Heterogeneity in human intestinal mononuclear phagocytes

Wulff, Line

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Heterogeneity in human intestinal mononuclear phagocytes

Line Wulff

PhD thesis

Technical University of Denmark, 2022

Preface

This PhD thesis has been submitted to the Health Technology department of the Technical University of Denmark (DTU) where the research was performed apart from a virtual exchange with Professor Eugene Butcher's group at Stanford University. The PhD project was initiated in April 2019 and has come to its' conclusion by the end of March 2022 all under the department of DTU Health Technology. The work presented here has been supervised by Professor William Agace, employed at DTU Health Technology and Lund University. Co-supervision was initiated by associate Professor Jose MG. Izarzugaza, but responsibilities were moved to Lars Rønn Olesen associate Professor at DTU Health Technology in January 2020.

The thesis consists of a general introduction to the different research topics and the common technology utilised in these studies, this is followed by three manuscripts, and a discussion with future perspectives of some important findings.

Line Wulff

Table of Contents

Abstract

The intestine is a unique environment which can broadly be separated into the small intestine (SI) and colon. These segments are responsible for specific functions, such as nutrient absorption in the SI and waste disposal, water reabsorption and hosting the commensal microbiota in the colon. The intestine has immune effector sites such as the lamina propria (LP), where immune cells maintain tolerance and fight infection, and immune inductive sites where adaptive immune responses are initiated. Residing in the gut wall, the gut associated lymphoid tissues (GALT) consist of isolated lymphoid follicles (ILF) and Peyer's patches, and are inductive sites for mucosal immunity.
Mononuclear phagocytes (MNP) are important immune cells that bridge the innate and adaptive immune system. They consist of two major subsets: macrophages and conventional dendritic cells (cDC). Macrophages are tissue resident cells and have important functions in clearing apoptotic cells, tissue remodelling and clearance of pathogens. They can arise from embryonic progenitors before birth, or from bone marrow derived monocytes. cDCs are derived from specific precursors in the bone marrow and patrol tissues to sample antigen; upon activation, they migrate to lymphoid tissue to prime naïve T cells.
The intestinal tissues also contain multiple cell types collectively termed mesenchymal stromal cells (MSC) which provide support for the epithelium lining the lumen and produce proteins for the LP's structural matrix.

The aim of this thesis was to unravel the heterogeneity of the human intestinal MNP compartment. We approached this by isolating GALT and LP separately from healthy intestinal tissue taken from individuals undergoing colorectal cancer surgery, using a novel protocol developed recently in the Agace lab. After sorting for MNPs, the cells were subjected to single cell RNA sequencing (scRNA-seq) to unbiasedly characterize all intestinal MNPs. scRNA-seq was also applied to murine MSC LP from SI and colon to characterize MSC diversity and ontogeny.

In the first manuscript, I show that there are three distinct subtypes of cDCs in the LP of both human SI and colon. These are cDC1, cDC2 and the more recently discovered cDC3, which has similarities to cDC2. To examine the expression of surface proteins, we used CITE-seq antibodies and found that cDC2 could be enriched as $CD1c^+$ $CD207^+$ $CD11a^-$, whereas cDC3 were $CD1c^+$ $CD207^-$ $CD11a^+$. These phenotypic markers allowed us to show that SI was relatively enriched in $CD207^+$ $CD11a^-$ cDC2, while the colon had a higher frequency of $CD207^-$ $CD11a^+$ cDC3. We also identified a putative common precursor for all the cDC in the LP, as well as downstream committed precursors for each cDC lineage which become increasingly similar to their mature counterparts.
Resident macrophages in the intestine are mainly derived from circulating monocytes. In the first manuscript, I was able to recreate this process by bioinformatic analysis and also showed that there are multiple distinct populations of mature macrophages. These may have different functions or they may have different locations within the LP. Furthermore, in inflammatory bowel disease the proportional distribution of monocytes and macrophages changed; the frequency of mature macrophages was reduced due to an influx of monocytes.

In the second manuscript, I compared the MNP populations in LP with GALT and observed that all major subsets of cDCs and macrophages were present in both locations. However,

GALT, especially Peyer's patches, had a higher frequency of cDCs than the surrounding LP, while the LP had a higher abundance of mature macrophages. By examining LP and GALT from a patient with Crohn's disease, we found substantial recruitment of monocytes in inflamed tissues and this was associated with the appearance of a pro-inflammatory monocyte subset that was only rarely observed in healthy tissues.

In the third manuscript, I contributed to the bioinformatic analysis of MSC subsets in the murine intestine and together, the results from my thesis illustrate how unbiased computational approaches can be used to unravel cellular heterogeneity in the intestine. Our results add to the growing knowledge on cDCs, macrophage and stromal populations in healthy intestinal tissues, providing insights into their ontogeny, functions and behaviour during inflammation that may help develop new therapies.

Dansk resumé

Tarmen er et unikt miljø der groft kan inddeles i tyndtarm og tyktarm. Disse dele varetager særlige funktioner såsom optagelse af næringsstoffer, bortskaffelse afføring og danne grobund for tarmfloraen. Tarmen har både immuneffektorvæv såsom lamina propia (LP), hvor immunceller opretholder tolerance og bekæmper infektioner, og immuninducerende områder hvor det adaptive immunrespons initieres. I selve tarmvæggen findes tarmassocierede lymfevæv (GALT) som består af isolerede lymfefollikler (ILF) og "Peyer's patches", som er inducerende områder for slimhindeimmunitet.
Mononukleære fagocytter (MNP) er vigtige immunceller der danner forbindelse mellem det medfødte og det adaptive immunsystem. De består af to overordnede grupper: makrofager og konventionelle dentritceller (cDC). Makrofager opholder sig permanent i væv og spiller en vigtig rolle i bortskaffelse af apoptotiske celler, modellering af væv samt fjernelse af patogener. De kan opstå fra embryoniske forløbere eller fra knoglemarvens monocytter. cDC'er udvikles fra specifikke knoglemarvs forløbere og overvåger væv for at opsamle antigener, ved aktivering kan de migerere til lymfevæv for at ingangsætte naïve T celler.
Tarmens væv indeholder også adskillige celle typer sammen kaldet mesenkymale stromale celler (MSC), disser celler støtter epitelcellelaget der fore lumen og producerer proteiner til den LPs strukturelle matrix.

Målet med denne afhandling var at afdække heterogeniteten af mennesketarmens MNP-undertyper. Vores tilgang var at adskille GALT og LP fra raskt væv taget fra individer som blev opereret for tyktarmskræft, til det brugte vi den nyudviklede protokol udviklet i Agace lab. Efter at have sorteret for MNP'er, gennemgik cellerne enkeltcelle-RNA-sekventering (scRNA-seq) for at karakterisere alle tarm-MNP'erne uden bias. scRNA-seq blev også udnyttet til MSC LP fra tynd- og tyktarm for at kunne karakterisere MSC diversitet og ontogeni.

I det første manuskript viser jeg, at der eksisterer tre forskellige undertyper af cDC'er i menneskets LP i både tynd- og tyktarm. Disse er cDC1, cDC2 og de nyligt observerede cDC3, som er fænotypisk lignende cDC2. For at undersøge ekspression af overflademarkører brugte vi CITE-seq anti-stoffer og opdagede, at cDC2 er beriget i $CD1c^+$ $CD207^+$ $CD11a^-$, hvorimod cDC3 er beriget i $CD1c^+$ $CD207^-$ $CD11a^+$. Disse fænotypiske markører tillod os at vise at tyndtarmen er relativt beriget for $CD1c^+$ $CD207^+$ $CD11a^-$ cDC2'er, hvor tyktarmen har en højere frekvens af $CD207^-CD11a^+$ cDC3'er. Vi fandt endvidere en formodet fælles forløber for alle cDC'erne i LP, og de downstream-engagerede forløbere som i stigende grad ligner deres færdigudviklede modparter.
Makrofager der befinder sig i tarmen stammer hovedsageligt fra cirkulerende monocytter. I det første manuskript har jeg genskabt denne proces transskriptomisk og vist at der er flere særskilte færdigudviklede makrofager. Disse kan muligvis have forskellige funktioner eller de kan have forskellige lokationer indenfor LP. Ydermere viste det sig at ved kroniske tarminflammatoriske sygdomme ændrede monocytter og makrophager frekvens, således at frekvensen af makrophager faldt grundet et influx af monocytter,

I det andet manuskript sammenlignede jeg MNP'er i LP med GALT og observerede at de hyppigst forekommende typer af cDC'er og makrofager var til stede i både LP og GALT. For GALT, særligt i "Peyer's patches", var forekomsten af cDC'er højere end i det omgivende LP, mens LP viste større forekomster af færdigudviklede makrofager. Ved at udforske LP og GALT

fra en Crohns patients væv observerede vi en væsentlig rekruttering af monocytter i betændte væv, disse var associerede med tilstedeværelsen af en pro-inflammatorisk monocyt type som kun sjældent er blevet observeret i raskt væv.

I det tredje manuskript har jeg bidraget med bioinformatisk analyse af MSC undergrupper i musetarm og tilsammen illustrerer resultaterne fra min afhandling hvordan datadreven tilgang uden bias kan bruges til at afdække heterogenitet i tarmvævene. Vores resultater bidrager til det voksende vidensgrundlag om cDC'er, makrofager og stromale cellepopulationer i raske tarmvæv sammen med deres ontogeni, funktioner og opførsel ved inflammation som måske kan fremme udvikling nye terapi muligheder.

Acknowledgements

What a rollercoaster! It has definitely not always easy to be the only bioinformatician in a lab of wet lab immunologists, but though the responsibility at times have felt heavy, I have also had the pleasure of working with and learning from a really great bunch of people. I got my master's in bioinformatics because it felt effortless and fun, but the last three years have added another layer. My passion was never in computer science but in human biology, and I got to do what I am good at while solving the mysteries of my very favourite biology topic. How lucky can one be? So, without further rambling, I would like to thank the people who have supported me and made this possible:

Firstly, I would like to give my thanks to **Bill**, for being willing to take on a bioinformatic master student and giving me a chance to do research within my dream field. Thank you for great inputs, limitless ideas, feedback and support whenever I have needed it. Next up is **Tom**, whom I would like to thank for a great collaboration and mentorship, through which I have learned so much both about the field of immunology but also about my own capabilities. To **Allan**, I am so grateful, that you could step in and help me with the last bits. And for all the great tutorials, which have taught me so much both on immunology and why it is my favourite biology field of all. To **Simone**, thank you for being in my corner, for cheering me on and always letting me know how much you appreciate the work I have done for our collaboration. It means the world to my confidence, to have you tell me out loud that you think I do great work. A big thanks to my co-supervisor **Lars**, I have appreciated the interaction with other bioinformaticians through your group, and the knowledge that I had support should I ever need it. A big thanks to the team of Agace lab members **Urs, Peter, Julien,** and **Venla** who has acquired data sets for me, letting me mainly maintain the sides I am actually good at. And especially thanks to **Urs** for helping me out with my thesis drafts and general GALT confusion. Thank you to **Fred** for helping me out when time got limited and letting me feel less responsible for everyone's data. To **Annie**, **Sara**, and **Isabel** thank you for sharing the rollercoaster that is a PhD journey, for a lot of moral support, when things have been difficult but also for sharing the best of times the PhD has had to offer. Thank you to my collaborators in Sweden, **Kristina**, **Julia,** and **Fredric** for interesting project collaborations and expanding my communicational skills of bioinformatics. An extra thanks to **Isabel, Annie,** and **Amalie**, who have used some of their own time to help me better my thesis drafts. And a great thank you to all the former colleagues in the Agace lab, **Telma**, **Nicole**, **Sophie**, and **Nikita**, who set the environment and welcomed me in. The same goes for **Rasmus, Imran, Vicky,** and **Elisa**, for have sharing labmeetings, office spaces and lunches, essentials to make DTU much preferable to the home office. Thank you to **Adam**, for being so very excited about single cell data and appreciating my skills. Thanks to **Marianne**, for answering my endless stream of questions on admin. And thanks to **Katha** and **Vaseileos** for asking difficult questions to improve my research and me as a scientist. Thank you to **Sine Reker** and her group for sharing and making up the environment of our department.

I am very grateful to our fundings: The **Lundbeck foundation** and the **Helmsley Charitable Trust**, without the grants my projects would have never happened.

And finally big thanks to my family and friends (especially: **mor, far, Lille Pus, Alexandra, Mette, Caroline,** and **Katrine**) for supporting me, helping me manage my stress and anxiety,

getting me to eat, read and do yoga to survive even the last bit of this journey. And finally to **Frede**, I couldn't possibly have done this without you by my side, more importantly, I wouldn't have wanted to.

Co-authored paper not included in thesis

(preprint) Julia Lienard, Kristina Munke, Line Wulff, Clément Da Silva, Julien Vandamme, Katie Laschanzky, Thorsten Joeris, William Agace, Fredric Carlsson. Intra-granuloma accumulation and inflammatory differentiation of neutrophils underlie mycobacterial ESX-1-mediated immunopathology . *BioRxiv.* (2022) doi: https://doi.org/10.1101/2022.01.19.477031

Abbreviations

APC – Antigen Presenting Cell
BM – Bone marrow
CD – Crohn's Disease
cDC – Conventional Dendritic Cell
cDC1 – Conventional Dendritic Cell Type 1
cDC2 – Conventional Dendritic Cell Type 2
cDC3 – Conventional Dendritic Cell Type 3
CLP – Common lymphoid progenitor
CDP – Common dendritic cell progenitor
CMP – Common myeloid progenitor
DAMP – Damage associated molecular pattern
DSB – Denoised and scaled by background
DSS – Dextran sodium sulphate
ECM – Extracellular matrix
ep-Mac – epithelium associated macrophages
FB – Fibroblast
FDC – Follicular dendritic cell
FL-monocyte – Fetal liver monocyte
GALT – Gut associated lymphoid tissue
GC – Germinal center
GMP – Granulocyte-monocyte progenitor
HLA – Human leukocyte antigen
HSPC – Hematopoietic stem/progenitor cell
IBD – Inflammatory Bowel Disease
ICC – Interstitial cells of Cajal
IEL – Intraepithelial lymphocyte
IHC – Immunohistochemistry
IL – Interleukin
ILC – Innate lymphoid cell
iMSC –Intestinal mesenchymal stromal cell
KNN – k-nearest neighbourhood
IFN – interferon
IL10R – IL10 receptor
Infl-cDC – Inflammatory cDC
LI – Large intestine
LP – Lamina Propria
LMPP – Lymphomyeloid-primed progenitor
Lyso-DC – Lysozyme expressing DC
Lyso-Mac – Lysozyme expressing macrophage
MDP – Macrophage dendritic cell progenitor
MHC – Major histocompatibility complex
MNP – Mononuclear phagocyte
MPP – Multipotent progenitors
mregDC – migratory mature immunoregulatory DC
Neu-Mac – Neuron-associated macrophage
PP – Peyer's patches
M-ILF – Mucosal Isolated Lymphoid Follicles
MLN – Mesenteric lymph node
PAMP – Pathogen associated molecular pattern
pIgR – Polymeric immunoglobulin receptor
pDC – Plasmacytoid Dendritic Cell
pre-cDC – pre-conventional dendritic cell
PRR – Pattern Recognition Receptors
RALDH – Retinoic acid aldehyde dehydrogenase
scRNA-seq – Single cell RNA sequencing
SED – Subepithelial dome
SI – Small intestine
SLE – Systemic lupus erythematosus
SMC – Smooth muscle cell
SM-ILF – Submucosal Isolated Lymphoid Follicles
TB-Mac – Tangible body macrophage
TCR – T cell receptor
TED – Transepithelial dendrite
TF – Transcription factor
Tfh – T follicular helper cells
TGFβ – Transforming growth factor β
Th1 – T helper cell type 1
Th2 – T helper cell type 2
Th17 – T helper cell type 17
TI – Trajectory inference
TLR – Toll like receptor
TNF – Tumour necrosis factor
Treg – Regulatory T helper cells
UC – Ulcerative Colitis
UMI – Unique Molecular Identifier

vas-Mac – Vasculature-associated macrophages

YS-macrophage – Yolk sack derived macrophage

Introduction

The intestine is the biggest surface of the human body in direct contact with the external environment. As a consequence, the intestine harbours a large reservoir of specialised immune cells[1]. These cells have to stay tolerant towards food antigens and commensal bacteria while simultaneously remaining alert towards infectious pathogens. When this balance is disturbed diseases such as food allergies, coeliac disease, colon cancer and inflammatory bowel disease (IBD) can develop[2].

1 Anatomy of the intestine

The human gastrointestinal tract extends from the oral cavity to the stomach, and further into the small (SI) and large intestine (LI). The SI starts after the stomach and comprises of three major segments which, in descending order, are the duodenum, jejunum and the ileum[1,3]. The LI is further segmented by cecum, ascending, transverse, descending and sigmoid colon and terminating with the rectum[1,3].

The SI is the major site of food digestion and nutrient uptake while the main function of the LI is to absorb water and salts. The colon is also home to the greatest number and diversity of microbes including bacteria, viruses and fungi, many of which have a beneficial relationship with the host. For example, commensal bacteria play a key role in degrading undigestible fiber, the biosynthesis of vitamin K, influencing epithelial metabolism and proliferation, and in the development and homeostasis of the intestinal immune system[1,3,4].

1.1 The layers of the intestine

The epithelium: Facing the lumen of the intestine is the mucosa, which comprises the intestinal epithelium and the intestinal lamina propria (LP). The intestinal epithelium consists of a single layer of columnar epithelial cells, each closely connected by tight junctions, that provides a physical barrier to the underlying tissue[5]. The majority of intestinal epithelial cells are absorptive enterocytes, that are essential for the digestion and uptake of nutrients and foodstuffs. In the SI the luminal side of enterocytes is covered in microvilli, small membrane protrusions that together form the brush border. Digestive enzymes and ion transporters are embedded in the brush border to support the final digestion and absorption of nutrients into the enterocytes[6]. Other specialized epithelial cells also exist within the intestinal epithelium including mucous producing goblet cells, hormone secreting enteroendocrine cells, tuft cells that are chemosensory and can secrete biological mediators, as well as Paneth cells of the SI, that sit in the base of epithelial crypts and are a rich source of antimicrobial peptides and growth factors for epithelial stem cells[7](***Figure 1***).

The lamina propria: The underlying LP consists of a complex network of non-hematopoietic cells including mesenchymal stromal cells (MSC), vascular and lymphatic endothelial cells, and glial cells[8,9]. Vascular endothelial cells located in the LP compose microvasculature which maintains the transport of oxygen and nutrients into the tissue and carries waste and metabolites out. Furthermore, the vasculature also plays a role in immune homeostasis and defence by recruiting immune cells from the bloodstream into the tissues mediated by the expression of chemokines and adhesion molecules to support leukocyte extravasation[10,11]. The lymphatics located the LP, similarly to other tissues, have crucial functions in immune surveillance by transporting immune cells to the draining lymph nodes. Additionally, the SI has blind-ending lymphatics vessels called lacteals which are important in the transport of

dietary lipids[12,13]. Subsets of MSCs, such as fibroblasts, provide proteins for the extracellular matrix (ECM), which function as a scaffold and framework for immune cells. In the last decade however, it has also been recognised that MSCs play essential roles in maintaining the epithelium [8,9]. Collectively these cells provide important structural support to the LP, as well as communication with the blood, lymph and enteric nervous systems (***Figure 1***).

Together the epithelium and LP form what is termed the mucosa. Both mucosa of the small intestine (SI) and colon have indentations away from the lumen known as crypts of Lieberkühn which is a source of stem cells developing into new epithelial and secretory cells through intermediates moving up towards the lumen[1,14]. In the SI the surface area is drastically increased by finger-like projections into the lumen called villi. These villi markedly increase the surface area of the SI supporting optimal absorption from the incoming foodstuff [1].

The submucosa and muscle: The submucosa is a dense layer of connective tissue separated from the mucosa by a thin muscle layer called the mucosa muscularis. Though small capillaries, as described above, extend into the LP, the main vessels, both blood and lymphatic, are situated in the submucosa [13]. Beneath the submucosa are two layers of muscles responsible for peristaltic activity moving along the content of the lumen. The outer layer of the intestine is a thin smooth connective tissue layer called the serosa [1].

2 The intestinal immune system

Given the intestines' continual exposure to antigen and metabolites from the microbiota and foodstuffs, it is perhaps unsurprising that the intestine contains the greatest number and diversity of immune cells and immune compartments in the body [1]. The intestinal immune system can be broadly divided into two main compartments: 1) Intestinal inductive sites, where cells of the adaptive immune system undergo initial priming and differentiation, and 2) effector sites where innate and previously active adaptive immune cells localize to promote barrier integrity, protective immunity, and tolerance against food antigens and commensal microbes [14].

2.1 Intestinal inductive sites

Intestinal inductive sites are sites where adaptive immune cells are primed towards intestinal derived antigens; these sites include the mesenteric lymph nodes (MLN) and the gut associated lymphoid tissues (GALT). Similar to other peripheral lymph nodes, MLNs receive antigens via tissue draining afferent lymphatic vessels. The different MLNs drain specific areas of the SI and LI, which have been suggested to be reflected in differing ratios of specific immune cells[15]. The afferent lymphatic vessels from the mucosa and the muscle layers of the intestine meet in bigger lymphatic vessels of the submucosa and are collectively transferred to the appropriate MLN. The afferent lymphatics are an important distinguishing factor of MLN and GALT, since GALT does not have afferent lymphatics but can sample antigens directly from the lumen[16,17]. GALT has access to lumenal antigens via the follicular associated epithelium (FAE) including specialised microfold cells (M-cells), which in mice have been shown to shuttle IgA covered antigens into the GALT [18].

GALT exists as multifollicular structures and as isolated lymphoid follicles (ILF). The best characterised multifollicular GALT is the Peyer's Patches (PP), which in humans can consist of tens to hundreds follicles[17]. The PPs are located throughout the SI but more commonly found in the ileum. They structurally consist of the FAE covered by only a thin layer of mucus, an underlying subepithelial dome (SED) and T cell zones surrounding the B cell follicles. High endothelial venules (HEV) are dispersed in the T cell zones which allow circulating lymphocytes entry into the PP, a process which in mice is governed by MAdCAM-1 on HEV and $\alpha_4\beta_7$ on lymphocytes. MAdCAM-1 is similarly expressed on human PP HEV proposing that a similar mechanism exists for lymphocyte migration into human PP (***Figure 1***)[19]. Each niche of the PP is characterised by different cells caretaking specific functions[17,20]. The SED has a high abundance of mononuclear phagocytes (MNP), which in murine models have been shown to be able to extend dendrites through the M-cells to directly sample lumenal content[21–23]. Follicles are B cell zones that can contain active germinal centers (GC), supporting B cell class switch and somatic hypermutation. Apart from naïve, memory and GC B cells, the follicles also harbour other cells such as follicular dendritic cells (FDC) which facilitate these processes [16,24,25]. The T cells found in PP have been shown to be enriched in naïve and central memory T cells and they include fewer cytokine producing effector T cells compared to surrounding LP[16]. PP are in mice considered crucial for the induction of adaptive SI immune responses by contributing to IgA plasma cell generation and migration. This is attributed by the induction of the SI homing molecule CCR9 on IgA+ plasma cells in the PP and SI MLN. As many of the structural features and immune populations are the same between human and murine PP, it is thought that PPs are also important for adaptive responses in the human setting[24–26].

ILFs in humans can be generally divided into two categories depending on their location in the intestinal wall; mucosal ILFs (M-ILF) directly in mucosa and submucosal ILFs (SM-ILF) mainly located in the submucosa but extending into the overlying mucosa. M-ILFs seem to be largely restricted to the ileum and distal colon (descending and sigmoid colon), while SM-ILFs are found throughout the colonic tissues and in the cecum[16]. SM-ILFs are on average larger than the M-ILFs, and though both have more T cells than the murine ILFs, the lymphocyte compartment of SM-ILFs contain higher proportions of naïve T cells[16,27]. The M-ILFs have higher B to T cell ratios including higher proportions of GC B cells[16]. Both structures have an overlying FAE containing M-like cells and an underlying SED containing CD11c+ MNPs, but the FAE and SED region is larger in the M-ILF[28–30]. Furthermore, the M-ILF M-cells have been shown to express GP-2, a marker which in mice have been associated with the uptake of fimbriated bacteria[16,31]. Both ILF types are associated with MAdCAM-1+ HEVs and a lymphatic network which could support functions like those seen for PP[16,19,24]. With an estimated number of 30.000 ILFs, the ILFs could collectively contribute significantly to the induction of adaptive immune responses in the human intestine[32,33]. This is supported by the presence of all the necessary components for T-cell-dependent B cell responses in form of T follicular helper cells (Tfh), GC, CD40L expression and expression of AID (enzyme necessary for somatic hypermutation) in GC B cells[16,34–36]. And further by the finding of IgA+ clones in the ILFs being overrepresented in the surrounding colonic LP compared with SI LP, suggesting local B cell induction by ILFs[16].

However, how the individual GALT compartments contribute to the intestinal immune responses and whether they sample and initiate immune responses towards specific

intestinal antigens is unknown. Furthermore, the observation of increased size of SM-ILFs in Crohn's disease (CD) poses questions on the role of ILF in the initiation and maintenance of IBD[29]. The recent protocol development to isolate both GALT and GALT-free LP in Agace laboratory[29] has enabled us to start to address these important questions.

2.2 Intestinal effector sites

The main effector tissues of the SI and LI are the LP and epithelium. These tissues harbour a wide range of innate and adaptive immune cells like eosinophils, mast cells, different types of T and B cells, innate lymphoid cells (ILC) and mononuclear phagocytes (MNP) [1,14].
Intraepithelial lymphocytes (IEL) sit along the basement membrane of the epithelium and can be found at levels ranging 10-15 IELs/100 enterocytes in the duodenum and have decreasing rates along the length of intestine. Within the IEL compartment, different subsets of T cells can be distinguished based on their expression of the $\alpha\beta$TCR co-expressed with either the CD8$\alpha\beta$ or CD4 or by the CD8$\alpha\alpha$ co-expressed with either $\gamma\delta$TCR or $\alpha\beta$TCR. All of these subsets are represented in the SI but along the length of the intestine the CD8$\alpha\alpha^+$ becomes increasingly rare[1,3]. The IELs are important in protective immunity where they limit damage by infectious pathogens and preserve the integrity of the epithelial barrier [1,37].

In the LP lymphocytes exist in both the form of T and B cells. The T cell compartment primarily consists if antigen-experienced $CD4^+$ helper T cells and $CD8^+$ cytotoxic T cells which can be stimulated to produce cytokines [16]. Within the $CD4^+$ T cell compartment, different effector types in form of T helper 17 (Th17) cells and FoxP3+ regulatory T helper cells ($FoxP3^+$ Treg) have an inverse proportional relationship going along the length of the intestine, the former with highest abundance in the SI and the latter in the LI [1,3]. The $FoxP3^+$ Tregs are recognized to be a major source of the anti-inflammatory interleukin (IL)10 in the intestinal tissues and play an important role in promoting tolerance [3,38]. The B cells found in the LP are mainly IgA^+ plasma cells, which can readily secrete antibodies [16]. The secreted IgA can be transported into the lumen via specific polymeric immunoglobulin receptor (pIgR) expressed on the basolateral side of enterocytes, where it can neutralize bacteria to mediate immune exclusion and help antigen clearance [39]. ILCs, which are the innate counterpart of lymphocytes, exist in the LP in different subsets resembling effector T cell subsets. However, these are reported to express effector molecules even at steady-state conditions and contribute to organ development and antimicrobial peptide defense[1,40,41]. Granulocytes can be identified in the healthy mucosa mainly in the form of mast cells and eosinophils, whereas an influx of neutrophils and basophils can be observed during intestinal inflammation [42–44]. Finally, the LP also harbours a range of resident and migratory mononuclear phagocytes (MNP), the overall function and their distinguishment in the intestine will be discussed in detail in the following sections.

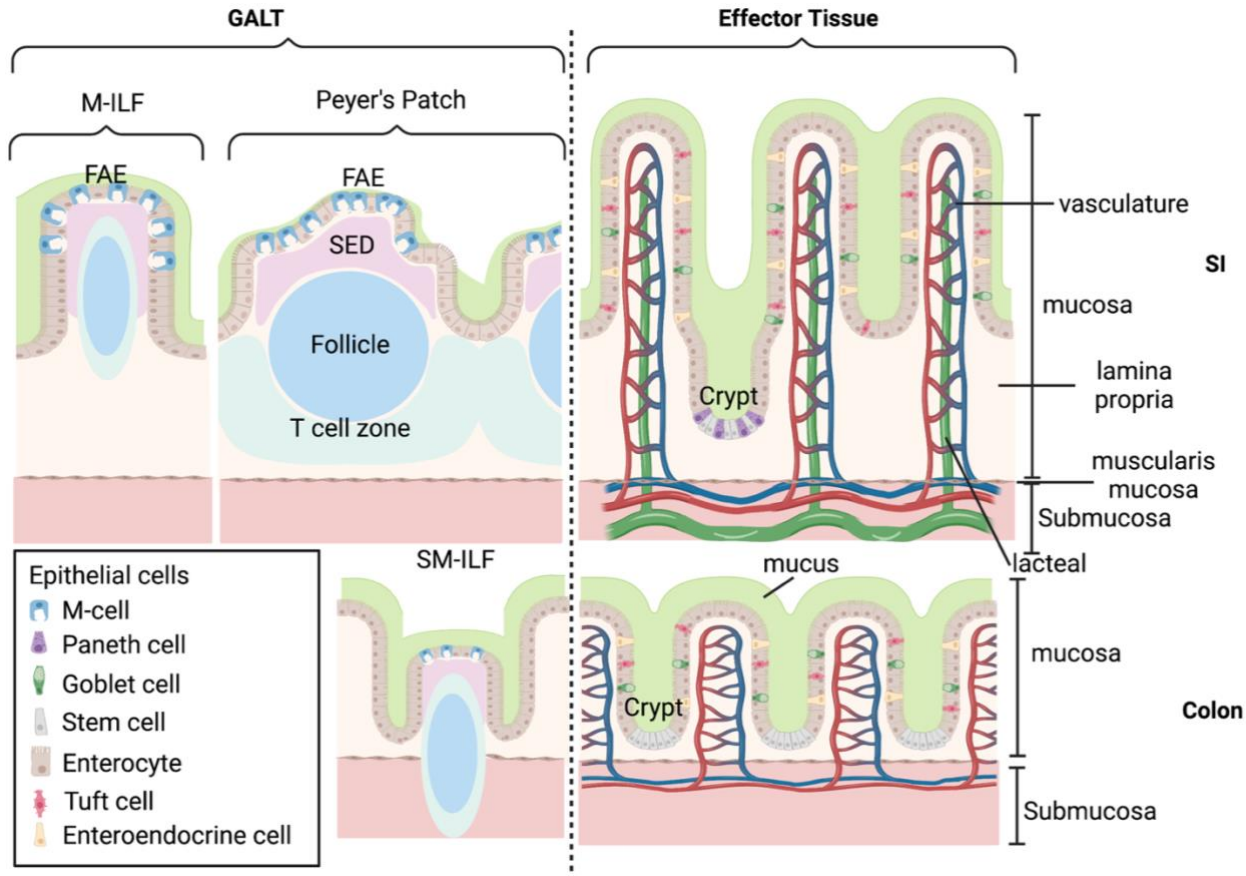


Figure 1 – Structural schematics of intestinal effector and lymphoid tissues: *Schematics of the inductive sites directly located in the mucosa and submucosa and the effector tissue. In human SI there are two main GALT; Mucosal isolated lymphoid follicles (M-ILF) and Peyer's patches. M-ILFs mainly consist of B cells and a small surrounding T cell zone. The Peyer's patch is a large multi-follicular GALT containing follicles connected via T cell rich interfollicular regions. Subepithelial dome (SED) underlies the FAE in GALT. In the colon, the main GALT are submucosal ILFs (SM-ILF), which as name suggest mainly occupy the submucosal layer but also indents into the mucosa. The structure of the effector tissues also varies between SI and colon. In the SI the surface areas is optimised by villi structures protruding out into the luminal space. In between the villi structure are crypts with many specialised cell types, which maintain the development of epithelial cells. In the colon effector tissues has no villi but still has crypts. The colon has a thicker mucus layer compared to the SI. Figure created with BioRender.*

3 Mononuclear Phagocytes

Mononuclear phagocytes (MNP) is the term given to a large family of innate immune cells, that includes monocytes, macrophages and conventional dendritic cells (cDC). They have classically been grouped together because both macrophages and cDC-like cells can be differentiated from monocytes *in vivo* [45]. As the name suggests MNPs are characterised by a single nucleus, however, the shape ranges from bilobed kidney shape in monocytes to a similar elongated kidney shape or rounded in dendritic cells and macrophages, respectively [46,47]. They are further distinguished morphologically by stellate shaped cDC and big vacuolar macrophages [45]. Collectively MNPs have a wide range of functions in the immune system including phagocytosis of foreign microbes, uptake of apoptotic epithelial cells, regulation of innate and adaptive immune cell responses, tissue repair and antigen presentation. As such MNP are essential for host homeostasis, defense against invading pathogens and, if dysregulated, in promoting immune-mediated pathology.

MNP recognize microbes and damaged cells through germline-encoded pattern recognition receptors (PRR) that detect conserved molecular motifs on pathogens (pathogen-associated

molecular pattern - PAMPs) or damaged cells (damage-associated molecular patterns - DAMPs). A major family of PRR involved in the detection of conserved microbial motifs are the toll like receptors (TLR) expressed on different sites of the cell to allow for recognition of different types of molecular patterns[48]. The TLRs cover a range of motifs expressed on extracellular bacteria like lipopolysaccharide and flagellin recognised by TLR4 and TLR5, respectively. TLRs can also detect motifs from intracellular infections such as viral genomic material recognised as double-stranded and single-stranded RNA by TLR3 and TLR7/8 and unmethylated CpG islands on DNA, a hallmark of bacterial and viral DNA, by TLR9. TLR4 and 5 in these examples are expressed on cell surfaces of MNPs, while TLR3, 7, 8 and 9 are all situated intracellularly on endosomal membranes[48,49]. Signalling through recognition of the TLRs recruit adaptor proteins such as Myd88, which can lead to downstream effects like NFkB activation and subsequent production of inflammatory cytokines or the production of type I interferons[49]. In addition to direct recognition of PAMPs, MNPs also express receptors recognising apoptotic cells and opsonic receptors like Fc receptors and complement receptors. The latter receptors enable the recognition and uptake of particles coated in either antibodies or complement proteins [50]. The apoptotic receptors are important as cells in healthy tissues are constantly dying by apoptosis. The apoptotic cells express molecules signalling their status in form of lysophosphatidylcholine and phosphatidylserine. Receptors involved in recognition of these and thus in recognition of apoptotic cells include TIM-1, TIM-4, and CD14[50–52].

The above-mentioned examples of detection highlight how MNPs can recognize a variety of patterns by germline-encoded receptors which in turn can lead to internal signalling events. For MNPs a downstream event of such signalling can be the internalization process referred to as phagocytosis. The MNPs are as their name suggest highly efficient phagocytes and their major functions revolve around this capacity. Phagocytotic receptors such as the apoptotic receptors and the opsonic receptors can directly activate the internalization process. In contrast, TLR signalling does not directly activate phagocytosis but can lead to priming of the MNP, which in turn can upregulate phagocytotic integrins leading to phagocytosis [53]. After detection, the internalization process is initialized by rearrangement of the cytoskeleton leading to encapsulation in a specialized vacuole called a phagosome. The phagosome undergoes maturation and fuses with a lysosome giving rise to the phagolysosome where the internalized pathogen or cell is degraded. Degraded peptides are then presented on the cell surface by major histocompatibility complexes (MHC)[50,54].

Two types of MHC exist: MHCI and MHCII, in humans they are referred to as human leukocytes antigen (HLA). The MHC:peptide complexes can interact with appropriate T cell receptors (TCR) and their co-receptor CD8 or CD4, expressed mutually exclusive on $CD4^+$ and $CD8^+$ T cells[54]. The two MHC complexes share a similar 3D structure, but are composed of different subunits, which leads to differences in the peptide binding groove and importantly in the Ig-like domain closest to the cell membrane. The difference between the Ig-like domains decide whether the MHC can bind the co-receptor CD4 or CD8, thus, MHCI only interacts with $CD8^+$ T cells and MHCII with $CD4^+$ T cells[54,55]. MHCI is expressed by all nucleated cells where it presents endogenous antigens including auto-antigens. These antigens presented on MHCI are surveyed by $CD8^+$ T cells, which function as a monitoring system to determine whether cells are healthy or have become infected or cancerous. However, MHCI can also present exogenous antigens by a process known as cross-presentation. Cross-

presented antigens on MHCI are used by MNPs to prime the cytotoxic T cells in adaptive immune responses against viral and other intracellular infections [56,57]. MHCII is displayed on specific cell subsets including MNPs, B cells and thymic epithelial cells (TEC). In B cells MHCII regulates activation, proliferation and differentiation by interaction with Tfh cells, whereas TEC expresses MHCII for the initial selection process of non-self-reactive T cells which occurs in the thymus during early development of T cell repertoires[58,59]. In MNPs MHCII expression is used to prime and activate naïve T cells, this will be described in further detail in a later section.

All MNPs can differentiate from bone marrow (BM) derived hematopoietic stem/progenitor cells (HSPC) with the exception of certain tissue resident macrophages which are derived from embryonic precursors, which will be discussed further below. In the classic model of hematopoiesis, HSPCs differentiate into multipotent progenitors (MPP) and downstream progenitors that become more and more committed to specific hematopoietic lineages (**Figure 2**). MPP give rise to progenitors for the two major lineages of the immune system: common lymphoid progenitors (CLP) and common myeloid progenitors (CMP). In this model the CLPs give rise to all lymphocytes including B cells, T cells, and other innate lymphoid cells [60–63], while CMP gives rise to the granulocyte-monocyte progenitor (GMP). GMP in turn generate a granulocyte progenitor committed to the granulocyte lineages (eosinophils, basophils, neutrophils, and mast cells) or macrophage dendritic cell progenitors (MDP) committed to the MNP lineage. Monocytes and monocyte derived macrophages derive from a common monocyte progenitor while DCs share the common DC progenitor (CDP) and the pre-conventional dendritic cell (pre-cDC) (***Figure 2***). However, the advent of single cell sequencing technologies has highlighted considerable heterogeneity within many of these progenitors and this classical model of hematopoiesis is currently under re-evaluation[62–67]. This was clearly represented in a study of single cell RNA sequencing (scRNA-seq) data on sorted BM precursors from different subsets of the schematics outlined in ***Figure 2***, demonstrating considerable transcriptional heterogeneity within these populations. Furthermore, the individually sorted populations included cells that were transcriptionally similar to downstream populations suggesting commitment at earlier stages than previously assumed [68].

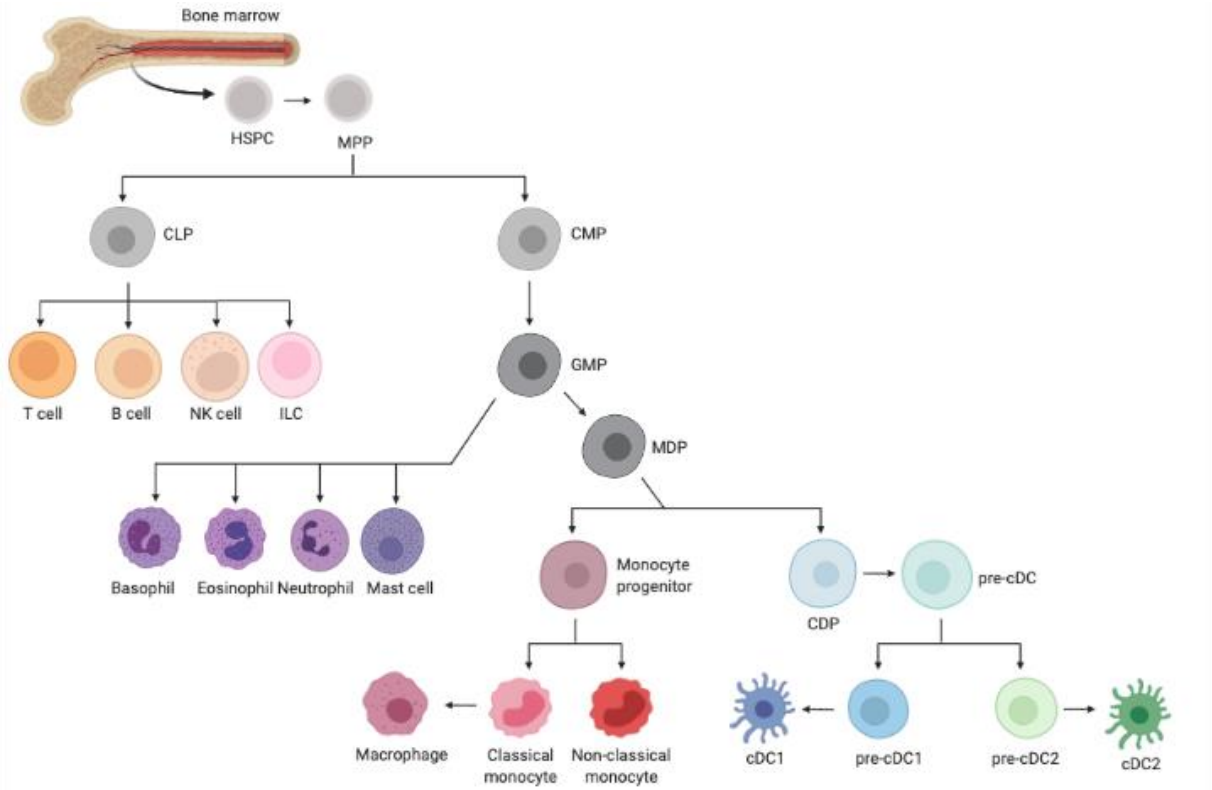


Figure 2 – Classic hierarchical schematics of hematopoiesis of immune cells focused on MNPs: Hematopoietic stem/progenitor cells (HSPC) from the bone marrow gives rise multipotent progenitors (MPP). These in turn gives rise to the common lymphoid progenitor (CLP) and all the lymphocytes and other innate lymphoid cells. The common myeloid progenitor (CMP) differentiate into the granulocyte monocyte progenitor (GMP) which gives rise to all granulocytes and to the macrophage dendritic cell progenitor (MDP). The MDP gives rise to all MNPs by either the monocyte progenitor or the common dendritic cell progenitor (CDP). Figure created with BioRender. Adapted from [60–63].

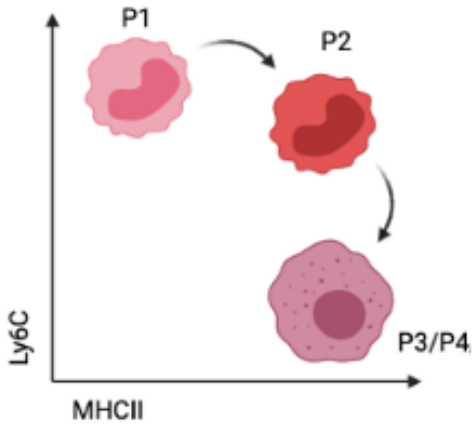
3.1 Monocytes and macrophages

3.1.1 Ontogeny

Studies in mice have demonstrated that tissue resident macrophages arise from three distinct precursors during embryogenesis and in adult life. The first wave consists of yolk sack derived macrophages (YS-macrophage) during embryogenesis; these cells do not develop through a monocyte intermediate like later macrophages[69]. Cells from the YS also seed the fetal liver, where they proceed to give rise to fetal hematopoiesis including fetal monocytes (FL-monocytes) which in turn forms the second wave of macrophages during fetal development[70,71]. The FL-monocytes develop into macrophages through a monocyte intermediate and, similar to YS-precursors, seed tissues before birth. Both of these events precedes seeding of BM and the eventual generation of HSPC used in adult hematopoiesis, supporting macrophages of different origin than BM derived monocytes[72]. A common feature of YS and FL derived macrophages is the proliferative capacity enabling them to self-renew and be maintained within tissues for long periods of time even in adult life[73]. Indeed, macrophages in the adult mouse brain (microglia) are thought primarily to originate from YS-derived precursors[69,72,74,75]. Similarly, adult tissue-resident alveolar macrophages are thought to mainly derive from FL-monocytes backed by an experiment showing that when transferring 1:1:1 ratios of YS-macrophages, FL-monocytes and BM derived monocytes into the neonatal lung, 80% of macrophages were derived from the FL-monocytes two weeks later[73]. After birth, many tissues like the skin, heart, pancreas and intestine maintain only small populations

of FL-monocyte derived and YS-macrophages. Maintenance of the macrophage niche within these compartments instead requires continual input from BM derived monocytes[72]. BM derived monocytes differentiate into tissue macrophages through a series of intermediates known as the "monocyte waterfall"[76–78], where cells progressively lose several phenotypic markers associated with monocytes and gain those associated with tissue resident macrophages (Figure 3)[79]. In mice, upon entry into tissues, Ly6C^{hi}MHCII^{neg}-monocytes (P1) gain expression of MHCII (P2). These cells then downregulate Ly6C and start to express intermediate levels of CX3CR1 (P3). Finally, these cells differentiate into CX3CR1^{hi} (P4) phenotypically resembling tissue resident macrophages. Further markers utilised for their characterisation include CD64, which is upregulated to lower levels on P2 and fully upregulated at P3[76,80] (***Figure 3***). The monocyte waterfall has also been suggested to occur in humans albeit using different sets of phenotypic markers including CD14, HLA-DR, CD11b and CD11c (***Figure 3***) [81,82]. Though the above observations have been made in intestinal tissues a similar process is proposed to contribute to specific macrophage subsets in multiple tissues such as skin, heart, peritoneal cavity, and interstitial lung[78,83–87]. As might be expected, the phenotypic changes, associated with the monocyte waterfall during transition towards tissue resident macrophages, are also accompanied by changes in cellular morphology and cellular transcription. P1 in both mice and humans are transcriptionally and morphologically closely related to peripheral classical blood monocytes, with smaller size and kidney shaped nuclei (***Figure 3***). Later stages are bigger, have more cytoplasm, distinguished nuclei and are auto fluorescent [88,89]. Gene expression in later stages of monocyte-derived macrophages are as with mature macrophages dependent on local environmental stimuli [78,89].

Though mature tissue resident macrophages established before birth can maintain by self-renewal, circumstances can arise where BM-derived monocytes are recruited into tissues to replenish macrophage populations. This is observed in Kupffer cells (liver macrophages), which are self-maintaining in homeostasis, but during infection with listeria monocytogenes or in targeted depletion, BM derived monocytes enter the liver and can replace the mature macrophages[90–92]. Similarly, can be observed in a herpes lung infection model and experimental autoimmune encephalitis causing loss of the alveolar macrophages and the microglia (brain macrophages), respectively[93,94]. Collectively these findings highlight that the loss of self-renewing tissue resident macrophages in multiple settings like depletion, inflammation and infection can be overcome by the recruitment of BM derived monocytes.



mouse	P1	P2	P3	P4
CD64	neg	low	pos	pos
Ly6C	high	high/pos	neg	neg
MHCII	low	pos	high	high
CX3CR1	int	int	int	high
CD11b	high	high	high	high
human				
CD14	pos/high	pos	pos	high
HLA-DR	int	high	high	high
CD11b	high	pos	neg	pos
CD11c	pos	pos	neg	neg

Figure 3 - Monocyte Waterfall: Circulating classical monocytes enter the tissues from the blood stream as Ly6C^{hi} MHCII^{neg} (P1), at the next stage MHCII is upregulated (P2) but monocyte still has similar morphology to that of blood monocytes. Ly6C is downregulated (P3) and morphological changes start being visible. Figure created with BioRender. Table indicates differential protein expression in P1-P4 in murine and human counterparts, as shown in [76,80,88].

3.1.2 Functions

Macrophages are a diverse population of tissue-resident MNPs, found in both lymphoid and non-lymphoid tissues throughout the body[95–97]. Some of the main functions of tissue resident macrophages are clearing pathogens and apoptotic cells by phagocytosis and maintaining tissue integrity by contributing to remodelling and repair[98]. Macrophage tissue remodelling is for instance essential in development by removing apoptotic cells during the formation of digits[99]. In wound repair, macrophages are important for tissue remodelling directly by producing matrix metalloproteinases and their inhibitors which adjust ECM turnover, and indirectly by secretion of transforming growth factor β (TGFβ) promoting fibroblast differentiation[100]. During homeostasis bone is constantly remodelled, here osteoclasts, a bone resident macrophage, play an important role in resorption of bone by releasing hydrochloride acid to dissolve mineral and metalloproteinases to disintegrate the collagenous matrix[101]. Together, these examples highlight how macrophages perform tissue remodelling in many settings through both phagocytic, enzymatic, chemical and cell mediating effects. Macrophages also function as tissue sentinels which can initiate inflammatory responses in case of infection. Inflammatory responses can be initiated in macrophages through the NFκB-pathway downstream of PAMP or DAMP detection leading to the production of pro-inflammatory cytokines such as IL1β and tumour necrosis factor (TNF)[102,103]. Tissue inflammation has in turn been shown to recruit classical monocytes, which can further aggravate inflammation or help clear off infection by production of reactive oxygen species and further production of pro-inflammatory cytokines[104]. Most of these incoming monocytes will disappear again after inflammation is resolved and will not contribute to the resident mature macrophage populations[91,92,105].

Though many tissue resident macrophages share common features and functions the last decade has revealed that macrophages are highly diverse and tailored by local signals to the needs of the niches in which they reside[89,106–108]. Kupffer cells do not only phagocytose pathogens and toxins from the bloodstream but are also important in the metabolism of

micronutrients, lipids and the recycling of iron[109]. The niche supporting Kupffer cell specific functions have been associated with signals deriving from the microenvironment, specifically from endothelial and stellate cells. When Kupffer cells die they activate endothelial and stellate cells in their environment by releasing TNF. This causes a temporary upregulation of receptors for cell adhesion and migration of blood monocytes followed by an imprinting of blood monocytes to develop into new Kupffer cells[91]. Together, this provides an example of how tissue resident macrophages can interact with local tissue cells, which in in turn provide imprint on incoming monocytes.
Alveolar macrophages are important for clearance and recycling of surfactant in the lung and also in clearing debris and inhaled particles. This unique capacity is lost and subsequent leads to pulmonary alveolar proteinosis (defective clearing of surfactant), when alveolar macrophages are depleted and replaced by tissue-resident macrophages from colon, liver or peritoneal cavity[73]. Underlining that macrophages in distinct niches adopt niche specific functions but also that specialization causes the loss of plasticity in tissue resident macrophages. Microenvironments such as the germinal centers in lymphoid tissues also support niche specific macrophage populations. In germinal centers a macrophage subset known as “tangible body” macrophages (TB-Mac) are recognised microscopically due to their distinct morphology with “tangible bodies” representing states of degraded apoptotic cells. These macrophages uptake apoptotic B cells to remove sources of auto-antigens supported by the finding that in patients with systemic lupus erythematosus (SLE) germinal centers have fewer TB-Mac and increased amounts of apoptotic cells[110,111]. However, in many tissues and microenvironments like the germinal centers we do not yet fully understand the exact cues and interactions driving functional specialization[95,106].

Circulating monocytes, in both mice and humans, consist of two major subsets; classical and non-classical monocytes[112]. Classical monocytes, that in humans are $CD14^{+}$ $CD16^{-}$ and in mice $Ly6C^{hi}CX3CR1^{int}CCR2^{+}$ cells, serve under homeostasis as the immediate precursors of tissue resident macrophages and are short-lived with a lifespan of 1 day[112,113]. Classical monocytes are dependent on their CCR2 expression for egression from the bone marrow and the receptor is also used to migrate into peripheral tissues[114]. Non-classical monocytes, defined in humans as $CD14^{-}CD16^{+}$ cells and mice as $Ly6C^{low}CX3CR1^{hi}$ $CCR2^{-}$ cells, are also called patrolling monocytes due to the characteristic manner of “crawling” along the vascular endothelium, which enables them to survey the vascular endothelium for abnormalities[112]. They have a relative long lifespan of 2 days in mice and 7 days in human compared to the classical monocytes[115,116]. In humans, the non-classical monocytes have proposed to be derived from the classical monocytes through a $CD14^{+}CD16^{+}$ intermediate monocyte[115]. In mice, this was supported by adoptive transfer of the monocyte progenitor giving rise to first classical monocytes and delayed rise of non-classical monocytes[117]. The above observations has given rise to the suggestion that non-classical monocytes are circulation analogues of tissue resident macrophages[64].

3.2 Dendritic cells

Dendritic cells (DC) are the second major branch of MNPs. These have classically been divided into two subsets: plasmacytoid DC (pDC) and conventional DC (cDC), which have considerable differences in both function, morphology and ontogeny [118]. The recent observations on pDCs have led to question whether these are actual DCs and the population will only be briefly covered in this thesis.

3.2.1 cDC function

cDCs patrol tissues where they sample self and foreign antigen. At this stage cDCs are considered immature and do not express co-stimulatory molecules CD80, CD86 and CD40, they have high phagocytic capacity but are slower in migrating compared to their mature counterparts[119]. Capture of antigens by cDC can also be performed by receptor mediated endocytosis or macropinocytosis, both are also lowered upon cDC activation[120]. Another feature of immature cDC is the low expression of MHCII:peptide complexes on the cell surface, which are mainly loaded with auto-antigens[121]. When an immature cDC encounters danger signals in form of PAMP or DAMPs they can initiate phagocytosis or endocytosis of the particle and activation of the cDC. Specifically, TLR signalling induces maturation of the phagolysosome, which leads to increased acidification and proteolytic activity which leads to the degradation of engulfed particles into peptide fragments[122]. MHCII molecules are not very stable without a bound peptide fragment, therefore they are stabilized after synthesis in the endoplasmatic reticulum by a molecule called the invariant chain. After transferal to endosomal or phagosome compartments, the invariant chain is slowly degraded until the last fragment, called CLIP, is exchanged for an appropriate peptide[121]. Under homeostatic conditions many MHCII:peptide will be sorted and targeted for direct ubiquination and subsequent degradation in the lysosome. However, during cDC activation ubiquination of newly synthesized MHCII:peptide complexes is reduced leading to an increased expression of MHCII:peptide associated to the danger signal on the cell surfaces of activated cDCs[123]. cDC activation is further accompanied by expression of CCR7 triggering cDC migration to lymphoid tissues. This is accomplished by expression of the CCR7 ligand, CCL21, expression on afferent lymphatics in peripheral tissues as well as the T cell zone of lymphoid tissues[124].

After arrival in the T cell zone the cDCs can mediate activation and differentiation of naïve T cells. For this to occur the cDC must provide three distinct signals: Firstly, cDCs and T cells interact by the T cell receptor (TCR) on the T cell and the MHC:peptide complex on the cDC. This a specific interaction and requires that the TCR is compatible with the MHC:peptide complex, e.g. the TCR is specific for the particular antigen presented by the cDC[55]. Secondly, in order for the T cells to become activated there has to be a co-stimulatory signal otherwise the T cells will become anergic. The co-stimulatory signal is provided by an interaction between CD28 on the T cell and CD80/CD86 on the cDC. Other ligand-receptor pairs can also provide both co-stimulatory or inhibitory signals to enhance or inhibit T cell responses [125]. The last signal for T cell activation determines what type of effector function the activated T cell adopts. This is accomplished by soluble signals from the environment that the T cell receives during activation, including for example cDC derived cytokines (***Figure 4***). There are five well-described sets of $CD4^+$ T helper cell polarizing factors. Briefly, these include IL12, IL27 and IFNγ inducing expression of transcription factor (TF) T-bet, which gives rise to T helper cell 1 (Th1) effective in intracellular infections and involved in auto-immune diseases. IL12 inhibition and IL4 induces TF GATA-3 and subsequent T helper cell 2 (Th2) differentiation effective against parasitic infections and involved in allergies. IL6 and IL23 induce TF RORγT which gives rise to Th17 cells effective against extracellular bacteria and also involved in auto-immunity. IL2, TGFβ and RA can induce TF FoxP3 expression and subsequent differentiation of Treg cells. Finally, IL6 causes induction of TF Bcl-6 followed by differentiation of Tfh cells[126].

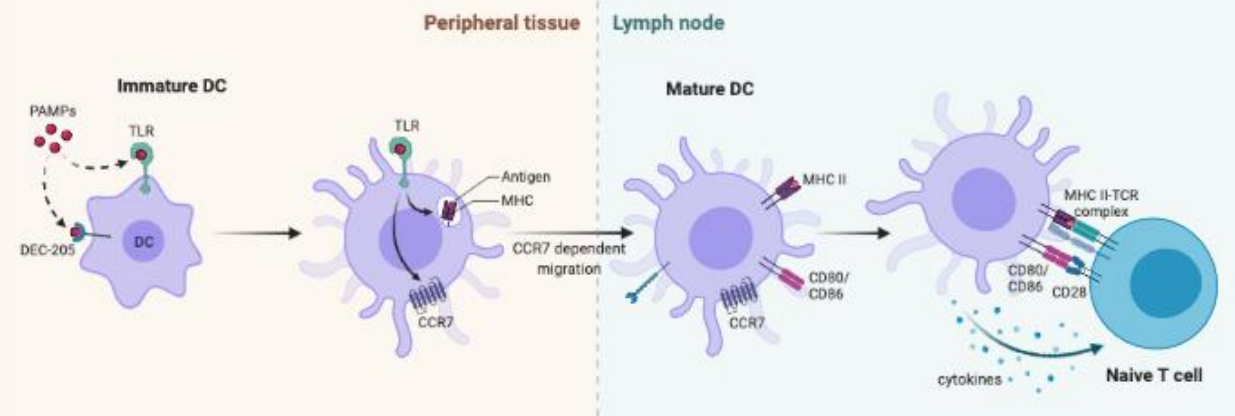


Figure 4 - cDC priming of naïve T cells: Immature cDCs in peripheral tissues express different PRRs to recognize different molecular associated patterns. After receptor mediated endocytosis or phagocytosis and triggered by PRR stimulation the cDC mature and start upregulating CCR7, which causes the cell to migrate to lymphoid tissues. In the T cell zone of lymphoid tissues cDCs can interact with naïve T cells with appropriate TCRs for the antigen captured in the peripheral tissue. In order to prime the cell there must be interactions between TCR and MHCII-Antigen, co-stimulatory signal by CD80/CD86 interacting with CD28 and cytokine secretion for directing the appropriate T helper cell lineage. Figure created with BioRender.

3.2.2 cDC subsets

Though cDCs share capacity of the overall function described above, they consist of several transcriptionally and functionally distinct subsets.

cDC1: The cDC type 1 (cDC1) are well characterised in both human and mouse sharing expression of XCR1 and DNGR-1/CLEC9A across species, common human cDC1 markers also include CADM1 and CD141, whereas murine lymphoid cDC1 are also characterised by CD8α[124]. cDC1s develop from the pre-cDC in circulation or from the BM in a manner dependent on the TFs IRF8 and BATF3 by a cDC1 committed intermediate called pre-cDC1[124]. In mice, BATF3 deletion causes the loss or reduction of cDC1[127], and similarly, when BATF3 is silenced in *in vitro* cultures of developing human cDC, there is reduced fractions of cDC1 in the total cDC population[128], indicating an essential role of BATF3 for cDC1 development across species. Similarly, in mice deletion of IRF8 in cDCs lead to significant reduction of cDC1[129]. cDC1 are especially effective in driving $CD4^+$ Th1 and $CD8^+$ cytotoxic T cells responses[130–132], both of which are important in immune responses towards intracellular infections and tumours. The $CD4^+$ Th1 priming is accomplished via the MHCII:peptide, and is proposed to be superior in cDC1 due to their high production of IL12[133]. The induction $CD8^+$ T cell responses are initiated by peptides cross-presented on MHCI which is a highly specific function of cDC1, which are uniquely adapted in subcellular machinery and vesicle trafficking for cross-presentation[127]. cDC1s have for instance been shown to express higher levels of several molecules related to MHCI pathways and cross-presentation compared to other cDCs[134]. The importance of cross-presentation by cDC1s is underlined in a study of Batf3 deficient mice with a subsequent loss of cDC1s, consequently, the mice could not mount virus specific $CD8^+$ T cell responses or reject highly immunogenic tumours[135]. cDC1 also have implications in induction of Treg responses, which will be discussed later in the gut specific context.

cDC2: cDC type (2) have more differences between human and mice than cDC1, however across species they consistently express SIRPα and CLEC4A. human cDC2 are further distinct from cDC1 by expressing CLEC10A and CD1c, whereas the murine cDC2 express CD11b and in spleen CD4[124]. cDC2s, like cDC1s, develop from the CDP via the pre-cDC intermediate. The pre-cDC development into pre-cDC2 is proposed to be maintained by different TFs depending

on tissues and subtype. For instance IRF4 deletion in cDCs only affect CD11b^{+} cDC2 frequencies in some tissues like the SI and lung, but not in the spleen[136]. Other TF with possible involvement in cDC2 development include Notch2 and KLF4[137,138]. The discrepancies between these pinpoints increased heterogeneity within the cDC2 subset compared to the cDC1. Functionally, cDC2 excel in activating CD4^{+} Th2 and Th17 responses, which are effective in responding to respectively parasitic and extracellular bacterial infections. Production of thymic stromal lymphopoietin, type I IFN, CCL17 and CCL22 by cDC2 have been shown in correlation to promoted Th2 differentiation, however, none of these are essential in Th2 differentiation and also there is no evidence of cDC2 secreting IL4[139,140]. On the other hand, several Th17 polarizing factors have been attributed to cDC2, for instance deletion of TGFβ-activating αvβ8 on cDC2s significantly reduces Th17 responses in the LP, proposing a prominent role on Th17 induction[141].

cDC3: Recently studies of unsupervised scRNA-seq data have suggested the presence of new type of cDCs in both human and murine tissues. The emerging subset has been reported with different sets of markers, in different conditions and tissues, and it is thus difficult to assess whether all the reported subsets represent the same cell type. However, a commonality of the newly described subsets is a transcriptomic profile with similarities to both cDC2s and monocytes (expressing genes such as *CD1C, FCER1A, S100A9*, *S100A8*, *FCN1*, *VCAN*, *CD14* and *CD163*). This emerging subset of cDCs has been labelled cDC3, DC3, or cDC2B[142–146]. Bourdely *et al.* separated cDC3 from monocytes with CD88 and FCERI, where monocytes were CD88^{pos} and cDC3 were FCER1^{pos}. cDC3 in this study diverted from cDC2 on the transcriptional level in a direct comparison by expressing genes like *S100A8, S100A9, S100A12,* and *CD14,* in this comparison cDC2 had DEGs such as *LTB, CD1E,* and *CD1A*[143]. A shared feature for the putative cDC3 subset described by various groups is surface expression of both CD1c and low CD14[143,144,146]. In blood, BM, skin and spleen of humans Cytlak *et al.* reported a subset differing from cDC2 by being CD5^{neg} and CD163^{pos}, these were described to differentiate from a separate precursor from CDP with IRF8^{low} expression and the cDC3 were not lost in patients with genetic loss of IRF8 affecting both cDC2 and cDC1[146]. In a different study, potential precursors were sorted for *in vitro* differentiation and here cDC3 were found to be derived mainly from CLEC12A^{pos} GMPs[146]. Both of the above studies suggest that cDC3 differentiate from a separate precursor from other cDCs, splitting off just after MDP stage. Dutertre *et al.* identified a similar subset and showed that these expand in patients in vivo with FLT3L injections, which has long been acknowledged to regulate cDC expansion[147], confirming cDC3 identity as a cDC subset[144]. Though all of the above suggest existence of a cDC3 subset separate from cDC2 by transcriptome, phenotype and ontogeny, the function of cDC3 is not well understood. However, cDC3s have been shown to increase in frequency in SLE[144] and have been proposed to contribute to Th1 and Th17 polarisation[142,144,148]. Upon TLR stimulation cDC3 can mature in a similar fashion to cDC2s expressing genetic markers of migration, co-stimulation, T cell attracting chemokines and secretion of T cell polarizing cytokines all consistent with a role in adaptive immune induction[143].

3.2.3 Plasmacytoid DC function, phenotype and ontogeny

The main function of pDC is to produce type I interferon (IFN), which they do in response to intracellular TLR7 and TLR9 stimulation by single-stranded RNA and DNA with unmethylated CpG islands; both common PAMPs for viral infections [149,150]. Type I IFN is important in antiviral defense through NK activation, by interfering with the viral replication cycle, and by

promoting cDC mediated Th1 responses [151,152]. Furthermore, cells like cDCs, stroma, monocytes and gut stem cells depend on low continuous type I IFN to function normally in homeostasis [153]. Classically, CD123 and CD45RA have been used to identify pDC in human blood. However, in 2017 See *et. al* found that putative pre-cDC are also $CD123^{pos}$ $CD45RA^{pos}$, leading to the discovery that many classic cDC functions also associated with pDC are actually contributed by pre-cDC in this mixed populations[154]. Later followed the discovery that pDCs mainly develop through CLP and a common progenitor with B-cells and only a small part of pDCs are derived from the CDP [155]. These observations together propose that pDCs, if based on ontogeny and function, might not actually be a dendritic cell.

3.3 Intestinal monocytes and macrophages:

Intestinal monocytes and macrophages are found throughout the intestinal effector and inductive sites. Here, distinct macrophage populations have been described in the submucosa, muscularis and serosa and in different niches of the GALT[82,156].

3.3.1 Macrophages in intestinal LP

Most LP macrophages in adults derive from circulating monocytes that transition to mature macrophages within the intestine following the monocyte waterfall[76,78,88]. In the first stages of differentiation, macrophage intermediates in the intestine are able to produce pro-inflammatory cytokines in response to TLR stimulation. However, during their differentiation to mature macrophages, these cells are exposed to local environmental signals that drive them to a hyporesponsive 'tolerogenic' phenotype. As a result intestinal LP macrophages become hyporesponsive to TLR stimulation and IFNγ[76,81], which would normally lead to production of pro-inflammatory cytokines, nitric oxide and reactive oxygen species[79,157–159]. However, this hyporesponsiveness is not thought to be due to downregulated PRR expression, but rather changes of downstream signalling components underpinned by abrogated inflammation in Myd88 deletion in a specific colitis models[160]. Mature intestinal macrophages also upregulate a range of receptors for detection and engulfment of opsonized material and apoptotic cells such as CD163 (heme scavenger receptor), CD206 (mannose receptor), CD36 (scavenger receptor for dietary long-chained lipids and cholesterol), the tyrosine kinase receptors AXL and MERTK and complement C1Q components[79], and metalloproteases utilized in tissue remodelling[78]. Though specific functions and hyporesponsiveness is not observed in the early stages of the monocyte waterfall, it has been shown that monocytes in mice already acquire a differential program 24 hours after recruitment into the mucosa[76]. One mediator of the intestinal specific differentiation is TGFβ, which has been shown to locally impact macrophage development and function. When TGFβ signaling is absent the monocyte waterfall is disrupted with increases in P1 and P2 and decreases in mature macrophages. Furthermore, transcriptional differences, including *IL-10* and *CXC3CR1,* found to be upregulated through the colon specific monocyte waterfall are not upregulated in the mature macrophages in the absence of TGFβ signalling[78]. These results indicate that TGFβ is important for imprinting specific intestinal differentiation of macrophages.

