Unbiased molecular definition of epithelial barrier formation and defects driving inflammatory bowel disease

David Fawkner-Corbett

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

4th October 2021

Acknowledgements

This thesis represents a huge amount of work over a number of years working closely with a number of hugely talented individuals who have provide help, advice and support.

I would like to thank Professor Alison Simmons for supervising this work and providing a laboratory environment that was both friendly and scientifically stimulating. I appreciate the help from all members of the Simmons lab in discussing and giving me feedback for this work. I also thank Professor Paul Johnson for help, advice and support throughout the DPhil. Also the stimulating discussions and help I had from my thesis committee, Professor Todd and Professor Sauka-Spengler, as the project developed.

Within the lab I would like to thank Dr Agne Antanaviciute, the project as a whole closely merged the most recent wet lab techniques alongside computational methods and so it has been fantastic to work closely to achieve this. I am also hugely grateful to Dr Kaushal Parikh and Marta Jagielowicz who helped with developing the project and techniques as well as for ongoing training and advice.

I appreciate the help provided by collaborators to help generate this work including: Neil Ashley at the WIMM single cell facility, Simona Fourie and the ORNIID nurses, Berta Crespo and the team at the HDBR, Dr Fowler in paediatric pathology, along with all co-authors from the publications that have resulted from this work.

I also acknowledge and appreciate the help of funders for my DPhil – the Wellcome trust, the Royal College of Surgeons of England through the British Association of Paediatric Surgeons and the Oxford Health Services Research Committee.

Finally I would like to thanks my wife Emily for all the encouragement and support during this DPhil .

Attributions

Many aspects of this project involved working together with other researchers. Throughout this project I worked very closely with Dr A Antanaviciute for bioinformatic analysis/interpretation of the large multi-omic data methods that formed a core component of this work. The methods that generated the output for these is described and also attributed in the methods section (**Methods: 6.11**).

Dr K Parikh was also involved closely in both projects, studying the effect of *WFDC2* in IBD described in **Chapter 2**, and also for interpretation of the large transcriptomic datasets which constitute a large part of this work. This was approached collaboratively throughout the DPhil and work generated.

This thesis is a result of my own work with the exception of the data below, which was generated by my colleagues as indicated. Specific attributions of results presented in this thesis include:

- **Figure 1.1** is from a publication (Barker et al 2014[1]) and is re-used with requested permission from the publisher (Springer Nature).
- Dr K Parikh provided immunofluorescence images used in **Figure 2.6**.
- Marta Jagielowicz undertook the staining for *in situ* hybridisation stains (**Figure 3.2** and **4.5**), immunohistochemistry stains (**Figure 2.7,** also with Dr Parikh), infection assay (**Figure 2.7,** also with Dr Aulicino**)** and scanning electron microscopy (**Figure 2.7**), the latter alongside Dr Errin Johnson. Quantification of staining intensity was performed by N. Khalid Alham (Nuffield department of surgery) in **Figure 2.7**.
- I worked alongside Dr Anna Aulicino to undertake the Smartseq-2 protocol in **Figure 3.3** in which she had expertise.

- For mass spectrometry performed in **Figure 3.3** the running of samples was performed by Dr S. Davis and Dr R. Fisher (TDI, Oxford)
- Dr Chris Lagerholm undertook the live imaging that contributed to **Figure 3.4** with Dr K Parikh undertaking analysis.

Publication statement

Work that contributed to this DPhil has been published before thesis submission. Due to the nature of these experiments and the generation of large datasets and resources, work in the DPhil thesis has therefore been included in these publications. The inclusion of this work that has been used in this thesis meets the requirements of the "re-use of published material" guidance in that; attributions have been detailed, I was joint first author, I was involved in creation of all figures and I have permission from my co-authors and supervisors to re-use.

Details of publications directly linked with the results of this work include (* indicates equal authorship):

- Parikh, K.*, Antanaviciute, A.*, Fawkner-Corbett, D.* *et al*. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature 567, 49–55 (2019). https://doi.org/10.1038/s41586-019-0992-y. [2]
- Fawkner-Corbett, D.*, Antanaviciute, A.* Parikh, K.* *et al* Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell 184(3), 810-826 (2021) https://doi.org/10.1016/j.cell.2020.12.016.[3]

In keeping with the publishers reuse statements (Elsevier and Nature publishing group) the results and parts of figures presented in the thesis are acceptable (statement available from https://www.elsevier.com/about/policies/copyright/permissions and https://www.nature.com/nature-portfolio/reprints-and-permissions/permissions-requests)

Coronavirus-19 statement

A significant proportion of this DPhil was undertaken during the Coronavirus 19 (CoV-19) pandemic. Due to the generation of large datasets for this work a great deal of work was done on building on these and the findings. However some aspects of the planned DPhil project were altered or were not able to be included in final results. A full overview of alterations due to pandemic is included in **Appendix: A.1**

Abstract

The intestinal epithelial barrier is one of the body's largest mucosal surfaces. The cells involved reflect the numerous functions the epithelium must perform. The barrier requires time critical co-ordination with other intestinal cell compartments *in utero* for normal development. In maturity, dysregulation of the barrier or cross-talk can lead to disease such as inflammatory bowel disease (IBD).

Despite the importance of the intestinal epithelium in development and health, characterisation of the origins of dysregulation are lacking. Questions also remain about the full spectrum of epithelial cell diversity and its mutualistic relationship with other intestinal compartments.

This project characterises the development of the epithelial and mesenchymal cross talk *in utero*, in health and in IBD at high resolution using single cell RNA-sequencing (scRNA-seq). Studying intestinal epithelial diversity in health and disease, isolating epithelial cells from IBD and controls and mapping over 11,000 cells from health and UC inflammation, mapped cell diversity through the maturation of epithelium. This identified a hitherto unappreciated BEST4/OTOP2 cell that sensed pH and was dysregulated in inflammation and cancer.

Furthermore, a second project mapped human intestinal development from 8-22 post conceptual weeks. This charted 101 cell types across developmental time and through spatial transcriptomics (ST) could map these to tissue revealing origins of diverse cellular compartments along with fibroblast and intestinal stem cells across space and time.

These results provide a platform that maps the human intestinal epithelium and previously unappreciated resolution and from which the normal developmental cues can be established and allow identification of the drivers of dysregulation in IBD.

Contents

List of Figures

List of Abbreviations

ALP	Alkaline phosphatase
AMPs	Antimicrobial peptides
APC	Antigen presenting cell
AUC	Area under curve
BMP	Bone morphogenetic protein
CCK	Cholecystokinin
CD	Crohn's disease
DCs	Dendritic cells
DEGs	Differentially expressed genes
ECs	Endothelial cells
EDTA	Ethylenediaminetetraacetic acid
EECs	Enteroendocrine cells
EGF	Epidermal growth factor
ENS	Enteric nervous system
FACS	Fluorescence-activated cell sorting
FAE	Follicle associated epithelium
FDR	False discovery rate
FGF	Fibroblast growth factor
GALT	Gut associated lymphoid tissue
GEO	Gene expression omnibus
GIP	Glucose dependent insulinotropic peptide
GLP1/2	Glucagon-like peptide 1/2
GO(E)	Gene ontology (enrichment)

H&E	Haemotoxylin and eosin staining
HB-EGF	Heparin binding epidermal growth factor
HC	Healthy control
HDBR	Human Developmental Biology Resource[4]
HPO	Human Phenotype Ontology
HTO	Hashtag oligonucleotide
IBD	Inflammatory bowel disease
ICC	Interstitial cells of Cajal
IECs	Intestinal epithelial cells
IHC	Immunohistochemistry
IL	Interleukin
ILCs	Innate lymphoid cells
ISC	Intestinal stem cell
IF	Immunofluorescence
LTi	Lymphoid tissue inducer cell
MP	Muscularis propria
NK	Natural killer
NEC	Necrotising enterocolitis
PAS	Periodic acid-Schiff stain
PBMCs	Peripheral blood mononuclear cells
PCW	Post conceptual weeks
PPs	Peyer's Patches
PRRs	Pattern recognition receptors
PYY	Peptide YY
QC	Quality control

(q)PCR	(quantitative) polymerase chain reaction
R:L	Receptor : ligand
RNA-seq	Ribonucleic acid-sequencing
scRNA-seq	Single cell RNA-sequencing
SMCs	Smooth muscle cells
smISH	single molecule *in situ* hybridisation
SSII	Smart-seq 2[5]
ST	Spatial transcriptomics
TA	Transit amplifying cells
TCGA	Tissue cancer genome atlas[6]
TEM	Transmission electron microscopy
TNF	Tumour necrosis factor
***t*-SNE**	*t*-distributed stochastic neighbour embedding
UC	Ulcerative colitis
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier

Chapter 1: Introduction

1.1 The intestinal epithelial barrier

The intestinal epithelium will represent one of the largest barrier surfaces in the adult human body, in the region of $32m^2$ if flattened out[7]. It consists of a continuous monolayer of polarized columnar intestinal epithelial cells (IECs) throughout the gastrointestinal tract which, once mature, will organise into a crypt-villus axis to increase volume for function. The cells are joined together through tight junctions to prevent permeability and maximise absorption. There is rapid turnover of cells constituting the barrier, being replaced every 3 – 5 days[8]. Within this time window all cells required for epithelial function will form from a dividing intestinal stem cell (ISC), demarcated by the G-protein coupled receptor LGR5 at the base of each of the deep invaginations called crypts[9].

It is through this interface the intestine must perform a variety of diverse roles to maintain health; efficient digestion, absorption, symbiosis with commensal microflora to name but a few[1]. Diverse types of IECs are required to fulfil these functions, but the relative proportion composition and interaction of these cell types vary along the alimentary canal along with key differences in morphology; such as the small intestine forming finger-like villi and the colon forming crypts (**Figure 1.1**)

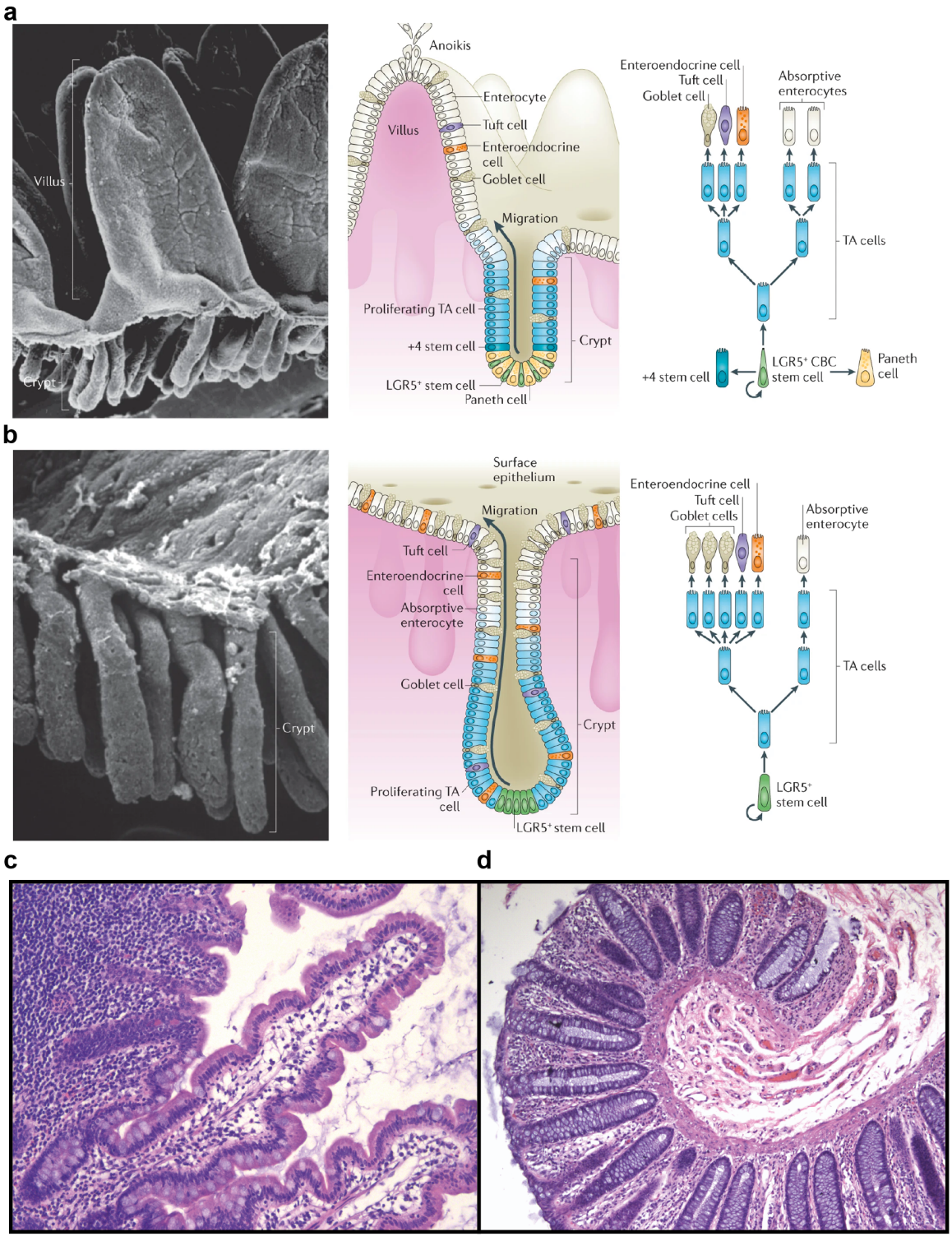

Figure 1.1: Structural organisation of the intestine

Representative scanning electron microscopy images of murine small intestine (**a**) and colon (**b**) demonstrating regional topographical differences alongside cell type composition differences through the crypt (centre) and schematic of ISC differentiation stages (Figures **a** and **b** from Barker 2014[1] and re-used with permission from Springer Nature). Comparison is also made to human histological (H&E) staining of small intestine, with villi (**c**), and colon (**d**) with crypts.

As well as the physical defence provided by this continual layer it performs specialised protective functions including the production of mucus. The secretion of these highly glycosylated mucins (namely MUC2) provide additional support and capture commensal bacteria[10]. This can not only shield the host from digestive enzymes and microbial invasion, but also acts as a site of maintenance for important commensal bacteria that play a homeostatic role in metabolic and immune tolerance[11]. By sensing microbial components in the mucous layer, most well described through pattern recognition receptors (PRRs), epithelial cells can co-ordinate responses to pathological bacteria whilst establishing tolerance to symbiotic commensals[12].

In fact, the intestinal microbiome will be so vast in adulthood that it will constitute over 10 times the number of human cells in the body with colonisation beginning immediately after birth and the microbial diversity being unique to an individual and dependent on both genetics and environmental factors[13]. Similar to IECs of the epithelial barrier, the mucus compartment will also vary in structure across the alimentary canal, being two layers thick in the colon and only one in the small intestine – again reflecting a difference in function and bacterial response[14].

As well as interacting and generating intra-luminal components the epithelium will interact closely with deeper intestinal structures. IECs can co-ordinate responses to infection or inflammation by signalling through local immunoregulatory circuits or distant effector circuits through immune, vasculature or neural connections[11]. Reciprocally, underlying cellular compartments can provide vital support and spatial context through homeostatic circuits to help the maintenance and maturation of the barrier.

Morphogens are secreted substances which can have distant effects from their source and drive maturation of cells in a gradient dependent manner[15]. These are vital in the turnover of the epithelium and patterning of the crypt-villus shape through a variety of

pathways that signal through these morphogens. Key intestinal mechanisms include the bone morphogenetic protein (BMP) and Wnt pathways[16].

Wnt pathway signalling consists of a family of genes (in mammals 19 Wnt ligands and 10 Frizzled (Fzd) receptors) that are important in many embryological and homeostatic processes, which are conserved throughout various species[17]. Wnts travel a short distance and initiate downstream signalling in effector cells, which can be via canonical or non-canonical pathways. Wnt has been shown to be key to maintaining ISCs homeostasis and regeneration within the base of the crypt, promoting ISC turn-over, aberration of which can lead to oncological processes through downstream effectors such *c-Myc* - a known oncogene in colorectal cancer[18]. Conversely the effect of Wnt signalling can be effected by external sources, for example inactivation of *APC* causing nuclear re-localisation of the Wnt signal transducer β-catenin also resulting in colorectal cancer, and demonstrating the inter-connected nature of many genes and pathways in intestinal crypt homeostasis[19].

In contrast, BMP balances the Wnt pathway in health with expression being seen higher up the crypt and it restricts the stemness of ISCs by preventing the hyperproliferation of stem cells so the two pathways normally act in balance[20]. Other key IEC regulators can exhibit their effects through BMP signalling, modulators Gremlin 1 and 2 (*GREM1/2*) and Chordin-like-2 (*CHRDL2*) can also do this from the surrounding mesenchyme with dysregulation leading to enteropathy[21]. Another important, although rarely seen, BMP modulator is Noggin which if overexpressed can lead to ectopic crypt formation and can also be seen in the surrounding non-IEC cells further highlighting the complexity of crypt signalling[22]. Epidermal growth factors (EGF) also stimulate proliferation[23] and it has recently been shown in combination that these components can maintain ISCs in culture for *in vitro* modelling as organoids[24].

Other signalling mechanisms, such as NOTCH occur in a contact-dependent manner, deep in the crypt with ligands being presented directly to inhibit direct maturation signals to a laterally placed neighbour[25]. Although components of these multiple pathways are on IECs and activated at key maturation decisions, it is apparent the underlying mesenchyme is also a key source of ligands which are necessary for ISC maintenance, highlighting the importance of cross-compartmental cross-talk in normal intestinal health[26,27].

Taken together it is clear the epithelial barrier has a broad range of roles that require both locational and morphological context. It has been demonstrated that a number of key pathways and interactions, generated from both the lumen and deeper tissue, are required to maintain a delicate homeostatic balance in health. However, the ways in which these pathways become dysregulated in disease, and the exact mechanisms that direct how they are formed from a flat gut tube in human development remains elusive.

1.2 Diversity of cellular architecture

In adult health all IECs will be formed from ISCs through a proliferation zone at the base of each crypt, a stereotypical pattern which is repeated throughout the gastrointestinal tract despite regional specific differences in morphology[1]. The classification of many of these cell types have historically been done by definition of clear histological features on routine stains. These include haematoxylin & eosin (H&E) which are able to identify the shape and structures that could identify location (Brunner's glands in duodenum, villi in small intestine, crypts in colon) and cell shape such as columnar absorptive enterocytes[28]. Specific immunohistochemical stains for the "four core differentiated cell types" of the intestine[29] could then be performed to identify; goblet cells by periodic acid Schiff stain (PAS) demonstrating mucopolysacharides[30]; anti-synaptophysin to identify

enteroendocrine cells (EECs) through synaptic-like micro-vesicles[31,32]; the antibacterial protein lysozyme to stain Paneth cells[33]; the brush border enzyme alkaline phosphatase (ALP) to identify differentiated enterocytes[34].

Though these techniques demonstrated differentiated cell types historically, it has since been shown that there is cellular diversity within these core cell types. Further classification has identified specific roles in IECs or cell sub-types. EECs have shown to have a diverse repertoire of responses to luminal stimuli and expression of various peptide hormones (cholecystokinin (CCK), glucagon-like peptide 1 and 2(GLP1/2), glucose dependent insulinotropic peptide (GIP), peptide YY (PYY), gastrin, secretin, somatostatin motilin, leptin, nesfatin-1 and ghrelin) and biologically active amines (histamine and serotonin)[35]. This led to descriptions of greater than 11 EEC sub-type populations reported, with specific functions and frequencies throughout the alimentary canal, with some overlap between classifications and secretome profiles[35].

Similarly, goblet cells have been shown to have differing roles in specific contexts fulfilling a ISC protective role in the colon where Paneth cells are lacking[36] or having dynamic interactions with microbiota and the immune system and presenting antigen in a tolerogenic or non-tolerogenic manner depending on context[37,38]. Together these findings raise the question of what truly defines a cell type or transitional state within the intestine

Alongside functional roles much information has been gleaned about the transcriptional drivers of IEC diversity using targeted gene tagging approaches. This has revealed that all types of IECs will form from an ISC which has self-renewing capacity at the base of the crypt and expresses the G-protein coupled receptor LGR5[8]. Conditional mapping of transiently expressed transcriptional regulators have then identified subsequent control of mature cell types, such as *Neurod1* being shown to control gene expression in maturing EECs in a targeted manner[39]. Some transcriptional regulators are activated

through interaction and lateral inhibition of other cells within the crypt, such as NOTCH signalling acting through direct membrane contact (ligands DLL1 or DLL4) in undifferentiated IECs to express the transcription factor *HES1* which will repress *ATOH1* blocking secretory cell type differentiation and generating an absorptive phenotype as the cell matures out from the crypt[40]. There will then be subsequent further specialisation in *ATOH1* secretory progenitors to define the final phenotype of goblet, EEC or Paneth cell, with *SPDEF* being one factor that defines terminal goblet cell differentiation[41].

The current overview of cell types that are described in literature to constitute IECs, including their morphology function and transcriptional regulators, are shown in **Table 1.1**. Although the majority (>80%) of IECs will be of the enterocyte class, it is important to highlight that great diversity of sub-types can be seen even in the rare cell types. This can be demonstrated in the functional roles carried out by differing neuropeptide secretion that can differ throughout the intestine in EECs[35].

Some roles can be specific to the interaction with other anatomical features – M-cell markers have been reported individually in other cell sub-sets but the cell type being defined by its proximity to Peyer's patches (PPs) or immune follicles[42]. Some types may also be entirely absent in some locations of the gastrointestinal tract. Paneth cells are seen in the small intestine where they have a key role in ISC maintenance and bacterial protection through secretion of lysozyme, but they are not seen in the distal colon[29]. In contrast goblet cells are seen throughout although the frequency differs and they constitute a larger proportion of IECs in the colon. Despite this difference their role of maintaining a mucus barrier and forming the tight junctions needed to maintain IEC permeability is still necessary throughout the whole gastrointestinal tract[43]. Dysregulation of either of these functions in goblet cells is a hallmark of intestinal inflammation[44]. These locational and cell type differences in IECs have important implications for the application of

differentiation and cell type interaction models used to understanding the maintenance of the intestine .

It should also be appreciated that the definition of specific cell types is on the background of an ongoing maturation through the rapidly regenerating stereotypical pattern of the crypt-villus axis. Thus +4 cells can exhibit primitive features of a number of cell types[45], and indeed there is ongoing debate with regards their regenerative potential. They can exhibit plasticity and can replenish the ISC pool in murine models of radiation damage[46] and reports have shown they can be de-differentiated from both secretory[47] or enterocyte lineages[48]. Indeed the concept of plasticity adds a further dimension to the proposed unilateral progression of stem cell maturation, as it has been shown that ablation of ISCs can be accounted for by de-differentiation of more mature cell types[49], and raises questions of whether multiple ISC sub-types exist (such as *BMI*+ and *LGR5*+ distint cell types[50]) and if cells can change their protein, RNA and even chromatin expression patterns in a context dependent manner[51].

Thus, it can be difficult to ascertain specific transcriptional regulation and function across cell types, especially in the transit amplifying (TA) cell zone where markers are lowly expressed. There are also limitation on modelling differentiation using murine modelling – CreER and tamoxifen can cause toxicity and off target effects - and there can be incongruity between reported and protein expression or markers so some models of injury can be inconsistent[52]. Thus, it is vital to apply the principles developed in these models to human intestinal tissue, ideally at the molecular level.

Overall, IECs represent a varied and dynamic selection of cells which are performing a range of functions with signals from surrounding IECs, along with those from further afield. Much of the definition and annotation historically has been done by their roles and structure, by using targeted techniques or inferring fate decisions with modelling.

However, there is still not consensus on the discrete types and subtypes of all epithelial cells in the gut, how they change in differentiation and response to spatial or pathological stimuli or how the crypt axis is originally established.

Cell Type	Morphology	Function	Markers	Transcription factor(s)	Current literature themes
ISCs/Crypt base columnar cells	Slender cells seen at crypt base.	Exhibit self-renewal and can differentiate to IECs of all other lineages[53].	LGR5[8]	N/A – primary stem cell	Progeny forming all lineages and producing life-long clones was demonstrated using longitudinal *Lgr5* tracing in *Lgr5*$^{EGFP\text{-}IRES\text{-}CreERT}$ mice crossed with *R26R-lacZ* reporter.[8]
+4 cells	Small, undifferentiated cycling cells placed between ISC and Paneth cell (in small intestine) or bottom 4 positions of crypt (colon) – the TA zone.	Highly resistant to radiation and can replenish ISC pool when required[54].	*Bmi1*[55], *Tert*[56], *Hopx*[57], *Lrig1*[58]	c-MYC	Isolation by laser capture microdissection demonstrated components of *c-myc* signalling[59]
Paneth cell	Localised to the small intestine, above +4 cells. Pyramidal in shape and have acidophilic granules[60].	Secretes antimicrobial peptides (AMPs) and defensins to protect and support the ISC crypt niche[61].	LYZ, *WNT3*[16], *CD24*[62], *DEFA2*	*ATOH1**, *SOX9*	Appear to move downward or be uncoupled from other cells in the crypt-villus axis. Supply vital morphogens in small intestine organoid culture[62].
Goblet cell	Polarized with nucleus at base, remainder of the cell is filled with secretory granules.	Coating the intestine with protective mucus – single layer in small intestine and dual layered in colon.	MUC2, *TFF3*	*SPDEF*[41,63], *ATOH1**, *KLF4*[64]	Genetic or pharmacological ablation of Notch signalling will turn proliferative crypt cells into goblet cells[65].
Enteroendocrine cell	Typically tall and columnar in appearance, can span the epithelial barrier, large variation reported between sub-types of EEC.	Specialised for response to local environment by secreting peptide hormones to control appetite, enzyme secretion and appetite.	CHGA, CLDN4[66]	*ATOH1**, *NEUROG3*, *NEUROD1*	Numerous types described, unclear if each represents a distinct cell type or is sub-sets may express a "cocktail" or hormones. EECs may have distant functions interacting with immune or enteric nervous system (ENS)[35]

Enterocyte	Tall cells, columnar with oval nucleus placed basally. Apical surface composed of microvilli and glycocalyx. Junctional complexes (tight junctions and adherens junctions) fix to other enterocytes laterally[67].	Nutrient uptake and absorption.	ALP, FABP1, VIL1	*ATOH1* negative, *CDX2, TCF4*[68], *HES1*	Constitute majority (80%) or IECs. Fate determination in crypt based on notch dependent *Atoh1*[69] repression. Maturation is through loss of Wnt and increase of BMP signalling that have expression gradients when leaving the crypt[21].
M-Cell	Appearance similar to enterocytes but found overlying FAE of PP, isolated lymphoid follicles or in the appendix. "Microfolds" that house lymphocytes and can be seen on apical surface with electron microscopy[42].	Absorptive cells that overly PPs. Sample antigens in the intestinal lumen and transport to immune cells[70].	GP2, UEA-1**, Annexin V[71]	*SPIB*	Differentiation is determined by RANKL expression in underlying stroma[72].
Tuft cell	Long and thin with a visible "tuft" of microvilli projecting from apical surface.	Immune regulation and defence in helminth infection	DCLK1, TRPM5, PTGS1	*Pou2f3*[73], originate from DLL^+ secretory progenitors but thought to be distinct.	Very rare (<0.4%)[74]. Differentiation can be activated to IL-4 and IL-13 produced by ILC-2 in response to helminths[75].

Table 1.1: Overview of current understanding in diversity of intestinal epithelial cells(*$ATOH1$($Math1$) shown to demarcate shared early progenitor of multiple secretory lineages, regulated by *Hes1*[76], **UEA-1 also described as M-Cell marker only when present in follicle associated epithelium (FAE))

1.3 Formation of the intestinal epithelium

Similar to the complex interactions of IECs to maintain homeostasis, the intestine itself will undergo a number of time critical interactions *in utero* in order to create the complex morphology that is seen in maturity. It is challenging to study this process dynamically in humans, due to the scarcity of tissue at early time points, and so much is inferred from animal models.

The process of gastrulation will occur at 16-19 days post conception (Carnegie stage 7) in the human[77] and this process will give rise the three germ layers – ectoderm, mesoderm and endoderm. The complexity of the intestine is further demonstrated in organogenesis, where it will receive aspects of all these layers during development: the endoderm giving rise to the epithelium; the mesoderm the mesenchyme and smooth muscle; and the ectoderm the enteric nervous system[78]. These components will allow the primitive gut tube, which is a few cell layers thick initially, to eventually form the full intestine.

Following gastrulation the posterior endoderm undergoes extensive folding to generate the embryonic gut tube that will eventually give rise to the small and large intestine; and between eight and twelve post conceptual weeks (PCW) the tube will proceed from a flat tube with pseudo-stratified epithelium to the expanding and lengthening folds that herald the formation of crypts and villi[79,80]. The establishment of these morphologies will proceed in a craniocaudal manner, with animal modelling demonstrating patterning is established by drivers such as *Cdx2* and *Sox2* with cascades of inter-related morphogen pathways defining gradients of differentiation including Wnt, hedgehog, BMP, fibroblast growth factor (FGF)[81,82]. By 12PCW onwards the intestinal structure will become akin to adult

like mature morphology, and differentiated IECs will be seen, suggesting the establishment of self-renewing ISC circuits[79].

To date, much of the knowledge of how the intestine undergoes developmental morphogenesis is derived from studies of chick and mouse, in order to obtain time-critical samples for mapping[83–85]. Human tissue has been used to derive information from historical fixed samples with descriptions of morphological features in 2D[86] or more recently 3D[77], targeting known markers of development.

Large unbiased studies of human IEC development is still lacking however, and will be important when applying animal models as some analyses have highlighted species specific differences in both murine and chick models of intestinal development[87]. For example, mouse models have highlighted PDGFR-A expressing mesenchymal cells forming "hillocks" to direct epithelial differentiation, whereas buckling forces applies from the external muscularis are key in chick[84,85].

It is also apparent that although histological features may appear mature late on in human development, they may not represent functional maturity. In mice it has been shown that the epithelium of the colon will delay expression of certain genes until the suckling-weaning transition at postnatal day 17[64]. This can be seen in the case of extremely premature infants, the epithelial barrier and its morphology may have mature appearance but in some cases exposure to feed can result in potentially fatal inflammatory processes seen in necrotising enterocolitis (NEC) which is much more common in earlier gestations[88]. The exact causation remains unclear but is likely to involve microbial triggers in the context of a naïve intestinal barrier and immune system.

1.4 Intestinal compartments and their cross-talk

The intestinal epithelial barrier and it's locational and maturation axis exist in steady state in health to complete the IECs diverse role. The epithelial compartment cannot function or maintain itself in isolation however. IECs also communicate more deeply to other cell type, such as goblet cells sensing microbiota and generating antigen passages in response to ENS signals to arm immune responses[37]. Bidirectional signalling and support are vital from other compartments in order to generate and maintain the barrier. These include cells of the fibroblast, endothelial, pericyte, neural, immune, muscle and mesothelial/serosal lineages.

Recently our understanding of these interactions has increased, and it has been shown that support from immune cell signals is vital to form the normal epithelium[89]. In murine models, loss of this support leads to worse intestinal damage in a DSS colitis model, showing these circuits to potentially be lost in IBD[90].

In adulthood the maintenance of epithelial proliferation through ISC division is supported by ligands secreted from fibroblast cells, with peri-cryptal specialised fibroblasts recently termed "telocytes" being shown in murine models to express *Foxl1* and produce Wnt ligands which are required for normal crypt morphogenesis [27]. It has also been shown that this mesenchymal supportive role may be even more essential in the colon, where morphogenic fibroblast cells marked by *Gli1* were vital in the colon and could compensate for Paneth cell loss in the small intestine, as well as support murine intestinal organoids[90]. The exact role of these fibroblast ISC supportive cells in human remains unclear.

Cross talk with cells from the more physically distant compartment of the ENS, including both glial and neuronal cells, will also provide key supportive signals including the regulation of tight junctions through glia-derived s-nitrosogluthanine[91]. Loss of

neuroprotective and neuro-modulatory support from the ENS is hypothesised to be a driver of intestinal inflammation[92].

Overall it is apparent that IEC maintenance cannot be considered in isolation. These cross-talk events can occur in a time critical manner, errors in which during development or adulthood can lead to disease. These interactions may also occur in a different manner during development due to the vast differences in distance between cells, and the absence or presence of specific cell types at differing time points (**Figure 1.2**).

Although size and cellular content may differ, the anatomical landmarks of the intestine remain similar throughout. The IECs will form a monolayer with overlying mucus, underneath which is a basement membrane consisting of a sheet of connective tissue. Beneath this is a mixed cellular compartment known as the lamina propria which will contain diverse cell types and structures– fibroblasts, vessels and endothelial cells, nervous system cells[93] (**Figure 1.2**). At the base of this is the muscular layers of the muscularis mucosae and then the thicker muscularis propria (MP) with more organised neural plexuses in between. All of this is then enclosed in layers of circular and longitudinal smooth muscle that controls peristalsis through pacemaker cells with the final layer being a serosa that is in continuity with the peritoneum of the abdominal cavity. This section will seek to review the existing classification of non-epithelial cells of the intestine which will interact in these anatomical spaces**.**

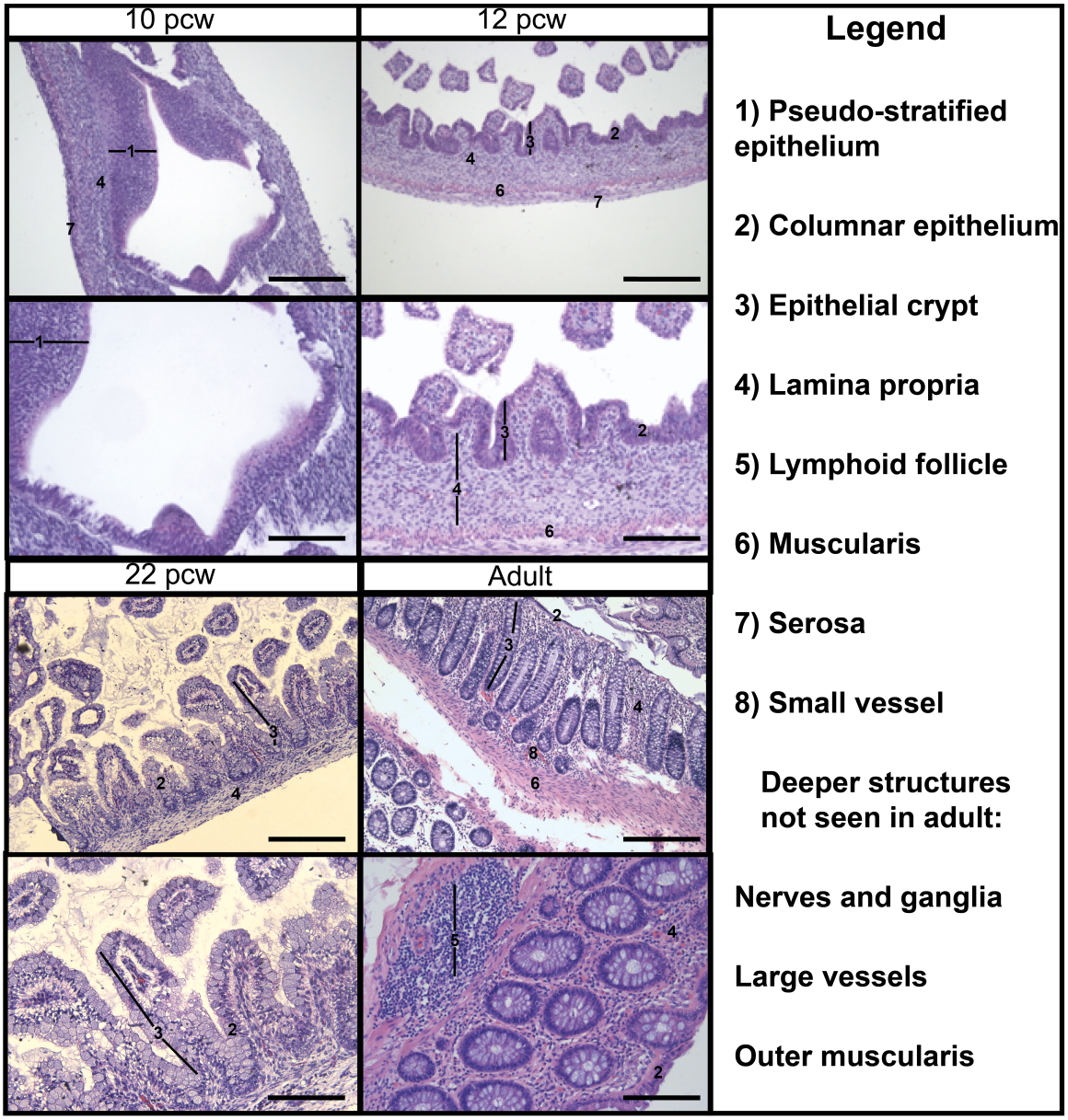

Figure 1.2: Key structures of the intestine over time

H&E stains of colonic intestinal samples over developmental time and adulthood to highlight location of key anatomical areas and structures that correspond to cell-type locations and demonstrating the large differences in relative distance between structures (Legend identifying histological or anatomical feature when seen in section, note some features are not demonstrated due to depth in adult tissue or small size in developmental tissue, differing images used in adult to demonstrate lymphoid follicles, scale bar represents 180 μm in upper images and 90 μm in lower images)

1.4.1 Fibroblast, myofibroblast and pericyte cells

Mesenchymal cells are described as non-epithelial, non-haematopoietic, non-endothelial structural cells and are present in almost all organs of the body where they underlie the epithelium and can play diverse roles in undertaking and co-ordinating organ functions and innate responses[94]. Their definition by the absence of markers reflects they were originally identified as purely structural, the origin of nomenclature being *mesos* (intermediate) and *enchyma* (infusion), thus "filling the gap" between other functional cellular structures[95].

Within the intestine the three major sub-sets of mesenchymal cells seen within the lamina propria are fibroblasts, myofibroblasts and pericytes. These form a dense cellular network beneath the epithelium. Each are reported to exhibit relatively higher expression for distinct marker proteins; α-SMA for myofibroblasts, Thy1 (CD90) for fibroblasts, desmin for pericytes[96] (**Table 1.2**). However, there is a large amount of overlap in expression patterns and the relative expression of markers can change in a depth related manner along the crypt-villus axis leading to contention over exact classifications and origins[95]. This along with these cell types exhibiting plasticity in a spatial, physiological or pathological context results in some literature reporting that confident identification can only be performed histologically with the use of transmission electron microscopy (TEM) to allow ultrastructure appreciation[97].

Cell Type	Reported markers[96]				Alternative identifiers	Morphology
	α-SMA	THY1 (CD90)	Desmin	Vimentin		
Fibroblast	-	+	-	+	FSP1, PDGFRA[84]	Collagen granules, spindle cell type or enlarged with expanded endoplasmic reticulum
Myofibroblast	+	+	-	+	AOC3, NG2, FAP[98]	Myofilaments, fibronexus structure that connects extracellular matrix to fibronectin fibrils[97]
Pericyte	+	+	+/-	+	NG2, PDGFR-B,RGS5[99]	Perivascular, closely related to ECs, peg-socket contacts, prominent nuclei

Table 1.2: Overview of peri-epithelial mesenchymal cells of the intestine

Sub-epithelial fibroblasts are described to have two morphologies including spindle shaped cells with slender cytoplasmic processes or with a large cytoplasm and a well-developed rough endoplasmic reticulum and golgi apparatus, intermediate morphologies have also described[93]. They can also be identified by their lack of smooth muscle myofilaments[100]. They can be seen both adjacent to the epithelium, where a sub-type of PDGFRA expressing fibroblasts was shown to transduce epithelial signals in the murine colon[101], or deeper where they can play a role in directing immune activation signals[102]. There is contention over the deeper fibroblast cells and their distinct mesenchymal classification, it has been reported that superficial fibroblasts can undergo transition to epithelial cell types in inflammation when ISCs need replacement[103] suggesting

differential roles depending on depth. Thus fibroblast roles in the lamina propria are diverse and may represent a spectrum of states, but remain poorly defined.

Myofibroblasts are also seen in a sub-epithelial position but identification is aided by intermediate staining of α-SMA, albeit much lower than muscularis cells (**Table 1.2**). They have bundles of myofilament reflecting involvement in contractile processes, highlighting them as important in fibrosis when excess conversion to stress fibres occurs[95]. The presence of definitive myofibroblasts in the steady state intestine is contentious, some reports state only fibroblast or muscularis cells are seen[100]. This emphasises these cells to exhibit dynamic expression of marker proteins, and that may implicate a close relation to fibroblast cells in differentiation. They are known to be important in directing responses in diseases, serving as antigen presenting cells (APCs) in graft vs host disease[104] and regulating cytokine and growth factor expression to regulate epithelial proliferation[94].

Pericytes, also termed mural cells or vascular smooth muscle cells due to their close inter-relationship with vessels and endothelial cells (ECs), are described as α-SMA positive myofibroblast-like cells. They wrap closely around capillaries as a single cell layer, through which they communicate with ECs via paracrine signalling[105]. With ECs they form a thin layer and although previously thought to be structural, have since been shown to signal through heparin-binding epidermal growth factor (HB-EGF), NOTCH or Ephrin signalling to co-ordinate vessel contractility and permeability based responses in vessels and further afield[106]. They can direct vascular remodelling and angiogenesis through reciprocal expression of angiopoietin-Tie2 pathway with ECs[107]. Their close relationship to other mesenchymal cells is demonstrated in other organs where pericytes can transform from myofibroblasts in disease settings[108].

Thus mesenchymal cells represent a number of distinct yet closely related cellular populations, to the extent that transition between each is reported and historic

classification by selected markers is contentious[100]. Their exact origins also remain elusive, in murine models it has been reported the serosal mesothelium (Wt1+) is the shared progenitor for myofibroblasts and pericytes[109]. This is yet to be confirmed in the intestine of humans, or whether fibroblasts are of the same lineage and if true it is unclear at what point they become distinct[105].

1.4.2 Endothelial cells

ECs, identified by markers of adhesion proteins including PECAM1 (CD31)[110], are key structural components of the vasculature, forming the vast networks of blood supply required for normal intestinal function - one third of cardiac output being directed at the alimentary canal[111]. They form tight seals through elaborate tight junctions and adherens junctions to transport blood and complete the intestines vital role for the intake of nutrients and excretion of electrolytes, albeit this seal is selectively porous to achieve important intestinal functions[112].

As well as the role of vascular endothelial cells in blood supply, the lymphatic system forms a one way drainage system for interstitial fluid and immune cells in all organs which will transport to larger lymph vessels then back into circulation through the thoracic duct, and lymphatic vessels are highly prevalent throughout the intestine. Unlike capillaries intestinal lymphatic vessels are more porous, they have valves and are surrounded by SMCs to help pump the contents in low pressure environment. They consist of lacteals which project up into the villi or crypt tops, collecting vessels in the lamina propria which pool contents from the epithelium and pass through mesenteric lymph nodes[113]. Through this structure the lymphatic system can plan key roles in many intestinal functions such as being the only source of fat-soluble vitamin and cholesterol

absorption[114]. They are also a key source of priming immune responses through cells travelling to lymph nodes (e.g. DCs) or via bacterial endotoxins (e.g. LPS) trafficking on chylomicrons[115].

Lymphatic ECs, exhibiting specific expression of hyaluronon (LYVE1)[116], have been shown to be actively involved in intestinal defence by selectively controlling antigens that are presented to the intestinal immune system and more systemically through circulation[117]. The vessels of the gastrointestinal tract can control systemic blood pressure through the renin-angiotensin system attuning the uptake and, in some conditions, pathological loss of fluids[118]. ECs direct angiogenesis and regeneration signals in other organs including the eye and liver[119,120]. Targeted approaches have shown ECs can direct intestinal inflammation by secretion of cytokines including tumour necrosis factor (TNF)-α, interleukin (IL)-6 and IL1- β and signal more distantly to the epithelium through the aryl hydrocarbon receptor pathway[111,120]. Together this identifies ECs to have dynamic roles in tissue dependent and disease-based contexts but whether heterogenous populations exist in the human intestine is still undefined.

1.4.3 Enteric nervous system

The ENS entails the cells from the nerves of the intestinal tract and their covering glia, constituting one of the largest components of the peripheral nervous system with as many neurons as the spinal cord[121]. Initial understanding of the ENS was its role in intestinal luminal contractility, with local signalling circuits generating continual peristalsis of food. This was evident by the demonstration that contractility would continue even when the intestine was severed from the nervous system[122]. Much original study of the ENS was therefore as a mechanism of propulsion and identified that neurons interact

with the smooth muscle and are arranged between the muscularis layers, within myenteric plexuses that are densely innervated by excitory or inhibitory neurons and connect by internodal strands[123]. The ENS colonises the bowel in a caudo-cranial manner and absence of these plexus in genetic murine models caused lethal obstruction after birth due to increased contraction of the smooth muscle, and was shown to link with defects in *RET* or *PHOX2B* genes in humans which cause the ENS disorder Hirschsprung's disease[124].

In recent years it has been demonstrated the intestine also functions as a sensory organ through the ENS[123]. Neurons are closely adjacent to the lumen of the intestine, with many abutting the epithelium and interacting with EECs to signal centrally through mediators such as serotonin[125]. It has been also been shown through rabies tracing studies in mice that these afferent signals can travel rapidly, through one synapse via glutamate release[126], and so the gut-brain axis is an important area of ongoing research into how luminal signals or microbiome may alter the central nervous system.

The neurons and ganglia of the ENS do not exist in isolation, and are enclosed within glial cells. Glial cells execute an important supportive role encasing the long cellular bodies of the afferent and efferent ENS neurons, providing nutrients and insulating signalling pathways. It is beginning to be understood they have bi-directional communication with underlying neurons, and that they can direct inflammatory cascades in the setting of colitis or infection and conversely can generate regenerative signals within the ENS[127,128]. The mediation of this balance has highlighted glial as representing heterogeneous cell sub-types with distinct roles, which have been reported through histochemical analysis of their structure and transmitters along with bulk studies of their active transcriptional profile[129,130]. How many types of glial exist within the intestine and their differing functional roles is still not fully appreciated.

Thus, the neurons and glial forming the ENS demonstrate not only the importance of cellular cross talk within the intestine, with rapid signals travelling from epithelium centrally, but also how dysregulation of the intestine can have far reaching effects throughout the body.

1.4.4 Immune cells of the intestine

To maintain and protect the intestinal barrier's large surface area, it houses 70% of the body's immune cells in adulthood[1]. This reflects the large amount of antigen exposure occurring at this interface, the intestinal lumen being populated by approximately 40 trillion microbes in adulthood[131,132]. Exposure occurs immediately after birth and in the first year of life this environmental colonization moulds an individual's immunity in a manner that is key for future health[133] with far reaching effects across cell compartments, such the enteric nervous system[134,135]. The immune system populates the proximal intestine during fetal development in preparation[136]. Then in the first year intestinal homing CD8+ immune cells in the peripheral blood are the most dynamically changing of any immune population, and CD4+ T-cells can become activated in disease causing profound systemic pathological effects[137,138].

Compartmental immune cell cross-talk is vital for normal epithelial maturation exemplified by goblet cells co-ordinating antigen presentation by dendritic cells (DCs) in the lamina propria to drive tolerogenic signalling[37,139]. Once mature the intestine will contain a diverse repertoire of immune cells which will be anatomically localised into sites adjacent to the epithelium to undergo antigen priming and differentiation – namely the gut associated lymphoid tissue (GALT) and the mesenteric lymph nodes[140]. GALT will differ in location, forming PPs in the small intestine and crypto-patches in the colon, although

both will be supported by numerous smaller isolated lymphoid follicles that are distributed throughout the GI tract[141].

These structures and the cells within must delicately balance the response to pathogens alongside tolerance of the diverse microbiome, especially within the colon which has the greatest microbial burden and forms a symbiotic relationship with many of these bacteria. Key players in this mechanism have been identified, and often studied individually by isolation by surface markers. This includes tissue resident immune cells, which can be T-cells of numerous types (CD4+, CD8+, natural killer (NK), or innate lymphoid cells (ILCs)) which exhibit increased expression of residency markers (CD69, CD103) when migrating to barrier sites[142]. The differential regulation of key immune cell markers can make identification of discrete immune cell functions challenging, and previous work has often focused on bulk approaches to study these[143].

Antigen presented to the APCs of the lamina propria can then be subjected to tolerance signals from numerous regulatory T-cells through release of cytokines such as IL-10 cells in a context dependent manner[144]. Imbalance in any of the functions of these counterpoised heterogeneous populations can result in pathological inflammation including both forms of IBD[142].

Taken together it is clear that the intestine can be considered as an immune organ in its function, and the epithelium is a key component of immune tolerance through antigen passage[37] and skewing of tolerance in CD103+ DCs of the lamina propria[145]. How this balance is dysregulated in diseases such as IBD is still not fully appreciated, and understanding of the key immune maturation events that prepare the intestine for performing these functions in adulthood is lacking.

1.4.5 Intestinal muscularis

The intestine is tightly wrapped in continuous smooth muscle throughout its tract, reflecting the ongoing requirement for peristalsis. Intestinal smooth muscle cells (SMCs) arise early in development from the splanchnic mesoderm and develops throughout embryology so that in the mature small and large intestine it will consist of two layers - the inner muscularis mucosa and the more outwardly placed MP (between submucosa and serosa). The MP is further divided into an inner circular layer and outer longitudinal layer[146].

The exact developmental sequence of the muscle layers forming is debated, with either both developing at the same time or the inner layer preceding the outer layer[147,148]. In either case these events must time closely and interact with the forming ENS as the myenteric plexuses will come to lie between the layers. Once mature, the majority of the cellular content of the muscle layers will be SMCs which will be elongated and exhibit thin (actin and calponin) and thick (myosin) filaments with a different arrangement to their skeletal muscle counterparts[149]. They highly express the gene actin (*ACTA2*) and in the GI tract also the isoform gamma enteric actin (*ACTG2*)[149].

SMCs cannot perform their role in isolation, and they are electrically coupled to two specialised interstitial cells that can orchestrate contraction signals from the ENS - Interstitial Cells of Cajal (ICCs) and PDGFRα+ cells – collectively known as the SIP syncytium[150]. These form pacemakers for the intestines contractions, and are tightly adhered to the muscle through tight junctions and can express a variety of neurotransmitters, hormones, paracrine substances and inflammatory mediators[151].

The muscularis also interacts closely with the immune system, specialised muscularis macrophages (CD11b+CXCR3+) are reported to be localised in the muscularis

externa and have a protective phenotype through expression of *Arg1, Il10* and *Retlna* among others[152]. These muscularis macrophages are placed to be closely related to neurons and can mediate contraction, or clear pathogens in a neuro-protective setting[152]. Together this demonstrates the muscularis to be actively interacting with other cell types as well as performing its role of continual peristalsis, the importance of the muscularis in inflammation is also apparent with thickening of the MP being a prominent histological feature in IBD[153].

1.4.6 Mesothelial and serosal cells

The serosa is a loose layer of connective tissue and mesothelial cells that surrounds the outer muscularis and is the final layer of the intestine. It covers the alimentary canal and protects it within abdominal cavity by producing a lubricating fluid, although it lost at some points such as in the oesophagus[154].

Despite being the most distant cell type from the epithelial barrier, it's role in development spans the thickness of the intestine. It has been shown that mesothelial cells are involved in the process of vessel formation within the intestine, sending branches down to muscularis, mucosa and lamina propria. They can be identified by the marker *WT1* and are the major source of vasculogenic cells of the gut[109]. This is driven by the epithelial-to-mesenchymal transition process, demonstrating a close relationship with the more central mesenchyme, which is shown to be an ongoing process in postnatal mice following irradiation based injury[155].

Even in maturity the mesothelium can exhibit regenerative potential, showing effective reduction of adhesion formation[156]. Conversely the serosa is reported to be involved in Crohn's disease (CD) and can act as a lead-point for fistula formation to other viscera, in

some rare instances resulting in florid serosal reactivity despite the site of inflammation activation being the epithelial barrier[157].

This exemplifies that even the most distant cell compartments can affect each other, and that cell communications and behaviour cannot be considered purely in isolation. Diverse roles of these and other intestinal cell types, which may change in a spatial context, make bulk processing and techniques challenging to decipher spectrums of change. To fully appreciate these interactions higher resolution techniques would be needed.

1.5 Dysregulation of barrier functions in disease

As described there are a large diversity of cells that form and maintain the intestine, and their interactions and functions have a co-ordinated effect in balance with their surroundings. Any deviation from these tasks at any point in life, or a breakdown in the protection afforded by the epithelial barrier can result in disease. These may be developmental if occurring *in utero*, inflammatory in adult life, or oncological processes in later life when homeostatic mechanisms falter.

