, consistent with regenerative or progenitor-like phenotypes, whereas KRT20 is markedly reduced, indicating loss of mature absorptive cell identity. These results suggest inflammatory reprogramming and impaired differentiation in UCI epithelium.

GO enrichment analysis reinforced the observed transcriptional differences, revealing distinct biological processes enriched in healthy versus inflamed colonic epithelium (Figure 62). In inflamed UC samples, the most significantly enriched pathways were predominantly immune-related, including interferon responses, T cell receptor signalling, leukocyte-mediated immunity, and receptor-mediated endocytosis. These findings are consistent with an activated inflammatory state. In contrast, healthy colon samples showed enrichment for more homeostatic processes, such as regulation of sodium ion transport, cell–cell adhesion, and canonical Wnt signalling, reflecting intact epithelial function and tissue organisation. These results highlight the shift from epithelial maintenance to immune activation in the inflamed state.

Figure 62 - GO enrichment analysis of differentially expressed genes in healthy control (HC) and ulcerative colitis inflamed (UCI) epithelial samples



Legend: Gene Ontology (GO) biological process enrichment for differentially expressed genes in intestinal epithelial cells derived from healthy controls (HC) and ulcerative colitis samples (UCI). Key enriched pathways in HC include taxis, positive regulation of cell-cell adhesion, and chemotaxis, suggesting an emphasis on epithelial structural organisation and immune cell guidance. By contrast, UCI-enriched processes include defence response to virus, T cell receptor signalling pathway, and response to type II

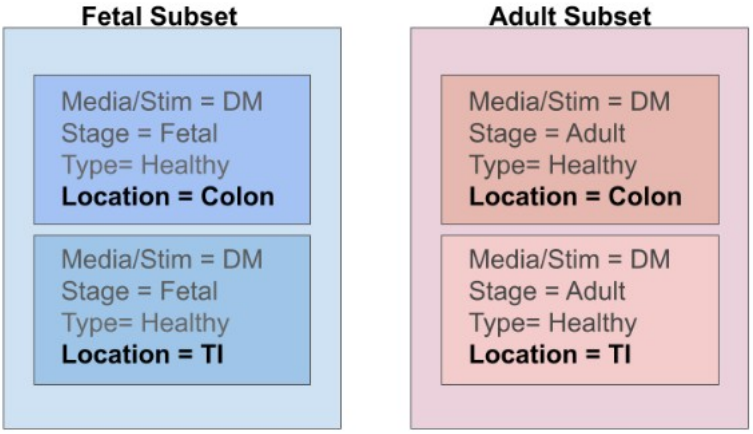
interferon, consistent with a heightened inflammatory and antiviral state. Shared terms such as receptor-mediated endocytosis and regulation of sodium ion transport highlight common epithelial functional pathways affected across conditions. These findings reveal distinct biological signatures between homeostatic and inflamed intestinal epithelia.

Together, these findings support the notion that epithelial components of the UC inflammatory response are at least partially encoded at the tissue-of-origin level and remain detectable under standard differentiation conditions, validating the use of hIEOs for modelling basal disease signatures and epithelial responses to inflammation.

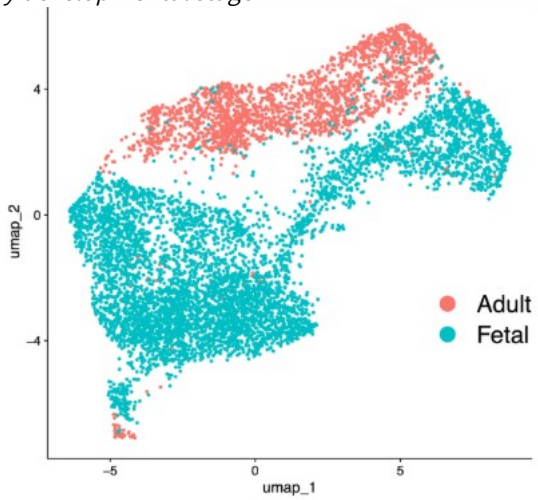
5.2.4 Maintenance of Developmental Stage-Specific Signatures in House-Protocol Cultured hIEOs

To evaluate whether hIEOs preserve transcriptional differences associated with developmental stage, differential expression analysis was performed between fetal and adult-derived hIEOs cultured under unstimulated differentiation conditions (DM; Figure 63A). By focusing on a non-stimulated baseline, this comparison isolates intrinsic stage-specific transcriptional programs without confounding external signals.

Figure 63 - Transcriptional differences between fetal and adult intestinal epithelial organoids
A) Metadata for fetal and adult subsets



B) UMAP of epithelial cells by developmental stage



Legend: Metadata overview (top) and UMAP projection (bottom) showing single-cell transcriptomes from fetal and adult intestinal epithelial organoids derived from both terminal ileum (TI) and colon, all cultured under the same differentiation medium (DM) conditions. Cells are coloured by developmental stage: fetal (blue) and adult (red). The UMAP shows a clear separation between fetal and adult cells, indicating substantial transcriptomic differences driven by developmental stage. Despite being cultured ex vivo under identical conditions, fetal cells occupy distinct transcriptional territories compared to their adult counterparts, reflecting intrinsic differences in lineage specification, proliferative potential, and cellular identity across developmental timepoints.

As shown in Figure 64A, fetal epithelial samples displayed a distinct transcriptional landscape indicative of early developmental states. Genes such as *NEUROD1*, *OLFM4*, *SMOC2*, and *CLDN10* were significantly upregulated in fetal cells, consistent with a progenitor-rich, immature epithelial phenotype: *NEUROD1* is a transiently expressed neurogenic transcription factor in early endocrine progenitors, *OLFM4* marks crypt-based stem and progenitor populations,

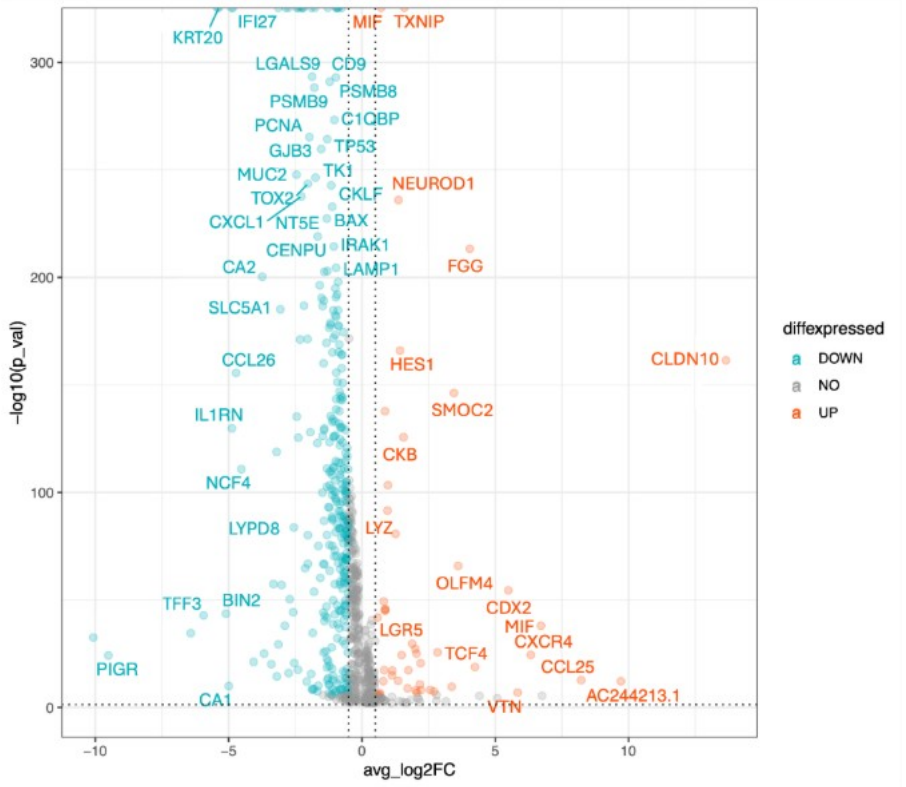
SMOC2 is implicated in developmental matrix remodelling, and *CLDN10* is associated with epithelial barrier formation and fetal extracellular matrix organisation. By contrast, adult epithelial cells expressed higher levels of genes linked to terminal differentiation and functional maturation, including *KRT20*, *SLC5A1*, *TFF3*, and *LYPD8*. These markers reflect absorptive and secretory lineages characteristic of a homeostatic adult colon: *KRT20* defines mature enterocytes, *SLC5A1* encodes a key transporter for nutrient uptake, *TFF3* is secreted by goblet cells to maintain mucosal integrity, and LYPD8 is a colonocyte-derived antimicrobial factor.

Although the volcano plot shows an apparent skew towards downregulated genes, this pattern is biologically plausible and does not indicate a normalisation error. The comparison performed here contrasts fetal versus adult colon epithelium, two conditions known to differ markedly in global transcriptional activity, cell identity programmes, and proliferative states. Adult tissue displays a more differentiated, lineage-restricted transcriptome, whereas fetal epithelium is characterised by broad developmental gene expression, less transcriptional repression, and extensive cellular plasticity. Consequently, when adult is used as the reference group, a large number of genes appear downregulated simply because fetal tissue expresses a wider set of genes at higher baseline levels. This phenomenon has been reported before in different studies comparing developmental and adult intestinal epithelium and is not attributable to sequencing depth or normalisation artefacts.

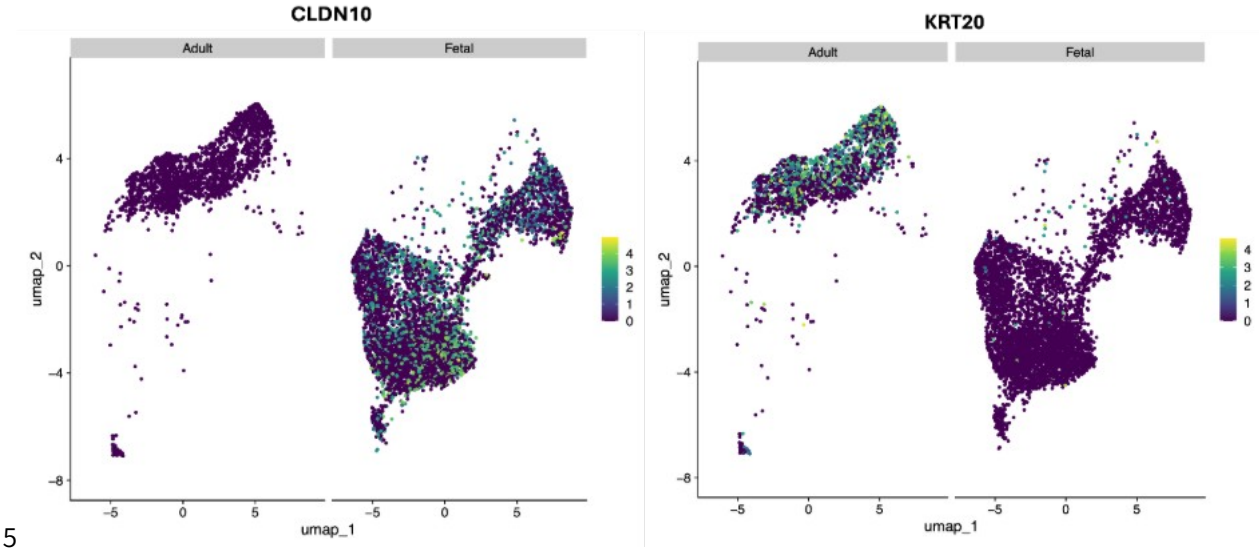
The UMAP projections of *CLDN10* and *KRT20* expression (Figure 64B) further illustrate these developmental distinctions, with *CLDN10* predominantly expressed in fetal cells and *KRT20* in adult cells. Together, these results underscore the fidelity of the system in capturing ontogeny-associated transcriptional programs and support the preservation of stage-specific epithelial identity during *in vitro* expansion.

Figure 64 - Transcriptomic comparison between fetal and adult human colonic epithelium

A) Differential gene expression between fetal and adult colon



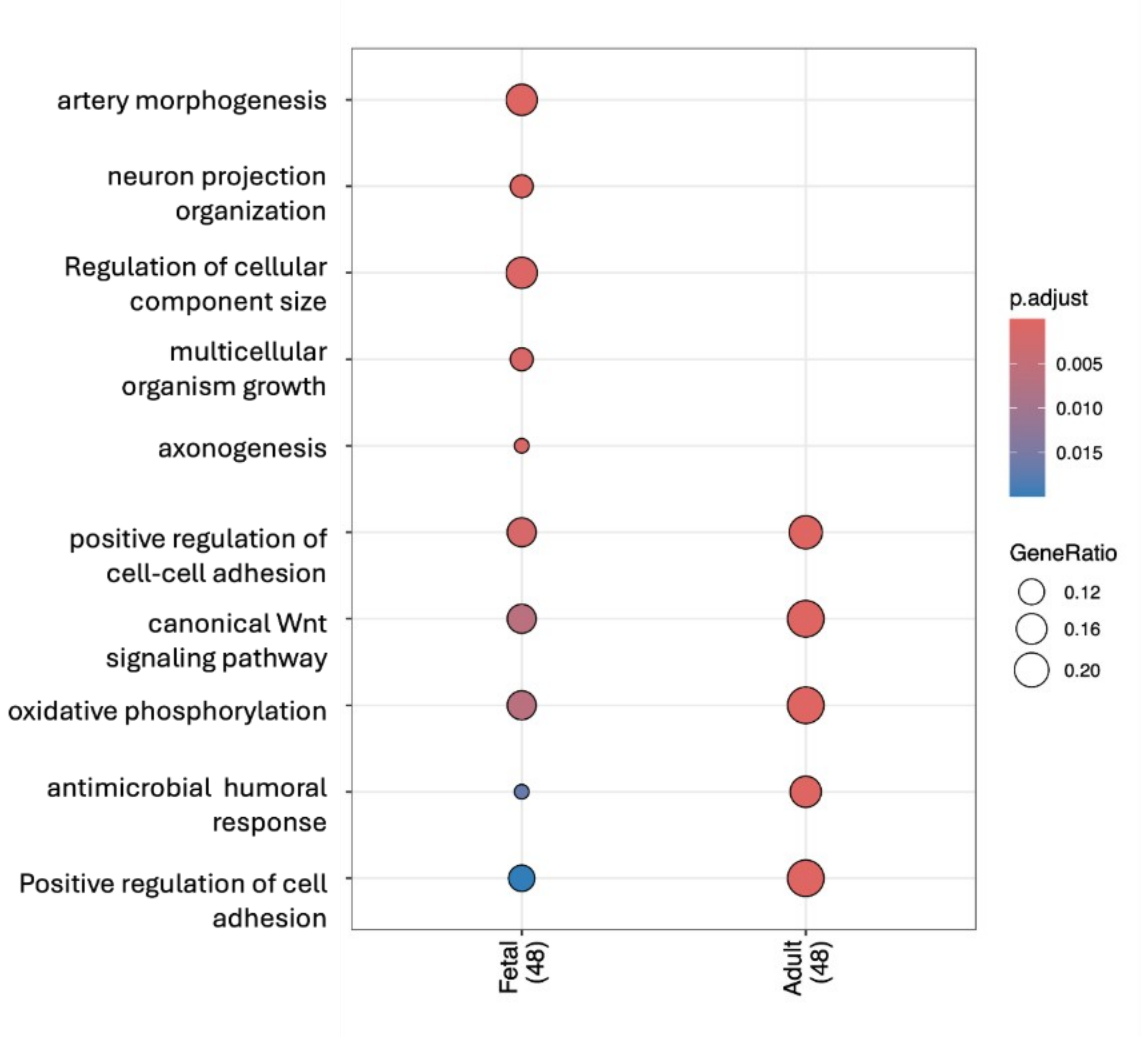
B) Expression of CLDN10 and KRT20 in adult and fetal epithelium



Legend: (A) Differentially expressed genes between fetal and adult colonic epithelial cells. Genes upregulated in the fetal epithelium are shown in orange and include developmental regulators such as LGR5, OLFM4, TCF4, and CLDN10, while genes upregulated in the adult colon, shown in blue, include markers of mature epithelial differentiation such as KRT20, MUC2, and PIGR. (B) UMAP projections of adult and fetal epithelial cells coloured by expression levels of CLDN10 (left) and KRT20 (right). CLDN10 is selectively enriched in fetal cells, reflecting a developmental transcriptional programme, whereas KRT20 is predominantly expressed in the adult compartment, consistent with terminal differentiation of absorptive lineages. These data demonstrate a clear shift in gene expression profiles between fetal and adult stages, highlighting the immaturity and proliferative bias of the fetal epithelium.

GO enrichment analysis of differentially expressed genes between fetal and adult epithelial cells revealed distinct pathway-level signatures reflective of developmental stage (Figure 65). In fetal samples, enriched terms included artery morphogenesis, axonogenesis, neuron projection organisation, and multicellular organism growth, consistent with a transcriptional landscape geared toward developmental plasticity, tissue patterning, and cellular expansion. These findings align with the elevated expression of progenitor-associated genes such as *OLFM4*, *NEUROD1*, and *SMOC2* in fetal cells. By contrast, adult epithelial cells displayed significant enrichment for pathways associated with epithelial homeostasis and function, including canonical Wnt signalling, oxidative phosphorylation, positive regulation of cell–cell adhesion, and antimicrobial humoral response. These terms reflect a mature, metabolically active, and immunocompetent epithelium, characteristic of the adult intestinal barrier. Collectively, these results confirm that the transcriptional differences observed at the gene level are mirrored by broader functional programmes, reinforcing the biological fidelity of hIEO-derived cells in capturing developmental stage-specific epithelial identities.

Figure 65 - GO enrichment analysis of differentially expressed genes in fetal versus adult human colonic epithelium



Legend: Top enriched GO biological processes among genes differentially expressed in fetal (left) and adult (right) epithelial cells. Fetal-enriched pathways include developmental and morphogenetic processes such as artery morphogenesis, axonogenesis, and neuron projection organisation, while adult-enriched pathways include oxidative phosphorylation, canonical Wnt signalling, positive regulation of cell adhesion, and antimicrobial humoral response. These findings illustrate distinct biological programmes active in fetal versus adult epithelium, reflective of developmental stage and functional maturation.

These findings highlight that the developmental origin of epithelial cells exerts a lasting influence on their transcriptional identity, even when cultured under standardised conditions. The retention of stage-specific gene expression programmes — including immune-related signatures in fetal samples, and differentiation-associated markers in adult samples — underscores the importance of considering donor age as a key biological variable in organoid-

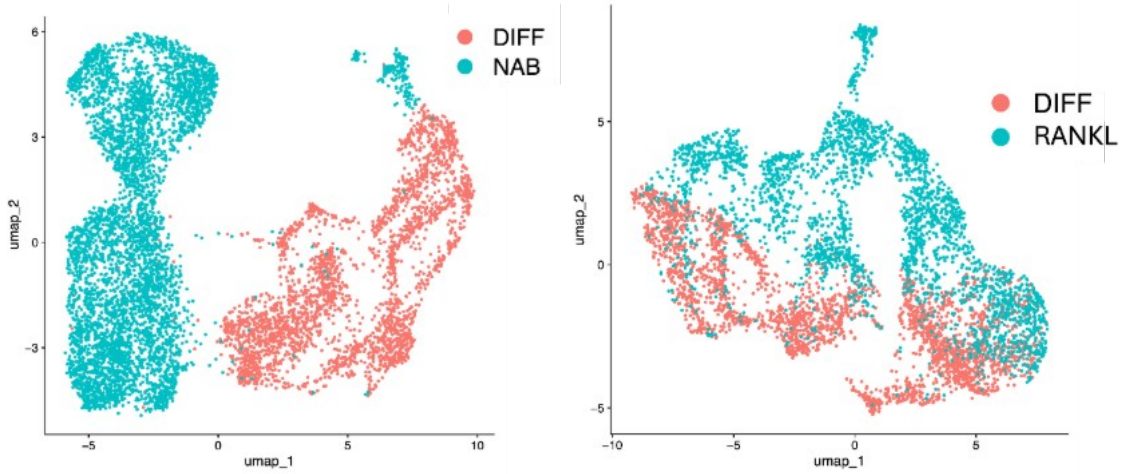
based studies. Together, these results confirm that hIEOs preserve intrinsic developmental cues, supporting their utility for modelling stage-specific epithelial states *in vitro*.

5.2.5 Differential Responses to Stimulation in House-Protocol Cultured hIEOs: NaB and RANKL as Divergent Modulators

To assess the transcriptional impact of specific perturbations under uniform differentiation conditions, a donor-matched comparison was carried out using two stimuli of interest—sodium butyrate (NaB) and RANKL—relative to baseline differentiation (DM, or DIFF). The analysis included all donor samples previously described in Figure 2: two fetal TI, two fetal colon, one adult UC inflamed, one adult UC non-inflamed, one adult healthy colon, and one adult healthy terminal ileum. Combined UMAP projections for each stimulation condition across all donors are shown in Figure 66.

Figure 66 - UMAP projections showing transcriptional responses to sodium butyrate (NaB) and RANKL stimulation in hIEOs

A) NaB stimulation *B) RANKL stimulation*



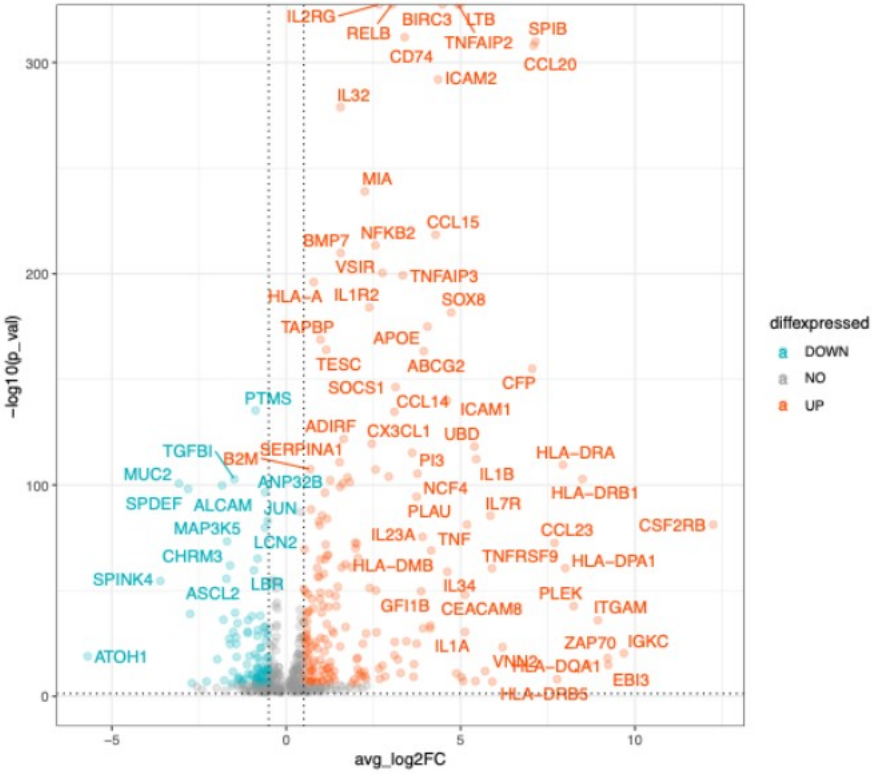
Legend: UMAP plots visualising single-cell transcriptomes coloured by treatment condition. (A) Comparison between standard differentiation media (DIFF, red) and sodium butyrate-treated samples (NaB, blue), showing a pronounced shift in transcriptional landscape indicative of widespread cellular reprogramming. (B) Comparison between DIFF (red) and RANKL-treated conditions (blue), demonstrating more subtle but distinct changes in cell state distribution.

NaB treatment induced a pronounced shift in the transcriptional landscape compared to standard differentiated controls (Figure 66A). Gene set enrichment analysis revealed robust activation of pathways related to cell cycle progression, DNA replication, and nuclear processes associated with transcription factor activity and chromatin remodelling (data not shown), as well as aberrant expression of B-cell–associated genes (see Supplementary Figure S27). These changes are consistent with the known histone deacetylase (HDAC) inhibitory effects of NaB, suggesting that the treatment triggered extensive transcriptional reprogramming. However, the scale and profile of the response appeared non-physiological and indicative of cellular stress or dysfunction. The high NaB concentration used likely contributed to cytotoxicity, reducing cell viability and disrupting normal lineage specification. Due to these confounding effects, NaB-treated samples were excluded from downstream analyses to preserve the biological interpretability of the dataset.

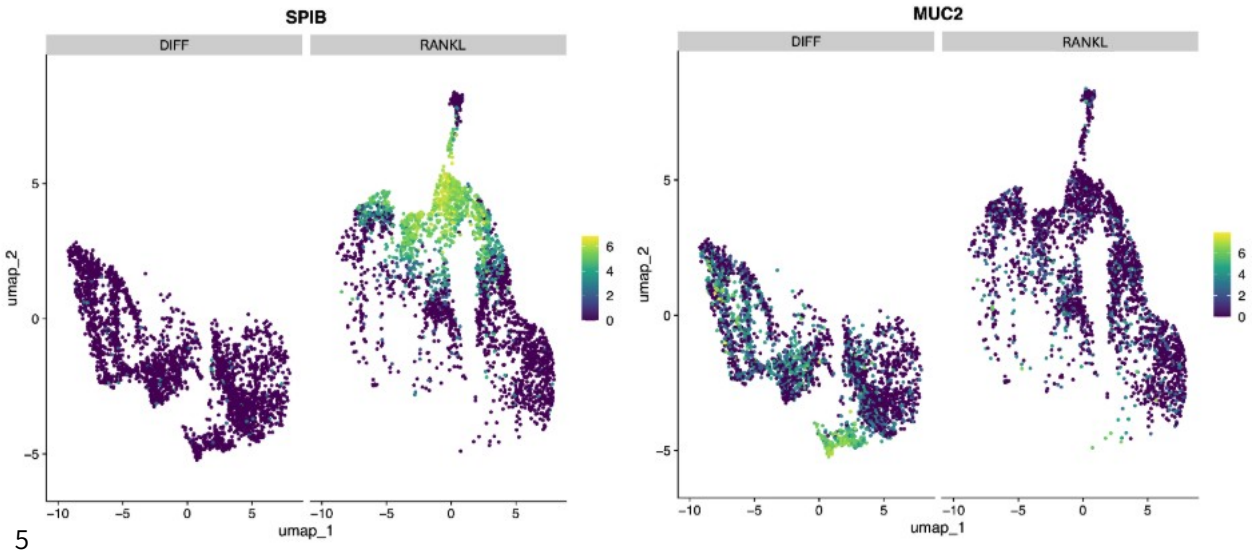
In contrast to the widespread and dysregulated transcriptional effects observed with NaB, RANKL stimulation resulted in a more targeted and physiologically relevant transcriptional shift (Figure 66B). As illustrated in Figure 67A, RANKL-treated hIEOs showed upregulation of canonical M-cell markers, including *SPIB*, *TNFAIP2*, *CCL15*, and *CCL20*, alongside enrichment of gene sets linked to antigen sampling, lymphoid signalling, and epithelial remodelling. These signatures were consistently observed across multiple donors and culture conditions, underscoring the robustness and reproducibility of the response. This is further exemplified in the UMAP projections (Figure 67B), where *MUC2*, a goblet-cell marker associated with mucin production, is enriched in unstimulated differentiating conditions (DM, or DIFF), while *SPIB*, a transcription factor essential for M-cell identity, is specifically induced upon RANKL treatment. Together, these results validate RANKL as a functional cue for directing lineage specification within the hIEO model.

Figure 67 - Transcriptomic effects of RANKL stimulation on hIEOs

A) Differential gene expression between RANKL-stimulated and control (DIFF) conditions



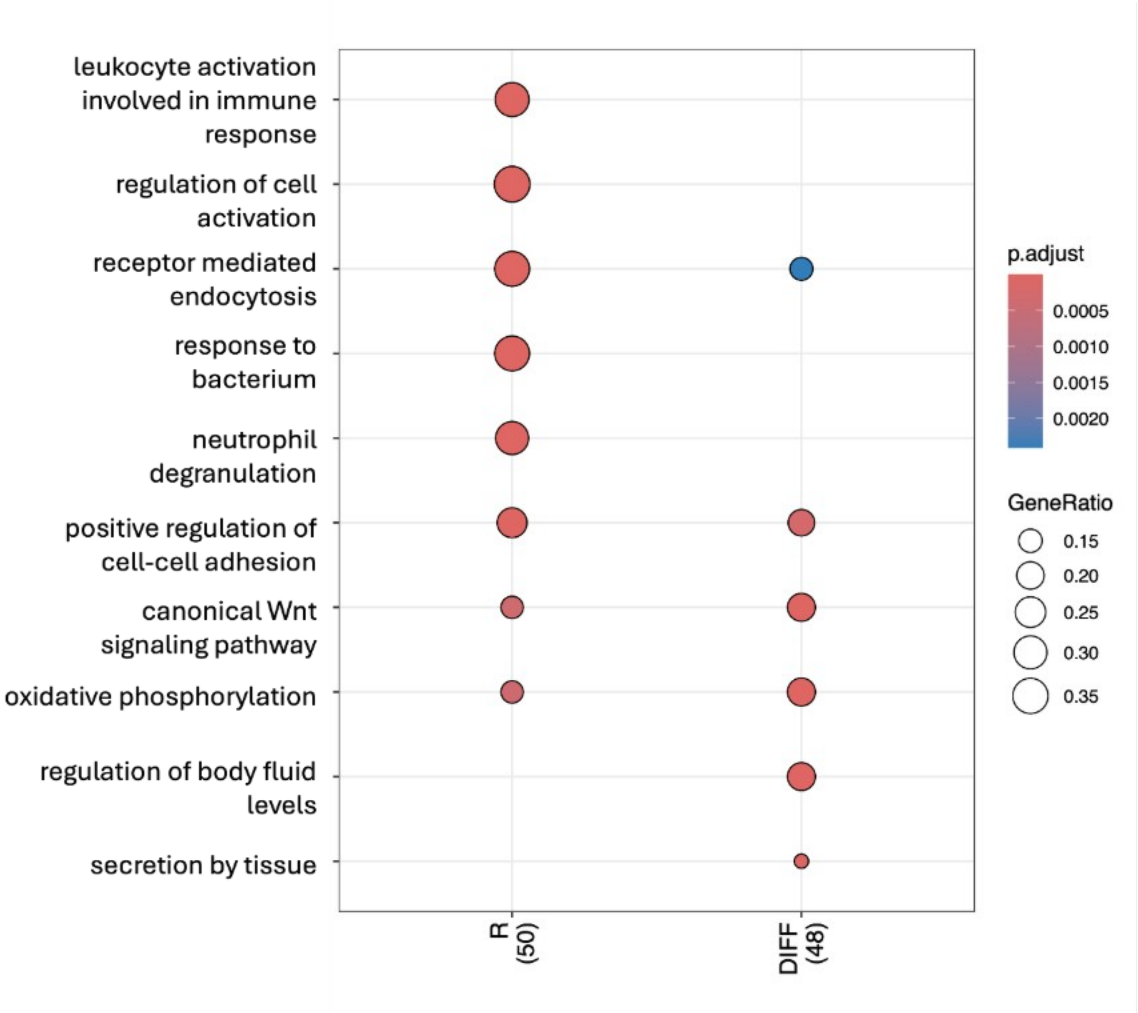
B) Expression of SPIB and MUC2 in DIFF and RANKL conditions



Legend: (A) Comparative gene expression profiles between RANKL-stimulated epithelial cells and unstimulated differentiation controls (DIFF). RANKL exposure led to significant upregulation of genes involved in M-cell specification, pro-inflammatory signalling, and antigen presentation, including SPIB, TNFAIP2, REL, IL32, and HLA family genes (highlighted in orange). Conversely, several epithelial differentiation markers, such as MUC2, TFF3, and SPDEF, were significantly downregulated (highlighted in blue), suggesting a shift in cellular identity. (B) UMAP projections showing the expression of SPIB (left), a key transcription factor required for M-cell differentiation, and MUC2 (right), a marker of goblet cells. SPIB expression is strongly induced in RANKL-treated cells, while MUC2 expression is reduced, consistent with a lineage bias away from goblet cells toward an M-cell–like phenotype. These results confirm the ability of RANKL to reprogramme epithelial fate in vitro and promote M-cell–associated transcriptional profiles.