Another key cytokine thought to drive intestinal macrophage function is IL-10. Mechanistically, IL-10 signals through the IL-10R on maturing macrophages rendering these cells hyporesponsive[161]. Such signalling is key to intestinal homeostasis as mice with a selective deletion of IL-10R in macrophages develop spontaneous colitis[162]. The importance

of IL-10 in intestinal homeostasis in humans is further underscored by the findings that both IL-10 and IL-10R deficiency has been associated with early onset IBD[163]. Intestinal macrophages can themselves produce IL-10, however, they have not been found to be the main source of secreted IL-10 in the gut, and IL-10 deletion in macrophages does not cause colitis in mice[162].

Studies have highlighted considerable functional and ontogenic heterogeneity in intestinal macrophages[82,164]. Indeed, a picture is emerging in which functionally specialized subsets of macrophages inhabit different microenvironments within the intestinal LP, submucosa and muscle layer, where they provide key niche specific support for local cells. These include, (1) epithelium associated macrophages (ep-Mac), that support enterocyte growth and epithelial barrier integrity by constitutive low secretion of TNFα, and in the colon signal to enterocytes to stop absorption in the presence of fungal toxins[165], (2) vasculature-associated macrophages (vas-Mac), where two subsets have been described, a subset expressing $CD169^{+}$ located towards crypt regions away from epithelium[166] and a perivascular subset associated to the microvasculature in the villi. The latter forms a network in the villi which is suggested to function as a anatomical barrier that can prevent bacterial entry[167]. (3) Neuron-associated macrophages (neu-Mac) that on transcriptomic level show many similar expression patterns to the microglia of the brain, they have been reported in both the submucosa and the muscularis, where they maintain neuronal survival[164]. The neuronal maintenance have been proposed to be performed via specific interactions between neurons and neu-Mac for instance; complement cascade genes from the neu-Mac are involved in synaptic pruning, and neural function such as peristalsis is regulated by neu-Mac BMP2 expression[82,156,168].

Interestingly, in mice several of these populations including neu-Mac and vas-Mac contain higher proportions of embryonically derived macrophages that are self-renewing in adult steady state, whereas the ep-Mac are seeded embryonically but quickly after birth become replaced by monocyte derived macrophages which maintain this population through life[106]. In effector sites TIM-4 has been suggested to be a marker of self-renewing macrophages, however, the expression was not correlated to localization of the local niche macrophages, leaving unclarity of whether neu-Mac and vas-Mac exclusively express TIM-4[169]. Though it is difficult to assess the ontogeny of human intestinal macrophages, muscularis mucosa macrophages are found to be highly proliferative, which could indicate self-renewing capacity[82]. Finally, in mice irrespective of macrophage subset, all populations are decreased in the absence of a microbiota, indicating a role of microbiome interactions in regulation of the intestinal macrophages no matter their ontogenetic origin[169].

3.3.2 Myeloid populations in GALT

The PP contain two main classes of macrophages based on location: One that is located in the SED directly underlying the FAE, and the other in the deeper interfollicular and follicular regions of the lymphoid tissues. All the known macrophages of the PP are characterised by lysozyme expression[21], a protein known to have bactericidal action by disrupting the cell wall of gram positive bacteria through hydrolysis[170]. SED macrophages are referred to as $TIM4^{-}$ lysozyme expressing macrophages (lyso-Mac), where lysozyme distinguishes them from LP macrophages in surrounding villi and $TIM4^{-}$ makes them distinct from macrophages in interfollicular and follicular regions of the PP[21]. Their innate function is also supported by expression of host antiviral restriction factor BST2[171]. The $TIM4^{-}$ lyso-Macs have further been

reported to express *IL22RA2* (coding for IL-22-BP), affecting the FAE by IL-22 antagonism, however expression of IL-22 BP in SED macrophages is not yet confirmed at the protein level[21]. Inhibition of IL-22 signalling in the FAE is needed for FAE associated characteristics such as lowered mucin and antimicrobial peptide secretion and changes in surface glycosylation[21,172]. Indicating that producers of IL-22-BP could support FAE characteristics. In comparison to other MNPs in the PP, TIM4$^-$ lyso-Mac express high levels of genes involved in innate responses like TLRs, sequestering of metal, inflammasome formation and inhibition of replication, all consistent with an important role in antiviral and bacterial reponses[171]. The above propose that TIM4$^-$ lyso-Mac positioned directly in the SED could be important in both innate functions and by affecting the FAE characteristics by possible IL-22-BP expression.

The second general macrophage subset of the PP can be found in the interfollicular regions and in germinal centers where they phagocytose apoptotic T and B cells, respectively. The germinal center macrophages are TB-macs like those found in other secondary lymphoid tissues. Across species, both TB-macs and macrophages in the interfollicular regions express TIM-4, MERTK and lysozyme. The interfollicular region macrophages are referred to as TIM4$^+$ lyso-Macs and like the SED TIM4$^-$ lyso-Mac they express BST2, which is not expressed by TB-macs. Apart from BST2 and location, it is not well understood how the TIM-4$^+$ lyso-Mac differ from PP TB-Macs[22,82,171]. The commonly expressed TIM-4 in lymphoid macrophages is reported to be a receptor for phosphatidylserine, which as mentioned is a marker of apoptotic cells, and has been shown to cooperate with MERTK for binding and subsequent phagocytosis of apoptotic cells[173,174]. The expression of these factors by TB-Macs and lyso-Macs aligns well with their major functions in PP: clearance of apoptotic adaptive cells. All of these subsets can also be identified in the human PP with a few differences; TB-Macs in human PP express TIM4, but they do not express lysozyme and BST2. Lyso-Macs can like their murine counterparts be distinguished with TIM4$^-$ in SED and TIM4$^+$ in interfollicular region. However, in humans, interfollicular TIM4$^+$ lyso-Mac express CD163, which in mice is used to distinguish contaminating LP macs from true PP macrophages[22,175].

Importantly, the PP SED also harbours lysozyme expressing DCs (lyso-DC), which like the lyso-Mac are derived from monocytes albeit with a faster turnover[21]. Lyso-DC express BST2 and lysozyme like the lyso-Mac, but are distinguished by their expression of CD11c and high levels of MHCII. PP lyso-DC been shown to acquire antigens from the lumenal space by extending dendrites through pores in the M cells. Similar to lyso-Mac, lyso-DC have been proposed to be more involved in innate responses compared to other PP DCs as they have a similar genetic signature to lyso-Mac[171,176]. Lyso-DC also express the highest level of *IL22RA2* reported in the PP, and are thus thought to support FAE characteristics[21]. Lyso-DC are not purely associated with innate immune functions: In mice, lyso-DC and pDC in PP has been shown to exclusively express TLR7, and upon *in vivo* stimulation lyso-DC can phagocytise microspheres, upregulate CCR7 and travel to interfollicular regions to prime naïve T cells[22,177]. Lyso-DC have also been identified in human PP and here can be distinguished from lyso-Mac through their expression of SIRPα, CD24 and JAM-A, and from cDC2 by their lack of expression of CD1c[22].

While the above examples serve to highlight that myeloid cells adapt unique functions to support cells and responses within their specific microenvironments, the environmental factors driving these functional specificities remain largely unclear.

3.4 Intestinal conventional dendritic cells

3.4.1 cDCs in the intestinal LP

In mice, three FLT3L dependent cDC subsets have been identified in the intestinal LP that can be distinguished from one another by the markers CD103 and CD11b. $CD103^+$ $CD11b^-$ cDC represent cDC1, while $CD103^+$ $CD11b^+$ and $CD103^-$ $CD11b^+$ represent cDC2. cDC1 and $CD103^+$ and CD103- cDC2 have also been identified in the human intestinal LP[178,179]. The three observed intestinal subsets have been shown to be bona fide cDC by all expanding to FLT3L, have similar rapid turnover in steady state (unlike slow turnover in macrophages) and differentiate from pre-cDC not monocytes[180]. Interestingly, the frequency of these three cDC populations differ between the SI and colon; $CD103^+$ $CD11b^-$ cDC1 is the most common cDC subset in the colon followed by the $CD103^-$ $CD1b^+$ cDC2, whereas the $CD103^+$ $CD11b^+$ cDC2 are most abundant in the SI[15]. Similar trends in the cDC2 subsets have been observed in the human intestine[179]. Whether cDC3 are present in the human and mouse intestine remains to be resolved.

Intestinal LP cDC scan the LP for self and foreign derived antigen prior to their migration to intestinal draining MLN. LP cDC been shown to acquire luminal antigen via several routes. Firstly, cDC have been proposed to extend transepithelial dendrites (TEDs) by opening the tight junctions between epithelial cells without disrupting the barrier integrity, and into the lumen to directly sample antigen[181,182]. Secondly, ep-Mac are high expressors of CX3CR1, and antigen captured in lumen by ep-MAC TEDs can be transferred to cDCs. This process is called trogocytosis and allows transfer of soluble antigens by creating gap junctions between cDC and macrophages[183]. Thirdly, the mucin producing goblet cells can function as passages for low molecular weight lumenal antigens and deliver these to underlying cDCs [184]. Other mechanisms by which cDCs acquire intestinal antigen include sampling of the LP, epithelium and interstitial fluid via receptor mediated phagocytosis or endocytosis or via micropinocytosis.

Intestinal cDC1 functions: Given their high turnover, apoptotic epithelial cells are a major source of self-antigen within the intestine. Intestinal cDC1 represent the major cDC population that cross present epithelial derived antigen to $CD8^+$ T cells in MLN[185]. Interestingly, cDC1 mediated cross presentation of epithelial antigen was recently shown to drive the generation of $FoxP3^+$ $CD8^+$ T cells that were required for cross tolerance. This function was shown to be be regulated by cDC1 derived by PD-L1, TGFβ and retinoic acid (RA) where RA specifically induced expression of the SI homing CCR9 on the $FoxP3^+$ $CD8^+$ T cells and TGFβ induced expression of CD103[186]. Intestinal cDC1 have also been shown to play an essential role in the development of the intestinal IEL compartment and in intestinal Th1 responses [187], although the underlying mechanism remain to be fully elucidated. Thus, cDC1 play key non redundant roles in intestinal immune homeostasis.

Intestinal cDC2 functions: Similar to cDC1, intestinal cDC2 have been shown to play key non-redundant roles in intestinal adaptive immune homeostasis, most notably in the generation of intestinal Th17 and Th2 responses. The results vary based on different model systems, and the associated functions are thus not always as clear cut as has been observed in cDC1. This also suggests heterogeneity within the intestinal cDC2 compartment and the need for unbiased cDC characterisation in humans and mice alike. Both $CD103^+$ and $CD103^-$ cDC2 have been implicated in the induction of intestinal Th17responses[180,188–190]. For instance, Persson

et al. found that mice whose CD11c$^+$ cells lacked *Irf4* had a marked reduction of CD103$^+$ cDC2 and reduced numbers of Th17 cells in the SI, colon and MLN[191]. Contradictively, Scott and colleagues identified two separate subsets within the CD103$^-$ cDC, of these the CD103$^-$ CD11b$^+$ CCR2$^+$ subset was more effective in driving differentiation of Th17 *in vitro*, and this was masked when the CD103$^-$ CD11b$^+$ CCR2$^{+/-}$ were not separately gated[180]. Furthermore, in a cDC CD11c-specific deletion of the receptor Notch2 CD103$^+$ CD11b$^+$ cDC2 were lost and correlated with a loss of intestinal IL17 secreting CD4$^+$ T cells[138], suggesting an *in vivo* role of CD103$^+$ cDC2s in Th17 induction. Interestingly, the importance of cDC2 subsets to drive intestinal Th2 responses appears to differ between the SI and colon, with CD103$^+$ cDC2 driving these responses in the SI and CD103$^-$ cDC2 driving these responses in the colon[192].

3.4.2 cDCs in GALT

In PP cDC are mainly located in the SED and in the T cell rich interfollicular regions. The cDC compartment differs from the lyso-DC by their lack of expression of lysozyme and BST2. Like in other lymphoid tissues, murine cDC1 are mainly characterised by CD8α and XCR1 expression, with the human counterpart expressing XCR1. cDC1 are mainly located in the interfollicular regions of the PP[177,193]. PP cDC2, both of which express SIRPα, are found in two maturation states. The immature cDC2 do not express CD11b or CCR7, but express CCR6, whose ligand CCL20 is expressed by the FAE, and these cells are predominantly located within the SED where they can sample incoming antigen. The mature cDC2 are CCR6$^-$ and have upregulated CD11b and CCR7, the latter enabling them to migrate to the interfollicular regions where they can prime naïve T cells[21,177]. Half of the mature CD11b$^+$ cDC2 also express *IL22RA2*, though at lower levels to what is found in lyso-DC, which as earlier discussed could be indicative of an FAE support function by this subset[21,22], a puzzling observation considering their location is mainly in the interfollicular regions. While lyso-DC can extend dendrites through the M cell pores, cDC subsets are mainly dependent on M cell shuttling of lumenal antigens into the SED. Here, cDCs can engulf the particles before migrating to the interfollicular regions to prime naïve T cells[18].

A major and common property SI cDC1 and cDC2 is their enhanced ability to induce the gut homing receptors CCR9 and $\alpha_4\beta_7$ on responding T cells subsequent to their migration to MLN or GALT. This is accomplished because SI cDC express high levels of retinaldehyde dehydrogenase2 (*RALDH2*), compared to cDC in other tissues, which allows these cells to generate the Vitamin A metabolite, retinoic acid (RA) [188,194]. Indeed RA is sufficient to induce gut homing receptors on anti-CD3 activated T cells *in vitro*[195,196]. A cDC role for IgA switch was first described to be associated with the LTβ signalling dependent cDCs, mainly CD11b$^+$ cDC2 and CD103$^-$ CD11b$^-$ cDC, where deletion of *Ltbr* led to a subsequent decrease in IgA titers. This interaction was mainly thought to occur in the SED and was dependent on cDC $\alpha_v\beta_8$ expression catalyzing the activation of TGFβ[197]. As mentioned, the SED is rich in immature cDC2 and lyso-DC, however it was not clarified whether they expressed differential capacity for IgA switch induction. After the study on cDC B cell interaction in the SED, both cDC1 and cDC2 have implicated in the generation of intestinal IgA responses where their roles appear highly context dependent. For example, intestinal CD103$^+$ cDC2 appear essential for IgA responses to soluble flagellin[198], while cDC1 are important for the initial induction of IgA responses to rotavirus[199]. Interestingly, optimal generation of IgA also appears to require RA[200], and short chain fatty acid (SCFA) produced by commensal bacteria were recently implicated in enhancing RALDH activity in intestinal cDC[201]. Further studies proposed that the

SCFA acetate was responsible for inducing B cell IgA switching and that this was mediated by RALDH activity in cDC[202].

MNP populations in murine and human ILFs are less well characterised, though a population of $CD11^+$ MNPs have been observed many years ago in murine context[203]. Recently, murine ILFs have been shown to contain a major population of $CD103^-$ $CD11b^+$ cDC2 that express high levels of IL-22-BP and lysozyme[204]. These were distinct from lyso-DC from the PP in expression of ZBTB46, a major transcription factor of cDC, which is not expressed by the monocyte derived lyso-DC[22,204]. In these ILF associated cDCs IL-22-BP was suggested to be an important regulator of lipid absorption by enterocytes by blocking IL3 derived IL-22. Mice deficient in IL-22-BP had higher levels of bioavailable IL-22 and subsequent reduced absorption of lipids from the lumen due to lowered expression of lipid transporters on enterocytes[204]. While $CD11c^+$ MNPs are also present in the SED of humans M-ILF and SM-ILF[16], the broader MNP composition of human ILF remains unclear.

The above examples illustrate important roles for the intestinal cDC in both immunogenic and tolerogenic responses. They further demonstrate that environmental factors present within the intestine including vitamin A and SCFA regulate cDC function.

4 Intestinal Mesenchymal Stromal Cells

Intestinal mesenchymal stromal cells (iMSC) represent a non-hematopoietic tissue resident population of cells present throughout intestinal inductive and effector sites. iMSC are an essential source of extra cellular matrix (ECM) proteins that provide key structure to the mucosa[205]. Apart from these functions iMSCs play key pathogenic roles in different diseases like inflammatory disorders and cancer, where they are crucial in the tumour environment supporting cell proliferation[8,206,207]. Traditionally, iMSCs have been defined by the lack of the major markers of other cell types including CD45 (immune cells), EpCAM (epithelial cells), CD31 (endothelial cells), Ter119 (erythrocytes), L1CAM (neurons and glial cells). At the beginning of this thesis little was known regarding LP iMSC diversity and ontogeny. However, with the emergence of single cell technologies our understanding of iMSC has grown substantially and revealed several distinct MSC populations in the LP[206–209].

4.1 Pericytes

Pericytes are closely associated with the vasculature sitting in close proximity to endothelial cells, which form the actual blood vessels. They have been shown to regulate the diameter of the vasculature to regulate blood flow and provide growth factors supporting the blood vessels[8,9,206]. Indeed, loss of pericytes is associated with leaky vasculature and hyperdilation[210–212]. Pericytes have been shown to express high amounts of PDGFRβ, RGS-5, filaments associated with contraction like Desmin, NG2, NOTCH3, ESAM1, CD146, and some additionally express α-SMA (gene: *Acta2*). However, several of these proteins are also expressed by smooth muscle cells (SMC)[210].

4.2 SMCs

SMC is a major cell type of the different muscle layers in the intestinal tissues. Furthermore, they envelop the larger vessels in the submucosa of both SI and colon, and additionally sit around lacteals in the SI villi[210,213]. The main marker used for identifying SMCs is α-SMA ,

however as mentioned aSMA is also expressed by some pericytes and by a FB subset sometimes referred to as myofibroblasts, which has led to possible misclassifications in the literature [8,9].

4.3 Fibroblasts

Shared and well-acknowledged functions of all FB subsets is the production of ECM proteins as well as an important role in tissue repair and wound healing. More recently, fibroblasts have been demonstrated to play an essential role in supporting the intestinal epithelial niche including Wnt-signalling supporting the intestinal epithelial stem cell niche in crypts and BMPs supporting the differentiation of mature epithelial cells[209,211,213,214]. A shared feature of FB is the expression of PDGFRα, albeit at different levels [9,208,209]. While several subsets of FBs have been identified within the intestine, the nomenclature and the phenotypic markers used to identify these subsets has yet to be unified. In the following sub sections I discuss intestinal FB subsets based on their proposed locations.

Subepithelial FB: A thin layer of subepithelial FB directly underline the epithelial layers, and it is generally acknowledged that they can be separated from other FB by the highest expression of PDGFRα observed. They have been shown to produce ECM specifically for sustaining the basal lamina (a thin membrane just below the epithelial cells) and they secrete factors for both maintaining the intestinal epithelial stem cell niche and for driving their differentiation along the length of the crypt to villus tip[215]. However, several different subsets have been described in this context and though they share several features, many are inconsistent between studies. These investigated subsets include so-called interstitial cells of Cajal (ICC)-like cells, which can provide electric stimuli to the muscle layer, and telocytes involved in intercellular signalling to surrounding cell types via long processes[213,216–219]. Adding to the confusion, studies characterising subepithelial FB spatially with electron microscopy and immunohistochemistry (IHC) are in disagreement on whether some subepithelial FBs themselves or just a closely associated cell type expresses low amounts of α-SMA[220–222]. The genetic counterpart of α-SMA, *Acta2*, has also been reported to be expressed in FB subsets expressing high levels of PDGFRa in scRNA-seq studies, in some cases these are referred to as myofibroblasts. The term myofibroblasts is, however, used rather broadly and could in some cases include or denote *bona fide* SMC, albeit these should be devoid of PDGFRa and thus has led to misleading classifications[223–225]. Consequently, the heterogeneity of PDGFRα^{high} subepithelial FB is still largely undetermined.

Crypt/perivascular FB:

In colon a subset of submucosal ACKR4 expressing FB are known to surround the intestinal vessels[226]. Hong *et al*. observed a similar cluster of *Ackr4*$^{+}$ in the SI and suggested that these were equivalent to the submucosal FB reported in association with submucosal vessels[208]. McCarthy et al. did not report Ackr4 expression in the SI, but could identify a populations of *Pdgfra*^{low} cells in the SI whith selective expression of *Grem1* and *CD81* and was located below crypts[209]. And in colon, Roulis *et al.* observed a cluster of *Pdgfra*$^{+}$ co-expressing *Ptgs2* which were also located just below crypts[207]. Similarly, Brügger *et al.* reported two subset of colonic subcrypt associated FB with *Pdgfra* expression[227]. The above subsets were all in proximity to the crypts and shared suggested roles of supporting the epithelial stem cell niche in the crypts. This role was suggested to be maintained by expression of canonical Wnt signalling elements and BMP inhibitors. The former directly supporting the “stemness” preservation in

crypts and the latter inhibiting BMP driven maturation of epithelial cells[207–209,227,228]. Kinchen *et al.* also observed a murine subset of *Ackr4*$^{+}$ *Cd81*$^{+}$ cells, their localization was not clarified. However, their main involved gene ontologies included BMP and Wnt signalling and together with their gene expression could indicate position in proximity to crypts[213].

Interstitial FB: PDGFRα$^{+/low}$ FB can also be found interstitially in SI villi and intercryptal regions in colonic LP. One such subset was identified in the SI by McCarthy et al. as a *Pdgfra*^{low} subset, these did not express the selective genes of the subcryptal FB from the same dataset, but had expression of several BMPs and were located to the base of the villi[209]. Similarly, Roulis et. al identified a major murine colonic interstitial *Pdgfra*^{low} FB subset expressing *Cd34* and differentiated based on expression of *Fgfr2* expressing FB[207]. The *Fgf2*$^{+}$ FB have been observed to locate to the interstitial villus region of the SI and closer toward the lumen of the crypt in colonic tissue[229]. The interstitial FBs have been proposed to promote differentiation of mature epithelial cells by expression of BMPs and non-canonical Wnt signalling[230]. Together the BMP and Wnt signalling cues create a gradient where crypts remain stem like and proliferating epithelial cells are pushed out of the niche and meet cues moderating their differentiation into mature cells[209,231,232]. Hong et al. also identified three populations of interstitial villi FB, however, only one of these were by gene ontology interpreted to play a role in BMPs or non-canonical Wnt signalling by having a transcriptomic profile aligning to tissue morphogenesis[208]. Interestingly, Kinchen *et al.* found the interstitial FB populations were overlapping between human and mice colonic tissues, with the exception of a CCL19^{+} subset in the human dataset [213]. CCL19 is typically used as a marker of lymphoid associated MSC, suggesting contaminating GALT tissue in the human data set[213].

4.4 Maintenance of iMSC

The intestinal MSCs appear to turn over slowly in adult life [9], and the mechanisms contributing to their maintence remain largely unknown. Several studies have suggested the presence of MSC progenitors within the intestine[8,213,233], however the requirement for these progenitors in maintaining intestinal FB subsets in steady state remains unclear. These potential progenitors include pericytes [219,234,235], though disproved by Guimarães-Camboa *et al.* [236]. Evidence that CD34^{+} adventitial FB contain potential MSC progenitors comes from observations that CD34^{+} adventitial cells in adipose tissues can give rise to pericytes [213,237–244], lymphoid MSC subsets [243,244] and that these cells display tri-lineage potential in vitro, giving rise to adipocytes, chondrocytes and osteocytes [237,244]. Furthermore, in other tissues such as the lung, CD34^{+} mesenchymal stem cells inhabit a perivascular niche and can proliferate and give rise to myofibroblast after injury[241]. The intestinal adventitia of the submucosal vasculature is structurally similar to the perivascular niche of other tissues. The main subset which has been associated with this area are the *Ackr4*$^{+}$ FB, which are located mainly in submucosa, the deep mucosa and vessel-associated adventitia, these also express high levels of *Cd34* [239,240]. Interestingly, Kinchen *et al.* found that ACKR4^{+} FB expanded after dextran sulfate sodium (DSS) induced colitis [207,213,227,233]. Another SC study had similar findings, but identified the subset as RBP1^{+}*Cd34*$^{+}$, but expressed several other comparable genes to ACKR4+ subset from Kinchen *et al.*[213]. Further computational trajectory analysis proposed the *Ackr4*$^{+}$ FB give rise to both subsets of *Pdgfra* expressing FB and interstitial? [233]. Other suggestions have been that mesothelium could provide a source for iMSC during injury[245]. This was suggested by a report finding lineage tracing of *Msln*$^{+}$ cells in a radiation model gave rise to vascular SMCs and some unidentified FB subsets[209]. However, this process

might not be a progenitor in homeostasis as differentiation did not take place in a four month steady-state period[209]. Together these underline the lack of consensus and need for further investigations in the intestinal MSC development in homeostasis.

5 Inflammatory Bowel Disease

Inflammatory bowel disease (IBD) are chronic incurable inflammatory diseases of the intestine. IBD has been been classified into two major disease types Crohn's disease (CD) and ulcerative colitis (UC), although it is clear that disease heterogeneiety exits within both conditions. While CD can affect all layers of the gut wall and any site of the intestinal tract, it most commonly involves the ileum and the proximal colon[17]. In contrast, UC is usually restricted to the large intestine and inflammation is limited to the mucosa[17]. The incidence of both CD and UC has increased worldwide over several decades[245,246]. The most effective biological therapies for the treatment of IBD are TNFα inhibitors however many patients do not respond to this therapy or lose responsiveness over time. There is thus an urgent need for novel therapeutic approaches for the managent of CD and UC. While much research and progress has been made in understanding disease phenotype and underlying risk factors [247] we still do not know why or how these debilitating diseases develop, although they are likely to be multifactorial. In order to understand IBD and possible novel therapy targets we need to understand the cells existing in the affected tissues and how they differ compared to the healthy gut.

GWAS studies have revealed risk alleles associated with increased incidence of IBD[247,248], however, these do not account for all the heterogeneity of the disease risk, e.g. many people who have these risk alleles will never develop IBD[249–251]. IBD is a more complex set of disorders and associated with several layers of disease-susceptibility, where it seems several must be breached in order for disease to occur. It is also not yet clear what defects initiate IBD and which are simply results of active IBD. The layers include genetic factors, environmental factors, barrier deficiencies, dysbiosis and infection. Furthermore, the innate and adaptive immune systems must be dysregulated to maintain the disease [2].

5.1 Monocyte and macrophages in IBD

A common observation in murine models of colitis and IBD is the accumulation of P1 and especially P2 monocytes from the monocyte waterfall within the intestinal LP[76,252–255]. Interestingly, in healthy intestine monocyte recruitment is mediated by CCR2 and $\alpha_4\beta_7$, but in inflamed tissues the homing is mainly mediated by CCR2[256]. Notably, in DSS induced colitis, inflammation can be ameliorated by monocyte specific deletion of CCR2, indicating a key role for recruited monocytes in this colitis model[157]. Intestinal inflammation have also proposed to be maintained by the CD169$^+$ vas-Mac, and in a DSS colitis model the inflammation could be ameliorated by treating with anti-CCL8, indicating an important role for CCL8 in monocyte recruitment during inflammation[166]. Furthermore, the influence of inflammation on monocyte differentiation was investigated using a surgical induced model of inflammation and transfer of monocytes to determine differentiation process. Here, it was observed that though differentiation of monocytes phenotypically only had small gradual changes, the transcriptional expression of several genes and consistently *Ccl2, Cxcl1, Il1a*, *Il23a*, and *Cd14* were distinctively upregulated in inflamed tissue[256]. This suggest that a distinct development of monocytes could occur in inflamed tissue of mice during inflammation. Consistent with

early stages of the monocyte waterfall accumulating in IBD as well as colitis models in mice, infiltrating $CD14^+$ cells are CD163-/dim and produce IL-23 and IL1β in CD lesions[253].

Interestingly the accumulation of early and intermediate monocytes does not result in a reduction in total number of tissue resident macrophages[257], furthermore, in CD lesions and DSS colitis the mature macrophages are generally found to remain anergic, indicating that they do not contribute to the inflammatory process[258].

The accumulating P1 and P2 monocytes in colitis express a wide range of inflammatory effectors including; (1) CXCL10 and CXCL9, which functions as chemokines to recruit CXCR3+ effector T cells [258] (2) IL-1β and IL-23 that promote Th17 and Th17/Th1 polarization and survival [252,253] and (3) IL-1β, IL-6 and TNFα which, amongst many pro-inflammatory effects, have been described to lead to barrier defects by compromising epithelial barrier integrity[259]. Finally, it has been suggested that the inflammatory monocytes via cytokine production promotes FB activation, and FB production of CCL2, the ligand for CCR2 expressed on blood monocytes thereby creating a positive feedback loop for inflammation[258].

5.2 cDC in IBD

Though MNPs change in abundance in IBD, an additional established phenotype of IBD is the accumulation of activated effector T cells[260], and neutrophils[43] within the inflamed intestinal tissue and skewed IgA/IgG antibody production by plasma B cells[261]. Indeed an important current biological therapy in IBD is to target the T and B cell gut homing receptor $\alpha_4\beta_7$[262]. Given their key role in the initiation and differentiation of intestinal adaptive immune responses, surprisingly little is known as to the role of cDC and cDC subsets in IBD. In a recent scRNA-seq study it was reported that mature or 'activated' cDC are found more commonly in CD lesions compared with uninvolved intestinal tissue, however this observation was not consistent across patients. In patients with increased proportion of activated cDC, these co-localized with aggregates of T and B cells, which could indicate that the inspected lesion in these cases included a lymphoid structure such as an ILF[263]. This enhanced activation state is in line with an older study, which reported upregulation of TLR2, TLR4, and CD40 on intestinal lamina propria cDC in IBD, however here there was no indication of whether these cells were associated with GALT[264]. In another study, LP $CD103^+$ cDCs from UC patients were found to have impaired induction of $FoxP3^+$ $CD4^+$ T cells *in vitro*, and instead promoted the induction of IL-17, IL13, and IFNγ producing $CD4^+$ T cells[265]. The proportions of $CD141^+$ cDC1 and $CD1c^+$ cDC2 does not appear to change between healthy controls, uninvolved, and inflamed tissues from IBD patients, however, the $CD103^+$ proportion of both $CD1c^+$ and $CD141^+$ cDCs are reduced in IBD[257].

As established, environmental priming of cDCs can affect downstream T and B effector cell responses. In IBD abnormal cDC conditioning has been suggested to affect the disease state. Both TGFβ and RA expressed by epithelial cells are downregulated in IBD potentially affecting the environmental conditioning of cDCs[258,266].

In conclusion, there remains much to be learnt on the diversity and function of cDC subsets in IBD.

6 Single Cell Technology

Single cell transcriptomic sequencing and the aligning technologies first emerged in 2009 with Tang *et al*. describing a single cell gene expression profiling assay[267] and the technology has been on the rise ever since. The technology allows us to distinguish single cells in thousands of expression patterns and unravel heterogeneity which cannot be addressed in low dimensionality data sets like flow cytometry. Within recent years the technology has allowed for additional dimensionality on single cell level by adding data such as chromatin accessibility (ATAC-seq)[268], spatial transcriptomics (albeit still with limited resolution), DNA methylation [269] and surface protein expression (CITE-seq and cell hashing)[270,271]. These data types are part of great efforts all over the world in creating reference databases of the entire human body in health and disease. These include efforts such as the Human Cell Atlas, the Allen Brain Atlas, the Gut Cell Atlas and the Human Tumour Atlas Network.

Multiple different platforms exist for single cell RNA sequencing (scRNA-seq). These methods differ in several stages of the processing like single cell isolation, lysis, amplification, cDNA formation, tagging and transcript coverage. The outputs differ depending on the possible number of single cells, and in the depth and coverage of sequencing[272]. The commercial platform 10x Chromium is a popular choice since it allows recovery of a large proportion of input cells, reasonable costs per captured cell, and tagging at either 3' or 5'. The output transcripts are not full-length coverage, however, this type of tagging allows for the use of unique molecular identifiers (UMI)[273], which can account for amplification bias during PCR amplification.

The biggest advantage using scRNA-seq to bulk RNA-seq is enabling identification of rare cell subsets, unknown heterogeneity and capturing multidimensional dynamics of cellular processes such as cell differentiation. An important example of how scRNA-seq data is changing our understanding of biology is the conceptual change that hematopoiesis as a continuous and not a stepwise process. However, the biggest disadvantage of the technique is still the sparsity of the data, even with novel full-length, well-based platforms for scRNA-seq that allows for much deeper sequencing per cell. scRNA-seq data is extremely sparse, it has been estimated that up to 90% of a raw output matrix consist of zero values. Even with computational approaches, it cannot be completely clarified what is a true biological zero and what is caused by drop-outs[274]. Compared to flow cytometry scRNA-seq holds the advantage of giving an unbiased and much higher dimensionality view of cells of interest. Where flow cytometry is limited by fluorescent spill-over, scRNA-seq experiments, including the addition of a large panel of surface marker labels, is mainly limited by cost. The addition of surface protein information to scRNA-seq also means the technology is one step closer to figuring out the correlation of gene expression and actual protein expression[275,276].

6.1 Computational Pipeline

The emergence of scRNA-seq has been followed by an extensive wave of novel tools to process the increasing amount of data and handle the different challenges scRNA-seq data represents. Currently, scRNA-tools.org reports more than 1100 tools in 30 different categories[277]. These tools handle different types of data analysis and processing on different platforms. The most appropriate tool to use in a given case, will depend on the questions and hypotheses posed[278,279]. However, some tools like Seurat[280] and Scanpy[281] are widely used

and handle initial steps which are necessary in most cases, these steps could include preprocessing, normalization, dimensionality reduction and more (***Figure 5***, blue box).

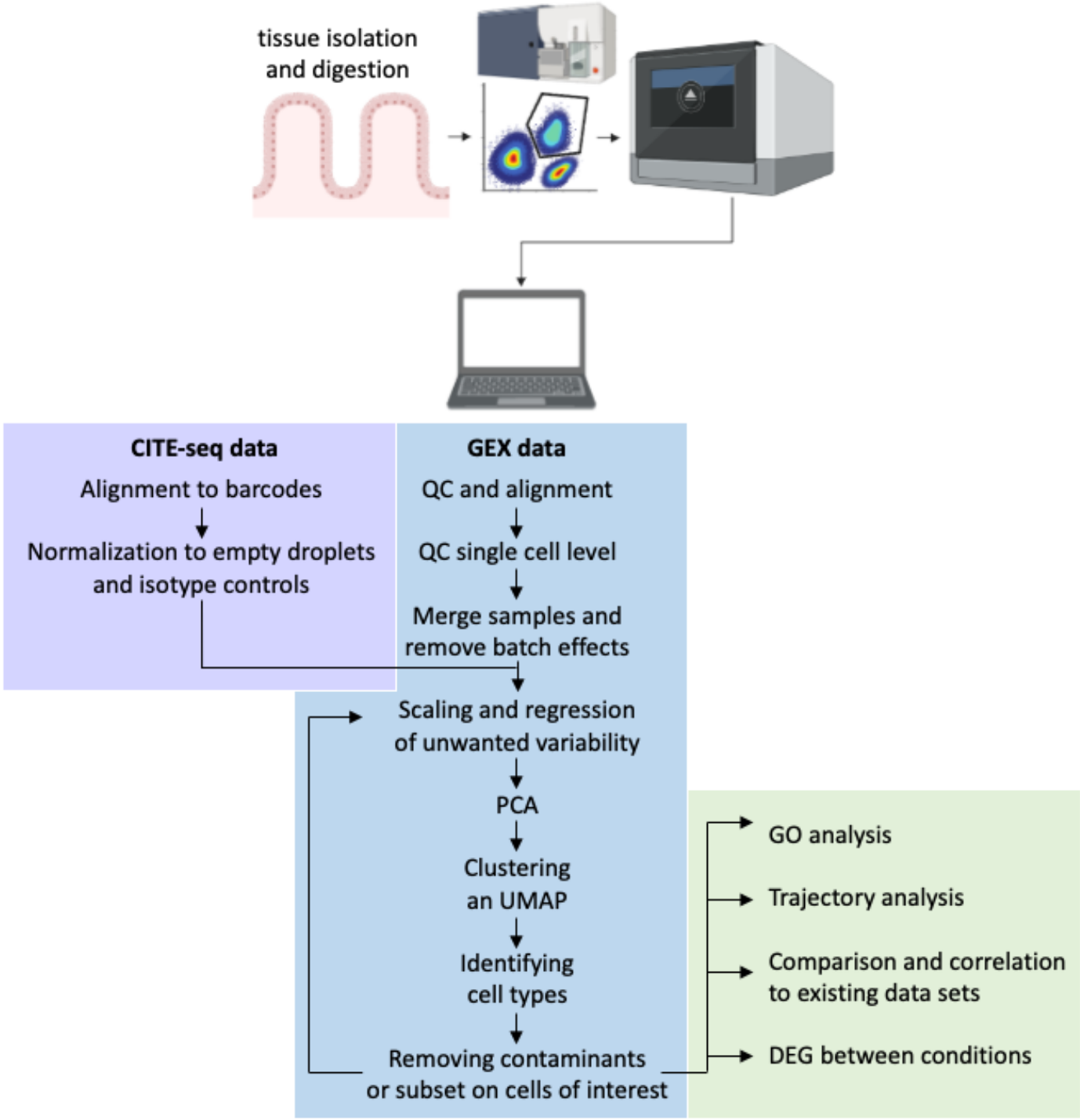


Figure 5 - Computational single cell pipeline utilised in this thesis: Isolated and digested single cell suspensions are sorted to enrich specific cell types. The cells are subjected to the 10x Chromium platform and sent off for sequencing. Fastq files from sequencing are subjected to the outlined computational pipeline. GEX data is always processed with the steps outlined in the blue box. If CITE-seq antibodies are used they are integrated into the GEX data after the steps outlined in the purple box. After cell subsetting, the green box outlines some of the possible downstream analyses. Created with BioRender.

Batch effects: In scRNA-seq, like other experimental protocols, there are both technical and biological sources of variation which are unwanted and can confound downstream analysis. Technical sources of variation can be anything from the machine used for sequencing, the single platform or differences in experimental designs. Some of the main biological variables

are interpatient variability, different conditions in mice models and cell states like cell cycling[280,282]. The former two are often referred to as batch effects. Depending on the experimental setup any of these bias can exclusively drive clustering and thus identification of cell types. In the last years several tools have been developed which can address this issue with differing computational usage and scalability for larger data sets. Two approaches that have scored well in benchmark studies are Harmony and Seurat v3 anchor integration [280,283,284].

Clustering: A common step in single cell analysis is clustering, which is supposed to assign cells into biologically significant groups in an unsupervised manner. This can be driven by cell type or cell state. A popular approach with single cell data uses k-nearest neighbourhood (KNN) graphs, where cells are nodes and the cell-to-cell distances represent the edges. The KNN graph is then used to iteratively find cluster centers and assign each cell until cluster centers are stabilized. Clusterings will most often require an input parameter which decides either directly or indirectly the number of output clusters[285]. If the data consist of highly heterogeneous populations, like total CD45+ immune compartment, it can be advantageous to either do subclusterings or reclusterings when focusing on heterogeneity of specific areas of interest[286]. Another approach commonly used in CyTOF data is overclustering where the output is set to an arbitrarily high number of clusters. This can help to split clusters which are driven by unwanted effects such as cell cycle into biologically relevant subsets, this also ensures that rare cell types are not overlooked[287,288]. Cell types can then be decided on the high number of clusters and merged resulting in a semi-supervised cell clustering. A limitation to clustering is that it is run under the assumption that discrete subpopulations exist within a data set, meaning the KNN graphs will consist of more or less disconnected neighbourhoods. When dealing with transitional data like cell differentiation the connecting edges will be dense across the entire differentiation process. In these instances it can be better to order cells according to their differentiation with trajectory inference (TI) type analyses [286].

Trajectory inference: TI orders cells based on transient gene patterns rather than focusing on finding the sources of most variability in a data set. TI consists of two overall building stones: dimensionality reduction and trajectory modelling where cells are ordered in the lower dimensionality space. How this is accomplished depends on the tool. A common disadvantage to TI tools is the assumption that underlying trajectories exist in the data, meaning even if no biological meaningful trajectories exist one will still be found. Another limitation is the need for all intermediates to be present in order to construct a meaningful TI[278]. In 2018, a different angle to TI was developed when La Manno *et al.* found that 15-25% of reads from different scRNA-seq platforms aligned to unspliced intronic sequences. The intronic regions correlated with the exonic counts and were suggested to represent unspliced precursor mRNA for the exonic regions of the same genes. They showed that this was true for circadian-associated genes in mouse liver, where intronic transcripts predicted exonic transcripts at the next time point. This was the foundation to develop 'RNA velocity', which calculates cell trajectories based on the unspliced to spliced gene patterns[289].

CITE-seq data: CITE-seq data adds invaluable protein information on top of what the RNA transcripts normally obtained in scRNA-seq experiments. However, CITE-seq data like flow cytometry comes with certain challenges like non-specific binding of antibodies and background signals. These challenges can be overcome in flow cytometry by using isotype

controls and FMOs. Isotype controls are also available for CITE-seq data, however not many tools utilize the information contained in the isotype controls. These issues prompted the development of the tool Denoised and Scaled by Background (DSB). DSB exploits empty droplets to correct for protein-specific background noise and includes isotype controls to correct for non-specific binding of the antibodies[290] (***Figure 5***, purple box). Although this method exploits the cell specific background, it is not yet capable of utilising the specific isotype controls paired to the CITE-seq antibody.

Aims of the thesis

MNPs are crucial immune cells in both homeostasis and infection. In the intestinal tissues they harbor many specific functions and several of these could potentially maintain or initiate disease states in IBD. However, previous methods have been performed on tissue sections potentially containing both GALT and effector tissues as methods for dissecting these simply did not exist. With the protocol developed by Jørgensen *et al.*[29] and unbiased single cell sequencing of sorted MNPs we aimed to:

- **Manuscript I:** Identify the different MNP subsets in the uncontaminated human intestinal lamina propria and to investigate whether development from progenitors into mature intestinal subsets happens during homeostasis. We also aim to resolve whether there are differences in abundance and gene expression between the SI and the colon. This can serve as a reference of the MNPs in healthy intestinal LP in any comparison.
- **Manuscript II:** Unravel differences in abundance between MNP subsets of the isolated GALT compartments to the MNPs found in effector tissues. Investigate whether we can identify transcriptomically indicated functions in cDC1 and cDC2 between GALT and surrounding LP. Furthermore, with patient data from SM-ILF and colon LP from one CD patient we also wanted to preliminary identify whether there are any changes in the MNP subsets between healthy and inflamed tissue.

Furthermore, at the time this PhD project was commenced, there was a lack of understanding of the importance of stromal cells and their diversity and ontogeny in the mucosa and how these related between the SI and colon. It was also not clear whether specific subsets of MSC are responsible for the different support functions of gut homeostasis. In this manuscript I contributed with computational analysis of bulk- and scRNA-seq, in order to:

- **Manuscript III:** Investigate diversity and ontogeny of intestinal lamina propria stromal cells from murine LP and how these subsets differ between SI and colon. We also sought to design a phenotypical panel, which could help clarify the distinctions in existing studies and combine with IHC for spatial understanding of the MSC subsets.

Manuscript I

Single-cell characterization of mononuclear phagocytes in the human intestinal mucosa

Single-cell characterisation of mononuclear phagocytes in the human intestinal mucosa

Authors: Thomas M. Fenton[1, 2,†*], Line Wulff[1,†], Gareth-Rhys Jones[3], Julien Vandamme[1], Peter B. Jørgensen[1], Calum C. Bain[3], Julie Lee[4], Jose MG. Izarzugaza[5], Kirstine G. Belling[6], Gwo-Tzer Ho[3], Ole H. Nielsen[7], Lene B. Riis[8], Tune H. Pers[9], Henrik L. Jakobsen[10], Allan M. Mowat[11], Søren Brunak[6], William W. Agace[1,12,*]

Affiliations:
[1]Mucosal Immunology group, Department of Health Technology, Technical University of Denmark, Kemitorvet, 2800 Kgs. Lyngby, Denmark
[2]Institute of Immunology and Infection Research, University of Edinburgh, Ashworth buildings, Edinburgh, EH9 3FL, UK
[3]University of Edinburgh Centre for Inflammation Research, Queens Medical Research institute, Edinburgh, EH16 4TJ, UK
[4]Novo Nordisk Foundation Centre for Stem Cell Biology, DanStem, University of Copenhagen, 2200 Copenhagen N, Denmark
[5]Center for Biological Sequence Analysis, Technical University of Denmark, 2800 Lyngby, Denmark
[6]Translational Disease Systems Biology, Novo Nordisk Foundation Centre for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
[7]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark.
[8]Department of Pathology, Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark
[9]Novo Nordisk Foundation Centre for Basic Metabolic Research, University of Copenhagen, 2200 Copenhagen N, Denmark
[10]Department of Gastroenterology, Surgical Section, Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark.
[11]Institute of Infection, Immunity and Inflammation, College of Medicine, Veterinary Medicine and Life Sciences, University of Glasgow, Glasgow, UK
[12]Immunology Section, Lund University, BMC D14, 221-84 Lund, Sweden

[†]These authors contributed equally to this work

[*]Corresponding authors: thomas.m.fenton@gmail.com, wiag@dtu.dk

Abstract

Subsets of mononuclear phagocytes (MNP), including macrophages and classical dendritic cells (cDC), are highly heterogeneous, with each subset playing distinct roles in immune responses. Understanding this complexity of intestinal MNP at the cellular level has proven difficult due to the expression of overlapping phenotypic markers and the inability to isolate leukocytes of the mucosal lamina propria (LP) effector site, without contamination from intestinal lymphoid tissues, which are embedded in the mucosa, and which are responsible for the induction of immunity. Here we exploit our novel method for separating LP from intestinal lymphoid tissues to explore the transcriptional, phenotypic and developmental diversity of MNP populations in the human ileum and colon LP. Amongst cDC we identify cDC1, cDC2, and a population of $CD1c^+$ cDC subset that shares features of cDC3, recently described in human peripheral blood, and demonstrate transcriptional diversity within these populations between the ileum and colon. We further provide evidence that each of these subsets originate from distinct subset-specific precursors present within the intestine. Within the monocyte/macrophage compartment we identify monocytes, non-classical monocytes, and mature macrophage subsets as well as developmental intermediates along the monocyte 'waterfall'. Finally, we show that MNP proportions vary between the ileum and colon and in the setting of inflammatory bowel disease (IBD). Collectively our results provide novel insights into intestinal MNP heterogeneity, development, and function of relevance to intestinal immune homeostasis and disease.

Introduction

The mononuclear phagocyte (MNP) family consists of classical dendritic cells (cDC), classical monocytes, non-classical monocytes, and macrophages, each of which play specific roles in the induction and function of immune responses, tissue homeostasis and inflammation [1–3]. Whereas cDC are the main cells involved in the induction and shaping of adaptive immune responses and tolerance [4], macrophages are primarily involved in maintaining tissue homeostasis, and in promoting inflammation in response to infection or tissue damage [5,6]. Thus, characterising these populations and understanding their functions under different conditions will be important for improving the treatment of disease and for developing preventative strategies such as vaccines. However, this has proved difficult due to the expression of overlapping phenotypic markers and the increasingly evident heterogeneity within each subset [7–10]. This is particularly so in non-lymphoid tissues such as the intestine, where local factors imprint unique adaptations and functions, meaning that each tissue has to be explored independently and at the single-cell level [11].

Numerous MNP populations have been described in the human intestine, where they are thought to play important roles in disorders such Inflammatory Bowel Disease (IBD) [12,13]. However, much of what is known about intestinal MNPs comes from work in mice and there are very few direct comparisons of these cells in the small and large intestine, which represent anatomical compartments with distinct functions and properties [14]. Furthermore, although recent work has examined human intestinal leukocytes at the single-cell level [15,16], these studies did not discriminate between cells in the mucosal lamina propria (LP) and those in the secondary lymphoid organs of the gut-associated lymphoid tissues (GALT), which include the isolated lymphoid follicles (ILFs) which are embedded in the LP [17–19]. As these compartments have distinct roles in local immunity, with the GALT being responsible for the initiation of antigen-specific B and T cell responses, while the effector responses take place in the LP and epithelium [20–22], it is critical to assess their cellular components independently. Recently, we developed a novel technique which allows ILFs of the GALT and the surrounding LP to be

isolated separately from human intestinal tissue [23]. By combining this method with CyTOF, flow cytometry, and IgA-seq, we recently generated an atlas of effector lymphocytes in the LP of the small intestine and colon, and showed that these compartments contain very different populations of T and B cells [24]. Here, we have applied the same approaches to analyse the heterogeneity of MNP that are derived unequivocally from the LP of healthy human ileum and colon. As well as separate clusters of mature macrophages, we show that mucosal cDC can be divided into three distinct subsets and, for the first time, describe the presence of putative subset specific cDC precursors (pre-cDC) in the human intestine. Trajectory analysis revealed clear developmental relationships between tissue monocytes and macrophages, and between pre-cDC1, 2 and 3 and their mature cDC counterparts. By providing novel insights into the heterogeneity and development of intestinal MNP, our findings should help target approaches for modulating intestinal immune responses.

Results

Single-cell sequencing of MNP subsets from the human intestinal mucosa

To characterise MNP diversity within the human intestinal LP, surgical samples of ileum and colon from uninvolved areas of colorectal cancer patients were processed to remove contaminating GALT and submucosa (SM), as we recently described [23,24]. Following LP digestion, single-cell RNA sequencing (scRNA-seq) was performed on flow cytometry-sorted $CD45^+CD3^-CD19^-HLADR^{int/+}$ cells from ileum and colon LP cell suspensions, using the 10x Chromium system (**Fig. 1A**). Sequences were obtained from a total of six colon LP and four paired ileum LP samples (**Table S1**). Distinct clusters of contaminating *CD3E*$^+$ T cells, *CD79A*$^+$ B cells, *VWF*$^+$ endothelial cells, *MS4A2*$^+$ mast cells, *COL3A1*$^+$ stromal cells, and *NRXN1*$^+$ glia, were identified and excluded from further analysis (**Fig. S1A**). Expression of *HLA-DQA1* (MHCII) identified one 'supercluster' and two peripheral clusters (**Fig. S1B**), which were computationally isolated and re-clustered. These 28,758 $MHCII^+$ cells comprised distinct clusters of *IL3RA*$^+$ plasmacytoid DC (pDC), *CLEC9A*$^+$ cDC1, and *FCGR3A*$^+$ non-classical monocytes (**Fig. S1C and 1B**), together with a large cluster that contained cells expressing either the cDC2-associated marker *CD1C,* the monocyte/macrophage (Mono/Mac)-associated marker *CD14*, or both (**Fig. 1C**). Flow cytometry analysis of colon LP $CD45^+$ $HLA\text{-}DR^+$ $lineage^-$ cells confirmed the presence of CD1c and CD14 single positive cells, as well as cells expressing variable levels of both CD1c and CD14 (**Fig. 1D**).

To further characterise subsets within this MNP supercluster, these cells were re-clustered at high resolution and were analysed for expression of a panel of monocyte, macrophage, and cDC2/3 signature genes (**Fig. 1E**). Each of these high-resolution clusters was present in both ileum and colon LP preparations albeit in slightly different proportions (**Fig S1D**). These clusters could be broadly divided into two groups. Clusters X1-20 expressed either the classical monocyte transcription factor *ZBTB16*[25], the macrophage-associated genes *SEPP1, MERTK* and *MAF,* or a mixture of both, as well as other TFs involved in tissue-resident macrophage development such as *ID3*[26], indicating they represented monocytes

(Mono), macrophages (Mac) or Mono/Mac intermediates. In contrast, clusters X21-X35 expressed lower levels of Mono/Mac associated genes, but high levels of the cDC2/3-associated genes *AP1S3*, *FLT3, SEPT6* and *IRF4* [9,27] (**Fig. 1E**). The remaining clusters X36-39 expressed both cDC2-associated genes such as *FLT3* and *IRF4,* and Mono/Mac-associated genes, including *CD14* and *C5AR1* (CD88) (**Fig. 1E**). Since cDC3 have been reported to co-express these markers [7,25,27–30] we assigned these cells to the cDC2/cDC3 sub-cluster. Some genes previously used as signature markers to discriminate cDC2/3 from monocytes in blood, such as *FCER1A* and *CLEC10A* [7], were broadly expressed amongst most clusters, and thus were not useful in segregating these populations in the human intestine (**Fig. 1E**). As a result of this analysis, we could tentatively divide the MNP supercluster into Mono/Mac and cDC2/cDC3 sub-clusters (**Fig. 1F**).

To assess the accuracy of designating these clusters as Mono/Mac and cDC2/3, pseudo-bulk PCA analysis of the clusters was performed using three datasets of MNP gene signatures that included blood cDC2, classical monocytes, and *in vitro* monocyte-derived macrophages [31,32]; signatures that have been used previously in combination to define subsets [33]. Consistent with our supervised approach, the clusters separated into three groups, with PC1 driving separation of cDC2/3 and PC2 separating monocytes from macrophages (**Fig. 1G**). In summary, human intestinal monocytes, macrophages, and cDC2/3 showed a high degree of transcriptional overlap. However, by combining high resolution clustering and gene expression analysis of published data sets, we were able to separate them into Mono/Mac, cDC1, pDC and cDC2/3.

Identification of human intestinal cDC1, cDC2 and cDC3

To further explore intestinal LP cDC diversity and origins, the cDC2/3 clusters were computationally isolated and recombined with cDC1. Cells were clustered at high resolution to enable accurate designation of small cluster identity, based on unique gene expression patterns. Trajectory-based clustering was then performed using tSpace[34], with the clusters

being visualised by 2D representation of 3-dimensional tSpace UMAP (Flat tUMAP) (**Fig. S2A**). All sub-clusters were present in both ileum and colon datasets, albeit in slightly different proportions (**Fig. S2B**). Seven clusters, located together at the top of the tUMAP, were enriched for low expression of MHCII genes and/or high expression of mitotic G2M/S genes (**Fig. S2C**), indicating they are not mature cDC, and which will be considered later. Of the remaining clusters, seven showed a clear cDC1 signature, expressing the canonical cDC1 genes *CLEC9A, CADM1, XCR1, BATF3,* and *IRF8* (**Fig. S2D**). The remaining clusters were ordered according to increasing expression of a cDC3 signature score, which was based on genes identified by Bourdeley et al[30] (**Fig. S2E**). This revealed a contiguous group of clusters with a high cDC2 score and lower expression of cDC3 genes, which we characterised as likely cDC2 (**Fig. S2E**). Conversely, a group of contiguous clusters with high cDC3-scores and lower cDC2 scores was identified as likely cDC3 (**Fig. S2E**). The remaining clusters displayed ambiguous identity scores (**Fig. S2E**) and formed a contiguous group sitting between cDC2 and cDC3 on the tUMAP plot (**Fig. S2A**). Based on this analysis, we classified the mature cDC clusters as cDC1, cDC2, ambiguous cDC2/3, and cDC3 (**Fig. 2A**). The definition of these clusters was supported by their expression of cDC subset-specific signature genes, including expression of *CLEC9A* and *CADM1* by the cDC1 cluster, *IRF4* and *PLAC8* by the cDC2 cluster, and *C1QA* and *CD163* by the cDC3 (**Fig. 2B**). Notably, the cDC2, cDC3 and ambiguous clusters all expressed comparable levels of genes that have been used previously to differentiate cDC2 from cDC3 in human blood, including *CD1C*, *CLEC10A* and *FCER1A* (**Fig. 2B**), indicating the possible importance of tissue-specific conditioning in cDC subset identification.

The differences between intestinal cDC subsets were underlined when transcription factor expression was analysed. As expected, cDC1 expressed the TFs *BATF3, IRF8* and *ID2*[35]*,* and they also expressed *ZEB1* which has been reported to regulate Th1 induction by cDC1[36] (**Fig. 2C**). In contrast, only a few TFs were expressed preferentially by cDC2, including the DC2 signature TF *IRF4*, as well as the developmental genes *ARID3A*, *FOXC1*, *HES4*, and

MSX1 (**Fig. 2C**). cDC2 also expressed high levels of *NR4A3,* which has been implicated in DC activation[37] and in differentiation of moDC *in vitro*[38]. The cDC3 clusters expressed macrophage-associated genes including *MAF, MAFB, MAF* and *ZBTB16*, as well as the inhibitory transcription factors *NFKBIA* and *NFKBIZ* (**Fig. 2C**). The ambiguous cDC2/3 clusters expressed high levels of TFs associated with activation, including ATF3 and JUNB (**Fig. 2C**). The transcription factors GATA2, NR2F2, and EGR2 were differentially expressed between the ileum and colon (**Fig 2C**), with GATA2 and EGR2 being expressed at higher levels in the colon, while the retinoic acid-inducible gene NR2F2 was expressed more by ileal than colonic cDC1 (**Fig. 2C**).

To explore the functional potential of the cDC subsets, we assessed their expression of cytokine mRNA. cDC1 expressed high levels of the TNF family members *TNF, TNFSF9* (4-1BBL) and *TNFSF11* (RANK-L), cDC2 highly expressed *CCL19* and *CCL22*, and cDC3 expressed a wide range of cytokines and chemokines, including *IL10*, *IL1B* and *IL6,* and the interferon-inducible chemokines *CXCL9* and *CXCL1* (**Fig 2C**). A number of cytokines (*CSF1*, *CCL17,* and *TNFSF10*) were expressed at higher levels by colonic versus ileal cDC2, while *CCL3, BMP2, CCL4* and *CXCL3* were expressed more by the ambiguous cells of the colon versus the ileum, perhaps reflecting increased activation in response to the higher microbial load of the colon. The only mediator to be expressed at higher levels in the ileum than in the colon was BMP4 in ileal cDC2 (**Fig 2C**). We also investigated expression of genes involved in antigen presentation (**Fig 2C**). Multiple MHCII genes were differentially expressed by either cDC1 or cDC2. Genes involved in lipid presentation, including *CD1A, CD1B, CD1C,* and *CD1E,* were higher in both cDC2 and cDC3 than in cDC1, whereas *TAP1* (involved in MHC-I peptide loading) was expressed selectively by cDC1 (**Fig 2C**).

cDC subsets were further analysed using the Discriminant Regulon Expression Analysis package (DoRothea), which infers transcription factor activity from expression of downstream target genes (**Fig 2D**), and the Pathway RespOnsive GENes package (PROGENy), which infers pathway activity in cells based on expression levels of pathway

response genes (**Fig 2E**). Dorothea showed activity of *SOX2*, *FLI1*, *LEF1*, and *FOXA1* in cDC1 (**Fig 2D**), and PROGENy suggested cDC1-specific activation of the PI3K pathway (**Fig 2E**). There were no TFs or pathways specifically activated in cDC2, but there were TFs active in multiple ileal cDC compared to colonic cDC subsets, including *SREBF1, SREBF2* (both involved in sterol/cholesterol metabolism)*, HNFA1, TAL1,* and *CDX2* (which drives intestine-specific transcriptional programs[39]) (**Fig 2D**). There were also TFs specifically active in colonic compared to ileal cDCs, most notably the pioneer TF *FOXA1* and its associated *ESR1,* which has been previously implicated in cDC development[40] (**Fig 2D**). In agreement with TF and cytokine gene expression (**Fig 2C**), TF activity in cDC3 included activation of the stress/inflammation-induced TFs *JUN, JUND, NFKB1, REL,* and *RELA* (**Fig 2D**). cDC3 also showed relative activation of multiple signalling pathways, including estrogen, androgen, WNT, trail, VEGF, p53, JAK-STAT, hypoxia, NFκB, and TNFα pathways (**Fig 2E**).

Because cDC3 showed activity of interferon regulated transcription factors including *STAT1* and *STAT3* (**Fig 2D**), we checked for expression of interferon regulated genes in the cDC subsets. One cluster within the cDC3 group expressed high levels of interferon regulated genes, including *CXCL11, ISG15, CXCL10, and IRF1* (**Fig 2F and S2F**). Thus, only a subset of cDC3 appeared to be responsible for the interferon-stimulated gene signature of cDC3. *CCR7* was expressed by distinct subsets of cDC2 and cDC3, and this was accompanied by expression of genes associated with cDC activation including *LAMP3*, *CD40* and *CCL19* (**Fig 2G and S2G**).

When gene expression was compared between ileum and colon, more genes were upregulated in the colon for each subset of mature cDC (**Fig 2H**); most upregulated genes were unique to one subset of cDC, although there were also significant numbers of genes that were shared across subsets (**Fig 2I**). Genes upregulated in the colon versus the ileum were involved in 'response to LPS/cytokines' (cDC2 and cDC3) and 'heat shock/stress' (all subsets), together with various transcription factors including *JUN, MAFF, KLF6* (all subsets), 'adhesion/extracellular vesicles' (all subsets), and TLR/IRF3/IRF7 signalling (cDC1 only) (**Fig**

S2H). In the ileum, cDC1 and cDC3 shared increased expression of the 'response to external stimulus' pathway, while cDC1 and cDC2 shared increased expression of ribosomal genes, and cDC2 and cDC3 shared increased expression of lipid metabolism genes (**Fig S2H**). Ileal cDC3 uniquely upregulated genes involved in lysosome and iron uptake/clathrin-coated vesicles, compared with the colon (**Fig S2H**).

Site- and disease-specific differences in DC subset composition

CD14 expression has been used previously to distinguish subsets of cDC2/cDC3, but here both cDC2 and cDC3 expressed low levels of CD14 (**Fig. 3A and B**) and equivalent levels of CD1c by Denoised and Scaled by Background (DSB)[41] normalized CITE-seq expression (**Fig. S3A**). To identify surface markers which could potentially be used to distinguish intestinal cDC2 from cDC3 by flow cytometry, colonic $CD1c^{+}CD14^{-}$ MNP were screened using a large antibody panel. Of 288 surface markers tested, CD1a, CD11a, CD103, CD206, and CD207 showed bimodal expression on colonic $CD1c^{+}CD14^{-}$ MNP (**Fig. S3B**). To map expression of these surface markers to the cDC2/3 scRNA-seq clusters, antibodies against these targets, as well as CD1c, were used to generate a panel of bar-coded antibodies for CITE-seq analysis of colonic LP cells. CD1a was not included in the CITE-seq panel, as it almost exactly matched the expression pattern of CD103 (data not shown). This approach showed that $CD14^{-}CD1c^{+}$ MNP could be split into 4 quadrants based on their expression of CD11a and CD207 (**Fig. 3C**). While ambiguous cells were split evenly between quadrants, the remaining cells in the $CD207^{+}$ $CD11a^{-}$ Q1 were almost entirely cDC2 while the majority of cells in the $CD207^{-}CD11a^{+}$ Q4 were cDC3 (**Fig. 3D**). CITE-seq analysis of ileal cDC paired with colonic cDC from a single patient also showed enrichment of cDC2 in $CD207^{+}$ $CD11a^{-}$ MNP and cDC3 in $CD207^{-}CD11a^{+}$ MNP (**Fig. S3C**). Based on these findings we designed a simple antibody panel that could be used for flow cytometric assessment of cDC subset distribution in different segments of the intestine and in response to inflammation. Examination of 10 matched ileal and colonic resection samples showed that the proportion of $CD207^{+}CD11a^{-}$ cells (primarily cDC2) were

enriched in the ileum, whereas CD207$^-$CD11a$^+$ cells (primarily cDC3) were enriched in the colon (**Fig. 3E and F**). To investigate subset frequencies in disease, colonic biopsies were taken from patients undergoing endoscopic screening for IBD diagnosis. Biopsies were classified by the gastroenterologist as showing active or quiescent IBD inflammation. Flow cytometric gating of CD1c$^+$ cDC in these biopsies showed that biopsies with active inflammation had a significantly higher proportion of CD207$^+$ CD11a$^+$ cDC (Q3) and showed a trend toward a higher proportion of CD207$^-$ CD11a$^+$ cDC, as well as a significant reduction in CD207$^+$ CD11a$^-$ cDC (Q1) (**Fig. 3G and H**). Thus, there seems to be an enrichment of cDC3 over cDC2 during active inflammation. Of note, some of the IBD patients were undergoing treatment with anti-α4β7 antibody (Vedolizumab) and biopsies from these patients showed a dramatic reduction in total CD1c$^+$ MNPs which affected the CD11a$^+$ and CD11a$^-$ subsets equally (**Fig. 3 I, J and K**). Importantly, Vedolizumab therapy had no effect on CD14$^+$ monocyte/macrophages, supporting the view that human intestinal cDC3 do not derive from monocytes.

The human intestinal LP contains precursors of cDC1, cDC2 and cDC3

Recent studies have identified committed precursors of cDC1 (pre-cDC1), cDC2 (pre-cDC2) and cDC3 (pre-cDC3), as well as uncommitted pre-cDC precursors in human bone marrow, blood [25,28,29,42] and tonsils [43]. To explore whether precursors of these kinds might also be present in human intestine, we utilised the high resolution clustering of the HLA-DR^{low} populations of cDC we had identified previously, revealing 8 distinct clusters (**Fig. 4A**). These cells were highly proliferative compared with mature cDC (**Fig. 4B**) and expressed low levels of the CD11c gene *ITGAX* (**Fig. 4C**), features consistent with previous studies of pre-cDCs in mice and humans [44–47]. As these proliferating clusters formed three distinct branches that aligned with cDC1, cDC2 and cDC3, we explored the hypothesis that each branch represented cDC subset-specific precursors. To do this, we generated signatures composed of the top 50 DEGs which distinguished the mature cDC subsets from each other and examined how these

were expressed by the various clusters of HLA ^{low} cDCs. This analysis showed that cluster 4 and 5 shared expression with cDC1, while cluster 7 expressed cDC2 DEGs and cluster 3 and 8 had similar expression patterns to cDC3 (**Fig. 4D**), consistent with each lineage of mature cDCs having separate committed precursors. RNA velocity analysis of mRNA splicing patterns[48] supported this idea further, with cluster 4 appearing to be at the beginning of a trajectory which showed directionality into cluster 5 and thereafter into the mature cDC1 clusters (**Fig. 4E**). Similarly, cluster 7 showed a trajectory into the mature cDC2 clusters, while cluster 3 showed a trajectory towards cluster 8 and then into mature cDC3 (**Fig 4E**). These patterns remained when the dataset was split into ileal and colon LP samples (**Fig. S4A**). Together these results suggest that clusters 4 and 5 are pre-cDC1, cluster 7 are pre-cDC2 and clusters 3 and 8 are pre-cDC3.

Three clusters adjacent to the putative committed pre-cDC (clusters 1, 2, and 6) did not express DEG specific for the mature DC subsets, and we hypothesised that they may be earlier, less-committed precursors. To determine whether any of these clusters showed evidence of commitment to any of the cDC lineages, we used the top 50 DEGs from each of the committed clusters of pre-cDC1 (cluster 5), pre-cDC2 (cluster 7) and pre-cDC3 (cluster 8) as input for a PCA containing all the HLA^{low} clusters. Of these 150 DEGs from the committed precursors, 79 were also DEGs in the mature cDC populations. PC1-2 split the cells into 3 distinct areas, each of which represented one of the committed precursor populations (**Fig. 4F**). Using this approach, cluster 6 aligned clearly with pre-cDC2, while one subset of cluster 2 (2A) aligned with pre-cDC1 and another subset of cluster 2 aligned with pre-cDC3 (2B). Most of cluster 3 also aligned with pre-cDC3 (3B), with a few cells in the pre-cDC1 region (3A) (**Fig. 4G**). In contrast, cluster 1 and a small subset of cluster 2 did not overlap clearly with any of the pre-cDC groups and did not express markers associated with any of the cDC lineages, suggesting they may contain non-committed precursors of the cDC lineage. These patterns remained when the dataset was split into ileum and colon LP (**Fig. S4B**).

