Intestinal Stromal Cell Types in Health and Inflammatory Bowel Disease uncovered by Single-Cell Transcriptomics



James Kinchen

St Cross College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

This thesis is dedicated to
my daughter Isabel

Acknowledgements

I would like to thank Professor Alison Simmons for supervising this work, and members of the Simmons Group for providing a friendly and stimulating work environment.

Particular thanks are due to Dr Kaushal Parikh and Dr Hannah Chen for their contribution to the wet lab validation of predictions generated from single-cell transcriptomic models and for supplying immunohistochemistry (IHC), single-molecule in-situ hybridisation (smISH) and flow cytometry data featured in this document. I would also like to thank Stina Lindberg and Felix Richter, who assisted with individual experiments under my supervision.

Single cell capture, library preparation and sequencing were conducted in collaboration with the Single Cell Facilities at both the Wellcome Trust Centre for Human Genetics and the Weatherall Institute of Molecular Medicine with significant contributions from Drs Neil Ashley, Rory Bowden, Esther Mellado, Moustafa Attar, Jerome Nicod, and Eshita Sharma.

I am grateful to the many endoscopy patients who donated tissue for use in this study and to Dr James East, Dr Alissa Walsh, and the Oxford Research Nurses for Immunology and Inflammatory Diseases (ORNIID) for their help in obtaining these samples.

In the bioinformatic aspects of this work, I received invaluable advice and guidance from Dr Quin Wills and Dr Davis McCarthy. Funding for my fellowship was generously provided by the Lee Placito Medical Fund, and I am grateful to Oxfordshire Health Services Research Committee for additional financial support.

Attributions

This thesis is the result of my own work with the exception of the data listed below, which was generated by colleagues as indicated. The choice of target molecules selected for validation was the result of a collaborative discussion based around predictions generated from the single cell transcriptomic models.

Dr Kaushal Parikh supplied the IHC and smISH images featured in Figures 3.16, 4.4, 4.5, 4.6, 4.14, 5.7, and Appendix B, Figure B.2.

Dr Hannah Chen produced the flow cytometric data that appears in Figures 4.7, 4.14, 4.15, 4.16, 4.18, and Appendix B, Figure B.3.

The murine DSS challenge was performed by Epistem®, who supplied colon tissue for use in single-cell experiments and recorded in-life and post mortem observations.

Abstract

Colonic stromal cells provide critical structural support but also regulate immunity, tolerance and inflammatory responses in the mucosa. Substantial variability and plasticity of mucosal stromal cells has been reported but a paucity of distinct marker genes exist to identify distinct cell states. Here single-cell RNA-sequencing is used to document heterogeneity and subtype specific markers of individual colonic stromal cells in human and mouse.

Marker-free transcriptional clustering of fibroblast-like cells derived from healthy human tissue reveals distinct populations corresponding to myofibroblasts and three transcriptionally and functionally dissimilar populations of fibroblasts. A *SOX6* high fibroblast subset occupies a position adjacent to the epithelial basement membrane and expresses multiple epithelial morphogens including *WNT5A* and *BMP2*. Additional fibroblast subtypes show specific enrichment for chemokine signalling and prostaglandin E_2 synthesis respectively.

In ulcerative colitis, substantial remodelling occurs with depletion of the *SOX6* high population and emergence of an immune enriched population expressing genes associated with fibroblastic reticular cells including *CCL19*, *CCL21* and *IL33*.

A large murine dataset comprising over 7,000 colonic mesenchymal cells from an acute colitis model and matched healthy controls reveals strong preservation of the *SOX6* high and myofibroblast transcriptional signatures. Unsupervised pseudotemporal ordering is used to relate fibroblast subsets to one another producing a branched developmental hierarchy that includes a potential progenitor population with mesothelial characteristics at its origin. This work provides a molecular basis for re-classification of colonic stromal cells and identifies pathological changes in these cells underpinning inflammation in UC.

Contents

List of Figures **xi**

List of Abbreviations **xv**

1 Introduction **1**

1.1 The intestinal mucosa 2
1.2 Anatomy of the colon 3
1.3 Existing classification of colonic stromal cells 5
1.3.1 Fibroblasts 6
1.3.2 Pericryptal myofibroblasts 7
1.3.3 Pericytes 7
1.3.4 Smooth muscle elements 8
1.3.5 Bone-marrow derived stromal stem cells 8
1.4 Candidate markers for additional mesenchymal cell subsets 9
1.4.1 Thymocyte differentiation antigen-1 (THY1 / CD90) 9
1.4.2 Platelet-derived growth factor receptor alpha (PDGFRA) . . 10
1.4.3 Amine oxidase, copper containing 3 (AOC3) 10
1.4.4 Prostaglandin-endoperoxide synthase 2 (PTGS2 / COX2) . . 10
1.4.5 Forkhead Box L1 (FOXL1) 11
1.5 Knockout models and ablation approaches 11
1.6 Developmental origins 13
1.7 Replacement of mesenchymal cell types in the adult 14
1.8 Stromal interaction with the epithelial compartment 16
1.8.1 Hedgehog signalling 17
1.8.2 Notch signalling 18
1.8.3 Fibroblast growth factors 19
1.8.4 Platelet-derived growth factors 20
1.8.5 Bone morphogenetic proteins 20
1.8.6 Maintenance of the epithelial stem cell niche 21
1.9 Stromal cells in mucosal immunity 23
1.9.1 Innate sensors 23

1.9.2 Cytokine-mediated effects . . . 24
1.9.3 Interaction with professional immune cells . . . 25
1.9.4 Contribution to lymphoid tissues . . . 27
1.10 Stromal contribution to disease states . . . 29
1.10.1 Inflammatory bowel disease . . . 29
1.10.2 Colorectal cancer . . . 32
1.11 Single-cell RNA sequencing . . . 34
1.12 Discussion . . . 35

2 Materials and Methods **37**
2.1 Cell preparation — Fluidigm C1 . . . 38
2.1.1 Human sample collection . . . 38
2.1.2 Isolation of human colonic epithelial cells . . . 38
2.1.3 Isolation of human colonic stromal cells . . . 39
2.1.4 Cell capture, library preparation and sequencing . . . 39
2.1.5 Modifications for multiplexed experiments . . . 40
2.2 Bioinformatic approach — Fluidigm C1 . . . 40
2.2.1 Alignment and transcriptome assembly . . . 40
2.2.2 Cell and gene quality control . . . 41
2.2.3 Differential expression analysis . . . 42
2.2.4 Library size normalisation . . . 42
2.2.5 Technical batch effect adjustment . . . 44
2.2.6 Weighted gene co-expression network analysis . . . 44
2.2.7 Cell clustering . . . 45
2.2.8 Marker gene identification and ontology enrichment . . . 46
2.2.9 Module preservation . . . 47
2.2.10 Reduced dimensionality representations . . . 47
2.3 Cell preparation — 10X genomics . . . 47
2.3.1 Dextran sodium sulphate (DSS) challenge . . . 47
2.3.2 Isolation of murine colonic stromal cells . . . 48
2.3.3 Single cell capture, library preparation and sequencing . . . 49
2.4 Bioinformatic approach — 10X genomics . . . 50
2.4.1 Cellranger pipeline . . . 50
2.4.2 Identification of cellular barcodes . . . 50
2.4.3 Initial clustering and QC . . . 50
2.4.4 Cell cycle annotation, normalisation and final clustering . . . 51
2.4.5 Further Seurat analysis functions . . . 51
2.4.6 Diffusion maps and diffusion pseudotime . . . 52
2.5 Statistical methods . . . 52

2.6 Common validation methods . . . 53
2.6.1 Flow cytometry . . . 53
2.6.2 Immunohistochemistry (IHC) . . . 54
2.6.3 Single molecule in-situ hybridisation (smISH) . . . 55
2.6.4 qPCR . . . 55
2.7 Key Resources Table . . . 56

3 Single cell RNAseq of human colonic mucosal cells 59
3.1 Introduction . . . 59
3.1.1 Motivation . . . 61
3.2 Results . . . 61
3.2.1 Cell isolation . . . 61
3.2.2 Single cell capture, imaging and cDNA library generation . . . 63
3.2.3 Single cell transcriptome assembly . . . 64
3.2.4 Cell and gene QC process . . . 64
3.2.5 Differential gene expression . . . 65
3.2.6 Epithelial mesenchymal reciprocal signalling . . . 69
3.2.7 Innate immune function in the the mesenchymal compartment 71
3.2.8 Correlation network analysis . . . 76
3.2.9 Cell clustering based on module expression . . . 81
3.2.10 Evaluation of clustering in epithelial dataset . . . 84
3.2.11 Comparison with other clustering methods . . . 87
3.3 Discussion . . . 89

4 Intestinal stromal cell types 91
4.1 Introduction . . . 92
4.1.1 Motivation . . . 93
4.2 Results . . . 93
4.2.1 WGCNA-based clustering captures an important source of heterogeneity in human colonic stromal cells . . . 93
4.2.2 Characterization of human colonic stromal subtypes . . . 95
4.2.3 Single-cell transcriptomes of UC-derived colonic stromal cells 104
4.2.4 Altered global expression of epithelial morphogens in UC stromal cells . . . 105
4.2.5 An additional gene co-expression module and cell cluster is detected in the stroma in acute UC . . . 107
4.2.6 Validation of the UC-specific Stromal 4 subset . . . 111
4.2.7 Inflammation-associated changes in stromal composition in UC 113
4.2.8 Single-cell transcriptomes of colonic stromal cells from patients with UC-associated dysplasia . . . 115

4.2.9 Transcriptional changes induced by TNF-alpha stimulation of colonic stromal cells . . . 119
4.3 Discussion . . . 121

5 Murine colonic stroma **125**
5.1 Introduction . . . 126
5.1.1 Motivation . . . 128
5.2 Results . . . 129
5.2.1 DSS challenge . . . 129
5.2.2 Single-cell dissociation and capture . . . 130
5.2.3 Sequencing results and assessment of batch effect . . . 132
5.2.4 Unsupervised clustering of stromal cells from the control group 132
5.2.5 Conservation of human Stromal 1–3 transcriptional signatures in mouse . . . 136
5.2.6 DSS-induced changes in stromal composition . . . 141
5.2.7 Expansion of a Stromal 4-like fibroblast subset on DSS challenge 144
5.2.8 Biological activities of murine stromal subsets in health and DSS colitis . . . 149
5.2.9 Ontogeny of stromal fibroblast subsets . . . 151
5.2.10 Subset-specific transcriptional responses to DSS challenge . . 154
5.2.11 Specificities of existing fibroblast markers and novel marker candidates . . . 156
5.3 Discussion . . . 157

6 Concluding Discussion **160**
6.1 Introduction . . . 160
6.2 Generation of interpretable stromal single-cell transcriptomes from primary human and murine tissue samples . . . 162
6.3 Development of an unsupervised cell classification method for single cell RNAseq data using WGCNA and NMF . . . 163
6.4 Single cell RNAseq identifies aspects of colonic stromal gene expression not accessible to bulk methods . . . 166
6.5 Multiple transcriptionally and spatially distinct fibroblast subsets are detected within the human colonic lamina propria . . . 167
6.6 Changes in stromal composition and gene expression associated with ulcerative colitis . . . 169
6.7 Similar fibroblast transcriptional heterogeneity is observed in a murine model . . . 171
6.8 Defining the stromal response to DSS-induced acute colitis . . . 173

6.9 Specificities of existing murine Cre models and identification of new candidates 175
6.10 Wider significance — stromal biology 175
6.11 Wider significance — Inflammatory Bowel Disease 176
6.12 Future directions 177

Appendices

A Supplementary Material for Chapter 3 **180**
A.1 Supplementary figures 180
A.2 Supplementary tables 194

B Supplementary Material for Chapter 4 **196**
B.1 Supplementary figures 196
B.2 Supplementary tables 201

C Supplementary Material for Chapter 5 **204**
C.1 Supplementary figures 204

References **224**

List of Figures

1.1 Anatomy of the colon 4

3.1 Isolation of human colonic epithelial cells 62
3.2 Isolation of human colonic mesenchymal cells 63
3.3 Identification of genes showing differential expression in the epithelial and mesenchymal compartments 66
3.4 Differentially expressed marker genes and processes in the colonic epithelial and mesenchymal compartments 67
3.5 Reciprocal paracrine signalling pathways identified from single-cell RNAseq data 68
3.6 Enrichment for innate-immune processes in colonic mesenchymal cells 72
3.7 KEGG pathway enrichment: NOD-Like receptor signaling 73
3.8 KEGG pathway enrichment: Antigen processing and presentation . 74
3.9 KEGG pathway enrichment: Phagosome 75
3.10 Detection of gene co-expression modules 78
3.11 Fundamental network concepts in detected modules 79
3.12 Functional ontology enrichment analysis in detected modules 80
3.13 Module eigengene correlation 81
3.14 Consensus clustering of colonic single cells 83
3.15 Epithelial marker gene expression in detected cell clusters 85
3.16 Validation of known and novel goblet cell markers by immunohistochemistry 86
3.17 Comparison of clustering methods in the epithelial dataset 88

4.1 Four distinct fibroblast subsets are identified in healthy colonic samples 94
4.2 Transcription factor and ontology enrichment in the myofibroblast (MF) subset 96
4.3 Transcription factor and ontology enrichment in stromal 1–3 97
4.4 Tissue distribution of detected stromal subsets by sm-ISH 100
4.5 Tissue distribution of detected stromal subsets by IHC 102
4.6 CD45$^-$ stromal subsets are anatomically separated 103

4.7 Distinct fibroblast sub-populations identified by FACS 104
4.8 Sample multiplexing on Fluidigm C1 IFCs 105
4.9 Differential gene expression between healthy and UC lamina propria stromal cells 107
4.10 WGCNA module detection and preservation in UC stromal cells . . 109
4.11 Ontology enrichment for UC-specific 'green' module 110
4.12 NMF clustering of UC stromal cells 110
4.13 Stromal subset 4 marker genes 111
4.14 Stromal FDCSP expression in healthy and UC specimens 112
4.15 Stromal PDPN and CD74 expression in healthy and UC specimens 112
4.16 Intracellular cytokine staining in colonic stromal cells from patients with active UC 113
4.17 Differentially expression analysis comparing sites of inflammation in UC to adjacent uninvolved mucosa 114
4.18 Alterations in stromal composition at sites of active inflammation compared to adjacent uninvolved mucosa 115
4.19 Differential stromal cell gene expression in dysplastic UC 117
4.20 Stromal composition by cluster assignment in dysplastic UC 118
4.21 KEGG pathway enrichment: TNFα signaling pathway 120
4.22 Transcriptional response to TNFα stimulation in freshly-isolated human stromal cells 121

5.1 DSS colitis and control mice, in-life and post-mortem observations . 130
5.2 Murine DSS challenge: summary of experimental design 131
5.3 Detection of cell clusters in healthy control mice 133
5.4 Correlation of clusters to expected cell types in healthy controls . . 135
5.5 Significant overlap exists between murine and human fibroblast subsets. 137
5.6 Expression plots of human fibroblast subset markers in the murine dataset 139
5.7 smISH localisation of Stromal 2 marker genes 140
5.8 Phylogenetic tree and identities of murine stromal cell clusters in health 141
5.9 Cell clustering and cluster identity in DSS colitis 143
5.10 An increase in the relative abundance of the Stromal 3 subset is seen in DSS colitis 144
5.11 An immunologically specialised fibroblast subset analogous to human Stromal 4 is identified in the murine DSS model 146
5.12 Phylogenetic tree and identities of murine stromal cell clusters in DSS 147
5.13 Fibroblast subsets show differential proliferative activity on DSS challenge 148