Abnormal intestinal development can lead to significant disease, with congenital defects being the fifth most common cause of death in children under five years worldwide; intestinal surgery represents 40% of emergency neonatal operations[158,159]. A great deal remains to be understood about the origins of these disorders, and an understanding of the key events of intestinal formation would help address this. The molecular pathogenesis is often unclear as deviation from normal development happens before birth when the opportunity to obtain tissue is rare. It is clear that these diseases can present in a variety of ways: antenatally on anomaly scans such as ventral wall defects;

perinatally with acute disease in the setting of atresia's or anatomical malformations; or postnatally with congenital sequelae such as malrotation and volvulus. Although ventral wall defects appear to be distinct from epithelial barrier regulation in presentation, the closure or the abdominal wall times with other key hallmarks of developmental health such as SMC and villi formation, and so many diverse pathologies may be closely interlinked with intestinal developmental milestones[160].

There is evidence that a large number of these conditions occur early in development and involve multiple cellular compartments; defects in recanalization of the lumen can result in stenoses or duplications and will occur in the fetal period before 12PCW[161], and defects in villi formation can occur in animal studies when there are errors in signalling pathways including BMP, hedgehog, PDGF and Wnt which orchestrate endodermal-mesodermal cross talk[161]. Other conditions such as Hirschsprung's disease is identified as an error in the migration of the ENS, but the effect of this is far reaching with inflammation occurring at the barrier surface in non-affected intestinal tissue after causing Hirschsprung's enterocolitis[162]. Exact aetiology is unclear although defects in barrier maturity and microbial response are hypothesised; interestingly it affects the epithelial barrier in sections where ENS migration has been macroscopically normal even after diseased tissue is resected[163].

Even after full maturation of the intestine, breakdown of the symbiotic relationship between cell types and the outside environment can occur. IBD, consisting of Ulcerative colitis (UC) and CD, results from a breakdown of tolerance and in the mutualistic relationship between intestinal compartments in a genetically predisposed individual[164]. Intestinal inflammation of UC is restricted to the colon, whereas CD can affect any part of the alimentary canal, although both forms can have systemic extra-intestinal manifestations. Despite best medical management, and the development of novel

therapeutics including TNF-α inhibitors, up to 40% of patients will fail to respond to current best medical management[165].

The prevalence of IBD in the UK is estimated to be between 0.5-1%[166] and a significant proportion of these are children, leading to lifelong disease burden, with the prevalence of paediatric IBD increasing[167].

Although UC and CD are eponymized as disparate clinical entities, their pathology and features closely intersect. Distinction can be challenging at first presentation and 110 of 163 IBD associated genes overlap in both forms, with IBD "undefined" being a clinical phenotype seen especially in children[168]. This reflects a lack of understanding in the drivers of each at a cellular level. Epithelial barrier breakdown is a common feature in both forms of IBD, both showing destruction of the epithelial layer and microbial breach of mucosal defences[43]. The underlying mesenchyme has also been shown to direct inflammatory stimuli in both[95].

Although the impact of intestinal diseases is clear, a great deal is still unknown of the molecular drivers of disease processes. Identification of this may reveal new targets for therapeutics or new insights into pathogenesis.

1.6 Cataloguing cellular dysregulation at high resolution

Given the complexity and variety of changes in physiological processes that may contribute to disease, studying of a cells transcriptomic features has long been a useful tool for charting these. Although initial approaches required a targeted approach such as quantitative polymerase chain reaction (qPCR) or northern blots for expression of a gene of interest, the application was vastly broadened through the use of high throughput sequencing. This method, RNA sequencing (RNA-seq), allowed mapping of the

expression profile in a tissue or large number of cells in an unbiased manner[169]. Application of RNA-seq increased exponentially and has revealed many new insights into gene regulation and differential expression across tissue, organs and in disease context including gene expression, transcriptional regulation and protein correlation[170–172].

Within the intestine new insights into both the development and dysregulation of the intestine have been revealed using a RNA-seq approach such as the conserved features of human IBD compared to murine models of inflammation[173], or the gene signatures and pathways that correlate with disease severity[174]. Interrogation of RNA-seq data can apply computational approaches to link with other features of intestinal dysregulation including polarisation to a Th1 immune phenotype or response to microbial dysbiosis in IBD[175], and apply dynamic longitudinal changes in development revealing differential timing of expression patterns along the human hindgut[176].

Although RNA-seq has become a tool applied extensively to many fields of molecular biology, it does have some inherent limitations. Namely that a relatively large amount (e.g. $1x10^5$ or $1x10^6$) of cells are required to produce the libraries for sequencing, and that the signature of rare or unique cell types can be lost within these. This is important within the intestine where some cells, such as EECs can constitute 1-2% of the epithelium but have wide reaching effector roles. Methods such as fluorescence protein knock-in and flow sorting have been shown to help account for this in EECs and tuft cells, but methods for more global analysis across cell types was previously elusive[75,177].

The advent of methods to undertake transcriptome based approaches on single cells would overcome these limitations but methodology proved more challenging, and initially the concept was proven in a small number of cells with a panel of targets requiring laborious manual separation and processing[178]. This was then applied to the whole transcriptome through sequencing which set the premise for other methods of RNA

isolation and identifier barcoding, which could then be amplified and sequenced with computational deconvolution to identify cell of origin to truly allow single cell resolution unbiased RNA sequencing (scRNA-seq)[179]. From this principle a variety of encapsulation, multiplexing and protocol miniaturisations heralded an exponential increase in cell throughput in recent years with methods to sort cells and undertake plate based sequencing[5], separate cells with integrated fluidic circuits[180], droplet encapsulation[181] or with in situ barcoding[182] to name but a few. Alongside the advent of various techniques it has become clear that the outputs generated from these investigations have been increasing exponentially with studies now at the organ or organism level[182,183] revealing new insights into cellular diversity and the interaction between cells at high resolution. The ability to measure gene expression in a single cell mean that results are cell type specific (unlike DNA), more easily interpretable compared to epigenetic characterisation, and scalable to thousands of features (compared to targeting proteins) whilst also being able to be integrated with each of these important counterparts.

At the time of commencing this study, large scale reports of scRNA-seq application had been limited to model organisms or preliminary human investigations. The application had shown hitherto unappreciated diversity in the human lung and brain, with new cell types identified in each by using an unbiased approach – the *CFTR* expressing pulmonary ionocyte that may have a role in cystic fibrosis pathology[184], and the elongated "rosehip" interneuron in human brain that had no murine counterpart[185]. It is expected that this trend of exponential expansion of applications will continue[186], with studies on an organism scale expected to identify new features of human biology[187].

Within the intestine this upscaling was also reflected. Studies initially showed differential expression in hundreds of cells identifying ISC heterogeneity or different EEC lineages in murine tissue or organoids, and number allowed plotting of their maturation

through transitions of RNA expression with trajectory analysis[188–190]. Subsequently tens of thousands of murine intestinal cells were used to reveal novel features of Wnt and RSPO modulation in the crypt[191] and even over 50,000 cells undergoing scRNA-seq to map the entire diversity of the mouse small intestine showing new EEC sub-sets and tuft cell diversity in response to helminth infection[192].

When starting this research however, this scale had not yet been undertaken in the human intestinal epithelium. Previous work had reported transcriptomic signatures in cancer fibroblasts associated with survival from analysing hundreds of cells[193]. It had also been applied in the developmental setting, with over 5,000 cells sequenced and showing locational and time point differences, although application of these results to rare cell types was difficult as it studied the whole alimentary canal over time[194]. Building on this, in work leading up to this project, scRNA-seq had also been demonstrated as a powerful tool in charting the diversity of fibroblasts in IBD with four sub-types shown (annotated S1-4) each with specialist functions including a crypt-niche supportive population (S2) and a immune ligand expressing population (S4) that was expanded in inflammation seen in UC compared to healthy control (HC) samples[195]. Methods had also been developed when starting to adapt and apply the unbiased approach of sequencing to tissue sections using oligo-barcoded slides, in order to attempt to overcome the limitation of having to digest tissue and lose spatial context[196]. Understanding the interplay of spatial cues is particularly important in development and dysregulation, where patterning and location-specific morphogen gradients and cytokines shape these processes. At commencing this project this technique had not been applied to human intestinal tissue either.

Together this demonstrates the exponential rise of high-resolution techniques in charting cellular behaviour at the molecular level, and how they can be applied to developmental and disease settings. Due to the huge amount of cell types constituting the

intestine, the complex nature of cross talk in its formation, and a lack of understanding in the pathways of its dysregulation it is important area of further study.

1.7 Project aims

Overall it has been shown that the intestine and the epithelial barrier through which is interfaces with the outside world represents a dynamic spectrum of cell types and states. These interact in a co-ordinated manner in order to maintain health and in cross-talk with other compartments during embryological formation. There is a great unmet need to understand the causes of dysregulation and to develop greater insight into barrier breakdown in conditions such as IBD. Previous attempts to study this have been limited by the restriction to *a priori* targets. Use of an unbiased approach could help develop new therapies and better models of disease.

Conversely, in development of the intestine a great deal of information has had to be gleaned from animal models due to the scarcity of human tissue. There are distinct differences in some findings from these and so it is important to study the formation of the human intestine which may gain greater insight to hallmarks of adult maturity such as the stem cell axis, or into rare developmental diseases.

This project will seek to address some of these unmet needs, by using recently developed high dimensional techniques to achieve the following aims:

1) Undertake a census of the colonic epithelium using scRNA-seq in health and IBD to study the drivers of epithelial barrier breakdown

2) Characterise human intestinal development using transcriptomic and spatial methods to determine epithelial fate decisions and place mesenchymal sub-sets in development
3) Map the nature of cross-talk between colonic mesenchyme and epithelium

Chapter 2: Colonic epithelial cell diversity in health and disease

2.1 Introduction

The intestinal epithelial barrier represents the true interface of the intestines numerous functions – absorption, pathogen sensing, immune tolerance and signalling pathway activation. Thus the cell types included in the colonic epithelium are highly diverse and have a rapid turnover every few days[1]. As epithelial barrier breakdown is a hallmark of both forms of IBD[43], dysregulation in this highly co-ordinated process is implicated. The molecular drivers of barrier breakdown in IBD remain poorly understood, and charting these may identify new pathways for future treatments.

Upon starting this project high resolution techniques to map intestinal cell defects at the molecular level had been performed in mouse tissue[192] or in low throughput scRNA-seq methods on cultured epithelial cells[190]. These endeavours have been highly revealing for new aspects of cellular heterogeneity and function, but application to human samples had been limited. Low throughput techniques had highlighted the approach as feasible but included in the region of hundreds of cell which could not fully recapitulate the entire epithelium[193].

Building on previous work to generate a cohesive structural network of mesenchymal cells using these techniques[195] this project first endeavoured to apply a similar approach to IECs to chart the barrier in health and dysregulation of IBD. This constituted optimising a workflow to isolate epithelial cells from healthy donors and those with active UC, and

undertaking scRNA-seq to categorise cellular heterogeneity and gene expression patterns in homeostasis and in disease.

2.2 Mapping heterogeneity of intestinal epithelial cells

Before undertaking scRNA-seq of adult intestinal samples a protocol was optimised to isolate viable epithelial cells from colonic biopsies collected during endoscopy. Following informed consent (REC GI 16/YH/0247 and IBD 09/H1204/30) samples were collected from patients with active UC who were immunotherapy naïve, and from HCs (undergoing endoscopy for a non-inflammatory indication). These were then transferred to the laboratory and processed immediately.

Adaptions were made to a previously published protocol using Ethylenediaminetetraacetic acid (EDTA) based crypt chelation followed by single cell dissociation[197] (**Methods:6.1**). Flow cytometry was used to demonstrate the single-cell suspension had a high viability of epithelial enriched cells, with only a small fraction of immune or stromal cells (**Methods: 6.2**, **Figure 2.1**). Once the protocol was confirmed as effective for isolation of the cell type of interest, subsequent samples were collected with UC (sampled from inflamed area and adjacent non-inflamed intestine) and HC on the same day, within a few hours of each other, and processed immediately in the same manner with isolated cells being processed directly for scRNA-seq (**Methods 6.2**).

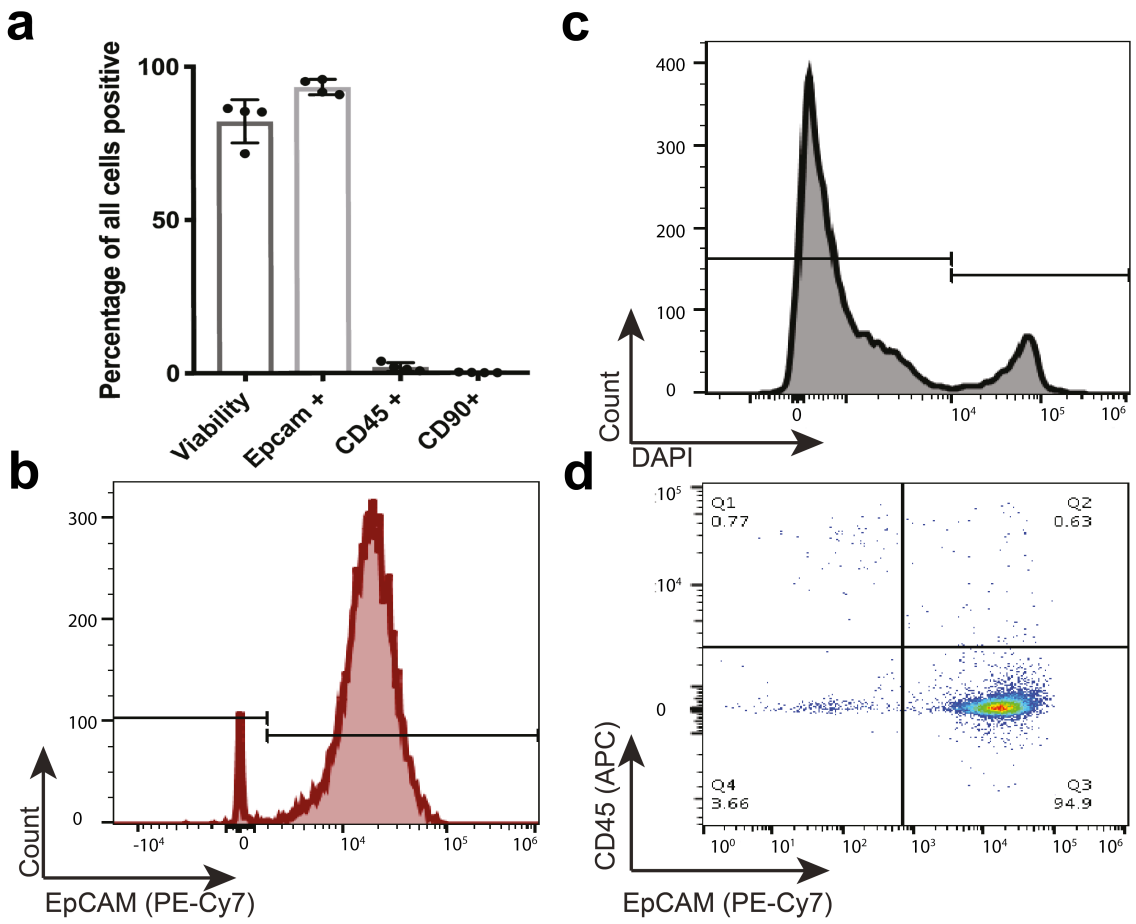

Figure 2.1: Flow-cytometry analysis of cells isolated from colonic biopsies.

a, Percentage of single cells (defined on FSC and SSC) measuring epithelial viability (DAPI-), purity (EPCAM+), immune (CD45+) and stromal (CD90+) sub-sets (n=4, showing mean +- S.D.), demonstrating the gating strategy for known epithelial markers (**b**), viability (**c**) and immune cells (**d**).

2.3 Crypt gradients of absorptive and secretory cells

Using the optimised digestion protocol for IECs, patient samples were collected and proceeded directly to scRNA-seq using 3' droplet-based targeted poly-A RNA targeting and amplification (Chromium, 10X Genomics, **Methods:6.3**) with 3 conditions (HC, UC inflamed and UC non-inflamed) from 2 patients run in parallel on 3 separate occasions (Patients n=6, conditions n=9, **Appendix: A2**). Those patients with UC had active inflammation identified on endoscopy as judged by the UCEIS scoring system[198] and were

immunotherapy naïve. The non-inflamed tissue was taken from adjacent non-inflamed colon which appeared macroscopically normal on endoscopy.

After completion of the scRNA-seq protocol (which constituted of reverse transcription of cells in emulsion, PCR amplification, clean up and indexing) final libraries were pooled and samples were sequenced to an aimed depth of 50,000 reads per cell using Illumina sequencing (**Methods: 6.3**).

Raw sequencing files underwent QC and alignment which constituted of FastQC, alignment of UMIs to human genome with CellRanger and export of count matrices for further computational analysis using Seurat (**Methods: 6.11**). In total, after stringent quality control (QC), 11,175 cells were captured for subsequent downstream analysis with average cell recovery of 1400 cells per sample and 1,736 genes per cell.

Studying the differential expression of the transcriptome of cells obtained from healthy tissue first to confirm utility of the technique, dimensionality reduction identified there were ten clusters of cells identified using *t*-distributed stochastic neighbourhood embedding (*t*-SNE) (**Figure 2.2a, Methods: 6.11)**. The majority of these were readily identifiable as representing known cell types of the epithelium, based on their cluster specific transcriptional signature matching literature on previous bulk RNA-seq marker genes or proteins which the RNA forms. In this manner the clusters were annotated upon their marker genes (**Figure 2.2b Appendix:A.3**). For example; goblet cells on *MUC2, SPINK4, TFF3, LRRC26* and EECs on *CHGA* and *CHGB*[190]. Three clusters expressed genes in keeping with the annotation "colonocyte" with expression of key genes such as *CA1, CA2, SLC6A2, FABP1* and *KRT20*[199,200]. One of these clusters exhibited relatively higher expression of mature colonocyte markers (*AQP8, TSPAN1, CEACAM7* and *CA4*) and were therefore annotated as a mature "crypt top" colonocytes[201].

Given the nature of the highly dimensional scRNA-seq data in showing dynamic transcriptional changes across clusters, the maturation axis of the colonic crypt was reflected in RNA expression gradients and could be studied in greater detail. By mapping the gene expression gradients of maturation within and between clusters by differential expression of known crypt axis markers such as *SEPP1*[202] and 14 other genes allowed assignment of each cell to a crypt-axis score (**Figure 2.2c, Methods:6.11)** and placed stem cells at the base and crypt-top colonocytes at the top; recapitulating the placement of known cell types of the intestinal epithelium.

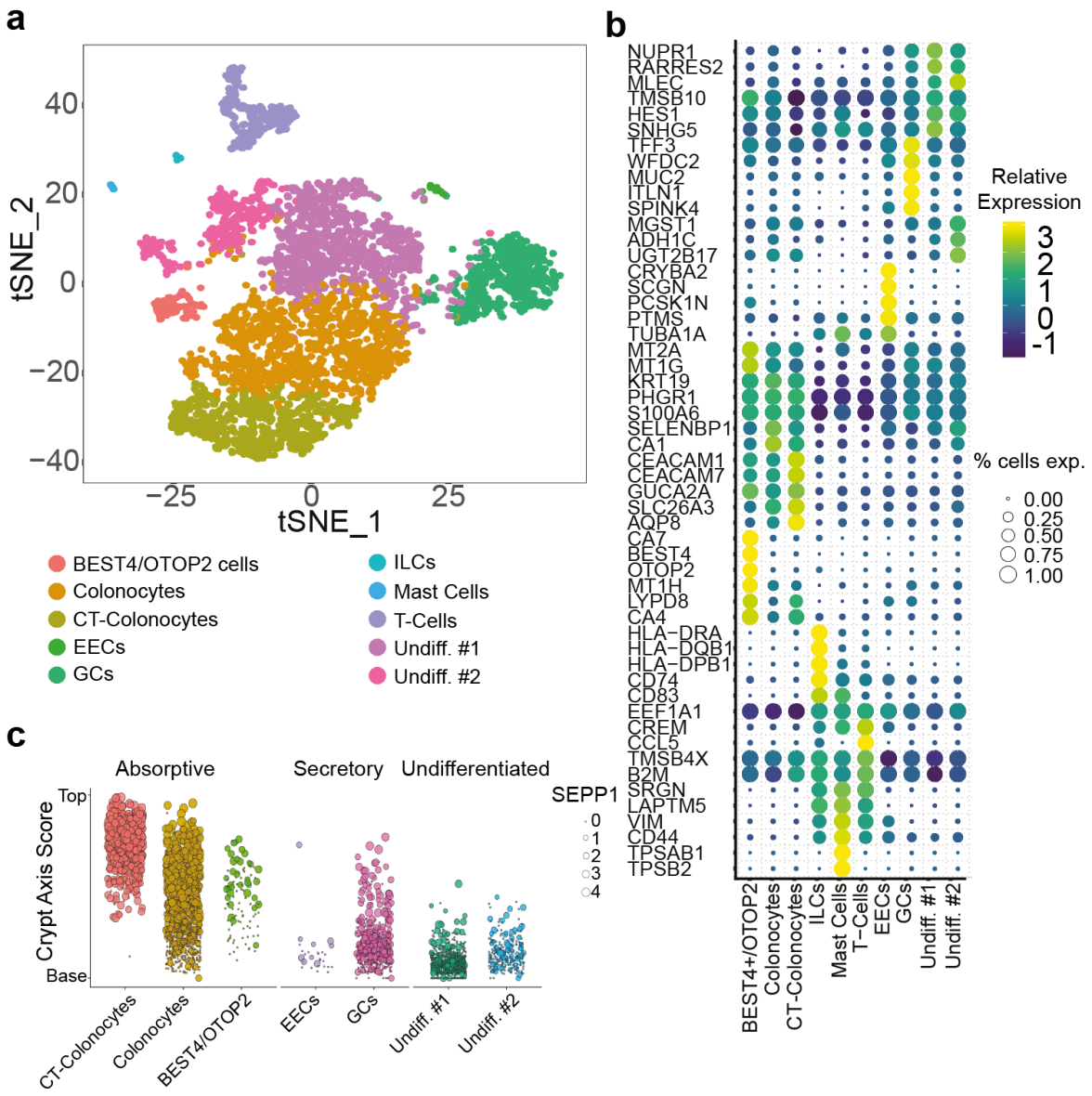

Figure 2.2: Colonic epithelial cell diversity in health.

a, *t-sne* of healthy epithelium (n=3 patients with n=9 conditions; CT – Crypt Top). **b,** Heat map showing gene cluster markers for **a** (coloured by relative gene expression; relative size represents fraction of cells expressing marker). **c,** Spatial segregation across the crypt of cell clusters (y-axis) using crypt-axis score of 15 markers (**Methods:6.11**, size of dot representing *SEPP1* expression).

In some instances, the biological cell types within these clusters could be resolved further; such as in the case of EECs which constituted <2% of all cells but had highly specific expression of neuro-peptide related genes. These could be distinguished on transcriptional signature to the EEC sub-types which are seen within the colon – Enterochromaffin cells and L-cells – alongside an EEC precursor population (**Figure 2.3a,**

Appendix:A.4)[203]. The precursor population had lower expression of mature neuropeptides such as *PYY* and *CHGA* but higher expression of genes such as the important secretory and developmental transcription factor *SOX4*[204], reflecting the shared origin of secretory populations as progenitors[205]. Previously undescribed EEC marker genes were revealed including *CPE* and *PCSK1N* which were validated at the protein level using IHC to co-localise with Chromogranin A (CHGA) (**Figure 2.3b, Appendix:A.4**). Overall this showed scRNA-seq was able to identify human colonic IEC heterogeneity at high resolution in even rare cell types.

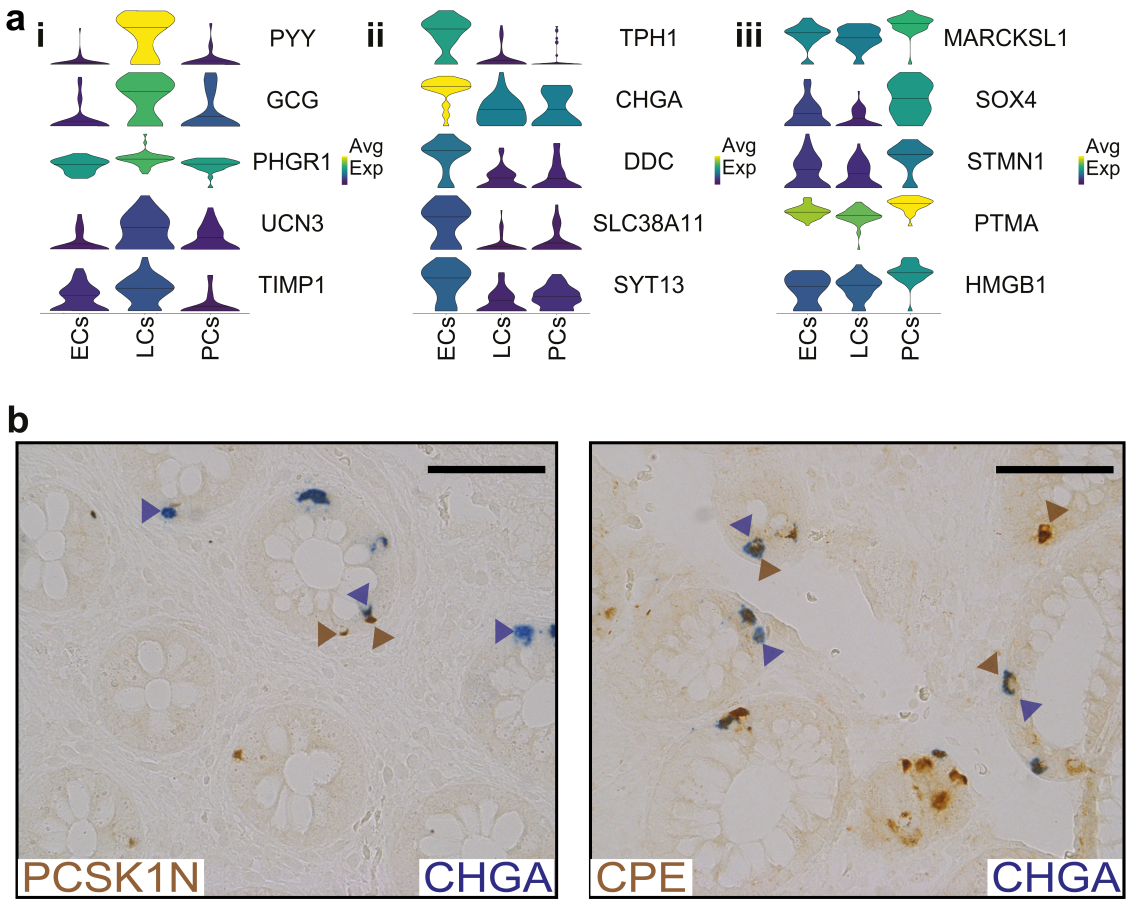

Figure 2.3: Enteroendocrine cells of the colon

a, Cells of the EEC cluster were analysed in isolation to identify two known colonic EC sub-type (Enterochromaffin(ECs) and L-cells (LCs)) with a precursor population (PCs). **b,** EEC marker genes shown from data (**Figure 2.2b**) were validated by IHC co-staining for new marker (brown) alongside known marker CHGA (blue) demonstrating single and double stained EECs (arrows, 20x magnification, scale bar represents 90 µm, representative stain of >3 individual experiments).

Given this resolution and that the resource was able to identify the maturation of the colonic crypt axis, was then applied to study cells in the spectrum of their fate decisions. ISCs of the colon could be identified by partitioning the previously identified undifferentiated cluster which reflected expression of their known markers (*LGR5, ASCL2*)[8,199] (**Figure 2.4a, Appendix:A.5**). Furthermore, all epithelial cells could be placed in pseudo-time using trajectory analysis (**Methods 6.11)** which reflected the maturation axis across the crypt (**Figure 2.4b**).

This generated a bifurcating trajectory with division of primitive ISCs and TA cells into absorptive and secretory lineages, showing mature crypt top colonocytes and goblet cells as the final stage of their respective lineages (**Figure 2.4c**). The fate decisions at bifurcation could also be mapped (**Figure 2.4d**), revealing the loss of ISC markers (e.g. *LGR5)* alongside the gain of lineage specific transcription factors such as *SOX4* of the secretory lineage[206] or *ADIRF* for absorptive which is known to be highly expressed in mature enterocytes[207] (**Appendix: A.6**).

Taken together these results demonstrated the scRNA-seq data was able to reflect known features of epithelial maturation, with placement of ISCs and their subsequent mature cell-types as well reflecting the effects of transcriptional regulators – something that has not previously been possible outside *in-vivo* modelling. It also showed that the approach was able to chart both known and novel features of IECs in health, and provide a platform to compare these changes to the inflammation of IBD.

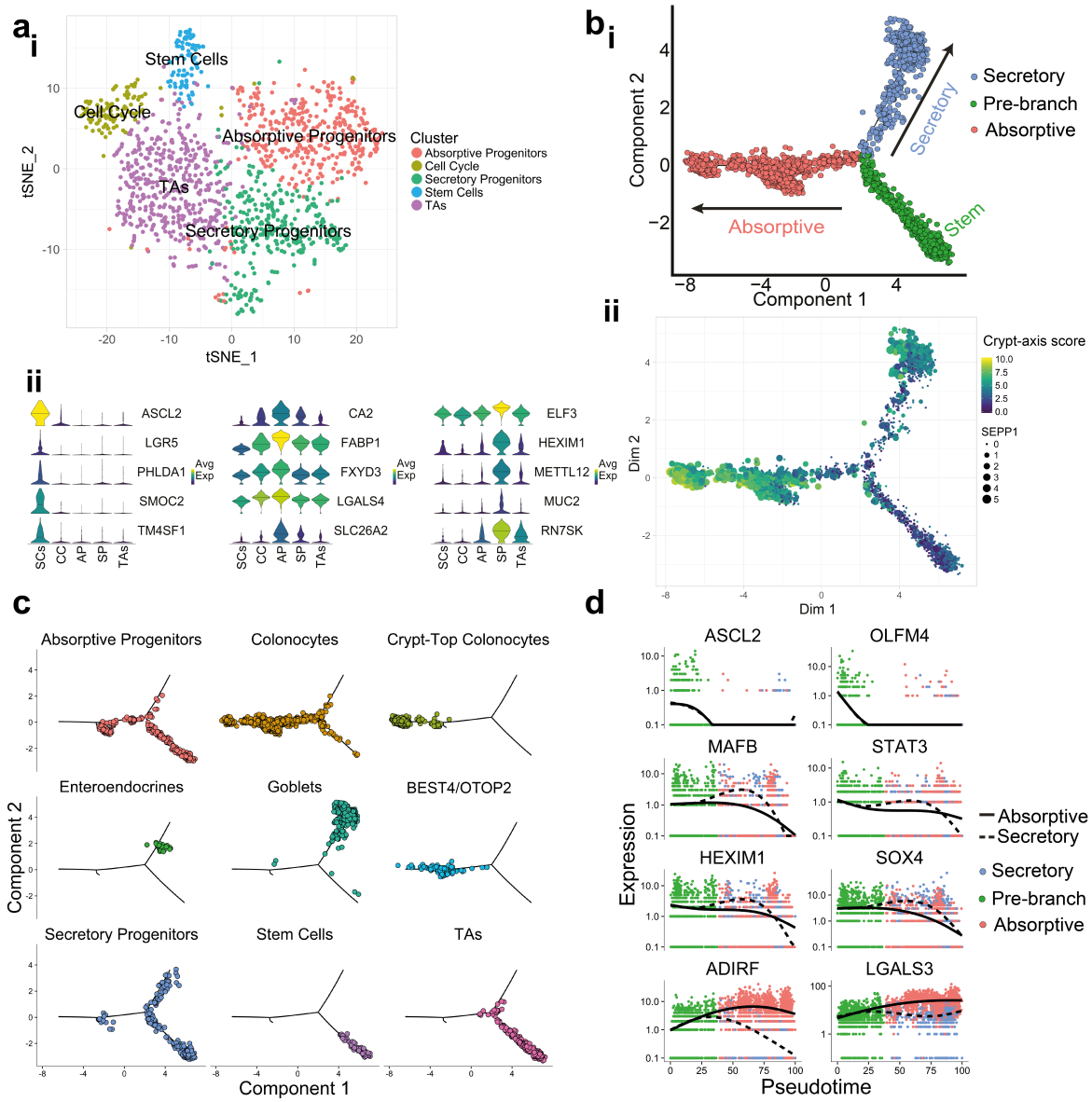

Figure 2.4: Epithelial cell differentiation and maturation

Undifferentiated cells were partitioned with **a,** *t*-sne plot visualising colonic epithelial cell sub-clusters (**i**) and differentially expressed genes (**ii**) in keeping with stem cell (SC), absorptive (AP) and secretory progenitors (SP). **b**, Differentiation pseudo-time trajectory analysis of epithelial cells, coloured by predicted lineage which correlated with previously described crypt axis score (**ii**). Individual cell types are shown on trajectory analysis (**c**) and branch-specific expression of selected SC markers are plotted over pseudo-time (**d**) with lineage transcriptional regulators.

2.4 Gene responses in inflammation

Building upon the findings of IECs in health, we next sought to utilise this as a reference to study the cell-specific changes in the barrier in IBD by comparing relatively

healthy IECs, alongside inflamed and non-inflamed samples from the same individual (total n=9 samples from n=6 patients). Performing clustering analysis on the inflamed cells was able to recapitulate the cells seen in health, along with two additional clusters – inflammation associated goblet cells and intraepithelial immune cells (**Figure 2.5a**) This was able to identify 1147 genes (<1% FDR) that were dysregulated in UC at a cell type specific level (**Figure 2.5b, Methods 6.11**, **Appendix: A.7**). The greatest number of these changes were observed in the luminal facing cells: colonocytes and crypt-top colonocytes (n=734 and 676 respectively) with smaller numbers of genes seen to be differentially regulated in goblet cells, stem cells and EECs (n=140, 65 and 4 respectively).

Gene ontology enrichment (GOE) of DEGs showed that some processes were conserved across all cell types such as response to interferon- γ signalling or neutrophil activation, whilst some were more cell type specific – antigen processing having a higher gene ratio in ISCs (**Figure 2.5c , Methods 6.11**). This may represent recently reported findings in the mouse intestine – where ISCs can act as non-conventional APCs to play an active role in immune recruitment and epithelial cell remodelling[208].

The ability to then map these processes to cell types gave the means to dissect the components and identify the cell-type-specific responses to UC inflammation. Colonocyte populations showed downregulation of metabolic processes such as *HMGCS2* a mediator of lipid processing[209]. They also simultaneously inducing genes involved in the production of reactive oxygen species and microbial killing (*SAA1, DMBT1, PLA2G2A*)[210,211]. Goblet cells upregulated stress response genes that promote cell survival over apoptosis such as *LYZ* – normally a Paneth cell gene in the small intestine – was upregulated in the lower crypt goblet cells in inflammation (**Figure 2.5d)** and could represent a change in role similar to the previously described "deep crypt secretory cells" of the colon that protect the ISC niche producing lysozyme to prevent bacterial damage to the crypt during colitis[36]. In

contrast ISCs showed downregulation of Heparin binding EGF-like growth factor (*HBEGF*) (**Figure 2.5e)** which has been shown to protect the intestine from injury by preserving Wnt/ β-catenin signalling in ISCs[212]. Given this role in health, the inability to upregulate expression in UC may negatively affect intestinal regeneration.

Taken together, studying the dynamic transcriptional changes of IECs in inflammation revealed the outcomes to these inflammatory events depends on how individual cell subtypes balance the dual requirement to restore health and tissue integrity whilst simultaneously responding to aberrant tissue homeostasis.

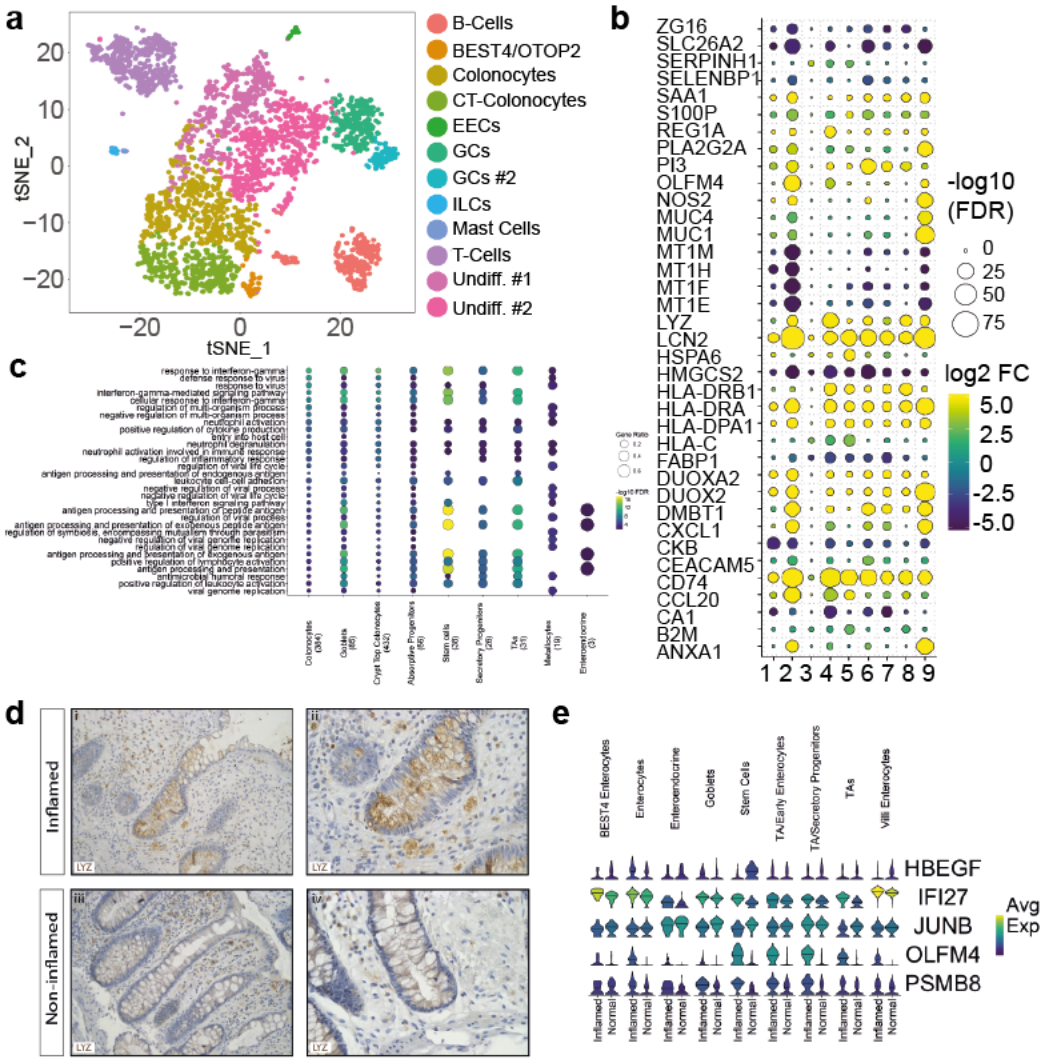

Figure 2.5 Response of epithelium in active colitis

a, *t*-sne plot of single cell clusters in active UC. **b,** Heatmap showing key differentially expressed genes (FDR <1%) between health and active colitis (colour indicating up/downregulation, size of dot representing significance. Cell types; 1-BEST4/OTOP2, 2-colonocytes, 3-EECs, 4-goblet cells, 5-ISCs, 6-absorptive progenitors, 7-secretory progenitors, 8-TA cells, 9-crypt top colonocytes). **c,** dot plot of GO enrichment in upregulated genes (<1% FDR) comparing cell clusters DEGs active processes in comparison to health (points coloured by enrichment confidence, size representing proportion of all genes within cluster annotated with GO). **d,** Immunohistochemistry (IHC) images (representative of n = 3) showing LYZ expression (brown) in inflamed and noninflamed colonic tissue sections. (**i, iii**, ×20 magnification; **ii, iv**, ×40 magnification).**e,** Violin plots showing expression (y-axis) of selected genes dysregulated in inflammation compared with health in ISCs and/or other undifferentiated populations (colour indicates mean expression).

2.5 Heterogeneity of goblet cells in health and IBD

In the colon the lack of Paneth cells in the crypt is reported to be alleviated by goblet cells playing a ISC supportive role, namely deep goblet secretory cells marker by *REG4*[36]. Dysregulated goblet cell function has been reported in both forms of IBD – depleted in UC but expanded in CD[43] – and it has also been shown that they can direct immune tolerance through antigen presentation to tolerogenic (CD103+) DCs[38]. Despite the importance of goblet cells in these processes it is still not fully understood what pathways underpin goblet cell dysregulation in IBD or how they perform these different functions, so we sought to interrogate the pathways that might drive this.

Pooling scRNA-seq profiles of all goblet cells across health and disease revealed that they divided into 5 clusters (**Figure 2.6a, Appendix A.8)**. Key genes that were differentially expressed across these clusters suggested that one aspect of the heterogeneity could be accounted for by the crypt depth of the goblet cells, with cluster 3 having expression of secretory progenitor markers (*SOX4*) and cluster 5 with a more mature, luminal facing, phenotype demonstrated by relatively higher expression of *MUC2* (**Figure 2.6b**). Linking the goblet cells with the previously derived score of crypt top axis quantification this supported this, with clusters placed across depth and maturity. (**Figure 2.6c)**.

These findings showed that although common inflammatory responses exist throughout goblet cells there is also spatially distinct roles. This was shown in genes such as *LCN2* and *REG1A* induced throughout the crypt, compared to *CD74* and *LAMB3* which showed locational limitations (**Figure 2.6d)**. Both of the latter genes have been shown to have roles in colorectal cancer progression, but our data suggests their response to inflammation was location specific[213,214]. Consistent with these observations we saw that

one cluster was inflammation associated, somewhat homologous to cluster 4 but seen predominantly in active UC and differentially expresses laminins (*LAMC2, LAMB3*) which have been shown to essential for the integrity of the epithelial barrier (**Figure 2.6b**)[215].

Finally, to confirm the differential spatial localisation of key markers of goblet cell cluster immunofluorescence (IF) was used to show BCAS1 (cluster 5), CLCA1 (cluster 1), REG4 (cluster1) alongside MUC2, with CLCA1 seem predominantly at the bottom of the crypt and REG4 seen at the middle and upper crypt (**Figure 2.6f**). Thus, confirming the differential expression seen in the resource, likely reflecting that goblet cells have specific spatial roles when appreciating them with a single cell approach.

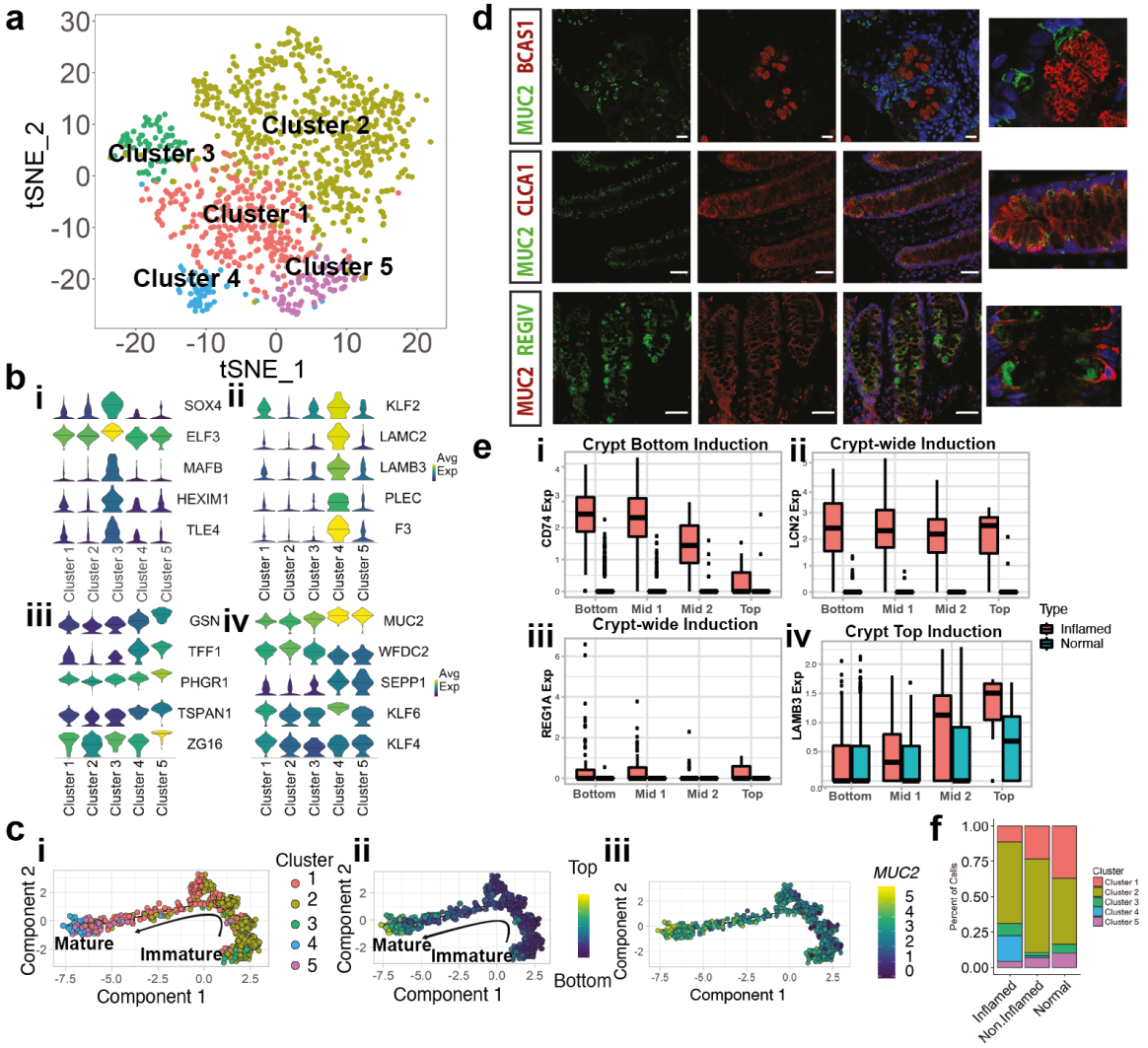

Figure 2.6: Goblet cell heterogeneity in health and disease

a, *t*-sne plot of goblet cell sub-clusters across all captured cells. **b,** Violin plots showing cluster gene expression (y-axis) for marker genes in cluster 1(**i**), 2 (**ii**), 3 (**iii**) and common to cluster 4 and 5 (**iv**) (centre bar shows median value and colour showing relative expression). **c,** Pseudo-temporal ordering of goblet cells shown by cluster (**i**), crypt-axis score (**ii**) and expression of mature goblet cell marker *MUC2* (**iii**) demonstrating differences in maturity. **d,** IF showing protein expression goblet cell markers which stain similarly (MUC2) or differentially (BCAS1, CLCA1, REG4) across clusters in healthy colonic tissue (representative images of 3 stains, colour of marker reflected in text, nuclei stained with DAPI, scale bars represent 20µm top, and 50µm in others). **e,** Box plots showing expression of selected goblet cell genes, divided along crypt axis by division into four ranges in health and active colitis (bottom, mid1, mid2 and top, n = 3 per group; 25th, 50th and 75th percentiles shown, colour reflecting inflammation status). **f,** Stacked bar chart showing relative frequency distribution of goblet-cell sub-clusters (percentage of all goblet cells captured) in health, inflamed and inactive non-inflamed colitis.

2.6 The serine protease WFDC2 is required for barrier integrity in health

A gene that was present in the deeper goblet cells, with expression downregulated in the more mature cell types was *WFDC2*(**Figure 2.6b**). This gene had not previously been well described in the intestine, although previous work reported its presence in the testis and ovary[216]. This encoded for a whey acidic protein, which although previously thought to be biologically inert recent studies have shown family members such as SLPI to be important in intestinal homeostasis[217] and propose that it may have an antibacterial role[218]. Our data showed that *WFDC2* was selectively downregulated in inflammation (**Figure 2.7a**). This finding was validated histologically in a larger cohort of patients with active UC using paired sampling from inflamed and non-inflamed sections confirming the selective loss of protein expression in active inflammation (**Figure 2.7b**). Additionally, *in-vitro* culture of *WFDC2* with a panel of intestinal microbes demonstrated that is had a selective anti-bacterial effect suggesting a protective role in the colon (**Figure 2.7c**).

Furthermore, to confirm this effect *in-vivo*, we used a heterozygous knock-out murine model of $Wfdc2^{+/-}$ (required as homozygous knock out is embryologically lethal[219]). These mice exhibited abnormal histology, with epithelial hyperplasia (**Figure 2.7d**) along with increased bacterial adherence using Gram staining (**Figure 2.7e**). Studying this in detail using TEM confirmed increased bacterial attachment and invasion in goblet cells compared to wild type mice (**Figure 2.7f**). Showing *WFDC2* was a goblet cell protein with anti-bacterial effects that is selectively lost in UC associated inflammation.

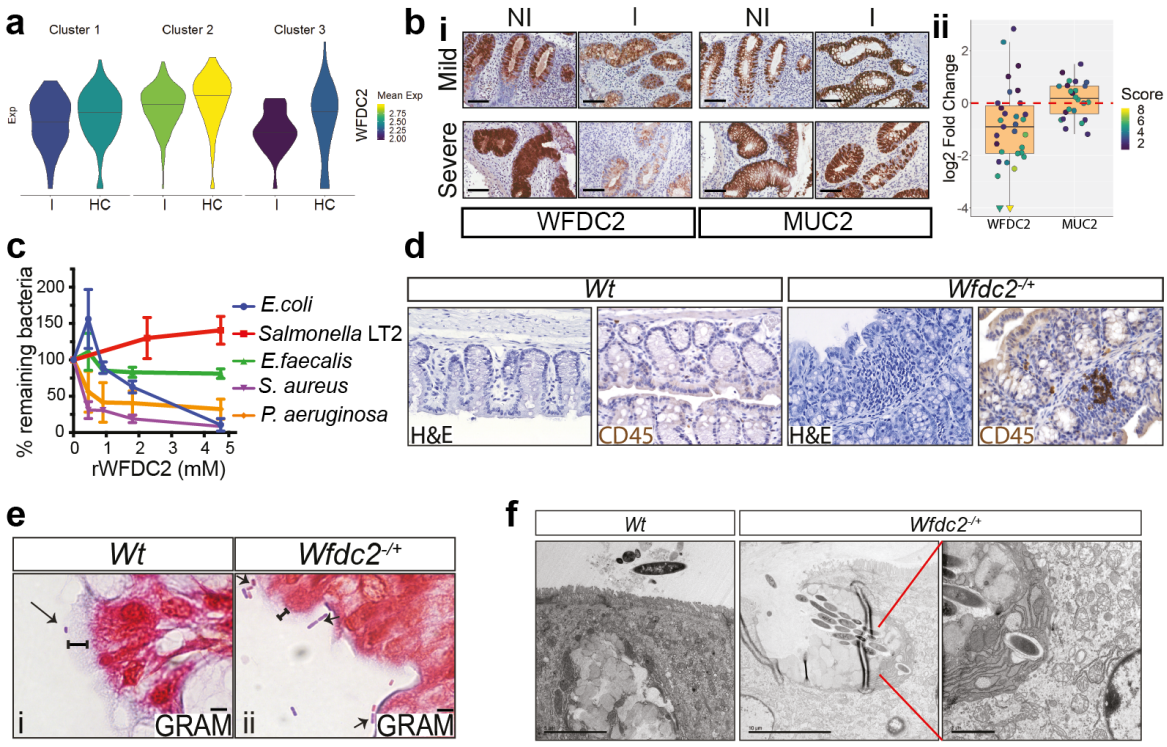

Figure 2.7: *WFDC2* is lost in inflammation and involved in barrier defence

a, Violin plots showing expression (y-axis) of *WFDC2* in goblet cell clusters in health (HC) and inflammation (I) (centre bar representing median, colour representing mean expression). **b**, IHC staining for WFDC2 protein in severe and mild inflammation showing selective depletion with MUC2 expression conserved. Representative images of a cohort of stains (WFDC2 n=31 and MUC2 n=24)which is quantified (**ii**) using automated quantification methods (**Methods: 6.8**) showing fold change (y-axis expressed $\log_2$) from inflamed and non-inflamed paired samples (point coloured by severity score / UCEIS, triangles represent outliers with >16-fold depletion, two-sided paired t-test p=0.00018). **c,** Recombinant WFDC2 (rWFDC2) was added to mid-logarithmic-phase bacteria for 4h and quantified by dilution plating (means +-SD, n=3). **d,** Histopathological changes in intestinal samples from WT and *Wfdc2*$^{+/-}$ mice showing irregular crypts (H&E) and focal leukocyte infiltrated (CD45+, magnification 20x in all). **e,** Gram staining of murine intestinal tissue showing increased colonisation and adherence (arrow) in heterozygous *Wfdc2*$^{+/-}$ mice along with reduced thickness of mucus barrier (depth marker). **f,** TEM showing bacterial invasion in *Wfdc2*$^{+/-}$ mice compared to WT with higher resolution example showing penetration through goblet cells (scale bar 5µm in left image, 10µm in middle, 2µm in right. Representative images of n=4 experiments in each group).