To further investigate the identity of the cells in clusters 1 and 2, we correlated their gene expression with a published single cell dataset of human BM that contained different stages of hematopoietic cell development[49]. This analysis showed that cluster 1 had some overlap with mature cDC1 and cDC2 in human BM, as well as with hematopoietic stem cells, lympho-myeloid precursors and early, but not late myeloid progenitors or monocytes (**Fig. 4H**). Together, these findings support the hypothesis that cluster 1 contains precursors of myeloid origin which are part of the cDC lineage, but are not yet committed to any of the mature subsets. Interestingly, the committed pre-cDC1 in clusters 2/2A retained some correlation with the lympho-myeloid progenitors and early promyelocytes, as well as showing more overlap with mature cDC1 than cluster 1. Thus clusters 2/SA may be more advanced in the trajectory towards cDC1 commitment (**Fig. 4H**). Similarly, clusters 6 and 7 retained correlation with HSC & MPP but correlated more strongly with cDC2 in bone marrow than cluster 1, consistent with them being pre-cDC2 (**Fig. 4H**). In contrast, the putative pre-cDC3 clusters 2B and 3B did not correlate with any of the progenitor populations in the bone marrow dataset (**Fig. 4H**).

The putative trajectories derived from this analysis (**Fig. 4I**) were confirmed by analysis showing that the expression of cDC subset signature genes increased progressively across different clusters (**Fig 4J**). Thus, expression of the cDC1-related genes *BATF3*, *IRF8, XCR1, CLEC9A* and *CADM1* showed progressive increases that suggested a trajectory from cluster 1 to mature cDC1 via clusters 2A, 3A, 4 and 5. In parallel, these clusters showed downregulation of the cDC2 associated genes *LTB, CD207, CD1C* and *IRF4*; there was no clear pattern of regulated expression of the cDC3 genes *CD163, CD14, S100A9, C1QA* and *MERTK* (**Fig 4J**). Analysis of the same genes showed similar evidence that that there was progressive expression of cDC2-associated genes from cluster 1 via clusters 6 and 7 to mature cDC2, and of cDC3-associated genes from cluster 1 to mature cDC3 via clusters 2B, 3B and 8 (**Fig 4J**). Together, these data indicate that committed precursors of all three cDC lineages exist in the human intestinal LP, and there may also be a rare population of non-committed cDC precursors.

Characterising intestinal LP monocyte and macrophage populations

Murine intestinal monocytes develop into macrophages along a 'waterfall' of phenotypic intermediates [50–53]. To investigate whether a similar monocyte 'waterfall' exists in the human intestinal LP, ileum and colon LP cells identified as Mono/Mac in Figure 1 were isolated and trajectory analysis was performed with tSpace [34]. The tSpace principal components were used for trajectory-based re-clustering and tUMAP embedding, which resulted in 11 Mono /Mac clusters (M1-M11) (**Fig. 5A**). DEG analysis of the trajectory-based clusters identified high levels of monocyte-associated genes, such as *S100A9*, *FCN1* and *VCAN* [28,54], in cluster M1 (**Fig. 5B and C**). To investigate the downstream developmental trajectories, pseudotime scoring was performed on the tSpace trajectories using the M1 monocyte cluster as the pseudotime starting point and averaging all trajectories from there. This demonstrated a pseudotime trajectory from cluster M1, via M2-M3, then M4-M6, and ending at M7, M8, M9, and M10 (**Fig. 5D**). Cluster M11 was considered too small to analyse by pseudotime with confidence.

To investigate heterogeneity of the monocyte-derived clusters, DEG analysis showed that clusters M2 and M3 expressed intermediate levels of *S100A9*, *VCAN*, and *ITGAX* and low levels of *C1QC*, which we classified as early intermediate cells; clusters M4-6, which lacked expression of *S100A9* and *FCN1* and expressed intermediate levels of *ITGAX* and *C1QC*, were classified as late intermediate cells; and cluster M7 and M8 that expressed high levels of *CD209* and *C1QC*, were consistent with mature macrophages [54,55] (**Fig. 5B and C**). The two small clusters M9 and M10 shared some features with mature macrophages, including high expression of MHCII and C1Q genes, but expressed low levels of *CD209* and *CD163* (**Fig. 5B and C**). Finally, the smallest cluster, M11, expressed very high levels of cell-cycle associated genes including *MKI67* and *KIAA0101*, indicating that these represented a population of proliferating cells (**Fig. 5C**).

In recent single-cell data, murine monocytes were found to develop into resident macrophages along two distinct paths depending on the inflammatory state of the tissue[53]. To investigate whether this might be occurring with human intestinal Mono/Mac, signalling pathway activity was analysed using PROGENy [56] (**Fig. 5E**). PROGENy analysis showed activation of NFκB and TNFα signalling pathways in the monocyte cluster M1 and in the early intermediate cluster M3, but not in the parallel early intermediate cluster M2 (**Fig. 5E**). Consistent with this, cluster M3 expressed high levels of pro-inflammatory cytokines including *IL1B, CXCL8* (IL8), and *CCL4* (MIP-1β), as well as the signalling protein *NFKBIA* (**Fig. 5C** and **F**). This dichotomy extended to late intermediate as well as mature macrophage clusters, with the late intermediate cluster M5 and the mature macrophage cluster M7 showing indications of enhanced NFκB and TNF signalling, and enhanced expression of higher pro-inflammatory cytokines, compared with other late intermediate and mature macrophage clusters (**Fig. 5C and F**). Clusters M3, M5 and M7 also showed higher activation of MAPK and EGFR pathways (**Fig. 5E**). In contrast, the late intermediate clusters M4 and M6, and the mature macrophage clusters M8 and M9, all showed evidence of increased activation of the JAK-STAT and WNT signalling pathways (**Fig. 5E**). Interestingly, three of the 4 mature macrophage clusters (M7-M9) showed evidence of responding to TGFβ signalling (**Fig. 5E**), consistent with previous work in mice [55]. These clusters also showed evidence of p53 pathway activation, which is involved in multiple aspects of cell function including cell-cycle arrest, and which was not active in the proliferating cluster M11 (**Fig. 5E**). In contrast, M11 showed evidence of responding to oestrogen (**Fig. 5E**), consistent with findings that oestrogen may drive proliferation of macrophages [57].

As well as showing a putative pro-inflammatory phenotype, clusters M5 and M7 expressed the stimulus-response genes *FOS, JUN,* and *EGR1* [58] (**Fig. 5C and G**), and other transcription factors including *KLF2, KLF4* and *KLF6* (**Fig. 5C, G and H**). Using DoRothea [59,60], we found evidence of Nuclear Factor kappa-light-chain-enhancer of activated B cells (NFκB) and Activator Protein-1 (AP-1) activity in clusters M1, M3, M5, and M7, and activity of Activating

Transcription Factor (ATF) in clusters M5 and M7 (**Fig. S5A**). In contrast, clusters M4, M6, M8 and M9 showed inferred activity of the transcriptional repressor RE1-Silencing Transcription factor (REST), and the MHCII promoter-associated regulatory factor X5 (RFX5) (**Fig. S5A**). Interestingly, cluster M6 also expressed the stimulus-response genes and KLF genes (**Fig. 4C, G and H**), but this cluster showed activation of the JAK-STAT and WNT pathways instead of the NFκB, TNFα MAPK, and EGFR pathways (**Fig. 5E**), it expressed low levels of pro-inflammatory cytokines (**Fig. 5F**) and showed particularly low activation of *JUN* and *FOS* (**Fig. S5A**). Thus, clusters M3, M5, and M7 seem to share a stimulus-response and pro-inflammatory profile, while clusters M2, M4, and M6 seemed to share a less-inflammatory profile.

Clusters M5, M6, M7, and M10 also expressed the transcription factor *KLF2*, which has been implicated in efferocytosis, and clusters M6 and M7 expressed the efferocytosis-associated gene *STAB1*, suggesting they may be specialised for uptake of apoptotic cells (**Fig. 5H**)[61]. Two other apoptotic uptake-associated genes, *AXL* and *MERTK*, were expressed by clusters M4-M9 and M7-M9, respectively (**Fig. 5H**). Thus, the late intermediate clusters M4-M6 and mature macrophages M7-M9 may be involved in apoptosis, but, in line with a recent study [62], there is heterogeneity in the expression of genes involved.

To understand how the intestinal macrophage subsets (M7-10) may be related to previously described tissue resident macrophage subsets, we further explored genes differentially expressed by these subsets. The macrophage clusters M7, and, to a lesser extent, M8, were both enriched in gene signatures of perivascular macrophages [63], including *LYVE1, SIGLEC1* (CD169)*, FOLR2,* and *MAF* (**Fig. 5C and I**), consistent with a previous report that the majority of mature macrophages in the LP are found in a perivascular niche [64]. $CD169^{+}$ macrophages have also been observed in mouse intestine [65], where they seem to play a role in surveillance of the vasculature [66].

The small $CD163^{lo}$ mature macrophage cluster, M9, expressed relatively high levels of *CD4, C2*, and *ADAMDEC1* (**Fig. 5C and J**), and in this respect resembled the non-monocyte-

derived, self-maintaining resident macrophages described in the murine intestine [67–69]. These cells also highly expressed matrix metalloproteinases, including *MMP9* and *MMP12* (**Fig. 5C** and **J**). Expression of other MMPs, MMP2 and MMP14, was previously reported as a signature of self-maintaining murine intestinal macrophages [68]. The other *CD163*^{lo} macrophages, cluster M10, expressed genes associated with iron sequestration (*FTL*), vitamin A transport (*RBP1*), microglia (*PLP1*, *PMEPA1, GPM6B*) and fibrosis (*SPARC*), as well as the adipocyte-associated genes *ADIRF* and *SCD* (**Fig. 5C**), and showed poor activation of most of the response pathways (**Fig. 5E**) and transcription factor activity (**Fig. S5A**), so did not have a clear phenotype. They did however resemble adipose tissue macrophages in several respects[70,71], including activity of the TRAIL pathway (**Fig. 5E**) [72].

Finally, the genes differentially expressed in each subset between paired ileum and colon LP samples was assessed. There were few genes differentially expressed between colon and ileum for the earliest clusters, M1 and M2 (**Fig. 5K**). Clusters M3, M4, M5, M7, and M8 all showed more genes differentially expressed in the colon than in the ileum (**Fig. 5K**). This may reflect increased activation or cytokine production in response to the higher microbial load of the colon. Mature clusters M6, M9, and M10 did not show this pattern, which suggests they may be found in different micro-anatomical locations or may be transcriptionally repressed.

In summary, ileum and colon LP Mono/Mac are highly heterogeneous, encompassing a differentiation spectrum from monocytes (M1) through early (M2, M3) and late (M4, M5, M6) intermediate cells, to mature macrophages (M7, M8), as well as small populations resembling self-maintaining macrophages (M9), adipose tissue-like macrophages (M10), and proliferating cells without a clear identity (M11). Monocytes appear to develop into macrophages via both a pro-inflammatory route (M1- M3- M5- M7) and a parallel less-inflammatory route (M1- M2- M4- M6- M8).

Analysis of Mono/Mac subsets by surface marker expression

To identify surface antigens that could be used to distinguish the different stages of intestinal monocyte development by flow cytometry, colonic $CD14^{+}CD1c^{lo}$ cells were screened for surface antigens which showed heterogeneous expression, using the Legendscreen containing 288 antibodies. Using this approach, colonic $CD14^{+}CD1c^{lo}$ cells were found to express heterogenous levels of CD11a (LFA1/*ITGAL*), CD11b (*ITGAM*), CD11c (*ITGAX*), CD13 (*ANPEP*), CD18 (*ITGB2*), CD39 (*ENTPD1*), CD50 (*ICAM3*), CD55 (DAF), CD81, CD89 (FCAR), CD90, CD163, CD206 (mannose receptor/*MRC1*), CD274 (PDL1), CD276, and CD301 (CLEC10A) (**Fig. S5B**), indicating that these could be used as surface markers to distinguish distinct subsets. Of these, CD11c, CD11a, CD206, and CD55; together with CD14 and CD1c, were used to construct a CITE-seq panel to determine their levels of expression on clusters M1-M11 (**Fig. 6A**). CD55 was highly expressed by the monocyte cluster M1, expressed at intermediate levels by the early intermediate clusters M2 and M3, and was not expressed by the late intermediate clusters M4 and M5, or mature Mac (**Fig. 6A**). CD11a was highly expressed by monocytes (M1) and early intermediate clusters (M2 and M3) and was expressed at intermediate levels by the late intermediate clusters (M4 and M5) (**Fig. 6A**). In contrast, CD206 expression appeared at the late intermediate stage of development and was maintained on mature macrophage clusters but was not present on the small M9 and M10 clusters (**Fig. 6A**). CD1c was poorly expressed by monocytes and macrophages but was expressed at moderate levels by all the intermediate clusters (M2-M6) (**Fig. 6A**), consistent with these clusters also expressing other cDC2-associated genes such as *CD1E*, *CLEC10A*, and *FCER1A* (**Fig. 5C**). CD11c was expressed at moderate levels by monocytes and increased during their transition to the intermediate clusters M2-M6; it was then expressed at moderate-to-high levels by clusters M8 and M9 and at low levels by M7 and M10 (**Fig. 6A**). Cluster M7 was distinguished from other mature macrophages by particularly high CD14 expression, while M9 and M10 were uniquely $CD206^{-}$ $CD11c^{+}$ and $CD206^{-}$ $CD11c^{-}$ respectively (**Fig. 6A**). In summary, the expression of surface markers correlated with gene expression and

confirmed the heterogeneity and likely developmental stages within intestinal monocytes and macrophages.

Site-specific alterations in mono/mac composition

To identify the different Mono/Mac subsets by flow cytometry, CD14$^{int/+}$ cells were divided into two fractions comprising CD55^{+} CD206^{-} and CD55^{-} CD206$^{+/-}$ cells (**Fig. 6B**). The CD55^{+} CD206^{-} cells contained HLA-DR^{int} monocytes and HLA-DR^{+} early intermediate cells. The CD55^{-} CD206$^{+/-}$ fraction was first divided into CD11a^{+} CD163^{-} and CD11a^{-} CD163$^{int/hi}$ populations representing late intermediate and mature macrophages respectively (**Fig. 6B**). Finally, the CD11a^{-} macrophages were divided into CD163^{int} CD14^{int} and CD163^{hi} CD14^{hi} subsets. Although CD14 and CD163 showed a continuous spectrum of expression, the proportions of the subsets based on these markers were clearly different between intestinal sites, with significantly more CD163^{int} CD14^{int} macrophages being found in the ileum than in the colon (**Fig. 6C**).

Inflammation-associated changes to mono/mac subset frequencies

To assess the impact of intestinal inflammation on the relative abundance of MNP subsets in the intestinal LP, we analysed biopsies taken during routine endoscopic procedures for initial assessment or disease surveillance. Biopsies were taken from multiple sites in each patient, and the extent of inflammation was classified as either quiescent, mild, or moderate, before analysis using flow cytometry. Because cell numbers were lower in biopsies than in resection samples, CD55^{+} CD206^{-} HLA-DR^{int} monocytes could not be reliably distinguished from CD55^{+} CD206^{-} HLA-DR^{+} early intermediate cells, so both were collected together in the CD55^{+} CD206^{-} 'Mono & Early Intermediate' gate (**Fig. 6D**). Increased severity of endoscopic inflammation was associated with an increase in both combined monocyte/early intermediates and in late intermediate cells in the colonic LP (**Fig. 6D**). The frequency of late intermediate cells was slightly increased in quiescent IBD samples compared with non-IBD controls, and

was equivalent between mild and moderate IBD inflammation. However, the frequency of combined monocyte/early intermediate cells was only increased when inflammation was present (**Fig. 6D**). Human intestinal monocyte-derived cells in IBD therefore act similar to those of the mouse, with an increase of monocyte and intermediate cells, and a reduction of mature macrophages, only during active inflammation.

Discussion

MNPs play critical roles in tolerance, immunity and inflammation at the intestinal barrier and characterizing the individual members of this family will be crucial for understanding immune responses in the gut during health and disease. Although there has been considerable recent research on deciphering MNP subset identities in mice and humans, this has been complicated by their overlapping phenotypes [7,9,25,28,73–77].Here, we used scRNA-seq to carry out unbiased characterisation of MNPs in the human intestinal LP, using a relatively non-stringent gating strategy to enrich cells by FACS sorting before sequencing. This allowed us to capture all $CD3^-$ $CD19^-$ MNP subsets, including those which expressed low levels of HLA-DR such as pre-cDC and monocytes.

This approach revealed clear clusters of pDC and cDC1, together with heterogeneous and transcriptionally overlapping groups of monocytes, macrophages and cells with the appearance of cDC2 and the recently defined population of cDC3. Resolution of these cells into distinct clusters required high resolution analysis, together with regression out of shared stimulus-response-associated genes, MHCII and cell-cycle-associated genes. Using this strategy in an iterative process involving both biased and unbiased approaches, we were able to define populations expressing signature genes of classical monocytes, macrophages and cDC2. In addition, we found a group of cells which expressed many monocyte-associated genes, but also genes normally associated with cDC2, including *CD1C*. Human MNPs with these characteristics have historically been identified as monocyte-derived DC ("moDC") [78] and DC-like cells can be derived from human monocytes *in vitro* [79]. However, recent scRNA-seq analyses have revealed a unique population of this kind, which has been termed cDC3 [7,25,28,30]. By extending our scRNA-seq findings with flow cytometry and CITE-seq, we confirmed previous reports that two subsets of $CD1c^+$ intestinal cDC2-like cells could be defined based on their expression of CD207 vs. CD11a, of which one is likely to be cDC3 [80–83].

The cells we identified as cDC3 not only expressed a number of monocyte-associated genes, but like monocytes, they were enriched in the colon compared with the ileum and

expanded relative to cDC2 during active IBD. However it seems unlikely that they are derived from monocytes, as their numbers were significantly reduced in IBD patients who had been treated with anti-integrin α4β7 antibody (Vedolizumab), an intervention which we and others have shown does not affect monocyte proportions in the LP [84]. In contrast, α4β7 has been shown to play a role in the homing of cDC to the murine intestine [85] and our Vedolizumab-treated patients showed reduced proportions of cDC2. These findings have potential implications for the use of anti-integrin α4β7 as a long-term IBD therapeutic, as although blocking the migration of pro-inflammatory cDC activity into the gut could be beneficial, cDC are also crucial for the induction of tolerance to the microbiota in the gut [86–88].

As shown in other tissues, intestinal cDC2 and cDC3 shared many transcriptional features, raising the possibility that these represent different states of maturation or activation of the same lineage. However recent work has suggested that cDC3 develop along a developmental trajectory distinct from cDC1 and cDC2 [25]. In support of this possibility, we identified distinct populations of potential cDC precursors in the LP, including pre-cDC1, pre-cDC2, and pre-cDC3. These clusters showed low expression of MHCII genes, high expression of cell-cycle-associated genes and mixed expression of other hematopoietic and pre-cDC-associated genes. However, they also expressed a variety of the DEGs associated with the individual cDC subsets, indicating that these were already committed to one of the lineages. As well as these MHCII^{lo} clusters, there were other proliferating, closely related clusters which also expressed cell-cycle genes to different extents, which could be developmental intermediates or immature cDC. These findings are consistent with work in humans and mice showing progressive differentiation of cDC subsets from early myeloid progenitors via increasingly committed precursors, and suggest that a variety of these different cDC precursors may be present in the normal human intestinal LP. However, confirmation of this idea will require formal exploration of progenitor capacity by appropriate methods *in vitro* or *in vivo*.

In the mouse intestine, monocytes differentiate into macrophages through intermediate stages in what is known as the macrophage 'waterfall' [50,51,55,73]. It has recently been suggested that a similar monocyte-macrophage differentiation continuum may also exist in the human ileal mucosa [54]. In support of this, we identified 11 clusters of Mono/Macs which included classical monocytes, three clusters of mature macrophages, and five clusters which appeared to be intermediate between them. These were defined as intermediates based on their co-expression of genes associated with both monocytes and mature macrophages and this was confirmed by our pseudotime analysis. CITE-seq and flow cytometry analyses suggested a continuum in which $CD55^{+}$ $CD206^{-}$ HLA-DR^{lo} monocytes differentiated into $CD55^{+}$ $CD206^{-}$ HLA-DR^{+} early intermediates, followed by $CD55^{-}$ $CD11a^{int}$ $CD206^{+}$ late intermediates and ultimately, two populations of $CD11a^{-}$ $CD14^{int}$ $CD163^{int}$ and $CD14^{hi}$ $CD163^{hi}$ mature macrophages. However, a complete definition of these cells as developmentally related will depend on further analysis, such as tracing of donor cells in transplant patients [54].

Our trajectory analysis suggested that monocytes entering the LP may give rise to two parallel pathways of macrophage development. One pathway (M3-M5-M7) was characterised by clusters expressing higher levels of pro-inflammatory cytokines and stimulus response genes, as well as inferred activity of NFκB/TNFa/EGFR signalling pathways and NFκB/ATF/AP-1 transcription factors. The other pathway (M2-M4-M6-M8)) included clusters which were phenotypically and transcriptionally similar to the first, but did not express proinflammatory cytokines and did not show evidence of NFκB/TNFa/EGFR signalling or NFκB/ATF/AP-1 transcriptional activity. Instead, these clusters showed inferred activity of JAK-STAT and WNT pathways, and of the transcriptional repressor REST. The transcriptional differences between subsets were reflected in heterogeneity in expression of surface markers, efferocytosis genes and cytokines, indicating potential functional differences between the subsets. The reasons for these distinctive properties are unclear, but they could reflect monocyte-macrophage differentiation occurring within different anatomical or immunological niches in the LP. For instance, work in mice has shown that macrophages in the LP and

muscularis layers of the intestine may be distinguished by their pro- or anti-inflammatory phenotype respectively [13,89]. Although our isolation technique allows us to exclude the muscularis as a source of macrophages, a population of $CD169^{+}$ macrophages with pro-inflammatory properties has also been described near the crypt base of the LP [65]. Furthermore, there is increasing evidence from other tissues that monocyte-derived resident macrophages can acquire very different properties depending on the exact niche which they locate to, such as close to nerves, blood vessels or airways [11]. Whether similar anatomical factors constrain the development of these heterogeneous populations of macrophages identified in human intestinal LP remains to be elucidated. In addition to these putative monocyte-derived macrophages, we also observed a small cluster of macrophages expressing low levels of *CD163* (M9) which transcriptionally resembled self-maintaining resident macrophages. Although rare in the intestine, a population of this kind has been found in mouse intestine, where it expresses CD4 and TIM4[67]. Analogous macrophages in other tissues are thought to play critical roles in tissue homeostasis, but confirmation of the origin and functions of this population in human intestine awaits validation using other methods. One other cluster of mature macrophages which we could identify was M10, which had similarities to adipose tissue macrophages, but again, these cells will require further investigation to clarify their role fully.

In addition to mature macrophages and classical $CD14^{hi}$ monocytes, we also detected a small population of non-classical monocytes expressing *FCGR3A* (CD16) in our LP preparations. Non-classical monocytes are generally thought to function as MNP patrolling the vasculature [90] and we identified substantial numbers of vascular and lymphatic endothelial cells in our preparations. Although this suggests that at least some of the non-classical monocytes we found may have been within the vasculature, there is also evidence that non-classical monocytes can migrate into tissues including the intestine [84]. Thus, the exact location of these non-classical monocytes remains an open question.

Accumulation of pro-inflammatory, $CD14^{+}$ $HLADR^{int}$ cells is a characteristic feature of active inflammatory bowel disease [16,91] and recent studies have reported single-cell analyses

of human intestinal MNP in IBD [16,92,93]. However, the number of MNP was small in these studies and there was no consensus on their exact nature, nor how they correspond to putative counterparts in the healthy gut. Here we confirmed the accumulation of $CD14^{+}$ $HLADR^{int}$ monocytes and monocyte-derived intermediates in both colonic and ileal LP during active IBD, together with lower proportions of mature $CD14^{int}$ and $CD14^{hi}$ macrophages. The exact nature of the monocyte infiltrate appeared to depend on the degree of inflammation, as there was a bias towards late intermediates during milder inflammation, whereas monocytes and early intermediates dominated the response in more severe inflammation. The failure of monocytes to develop into mature macrophages in inflamed intestine was initially thought to reflect a block during the intermediate stages of differentiation [50,51]. However, more recent work in mice has suggested that monocytes begin to follow an alternative pathway of development as soon as they enter the inflamed intestine [53]. Our findings support a model in which monocytes immediately respond to different environments as soon as they enter the gut, and as discussed above, this may also be the case in the steady-state intestine. The novel transcriptional patterns and developmental trajectories we have defined here should provide an important resource to interrogate the development and biology of intestinal MNP in more detail.

In summary, our single-cell studies of MNP in carefully isolated LP emphasise the heterogeneity of these cells in both health and disease, and show that this may also differ depending on the gut segment being examined. By identifying novel genetic and phenotypic markers, our work should provide useful information for further studies of how monocytes, macrophages and DC subsets might contribute to regulation of immune responses in the intestine.

Acknowledgments

We thank all patients and collaborating staff at Herlev hospital, and in particular the Gastroenterology Team (Department of Pathology) for help in providing tissue samples. Sequencing was performed at the National Genomics Infrastructure (NGI) and Science for Life

Laboratory SNP&SEQ Technology Platform in Uppsala (supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation). We thank Dr J Rizk (Copenhagen University) for valuable guidance regarding signalling pathways. This work was supported by a grant from the Lundbeck foundation (R155-2014-4184), Denmark, to WW.A and S.B, the Danish Research Council (Sapere Aude III senior researcher grant 1331-00136B to W.W.A), and the Swedish Medical Research Council (2017-02072) and the Swedish Cancerfonden (18 0598) to W.W.A.

Materials and Methods

HUMAN SUBJECTS

Resection samples were taken with informed consent from patients undergoing colorectal cancer surgery, as approved by the Videnskabsetiske Komité for Region Hovedstaden, ethical permission H-3-2013-118. Resection tissues were taken at least 20 cm distant from any tumours present.

Biopsy samples were obtained with informed consent from adult IBD patients attending routine colonoscopy for initial disease surveillance or ongoing disease assessment (Table S2 for anonymised patient information) at the Western General Hospital, Edinburgh, UK, after informed consent under existing approvals (REC:19/ES/0087). A diagnosis of IBD was made using standard criteria (Leonard-Jones), with all patients part of the Lothian IBD registry [94]. Endoscopic assessment of disease severity for each biopsy site was performed by clinicians using the Mayo endoscopic subscore for UC or the simple endoscopic score (SES-CD) for CD. Each biopsy site was classed as quiescent, mild, moderate or severe endoscopic activity to enable comparison across IBD subtypes, and 3-5 biopsies were taken for analysis per site.

Tissue processing

Samples were processed as described previously [23]. Briefly, resection samples were taken at least 10 cm distant from tumours, where present. The muscularis externa was removed using curved surgical scissors and the remaining tissue was incubated in a 370rpm shaking incubator twice for 10 min at 37°C in RPMI-5 (RPMI/5% FCS/1% penicillin and streptomycin) containing 4 mM DTT, to remove mucus. Macroscopically visible submucosa (SM) was trimmed away using scissors. The mucosa was then separated from the underlying SM under a stereo microscope using forceps. Epithelial cells were removed from the remaining mucosa by shaking four times in $Ca2^{+}$ and $Mg2^{+}$ - free HBSS containing 1% penicillin and streptomycin and 5 mM EDTA at 37°C, for 10 min each time. Any remaining ILF present in the mucosa were then dissected out using a scalpel under a stereo microscope with a transmitted light source.

The remaining GALT-free LP was cut into 2-4 mm^2 pieces prior to digestion. A single-cell suspension of isolated LP was generated by shaking tissues at 370 rpm in a 37^oC incubator in RPMI-5 containing 30 µg/ml DNase and 5 mg/ml collagenase D or 2.5 mg/ml Liberase TM for 45 min, at 37^oC. The suspension was passed through a 100 µm filter before washing twice in fresh media and centrifuging to form a pellet.

Flow cytometry

Cell suspensions were stained with the antibodies indicated (Table S3) in Brilliant stain buffer (BD Biosciences) containing 4% normal mouse serum according to standard techniques. The cells were stained with 7-AAD and analyzed on an LSR Fortessa 2 (BD Biosciences) using Flowjo software (Treestar).

For the Legendscreen assay (Biolegend), each antibody was first resuspended in 35 µl of FACS buffer (PBS with 5% FCS and 0.05% sodium azide). Up to $300x10^6$ cells were stained with the antibody backbone panel (Table S4), washed and resuspended in PBS with 5% FCS. 40 µl of cell suspension was aliquoted into each well of V-bottom 96-well plates containing 10 ul of the PE-conjugated antibody per well. The cells were analysed using an LSR Fortessa 2 with High Throughput Sampler (BD).

Magnetic enrichment of MNP

Single-cell suspensions to be used for single-cell sequencing were enriched for HLA-DR^+ cells using anti-HLA-DR microbeads (Miltenyi Biotec), as per the manufacturer's instructions. Briefly, cells were counted then centrifuged and resuspended in 80 µl cold MACS buffer (PBS with 0.5% BSA and 2mM EDTA) and 20 µl HLA-DR microbeads, per $1x10^7$ total cells. Cells were stored on ice for 15 minutes then washed in 1 ml MACS buffer per $1x10^7$ cells. Cells were centrifuged and resuspended in 500 µl fresh MACS buffer before adding to an LS MACS column. The flow-through was discarded and the HLA-DR^+ cells were washed through the column with 5 ml fresh buffer.

Cell sorting

Cells were counted, pelleted, and stained for flow cytometry as above. The cells were stained with anti-human CD45-BV421, CD3-PECF594, CD19-PECF594, HLA-DR-AF700, and 7-AAD was added before sorting to exclude dead cells. Viable $CD45^+$ $HLA\text{-}DR^{int/+}$ $CD3^-$ $CD19^-$ cells were sorted on a FACSMelody sorter (BD) into MACS buffer, before being counted and re-suspended in PBS with 0.4% BSA.

CITE-seq

For some samples, cells were first stained with barcode-labelled antibodies together with the antibodies used for sorting, before running the 10x protocol. The Totalseq-A antibodies used were anti-human CD1c (clone L161), CD14 (clone M5E2), CD11a (clone TS2/4), CD55 (clone JS11), CD5 (clone UCHT2), CD206 (clone 15-2), CD209 (clone 9E9A8), and CD11c (clone S-HCL-3), all from Biolegend.

Single-cell 10x protocol

Suspensions of sorted single cells were subjected to droplet-based massively parallel single-cell RNA sequencing using the Chromium Single Cell 3′ Reagent Kit v3 following the manufacturer's instructions (10x Genomics). The 10x Chromium Controller generated nanoliter-scale Gel Bead-In Emulsions (GEMs) droplets, where each cell was labeled with a specific barcode, and each transcript labeled with a unique molecular identifier (UMI). After reverse transcription (55°C for 45min and 85 °C for 5min), the GEMs were broken down and the barcoded cDNA was purified with Dynabeads MyOne Silane beads (Thermofisher). The cDNA was amplified by PCR with 10x genomics and ADT additive primer (98 °C for 3min; 12 cycles of 98 °C for 15sec, 67 °C for 20sec and 72 °C for 1; 72 °C for 1min, end at 4 °C). The products were size separated with SPRIselect beads (Beckman Coulter) into fragments <

300nt containing antibody derived tags (ADTs) and > 300nt containing cDNAs derived from cellular mRNA.

For sc-RNA libraries, 50 ng of amplified cDNA was used for final library preparation, consisting of enzymatic fragmentation, end repair, A-tailing, adaptor ligation and sample index PCR as per the manufacturer's instructions.

For epitope sequencing (CITE-seq), 2ul of ADTs were PCR amplified with Illumina compatible SI PCR and TrueSeq Small RNA RPIx primers (15 cycles). Final libraries were purified using SPRIselect beads. Quality and quantity of the final libraries (sc-RNA and ADT) were measured using the Agilent 2100 Bioanalyzer equipped with High Sensitivity DNA chip. Libraries were pooled according to the starting number of cells of each run (10% ADT + 90% sc-RNA).

SI-PCR primer:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC*T*C

ADT additive primer: CCTTGGCACCCGAGAATT*C*C

RPI1:CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCG AGAATTC*C*A

RPI2:CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCG AGAATTC*C*A

RPI3:CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCCG AGAATTC*C*A

RPI4:CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCCG AGAATTC*C*A

* indicates a phosphorothioate bond

Sequencing

Illumina sequencing was carried out at the Genomics Core Unit: Center of Excellence for Fluorescent Bioanalytics (KFB, University of Regensburg, Germany) and at the SNP&SEQ Technology Platform in Sweden which is part of the National Genomics Infrastructure (NGI), funded by the Swedish Council for Research Infrastructures and Science for Life Laboratory. Libraries were sequenced using HiSeq, NextSeq and NovaSeq systems (300 cycles), targeting a minimum of 30,000 read pairs per cell for sc-RNA and 3000 read pairs per cell for ADTs.

Bioinformatic analysis

Sequencing data was pre-processed and aligned with CellRanger (version 2.2.0 for the first patient samples and version 3.1.0 for all other samples)[95,96]. The sequencing data from the samples stained with TotalSeq antibodies was processed with cite-seq count [97]. Each sample was then read into a Seurat (version 3.1.5) [98] object in R (versions 3.5.1/4.0.1) [99] and processed by removing cells with exceptionally low or high UMI and gene counts (debris and doublets). The lower thresholds were set from 500-1000 genes per cell and the upper threshold were set from 3000-6000. Furthermore, cells with mitochondrial gene content > 10% were also removed. The thresholds were set by studying UMI and gene counts, and mitochondrial gene content according to current best practise [100]. The debris removed from each sample was used as empty droplet information to normalize the protein level in the individual samples together with the isotype controls using the tool Denoised and Scaled by Background (DSB)[41]. The normalized protein data was added to each of the relevant samples.

The scRNA data for each sample was normalized and cell cycle gene modules were calculated using the Seurat CellCycleScoring function and additionally a gene module representing cell cycle genes from the tool ccRemover by summing the raw counts of these genes per cell divided by the total number of reads per cell[101]. The top 3000 most variable genes were identified with the selection method vst. After this initial data processing, all the samples were integrated with Seurat anchor integration and gene expression was scaled while regressing out effect of cell cycle, UMI counts, and mitochondrial gene content. The merged

data were dimensionality reduced with PCA and the 15 first PCs were chosen for downstream analysis. A shared nearest neighbour graph was constructed and used to cluster the data with Louvain clustering. The PCs were also used to dimensionality reduce the data further with UMAP for visulization purposes. The data were reclustered on described populations after removal of contaminating cell types and new UMAPs calculated.

The pseudobulk for heatmaps and PCA was performed with the Seurat AverageExpression function. The PCA on pseudobulk of clusters was performed only on genes from published gene signatures of blood DC2, blood monocytes and in vitro mo-Macs[33]. Differential gene expression was calculated with Seurat FindMarkers for comparisons between specific groups or FindAllMarkers for DEGs for all clusters both using the standard non-parameteric Wilcoxon rank sum test.

Trajectory inference was performed with tSPACE on PC spaces of indicated populations [34]. The outputs were further dimensionality reduced with UMAP[102,103] to 2 and 3 dimensions with distance metric set to Pearson. Clustering was performed with Louvain clustering for Seurat on the tSPACE PC output as input. RNA velocity estimates were calculated for T=1 and only included genes with splicing information also present in the variable genes and only on cells of interest (e.g. precursors). The information was embedded on top of 2D UMAPs from the tSPACE trajectories using n=400, scale=sqrt, grid.n=50 and arrow.scale=2 [48].

Variable genes overlapping between healthy bone marrow data from Triana et al [49] and LP data were identified and average expression of these were calculated for each data set. The pseudobulk LP and BM data was then correlated using pearson.

QUANTIFICATION AND STATISTICAL ANALYSIS

Flow cytometry analyses were performeddata were analysed using in Prism software (GraphPad). Details for, with the statistical tests used in each experiment can be foundbeing shown in the relevant figure legends. Statistical significance was defined as *p<0.05, **p<0.01,

***p<0.001. Statistical analysis of sequencing data was performed in R as described in the bioinformatics section above.

Figure legends

Figure 1. Untangling intestinal monocytes and macrophages from transcriptionally similar dendritic cells. (**A-C**) scRNA-seq analysis of 28,758 combined ileal and colonic LP cells, identified as MNP clusters by high average expression of MHCII genes. n = 6 colonic and 4 paired ileal CRC patient samples. (**A**) scRNA-seq experimental outline. (**B**) UMAP of $HLADR^+$ cell clusters with known cell identities. Coloured by Louvain clustering. (**C**) UMAP of normalized gene co-expression of *CD1C* and *CD14* co-expression. Dashed line encompasses MNP subsets which are not readily identifiable as pDC, cDC1, or non-classical monocytes. (**D**) Representative flow cytometry gating strategy to identify $CD1C^+/CD14^+$ MNP in colon LP. (**E-G**) scRNAseq data of 26,578 combined ileal and colonic LP cells identified by expression of *CD1C* and/or *CD14*. n = 6 colonic and 4 paired ileal CRC patient samples. (**E**) Curated heatmap of re-clustered $CD1C^+/CD14^+$ cells showing scaled gene expression of macrophage/DC signature genes averaged per cluster. (**F**) UMAP of $CD1C^+/CD14^+$ cells re-clustered at high resolution from clusters within dotted line in Fig. 1C. Dashed line represents major division between clusters identified as Mono/Mac (top) and cDC2/3 (bottom). (**G**) PC1 and PC2 from pseudo-bulk based PCA of $CD1C^+/CD14^+$ clusters using signature gene lists from blood[78] cDC2, classical blood monocytes, and monocyte-derived macrophages

Figure 2. Single-cell characterisation of human intestinal cDC2 and cDC3. (**A-I**) scRNA-seq of mature DC. n = 6 colonic (red) and 4 paired ileal (turquoise) CRC patient samples. (**A**) Flattened 3D tspace UMAP (tUMAP) plot of cDC clusters grouped into cDC1 (pink), cDC2 (purple), cDC3 (yellow), and an ambiguous cDC2/3 group (light grey). (**B**) Heatmap of scaled gene expression of top 25 DEG expressed by mature cDC1, cDC2, and cDC3 subsets. (**C**) Manually curated heatmaps of scaled DEG expression classified by the Mouse Genome Informatics group as GO terms associated with transcription factors, cytokines and antigen

processing and presentation, respectively. (**D**) TF activity inferred by expression of TF target genes in DC subsets using the DoRothea package. (**E**) Signalling pathway activity in DC subsets inferred from pathway gene activity using the PROGENy package. (**F-G**) Flattened 3D tUMAP plots overlayed with expression of (**G**) *CXCL11* and (**F**) *LAMP3.* (**H** and **I**) DEG enrichment in cLP vs siLP for (**H**) each subset and (**I**) overlap of these segment-specific DEG between subsets.

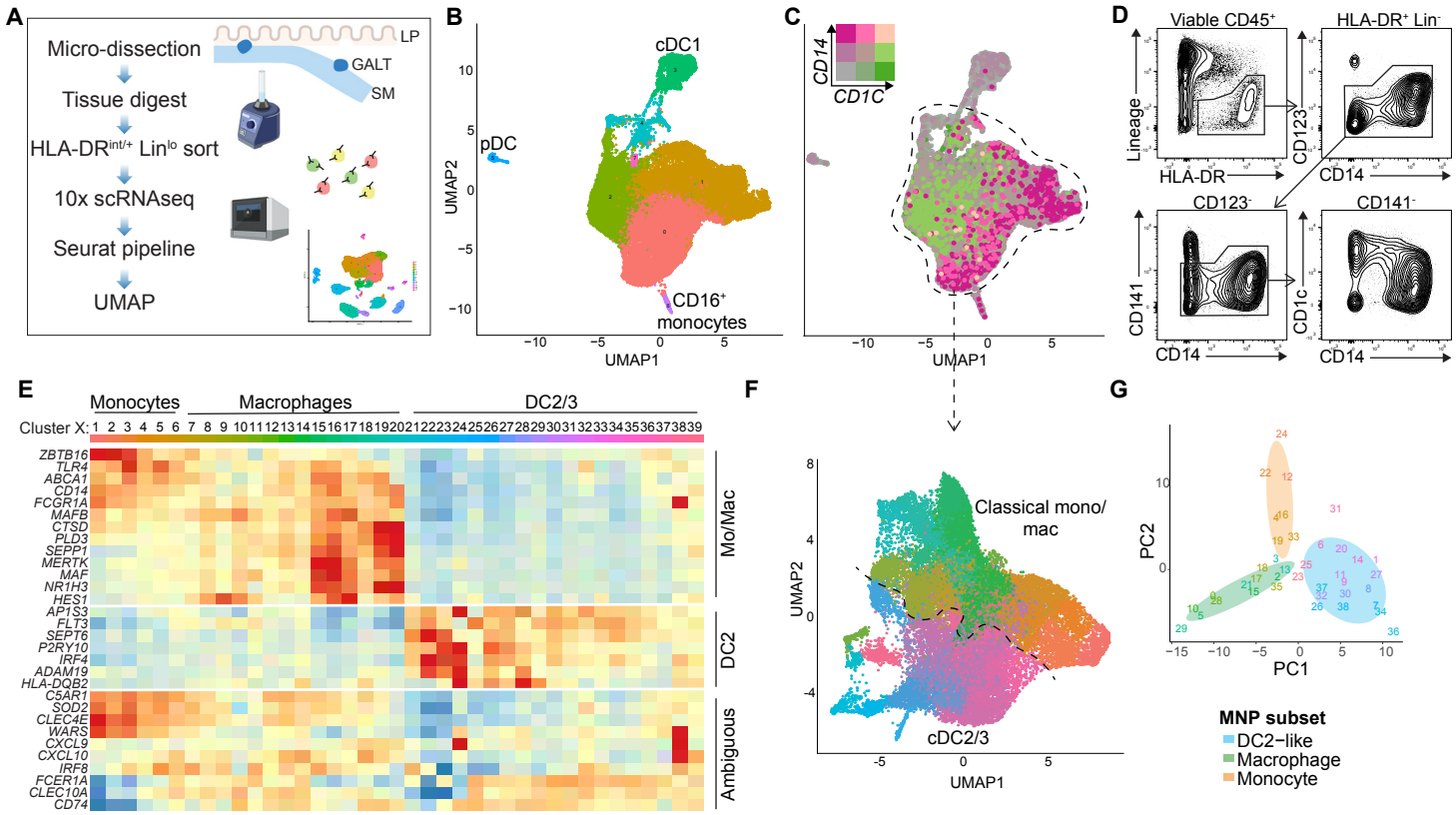
Figure 3. Site- and disease-specific changes in cDC subset composition. (**A**) Denoised and Scaled by Background (DSB) normalized CITE-seq expression of CD14 and (**B**) MNP subsets overlayed on UMAP of re-clustered mono/Mac/cDC subsets. (**C**) DSB normalized CITE-seq based surface expression of CD207 and CD11a on cDC2/cDC3/ambiguous cDC. (**D**) cDC2/cDC3/ambiguous DC within each CD207 vs. CD11a quadrant as a proportion of colon LP $CD1c^+$ MNP, results are combined from 6 CRC resections. (**E** and **F**) Representative gating (**E**) and bar plot (**F**) of CD207 vs CD11a expression within total live $CD45^+$ $HLADR^+$ $lineage^-$ $CD1c^+$ cells, from paired ileal and colonic CRC resections. n = 9 CRC patients. (**G** and **H**) Representative gating (**G**) and bar plot (**H**) of CD207 vs CD11a expression within total live $CD45^+$ $HLADR^+$ $lineage^-$ $CD1c^+$ cells, from unpaired colonic biopsies (IBD) and from CRC (non-IBD). n = 9 CRC patients, 4 quiescent IBD patients, and 5 active IBD patients. IBD activity score per site was determined by a gastroenterologist at time of removal. Statistical significance was determined using 2-way ANOVA with Dunnett's multiple comparisons, **p<0.01. (**I**) Representative gating and (**J**) bar plot of CD1c vs CD14 expression within total live $CD45^+$ $HLADR^+$ $lineage^-$ cells, from unpaired colonic biopsies (IBD). (**K**) Percentage of $CD11a^+$ cells within $CD1c^+$ cells in quiescent IBD biopsy samples from patients treated or untreated with vedolizumab. n = 6-10 biopsies from sites of quiescent IBD inflammation. Statistical significance was determined using 2-way ANOVA with Dunnett's multiple comparisons, *p<0.05, **p<0.01, ***p<0.001.

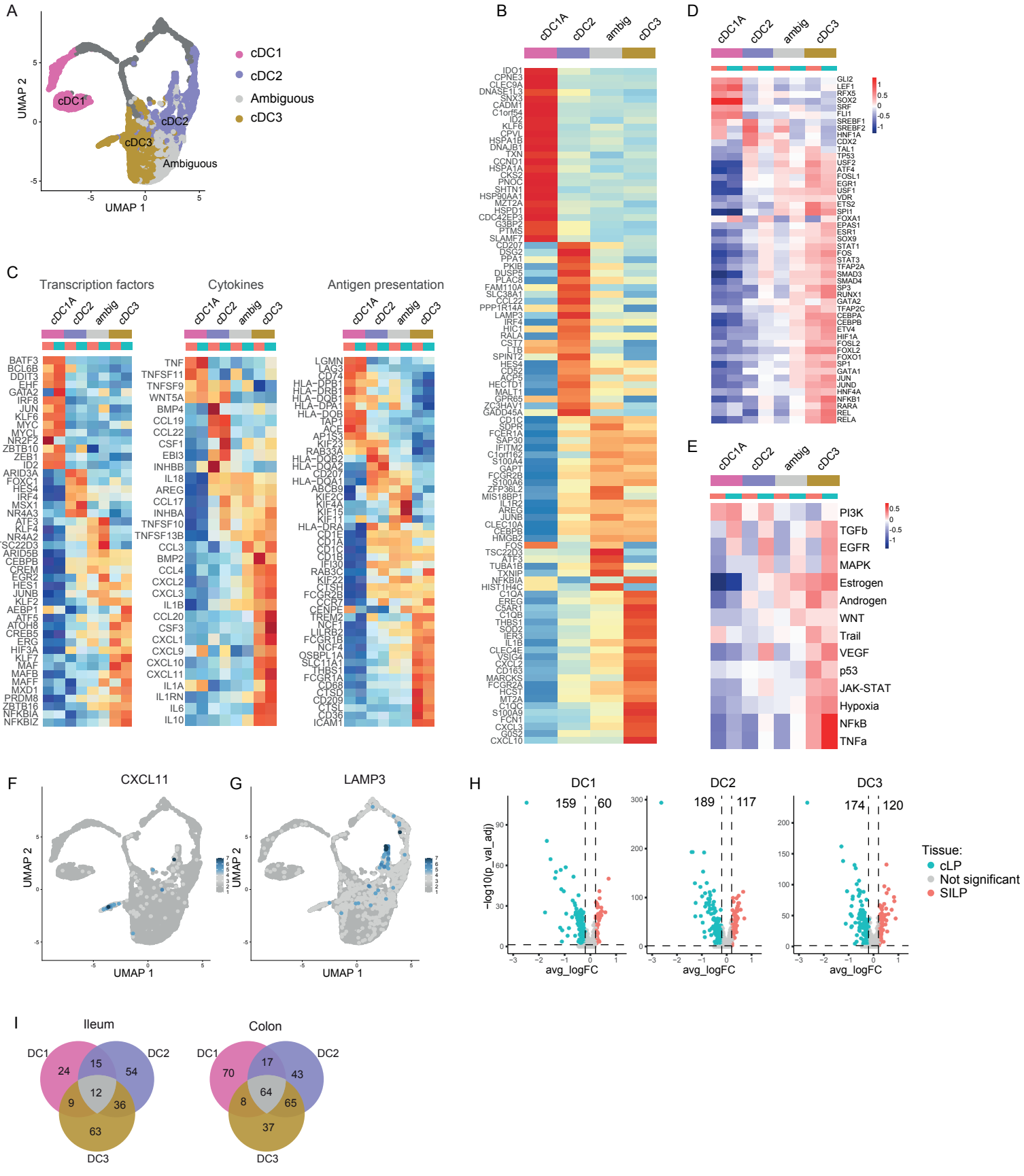
Figure 4. cDC precursors in the human intestine. (**A**-**C**) Flattened 3D tspace UMAPs (tUMAP) (calculated with pearson on tspace PC (tPC) 1-15 and metric set to pearson) of combined cDC1, cDC2 and cDC3 scRNAseq data. (**A**) Immature HLA-DR^{lo} DC clusters identified in Figure 2, overlaid on tUMAP. (**B**) Proliferation module score calculated using averaged expression of cell-cycle-associated genes *KIAA0101*, *TUBA1B*, *MKI67*, *HIST1H4C*, *TUBB*, *UBE2C*, *STMN1*, and *H2AFZ*. (**C**) Expression of CD11c gene *ITGAX* overlayed on DC tUMAP. (**D**) Scaled gene expression of top 50 DEG in immature cDC subsets sorted by avg. logFC for mature cDC1, cDC2, cDC3 subsets, using p = 0.05. (**E**) RNA velocities of immature DC clusters represented by arrows, calculated with Velocyto package on flat tUMAP. (**F**) PCA plot of the clusters identifiable by shared DEG as either pre-cDC1 (clusters 4 and 5, red and pink), pre-cDC2 (cluster 7, green), or pre-cDC3 (cluster 8, blue). (**G**) Each of the immature clusters which were not identifiable by mature DC DEG (clusters 1, 2, 3, and 6) overlaid on the PCA plot in (**F**). (**H**) Pearson correlation of each pre-cDC cluster and *CD14*$^{+}$ classical monocytes with different progenitor populations from Triana et al. [49]. (**I**) Diagrammatic representation of clusters in cDC development pathways. (**J**) Expression of cDC1-, cDC2-, and cDC3-specific genes in each cDC subset development pathway.

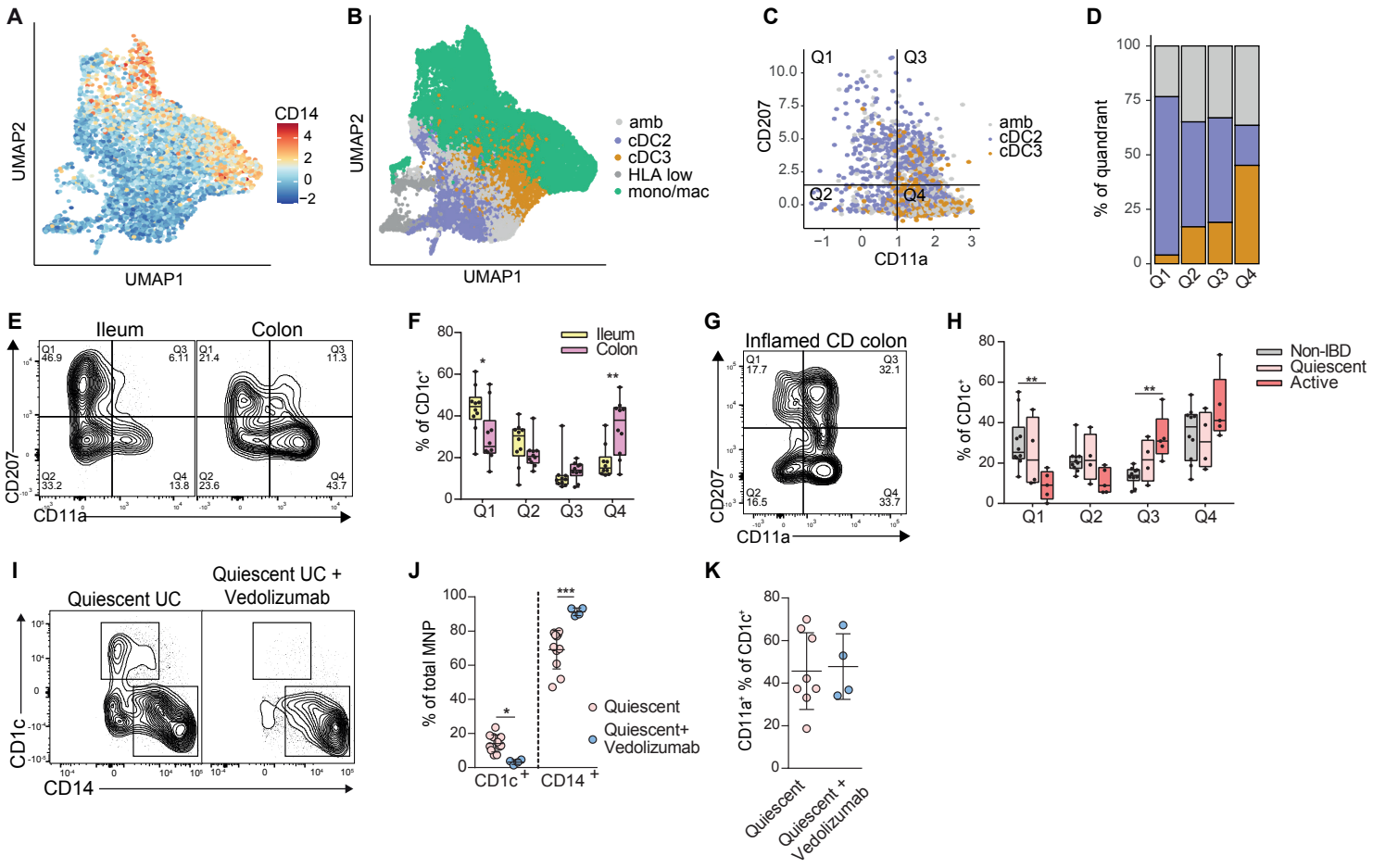
Figure 5. Characterisation of human intestinal LP monocyte/macrophage populations. (**A**) tSpace analysis on computationally isolated scRNA-seq data of ileal and colonic LP cells identified as belonging to the mono/Mac lineage. Data presented as a tUMAP (with Seurat function) and Louvain clustering (resolution = 0.55) based tPC 1-10 of cells. (**B**) Violin plots of normalized gene expression on Mono/Mac trajectory clusters M1-M11. DEG shown are associated with monocyte-macrophage differentiation.(**C**) Heatmap of select scaled gene expression by Mono/Mac trajectory clusters. (**D**) Pseudotime of Mono/Mac clusters calculated from tSpace output with means of trajectories starting at cluster M1. (**E**) Analysis of response

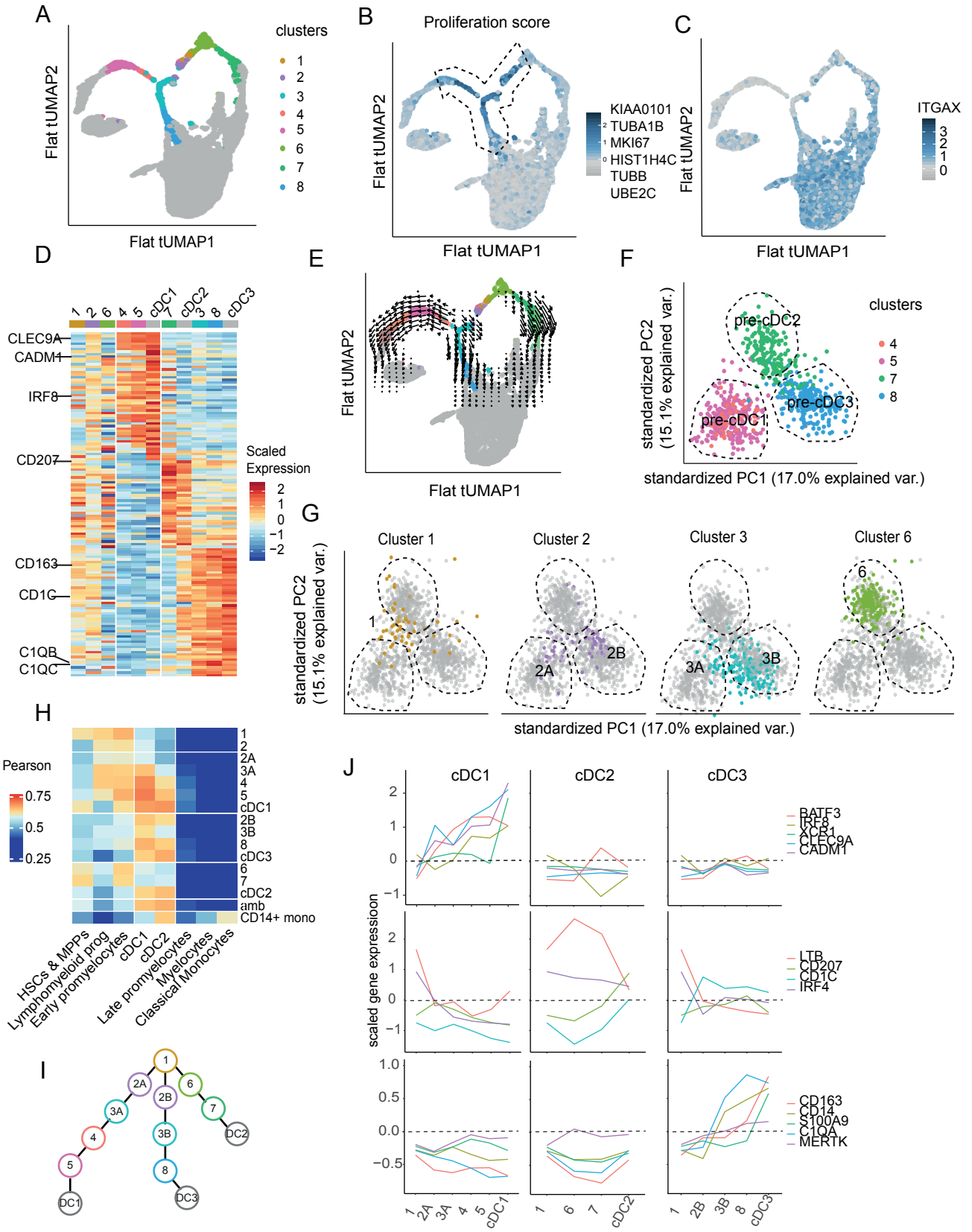
pathway activity in Mono/Mac trajectory clusters M1-M11 using the PROGENy package. (**F**-**J**) Violin plots of normalized gene expression by Mono/Mac trajectory clusters. DEG shown are associated with (**F**) pro-inflammatory cytokine secretion, (**G**) response to stimuli, (**H**) efferocytosis, (**I**) $CD169^{+}$ perivascular macrophages, and (**J**) self-maintaining/wound healing macrophages. (**K**) Number of genes upregulated in either ileum or colon for each mono/mac subset, based on 4 paired resection samples.

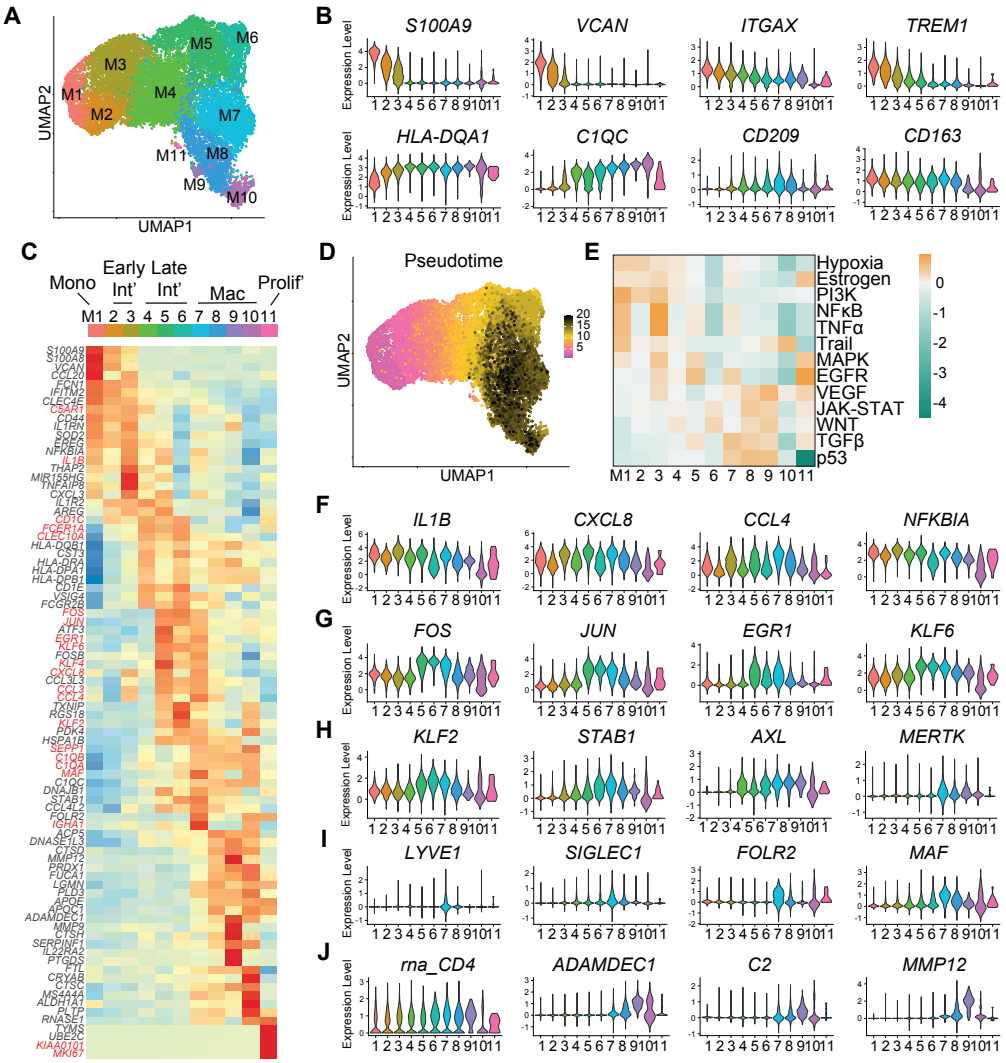
Figure 6. Flow cytometry analysis of mono/mac subsets. (**A**) Denoised and Scaled by Background (DSB) normalized expression of surface markers on combined ileal and colonic LP Mono/Mac trajectory clusters using CITE-seq, after exclusion of the small proliferating cluster M11. (**B**) Gating strategy to identify monocytes, monocyte-derived early and late intermediates, and two subsets of $CD11a^{-}$ macrophage. Representative data from 9 CRC resection samples. (**C**) Analysis of monocytes and monocyte-derived MNP subsets as a proportion of total $CD14^{+}$ cells in colonic and ileal LP. Data from 9 paired CRC resection samples. (**D**) Flow cytometry analysis of mono/mac subsets as proportion of $CD14^{+}$ $CD1c^{lo}$ MNP, from colonic biopsies. Extent of inflammation was scored at the time of biopsy collection by the clinician as quiescent, mild, or moderate IBD. n = 6-14 biopsies from sites of quiescent, mild, or moderate IBD inflammation, or from non-IBD controls. Some points represent single patients with samples taken from sites with different levels of inflammation. Statistical significance was determined using 2-way ANOVA with Dunnett's multiple comparisons, *p<0.05, **p<0.01 ***p<0.001.