5.14 Common and differential functional ontology enrichment among fibroblast subsets seen in health 150
5.15 Progressive enrichment for stress, activation and proliferation seen within Stromal 3 subset in DSS 151
5.16 Diffusion pseudotime provides a framework to relate observed fibroblast clusters 153
5.17 Fibroblast subsets show differential responses to DSS challenge . . . 155
5.18 Candidate molecular markers for further subset characterisation and functional study 157

A.1 IFC imaging to detect viable single cell capture 181
A.2 Assignment of reads to genomic features. 182
A.3 Effect of duplicate removal on read counts 183
A.4 Single cell library quality measures: depth, coverage and saturation 184
A.5 Single cell library quality measures: spike-in RNAs 185
A.6 Single cell library quality measures: mitochondrial genes 185
A.7 Cell quality control summary 186
A.8 Gene quality control summary 187
A.9 Removal of contaminating non-fibroblast-like cells 188
A.10 Effect of library size normalisation on per-cell median relative log expression value 189
A.11 Adjustment of expression values to reduce unwanted batch effects . 190
A.12 Decomposition of gene variances into biological and technical components 191
A.13 NMF factorisation rank surveys 192
A.14 Module eigengene expression by cell cluster 193

B.1 Additional subset marker violin plots (healthy stroma) 196
B.2 Single molecule ISH technical controls 197
B.3 Representative FACS gating strategies 197
B.4 UC dataset NMF factorisation rank survey and eigengene expression 198
B.5 Expression profiles of selected FDC-related genes in the UC dataset 199
B.6 Module preservation and NMF clustering in dysplastic UC 200

C.1 Further in-life and post-mortem outcomes of DSS challenge 205
C.2 FACS assessment of murine colonic stromal cell isolation 206
C.3 Identification of cellular barcodes — healthy controls 207
C.4 Identification of cellular barcodes — DSS colitis 209
C.5 Identification of spike-in control RAW 264.7 cells 211
C.6 Assessment of channel-associated batch effect using control cells . . 213

C.7 Cell-free RNA produces a small batch effect between conditions . . 215
C.8 Cellular QC 216
C.9 Determination of important principle components for cell clustering 217
C.10 Merging of low-confidence clusters, healthy dataset 218
C.11 Heterogeneity within the smooth muscle / ICC cell cluster 219
C.12 Merging of similar clusters, DSS dataset 220
C.13 Random-forest-based cluster identification in the DSS dataset . . . 221
C.14 Major fibroblast subsets as defined for differential expression analysis 222
C.15 Additional marker t-SNEs 223

List of Abbreviations

AOM/DSS . . Azoxymethane / dextran sodium sulphate. Used in combination in a murine chemically induced model of inflammation-associated carcinogenesis.

AUROC . . . . Area under receiver-operator characteristic.

BMP . . . . . Bone morphogenetic protein.

CAF . . . . . Cancer-associated fibroblast.

CCL/CCR . . C-C motif ligand or receptor. A chemokine signalling family.

CD . . . . . Crohn's disease.

cDNA . . . . . Complementary DNA. Synthesized from a single stranded RNA.

CD31 . . . . . Cluster of differentiation 31 antigen. Also known as Platelet endothelial cell adhesion molecule (PECAM-1). A marker of endothelial cells.

CD45 . . . . . Cluster of differentiation 45 antigen. Also known as Protein tyrosine phosphatase, receptor type C (PTPRC) or leukocyte common antigen (LCA). A pan-leukocyte marker protein.

CD90 . . . . . Cluster of differentiation 90 antigen. Also known as thymocyte differentiation antigen-1 (THY1). A member of the immunoglobulin superfamily and candidate stromal stem cell marker.

CD235a . . . . Cluster of differentiation 235a antigen. Also known as Glycophorin A. A marker of mature erythrocytes.

CRC . . . . . Colorectal cancer.

Cre . . . . . Cre recombinase. A tyrosine recombinase enzyme derived from the P1 bacteriophage. Performs site-specific recombination between two DNA recognition sites (LoxP sites). Used in construction of genetic models.

CXCL/CXCR C-X-C motif ligand or receptor. A chemokine signalling family.

CyTOF . . . . Time of flight mass cytometry.

DTR . . . . . Diphtheria toxin receptor. Conditional transgenic expression of this protein renders cells susceptible to ablation on exposure to diphtheria toxin.

DSS . . . . . Dextran sodium sulphate. A colitogenic chemical.

ECM . . . . . Extracellular matrix.

EMT . . . . . Epithelial-to-mesenchymal transition.

EPCAM . . . Epithelial cell adhesion molecule. Also known as CD326. A marker of epithelial cells.

ERCC . . . . . External RNA Controls Consortium.

Fl . . . . . 'Floxed.' Flanked by LoxP recognition sites and therefore a target for Cre-mediated recombination.

FACS . . . . . Fluorescence-activated cell sorting.

FDC . . . . . Follicular dendritic cell.

FGF . . . . . Fibroblast growth factor.

FRC . . . . . Fibroblastic reticular cell.

FOX . . . . . Forkhead box. A transcription factor family sharing a winged helix DNA binding motif.

GALT . . . . . Gut-associated lymphoid tissue.

GFP . . . . . Green-fluorescent protein.

GO . . . . . Gene ontology.

GVHD . . . . Graft-versus-host disease.

HGF . . . . . Hepatocyte growth factor.

HOX . . . . . Homeobox. A family of transcription factors sharing a DNA binding motif and involved in regulation of anatomical development.

HSC . . . . . Haematopoietic stem cell.

IBD . . . . . Inflammatory Bowel Disease.

ICC . . . . . Interstitial Cell of Cajal.

IFC . . . . . Integrated fluidic circuit.

IHC . . . . . Immunohistochemistry.

IL . . . . . Interleukin.

ILF . . . . . Isolated lymphoid follicle.

ISEMF . . . . Intestinal sub-epithelial myofibroblast.

KEGG . . . . . Kyoto Encyclopaedia of Genes and Genomes.

LGR5 . . . . . Leucine-rich repeat-containing G-protein coupled receptor 5. An important biomarker of intestinal epithelial stem cells.

LPS . . . . . Lipopolysaccharide. A component of the outer membrane of Gram-negative bacteria.

MACS . . . . . Magnetic activated cell sorting.

MHC . . . . . Major histocompatibility complex.

MKI67 . . . . Marker of proliferation Ki-67. A nuclear protein associated with cellular proliferation.

MMP . . . . . Matrix metalloproteinase.

MSC . . . . . Mesenchymal stem cell.

NFκB . . . . . Nuclear factor kappa-light-chain-enhancer of activated B cells.

NG2 . . . . . Neuron-glial antigen 2. Also known as Chondroitin sulphate proteoglycan 4 (CSPG4). A candidate marker of pericytes.

NLR . . . . . NOD-like receptor.

NMF . . . . . Non-negative matrix factorisation.

NOD . . . . . Nucleotide-binding oligomerization domain-containing protein.

PAMP . . . . . Pathogen associated molecular pattern.

PCA . . . . . Principle component analysis.

PDPN . . . . . Podoplanin. Also known as gp38.

PDGF . . . . . Platelet-derived growth factor.

PGE_2 . . . . . Prostaglandin E_2.

PRR . . . . . Pattern recognition receptor.

PTGS2 . . . . Prostaglandin-endoperoxide synthase 2. Also known as cyclooxygenase-2 (COX2).

RNAseq . . . . RNA sequencing.

ROC . . . . . Receiver-operator characteristic.

ROS . . . . . Reactive oxygen species.

RT-qPCR . . . Quantitative reverse transcription polymerase chain reaction.

α-SMA . . . . α-smooth muscle actin. Also known as alpha-actin-2 (ACTA2). A widely used myofibroblast marker.

SILT . . . . . Solitary isolated lymphoid tissue.

sm-ISH . . . . Single molecule in situ hybridisation.

SNP . . . . . Single-nucleotide polymorphism.

t-SNE . . . . . T-distributed stochastic neighbour embedding.

TIMP . . . . . Tissue inhibitor of metalloproteinase.

TGFβ . . . . . Transforming growth factor β.

TNFα . . . . . Tumour necrosis factor α.

TLR . . . . . Toll-like receptor.

TNBS . . . . . 2,4,6-trinitrobenzene sulfonic acid. A colitogenic chemical used in animal models of IBD.

UC . . . . . Ulcerative colitis.

UMI . . . . . Unique molecular identifier.

WGCNA . . . Weighted Gene Co-expression Network Analysis.

Chapter 1

Introduction

Contents

1.1 The intestinal mucosa 2
1.2 Anatomy of the colon 3
1.3 Existing classification of colonic stromal cells 5
1.3.1 Fibroblasts 6
1.3.2 Pericryptal myofibroblasts 7
1.3.3 Pericytes 7
1.3.4 Smooth muscle elements 8
1.3.5 Bone-marrow derived stromal stem cells 8
1.4 Candidate markers for additional mesenchymal cell subsets 9
1.4.1 Thymocyte differentiation antigen-1 (THY1 / CD90) . . 9
1.4.2 Platelet-derived growth factor receptor alpha (PDGFRA) 10
1.4.3 Amine oxidase, copper containing 3 (AOC3) 10
1.4.4 Prostaglandin-endoperoxide synthase 2 (PTGS2 / COX2) 10
1.4.5 Forkhead Box L1 (FOXL1) 11
1.5 Knockout models and ablation approaches 11
1.6 Developmental origins 13
1.7 Replacement of mesenchymal cell types in the adult . 14
1.8 Stromal interaction with the epithelial compartment . 16
1.8.1 Hedgehog signalling 17
1.8.2 Notch signalling 18
1.8.3 Fibroblast growth factors 19
1.8.4 Platelet-derived growth factors 20
1.8.5 Bone morphogenetic proteins 20
1.8.6 Maintenance of the epithelial stem cell niche 21
1.9 Stromal cells in mucosal immunity 23
1.9.1 Innate sensors 23
1.9.2 Cytokine-mediated effects 24
1.9.3 Interaction with professional immune cells 25
1.9.4 Contribution to lymphoid tissues 27
1.10 Stromal contribution to disease states 29

1.10.1 Inflammatory bowel disease 29
1.10.2 Colorectal cancer 32
1.11 Single-cell RNA sequencing 34
1.12 Discussion 35

1.1 The intestinal mucosa

Epithelial surfaces form the interface between multicellular organisms and their environment. While the skin is the the most obvious of these interfaces, it is less significant in terms of area than mucosal surfaces, of which the largest is the intestinal mucosa. Under homeostatic conditions, the intestinal mucosa mediates a number of functions critical to health, including acting as a physical barrier and maintaining nutrient and water homeostasis. It is also an important site of interaction between the immune system, dietary antigens, and the complex microbial community that exists in the gut lumen (the intestinal microbiome). Performance of these functions involves the coordinated activity of multiple diverse cell types, including epithelial cells, professional immune cells, neural cells, and the endothelial cells of capillaries and lymphatic channels. Much progress has been made in elucidating the specific pathways utilized by these players in regulating intestinal immunity and barrier function. Areas of intense interest include development and regulation of the epithelial compartment[1], the role of the microbiota in modulating intestinal immune responses[2], and the composition and function of resident intestinal immune cells[3].

The cellular complement of the intestinal mucosa is completed by a relatively poorly defined group of stromal cells, predominantly fibroblasts and myofibroblasts, which were once thought to be relatively inert structural supports. Several lines of evidence, however, now implicate this group as active participants both in homeostatic processes, such as the maintenance of the epithelial stem cell niche and regulation of epithelial repair following injury, and in clinically important disease states. Therapeutic targeting of pathways specific to the stromal compartment has shown potential in the treatment of both inflammatory bowel disease (IBD) and

cancer[4, 5]. However, much remains to be determined regarding the true nature of these cells, with key questions relating to their origins, heterogeneity, differentiation hierarchy, and functional specialisation remaining unanswered.

1.2 Anatomy of the colon

Starting from the bowel lumen, the colonic mucosal barrier consists of an mucus layer overlying a monolayer of epithelial cells that forms multiple invaginations, or crypts. Villi, the finger-like luminal projections characteristic of the small bowel mucosa, are absent in the colon. Beneath the epithelial monolayer lies the basement membrane, a fenestrated sheet of connective tissue fibrils. Fibroblasts are present in this layer and their cytoplasmic processes surround its fenestrations[6]. Next, there is a mixed cellular compartment known as the lamina propria containing immune cells, stromal cells, vascular, lymphatic and neural elements. At the basal boundary of the lamina propria lie the muscularis mucosae and muscularis propria, thin layers of enteric smooth muscle against which the crypt bases abut. Collectively, the muscularis mucosae, lamina propria and epithelium form the mucosa, which itself overlies the sub-mucosa, a layer of loose connective tissue. This is in turn enclosed in successive layers of smooth muscle fibres in both concentric and longitudinal orientations that, under the influence of local pacemaker cells (the interstitial cells of Cajal, ICCs) and parasympathetic innervation, generate the peristaltic movements that propel the luminal contents. The outermost serosal layer, in continuity with the peritoneum, forms the interface with the abdominal cavity. An overview of this structure is presented in Figure 1.1.

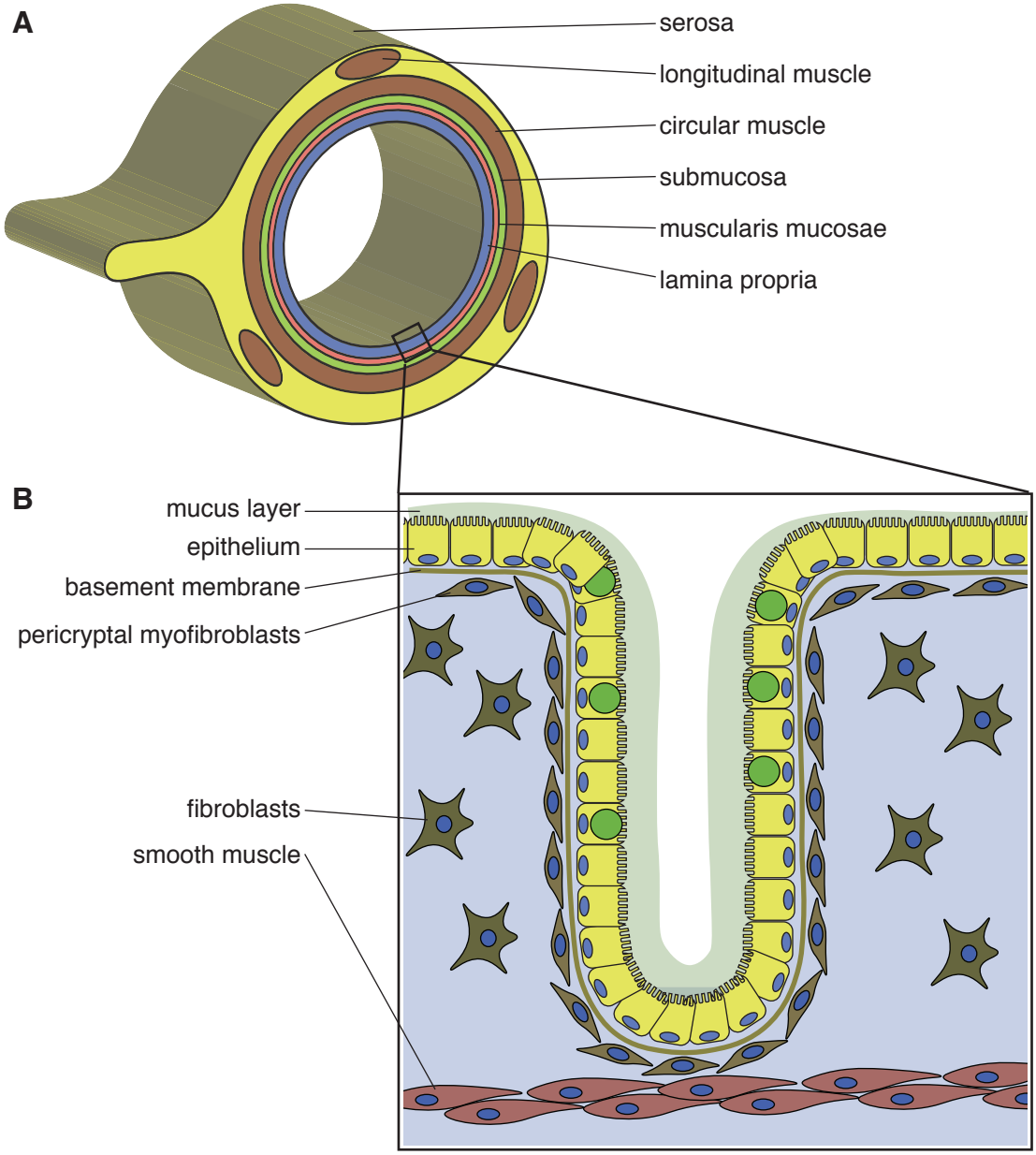


Figure 1.1: Anatomy of the colon. (A) Tubular structure of the colon with distinct tissue layers indicated. (B) Schematic representation of a colonic crypt with stromal cell types indicated.