2.7 Interim conclusion

This work represents one of the first single cell surveys of the human intestinal epithelium using high resolution single cell approaches[2]. It demonstrates that this approach is able to characterise the dynamic features of the epithelium in health and its response to

UC-associated inflammation with the ability to demonstrate cell sub-type specific gene expression in cells such as EECs and goblet cells, which were validated *in situ*.

Results were able to identify gradients of progenitor cells and novel transcriptional regulators induced prior to lineage commitment. These gradients may act as fate specifiers of intestinal epithelium in health.

The finding that goblet cells have heterogeneous sub-populations which can be spatially located across the crypt may reflect adaption to different microflora which have been shown to be differently localised to reside at the crypt top and crypt base[220–222]. Across the crypt all cell types downregulated metabolic pathway genes and induced specific innate antibacterial defence programmes, such as induction of *LCN2* that binds to bacterial siderophores to limit bacterial growth[223]. Crypt top colonocytes specifically induced *MUC1* and *MUC4* that limit mucosal infection[224]. The induction of a broad repertoire of innate defence pathways in diverse epithelial cell types confirms current views of UC pathogenesis resulting from impaired immune responses to dysbiotic microflora originating in barrier cells. To further support this, the work has delineated a functional role for a novel goblet-cell secreted antibacterial protease *WFDC2*. This was expressed in deep goblet cells and was markedly downregulated in the context of inflammation in a severity dependent manner. *Wfdc2*$^{+/-}$ mice distorted villi and were vulnerable to bacterial invasion showing it to be a novel AMP and demonstrating that the approach could identify new facets of epithelial barrier breakdown.

In addition to the dysregulation of behaviour in known cell types there is one area of the healthy intestinal survey which has not been mentioned until now. The use of an unbiased approach has been powerful in identifying new cell types and states in other organs before this work was undertaken – a *CFTR* expressing pulmonary ionocyte in the lung, or a rosehip cell in the brain[184,185] – and one cluster in healthy epithelium did not

match the markers of previously described enterocytes which forms the focus of the next part of work.

Overall this work to date demonstrates this approach is able to characterise dynamic epithelial during inflammation of the colon. It also raises further questions of how these new features interact with the adjacent mesenchymal compartment, whether key differences occur in different locations of the intestine and how these findings apply more widely to the maintenance of health in the intestine as a whole.

Chapter 3: Characterisation of a novel epithelial cell type

3.1 Introduction

The first part of work was able assess the cellular constituents of the colonic epithelium at high resolution in health and UC. This showed pathways of epithelial differentiation and maturation in steady state, and some of the cell type specific pathways that are dysregulated in IBD. In this initial assessment there was a cluster of cells that were seen across the disease states samples which was annotated based on its top expressed genes – the *BEST4/OTOP2* cell. This next section will seek to confirm the presence of the cell and characterise this population more formally. This will aim to address whether it may represent a novel cell type in line in line with other reports describing application of scRNA-seq to other organs[184,185]. In addition, it will aim to address what functional role this cell type may have within the diverse cells seen in the intestinal epithelium.

3.2 Identification of *BEST4/OTOP2* cells on transcriptional profile

Alongside those cell types which were seen in our study of IEC diversity, it was surprising to identify one cluster of cells which had a transcriptional profile that was not able to be readily annotated based on previously described IEC markers. This cell cluster, which was termed BEST4/OTOP2 cells due to its marker genes (**Figure 2.2**, **Appendix A.3**). They expressed colonocyte markers such as *CA7*[225] and *CA4*[226], but were also

uniquely distinguishable from other enterocytes by genes including *BEST4, OTOP2, LYPD8,* and *GUCA2B* (**Figure 2.2**).

BEST4 encodes a protein of the bestrophin family, with members such as the more studied BEST1 and BEST2 proteins shown to be ion channels which can function as Cl^- channels, HCO_3^- channels of voltage gated $Ca2^+$ channels[227]. BEST4 expression has been reported in the human colon and small intestine but the functional role remains to be elucidated[228]. *OTOP2* has been recently described to encode an ion channel that functions selectively in a low pH environment in murine taste receptors[229]. In contrast, *LYPD8* is known to have a differing functional role where secretion in the colon can promote the segregation of flagellated bacteria and may have a potential role in prevention of enteric infection[230].

The survey showed these cells also specifically expressed components of the uroguanylin pathway (*GUCA2A* and *GUCA2B*), uroguanylin being a endogenous paracrine hormone and satiety peptide which activates the guanylate cyclase 2C pathway and modulates epithelial cyclic GMP activity[231]. Other highly expressed genes included those of the metallothionein family (*MT1H, MT1G, MT2A*) which contribute to metal ion processing and storage to provide defence from free radicals[232].

Although these individual genes and pathways have been reported in murine and human intestinal studies, taken together they represent diverse potential roles for the cluster of BEST4/OTOP2 cells, so expression was summarised with GOE (**Figure 3.1a)**.This supported a role in metal (terms with responses to zinc, cadmium and metal ions) and electrolyte transportation (bicarbonate transport, response to inorganic substances). These are important processes in the intestine, with strict pH and metal ion control over the barrier required to maintain intestinal permeability[233] and maintain tightly regulated symbiotic bacterial colonies to prevent infection[234,235]. Trajectory analysis was

undertaken to gain a greater understanding of how this cell was placed in the context of crypt-villus differentiation. This identified it as within the absorptive lineage (**Figure 3.1b**) and previous analysis of crypt-depth placed them at the crypt top (**Figure 2.2c**).

Study of BEST4/OTOP2 differentially expressed genes showed that the transcription factors *SPIB* and *HES4* were specific to the cluster (**Figure 3.1c)**. *SPIB* has previously been shown to be an important transcription factor in the differentiation of M-cells and that it is up-regulated in response to RANKL (TNFSF11) in organoid modelling[71]. The BEST4/OTOP2 cells were however lacking in other reported M-cell markers such as GP2 and annexin-V (*ANXA5*)[71]. *HES4* is a transcription factor that is activated in response to NOTCH signalling a manner akin to the described enterocyte transcriptional regulator *HES1*[236], although less well described in IECs. *HES4* has not been reported in M-cells, even in murine organoids stimulated with RANKL and characterised with scRNA-seq[192].

As this was the first report of the cell type in colonic epithelium with relatively large cell number, we also studied previously reported scRNA-seq and bulk RNA datasets to see if the cell-type signature could be identified. First looking at a previous study of fetal intestine including 5,227 cells from across the whole of the digestive tract and analysed using smart-seq2[194]. Although the number of cells in the study did not identify a distinct cluster in their initial assessment, a semi-supervised approach to colonic epithelium identified BEST4/OTOP2 cells on core transcriptomic signature (**Figure 3.1d, Appendix: A.9)**. In addition, this signature could be seen to be downregulated in bulk-RNA analysis of cancer tissue compared to HCs, in publicly available data from the tissue cancer genome atlas(TCGA)[6] (**Figure 3.1e, Methods: 6.11**) reflecting that BEST4/OTOP2 cells signature could be downregulated in colorectal cancer – a feature which is not clear from our work whether it would be from a loss of cell type or the signature being drowned out by other cells, but something that would be put down to individual genes rather than a cell

type when appreciating only bulk techniques. Together this highlighted that BEST4/OTOP2 cells may represent a novel and distinct cell type and so was studied further to confirm its role.

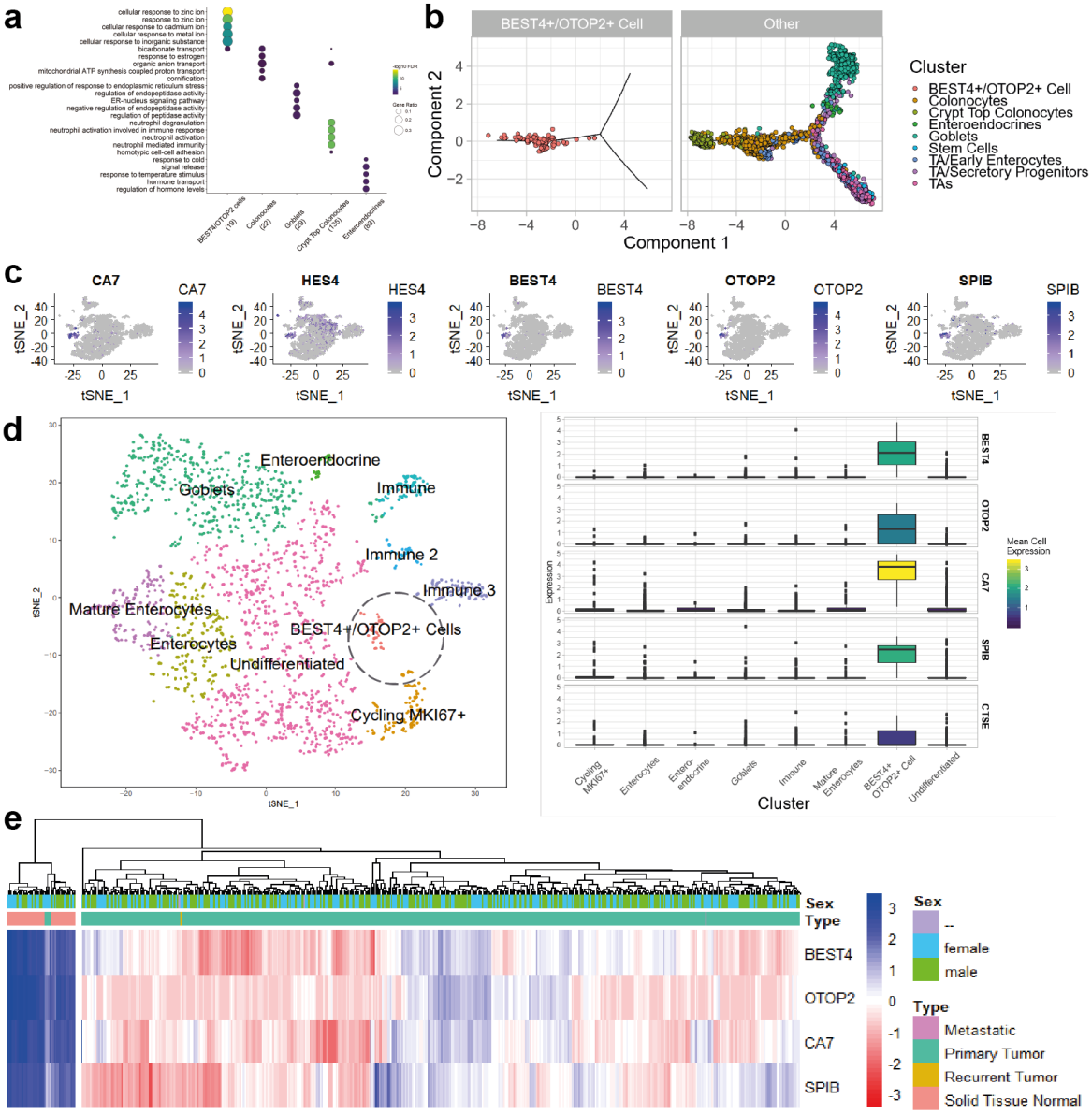

Figure 3.1: Transcriptional hallmarks of BEST4/OTOP2 cells

a, Selected GO terms that show specific enrichment in epithelial clusters and highlight potential functional role for BEST4/OTOP2 cells. (number of markers identified for each on x-axis, size of circle corresponding to proportion of cells expressing and colour representing significance (FDR)) **b,** Cluster distribution along the differentiation trajectory or epithelial cells captured by Monocle[237]. BEST4/OTOP2 cells only are highlighted in left panel. **c,** *t*-sne gene expression overlay across all epithelial cells of core BEST4/OTOP2 markers including the transcription factors *SPIB* and *HES4*. **d,** *t*-sne visualisation of epithelial cells from scRNA-seq dataset of fetal intestine showing semi-supervised clustering for annotation of BEST4/OTOP2 cells (highlighted) and box-plot showing expression of core transcriptomic signature (box-plot showing 25th,50th and 75th quartiles, colour showing mean cell expression[194]. **e**, Heat-map showing the expression of core BEST4/OTOP2 cell gene signature in TCGA bulk RNA-seq in patients with cancer tissue and matched normal tissue[6].

3.3 Validation of *BEST4/OTOP2* cells in tissue

To initially validate the presence of BEST4/OTOP2 cells as a distinct enterocyte tissue samples were stained with IHC and single molecule *in situ* hybridisation (smISH) to confirm the localisation of the core transcriptional signature. IHC for BEST4 and Cathepsin E (*CTSE*) showed that there was a distinct enterocyte population, often visualised at the opening of the crypt (**Figure 3.2a**) in keeping with the *in-silico* analysis of crypt depth. BEST4 was also seen in the underlying stroma which may represent expression in granulocytes, that can be poorly captures in droplet based scRNA-seq methods[238,239]. Cathepsin E has previously been described to be expressed in M-Cells when overlying FAEs but could also be seen on IHC to specifically stain tissue without underlying lymphoid associations[240].

Validation at RNA level using smISH showed a similar pattern of staining, with specific enterocytes expressing *OTOP2, GUCA2B,* along with the aforementioned transcription factors *SPIB* and *HES4* to be localised in a comparable manner (**Figure 3.2b)**.

Co-localisation of both protein and RNA targets was performed to confirm that these marker genes identified the same specific cell type. This showed that BEST4 was co-expressed with both *OTOP2* and *GUCA2B* in specific enterocytes, with a thin morphology, that were located at the opening of the crypt (**Figure 3.2c**).

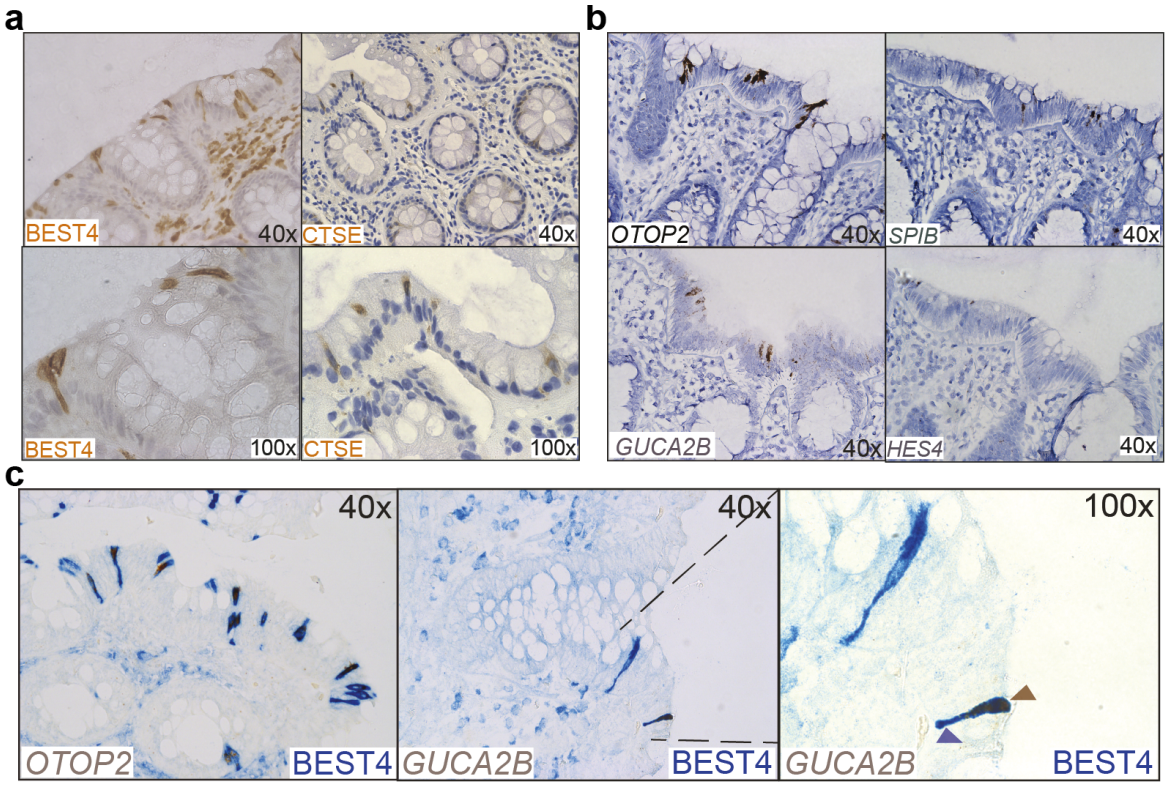

Figure 3.2 Validation of BEST4/OTOP2 cell *in-situ*

a, IHC showing protein markers of BEST4/OTOP2 cells BEST4 and CTSE. **b,** smISH demonstrating RNA targets of BEST4/OTOP2 cells including *OTOP2, SPIB, GUCA2B* and *HES4*. **c,** Combinatorial smISH and IHC staining to demonstrate co-localisation of markers identifying BEST4/OTOP2 cells (all representative of ≥3 stains, protein or RNA target and magnification show in pane, arrows added for clarity)

3.4 Deep characterisation of BEST4/OTOP2 cells

After showing BEST4/OTOP2 cells represented a distinct cell type in tissue, it was then aimed to isolate them from other enterocytes to characterise their role in more detail. A fluorescence-activated cell sorting (FACS) approach was undertaken to achieve this with epithelial cells isolated from HCs in a similar manner and then epithelial cells (EPCAM+) expressing the surface marker BEST4 were separated (**Figure 3.3a, Methods 6.2**). Key RNA markers were validated in the sorted cells using microfluidic-PCR to ensure that marker genes were differentially expressed between EPCAM+BEST4- and EPCAM+BEST4+ cells (**Figure 3.3b, Methods 6.4**).

After ensuring the validity of the FACS strategy, separate experiments were undertaken to gain a more in-depth multi-omic understanding of the new cell type. This included sorting each population and undertaking quantitative proteomic analysis with mass spectrometry proteomics (**Figure 3.3c, Methods 6.5, Appendix A.10)** and also to perform deep scRNA-seq using the smart-seq 2(SSII) protocol[5] (**Figure 3.3d, Appendix A.11**).

The greater read depth of a SSII approach helped identify additional protein and mRNA markers of BEST4/OTOP2 cells, such as mesothelin and DMBT1. Mesothelin in normally expressed in primitive mesothelial cells, although it's overexpression has been reported in intestinal epithelial cells in colorectal cancer where it is associated with poor prognosis[241]. Pathway analysis of proteomic data showed differentially expressed proteins in BEST4/OTOP2 cells that supported a role in oxidative stress and pH balance such as BAG1 and XDH[242]. This was further supported by GO analysis of proteomic data that highlighted the cell type to be involved in processes such as ethanol, small molecule and lipid metabolism alongside icosonoid and fatty acid metabolism (**Appendix A.12**).

Together this produced an in-depth characterisation of the transcriptomic and proteomic features of BEST4/OTOP2 cells, and suggested from many methods of analysis of these data that they may play a role in pH and metal ion balance along with modulating oxidative stress.

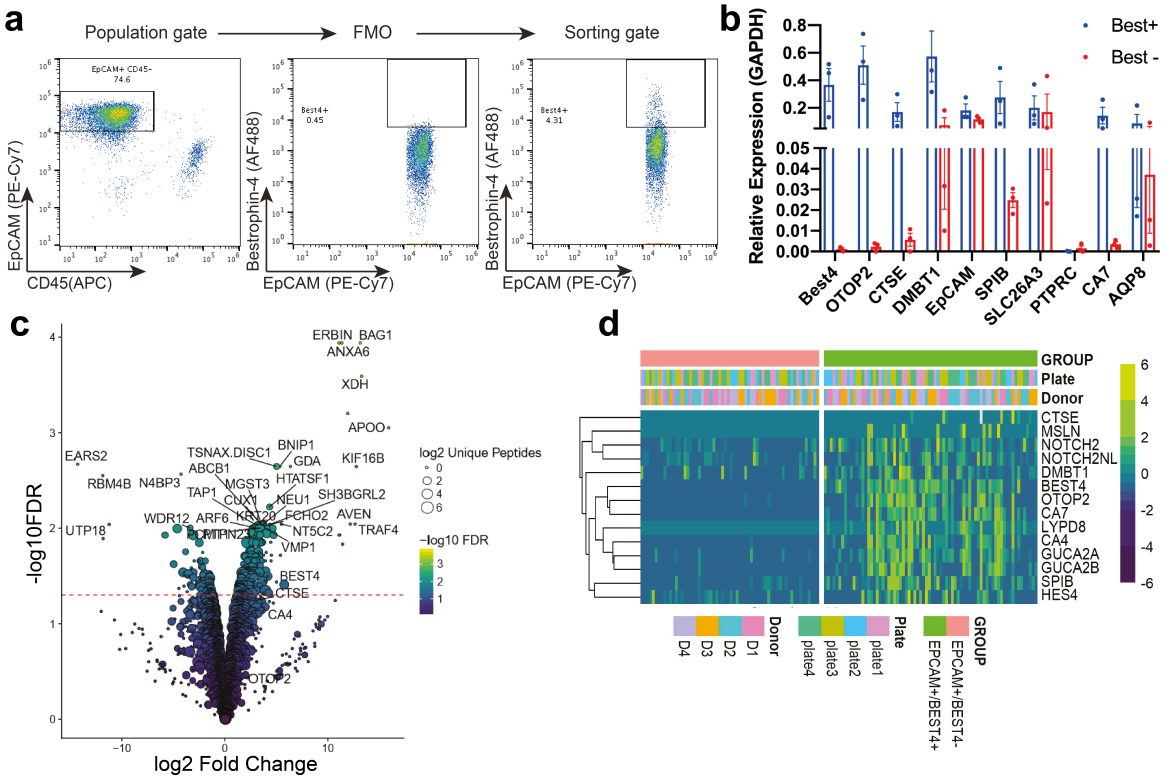

Figure 3.3: Deep characterisation of BEST4/OTOP2 cells

a, Development of a flow cytometry strategy to isolate epithelial (live (DAPI-), single cells, EPCAM+, CD45-, left pane – the population gate) cells which expressed BEST4 (right pane with gate applied by sample specific comparison to fluorescence minus one control (FMO) in centre, validated on n=3 samples by flow cytometry before proceeding to FACS). **b,** one hundred BEST4+/- cells (n=3 individuals) were isolated using FACS and underwent microfluidic qPCR for increased expression of known BEST4/OTOP2 markers (error bars showing SEM). **c**, Volcano plot showing proteins that were differentially expressed between BEST4+/- epithelial cells and analysed using mass spectrometry (n=3 in each group, BEST4+ cells positive fold change and BEST4- negative fold change, red line indicates 5% FDR). **d,** Heat maps showing differentially expressed genes between BEST4+ and BEST4- cells analysed by SSII (n=4 individuals per group, n=94 cells sorted per individual).

3.5 BEST4/OTOP2 cells act as pH sensing cells

A specific marker of the BEST4/OTOP2 cells that was identified throughout the analysis performed was the expression of *OTOP2*. The OTOP family of proteins have been shown to be involved in taste by channelling protons in a pH dependent manner[229].Given that a similar role was identified in GO and pathway analysis we sought to identify it intestinal BEST4/OTOP2 cells expressed a functional proton pump.

Using a method similar to that described in other members of the otoptrin family it was tested whether these cells conducted protons into the cell cytosol in response to lowering extracellular pH (pH0)[229] (**Methods 6.6**). FACS was used to separate BEST4+ve and BEST4-ve cells and then we performed microfluorimetry analysis using a membrane permeant pH indicator, pHrodo Red. Lowering extracellular pH from 7.4 to 5.0 caused an increase in emission of pHrodo Red (**Figure 3.4a)** seen only in BEST4+ve cells, corresponding to a significant change in the intracellular pH (pHi) (**Figure 3.4b**).

Intracellular acidification can be cytotoxic and induce apoptosis; however, our proteomics data suggest that BEST4/OTOP2 cells express high levels of anti-apoptotic protein BAG1 (**Figure 3.3c**) which suggests a potential homeostatic mechanism that allows the cell to tolerate substantial pH changes.

Taken together, these data identified a new absorptive colonic epithelial cell sub-type, the BEST4/OTOP2 cell. After identifying this in initial scRNA-seq a full multi-omic characterisation was performed - 10x scRNA-seq, SSII, proteomics and semi-supervised analysis of publicly available data – that generated a full characterisation of its markers (**Figure 3.4c, Methods 6.11**).

A shared feature of these methods was that the cell type likely specialises in ion transport, metal storage, and pH sensing that is depleted in cancer. It was demonstrated that the cells expressed a functional proton pump which exhibited effect in a pH dependent manner. The placement of these cells at the entrance to the crypt alongside the expression of Uroguanylin coincident with the ability to sense pH suggests these cells play a role in setting colonic epithelial cGMP tone in response to luminal pH.

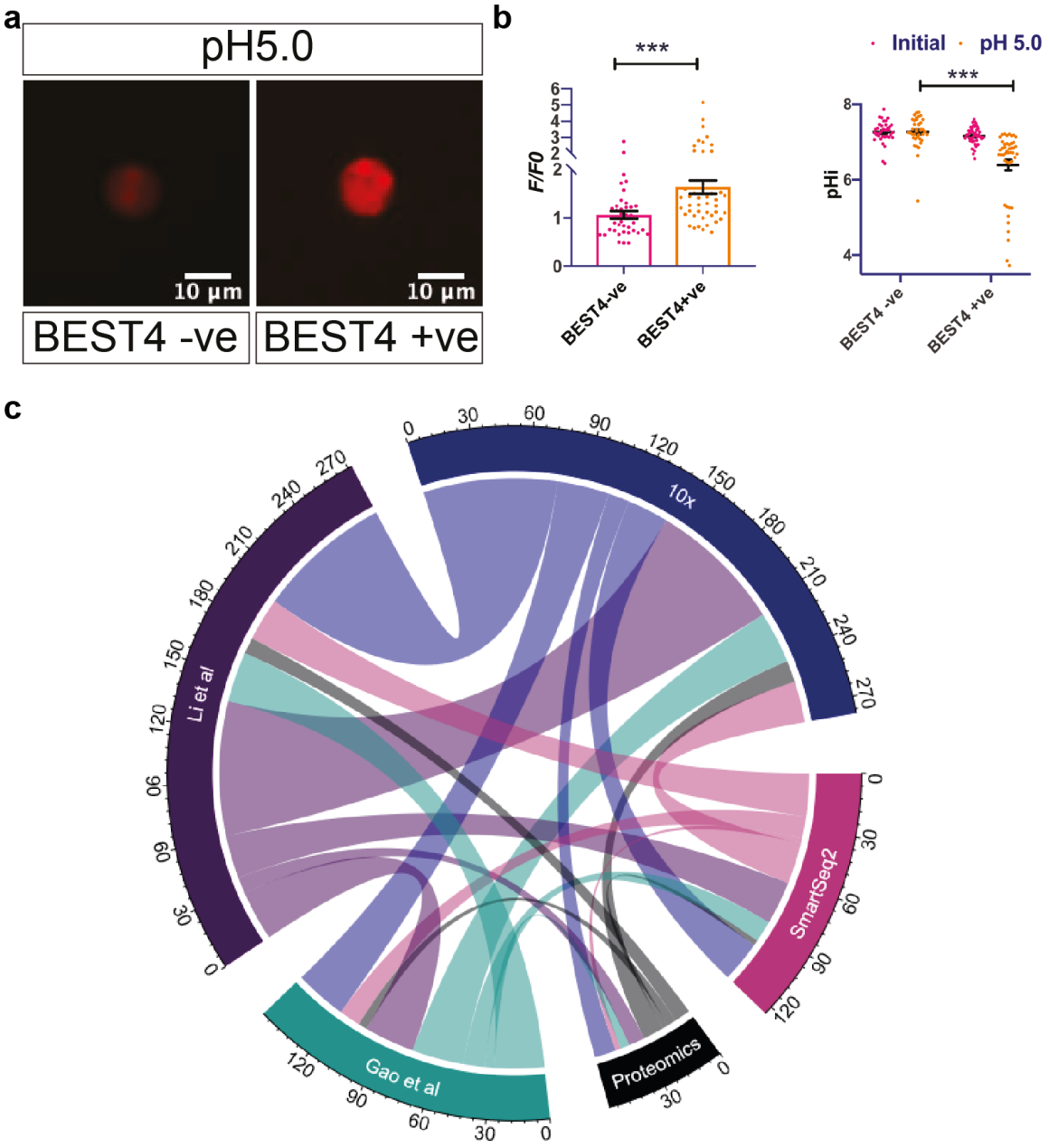

Figure 3.4: BEST4/OTOP2 cells involved in pH sensing

a, Representative images showing intracellular change in pH of BEST4+ and BEST4- epithelial cells exposed to buffer of pH 5.0 **b,** Normalized fluorescence emission (F/F0) from pH indicator pHrodo Red in BEST4− cells (n = 39; pink) and BEST4+ (n = 45; orange) cells (error bars represent mean ± s.e.m, P value <0.001) and intracellular (pHi) for the same populations showing the peak response during each stimulus (two-sided paired t-test, P value 0.9873310; BEST+EPCAM+ P value <0.000001; mean ± s.e.m).**c**, Circos plot showing combined overlap of the top 200 BEST4/OTOP2 makers detected between 10x, smart-seq2, quantitative proteomics and previously published data

3.6 Interim conclusion

Through surveying the diversity of the intestinal epithelium, a potentially novel type of enterocytes was identified in the colon. This shows the potential advantage of using an unbiased approach to glean new information. There is precedence in novel cell types being identified in other recent work[184,185], but this part of the project sought to confirm this possibility in the intestine.

Through combined staining, isolation, deep sequencing and proteomic analysis alongside comparison to public data it was shown that BEST4/OTOP2 cells represent a distinct cell type of the absorptive lineages. They differentially express genes involved in metal ion and electrolyte processing which was confirmed using multiple modalities. This highlighted that the cell type was capable of a pH sensing role and have a functional pH channel.

BEST4/OTOP2 cells therefore may maintain luminal homeostasis, potentially through regulation of the GC-C signalling pathway, of which is it unique in expression of key pathway components such as *GUCA2A* and *GUCA2B* which encode uroguanylin – the endogenous paracrine hormone required for GC-C activation[243]. This pathway is pH dependent for activation and modulates key physiological functions such as fluid and electrolyte balance and changing epithelial proliferation as well as sensing microbial signals[244].

Uroguanylin mimetics have been developed and are sought to act as novel treatments for functional gastrointestinal disorders and IBD[245]. Expression of individual genes that comprise this cell type have been reported in isolation such as for BEST4 or components of the uroguanylin pathway[228,231], but they had not before been linked to one distinct cell type.

It will be an area for further work to see if localisation of what was previously thought to be a generic expression of pathway components, or attributed to a different cell type, can reveal new insights to bowel pathology and homeostasis. In addition, application of our core gene signature to bulk RNA data identified BEST4/OTOP2 signature was also found to be deficient in studies of colorectal cancer.

This highlights that scRNA-seq is able to add a layer of information to previous bulk approaches, and the importance of being able to assign cell type specific effects. Another important question raised by this finding however is how a new cell type is placed in the previous understanding of intestinal epithelial crypt maturation, or indeed when it and other fibroblast sub-types appear[195]. Also, by studying the epithelium in isolation a large number of important interactions across cell compartments will not be appreciated. This forms the basis of the next part of work where it will seek to build on these findings and understand the cellular landscape across the human intestine as it forms *in utero*.

Chapter 4: Human intestinal development across location and time

4.1 Introduction

In work up until now the project has endeavoured to categorise the epithelial barrier in steady state and identify the drivers of breakdown in inflammation. This has been able to identify novel features of goblet cell diversity and WFDC2 as important in epithelial maintenance demonstrated by its loss in UC. This also showed a novel type of enterocyte, the BEST4/OTOP2 cell, that is involved in pH balance.

Together this demonstrates scRNA-seq to be a powerful tool in mapping the epithelium in the context of the mature intestine and its breakdown. However, the epithelium does not function in isolation, and a large number of other compartments will provide support and cross-talk. Recently, studies have shown that the understanding of the interaction between the epithelial, mesenchymal and immune compartments is vital when considering colonic crypt development[89] and maintenance[27]. It has also become clear that colonic maintenance differs from small intestinal in inflammation[90].

This raises the question of how the mature epithelium forms and interacts with other compartments during development. The interpretation of scRNA-seq data can sometimes be challenging across organs, as the tissue-based context of cells are lost during the digestion process. This is especially important when considering cross-talk as intracellular communication will be dependent on spatial context. Recently methods have been developed to map transcriptome based expression in tissue and one, spatial transcriptomics

(ST), can do this in a transcriptome wide manner[196] and this has been shown to be extremely powerful in the setting of human organogenesis[246].

Within intestinal development it is known that at the end of gastrulation, the endoderm undergoes craniocaudal patterning directed by specific transcription factors[80–82]. The posterior endoderm gives rise to the small and large intestines with subsequent folding generating the embryonic gut tube and primitive epithelium[79]. This simple pseudo-stratified epithelium proliferates, lengthens and expands to establish mature villus and crypt morphogenesis with regional variation along the alimentary canal. The tightly-controlled timing of differentiation and co-ordination with other cell type is a complex process requiring strictly regulated cross-talk.

To date, most of our knowledge on intestinal development comes from seminal studies in model organisms such as the chick and mouse[83–85,247]. However, it remains unclear which aspects are directly applicable to human development, as these studies highlight several species-specific mechanisms and key differences such as chick vilification relying on buckling muscle forces whilst murine processes are co-ordinated by morphogen rich "hillocks"[248]. Recent access to human fetal tissue has allowed high-resolution 3D studies of early stage human embryos[77] and along with histology[87] have shed some light on human intestinal development. However, questions remain about how specific cells types appear, localize and interact during human intestinal development and what are the molecular cues that control their fate.

The next part of this work will exploit multiplexed high-throughput scRNA-seq incorporated with ST to create the first large scale single cell spatio-temporal atlas of human intestinal development, charting developmental processes across time, location and compartments. This compiles a multi-modal resource that catalogues cellular diversity,

explores cell-cell signalling, and highlights progenitor origins and locational fate decisions during organogenesis.

4.2 Generating a resource of 101 intestinal cell types across development

To chart the development of the human intestine it was aimed to collect tissue from multiple locations (small intestine and colon), from all cellular compartments (full thickness) and across fetal maturation. Following informed consent tissue was collected from the Human Developmental Biology Resource (HDBR) and two previous protocols were optimised and adapted[2,195] to generate pure, viable single cell populations of epithelial crypt disassociated cells (EPCAM+) and the remaining crypt depleted fraction (EPCAM-) from the same tissue sample (**Figure 4.1a, Methods 6.1**).

scRNA-seq was then performed with use of a combinatorial antibody-oligonucleotide technique previously described in peripheral blood mononuclear cell (PBMCs)[249] in order to increase the throughput and run up to 9 samples per reaction of 10x Genomics scRNA-seq (**Methods 6.1**, **Appendix A.13**). This allowed 77 samples to be collected from 17 individual embryos from multiple locations (small intestine and colon/hindgut) at 8-22 PCW; representing timepoints from prior to crypt formation up to generation of adult-like crypt/villus morphology (**Figure 4.1b, Appendix A.14**)

Together, following deconvolution protocols and stringent QC, this generated a resource of 76,592 high quality single cells across time and location (**Figure 4.1c, Methods 6.11**). Clustering analysis readily identified these as populations across nine intestinal compartments that were annotated on transcriptional signature: epithelial (*EPCAM, FABP1*), fibroblast (*THY1, COL1A1*), endothelial cells (*PECAM1, RAMP2*),

pericytes (*RGS5, KCNJ8*), ENS (*PHOX2B, S100B*), muscularis (*MYH11 TAGLN*) , mesothelium (*MSLN, WT1*), myofibroblast (*PITX1*, *ACTGT2*) and immune (*PTPRC*) (**Figure 4.1d, Appendix A.14**).

Even studying cells at the compartmental level distinct time point and locational differences could be appreciated. The epithelium being divided into a proximal and distal counterpart, and myofibroblasts not appearing as a distinct cluster until after 16PCW (**Figure 4.2e and f**).

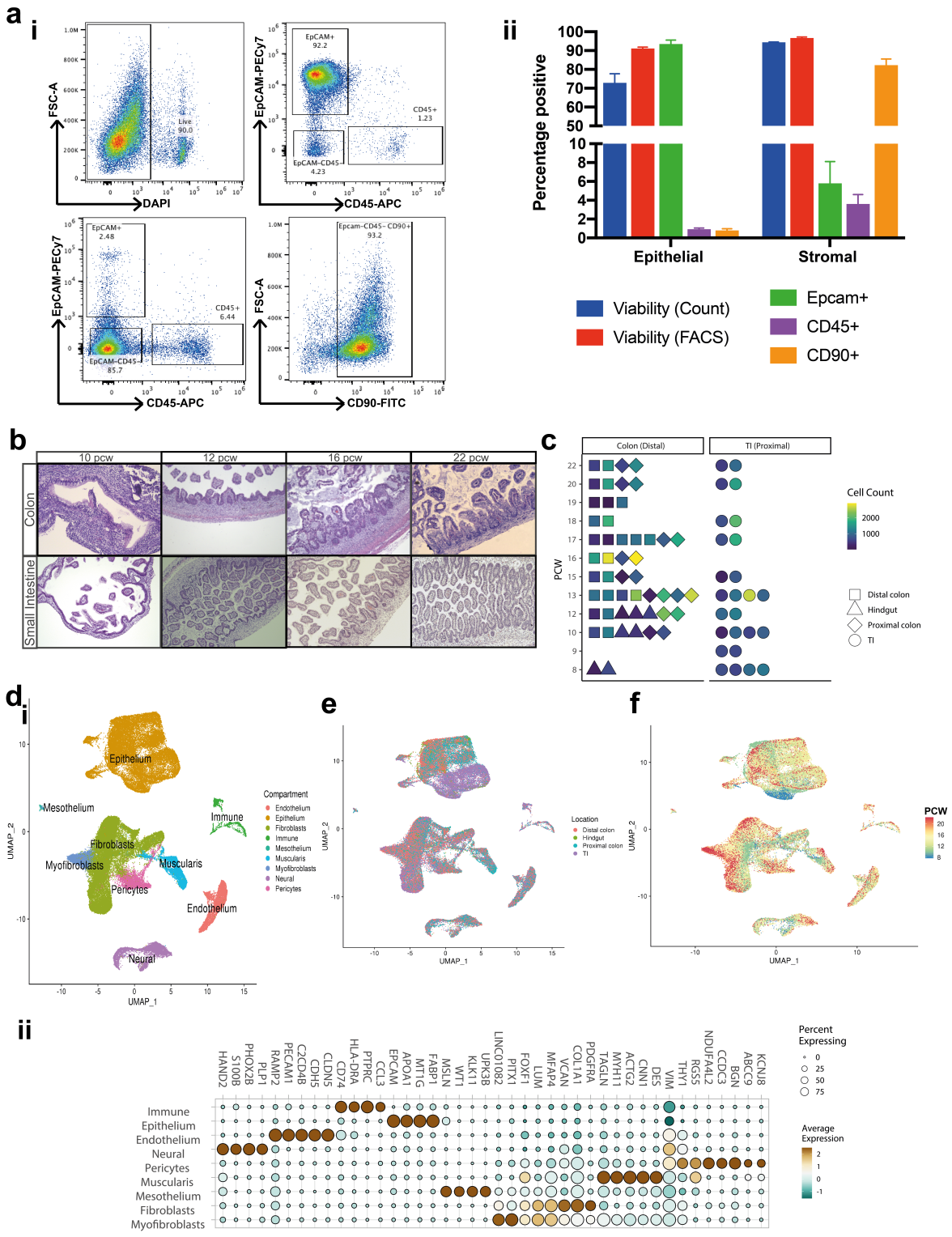

Figure 4.1: Mapping fetal intestinal cell types

a, Validation of digestion technique using flow cytometry with gating strategy (**i**) for measuring viable epithelial (EPCAM+) cells (top panels) and non-epithelial cell types including immune (CD45+) and stromal (EPCAM-CD45-) with quantification (**ii**, n=3 experiments, error bars S.E.M). **b,** H&E stained sections of fetal intestine through timepoints captured in resource showing morphology. **c,** Sample overview of tissues used to generate resource depicting sample location, time-point and cell recovery post QC. **d,** UMAP embedding of cells across 9 compartments with key marker genes shown in heat dot-plot(**ii**) with representative overlays of location (**e**) and gestational age (**f**, PCW)

Further analysis was then performed within each compartment, to appreciate individual cell types and states based on marker genes described within the literature (**Methods 6.10)**. Through this the resource was able to delineate 101 distinct sub-populations (23 epithelial, 16 fibroblast, 12 endothelial, 8 pericyte, 13 neural, 11 muscle, 4 myofibroblast, 2 mesothelial and 12 immune. These were annotated based on key marker genes and literature and their relationships demonstrated using partition based graph abstraction (**Figure 4.2a, Methods 6.10, Appendix A.14**). The largest connected component encompassed diverse mesenchymal cell clusters (fibroblast, muscularis, myofibroblast and pericytes) reflecting their shared supportive roles and high similarity of signature, especially in the early progenitor cells (**Figure 4.2b i-iii)**. While immune, endothelial, epithelial and neuronal lineages formed their own partitions (**Figure 4.2b iv-vi)**. The smallest of the annotated clusters constituted only 21 cells (EEC I-Cell subtype) demonstrating the utility of the approach in capturing comprehensive cellular anatomy such as very rare cell types.

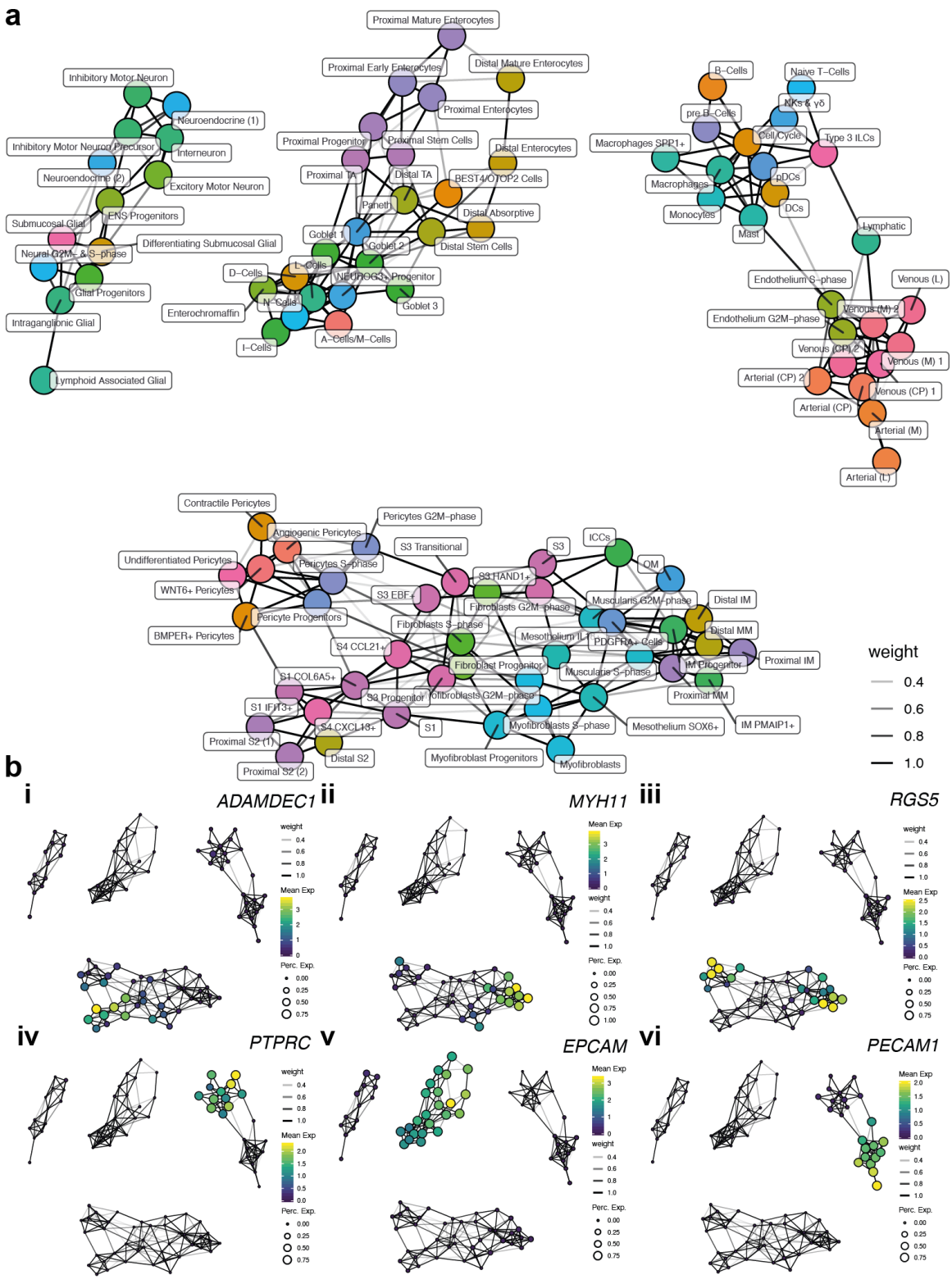

Figure 4.2: Relationship of 101 cell types and states

a, Fine annotation of 101 cell clusters was mapped using partition-based graph abstraction (**Methods 6.11**, line representing weight of interaction) with example genes shown (**b**) to demonstrate fibroblast (**i**), muscle/myofibroblast (**ii**), pericyte (**iii**), immune (**iv**), epithelial (**v**) and endothelial (**vi**) compartments (colour showing relative expression, gene shown top right)

4.3 Spatial localisation of transcriptional hallmarks

To compliment the fine resolution scRNA-seq map, and facilitate mapping of single cell signatures to tissue, we undertook ST from across fetal development and adult intestinal tissue for comparison[196] (**Methods 6.9**, 8 sections of 5 samples; 12PCW, n=5; 19PCW, n=1; adult, n=2). Tissue sections were cut onto an oligo-barcoded slide where they were permeabilised and RNA was reverse transcribed and amplified to undergo sequencing (**Methods 6.9, Appendix A.14**).

This generated a dataset of 9330 tissue-covered spots with an average of 2480 genes detected per spot, and some of the smaller tissues slides sequenced to a higher depth (**Methods 6.11**). Each spot covered a 55μm area with a centre-to-centre distance of 100μm which, upon examination of H&E images appeared to encompass in the region of 6-10 cells depending on location and gestational age.

Initial validation of the technique could be seen by linking gene expression of each spot and confirming it aligned with known morphological features. Examples included *RET*, a marker of enteric neurons, being seen with spot-specific expression aligned to myenteric plexuses visible in the H&E stain (**Figure 4.3ai**). Similarly *PTPRC* which encodes for CD45 localised to immune cells within a submucosal lymphoid follicle (**Figure 4.3 aii**).

Performing dimensionality reduction and clustering on the tissue spots in a manner akin to scRNA-seq was able to identify 5-13 spot clusters in each slide, which were mapped back to spatial locations. Clustering reflected anatomical features, one showing the adult immune follicle and another cluster representing neural genes and linking with myenteric plexuses between the inner and outer muscularis in a late (19PCW) fetal sample (**Figure 4.3b**). Although the resolution wasn't single cell, using the scRNA-seq dataset as

a reference tool along with that generated in adult epithelial above and other published scRNA-seq data[2,195,250,251] it was possible to carry out factor analysis (**Methods 6.11**) and determine the likely single cell composition of each spot. This could localise well characterised cell types such as crypt top colonocytes at the top of the crypt, myofibroblasts in place at the muscularis mucosae and BEST4/OTOP2 cells at the top of the crypt in keeping with previous spatial validations (**Figure 4.3c**).

The ability to infer cell-type content of each spot also enabled the use of pairwise cell type signal correlation analysis to highlight significant same-spot co-occurrence of cell types (**Methods 6.11**). This could show co-localisation of cells in tissue including BEST4/OTOP2 cells occurring with crypt top colonocytes, or goblet cells being placed alongside undifferentiated cell types reflecting their placement deeper within the crypt (**Figure 4.3d**).

It was notable that many of the clusters were arranged in layered ring-like formations, highlighting that the greatest determinant of transcriptional and cell type variability across tissue was depth especially in fetal samples where the entire thickness of intestine could fit on the slide. This could be represented by assigning a depth score from the serosa (in fetal) or muscularis mucosae (in adult). This distance measure identified 2893 significantly (<5% FDR) depth associated genes and GO enrichment defined the molecular pathways activated at different layers of tissue; reflecting deeper structures were more related to muscular and neural functions (GO terms: muscle contraction, axonogenesis), whereas more superficial clusters exhibited more luminal functions (GO terms: microvillus organisation, digestive system processes, **Appendix A.14**). In addition the cell types identified from the pooled scRNA-seq and ST resource could be placed within the depth score, showing sequential enrichment of cell type signatures; mesenchymal cells, myofibroblast and fibroblasts, goblets, colonocytes and crypt top colonocytes in adult

(**Figure 4.3e**). In fetal tissue a similar ordering of cell types by depth was seen for many of the cell types at 12PCW and 19PCW but there were notable exceptions such as inner and outer muscularis signatures exhibiting no depth separation until 19PCW when they would occupy distinct layers, something which was also reflected in the scRNA-seq data of when outer muscularis would appear (**Figure 4.3f**). Together both modalities showed that the muscularis matured in a proximal to distal manner, forming both layers in the latter part of our sampling window.

Thus, the combination of both scRNA-seq and ST was able to place cell types and their relationships into tissue, increasing the utility of identifying functional roles to the 101 cell types and states.

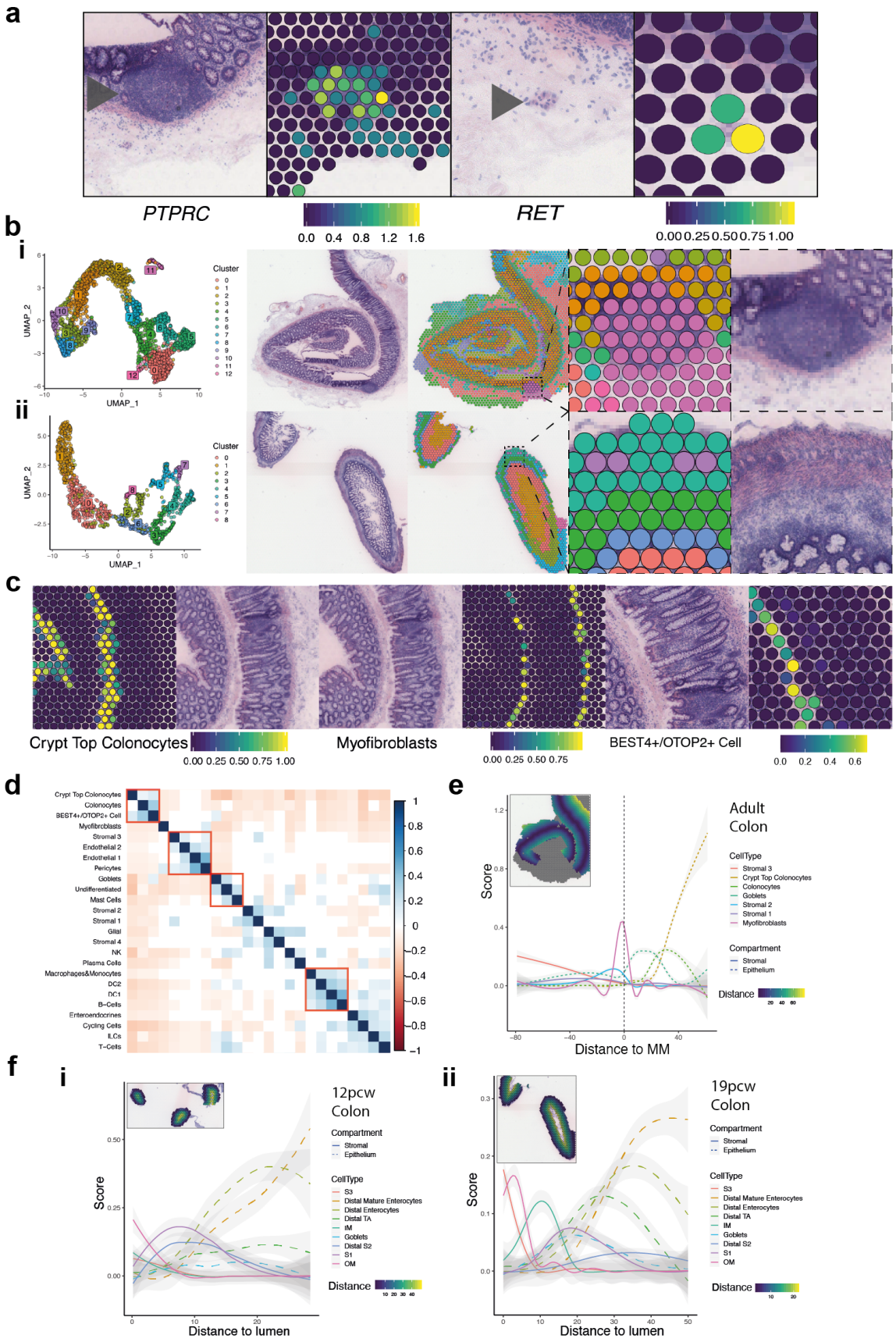

Figure 4.3: Spatial mapping of fetal development with ST and scRNA-seq integration

a, Validation that ST is able to identify histological landmarks matching expression (overlay of expression in ST arrays, colour corresponding to relative gene expression) of key marker genes for immune follicle (*PTPRC*) and for neural marker (*RET*) with corresponding H&E showing anatomical features. **b,** UMAP plots showing clustering of two example slides, one adult (**i**) and one fetal (**ii**) showing tissue, and location of clusters, then also an example higher resolution area of cluster overlaying an anatomical area. **c,** Examples of cell type localisation into tissue, with cell-type spot expression viewed alongside tissue, localising crypt top colonocytes to top of crypt, myofibroblasts to muscularis mucosae and BEST4/OTOP2 cells to crypt top. **d,** Pairwise cell type prediction signal correlation heatmap in adult ST spots. (Non-significant correlations (<0.05 adjusted p value) are shown in white; colour bar indicates Pearson's r value. Red boxes highlight selected biologically relevant correlation groups.) **e**, Selected cell type prediction distribution over distance/depth score (inset and legend: ST slide overlay showing distance measures from muscularis mucosa used to assign each spot a distance gradient coloured by depth score overlaid over H&E image **f**, as in (**e**) with fetal tissue showing sequential cell type expression over depth at 12PCW (**i**) and 19PCW (**ii**)

4.4 Human intestinal epithelial and stem cell development

The formation of the epithelial crypts and villi *in utero* from a flat surface represents the establishment of life long circuits which will control epithelial maturation and barrier function. Our resource included a total of 17,622 epithelial cells to study these events, that were discriminated on their transcriptional profiles and divided into 15 clusters with clear lineage separation; absorptive (8 enterocyte populations), secretory (EECs, goblet cells and secretory progenitors), undifferentiated (distal and proximal TA cells) and ISCs (proximal and distal) (**Figure 4.4a, Appendix A.14**).