A
Micro-dissection
Tissue digest
HLA-DR^int/+ Lin^lo sort
10x scRNAseq
Seurat pipeline
UMAP
LP
GALT
SM
B
cDC1
pDC
CD16+ monocytes
UMAP1
UMAP2
C
CD14
CD1C
UMAP1
D
Viable CD45+
HLA-DR+ Lin-
Lineage
CD123
HLA-DR
CD14
CD123-
CD141-
CD141
CD14
CD1c
CD14
E
Monocytes
Macrophages
DC2/3
Cluster X: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
ZBTB16
TLR4
ABCA1
CD14
FCGR1A
MAFB
CTSD
PLD3
SEPP1
MERTK
MAF
NR1H3
HES1
AP1S3
FLT3
SEPT6
P2RY10
IRF4
ADAM19
HLA-DQB2
C5AR1
SOD2
CLEC4E
WARS
CXCL9
CXCL10
IRF8
FCER1A
CLEC10A
CD74
Mo/Mac
DC2
Ambiguous
F
Classical mono/ mac
cDC2/3
UMAP1
UMAP2
G
PC1
PC2
MNP subset
DC2−like
Macrophage
Monocyte


A
UMAP 2
UMAP 1
cDC1
cDC2
cDC3
Ambiguous
cDC1
cDC2
Ambiguous
cDC3
B
cDC1A
cDC2
ambig
cDC3
D
cDC1A
cDC2
ambig
cDC3
C
Transcription factors
Cytokines
Antigen presentation
cDC1A
cDC2
ambig
cDC3
BATF3
BCL6B
DDIT3
EHF
GATA2
IRF8
JUN
KLF6
MYC
MYCL
NR2F2
ZBTB10
ZEB1
ID2
ARID3A
FOXC1
HES4
IRF4
MSX1
NR4A3
ATF3
KLF4
NR4A2
TSC22D3
ARID5B
CEBPB
CREM
EGR2
HES1
JUNB
KLF2
AEBP1
ATF5
ATOH8
CREB5
ERG
HIF3A
KLF7
MAF
MAFB
MAFF
MXD1
PRDM8
ZBTB16
NFKBIA
NFKBIZ
TNF
TNFSF11
TNFSF9
WNT5A
BMP4
CCL19
CCL22
CSF1
EBI3
INHBB
IL18
AREG
CCL17
INHBA
TNFSF10
TNFSF13B
CCL3
BMP2
CCL4
CXCL2
CXCL3
IL1B
CCL20
CSF3
CXCL1
CXCL9
CXCL10
CXCL11
IL1A
IL1RN
IL6
IL10
LGMN
LAG3
CD74
HLA−DPB1
HLA−DRB1
HLA−DQB1
HLA−DPA1
HLA−DOB
TAP1
ACE
AP1S3
KIF23
RAB33A
HLA−DQB2
HLA−DQA2
CD207
HLA−DQA1
ABCB9
KIF2C
KIF4A
KIF15
KIF11
HLA−DRA
CD1E
CD1A
CD1C
CD1B
IFI30
RAB3C
KIF22
CTSH
FCGR2B
CCR7
CENPE
TREM2
NCF1
LILRB2
FCGR1B
NCF4
OSBPL1A
SLC11A1
THBS1
FCGR1A
CD68
CTSD
CD209
CTSL
CD36
ICAM1
IDO1
CPNE3
CLEC9A
DNASE1L3
SNX3
CADM1
C1orf54
ID2
KLF6
CPVL
HSPA1B
DNAJB1
TXN
CCND1
HSPA1A
CKS2
PNOC
SHTN1
HSP90AA1
MZT2A
HSPD1
CDC42EP3
G3BP2
PTMS
SLAMF7
CD207
DSG2
PPA1
PKIB
DUSP5
PLAC8
FAM110A
SLC38A1
CCL22
PPP1R14A
LAMP3
IRF4
HIC1
RALA
CST7
LTB
SPINT2
HES4
CD52
ACP5
HECTD1
MALT1
GPR65
ZC3HAV1
GADD45A
CD1C
SDPR
FCER1A
SAP30
IFITM2
C1orf162
S100A4
GAPT
FCGR2B
S100A6
ZFP36L2
MIS18BP1
IL1B2
AREG
JUNB
CLEC10A
CEBPB
HMGB2
FOS
TSC22D3
ATF3
TUBA1B
TXNIP
NFKBIA
HIST1H4C
C1QA
EREG
C5AR1
C1QB
THBS1
SOD2
IER3
IL1B
CLEC4E
VSIG4
CXCL2
CD163
MARCKS
FCGR2A
HCST
MT2A
C1QC
S100A9
FCN1
CXCL3
G0S2
CXCL10
GLI2
LEF1
RFX5
SOX2
SRF
FLI1
SREBF1
SREBF2
HNF1A
CDX2
TAL1
TP53
USF2
ATF4
FOSL1
EGR1
USF1
VDR
ETS2
SPI1
FOXA1
EPAS1
ESR1
SOX9
STAT1
FOS
STAT3
TFAP2A
SMAD3
SMAD4
SP3
RUNX1
GATA2
TFAP2C
CEBPA
CEBPB
ETV4
HIF1A
FOSL2
FOXL2
FOXO1
SP1
GATA1
JUN
JUND
HNF4A
NFKB1
RARA
REL
RELA
1
0.5
0
-0.5
-1
E
cDC1A
cDC2
ambig
cDC3
PI3K
TGFb
EGFR
MAPK
Estrogen
Androgen
WNT
Trail
VEGF
p53
JAK-STAT
Hypoxia
NFkB
TNFa
0.5
0
-0.5
-1
F
CXCL11
UMAP 2
UMAP 1
G
LAMP3
UMAP 2
UMAP 1
H
DC1
DC2
DC3
159
60
189
117
174
120
−log10(p_val_adj)
avg_logFC
avg_logFC
avg_logFC
Tissue:
cLP
Not significant
SILP
I
Ileum
DC1
DC2
DC3
24
15
54
12
9
36
63
Colon
DC1
DC2
DC3
70
17
43
64
8
65
37

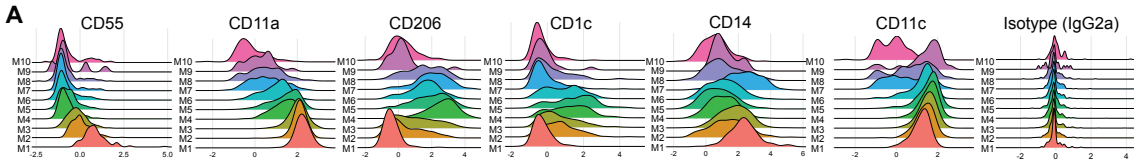

A
UMAP2
UMAP1
CD14
4
2
0
−2
B
UMAP2
UMAP1
amb
cDC2
cDC3
HLA low
mono/mac
C
Q1
Q3
Q2
Q4
CD207
CD11a
amb
cDC2
cDC3
D
% of quandrant
Q1
Q2
Q3
Q4
E
Ileum
Colon
Q1 46.9
Q3 6.11
Q2 33.2
Q4 13.8
Q1 21.4
Q3 11.3
Q2 23.6
Q4 43.7
CD207
CD11a
F
Ileum
Colon
% of CD1c+
Q1
Q2
Q3
Q4
G
Inflamed CD colon
Q1 17.7
Q3 32.1
Q2 16.5
Q4 33.7
CD207
CD11a
H
% of CD1c+
Non-IBD
Quiescent
Active
Q1
Q2
Q3
Q4
I
Quiescent UC
Quiescent UC + Vedolizumab
CD1c
CD14
J
% of total MNP
Quiescent
Quiescent+ Vedolizumab
CD1c+
CD14+
K
CD11a+ % of CD1c+
Quiescent
Quiescent + Vedolizumab

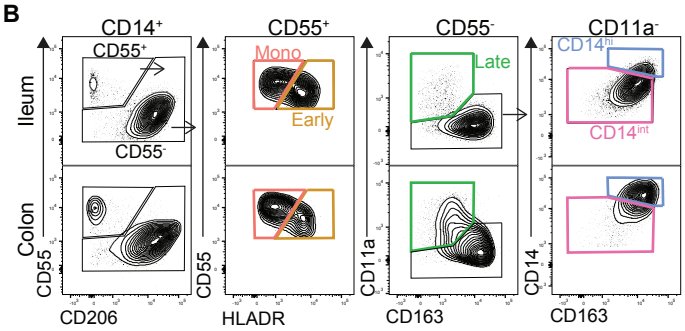

A
Flat tUMAP2
Flat tUMAP1
clusters
1
2
3
4
5
6
7
8
B
Proliferation score
KIAA0101
TUBA1B
MKI67
HIST1H4C
TUBB
UBE2C
C
ITGAX
D
CLEC9A
CADM1
IRF8
CD207
CD163
CD1C
C1QB
C1QC
Scaled Expression
E
F
pre-cDC2
pre-cDC3
pre-cDC1
standardized PC2 (15.1% explained var.)
standardized PC1 (17.0% explained var.)
G
Cluster 1
Cluster 2
Cluster 3
Cluster 6
H
Pearson
HSCs & MPPs
Lymphomyeloid prog
Early promyelocytes
cDC1
cDC2
Late promyelocytes
Myelocytes
Classical Monocytes
CD14+ mono
I
DC1
DC2
DC3
J
cDC1
cDC2
cDC3
scaled gene expression
BATF3
IRF8
XCR1
CLEC9A
CADM1
LTB
CD207
CD1C
IRF4
CD163
CD14
S100A9
C1QA
MERTK

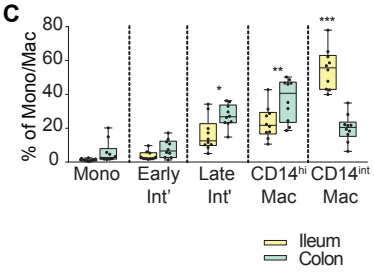


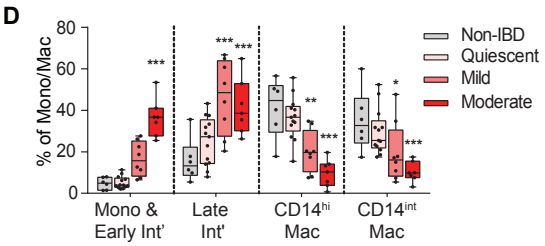
K

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
Ileum	0	5	3	0	2	15	1	0	1	84	n.d.
Colon	3	4	34	41	64	18	37	37	0	36	n.d.


A
CD55
CD11a
CD206
CD1c
CD14
CD11c
Isotype (IgG2a)


B
CD14+
CD55+
CD55-
Ileum
Colon
CD206
CD55+
Mono
Early
HLADR
CD55-
Late
CD11a
CD163
CD11a-
CD14hi
CD14int
CD14
CD163


C
% of Mono/Mac
Mono
Early Int'
Late Int'
CD14hi Mac
CD14int Mac
Ileum
Colon


D
% of Mono/Mac
Mono & Early Int'
Late Int'
CD14hi Mac
CD14int Mac
Non-IBD
Quiescent
Mild
Moderate

References

1. Joeris, T., Müller-Luda, K., Agace, W. W. & Mowat, A. M. I. Diversity and functions of intestinal mononuclear phagocytes. *Mucosal Immunol.* **10**, 845–864 (2017).
2. Caër, C. & Wick, M. J. Human Intestinal Mononuclear Phagocytes in Health and Inflammatory Bowel Disease. *Front. Immunol.* **11**, 410 (2020).
3. Arroyo Portilla, C., Tomas, J., Gorvel, J. P. & Lelouard, H. From Species to Regional and Local Specialization of Intestinal Macrophages. *Front. Cell Dev. Biol.* **8**, (2021).
4. Cabeza-Cabrerizo, M., Cardoso, A., Minutti, C. M., Pereira da Costa, M. & Reis E Sousa, C. Dendritic Cells Revisited. *Annu. Rev. Immunol.* 131–166 (2021).doi:10.1146/annurev-immunol-061020-053707
5. Na, Y. R., Stakenborg, M., Seok, S. H. & Matteoli, G. Macrophages in intestinal inflammation and resolution: a potential therapeutic target in IBD. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 531–543 (2019).
6. Bain, C. C. & Schridde, A. Origin , Differentiation , and Function of Intestinal Macrophages. **9**, 1–15 (2018).
7. Dutertre, C. A. *et al.* Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity* **51**, 573-589.e8 (2019).
8. Brown, C. C. *et al.* Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell* **179**, 846-863.e24 (2019).
9. Guilliams, M. *et al.* Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. *Immunity* **45**, 669–684 (2016).
10. Villar, J. & Segura, E. Decoding the Heterogeneity of Human Dendritic Cell Subsets. *Trends Immunol.* **41**, 1062–1071 (2020).
11. Blériot, C., Chakarov, S. & Ginhoux, F. Determinants of Resident Tissue Macrophage Identity and Function. *Immunity* **52**, 957–970 (2020).
12. Bernardo, D., Chaparro, M. & Gisbert, J. P. Human Intestinal Dendritic Cells in Inflammatory Bowel Diseases. *Mol. Nutr. Food Res.* **62**, 1–8 (2018).
13. Viola, M. F. & Boeckxstaens, G. Intestinal resident macrophages: Multitaskers of the gut. *Neurogastroenterol. Motil.* **32**, (2020).
14. Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* **14**, 667–685 (2014).
15. Chapuy, L. *et al.* Two distinct colonic CD14+ subsets characterized by single-cell RNA profiling in Crohn's disease. *Mucosal Immunol.* **12**, 703–719 (2019).
16. 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).
17. Mörbe, U. M. *et al.* Human gut-associated lymphoid tissues (GALT); diversity, structure, and function. *Mucosal Immunol.* **14**, 793–802 (2021).
18. Spencer, J., Siu, J. H. Y. & Montorsi, L. Human intestinal lymphoid tissue in time and space. *Mucosal Immunol.* **12**, 296–298 (2019).
19. Senda, T. *et al.* Microanatomical dissection of human intestinal T-cell immunity reveals site-specific changes in gut-associated lymphoid tissues over life. *Mucosal Immunol.* **12**, 378–389 (2019).
20. Boursier, L., Gordon, J. N., Thiagamoorthy, S., Edgeworth, J. D. & Spencer, J. Human intestinal IgA response is generated in the organized gut-associated lymphoid tissue but not in the lamina propria. *Gastroenterology* **128**, 1879–1889 (2005).
21. Barone, F. *et al.* IgA-producing plasma cells originate from germinal centers that are induced by B-cell receptor engagement in humans. *Gastroenterology* **140**, 947–956 (2011).
22. Masahata, K. *et al.* Generation of colonic IgA-secreting cells in the caecal patch. *Nat. Commun.* **5**, 3704 (2014).
23. Jørgensen, P. B. *et al.* Identification, isolation and analysis of human gut-associated

lymphoid tissues. *Nat. Protoc.* **16**, (2021).
24. Fenton, T. M. *et al.* Immune Profiling of Human Gut-Associated Lymphoid Tissue Identifies a Role for Isolated Lymphoid Follicles in Priming of Region-Specific Immunity. *Immunity* **52**, 557-570.e6 (2020).
25. Cytlak, U. *et al.* Differential IRF8 Transcription Factor Requirement Defines Two Pathways of Dendritic Cell Development in Humans. *Immunity* **53**, 353-370.e8 (2020).
26. T'Jonck, W., Guilliams, M. & Bonnardel, J. Niche signals and transcription factors involved in tissue-resident macrophage development. *Cell. Immunol.* **330**, 43–53 (2018).
27. Yin, X. *et al.* Human Blood CD1c + Dendritic Cells Encompass CD5 high and CD5 low Subsets That Differ Significantly in Phenotype, Gene Expression, and Functions. *J. Immunol.* **198**, 1553–1564 (2017).
28. Villani, A. C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science (80-.).* **356**, (2017).
29. See, P. *et al.* Mapping the human DC lineage through the integration of high-dimensional techniques. *Science (80-.).* **356**, eaag3009 (2017).
30. Bourdely, P. *et al.* Transcriptional and Functional Analysis of CD1c+ Human Dendritic Cells Identifies a CD163+ Subset Priming CD8+CD103+ T Cells. *Immunity* **53**, 335-352.e8 (2020).
31. Carpentier, S. *et al.* Comparative genomics analysis of mononuclear phagocyte subsets confirms homology between lymphoid tissue-resident and dermal XCR1+ DCs in mouse and human and distinguishes them from Langerhans cells. *J. Immunol. Methods* **432**, 35–49 (2016).
32. Goudot, C. *et al.* Aryl Hydrocarbon Receptor Controls Monocyte Differentiation into Dendritic Cells versus Macrophages. *Immunity* **47**, 582-596.e6 (2017).
33. Tang-Huau, T. L. *et al.* Human in vivo-generated monocyte-derived dendritic cells and macrophages cross-present antigens through a vacuolar pathway. *Nat. Commun.* **9**, (2018).
34. Dermadi, D. *et al.* Exploration of Cell Development Pathways through High-Dimensional Single Cell Analysis in Trajectory Space. *iScience* **23**, 100842 (2020).
35. Murphy, K. M. *Transcriptional Control of Dendritic Cell Development*. *Adv. Immunol.* **120**, (Elsevier Inc., 2013).
36. Smita, S. *et al.* Importance of EMT factor ZEB1 in cDC1 'mutuDC line' mediated induction of Th1 immune response. *Front. Immunol.* **9**, 9–12 (2018).
37. Nagaoka, M. *et al.* The Orphan Nuclear Receptor NR4A3 Is Involved in the Function of Dendritic Cells. *J. Immunol.* **199**, 2958–2967 (2017).
38. Boulet, S. *et al.* The orphan nuclear receptor NR4A3 controls the differentiation of monocyte-derived dendritic cells following microbial stimulation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15150–15159 (2019).
39. Chewchuk, S., Jahan, S. & Lohnes, D. Cdx2 regulates immune cell infiltration in the intestine. *Sci. Rep.* (2021).doi:10.1038/s41598-021-95412-w
40. Laffont, S., Seillet, C. & Guéry, J. C. Estrogen receptor-dependent regulation of dendritic cell development and function. *Front. Immunol.* (2017).doi:10.3389/fimmu.2017.00108
41. Mulè, M. P., Martins, A. J. & Tsang, J. S. Normalizing and denoising protein expression data from droplet-based single cell profiling. (2020).
42. Breton, G. *et al.* Circulating precursors of human CD1c+ and CD141+ dendritic cells. *J. Exp. Med.* **212**, 401–413 (2015).
43. Durand, M. *et al.* Human lymphoid organ cDC2 and macrophages play complementary roles in T follicular helper responses. *J. Exp. Med.* **216**, 1561–1581 (2019).
44. Cabeza-Cabrerizo, M. *et al.* Tissue clonality of dendritic cell subsets and emergency DCpoiesis revealed by multicolor fate mapping of DC progenitors. *Sci. Immunol.* **4**, 1–14 (2019).

45. Chen, J., Schlitzer, A., Chakarov, S., Ginhoux, F. & Poidinger, M. Mpath maps multi-branching single-cell trajectories revealing progenitor cell progression during development. *Nat. Commun.* **7**, 1–15 (2016).
46. Ma, W. *et al.* Single cell RNA-Seq reveals pre-cDCs fate determined by transcription factor combinatorial dose. *BMC Mol. Cell Biol.* **20**, 1–14 (2019).
47. See, P. *et al.* Mapping the human DC lineage through the integration of high-dimensional techniques. *Science (80-.).* **356**, (2017).
48. Manno, G. La *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
49. Triana, S., Vonficht, D., Jopp-saile, L., Raffel, S. & Lutz, R. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* **22**, 1577–1589 (2021).
50. Bain, C. C. *et al.* Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6C hi monocyte precursors. **6**, (2013).
51. Rivollier, A., He, J., Kole, A., Valatas, V. & Kelsall, B. L. Inflammation switches the differentiation program of Ly6Chi monocytes from antiinflammatory macrophages to inflammatory dendritic cells in the colon. *J. Exp. Med.* **209**, 139–55 (2012).
52. Bain, C. C. *et al.* Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nat. Immunol.* **15**, 929–937 (2014).
53. Desalegn, G. & Pabst, O. Inflammation triggers immediate rather than progressive changes in monocyte differentiation in the small intestine. *Nat. Commun.* **10**, 3229 (2019).
54. Bujko, A. *et al.* Transcriptional and functional profiling defines human small intestinal macrophage subsets. (2018).
55. Schridde, A. *et al.* Tissue-specific differentiation of colonic macrophages requires TGFβ receptor-mediated signaling. *Mucosal Immunol.* **10**, 1387–1399 (2017).
56. Schubert, M. *et al.* Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* **9**, 20 (2018).
57. Pepe, G. *et al.* The estrogen-macrophage interplay in the homeostasis of the female reproductive tract. *Hum. Reprod. Update* **24**, 652–672 (2018).
58. Bahrami, S. & Drabløs, F. Gene regulation in the immediate-early response process. *Adv. Biol. Regul.* **62**, 37–49 (2016).
59. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* (2019).doi:10.1101/gr.240663.118
60. Holland, C. H. *et al.* Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, 36 (2020).
61. Elliott, M. R., Koster, K. M. & Murphy, P. S. Efferocytosis Signaling in the Regulation of Macrophage Inflammatory Responses. *J. Immunol.* **198**, 1387–1394 (2017).
62. Lantz, C., Radmanesh, B., Liu, E., Thorp, E. B. & Lin, J. Single-cell RNA sequencing uncovers heterogenous transcriptional signatures in macrophages during efferocytosis. *Sci. Rep.* **10**, 14333 (2020).
63. Silva, H. M., Kitoko, J. Z. & Queiroz, C. P. c-MAF dependent perivascular macrophages regulate diet induced metabolic syndrome. *Preprint* (2021).doi:10.1101/2021.02.07.430147
64. Honda, M. *et al.* Perivascular localization of macrophages in the intestinal mucosa is regulated by Nr4a1 and the microbiome. *Nat. Commun.* **11**, 1329 (2020).
65. Hiemstra, I. H. *et al.* The identification and developmental requirements of colonic CD169+ macrophages. *Immunology* **142**, 269–278 (2014).
66. Kang, B. *et al.* Commensal microbiota drive the functional diversification of colon macrophages. *Mucosal Immunol.* **13**, 216–229 (2020).
67. Shaw, T. N. *et al.* Tissue-resident macrophages in the intestine are long lived and defined by Tim-4 and CD4 expression. **215**, 1507–1518 (2018).
68. Schepper, S. De *et al.* Self-Maintaining Gut Macrophages Are Essential for Intestinal

Homeostasis. *Cell* **175**, 400-415.e13 (2018).
69. Liu, Z. *et al.* Fate Mapping via Ms4a3-Expression History Traces Monocyte-Derived Cells. *Cell* **178**, 1509-1525.e19 (2019).
70. Li, Y., Yun, K. & Mu, R. A review on the biology and properties of adipose tissue macrophages involved in adipose tissue physiological and pathophysiological processes. *Lipids Health Dis.* **19**, 1–9 (2020).
71. Frey, S. K. & Vogel, S. Vitamin A Metabolism and Adipose Tissue Biology. *Nutrients* **3**, 27–39 (2011).
72. Maixner, N. *et al.* A trail-tl1a paracrine network involving adipocytes, macrophages, and lymphocytes induces adipose tissue dysfunction downstream of e2f1 in human obesity. *Diabetes* **69**, 2310–2323 (2020).
73. Tamoutounour, S. *et al.* CD64 distinguishes macrophages from dendritic cells in the gut and reveals the Th1-inducing role of mesenteric lymph node macrophages during colitis. *Eur. J. Immunol.* **42**, 3150–3166 (2012).
74. Schlitzer, A., McGovern, N. & Ginhoux, F. Dendritic cells and monocyte-derived cells: Two complementary and integrated functional systems. *Semin. Cell Dev. Biol.* **41**, 9–22 (2015).
75. Satpathy, A. T. *et al.* Zbtb46 expression distinguishes classical dendritic cells and their committed progenitors from other immune lineages. *J. Exp. Med.* **209**, 1135–52 (2012).
76. Varol, C. *et al.* Intestinal lamina propria dendritic cell subsets have different origin and functions. *Immunity* **31**, 502–12 (2009).
77. Bogunovic, M. *et al.* Origin of the lamina propria dendritic cell network. *Immunity* **31**, 513–25 (2009).
78. Tang-Huau, T. L. & Segura, E. Human in vivo-differentiated monocyte-derived dendritic cells. *Semin. Cell Dev. Biol.* **86**, 44–49 (2019).
79. Sallusto, F. & Lanzavecchia, A. Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor alpha. *J. Exp. Med.* **179**, 1109–1118 (1994).
80. Jaensson, E. *et al.* Small intestinal CD103+ dendritic cells display unique functional properties that are conserved between mice and humans. *J. Exp. Med.* **205**, 2139–49 (2008).
81. Watchmaker, P. B. *et al.* Comparative transcriptional and functional profiling defines conserved programs of intestinal DC differentiation in humans and mice. **15**, (2014).
82. Mann, E. R. *et al.* Compartment-specific immunity in the human gut: properties and functions of dendritic cells in the colon versus the ileum. *Gut* **140**, 169 (2015).
83. Bernardo, D. *et al.* Chemokine (C-C Motif) Receptor 2 Mediates Dendritic Cell Recruitment to the Human Colon but Is Not Responsible for Differences Observed in Dendritic Cell Subsets, Phenotype, and Function Between the Proximal and Distal Colon. *C. Cell. Mol. Gastroenterol. Hepatol.* **2**, 22-39.e5 (2016).
84. Schleier, L. *et al.* Non-classical monocyte homing to the gut via α4β7 integrin mediates macrophage-dependent intestinal wound healing. *Gut* **69**, 252–263 (2020).
85. Villablanca, E. J. *et al.* B7 Integrins Are Required To Give Rise To Intestinal Mononuclear Phagocytes With Tolerogenic Potential. *Gut* **63**, 1431–1440 (2014).
86. Worbs, T. *et al.* Oral tolerance originates in the intestinal immune system and relies on antigen carriage by dendritic cells. *J. Exp. Med.* **203**, 519–27 (2006).
87. Ohnmacht, C. *et al.* Constitutive ablation of dendritic cells breaks self-tolerance of CD4 T cells and results in spontaneous fatal autoimmunity. *J. Exp. Med.* **206**, 549–559 (2009).
88. Travis, M. A. *et al.* Loss of integrin alpha(v)beta8 on dendritic cells causes autoimmunity and colitis in mice. *Nature* **449**, 361–5 (2007).
89. Gabanyi, I. *et al.* Neuro-immune Interactions Drive Tissue Programming in Intestinal Macrophages. *Cell* **164**, 378–391 (2016).

90. Carlin, L. M. *et al.* Nr4a1-dependent Ly6C(low) monocytes monitor endothelial cells and orchestrate their disposal. *Cell* **153**, 362–75 (2013).
91. Jones, G. *et al.* Dynamics of Colon Monocyte and Macrophage Activation During Colitis. **9**, (2018).
92. Rubin, S. J. S. *et al.* Mass cytometry reveals systemic and local immune signatures that distinguish inflammatory bowel diseases. *Nat. Commun.* **10**, 2686 (2019).
93. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730.e22 (2019).
94. Jones, G.-R. *et al.* IBD prevalence in Lothian, Scotland, derived by capture-recapture methodology. *Gut* gutjnl-2019-318936 (2019).doi:10.1136/gutjnl-2019-318936
95. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, (2017).
96. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
97. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
98. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
99. R Core Team R: A Language and Environment for Statistical Computing. *Vienna, Austria* (2018).
100. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, (2019).
101. Barron, M. & Li, J. Identifying and removing the cell-cycle effect from single-cell RNA-Sequencing data. *Sci. Rep.* **6**, 33892 (2016).
102. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* (2018).doi:10.21105/joss.00861
103. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–47 (2019).

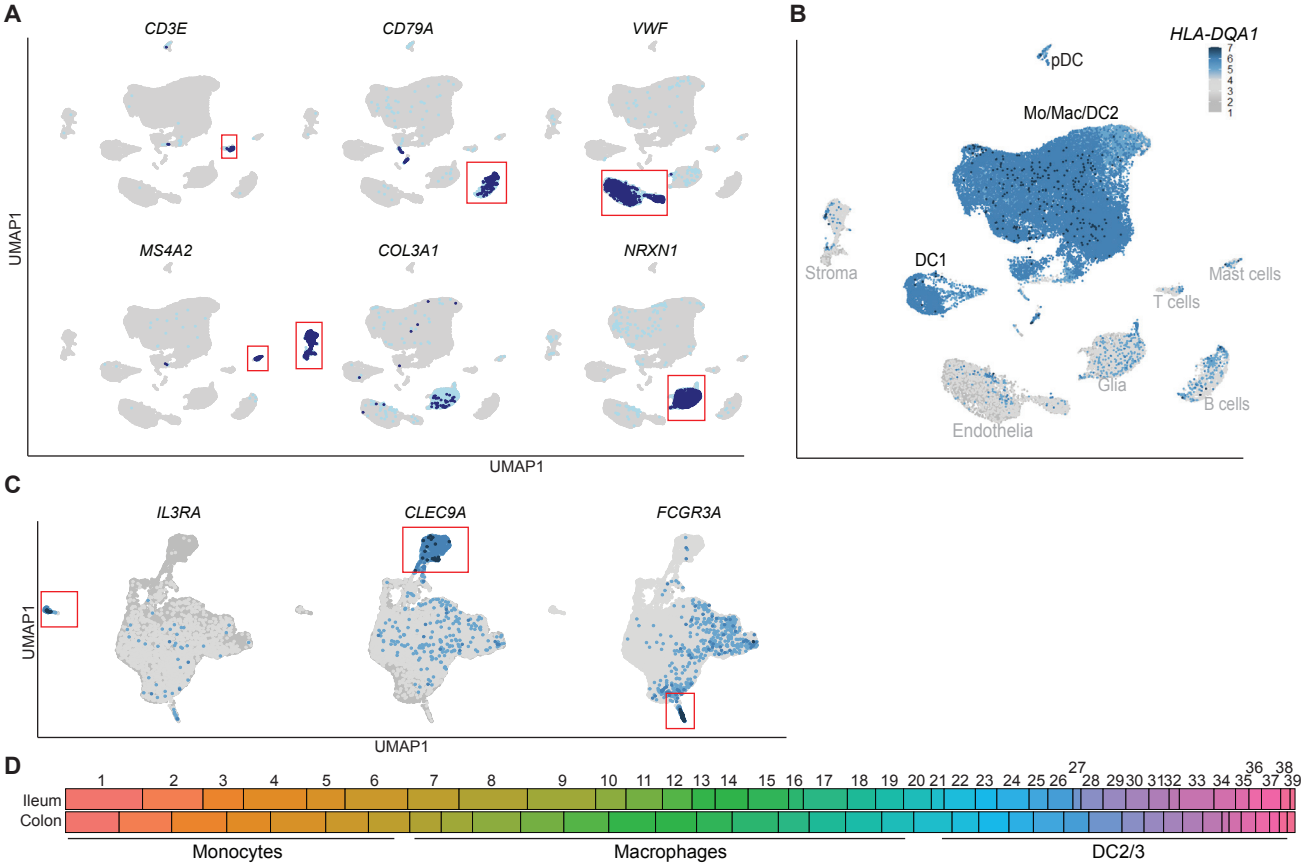


Figure S1. Initial characterisation of scRNA-seq data sets. (A and B) scRNAseq data of 42,506 combined ileal and colonic LP cells, digested and enriched for HLA-DR+ cells by MACS separation and FACS sorting. Data are from 6 colonic and 4 paired ileal CRC patients. (A) Normalized gene expression of signature genes for non-MNP cells within the dataset, including T cells, B cells, endothelia, mast cells, stroma, and glia. (B) Normalized HLA-DQA1 expression. (C) scRNA-seq data of 28,758 computationally isolated and re-clustered MNP from combined ileal and colonic LP cells, showing normalized gene expression of signature genes for pDC, cDC1, and non-classical monocytes. (D) Proportion of each cluster present in ileal vs colonic samples. Data are from 4 paired ileal and colonic samples. Related to Figure 1.

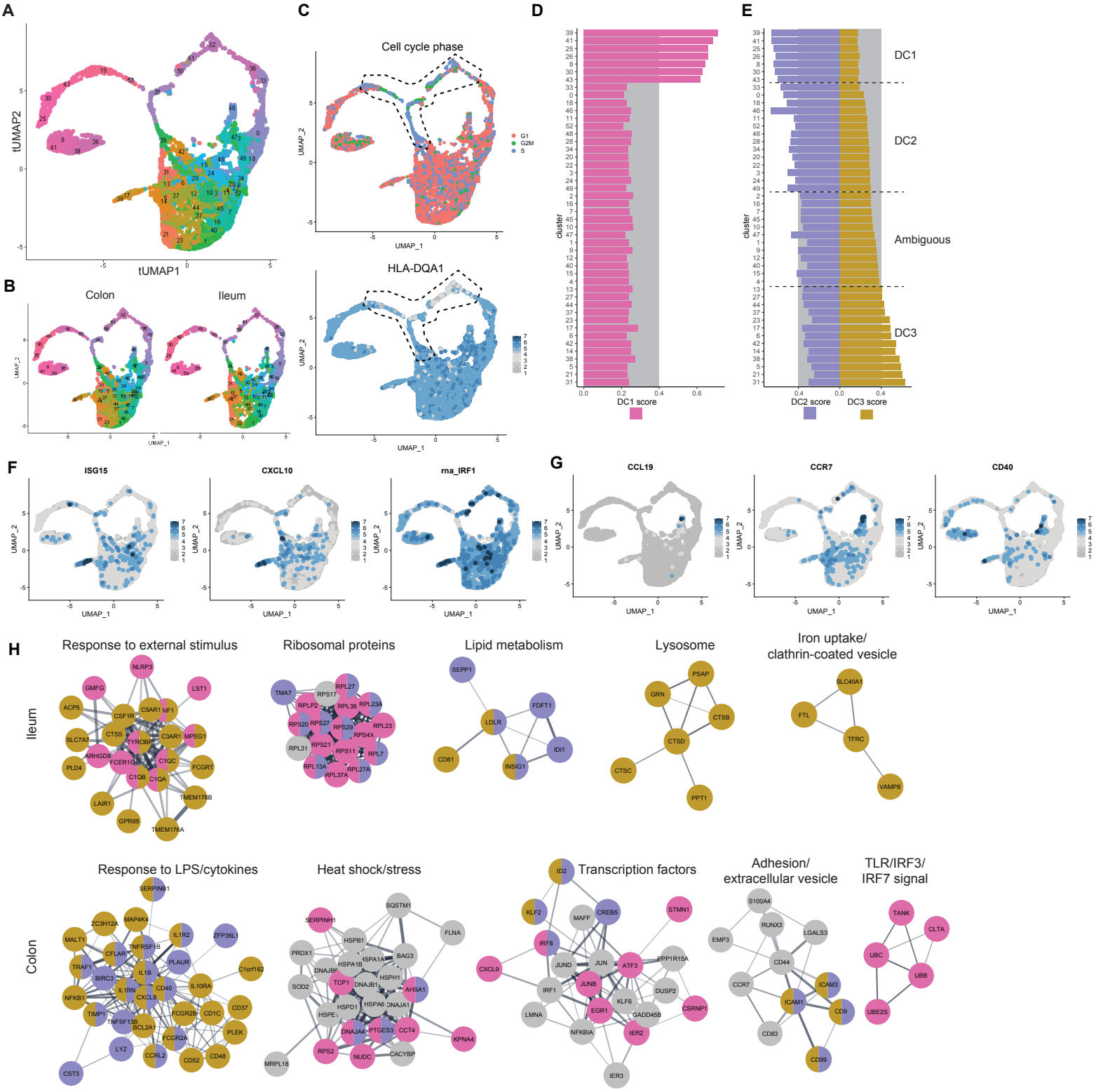


Figure S2. scRNA-seq characterisation of cDC2/3. (A-H) scRNA-seq data of bioinformatically isolated and re-clustered cDC showing combined colon and ileum LP unless otherwise stated. (A and B) Flattened 3D tspace UMAP (tUMAP) plot of 52 cDC clusters obtained at resolution 3.0, showing (A) combined colon and ileum LP and (B) separated colon and ileum LP. (C) tUMAP plots overlayed with cell cycle phase and HLA-DQA1 expression, respectively. (D and E) 44 mature (non-proliferating, HLA-DQA1hi) cDC clusters sorted first according to (D) decreasing expression of cDC1 score, composed of core DC1 signature genes then according to (E) increasing expression of a cDC3 score, composed of signature genes identified by Bourdeley et al.30 (F and G) tUMAP plots overlayed with selected (F) interferon stimulated genes and (G) migratory cDC genes. (H) STRING analysis of overlapping DEG enriched in cDC subsets in ileal vs. colonic LP. GO terms from a mix of KEGG pathways, Reactome pathways, and Gene Ontology Biological Processes. Related to Figure 2

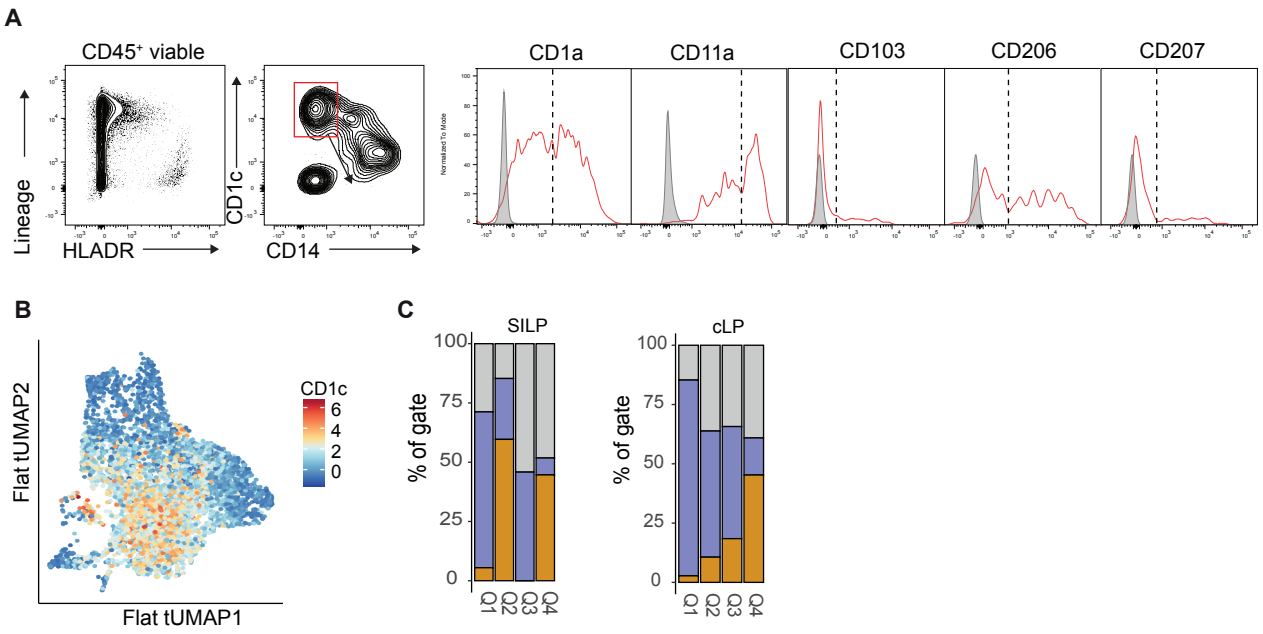


Figure S3. Phenotypic cDC subset characterisation. (A) A 282-antibody screen (Legendscreen) showing representative gating and plots for all antibodies which had bimodal expression on digested resected colon LP CD1c+ MNP. Representative plots shown of n = 1 or 2 CRC patients analysed per antibody. (B) Flattened 3D tspace UMAP (tUMAP) plot of cDC clusters overlayed with DSB normalized CITE-seq CD1C expression. (C) cDC2/cDC3/ambiguous DC within each CD207 vs. CD11a quadrant as a proportion of colonic and ileal LP CD1c+ MNP, both tissues from a single CRC resection patient. Related to Figure 3.

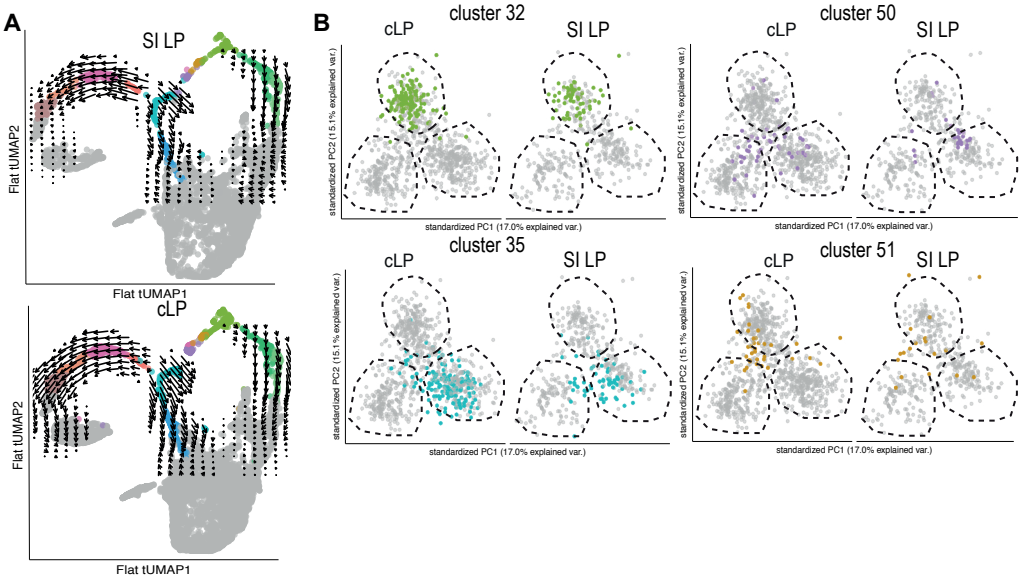


Figure S4. Characterisation of Pre-cDC in the small intestine and colon LP. (A) RNA velocities of immature DC clusters represented by arrows, calculated with Velocyto package on flattened 3D tUMAP, split into small intestine and colon LP. (B) PCA plot of the immature clusters which were not identifiable by mature DC DEG (clusters 1, 2, 3, and 6) overlaid on the PCA plot from Figure 4F, split into small intestine and colon LP. Related to Figure 4.

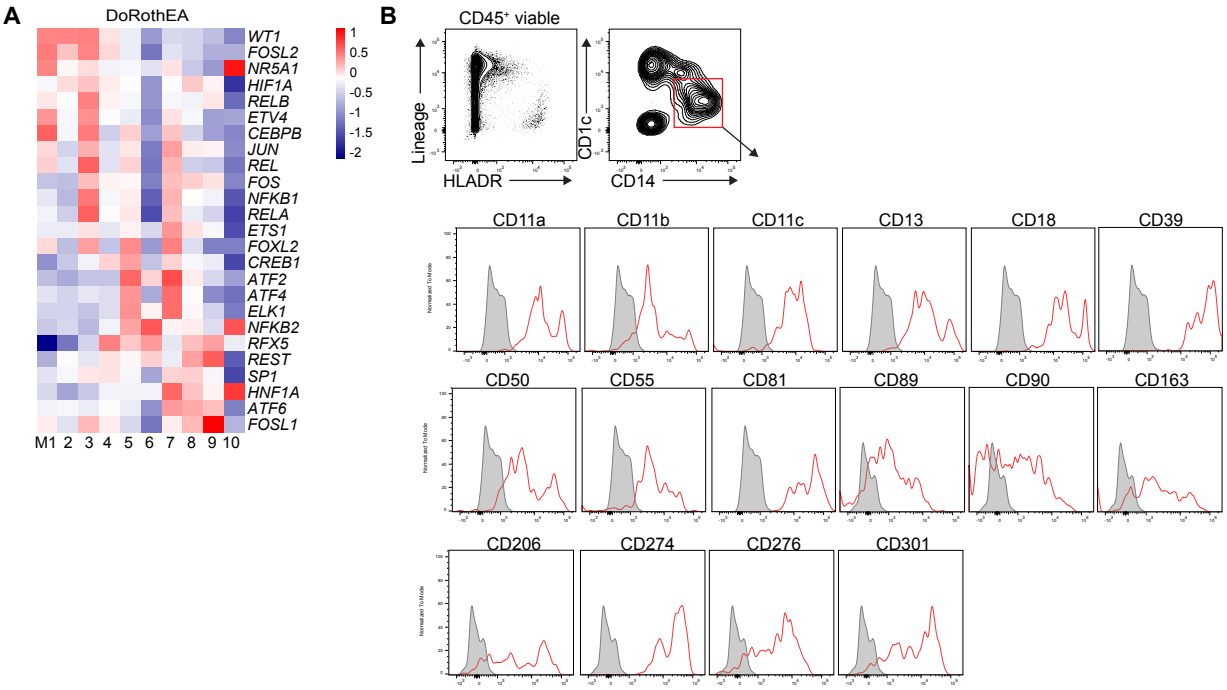


Figure S5. Intestinal Mono/Mac development analysis. (A) Transcription factor activity inferred from target gene expression in Mono/Mac trajectory clusters M1-M10, using the DoRothEA package. (B) Representative staining of all antibodies showing a bimodal expression pattern on colon CD14+ MNP in a 282-antibody based Legendscreen. 1-2 patients tested per antibody. Gated on viable CD45+ HLADR+ lineage- cells. Grey line = FMO, red line= specific antibody stain. Related to Figure 5.

Patient	Age	Sex	Diagnosis	Cancer TNM	Cancer stage	Site of colon resection	Site of SI resection	CITEseq
SC1	66	F	Adenocarcinom	T1 N0 V0 M0	1	n.d.	Ileum	-
SC2	74	M	Adenocarcinom	T4a N1a V0 M0	3b	Transverse	Ileum	-
SC3	63	F	Polyp	n.a.	n.a.	Ascending	Ileum	-
SC4	76	F	Adenocarcinom	T4 N0	2	Transverse	Ileum	Colon+Ileum
SC5	63	F	Polyp	n.a.	n.a.	Ascending	-	Colon
SC6	83	M	Adenocarcinom	T3 N1 V2 M0	3	Cecum	-	Colon

Table S1 - Characteristics of patient samples used for scRNAseq. n.a = not applicable, n.d. = no data.

Patient	Diagnosis	Site	Severity	Vedolizumab
1	CD	Colon	Active- mild	YES
1	CD	Colon	Quiescent	YES
1	CD	Ileum	Quiescent	YES
2	CD	Colon	Active- mod	
2	CD	Colon	Quiescent	
2	CD	Ileum	Active- mild	
3	CD	Colon	Active- mild	
3	CD	Colon	Quiescent	
4	CD	Colon	Active- mild	
4	CD	Colon	Active- mild	
4	CD	Colon	Active- mod	
5	IBDU (CD)	Colon	Active- mild	
5	IBDU (CD)	Colon	Quiescent	
6	CD	Colon	Active-mod	
7	IBDU (CD)	Colon	Active- mild	
7	IBDU (CD)	Colon	Quiescent	
8	UC	Colon	Active- mod	
8	UC	Colon	Quiescent	
9	UC	Colon	Active- mod	YES
9	UC	Colon	Quiescent	YES
9	UC	Ileum	Quiescent	YES
10	CD	Colon	Quiescent	
10	CD	Ileum	Active- mild	
11	IBDU (UC)	Colon	Quiescent	
11	IBDU (UC)	Ileum	Quiescent	
12	UC	Colon	Quiescent	
13	UC	Colon	Active- mod	
13	UC	Colon	Quiescent	
14	Healthy	Colon	Healthy	
14	Healthy	Ileum	Healthy	
15	Healthy	Colon	Healthy	
16	Healthy	Colon	Healthy	
16	Healthy	Ileum	Healthy	
17	Healthy	Colon	Healthy	
17	Healthy	Ileum	Healthy	
18	UC	Colon	Active- mild	
18	UC	Ileum	Active-mild	
19	CD	Colon	Quiescent	
20	CD	Colon	Active- mild	
20	CD	Ileum	Active- mod	
21	UC pan	Colon	Active- mod	YES
21	UC pan	Ileum	Quiescent	YES
22	UC pan	Colon	Quiescent	
22	UC pan	Ileum	Quiescent	
23	Healthy	Colon	Healthy	
23	Healthy	Ileum	Healthy	
24	Healthy	Colon	Healthy	
25	CD	Ileum	Quiescent	
25	CD	Ileum	Quiescent	
26	UC	Colon	Quiescent	YES
26	UC	Colon	Quiescent	YES

26	UC	Rectum	Active- mild	YES

Flurophore	Target	Clone
BUV737	CD38	HB-7
BUV393	CD45	HI30
BV786	CD1c	F10/21A3
BV711	CD141	1A4
BV650	CD163	
BV605	CD103	B-Ly7
BV480	CD1a	
BV421	CD14	MφP9
BB700	CD123	7G3
BB515	CYTOX	F10/21A3
PE/Cy7	CD55	JS11
PE/Cy5	CD206	15-2
PE/CF594	CD3	UCHT1
PE/CF594	CD19	HIB19
PE	CD5	UCHT2
APCFire750	CD11a	TS2/4
AF700	HLADR	G46-6
APC	CD207	10E2

Flurophore	Target	Clone
BUV393	CD45	HI30
BV650	HLADR	G46-6
FITC	via	via
PE/CF594	CD19	HIB19
AF647	CD3	UCHT1

Manuscript II

Transcriptomic characterization of mononuclear phagocytes from human gut-associated lymphoid tissues

Transcriptomic characterization of mononuclear phagocytes from human gut-associated lymphoid tissues

Line Wulff[1], Thomas M. Fenton[1,2], Urs M. Mörbe[1], Peter B. Jørgensen[1], Fredrik Junghus[1], Julien Vandamme[1], Flemming Bendtsen[3], Ole H. Nielsen[4], Allan M. Mowat[5], William W. Agace[1,6]

Affiliations

[1]Mucosal Immunology group, Department of Health Technology, Technical University of Denmark, Kemitorvet, 2800 Kgs. Lyngby, Denmark
[2]University of Edinburgh Centre for Inflammation Research, Queens Medical Research institute, Edinburgh, EH16 4TJ, UK
[3]Gastro Unit, Medical Division, Amager-Hvidovre University Hospital, Hvidovre, Denmark.
[4]Department of Gastroenterology, Medical Section, Herlev Hospital, University of Copenhagen, 2730 Herlev, Denmark.
[5]Institute of Infection, Immunity and Inflammation, College of Medicine, Veterinary Medicine and Life Sciences, University of Glasgow, Glasgow, UK
[6]Immunology Section, Lund University, BMC D14, 221-84 Lund, Sweden

Abstract

Mononuclear phagocytes (MNPs) play crucial roles in intestinal immunity, with conventional dendritic cells (cDC) being important for initiating the response and macrophages having several effector functions, as well as being involved in maintaining tissue homeostasis. Immune responses in the intestine are initiated in the gut associated lymphoid tissues (GALT), whereas the lamina propria (LP) of the mucosa is where these responses are expressed. Although MNP are found in both locations, the exact composition of the relevant populations is not well understood in humans, partly because it was impossible to isolate cells from each site without contamination from the other. Here we have exploited our novel technique to perform unbiased scRNA-seq analysis of MNPs from precisely defined samples of LP and GALT from healthy human intestine. We show that there are significant differences in the distribution and transcriptomes of cDC subsets and macrophages in the inductive sites and LP of both small intestine and colon, suggesting distinct roles in the two locations. Preliminary analysis indicates that these features are also altered in the presence of active inflammation during Crohn's disease (CD). Thus it is essential to analyse inductive and effector sites separately when assessing MNP populations in the intestine.

Introduction

Immune tissues in the intestine can be divided into effector or inductive sites. The immune effector compartments consist of the mucosal lamina propria and its overlying epithelium, where innate and primed adaptive immune cells cooperate to promote epithelial integrity, protective immunity and to maintain tolerance against food antigens and commensal microbes[1]. The inductive sites are the mesenteric lymph node and the gut-associated lymphoid tissues (GALT) which are secondary lymphoid organs residing within the wall of the intestine itself and are where T and B lymphocytes are primed by antigen. The GALT include the macroscopically visible Peyer's patches (PP) in the small intestine (SI) and the much more numerous isolated lymphoid follicles (ILF), which are located either in the mucosa (M-ILF) or mostly in the submucosa (SM-ILF)[2].
Mononuclear phagocytes (conventional dendritic cells (cDC) and macrophages) are abundant both in the GALT and mucosa, each playing crucial, but distinct roles in intestinal immunity. Whereas cDC are primarily responsible for initiating immunogenic and tolerogenic immune responses[3–5], macrophages are involved as effector cells in protective immunity and inflammation, as well as being important for

maintaining tissue homeostasis[4,6]. However the exact roles of each MNP population in the different immune compartments of the intestine remain uncertain, particularly in humans. In part, this reflects their shared expression of many phenotypic markers, together with the increasing awareness of the complex heterogeneity within each population[7–10]. Of equal importance is the fact that it has been impossible to isolate cells from GALT and mucosa without contamination from the other site. While PPs are easily seen and excluded by the naked eye, this cannot be done with the smaller and much more numerous ILFs that are present in all samples of mucosa obtained using standard techniques. Recently, we have published a novel method which allows LP, M-ILF and SM-ILF to be isolated separately from samples of human intestine[11] and here we have exploited this technique to characterize and compare the MNP populations found in healthy and Crohn's disease GALT and mucosa using unbiased scRNA-seq analysis. Our findings show clear differences in the MNP subsets present in the two compartments and underline the need to examine these tissues separately.

Results

Characterization of MNP populations in intestinal LP and GALT

To characterize the MNPs in the different immune compartments of the steady state human intestine, we isolated cells from resections of terminal ileum and colon taken from uninvolved areas during surgery for colon cancer. MNPs were sorted as $CD45^+HLADR^+CD3^-CD19^-$ and single cell libraries were prepared for scRNA-seq analysis with the 10x Chromium platform. To study MNPs in fine detail, we first bioinformatically excluded remaining B cells, stromal cells, endothelial cells, and T cells, resulting in a data set consisting of 45,987 MNPs pooled from five patients and containing paired samples of SM-ILF from colon, colonic LP (cLP) and small intestinal LP (SILP); three of these samples also included PP (Fig. 1A). High resolution clustering was used to guide lower semi-supervised clustering and identified 14 MNP clusters. Assessment of their differentially expressed genes (DEG) showed these 14 clusters comprised four clusters of mature macrophages identified by *C1Q* gene expression, a *CD14*$^+$ monocyte cluster and a cluster of *CD16*$^+$ non-classical monocytes. There were also two clear clusters of cDC1 and plasmacytoid dendritic cells (pDC), together with a larger, heterogeneous group of cells expressing *CD1C* and lacking *CD14* (Fig. 1A, Sup. Fig 1A). Based on our recent work on human LP, we could identify these as cDC2 and cDC3, based on theie mutually exclusive expression of *CD207* and *CD163* respectively (Fig. 1A, Sup. Fig. 1B)[7,12,13]. There were three cDC3 clusters and one cDC2 cluster (Fig 1A, Sup. Fig. 1C). A cluster of migratory (mig) cDC was also identified by its expression of *CCR7* and *LAMP3*. Finally, there was a small cluster of *CD207*$^-$ cells close to the cDC2/cDC3 clusters that expressed DEGs similar to those we found expressed by putative pre-cDC in LP (Fig. 1A, Sup. Fig. 1A-C)[7].

The MNP subsets were distributed differently between tissues, with mature macrophages being relatively enriched in SILP, cDCs in PP and pDCs in SM-ILFs. In contrast, cLP showed a more even distribution of the major populations (Fig. 1B-C). These tissue-specific patterns were consistent across the patient samples, although the exact proportions of the MNP subsets varied between individual samples (Sup. Fig. 1D). When we focused on the monocyte and macrophage populations specifically, it was clear that mature macrophages were more highly represented in the SILP, whereas the PP had consistently higher numbers of *CD14*$^+$ monocytes (Fig. 1D). A similar difference was seen when comparing the cLP to SM-ILF, with *CD14*$^+$ monocytes consistently accounting for more than 50% of the monocyte and macrophage compartment in the SM-ILFs (Fig. 1D). In the cDC compartment, PP consistently had more putative pre-cDCs and fewer migratory cDC than the paired SILP. Similarly, the cLP had more migratory cDCs than SM-ILF, with more putative pre-cDCs in SM-ILF. There were also higher proportions of cDC1 amongst cDC in ILF (Fig. 1E).

Together these data indicate that all the monocyte, macrophage and cDC populations can be found in the GALT and LP of both SI and colon, but that the populations have different distributions between the tissues.

Transcriptional differences between cDC1 and cDC2 in LP and GALT

We next explored whether the cDC1 and cDC2 subsets in GALT and LP compartments might show transcriptional differences. To address this, we identified DEGs that differed between matched samples for the three patients where we had tissues from all compartments and used only those DEGs that were significantly different (p-adjusted < 0.05 and average log fold change > 0.2) in every patient. This analysis was performed separately for cDC1 and cDC2 and was used to compare SI LP with PP, as well as colon LP versus SM-ILF (Fig. 2A-D).

29 DEGs distinguished between cDC1 in PP and SILP, with 8 of these being specific to PP and 21 to SILP. There was a larger number of DEGs distinguishing cDC1 in SM-ILF and cLP, of which 43 were specific to SM-ILF cDC1 and 50 to cLP (Fig. 2A-B). Interestingly, one of the cDC1 genes specific to SM-ILF was *LYZ,* which codes for lysozyme and has been shown previously to be expressed by a unique population of monocyte-derived, cDC-like MNP in mouse and human PP[14] (Fig. 2B, Sup. Fig. 2A).Only 4 DEGs distinguished between cDC2 in PP and SILP, all of which were specific to PP (Fig. 2C, Sup. Fig. 2B). One of these was *IL22RA2*, which codes for IL22 binding protein (IL22BP), which is expressed by a subset of cDC2 in murine SI and by the lysozyme expressing, monocyte-derived cells in human PP[10,15]. Again, there were more DEGs separating cDC2 in cLP and SM-ILF, with 168 being specific to cLP and 65 to SM-ILF (Fig. 2D, Sup. Fig. 2B). To explore the possible functional significance of these transcriptional differences, we used EnrichR with Biological Process 2021 to create a list of related gene ontology (GO) terms, focusing on pathways of immunological relevance[16–18]. GO terms relating to innate recognition of microbes were a common feature of cDC1 in both SI and colonic LP (Fig. 2E). Converserly, cDC1 in SM-ILF showed terms related to positive regulation of $CD4^+$ T cell proliferation and differentiation (Fig. 2E). cDC2 in both small intestinal and colonic LP showed expression of pathways involved in antigen presentation and T cell activation, whereas GO terms for cDC2 in SM-ILF were related to regulation of lymphocyte differentiation and TGFβ response (Fig. 2F). It was not possoble to generate precise information of this kind for cDC1 and cDC2 in PP due to the low number of DEGs and there were few if any GO terms of direct immunological relevance (Fig. 2E, F).

Finally, we explored whether cDC1 and cDC2 in the GALT compartments of small intestine and colon shared DEGs that distinguished them from their counterparts in the effector sites of the equivalent gut segment. For cDC1, most DEGs that distinguished GALT from their respective LP were exclusive to either PP or SM-ILF, with only two DEGs being shared (Fig. 2G). Similarly, cDC1 from SI and colonic LP shared only three of the DEGs that distinguished them from their equivalent GALT (Fig. 2G). Interestingly, both LP and ILF from colon showed larger numbers of unique DEGs compared with the equivalent tissues in the small intestine. There were no shared genes when cDC2 from SM-ILF and PP were compared with their effector sites, although again this could reflect the low amounts of DEGs found in PP. Even more than with cDC1, the vast majority of DEGs that were selective to cDC2 from colonic GALT or LP were unique to their tissue of origin, although cDC2 from cLP and SILP shared 16 DEGs that distinguished them from SM-ILF or PP respectively. These commonalities included *LGALS1/3* which code for galectins 1 and 3 and which are associated with tolerogenic DC[19,20], together with *FCGR2B*, *FCER1A*, *C1Q* genes and *CXCL8* (Fig. 2H).

Together these observations suggest that there are functional differences between the cDC subsets found in the LP and GALT from the small and large intestine, many of which are reflect their localization in inductive or effector compartments. Furthermore, the results indicate that most genes

that differentiate cDC in GALT or LP from their counterparts in the opposite compartment are unique to either the SI or colon.

Behaviour of intestinal MNP populations during Crohn's disease

The distribution of MNP populations in LP changes during colitis, mostly due to an influx of CD14$^+$ monocytes from the bloodstream[21,22]. However we have recently found that the SM-ILFs increase in size in patients with active Crohn's disease (CD), indicating that changes can occur in both inductive and effector sites during inflammation[11]. Given the fact that the samples of "LP" used in earlier studies were probably contaminated by ILFs, here have exploited our ability to examine these sites separately to investigate how MNPs might change in SM-ILFs and colonic LP. To do this, we conducted a preliminary study of one CD patient in whom we could obtain samples of uninvolved and inflamed tissue (Sup. Fig. 3A). To this end, we integrated data from MNPs obtained from the CD patient with those we had obtained from the healthy SM-ILFs and colonic LP as described above (Fig. 1). By applying the same meta data, we could re-identify all clusters we had found previously, except for two of the macrophage populations which appeared to be interspersed with two of the original cDC3 clusters (5 and 7) and with the two macrophage clusters that could be distinguished (1 and 3)(Fig. 3A, Sup. Fig. 3A, B). An additional cluster that was present only when inflamed MNPs were included had a high mitochondrial gene content and is probably debris[23,24]. The distribution of monocytes, macrophages and cDCs in uninvolved colonic LP was very similar to that in healthy colonic LP, although there was some evidence of more CD14$^+$ monocytes in the uninvolved tissue (Fig. 3B-E). In contrast, the LP of inflamed colon showed a dramatic change in the distribution on MNPs, with a large increase in a cluster of monocytes which was rarely found in either healthy or uninvolved LP. Together with the original population of CD14$^+$ monocytes, these comprised more than 70% of the monocyte/macrophage subsets in inflamed colonic LP (Fig 3D). In parallel, there was a decrease in the proportion of cDC amongst total MNP in inflamed LP, but no change in the frequency of macrophages (Fig. 3B, C). Consistent with our recent findings[7], cDC3 in in cluster 7 appeared to make up a larger proportion of cDC in inflamed LP (Fig. 3E). Unfortunately too few MNPs were obtained from uninvolved or inflamed ILFs (124 and 702 cells respectively) to allow analysis of the scRNA-seq data from these tissues (Fig. 3B).

Studies in mice have shown that monocytes recruited to the intestine undergo local differentiation into macrophages through a continuum known as the monocyte "waterfall" and that this process is altered during inflammation[25,26]. Under these circumstances, the biggest change is expansion of a population of inflammatory monocytes that are somewhat more mature than those that have recently left the bloodstream[21,26,27]. We therefore hypothesised that the novel population of inflammatory monocytes we had found in human Crohn's disease was analogous to the subset that expands preferentially in experimental colitis. To examine this idea, we used a gene signature obtained from human CD14$^+$ blood monocytes[28] to generate a score for the monocytes and macrophages pooled from healthy, uninvolved and inflamed LP. We also included cDC3 in this analysis, as these also appear to increase in IBD and they share some features of monocytes[8,13]. This showed that although the inflammatory monocytes had a lower score than CD14$^+$ monocytes, this was higher than for any of the cDC3 or macrophage subsets (Fig. 3F). Furthermore, the inflammatory monocytes and CD14$^+$ monocytes expressed equally low levels of *FCER1A*, which was expressed by the cDC3 clusters (Fig. 3G). Together these findings indicate that the novel population of inflammatory MNPs is related to monocytes rather than DC.

The inflammatory nature of these novel monocytes was confirmed by DEG analysis of the monocyte and macrophage populations in healthy and inflamed colonic LP, which showed the expression of many inflammatory genes including the chemokines *CXCL1*, *CXCL9* and *CXCL10*, together with IFNγ inducible genes such as *IDO1*, *GPB1*, *GPB2* and *GPB5* (Fig. 3H, Sup. Fig. 3C). This also aligned the

inflammatory cluster with a recently described small population of pro-inflammatory monocytes found in healthy colonic mucosa suggested to represent a monocyte undergoing an distinctive pro-inflammatory trajectory[9]. Notably, the expression of the DEGs associated with these monocytes correlated with the level of inflammation, being most highly expressed in inflamed tissue, slightly lower in uninvolved samples and at very low levels in cells from the healthy controls (Sup. Fig. 3C).

Together, this preliminary study suggests that a population of inflammatory monocytes expands in CD. As well as being distinct from the classical monocytes found in blood and from cDC, their inflammatory profile may allow them to contribute to the pathology of CD.

Discussion

This study represents the first comparison of the MNP compartments found within the inductive and effector sites of the human intestine without cross-contamination of the two tissues. Our results show that whereas the effector sites of small and large intestinal LP are enriched in cells of the monocyte-macrophage lineage, the inductive sites represented by PP and SM-ILF contain a high proportion of cDCs. These findings are consistent with the known roles of cDCs and macrophages in the iniation and expression of immune responses.

In addition to the overall abundance of cDCs in GALT, the SM-ILF cDC1 in this location had differential expression of *LYZ* compared with cDC1 in the colonic LP. *LYZ* was also expressed in certain subsets of macrophages and cDC3s. As well as being a typical product of monocytes and macrophages, recent work has identified an unusual subset of monocyte-derived DC (lyso-DC) in mouse and human PP which are believed to have unique roles in sampling and presenting local antigen [14,15,29,30]. Other studies have shown that lyso-DC and a subset of $CD11b^+$ cDC2 express IL22BP in both ILF and PP[10,14]. Here, we could identify a small subset of cDC2 expressing *IL22RA2* (coding for IL22BP) which was specific to PP, although these did not express *LYZ*. As IL22BP is thought to regulate the ability of IL22 to induce epithelial cell turnover in the intestine, these results could indicate that subsets of cDC2 may contribute to the homeostasis of the follicle associated epithelium that overlies human GALT. An important feature of our work was that we could identify the recently characterised subset of cDC3 in both LP and GALT. These DCs have similarities to both monocytes and macrophages, but appear to be distinct from either and develop from a separate precursor[8,12]. Their precise contribution to immune responses remains uncertain[8,13,31], but the current findings indicate that their functions must extend to both lymphoid and non-lymphoid tissues in the intestine. Further studies are needed to fully characterize cDC3 in different anatomical compartments and to explore the division of labour between the various cDC subsets in local immunity.

An interesting finding from our analysis of DEGs in cDCs was that there were many more differences between cDC1 and cDC2 in colonic LP and their counterparts in colonic SM-ILF than there were when comparing the same subsets in SILP and PP. The reasons for this are unclear, but one possibility could be that the environments and immunological pressures are more distinct for the inductive and effector compartments in the colon than in the small intestine. Alternatively, there may have been more cells from LP contaminating our PP samples than in SM-ILF. There are always small amounts of LP present in the interfollicular regions of PP which are impossible to separate during processing, whereas SM-ILF can be punched out from the surrounding mucosa without significanct contamination. This could be solved by cutting out the individual follicles from the PP, and thus avoiding the contaminating interfollicular regions.

As expected, we found that mature macrophages were much more abundant in the the LP of both colon and small intestine than in the equivalent GALT, consistent with the role of these cells in the effector phase of immunity and in maintaining tissue homeostasis. However a surprising finding was

that up to 50% of the monocyte/macrophage lineage in the healthy GALT sites were CD14$^+$ monocytes. This was particularly the case for SM-ILF and is quite different to the steady state LP, where monocytes are rare. The reasons for this are unclear and it is unknown how monocytes might contribute to the function of normal secondary lymphoid organs. One possibility which will be important to exclude is that these monocytes are derived from the bloodstream rather than the GALT parenchyma itself. That this could be a confounding factor is suggested by our recent findings that SM-ILF are particularly rich in blood vessels[32]. Having gated out neutrophils and lymphocytes from the MNPs we sorted for analysis, we could not use the numbers of these cells as indicators of blood contamination and therefore this issue will require more direct attention using eg immunohistochemistry or detailed transcriptional comparison of the CD14$^+$ monocytes in blood and GALT.

As has been shown in experimental models of colitis and in human IBD using flow cytometry[21,26], we observed a large increase of monocytes in inflamed LP from the patient with Crohn's disease we examined. Although some of these were similar to the CD14$^+$ monocytes seen in healthy GALT, the majority was distinct from blood monocytes and expressed higher amounts of inflammatory genes. Closer analysis showed that small numbers of these "inflammatory monocytes" were also present in healthy LP and in uninvolved mucosa in Crohn's disease. Together these findings are reminiscent of the population of monocyte-derived cells found in the monocyte waterfall in healthy mucosa of mice and humans ("P2") and which has been proposed to represent an intermediary stage in the process of differentiation that allows continuous replenishment of mucosal macrophages from blood monocytes [22,26,33]. As we confirm here, this subset predominates during inflammation and recent work suggests that this accumulation represents an alternative trajectory of differentiation to that seen in healthy intestine[9,34]. This idea could explain the clear differences we found between "inflammatory monocytes" and conventional CD14$^+$ monocytes, but confirmation will require more detailed comparison of all monocyte subsets in healthy and inflamed intestine to establish the develoomental relationship between them. Interestingly, the accumulation of monocytes in active Crohn's disease was accompanied by expansion of one of the clusters of cDC3 in the inflamed cLP, whereas the distribution of cDC1 and cDC2 was not influenced by the presence of inflammation. This finding serves to underline the similarities between cDC3 and the monocyte/macrophage lineage and further analysis is required to confirm the nature of the cDC3 that expands in inflammation and its relationship to the other populations of MNP under these circumstances.

Together our work emphasises the differences between the cDC and monocyte/macrophage populations present in different immune compartments of the intestine underlines the need to examine these tissues separately when analysing local immune cells and their functions

Author contribution

The study was designed by L.W., T.M.F. and W.W.A. Patient resection material was provided by F.B. and O.H.N. The experiments were performed by T.M.F., U.M.M., P.B.J., and J.V., while the bioinformatic analysis was performed by L.W. with assistance from F.J. The manuscript was written by L.W. and A.M. with input from all authors

Acknowledgements

The authors thank the staff and patients at the Gastro Unit at Amager-Hvidovre University Hospital and from the Department of Gastroenterology at Herlev Hospital. The sequencing was performed at the National Genomics Infrastructure (NGI), Science for Life Laboratory SNP&SEQ Technology Platform in Uppsala. The work was supported by grants from Lundbeck foundation and the Helmsley Charitable Trust.