1.3 Existing classification of colonic stromal cells

The presence of a sheath of fibroblast-like cells within the lamina-propria abutting the colonic epithelium was first described almost 50 years ago[7]. Equivalent cell populations were subsequently identified in other organs of the gastrointestinal tract, including the small bowel and gallbladder, and they were shown to constitute a syncytial cell network connected by gap junctions that extends throughout the lamina propria[8]. They were initially considered to be a single cell population, termed intestinal sub-epithelial myofibroblasts (ISEMFs). This model was subsequently challenged by ultrastructural and marker-based studies that identify a number of distinct cell types, albeit with some overlap in marker expression, contributing to this network and its biological functions[9]. While considerable uncertainty regarding the ontogeny and function of these populations remains, the mesenchymal complement of the lamina propria is now thought to include pericryptal myofibroblasts, fibroblasts, perivascular mural cells (pericytes), distinct populations of smooth muscle cells in the muscularis mucosae and lymphatic lacteals, and bone marrow-derived stromal stem cells[9–12]. The evidence supporting these entities as distinct cell types is reviewed below, while a summary of their morphological characteristics and marker expression is presented in Table 1.1.

Ultrastructure	α-SMA	Vim	Des	Other markers
Fibroblast				
abundant ER, collagen granules	−	+	−	$CD90^+$, $FSP1^+$, $PDGFRA^+$
Myofibroblast				
abundant ER, collagen granules, myofilaments, fibronexus	+	+	−	$CD90^+$, $AOC3^+$, $NG2^+$, FAP^+
Pericyte				
perivascular processes, 'peg-socket' contacts, prominent nucleus	+	+	+	$PDGFRA^+$, $PDGFRB^+$, $VCAM1^+$, $NG2^+$, $RGS5^+$
Smooth muscle cell				
scarce ER, abundant myofilaments, external lamina	++	−	++	$PDGFRA^+$, $AOC3^+$, $NG2^+$

Table 1.1: Comparative features of fibroblasts, myofibroblasts, pericytes and smooth muscle cells. ER, endoplasmic reticulum; SMA, smooth muscle actin; Vim, vimentin; Des, desmin.

1.3.1 Fibroblasts

Intestinal fibroblasts are defined as non-haematopoietic (CD45−), non-endothelial (CD31−) cells that express the intermediate filament vimentin but not α-smooth muscle actin (α-SMA) or desmin. At least two morphologically distinct variants are seen in ultrastructural studies, spindle-shaped cells with slender cytoplasmic processes and enlarged cells with expansion of the endoplasmic reticulum and Golgi apparatus[6]. Intermediate forms were also observed. Fibroblasts contain collagen secretory granules and lack myofilaments and an external lamina, which are characteristic of smooth muscle cells[13]. They are seen in deeper positions within the lamina propria, not immediately adjacent to either the epithelium or microvasculature, as these niches are occupied by other mesenchymal subtypes discussed below. These structural observations suggest a role in modelling the

extracellular matrix (ECM) but no intrinsic contractility.

1.3.2 Pericryptal myofibroblasts

The pericryptal myofibroblasts, also referred to α-SMA positive or sub-epithelial myofibroblasts, occupy a position immediately sub-adjacent to the epithelial basement membrane into which they are embedded. They are characterised by intermediate α-SMA staining in comparison to the high staining seen in smooth muscle cells and the negative staining observed in fibroblasts. Like fibroblasts, they express vimentin but not desmin[14]. While intermediate α-SMA expression is the most widely used defining characteristic of this population, it is not specific as pericytes and lymphatic mural cells also express intermediate levels of this marker.

The term 'myofibroblast' is somewhat controversial in its application to this cell subset and not universally accepted. True myofibroblasts were first defined in the context of granulation tissue as fibroblasts exhibiting some properties of smooth muscle cells, such as myofilaments and gap junctions[15]. In this setting, they differentiate from fibroblast progenitors in a stepwise process that involves the induction of stress fibres composed of β- and γ-cytoplasmic actins and finally α-SMA[16]. They have an intrinsic contractility that is important for granulation tissue contraction and wound healing. While the pericryptal fibroblast-like cells of the intestinal lamina propria are similar to classical myofibroblasts in terms of α-SMA expression, they lack a structure called the fibronexus that connects ECM fibronectin fibrils to aligned intracellular myofilament bundles and is a central part of myofibroblastic differentiation[17]. Based on a strict definition incorporating both ultrastructural and marker based approaches, the sub-epithelial α-SMA positive fibroblast-like cells of the intestinal lamina propria are not true myofibroblasts[13].

1.3.3 Pericytes

Pericytes, or mural cells, are embedded within and contribute to the basement membrane of microvessels[18]. Structurally, they are characterised by a prominent nucleus, long processes extending parallel to the vessel and smaller circumferential processes partially encircling it. Gap junctions interconnect pericytes, and they

interface with underlying endothelial cells via 'peg-socket' contacts and adhesion plaques[19]. Candidate immunohistochemical markers include α-SMA, desmin, neuron-glial antigen 2 (NG2), the platelet-derived growth factor receptor (PDGFR)-β, and regulator of G protein signalling-5 (RGS5)[20]. However, no single marker is specific and the pericyte distribution pattern and molecular identifiers vary in different tissues. Paracrine signalling between pericytes and endothelial cells is important in the regulation of angiogenesis and vascular permeability. This interaction across a basement membrane is analogous to that between pericryptal myofibroblasts and epithelial cells. Furthermore, these two cell subsets share expression of some marker proteins and respond to similar cytokines and growth factors (such as TGFβ and hedgehog ligands)[20, 21]. On this basis, it has been suggested they may share a common origin[9].

1.3.4 Smooth muscle elements

The primary anatomical site of desmin$^+$, α-SMA$^+$, vimentin$^-$ enteric smooth muscle cells within the intestinal mucosa is the muscularis mucosae[22, 23]. This structure forms the boundary between the lamina propria and submucosa and is in close proximity to the bases of the epithelial crypts. This muscle layer is innervated by the submucosal plexus and can respond to changes in the luminal environment via a neurally mediated afferent-efferent link[24]. In this way, mechanical forces at the epithelial surface are coordinated with secretory activity. In the small intestine, a further population of smooth muscle cells are seen surrounding the lymphatic lacteal in the villus core, where their position would suggest a role in the propulsion of lymphatic fluid[25, 26]. Projections of smooth muscle cells are also seen in the inter-crypt spaces in the colon, although the relation to lymphatic vessels in this setting is less clear.

1.3.5 Bone-marrow derived stromal stem cells

The presence of a population bone-marrow derived mesenchymal cells in the lamina propria has been demonstrated using in-situ hybridisation for the Y chromosome in murine bone marrow chimeric models with male donor animals and female recipients

and in human subjects with graft-versus-host disease (GVHD) following allogenic bone marrow transplantation[27]. Whether these cells have true stem cell properties such self renewal and the extent of their differentiation potential have not been fully determined. In the absence of chimerism, bone-marrow derived stromal stem cells are difficult to identify by histological methods.

1.4 Candidate markers for additional mesenchymal cell subsets

While the levels of α-SMA, vimentin and desmin are helpful in delineating the above classification, there is considerable overlap in expression of these markers and none can be considered specific for any given cell subset. A number of additional candidate markers have been proposed and linked to particular cell populations or aspects of fibroblast biology.

1.4.1 Thymocyte differentiation antigen-1 (THY1 / CD90)

Heterogenous expression of the surface marker CD90, or thymocyte differentiation antigen-1 (THY1), has been observed among fibroblasts isolated from the orbit and myometrium. In both cases, only the $CD90^+$ population had the potential to differentiate into myofibroblasts when stimulated with TGFβ/ *in vitro*, while only the $CD90^-$ population demonstrated lipofibroblastic differentiation when stimulated with 15-deoxy-δ-prostaglandin-J_2[28]. The two populations also showed differing patterns of cytokine, prostaglandin, and collagen secretion on stimulation, and differential expression of major histocompatibility complex (MHC) class II[29]. Heterogeneity in CD90 expression is also observed in colonic mesenchymal cells, with the majority of α-SMA positive cells co-expressing this marker together with a subset of the α-SMA negative fibroblasts[30].

1.4.2 Platelet-derived growth factor receptor alpha (PDGFRA)

The discovery of a distinct population of PDGFR-α^+ myofibroblasts associated with enteric muscle fibres provided impetus to test the distribution of this marker among colonic lamina propria stromal cells in human tissues and in a PDGFR-α-eGFP reporter mouse[31]. PDGFR-α staining was observed in cells surrounding the upper crypt in close proximity to the epithelial basement membrane, a position consistent with pericryptal myofibroblasts. The PDGFR-α cells were, however, proposed to be distinct from α-SMA^+ myofibroblasts on the basis of co-stainings and showed differential expression of purinergic receptors, toll-like receptors (TLRs) and hedgehog signalling genes when compared to the mesenchymal compartment as a whole.

1.4.3 Amine oxidase, copper containing 3 (AOC3)

Early supporting evidence for the classification of the α-SMA+ pericryptal cells of the intestinal lamina propria as myofibroblasts came from the discovery of a mouse monoclonal antibody, PR2D3, raised against fresh samples of normal large intestine[32]. This demonstrated selective binding to the pericryptal cells and smooth muscle but not to underlying connective tissue fibroblasts. Positive staining was also observed in presumptive myofibroblasts from other tissues. The target of PR2D3 has recently been identified as amine oxidase, copper containing 3 (AOC3), a surface-bound enzyme of the semicarbazide-sensitive amine oxidase / copper-containing amine oxidase (SSAO) family[33]. The transcription factor NKX2-3 was identified as differentially expressed in the colonic $AOC3^+$ cell subset, and knockdown of this gene in myofibroblasts lead to a decrease in myofibroblast-related gene expression.

1.4.4 Prostaglandin-endoperoxide synthase 2 (PTGS2 / COX2)

Prostaglandin E_2 (PGE_2) has been identified as an important regulator of the proliferative activity of the colonic epithelium[34]. Prostaglandin-endoperoxide synthase 2 (PTGS2 / COX2) knockout mice, which are defective in PGE_2 synthesis, show profound defects in epithelial regeneration following injury that can be rescued

by exogenous PGE_2 administration[35]. PTGS2 expression in the colon is restricted to a subset of fibroblast-like cells that co-express the surface markers CD44, CD29 and CD54 within the mesenchymal compartment[35, 36]. These markers have previously been associated with mesenchymal stem cells in multiple tissues[37]. Interestingly in the presence of inflammation, these cells appear to migrate to positions adjacent to the crypt bases from deeper within the lamina propria, a process that also requires the innate signalling adaptor MYD88. It remains unclear whether these cells are a resident population or arise from bone marrow progenitors[38].

1.4.5 Forkhead Box L1 (FOXL1)

The winged-helix transcription factor Forkhead Box L1 (FOXL1) is expressed specifically in the gastrointestinal mesenchyme[39]. Antibodies to FOXL1 protein identify a small subset of α-SMA^- mesenchymal cells in close proximity to the epithelial basement membrane[40]. The absence of α-SMA staining suggests that these cells are distinct from pericryptal myofibroblasts. In the same study, the $FOXL1^+$ cells were shown to express critical morphogens associated with maintenance of the epithelial stem cell niche including gremlin 1 (GREM1), a bone morphogenetic protein (BMP) antagonist that has itself been suggested as a candidate marker of a mesenchymal stem cell population localised at the crypt-villus junction in murine ileum[41]. The distributions of GREM1 and FOXL1 positive cells from these two sources in murine ileum are similar.

1.5 Knockout models and ablation approaches

With the identification of candidate subset marker genes has come an increased appreciation of the complexity of the intestinal mesenchymal compartment. This parallels the discovery of complex networks of specialised fibroblast-like cells in other tissues, such as lymphoid tissues and the skin[42, 43]. In both the skin and the gastrointestinal tract, the gene expression profiles of fibroblasts also show variation by anatomical site, adding a further layer of diversity[44, 45]. Within the skin it has

been suggested that topographic differentiation and positional memory are directed by the Homeobox (HOX) family of transcriptional regulators[44].

While marker based approaches have been helpful in defining stromal heterogeneity, a true definition of distinct cell types would ideally include an understanding of their ontogeny, mechanistic insight into the differentiation process, and assignment of specific non-redundant biological functions. Knockout models and subset specific ablation approaches can be helpful in elucidating this last aspect, although the lack of specificity of many of the widely used stromal markers is problematic. α-SMA knockout mice are able to feed and reproduce normally although they demonstrate abnormalities in vascular tone[46]. Alternate actin isoforms were observed to substitute for α-SMA in the vascular tissues in this model and the same may be true for the gastrointestinal tract although it was not studied in detail.

Mitogen-activated protein kinase 2 (MK2) is involved in the phenotypic differentiation from fibroblast to myofibroblast *in vitro*, and a null mouse model showed reduced numbers of α-SMA positive myofibroblasts in the lung and increased susceptibility to pulmonary fibrosis[47]. Chondroitin Sulfate Proteoglycan 4 (CSPG4 / NG2) is a marker of both pericytes and pericryptal myofibroblasts in the intestine[48]. A murine knockout model shows defects in adipogenesis but no abnormal intestinal phenotype was reported[49]. Fibroblast specific protein 1 (FSP1 / S100A4) has been used for fibroblast depletion in an animal model of peritoneal fibrosis, but unfortunately FSP1 is neither specific for fibroblasts in general nor highly expressed in the pericryptal myofibroblasts of the intestine[50].

The extracellular matrix protein periostin (POSTN) is another candidate marker of the pericryptal myofibroblasts in intestinal tissues, and a murine knockout model shows reduced epithelial proliferation[51]. Selective ablation of the FOXL1^{+} pericryptal cells using a Foxl1-Cre;Rosa-iDTR mouse model produced a similar loss of epithelial proliferative activity[40].

