

**THE PRIMARY CILIA OF THE GASTROINTESTINAL TRACT  
IN HOMEOSTASIS AND DISEASE AT THE SINGLE-CELL  
LEVEL**

A THESIS SUBMITTED TO  
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE  
OF BILKENT UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR  
THE DEGREE OF  
MASTER OF SCIENCE  
IN  
MOLECULAR BIOLOGY AND GENETICS

By  
Deniz Esen  
July 2021

THE PRIMARY CILIA OF THE GASTROINTESTINAL TRACT IN  
HOMEOSTASIS AND DISEASE AT THE SINGLE-CELL LEVEL

By Deniz Esen

July 2021

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

\_\_\_\_\_  
Bahar Değirmenci Uzun (Advisor)

\_\_\_\_\_  
Serkan İsmail Göktuna

\_\_\_\_\_  
Nihal Terzi Çizmecioglu

Approved for the Graduate School of Engineering and Science:

\_\_\_\_\_  
Ezhan Karaşan ✓

Director of the Graduate School

# **ABSTRACT**

## **The Primary Cilia of the Gastrointestinal Tract in Homeostasis and Disease at the Single-Cell Level**

**Deniz Esen**

**M.Sc. in Molecular Biology and Genetics**

**Advisor: Bahar Değirmenci Uzun**

**July 2021**

The gastrointestinal tract is regularly renewed by stem cells which divide and differentiate into functionally and morphologically distinct cell types. Several key pathways, such as Wnt, Hedgehog and Bmp, regulate the cell fate. However, it remains a mystery how the associated signaling molecules are relayed between cells to coordinate stemness and differentiation cues. Primary cilia are small antenna-like organelles that harbors many receptors for these pathways. Here public single-cell RNA sequencing data is re-analyzed to show that primary cilia expression is heterogeneous in the intestinal mesenchyme and liver. Presence of primary cilia is also validated using immunofluorescence in the stroma and muscle cells of the mouse colon, using known markers of the primary cilia. *Acot7* is identified as a primary cilium associated marker and found to be expressed in myenteric ganglia. In mice challenged with DSS to model ulcerative colitis, primary cilia are observed more abundant as the area covered by crypt structures become reduced due to the loss of epithelium. *Acot7* expressing ganglial cells were observed more frequently and displayed morphological differences. Additionally, mice fed a high-fat diet over 16 weeks had shortening of the colon crypts and an increase in the primary cilia. This work suggests that primary cilia exist in the gastrointestinal tract during homeostasis and participate in inflammation and diet-based adaptations.

*Keywords:* colon, primary cilia, *Acot7*, DSS, high-fat diet, myenteric ganglion, immunofluorescence, scRNA-seq, small intestine, liver

# ÖZET

## Homeostaz ve Hastalıkta Tek Hücre Düzeyinde Gastrointestinal Sistemin Primer Silyaları

Deniz Esen

Moleküler Biyoloji ve Genetik, Yüksek Lisans

Tez Danışmanı: Bahar Değirmenci Uzun

Temmuz 2021

Gastrointestinal sistem, kök hücrelerin bölünerek fonksiyonel ve morfolojik olarak farklı hücre tiplerine dönüşmesiyle düzenli olarak yenilenir. Hücre kaderi, Wnt, Hedgehog ve Bmp gibi birkaç önemli sinyal yolağı aracılığıyla belirlenir. Ancak kök hücre karakteri ve farklılaşma işaretlerinin koordinasyonu için bu sinyal moleküllerinin hücreler arasında nasıl iletildiği halen bir sırdır. Primer silyalar, bu sinyal yolakları için birçok reseptör barındıran anten benzeri küçük organellerdir. Bu çalışmada, umuna açılmış tek hücreli RNA dizileme verileri yeniden analiz edilerek primer silya ekspresyonunun bağırsak mezenkiminde ve karaciğerde heterojen olduğu gösterilmektedir. Primer silyaların varlığı bilinen belirteçleri kullanılarak immünfloresan tekniği ile fare kalın bağırsağının stroma ve kas hücrelerinde doğrulanmaktadır. Acot7 primer silyalarla ilişkili işaretleyici olarak ortaya çıkmaktadır ve miyenterik gangliyonlarda eksprese edildiği tespit edilmiştir. Ülseratif koliti modellemek için DSS verilen farelerde, epitel kaybı nedeniyle kript yapılarının kapladığı alan küçüldükçe primer silyaların daha bol olduğu gözlemlenmiştir. Acot7 eksprese eden gangliyal hücreler daha sık ortaya çıkmış ve morfolojik farklılıklar sergilemiştir. Ek olarak, 16 hafta boyunca yüksek yağlı diyetle beslenen farelerde, kolon kriptlerinde kısalma ve primer silyalarda bir artış görülmüştür. Bu çalışma, homeostaz sırasında gastrointestinal sistemde primer silyaların bulunduğunu, ayrıca bunların iltihaplanma ve diyetle dayalı adaptasyonlara katıldığını göstermektedir.

*Anahtar sözcükler:* kalın bağırsak, primer silya, Acot7, DSS, yüksek yağlı diyet, miyenterik gangliyon, immünfloresan, tek hücre RNA sekanslama, ince bağırsak, karaciğer

## Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Bahar Degirmenci Uzun, for leading and supervising this work. Her creativity and enthusiasm were truly inspiring from start to finish. I feel privileged to be one of the first members of her lab, and the first student to graduate. I appreciate her encouragement and understanding through the best and worst times of my journey.

I would like to extend my best regards to Dr. Nihal Terzi Çizmecioglu and Dr. Serkan İsmail Göktuna for being a part of my thesis committee and their feedback for my thesis work.

I would also like to show gratitude to Prof. Dr. Petek Korkusuz for making it possible for us to conveniently prepare our mouse tissue samples.

I am grateful to have had the opportunity to work with Nagihan Gizay Gönüllü and Müge Bozlar at the BD Lab. I could not have asked for more devoted and skilled colleagues. I am thankful for all their support, help and input to my work.

I also owe thanks to my partner, Abbas Güven Akçay, for his constant support. Having him in my life encouraged me to continue forward to my best abilities in countless situations. Thank you for all the times we shared together and will share in the future.

I extend my thanks to my parents, Fügen and Melih Esen, who have supported me patiently for my whole life to this point. I would not have been able to receive the education I did and become the person that I am without them.

Last but not least, I would like to acknowledge all the mice sacrificed for this study. I hope our culturing methods will improve enough in my lifetime to eliminate the necessity for their use.

This work was funded by EMBO installation grant and TÜBİTAK 1001 grant 219Z163 awarded to Dr. Bahar Degirmenci Uzun.

## **Attributions**

The data presented in this thesis is the result of my own work with the exceptions listed here.

The primary cilia associated marker lists generated by scRNA-seq re-analysis were studied by Nagihan Gizay Gönüllü and Müge Bozlar to select candidates for intestines and liver, respectively.

Figure S1 was generated by Ece Seyrek. Figure 4.7 was generated by Burak Köksal.

Mouse transcardial perfusion was performed by Müge Bozlar.

Mice used for the HFD model were fed and monitored by Nagihan Gizay Gönüllü.

## Table of Contents

Acknowledgements.....	v
Attributions .....	vi
Table of Contents.....	vii
List of Figures .....	xi
List of Tables .....	xiii
Abbreviations.....	xiv
1.INTRODUCTION .....	1
1.1 Anatomy of the Intestines .....	1
1.1.1 Homeostasis and Cell Turnover of The Intestinal Epithelium.....	1
1.1.2 Stem Cell Recovery Upon Injury .....	2
1.1.3 Regulatory Signaling Pathways .....	3
1.1.3.1 Wnt.....	3
1.1.3.2 Hedgehog .....	4
1.1.3.3 Bone Morphogenic Proteins (BMPs).....	4
1.1.3.4 Notch.....	5
1.1.3.5 Platelet-derived Growth Factor A .....	5
1.1.3.6 Hippo.....	6
1.2 Primary Cilia.....	6
1.2.1 Signaling Pathways and Primary Cilia .....	7
1.3 Enteric Nervous System.....	8
1.3.1 Anatomy of the ENS .....	8
1.4 Diseases of The Intestines.....	10
1.4.1 ENS and Disease.....	10
1.4.2 Inflammatory Bowel Disease.....	10
1.4.2.1 Ulcerative Colitis .....	11
1.4.2.1.1 Dextran Sodium Sulfate Colitis Model.....	12

1.4.2.2 Crohn's Disease .....	13
1.4.3 Diet and Intestines.....	13
1.4.3.1 Fasting.....	13
1.4.3.1.1 Fasting Model of Mouse .....	14
1.4.3.2 Obesity .....	14
1.4.3.2.1 High-fat Diet Model.....	15
1.4.3.2.2 Western Diet Model.....	15
1.5 Liver.....	16
1.5.1 Liver, Diet and Disease.....	16
1.6 Single-cell RNA Sequencing .....	17
1.6.1 Public Datasets.....	18
1.6.1.1 Intestine scRNA-seq Data.....	18
1.6.1.2 Liver scRNA-seq Data.....	18
1.6.1.3 Primary Cilia in scRNA-seq Data.....	19
1.7 Discussion .....	19
2. AIMS OF THE STUDY .....	20
2.1 Characterization of the primary cilia in the colon, small intestine and liver .....	20
2.2 Identification of the marker genes for primary cilia expressing cells.....	20
2.3 Monitoring responses in the colon following diet and DSS challenges .....	20
3. MATERIALS AND METHODS.....	21
3.1 Re-analysis of public single-cell RNA sequencing datasets.....	21
3.2 Marker validation with mouse models.....	22
3.2.1 Establishment of mouse models.....	22
3.2.1.1 DSS challenge and recovery .....	23
3.2.1.2 Fasting and Refeeding.....	23
3.2.1.3 High-fat Diet .....	23
3.2.1.4 Western Diet .....	26



3.2.2 Marker Validation with Immunofluorochemistry .....	26
3.2.2.1 Mouse Colon Isolation .....	26
3.2.2.2 Tissue Fixation, Dehydration, Embedding and Sectioning .....	28
3.2.2.3 Immunofluorescent Staining .....	29
3.2.2.3.1 Deparaffinization and Rehydration .....	29
3.2.2.3.2 Antigen Retrieval .....	29
3.2.2.3.3 Blocking and Immunostaining .....	29
3.2.2.4 Imaging .....	30
3.2.2.4.1 Confocal Microscopy .....	30
3.2.2.4.2 Image Processing .....	31
4. RESULTS .....	35
4.1 Primary Cilia Marker Gene Expression .....	35
4.1.1 Primary Cilia in the Colon and Small Intestine .....	35
4.1.2 Primary Cilia in the Liver .....	41
4.2 Primary Cilia Associated Marker Discovery in the Colon Mesenchyme .....	44
4.3 Validation of the Primary Cilia in The Mouse Colon .....	48
4.3.1 Validation of Candidate Primary Cilia Associated Marker Acot7 .....	49
4.4 Mouse DSS-induced Colitis Model .....	49
4.4.1 Primary Cilia and Acot7 Response to DSS Challenge .....	50
4.5 Mouse Fasting and Refeeding Model .....	54
4.5.1 Changes to Primary Cilia and Acot7 in Fasted and Refed Mice .....	55
4.6 Mouse High-fat Diet Model .....	55
4.7 Mouse Western Diet Model .....	58
5. DISCUSSION & FUTURE PERSPECTIVES .....	59
5.1 Primary Cilia Locations and Abundance According to mRNA Expression Profiles .....	59
5.2. Markers of the Primary Cilia Expressing Cells .....	60

5.3 Primary Cilia in The Healthy Mouse Colon .....	61
5.3.1 Acot7 <sup>+</sup> Cells in the Mouse Colon .....	62
5.3.2 Prrx2 Expression in Colon Mesenchyme.....	63
5.4 DSS Challenge and Recovery Induced Changes to Primary Cilia and Acot7 Expression.....	63
5.5 Fasting and Refeeding Induced Changes to Primary Cilia and Acot7 Expression .....	64
5.6 High-fat Diet Induced Changes to the Colon.....	65
5.7 Western Diet Model.....	66
6. APPENDIX.....	67
A. Codes Used for scRNA-seq Data Analysis.....	67
B. Supplementary Figures.....	102
C. Supplementary Tables .....	113
7. REFERENCES.....	115

## List of Figures

1.1	Schematic anatomy of the colon.	22
4.1	ARL13B mRNA expression of biopsied human healthy, injured or uninjured colon mesenchyme	36
4.2	Arl13b mRNA expression of colon mesenchymal cells in healthy or DSS treated mice	38
4.3	Ift88 mRNA expression in colon mesenchyme of human and mouse	39
4.4	ARL13B mRNA expression of biopsied human colon epithelium	40
4.5	Arl13b mRNA expression in Pdgfra <sup>+</sup> cells from small intestine mesenchyme	41
4.6	ARL13B mRNA expression profile of human uninjured and cirrhotic liver biopsies	42
4.7	Arl13b mRNA expression profile in Pdgfrb <sup>+</sup> mouse liver mesenchyme	43
4.8	Acot7 mRNA expression in colon mesenchyme	46
4.9	Prrx2 mRNA expression in colon mesenchyme	47
4.10	Appearance of primary cilia around the colon crypts	48
4.11	Acot7 expression in the healthy mouse colon	49
4.12	Changes to primary cilia in the crypt top region DSS challenge	50
4.13	Changes to colon morphology and Acot7 expression following the DSS challenge	51
4.14	Primary cilia and Acot7 response to DSS challenge	52
4.15	Myenteric ganglia in DSS challenge and recovery	53
4.16	Cell proliferation in DSS challenge and recovery	54
4.17	Arl13b and Acot7 in colons of 24 hours fasted and refed mice	55
4.18	Crypt morphology and Ki67 expression after 16 weeks on a high-fat diet	57
4.19	Colonic crypt lengths in 16-week HFD mice and same aged controls	57
4.20	Primary cilia expression after 16 weeks on a high-fat diet	58

S1	Distribution of cell counts across clusters in human colon mesenchyme scRNA-seq data	102
S2	Arl13b mRNA expression in small intestine epithelium	103
S3	Arl13b mRNA expression profile in mouse liver endothelium and immune cells	103
S4	Subclusters generated by human liver scRNA-seq re-analysis	104
S5	Acot7 - Pdgfra mRNA co-localization in human colon mesenchyme	104
S6	Acot7 - Pdgfra mRNA co-localization in mouse colon mesenchyme	105
S7	Prrx2 - Pdgfra mRNA co-localization in human colon mesenchyme	105
S8	Prrx2 - Pdgfra mRNA co-localization in mouse colon mesenchyme	106
S9	Acot7 expression in mouse colon ENS	106
S10	Acot7 expression in mouse small intestine epithelium	107
S11	Mice colons after DSS challenge and recovery	108
S12	Acot7 and Arl13b expression after 3 and 5 days of recovery from DSS	108
S13	Primary cilia and Acot7 expression after 2 weeks on a high-fat diet	109
S14	Primary cilia expression after 8 weeks on a high-fat diet	110
S15	Primary cilia expression after 12 weeks on a high-fat diet	111
S16	Ki67 expression after 12 weeks on a high-fat diet	112

## List of Tables

3.1	List of publications scRNA-seq data was obtained and analyzed from	21
3.2	Weight chart of mice used in DSS experiments	24
3.3	Weight chart of mice used in fasting and refeeding experiments	25
3.4	Weight chart of mice used in HFD experiments	26
3.5	Weight chart of mice used in WD experiments	27
3.6	List of utilized reagents and resources	31
3.7	List of antibodies used in immunofluorochemistry and optimal working conditions	33
4.1	Number of cells in the human UC colon mesenchyme dataset	37
4.2	Number of cells in the mouse colon mesenchyme dataset	38
4.3	Number of cells in the human liver dataset, non-immune cells only	42
4.4	Top 30 most differentially expressed genes in the human colon mesenchyme WNT2B <sup>+</sup> Fos-hi cluster between healthy and inflamed datasets	45
S1	Colon crypt lengths after 16 weeks of HFD	113

## Abbreviations

Acot7	Acyl coenzyme A thioesterase 7
ALD	Alcoholic Liver Disease
Arl13b	ADP-ribosylation factor-like protein 13B
BMP	Bone morphogenic protein
CCl <sub>4</sub>	Carbon tetrachloride
CD	Crohn's Disease
ddH <sub>2</sub> O	double distilled water
dH <sub>2</sub> O	single distilled water
DSS	Dextran sodium sulfate
ENS	Enteric Nervous System
Gli1	Glioma-Associated Oncogene Homolog 1
HFD	High-fat diet
IBD	Inflammatory Bowel Disease
ICC	Interstitial cells of Cajal
Ift88	Intraflagellar transport protein 88
Lgr5	Leucine-rich repeat-containing G-protein coupled receptor 5
MAPK	Mitogen-activated protein kinase
mRNA	Messenger Ribonucleic Acid
NAFLD	Non-alcoholic Fatty Liver Disease
PBS	Phosphate Buffered Saline
PBS-T	PBS-Triton X
Pdgfra	Platelet-derived growth factor receptor A

Pdgfrb	Platelet-derived growth factor receptor B
PFA	Paraformaldehyde
Prrx2	Paired related homeobox 2
scRNA-seq	Single-cell Ribonucleic acid sequencing
Taz	PDZ binding motif
TGFβ	Tumor growth factor beta
UC	Ulcerative Colitis
v/v	volume (in milliliter) per total volume (milliliter)
w/v	weight (in grams) per total volume (milliliter)
WD	Western Diet
Yap1	Yes-associated protein 1

# **1. INTRODUCTION**

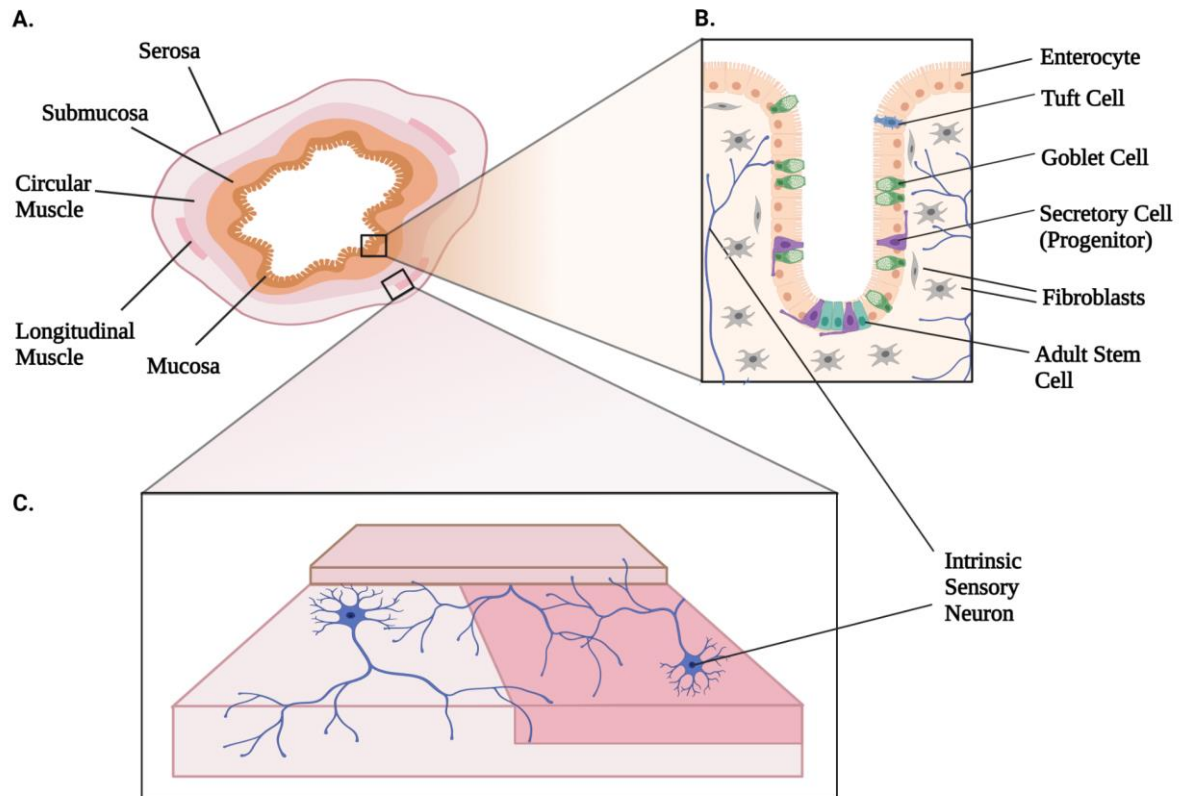
## **1.1 Anatomy of the Intestines**

The gastrointestinal tract is a continuous tube that consists of several tissue layers. The innermost monolayer is a sheet of epithelial cells that encloses the lumen, acting as a barrier between the body and the microbiota, while also performing the absorptive functions. A layer of mucus, produced by the epithelial layer, acts as a barrier between the intestinal contents, including the microbiota, and the epithelial cells. The microscopic architecture of the intestine consists of invaginations of the epithelial cell layer into the lamina propria, the inner layer of the mucosa. These structures, named crypts of Lieberkühn, house the stem cells at the bottom, and the differentiated cells towards the lumen. The small intestinal epithelium extends into the lumen in the form of protrusions called villi, which are absent in the colon. Instead, the colon lumen is encircled with a continuous, mostly flat surface epithelium barring the crypt mouths. The lamina propria underneath the epithelial monolayer consists of a heterogeneous population of cells which can be broadly classified into fibroblasts, immune cells, neurons and endothelial cells. These mucosal layers are encircled by muscular layers, which also house specialized structures, notably the enteric neurons located in the myenteric plexus between the circular and longitudinal muscle layers. Lastly, the mesothelial cell monolayer named serosa surrounds the entire outer side of the intestinal tube [1] (Fig. 1.1).

### **1.1.1 Homeostasis and Cell Turnover of The Intestinal Epithelium**

The different functions of the intestines are performed by a variety of specialized cells. All of these specialized cells arise from the stem cells, and they rapidly differentiate to cover the wide array of intestinal functions. The Lgr5<sup>+</sup> crypt base columnar cells are the stem cells that sit at the very bottom of the crypt both in the small intestine and in the colon [2]. These stem cells divide about once per day to be able to completely renew the epithelial layer every 3 to 5 days [1]. Paneth cells reside between the stem cells in the small intestine, but they are absent in the colon, and other secretory cells populate the crypt bottom in their stead [3]. Once an epithelial stem cell exits the niche, it starts losing its multipotency and starts differentiating into progenitors of different cell types. These progenitors show quick proliferation while they exist in the transit amplifying region directly above the intestinal stem cell niche, in the upper crypt region. All maturing cells travel away from the crypt and complete their differentiation





**Figure 1.1. Schematic anatomy of the colon.** A) Cross-section of the colon indicating the major layers. B) Close-up of the mucosal layers. The stem cells sit at the bottom of the crypt with progenitor cells in the proximity. The progenitor cells differentiate into the mature cell types, such as enterocytes, goblet cells and tuft cells. The epithelial layer is surrounded by fibroblasts and other cell types constituting the lamina propria, such as immune and endothelial cells (not shown). C) Close-up of the myenteric plexus located between the circular and longitudinal muscle layers. Neurons of the enteric nervous system are in this layer along the gut tube, with protrusions of the intrinsic sensory neurons extending into the lamina propria shown in B). Illustration created with BioRender.com.

with the exception of Paneth cells in the small intestine, which move back to the crypt bottom. Ultimately, the mature epithelial cells are shed into the lumen, and expire. These differentiation and self-renewal programs are under strict, coordinated control, and multiple signaling pathways supported by various cell types within the microenvironment play regulatory roles [1].

### 1.1.2 Stem Cell Recovery Upon Injury

The intestinal stem cells in the adult tissue can be lost and recovered from an injury. The crypt bottom cells can be replenished by other cell types of the epithelium via promotion of a more fetal-like transcription program [3]. Recovery of Lgr5 expressing cells appears as the key to restoration of the stem cell population. One of the populations able to dedifferentiate

into Lgr5<sup>+</sup> crypt base columnar cells are the secretory progenitors, previously dubbed as the reserve stem cells [1]. A more recent discovery defined a population of revival stem cells, which are Clusterin expressing multipotent cells. These cells are mostly absent under homeostasis, but they expand upon injury. This population was found to be able to give rise to new Lgr5<sup>+</sup> cells and re-establish the epithelial stem cell population [4]. Yap1 emerges as a key factor in incorporating chemical and mechanical cues from the microenvironment exclusively in the recovery state [5] and contributes to inducing Lgr5 expressing cell production from the revival stem cells [4].

### **1.1.3 Regulatory Signaling Pathways**

#### **1.1.3.1 Wnt**

Wnt molecules are a family of secreted proteins that are capable of inducing a plethora of signaling cascades crucial for many transcriptional programs. The Wnt pathways can be classified as  $\beta$ -catenin dependent (canonical) or independent (non-canonical). The pathways to be activated are dictated by the receptor availability in the cell. The non-canonical pathways are chiefly associated with Wnt3, Wnt5a and Wnt 11, and can be divided further into two categories. The first signaling cascade plays a crucial role in establishing polarity and is named planar cell polarity [2]. The other, named Wnt/calcium pathway, appears to contribute to cell migration during development, mediate inflammatory response and act as a proto-oncogene or tumor suppressor based on receptor composition of the cell [6]. On the other hand, canonical pathway activation depends on the Wnt ligand binding to a member of the Frizzled family of transmembrane receptors [2]. The Wnt agonists R-spondins help amplify Wnt signaling by stabilizing Frizzled [1,2]. Downstream of this activation,  $\beta$ -catenin escapes degradation and translocates into the nucleus. Finally, Wnt target genes are transcribed when  $\beta$ -catenin interacts with the necessary cofactors [2].

Canonical Wnt signaling is indispensable for intestinal stem cell renewal. In the small intestine, Paneth cells within the crypt can provide the Wnt molecules to the neighboring stem cells. Since Paneth cells are absent in the colon crypts, Wnt is sourced completely from outside the epithelial layer. It has been shown in both small intestine and colon that Gli1<sup>+</sup> mesenchymal cells provide Wnt to the crypt. They constitute the essential niche for Wnt signaling in the colon, and act as the reserve source in the small intestine in the event of epithelial injury [7]. Foxl1<sup>+</sup> mesenchymal cells also play an essential role in intestinal stem cell renewal, and a part

of the population likely overlaps with the Gli1<sup>+</sup> cells [2]. Another source for Wnt among the mesenchymal cells is the Pdgfra<sup>+</sup> population, as disruption of Wnt processing results in impaired crypt formation, and R-spondin 3 knockout in these cells increase epithelial damage susceptibility [2]. In colonic wound repair, Wnt5a has been identified as a crucial factor in suppressing proliferation and promoting a fetal-like state to allow recovery of the epithelial layer.

#### **1.1.3.2 Hedgehog**

Hedgehog signaling plays a key role in intestine organogenesis during early stages of development in terms of determining cellular identity as a cell communicator. It is also capable of promoting secretion of other signaling factors such as Wnts and BMPs. In adult intestine homeostasis, hedgehog signaling is significantly reduced, but still important to maintain the homeostasis and respond to injury. Hedgehog signaling appears as an important method of communication between the epithelial and mesenchymal cells, as the ligands are secreted from the epithelial cells and bound to receptors on primary cilia of mesenchymal cells [10]. Gli1 is a component of Hedgehog signaling, and an increase in the number of Sonic Hedgehog (Shh) ligands appears as a driver for Gli1<sup>+</sup> mesenchymal cell population expansion [7].

#### **1.1.3.3 Bone Morphogenic Proteins (BMPs)**

BMPs are a part of the tumor growth factor  $\beta$  (TGF $\beta$ ) family of signaling molecules [11]. In the intestines, BMP signaling establishes a gradient working opposite to Wnt signaling, promoting intestinal stem cell differentiation over proliferation [1]. Similar to Wnt, the pathway can be divided into two major branches. The canonical branch operates by enabling Smad translocation in the nucleus upon receptor activation by extracellular Bmp binding, leading to transcriptional regulation of Bmp target genes. This pathway can be interrupted by Bmp inhibitors, such as Gremlin1 and Noggin. Less is known about the non-canonical pathway, except for the contribution of other TGF $\beta$  family molecules, and consequent activation of the MAPK pathway [11].

In the small intestine, villus formation relies on communication between the epithelial and mesenchymal cells via the interplay of Bmp and Shh signaling [11]. Loss of Bmp signaling leads to increased stemness and ectopic crypt formation, as well as crypts forming on villi [1,12]. During development, Bmp establishes a gradient along the primitive gut tube to define the character of the gastrointestinal tract compartments and is required for the formation of the

intestines. Bmp molecules are more abundant towards the distal end of the primitive gut tube, which develops into the colon [11]. In the adult small intestine, it has been recently shown that Bmp gradient along the crypt-villus axis does not rely on Bmp diffusion alone but is coupled with Bmp inhibition by Gremlin1. Crypt bottom mesenchyme provides Gremlin1 to promote stemness, while Bmp pathway is more active towards the top of the crypt region and encourages differentiation of the epithelial cells [12].

In the adult colon, Bmp ligands appear as an anti-inflammatory agent in IBD. Myofibroblasts and smooth muscle cells supply the tissue with Bmp antagonists and expand if Bmp signaling is blocked in the mesenchyme. Immune cell populations also increase during the same blocking event [11].

#### **1.1.3.4 Notch**

Similar to Wnt, Notch ligands are promoters of stemness for the intestinal stem cells and are secreted by Paneth cells in the small intestine. Notch signaling is relayed directly over cell-to-cell contact, as the Lgr5<sup>+</sup> stem cells, direct neighbors of the Paneth cells, express Notch receptors [13]. Notch signal inhibition leads to stem cell loss, on the other hand, its amplification leads to an increase in the number of progenitor cells [1]. Due to the absence of Paneth cells in the colon, it is unclear how Notch signaling occurs. However, it has been observed that Notch signaling is activated in IBD patients in a manner correlating with epithelial regeneration and could also be playing a role in regulating the immune response. Most notably, overexpression of Notch has an oncogenic effect and has been considered a target for colorectal cancer treatments. All in all, the evidence suggests that Notch signaling is under strict control during homeostasis throughout the intestines [14].

#### **1.1.3.5 Platelet-derived Growth Factor A**

Pdgf-A is sensed by its transmembrane receptor, Pdgfra, which has gathered increasing attention as a marker for distinguishing mesenchymal cell populations in the intestines [1,7,12,15,16]. Pdgfra activation results in transcription of genes related to cell motility, viability, survival and proliferation. Traditionally, mutations in Pdgfra have been linked to cancers in a variety of organs and tissues, such as gastrointestinal stromal tumors and brain tumors. Pdgfra is essential for embryonic development, and knockout mice models were found not viable due to several defects in gastrointestinal tract, kidneys, nervous system and many

other tissues. In adult tissues, their main role is correlated with homeostasis maintenance and wound repair [17].

In intestinal lamina propria, *Pdgfra* is expressed in a majority of the cells at different levels and has been used for differentiating between populations. Highest *Pdgfra* expression is associated with fibroblasts closest to the lumen in both small intestine and colon. The expression decreases towards the crypt bottom [12,15,16]. *Pdgfra* expressing cells have been found to be the source of many signaling molecules crucial for homeostasis and injury response pathways, such as Wnts, R-spondin 3, *Gli1*, *Gremlin1*. The only known intestinal fibroblast population not expressing *Pdgfra* are the myofibroblasts [1].

#### **1.1.3.6 Hippo**

The Hippo pathway plays an important role in regulating embryonic development and regeneration upon injury in adult tissues. Its most central members, Yap/Taz, are transcriptional co-activators and can directly regulate gene expression. The Hippo pathway partially overlaps with the Wnt pathway due to Yap/Taz's presence in the  $\beta$ -catenin destruction complex. Signals from other pathways are also able to contribute to the Hippo pathway, giving it a signal network character instead of a cascade. Due to its potency in regulating stemness, organ growth and other fetal-like functions, Yap/Taz activation can lead to tissue and organ overgrowth and cancer [18].

As mentioned in section 1.1.3.1, Yap1 is present in low quantities in the intestine during homeostasis but emerges as a key factor in epithelium regeneration after injury, to the extent that its loss during homeostasis produces no phenotype, but severely impairs regeneration in an injured state [1]. Yap1 and Wnt signaling work in opposition to each other, where Yap1 promotes more fetal-like transcription programs, promotes cell survival and proliferation, instead of the self-renewal and differentiation program typical to adult stem cells [3,18]. Other than the Wnt signaling pathway, Yap1 is able to incorporate mechanical signaling that stems from the extracellular matrix remodeling subsequent to injury. When the recovery is complete, the fetal-like program promoted by Yap1 is switched with the regular, adult stem cell program [19].

### **1.2 Primary Cilia**

Primary cilia are small, non-motile, antenna-like organelles protruding from the cell membrane and can be found in most cell types in the human body. Only one primary cilium

exists per cell. It consists of 9 peripheral microtubule pairs built as an extension of the centriole, and a portion of the plasma membrane enclosing the axoneme [20,21]. Receptors for multiple signaling pathways can be found on the primary cilium membrane, implicating a crucial role as a relay structure in terms of controlling development, proliferation and stemness of the cell. Signaling pathways associated with primary cilia include Hedgehog, Wnt, Bmp, Notch, Pdgfra and Hippo. Aside from signaling proteins, primary cilia also act as a sensor for pH, Calcium, Oxygen, and other molecules, as well as light [21].

Primary cilia are disassembled and generated anew during the cell cycle. Prior to mitosis, centrosomes at the cell membrane are released, thus the primary cilia disappear. The mitotic spindles disassemble after mitosis, allowing centrosomes to attach back to the cell membrane and start ciliogenesis. The primary cilia are built out of the mother centriole, also called the basal body, and stabilized via post-translational modifications [22].

As an incorporator of many signals, defects surrounding primary cilia are associated with various disorders classified as ciliopathies. These diseases can affect a wide range of organs and systems in the mammalian body, such as kidneys and nervous system [21].

### **1.2.1 Signaling Pathways and Primary Cilia**

Major components of the Hedgehog pathway are situated on the primary cilium membrane. Receptor Patched (Ptch1) acts as an inhibitor of the Smoothened (Smo) protein and is expelled from the membrane upon Hedgehog signal initiation. Smo accumulates on the membrane and allows passage of Gli proteins into the cell and the nucleus [23].

Several Wnt pathway components are located on the primary cilia, however, the link between the primary cilia and Wnt signal transduction remains unclear. Several studies show that the primary cilia may be unnecessary for canonical or non-canonical Wnt signaling, however, some findings suggest that the primary cilia may downregulate Wnt signaling, or act as a switch between canonical and non-canonical pathway commitments [21].

Pdgfra can be found on the primary cilium membrane, although some cell types can have the receptor located elsewhere as well. Disruption of Pdgfra signaling has been linked to various chronic disorders, some of which coincides with disorders of the primary cilia [24]. Defective Pdgfra signaling has been linked to tumorigenesis in a variety of tissues, one of which promotes disassembly of the primary cilium, leading to uncontrolled cell division in gastrointestinal stromal tumors [21].

Notch signaling may also be dependent on the primary cilium due to the Notch transmembrane receptor's interaction with Calcium. Ciliary basal body defects can lead to defects in Notch signaling as well. In return, Notch signaling is important for regulating the cilia length, as well as promoting Hedgehog signaling along the primary cilia [25].

Primary cilia are a coordinator of Bmp signaling and possesses receptors for the TGF- $\beta$  family of molecules. Based on the concentration of related receptors, feedback signalization also appears to run along the ciliary axis. Therefore, primary cilia emerge as a key organelle in cell identity control, which is supported by the fact that ciliary defects can lead to cells transitioning into a mesenchymal profile [21].

The Hippo pathway not only has close links with primary cilia signaling as an incorporator of other related pathways, but also as a regulator of ciliogenesis. Moreover, the Hippo pathway works as a network capable of responding to environmental cues sensed by the primary cilium. Yap/Taz inhibits ciliogenesis, and ciliogenesis or cilia disassembly signals regulate Yap/Taz activity in return. However, it remains unclear whether Yap/Taz and primary cilia maintenance are linked correlatively or causatively [26].

### **1.3 Enteric Nervous System**

The gastrointestinal tract has its own nervous system independent of the central nervous system. The enteric nervous system governs the gut motility, secretion, blood flow, interacts with the microbiome, epithelium and immune system. The Enteric Nervous System (ENS) is therefore an important regulator of intestinal function and homeostasis [27].

#### **1.3.1 Anatomy of the ENS**

The ENS consists of multiple types of neurons. Their cell bodies surrounded by enteric glia comprise the ganglia within the myenteric and submucosal plexuses. Neurons constituting the ENS are highly heterogeneous. The diversification into distinct types of stems from differences in intestinal regions, species, sex, age and diseases [28]. The myenteric plexus is located within the muscle layer, between the longitudinal and circular muscles, which is a convenient position to be able to regulate muscle movements to control gut motility. The submucosal plexus lies above the muscle layer and underneath the epithelial layer. Neurons of the submucosal plexus contribute to regulating blood flow and secretions [27]. ENS works

together with enteroendocrine cells of the epithelium, which act as a link between the gut lumen and ENS [29].

There are three main categories of neurons in the ENS. First are intrinsic sensory neurons (IPANs)(Fig.1.1), which can both sense chemical and mechanical cues and excite other neurons. They are classified as after-hyperpolarization (AH) and Dogiel type II neurons due to their high action potential and long hyperpolarization time. IPANs envelope almost all the intestinal walls, extending into upper mucosal layers, and can release transmitters from all their extensions [27]. IPANs are a heterogeneous group of neurons that have different localization, different sensory targets and transcriptional profiles [27,28]. However, it is also suggested that not all of these groups should be considered as sensory neurons [30].

The next group are the interneurons, which exist only in the myenteric plexus. Their protrusions extend along the longitudinal axis of the intestines. There are two major subtypes according to the direction of their extensions: ascending or descending. Further diversifications occur according to the region of the gut. Descending interneurons create an exception to their localization by having projections into the submucosal ganglia [27].

Motor neurons, the last major group of ENS, only have ascending projections along the gut tube, but have vertical extensions into both surrounding muscle layers, as well as the submucosa. Specialized subtypes regulate muscle movements [27].

Glial cells often outnumber the enteric neurons within the myenteric ganglia and contribute to gut homeostasis [27,29]. Like the neurons, several subtypes of glia exist. However, there is yet to be a consensus on their classification [27-30].

Interstitial cells of Cajal (ICCs) are found in close proximity to the myenteric ganglia, either in the myenteric layer or in the muscle layers surrounding it. [31] They are derived from a mesenchymal lineage and act as the pacemaker of the gut, which initially led them to be classified as a neuron subtype. They interact with enteric neurons and smooth muscles [32]. ICC cells are also linked with  $\text{Pdgfra}^+$  cells that envelop myenteric ganglia. Due to this positioning, these  $\text{Pdgfra}^+$  cells likely contribute to the ENS network's communication with other surrounding cells. [31]

Several populations of enteric glial cells have been found to possess primary cilia. These populations also resemble the neural stem cells of the adult central nervous system, suggesting that they could be the source of neurons in the gastrointestinal tract [20].



## **1.4 Diseases of The Intestines**

The intestines are susceptible to various diseases as a complex tissue under pressure from internal and external factors. The mechanisms, severity, phenotype and area of effect changes depending on the initiating factors and affected areas [3].

### **1.4.1 ENS and Disease**

Due to the ENS' central role in coordinating events in the gastrointestinal tract, defects in the system or the associated tissues lead to damage and repair mechanism activation in the other. A classic example of a congenital disorder of ENS is Hirschsprung's disease, a condition caused by the absence of ganglia, typically in the distal portion of the intestines, and results in a lack of peristalsis of the gut muscles [29]. ICC networks are also affected in this disorder. [32].

### **1.4.2 Inflammatory Bowel Disease**

IBD is the general term for two similar diseases which stem from repetitive inflammation of the intestines. Ulcerative colitis (UC) occurs exclusively in the colon and the rectum, while in Crohn's disease (CD) the inflammation could be present anywhere in the gastrointestinal tract, but more commonly in the colon and the terminal ileum. Both forms of IBD are associated with defects of intestinal barrier integrity, which can be linked to genetic and environmental factors. IBD is a condition that lowers quality of life, increases mortality and risk of cancer in the long term [33].

IBDs have a strong immune component, but there is increasing evidence that ENS is also a contributor. During infection, neuron cells and fibers undergo physical changes such as hypertrophy and retraction, and show functional defects, such as hyperexcitability. Moreover, during inflammation, mast cells appear near enteric neurons in larger numbers than controls, hinting that an immune-ENS interaction could be occurring in IBD patients. An even stronger link may be established over vasoactive intestinal peptide (VIP), one of the major transmitters secreted by ENS. VIP acts as a stimulant for innate lymphoid cells, which respond rapidly to generate an immune response in case the intestinal barrier integrity is disrupted. Loss of VIP has also been associated with an increased inflammation susceptibility [29]. Although the role of ENS remains unclear in inflammation, there is evidence that several genes associated with elevated risk of IBD are expressed in ENS [28], and glial cells could have a responsibility in regulating tumor necrosis factor (TNF) signaling in immune system cells [34].

#### **1.4.2.1 Ulcerative Colitis**

UC is an inflammatory disease which typically starts in the rectum and ascends into the colon. The inflammation is restricted to the mucosal layer of the colon. It typically presents with blood in the stool. In recent years, the incidence of UC has been on the increase, showing a prevalence especially in developed and developing countries. English-speaking countries and Scandinavia have shown the highest number of cases (>7.71) per 100,000 people per year in the last two decades, while India was the leading region out of developed countries with 5.41 cases per 100,000. The incidence rate shows no link to ethnicity or sex [35].

It remains unclear how an ulcerative colitis episode begins, but disruption of the epithelial barrier is one of the key events. In a healthy colon, bacteria cannot pass through the mucus layer covering the underlying intestinal epithelium, but the mucosal layer is invaded by bacteria in most UC patients [36]. These bacteria are less diverse than the typical gut microbiome population. Less Mucin2, a key component of the mucus, is produced by the epithelium, and apoptosis is more prevalent among the epithelial cells. Tight junctions are weakened, which allows more bacteria passage, and subsequent recruitment of immune cells to start the inflammatory response. The intestinal barrier defects could have underlying genetic mechanisms, as healthy relatives of UC patients also have reduced intestinal barrier integrity, but it remains unclear how much the shared environment contributes [35].

UC is classified into three states, mild, moderate, and severe, which show distinctly different morphologies in colonoscopy examination. The area of the colon undergoing inflammation is also used for classification and can range from only the anus and rectum (proctitis) to a whole colon inflammation (pancolitis). The inflammation can occur at any level of severity for any of the episodes a patient experience. However, a patient will typically start experiencing proctitis and the affected region can expand with repeated episodes until the whole colon is inflamed [37]. At a microscopic level, crypt structure is mostly lost, and the submucosa is invaded by immune cells. Cellular composition of the epithelial layer also changes, as goblet cells are depleted and differentiated epithelial cells transition into other cell types. [35]

UC patients are susceptible to polyp formation in more severe states of infection, which may require surgical removal due to an increased risk of cancer. Incidence of colorectal cancer in UC patients is much higher, and some studies propose the risk may be twice as much

compared to the general population. The risk increases over the years a patient has lived with the disease [35].

5-aminosalicylic acid is used for standard treatment and is sufficient in mild to moderate cases. Anti-TNF and anti-inflammatory drug therapies are available for advanced treatment. Inhibitors for Janus kinase (JAK) are also emerging options for UC treatment. If drug therapies are ineffective or not applicable, surgical removal of the affected area could be necessary, with which a portion or the whole of the colon will be removed via colectomy. The remaining intestines can be redirected to outside of the abdomen with an ileostomy surgery, which could be permanent or temporary, if the anus is healthy enough for a reattachment [35,37].

#### **1.4.2.1.1 Dextran Sodium Sulfate Colitis Model**

DSS is a polysaccharide with anticoagulant properties capable of inducing colitis in mice. Its impact depends on the dosage and can mimic human UC. It is highly variable in its molecular weight (5 to 1400 kDa). Most commonly, 40 to 50 kDa is administered via drinking water over several days to induce epithelial damage. Regular drinking water can be given afterwards for a few days to allow recovery. The dosage, treatment length and frequency can be altered to generate models of acute, chronic or relapsing infections [38,39] Mouse strain used for the model plays a role in the progress of infection. Some commonly available strains, such as C57/BL6 and BALB/C, are more susceptible to DSS colitis [39]. Although sex does not appear as a confounder in human UC, female mice display less severe symptoms of colitis and faster recovery in DSS model, but mortality is not affected [40].

DSS damages the colonic epithelial cells directly [39]. 12 hours after DSS administration, colonic mucus composition is already more permeable to bacteria, to the extent that some can already invade the epithelial cells underneath [36]. In Mucin2 deficient mice, colitis can occur spontaneously without DSS or other toxin administration, underlying the importance of mucus production and goblet cells in maintaining colon health. Another early event in colitis development is the compositional changes occurring in the tight junctions, which result in increased permeability of the epithelial layer as well. Claudin-2 is one of the tight junction proteins found more abundant in colitis, and it has been suggested that the intestinal permeability is deliberately increased via its expression to allow more immune cell migration to the area of infection. Claudin-2 deficient mice also show severe colitis upon DSS treatment [39].

Coupling DSS treatment with Azoxymethane (AOM) injections is a commonly used method for generating a UC-associated colorectal cancer model. AOM is a potent carcinogen which can induce cancer in colon when administered on its own as well. AOM/DSS is a powerful model as it closely resembles the UC-associated cancers, albeit with a lower rate of metastasis [41].

#### **1.4.2.2 Crohn's Disease**

CD is also an inflammatory disease very similar to UC, but with unique features. The patient profiles and experiences are very similar, which poses a challenge in diagnosing a case of IBD correctly. The hallmarks used for diagnosis are the presence of skip lesions as opposed to one continuous ulceration, involvement of all gut layers and possibly of gastrointestinal tract outside the colon [42]. Gut walls thicken and constrict the lumen, thus preventing bowel movement. There is a possibility of fistula and abscess formation over time, which typically require surgical correction [43].

CD also differs from UC in terms of treatment options. While steroids can be utilized in both forms of IBD to control infection, 5-aminosalicylic acid treatment does not help with mucosal healing in CD. It is still recommended in some cases for symptom relief. [43]

#### **1.4.3 Diet and Intestines**

The gastrointestinal tract is responsible for nutrient absorption, therefore has to be able to adapt to nutrient availability. Tissue regeneration, stemness and ability to function is directly affected by diet [44]. The amount of calories consumed, and the type of nutrients those calories were available to the body pose different challenges in terms of maintaining the homeostasis. Certain foods are considered proinflammatory, while others can have anti-inflammatory effects. [45]. As a result, an individual's diet can cause adverse effects on various mechanisms of the body, promoting emergence of predisposed disease phenotypes, such as IBD [39] and cancer [46].