Enterocytes represented a spectrum of maturity in keeping with the previous description of colonic adult epithelia and in line with that reported in the small intestine[252] with division into early enterocytes, enterocytes and mature enterocytes. There was also a clear distinction between proximal/small intestine and distal/colonic epithelial cells, by genes such as *CCL25, APOE* and *FXYD3* with these determinants being identifiable in even the earliest collected samples prior to visible locational features in tissue (**Figure 4.4b, Appendix A.14**).

The proportions of epithelial types exhibited significant changes throughout developmental time, reflecting the ongoing remodelling. Very few mature absorptive and

secretory lineages were seen prior to 12PCW, when TA and progenitor cells dominated (**Figure 4.4c, Appendix A.14**).

ISCs gradually increased in frequency over developmental time in both proximal and distal intestine, but even in late development the frequency was much greater than that seen in adult tissue even though morphology appeared much more akin to mature epithelium. Up to 25% of IECs in some samples after 19PCW expressed markers in keeping with ISCs (**Figure 4.4d**). It was striking that even in the earliest samples, the ISCs expressed clear locational differences reflected by their transcriptional profiles. Colonic ISCs were marked by specific expression of *BMP7* and *LEFTY1* whilst proximal ISCs expressed *DMBT1* and the transcription factor *FOXD1* (**Figure 4.4e, Appendix A.14**). Receptor : Ligand (R:L) analysis was undertaken to investigate the relationships of cells and the pathways through with they may interact based on curated cell interactome databases (**Methods 6.11**). In the earlier time-points a large proportion of neural cells were present, and as the ENS migrates early it could feasibly interact with the developing epithelium. This method highlighted potential interactions such as colonic ISCs interacting with neural cells, that were also present at this time, through *LEFTY1* and the Activin A receptor (*ACVR2B*), both of which had high specificity for their respective cell types (**Figure 4.4f**). *LEFTY1* is an important developmental gene in deciding anatomical asymmetry, so this link potentially suggests how the ENS may also interact at with these decision in early developmental time[253].

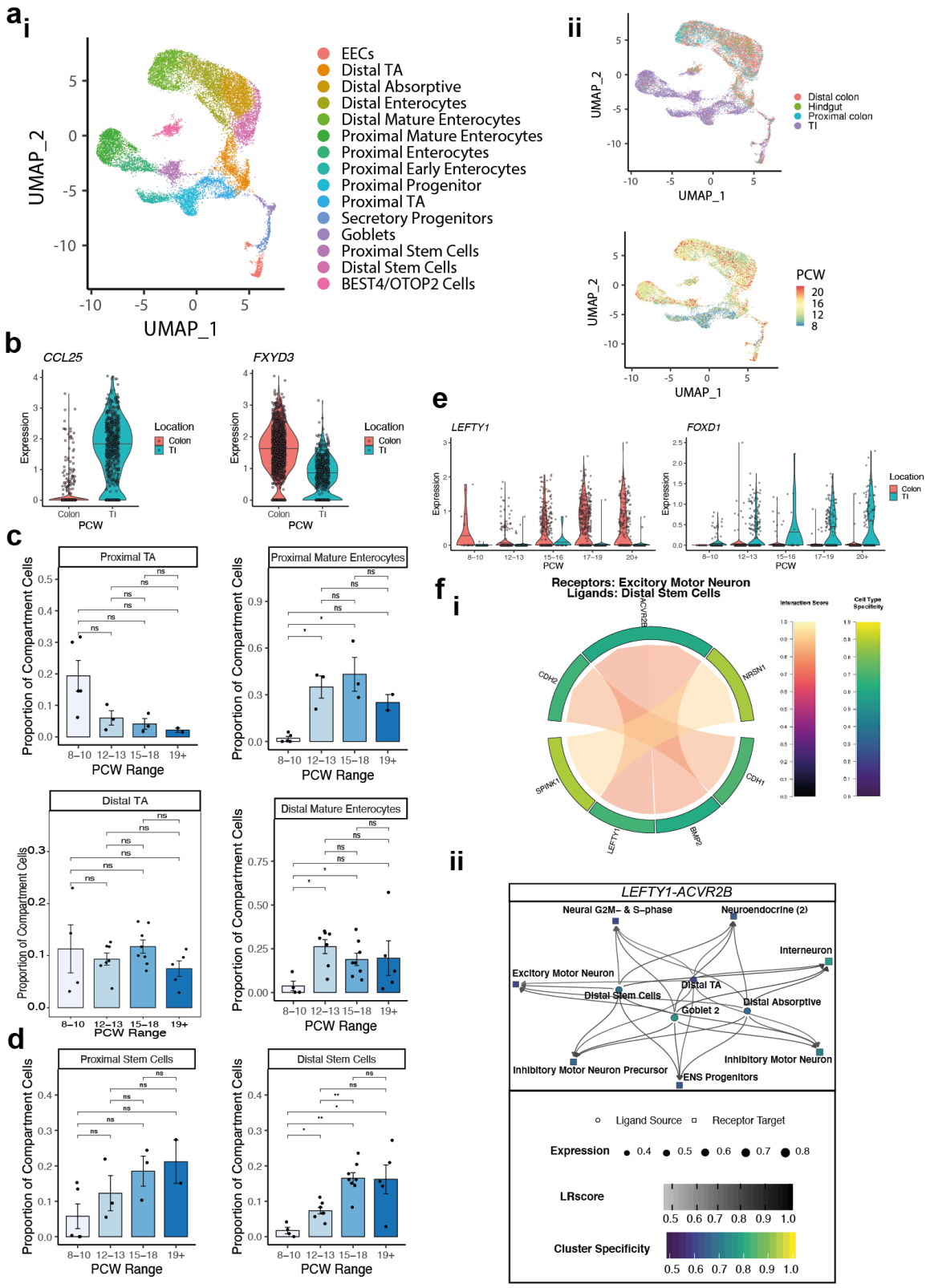

Figure 4.4: Epithelial generation and localisation in human development

a, UMAP plot showing visualisation of expression profiles of all cells epithelial captured in scRNA-seq (**i**) with cell type annotations (**Appendix A.14**) and distribution based on location (**ii**) and time point (**iii**). **b,** Examples of differential gene expression (y-axis, centre median) based on intestinal location (colour). **c,** Selected population abundance of epithelial cell types shown over time (bar plots, Wilcox rank test,

*=p<0.05; ** <0.01; ***=<0.001) for TA and enterocytes, along with location ISCs (**d**). **e,** Violin plots showing expression of selected genes within stem cell clusters only (divided by time point, coloured by location). **f,** Example of inter-related R:L pathways highlighted at a cellular cluster level (**i,** circos plot visualising putative cross-talk between clusters) and within a specific pathway (**ii**) between motor neurons and distal ISCs.

Studying the trajectory of epithelial cell maturation *in silico* (**Methods 6.11**) showed a clear divide between proximal and distal epithelium with fate decisions made early in development (**Figure 4.5a)**. Direct comparison between the TI and colonic intestinal stem cells highlighted transcriptional differences such as *CCL25* and *DMBT1* (proximal) or *FXYD3* and *LEFTY1* (distal) (**Figure 4.5b, Appendix A.14)**. These could be observed even before 12PCW showing fate decision preceded morphological changes and when the mature crypt will have formed.

When further comparing the primitive ISC-like populations in early development (<12PCW) it was identified that there was a unique epithelial proximal stem-like cell population that did not have a distal counterpart. These were present only when the epithelium would be pseudostratified and exhibited many primitive features and expressed genes such as *VTN* which has been shown to be important in vilification in animal models and that is important in mesodermal differentiation[254] (**Figure 4.1d**). Direct comparison between this and *de facto* ISCs demonstrated this population differed in that it had lower expression of core stem cell marker *LGR5* and differentially expressed genes such as *ONECUT2* which is necessary for normal epithelial development (**Figure 4.5c, Appendix A.14**). Furthermore this population exhibited time dependent and unique expression of the transcription factor *GATA4* and *TF* which was validated with IHC (**Figure 4.5d**). *Gata4* is known to be an early endodermal regulator in mice and *TF* is involved in iron transportation and uptake[255]. In animal models iron has been shown to be important in the formation of villi with deficiency leading to reduced crypt length, so this may represent a

population of cells that is actively involved in the elongation of villi in proximal tissue that is superseded as the mature morphology and villi shape is established[256,257].

ISCs have been well described in the setting of the adult crypt, but when they are established in human embryogenesis is unknown. *LGR5* was detected diffusely at very low levels in early gestation (<12 PCW) across ISCs and progenitors in both proximal and distal epithelium. This finding was confirmed using smISH showing low and diffuse expression of *LGR5* at 10 PCW which was localised to the crypt base as the crypts formed (**Figure 4.5e**). This is analogous to behaviour of *Lgr5* reported in chick and mouse development but has not previously been seen in humans[248].

Even after crypt morphology was established (e.g. post-19 PCW) the proportion of stem cells constituted a mean of 18-22% of all captured epithelial cells in distal and proximal samples respectively (**Figure 4.4d**), much higher than the reported 3-4% in adult colon with similar methodology[2,251] (**Figure 2.4**). This may help explain why even though morphology appears mature in later samples, extremely premature infants are at high risk of conditions such as NEC, as the crypt still has ongoing remodelling of the stem cell axis[88].

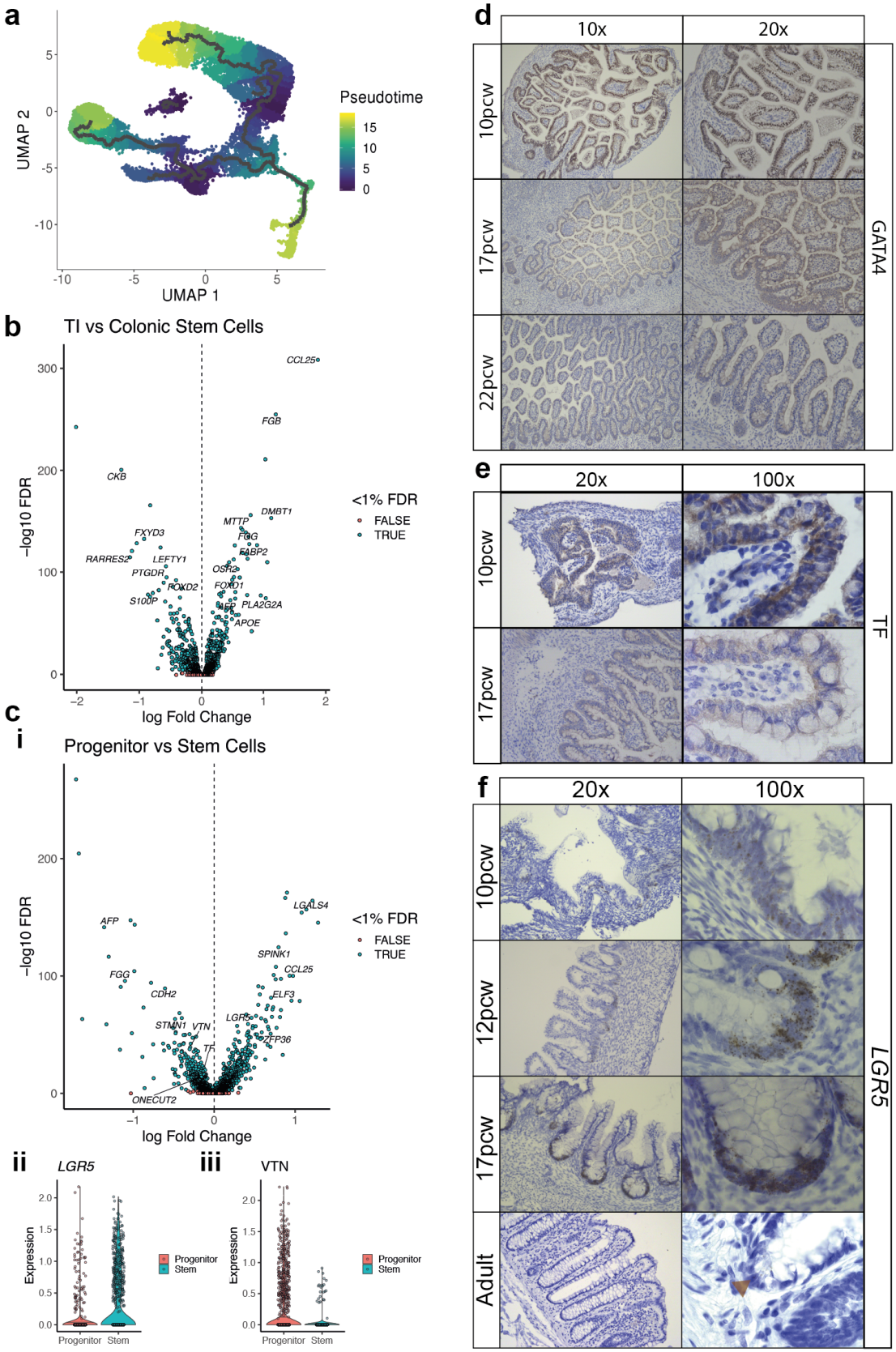

Figure 4.5: Identification of a proximal progenitor ISC-like population

a, Trajectory analysis of epithelial cells using Monocle algorithm shown over a UMAP embedding with colour representing pseudo-time. **b-c,** volcano plots showing differentially expressed genes between colonic and TI stem cells (**b**) or between stem cells and stem-like progenitor cells (**c**) with selected genes labelled or, in the latter comparison *LGR5* and *VTN* shown by violin plot (**c ii-iii**). **d – e**, IHC of small intestinal tissue across developmental time stained for the transcription factor GATA4 and protein Transferrin (TF) (gestational age and magnification highlighted, images representative of >3 individual stains within +-1PCW of label). **F**, smISH for *LGR5* in fetal colonic and adult tissue demonstrating the transition from diffuse early (10PCW) expression to focal expression within crypts, and relative higher expression in late gestation compared to adult (each image representative of n=3 stains, arrow in adult to highlight area of expression).

4.5 Compartmental cross-talk prefaces morphology

The maturation of the developing epithelium could be seen throughout the timepoints included in the joint scRNA-seq and ST atlas. The secretory lineage could be further sub-divided into 11 distinct clusters based on expression of goblet markers, EEC subtype markers (A/M, D, and Enterochromaffin/I/L/N-cells) and the SI specific Paneth cells alongside a *NEUROG3*+ progenitor population[189] (**Figure 4.6a, Methods 6.10**). Early in development (8-10PCW) very few goblet cells were observed, but there was presence of I- and A- cells alongside Paneth cells in the proximal intestine (**Figure 4.6a)**.

R:L analysis was able to highlight some of the primitive circuits through which these early time-point cells communicate, and how core EEC signalling circuits interacted with other compartments. For instance, Enterochromaffin cells, which were present mainly in the colon from 12 PCW (**Figure 4.6a**) interacted with inhibitory motor neurons through a number of pathways and most specifically through *TPH1-HTR2B*, a known serotonin signalling pathway in adult appetite and motility[258] (**Figure 4.6b**). This highlights EEC diversity and their related neural networks are established early in human intestinal development.

Strikingly, the previously identified BEST4+/OTOP2+ cells were already detected as transcriptionally distinct clusters in the earliest time points prior to crypt formation, which was confirmed at tissue level with IHC (**Figure 4.6c**). Their frequency did not increase

with developmental time course, in sharp contrast to other maturing secretory and absorptive lineage cells, suggesting that their development may be uncoupled from normal crypt-villus circuitry. Loss of the uroguanylin pathway in development, a hallmark of BEST4+/OTOP2 cells, gives rise to intestinal crypt hyperplasia[259,260]. These findings highlight that BEST4+/OTOP2+ cells are present at a timepoint where crucial mechanisms are being undertaken as part normal intestinal development, and so it is possible that they may have a role in these. It will be of future interest to investigate what necessitates their presence much earlier than other differentiated epithelial cells, and how they are interacting at this time. Interestingly, querying our R:L database predicted a strong interaction between BEST4/OTOP2 and neuronal cells, both of which are established early in development. For instance, inhibitory motor neurons showed highly specific expression of neurotransmitter VIP (ligand), which mediates intestinal epithelial secretion through the receptor VIPR1 (receptor), expressed exclusively by the BEST4/OTOP2 cells showing how different compartments interact at early timepoints(**Figure 4.6d-e**).

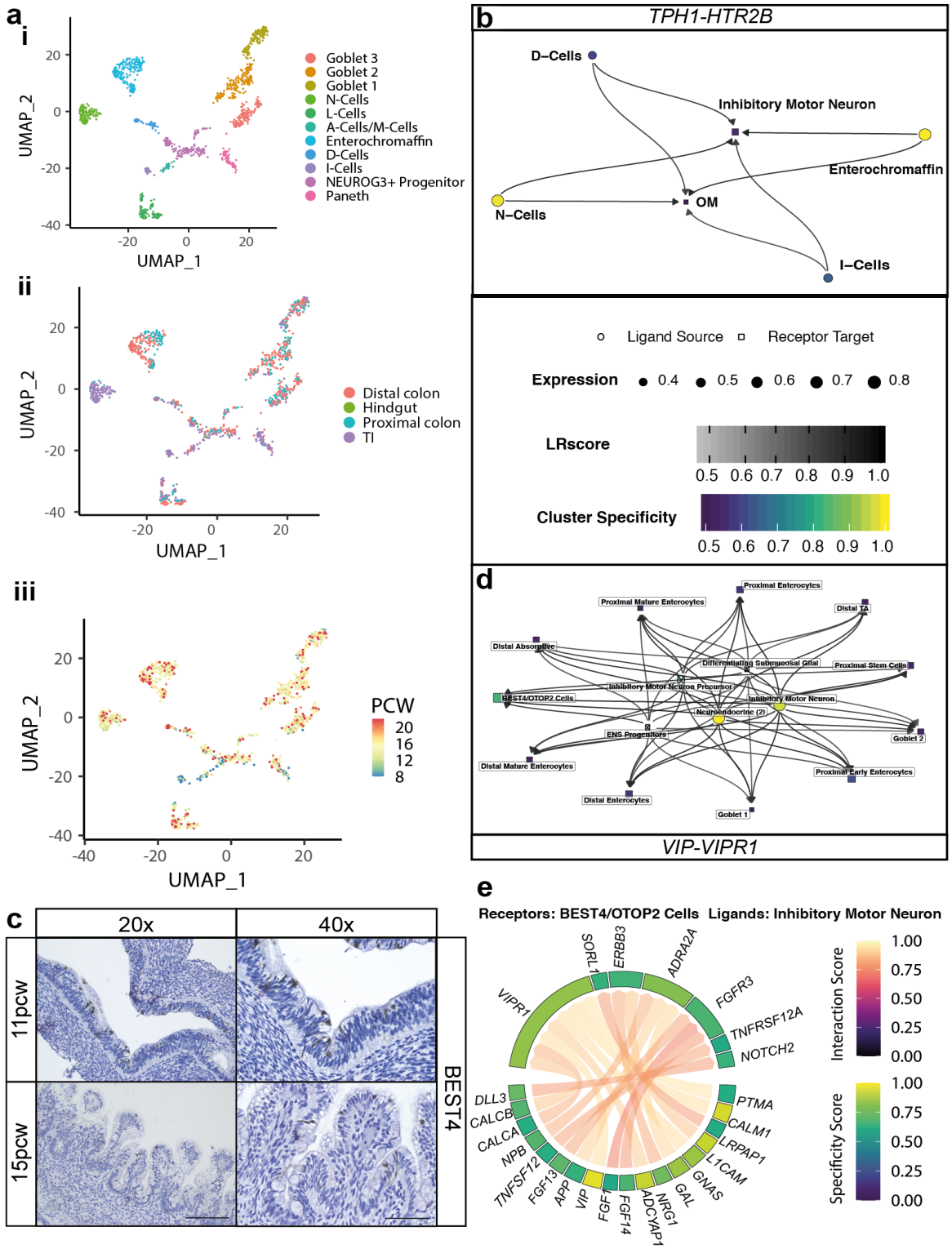

Figure 4.6: Early routes of epithelial cell cross-talk

a, UMAP embedding of secretory epithelial cell sub-types demonstrating goblet and EEC diversity with comparison over location (**ii**) and time-point (**iii**). **b,** interacting cell network plot showing cell type cross talk via specified R:L analysis for *TPH1-HT2RB*. (Ligand source in circle, receptor target is square, node colour specificity and size indicates percentage of cells expressing). **c,** IHC of BEST4 protein in colonic samples in early and late development demonstrating BEST4+/OTOP2+ cells present before crypt formation (n>3 individual stains of time points +/- 1PCW). **d-e**, R:L plot of pathway *VIP1-VIPR1* with key as per (**b**) and circos plot visualising R:L interactions between cells types.

In non-epithelial compartments it could also be seen that transcriptional signatures often correlated with structures as they formed – such as in ECs where expression of genes including *HEY1* and *SOX13* divided populations into arterial, venous and lymphatic components (**Figure 4.7a, Methods 6.10, Appendix A.14**). Although large vessel EC's increased throughout developmental time, they were also identified in the earliest time points reflecting compartment maturity (**Figure 4.7b**). This contrasted to pericytes which had a primitive angiogenic sub-type early in time expressing drivers of vessel formation *PRRX1, THBS4* and *ANGPT2* – the latter being visualised in developing vessels – reflecting differing maturation points in ECs and pericytes despite shared functional roles in angiogenesis (**Figure 4.7c, Methods 6.10**). Moreover, a number of structural cell types had close relationships when studying graph abstraction (**Figure 4.2a**), and this could be seen with shared transcriptional features and trajectory analysis between ECs, pericytes, fibroblasts and primitive *ACTA2*+ muscularis cells (**Figure 4.7d**). This suggested an early fibroblast-to-pericyte transition followed by a subsequent wave of pericyte differentiation. Myofibroblasts, the eventual result of the computed trajectory, were not seen until after 16PCW (**Figure 4.7biii**), which contrasted to the ENS compartment which had 12 clusters of cells (5 glial and 7 neuron) that were present by 7-8PCW, in line with previous literature[148] (**Figure 4.7e, Appendix A.14**). There was however closely related glial and neuronal progenitor populations, expressing genes such as *PHOX2B* and *DLL3/4* respectively leading to their annotation[261,262] (**Methods 6.10**), which did vanish across developmental time (**Figure 4.7b iv-v).**

Together this shows how the different compartments exhibited differential spectrums of maturity within the window captured in the atlas, with ENS progenitors being

lost as EC and later myofibroblasts appear, and also how the interactions between these compartments could be seen within the data.

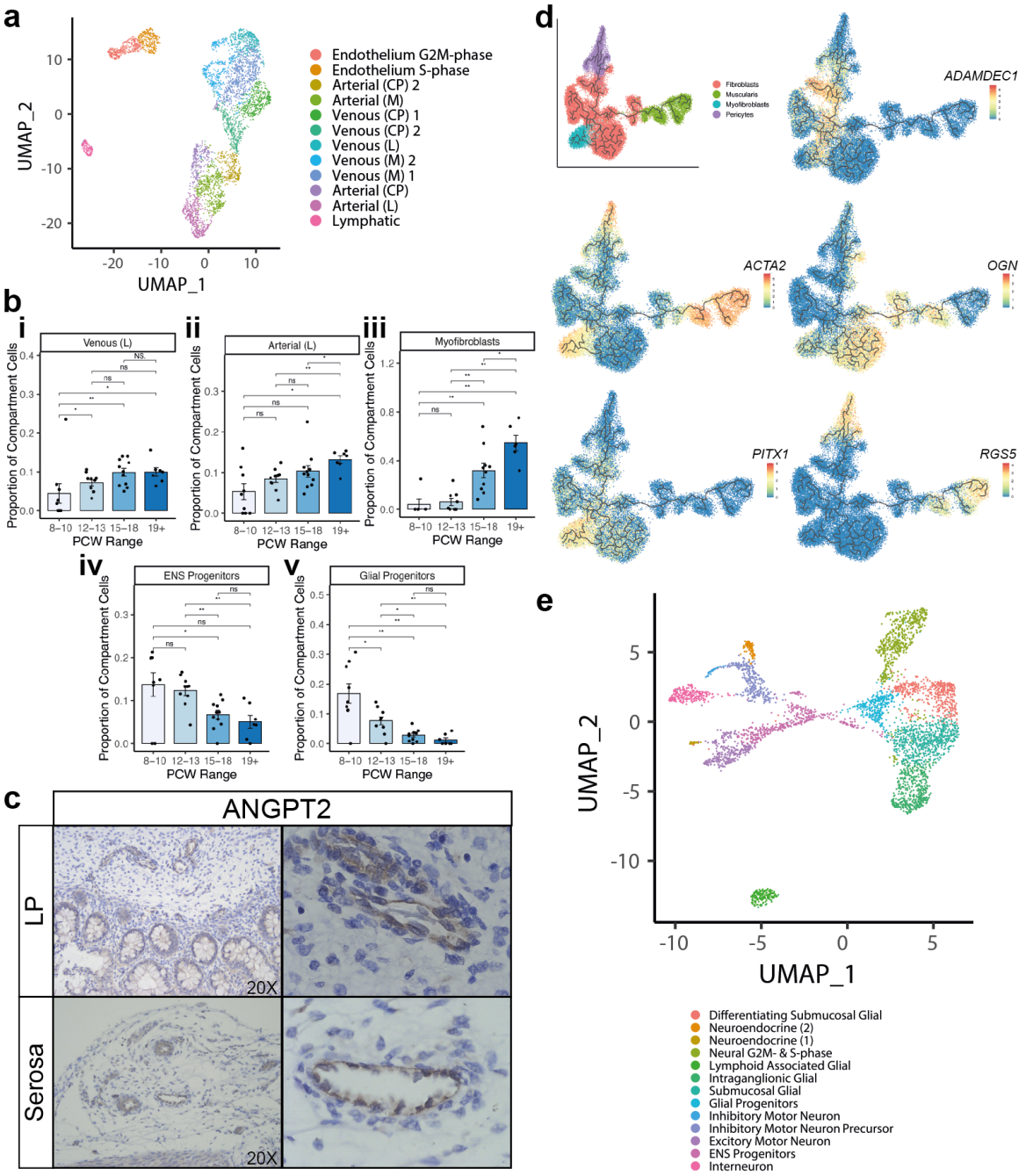

Figure 4.7: Maturation of non-epithelial compartments in development

a, UMAP embedding of EC cells showing clusters representing arterial, venous and lymphatic types along with size related expression patterns (**Methods 6.10**). **b,** Bar charts showing changes in cell type population abundance, as a proportion of compartment cells, over developmental time (p-value: <0.05 *, <0.01 **, <0.001 ***. Error bar representing standard error of mean). **c,** Representative images of IHC staining of late (>19PCW) colonic sections stained for ANGPT2 (brown) and localised to lamina propria(LP) and serosal vessels. **d,** Trajectory analysis of shared structural cells to highlight origins, with differentiation of cell type marker genes. **e,** UMAP embedding of ENS clusters with differentiation into glial and neuronal cell types along with an early progenitor of each.

4.6 Fibroblast diversity and context in development

Use of a full thickness digestion approach was able to identify the differential maturation of cell types, which reflected ongoing processes such as angiogenesis and neurogenesis. But the structural cell type most closely related to the forming epithelium are fibroblast cells. Previous work in the adult colon in health and in UC had shown that rather than being a homogenous cell type, intestinal fibroblasts exhibited heterogeneity and plasticity in relation to spatial location and disease state[195]. This showed four fibroblast sub-types, termed S1-4, with S4 being expanded in inflammation. Similar high resolution techniques have been used in other organs and diseases to support this in the intestine[27,263,264]. However, questions remain about the cues that drive this differentiation along with the role and spatial localisation of these sub-types, and so the resource was utilised to explore this.

The scRNA-seq resource captured 24,081 fibroblast cells that were identified by expression of key markers such as *THY1* and *VIM* (**Appendix A.14**). These divided into 16 clusters which were annotated in line with the S1-S4 types reported in adulthood[195] along with further distinct sub-types with locational, temporal or cycling features (**Figure 4.8a, Methods 6.10**).

As mentioned previously a key process being undertaken in the non-epithelial compartments was angiogenesis, with distinct EC types seen along with their staggered maturation and R:L interaction with pericytes. Studying the ST data which placed depth scores alongside cell types it was apparent that the S3 fibroblast cells lay deep in the developing intestine and in close proximity to ECs (**Figure 4.3d-f**). In contrast to adult data, S3-like cells comprised of a major proportion of fibroblast cells in the developing

intestine. Moreover, some of the S3-like clusters had high expression of primitive markers such as *HAND1, DLK1* and *SFRP1* along with the adult hallmark genes of S3 (**Figure 4.8b**).

Trajectory analysis suggested early S3 populations, such as S3 progenitors and S3 transitional, were the possible starting point of fibroblast maturation, a feature supported by the progenitor population being most frequent before 13PCW (**Figure 4.8c-d**). Studying the adult ST slides, where morphology was more readily appreciated, it was noted that the hallmark genes of S3 – such as *C7* – were seen around vascular structures and this matched with cell-type spot deconvolution (**Figure 4.8e**). S3 often co-occurred with endothelial cells in the depth score (**Figure 4.3d**), and R:L showed a number of cross-talk interactions between these cell types including *LRP1* and *HSPG2* with the former reported as a key regulator of plasminogen activation in fibroblasts[265] (**Figure 4.8f**). Thus S3 appeared to be a primitive vessel-associated fibroblast sub-type involved in angiogenic processes in the early intestine, a role that was not previously appreciated in adult data due to the absence of spatial integration and possible the lack of depth in endoscopic biopsy sampling.

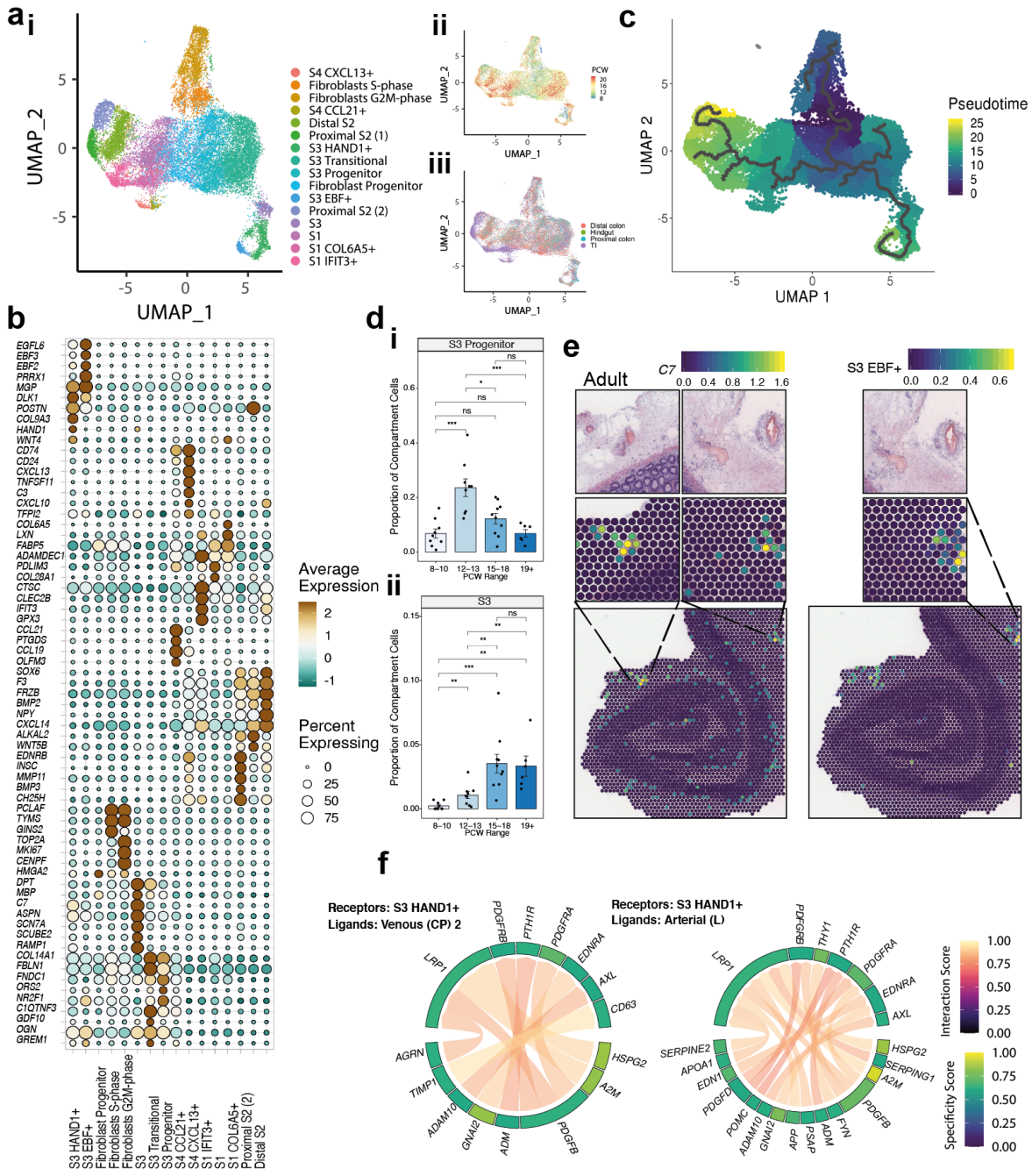

Figure 4.8: Fibroblast diversity, functions and cellular interactions

a, UMAP embedding of fibroblast cells from scRNA-seq with 16 clusters annotated (**Methods 6.10**) and also presented by maturation (PCW, **ii**) and location (**iii**) with heatmap of highly expressed genes in each cluster (**b**). **c,** Trajectory analysis of fibroblast cells using Monocle algorithm shown over a UMAP embedding with colour representing pseudo-time. **d,** Bar charts showing changes in cell type population abundance, as a proportion of compartment cells, over developmental time (p-value: <0.05 *, <0.01 **, <0.001 ***. Error bar representing standard error of mean). **e,** Co-localization of S3 marker *C7* with large vessels in adult ST slide (**i**) and localization of cell-type deconvolution of S3 EBF+ (**ii**) showing gene and cell expression in areas surrounding vessels in adult slides. **f,** Circos plot visualizing putative cross-talk between S3+ HAND1+ cells and Arterial and Venous endothelial cells.

Another sub-type of fibroblast described in the adult was the S4 type which was expanded in colonic inflammation of UC[195]. It was surprising to see that cells with the markers of this cell type - *CCL19, C3* and *CD74* (**Appendix A.14**) – were identified in the healthy fetal intestine outside of an inflammatory context. *CCL19*, *CCL21* and *CXCL13* are reported to be key signalling mechanisms through which immune cells can congregate to forming lymphoid tissue in the developing murine intestine[136] and these were all specifically expressed in S4 cells, with expression being absent before 10PCW and increasing during development (**Figure 4.9a, Appendix A.14).** S4 types were also present predominantly in the small intestinal tissue (**Figure 4.8a**). The two sub-types of S4 cells seen also appeared to have transcriptional differences in line with a depth dependent variants of the fibroblast cells. S4 *CXCL13*+ exhibiting higher relative expression of *F3*, *DLL1* and *PDGFRA* which are all peri-cryptal S2 telocyte genes when direct comparisons between S4 types were made (**Figure 4.9b, Appendix A.14**).

Alongside S4, another compartment of cells that were also much more prevalent in the small intestine were immune cells. In this regard we captured 2,199 immune cells across development (*PTPRC*+) that were divided into 12 clusters across all 6 immune lineages (macrophage, monocyte, DCs), eosinophils, adaptive and ILCs) (**Figure 4.9c**). There was a paucity of immune cells before 10PCW, at which time the majority were myeloid cells, but the influx of proximal immune cell colonisation could be seen at 12PCW reflected by appearance of CD4+ and CD8+ cells (**Figure 4.9c**). This feature was in keeping with literature that immune cells and GALT appear *in utero* in the developing intestine alongside the development of proximal lymphoid tissue such as PPs, whereas colonic lymphoid structures mature after birth[136].

Three clusters of immune cells expressed ILC related genes (*ETS1, TOX*)[266] (**Appendix A.14**) but one specifically had high expression of *IL7RA* and *ID2* which are

both hallmarks of Type 3 ILCs, specifically those described to be lymphoid tissue inducer (LTi) cells in the developing gut which organise the recruitment of immune cells and formation of GALT[136,267] (**Appendix A.14**). In murine models the interactions of LTi cells that have been shown to be key to PP formation have been with ECs, fibroblast cells and neural cells expressing *GFRA3* – which represented lymphoid associated glial cells in our resource[268] (**Figure 4.7e, Appendix A.14**). Utilising R:L interaction analysis we could pick specific reported pathways of *in utero* GALT formation, that if absent result in lack of intestinal lymphoid tissue in mice: *IL7-IL7R, CCL21-CCR7, CCL19-CCR7*[269] (**Figure 4.9e**). All these interactions reflected pathways involving S4 sub-types with cell types previously described in murine models, thus highlighting S4 as a keystone of human GALT formation.

The interaction of S4 with immune cells was validated in the adult ST, where tissue size allowed visualisation of submucosal lymphoid follicles. Around these the expression of hallmark S4 genes (*CCL19*); their R:L interactions highlighted in scRNA-seq data (*CCR7-CCL19*) and the cell-type deconvolution of S4, T-cells and B-cells could be visualised within tissue demonstrating the application of a combined multi-omic approach (**Figure 4.9f**).

Overall S4 was shown to be a keystone of *in utero* proximal intestinal lymphoid formation and the resource was able to model interactions and mechanisms of formation. The resource was thus able to resolve the role of S3 and S4, with S1 being the core structural cells. The remaining sub-type that remained worthy of more focused study in relation to the epithelium was peri-cryptal S2, which in initial census of colonic fibroblasts was rich in morphogens, thus raising questions of how it supported the forming epithelium in the developing intestine.

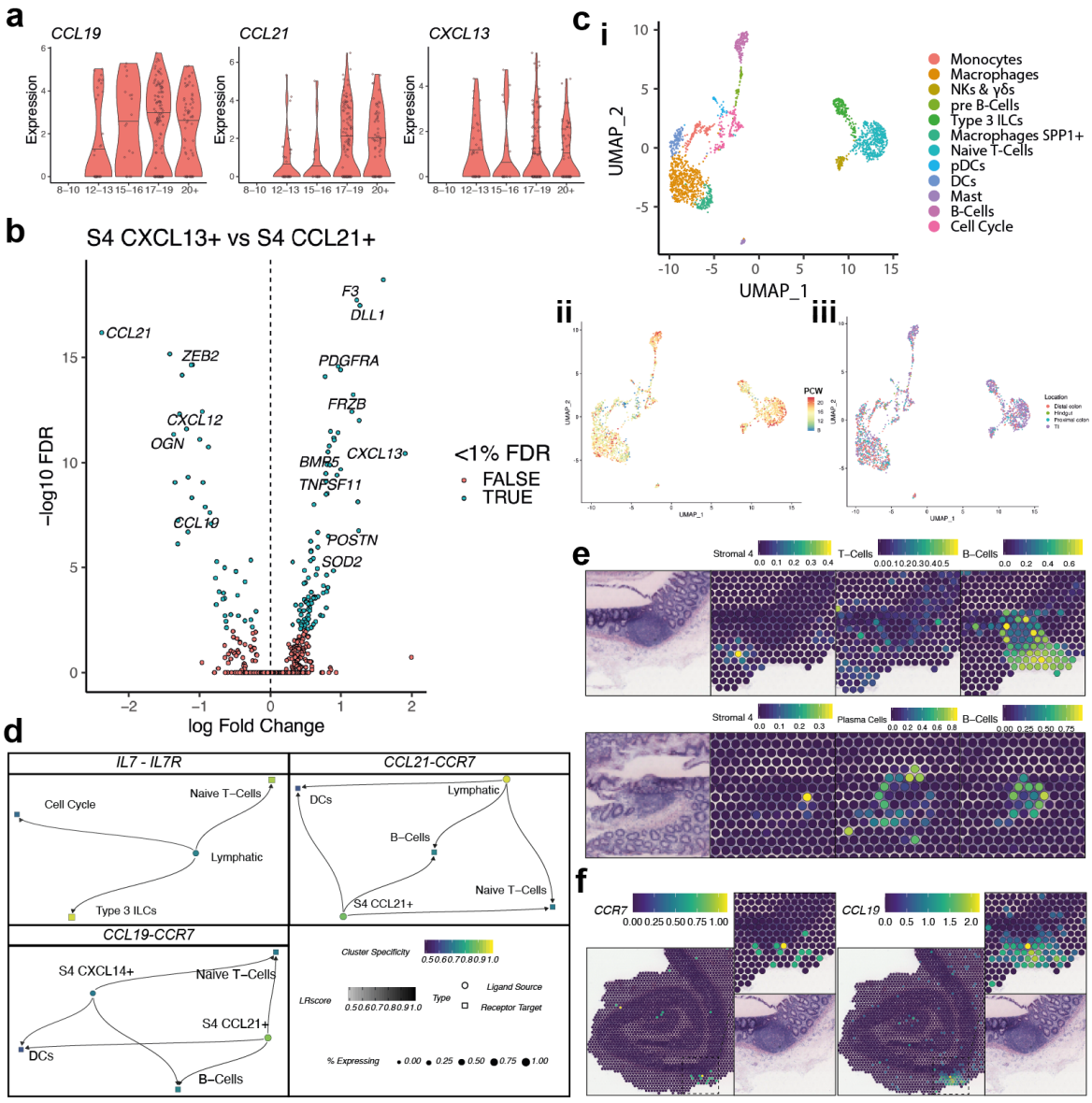

Figure 4.9: S4 fibroblasts are a keystone of fetal immune lymphoid formation

a, Violin plots showing expression of *CCL19*, *CCL21* and *CXCL13*, known mediators of GALT formation, in S4 fibroblast cells over developmental time course. **b,** Volcano plot visualizing differentially expressed genes between S4 CXCL13+ and S4 CCL21+ cell subpopulations. **c,** UMAP embedding of immune cells from scRNA-seq with 12 clusters annotated (**Methods 6.10**) and also presented by maturation (PCW, **ii**) and location (**iii**). **d,** Interacting cell network plots showing cell type cross talk via specific receptor-ligand pairs, *IL7/IL7R, CCL21/CCR7* and *CCL19/CCR7* **e,** ST cell type predictions in spots overlaying a submucosal lymphoid aggregate in adult tissue slide showing scores for adult Stromal 4, T-cells, B-cells and Plasma cells in the bottom of adult slide A1 and centre of slide A2. H&E images and spot overlays show selected regions from ST sections. **f,** Expression of CCR7 and CCL19 in ST adult slides. H&E images and all spot overlays shown are plotted over selected regions of ST H&E section A1 (**Appendix A.14**)

4.7 Morphogen gradients in relation to the forming epithelium

In adults, the self-renewal of the intestinal epithelium alongside the cell-fate decisions across the crypt-axis are maintained by morphogens – soluble factors that form a gradient of activity away from their source[205]. Morphogen gradients are controlled through blockade by antagonists, with the most well characterised example being the non-canonical Wnt/ BMP polarized gradient acting to repress the expression of ISCs[26,248]. Dysregulation of any of these pathways can result in abnormal intestinal cell turnover and lead to malignant diseases such as colorectal cancer[270]. Components of morphogen pathways are often studied in isolation; although concerted activity of diverse morphogens is required for co-ordinated development of all intestinal compartments. For instance, in our data expression of BMP ligands was highly restricted to specific cell types segregating with discrete spatial layers/localisations - S2 (*BMP3/4/5*), S4 (*BMP7*), S3 (*BMP1*), muscle (*BMP6*), endothelial (*BMP2*) and neural (*BMP8B*) cell types.

Given the importance of these circuits in facilitating co-ordination between the epithelium and deeper cell types in development, we leveraged both the multi-omic data to generate a morphogen map (**Methods 6.11**) to better understand how these key molecules act in tandem. First generating a curated list of morphogens expressed in the fetal intestine from several pathways known to contribute to patterning, organogenesis and crypt formation in humans (Hedgehog, NOTCH, Wnt, HIPPO, RTK, TGF-Beta, FGF and EGFR signalling), encompassing their ligands, receptors, co-receptors, antagonists and transcription factors (**Appendix A.14**).

This curated list then underwent co-expression analysis of these factors and identified 11 cell-type specific and 13 spatially co-localising modules, from scRNA-seq and ST data respectively, which we scored in spots across all ST slides for individual

module activity (**Figure 4.10a, Methods 6.11, Appendix A.14**). These modules often aligned with depth in ST data. Spatial module 3 consisting of morphogens from ECs, fibroblast and pericytes and contained *LRP1* which was most highly expressing in the S3 sub-set and identified as involved in angiogenesis[271] (**Figure 4.10bi, Appendix A.14)**. Epithelial specific modules expressed genes from Hedgehog pathway (*IHH*), the Frizzled co-receptor LRP5 that is important in Wnt signalling[272], and *FGF3* – all of which have been identified as vital to ISC differentiation[273] - and this module localized near the lumen (**Figure 4.10bii**).

Another module containing myofibroblast components, with the morphogen *WNT2B* and Wnt receptor *RSPO2*, appeared relatively diffusely throughout tissue at 12 PCW but became more localised in later development (**Figure 4.10biii**). This was in line with the earlier finding that mature myofibroblasts appeared much later than epithelial components, suggesting that the ISC-myofibroblast signalling circuit encompassing *RSPO2-LGR5* interactions is not established until several weeks post-crypt formation. Thus this suggests a paradigm where, as the developing intestine grows, morphogen gradients that may be otherwise be broken by increasing physical distance between intestinal layers can be restored by the emergence of new cell types.

The morphogen map highlighted how signalling pathways may dynamically change based on depth and time point maturation of the intestine, during ongoing epithelial maturation. Previous work using scRNA-seq in the adult intestine showed that in developed steady state a large number of morphogens were expressed by the remaining fibroblast subtype S2[195]. This resides in the peri-cryptal region of the crypt and expresses ligands of the TGF- β superfamily and Wnt pathways, a finding supported by other work reporting this as a peri-cryptal telocyte[26,27,90]. The resource identified three sub-types of S2 cells (**Figure 4.8a, Methods 6.10**). Genes expressed by S2 cells included key morphogens

from both TGF-β (*BMP2, BMP4, BMP5*) and non-canonical Wnt pathway components(*WNT5A, WNT5B*) and formed part of a predominantly submucosal fibroblast module encompassing *DLL1*, *BMP5* and *NGR1*, highlighting that these cells and the morphogen-rich niche they provide are present early during intestinal development (**Figure 4.10biv**).

Unlike other fibroblast populations, proximal and distal cluster counterparts of S2 were surprisingly different, and could be distinguished by 885 differentially expressed genes (<5% FDR), including colon-specific expression of *POSTN*, a key gene identified in adult colonic S2 in IBD, and morphogens such as *PDGFRA* which was more prominent in the proximal S2 (**Figure 4.10c, Appendix A.14**). Many location specific differences could be appreciated even before 10 PCW, representing the establishment of a strong locational identity of these cells before crypts and villi are formed (**Fig 4.10d**). These key differences in peri-cryptal fibroblast morphogen profiles between TI and colonic cells suggest the mechanisms by which vastly different morphologies of small and large intestinal epithelium can be achieved during development.

Finally to see the early determination of the S2 sub-types it was confirmed that F3, one of the key markers of this sub-set, was present before crypt formation and formed hillocks which increased upwards as vilification occurred (**Figure 4.10e**). Thus showing that in human intestine, S2 type fibroblasts are required not only for epithelial crypt-niche maintenance, but also play an active role in its formation.

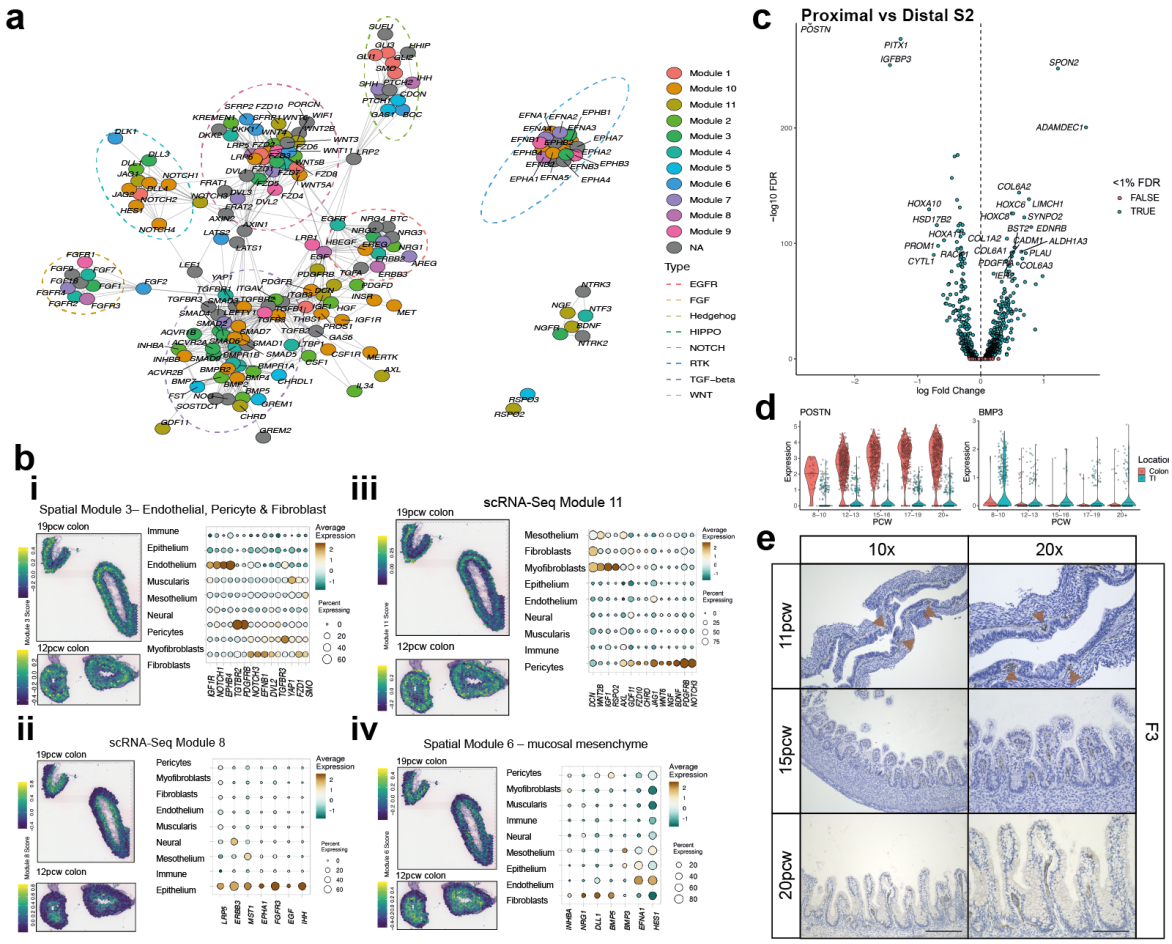

Figure 4.10: Charting morphogens and the barrier supportive role of fibroblasts

a, Graph visualization of morphogen molecule STRING interactome. (Communities enriched for EGFR, FGF, Hedgehog, HIPPO, NOTCH, RTK, TGF-beta and Wnt signalling pathways are highlighted in dashed ellipses. Nodes are coloured by scRNA-Seq module.) **b,** Individual morphogen module overview shown as a module score overlay in ST spots in slides at 19 and 12 PCW colon and a dotplot showing module gene expression at compartment level (ST/scRNA-seq module in title, map and tissue coloured by expression and sized by percent expression). **c,** Volcano plot visualizing differentially expressed genes between colonic and TI S2 populations. **d,** Violin plots depicting selected S2 locational specific genes *POSTN* and *BMP3* that show locational and time-course differences in expression. **e,** Representative images of colonic sections from 11, 15 and 20 PCW embryonic tissue stained for F3 protein by IHC. Brown arrows indicate positive F3 staining below newly developing epithelial invaginations/hillocks (n = 3 for each individual image, 10x/20x magnification scale bar=360/180 μm)

4.8 Application of fetal intestinal atlas to neonatal disease

Congenital anomalies account for a large disease burden in childhood, being the 5th most common cause of death in children under 5 years of age worldwide[159]. Congenital intestinal diseases can present antenatally or acutely after birth, and can be seen in

isolation or associated with complex genetic syndromes. Studying the drivers of these conditions is extremely difficult, given that the genetic associations can be exceedingly rare and that some abnormalities occur early *in utero* such as normal intestinal rotation which will occur in the first trimester[274].