1.6 Developmental origins

Understanding the developmental origin of the fibroblast-like cells of the intestine is a helpful step in defining a cell-type classification. The development of the gastrointestinal tract is relatively well conserved between species. A key initiating event is gastrulation, the formation of the primitive gut tube from the endoderm and the apposition of the inner leaflet of the lateral plate of the mesoderm against this structure. The endoderm, which ultimately gives rise to the epithelial compartment, becomes encircled by the expanding visceral mesoderm, and there follows a period of rapid enlargement and patterning along the dorsoventral, anterior-posterior, left-right and radial axes, leading to the differentiation of foregut, midgut and hindgut structures. Myofibroblasts are detectable as early as embryonic day (E)18.5 in mouse and by 11 weeks in human embryos[52, 53]. Cross-talk between the mesoderm and endoderm is essential in regulating this developmental trajectory. For example, endodermal hedgehog signalling induces mesodermal BMP4 and FOXF1 expression resulting in repression of the transcription factor Iroquois (IRO) and the somatic mesoderm transcriptional program[52].

In addition to the visceral mesoderm, other possible contributors to the intestinal mesenchymal compartment include the neural crest and the mesothelium. Neural crest cells migrate caudally through the embryonic gut giving rise to the enteric nervous system[54]. While at least a subset of neural crest cells possess a degree of multipotency and are able to transdifferentiate into fibroblast-like cells, lineage tracing experiments using Wnt1-Cre, Sox9-Cre and Mzp-Cre selective recombinase models showed that their contribution to the intestinal mesenchyme was limited to neural cells of the enteric nervous system[55].

Mesothelial cells form a monolayer that covers the abdominal cavity and internal organs. Lineage tracing experiments using the marker gene Wilms' tumour 1 (*Wt1*) demonstrated that the progeny of *Wt1*-expressing meothelial cells undergo epithelial-to-mesenchymal transition (EMT) and migrate from the serosal surface towards the gut lumen, where they contribute to the perivascular smooth muscle[56]. Subsequent

similar experiments performed using a murine mesothelin (*Msln*) knock-in model again identified mesothelial cells as not only the major progenitors of perivascular smooth muscle cells but also the majority of smooth muscle and fibroblast-like mesenchymal cell types in multiple tissues[55]. The model proposed to explain these observations placed the *Msln*-expressing population at the apex of a differentiation hierarchy that included *Wt1*-expressing cells as a subsequent, lineage-restricted, intermediate step. While the evidence supporting this model is convincing, the precise relationships between progenitors and differentiated fibroblast-like cell types in the intestine remain to be determined.

1.7 Replacement of mesenchymal cell types in the adult

How mesenchymal cells are replaced in adult tissues also remains uncertain. One possibility is that resident fibroblast-like cells retain the ability to expand *in situ* and replace cells that are lost. Early studies based on ^{3}H-thymidine pulse labelling in rabbit colon suggested that, analogous to the epithelial compartment, fibroblast-sheath cells are in a process of continuous migration, differentiation and apoptosis as they progress towards the luminal surface[57, 58]. Subsequent investigators, however, were unable to convincingly reproduce these findings in mouse small intestine and colon[59].

Transdifferentiation from a resident non-fibroblast cell type is another potential source. Epithelial-to-mesenchymal transition (EMT) has been demonstrated in the context of embryonic development, in fibrosis, and in cancer progression and metastasis[60]. In inflammatory bowel disease, EMT has been implicated in fistula formation in Crohn's disease (CD) and in intestinal fibrosis in animal models[61, 62]. However, while colonic epithelial cells exhibit EMT in the setting of 2,4,6-trinitrobenzene sulfonic acid (TNBS) induced colitis or transforming growth factor β (TGFβ) stimulation (characterised by a switch in marker expression from epithelial markers such as E-cadherin to mesenchymal markers such as FSP1), the mesenchymal markers displayed are non-specific[62]. Equivalence between the

cellular products of EMT and the resident fibroblast populations in the lamina propria has not been demonstrated. Furthermore, rigorous lineage tracing has failed to demonstrate either epithelial or endothelial to mesenchymal transition as major contributors to the mesenchymal expansion seen in renal or hepatic fibrosis[63–65].

The idea that cells originating in the bone marrow (BM) can migrate via the circulation, engraft in the intestinal mesenchyme, and differentiate to a fibroblast-like cells is supported by experimental evidence. Using in-situ hybridisation to detect the Y-chromosome in female mice engrafted with bone marrow from male donors and in human female recipients of allogenic bone marrow transplants, it is possible to identify BM-derived cells in the lamina propria that are morphologically compatible with intestinal fibroblasts and co-stain for operationally useful but non-specific myofibroblast markers such as α-SMA[27]. Two distinct stem cell populations exist within the bone marrow, haematopoietic stem cells (HSCs) and mesenchymal stem cells (MSCs)[66]. Both are candidates for the origin of BM-derived fibroblast-like cells in the gut. Following single-cell cloning and transplantation, GFP-labelled HSCs have been shown to repopulate the pericyte niche in kidney, cardiac and brain tissue[67]. This is likely to occur via a circulating monocyte-derived cell termed the fibrocyte, which can be identified based on surface marker expression in the vascular compartment[68]. In inflamed tissue and when exposed to TGFβ, fibrocytes adopt a myofibroblast phenotype, downregulating CD45 and CD34 and upregulating α-SMA[69]. They are thought to be involved in several pathological processes, including hypertrophic scar formation, airway remodelling in asthma and atherosclerosis[68, 70].

MSCs are cultured from bone marrow stromal cells by adhesion to plastic in 10% calf serum. They are able to support HSC growth in culture, and with addition of appropriate growth factors are able to differentiate into a spectrum of mesenchymal cell types including adipocytes, osteocytes and chondrocytes[71]. MSCs have immunomodulatory properties and are capable of migrating to sites of inflammation and injury[72]. They are currently being tested as a cellular therapy in IBD and enteric fistulae[73].

While replacement of colonic mesenchymal cells by a circulating progenitor is possible, Rinkevich and colleagues provided compelling evidence that this process is not a major contributor to mesenchymal cell populations under homeostatic conditions[55]. A surgically conjoined mouse model consisting of one donor green-fluorescent protein $(GFP)^+$ mouse and a second GFP^- animal was used to trace the fate of all cells transiting through the anastamosed circulatory system. After one year conjoined, the animals were sacrificed and the recipient tissues examined. In the intestine, all GFP^+ cells examined co-stained for the haematopoietic marker CD45. The authors go on to provide evidence that, as in development, the mesothelium is a major source of replacement for fibroblast-like cells in the adult intestine. Mesothelial labelling was performed using a tamoxifen-inducible Msln-CreER followed by whole-body irradiation and follow up for up to 4 months. At the end of this period, labelled cells made up a substantial part of the interstitial fibroblasts and smooth muscle cells of the intestine.

1.8 Stromal interaction with the epithelial compartment

Bidirectional signalling between the endoderm and mesoderm is essential for normal gut development and continues in adult life in the form of epithelial-mesenchymal cross-talk. The importance of multiple signalling pathways has been described through the use of knockout models, with essential roles defined for the hedgehog pathway, the notch pathway, the fibroblast growth factor (FGF) pathway, the bone morphogenetic protein (BMP) pathway, and the platelet-derived growth factors (PDGFs). A summary of the models and intestinal phenotypes observed is shown in Table 1.2. Expression of pathway components and pathway activation is tightly controlled across developmental time, cellular compartment and anatomical location within the intestine.

Model	Specificity	Action	Phenotype
Indian hedgehog			
Villin-Cre; $Ihh^{Fl/Fl}$	epithelium	blockade	Loss of the muscularis mucosae, reduced crypt myofibroblasts, reduced expression of extracellular matrix and basement membrane genes[74]
Villin-Hhip	epithelium	blockade	Perturbed formation of villi, mislocalisation of intestinal myofibroblasts[26]
Bapx1-Cre; R26-SmoM2	mesenchyme	activate	Expansion of the mesenchyme and of α-SMA expressing cells at E14.5[75]
Notch			
$Bapx1^{+/Cre}$; $Rbp\text{-}J^{Fl/Fl}$	mesenchyme	blockade	Shortened gut, reduced thickness of the submucosal mesenchyme, progressive loss of subepithelial fibroblasts[76]
$Bapx1^{+/Cre}$; $Rosa26^{+/Nicd}$	mesenchyme	activate	Shortened gut, reduced thickness of the submucosal mesenchyme, radial asymmetry[76]
Fibroblast Growth Factor			
$Fgf9^{-/-}$	global	blockade	Lack of a caecum, shortened small intestine, decreased mesenchymal proliferation, enhanced expression of α-SMA[77]
Platelet-derived Growth Factor			
$Pdgfa^{-/-}$, $Pdgfra^{-/-}$	global	blockade	Reduced numbers of villi, loss of pericryptal mesenchyme, enhanced expression of α-SMA[78]
Bone Morphogenetic Protein			
Villin-X-noggin	epithelium	blockade	Ectopic crypt formation, Juvenile Polyposis-like appearance[79]

Table 1.2: Genetic evidence for key epithelial-mesenchymal signalling pathways. Adapted from [12].

1.8.1 Hedgehog signalling

The hedgehog proteins, Indian hedgehog (IHH) and Sonic hedgehog (SHH), exert their signalling actions by interacting with the transmembrane receptors Patched 1 (PTCH1) and Smoothened (SMO) on target cells[80]. In the absence of a hedgehog

ligand, PTCH1 and SMO form an inactive complex. On ligand binding to PTCH1, SMO is released and recruits the Glioblastoma transcription factors (GLI1, GLI2, GLI3) to mobilise the cellular transcriptional response. The Hedgehog inhibitor HHIP also sequesters Hedgehog ligands, but unlike PTCH1 has no effect on SMO activity, acting as a negative regulator of pathway activation[26]. In the intestine, Hedgehog signalling is predominantly from the epithelial to the mesenchymal compartments, with colonic epithelial cells expressing predominantly IHH[81]. *Ihh* gene knockout in murine models is lethal, and mutations resulting in reduced activity or expression of IHH or SHH result in gross malformations of alimentary tract in humans[9]. In adults, the hedgehog pathway appears critical for the development of the mesenchymal cells of the lamina propria. Two different tissue-specific knockout models have been described, a Villin-Cre;$Ihh^{Fl/Fl}$ mouse and a Villin-Hhip mouse, in which the negative regulator is overexpressed in the epithelial compartment[26, 74]. In both cases, mislocalisation of the subepithelial myofibroblasts and progressive loss of the muscularis mucosae was observed. Early death occurred through a combination of villus atrophy, diarrhoea, malabsorption and malnutrition. Increased epithelial proliferation was also observed with $MKI67^{+}$ epithelial transit amplifying cells extending from their usual position towards the crypt bases to the upper crypt and luminal surface epithelium. The intestinal inflammation seen in these models may be due to impaired epithelial barrier function, structural disruption of the lymphatic lacteals or loss of a direct anti-inflammatory property of the hedgehog pathway. Conversely, expression of a constitutively active Smoothened isoform (SmoM2) targeted to the embryonic gut mesenchyme in a Bapx1-Cre model resulted in a marked expansion of the α-SMA^{+} pericryptal myofibroblast population[75].

1.8.2 Notch signalling

The Notch signalling pathway relies on direct cell-cell contact between cells expressing membrane-tethered Notch ligands (Delta-like 1–3, Jagged 1–2) and cells expressing Notch receptors (Notch 1–4). The study of Notch signalling in the intestine has focused on epithelial stem cells, which are dependent on this pathway

for their maintenance[82, 83]. However, structural abnormalities are also observed in animal models where Notch signalling is specifically perturbed in the intestinal mesenchyme, suggesting that the pathway also operates between compartments. Deletion of the Notch effector gene *Rbpj* in the developing gut mesenchyme (targeted using Bapx1-Cre) resulted in a loss of subepithelial fibroblasts and a shortened gut[76]. Hyper-activation of the pathway, achieved by expressing a constitutively active Notch1 intracellular domain (NICD) using the same Cre driver, produced a similar phenotype, suggesting that homeostatic Notch signalling is tightly regulated.

1.8.3 Fibroblast growth factors

The fibroblast growth factors form an important autocrine and paracrine signalling axis in development of many tissues[84]. In the canonical signalling pathway, FGF ligands bind in concert with heparin sulphate proteoglycans to transmembrane FGF receptors, leading to receptor dimerisation, activation of an intracellular receptor kinase subunit and downstream signal transduction predominantly via the mitogen-activated protein kinase (MAPK) pathway[85]. FGF9 is expressed in the intestinal epithelium in mice. *Fgf9*$^{-/-}$ animals fail to develop a caecum and demonstrate reduced mesenchymal proliferation and an altered fibroblast phenotype characterised by reduced BMP4 and enhanced α-SMA expression[77]. Similar findings are observed in *Fgf10*$^{-/-}$ animals. In explant culture, FGF9 and FGF10 induce proliferation in the mesenchymal and epithelial compartments respectively, suggesting a reciprocal signalling arrangement[86]. The mechanism by which fibroblast growth factors direct fibroblast differentiation has not been fully elucidated, but likely involves interaction with the TGFβ signalling pathway, as differential phosphorylation of TGFβ pathway intermediates (phospho-SMAD2/3) is observed in knockout animals[77]. The FGF pathway is critical for multiple aspects of organogenesis in early development, and a lack of specific genetic tools has to limit the off-target effects of mutations and avoid embryonic lethality has held back progress in elucidating its function in adult intestinal homeostasis.

1.8.4 Platelet-derived growth factors

Platelet derived growth factors are predominately homodimeric polypeptides that signal via transmembrane tyrosine kinase receptors[87]. PDGF-α signalling has a role in the development of many tissues including the neural crest, gonads, lung, and intestine. PDGF-β is implicated in haematopoeisis and vascular development. Upregulation of PDGF signalling is seen in many epithelial cancers, where it triggers stromal recruitment, and in gastrointestinal stromal tumours, where it is associated with progression[88–90]. The receptor PDGFRA has been identified as a candidate marker for a subset of pericryptal fibroblast-like cells in the intestine (see above), and together with its ligand, PDGFA, is essential for villus formation in the small intestine[31, 78]. Clusters of $PDGFRA^{+}$ cells are observed at the tip of the mesenchymal core of the developing villus. *Pdgfa*$^{-/-}$ and *Pdgfra*$^{-/-}$ mice show reduced villus formation, a paucity of mesenchymal cells and increased fibroblast α-SMA expression[78].

1.8.5 Bone morphogenetic proteins

The bone morphogenetic proteins (BMPs) are members of the TGFβ superfamily of signalling molecules. Ligand binding induces phosphorylation of the type 1 receptor subunit by the type 2 receptor subunit and release of a Mothers Against Decapentaplegic homologue (SMAD) family signalling intermediary. SMADs go on to form heterodimers which translocate to the nucleus and act as transcriptional regulators[91]. SMAD-independent signalling also occurs via the MAP kinase pathway. The inter-crypt mesenchyme is the major source of BMP4 in the intestine, and BMP signalling is intimately involved in the maintenance of epithelial homeostasis[79, 92]. The colonic crypt epithelium forms a prototypical adult stem cell compartment, with epithelial stem cells located at the crypt base undergoing infrequent, symmetrical divisions[1]. As stem-cell progeny are displaced towards the luminal surface, they transform into transit amplifying cells and proliferate rapidly before differentiating in the upper crypt, first into broad secretory and absorptive

lineages, then into all differentiated epithelial cell types before ultimately being shed into the lumen. This process occurs continuously in the adult intestine.