##### **1.4.3.1 Fasting**

Restrictive diets, either in terms of timely limitations to consumption, or elimination of certain types of foods have been rapidly emerging as therapeutic options for a variety of conditions in recent years. In general, short term caloric restriction and fasting increases the

regenerative capacity of intestinal crypts, enhancing both stem cell function and proliferation, while not increasing the incidence of tumor formation [47].

#### **1.4.3.1.1 Fasting Model of Mouse**

Several different feeding regimens are utilized to study effects of fasting in mice models. Caloric restriction can be enforced by offering only a certain percentage of daily required calorie intake over a few days. Another alternative is short-term fasting, which is induced by complete food withdrawal for 24 hours [47].

In 24-hour fasted mice, fatty acid oxidation is promoted over carbohydrate metabolism, which enhances regeneration of the intestinal epithelium [44]. Free fatty acids are supplied from the adipose tissue and oxidized, which promotes the peroxisome proliferator-activated receptor (PPAR)- $\delta$  pathway [44,47]. Free fatty acid oxidation also promotes self-renewal of intestinal stem cells. The architecture of the intestines is also affected over longer periods of caloric restriction. Crypts become enlarged, whereas the small intestinal villi become smaller [47].

The intestinal epithelium responds to fasting and refeeding cycles by adapting cellular machineries. The intestinal epithelial cells are already highly polarized cells, which show a preference for performing RNA translation at the apical section of the cell, which produces proteins more efficiently. This type of adjustment allows the intestines to rapidly respond to environmental changes occurring in the gut lumen [48].

#### **1.4.3.2 Obesity**

Obesity is a risk factor for many diseases and is often linked with excessive consumption of sugar and fat [47]. It is classified as a cause of chronic inflammation [49], and high fat consumption alone can be a driving factor for ulcerative colitis [39]. Colorectal cancer risk, as well as risk for diseases of cardiovascular system and diabetes have been linked to obesity as well [50]. The destructive nature of chronic inflammation could be the cause of increased colorectal cancer occurrence in obese patients and could be preventable by changing dietary patterns and promoting an overall healthier lifestyle [51].

Obesity is increasingly being associated with IBD, since the chronic inflammation state could promote lesions typical to IBD, as well as pathways contributing to damage responses. Obesity could contribute to development of CD, but an associated risk factor for UC has not

yet been studied. Other comorbidities of obesity also complicate treatment of IBD [52]. Mouse models have revealed that both the amount and composition of fat matters in terms of susceptibility to colitis. Alterations in bile acid characteristics, such as viscosity and hydrophobicity, are indicative of changes in intestinal permeability. Omega-3 and Omega-6 fatty acid intake has been linked to such changes in bile acid composition [39].

#### **1.4.3.2.1 High-fat Diet Model**

Mouse high-fat diets (HFD) typically offer 32% to 60% of total calories in the form of fats and can be used for inducing obesity over the course of several weeks or months [47]. Mice consuming HFD will typically gain more weight at a faster pace than their counterparts offered comparable diets with less fat. Gene expression profile in the colon is significantly influenced by HFD. Male and female mice exhibit differences in their susceptibility to obesity-associated diseases, such as diabetes, which is suspected to be an effect of estrogen [53].

HFD induces morphological changes in the intestines. In the small intestine, a decrease in the number of Paneth cells occurs, while the number of stem cells increases, along with the crypt depth. In contrast, villi are shortened and therefore have less enterocytes [46,47]. The intestinal stem cells become less dependent on Paneth cell signaling to sustain their stemness. Mechanistically, this phenotype stems from PPAR- $\delta$  signaling activation, which results in an increase of  $\beta$ -catenin activity in the nucleus and elevated expression of Wnt target genes [46]. The small intestine length changes are unclear and appear to be affected by the duration of HFD and the age of the mice. However, the colon becomes significantly shortened and shows shallower crypts with increased stemness and proliferation marker expression. The intestinal barrier function is disrupted in the colon, but not in the small intestines, which could be linked to an increase in apoptosis rate of colonic epithelial cells [54].

#### **1.4.3.2.2 Western Diet Model**

The Western diet (WD) mouse model is utilized to be able to resemble the effects of the current average dietary regime of Western populations. It is also considered a high-fat diet (approximately 35% of calories), but typically contains less fat and has increased carbohydrate content (approximately 49% of calories) instead [47]. WD consumption also results in faster, and more weight gained compared to same-aged control mice and can lead to metabolic syndrome as well. Similar to HFD, the effects of the diet can differ between sexes [55]. Since WD is also a high-carbohydrate diet, some of the observed phenotypes differ from HFD. Unlike HFD, WD hinders intestinal stem cell proliferation, as well as differentiation [47].

## **1.5 Liver**

The liver is an organ tightly associated with the intestines. Their role in metabolism, nutrient absorption and utilization, waste generation and expulsion require highly coordinated actions, which is facilitated over the gut-liver axis. Portal veins establish a direct link from the intestines to the liver, transporting nutrients and metabolites. The liver, in return, secretes bile and antibodies to the intestines, completing the feedback loop. The microbiome of the intestines also influences this communication, meaning changes in intestinal barrier integrity and microbiome composition enforces responses from the liver to restore homeostasis [56].

The microscopic architecture of the liver consists of repetitive, structural units organized around blood and bile vessels. These units, called lobules, have a hexagonal shape around a single central vein, draining blood from the liver to the body's circulation. The blood carrying contents from the intestines enter the liver from the portal veins, which are positioned around the central vein. These six corners of the hexagonal unit are also called the portal triad, since the portal vein is accompanied by the portal artery and bile duct. The liver functions are spatially segregated in concentric layers radiating out of the central vein. The distribution is based on the availability of key substances and factors, such as Oxygen and hormones. This separation allows the liver to perform opposing tasks, like synthesizing cholesterol in the periportal zone, and cholesterol consumption in the pericentral zone [57].

The primary cilia are crucial organelles in monitoring and regulating liver function and responses. Their presence in cholangiocytes, which form the bile ducts, has been well established. The liver primary cilia extend into the lumen of the bile duct and join in regulation of bile composition by monitoring its flow, acidity and ion concentrations [58].

### **1.5.1 Liver, Diet and Disease**

Due to its dominant role in metabolism and close link to the intestines, the liver is susceptible to dietary ailments as well. The liver can be injured through impaired intestinal barrier function, and changes in bile flow and composition [56]. Chronic inflammation is required to induce the damage altering the cellular composition. A dominant mechanism for hepatic injury repair is fibrillar collagen synthesis by myofibroblasts, resulting in extracellular matrix accumulation [59,60]. Myofibroblasts become more abundant in response to injury, and the majority of the population stems from differentiation of hepatic stellate cells (HSCs), but also of other types of cells, such as portal fibroblasts and epithelial cells [60]. Defects of the

primary cilia in the liver are detrimental to the organ and the whole organism and can be a driving cause of fibrosis as well [58]. If fibrosis persists over a long time, it can progress into cirrhosis and eventually liver failure, or form hepatocellular carcinoma [60].

The transition stages between fibrosis and cirrhosis can be categorized by the drivers of the injury state. Non-alcoholic fatty liver disease (NAFLD) is one of the most common precursors for end-stage liver disease, and very prevalent among type 2 diabetic and obese patients. The distinguishing feature of NAFLD, as opposed to other fibrotic diseases such as alcoholic liver disease (ALD), is the accumulation of excess free fatty acids in the liver, which has a toxic effect on the tissue [60,61].

### **1.6 Single-cell RNA Sequencing**

In recent years, the emergence of tools enabling investigation of individual cells constituting a tissue has been pivotal in understanding the heterogeneity, communication, and cooperation present within the individual units of multicellular organisms. scRNA-seq has been one of these core technologies, which provides the transcriptomic profile of any individual cell dissociated from a tissue, enabling the further resolution of a cell's identity, signaling pathways necessary for different stages of development, tissue homeostasis and adaptations under disease states [62].

The technique relies on barcoding of individual mRNA molecules and the cells they stem from, which is achieved by addition of short RNA sequences to the mRNA sequences harvested from the cells. The sequencing part of the procedure is mostly identical to bulk-RNA sequencing. Based on the tissue type and characteristics of the targeted cells, the upstream procedures of cell dissociation and sorting can be modified to achieve a greater resolution, a greater sequencing depth, or a wider array of cells [62-64].

As with all other sequencing techniques, there is a tradeoff between sequencing depth, volume, cost and time. Other limitations more specific to the single-cell techniques arise from the high dimensionality of the data, and challenges in ensuring the data does stem from a single cell only. The latter is usually remedied by defining plausible ranges for RNA detected per cell based on the cell type, size and sequencing conditions. Multiple bioinformatics tools have been developed to both perform this quality control and incorporate the different dimensions into the data analysis [63,64].



Seurat, one such tool developed by Satija et al., is widely used for integrating most common algorithms at all stages of analysis. The standard workflow entails normalizing and scaling the expression data matrix, clustering the cells, and biologically interpreting the common expression profiles of cells within the clusters to identify the detected populations [65].

### **1.6.1 Public Datasets**

Publications featuring data generated by scRNA-seq often share the RNA expression dataset [7,12,15,16,28,34,66-70]. Similar to the Human Genome Project, there is an incentive for mapping individual cells of any human and mouse tissue at a high quality, both in healthy and diseased states, and making these expression datasets easily accessible for further research [71,72]. With this in mind, many publications will present the associated data in an already curated state, reducing the requirements to be able to utilize the data in further analysis, and also generate online apps enabling quick investigation of the data with an easy-to-use user interface [28,68,70,71]

#### **1.6.1.1 Intestine scRNA-seq Data**

The intestines have been studied using scRNA-seq approaches targeting the different compartments of the tissue. Often the layers which are not of interest to the research question are removed before cell sorting [7,15,16,28,66]. The cell dissociation and sorting protocols are optimized to enhance for the cell type of interest [16,66]. Currently, healthy small intestinal epithelium of mice [66] and human [73] and mouse small intestine mesenchyme [12] expression data are available. The different layers of the human colon have been also profiled, from both healthy tissue samples [73,74,75] and from donors experiencing acute colitis [75]. The mouse colon mesenchyme expression data is also available for healthy animals [7,15] and from a DSS colitis model [15].

#### **1.6.1.2 Liver scRNA-seq Data**

Like the intestines, the RNA expression profile of liver cells has been studied with various approaches. Unlike the intestines, the liver tissue cannot be mechanically separated into different layers. Coupled with the vast differences in cell sizes and abundance, enhancing scRNA-seq data output for some cell populations pose a great challenge. Development of novel dissociation techniques or tagging target cells if a known marker is available is necessary, both of which have been utilized in mouse experiments [68,69]. Similar to biopsies from human UC

patients, human liver RNA expression data is also available for both healthy and cirrhotic cells [70].

### **1.6.1.3 Primary Cilia in scRNA-seq Data**

scRNA-seq data can be subsetted using multiple constraints, allowing a more detailed look into certain groups [84]. Since several markers specific for the primary cilia, such as *Arl13b* and *Ift88*, have been characterized over the years, the possible existence of a primary cilia can be argued based on the expression of these markers [21,85].

## **1.7 Discussion**

The presence of primary cilia in the gastrointestinal tract has been documented in the literature, but the identities of cells expressing a primary cilium remains unknown. Due to the primary cilia's ability role in environmental sensing and signal transduction, it is probable that they are required for establishing the stem cell niche of the intestinal epithelium, and direct differentiation. In the liver, the cholangiocytes are defined as the cells possessing primary cilia, but it remains to be determined whether other types of cells also possess the organelle. We suggest that cells responsible for maintenance of the stem cell niche in the colon, small intestine and liver utilize the primary cilium as a signaling hub. Therefore, we aim to test this hypothesis by validating primary cilia presence at transcript level by checking *Arl13b* expression in public scRNA-seq datasets. Then, we will strive to define the primary cilia expressing cells by investigating candidate marker gene expression, focusing on those which could be contributing to the injury response generated by these cells. Lastly, we will be verifying the existence of primary cilia and these marker gene products in vivo, as well as follow them over various models of injury.

## **2. AIMS OF THE STUDY**

### **2.1 Characterization of the primary cilia in the colon, small intestine and liver**

- a) Re-analyze scRNA-seq expression data from previous studies of the colon, small intestine and liver
- b) Investigate expression of primary cilia marker genes

### **2.2 Identification of the marker genes for primary cilia expressing cells**

- a) Find and survey the expression of candidate marker genes in scRNA-seq data from healthy and diseased tissues
- b) *In vivo* validation of primary cilia and associated candidate markers in healthy mouse colon at the protein level.

### **2.3 Monitoring responses in the colon following diet and DSS challenges**

- a) Generate mouse models
  - i) DSS challenge and recovery
  - ii) Fasting and refeeding
  - iii) High-fat diet
  - iv) Western diet
- b) Investigate changes in the number of primary cilia harboring cells via immunohistological analysis of associated marker expressions in the mouse models

### 3. MATERIALS AND METHODS

#### 3.1 Re-analysis of public single-cell RNA sequencing datasets

scRNA-seq datasets (Table 3.1) for mouse or human intestine and liver were obtained from repositories (Table 3.2) named in their respective publications. All code for re-analysis was written in R [75] using Seurat package version 3.1 [65]. RStudio [76] was used for quick visualization of data structures and code outputs. No new software was created. All datasets were available as either Seurat v2 objects, or expression matrices, and made into Seurat v3 format for processing. Where required, Matrix [77] and tidyr [78] packages were utilized for adjusting the matrix to suit Seurat functions, or retrieve information from the matrices within the Seurat objects. To the best of our abilities, clusters were identified as faithfully as possible to their published presentations in the literature. In cases where cluster identities were not present within the data at hand, top markers for the clusters were submitted to Gene Ontology Enrichment Analysis tool [79] to match up with known cell identities. The colon mesenchyme and liver datasets were subsetting to only include cells with Arl13b gene expression greater than 0. Cluster names from the starting dataset were relayed to the Arl13b+ subset.

Using Seurat FindMarkers function, cluster markers were found for each cluster out of genes expressed in at least 1% of the cells, using Wilcoxon rank sum test. Additional information such as health status were also included in the calculation where relevant. The output was investigated to compile a list of candidate marker genes with p-values 0.05 and lower. Gene expressions were visualized with t-SNE and violin plots using default Seurat functions dependent on ggplot2 [80]. The plots were used for visualizing the predicted expression patterns of candidate marker genes and selecting the markers of interest. The codes used for the re-analysis are available in Appendix A.

**Table 3.1.** List of publications scRNA-seq data was obtained and analyzed from

Publication	Tissue	Analyzed data format
Smillie et al. (2019), <i>Intra- and Inter-cellular Rewiring of the Human Colon During Ulcerative Colitis</i>	Human colon biopsies from 12 healthy and 18 UC patients	Expression and training datasets, and code
Kinchen et al. (2018), <i>Structural Remodeling of the Human Colonic Mesenchyme in</i>	Colon mesenchyme of 3 control and 3 DSS-challenged	Curated expression

<i>Inflammatory Bowel Disease</i>	mice	matrix and code
Haber et al. (2017), <i>A Single-cell Survey of The Intestinal Epithelium</i>	Mouse small intestine epithelium from 4 mice per experiment	Curated expression matrix
McCarthy et al. (2020), <i>Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient</i>	Mouse intestine fibroblasts expressing Pdgfra	Curated expression matrix
Roulis et al. (2020), <i>Paracrine Orchestration of Intestinal Tumorigenesis By A Mesenchymal Niche</i>	Mouse small intestinal organoids grown over fibroblasts	Curated expression matrix
Drokhlyansky et al. (2020), <i>The Human and Mouse Enteric Nervous System at Single-Cell Resolution</i>	Mouse large intestine neuron and glial cells	Curated expression matrix
Halpern et al. (2018), <i>Paired-cell Sequencing Enables Spatial Gene Expression Mapping Of Liver Endothelial Cells</i>	Mouse liver endothelial and immune cells from 3 mice	Seurat v2 object
Ramachandran et al. (2019), <i>Resolving the Fibrotic Niche Of Human Liver Cirrhosis At Single-cell Level</i>	Human liver biopsies from 5 healthy and 5 cirrhosis patients	Seurat v2 object
Dobie et al. (2019), <i>Single-Cell Transcriptomics Uncovers Zonation of Function in the Mesenchyme During Liver Fibrosis</i>	Mouse liver mesenchymal (Pdgfrb-expressing) cells from healthy and CCl4-challenged mice	Seurat v2 object

### 3.2 Marker validation with mouse models

Multiple mouse models were established for qualitatively evaluating the changes of interest. C57/BL6 mice, supplied from Bilkent University Animal Facility were used in all models. Each individual experimental group was matched in age and gender with the corresponding control group. All mouse experiment protocols were approved by the Bilkent Animal Ethics Committee. Protein products of marker genes selected from the scRNA-seq expression analysis were visualized in vitro using immunofluorescence.

#### 3.2.1 Establishment of mouse models

Multiple mouse models were set up to apply independent treatments and observe their effects on the mice in terms of weight, as well as macroscopic and microscopic changes occurring in organs of interest.

### **3.2.1.1 DSS challenge and recovery**

DSS, a colitogenic chemical, was administered to mice with their drinking water. 2.5% w/v DSS solution was prepared in drinking water and presented to mice for *ad libitum* consumption over 5 days. At the end of 5th day, DSS water was swapped with drinking water to allow recovery from the challenge.

4 groups, 2 male and 2 female, were used for the model. All mice were 8-12 weeks old. All experimental groups consisted of 3 littermate mice per cage, from which one mouse was collected per time point. The first group was allowed to recover for 1 day, 3 days and 5 days. The collection points were adjusted to no recovery, 1-day and 2-day recovery for the following groups. 3 mice per time point were used as control for the first group, 1 littermate mouse was used as control for the consecutive groups.

The weights of individual mice were measured at the start and end of DSS treatment. DSS water consumption was monitored daily (Table 3.2).

### **3.2.1.2 Fasting and Refeeding**

8- to 12-week-old C57/BL6 mice were starved over a 24-hour period. For the fasting model, mice were sacrificed at the end of the 24 hours. For the refeeding, mice were provided with regular chow food (3.0 kcal/g) for *ad libitum* consumption over 3 hours or 6 hours. Access to water was not restricted throughout the experiment.

Mice were weighted at the start and end of the fasting period (Table 3.3).

### **3.2.1.3 High-fat Diet**

8- to 12-week-old C57/BL6 mice were provided with High-fat diet food (Research Diets, D12492), with an energy density of 5.21 kcal/g. 60% of the total energy was available in the form of fat, 20% as protein, and 20% as carbohydrate. Food was available at all times in small amounts and restocked frequently to prevent spoilage due to high fat content.

**Table 3.2.** Weight chart of mice used in DSS experiments.

Group 1	Date of birth:	23.11.2020	Experiment start:	07.02.2021	Age:	10 wk 6 days	Gender:	male
	Start weights (g)	Control	DSS		End weights (g)	Control	DSS	
	No tag	24	23		No tag	24	23	
	R1	21	23		R1	22	24	
	L1	23	23		L1	24	23	
Group 2	Date of birth:	09.01.2021	Experiment start:	03.03.2021	Age:	7wk 4 days	Gender:	female
	Start weights (g)	Control	DSS		End weights (g)	Control	DSS	
	No tag	17	18		No tag	17	18	
	R1		17		R1		17	
	L1		17		L1		19	
Group 3	Date of birth:	15.03.2021	Experiment start:	11.05.2021	Age:	8 wk 2 days	Gender:	male
	Start weights (g)	Control	DSS		End weights (g)	Control	DSS	
	RL	22	24		RL	22	22	
	R1		26		R1		25	
	L2		24		L2		24	
Group 4	Date of birth:	24.03.2021	Experiment start:	17.05.2021	Age:	8 wk 2 days	Gender:	female
	Start weights (g)	Control	DSS		End weights (g)	Control	DSS	
	R2	18	18		R2	18	18	
	L1		18		L1		17	
	RL		19		RL		19	
Group 5	Date of birth:	10.04.2021	Experiment start:	27.06.2021	Age:	11 wk 1 day	Gender:	female
	Start weights (g)	Control	DSS Cage 1		End weights (g)			
	No tag	20	16		No tag		16	
	R1	19	20		R1		19	
	L1	17	18		L1		18	
	R2		19		R2		18	
	Date of birth:	22.04.2021	DSS Cage 2		Age:	9 wk 3 days		
	No tag		19		No tag		19	
	R1		17		R1		17	
	L1		18		L1		17	
	R2		18		R2		19	
	Date of birth:	22.04.2021	DSS Cage 3		Age:	9 wk 3 days		
	No tag		18		No tag		19	
	R1		18		R1		18	
	L1		16		L1		15	
	R1L1		20		R1L1		20	

**Table 3.3.** Weight chart of mice used in fasting and refeeding experiments.

Date of Measure	Number of Animal	Fasting Group (M)	Control Group (M)		
5.01.2021 @9.00	No Tag	23 g	23 g		
	R1	24 g	24 g		
	L1	25 g			
		<b>Fasting Group @9.00</b>	<b>3h Refeeding @ 12.00</b>	<b>6h Refeeding @ 15.00</b>	<b>Control Group @ 15.00</b>
06.01.2021	No Tag	19 g	Isolated	Isolated	23 g
	R1	21 g	24 g	Isolated	24 g
	L1	22 g	24 g	23 g	

25.01.2021 @9.00	No Tag	17 g	21 g	21 g	18 g
	R1	17 g	19 g	18 g	19 g
	L1	20 g	17 g	18 g	20 g
		<b>Fasting Group (F)</b>	<b>3h Refeeding Group (F)</b>	<b>6h Refeeding Group (F)</b>	<b>Control Group (F)</b>
26.01.2021 @9.00	No Tag	14 g	18 g	18 g	
	R1	15 g	17 g	16 g	
	L1	17 g	14 g	16 g	
26.01.2021 @12.00	No Tag	Isolated	19 g		
	R1	Isolated	18 g		
	L1	Isolated	16 g		
26.01.2021 @15.00	No Tag		Isolated	17 g	17 g
	R1		Isolated	19 g	19 g
	L1		Isolated	19 g	19 g

All mice were weighted at the start of the experiment, and every second week afterwards (Table 3.4).

Mice were housed in groups of 3 per cage, and the cage containing the heaviest mice was selected for tissue isolation at all time points. Tissue isolation was performed after 2, 8, 12 and 16 weeks of high-fat diet consumption. Age-matched control mice, fed on a chow diet, were isolated at the same time.



**Table 3.4.** Weight chart of mice used in HFD experiments.

HFD 2 weeks (F)	Control (F)	HFD 8 weeks (M)	Control (M)	HFD 12 weeks (M)	Control (M)	HFD 16 weeks (M)	Control (M)
18 g	18 g	40 g	28 g	33 g	26 g	40 g	30 g
19 g	18 g	33 g	29 g	46 g	29 g	37 g	29 g
18 g	16 g	28 g	25 g	39 g	27 g	43 g	29 g

#### 3.2.1.4 Western Diet

7-week-old C57/BL6 male mice, and 6- or 7-week-old C57/BL6 female mice were given Western diet (Research Diets, 12079B), for *ad libitum* consumption, with an energy density of 4.67 Kcal/g. 40% of calories were available as fat, 17% as protein, and 43% as carbohydrate. Similar to the high-fat diet model, food was restocked to have at least 10 pellets (avg. weight: 3.3 g) for each cage every 2 days while avoiding spoilage.

All mice were weighted at the start of the experiment, and every second week afterwards (Table 3.5).

#### 3.2.2 Marker Validation with Immunofluorochemistry

Protein products of the genes selected by their expression in scRNA-seq data were targeted using commercially available primary antibodies in order to validate their presence in selected mouse tissues. The primary antibodies were targeted with secondary antibodies tagged with fluorophores to obtain a high-resolution light signal providing positional information.

##### 3.2.2.1 Mouse Colon Isolation

Mice were sacrificed via anesthetic overdose for tissue isolation at key time points of each experimental model. Transcardial perfusion with 10% formalin or 4% w/v PFA was performed for high-fat diet and fasting & refeeding experiments to achieve better fixation of the liver, which were isolated simultaneously for other experiments. The colons were removed from mice, excluding the regions nearest to caecum and rectum, and put in a petri dish filled with ice cold PBS. For DSS experiments, the colon was first placed on a paper towel to measure the length.

**Table 3.5.** Weight chart of mice used in WD experiments. (All weights in grams)

Week:	0		2		4		6		8		10		12		14		
	HFD	Ctrl	HFD	Ctrl	HFD	Ctrl	HFD	Ctrl	HFD	Ctrl	HFD	Ctrl	HFD	Ctrl	HFD	Ctrl	
<b>No tag</b>	20	21	24	25	26	27	28	29	30	31	32	31	33	30	35	32	Male
<b>R1</b>	23	21	27	24	28	26	31	28	35	28	36	28	36	29	37	30	
<b>L1</b>	20	20	23	24	26	28	28	29	30	30	30	30	32	31	33	31	
Cage total	63	62	74	73	80	81	87	86	95	89	98	89	101	90	105	93	
<b>No tag</b>	24	22	28	26	32	27	35	28	38	29	41	30	41	30	44	31	
<b>R1</b>	20	23	25	26	28	27	32	27	36	29	40	29	41	30	44	30	
<b>L1</b>	22	23	26	26	28	27	30	29	33	30	36	31	32	32	30	30	
Cage total	66	68	79	78	88	81	97	84	107	88	117	90	114	92	118	91	
<b>No tag</b>	23	25	26	28	28	30	30	30	33	32	35	33	35	33	36	33	
<b>R1</b>	25	21	28	22	30	26	33	28	38	29	40	31	42	31	42	32	
<b>L1</b>	22	27	26	28	28	30	31	31	33	32	35	34	37	35	37	34	
Cage total	70	73	80	78	86	86	94	89	104	93	110	98	114	99	115	99	
<b>No tag</b>	18	18	24	23	25	25	27	26	30	27	34	28	37	29	38	29	
<b>R1</b>	19	20	24	25	27	27	30	29	35	30	37	31	40	32	41	31	
<b>L1</b>	17	19	21	23	22	24	24	23	26	27	28	27	30	27	31	28	
Cage total	54	57	69	71	74	76	81	78	91	84	99	86	107	88	110	88	
<b>No tag</b>	21	21	26	23	26	24	30	26	32	26	35	27	37	28	38	28	
<b>R1</b>	20	21	25	24	27	25	31	25	33	26	36	27	37	28	39	28	
<b>L1</b>	23	20	26	25	31	26	33	27	36	28	37	29	39	28	39	28	
Cage total	64	62	77	72	84	75	94	78	101	80	108	83	113	84	116	84	
<b>No tag</b>	16	17	19	18	20	20	20	21	22	23	22	21	22	22	22	23	Female
<b>R1</b>	17	13	17	18	18	20	20	19	20	21	21	21	21	21	22	22	
<b>L1</b>	15	17	18	20	20	21	19	22	20	23	22	23	22	23	22	24	
Cage total	48	47	54	56	58	61	59	62	62	67	65	65	65	66	66	69	
<b>No tag</b>	19	19	21	21	24	22	25	23	26	23	29	23	31	24	30	24	
<b>R1</b>	16	17	17	18	19	19	20	19	21	20	22	20	22	21	22	20	
<b>L1</b>	17	18	17	19	18	20	17	22	20	21	19	22	20	23	20	22	
<b>R2</b>	18	17	18	18	19	20	20	21	21	20	21	22	21	22	21	22	
Cage total	70	71	73	76	80	81	82	85	88	84	91	87	94	90	93	88	
<b>No tag</b>	17	16	19	19	20	20	22	20	24	20	24	21	25	21	26	22	
<b>R1</b>	19	13	19	16	21	16	23	18	24	19	26	20	28	20	31	21	
<b>L1</b>	17	16	20	19	20	20	22	21	21	22	24	21	24	22	25	22	
Cage total	53	45	58	54	61	56	67	59	69	61	74	62	77	63	82	65	

A 10 ml sterile syringe was fitted with a 200  $\mu$ L micropipette tip cut at both ends. The syringe was filled with ice cold PBS, and the micropipette tip was inserted into the colon tube. The gut contents were flushed out with the PBS. The clean colon was cut into small pieces and placed in the wells of a 12-well plate, to be kept on ice until dehydration. The tissue was washed twice for 5 minutes in PBS.

### **3.2.2.2 Tissue Fixation, Dehydration, Embedding and Sectioning**

PBS used for washing the tissue was exchanged for 4% PFA.

For frozen sectioning, the tissues were washed 3 times with PBS for 5 minutes after allowing PFA fixation for 1 hour. PBS was exchanged for 30% w/v Sucrose solution in PBS. The tissues were stored overnight at 4°C.

On the next day, the tissues were transferred into Sakura Tissue-Tek O.C.T Compound poured on a petri dish and coated thoroughly. Plastic embedding molds were filled halfway with the same compound, and the coated tissues were transferred into the mold, which was filled up right afterwards. The tissues were rapidly placed in a -80°C freezer for the compound to solidify for a few hours. For long-term storage, the frozen blocks were moved to -20°C.

For paraffin embedding, 4% PFA was refreshed right before placing tissues in overnight storage at 4°C. The tissues were first washed in PBS at room temperature 3 times for 5 minutes, then placed in 30% v/v Ethanol in PBS solution for 1 hour at room temperature. The solution was replaced with 50% v/v Ethanol in ddH<sub>2</sub>O for the next hour. Finally, the tissues were transferred into embedding cassettes and put in 70% v/v Ethanol in ddH<sub>2</sub>O. If the subsequent tissue processing could not be performed on the same day, the tissues were kept in 70% Ethanol solution at 4°C, which was refreshed every 2 days to prevent contamination.

Tissue processing was done using a tissue processor overnight with the following program: 70%, 80%, 90%, 96% Ethanol (in ddH<sub>2</sub>O) for 1 hour each, 100% Ethanol for 1 hour three times, Xylene for 1 hour two times, liquid paraffin for 1 hour, and 1.5 hours. The processed tissues were embedded into liquid paraffin and allowed to solidify for 1 hour on a cool plate. The paraffin blocks were stored at room temperature.

Both paraffin and frozen blocks were sectioned into 5 $\mu$ m thick layers using a microtome and cryotome respectively. The frozen sections were kept at room temperature after transfer onto the slides for improved adherence. The slides were preheated to 38°C for paraffin

sections, which were allowed to float on 38°C ddH<sub>2</sub>O to flatten, and then placed on a stretching table at 38°C to dry overnight.

### **3.2.2.3 Immunofluorescent Staining**

Tissue sections were prepared to accept antibody binding and tagged with antibodies specific to target proteins using the procedure described below.

#### **3.2.2.3.1 Deparaffinization and Rehydration**

Paraffin-embedded tissue sections were dehydrated by dipping slides in Xylene twice for 7.5 minutes. Then, the tissues were rehydrated with serial Ethanol washes. 100%, 96%, 80%, 70% and 50% v/v Ethanol solutions were used. Every wash was done twice, for 4 minutes each.

Frozen section slides were instead fixed with 4% PFA for 10 minutes.

The slides were then washed with PBS 3 times for 5 minutes.

#### **3.2.2.3.2 Antigen Retrieval**

Antigen retrieval was performed only for paraffin-embedded tissue sections. Depending on the antibody (Table 3.3), either citrate buffer (1.8% stock 0.1 M citric acid anhydrous, 8.2% stock 0.1 M trisodium citrate dihydrate, mixed to 10 mM working dilution, pH 6.0) or Tris-EDTA buffer (10 mM Trizma base, 1 mM EDTA, pH 9.0) was used. First, the buffer was preheated to boil in a microwave and poured into a glass container with a lid. Then, slides were dipped into the buffer with a slide holder and quickly put into a vegetable steamer with water pre-boiled for steam build-up. The slides were kept in the steamer for 35 minutes, then allowed to cool together with the buffer for at least 20 minutes at room temperature. The cooled tissues were washed twice for 5 minutes with PBS.

#### **3.2.2.3.3 Blocking and Immunostaining**

The blocking buffer was prepared with 10% donkey serum and 0.3% Triton-X in PBS. Tissue slides were taken out of PBS wash and briefly allowed to dry before being placed in a dark wet chamber. Tissues on the slide were coated with the blocking buffer, and carefully covered with a Parafilm strip smaller than the slide area, but large enough to cover all tissues. The tissues were allowed to incubate in blocking buffer for 1 hour at room temperature.

The primary antibodies were prepared in the blocking buffer diluted to the factors listed in (Table 3.3). Antibodies were added to the same blocking buffer for double or triple staining. Parafilms were removed and the blocking buffer was drained at the end of the blocking step. Primary antibody was applied to the tissues in the same manner as the blocking buffer. The tissues were allowed to incubate in primary antibodies overnight at 4°C.

The secondary antibodies were prepared in the same manner as primary antibodies, but in aluminium foil wrapped tubes to avoid bleaching of the fluorophores. The antibody tubes were very briefly centrifuged to avoid transferring fluorophore clumps onto the tissues. The slides were washed 3 times in PBS for 5 minutes and incubated with secondary antibodies for 1 hour at room temperature.

All washes after this point were done in aluminium foil wrapped containers. The slides were first washed twice in PBS for 5 minutes, and then twice in PBS-T (0.3% v/v Triton X in PBS) for 5 minutes. DAPI was diluted in PBS-T in a foil wrapped tube and applied on the tissues as before. A parafilm cover was not used as opposed to the previous incubation steps, and the tissues were allowed to incubate for 10 minutes. The slides were washed in reverse; twice in PBS-T and twice in PBS for 5 minutes each.

The slides were allowed to dry shortly after the final wash and then laid flat on a paper towel. Mounting medium was applied directly over the tissues. The coverslip was applied over the slide by carefully pressing on the glass with the tip of a forceps, and tracing from bottom to top to push air bubbles out and achieve full coating. The slides were stored at 4°C.

#### **3.2.2.4 Imaging**

The microscopic structures of the colon were visualized via confocal microscopy. Fluorescent signals from the antibody-tagged proteins were studied qualitatively to determine the location and morphology of the respective positive cells.

##### **3.2.2.4.1 Confocal Microscopy**

All microscopy images were obtained using a White Light Laser Leica TCS SP8 confocal microscope. LAS X software provided by the manufacturer was used for the setup and image recording. DAPI staining was visualized using the 405 nm diode and used for locating the tissue and target structures. White light laser (70% power) was used for excitation of the fluorophores at their appropriate wavelengths. Default settings were used for each of the

fluorophores. The signal was collected using a PMT or HyD detector, which was selected automatically by the software.

The intensity, gain and offset settings were determined at the start of an imaging session and kept unchanged between experimental and control tissue slides for all markers of interest. DAPI and E-cadherin signal intensities were seldom adjusted while scanning different regions.

All images were taken over a z-stack size determined visually according to the regions where signals were detected. To cover the full section thickness of 5  $\mu\text{m}$ , the total z-size was set to be close to 5  $\mu\text{m}$  as well. The start and end points were determined according to the signal strength of dyes marking the proteins of interest. The z-step size was set to 0.30  $\mu\text{m}$  or 0.35  $\mu\text{m}$ .

All images were saved as .lif files.

### 3.2.2.4.2 Image Processing

All images were processed using the Fiji [81] distribution of ImageJ [82], version 1.53c.

Files containing the microscopy images were imported using the Bio-Formats plugin [83] with the following options: View stack with: Hyperstack, Stack order: XYZCT; Color mode: Colorized, Autoscale checked; Split into separate channels: Split channels.

Z-stack projection for each color channel of an image was calculated individually. The method producing the strongest signal and the least background was selected for each experiment. Preferred projection methods for each target are listed in (Table 3.3). For DAPI and E-cadherin images, some of the slices in the channel were omitted to produce a better focused image when necessary. All channels were merged to produce the final image, and also merged in different combinations for better visibility. All images were converted to RGB Color type and exported in TIFF format. Crypt lengths were measured by drawing a line from the bottom of the crypt to the mouth of the crypt and using the native ImageJ function.

**Table 3.6.** List of utilized reagents and resources.

Reagent or resource	Source	Identifier
High-fat diet	Research Diets	D12492
Western diet	Research Diets	12079B
Regular chow diet	Optima	Rat Yemi 24/30

Dextran sodium sulfate	MP Biomedicals	160110
Paraformaldehyde	Fluka	76240
Formalin	Sigma-Aldrich	HT501128
10x PBS	Biowest	Dulbecco's Phosphate Buffered Saline 10x w/o Calcium, w/o Magnesium
PBS, tablet	Sigma-Aldrich	P4417
O.C.T. Compound	Sakura	Tissue-Tek O.C.T. Compound
Paraffin	Sakura	Tissue-Tek III Paraffin Wax (Polymer added)
Microscope Slides	Thermo Scientific	Menzel Gläser Superfrost Plus
Cover Slips	Thermo Scientific	Menzel Gläser Cover Slips 24*50mm
Citric acid anhydrous	Sigma-Aldrich	55H0918
Tri-sodium citrate dihydrate	Merck	1.06448
Trizma base	Sigma-Aldrich	T1503
EDTA disodium salt dihydrate	AppliChem	A2937
Triton X-100	Sigma-Aldrich	T-8787
Donkey serum	Sigma-Aldrich	D9663
Mounting medium	Abcam	ab104135
Xylene	Sigma-Aldrich	16446

<b>Data</b>	<b>Source</b>
Smillie et al. (2019)	<a href="https://singlecell.broadinstitute.org/single_cell/study/SCP259">https://singlecell.broadinstitute.org/single_cell/study/SCP259</a>
Kinchen et al. (2018)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11437">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11437</a>
Halpern et al. (2018),	Provided by author
Ramachandran et al. (2019)	<a href="https://datashare.is.ed.ac.uk/handle/10283/3433">https://datashare.is.ed.ac.uk/handle/10283/3433</a>
Roulis et al. (2020)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142431">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE142431</a> , GSM4227191 and GSM4227192 only
Dobie et al. (2019)	<a href="https://datashare.ed.ac.uk/handle/10283/3573">https://datashare.ed.ac.uk/handle/10283/3573</a>
Haber et al. (2017)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92332">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92332</a>
McCarthy et al. (2020)	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130681">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130681</a> , GSM3747599_PdgfraEGFPsc_1_mm10 only

Drokhlyansky et al. (2020)	<a href="https://singlecell.broadinstitute.org/single_cell/study/SCP1038">https://singlecell.broadinstitute.org/single_cell/study/SCP1038</a>
<b>Software</b>	<b>Source</b>
R (v3.6) [75]	<a href="https://www.r-project.org/">https://www.r-project.org/</a>
Seurat (v3) [65]	<a href="http://www.satijalab.org/seurat">http://www.satijalab.org/seurat</a>
Rstudio (v1.3) [76]	<a href="http://www.rstudio.com/">http://www.rstudio.com/</a>
Matrix (v1.18) [77]	<a href="https://cran.r-project.org/web/packages/Matrix/index.html">https://cran.r-project.org/web/packages/Matrix/index.html</a>
tidyr (v1.1.2) [78]	<a href="https://tidyr.tidyverse.org/">https://tidyr.tidyverse.org/</a>

**Table 3.7:** List of antibodies used in immunofluorochemistry and optimal working conditions.

<b>Antibody and/or Dye</b>	<b>Source</b>	<b>Identifier</b>	<b>Dilution factor</b>	<b>Antigen retrieval buffer</b>	<b>Percent laser intensity</b>	<b>Z-stack projection method</b>
Goat polyclonal anti human/mouse Grem1	R&D Systems	AF956	1:50	Tris-EDTA or citrate	15	Standard deviation
Goat polyclonal anti mouse Pdgfra	R&D Systems	AF1062	1:50 or 1:100	Tris-EDTA or citrate	5-15	Maximum Projection or Sum slices
Rabbit polyclonal anti Arl13b	Proteintech	17711-1-AP	1:1000	Tris-EDTA or citrate	3-15	Standard deviation or Sum slices
Mouse monoclonal anti Arl13b	Abcam	ab136648	1:1000	Tris-EDTA or citrate	3-15	Standard deviation or Sum slices
Mouse monoclonal anti acetylated tubulin	Sigma-Aldrich	T7451	1:100	citrate	10	Sum slices
Rabbit polyclonal anti human Acot7	Sigma-Aldrich	HPA025735	1:100 or 1:150	Tris-EDTA	3-10	Sum slices
Purified mouse monoclonal anti human E-cadherin	BD Biosciences	610182	1:100 or 1:50	Tris-EDTA or citrate	25-100	Sum slices
Rabbit monoclonal recombinant anti Ki67	Abcam	ab16667	1:100	citrate	9- 10	Maximum Projection



Rabbit polyclonal anti pericentrin (PCNT)	Abcam	ab4448	1:100	Tris-EDTA	2	Sum slices
DAPI	Thermo Scientific	62248	1:3000	N/A	1-7	Maximum Projection or Sum slices
Donkey anti-rabbit IgG, Alexa Fluor 488	Thermo Scientific	A-21206	1:400	N/A	N/A	N/A
Donkey anti-mouse IgG, Alexa Fluor 647	Thermo Scientific	A-31571	1:400	N/A	N/A	N/A
Cy2 donkey anti-mouse IgG	Jackson ImmunoResearch	715-225-151	1:400	N/A	N/A	N/A
Cy3 donkey anti-rabbit IgG	Jackson ImmunoResearch	711-165-152	1:400	N/A	N/A	N/A
Cy5 donkey anti-goat IgG	Jackson ImmunoResearch	705-175-147	1:400	N/A	N/A	N/A

## 4. RESULTS

### 4.1 Primary Cilia Marker Gene Expression

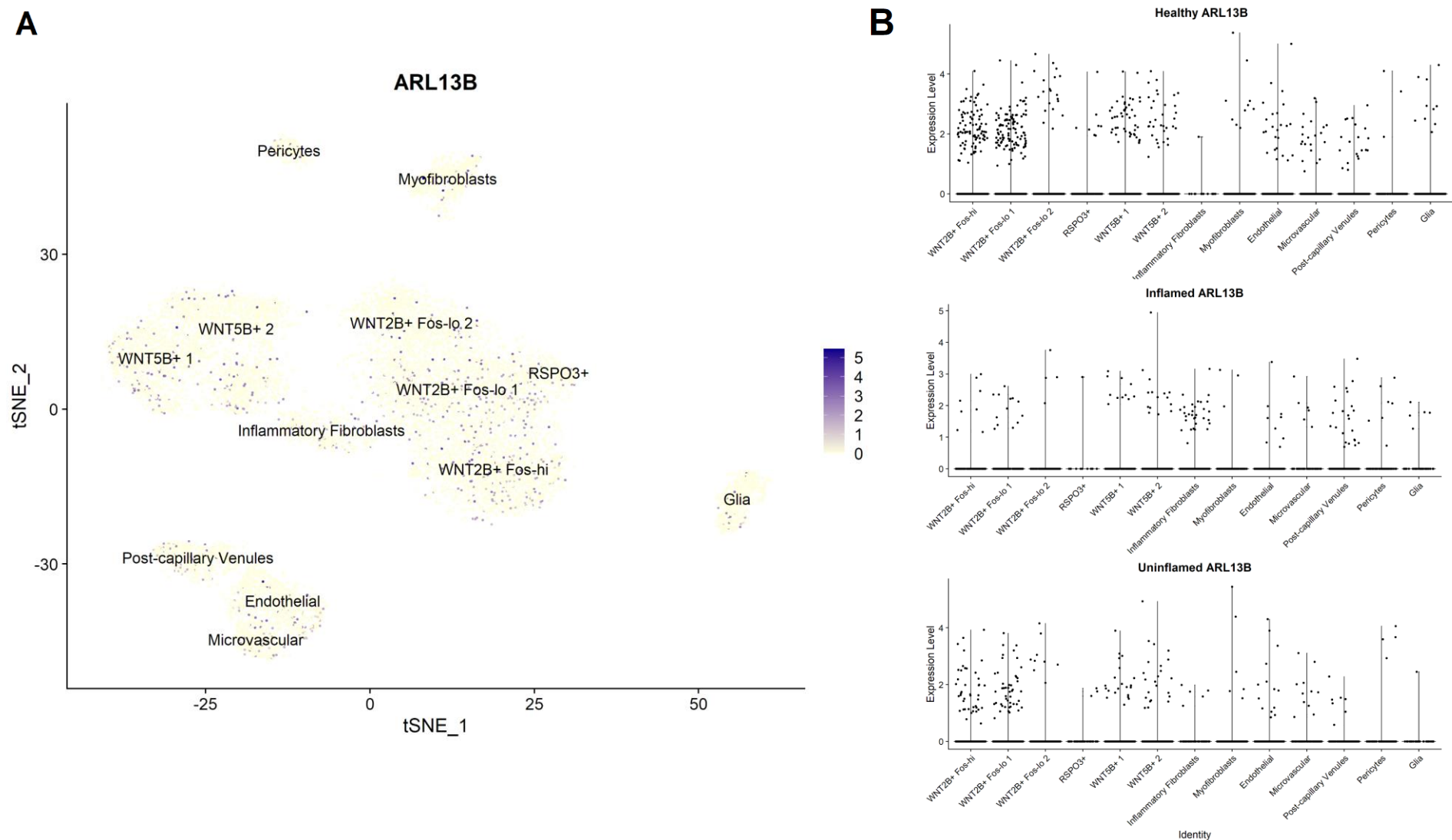
The existence and distribution of primary cilia in intestine and liver tissues was first investigated *in silico* using the expression of primary cilia marker gene Arl13b. Arl13b transcripts were found in all datasets. Ifit88, another marker for primary cilia, was also present in a similar pattern to Arl13b, albeit in lower numbers.

#### 4.1.1 Primary Cilia in the Colon and Small Intestine

Arl13b transcripts were found in all clusters of the colon mesenchyme in both human and mouse datasets. Moreover, both healthy and inflamed tissues showed Arl13b expression. The expression appears mostly uniform throughout all clusters.

In the human colon mesenchyme, both fibroblast and other stromal cell populations have ARL13B transcripts. Among the fibroblast populations, (WNT2B<sup>+</sup> Fos-hi, WNT2B<sup>+</sup> Fos-lo 1, WNT2B<sup>+</sup> Fos-lo 2, WNT5B<sup>+</sup> 1, WNT5B<sup>+</sup> 2, RSPO3<sup>+</sup> and Inflammatory Fibroblasts) WNT2B<sup>+</sup> Fos-hi and WNT2B<sup>+</sup> Fos-lo 1 show more abundant ARL13B expression (Fig. 4.1A, Table 4.1). WNT2B<sup>+</sup> Fos-lo 2 and WNT 5B<sup>+</sup> 2 stand out as clusters with less expression than the other WNT<sup>+</sup> clusters (Fig. 4.1B, Table 4.1). The Inflammatory Fibroblast population is larger in the inflamed tissue compared to the healthy and uninjured tissues (Fig. S1) and show slightly more ARL13B expression by percentage (Fig. 4.1B, table 4.1). Overall, the percentage of fibroblast cells positive for ARL13B expression is higher in UC patients compared to the healthy controls (Table 4.1)

Arl13b expression in the mouse colon mesenchyme is similar to the human pattern, but the transcripts are much more abundant (Table 4.2). Similar to the increase observed between human healthy and UC patients, mice challenged with DSS show more Arl13b expression compared to the healthy control animals. A cluster of cells, labeled Str4 and detected in the DSS-challenged mouse colon mesenchyme only, also express Arl13b like other stromal cell clusters. (Fig.4.2, Table 4.2).