The GO enrichment analysis revealed that RANKL induces a distinct and functionally coherent transcriptional programme consistent with M-cell specification and immune-epithelial interaction (Figure 68). In the RANKL condition, there is significant enrichment of immune-associated processes, including “leukocyte activation involved in immune response,” “neutrophil degranulation,” and “response to bacterium” — key hallmarks of M-cell biology, which is defined by its role in antigen sampling and immune surveillance. Enrichment of “receptor-mediated endocytosis” further supports the acquisition of transcytotic capacity, a defining feature of mature M-cells. Additional terms such as “regulation of cell activation” and “positive regulation of cell–cell adhesion” suggest epithelial remodelling and heightened immune responsiveness. Enrichment of “canonical Wnt signalling pathway” and “oxidative phosphorylation” may reflect underlying shifts in metabolic or developmental signalling states associated with lineage commitment and cellular specialisation, but, importantly, show that RANKL stimulation preserves overall epithelial viability, function and structural integrity, in stark contrast to the cytotoxic effects seen with NaB. By contrast, the unstimulated condition showed a more limited enrichment profile dominated by general epithelial functions such as “oxidative phosphorylation,” “secretion by tissue,” and “regulation of body fluid levels,” aligning with homeostatic absorptive and secretory programmes typical of enterocyte differentiation. Notably absent were immune-focused processes, reinforcing the distinction between these two conditions. Together, these results demonstrate that RANKL stimulation activates a robust and immune-specialised transcriptional state distinct from standard differentiation, providing strong molecular evidence for its utility in promoting M-cell-like identity within the hIEO model. The induction of M-cell–like features without compromising cell health supports the use of RANKL as a tractable experimental tool to study rare epithelial lineages *in vitro.*

Figure 68 - GO enrichment analysis of genes upregulated in RANKL-stimulated versus unstimulated intestinal epithelial organoids



Legend: GO biological processes enriched among genes upregulated in intestinal epithelial organoids cultured with RANKL stimulation (R) or under standard differentiation conditions (DIFF). RANKL-stimulated cultures show strong enrichment for immune-related processes such as leukocyte activation, neutrophil degranulation, and response to bacterium, consistent with induction of M-cell–like phenotypes. By contrast, unstimulated conditions (DIFF) are associated with enrichment in epithelial and metabolic processes such as canonical Wnt signalling, oxidative phosphorylation, and secretion by tissue. These results highlight the transcriptional reprogramming driven by RANKL stimulation and the emergence of specialised epithelial states linked to immune surveillance.

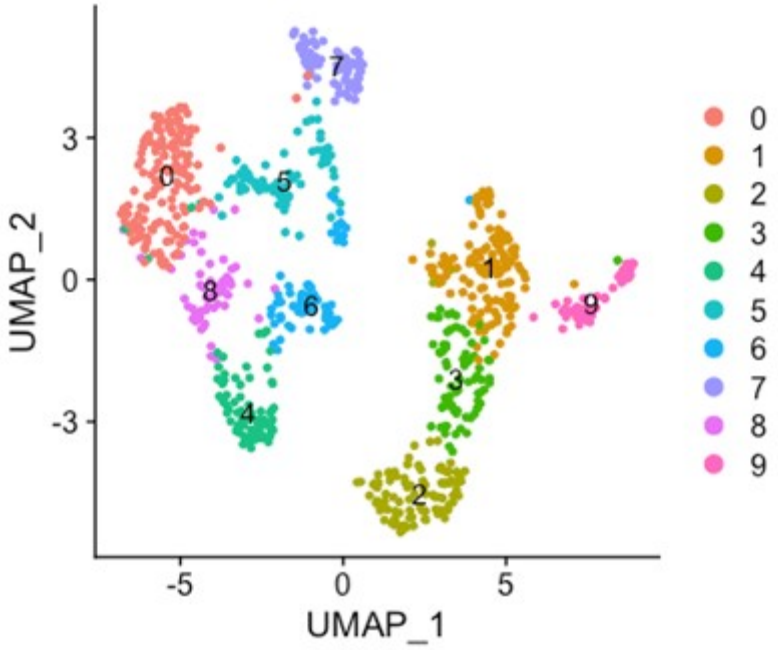
5.2.5.1 Outcomes of RANKL stimulation on rare cell types in House-Protocol Cultured hIEOs

To investigate the cell states induced specifically by RANKL stimulation, the subset of cells derived exclusively from the RANKL-only condition were selected and analysed independently. UMAP-based clustering revealed substantial transcriptional heterogeneity, with the emergence of several distinct subpopulations (Figure 69A). Expression analysis of canonical M-cell markers

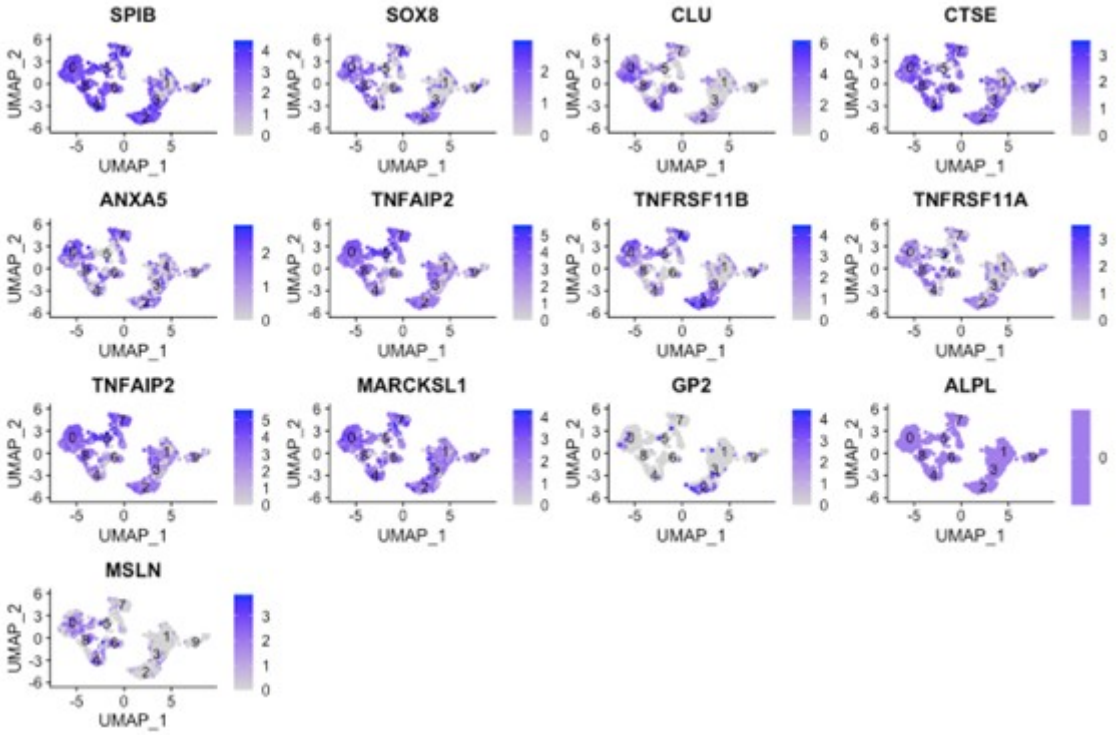
showed that genes such as *SPIB*, *SOX8*, *TNFAIP2*, *TNFRSF11B*, *GP2*, *CLU*, *CTSE*, *MARCKSL1*, and *MSLN* were selectively enriched in defined clusters (Figure 69B). These markers span transcription factors, immune-related receptors, and membrane-associated effectors typically associated with M-Cell maturation and specialisation, collectively suggesting the presence of M-cell subpopulations within the RANKL-stimulated epithelium. Additional genes associated with chemokine signalling and immune crosstalk — including *CCL20*, and *CCL25* (see Supplementary Figure S28) — were also upregulated, reinforcing the immunologically specialised phenotype characteristic of M-cells. Taken together, these data confirm that RANKL stimulation induces a robust and spatially restricted M-cell transcriptional programme, validating its effectiveness in directing rare epithelial lineage specification *in vitro*.

Figure 69 - UMAP clustering and gene expression profiles of RANKL-stimulated epithelial cells showing marker expression across transcriptional clusters

A) UMAP clustering of cells from the RANKL-only condition



B) Expression of selected M-cell–associated markers across RANKL-stimulated clusters

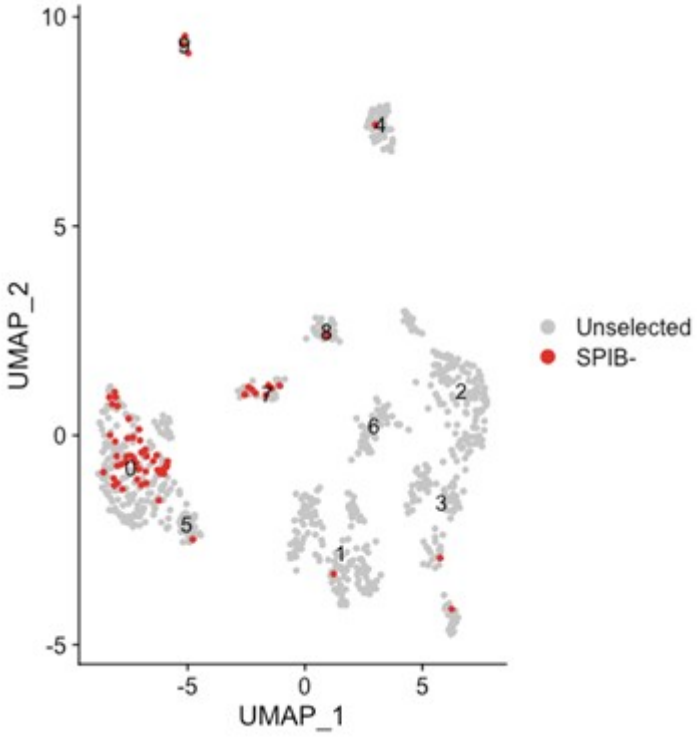


Legend: (A) UMAP projection of single cells isolated from RANKL-stimulated hIEOs, revealing transcriptional heterogeneity with multiple distinct subclusters. Each colour represents a different transcriptional cluster. (B) Expression of key M-cell–associated genes across these clusters. Markers include transcriptional regulators (SPIB, SOX8), immunoregulatory genes (TNFAIP2, TNFRSF11A, TNFRSF11B), and membrane-associated effectors (GP2, CLU, ANXA5, MARCKSL1, MSLN). Expression is concentrated in specific clusters, indicating the emergence of M-cell–like subpopulations. These results confirm that RANKL robustly induces an M-cell differentiation programme within the organoid epithelium.

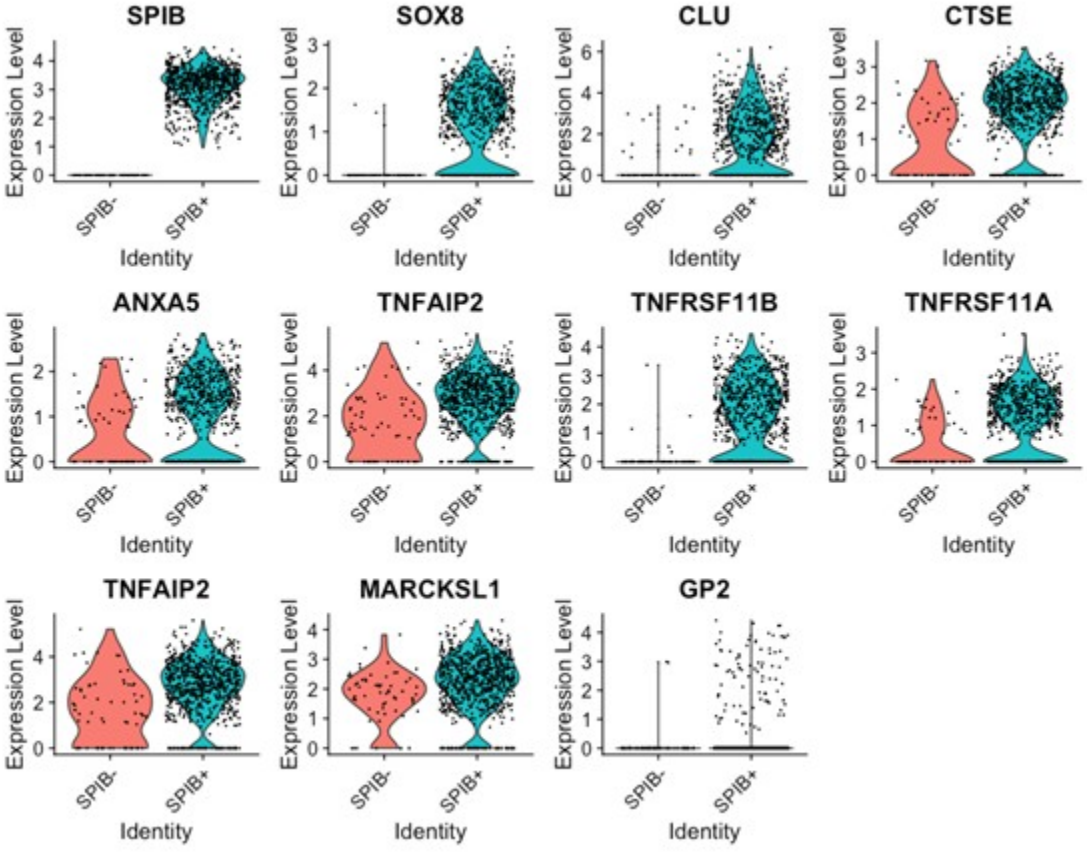
To further investigate the transcriptional diversity within RANKL-stimulated epithelial cultures, the subset of cells lacking *SPIB* expression (*SPIB*-) were isolated to determine whether they represented a distinct cell fate or an earlier developmental stage. As shown in the UMAP projection (Figure 70A), *SPIB*- cells were broadly distributed across multiple transcriptional clusters, rather than forming a separate population, suggesting a continuum of cell states. Expression profiling revealed that *SPIB*- cells lacked canonical M-cell markers, including *CLU*, *TNFRSF11B*, and *SOX8* (Figure 70B), consistent with an absence of mature M-cell identity. Instead, *SPIB*- cells showed elevated expression of genes linked to proliferative and early epithelial programmes, such as *OLFM4*, *GDF15*, *TOP2A*, and *STMN1* (Figure 70C). Additional enrichment of genes like *FABP5*, *NUPR1*, *CAMK2N1*, and *UCA1* further supported a bias toward undifferentiated or stress-associated states. Together, these results suggest that *SPIB*- cells within the RANKL-stimulated compartment are not committed to a separate lineage, but likely represent an immature or progenitor-like population poised for transition toward M-cell differentiation.

Figure 70 - UMAP projection and differential gene expression analysis of SPIB⁻ cells within RANKL-stimulated epithelial monolayers

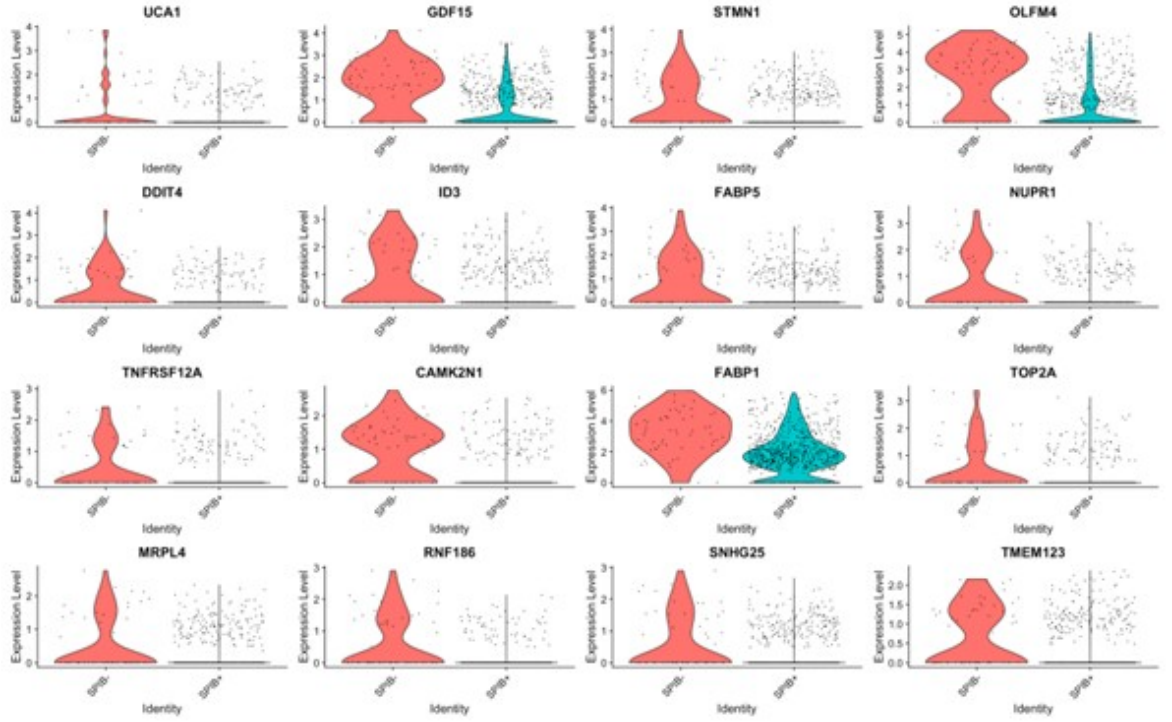
A) UMAP projection showing SPIB⁻ cells (red) among all RANKL-stimulated cells



B) Gene expression levels for differentially upregulated genes in in SPIB+ versus SPIB- populations



C) Gene expression levels for differentially upregulated genes in in SPIB- versus SPIB+ populations

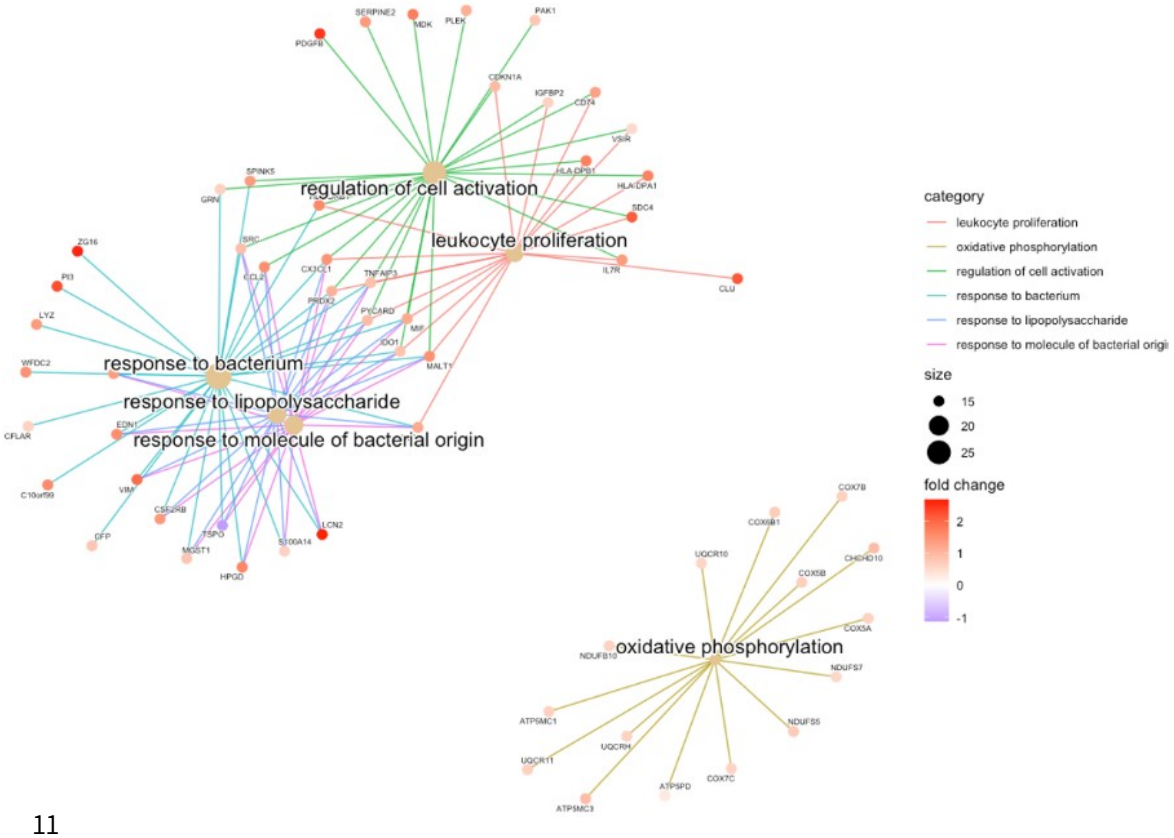


Legend: (A) UMAP embedding of RANKL-stimulated epithelial cells, highlighting cells that do not express SPIB (SPIB⁻, red) versus all other cells (grey). These cells are distributed across several clusters, with enrichment in regions associated with early or proliferative states. (B) Expression patterns of M-cell markers (SPIB, SOX8, CLU, CTSE, TNFAIP2, TNFRSF11A/B, GP2, MARCKSL1), which are largely absent in SPIB⁻ cells, confirming their lack of M-cell identity. (C) Genes enriched in the SPIB⁻ population, including early/progenitor-associated markers (OLFM4, GDF15, TOP2A, STMN1, CAMK2N1, NUPR1) and stress-response or undifferentiated state markers (UCA1, RNF186, SH3G25, FABP5, FABP1), indicating that SPIB⁻ cells predominantly represent immature or transit-amplifying epithelial states.

To further explore the functional impact of RANKL stimulation, a network-based representation of enriched GO terms was constructed (Figure 71). The analysis revealed two distinct functional modules. The first and most interconnected cluster was dominated by immune-related processes, including “leukocyte proliferation,” “regulation of cell activation,” and “response to bacterium,” as well as responses to lipopolysaccharide and other bacterial molecules. Central genes within this module — such as *TNFAIP2/3*, *VIM*, *CLU*, *PI3*, and *MALT1*— highlight a strong epithelial transcriptional response to immunological and microbial cues. This is consistent with the known role of RANKL in promoting M-cell differentiation and enhancing mucosal immune

surveillance. A second, spatially distinct module was enriched for “oxidative phosphorylation” and comprised multiple mitochondrial genes, including *COX6A1*, *COX7C*, *NDUFS7*, *UQCRC1*, and *ATP5MC1*. This suggests that RANKL stimulation may also drive metabolic reprogramming, potentially to meet the energetic demands associated with differentiation and immune function. Node colour intensity in the network reflects fold-change magnitude, illustrating coordinated upregulation across both immune and metabolic modules. Collectively, these findings indicate that RANKL elicits a broad transcriptional programme that integrates immune activation with metabolic support, reinforcing its role in directing epithelial specialisation.

Figure 71 - Network-based visualisation of enriched Gene Ontology terms and associated genes in RANKL-stimulated epithelial cells

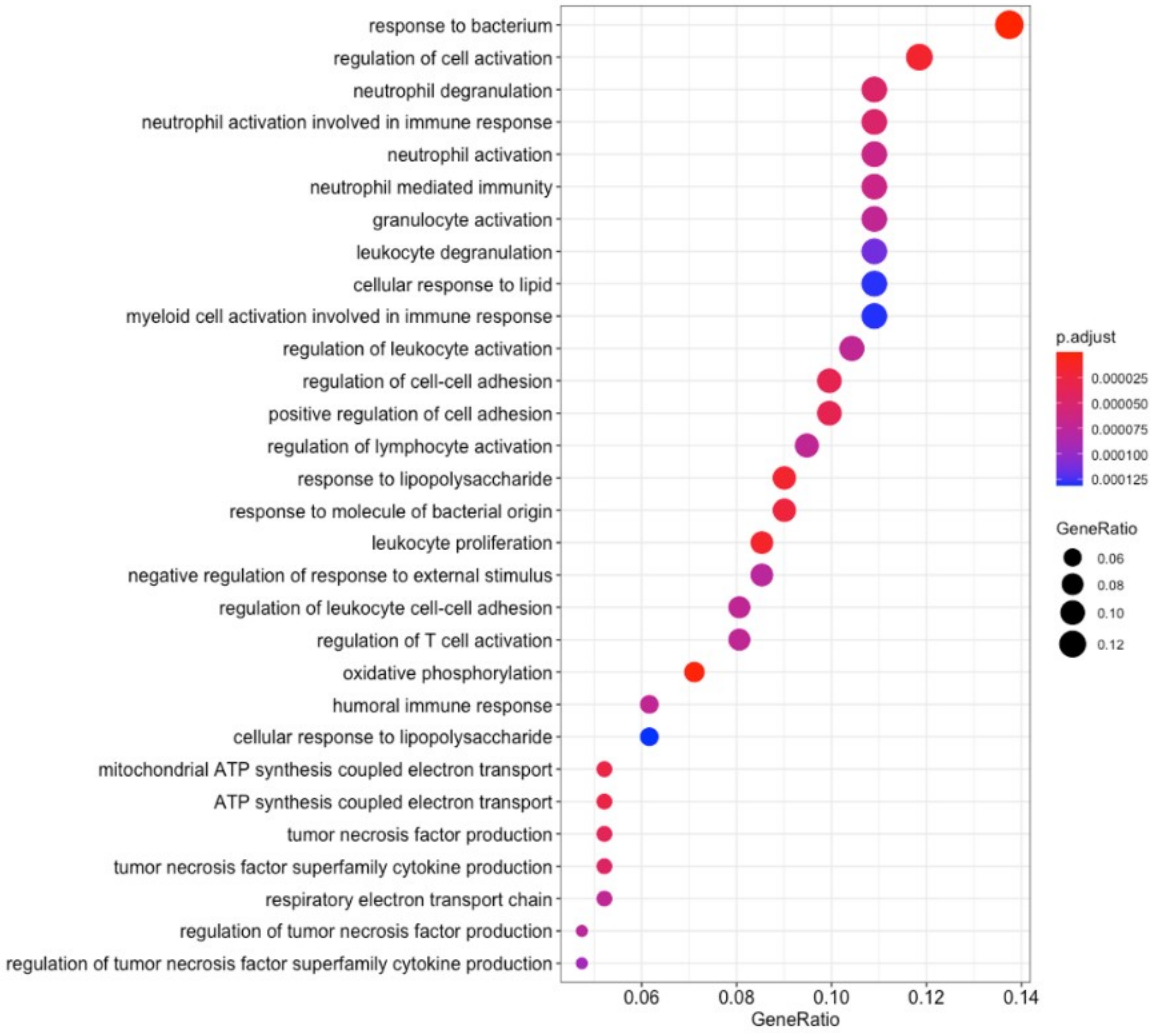


Legend: Network representation of enriched GO biological processes in RANKL-treated hIEOs. Nodes represent genes, and edges indicate functional associations within GO terms. Colours correspond to GO categories (e.g. immune response, oxidative phosphorylation), while node size reflects gene connectivity and colour intensity indicates fold change. Two main functional modules are evident: at the top, an immune-related module encompassing “regulation of cell activation,” “leukocyte proliferation,” and responses to bacterial molecules, including TNFAIP2, CXCL1, CXCL2, PIGR, and MMP1, reflecting epithelial priming for immune surveillance; at the bottom, a distinct metabolic module enriched for “oxidative

phosphorylation", composed of mitochondrial genes (COX6A1, NDUFS7, UQCRC1, etc.), suggesting RANKL-driven metabolic reprogramming.

To complement the network-based GO analysis, a ranked dot plot of significantly enriched GO terms among upregulated genes in RANKL-treated epithelial cells was generated (Figure 72). This analysis quantitatively confirmed the dominance of immune-related processes, with "response to bacterium" and "regulation of cell activation" emerging as the top-ranked terms, based on both statistical significance and gene ratio. Additional immune-relevant categories — such as "neutrophil degranulation," "myeloid cell activation involved in immune response," and "granulocyte activation" — highlighted the activation of innate immune pathways, particularly those associated with antimicrobial defence and immune cell recruitment. In parallel, several terms associated with epithelial function and metabolic activity were also significantly enriched. These included "positive regulation of cell adhesion" and mitochondrial processes such as "oxidative phosphorylation," "ATP synthesis coupled electron transport," and "respiratory electron transport chain," suggesting that RANKL stimulation also enhances bioenergetic processes and overall metabolic reprogramming, where mitochondrial function is likely upregulated to meet energy demands. Together, these data reinforce the role of RANKL in driving immune-specialised programmes aligned with M-cell differentiation.

Figure 72 - Enriched GO terms among upregulated genes in RANKL-stimulated epithelial cells



Legend: Top enriched GO biological processes associated with genes upregulated in RANKL-treated hIEO monolayers. Immune-related terms—such as "response to bacterium," "regulation of cell activation," "neutrophil degranulation," and "myeloid cell activation"—dominate the upper ranks, highlighting a strong immunological response. Additional enrichment of epithelial and metabolic processes—such as "positive regulation of cell adhesion," "oxidative phosphorylation," and "ATP synthesis coupled electron transport"—suggests coordinated transcriptional remodelling that integrates immune functionality with metabolic and structural readiness.

5.2.6 Media-Dependent Variation in hIEO Responses to Location and Stimulation

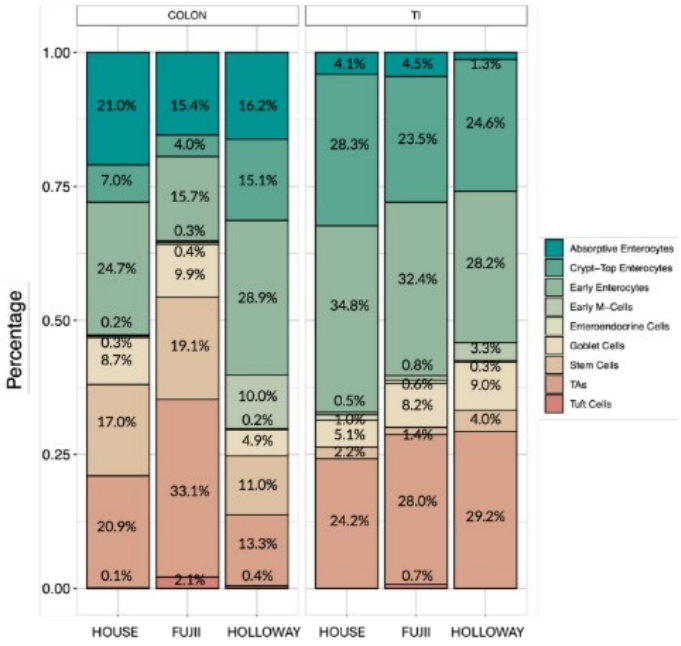
Throughout this thesis, multiple human intestinal epithelial organoid (hIEO) culture media—namely the House, Holloway, and Fujii protocols—have been systematically evaluated. The House protocol represents a standard culture medium, the reference baseline against which the performance of other protocols is benchmarked. This comparative framework is based on the premise that the House protocol reliably captures key transcriptional signatures characteristic of primary intestinal tissues across diverse contexts, including anatomical regions (colon vs TI), developmental stages (fetal vs adult), and inflammatory states (healthy vs UC inflamed), as well as under different stimulatory conditions. The previous section 5.2.1 established the fidelity of the House medium in these respects. Building on that foundation, the present section delves into the data generated from EXP03 (Figure 20) to examine how each culture medium modulates epithelial lineage representation and functional responses in relation to tissue origin and external stimuli, thereby unpicking the media-dependent plasticity of hIEOs.