The hedgehog, BMP and Wnt pathways act in unison to define the proliferative and differentiating zones of the epithelial crypt 'escalator'. Hedgehog, released predominantly from differentiated enterocytes, acts on the pericryptal mesenchyme to induce BMP production[81, 93]. Mesenchymal BMPs then oppose canonical Wnt signalling, the fundamental pathway driving epithelial proliferation, by stabilising phosphatase and tensin homolog (PTEN), decreasing AKT activity and ultimately reducing nuclear β-catenin[94]. Transgenic mice expressing Noggin, which opposes BMP2, 4 and 7, show dysregulated epithelial proliferation characterised by formation of ectopic crypts and a Juvenile-Polyposis like phenotype[79, 95]. The increased epithelial proliferation observed in *Ihh* knockout animals can also be explained by this model and is supported by microarray data showing reduced global expression of BMPs 2, 4 and 5 and increased expression of BMP antagonists such as GREM1 and Chordin-like 2 (CHRDL2)[74].

The mesenchymally expressed vesicle-docking protein Syntaxin 2 (STX2, also known as epimorphin) may also contribute to this epithelial-mesenchymal feedback loop, either by promoting the release of other mesenchymal growth factors or by an as-yet-undefined direct mechanism. Epimorphin-transfected myofibroblasts show increased BMP4 secretion[96]. *Stx2*$^{-/-}$ mice are resistant to DSS colitis induction and show augmented crypt epithelial proliferation[97]. Interestingly, they also show reduced intestinal dysplasia in the azoxymethane / dextran sodium sulphate (AOM/DSS) model of inflammation-associated carcinogenesis, possibly as a result of increased epithelial turnover[98].

1.8.6 Maintenance of the epithelial stem cell niche

Epithelial stem cells reside at the crypt bases and undergo clonal expansion to replenish the epithelial compartment. In an *in-vitro* system, a single LGR5^{+} epithelial stem cell can give rise to an epithelial organoid with a full representation of differentiated epithelial cell types[99]. While this might suggest that 'stemness' is

a cell-intrinsic property, it should be noted that such systems require an exogenous Matrigel matrix and growth factor supplementation. *In vivo*, stem cell properties are maintained by the local microenvironment, the stem cell niche, to which neighbouring epithelial cells, pericryptal fibroblasts, the extracellular matrix and basement membrane, and soluble growth factors all contribute[100]. A major pathway governing stem cell identity is Wnt signalling, involving the canonical Wnts (3, 6, 9b), β-catenin and Transcription Factor 4 (TCF4)[1, 101]. Following Wnt ligand ligation, β-catenin translocates to the nucleus and acts with TCF4 to induce Wnt target genes. The importance of Wnt signalling in stem cell identity was demonstrated in cancer-derived epithelial cell lines by differential expression analysis following overexpression of a dominant-negative form of TCF4[102]. Among the downregulated genes identified were a large number of stem cell markers including *Lgr5* and *Olfm4*. Specialised secretory epithelial cells known as Paneth cells are the major source of WNT3 in the small intestine[103]. The identity of the equivalent cell subset in the colon, which lacks Paneth cells, is less clear. Co-embedding of Paneth cells and $LGR5^+$ cells in culture results in enhanced organoid formation[104]. The other critical signalling molecules required for the maintenance of the $LGR5^+$ stem cells are the R-spondin family of signalling proteins. These act via receptors including LGR5 itself to potentiate Wnt signalling[105]. A recent elegant study used single-cell RNA sequencing (RNAseq) to define the essential role of R-spondin in maintaining the stem cell population in the small bowel with overexpression of R-spondin causing a failure of differentiation, and blockade of the R-spondin signalling pathway resulting in a loss of stem cells[106].

Mesenchymal cells have been shown both to contribute to and to constrain the epithelial stem cell niche. When whole crypts are co-embedded with intestinal myofibroblasts in culture, organoids form and can be maintained over several weeks in contrast to crypt-only systems where they break down after a few days[107]. Furthermore, co-culture with myofibroblasts bypasses the requirement for the Wnt agonist R-spondin 1 in organoid culture, and the myofibroblasts

that support organoid growth in this setting show increased expression of R-spondin 2[108]. Paneth-cell derived Wnt3 is also required for organoid expansion in culture, but dispensable *in vivo*, with mesenchymal WNT2B the likely redundant source[109]. Similarly, a murine epithelial-specific *Porcn* knockout, which abolishes all epithelial Wnt secretion, has near-normal epithelial morphology[110]. Conversely, mesenchymal cells are the major source of BMPs, which antagonise Wnt signalling, and additional Wnt antagonists such as SFRP2 and DKK3[111]. They also express non-canonical Wnts (2b, 4, 5a, 5b), some of which act via β-catenin independent pathways to oppose canonical Wnt signals[112, 113]. The presence of heterogeneity among pericryptal fibroblast-like cells offers a possible reconciliation for these apparent contradictory properties and experimental evidence supports this. Expression of BMP antagonists *Noggin*, *Grem1* and *Chrdl2* is not only restricted to the mesenchyme, but also further limited to a subset of cells surrounding the crypt bases and embedded in the adjacent smooth muscle of the muscularis mucosae[92, 94].

1.9 Stromal cells in mucosal immunity

The colonic mucosal surface is continuously exposed to luminal antigens and the gut microbiota. Maintenance of immune homeostasis therefore requires a balance of pro- and anti-inflammatory processes. The mucus layer and epithelium do not present an absolute barrier, and the underlying stromal cell types are endowed with a range of cell-intrinsic responses to leaked antigenic stimuli originating in the lumen and epithelial compartments.

1.9.1 Innate sensors

The innate immune system is an ancient, evolutionarily conserved branch of immunity that confers an ability to mount a rapid protective response to conserved pathogen or damage associated molecular patterns (PAMPs). These are sensed by pattern recognition receptors (PRRs) such as the transmembrane toll-like receptors (TLRs) or the cytosolic NOD-like receptors (NLRs). PRRs are widely distributed

among cell types, including both professional immune cells, where they may prime or modulate the adaptive immune response, and non-immune cells such as those of the stroma or epithelial surfaces[114].

Colonic fibroblasts and myofibroblasts express a range of PRRs including NOD1–2 and TLR1–9[115, 116]. There is experimental evidence that these sensors are functional and contribute to the mucosal immune response, albeit primarily from *in vitro* model systems. *Nod2*$^{-/-}$ mice exhibit impaired intestinal pathogen clearance when infected with *Citrobacter rodentium* in a murine model of human enterohaemorrhagic *Escherichia coli* infection[117]. Mechanistically, the defect was localised to stromal CCL2 production, which was necessary for inflammatory monocyte recruitment. Intestinal myofibroblasts respond to *in vitro* stimulation with bacterial lipopolysaccharide (LPS), a ligand for TLR4, by activation of a number of important intracellular signalling pathways including NF-$\varkappa$B, Akt and p38[118]. This results in upregulation of both cytokines, including interleukin-1 (IL1), IL6, IL8, and granulocyte-macrophage colony stimulating factor (GMCSF), and the adhesion molecules ICAM1 and VCAM1[119]. They also show induction of PTGS2 (COX2), a critical enzyme in the synthesis of PGE_2, which is a major mediator of epithelial repair[35, 118, 120]. Exposure to Staphylococcal toxins produces a similar release of inflammatory mediators including IL8 and CCL2[121]. Fibroblasts and myofibroblasts have also been identified as the main intestinal cells expressing NLRP6, and *Nlrp6*$^{-/-}$ mice show impaired epithelial healing after injury and increased susceptibility to carcinogenesis in an azoxymethane / dextran sodium sulphate (AOM/DSS) model[122]. Triggers for innate responses may derive from neighbouring damaged cells as well as directly from pathogens. Epithelial cell disintegration is associated with the release of IL36 receptor ligands, which are sensed by IL36R^{+} subepithelial fibroblasts leading to production of chemokines, IL6 and GMCSF[123].

1.9.2 Cytokine-mediated effects

In addition to direct innate sensing, colonic stromal cells respond to a wide range of cytokines originating from other cell types. Epithelial Galectin-3 has been shown to

induce myofibroblast IL8 secretion *in vitro*[124]. A reciprocal signalling arrangement is seen in the case of IL24, which is secreted by myofibroblasts and induces goblet cell mucin secretion[125]. Activated myofibroblasts also induce epithelial C-X-C motif ligand 7 (CXCL7), involved in neutrophil recruitment[126]. Inflammatory cytokines such as IL1α and IL1β induce stromal production of chemokines, matrix metalloproteinases (MMPs) and PTGS2[127, 128]. There is also evidence of interaction with multiple T-lymphocyte subsets. The Th1 cytokine interferon-γ induces stromal upregulation of CD40 and MHC class II molecules, suggesting a role in antigen presentation[129]. Conversely, stimulation with IL1β and TGFβ induces secretion of IL11, which primes Th2 responses[130]. IL17 produced by the Th17 lymphocyte subset, induces a stromal response similar to that induced by LPS, characterised by increased IL6, IL8 and CCL2[131]. Finally, the smooth muscle cells of the lamina propria posses some further, specialised functions, with IL4 receptor-α signalling required for expulsion of helminths in murine models[132, 133].

1.9.3 Interaction with professional immune cells

The fibroblast-like cells of the lamina propria influence mucosal immunity though direct interactions with the professional immune cells they are in close proximity to within the subepithelial compartment. Bidirectional fibroblast-granulocyte interactions have been demonstrated and include stromal IL8 production in response to eosinophil-derived major basic protein (MBP) and fibroblast-mediated promotion of mast cell survival in culture via an IL6-dependent mechanism[134, 135]. The influence of stromal cells on myeloid biology is another area of interest. Through the production of growth factors such as GM-CSF and macrophage colony-stimulating factor (M-CSF), stromal cells provides a niche for myeloid differentiation that is expanded by stromal exposure to IL17 and TNFα[136, 137]. A subset of podoplanin $(PDPN)^{high}$ lamina propria fibroblasts have been shown to act in synergy with local dendritic cells to produce retinoic acid (RA), which is in turn involved in controlling T-cell activation and migration to the intestine[138]. The relevance of the stroma in maintaining a balance between pro- and anti-inflammatory pathways

is again in evidence, as stromal factors such as PGE_2 and IL10 have been shown to induce tolerogenic phenotypes in dendritic cells[139].

Intestinal myofibroblasts are a major local source of the IL1-family alarmin IL33, which is released on barrier injury and involved in mobilising Type 2 immune responses[140, 141]. In the skin, fibroblast-derived IL33 has been shown to induce an inflammatory response in eosinophils, while in animal models of synovial inflammation IL33-induced mast-cell activation is a major pathological axis[142, 143]. The precise role of IL33 in intestinal immunity has not been fully determined, with both protective and pathological roles described in model systems[144, 145]. Importantly, it has recently been shown to enhance the TGFβ-mediated differentiation of regulatory T-cells (Tregs), which maintain tolerance in the intestinal mucosa and abrogate colitis when adoptive transfer is performed in murine models of IBD[146]. IL33 maintains Treg populations within inflamed tissues and is antagonised by IL23, a key pro-inflammatory cytokine implicated in the pathogenesis of IBD.

The understanding of stromal-lymphoid interactions in the intestine is far from complete. However, important roles have been described both in humoral and cellular immunity. Immunoglobulin A (IgA) is the major immunoglobulin isotype produced within the intestinal mucosa and is transcytosed to the gut lumen. There, pathogen-specific IgA provides protection against invasive components of the microbiota through neutralisation and exclusion[147]. A combination of both stromal and microbial factors is required for the generation of multifunctional IgA^+ plasma cells that also produce TNFα and nitric oxide (NO)[148]. Adherence of stromal cells to IgM^+ B-cells has been shown to induce IgA class-switching, while stromal lymphotoxin beta receptor (LTβR)-mediated interactions enhance IgA production[149, 150].

The catalogue of stromal-T-cell interactions described in the intestinal mucosa is extensive. One key function is the maintenance of resident T-cell populations. Effector T-cell apoptosis is reduced in culture with the use of conditioned growth medium from autologous mucosal fibroblasts, an effect mediated by IL10[151].

Stromal cells also function as non-professional antigen presenting cells. They express the co-stimulatory molecules CD80 and CD86 and are able to present model antigen via HLA-DR/MHCII and induce T-cell activation both in culture and animal models[116, 152]. Stromal antigen presentation in the intestine following ablation of professional antigen presenting cells is sufficient to induce acute graft-versus-host disease in a T-cell-transfer model[153]. Whether antigen presentation results in activation or tolerance induction under homeostatic conditions is unknown, but the balance of evidence supports the latter. Colonic fibroblasts an myofibroblasts express the inhibitory molecules program death receptor (PD)-L1 and PD-L2 and suppress activated $CD4^+$ T-helper cell proliferative responses[30, 154]. Stromal TLR4 stimulation increases PD-L1 expression and reinforces this activity[155]. In the absence of inflammation, stromal cells also direct T-cell differentiation towards tolerant states. Co-culture of mucosal fibroblasts with naive T-cells results in induction of a Treg phenotype and expansion of a subpopulation of T-cells expressing cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), IL10 and TGFβ with suppressive properties[156]. This action is abrogated by addition of blocking antibodies to MHCII or the PGE_2 synthesis inhibitor indomethacin. However, there is some evidence that stromal cells may potentiate pro-inflammatory Th17 differentiation under pathological conditions through expression of IL23, as upregulation of the IL23 subunit IL23p19 with inflammatory stimuli has been documented[157].

1.9.4 Contribution to lymphoid tissues

The immune cells of the gut exist both as disseminated individual effectors within the epithelial compartment and lamina propria and in organised aggregates, known collectively as gut-associated lymphoid tissue (GALT)[3]. Stromal cells play a major role in GALT formation, maintenance and function, and the fibroblasts and myofibroblasts found within these structures show differential marker expression and functional properties compared to the equivalent cell types found elsewhere in the mucosa. The GALT can be divided into two main classes of aggregate.

The larger type, including the macroscopic Peyer's patches of the ileum and the caecal and colonic patches in the colon, contain defined B and T-cell zones, a subepithelial dome (SED) of antigen-presenting cells, and overlying follicular epithelium containing M-cells that are specialised in reverse-transcytosis of luminal antigen. They differ from lymph nodes in that they always contain B-cell germinal centres, consistent with continuous antigen exposure, and lack a capsule. The smaller GALT aggregates, which can only be detected microscopically, are termed solitary isolated lymphoid tissues (SILTs). These contain predominantly B-cells and lack clear zonation. In the mouse, there is a further subdivision of this group into cryptopatches and isolated lymphoid follicles (ILFs), which probably represent a developmental continuum[158]. Cryptopatches have not yet been identified in human tissue. Within the human colon, ILFs are most abundant in the distal colon, reaching a maximal density of 0.6 follicles/cm section[159].

The initial formation of Peyer's patches involves interaction of $CD11c^+$ dendritic cells and stromal cells expressing LTβR[160]. This results in stromal upregulation of the cytokines IL7, CCL19 and CCL21, and the adhesion molecules VCAM1, ICAM1 and MADCAM1, and recruitment of lymphoid tissue inducer (LTi) cells, a class of innate lymphocyte characterised by expression of the transcription factor RORγt[161]. Epithelial CCL20 production induces migration of dendritic cells into the subepithelial dome[162]. A recently described, specialised stomal cell, the M-cell inducer (MTi) cell, in a pericryptal location within the SED drives epithelial M-cell differentiation and CCL20 secretion via a mechanism that depends on cell-cell contact and the cytokine receptor activator of NFκB ligand (RANKL)[163]. Cryptopatches are detectable in the first 2-weeks of life in mice and also depend on an interaction between $LT\beta R^+$ stromal cells and LTi cells for their genesis[164]. In the small intestine, maturation to ILFs occurs in parallel with diversification of the microbiome and requires expression of the innate sensor NOD1 and production of CCL20 and RANKL by the stromal component[165]. A different mechanism is operational in the colon, as both CCL20 and RANKL are dispensable for colonic ILF formation[166, 167].