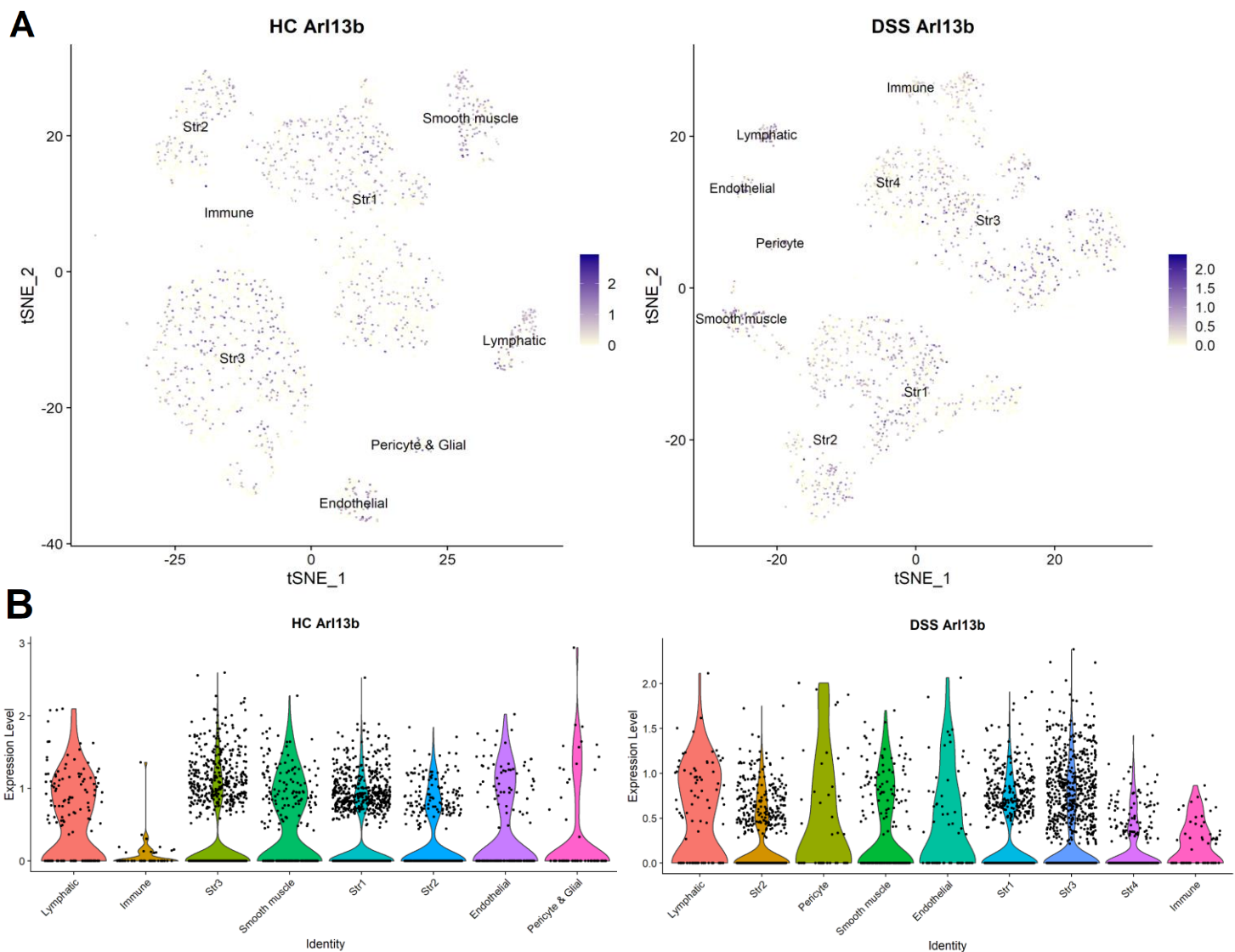
**Figure 4.1. Single-cell RNA expression profile of biopsied human healthy, injured or uninjured colon mesenchyme.** Data from Smillie et al. (2019). Each dot represents an individual cell. A) Feature plot of ARL13B distribution across whole data t-SNE plot. Colors show the relative expression level, scale available on the right-hand side. B) Violin plots showing ARL13B expression per condition. Healthy: tissue from donor with no UC history, Inflamed: actively inflamed tissue from UC patient, Uninflamed: healthy tissue biopsied proximal to active UC sites.

**Table 4.1.** Number of cells in the human UC colon mesenchyme dataset from Smillie et al. (2019)

	WNT2B+ Fos-hi	WNT2B+ Fos-lo1	WNT2B+ Fos-lo2	RSPO3+	WNT5B+ 1	WNT5B+ 2	Inf. Fib	Myofibroblasts	Endothelial	Microvascular	Post-capillary venules	Pericytes	Glia
All	1746	1960	1124	194	1254	1555	475	621	738	406	646	249	588
Arl13b+ cells	145	167	33	12	78	72	37	19	47	38	50	14	16
%Arl13b+ cells	%8,30	%8,52	%2,94	%6,19	%6,22	%4,63	%7,79	%3,06	%6,37	%9,36	%7,74	%5,62	%2,72
Healthy	1171	1398	712	135	738	905	16	256	403	258	255	81	492
Arl13b+ cells	95	107	20	8	46	34	1	10	24	21	20	3	10
%Arl13b+ cells	%8,11	%7,65	%2,81	%5,93	%6,23	%3,76	%6,25	%3,91	%5,96	%8,14	%7,84	%3,70	%2,03
Inflamed	95	110	140	12	120	197	384	123	166	52	253	81	50
Arl13b+ cells	8	13	4	1	11	15	31	3	9	6	23	7	5
%Arl13b+ cells	%8,42	%11,82	%2,86	%8,33	%9,17	%7,61	%8,07	%2,44	%5,42	%11,54	%9,09	%8,64	%10,00
Uninflamed	480	452	272	47	396	453	75	242	169	96	138	87	46
Arl13b+ cells	42	47	9	3	21	23	5	6	14	11	7	4	1
%Arl13b+ cells	%8,75	%10,40	%3,31	%6,38	%5,30	%5,08	%6,67	%2,48	%8,28	%11,46	%5,07	%4,60	%2,17

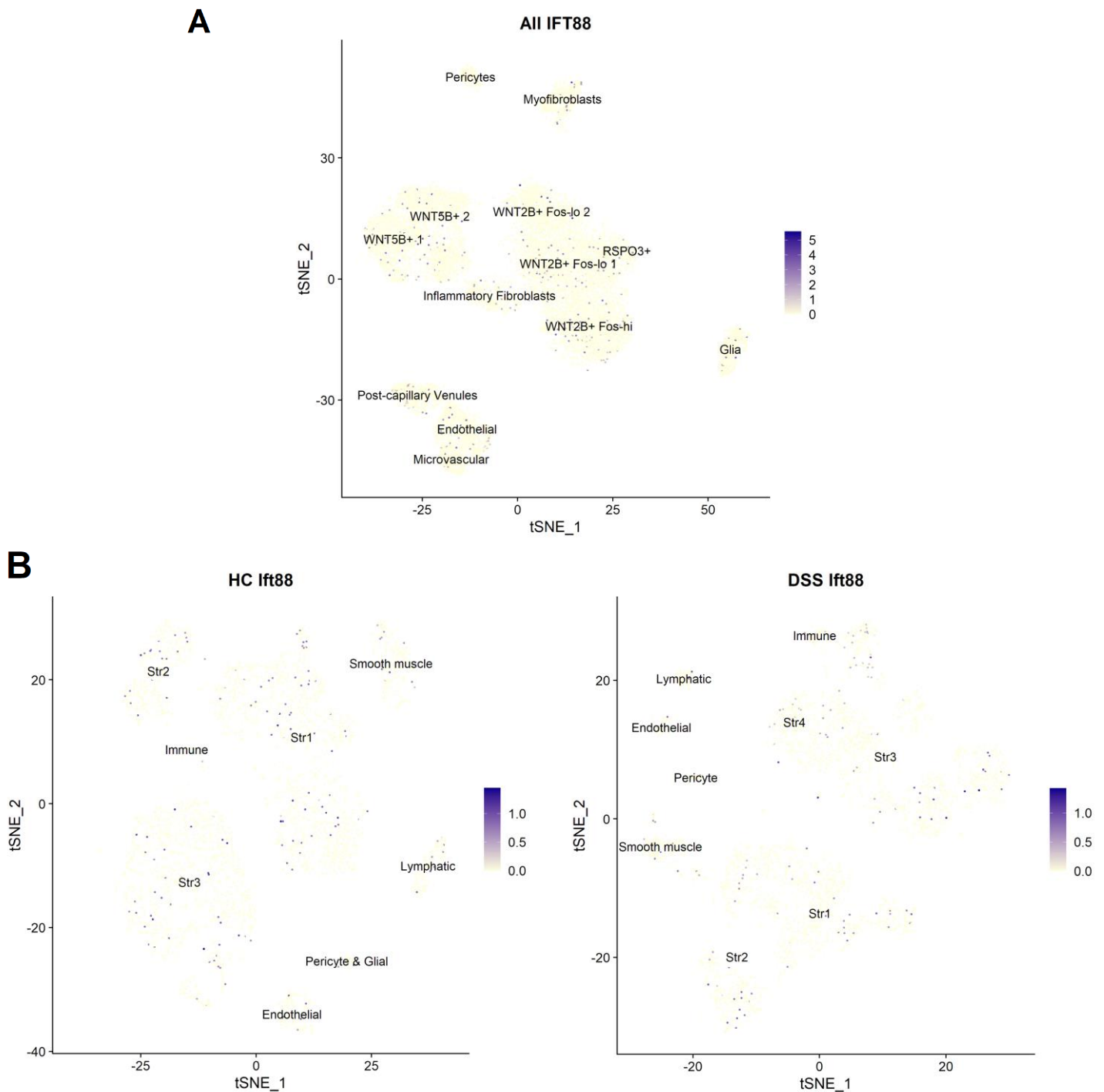
**Table 4.2** Number of cells in the mouse colon mesenchyme dataset from Kinchen et al. (2018)

	Lymphatic	Immune	Str3	Smooth muscle	Str1	Str2	Endothelial	Pericyte & Glial	Str4
HC	148	42	1292	240	1444	380	150	64	
Arl13b+ cells	89	13	383	131	479	134	59	18	
%Arl13b+ cells	%60,14	%30,95	%29,64	%54,58	%33,17	%35,26	%39,33	%28,13	
DSS	87	58	1363	198	782	621	68	46	268
Arl13b+ cells	53	30	532	86	290	242	35	23	93
%Arl13b+ cells	%60,92	%51,72	%39,03	%43,43	%37,08	%38,97	%51,47	%50,00	%34,70



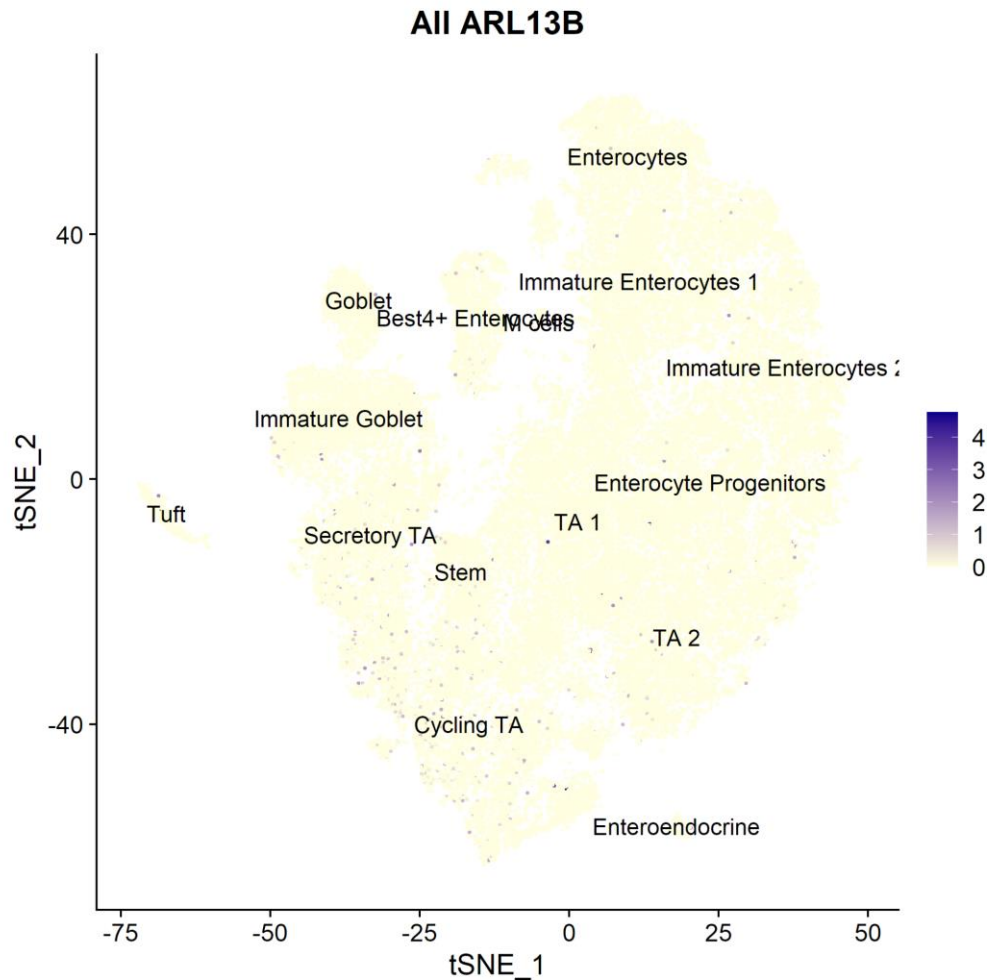
**Figure 4.2.** Arl13b mRNA expression of colon mesenchymal cells in healthy or DSS treated mice. Data re-analyzed from Kinchen et al. (2018). Each dot represents an individual cell. A) Feature plots of Arl13b distribution. Colors show the relative expression level, scale available on the right-hand side. B) Violin plots showing Arl13b expression per condition in each cluster. HC: healthy control, DSS: DSS-challenged. Str: Stromal.

The presence of primary cilia was also controlled with the expression of the *Ift88* gene. *Ift88* transcripts are present in both human and mouse scRNA-seq data, with a similar distribution to *Arl13b*. However, the *Ift88* transcript count is lower than *Arl13b* (Fig. 4.3).



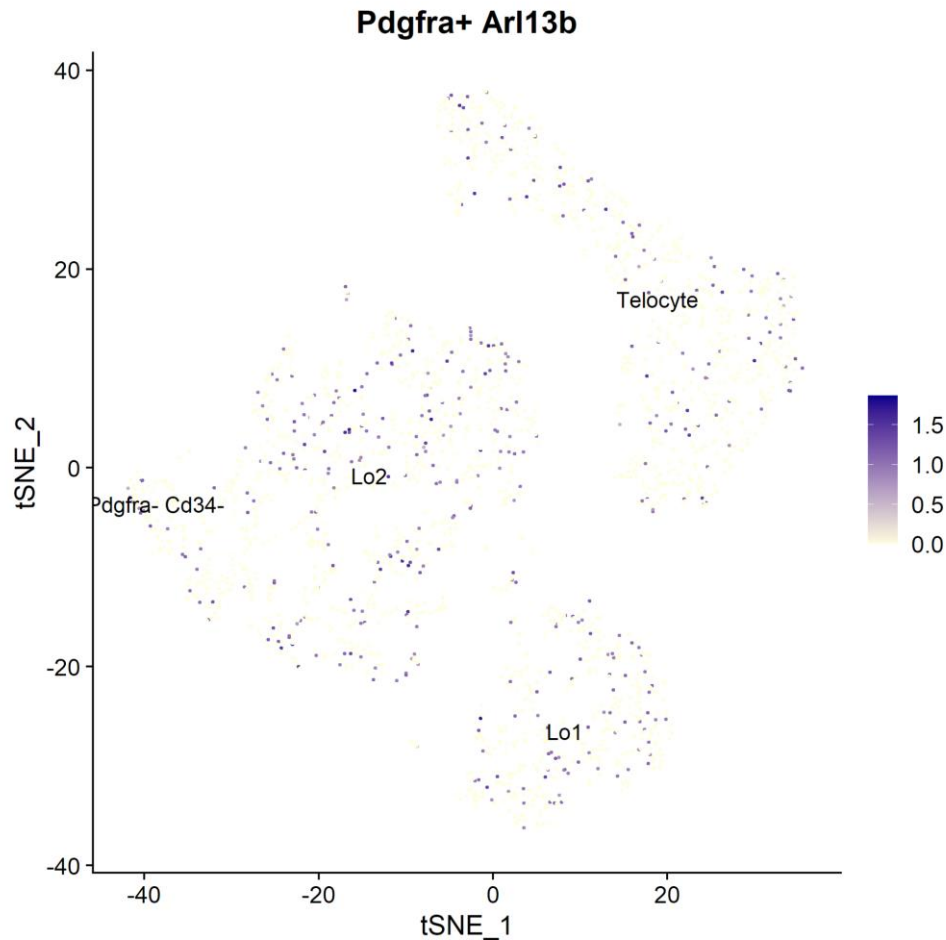
**Figure 4.3. *Ift88* mRNA expression in colon mesenchyme.** Feature plots of *Ift88* transcript distribution. Each dot represents an individual cell. Colors show the relative expression level, scales available on the right-hand side of each plot. A) Human colon mesenchymal dataset from Smillie et al.(2019) Plot shows cells from all health condition groups shown in Fig. 4.1. B) Mouse colon mesenchymal dataset from Kinchen et al. (2018), re-analyzed, same t-SNE plot as in Fig. 4.2. HC: healthy control DSS: DSS-challenged.

Arl13b expression in the human colon epithelium was also plotted to check whether primary cilia are present in the epithelial cells of the colon. Some transcripts were detected (Fig. 4.4). Similarly, mouse small intestinal epithelium also showed Arl13b transcripts (Fig. S2).



**Figure 4.4. ARL13B mRNA expression of biopsied human colon epithelium.** Data from Smillie et al. (2019). Feature plot from human colon epithelium scRNA-seq data. Plot shows cells from all health condition groups shown in Fig. 4.1. Each dot represents an individual cell. Colors show the relative expression level, scales available on the right-hand side of each plot.

Lastly, the distribution of the primary cilia in the small intestine mesenchyme was visualized in a *Pdgfra*<sup>+</sup> dataset. Similar to what was observed in the colon mesenchyme dataset, the *Arl13b* transcripts were present in all clusters. However, they were less abundant compared to the mouse colon mesenchyme (Fig. 4.5).



**Figure 4.5. Arl13b mRNA expression mouse small intestine mesenchymal cells FACS sorted for Pdgfra.** Data from McCarthy et al. (2020), re-analyzed. Feature plot showing Arl13b transcript distribution. Each dot represents an individual cell. Colors show the relative expression level, scales available on the right-hand side of each plot.

#### 4.1.2 Primary Cilia in the Liver

Primary cilia presence was investigated in the liver with the same approach. Arl13b transcripts were found in multiple clusters in liver scRNA-seq data generated from human or mouse tissues. Both healthy and fibrotic livers showed Arl13b transcription.

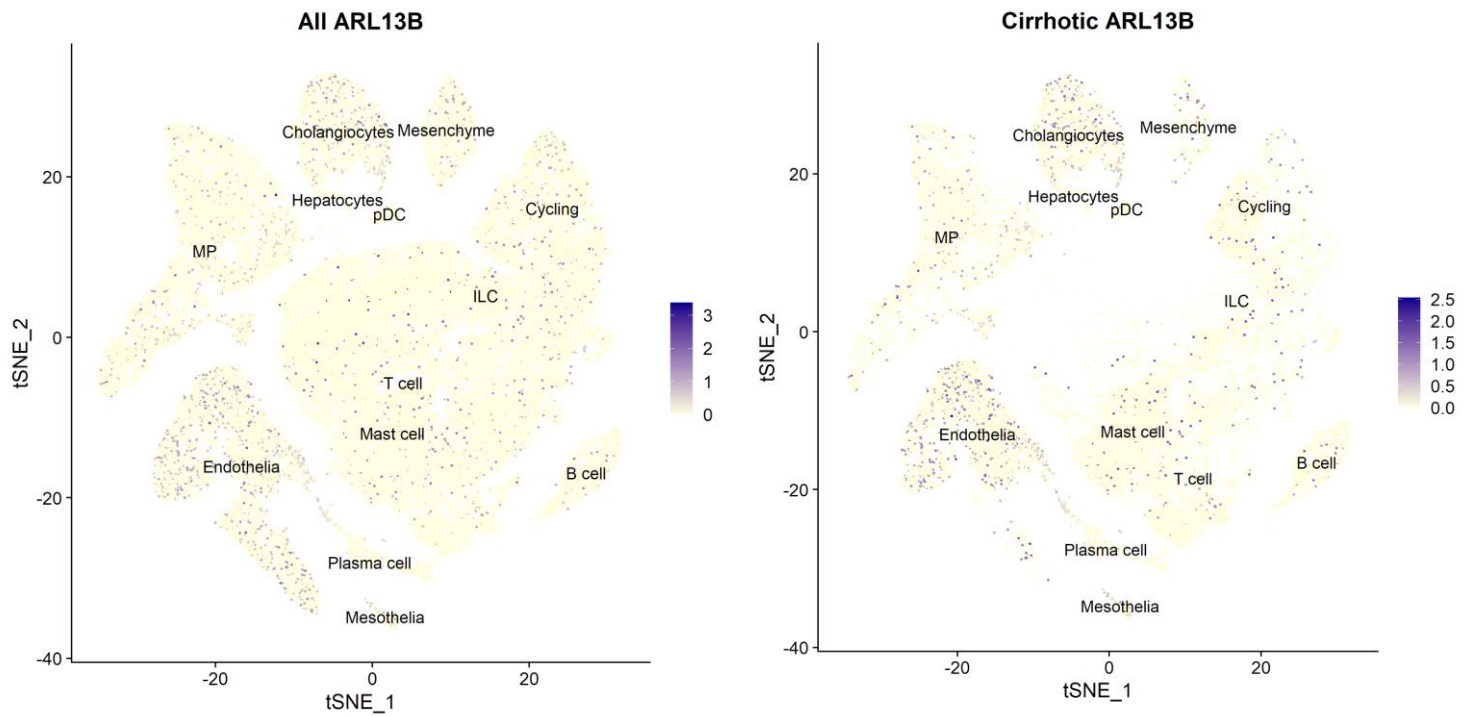
In the human liver, cholangiocytes show the highest abundance of primary cilia marker ARL13B (Table 4.3). However, Arl13b expression is also observed in other cell types (Fig. 4.6). After cholangiocytes, mesothelial cells are the most frequently ARL13B expressing cells. Moreover, the sub-clusters of the assigned cell groups show variation in their ARL13B frequency. (Table 4.3).

Primary cilia associated gene expression in the mouse liver mesenchyme was investigated using a population selected for Pdgfrb expression. Arl13b transcripts were



abundant. HSCs and fibroblasts were more often Arl13b<sup>+</sup> compared to vascular smooth muscle cells. Both healthy and CCl<sub>4</sub>-induced fibrotic livers expressed Arl13b. (Fig. 4.7).

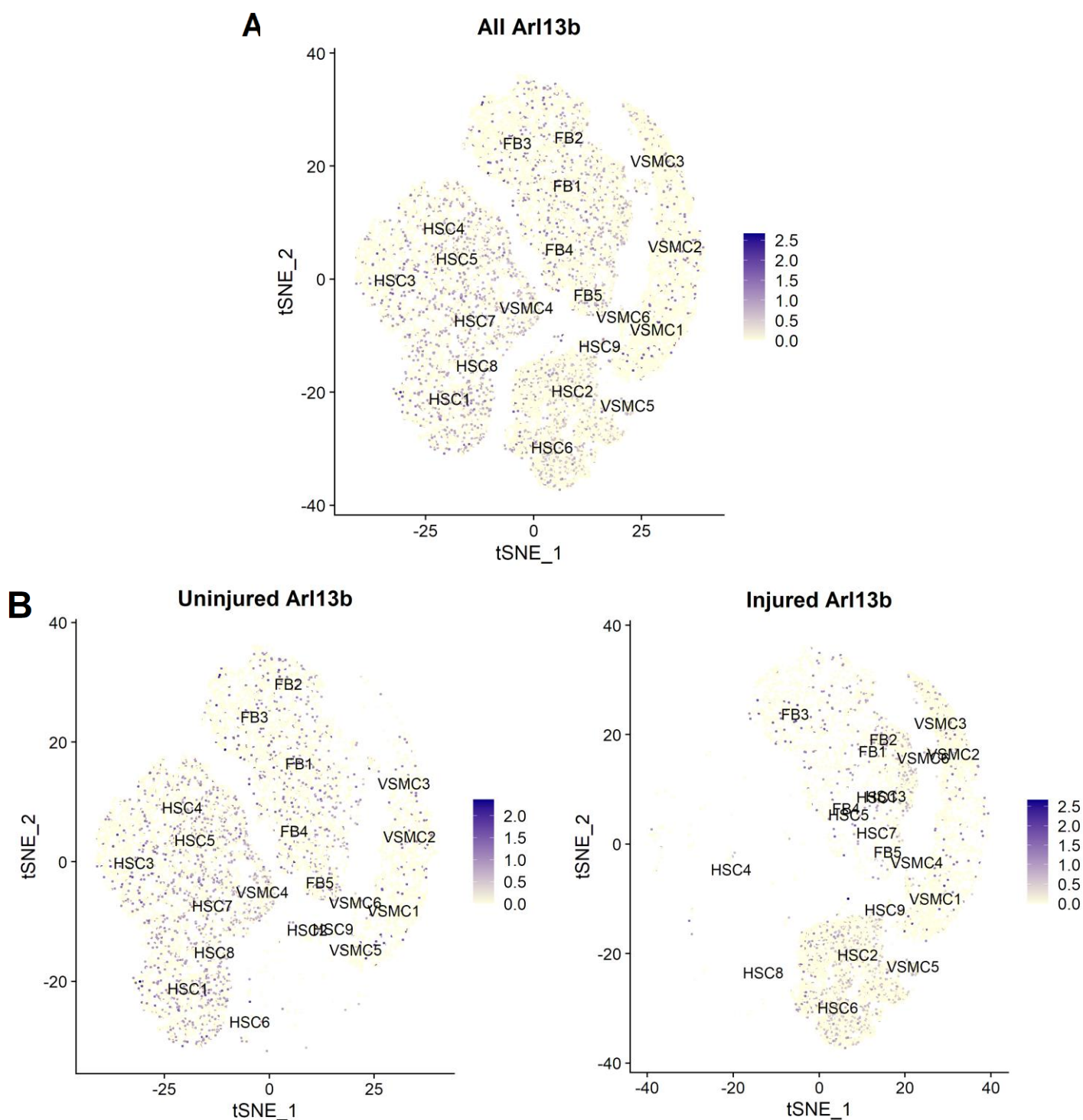
In two of the three datasets re-analyzed, immune cells were also present and showed little to no Arl13b expression (Fig. 4.6, Fig. S3). On the other hand, Arl13b expression is scarce in the mouse liver endothelium (Fig. S3), but frequent in the human liver (Fig. 4.6).



**Figure 4.6. ARL13B mRNA expression profile of human uninjured and cirrhotic liver biopsies.** Data from Ramachandran et al. (2019). Cells were re-clustered while preserving the original cluster names. Feature plot showing ARL13B transcript distribution. Each dot represents an individual cell. Colors show the relative expression level, scales available on the right-hand side of each plot. Left: Expression profiles of all sequenced biopsies, healthy and cirrhotic donor data combined. Right: ARL13B<sup>+</sup> cells from cirrhotic liver biopsies only.

**Table 4.3.** Number of cells in the Ramachandran et al. (2019) human liver dataset, non-immune cells only. Subcluster: PCA-assigned cluster number of a member cluster (Fig. S4)

	Hepatocytes	Cholangiocytes	Mesenchyme	Mesothelia	Endothelia
All	180	3467	1901	427	7938
subcluster	22	8	20	12	19
	180	3059	408	1901	427
ARL13B+	10	356	109	67	962
subcluster	10	282	74	109	67
	180	3059	408	1901	427
Cirrhotic all	137	2384	402	348	4875
subcluster	137	2154	230	402	348
	137	2154	230	402	348
ARL13B+	4	280	36	58	550
subcluster	4	220	60	36	58
	4	220	60	36	58



**Figure 4.7. Arl13b mRNA expression profile in *Pdgfrb*<sup>+</sup> mouse liver mesenchyme.** Data from Dobie et al. (2020), merged and re-clustered. Feature plots showing Arl13b transcript distribution. Each dot represents an individual cell. Colors show the relative expression level, scales available on the right-hand side of each plot. Uninjured: liver cells harvested from control mice. Injured: liver cells collected from fibrotic tissue. Plots in B) are subsets of A). HSC: Hepatic Stellate cell. FB: Fibroblast. VSMC: Vascular smooth muscle cell. Cluster names were carried over from original analysis, while the numbers denote subclusters generated by merging and re-clustering of the data.

## 4.2 Primary Cilia Associated Marker Discovery in the Colon Mesenchyme

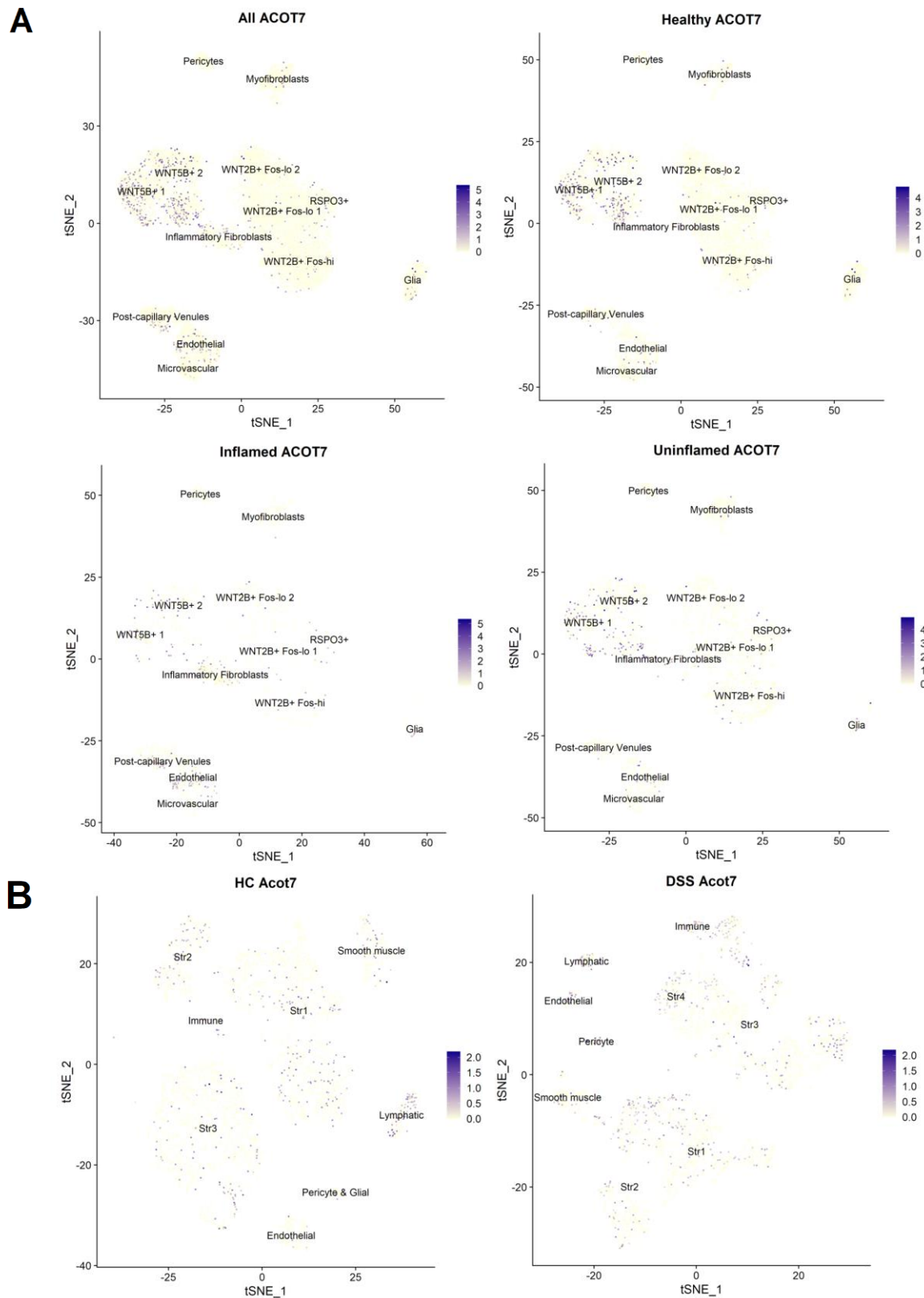
In order to be able to characterize primary cilia expressing cells better, we next looked for markers associated with Arl13b expressing cells in the re-analyzed datasets. From the comparison of healthy and inflamed cells for each cluster of the human colon mesenchyme dataset, we selected two candidate markers to validate in vivo, and follow in different mouse models (Table 4.4). Acot7 was of interest because of its long-chain fatty acid hydrolysing function and association with hyperinsulinemic hypoglycemia. Prrx2 was selected due to its localization to fetal fibroblasts and downregulation in healing dermal tissue.

Next, we investigated the expression and co-localization of these markers with Arl13b and other known markers of the intestinal mesenchyme. Acot7 transcripts were abundant in the WNT5B<sup>+</sup> clusters (Fig. 4.8), thus overlapping with Pdgfra expression (Fig. S5).

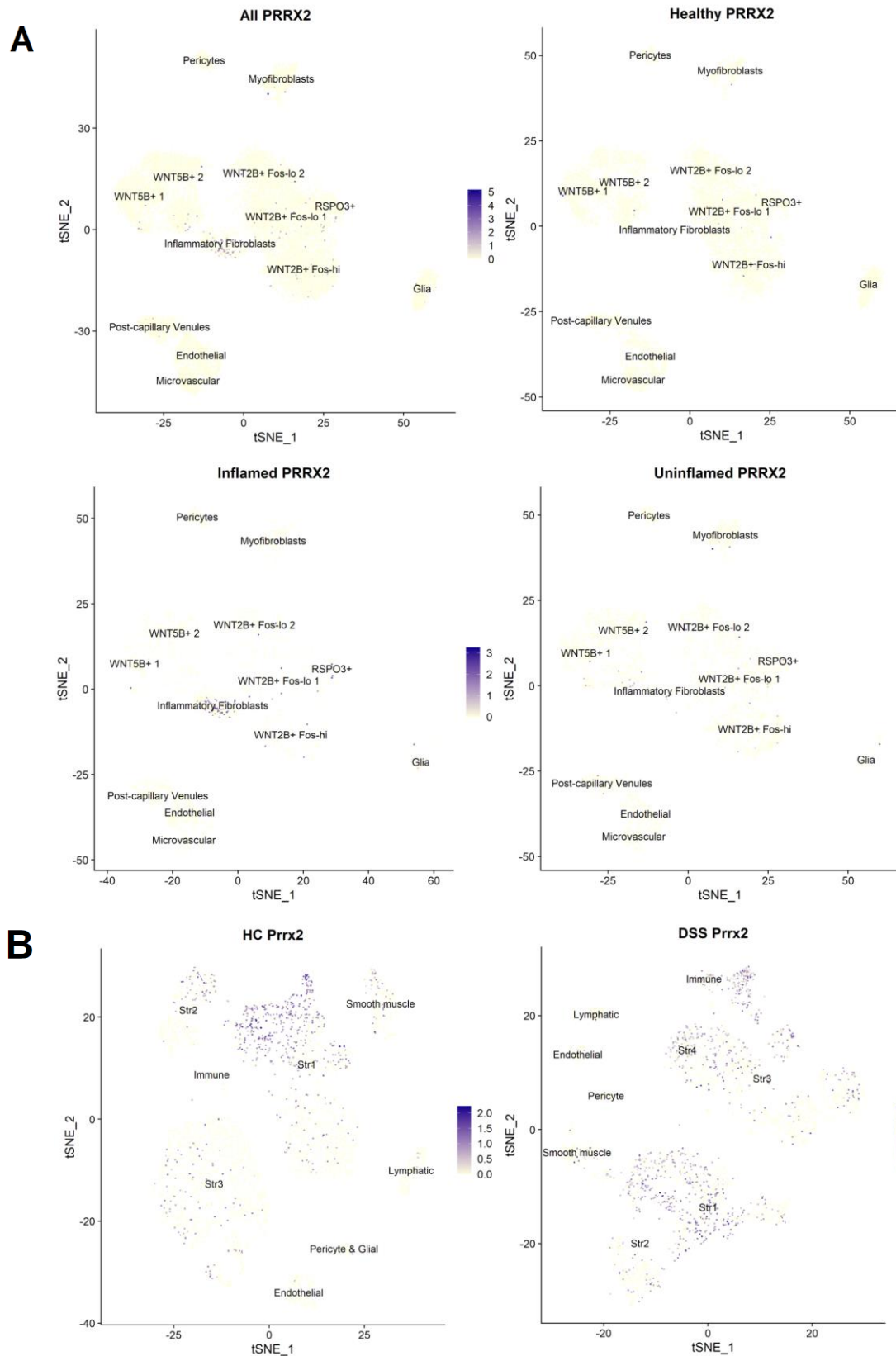
In mouse colon mesenchyme, the Acot7<sup>+</sup> cells appear more diffuse among the fibroblasts. In the DSS-challenged mice, they also appear in the inflammation-associated Str4 and Immune clusters (Fig. 4.9). Like in the human colon, Acot7 expression overlaps with Pdgfra expression in the mouse colon, however, the overlap in Str2, where Pdgfra expression is highest, is not as extensive as in the human colon (Fig. S6). Prrx2 transcripts on the other hand, are most common in inflammatory fibroblasts of the inflamed human colon mesenchyme (Fig. 4.10) and not strongly associated with high Pdgfra expression (Fig. S7). A similar situation is present in the mouse colon mesenchyme (Fig. S8). Prrx2 appears to be expressed more in healthy mouse colon mesenchyme compared to the human, and Str1-type fibroblasts show the most expression. In the DSS-challenged mouse colon mesenchyme, it is seen that Str4 cells also express Prrx2 (Fig. 4.11).

**Table 4.4.** Top 30 most differentially expressed genes in the human colon mesenchyme WNT2B+ Fos-hi cluster between healthy and inflamed datasets. p\_val: P-value calculated by FindMarkers function. P\_val\_adj: P-value adjusted for batch effect. avg\_logFC: average log fold change. Negative values indicate lower expression in the healthy group. pct.1,2: percent of cells showing the gene's expression in healthy or inflamed group, respectively.

<b>WNT2B+ Fos-hi Healthy vs Inflamed</b>					
<b>Gene names</b>	<b>p_val</b>	<b>avg_logFC</b>	<b>pct.1</b>	<b>pct.2</b>	<b>p_val_adj</b>
GMDS	3.05E-12	-1.223895796	0	0.5	4.73E-08
SMCR5	1.03E-09	-1.20761699	0.011	0.5	1.60E-05
EPHA3	1.90E-09	-0.9280970892	0	0.375	2.95E-05
ZNHIT6	1.90E-09	-1.166743388	0	0.375	2.95E-05
F3	4.79E-08	-1.111945106	0.042	0.625	0.0007445118465
SLC16A3	6.02E-08	-0.7500750075	0.021	0.5	0.0009357894997
PHGR1	6.02E-08	-0.9924510831	0.021	0.5	0.0009357894997
SMARCD1	2.82E-07	-1.386407932	0.011	0.375	0.004382280201
INTS8	4.16E-07	-0.7638938538	0.011	0.375	0.00646499042
HMGCS1	4.16E-07	-1.005855979	0.011	0.375	0.00646499042
SSH3	5.04E-07	-0.9337870826	0.011	0.375	0.007836952698
ANKRD13D	1.11E-06	-0.4416745044	0	0.25	0.01720973401
FOXRED1	1.11E-06	-0.4542444861	0	0.25	0.01720973401
RGS11	1.11E-06	-0.4542444861	0	0.25	0.01720973401
RAD9A	1.11E-06	-0.4570059712	0	0.25	0.01720973401
CARD8	1.11E-06	-0.6784851134	0	0.25	0.01720973401
ARMC6	1.11E-06	-0.7584149068	0	0.25	0.01720973401
EHMT2	1.11E-06	-0.7584149068	0	0.25	0.01720973401
CDK20	1.11E-06	-0.7604528887	0	0.25	0.01720973401
DDAH1	1.11E-06	-0.8001296813	0	0.25	0.01720973401
PRRX2	1.11E-06	-0.8001296813	0	0.25	0.01720973401
RAB3A	1.11E-06	-0.8001296813	0	0.25	0.01720973401
ACOT7	1.11E-06	-0.8020844794	0	0.25	0.01720973401
NRBP2	1.11E-06	-0.8108673081	0	0.25	0.01720973401
BCL9L	1.11E-06	-0.9619577866	0	0.25	0.01720973401
NR1D2	1.11E-06	-0.9619577866	0	0.25	0.01720973401
NEK9	1.11E-06	-0.9636207559	0	0.25	0.01720973401
HTR2A	1.11E-06	-0.9710983848	0	0.25	0.01720973401
RP11-736K20.5	1.11E-06	-0.9710983848	0	0.25	0.01720973401



**Figure 4.8. ACOT7 mRNA expression in colon mesenchyme.** Each dot represents an individual cell. Colors show the relative expression level, scales available on the right-hand side of each plot. A) Human colon mesenchyme, data from Smillie et al. (2019). Feature plots of ACOT7 across the whole data t-SNE plot. Plots show cells from all health condition groups shown in Fig. 4.1, All shows the other three plots together. B) Mouse colon mesenchyme, data re-analyzed from Kinchen et al. (2018). t-SNE plot identical to Fig. 4.2, feature plots show Acot7 expression. HC: healthy control, DSS: DSS-challenged. Str: Stromal.

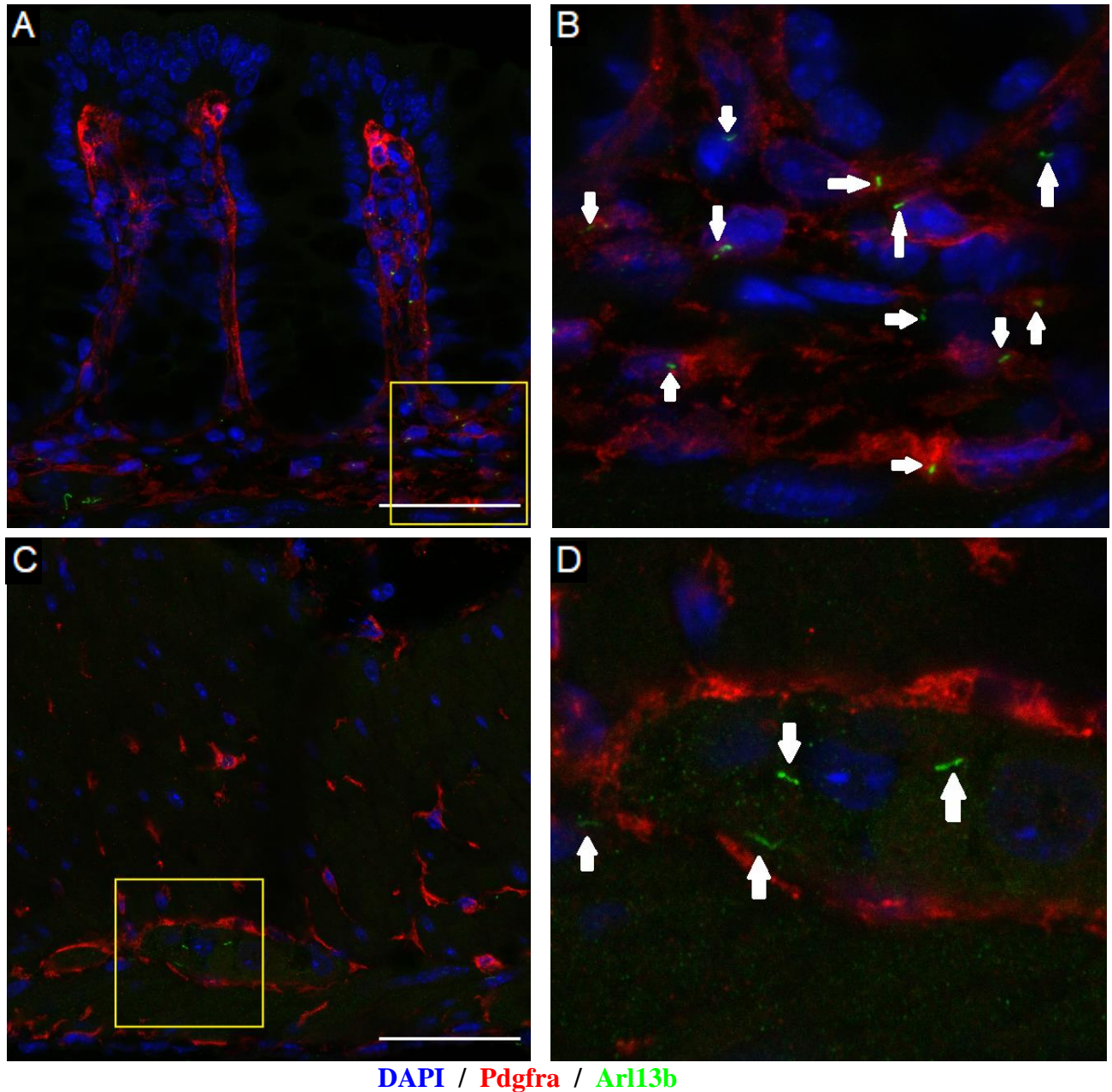


**Figure 4.9. PRRX2 mRNA expression in colon mesenchyme.** Each dot represents an individual cell. Colors show the relative expression level, scales available on the right-hand side of each plot. A) Human colon mesenchyme, data from Smillie et al. (2019). Feature plots of PRRX2 across the whole data t-SNE plot. Plots show cells from all health condition groups shown in Fig. 4.1, All shows the other three plots together. B) Mouse colon mesenchyme, data re-analyzed from Kinchen et al. (2018). t-SNE plot identical to Fig. 4.2, feature plots show Prrx2 expression. HC: healthy control, DSS: DSS-challenged. Str: Stromal.