5.2.6.1 Comparative Retention of Location and Inflammatory Signatures Across House, Fujii, and Holloway Protocols

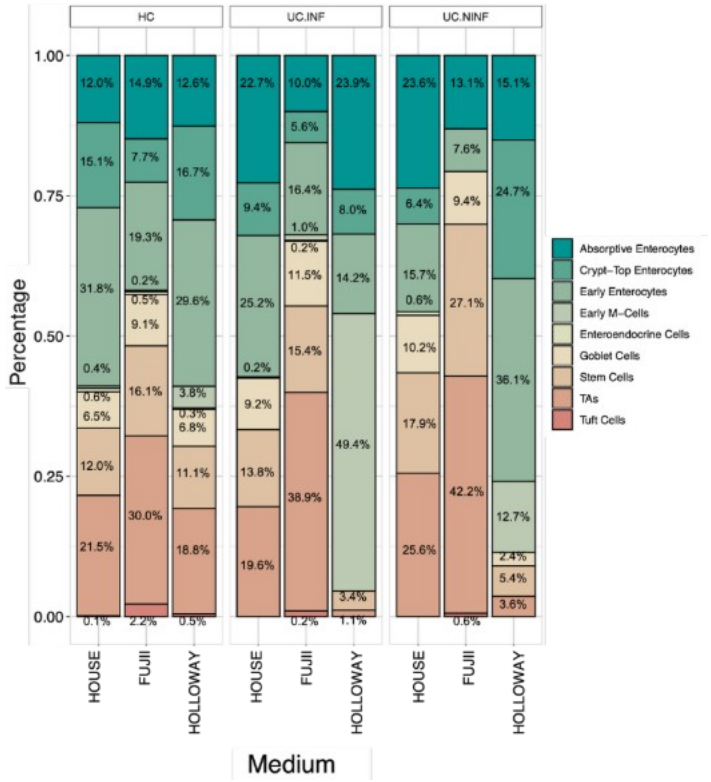
Figure 73A show the proportional representation of major epithelial cell types across colon and TI hIEOs, generated using three different culture protocols: House, Fujii, and Holloway. The aim of this analysis is to evaluate how effectively each protocol retains location-specific epithelial signatures, and to assess their ability to support differentiation into diverse intestinal lineages.

Figure 73 - Compositional differences in epithelial cell types across media conditions, intestinal regions, and disease states

A) Comparison across locations



B) Comparison across disease states



Legend: Proportional abundance of epithelial cell types across experimental conditions varying by culture medium (House, Fujii, Holloway), intestinal region (colon vs terminal ileum), and disease state (healthy, UC inflamed, UC non-inflamed). Terminal ileum organoids show more absorptive and early enterocytes, while colon organoids are enriched for secretory lineages, depending on the medium. Fujii medium favours stem and TA cell expansion; House and Holloway support more differentiated types. UC-inflamed samples display a shift toward progenitor-like states and reduced mature lineages, particularly under House medium, while non-inflamed UC samples resemble healthy profiles.

In colon-derived organoids, clear differences emerge between the three protocols. The House medium promotes a relatively balanced composition, characterised by high proportions of Enterocytes (Absorptive - 21.0%, Early - 24.7%, Crypt top – 7%). Goblet Cells are moderately represented (17.0%), although other secretory lineages are nearly absent (EECs at 0.3% and Tuft Cells at 0.1%). This suggests that while House medium supports absorptive differentiation effectively, its capacity to sustain secretory or rare epithelial phenotypes is limited in the colon. The Fujii protocol exhibits a strongly proliferative bias, with TAs (33.1%) and Stem Cells (19.1%) comprising over half of the total population, with Enterocytes representing <40% of the overall population. This medium shows the highest representation of secretory cells in colonic organoids, including Goblet (9.9%), EECs (0.4%), and Tuft cells (2.1%). These findings point to a culture condition that maintains stemness and multipotency, while also supporting secretory lineage commitment more effectively than the other two protocols. The Holloway medium favours the most differentiated epithelial state in the colon, with substantial enrichment in terminally differentiated enterocytes (Crypt-top, 15.1%). Absorptive lineages comprise around 70% of the full epithelial landscape, with a marked shift towards M-Cell differentiation (9.9%). Regarding secretory lineages, Goblet Cells are moderately represented (4.9%), while Tuft Cells (0.4%) and EECs (0.2%) are present at very low levels. Notably, Holloway's capacity to maintain proliferative compartments is reduced, as reflected in the lower proportions of Stem and TA cells.

In TI-derived organoids, the inter-protocol differences become less pronounced, with all three showing much similar patterns of cell type representation. As in the colon, the House medium favours more Enterocyte differentiation, with the highest proportions of terminally differentiated enterocytes (28.3%), and much higher than that seen in the colon (7.0%). Interestingly, Stem Cell representation is conversely much reduced in TI vs colon (2.2% vs 17%),

highlighting a significant region-specific response. Goblet Cells were less abundant than in the colon, while EECs show an important increase from 0.3 to 1%. Compared to colon, in TI the Fujii protocol shows a shift away from stemness and proliferation and towards Enterocyte lineage commitment. Secretory lineages, including Goblet (8.2%), EECs (0.6%), and Tuft cells (0.7%) remain modestly represented, with a slight increase in EECs and reduction in Tuft. Nonetheless, Fujii yielded a more absorptive biased differentiation profile in TI vs colon. As for Holloway, it maintained a strong Enterocyte representation in TI, but interestingly the proportion of secretory versus absorptive cells increased vs colon, as shown by an apparent reduction in Early M-cell populations (10.0% to 3-3%) and increase in Goblet Cells (4.9% to 9.0%).

Overall, these results suggest that the choice of medium should depend on the specific experiment aim. The Fujii protocol consistently supports proliferative compartments in both regions, and excels in preserving secretory lineages — particularly Tuft Cells, and Goblet cells in the colon — making it the most effective protocol for rare secretory cell recovery. However, its differentiation bias weakens in TI, where absorptive enterocytes become more prominent. The Holloway protocol consistently supports the most complete differentiated absorptive state across both regions, particularly in TI, along with slight improvements in secretory lineage proportions (particularly M-Cells) and enhanced proliferative capacity (but with little EEC or Tuft Cell representation). Finally, the House medium offers a useful compromise by promoting strong absorptive lineage differentiation and maintaining some proliferative potential, especially in TI (and uniquely supporting relatively higher EEC recovery). In sum, media selection should align with the specific lineage or biological question being addressed: Fujii for rare secretory populations, Holloway for absorptive differentiation, and House for balanced output with an emphasis on terminal differentiation and regional fidelity.

1

Regarding medium performance in health versus disease (i.e. HC vs UC.INF and UC.NINF; Figure 73B), the House medium showed a relatively stable and consistent cell type profile across health states. In HC, House supports a balanced epithelial composition with strong Enterocyte representation (Absorptive: 12%; Crypt-Top: 15.1%; Early: 31.8%). TAs and Stem Cells together accounted for 33.5%, reflecting a healthy proliferative compartment. In UC.INF, the proportion of terminally differentiated enterocytes decreases slightly, although the overall Enterocyte proportion remains stable. There is a moderate increase in Goblet Cells (6.5% to 9.2%), but the few rarer cell types present in HC mostly disappear, potentially as a result of inflammation-induced epithelial shifts. In UC.NINF, cell profiles are more resemblant of the inflamed state vs HC, but show the highest proportions of TA/Stem Cells, Goblet Cells, and Early M-cells, and the lowest representation of Enterocytes. This pattern possibly indicates delayed priming of more specialized antimicrobial populations or effector cells, and primed proliferation, but with lower representation of cells specialising in nutrient uptake under homestatic conditions, potentially representing early adaptations to a heightened (but not yet overwhelming) inflammatory state. In sum, the House medium maintains a robust epithelial balance and reflects subtle disease-associated alterations in lineage composition, particularly through a moderate expansion of Goblet Cells and proliferative/multipotent populations, a shifting absorptive-to-proliferative balance.

By contrast, Fujii exhibits strong sensitivity to inflammatory status, with distinct shifts in epithelial composition across conditions. In HC, it promotes a modest absorptive and secretory representation. Nonetheless, for example, Goblet and Tuft Cells represent 9.1% and 2.2%, respectively, indicating preserved secretory potential. In UC.INF, proliferative cell types become even more dominant (almost 75% of the overall population) and with lower representation of Enterocytes (41.9% to 32%), Tuft Cells (2.2% to 0.2%), and EECs (0.5% to 0.2%). However, Early

M-cells and Goblet Cells increase (0.2% to 1.0%, and 9.1% to 11.5%), suggesting a possible antimicrobial response. Curiously, and contrarily to what is seen in House, UC.NINF showed a less immune-related profile, with early M cells disappearing and Goblet Cells returning to HC levels, while Tuft Cells show a small increase (0.2% to 0.6%). Enterocyte populations are even more reduced than in UC INF, and the few EECs disappear. Overall, this pattern may reflect inflammation-induced epithelial regeneration and delayed differentiation, instead of the more active anti-inflammatory response seen in House. In sum, Fujii medium also amplifies proliferative phenotypes under inflammatory conditions, as seen with House. While it performs well for retaining secretory lineages, it does so at the cost of balanced differentiation, especially absorptive lineages.

Finally, Holloway showed a stark inflammation-induced shift, but with an opposite behaviour to the former media. In HC, Holloway shows an overall pattern similar to House, with good representation of absorptive populations, especially Enterocytes, and a moderate undifferentiated fraction (Stem Cells and TAs). Tuft Cells and EECs are almost negligible, but there is a modest fraction of M-Cells (3.8%) and Goblet Cells. In UC.INF, the fraction of multipotent/cycling cells drops abruptly, and all secretory lineages disappear. Most cells (49.4%) are instead driven towards Early M-Cell differentiation, highlighting an extreme activation of epithelial immune effector cells. In UC.NINF, there is an apparent intermediate behaviour between HC and UC.INF, with an expansion of Early M-cell population (12.7%), but predominance of Enterocytes (75.9%) and some preservation of proliferative populations (9.0%) and Goblet Cells (2.4%). In sum, Holloway promotes an exacerbated Early M-Cell expansion under inflammation, while limiting proliferative skewing.

Overall, this analysis highlighted marked differences in cell composition response to inflammation under different culture protocols. House and Fujii promoted a shift towards a

more reparative phenotype, by increasing proliferation but also slightly expanding cells related with antimicrobial activity (with Fujii retaining more secretory cell types). Conversely, Holloway seemed to promote an exaggerated shift towards an immune priming or inflammatory profile dominated by M-like cells, while reducing both absorptive and proliferative lineages. As a result, Fujii seems most suitable for modelling secretory behaviour or epithelial plasticity in inflammation. In turn, Holloway is preferred for robust absorptive differentiation and M-cell priming (although the responses seem may be too extreme and lead to poor cell recovery). Finally, House performs best for maintaining balanced differentiation and epithelial integrity.

5.2.6.2 Impact of Culture Medium on hIEO Response to RANKL Stimulation

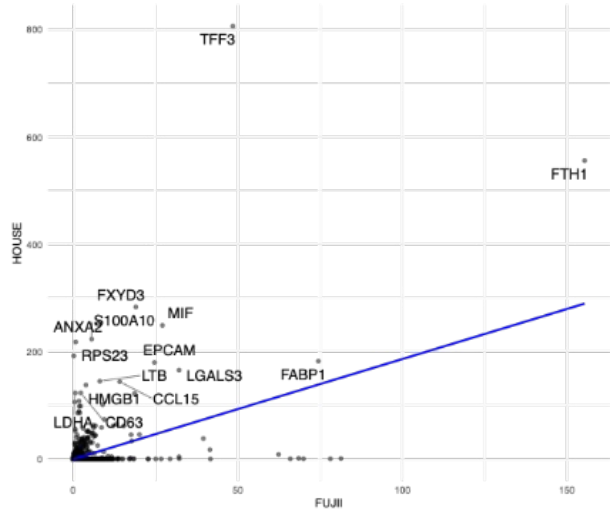
Another objective of EXP03 was to compare how different culture media (Fujii vs Holloway) influenced hIEO response to RANKL stimulation - used here as differentiation driver for a rare cell type (M-Cells) - while aiming to maintain balanced epithelial compositions. As RANKL stimulation had previously been applied under House conditions (EXP01) – as assessed in Section 5.2.2.3 – no direct donor-matched comparisons could be performed here. To first explore overall differences in transcriptional landscape, a comparative gene expression analysis was performed by combining all donor samples cultured in House media supplemented with RANKL (from EXP01), and comparing these with all donor samples cultured in Fujii or Holloway media supplemented with RANKL (from EXP03). As shown in Figure 74A, differentially-enriched genes in House when compared to either Fujii and Holloway were mostly stress-associated *(S100A10, ANXA2, HMGB1, FTH1)*, immune modulatory *(MIF, LTB)* and differentiation-related genes *(TFF3, FABP1)*. However, this difference was more accentuated in House vs Holloway when compared with House vs Fujii (as shown by larger differences in transcript numbers).

As some of these patterns could be at least partly due to different donor profiles, Figure 74B presents a genome-wide correlation analysis of $\log_2$ fold changes in gene expression between the Fujii and Holloway media, each relative to House medium. The tight clustering along the identity lines indicates high agreement in transcriptional profiles between Fujii and Holloway conditions, suggesting that both media elicit largely overlapping gene expression responses when benchmarked against House. However, several genes deviate from the line of identity, highlighting media-specific effects. For example, genes upregulated in Fujii but not Holloway (bottom right quadrant) include *IL23R, S100A7, APOA2, SH2D1B,* and *TNFSF4* — many of which are inflammation-related or associated with immune cell recruitment. By contrast, genes upregulated in Holloway but not Fujii (top left quadrant) include *NOTCH2NLB, MYCN*, and

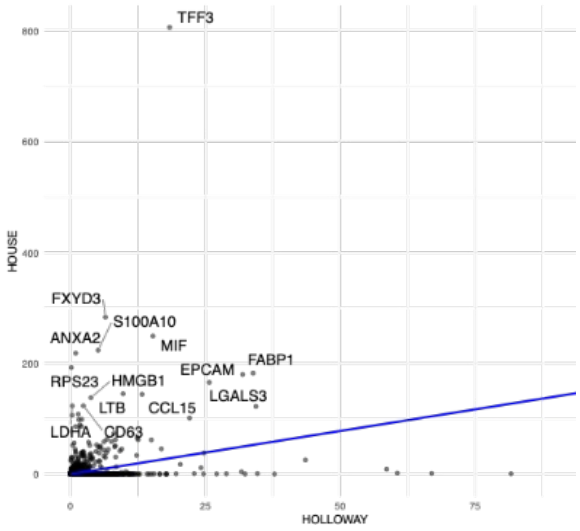
FCRL2, which are associated with absorptive lineage specification, progenitor states, and immune modulation. Overall, the data suggest that while the overall transcriptional impact of RANKL is conserved across media, certain immune and epithelial response genes diverge, potentially due to differences in baseline epithelial state or medium-specific cues affecting cell fate dynamics.

Figure 74 - Comparative analysis of gene expression in hIEOs across different media

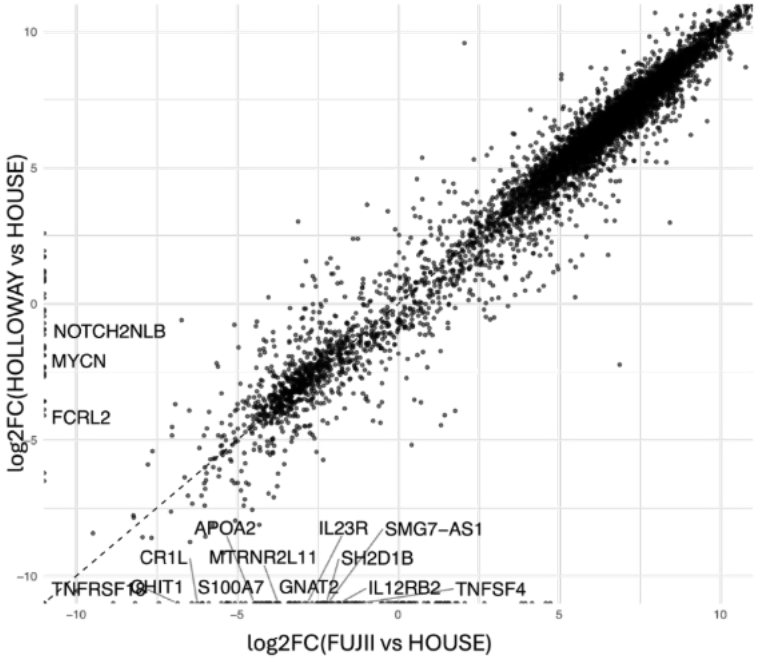
A) House vs Fujii medium



B) House vs Holloway medium



C) Correlation of log₂ fold changes between Fujii and Holloway relative to House medium



Legend: Transcript abundance across different media conditions in hIEOs. Panels A and B show raw expression values for selected genes in House medium versus Fujii (A) or Holloway (B), with key upregulated genes labelled. The blue regression line indicates overall trend direction, highlighting divergence in expression for genes such as TFF3, FTH1, FABP1, and EPCAM between media types. Panel C displays the correlation of log₂ fold changes for all genes comparing Fujii vs House (x-axis) and Holloway vs House (y-axis). Genes along the diagonal are similarly regulated in both comparisons, while those deviating from the diagonal (e.g. IL23R, NOTCH2NLB, FCGRL2) exhibit medium-specific transcriptional differences.

Having established that no substantial overall differences emerged with regards to the impact of RANKL stimulation across Fujii and Holloway media when compared with House medium, a more detailed exploration of the differential impact of each media across donor settings was performed. In the colon, marked M-Cell expansion was seen under both media conditions (Figure 75A). In Fujii, M-Cells account for 44.1% (vs 0.3% in baseline conditions; Figure 73A), with the remaining epithelial landscape represented by Enterocytes (Absorptive: 21.7%; Crypt-Top: 17.%1; Early: 0.1%) and proliferative populations (Stem-Cells: 4.8%; TAs: 11%). Secretory lineages (Goblet cells, EECs, Tuft Cells) were nearly absent (0.1% each). In Holloway, M-Cells were slightly less abundant (39.2%, vs 10% in baseline) while Enterocytes where more frequent (especially terminally differentiated Crypt-Top enterocytes; 44.2% of total). However, proliferating and secretory populations were either negligible or absent, respectively.

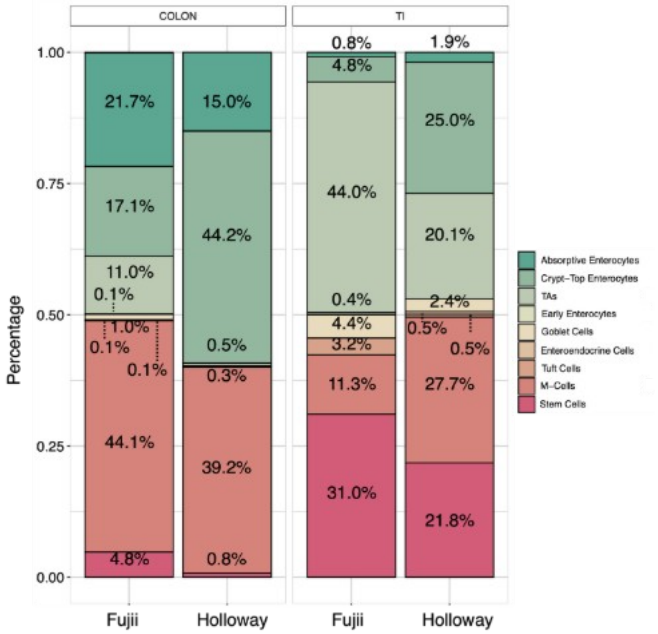
In TI, M-Cell induction was substantially less accentuated (Fujii: 11.3% RANKL vs 0.8% baseline; Hollway: 27.7% RANKL vs 3.3% baseline). This contrasts with what could be expected based on physiological, *in vivo* conditions, where M-Cells are a well-established population in TI (and mostly associated with Peyer's Patches) but almost not seen in the colon (without stimulation). The overall epithelial landscape was much broader in the RANKL-stimulated TI vs colon, with Fujii showing a modest proportion of Goblet Cells (4.4%), and Tuft Cells (3.2%), no EECs, and a very high proportion of Stem Cells and TAs (75%). Holloway originated a much more prominent M-Cell population vs Fujii, while retaining several secretory populations - but at lower levels than Fujii (Goblet Cells: 2.4%; EECs: 0.5%; Tuft Cells: 0.5%) – and a significant fraction of Stem Cells and TAs (41.9%) which were absent in colon.

In conclusion, this analysis showed that regional context was critical in determining the impact of RANKL stimulation on the epithelial landscape. M-Cell induction was far more effective in the colon for both media, with Fujii preserving some proliferative populations while Holloway

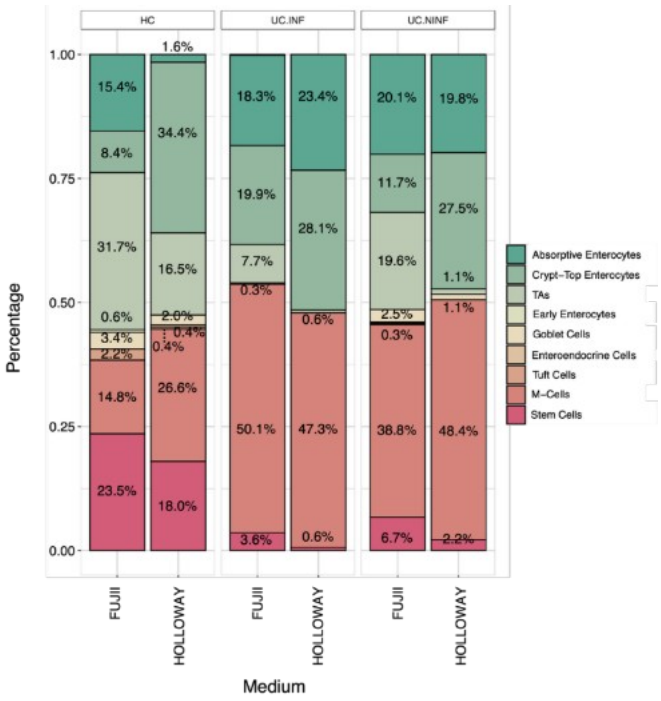
prioritises enterocyte differentiation. While more nuanced, the response in TI yielded a more balanced epithelial landscape, which is more consistent with physiological distribution and regional susceptibility. When compared with Holloway, Fujii generated a lower M-Cell fraction, but with a broader spectrum of epithelial lineages (including secretory subtypes) and higher proliferative capacity. Comparison of transcriptomic profiles of M-Cells from colon vs TI did not show significant differences in expression of canonical M-Cell markers, while indicating that colonic M-Cells adopt a TI-regional M-Cell identity when exposed to RANKL stimulation (data not shown). Overall, for maximising RANKL-induced M-Cell differentiation, while keeping a decent representation of a broader epithelial landscape, Fujii medium is therefore preferred in both TI and colon.

Figure 75 - Effect of medium, intestinal region, and inflammation on epithelial lineage composition in colon organoids subject to RANKL stimulation

A) Comparison across locations



B) Comparison across disease states



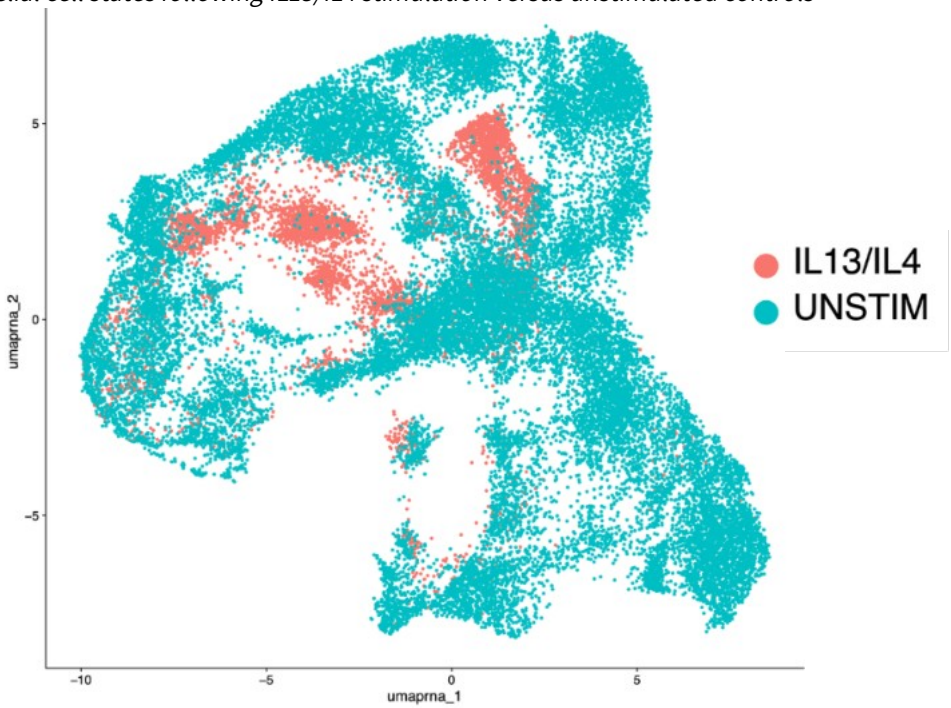
Legend: Proportional distribution of epithelial cell types in colon-derived organoids across different culture media (Fujii, Holloway), intestinal regions (left vs right colon), and inflammatory states (healthy control, UC inflamed, UC non-inflamed). Fujii medium favours progenitor-rich populations such as stem and TA cells, particularly in inflamed UC samples. By contrast, Holloway medium supports higher proportions of differentiated lineages including absorptive and crypt–top enterocytes. Inflamed UC-derived organoids exhibit reduced differentiation and increased progenitor expansion, while non-inflamed and healthy samples retain more balanced epithelial compositions. These data highlight how culture conditions and disease state jointly influence cell fate output in colon organoids.

Figure 75B compares the effect of RANKL stimulation under both culture media across different health contexts. Overall, there is a decrease in Stem Cells for UC.NINF and UC.INF in both media when compared with HC (but more accentuated in UC.INF), along with an increase in M-Cells (of a similar extent in UC.NINF and UC.INF for Holloway, but more marked in UC.INF vs UC.NINF for Fujii). Other differentiated populations were overall substantially decreased in UC states across both media. Broadly, the impact of RANKL stimulation in disease states seemed to reflect that seen for both media without stimulation (i.e. reduced cell diversity) but with substantially higher M-Cell representation. The latter was potentially already primed in Holloway without RANKL (Figure 73B), while in Fujii it might have replaced the exaggerated increase in proliferative cells seen earlier. In sum, Fujii seems best suited for experiments requiring Stem Cell maintenance, modelling of inflammation-induced plasticity, and detection of rare epithelial lineages (including M-Cells, EECs, and Tuft Cells), particularly in healthier states (UC.NINF, HC), while Holloway is best at preserving overall absorptive/enterocyte identity and maintaining structural fidelity under disease stress (but suboptimal for Stem Cell expansion and rare cell differentiation).

5.2.7 Impact of IL13/IL4 Stimulation on the Emergence of Secretory Cell Populations

In EXP06, perturbation of a single adult healthy colon sample with the type 2 cytokines IL-13 and IL-4 was employed as a tentative approach to engage SPIB pathway signalling, with the aim of driving differentiation of Best4 Cells or rare secretory cell types. This stimulation resulted in a distinct redistribution of epithelial cell states (Figure 76) – when compared to unstimulated controls (blue), IL-13/IL-4-stimulated cells (red) occupy new or expanded transcriptional niches, reflecting cytokine-driven epithelial plasticity.

Figure 76 - Epithelial cell states following IL13/IL4 stimulation versus unstimulated controls



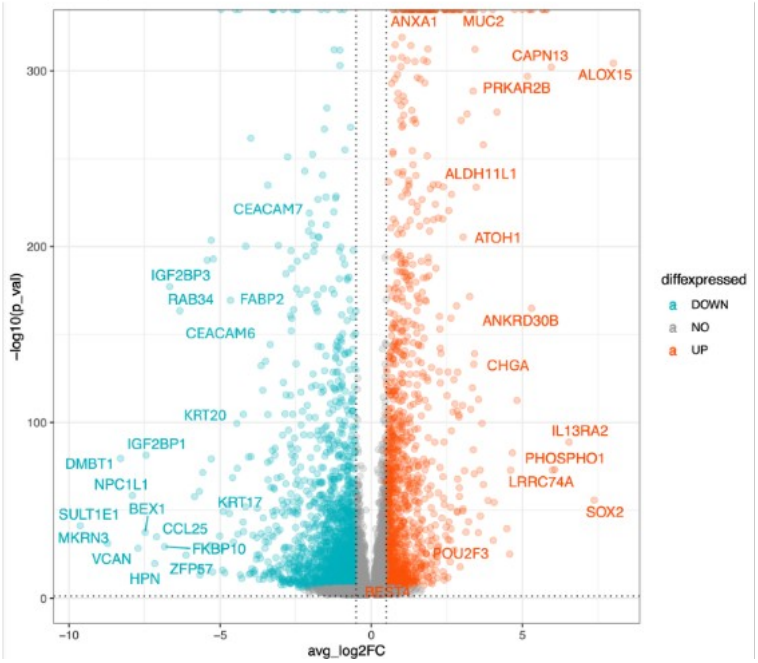
Legend: Single-cell transcriptomes from human intestinal epithelial organoids cultured under either IL13/IL4 stimulation (pink) or unstimulated (turquoise) conditions. Stimulation with IL13/IL4 leads to a redistribution of cells across transcriptional space, suggesting the emergence or expansion of distinct epithelial states. Several regions show marked enrichment for cytokine-stimulated cells, consistent with activation of type 2 inflammatory or goblet-cell–biased programmes.

As shown in Figure 77A, upregulated genes (orange) in the stimulated sample included *MUC2, ANXA1, IL13RA2, POU2F3*, and *ATOH1* - markers associated with secretory lineage priming, namely Tuft and Goblet Cell differentiation, type 2 immune signalling, and mucosal remodeling.

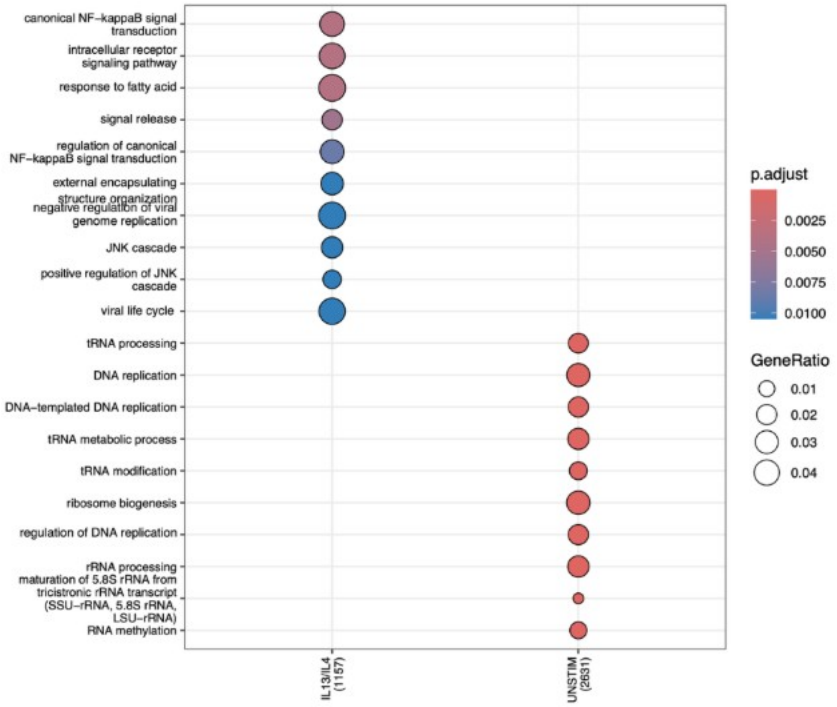
Notably, *MUC2* and *ATOH1* underscore a strong secretory lineage activation, while *IL13RA2* reflects cytokine responsiveness. Conversely, downregulated genes (blue) such as *CEACAM6, FABP2, KRT17*, and *IGF2BP1* highlight suppression of absorptive enterocyte programs and epithelial homeostasis. The overall gene expression landscape suggests IL13/IL4-driven induction of specialised programs, likely involving type 2 inflammatory responses or secretory lineage commitment, such as goblet cell bias. The emergence of these distinct clusters underscores the potent effect of IL-13/IL-4 on reshaping the epithelial landscape, potentially mimicking immune-epithelial interactions occurring during allergic inflammation, parasitic infection, or mucosal barrier modulation.