To reveal how some of the time-critical transcriptional defects may contribute, we leveraged the developmental resource and correlated it with a curated list of perinatal intestinal diseases from the Human Phenotype Ontology (HPO), encompassing developmental disorders that were linked to intestinal, ventral, perineal, aganglionic, inflammatory or oncological pathology and were annotated with hereditary phenotypes (**Figure 4.11a, Methods 6.11**, **Appendix A.14**). Integration of 749 disease genes with scRNA-seq data, connected congenital disorders with phenotypes that likely manifest through highly cell-type specific defects (**Figure 4.11b**).

Localised expression could be seen at compartment level such as *SPINT2* which was expressed across the epithelium, while *DGAT1* localised within distal enterocytes (area under the curve (AUC) = 0.80, **Figure 4.11ci**). Defects in both genes can result in forms of congenital diarrhoea so severe that they present within the first few days of life and require life-long parenteral nutrition[275,276]. Genetic defects in *SOX10* and *RET* are both associated with Hirschsprung's disease (intestinal aganglionosis)[277] but were differentially expressed in the glial and neuronal cells (**Figure 4.11cii**). The most specific cell types in which these genes were expressed were lymphoid associated glial (AUC 0.814, **Appendix A.14**), highlighting cell types as important in Hirschsprung's disease development in utero. Given that the former also expressed *GFRA3* which has been shown to play a role in the formation of lymphoid follicles[268], our data suggests a potential link in explaining one of the important clinical complications of Hirschsprung's disease – Hirschsprung's

enterocolitis – that originates through complex neuro-immune interplays and occurs even after aganglionic segments of intestine are removed[277].

Within the fibroblast cells linkage was seen with expression in S3 subtypes of *COL3A1* (S3 transitional AUC = 0.965, S3 EBF+=0.95, S3 HAND1+=0.943, S3 progenitor=0.893, S3=0.891) which leads to the phenotype of the vascular variant of Ehlers-Danlos syndrome (ORPHA 286)[278] (**Figure 4.11ciii**). The two major diagnostic criteria are vascular and intestinal disorders with intestinal perforation and arterial aneurysms[279]. This is intriguing, given the previously highlighted role of S3 in normal endothelial cross-talk and angiogenesis within the intestine, and therefore shows how application to human disease may represent the effects of abnormalities in this cell sub-type.

The temporal aspect of our data also allowed the study of disease genes that significantly varied over developmental time, and could provide insight to linked conditions at times when the compartments are forming. *CAVIN1* was associated with lipodystrophy presenting with hypertrophic pyloric stenosis[280] (OMIM:613327, **Figure 4.11d**) and was differentially expressed over developmental time in the muscularis (coeff= -0.284, p=2.12 x10^{-30}). *HMGA2* was most highly expressed in the IM progenitor population (AUC=0.796) associated with pre-12 PCW time points and subsequently decreased in expression over time in the muscularis and neural compartments (coeff=-0.298, p=2.20x10^{-149} and coeff=0.087, p=6.71x10^{-24} respectively)(**Figure 4.11d**). *HMGA2* is located on chromosome 12 and lost with 12q14 microdeletion syndrome (ORPHA 94063)[281] that can manifest in intestinal malrotation. It is revealing to see that *HMGA2* expression coincides with the time point where the intestine returns to the abdominal cavity in a tightly regulated process to prevent malrotation(**Figure 4.11e**). Our results highlight that glial cells and muscular progenitors may be vital in this process. Failure of the intestine to return at this time can also be seen in the condition of omphalocele and is a

feature of Robinow Syndrome (OMIM 618529). It has been reported pathogenic variants of *NXN* can lead to this syndrome[282] and *NXN* was also expressed at a similar time point to HMGA2 in inhibitory motor neurons (AUC=0.751) (**Figure 4.11eii**).

Taken together these conditions highlight the timing of neural and musculature interaction to ensure normal intestinal health. This demonstrates how our resource can be rationally linked with disease and reveal new information on where and when these abnormalities might occur during development, in a way which has not previously been possible due to the pre-natal nature of these defects.

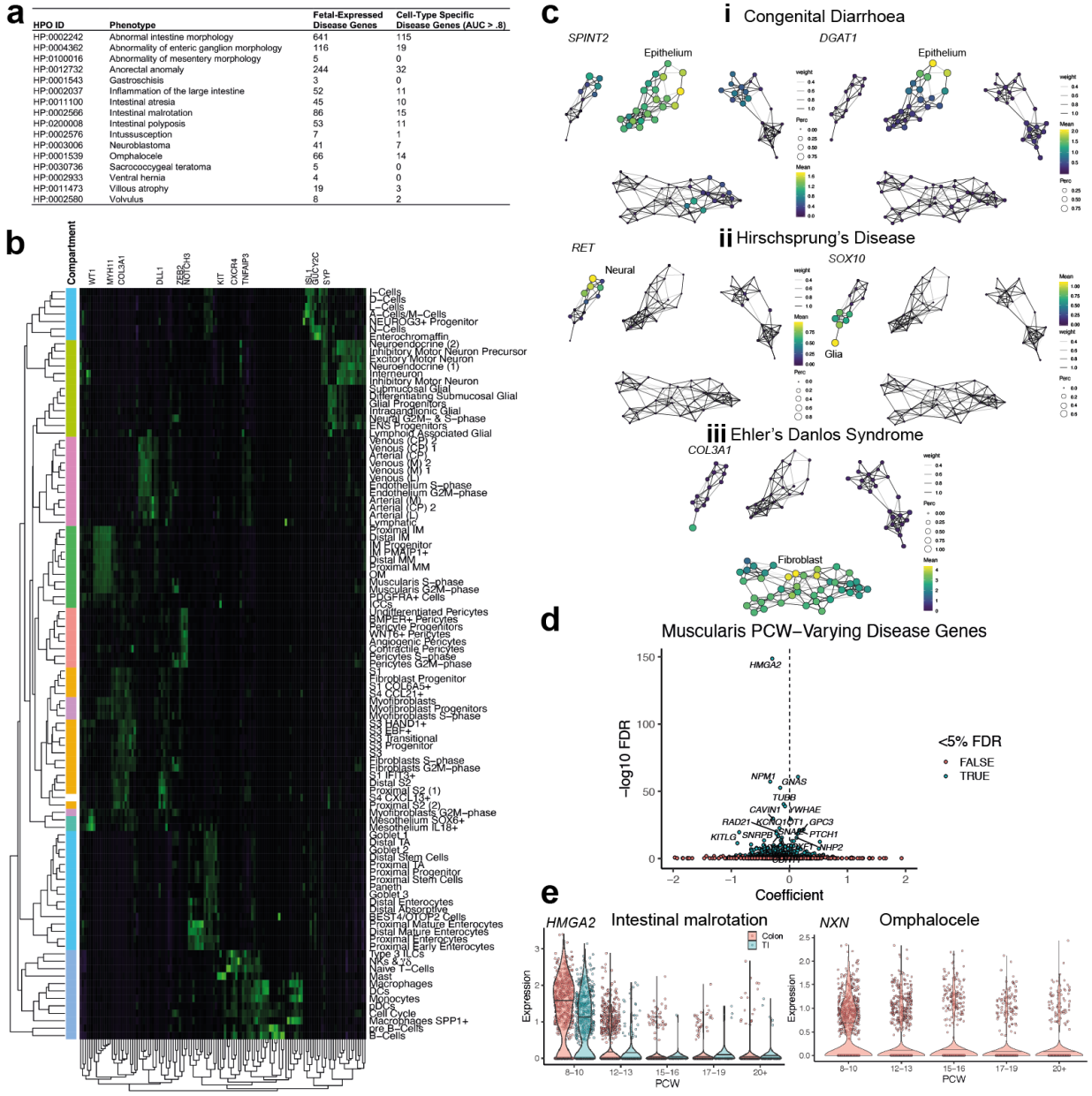

Figure 4.11: Application of resource to paediatric intestinal disease

a, Table summarizing intestinal disease HPO phenotype terms used for comparison to resource, and the number of associated genes which are expressed in scRNA-Seq dataset or highly cell type specific. **b,** Heatmap visualizing mean scaled cluster expression of cell-type specific disease genes summarized in (a). **c,** Graph abstraction overlay of cluster expression of disease genes associated with (**i**) congenital diarrhoea, (**ii**) Hirschsprung's Disease and Ehler's Danlos syndrome (**iii**, cell compartment labels added for clarity, cell types correlate to labels in **Figure 4.2,** node colour represent relative expression and size percentage expression). **d,** Volcano plots highlighting top time-course varying disease-associated genes in muscularis compartment. **e,** Violin plots showing individual time-course varying disease-associated genes in intestinal malrotation (*HMGA2* expression over time in muscularis compartment of TI and colon, left); and omphalocele (*NXN* expression over time in neuronal compartment).

4.8 Interim conclusion

This part of the project reports the first large scale study of human intestinal development from 8-22 PCW at single cell resolution. It also represents the first ST exploration of adult and fetal intestine, providing crucial spatial context to our findings. scRNA-seq of over 95,000 cells was able to characterise 101 intestinal cell types and states, capturing cellular interplay in tandem during several critical developmental events - epithelial crypt formation, maturation of the crypt/villus axis, establishment of muscle layers, expansion of vascular systems and early immune colonisation following the emergence of GALT.

The formation of the intestinal epithelial barrier is a complex and co-ordinated process. This study captured both the first stages of crypt development and subsequent maturation into adult-like morphology seen at birth. This helped place the previously identified new cell type BEST4/OTOP2 cells, which were in sharp contrast to most other specialised epithelial cells and were already established prior to crypt formation and appeared independent from the crypt-villus axis dynamics at later time points. The resource identified location-specific transcription factors, such as epithelial proximal *FOXD1* and distal *LEFTY1*. We also observed time-critical differences in early, pseudostratified epithelium where a population of TI specific cells termed proximal progenitors could be distinguished by several transcriptional hallmarks such as expression of *TF* and *VTN*, genes that have individually been shown to play a role in vilification in animal models[254,255].

The time window included formation of *LGR5*+ hillocks at 12 PCW which was validated in tissue; however, even several weeks post crypt formation, a substantially larger proportion of ISCs remained when compared to adult tissue. It is interesting to note

that even though the variety of cell types observed at 22PCW was in line to those found in adult tissue, the proportions of immature cells were higher. This may help explain why even though mature morphology is present, the epithelial barrier may still not be primed, and consequently extremely premature infants (<30PCW) are at high risk of disease characterised by barrier breakdown, such as NEC[283,284].

A key advantage of the method in this part of the project lies in the capture of full thickness of the intestinal tissue, whereas past studies in adult were reliant on intestinal biopsies with limited depth. Here, for the first time we were able to chart all intestinal compartments and delineating. The temporal dynamics of these cells shed light on how the morphology of the intestine is achieved, where while some populations emerge in tandem, other early cell types create niches paving the way for differentiation or colonisation by lagging populations. This is evident in GALT formation, where the coordinated emergence of LTi and "S4" populations at 12 PCW was followed by increasing B and T cell colonisation. S4 type fibroblasts were first reported as UC-associated cells in adult colon, bearing features of follicular reticular cells[195]. Here, we definitively localise them to submucosal lymphoid aggregates in adult colon using ST and highlight the cell types and pathways involved in formation of GALT and PPs. This also raises future questions about the potential dichotomous function of this cell sub-type; how it can be involved in normal lymphoid tissue formation in development and also direction of inflammation in UC

Furthermore, we also establish a putative new function for "S3" type fibroblasts as vasculature niche cells, where ST was able to localise these cells congregating around large vessels. These cells interacted with endothelium via *ANGPT2* and other R-L pairs and expressed complement components (e.g. C7) which suggests involvement in local tissue remodelling during vessel development. In fibroblasts it was also shown that the peri-cryptal fibroblast/telocyte population S2, a source of key morphogens required for

epithelial crypt support, emerges prior to crypt formation and localises to “hillocks” as epithelium begins to warp during this process. Furthermore it was identified how a number of key morphogen pathways intersect with S2 and across other compartments in time. These early niche cells were also already highly distinct between TI and colon.

Overall, this reflects the ability of utilising a multi-compartment resource to study co-ordinated development of cellular cross talk and dynamics of differentiation across the fetal intestine. We find some populations appearing first and thus likely establishing critical niches that pave the way for other cells, whereas in others co-localising populations evolve in tandem. The approach is able to map the formation of the human intestine at high resolution and also apply it to important neonatal diseases such as Hirschsprung’s disease, malrotation and omphalocele among others.

Chapter 5: Concluding Discussion

5.1 Summary of findings

The intestinal barrier is vital in numerous key homeostatic roles, dysregulation of any can result in disease. Since the original descriptions based on histology and targeted stains it has been clear that the historical "four core differentiated cell types" of the intestinal epithelium have been bookmarks for deeper, more specific roles which are required to complete the barriers numerous functions[29]. Previous intricate modelling approaches helped gain insights into how these cell types were placed on a maturation axis - from a primitive ISC to developed cell type with subsequent anoikis every few days – and the transcriptional drivers that define mature phenotypes[8] (**Table 1.1**). However, these findings also raised questions of what constitutes the basis of normal cellular diversity, cell states, and the lineage of maturation between cells within the epithelium between the core cell types.

Given that previous targeted techniques limited *de novo* discovery and that bulk unbiased approaches would mask the contribution of individual or rare cell types, it has been challenging to formally address this until now[174]. Recent work when commencing this project in animal models[192], lower throughput human samples[193,194] and work preceding this in colonic fibroblasts[195] highlighted scRNA-seq as a technique of choice to map diversity and highlight the drivers of epithelial barrier breakdown at the molecular level. For example, it was shown that this could be applied to generate a classification profile of intestinal fibroblasts, how they play diverse roles such as epithelial maintenance (S2) or plasticity and expansion in inflammation (S4)[195].

Thus, to build on these foundations, this work first sought to apply scRNA-seq to intestinal epithelium of samples taken during endoscopy from HCs and those with IBD. Sampling from the distal colon with a validated digestion protocol generated a resource of over 11,000 cells from 6 patients and 9 conditions (healthy, inflamed, non-inflamed). This highlighted clusters of known epithelial cell types in health based on gene expression, and the resolution allowed "zooming in" to both rare cells (EECs) and the ISC niche and could identify novel marker genes of some epithelial cell types that were validated in tissue (**Figure 2.3**). Moreover, alongside discrete clusters the spectrum of gene expression allowed mapping *in silico* of maturation axis which correlated to crypt depth (**Figure 2.2**) and maturation from ISC to enterocyte or absorptive lineage (**Figure 2.4**).

After confidently mapping the steady-state in health comparisons were made to UC, a condition with distal intestinal inflammation and in which up to 40% of patients will fail to respond to current medical therapies[165]. This identified global transcriptional changes across the epithelium, such as universal upregulation of interferon gamma signalling, antigen presentation and cytokine production across cell types (**Figure 2.5**). The approach used allowed further information on cell-type specific responses to colitis, such as *LYZ* expression being specifically upregulated in the deep goblet cells of the distal colon (**Figure 2.5**). This demonstrated that although Paneth cells, the de-facto *LYZ* expressing cell in the small intestine, are absent in the distal colon goblet cells can adopt their bacterial protective role in colitis. This is similar to the described function of a deep secretory goblet cell in mouse models, and suggests the barrier re-organises cellular roles in inflammation to protect the ISC niche[36]. The counterpoint to this was other roles were down-regulated, such as *HB-EGF* in progenitor cells (**Figure 2.5**). *HB-EGF* in health maintains Wnt/β- catenin signalling to facilitate ISC regeneration[212], and this suggests a loss of tissue homeostatic to further protective, bactericidal roles.

The combination of novel cell-type markers gleaned from an unbiased approach, and the cell-type specific dysregulation that could be charted in UC highlighted the previously unappreciated goblet cell gene *WFDC2*. Although goblet cells are seen throughout the small intestine and colon, their interactions with bacteria differ by forming two layers in the colon with a sterile inner layer and a single layer in the small intestine[43]. This highlights that colonic goblet cells and the mucus they form will have a more direct link with bacterial pathogens and commensals. *WFDC2* was expressed in crypt-bottom goblet cells in health with selective reduction of this in active UC, a feature which was validated in a cohort of patients (**Figure 2.6**). WFDC2 was a secreted anti-bacterial protein which had previously been shown in the testis to have an innate immune regulatory affect through protease inhibition[285], so loss in UC could signify absence of this role. This was validated in a homozygous murine model (*Wfdc2*$^{+/-}$) which showed leukocyte infiltration and barrier breakdown, with bacterial invasion (**Figure 2.7**). Thus the first aim of the project was able to chart the diversity of the epithelium, identify new features of differentiation and also compare the cell type specific dysregulation of UC. The demonstration of goblet cell expressed protein WFDC2 as a new and important anti-bacterial defence factor shows the approach was able to identify molecular drivers of barrier breakdown, and it is envisioned the data generated will provide other targets for the future[2].

In the epithelial census one cluster of cells, the BEST4/OTOP2 cell did not match to previous annotations, and so extra care was taken to study this in detail. This was to see if, in a similar manner to other organs, this may represent a new cell type identified due to the unbiased approach[184,185]. Protein and RNA staining confirmed that the signature of this cell was localised to a distinct enterocyte sub-set (**Figure 3.2**). Use of a membrane maker allowed isolation of the cell type which could then be studied with in depth RNA and

proteomic analysis (**Figure 3.3**). Comparison to previous smaller scRNA-seq datasets could identify the matching signature of this cell type, and most intriguingly analysis of colonic cancer samples from the tissue cancer genome atlas showed loss of the individual genes that constitute the markers of the BEST4/OTOP2 cell type (**Figure 3.1**). Functional studies of this cell type highlighted a role in pH balance with an active proton channel (**Figure 3.4**), and so it is possible that the loss of BEST4/OTOP2 cells in cancer may be linked with a loss in pH balance. The recent discovery that otopterin 2 (OTOP2) is a channel involved in taste sensing further supports that the cell type is likely involved in a defensive epithelial role[229].

It is known that mutations in the uroguanylin pathway (*GUC2YC)* can result in profuse diarrhoea, an increased susceptibility to IBD and obstruction secondary to meconium ileus at birth[286,287] . Loss of this signalling cascade also is shown to result in increased intestinal tumorigenesis in murine models and to this end uroguanylin mimetics have been developed to target this depletion[244]. Our results put into context that the loss of expression in these disease contexts could be co-ordinated from one type of cell, rather than generalised epithelial dysregulation. This will be important for the application of these treatments and to further understanding of these diseases.

Discovery of a new enterocyte cell demonstrated the power of the approach, but also raised questions about how to place that and previously described fibroblast sub-types in the spectrum of intestinal maturation[195]. Although it is clear that the intestine appears morphologically mature at birth, much previous work characterising intestinal development had been elucidated from animal models of from archived fixed human specimens[77]. There are key differences in animal models of maturation, with many forming villi but the drivers being hypothesised as the buckling force of the muscularis in chick and the localisation of morphogen rich “hillocks” in mice[84,248] and the drivers in

humans are still not fully understood. Moving forward from the first census, the project next sought build upon the continued exponential growth of scRNA-seq techniques[186] with a multiplexed approach hugely increasing the previous throughput[249] in order to chart human intestinal development across time.

By sampling across the time points of intestinal maturation (**Figure 4.1**) with an adapted digestion protocol with sample multiplexing, this census was able to capture cells from the entire thickness of the intestine (**Methods 6.3**) which allowed a holistic view of formation and highlighted compartmental differences over time. Within the epithelium this identified key differences in locational transcriptional features, with ISCs showing regional markers before morphology such as *GATA4* in the small intestine (**Figure 4.5**). Comparison of locational differences revealed a proximal specific progenitor population, that heralded the formation of *LGR5* expressing ISCs (**Figure 4.5**). This expressed genes such as *CDH2* which has been shown to be a regulator of stem cell decision making in mesenchymal stem cells[288] along with *VTN* and *ONECUT2* which have role in animal models of villi formation[254,289]. Together this reflects an early fate decision with transient high expression of vilification processes specific to the small intestine before ISCs form. In mice it has been described that *Lgr5,* the Wnt target receptor that gives rise to all lineages of epithelial cells[46], is diffusely expressed before localising[248]. The resource was able to capture this in human development and we validated this finding in tissue (**Figure 4.5**) thus capturing the appearance of normal crypt-villus architecture and placing each cell within it.

This identified the previously described BEST4/OTOP2 cell as presenting in the epithelium early, and being transcriptionally distinct in even the earliest time window. (**Figure 4.6**). There were few other mature compartments or cell types at this early stage, but among those were seen neural cells and EECs appeared to interact with

BEST4/OTOP2 cells through R:L analysis. One of these pathways included vasoactive peptide secreted from neurons and the receptor on BEST4/OTOP2 cells. VIP is reported to be important in immune defence through signalling with ILC3s in an IL-22 dependent manner in maturity[290]. It is interesting to see that these signals may occur directly with the epithelium in the primitive intestine. These interactions also suggested links with *NOTCH2* expression in BEST4/OTOP2 cells and the notch ligand *DLL3* from neuronal cells (**Figure 4.6**). NOTCH signalling is a key pathway of intestinal epithelial cell differentiation in the mature crypt-villus axis, with fate decisions into secretory or epithelial lineages being decided based upon it through the mediator *ATOH1*[291,292]. NOTCH signalling is also reported to activate downstream transcription factors of the Hairy and Enhancer of Split (HES) family, and *HES4* is also a specific marker of BEST4/OTOP2[236]. Together this shows the application of multiple analysis across developmental time could identify new avenues for understanding the maturation of this novel cell type, which appeared to be uncoupled in time from other epithelial cells.

As well as charting interactions from the epithelium, the full-thickness digestion approach allowed this to be expanded throughout the other compartments to chart cross-talk. The application of ST also gave a richer view of the emergence of cell types and trajectories within a tissue based context. A key example of this was within the fibroblast cells, although diversity had been demonstrated before in IBD with annotations S1-4[195], some of the sub-type roles were not confirmed. Interactions with pericytes and endothelial cells highlighted S3 to be vascular associated and involved in the process of angiogenesis (**Figure 4.7**). Recent work characterising fibroblasts across multiple organs in mice and humans, with application to diseases showed fibroblasts expressing *Dpk* and *Col15a1* represented a possible progenitor population using lineage tracing[293]. Both of these genes were highly expressed in fetal S3 EBF+ and S3 respectively, suggesting these primitive

processes such as angiogenesis may also show this sub-type to be a fibroblast progenitor population in the fetal intestine (**Figure 4.8**). Spatio-temporal maturation of immune cells in the small intestine with S4 and LTi-type ILC3s showed this to be involved in the appearance and formation of GALT[136] (**Figure 4.9**). There is still a great deal of debate about whether the intestine is sterile before birth, with conflicting reports of bacterial absence and of bacterial priming before birth[133,294]. Although we did not look for bacteria in this study, the finding that immune system in the small intestine appears formed in early development with mature cell types is interesting and raises the question of at what time point between our resource and birth that the colon follows suit.

The interrogation of cell-cell and compartmental interactions was a rich resource, but it was also able to be interrogated with broader questions to chart cross-talk across cell compartment sand apply the resource to disease. This included charting morphogens. Where previous work involved looking at specific components of pathways such as NOTCH, Wnt, EGFR, FGF, Hedgehog, HIPPO and TGF-beta our resource could chart them in tandem (**Figure 4.10**). It is known that these pathways are important in maintenance of the ISC niche[248,295]. The data generated could look individual genes over location how they formed modules with other morphogen components across compartments, and how the module expression would change over time and location (**Figure 4.10**). Thus *BMP3* was seen in proximal specific S2 telocyte like cells, but more broadly a morphogen module including it was passed from the serosa early in development onto fibroblasts/myofibroblasts later on in developmental time as these mature cell types appeared and the distance between cells increased. These associated modules will hopefully form the basis of future work to look at the locational specific determinants of crypt and villus structure, and shows that it may be more complicated than looking at one morphogen pathway in isolation.

Finally, congenital anomalies account for a large disease burden in childhood, being the 5th most common cause of death in children under 5 years of age worldwide[159]. Congenital disease of the intestine can present in a variety of ways- antenatally on anomaly scans such as ventral wall defects; perinatally with acute disease in the setting of atresia or anatomical malformations; or postnatally with congenital sequelae such as malrotation and volvulus. These can be seen in isolation or associated with complex genetic syndromes. Studying the drivers of these conditions is extremely difficult, being exceedingly rare and that some abnormalities occur early in utero. Intestinal ventral herniation, elongation, rotation and repositioning all occur in the first trimester and deviation from the co-ordinated processes driving this will lead to defects such as omphalocele[274]. By linking the developmental resource with HPO terms that encompassed a variety of congenital intestinal diseases the time a cell-type expression of disease associated genes could be seen (**Figure 4.11**). This demonstrates the power of integrating high resolution data to provide insights into neonatal disease. Such as Hirschsprung's disease being linked with known expression in *SOX10* and *RET* expressing neurons[277], and also time-point specific expression of *HMGA2* in IM progenitor cells at the time that intestinal rotation is occurring, errors in this being seen in 12q14 microdeletion syndrome of which malrotation is a feature (ORPHA 94063). This approach also helped validate earlier results of the resource, with *COL3A1* being seen in the vascular fibroblast sub-type S3, and defects of which present with a variant of Ehler's Danlos syndrome where intestinal perforation and aneurysms are two diagnostic features[279]. We tracked genes linked to these genetic defects to highly specific time points and cell types, and thus reveal information about disease pathology that is otherwise difficult to study in human tissue due to in utero development of these abnormalities. The field of pre-natal diagnosis is rapidly expanding; thus, this may aid both further study of these diseases and the

application of in utero intestinal anomaly screening by highlighting pathways of dysregulation and their cellular origins.

5.2 Wider significance and evolution of research

The work outlined in this project highlights the diversity of epithelial cells in maturity, and the full spectrum of intestinal cells within development. Since starting this work the field for both of these avenues has expanded exponentially, with studies applying similar new high resolution techniques in both contexts. This highlights the importance of furthering understanding in this area and also generates a number of comparators for applications of the projects results.

5.2.1 Wider significance in epithelial biology and IBD

Within IBD a number of other studies have used similar or complimentary techniques to reveal new insights into barrier breakdown. Some of these have used a similar approach to the latter parts of this project to undertake scRNA-seq on the epithelium and underlying stroma in tandem. Smillie et al compared a larger number of patient samples (18 UC and 12 healthy) and used a digestion technique obtaining a full biopsy digestion[251]. Within the epithelium BEST4/OTOP2 cells were seen, representing independent confirmation of our findings. They were shown to be slightly decreased in frequency during UC inflammation and a predicted lineage inference for it being associated with M-like cells, a counterpart of M-cells that were grossly expanded in inflammation[251]. This would support previous investigations of the uroguanylin pathway being downregulated in IBD, a specific gene expressed by BEST4/OTOP2 cells[296].

Other reports using scRNA-seq but with different digestion techniques also identified BEST4/OTOP2 cells in paediatric IBD[297] and in adult UC[250]. The latter termed this a colonic Paneth like cell, but the transcriptional profile was identical and highlights the challenges in annotating new cell types from unbiased techniques. A similar cell type was also identified in the more proximal alimentary canal, with expression of *CFTR* alongside *BEST4* in the duodenum of a rare enterocyte subset[298]. The expression of the key gene linked to cystic fibrosis is of note as sufferers will have a number of intestinal complication alongside the well documented respiratory compromise – such as meconium ileus, nutrient malabsorption and distal intestinal obstructive syndrome[299]. Together this finding shows differential expression in BEST4/OTOP2 cells across location, but a potential important role in a number of diseases; CF in the proximal intestine, UC in the colon through uroguanylin dysregulation, and selective loss of cell type signature in colorectal cancer.

In the larger study of UC inflammation the epithelium differentially expressed genes were in line with a number of metabolic pathways, including *DUOX* and *NOS2* that were seen in our data (**Figure 2.5**) although the larger number of samples allowed mapping to metabolic pathways that were dysregulated and associated with GWAS risk genes – such as *GPR35* in the tryptophan pathway[251]. Risk modules were generated across cell types to show epithelial associated inflammation specific pathway expression, a similar approach was used to characterise cross cellular responses in CD to summarise synchronised dysregulation of plasma cells, monocytes, T-cells and stromal cells but unfortunately epithelial cells were not included in this analysis[300]. It is of interest that GWAS inflammation modules did not identify goblet cells as highly connected, with Macrophages and CD8 cells incorporating a larger percentage of GWAS loci. This may be due to GWAS genes being more highly correlated with immune cells, and should be considered

when applying results of these methods given that *WFDC2* was not a GWAS associated gene but was shown to be a driver of barrier breakdown.

Goblet cell diversity has been well reported across different mucosal barriers, where the specificity of defensive roles will differ alongside mucus production, with a single layer of mucus in the small intestine and double mucus layer in the microbial laden colon[43]. Alongside this project recent papers also studied goblet cell diversity in the mouse using an mCherry-MUC2 reporter mice[301]. This showed 8 clusters of GCs with a spectrum from proliferation to mature inter-crypt goblet cells via a canonical and non-canonical route, with *Wfdc2* seen in the former and the transcription factor *Hes1* identifying the latter. This suggested that it was mature inter-crypt goblet cells that are lost in UC, although the goblet cell driven breakdown of the mucus barrier was a shared feature to what we identified in *Wfdc2*$^{+/-}$ mice.

Overall these reports show that the projects key findings appear externally validated and are placed well within a rapidly developing field. The epithelial barrier breaking down is a key feature of both forms of IBD so these results and other reports of new findings of epithelial diversity and goblet cells function help highlight the drivers of mucosal barrier function and breakdown in unprecedented resolution, with outstanding questions on the role of BEST4/OTOP2 cells in disease and the drivers of goblet cell diversity.

5.2.2 Wider significance within intestinal development

Within development there has been a similar abundance of new studies. When first starting this project there was reports of scRNA-seq studies of the developing intestine, with over 5,000 cells taken from throughout the digestive tract and across time (6-25PCW) which could identify broad transcriptional changes over time or location but lacked

resolution to make robust comparisons within cellular compartments[194]. By focusing on specific anatomical locations and increasing throughput with a multiplexed droplet based technique with batch processing this study managed to gain greater insight into key maturation events in the small intestine and colon.

Alongside the completion of our project other reports detailed the formation of the intestine at multiple locations between 6 and 10PCW in 9 developmental samples which were also compared to a cohort of paediatric CD samples collected from the terminal ileum[302]. This resource showed full thickness cell contents in fetal samples with similar annotations to ours, and identified a proximal distal wave of epithelial cell maturation in secretory cells, something our time window mirrored in the early time points and in later more mature cell types such as muscularis and myofibroblasts. In addition this study also identified a proximal *BEX5* progenitor cell type which was in keeping with our proximal progenitor cell type, and highlighted cross talk between S2 or FOXL1+ telocytes and the maturing ISCs in epithelial formation[302]. Together this shows how complimentary overlapping time point windows in different studies can identify shared features of maturation.

A similar study took samples from earlier time points than ours (6–18PCW, n=8 samples) but focused on the proximal intestine namely the duodenum and jejunum. In line with these studies in series, vilification appeared to take place earliest in this proximal location, being seen histologically at 7PCW[303]. This also allowed capture of the maturation of a peri-cryptal fibroblast population with expression of *F3*, *PDGFRA* and *DLL1* in keeping with our S2 annotation, showing it was not present before 8PCW but matured from that point. This was similar to what we identified with key S2 genes such as *POSTN* increasing in expression after 8PCW in the colon (**Figure 4.10**). This study identified the EGF family member NRG1 to be secreted from S2 and its receptor ERBB2/3 being

expressed on developing epithelium, demonstrated by the addition of NRG1 to organoid culture increasing diversity of mature secretory cells[303]. Together this demonstrates how similar techniques to this project have found preserved cellular features across the intestine, such as S2 being crypt supportive, and the method can be used to link with modelling in techniques such as organoid culture. This has also been carried forward to murine modelling where study of S2 markers in more detail showed further heterogeneity in mice related to depth, and abrogation of *Grem1* expression in a Cre dependent manner showed crypt degeneration[26].

Results from recent studies have also reiterated key features in other compartments of the developmental resources. Within ECs large resources have shown key differences in the transcriptional signature of large and small vessels, or across organs[304,305].This showed transcriptional differences across vessels (confirming *HEY1* in arterial an *VWF* in large veins), and also that different sub-sets had differing metabolic and angiogenic profiles, in line with what we saw with ECs related to S3 during intestinal angiogenesis, and that some of these roles are conserved across organs whilst some are organ specific[304,305]. In the murine intestine, when compared to other organs, ECs had some expression of lymphatic like genes (*Tcf15, CD36, Fabp5*) but without the key hallmarks of these (lacking *Prox1* and *Pdpn*) suggesting a lacteal role for transporting dietary fats in venous circulation[305]. Within our developmental data the marker gene of this murine cluster, *Aqp7*, was seen in capillary ECs which would be the site of these nutritional uptake.

Our study identified a range of neural and glial cells that were seen to be distinct at the earliest part of the time window, in keeping with animal models that ENS colonisation and fate decisions occur very early in development[306]. This finding was contrary to other reports, where ENS cells required specialised protocols with nuclear isolation to be captured by scRNA-seq[307]. In development this is likely due to the size of the neuronal

cells, with other fetal samples being processed without modification[302]. A number of neurotransmitters used for annotation were conserved between nucleus sequencing, and in turn with mouse, such as acetylcholine receptors (*GFRA2, HTR4*) in excitory neurons[307]. Disease associated risk genes from GWAS studies were also mapped to the ENS of adult patients with this work, highlighting targets such as *RET* and *PHOX2B* from a different method to our HPO mapping[307]. Together this shows a rapidly emerging field, with a large number of new insights being seen about the intestine and its formation and dysregulation. Some studies approaching areas of interest in a compartmental level, broad organ level or from findings of animal modelling. It is reassuring that across the field the use of high resolution techniques seem to be confirming a number of the key findings of this project.

5.3 Future directions

The work detailed in this thesis sought to address the molecular drivers of epithelial barrier breakdown and formation using recently developed high resolution techniques. This has revealed how cells interact across the depth of the colonic crypts, placed a new cell type with a role in pH balance within the diversity of epithelial cell types and shown that an unbiased approach can identify previously unappreciated mechanisms of dysregulation in IBD[2]. Within human development it showed how the epithelium and it's stem cell axis formed, applied cellular signals to tissue and also charted cross talk across intestinal compartments during key maturation events[3]. During the undertaking of this work the field has also progressed, placing many of the key findings in context and highlighting them as important within the field[308]. Although many questions remain unanswered which will be of interest in future work.

In line with the exponential increase in the throughput of scRNA-seq within this project, other studies have reflected this with new insights being gained on an organ[309] or organism level[310]. These have highlighted the level of study using unbiased RNA and associated sequencing techniques that see the aim of a human level cellular level resource being achievable within the coming years, in a venture compared to the human genome project in scale[187]. This thesis has represented one small part towards this goal by describing the determinants of intestinal formation and dysregulation, but further meta-analysis of data as these studies appear will be important. Namely the study of maturation from fetal to adult intestine is missing the paediatric and immediately post-natal phase. A number of pathologies can occur after birth, such as NEC with extremely high mortality in premature infants[88]. Our data showed the appearance of proximal immune follicles but the colon will not fully colonise until after birth[136] and within the first year of life the intestine will be a key site of immune cell priming[137]. A cross comparison of the maturation of the intestine and immune maturation through childhood is greatly needed.

On a cell type level questions also remain about the origins of cells through maturation and the accuracy of lineage prediction in epithelial and non-epithelial compartments. Specific cell sub-types are beginning to approach these issues, such as a combined cross-organism data meta-analysis of fibroblasts have shown that a possible progenitor with markers *Pi16* and *Col5a1* identifying potential progenitor clusters[293]. Within the near future it will be possible to build on such approaches to integrate across diverse conditions, such as CD and UC across age and disease sub-type, or varied spectrums of age such as fetal through to adulthood. Some studies are still missing key features to address these questions, such as not incorporating epithelial cells[311], only sorting immune cells in health, UC and checkpoint colitis[312] or focusing on only one cell type[307]. Future work will seek to fill the unmet need of these omissions with increasing

throughput to begin approaching a cross comparison of origins on intestinal dysfunction across maturity. These will also seek to integrate other omic approaches to intestinal cells, such as epigenetic regulation of RNA expression through maturity and disease[313] and other multimodal approaches such as genetic mutations and protein expression, which have been vitally revealing in bulk analysis[314,315] and are beginning to dissect cellular behaviour and states at the molecular level[316,317].Together this will help detect the earliest cellular states in developmental time and response to disease stimuli.

Future directions shall also seek to build upon this project and fully characterise the role and maturation of the BEST4/OTOP2 cell. Extra care was taken to confirm this represented a distinct cell type due to the expression of *SPIB,* a de-facto M-cell marker[71], and so it was validated to stain a specific cell type and shown to have a role in pH balance (**Figure 3.4**). It is intriguing that this cell type had not been seen before from targeted approaches, or in unbiased approaches studying the murine intestine[192], perhaps being related to there being no ortholog of one of the key marker genes *BEST4*. It has been shown that zebrafish do express a counterpart of the cell type and that there is a distinction between *Spib* expressing epithelial cell types[109] providing a tantalising avenue for future *in vivo* work. Other modelling approaches include the culture and expansion of intestinal epithelium using *in vitro* primary organoid models that are now able to recapitulate epithelial homeostasis across the gastrointestinal tract[24]. Studies have shown that a BEST4/OTOP2 cell type is identifiable in human intestinal organoids cultured in this manner[302,303] including in more proximal intestinal organoids where *CFTR* expression is higher in the BEST4/OTOP2 population[298], a gene which has been a favourable target for gene editing in respiratory organoids[318]. These models are continuing to advance with multiple cell compartments and all germ layers being able to be engineered and so cross talk from human epithelium could be applied to fully study BEST4/OTOP2 cells as well as

other targets generated from this projects resource[319–321]. This is also accompanied by recent breakthroughs in lineage tracing methods within organoids that could aid the clarification of the origins of BEST4/OTOP2 cells maturation axis[322–324].

Finally, the results of the project in mapping the development of the fetal intestine showcased the power of integrating scRNA-seq with spatial transcriptomics, helping balance the strength of single cell analysis with mapping signature to tissue, something which is especially important in development where signalling may be context dependent within tissue. Other studies have also shown the benefit of these combined approaches[325–328] and this has been used on the organ scale in the heart during human development[246].

Similar to scRNA-seq this field represents a future prospect of an explosion of techniques to further the understanding of the roles of individual cells *in situ*. There is currently a dichotomy of largely expanding techniques with targeted approaches brining methods to surpass the limit of fluorochromes in visible light. These include examples such as imaging mass cytometry or CODEX which have expanded panels of proteins over 36[329], and shown highly dimensional cellular relationships in the intestine during development and cancer[330,331]. These have a differing approach to other expanding techniques with large panels targeting hundreds or thousands of genes such as in situ sequencing[246], MERFISH[332], seqFISH[333] and FISSEQ[334] that will allow targeting many genes whilst retaining cellular or sub-cellular information but require more complex imaging methods. A previous limitation of the truly unbiased techniques such as ST, which was used in this project, is the loss of that cellular resolution. Although partly overcoming that with integration of scRNA-seq data future techniques appear to promise both with methods including high density ST[335], Slide-seq[336], Stereo-seq[337] and seq-scope[338] all purporting sub-cellular unbiased resolution.

Although this increases the diversity of techniques it remains to be seen which will best address the ongoing limitation of ST which will be to study cell in tissue and identify their communication by assessing R:L interactions. Currently a great deal of emphasis is put on cell-cell contact, where some key signalling circuits such as morphogens may have distant effects such as across the whole embryo in development. Similarly in IBD inflammatory cytokine signalling and their subsequent pathways can have distant effects, and methods to address these questions will be of great interest in the future. This again highlights the future benefit of the continuing growth of the approaches used in this project to help understand the intestines growth and dysregulation.

Chapter 6: Methods

6.1. Processing of adult and fetal intestinal samples

For adult samples used in scRNA-seq, after informed patient consent (REC reference: 16/YH/0247 and 09/H1204/30) 3-4 pairs of biopsies were collected from volunteers attending endoscopy for routine colonoscopic screening with no inflammatory diagnosis ("Healthy controls") or as part of ongoing clinical care for IBD. In those cases with active inflammation sampling would be from an inflamed area and adjacent non-inflamed tissue. UC samples for scRNA-seq was obtained from immunotherapy naïve patients with a proven histological diagnosis. Full details of adult samples involved in this study is details in **Appendix A.2**

Samples from adult tissue were processed fresh and were transported to the lab on ice in transport media (HBSS media supplemented with penicillin, streptomycin and 10mM HEPES). They were then transferred to 5ml pre-warmed (37°C) chelation medium (transport media supplemented with 1mM EDTA) for 80 minutes with regular agitation. The supernatant, which contained epithelial crypts, was removed and chelation media replaced every 20 minutes. Each supernatant was spun down (300G, 5 minutes – similar for all digestion steps unless detailed otherwise) and re-suspended in 2ml transport medium at 4°C until four pooled fractions were merged. This was then centrifuged and placed in TrypLE Express containing 50ug/ml DNAse for 1 hour at 37°C on constant rotation (MACSMix). The resulting single cell suspension was washed and filtered through a 70µm and 40µm filter. Cell counts and viability were checked with a Countess II automated cell counter (Thermo Fisher) with confirmation by manual haemocytometer if required. The epithelial enriched suspension would then immediately undergo scRNA-seq,

FACS or other downstream technique as required. Following publication of results for human intestinal biopsies[2] full detailed methods with step by step processes and key resources were also made available to aid reproducibility[339].

Additional adult samples were collected for comparison to developmental tissue from surgical patients using ST. This included patients undergoing non-emergency intestinal surgery for resection of IBD pathology (e.g. stricture) or resection for colorectal cancer (with sampling taken at non-oncological area). Patients underwent informed consent (REC: 18/WM/0237) with small amount of tissue (e.g. $2cm^2$) being placed in RPMI (Sigma Aldrich) and on ice for transfer to the lab for further processing in the same manner as fetal tissue with one additional step being the removal of muscularis layers which was done manually with cutting aided by dissection microscope or magnification lens.

For embryological samples tissues of fetal intestine were initially collected, after consent, and processed at the Human Developmental Biology Resource (HDBR) London (HDBR project 200462, REC 18/LO/0822). This involved entire embryos or piecemeal products being divided by organ among projects, and intestinal tissue being place in Lebowitz medium (L-15, Sigma) and sent immediately to Oxford on ice for further processing.

In Oxford, all intestinal tissue was examined with operating microscope or magnifying lens to identify anatomical landmarks (stomach, Meckel's diverticulum and/or appendix) and these were used to divide sampling among the terminal ileum, proximal colon and distal colon. In instances of low gestational age (e.g. <12PCW) where the entire colon represented a small amount of tissue (e.g. <2cm) the entire colon would be processed and termed "hindgut". In samples that were piecemeal in nature and the distal intestine (TI-distal colon) could not be confidently identified in continuity. Some tissue

without macroscopic landmarks would be placed in formalin to check signs of anatomical location with a paediatric pathologist at a later date.

Once anatomy was defined, these tissue areas would be separated and washed out with cold PBS and then either; proceeded directly to digestion; cryopreserved in CryoStorCS10 (Sigma Aldrich) and frozen to -80°C for later digestion; fixed in 10% neutral buffered formalin for 48 hours then processed for paraffin embedding; embedded in OCT media (Thermo Scientific) by submersion in isopentane (2-methylbutane, Sigma Aldrich) pre-cooled to -80°C in dry ice for Spatial Transcriptomics (**Methods: 6.9**).

Digestion of tissue used an adapted protocol with similar buffers encompassing that previously described for adult biopsies above, and subsequent digestion of epithelial crypt depleted tissue using a stromal digestion protocol[195]. In an initial optimisation experiment a full digestion without crypt disassociation was tried but had a low yield of epithelial cells (**Appendix A.14**). In detail this protocol entailed direct processing of fresh samples; or for cryopreserved specimens thawing for 2 minutes in 37°C waterbath and washing with 30ml of transport media (as per adult biopsies) supplemented with 5% FCS before centrifugation and processing both conditions in the same manner. Samples would then undergo crypt dissociation with amendment of the timing to two 20min incubations and increasing the EDTA concentration of chelation medium to 5nM. The crypt enriched supernatants would be pooled in a similar manner to adult, and undergo single cell disassociation with TrypLE but the remaining tissue would be immediately placed in umbilical cord digestion kit (Milteyni Biotech) at 37°C for 45 minutes with regular disruption using a blunt needle and syringe until a single cell suspension was seen under microscope.

Both the epithelial (EPCAM+) and non-epithelial (EPCAM-) digestions strategies would then be processed in tandem undergoing two washes with transport medium supplemented with 5% FCS, sequential filtering through a 70μM and 40μM filter pre-

treated with FCS and then being re-suspended in media and counted twice to ensure high (>75%) viability (Countess II, Thermo Fisher). This method was validated with flow cytometry on n=4 occasions, and then in subsequent runs would proceed directly to downstream processing. This involved up to 500,000 cells being brought forward into a hashtag-oligo (HTO)antibody stain; whereby an antibody was conjugated to a custom oligonucleotide that could be recovered during library preparation and sequenced[249]. This was performed with either a commercially available HTO antibody (Total seq, BioLegend) or, in later runs, with a combination of that and an in house custom HTO created from and EpCAM antibody (BioLegend) that was divided into 10µg aliquots and conjugated to 9 oligos using iEDDA-click chemistry in keeping with a published methods[249]. Staining was performed for 30 minutes at 4°C and then washed with PBS supplemented with 0.04% BSA before repeated count, viability check and re-suspension at 1,000 cells/µl and transported to single cell facility for pooling at the same time as preparation of 10x reagents (**Methods 6.3**).

6.2 Flow Cytometry analysis and sorting

Before processing samples for scRNA-seq purity of epithelial (adult) or epithelial and non-epithelial (fetal) was confirmed using flow cytometry. FACS was also used for sorting of populations of interest in specific experiments. For this cells were digested in a similar manner to those described, then centrifuged and washed in FACS buffer (Sterile PBS supplemented with 2% FBS and 2mM EDTA) before staining with appropriate antibody cocktail for 30 minutes at 4°C. In the case of BEST4+ staining and sorting a secondary antibody was added (IgG Rabbit anti Goat AF488, InVitrogen) for a second 25 minute stain. In both instances antibodies were then washed off and a viability stain (DAPI, BD

Bioscience) for 5 minutes before either flow cytometry (Attune NTx, ThermoFisher) or FACS (SonySH800, Sony).

When processing samples compensations were generated with singularly stained samples or in sorting experiments fluorescence minus one (FMO) samples were used to delineate sorting gates with at least 20,000 cells processed. Flow cytometry data was analysed with FlowJo (FlowJo v10.5.3)

6.3 Droplet based scRNA-seq

For adult epithelial experiments cells were centrifuged and re-suspended in PBS supplemented with 0.04% BSA and loaded onto the 10X Chromium single cell platform at a concentration of 1,000cells/µl as described in the manufacturer's instructions (10x Genomics, Single cell 3' kit version 2). On average 8,000 cells were loaded in each reaction.

Generation of single cells within gel beads in emulsion (GEMs), barcoding, GEM reverse transcription clean up, complimentary DNA amplification and library quantification and pooling were performed as per manufacturer's instructions. At key steps library quality was assessed using Bioanalyzer tapestation (Agilent, DS high sensitivity) and concentration measured using Qubit 2.0 (ThermoFisher). The final libraries (n=9) were pooled at a concentration of 4nM with subsequent library pool being denatured and diluted for analysis on Novaseq 6000 (Illumina) using 150 base-pair pair-end reads.

A total of 11,175 cells were captured across n=9 adult samples and passes QC (**Methods 6.11**) with an average of 1,400 cells per sample. The mean read depth of sequencing was 163,882 reads and 1,736 genes per cell.

For fetal samples digested single cell suspensions, with EPCAM+ and EPCAM- fractions for each tissue, were pooled at a ratio of 1:1 with up to 9 samples identified by HTO/HTOs run in each reaction whilst setting up the reagents for the first step of 10x Chromium (during GEM thawing). They pooled samples were then run with 30-37,000 cells loaded per experimental reaction. The protocol proceeded as per the manufacturer's instructions (10x genomics, 3' v3 chemistry revision 3) including generation of GEMs, GEM reverse transcription, clean up, complimentary cDNA amplification, library construction and index PCR. Some modifications were required to retrieve the HTO sequencing libraries which included addition of the HTO additive primer at initial cDNA amplification, HTO-containing supernatant being removed during SPRI bead clean up and these small fragments being cleaned up, amplified and indexed with custom Tru-seq indexes as described in original protocol[249] with 10 cycles of PCR amplification. Library quantification and pooling was done in a similar manner to adult samples. Sequencing was performed on a Nextseq or Novaseq6000 (Illumina) using 150base-pair paired end reads with sequencing depth determined on expected cell recovery or a separate run of HTO library for cell count estimate. Final gene expression libraries were sequenced to a depth of 40,000 reads per cell (EPCAM-) and 50,000 reads per cell (EPCAM+) and HTO library to 5% of gene expression depth.

6.4 Plate based scRNA-seq, real time PCR and RNA amplification

For experiments to more deeply characterise BEST4/OTOP2 cells, epithelial cells were isolated from biopsies and stained to undergo FACS. EPCAM+BEST4+ and EPCAM+BEST4- cells were isolated at a ratio of 1:1 and sorted into 96 well plates to undergo Smart-seq II protocol as previously published[5]. Cells were sorted into lysis buffer

(0.4% Triton X and RNAse inhibitor (Clonotech), dNTPs (10mM) and oligo dT (7.5uM)) with a negative control in one well per plate. Plates were immediately spun after sorting and placed on dry ice, with storage at -80°C until further processing.

Reverse transcription was carried out with four plates in tandem using the published reaction constituents[5] with 0.75 units of SMARTScribe (Clonotech) per reaction, followed by PCR pre-amplification with 5'-biotinylated IS PCR primers (Biomers) for 25 cycles. Post-PCR cleanup was done with Ampure XP beads (Beckman Coulter) at a ratio of 0.8/1 beads:cDNA. Cleaned cDNA was eluted (Qiagen) and QC performed with high sensitivity DS DNA chip (Agilent) on random wells and negative control before library preparation.