### 4.3 Validation of the Primary Cilia in The Mouse Colon

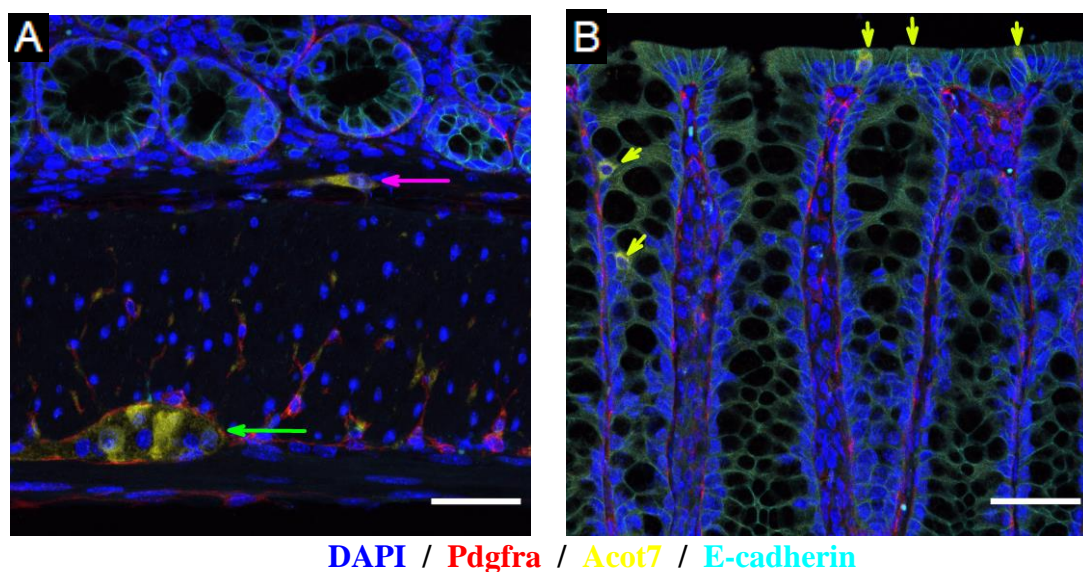
The presence of primary cilia in mouse colon tissue was confirmed using immunofluorescence. Primary cilia were observed in a variety of morphologically different cells and in different layers of the colon mesenchyme. No colon epithelial cell showed primary cilia. The primary cilia could be resolved best with 63x magnification (Figure 4.10).



**Figure 4.10 Appearance of primary cilia in the mouse colon.** Primary cilia appear in the colon mesenchyme, but not in the epithelial layer. Only one primary cilium is present per cell nucleus. A) Crypt structure, from bottom to top. B) Magnification of the area marked with a yellow square in A). C) Primary cilia in the muscle and myenteric plexus. The myenteric plexus is enveloped by Pdgfra. D) Magnification of the area marked with a yellow square in C) Arrows: Primary cilia. Scale bars: 50 μm. Images taken with 63x magnification.

#### 4.3.1 Validation of Candidate Primary Cilia Associated Marker Acot7

The expression of Acot7 in the colon was also confirmed using immunofluorescence. Acot7 expression was located in the myenteric ganglion and within the muscle layer. Some Acot7<sup>+</sup> cells were also located in the submucosal plexus. Lastly, Acot7 expression was also discovered in tuft cells (Figure 4.11). Upon these observations, Acot7 expression at mRNA level was investigated in different scRNA-seq datasets. Acot7's expression pattern was investigated in the mouse colon ENS, where neurons showed more Acot7 transcripts compared to glial cells (Figure S9). In the small intestinal epithelium, the higher expression by the tuft cells was confirmed at the mRNA level as well (Figure S10).



**Figure 4.11. Acot7 expression in the healthy mouse colon.** Acot7-expressing cells were detected in myenteric ganglion, submucosal plexus, muscle and tuft cells. A) Muscle and bottom of crypts. Acot7 expression is seen in the submucosal plexus (pink arrow) and myenteric plexus (green arrow). B) Crypt openings to the lumen. Acot7<sup>+</sup> Tuft cells are highlighted with the yellow arrows. 40x magnification, scale bars: 50  $\mu$ m.

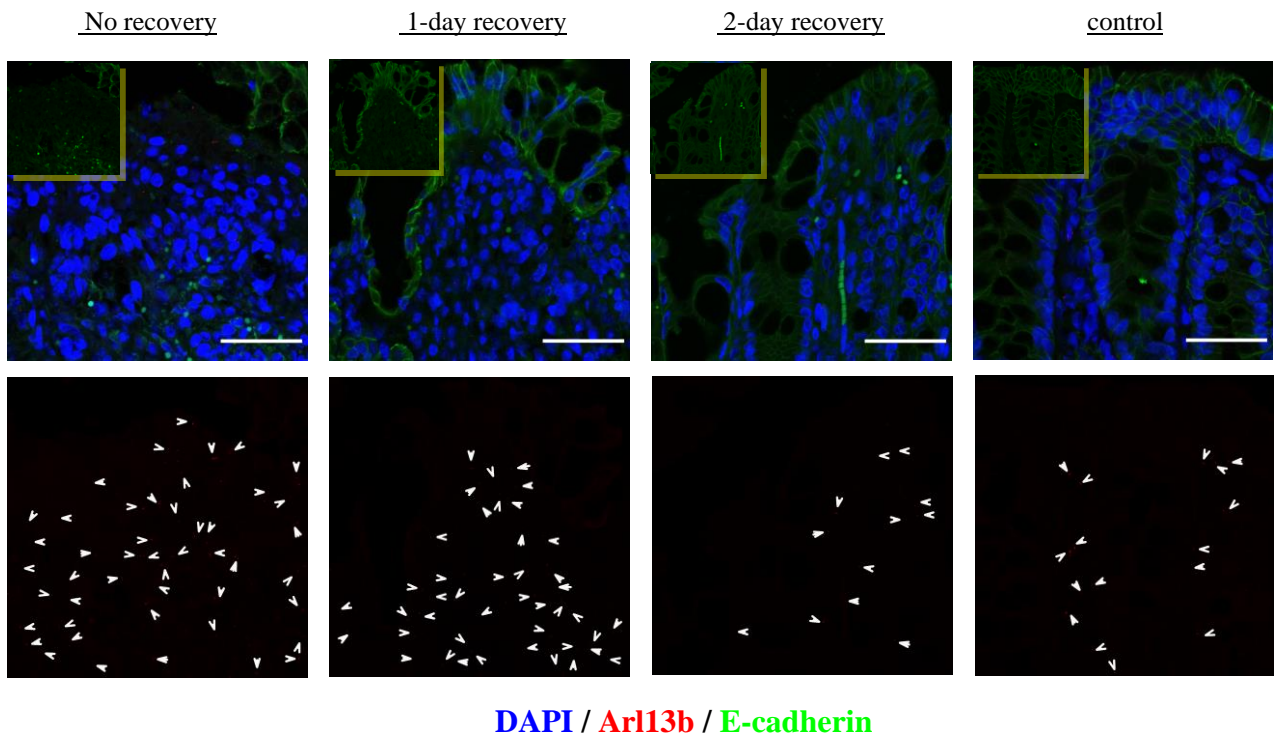
#### 4.4 Mouse DSS-induced Colitis Model

Acute colitis was induced in the mouse colon via administration of 2.5% DSS in drinking water over 5 days. All animals tolerated the DSS challenge. There were no visible changes to the stool consistency, nor any bleeding was detected. The weight of the animals remained mostly unchanged; the greatest loss or gain was 2 grams (Table 3.2). The colons of the animals were shortened (Figure S11).

DSS challenge disrupted the epithelial integrity and crypt structure in all animals. The epithelial loss did not comprise the whole colon, and uninjured crypts were present between injured regions. New crypts were able to emerge after the end of the DSS challenge. Changes to Arl13b abundance and Acot7 morphology were observed over the recovery period.



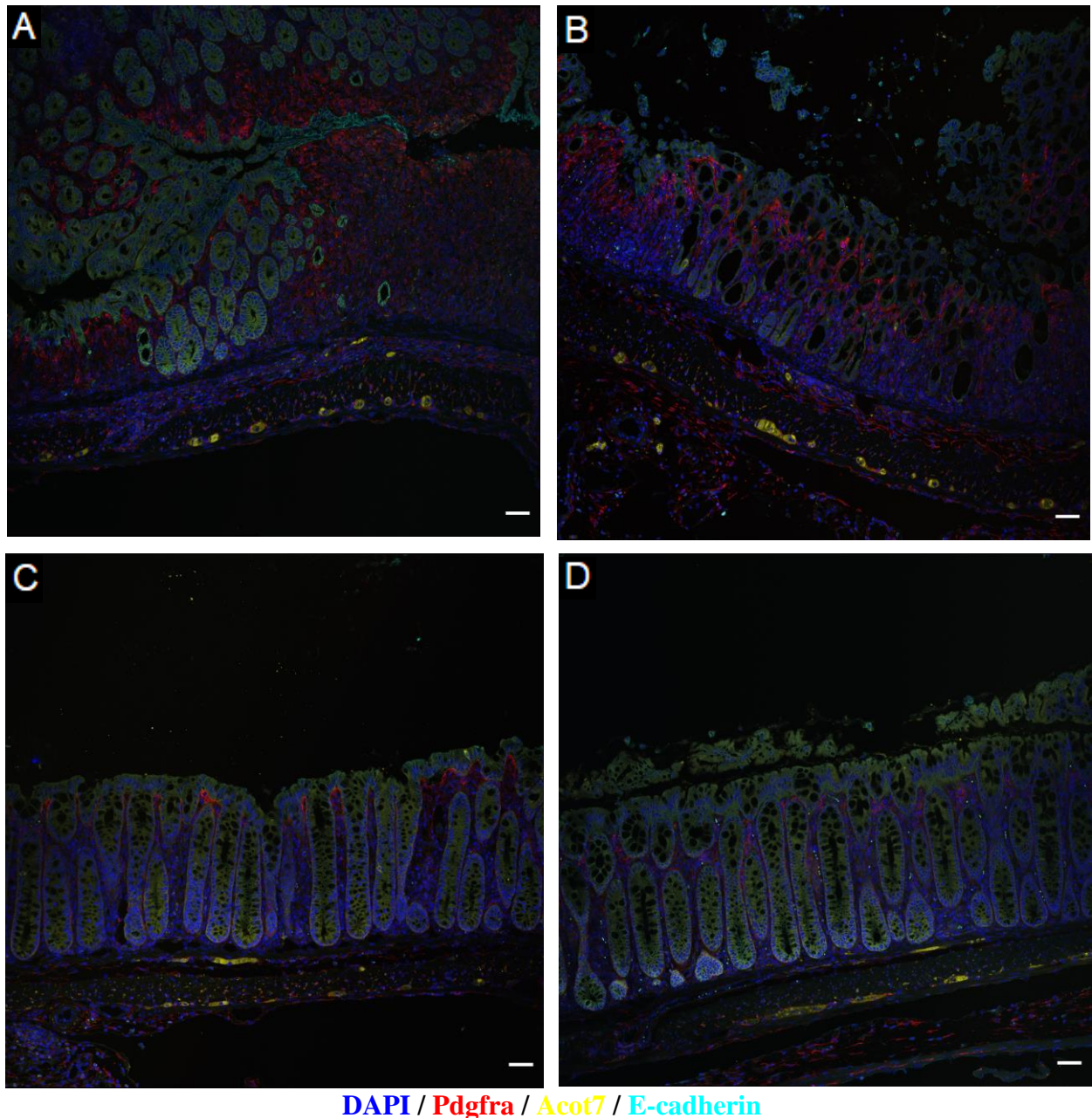
#### 4.4.1 Primary Cilia and Acot7 Response to DSS Challenge



**Figure 4.12. Changes to primary cilia in the crypt top region following the DSS challenge.** The colon lumen is towards the top of the images. Bottom frame shows primary cilia without other fluorescent dyes and the arrows point to individual primary cilia. No recovery shows a colon section obtained from a mouse at the end of the 5-day DSS treatment, 1- and 2-day recovery show sections from mice allowed to recover the respective durations by replacement of DSS solution with normal drinking water. No recovery shows minimum E-cadherin staining, thus a lack of epithelial cells. 1 day recovery shows the re-emergence of epithelial cells before the establishment of the crypt morphology. 2-day recovery shows newly formed crypts narrower than the control.  $n = 2$ . 63x magnification, scale bars: 50  $\mu\text{m}$ .

Since the epithelial layer is lost upon DSS challenge, the invaginations of the crypt structures are also lost, and the area is occupied by mesenchymal tissue instead. This expansion in terms of area also increases the number of primary cilia which can be observed at a glance (Fig. 4.12).

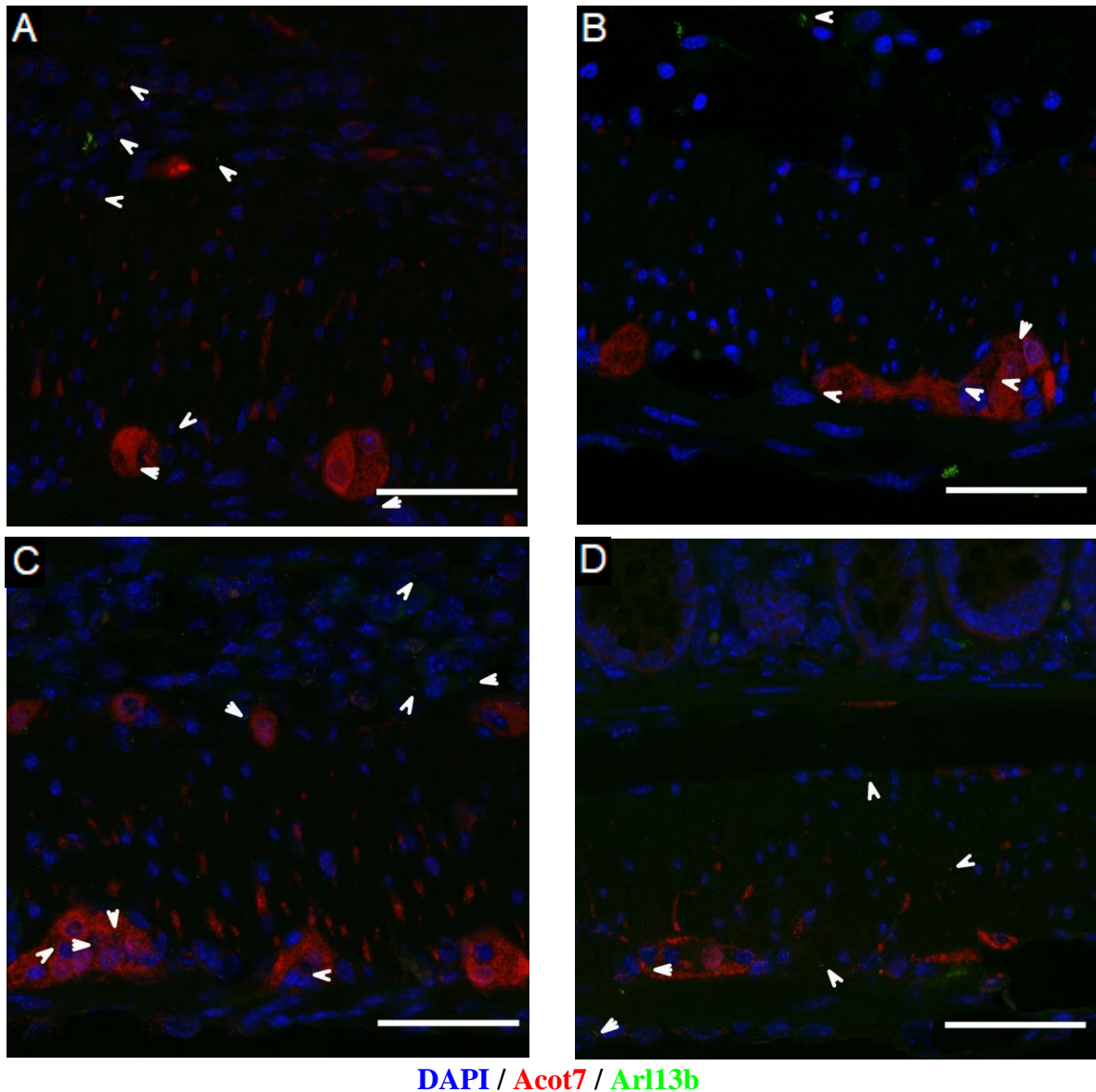
Acot7 in the epithelium is lost upon DSS challenge due to the injury to epithelial cells, including the tuft cells. Myenteric Acot7 expression pattern changes over the course of recovery. Acot7<sup>+</sup> areas in the myenteric plexus appear thicker and rounder compared to the control and appear within smaller distances of each other (Fig. 4.13).



**Figure 4.13. Changes to colon morphology and Acot7 expression following the DSS challenge.** Overview of all colon layers in A) no recovery, B) 1 day recovery, C) 2-day recovery D) control. No recovery means a colon section obtained from a mouse at the end of the 5-day DSS treatment, 1- and 2-day recovery show sections from mice allowed to recover the respective durations by replacement of DSS solution with normal drinking water. In A), crypt structure is lost regionally while some areas retain mature crypts. B) shows new crypt invaginations forming within the mucosal layer. The newly and fully formed crypts can be seen in C). Acot7 in the myenteric plexus appears as round, shortly interspersed regions in A) and B), whereas they appear more elongated in C) and D). 4 images taken with 20x magnification and stitched together. n = 2. Scale bars: 50  $\mu$ m

The patterns of Acot7 expression and primary cilia abundance become more akin to the control within a few days (Fig. 4.12, 4.13). After 3 days of recovery, no injured sections were seen. Primary cilia and Acot7 distribution were mostly identical to the healthy colon (Fig. S12).

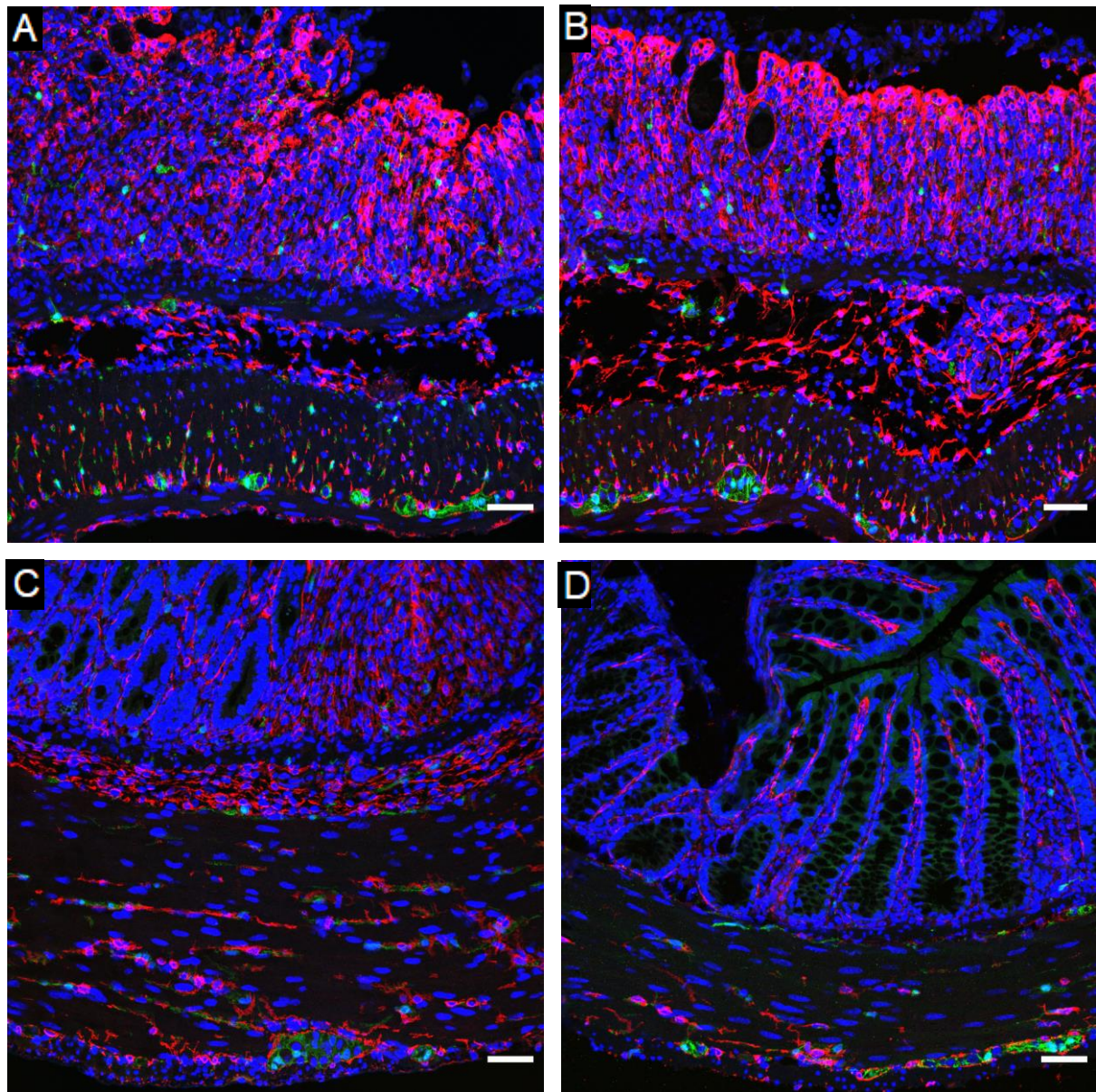
Over the course of DSS treatment and recovery, primary cilia present within the myenteric ganglia do not show any noticeable changes to their abundance or morphology despite the changes observed in the mucosal layer and the morphological changes to Acot7<sup>+</sup> cells (Fig. 4.14).



**Figure 4.14. Primary cilia and Acot7 response to DSS challenge.** The appearance of Acot7 in the myenteric plexus was altered after the DSS challenge. Acot7<sup>+</sup> cells appear thicker compared to control. The distance between two Acot7<sup>+</sup> cell is decreased. The primary cilia distribution within the myenteric plexus and the surrounding muscle do not show any noticeable changes. A) No recovery B) 1-day recovery C) 2-day recovery D) Control. n = 2. 40x magnification, scale bars: 50 μm.



The myenteric ganglion was visualized with the enteric glia marker S100b to verify the Acot7<sup>+</sup> regions' identity in the DSS challenge and recovery response.



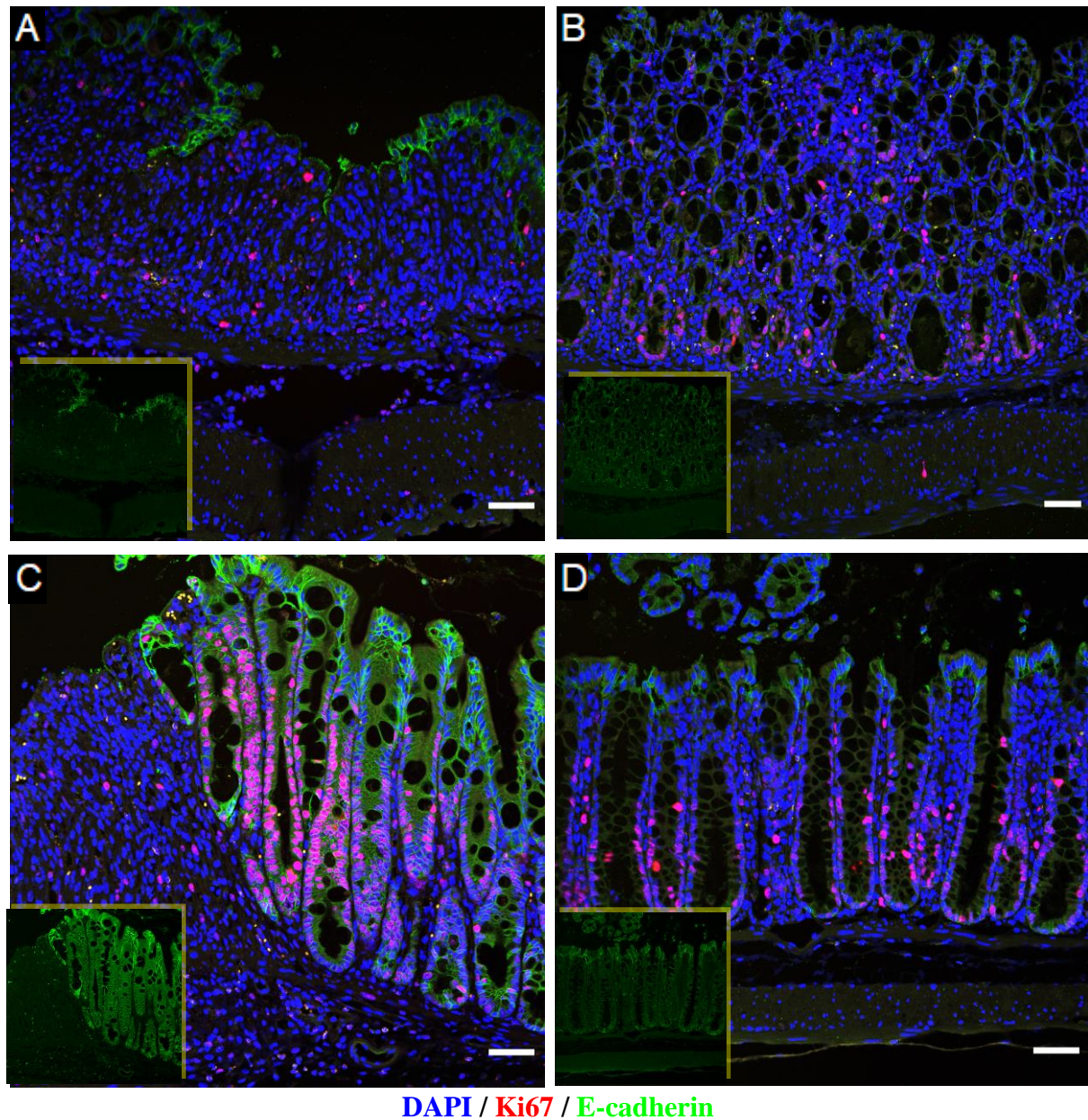
DAPI / Pdgfra / S100b

**Figure 4.15. Myenteric ganglia in DSS challenge and recovery.** S100b marks the glial cells in the myenteric ganglia, which resemble Acot7<sup>+</sup> cells shown in Fig. 4.13 and 4.14. Pdgfra envelopes myenteric ganglia. A) No recovery B) 1-day recovery C) 2-day recovery D) Control. n = 2. 20x magnification, scale bars: 50  $\mu$ m.

Lastly, the proliferation of the epithelial cells required to re-establish the epithelial layer and crypt structure was demonstrated with Ki67 staining. Ki67<sup>+</sup> cells were observed in all stages of recovery from the DSS challenge. While these proliferating cells were scattered in the absence of crypts, epithelial cells forming new crypts starting from 1 day of recovery



followed the crypt structure. Similar to the control, crypts formed within 2 days of recovery had more localized Ki67 expression, with the crypt top cells being Ki67<sup>+</sup>.



**Figure 4.16. Cell proliferation in DSS challenge and recovery.** Ki67 marks the proliferating cells and is localized to the epithelial cells lining the crypts, except the crypt tops where mature cells are typically located. Proliferation activity is high in newly forming crypts after injury. A) No recovery B) 1-day recovery C) 2-day recovery D) Control. n = 2. 20x magnification, scale bars: 50  $\mu$ m.

#### 4.5 Mouse Fasting and Refeeding Model

Mice were fasted over 24 hours and some of the fasted mice were allowed to eat for the following 3 or 6 hours to establish the fasting and refeeding model. All animals tolerated the challenge. The fasted mice had more solid stools compared to other animals. The weights of

each mouse before and after fasting and after refeeding are presented in Table 3.3. All animals had lost weight at the end of the fasting period and regained some of it through refeeding.

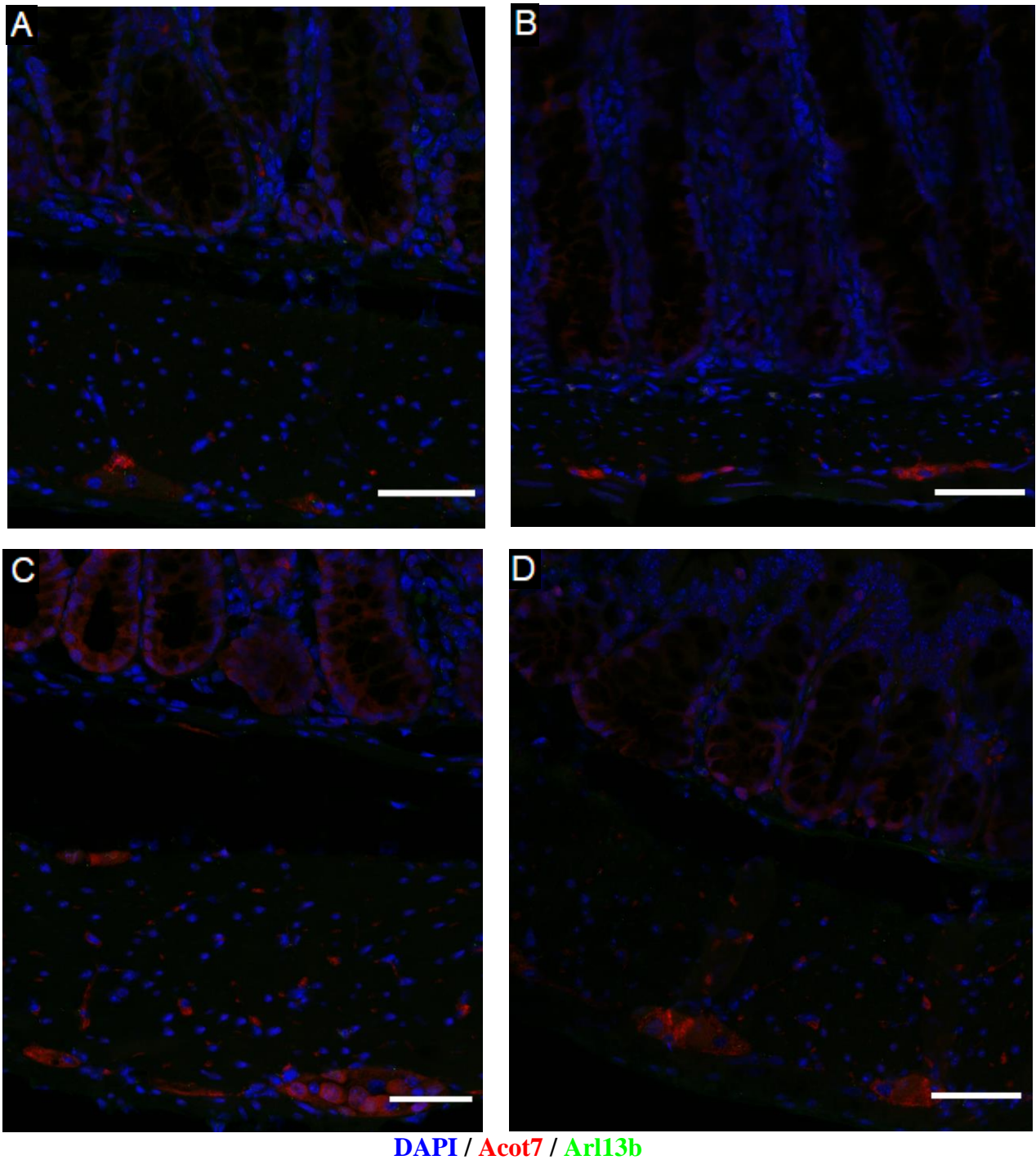
#### **4.5.1 Changes to Primary Cilia and Acot7 in Fasted and Refed Mice**

No changes were observed in either primary cilia presence and abundance, or expression and morphology of Acot7 upon fasting challenge, or the following eating period (Fig. 4.17).

#### **4.6 Mouse High-fat Diet Model**

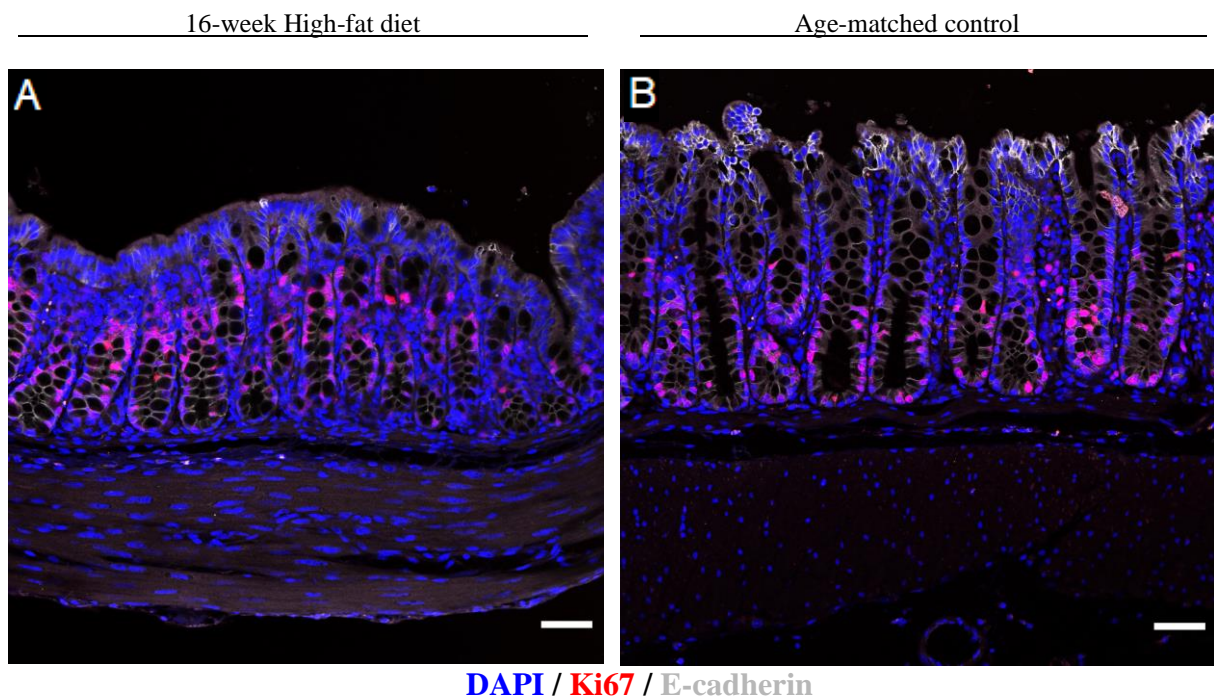
Mice were provided with high-fat diet food instead of regular chow over the course of 2, 8, 12 or 16 weeks. Weights at the time of isolation, including age and sex-matched control group weights, are given in Table 3.4.

The experimental cages on HFD gained more weight compared to the control cages. However, an effect of the diet on the colon was not noticeable prior to the 16th week of HFD (Figures S13-S16). At 16th week of HFD, the colon crypts were shorter in HFD animals compared to the age-matched controls (Fig. 4.18, 4.19 Table S2). Proliferating cells, as indicated by Ki67 expression, were closer to the crypt mouth in HFD colons (Fig. 4.18). Primary cilia could be observed more frequently in HFD-fed colons compared to the controls (Fig. 4.20).

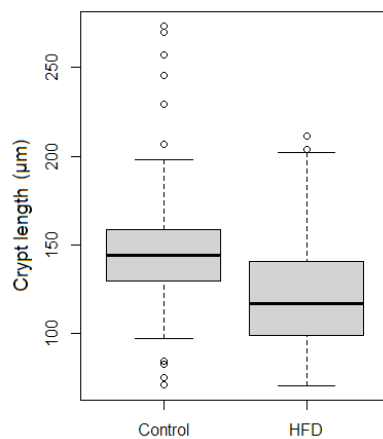


**Figure 4.17. Arl13b and Acot7 in colons of 24 hours fasted and refed mice.** No differences between control fasted, or refed groups were observed. A) Colon after 24 hours of fasting. B) Colon after 3 hours of refeeding following 24-hour fasting. C) Colon after 3 hours of refeeding following 24-hour fasting. D) Control. n = 4. 40x magnification, scale bars: 50  $\mu$ m.



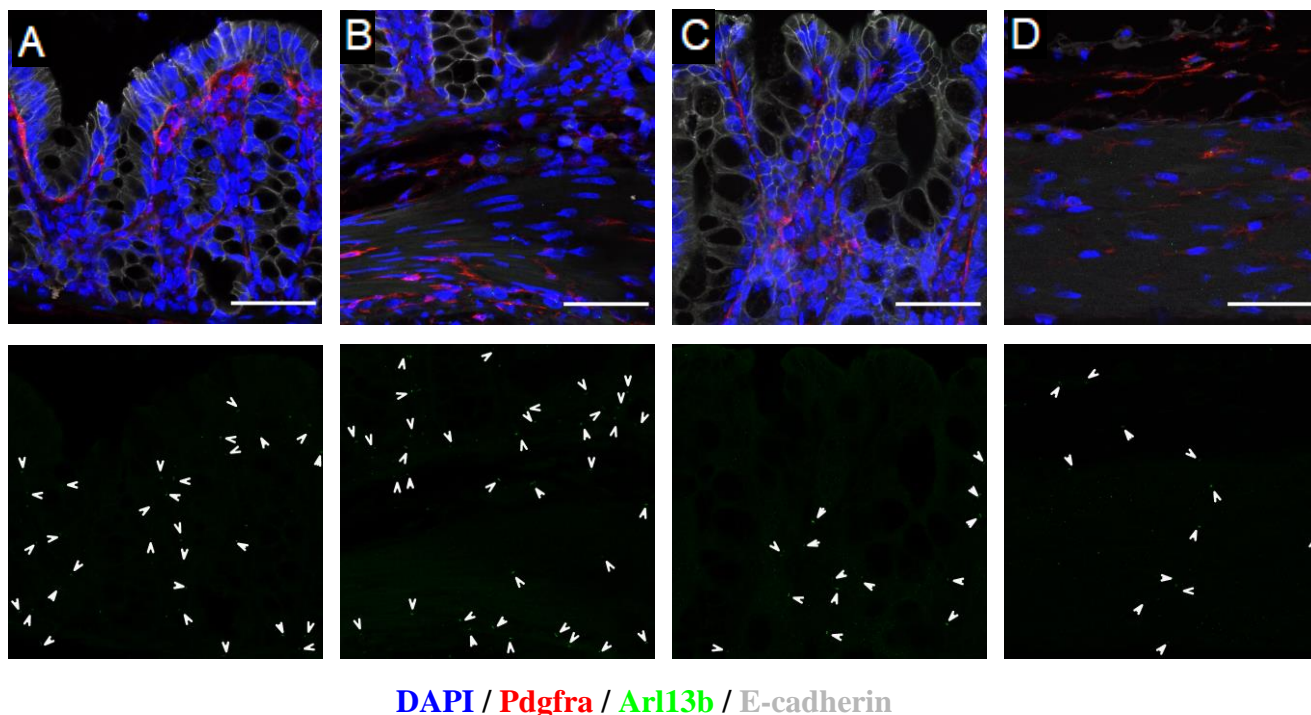


**Figure 4.18. Crypt morphology and Ki67 expression after 16 weeks on a high-fat diet.** A) HFD B) Control. Crypts appear shorter in HFD mice compared to controls. Ki67 expression is restricted to the bottom of the crypt. n = 2. 20x magnification, scale bars: 50  $\mu$ m.



**Figure 4.19. Colonic crypt lengths in 16-week HFD mice and same aged controls.** Measurement of 100 random crypts per group, lengths listed in Table S3. Crypt length is defined as the straight line drawn from the nucleus at the middle of the crypt bottom to the crypt mouth, level with the epithelial cells on either side of the crypt. Bold line shows the average. Crypts in HFD-fed mice colons appear shorter than same-age control mice on average.





**Figure 4.20. Primary cilia expression after 16 weeks on a high-fat diet.** Bottom frame shows primary cilia without other fluorescent dyes and the arrows point to individual primary cilia. A), B), colon of mice fed HFD for 16 weeks. C), D) colon of age-matched control mice. A), C) show the crypt top region, with the colon lumen towards the top of the images. B), D) are images of the muscle and submucosal region with the muscle located at the bottom of the images. Primary cilia appear more abundant in HFD compared to the control group.  $n = 2$ . 63x magnification, scale bars: 50  $\mu\text{m}$ .

#### 4.7 Mouse Western Diet Model

Male and female mice aged 6 or 7 weeks were offered a Western Diet instead of regular chow and weighted biweekly. Over the course of 14 weeks, mice fed WD gained more weight compared to their littermate control groups (Table 3.5).

## 5. DISCUSSION & FUTURE PERSPECTIVES

### 5.1 Primary Cilia Locations and Abundance According to mRNA Expression Profiles

Arl13b, the primary cilia marker used in this study, is found at the mRNA level in all detected mesenchymal cell types of the intestines in a diffuse pattern. This suggests that the primary cilium is not specific to a single cell type, which introduces a challenge to clustering and further identifying the associated cells and uncovering its role in homeostasis, tissue renewal and injury response.

In this study, we have used t-SNE and violin plots to visualize the amount and distribution of Arl13b and other marker genes. While these plots are powerful visualization tools used in many studies [7,12,15,16,28,34,66-70], they can be misleading when being used for making a comparison between multiple studies. Within a singular study, several parameters such as cell size and stage of cell cycle are accounted for using batch correction methods. However, introduction of more parameters such as the number of cells present in a study and the sequencing depth are prone to creating more bias when attempting to compare and contrast different studies, which is an issue identified as an important obstacle before cell atlas type projects [86]. We have noticed that interpreting different treatment groups within a study also requires attention to such batch effects, even though clustering has been performed with care. As a concrete example, Fig 4.1B shows a significant reduction in number of ARL13B expressing WNT<sup>+</sup> cells from a healthy to inflamed state, and an increase of ARL13B transcripts in the Inflammatory Fibroblasts. However, the fraction of cells assigned to these two groups also change between the health states. Therefore, there is not only a reduced number of ARL13B expressing fibroblasts in an inflamed state, there are also less fibroblasts not associated with an inflammatory profile altogether. Similarly, there are only a few inflammatory fibroblasts in the healthy mesenchyme, and the ARL13B expression increases together with the increasing numbers of inflammatory fibroblasts during an ulcerative colitis episode. Therefore, to assist with interpretation of Arl13b expression changes, we also looked at the overall cell numbers, which indicated a small increase in the population of Arl13b expressing cells upon inflammation of the tissue. This trend is present in both human and mouse, which is a promising sign in the mouse model's ability to mimic the human tissue. However, the fraction of Arl13b expressing cells in the mouse is much higher than in the human

tissue, which may still indicate differences in homeostasis and the injury response between the two organisms.

Another source of bias to comparing scRNA-seq dataset comes from the differences in the cell preparation and sorting protocols, which may be impossible to overcome with batch effect correction. One of the requirements of scRNA-seq is the dissociation of the harvested tissue into single cells. This dissociation procedure could be detrimental to locating the primary cilia, since it is an organelle protruding out of the cell, and it is possible that the structure could detach and move with a different neighboring cell. Such a possibility could be investigated to an extent with other techniques used for localizing mRNA within the cells, such as smFISH. Although some *Arl13b* transcripts were spotted in the intestine epithelium scRNA-seq data, we observed no primary cilia extending out of a cell in the epithelial lining in our validation with immunofluorescence, which supports this possibility. However, it cannot be ruled out that *Arl13b* might be expressed independent of a primary cilia [87]. Same concerns also apply to the liver.

The primary cilia of the liver have been previously characterized as an important organelle in monitoring bile. However, single-cell *Arl13b* expression data reveals that they may not be restricted to the cholangiocytes as was previously reported in the literature [58]. Indeed, our lab has validated this observation *in vivo* using immunofluorescence (Data not shown), however, the characterization part of the study is not in scope of this thesis. In light of our observations, we propose that a non-cholangiocyte associated population of primary cilia exists in the liver, which may be contributing to homeostasis and tissue renewal.

## **5.2. Markers of the Primary Cilia Expressing Cells**

Seurat's FindMarkers function is a powerful tool utilized often in biological interpretation of scRNA-seq data. However, the output list of genes requires in-depth literature research and consideration of the research question to select marker genes which could be interesting for the particular study. Since we were aiming to find marker genes whose expression would change in response to injury by colitis, as well as to changes in diet, we chose two candidates which had a role in metabolism and proliferation.

Other marker candidates selected from the same FindMarkers outputs, as well as genes featured in associated literature were also considered by looking at expression patterns in the same manner as demonstrated here. *Pdgfra* has been documented in the recent literature as a

fibroblast marker capable of distinguishing between mesenchymal cells based on its expression level [12,15,16], therefore we tried to utilize its characteristics to identify our populations of interest. However, we were unable to find any distinguishing correlations at the RNA level.

Since Arl13b<sup>+</sup> cells are diverse and do not emerge as a cluster marker itself, no strong candidate gene showing a parallel expression pattern could be determined. In the future this could be remedied by sequencing cells sorted for Arl13b expression beforehand, thus allowing for enhancement of primary cilia possessing cells and possible prevention of contamination by Arl13b mRNA travelling across single cell suspensions. Such a focused dataset could provide more insight into the different populations of cells in the colon mesenchyme which utilize primary cilia. All in all, although re-analysis of public datasets was useful for selecting candidate genes, as well as showing where to expect our markers to be expressed, designing a sequencing experiment tailored to fit a specific research question remains necessary. Moreover, the advances in scRNA-seq technology, such as integration of spatial data, will allow discovery of more features from tissues which remained a mystery due to technical limitations.

### **5.3 Primary Cilia in The Healthy Mouse Colon**

The primary cilia in the mouse colon could be detected using Arl13b specific antibodies. As reported in the literature, the immunofluorescent signal consistently revealed rod-like structures, the tallest reaching only 4  $\mu\text{m}$  in length. Furthermore, no cell, as marked with DAPI staining, possesses more than one cilia, which is another defining feature of the organelle.

Arl13b signals were closely inspected to see which types of cells were the ones possessing the primary cilia. The organelle was clearly associated with the nuclei of mesenchymal cells in most cases. Some primary cilia close to the epithelial layer required closer investigation. As we were able to scan our tissue sections using confocal microscopy, we were able to trace the direction and reach of most primary cilia, which allowed us to conclude that no primary cilium extended from an epithelial cell.

Primary cilia are not exclusive to mesenchymal cells. Albeit in lower amounts, the muscle layers under the mesenchymal cells also showed the presence of primary cilia. Notably, primary cilia were visible within the myenteric plexus, in the myenteric ganglia.

Pdgfra was used as a fibroblast marker to aid in distinguishing primary cilia expressing cells. Although Pdgfra has been associated with the primary cilia, it is much more abundant throughout the mesenchyme. Previously, Pdgfra has been defined as a receptor showing differential expression between different types of fibroblasts in the colon mesenchyme, and the highest expression is located at the crypt top region [16]. However, we did not observe a clear gradient from the crypt top to bottom showing a decrease in Pdgfra expression and were unable to characterize our Arl13b<sup>+</sup> cells based on Pdgfra expression. Moreover, not all Arl13b<sup>+</sup> cells are Pdgfra<sup>+</sup>.