Accordingly, GO enrichment analysis revealed distinct biological process activation profiles (Figure 77B). Upregulated genes are significantly enriched in pathways linked to canonical NF-κB signalling, intracellular receptor signalling, and JNK cascade activation, all of which are hallmarks of type 2 immune responses and epithelial stress adaptation. Additionally, processes such as signal release and response to fatty acids indicate broad reprogramming of secretory and metabolic functions. By contrast, unstimulated cells show enrichment in fundamental cellular processes, including DNA replication, rRNA and tRNA processing, and ribosome biogenesis, reflecting a baseline proliferative and translational state. These findings underscore the potent immunomodulatory effects of IL13/IL4, shifting epithelial programs from homeostasis toward inflammation-induced secretory and immune-responsive states.

Figure 77 - Transcriptomic and pathway-level responses of epithelial organoids to IL13/IL4 stimulation
A) Differentially expressed genes in IL13/IL4-stimulated versus unstimulated conditions



B) GO enrichment of biological processes associated with IL13/IL4 upregulated genes



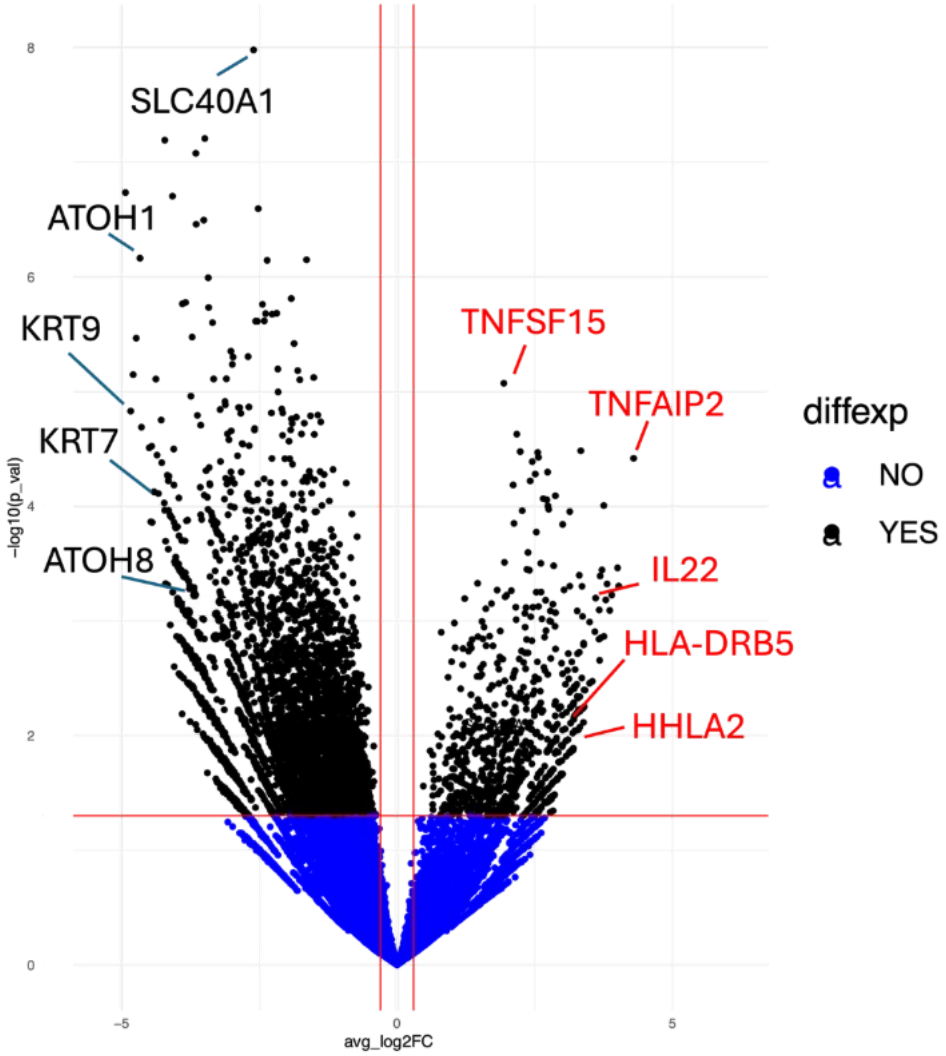
Legend: Panel A shows differential gene expression in intestinal epithelial organoids exposed to IL13/IL4 stimulation versus unstimulated controls. Upregulated genes (red) include goblet-cell markers (MUC2, SPDEF, TFF3), type 2 inflammation genes (IL13RA2, ALOX15, ANXA1), and differentiation regulators (ATOH1, CHGA), indicating secretory lineage bias. Downregulated genes (blue) include absorptive and barrier-associated genes (CEACAM6, KRT20, IGFBP3). Panel B shows gene ontology enrichment for IL13/IL4-upregulated genes, highlighting NF-kB signalling, cytokine response, and RNA metabolism. These findings indicate robust transcriptional reprogramming and secretory skewing under type 2 cytokine stimulation.

5.3 Integrated Transcriptomic and Epigenomic Profiling (Multiome)

EXP05 (Figure 25) was designed to enable simultaneous assessment of transcriptomic and epigenomic landscapes (scMultiome) in a multiplexed approach across multiple disease states in adult colon samples, allowing the exploration of more nuanced cell differentiation trajectories. To ensure that inflammatory signatures were adequately retained in this experimental setup, and identify suitable use-cases to explore epigenetic priming in these conditions, differentially expressed genes in inflamed UC (UC.INF) were compared to healthy controls (HC) (Figure 78). Several inflammation- and immune-associated genes are significantly upregulated in inflamed UC epithelium: *TNFSF15* and *TNFAIP2*, classical pro-inflammatory genes involved in TNF signalling pathways; *IL22*, a cytokine critical in mucosal immunity and epithelial regeneration; *HLA-DRB5* and *HHLA2*, immune-regulatory genes, indicating heightened antigen presentation and potential T cell interaction in the UC microenvironment. These upregulated genes reflect a robust immune activation and epithelial-immune crosstalk during active UC inflammation. By contrast, several genes downregulated in UC suggest loss of differentiation or epithelial barrier function: *ATOH1* and *ATOH8*, transcription factors essential for secretory lineage differentiation, including Goblet and Tuft Cells; *KRT7* and *KRT9*: cytokeratins typically associated with epithelial structure and stratification, commonly found in maturing/matured enterocytes; *SLC40A1*, involved in iron transport and epithelial homeostasis. These patterns indicate impaired differentiation and epithelial integrity in UC inflammation, consistent with known disease pathology. In sum, this analysis highlights that inflamed UC epithelium exhibits a dual signature, marked by upregulation of immune mediators (particularly those involved in cytokine signalling and antigen presentation) and downregulation of key epithelial identity and differentiation markers (particularly secretory transcription factors and structural proteins).

These findings support the view that inflammation disrupts epithelial lineage fidelity while amplifying immune-epithelial signalling.

Figure 78 - Differential gene expression in inflamed UC versus healthy epithelial samples

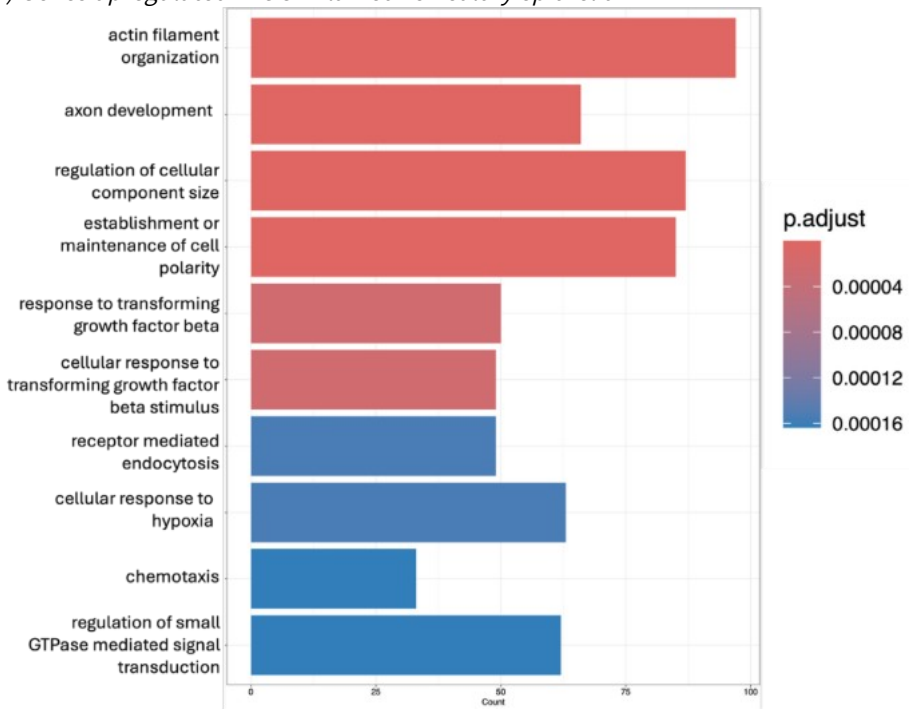


Legend: Differentially expressed genes between inflamed UC and healthy intestinal epithelial samples derived from the multi-condition, multiplexed hIEO single-cell RNA-seq dataset. Selected genes enriched in the UC inflamed condition (e.g. TNFSF15, TNFAIP2, IL22, HLA-DRB5, HHLA2) are highlighted in red, while genes downregulated in inflammation (e.g. KRT7, KRT9, ATOH1, SLC40A1) are shown in teal. These results underscore key immune-epithelial shifts and disruption of absorptive/secretory lineage markers in active inflammation.

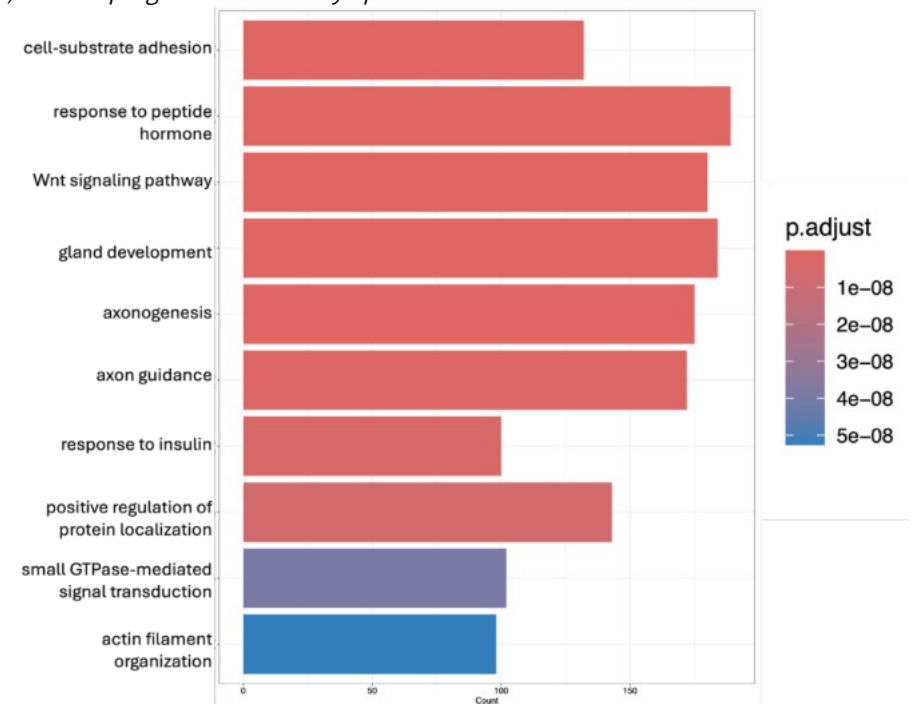
GO analysis highlighted biological processes enriched among genes upregulated in UC-inflamed epithelium, pointing particularly to inflammatory and stress-responsive pathways, with top relevant processes including “Actin filament organisation”, “Cell polarity and component size regulation”, "Transforming growth factor beta (TGF-β) response”, and “Cellular response to hypoxia" (Figure 79A). These pathways indicate cytoskeletal remodelling, stress responses, and immune signalling during inflammation. The enrichment of TGF-β-related terms and receptor-mediated endocytosis further supports a role for inflammatory remodelling and epithelial-immune crosstalk in UC. By contrast, biological processes suppressed in UC-inflamed tissue, or more active in healthy epithelium, included “cell-substrate adhesion”, “Wnt signalling”, “response to peptide hormone” and “positive regulation of protein localization” (Figure 79B). These pathways are consistent with epithelial integrity, morphogenesis, and homeostatic signalling, all of which are known to be disrupted during UC flares. The differences seen in Wnt signalling and gland development suggest impaired epithelial renewal and functional maturation in disease.

Taken together, these results confirm a clear inflammatory signature in UC epithelium, characterised by activation of remodelling, chemotactic, and hypoxia-associated programmes, and suppression of differentiation, adhesion, and developmental processes. This pattern reflects a shift from stable epithelial maintenance to an injury-responsive state, with potential consequences for barrier function, regeneration, and lineage fidelity.

Figure 79 - Gene ontology enrichment of biological processes altered in ulcerative colitis inflammation
A) Genes upregulated in UC inflamed vs healthy epithelium



B) Genes upregulated in healthy epithelium vs UC inflamed



Legend: enriched Gene Ontology (GO) biological processes among differentially expressed genes between inflamed UC and healthy epithelial samples. Panel A highlights pathways upregulated in inflammation, including actin filament organisation, axon development, and TGF-β signalling response, while Panel B presents pathways suppressed in UC inflammation, such as cell-substrate adhesion, Wnt signalling, and gland development. Bars are coloured by adjusted p-value, with red tones indicating stronger enrichment.

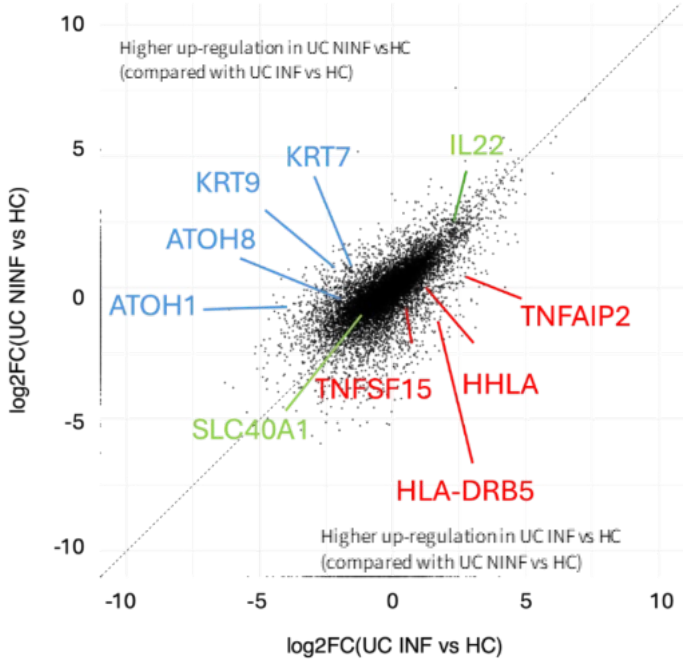
These results reveal extensive cytoskeletal remodelling, altered morphogenesis, and suppression of homeostatic epithelial programmes during inflammation.

Building on the observations above, transcriptomic and chromatin accessibility profiles were compared across UC states by focusing on the differentially-expressed genes highlighted in Figure 78. Transcriptomic comparison (Figure 80A) showed clear up-regulation of inflammatory pathways in UC.INF vs HC when compared with UC non-inflamed (UC.NINF) vs HC - namely *TNFSF15*, *TNFAIP2*, *HHLA2*, and *HLA-DRB5* (right-shifted, in red. By contrast, epithelial identity and differentiation markers - such as *ATOH1*, *ATOH8*, *KRT7*, and *KRT9* (left-shifted, in blue) – are downregulated to higher degree in UC.INF vs HC when compared with UC.NINF vs HC. This highlights a transcriptional signature of active inflammation, with pronounced immune activation and suppressed epithelial lineage programmes, which are at least partially preserved in non-inflamed tissue. Interestingly, *IL-22* and *SLC40A1* were differentially expressed to a similar degree between UC.INF and UC.NINF when compared with HC (the first upregulated, the second downregulated), suggesting a role in driving early inflammatory or immune-mediated activation (before overt pathophysiological changes become apparent) – or else reflecting some restoration of a more homeostatic landscape.

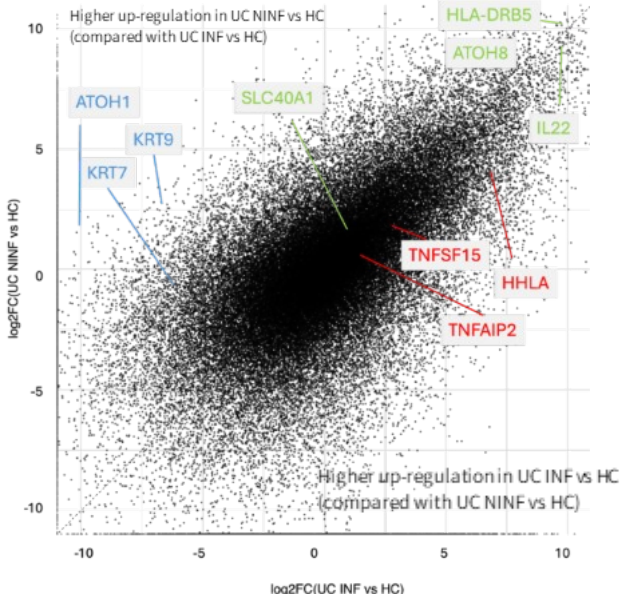
The chromatin landscape echoes the transcriptional trends, but also reveals priming effects. For example, genes like *TNFSF15*, *TNFAIP2*, and *HHLA2* are both more expressed and more accessible in inflamed vs non-inflamed states, suggesting increasing epigenetic activation and perpetuation or amplification or inflammation. By contrast, *ATOH1*, *KRT7*, and *KRT9* are both less expressed and less accessible in UC.INF, reflecting persistent downregulation and disruption of epithelial homeostasis. *IL22* and *SLC40A1* show comparable degrees of accessibility in both UC.INF and UC.NINF, reflecting the similar transcriptional activation seen above. However, interestingly, HLA-*DRB5* and *ATOH8* showed similar chromatin accessibility in

UC.NINF vs UC.INF, despite respectively higher and lower transcriptomic activity in UC.INF when compared with UC.NINF – suggesting epigenetic priming in the absence of overt inflammation that will eventually drive increased immune activation and loss of differentiation capacity (or lasting imprinting or ‘memory’ of inflammation that could influence relapse or treatment response). These findings support the hypothesis that epigenomic priming precedes and possibly predisposes transcriptomic reactivation, offering mechanistic insight into UC persistence, flare risk, and therapeutic targeting.

Figure 80 - Comparative transcriptomic and chromatin accessibility changes in inflamed versus non-inflamed UC epithelium relative to healthy controls
A) Transcriptomic comparison



B) Chromatin accessibility comparison



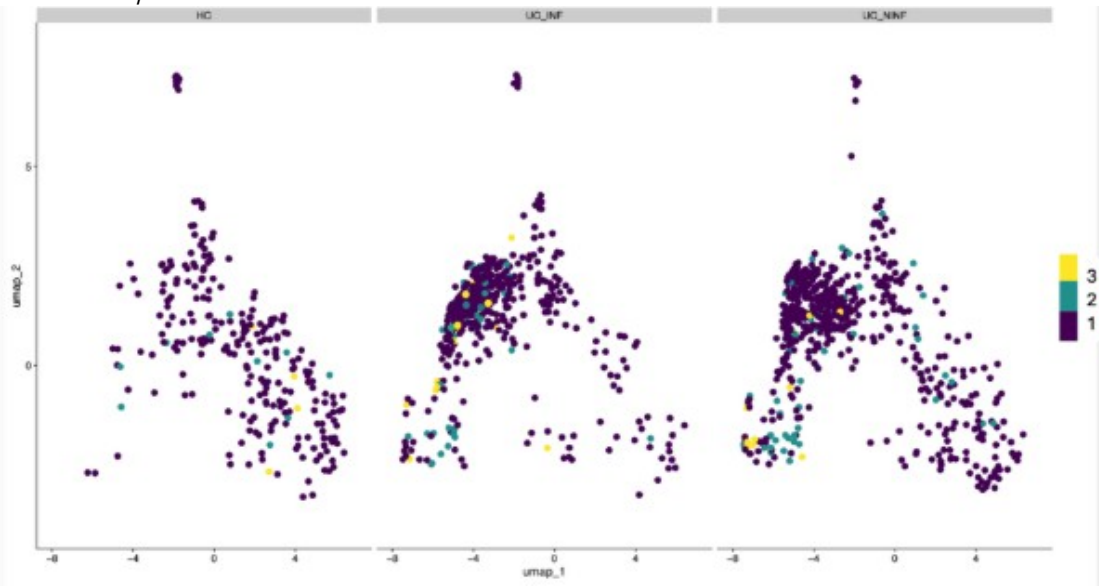
Legend: Differential regulation between inflamed UC (UC INF) and non-inflamed UC (UC NINF) relative to healthy control (HC) epithelium, across transcriptomic (Panel A) and chromatin accessibility (Panel B) profiles. The x-axis indicates log_2 fold change in UC INF vs HC, while the y-axis shows log_2 fold change in UC NINF vs HC. Genes or peaks along the diagonal show similar regulation in both UC areas; deviations suggest changes specific to inflammatory status. Inflammatory mediators (TNFSF15, TNFAIP2, HLA-DRB5, HHLA2, IL22) are selectively enriched in inflamed tissue, while epithelial lineage and absorptive markers (KRT7, KRT9, ATOH1, ATOH8, SLC40A1) are depleted. However, some enriched inflammatory markers

(such as HLA-DRB5) have similar chromatin level accessibility, suggesting genomic tissue priming that precedes transcriptomic signatures.

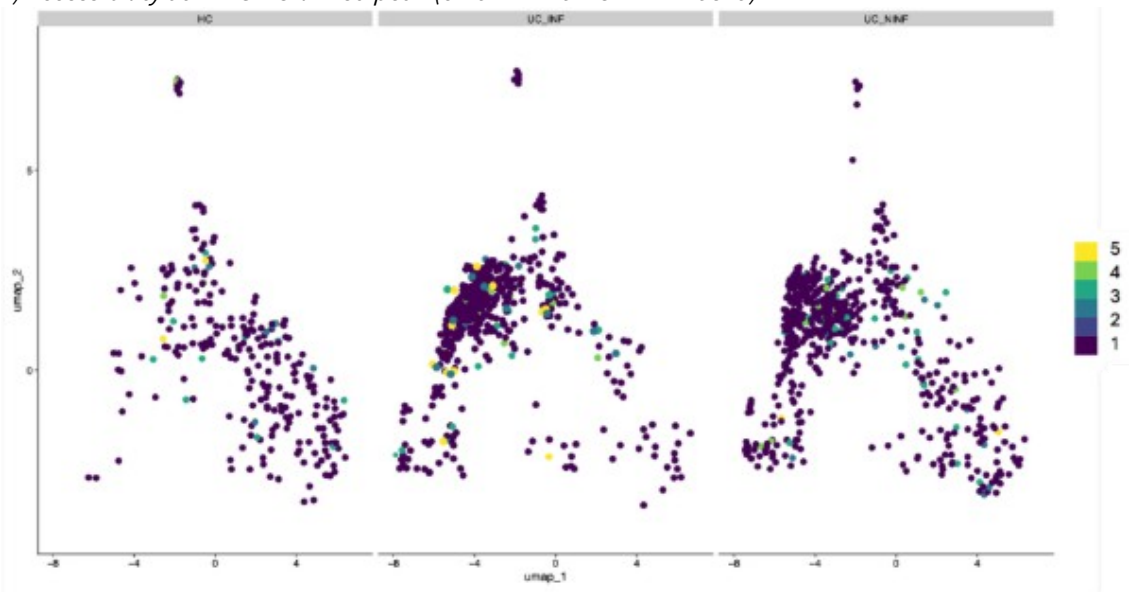
Figure 81 collectively elucidates how *TNFSF15*, a key inflammatory cytokine implicated in UC, is regulated at both transcriptional and chromatin accessibility levels. *TNFSF15* expression is markedly elevated in UC.INF samples, with a slightly reduced but still detectable signal in UC.NINF compared to HC (Figure 81A). Expression is not largely restricted to specific epithelial subsets, implying an overall inflammatory background in the whole epithelium. Accordingly, ATAC-seq signal across the peak near chr9:114779175–114779829 — annotated as a putative regulatory enhancer of *TNFSF15* — shows a stark contrast between healthy and disease states (Figure 81B). Both UC.INF and UC.NINF exhibit higher accessibility at this locus compared to HC, strongly indicating active regulatory priming in diseased tissues. Notably, UC.NINF retains accessibility even without overt inflammation, suggesting pre-conditioning of the epithelial genome, or “epigenetic memory”. Figure 81C integrates gene tracks and ATAC-seq data, revealing that UC-specific peaks (boxed sections) are either entirely absent or significantly weaker in healthy controls. Specifically, UC.INF samples show the most robust accessibility at the *TNFSF15* regulatory region, correlating well with transcriptional upregulation. UC.NINF displays intermediate accessibility, reinforcing the notion that epigenetic changes persist beyond active inflammation, potentially priming the tissue for rapid reactivation upon future inflammatory stimuli. These data support a model wherein *TNFSF15* transcription in UC is tightly linked to enhancer accessibility, with chromatin-level priming preceding or accompanying transcriptional induction. The presence of epigenetic activation in non-inflamed UC suggests persistent rewiring of epithelial regulatory networks. This highlights *TNFSF15* not just as a downstream inflammatory effector, but also as a marker of durable regulatory reprogramming, potentially relevant for disease relapse or chronicity.

Figure 81 - Multiomic analysis of TNFSF15 gene expression and regulatory accessibility in UC-derived epithelial organoids

A- TNFSF15 expression across conditions

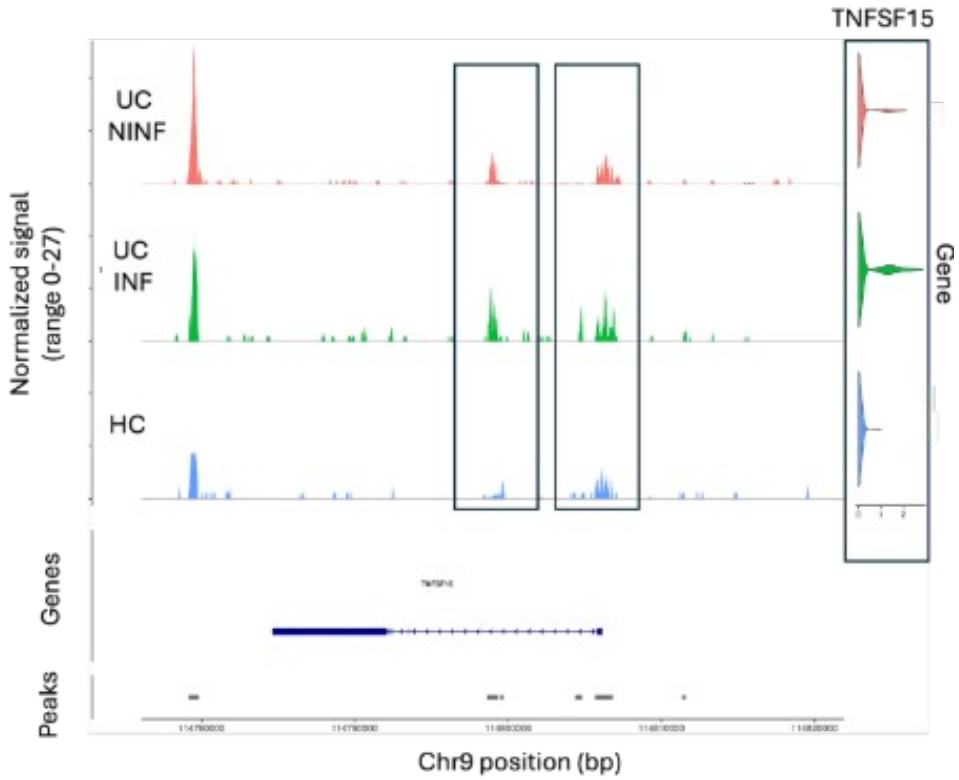


B) Accessibility at TNFSF15-linked peak (chr9-114779175-1144779829)



1

C) ATAC-seq signal at TNFSF15 locus



Legend: Integrated multiomic visualisation of TNFSF15 transcription and local chromatin accessibility in epithelial cells derived from healthy controls (HC), inflamed UC (UC.INF), and non-inflamed UC (UC.NINF) regions. The top panel shows TNFSF15 gene expression projected onto UMAP space, with elevated expression observed in discrete epithelial subsets from UC.INF and UC.NINF samples. The middle panel displays chromatin accessibility at a regulatory peak near TNFSF15 (chr9:114797176–114797829), showing similar localisation and intensity patterns, suggestive of regulatory activation. The bottom panel presents aggregated ATAC-seq signal tracks at the TNFSF15 locus across conditions. Highlighted regions indicate UC-specific regulatory peaks absent or less accessible in healthy epithelium. Together, these data support a model in which UC-associated inflammatory signalling involves transcriptional upregulation of TNFSF15 driven by condition-specific chromatin accessibility at upstream regulatory elements.

5.4 Comparing Organoid Cell States with Primary Epithelial Phenotypes

Throughout this work, multiple experimental approaches were developed with the aim of recapitulating, enhancing, or uncovering the presence and role of Best4 Cells, a rare intestinal cell type with yet poorly characterised function, but which has been implicated in biological processes ranging from fetal development to pathological conditions. Although the initial aim of this project could not be completely fulfilled – i.e. the detection of Best4 Cells remained sparse and elusive – the work developed led to the engineering of novel experimental and analytical approaches, and the refinement of tailored organoid culture conditions.