References

1. Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* **14**, 667–685 (2014).
2. Mörbe, U. M. *et al.* Human gut-associated lymphoid tissues (GALT); diversity, structure, and function. *Mucosal Immunol.* **14**, 793–802 (2021).
3. Watchmaker, P. B. *et al.* Comparative transcriptional and functional profiling defines conserved programs of intestinal DC differentiation in humans and mice. **15**, (2014).
4. Luciani, C., Hager, F. T., Cerovic, V. & Lelouard, H. Dendritic cell functions in the inductive and effector sites of intestinal immunity. *Mucosal Immunol.* **15**, 40–50 (2022).
5. Joeris, T., Müller-Luda, K., Agace, W. W. & Mowat, A. M. I. Diversity and functions of intestinal mononuclear phagocytes. *Mucosal Immunol.* **10**, 845–864 (2017).
6. Viola, M. F. & Boeckxstaens, G. specific functional heterogeneity of intestinal resident macrophages. 1383–1395 (2021) doi:10.1136/gutjnl-2020-323121.
7. Fenton, T. M. *et al.* Single-cell characterisation of mononuclear phagocytes in the human intestinal mucosa. *Unpublished*.
8. Bourdely, P. *et al.* Transcriptional and Functional Analysis of CD1c+ Human Dendritic Cells Identifies a CD163+ Subset Priming CD8+CD103+ T Cells. *Immunity* **53**, 335-352.e8 (2020).
9. Domanska, D. *et al.* Single-cell transcriptomic analysis of human colonic macrophages reveals niche-specific subsets. *J. Exp. Med.* **219**, e20211846 (2022).
10. Silva, C. Da, Wagner, C., Bonnardel, J., Gorvel, J. & Lelouard, H. The Peyer ' s Patch Mononuclear Phagocyte System at Steady State and during Infection. **8**, (2017).
11. Jørgensen, P. B. *et al.* Identification, isolation and analysis of human gut-associated lymphoid tissues. *Nat. Protoc.* **16**, (2021).
12. Cytlak, U. *et al.* Differential IRF8 Transcription Factor Requirement Defines Two Pathways of Dendritic Cell Development in Humans. *Immunity* **53**, 353-370.e8 (2020).
13. Dutertre, C. A. *et al.* Single-Cell Analysis of Human Mononuclear Phagocytes Reveals Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity* **51**, 573-589.e8 (2019).
14. Wagner, C. *et al.* Differentiation Paths of Peyer ' s Patch LysoDCs Are Linked to Sampling Site Positioning , Migration , and T Article Differentiation Paths of Peyer ' s Patch LysoDCs Are Linked to Sampling Site Positioning , Migration , and T Cell Priming. *Cell Rep.* **31**, (2020).
15. Martin, J. C. *et al.* Interleukin-22 binding protein (IL-22BP) is constitutively expressed by a subset of conventional dendritic cells and is strongly induced by retinoic acid. *Mucosal Immunol.* **7**, 101–113 (2014).
16. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
17. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016

update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
18. Moiyadi, A. V. & Sridhar, E. δ-Aminolevulinic acid-induced fluorescence unmasks biological intratumoral heterogeneity within histologically homogeneous areas of malignant gliomas. *Acta Neurochir. (Wien).* **157**, 617–619 (2015).
19. Poncini, C. V. *et al.* Trypanosoma cruzi Infection Imparts a Regulatory Program in Dendritic Cells and T Cells via Galectin-1–Dependent Mechanisms . *J. Immunol.* **195**, 3311–3324 (2015).
20. Ilarregui, J. M. *et al.* Tolerogenic signals delivered by dendritic cells to T cells through a galectin-1-driven immunoregulatory circuit involving interleukin 27 and interleukin 10. *Nat. Immunol.* **10**, 981–991 (2009).
21. Jones, G. *et al.* Dynamics of Colon Monocyte and Macrophage Activation During Colitis. **9**, (2018).
22. Tamoutounour, S. *et al.* CD64 distinguishes macrophages from dendritic cells in the gut and reveals the Th1-inducing role of mesenteric lymph node macrophages during colitis. *Eur. J. Immunol.* **42**, 3150–3166 (2012).
23. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 1–15 (2016).
24. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, (2019).
25. Tamoutounour, S. *et al.* CD64 distinguishes macrophages from dendritic cells in the gut and reveals the Th1-inducing role of mesenteric lymph node macrophages during colitis. *Eur. J. Immunol.* **42**, 3150–3166 (2012).
26. Bain, C. C. *et al.* Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6C hi monocyte precursors. **6**, (2013).
27. El Sayed, S. *et al.* CCR2 promotes monocyte recruitment and intestinal inflammation in mice lacking the interleukin-10 receptor. *Sci. Rep.* **12**, 1–12 (2022).
28. Tang-Huau, T. L. *et al.* Human in vivo-generated monocyte-derived dendritic cells and macrophages cross-present antigens through a vacuolar pathway. *Nat. Commun.* **9**, (2018).
29. Bonnardel, J. *et al.* Gene expression profiling of the Peyer's patch mononuclear phagocyte system. *Genomics Data* **5**, 21–24 (2015).
30. Bonnardel, J. *et al.* Distribution, location, and transcriptional profile of Peyer's patch conventional DC subsets at steady state and under TLR7 ligand stimulation. *Mucosal Immunol.* **10**, 1412–1430 (2017).
31. Brown, C. C. *et al.* Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell* **179**, 846-863.e24 (2019).
32. Fenton, T. M. *et al.* Immune Profiling of Human Gut-Associated Lymphoid Tissue Identifies a Role for Isolated Lymphoid Follicles in Priming of Region-Specific Immunity. *Immunity* **52**, 557-570.e6 (2020).
33. Bujko, A. *et al.* Transcriptional and functional profiling defines human small intestinal macrophage subsets. (2018).
34. Desalegn, G. & Pabst, O. Inflammation triggers immediate rather than progressive changes in monocyte differentiation in the small intestine. *Nat. Commun.* **10**, 3229 (2019).
35. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
36. Li, Jun; Barron, M. ccRemover: Removes the Cell-Cycle Effect from Single-Cell RNA-Sequencing Data. *CRAN* (2017).
37. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, (2018).

Legends

Figure 1:

A) UMAP of total MNPs pooled from SILP (n=5), cLP (n=5), PP (n=3) and SM-ILF (n=5) of resected intestine after removal of contaminating stromal cells, endothelial cells, B cells and T cells. Color coding indicates individual populations of MNP identified by semi-supervised clustering.
B) Contoured distribution of the UMAP embedded clusters shown in A across different anatomical sites of the intestine. Numbers show the total number of MNPs analysed in each tissue.
C) Distribution of the individual clusters shown in A above in paired SILP and PP, cLP and ILF, illustrating the contribution of the monocyte/macrophage, cDC and pDC lineages. The results show the proportion of each cluster amongst total MNPs in each tissue.
D) Distribution of monocyte and macrophage subsets as a proportion of the total monocyte/macrophage compartment pooled from SILP, PP, cLP and SM-ILF of all patients (left panel) and amongst the total monocyte/macrophage compartment in paired tissues from individual patients (right panels).
E) Distribution of cDC1, cDC2 and cDC3 subsets as a proportion of total cDCs pooled from SILP, PP, cLP and SM-ILF of all patients (left panel) and amongst total cDCs in paired tissues from individual patients (right panels).

Figure 2:

A, B) DEGs between cDC1 in paired SILP and PP (A) and between cDC1 in paired cLP and SM-ILF (B) in patients 3, 7 and 8. The results show log. normalized scaled expression across cDC1.
C, D) Top 100 DEGs between cDC2 in paired SILP and PP (A) and between cDC2 in paired cLP and SM-ILF (B) in patients 3, 7 and 8. The results show log. normalized scaled expression across cDC2.
E, F) Enrichr analysis with Biological Process 2021 of DEGs for cDC1 (E) and cDC2 (F) from SILP, PP, cLP and SM-ILF, highlighting pathways of immunological relevance. The Y-axis denotes -log(p-adj.) for each GO term. Colored lines indicate individual tissue values for each GO term.
G, H) DEGs that distinguished between cDC1 (G) and cDC2 (H) in GALT and LP compartments of both small intestine and colon were analysed for their shared or unique expression in PP versus SILP (left panels) and SM-ILF versus cLP (right panels). The numbers shown represent the genes which were consistently differentially expressed in all three patients from whom paired tissues were available and shared genes are annotated.

Figure 3:

A) UMAP of total MNPs pooled from healthy cLP (n=5), uninvolved and inflamed CD cLP (n=1), healthy SM-ILF (n=5), uninvolved CD SM-ILF (n=1), and inflamed CD SM-ILF (n=1) of resected intestine after removal of contaminating stromal cells, endothelial cells, B cells and T cells. Colored by semisupervised clustering guided by the labelling in Fig. 1A.
B) Contoured distribution of the UMAP embedded clusters shown in A across cLP and SM-ILF in healthy controls, uninvolved CD and inflamed CD. Numbers show the total number of MNPs analysed in each tissue.
C) Distribution of monocyte/macrophages and cDC clusters amongst total MNPs in healthy, uninvolved and inflamed cLP.
D) Distribution of monocyte/macrophage subsets amongst total monocytes and macrophages in healthy, uninvolved and inflamed cLP.
E) Distribution of cDC1, cDC2 and cDC3 subsets amongst total cDC in healthy, uninvolved and inflamed cLP.
F) Module score of macrophages, monocytes and cDC3 subsets from pooled cLP from healthy, uninvolved and inflamed based on the $CD14^+$ blood monocyte signature derived by Tang-Huau et al[28].
G) Normalized expression of *FCER1A* by monocytes, macrophages and cDC3 subsets from pooled cLP of healthy, uninvolved and inflamed tissues.

H) Top 20 DEGs distinguishing monocyte and macrophage subsets in healthy, uninvolved and inflamed cLP from individual patients. The results are shown as scaled log. normalized expression across the subsets.

Supplementary Figure 1:

A) Top 5 DEGs identifying the monocyte, macrophage and DC clusters from the pooled small intestinal and colonic GALT and LP clusters shown in Figure 1A. Genes were selected based on highest average log fold change and are shown as integrated scaled gene expression on downsampled clusters.
B, C) Overlaid expression of *CD1C*, *CD14* (B), *CD163* and *CD207* (C) by the clusters embedded in the UMAP from A above
D) Distribution of MNP clusters within paired samples from SILP, PP, cLP and SM-ILF from individual patients.

Supplementary Figure 2:

A) Overlaid expression of *LYZ*, *BST2*, *LYVE1* and *IL22RA2* by the clusters embedded in the UMAP of total MNPs (see Supplementary Figure 1A), identifying subsets of macrophages and cDC2.
B) Heatmap showing remaining DEGs that distinguish cDC2 in cLP and SM-ILF and which were consistent in all 3 patients in which paired tissues were available, Data shown are scaled log. normalized between the cDC2 samples.

Supplementary figure 3:

A) UMAP of healthy control SM-ILF and cLP cells, with cell identities being transferred from Fig. 1A.
B) UMAP of all healthy, uninvolved and inflamed SM-ILF and cLP samples colored by res.1 clustering, with identities defined in A) above.
B) Histological appearance of inflamed and uninvolved colon from patient with Crohn's disease, together with healthy control colon.
C) Full list of DEGs distinguishing inflammatory monocytes from macrophage and monocyte subsets in healthy, uninvolved and inflamed cLP visualized per patient with scaled normalized expression across samples (thresholds: average logFC > 0.3 and p. adj > 0.05)

Methods

Human subjects and tissues

For scRNA-sequencing, surgical resection samples were obtained from five treatment naïve, colorectal cancer patients and one patient with active Crohn's disease after informed consent. All tissues from cancer patients were taken >10 cm distant to the cancer lesion and were histologically cancer-free as confirmed by a trained pathologist. Macroscopically inflamed as well as adjacent uninvolved resection samples were taken from the Crohn's disease patient by a trained pathologist, with the presence of inflammation being confirmed by H&E assessment of mucosal architecture, inflammatory cell infiltrates, edema and epithelial erosions. The studies were approved by the local ethical committee (Videnskabsetiske Komite for Region Hovedstaden, Denmark) and were performed under the ethical permissions H-3-2013-118 and H-20054066.

Histological analysis

Samples of fresh intestine were fixed for ≥12 h in PBS containing 4% PFA (Sigma Aldrich), washed in 70% ETOH in water and stored at 4°C until further use. For H&E staining, tissues were embedded in paraffin, sectioned into 3 µm slices, stained on glass slides with H&E and analyzed using a Panoramic Midi II slide scanner (3DHistech). Acquired images were analyzed with the software SlideViewer v2.5 (3D Histech).

Singe-cell (sc) RNA-sequencing

For scRNA-sequencing, live EPCAM$^-$, CD235ab$^-$, CD45$^+$, CD19$^-$, CD3$^-$, HLA-DR$^+$ cells were sorted using a FACSAria III cell sorter (BD Biosciences). The sorted cells were washed once in PBS containing 2% BSA and subjected to scRNA-sequencing using the 10x Chromium scRNA-seq platform. Specifically, the Single Cell 3' Library & Gel Bead Kit v3 (10x Genomics) in combination with the Chromium Chip B Single Cell Kit (10x Genomics) or the Single Cell 5' Library & Gel Bead Kit v2 (10x Genomics) in combination with the Chromium Chip K Single Cell Kit (10x Genomics) were used, following the manufacturers' instructions (10x Genomics; Chromium Single Cell 3' Reagent Kits v3 User Guide, Rev C or Chromium Next GEM Single Cell 5' v2 (Dual Index) User Guide, Rev A). cDNA amplification and purification, library construction and sample indexing were performed as recommended by the manufacturer (10x Genomics). Library purities and sizes were checked with a KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems) and the 2100 Bioanalyzer, using a High Sensitivity DNA kit (Agilent).
The final cDNA indexed libraries were pooled and sequenced either with a NextSeq 500/550 High Output v2.5 kit (150 cycles) at the Center of Excellence for Fluorescent Bioanalytics (KFB, University of Regensburg, Germany) or using the Illumina Novaseq 6000 system (200 cycles) at the SNP&SEQ Technology Platform (Uppsala, Sweden).

Analysis of 10x data

The sequencing data were aligned with CellRanger (Version 2.2.0 for patients 1, 2 & 3 and Version 6.0.1 patients 7, 8 & 9). After alignment, the individual samples were quality checked by read, gene count and mitochondrial content[23,24] (reads aligned to genenames starting with "MT-" compared with the total number of reads per cell). The lowest cut-off was 458 genes/809 reads and the highest was 4300 genes/41724 reads, while MT content had to be below 10%. The data sets were integrated using Seurat (v. 3.1.5,[35]) anchors and the variable genes for the integrated data sets were identified. These were scaled, while linearly regressing out the cell cycle effect assigned by Seurat CellCycleScoring using the genes from Li *et al.*[36] and remaining mitochondrial content. Louvain clustering, UMAP dimensionality reduction and differential gene expression of all clusters against one another (using FindAllMArkers) were used to identify contaminating immune and non-immune populations including B cells, T cells, stroma cells, ILCs and endothelial cells. These clusters were removed and the data were rescaled based on the new set of variable genes calculated only on MNPs. The rescaled data were then used to calculate new clustering and UMAP. Average expression of signature genes from blood monocytes and cDC2 and monocyte-derived macrophages from Tang-Huau et al[28] was calculated for high resolution clusters (res.3) and used as input in a principal component analysis to separate monocytes, macrophages and cDC2/cDC3 clusters. Semi-supervised clustering with these were mapped back to lower resolution (res.0.5) with R package clustree[37] and split if necessary. DEGs were calculated using FindMarker function on the log normalised data comparing SILP and PP or cLP and ILF data respectively after subsetting for each individual patient and the individual cell identity. Significance levels were set to average logFC 0.2 and p-adjusted value<0.05. DEGs were only included in the final results if they were consistent for all 3 patient samples in whom there was a complete set of paired data (pat3, pat7 and pat9). The DEGs were plotted with a modified version of DoHeatmap to allow for multiple groupings. The outputted gene lists were submitted to EnrichR[16–18] and the output tables based on Biological Processes 2021 were searched for biologically relevant terms that showed preferential overlap between the effector and inductive tissues.
A CD14+ blood mono signature derived from Tang-Huau et al.[28] was assigned with AddModuleScore.
All analyses after CellRanger, were performed in R (4.0.1) and all distributions and plot moderations were made with ggplot2 (3.3.1).

Table S1: Reagents

Reagent	Source	Identifier/Cat.no.

Paraformaldehyde (PFA)	Sigma-Aldrich (St.Louis, USA)	P6148

Kits used for scRNA-sequencing and specific components

Material	**Supplier**	**Cat. no.**
Chromium Single Cell 3’ Library & Gel Bead Kit v3	10x Genomics	PN-1000092
Chromium Single Cell 5’ Kit v2	10x Genomics	PN-1000263
Chromium Chip B Single Cell Kit	10x Genomics	PN-1000074
Chromium Chip K Single Cell Kit	10x Genomics	PN-1000287
Chromium Controller & Next GEM Accessory Kit	10x Genomics	PN-120223
Dynabeads MyOne silane	10x Genomics	PN-2000048
SPRIselect Reagent kit	Beckman Coulter	B23318
Chromium i7 Sample Index	10x Genomics	PN-220103
Dual Index Kit TT Set A	10x Genomics	PN-1000215
KAPA Library Quantification Kit for Illumina Platforms	Kapa Biosystems	KK4873
Agilent High Sensitivity DNA kit	Agilent	5067-4626

Figure 1

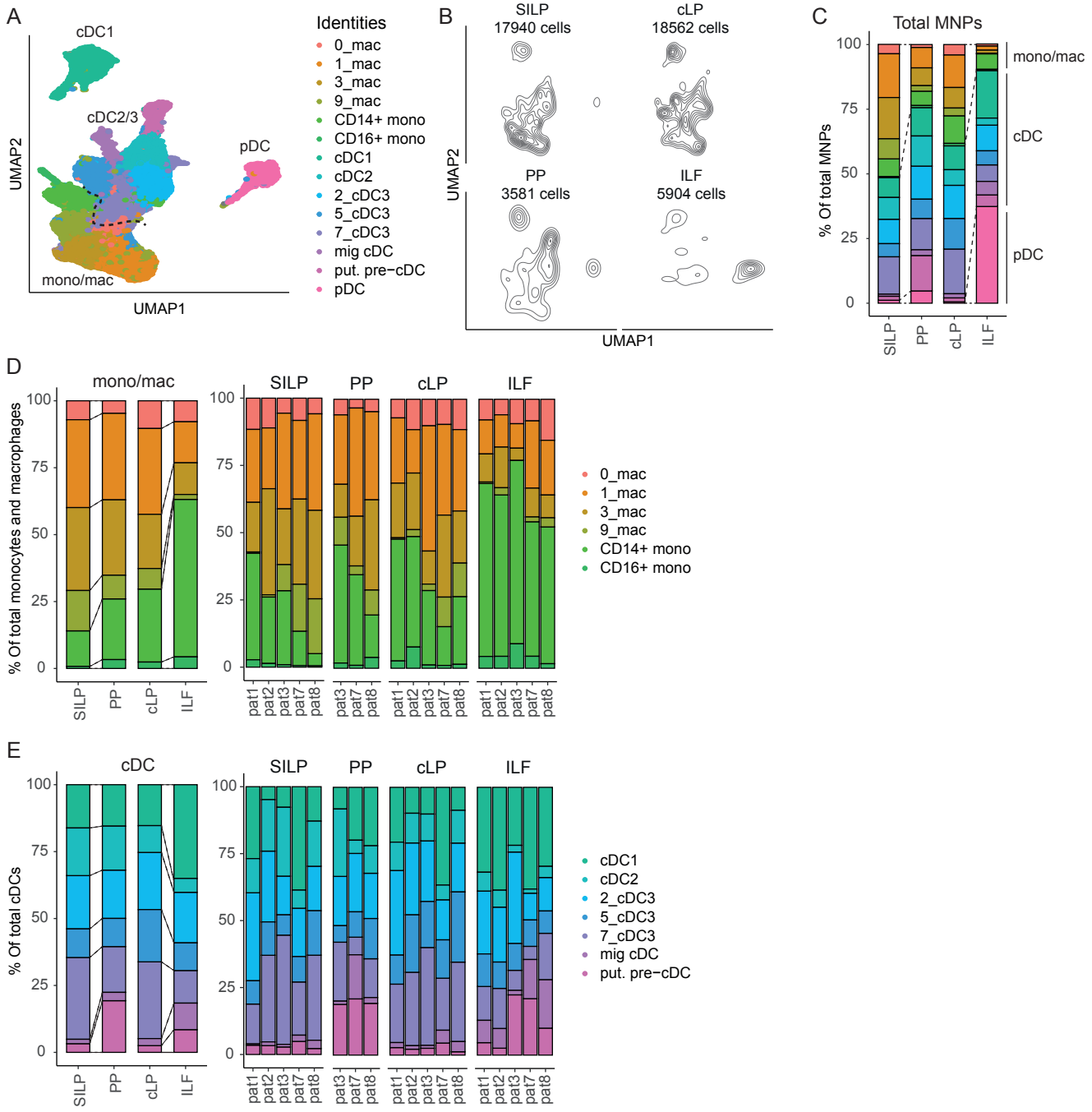

A
cDC1
cDC2/3
pDC
mono/mac
UMAP2
UMAP1
Identities
0_mac
1_mac
3_mac
9_mac
CD14+ mono
CD16+ mono
cDC1
cDC2
2_cDC3
5_cDC3
7_cDC3
mig cDC
put. pre−cDC
pDC
B
SILP
17940 cells
cLP
18562 cells
PP
3581 cells
ILF
5904 cells
UMAP2
UMAP1
C
Total MNPs
% Of total MNPs
mono/mac
cDC
pDC
SILP
PP
cLP
ILF
D
mono/mac
% Of total monocytes and macrophages
SILP
PP
cLP
ILF
pat1
pat2
pat3
pat7
pat8
0_mac
1_mac
3_mac
9_mac
CD14+ mono
CD16+ mono
E
cDC
% Of total cDCs
cDC1
cDC2
2_cDC3
5_cDC3
7_cDC3
mig cDC
put. pre−cDC

Figure 2

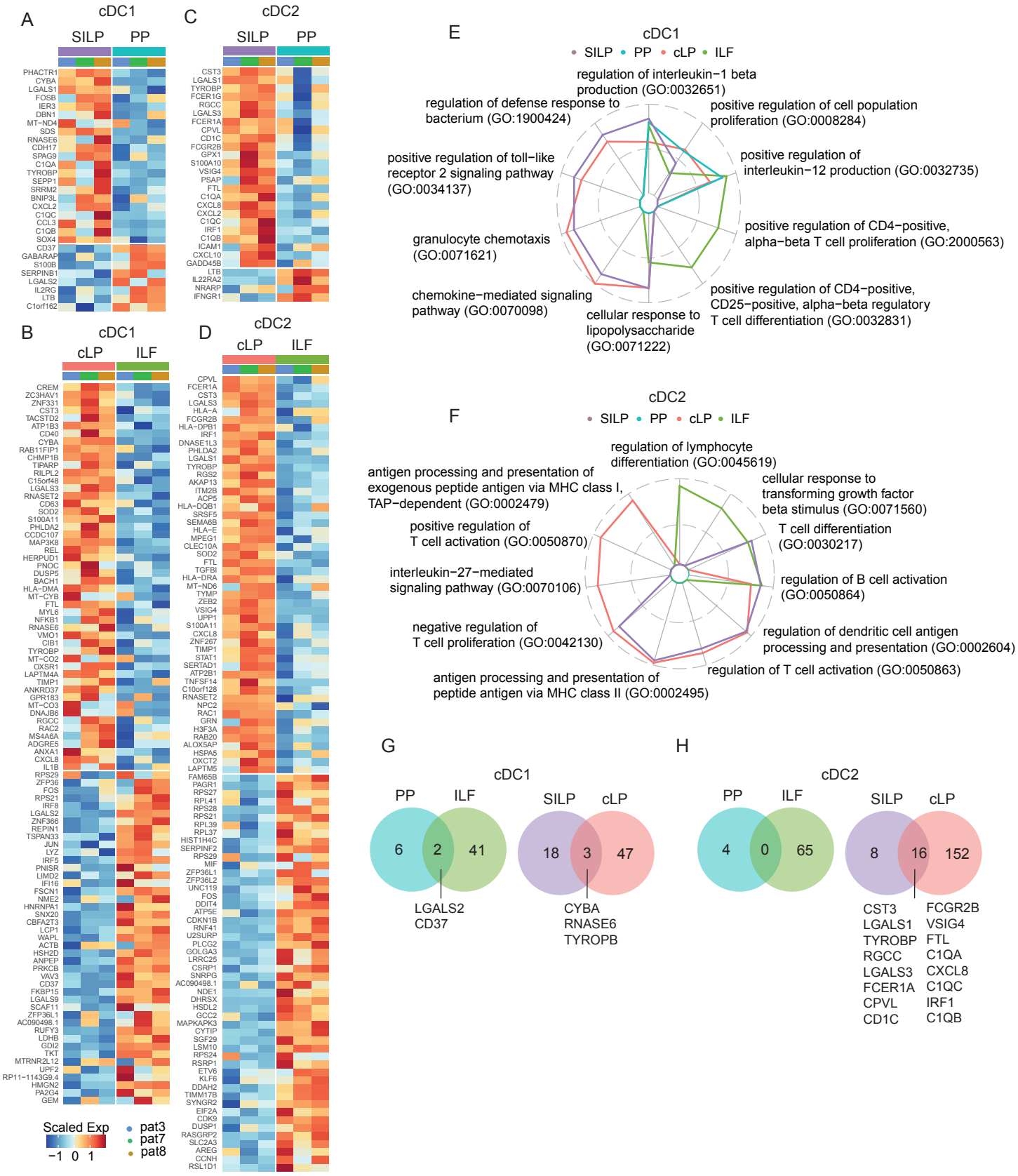
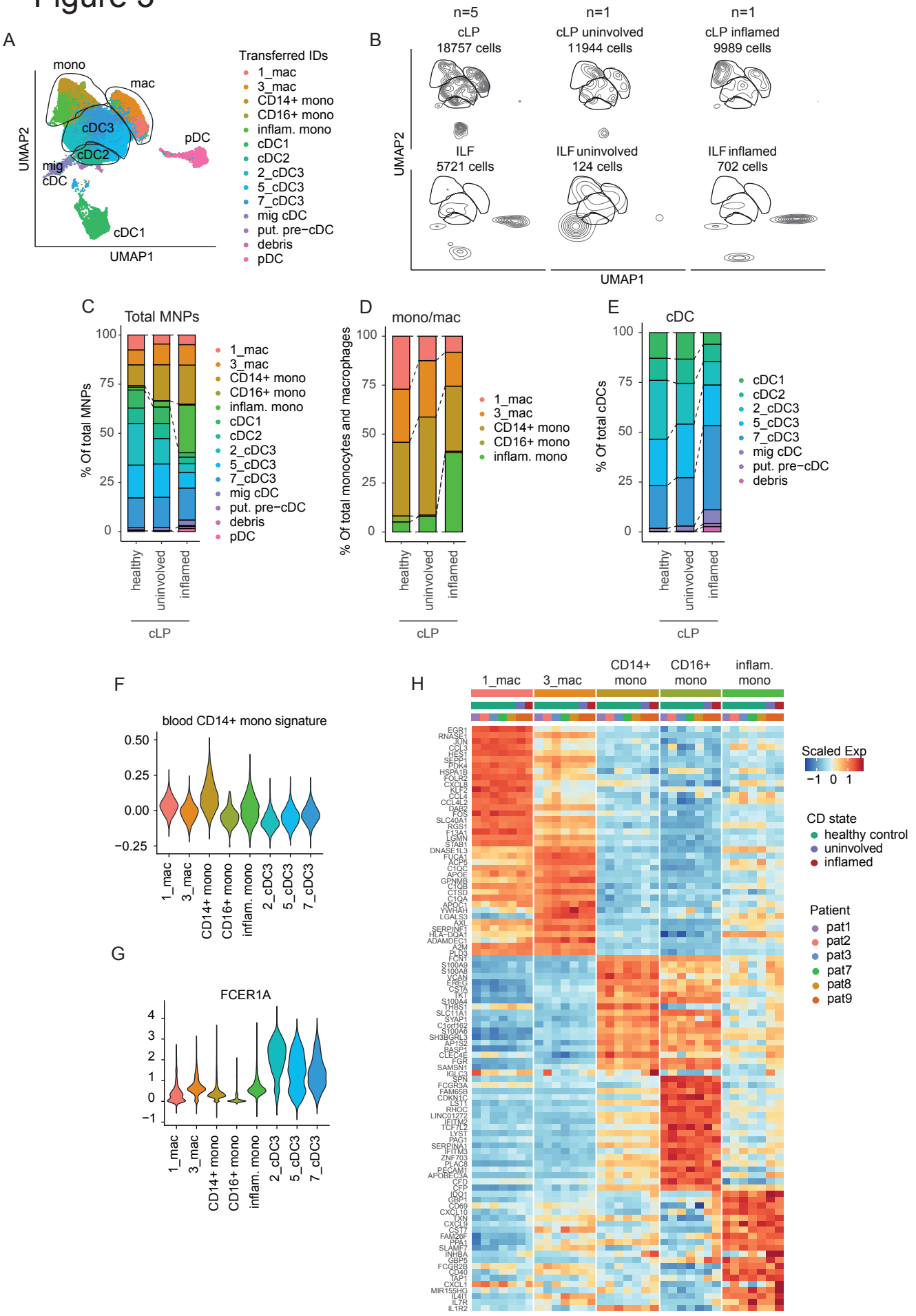
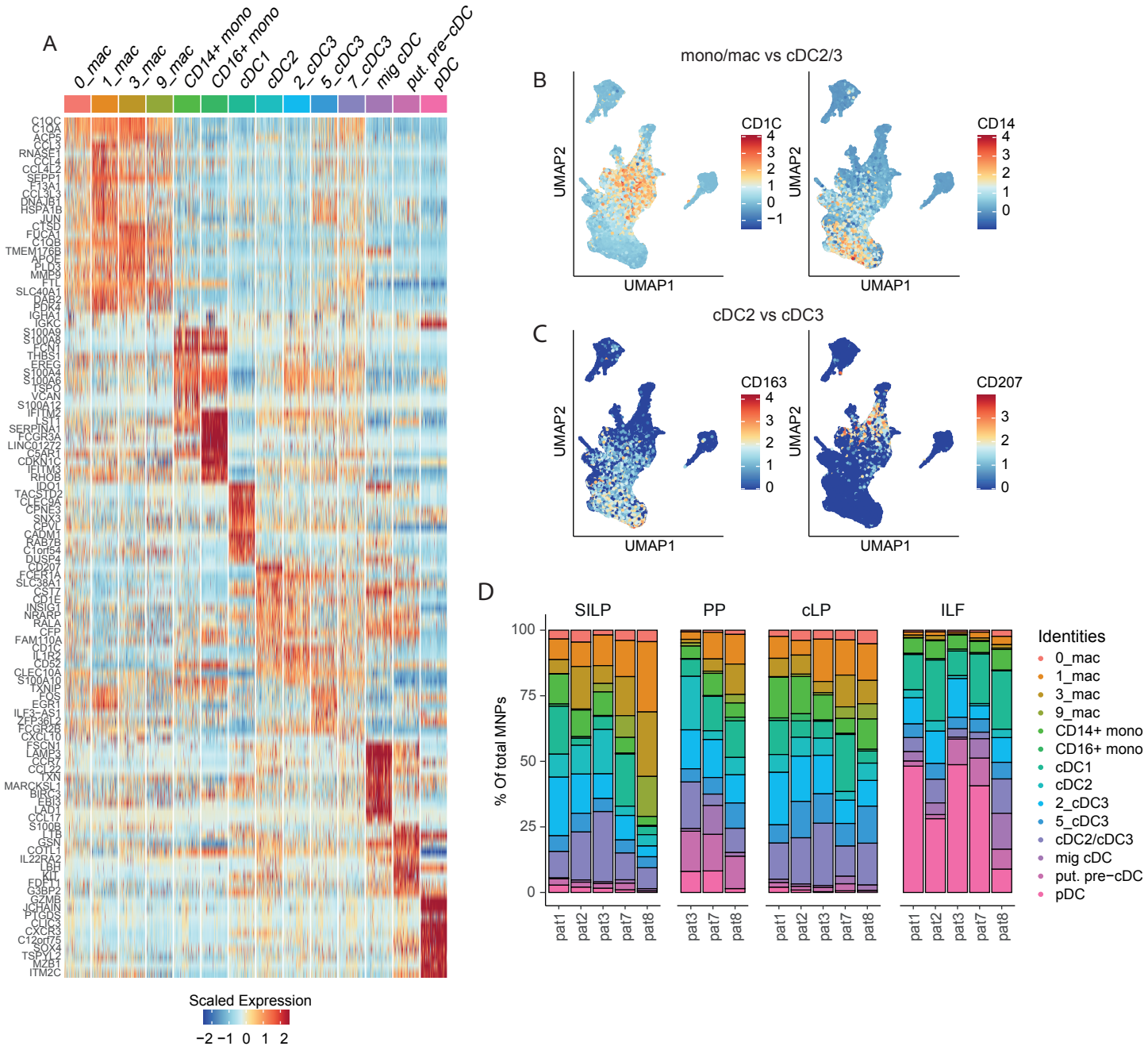

A
cDC1
SILP
PP
PHACTR1
CYBA
LGALS1
FOSB
IER3
DBN1
MT−ND4
SDS
RNASE6
CDH17
SPAG9
C1QA
TYROBP
SEPP1
SRRM2
BNIP3L
CXCL2
C1QC
CCL3
C1QB
SOX4
CD37
GABARAP
S100B
SERPINB1
LGALS2
IL2RG
LTB
C1orf162
C
cDC2
SILP
PP
CST3
LGALS1
TYROBP
FCER1G
RGCC
LGALS3
FCER1A
CPVL
CD1C
FCGR2B
GPX1
S100A10
VSIG4
PSAP
FTL
C1QA
CXCL8
CXCL2
C1QC
IRF1
C1QB
ICAM1
CXCL10
GADD45B
LTB
IL22RA2
NRARP
IFNGR1
B
cDC1
cLP
ILF
CREM
ZC3HAV1
ZNF331
CST3
TACSTD2
ATP1B3
CD40
CYBA
RAB11FIP1
CHMP1B
TIPARP
RILPL2
C15orf48
LGALS3
RNASET2
CD63
SOD2
S100A11
PHLDA2
CCDC107
MAP3K8
REL
HERPUD1
PNOC
DUSP5
BACH1
HLA−DMA
MT−CYB
FTL
MYL6
NFKB1
RNASE6
VMO1
CIB1
TYROBP
MT−CO2
OXSR1
LAPTM4A
TIMP1
ANKRD37
GPR183
MT−CO3
DNAJB6
RGCC
RAC2
MS4A6A
ADGRE5
ANXA1
CXCL8
IL1B
RPS29
ZFP36
FOS
RPS21
IRF8
LGALS2
ZNF366
REPIN1
TSPAN33
JUN
LYZ
IRF5
PNISR
LIMD2
IFI16
FSCN1
NME2
HNRNPA1
SNX20
CBFA2T3
LCP1
WAPL
ACTB
HSH2D
ANPEP
PRKCB
VAV3
CD37
FKBP15
LGALS9
SCAF11
ZFP36L1
AC090498.1
RUFY3
LDHB
GDI2
TKT
MTRNR2L12
UPF2
RP11−1143G9.4
HMGN2
PA2G4
GEM
D
cDC2
cLP
ILF
CPVL
FCER1A
CST3
LGALS3
HLA−A
FCGR2B
HLA−DPB1
IRF1
DNASE1L3
PHLDA2
LGALS1
TYROBP
RGS2
AKAP13
ITM2B
ACP5
HLA−DQB1
SRSF5
SEMA6B
HLA−E
MPEG1
CLEC10A
SOD2
FTL
TGFBI
HLA−DRA
MT−ND6
TYMP
ZEB2
VSIG4
UPP1
S100A11
CXCL8
ZNF267
TIMP1
STAT1
SERTAD1
ATP2B1
TNFSF14
C10orf128
RNASET2
NPC2
RAC1
GRN
H3F3A
RAB20
ALOX5AP
HSPA5
OXCT2
LAPTM5
FAM65B
PAGR1
RPS27
RPL41
RPS28
RPS21
RPL39
RPL37
HIST1H4C
SERPINF2
RPS29
MIF
ZFP36L1
ZFP36L2
UNC119
FOS
DDIT4
ATP5E
CDKN1B
RNF41
U2SURP
PLCG2
GOLGA3
LRRC25
CSRP1
SNRPG
AC090498.1
NDE1
DHRSX
HSDL2
GCC2
MAPKAPK3
CYTIP
SGF29
LSM10
RPS24
RSRP1
ETV6
KLF6
DDAH2
TIMM17B
SYNGR2
EIF2A
CDK9
DUSP1
RASGRP2
SLC2A3
AREG
CCNH
RSL1D1
Scaled Exp
−1 0 1
pat3
pat7
pat8
E
cDC1
SILP PP cLP ILF
regulation of interleukin−1 beta production (GO:0032651)
positive regulation of cell population proliferation (GO:0008284)
positive regulation of interleukin−12 production (GO:0032735)
positive regulation of CD4−positive, alpha−beta T cell proliferation (GO:2000563)
positive regulation of CD4−positive, CD25−positive, alpha−beta regulatory T cell differentiation (GO:0032831)
cellular response to lipopolysaccharide (GO:0071222)
chemokine−mediated signaling pathway (GO:0070098)
granulocyte chemotaxis (GO:0071621)
positive regulation of toll−like receptor 2 signaling pathway (GO:0034137)
regulation of defense response to bacterium (GO:1900424)
F
cDC2
SILP PP cLP ILF
regulation of lymphocyte differentiation (GO:0045619)
cellular response to transforming growth factor beta stimulus (GO:0071560)
T cell differentiation (GO:0030217)
regulation of B cell activation (GO:0050864)
regulation of dendritic cell antigen processing and presentation (GO:0002604)
regulation of T cell activation (GO:0050863)
antigen processing and presentation of peptide antigen via MHC class II (GO:0002495)
negative regulation of T cell proliferation (GO:0042130)
interleukin−27−mediated signaling pathway (GO:0070106)
positive regulation of T cell activation (GO:0050870)
antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP−dependent (GO:0002479)
G
cDC1
PP
ILF
6
2
41
LGALS2
CD37
SILP
cLP
18
3
47
CYBA
RNASE6
TYROPB
H
cDC2
PP
ILF
4
0
65
SILP
cLP
8
16
152
CST3
LGALS1
TYROBP
RGCC
LGALS3
FCER1A
CPVL
CD1C
FCGR2B
VSIG4
FTL
C1QA
CXCL8
C1QC
IRF1
C1QB

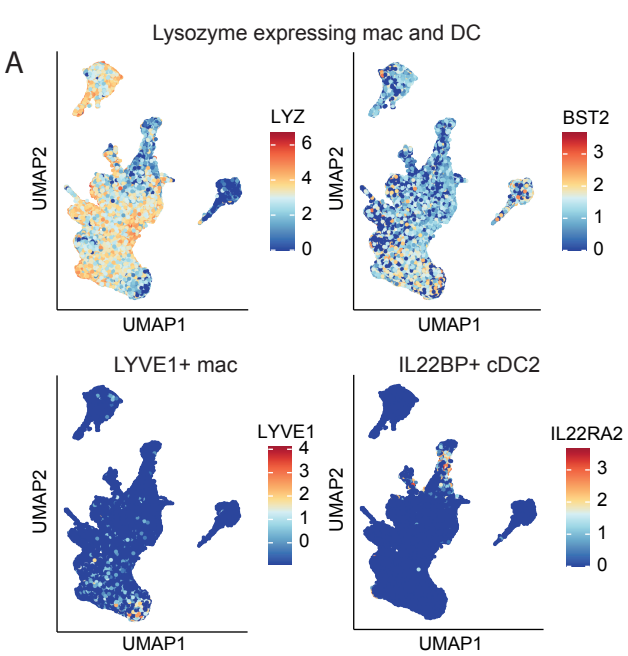
Figure 3

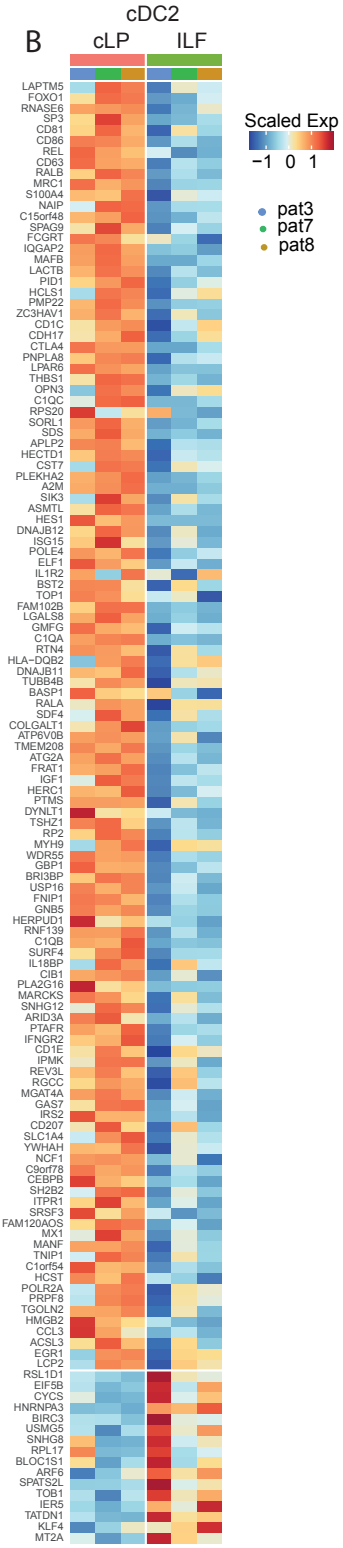

A
mono
mac
cDC3
cDC2
mig cDC
pDC
cDC1
UMAP2
UMAP1
Transferred IDs
1_mac
3_mac
CD14+ mono
CD16+ mono
inflam. mono
cDC1
cDC2
2_cDC3
5_cDC3
7_cDC3
mig cDC
put. pre−cDC
debris
pDC
B
n=5
cLP
18757 cells
n=1
cLP uninvolved
11944 cells
n=1
cLP inflamed
9989 cells
ILF
5721 cells
ILF uninvolved
124 cells
ILF inflamed
702 cells
UMAP2
UMAP1
C
Total MNPs
% Of total MNPs
100
75
50
25
0
healthy
uninvolved
inflamed
cLP
1_mac
3_mac
CD14+ mono
CD16+ mono
inflam. mono
cDC1
cDC2
2_cDC3
5_cDC3
7_cDC3
mig cDC
put. pre−cDC
debris
pDC
D
mono/mac
% Of total monocytes and macrophages
100
75
50
25
0
healthy
uninvolved
inflamed
cLP
1_mac
3_mac
CD14+ mono
CD16+ mono
inflam. mono
E
cDC
% Of total cDCs
100
75
50
25
0
healthy
uninvolved
inflamed
cLP
cDC1
cDC2
2_cDC3
5_cDC3
7_cDC3
mig cDC
put. pre−cDC
debris
F
blood CD14+ mono signature
0.50
0.25
0.00
−0.25
1_mac
3_mac
CD14+ mono
CD16+ mono
inflam. mono
2_cDC3
5_cDC3
7_cDC3
G
FCER1A
4
3
2
1
0
−1
1_mac
3_mac
CD14+ mono
CD16+ mono
inflam. mono
2_cDC3
5_cDC3
7_cDC3
H
1_mac
3_mac
CD14+ mono
CD16+ mono
inflam. mono
EGR1
RNASE1
JUN
CCL3
HES1
SEPP1
PDK4
HSPA1B
FOLR2
CXCL8
KLF2
CCL4
CCL4L2
DAB2
P2S
SLC40A1
RGS1
F13A1
LGMN
STAB1
DNASE1L3
FUCA1
XCR1
C1QC
APOE
GPNMB
C1QB
C1SD
C1QA
APOC1
YWHAH
LGALS3
AXL
SERPINF1
HLA−DQA1
ADAMDEC1
AZM
PLD3
FCN1
S100A9
S100A8
VCAN
EREG
CSTA
TKT
S100A4
THBS1
SLC11A1
SYAP1
C1orf162
S100A6
SH3BGRL3
AP1S2
BASP1
CLEC4E
FGR
SAMSN1
IGLC3
SPN
FCGR3A
FAM65B
CDKN1C
LST1
RHOC
LINC01272
IFITM2
TCF7L2
LYST
PAG1
SERPINA1
IFITM3
ZNF703
PLAC8
PECAM1
APOBEC3A
CFD
CFP
IDO1
GBP1
CD69
CXCL10
TXN
CXCL9
CST7
FAM26F
PPA1
SLAMF7
INHBA
GBP5
FCGR2B
CD40
TAP1
CXCL1
MIR155HG
IL4I1
IL7R
IL1R2
Scaled Exp
−1 0 1
CD state
healthy control
uninvolved
inflamed
Patient
pat1
pat2
pat3
pat7
pat8
pat9

Supplementary Figure 1

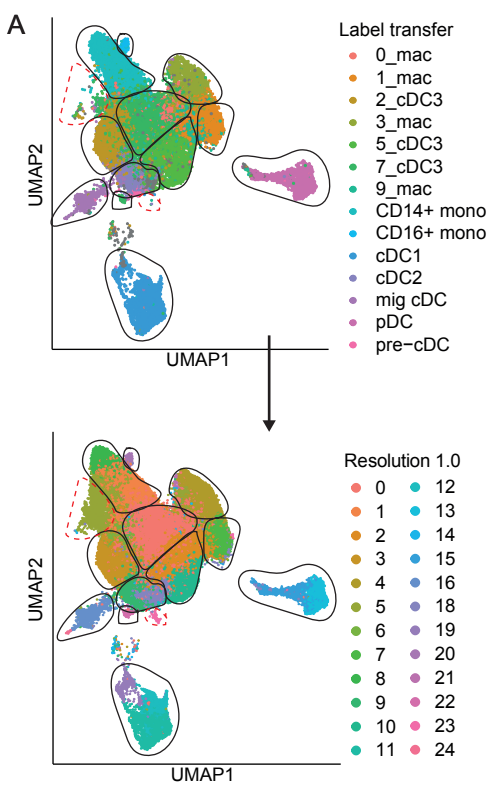


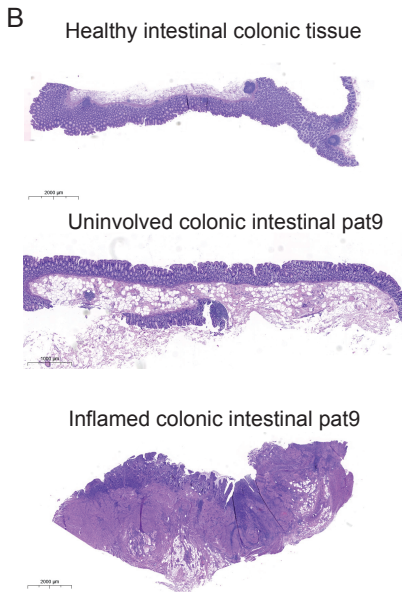
Supplementary Figure 2

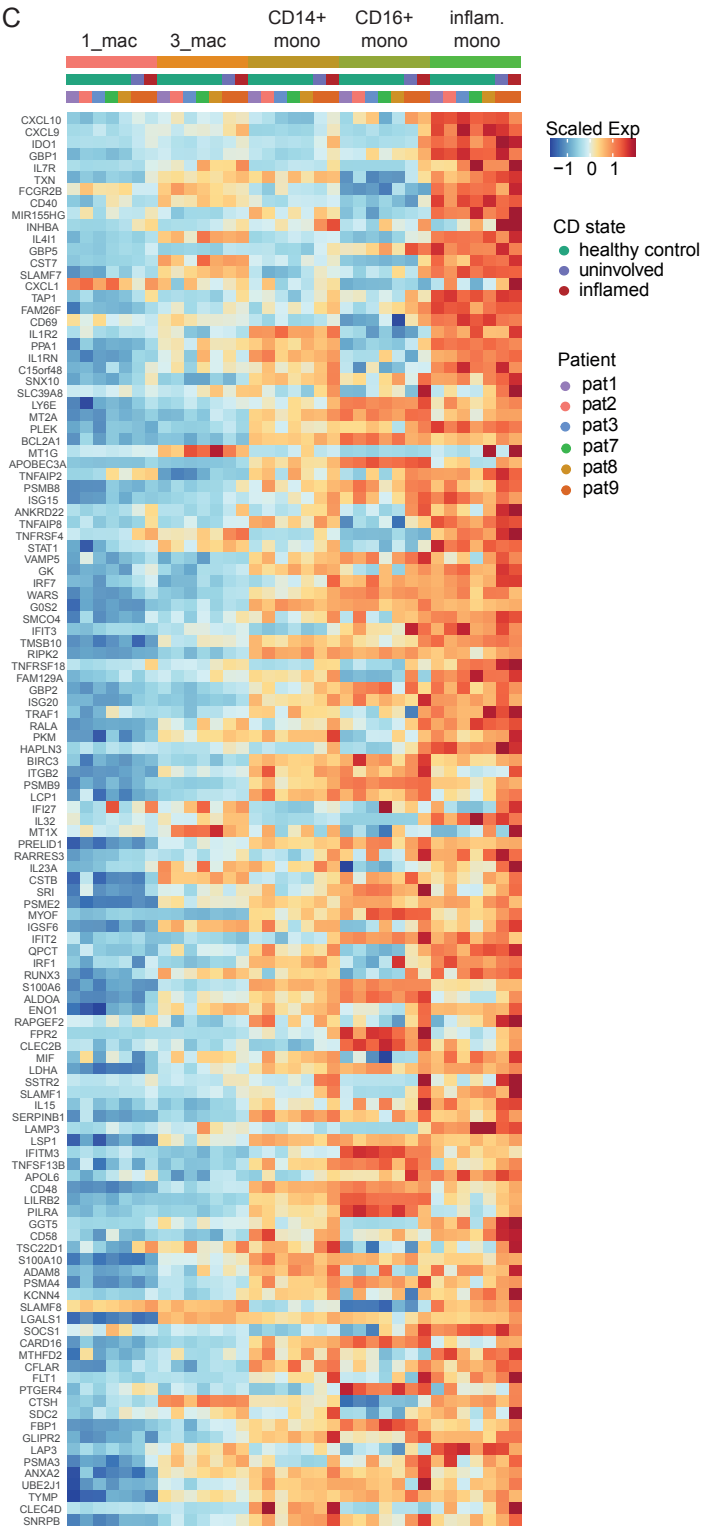




Supplementary Figure 3







Manuscript III

The small intestinal and colon mucosa contains similar mesenchymal stromal cell subsets that derive from Gli1+ embryonic precursors

The small and large intestine contain transcriptionally related mesenchymal stromal cell subsets that derive from embryonic *Gli1*$^{+}$ mesothelial cells

Simone Isling Pærregaard[1], Sophie Schussek[1, #], Line Wulff[1,#], Kristoffer Niss[2] , Urs Mörbe[1], Johan Jendholm[1], Kerstin Wendland[3], Anna T. Andrusaite[4], Kevin F. Brulois[5], Robert J. B. Nibbs[4], Katarzyna Sitnik[1], Allan McI Mowat[4], Eugene C. Butcher[5,6], Søren Brunak[2], William W. Agace[1,3,] *.

[1]Department of Health Technology, Technical University of Denmark, Kemitorvet, 2800 Kgs. Lyngby, Denmark.

[2]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen 2200, Denmark.

[3]Immunology Section, Lund University, Lund 221 84, Sweden.

[4]Institute of Infection, immunity and Inflammation, University of Glasgow, Glasgow, Scotland.

[5]Laboratory of Immunology and Vascular Biology, Department of Pathology, School of Medicine, Stanford University, Stanford, USA.

[6]The Center for Molecular Biology and Medicine, Veterans Affairs Palo Alto Health Care System and the Palo Alto Veterans Institute for Research (PAVIR), Palo Alto, USA

[#] Equal contribution

*Correspondence: William Agace, wiag@dtu.dk

Summary

Intestinal fibroblasts (FB) play essential roles in intestinal homeostasis. Here we show that the small and large intestinal lamina propria (LP) contain similar FB subsets that locate in specific anatomical niches and express distinct arrays of epithelial support genes. However, there were tissue specific differences in the transcriptional profile of intestinal FB subsets in the two sites. All adult intestinal LP mesenchymal stromal cells (MSC), including FB, smooth muscle cells (SMC) and pericytes derive from *Gli1*-expressing embryonic precursors which we identify as mesothelial cells. Trajectory analysis suggested that adult SMC and FB derive from distinct embryonic intermediates, and that adult FB subsets develop in a linear trajectory from $CD81^{+}$ FB. Finally, we show that colonic subepithelial $PDGFR\alpha^{hi}$ FB comprise several functionally and anatomically distinct populations that originate from an *Fgfr2*-expressing FB intermediate. Collectively our results provide novel insights into MSC diversity, location, function and ontogeny, with implications for our understanding of intestinal development, homeostasis and disease.

Key words

Mesenchymal stromal cells, regional specialization, fibroblast, development, small intestine, colon, trajectory, embryonic.

Introduction

The small and large intestines form a continuous tube from the stomach to the anus, but are functionally and anatomically distinct. The small intestine is the primary site of food digestion and nutrient absorption and is characterized by finger-like projections termed villi that protrude into the intestinal lumen and maximize the absorptive area of the epithelium. In contrast, the large intestine is primarily a site of water absorption and is a major niche for beneficial microbes; its surface consists of crypts linked by short regions of flat surface epithelium. The cellular composition of the intestinal mucosa also differs markedly between the small and large intestines (Mowat and Agace, 2014; Agace and McCoy, 2017). For example, the small and large intestines contain different numbers and proportions of innate and adaptive immune cells as well as epithelial subpopulations (Barker, 2014; Mowat and Agace, 2014; Agace and McCoy, 2017; Parikh *et al.*, 2019). These distinct segments are also exposed to different concentrations of microbial and food-derived metabolites that regulate the composition and function of local cells (Mowat and Agace, 2014; Agace and McCoy, 2017). However, the cellular and signaling components that determine the differences in tissue structure and composition are not fully understood.

The intestinal lamina propria (LP) contains a large population of tissue resident mesenchymal stromal cells (MSC) that include fibroblasts (FB), pericytes (PC) and smooth muscle cells (SMCs) that play an essential role in intestinal homeostasis (Degirmenci *et al.*, 2018; Kinchen *et al.*, 2018; Shoshkes-Carmel *et al.*, 2018; Brügger *et al.*, 2020; Hong *et al.*, 2020; McCarthy *et al.*, 2020; Wu *et al.*, 2021). For example, intestinal FB are major producers of extracellular matrix proteins that help provide structure to the mucosa (Furuya and Furuya, 2007; Roulis and Flavell, 2016). They also express factors essential for epithelial (Stzepourginski *et al.*, 2017; Degirmenci *et al.*, 2018; Shoshkes-Carmel *et al.*, 2018; Karpus *et*

al., 2019; McCarthy *et al.*, 2020; Wu *et al.*, 2021) and endothelial homeostasis (Thomson *et al.*, 2018; Hong *et al.*, 2020; Fawkner-Corbett *et al.*, 2021), as well as immune cell localization and function (Fagarasan *et al.*, 2001; Powell *et al.*, 2011; Beswick *et al.*, 2014; Vicente-Suarez *et al.*, 2015). Recent single cell (sc)RNA-seq studies have demonstrated considerable heterogeneity within the intestinal LP MSC compartment and have led to the identification of several FB clusters with non-redundant functions in intestinal homeostasis (Kinchen *et al.*, 2018; Smillie *et al.*, 2019; Brügger *et al.*, 2020; Hong *et al.*, 2020; McCarthy *et al.*, 2020; Roulis *et al.*, 2020; Wu *et al.*, 2021). A picture is also emerging whereby different intestinal FB subsets locate within distinct regions of the mucosa (Degirmenci *et al.*, 2018; Shoshkes-Carmel *et al.*, 2018; Thomson *et al.*, 2018; Parikh *et al.*, 2019; Hong *et al.*, 2020; McCarthy *et al.*, 2020; Fawkner-Corbett *et al.*, 2021), providing specialized support to cells in their local environment (Aoki *et al.*, 2016; Stzepourginski *et al.*, 2017; Degirmenci *et al.*, 2018; Kinchen *et al.*, 2018; Shoshkes-Carmel *et al.*, 2018; Karpus *et al.*, 2019; Hong *et al.*, 2020; McCarthy *et al.*, 2020; Wu *et al.*, 2021). However, the exact nature of these diverse LP MSC subsets and how they differ in the small and large intestine remains to be established.

scRNA-seq analyses have shown that the composition of human intestinal MSC populations changes markedly as the tissue develops in the embryo (Fawkner-Corbett *et al.*, 2021; Holloway *et al.*, 2021). Although the origin of these populations remains to be determined, lineage-tracing experiments in mice have suggested that the mesothelium, an epithelial monolayer that lines the serosal surface of the intestine (Koopmans and Rinkevich, 2018), can give rise to SMC and various FB in the intestinal serosa and muscle layers (Wilm *et al.*, 2005; Rinkevich *et al.*, 2012). Whether MSC subsets present within the adult intestinal LP derive from cells of common or distinct embryonic origin and the developmental relationship between adult MSC subsets remains unclear.

Here we demonstrate that LP MSC subset composition is similar in the small and large intestine and that each subset occupies distinct anatomical niches. Nevertheless, the transcriptional profile of the major LP FB subsets differed markedly between the small and large intestine, suggesting regional specific functions in intestinal homeostasis. Grafting and lineage-tracing experiments demonstrated that all MSC subsets in adult small intestinal and colonic LP derive from *Gli1*-expressing precursors present in embryonic day (E)12.5 intestine. Computational analysis suggested that all adult intestinal MSC derive from embryonic intestinal mesothelial cells and that adult SMC and FB arise from distinct mesothelial derived embryonic intermediates. We could also define a linear developmental trajectory for all adult FB subsets that originated from $CD81^{+}$ FB.

Results

The small intestine and colon LP contain diverse, but transcriptionally related MSC subsets.

To gain a broad understanding of MSC subset diversity in the intestinal LP, we performed scRNA-seq on MSC isolated from the small intestine and colon LP of 8-10 week old mice. Briefly, after removal of Peyer's patches, muscularis externa and epithelium, intestinal MSCs were enriched from digested intestinal LP cell suspensions by fluorescently activated cell sorting of live, single, lineage$^-$ (CD45$^+$, Ter119$^+$), non-epithelial (EpCAM$^+$), non-endothelial (CD31$^+$), non-lymphoid tissue-associated MSCs (BP3$^+$) (Taylor *et al.*, 2007) and non-glial cells (L1CAM$^+$), followed by gating on cells expressing the pan MSC marker Itgβ1 (Fig. S1A). After bioinformatic removal of contaminating c-kit$^+$ interstitial cells of Cajal (ICC), CD31$^+$ endothelial cells, plasma cells and CD45$^+$ immune cells, sequencing data of 16.964 small intestinal and 14.164 colonic MSC remained.

Louvain clustering identified six small intestinal MSC clusters (Fig. 1A) and differential gene expression (DEG) analysis of these clusters identified pericytes, SMC and four FB clusters (Fig. S1B). These were PDGFRα^{hi} FB, two PDGFRα^{lo}CD34^{hi} clusters that could be distinguished based on their expression of *Cd81* (hereafter called CD81$^+$ FB) and *Igbp5* (hereafter called Igfbp5$^+$ FB), and a PDGFRα^{lo}CD34^{lo} cluster that expressed higher levels of *Fgfr2* (hereafter called Fgfr2$^+$ FB) (Fig. S1C). To determine how these clusters might relate to those identified in other, recently published scRNA-seq studies of small intestinal MSC (Hong *et al.*, 2020; McCarthy *et al.*, 2020), DEGs from the previous MSC subsets were overlaid with our scRNA-seq dataset (Fig. S1D). The MSC population termed "mural cells" by Hong *et al* (Hong *et al.*, 2020) corresponded to our pericytes, while their FB subsets termed FB2, 3, 4 and 5 corresponded to our small intestinal Igfbp5$^+$ FB, Fgfr2$^+$ FB, CD81$^+$ FB and PDGFRα^{hi} FB clusters, respectively (Fig. S1D). The signature genes of FB1 identified by Hong

et al as activated FB based on their expression of *Junb* and *Fosb*, were expressed widely by several MSC subsets in our dataset (Fig. S1D), indicating that this cluster represents a cell state rather than an MSC subset. Similar analysis of the MSC datasets generated by McCarthy *et al* (McCarthy *et al.*, 2020) demonstrated that the PDGFRα^{hi} MSC subset they defined as "telocytes" corresponded to our PDGFRα^{hi} FB cluster, while their Lo-1 FB subset corresponded to our CD81^{+} FB cluster and their Lo-2 FB subset encompassed both our Fgfr2^{+} and Igfbp5^{+} FB clusters (Fig. S1D) (McCarthy *et al.*, 2020). Thus, our results confirm and extend recent findings and highlight the complexity of MSC subsets in the small intestinal LP.

Louvain clustering also identified six MSC clusters in colon LP (Fig. 1B), which DEG analysis identified as pericytes, SMC, and four FB clusters (Fig. S1E). These were PDGFRα^{hi} FB and three PDGFRα^{lo}CD34^{+} clusters that could be distinguished based on their expression of *Cd81* (hereafter called CD81^{+} FB), *CD90* (hereafter called CD90^{+} FB) or *Fgfr2* (hereafter called Fgfr2^{+} FB) (Fig. S1F). To determine the relationship between the colonic and small intestinal MSC subsets, Pearson correlation analysis was performed based on the pseudo-bulk of overlapping variable genes between the two data sets. This showed that colonic pericytes, SMC, PDGFRα^{hi} FB, CD81^{+} FB closely correlate with their counterparts in the small intestine, that colonic CD90^{+} FB most closely correlate with small intestinal Igfbp5^{+} FB and that colonic Fgfr2^{+} FB closely correlate with both Fgfr2^{+} and Igfbp5^{+} FB (Fig. 1C).

FB subsets are located in distinct niches along the crypt-villus axis

There is increasing evidence that subsets of small intestinal FBs may occupy distinct anatomical niches that overlap with the WNT/BMP signaling gradient along the crypt-villus axis (Shoshkes-Carmel *et al.*, 2018; Bahar Halpern *et al.*, 2020; McCarthy *et al.*, 2020). In line with a recent report (Stzepourginski *et al.*, 2017), we found that small intestinal CD34^{+} FB (which include CD81^{+} and Igfbp5^{+} FB) were located around crypts and in the submucosa, but

were largely excluded from the villus core (Fig. S1G). Of these, $CD34^{+}CD81^{+}$ FB located around $CD31^{+}$ vessels close to and within the submucosa, with some locating close to crypts (Fig. 1D), consistent with recent reports (Thomson *et al.*, 2018; Hong *et al.*, 2020; McCarthy *et al.*, 2020), while $Igfbp5^{+}$ ($CD34^{+}CD81^{-}$) FB located around crypts (Fig. 1D). Conversely, $PDGFR\alpha^{+}CD34^{-}$ FB (including both $PDGFR\alpha^{hi}$ FB and $PDGFR\alpha^{lo}Fgfr2^{+}$ FB) were located directly underlying the epithelium and within the villus core (Fig. S1G). Of these, the $Fgfr2^{+}$ FB were located within the villus core towards the tip of the villus (Fig. 1E); this was confirmed using CXCL14 as a marker for this subset (Fig. S1H and Fig. 1F). In contrast, the $PDGFR\alpha^{hi}$ FB lay directly under the epithelium and at the villus tip (Fig. 1E, Fig. 1F, Fig. S1G), supporting previous findings (Kurahashi *et al.*, 2013; Bahar Halpern *et al.*, 2020; Brügger *et al.*, 2020; Hong *et al.*, 2020; McCarthy *et al.*, 2020).

As in the small intestine, colonic $CD34^{+}$ FB subsets located beneath and surrounding intestinal crypts, while $PDGFR\alpha^{hi}$ FB formed a thin layer directly underlying the epithelium and were concentrated at the top of crypts (Fig. S1I). Colonic $CD90^{+}CD34^{+}$ FB, which expressed high levels of *Pparg* (Fig. S1H) could be identified after staining for PPARγ and were located at the base of colonic crypts (Fig. 1G). Colonic $Fgfr2^{+}$ FB localized preferentially between crypts (Fig. 1H), whereas the colonic $CD81^{+}$ FB located below the crypts and in the submucosa (Fig. 1I). Collectively these results demonstrate that the FB subsets identified by scRNA-seq locate within distinct niches of the small and large intestine.

Expression of epithelial support genes is conserved across FB subsets in the small intestine and colon

Recent studies have suggested a division of labor amongst small intestinal FB subsets in the production of epithelial support factors (Degirmenci *et al.*, 2018; Kinchen *et al.*, 2018; Brügger *et al.*, 2020; McCarthy *et al.*, 2020; Wu *et al.*, 2021) and we thus assessed the expression of

such genes in our small intestinal and colonic FB datasets. Consistent with previous studies (Shoshkes-Carmel *et al.*, 2018; McCarthy *et al.*, 2020), small intestinal PDGFRα^{hi} FB were major producers of BMPs and this property was shared by colonic PDGFRα^{hi} FB (Fig. 2A). *Bmp3*, *Bmp5* and *Bmp7* expression was largely restricted to PDGFRα^{hi} FB, while expression of *Bmp1*, *Bmp2* and *Bmp4* was found more broadly among FB MSC subsets in both tissues (Fig. 2A). Both small intestinal and colonic PDGFRα^{hi} FB were also the dominant source of the non-canonical WNT ligands, *Wnt4*, *Wnt5a* and *Wnt5b*, although Fgfr2^{+} FB also expressed *Wnt4*, particularly in the small intestine (Fig. 2A). Consistent with previous results (Brügger *et al.*, 2020; McCarthy *et al.*, 2020), CD81^{+} FB were the major source of the BMP antagonist *Grem1* in the small intestine and this was also highly expressed by colonic CD81^{+} FB. However, in the colon, Fgfr2^{+} and CD90^{+} FB also expressed *Grem1* (Fig. 2A). Thus, the specialization of MSC subsets in their expression of epithelial support genes is largely conserved between the small intestine and colon.

Small intestinal and colonic PDGFRα^{hi} FB and Fgfr2^{+} FB display regional transcriptional specificity

To gain a broader understanding of how the major PDGFRα^{hi} and Fgfr2^{+} FB subsets in the small and large intestine might be related, we examined our scRNA-seq datasets for surface markers that would allow us to identify and sort these cells for bulk RNA-seq analysis (Fig. S2A). For the small intestine, pericytes were identified and sorted as PDGFRα^{-}ESAM-1^{+}PDPN^{-} cells, SMC as PDGFRα^{-}ESAM-1^{+}PDPN^{+} cells, PDGFRα^{hi} FB as ESAM-1^{-} PDGFRα^{hi}CD34^{-} cells, Fgfr2^{+} FB as ESAM-1^{-}PDGFRα^{int}CD34^{-} cells and CD34^{+} FB (including CD81^{+} and Igfbp5^{+} FB) as ESAM-1^{-}PDGFRα^{int}CD34^{+} cells. For the colon, pericytes were sorted as for small intestine, PDGFRα^{hi} FB was sorted as ESAM-1^{-} PDPN^{+}CD34^{-} cells, Fgfr2^{+} FB were sorted as ESAM-1^{-}PDPN^{hi}CD34^{+}CD90^{-} cells, CD90^{+} FB

as ESAM-1^{-}CD34^{+}PDPN^{hi}CD90^{+} cells and CD81^{+} FB as ESAM-1^{-}PDPN^{int}CD34^{+}CD90^{-} cells, based on the fact that colonic CD81^{+} FB express low levels of PDPN compared with the other CD34^{+} colonic FB subsets (Fig. S2B). Correlation analysis of these bulk sorted intestinal FB subsets with the scRNA-seq data confirmed the accuracy of this staining strategy to identify small intestinal and colonic FB subsets by flow cytometry (Fig. 2B and Fig S2C). This initial panel was then refined for use in subsequent flow cytometry based analysis by including anti-CD81 to positively identify CD81^{+} FB directly, together with anti-CD146 (Fig. 2C and D), which can be used interchangeably with ESAM-1 (Fig. S2D). CD81^{+} FB also expressed the atypical chemokine receptor, ACKR4, as assessed using *Ackr4*.GFP reporter mice (Fig. 2C and D) (Thomson *et al.*, 2018), consistent with previous reports (Thomson *et al.*, 2018; Brügger *et al.*, 2020; Hong *et al.*, 2020; McCarthy *et al.*, 2020) and our scRNA-seq analysis (Fig. S1B).