### **5.3.1 Acot7<sup>+</sup> Cells in the Mouse Colon**

Acot7 was chosen as a marker for fibroblast cells responding to inflammation based on the scRNA-seq marker discovery results. However, Acot7<sup>+</sup> cells appeared to be a very different population. The most prominent Acot7 expression was located in the myenteric ganglia, a bundle of nerve cells which reside in the myenteric plexus. Some cells in the submucosa also show Acot7 expression. Moreover, smaller Acot7<sup>+</sup> cells appear within the muscle. While the cells in the myenteric and submucosal plexuses have a round to oblong shape, the cells within the muscle appear thin and long.

Pdgfra appears as a location marker for myenteric ganglia, since each individual ganglion is surrounded by Pdgfra expressing cells. As suggested in the literature, these Pdgfra<sup>+</sup> cells could be acting as the interface between ICC and neurons [31]. Moreover, since both Acot7 and Pdgfra signals within the muscle appear as vertical lines between the two plexuses, it can be suggested that the protrusions of Acot7<sup>+</sup> nerve cells extend into the muscle towards upper layers, and are either accompanied by Pdgfra<sup>+</sup> cells, or express Pdgfra themselves.

Acot7 has been previously identified in the neurons as an important enzyme for regulating lipid retention and metabolism, and its loss has been linked to neurotoxicity. As the enzyme hydrolyzes long chain fatty acids into free fatty acids and coenzyme A [88], it may be serving a role in adjusting the response to the concentration of fatty acids in the intestinal lumen. Moreover, it is known that multiple Acot7 isoforms exist, which may localize in the cytoplasm or mitochondria in different types of cells and tissues [89]. We observed that the Acot7 signal in myenteric ganglia may localize in the nucleus as well, suggesting that multiple isoforms could be expressed in these cells. Roles of the isoforms require further investigation in order to characterize Acot7 in depth.

### **5.3.2 Prrx2 Expression in Colon Mesenchyme**

Prrx2 presence in the mouse colon mesenchyme could not be verified via immunofluorescence. While this can indicate that the mouse colon mesenchymal cells do not express Prrx2, other factors should be considered as well. Prrx2 was selected as a potential primary cilia marker from a human RNA dataset, which leaves the possibility that it may have a species-specific function and warrants an attempt of verification in a human system at the protein level. Another reason is that since Prrx2 is a transcription factor, our methods of tissue fixation and antigen retrieval may be hindering our ability to target Prrx2 molecules immunologically. Lastly, since Prrx2 is expected to increase in UC, it may not be active in the mouse colon mesenchyme without prolonged DSS treatment.

### **5.4 DSS Challenge and Recovery Induced Changes to Primary Cilia and Acot7 Expression**

DSS injures the colonic mesenchyme as a colitogenic chemical. To recover from the injury and rebuild the epithelial layer, the remaining colon tissue needs to promote reconstitution of the epithelial cells.

The colon walls do not become thinner due to loss of the crypts. Non-epithelial cells fill the gaps left from the lost crypts, which can be seen from areas where wounds and uninjured crypts neighbor each other. The distance from the lumen of the colon to the submucosal plexus remains mostly constant. However, structural changes to the colon can be seen at a macroscopic scale. The colon was shortened with DSS challenge and increased in length through recovery. Colon length shortening due to DSS challenges has been extensively reported in the literature, and it has also been reported that mice fed a high-fat diet for a short term can experience more severe length loss when challenged with DSS [90]. Similarly, the generation of new crypts did not change the mucosal thickness.

We observed that Acot7, which we determined as a marker of the myenteric plexus, had a thicker morphology in mice challenged with DSS. This could be either due to an increase in cell size or number within the myenteric ganglia, or a change in the shape of the myenteric plexus itself. If the colon shortening is facilitated by a permanent contraction of the muscles, this could result in the latter possibility. Moreover, the effect of this morphological change in the myenteric ganglia remains a question.

Primary cilia patterning changes with the loss of the epithelial layer. The mesenchymal cells which fill the area left void by the absence of crypts possess primary cilia, which means the number of primary cilia in a given area of the colon is larger in injured tissue compared to the healthy one. This could have several implications and causes. If the primary cilia promote the establishment of an epithelial layer, its presence could be vital for crypt neogenesis. Primary cilia's contribution to this process could be due to its contribution to various regeneration and differentiation pathways. For example, the Hippo pathway is utilized for the initiation of a preliminary epithelial generation, which is patterned via Wnt5a. The primary cilia are known to contribute to both Wnt and Hippo signaling. Moreover, we have shown the co-expression of *Pdgfra* and *Arl13b*. Kinchen et al. have identified a cluster of stromal cells (Str2) which can be identified by both *Pdgfra* and Wnt5a expression [15] It remains to be determined whether the primary cilia play an active role in connecting these various signaling pathways to direct the injury response. Last but not least, the number of primary cilia in an area decreases as the area is covered more by crypts rather than the mesenchymal cells as the colon tissue recovers. This decrease in abundance could be an initiator or result of crypt maturation and could serve to modulate the dosage of some signaling molecules to re-establish essential gradients for cell fate determination, such as Bmp.

### **5.5 Fasting and Refeeding Induced Changes to Primary Cilia and *Acot7* Expression**

The colon did not show any detectable changes to adapt for fasting or the following feeding period. The primary cilia presence was indifferent across all specimens. *Acot7* expression and the appearance of myenteric ganglia also did not change across the experimental or control groups. As the final portion of the gastrointestinal tract and the organ mostly responsible for water resorption as opposed to nutrient absorption happening in the upper regions, it was expected that the colon would not respond to being fasted. However, our observations do not conclude that the colon is indifferent towards fasting and refeeding. The fasted mice had more solid stool compared to other groups, which indicates that the fasted mice also consumed less water. It is likely that the reduced availability of water may have induced a response in the colon out of scope of this work.

## 5.6 High-fat Diet Induced Changes to the Colon

HFD resulted in accelerated and amplified weight gain in male mice. Although the difference in weight gain could be seen between 2 to 8 weeks of diet change, it took until 16th week before morphological changes were observed in the colon. These morphological changes were accompanied by changes in the gene expression profile as well.

Our model includes female mice as well, however, they were not included in this thesis. Female mice gain less weight and at a slower pace [53], therefore we opted to observe the effects of HFD on female mice starting from later time points. Some animals were sacrificed due to complications with ulcerative dermatitis [55].

We have observed that the colonic crypts are shorter in 16-week male HFD mice compared to the control mice. Moreover, Ki67-expressing cells could be observed closer to the crypt mouth compared to the age-matched control. The crypt shortening could be due to a decrease in the mature epithelial cell population, rather than a shrinkage in the stem cell region. Studies using different models also reported a shortening of the crypts both in the colon and the small intestine. Furthermore, it has been suggested that intestinal stem cells and proliferating cells increase in numbers in HFD, and result in changes to crypt function [46,54]. The shortening of the crypt size despite an increase in proliferation has been justified with an increase in apoptosis [54]. With our model we show that the same changes occur in male mice fed 16 weeks of HFD starting from 6-8 weeks of age.

The primary cilia of the colon also showed an increase in numbers at the same time as the morphological changes were observed. This response suggests that the primary cilia may indeed play a role in proliferation, such that their numbers reflect the rate of proliferation, or differentiation, in the tissue. Unlike in the DSS model, the epithelium and the crypt structure are not lost upon HFD consumption, therefore the increase could not be justified with an areal expansion of the mesenchyme. It is a possibility that the increase in primary cilia numbers, or abundance, could be necessary to generate the same injury response to DSS treatment and HFD, since obesity is considered a chronic inflammation state, similar to how UC is a chronic condition with acute inflammatory relapses [52]. Moreover, the two models appear to have a cumulative impact on the tissue [90]. However, the role of the primary cilia in colon homeostasis needs to be uncovered first in order to suggest a common injury response mechanism. Therefore, a detailed survey of all known primary cilia associated pathways in context of the colonic crypts is required.



It must be noted that our observations were qualitative, not quantitative, and a robust method of quantification needs to be applied to be able to show a true correlation between primary cilia numbers, size of the intestinal stem cell niche, and the crypt size. Since primary cilia is a very small organelle, there is a risk that sectioning the tissue results in the loss of primary cilium of a cell and its emergence in the next section, therefore methods which allow observation of the tissue as a whole are needed. Likewise, characterization of receptor-signal interactions on the primary cilia will pose a challenge for the discovery of the colonic primary cilia associated signaling pathways. However, documentation of the primary cilia associated functions and adaptations in homeostasis and disease can open up novel treatment targets in the future.

### **5.7 Western Diet Model**

The Western diet is an alternative to the HFD in mice. Due to time constraints, the effects of WD in mice could not be documented at the tissue level. However, WD-fed mice show an accelerated weight gain compared to the control, similar to what was observed in HFD mice. Therefore, we expect that WD will induce similar changes as HFD. The lower fat content of WD could mean that the effects will appear at a later time point than HFD. However, since WD is also a high carbohydrate diet, some unexpected changes may occur as well. Using these two models of obesity could allow for the demonstration of effects exacerbated by sugar and fat more distinctively and allow for characterizing how the source of the calories taken into the body affects the tissues in correlation with the total body mass.

## APPENDIX

### A. Codes Used for scRNA-seq Data Analysis

The codes used for t-SNE and Violin plots were adapted from A.1.3 to suit the object structure.

#### A.1 Analyzing Smillie et al. (2019), Intra- and Inter-cellular Rewiring of the Human Colon During Ulcerative Colitis

##### A.1.1 Object construction

Seurat v3 objects were generated using the code and data published by the authors, which includes PCA clustering and t-SNE dimensionality reduction. No adjustments were made to the code.

##### A.1.2 Arl13b+ marker discovery

```
1 # ARL13B+ cluster marker searcher in lamina propria
2 # Separates cells into health subsets for only ARL13B+ cells
3 # Then into clusters for each group
4 # Take cell names as vectors
5 # Find markers for 1 health state in 1 cluster vs other 2 health states in same cluster
6 # Save cell name list as well as Marker output as .csv
7
8 # Deniz Esen      26.01.2020
9 # Edited 5.3.2020, now checks for markers in permutation
10 # R version 3.6.2 (2019-12-12) -- "Dark and Stormy Night"
11
12 # Get separation function and separate Seurat object for health state
13 library(Seurat)
14 source('drawfun.r')
15 fib.seur <- readRDS("fibroblast/fib.seur.RDS")
16
17 # Subset for ARL13B
18 fib.seur.ARL13B <- subset(fib.seur, subset = ARL13B >0)
19
20 # Get Cluster names
21 ClusterNames <- levels(fib.seur@meta.data$Cluster)
22
23 # Run for all clusters, manually continue loops after clusters it won't run for
24 # Healthy 7th cluster (Inflammatory Fibroblasts) is missing, continue from 8
25 # Inflamed 4th cluster (RSPO3+) is missing, continue from 5
26 # Uninflamed 13th (last) cluster (Glia) is missing, no extra loops needed
27
28 #####
29 # Healthy vs Inflamed
```

```

30 for (i in 1:length(ClusterNames)){
31   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Healthy", ident.2 = "Inflamed",
   group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
32   write.csv(markers, file =
   paste("fibroblast/healthMarkers/healthy/ARL13B+_Healthy_vs_Inflamed_",
   ClusterNames[i], ".csv", collapse = ""))
33 }
34 for (i in 5:length(ClusterNames)){
35   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Healthy", ident.2 = "Inflamed",
   group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
36   write.csv(markers, file =
   paste("fibroblast/healthMarkers/healthy/ARL13B+_Healthy_vs_Inflamed_",
   ClusterNames[i], ".csv", collapse = ""))
37 }
38 for (i in 8:length(ClusterNames)){
39   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Healthy", ident.2 = "Inflamed",
   group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
40   write.csv(markers, file =
   paste("fibroblast/healthMarkers/healthy/ARL13B+_Healthy_vs_Inflamed_",
   ClusterNames[i], ".csv", collapse = ""))
41 }
42
43 # Healthy vs Uninflamed
44 for (i in 1:length(ClusterNames)){
45   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Healthy", ident.2 = "Uninflamed",
   group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
46   write.csv(markers, file =
   paste("fibroblast/healthMarkers/healthy/ARL13B+_Healthy_vs_Uninflamed_",
   ClusterNames[i], ".csv", collapse = ""))
47 }
48 for (i in 8:length(ClusterNames)){
49   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Healthy", ident.2 = "Uninflamed",
   group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
50   write.csv(markers, file =
   paste("fibroblast/healthMarkers/healthy/ARL13B+_Healthy_vs_Uninflamed_",
   ClusterNames[i], ".csv", collapse = ""))
51 }
52
53 #####
54 # Repeat for Inflamed
55 # Inflamed vs Healthy
56 for (i in 1:length(ClusterNames)){
57   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Inflamed", ident.2 = "Healthy",
   group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
58   write.csv(markers, file =
   paste("fibroblast/healthMarkers/inflamed/ARL13B+_Inflamed_vs_Healthy_",
   ClusterNames[i], ".csv", collapse = ""))
59 }
60 for (i in 5:length(ClusterNames)){

```

```

61 markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Inflamed", ident.2 = "Healthy",
62   group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
63 write.csv(markers, file =
64   paste("fibroblast/healthMarkers/inflamed/ARL13B+_Inflamed_vs_Healthy",
65     ClusterNames[i], ".csv", collapse = ""))
66 }
67 for (i in 8:length(ClusterNames)){
68   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Inflamed", ident.2 = "Healthy",
69     group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
70   write.csv(markers, file =
71     paste("fibroblast/healthMarkers/inflamed/ARL13B+_Inflamed_vs_Healthy",
72       ClusterNames[i], ".csv", collapse = ""))
73 }
74 # Inflamed vs Uninflamed
75 for (i in 1:length(ClusterNames)){
76   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Inflamed", ident.2 =
77     "Uninflamed", group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
78   write.csv(markers, file =
79     paste("fibroblast/healthMarkers/inflamed/ARL13B+_Inflamed_vs_Uninflamed_",
80       ClusterNames[i], ".csv", collapse = ""))
81 }
82 for (i in 5:length(ClusterNames)){
83   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Inflamed", ident.2 =
84     "Uninflamed", group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
85   write.csv(markers, file =
86     paste("fibroblast/healthMarkers/inflamed/ARL13B+_Inflamed_vs_Uninflamed",
87       ClusterNames[i], ".csv", collapse = ""))
88 }
89 }
90 #####
91 # Repeat for Uninflamed
92 # Uninflamed vs Healthy
93 for (i in 1:length(ClusterNames)){
94   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Uninflamed", ident.2 = "Healthy",
95     group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
96   write.csv(markers, file =
97     paste("fibroblast/healthMarkers/uninflamed/ARL13B+_Uninflamed_vs_Healthy",
98       ClusterNames[i], ".csv", collapse = ""))
99 }
100 for (i in 8:length(ClusterNames)){
101   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Uninflamed", ident.2 = "Healthy",
102     group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
103   write.csv(markers, file =
104     paste("fibroblast/healthMarkers/uninflamed/ARL13B+_Uninflamed_vs_Healthy",
105       ClusterNames[i], ".csv", collapse = ""))
106 }
107 }
108 # Uninflamed vs Inflamed
109 for (i in 1:length(ClusterNames)){

```

```

93 markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Uninflamed", ident.2 =
  "Inflamed", group.by = "Health", subset.ident = ClusterNames[i], min.pct = 0.01)
94 write.csv(markers, file =
  paste("fibroblast/healthMarkers/uninflamed/ARL13B+_Uninflamed_v
    s_Inflamed", ClusterNames[i], ".csv", collapse = ""))
95 }
96 for (i in 5:length(ClusterNames)){
97   markers <- FindMarkers(fib.seur.ARL13B, ident.1 = "Uninflamed",
    ident.2 = "Inflamed", group.by = "Health", subset.ident =
    ClusterNames[i], min.pct = 0.01)
98   write.csv(markers, file =
    paste("fibroblast/healthMarkers/uninflamed/ARL13B+_Uninflamed_v
      s_Inflamed", ClusterNames[i], ".csv", collapse = ""))
99 }

```

### A.1.3 Short script to paint genes of interest

```

1 # Type gene names here:
2 genes = c("ACOT7", "ARL13B", "CDH3", "FABP4", "GLI1", "GREM1", "IGFBP7",
  "PDGFRA", "PRRX2", "TUBA4A", "WNT4", "EHD3", "DAAM1", "LYVE1", "PCNT",
  "IFT88")
3
4 library(Seurat)
5 library(ggplot2)
6
7 # Get Seurat Object
8 fib.seur <- readRDS("fibroblast/fib.seur.RDS")
9
10 # Check whether all genes are present in data
11 for (i in 1:length(genes))
12   print(which(fib.seur@assays$RNA@counts@Dimnames[[1]] == genes[i]))
13
14 # divide Seurat object into separate health states
15 health <- list(subset(fib.seur, subset = Health == "Healthy"),
  subset(fib.seur, subset = Health == "Inflamed"),
  subset(fib.seur, subset = Health == "Uninflamed"))
16 # subset for ARL13B, divide into health states
17 fib.seur.ARL13B <- subset(fib.seur, subset = ARL13B > 0)
18 ARL13B.health <- list( subset(fib.seur.ARL13B, subset = Health == "Healthy"),
1. subset(fib.seur.ARL13B, subset = Health == "Inflamed"),
2. subset(fib.seur.ARL13B, subset = Health == "Uninflamed"))
19
20 # Lazily making new folders on Desktop to save everything
21 dir.create("All")
22 dir.create("ARL13B")
23
24 # Draw and save the tSNE plots
25 for (i in 1:length(genes)){
26   ggsave(filename = paste("tSNE_All_", genes[i], ".png", collapse = "", sep = ""),

```

```

27     plot = FeaturePlot(fib.seur, genes[i], pt.size = 0.5, cols = c("lightyellow",
"darkblue"), label = T) + ggtitle(paste("All", genes[i])),
28     path = "All", device = "png")
29     ggsave(filename = paste("tSNE_Healthy_", genes[i], ".png", collapse = ""),
30     plot = FeaturePlot(health[[1]], genes[i], pt.size = 0.5, cols = c("lightyellow",
"darkblue"), label = T) + ggtitle(paste("Healthy", genes[i])), path = "All")
31     ggsave(filename = paste("tSNE_Inflamed_", genes[i], ".png", collapse = ""),
32     plot = FeaturePlot(health[[2]], genes[i], pt.size = 0.5, cols = c("lightyellow",
"darkblue"), label = T) + ggtitle(paste("Inflamed", genes[i])), path = "All")
33     ggsave(filename = paste("tSNE_Uninflamed_", genes[i], ".png", collapse = ""),
34     plot = FeaturePlot(health[[3]], genes[i], pt.size = 0.5, cols = c("lightyellow",
"darkblue"), label = T) + ggtitle(paste("Uninflamed", genes[i])), path = "All")
35
36     ggsave(filename = paste("tSNE_ARL13B+", genes[i], ".png", collapse = ""),
37     plot = FeaturePlot(fib.seur.ARL13B, genes[i], pt.size = 0.5, cols = c("lightyellow",
"darkblue"), label = T) + ggtitle(paste("ARL13B+", genes[i])), path = "ARL13B")
38     ggsave(filename = paste("tSNE_ARL13B+_Healthy_", genes[i], ".png", collapse = ""),
39     plot = FeaturePlot(ARL13B.health[[1]], genes[i], pt.size = 0.5, cols =
c("lightyellow", "darkblue"), label = T) + ggtitle(paste("ARL13B+ Healthy", genes[i])),
path = "ARL13B")
40     ggsave(filename = paste("tSNE_ARL13B+_Inflamed_", genes[i], ".png", collapse = ""),
41     plot = FeaturePlot(ARL13B.health[[2]], genes[i], pt.size = 0.5, cols =
c("lightyellow", "darkblue"), label = T) + ggtitle(paste("ARL13B+ Inflamed", genes[i])),
path = "ARL13B")
42     ggsave(filename = paste("tSNE_ARL13B+_Uninflamed_", genes[i], ".png", collapse =
""),
43     plot = FeaturePlot(ARL13B.health[[3]], genes[i], pt.size = 0.5, cols =
c("lightyellow", "darkblue"), label = T) + ggtitle(paste("ARL13B+ Uninflamed",
genes[i])), path = "ARL13B")
44
45     # Also Violin plots for each
46     ggsave(filename = paste("Violin_All_", genes[i], ".png", collapse = ""),
47     plot = VlnPlot(fib.seur, genes[i]) + ggtitle(paste("All", genes[i], collapse = "")) +
theme(legend.position = 'none'),
48     device = "png", path = "All", width = 20, height = 11.25, units = "in")
49     ggsave(filename = paste("Violin_Healthy_", genes[i], ".png", collapse = ""),
50     plot = VlnPlot(health[[1]], genes[i]) + ggtitle(paste("Healthy", genes[i], collapse =
"")) + theme(legend.position = 'none'),
51     device = "png", path = "All", width = 20, height = 11.25, units = "in")
52     ggsave(filename = paste("Violin_Inflamed_", genes[i], ".png", collapse = ""),
53     plot = VlnPlot(health[[2]], genes[i]) + ggtitle(paste("Inflamed", genes[i], collapse =
"")) + theme(legend.position = 'none'),
54     device = "png", path = "All", width = 20, height = 11.25, units = "in")
55     ggsave(filename = paste("Violin_Uninflamed_", genes[i], ".png", collapse = ""),
56     plot = VlnPlot(health[[3]], genes[i]) + ggtitle(paste("Uninflamed", genes[i], collapse
= "")) + theme(legend.position = 'none'),
57     device = "png", path = "All", width = 20, height = 11.25, units = "in")
58
59     ggsave(filename = paste("Violin_ARL13B+", genes[i], ".png", collapse = ""),

```

```

60     plot = VlnPlot(fib.seur.ARL13B, genes[i]) + ggtitle(paste("All", genes[i], collapse =
    "")) + theme(legend.position = 'none'),
61     device = "png", path = "ARL13B", width = 20, height = 11.25, units = "in")
62     ggsave(filename = paste("Violin_ARL13B+_Healthy_", genes[i], ".png", collapse = ""),
63     plot = VlnPlot(ARL13B.health[[1]], genes[i]) + ggtitle(paste("Healthy", genes[i],
    collapse = "")) + theme(legend.position = 'none'),
64     device = "png", path = "ARL13B", width = 20, height = 11.25, units = "in")
65     ggsave(filename = paste("Violin_ARL13B+_Inflamed_", genes[i], ".png", collapse = ""),
66     plot = VlnPlot(ARL13B.health[[2]], genes[i]) + ggtitle(paste("Inflamed", genes[i],
    collapse = "")) + theme(legend.position = 'none'),
67     device = "png", path = "ARL13B", width = 20, height = 11.25, units = "in")
68     ggsave(filename = paste("Violin_ARL13B+_Uninflamed_", genes[i], ".png", collapse =
    "")),
69     plot = VlnPlot(ARL13B.health[[3]], genes[i]) + ggtitle(paste("Uninflamed",
    genes[i], collapse = "")) + theme(legend.position = 'none'),
70     device = "png", path = "ARL13B", width = 20, height = 11.25, units = "in")
71 }
72
73 #####
74 # Antibody combinations
75 rabbit <- c("ACOT7", "FABP4", "GLI1", "IGFBP7", "PRRX2", "EHD3", "DAAM1",
    "PCNT", "IFT88")
76 goat <- c("GREM1", "PDGFRA", "WNT4")
77 mouse <- c("TUBA4A", "CDH1")
78 rat <- c("LYVE1")
79
80 dir.create("Combinations")
81 # Draw and save the tSNE plots
82 for (i in 1:length(rabbit)){
83   # rabbit&goat
84   for (j in 1:length(goat)){
85     ggsave(filename = paste("All", rabbit[i], goat[j], ".png", collapse = "", sep = " "),
86     plot = FeaturePlot(fib.seur, c(rabbit[i], goat[j]), blend = T, cols = c("lightyellow",
    "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] + ggtitle(paste("All",
    rabbit[i], "(blue)", goat[j], "(red)")),
87     path = "Combinations", device = "png")
88     ggsave(filename = paste("Inflamed", rabbit[i], goat[j], ".png", collapse = "", sep = " "),
89     plot = FeaturePlot(health[[2]], c(rabbit[i], goat[j]), blend = T, cols =
    c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
    ggtitle(paste("Inflamed", rabbit[i], "(blue)", goat[j], "(red)")), path = "Combinations",
    device = "png")
90     ggsave(filename = paste("Uninflamed", rabbit[i], goat[j], ".png", collapse = "", sep = "
    ")),
91     plot = FeaturePlot(health[[3]], c(rabbit[i], goat[j]), blend = T, cols =
    c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
    ggtitle(paste("Uninflamed", rabbit[i], "(blue)", goat[j], "(red)")), path = "Combinations",
    device = "png")
92   # Skipping Grem1 in healthy, expression is 0
93   if(goat[j] == "GREM1")
94     next

```

```

95  ggsave(filename = paste("Healthy", rabbit[i], goat[j], ".png", collapse = "", sep = " "),
96      plot = FeaturePlot(health[[1]], c(rabbit[i], goat[j]), blend = T, cols =
      c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
      ggtitle(paste("Healthy", rabbit[i], "(blue)", goat[j], "(red)")), path = "Combinations",
      device = "png")
97  }
98  #rabbit&mouse
99  for (j in 1:length(mouse)){
100     ggsave(filename = paste("All", rabbit[i], mouse[j], ".png", collapse = "", sep = " "),
101         plot = FeaturePlot(fib.seur, c(rabbit[i], mouse[j]), blend = T, cols =
            c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
            ggtitle(paste("All", rabbit[i], "(blue)", mouse[j], "(red)")), path = "Combinations", device
            = "png")
102     ggsave(filename = paste("Healthy", rabbit[i], mouse[j], ".png", collapse = "", sep = "
        "),
103         plot = FeaturePlot(health[[1]], c(rabbit[i], mouse[j]), blend = T, cols =
            c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
            ggtitle(paste("Healthy", rabbit[i], "(blue)", mouse[j], "(red)")), path = "Combinations",
            device = "png")
104     ggsave(filename = paste("Inflamed", rabbit[i], mouse[j], ".png", collapse = "", sep = "
        "),
105         plot = FeaturePlot(health[[2]], c(rabbit[i], mouse[j]), blend = T, cols =
            c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
            ggtitle(paste("Inflamed", rabbit[i], "(blue)", mouse[j], "(red)")), path = "Combinations",
            device = "png")
106     ggsave(filename = paste("Uninflamed", rabbit[i], mouse[j], ".png", collapse = "", sep = "
        "),
107         plot = FeaturePlot(health[[3]], c(rabbit[i], mouse[j]), blend = T, cols =
            c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
            ggtitle(paste("Uninflamed", rabbit[i], "(blue)", mouse[j], "(red)")), path =
            "Combinations", device = "png")
108  }
109  #rabbit&rat
110  for (j in 1:length(rat)){
111     ggsave(filename = paste("All", rabbit[i], rat[j], ".png", collapse = "", sep = " "),
112         plot = FeaturePlot(fib.seur, c(rabbit[i], rat[j]), blend = T, cols = c("lightyellow",
            "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] + ggtitle(paste("All",
            rabbit[i], "(blue)", rat[j], "(red)")), path = "Combinations", device = "png")
113     ggsave(filename = paste("Healthy", rabbit[i], rat[j], ".png", collapse = "", sep = " "),
114         plot = FeaturePlot(health[[1]], c(rabbit[i], rat[j]), blend = T, cols = c("lightyellow",
            "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] + ggtitle(paste("Healthy",
            rabbit[i], "(blue)", rat[j], "(red)")), path = "Combinations", device = "png")
115     ggsave(filename = paste("Inflamed", rabbit[i], rat[j], ".png", collapse = "", sep = " "),
116         plot = FeaturePlot(health[[2]], c(rabbit[i], rat[j]), blend = T, cols = c("lightyellow",
            "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
            ggtitle(paste("Inflamed", rabbit[i], "(blue)", rat[j], "(red)")), path = "Combinations",
            device = "png")
117     ggsave(filename = paste("Uninflamed", rabbit[i], rat[j], ".png", collapse = "", sep = "
        "),

```



```

118     plot = FeaturePlot(health[[3]], c(rabbit[i], rat[j]), blend = T, cols = c("lightyellow",
    "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
    ggtitle(paste("Uninflamed", rabbit[i], "(blue)", rat[j], "(red)")), path = "Combinations",
    device = "png")
119 }}
120
121 for (i in 1:length(goat)){
122   #goat&mouse
123   for (j in 1:length(mouse)){
124     ggsave(filename = paste("All", goat[i], mouse[j], ".png", collapse = "", sep = " "),
125       plot = FeaturePlot(fib.seur, c(goat[i], mouse[j]), blend = T, cols = c("lightyellow",
    "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] + ggtitle(paste("All",
    goat[i], "(blue)", mouse[j], "(red)")), path = "Combinations", device = "png")
126     ggsave(filename = paste("Inflamed", goat[i], mouse[j], ".png", collapse = "", sep = " "),
127       plot = FeaturePlot(health[[2]], c(goat[i], mouse[j]), blend = T, cols =
    c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
    ggtitle(paste("Inflamed", goat[i], "(blue)", mouse[j], "(red)")), path = "Combinations",
    device = "png")
128     ggsave(filename = paste("Uninflamed", goat[i], mouse[j], ".png", collapse = "", sep = "
    ")),
129     plot = FeaturePlot(health[[3]], c(goat[i], mouse[j]), blend = T, cols =
    c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
    ggtitle(paste("Uninflamed", goat[i], "(blue)", mouse[j], "(red)")), path = "Combinations",
    device = "png")
130   # Skipping Grem1 in healthy, expression is 0
131   if(goat[i] == "GREM1")
132     next
133   ggsave(filename = paste("Healthy", goat[i], mouse[j], ".png", collapse = "", sep = " "),
134     plot = FeaturePlot(health[[1]], c(goat[i], mouse[j]), blend = T, cols =
    c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
    ggtitle(paste("Healthy", goat[i], "(blue)", mouse[j], "(red)")), path = "Combinations",
    device = "png")
135 }
136 #goat&rat
137 for (j in 1:length(rat)){
138   ggsave(filename = paste("All", goat[i], rat[j], ".png", collapse = "", sep = " "),
139     plot = FeaturePlot(fib.seur, c(goat[i], rat[j]), blend = T, cols = c("lightyellow",
    "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] + ggtitle(paste("All",
    goat[i], "(blue)", rat[j], "(red)")),
140     path = "Combinations", device = "png")
141   ggsave(filename = paste("Inflamed", goat[i], rat[j], ".png", collapse = "", sep = " "),
142     plot = FeaturePlot(health[[2]], c(goat[i], rat[j]), blend = T, cols = c("lightyellow",
    "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
    ggtitle(paste("Inflamed", goat[i], "(blue)", rat[j], "(red)")), path = "Combinations", device
    = "png")
143   ggsave(filename = paste("Uninflamed", goat[i], rat[j], ".png", collapse = "", sep = " "),
144     plot = FeaturePlot(health[[3]], c(goat[i], rat[j]), blend = T, cols = c("lightyellow",
    "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
    ggtitle(paste("Uninflamed", goat[i], "(blue)", rat[j], "(red)")), path = "Combinations",
    device = "png")

```

```

145 # Skipping Grem1 in healthy, expression is 0
146 if(goat[i] == "GREM1")
147   next
148 ggsave(filename = paste("Healthy", goat[i], rat[j], ".png", collapse = "", sep = " "),
149       plot = FeaturePlot(health[[1]], c(goat[i], rat[j]), blend = T, cols = c("lightyellow",
150       "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] + ggtitle(paste("Healthy",
151       goat[i], "(blue)", rat[j], "(red)")), path = "Combinations", device = "png")
152 }
153 }
154 for (i in 1:length(mouse)){
155   #mouse&rat
156   for (j in 1:length(rat)){
157     ggsave(filename = paste("All", mouse[i], rat[j], ".png", collapse = "", sep = " "),
158         plot = FeaturePlot(fib.seur, c(mouse[i], rat[j]), blend = T, cols = c("lightyellow",
159         "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] + ggtitle(paste("All",
160         mouse[i], "(blue)", rat[j], "(red)")), path = "Combinations", device = "png")
161     ggsave(filename = paste("Healthy", mouse[i], rat[j], ".png", collapse = "", sep = " "),
162         plot = FeaturePlot(health[[1]], c(mouse[i], rat[j]), blend = T, cols =
163         c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
164         ggtitle(paste("Healthy", mouse[i], "(blue)", rat[j], "(red)")), path = "Combinations",
165         device = "png")
166     ggsave(filename = paste("Inflamed", mouse[i], rat[j], ".png", collapse = "", sep = " "),
167         plot = FeaturePlot(health[[2]], c(mouse[i], rat[j]), blend = T, cols =
168         c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
169         ggtitle(paste("Inflamed", mouse[i], "(blue)", rat[j], "(red)")), path = "Combinations",
170         device = "png")
171     ggsave(filename = paste("Uninflamed", mouse[i], rat[j], ".png", collapse = "", sep = "
172     "),
173         plot = FeaturePlot(health[[3]], c(mouse[i], rat[j]), blend = T, cols =
174         c("lightyellow", "darkblue", "darkred"), pt.size = 1, label = T, combine = F)[[3]] +
175         ggtitle(paste("Uninflamed", mouse[i], "(blue)", rat[j], "(red)")), path = "Combinations",
176         device = "png")
177   }
178 }

```

## A.2 Analyzing Kinchen et al. (2018), Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease

### A.2.1 Object construction

```

1 # Data retrieved from : https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE114374
2
3 # Create Seurat objects
4
5 library(Seurat)
6 library(scran)
7 library(ggplot2)
8
9 #####

```

```

10 # Healthy
11 # Get expression matrix, file downloaded from GEO, supplementary
12 HC <- as.sparse(read.table("GSE114374_Mouse_HC_expression_matrix.txt.gz"))
13
14 # Set up Seurat objects with expression matrix
15 HC.seur <- CreateSeuratObject(counts = HC)
16
17 # Finding Variable Features
18 {
19 ## Add batch information as metadata
20 HC.seur <- AddMetaData(HC.seur,
  as.factor(sapply(colnames(HC.seur@assays$RNA@data), function(x)
21   paste0("HC",substr(x,nchar(x),nchar(x))))), "batch")
22
23 ## Add cell cycle scores as metadata
24 {
25 ### Get pre-trained marker sets
26 mm.pairs <- readRDS(system.file("exdata", "mouse_cycle_markers.rds",
  package="scran"))
27 ### Get the expression data matrix
28 cc_mat <- as.matrix(HC.seur@assays$RNA@data)
29 ### Get ensembl gene id & mouse gene symbol list, replace relevant gene symbols with
  ids
30 genelist <- read.csv("mousegenetosymbol.csv", header = TRUE)
31 cc_ensembl <- sapply(rownames(cc_mat), function(x)
  genelist$ensembl[genelist$symbol==x][1])
32 ### Cell cycle anotation using the cyclone pairs method
33 cyc <- cyclone(cc_mat, mm.pairs, gene.names=cc_ensembl, verbose = TRUE)
34 ### Add back to Seurat object
35 HC.seur <- AddMetaData(HC.seur, metadata = cyc$phases, col.name = "phases")
36 for(scr in colnames(cyc$scores)) {
37   HC.seur <- AddMetaData(HC.seur, metadata = cyc$scores[[scr]], col.name = "scr")
38   HC.seur <- AddMetaData(HC.seur, metadata = cyc$normalized.scores[[scr]], col.name
    = paste0("norm_",scr) )
39 }}
40
41 ## Calculate variable genes. Old params: (MeanVarPlot)
42 ### fxn.x = expMean, fxn.y = logVarDivMean, (defaults)
43 ### x.low.cutoff = 0.0125, x.high.cutoff = 3, y.cutoff = 0.5 but high cutoff is 4 in the
  paper
44 HC.seur <- FindVariableFeatures(object = HC.seur, selection.method = "mvp",
  mean.cutoff = c(0.0125, 4), dispersion.cutoff = c(0.5, Inf), verbose = T)
45 # Regress out unwanted variation
46 HC.seur <- ScaleData(object = HC.seur, features = VariableFeatures(HC.seur), do.scale =
  TRUE, do.center = TRUE,
47   vars.to.regress = c("nUMI", "batch", "G1", "G2M"), model.use =
  "negbinom")
48
49 # Clustering
50 HC.seur <- RunPCA(object = HC.seur, features = VariableFeatures(HC.seur), npcs = 50)

```

```

51
52 # Quality check PCA
53 ElbowPlot(HC.seur, ndims = 30) # flattens at pc23
54
55 HC.seur <- JackStraw(HC.seur, dims = 30)
56 HC.seur <- ScoreJackStraw(HC.seur, dims = 1:30)
57 JackStrawPlot(HC.seur, dims = 1:30)
58
59 for (i in 1:30)
60   ggsave(filename = paste("DimHeatmap_PC", i, ".png", collapse = ""), device = "png",
61     plot = DimHeatmap(HC.seur, dims = i, nfeatures = 20))
62
63 HC.seur <- FindNeighbors(object = HC.seur, force.recalc = T)
64 HC.seur <- FindClusters(object = HC.seur, resolution = 1)
65 HC.seur <- RunTSNE(object = HC.seur, dims = 1:28, resolution = 1)
66 DimPlot(object = HC.seur, reduction = "tsne", pt.size = 0.5, label = T)
67
68 HC.seur <- BuildClusterTree(HC.seur)
69 PlotClusterTree(HC.seur)
70 DimPlot(object = HC.seur, reduction = "tsne", pt.size = 0.5, label = T)
71
72 # Start assigning ids to clusters
73 ## Markers from paper
74 marker_list <- c("Epcam", "Krt19", "Kit", "Ano1", "Myh11", "Des",
75   "Rgs5", "Pdgfrb", "Pecam1", "Lyve1",
76   "S100b", "Gfap", "Cd52", "Ptprc",
77   "Acta2", "Dpt", "Col6a2", "Col1a2")
78
79 ## Find markers for clusters
80 HC.markers <- FindAllMarkers(HC.seur)
81 write.csv(HC.markers, "HC.markers.csv")
82
83 ## Visualize marker dispersion
84 FeaturePlot(HC.seur, marker_list, pt.size = 0.5, label = T)
85 VlnPlot(HC.seur, features = marker_list, combine = F)
86 DoHeatmap(HC.seur, features = marker_list)
87 DotPlot(HC.seur, features = marker_list)
88
89 # Glial cell barcodes from Dr. Antanaviciute
90 glial <- c("AACCATGCACATGACT.1", "AGGCCGTGTCATACTG.1",
91   "ACGGGCTAGTTACCCA.2",
92   "AGCGTCGGTCGTGGCT.2", "AGGTCATCATGCTAGT.2",
93   "CCGTTCACTAGCGATG.2",
94   "CTGATAGGTTAGGGTG.2", "CTGTGCTTCCAAGTAC.2",
95   "TGCTGCTAGGGTTTCT.2",
96   "TGGGCGTCATGGTAGG.2", "TCAGCTCCAGTATGCT.3",
97   "TGTCCCATCAATCTCT.3")
98 glialloc <- vector(length = 12)
99 for (i in 1:12)
100   glialloc[i] <- which(HC.seur@assays$RNA@data@Dimnames[[2]] == glial[i])

```

```

97
98 DimPlot(HC.seur, cells = glialloc, pt.size = 0.5)
99 VlnPlot(HC.seur, c("S100b", "Sox6", "Gfap"))
100
101 # Subset for fibroblast clusters for closer inspection
102 HC.fib.seur <- subset(HC.seur, idents = c(0:5, 7:11))
103 # Rebuild the cluster tree
104 HC.fib.seur <- BuildClusterTree(HC.fib.seur, reorder = T, reorder.numeric = T)
105 PlotClusterTree(HC.fib.seur)
106 DimPlot(object = HC.fib.seur, reduction = "tsne", pt.size = 0.5, label = T)
107
108 # Find markers again for fibroblasts only
109 fib.markers <- FindAllMarkers(HC.fib.seur)
110 write.csv(fib.markers, "HC.fib.markers.csv")
111
112
113 # Rename fibroblast clusters according to paper
114 FeaturePlot(HC.seur, features = c("Cxcl12", "Bmp2", "Maff", "Adamdec1", "Wnt5a",
    "Cd55"))
115
116 HC.seur <- RenameIdents(HC.seur,
117     "Str1", "Str3", "Str3", "Str3",
118     "Str1", "Str1", "Smooth muscle & MF", "Str1",
119     "Str2", "Str2", "Str1", "Str3",
120     "Endothelial", "Lymphatic", "Pericyte & Glial", "Immune")
121 DimPlot(object = HC.seur, reduction = "tsne", pt.size = 0.5, label = T)
122
123 HC.fib.seur <- RenameIdents(HC.fib.seur,
124     '1' = "Str3.1", '2' = "Str3.2", '3' = "Str3.3", '4' = "Str3.4",
125     '5' = "Str1.1", '6' = "Str1.2", '7' = "Str1.3", '8' = "Str1.4",
126     '9' = "Str2.1", '10' = "Str1.5", '11' = "Str2.2")
127
128 DimPlot(object = HC.fib.seur, reduction = "tsne", pt.size = 0.5, label = T)
129
130 saveRDS(HC.seur, "HC.seur.rds")
131 saveRDS(HC.fib.seur, "HC.fib.seur.rds")

```

The DSS Seurat object was constructed with the same code.