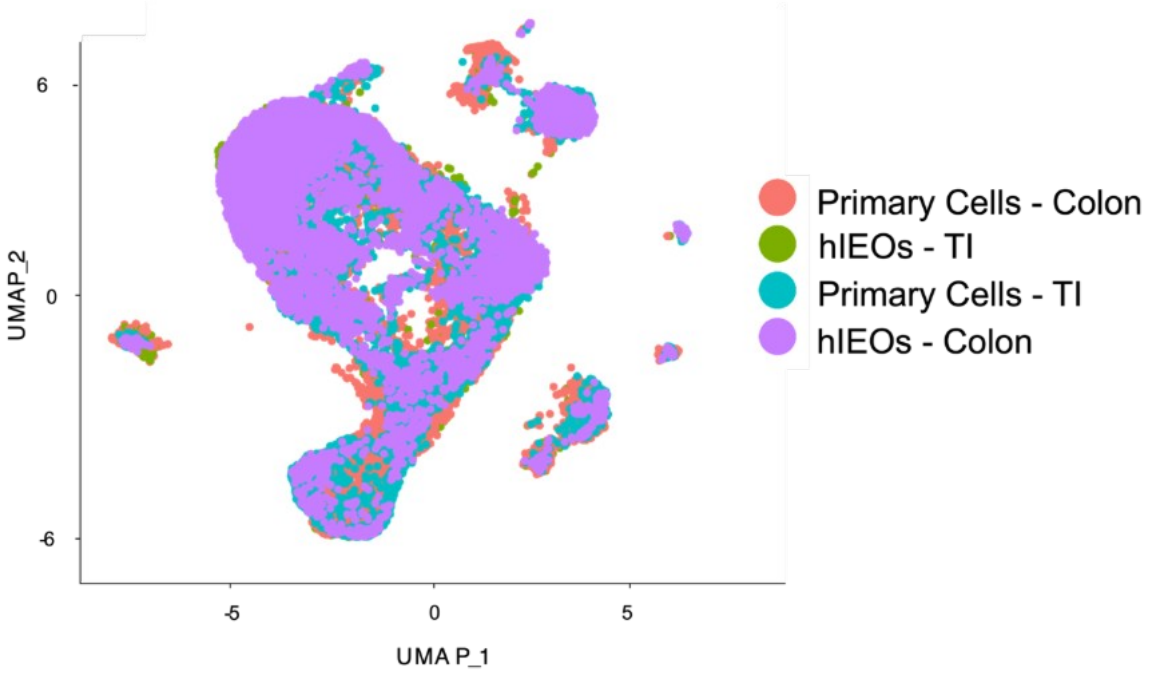
The analyses shown in previous sections established that hIEOs developed under either House, Fujii, or Holloway protocols preserved broad regional, pathophysiological, and developmental signatures associated with the donor tissues from which they derived. Nonetheless, a direct comparison of transcriptomic features between hIEOs and primary tissue was necessary to firmly establish the fidelity of the proposed experimental frameworks in recapitulating *in vivo* conditions, concluding this work with robust external validation of the methodology developed and its contributions to subsequent research.

For this purpose, a subset of healthy TI and colon hIEOs cultured under the three different protocols under study (from EXP03; Figure 20) were benchmarked through integration with well-established scRNAseq healthy intestinal datasets from primary tissue from Smilie *et al.* (2019) and Burclaff *et al.* (2022). This analysis allowed a broad evaluation of the organoid systems developed when compared with the biological conditions they aim to replicate, with the added benefit of high cell numbers (and therefore improved resolution) and enhanced representation of multiple cell types resulting from the combination of hIEOs cultured under the 3 different protocols (as shown in Section 5.2.3.1).

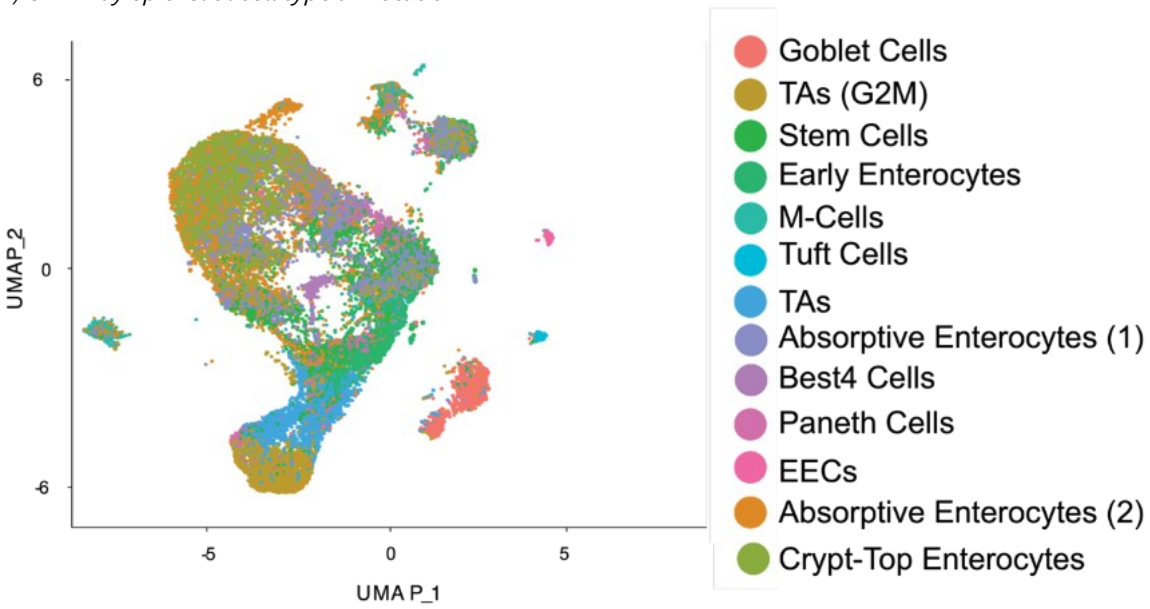
The combined transcriptional landscape revealed an exceptionally close correspondence between organoid-derived epithelial states and their primary tissue counterparts (Figure 82). Specifically, there is a near-complete phenotypic convergence between human intestinal epithelial organoids (hIEOs) and primary cells from both colon and TI (Figure 82A). This almost perfect match reflects the high fidelity of hIEOs in preserving location-specific epithelial identities post-integration, despite their *in vitro* expansion and differentiation. The fact that TI- and colon-derived organoids occupy nearly identical transcriptional manifolds as their tissue-derived counterparts underscores the effectiveness of the culture protocols and validates the use of organoids as physiologically relevant surrogates for *in vivo* epithelium. Likewise, annotation by canonical epithelial lineages demonstrates successful capture of diverse cell types, including Goblet Cells, TAs, Enterocytes, and rare secretory phenotypes like Tuft, M-Cells, EECs and – importantly - Best4 cells (Figure 82B).

Taken together, these findings strongly reinforce prior observations made throughout this study — from media-induced differentiation capacity to location-specific lineage representation and disease modelling — collectively demonstrating that intestinal organoids retain the essential architectural and molecular signatures of primary epithelium. This alignment justifies their utility in mechanistic and translational studies, offering a faithful and scalable proxy for human intestinal biology.

Figure 82 - Epithelial cell type annotation and culture format distribution
A) UMAP by sample type



B) UMAP by epithelial cell type annotation



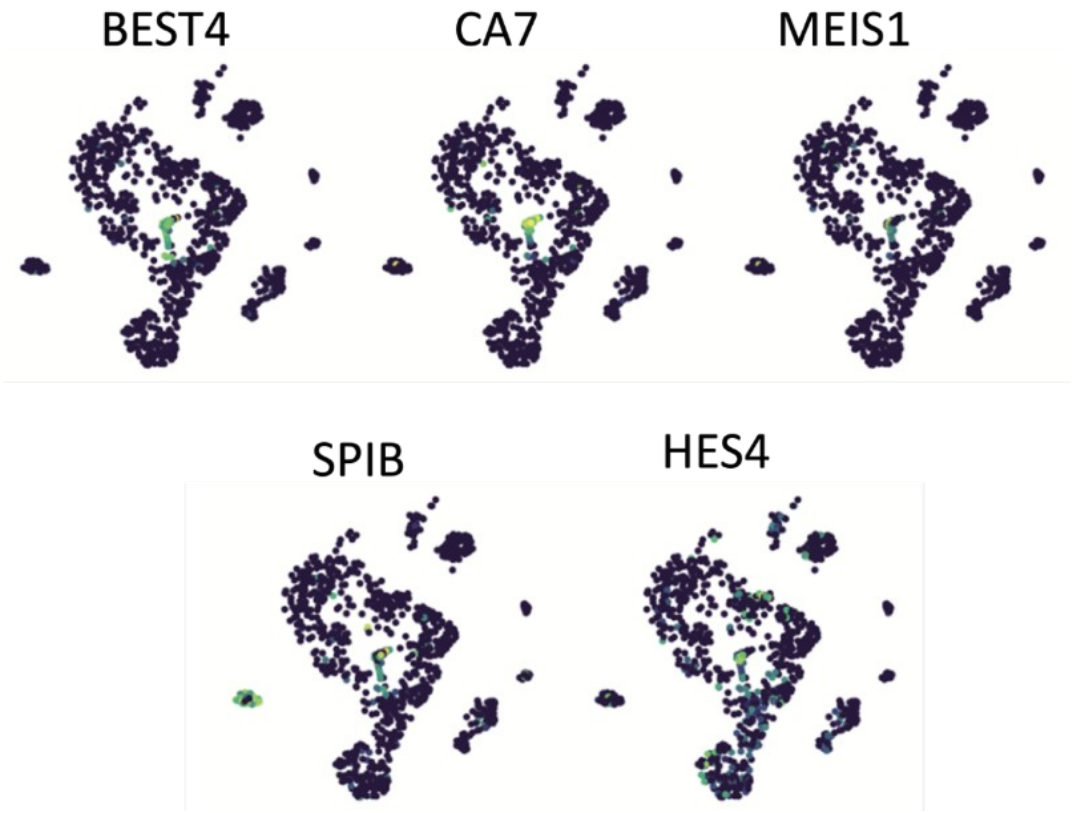
Legend: (A) Integrated UMAP projection of epithelial single-cell transcriptomes derived from both hIEOs and primary tissue, coloured by sample origin. The overlap between hIEO and primary cells suggests broad transcriptomic similarity and conservation of regional identity. (B) UMAP embedding coloured by epithelial cell type annotation. All major intestinal lineages are recovered across conditions, including stem cells, transit-amplifying cells (TAs), early enterocytes, M-cells, tuft cells, goblet cells, enteroendocrine cells (EECs), Paneth cells, and Best4 cells, supporting the fidelity of in vitro differentiation. The presence of rare lineages such as M-cells and Best4 cells confirms that the combined culture and sequencing strategy preserves epithelial diversity and enables lineage-resolved analysis.

One particularly noteworthy outcome of this integration was the identification of a near-perfect match between BEST4^{+} cells in organoids and their counterparts in primary tissue. Of note, these cells formed a distinct cluster with clear separation from all other cell types, and contributions from all four conditions under study (hIEOs and primary tissue derived from both TI and colon). Importantly, this congruence enabled a refined investigation into transcriptional regulation specific to this rare epithelial lineage when considered within the context of benchmarked *in-vivo* datasets, bringing with it a new perspective on the results shown earlier. As shown in Figure 83, Best4-Cell markers including *BEST4*, *CA7*, *MEIS*, *SPIB*, and *HES4* showed clear and high expression in the cell cluster identified above, with the first three demonstrating highly-specific expression mostly restricted to the Best4-Cell cluster, *SPIB* also showing high expression in the M-Cell cluster, and *HES4* demonstrating more sparse and widespread expression – providing further support to the observations undertaken throughout this work.

Together, the results presented throughout this work culminate in a robust demonstration of how intestinal organoid models — when rigorously benchmarked and strategically engineered — can faithfully recapitulate complex epithelial biology, including the emergence of rare and poorly characterised cell types such as Best4^{+} cells. While the incomplete recovery of this lineage reflects the biological challenges inherent to modelling such rare states, the technical advancements, culture refinements, and integrative single-cell strategies developed here lay critical groundwork for future efforts. By anchoring *in vitro* observations to primary human tissue, this study not only validates the translational relevance of hIEOs but also establishes a methodological blueprint for probing human epithelial differentiation at clonal and lineage-resolved resolution — a necessary step towards understanding how epithelial diversity arises, how it is perturbed in disease, and how it might be restored or manipulated for therapeutic benefit.

1

Figure 83 - Expression of Best4-Cell associated markers in integrated hIEO and primary tissue dataset



Legend: Expression of key marker genes associated with the Best4 Cell lineage across an integrated single-cell dataset comprising human intestinal epithelial organoids (hIEOs) and primary intestinal epithelium. All markers exhibit spatially localised expression within a discrete cluster corresponding to the Best4 cell population, supporting the consistent recovery of this rare epithelial lineage across in vitro and in vivo samples.

5.5 Chapter Summary

This chapter systematically evaluated the extent to which optimised hIEO culture models preserved intrinsic transcriptional signatures across biological dimensions — anatomical location (colon vs TI), developmental stage (fetal vs adult), and inflammatory state (healthy vs UC, inflamed vs non-inflamed) — and assessed responses to external stimuli, notably RANKL and IL13/IL4. The carefully structured multiplexed scRNA-seq approach enabled robust analyses, ensuring clear interpretability of how individual conditions influenced epithelial identity.

The findings demonstrated that hIEOs cultured in a standard protocol such as House medium retained substantial regional identity, specifically distinguishing colon from TI. Differential expression analyses recovered numerous validated region-specific markers, such as *SLC10A2*, *LYZ* and *CCL25* (TI-enriched) versus *LEFTY1*, *CFTR*, and *SPINK5* (colon-enriched). These patterns closely aligned with primary tissues, validating model fidelity. GO enrichment further underscored functional specialisation of each region, highlighting nutrient absorption and immune signalling in TI, and epithelial organisation and structural maintenance in the colon.

Likewise, hIEOs from inflamed UC tissues preserved a clear inflammatory profile, characterised by upregulated inflammatory markers *(IL18, IFITM2, OLFM4)* and downregulated differentiation markers *(FABP1, KRT20)*. This state mirrored the pathological epithelial phenotype in UC, marked by progenitor hyperplasia and reduced absorptive differentiation. GO analysis revealed heightened immune pathways and impaired differentiation, confirming that hIEOs intrinsically capture disease-specific signatures.

Developmental origin was also faithfully preserved. Fetal organoids displayed upregulated developmental markers *(OLFM4, NEUROD1, CLDN10)* indicative of progenitor-rich, immature phenotypes, while adult organoids expressed strong differentiation markers *(KRT20, SLC5A1,*

TFF3), confirming stage-specific epithelial maturity. GO enrichment highlighted fetal-specific programmes related to morphogenesis and plasticity, while adult samples showed homeostatic, differentiated functions.

RANKL stimulation potently induced M-cell differentiation in hIEOs, with canonical markers *(SPIB, SOX8, GP2, TNFAIP2)* clearly delineating lineage commitment. Distinct SPIB-negative subsets suggested earlier progenitor states. GO and network analyses reinforced immune and metabolic reprogramming associated with RANKL, supporting its value as a tool for studying rare lineages. By contrast, IL13/IL4 induced broad epithelial reprogramming toward secretory fates, enhancing Goblet and Tuft cell pathways *(MUC2, IL13RA2, ATOH1, POU2F3)*. GO analysis confirmed activation of immune and stress-adaptation pathways, highlighting epithelial plasticity under type 2 inflammatory stimuli.

Comparative analysis of three culture media (House, Fujii, Holloway) revealed distinct lineage biases. House medium provided a balanced baseline with absorptive and proliferative populations, modest responses to inflammation, and preserved homeostasis. Fujii strongly supported proliferation and rare secretory lineages (Goblet, Tuft, EEC) while mounting a pronounced inflammatory response via progenitor expansion. Holloway favoured absorptive differentiation and terminal enterocytes, with strong M-Cell priming but exaggerated inflammatory shifts toward immune-effector lineages. Each medium exhibited clear strengths — Fujii for secretory lineages and plasticity, Holloway for absorptive differentiation, and House as a balanced control — emphasising the need for tailored media selection.

By combining transcriptional and epigenetic profiling, the multiomic framework showed that UC inflammation drives immune activation and disrupted epithelial differentiation. Key mediators *(TNFSF15, TNFAIP2, IL22)* were upregulated, while the chromatin landscape revealed inflammation-specific priming persisting even in non-inflamed tissue, suggesting an epigenetic

"memory" contributing to disease chronicity. Conceptually, epigenetic memory refers to the stable retention of lineage-specific chromatin states, DNA methylation patterns, and histone modifications that arise during tissue development and remain imprinted in somatic cells long after they are removed from their native niche. This molecular memory constrains the transcriptional potential of cells and confers a form of "positional identity", enabling adult epithelia to retain region-specific features even in the absence of *in vivo* environmental cues. *In vitro* culture systems such as intestinal organoids consistently demonstrate this phenomenon: cells derived from different anatomical sites (e.g., ileum versus colon) maintain distinct transcriptional programs, enhancer landscapes and differentiation capacities despite being exposed to identical growth conditions. These persistent signatures arise from pre-established epigenetic architectures that dictate accessible regulatory elements, responsiveness to Wnt, BMP or cytokine signalling, and lineage biases encoded during development. Consequently, much of the regional heterogeneity observed in primary tissues remains detectable in long-term organoid cultures, highlighting the strong influence of epigenetic memory on epithelial identity. Understanding these stable chromatin features is fundamental for interpreting organoid biology, as they can sustain transcriptional divergence between tissues, modulate differentiation outcomes, and impose limits on cellular plasticity even when environmental conditions are experimentally controlled.

Integration of hIEO transcriptomes with benchmarked primary tissue datasets further validated these models. hIEOs preserved key regional and lineage-specific signatures with near-complete convergence onto primary states, enabling refined characterisation of rare lineages such as Best4^{+} Cells. A discrete Best4 cluster emerged consistently across *in vitro* and *in vivo* samples, reinforcing biological fidelity.

In conclusion, this chapter demonstrated that optimised, benchmarked hIEOs provide a robust and physiologically relevant platform for modelling human intestinal biology. Across anatomical, developmental, and pathological axes, hIEOs preserved epithelial signatures, responded predictably to stimuli, and supported recovery of both common and rare lineages. Integration with primary tissue datasets confirmed their translational relevance, underscoring the potential of organoids as tractable surrogates for human gut epithelium in health and disease.

6 Discussion

6.1 Rationale for Research

The intestinal epithelium is a complex and highly regenerative tissue that plays a central role in nutrient absorption, barrier integrity, and immune surveillance. Due to its capacity for continuous renewal, while maintaining a diverse repertoire of specialised cell types, dissecting intestinal epithelial biology in health and disease requires the development of appropriate methods to investigate lineage dynamics, cellular plasticity, and context-specific differentiation. In recent years, advances in organoid culture systems and single-cell genomics have markedly advanced our ability to model intestinal epithelial biology *in vitro*. However, despite these technical improvements, important challenges remain in standardising culture protocols, improving the resolution of rare epithelial cell types, and enabling robust tracking of dynamic cell fate transitions under controlled experimental perturbations.

This thesis was motivated by the need to develop, benchmark and apply reproducible, high-resolution methodologies for interrogating human intestinal epithelial differentiation within organoid-based systems. Particular emphasis was placed on rare epithelial subsets – such as Best4 Cells - which remain poorly characterised despite their emerging physiological relevance. Specifically, this worked aimed to address three major methodological gaps in the field:

(1) the absence of standardised, donor-agnostic organoid culture protocols capable of consistently generating the full spectrum of mature epithelial lineages observed *in vivo* across diverse experimental conditions;

(2) the difficulty in reliably identifying and validating rare or intermediate cell states, particularly under physiologically relevant stimuli; and

(3) the limited ability to reconstruct lineage trajectories in hIEOs over time using simple, rapid, and scalable strategies.

To address these limitations, this research integrated multiplexed single-cell transcriptomic profiling, chromatin accessibility assays, and clonal lineage tracing in human intestinal epithelial organoids. By systematically comparing differentiation protocols, developmental stages, disease states, and immune-modulatory cues, this thesis established a comprehensive framework for dissecting how intrinsic programmes and extrinsic signalling interact to guide epithelial fate. This work also assessed the capacity of *in vitro* systems to faithfully recapitulate *in vivo* epithelial identities, with particular focus on the emergence and maintenance of rare cell populations. Collectively, this thesis makes significant contributions to the development of an experimental toolkit for investigating the complex landscape of the intestinal epithelium. It offers validated methodologies to resolve differentiation trajectories and to characterise rare epithelial subtypes, thereby laying the groundwork for a deeper understanding of their roles in intestinal homeostasis and disease.

6.2 Summary of Findings

6.2.1 Multiplexed Single-cell RNA Sequencing Workflow

In Section 3.2, a high-throughput, multiplexed scRNA-seq platform was established for the characterisation of hIEOs, enabling the simultaneous profiling of 40 conditions across multiple donor lines, media formulations, and anatomical sources within a single experiment — substantially reducing cost, labour, and batch-dependent noise. The workflow integrated cell hashing, droplet-based capture, and a tailored computational pipeline to enable precise demultiplexing, quality control, and transcriptomic analysis. By facilitating parallel evaluation of

complex designs while controlling for technical effects, this approach provided a robust foundation for benchmarking organoid differentiation potential and epithelial fidelity.

The core aim was to assess how effectively standard hIEO culture protocols — particularly the House medium, a slight modification of the canonical organoid protocol developed by Sato *et al.* (2011) — recapitulate the cellular diversity of the native intestinal epithelium, with emphasis on underrepresented or rare lineages. The resulting dataset of over 12,000 high-quality cells revealed that while key undifferentiated and absorptive lineages were consistently recovered (e.g. *LGR5+/ASCL2+* Stem Cells, *MKI67+/TOP2A+* TAs, *LYPD8+/KRT20+* Crypt-Top Enterocytes, *SLC5A1+/CCL25+* Absorptive Enterocytes, *FABP1+/SLC5A1+* Early Enterocytes), secretory and rare cell types — including Best4 Cells, Tuft Cells, EECs, M-Cells, and Paneth Cells — were absent or poorly resolved. Canonical markers for these lineages (e.g. *BEST4, SPIB, CA7, MEIS1* for Best4; *SPIB, SOX8, CCL20, TNFAIP2* for M-Cells; *SPIB, POU2F3, ALDH1A3, TRPM5* for Tuft; *CHGA, PYY, NEUROD1, TPH1* for EECs; *DEFA5, DEFA6, PRSS2, LYZ* for Paneth) were sparsely expressed or detected only in transitional states, indicating incomplete differentiation.

These findings underscore the limitations of standard culture protocols in supporting the full epithelial spectrum. While organoids offer a powerful platform to study intestinal biology, rare lineages were markedly restricted under baseline House conditions. This is consistent with studies showing that rare subsets — particularly Best4 — are acutely sensitive to microenvironmental context, including precise Wnt, R-spondin, and BMP modulation[38,41,60,67,72]. The lack of robust specification suggests that default media formulations lack the necessary instructive cues or fail to sustain lineage commitment.

Importantly, this section also established scRNA-seq as a sensitive and scalable tool for evaluating epithelial fidelity *in vitro*. Compared to bulk transcriptomics or qPCR, single-cell resolution enabled direct assessment of representation, identification of transitional states, and

quantification of lineage biases. These features proved particularly valuable in exposing the underrepresentation of mature subsets, an insight not readily obtainable via conventional methods. Notably, sufficient resolution was achieved using an intermediate sequencing platform (NextSeq), obviating the need for more advanced and costly technologies. While higher-resolution data would be attainable with NovaSeq, the approach adopted here provided a rapid and cost-effective framework for robust evaluation of the epithelial landscape.

Collectively, this section provided the experimental and analytical foundation for the remainder of the thesis. It generated a comprehensive map of epithelial differentiation under baseline conditions, exposed critical limitations in rare lineage specification, and demonstrated the value of scRNA-seq as both a diagnostic and discovery tool. The platform's scalability and resolution establish it as a cornerstone for optimising hIEO culture strategies, validating differentiation protocols, and supporting translational research in disease modelling and regenerative medicine.

6.2.2 Comparative Meta-Analysis Of Organoid Single-Cell RNA Sequencing Datasets

The differentiation of rare intestinal epithelial lineages often requires the integration of highly specialised biological and mechanical cues that are absent from standard *in vitro* culture conditions. While conventional organoid media support robust expansion of stem and transit-amplifying populations, they insufficiently recapitulate the complex signalling gradients, physical constraints, and cell–cell interactions that operate *in vivo* to specify low-abundance cell types. Rare lineages such as M-cells, BEST4^{+} enterocytes, and certain enteroendocrine subtypes depend on temporally coordinated cytokine exposure, ligand-dependent activation of immune–epithelial signalling pathways, and defined mechanical environments that regulate cytoskeletal tension and apical–basal polarity. This thesis provides several examples of this requirement.

First, differentiation into M-cells was strongly enhanced only upon exposure to RANKL, demonstrating that exogenous activation of the RANK–RANKL axis is necessary to unlock this otherwise inaccessible fate *in vitro*. Second, the emergence of BEST4^{+} epithelial cells remained highly restricted under basal conditions but increased under selected stimulation regimes, underscoring their dependence on specific niche-derived signals. Third, mechanical context was shown to influence lineage composition: 3D cultures favoured certain progenitor outputs and goblet cell differentiation, whereas 2D monolayers—by altering substrate stiffness, planar spreading, and epithelial tension—modified transcriptional states and the abundance of rare lineages. Collectively, these findings demonstrate that rare epithelial fates cannot be generated reliably through media composition alone; instead, their specification requires deliberate reintroduction of key developmental and mechanical cues that are otherwise lost when tissues are removed from their native microenvironment.

Section 3.3 established a comparative meta-analysis framework for evaluating epithelial fidelity across published organoid scRNA-seq datasets, focusing on the emergence of Best4 Cells and other rare epithelial lineages. This approach enabled cross-study benchmarking of differentiation protocols and facilitated the identification of shared or divergent patterns in rare cell specification, complementing in-house experimental findings with broader external validation.

The rationale for selecting the Fujii *et al.*, Holloway *et al.*, and Elmentaite *et al.* datasets was their reported ability to recapitulate the full epithelial landscape, along with preliminary analysis indicating enhanced expression of *BEST4* transcripts. Re-analysis of these datasets demonstrated striking differences in both the abundance and resolution of rare populations. Unsupervised clustering revealed that the Fujii protocol most effectively supported a broader epithelial repertoire, with distinct populations of rarer secretory types (such as EECs and Tuft

Cells). By contrast, the Holloway protocol exhibited strong absorptive lineage priming, reflected in high representation of enterocyte stages but comparatively limited secretory differentiation. The Elmentaite dataset presented the most robust population of EECs, but showed reduced capacity to capture other differentiated or rare subtypes. Most notably, a discrete Best4 cluster — characterised by co-expression of *BEST4*, *SPIB*, *CA7* and *MEIS1* — was identifiable in the Fujii and Holloway datasets, but not in Elmentaite. In the latter — as in the House dataset — numerous *BEST4*-expressing cells were present, but dispersed across clusters. These cells likely failed to cluster due to insufficient transcriptional maturity or divergence in broader transcriptomic profiles. This discrepancy highlights the critical role of media composition, growth factor signalling, and transcriptional resolution in enabling rare epithelial lineages. For example, the inclusion of NRG1 (Holloway) and IGF-1/FGF2 (Fujii) — absent from both House and Elmentaite — appears to enhance lineage diversity, suggesting these niche signals may act synergistically with BMP and Wnt gradients to support rare fate specification. NRG1, a mesenchyme-derived member of the EGF family, plays a pivotal role in intestinal stem and progenitor regulation. In organoid cultures, NRG1 supplementation promotes robust proliferation and budding, significantly increasing colony formation and regenerative potential[309]. It induces a fetal-like transcriptional programme and supports expansion and maintenance of progenitor identity. Similarly, a high-throughput screen identified IGF-1 and FGF-2 as potent enhancers of stem cell activity and epithelial diversity[88]. In organoids, these factors markedly improve clonogenicity and enable long-term maintenance of diverse subtypes. Compared to standard media, IGF-1 and FGF-2 support differentiation of a broad array of cell types, highlighting their potential in promoting comprehensive specification.

Integration of the dataset generated in EXP01 (Section 3.2) with these three selected datasets further validated earlier findings. With enhanced statistical power from the increased dataset

size, some *BEST4*$^+$ cells from the House and Elmentaite datasets — previously dispersed across clusters — were now integrated within the discrete Best4 cluster. The consistent co-expression of *SPIB*, *BEST4*, *CA7*, and *MEIS1* confirmed the robustness of these markers for identifying Best4 Cells. In parallel, the broader, low-level expression of *BEST4* within undifferentiated and early enterocyte intermediates — often accompanied by *SPIB* upregulation — supported a model of progressive differentiation, where priming precedes substantial expression of effectors such as *CA7* and *BEST4*. However, while integration with previous datasets enhanced clustering resolution and enabled more confident identification of transitional Best4 states, it also diluted extremely rare lineages (e.g. Tuft Cells, only present in the Fujii dataset), cautioning against indiscriminate merging in rare cell studies.

6.2.3 Benchmarking Of Candidate Differentiation Protocols For Rare Cell Types

To interrogate whether Best4 Cells in organoid datasets were transcriptionally underrepresented due to technical limitations rather than biological absence, the scRNA-seq methodologies underlying each dataset analysed in the meta-analysis were considered. Notably, the Fujii dataset employed full-length Smart-seq2, a platform with superior sensitivity for low-abundance transcripts, potentially allowing more accurate detection of rare epithelial lineages. By contrast, both the Holloway and Elmentaite datasets — like the House dataset — used 10x Genomics, albeit at significantly higher sequencing depth than the House data. This raised the hypothesis that Best4 Cells may be present but too transcriptionally diluted or immature to form a distinct cluster in lower-sensitivity platforms.

To test this, Section 3.4 began with the use of Smart-seq2 in FACS-sorted EPCAM$^+$/BEST4$^+$ cells derived from both primary intestinal epithelium and hIEO cultures. While sorting successfully enriched for BEST4$^+$ cells in both contexts, transcriptomic profiling revealed a striking

divergence. BEST4^{+} cells from primary tissue robustly co-expressed canonical markers including *SPIB*, *CA7*, *MEIS1*, *OTOP2*, and *CA4*, confirming their identity. By contrast, BEST4^{+} organoid-derived cells lacked this signature entirely. This result underscored the risk of over-interpreting surface marker enrichment without transcriptomic confirmation, and highlighted the value of high-sensitivity platforms for rare lineage validation.

Moreover, these findings validated the *SPIB–BEST4–CA7–MEIS1* axis as a robust marker set for Best4 lineage identity. The complete absence of this signature in organoid-derived BEST4^{+} cells suggested that the lineage was truly absent in House datasets, probably due to insufficient niche cues or differentiation time. Of note, attempts to prolong differentiation to encourage Best4 Cell maturation led to increased cell death and downregulation of key markers, suggesting that the House protocol may impair epithelial maturation over extended culture periods.

The following experiment aimed to compare the two top-performing culture conditions identified in the meta-analysis — Fujii and Holloway — and directly benchmarked them against the standard House protocol in a controlled experimental setting using multiplexed scRNA-seq. Organoid lines subjected to each culture condition were derived from equally split, patient-matched parent cultures, thereby ensuring a consistent genetic background and eliminating inter-donor variability. This design provided a robust framework for assessing the intrinsic effects of each medium on epithelial differentiation. Multiple experimental conditions were evaluated, including anatomical origin, disease status, and cytokine stimulation, enabling a systematic comparison of the House (serving as control), Fujii, and Holloway protocols. By multiplexing these conditions into a single 10x Genomics experiment, batch effects were mitigated, and subtle differences in lineage output became readily interpretable. This design offered an affordable, scalable and biologically relevant model for evaluating differentiation protocols side-by-side.