As in the lamina propria, a syncytial network of fibroblast-like cells, named fibroblastic reticular cells (FRCs), exists within lymph nodes and multiple FRC subtypes have been described[42]. FRC subsets can be characterised by transcriptional profiling and surface marker expression and are spatially and functionally distinct[168]. They regulate immunity and self-tolerance via a complex web of cross-talk with haematopoietic cell types[169]. Whether a similar hierarchy of specialised fibroblasts exists within the lymphoid tissues of the colon is not known. However, the fibroblasts that are associated with lymphoid aggregates in the colon show increased expression of FRC markers such as podoplanin (PDPN/gp38) and VCAM1 and expression of these markers is seen to increase with inflammation in murine models[166, 170]. Lamina propria fibroblasts in general share important functional characteristics with lymph node FRCs, such as the ability to present antigen and to induce tolerance via PD-L1 expression, raising the possibility that both compartments are fundamentally part of the same interconnected immune-lymphatic network.

1.10 Stromal contribution to disease states

1.10.1 Inflammatory bowel disease

The idiopathic inflammatory bowel diseases, Crohn's disease (CD) and ulcerative colitis (UC), are thought to occur where there is a breakdown in tolerance to the normal intestinal microbiota leading to a sustained inflammatory response in genetically predisposed individuals[171, 172]. While they are classically described as distinct pathological entities, both conditions are heterogenous and intermediate forms exist. The most effective treatments are antibodies targeting tumour necrosis factor α (TNFα), but these agents typically induce remission in only 40% of patients[173]. The incidence of IBD is rising, and there remain patients refractory to all medical therapies. The effects of repeated surgeries, prolonged immunosuppression and a markedly increased risk of developing colorectal cancer contribute to considerable morbidity and mortality in this group. In this context, there is a need to delineate the immunopathogenesis of these conditions in greater

detail with a view both to rationalising the use of existing treatments and to identification of new therapeutic targets.

While a number of inflammatory axes driving IBD inflammation have be described, the major focus has been on the roles of the epithelial compartment and conventional immune cells[174]. Fibroblast-like cells directly interact with both of these cell types and, as described above, possess the cellular machinery to sense and respond to inflammatory stimuli. Evidence of direct involvement in IBD pathogenesis is growing, and is supported by evidence from the study of other inflammatory diseases. In rheumatoid arthritis (RA), synovial fibroblasts promote the progression from acute to chronic inflammation via activation of innate signalling, elution of inflammatory cytokines and promotion of angiogenesis[175, 176]. Furthermore, there is evidence from an animal model of systemic fibrosis that fibroblasts exposed to an inflammatory insult show persistent activation over multiple passages in culture[177]. The mechanism behind this 'imprinting' may be epigenetic, as persistent epigenetic changes have been described in fibroblasts originating from inflamed joint synovium, lung, kidney and skin in animal models and patients[178–181].

The importance of fibroblast-like cells in remodelling the extracellular matrix has long been appreciated, and much of the evidence around their role in IBD relates to this function. Fibrosis, the pathological deposition of ECM components and associated connective tissue hypertrophy, is a feature of both forms of IBD. Fibrotic strictures are a common complication of Crohn's disease and frequently require surgical resection due to an absence of effective medical therapies. ECM turnover is in part regulated by the balance of matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs). Myofibroblasts in cultures established from CD strictures showed increased basal expression of TIMP1, fibronectin and collagen type 1 compared to cultures derived from healthy tissue[182, 183]. Crohn's stricture-derived myofibroblasts also exhibit oxidative stress in culture and this is exacerbated by TNFα stimulation driving production of IL6 and MMP3[184, 185]. Conversely, treatment with anti-TNF monoclonal antibodies reduces MMP3 expression in myofibroblast cultures[186].

The pathway connecting inflammatory signalling to expression of structural components and accumulation of ECM, termed 'reparative inflammation', has been documented in multiple tissues[187]. TGFβ has been identified as a central driver of this process. In an antigen-driven model of lung inflammation, epithelial TGFβ secretion induces stromal PTGS2 (COX2) driving mucus secretion, stromal proliferation and collagen deposition[188]. Variation in TGFβ isoform expression is observed in intestinal myofibroblast cultures derived from healthy bowel, ulcerative colitis and fibrotic Crohn's disease, with only the CD-associated isoforms able to induce TIMP1 expression[182, 189]. While excessive activation of stromal reparative pathways can produce fibrosis, defective barrier repair can also produce pathological inflammation. The microRNA cluster miR143/145 is expressed in enteric smooth muscle and lamina propria fibroblasts. Specific deletion using an Myh11- or Twist2-Cre results in de-repression of stromal insulin-like growth factor binding protein 5 (IGFBP5), impaired IGF signalling, and increased susceptibility to colitis[190].

Aside from their effects on the ECM, there is evidence that stromal cells contribute to an independent inflammatory axis in IBD. The murine TNFΔARE model of Crohn's disease is characterised by epithelial overproduction of TNFα resulting in an ileitis phenotype that can be abolished by inhibition of TNF receptor signalling[191]. Selective expression of TNF receptor 1 (TNFR1) in the mesenchymal compartment using a ColVI-Cre is sufficient to fully reconstitute the ileal pathology[192]. Mechanistically, increased expression of ICAM1 is seen in α-SMA$^+$, CD90$^+$ pericryptal fibroblasts prior to inflammatory cell recruitment[193]. Unresolved stromal activation may in part explain the lack of efficacy of anti-TNF therapy. The stromal cytokine oncostatin-M (OSM) is highly upregulated in intestinal biopsies from IBD patients and its expression level shows a strong positive correlation with the risk of anti-TNF treatment failure[4]. The OSM receptor (OSMR) is expressed on lamina propria myofibroblasts and ligand ligation induces upregulation of ICAM1 and IL6. Anti-OSM blocking antibodies attenuate disease in the TNF-resistant Hh+αIL10R murine colitis model.

Whether the stromal pathways subverted in IBD are active in all fibroblast-like cells in the mucosa or specific to particular fibroblast subtypes has not been determined, but stromal heterogeneity is again apparent. The pericryptal fibroblast sheath, identified by staining for the markers α-SMA and tenascin C (TNC), is significantly attenuated in the presence of inflammation[194]. *Ex vivo* stimulation with IFNγ and TNFα supports the hypothesis that this occurs by apoptosis of the pericryptal cells rather than a change in marker expression. Other functionally important genes, such as PTGS2, show expression in only a relatively small subset of stromal cells occupying a restricted anatomical niche[35]. Dissecting the precise stromal contribution to these clinically important disease states will be facilitated by a more detailed understanding of stromal diversity.

1.10.2 Colorectal cancer

Colorectal cancer (CRC) is the third-most common form of cancer globally, with an incidence of 1.2 million cases per year[195]. Inflammatory bowel disease is a major risk factor, and the pathological progression of colitis-associated cancer differs from sporadic forms of the disease[196]. The tissue microenvironment in the intestinal mucosa, shaped in part by its stromal constituents, is an important contributor to cancer initiation, progression and metastasis[197]. Cancers themselves contain a significant stromal component, and within this the cancer-associated fibroblasts (CAFs, defined as fibroblastic, non-neoplastic, non-vascular, non-epithelial, non-inflammatory cells) have been the focus of extensive study due to their cancer-promoting properties, which include maintenance of cancer stem cells, prevention of immunosurveillance, promotion of angiogenesis, and ECM remodelling[198]. As in the healthy epithelium, considerable heterogeneity is evident in CAFs, with PDGFRA, FSP1 and FAP all having been used to identify fibroblast subsets with distinct functional properties[5, 199].

Tissue-resident fibroblast populations are the major source of CAFs, which undergo a differentiation program largely mediated by TGFβ signalling[200]. Fibroblasts isolated from normal tissue suppress cancer cell proliferation, so this

transformation has important functional consequences and is associated with worse cancer outcomes[201–203]. In addition to TGFβ, persistent CAF-specific patterns of gene expression can be induced by inflammatory cytokines, such as IL1 and TNFα, which may contribute to the increased cancer risk associated with IBD[127]. CAFs show increased expression of multiple factors including IL6, CCL2 and CCL8 that are associated with increased tumour cell proliferation, migration and adhesion[204]. They also have immunomodulatory properties. CCL2 has been shown to polarise macrophages to a tolerogenic M2 phenotype, and coexpression of CAF and M2 macrophage markers in tumours is associated with poor prognosis[205, 206]. Tumour cell-derived CCL21 induces the formation of a stromal fibroblastic network that supports Treg differentiation analogous to that seen in lymph nodes[207].

CAF activation is dependent on NF-κB signalling. A mesenchymal-targeted blockade of the inhibitor of kappa B kinase beta (IKKβ) with ColVI-Cre leads to reduced NF-κB signalling, attenuated inflammatory cytokine expression and reduced tumour burden in the AOM/DSS model of colitis induced cancer[208]. In contrast, however, deletion of IKKβ using an inducible Col1a2-CreER led to only modest reductions in inflammatory cytokines, increased TGFβ and hepatocyte growth factor (HGF) expression and increased tumour progression in the same AOM/DSS model[209]. These contradictory findings highlight the need to better define the specificity of the current Cre models, as the same pathway may have opposite effects in different fibroblast subtypes or developmental stages.

Stromal-derived PGE_2 and HGF have been identified as particularly important promoters of carcinogenesis. In the case of HGF, this may operate via potentiation of Wnt signalling in cancer stem cells, which express the HGF receptor MET and are found in close proximity to CAFs in colorectal tumours[210]. The serine/threonine kinase TPL2 (MAP3K8) is a shared upstream regulator of both pathways. TPL2 activation leads to upregulation of PTGS2 (COX2) and increased PGE_2 synthesis[34]. Knockout of Tpl2 using a fibroblast-specific ColVI-Cre results in reduced levels of HGF and reduced tumour burden in a cancer associated colitis model[211]. IGF2BP1 may act as an intermediate in HGF regulation, as its expression is

also reduced in the ColVI-Cre;Tpl2$^{Fl/Fl}$ animals and a similar HGF^{low}, tumour-suppressive phenotype is observed in a Twist2-Cre;Igf2bp1$^{Fl/Fl}$ model[212]. TGFβ and epimorphin have also been shown to regulate HGF expression in stromal-targeted selective knockout models[213, 214].

The cancer-associated stroma is essential to tumour development and an attractive target for therapeutic manipulation, particularly as CAFs are non-transformed genetically stable cells and therefore inherently less likely to develop resistance[215]. Both PTGS2 and HGF inhibitors have been evaluated in this setting[216, 217].

1.11 Single-cell RNA sequencing

RNA sequencing (RNAseq) offers a global perspective of transcriptional activity within a sample and complements marker-based approaches in unravelling the heterogeneity within the stromal compartment. Both RNAseq and gene-expression microarrays have been successfully applied to uncover important aspects of stromal biology, such as the roles of BMPs and OSM in regulating epithelial proliferation and inflammatory responses respectively[4, 92]. When applied to complex tissues samples, however, bulk sequencing approaches cannot define the contribution of individual cell types and, in failing to capture the relationship of component parts, can lead to inaccurate conclusions about the nature of a biological process[218, 219]. This is particularly problematic in the case of stromal cells, which make up a minority of the cellular population of the intestinal mucosa and are less transcriptionally active that other constituents such as epithelial and immune cells.

The relatively recent development of single-cell RNA-seq circumvents this limitation by first dissociating a heterogenous sample into single cells, then capturing the RNA from each in a separate reaction. The technique relies on appending molecular barcodes to RNA molecules so that they can be traced back to their cell of origin and on multiple amplification steps, either by quantitative reverse transcription polymerase chain reaction (RT-qPCR) or *in vitro* transcription (IVT), that amplify tiny amounts of starting material while maintaining relative levels of transcript abundance[220].

While single-cell RNAseq data offers a unique insight into the global activities of individual cells, analysis can be challenging due to increased dispersion in the gene counts generated relative to bulk RNAseq. Key contributors to this are drop-out events, transcriptional bursting and systematic biases introduced by the multiple rounds of amplification. Drop out events are characterised by a failure to capture an mRNA molecule that was present in the original cell due to the inefficiency of reverse transcriptase. Transcriptional bursting is a fundamental property of genes in which transcription occurs rapidly in discrete bursts, as opposed to over a continuum of rates[221]. Transcript abundance in influenced by burst frequency and duration, mRNA decay processes and the buffering properties of the nuclear envelope. In bulk RNAseq, the effect of transcriptional bursting on transcript counts is reduced by averaging across thousands or millions of cells but this advantage is lost when working at single cell level. Finally, small differences in PCR efficiency resulting, for example, from the presence of secondary structure, associated proteins, or variation in transcript length can all give rise to bias in final gene counts when extensive amplification is employed. Despite these limitations, single cell RNAseq has been used successfully to refine the classification of cells in the murine nervous system, spleen and lung, as well as intestinal epithelial cells and immune cells such as $CD4^+$ T cells and innate lymphoid cells[222–228].

1.12 Discussion

There has been a profound and dramatic shift in the perception of stromal cells from passive supports to active participants in maintaining homeostasis at the epithelial barrier surfaces of the intestine. As well as providing a structural framework and shaping the ECM, there is convincing evidence of a dynamic cross-talk with the epithelial compartment. They also appear capable of acting as local immune rheostats, sensing the local environment though an array of innate sensors and able to induce either tolerogenic or inflammatory responses from mucosal immune system. They are relatively abundant in comparison to professional antigen presenting cells and occupy highly relevant tissue niches directly below the epithelial barrier. Unlike

epithelial cells, they do not turn over rapidly and changes in their activation can persist over multiple passages in culture possibly mediated by epigenetic mechanisms. As such, they are a potential site of immune imprinting or memory, which is of great relevance in chronic relapsing conditions such as IBD.

As knowledge of the repertoire of stromal cell functions has expanded, so has evidence of heterogeneity. Both ultrastructural in immunohistochemical methods can be used to define anatomically distinct subsets of fibroblasts and myofibroblasts within the lamina propria and adjacent tissues. Numerous candidate markers for functionally distinct stromal subsets have been proposed. The existence of specialised fibroblast subsets seems probable given the specialisation seen in the haematopoietic-derived and epithelial constituents of the mucosa and the essential contribution of stromal cells to highly distinctive tissue niches such those found in mucosal lymphoid aggregates and surrounding epithelial stem cells.

There remains much work to be done in this area. The performance characteristics of widely used markers such as α-SMA are suboptimal and the specificities of murine Cre recombinase systems, such as those based on collagen types 1 and 6, have not been fully determined. In addition, functional assays based on *ex vivo* maintenance of primary stromal cells are predicated on the assumption that representative cross sections of endogenous stromal subtypes are amenable to capture and expansion by such methods and that they retain their native characteristics.

The transcriptome constitutes an essential piece of cell identity, and single-cell RNAseq has not yet been applied to classify and compare intestinal stromal cells in health and inflammation. The focus of this project is to construct a transcriptional classification of stromal cell types within the colonic lamina propria using single cell RNAseq. In doing so, I will address important unanswered questions in the field including the diversity of stromal cell types, their functional specialisation and ontogeny, and the specificity of existing genetic markers. The data generated will also provide a basis for the identification and validation of novel markers with improved performance characteristics.