PCA analysis of bulk sorted PDGFRα^{hi} FB and Fgfr2^{+} FB distinguished these subsets from one another in PC1, while PC2 separated small intestinal from colonic FB (Fig. 2E), suggesting that anatomical location has a major impact on the transcriptional profile of these FB subsets. Consistent with this, small intestinal and colonic PDGFRα^{hi} FB differed in their transcription of 698 genes, while the two populations of Fgfr2^{+} FB differed in their transcription of 716 genes (Fig. 2F, see Supplementary Table 1 (for PDGFRα^{hi} FB) and Supplementary Table 2 (for Fgfr2^{+} FB) for complete list). Of these, 149 genes were differentially expressed between the small intestine and colon in both FB subsets (Supplementary Table 3); this included numerous Hox genes (Fig. S2E), consistent with the role of mesoderm in specifying the development of the different intestinal segments (Yuasa, 2003). Enrichr based analysis (Bioplanet 2019 (Huang *et al.*, 2019)) showed that most of the upregulated pathways in the two subsets were in colon compared with small intestine, with few being upregulated in small intestine compared with colon (Fig. S2F). Irrespective of their location, PDGFRα^{hi} FB and Fgfr2^{+} FB showed very distinct expression of epithelial support

genes (Fig. 2G), suggesting these populations play discrete roles in maintaining the epithelium; many of these genes were expressed at significantly higher levels in colonic subsets compared with their small intestinal counterparts (Fig. 2G). PDGFRα^{hi} FB and Fgfr2^{+} FB also expressed a wide range of immunologically relevant genes in both a subset- and tissue-specific manner (Fig. 2H-J). This included several cytokine and cytokine receptors (Fig. 2H), while small intestinal but not colon Fgfr2^{+} FB expressed a wide range of chemokines (Fig. 2I). Both subsets of small intestinal FB also expressed enzymes implicated in vitamin A metabolism, including the generation of retinoic acid (Fig. 2J), a major regulator of small intestinal immune responses. Collectively, these results highlight the unique functions of intestinal PDGFRα^{hi} FB and Fgfr2^{+} FB and show that these vary depending on anatomical location.

Intestinal precursors in E12.5 intestine can give rise to all adult intestinal MSC subsets

While adult small intestinal and colonic LP contains multiple phenotypically, transcriptionally, and spatially distinct MSC subsets, the developmental relationship between these subsets and whether all derive from similar precursors remains unclear (Wilm *et al.*, 2005; Rinkevich *et al.*, 2012; Fawkner-Corbett *et al.*, 2021). To explore this, we first investigated which MSC might be present in the small intestine and colon of E12.5 embryos by flow cytometry (Fig. 3A and S3A). In contrast to adult mice (Fig. 2C and D), E12.5 small intestinal and colonic Itgβ1^{+} cells consisted of one major population of PDGFRα^{+}CD34^{-} MSCs, together with a small subset of PDGFRα^{-} cells that expressed the mesothelial markers dipeptidyl peptidase-4 (DPP4, CD26) and PDPN (Fig. 3A and B) (Meyerholz, Lambertz and McCray, 2016). To assess whether these populations could give rise to the MSC subsets found in the adult intestine, small and large intestine were dissected from E12.5 embryos and transplanted under the kidney capsule of adult WT recipient mice (Fig. 3C). Embryonic intestines from mice ubiquitously expressing EYFP were used for these experiments in order to trace the development of donor derived

(EYFP$^+$) MSC within grafted tissues. As expected (Ferguson, Parrott and Connor, 1972; Mosley and Klein, 1992; Yanai *et al.*, 2017), small intestinal and colonic grafts had increased markedly in size by 4-6 weeks post transplantation (Fig. 3C) and contained mucosa that histologically resembled that of adult small intestine and colon, respectively (Fig. S3B). To assess the phenotypic diversity of graft-derived MSC, small intestinal and colonic grafts were isolated 4 weeks after transplantation, digested, and the expression of MSC subset markers on embryonically derived (YFP$^+$) Itgβ1$^+$ MSC assessed by flow cytometry (Fig. 3D, Fig S3C). Both small intestinal and colonic grafts contained putative populations of graft-derived SMC (CD146$^+$PDGFRα$^-$PDPN$^+$), pericytes (CD146$^+$PDGFRα$^-$PDPN$^-$), PDGFRα^{hi} FB (CD146$^-$CD34$^{-/lo}$PDGFRα^{hi}), CD81$^+$ FB (CD146$^-$PDGFRα^{lo}CD81$^+$CD34$^+$) and CD81$^-$CD34$^+$ FB (CD146$^-$PDGFRα^{lo}CD34$^+$CD81$^-$) (Fig. 3D). To confirm the presence of these MSC subsets in the grafts, YFP$^+$Itgβ1$^+$ MSC were sorted from grafted colon and subjected to scRNA-seq (Fig. S3C). UMAP dimensionality reduction and Louvain clustering identified eight clusters (Fig. 3E), two of which (clusters 6 and 7) were identified as ICC and mesothelial cells, respectively (Kanamori-Katayama *et al.*, 2011; Lee *et al.*, 2017; Namvar *et al.*, 2018) (Fig. S3D). These clusters were not part of our adult MSC datasets, as ICC were removed bioinformatically and the mesothelium was removed together with the muscularis externa during tissue processing. Pearson correlation analysis based on the pseudobulk of overlapping variable genes identified cluster 3 as being similar to adult PDGFRα^{hi} FB, cluster 4 as pericytes and cluster 5 as SMC (Fig. 3E and F). The remaining three clusters (clusters 0-2) were more closely related to the three adult CD34$^+$ FB subsets, with cluster 1 most closely correlated to CD81$^+$ FB, cluster 0 most closely correlated to Fgfr2$^+$/CD90$^+$ FB and cluster 2 showing equivalent correlation to all three adult CD34$^+$ FB subsets (Fig. 3F). Furthermore, the distinct expression of epithelial support genes by each of the four FB subsets largely overlapped with the pattern seen in adult

intestine (Fig. 3G, Fig. 2A). Collectively, these results suggest that embryonal MSC precursors present in E12.5 intestine can give rise to all adult intestinal MSC subsets.

Adult intestinal MSC derive from *Gli1*$^+$ embryonic precursors

To explore further the origin of adult MSC, we next lineage-traced E12.5 MSC and mesothelium into adulthood. GLI1 is a transcription factor induced by active hedgehog-signaling and is expressed by MSC in multiple organs (Kramann *et al.*, 2015). PDGFRα$^+$CD34$^-$ MSC and PDGFRα$^-$PDPN$^+$ mesothelial cells from the small intestine and colon of E12.5 *Gli1*-EGFP embryos both expressed EGFP, whereas intestinal epithelial, endothelial and CD45$^+$ cells did not (Fig. 4A and B). To lineage trace *Gli1*-expressing cells into adulthood, female *R26R*.EYFP mice (Srinivas *et al.*, 2001) were mated with *Gli1*-Cre.ERT2 males expressing the estrogen receptor (ERT2) under control of *Gli1*-Cre, and pregnant dams injected i.p with 4-hydroxytamoxifen (4-OHT) at E11.5 (Fig. 4C). Two days later, YFP expression had been induced in a small but consistent proportion of Itgβ1$^+$PDGFRα$^+$ MSC and mesothelial cells in the small intestine and colon of *Gli1*-CreERT2$^{+/-}$.*R26R*.EYFP embryos, but not in Cre$^-$ embryos (Fig. S4A and B). Labeling was not observed in intestinal epithelial, endothelial, or CD45$^+$ immune cells of *Gli1*-CreERT2$^{+/-}$.*R26R*.EYFP mice and thus was specific to intestinal MSC and mesothelial cells (Fig. S4A and B). 5-7 weeks after birth, similar proportions of YFP-expressing cells were detected in all mature MSC subsets in both the small intestine and colon (Fig. 4D). Collectively, these results demonstrate that *Gli1*$^+$ cells present in the E12.5 intestine contain cells that give rise to all major adult intestinal MSC subsets.

Trajectory analysis indicates that adult intestinal MSC subsets originate from embryonic *Gli1*$^+$ mesothelial cells

To gain further insights into the relationship between embryonic intestinal *Gli1*$^+$ cells and adult intestinal MSC subsets, scRNA-seq was performed on fluorescently activated cell sorted Itgβ1$^+$ MSC from the colon of E12.5 embryos. Louvain clustering identified six clusters (Fig. 5A), one of which, cluster 4, was identified as mesothelial cells due to its expression of mesothelial associated markers (Kanamori-Katayama *et al.*, 2011; Namvar *et al.*, 2018) (Fig. 5B). Consistent with our flow cytometric analysis (Fig. 3B), this cluster expressed transcripts for DPP4 and PDPN, but lacked expression of PDGFRα (Fig. S5A). To determine the relationship between embryonic and adult MSC subsets, the embryonic and adult colonic datasets were integrated and tSPACE (Dermadi *et al.*, 2020) trajectory analysis was performed on MAGIC imputed sets of variable genes, as described previously (Brulois, 2020; Xiang *et al.*, 2020). Pericytes were removed from this analysis, as too few of these cells were present in the adult dataset to generate meaningful conclusions. Three-dimensional visualization of tSPACE principal components (tPC) 1-3 demonstrated that embryonic cells broadly clustered together and away from adult MSC subsets (Fig. 5C). Nevertheless, two clear connections were observed between embryonic and adult colonic MSC (Fig. 5C, arrow heads). The first was a direct and distinct connection between embryonic clusters 0 and 2 and adult SMC, while the second was a connection between embryonic clusters 4 (mesothelial cells) and 5 to adult CD81$^+$ FB and to a lesser extent adult CD90$^+$ FB (Fig. 5C). Supporting the idea that the mesothelium gives rise to SMC and some FB in the intestinal serosa and muscle layers of the intestine (Wilm *et al.*, 2005; Rinkevich *et al.*, 2012), we found that both mesothelial cells and cluster 5 expressed several genes previously associated with FB progenitors (Bae *et al.*, 2011; Dulauroy *et al.*, 2012; Castagnaro *et al.*, 2013; Driskell *et al.*, 2013; Worthley *et al.*, 2015; Vallecillo-García *et al.*, 2017) (Fig. S5B). We thus selected mesothelial cells as a tSPACE trajectory starting point for pseudotime analysis (Fig. 5D). This demonstrated a pseudotime trajectory of

mesothelial cells to adult SMC via embryonic clusters 0 and 2 and from mesothelial cells and cluster 5 to adult CD81$^+$ FB (Fig. 5D, arrows).

Interestingly, rather than branching immediately into distinct FB subsets, adult CD81$^+$ FB connected directly to CD90$^+$ FB that then connected to Fgfr2$^+$ FB and finally to PDGFRα^{hi} FB (Fig 5C), and tSPACE analysis of adult FB showed a similar linear connection between FB subsets (Fig. 5E). As adventitial CD81$^+$ FB have been suggested to contain FB precursors in adults (Buechler *et al.*, 2021), we used them as the starting population for a new pseudotime analysis, which again indicated a linear trajectory from adult CD81$^+$ FB via CD90$^+$ FB and Fgfr2$^+$ FB to PDGFRα^{hi} FB (Fig. 5F). This conclusion was further supported when we overlaid our trajectory on to the DEG genes for clusters generated by a recent pseudotime analysis of mouse tissue FB, which has suggested a developmental trajectory from *Pi16*$^+$ precursors through a population of *Col15α1*$^+$ FB that eventually gives rise to mature tissue specific FB that include *Fbln1*$^+$and then *Bmp4*$^+$ FB in the intestine (Buechler *et al.*, 2021). This analysis showed that the *Pi16*$^+$, *Col15α1*$^+$, *Fbln1*$^+$ and *Bmp*4$^+$ FB clusters defined by Buechler *et al* broadly overlapped with our colonic CD81$^+$, CD90$^+$, Fgfr2$^+$ and PDGFRα^{hi} FB subsets, respectively (Fig. S5C). Collectively, these results suggest that adult MSC subsets originate from the embryonic *Gli1*$^+$ mesothelium, with adult SMC deriving from an embryonic intermediate distinct from that which gives rise to adult FB. In addition, these results suggest that adult FB subsets arise sequentially from CD81$^+$ FB.

Each FB cluster had an extended appearance in tSPACE, with groups of cells streaming outwards from a central core (Fig 5E). To determine what processes might underlie this appearance, we assessed differences in gene expression between the start (core) and end (tip) of each FB cluster (Fig. S5D). While cells at the tip and core of the clusters had similar read counts and detected genes (Fig. S5E), those at the core of each subset expressed 108-121 genes at significantly higher levels than those at the tip, while tip cells expressed no or few (0-2)

genes at a significantly higher level. Enrichr based analysis (GO biological process 2018 (Ashburner *et al.*, 2000)) of the genes expressed preferentially by cells at the core demonstrated that 9 of the top 10 pathways were shared across FB subsets (Fig. S5F). These included processes involved in positive regulation of transcription, responses to cytokines, and responses to unfolded proteins, with the overwhelming majority of these genes being shared by the core cells in all the FB subsets (Fig. S5G). Together these results suggest that the tip cells within each FB subset are more quiescent than their core counterparts and hence may be more highly differentiated.

Colonic PDGFRα^{hi} FB consist of three transcriptionally distinct clusters originating from Fgfr2^{+} FB.

tSPACE analysis of adult FB subsets indicated that PDGFRα^{hi} FB originated from Fgfr2^{+} FB and then separated into three branches (Fig. 5E and F). To validate the idea that Fgfr2^{+} FB act as precursors of PDGFRα^{hi} FB we performed RNA velocity analysis (La Manno *et al.*, 2018) focusing on these subsets, which confirmed the directionality from Fgfr2^{+} FB to PDGFRα^{hi} FB (Fig. 6A). Re-clustering of only PDGFRα^{hi} FB uncovered three clusters that diverged along the three trajectory branches (Fig. 6B) that could be distinguished based on expression of *Cd9* and *Cd141* (thrombomodulin (Thbd)) (Fig. 6C). This generated clusters of CD9^{hi}CD141^{-}, CD9^{lo}CD141^{+} and CD141^{int} cells, all of which expressed the "telocyte" marker, *Fox1l* (Fig. S6A) (Shoshkes-Carmel *et al.*, 2018). Consistent with these findings, flow cytometric analysis of colonic PDGFRα^{hi} FB identified distinct clusters of CD9^{hi}CD141^{-} and CD9^{lo}CD141^{+} cells, together with CD9^{-} cells that expressed heterogeneous levels of CD141 and which we referred to as CD141^{int} FB (Fig. 6D). Analysis of the top DEG between these populations demonstrated that CD9^{hi}CD141^{-} cells expressed the highest levels of *Nrg1*, *Fgf7*, *Il1rl1* (ST2 (IL33 receptor)) and *Ptgs2*, that CD9^{lo}CD141^{+} cells expressed high levels of fibrosis-associated *Aspn* (Asporin),

Il11ra1 and *Cxcl12*, while CD141^{int} cells expressed high levels of *Cxcl10*, *Ly6c1*, *Adamdec1*, *Wnt4a* and *Plpp3* (Fig. S6B). The CD9^{lo}CD141^{+} cells, and to a lesser extent the CD141^{int} cells, expressed mRNA and protein for αSMA (Fig. S6C), a marker of myofibroblasts but not "telocytes". Immunohistochemical staining for PDGFRα and αSMA showed that αSMA^{+}PDGFRα^{hi} cells localize preferentially to the isthmus area just above colonic crypts, while αSMA^{-}PDGFRα^{hi} cells aligned directly underneath the epithelium at the top and bottom of crypts (Fig. 6F). The CD9^{lo}CD141^{+}, CD9^{hi}CD141^{-} and CD141^{int} FB also differentially expressed several epithelial support genes (Fig. 6G), suggesting that these populations may play distinct roles in supporting the epithelium at different stages of its development. Thus, adult colonic subepithelial PDGFRα^{hi} FB consist of spatially and transcriptionally distinct clusters that derive from Fgfr2^{+} FB.

Discussion

Recent studies have demonstrated considerable heterogeneity within the intestinal LP MSC compartment (Degirmenci *et al.*, 2018; Kinchen *et al.*, 2018; Bahar Halpern *et al.*, 2020; Hong *et al.*, 2020; McCarthy *et al.*, 2020; Roulis *et al.*, 2020) and suggested non-redundant roles for MSC subsets in intestinal homeostasis (Bahar Halpern *et al.*, 2020; David *et al.*, 2020; Hong *et al.*, 2020; McCarthy *et al.*, 2020; Wu *et al.*, 2021), inflammation (West *et al.*, 2017; Kinchen *et al.*, 2018; Smillie *et al.*, 2019) and cancer (Roulis *et al.*, 2020). As these studies have largely focused on single regions of the intestine (Degirmenci *et al.*, 2018; Kinchen *et al.*, 2018; Smillie *et al.*, 2019; Bahar Halpern *et al.*, 2020; David *et al.*, 2020; Hong *et al.*, 2020; McCarthy *et al.*, 2020; Roulis *et al.*, 2020; Wu *et al.*, 2021), it has been unclear whether there are regionally circumscribed differences in the composition of LP MSC subsets along the length of the intestine. By performing scRNA-seq analysis of small intestinal and colonic LP MSC from the same mice, we show here that both locations contain similar LP MSC subsets and that their pattern of expression of epithelial support genes by LP MSC subsets is largely conserved between these sites. Bulk RNA-seq analysis of sorted PDGFRα^{hi} subepithelial FB and interstitial Fgfr2^{+} FB confirmed that these subsets expressed distinct arrays of epithelial support genes irrespective of the tissue. However, both PDGFRα^{hi} and Fgfr2^{+} FB expressed higher levels of many epithelial support genes in the colon compared with their small intestinal counterparts, indicating a greater role for these FB in sustaining epithelial integrity in the colon. Consistent with this idea, WNT secretion by *Gli1*-expressing MSC is essential for homeostasis of the colonic epithelium (Degirmenci *et al.*, 2018; Karpus *et al.*, 2019; David *et al.*, 2020), but this is not the case in the small intestine where Paneth cells represent a major source of WNTs (Sato *et al.*, 2011).

Among the pathways significantly upregulated in colonic FB subsets compared with those in small intestine were *TGFβ regulation of extracellular matrix expression*, *epidermal growth factor receptor (EGFR1) signaling*, *brain derived neurotrophic factor (BDNF) signaling pathway* and *thyroid stimulating hormone (TSH) regulation of gene expression*. The relevance of these pathways in colonic versus small intestinal homeostasis remains to be determined. In contrast, few pathways were selectively upregulated in small intestinal FB subsets. Among these, several genes encoding chemokines, cytokines and cytokine receptors were significantly overexpressed by the Fgfr2^{+} FB population, suggesting that this subset may be involved in immune functions in the small intestine, but not in the colon. Interestingly, small intestinal FB also expressed higher levels of enzymes involved in vitamin A metabolism, consistent with previous findings that some small intestinal FB display aldehyde dehydrogenase activity and that there is increased retinoic acid receptor signaling in the small compared with the large intestine (Jaensson-Gyllenbäck *et al.*, 2011; Vicente-Suarez *et al.*, 2015). Collectively, these findings indicate that the local microenvironment plays a crucial role in regulating the transcriptional profile and specialization of intestinal FB in different regions of the intestine. The nature of the relevant factors and their importance in local homeostasis awaits further study.

Consistent with the idea that they may provide niche-specific support for local cells, the subsets of intestinal FB were located within distinct regions of the gut wall. As others have shown (Eyden, Curry and Wang, 2011; Kurahashi *et al.*, 2013; Shoshkes-Carmel *et al.*, 2018), we found that PDGFRα^{hi} FB directly underlie the intestinal epithelium in both the small and large intestine. In contrast to an earlier report that small intestinal CD81^{+} FB lie solely within the submucosa (Thomson *et al.*, 2018), we found these cells within both the submucosa and surrounding larger vessels deep in the mucosa, consistent with more recent studies (Hong *et al.*, 2020; McCarthy *et al.*, 2020). In addition, we demonstrate that CD81^{+} FB are found in

similar locations in the small intestine and colon. $CD81^+$ FB play an essential role in maintaining the epithelial stem cell niche in the small intestine, partly through their selective expression of the BMP antagonist, gremlin-1 (McCarthy *et al.*, 2020). Consistent with this, our scRNA-seq analysis demonstrated that $CD81^+$ FB were the major source of *Grem1* in the small intestine, although $CD81^+$ FB, $CD90^+$ FB and $Fgfr2^+$ FB all expressed *Grem1* in the colon, indicating potential redundancy between these subsets in supporting the colonic epithelial stem cell niche.

Less is known regarding the location of the intestinal $PDGFR\alpha^{lo}$ FB subsets that do not express CD81. Our immunohistochemical analysis demonstrated that $Fgfr2^+$ FB were located preferentially within the villus core towards the villus tip in the small intestine, while those in the colon were located between crypts. These findings are consistent with work on *Fgfr2*-mCherry reporter mice that suggested the $Fgfr2^+$ cells represent interstitial FB (Roulis *et al.*, 2020). Using PPARγ expression as a surrogate marker, we could show that $CD90^+$ ($PPAR\gamma^+$) FB were located near the base of colonic crypts, but were unable to do this in the small intestine, as we failed to identify a specific marker for this population. However, small intestinal $Igfbp5^+$ FB and colonic $CD90^+$ FB resembled one another transcriptionally and $Igfbp5^+$ FB were the only $CD34^+CD81^-$ FB subset in small intestine. As the cells with this phenotype were located primarily around small intestinal crypts, these results indicate that $CD90^+$ FB and $Igfbp5^+$ FB appear to represent a peri-cryptal population of FB in the colon and small intestine, respectively. Collectively, our findings confirm and extend previous work on the localization of intestinal FB subsets, highlighting their distinct transcriptional profiles and complex spatial organization within the mucosa.

The mesothelium is an epithelial monolayer that lines body cavities and internal organs, including the serosal surface of the intestine (Winters and Bader, 2013). Mesothelial cells undergo epithelial-mesenchymal transition (EMT) and can give rise to both SMC and FB in

response to injury in a number of tissues including the intestine (Miyoshi *et al.*, 2012; Rinkevich *et al.*, 2012; Koopmans and Rinkevich, 2018). During ontogeny, the intestinal mesothelium is a source of precursors for SMC in the intestinal muscle layers and vasculature (Wilm *et al.*, 2005), as well as for uncharacterized FB in the outer serosa of the intestine (Rinkevich *et al.*, 2012). In contrast, the origin(s) of the MSC subsets in the adult small intestine and colon LP has remained unclear. Here we used intestinal transplantation and lineage-tracing approaches to demonstrate that all adult small intestine and colon LP MSC subsets derive from *Gli1*-expressing progenitors present in the E12.5 intestine. At that time point, *Gli1* expression was restricted to mesothelial cells and an embryonic population of CD34$^-$PDGFRα$^+$ FB. scRNA-seq analysis showed that both mesothelial cells and a minor cluster of cells within the CD34$^-$PDGFRα$^+$ FB population expressed markers previously associated with FB progenitors, while tSPACE analysis suggested a direct trajectory connection between these two populations. These results provide strong evidence that the mesothelium is a source of FB precursors during early intestinal development and that these are capable of giving rise to all adult LP MSC subsets. Interestingly, while also orignating from embryonic mesothelium, smooth muscle cells (SMC) in adult LP developed along an embryonic trajectory that was distinct from that of adult FB in tSPACE. Thus, SMC and FB in the steady state adult LP appear to represent independent lineages that are specified during development.

tSPACE analysis revealed direct connections between embryonic MSC clusters and adult CD81$^+$ FB, suggesting that adult FB subsets arise from CD81$^+$ FB, rather than from distinct populations of intermediates that develop in the embryo. Consistent with this idea, CD81$^+$ FB locate in the submucosa and in the adventitia surrounding larger vessels at the base of the mucosa, an anatomical niche that contains MSC progenitors in other tissues (Sidney *et al.*, 2014; Díaz-Flores *et al.*, 2015; Kramann *et al.*, 2015, 2016; Sitnik *et al.*, 2016; Benias *et al.*, 2018; Merrick *et al.*, 2019). Furthermore, lineage tracing of *Grem1*$^+$ FB in the adult small

intestine has identified progenitors that give rise to subepithelial FB along the entire crypt-villus axis, in a process that is relatively rapid during the perinatal period, but takes around a year to be completed in adult intestine (Worthley *et al.*, 2015). In support of this idea, we found that *Grem1* expression is largely restricted to CD81^{+} FB in the small intestine.

While early trajectory analysis suggested a bifurcation downstream from CD81^{+} FB (Kinchen *et al.*, 2018), our tSPACE, pseudotime and Velocity analyses suggested that adult colonic CD81^{+} FB connected in a linear direction to adult CD90^{+} FB, then to Fgfr2^{+} FB and finally to PDGFRα^{hi} FB. Although we used different markers, our findings are consistent with those recently published by Buechler *et al* who suggested that adventitial *Pi16*$^{+}$ FB give rise first to *Col15α1*$^{+}$ FB and then to tissue specific FB clusters that in the intestine included *Fbln1*$^{+}$ FB and subsequently *Bmp4*$^{+}$ FB (Buechler *et al.*, 2021). Overlaying these clusters on our tSPACE trajectories demonstrated that the *Pi16*$^{+}$, *Col15α1*$^{+}$ and *Fbln1*$^{+}$ FB most closely resembled the CD81^{+}, CD90^{+} and Fgfr2^{+} FB we found in colon, while the *Bmp4*$^{+}$ FB were similar to our colonic PDGFRα^{hi} FB. The trajectory from CD81^{+} FB also correlates with the basal to apical localization of the downstream subsets in the colonic LP, indicating that this process may reflect factors present in distinct microenvironmental niches. Despite this strong evidence for linear differentiation from a single precursor, it should be noted that FB show evidence of slow turnover in the adult intestine (Worthley *et al.*, 2015; Kinchen *et al.*, 2018; Bahar Halpern *et al.*, 2020) and we cannot exclude the possibility that each FB subset might self-renew *in situ*.

Recent scRNA-seq studies have suggested that colonic subepithelial PDGFRα^{hi} FB are heterogeneous (Kinchen *et al.*, 2018; Roulis *et al.*, 2020) and here we found that PDGFRα^{hi} FB diverged into 3 clusters, which we could define as CD9^{lo}CD141^{+}, CD9^{hi}CD141^{-} and CD141^{int} FB. The CD9^{lo}CD141^{+} FB are likely related to the PDGFRα^{hi} FB sub-cluster S2a defined by Kinchen *et al*, as they expressed high levels of *Cxcl12*, while CD9^{hi}CD141^{-} FB

expressed high levels of *Nrg1* and so are likely related to the PDGFRα^{hi} FB subcluster S2b (Kinchen *et al.*, 2018). Although all three clusters expressed the telocyte marker *Foxl1* (Shoshkes-Carmel *et al.*, 2018), CD9^{lo}CD141^{+} FB and to a lesser extent CD141^{int} FB, expressed *Acta2*, coding for α–smooth muscle actin (αSMA), a marker often associated with myofibroblasts. αSMA^{+}PDGFRα^{hi} FB were located directly underneath the epithelium approximately half way up colonic crypts, suggesting that CD9^{lo}CD141^{+} and CD9^{hi}CD141^{-} subepithelial FB localise within distinct regions along the crypt axis. Of the three PDGFRα^{hi} FB clusters, CD9^{lo}CD141^{+} FB expressed the highest levels of the WNT antagonists *Wif1*, *Bmp3* and *Bmp4*. Therefore we speculate that the location of CD9^{lo}CD141^{+} FB half way up colonic crypts allows them to promote the terminal differentiation of epithelial cells as they migrate up the crypt (Qi *et al.*, 2017). In contrast, CD9^{hi}CD141^{-} FB expressed high levels of top of crypt-associated non-canonical *Wnt4, Wnt5a* and *Wnt5b* (Gregorieff *et al.*, 2005; Kosinski *et al.*, 2007) and *Tenascin C (Tnc)* (Probstmeier, Martini and Schachner, 1990; Bernier-Latmani *et al.*, 2015) and base of crypt-associated *Ptgs2* (the gene encoding COX-2) (Stzepourginski *et al.*, 2017; Roulis *et al.*, 2020), *Sema3a* (Karpus *et al.*, 2019). Thus, our findings indicate that each of the three PDGFRα^{hi} FB subsets may play distinct roles in colonic epithelial homeostasis.

In conclusion, our study provides a comprehensive mapping of intestinal MSC diversity, location and epithelial support function and highlights a central role for location along the intestinal length in regulating transcriptional profile and functional specialization. We also show that all adult MSC derive from *Gli1*-expressing embryonic mesothelial cells and we propose there is a linear developmental relationship between adult FB subsets that culminates in the development of a heterogeneous group of subepithelial PDGFRα^{hi} FB. Together our findings provide key insights into MSC diversity, development, function and interrelationships with relevance to intestinal development and homeostasis.

Acknowledgements

We thank Dr. J. Vandamme (DTU, Denmark) for performing scRNA-seq and library preparation, Dr. A.L Joyner (Memorial Sloan-Kettering Cancer Center) for providing *Gli1*-EGFP mice, Dr. S. Milling (University of Glasgow University, UK) for providing laboratory space and materials for experiments involving *Ackr4*^{tmlCcbl1} mice and Dr. R. Gentek (Edinburgh University, UK) for advice regarding embryonic lineage tracing. The SNP&SEQ Platform is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. This work was supported by grants awarded to W.W.A. from the Lundbeck foundation (R155-2014-4184), Denmark, and the Gut Cell Atlas, an initiative funded by the Leona M. and Harry B. Helmsley Charitable Trust, US.

Author contributions

The study was designed by S.I.P, S.S., and W.W.A. Experiments in Denmark were performed by S.I.P., S.S. U.M and J.J., the grafting experiments were performed by K.W. and S.I.P, experiments in Glasgow were performed by S.I.P. and A.T.A. with support from R.J.B.N., bioinformatics analyses was performed by L.W. with support from K.N., K.F.B, E.C.B and S.B. A.M provided valuable intellectual input throughout. The manuscript was written by S.I.P. and W.W.A after input from all authors.

Declaration of Interests

The authors declare no competing interests.

Figure legends

Figure 1. Intestinal MSC subsets are broadly conserved across intestinal segments. (**A-B**) Uniform Manifold Approximation and Projection (UMAP) colored by unsupervised Louvain clustering of murine small intestinal (**A**) and (**B**) colonic MSC. Results are from 2 independent experiments/organ with 3 pooled mice/experiment. (**C**) Pearson correlations between averaged cluster expressions of Louvain clusters from small intestinal and colonic MSC based on 1301 overlapping variable genes. Unsupervised hierarchical clustering indicate similarity of subsets within each tissue. (**D-I)** Immunohistochemical staining of mouse jejunum (**D-F**) or colon (**G-I**) for indicated antigens. (**D, F-I**) Region (R)1' and R2' represent magnifications of R1 and R2 quadrants (yellow squares). (**D**) Arrows indicate location of $CD81^+$ FB ($CD81^+CD34^+CD31^-$ cells) and stars, location of $Igfbp5^+$ FB (CD81-CD34+ $CD31^-$ cells). (**E**) Images of villus tip (left) and crypt (right). (**F-I**) Arrows indicate location of (**F**) $Fgfr2^+$ FB ($CXCL14^+PDGFR\alpha^+CD34^-$ cells), (**G**) $CD90^+$ FB ($PPAR\gamma^+CD34^+CD31^-$ cells), (**H**) $Fgfr2^+$ FB ($Fgfr2^+CD34^+PDGFR\alpha^+$ cells) and (**I**) $CD81^+$ FB ($CD81^+CD34^+CD31^-$ cells). Results are representative stains from (**D-F, H**) 2 and (**G** and **I**) 3 experiments. See also Figure S1.

Figure 2. Despite similar FB subset composition, small intestinal and colonic FB display regional transcriptional specialization. (**A**) Heatmaps showing scaled expression (integrated data) of selected epithelial support genes by indicated MSC subsets. (**B**) Pearson correlations between averaged cluster expressions of Louvain clusters from scRNA-seq and bulk RNA-seq datasets based on 1937 (small intestine, left) and 1925 (colon, right) overlapping variable genes. Bulk RNA-seq data is from sorted MSC subsets from 3 independent experiments. Unsupervised hierarchical clustering indicate similarities of bulk RNA-seq subsets within each tissue. (**C-D**) Flow cytometric analysis of adult small intestinal (**C**) and colonic (**D**) $Itg\beta1^+$

MSCs from *Ackr4.GFP* mice. Representative staining of 2 experiments with 2-4 mice/experiment. Colored gates represent indicated MSC subsets. PCs - pericytes; SMCs - smooth muscle cells, FB - fibroblast. (**E**) Principal component analysis (PCA) of bulk RNA-seq data from indicated sorted FB populations. Results are from 3 independent sorts/population. (**F**) Volcano plots showing differentially expressed genes (DEGs) between small intestinal and colonic PDGFRα^{hi} FB (left) and Fgfr2^{+} FB (right). Dotted horizontal line denotes significant adjusted p-value of 0.05, vertical dotted lines denote $log_2FC = 0$ and the log_2FC of +/- 1.5. (**G-J**) Heatmap representations of averaged transcription levels of indicated genes within sorted FB subsets. Data are averaged from 3 independent bulk RNA-seq datasets. (**G**) Epithelial support genes, (**H**) cytokines and cytokine receptors, (**I**) chemokines, (**J**) vitamin A metabolism. Gene lists for **I** were selected based on the epithelial support list in (**A**) while those in **H-J** were differentially expressed between either small intestinal and colonic PDGFRα^{hi} FB or between small intestinal and colonic Fgfr2^{+} FB. Identified DEG that are $1.5<|log_2FC|$ are listed to the right of (**G**) or below (**H-J**) the heat maps and. See also Figure S2.

Figure 3. Adult intestinal MSC subsets derive from intestinal precursors present in E12.5 intestine. (**A-B**) Flow cytometric analysis of Itgβ1^{+} MSCs isolated from indicated organs on embryonic day (E) 12.5. (**B**) Right hand plots show expression of DPP4 (CD26) on gated PDPN^{+}PDGFRα^{-} (blue) and PDPN^{+}PDGFRα^{+} (red) cells from plots on left. Data are representative of (**A**) 4 experiments with 2-8 embryos/experiment, or (**B**) 3 experiments with 6-8 individual embryos. (**C**) Workflow of transplantation of E12.5 intestine from YFP^{+} mice under the kidney capsule of WT recipients. (**D**) Flow cytometric analysis of YFP^{+}Itgβ1^{+} MSC in intestinal grafts 4 weeks after transplantation. Results are representative of 2 experiments with 4 (small intestine) or 2-3 (colon) grafts/experiment. (**E**) UMAP dimensionality reduction

of scRNA-seq data colored by Louvain clustering from FACS purified YFP^{+}Itgβ1^{+} MSC isolated from colonic grafts 4 weeks after transplantation. Data are from 8624 single cells from 3 pooled colonic grafts with an average of 2223 genes/cell. (**F**) Pearson correlations of averaged gene expression in colonic graft and adult colon MSC clusters based on 1486 overlapping variable genes. (**G**) Heatmap showing scaled transcription levels (integrated data) of selected epithelial support genes within the putative corresponding FB clusters identified in (**E**). See also Figure S3.

Figure 4. Adult intestinal MSC derive from *Gli1*$^{+}$ embryonic precursors. (**A**) Representative flow cytometric analysis and (**B**) Proportions of indicated cells expressing EGFP in the small intestine and colon of embryonic E12.5 *Gli1*-EGFP mice. Results are from 8 individual embryos, with each circle representing an individual embryo. (**C**) Workflow of lineage-tracing experiments. *R26R*.EYFP females were mated overnight with *Gli1*.CreERT2$^{+/-}$ males and pregnant dams injected i.p. with 4-Hydroxytamoxifen (4-OHT) at E11.5. (**D**) Proportions of indicated MSC subset expressing YFP in small intestine and colon of 5.5-7 week old *Gli1*.CreERT2$^{+/-}$.*R26R*.EYFP and *Gli1*.CreERT2$^{-/-}$*R26R*.EYFP littermates. Results are from 4 independent experiments with 2-8 mice/experiment. Each circle represents an individual mouse. Bars represent the means and SD. See also Figure S4.

Figure 5. Trajectory analysis indicates that adult intestinal MSC subsets derive from embryonic *Gli1*$^{+}$ mesothelial cells. (**A**) UMAP dimensionality reduction of scRNA-seq data colored by Louvain clustering from FACS purified Itgβ1^{+} MSC from the colon of embryonic day E12.5 mice. Data are from 9632 single cells from 2 pooled experiments using 3-5 embryonic colons/experiment, with an average of 2521 genes/cell. (**B**) UMAP of E12.5 large intestinal Itgβ1^{+} MSC overlaid with expression of the indicated mesothelium associated genes. (**C** and **D**) tSPACE principal component analysis (tPC 1-3) projection of pooled adult colonic

and E12.5 large intestinal MSC. (**C**) Clusters are color coded as in (**A**) for embryonic clusters or as in Fig. 1B for adult clusters. Arrow heads indicate connections between embryonic and adult clusters. (**D**) Pseudotime analysis using averaged values of the 9 trajectories with starting point in mesothelial cells superimposed on tPC 1-3. (**E** and **F**) tSPACE projections of adult colonic MSC in tPC1-3. (**F**) Pseudotime analysis superimposed on (**E**) using averaged values of the 215 trajectories starting in $CD81^+$ FB. See also Figure S5.

Figure 6. Subepithelial PDGFRα^{hi} FB consist of three transcriptionally distinct clusters originating from Fgfr2$^+$ FB. (**A**) tSPACE projection of adult colonic MSC in tPC1 and 3 highlighting $Fgfr2^+$ FB and PDGFRα^{hi} FB overlaid with RNA Velocity. (**B**) tSPACE projection of colonic MSC in tPC1-3, highlighting $Fgfr2^+$ FB and three PDGFRα^{hi} FB clusters. (**C**) UMAP dimensionality reduction of re-clustered colonic PDGFRα^{hi} FB (top panel), with *Cd9* (bottom left panel) and *Cd141* (bottom right panel). (**D**) Representative flow cytometric analysis of CD9 and CD141 expression by colonic PDGFRα^{hi} FB. Representative plots from 2 experiments with 3 mice/experiment. (**E**) *Acta2* (αSMA) gene expression projected onto UMAP of colonic PDGFRα^{hi} FB. (**F**) Immunohistochemical staining of colonic tissue for indicated antigens. R1','' and R2','' represent magnifications of R1 and R2 quadrants (yellow squares) on left image. Results are representative stains from 3 experiments with 3 mice/experiment. Arrows indicate $\alpha SMA^+PDGFR\alpha^{hi}$ FB. (**G**) Heatmap showing scaled transcription levels (integrated data) of significantly ($p<0.05$) differentially expressed epithelial support genes between the PDGFRα^{hi} FB clusters. See also Figure S6.

.

METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr WW Agace (wiag@dtu.dk).

Materials Availability

This study did not generate new unique reagents or mouse strains.

Data and Code Availability

Single-cell RNA-seq and bulk RNA-seq data has been deposited at NCBI GEO and are publicly available as of the date of publication. Accession numbers will be listed when published. Microscopy data reported in this paper will be shared by the lead contact upon request. This paper does not report original code. Any additional information required to re-analyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

Gli1$^{tm3(cre/ERT2)Alj}$ (*Gli1*-CreER^{T2}, 007913 Jackson laboratories), B6.129X1-Gt(ROSA)26Sor$^{tm1(EYFP)Cos}$/J (*R26R*.EYFP, 006148 Jackson laboratories), *Gli1*-EGFP (Garcia *et al.*, 2010) and EYFP mice (obtained by crossing *R26R*.EYFP with the relevant Cre mice) were bred and maintained at the Bio-Facility animal house (Technical University of Denmark). C57BL/6Nrj mice were purchased from Janvier Labs (Le Genest-Saint-Isle, France).

Ackr4^{tmlCcbl1} mice (*Ackr4*.EGFP) (Heinzel, Benz and Bleul, 2007) were bred and maintained in the Central Research Facility, Glasgow University. Adult mice were used between 5.5 and 12w of age. Mice of both genders were used in all experiments and littermates were used as controls. All experiments were approved by the Danish Animal Experiments Inspectorate, or with ethical approval under a Project Licence from the the UK Home Office.

METHOD DETAILS

Kidney grafting

EYFP male mice were mated overnight with C57BL/6Nrj females and the following morning was defined as gestational day 0.5 (E0.5). Pregnant dams were sacrificed at E12.5 and small and large intestine were dissected from embryos under a stereo microscope (VWR). Adult WT mice were anaesthetized by i.p injection of Ketaminol Vet. (100mg/kg, MSD animal health) and Rompun Vet. (10mg/kg, Bayer) and were injected subcutaneously with Bupaq (0.1mg/kg, Richter Pharma). Washed embryonic intestine was transplanted under the kidney capsule of anesthetized recipients as described previously (Ferguson, Parrott and Connor, 1972). Recipients were sacrificed at the time points indicated and grafts were dissected and cut into pieces prior to cell isolation as described below.

In vivo lineage tracing

Gli1-CreER^{T2} male mice were mated overnight with *R26R*.EYFP females and the following morning was defined as gestational day E0.5. At E11.5, pregnant dams were injected i.p. with 4-hydroxytamoxifen ((4-OHT), 1.6 mg, Sigma) and progesterone (0.8 mg, Sigma) in 160µl PBS with 25% Kolliphor (Sigma)/25% ethanol (Fischer Scientific). Small and large intestine were isolated from embryos or weaned offspring at the time points indicated.

Cell isolation

Intestinal cell suspensions were generated as described previously (Schulz *et al.*, 2009) with minor changes. Briefly, washed intestinal tissue was opened longitudinally and Peyer's patches removed. For scRNA-seq and bulk RNA-seq experiments on adult intestine, muscularis externa was stripped away using tweezers. Tissues were cut into 0.5-1 cm pieces and epithelial cells removed by 3 consecutive rounds of incubation in HBSS supplemented with HEPES (15mM), sodium pyruvate (1mM), penicillin/streptomycin (100 U/mL), gentamycin (0.05 mg/mL), EDTA (2mM) (all Invitrogen) and FCS (2.5%) (Sigma), for 15 min at 37°C with constant shaking at 350 rpm. After each incubation, samples were shaken for 10 sec and medium containing epithelial cells and debris was discarded. For colonic tissues, DL-dithiothreitol (5mM) (Sigma) was added at the first incubation step. Remaining tissue pieces were digested with collagenase P (0.6U/mL, Sigma) or with Liberase TM (0.325U/mL, Roche) and DNAse I (31 µg/mL, Roche) in R10 medium (RPMI 1640, sodium pyruvate (1mM), HEPES (10 mM), penicillin/streptomycin (100 U/mL), gentamycin (0.05mg/mL), and 10% FCS) for up to 30 min at 37°C with constant shaking at 550 rpm (small intestine) or with a magnetic stirrer and at 280 rpm (large intestine). For bulk RNA-seq cells were treated with ACK lysing buffer (Gibco) to lyse red blood cells prior to sorting. For isolation of cells from embryonic intestine, tissues were digested directly for 30 min at 37°C in Eppendorf tubes with constant shaking at 900 rpm. The resulting cell suspensions were filtered through a 70 µm filter and washed in MACS buffer (phosphate buffered saline (PBS) with 3% FCS and 2 mM EDTA) twice prior to subsequent analyses.

Flow cytometry and cell sorting

Cell suspensions were stained with fluorochrome labelled primary antibodies (see table 1) in Brilliant stain buffer (BD Biosciences) for 30 min on ice. Flow cytometry was performed on an LSR Fortessa II (BD Biosciences), FACSAria Fusion (BD Biosciences), or FACSMelody (BD Biosciences) and analysed with FlowJo software (TreeStar). Dead cells were identified by staining with either 7-AAD (eBioscience) or Zombie UV fixable viability dye (BD Biosciences) and cell doublets were excluded on the basis of FSC-A/FSC-H. For intracellular staining, cells were stained for surface antigens, fixed with FoxP3 Staining Buffer set (eBioscience) and stained for αSMA in FoxP3 Permeabilization buffer (eBioscience). After washing, cells were stained with antibodies to surface antigens not compatible with fixation according to the manufacturer's instructions.

Immunohistochemistry

Tissues were fixed in paraformaldehyde (4%, PFA) and sectioned (70 μm) using a Vibratome (Leica VT12000S). Sections were incubated in PBS containing BSA (1%) and Triton-X100 (0.3%) for 1 hour at room temperature (RT) to block non-specific staining and incubated with fluorochrome conjugated or unconjugated primary antibodies (see Table S1) overnight at 4ºC. After washing with PBS containing Triton-X100 (0.3%), tissues were incubated with secondary antibodies at RT for 2-4 hours (see table 1). For detection of CD81, staining with biotinylated anti-CD81 was enhanced using the Biotinyl Tyramide kit (Perkin Elmer) according to the manufacturer's instructions after blocking of endogenous biotin using Streptavidin/Biotin Blocking kit (Invitrogen). Endogenous peroxidase was inactivated by incubating tissues with 3% H_2O_2 for 30 min at RT before incubation with streptavidin-horse radish peroxidase (HRP). Sections were analysed under 40x magnification using a Zeiss LSM 710 confocal microscope and the images were processed using Zeiss Zen and Imaris software.

For histological analysis of kidney grafts, tissue pieces were fixed in 4 % paraformaldehyde for 8 h and paraffin-embedded sections were stained with hematoxylin and eosin.

Library Preparation and sequencing

Single cell RNA-seq. Sorted cells were washed in cold PBS containing bovine serum albumin (0.04%), counted and diluted to the desired concentration following 10X Genomics guidelines (10x Genomics, CG000053_CellPrepGuide_RevC). ScRNA-seq libraries were prepared according to the manufacturer's instructions using Chromium Single Cell 3′ Library & Gel Bead Kit v3 (10x Genomics, PN-1000092) or 5' kit Chromium Single Cell 5′ Library & Gel Bead Kit (10x Genomics, PN-1000006) and Chromium Chip B Single Cell Kit (PN-1000074) with the Chromium Controller & Next GEM Accessory Kit (10x Genomics, PN-120223). In brief, single cells, reverse transcription reagents, Gel Beads containing barcoded oligonucleotides, and oil were combined on a microfluidic chip to form Gel Beads in Emulsion (GEMs). Individual cells were lysed inside the GEMs and the released poly-A transcripts were barcoded with an Illumina R1 sequence, a 10X barcode and a Unique Molecular Identifier (UMI) during reverse transcription (RT). After RT, GEMs were broken, barcoded cDNA was purified using Dynabeads MyOne silane (10x Genomics, PN-2000048) and amplified by Polymerase Chain Reaction (PCR). Amplified cDNA were cleaned up with SPRIselect Reagent kit (Beckman Coulter, B23318). Indexed sequencing libraries were constructed by enzymatic fragmentation, end-repair and A-tailing, before a second and final PCR amplification using the Chromium i7 Sample Index (10x Genomics, PN-220103), introducing an Illumina R2 sequence, a unique sample index (allowing multiplex sequencing) and P5/P7 Illumina sequencing adaptors to each library. Library quality control and quantification were performed using a KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems, KK4873) and the 2100 Bioanalyzer equipped with a High Sensitivity DNA kit (Agilent, 5067-

4626). Muliplexed libraries were pooled and sequenced either by NextSeq 500/550 High Output v2.5 kit (150 cycles) at the Center of Excellence for Fluorescent Bioanalytics (KFB, University of Regensburg, Germany) or by Novaseq 6000 S1 or S2 (200 cycles) at the SNP&SEQ Technology Platform (Uppsala, Sweden) with the following sequencing run parameters: Read1 - 28 cyles; i7 index - 8 cycles; Read2 – 126 cycles at a depth of at least 100M reads/sample.

Bulk RNA-seq. Cells were sorted into RLT buffer and total RNA was isolated using the RNeasy Micro kit. Extraction was performed according to the manufacturer's protocol, with an on-column DNAse digestion step added after the first wash buffer step. RNA quality and quantity were measured using the 2100 BioAnalyzer equipped with RNA6000 Pico chip (Agilent Technologies). RNA was subjected to whole transcriptome amplification using Ovation RNA-Seq System V2 kit and amplified cDNA samples were purified using the MinElute Reaction Cleanup kit . The quantity and quality of the cDNA samples were measured using the 2100 BioAnalyzer equipped with DNA1000 chip (Agilent technologies) and the Nanodrop (ThermoFisher Scientific). Libraries were constructed with the Ovation Ultralow system V2 kit, following the manufacturer's instructions. Amplified cDNA (100 ng) was fragmented by sonication using a Bioruptor Pico (Diagenode), sheared cDNA end-repaired to generate blunt ends and ligated to Illumina adaptors with indexing tags followed by AMPure XP bead purification. Library size distribution was evaluated using a 2100 Bioanalyzer equipped with DNA1000 chip (Agilent technologies) and quantified using KAPA library Quantification Kit Illumina platforms (Kapa Biosystems). Libraries were diluted and pooled at equimolar concentration (10 nM final) and sequenced on the Hiseq2500 platform (Illumina) using 50bp single reads (Center for Genomic Regulation, Spain) with a depth of 15-20M reads per sample.

Computational analysis

Single cell RNA-seq. Alignment of scRNA-seq data to reference genome mm10 was performed with CellRanger (version 3.0.2 & 3.1.0) (Dobin *et al.*, 2013; Zheng *et al.*, 2017). The data was imported into R (version 4.0.1) (R Core Team, 2020) and processed to remove debris and doublets in individual samples by looking at gene, read counts and mitochondrial gene expression. Variable genes were calculated with Seurats FindVariableFeatures function and selection method set to "vst" (version 3.1.5) (Stuart and Satija, 2019). The respective samples and all their overlapping genes were then integrated with anchor integration for Seurat. Cell cycle effects were regressed out with linear regression using a combination of the build-in function in Seurat and scoring gene sets from ccremover per cell (Li, Jun; Barron, 2017) during scaling of the gene expression. The datasets were dimensionality reduced first with PCA and then UMAP and clustered with Louvain clustering all using Seurat. After initial clustering, contaminating cells were removed and an additional round of clustering and dimensionality reduction with UMAP was run on the cells of interest. DEGs were identified using Seurat FindAllMarkers function with the default test setting (non-parameteric Wilcoxon Rank Sum test). Expression of gene modules in the form of published signature gene sets were calculated with AddModuleScore (Seurat) taking the top DEG from telocytes (10 genes), Lo-1 FB (10 genes) and Lo-2 FB (7 genes) reported by McCarthy *et al* (McCarthy *et al.*, 2020), the top 10 DEG from FB1-5, MC and SMCs reported by Hong *et al* (Hong *et al.*, 2020), and the top 20 DEG from *Pi16*$^{+}$, *Col15a1*$^{+}$, *Fbln1*$^{+}$ and *Bmp4*$^{+}$ FB reported by Buechler *et al* (Buechler *et al.*, 2021). Pearson correlations between datasets were calculated based on average expressions per cluster of overlapping variable genes and plotted with heatmap.2 (version 3.0.3) (Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller and Venables,

2020). Heatmaps were constructed with a modified version of Seurats DoHeatmap to allow for multiple grouping variables. Data plotted in expression heatmaps was scaled based on the anchor integrated data. Data imputation was performed per dataset across samples on raw count data with magicBatch (Brulois, 2020) where the affinity matrix used was Seurat's batch-corrected PCA coordinates. Trajectories and trajectory spaces were determined with tSPACE (Dermadi *et al.*, 2020) on the top 2000 imputed variable genes for adult trajectories and top 1000 imputed variable genes for the integrated E12.5 and adult trajectory. GO analysis was performed using GO Biological Process 2018 (Ashburner *et al.*, 2000) from Enrichr computational biosystems (Chen *et al.*, 2013; Kuleshov *et al.*, 2016; Xie *et al.*, 2021).

Bulk RNA-seq. Raw RNA sequencing data from the 30 samples were pre-processed with TrimGalore (version 0.4.0) and FastQC (version 0.11.2). Pseudo-alignment of reads was performed with Kallisto (version 0.42.5) to obtain RNA expression information. To assess correlations between bulk-seq samples and SC clusters Pearson correlations based on SC variable genes were calculated between the bulk-seq samples and the pseudo-bulk of the SC clusters for the individual tissues and visualized with heatmap.2 (part of gplot package).

For all DESeq2 (1.26.0 (Love, Huber and Anders, 2014)) analysis, transcripts identified in less than 3 replicates and at levels below 6 reads were filtered out prior to further analysis. Heatmaps of bulk-seq data expression was created in R with the ComplexHeatmap package (version 2.7.11) and volcano plots with ggplot2 (version 3.3.1). For the comparison between tissues, DEGs were only classified as significant if they had a $|\log2FC| > 1.5$ and adjusted p-value < 0.05. GO analysis was performed using BioPlanet 2019 (Huang *et al.*, 2019) from Enrichr computational biosystems (Chen *et al.*, 2013; Kuleshov *et al.*, 2016; Xie *et al.*, 2021).

Statistical analysis

Statistical significance was determined with a 2-way ANOVA with Benjamini, Krieger and Yekutieli multiple comparisons and performed in Prism software (GraphPad). *p < 0.05, **p < 0.01, ***p < 0.001.

Table 1: Antibodies

Antibodies		
BV711 anti-mouse CD9 clone KMC8	BD Biosciences	Cat#740696, RRID: AB_2740380
AF700 anti-mouse CD11b clone M1/70	BioLegend	Cat#101222, RRID: AB_493705
AF700 anti-mouse CD11c clone N418	BioLegend	Cat#117320, RRID: AB_528736
PE anti-mouse CD26 clone H194-112	BioLegend	Cat#137804, RRID: AB_2293047
AF594 anti-mouse CD31 clone MEC13.3	BioLegend	Cat#102520, RRID: AB_2563319
AF647 anti-mouse CD31 clone MEC13.3	BioLegend	Cat#102515, RRID: AB_2161030
BV605 anti-mouse CD31 clone 390	BioLegend	Cat#102427, RRID: AB_2563982
BV650 anti-mouse CD31 clone 390	BD Biosciences	Cat#740483, RRID: AB_2740207
PerCP/Cy5.5 anti-mouse CD31 clone 390	BioLegend	Cat#102420, RRID: AB_10613644
BV421 anti-mouse CD34 clone RAM34	BD Biosciences	Cat#562608, RRID: AB_11154576
FITC anti-mouse CD34 clone RAM34	Thermo Fischer Scientific	Cat#11-0341-85, RRID: AB_465022
Rat anti-mouse CD34 clone RAM34	Thermo Fischer Scientific	Cat#14-0341-82, RRID: AB_467210
AF700 anti-mouse CD45 clone 30-F11	Thermo Fischer Scientific	Cat#56-0451-82, RRID: AB_891454
AF700 anti-mouse CD45.2 clone 104	Thermo Fischer Scientific	Cat#56-0454-82, RRID: AB_657752
APCCy7 anti-mouse CD45.2 clone 104	BioLegend	Cat#109824, RRID: AB_830789
Biotin anti-mouse CD81 clone Eat-2	BioLegend	Cat#104903, RRID: AB_313138
APCCy7 anti-mouse CD90.2 clone 53-2.1	BD Biosciences	Cat#561641, RRID: AB_10898013
FITC anti-mouse CD90.2 clone 53-2.1	Thermo Fischer Scientific	Cat#11-0902-82, RRID: AB_465154
PE anti-mouse CD141 clone REA964	Miltenyi	Cat#130-116-017, RRID: AB_2727308
BUV395 anti-mouse CD146 clone ME-9F1	BD Biosciences	Cat#740330, RRID: AB_2740063
AF488 anti-mouse αSMA clone 1A4	Abcam	Cat#ab184675, RRID: AB_2832195
AF700 anti-mouse B220 clone RA3-6B2	BioLegend	Cat#103232, RRID: AB_493717
APC anti-mouse BP3 clone BP3	BioLegend	Cat#140208, RRID: AB_10901172

BV650 anti-mouse BP3 clone BP-3	BD Biosciences	Cat#740611, RRID: AB_2740311
BV786 anti-mouse BP3 clone BP-3	BD Biosciences	Cat#741012, RRID: AB_2740634
AF555 anti-mouse CXCL14 rabbit polyclonal	BIOSS	Cat#bs-1503R-A555
AF647 anti-mouse EpCAM clone G8.8	BioLegend	Cat#118211, RRID: AB_1134104
APCCy7 anti-mouse EpCAM clone G8.8	BioLegend	Cat#118218, RRID: AB_2098648
BV510 anti-mouse EpCAM clone G8.8	BD Biosciences	Cat#563216, RRID: AB_2738075
PerCP-eF710 anti-mouse EpCAM clone G8.8	Thermo Fischer Scientific	Cat#46-5791-82, RRID: AB_10598205
PE anti-mouse ESAM clone 1G8	BioLegend	Cat#136204, RRID: AB_1953301
Rabbit anti-mouse FGFR2 polyclonal	Proteintech	Cat#13042-1-AP, RRID: AB_10642943
AF700 anti-mouse Gr-1 clone RB6-8C5	BioLegend	Cat#108422, RRID: AB_2137487
AF488 donkey anti-rat IgG	Jackson IR	Cat#712-545-153, RRID: AB_2340684
Cy3 donkey anti-rat IgG	Jackson IR	Cat#712-166-150, RRID: AB_2340668
AF647 donkey anti-goat IgG	Jackson IR	Cat#705-605-147, RRID: AB_2340437
AF647 donkey anti-rabbit IgG	Jackson IR	Cat#711-605-152, RRID: AB_2492288
BV605 anti-mouse Itgb1 clone HM β1-1	BD Biosciences	Cat#740365, RRID: AB_2740097
APC anti-mouse L1CAM clone 555	Miltenyi	Cat#130-102-221, RRID: AB_2655594
APC anti-mouse NCAM clone 809220	R&D	Cat#FAB7820A
AF700 anti-mouse NK1.1 clone PK136	BioLegend	Cat#108730, RRID: AB_2291262
BV421 anti-mouse PDGFRα clone APA5	BD Biosciences	Cat#566293, RRID: AB_2739666
Goat anti-mouse PDGFRα polyclonal	R&D	Cat#AF1062, RRID: AB_2236897
PE/CF594 anti-mouse PDGFRα clone APA5	BD Biosciences	Cat#562775, RRID: AB_2737786
PECy7 anti-mouse PDPN clone 8.1.1	Thermo Fischer Scientific	Cat#25-5381-82, RRID: AB_2573460
Rabbit anti-mouse PPARγ polyclonal	Invitrogen	Cat#PA5-25757, RRID: AB_2543257
AF700 anti-mouse Ter119 clone Ter119	BioLegend	Cat#116220, RRID: AB_528963
APC-eF780 anti-mouse Ter119 clone Ter119	Thermo Fischer Scientific	Cat#47-5921-82, RRID: AB_1548786
BV510 streptavidin	BD Biosciences	Cat#563261, RRID: AB_2869477

Supplementary Information

The manuscript contains 6 supplemental Figures and 3 supplemental Tables**.**

Supplemental Table 1: List of genes that are differentially expressed between small intestinal and colonic PDGFRα^{hi} FB ranked in order of significance.

Supplementary Table 2. List of genes that are differentially expressed between small intestinal and colonic Fgfr2^{+} FB ranked in order of significance.

Supplementary Table 3. List of common genes that are differentially expressed between both small intestinal and colonic PDGFRα^{hi} FB and small intestinal and colonic Fgfr2^{+} FB.

References

Agace, W. W. and McCoy, K. D. (2017) 'Regionalized Development and Maintenance of the Intestinal Adaptive Immune Landscape', *Immunity*. Elsevier Inc., 46(4), pp. 532–548. doi: 10.1016/j.immuni.2017.04.004.

Aoki, R. *et al.* (2016) 'Foxl1-Expressing Mesenchymal Cells Constitute the Intestinal Stem Cell Niche', *Cellular and Molecular Gastroenterology and Hepatology*, 2(2), pp. 175–188. doi: 10.1016/j.jcmgh.2015.12.004.

Ashburner, M. *et al.* (2000) 'Gene Ontology: tool for the unification of biology', *Nature Genetics*, 25(1), pp. 25–29. doi: 10.1038/75556.

Bae, S. *et al.* (2011) 'Combined omics analysis identifies transmembrane 4 L6 family member 1 as a surface protein marker specific to human mesenchymal stem cells', *Stem Cells and Development*, 20(2), pp. 197–203. doi: 10.1089/scd.2010.0127.

Bahar Halpern, K. *et al.* (2020) 'Lgr5+ telocytes are a signaling source at the intestinal villus tip', *Nature Communications*. Springer US, 11(1), p. 1936. doi: 10.1038/s41467-020-15714-x.

Barker, N. (2014) 'Adult intestinal stem cells: Critical drivers of epithelial homeostasis and regeneration', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 15(1), pp. 19–33. doi: 10.1038/nrm3721.

Benias, P. C. *et al.* (2018) 'Structure and distribution of an unrecognized interstitium in human tissues', *Scientific Reports*, 8(1), pp. 1–8. doi: 10.1038/s41598-018-23062-6.

Bernier-Latmani, J. *et al.* (2015) 'DLL4 promotes continuous adult intestinal lacteal regeneration and dietary fat transport', *Journal of Clinical Investigation*, 125(12), pp. 4572–4586. doi: 10.1172/JCI82045.

Beswick, E. J. *et al.* (2014) 'TLR4 Activation Enhances the PD-L1-Mediated Tolerogenic Capacity of Colonic CD90+ Stromal Cells.', *Journal of Immunology*, 193(5), pp. 2218–2229. doi: 10.4049/jimmunol.1203441.

Brügger, M. D. *et al.* (2020) 'Distinct populations of crypt-associated fibroblasts act as signaling hubs to control colon homeostasis', *PLoS biology*, 18(12), pp. 1–20. doi: 10.1371/journal.pbio.3001032.

Brulois, K. (2020) 'magicBatch: R wrapper for the original python implementation of the Marcov Affinity-based Graph Imputation of Cells (MAGIC) algorithm. R package version 0.1.0.', *github*. Available at: https://github.com/kbrulois/magicBatch.

Buechler, M. B. *et al.* (2021) 'Cross-tissue organization of the fibroblast lineage', *Nature*,

(September 2020). doi: 10.1038/s41586-021-03549-5.

Castagnaro, L. *et al.* (2013) 'Nkx2-5+islet1+ mesenchymal precursors generate distinct spleen stromal cell subsets and participate in restoring stromal network integrity', *Immunity*. Elsevier Inc., 38(4), pp. 782–791. doi: 10.1016/j.immuni.2012.12.005.

Chen, E. Y. *et al.* (2013) 'Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool Edward', *BMC Bioinformatics*, 128(4), pp. 617–619. doi: 10.1007/s00701-014-2321-4.

David, M. B. *et al.* (2020) 'Distinct populations of crypt-associated fibroblasts act as signaling hubs to control colon homeostasis', *PLoS Biology*, 18(12), pp. 1–20. doi: 10.1371/journal.pbio.3001032.

Degirmenci, B. *et al.* (2018) 'GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells', *Nature*. Springer US, 558(7710), pp. 449–453. doi: 10.1038/s41586-018-0190-3.

Dermadi, D. *et al.* (2020) 'Exploration of Cell Development Pathways through High-Dimensional Single Cell Analysis in Trajectory Space', *iScience*. Elsevier Inc., 23(2), p. 100842. doi: 10.1016/j.isci.2020.100842.

Díaz-Flores, L. *et al.* (2015) 'Human resident CD34+ stromal cells/telocytes have progenitor capacity and are a source of αSMA+ cells during repair', *Histology and Histopathology*, 30(5), pp. 615–627. doi: 10.14670/HH-30.615.

Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.

Driskell, R. R. *et al.* (2013) 'Distinct fibroblast lineages determine dermal architecture in skin development and repair', *Nature*. Nature Publishing Group, 504(7479), pp. 277–281. doi: 10.1038/nature12783.

Dulauroy, S. *et al.* (2012) 'Lineage tracing and genetic ablation of ADAM12 + perivascular cells identify a major source of profibrotic cells during acute tissue injury', *Nature Medicine*, 18(8), pp. 1262–1270. doi: 10.1038/nm.2848.

Eyden, B., Curry, A. and Wang, G. (2011) 'Stromal cells in the human gut show ultrastructural features of fibroblasts and smooth muscle cells but not myofibroblasts', *Journal of Cellular and Molecular Medicine*, 15(7), pp. 1483–1491. doi: 10.1111/j.1582-4934.2010.01132.x.

Fagarasan, S. *et al.* (2001) 'In situ class switching and differentiation to IgA-producing cells in the gut lamina propria', *Nature*, 413(6856), pp. 639–643. doi: 10.1038/35098100.

Fawkner-Corbett, D. *et al.* (2021) 'Spatiotemporal analysis of human intestinal development

at single-cell resolution', *Cell*, pp. 1–17. doi: 10.1016/j.cell.2020.12.016.

Ferguson, A., Parrott, M. V and Connor, O. (1972) 'GROWTH A N D DEVELOPMENT OF " ANTIGEN-FREE " GRAFTS O F FOETAL MOUSE INTESTINE *', *Journal of Pathology*, 106(7103).

Furuya, S. and Furuya, K. (2007) 'Subepithelial Fibroblasts in Intestinal Villi: Roles in Intercellular Communication', *International Review of Cytology*, 264(07), pp. 165–223. doi: 10.1016/S0074-7696(07)64004-2.

Garcia, A. D. R. *et al.* (2010) 'Sonic Hedgehog regulates discrete populations of astrocytes in the adult mouse forebrain', *Journal of Neuroscience*, 30(41), pp. 13597–13608. doi: 10.1523/JNEUROSCI.0830-10.2010.

Gregorieff, A. *et al.* (2005) 'Expression pattern of Wnt signaling components in the adult intestine', *Gastroenterology*, 129(2), pp. 626–638. doi: 10.1016/j.gastro.2005.06.007.

Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, M. S. and Venables, B. (2020) 'gplots: Various R Programming Tools for Plotting Data.', *CRAN*. CRAN. Available at: https://cran.r-project.org/package=gplots.

Heinzel, K., Benz, C. and Bleul, C. C. (2007) 'A silent chemokine receptor regulates steady-state leukocyte homing in vivo', *Proceedings of the National Academy of Sciences of the United States of America*, 104(20), pp. 8421–8426. doi: 10.1073/pnas.0608274104.

Holloway, E. M. *et al.* (2021) 'Mapping Development of the Human Intestinal Niche at Single-Cell Resolution', *Cell Stem Cell*. Elsevier Inc., 28(3), pp. 568-580.e4. doi: 10.1016/j.stem.2020.11.008.

Hong, S. P. *et al.* (2020) 'Distinct fibroblast subsets regulate lacteal integrity through YAP/TAZ-induced VEGF-C in intestinal villi', *Nature Communications*. Springer US, 11(1), p. 4102. doi: 10.1038/s41467-020-17886-y.

Huang, R. *et al.* (2019) 'The NCATS BioPlanet – An integrated platform for exploring the universe of cellular signaling pathways for toxicology, systems biology, and chemical genomics', *Frontiers in Pharmacology*, 10(APR), pp. 1–13. doi: 10.3389/fphar.2019.00445.

Jaensson-Gyllenbäck, E. *et al.* (2011) 'Bile retinoids imprint intestinal CD103+ dendritic cells with the ability to generate gut-tropic T cells.', *Mucosal immunology*, 4(4), pp. 438–47. doi: 10.1038/mi.2010.91.

Kanamori-Katayama, M. *et al.* (2011) 'LRRN4 and UPK3B are markers of primary mesothelial cells', *PLoS ONE*, 6(10), pp. 2–9. doi: 10.1371/journal.pone.0025391.

Karpus, O. N. *et al.* (2019) 'Colonic CD90+ Crypt Fibroblasts Secrete Semaphorins to

Support Epithelial Growth', *Cell Reports*. ElsevierCompany., 26(13), pp. 3698-3708.e5. doi: 10.1016/j.celrep.2019.02.101.

Kinchen, J. *et al.* (2018) 'Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease', *Cell*. Elsevier Inc., 175(2), pp. 372-386.e17. doi: 10.1016/j.cell.2018.08.067.

Koopmans, T. and Rinkevich, Y. (2018) 'Mesothelial to mesenchyme transition as a major developmental and pathological player in trunk organs and their cavities', *Communications Biology*. Springer US, 1(1), p. 170. doi: 10.1038/s42003-018-0180-x.