Results confirmed that the Fujii protocol generated the highest cell recovery and supported the broadest epithelial repertoire, particularly secretory and/or rare lineages such as Tuft Cells, EECs, and early M-Cells (expressing *SPIB* and low *SOX8*, *CCL25* and *CLU)*. Holloway media, by contrast, consistently skewed toward absorptive differentiation, while the House protocol yielded the lowest epithelial diversity. However, none of these protocols succeeded in producing transcriptionally mature and clustered Best4 Cells, with few and diffusely distributed *BEST4+* cells detected across all datasets. Progressive filtering based on co-expression of *SPIB*, *CA7*, and *MEIS1* rapidly reduced candidate cell counts, with no condition yielding more than three cells fulfilling all marker criteria. These findings affirmed that while niche-enriched media enhanced overall complexity, they failed to fully support Best4 lineage commitment *in vitro*.

Taken together, these findings underscore critical challenges in the reproducibility of organoid differentiation protocols. Despite attempts to standardise culture conditions, considerable variability was observed across datasets, particularly in the emergence and transcriptional definition of rare epithelial populations. While differences in donor origin, sequencing depth, and analytical resolution may partially account for these discrepancies, the results emphasise the inherent limitations of relying solely on scRNA-seq for lineage assessment. Most importantly, they highlight the need for rigorous, multi-modal validation frameworks — integrating transcriptomic, epigenetic, and phenotypic readouts — when interrogating rare or transcriptionally ambiguous populations.

6.2.4 Integrating Transcriptomic And Epigenetic Modalities To Resolve Rare And Transitional Epithelial Cell States

While single-cell RNA-seq has proven indispensable for characterising epithelial heterogeneity, it is constrained by its dependence on mRNA abundance — often failing to capture early lineage

priming, particularly in low-abundance or transcriptionally immature populations such as Best4 Cells. To address this, Section 4.2 introduced a multimodal strategy combining transcriptome and chromatin accessibility profiling from the same cells (scMultiome-seq). This dual-modality approach was hypothesised to improve state resolution by revealing regulatory activity not yet evident at the transcriptional level.

Standard nuclear-isolation-based scMultiome protocols suffer from RNA loss and limited multiplexing capacity [257,259]. To circumvent this, a modified workflow enabling whole-cell profiling with surface-compatible hashing was optimised. Two permeabilisation strategies — PFA and digitonin — were benchmarked, with digitonin yielding superior RNA and ATAC quality, including greater transcriptomic complexity, stronger nucleosome positioning, and more uniform antibody incorporation. This protocol was employed in a multiplexed multiome screen of eight hIEO lines across diverse donor and disease backgrounds.

Despite technical success — confirmed by high-quality RNA capture, preserved chromatin structure, and effective sample deconvolution — the ATAC modality alone showed limited power to resolve fine-grained epithelial lineages. Chromatin-based clustering poorly separated secretory, progenitor, and absorptive cell types, reflecting both low sequencing saturation and the sparse nature of scATAC-seq. By contrast, RNA-based clustering and integrated RNA+ATAC embeddings offered clearer resolution, recapitulating hierarchies from Stem Cells and TAs through to absorptive and secretory fates. Integrated clustering also refined rare populations such as Tuft Cells and EECs, highlighting the added value of combining modalities.

Nevertheless, Best4 Cells remained absent. Transcriptomic analysis identified only eight *BEST4*-expressing cells — one co-expressing *SPIB*, none meeting the stringent *BEST4*/*SPIB*/*CA7* criteria — with no discrete cluster resembling a mature Best4 identity. Chromatin accessibility, however, revealed open regions at Best4-relevant loci — including *SPIB*, *MEIS1*, and the *BEST4* promoter

— particularly in progenitor populations. This suggests that epigenetic groundwork for Best4 specification may be partly established, but transcriptional activation is not achieved, likely due to insufficient niche cues. Chromatin at these loci remained largely inaccessible in secretory populations, supporting absorptive-lineage specificity in Best4 identity.

Beyond Best4, chromatin profiling offered broader insights. ATAC-based GO enrichment analyses captured lineage-consistent themes — TNF responses in Enterocytes, CoA metabolism in progenitors, and adrenergic signalling in Tuft Cells. Several clusters also showed heterogeneous nucleosome and TSS enrichment, highlighting chromatin diversity even within transcriptionally coherent populations. Multiomic pseudotime trajectories further validated cluster annotations by reconstructing continuous absorptive and secretory pathways.

Despite its promise, scMultiome-seq was technically demanding and limited by sequencing depth. ATAC-seq requires far deeper coverage than scRNA-seq, particularly for rare or transitional states. Here, shallow chromatin depth — exacerbated by NextSeq constraints — precluded robust peak detection and reduced sensitivity to subtle lineage differences. The persistent failure to recover Best4 clusters, even in samples with epigenetic priming, underscores the need for refined differentiation protocols incorporating specific niche signals (e.g. BMP modulation, RANKL, microbiota-derived metabolites). These may be essential to drive transcriptional activation where regulatory elements are primed but inactive.

In sum, this section showed that scMultiome-seq is a valuable, feasible platform for refining epithelial classification. It revealed gaps between regulatory potential and realised transcription, underscoring the role of extrinsic cues in unlocking rare lineage programmes. While current protocols can induce partial epigenetic readiness for Best4 differentiation, full commitment remains rare — reflecting a biological limitation rather than a detection artefact.

Nonetheless, these findings provide a foundation for future efforts to engineer multi-layered frameworks for lineage tracing and epithelial modelling in organoids.

6.2.5 Development Of A Lineage Tracing Platform In Human Epithelial Organoids

Recognising the limitations of conventional scRNA-seq in resolving rare or transient epithelial lineages - and which could not be appropriately tackled by the scMultiome approach developed - a temporally structured CellTagging strategy was devised in Section 4.3 to infer clonal relationships and reconstruct lineage trajectories at single-cell resolution within hIEOs, leveraging the transcriptome-compatible CellTag lentiviral barcoding system to capture cell fate dynamics over time.

A key innovation was the transition from standard 3D cultures to 2D monolayer configurations, thereby enhancing lentiviral transduction efficiency by exposing the apical membrane of epithelial cells. This adaptation was crucial for delivering sequentially timed CellTag libraries (V1–V3), enabling temporally distinct barcoding of progenitor and differentiated cell populations. Preliminary transcriptomic comparisons between 2D and 3D cultures demonstrated that 2D monolayers preserved major epithelial lineage identities, including absorptive and secretory subtypes, albeit with minor shifts in global gene expression and reduced epithelial diversity at higher passage numbers. Pilot experiments established that the tagging process did not induce major transcriptional artefacts, and that CellTags remained detectable across timepoints. However, storage conditions (−80 °C vs 4 °C) adversely affected RNA integrity, limiting the feasibility of staggered sample processing. Optimised viral concentration via ultracentrifugation and refined sample handling improved transduction rates but imposed cellular stress, reducing RNA quality and hashing signal recovery.

The final proof-of-concept experiment integrated sequential CellTag barcoding with scRNA-seq across three timepoints, simultaneous with changes in experimental conditions (i.e. transitioning to CM, DM, and RANKL). Despite reduced cell recovery and moderate tag dropout, clonal structures were successfully reconstructed. Network-based visualisations revealed dynamic lineage bifurcations, including the emergence of M-Cells and Goblet Cells from shared absorptive-like progenitors, providing empirical evidence of temporal lineage branching in hIEOs. The majority of clones arose from early-tagged progenitors, highlighting the temporal resolution of the method. However, refinement of this method would still require further optimisation cycles, additional follow-up experiments, and a tailored computational pipeline. Similarly, while an exploratory Multiome experiment demonstrated the feasibility of integrating epigenetic profiles onto this novel approach, insufficient ATAC sequencing depth precluded meaningful chromatin-based integration. Overall, despite the technical limitations identified, this work represents the first demonstration of a multiplexed, high-resolution lineage tracing system in hIEOs, establishing the methodological foundation for future studies probing epithelial plasticity, differentiation, and rare lineage emergence using transcriptome-resolved clonal barcoding.

6.2.6 Biological Validation Of Human Epithelial Organoids As Models Of Intestinal Biology

Although the primary objective of this thesis was methodological - aiming to benchmark and optimise organoid-based workflows for rare epithelial lineage recovery - several biologically informative findings emerged as result of the multiplexed single-cell experiments conducted. These findings provide critical validation of the biological fidelity of hIEOs and offer valuable use cases illustrating the interpretive power of the methods developed.

Chapter 5 focused on the analytical strategies that can be leveraged in high-throughput multiplexed experiments, and used specific biological case studies to demonstrate how hIEOs respond to anatomical, developmental, disease-specific, and cytokine-driven cues. Notably, hIEOs displayed a high degree of transcriptomic fidelity in preserving core features of their tissue of origin. Organoids derived from the colon and TI reliably recapitulated region-specific transcriptional programs, including differences in secretory versus absorptive lineage balance and crypt-top gradients. Likewise, fetal and adult organoids maintained developmental hallmarks, supporting the validity of hIEOs for modelling age-specific epithelial biology. Disease-specific signatures were also preserved - for example, organoids derived from inflamed UC tissue maintained inflammatory transcriptional signatures, further underscoring their relevance for translational applications.

The cytokine stimulation experiments — originally designed to prime *SPIB+* progenitors with the aim of inducing Best4 Cell differentiation — provided additional biological insight. RANKL stimulation led to a robust and selective enrichment of M-Cell populations, consistent with known *in vivo* roles for RANKL in SPIB-mediated M-cell specification[54,349]. Similarly, IL13/IL4 treatment induced a marked expansion of Tuft Cell populations, reproducing cytokine-induced secretory skewing previously described in both murine and human models[72]. Although these stimulations did not achieve the intended differentiation of Best4 Cells, they validated the ability of hIEOs to respond predictably and reproducibly to external stimuli, making them ideal systems for probing epithelial plasticity and cytokine-mediated lineage remodelling.

Transcriptomic comparisons between stimulated organoids and primary tissues further reinforced these conclusions. In nearly all contexts, organoid-derived epithelial populations closely resembled their *in vivo* counterparts, with high correlation of gene expression profiles and conserved lineage-specific markers. This was particularly evident in the stimulation

conditions, where organoid-derived M-cells and Tuft Cells showed strong transcriptomic overlap with corresponding populations in primary intestinal tissue.

As a final case study, and to assess the overall biological fidelity of the organoid datasets generated in this study, a balanced subset of unstimulated, healthy organoids derived from both colon and TI —spanning all three tested media (House, Fujii, and Holloway) — was integrated with two publicly available, benchmark, primary epithelial single-cell datasets. This integrative analysis revealed a striking concordance between organoid-derived and primary epithelial cells across key lineages, with cells from both sources mapping closely within the same transcriptional clusters. This result affirmed the capacity of hIEOs to preserve general and region-specific epithelial identities, faithfully recapitulating features of both colonic and ileal differentiation programs. Furthermore, lineage-specific similarities extended to both abundant populations (e.g., Enterocytes and Goblet Cells) and rarer subsets, reinforcing the robustness of the proposed organoid models as platforms for studying native epithelial complexity. This high-resolution integration provided a rigorous external validation of organoid differentiation fidelity and established a foundation for more nuanced comparisons, such as the evaluation of rare cell types like Best4 populations in subsequent analyses.

Remarkably, the integrated atlas revealed that Best4 Cells derived from organoids co-clustered seamlessly with their primary counterparts. The consistent co-expression of markers such as *SPIB, BEST4, CA7,* and *MEIS1* across both datasets confirmed the transcriptional coherence of the Best4 program. Moreover, this integration enabled differential gene expression analysis between primary and organoid-derived Best4 cells. In sum, these results affirmed not only the biological relevance of the hIEO-derived Best4 cells, but also the analytical power of integrative single-cell approaches in dissecting rare epithelial states.

6.3 Relevance And Contributions Of The Current Research

This thesis makes a significant contribution to the field of intestinal epithelial biology, not by uncovering new mechanistic pathways or resolving longstanding biological debates, but by establishing and validating a suite of advanced methodological frameworks that expand the experimental and analytical toolkit available to researchers. Its central value lies in the rigorous implementation and benchmarking of scalable, multiplexed single-cell approaches — including droplet-based scRNA-seq, integrative multiomic pipelines, and high-resolution lineage tracing — tailored specifically to human intestinal epithelial organoids (hIEOs). These platforms were systematically applied across donor lines, anatomical locations (colon vs terminal ileum), developmental stages (fetal vs adult), and inflammatory states (healthy vs ulcerative colitis), yielding a robust and generalisable foundation for future organoid-based research and providing several key contributions to the field.

First, an optimised droplet-based single-cell RNA-seq workflow was established and benchmarked for use in hIEOs. This multiplexed approach enabled high-throughput, cost-effective, and high-resolution profiling across a wide range of conditions, offering a scalable method for interrogating epithelial differentiation at single-cell resolution.

Second, full-length SmartSeq2 sequencing was integrated alongside droplet-based scRNA-seq to validate the presence or absence of rare epithelial cell types. This revealed that recovery of such lineages is highly dependent on transcript coverage and sensitivity, underscoring the limitations of short-read technologies for detecting low-abundance populations and establishing the value of orthogonal validation platforms.

Third, the thesis identified critical limitations in standard hIEO differentiation protocols. Rare epithelial lineages such as Best4^{+} cells, M-cells, Tuft Cells, and Paneth Cells were inconsistently

recovered, even when culture conditions were carefully replicated. This finding emphasises the need for improved and better standardised protocols and highlights the inherent biological variability in epithelial lineage specification.

Fourth, a multiomic single-cell analysis framework was developed by integrating transcriptomic (RNA) and chromatin accessibility (ATAC) data within the same cells — to my knowledge, the first time this was achieved in human intestinal organoids. While technical limitations in sequencing depth restricted interpretability, the results demonstrated that with sufficient coverage, multiomic profiling can enhance cluster resolution and support the identification of rare or transitional epithelial states.

Fifth, this work adapted the CellTagging methodology for use in hIEO systems, including the development of a triple-tag temporal barcoding strategy (V1–V3). To my knowledge, this is also the first application of a lineage tracing platform in hIEOs. This enabled clonal reconstruction over time and revealed transcriptional divergence among clonally related cells, illustrating the potential of this approach for lineage-resolved fate mapping in epithelial tissues.

Sixth, a systematic perturbation framework was implemented using small molecules and cytokines — including NaB, RANKL, and IL13/IL4 — to modulate epithelial identity in hIEOs. This allowed controlled enrichment of specific lineages and the exploration of condition-specific transcriptional plasticity, offering valuable tools for studying epithelial responses to defined environmental cues.

Seventh, a comprehensive single-cell atlas was generated, encompassing perturbed and unperturbed hIEOs derived from multiple patients, anatomical sites, and culture conditions. This dataset constitutes a valuable reference for benchmarking studies and for evaluating differentiation fidelity across biological and technical variables.

Finally, this thesis highlighted important limitations to the current definition of BEST4$^+$ epithelial cells, which remains heavily anchored in a small set of marker genes — most prominently *BEST4* and *SPIB* — that capture only part of the transcriptional and functional identity of this lineage. Relying on a handful of markers risks oversimplifying a population whose emergence, maturation, and functional states may be far more heterogeneous than these canonical features suggest. Recent single-cell studies have shown that BEST4$^+$ cells occupy a distinct regulatory niche enriched for ion-transport programs, purinergic signalling, and pH-sensing pathways, and that their identity is supported by coordinated activity of transcriptional regulators rather than by expression of *BEST4* alone. In this thesis, multiple observations reinforce this view: for example, BEST4$^+$ cells displayed divergent transcriptional relationships depending on culture format, media composition, and differentiation stage, despite maintaining stable *BEST4* expression. Moreover, the relative preservation of *BEST4* expression across 2D and 3D cultures contrasted with shifts in broader regulatory networks revealed by differential expression and GO analyses, suggesting that the defining features of this lineage may lie in its regulatory architecture rather than solely in marker levels. Collectively, these findings support a broader definition of BEST4$^+$ cells that integrates transcription factor activity, chromatin accessibility, and pathway-level signatures. Such a regulatory perspective not only aligns more closely with functional identity but also provides a more robust framework for recognising precursor states, intermediate phenotypes, and condition-dependent variants of the BEST4$^+$ lineage.

By directly comparing media compositions, stimulation regimes, and sequencing platforms, this thesis provides a practical framework for evaluating epithelial differentiation fidelity and resolving context-dependent transcriptional variation. Importantly, it delineates current technical boundaries in recovering rare or transitional cell types and offers realistic expectations for future applications of these models. While the focus was not on delivering biological

discoveries per se, the work establishes validated protocols, scalable analytical pipelines, and curated datasets that enable more reproducible, high-resolution investigations of intestinal epithelial biology. Collectively, this thesis lays critical groundwork for future studies aiming to resolve lineage trajectories, disease mechanisms, and regenerative processes, positioning itself as a foundational resource to support mechanistic and translational advances in the field.

Despite robust experimental designs and advanced analytical strategies, this research is not without limitations. First, while hIEOs provide a powerful platform for modelling epithelial biology, they inherently lack the full complexity of the native intestinal niche. The absence of mesenchymal, immune, vascular, and neural components constrains physiological relevance. Processes such as epithelial–immune crosstalk, microbiome sensing, stromal feedback, and mechanical stimuli cannot be fully captured in standard monocultures. Although this thesis focused on epithelial-intrinsic programmes, future studies incorporating co-cultures, assembloids, or organ-on-chip platforms will be essential to extend these findings into more physiologically complete systems.

Second, this work relied heavily on single-cell transcriptomics as the principal readout for cell identity and state. While scRNA-seq offers unprecedented resolution, it remains an indirect proxy for function. Transcription does not always translate into protein activity, and post-transcriptional regulation, protein turnover, and subcellular localisation are not captured. Moreover, cell dissociation protocols may induce stress responses that confound interpretation, particularly when comparing monolayer and 3D cultures or inflamed samples. Complementary approaches such as proteomics, spatial transcriptomics, or live-cell imaging would provide orthogonal validation and contextualise transcriptomic findings within functional outputs.

Third, despite careful optimisation, recovery of rare epithelial lineages remained sparse and inconsistent. This may reflect genuine scarcity but also reveals limitations in sensitivity,

sampling depth, and library preparation biases. Clonal dropout and technical noise inherent to single-cell barcoding can obscure subtle transitions, compounded by variability in organoid output even under identical conditions. More robust protocols and batch correction strategies will be needed.

Fourth, the adaptation of CellTagging, though innovative and functional, faced limitations in efficiency and resolution. High viral titres were required, inducing toxicity and reducing recovery. Clonal coverage was sufficient for proof-of-concept but not dense enough to reconstruct lineage hierarchies across all conditions. The temporal spacing between tags, combined with asynchronous differentiation, limited the resolution of inferred trajectories. Improved viral delivery, selective tag enrichment, and higher-throughput sampling will be necessary to scale this approach.

Fifth, although a multiomic (scRNA + scATAC) approach was piloted, chromatin data did not reach the depth required for optimal resolution. Integration of transcriptomic and epigenetic data was only partially successful and not leveraged to its full potential. Future experiments using higher-yield protocols, targeted perturbations, or refined computational tools could better realise the promise of multiomic strategies.

Finally, the emphasis on transcriptomic characterisation precluded functional assessment of behaviours such as migration, polarity, barrier integrity, or secretory activity — all critical for understanding epithelial physiology. The exclusive use of *in vitro* models also limits generalisability to *in vivo* contexts, where spatial architecture, microbiota exposure, and systemic signalling shape epithelial behaviour.

In summary, while this thesis establishes a rigorous and versatile experimental framework for interrogating epithelial differentiation and plasticity, it also underscores the importance of integrating multi-modal data, physiologically complex systems, and functional assays to fully

capture the dynamics of intestinal epithelial biology. Recognising and addressing these limitations will be essential for advancing the translational potential of organoid-based research.

6.4 xxLate-Breaking Insights on BEST4/CA7$^+$ Cells

A central objective of this thesis was to evaluate the capacity of organoid-based systems to support the emergence and maintenance of rare epithelial lineages, with a specific focus on Best4 Cells. Despite comprehensive benchmarking of multiple differentiation protocols, integration of multiomic approaches, lineage tracing, and stimulation strategies, the generation of transcriptionally coherent Best4 clusters remained elusive in most experimental conditions tested. *BEST4*-expressing cells were detected in several datasets, including those produced under House, Fujii, and Holloway media, but their distribution was diffuse, lacking co-expression of canonical markers such as *SPIB*, *CA7*, and *MEIS1*. This raised questions as to whether these cells were genuinely absent, transcriptionally immature, or simply below the detection threshold of standard droplet-based scRNA-seq.

In this context, it is particularly informative to relate these findings to recent work shown in Wang *et al.* (2025)[350], published during the final stages of writing this thesis. Wang and colleagues have provided the most definitive characterisation to date of human intestinal BEST4$^+$/CA7$^+$ cells. In their study, the authors employed high-resolution single-cell techniques, including Smart-seq2, to identify BEST4$^+$ cells as a distinct absorptive lineage with functional relevance in ion transport and immune modulation. These cells were shown to co-express a consistent gene signature — including *BEST4*, *CA7*, *SPIB*, and *OTOP2* — and expanded in response to IFN-γ, indicating a cytokine-responsive behaviour with roles in mucosal homeostasis and defence. This discovery validates the biological relevance of the previously understudied BEST4$^+$ epithelial

lineage explored throughout this thesis. Moreover, the demonstration that these cells are primary epithelial targets for bacterial diarrheal toxins such as cholera toxin emphasizes their importance in host-pathogen interactions and epithelial barrier regulation.

Tables 21-22 below provide a direct comparison between the compositions of Fujii-based organoid media and the media used by Wang et al. This analysis underscores why Fujii-based organoids might yield different lineage outcomes (e.g., low abundance of $BEST4^+$ cells) compared to the Wang *et al*. 2025 small-intestine-derived BEST4 protocol: the absence of directed lineage-inducing cues (Notch activation + IFN-γ) and the continued presence of self-renewal/niche signals (Wnt, EGF, Noggin, A83-01, etc.) likely maintains a more stem- or progenitor-biased state and suppressed differentiation into rare, functionally specialised lineages. Overall, the contrast between the results of this thesis and those from Wang *et al*. underscores both the biological specificity and technical sensitivity required to robustly detect such rare cell types. Thus, the inability to detect coherent Best4 clusters in the datasets produced in this work is likely solely due to constraints posed by limited lineage-inducing cues. Importantly, this thesis demonstrated that *BEST4+* cells could be detected at low levels in single-cell datasets, especially after integration with primary data from Smillie *et al*. and Burclaff *et al*., where transcriptomic alignment improved. However, the absence of a well-defined cluster of BEST4+ Cells in the organoid datasets on their own highlighted their transcriptional immaturity and reinforced the importance of niche-specific cues — such as IFN-γ or other mesenchymal factors — in fully recapitulating this lineage (as alluded throughout this work). Ultimately, these parallel insights lay the foundation for iterative improvements in hIEO differentiation and underscore the importance of aligning biological ambition with technical capability. Interestingly, while the Wang *et al.* study demonstrated that IFN-γ stimulation is a potent inducer of $BEST4^+/CA7^+$ lineage specification in organoids, the Fujii and Holloway

datasets re-analysed in this thesis also yielded distinct BEST4$^+$ clusters despite the apparent absence of direct IFN-γ treatment – raising interesting questions as to what drove improved representation of the BEST4$^+$ lineage compared to the House or Elmentaite protocols, when none incorporated IFN stimulation.

Table 21 - List of reagents for the preparation of "Fujii" vs "Wang" conditioning media, as per Fujii et. al (2018) and Wang et. al (2025)

Reagent	**Final Concentration (Fujii)**	**Final Concentration (Wang)**
Recombinant mouse EGF	50 ng/mL	50 ng/mL
A83-01	500 nM	500 nM
Recombinant mouse Noggin	100 ng/mL	100 ng/mL
Hepes	10 mM	10 mM
GlutaMAX	2 mM	2 mM
B27 Supplement	1x	1x
Penicillin-Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin
Recombinant human R-spondin1*	1 µg/mL	200 µg/mL
N-Acetyl-L-cysteine*	1 mM	1.25 mM
[Leu^{15}]-Gastrin I human*	10 nM	-
WREN medium*	50%	-
Recombinant human IGF-1*	100 ng/mL	-
Recombinant human FGF-2*	50 ng/mL	-
Wnt Surrogate*	-	0.5 nM
SB202190*	-	1 µM
PGE2*	-	1 µM
Nicotinamide*	-	10 mM
In Advanced Dulbecco's Eagle's Medium/F12	-	-

**: Reagents that differ between Fujii and Wang protocols (in usage or concentration)*

Table 22 - List of reagents for the preparation of "Fujii" vs "Wang" differentiation media, as per Fujii et. al (2018) and Wang et. al (2025)

Reagent	**Final Concentration (Fujii)**	**Final Concentration (Wang)**
Hepes	10 mM	10 mM
GlutaMAX	2 mM	2 mM
B27 Supplement	1x	1x
Penicillin-Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin	100 U/mL Penicillin 0.1 mg/mL Streptomycin
Recombinant human R-spondin1*	1 µg/mL	200 µg/mL
N-Acetyl-L-cysteine*	1 mM	1.25 mM
[Leu^{15}]-Gastrin I human*	10 nM	-
WREN medium*	50%	-
Recombinant human IGF-1*	100 ng/mL	-
Recombinant human FGF-2*	50 ng/mL	-
A83-01*	500 nM	-
Recombinant mouse Noggin*	100 ng/mL	-
In Advanced Dulbecco's Eagle's Medium/F12	-	-
IFN-γ	-	10 ng/ml

**: Reagents that differ between Fujii and Wang protocols (in usage or concentration)*

Several non-mutually exclusive hypotheses can explain this observation. First, both the Fujii and Holloway media formulations include niche-enriching growth factors that may contribute to a more permissive chromatin and transcriptional landscape. Fujii media supplements IGF-1 and FGF2, which have been shown to support broader epithelial diversity and stem cell clonogenicity[88]. In turn, Holloway includes NRG1, a known enhancer of progenitor identity and epithelial expansion[309]. These factors could indirectly prime the epithelium toward rare fate specification by stabilizing intermediate cell states or by modulating signalling axes (e.g., BMP/Wnt) that intersect with pathways involved in BEST4$^+$ cell commitment. Second, the use of full-length transcriptome sequencing (Smart-seq2) in the Fujii dataset likely enhanced the resolution of transcriptionally ambiguous or rare populations, including BEST4$^+$ cells. In

contrast to droplet-based platforms (e.g., 10x Genomics), which may under-detect low-abundance transcripts, Smart-seq2 provides more comprehensive transcript coverage — especially useful for resolving small but transcriptionally distinct populations. Third, the precise timing and duration of organoid culture, passaging intervals, and endogenous cytokine levels (e.g., low-level IFN signalling from epithelial-mesenchymal interactions or media components) could vary subtly between labs and protocols. It is conceivable that low tonic levels of IFN-related signalling occurred in these datasets through unknown or uncontrolled factors, priming cells toward a *BEST4*$^{+}$ trajectory without exogenous stimulation. Thus, while the absence of IFN-γ in the Fujii and Holloway protocols suggests that this cytokine is not strictly necessary for baseline *BEST4* expression, it is likely that a combination of enhanced niche support, improved transcriptomic resolution, and subtle protocol-specific differences contributed to the appearance of distinct *BEST4*$^{+}$ clusters. These findings reinforce the complexity of rare lineage specification, and the importance of both biological context and technical sensitivity in capturing such states *in vitro*. Of note, the complementary results from this thesis and the work by Wang *et al.* not only reinforce the significance of BEST4^{+} cells but also open exciting avenues for future functional and translational research, which could leverage the organoid-based models developed and validated in this thesis, combined with the novel insights into Best4 Cell biology (particularly the importance of IFN stimulation) to further investigate their role in intestinal physiology, inflammation, and infectious diseases.

6.5 Implications For Future Research

The work presented in this thesis lays a robust methodological foundation for interrogating intestinal epithelial differentiation and plasticity using human organoid models. The development of multiplexed single-cell workflows, integrative multi-omic pipelines, and scalable lineage tracing strategies represents a major advance in tools for studying epithelial fate decisions. Nevertheless, several important directions emerge — particularly when considered alongside recent findings from Wang *et al.* (2025) — which together outline a clear roadmap for future investigation.

One immediate opportunity is the refinement of lineage tracing. The triple CellTagging framework demonstrated the feasibility of temporally resolved clonal reconstruction within hIEOs. Future studies should more densely sample clonal diversity, apply barcode enrichment, and integrate tags with spatial transcriptomics or live imaging. This would enable high-resolution tracking of fate bifurcations, particularly in response to niche signals or disease-relevant perturbations. The emergence of M-cells and Goblet Cells from shared progenitors provides a compelling proof-of-concept, but broader application to other rare lineages — such as Best4$^+$ cells — will be essential to uncover hidden fate pathways.

Second, the findings of Wang *et al.* highlight the need to refine culture conditions to support transcriptionally coherent *BEST4*$^+$*/CA7*$^+$ cells, which were largely absent in this thesis. Their demonstration that IFN-γ potently induces the Best4 lineage offers a concrete lever for revisiting *in vitro* protocols. Future work should directly test IFN-γ — alone or with factors such as BMP inhibitors or microbiota-derived metabolites — on Best4 emergence. Incorporating these insights into CellTag-based mapping could reveal whether IFN-driven Best4 cells arise via unique trajectories or represent terminal branches of absorptive intermediates.

Third, the multiomic framework offers another promising tool. Proof-of-concept integration of transcriptomic and chromatin accessibility data revealed latent readiness at Best4 loci (e.g. *SPIB*, *MEIS1*, *BEST4*). Sequencing at higher depth, alongside complementary modalities (e.g. CUT&Tag, scNOMe-seq)[351–353], could resolve the regulatory architecture of these primed but silent states. Combined with cytokine perturbations and lineage tracing, this approach could define how chromatin priming transitions into fate commitment.

Fourth, functional validation remains critical. While the single-cell transcriptomic and lineage-tracing analyses presented in this thesis provide a detailed molecular framework for understanding the specification of rare intestinal epithelial lineages, future work will require dedicated functional assays to substantiate the predicted roles, behaviours, and regulatory architectures of these cell types. One important avenue will be the functional validation of ion-transport properties in BEST4^{+} cells. Given their enrichment for pH-sensing and ion-channel genes such as *BEST4*, *OTOP2*, and associated transport regulators, electrophysiological characterisation is a logical next step. Approaches such as Ussing-chamber measurements on organoid-derived monolayers could quantify transepithelial ion transport and barrier properties under controlled stimulation, while pH-sensitive imaging or proton flux assays would allow direct assessment of the acid-sensing activities predicted by the single-cell data. These studies would be complemented by live calcium or ion imaging (for example using GCaMP or Fluo-4), which would enable real-time monitoring of purinergic, neuromodulatory, or mechanosensory responses that emerged as transcriptionally active pathways in the single-cell analyses. Similarly, the functional identity of M-cells, whose emergence was robustly induced by RANKL stimulation in this thesis, could be further validated using established transcytosis assays. The ability of RANKL-generated M-cells to internalise and transport luminal antigens—via fluorescent beads, soluble ovalbumin, or labelled bacterial particles—would provide a definitive

demonstration of their specialised immune-sampling function. Co-culture systems incorporating dendritic cells or macrophages in the basolateral compartment could additionally determine whether epithelial–immune communication is preserved *in vitro* and whether M-cell-derived antigens are capable of driving downstream immune activation. This would bridge the gap between transcriptomic prediction and physiological behaviour, establishing a direct link between RANKL-triggered transcriptional programmes and functional output. Further mechanistic work should also explore the regulatory basis of rare lineage emergence through direct perturbation experiments. CRISPR-based modulation of candidate transcription factors or signalling nodes identified in the single-cell and scATAC analyses—such as SPIB for M-cells or metabolic and ion-responsive regulators for BEST4$^+$ cells—would provide causal evidence for their roles in lineage specification. Coupling such perturbations with follow-up single-cell RNA-seq, ATAC-seq, or CellTag-based lineage tracing would allow changes in both fate trajectories and clonal dynamics to be quantified. Parallel chromatin profiling using CUT&Tag or CUT&RUN on enriched populations would validate the enhancer and transcription factor signatures inferred computationally, providing a mechanistic link between regulatory architecture and lineage identity. Finally, spatial validation—either through high-resolution smFISH (RNAscope, MERFISH) or spatial transcriptomics—would establish whether the rare states identified *in vitro* correspond to their anatomical counterparts in human tissue and would help define their microenvironmental context. Together, these functional and mechanistic assays would provide a comprehensive framework for bridging molecular signatures with physiological behaviour, thereby completing the transition from descriptive single-cell profiling to an experimentally grounded understanding of rare epithelial lineage function and regulation.