Library preparation was then performed using barcoded Illumina sequencing libraries (Nextera XT, Library preparation kit, Illumina) using an automated robotic platform for all 384 samples (Biomek FXp), and libraries were pooled and sequenced using Illumina nextseq 150 cycle hi-output kit (Illumina).

A similar method to SMARTSeq II was used for cells with low numbers sorted (e.g. 12,500- 25000 cells) with the alteration that RNA extraction done using an RNAeasy MicroKit (Qiagen) on sorted cells with on membrane RNA extraction, clean up and elution as per manufacturer's instructions. 1µl of extracted RNA would then be placed into pre-PCR cocktail and PCR cycles would be reduced to 20[5]. Then, after bead clean up cDNA libraries would be used for qPCR as described below.

For microfluidic qPCR of very small bulk cell numbers, to assess signature of the FACS strategy, 100 BEST4+ and 100+ BEST4- cells were isolated from three biological replicates of biopsies that underwent crypt digestion. RNA was amplified using a specific targeted amplification strategy with specified gene primers (Taqman, ThermoFisher) being present in the RT mix as per manufacturers protocol (Biomark, Fluidigm) with a negative

control containing no reverse transcriptase. This was then run on an integrated fluidic chip (Flex 6, Biomark HD) for quantification against a panel of 12 genes.

For qPCR with larger cell numbers (bulk >25,000 cells) total RNA would be isolated with the RNeasy minikit (Qiagen) according to manufacturer's instructions. cDNA was then synthesized using the high capacity RNA-to-cDNA kit (ThermoFisher). cDNA would then undergo qPCR using applicable Taqman primer for gene of interest and quantification on the QuantStudio 7-Flex system (ThermoFisher).

6.5 Proteomic analysis of *BEST4/OTOP2* cells

Proteomic analysis (**Figure 3.3**) was undertaken by using FACS to isolate EPCAM+BEST4+ and EPCAM+BEST4- cells at a 1:1 ration on the SonySH800 cell sorted as previously described. 6,250 cells of each population were sorted into 25 µl of lysis buffer (radioimmunoprecipitation assay (RIPA) buffer with 4% NP-40 (IPEGAL, Sigma)), being centrifuges and frozen to -80 °C. Samples were then transferred to the Target Discovery Institute (TDI), Oxford with subsequent steps being performed by S. Davis. These included thawing, addition of 1µl of benzonase (E1014, Sigma) and incubation on ice for 30mins. The protein lysates were then digested using a published SP3 protocol[340] and method adaptations that have been previously published to break down and purify peptides[2]. The peptides were analysed by nanoscale ultraperformance liquid chromatography coupled to tandem mass spectrometry (Dionex Ultimate 3000 couples to Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific)) with setting previously published[2].

6.6 pH imaging

pH imaging was carried out in a manner to previously published work[229]. Sorted EPCAM+BEST4- and EPCAM+BEST4+ cells were plated onto a poly-L-lysine coated coverslip and loaded with pHrodo Red AM as per manufacturer's instructions (Molecular Probes). Imaging was performed in collaboration with Dr C Lagerholm (Wolfson Imaging facility) using an Olympus DeltaVision II Microscope System. pHrodo fluorescence intensity was measured in response to pH5.0 solution (MES; 150mM NaCl, 10mM MES, 2mM $CaCl_2$) with intensity normalised to baseline fluorescence in pH 7.4 solution (F_0) before application to determine pH shift (F/F_0). Cells were measured at pH 4.5, 5.5, 6.5 and 7.5 to extrapolate a standard curve for pH_i quantification.

6.7 IHC, IF and smISH

Following informed consent tissue samples being used for IHC, IF and smISH would be collected from adults undergoing endoscopic assessment, or embryonic intestinal tissue as previously described. Samples would be washed and then placed in a cassette for fixation with 10% neutral buffered formalin for 48 hours, before transfer to 70% ethanol and subsequent batch paraffin embedding.

For IHC, paraffin embedded tissue was cut onto a slide at 5μm using a sectioning microtome. These were dried (37°C overnight) and used for long term storage. When stained, tissue sections were deparaffinized by passage through an ethanol gradient and then heat induced epitope retrieval (HIER) was performed at by boiling in a ph6 (Citrate) or ph9 (Tris/EDTA) buffer at 96°C for 25 minutes. Choice of buffer was dependent on antibody following optimisation. Peroxidase was blocked and primary antibody staining

was undertaken for 90 minutes at room temperature before addition of species-specific secondary antibody (Impact, Vector). The stain intensity was then developed using either a single or dual peroxidase / alkaline phosphatase kit as per manufacturer's instructions (Impact, Vector) before mounting (Surgipath mounting medium). Wash steps were performed between each of the steps with PBS prior to primary antibody staining or PBS supplemented with 0.1% TWEEN after this.

Haemotoxylin and Eosin staining was carried out for mapping of intestinal morphology or confirmation of structures/ clinical features. This involved processing and deparaffinisation in a similar manner, then sequential addition of Haemotoxylin, bluing buffer and eosin solution from a kit as per manufacturer's instructions (H&E Kit, Vector laboratories). In addition, haemotoxylin was added for 1 minute before mounting IHC stains when single stain was used.

For smISH all probes were purchased from Advanced Cell Diagnostics and developed with RNAscope 2.5HD assay-brown as per instructions (ACD, Bio-techne). In brief, this consisted of a sequential pre-treatment (Pretreat 1,2 and 3) followed by addition of warmed (40°C) probes with incubation for 2 hours at 40°C in HybEZ oven (ACD) then undergoing a six step signal amplification protocol and detection with DAB prior to Haemotoxylin counterstaining and mounting (Pertex mounting medium).

Double staining of smISH and IHC was performed in the same manner as smISH protocol with subsequent overnight staining of primary antibody and development of stain with VectorBlue (Vector).

6.8 Quantification of IHC staining intensity and bacterial activity

For quantification of WFDC2 intensity a cohort of patient slides underwent IHC staining for WFDC2 as described above. These were then scanned using a Leica ScanScope machine (Leica Biosystems). Image intensity quantification was done using Visiopharm (Visiopharm) with collaboration with N. Khalid Alham (Nuffield Department of Surgical Sciences) which entailed a programmed protocol calculated as follows: percentage of positive goblet-cell area = fraction of goblet-cell area × 100, in which the fraction of goblet-cell area was calculated as the goblet-cell area (area inside region of interest (ROI) 1, set on a defined image) as a fraction of the area of interest (total ROI2 in a defined image). The protocol was set to repeat the calculation randomly for 50% of the whole biopsy to generate an average. The result was equated to the percentage of positive/brown stain in goblet cells for each biopsy.

For quantification of WFDC2 antibacterial activity, experiments were performed in collaboration with M. Jagielowicz and A. Aulicino. Mid-logarithmic-phase cultures of ATCC12973 *S. aureus*, ATCC13379 *E. faecalis*, ATCC27853 *P. aeruginosa*, *S. Typhimurium LT2* and *E. coli* were incubated with diluted PBS containing Lucia Bertani (LB), tryptic soy broth (TSB) or brain and heart broth (all Sigma-Aldrich) to a concentration of 2×10^4 colony-forming units per ml. A final volume of 100 µl was used to measure antibacterial activity of recombinant WFDC2 by addition at a concentration of between 0.45–4.50 µM. Following a 4-h incubation at 37 °C, the frequency of bacteria was determined by serial dilution. Survival was calculated as the percentage of bacteria present at 4 h compared with baseline with all experiments being performed three times.

Imaging of *Wfdc2*$^{+/-}$ mice colons for adherence of bacteria was done in collaboration with E. Johnson (Sir William Dunn School of Pathology) with methods previously published[2].

6.9 Spatial transcriptomics

For ST fetal and adult samples were processed and frozen into OCT as previously described (**Methods 6.1**) then stored at -80°C in an air tight container until processed further.

Before undertaking a full ST protocol the method of freezing and RNA isolation was confirmed with testing RNA quality of blocks to be >7.0 RIN value (RNA Pico kit, Agilent) and then three tissue optimisation experiments were performed on adult and fetal tissue with a 10μM section cut in a cryostat onto and array that contained spatially organised oligonucleotides and processed as per manufacturer's instructions (10x Genomics, Visium Spatial Tissue Optimization, Rev A). The fluorescence footprint of on-slide amplified cDNA was visualised on the InCell6000 analyser (GE Healthcare) and image analysis was performed on FiJi (ImageJ v2.0.0) identifying 18 minutes as the optimum time for permeabilization.

Once optimum digestion time was identified, subsequent samples were then processed for full ST experiment as per manufacturer's instructions (10x Genomics, Visium Spatial, Rev B). Cut at 10 μM thickness onto four 6.5mm x 6.5mm capture areas with 5000 oligo-barcoded spots with spatial identifiers. For some samples undertaken in the first run (Adult, 12PCW TI and Colon) n = 2 technical replicates were performed whilst establishing the technique. Following sectioning the slides underwent methanol fixation and H&E staining with immediate imaging on Aperio Scanscope (Leica Biosystems) at 40x magnification. Tissue underwent permeabilization with proprietary enzyme at the identified optimum timepoint (18 minutes), reverse transcription and second strand synthesis performed on the slide with cDNA quantification undertaken with qPCR using KAPA SYBR FAST-qPCR kit (KAPA Biosystems) and targeted primers with

amplification traces analysed on the QuantStudio 7-Flex system (ThermoFisher). qPCR results (Cq value at 25% of peak fluorescence) which informed cDNA amplification.

Libraries were constructed as per protocol and quantified using either the KAPA-Illumina PCR quantification kit (KAPA Biosystems) or QuBit 3.0 (Thermo Fisher). Samples were pooled at 4nM concentration with a sample ratio corresponding to the surface area of tissue coverage obtained from the H&E imaging calculated as percentage of array marked by area in FiJi (ImageJ v2.0.0).

Pooled libraries were sequenced on a NextSeq (Illumina) using 150 base-pair paired-end dual-indexed set up (High output, v 2.5, Illumina) loaded at a concentration of 1.8pM. Four slides were sequenced to a manufacturer recommended depth of ~50,000 reads per tissue covered spot (mean achieved: 52,714 reads). Four subsequent slides were deep sequenced to the depth of 362,034, 364,566, 143,014 and 183,966 mean reads per tissue-covered spot in order to increase sequencing saturation and detection rates of low expression transcripts. This generated a total of 9330 tissue-covered spots across slides (average 2480 genes detected per spot or 3877 mean genes for deep sequencing slides). Each spot covered a 55μm area with a 100μm centre-to-centre distance, expected to encompass 6-10 total cells based on study of the H&E images. Areas highlighted as anatomical reference points in results had images discussed with a paediatric pathologist to confirm annotation.

6.10 Detailed characterisation of intestinal scRNA-seq clusters

Due to the large number of cell types and states (n=101) identified in the fetal scRNA-seq atlas an overview of the annotation rationale was placed here rather than in

main text. A full overview of annotations and marker genes is presented in **Appendix: A.14**. The literature and conclusions to support these are as follows.

Compartment annotations: Following initial QC all cells were studied as a whole to identify the cellular compartments of origin. This was divided based on known features and key marker genes, so that each could be focused for separate analyses. Compartments included differential expression of epithelial (*EPCAM, FABP1*), fibroblast (*THY1, COL1A2, VIM*), myofibroblast (*VIM, FOXF1, TAGLN*), endothelium (*PECAM1, CDH5, CLDN5*), pericyte (*KCNJ8, ABCC9, TGS5*), neural/glial (*PHOX2B, HAND2, TUBB2B*), immune (*PTPRC*) and muscle (*MYH11, ACTG2*) genes with each separated to study their constituent cell types.

Epithelial annotations: There were 15 epithelial clusters identified and annotated by their gene expression profile from literature of from the initial epithelial assessment. There were annotated as; Enteroendocrine cells (*CHGA, TPH1* and *NEUROD1*)[9], transit-amplifying cells (*MKI67, UBE2C* and *TOP2A*)[2], enterocytes (*FABP2, CEACAM1* and *EPCAM*)[199], goblet cells (*MUC2, SPDEF* and *WFDC2*)[2,43] and ISCs (*LGR5, ASCL2* and *SMOC2*)[1]. One cluster expressed genes with a distinct transcriptional signature in line with BEST4/OTOP2 cells (*BEST4*, *CA7, CA4*). Alongside the secretory population, a cluster expressed genes described in EEC (*NEUROD1*)[39] and goblet cell (*DLL1*) progenitors[65]. *FOXA2* expression was high which matched a combined secretory lineage and so was annotated as secretory progenitor[341].

Remaining differences in clusters were annotated as per location of cells of origin (proximal, distal), or maturity of enterocytes with a spectrum of maturity markers (*AQP8, FABP2*) from TA cells to higher expression in mature cells.

In this initial characterization intra-cluster expression of specific genes, most notably EEC hormones, could still be seen. Thus the secretory lineage was also analysed in

isolation this generated a further 11 clusters which included an EEC progenitor marked by *NEUROG3* and other genes similar to transcriptional analysis of organoid modelling[9,205]. Goblet cells divided into 3 clusters which appeared to represent a spectrum of goblet cell-types from cycling (Goblet 1, *UBE2C, MKI67*) to primitive stem-cell like (Goblet 2, *MYC, TRAB2A* and low expression *LGR5*) and relatively higher expression of mature goblet cell markers (Goblet 3, *MUC2, TFF3, SPDEF*). The EECs divided into 6 cluster with key hormones mimicking previously reported signatures[9,205]; Somatostatin (*SST*, D-cells), serotonin (*TPH1*, Enterochromaffin cells), CCK (I cells), PYY/ glucagon (*GCG* L-Cells), Neurotensin (*NTS*, N cells) and co-expression of Motilin and Ghrelin was seen in the same cluster and annotated as A-Cell/M-Cells similar to other studies. The final cluster was small intestinal specific and expressed defensins (*DEFA5*, *DEFA6*) in keeping with Paneth cells[192].

Fibroblast annotations: Fibroblast cells separated computationally into 16 distinct clusters. Two clusters demonstrated high expression of cell cycle phase genes and so were annotated Fibroblast G2M phase (*MKI67, TOP2A*) and Fibroblast S-phase (*PCLAF, TYMS*) in line with these. There was a small distinct number of contaminating erythroid cells (*HBB, GYPA*) that were clearly discrete and excluded from downstream analysis. The remaining fibroblast clusters were divided into fibroblast subtypes which mimicked those seen in scRNA-seq of adult intestinal tissue and were annotated as such[195]; fibroblast S1 (*ADAMDEC1, FABP5*), S2 (*FRZB, BMP5*), S3 (*C7, ASPN*) and S4 (*CCL19, C3*). Within these further transcriptional diversity was seen and annotations were then based on locational differences, such as in S2 (Distal S2, Proximal S2 1 and Proximal S2 2), or on key genes that differentiated sub-types – *COL6A5+* and *IFIT3* population in S1, and *CXCL13* and *CCL21* in S4 respectively. A relatively larger proportion of fibroblasts were of the S3 type and this was reflected in 5 clusters expressing S3 genes but differing in gene

expression, two populations were annotated on higher expression of transcription factors *HAND1* and *EBF* (*EBF3*) respectively. One S3 population had relatively lower expression of S3 marker genes and was placed between the fibroblast progenitor and mature S3 populations on trajectory analysis thus annotating these as "S3 Transitional" (**Appendix A.14**)

There was one cluster that did not express fibroblast subset-specific markers and had relatively lower expression of mature fibroblast genes (*THY1, IGFBP4*). Alongside this, the frequency of cells in this cluster was higher in early (< 12PCW) time points and one of the highly expressed genes *HMGA2* has been shown to be involved in embryonic fibroblast formation[342] leading to the annotation of "Fibroblast progenitor." Similar to this, an adjacent cluster had a relatively lower expression of S3 type genes (*OGN, FBLN1*) and was represented in earlier time-points leading to the annotation of S3 progenitor.

Myofibroblast / mesothelial annotations: Within the group annotated as myofibroblast cells there was a population representing cells within the G2M (*CENPF, MKI67*) and S-phase (*PCLAF, TYMS*) of cell cycle and were thus annotated accordingly. In the remaining clusters, 2 were characterized by *WT1* and *MSLN* which are both reported as markers of mesothelial layer in the early formation of abdominal viscera[109,155], so these clusters were annotated based on their marker genes; *SOX6* and *IL18* mesothelium which may arise due to locational contribution of the serosa. The remaining two clusters expressed more classical myofibroblast genes such as *NKX-3* and *ACTG2*[78,95] but one had higher expression of the mature markers *MYH11*/*ACTA2* and contained more cells from later in the developmental times samples, leading to the annotations of "myofibroblast progenitor" and "myofibroblast."

Endothelial annotations: There were 11 clusters that comprised the EC compartment and could be divided into five venous clusters which expressed known EC vein specific genes

such as *ACKR1* and *VWF*[343]. There were four arterial clusters expressing genes such as *GJA4* and *HEY1*[305].

Further annotation of these clusters related to expression of genes reported to reflect the size of the vessel in mice, which appeared to be recapitulated in the data. In venous EC's genes reported to be more highly expressed in capillaries (*CD83, RGCC*) supported the annotation of "venous capillaries (CP)" and contrasted to large vein specific genes such as *ACKR1* or *ADGRG6*[305,343].

Similarly, in arterial clusters expression was seen ranging from large arterial vessel genes (*HEY1, GJA5*) through to genes reported in capillary endothelial cells (*VWA1, RBP7*)[305,344,345]. Annotation in turn reflected this with large(L), medium sized (M) or two capillary (CP1, CP2) endothelial cell clusters.

Alongside cells expressing cell cycle genes (*MKI67, CENPF*) one further group of cells was distinct from the vascular ECs and had specific expression of genes such as *PROX1, PDPN* and very high expression of *LYVE1* which are known to be markers of lymphatic ECs in the murine intestine and so annotation reflected this[346].

Pericyte annotations: There were eight clusters of cells identified in the pericyte compartment, with two expressing genes identifying them as cycling (*MKI67, PCLAF*). One distinct group of cells was present mainly within the colon and had high expression of genes reported in pericytes involved in angiogenesis - *PRRX1* and *PROCR* - so was termed angiogenic colonic pericytes[347348]. The remaining cells formed a continuity of maturation from an undifferentiated pericyte population, with relatively low expression of pericyte marker genes, through to cells expressing muscle related genes (*MYH11, ACTG2*, "contractile pericytes"), or fibroblast-like genes (*THY1, SPON2, BMPER* "BMPER+ pericytes"). Alongside the undifferentiated cluster, a group of cells expressed *THBS1* that is seen in early differentiating pericytes[349] and this, in addition to the majority of cells

being from early (< 12PCW) time points led to the annotation "pericyte progenitors." *WNT6*, which has been shown in animal models to be important in muscle differentiation[350], was specifically expressed in cells placed between the undifferentiated and contractile pericytes and this may reflect intestinal pericyte development in this intermediate stage between differentiating pericytes (with loss of *THBS1*) and so this cluster was annotated "WNT6 pericytes."

Neural annotations: Along with a cycling cluster (*MKI67, CENPF*) there were 5 clusters of cells that expressed genes matching glial (*SOX10, S100B*), and 7 neural (*ELAVL4, CHRNA3*) cell types respectively.

Within glial cells there was a cluster with relatively lower expression of mature glial markers, represented by predominantly early time point samples. It also expressed *PHOX2B* which is seen in ENS progenitors and was thus termed "glial progenitors"[261]. *ENTPD2* expression has previously been reported in intra-ganglionic glia, so cells expressing this were annotated as such[351]. A separate cluster had relatively high expression of *TGFB1*, which is expressed by subsets of glial that are closely adherent to the epithelial surface. The function of this cluster was further supported by GOE which highlighted terms including "regulation of epithelial cell proliferation" and "epithelial cell proliferation." This cluster was thus termed "submucosal glia" and a separate cluster with lower expression of similar genes, an increase in early-time point cells and glial-specific expression of *HAND2* (reported to be important in glial precursors in mural intestinal ENS development) led to these cells being annotated as "submucosal precursors"[352]. The final glial cell cluster was transcriptionally distinct from other glial cell types and exhibited high expression of genes involved in immune infiltrate formation (*FGL2*), T cell transduction (*MAL*) and expressed TGF-beta receptors (*TGFBR3*) highlighting the possibility of these cells interacting with immune cells[353]. Furthermore this cluster expressed *GFRA3, ARTN*

and retinoic acid receptor (*RXRG*) which in mice are vital for Peyer's patch development which led to this cluster being termed "lymphoid associated glial"[268].

For neuronal clusters; interneurons were identified by expression of tachkinin (*TAC1*) and enkephalin (*PENK*), inhibitory motor neuron by expression of nitric oxide (*NOS1*) and VIP and excitory motor neurons by expression of *CASZ1*[351,354], *SLC5A7* and *GFRA2* – the latter two identified as part of the core transcriptional signature of excitory neurons in a recent scRNA-seq study of the ENS[307] . A cluster adjacent to the inhibitory motor cells expressed similar genes such as *SCGN* and *NOS1* at a lower level and was thus annotated as "inhibitory motor neuron precursor." Two cluster showed mature neuronal markers and specifically expressed a number of markers that matched enteroendocrine circuits including neuropeptides (*SST, NPY*) alongside their receptors (*SSTR2, NPY2R*) so were termed neuroendocrine 1 and neuroendocrine 2. The final neural cluster had a relatively lower expression of neuronal markers such as *ELAV4* and expressed genes seen in primitive neuronal progenitors such as *DLL3* and *DLL4* and annotated as "ENS progenitors"[262].

Immune annotations: Immune cells, expressing pan-immune markers (*PTPRC*) identified 12 clusters. Alongside a cycling cluster (*MKI67, CENPF*) cells were further subdivided based on 5 lymphoid clusters (Adaptive and innate) 5 of myeloid origin (monocyte, macrophage, DCs) with a final cluster exhibiting a distinct signature in keeping with intestinal mast cells (*TPSAB1, TPSB2, CPA3*)[2,251].

Specifically, monocytes were identified by high expression of CD14 in comparison to macrophages as well as *FCN1* a key gene expressed in colonic monocytes in colorectal cancer[355] alongside transcription of Myeloid Inhibitory C-type Lectin like receptor (MICL encoded by *CLEC12A*) reported in this cell type[356]. Macrophages exhibited expression or reported genes and markers such as *MERTK*[357] as well as Cathepsin C and D (*CTSC,*

CTSD)[300]. Two macrophage clusters differed in expression of SPP1, a recently described gene identifying polarized macrophages termed SPP1+ Macrophages[358], and the second also expressed genes involved in complement formation (*C1QA, C1QC*) and *IL1B* at a higher level – in keeping with a classically activated (M1) phenotype[355,359]. Two clusters represented DCs, by expression of the pan-marker *ETV6*[360] were divided into conventional DC (*FLT3*) and plasmocytic DCs (*PTCRA, LILRA4*)[300,361].

In the lymphoid compartment, B cells were seen with expression of genes including *CD19*, and clustered into cells which had more mature B cells markers (*CD19, CD79A, IGHD*) and a cluster with a more naive appearance; relatively lower expression of *CD19*, expressing light chain (*IGLL1*), and other genes reported in naive B cells such as *VPREB3*[362,363]. This and their appearance in early gestational age led to the annotation 'Pre-B cell'[364].

Three cell clusters had signatures in keeping with innate lymphoid cells, the first expressing markers of a naive T cell phenotype (*CD3G, CCR7, CD27*) and another representing a Gamma delta cells (*TRDC*), and NK cells (*IL2RB*)[300]. The final cluster expressed *KIT* highlighting it as a ILC cell which together with *KLRB1*, *RORC* and *ID2* demonstrated this to be of the ILC-3 subtype[300]. The expression signature, timing of appearance and interaction with other compartments highlighted this cell type as representing the Lymphoid Tissue Inducer (LTi) population of ILC-3 cells.

Muscle annotations: The muscle compartment comprised of 11 clusters. Two were readily identifiable as cycling in G2M (*MKI67, CENPF*) or S-phase (*PCLAF, TYMS*). Within the compartment there was a clear temporal divide, with two clusters being represented by predominantly early time point (< 12PCW) samples. The first expressed *PDGFRA* and *KIT* which was in keeping with PDGFRA+ interstitial cells[365], it was also closely related to a second cluster expressing *KIT* alongside *ANO1* and *SPON2* –

identifying ICCs[366]. The second early time-point cluster expressed primitive genes such as *HOX5C* alongside myocardin (*MYOCD*) which is a critical factor in early smooth muscle differentiation[367]. Along with low expression of *FOXF2*, a marker identifying the intestines internal musculature[368], highlighted this cluster as an internal muscle progenitor. Similar to IM progenitor *FOXF2* was expressed in a gradient with highest expression in two clusters termed proximal and distal muscularis mucosae, supported by expression of *IGF1*, *HHIP* and *NOG* which are known to be key to smooth muscle and lamina propria development mechanisms respectively[369,370]. Another inner muscle cluster specifically expressed *PMAIP1* (aka *NOXA*) which is important in smooth muscle differentiation, so annotation reflected this[371]. The other two IM clusters had intermediate expression of *FOXF2* and a lower expression of myofibroblast-like marker *VIM*, compared to muscularis mucosae, suggesting a slightly more peripheral placement and so was termed IM proximal and IM distal on locational differences. The final two clusters were negative for the internal muscular markers *FOXF2* and *ACTC1* and expressed *IGFBP5* which found at a higher level in outer intestinal muscle thus leading to the annotations proximal and distal outer muscularis (OM) based on locational differences[372,373].

6.11 Computational processing and analyses

As highlighted in the attributions section, the work in analysing the data generated in this work was done by Dr. A. Antanaviciute in close collaboration. A full overview of the methods used along with additional analysis (e.g. GWAS cross references with IBD epithelium) is now published in the associated papers for the adult epithelial[2] and fetal intestinal[3]. Raw sequencing data is also available by being deposited on the gene expression omnibus (GEO): Epithelial data - GSE116222; Fetal scRNA-seq -

GSE158702; Fetal ST - GSE158328. An interactive searchable resource of the fetal survey is available at: https://simmonslab.shinyapps.io/FetalAtlasDataPortal/. A brief overview of computational methods is as follows.

Raw 10x scRNA-seq and ST processing and QC: Raw sequence reads were demultiplexed from Illumina sequencer were quality-checked (FastQC software). Cell Ranger (10x Genomics) version 2.1.1 (Epithelial) or 3.1.0 (Fetal) was used to process, align and summarize unique molecular identifier (UMI) counts against the human genome (hg38 assembly reference genome analysis set)[374] to generate gene annotations (Ensembl). A similar process was undertaken for ST sequencing data using Spaceranger software (10x Genomics) with UMI counts generated against hg38 reference for each spot on the Visium ST array. In fetal work antibody libraries were processed separately using CITE-seq Count (Version 1.4.3) to obtain antibody tag UMI counts summarizing both TotalSeq Antibody libraries and in house conjugated libraries (**Appendix: A.14**) which were filtered against 10x cellular whitelist.

Count matrices were imported into R. Droplet barcodes were quality controlled for: empty droplets; index swapping; high mitochondrial reads; low total UMI counts[2,3]. This was done on a compartment based level in fetal work.

Seurat[375] was used to: normalise expression for UMIs/cell; identification of highly variable genes; determination of principle components in clustering; cell clustering as *t*-sne or UMAP. Cells from separate pools were merged and batch effect signal was corrected using harmony[376] with comparisons of cell types across pools to ensure heterogeneity was not lost. Merged data was clustered (epithelial) or clustered into compartments based on marker gene expression then into intra-compartmental cells (fetal).

Crypt axis score (Epithelial): The genes used to define the crypt-axis score included: *SEPP1, CEACAM7, PLAC8, CEACAM1, TSPAN1, CEACAM5, CEACAM6, IFI27,*

DHRS9, KRT20, RHOC, CD177, PKIB, HPGD and *LYPD8*. Each gene was normalised for expression across all cells to range between 0 and 1, and the score for each cell was defined as the sum of all normalised expression values.

Clustering of public scRNA-seq data and analysis of TCGA and microarray data: For testing if the presence of BEST4/OTOP2 cells in other single cell-datasets, data was downloaded from GEO and processing was done as previously described with the addition of comparisons to *a priori* markers from the epithelial scRNA-seq resource (GEO; accessions GSE103239 and GSE81861). Cluster markers were compared for BEST4/OTOP2 genes.

For TCGA comparisons data was downloaded for all colorectal cancer patients and matched normal tissue (https://cancergenome.nih.gov). Data was normalised and studied for the core BEST4/OTOP2 gene signature[2].

For comparison of publicly available inflamed UC and paired normal tissue data was downloaded (GSE59071) and differential expression analysis performed[2].

Ontology enrichment analysis: Biological-process GOE of cluster markers and differentially expressed genes/proteins was performed using the R package clusterProfiler[377] with an FDR cut-off of 0.05, using all expressed/detected genes as a background control.

Smart-seq2 data processing and analysis: Raw sequencing data were demultiplexed into one fastq file per plate well using bcl2fastq software. Reads were quality checked: FastQC software, Illumina Nextera sequencing adapters and Smart-seq2 oligonucleotide sequences. Poly-d(T) and poor-quality (< 20-base-pair) sequences were trimmed. Reads were aligned to the human hg38 reference. Raw read-count matrices were obtained. Poor-quality wells were filtered on the basis of: <60% of reads uniquely mapped, <500 genes detected, > 20% of reads mapping to mitochondrial features. Library normalization size

factors were computed. A small number of contaminating immune cells were filtered out. BEST4/OTOP2 cell marker genes were identified using the R package Seurat.

Proteomics data analysis: Label-free quantification of proteins was performed using Progenesis QI for Proteomics (version 4.1, Waters) and proteins were identified using MASCOT (Matrix Science) by searching against the Uniprot reference human proteome (95,128 sequences, database accessed on 18 July 2018). Peptide FDR was adjusted to 1% and low-scoring peptides (less than 20) were excluded. The R package limma was used for normalization of protein expression and differential expression analysis[378]. The Benjamini–Hochberg multiple testing correction was used to estimate the FDR.

BEST4/OTOP2 cell marker overlap: The intersection of the top 200 markers for BEST4/OTOP2 cells (10× single-cell data, SmartSeq2 data, quantitative proteomics assays and previous datasets[193,194] was visualized using the R package circos.

Trajectory and pseudo-time analysis: In epithelial work cell-differentiation trajectories were reconstructed using the R package monocle[237]. Non-epithelial-cell clusters were filtered out and dimensionality reduction was performed with cell trajectory reconstructed with DDRTree algorithm using the orderCells function and the starting state was denoted as the branch encompassing the previously identified stem cells at the most distal end.

To identify putative lineage regulators, genes that change between secretory and absorptive branches were identified, modelling pseudo-time as a covariate. Then genes that are induced before the trajectory bifurcation point were seen by performing a differential expression test between the cells in the earliest trajectory state. All significantly upregulated (less than 1% FDR, greater than 0 fold change (expressed in log2)) genes were then intersected with all previously identified genes. This subset identified genes with branch-specific expression that are also induced before lineage

divergence. In all statistical tests, patient/sample batches, cell-cycle scores, cell-size factors and percentage of mitochondrial gene expression were modelled as covariates.

For fetal work RNA velocity estimates were used to predict the directionality of biological processes, comparing the dynamics of spliced and unspliced/nascent RNA using the R packages velocyto.R[379] and SeuratWrappers using batch-corrected harmony dimensionality reduction[3]. Trajectories on compartment embeddings were calculated using Monocle 3 algorithm[237]. For each compartment, the start of the trajectory was selected as the node nearest to the identified progenitor populations and trajectory pseudo-time was then computed from the selected starting node.

To better understand the connectivity structure of 101 single cell populations identified in fetal work, a partition-based graph abstraction algorithm was used[380] using batch-corrected, dimension-reduced data and cell identities as defined from the clustering analyses.

Hashed sample de-multiplexing (Fetal): EPCAM- pools had HTO antibody and UMI matrices filtered against 10x barcodes from droplets passing QC, and then the filtered matrix was used to demultiplex samples as originally described[249]. Cell identity was assigned on hash tag and multiplets identified as cells positive for multiple tags which were filtered out along with cells negative for all hash tags.

This process was modified in the dual labelled EPCAM+ pools where TotalSeq and in-house hashing antibody UMI counts were normalised and clustering used matrices of both tags. When there was disagreement on cell of origin (6-10% cells) it was mainly due to cells from lower gestation where the in house antibody would locate cells not stained by totalseq antibody, or in non-epithelial contaminant cells which was identified by *EPCAM* and *THY1* expression on mRNA. Sample multiplexing was also examined for function by checking sex-specific genes such as *XIST* checked against sample of origin[3].

ST data analysis: Raw UMI count spot matrices, images, spot-image coordinates and scale factors were imported into R. Spot matrix was filtered out to keep only spots overlaying tissue. Raw UMI counts were normalized[381] to account for variability in total spot RNA content. Dimensionality reduction was performed using PCA and the optimum number of principal components was determined. Clustering was performed and clusters were visualized using UMAP. Clusters distributions were visualized over H&E images.

Cell type prediction probabilities were calculated for each spot using factor analysis in Seurat. In addition to single cell data generated here, single cell populations were obtained from adult tissue reference datasets[2,195,250,251] (GEO: GSE114374, GEO: GSE116222, DUOS-000110 & GEO:GSE125970) to predict spot content for all slides.

For all predictions, pairwise cell type prediction probability score were calculated using all tissue covered spots in a given slide to assess spatial cell type co-occurrence within the same spot[3].

ST distance analysis: For distance-based analyses, intestinal tissue in H&E images was demarcated and muscularis mucosa was marked (adult) or serosa (fetal) as a reference point. Spots were filtered for artefacts, including tissue folds. For each remaining spot, we then calculated Euclidean distance from the centre pixel. Adult tissue sample slides were segmented prior to this to account for helical positioning of the tissue. Spots in the adult mucosa were assigned positive distance values to indicate distance toward the lumen, and spots in the submucosa were assigned negative distance values to indicate distance away from the muscularis mucosa. Distance measures were then used to examine cell type score distributions relative to tissue depth for further analysis of cell type and gene expression over depth[3].

Marker gene detection and differential expression: For all marker gene expression, we used R packages Seurat and MAST for statistical testing[382]. For each identified cluster, we

compared the cells within that cluster versus all other cells. To identify markers within sub-populations in compartments, we compared cells from each cluster versus all other cells within a given compartment. Location-specific genes were identified within each cluster by comparing cells from TI or colon tissue samples. In cases where we observed location-driven clustering, we performed comparisons with the nearest location-matched counterpart cluster (e.g., distal stem cells versus proximal stem cells; distal S2 versus proximal S2). In each case, confounding sources of variation stemming from cellular detection rate, pool batch effects and donor genotype were included in the model formula as covariates. For visualizing and thresholding cell type specificity, we calculated gene AUC scores for all cell types using area under the curve analysis.

To identify time-course varying differentially expressed genes in scRNA-Seq data, we used zero inflated negative binomial models[383] to fit single cell count data for each gene and compute their observational weights. DESeq2 was then used to perform differential expression analyses[384].

To identify cell clusters significantly associated with location or time point differences, for each cluster for each sample we normalized the total number of cells detected within the compartment in that sample. To test for compartment-level differences, we normalized to the total number of cells detected for each sample in each 10x pool. Then, the percentages of cells in each sample were compared using a two-sided Wilcoxon test. P values < 0.05 were considered as significantly different. For location-specific clusters, we compared location-matched samples only.

Receptor-ligand analysis: Analysis was performed with receptor-ligand database SingleCellSignalR[385]. We then scored each receptor-ligand pair in our dataset using the scoring procedure described in the package[385]. Both paracrine and autocrine interactions were computed across all pairwise cluster combinations (101x101 = 10201 possible cluster

pairs), resulting in signalling network spanning 4452511 total putative interactions across all clusters involving 2552 unique receptor ligand pairings were identified (654 unique ligands and 646 unique receptors). We further prioritised interactions by assigning a cell type specificity AUC score to each receptor and ligand for every interaction. To visualize interactions between pairs of clusters used circos plots via R package 'circlize' (version 0.4.8). Interaction scores <0.5 and cell type specificity AUC < 0.7 were filtered out. Individual R:L networks were constructed using R package 'ggraph' (version 2.0.3).

Morphogen module identification: We curated a list of morphogens from pathways that contribute to patterning, organogenesis and crypt formation. This included Hedgehog, NOTCH, Wnt, HIPPO, RTK, TGF-Beta, FEGF and EGFR signaling pathways and encompassed their ligands, receptors, co-receptors, antagonists and transcription factors. (**Appendix: A.14**). With the curated gene list, weighted gene co-expression network analysis was performed using R package WGCNA[386]. A signed network was constructed and used to compute a pairwise distance matrix for the morphogens. Dynamic hybrid tree cut method partitioned the genes into modules[3]. A score for each detected module for all spots in ST was generated by averaging expression levels of each module in each spot and subtracting similarly aggregated expression of randomly selected control features. To visualize morphogen pathway interactions, we downloaded high confidence interactions from STRING database[387] for the morphogens. A network was constructed using R package igraph and laid out using a force-directed layout.

Disease gene analysis: HPO[388] was curated to select human disease phenotype terms pertaining to congenital or developmental gastrointestinal disease. The curated HPO phenotypes and ontology annotations were linked with selected phenotypes to Online Mendelian Inheritance in Man (OMIM) and OrphaNet human diseases and their associated causative disease genes. We identified 749 disease genes linked to 18 phenotypes, 718 of

which were found to be at least minimally expressed in our single cell data. AUC analysis identified cell type specific expression for disease gene in each cluster (> 0.8 were considered highly cell type specific). To identify developmental time-course varying disease genes, we used differential expression analyses as described above.

6.12 Data availability

Epithelial work: Raw and processed sequencing data files are deposited under the GEO accession number GSE116222. The source code for analyses has been deposited at https://github.com/agneantanaviciute/colonicepithelium. Proteomics data have been deposited at the ProteomeXchange Consortium with the dataset identifiers PXD011655 and 10.6019/PXD011655.

Fetal work: Raw and processed sequencing data for the project are deposited on GEO and is publicly available (ST: GEO: GSE158328 and scRNA-seq GEO: GSE158702). Interactive analysed data with searchable functions has also been made available as an online resource – the Spatio-Temporal Analysis Resource of Fetal Intestinal Development: STAR-FINDer (https://simmonslab.shinyapps.io/FetalAtlasDataPortal/). STAR-FINDer has features including: gene expression, ST, gene regulator networks, trajectory analysis, time-course differences; morphogen expression; RL interactions.

Appendices

Details of supplementary information to the main thesis are detailed here. For large supplementary information such as gene lists, differentially expressed genes or GO terms only a selection of the analysis (e.g. the top 30 features) are given to save space. Full details are uploaded to a Mendeley Data upload (available from: https://data.mendeley.com/datasets/62z54f5yf3/1 with DOI: 10.17632/62z54f5yf3.1) with signposting by referencing[389] where appropriate.

A.1: COVID-19 log

The COVID-19 pandemic caused disruption to the undertaking of the planned experiments in this DPhil. Although the project and its aims were completed some adaptions were made to experiment plans to account for the lack of laboratory access from national lockdowns or a period of self-isolation. This resulted in delays of planned *in vivo* modelling of the intestine using organoids which therefore did not feature in the thesis, and some other methods. During the pandemic students were requested to keep a contemporary log of how the pandemic affected their laboratory work and this is placed here as a record of how the project was affected.

Event start date	Event end date	Brief description	Level of impact	Additional details
25/3/20	25/3/20	Institute closed for coronavirus	High	The institute closed and advice was given to work from home or complete urgent experiments with permission in the next 48 hours. I remained for this to finish the last sequencing experiment of the developmental resource project (ST).
27/03/20	27/03/20	Last lab experiment	High	This was the last date where I undertook a final sequencing experiment with the aim of generating data to analyse during working from time at home due to national restrictions.
28/03/20	onwards	Working from home	High	During lockdown I have been working from home with analysis and writing of data with the aim of producing a paper draft for my work to date. This alongside generating reading and literature for my eventual thesis. My aim at this time is to hopefully not delay my hand in and to continue working on non-lab work at this time provided the lock down does not go on for too long.
30/06/20	onwards	Approval for socially distanced return to work	Medium	Following the new guidance on return to safe working and new risk assessments I undertook the required inductions, risk assessments and approvals from supervisor and DGS. I began a period or phased return to work on a timetable to meet social distancing requirements. I will aim to have 2 or 3 days a week undertaking experiments for validation of fetal work so hope to continue progressing with my project. Preparing for cell culture (organoid) techniques once reliable return to work pattern is established.
08/07/2020	December 2020	Lab work in a shift based pattern	Medium	Lab began working in a shift based pattern. This would mean access to the lab on average 2-3 days a week initially which improved throughout the year until December where I could be in, for at least some period of the day, every day. This allowed tasks like cell culture (organoids) to resume. Developmental samples were still limited or less frequent at the end of the year and had not been collected in earlier months for validation of results.
26/12/2020	March 2021	Varied access to lab	Medium	There was a further national lockdown and although the lab remained open the number of users was reduced to maximise social distancing. This again reduced lab work from full time. Still maintained cell culture techniques. It improved slowly after the spring so that by June 2021 I was having full weeks back in the lab.
30/06/2021	08/07/2021	Self-isolation period (Contact)	Medium	At short notice I received a notification that I had been in proximity at work with someone that had gone on to have a positive CoVID test. This triggered an 8 day isolation through the NHS app. A large organoid stimulation that was planned for scRNA-seq and comparison to fetal and adult epithelial data had to be abandoned.

Table A.1: Post graduate researchers (Medical Sciences Division) impact of the coronavirus (COV-19) student log

A.2: Patient information (IBD study)

Patient group	N	Age (years) (median [range])	Gender (n(%) male)
Healthy 10x scRNA-seq	3	50 [47-74]	1 (33%)
UC 10x scRNA-seq	3	55 [36-80]	2 (66%)
Smartseq2 FACS cells	4 Healthy	68 [60-77]	3 (66%)
Proteomic FACS cells	4 Healthy	56 [45-64]	1 (25%)
Intracellular pH measurement	2 Healthy	n/a [17-83]	2 (100%)
Microfluidic RT-PCR	3 Healthy	53[49-78]	2 (66%)
IHC quantification (Mild Inflammation)	22	47 [23-65]	8 (36%)
IHC quantification (Severe Inflammation)	11	45 [29-81]	6 (55%)

Table A.2: Overview of patient characteristics for samples used in study of adult intestinal epithelium

A3: Adult IEC major cluster markers

Presented here are the top 15 genes (highest AUC value) for each cluster in the IEC adult experiment, full genes lists are available in the uploaded supplementary document[389]. Only the top positive markers are shown.

Colonocytes			Enteroendocrines			ILCs			Mast Cells			BEST4/OTOP2 Cells			T-Cells		
AUC	logFC	gene	AUC	logFC	gene	AUC	logFC	gene	AUC	logFC	gene	AUC	logFC	gene	AUC	logFC	gene
0.892	1.27	*KRT19*	0.964	3.82	*PCSK1N*	1	4.53	*HLA-DRA*	1	5.94	*TPSB2*	0.963	3.07	*CA7*	0.989	3.07	*SRGN*
0.87	1.09	*PHGR1*	0.948	2.65	*CRYBA2*	1	4.32	*CD74*	1	5.16	*TPSAB1*	0.941	2.70	*CA4*	0.981	3.17	*CREM*
0.85	1.16	*SELENBP1*	0.947	1.90	*SCGN*	1	3.11	*HLA-DQB1*	0.995	3.62	*VIM*	0.926	2.39	*BEST4*	0.977	4.31	*CCL5*
0.85	0.81	*S100A6*	0.943	1.53	*PTMS*	0.998	3.45	*HLA-DPB1*	0.993	2.49	*LAPTM5*	0.907	2.16	*MT1G*	0.974	1.41	*TMSB4X*
0.847	1.39	*CA1*	0.939	2.51	*TUBA1A*	0.97	2.77	*CD83*	0.991	2.33	*CD44*	0.892	2.70	*OTOP2*	0.969	1.58	*B2M*
0.84	1.06	*FABP1*	0.931	2.05	*SCG5*	0.969	3.66	*HLA-DRB1*	0.99	2.61	*ANXA1*	0.888	1.84	*SPIB*	0.968	1.42	*EEF1A1*
0.827	0.75	*TCEB2*	0.929	1.91	*CPE*	0.959	1.62	*NAP1L1*	0.985	2.51	*GLUL*	0.867	1.63	*MT2A*	0.967	2.81	*CD52*
0.812	0.94	*CA2*	0.927	1.68	*MS4A8*	0.947	1.94	*LAPTM5*	0.984	1.91	*ALOX5AP*	0.863	2.06	*MT1H*	0.959	1.74	*BTG1*
0.812	0.77	*S100A14*	0.896	1.70	*INSM1*	0.944	2.93	*HLA-DQA1*	0.977	2.55	*SRGN*	0.855	1.57	*HES4*	0.953	2.99	*RGS1*
0.809	0.87	*SLC26A2*	0.894	1.03	*EID1*	0.926	2.80	*CXCR4*	0.976	2.77	*GATA2*	0.823	1.61	*LYPD8*	0.951	2.87	*CD7*
0.806	0.97	*ADIRF*	0.891	1.47	*MARCKSL1*	0.926	2.19	*CD37*	0.976	2.40	*HPGDS*	0.822	0.89	*TMSB10*	0.95	1.60	*PABPC1*
0.801	0.81	*C10orf99*	0.885	1.02	*ACTG1*	0.918	1.11	*RPS25*	0.975	1.91	*SAMSN1*	0.803	1.73	*GUCA2A*	0.948	2.19	*ARHGDIB*
0.798	0.95	*CKB*	0.882	1.34	*HOXB9*	0.915	3.26	*HLA-DPA1*	0.967	2.32	*RGS1*	0.782	1.37	*MT1X*	0.943	2.21	*HCST*
0.798	0.65	*ATP5D*	0.879	2.07	*FEV*	0.903	0.94	*H3F3B*	0.966	2.65	*S100A4*	0.78	1.08	*CKB*	0.941	0.99	*EIF1*
0.795	0.67	*C19orf33*	0.875	1.25	*PAM*	0.902	1.10	*RPL4*	0.964	2.15	*LMNA*	0.767	0.97	*PHLDA2*	0.939	2.05	*CD3E*

Undifferentiated #1			Undifferentiated #2			Crypt top colonocytes		
AUC	logFC	gene	AUC	logFC	gene	AUC	logFC	gene
0.916	1.02	*RPS18*	0.805	0.66	*EEF1A1*	0.954	1.95	*SLC26A3*
0.915	1.07	*RPL35*	0.8	0.73	*MLEC*	0.95	1.78	*CEACAM1*
0.91	1.10	*RPS19*	0.796	0.82	*UGT2B17*	0.945	1.78	*GUCA2A*
0.909	1.04	*RPL13*	0.79	0.80	*MGST1*	0.943	2.27	*AQP8*
0.902	1.12	*RPS2*	0.787	0.86	*ADH1C*	0.936	1.89	*CEACAM7*
0.9	0.96	*RPS15*	0.783	0.67	*RPS6*	0.935	1.62	*PLAC8*
0.899	0.96	*RPL36*	0.781	0.61	*PPIA*	0.933	1.63	*FTH1*
0.886	0.81	*RPL28*	0.78	0.74	*C1QBP*	0.923	1.50	*CEACAM5*
0.879	0.80	*RPS9*	0.777	0.69	*RPL7*	0.918	1.55	*LINC01133*
0.862	0.89	*RPS8*	0.777	0.66	*CD24*	0.916	1.18	*GUCA2B*
0.862	0.80	*RPL18A*	0.776	0.71	*RPL5*	0.915	1.29	*TSPAN1*
0.862	0.76	*RPL37A*	0.772	0.70	*KLF5*	0.915	1.14	*MYL12B*
0.861	0.77	*RPL13A*	0.766	0.60	*RPSA*	0.909	1.58	*EMP1*
0.856	1.30	*NUPR1*	0.762	0.64	*ATP5A1*	0.909	1.36	*IFI27*
0.854	0.84	*RPL7A*	0.761	0.70	*HSPD1*	0.908	1.49	*PRDX6*

Table A.3: Overview of top 15 positive cluster makers with highest AUC for adult IECs in health (n=9 clusters, logFC = average log fold change)

A.4: EEC markers in adult

Presented here are the top 15 genes (highest value for AUC) for each cluster in the sub-analysis of EEC heterogeneity in adult scRNA-seq experiment, full genes lists are available in the uploaded supplementary document[389]. Only the top positive markers are shown.

Enterochromaffin Cells			L-Cells			Precursor Cells		
AUC	logFC	gene	AUC	logFC	gene	AUC	logFC	gene
0.858	1.50	*CHGA*	0.863	4.19	*PYY*	0.864	0.78	*MARCKSL1*
0.845	2.39	*TPH1*	0.81	0.76	*S100A6*	0.859	0.76	*H3F3A*
0.838	1.45	*DDC*	0.793	0.66	*S100A10*	0.856	0.85	*PTMA*
0.82	1.35	*SLC38A11*	0.772	1.72	*VIM*	0.851	0.78	*HMGB1*
0.796	1.18	*SYT13*	0.769	1.27	*GCG*	0.833	0.74	*RAN*
0.786	1.03	*RGS2*	0.753	0.88	*PHGR1*	0.824	0.75	*HNRNPA1*
0.778	0.66	*FEV*	0.731	0.82	*SLPI*	0.822	0.51	*RPL36A*
0.773	2.17	*CHGB*	0.731	0.60	*ITM2C*	0.821	0.46	*EEF1A1*
0.763	1.19	*GULP1*	0.728	0.74	*TIMP1*	0.817	0.57	*RPL5*
0.762	0.85	*CISD1*	0.717	0.80	*FAM3D*	0.806	0.89	*STMN1*
0.744	1.18	*RAMP1*	0.713	0.52	*CHCHD10*	0.802	0.63	*NIPSNAP1*
0.743	0.85	*PRAC1*	0.707	0.87	*UCN3*	0.802	0.53	*BTF3*
0.738	0.66	*AGR2*				0.801	1.66	*SOX4*
0.732	1.23	*SYT1*				0.801	0.77	*TUBB*
0.727	1.07	*SLC18A1*				0.8	0.69	*BNIP3L*

Table A.4: Overview of top 15 positive cluster makers with highest AUC for analysis of adult EECs in health (n=3 clusters, logFC = average log fold change, L-Cells positive markers n=12 only)

A.5 Undifferentiated markers adult

Presented here are the top 15 genes (using negative binomial analysis) for each cluster in the sub-analysis of undifferentiated IECs in adult scRNA-seq experiment, full genes

lists are available in the uploaded supplementary document[389]. Only the top positive markers are shown.