Kosinski, C. *et al.* (2007) 'Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors', *Proceedings of the National Academy of Sciences*, 104(39), pp. 15418–15423. doi: 10.1073/pnas.0707210104.

Kramann, R. *et al.* (2015) 'Perivascular Gli1+ progenitors are key contributors to injury-induced organ fibrosis', *Cell Stem Cell*. Elsevier Inc., 16(1), pp. 51–66. doi: 10.1016/j.stem.2014.11.004.

Kramann, R. *et al.* (2016) 'Adventitial MSC-like Cells Are Progenitors of Vascular Smooth Muscle Cells and Drive Vascular Calcification in Chronic Kidney Disease', *Cell Stem Cell*. Elsevier Inc., 19(5), pp. 628–642. doi: 10.1016/j.stem.2016.08.001.

Kuleshov, M. V. *et al.* (2016) 'Enrichr: a comprehensive gene set enrichment analysis web server 2016 update', *Nucleic acids research*, 44(W1), pp. W90–W97. doi: 10.1093/nar/gkw377.

Kurahashi, M. *et al.* (2013) 'A novel population of subepithelial platelet-derived growth factor receptor α-positive cells in the mouse and human colon.', *American journal of physiology. Gastrointestinal and liver physiology*, 304(9), pp. G823-34. doi: 10.1152/ajpgi.00001.2013.

Lee, M. Y. *et al.* (2017) 'Transcriptome of interstitial cells of Cajal reveals unique and selective gene signatures', *PLoS ONE*, 12(4). doi: 10.1371/journal.pone.0176031.

Li, Jun; Barron, M. (2017) 'ccRemover: Removes the Cell-Cycle Effect from Single-Cell RNA-Sequencing Data', *CRAN*.

Love, M. I., Huber, W. and Anders, S. (2014) 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), pp. 1–21. doi: 10.1186/s13059-014-0550-8.

La Manno, G. *et al.* (2018) 'RNA velocity of single cells', *Nature*. Springer US, 560(7719), pp. 494–498. doi: 10.1038/s41586-018-0414-6.

McCarthy, N. *et al.* (2020) 'Distinct Mesenchymal Cell Populations Generate the Essential

Intestinal BMP Signaling Gradient', *Cell Stem Cell*. Elsevier Inc., 26(3), pp. 391-402.e5. doi: 10.1016/j.stem.2020.01.008.

Merrick, D. *et al.* (2019) 'Identification of a mesenchymal progenitor cell hierarchy in adipose tissue', *Science*, 364(6438), p. eaav2501. doi: 10.1126/science.aav2501.

Meyerholz, D. K., Lambertz, A. M. and McCray, P. B. (2016) 'Dipeptidyl Peptidase 4 Distribution in the Human Respiratory Tract Implications for the Middle East Respiratory Syndrome', *American Journal of Pathology*. American Society for Investigative Pathology, 186(1), pp. 78–86. doi: 10.1016/j.ajpath.2015.09.014.

Miyoshi, H. *et al.* (2012) 'Wnt5a Potentiates TGF- b Signaling to promote colonic crypt regeneration after tissue injury', *Science*, 338(October), pp. 108–113.

Mosley, L. and Klein, J. R. (1992) 'Peripheral engraftment of fetal intestine into athymic mice sponsors T cell development: direct evidence for thymopoitic function of murine small intestine', *J Exp Med*, 176(November), pp. 1365–1373.

Mowat, A. M. and Agace, W. W. (2014) 'Regional specialization within the intestinal immune system', *Nature Reviews Immunology*, 14, pp. 667–685. doi: 10.1038/nri3738.

Namvar, S. *et al.* (2018) 'Functional molecules in mesothelial-to-mesenchymal transition revealed by transcriptome analyses', *Journal of Pathology*, 245(4), pp. 491–501. doi: 10.1002/path.5101.

Parikh, K. *et al.* (2019) 'Colonic epithelial cell diversity in health and inflammatory bowel disease', *Nature*, p. 1. doi: 10.1038/s41586-019-0992-y.

Powell, D. W. *et al.* (2011) 'Mesenchymal Cells of the Intestinal Lamina Propria', *Annual Review of Physiology*, 73(1), pp. 213–237. doi: 10.1146/annurev.physiol.70.113006.100646.

Probstmeier, R., Martini, R. and Schachner, M. (1990) 'Expression of J1/tenascin in the crypt-villus unit of adult mouse small intestine: Implications for its role in epithelial cell shedding', *Development*, 109(2), pp. 313–321.

Qi, Z. *et al.* (2017) 'BMP restricts stemness of intestinal Lgr5+ stem cells by directly suppressing their signature genes', *Nature Communications*. Nature Publishing Group, 8(1), p. 13824. doi: 10.1038/ncomms13824.

R Core Team (2020) 'R: A Language and Environment for Statistical Computing.', *R Foundation for Statistical Computing*. Vienna, Austria.

Rinkevich, Y. *et al.* (2012) 'Identification and prospective isolation of a mesothelial precursor lineage giving rise to smooth muscle cells and fibroblasts for mammalian internal organs, and their vasculature', *Nature Cell Biology*. Nature Publishing Group, 14(12), pp. 1251–1260. doi: 10.1038/ncb2610.

Roulis, M. *et al.* (2020) 'Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche', *Nature*. Springer US, 580(7804), pp. 524–529. doi: 10.1038/s41586-020-2166-3.

Roulis, M. and Flavell, R. A. (2016) 'Fibroblasts and myofibroblasts of the intestinal lamina propria in physiology and disease', *Differentiation*. Elsevier, 92(3), pp. 116–131. doi: 10.1016/j.diff.2016.05.002.

Sato, T. *et al.* (2011) 'Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts', *Nature*. Nature Publishing Group, 469(7330), pp. 415–418. doi: 10.1038/nature09637.

Schulz, O. *et al.* (2009) 'Intestinal CD103+, but not CX3CR1+, antigen sampling cells migrate in lymph and serve classical dendritic cell functions', *Journal of Experimental Medicine*, 206(13), pp. 3101–3114. doi: 10.1084/jem.20091925.

Shoshkes-Carmel, M. *et al.* (2018) 'Subepithelial telocytes are an important source of Wnts that supports intestinal crypts', *Nature*, 557(7704), pp. 242–246. doi: 10.1038/s41586-018-0084-4.

Sidney, L. E. *et al.* (2014) 'Concise Review: Evidence for CD34 as a Common Marker for Diverse Progenitors', *STEM CELLS*, 32(6), pp. 1380–1389. doi: 10.1002/stem.1661.

Sitnik, K. M. *et al.* (2016) 'Context-Dependent Development of Lymphoid Stroma from Adult CD34+ Adventitial Progenitors', *Cell Reports*, 14(10), pp. 2375–2388. doi: 10.1016/j.celrep.2016.02.033.

Smillie, C. S. *et al.* (2019) 'Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis', *Cell*, 178(3), pp. 714-730.e22. doi: 10.1016/j.cell.2019.06.029.

Srinivas, S. *et al.* (2001) 'Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus', *BMC Dev Biol*, 1, p. 4. doi: 10.1186/1471-213X-1-4.

Stuart, T. and Satija, R. (2019) 'Integrative single-cell analysis', *Nature Reviews Genetics*. Springer US, p. 1. doi: 10.1038/s41576-019-0093-7.

Stzepourginski, I. *et al.* (2017) 'CD34 + mesenchymal cells are a major component of the intestinal stem cells niche at homeostasis and after injury', *Proceedings of the National Academy of Sciences*, 114(4), pp. E506–E513. doi: 10.1073/pnas.1620059114.

Taylor, R. T. *et al.* (2007) 'Lymphotoxin-Independent Expression of TNF-Related Activation-Induced Cytokine by Stromal Cells in Cryptopatches, Isolated Lymphoid Follicles, and Peyer's Patches', *The Journal of Immunology*, 178(9), pp. 5659–5667. doi: 10.4049/jimmunol.178.9.5659.

Thomson, C. A. *et al.* (2018) 'Expression of the Atypical Chemokine Receptor ACKR4

Identifies a Novel Population of Intestinal Submucosal Fibroblasts That Preferentially Expresses Endothelial Cell Regulators', *The Journal of Immunology*, 201(1), pp. 215–229. doi: 10.4049/jimmunol.1700967.

Vallecillo-García, P. *et al.* (2017) 'Odd skipped-related 1 identifies a population of embryonic fibro-adipogenic progenitors regulating myogenesis during limb development', *Nature Communications*. Springer US, 8(1). doi: 10.1038/s41467-017-01120-3.

Vicente-Suarez, I. *et al.* (2015) 'Unique lamina propria stromal cells imprint the functional phenotype of mucosal dendritic cells', *Mucosal Immunology*. Nature Publishing Group, 8(1), pp. 141–151. doi: 10.1038/mi.2014.51.

West, N. R. *et al.* (2017) 'Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor–neutralizing therapy in patients with inflammatory bowel disease', *Nature Medicine*, 23(5), pp. 579–589. doi: 10.1038/nm.4307.

Wilm, B. *et al.* (2005) 'The serosal mesothelium is a major source of smooth muscle cells of the gut vasculature', *Development*, 132(23), pp. 5317–5328. doi: 10.1242/dev.02141.

Winters, N. I. and Bader, D. M. (2013) 'Development of the serosal mesothelium', *Journal of Developmental Biology*, 1(2), pp. 64–81. doi: 10.3390/jdb1020064.

Worthley, D. L. *et al.* (2015) 'Gremlin 1 identifies a skeletal stem cell with bone, cartilage, and reticular stromal potential', *Cell*. Elsevier Inc., 160(1–2), pp. 269–284. doi: 10.1016/j.cell.2014.11.042.

Wu, N. *et al.* (2021) 'MAP3K2-regulated intestinal stromal cells define a distinct stem cell niche', *Nature*. Springer US. doi: 10.1038/s41586-021-03283-y.

Xiang, M. *et al.* (2020) 'A Single-Cell Transcriptional Roadmap of the Mouse and Human Lymph Node Lymphatic Vasculature', *Frontiers in Cardiovascular Medicine*, 7(April). doi: 10.3389/fcvm.2020.00052.

Xie, Z. *et al.* (2021) 'Gene Set Knowledge Discovery with Enrichr', *Current Protocols*, 1(3). doi: 10.1002/cpz1.90.

Yanai, H. *et al.* (2017) 'Intestinal stem cells contribute to the maturation of the neonatal small intestine and colon independently of digestive activity', *Scientific Reports*. Springer US, 7(1), pp. 2–5. doi: 10.1038/s41598-017-09927-2.

Yuasa, Y. (2003) 'Control of gut differentiation and intestinal-type gastric carcinogenesis', *Nature Reviews Cancer*, 3(8), pp. 592–600. doi: 10.1038/nrc1141.

Zheng, G. X. Y. *et al.* (2017) 'Massively parallel digital transcriptional profiling of single cells', *Nature Communications*. Nature Publishing Group, 8. doi: 10.1038/ncomms14049.

Figure 1

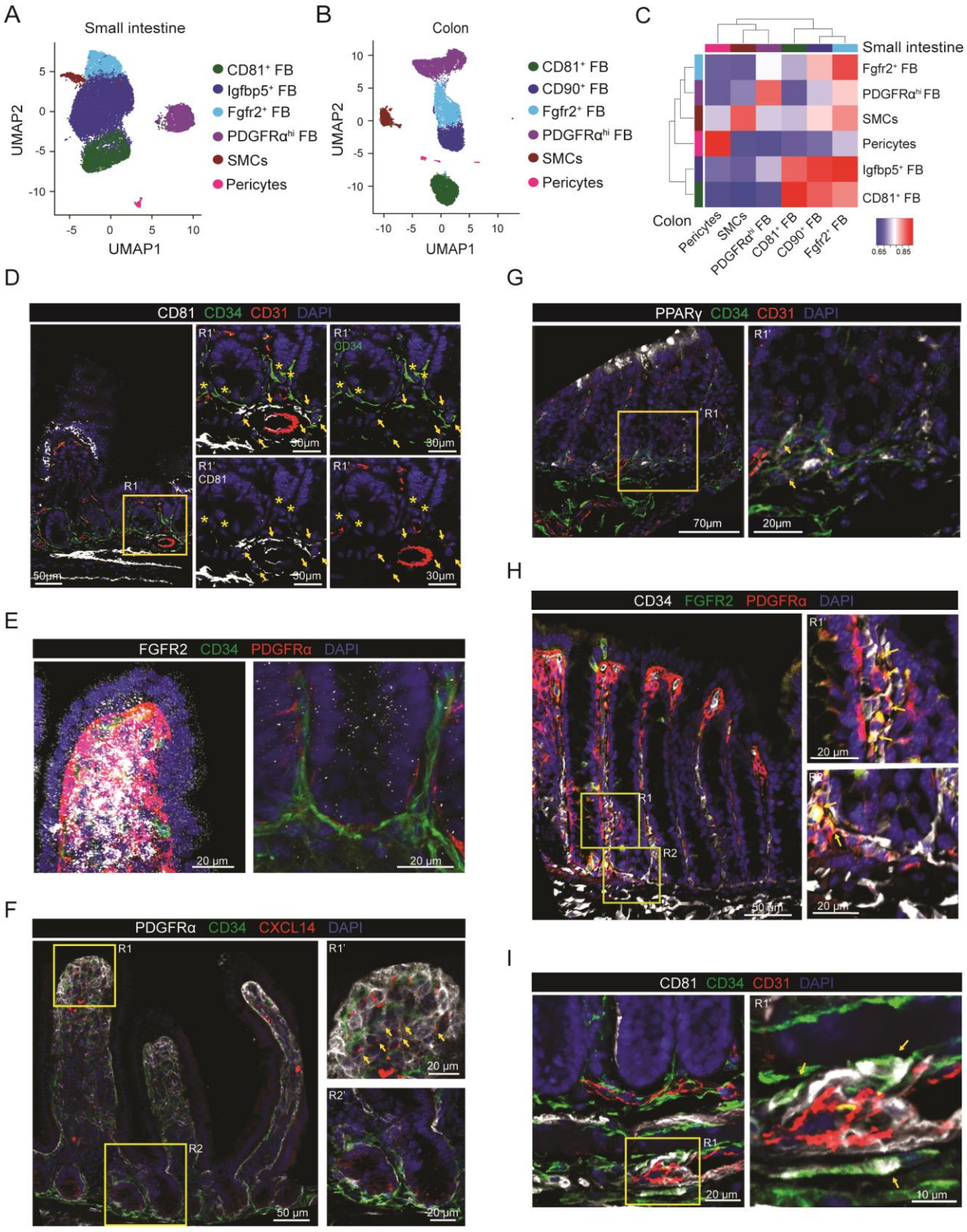

A
Small intestine
UMAP2
UMAP1
CD81⁺ FB
Igfbp5⁺ FB
Fgfr2⁺ FB
PDGFRα^hi FB
SMCs
Pericytes
B
Colon
UMAP2
UMAP1
CD81⁺ FB
CD90⁺ FB
Fgfr2⁺ FB
PDGFRα^hi FB
SMCs
Pericytes
C
Small intestine
Fgfr2⁺ FB
PDGFRα^hi FB
SMCs
Pericytes
Igfbp5⁺ FB
CD81⁺ FB
Colon
Pericytes
SMCs
PDGFRα^hi FB
CD81⁺ FB
CD90⁺ FB
Fgfr2⁺ FB
0.65
0.85
D
CD81 CD34 CD31 DAPI
R1
R1'
CD34
CD81
50μm
30μm
E
FGFR2 CD34 PDGFRα DAPI
20 μm
F
PDGFRα CD34 CXCL14 DAPI
R1
R2
R1'
R2'
50 μm
20 μm
G
PPARγ CD34 CD31 DAPI
R1
R1'
70μm
20μm
H
CD34 FGFR2 PDGFRα DAPI
R1
R2
R1'
R2'
50 μm
20 μm
I
CD81 CD34 CD31 DAPI
R1
R1'
20 μm
10 μm

Figure 2

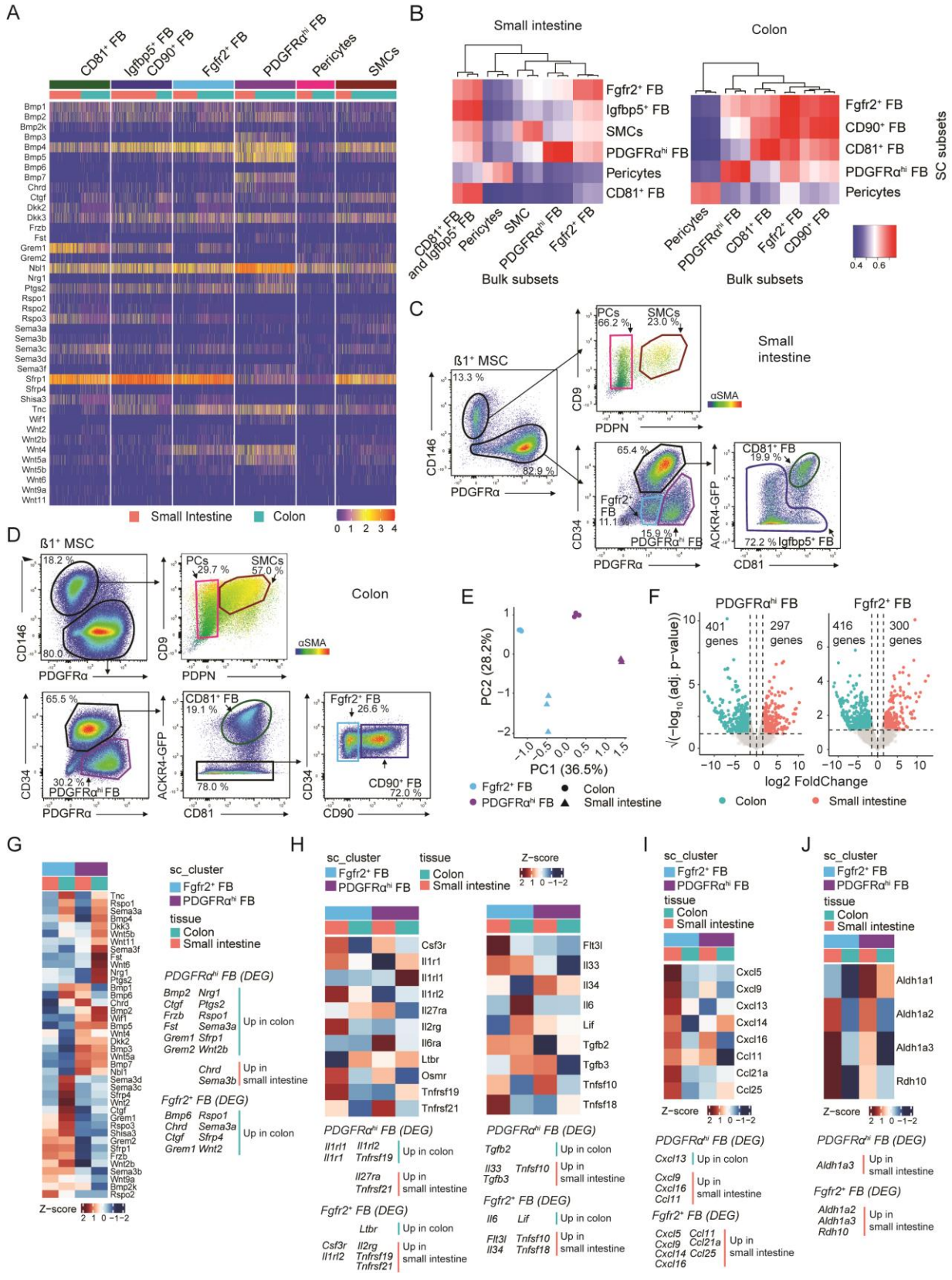

A
CD81+ FB
Igfbp5+ FB
CD90+ FB
Fgfr2+ FB
PDGFRαhi FB
Pericytes
SMCs
Small Intestine
Colon
B
Small intestine
Colon
Fgfr2+ FB
Igfbp5+ FB
SMCs
PDGFRαhi FB
Pericytes
CD81+ FB
CD90+ FB
SC subsets
Bulk subsets
C
ß1+ MSC
13.3 %
82.9 %
CD146
PDGFRα
PCs
66.2 %
SMCs
23.0 %
CD9
PDPN
αSMA
Small intestine
65.4 %
Fgfr2+ FB
11.1 %
15.9 %
PDGFRαhi FB
CD34
CD81+ FB
19.9 %
72.2 %
Igfbp5+ FB
ACKR4-GFP
CD81
D
ß1+ MSC
18.2 %
80.0
PCs
29.7 %
SMCs
57.0 %
Colon
65.5 %
30.2 %
PDGFRαhi FB
CD81+ FB
19.1 %
78.0 %
Fgfr2+ FB
26.6 %
CD90+ FB
72.0 %
CD90
E
PC2 (28.2%)
PC1 (36.5%)
Fgfr2+ FB
PDGFRαhi FB
Colon
Small intestine
F
PDGFRαhi FB
Fgfr2+ FB
401 genes
297 genes
416 genes
300 genes
√(−log10 (adj. p−value))
log2 FoldChange
Colon
Small intestine
G
sc_cluster
Fgfr2+ FB
PDGFRαhi FB
tissue
Colon
Small intestine
Z−score
PDGFRαhi FB (DEG)
Bmp2 Nrg1
Ctgf Ptgs2
Frzb Rspo1
Fst Sema3a
Grem1 Sfrp1
Grem2 Wnt2b
Up in colon
Chrd Sema3b
Up in small intestine
Fgfr2+ FB (DEG)
Bmp6 Rspo1
Chrd Sema3a
Ctgf Sfrp4
Grem1 Wnt2
Up in colon
H
sc_cluster
Fgfr2+ FB
PDGFRαhi FB
tissue
Colon
Small intestine
Z−score
Csf3r
Il1r1
Il1rl1
Il1rl2
Il27ra
Il2rg
Il6ra
Ltbr
Osmr
Tnfrsf19
Tnfrsf21
Flt3l
Il33
Il34
Il6
Lif
Tgfb2
Tgfb3
Tnfsf10
Tnfsf18
PDGFRαhi FB (DEG)
Il1rl1 Il1rl2
Il1r1 Tnfrsf19
Up in colon
Il27ra
Tnfrsf21
Up in small intestine
Fgfr2+ FB (DEG)
Ltbr
Up in colon
Csf3r Il2rg
Il1rl2 Tnfrsf19
Tnfrsf21
Up in small intestine
PDGFRαhi FB (DEG)
Tgfb2
Up in colon
Il33 Tnfsf10
Tgfb3
Up in small intestine
Fgfr2+ FB (DEG)
Il6 Lif
Up in colon
Flt3l Tnfsf10
Il34 Tnfsf18
Up in small intestine
I
sc_cluster
Fgfr2+ FB
PDGFRαhi FB
tissue
Colon
Small intestine
Cxcl5
Cxcl9
Cxcl13
Cxcl14
Cxcl16
Ccl11
Ccl21a
Ccl25
Z−score
PDGFRαhi FB (DEG)
Cxcl13
Up in colon
Cxcl9
Cxcl16
Ccl11
Up in small intestine
Fgfr2+ FB (DEG)
Cxcl5 Ccl11
Cxcl9 Ccl21a
Cxcl14 Ccl25
Cxcl16
Up in small intestine
J
sc_cluster
Fgfr2+ FB
PDGFRαhi FB
tissue
Colon
Small intestine
Aldh1a1
Aldh1a2
Aldh1a3
Rdh10
Z−score
PDGFRαhi FB (DEG)
Aldh1a3
Up in small intestine
Fgfr2+ FB (DEG)
Aldh1a2
Aldh1a3
Rdh10
Up in small intestine

Figure 3

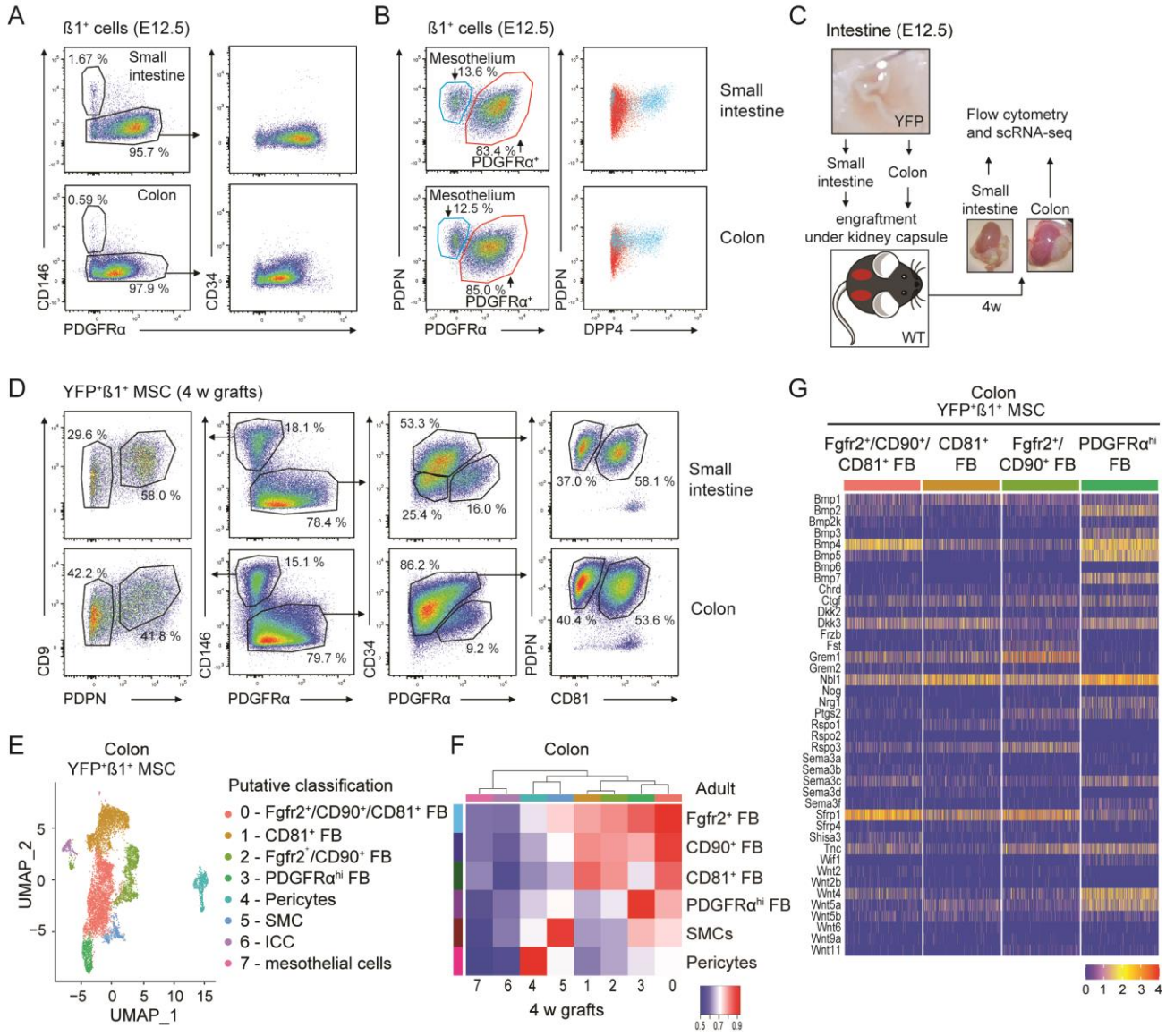

A
ß1⁺ cells (E12.5)
1.67 %
Small intestine
95.7 %
0.59 %
Colon
97.9 %
CD146
PDGFRα
CD34
B
ß1⁺ cells (E12.5)
Mesothelium
13.6 %
83.4 %
PDGFRα⁺
Small intestine
Mesothelium
12.5 %
85.0 %
PDGFRα⁺
Colon
PDPN
PDGFRα
PDPN
DPP4
C
Intestine (E12.5)
YFP
Small intestine
Colon
engraftment under kidney capsule
WT
4w
Flow cytometry and scRNA-seq
Small intestine
Colon
D
YFP⁺ß1⁺ MSC (4 w grafts)
29.6 %
58.0 %
42.2 %
41.8 %
CD9
PDPN
18.1 %
78.4 %
15.1 %
79.7 %
CD146
PDGFRα
53.3 %
25.4 %
16.0 %
86.2 %
9.2 %
CD34
PDGFRα
37.0 %
58.1 %
40.4 %
53.6 %
PDPN
CD81
Small intestine
Colon
E
Colon
YFP⁺ß1⁺ MSC
UMAP_2
UMAP_1
Putative classification
0 - Fgfr2⁺/CD90⁺/CD81⁺ FB
1 - CD81⁺ FB
2 - Fgfr2⁻/CD90⁺ FB
3 - PDGFRα^hi FB
4 - Pericytes
5 - SMC
6 - ICC
7 - mesothelial cells
F
Colon
Adult
Fgfr2⁺ FB
CD90⁺ FB
CD81⁺ FB
PDGFRα^hi FB
SMCs
Pericytes
7 6 4 5 1 2 3 0
4 w grafts
0.5 0.7 0.9
G
Colon
YFP⁺ß1⁺ MSC
Fgfr2⁺/CD90⁺/ CD81⁺ FB
CD81⁺ FB
Fgfr2⁺/ CD90⁺ FB
PDGFRα^hi FB
Bmp1
Bmp2
Bmp2k
Bmp3
Bmp4
Bmp5
Bmp6
Bmp7
Chrd
Ctgf
Dkk2
Dkk3
Frzb
Fst
Grem1
Grem2
Nbl1
Nog
Nrg1
Ptgs2
Rspo1
Rspo2
Rspo3
Sema3a
Sema3b
Sema3c
Sema3d
Sema3f
Sfrp1
Sfrp4
Shisa3
Tnc
Wif1
Wnt2
Wnt2b
Wnt4
Wnt5a
Wnt5b
Wnt6
Wnt9a
Wnt11
0 1 2 3 4

Figure 4

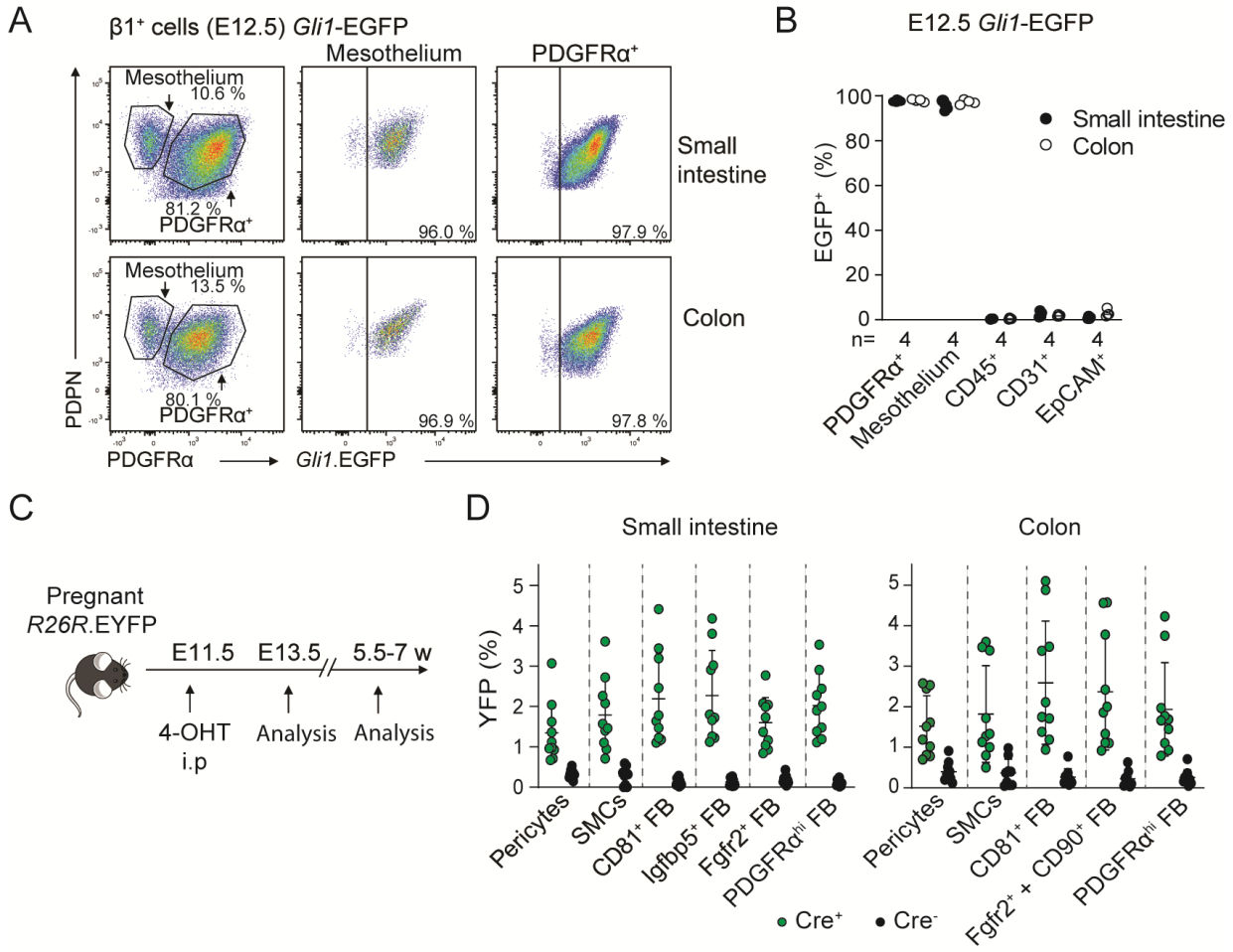

A
β1⁺ cells (E12.5) Gli1-EGFP
Mesothelium
PDGFRα⁺
Small intestine
Colon
Mesothelium 10.6 %
81.2 % PDGFRα⁺
96.0 %
97.9 %
Mesothelium 13.5 %
80.1 % PDGFRα⁺
96.9 %
97.8 %
PDPN
PDGFRα
Gli1.EGFP
B
E12.5 Gli1-EGFP
Small intestine
Colon
EGFP⁺ (%)
n= 4 4 4 4 4
PDGFRα⁺ Mesothelium
CD45⁺
CD31⁺
EpCAM⁺
C
Pregnant R26R.EYFP
E11.5
E13.5
5.5-7 w
4-OHT i.p
Analysis
Analysis
D
Small intestine
Colon
YFP (%)
Pericytes
SMCs
CD81⁺ FB
Igfbp5⁺ FB
Fgfr2⁺ FB
PDGFRα^hi FB
Fgfr2⁺ + CD90⁺ FB
Cre⁺
Cre⁻

Figure 5

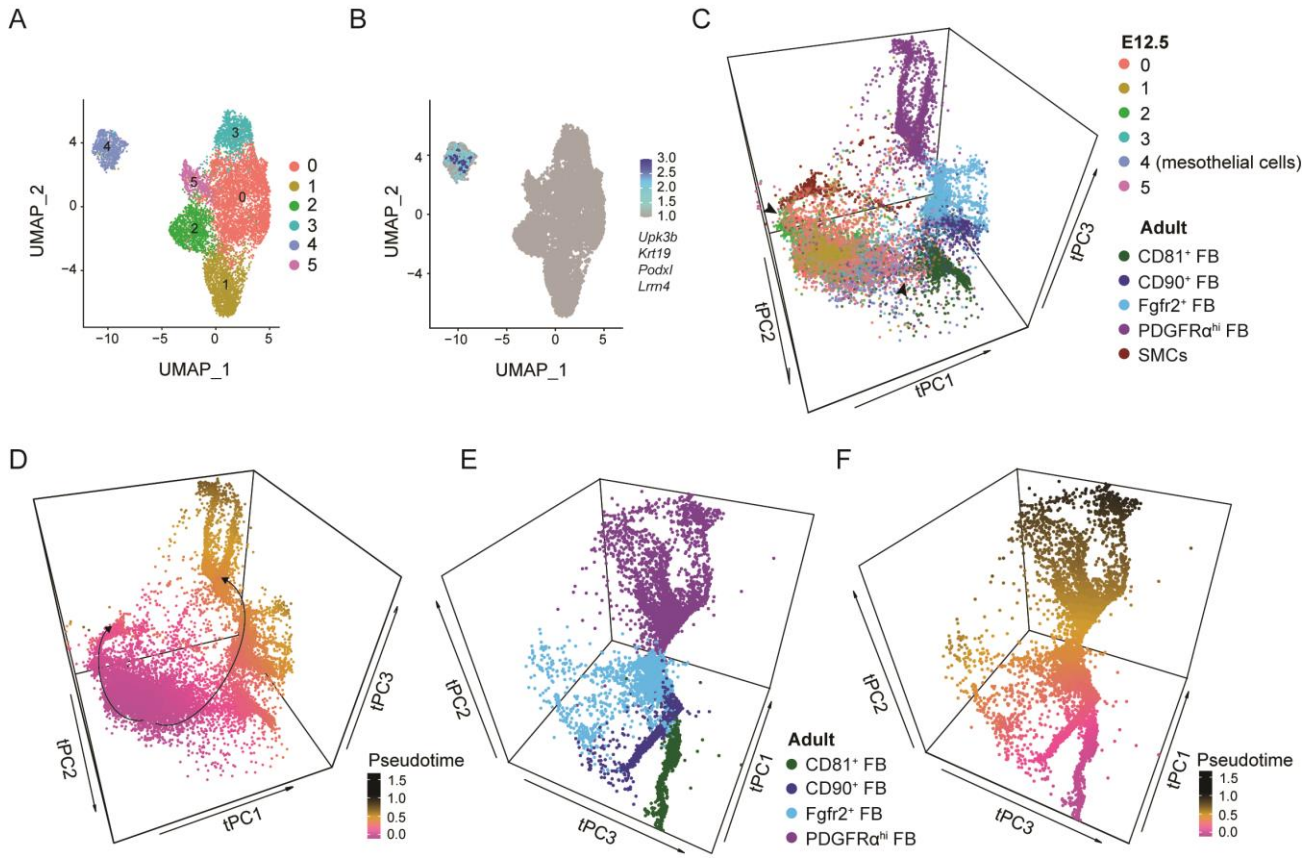
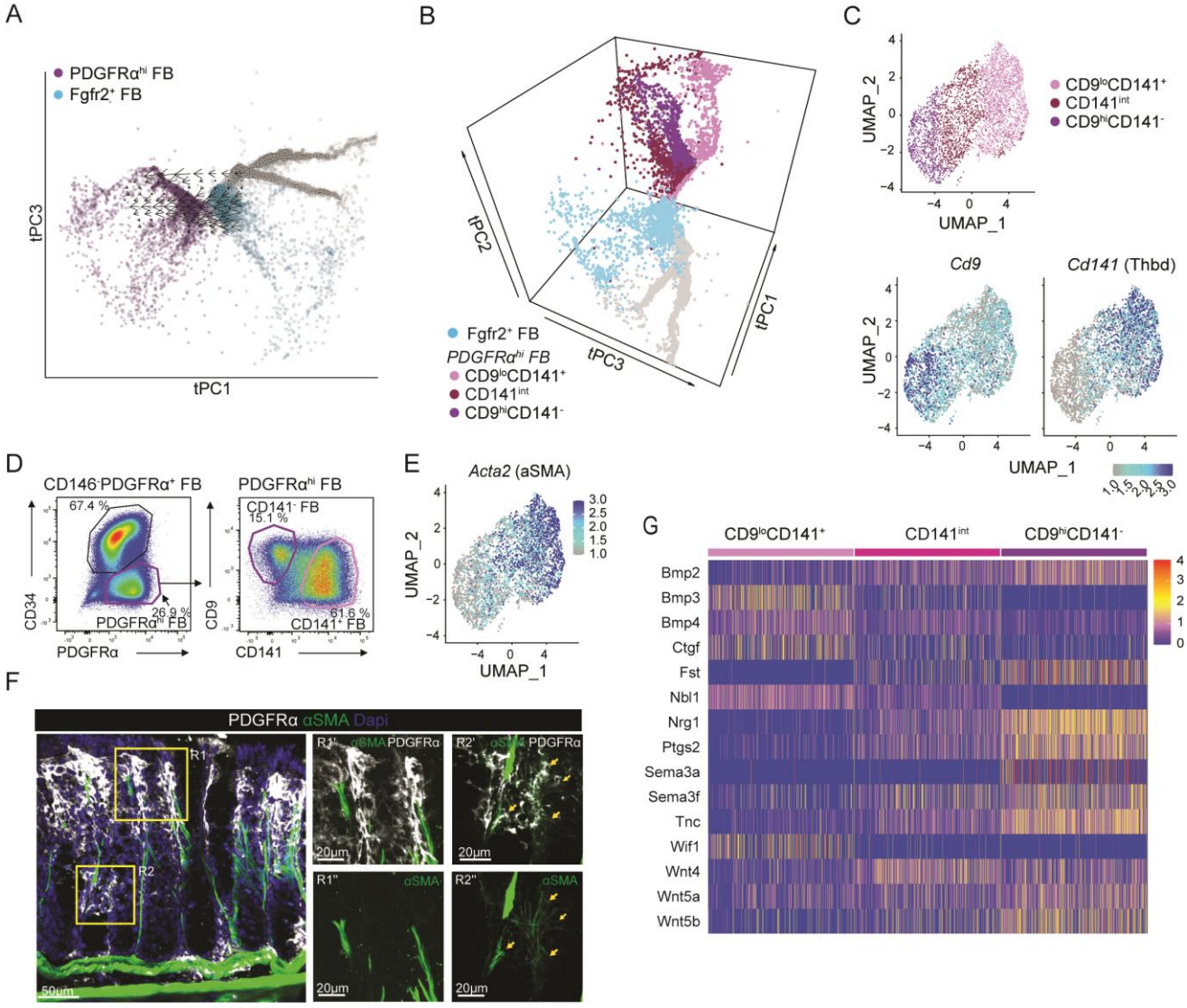

A
B
C
UMAP_2
UMAP_1
0
1
2
3
4
5
Upk3b
Krt19
Podxl
Lrrn4
E12.5
4 (mesothelial cells)
Adult
CD81⁺ FB
CD90⁺ FB
Fgfr2⁺ FB
PDGFRα^hi FB
SMCs
tPC1
tPC2
tPC3
D
E
F
Pseudotime

Figure 6


A
PDGFRα^hi FB
Fgfr2+ FB
tPC3
tPC1
B
tPC2
tPC3
tPC1
Fgfr2+ FB
PDGFRα^hi FB
CD9^loCD141+
CD141^int
CD9^hiCD141-
C
UMAP_2
UMAP_1
CD9^loCD141+
CD141^int
CD9^hiCD141-
Cd9
Cd141 (Thbd)
D
CD146-PDGFRα+ FB
PDGFRα^hi FB
67.4 %
26.9 %
CD34
PDGFRα
CD141- FB
15.1 %
61.6 %
CD9
CD141
E
Acta2 (aSMA)
UMAP_2
UMAP_1
F
PDGFRα αSMA Dapi
R1
R2
R1' αSMA PDGFRα
R2' αSMA PDGFRα
R1'' αSMA
R2'' αSMA
20μm
50μm
G
CD9^loCD141+
CD141^int
CD9^hiCD141-
Bmp2
Bmp3
Bmp4
Ctgf
Fst
Nbl1
Nrg1
Ptgs2
Sema3a
Sema3f
Tnc
Wif1
Wnt4
Wnt5a
Wnt5b

Supplemental Information

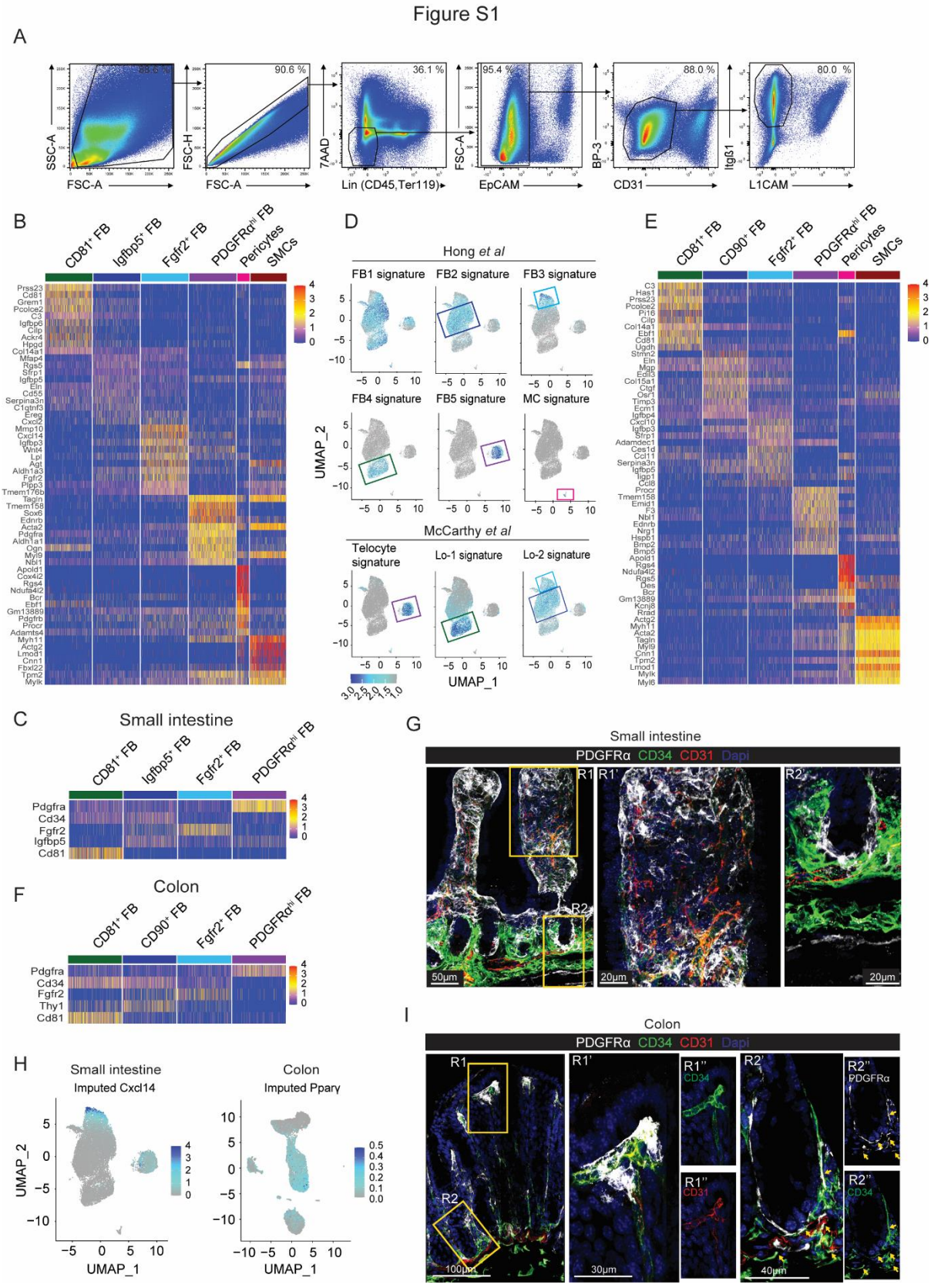
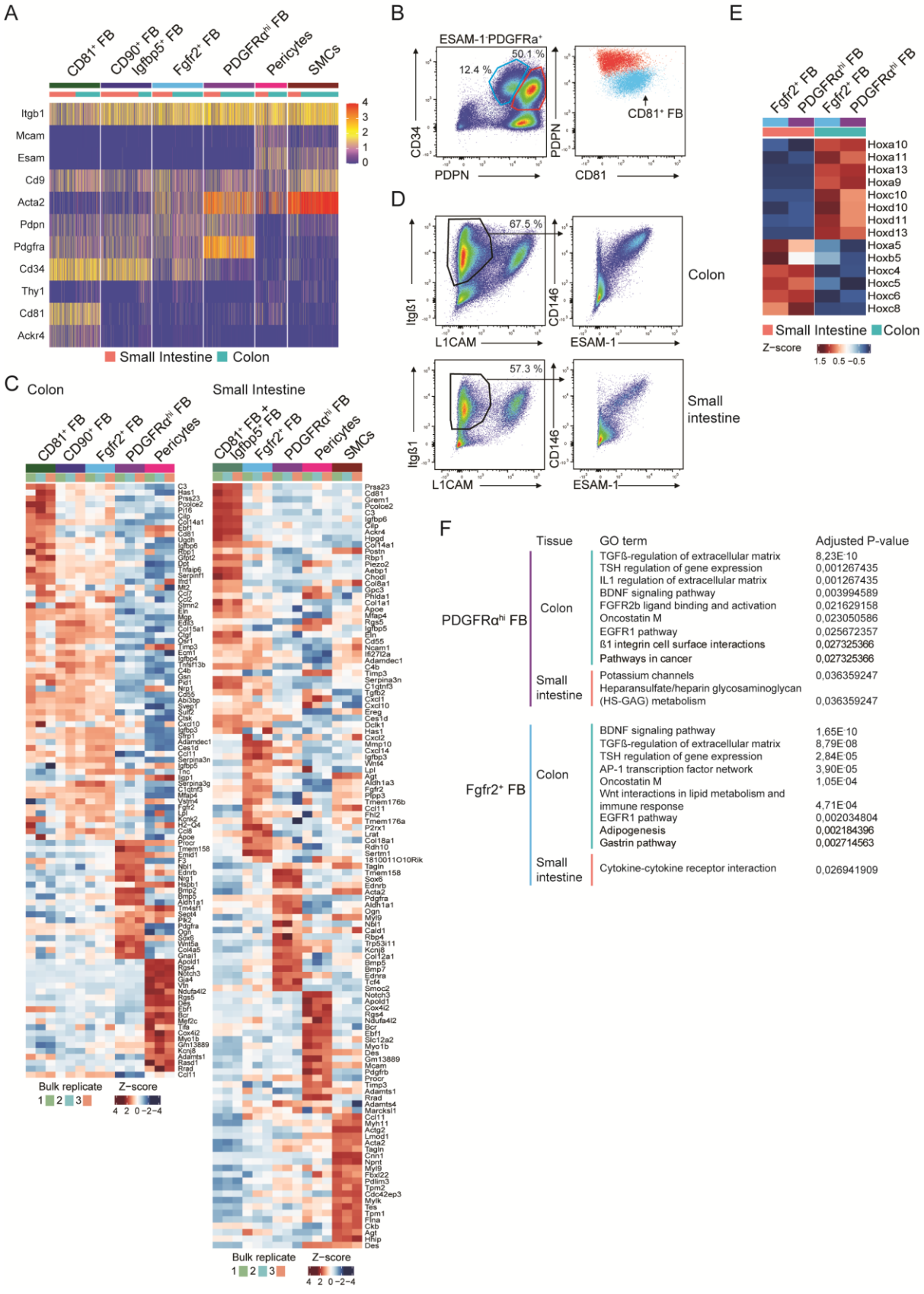


Figure S1.

Related to Figure 1. (**A**) Flow cytometric gating strategy for sorting adult small intestinal and colonic Itgβ1^{+} MSC. (**B**) Heatmap with scaled transcription levels (integrated data) of the top 10 differentially expressed genes (DEG) between adult small intestinal MSC subsets. (**C**) Heatmap with scaled transcription levels (anchor integrated data) of markers used to distinguish small intestinal FB subsets. (**D**) Signature genes for MSC subsets identified in Hong *et al* (Hong *et al.,* 2020) and McCarthy *et al* (McCarthy *et al.,* 2020) projected onto the small intestinal MSC UMAP as gene modules. Boxes represent the 6 MSC clusters and are color coded as in (**B**). (**E**) Heatmap with scaled transcription levels (integrated data) of the top 10 DEGs between adult colonic MSC subsets. (**F**) Heatmap with scaled transcription levels (anchor integrated data) of markers used to distinguish colonic FB subsets. (**G)** Immunohistochemical staining of mouse jejunum. R1' and R2' are high magnifications of the R1 and R2 quadrants (yellow squares) in the left image. (**H**) MAGIC imputed projection of (left) *Cxcl14* expression onto the small intestinal MSC UMAP and (right) *Pparγ* expression projected onto the colon MSC UMAP. (**I**) Immunohistochemical staining of mouse colon for indicated antigens. R1'' and R2'' depict single stains. Arrows indicate CD34^{+} FB (CD34^{+}PDGFRα^{+}CD31^{-} cells). (**H** and **I**) Results are representative stains from (**H**) 3 and (**I**) 2 experiments analyzing intestinal sections from 3 mice/experiment.

Figure S2



F

	Tissue	GO term	Adjusted P-value
PDGFRα^{hi} FB	Colon	TGFß-regulation of extracellular matrix	8,23E-10
		TSH regulation of gene expression	0,001267435
		IL1 regulation of extracellular matrix	0,001267435
		BDNF signaling pathway	0,003994589
		FGFR2b ligand binding and activation	0,021629158
		Oncostatin M	0,023050586
		EGFR1 pathway	0,025672357
		ß1 integrin cell surface interactions	0,027325366
		Pathways in cancer	0,027325366
	Small intestine	Potassium channels	0,036359247
		Heparansulfate/heparin glycosaminoglycan (HS-GAG) metabolism	0,036359247
Fgfr2^{+} FB	Colon	BDNF signaling pathway	1,65E-10
		TGFß-regulation of extracellular matrix	8,79E-08
		TSH regulation of gene expression	2,84E-05
		AP-1 transcription factor network	3,90E-05
		Oncostatin M	1,05E-04
		Wnt interactions in lipid metabolism and immune response	4,71E-04
		EGFR1 pathway	0,002034804
		Adipogenesis	0,002184396
		Gastrin pathway	0,002714563
	Small intestine	Cytokine-cytokine receptor interaction	0,026941909

Figure S2.

Related to Figure 2. (**A**) Heatmap showing scaled transcription levels (integrated data) of genes used as markers to distinguish the 6 MSC subsets by flow cytometry. (**B**) Flow cytometric analysis verifying that colonic PDPN^{lo}CD34^{+}FB (blue gate, left plot) express CD81 (blue cells right plot). Representative gating of at least 3 experiments with 4-7 mice/experiment. (**C**) Heatmaps showing transcription levels of the top 20 DEGs defining the scRNA-seq subsets within the bulk RNA-seq subsets from colon (left) and small intestine (right). (**D**) Flow cytometric gating strategy of colon (top panels) and small intestine (bottom panels) showing co-staining of ESAM-1 and CD146 by Itgβ1^{+} MSCs. Representative stains of at least 6 experiments with 1 mouse/experiment. (**E**) Heatmap of transcription levels (averaged between bulk RNA-seq triplicates) of Hox genes that were differentially expressed between small and large intestine in the indicated FB subset. (**F**) Gene ontology (GO) analysis using Enrichr (Bioplanet 2019) of DEGs between small intestinal and colonic PDGFRα^{hi} FB (purple) and between small intestinal and colonic Fgfr2^{+} FB (light blue). Top 9 significant pathways based on adjusted p values for colon and all significant pathways for small intestine are shown.

Figure S3

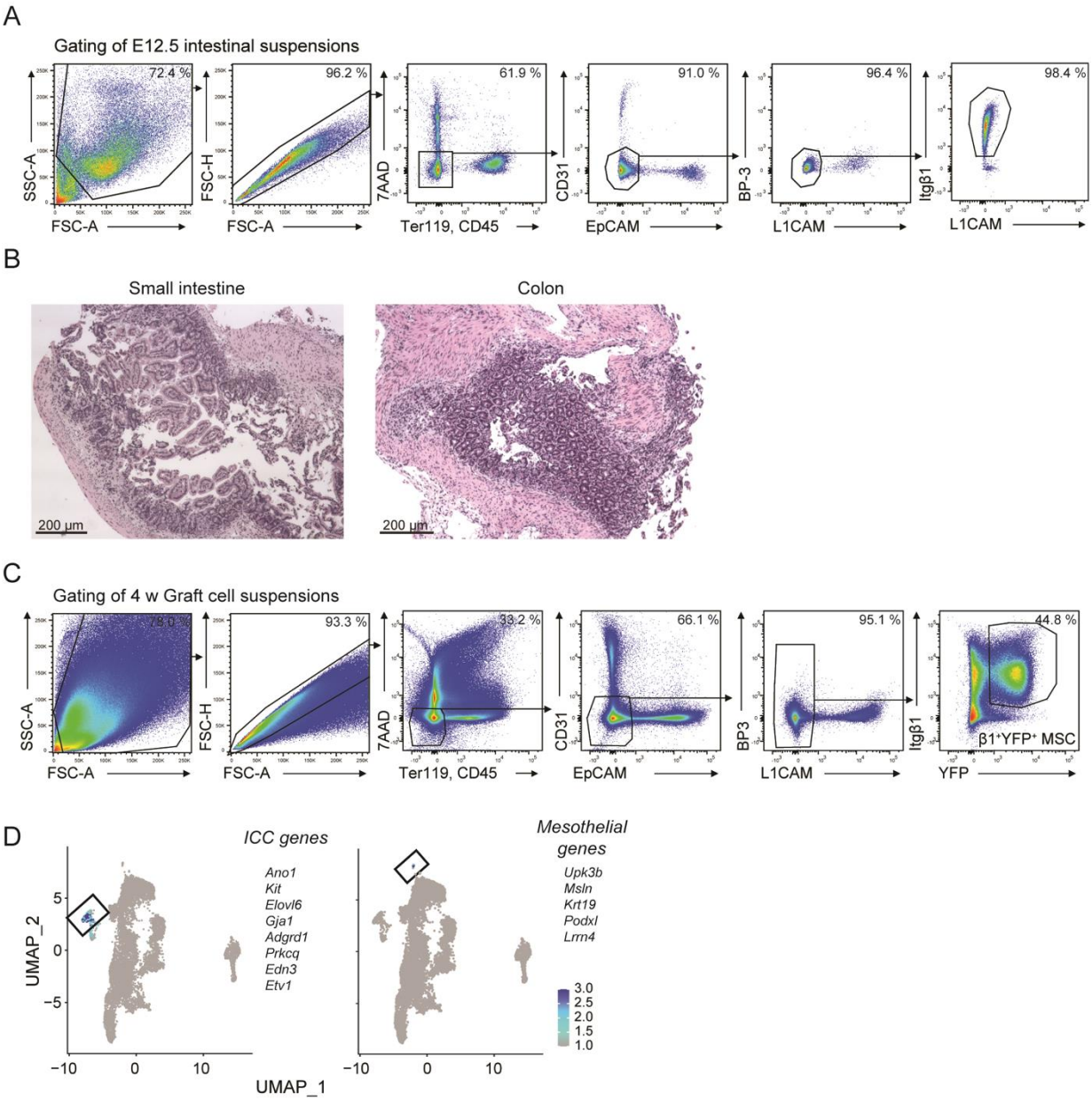


Figure S3.

Related to Figure 3. (**A**) Flow cytometric gating strategy for the identification and sorting of intestinal Itgβ1^{+} MSC from embryonic day (E)12.5 mice. (**B**) Hematoxylin and eosin staining of small intestinal and colonic grafts 6 weeks post transplantation. Results are representative stains of 2 grafts for each tissue. (**C**) Flow cytometric gating strategy for the identification and

sorting of YFP^{+}Itgβ1^{+} MSC from intestinal grafts. (**D**) Projections of module score of signature genes for Interstitial cells of Cajal (ICC) and mesothelium onto colon graft UMAP. Boxes identify the ICC and mesothelial clusters.

Figure S4

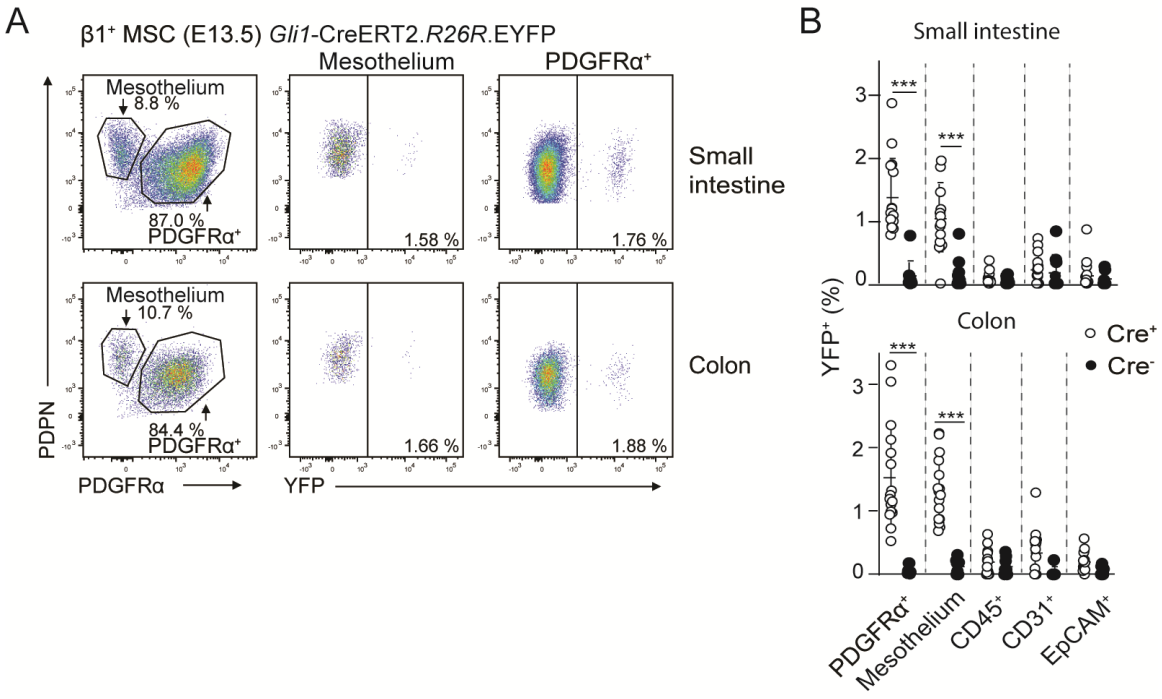


Figure S4.

Related to Figure 4. (**A**) Representative flow cytometric analysis and (**B**) pooled data of the proportions of YFP-expressing cells in indicated intestinal populations of E13.5 *Gli1*.CreERT2$^{+/-}$.*R26R*.EYFP mice 2 days after injection with 4-OHT. Pre-gating strategy as in Fig. S3A. Results are from 4 pooled experiments with 1-6 embryos/experiment. Bars, mean (SD). ***p<0.0001, 2-way ANOVA with Benjamini, Krieger and Yekutieli multiple comparisons.

Figure S5

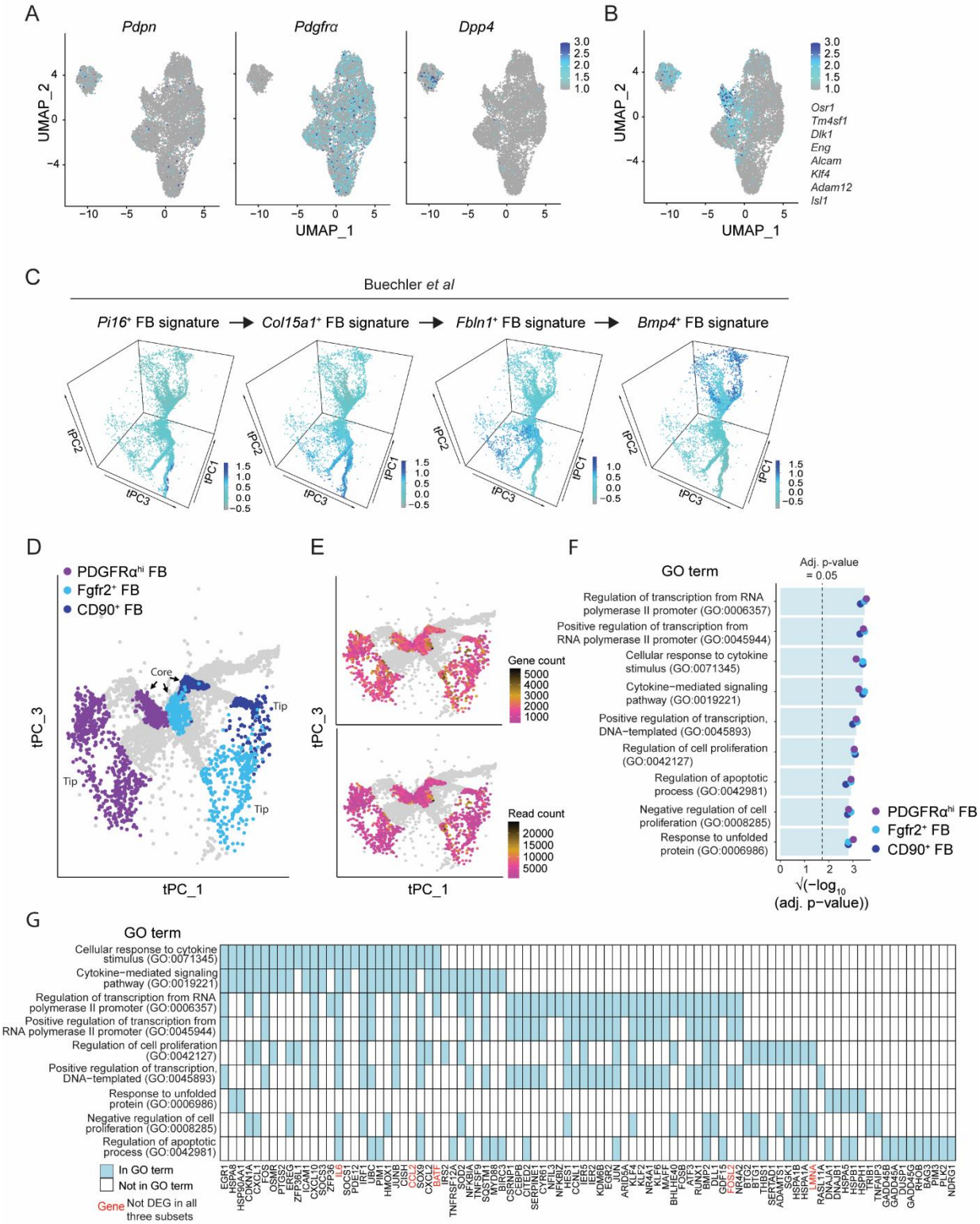

A
Pdpn
Pdgfrα
Dpp4
UMAP_2
UMAP_1
B
Osr1
Tm4sf1
Dlk1
Eng
Alcam
Klf4
Adam12
Isl1
C
Buechler et al
Pi16+ FB signature
Col15a1+ FB signature
Fbln1+ FB signature
Bmp4+ FB signature
tPC1
tPC2
tPC3
D
PDGFRα^hi FB
Fgfr2+ FB
CD90+ FB
Core
Tip
tPC_3
tPC_1
E
Gene count
Read count
F
GO term
Adj. p-value = 0.05
Regulation of transcription from RNA polymerase II promoter (GO:0006357)
Positive regulation of transcription from RNA polymerase II promoter (GO:0045944)
Cellular response to cytokine stimulus (GO:0071345)
Cytokine−mediated signaling pathway (GO:0019221)
Positive regulation of transcription, DNA−templated (GO:0045893)
Regulation of cell proliferation (GO:0042127)
Regulation of apoptotic process (GO:0042981)
Negative regulation of cell proliferation (GO:0008285)
Response to unfolded protein (GO:0006986)
√(−log10 (adj. p−value))
G
GO term
In GO term
Not in GO term
Gene Not DEG in all three subsets

Figure S5.