Finally, the translational potential of the platforms developed should be explored. The multiplexed, donor-matched hIEO systems described here provide an ideal scaffold for patient stratification in inflammatory bowel disease, colorectal cancer, and epithelial repair disorders. Integrating patient-specific responses to cytokines or drugs — including IFN-γ and RANKL — with single-cell readouts and clonal dynamics could enable discovery of predictive biomarkers and personalised therapies. Given the responsiveness of hIEOs to inflammatory cues, their use in immune–epithelial co-cultures or organ-on-chip systems also represents a promising avenue.

In summary, this thesis provides both the technical infrastructure and conceptual scaffolding for a next generation of studies in epithelial lineage dynamics. By integrating recent insights into rare cell biology — such as the IFN-driven emergence of Best4$^+$ cells — with the multi-layered approaches developed here, future research can move beyond descriptive cell atlases toward a mechanistic and functionally grounded understanding of epithelial fate, resilience, and dysfunction.

6.6 Final Summary

This thesis presents a comprehensive methodological framework for investigating human intestinal epithelial differentiation using organoid-based systems. Through the integration of multiplexed scRNA-seq, chromatin accessibility profiling, and temporal CellTag-based lineage tracing, it enables high-resolution analysis of epithelial identity and fate transitions. Systematic benchmarking revealed key limitations in standard protocols — particularly in generating rare lineages such as Best4^{+} cells — but also identified targeted perturbations and regulatory cues that may overcome these barriers. The successful adaptation of lineage tracing and multiomic approaches demonstrated that even transcriptionally elusive populations exhibit epigenetic priming, offering new strategies for rare cell recovery. Organoid-derived cells retained strong fidelity to primary tissue across regional, developmental, and disease contexts, confirming the biological validity of the system. In conclusion, this work makes a substantive contribution to the field by advancing the methodological toolkit for intestinal biology, introducing new strategies to interrogate epithelial fate, and establishing a robust foundation for future discoveries in rare cell biology and human epithelial modelling.

7 References

1. Shalon, D. *et al.* Profiling the human intestinal environment under physiological conditions. *Nature* **617**, 581–591 (2023).
2. Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nat Rev Immunol* **14**, 667–685 (2014).
3. Groschwitz, K. R. & Hogan, S. P. Intestinal barrier function: molecular regulation and disease pathogenesis. *J Allergy Clin Immunol* **124**, 3–20 (2009).
4. Sylvestre, M., Di Carlo, S. E. & Peduto, L. Stromal regulation of the intestinal barrier. *Mucosal Immunol* **16**, 221–231 (2023).
5. Wang, Y. *et al.* Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J Exp Med* **217**, (2020).
6. Hickey, J. W. *et al.* Organization of the human intestine at single-cell resolution. *Nature 2023 619:7970* **619**, 572–584 (2023).
7. Soderholm, A. T. & Pedicord, V. A. Intestinal epithelial cells: at the interface of the microbiota and mucosal immunity. *Immunology* **158**, 267 (2019).
8. Chesné, J., Cardoso, V. & Veiga-Fernandes, H. Neuro-immune regulation of mucosal physiology. *Mucosal Immunology* vol. 12 10–20 Preprint at https://doi.org/10.1038/s41385-018-0063-y (2019).
9. Barbara Young, James S. Lowe, Alan Stevens, John W. Heath & Philip J. Deakin. *Wheater's Functional Histology : A Text and Colour Atlas* . (Elsevier, 2006).
10. Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space and time. *Nature* **597**, 250–255 (2021).
11. Joeris, T., Müller-Luda, K., Agace, W. W. & Mowat, A. M. I. Diversity and functions of intestinal mononuclear phagocytes. *Mucosal Immunology 2017 10:4* **10**, 845–864 (2017).

12.1 David Brügger, M. & Basler, K. The diverse nature of intestinal fibroblasts in development, 2 homeostasis, and disease. *Trends Cell Biol* **33**, 834–849 (2023).

13.3 James, K. R., Elmentaite, R., Teichmann, S. A. & Hold, G. L. Redefining intestinal immunity with 4 single-cell transcriptomics. *Mucosal Immunology 2021 15:4* **15**, 531–541 (2021).

14.5 Thaiss, C. A., Zmora, N., Levy, M. & Elinav, E. The microbiome and innate immunity. *Nature 2016* 6 *535:7610* **535**, 65–74 (2016).

15.7 Mörbe, U. M. *et al.* Human gut-associated lymphoid tissues (GALT); diversity, structure, and 8 function. *Mucosal Immunology 2021 14:4* **14**, 793–802 (2021).

16.9 Man, K. & Kallies, A. Recreating immune and epithelial interactions in organoids. *Nat Immunol* 10 **25**, 1778–1780 (2024).

17.11 Okumura, R. & Takeda, K. Roles of intestinal epithelial cells in the maintenance of gut 12 homeostasis. *Experimental & Molecular Medicine 2017 49:5* **49**, e338–e338 (2017).

18.13 Y, K. & H, K. Mucosal Ecological Network of Epithelium and Immune Cells for Gut Homeostasis 14 and Tissue Healing. *Annu Rev Immunol* **35**, 119–147 (2017).

19.15 Kobayashi, N., Takahashi, D., Takano, S., Kimura, S. & Hase, K. The Roles of Peyer's Patches and 16 Microfold Cells in the Gut Immune System: Relevance to Autoimmune Diseases. *Front Immunol* 17 **10**, 2345 (2019).

20.18 Mehandru, S. & Colombel, J.-F. The intestinal barrier, an arbitrator turned provocateur in IBD. 19 *Nature Reviews Gastroenterology & Hepatology 2020 18:2* **18**, 83–84 (2020).

21.20 Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at 21 the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).

22.22 Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature 2017 551:7680* 23 **551**, 333–339 (2017).

MA Odenwald, J. T. The intestinal epithelial barrier: a therapeutic target? *Nat. Rev. Gastroenterol. Hepatol.* **14**, 9–21 (2017).

De Schepper, S. *et al.* Self-Maintaining Gut Macrophages Are Essential for Intestinal Homeostasis. *Cell* **175**, 400-415.e13 (2018).

Eshleman, E. M. & Alenghat, T. Epithelial sensing of microbiota-derived signals. *Genes and Immunity* 1–10 Preprint at https://doi.org/10.1038/s41435-021-00124-w (2021).

McCauley, H. A. *et al.* Enteroendocrine cells couple nutrient sensing to nutrient absorption by regulating ion transport. *Nat Commun* **11**, (2020).

Citi, S. *et al.* A short guide to the tight junction. *J Cell Sci* **137**, jcs261776 (2024).

Suzuki, T. Regulation of intestinal epithelial permeability by tight junctions. *Cellular and Molecular Life Sciences 2012 70:4* **70**, 631–659 (2012).

Zihni, C., Mills, C., Matter, K. & Balda, M. S. Tight junctions: from simple barriers to multifunctional molecular gates. *Nat Rev Mol Cell Biol* **17**, 564–580 (2016).

Gustafsson, J. K. & Johansson, M. E. V. The role of goblet cells and mucus in intestinal homeostasis. *Nature Reviews Gastroenterology & Hepatology 2022 19:12* **19**, 785–803 (2022).

Van Der Post, S. *et al.* Structural weakening of the colonic mucus barrier is an early event in ulcerative colitis pathogenesis. *Gut* **68**, 2142–2151 (2019).

Wehkamp, J. *et al.* Reduced Paneth cell α-defensins in ileal Crohn's disease. *Proc Natl Acad Sci U S A* **102**, 18129 (2005).

Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339-1353.e21 (2016).

Earle, K. A. *et al.* Quantitative Imaging of Gut Microbiota Spatial Organization. *Cell Host Microbe* **18**, 478–488 (2015).

35. Koropatkin, N. M., Cameron, E. A. & Martens, E. C. How glycan metabolism shapes the human gut microbiota. *Nature Reviews Microbiology 2012 10:5* **10**, 323–335 (2012).

36. Wu, W., Xiao, Z., An, W., Dong, Y. & Zhang, B. Dietary sodium butyrate improves intestinal development and function by modulating the microbial community in broilers. *PLoS One* **13**, e0197762 (2018).

37. Salvi, P. S. & Cowles, R. A. Butyrate and the Intestinal Epithelium: Modulation of Proliferation and Inflammation in Homeostasis and Disease. *Cells* **10**, (2021).

38. Parikh, K. *et al.* Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature* **567**, 49–55 (2019).

39. Clevers, H. Modeling Development and Disease with Organoids. *Cell* **165**, 1586–1597 (2016).

40. Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nature Reviews Gastroenterology & Hepatology 2018 16:1* **16**, 19–34 (2018).

41. Fawkner-Corbett, D. *et al.* Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810-826.e23 (2021).

42. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).

43. Barker, N. Adult intestinal stem cells: Critical drivers of epithelial homeostasis and regeneration. *Nature Reviews Molecular Cell Biology* vol. 15 19–33 Preprint at https://doi.org/10.1038/nrm3721 (2014).

44. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature 2009 459:7244* **459**, 262–265 (2009).

45. Choi, J. & Augenlicht, L. H. Intestinal stem cells: guardians of homeostasis in health and aging amid environmental challenges. *Experimental & Molecular Medicine 2024 56:3* **56**, 495–500 (2024).

46.1 Elmentaite, R. *et al.* Single-cell sequencing of developing human gut reveals transcriptional links
2 to childhood Crohns disease. *bioRxiv* 2020.02.06.937110 (2020) doi:10.1101/2020.02.06.937110.
47.3 Elmentaite, R. *et al.* Cells of the human intestinal tract mapped across space and time. *Nature*
4 *2021 597:7875* **597**, 250–255 (2021).
48.5 Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along
6 the Intestinal Villus Axis. *Cell* **175**, 1156-1167.e15 (2018).
49.7 Barrenetxe, J. *et al.* Distribution of the long leptin receptor isoform in brush border, basolateral
8 membrane, and cytoplasm of enterocytes. *Gut* **50**, 797–802 (2002).
50.9 McConnell, R. E. *et al.* The enterocyte microvillus is a vesicle-generating organelle. *J Cell Biol*
10 **185**, 1285 (2009).
51.11 Beumer, J. *et al.* BMP gradient along the intestinal villus axis controls zonated enterocyte and
12 goblet cell states. *Cell Rep* **38**, (2022).
52.13 Dillon, A. & Lo, D. D. M cells: Intelligent engineering of mucosal immune surveillance. *Front*
14 *Immunol* **10**, 461052 (2019).
53.15 Mabbott, N. A., Donaldson, D. S., Ohno, H., Williams, I. R. & Mahajan, A. Microfold (M) cells:
16 Important immunosurveillance posts in the intestinal epithelium. *Mucosal Immunol* **6**, 666–677
17 (2013).
54.18 Del Castillo, D. & Lo, D. D. Deciphering the M-cell niche: insights from mouse models on how
19 microfold cells “know” where they are needed. *Front Immunol* **15**, 1400739 (2024).
55.20 Rios, D. *et al.* Antigen sampling by intestinal M cells is the principal pathway initiating mucosal
21 IgA production to commensal enteric bacteria. *Mucosal Immunol* **9**, 907–916 (2016).
56.22 Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature*
23 **469**, 415–418 (2011).

Farin, H. F., Van Es, J. H. & Clevers, H. Redundant sources of Wnt regulate intestinal stem cells and promote formation of paneth cells. *Gastroenterology* **143**, (2012).

de Lau, W. *et al.* Peyer's patch M cells derived from Lgr5(+) stem cells require SpiB and are induced by RankL in cultured 'miniguts'. *Mol Cell Biol* **32**, 3639–3647 (2012).

dos Reis, D. C. *et al.* CFTR High Expresser BEST4+ cells are pH-sensing neuropod cells: new implications for intestinal physiology and Cystic Fibrosis disease. *bioRxiv* (2025) doi:10.1101/2025.01.24.634747.

Malonga, T., Vialaneix, N. & Beaumont, M. BEST4+ cells in the intestinal epithelium. *Am J Physiol Cell Physiol* **326**, C1345–C1352 (2024).

Buczacki, S. J. A. *et al.* Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature 2013 495:7439* **495**, 65–69 (2013).

Qi, Z. *et al.* BMP restricts stemness of intestinal Lgr5+ stem cells by directly suppressing their signature genes. *Nature Communications 2017 8:1* **8**, 1–14 (2017).

KA, K. & RD, N. Goblet cells: multifaceted players in immunity at mucosal surfaces. *Mucosal Immunol* **11**, 1551–1557 (2018).

Beumer, J. *et al.* BMP gradient along the intestinal villus axis controls zonated enterocyte and goblet cell states. *Cell Rep* **38**, (2022).

Zeve, D. *et al.* Robust differentiation of human enteroendocrine cells from intestinal stem cells. *Nature Communications 2022 13:1* **13**, 1–20 (2022).

Bohórquez, D. V. *et al.* An Enteroendocrine Cell – Enteric Glia Connection Revealed by 3D Electron Microscopy. *PLoS One* **9**, e89881 (2014).

Yan, K. S. *et al.* Intestinal Enteroendocrine Lineage Cells Possess Homeostatic and Injury-Inducible Stem Cell Activity. *Cell Stem Cell* **21**, 78-90.e6 (2017).

68. McRae, A. N. *et al.* Regulator of G-protein signaling expression in human intestinal enteroendocrine cells and potential role in satiety hormone secretion in health and obesity. *EBioMedicine* **107**, 105283 (2024).

69. Xiong, Z. *et al.* Intestinal Tuft-2 cells exert antimicrobial immunity via sensing bacterial metabolite N-undecanoylglycine. *Immunity* **55**, 686-700.e7 (2022).

70. Schneider, C., O'Leary, C. E. & Locksley, R. M. Regulation of immune responses by tuft cells. *Nature Reviews Immunology 2019 19:9* **19**, 584–593 (2019).

71. Feng, X., Flüchter, P., De Tenorio, J. C. & Schneider, C. Tuft cells in the intestine, immunity and beyond. *Nature Reviews Gastroenterology & Hepatology 2024 21:12* **21**, 852–868 (2024).

72. Huang, L. *et al.* Tuft cells act as regenerative stem cells in the human intestine. *Nature 2024 634:8035* **634**, 929–935 (2024).

73. Lei, H. *et al.* Tuft cells utilize taste signaling molecules to respond to the pathobiont microbe Ruminococcus gnavus in the proximal colon. *Front Immunol* **14**, 1259521 (2023).

74. Yin, X. *et al.* Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny. *Nat Methods* **11**, 106–112 (2014).

75. Campàs, O., Noordstra, I. & Yap, A. S. Adherens junctions as molecular regulators of emergent tissue mechanics. *Nature Reviews Molecular Cell Biology 2023 25:4* **25**, 252–269 (2023).

76. Schmitz, H. Altered tight junction structure contributes to the impaired epithelial barrier function in ulcerative colitis. *Gastroenterology* **116**, 301–309 (1999).

77. Hoffman, B. D. & Yap, A. S. Towards a Dynamic Understanding of Cadherin-Based Mechanobiology. *Trends Cell Biol* **25**, 803–814 (2015).

78. Takeichi, M. Dynamic contacts: Rearranging adherens junctions to drive epithelial remodelling. *Nat Rev Mol Cell Biol* **15**, 397–410 (2014).

79. McCutcheon, S., Stout, R. F. & Spray, D. C. The dynamic Nexus: gap junctions control protein localization and mobility in distinct and surprising ways. *Scientific Reports 2020 10:1* **10**, 1–18 (2020).

80. Goodenough, D. A. & Paul, D. L. Beyond the gap: Functions of unpaired connexon channels. *Nat Rev Mol Cell Biol* **4**, 285–294 (2003).

81. Delva, E., Tucker, D. K. & Kowalczyk, A. P. The desmosome. *Cold Spring Harb. Perspect. Biol.* **1**, a002543 (2009).

82. Nekrasova, O. & Green, K. J. Desmosome assembly and dynamics. *Trends Cell Biol.* **23**, 537–546 (2013).

83. Price, A. J. *et al.* Mechanical loading of desmosomes depends on the magnitude and orientation of external stress. *Nature Communications 2018 9:1* **9**, 1–11 (2018).

84. Kowalczyk, A. P., Bornslaeger, E. A., Norvell, S. M., Palka, H. L. & Green, K. J. Desmosomes: Intercellular Adhesive Junctions Specialized for Attachment of Intermediate Filaments. *Int Rev Cytol* **185**, 237–302 (1998).

85. Garrod, D. & Chidgey, M. Desmosome structure, composition and function. *Biochim Biophys Acta Biomembr* **1778**, 572–587 (2008).

86. Kai, Y. Intestinal villus structure contributes to even shedding of epithelial cells. (2021) doi:10.1016/j.bpj.2021.01.003.

87. Meran, L., Baulies, A. & Li, V. S. W. Intestinal Stem Cell Niche: The Extracellular Matrix and Cellular Components. *Stem Cells Int* **2017**, 7970385 (2017).

88. Fujii, M. *et al.* Human Intestinal Organoids Maintain Self-Renewal Capacity and Cellular Diversity in Niche-Inspired Culture Condition. *Cell Stem Cell* **23**, 787-793.e6 (2018).

89. Liu, J. *et al.* Wnt/β-catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduction and Targeted Therapy 2021 7:1* **7**, 1–23 (2022).

90. Van Es, J. H. *et al.* Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature 2005 435:7044* **435**, 959–963 (2005).

91. Sancho, R., Cremona, C. A. & Behrens, A. Stem cell and progenitor fate in the mammalian intestine: Notch and lateral inhibition in homeostasis and disease. *EMBO Rep* **16**, 571–581 (2015).

92. Buczacki, S. Fate plasticity in the intestine: The devil is in the detail. *World Journal of Gastroenterology* vol. 25 3116–3122 Preprint at https://doi.org/10.3748/wjg.v25.i25.3116 (2019).

93. Tetteh, P. W., Farin, H. F. & Clevers, H. Plasticity within stem cell hierarchies in mammalian epithelia. *Trends Cell Biol* **25**, 100–108 (2015).

94. Holzem, M., Boutros, M. & Holstein, T. W. The origin and evolution of Wnt signalling. *Nature Reviews Genetics 2024 25:7* **25**, 500–512 (2024).

95. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev* **17**, 1709–1713 (2003).

96. Bray, S. J. Notch signalling: A simple pathway becomes complex. *Nature Reviews Molecular Cell Biology* vol. 7 678–689 Preprint at https://doi.org/10.1038/nrm2009 (2006).

97. Bray, S. J. & Bigas, A. Modes of Notch signalling in development and disease. *Nature Reviews Molecular Cell Biology 2025* 1–16 (2025) doi:10.1038/s41580-025-00835-2.

98. Xie, Z. *et al.* Recent developments on BMPs and their antagonists in inflammatory bowel diseases. *Cell Death Discovery 2023 9:1* **9**, 1–10 (2023).

99. Khoramjoo, S. M. *et al.* Overview of Three Proliferation Pathways (Wnt, Notch, and Hippo) in Intestine and Immune System and Their Role in Inflammatory Bowel Diseases (IBDs). *Front Med (Lausanne)* **9**, 865131 (2022).

100. Xie, Z. *et al.* Emerging roles of the Hedgehog signalling pathway in inflammatory bowel disease. *Cell Death Discovery 2021 7:1* **7**, 1–9 (2021).

101. Neunlist, M. *et al.* Enteric glia inhibit intestinal epithelial cell proliferation partly through a TGF-β1-dependent pathway. *Am J Physiol Gastrointest Liver Physiol* **292**, (2007).

102. Buongusto, F. *et al.* Disruption of the Hedgehog signaling pathway in inflammatory bowel disease fosters chronic intestinal inflammation. *Clin Exp Med* **17**, 351–369 (2017).

103. Deng, Z. *et al.* TGF-β signaling in health, disease and therapeutics. *Signal Transduction and Targeted Therapy 2024 9:1* **9**, 1–40 (2024).

104. Flanagan, D. J. *et al.* Epithelial TGFβ engages growth-factor signalling to circumvent apoptosis and drive intestinal tumourigenesis with aggressive features. *Nature Communications 2022 13:1* **13**, 1–18 (2022).

105. Deng, F., Wu, Z., Zou, F., Wang, S. & Wang, X. The Hippo–YAP/TAZ Signaling Pathway in Intestinal Self-Renewal and Regeneration After Injury. *Front Cell Dev Biol* **10**, 894737 (2022).

106. Xie, Z. *et al.* The role of the Hippo pathway in the pathogenesis of inflammatory bowel disease. *Cell Death & Disease 2021 12:1* **12**, 1–14 (2021).

107. Nowarski, R., Jackson, R. & Flavell, R. A. The Stromal Intervention: Regulation of Immunity and Inflammation at the Epithelial-Mesenchymal Barrier. *Cell* vol. 168 362–375 Preprint at https://doi.org/10.1016/j.cell.2016.11.040 (2017).

108. Colozza, G., Park, S. Y. & Koo, B. K. Clone wars: From molecules to cell competition in intestinal stem cell homeostasis and disease. *Experimental & Molecular Medicine 2022 54:9* **54**, 1367–1378 (2022).

109. de Sousa e Melo, F. & de Sauvage, F. J. Cellular Plasticity in Intestinal Homeostasis and Disease. *Cell Stem Cell* **24**, 54–64 (2019).

110. Yan, Z., Liu, Y. & Yuan, Y. The plasticity of epithelial cells and its potential in the induced differentiation of gastric cancer. *Cell Death Discovery 2024 10:1* **10**, 1–16 (2024).

1. Van Der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu Rev Physiol* **71**, 241–260 (2009).

2. Sansom, O. J. *et al.* Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev* **18**, 1385–1390 (2004).

3. Cuomo, A. S. E. *et al.* Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nature Communications 2020 11:1* **11**, 1–14 (2020).

4. Dydensborg, A. B., Herring, E., Auclair, J., Tremblay, E. & Beaulieu, J. F. Normalizing genes for quantitative RT-PCR in differentiating human intestinal epithelial cells and adenocarcinomas of the colon. *Am J Physiol Gastrointest Liver Physiol* **290**, 1067–1074 (2006).

5. Krausova, M. & Korinek, V. Wnt signaling in adult intestinal stem cells and cancer. *Cell Signal* **26**, 570–579 (2014).

6. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009).

7. Zhou, D. *et al.* Mst1 and Mst2 maintain hepatocyte quiescence and suppress the development of hepatocellular carcinoma through inactivation of the Yap1 oncogene. *Cancer Cell* **16**, 425 (2009).

8. Yui, S. *et al.* YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell* **22**, 35-49.e7 (2018).

9. Haerinck, J., Goossens, S. & Berx, G. The epithelial–mesenchymal plasticity landscape: principles of design and mechanisms of regulation. *Nature Reviews Genetics 2023 24:9* **24**, 590–609 (2023).

10. Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation* **119**, 1420–1428 (2009).

11. Su, J. *et al.* Cell–cell communication: new insights and clinical implications. *Signal Transduction and Targeted Therapy 2024 9:1* **9**, 1–52 (2024).

121. Ko, S. C. W. *et al.* Paracrine cyclooxygenase-2-mediated signalling by macrophages promotes tumorigenic progression of intestinal epithelial cells. *Oncogene 2002 21:47* **21**, 7175–7186 (2002).

123. Degirmenci, B., Valenta, T., Dimitrieva, S., Hausmann, G. & Basler, K. GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells. *Nature 2018 558:7710* **558**, 449–453 (2018).

124. De Lau, W. *et al.* Lgr5 homologues associate with Wnt receptors and mediate R-spondin signalling. *Nature* **476**, 293–297 (2011).

125. Clark, H. R. *et al.* Epigentically regulated digital signalling defines epithelial innate immunity at the tissue level. *Nat Commun* **12**, 1–13 (2021).

126. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

127. Moussa, L. *et al.* BMP Antagonists Secreted by Mesenchymal Stromal Cells Improve Colonic Organoid Formation: Application for the Treatment of Radiation-induced Injury. *Cell Transplant* **29**, 0963689720929683 (2020).

128. Kinchen, J. *et al.* Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. *Cell* **175**, 372-386.e17 (2018).

129. Ehata, S. & Miyazono, K. Bone Morphogenetic Protein Signaling in Cancer; Some Topics in the Recent 10 Years. *Front Cell Dev Biol* **10**, 883523 (2022).

130. Hofmann, C. *et al.* Cell-Cell Contacts Prevent Anoikis in Primary Human Colonic Epithelial Cells. *Gastroenterology* **132**, 587–600 (2007).

131. Su, J. *et al.* Cell–cell communication: new insights and clinical implications. *Signal Transduction and Targeted Therapy 2024 9:1* **9**, 1–52 (2024).

132. Armingol, E., Officer, A., Harismendy, O. & Lewis, N. E. Deciphering cell–cell interactions and communication from gene expression. *Nat Rev Genet* **22**, 71–88 (2021).

1. Bazellières, E. *et al.* Control of cell-cell forces and collective cell dynamics by the intercellular adhesome. *Nat Cell Biol* **17**, 409–420 (2015).

4. Staehelin, L. A. Three types of gap junctions interconnecting intestinal epithelial cells visualized by freeze-etching. *Proc Natl Acad Sci U S A* **69**, 1318–1321 (1972).

5. Green, K. J. & Simpson, C. L. Desmosomes: New perspectives on a classic. *Journal of Investigative Dermatology* **127**, 2499–2515 (2007).

7. Dupont, S. *et al.* Role of YAP/TAZ in mechanotransduction. *Nature* **474**, 179–184 (2011).

8. Vooijs, M., Liu, Z. & Kopan, R. Notch: Architect, Landscaper, and Guardian of the Intestine. *Gastroenterology* **141**, 448 (2011).

10. Bray, S. J. Notch signalling in context. *Nat Rev Mol Cell Biol* **17**, 722–735 (2016).

11. Bonnans, C., Chou, J. & Werb, Z. Remodelling the extracellular matrix in development and disease. *Nature Reviews Molecular Cell Biology 2014 15:12* **15**, 786–801 (2014).

13. Vilardi, A., Przyborski, S., Mobbs, C., Rufini, A. & Tufarelli, C. Current understanding of the interplay between extracellular matrix remodelling and gut permeability in health and disease. *Cell Death Discovery 2024 10:1* **10**, 1–11 (2024).

16. Hynes, R. O. Integrins: Bidirectional, Allosteric Signaling Machines. *Cell* **110**, 673–687 (2002).

17. Schaller, M. D. Biochemical signals and biological responses elicited by the focal adhesion kinase. *Biochim Biophys Acta Mol Cell Res* **1540**, 1–21 (2001).

19. Mitra, S. K., Hanson, D. A. & Schlaepfer, D. D. Focal adhesion kinase: In command and control of cell motility. *Nat Rev Mol Cell Biol* **6**, 56–68 (2005).

21. Wörthmüller, J. & Rüegg, C. The Crosstalk between FAK and Wnt Signaling Pathways in Cancer and Its Therapeutic Implication. *Int J Mol Sci* **21**, 9107 (2020).

23. Parsons, J. T. Focal adhesion kinase: The first ten years. *J Cell Sci* **116**, 1409–1416 (2003).

1. Geiger, B., Bershadsky, A., Pankov, R. & Yamada, K. M. Transmembrane extracellular matrix-cytoskeleton crosstalk. *Nat Rev Mol Cell Biol* **2**, 793–805 (2001).

3. Gjorevski, N. *et al.* Designer matrices for intestinal stem cell and organoid culture. *Nature* **539**, 560–564 (2016).

5. Spence, J. R. *et al.* Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature* **470**, 105–110 (2011).

7. Kostic, A. D., Xavier, R. J. & Gevers, D. The Microbiome in Inflammatory Bowel Diseases: Current Status and the Future Ahead. *Gastroenterology* **146**, 1489 (2014).

9. Chelakkot, C., Ghim, J. & Ryu, S. H. Mechanisms regulating intestinal barrier integrity and its pathological implications. *Experimental & Molecular Medicine 2018 50:8* **50**, 1–9 (2018).

11. Fink, M. P. Intestinal epithelial hyperpermeability: Update on the pathogenesis of gut mucosal barrier dysfunction in critical illness. *Curr Opin Crit Care* **9**, 143–151 (2003).

13. Neurath, M. F., Artis, D. & Becker, C. The intestinal barrier: a pivotal role in health, inflammation, and cancer. *Lancet Gastroenterol Hepatol* **10**, 573–592 (2025).

15. Wilkins, L. J., Monga, M. & Miller, A. W. Defining Dysbiosis for a Cluster of Chronic Diseases. *Scientific Reports 2019 9:1* **9**, 1–10 (2019).

17. Okumura, R. & Takeda, K. Roles of intestinal epithelial cells in the maintenance of gut homeostasis. *Exp Mol Med* **49**, e338–e338 (2017).

19. Zhao, M., Tang, S., Xin, J., Wei, Y. & Liu, D. Reactive oxygen species induce injury of the intestinal epithelium during hyperoxia. *Int J Mol Med* **41**, 322 (2017).

21. Wang, H., Zhai, N., Chen, Y., Fu, C. & Huang, K. OTA induces intestinal epithelial barrier dysfunction and tight junction disruption in IPEC-J2 cells through ROS/Ca2+-mediated MLCK activation. *Environmental Pollution* **242**, 106–112 (2018).

157 1. Muro, P. *et al.* The emerging role of oxidative stress in inflammatory bowel disease. *Front*
2 *Endocrinol (Lausanne)* **15**, 1390351 (2024).
158 3. Camilleri, M. The Leaky Gut: Mechanisms, Measurement and Clinical Implications in Humans.
4 *Gut* **68**, 1516 (2019).
159 5. Mu, Q., Kirby, J., Reilly, C. M. & Luo, X. M. Leaky gut as a danger signal for autoimmune diseases.
6 *Front Immunol* **8**, 269575 (2017).
160 7. Aleman, R. S., Moncada, M. & Aryana, K. J. Leaky Gut and the Ingredients That Help Treat It: A
8 Review. *Molecules 2023, Vol. 28, Page 619* **28**, 619 (2023).
161 9. Turner, J. R. Intestinal mucosal barrier function in health and disease. *Nature Reviews*
10 *Immunology 2009 9:11* **9**, 799–809 (2009).
162 11. Kaminsky, L. W., Al-Sadi, R. & Ma, T. Y. IL-1β and the Intestinal Epithelial Tight Junction Barrier.
12 *Front Immunol* **12**, 767456 (2021).
163 13. Ruder, B., Atreya, R. & Becker, C. Tumour necrosis factor alpha in intestinal homeostasis and gut
14 related diseases. *Int J Mol Sci* **20**, (2019).
164 15. Pabst, O. & Cerovic, V. Interferon-γ sensing by epithelial cells tames gut inflammation. *Nat*
16 *Immunol* **25**, 9–10 (2024).
165 17. Giuffrida, P., Caprioli, F., Facciotti, F. & Di Sabatino, A. The role of interleukin-13 in chronic
18 inflammatory intestinal disorders. *Autoimmun Rev* **18**, 549–555 (2019).
166 19. Górecka, A. *et al.* Biochemical Modulators of Tight Junctions (TJs): Occludin, Claudin-2 and
20 Zonulin as Biomarkers of Intestinal Barrier Leakage in the Diagnosis and Assessment of
21 Inflammatory Bowel Disease Progression. *Molecules 2024, Vol. 29, Page 4577* **29**, 4577 (2024).
167 22. Ding, N. S. *et al.* Metabonomics and the gut microbiome associated with primary response to
23 anti-TNF therapy in Crohn's disease. *J Crohns Colitis* **14**, 1090–1102 (2020).