Chapter 2

Materials and Methods

Contents

2.1 Cell preparation — Fluidigm C1 38
2.1.1 Human sample collection 38
2.1.2 Isolation of human colonic epithelial cells 38
2.1.3 Isolation of human colonic stromal cells 39
2.1.4 Cell capture, library preparation and sequencing 39
2.1.5 Modifications for multiplexed experiments 40
2.2 Bioinformatic approach — Fluidigm C1 40
2.2.1 Alignment and transcriptome assembly 40
2.2.2 Cell and gene quality control 41
2.2.3 Differential expression analysis 42
2.2.4 Library size normalisation 42
2.2.5 Technical batch effect adjustment 44
2.2.6 Weighted gene co-expression network analysis 44
2.2.7 Cell clustering 45
2.2.8 Marker gene identification and ontology enrichment . . 46
2.2.9 Module preservation 47
2.2.10 Reduced dimensionality representations 47
2.3 Cell preparation — 10X genomics 47
2.3.1 Dextran sodium sulphate (DSS) challenge 47
2.3.2 Isolation of murine colonic stromal cells 48
2.3.3 Single cell capture, library preparation and sequencing . 49
2.4 Bioinformatic approach — 10X genomics 50
2.4.1 Cellranger pipeline 50
2.4.2 Identification of cellular barcodes 50
2.4.3 Initial clustering and QC 50
2.4.4 Cell cycle annotation, normalisation and final clustering 51
2.4.5 Further Seurat analysis functions 51
2.4.6 Diffusion maps and diffusion pseudotime 52
2.5 Statistical methods 52
2.6 Common validation methods 53
2.6.1 Flow cytometry 53

2.6.2 Immunohistochemistry (IHC) 54
2.6.3 Single molecule in-situ hybridisation (smISH) 55
2.6.4 qPCR 55
2.7 Key Resources Table 56

2.1 Cell preparation — Fluidigm C1

2.1.1 Human sample collection

Colonic biopsy samples were collected, following written informed consent, from patients attending for clinically indicated endoscopy procedures at Oxford University Hospitals NHS Foundation Trust. Inclusion criteria were male or female adults aged 18–90 years, mental capacity to give informed consent for study participation, and proficient in understanding written and verbal English. Exclusion criteria were the presence of a defined genetic syndrome predisposing to colorectal cancer (Hereditary Non-Polyposis Colorectal Cancer (HNPCC), Familial Adenomatous Polyposis (FAP), Peutz-Jeghers syndrome, Juvenile polyposis, MUTYH associated polyposis, serrated polyposis), a family history of colorectal cancer defined as high or moderate risk, any contraindication to endoscopic forceps biopsy, or significant co-morbidity. Ethical approval for the study was obtained from NRES Committee Yorkshire & The Humber — Leeds West (REC reference: 14/YH/1116).

2.1.2 Isolation of human colonic epithelial cells

A previously described method was used with some modifications[229]. Up to eight endoscopic forceps biopsies were collected in Dulbecco Modified Eagle Medium (DMEM) supplemented with 100 U/ml penicillin, 100 µg/ml streptomycin and 10 mM HEPES on ice, and washed in the same media supplemented with 1 mM EDTA and 1 mM DTT. Loosely attached cells were removed by incubation in the presence of 1 mM EDTA for 20 minutes at 37 °C with vigorous shaking every 10 minutes. Tissue fragments were then transferred to a fresh tube and incubated in the presence of 1 mM EDTA in 10 minute cycles, with vigorous shaking to detach intact crypt units at the end of each cycle and harvesting of the crypt-containing

supernatant, until no further crypts could be detached (typically 3–4 cycles). The resultant supernatants were pooled and intact crypts separated from single cells and debris by centrifugation at 150 G for 4 minutes. The crypt pellet was re-suspended in TrypLE Express (Thermo Fisher) supplemented with 50 µg/mL DNAse and incubated at 37C for 60 minutes, vortexing for 15 seconds every 15 minutes. A pellet of dissociated single cells was obtained by passing the resultant suspension through a 35 µm cell strainer and centrifugation of the filtrate at 500 G for 5 minutes.

2.1.3 Isolation of human colonic stromal cells

Endoscopic forceps biopsies were collected as before and incubated for 5 minutes at room temperature in ACK lysing buffer (Thermo Fisher). After washing with PBS, the biopsies were dissociated using a human Umbilical Cord Dissociation Kit (Miltenyi Biotec) according to the manufacturer's protocol for fresh tissue with some modifications. Incubation time was reduced to 2 hours and mechanical dissociation was achieved by passing the suspension 5–10 times through an 18 gauge drawing up needle every 60 minutes. The resultant cell suspension was passed through a 70 µm cell strainer and centrifuged at 500 G for 5 minutes. Depletion of non-stromal cell types was achieved using MACS LD separation columns (Miltenyi Biotec) following the manufacturer's protocol with some modifications. The cell re-suspension buffer was substituted for HBSS with added penicillin 100 U/mL, streptomycin 100 µg/mL, HEPES 10 mM, EDTA 1 mM, and BSA 0.5% weight/volume. CD326 (EpCAM), CD45 and CD235a conjugated micro-beads (Miltenyi Biotec) were added to the cell suspension, and the mixture was incubated at 4 °C for 15 minutes. Cells were then washed and loaded onto an the column via a 35 µm pre-separation filter. The flow-through fraction was collected, and single cells pelleted by centrifugation at 500 G for 8 minutes.

2.1.4 Cell capture, library preparation and sequencing

Isolated single cells were re-suspended at a density of 700 live cells/µl in DMEM with 5% foetal calf serum (FCS). Cells were stained for DNA content and viability by supplementing the re-suspension buffer with NucBlue live cell stain (Life technologies,

following the manufacturer's protocol) and propidium iodide 10 µg/ml final. Cells were captured on the C1 system (Fluidigm) and processed using the SMARTer chemistry (Clontech) according to the Fluidigm protocol. External RNA Controls Consortium (ERCC) RNA spike-in mix was added to the lysis buffer 1:1000 for epithelial cells and 1:4000 for mesenchymal cells. The spike-in concentration varied between cell types reflecting both differences in cell size and endogenous mRNA content and iterative refinement to the experimental protocol in light of the measured proportion of reads mapping to ERCC transcripts in initial sequencing runs. C1 integrated fluidic circuits (IFCs) were imaged prior to cell lysis to identify sites occupied by single viable cells for downstream analyses. cDNA samples were selected after analysing the cell images and prepared for sequencing using the Nextera XT library prep kit (Illumina) with in-house primers[230]. Libraries were sequenced using either Illumina HiSeq2500 100 bp paired-end sequencing or Illumina HiSeq4000 75 bp paired-end sequencing.

2.1.5 Modifications for multiplexed experiments

Biopsies from inflamed and non-inflamed bowel regions were collected separately from the same donor, and single cell isolation was performed for both sets of biopsies in parallel. The resulting single cell suspensions were counter-stained with either CellTracker Green CMFDA or CellTracker Orange CMRA in addition to nuclear and viability staining with NucBlue live cell stain and Live/Dead fixable Far Red cell stain (all Life technologies). Counter-stained cell suspensions were mixed 1:1 immediately prior to loading onto the C1 IFC. Subsequent imaging and downstream processing was performed as above.

2.2 Bioinformatic approach — Fluidigm C1

2.2.1 Alignment and transcriptome assembly

Reads were demultiplexed using the index barcodes to generate single cell libraries. Adaptor sequences (IlluminaP7, Nextera, SMARTer) were trimmed using the Skewer package[231], then aligned to an Ensembl GRCh38 human genome index with

HISAT[232]. Sorting and duplicate read detection was performed using Picard Tools (`http://broadinstitute.github.io/picard/`) prior to summarisation with featureCounts using Ensembl Release 84 (March 2016) transcriptome annotation[233, 234]. Reads flagged as optical duplicates (originating from adjacent tiles on the same slide), multi-mapping and ambiguously mapping reads were discarded. Reads flagged as potential PCR duplicates (having the same alignment start and end positions) were retained, as this method of duplicate flagging can over-estimate the true duplication rate and introduce bias especially in the presence of abundant transcripts with short exons[235]. The effect of PCR duplicate removal on the final gene counts was small (Appendix A, Figure A.3)

2.2.2 Cell and gene quality control

Raw sequence QC reports were generated with FastQC using default settings (`http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`). These and other QC outputs generated by the pipeline tools above were collated using multiQC[236], imported together with gene count tables into R, and assembled into an SCESet object using the package scater[237]. Single cell libraries were binned into 'pass', 'warn', or 'fail' categories for each QC output using thresholds determined by manual inspection of the distributions with reference to bulk and empty capture site controls where appropriate.

No single metric can be used as a proxy for sample quality, and aggregate performance over a number of domains will provide a more robust approximation. Among the most useful measures are sequencing depth, which in this context refers to the total number of reads that align to genomic features, library complexity (the number of unique genes detected), and measures that capture the degree of sequencing saturation that has been achieved. These outputs and the thresholds used are shown in Appendix A, Figure A.4.

The ERCC spike-in RNA mixture, which contains 92 different RNA sequences over a wide range of concentrations, is present at the same overall concentration in each capture site and therefore provides purely technical insight into the efficiency

of the reverse transcription, amplification and sequencing steps. The number of unique spike-in molecules detected can be used as a measure of sensitivity, while the percentage of total counts derived from ERCC molecules can be used to approximate the relative amounts of cellular RNA captured. Data and QC thresholds used are shown in Appendix A, Figure A.5.

Another potentially useful measure is the percentage of counts derived from genes on the mitochondrial chromosome (Appendix A, Figure A.6). While mRNA molecules in the nuclear and cytoplasmic compartments may be degraded or lost in the event of cellular stress, early apoptosis or cell damage, mRNA molecules generated within the mitochondrial syncytium are relatively protected and therefore make up a greater part of the library in low quality cells.

An overall QC score was calculated and cells failing any single metric or generating 2 or more warnings were discarded (Appendix A, Figure A.7). Following cell QC, gene QC was performed in two stages. Genes that were either undetected or not detected in all three donors were removed first, then genes that were not detected at a level of at least 5 counts in at least 2 cells that passed the cell QC process (Appendix A, Figure A.8).

2.2.3 Differential expression analysis

Differential expression analysis was performed using the R packages 'DESeq2' and 'scde'[238, 239]. Raw read counts were used as input. Input matrices were filtered to remove non-protein coding features, features on the Y chromosome and low expressed features (detected in < 2 batches or with a mean expression level < 2 reads/cell) prior to running differential expression testing. A false-discovery rate threshold of 5% was used to define differentially expressed gene sets.

2.2.4 Library size normalisation

Having excluded low-quality cells, normalisation was employed to reduce the remaining cell-cell differences in library size resulting from variation in capture efficiency, sequencing depth and other technical confounders. The simplest form of normalisation (scaling all libraries to uniform size, the basis of measures such

as counts-per-million) is not sufficient in the context of a heterogenous population of cells as it may introduce spurious differences between cells in the presence of differentially expressed genes. Neither is there any validated set of invariant 'housekeeping' genes that can be used as a basis for normalising single cell data. Many accepted housekeeping genes from bulk approaches show significant biological variation at single cell level between cell types or cell-cycle stages[240]. Normalisation to equalise the ERCC counts can be employed, but relies on the assumption that technical factors affect the spike-ins in the same way as they do endogenous genes. This is unlikely as spike-ins are not subject to some sources of technical variation such as the efficiency of cell lysis. Spike-ins also have shorter poly-A tails, on average, than mammalian mRNA molecules and are not associated with bound proteins or secondary structure that may influence capture and amplification efficiency[241].

Approaches have been developed for normalising RNAseq data in the presence of differentially expressed genes. Widely used examples include trimmed mean of M values (TMM) normalisation and the relative log expression method first introduced with the package DESeq. These rely on calculation of gene expression ratios between samples with the assumption that the majority of genes are not differentially expressed. Counts are then scaled such that on average there is no fold difference in expression for most genes. In both approaches, however, the sparsity of single-cell data is problematic, with the relatively high proportion of zero expression values making scale factors either impossible to estimate or inaccurate. The recently published 'scran' package provides a single-cell specific normalisation method that addresses this problem[240]. Individual cells are first randomly assigned to pools and their counts summed. Size factors can then be calculated for pools of cells against an average reference pool. Multiple permutations are performed to construct a linear system from which the cell-based size factors can be deconvolved. The results of using this method in the epithelial and mesenchymal datasets are shown in Appendix A, Figure A.10. Measures of normalisation efficacy including the median relative log expression (RLE) value and dispersion around zero were reduced relative to naive cpm normalisation.

2.2.5 Technical batch effect adjustment

The cell capture, cDNA synthesis, amplification and library preparation steps were performed in batches and are potential sources of non-biological variability (technical batch effects). Library size normalisation alone does not specifically address this problem and batch effects can clearly be seen in the Fluidigm C1 datasets following library size normalisation (Appendix A, Figure A.11). While a number of methods exist for removing batch effects from replicate samples, none is applicable to this experimental design where biological differences between donors and variability arising from sampling small cell numbers from heterogenous starting populations confound the technical differences between batches.

In the absence of an optimal solution, adjustment for batch effect was performed by first identifying control genes *in silico*. A mean-variance trend was fitted to the expression values and used as a basis for decomposition of each gene's total variance into biological and technical components (Appendix A, Figure A.12). Genes with total variance below the fitted value were considered to show only technical variability and used as controls for correction of technical batch effect using the 'RUVg' method from the R package 'RUVseq'[242]. This was effective in reducing the inter-batch differences in median cell RLE values (Appendix A, Figure A.11). Decomposition of variance into technical and biological components was then repeated, and genes with total variance above the fitted value (biologically variable genes) were used in downstream analyses. Variance-stabilised expression values were employed throughout these steps, calculated from normalised read counts by log2 transformation with an offset of 1.

2.2.6 Weighted gene co-expression network analysis

Gene co-expression network construction was performed in R using the package 'WGCNA'[243, 244]. An expression matrix of biologically variable genes and cells passing QC was converted to a pairwise gene adjacency matrix by Pearson correlation. Negative correlations were set to zero (a 'signed-hybrid' network). This

was, in turn, converted to a dissimilarity matrix by subtraction from 1 and a soft-thresholding power of 4 was applied. As a further noise-filtering step, an expression level threshold was set that represented 'high confidence' detection of a gene, and pairwise connectivities based on few cells with 'high confidence' detection of both genes were removed from the network prior to clustering and module detection. The 'high confidence' thresholds used were between 5 and 8 normalised counts while the minimum cell number was 3 or 4 depending on the dataset. Genes remaining in the filtered dissimilarity matrix were hierarchically clustered by average linkage, and the resulting clustering tree was cut to find co-expression modules using the 'cutreedynamic' function with a deep split setting of 3 and a minimum module size of between 30 and 50 genes. Module eigengenes (the first principal component of an expression matrix of all module genes) were calculated to provide a summary of total module gene expression within individual cells.