### **A.2.2. Arl13b+ Marker Discovery**

```

1 library(Seurat)
2 library(ggplot2)
3
4 HC.seur <- readRDS("HC.seur.rds")
5 DSS.seur <- readRDS("DSS.seur.rds")
6
7 # Subset for Arl13b
8 HC.seur.Arl13b <- subset(HC.seur, subset = Arl13b > 0)
9 DSS.seur.Arl13b <- subset(DSS.seur, subset = Arl13b > 0)

```

```

10
11 ggsave(filename = "tSNE_HC_All_Arl13b.png",
12       plot = FeaturePlot(HC.seur, "Arl13b", pt.size = 0.2,
13       cols = c("lightyellow", "darkblue"), label = T)
14       + ggtitle(paste("HC", "Arl13b")) + theme(legend.position="none"),
15       width = 9, height = 7, units = "in", dpi = 300, device = "png")
16
17 ggsave(filename = "tSNE_HC_Arl13b_subset.png",
18       plot = DimPlot(HC.seur.Arl13b, reduction = "tsne", pt.size = 0.2, label = T)
19       + ggtitle("HC Arl13b+") + theme(legend.position="none"),
20       width = 9, height = 7, units = "in", dpi = 300, device = "png")
21
22 ggsave(filename = "tSNE_DSS_All_Arl13b.png",
23       plot = FeaturePlot(DSS.seur, "Arl13b", pt.size = 0.2,
24       cols = c("lightyellow", "darkblue"), label = T)
25       + ggtitle(paste("DSS", "Arl13b")) + theme(legend.position="none"),
26       width = 9, height = 7, units = "in", dpi = 300, device = "png")
27
28 ggsave(filename = "tSNE_DSS_Arl13b_subset.png",
29       plot = DimPlot(DSS.seur.Arl13b, reduction = "tsne", pt.size = 0.2, label = T)
30       + ggtitle("DSS Arl13b+") + theme(legend.position="none"),
31       width = 9, height = 7, units = "in", dpi = 300, device = "png")
32
33 # Get unique cluster markers between stromal populations
34 ## HC has 3, DSS has 4
35
36 dir.create("HC markers")
37 markers <- FindMarkers(HC.seur, ident.1 = "Str1", ident.2 = c("Str2", "Str3"))
38 write.csv(markers, file = "HC markers/HC_Str1.csv")
39 markers <- FindMarkers(HC.seur, ident.1 = "Str2", ident.2 = c("Str1", "Str3"))
40 write.csv(markers, file = "HC markers/HC_Str2.csv")
41 markers <- FindMarkers(HC.seur, ident.1 = "Str3", ident.2 = c("Str1", "Str2"))
42 write.csv(markers, file = "HC markers/HC_Str3.csv")
43
44 dir.create("DSS markers")
45 markers <- FindMarkers(DSS.seur, ident.1 = "Str1", ident.2 = c("Str2", "Str3", "Str4"))
46 write.csv(markers, file = "DSS markers/DSS_Str1.csv")
47 markers <- FindMarkers(DSS.seur, ident.1 = "Str2", ident.2 = c("Str1", "Str3", "Str4"))
48 write.csv(markers, file = "DSS markers/DSS_Str2.csv")
49 markers <- FindMarkers(DSS.seur, ident.1 = "Str3", ident.2 = c("Str1", "Str2", "Str4"))
50 write.csv(markers, file = "DSS markers/DSS_Str3.csv")
51 markers <- FindMarkers(DSS.seur, ident.1 = "Str4", ident.2 = c("Str1", "Str2", "Str3"))
52 write.csv(markers, file = "DSS markers/DSS_Str4.csv")
53
54 # Repeat for Arl13b+
55 markers <- FindMarkers(HC.seur.Arl13b, ident.1 = "Str1", ident.2 = c("Str2", "Str3"))
56 write.csv(markers, file = "HC markers/HC.Arl13b_Str1.csv")
57 markers <- FindMarkers(HC.seur.Arl13b, ident.1 = "Str2", ident.2 = c("Str1", "Str3"))
58 write.csv(markers, file = "HC markers/HC.Arl13b_Str2.csv")
59 markers <- FindMarkers(HC.seur.Arl13b, ident.1 = "Str3", ident.2 = c("Str1", "Str2"))

```

```

60 write.csv(markers, file = "HC markers/HC.Arl13b_Str3.csv")
61
62 markers <- FindMarkers(DSS.seur.Arl13b, ident.1 = "Str1", ident.2 = c("Str2", "Str3",
  "Str4"))
63 write.csv(markers, file = "DSS markers/DSS.Arl13b_Str1.csv")
64 markers <- FindMarkers(DSS.seur.Arl13b, ident.1 = "Str2", ident.2 = c("Str1", "Str3",
  "Str4"))
65 write.csv(markers, file = "DSS markers/DSS.Arl13b_Str2.csv")
66 markers <- FindMarkers(DSS.seur.Arl13b, ident.1 = "Str3", ident.2 = c("Str1", "Str2",
  "Str4"))
67 write.csv(markers, file = "DSS markers/DSS.Arl13b_Str3.csv")
68 markers <- FindMarkers(DSS.seur.Arl13b, ident.1 = "Str4", ident.2 = c("Str1", "Str2",
  "Str3"))
69 write.csv(markers, file = "DSS markers/DSS.Arl13b_Str4.csv")
70 #####
71 # merge the objects to compare with each other
72 HC.seur <- AddMetaData(HC.seur, metadata = "HC", col.name = "Health")
73 DSS.seur <- AddMetaData(DSS.seur, metadata = "DSS", col.name = "Health")
74
75 merge.seur <- merge(HC.seur, DSS.seur)
76
77 # Get Cluster names
78 merge.seur <- RenameIdents(merge.seur, 'Pericyte' = 'Pericyte & Glial') # couldn't
  identify glial cells in DSS
79 ClusterNames <- levels(merge.seur)
80
81 # Markers for each cluster grouped by health group
82 dir.create("healthMarkers")
83 for (i in 1:length(ClusterNames)){
84 markers <- FindMarkers(merge.seur, ident.1 = "DSS", ident.2 = "HC", group.by =
  "Health",
85           subset.ident = ClusterNames[i])
86 write.csv(markers, file = paste("healthMarkers/DSS_vs_HC_", ClusterNames[i], ".csv",
  collapse = ""))
87 }
88
89 # Same for Arl13b+
90 merge.seur.Arl13b <- subset(merge.seur, Arl13b > 0)
91 for (i in 1:length(ClusterNames)){
92 markers <- FindMarkers(merge.seur.Arl13b, ident.1 = "DSS", ident.2 = "HC", group.by
  = "Health",
93           subset.ident = ClusterNames[i])
94 write.csv(markers, file = paste("healthMarkers/Arl13b+_DSS_vs_HC_",
  ClusterNames[i], ".csv", collapse = ""))
95 }
96
97 # Find markers for Arl13b+ vs - cells
98 idx <- which(HC.seur@assays$RNA@counts@Dimnames[[1]]=="Arl13b")
99 Arl13b <- HC.seur@assays$RNA@counts[idx,] > 0
100 HC.seur <- AddMetaData(HC.seur, metadata = Arl13b, col.name = "filter")

```

```

101
102   idx <- which(DSS.seur@assays$RNA@counts@Dimnames[[1]]=="Arl13b")
103   Arl13b <- DSS.seur@assays$RNA@counts[idx,] > 0
104   DSS.seur <- AddMetaData(DSS.seur, metadata = Arl13b, col.name = "filter")
105
106   idx <- which(merge.seur@assays$RNA@counts@Dimnames[[1]]=="Arl13b")
107   Arl13b <- merge.seur@assays$RNA@counts[idx,] > 0
108   merge.seur <- AddMetaData(merge.seur, metadata = Arl13b, col.name = "filter")
109
110   # whole data
111   markers <- FindMarkers(HC.seur, ident.1 = T, ident.2 = F, group.by = "filter")
112   write.csv(markers, "HC_Arl13bmarkers_All.csv")
113
114   markers <- FindMarkers(DSS.seur, ident.1 = T, ident.2 = F, group.by = "filter")
115   write.csv(markers, "DSS_Arl13bmarkers_All.csv")
116
117   markers <- FindMarkers(merge.seur, ident.1 = T, ident.2 = F, group.by = "filter")
118   write.csv(markers, "merge_Arl13bmarkers_All.csv")
119
120   # per cluster
121   dir.create("Arl13b markers")
122   ClusterNames <- levels(HC.seur)
123   for (i in 1:length(ClusterNames)){
124     markers <- FindMarkers(HC.seur, ident.1 = T, ident.2 = F, group.by = "filter",
125       subset.ident = ClusterNames[i])
126     write.csv(markers, file = paste("Arl13b markers/HC_Arl13b+vs-", ClusterNames[i],
127       ".csv", collapse = ""))
128   }
129   ClusterNames <- levels(DSS.seur)
130   for (i in 1:length(ClusterNames)){
131     markers <- FindMarkers(DSS.seur, ident.1 = T, ident.2 = F, group.by = "filter",
132       subset.ident = ClusterNames[i])
133     write.csv(markers, file = paste("Arl13b markers/DSS_Arl13b+vs-",
134       ClusterNames[i], ".csv", collapse = ""))
135   }
136   ClusterNames <- levels(merge.seur)
137   for (i in 1:length(ClusterNames)){
138     markers <- FindMarkers(merge.seur, ident.1 = T, ident.2 = F, group.by = "filter",
139       subset.ident = ClusterNames[i])
140     write.csv(markers, file = paste("Arl13b markers/merge_Arl13b+vs-",
141       ClusterNames[i], ".csv", collapse = ""))
142   }

```

### **A.3 Analyzing Haber et al. (2017), A Single-cell Survey Of The Intestinal Epithelium**

#### **A.3.1 Object construction**

```
1 # Haber et al., 2017: A single-cell survey of the small intestinal epithelium
```



```

2 # Data from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE92332
3 # GSE92332_EEC_UMIcounts.txt.gz    1.1 Mb : Characterizing subsets of
   enteroendocrine cells
4 # GSE92332_FAE_UMIcounts.txt.gz    9.6 Mb : Characterizing microfold cells (follicle-
   associated epithelia)
5 # GSE92332_LargeCellSort_UMIcounts.txt.gz 24.1 Mb      : Extended Data Figure 3 |
   Regional variation in Paneth cell subtypes and stem cell markers
6 # GSE92332_Org_RANKL_UMIcounts.txt.gz 11.2 Mb      : Characterizing microfold
   cells (to compare FAE)
7 # GSE92332_Regional_UMIcounts.txt.gz 21.2 Mb      : Regional cell-type
   diversity
8 # GSE92332_atlas_UMIcounts.txt.gz: Single-cell survey of small intestinal epithelial
   cells
9
10 library(Seurat)
11 library(dplyr)
12 library(tidyr)
13
14 # atlas_UMIcounts (Fig. 1b)
15 #####
16 GSE92332_atlas_UMIcounts <- read.delim("GSE92332_atlas_UMIcounts.txt",
   row.names=1)
17 # name format: batch_barcode_cluster. Separating strings
18 names <- data.frame(x = colnames(GSE92332_atlas_UMIcounts))
19 names <- names %>% separate(x, c("batch", "barcode", "cluster"), sep = "_")
20
21 SI.epi.seur <- CreateSeuratObject(as.sparse(GSE92332_atlas_UMIcounts))
22 SI.epi.seur[["batch"]] <- names$batch
23 SI.epi.seur[["barcode"]] <- names$barcode
24 SI.epi.seur[["cluster"]] <- names$cluster
25
26 SI.epi.seur <- NormalizeData(SI.epi.seur)
27 SI.epi.seur <- FindVariableFeatures(SI.epi.seur)
28 SI.epi.seur <- ScaleData(SI.epi.seur, do.scale = T, do.center = T, vars.to.regress =
   c("nCount_RNA", "batch"))
29 SI.epi.seur <- RunPCA(SI.epi.seur)
30 ElbowPlot(SI.epi.seur, ndim = 50) # jumps, last at 22
31 SI.epi.seur <- FindNeighbors(SI.epi.seur, dims = 1:22)
32 SI.epi.seur <- FindClusters(SI.epi.seur)
33 SI.epi.seur <- RunTSNE(SI.epi.seur)
34 DimPlot(SI.epi.seur, group.by = "cluster", label = T, pt.size = 1)
35 DimPlot(SI.epi.seur, label = T, pt.size = 1)
36
37 # Refining the clusters
38 # Cluster 3 is a mix of other clusters but segregated here (Stem, TA G1, TA G2,
   Enterocyte.Progenitor.Early)
39 markers3 <- FindMarkers(SI.epi.seur, ident.1 = 3, ident.2 = c(5,0,8,2,1), only.pos = T)
40 # lipoproteins (high density), lipid catabolism
41
42 # Change active ident to paper clusters, but separate 4th cluster

```

```

43 # Get cell names
44 names3 <- CellsByIdentities(SI.epi.seur, idents = '3')
45 SI.epi.seur <- SetIdent(SI.epi.seur, value = "cluster")
46 SI.epi.seur <- SetIdent(SI.epi.seur, cells = unlist(names3), value = "lipid metabolizing")
47
48 #####
49 # LargeCellSort (Fig Ext 3a)
50 GSE92332_LargeCellSort_UMIcounts <-
  read.delim("GSE92332_LargeCellSort_UMIcounts.txt", row.names=1)
51 SI.largecell.epi.seur<-
  CreateSeuratObject(as.sparse(GSE92332_LargeCellSort_UMIcounts))
52
53 names <- data.frame(x = colnames(GSE92332_LargeCellSort_UMIcounts))
54 names <- names %>% separate(x, c("barcode", "mouse", "batch", "cluster"), sep = "_")
55
56 SI.largecell.epi.seur[["batch"]] <- names$batch
57 SI.largecell.epi.seur[["mouse"]] <- names$mouse
58 SI.largecell.epi.seur[["barcode"]] <- names$barcode
59 SI.largecell.epi.seur[["cluster"]] <- names$cluster
60
61 SI.largecell.epi.seur <- NormalizeData(SI.largecell.epi.seur)
62 SI.largecell.epi.seur <- FindVariableFeatures(SI.largecell.epi.seur)
63 SI.largecell.epi.seur <- ScaleData(SI.largecell.epi.seur, do.scale = T, do.center = T,
  vars.to.regress = c("nCount_RNA", "batch"))
64 SI.largecell.epi.seur <- RunPCA(SI.largecell.epi.seur)
65 ElbowPlot(SI.largecell.epi.seur, ndim = 50)
66 SI.largecell.epi.seur <- FindNeighbors(SI.largecell.epi.seur, dims = 1:17)
67 SI.largecell.epi.seur <- FindClusters(SI.largecell.epi.seur)
68 SI.largecell.epi.seur <- RunTSNE(SI.largecell.epi.seur)
69 DimPlot(SI.largecell.epi.seur, group.by = "cluster", label = T, pt.size = 1)
70 SI.largecell.epi.seur <- SetIdent(SI.largecell.epi.seur, value = "cluster")

```

## **A.4 Analyzing McCarthy et al. (2020), Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient**

### **A.4.1 Object construction**

```

1 # McCarthy et al., Distinct Mesenchymal Cell Populations Generate the Essential
  Intestinal BMP Signaling Gradient,
2 # Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.01.008
3 # Data from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130681
4 # GSM3747599_PdgfraEGFPsc_1_mm10: PDGFRA+ cells
5
6 library(Seurat)
7 library(ggplot2)
8 library(Matrix)
9
10 # read and assemble counts matrix
11 counts <- readMM("GSM3747599_PdgfraEGFPsc_1_mm10_matrix.mtx.gz")

```

```

12 barcodes <- read.table("GSM3747599_PdgfraEGFPsc_1_mm10_barcodes.tsv.gz",
    quote="", comment.char="")
13 genes <- read.delim("GSM3747599_PdgfraEGFPsc_1_mm10_genes.tsv.gz",
    header=FALSE)
14
15 counts@Dimnames[1] <- genes[2]
16 counts@Dimnames[2] <- barcodes
17
18 # Filter and construct Seurat object
19 # "PDGFRA+ cells with R 1,500 detected transcripts and genes expressed in R 100 single
    cells were retained,
20 # resulting in 9,334 detectable genes in 2,894 single cells"
21 Pdgfra.seur <- CreateSeuratObject(counts, min.cells = 100, min.features = 1500)
22
23 # Further exclusion of cells with > 3% mitochondrial transcripts gave a final count of
    2,595 informative single cells.
24 mito.genes <- grep(pattern = "^mt-", x = rownames(Pdgfra.seur@assays[["RNA"]]),
    value = TRUE)
25 percent.mito <- Matrix::colSums(Pdgfra.seur@assays[["RNA"]][mito.genes,
    ])/Matrix::colSums(Pdgfra.seur@assays[["RNA"]])
26 Pdgfra.seur <- AddMetaData(object = Pdgfra.seur, metadata = percent.mito, col.name =
    "percent.mito")
27 VlnPlot(Pdgfra.seur, features = c("nFeature_RNA", "nCount_RNA", "percent.mito"),
    ncol = 3)
28
29 Pdgfra.seur <- subset(Pdgfra.seur, subset = percent.mito < 0.03)
30
31 # Normalize & log transform
32 Pdgfra.seur <- NormalizeData(Pdgfra.seur)
33
34 # top 714 variable genes using the "FindVariableGenes" function in Seurat using
    parameters:
35 # x.low.cutoff 0.0125, x.high.cutoff 3, and y.cutoff 1 for principal component analysis
36 Pdgfra.seur <- FindVariableFeatures(Pdgfra.seur, mean.cutoff = c(0.0125, 3),
    dispersion.cutoff = c(1, Inf), nfeatures = 714)
37
38 # The top 7 principal components were selected for PDGFRA+ cells
39 Pdgfra.seur <- ScaleData(Pdgfra.seur)
40 Pdgfra.seur <- RunPCA(Pdgfra.seur)
41
42 Pdgfra.seur <- JackStraw(Pdgfra.seur)
43 Pdgfra.seur <- ScoreJackStraw(Pdgfra.seur)
44 JackStrawPlot(Pdgfra.seur)
45
46 Pdgfra.seur <- FindNeighbors(Pdgfra.seur, dims = 1:7)
47 # resolution 0.2 for PDGFRA+ cells
48 Pdgfra.seur <- FindClusters(Pdgfra.seur, resolution = 0.2)
49
50 Pdgfra.seur <- RunTSNE(Pdgfra.seur)
51 DimPlot(Pdgfra.seur)

```

```

52
53 # Cluster identification
54 FeaturePlot(Pdgfra.seur, "Pdgfra") # Cluster 1 -> Telocyte
55 FeaturePlot(Pdgfra.seur, "Cd34") # Cluster 0 -> Lo2
56 FeaturePlot(Pdgfra.seur, "Grem1") # Cluster 2 -> Lo-1
57 # Cluster 3 -> Pdgfra- Cd34-, also visible in Fig. 3a but not appearing distinct
58 new.cluster.ids <- c("Lo2", "Telocyte", "Lo1", "Pdgfra- Cd34-")
59 names(new.cluster.ids) <- levels(Pdgfra.seur)
60 Pdgfra.seur <- RenameIdents(Pdgfra.seur, new.cluster.ids)
61
62 DimPlot(Pdgfra.seur)
63 saveRDS(Pdgfra.seur, "Pdgfra.seur.rds")
64
65
66 # Arl13b Acetylated tubulin Acot7 positive cells
67 Pdgfra.seur.Arl13b <- subset(Pdgfra.seur, subset = Arl13b > 0)
68 Pdgfra.seur.Arl13b.Tubal1a <- subset(Pdgfra.seur.Arl13b, subset = Tubal1a > 0)
69 Pdgfra.seur.Arl13b.Tubal1a.Acot7 <- subset(Pdgfra.seur.Arl13b.Tubal1a, subset = Acot7 >
  0)
70 # 38 cells. Barcodes:
71 quadpos <- c("AAACCTGAGGCGATAC-1", "AAAGATGTCTCGATGA-1",
  "AACGTTGCACACCGAC-1", "ACCCACTGTGCAGGTA-1",
72 "AGAGCGACACCTGGTG-1", "AGAGTGGTCGTCCAGG-1",
  "AGCAGCCTCATAGCAC-1", "AGCATACGTCGATTGT-1",
73 "AGCTCTCAGCAGACTG-1", "ATCCGAAGTACCATCA-1",
  "CAACCTCTCTGCGACG-1", "CAAGAAACAGGCAGTA-1",
74 "CCGTACTAGGTAGCTG-1", "CCTCAGTTCTGCAGTA-1",
  "CGCTTCATCAGTGCAT-1", "CGGAGCTGTGCTAGCC-1",
75 "CGTCACTTCTCCCTGA-1", "CTTAGGAAGGTGACCA-1",
  "GAACGGATCTCTGAGA-1", "GAGGTGATCGCGTAGC-1",
76 "GCGCAGTTCCGCTGTT-1", "GCTCCTAAGAGACTTA-1",
  "GCTCTGTTCTAACCGA-1", "GCTGCAGGTGGGTCAA-1",
77 "GCTTCCAGTGCCTTGG-1", "GGAACCTCAATCAGAA-1",
  "GGCGACTGTCATGCCG-1", "GTGGGTCAGTCTCAAC-1",
78 "GTTCTCGTCCACGACG-1", "TATTACCATGGAATA-1",
  "TCAACGAGTGTTTCGAT-1", "TCATTACCAGGGATTG-1",
79 "TCGAGGCAGGGAACGG-1", "TCGAGGCTCTGCCCTA-1",
  "TGCACCTCAGACGTAG-1", "TTGGCAAGTATCAGTC-1",
80 "TTTCCTCGTTGGTAAA-1", "TTTGCGCCATCTGGTA-1")
81
82 counts <- Pdgfra.seur.Arl13b.Tubal1a.Acot7@assays$RNA@counts
83 write.csv(cbind(rowMeans(counts), rowSums(counts), counts), "Pdgfra Arl13b acTub
  Acot7 cells expression matrix.csv")
84
85 # assign new identity to metadata
86 location <- vector(length = 38)
87 for (i in 1:38)
88   location[i] <- which(Pdgfra.seur@assays$RNA@data@Dimnames[[2]] == quadpos[i])
89
90 Pdgfra.seur <- StashIdent(Pdgfra.seur, "clusters")

```

```

91 expressing <- rep("negative", Pdgfra.seur@assays$RNA@counts@Dim[2])
92 for(i in 1:38)
93   expressing[location[i]] <- "positive"
94
95 Pdgfra.seur[["expressing"]] <- expressing
96 DimPlot(Pdgfra.seur, group.by = "expressing", pt.size = 1)
97
98 # Find markers for this group
99 markers <- FindMarkers(Pdgfra.seur, ident.1 = "positive", ident.2 = "negative", group.by
   = "expressing")
100   write.csv(markers, "Acot7 Arl13b AcTub Pdgfra markers.csv")

```

## **A.5 Analyzing Roulis et al. (2020), Paracrine Orchestration Of Intestinal Tumorigenesis By A Mesenchymal Niche**

### **A.5.1 Object construction**

```

1 # Adapted from Roulis_2020_organoids.R
2 # Article: https://www.nature.com/articles/s41586-020-2166-3
3 # Related figures: 3d and extended 4d, e
4 # Data: GSM4227191 and GSM4227192
5 # Original code: https://github.com/KlugerLab/Scripts\_Roulis\_et\_al\_2020
6 # has a copy of combineData.R
7
8 # sc-RNAseq data from mouse intestinal organoid culture and crypt fibroblast
   organotypic coculture
9 # Crypts: harvested from last 3/4 of SI, villi scraped off, placed in organoid domes
10 # Fibroblasts: harvested from mice SI, epithelial shaken off, digested and strained
11 # fibroblasts were seeded onto culture plates, grown overnight, cryps in 1:1 Matrigel
   were overlaid with OGM
12 # Combines PtgerON (C) and PtgerOFF (T) data
13
14 library(Seurat)
15 library(Matrix)
16 library(ggplot2)
17
18 ## Combine data function
19 combineData <- function(data1, data2, name1 = NULL, name2 = NULL, delim = "_"){
20   # read in data
21   exp_pre <- data1
22   exp_post <- data2
23   if(!is.null(name1)){
24     colnames(exp_pre) <- paste(colnames(exp_pre), name1, sep = delim)
25   }
26   if(!is.null(name2)){
27     colnames(exp_post) <- paste(colnames(exp_post), name2, sep = delim)
28   }
29
30   # find all genes

```

```

31 allgene <- unique(c(rownames(exp_pre), rownames(exp_post)))
32 missG_pre <- setdiff(allgene, rownames(exp_pre))
33 missG_post <- setdiff(allgene, rownames(exp_post))
34
35 # create matrix with 0 for missing genes
36 miss_pre <- matrix(0, ncol = dim(exp_pre)[2], nrow = length(missG_pre))
37 rownames(miss_pre) <- missG_pre
38 colnames(miss_pre) <- colnames(exp_pre)
39 miss_post <- matrix(0, ncol = dim(exp_post)[2], nrow = length(missG_post))
40 rownames(miss_post) <- missG_post
41 colnames(miss_post) <- colnames(exp_post)
42
43 # bind data
44 new_pre <- rbind(exp_pre, miss_pre)
45 new_post <- rbind(exp_post, miss_post)
46 new_pre <- new_pre[order(rownames(new_pre)),]
47 new_post <- new_post[order(rownames(new_post)),]
48 print(mean(rownames(new_pre) == rownames(new_post)))
49 data_new <- cbind(new_pre, new_post)
50 return(data_new)
51 }
52
53
54 data_S <- CreateSeuratObject(counts = combineData(data1 =
  read.table("/media/deniz/Dosyalar/Bilkent/Master/Thesis/Articles/sc-
  RNAseq/Mesenchyme+Organoid/Scripts_Roulis_et_al_2020-master/data/DS-014B-C-
  56099208_gene_exon_tagged_dge.txt.gz",
55                               header = T, sep = "\t", stringsAsFactors = F,
  row.names = 1),
56                               data2 =
  read.table("/media/deniz/Dosyalar/Bilkent/Master/Thesis/Articles/sc-
  RNAseq/Mesenchyme+Organoid/Scripts_Roulis_et_al_2020-master/data/DS_014B_T-
  56094214_gene_exon_tagged_dge.txt.gz",
57                               header = T, sep = "\t", stringsAsFactors = F,
  row.names = 1),
58                               name1 = "C", name2 = "T"),
59                               project = "organoid", min.cells = 0, min.features = 0, names.field = 2,
  names.delim = "_")
60 table(data_S@meta.data$orig.ident)
61
62
63 # filter cells: 1000+ transcripts <10% mitochondrial genes
64 sum(data_S@meta.data$nCount_RNA >= 1000)
65 data_S <- subset(x = data_S, subset = nCount_RNA > 999)
66
67 mito.genes <- grep(pattern = "^mt-", x = rownames(x = GetAssayData(data_S)), value =
  TRUE, ignore.case = T)
68 data_S <- AddMetaData(data_S, col.name = "mito.ratio",
  colSums(data_S@assays$RNA@counts[mito.genes, colnames(data_S)])/
69          colSums(data_S@assays$RNA@counts[, colnames(data_S)]))

```

```

70 max(data_S@meta.data$mito.ratio)
71 VlnPlot(data_S, "mito.ratio", group.by = "orig.ident")
72 data_S <- subset(data_S, subset = mito.ratio < 0.1)
73 VlnPlot(data_S, "mito.ratio", group.by = "orig.ident")
74
75 table(data_S@meta.data$orig.ident)
76 length(colnames(data_S))
77
78 # Normalization
79 data_S <- NormalizeData(data_S)
80
81 # Imputation
82 # data_S <- RunALRA(data_S)
83
84 # analysis
85 data_S <- ScaleData(data_S, do.scale = F, do.center = T)
86 data_S <- RunPCA(data_S, npcs = 29, features = rownames(data_S), nfeatures.print = 5)
87
88 data_S <- RunTSNE(data_S, dims.use = 1:29, seed.use = 1)
89 DimPlot(data_S, group.by = "orig.ident", pt.size = 0.5)
90
91 data_S <- FindNeighbors(data_S)
92 data_S <- FindClusters(data_S, resolution = 0.8)
93 DimPlot(data_S, label = T, pt.size = 0.5)
94
95 # Find Epithelial
96 FeaturePlot(data_S, "Epcam")
97 # Clusters 1, 2, 3, 4, 5, 6, 8, 9
98
99 # Add culture info as metadata
100 culture <- c("fibroblast", "epithelial",
  "epithelial", "epithelial", "epithelial", "epithelial", "fibroblast",
  "epithelial", "epithelial", "fibroblast", "fibroblast")
101 names(culture) <- levels(data_S)
102 data_S <- RenameIdents(data_S, culture)
103 DimPlot(data_S, label = T, pt.size = 0.5)
104 data_S <- StashIdent(data_S, save.name = "culture")
105 Idents(data_S) <- data_S@meta.data$seurat_clusters
106
107 epithelial <- subset(data_S, subset = culture == "epithelial")
108 sum(table(epithelial@active.ident))
109 # 1584 cells (1 less than paper)
110 fibroblast <- subset(data_S, subset = culture == "fibroblast")
111
112 ## Identifying Epithelial clusters
113 {
114   # Clusters 1, 2, 5 are too mitosis related. Regress out cell cycle and try clustering
  again.
115   # Applying the method in https://satijalab.org/seurat/v3.0/cell\_cycle\_vignette.html
116   # Get cell cycle phase markers

```

```

117 s.genes <- cc.genes$s.genes
118 g2m.genes <- cc.genes$g2m.genes
119
120 # gene names are for human, convert to mouse
121 # https://www.r-bloggers.com/converting-mouse-to-human-gene-names-with-
    biomaRt-package/
122 convertHumanGeneList <- function(x){
123
124     require("biomaRt")
125     human = useMart("ensembl", dataset = "hsapiens_gene_ensembl")
126     mouse = useMart("ensembl", dataset = "mmusculus_gene_ensembl")
127
128     genesV2 = getLDS(attributes = c("hgnc_symbol"), filters = "hgnc_symbol", values
    = x , mart = human, attributesL = c("mgi_symbol"), martL = mouse, uniqueRows=T)
129
130     humanx <- unique(genesV2[, 2])
131
132     # Print the first 6 genes found to the screen
133     return(humanx)
134 }
135
136 g2m.genes <- convertHumanGeneList(g2m.genes)
137 s.genes <- convertHumanGeneList(s.genes)
138 epithelial <- CellCycleScoring(epithelial, s.features = s.genes, g2m.features =
    g2m.genes)
139
140 epithelial <- ScaleData(epithelial, do.scale = F, do.center = T,
141     vars.to.regress = c("S.Score", "G2M.Score"), features =
    rownames(epithelial))
142 epithelial <- FindVariableFeatures(epithelial)
143 epithelial <- RunPCA(epithelial, npcs = 29, features = VariableFeatures(epithelial),
    nfeatures.print = 5, force.recalc = T)
144
145 epithelial <- RunTSNE(epithelial, dims.use = 1:29, seed.use = 1, perplexity = 60)
146 epithelial <- FindNeighbors(epithelial)
147 epithelial <- FindClusters(epithelial, resolution = 2)
148 DimPlot(epithelial, label = T, pt.size = 0.5)
149
150 markers <- FindAllMarkers(epithelial, only.pos = T)
151 write.csv(markers, file = "newallmarkers.csv")
152
153
154 # Plot markers from Haber et al, 2017
155 VlnPlot(epithelial, c("Chgb", "Cpe", "Gfra3")) # Enteroendocrine
156 VlnPlot(epithelial, c("Chga", "Tac1", "Tph1", "Neurog3")) # Enteroendocrine
157 VlnPlot(epithelial, c("Bmi1", "Prox1")) # Enteroendocrine
158 VlnPlot(epithelial, c("Lyz1", "Mptx2", "Defa-rs1")) # Paneth
159 VlnPlot(epithelial, c("Defa17", "Defa22", "Defa24", "Ang4")) # Paneth
160 VlnPlot(epithelial, c("Tff3", "Manf", "Ccl9")) # Goblet
161 VlnPlot(epithelial, c("Muc2", "Clca3", "Agr2")) # Goblet

```



```

162 VlnPlot(epithelial, c("Lrmp", "Dclk1", "Cd24a")) # Tuft
163 VlnPlot(epithelial, c("Trpm5", "Bmi1")) # Tuft
164 VlnPlot(epithelial, c("Mep1a", "Fgf15", "Clec2h")) # Distal enterocyte
165 VlnPlot(epithelial, c("Lct", "Cbr1", "Ephx2")) # Proximal enterocyte
166 VlnPlot(epithelial, c("Alpi", "Apoa1", "Apoa4", "Fabp1")) # Enterocyte
167 VlnPlot(epithelial, c("Lgr5", "Ascl2", "Slc12a2", "Axin2", "Olfm4", "Gkn3")) #
Stem
168 VlnPlot(epithelial, c("Mki67", "Cdk4", "Mcm5", "Mcm6", "Pcna")) # Cell cycle
169 VlnPlot(epithelial, c("Ly6a", "Clu", "Msln", "Il1rn")) # Reserve stem cells
170 VlnPlot(epithelial, c("Vil1", "Atoh1", "Dclk1", "Ptgs1", "Neurog3")) # Tuft
171 VlnPlot(epithelial, c("Lgr5", "Slc12a2"))
172
173 # View top expressed genes in each cluster, but most were just ribosomal ot
mitochondrial
174 cluster12 <- subset(epithelial, idents = 12)
175 topgenes12 <- rowSums(cluster12@assays$RNA@data)
176 tail(sort(topgenes12), 20)
177
178 # name clusters
179 new.cluster.ids <- c("Stem", "Reserve Stem", "Progenitors", "Stem",
"Paneth/Goblet", "Respiration", "Immune/adhesion", "Hypoxic/Gluconeogenesis",
"Stem", "Respiration/Glycolysis", "Tuft", "Enterocyte", "Metabolizing/arachidonic acid")
180 names(new.cluster.ids) <- levels(epithelial)
181 epithelial <- RenameIdents(epithelial, new.cluster.ids)
182 DimPlot(epithelial, label = T)
183 DimPlot(epithelial, group.by = "orig.ident")
184
185 # To revert:
186 # Idents(epithelial) <- epithelial@meta.data$seurat_clusters
187
188 # Save epithelial Seurat object
189 saveRDS(epithelial, file = "organoid.RDS")
190 epithelial <- readRDS("organoid.RDS")
191
192 # Divide epithelial into Ptger-ON and OFF
193 epi.PtgerON <- subset(epithelial, subset = orig.ident == "C")
194 epi.PtgerOFF <- subset(epithelial, subset = orig.ident == "T")
195
196 # Create Arl13b+ subsets for the groups
197 epi.Arl13b <- subset(epithelial, subset = Arl13b > 0)
198 epi.PtgerON.Arl13b <- subset(epi.PtgerON, subset = Arl13b > 0)
199 epi.PtgerOFF.Arl13b <- subset(epi.PtgerOFF, subset = Arl13b > 0)
200
201 # Arl13b present almost equally between treatment groups, continue with epi.Arl13b
only
202 rm(epi.PtgerON.Arl13b)
203 rm(epi.PtgerOFF.Arl13b)
204 gc()
205
206 # Find markers for Arl13b+ cells

```

```

207     arl13bmarkers <- FindAllMarkers(epi.Arl13b)
208
209     # Some cluster markers are new. Conclusion: Pack markers and expression matrices
    for filtered and unfiltered and present as result.
210     markers <- FindAllMarkers(epithelial)
211     write.csv(markers, file = "namedallmarkers.csv")
212     write.csv(arl13bmarkers, file = "epithelial arl13bmarkers.csv")
213     write.csv(epithelial@assays$RNA@counts, file = "epithelial expression matrix.csv")
214     write.csv(epi.Arl13b@assays$RNA@counts, file = "epithelial Arl13b+ expression
    matrix.csv")
215
216     ggsave(filename = "epithelial Arl13b+.png",
217           plot = FeaturePlot(epithelial, "Arl13b", pt.size = 0.5, cols = c("lightyellow",
    "darkblue"), label = T) + ggtitle("Epithelial Arl13b+"),)
218     ggsave(filename = "epithelial Ift88+.png",
219           plot = FeaturePlot(epithelial, "Ift88", pt.size = 0.5, cols = c("lightyellow",
    "darkblue"), label = T) + ggtitle("Epithelial Ift88+"),)
220   }
221
222   ## Identifying Fibroblast clusters
223   {
224     # Regress out cell cycle
225     fibroblast <- CellCycleScoring(fibroblast, s.features = s.genes, g2m.features =
    g2m.genes)
226     fibroblast <- ScaleData(fibroblast, do.scale = F, do.center = T, vars.to.regress =
    c("S.Score", "G2M.Score"), features = rownames(fibroblast))
227
228     # Recluster
229     fibroblast <- FindVariableFeatures(fibroblast)
230     fibroblast <- RunPCA(fibroblast, npcs = 50, features = VariableFeatures(fibroblast),
    nfeatures.print = 5, force.recalc = T)
231     ElbowPlot(fibroblast, ndims = 50)
232     fibroblast <- RunTSNE(fibroblast, dims.use = 1:25, seed.use = 1)
233     fibroblast <- FindNeighbors(fibroblast)
234     fibroblast <- FindClusters(fibroblast)
235     DimPlot(fibroblast, label = T)
236     DimPlot(fibroblast, group.by = "orig.ident")
237     markers <- FindAllMarkers(fibroblast, only.pos = T)
238     write.csv(markers, "fibroblast allmarkers.csv")
239
240     # Check markers (used in Fig 1)
241     VlnPlot(fibroblast, c("Pdgfra", "Ptgs2", "Cd34", "Vcam1", "Fgfr2", "Lama1",
    "Acta2", "Rspo1"))
242     # Col6a5, Ano2 not in, Has1, Ces1g has 0 expression
243     VlnPlot(fibroblast, c("Pdgfrb", "Myh11", "Dpt"))
244     VlnPlot(fibroblast, c("Wnt5a", "Wnt5b", "Wnt2b", "Fos"))
245     VlnPlot(fibroblast, c("Pdgfra", "Pdpn"))
246     VlnPlot(fibroblast, c("Serpine2", "Grem1", "Rspo3"))
247
248     collagens <- rownames(fibroblast)[3719:3755]

```

```

249   DotPlot(fibroblast, features = collagens)
250
251   # Clusters 4 and 6 don't have unique markers and are similar.
252   fibroblast <- BuildClusterTree(fibroblast)
253   PlotClusterTree(fibroblast)
254
255   new.cluster.ids <- c("Metal Ion", "Myofibroblast", "Fos/Serpine 2/Fgf10",
    "Grem1/Rspo3", "NA", "Collagens", "NA")
256   names(new.cluster.ids) <- levels(fibroblast)
257   fibroblast <- RenameIdents(fibroblast, new.cluster.ids)
258   DimPlot(fibroblast, label = T)
259   DimPlot(fibroblast, group.by = "orig.ident")
260
261   markers <- FindMarkers(fibroblast, ident.1 = "NA", only.pos = T)
262   write.csv(markers, "NA cluster.csv")
263
264   VlnPlot(fibroblast, "Mki67")
265
266   new.cluster.ids <- c("Metal Ion", "Myofibroblast", "Fos/Serpine 2/Fgf10",
    "Grem1/Rspo3", "Proliferating", "Collagens")
267   names(new.cluster.ids) <- levels(fibroblast)
268   fibroblast <- RenameIdents(fibroblast, new.cluster.ids)
269   DimPlot(fibroblast, label = T)
270   DimPlot(fibroblast, group.by = "orig.ident")
271
272   VlnPlot(fibroblast, c("Grem1", "Arl13b", "Ift88"))
273   FeaturePlot(fibroblast, "Arl13b")
274   FeaturePlot(fibroblast, "Ift88")
275
276   # To revert:
277   # Idents(fibroblast) <- fibroblast@meta.data$seurat_clusters
278
279   # Save fibroblast Seurat object
280   saveRDS(fibroblast, file = "fibroblast.RDS")
281   fibroblast <- readRDS("fibroblast.RDS")
282
283   # Divide fibroblast into Ptger-ON and OFF
284   fib.PtgerON <- subset(fibroblast, subset = orig.ident == "C")
285   fib.PtgerOFF <- subset(fibroblast, subset = orig.ident == "T")
286
287   # Create Arl13b+ subsets for the groups
288   fib.Arl13b <- subset(fibroblast, subset = Arl13b > 0)
289   fib.PtgerON.Arl13b <- subset(fib.PtgerON, subset = Arl13b > 0)
290   fib.PtgerOFF.Arl13b <- subset(fib.PtgerOFF, subset = Arl13b > 0)
291
292   # Arl13b present almost equally between treatment groups, continue with fib.Arl13b
    only
293   rm(fib.PtgerON.Arl13b)
294   rm(fib.PtgerOFF.Arl13b)
295   gc()

```

```

296
297   # Find markers for Arl13b+ cells
298   arl13bmarkers <- FindAllMarkers(fib.Arl13b)
299
300   # Some cluster markers are new. Conclusion: Pack markers and expression matrices
    for filtered and unfiltered and present as result.
301   markers <- FindAllMarkers(fibroblast)
302   write.csv(markers, file = "fibroblast namedallmarkers.csv")
303   write.csv(arl13bmarkers, file = "fibroblast arl13bmarkers.csv")
304   write.csv(fibroblast@assays$RNA@counts, file = "fibroblast expression
    matrix.csv")
305   write.csv(fib.Arl13b@assays$RNA@counts, file = "fibroblast Arl13b+ expression
    matrix.csv")
306
307   ggsave(filename = "fibroblast Arl13b+.png",
308         plot = FeaturePlot(fibroblast, "Arl13b", pt.size = 0.5, cols = c("lightyellow",
    "darkblue"), label = T) + ggtitle("fibroblast Arl13b+"),)
309   ggsave(filename = "fibroblast Ift88+.png",
310         plot = FeaturePlot(fibroblast, "Ift88", pt.size = 0.5, cols = c("lightyellow",
    "darkblue"), label = T) + ggtitle("fibroblast Ift88+"),)
311   }

```

## **A.6 Analyzing Drokhlyansky et al. (2020), The Human and Mouse Enteric Nervous System at Single-Cell Resolution**

### **A.6.1 Object construction**

Seurat v3 objects were generated using the code and data published for the Smillie et al. (2019) article, as was done for the original study.

## **A.7 Analyzing Halpern et al. (2018), Paired-cell Sequencing Enables Spatial Gene Expression Mapping Of Liver Endothelial Cells**

### **A.7.1 Object construction**

The Seurat v2 object was generously provided by Dr. Keren Bahar Halpern and only updated to Seurat v3.

## **A.8 Analyzing Ramachandran et al. (2019), Resolving The Fibrotic Niche Of Human Liver Cirrhosis At Single-cell Level**

### **A.8.1 Object Construction**

```

1 library(Seurat)
2 library(dplyr)
3 library(ggplot2)
4

```

```

5 # Data from https://datashare.is.ed.ac.uk/handle/10283/3433
6 # Readme:
7 # tissue.rdata: RData file containing a Seurat object "tissue", which in turn contains the
  raw
8 # and normalised expression values for 58,358 human single cell transcriptomes from
  healthy and
9 # cirrhotic liver. This represents the data after application of all relevant, stringent quality
10 # control and doublet removal steps, and thus while it does not wholly correlate to any
  figure
11 # shown in the related publication, may be thought of as the 'final' processed and
  annotated
12 # version of Figure 1.
13
14 # The Seurat object "tissue" also contains metadata fields as follows: <-
  ScaleData(humanLiver.seur, vars.to.regress = "nUMI",
15
16 # - nGene, nUMI, percent.mito: quality control metrics
17 # - liver: liver of origin, i.e. Healthy1, Cirrhotic4 (see Extended Data Figure 2f)
18 # - dataset: dataset of origin, i.e. Healthy1_Cd45+, Cirrhotic4_Cd45+
19 # - condition: condition of liver of origin, i.e. Uninjured, Cirrhotic
20 # - aetiology: aetiology of liver of origin, i.e. NAFLD, PBC
21 # - annotation_original: results of initial high-level clustering, i.e. 1-21
22 # - annotation_lineage: results of lineage annotation, i.e. Endothelia, MPs
23 # - annotation_indepth: results of detailed low-level clustering, i.e. MPs (1), Endothelia
  (4)
24
25 # Update original seurat v2 object to v3
26 {
27   load("tissue.rdata")
28   humanLiver.seur <- UpdateSeuratObject(tissue)
29   # This didn't work. (Error: None of the features provided are in this Assay object)
30
31   # Alternative: make a new Seurat object using the count matrix
32   # in the old one and add the metadata
33   humanLiver.seur <- CreateSeuratObject(tissue@data, meta.data = tissue@meta.data,
34     project = tissue@project.name)
35   humanLiver.seur@assays$RNA@counts <- tissue@raw.data
36   saveRDS(humanLiver.seur, "humanLiver.seur.RDS")
37   rm(tissue)
38   gc()
39 }
40 # If new Seurat object already available
41 # humanLiver.seur <- readRDS("humanLiver.seur.RDS")
42
43 # Find variable features & Scale Data
44 {
45   # Post normalization preprocessing:
46   # Variation in UMI counts between cells was regressed according to a negative binomial
47   # model, before scaling and centring the resulting value by subtracting the mean
  expression

```

```

48 # of each gene and dividing by its standard deviation (En), then calculating
49 #  $\ln(10^4 \times En + 1)$ .
50
51 humanLiver.seur <- FindVariableFeatures(humanLiver.seur, selection.method = "mvp")
52
53 # ScaleData needs to allocate vector of size 10.2 GB.
54 humanLiver.seur.scaled <- ScaleData(humanLiver.seur, vars.to.regress = "nUMI",
55                                     model.use = "negbinom", do.scale = T, do.center = T, verbose =
    T)
56 saveRDS(humanLiver.seur.scaled, "humanLiver.seur.scaled.RDS")
57 humanLiver.seur <- humanLiver.seur.scaled
58 rm(humanLiver.seur.scaled)
59 gc()
60 }
61
62 # PCA clustering
63 {
64 # We performed unsupervised clustering and differential gene expression analyses in
65 # the Seurat R package v.2.3.039. In particular, we used shared nearest neighbour
66 # graph-based clustering, in which the graph was constructed using from 2 to 11
67 # principal components as determined by dataset variability shown in principal
    component
68 # analysis (PCA); the resolution parameter to determine the resulting number of clusters
69 # was also tuned accordingly. To assess cluster similarity we used the
70 # 'BuildClusterTree' function from Seurat.
71
72 humanLiver.seur <- RunPCA(humanLiver.seur, features =
    VariableFeatures(humanLiver.seur), npcs = 20)
73
74 ElbowPlot(humanLiver.seur, ndims = 20) # drop at 4th, 6th, 10th, 14th, 10th looks like
    the elbow
75 humanLiver.seur <- JackStraw(humanLiver.seur, dims = 20)
76 humanLiver.seur <- ScoreJackStraw(humanLiver.seur, dims = 1:20, do.plot = T) # all
    very low p
77
78 humanLiver.seur <- FindNeighbors(humanLiver.seur)
79 # Clusters in paper = 21
80 # res = 0.8 -> 20 | res = 1 -> 27 | res = 0.9 -> 23
81 humanLiver.seur <- FindClusters(humanLiver.seur, resolution = 0.9)
82 humanLiver.seur <- BuildClusterTree(humanLiver.seur, reorder = T, reorder.numeric =
    T)
83 PlotClusterTree(humanLiver.seur)
84 }
85
86 # Cluster identification
87 humanLiver.seur <- RunTSNE(object = humanLiver.seur, dims = 1:11, perplexity = 70)
88 ggsave(filename = "tsNE All cluster IDs.png",
89         plot = DimPlot(object = humanLiver.seur, reduction = "tsne", pt.size = 0.2, label =
    T),
90         width = 9, height = 7, units = "in", dpi = 300, device = "png")

```

```

91
92 ggsave(filename = "tSNE Patient IDs.png",
93       plot = DimPlot(object = humanLiver.seur, group.by = "orig.ident",
94       reduction = "tsne", pt.size = 0.2, label = F),
95       width = 9, height = 7, units = "in", dpi = 300, device = "png")
96
97 ggsave(filename = "tSNE All condition.png",
98       plot = DimPlot(object = humanLiver.seur, reduction = "tsne",
99       group.by = "condition", pt.size = 0.2, label = T),
100      width = 9, height = 7, units = "in", dpi = 300, device = "png")
101
102 ggsave(filename = "tSNE aetiology.png",
103       plot = DimPlot(object = humanLiver.seur, group.by = "aetiology",
104       reduction = "tsne", pt.size = 0.2, label = F),
105       width = 9, height = 7, units = "in", dpi = 300, device = "png")
106
107 ggsave(filename = "tSNE original clusters.png",
108       plot = DimPlot(object = humanLiver.seur, group.by = "annotation_lineage",
109       reduction = "tsne", pt.size = 0.2, label = T),
110       width = 9, height = 7, units = "in", dpi = 300, device = "png")
111
112 # Find all markers
113 humanLiver.markers <- FindAllMarkers(humanLiver.seur, test.use = "roc",
114                                     logfc.threshold = 0.25, min.pct = 0.25)
115 write.csv(humanLiver.markers , "humanLiver_allMarkers.csv")
116
117 # Show top 10 markers of each cluster
118 top10 <- humanLiver.markers %>% group_by(cluster) %>% top_n(n=10, wt =
119   avg_logFC)
120
121 # markers from paper
122 marker_list <- c(# MP
123   "FCN1", "C1QC", "VCAN", "IL1B", #1
124   "MND A", "MS4A7", "CD163", "CPVL", "MSR1", "CYBB", #2
125   "CSF2RA", "CLEC10A", #3
126   # pDC
127   "LILRA4", "PTCRA", "LRRC26", "CLEC4C", #4
128   # ILC
129   "KLRF1", "KLRC1", #5
130   "PRF1", "FGFBP2", #6/7
131   # Cycling
132   "MKI67", "UBE2C", "TOP2A", "RRM2", #8
133   # T cell
134   "TRAC", "CD2", #9&10&11&12
135   # B cell
136   "CD79A", "MS4A1", #13
137   # Plasma cell
138   "FCRL5", "JSRP1", #14
139   # Mast cell

```

```

140     "TPSAB1","CPA3",#15
141     # Endothelia
142     "CLEC14A",
      "EGFL7","HSPG2","FCN3","CLEC4G","PLVAP","VWF","EMCN","OIT3","CLEC4M"
      ,#16&17
143     # Mesenchyme
144     "TAGLN","ACTA2","TIMP2","COL3A1","COL1A2","RGS5","PDGFRB","LUM","PD
      GFRA","CCL19",#18&19
145     # Epithelia
146     "ALB","KNG1","TTR","ORM1","TF","APOC2","F2","PRAP1","HSD17B6","ITIH1","
      CYP4A11","AGXT",#20
147     "KRT19","CXCL6","CXCL1","SOX9" #21
148   )
149
150   # Show markers in clusters
151   dir.create("paper marker violins")
152   for (i in 1:length(marker_list)){
153     ggsave(VlnPlot(humanLiver.seur, features = marker_list[i]), device = "png",
154             path = "paper marker violins/", filename = paste(marker_list[i], ".png", collapse
      = ""))
155   }
156
157   # Clusters were identified based on the marker expressions used in the paper.
158   # MP: 3, 5, 6 Epithelia: 2, 7 Mesenchyme: 8, 9 B cell: 10 Endothelia: 15, 16
159   # pDC: 11, Plasma: 1 Mast: 19
160   # Some markers were expressed in too many clusters. Used Cluster tree for grouping
161   # ILC/Cycling: 12, 13, 14 T cells: 17, 18, 20
162   # Cluster 4 didn't have any markers dominantly expressed (70 cells only)
163   # Marker genes are for metal ion processes, heat shock and immunity
164   # Added to T cells, because the other markers showed almost no expression
165
166   # Stash original cluster numbers as metadata
167   humanLiver.seur <- AddMetaData(humanLiver.seur, metadata =
      humanLiver.seur@meta.data$seurat_clusters, col.name = "initial_clusters")
168
169   # Assign new cluster names
170   new.cluster.ids <- c("Plasma cell", "Epithelia", "MP","Epithelia","Unknown",
171                       "Mesenchyme", "Mesenchyme", "Endothelia", "ILC", "Cycling",
172                       "ILC", "B cell", "pDC", "Endothelia", "Endothelia", "T cell",
173                       "T cell", "T cell", "Mast cell", "MP", "MP", "MP", "MP")
174   names(new.cluster.ids) <- levels(humanLiver.seur)
175   humanLiver.seur <- RenameIdents(humanLiver.seur, new.cluster.ids)
176
177   ggsave(filename = "tSNE All named clusters.png",
178           plot = DimPlot(object = humanLiver.seur, reduction = "tsne", pt.size = 0.2, label
      = T),
179           width = 9, height = 7, units = "in", dpi = 300, device = "png")
180