Stem Cells				TA				Absorptive Progenitors			
p_val	logFC	FDR	gene	p_val	logFC	FDR	gene	p_val	logFC	FDR	gene
6.69E-110	1.60	3.93E-105	*ASCL2*	6.24E-29	0.27	3.66E-24	*TMSB10*	3.38E-83	1.42	1.98E-78	*FABP1*
4.13E-100	1.03	2.42E-95	*SMOC2*	3.66E-28	0.28	2.15E-23	*FAU*	1.41E-66	1.07	8.26E-62	*FXYD3*
1.15E-51	0.33	6.73E-47	*LGR5*	8.64E-28	0.41	5.07E-23	*HMGB1*	5.93E-62	0.88	3.48E-57	*LGALS4*
1.16E-41	0.52	6.80E-37	*PHLDA1*	4.14E-27	0.46	2.43E-22	*MIF*	1.01E-61	0.78	5.91E-57	*LGALS3*
6.32E-35	0.28	3.71E-30	*PTPRO*	4.75E-27	0.46	2.79E-22	*NUPR1*	5.00E-61	0.82	2.94E-56	*PHGR1*
7.49E-33	0.93	4.39E-28	*TM4SF1*	4.79E-26	0.49	2.81E-21	*RANBP1*	1.33E-59	0.66	7.83E-55	*TMSB4X*
9.83E-33	0.87	5.77E-28	*EEF1A1*	1.06E-23	0.36	6.25E-19	*SNRPD2*	5.21E-57	1.19	3.06E-52	*CA2*
2.14E-31	0.80	1.26E-26	*SLC12A2*	4.04E-23	0.48	2.37E-18	*RARRES2*	9.10E-55	0.90	5.34E-50	*KRT19*
2.76E-31	0.28	1.62E-26	*TUBA1A*	7.94E-22	0.32	4.66E-17	*EEF1B2*	6.18E-53	0.63	3.63E-48	*KRT8*
1.61E-28	1.22	9.47E-24	*EDN1*	1.55E-21	0.43	9.11E-17	*SNHG6*	5.33E-50	0.98	3.13E-45	*CD24*
2.34E-28	0.79	1.38E-23	*GAS5*	2.17E-20	0.40	1.27E-15	*STRA13*	6.50E-48	1.21	3.81E-43	*SLC26A2*
4.80E-28	0.70	2.81E-23	*LRRC75A-AS1*	3.69E-20	0.44	2.17E-15	*SIVA1*	9.36E-48	0.79	5.50E-43	*SELENBP1*
5.14E-28	0.61	3.02E-23	*ALDH1B1*	1.04E-18	0.28	6.08E-14	*PFDN5*	1.46E-45	0.65	8.59E-41	*S100A14*
3.73E-27	0.88	2.19E-22	*CXCL8*	2.46E-18	0.35	1.44E-13	*MZT2A*	4.17E-45	1.29	2.45E-40	*CA1*
6.24E-27	0.26	3.66E-22	*PROX1*	3.68E-18	0.36	2.16E-13	*GPX4*	7.43E-45	0.74	4.36E-40	*HSD11B2*

Cell Cycle/TAs				Secretory Progenitors			
p_val	logFC	FDR	gene	p_val	logFC	FDR	gene
5.82E-47	0.60	3.42E-42	*CCNA2*	2.37E-94	1.07	1.39E-89	*ELF3*
1.74E-42	1.20	1.02E-37	*STMN1*	3.12E-80	1.22	1.83E-75	*METTL12*
5.52E-41	1.26	3.24E-36	*TUBB*	2.38E-77	1.20	1.40E-72	*HEXIM1*
4.92E-39	1.09	2.89E-34	*H2AFZ*	1.80E-62	1.37	1.06E-57	*RN7SK*
4.81E-37	0.72	2.82E-32	*CDK1*	1.51E-58	0.64	8.87E-54	*MALAT1*
2.54E-35	1.02	1.49E-30	*PTGES3*	9.82E-56	0.91	5.76E-51	*SAT1*
4.25E-35	0.91	2.50E-30	*H3F3A*	1.10E-55	0.99	6.48E-51	*HIST1H4C*
4.50E-35	0.70	2.64E-30	*LMNB1*	2.49E-54	0.82	1.46E-49	*NEAT1*
3.13E-33	0.98	1.84E-28	*KIAA0101*	4.15E-50	0.69	2.43E-45	*JUN*
3.98E-32	0.84	2.34E-27	*TUBA1B*	8.09E-47	0.84	4.75E-42	*BTG1*
1.35E-31	0.86	7.91E-27	*PPIA*	7.92E-46	0.80	4.65E-41	*FADS3*
2.12E-31	0.69	1.24E-26	*BIRC5*	4.21E-44	1.17	2.47E-39	*MAFB*
1.71E-30	0.95	1.00E-25	*RAN*	1.64E-43	0.76	9.61E-39	*ZFP36L2*
1.77E-30	0.67	1.04E-25	*MAD2L1*	3.74E-41	0.82	2.19E-36	*YME1L1*
3.29E-30	0.96	1.93E-25	*NPM1*	8.80E-41	0.83	5.17E-36	*EIF4A1*

Table A.5 Overview of top 15 positive cluster makers with lowest FDR for analysis of undifferentiated cells in health (n=5 clusters, logFC = average log fold change, p_val = p-value, FDR = false discovery rate)

A.6: Branch specific genes in adult IECs

Presented here are selected genes from the analysis of branch specific expression of IEC genes. These represent genes that change between secretory and absorptive lineage branches which were significantly upregulated (<1% FDR, >0 fold change expressed in log_2) and intersected with significant pseudo-time varying and branch specific expression. It also identified genes that were expressed before lineage divergence and thus suggested as putative lineage regulators. The full analysis list is available in supplementary upload[389].

p-value	Gene	Number cells expressing	Putative Regulator/Pre-branch marker
0	*ADIRF*	2735	TRUE
0	*LGALS3*	3667	TRUE
0	*REG4*	523	FALSE
0	*ITLN1*	824	FALSE
0	*EPCAM*	3375	FALSE
0	*FABP1*	3693	TRUE
0	*SPINK1*	1891	FALSE
0	*CHGA*	62	FALSE
0	*TFF3*	3229	FALSE
0	*SEPP1*	1636	FALSE
1.22E-153	*ATOH1*	228	FALSE
5.16E-107	*KLF4*	2694	FALSE
1.43E-105	*SOX4*	1759	TRUE
1.02E-35	*HEXIM1*	1847	TRUE
1.37E-34	*MAFB*	788	TRUE
2.35E-27	*NEUROD1*	17	FALSE
1.03E-10	*ASCL2*	173	FALSE
2.43E-09	*STAT3*	1385	FALSE
0.05776006	*OLFM4*	195	FALSE

Table A.6: Selected genes shown in **Figure 2.4** along with other known branch regulators of mature genes, all identified from branch specific gene analysis.

A.7: DEGs in health and inflammation

Presented are a selection of 30 DEGs in inflammation separated by cell type. Full gene list is available[389].

Gene	Colonocytes	Goblets	Crypt Top Colonocytes	Absorptive Progenitors	Stem Cells	Secretory Progenitors	TAs	BEST4/OTOP2 Cells	EECs
HES4	0.33785	0.99982	0.00001	0.46006	0.99989	0.99714	0.99884	0.99787	0.99933
ISG15	0.04598	0.00903	0.00000	0.00066	0.66611	0.01115	0.10917	0.00009	0.99933
AGRN	0.00801	0.99982	0.00391	0.97121	1.00000	1.00000	1.00000	1.00000	0.99933
DVL1	0.00176	0.99982	0.02136	0.99992	1.00000	0.99714	0.99884	1.00000	0.99933
VWA1	0.00163	0.14644	0.00000	0.07671	1.00000	0.23679	0.99884	1.00000	0.99933
ERRFI1	0.00157	0.99982	0.00018	0.99992	0.99989	0.99714	0.99884	0.86459	0.99933
PADI2	0.00000	0.16364	0.05539	0.00037	1.00000	1.00000	1.00000	0.61748	0.99933
IFFO2	0.00631	1.00000	0.00050	1.00000	1.00000	1.00000	1.00000	1.00000	0.99933
PLA2G2A	0.00000	0.00000	0.00000	0.05034	0.00001	0.00367	0.53062	0.00001	0.99933
CAMK2N1	0.87204	0.00275	0.80184	0.66563	0.21043	0.20625	0.99884	0.99754	0.99933
CELA3B	0.00504	1.00000	0.00083	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
TCEA3	0.00005	0.64480	0.04220	0.10823	0.99989	0.54198	0.79962	1.00000	0.99933
ID3	0.00087	0.00000	0.00388	0.00197	0.00000	0.07104	0.00067	0.30007	0.99933
GALE	0.36217	0.81716	0.00051	0.99992	0.99989	0.99714	0.99884	0.97276	0.99933
HMGCL	0.14836	0.99982	0.00830	0.99992	0.99989	1.00000	0.99884	0.96949	0.99933
IL22RA1	0.00012	0.99982	0.56613	0.99992	1.00000	1.00000	1.00000	1.00000	0.99933
IFI6	0.54695	0.00000	0.01757	0.37553	0.00031	1.00000	0.01770	1.00000	0.99933
ZC3H12A	0.02078	0.79644	0.00037	0.66971	0.99989	0.99714	0.99884	0.94152	0.99962
YBX1	0.03677	0.99982	0.00093	0.99992	0.99989	0.99714	0.99884	0.15113	0.99933
ERI3	0.22180	0.00431	0.77656	0.99992	0.99989	0.99714	0.99884	0.99406	0.99933
UQCRH	0.28716	0.75429	0.00354	0.59846	0.99989	0.30574	0.99884	0.95598	0.99933
PDZK1IP1	0.00000	0.00004	0.00000	0.04032	0.22389	0.41296	0.85744	0.02240	0.99933
FAM151A	0.00223	1.00000	0.06735	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
DNAJB4	0.06465	0.00611	0.31420	0.37941	0.00046	0.99714	0.99884	0.63443	0.99933
CLCA1	0.80529	0.53824	0.89477	0.99992	1.00000	0.00010	0.99884	1.00000	0.99933
CLCA4	0.13551	0.99982	0.00005	0.99992	1.00000	1.00000	0.99884	0.05564	0.99933
WFDC2	0.15809	0.15150	0.00003	0.99992	0.27656	0.05709	0.99884	0.87628	0.99933
GBP1	0.00000	1.00000	0.00001	0.00064	0.28939	1.00000	1.00000	0.14599	0.99933
SERPINH1	0.99458	0.00002	0.99407	0.77065	0.00004	0.98678	0.99884	1.00000	0.00985
F3	0.00000	0.00083	0.00000	0.99992	0.99989	1.00000	1.00000	1.00000	0.99933

Table A.7 Selected genes differentially expressed in inflammation compared to HC by cell type. 30 genes are selected from n=1784 DEGs as examples of cell type specific effect. (FDR shortened to 5dp, Green colour highlighting significant differences)

A.8 Goblet cell sub-type markers

Presented are the top 15 genes for each cluster in the sub-analysis of goblet cells in adult scRNA-seq experiment, full genes lists are available in the uploaded supplementary document[389].

Cluster 1			Cluster 2			Cluster 3			Cluster 4			Cluster 5		
AUC	logFC	Gene	AUC	logFC	Gene	AUC	logFC	Gene	AUC	logFC	Gene	AUC	logFC	Gene
0.749	0.65	*HERPUD1*	0.214	-0.99	*PHGR1*	0.816	1.15	*SOX4*	0.908	1.29	*CEACAM5*	0.963	2.28	*TFF1*
0.741	0.52	*SELK*	0.225	-0.64	*S100A6*	0.808	0.82	*SAT1*	0.901	1.38	*LMO7*	0.96	1.85	*GSN*
0.728	0.57	*KLF6*	0.238	-0.91	*KRT20*	0.801	0.75	*ELF3*	0.889	1.26	*PLAUR*	0.957	1.66	*PHGR1*
0.723	0.44	*LGALS4*	0.26	-1.93	*TFF1*	0.799	1.31	*RN7SK*	0.887	1.08	*MCL1*	0.941	1.35	*TSPAN1*
0.71	0.70	*AREG*	0.263	-0.77	*TSPAN1*	0.786	0.74	*MALAT1*	0.883	0.93	*KLF6*	0.938	1.03	*S100A6*
0.71	0.58	*FOSB*	0.734	0.63	*WFDC2*	0.754	1.23	*MAFB*	0.882	1.18	*DST*	0.92	1.74	*ZG16*
0.71	0.38	*ACTG1*	0.269	-0.75	*MUC13*	0.251	-0.59	*ACTB*	0.881	1.01	*MUC13*	0.91	1.16	*FXYD3*
0.709	0.51	*KLF4*	0.27	-1.04	*GSN*	0.256	-0.69	*NDUFA4*	0.878	1.08	*CYP3A5*	0.108	-1.01	*GNB2L1*
0.702	0.40	*EZR*	0.272	-1.09	*ZG16*	0.742	0.62	*NDUFB7*	0.878	1.00	*SSFA2*	0.121	-0.99	*PTMA*
			0.275	-0.58	*C19orf33*	0.741	0.59	*CYBA*	0.874	1.23	*F3*	0.876	1.09	*IFI27*
			0.282	-0.77	*AREG*	0.259	-0.66	*BTF3*	0.869	1.21	*ARL14*	0.124	-1.16	*EEF1A1*
			0.287	-0.64	*KLF6*	0.259	-0.68	*LRRC75A*	0.869	1.19	*LAMC2*	0.861	1.42	*LYPD8*
			0.291	-0.51	*S100A14*	0.734	0.57	*RRBP1*	0.866	1.13	*CAPN8*	0.152	-1.12	*LRRC75A*
			0.295	-0.66	*NEAT1*	0.733	0.40	*STARD10*	0.864	1.18	*B4GALT1*	0.846	1.65	*CKB*
			0.299	-0.79	*ARL14*	0.27	-0.58	*EEF1A1*	0.853	1.08	*S100P*	0.844	0.93	*C19orf33*

Table A.8: Overview of top 15 positive cluster makers with highest/lowest AUC for analysis of adult goblet cell sub-types in health (n=5 clusters, logFC = average log fold change, Cluster 1 n=12 markers only, positive markers highlighted green and negative in red)

A.9 BEST4/OTOP2 markers Gao et al

Presented are the top 20 markers from supervised analysis of scRNA-seq study by Gao et al[194], with full analysis available in supplementary materials[389].

AUC	logFC	gene
0.987	3.53349783	*CA7*
0.952	2.37302435	*SPIB*
0.95	2.71707212	*BEST4*
0.904	1.03720122	*TMSB10*
0.892	1.12268963	*NBPF10*
0.856	0.97611914	*RPL37A*
0.834	1.64200839	*NOTCH2NL*
0.829	1.9206064	*DMBT1*
0.828	1.04424738	*RPS8*
0.822	2.04738713	*OTOP2*

Table A.9: Overview of top 10 positive cluster makers with highest AUC for supervised analysis Gao et al[194] (logFC = average log fold change, only positive markers shown, highlighted genes represent those marking core signature from adult scRNA-seq study)

A.10 BEST4/OTOP2 proteomics results

Presented are the top 20 positive proteins detected by proteomics when comparing BEST4+EPCAM+ to BEST4-EPCAM+ cells isolated by FACS. Analysis performed in only proteins with 2+ unique peptides detected, full data for 2+ peptides and all proteins is available in supplementary materials[389]

Gene	logFC	P-value	Adjusted P-value	Unique peptides
TSNAX.DISC1	4.964039	6.33E-06	0.00226	2
HTATSF1	4.283275	2.43E-05	0.005998	2
MGST3	3.514067	5.80E-05	0.009123	4
NEU1	3.668953	5.14E-05	0.009123	3
ABCB1	3.516708	5.26E-05	0.009123	2
TAP1	3.086871	6.76E-05	0.009353	4
CUX1	3.530826	6.56E-05	0.009353	4
KRT20	3.504441	7.38E-05	0.009809	85
PTPN23	3.401646	8.88E-05	0.00993	5
PCMT1	3.135493	8.49E-05	0.00993	5
ARF6	2.984004	8.70E-05	0.00993	3
VMP1	3.837668	8.90E-05	0.00993	2
CLCA4	2.804737	9.93E-05	0.010088	12
PYCARD	4.545151	0.000103146	0.010088	4
DGKQ	2.659251	0.00010502	0.010088	2
HPGD	4.03236	9.90E-05	0.010088	2
AHNAK	3.2756	0.000116383	0.010591	175
GALT	3.760976	0.000125444	0.011123	2
SULT1B1	2.553437	0.000129166	0.011166	5
ACAT2	4.007407	0.000134329	0.011329	9

Table A.10: Top 20 positive peptides detected using proteomics with ≥2 unique peptides detected (Comparison using BEST4+EPCAM+ n=3 and BEST4-EPCAM+ n=2, 20 results with lowest adjusted p-value presented)

A.11 BEST4/OTOP2 SSII results

Presented are 15 genes identified from SSII protocol of BEST4+EPCAM+ and BEST4-EPCAM+ cells. Full results are available in supplementary materials[389]

AUC	logFC	Gene
0.153	-0.83872	*CH507-528H12.1*
0.154	-0.83872	*CH507-513H4.1*
0.189	-3.38898	*BEST4*
0.201	-1.10163	*MTRNR2L8*
0.213	-1.32171	*LL22NC03-86G7.1*
0.213	-3.38792	*CA4*
0.218	-3.17001	*GUCA2A*
0.223	-3.7533	*CA7*
0.232	-0.62866	*RP3-323A16.1*
0.236	-0.60977	*CTA-228A9.4*
0.242	-2.88996	*OTOP2*
0.261	-1.84027	*SPIB*
0.262	-4.19394	*GUCA2B*
0.265	-1.98274	*LYPD8*
0.308	-1.00419	*NOTCH2*

Table A.11: BEST4/OTOP2 markers identified using SSII (comparison made is BEST4- vs BEST4+ cells so top 10 genes with lowest AUC shown alongside 5 selected genes in the core BEST4/OTOP2 signature identified by initial experiment.

A.12 GOE of proteomic data for BEST4/OTOP2 cells

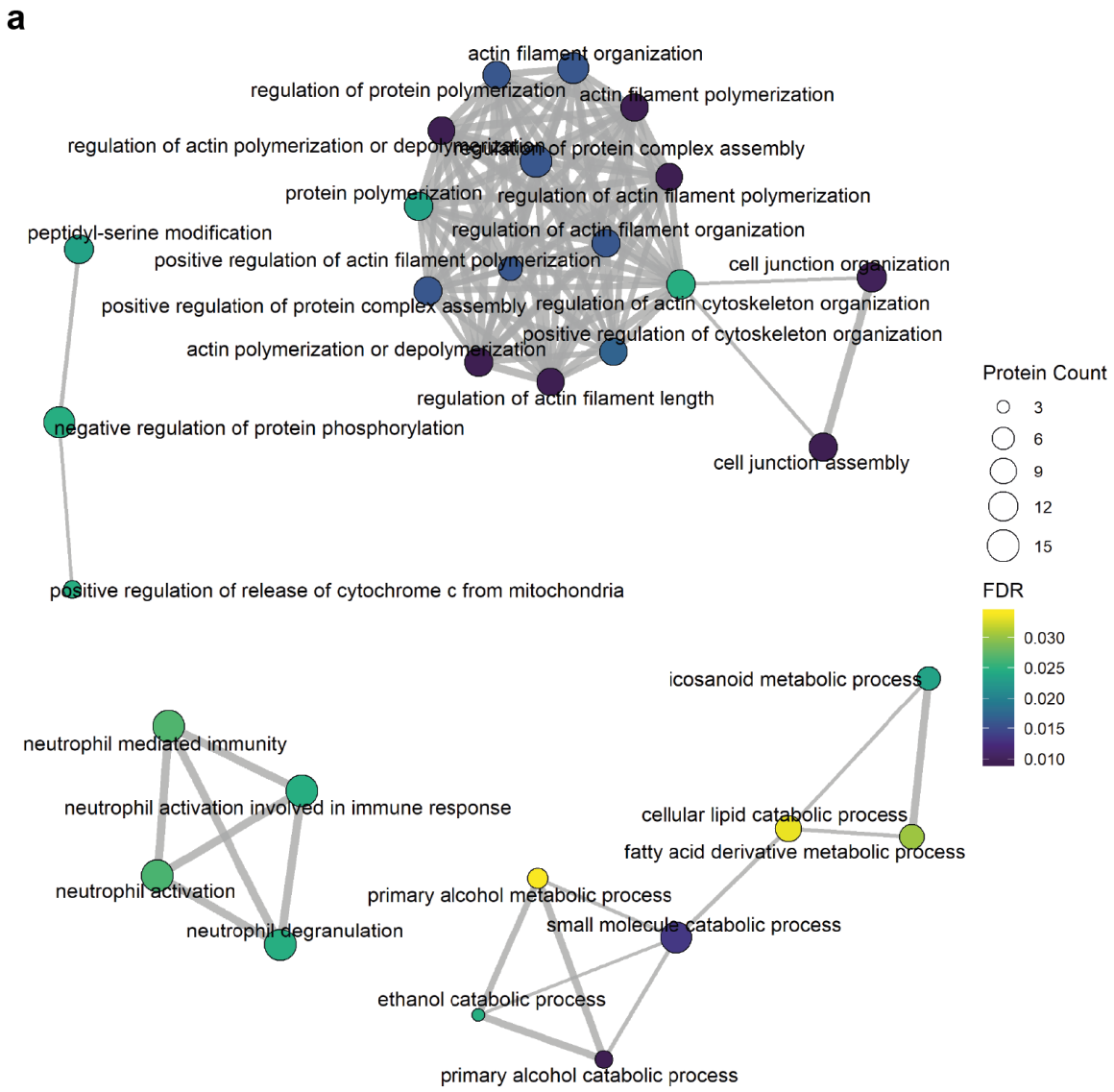

Figure A.12: Over-represented GO terms in the significantly upregulated protein set in BEST4/OTOP2 cells as identified by quantitative proteomics (n=2 BEST4- vs n=3 BEST4+)

A.13 Multiplexed scRNA-seq methodology

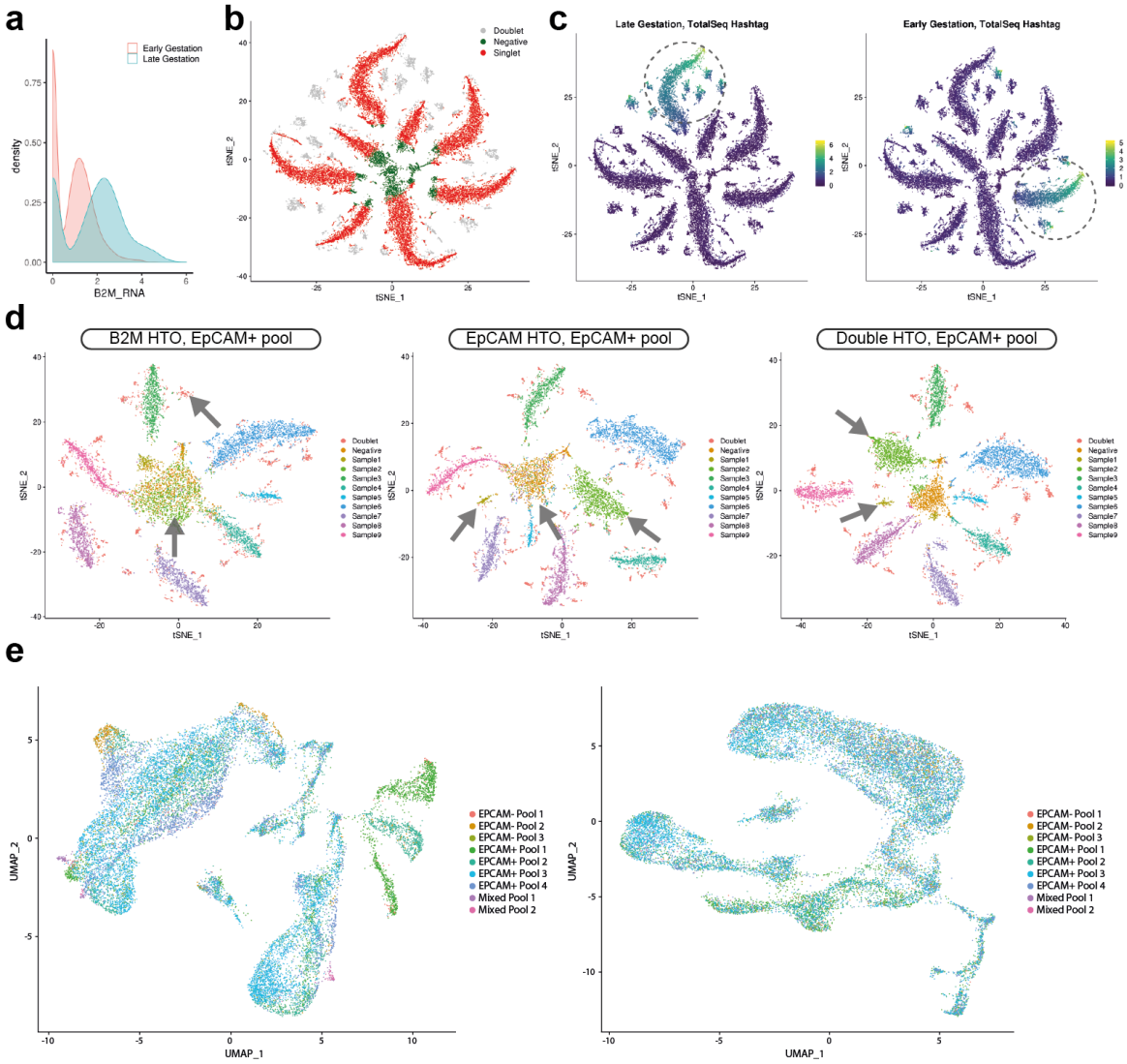

Figure A.13: Overview of multiplexing methodology

a, Example of mRNA distribution of *B2M* expression in EPCAM+ cells from early (8PCW) and late (19PCW) samples from samples in same pool demonstrating reduced *B2M* expression in early gestation samples. **b,** t-sne of cells from a representative EPCAM- pool showing de-hashing of cells into singlets, doublets or unstained cells with commercial (TotalSeq) antibodies. **c,** t-sne of demultiplexing in EPCAM- pool in both early and late gestation **d,** t-sne of EPCAM+ cells from a representative pool showing comparison of commercial and custom (EPCAM+) antibody alongside double HTO demultiplexing (arrows indicate regions unable to be resolved with B2M antibody, but able to be with in house antibody). **e,** example of pool effect batch correction described in **Methods 6.11** (UMAP shows epithelial cells and pools prior to and after correction)

A.14 Supplementary materials and analyses of fetal atlas

Due to the large amount of multi-dimensional data and analysis that were performed as part of the developmental study, information supporting the results have been outlined here or signposted to online supplementary materials[389] which will contain large analysis and outputs. Core components of the results will also be selected and presented when appropriate. An overview of these with descriptions is as follows.

Tissue samples used

An overview of gestational ages and locations of tissue sampling used in the scRNA-seq resource is given in **Figure 4.2**. Full demographic details and gender genotype information alongside run ID and antibody labelling is presented in Mendeley data resource[389] alongside those samples used for ST.

Compartment markers

For initial clustering of all cells used for the annotation of 9 cluster markers, markers of each respective compartment are given in full in supplementary resource[389].

Cell type markers

Annotation of 101 cell types was performed as described in **Methods 6.10**. A full overview of marker genes, divided by each compartment is listed in supplementary materials[389], a selected overview of specific markers is presented here.

Compartment	Compartment Genes	Cell Type	Key Genes (Cell Type)	Cluster Name (Graph abstraction number)	Key Genes (Cluster)	Locational Differences	Time-point Differences[b]
Epithelial	*EPCAM, FABP1*	Stem	*LGR5, ASCL2, MYC*	Proximal Stem Cells (1)	*CASP1, GRIA4, PLA2G2A*	Increased in small intestine (****)	No significant differences
				Distal Stem Cells (2)	*LEFTY1, PTGDR, SMOC2*	Increased in Colon (****)	Increases over time in colon and overall
		TA	*MKI67, TOP2A, UBE2C*	Proximal TA (3)	*CENPM, TK1, UBE2T*	Increased in small intestine (****)	Decreases over time overall
				Distal TA (4)	*HMGB3, MND1, TACC3*	Increased in colon (****)	No significant differences
		Secretory[a]	*FOXA2, INSM1, HES6*	Secretory Progenitors (5-15)	*ISL1, NEUROG3, GHRL*	Increased in small intestine (**)	No significant differences
				EECs (5-15)	*CHGA, TPH1, NEUROD1*	No locational preference	Increases over time in colon, TI and overall
				Goblet cells (5-15)	*MUC2, SPINK4, WFDC2*	No locational preference	Increases over time in colon and overall
		Absorptive	*FABP2, APOE* (proximal), *FAM3D* (distal)	Proximal Progenitors (16)	*FGG, SOX11, FGB*	Increased in small intestine (**)	Decreases over time in TI and overall
				Proximal Early Enterocytes (17)	*TF, AFP, VTN*	Increased in small intestine (**)	No significant differences
				Proximal Enterocytes (18)	*CCL25, HEBP1, OSR2*	Increased in small intestine (***)	No significant differences
				Proximal Mature Enterocytes (19)	*APOC3, ACE, FABP2*	Increased in small intestine (**)	Increases over time in TI and overall
				Distal Absorptive (20)	*CA12, HMGCS2, SLC13A2*	Increased in colon (****)	No significant differences
				Distal Enterocytes (21)	*SLC26A2, FABP1, CKB*	Increased in colon (****)	No significant differences
				Distal Mature Enterocytes (22)	*AQP8, FABP6, IL32*	Increased in colon (****)	Increases over time in colon and overall
			N/A	BEST4/OTOP2 cells (23)	*BEST4, OTOP2, CA7*	No locational preference	Increases over time in TI
Secretory[a]	*N/A*	Progenitor	N/A	*NEUROG3*+ Progenitor (5)	*NEUROG3, CLPS, MDK*	N/A	
		Goblet	*MUC2, TFF3*	Goblet 1(6)	*UBE2C, PBK, MKI67*		
				Goblet 2 (7)	*TRABD2A, PTGDR, MYC*		
				Goblet 3 (8)	*SPDEF, WFDC2, ATOH1*		
		EEC	Cluster specific	A-Cells/M-Cells (9)	*CLEC4G, MLN, GHRL*		
				D-Cells (10)	*SST, HHEX, PCP4*		
				Enterochromaffin Cells (11)	*TPH1, RXFP4, CHGA*		
				I-Cells (12)	*CCK, GC, TDO2*		
				L-Cells (13)	*PYY, GCG, UCN3*		
				N-Cells (14)	*NTS, HOPX, MEP1A*		

		Paneth	N/A	Paneth (15)	*PRSS2, DEFA5, DEFA6*		
Fibroblast	*THY1, COL1A2, VIM*	Cycling	N/A	Fibroblast G2M phase (24)	*CENPF, TOP2A, MKI67*	No locational preference	Decreases over time in colon, TI and overall
			N/A	Fibroblast S-Phase (25)	*PCLAF, TYMS, GINS2*	No locational preference	Decreases over time in colon, TI and overall
		Progenitor	N/A	Fibroblast progenitor (26)	*HMGA2, MBP, THY1(low)*	No locational preference	Decreases over time in colon, TI and overall
		S1	*ADAMDEC1, FABP5, CTSC*	S1 (27)	*PDLIM3, COL28A1, TFPI2*	Increased in colon (**)	Increases over time in colon and overall
				S1 *COL6A5*+ (28)	*COL6A5, LXN, WNT4*	Increased in small intestine (***)	Increases over time in TI
				S1 *IFIT3*+ (29)	*IFIT3, CLEC2B, GPX3*	No locational preference	Increases over time in colon, TI and overall
		S2	*FRZB, SOX6, F3*	Proximal S2 1 (30)	*NPY, CXCL10, BMP2*	Increased in small intestine (****)	Increases over time in TI
				Proximal S2 2 (31)	*MMP11, INSC, BMP3*	Increased in small intestine (****)	No significant differences
				Distal S2 (32)	*POSTN, ALKAL2, WNT5B*	Increased in colon (****)	Increases over time in colon and overall
		S3	*C7, ASPN, OGN*	S3 progenitor (33)	*OSR2, NR2F1, PTN(low)*	No locational preference	Increases then decreases over time in colon and overall
				S3 Transitional (34)	*GREM1, GDF10, FBLN1*	No locational preference	Increases over time in colon, TI and overall
				S3 (35)	*RAMP1, SCN7A, SCUBE2*	No locational preference	Increases over time in colon, TI and overall
				S3 EBF+ (36)	*PRRX1, EBF3, EBF2*	Increased in colon (*)	Increases over time in colon and overall
				S3 HAND1+ (37)	*COL9A3, DLK1, HAND1*	Increased in colon(*)	Increases over time in colon and overall
		S4	*CCL19, C3, CD74*	S4 CXL13+ (38)	*CXCL13, TNFSF11, CD24*	Increased in small intestine (*)	Increases over time in TI and overall
				S4 CCL21+ (39)	*CCL21, OLFM3, PTGDS*	No locational preference	Increases over time in colon, TI and overall
Myofibroblast /Mesothelium	*VIM, FOXF1, TAGLN, RSPO2, RGS5(-)*	Cycling	N/A	Myofibroblast G2M phase (40)	*MKI67, CENPF, TOP2A*	No locational preference	No significant differences
			N/A	Myofibroblast S-phase (41)	*PCLAF, TYMS, GINS2*	No locational preference	Increases over time overall
		Myofibroblast	*THY1, NKX2-3, ACTG2*	Myofibroblast progenitors(42)	*ADAMDEC1, MFAP4, ADAM28*	No locational preference	Decreases over time in colon and overall
				Myofibroblasts (43)	*ACTA2, MYH11, GREM2*	No locational preference	Increases over time in colon, TI and overall
		Mesothelium	*MSLN, WT1, RSPO1*	Mesothelium *IL18*+(44)	*IL18, GATA6, WNT10A*	No locational preference	No significant differences
				Mesothelium *SOX6*+ (45)	*SOX6, HPGD, STX2*	Increased in colon (**)	No significant differences
Endothelium	*PECAM1, CDH5, CLDN5*	Cycling	N/A	Endothelium G2M-phase (46)	*MKI67, CENPF, TOP2A*	No locational preference	Decreases over time in colon, TI and overall
				Endothelium S-phase (47)	*PCLAF, TYMS, GINS2*	No locational preference	Decreases over time in colon and overall
		Arterial	*GJA4, IGFBP3, UNC5B*	Arterial (L) (48)	*HEY1, GJA5, ITGB4*	No locational preference	Increases over time in colon and overall

				Arterial (M) (49)	*UNC5B, DLL4, PRND*	No locational preference	Increases over time in colon
				Arterial (CP) (50)	*VWA1, AFG1L, EDNRB*	No locational preference	Increases over time in colon, TI and overall
				Arterial (CP) 2 (51)	*CHST1, LGALS1, IGFBP3*	No locational preference	Decreases over time in colon and overall
		Venous	*MADCAM1, APLNR, PRCP*	Venous (L) (52)	*ACKR1, VWF, ADGRG6*	No locational preference	Increases over time in colon and overall
				Venous (M) 1 (53)	*LRG1, PLVAP, STMN1(-)*	No locational preference	Increases over time in colon, TI and overall
				Venous (M) 2 (54)	*CALCRL, APLNR, PRCP*	No locational preference	Decreases over time overall
				Venous (CP) 1 (55)	*CCL2, CD69, RGCC*	No locational preference	Decreases over time in colon and overall
				Venous (CP) 2 (56)	*LDHB, JUNB(-), ZFP36(-)*	No locational preference	Decreases over time in colon, TI and overall
		Lymphatic	N/A	Lymphatic (57)	*LYVE1, PROX1, RELN*	No locational preference	Increases over time in colon and overall
Pericytes	*KCNJ8, ABCC9, RGS5, CSP4G*	Cycling	N/A	Pericyte G2M-phase (58)	*MKI67, CENPF, TOP2A*	No locational preference	Decreases over time in TI and overall
			N/A	Pericyte S-phase (59)	*PCLAF, TYMS, GINS2*	No locational preference	Decreases over time in colon and overall
		Pericytes	See compartment genes	Pericyte progenitors (60)	*THBS1, MBP,CEBPD*	No locational preference	Decreases over time in colon and overall
				Undifferentiated pericytes (61)	*CYGB, NFKBIA, TESC*	No locational preference	Increases over time in colon and overall
				WNT6+ pericytes (62)	*WNT6, GPAT2, THBS1(-)*	No locational preference	Increases over time in colon and overall
				Angiogenic pericyte (63)	*PROCR, EBF2, THBS4*	Increased in colon (**)	Increases over time in colon and overall
				BMPER+ pericytes (64)	*BMPER, CYGB, SPON2*	No locational preference	Increases over time in colon, TI and overall
				Contractile pericyte (65)	*ADIRF, MYH11, ACTA2*	No locational preference	Increases over time in colon and overall
Neural	*PHOX2B, HAND2, TUBB2B*	Cycling	N/A	Neural G2M & S-Phase (66)	*MKI67, CENPF, TOP2A*	No locational preference	Decreases in time in colon and overall
		Glial	*S100B, SOX10, PLP1*	Lymphoid associated glial (67)	*FGL2, GFRA3, RXRG*	Increased in colon (**)	Increases over time in colon and overall
				Intraganglionic glial (68)	*ENTPD2, SPRY1, SOX6*	No locational preference	Increases over time in colon, TI and overall
				Submucosal glia (69)	*CEBPD, SOCS3, ZFP36*	No locational preference	Increases in time in colon and overall
				Differentiating submucosal glial (70)	*BCAN, APOE, LAMP5*	No locational preference	Increases over time in colon and overall
				Glial progenitors (71)	*PHOX2B, HMGA2, CTGF*	No locational preference	Decreases over time in colon, TI and overall
		Neuronal	*ELAVL4, CHRNA3, GAP43*	ENS progenitors(72)	*DLL3, DLL1, ELAVL4(low)*	No locational preference	Decreases over time in colon and overall
				Neuroendocrine 1 (73)	*NEUROD6, SCGN, SSTR2, NPY2R*	No locational preference	Increases in time in colon and overall

				Neuroendocrine 2 (74)	*DGKG, SST, KCTD16*	No locational preference	Increases in time in colon, TI and overall
				Excitory motor neuron (75)	*CASZ1, RAMP1, GFRA2*	No locational preference	No significant differences
				Inhibitory motor neuron precursor (76)	*SCGN, CRABP1, NOS1, VIP, ADCYAP1*	No locational preference	Decreases over time in TI and overall
				Inhibitory motor neuron (77)	*NOS1, VIP, HTR2B, ETV1*	No locational preference	Increases then decreases in time in colon and overall
				Interneuron (78)	*PENK, ACHE, TAC1, ONECUT2*	No locational preference	Decreases over time
Immune	*PTPRC*	Macrophage	*MERTK, CTSC, CTSD, CD14 (hi)*	Macrophage (79)	*FGL2, HES1, SLC40A1*	Increased in colon (*)	Increases over time in colon and overall
				Macrophage SPP1+ (80)	*SPP1, RNASE1, FOLR2*	Increased in colon (*)	No significant differences
		Monocyte	*CD14 (lo)*	Monocyte (81)	*FCN1, S100A8, CLEC12A*	No locational preference	No significant differences
		Dendritic cell	*ETV6, FLT3*	pDCs (82)	*IL3RA, LILRA4, PTCRA*	No locational preference	No significant differences
				DCs (83)	*CD207, BATF3, CD1D*	No locational preference	No significant differences
		Adaptive lymphoid cells	*CD19, EBF1, PAX5*	Pre B Cell (84)	*IGLL1, VPREB1, RAG1*	No locational preference	No significant differences
				B cell (85)	*BANK1, CD79B, IGHD*	No locational preference	Increases over time in colon, TI and overall
		Innate lymphoid cells	*ETS1, TOX*	Naïve T Cell (86)	*CD3D, CD3G, CD27*	No locational preference	Increases over time in colon, TI and overall
				NKs and γδ (87)	*CD8A, IL2RB, TRDC*	Increased in small intestine (*)	No significant differences
				Type 3 ILCs (88)	*IL7R, ID2, RORC, KRT86*	No locational preference	Increases over time in colon and overall
		Eosinophils	N/A	Mast (89)	*TPSAB1, CPA3, MS4A2*	No locational preference	Increases over time in TI and overall. Decreases in colon.
		Cycling	N/A	Immune Cell cycle (90)	*MKI67, CENPF, TOP2A*	No locational preference	Decreases over time in colon, TI and overall
Muscle	*MYH11, ACTG2, TAGLN*	Cycling	N/A	Muscularis G2M phase (91)	*MKI67, CENPF, TOP2A*	No locational preference	Decreases then increases over time in colon
				Muscularis S-phase (92)	*PCLAF, TYMS, GINS2*	No locational preference	No significant differences
		Interstitial Cells	*KIT*	Interstitial Cells of Cajal / ICCs (93)	*ANO1, ETV1, SPON2*	No locational preference	No significant differences
				PDGFRA+ cells (94)	*PDGFRA, ITGA8, BMP4*	No locational preference	Decreases over time in colon and overall
		Outer muscularis	*FOXF2(-)*	OM (95)	*BCHE, CAPN3, IGFBP5*	No locational preference	Increases over time in colon and overall
		Inner muscularis (IM)	*FOXF2, ACTC1*	Proximal IM (96)	*IGFBP2, HOXC8, IGFBP7*	Increased in small intestine (****)	No significant differences
				Distal IM (97)	*PNCK, MORN5, AHNAK2*	Increased in colon (***)	Increases over time in colon and overall
				Proximal MM (98)	*PLCG2, NPNT, IGF1*	Increased in small intestine (****)	No significant differences
				Distal MM (99)	*PLN, MOXD1, TMEM158*	Increased in colon (***)	Increases over time in colon and overall

				IM *PMAIP1+* (100)	*PMAIP1, HMGCS1, KRT17*	No locational preference	Increases then decreases over time in colon and overall. Increases in TI.
		IM progenitor	*N/A*	IM progenitor (101)	*HOXC5, PLPP2, MYOCD*	No locational preference	Decreases over time overall.

Table A.14 Overview of key genes from scRNA-seq data, divided by each cell compartment, that exhibit relatively high or specific expression in each cell type used for characterisation annotation ([a]further sub-division of secretory cells is also shown, [b]Comparison made as proportion of each compartment cells, if cluster is locational specific comparison of time-point difference made only in that location and overall. Locational and time-point differences: Wilcox rank test, p-value <0.05 *; p-value <0.01 **; p-value <0.01 ***, p-value <0.001****)

ST images

Within the thesis selected examples of areas from tissue are given alongside histology and gene expression or another feature. Full high resolution images of all slides are presented for completeness in supplementary materials[389]. Interrogation of specific genes, cell types or features of interest across slides can be performed interactively with the resource that accompanied the study available from: https://simmonslab.shinyapps.io/FetalAtlasDataPortal/

ST images DEGs for depth and GO enrichment

For calculation of genes expressed related to depth in ST data and the GEO enrichment of these genes, full results are displayed in supplementary materials[389].

Comparison of ISCs in TI and Colon

Direct comparison of genes expressed in proximal and distal ISCs was performed (**Figure 4.5**) and full genes list is available in supplementary materials[389].

Comparison between ISCs and proximal progenitors

Direct comparison was made between proximal progenitors and ISCs shown in **Figure 4.5**. Full genes list for this comparison is available in supplementary materials[389].

Comparison of S4 sub-types

Direct comparison was made between the two S4 clusters identified in fetal scRNA-seq data (S4 CXCL13+ and S4 CCL21), volcano plot is shown in **Figure 4.9** and full genes list is available in supplementary materials[389].

Morphogen list and morphogen modules

To study the expression and interaction of difference morphogens shown in **Figure 4.10** and described in **Methods 6.11** a list was curated of the receptors, ligands, co-receptors and antagonists of the pathways: Hedgehog, NOTCH, Wnt, HIPPO, RTK, TGF-Beta, FGF and EGFR signalling (n=182). This analysis generated co-expressed morphogen modules seen in scRNA-seq (n=11) and ST (n=13). Both the curated list and module list is available in supplementary materials[389].

Comparison of S2 locational differences

Direct comparison was made between S2 clusters identified in fetal scRNA-seq data to look at locational differences, volcano plot is shown in **Figure 4.10** and full genes list is available in supplementary materials[389].

Linking of resource to congenital disease

The HPO was searched for a curated list of intestinal congenital disorders and their associated phenotypes (**Figure 4.11, Methods 6.11**). A full list of analysis outcomes for genes which were cluster specific and those which varied with time, along with associated pathology are documented in supplementary materials[389].

References

1. Barker, N. Adult intestinal stem cells: Critical drivers of epithelial homeostasis and regeneration. *Nature Reviews Molecular Cell Biology* vol. 15 19–33 (2014).
2. Parikh, K. *et al.* Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature* **567**, 49–55 (2019).
3. Fawkner-Corbett, D. *et al.* Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* **184**, 810-826.e23 (2021).
4. General Information | HDBR. https://www.hdbr.org/general-information.
5. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
6. Weinstein, J. N. *et al.* The cancer genome atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
7. Helander, H. F. & Fändriks, L. Surface area of the digestive tract-revisited. *Scand. J. Gastroenterol.* **49**, 681–689 (2014).
8. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).
9. Gehart, H. & Clevers, H. Tales from the crypt: new insights into intestinal stem cells. *Nature Reviews Gastroenterology and Hepatology* vol. 16 19–34 (2019).
10. Ali, A., Tan, H. Y. & Kaiko, G. E. Role of the Intestinal Epithelium and Its Interaction With the Microbiota in Food Allergy. *Frontiers in Immunology* vol. 11 3222 (2020).
11. Peterson, L. W. & Artis, D. Intestinal epithelial cells: Regulators of barrier function and immune homeostasis. *Nature Reviews Immunology* vol. 14 141–153 (2014).
12. Fawkner-Corbett, D., Simmons, A. & Parikh, K. Microbiome, pattern recognition receptor function in health and inflammation. *Best Practice and Research: Clinical Gastroenterology* vol. 31 683–691 (2017).
13. Nicholson, J. K. *et al.* Host-gut microbiota metabolic interactions. *Science* vol. 336 1262–1267 (2012).
14. Johansson, M. E. V. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15064–15069 (2008).
15. Vincent, J. P. & Briscoe, J. Morphogens. *Current biology : CB* vol. 11 R851–R854 (2001).
16. Farin, H. F. *et al.* Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. *Nature* **530**, 340–343 (2016).
17. Mah, A. T., Yan, K. S. & Kuo, C. J. Wnt pathway regulation of intestinal stem cells. *Journal of Physiology* vol. 594 4837–4847 (2016).
18. Muncan, V. *et al.* Rapid Loss of Intestinal Crypts upon Conditional Deletion of the Wnt/Tcf-4 Target Gene c- Myc. *Mol. Cell. Biol.* **26**, 8418–8426 (2006).
19. Sansom, O. J. *et al.* Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev.* **18**, 1385–1390 (2004).
20. Qi, Z. *et al.* BMP restricts stemness of intestinal Lgr5 + stem cells by directly suppressing their signature genes. *Nat. Commun.* **8**, 1–14 (2017).
21. Davis, H. *et al.* Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche. *Nat. Med.* **21**, 62–70 (2015).
22. Haramis, A. P. G. *et al.* De Novo Crypt Formation and Juvenile Polyposis on BMP Inhibition in Mouse Intestine. *Science (80-.).* **303**, 1684–1686 (2004).

23. Ulshen, M. H., Lyn-Cook, L. E. & Raasch, R. H. Effects of Intraluminal Epidermal Growth Factor on Mucosal Proliferation in the Small Intestine of Adult Rats. *Gastroenterology* **91**, 1134–1140 (1986).
24. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).
25. Crosnier, C., Stamataki, D. & Lewis, J. Organizing cell renewal in the intestine: Stem cells, signals and combinatorial control. *Nature Reviews Genetics* vol. 7 349–359 (2006).
26. McCarthy, N. *et al.* Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient. *Cell Stem Cell* **26**, 391-402.e5 (2020).
27. Shoshkes-Carmel, M. *et al.* Subepithelial telocytes are an important source of Wnts that supports intestinal crypts. *Nature* **557**, 242–246 (2018).
28. Morson, B. C. Histopathology of the Small Intestine. *J. R. Soc. Med.* **52**, 6–10 (1959).
29. Van Der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annual Review of Physiology* vol. 71 241–260 (2009).
30. SPICER, S. S. A correlative study of the histochemical properties of rodent acid mucopolysaccharides. *J. Histochem. Cytochem.* **8**, 18–35 (1960).
31. Wiedenmann, B., Franke, W. W., Kuhn, C., Moll, R. & Gould, V. E. Synaptophysin: A marker protein for neuroendocrine cells and neoplasms. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3500–3504 (1986).
32. Rindi, G., Leiter, A. B., Kopin, A. S., Bordi, C. & Solcia, E. The 'normal' endocrine cell of the gut: Changing concepts and new evidences. *Annals of the New York Academy of Sciences* vol. 1014 1–12 (2004).
33. Elphick, D. A. & Mahida, Y. R. Paneth cells: Their role in innate immunity and inflammatory disease. *Gut* vol. 54 1802–1809 (2005).
34. Goldberg, R. F. *et al.* Intestinal alkaline phosphatase is a gut mucosal defense factor maintained by enteral nutrition. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3551–3556 (2008).
35. Worthington, J. J., Reimann, F. & Gribble, F. M. Enteroendocrine cells-sensory sentinels of the intestinal environment and orchestrators of mucosal immunity. *Mucosal Immunology* vol. 11 3–20 (2018).
36. Sasaki, N. *et al.* Reg4+ deep crypt secretory cells function as epithelial niche for Lgr5+ stem cells in colon. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5399–E5407 (2016).
37. Knoop, K. A., McDonald, K. G., McCrate, S., McDole, J. R. & Newberry, R. D. Microbial sensing by goblet cells controls immune surveillance of luminal antigens in the colon. *Mucosal Immunol.* **8**, 198–210 (2015).
38. McDole, J. R. *et al.* Goblet cells deliver luminal antigen to CD103 + dendritic cells in the small intestine. *Nature* **483**, 345–349 (2012).
39. Li, H. J. *et al.* Intestinal Neurod1 expression impairs paneth cell differentiation and promotes enteroendocrine lineage specification. *Sci. Rep.* **9**, 1–11 (2019).
40. 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).
41. Noah, T. K., Kazanjian, A., Whitsett, J. & Shroyer, N. F. SAM pointed domain ETS factor (SPDEF) regulates terminal differentiation and maturation of intestinal goblet cells. *Exp. Cell Res.* **316**, 452–465 (2010).
42. Kanaya, T. *et al.* The Ets transcription factor Spi-B is essential for the

differentiation of intestinal microfold cells. *Nat. Immunol.* **13**, 729–736 (2012).
43. McCauley, H. A. & Guasch, G. Three cheers for the goblet cell: Maintaining homeostasis in mucosal epithelia. *Trends in Molecular Medicine* vol. 21 492–503 (2015).
44. Deplancke, B. & Gaskins, H. R. Microbial modulation of innate defense: Goblet cells and the intestinal mucus layer. in *American Journal of Clinical Nutrition* vol. 73 1131S-1141S (Oxford Academic, 2001).
45. Ritsma, L. *et al.* Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging. *Nature* **507**, 362–365 (2014).
46. Barker, N., De Wetering, M. Van & Clevers, H. The intestinal stem cell. *Genes and Development* vol. 22 1856–1864 (2008).
47. Buczacki, S. J. A. *et al.* Intestinal label-retaining cells are secretory precursors expressing lgr5. *Nature* **495**, 65–69 (2013).
48. Tetteh, P. W. *et al.* Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell* **18**, 203–213 (2016).
49. Tian, H. *et al.* A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* **478**, 255–259 (2011).
50. Yan, K. S. *et al.* The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 466–471 (2012).
51. Jadhav, U. *et al.* Dynamic Reorganization of Chromatin Accessibility Signatures during Dedifferentiation of Secretory Precursors into Lgr5+ Intestinal Stem Cells. *Cell Stem Cell* **21**, 65-77.e5 (2017).
52. Buczacki, S. Fate plasticity in the intestine: The devil is in the detail. *World Journal of Gastroenterology* vol. 25 3116–3122 (2019).
53. Winton, D. J., Blount, M. A. & Ponder, B. A. J. A clonal marker induced by mutation in mouse intestinal epithelium. *Nature* **333**, 463–466 (1988).
54. Tetteh, P. W., Farin, H. F. & Clevers, H. Plasticity within stem cell hierarchies in mammalian epithelia. *Trends in Cell Biology* vol. 25 100–108 (2015).
55. Sangiorgi, E. & Capecchi, M. R. Bmi1 is expressed in vivo in intestinal stem cells. *Nat. Genet.* **40**, 915–920 (2008).
56. Montgomery, R. K. *et al.* Mouse telomerase reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 179–184 (2011).
57. Takeda, N. *et al.* Interconversion between intestinal stem cell populations in distinct niches. *Science (80-.).* **334**, 1420–1424 (2011).
58. Powell, A. E. *et al.* The pan-ErbB negative regulator lrig1 is an intestinal stem cell marker that functions as a tumor suppressor. *Cell* **149**, 146–158 (2012).
59. Stappenbeck, T. S., Mills, J. C. & Gordon, J. I. Molecular features of adult mouse small intestinal epithelial progenitors. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1004–1009 (2003).
60. Garabedian, E. M., Roberts, L. J. J., McNevin, M. S. & Gordon, J. I. Examining the role of Paneth cells in the small intestine by lineage ablation in transgenic mice. *J. Biol. Chem.* **272**, 23729–23740 (1997).
61. Wehkamp, J., Fellermann, K., Herrlinger, K. R., Bevins, C. L. & Stange, E. F. Mechanisms of disease: Defensins in gastrointestinal diseases. *Nature Clinical Practice Gastroenterology and Hepatology* vol. 2 406–415 (2005).
62. Sato, T. *et al.* Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
63. Gregorieff, A. *et al.* The ets-domain transcription factor Spdef promotes maturation

of goblet and paneth cells in the intestinal epithelium. *Gastroenterology* **137**, 1333 (2009).
64. Katz, J. P. *et al.* The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. *Development* vol. 129 2619–2628 (2002).
65. Van Es, J. H. *et al.* Dll1 + secretory progenitor cells revert to stem cells upon crypt damage. *Nat. Cell Biol.* **14**, 1099–1104 (2012).
66. Nagatake, T., Fujita, H., Minato, N. & Hamazaki, Y. Enteroendocrine cells are specifically marked by cell surface expression of claudin-4 in mouse small intestine. *PLoS One* **9**, e90638 (2014).
67. Boudry, G., Yang, P.-C. & Perdue, M. H. Small Intestine, Anatomy. in *Encyclopedia of Gastroenterology* 404–409 (Elsevier, 2004). doi:10.1016/b0-12-386860-2/00648-1.
68. Verzi, M. P. *et al.* TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15157–15162 (2010).
69. Bjerknes, M. *et al.* Origin of the brush cell lineage in the mouse intestinal epithelium. *Dev. Biol.* **362**, 194–218 (2012).
70. Mabbott, N. A., Donaldson, D. S., Ohno, H., Williams, I. R. & Mahajan, A. Microfold (M) cells: Important immunosurveillance posts in the intestinal epithelium. *Mucosal Immunology* vol. 6 666–677 (2013).
71. 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).
72. Knoop, K. A. *et al.* RANKL Is Necessary and Sufficient to Initiate Development of Antigen-Sampling M Cells in the Intestinal Epithelium. *J. Immunol.* **183**, 5738–5747 (2009).
73. Gerbe, F. *et al.* Intestinal epithelial tuft cells initiate type 2 mucosal immunity to helminth parasites. *Nature* **529**, 226–230 (2016).
74. Gerbe, F., Legraverend, C. & Jay, P. The intestinal epithelium tuft cells: Specification and function. *Cellular and Molecular Life Sciences* vol. 69 2907–2917 (2012).
75. Von Moltke, J., Ji, M., Liang, H. E. & Locksley, R. M. Tuft-cell-derived IL-25 regulates an intestinal ILC2-epithelial response circuit. *Nature* **529**, 221–225 (2016).
76. Yang, Q., Bermingham, N. A., Finegold, M. J. & Zoghbi, H. Y. Requirement of Math1 for secretory cell lineage commitment in the mouse intestine. *Science (80-.).* **294**, 2155–2158 (2001).
77. De Bakker, B. S. *et al.* An interactive three-dimensional digital atlas and quantitative database of human development. *Science (80-.).* **354**, (2016).
78. McLin, V. A., Henning, S. J. & Jamrich, M. The Role of the Visceral Mesoderm in the Development of the Gastrointestinal Tract. *Gastroenterology* vol. 136 2074–2091 (2009).
79. Spence, J. R., Lauf, R. & Shroyer, N. F. Vertebrate intestinal endoderm development. *Developmental Dynamics* vol. 240 501–520 (2011).
80. Sherwood, R. I., Chen, T. Y. A. & Melton, D. A. Transcriptional dynamics of endodermal organ formation. *Dev. Dyn.* **238**, 29–42 (2009).
81. Gao, N., White, P. & Kaestner, K. H. Establishment of Intestinal Identity and Epithelial-Mesenchymal Signaling by Cdx2. *Dev. Cell* **16**, 588–599 (2009).
82. Que, J. *et al.* Multiple dose-dependent roles for Sox2 in the patterning and differentiation of anterior foregut endoderm. *Development* **134**, 2521–2531 (2007).