Related to Figure 5. (**A**) *Pdpn, Dpp4, Pdgfrα* integrated and normalized gene expression and (**B**) mesenchymal precursor associated gene expression module score projected onto UMAP of E12.5 large intestinal Itgβ1^{+} MSC. (**C**) Top 20 DEGs for indicated FB clusters identified in Buechler *et al* projected as gene modules onto colonic tSPACE projections of adult colonic MSC in tPC1-3. Arrows indicate predicted trajectory between FB clusters according to Buechler *et al* (Buechler *et al.*, 2021). (**D**) tSPACE principal component analysis projection (tPC1 and tPC3) of adult colonic FB subsets. Colors represent the indicated FB subsets, where core and tip cells within each FB subset were compared for differential gene expression. (**E**) tPC1 vs tPC3 projection showing overlaid gene count (top) and read count (bottom) on tip and core cells highlighted in (**D**). (**F**) GO term analysis using Enrichr (GO Biological Process 2018) of DEGs enriched in core versus tip cells within each FB subset. The data shown are the 9 GO terms shared by core cells within each FB subset (ranked by adjusted p value). Dotted line denotes significant adjusted p value of 0.05. (**G**) Discrete heatmap indicating the DEGs accounting for the top 9 processes. Blue squares represent the genes in GO term. White squares represent the genes not in GO term. Genes labelled black are expressed at significantly higher levels in core compared with tip cells within each FB subsets. Genes in red are expressed at significantly higher levels in core cells from 1-2 of the three FB subsets.

Figure S6

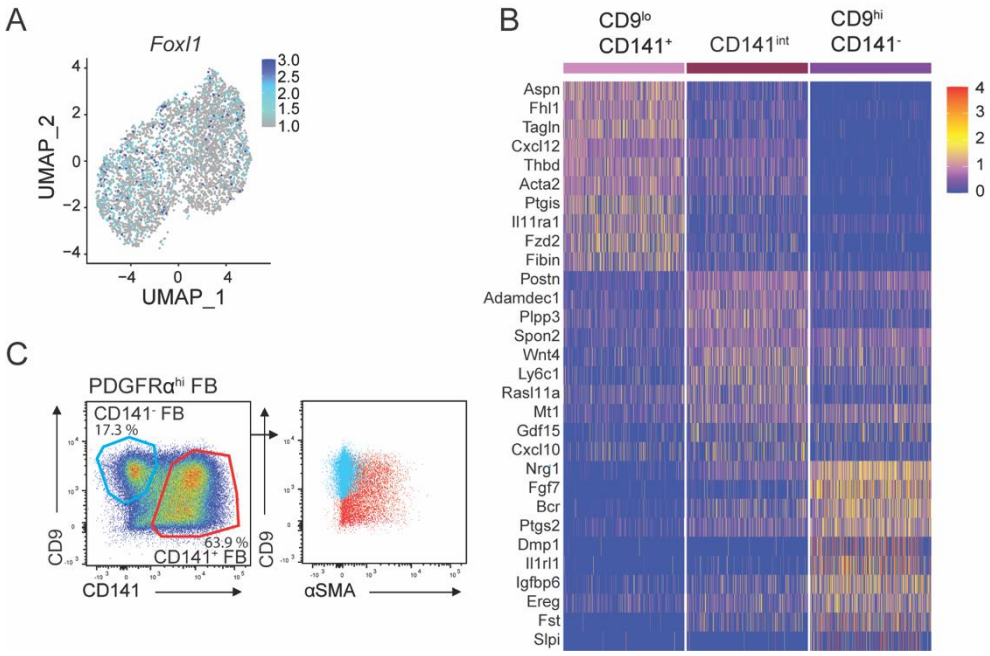


Figure S6.

Related to Figure 6. (**A**) Integrated and normalized *Foxl1* expression projected onto UMAP of colonic PDGFRα^{hi} FB. (**B**) Heatmap with scaled expression (integrated) of top 10 differentially expressed genes (DEGs) between the three colonic PDGFRα^{hi} subsets. (**C**) Flow cytometric analysis of PDGFRα^{hi} FB showing expression of αSMA (right hand plot) by CD141^{+} (red) and CD141^{-} (blue) cells using the gates depicted in the left hand panel. Results are representative staining of 2 experiments with 3 mice/experiment.

Supplementary tables:
https://www.biorxiv.org/content/10.1101/2021.08.13.456086v1.supplementary-material

Discussion and future perspectives

In this thesis I have applied bioinformatics to solve questions of heterogeneity within MNPs and iMSCs. Though general pipelines, like Seurat[280], exist for purposes of doing normalization, batch correction, linear regression of unwanted effects, clustering and differential gene expression analysis, these pipelines do not enlighten on the best uses with different types of data sets[291]. Many of the widely used tools like Seurat and the TI tool Monocle[292] are developed based on high heterogeneity data sets containing widely different cell populations. By working with the data sets on iMSC and MNP I have found that to increase the biologically relevant outputs the data sets should be cleaned from contaminating data. Contaminating cell types such as endothelial cells or T cells cause a high degree of variability to explain how these are different from our cells of interest. I have found that by reclustering (the computational analogue of gating) only the interesting sources of variability are included for downstream analysis, which ultimately helps us explore the biologically relevant differences between tissues and in populations that with the glance of an eye look homogeneous.

7 Trajectory inference of iMSC ontogeny

In manuscript 3, we used imputation to smooth differences between adult and embryonic data before performing tSpace TI, to understand the ontogeny of iMSCs. In other tissues SMC and FB can be derived from mesothelium (a single cell layer lining organs and cavities) after injury and the same has been observed for intestinal SMCs in muscle layer[245,246]. By integrating the embryonic and adult data and smoothing with MAGIC imputation[293] before running TI, we found that adult SMCs and CD81$^+$ FB connected directly to the *Itgβ1*$^+$ iMSC from E12.5 and proposing ontogeny from the *Gli*$^+$ mesothelium lying in between. This was confirmed with lineage tracing and intestinal transplantation, which showed that all the adult iMSC populations could be derived from the *Gli*$^+$ progenitor. This demonstrate that even though the adult iMSC and embryonic data had vast differences, integration and imputation combined with TI could indicate biologically relevant development.

8 Characterization of the cDC3 in the human LP

The first manuscript in this thesis, focused on the LP MNPs and adds information to the growing field of cDC complexity in the human tissues. Specifically, we could identify that the newly described cDC3[142–144] also exist in the human LP. We found, consistently with blood data, that intestinal cDC3 expresses CD1C, CD163 and low levels of CD14 compared to cDC2[105,143,145,146]. Furthermore, using CITE-seq we could show CD1c$^+$ CD11a$^+$ CD207$^-$ is selectively enriched on LP cDC3, where CD1c$^+$ CD11a$^-$ CD207$^+$ are enriched for cDC2. The CD11a$^+$ population was enriched in the colon LP compared to the SI LP by flow cytometry, whereas CD207$^+$ were more abundant in the SI LP, together suggesting that the SI LP has a higher frequency of cDC2 whereas the cDC3 could be proportionally more abundant in the colon LP. The previous studies on intestinal cDCs have observed higher frequency of CD103$^+$ cDC2s in SI LP aligning with our observation for CD207$^+$ cDC2[179]. However, we cannot be completely sure that these are the same populations as CD103 could only somewhat enrich for cDC2. Furthermore, CITE-seq expression of CD103 did not separate cDC2 and cDC3 as well as combining CD11a and CD207 did. With these results we suggest that CD207 is a better marker for *bona fide* human LP cDC2. The role of human intestinal cDC2 and cDC3 is not yet clear, but blood cDC3s can be activated upon TLR stimulation and initiate T cell responses as well as cDC2 do[143]. Indicating that cDC3s can actively participate in priming immune cells,

furthermore, our finding that these cells also exist in the GALT adds their location to the *in vivo* derived results.

With the data presented here, we had a limited set of phenotypic markers integrated with the scRNA-seq data. In retrospect, it could have been beneficial to include a much broader spectrum, for instance CD163, to check how well the protein expression match the gene expression in cDC3. CD163 is potentially a very strong candidate for selective expression on cDC3 against cDC2[143,144]. With a broader panel we could also inspect whether there are any proteins with gradient expression from cDC2 to ambiguous to cDC3. It remains clear that even with our efforts the ambiguous cells still take up a large proportion of the total $CD1c^+$ cells, however markers used in literature to distinguish gradient patterns like CD5[143,144] did not work in our hands in the LP. We would not expect the ambiguous cells to have a distinct phenotypic profile from the transcriptional convergence we have observed, but there might be negative-low-high expression patterns in markers we have not explored. This was explored in blood where all $CD5^-$, thought to be cDC3, were inspected as a continuum of CD14 and CD163 expression with double positive cDC3 being furthest from the profile of cDC2[144]. It is possible that a similar continuum could exist from cDC3 through ambiguous cDC to cDC2 in the intestine.

Though we have identified the existence of cDC2 and cDC3 in the intestinal LP, we do not know anything on their location in the LP. This could be explored with the new spatial transcriptomics, which would also allow us to interrogate the surrounding cell types as well. This sort of data would allow us to predict interaction networks with not just other immune cells but also structural cells such as endothelium. Together, these could give us an indication of whether cDC2 and cDC3 are located in the same microenvironments and if they specific immunogenic or tolerogenic roles there. Finally, spatial transcriptomics[294] would allow us to question if the "ambiguous" cDCs are microenvironmentally primed to drive the transcriptional convergence we observe.

9 cDC precursors in the human lamina propria

Pre-cDC have been shown to be proliferating in peripheral tissues like the lung and the SI, where the can give rise to mature cDCs. In an elegant study in mice Cabeza-Cabrerizo *et al.* demonstrated that most of the pre-cDC arrive from the blood already committed to either cDC2 or cDC1 and that each gives rise to multiple sister clones by proliferating in the tissue[295]. Consistent with these data, most of the proliferating pre-cDC in our data could be computationally shown to have commitment to a downstream lineage, whereas only very few cells did not show preferential commitment similar to what was shown in blood[296]. By studying the correlation of overlapping variable genes between our cDCs and the genes from a BM study on hematopoiesis[297] we found that the small population of uncommitted putative pre-cDC also correlated in gene expression with earlier precursors in hematopoiesis compared to any of the mature or pre-committed cDCs. Consistently, we also found that the latest pre-committed subsets showed preferential correlation to the BM mature counterparts. In blood, uncommitted pre-cDC have been reported to differential express *IRF4* compared to the mature cDC1 and *IRF8* compared to the mature cDC2[155], which was consistent in our data, where the uncommitted cells express both *IRF4* and *IRF8* to some degree. *IRF4* remains upregulated in our cDC2 pseudo-timeline, whereas our cDC1 loose *IRF4*, but strongly upregulate *IRF8*. In both mature cDC1 and cDC2 different HLA genes were also

upregulated compared to the uncommitted precursors in blood[296], which again is consistent with the results we have shown. However, the blood cDC data from See *et al.'s* study could not attribute with any knowledge on cDC3 as they did not separate a blood cDC3 from the cDC2s. Later studies have investigated the ontogeny of the cDC3 and found that cDC3 arise separately from cDC1 and cDC2 specific CDP precursors[143,298]. It was suggested that the precursor for cDC3 express low *IRF8*, which is also true for the cDC3 trajectory we have shown in manuscript 1[298]. However, the trajectory data on putative cDC development in LP, which we have presented here does not infer nor disprove that the split to cDC3 happens before the CDP stage. Our cDC trajectory proposes that an uncommitted progenitor in the LP can give rise to all three types of committed precursors. The earliest cDC precursor observed in the blood is the pre-cDC[296], which means blood pre-cDC are the earliest contribution of cDC progenitors from blood. In order to confirm this trajectory in LP we would need to show that the progenitors can give rise to all three cDCs, however these cells are very rare and we currently have no definitive way of separating them from their mature counterparts in LP.

10 Myeloid cells in GALT

In PP several specific myeloid populations and their location has been observed. These include TIM4$^-$ lyso-Mac and lyso-DC, both located in the SED, TB-Macs in GC and TIM4$^+$ lyso-Mac in the interfollicular regions. In our second manuscript we found that though *BST2* and *LYZ*[22] were expressed selectively in one macrophage population and mostly in one cDC3 population, these were not exclusive populations to the GALT. However, the anchor integration, which was used here will mainly include variance which is replicated across the samples[280], in this case variance consistent across tissues. Subsequently, transcriptomically similar macrophage populations will be clustered together across tissues. The differences in these and whether *LYZ* and *BST2* are preferentially expressed in specific subsets in GALT can be analysed in future analyses comparing the individual tissues across compartments. Proportionally, the PP were found to have more mature macrophages compared to SM-ILFs, however, the isolation of PP includes interfollicular regions, which could include contaminating LP[29]. Though it is possible to cut out the follicles individually this would mostly exclude the T cell rich interfollicular regions, which also are an important inductive sites of the PP[299].

Interestingly, we found that the monocyte and macrophage compartment in both ILF and PP had higher frequencies of CD14$^+$ monocytes compared to surrounding LP. We hypothesize that this could be due to the GALT vasculature contributing more blood vessels and thus peripheral blood cells than the surrounding LP. This hypothesis is consistent with recent findings from the Agace lab, where ILF were shown to be tightly surrounded by blood vessels[299]. This could propose that the CD14$^+$ monocytes actually reflect contribution from blood and not necessarily that they are the most abundant myeloid population inside GALT. In the future this could be confirmed if CD14$^+$ monocytes are localized mainly to the endothelial vessels. This can also be investigated by studying the transcriptomic differences between CD14$^+$ blood monocytes and P1 monocytes[81], which are thought to be induced by transmigration across the endothelium[300]. If the transcriptomic program induced by the endothelium is upregulated in GALT CD14+ monocytes, it could indicate that these are actual present in the GALT parenchyme.

11 Pro-inflammatory monocytes in Crohn's Disease

Our current understanding of MNPs in colitis is very limited. Here, we took the first steps in an unbiased characterization of major changes within the MNP compartment. Our results are of course limited by only having one CD sample, but we plan to do similar analysis including more patient material. This will also allow us to interrogate whether there any specific changes in the different populations, for instance if the resident mature macrophages undergo any changes in IBD. This has been approached in a study where monocytes, macrophages and their inter-developmental relation was compared in DSS induced colitis and naïve mice. Some responses like endoplasmatic reticulum stress response were described to be upregulated in both colon monocytes and macrophages in colitis. However, overall the main biologically relevant difference between colitis and naïve state was found in the monocyte compartment: monocytes in colitis expressed CXCL1, CCL7, and CCL8 compared to the naïve mice, and consistently, these were also upregulated in active human IBD biopsies. Though this observation was made for monocytes, the macrophages in both colitis and naïve mice had higher expression of CCL8, CCL7, and CXCL1[252]. Consistent with this observation CCL8 has been shown to be expressed by $CD169^{+}$ vas-Mac DSS induced colitis, leading to a recruitment of monocytes [166]. However, whether a human analogue to vas-Mac expresses CCL8 in IBD is yet to be clarified but will be possible with by comparing healthy and IBD mac populations directly.

From manuscript I, we found that early and late intermediates of monocytes to macrophage development were proportionally enriched in IBD biopsies. The preliminary results from manuscript 2 indicate that the majority of the monocyte contribution in CD is due to a specific subset of monocytes, which we have observed to have a pro-inflammatory profile with expression of *CXCL9* and *CXCL10*. Domanska *et al.* recently showed that a small subset strikingly similar to these and their TI analysis could indicate that these pro-inflammatory monocytes have undergone a different trajectory to the normal waterwall[82]. In the future we would like to question this alternative trajectory in IBD and if we can observe indication of signalling or transcription factor events at the branch point, which could cause the diverging differentiation path. Such a diverging trajectory has already been described in mice, where surgically induced colitis lead to induction of a range of genes including *Cxcl1*[256], which we also observed as differentially expressed in the pro-inflammatory monocytes. In the mice model there can only be observed small changes phenotypical changes in the monocyte waterfall between naïve mice and colitis[256]. It would also be of interest to interrogate these phenotypically in humans, for instance using a broad CITE-seq panel to examine whether there are any markers that phenotypically distinguishes these from "normal" P1 and P2 monocytes. Use of a CITE-seq panel is advantageous by the possibility of using a wide panel of surface markers without the limitations of fluorescence from flow cytometry. Finally, we also found that the profile of these inflammatory macrophages is even more striking in the inflamed cLP compared to healthy and uninvolved tissues. This seems to be a gradient upregulation as many genes followed a pattern with slight upregulation from healthy to uninvolved tissue before high upregulation in inflamed cLP. This profile regulation could also explain why we did not observe the pro-inflammatory monocytes in our first manuscript, with a less distinct profile, they are harder to identify, increasingly so being a rare cell subset in healthy tissues. Moving forward it would be of interest to compare the pro-inflammatory monocytes between healthy, uninvolved and inflamed tissue to understand how they in IBD contribute to disease.

References

1. Mowat, A. M. & Agace, W. W. Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* **14**, 667–685 (2014).
2. Maloy, K. J. & Powrie, F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. *Nature* **474**, 298–306 (2011).
3. Agace, W. W. & McCoy, K. D. Regionalized Development and Maintenance of the Intestinal Adaptive Immune Landscape. *Immunity* **46**, 532–548 (2017).
4. Shi, N., Li, N., Duan, X. & Niu, H. Interaction between the gut microbiome and mucosal immune system. *Mil. Med. Res.* **4**, 1–7 (2017).
5. Lee, S. H. Intestinal Permeability Regulation by Tight Junction: Implication on Inflammatory Bowel Diseases. *Intest. Res.* **13**, 11 (2015).
6. Delacour, D., Salomon, J., Robine, S. & Louvard, D. Plasticity of the brush border-the yin and yang of intestinal homeostasis. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 161–174 (2016).
7. Peterson, L. W. & Artis, D. Intestinal epithelial cells: Regulators of barrier function and immune homeostasis. *Nat. Rev. Immunol.* **14**, 141–153 (2014).
8. 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).
9. Roulis, M. & Flavell, R. A. Fibroblasts and myofibroblasts of the intestinal lamina propria in physiology and disease. *Differentiation* **92**, 116–131 (2016).
10. Habtezion, A., Nguyen, L. P., Hadeiba, H. & Butcher, E. C. Leukocyte Trafficking to the Small Intestine and Colon. *Gastroenterology* **150**, 340–354 (2016).
11. Danese, S. Nonimmune cells in inflammatory bowel disease: from victim to villain. *Trends Immunol.* **29**, 555–564 (2008).
12. Cifarelli, V. & Eichmann, A. The Intestinal Lymphatic System: Functions and Metabolic Implications. *Cmgh* **7**, 503–513 (2019).
13. Bernier-Latmani, J. & Petrova, T. V. Intestinal lymphatic vasculature: Structure, mechanisms and functions. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 510–526 (2017).
14. Smith , MacDonald, Thomas T.,, Blumberg, R. S., Society for Mucosal Immunology., P. D. *Principles of mucosal immunology*. (Garland Science, 2013).
15. Houston, S. A. *et al.* The lymph nodes draining the small intestine and colon are anatomically separate and immunologically distinct. *Mucosal Immunol.* **9**, 468–478 (2016).
16. Fenton, T. M. *et al.* Immune Profiling of Human Gut-Associated Lymphoid Tissue Identifies a Role for Isolated Lymphoid Follicles in Priming of Region-Specific Immunity. *Immunity* **52**, 557-570.e6 (2020).
17. Mörbe, U. M. *et al.* Human gut-associated lymphoid tissues (GALT); diversity, structure, and function. *Mucosal Immunol.* **14**, 793–802 (2021).
18. Mabbott, N. A., Donaldson, D. S., Ohno, H., Williams, I. R. & Mahajan, A. Microfold (M) cells: Important immunosurveillance posts in the intestinal epithelium. *Mucosal Immunol.* **6**, 666–677 (2013).
19. Farstad, I. N., Halstensen, T. S., Kvaie, D., Fausa, O. & Brandtzaeg, P. Topographic distribution of homing receptors on B and T cells in human gut-associated lymphoid tissue: Relation of L-selectin and integrin α4β7 to naive and memory phenotypes. *Am. J. Pathol.* **150**, 187–199 (1997).
20. Brandtzaeg, P., Kiyono, H., Pabst, R. & Russell, M. W. Terminology: Nomenclature of

mucosa-associated lymphoid tissue. *Mucosal Immunol.* **1**, 31–37 (2008).
21. Silva, C. Da, Wagner, C., Bonnardel, J., Gorvel, J. & Lelouard, H. The Peyer ' s Patch Mononuclear Phagocyte System at Steady State and during Infection. **8**, (2017).
22. Wagner, C. *et al.* Differentiation Paths of Peyer ' s Patch LysoDCs Are Linked to Sampling Site Positioning , Migration , and T Article Differentiation Paths of Peyer ' s Patch LysoDCs Are Linked to Sampling Site Positioning , Migration , and T Cell Priming. *Cell Rep.* **31**, (2020).
23. Lelouard, H., Fallet, M., De Bovis, B., Méresse, S. & Gorvel, J. Peyer's patch dendritic cells sample antigens by extending dendrites through M cell-specific transcellular pores. *Gastroenterology* **142**, 592-601.e3 (2012).
24. Spencer, J. O., Finn, T. & Isaacson, P. G. Human Peyer ' s patches : study immunohistochemical. *Gut* **27**, 405–410 (1986).
25. Reboldi, A. & Cyster, J. G. Peyer's patches: Organizing B-cell responses at the intestinal frontier. *Immunol. Rev.* **271**, 230–245 (2016).
26. Gibbons, D. L. & Spencer, J. Mouse and human intestinal immunity: Same ballpark, different players; Different rules, same score. *Mucosal Immunol.* **4**, 148–157 (2011).
27. Pabst, O. *et al.* Cryptopatches and isolated lymphoid follicles: Dynamic lymphoid tissues dispensable for the generation of intraepithelial lymphocytes. *Eur. J. Immunol.* **35**, 98–107 (2005).
28. O'Leary, A. D. & Sweeney, E. C. Lymphoglandular complexes of the normal colon: Histochemistry and immunohistochemistry. *Ir. J. Med. Sci.* **156**, 142–148 (1987).
29. Jørgensen, P. B. *et al.* Identification, isolation and analysis of human gut-associated lymphoid tissues. *Nat. Protoc.* **16**, (2021).
30. Spencer, J., Siu, J. H. Y. & Montorsi, L. Human intestinal lymphoid tissue in time and space. *Mucosal Immunol.* **12**, 296–298 (2019).
31. Hase, K. *et al.* Uptake through glycoprotein 2 of FimH + bacteria by M cells initiates mucosal immune response. *Nature* **462**, 226–230 (2009).
32. Langman, J. M. & Rowland, R. The number and distribution of lymphoid follicles in the human large intestine. *J. Anat.* **149**, 189–194 (1986).
33. Gebbers, J. O., Kennel, I. & Laissue, J. A. [Lymphoid follicles of the human large bowel mucosa: structure and function]. *Verh. Dtsch. Ges. Pathol.* **76**, 126–130 (1992).
34. Gustafson, C. E. *et al.* Limited expression of APRIL and its receptors prior to intestinal IgA plasma cell development during human infancy. *Mucosal Immunol.* **7**, 467–477 (2014).
35. Meier, D. *et al.* Immunological status of isolated lymphoid follicles after intestinal transplantation. *Am. J. Transplant. Off. J. Am. Soc. Transplant. Am. Soc. Transpl. Surg.* **14**, 2148–2158 (2014).
36. Boursier, L., Gordon, J. N., Thiagamoorthy, S., Edgeworth, J. D. & Spencer, J. Human intestinal IgA response is generated in the organized gut-associated lymphoid tissue but not in the lamina propria. *Gastroenterology* **128**, 1879–1889 (2005).
37. Cheroutre, H., Lambolez, F. & Mucida, D. The light and dark sides of intestinal intraepithelial lymphocytes. *Nat. Rev. Immunol.* **11**, 445–456 (2011).
38. Barnes, M. J. & Powrie, F. Regulatory T Cells Reinforce Intestinal Homeostasis. *Immunity* **31**, 401–411 (2009).
39. Spencer, J. & Sollid, L. M. The human intestinal B-cell response. *Mucosal Immunol.* **9**, 1113–1124 (2016).
40. Panda, S. K. & Colonna, M. Innate lymphoid cells in mucosal immunity. *Front.*

Immunol. **10**, 1–13 (2019).
41. Saez, A. *et al.* Innate lymphoid cells in intestinal homeostasis and inflammatory bowel disease. *Int. J. Mol. Sci.* **22**, (2021).
42. Albert-Bayo, M. *et al.* Intestinal Mucosal Mast Cells: Key Modulators of Barrier Function and Homeostasis. *Cells* **8**, 135 (2019).
43. Fournier, B. M. & Parkos, C. A. The role of neutrophils during intestinal inflammation. *Mucosal Immunol.* **5**, 354–366 (2012).
44. Yantiss, R. K. Eosinophils in the GI tract: How many is too many and what do they mean? *Mod. Pathol.* **28**, S7–S21 (2015).
45. Guilliams, M. *et al.* Dendritic cells, monocytes and macrophages: a unified nomenclature based on ontogeny. *Nat. Publ. Gr.* **14**, 571–578 (2014).
46. Skinner, B. M. & Johnson, E. E. P. Nuclear morphologies: their diversity and functional relevance. *Chromosoma* **126**, 195–212 (2017).
47. Lugo-Villarino, G. *et al.* Identification of dendritic antigen-presenting cells in the zebrafish. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15850–15855 (2010).
48. Li, D. & Wu, M. Pattern recognition receptors in health and diseases. *Signal Transduct. Target. Ther.* **6**, 1–24 (2021).
49. Kawasaki, T. & Kawai, T. Toll-like receptor signaling pathways. *Front. Immunol.* **5**, 1–8 (2014).
50. Uribe-Querol, E. & Rosales, C. Phagocytosis: Our Current Understanding of a Universal Biological Process. *Front. Immunol.* **11**, 1–13 (2020).
51. Gordon, S. Phagocytosis: An Immunobiologic Process. *Immunity* **44**, 463–475 (2016).
52. Underhill, D. M. & Goodridge, H. S. Information processing during phagocytosis. *Nat. Rev. Immunol.* **12**, 492–502 (2012).
53. Freeman, S. A. & Grinstein, S. Phagocytosis: Receptors, signal integration, and the cytoskeleton. *Immunol. Rev.* **262**, 193–215 (2014).
54. Rock, K. L., Reits, E. & Neefjes, J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol.* **37**, 724–737 (2016).
55. Hennecke, J. & Wiley, D. C. T cell receptor-MHC interactions up close. *Cell* **104**, 1–4 (2001).
56. Joffre, O. P., Segura, E., Savina, A. & Amigorena, S. Cross-presentation by dendritic cells. *Nat. Rev. Immunol.* **12**, 557–569 (2012).
57. Gutiérrez-Martínez, E. *et al.* Cross-presentation of cell-associated antigens by MHC class I in dendritic cell subsets. *Front. Immunol.* **6**, (2015).
58. Katikaneni, D. S. & Jin, L. B cell MHC class II signaling: A story of life and death. *Hum. Immunol.* **80**, 37–43 (2019).
59. Jenkinson, S. R. *et al.* Importance of MHCII expression on thymic epithelium versus dendritic cells for the positive and negative selection of CD4 T Cells (138.28). *J. Immunol.* **182**, 138.28 LP-138.28 (2009).
60. Akashi K, Traver D, Miyamoto T & IL, W. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
61. Seita, J. & Weissman, I. L. Hematopoietic stem cell: Self-renewal versus differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2**, 640–653 (2010).
62. Haas, S., Trumpp, A. & Milsom, M. D. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell* **22**, 627–638 (2018).
63. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418–426 (2018).

64. Guilliams, M., Mildner, A. & Yona, S. Review Developmental and Functional Heterogeneity of Monocytes. *Immunity* **49**, 595–613 (2018).
65. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
66. Jacobsen, S. E. W. & Nerlov, C. Haematopoiesis in the era of advanced single-cell technologies. *Nat. Cell Biol.* **21**, 2–8 (2019).
67. Loughran, S. J., Haas, S., Wilkinson, A. C., Klein, A. M. & Brand, M. Lineage commitment of hematopoietic stem cells and progenitors: insights from recent single cell and lineage tracing technologies. *Exp. Hematol.* **88**, 1–6 (2020).
68. Pellin, D. *et al.* A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat. Commun.* **10**, (2019).
69. Ginhoux, F. *et al.* Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science (80-.).* **330**, 841–845 (2010).
70. Hoeffel, G. *et al.* C-Myb+ Erythro-Myeloid Progenitor-Derived Fetal Monocytes Give Rise to Adult Tissue-Resident Macrophages. *Immunity* **42**, 665–678 (2015).
71. Guilliams, M. *et al.* Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. *J. Exp. Med.* **210**, 1977–1992 (2013).
72. Ginhoux, F. & Guilliams, M. Tissue-Resident Macrophage Ontogeny and Homeostasis. *Immunity* **44**, 439–449 (2016).
73. van de Laar, L. *et al.* Yolk Sac Macrophages, Fetal Liver, and Adult Monocytes Can Colonize an Empty Niche and Develop into Functional Tissue-Resident Macrophages. *Immunity* **44**, 755–768 (2016).
74. Ginhoux, F. & Jung, S. Monocytes and macrophages: Developmental pathways and tissue homeostasis. *Nat. Rev. Immunol.* **14**, 392–404 (2014).
75. Ajami, B., Bennett, J. L., Krieger, C., Tetzlaff, W. & Rossi, F. M. V. Local self-renewal can sustain CNS microglia maintenance and function throughout adult life. *Nat. Neurosci.* **10**, 1538–1543 (2007).
76. Bain, C. C. *et al.* Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6C hi monocyte precursors. **6**, (2013).
77. Bain, C. C. *et al.* Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. **15**, (2014).
78. Schridde, A. *et al.* Tissue-specific differentiation of colonic macrophages requires TGFβ receptor-mediated signaling. *Mucosal Immunol.* **10**, 1387–1399 (2017).
79. Joeris, T., Müller-Luda, K., Agace, W. W. & Mowat, A. M. I. Diversity and functions of intestinal mononuclear phagocytes. *Mucosal Immunol.* **10**, 845–864 (2017).
80. Tamoutounour, S. *et al.* CD64 distinguishes macrophages from dendritic cells in the gut and reveals the Th1-inducing role of mesenteric lymph node macrophages during colitis. *Eur. J. Immunol.* **42**, 3150–3166 (2012).
81. Bujko, A. *et al.* Transcriptional and functional profiling defines human small intestinal macrophage subsets. (2018).
82. Domanska, D. *et al.* Single-cell transcriptomic analysis of human colonic macrophages reveals niche-specific subsets. *J. Exp. Med.* **219**, e20211846 (2022).
83. Tamoutounour, S. *et al.* Origins and functional specialization of macrophages and of conventional and monocyte-derived dendritic cells in mouse skin. *Immunity* **39**, 925–938 (2013).

84. Bain, C. C. *et al.* Long-lived self-renewing bone marrow-derived macrophages displace embryo-derived cells to inhabit adult serous cavities. *Nat. Commun.* **7**, 1–14 (2016).
85. Bajpai, G. *et al.* The human heart contains distinct macrophage subsets with divergent origins and functions. *Nat. Med.* **24**, 1234–1245 (2018).
86. Epelman, S. *et al.* Embryonic and adult-derived resident cardiac macrophages are maintained through distinct mechanisms at steady state and during inflammation. *Immunity* **40**, 91–104 (2014).
87. Tan, S. Y. S. & Krasnow, M. A. Developmental origin of lung macrophage diversity. *Dev.* **143**, 1318–1327 (2016).
88. Richter, L. *et al.* Transcriptional profiling reveals monocyte-related macrophages phenotypically resembling DC in human intestine. *Mucosal Immunol.* **11**, 1512–1523 (2018).
89. Mass, E. *et al.* Specification of tissue-resident macrophages during organogenesis. *Science (80-.).* **353**, (2016).
90. Blériot, C. *et al.* Liver-Resident Macrophage Necroptosis Orchestrates Type 1 Microbicidal Inflammation and Type-2-Mediated Tissue Repair during Bacterial Infection. *Immunity* **42**, 145–158 (2015).
91. Bonnardel, J. *et al.* Stellate Cells, Hepatocytes, and Endothelial Cells Imprint the Kupffer Cell Identity on Monocytes Colonizing the Liver Macrophage Niche. *Immunity* **51**, 638-654.e9 (2019).
92. Scott, C. L. *et al.* Bone marrow-derived monocytes give rise to self-renewing and fully differentiated Kupffer cells. *Nat. Commun.* **7**, 1–10 (2016).
93. Machiels, B. *et al.* A gammaherpesvirus provides protection against allergic asthma by inducing the replacement of resident alveolar macrophages with regulatory monocytes. *Nat. Immunol.* **18**, 1310–1320 (2017).
94. Ajami, B., Bennett, J. L., Krieger, C., McNagny, K. M. & Rossi, F. M. V. Infiltrating monocytes trigger EAE progression, but do not contribute to the resident microglia pool. *Nat. Neurosci.* **14**, 1142–1150 (2011).
95. Gordon, S. & Plüddemann, A. Tissue macrophages : heterogeneity and functions. 1–18 (2017) doi:10.1186/s12915-017-0392-4.
96. Davies, L. C., Jenkins, S. J., Allen, J. E. & Taylor, P. R. Tissue-resident macrophages. *Nat. Immunol.* **14**, 986–995 (2013).
97. Lavin, Y., Mortha, A., Rahman, A. & Merad, M. Regulation of macrophage development and function in peripheral tissues. *Nat. Rev. Immunol.* **15**, 731–744 (2015).
98. Lavin, Y. & Merad, M. Macrophages: gatekeepers of tissue integrity. *Cancer Immunol. Res.* **1**, 201–209 (2013).
99. Hopkinson-Woolley, J., Hughes, D., Gordon, S. & Martin, P. Macrophage recruitment during limb development and wound healing in the embryonic and foetal mouse. *J. Cell Sci.* **107**, 1159–1167 (1994).
100. Murray, P. J. & Wynn, T. A. Protective and pathogenic functions of macrophage subsets. *Nat. Rev. Immunol.* **11**, 723–737 (2011).
101. Boyce, B. F., Li, J., Xing, L. & Yao, Z. Bone Remodeling and the Role of TRAF3 in Osteoclastic Bone Resorption. *Front. Immunol.* **9**, 1–12 (2018).
102. Dorrington, M. G. & Fraser, I. D. C. NF-κB signaling in macrophages: Dynamics, crosstalk, and signal integration. *Front. Immunol.* **10**, (2019).
103. Zhou, L., Cao, X., Fang, J., Li, Y. & Fan, M. Macrophages polarization is mediated by

the combination of PRR ligands and distinct inflammatory cytokines. *Int. J. Clin. Exp. Pathol.* **8**, 10964–10974 (2015).
104. Shi, C. & Pamer, E. G. Monocyte recruitment during infection and inflammation. *Nat. Rev. Immunol.* **11**, 762–774 (2011).
105. Guilliams, M. & Svedberg, F. R. Does tissue imprinting restrict macrophage plasticity? *Nat. Immunol.* **22**, 118–127 (2021).
106. Viola, M. F. & Boeckxstaens, G. Niche-specific functional heterogeneity of intestinal resident macrophages. *Gut* **70**, 1383–1395 (2021).
107. Gosselin, D. *et al.* Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
108. Gautiar, E. L. *et al.* Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat. Immunol.* **13**, 1118–1128 (2012).
109. Scott, C. L. & Guilliams, M. The role of Kupffer cells in hepatic iron and lipid metabolism. *J. Hepatol.* **69**, 1197–1199 (2018).
110. Gotur, S. P. & Wadhwan, V. Tingible body macrophages. *J. Oral Maxillofac. Pathol.* **24**, 418–420 (2020).
111. Rahman, Z. S. M., Shao, W.-H., Khan, T. N., Zhen, Y. & Cohen, P. L. Impaired Apoptotic Cell Clearance in the Germinal Center by Mer-Deficient Tingible Body Macrophages Leads to Enhanced Antibody-Forming Cell and Germinal Center Responses. *J. Immunol.* **185**, 5859–5868 (2010).
112. Narasimhan, P. B., Marcovecchio, P., Hamers, A. A. J. & Hedrick, C. C. Nonclassical Monocytes in Health and Disease. (2019).
113. Guilliams, M., Thierry, G. R., Bonnardel, J. & Bajenoff, M. Review Establishment and Maintenance of the Macrophage Niche. *Immunity* **52**, 434–451 (2020).
114. Tsou, C. L. *et al.* Critical roles for CCR2 and MCP-3 in monocyte mobilization from bone marrow and recruitment to inflammatory sites. *J. Clin. Invest.* **117**, 902–909 (2007).
115. Patel, A. A. *et al.* The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* **214**, 1913–1923 (2017).
116. Yona, S. *et al.* Article Fate Mapping Reveals Origins and Dynamics of Monocytes and Tissue Macrophages under Homeostasis. *IMMUNI* **38**, 79–91 (2013).
117. Hettinger, J. *et al.* Origin of monocytes and macrophages in a committed progenitor. *Nat. Immunol.* **14**, 821–830 (2013).
118. Murphy, K., Travers, P., Walport, M. & Janeway, C. *Janeway's Immunobiology*. (New York: Garland Science., 2008).
119. Kim, M. K. & Kim, J. Properties of immature and mature dendritic cells: phenotype, morphology, phagocytosis, and migration. *RSC Adv.* **9**, 11230–11238 (2019).
120. Banchereau, J. *et al.* Immunobiology of Dendritic Cells. *Annu. Rev. Immunol.* **18**, 767–811 (2000).
121. ten Broeke, T., Wubbolts, R. & Stoorvogel, W. MHC class II antigen presentation by dendritic cells regulated through endosomal sorting. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
122. Ramachandra, L., Simmons, D. & Harding, C. V. MHC molecules and microbial antigen processing in phagosomes. *Curr. Opin. Immunol.* **21**, 98–104 (2009).
123. Shin, J. S. *et al.* Surface expression of MHC class II in dendritic cells is controlled by regulated ubiquitination. *Nature* **444**, 115–118 (2006).

124. Cabeza-Cabrerizo, M., Cardoso, A., Minutti, C. M., Pereira Da Costa, M. & Reis E Sousa, C. Dendritic Cells Revisited. *Annu. Rev. Immunol.* **39**, 131–166 (2021).
125. Kumar, P., Bhattacharya, P. & Prabhakar, B. S. A comprehensive review on the role of co-signaling receptors and Treg homeostasis in autoimmunity and tumor immunity. *J. Autoimmun.* **95**, 77–99 (2018).
126. Hilligan, K. L. & Ronchese, F. Antigen presentation by dendritic cells and their instruction of CD4+ T helper cell responses. *Cell. Mol. Immunol.* **17**, 587–599 (2020).
127. Hildner, K. *et al.* Batf3 deficiency reveals a critical role for CD8α+ dendritic cells in cytotoxic T cell immunity. *Science (80-.).* **322**, 1097–1100 (2008).
128. Poulin, L. F. *et al.* DNGR-1 is a specific and universal marker of mouse and human Batf3-dependent dendritic cells in lymphoid and nonlymphoid tissues. *Blood* **119**, 6052–6062 (2012).
129. Tamura, T. *et al.* IFN Regulatory Factor-4 and -8 Govern Dendritic Cell Subset Development and Their Functional Diversity. *J. Immunol.* **174**, 2573–2581 (2005).
130. Ma, W. *et al.* Single cell RNA-Seq reveals pre-cDCs fate determined by transcription factor combinatorial dose. *BMC Mol. Cell Biol.* **20**, 1–14 (2019).
131. Amon, L., Lehmann, C. H. K., Heger, L., Heidkamp, G. F. & Dudziak, D. The ontogenetic path of human dendritic cells. *Mol. Immunol.* **120**, 122–129 (2020).
132. Guilliams, M. *et al.* Unsupervised High-Dimensional Analysis Aligns Dendritic Cells across Tissues and Species. *Immunity* **45**, 669–684 (2016).
133. Everts, B. *et al.* Migratory CD103+ dendritic cells suppress helminth-driven type 2 immunity through constitutive expression of IL-12. *J. Exp. Med.* **213**, 35–51 (2016).
134. Blander, J. M. Regulation of the Cell Biology of Antigen Cross-Presentation. *Annu. Rev. Immunol.* **36**, 717–753 (2018).
135. Collin, M. Human dendritic cell subsets : an update. 3–20 (2018) doi:10.1111/imm.12888.
136. Schlitzer, A. *et al.* IRF4 Transcription Factor-Dependent CD11b+ Dendritic Cells in Human and Mouse Control Mucosal IL-17 Cytokine Responses. *Immunity* **38**, 970–983 (2013).
137. Tussiwand, R. *et al.* Klf4 Expression in Conventional Dendritic Cells Is Required for T Helper 2 Cell Responses. *Immunity* **42**, 916–928 (2015).
138. Lewis, K. L. *et al.* Notch2 receptor signaling controls functional differentiation of dendritic cells in the spleen and intestine. *Immunity* **35**, 780–791 (2011).
139. Connor, L. M. *et al.* Th2 responses are primed by skin dendritic cells with distinct transcriptional profiles. *J. Exp. Med.* **214**, 125–142 (2017).
140. Janss, T. *et al.* Interferon response factor-3 promotes the pro-Th2 activity of mouse lung CD11b+ conventional dendritic cells in response to house dust mite allergens. *Eur. J. Immunol.* **46**, 2614–2628 (2016).
141. Steel, N. *et al.* TGFβ-activation by dendritic cells drives Th17 induction and intestinal contractility and augments the expulsion of the parasite Trichinella spiralis in mice. *PLoS Pathog.* **15**, 1–25 (2019).
142. Brown, C. C. *et al.* Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell* **179**, 846-863.e24 (2019).
143. Bourdely, P. *et al.* Transcriptional and Functional Analysis of CD1c+ Human Dendritic Cells Identifies a CD163+ Subset Priming CD8+CD103+ T Cells. *Immunity* **53**, 335-352.e8 (2020).
144. Dutertre, C. A. *et al.* Single-Cell Analysis of Human Mononuclear Phagocytes Reveals

Subset-Defining Markers and Identifies Circulating Inflammatory Dendritic Cells. *Immunity* **51**, 573-589.e8 (2019).
145. Villani, A. C. *et al.* Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science (80-.).* **356**, (2017).
146. Cytlak, U. *et al.* Differential IRF8 Transcription Factor Requirement Defines Two Pathways of Dendritic Cell Development in Humans. *Immunity* **53**, 353-370.e8 (2020).
147. Karsunky, H., Merad, M., Cozzio, A., Weissman, I. L. & Manz, M. G. Flt3 ligand regulates dendritic cell development from Flt3+ lymphoid and myeloid-committed progenitors to Flt3+ dendritic cells in vivo. *J. Exp. Med.* **198**, 305–313 (2003).
148. Segura, E. *et al.* Human Inflammatory Dendritic Cells Induce Th17 Cell Differentiation. *Immunity* **38**, 336–348 (2013).
149. Miller, J. C. *et al.* Deciphering the transcriptional network of the dendritic cell lineage. *Nat. Immunol.* **13**, 888–899 (2012).
150. Ardouin, L. *et al.* Broad and Largely Concordant Molecular Changes Characterize Tolerogenic and Immunogenic Dendritic Cell Maturation in Thymus and Periphery. *Immunity* **45**, 305–318 (2016).
151. Swiecki, M. & Colonna, M. The multifaceted biology of plasmacytoid dendritic cells. *Nat. Rev. Immunol.* **15**, 471–485 (2015).
152. Musumeci, A., Lutz, K., Winheim, E. & Krug, A. B. What makes a PDC: Recent advances in understanding plasmacytoid DC development and heterogeneity. *Front. Immunol.* **10**, 1–7 (2019).
153. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious disease. *Nat. Rev. Immunol.* **15**, 87–103 (2015).
154. Gough, D. J., Messina, N. L., Clarke, C. J. P., Johnstone, R. W. & Levy, D. E. Constitutive Type I Interferon Modulates Homeostatic Balance through Tonic Signaling. *Immunity* **36**, 166–174 (2012).
155. See, P. *et al.* Mapping the human DC lineage through the integration of high-dimensional techniques. *Science (80-.).* **356**, (2017).
156. Gabanyi, I. *et al.* Neuro-immune Interactions Drive Tissue Programming in Intestinal Macrophages. *Cell* **164**, 378–391 (2016).
157. Platt, A. M., Bain, C. C., Bordon, Y., Sester, D. P. & Mowat, A. M. An Independent Subset of TLR Expressing CCR2-Dependent Macrophages Promotes Colonic Inflammation. *J. Immunol.* **184**, 6843–6854 (2010).
158. Smythies, L. E. *et al.* Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J. Clin. Invest.* **115**, 66–75 (2005).
159. Roberts, P. J. *et al.* The physiological expression of inducible nitric oxide synthase (iNOS) in the human colon. *J. Clin. Pathol.* **54**, 293–297 (2001).
160. Hoshi, N. *et al.* MyD88 signalling in colonic mononuclear phagocytes drives colitis in IL-10-deficient mice. *Nat. Commun.* **3**, (2012).
161. Shouval, D. S. *et al.* Interleukin-10 receptor signaling in innate immune cells regulates mucosal immune tolerance and anti-inflammatory macrophage function. *Immunity* **40**, 706–719 (2014).
162. Zigmond, E. *et al.* Macrophage-restricted interleukin-10 receptor deficiency, but not IL-10 deficiency, causes severe spontaneous colitis. *Immunity* **40**, 720–733 (2014).
163. Glocker, E.-O. *et al.* Inflammatory Bowel Disease and Mutations Affecting the Interleukin-10 Receptor. *N. Engl. J. Med.* **361**, 2033–2045 (2009).

164. De Schepper, S. *et al.* Self-Maintaining Gut Macrophages Are Essential for Intestinal Homeostasis. *Cell* **175**, 400-415.e13 (2018).
165. Chiaranunt, P., Tai, S. L., Ngai, L. & Mortha, A. Beyond Immunity: Underappreciated Functions of Intestinal Macrophages. *Front. Immunol.* **12**, 1–22 (2021).
166. Asano, K. *et al.* Intestinal CD169 + macrophages initiate mucosal inflammation by secreting CCL8 that recruits inflammatory monocytes. *Nat. Commun.* **6**, 1–7 (2015).
167. Honda, M. *et al.* Perivascular localization of macrophages in the intestinal mucosa is regulated by Nr4a1 and the microbiome. *Nat. Commun.* **11**, 1329 (2020).
168. Stephan, A. H., Barres, B. A. & Stevens, B. The complement system: An unexpected role in synaptic pruning during development and disease. *Annu. Rev. Neurosci.* **35**, 369–389 (2012).
169. Shaw, T. N. *et al.* Tissue-resident macrophages in the intestine are long lived and defined by Tim-4 and CD4 expression. **215**, 1507–1518 (2018).
170. Ferraboschi, P., Ciceri, S. & Grisenti, P. Applications of lysozyme, an innate immune defense factor, as an alternative antibiotic. *Antibiotics* **10**, 1–55 (2021).
171. Bonnardel, J. *et al.* Innate and Adaptive Immune Functions of Peyer's Patch Monocyte-Derived Cells. *Cell Rep.* **11**, 770–784 (2015).
172. Jinnohara, T. *et al.* IL-22BP dictates characteristics of Peyer's patch follicleassociated epithelium for antigen uptake. *J. Exp. Med.* **214**, 1607–1618 (2017).
173. Nishi, C., Toda, S., Segawa, K. & Nagata, S. Tim4- and MerTK-Mediated Engulfment of Apoptotic Cells by Mouse Resident Peritoneal Macrophages. *Mol. Cell. Biol.* **34**, 1512–1520 (2014).
174. Miyanishi, M. *et al.* Identification of Tim4 as a phosphatidylserine receptor. *Nature* **450**, 435–439 (2007).
175. Wagner, C. *et al.* Some news from the unknown soldier, the Peyer's patch macrophage. *Cell. Immunol.* **330**, 159–167 (2018).
176. Bonnardel, J. *et al.* Gene expression profiling of the Peyer's patch mononuclear phagocyte system. *Genomics Data* **5**, 21–24 (2015).
177. Bonnardel, J. *et al.* Distribution, location, and transcriptional profile of Peyer's patch conventional DC subsets at steady state and under TLR7 ligand stimulation. *Mucosal Immunol.* **10**, 1412–1430 (2017).
178. Luciani, C., Hager, F. T., Cerovic, V. & Lelouard, H. Dendritic cell functions in the inductive and effector sites of intestinal immunity. *Mucosal Immunol.* **15**, 40–50 (2022).
179. Watchmaker, P. B. *et al.* Comparative transcriptional and functional profiling defines conserved programs of intestinal DC differentiation in humans and mice. **15**, (2014).
180. Scott, C. L. *et al.* CCR2 + CD103 - Intestinal dendritic cells develop from DC-committed precursors and induce interleukin-17 production by T cells. *Mucosal Immunol.* **8**, 327–339 (2015).
181. M, R. *et al.* Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nat. Immunol.* **2**, 361–7 (2001).
182. Farache, J. *et al.* Luminal Bacteria Recruit CD103+ Dendritic Cells into the Intestinal Epithelium to Sample Bacterial Antigens for Presentation. *Immunity* **38**, 581–595 (2013).
183. Mazzini, E., Massimiliano, L., Penna, G. & Rescigno, M. Oral Tolerance Can Be Established via Gap Junction Transfer of Fed Antigens from CX3CR1+ Macrophages to CD103+ Dendritic Cells. *Immunity* **40**, 248–261 (2014).

184. McDole, J. R. *et al.* Goblet cells deliver luminal antigen to CD103 + dendritic cells in the small intestine. *Nature* **483**, 345–349 (2012).
185. Cerovic, V. *et al.* Lymph-borne CD8α+ dendritic cells are uniquely able to cross-prime CD8+ T cells with antigen acquired from intestinal epithelial cells. *Mucosal Immunol.* **8**, 38–48 (2015).
186. Joeris, T. *et al.* Intestinal cDC1 drive cross-tolerance to epithelial-derived antigen via induction of FoxP3+CD8+ Tregs. *Sci. Immunol.* **6**, (2021).
187. Luda, K. M. *et al.* IRF8 Transcription-Factor-Dependent Classical Dendritic Cells Are Essential for Intestinal T Cell Homeostasis. *Immunity* **44**, 860–874 (2016).
188. Cerovic, V. *et al.* Intestinal CD103- dendritic cells migrate in lymph and prime effector T cells. *Mucosal Immunol.* **6**, 104–113 (2013).
189. Welty, N. E. *et al.* Intestinal lamina propria dendritic cells maintain T cell homeostasis but do not affect commensalism. *J. Exp. Med.* **210**, 2011–2024 (2013).
190. Persson, E. K., Scott, C. L., Mowat, A. M. & Agace, W. W. Immunity at the Barrier Review Series Dendritic cell subsets in the intestinal lamina propria : Ontogeny and function. 3098–3107 (2013) doi:10.1002/eji.201343740.
191. Persson, E. K. *et al.* IRF4 Transcription-Factor-Dependent CD103+CD11b+ Dendritic Cells Drive Mucosal T Helper 17 Cell Differentiation. *Immunity* **38**, 958–969 (2013).
192. Mayer, J. U. *et al.* Different populations of CD11b+ dendritic cells drive Th2 responses in the small intestine and colon. *Nat. Commun.* **8**, 15820 (2017).
193. Becker, M. *et al.* Ontogenic, phenotypic, and functional characterization of XCR1+ dendritic cells leads to a consistent classification of intestinal dendritic cells based on the expression of XCR1 and SIRPα. *Front. Immunol.* **5**, 1–12 (2014).
194. Jaensson, E. *et al.* Small intestinal CD103+ dendritic cells display unique functional properties that are conserved between mice and humans. *J. Exp. Med.* **205**, 2139–49 (2008).
195. Iwata, M. *et al.* Retinoic acid imprints gut-homing specificity on T cells. *Immunity* **21**, 527–538 (2004).
196. Johansson-Lindbom, B. *et al.* Selective generation of gut tropic T cells in gut-associated lymphoid tissue (GALT): Requirement for GALT dendritic cells and adjuvant. *J. Exp. Med.* **198**, 963–969 (2003).
197. Reboldi, A. *et al.* Mucosal immunology: IgA production requires B cell interaction with subepithelial dendritic cells in Peyer's patches. *Science (80-.).* **352**, (2016).
198. Flores-Langarica, A. *et al.* CD103 + CD11b + mucosal classical dendritic cells initiate long-term switched antibody responses to flagellin. *Mucosal Immunol.* **11**, 681–692 (2018).
199. Nakawesi, J. *et al.* αvβ8 integrin-expression by BATF3-dependent dendritic cells facilitates early IgA responses to Rotavirus. *Mucosal Immunol.* **14**, 53–67 (2021).
200. Pantazi, E. *et al.* Cutting Edge: Retinoic Acid Signaling in B Cells Is Essential for Oral Immunization and Microflora Composition. *J. Immunol.* **195**, 1368–1371 (2015).
201. Goverse, G. *et al.* Diet-Derived Short Chain Fatty Acids Stimulate Intestinal Epithelial Cells To Induce Mucosal Tolerogenic Dendritic Cells. *J. Immunol.* **198**, 2172–2181 (2017).
202. Wu, W. *et al.* Microbiota metabolite short-chain fatty acid acetate promotes intestinal IgA response to microbiota which is mediated by GPR43. *Mucosal Immunol.* **10**, 946–956 (2017).
203. Kanamori, Y. *et al.* Identification of novel lymphoid tissues in murine intestinal

mucosa where clusters of c-kit+ IL-7R+ Thy1+ lympho-hemopoietic progenitors develop. *J. Exp. Med.* **184**, 1449–1459 (1996).
204. Guendel, F. *et al.* Group 3 Innate Lymphoid Cells Program a Distinct Subset of IL-22BP-Producing Dendritic Cells Demarcating Solitary Intestinal Lymphoid Tissues. *Immunity* **53**, 1015-1032.e8 (2020).
205. LeBleu, V. S. & Neilson, E. G. Origin and functional heterogeneity of fibroblasts. *FASEB J.* **34**, 3519–3536 (2020).
206. Koliaraki, V., Prados, A., Armaka, M. & Kollias, G. The mesenchymal context in inflammation, immunity and cancer. *Nat. Immunol.* **21**, 974–982 (2020).
207. Roulis, M. *et al.* Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche. *Nature* **580**, 524–529 (2020).
208. Hong, S. P. *et al.* Distinct fibroblast subsets regulate lacteal integrity through YAP/TAZ-induced VEGF-C in intestinal villi. *Nat. Commun.* **11**, (2020).
209. McCarthy, N. *et al.* Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient. *Cell Stem Cell* **26**, 391-402.e5 (2020).
210. Bergers, G. & Song, S. The role of pericytes in blood-vessel formation and maintenance. *Neuro. Oncol.* **7**, 452–464 (2005).
211. Guimarães-Camboa, N. *et al.* Pericytes of Multiple Organs Do Not Behave as Mesenchymal Stem Cells In Vivo. *Cell Stem Cell* **20**, 345-359.e5 (2017).
212. Benabid, A. & Peduto, L. Mesenchymal perivascular cells in immunity and disease. *Curr. Opin. Immunol.* **64**, 50–55 (2020).
213. Kinchen, J. *et al.* Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. *Cell* **175**, 372-386.e17 (2018).
214. Furuya, S. & Furuya, K. Subepithelial Fibroblasts in Intestinal Villi: Roles in Intercellular Communication. *Int. Rev. Cytol.* **264**, 165–223 (2007).
215. Thomson, C. A., Nibbs, R. J., McCoy, K. D. & Mowat, A. M. Immunological roles of intestinal mesenchymal cells. *Immunology* **160**, 313–324 (2020).
216. Aoki, R. *et al.* Foxl1-Expressing Mesenchymal Cells Constitute the Intestinal Stem Cell Niche. *Cmgh* **2**, 175–188 (2016).
217. Greicius, G. *et al.* PDGFRα+ pericryptal stromal cells are the critical source of Wnts and RSPO3 for murine intestinal stem cells in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3173–E3181 (2018).
218. Jardé, T. *et al.* Mesenchymal Niche-Derived Neuregulin-1 Drives Intestinal Stem Cell Proliferation and Regeneration of Damaged Epithelium. *Cell Stem Cell* **27**, 646-662.e7 (2020).
219. Bahar Halpern, K. *et al.* Lgr5+ telocytes are a signaling source at the intestinal villus tip. *Nat. Commun.* **11**, 3–14 (2020).
220. Popescu, L. M. & Faussone-Pellegrini, M. S. TELOCYTES - a case of serendipity: The winding way from Interstitial Cells of Cajal (ICC), via Interstitial Cajal-Like Cells (ICLC) to TELOCYTES. *J. Cell. Mol. Med.* **14**, 729–740 (2010).
221. 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).
222. Lee, M. Y. *et al.* Transcriptome of interstitial cells of Cajal reveals unique and selective gene signatures. *PLoS One* **12**, 1–25 (2017).
223. Mantani, Y. *et al.* Ultrastructural and Immunohistochemical Study on the Lamina Propria Cells Beneath Paneth Cells in the Rat Ileum. *Anat. Rec.* **301**, 1074–1085

(2018).
224. Kurahashi, M. *et al.* Novel Population of Subepithelial Platelet-Derived Growth Factor Receptor A-Positive Cells in the Mouse and Human Colon. *Am. J. Physiol. - Gastrointest. Liver Physiol.* **304**, 823–834 (2013).
225. 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).
226. Thomson, C. A. *et al.* Expression of the Atypical Chemokine Receptor ACKR4 Identifies a Novel Population of Intestinal Submucosal Fibroblasts That Preferentially Expresses Endothelial Cell Regulators. *J. Immunol.* **201**, 215–229 (2018).
227. David, M. B., Valenta, T., Fazilaty, H., Hausmann, G. & Basler, K. Distinct populations of crypt-associated fibroblasts act as signaling hubs to control colon homeostasis. *PLoS Biol.* **18**, 1–20 (2020).
228. Stzepourginski, I. *et al.* CD34+ mesenchymal cells are a major component of the intestinal stem cells niche at homeostasis and after injury. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E506–E513 (2017).
229. Molotkov, A., Mazot, P., Brewer, J. R., Cinalli, R. M. & Soriano, P. Distinct Requirements for FGFR1 and FGFR2 in Primitive Endoderm Development and Exit from Pluripotency. *Dev. Cell* **41**, 511-526.e4 (2017).
230. Vanuytsel, T., Senger, S., Fasano, A. & Shea-Donohue, T. Major signaling pathways in intestinal stem cells. *Biochim. Biophys. Acta - Gen. Subj.* **1830**, 2410–2426 (2013).
231. Gómez-Orte, E., Sáenz-Narciso, B., Moreno, S. & Cabello, J. Multiple functions of the noncanonical Wnt pathway. *Trends Genet.* **29**, 545–553 (2013).
232. Kosinski, C. *et al.* Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15418–15423 (2007).
233. 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).
234. Ruchti, C., Haller, D., Nuber, M. & Cottier, H. Regional differences in renewal rates of fibroblasts in young adult female mice. *Cell Tissue Res.* **232**, 625–636 (1983).
235. Worthley, D. L. *et al.* Gremlin 1 identifies a skeletal stem cell with bone, cartilage, and reticular stromal potential. *Cell* **160**, 269–284 (2015).
236. Crisan, M. *et al.* A Perivascular Origin for Mesenchymal Stem Cells in Multiple Human Organs. *Cell Stem Cell* **3**, 301–313 (2008).
237. Sitnik, K. M. *et al.* Context-Dependent Development of Lymphoid Stroma from Adult CD34+ Adventitial Progenitors. *Cell Rep.* **14**, 2375–2388 (2016).
238. Benias, P. C. *et al.* Structure and distribution of an unrecognized interstitium in human tissues. *Sci. Rep.* **8**, 1–8 (2018).
239. Merrick, D. *et al.* Identification of a mesenchymal progenitor cell hierarchy in adipose tissue. *Science (80-.).* **364**, (2019).
240. Wörsdörfer, P. *et al.* The vascular adventitia: An endogenous, omnipresent source of stem cells in the body. *Pharmacol. Ther.* **171**, 13–29 (2017).
241. Kramann, R. *et al.* Adventitial MSC-like Cells Are Progenitors of Vascular Smooth Muscle Cells and Drive Vascular Calcification in Chronic Kidney Disease. *Cell Stem Cell* **19**, 628–642 (2016).
242. Kramann, R. *et al.* Perivascular Gli1+ progenitors are key contributors to injury-

induced organ fibrosis. *Cell Stem Cell* **16**, 51–66 (2015).
243. Corselli, M. *et al.* The tunica adventitia of human arteries and veins as a source of mesenchymal stem cells. *Stem Cells Dev.* **21**, 1299–1308 (2012).
244. Cheng, H. W. *et al.* Origin and differentiation trajectories of fibroblastic reticular cells in the splenic white pulp. *Nat. Commun.* **10**, (2019).
245. 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).
246. Miyoshi, H., Ajima, R., Luo, C. T., Yamaguchi, T. P. & Stappenbeck, T. S. Wnt5a potentiates TGF-beta SIgnaling to promote colonic crypt regeneration after injury. *Sci. Reports* (2012) doi:10.1002/path.2474.
247. Koma, A. E. M. Clinical Medicine Insights : Gastroenterology Inflammatory Bowel Disease : An Expanding Global Health Problem. *Clin. Med. Insights Gastroenterol.* **6**, 33–47 (2013).
248. Mayo Clinic. Inflammatory bowel disease (IBD). https://www.mayoclinic.org/diseases-conditions/inflammatory-bowel-disease/symptoms-causes/syc-20353315 (2020).
249. De Lange, K. M. *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017).
250. Momozawa, Y. *et al.* IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. *Nat. Commun.* **9**, 1–18 (2018).
251. Digby-Bell, J. L., Atreya, R., Monteleone, G. & Powell, N. Interrogating host immunity to predict treatment response in inflammatory bowel disease. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 9–20 (2020).
252. Jones, G. *et al.* Dynamics of Colon Monocyte and Macrophage Activation During Colitis. **9**, (2018).
253. Chapuy, L. *et al.* Two distinct colonic CD14+ subsets characterized by single-cell RNA profiling in Crohn's disease. *Mucosal Immunol.* **12**, 703–719 (2019).
254. Bernardo, D. *et al.* Human intestinal pro-inflammatory CD11chighCCR2+CX3CR1+ macrophages, but not their tolerogenic CD11c-CCR2-CX3CR1- counterparts, are expanded in inflammatory bowel disease article. *Mucosal Immunol.* **11**, 1114–1126 (2018).
255. Dige, A. *et al.* Reduced numbers of mucosal DRint macrophages and increased numbers of CD103+ dendritic cells during anti-TNF-α treatment in patients with Crohn's disease. *Scand. J. Gastroenterol.* **51**, 692–699 (2016).
256. Desalegn, G. & Pabst, O. Inflammation triggers immediate rather than progressive changes in monocyte differentiation in the small intestine. *Nat. Commun.* **10**, 3229 (2019).
257. Magnusson, M. K. *et al.* Macrophage and dendritic cell subsets in IBD: ALDH+ cells are reduced in colon tissue of patients with ulcerative colitis regardless of inflammation. *Mucosal Immunol.* **9**, 171–182 (2016).
258. 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).
259. Lissner, D. *et al.* Monocyte and M1 macrophage-induced barrier defect contributes to chronic intestinal inflammation in IBD. *Inflamm. Bowel Dis.* **21**, 1297–1305 (2015).

260. Imam, T., Park, S., Kaplan, M. H. & Olson, M. R. Effector T helper cell subsets in inflammatory bowel diseases. *Front. Immunol.* **9**, (2018).
261. Uzzan, M. *et al.* Ulcerative colitis is characterized by a plasmablast-skewed humoral response associated with disease activity. *Nat. Med.* (2022) doi:10.1038/s41591-022-01680-y.
262. Ley, K., Rivera-Nieves, J., Sandborn, W. J. & Shattil, S. Integrin-based therapeutics: Biological basis, clinical use and new drugs. *Nat. Rev. Drug Discov.* **15**, 173–183 (2016).
263. 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).
264. Hart, A. L. *et al.* Characteristics of intestinal dendritic cells in inflammatory bowel diseases. *Gastroenterology* **129**, 50–65 (2005).
265. Matsuno, H. *et al.* CD103+ Dendritic Cell Function Is Altered in the Colons of Patients with Ulcerative Colitis. *Inflamm. Bowel Dis.* **23**, 1524–1534 (2017).
266. Ahluwalia, B., Moraes, L., Magnusson, M. K. & Öhman, L. Immunopathogenesis of inflammatory bowel disease and mechanisms of biological therapies. *Scand. J. Gastroenterol.* **53**, 379–389 (2018).
267. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
268. Cusanovich, D. A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science (80-.).* **348**, 910–914 (2015).
269. Smallwood, S. A. *et al.* Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
270. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
271. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 1–12 (2018).
272. Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat. Protoc.* **13**, 599–604 (2018).
273. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2012).
274. Jiang, R., Sun, T., Song, D. & Li, J. J. Statistics or biology: the zero-inflation controversy about scRNA-seq data. *Genome Biol.* **23**, 1–24 (2022).
275. Payne, S. H. The utility of protein and mRNA correlation. *Trends Biochem. Sci.* **40**, 1–3 (2015).
276. Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J. & Smith, V. A. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci. Rep.* **5**, 1–9 (2015).
277. Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput. Biol.* **14**, (2018).
278. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).
279. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
280. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

281. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
282. Li, Jun; Barron, M. ccRemover: Removes the Cell-Cycle Effect from Single-Cell RNA-Sequencing Data. *CRAN* (2017).
283. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
284. Thi, H. *et al.* A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* 1–32 (2020).
285. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, (2019).
286. Kiselev, V. Y., Andrews, T. S. & Hemberg, M. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* **20**, 273–282 (2019).
287. Hu, Z. *et al.* MetaCyto: A Tool for Automated Meta-analysis of Mass and Flow Cytometry Data. *Cell Rep.* **24**, 1377–1388 (2018).
288. Robinson, M. D. *et al.* CyTOF workflow: Differential discovery in high-throughput high-dimensional cytometry datasets. *F1000Research* **6**, 1–69 (2017).
289. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
290. Mulè, M. P., Martins, A. J. & Tsang, J. S. Normalizing and denoising protein expression data from droplet-based single cell profiling. (2020).
291. Kharchenko, P. V. The triumphs and limitations of computational methods for scRNA-seq. *Nat. Methods* **18**, 723–732 (2021).
292. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
293. Brulois, K. magicBatch: R wrapper for the original python implementation of the Marcov Affinity-based Graph Imputation of Cells (MAGIC) algorithm. R package version 0.1.0. *github* (2020).
294. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat. Methods* **18**, 9–14 (2021).
295. Cabeza-Cabrerizo, M. *et al.* Tissue clonality of dendritic cell subsets and emergency DCpoiesis revealed by multicolor fate mapping of DC progenitors. *Sci. Immunol.* **4**, 1–14 (2019).
296. See, P. *et al.* Mapping the human DC lineage through the integration of high-dimensional techniques. *Science (80-.).* **356**, eaag3009 (2017).
297. Triana, S., Vonficht, D., Jopp-saile, L., Raffel, S. & Lutz, R. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* **22**, 1577–1589 (2021).
298. Cytlak, U. *et al.* Differential IRF8 Transcription Factor Requirement Defines Two Pathways of Dendritic Cell Development in Humans. *Immunity* **53**, 353-370.e8 (2020).
299. Fenton, T. M. *et al.* Immune Profiling of Human Gut-Associated Lymphoid Tissue Identifies a Role for Isolated Lymphoid Follicles in Priming of Region-Specific Immunity. *Immunity* **52**, 557-570.e6 (2020).
300. Williams, M. R., Sakurai, Y., Zughaier, S. M., Eskin, S. G. & McIntire, L. V. Transmigration across activated endothelium induces transcriptional changes, inhibits apoptosis, and decreases antimicrobial protein expression in human monocytes. *J. Leukoc. Biol.* **86**, 1331–1343 (2009).