```



```

181 # Compare with given cluster names, revert to initial IDs, apply new
182 new.cluster.ids <- c("Plasma cell", "Hepatocytes", "MP",
  "Cholangiocytes", "Cholangiocytes", "Mesenchyme", "Mesothelia", "Endothelia", "ILC",
  "Cycling", "ILC", "B cell", "pDC", "Endothelia", "Endothelia", "T cell", "T cell", "T
  cell", "Mast cell", "MP", "MP", "MP", "MP")
183 Idents(humanLiver.seur) <- humanLiver.seur@meta.data$seurat_clusters
184 humanLiver.seur <- BuildClusterTree(humanLiver.seur, reorder = T, reorder.numeric
  = T)
185
186 names(new.cluster.ids) <- levels(humanLiver.seur)
187 humanLiver.seur <- RenameIdents(humanLiver.seur, new.cluster.ids)
188 ggsave(filename = "tSNE All final clusters.png",
189       plot = DimPlot(object = humanLiver.seur, reduction = "tsne", pt.size = 0.2, label
  = T),
190       width = 9, height = 7, units = "in", dpi = 300, device = "png")
191
192 # Save Seurat object for marker discovery
193 saveRDS(humanLiver.seur, "humanLiver.seur.RDS")

```

## **A.9 Analyzing Dobie et al. (2019), Single-Cell Transcriptomics Uncovers Zonation of Function in the Mesenchyme During Liver Fibrosis**

### **A.9.1 Object construction**

```

1 library(Seurat)
2 library(ggplot2)
3
4 #loading the liver seurat object
5 livermarkers.seur <- readRDS("markeranalysis_livermerged.seur.rds")
6 liver.all <- table(livermarkers.seur@active.ident)
7
8 #Renaming the clusters
9 head(x = livermarkers.seur@meta.data)
10 new.cluster.ids <- c("HSC1", "HSC2", "FB1", "FB2", "HSC3", "HSC4", "VSMC1",
  "HSC5", "HSC6", "FB3", "FB4", "VSMC2", "VSMC3", "HSC7", "FB5", "HSC8",
  "VSMC4", "VSMC5", "HSC9", "VSMC6")
11 names(new.cluster.ids) <- levels(livermarkers.seur)
12 livermarkers.seur <- RenameIdents(livermarkers.seur, new.cluster.ids)
13 livermarkers.seur[["seurat_clusters"]] <- Idents(object = livermarkers.seur)
14 DimPlot(livermarkers.seur, label = TRUE)
15
16 #subsetting the healthy cells from the liver.seur object and detecting Arl13b+
17 liver.all <- table(livermarkers.seur@active.ident)
18 liver.uninjured <- subset(livermarkers.seur, subset = condition == "Uninjured")
19 liver.uninjured.t <- table(liver.uninjured@active.ident)
20 liver.uninjured.Arl <- subset(liver.uninjured, Arl13b > 0)
21 liver.uninjured.subs <- table(liver.uninjured.Arl@active.ident)
22 DimPlot(liver.uninjured.Arl)
23

```

```

24 #subsetting the injured cells from the liver.seur object and detecting Arl13b+
25 liver.all <- table(livermarkers.seur@active.ident)
26 liver.injured <- subset(livermarkers.seur, subset = condition == "Peak CCl4")
27 liver.injured.t <- table(liver.injured@active.ident)
28 liver.injured.Arl <- subset(liver.injured, Arl13b > 0)
29 liver.injured.subs <- table(liver.injured.Arl@active.ident)
30 DimPlot(liver.injured.Arl)
31
32 #subset of all Arl13b+ cells
33 liver.all.Arl <- subset(livermarkers.seur, Arl13b > 0)
34 liver.all.Arl.t <- table(liver.all.Arl@active.ident)
35 DimPlot(liver.all.Arl)
36
37 #the matrix of all, healthy, and injured cells
38 matrix_1 <- rbind(liver.all, liver.uninjured.t)
39 matrix_2 <- rbind(liver.injured.t)
40 merged.matrix <- merge(matrix_1, matrix_2, all=TRUE, sort=F)
41 names(merged.matrix)
42 rownames(merged.matrix) <- c("liver.all", "liver.uninjured.t", "liver.injured.t")
43
44 #the matrix of all Arl13b+, uninjured Arl13b+, and injured Arl13b+ cells
45 matrix_3 <- rbind(liver.all.Arl.t, liver.uninjured.subs )
46 matrix_4 <- rbind(liver.injured.subs)
47 merged.matrix_arl <- merge(matrix_3, matrix_4, all=TRUE, sort=F)
48 names(merged.matrix_arl)
49 rownames(merged.matrix_arl) <- c("liver.all.Arl13b+", "liver.uninjured.Arl13b+",
  "liver.injured.Arl13b+")
50
51 #Creating a File for Markers
52 dir.create("HealthMarkers_PeakCCl4vsUninjured_All_PeakCCl4vsUninjured_")
53
54 #Markers for Peak CCl4 vs Uninjured (If it gives an error message, skip that number by
  changing 1 to desired number.)
55 for (i in 1:length(new.cluster.ids)){
56   markers <- FindMarkers(livermarkers.seur, ident.1 = "Peak CCl4", ident.2 =
  "Uninjured", group.by = "condition", subset.ident = new.cluster.ids[i], min.pct = 0.01)
57   write.csv(markers, file =
  paste("HealthMarkers_PeakCCl4vsUninjured_All_PeakCCl4vsUninjured_",
  new.cluster.ids[i], ".csv", collapse = ""))
58 }
59
60 #Markers for Peak CCl4 vs Uninjured Arl13b+ (If it gives an error message, skip that
  number by changing 1 to desired number.)
61 for (i in 1:length(new.cluster.ids)){
62   markers <- FindMarkers(liver.all.Arl, ident.1 = "Peak CCl4", ident.2 = "Uninjured",
  group.by = "condition", subset.ident = new.cluster.ids[i], min.pct = 0.01)
63   write.csv(markers, file = paste("PeakCCl4vsUninjured_Arl13b_Markers_",
  new.cluster.ids[i], ".csv", collapse = ""))
64 }
65

```

```

66 #subsetting the PaHSC cells from the liver.seur object and detecting Arl13b+
67 liver.all <- table(livermarkers.seur@active.ident)
68 liver.PaHSC <- subset(livermarkers.seur, subset = hsc_zonation == "Portal")
69 liver.PaHSC.t <- table(liver.PaHSC@active.ident)
70 liver.PaHSC.Arl <- subset(liver.PaHSC, Arl13b > 0)
71 liver.PaHSC.subs <- table(liver.PaHSC.Arl@active.ident)
72 DimPlot(liver.PaHSC.Arl)
73
74 #subsetting the CaHSC cells from the liver.seur object and detecting Arl13b+
75 liver.all <- table(livermarkers.seur@active.ident)
76 liver.CaHSC <- subset(livermarkers.seur, subset = hsc_zonation == "Central")
77 liver.CaHSC.t <- table(liver.CaHSC@active.ident)
78 liver.CaHSC.Arl <- subset(liver.CaHSC, Arl13b > 0)
79 liver.CaHSC.subs <- table(liver.CaHSC.Arl@active.ident)
80 DimPlot(liver.CaHSC.Arl)
81
82 #the matrix of all, PaHSC, and CaHSC cells
83 matrix_5 <- rbind(liver.all, liver.PaHSC.t)
84 matrix_6 <- rbind(liver.CaHSC.t)
85 merged.matrix_PaCaHSC <- merge(matrix_5, matrix_6, all=TRUE, sort=F)
86 names(merged.matrix_PaCaHSC)
87 rownames(merged.matrix_PaCaHSC) <- c("liver.all.t", "liver.PaHSC.t", "liver.CaHSC.t")
88
89 #the matrix of all Arl13b+, CaHSC Arl13b+, and PaHSC Arl13b+ cells
90 matrix_7 <- rbind(liver.all.Arl.t, liver.PaHSC.subs )
91 matrix_8 <- rbind(liver.CaHSC.subs)
92 merged.matrix_PaCaHSC_arl <- merge(matrix_7, matrix_8, all=TRUE, sort=F)
93 names(merged.matrix_PaCaHSC_arl)
94 rownames(merged.matrix_PaCaHSC_arl) <- c("liver.all.Arl13b+",
    "liver.PaHSC.Arl13b+", "liver.CaHSC.Arl13b+")
95
96
97 #Markers for PaHSC vs CaHSC (If it gives an error message, skip that number by
    changing 1 to desired number.)
98 for (i in 1:length(new.cluster.ids)){
99   markers <- FindMarkers(livermarkers.seur, ident.1 = 'Portal', ident.2 = 'Central',
    group.by = 'hsc_zonation', subset.ident = new.cluster.ids[i], min.pct = 0.01)
100   write.csv(markers, file = paste("PaHSCvsCaHSC_Markers_", new.cluster.ids[i],
    ".csv", collapse = ""))
101 }
102
103 #Markers for PaHSC vs CaHSC Arl13b+ (If it gives an error message, skip that
    number by changing 1 to desired number.)
104 for (i in 1:length(new.cluster.ids)){
105   markers <- FindMarkers(liver.all.Arl, ident.1 = "Portal", ident.2 = "Central",
    group.by = "hsc_zonation", subset.ident = new.cluster.ids[i], min.pct = 0.01)
106   write.csv(markers, file = paste("PaHSCvsCaHSC_Arl13b_Markers_",
    new.cluster.ids[i], ".csv", collapse = ""))
107 }
108

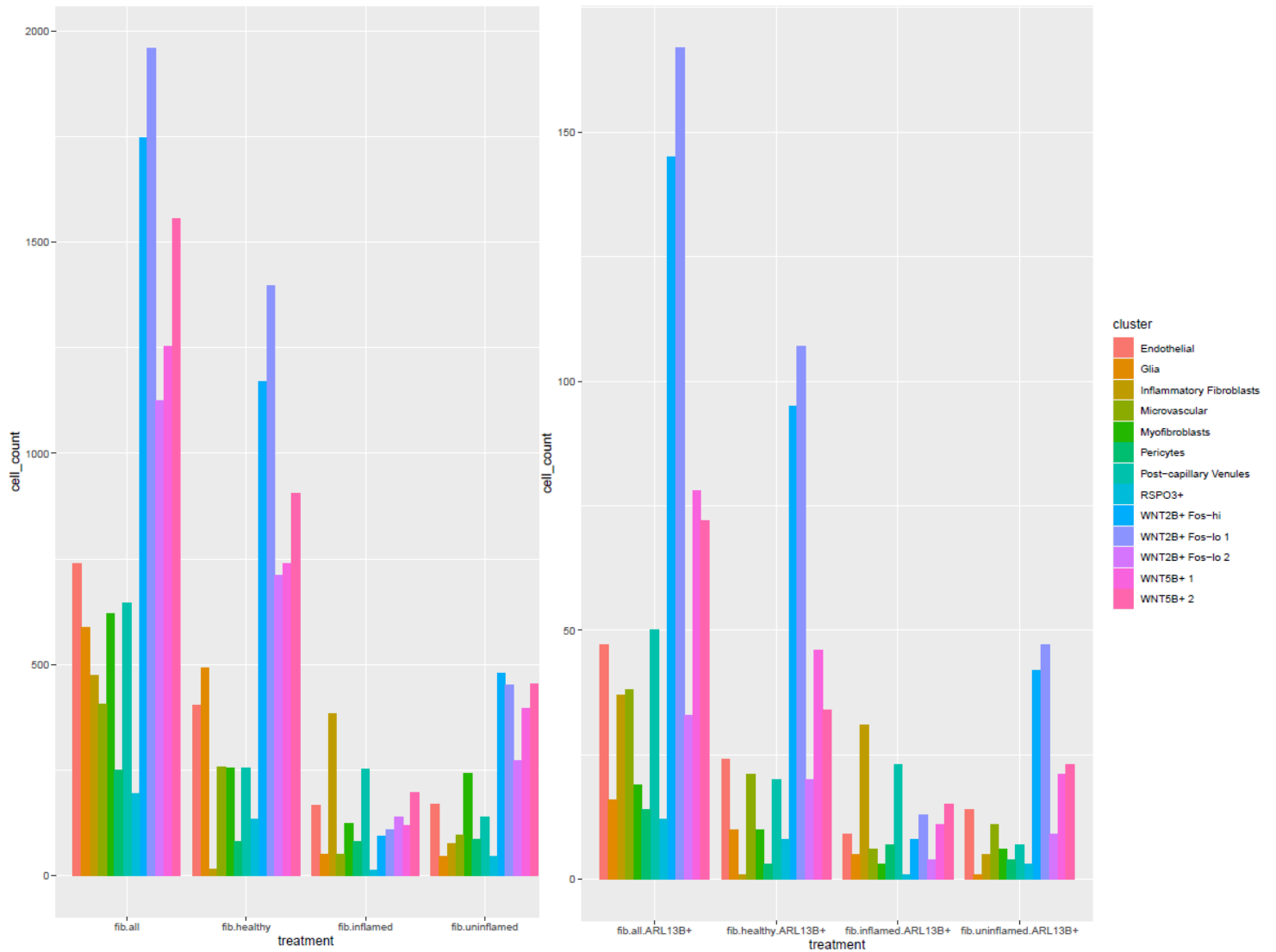
```

```

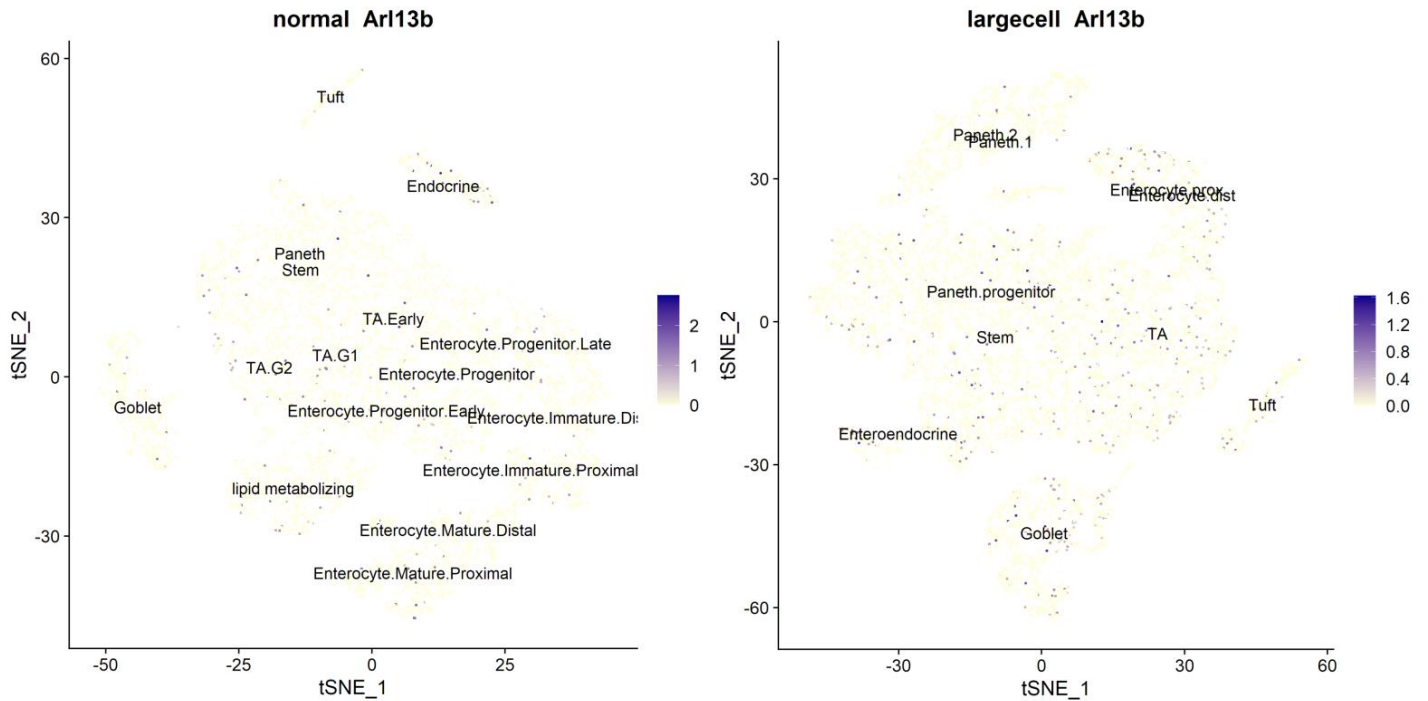
109 #Creating a subset for all Arl13b+ cells and trial
110 arl13b.seur <- subset(livermarkers.seur, subset = Arl13b>0)
111 DimPlot(arl13b.seur)
112 names(new.cluster.ids) <- levels(arl13b.seur)
113 arl13b.seur <- RenameIdents(arl13b.seur, new.cluster.ids)
114 Idents(arl13b.seur) <- arl13b.seur[["hsc_zonation"]]
115 markers_trial <- FindMarkers(arl13b.seur, ident.1 = c("HSC2", "HSC6"), ident.2 =
    c("HSC1", "HSC8", "HSC7", "HSC3", "HSC4", "HSC5"), group.by = "seurat_clusters",
    subset.ident = "Portal", min.pct = 0.01)
116
117
118 #Creating a subset for hscs, finding markers for Central and Portal, and saving them
    as csv
119 liver.hsc.seur <- subset(livermarkers.seur, subset = lineage == "HSC")
120 Idents(liver.hsc.seur) <- liver.hsc.seur[["hsc_zonation"]]
121 hsc_markers_central <- FindMarkers(liver.hsc.seur, ident.1 = "Peak CCl4", ident.2 =
    "Uninjured", group.by = "condition", subset.ident = "Central")
122 hsc_markers_portal <- FindMarkers(liver.hsc.seur, ident.1 = "Peak CCl4", ident.2 =
    "Uninjured", group.by = "condition", subset.ident = "Portal")
123 write.csv(hsc_markers_central, file = "HSC_Markers_Central.csv")
124 write.csv(hsc_markers_portal, file = "HSC_Markers_Portal.csv")
125
126 #Creating a subset for HSCs that are Arl13b positive, finding markers for Central and
    Portal, and saving them as csv
127 liver.hsc.arl13b.seur <- subset(livermarkers.seur, subset = Arl13b>0)
128 Idents(liver.hsc.arl13b.seur) <- liver.hsc.arl13b.seur[["hsc_zonation"]]
129 hsc_arl13b_markers_central <- FindMarkers(liver.hsc.arl13b.seur, ident.1 = "Peak
    CCl4", ident.2 = "Uninjured", group.by = "condition", subset.ident = "Central")
130 hsc_arl13b_markers_portal <- FindMarkers(liver.hsc.arl13b.seur, ident.1 = "Peak
    CCl4", ident.2 = "Uninjured", group.by = "condition", subset.ident = "Portal")
131 write.csv(hsc_arl13b_markers_central, file = "HSC_Arl13b_Markers_Central.csv")
132 write.csv(hsc_arl13b_markers_portal, file = "HSC_Arl13b_Markers_Portal.csv")
133
134 #Saving the RDS
135 saveRDS(livermarkers.seur, "Arl13bAnalysis.seur.rds")

```

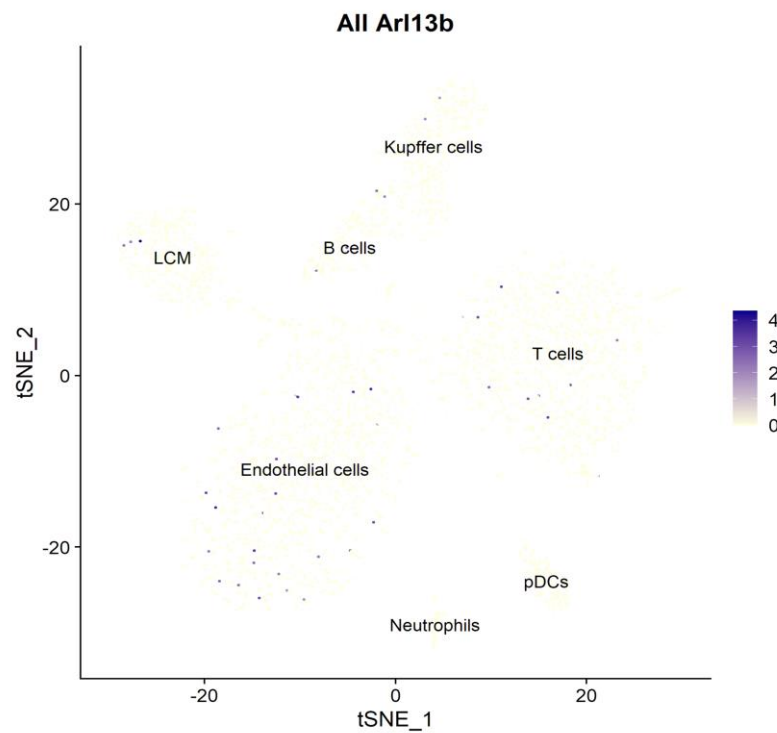
## B. Supplementary Figures



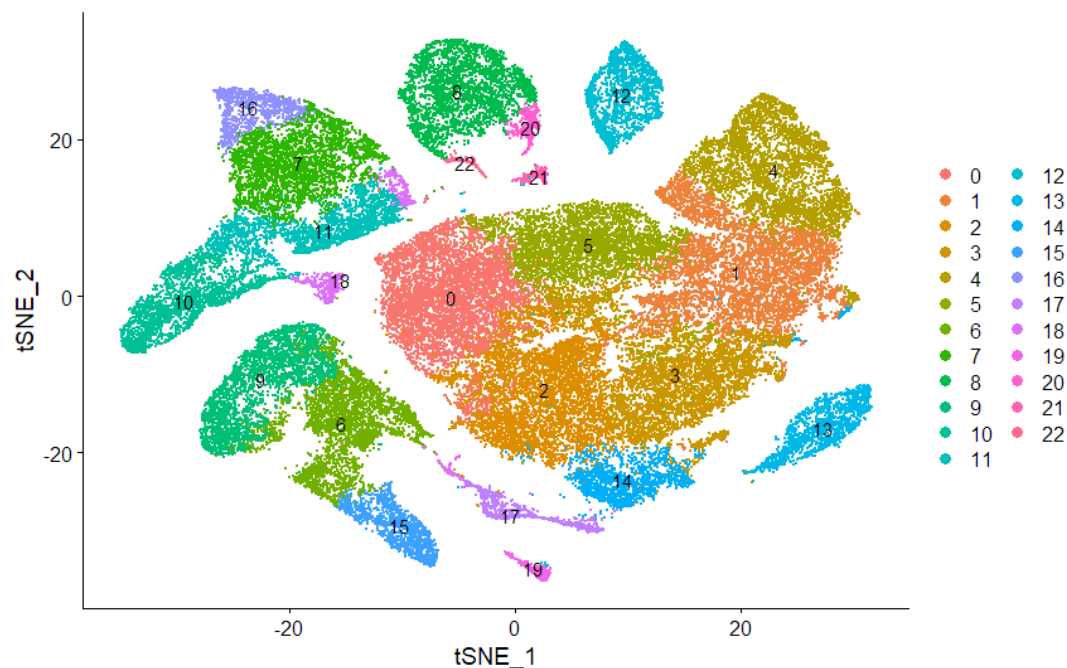
**Figure S1. Distribution of cell counts across clusters in human colon mesenchyme scRNA-seq data.** Data from Smillie et al. (2019). Left: Cell counts from whole data. Right: Cell counts from ARL13B<sup>+</sup> subset. fib: Fibroblast all: All health states. Uninflamed is defined as a healthy tissue biopsy obtained proximal to active ulcerative colitis.



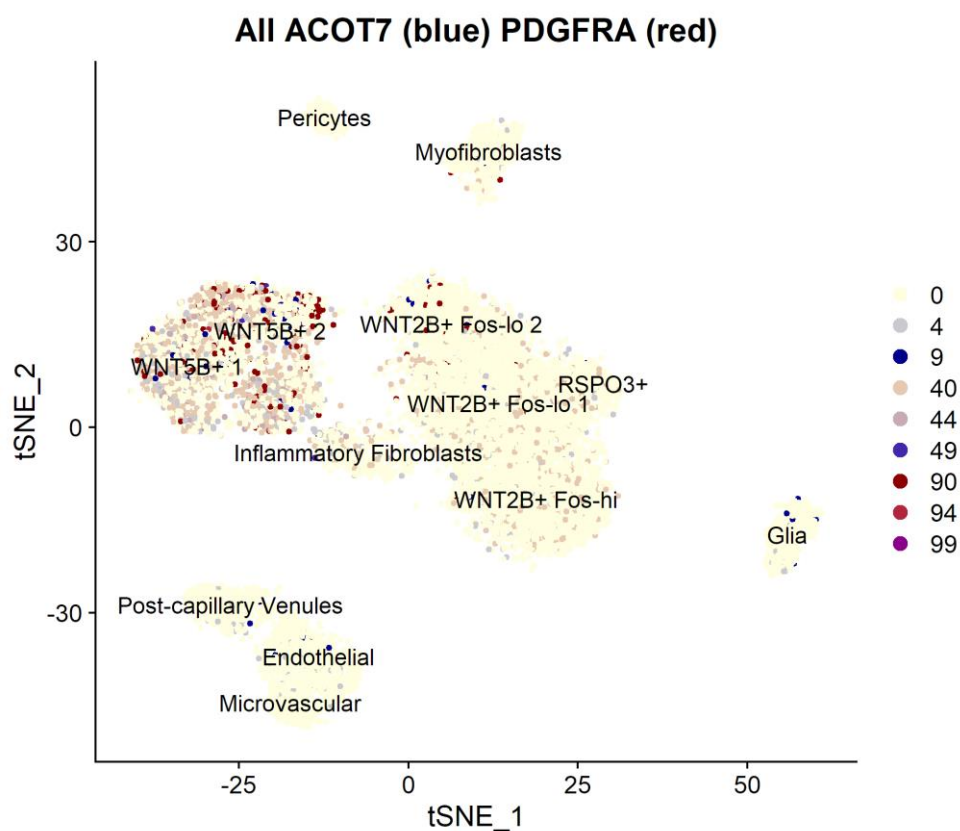
**Figure S2. Arl13b mRNA expression in small intestine epithelium.** Data from Haber et al. (2017). Feature plots of Arl13b transcript distribution in mouse small intestine epithelium scRNA-seq data. Plot shows cells from all health condition groups shown in Fig. 4.1. Normal: single cell suspension prepared with standard protocol. Largecell: Protocol altered to select for large cell types.



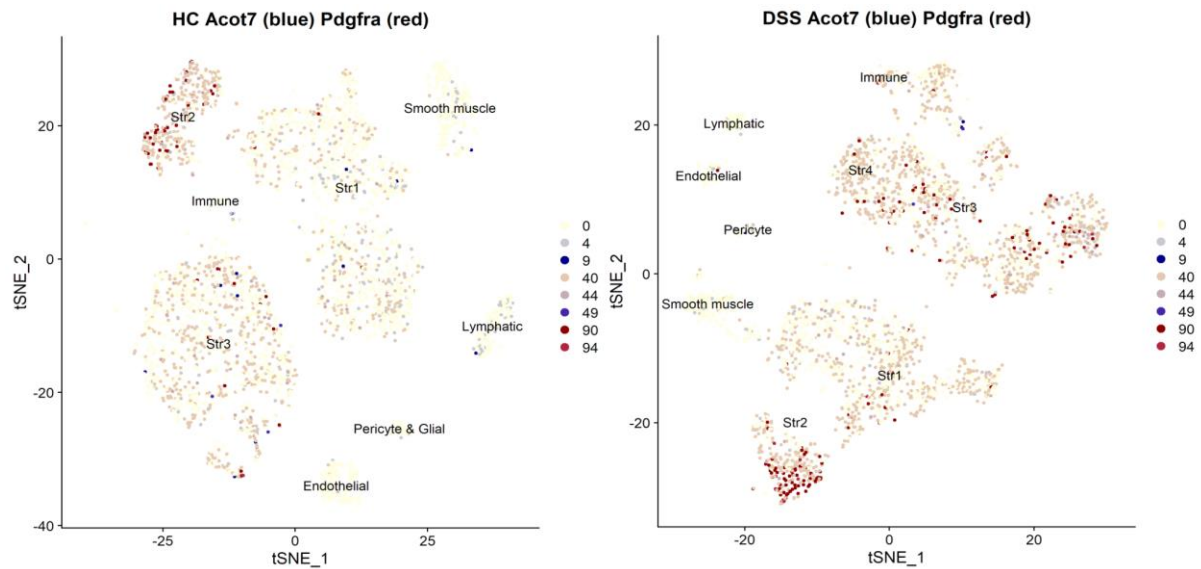
**Figure S3. Arl13b mRNA expression profile in mouse liver endothelium and immune cells.** Data from Halpern et al. (2018). Feature plot of Arl13b expression



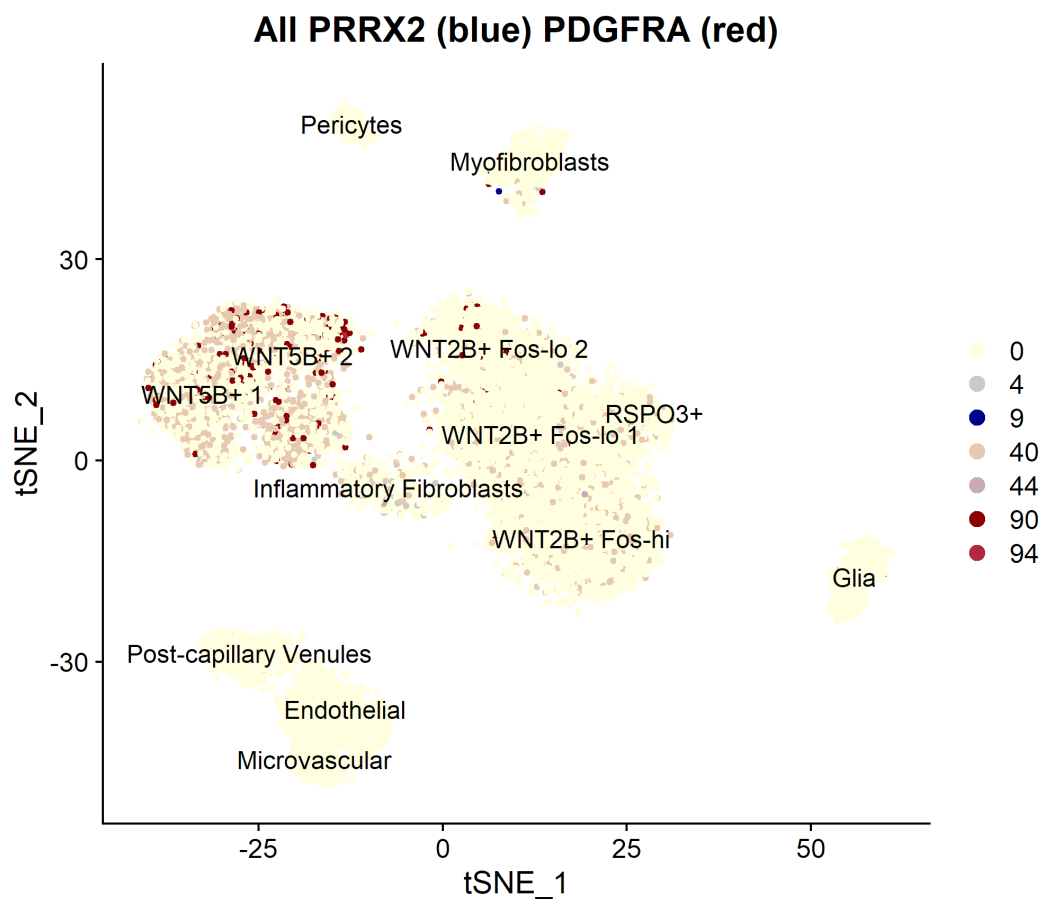
**Figure S4. Subclusters generated by human liver scRNA-seq re-analysis.** Data from Ramachandran et al. (2019). Labeled clusters are shown in Fig. 4.6. Table 4.3 subcluster numbers correspond to the numbers assigned in this t-SNE plot. Each dot represents an individual cell while the color denotes the cluster identity.



**Figure S5. ACOT7 - PDGFRA mRNA co-localization in human colon mesenchyme.** Data from Smillie et al. (2019). Feature plot of ACOT7 across the whole data t-SNE plot.

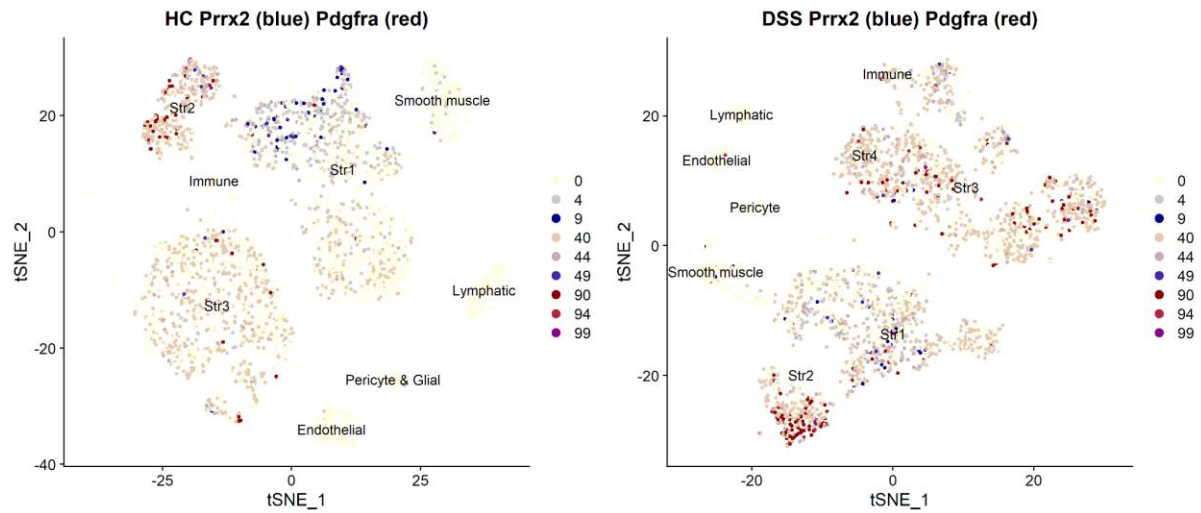


**Figure S6. Acot7 - Pdgfra mRNA co-localization in mouse colon mesenchyme.** Data from Kinchen et al. (2018). HC: healthy control DSS: DSS-challenged.

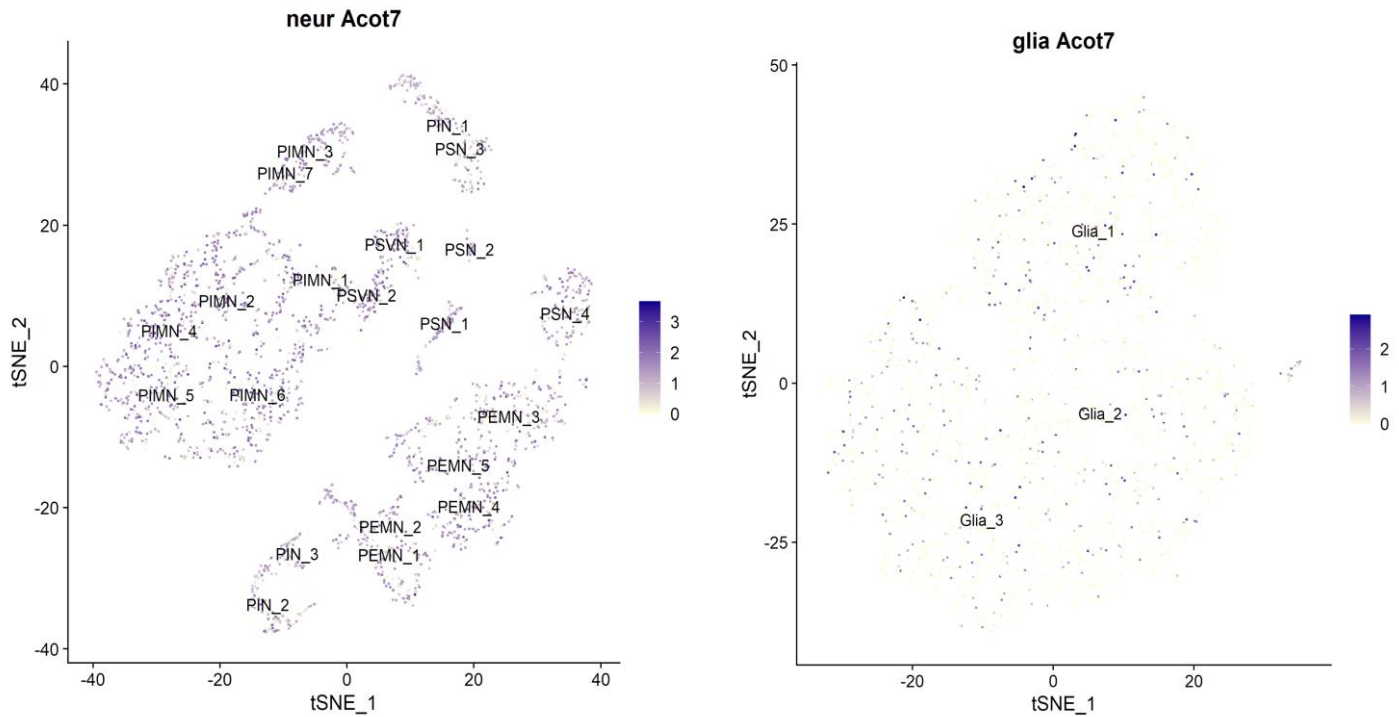


**Figure S7. PRRX2 - PDGFRA mRNA co-localization in human colon mesenchyme.** Data from Smillie et al. (2019). Feature plot of ACOT7 across the whole data t-SNE plot (All).

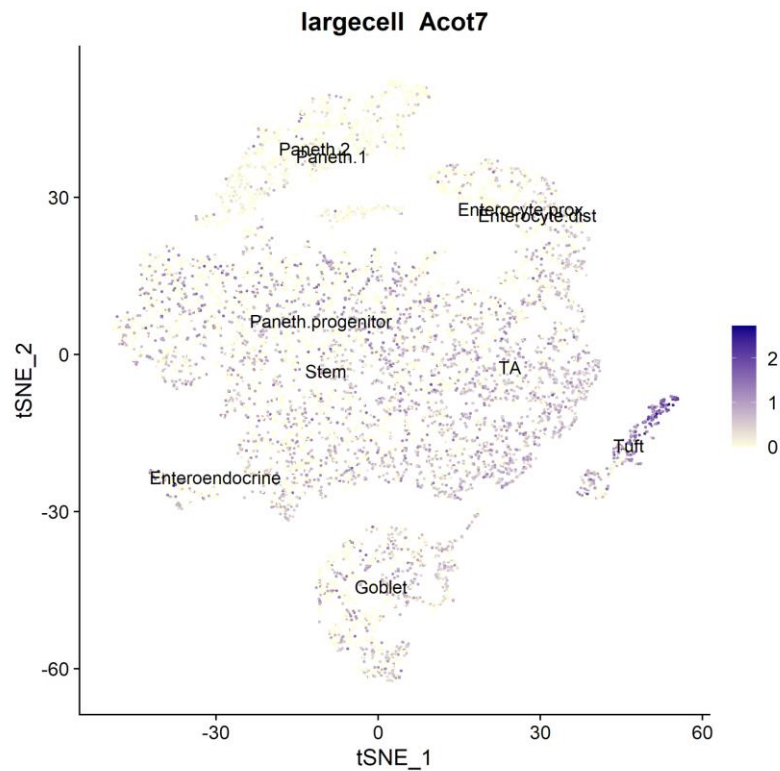




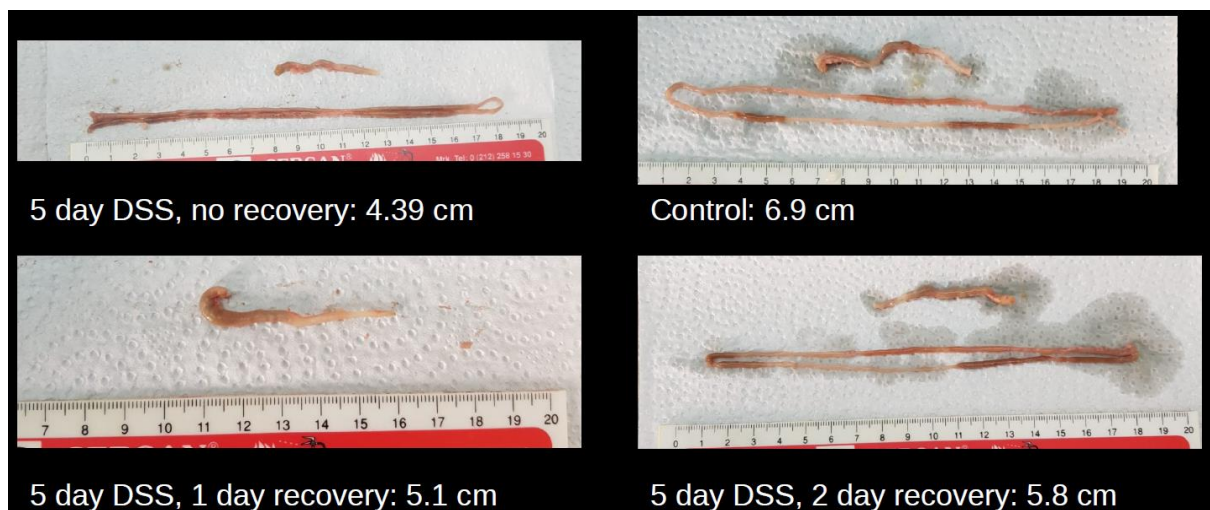
**Figure S8. *Prrx2* - *Pdgfra* mRNA co-localization in mouse colon mesenchyme.** Data from Kinchen et al. (2018). HC: healthy control DSS: DSS-challenged.



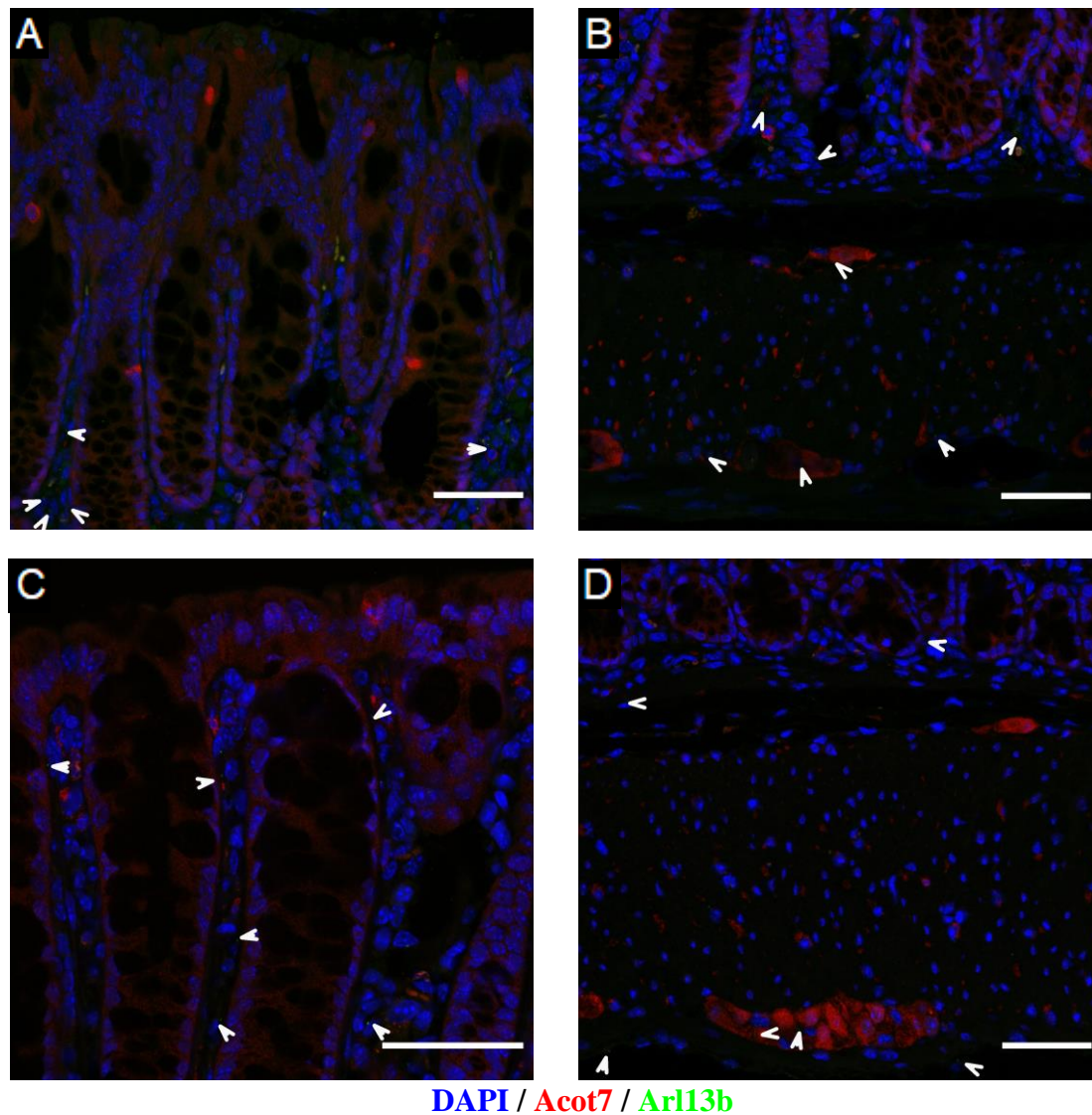
**Figure S9 *Acot7* expression in mouse colon ENS.** Data from Drokhlyansky et al. (2020). *Acot7* is present in both neuronal (left) and glial cells (right), but more abundant in neurons.



**Figure S10. Acot7 expression in mouse small intestine epithelium.** Data from Haber et al. (2017). Acot7 transcript was detected in all cell types, but Tuft cells showed the highest concentration.