1. Zeissig, S. *et al.* Changes in expression and distribution of claudin 2, 5 and 8 lead to discontinuous tight junctions and barrier dysfunction in active Crohn's disease. *Gut* **56**, 61 (2006).
4. Bhatnagar, G. *et al.* Inflammation and fibrosis in Crohn's disease: location-matched histological correlation of small bowel ultrasound features. *Abdom Radiol (NY)* **46**, 144 (2020).
6. Rieder, F., Brenmoehl, J., Leeb, S., Schölmerich, J. & Rogler, G. Wound healing and fibrosis in intestinal disease. *Gut* **56**, 130 (2007).
8. Blanpain, C. & Fuchs, E. Plasticity of epithelial stem cells in tissue regeneration. *Science* **344**, 1242281 (2014).
10. Rieder, F., Mukherjee, P. K., Massey, W. J., Wang, Y. & Fiocchi, C. Fibrosis in IBD: from pathogenesis to therapeutic targets. *Gut* **73**, 854 (2024).
12. Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell* **139**, 871–890 (2009).
14. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178–196 (2014).
16. Grivennikov, S. I. *et al.* Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature 2012 491:7423* **491**, 254–258 (2012).
18. Michels, B. E. *et al.* Human colon organoids reveal distinct physiologic and oncogenic Wnt responses. *Journal of Experimental Medicine* **216**, 704–720 (2019).
20. Fre, S. *et al.* Notch signals control the fate of immature progenitor cells in the intestine. *Nature* **435**, 964–968 (2005).
22. VanDussen, K. L. *et al.* Notch signaling modulates proliferation and differentiation of intestinal crypt base columnar stem cells. *Development* **139**, 488–497 (2012).

1. Zacharias, W. J. *et al.* Hedgehog is an anti-inflammatory epithelial signal for the intestinal lamina propria. *Gastroenterology* **138**, 2368 (2010).

3. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).

5. Stephani, J., Radulovic, K. & Niess, J. H. Gut microbiota, probiotics and inflammatory bowel disease. *Arch Immunol Ther Exp (Warsz)* **59**, 161–177 (2011).

7. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol* **14**, (2016).

9. Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science (1979)* **336**, 1262–1267 (2012).

11. Thaiss, C. A., Zmora, N., Levy, M. & Elinav, E. The microbiome and innate immunity. *Nature 2016 535:7610* **535**, 65–74 (2016).

13. Donohoe, D. R. *et al.* The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab* **13**, 517–526 (2011).

15. Ahmad, R., Sorrell, M. F., Batra, S. K., Dhawan, P. & Singh, A. B. Gut permeability and mucosal inflammation: bad, good or context dependent. *Mucosal Immunology 2017 10:2* **10**, 307–317 (2017).

17. Allegretti, J. R., Khanna, S., Mullish, B. H. & Feuerstadt, P. The Progression of Microbiome Therapeutics for the Management of Gastrointestinal Diseases and Beyond. *Gastroenterology* **167**, 885–902 (2024).

21. Andrews, C., McLean, M. H. & Durum, S. K. Cytokine tuning of intestinal epithelial function. *Front Immunol* **9**, 368738 (2018).

1. Cani, P. D., Depommier, C., Derrien, M., Everard, A. & de Vos, W. M. Akkermansia muciniphila: paradigm for next-generation beneficial microorganisms. *Nature Reviews Gastroenterology & Hepatology 2022 19:10* **19**, 625–637 (2022).

4. Biagioli, M. *et al.* Metabolic variability of a multispecies probiotic preparation impacts on the anti-inflammatory activity. *Front Pharmacol* **8**, 278842 (2017).

6. Tseng, C. H. *et al.* Development of live biotherapeutic products: a position statement of Asia-Pacific Microbiota Consortium. *Gut* **74**, 706–713 (2025).

8. Zepeda-Rivera, M. *et al.* A distinct Fusobacterium nucleatum clade dominates the colorectal cancer niche. *Nature 2024 628:8007* **628**, 424–432 (2024).

10. Kahn, M. Can we safely target the WNT pathway? *Nature Reviews Drug Discovery 2014 13:7* **13**, 513–532 (2014).

12. Yu, F. *et al.* Wnt/β-catenin signaling in cancers and targeted therapies. *Signal Transduction and Targeted Therapy 2021 6:1* **6**, 1–24 (2021).

14. Yin, X. *et al.* Engineering Stem Cell Organoids. *Cell Stem Cell* **18**, 25–38 (2016).

15. Lancaster, M. A. & Knoblich, J. A. Organogenesis in a dish: Modeling development and disease using organoid technologies. *Science (1979)* **345**, (2014).

17. Jung, P. *et al.* Isolation and in vitro expansion of human colonic stem cells. *Nat Med* **17**, 1225–1227 (2011).

18. Verstegen, M. M. A. *et al.* Clinical applications of human organoids. *Nat Med* **31**, 409–421 (2025).

20. Passaniti, A., Kleinman, H. K. & Martin, G. R. Matrigel: history/background, uses, and future applications. *J Cell Commun Signal* **16**, 621–626 (2022).

22. Kaur, S., Kaur, I., Rawal, P., Tripathi, D. M. & Vasudevan, A. Non-matrigel scaffolds for organoid cultures. *Cancer Lett* **504**, 58–66 (2021).

1. Urbischek, M. *et al.* Organoid culture media formulated with growth factors of defined cellular activity. *Sci Rep* **9**, 1–11 (2019).

2. Dotti, I. *et al.* P0204 Impact of crypt niche factors and oxygen gradient on intestinal epithelial differentiation and inflammatory response. *J Crohns Colitis* **19**, i608–i609 (2025).

3. He, S. *et al.* Stiffness Restricts the Stemness of the Intestinal Stem Cells and Skews Their Differentiation Towards Goblet Cells. *Gastroenterology* **164**, 1137 (2023).

4. Ye, S. *et al.* Accelerated production of human epithelial organoids in a miniaturized spinning bioreactor. *Cell Reports Methods* **4**, 100903 (2024).

5. Hetzel, L. A., Ali, A., Corbo, V. & Hankemeier, T. Microfluidics and Organoids, the Power Couple of Developmental Biology and Oncology Studies. *Int J Mol Sci* **24**, 10882 (2023).

6. Urresti, J. *et al.* Cortical organoids model early brain development disrupted by 16p11.2 copy number variants in autism. *Molecular Psychiatry 2021 26:12* **26**, 7560–7580 (2021).

7. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature 2013 501:7467* **501**, 373–379 (2013).

8. Taelman, J., Diaz, M. & Guiu, J. Human Intestinal Organoids: Promise and Challenge. *Front Cell Dev Biol* **10**, 854740 (2022).

9. Rauth, S., Karmakar, S., Batra, S. K. & Ponnusamy, M. P. Recent advances in organoid development and applications in disease modeling. *Biochim Biophys Acta Rev Cancer* 188527 (2021) doi:10.1016/j.bbcan.2021.188527.

10. Dekkers, J. F. *et al.* A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nature Medicine 2013 19:7* **19**, 939–945 (2013).

11. Tian, C. mei *et al.* Stem cell-derived intestinal organoids: a novel modality for IBD. *Cell Death Discovery 2023 9:1* **9**, 1–16 (2023).

1. Aguilar, C. *et al.* Organoids as host models for infection biology – a review of methods. *Exp Mol Med* **53**, 1471–1482 (2021).
3. Sarvestani, S. K. *et al.* Induced organoids derived from patients with ulcerative colitis recapitulate colitic reactivity. *Nature Communications 2021 12:1* **12**, 1–18 (2021).
5. Dennison, T. W. *et al.* Patient-derived organoid biobank identifies epigenetic dysregulation of intestinal epithelial MHC-I as a novel mechanism in severe Crohn's Disease. *Gut* **73**, 1464–1477 (2024).
8. Murthy, S. *et al.* Single-cell transcriptomics of rectal organoids from individuals with perianal fistulizing Crohn's disease reveals patient-specific signatures. *Sci Rep* **14**, 26142 (2024).
10. Paužuolis, M. *et al.* SARS-CoV-2 tropism to intestinal but not gastric epithelial cells is defined by limited ACE2 expression. *Stem Cell Reports* **19**, 629–638 (2024).
12. Yan, J. *et al.* Intestinal organoids to model Salmonella infection and its impact on progenitors. *Sci Rep* **14**, 1–10 (2024).
14. Saavedra, P. H. V. *et al.* Apoptosis of intestinal epithelial cells restricts Clostridium difficile infection in a model of pseudomembranous colitis. *Nature Communications 2018 9:1* **9**, 1–10 (2018).
17. Han, X. *et al.* Creating a More Perfect Union: Modeling Intestinal Bacteria-Epithelial Interactions Using Organoids. *Cell Mol Gastroenterol Hepatol* **12**, 769 (2021).
19. Kim, J., Koo, B. K. & Knoblich, J. A. Human organoids: model systems for human biology and medicine. *Nature Reviews Molecular Cell Biology 2020 21:10* **21**, 571–584 (2020).
21. LeSavage, B. L., Suhar, R. A., Broguiere, N., Lutolf, M. P. & Heilshorn, S. C. Next-generation cancer organoids. *Nature Materials 2021 21:2* **21**, 143–159 (2021).
23. Ooft, S. N. *et al.* Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. *Sci Transl Med* **11**, 2574 (2019).

1. Poling, H. M. *et al.* Human pluripotent stem cell-derived organoids repair damaged bowel in vivo. *Cell Stem Cell* **31**, 1513-1523.e7 (2024).

4. Jalan-Sakrikar, N., Brevini, T., Huebert, R. C. & Sampaziotis, F. Organoids and regenerative hepatology. *Hepatology* **77**, 305 (2022).

5. Roberto de Barros, N. *et al.* Engineered Organoids for Biomedical Applications. *Adv Drug Deliv Rev* **203**, 115142 (2023).

7. Yip, S., Wang, N. & Sugimura, R. Give Them Vasculature and Immune Cells: How to Fill the Gap of Organoids. *Cells Tissues Organs* **212**, 369 (2023).

9. Naderi-Meshkin, H. *et al.* Vascular organoids: unveiling advantages, applications, challenges, and disease modelling strategies. *Stem Cell Res Ther* **14**, 1–14 (2023).

11. Min, S., Kim, S. & Cho, S. W. Gastrointestinal tract modeling using organoids engineered with cellular and microbiota niches. *Experimental & Molecular Medicine 2020 52:2* **52**, 227–237 (2020).

13. Kasendra, M. *et al.* Development of a primary human Small Intestine-on-a-Chip using biopsy-derived organoids. *Scientific Reports 2018 8:1* **8**, 1–14 (2018).

15. Huang, Y. *et al.* Research Progress, Challenges, and Breakthroughs of Organoids as Disease Models. *Front Cell Dev Biol* **9**, 740574 (2021).

17. Wang, Q., Yuan, F., Zuo, X. & Li, M. Breakthroughs and challenges of organoid models for assessing cancer immunotherapy: a cutting-edge tool for advancing personalised treatments. *Cell Death Discovery 2025 11:1* **11**, 1–13 (2025).

20. Ethics need to keep up with human brain organoid research. *Nature Reviews Bioengineering* **2**, 711 (2024).

22. Li, X. & Wang, C. Y. From bulk, single-cell to spatial RNA sequencing. *Int J Oral Sci* **13**, 1–6 (2021).

24. De Rop, F. V. *et al.* Systematic benchmarking of single-cell ATAC-sequencing protocols. *Nat Biotechnol* **42**, 916–926 (2024).

1. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* **50**, 1–14 (2018).

3. Brazovskaja, A., Treutlein, B. & Camp, J. G. High-throughput single-cell transcriptomics on organoids. (2019) doi:10.1016/j.copbio.2018.11.002.

5. Ziegenhain, C. *et al.* Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell* **65**, 631-643.e4 (2017).

7. Sant, P., Rippe, K. & Mallm, J. P. Approaches for single-cell RNA sequencing across tissues and cell types. *Transcription* **14**, 127 (2023).

9. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell* **157**, 714–725 (2014).

11. Kelly, N. H., Huynh, N. P. T. & Guilak, F. Single cell RNA-sequencing reveals cellular heterogeneity and trajectories of lineage specification during murine embryonic limb development. *Matrix Biol* **89**, 1–10 (2020).

14. Villani, A. C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science (1979)* **356**, (2017).

16. Xu, Q. *et al.* An integrated transcriptomic cell atlas of human endoderm-derived organoids. *Nature Genetics 2025 57:5* **57**, 1201–1212 (2025).

18. Pleguezuelos-Manzano, C. *et al.* Dual RNA sequencing of a co-culture model of Pseudomonas aeruginosa and human 2D upper airway organoids. *Scientific Reports 2025 15:1* **15**, 1–14 (2025).

20. Lake, B. B. *et al.* An atlas of healthy and injured cell states and niches in the human kidney. *Nature 2023 619:7970* **619**, 585–594 (2023).

22. Cao, M. *et al.* Single-cell transcriptomic analysis reveals gut microbiota-immunotherapy synergy through modulating tumor microenvironment. *Signal Transduct Target Ther* **10**, 140 (2025).

246 1. Hong, D. *et al.* Integrative analysis of single-cell RNA-seq and gut microbiome metabarcoding
2 data elucidates macrophage dysfunction in mice with DSS-induced ulcerative colitis.
3 *Communications Biology 2024 7:1* **7**, 1–14 (2024).
247 4. Tanaka, Y., Cakir, B., Xiang, Y. & Sullivan, G. J. Synthetic Analyses of Single-Cell Transcriptomes
5 from Multiple Brain Organoids and Fetal Brain. *CellReports* **30**, 1682-1689.e3 (2020).
248 6. Chen, J. *et al.* Single-cell transcriptome analysis identifies distinct cell types and niche signaling
7 in a primary gastric organoid model. *Sci Rep* **9**, 1–12 (2019).
249 8. Pavlidis, P. *et al.* Cytokine responsive networks in human colonic epithelial organoids unveil a
9 molecular classification of inflammatory bowel disease. (2022) doi:10.1016/j.celrep.2022.111439.
250 10. Niklinska-Schirtz, B. J. *et al.* Ileal Derived Organoids From Crohn's Disease Patients Show Unique
11 Transcriptomic and Secretomic Signatures. *CMGH* **12**, 1267–1280 (2021).
251 12. Oliver, A. J. *et al.* Single-cell integration reveals metaplasia in inflammatory gut diseases. *Nature*
13 *2024 635:8039* **635**, 699–707 (2024).
252 14. Grandi, F. C., Modi, H., Kampman, L. & Corces, M. R. Chromatin accessibility profiling by ATAC-
15 seq. *Nature Protocols 2022 17:6* **17**, 1518–1552 (2022).
253 16. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory
17 epigenome. *Nature Reviews Genetics 2018 20:4* **20**, 207–220 (2019).
254 18. Rispal, J., Escaffit, F. & Trouche, D. Chromatin Dynamics in Intestinal Epithelial Homeostasis: A
19 Paradigm of Cell Fate Determination versus Cell Plasticity. *Stem Cell Rev Rep* **16**, 1062 (2020).
255 20. Yin, S. *et al.* Functional genomics analysis of human colon organoids identifies key transcription
21 factors. *Physiol Genomics* **52**, 234 (2020).
256 22. Deng, L. *et al.* Frizzled5 controls murine intestinal epithelial cell plasticity through organization
23 of chromatin accessibility. *Dev Cell* **60**, 352-363.e6 (2024).

1. Mimitou, E. P. *et al.* Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nature Biotechnology 2021 39:10* **39**, 1246–1258 (2021).
3. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol* **19**, 1–12 (2018).
5. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).
6. Lee, J., Hyeon, D. Y. & Hwang, D. Single-cell multiomics: technologies and data analysis methods. *Experimental and Molecular Medicine* vol. 52 1428–1442 Preprint at https://doi.org/10.1038/s12276-020-0420-2 (2020).
9. Zhao, T. *et al.* Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. *Nature 2021* 1–7 (2021) doi:10.1038/s41586-021-04217-4.
11. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods 2017 14:9* **14**, 865–868 (2017).
13. Zhou, Z., Ye, C., Wang, J. & Zhang, N. R. Surface protein imputation from single cell transcriptomes by deep neural networks. *Nature Communications 2020 11:1* **11**, 1–10 (2020).
15. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol* **19**, 1–12 (2018).
17. Mead, B. E. *et al.* Screening for modulators of the cellular composition of gut epithelia via organoid models of intestinal stem cell differentiation. *Nature Biomedical Engineering 2022 6:4* **6**, 476–494 (2022).
20. Tian, Y. *et al.* A multi-omic single-cell landscape of cellular diversification in the developing human cerebral cortex. *Comput Struct Biotechnol J* **23**, 2173–2189 (2024).
22. Brancati, G., Treutlein, B. & Camp, J. G. Resolving Neurodevelopmental and Vision Disorders Using Organoid Single-Cell Multi-omics. *Neuron* **107**, 1000–1013 (2020).

1. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat Biotechnol* **38**, 708–714 (2020).

3. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols 2013 9:1* **9**, 171–181 (2014).

5. Wang, X., He, Y., Zhang, Q., Ren, X. & Zhang, Z. Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. *Genomics Proteomics Bioinformatics* **19**, 253–266 (2021).

7. Mulqueen, R. M. *et al.* Highly scalable generation of DNA methylation profiles in single cells. *Nat. Biotechnol.* **36**, 428–431 (2018).

9. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).

11. Nichols, R. V. *et al.* High-throughput robust single-cell DNA methylation profiling with sciMETv2. *Nature Communications 2022 13:1* **13**, 1–10 (2022).

14. Mehrmohamadi, M., Sepehri, M. H., Nazer, N. & Norouzi, M. R. A Comparative Overview of Epigenomic Profiling Methods. *Front Cell Dev Biol* **9**, 714687 (2021).

15. Ramos, R. *et al.* Parsing patterns: Emerging roles of tissue self-organization in health and disease. *Cell* **187**, 3165–3186 (2024).

16. Gjorevski, N. *et al.* Tissue geometry drives deterministic organoid patterning. *Science (1979)* **375**, (2022).

19. Williams, C. G., Lee, H. J., Asatsuma, T., Vento-Tormo, R. & Haque, A. An introduction to spatial transcriptomics for biomedical research. *Genome Med* **14**, 1–18 (2022).

21. De Jonghe, J. *et al.* scTrends: A living review of commercial single-cell and spatial 'omic technologies. *Cell Genomics* **4**, 22 (2024).

23. Gao, S. *et al.* Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat Cell Biol* **20**, 721–734 (2018).

1. Legnini, I. *et al.* Spatiotemporal, optogenetic control of gene expression in organoids. *Nature Methods 2023 20:10* **20**, 1544–1552 (2023).
3. Shin, W., Su, Z., Yi, S. S. & Kim, H. J. Single-cell transcriptomic mapping of intestinal epithelium that undergoes 3D morphogenesis and mechanodynamic stimulation in a gut-on-a-chip. *iScience* **25**, 105521 (2022).
6. Bein, A. *et al.* Microfluidic Organ-on-a-Chip Models of Human Intestine. *CMGH* vol. 5 659–668 Preprint at https://doi.org/10.1016/j.jcmgh.2017.12.010 (2018).
8. Beaurivage, C. *et al.* Development of a human primary gut-on-a-chip to model inflammatory processes. *Scientific Reports 2020 10:1* **10**, 1–16 (2020).
10. Shin, W., Su, Z., Yi, S. S. & Kim, H. J. Single-cell transcriptomic mapping of intestinal epithelium that undergoes 3D morphogenesis and mechanodynamic stimulation in a gut-on-a-chip. *iScience* **25**, 105521 (2022).
13. Xia, C., Colognori, D., Jiang, X. S., Xu, K. & Doudna, J. A. Single-molecule live-cell RNA imaging with CRISPR–Csm. *Nature Biotechnology 2025* 1–8 (2025) doi:10.1038/s41587-024-02540-5.
15. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e17 (2016).
17. Yao, D. *et al.* Scalable genetic screening for regulatory circuits using compressed Perturb-seq. *Nat Biotechnol* **42**, 1282–1295 (2024).
19. Zheng, X. *et al.* Organoid cell fate dynamics in space and time. *Sci Adv* **9**, (2023).
20. Wagner, D. E. & Klein, A. M. Lineage tracing meets single-cell omics: opportunities and challenges. *Nature Reviews Genetics 2020 21:7* **21**, 410–427 (2020).
22. He, Z. *et al.* Lineage recording in human cerebral organoids. *Nature Methods 2021 19:1* **19**, 90–99 (2021).

291 1. Lindenhofer, D. *et al.* Cerebral organoids display dynamic clonal growth and tunable tissue
2 replenishment. *Nature Cell Biology 2024 26:5* **26**, 710–718 (2024).
292 3. He, L. *et al.* Enhancing the precision of genetic lineage tracing using dual recombinases. *Nat Med*
4 **23**, 1488 (2017).
295 5. Capdevila, C. *et al.* Cellular origins and lineage relationships of the intestinal epithelium. *Am J*
6 *Physiol Gastrointest Liver Physiol* **321**, G413 (2021).
297 7. Barker, N. & Clevers, H. Lineage tracing in the intestinal epithelium. *Curr Protoc Stem Cell Biol*
8 **Chapter 5**, (2010).
299 9. Zhang, Y., Zeng, F., Han, X., Weng, J. & Gao, Y. Lineage tracing: technology tool for exploring the
10 development, regeneration, and disease of the digestive system. *Stem Cell Research & Therapy*
11 *2020 11:1* **11**, 1–16 (2020).
2912. Kong, W. *et al.* CellTagging: combinatorial indexing to simultaneously map lineage and identity
13 at single-cell resolution. *Nat Protoc* **15**, 750–772 (2020).
2917. Jindal, K. *et al.* Single-cell lineage capture across genomic modalities with CellTag-multi reveals
15 fate-specific gene regulatory changes. *Nat Biotechnol* **42**, 946–959 (2024).
2918. Ludwig, L. S. *et al.* Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-
17 Cell Genomics. *Cell* **176**, 1325-1339.e22 (2019).
2919. Hopkins, J. F. *et al.* Mitochondrial mutations drive prostate cancer aggression. *Nature*
19 *Communications 2017 8:1* **8**, 1–8 (2017).
3020. Weng, C. *et al.* Deciphering cell states and genealogies of human haematopoiesis. *Nature 2024*
21 *627:8003* **627**, 389–398 (2024).
3022. Naik, S. H. *et al.* Diverse and heritable lineage imprinting of early haematopoietic progenitors.
23 *Nature 2013 496:7444* **496**, 229–232 (2013).

301. Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D. & Klein, A. M. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science (1979)* **367**, (2020).

303. Spanjaard, B. *et al.* Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nature Biotechnology 2018 36:5* **36**, 469–473 (2018).

305. Raj, B. *et al.* Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nature Biotechnology 2018 36:5* **36**, 442–450 (2018).

307. Wu, S. H. S. *et al.* Red2Flpe-SCON: a versatile, multicolor strategy for generating mosaic conditional knockout mice. *Nature Communications 2024 15:1* **15**, 1–12 (2024).

309. Laurie, C. C. *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat Genet* **44**, 642–650 (2012).

311. Serra, E. G. *et al.* Somatic mosaicism and common genetic variation contribute to the risk of very-early-onset inflammatory bowel disease. *Nature Communications 2020 11:1* **11**, 1–15 (2020).

313. Kok, R. N. U. *et al.* OrganoidTracker: Efficient cell tracking using machine learning and manual error correction. *PLoS One* **15**, e0240802 (2020).

315. Holloway, E. M. *et al.* Mapping Development of the Human Intestinal Niche at Single-Cell Resolution. *Cell Stem Cell* **28**, 568-580.e4 (2021).

317. Production of R-Spondin 1-conditioned Medium using Cultrex ® HA-R-Spondin1-Fc 293T Cells PROTOCOL Production of R-Spondin 1-conditioned Medium.

319. Protocol: Generate Human Intestinal Organoid-Derived Monolayers | STEMCELL Technologies. https://www.stemcell.com/how-to-generate-human-intestinal-organoid-derived-monolayers-using-intesticult.html.

322. Hendriks, G. J. *et al.* NASC-seq monitors RNA synthesis in single cells. *Nature Communications 2019 10:1* **10**, 1–9 (2019).

1. Addgene: Protocol - How to Create a Bacterial Glycerol Stock. https://www.addgene.org/protocols/create-glycerol-stock/.

3. Addgene: Protocol - How to Inoculate a Bacterial Culture. https://www.addgene.org/protocols/inoculate-bacterial-culture/.

5. FACS Titration of Lentivirus Protocol. https://www.sigmaaldrich.com/GB/en/technical-documents/protocol/genomics/gene-expression-and-silencing/facs-titration-of-lentivirus?srsltid=AfmBOoo46u-TW1WDWOn9WQ68lAc2j1U27MfCizEy072NeMhNJ_6IYTVJ.

8. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

10. Download Cell Ranger - 10x Genomics. https://www.10xgenomics.com/support/software/cell-ranger/downloads.

12. Lun, A. T. L. *et al.* EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* **20**, (2019).

14. CITE-seq & Cell Hashing Protocol.

15. Mulè, M. P., Martins, A. J. & Tsang, J. S. Normalizing and denoising protein expression data from droplet-based single cell profiling. *Nature Communications 2022 13:1* **13**, 1–12 (2022).

17. Nie, X. *et al.* Clustering ensemble in scRNA-seq data analysis: Methods, applications and challenges. *Comput Biol Med* **159**, 106939 (2023).

19. Kobak, D. & Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nat Commun* **10**, 1–14 (2019).

21. C Corro, L. N. V. L. A brief history of organoids. *Am. J. Physiol. Cell Physiol.* **319**, C151–C165 (2020).

22. Xiang, T., Wang, J. & Li, H. Current applications of intestinal organoids: a review. *Stem Cell Research & Therapy 2024 15:1* **15**, 1–17 (2024).

1. Jensen, K. B. & Little, M. H. Organoids are not organs: Sources of variation and misinformation in organoid biology. *Stem Cell Reports* **18**, 1255 (2023).
3. Xu, Q. *et al.* An integrated transcriptomic cell atlas of human endoderm-derived organoids. *Nature Genetics 2025 57:5* **57**, 1201–1212 (2025).
5. Grouls, M. *et al.* Differential gene expression in iPSC-derived human intestinal epithelial cell layers following exposure to two concentrations of butyrate, propionate and acetate. *Scientific Reports 2022 12:1* **12**, 1–15 (2022).
8. Kittredge, A., Ward, N., Hopiavuori, A., Zhang, Y. & Yang, T. Expression and Purification of Mammalian Bestrophin Ion Channels. *J Vis Exp* **2018**, 57832 (2018).
10. Whitehead, R. H., Young, G. P. & Bhathal, P. S. Effects of short chain fatty acids on a new human colon carcinoma cell line (LIM1215). *Gut* **27**, 1457 (1986).
12. Sun, Y. *et al.* Assessing the impact of sequencing platforms and analytical pipelines on whole-exome sequencing. *Front Genet* **15**, 1334075 (2024).
14. Zhang, M. J., Ntranos, V. & Tse, D. Determining sequencing depth in a single-cell RNA-seq experiment. *Nat Commun* **11**, 1–11 (2020).
16. Ahlmann-Eltze, C. & Huber, W. Comparison of transformations for single-cell RNA-seq data. *Nature Methods 2023 20:5* **20**, 665–672 (2023).
18. Isakova, A., Neff, N. & Quake, S. R. Single cell profiling of total RNA using Smart-seq-total. *bioRxiv* 2020.06.02.131060 (2020) doi:10.1101/2020.06.02.131060.
20. Isakova, A., Neff, N. & Quake, S. R. Single-cell quantification of a broad RNA spectrum reveals unique noncoding patterns associated with cell types and states. *Proceedings of the National Academy of Sciences* **118**, e2113568118 (2021).
23. Li, J. H., Trivedi, V. & Diz-Muñoz, A. Understanding the interplay of membrane trafficking, cell surface mechanics, and stem cell differentiation. *Semin Cell Dev Biol* **133**, 123 (2023).

336 1. Zhang, H. *et al.* Extensive evaluation of ATAC-seq protocols for native or formaldehyde-fixed
2 nuclei. *BMC Genomics* **23**, 1–18 (2022).
337 3. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory
4 variation. *Nature 2015 523:7561* **523**, 486–490 (2015).
338 5. Dong, A. *et al.* Global chromatin accessibility profiling analysis reveals a chronic activation state
6 in aged muscle stem cells. *iScience* **25**, 104954 (2022).
339 7. Wang, X. *et al.* Using transcriptomics to investigate the effects of infant formulae on the growth
8 and development of small intestinal organoids and Caco2 cells. *J Funct Foods* **131**, 106951
9 (2025).
340 10. Li, X. J. & Li, S. H. HAP1 and intracellular trafficking. *Trends Pharmacol Sci* **26**, 1–3 (2005).
341 11. Haghverdi, L. & Ludwig, L. S. Single-cell multi-omics and lineage tracing to dissect cell fate
12 decision-making. *Stem Cell Reports* **18**, 13–25 (2023).
342 12. Jindal, K. *et al.* Single-cell lineage capture across genomic modalities with CellTag-multi reveals
14 fate-specific gene regulatory changes. *Nat Biotechnol* **42**, 946–959 (2024).
343 15. VanHorn, S. & Morris, S. A. Next-Generation Lineage Tracing and Fate Mapping to Interrogate
16 Development. *Dev Cell* **56**, 7–21 (2021).
344 17. Maru, Y., Orihashi, K. & Hippo, Y. Lentivirus-based stable gene delivery into intestinal organoids.
18 *Methods in Molecular Biology* **1422**, 13–21 (2016).
345 19. Fujii, M. *et al.* A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche
20 Factor Requirements during Tumorigenesis. *Cell Stem Cell* **18**, 827–838 (2016).
346 21. VanDussen, K. L. *et al.* Development of an enhanced human gastrointestinal epithelial culture
22 system to facilitate patient-based assays. *Gut* **64**, 911–920 (2015).
347 23. Schambach, A. & Baum, C. Clinical Application of Lentiviral Vectors – Concepts and Practice. *Curr*
24 *Gene Ther* **8**, 474–482 (2008).

1. Hauck, B. *et al.* Efficient AAV1-AAV2 hybrid vector for gene therapy of hemophilia. *Hum Gene Ther* **17**, 46–54 (2006).

3. de Lau, W. *et al.* Peyer's patch M cells derived from Lgr5(+) stem cells require SpiB and are induced by RankL in cultured 'miniguts'. *Mol Cell Biol* **32**, 3639–3647 (2012).

5. Wang, D. *et al.* Interferon-responsive intestinal BEST4/CA7+ cells are targets of bacterial diarrheal toxins. *Cell Stem Cell* **32**, 598-612.e5 (2025).

7. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nature Biotechnology 2021 39:7* **39**, 825–835 (2021).

10. Wasney, M. & Pott, S. Simultaneous Measurement of DNA Methylation and Nucleosome Occupancy in Single Cells Using scNOMe-Seq. *Methods in Molecular Biology* **2611**, 231–247 (2023).

13. Wang, Y. *et al.* Single-cell multiomics sequencing reveals the functional regulatory landscape of early embryos. *Nature Communications 2021 12:1* **12**, 1–14 (2021).