83. Grey, R. D. Morphogenesis of intestinal villi. I. Scanning electron microscopy of the duodenal epithelium of the developing chick embryo. *J. Morphol.* **137**, 193–213 (1972).
84. Shyer, A. E. *et al.* Villification: How the gut gets its villi. *Science (80-.).* **342**, 212–218 (2013).
85. Walton, K. D. *et al.* Villification in the mouse: Bmp signals control intestinal villus patterning. *Dev.* **143**, 427–436 (2016).
86. Montgomery, R. K., Mulberg, A. E. & Grand, R. J. Development of the human gastrointestinal tract: Twenty years of progress. *Gastroenterology* **116**, 702–731 (1999).
87. Lim, A. A., Nadkarni, R. R., Courteau, B. C. & Draper, J. S. Comparison of human and mouse fetal intestinal tissues reveals differential maturation timelines. *bioRxiv* 2020.06.18.157818 (2020) doi:10.1101/2020.06.18.157818.
88. Neu, J. & Walker, W. A. Necrotizing Enterocolitis. *N. Engl. J. Med.* **364**, 255 (2011).
89. Schreurs, R. R. C. E. *et al.* Human Fetal TNF-α-Cytokine-Producing CD4 + Effector Memory T Cells Promote Intestinal Development and Mediate Inflammation Early in Life. *Immunity* **50**, 462-476.e8 (2019).
90. 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* **558**, 449–453 (2018).
91. Savidge, T. C. *et al.* Enteric Glia Regulate Intestinal Barrier Function and Inflammation Via Release of S-Nitrosoglutathione. *Gastroenterology* **132**, 1344–1358 (2007).
92. Cabarrocas, J., Savidge, T. C. & Liblau, R. S. Role of enteric glial cells in inflammatory bowel disease. *GLIA* vol. 41 81–93 (2003).
93. Toyoda, H., Ina, K., Kitamura, H., Tsuda, T. & Shimada, T. Organization of the lamina propria mucosae of rat intestinal mucosa, with special reference to the subepithelial connective tissue. *Cells Tissues Organs* **158**, 172–184 (1997).
94. 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 (2017).
95. Roulis, M. & Flavell, R. A. Fibroblasts and myofibroblasts of the intestinal lamina propria in physiology and disease. *Differentiation* vol. 92 116–131 (2016).
96. Pinchuk, I. V., Mifflin, R. C., Saada, J. I. & Powell, D. W. Intestinal mesenchymal cells. *Current Gastroenterology Reports* vol. 12 310–318 (2010).
97. Eyden, B. The myofibroblast: Phenotypic characterization as a prerequisite to understanding its functions in translational medicine: Translational Medicine. *J. Cell. Mol. Med.* **12**, 22–37 (2008).
98. Vannucchi, M. G., Traini, C., Manetti, M., Ibba-Manneschi, L. & Faussone-Pellegrini, M. S. Telocytes express PDGFRα in the human gastrointestinal tract. *J. Cell. Mol. Med.* **17**, 1099–1108 (2013).
99. Díaz-Flores, L. *et al.* Pericytes. Morphofunction, interactions and pathology in a quiescent and activated mesenchymal cell niche. *Histology and Histopathology* vol. 24 909–969 (2009).
100. Eyden, B., Curry, A. & Wang, G. Stromal cells in the human gut show ultrastructural features of fibroblasts and smooth muscle cells but not myofibroblasts. *J. Cell. Mol. Med.* **15**, 1483–1491 (2011).
101. Kurahashi, M. *et al.* novel population of subepithelial platelet-derived growth factor receptor α-positive cells in the mouse and human colon. *Am. J. Physiol.* -

Gastrointest. Liver Physiol. **304**, (2013).
102. Bernardo, M. E. & Fibbe, W. E. Mesenchymal stromal cells: Sensors and switchers of inflammation. *Cell Stem Cell* vol. 13 392–402 (2013).
103. Flier, S. N. *et al.* Identification of epithelial to mesenchymal transition as a novel source of fibroblasts in intestinal fibrosis. *J. Biol. Chem.* **285**, 20202–20212 (2010).
104. Koyama, M. *et al.* Recipient nonhematopoietic antigen-presenting cells are sufficient to induce lethal acute graft-versus-host disease. *Nat. Med.* **18**, 135–142 (2012).
105. Powell, D. W., Pinchuk, I. V., Saada, J. I., Chen, X. & Mifflin, R. C. Mesenchymal cells of the intestinal lamina propria. *Annu. Rev. Physiol.* **73**, 213–237 (2011).
106. Ramirez, M., Pell, N., Mejias, M. & Fernandez, M. Pericytes in the gut. in *Advances in Experimental Medicine and Biology* vol. 1122 73–100 (Springer, Cham, 2019).
107. Armulik, A., Genové, G. & Betsholtz, C. Pericytes: Developmental, Physiological, and Pathological Perspectives, Problems, and Promises. *Developmental Cell* vol. 21 193–215 (2011).
108. Wu, C. F. *et al.* Transforming growth factor β-1 stimulates profibrotic epithelial signaling to activate pericyte-myofibroblast transition in obstructive kidney fibrosis. *Am. J. Pathol.* **182**, 118–131 (2013).
109. Wilm, B., Ipenberg, A., Hastie, N. D., Burch, J. B. E. & Bader, D. M. The serosal mesothelium is a major source of smooth muscle cells of the gut vasculature. *Development* **132**, 5317–5328 (2005).
110. Privratsky, J. R. & Newman, P. J. PECAM-1: Regulator of endothelial junctional integrity. *Cell and Tissue Research* vol. 355 607–619 (2014).
111. Stockinger, B., Shah, K. & Wincent, E. AHR in the intestinal microenvironment: safeguarding barrier function. *Nature Reviews Gastroenterology and Hepatology* vol. 18 559–570 (2021).
112. Daneman, R. & Rescigno, M. The Gut Immune Barrier and the Blood-Brain Barrier: Are They So Different? *Immunity* vol. 31 722–735 (2009).
113. Bernier-Latmani, J. & Petrova, T. V. Intestinal lymphatic vasculature: structure, mechanisms and functions. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 510–526 (2017).
114. Dash, S., Xiao, C., Morgantini, C. & Lewis, G. F. New Insights into the Regulation of Chylomicron Production. *Annual Review of Nutrition* vol. 35 265–294 (2015).
115. Neal, M. D. *et al.* Enterocyte TLR4 Mediates Phagocytosis and Translocation of Bacteria Across the Intestinal Barrier. *J. Immunol.* **176**, 3070–3079 (2006).
116. Zheng, M., Kimura, S., Nio-Kobayashi, J. & Iwanaga, T. The selective distribution of LYVE-1-expressing endothelial cells and reticular cells in the reticulo-endothelial system (RES). *Biomed. Res.* **37**, 187–198 (2016).
117. Spadoni, I. *et al.* A gut-vascular barrier controls the systemic dissemination of bacteria. *Science (80-.).* **350**, 830–834 (2015).
118. Fätsignndriks, L. The angiotensin II type 2 receptor and the gastrointestinal tract. *JRAAS - Journal of the Renin-Angiotensin-Aldosterone System* vol. 11 43–48 (2010).
119. Uda, Y. *et al.* Angiogenesis is crucial for liver regeneration after partial hepatectomy. *Surg. (United States)* **153**, 70–77 (2013).
120. Potente, M. & Mäkinen, T. Vascular heterogeneity and specialization in development and disease. *Nature Reviews Molecular Cell Biology* vol. 18 477–494 (2017).
121. Gershon, M. D. & Rothman, T. P. Enteric glia. *Glia* **4**, 195–204 (1991).
122. Von Haller, A. & Temkin, O. *A dissertation on the sensible and irritable parts of animals*. (The John Hopkins press, 1936).

123. Spencer, N. J. & Hu, H. Enteric nervous system: sensory transduction, neural circuits and gastrointestinal motility. *Nature Reviews Gastroenterology and Hepatology* vol. 17 338–351 (2020).
124. Mckeown, S. J., Stamp, L., Hao, M. M. & Young, H. M. Hirschsprung disease: A developmental disorder of the enteric nervous system. *Wiley Interdiscip. Rev. Dev. Biol.* **2**, 113–129 (2013).
125. Alcaino, C. *et al.* A population of gut epithelial enterochromaffin cells is mechanosensitive and requires Piezo2 to convert force into serotonin release. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E7632–E7641 (2018).
126. Kaelberer, M. M. *et al.* A gut-brain neural circuit for nutrient sensory transduction. *Science (80-.).* **361**, (2018).
127. Delvalle, N. M. *et al.* Communication Between Enteric Neurons, Glia, and Nociceptors Underlies the Effects of Tachykinins on Neuroinflammation. *CMGH* **6**, 321–344 (2018).
128. Selgrad, M. *et al.* JC virus infects the enteric glia of patients with chronic idiopathic intestinal pseudo-obstruction. *Gut* **58**, 25–32 (2009).
129. Rao, M. *et al.* Enteric glia express proteolipid protein 1 and are a transcriptionally unique population of glia in the mammalian nervous system. *Glia* **63**, 2040–2057 (2015).
130. Boesmans, W., Lasrado, R., Vanden Berghe, P. & Pachnis, V. Heterogeneity and phenotypic plasticity of glial cells in the mammalian enteric nervous system. *Glia* **63**, 229–241 (2015).
131. Sender, R., Fuchs, S. & Milo, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* vol. 164 337–340 (2016).
132. Sender, R., Fuchs, S. & Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol.* **14**, e1002533 (2016).
133. Shao, Y. *et al.* Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574**, 117–121 (2019).
134. Bäckhed, F. *et al.* Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
135. De Vadder, F. *et al.* Gut microbiota regulates maturation of the adult enteric nervous system via enteric serotonin networks. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6458–6463 (2018).
136. Van De Pavert, S. A. & Mebius, R. E. New insights into the development of lymphoid tissues. *Nature Reviews Immunology* vol. 10 664–674 (2010).
137. Olin, A. *et al.* Stereotypic Immune System Development in Newborn Children. *Cell* **174**, 1277-1292.e14 (2018).
138. Zhou, Q. *et al.* Necrotizing enterocolitis induces T lymphocyte–mediated injury in the developing mammalian brain. *Sci. Transl. Med.* **13**, (2021).
139. Kulkarni, D. H. *et al.* Goblet cell associated antigen passages support the induction and maintenance of oral tolerance. *Mucosal Immunol.* **13**, 271–282 (2020).
140. Mörbe, U. M. *et al.* Human gut-associated lymphoid tissues (GALT); diversity, structure, and function. *Mucosal Immunology* vol. 14 793–802 (2021).
141. Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nature Reviews Immunology* vol. 14 667–685 (2014).
142. Mueller, S. N. & Mackay, L. K. Tissue-resident memory T cells: Local specialists in immune defence. *Nature Reviews Immunology* vol. 16 79–89 (2016).
143. Corridoni, D., Chapman, T., Antanaviciute, A., Satsangi, J. & Simmons, A. Inflammatory bowel disease through the lens of single-cell RNA-seq technologies. *Inflammatory Bowel Diseases* vol. 26 1658–1668 (2020).

144. Harrison, O. J. & Powrie, F. M. Regulatory T cells and immune tolerance in the intestine. *Cold Spring Harb. Perspect. Biol.* **5**, a018341 (2013).
145. Scott, C. L., Aumeunier, A. M. & Mowat, A. M. I. Intestinal CD103 + dendritic cells: Master regulators of tolerance? *Trends in Immunology* vol. 32 412–419 (2011).
146. Rao, J., Wang, J. & Rafael, S. Intestinal Architecture and Development. *Regul. Gastrointest. Mucosal Growth.* 1–5 (2010).
147. Fu, M., Tam, P. K. H., Sham, M. H. & Lui, V. C. H. Embryonic development of the ganglion plexuses and the concentric layer structure of human gut: A topographical study. *Anat. Embryol. (Berl).* **208**, 33–41 (2004).
148. Wallace, A. S. & Burns, A. J. Development of the enteric nervous system, smooth muscle and interstitial cells of Cajal in the human gastrointestinal tract. *Cell Tissue Res.* **319**, 367–382 (2005).
149. Sanders, K. M. Nerves, smooth muscle cells and interstitial cells in the gi tract: Molecular and cellular interactions. in *Clinical and Basic Neurogastroenterology and Motility* 3–16 (Academic Press, 2019). doi:10.1016/B978-0-12-813037-7.00001-7.
150. Sanders, K. M., Ward, S. M. & Koh, S. D. Interstitial cells: Regulators of smooth muscle function. *Physiological Reviews* vol. 94 859–907 (2014).
151. Akiho, H. Cytokine-induced alterations of gastrointestinal motility in gastrointestinal disorders. *World J. Gastrointest. Pathophysiol.* **2**, 72 (2011).
152. Muller, P. A. *et al.* Erratum: Crosstalk between muscularis macrophages and enteric neurons regulates gastrointestinal motility. *Cell* vol. 158 1210 (2014).
153. Chen, W. *et al.* Smooth muscle hyperplasia/hypertrophy is the most prominent histological change in Crohn's fibrostenosing bowel strictures: A semiquantitative analysis by using a novel histological grading scheme. *J. Crohn's Colitis* **11**, 92–104 (2017).
154. Bosman, F. T. & Yan, P. Development, Structure, and Function of the Tubal Gut. in *Pathobiology of Human Disease: A Dynamic Encyclopedia of Disease Mechanisms* 1242–1254 (Academic Press, 2014). doi:10.1016/B978-0-12-386456-7.03801-6.
155. Rinkevich, Y. *et al.* Identification and prospective isolation of a mesothelial precursor lineage giving rise to smooth muscle cells and fibroblasts for mammalian internal organs, and their vasculature. *Nat. Cell Biol.* **14**, 1251–1260 (2012).
156. Bertram, P. *et al.* Intraperitoneal transplantation of isologous mesothelial cells for prevention of adhesions. *Eur. J. Surg.* **165**, 705–709 (1999).
157. Wilkinson, L., De, P. & Bloxham, C. Mesothelial reaction in longstanding Crohn's ileitis simulating papillary mesothelioma. *J. Clin. Pathol.* **61**, 1119–1121 (2008).
158. Wright, N. J. *et al.* Management and outcomes of gastrointestinal congenital anomalies in low, middle and high income countries: Protocol for a multicentre, international, prospective cohort study. *BMJ Open* **9**, e030452 (2019).
159. Wright, N. J. *et al.* Mortality from gastrointestinal congenital anomalies at 264 hospitals in 74 low-income, middle-income, and high-income countries: a multicentre, international, prospective cohort study. *Lancet* **398**, 325–339 (2021).
160. Beaudoin, S. Insights into the etiology and embryology of gastroschisis. *Semin. Pediatr. Surg.* **27**, 283–288 (2018).
161. Huff, D. & Russo, P. *Chapter 8 – Congenital and Developmental Disorders of the Gastrointestinal Tract*. *Odze and Goldblum Surgical Pathology of the GI Tract, Liver, Biliary Tract and Pancreas* (Elsevier Inc., 2015).
162. Frykman, P. K. & Short, S. S. Hirschsprung-associated enterocolitis: Prevention and therapy. *Semin. Pediatr. Surg.* **21**, 328–335 (2012).

163. Gosain, A. & Brinkman, A. S. Hirschsprung's associated enterocolitis. *Current Opinion in Pediatrics* vol. 27 364–369 (2015).
164. Geremia, A., Biancheri, P., Allan, P., Corazza, G. R. & Di Sabatino, A. Innate and adaptive immunity in inflammatory bowel disease. *Autoimmunity Reviews* vol. 13 3–10 (2014).
165. Creed, T. J. & Probert, C. S. J. Review article: Steroid resistance in inflammatory bowel disease - Mechanisms and therapeutic strategies. *Alimentary Pharmacology and Therapeutics* vol. 25 111–122 (2007).
166. Ghosh, N. & Premchand, P. A UK cost of care model for inflammatory bowel disease. *Frontline Gastroenterol.* **6**, 169–174 (2015).
167. Benchimol, E. I. *et al.* Trends in epidemiology of pediatric inflammatory bowel disease in Canada: Distributed network analysis of multiple population-based provincial health administrative databases. *Am. J. Gastroenterol.* **112**, 1120–1134 (2017).
168. Birimberg-Schwartz, L. *et al.* PANCA and ASCA in Children with IBD-Unclassified, Crohn's Colitis, and Ulcerative Colitis - A Longitudinal Report from the IBD Porto Group of ESPGHAN. *Inflamm. Bowel Dis.* **22**, 1908–1914 (2016).
169. Emrich, S. J., Barbazuk, W. B., Li, L. & Schnable, P. S. Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res.* **17**, 69–73 (2007).
170. Ravasi, T. *et al.* An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man. *Cell* **140**, 744–752 (2010).
171. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).
172. Yu, N. Y. L. *et al.* Complementing tissue characterization by integrating transcriptome profiling from the human protein atlas and from the FANTOM5 consortium. *Nucleic Acids Res.* **43**, 6787–6798 (2015).
173. Czarnewski, P. *et al.* Conserved transcriptomic profile between mouse and human colitis allows unsupervised patient stratification. *Nat. Commun.* **10**, 1–11 (2019).
174. Haberman, Y. *et al.* Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. *Nat. Commun.* **10**, 1–13 (2019).
175. Haberman, Y. *et al.* Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J. Clin. Invest.* **124**, 3617–3633 (2014).
176. Tremblay, E. *et al.* Gene-expression profile analysis in the mid-gestation human intestine discloses greater functional immaturity of the colon as compared with the ileum. *J. Pediatr. Gastroenterol. Nutr.* **52**, 670–678 (2011).
177. 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).
178. Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3010–3014 (1992).
179. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
180. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1098 (2013).
181. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
182. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science (80-.).* **357**, 661–667 (2017).
183. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula

Muris. *Nature* **562**, 367–372 (2018).
184. Plasschaert, L. W. *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 (2018).
185. Boldog, E. *et al.* Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nat. Neurosci.* **21**, 1185–1195 (2018).
186. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nature Protocols* vol. 13 599–604 (2018).
187. Regev, A. *et al.* The human cell atlas. *Elife* **6**, (2017).
188. Barriga, F. M. *et al.* Mex3a Marks a Slowly Dividing Subpopulation of Lgr5+ Intestinal Stem Cells. *Cell Stem Cell* **20**, 801-816.e7 (2017).
189. Basak, O. *et al.* Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone-Producing Enteroendocrine Cells. *Cell Stem Cell* **20**, 177-190.e4 (2017).
190. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
191. Yan, K. S. *et al.* Non-equivalence of Wnt and R-spondin ligands during Lgr5 + intestinal stem-cell self-renewal. *Nature* **545**, 238–242 (2017).
192. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
193. Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.* **49**, 708–718 (2017).
194. 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).
195. Kinchen, J. *et al.* Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. *Cell* **175**, 372-386.e17 (2018).
196. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* vol. 353 78–82 (2016).
197. Grossmann, J. *et al.* Progress on isolation and short-term ex-vivo culture of highly purified non-apoptotic human intestinal epithelial cells (IEC). *European Journal of Cell Biology* vol. 82 262–270 (2003).
198. Ikeya, K. *et al.* The ulcerative colitis endoscopic index of severity more accurately reflects clinical outcomes and long-term prognosis than the mayo endoscopic score. *J. Crohn's Colitis* **10**, 286–295 (2016).
199. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* **29**, 1120–1127 (2011).
200. Comelli, E. M. *et al.* Biomarkers of human gastrointestinal tract regions. *Mamm. Genome* **20**, 516–527 (2009).
201. Holgersen, K. *et al.* High-resolution gene expression profiling using RNA sequencing in patients with inflammatory bowel disease and in mouse models of colitis. *J. Crohn's Colitis* **9**, 492–506 (2015).
202. Speckmann, B., Bidmon, H. J., Borchardt, A., Sies, H. & Steinbrenner, H. Intestinal selenoprotein P in epithelial cells and in plasma cells. *Arch. Biochem. Biophys.* **541**, 30–36 (2014).
203. Gunawardene, A. R., Corfe, B. M. & Staton, C. A. Classification and functions of enteroendocrine cells of the lower gastrointestinal tract. *International Journal of Experimental Pathology* vol. 92 219–231 (2011).
204. Jafarnejad, S. M., Ardekani, G. S., Ghaffari, M. & Li, G. Pleiotropic function of SRY-related HMG box transcription factor 4 in regulation of tumorigenesis.

Cellular and Molecular Life Sciences vol. 70 2677–2696 (2013).
205. Beumer, J. *et al.* High-Resolution mRNA and Secretome Atlas of Human Enteroendocrine Cells. *Cell* **181**, 1291-1306.e19 (2020).
206. Gracz, A. D. *et al.* Sox4 Promotes Atoh1-Independent Intestinal Secretory Differentiation Toward Tuft and Enteroendocrine Fates. *Gastroenterology* **155**, 1508-1523.e10 (2018).
207. Müller, S. *et al.* Galectin-3 modulates T cell activity and is reduced in the inflamed intestinal epithelium in IBD. *Inflamm. Bowel Dis.* **12**, 588–597 (2006).
208. Biton, M. *et al.* T Helper Cell Cytokines Modulate Intestinal Stem Cell Renewal and Differentiation. *Cell* **175**, 1307-1320.e22 (2018).
209. Narayan, S. B. *et al.* Short-chain 3-hydroxyacyl-coenzyme a dehydrogenase associates with a protein super-complex integrating multiple metabolic pathways. *PLoS One* **7**, e35048 (2012).
210. Fukui, H. *et al.* DMBT1 is a novel gene induced by IL-22 in ulcerative colitis. *Inflamm. Bowel Dis.* **17**, 1177–1188 (2011).
211. Tang, M. S. *et al.* Integrated Analysis of Biopsies from Inflammatory Bowel Disease Patients Identifies SAA1 as a Link between Mucosal Microbes with T H 17 and T H 22 Cells. *Inflamm. Bowel Dis.* **23**, 1544–1554 (2017).
212. Chen, C. L., Yang, J., James, I. O. A., Zhang, H. Y. & Besner, G. E. Heparin-binding epidermal growth factor-like growth factor restores Wnt/β-catenin signaling in intestinal stem cells exposed to ischemia/reperfusion injury. *Surg. (United States)* **155**, 1069–1080 (2014).
213. Ruan, P. *et al.* Expression and clinical significance of CD74 and MMP-9 in colon adenocarcinomas. *J. BUON.* **25**, 927–932 (2020).
214. Guess, C. M., Lafleur, B. J., Weidow, B. L. & Quaranta, V. A decreased ratio of laminin-332 β3 to γ2 subunit mRNA is associated with poor prognosis in colon cancer. *Cancer Epidemiol. Biomarkers Prev.* **18**, 1584–1590 (2009).
215. Spenlé, C. *et al.* Dysregulation of laminins in intestinal inflammation. *Pathol. Biol.* **60**, 41–47 (2012).
216. Kirchhoff, C., Habben, I., Iveli, R. & Krull, N. A major human epididymis-specific cDNA encodes a protein with sequence homology to extracellular proteinase inhibitors. *Biol. Reprod.* **45**, 350–357 (1991).
217. Motta, J. P. *et al.* Food-grade bacteria expressing elafin protect against inflammation and restore colon homeostasis. *Sci. Transl. Med.* **4**, (2012).
218. Hua, L. *et al.* Expression and biochemical characterization of recombinant human epididymis protein 4. *Protein Expr. Purif.* **102**, 52–62 (2014).
219. Zhang, T. *et al.* WFDC2 gene deletion in mouse led to severe dyspnea and type-I alveolar cell apoptosis. *Biochem. Biophys. Res. Commun.* **522**, 456–462 (2020).
220. Lee, S. M. *et al.* Bacterial colonization factors control specificity and stability of the gut microbiota. *Nature* **501**, 426–429 (2013).
221. Swidsinski, A., Loening-Baucke, V., Lochs, H. & Hale, L. P. Spatial organization of bacterial flora in normal and inflamed intestine: A fluorescence in situ hybridization study in mice. *World J. Gastroenterol.* **11**, 1131–1140 (2005).
222. Tropini, C., Earle, K. A., Huang, K. C. & Sonnenburg, J. L. The Gut Microbiome: Connecting Spatial Organization to Function. *Cell Host and Microbe* vol. 21 433–442 (2017).
223. Huang, X. *et al.* A Unique Conformational Distortion Mechanism Drives Lipocalin 2 Binding to Bacterial Siderophores. *ACS Chem. Biol.* **15**, 234–242 (2020).
224. Lindén, S. K., Florin, T. H. J. & McGuckin, M. A. Mucin dynamics in intestinal bacterial infection. *PLoS One* **3**, e3952 (2008).

225. Yang, G. Z. *et al.* Prognostic value of carbonic anhydrase VII expression in colorectal carcinoma. *BMC Cancer* **15**, 1–11 (2015).
226. Zhang, J. *et al.* Carbonic anhydrase IV inhibits colon cancer development by inhibiting the Wnt signalling pathway through targeting the WTAP-WT1-TBL1 axis. *Gut* **65**, 1482–1493 (2016).
227. Hartzell, H. C., Qu, Z., Yu, K., Xiao, Q. & Chien, L. T. Molecular physiology of bestrophins: Multifunctional membrane proteins linked to best disease and other retinopathies. *Physiological Reviews* vol. 88 639–672 (2008).
228. Ito, G. *et al.* Lineage-specific expression of bestrophin-2 and bestrophin-4 in human intestinal epithelial cells. *PLoS One* **8**, e79693 (2013).
229. Tu, Y. H. *et al.* An evolutionarily conserved gene family encodes proton-selective ion channels. *Science (80-.).* **359**, 1047–1050 (2018).
230. Okumura, R. *et al.* Lypd8 promotes the segregation of flagellated microbiota and colonic epithelia. *Nature* **532**, 117–121 (2016).
231. Ikpa, P. T. *et al.* Guanylin and uroguanylin are produced by mouse intestinal epithelial cells of columnar and secretory lineage. *Histochem. Cell Biol.* **146**, 445–455 (2016).
232. Sato, M. & Bremner, I. Oxygen free radicals and metallothionein. *Free Radic. Biol. Med.* **14**, 325–337 (1993).
233. Xiao, F. *et al.* Slc26a3 deficiency is associated with loss of colonic HCO3- secretion, absence of a firm mucus layer and barrier impairment in mice. *Acta Physiol.* **211**, 161–175 (2014).
234. Farmer, A. D., Mohammed, S. D., Dukes, G. E., Mark Scott, S. & Hobson, A. R. Caecal pH is a biomarker of excessive colonic fermentation. *World J. Gastroenterol.* **20**, 5000–5007 (2014).
235. Kumar, M. *et al.* Increased intestinal permeability exacerbates sepsis through reduced hepatic SCD-1 activity and dysregulated iron recycling. *Nat. Commun.* **11**, 1–15 (2020).
236. VanDussen, K. L. *et al.* Notch signaling modulates proliferation and differentiation of intestinal crypt base columnar stem cells. *Development* **139**, 488–497 (2012).
237. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
238. Blood atlas - FFAR4 - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000142959-BEST4/blood/granulocytes#hpa_neutrophil.
239. Can I process neutrophils (or other granulocytes) using 10x Single Cell applications? – 10X Genomics. https://kb.10xgenomics.com/hc/en-us/articles/360004024032-Can-I-process-neutrophils-or-other-granulocytes-using-10x-Single-Cell-applications-.
240. Finzi, G. *et al.* Cathepsin E in follicle associated epithelium of intestine and tonsils: localization to M cells and possible role in antigen processing. *Histochemistry* **99**, 201–211 (1993).
241. Shiraishi, T. *et al.* Mesothelin expression has prognostic value in stage II/III colorectal cancer. *Virchows Arch.* **474**, 297–307 (2019).
242. D'Arcangelo, D. *et al.* WIPI1, BAG1, and PEX3 autophagy-related genes are relevant melanoma markers. *Oxid. Med. Cell. Longev.* **2018**, (2018).
243. Cappelli, K. *et al.* Guanylin, uroguanylin and guanylate cyclase-c are expressed in the gastrointestinal tract of horses. *Front. Physiol.* **10**, (2019).
244. Waldman, S. A. & Camilleri, M. Guanylate cyclase-C as a therapeutic target in gastrointestinal disorders. *Gut* **67**, 1543–1552 (2018).

245. Rappaport, J. A. & Waldman, S. A. An update on guanylyl cyclase C in the diagnosis, chemoprevention, and treatment of colorectal cancer. *Expert Review of Clinical Pharmacology* vol. 13 1125–1137 (2020).
246. Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647-1660.e19 (2019).
247. COULOMBRE, A. J. & COULOMBRE, J. L. Intestinal development. I. Morphogenesis of the villi and musculature. *J. Embryol. Exp. Morphol.* **6**, 403–411 (1958).
248. Shyer, A. E., Huycke, T. R., Lee, C., Mahadevan, L. & Tabin, C. J. Bending Gradients: How the intestinal stem cell gets its home. *Cell* **161**, 569–580 (2015).
249. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 1–12 (2018).
250. Wang, Y. *et al.* Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J. Exp. Med.* **217**, (2020).
251. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).
252. Moor, A. E. *et al.* Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the Intestinal Villus Axis. *Cell* **175**, 1156-1167.e15 (2018).
253. Hamada, H., Meno, C., Watanabe, D. & Saijoh, Y. Establishment of vertebrate left-right asymmetry. *Nature Reviews Genetics* vol. 3 103–113 (2002).
254. Brafman, D. A., Phung, C., Kumar, N. & Willert, K. Regulation of endodermal differentiation of human embryonic stem cells through integrin-ECM interactions. *Cell Death Differ.* **20**, 369–381 (2013).
255. Kohlnhofer, B. M., Thompson, C. A., Walker, E. M. & Battle, M. A. GATA4 Regulates Epithelial Cell Proliferation to Control Intestinal Growth and Development in Mice. *CMGH* **2**, 189–209 (2016).
256. Anderson, G. J., Walsh, M. D., Powell, L. W. & Halliday, J. W. Intestinal transferrin receptors and iron absorption in the neonatal rat. *Br. J. Haematol.* **77**, 229–236 (1991).
257. Pu, Y. *et al.* Iron promotes intestinal development in neonatal piglets. *Nutrients* **10**, 726 (2018).
258. Borman, R. A. & Burleigh, D. E. Functional evidence for a 5-HT2B receptor mediating contraction of longitudinal muscle in human small intestine. *Br. J. Pharmacol.* **114**, 1525–1527 (1995).
259. Li, P. *et al.* Homeostatic control of the crypt-villus axis by the bacterial enterotoxin receptor guanylyl cyclase C restricts the proliferating compartment in intestine. *Am. J. Pathol.* **171**, 1847–1858 (2007).
260. Steinbrecher, K. A., Wowk, S. A., Rudolph, J. A., Witte, D. P. & Cohen, M. B. Targeted inactivation of the mouse guanylin gene results in altered dynamics of colonic epithelial proliferation. *Am. J. Pathol.* **161**, 2169–2178 (2002).
261. Lasrado, R. *et al.* Neurodevelopment: Lineage-dependent spatial and functional organization of the mammalian enteric nervous system. *Science (80-.).* **356**, 722–726 (2017).
262. Nelson, B. R., Hodge, R. D., Bedogni, F. & Hevner, R. F. Dynamic interactions between intermediate neurogenic progenitors and radial glia in embryonic mouse neocortex: Potential role in Dll1-notch signaling. *J. Neurosci.* **33**, 9122–9139 (2013).
263. Muhl, L. *et al.* Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat. Commun.* **11**, 1–18 (2020).

264. Croft, A. P. *et al.* Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature* **570**, 246–251 (2019).
265. Hardy, M. M., Feder, J., Wolfe, R. A. & Bu, G. Low density lipoprotein receptor-related protein modulates the expression of tissue-type plasminogen activator in human colon fibroblasts. *J. Biol. Chem.* **272**, 6812–6817 (1997).
266. Zook, E. C. *et al.* The ETS1 transcription factor is required for the development and cytokine-induced expansion of ILC2. *J. Exp. Med.* **213**, 687–696 (2016).
267. Boos, M. D., Yokota, Y., Eberl, G. & Kee, B. L. Mature natural killer cell and lymphoid tissue-inducing cell development requires Id2-mediated suppression of E protein activity. *J. Exp. Med.* **204**, 1119–1130 (2007).
268. Veiga-Fernandes, H. *et al.* Tyrosine kinase receptor RET is a key regulator of Peyer's Patch organogenesis. *Nature* **446**, 547–551 (2007).
269. Ohl, L. *et al.* Cooperating mechanisms of CXCR5 and CCR7 in development and organization of secondary lymphoid organs. *J. Exp. Med.* **197**, 1199–1204 (2003).
270. Davis, H. *et al.* Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche. *Nat. Med.* **21**, 62–70 (2015).
271. Gaultier, A., Hollister, M., Reynolds, I., Hsieh, E. hui & Gonias, S. L. LRP1 regulates remodeling of the extracellular matrix by fibroblasts. *Matrix Biol.* **29**, 22–30 (2010).
272. Fenderico, N. *et al.* Anti-LRP5/6 VHHs promote differentiation of Wnt-hypersensitive intestinal stem cells. *Nat. Commun.* **10**, (2019).
273. Arnaud-Dabernat, S., Yadav, D. & Sarvetnick, N. FGFR3 contributes to intestinal crypt cell growth arrest. *J. Cell. Physiol.* **216**, 261–268 (2008).
274. Adams, S. D. & Stanton, M. P. Malrotation and intestinal atresias. *Early Human Development* vol. 90 921–925 (2014).
275. Haas, J. T. *et al.* DGAT1 mutation is linked to a congenital diarrheal disorder. *J. Clin. Invest.* **122**, 4680–4684 (2012).
276. Holt-Danborg, L. *et al.* SPINT2 (HAI-2) missense variants identified in congenital sodium diarrhea/tufting enteropathy affect the ability of HAI-2 to inhibit prostasin but not matriptase. *Hum. Mol. Genet.* **28**, 828–841 (2019).
277. Heuckeroth, R. O. Hirschsprung disease - Integrating basic science and clinical medicine to improve outcomes. *Nature Reviews Gastroenterology and Hepatology* vol. 15 152–167 (2018).
278. Germain, D. P. Ehlers-Danlos syndrome type IV. *Orphanet Journal of Rare Diseases* vol. 2 1–9 (2007).
279. Byers, P. H. Vascular Ehlers-Danlos Syndrome. *GeneReviews®* (2019).
280. Rajab, A. *et al.* Fatal cardiac arrhythmia and long-QT syndrome in a new form of congenital generalized lipodystrophy with muscle rippling (CGL4) due to PTRF-CAVIN mutations. *PLoS Genet.* **6**, (2010).
281. 12q14 microdeletion syndrome. *Definitions* https://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=EN&Expert=94063 (2020) doi:10.32388/q9kkjg.
282. White, J. J. *et al.* WNT Signaling Perturbations Underlie the Genetic Heterogeneity of Robinow Syndrome. *Am. J. Hum. Genet.* **102**, 27–43 (2018).
283. Fitzgibbons, S. C. *et al.* Mortality of necrotizing enterocolitis expressed by birth weight categories. *J. Pediatr. Surg.* **44**, 1072–1076 (2009).
284. Ravisankar, S. *et al.* Necrotizing enterocolitis leads to disruption of tight junctions and increase in gut permeability in a mouse model. *BMC Pediatr.* **18**, 372 (2018).
285. Chhikara, N. *et al.* Human Epididymis Protein-4 (HE-4): A Novel Cross-Class Protease Inhibitor. *PLoS One* **7**, e47672 (2012).
286. Fiskerstrand, T. *et al.* Familial Diarrhea Syndrome Caused by an Activating

GUCY2C Mutation. *N. Engl. J. Med.* **366**, 1586–1595 (2012).
287. Romi, H. *et al.* Meconium ileus caused by mutations in GUCY2C, encoding the CFTR-activating guanylate cyclase 2C. *Am. J. Hum. Genet.* **90**, 893–899 (2012).
288. Alimperti, S. & Andreadis, S. T. CDH2 and CDH11 act as regulators of stem cell fate decisions. *Stem Cell Research* vol. 14 270–282 (2015).
289. Dusing, M. R., Maier, E. A., Aronow, B. J. & Wiginton, D. A. Onecut-2 knockout mice fail to thrive during early postnatal period and have altered patterns of gene expression in small intestine. *Physiol. Genomics* **42**, 115–125 (2010).
290. Talbot, J. *et al.* Feeding-dependent VIP neuron–ILC3 circuit regulates the intestinal barrier. *Nature* **579**, 575–580 (2020).
291. Lo, Y. H. *et al.* Transcriptional Regulation by ATOH1 and its Target SPDEF in the Intestine. *CMGH* **3**, 51–71 (2017).
292. Demitrack, E. S. & Samuelson, L. C. Notch regulation of gastrointestinal stem cells. *Journal of Physiology* vol. 594 4791–4803 (2016).
293. Buechler, M. B. *et al.* Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021).
294. Mishra, A. *et al.* Microbial exposure during early human development primes fetal immune cells. *Cell* **184**, 3394-3409.e20 (2021).
295. Biswas, S. *et al.* Microenvironmental control of stem cell fate in intestinal homeostasis and disease. *Journal of Pathology* vol. 237 135–145 (2015).
296. Brenna, Ø. *et al.* The guanylate cyclase-C signaling pathway is down-regulated in inflammatory bowel disease. *Scand. J. Gastroenterol.* **50**, 1241–1252 (2015).
297. Huang, B. *et al.* Mucosal Profiling of Pediatric-Onset Colitis and IBD Reveals Common Pathogenics and Therapeutic Pathways. *Cell* **179**, 1160-1176.e24 (2019).
298. Busslinger, G. A. *et al.* Human gastrointestinal epithelia of the esophagus, stomach, and duodenum resolved at single-cell resolution. *Cell Rep.* **34**, (2021).
299. O'Neal, W. K. & Knowles, M. R. Cystic fibrosis disease modifiers: Complex genetics defines the phenotypic diversity in a monogenic disease. *Annual Review of Genomics and Human Genetics* vol. 19 201–222 (2018).
300. Martin, J. C. *et al.* Single-Cell Analysis of Crohn’s Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493-1508.e20 (2019).
301. Nyström, E. E. L. *et al.* An intercrypt subpopulation of goblet cells is essential for colonic mucus barrier function. *Science (80-.).* **372**, (2021).
302. Elmentaite, R. *et al.* Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn's Disease. *Dev. Cell* **55**, 771-783.e5 (2020).
303. Holloway, E. M. *et al.* Mapping Development of the Human Intestinal Niche at Single-Cell Resolution. *Cell Stem Cell* **28**, 568-580.e4 (2021).
304. Nolan, D. J. *et al.* Molecular Signatures of Tissue-Specific Microvascular Endothelial Cell Heterogeneity in Organ Maintenance and Regeneration. *Dev. Cell* **26**, 204–219 (2013).
305. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764-779.e20 (2020).
306. Soldatov, R. *et al.* Spatiotemporal structure of cell fate decisions in murine neural crest. *Science* **364**, (2019).
307. Drokhlyansky, E. *et al.* The Human and Mouse Enteric Nervous System at Single-Cell Resolution. *Cell* **182**, 1606-1622.e23 (2020).
308. Cleynen, I. & Laukens, D. Cellular diversity in the colon: another brick in the wall. *Nature Reviews Gastroenterology and Hepatology* vol. 16 391–392 (2019).

309. La Manno, G. *et al.* Molecular architecture of the developing mouse brain. *Nature* **596**, 92–96 (2021).
310. Qiu, C. *et al.* Systematic reconstruction of the cellular trajectories of mammalian embryogenesis. *bioRxiv* 2021.06.08.447626 (2021) doi:10.1101/2021.06.08.447626.
311. James, K. R. *et al.* Distinct microbial and immune niches of the human colon. *Nat. Immunol.* **21**, 343–353 (2020).
312. Luoma, A. M. *et al.* Molecular Pathways of Colon Inflammation Induced by Cancer Immunotherapy. *Cell* **182**, 655-671.e22 (2020).
313. Mimitou, E. P. *et al.* Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* **10**, (2021).
314. Kraiczy, J. *et al.* Assessing DNA methylation in the developing human intestinal epithelium: Potential link to inflammatory bowel disease. *Mucosal Immunol.* **9**, 647–658 (2016).
315. 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).
316. Nam, A. S., Chaligne, R. & Landau, D. A. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nature Reviews Genetics* vol. 22 3–18 (2021).
317. Zhu, C., Preissl, S. & Ren, B. Single-cell multimodal omics: the power of many. *Nature Methods* vol. 17 11–14 (2020).
318. Schwank, G. *et al.* Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**, 653–658 (2013).
319. Yu, Q. *et al.* Charting human development using a multi-endodermal organ atlas and organoid models. *Cell* **184**, 3281-3298.e22 (2021).
320. Eicher, A. K. *et al.* Engineering functional human gastrointestinal organoid tissues using the three primary germ layers separately derived from pluripotent stem cells. *bioRxiv* 2021.07.15.452523 (2021) doi:10.1101/2021.07.15.452523.
321. Meran, L. *et al.* Engineering transplantable jejunal mucosal grafts using patient-derived organoids from children with intestinal failure. *Nat. Med.* **26**, 1593–1601 (2020).
322. He, Z. *et al.* Lineage recording reveals dynamics of cerebral organoid regionalization. *bioRxiv* 2020.06.19.162032 (2020) doi:10.1101/2020.06.19.162032.
323. Wang, J. *et al.* Nascent RNA sequencing analysis provides insights into enhancer-mediated gene regulation. *BMC Genomics* **19**, 1–18 (2018).
324. Biddy, B. A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219–224 (2018).
325. Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
326. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).
327. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
328. Karaiskos, N. *et al.* The Drosophila embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017).
329. Rendeiro, A. F. *et al.* The spatial landscape of lung pathology during COVID-19 progression. *Nature* **593**, 564–569 (2021).
330. Bertocchi, A. *et al.* Gut vascular barrier impairment leads to intestinal bacteria dissemination and colorectal cancer metastasis to liver. *Cancer Cell* **39**, 708-

724.e11 (2021).
331. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).
332. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science (80-.).* **348**, (2015).
333. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* **92**, 342–357 (2016).
334. Lee, J. H. *et al.* Highly multiplexed subcellular RNA sequencing in situ. *Science (80-.).* **343**, 1360–1363 (2014).
335. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
336. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).
337. Chen, A. *et al.* Large field of view-spatially resolved transcriptomics at nanoscale resolution Short title: DNA nanoball stereo-sequencing. *bioRxiv* 2021.01.17.427004 (2021) doi:10.1101/2021.01.17.427004.
338. Cho, C. S. *et al.* Microscopic examination of spatial transcriptome using Seq-Scope. *Cell* **184**, 3559-3572.e22 (2021).
339. Fawkner-Corbett, D. Colonic epithelial cell isolation for Single Cell RNA-sequencing. *protocols.io* (2019).
340. Sielaff, M. *et al.* Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J. Proteome Res.* **16**, 4060–4072 (2017).
341. Ye, D. Z. & Kaestner, K. H. Foxa1 and Foxa2 Control the Differentiation of Goblet and Enteroendocrine L- and D-Cells in Mice. *Gastroenterology* **137**, 2052–2062 (2009).
342. Tessari, M. A. *et al.* Transcriptional Activation of the Cyclin A Gene by the Architectural Transcription Factor HMGA2. *Mol. Cell. Biol.* **23**, 9104–9116 (2003).
343. Thiriot, A. *et al.* Differential DARC/ACKR1 expression distinguishes venular from non-venular endothelial cells in murine tissues. *BMC Biol.* **15**, (2017).
344. Buschmann, I. *et al.* Pulsatile shear and Gja5 modulate arterial identity and remodeling events during flow-driven arteriogenesis. *Development* **137**, 2187–2196 (2010).
345. Hu, C. *et al.* Retinol-binding protein 7 is an endothelium-specific PPARγ cofactor mediating an antioxidant response through adiponectin. *JCI insight* **2**, e91738 (2017).
346. Kyung, E. K., Sung, H. K. & Gou, Y. K. Lymphatic development in mouse small intestine. *Dev. Dyn.* **236**, 2020–2025 (2007).
347. Higuchi, M. *et al.* PRRX1- and PRRX2-positive mesenchymal stem/progenitor cells are involved in vasculogenesis during rat embryonic pituitary development. *Cell Tissue Res.* **361**, 557–565 (2015).
348. Yu, Q. C., Song, W., Wang, D. & Zeng, Y. A. Identification of blood vascular endothelial stem cells by the expression of protein C receptor. *Cell Res.* **26**, 1079–1098 (2016).
349. Niimi, K. *et al.* Endothelial specific deletion of FOXO1 alters pericyte coverage in the developing retina. *Biochem. Biophys. Res. Commun.* **520**, 304–310 (2019).
350. Lavery, D. L., Martin, J., Turnbull, Y. D. & Hoppler, S. Wnt6 signaling regulates heart muscle development during organogenesis. *Dev. Biol.* **323**, 177–188 (2008).
351. Gulbransen, B. D. & Sharkey, K. A. Novel functional roles for enteric glia in the

gastrointestinal tract. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 625–632 (2012).
352. Lei, J. & Howard, M. J. Targeted deletion of Hand2 in enteric neural precursor cells affects its functions in neurogenesis, neurotransmitter specification and gangliogenesis, causing functional aganglionosis. *Development* **138**, 4789–4800 (2011).
353. Dong, X.-L. *et al.* Fibrinogen-like protein 2 prothrombinase may contribute to the progression of inflammatory bowel disease by mediating immune coagulation. *Int. J. Clin. Exp. Pathol.* **11**, 1629–1636 (2018).
354. Monteiro, C. B. *et al.* Zinc finger transcription factor Casz1 expression is regulated by homeodomain transcription factor Prrxl1 in embryonic spinal dorsal horn late-born excitatory interneurons. *Eur. J. Neurosci.* **43**, 1449–1459 (2016).
355. Zhang, L. *et al.* Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell* **181**, 442-459.e29 (2020).
356. Marshall, A. S. J. *et al.* Identification and characterization of a novel human myeloid inhibitory C-type lectin-like receptor (MICL) that is predominantly expressed on granulocytes and monocytes. *J. Biol. Chem.* **279**, 14792–14802 (2004).
357. Bain, C. C. & Schridde, A. Origin, differentiation, and function of intestinal macrophages. *Front. Immunol.* **9**, (2018).
358. Zhang, Y., Du, W., Chen, Z. & Xiang, C. Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. *Exp. Cell Res.* **359**, 449–457 (2017).
359. Azizi, E. *et al.* Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* **174**, 1293-1308.e36 (2018).
360. Robbins, S. H. *et al.* Novel insights into the relationships between dendritic cell subsets in human and mouse revealed by genome-wide expression profiling. *Genome Biol.* **9**, (2008).
361. Villani, A. C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science (80-.).* **356**, (2017).
362. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell* **157**, 714–725 (2014).
363. Hay, S. B., Ferchen, K., Chetal, K., Grimes, H. L. & Salomonis, N. The Human Cell Atlas bone marrow single-cell interactive web portal. *Exp. Hematol.* **68**, 51–61 (2018).
364. O'Byrne, S. *et al.* Discovery of a CD10-negative B-progenitor in human fetal life identifies unique ontogeny-related developmental programs. *Blood* **134**, 1059–1071 (2019).
365. Sanders, K. M., Kito, Y., Hwang, S. J. & Ward, S. M. Regulation of gastrointestinal smooth muscle function by interstitial cells. *Physiology* vol. 31 316–326 (2016).
366. Lee, M. Y. *et al.* Transcriptome of interstitial cells of Cajal reveals unique and selective gene signatures. *PLoS One* **12**, (2017).
367. Du, K. L. *et al.* Myocardin Is a Critical Serum Response Factor Cofactor in the Transcriptional Program Regulating Smooth Muscle Cell Differentiation. *Mol. Cell. Biol.* **23**, 2425–2437 (2003).
368. Bolte, C. *et al.* Forkhead box F2 regulation of platelet-derived growth factor and myocardin/serum response factor signaling is essential for intestinal development. *J. Biol. Chem.* **290**, 7563–7575 (2015).
369. Kuemmerle, J. F. Occupation of αvβ3-integrin by endogenous ligands modulates IGF-I receptor activation and proliferation of human intestinal smooth muscle. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **290**, (2006).

370. Mifflin, R. C., Pinchuk, I. V, Saada, J. I. & Powell, D. W. Intestinal myofibroblasts: targets for stem cell therapy. *Am. J. Physiol. Gastrointest. Liver Physiol.* **300**, G684-96 (2011).
371. Bai, Y. *et al.* Interferon-gamma induces X-linked inhibitor of apoptosis-associated factor-1 and Noxa expression and potentiates human vascular smooth muscle cell apoptosis by STAT3 activation. *J. Biol. Chem.* **283**, 6832–6842 (2008).
372. Gurdziel, K., Vogt, K. R., Walton, K. D., Schneider, G. K. & Gumucio, D. L. Transcriptome of the inner circular smooth muscle of the developing mouse intestine: Evidence for regulation of visceral smooth muscle genes by the hedgehog target gene, cJun. *Dev. Dyn.* **245**, 614–626 (2016).
373. Zimmermann, E. M., Li, L., Hou, Y. T., Mohapatra, N. K. & Pucilowska, J. B. Insulin-like growth factor I and insulin-like growth factor binding protein 5 in Crohn's disease. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **280**, (2001).
374. Kuhn, R. M., Haussler, D. & James Kent, W. The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).
375. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
376. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
377. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).
378. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
379. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
380. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 1–9 (2019).
381. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 1–15 (2019).
382. Finak, G. *et al.* MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 1–13 (2015).
383. Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S. & Vert, J. P. A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat. Commun.* **9**, 1–17 (2018).
384. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).
385. Cabello-Aguilar, S. *et al.* SingleCellSignalR: Inference of intercellular networks from single-cell transcriptomics. *Nucleic Acids Res.* **48**, e55 (2021).
386. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 1–13 (2008).
387. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
388. Köhler, S. *et al.* Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* **47**, D1018–D1027 (2019).
389. Fawkner-Corbett, D. Unbiased molecular definition of epithelial barrier formation

and defects driving inflammatory bowel disease (DPhil Thesis D Fawkner-Corbett) - Mendeley Data. *Mendeley Data* https://data.mendeley.com/datasets/62z54f5yf3/1 (2021) doi:10.17632/62z54f5yf3.1.