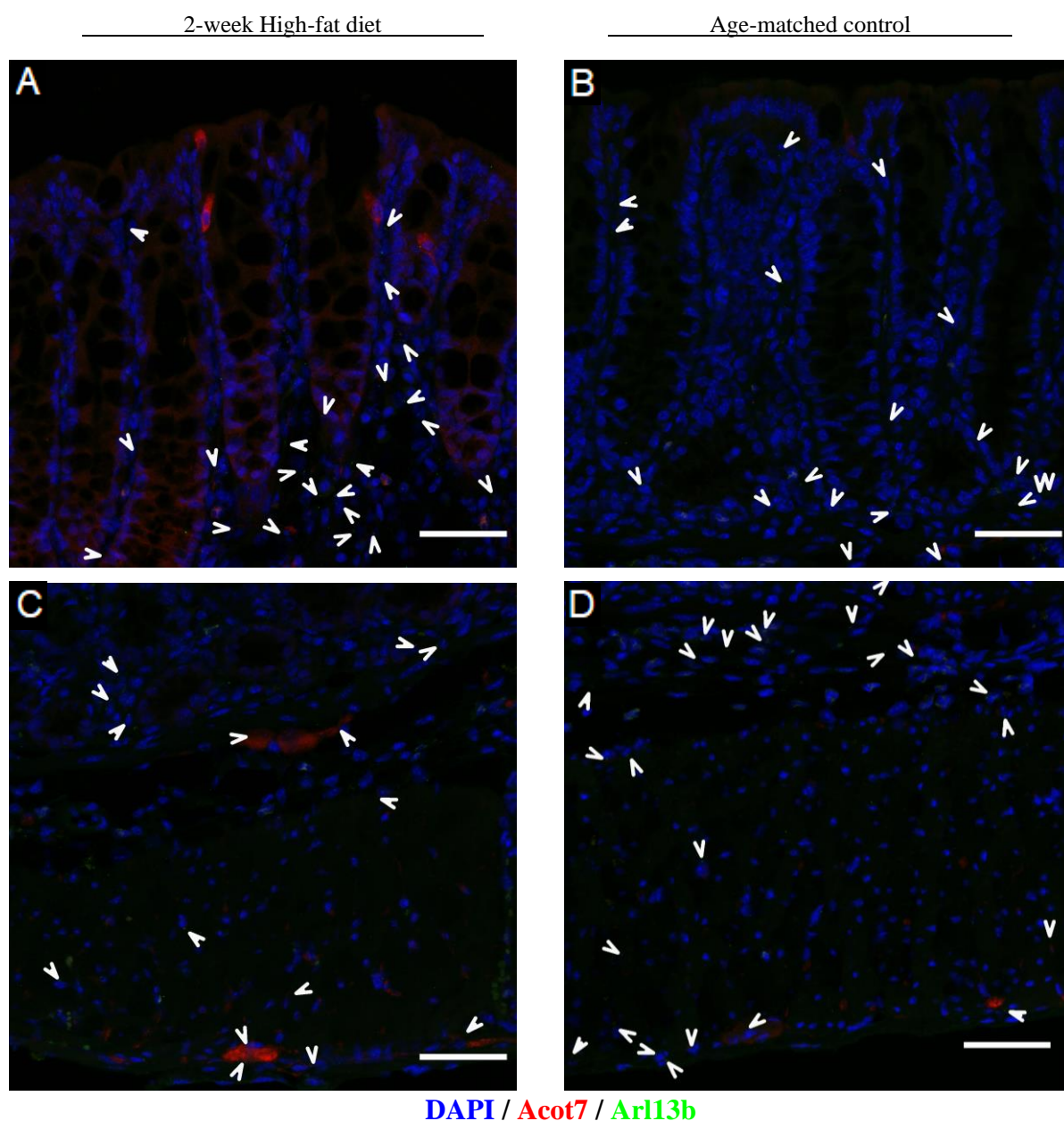


**Figure S11. Mice colons after DSS challenge and recovery.** Mice colon and small intestine were measured in length immediately after isolation. Colon lengths are denoted in the figure.

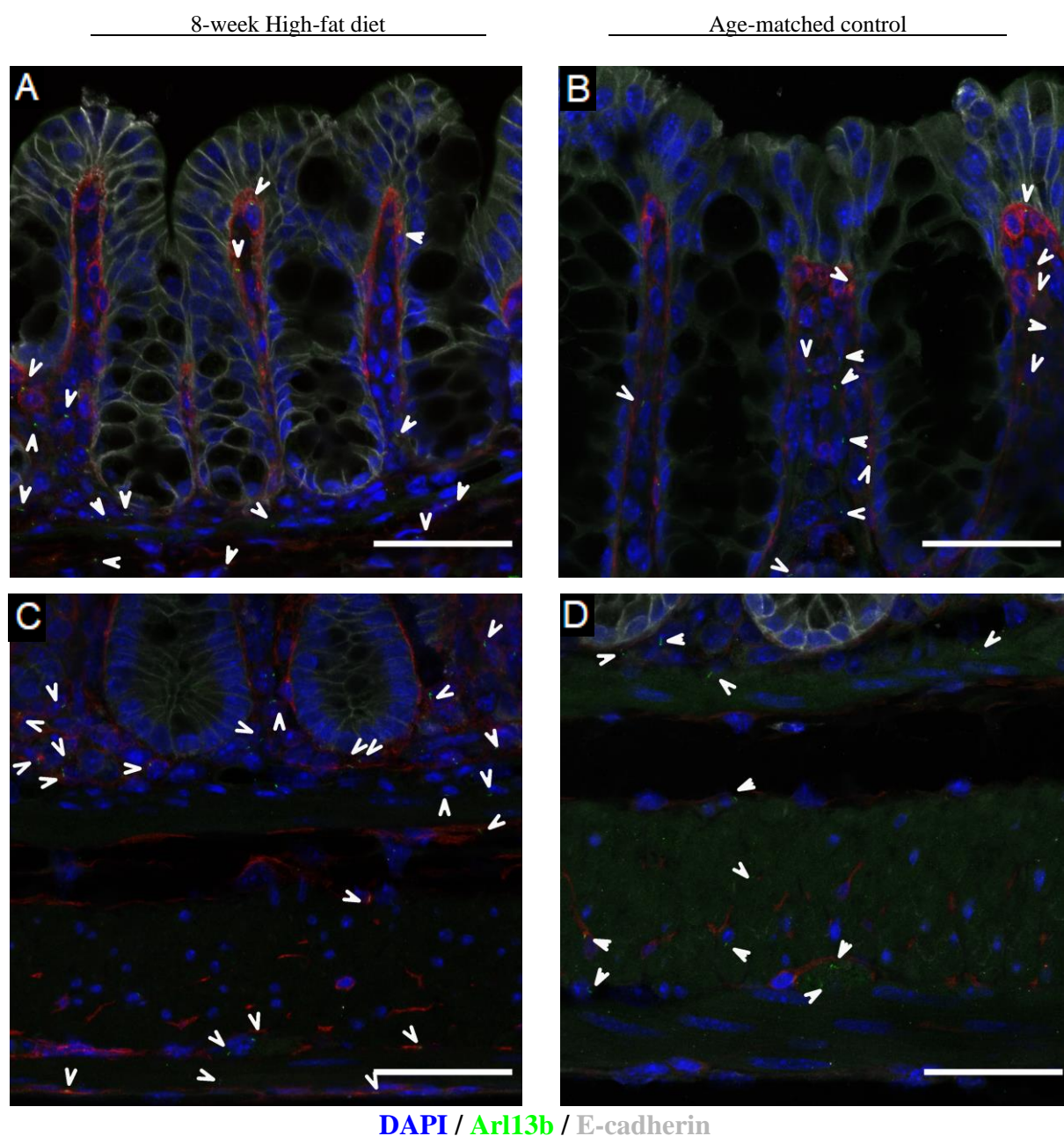


**Figure S12. Acot7 and Arl13b expression after 3 days of recovery from DSS.** There are no discerning differences between healthy and DSS-challenged, 3 day recovered mice colon. A) 3-day recovery, crypt top. B) 3-day recovery, muscle. C) control, crypt top. D) control, muscle. Arrows point to the primary cilia. n = 1. C) 63x magnification, A, B, D) 40x magnification. Scale bars: 50  $\mu$ m.



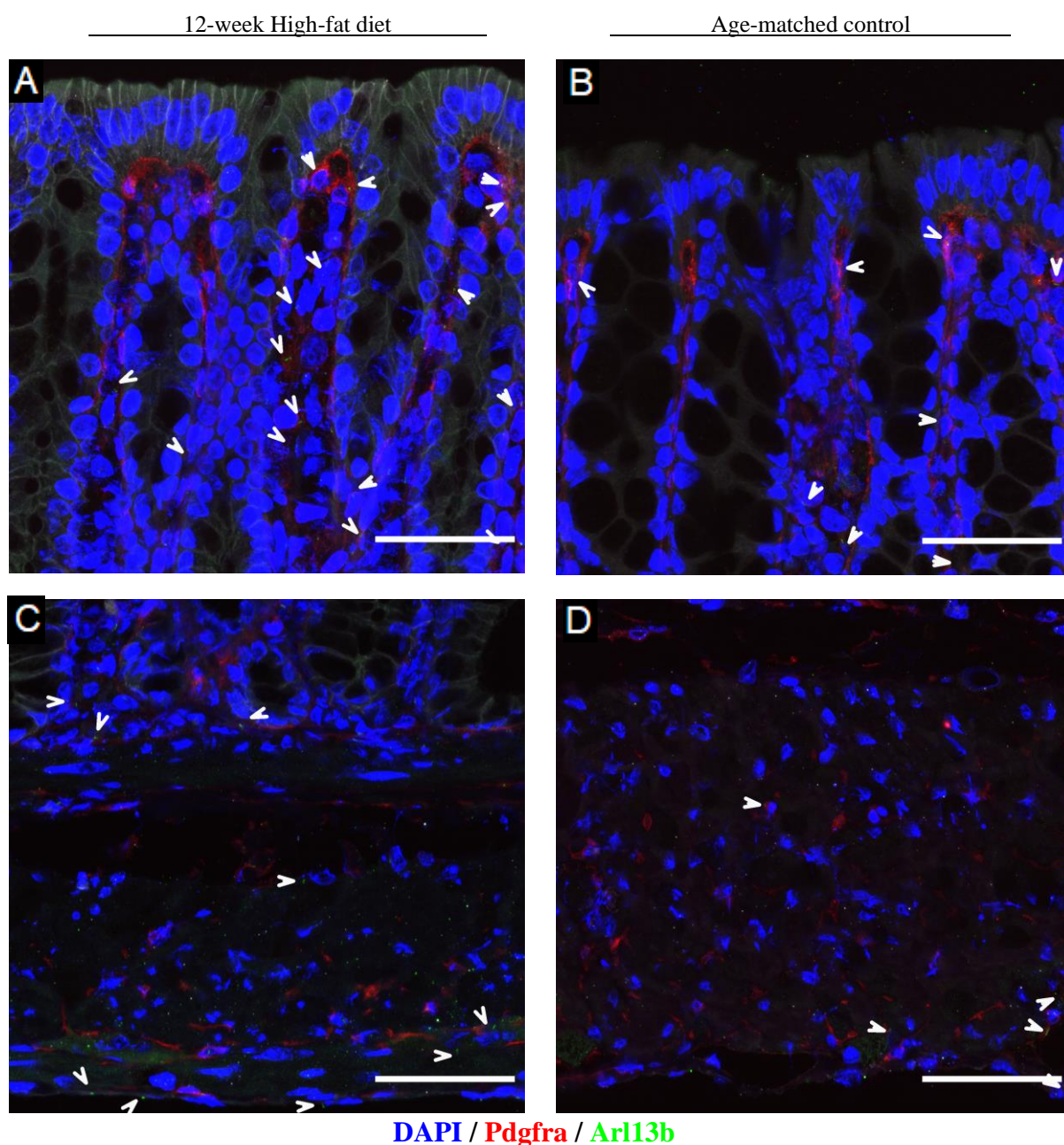


**Figure S13. Primary cilia and *Acot7* expression after 2 weeks on a high-fat diet.** A), C), colon of mice fed HFD for 2 weeks. B), D) colon of age-matched control mice. A), B) show the crypt top region, with the colon lumen towards the top of the images. C), D) are images of the muscle and submucosal region with the muscle located at the bottom of the images. No differences were noted between images obtained from experimental and control group colon sections. n = 2. 40x magnification. Scale bars: 50  $\mu$ m.

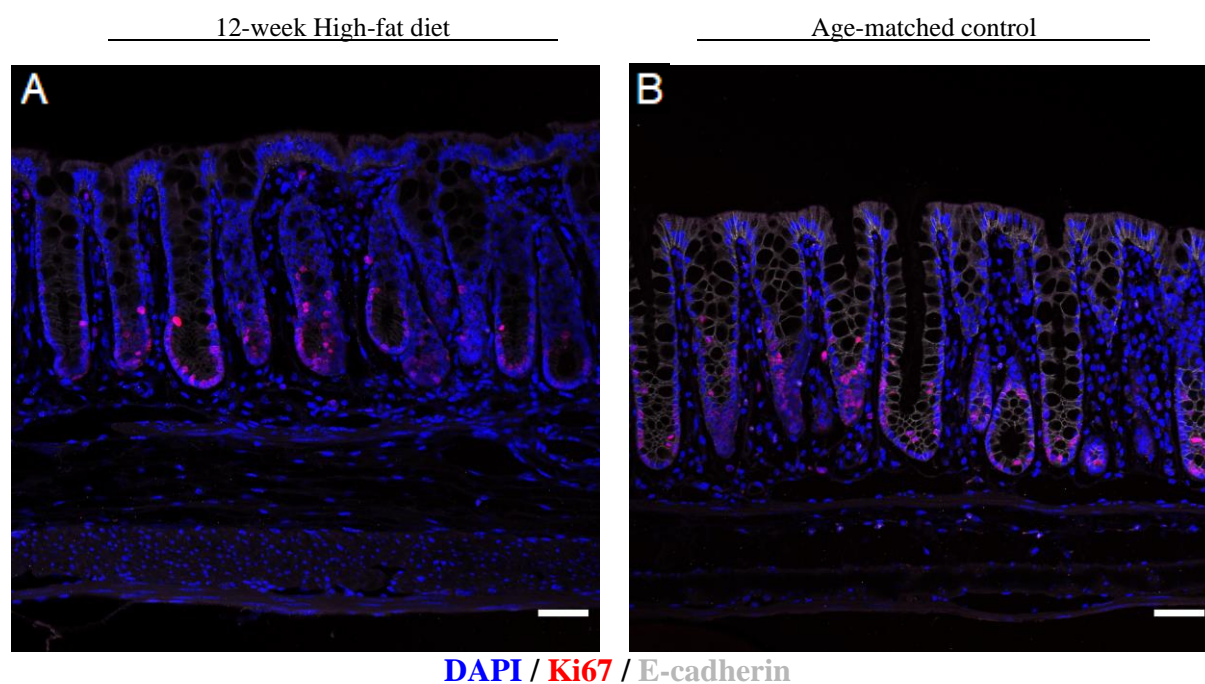


**Figure S14. Primary cilia expression after 8 weeks on a high-fat diet.** A), C), colon of mice fed HFD for 8 weeks. B), D) colon of age-matched control mice. A), B) show the crypt top region, with the colon lumen towards the top of the images. C), D) are images of the muscle and submucosal region with the muscle located at the bottom of the images. No differences were noted between images obtained from experimental and control group colon sections. n = 2. 63x magnification. Scale bars: 50  $\mu$ m.





**Figure S15. Primary cilia expression after 12 weeks on a high-fat diet.** A), C), colon of mice fed HFD for 12 weeks. B), D) colon of age-matched control mice. A), B) show the crypt top region, with the colon lumen towards the top of the images. C), D) are images of the muscle and submucosal region with the muscle located at the bottom of the images. No differences were noted between images obtained from experimental and control group colon sections.  $n = 2$ . 63x magnification. Scale bars: 50  $\mu\text{m}$ .



**Figure S16. Ki67 expression after 12 weeks on a high-fat diet.** Ki67 is restricted to the crypt bottom. A) Representative image of mice colon fed HFD for 12 weeks. B) colon of age-matched control mouse. No differences were noted between images obtained from experimental and control group colon sections. n = 2. 20x magnification. Scale bars: 50  $\mu$ m.

### C. Supplementary Tables

**Table S1.** Mouse colon crypt lengths after 16 weeks of HFD, random measurement of 100 clearly distinguishable crypts on confocal microscopy images, sample image available in Fig. 4.18.

Crypt no.	Length (control 16 weeks)	Length (HFD 16 weeks)	Crypt no. (ctd.)	Length (control 16 weeks)	Length (HFD 16 weeks)	Crypt no. (ctd.)	Length (control 16 weeks)	Length (HFD 16 weeks)
1	179,044	85,118	32	156,817	131,317	63	129,215	127,185
2	187,759	95,596	33	181,324	138,562	64	142,344	122,091
3	186,375	112,224	34	168,339	135,094	65	137,418	98,809
4	189,668	98,299	35	147,425	131,308	66	133,977	84,114
5	158,367	105,689	36	158,001	75,431	67	124,362	77,231
6	136,316	74,311	37	127,8	98,946	68	165,602	151,742
7	149,343	81,365	38	128,096	186,134	69	158,255	118,532
8	139,819	113,732	39	152,42	196,854	70	184,016	150,756
9	116,722	98,299	40	144,291	134,213	71	196,651	136,953
10	120,773	90,148	41	115,019	203,611	72	134,058	181,845
11	82,377	95,444	42	84,288	201,922	73	97,731	186,343
12	207,029	95,444	43	111,157	98,834	74	70,8	199,997
13	197,839	123,8	44	97,011	88,113	75	132,879	187,255
14	161,559	70,636	45	140,235	155,628	76	133,539	159,703
15	179,892	97,521	46	149,774	131,518	77	164,594	136,256
16	162,798	96,297	47	157,956	127,789	78	132,884	127,568
17	155,794	88,53	48	134,706	112,449	79	155,677	108,546
18	146,954	88,201	49	132,322	117,99	80	151,15	138,672
19	129,783	76,305	50	110,756	124,113	81	139,859	111,705
20	148,516	92,184	51	153,748	125,259	82	150,98	127,799
21	145,159	83,328	52	161,281	125,259	83	153,894	111,913
22	135,323	103,912	53	111,373	106,854	84	157,44	159,617
23	134,42	106,85	54	146,168	84,229	85	123,706	112,452
24	107,804	97,016	55	136,764	121,011	86	152,834	104,716
25	75	111,701	56	144,166	127,211	87	144,413	115,655
26	120,232	94,292	57	132,938	112,132	88	162,676	111,6
27	119,919	104,708	58	119,809	149,968	89	159,094	72,747
28	114,941	105,794	59	115,482	130,187	90	141,501	211,305
29	137,736	102,254	60	125,502	132,84	91	141,449	168,51
30	105,129	107,553	61	140,902	142,108	92	134,453	174,392
31	166,663	111,35	62	144,743	144,348	93	147,622	182,284



Crypt no. (ctd.)	Length (control 16 weeks)	Length (HFD 16 weeks)
94	132,389	186,135
95	270,05	177,956
96	257,283	187,866
97	273,375	195,98
98	229,434	191,149
99	229,436	106,063
100	245,745	121,225

Mean	148,225	125,238
SD	36,535	35,561
Min	70,8	70,636
Max	273,375	211,305

## 7. REFERENCES

1. Zhu, G., Hu, J., & Xi, R. (2021). The cellular niche for intestinal stem cells: a team effort. In *Cell Regeneration* (Vol. 10, Issue 1, pp. 1–16). Springer. <https://doi.org/10.1186/s13619-020-00061-5>
2. Degirmenci, B., Hausmann, G., Valenta, T. & Basler, K. Wnt ligands as a part of the stem cell niche in the intestine and the liver. *Prog. Mol. Biol. Transl. Sci.* 153, 1–19 (2018)
3. Larsen, H. L., & Jensen, K. B. (2021). Reprogramming cellular identity during intestinal regeneration. *Current Opinion in Genetics & Development*, 70, 40–47. <https://doi.org/10.1016/j.gde.2021.05.005>
4. Ayyaz, A., Kumar, S., Sangiorgi, B., Ghoshal, B., Gosio, J., Ouladan, S., Fink, M., Barutcu, S., Trcka, D., Shen, J., Chan, K., Wrana, J. L., & Gregorieff, A. (2019). Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature*, 569(7754), 121–125. <https://doi.org/10.1038/s41586-019-1154-y>
5. Yui, S., Azzolin, L., Maimets, M., Pedersen, M. T., Fordham, R. P., Hansen, S. L., Larsen, H. L., Guiu, J., Alves, M. R. P., Rundsten, C. F., Johansen, J. V., Li, Y., Madsen, C. D., Nakamura, T., Watanabe, M., Nielsen, O. H., Schweiger, P. J., Piccolo, S., & Jensen, K. B. (2018). YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell*, 22(1), 35–49.e7. <https://doi.org/10.1016/j.stem.2017.11.001>
6. De, A. (2011). Wnt/Ca<sup>2</sup> signaling pathway: A brief overview. In *Acta Biochimica et Biophysica Sinica* (Vol. 43, Issue 10, pp. 745–756). Oxford Academic. <https://doi.org/10.1093/abbs/gmr079>
7. Degirmenci, B., Valenta, T., Dimitrieva, S., Hausmann, G., & Basler, K. (2018). GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells. *Nature*, 558(7710), 449–453. <https://doi.org/10.1038/s41586-018-0190-3>
8. Harnack, C., Berger, H., Antanaviciute, A., Vidal, R., Sauer, S., Simmons, A., Meyer, T. F., & Sigal, M. (2019). R-spondin 3 promotes stem cell recovery and epithelial regeneration in the colon. *Nature Communications*, 10(1), 1–15. <https://doi.org/10.1038/s41467-019-12349-5>
9. Miyoshi, H., Ajima, R., Luo, C. T., Yamaguchi, T. P., & Stappenbeck, T. S. (2012). Wnt5a potentiates TGF- $\beta$  signaling to promote colonic crypt regeneration after tissue injury. *Science*, 338(6103), 108–113. <https://doi.org/10.1126/science.1223821>
10. Walton, K. D., & Gumucio, D. L. (2021). Hedgehog Signaling in Intestinal Development and Homeostasis. In *Annual Review of Physiology* (Vol. 83, pp. 359–380). *Annual Reviews Inc.* <https://doi.org/10.1146/annurev-physiol-031620-094324>

11. Zhang, Y., & Que, J. (2020). BMP Signaling in Development, Stem Cells, and Diseases of the Gastrointestinal Tract. In *Annual Review of Physiology* (Vol. 82, pp. 251–273). *Annual Reviews Inc.* <https://doi.org/10.1146/annurev-physiol-021119-034500>
12. McCarthy, N., Manieri, E., Storm, E. E., Saadatpour, A., Luoma, A. M., Kapoor, V. N., Madha, S., Gaynor, L. T., Cox, C., Keerthivasan, S., Wucherpennig, K., Yuan, G. C., de Sauvage, F. J., Turley, S. J., & Shivdasani, R. A. (2020). Distinct Mesenchymal Cell Populations Generate the Essential Intestinal BMP Signaling Gradient. *Cell Stem Cell*, 26(3), 391-402.e5. <https://doi.org/10.1016/j.stem.2020.01.008>
13. Kurokawa, K., Hayakawa, Y., & Koike, K. (2021). Plasticity of intestinal epithelium: Stem cell niches and regulatory signals. *International Journal of Molecular Sciences*, 22(1), 1–13. <https://doi.org/10.3390/ijms22010357>
14. Noah, T. K., & Shroyer, N. F. (2013). Notch in the intestine: Regulation of homeostasis and pathogenesis. In *Annual Review of Physiology* (Vol. 75, pp. 263–288). *Annual Reviews* . <https://doi.org/10.1146/annurev-physiol-030212-183741>
15. Kinchen, J., Chen, H. H., Parikh, K., Antanaviciute, A., Jagielowicz, M., Fawcner-Corbett, D., ... Simmons, A. (2018). Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. *Cell*, 175(2), 372-386.e17. <https://doi.org/10.1016/j.cell.2018.08.067>
16. Brugger, M. D., Valenta, T., Fazilaty, H., Hausmann, G., & Basler, K. (2020). Distinct populations of crypt-associated fibroblasts act as signaling hubs to control colon homeostasis. *PLoS Biology*, 18(12). <https://doi.org/10.1371/journal.pbio.3001032>
17. Guérit, E., Arts, F., Dachy, G., Boulouadnine, B., & Demoulin, J. B. (2021). PDGF receptor mutations in human diseases. In *Cellular and Molecular Life Sciences* (Vol. 78, Issue 8, pp. 3867–3881). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s00018-020-03753-y>
18. Moya, I. M., & Halder, G. (2019). Hippo–YAP/TAZ signalling in organ regeneration and regenerative medicine. In *Nature Reviews Molecular Cell Biology* (Vol. 20, Issue 4, pp. 211–226). Nature Publishing Group. <https://doi.org/10.1038/s41580-018-0086-y>
19. Yui, S., Azzolin, L., Maimets, M., Pedersen, M. T., Fordham, R. P., Hansen, S. L., Larsen, H. L., Guiu, J., Alves, M. R. P., Rundsten, C. F., Johansen, J. V., Li, Y., Madsen, C. D., Nakamura, T., Watanabe, M., Nielsen, O. H., Schweiger, P. J., Piccolo, S., & Jensen, K. B. (2018). YAP/TAZ-Dependent Reprogramming of Colonic Epithelium Links ECM Remodeling to Tissue Regeneration. *Cell Stem Cell*, 22(1), 35-49.e7. <https://doi.org/10.1016/j.stem.2017.11.001>
20. Iruzubieta, P., Cantarero, I., Monzón, M., Lahoz, M., & Junquera, C. (2020). Supporting Evidence of Human Enteric Nervous System Adult Neurogenesis: Presence of Primary Cilia and Adult Neurogenesis Markers. *Cellular and Molecular Neurobiology*, 1–9. <https://doi.org/10.1007/s10571-020-01017-8>

21. Anvarian, Z., Mykytyn, K., Mukhopadhyay, S., Pedersen, L. B., & Christensen, S. T. (2019). Cellular signalling by primary cilia in development, organ function and disease. In *Nature Reviews Nephrology*
22. Rocha, C., Papon, L., Cacheux, W., Marques Sousa, P., Lascano, V., Tort, O., Giordano, T., Vacher, S., Lemmers, B., Mariani, P., Meseure, D., Medema, J. P., Bièche, I., Hahne, M., & Janke, C. (2014). Tubulin glycosylases are required for primary cilia, control of cell proliferation and tumor development in colon. *The EMBO Journal*, 33(19), 2247–2260. <https://doi.org/10.15252/embj.201488466>
23. Niewiadomski, P., Niedziółka, S. M., Markiewicz, Ł., Uśpieński, T., Baran, B., & Chojnowska, K. (2019). Gli Proteins: Regulation in Development and Cancer. *Cells*, 8(2), 147. <https://doi.org/10.3390/cells8020147>
24. Moore, K., Fulmer, D., Guo, L., Koren, N., Glover, J., Moore, R., Gensemer, C., Beck, T., Morningstar, J., Stairley, R., & Norris, R. A. (2021). PDGFR $\alpha$ : Expression and function during mitral valve morphogenesis. *Journal of Cardiovascular Development and Disease*, 8(3). <https://doi.org/10.3390/JCDD8030028>
25. Pala, R., Alomari, N., & Nauli, S. M. (2017). Primary cilium-dependent signaling mechanisms. In *International Journal of Molecular Sciences* (Vol. 18, Issue 11, p. 2272). MDPI AG. <https://doi.org/10.3390/ijms18112272>
26. Rausch, V., & Hansen, C. G. (2020). The Hippo Pathway, YAP/TAZ, and the Plasma Membrane. In *Trends in Cell Biology* (Vol. 30, Issue 1, pp. 32–48). Elsevier Ltd. <https://doi.org/10.1016/j.tcb.2019.10.005>
27. Fung, C., & Vanden Berghe, P. (2020). Functional circuits and signal processing in the enteric nervous system. In *Cellular and Molecular Life Sciences* (Vol. 77, Issue 22, pp. 4505–4522). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s00018-020-03543-6>
28. Drokhlyansky, E., Smillie, C. S., Van Wittenberghe, N., Ericsson, M., Griffin, G. K., Eraslan, G., Dionne, D., Cuoco, M. S., Goder-Reiser, M. N., Sharova, T., Kuksenkov, O., Aguirre, A. J., Boland, G. M., Graham, D., Rozenblatt-Rosen, O., Xavier, R. J., & Regev, A. (2020). The Human and Mouse Enteric Nervous System at Single-Cell Resolution. *Cell*, 182(6), 1606–1622.e23. <https://doi.org/10.1016/j.cell.2020.08.003>
29. Holland, A. M., Bon-Frauches, A. C., Keszthelyi, D., Melotte, V., & Boesmans, W. (2021). The enteric nervous system in gastrointestinal disease etiology. In *Cellular and Molecular Life Sciences* (Vol. 1, p. 3). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s00018-021-03812-y>
30. Morarach, K., Mikhailova, A., Knoflach, V., Memic, F., Kumar, R., Li, W., Ernfors, P., & Marklund, U. (2021). Diversification of molecularly defined myenteric neuron classes revealed by single-cell RNA sequencing. *Nature Neuroscience*, 24(1), 34–46. <https://doi.org/10.1038/s41593-020-00736-x>

31. Iino, S., Horiguchi, K., Horiguchi, S., & Nojyo, Y. (2009). C-Kit-negative fibroblast-like cells express platelet-derived growth factor receptor  $\alpha$  in the murine gastrointestinal musculature. *Histochemistry and Cell Biology*, 131(6), 691–702. <https://doi.org/10.1007/s00418-009-0580-6>
32. Foong, D., Zhou, J., Zarrouk, A., Ho, V., & O'Connor, M. D. (2020). Understanding the biology of human interstitial cells of cajal in gastrointestinal motility. In *International Journal of Molecular Sciences* (Vol. 21, Issue 12, pp. 1–18). MDPI AG. <https://doi.org/10.3390/ijms21124540>
33. McDowell C, Farooq U, Haseeb M. Inflammatory Bowel Disease. [Updated 2020 Jun 28]. In: StatPearls [Internet]. Treasure Island (FL): *StatPearls Publishing*; 2021 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK470312/>
34. Smillie, C. S., Biton, M., Ordovas-Montanes, J., Sullivan, K. M., Burgin, G., Graham, D. B., ... Regev, A. (2019). Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell*, 178(3), 714-730.e22. <https://doi.org/10.1016/j.cell.2019.06.029>
35. Kobayashi, T., Siegmund, B., Le Berre, C., Wei, S. C., Ferrante, M., Shen, B., Bernstein, C. N., Danese, S., Peyrin-Biroulet, L., & Hibi, T. (2020). Ulcerative colitis. *Nature Reviews Disease Primers*, 6(1), 1–20. <https://doi.org/10.1038/s41572-020-0205-x>
36. Yao, D., Dong, M., Dai, C., & Wu, S. (2019). Inflammation and Inflammatory Cytokine Contribute to the Initiation and Development of Ulcerative Colitis and Its Associated Cancer. *Inflammatory Bowel Diseases*, 25(10), 1595–1602. <https://doi.org/10.1093/ibd/izz149>
37. Gajendran, M., Loganathan, P., Jimenez, G., Catinella, A. P., Ng, N., Umapathy, C., Ziade, N., & Hashash, J. G. (2019). A comprehensive review and update on ulcerative colitis,. In *Disease-a-Month* (Vol. 65, Issue 12, p. 100851). Mosby Inc. <https://doi.org/10.1016/j.disamonth.2019.02.004>
38. Chassaing, B., Aitken, J. D., Malleshappa, M., & Vijay-Kumar, M. (2014). Dextran sulfate sodium (DSS)-induced colitis in mice. *Current Protocols in Immunology*, 104(SUPPL.104), 15.25.1-15.25.14. <https://doi.org/10.1002/0471142735.im1525s104>
39. Eichele, D. D., & Kharbanda, K. K. (2017). Dextran sodium sulfate colitis murine model: An indispensable tool for advancing our understanding of inflammatory bowel diseases pathogenesis. In *World Journal of Gastroenterology* (Vol. 23, Issue 33, pp. 6016–6029). Baishideng Publishing Group Co. <https://doi.org/10.3748/wjg.v23.i33.6016>
40. Bábíčková, J., Tóthová, L., Lengyelová, E., Bartoňová, A., Hodosy, J., Gardlík, R., & Celec, P. (2015). Sex Differences in Experimentally Induced Colitis in Mice: a Role for

Estrogens. *Inflammation*, 38(5), 1996–2006. <https://doi.org/10.1007/s10753-015-0180-7>

41. Nascimento-Gonçalves, E., Mendes, B. A. L., Silva-Reis, R., Faustino-Rocha, A. I., Gama, A., & Oliveira, P. A. (2021). Animal models of colorectal cancer: From spontaneous to genetically engineered models and their applications. In *Veterinary Sciences* (Vol. 8, Issue 4). MDPI AG. <https://doi.org/10.3390/vetsci8040059>
42. Veauthier, B., & Hornecker, J. R. (2018). Crohn's Disease: Diagnosis and Management. In *American Family Physician* (Vol. 98, Issue 11). [www.aafp.org/afp](http://www.aafp.org/afp)
43. Torres, J., Mehandru, S., Colombel, J. F., & Peyrin-Biroulet, L. (2017). Crohn's disease. In *The Lancet* (Vol. 389, Issue 10080, pp. 1741–1755). Lancet Publishing Group. [https://doi.org/10.1016/S0140-6736\(16\)31711-1](https://doi.org/10.1016/S0140-6736(16)31711-1)
44. Mihaylova, M. M., Cheng, C. W., Cao, A. Q., Tripathi, S., Mana, M. D., Bauer-Rowe, K. E., Abu-Remaileh, M., Clavain, L., Erdemir, A., Lewis, C. A., Freinkman, E., Dickey, A. S., La Spada, A. R., Huang, Y., Bell, G. W., Deshpande, V., Carmeliet, P., Katajisto, P., Sabatini, D. M., & Yilmaz, Ö. H. (2018). Fasting Activates Fatty Acid Oxidation to Enhance Intestinal Stem Cell Function during Homeostasis and Aging. *Cell Stem Cell*, 22(5), 769–778.e4. <https://doi.org/10.1016/j.stem.2018.04.001>
45. Tan, J., Ni, D., Ribeiro, R. V., Pinget, G. V., & Macia, L. (2021). How changes in the nutritional landscape shape gut immunometabolism. In *Nutrients* (Vol. 13, Issue 3, pp. 1–36). MDPI AG. <https://doi.org/10.3390/nu13030823>
46. Beyaz, S., Mana, M. D., Roper, J., Kedrin, D., Saadatpour, A., Hong, S. J., Bauer-Rowe, K. E., Xifaras, M. E., Akkad, A., Arias, E., Pinello, L., Katz, Y., Shinagare, S., Abu-Remaileh, M., Mihaylova, M. M., Lamming, D. W., Dogum, R., Guo, G., Bell, G. W., ... Yilmaz, Ö. H. (2016). High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature*, 531(7592), 53–58. <https://doi.org/10.1038/nature17173>
47. Calibasi-Kocal, G., Mashinchian, O., Basbinar, Y., Ellidokuz, E., Cheng, C. W., & Yilmaz, Ö. H. (2021). Nutritional Control of Intestinal Stem Cells in Homeostasis and Tumorigenesis. In *Trends in Endocrinology and Metabolism* (Vol. 32, Issue 1, pp. 20–35). Elsevier Inc. <https://doi.org/10.1016/j.tem.2020.11.003>
48. Moor, A. E., Golan, M., Massasa, E. E., Lemze, D., Weizman, T., Shenhav, R., Baydatch, S., Mizrahi, O., Winkler, R., Golani, O., Stern-Ginossar, N., & Itzkovitz, S. (2017). Global mRNA polarization regulates translation efficiency in the intestinal epithelium. *Science*, 357(6357), 1299–1303. <https://doi.org/10.1126/science.aan2399>
49. Tarasiuk, A., Mosińska, P., & Fichna, J. (2018). The mechanisms linking obesity to colon cancer: An overview. In *Obesity Research and Clinical Practice* (Vol. 12, Issue 3, pp. 251–259). Elsevier Ltd. <https://doi.org/10.1016/j.orcp.2018.01.005>



50. Takahashi, H., Hosono, K., Endo, H., & Nakajima, A. (2013). Colon epithelial proliferation and carcinogenesis in diet-induced obesity. *Journal of Gastroenterology and Hepatology* (Australia), 28(S4), 41–47. <https://doi.org/10.1111/jgh.12240>
51. Pan, P., Yu, J., & Wang, L. S. (2019). Diet and colon: What matters? In *Current Opinion in Gastroenterology* (Vol. 35, Issue 2, pp. 101–106). Lippincott Williams and Wilkins. <https://doi.org/10.1097/MOG.0000000000000501>
52. Singh, S., Dulai, P. S., Zarrinpar, A., Ramamoorthy, S., & Sandborn, W. J. (2017). Obesity in IBD: Epidemiology, pathogenesis, disease course and treatment outcomes. In *Nature Reviews Gastroenterology and Hepatology* (Vol. 14, Issue 2, pp. 110–121). Nature Publishing Group. <https://doi.org/10.1038/nrgastro.2016.181>
53. Hases, L., Archer, A., Indukuri, R., Birgersson, M., Savva, C., Korach-André, M., & Williams, C. (2020). High-fat diet and estrogen impacts the colon and its transcriptome in a sex-dependent manner. *Scientific Reports*, 10(1), 16160. <https://doi.org/10.1038/s41598-020-73166-1>
54. Xie, Y., Ding, F., Di, W., Lv, Y., Xia, F., Sheng, Y., Yu, J., & Ding, G. (2020). Impact of a high-fat diet on intestinal stem cells and epithelial barrier function in middle-aged female mice. *Molecular Medicine Reports*, 21(3), 1133–1144. <https://doi.org/10.3892/mmr.2020.10932>
55. Hasegawa, Y., Chen, S. Y., Sheng, L., Jena, P. K., Kalanetra, K. M., Mills, D. A., Wan, Y. J. Y., & Slupsky, C. M. (2020). Long-term effects of western diet consumption in male and female mice. *Scientific Reports*, 10(1), 14686. <https://doi.org/10.1038/s41598-020-71592-9>
56. Albillos, A., de Gottardi, A., & Rescigno, M. (2020). The gut-liver axis in liver disease: Pathophysiological basis for therapy. In *Journal of Hepatology* (Vol. 72, Issue 3, pp. 558–577). Elsevier B.V. <https://doi.org/10.1016/j.jhep.2019.10.003>
57. Ben-Moshe, S., & Itzkovitz, S. (2019). Spatial heterogeneity in the mammalian liver. *Nature Reviews Gastroenterology and Hepatology*, 16(7), 395–410. <https://doi.org/10.1038/s41575-019-0134-x>
58. Masyuk, A. I., Masyuk, T. V., & LaRusso, N. F. (2008). Cholangiocyte primary cilia in liver health and disease. In *Developmental Dynamics* (Vol. 237, Issue 8, pp. 2007–2012). John Wiley & Sons, Ltd. <https://doi.org/10.1002/dvdy.21530>
59. Khanam, A., Saleeb, P. G., & Kottlil, S. (2021). Pathophysiology and Treatment Options for Hepatic Fibrosis: Can It Be Completely Cured? *Cells*, 10(5), 1097. <https://doi.org/10.3390/cells10051097>
60. Parola, M., & Pinzani, M. (2019). Liver fibrosis: Pathophysiology, pathogenetic targets and clinical issues. In *Molecular Aspects of Medicine* (Vol. 65, pp. 37–55). Elsevier Ltd. <https://doi.org/10.1016/j.mam.2018.09.002>

61. Kirpich, I. A., Marsano, L. S., & McClain, C. J. (2015). Gut-liver axis, nutrition, and non-alcoholic fatty liver disease. In *Clinical Biochemistry* (Vol. 48, Issues 13–14, pp. 923–930). Elsevier Inc. <https://doi.org/10.1016/j.clinbiochem.2015.06.023>
62. Svensson, V., Vento-Tormo, R., & Teichmann, S. A. (2018). Exponential scaling of single-cell RNA-seq in the past decade. In *Nature Protocols* (Vol. 13, Issue 4, pp. 599–604). *Nature Publishing Group*. <https://doi.org/10.1038/nprot.2017.149>
63. Lafzi, A., Moutinho, C., Picelli, S., & Heyn, H. (2018). Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. In *Nature Protocols* (Vol. 13, Issue 12). <https://doi.org/10.1038/s41596-018-0073-y>
64. Wolfien, M., David, R., & Galow, A.-M. (2021). Single-Cell RNA Sequencing Procedures and Data Analysis. In *Bioinformatics* (pp. 19–35). Exon Publications. <https://doi.org/10.36255/exonpublications.bioinformatics.2021.ch2>
65. Stuart and Butler et al. Comprehensive Integration of Single-Cell Data. *Cell* (2019).
66. Haber, A. L., Biton, M., Rogel, N., Herbst, R. H., Shekhar, K., Smillie, C., ... Regev, A. (2017). A single-cell survey of the small intestinal epithelium. *Nature*, 551(7680), 333–339. <https://doi.org/10.1038/nature24489>
67. Roulis, M., Kaklamanos, A., Scherthanner, M., Bielecki, P., Zhao, J., Kaffe, E., ... Flavell, R. A. (2020). Paracrine orchestration of intestinal tumorigenesis by a mesenchymal niche. *Nature*, 580(7804), 524–529. <https://doi.org/10.1038/s41586-020-2166-3>
68. Dobie, R., Wilson-Kanamori, J. R., Henderson, B. E. P., Smith, J. R., Matchett, K. P., Portman, J. R., ... Henderson, N. C. (2019). Single-Cell Transcriptomics Uncovers Zonation of Function in the Mesenchyme during Liver Fibrosis. *Cell Reports*, 29(7), 1832-1847.e8. <https://doi.org/10.1016/j.celrep.2019.10.024>
69. Halpern, K. B., Shenhav, R., Massalha, H., Toth, B., Egozi, A., Massasa, E. E., ... Itzkovitz, S. (2018). Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells. *Nature Biotechnology*, 36(10), 962–970. <https://doi.org/10.1038/nbt.4231>
70. Ramachandran, P., Dobie, R., Wilson-Kanamori, J. R., Dora, E. F., Henderson, B. E. P., Luu, N. T., ... Henderson, N. C. (2019). Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature*, 575(7783), 512-518. <https://doi.org/10.1038/s41586-019-1631-3>
71. Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., Huang, D., Xu, Y., Huang, W., Jiang, M., Jiang, X., Mao, J., Chen, Y., Lu, C., Xie, J., ... Guo, G. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*, 172(5), 1091-1107.e17. <https://doi.org/10.1016/j.cell.2018.02.001>



72. Schiller, H. B., Montoro, D. T., Simon, L. M., Rawlins, E. L., Meyer, K. B., Strunz, M., Vieira Braga, F. A., Timens, W., Koppelman, G. H., Budinger, G. R. S., Burgess, J. K., Waghray, A., Van Den Berge, M., Theis, F. J., Regev, A., Kaminski, N., Rajagopal, J., Teichmann, S. A., Misharin, A. V., & Nawijn, M. C. (2019). The human lung cell atlas: A high-resolution reference map of the human lung in health and disease. In *American Journal of Respiratory Cell and Molecular Biology* (Vol. 61, Issue 1, pp. 31–41). American Thoracic Society. <https://doi.org/10.1165/rcmb.2018-0416TR>
73. Wang, Y., Song, W., Wang, J., Wang, T., Xiong, X., Qi, Z., Fu, W., Yang, X., & Chen, Y.-G. (2020). Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *Journal of Experimental Medicine*, 217(2). <https://doi.org/10.1084/jem.20191130>
74. Parikh, K., Antanaviciute, A., Fawcner-Corbett, D., Jagielowicz, M., Aulicino, A., Lagerholm, C., Davis, S., Kinchen, J., Chen, H. H., Alham, N. K., Ashley, N., Johnson, E., Hublitz, P., Bao, L., Lukomska, J., Andev, R. S., Björklund, E., Kessler, B. M., Fischer, R., ... Simmons, A. (2019). Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature*. <https://doi.org/10.1038/s41586-019-0992-y>
75. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
76. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
77. Douglas Bates and Martin Maechler (2019). Matrix: Sparse and Dense Matrix Classes and Methods. R package version 1.2-18. <https://CRAN.R-project.org/package=Matrix>
78. Hadley Wickham (2020). tidyr: Tidy Messy Data. R package version 1.1.2. <https://CRAN.R-project.org/package=tidyr>
79. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* Jan 2019;47(D1):D419-D426
80. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
81. Schindelin, J.; Arganda-Carreras, I. & Frise, E. et al. (2012), "Fiji: an open-source platform for biological-image analysis", *Nature methods* 9(7): 676-682, PMID 22743772, doi:10.1038/nmeth.2019
82. Schneider, C. A.; Rasband, W. S. & Eliceiri, K. W. (2012), "NIH Image to ImageJ: 25 years of image analysis", *Nature methods* 9(7): 671-675, PMID 22930834
83. Linkert, M.; Rueden, C. T. & Allan, C. et al. (2010), "Metadata matters: access to image data in the real world", *The Journal of cell biology* 189(5): 777-782, PMID 20513764

84. Zhu, G., Hu, J., & Xi, R. (2021). The cellular niche for intestinal stem cells: a team effort. In *Cell Regeneration* (Vol. 10, Issue 1, pp. 1–16). *Springer*. <https://doi.org/10.1186/s13619-020-00061-5>
85. Delling, M., Decaen, P. G., Doerner, J. F., Febvay, S., & Clapham, D. E. (2013). Primary cilia are specialized calcium signalling organelles. *Nature*, 504(7479), 311–314. <https://doi.org/10.1038/nature12833>
86. Haghverdi, L., Lun, A. T. L., Morgan, M. D., & Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, 36(5), 421–427. <https://doi.org/10.1038/nbt.4091>
87. Ferent, J., Constable, S., Gigante, E. D., Yam, P. T., Mariani, L. E., Legué, E., Liem, K. F., Caspary, T., & Charron, F. (2019). The Ciliary Protein Arl13b Functions Outside of the Primary Cilium in Shh-Mediated Axon Guidance. *Cell Reports*, 29(11), 3356–3366.e3. <https://doi.org/10.1016/j.celrep.2019.11.015>
88. Ellis, J. M., Wong, G. W., & Wolfgang, M. J. (2013). Acyl Coenzyme A Thioesterase 7 Regulates Neuronal Fatty Acid Metabolism To Prevent Neurotoxicity. *Molecular and Cellular Biology*, 33(9), 1869–1882. <https://doi.org/10.1128/mcb.01548-12>
89. Hunt, M. C., Greene, S., Hultenby, K., Svensson, L. T., Engberg, S., & Alexson, S. E. H. (2007). Alternative exon usage selectively determines both tissue distribution and subcellular localization of the acyl-CoA thioesterase 7 gene products. *Cellular and Molecular Life Sciences*, 64(12), 1558–1570. <https://doi.org/10.1007/s00018-007-7062-6>
90. Cheng, L., Jin, H., Qiang, Y., Wu, S., Yan, C., Han, M., Xiao, T., Yan, N., An, H., Zhou, X., Shao, Q., & Xia, S. (2016). High fat diet exacerbates dextran sulfate sodium induced colitis through disturbing mucosal dendritic cell homeostasis. *International Immunopharmacology*, 40, 1–10. <https://doi.org/10.1016/j.intimp.2016.08.018>