2.2.7 Cell clustering

Following module detection, a matrix comprising expression values for all genes assigned to modules was used as a basis for cell clustering by non-negative matrix factorisation using the R package 'NMF'[245]. Briefly, the method aims to factor an initial matrix of non-negative expression values, A (N genes, M cells, dimensions $N \times M$), into a coefficient matrix W (dimensions $N \times k$) and a basis matrix H (dimensions $k \times M$) such that $A \approx W \times H$. When applied to gene expression data, the rank k can be considered as a small number of metagenes, with the coefficient matrix defining the contribution of individual genes to each metagene, and the basis matrix defining cells as positive linear combinations of these metagenes and providing a natural basis for clustering[246]. The method is non-deterministic, and 50 randomly initiated runs were performed to generate a consensus clustering matrix ($M \times M$) with values between 1 (cells always in the same cluster) and 0 (cells never in the same cluster). The choice of rank k is critical to the clustering result, and a number of measures have been suggested to determine the optimal value. These include the cophenetic correlation (the correlation between cell-cell distances

in the consensus clustering matrix and the induced distances in the hierarchical clustering tree) and the variation in the residual sum of squares (RSS) between the data matrix A and the estimated matrix WH[246, 247]. We surveyed a range of possible rank values, calculating cophenetic correlation, average silhouette width, factor dispersion, and RSS, seeking the first local maximum in the first three measures and the inflection point in the last (Appendix A, Figure A.13). The same metrics were calculated for randomised data by independent permutation of the gene expression values as a control. In the epithelial dataset, the calculated metrics suggest an optimal rank of 4. There was, however, a clear division within one cluster at rank 4 marked by expression of G2/M versus G1/S related cell cycle genes. A rank of 5 was selected to capture this. The other three clusters were unchanged. For stromal cells, the optimal rank values of 4 and 5 were used for the healthy and UC datasets respectively. In all datasets, there was little overlap in expression of the detected gene modules and cell clusters were assigned the colour identifier of the gene module with the highest median eigengene expression within that group (Appendix A, Figure A.14).

2.2.8 Marker gene identification and ontology enrichment

Marker genes are expressed highly in one cell cluster and can be used to distinguish that cluster from all other clusters. For each gene, mean expression levels were calculated in each cell cluster, and a binary classifier was constructed to test the ability of that gene to distinguish cells in the cluster with the highest mean expression from the remaining cells. This was quantified in terms of the area under the ROC curve[248]. The top 200 genes for each cluster, ranked in this way, were selected as candidate marker genes. Biological process ontology enrichment was performed both on detected gene modules and cluster marker genes using a hyper-geometric test with correction for multiple testing using the Benjamini & Hochberg (BH) method with both p-value threshold and false discovery rate set to 0.05. Dot and network plots were produced with the package ‘clusterProfiler’ in R[249].

2.2.9 Module preservation

Module preservation between datasets was performed both by a cross-tabulation approach, calculating the numbers of overlapping genes and the significance level using a Fisher's exact test, and by calculation of a 'Z-summary' score, a composite of three connectivity-based and three density-based network module preservation statistics[250]. For the latter approach, 30 permutations were performed and a 'gold' module of 100 randomly selected genes was included as an additional control.

2.2.10 Reduced dimensionality representations

Input matrices that included all biologically variable genes were used in producing reduced dimensionality representations. Principle component analysis (PCA) was performed using the function 'prcomp' in R. T-distributed stochastic neighbour embedding (t-SNE) was performed using the package 'Rtsne' with an initial PCA step[251, 252]. Ten randomly initiated t-SNE runs were performed and the solution with the lowest Kullback-Leibler divergence was selected for visualisation.

2.3 Cell preparation — 10X genomics

2.3.1 Dextran sodium sulphate (DSS) challenge

Murine tissue was generated in a DSS challenge model conducted by Epistem® Ltd. All procedures were certified according to the UK Home Office Animals (Scientific Procedures) Act 1986. C57BL/6 (*Helicobacter pylori*-free, murine norovirus-free) male mice (Envigo Laboratories, UK) were used in the study. Animals were 8–10 weeks old on supply and used at 10–12 weeks of age. All mice were held in individually ventilated cages (IVCs) in an SPF (Specific Pathogen Free) barrier unit. The animals were identified by numbered cages and by ear punches.

A total of 10 mice were randomised into two treatment groups of five mice each. One group received no treatment and the other received 1.75% DSS (36–50 kDa MW, MP Biomedicals, lot #M9147) in their drinking water, from study day 0 until mice were euthanised on the morning of study day 7. DSS was made fresh daily.

Animal well-being was monitored once daily from day 0 until the end of the study. All mice were weighed and assessed for stool consistency and the presence of overt blood in the stool or around the anus according to the criteria in Table 2.1. The scores for the criteria listed in Table 2.1 were summed in order to calculate a disease activity index (DAI) for each mouse, on the final study day.

Score	**Weight loss** (% day 0 weight)	**Stool consistency**	**Overt blood** (in stool/around anus)
0	<1%	Normal	None
1	⩾1% <5%	Slight diarrhoea	Slight
2	⩾5% <10%	Moderate diarrhoea	Moderate
3	⩾10% <15%	Heavy diarrhoea	Severe
4	⩾15%		

Table 2.1: Scoring criteria for murine DSS challenge in-life disease parameters

Mice were euthanised by cervical dislocation on study day 7. The large intestine was removed and flushed with ice-cold Dulbecco's (Ca^{2+} & Mg^{2+}-free) phosphate buffered saline (D-PBS, Sigma, UK) containing 100 U/mL penicillin and 100 µg/mL streptomycin (Sigma, UK). The length and wet weight of the large bowel were recorded prior to transferring the bowel to a 50 mL centrifuge tube, filled with RPMI 1640 medium with Glutamax™ (Thermo Fisher, UK) and 100 U/mL penicillin and 100 µg/mL streptomycin. Large bowel samples were maintained at 2–6 °C and processed the same day.

2.3.2 Isolation of murine colonic stromal cells

A previously published protocol was utilised with some modifications[253]. Colons were processed individually in parallel. Caecal pouches, mesenterium and fat were first detached and discarded. The remaining colon was opened longitudinally and cut into 1 cm fragments. These were incubated at 37 °C in RPMI with 0.1% BSA and 5 mM EDTA with horizontal shaking for 40 minutes to detach epithelial crypts. The crypt containing supernatant was discarded. Fresh RPMI with 0.1 % BSA and

5 mM EDTA was added and a further 15 minute incubation at 37 °C with horizontal shaking was performed to further deplete the epithelium. The tissue fragments were then washed and incubated in RPMI with added FCS (10%), HEPES (15 mM) and Collagenase VIII (100 U/mL, Sigma Aldrich) for 60 minutes at 37 °C with horizontal shaking. The resulting supernatant was passed through a 70 µm strainer and single cells were pelleted by centrifugation at 500 G for 8 minutes. Percoll gradient centrifugation was performed to remove non-cellular debris. Physiological 100% Percoll was made by combining 9 parts Percoll with 1 part 10X PBS. The cell pellets were resuspended in RPMI with 30% Percoll (GE Healthcare) and the resulting suspension layered over PBS with 70% Percoll in a 15 mL Falcon. Centrifugation at 900 G for 20 minutes (4 °C) was performed and the 30% / 70% interface layer was collected. MACS depletion of epithelial and haematopoietic cells was performed as for human stromal cells using in this case antibody-conjugated microbeads to murine EpCAM and CD45 (Miltenyi Biotec).

2.3.3 Single cell capture, library preparation and sequencing

Single cell suspensions were captured using the 10X Genomics® Chromium™ Single Cell 3′ Solution according to the manufacturers protocol. Cells from 3 DSS and 3 control mice were resuspended separately in PBS with 0.04% BSA at a density of 500 cells per µL. 5% murine RAW 264.7 macrophages stably transduced with a lentiviral expression construct containing an *Aspergillus fumigatus* blasticidin resistance gene and a monovalent citrine (mCitrine) yellow fluorescent protein (YFP) gene were spiked into the primary cell suspensions immediately prior to the barcoding reactions, which were performed simultaneously in separate channels on a single Chromium Single Cell A Chip using the v2 Reagent Kit (10X Genomics). 3,500 cells were loaded into each reaction, to give a predicted capture of 2,000 cells. Reverse transcription, PCR amplification, cDNA fragmentation, size selection and library indexing (Nextera XT Index Kit, Illumina) were performed in parallel for the six samples, prior to qPCR quantification and sample pooling to generate a single multiplexed library that was sequenced over 3 Illumina HiSeq 4000 lanes.

2.4 Bioinformatic approach — 10X genomics

2.4.1 Cellranger pipeline

The Cell Ranger software suite was obtained from 10X Genomics (`https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest`). Raw sequencing data was first demultiplexed using the Illumina indices to generate separate read files for each of the 6 samples. Alignment and transcript quantification was performed with the standard Cell Ranger 'count' script against a custom genome reference incorporating the University of California Santa Cruz (UCSC) mm10 murine genome assembly and the lentiviral construct present in the control cells. The 3 replicate control and DSS samples were then aggregated using the Cell Ranger 'aggr' script with the default normalisation step (by down sampling) disabled. The raw (unfiltered) count matrices were imported into R for further processing.

2.4.2 Identification of cellular barcodes

Raw expression matrices output from the Cell Ranger pipeline (starting dimensions 28,000 genes $\times$ 2.2 million possible barcodes) were first filtered to remove barcodes with fewer than 250 unique molecules detected. A density plot of total unique molecular identifiers (UMIs) was then constructed for the remaining barcodes (Appendix C, Figures C.3, C.4). This revealed a bimodal distribution, with the higher peak representing Gel Beads in emulsion (GEMs) that had successfully captured cells. The UMI threshold was set at the first local minimum on the density plot and barcodes with fewer UMIs were removed. This method proved superior to the default Cell Ranger cell detection algorithm, which tended to underestimate the true cell number in the presence of the control cells, as these produced libraries an order of magnitude more complex than the primary stromal cells.

2.4.3 Initial clustering and QC

Clustering was performed using the R package 'Seurat'[254, 255]. Low-expressed genes (detected in fewer than 3 cells) were first removed, then variable genes were annotated by examination of the mean-variance relationship. Variable genes met the fol-

lowing criteria: 0.0125 < mean of non-zero values < 4 AND standard deviation > 0.5. Dimensionality reduction was then performed using PCA. The first 10 principle components were used to generate an initial clustering using the Seurat community detection algorithm. A t-SNE representation was also generated for visualisation.

The control cell cluster was detected based on expression of the lentiviral blasticidin resistance gene and known macrophage markers (Appendix C, Figure C.5). QC metrics used were total UMI counts and percentage of UMIs originating from mitochondrial features. Outlying cells (> 5 cluster median absolute deviations from the cluster median) were presumed to be either multiplets or of low quality and removed (Appendix C, Figure C.8). Control cells were also removed at this stage and analysed separately.

2.4.4 Cell cycle annotation, normalisation and final clustering

Cell cycle stage annotation was performed using the 'cyclone' function from the R package 'scran', which implements a previously published robust gene-pair based prediction method[256]. Unwanted variation due to library size (total UMIs), experimental batch and the G1 and G2M cell cycle scores was then regressed out. PCA was repeated using the adjusted expression values. The number of principle components to use as a basis for clustering was determined by examination of the scree plot and the statistical jackstraws test, which determines the robustness of a principle component by repeated permutation of small fractions of the dataset (Appendix C, Figure C.9). The first 28 principle components were used to generate a final Seurat clustering and accompanying t-SNE plot for visualisation.

2.4.5 Further Seurat analysis functions

A listing of the Seurat analysis functions used in generating the figures in Chapter 5 and Appendix C follows. Default settings were used except where indicated. Hierarchical phylogenetic trees were constructed using the 'BuildClusterTree' function. Out of bag error rates (OOBEs) for internal nodes in the tree, representing the accuracy with which cells could be allocated to branches by a random forest classifier

and therefore cluster confidence, were generated using the 'AssessNodes' function. The random forest classifier used to relate the clusters detected in the DSS dataset to corresponding clusters in the control dataset was trained and applied using the 'BuildRFClassifier' and 'ClassifyCells' functions respectively. Cluster marker gene lists for cross tabulation with the human dataset and ontology enrichment were generated using the 'FindMarkers' and 'FindAllMarkers' functions with the following additional settings: min.pct = 0.25, thresh.use = 0.25, return.thresh = 0.3, test.use = 'roc'. Cluster-specific differentially expressed genes comparing the DSS and control datasets were generated with the 'FindMarkers' function using a negative binomial differential expression test (test.use='negbinom') and the total UMI counts, G1 and G2M cell cycle scores and batch annotation as latent variables. Biological process ontology enrichment was performed as for the Fluidigm datasets using the R package 'clusterProfiler' with a BH multiple testing adjustment and a false-discovery rate cut-off of 0.05.

2.4.6 Diffusion maps and diffusion pseudotime

Diffusion maps and diffusion pseudotime analysis was performed with the R package 'destiny'[257, 258]. An expression matrix consisting of normalised variable gene UMI counts for cells annotated to fibroblast clusters in the control dataset (Str 1.1, 1.2, 1.3, 2, and 3) was input to the function 'DiffusionMap' and a diffusion map generated with a local scale parameter (sigma='local'), rotated eigenvalues (rotate=TRUE) and considering each cell's 500 nearest neighbours (k=500). Diffusion pseudotime was was calculated using the function 'DPT' with default settings.

2.5 Statistical methods

Differential expression analysis in the human datasets performed using generalised linear models and mixture models as implemented in the 'DESeq2' and 'scde' packages. In the murine datasets, the equivalent negative-binomial differential expression test implemented in 'Seurat' was employed. In both cases, the significance

of differentially expressed genes was adjusted for multiple testing using the Benjamini & Hochberg (BH) method with a false discovery rate (FDR) of 0.05.

Testing for significant ontology enrichment among candidate marker genes was performed using a hyper-geometric test, again with correction for multiple testing using the BH method with both p-value threshold and false discovery rate set to 0.05. The significance of marker gene overlaps between datasets was determined using Fisher's exact test. WGCNA module preservation was also quantified in the form of a 'Z-summary' score, a composite of density and connectivity based network module preservation statistics[250]. The significance of changes in the abundance of stromal cell subtypes across experimental conditions was determined using a two-tailed Student's t-test for independent samples.

2.6 Common validation methods

2.6.1 Flow cytometry

Cells were extracted from biopsies obtained from healthy donors or ulcerative colitis patients with active inflammation. Sub-populations of cells were visualized using antibodies against cell surface markers (see Key Resources Table), then fixed and permeabilised using either the Foxp3 / Transcription Factor Staining Buffer Set (eBioscience) for the detection of nuclear targets, or Cytofix/Cytoperm™ (BD) for the detection of cytoplasmic targets. Intracellular targets were stained using either primary antibodies pre-conjugated using Lightning-Link® (Innova Biosciences), or a combination of primary antibody followed by secondary antibody staining (see Key Resources Table). For the detection of CCL19 and IL33, freshly-isolated cells were stimulated with PMA (0.2 µM) / ionomycin (1 µg/ml) for 4 hours in the presence of Brefeldin A (10 µg/ml), or in Brefeldin A alone without stimulation, respectively. Where appropriate, an anti-fibroblast antibody (Miltenyi) was used for fibroblast visualization in combination with other cell surface markers, as expression of this antigen is not fibroblast-specific. Data were acquired on the BD LSRII or BD Fortessa Instruments with FACS Diva. All data analyses were performed with FlowJo (Tree Star). Representative gating strategies are shown in Appendix C,

Figure B.3. Alternate gating strategies were used for nuclear and cytoplasmic targets as the nuclear permeabilisation protocol was found to alter the forward-scatter (FSC) and side-scatter (SSC) properties of these samples. The diffuse population of high FSC/SSC events were predominantly apoptotic $EpCAM^+$ epithelial cells, which were excluded at later stages in both approaches.

2.6.2 Immunohistochemistry (IHC)

Tissue slides were de-paraffinized and antigen retrieval was performed by boiling slides for 20 minutes in citrate buffer, pH 6 in a vegetable steamer. Slides were allowed to cool and endogenous peroxidase was blocked by incubation with 0.3% H_2O_2 in PBS for thirty minutes. Slides were washed in PBS and blocked with 2.5% Normal Goat Serum (Vector Laboratories, Peterborough,U.K) for thirty minutes. Slides were incubated with primary antibodies dissolved in 1% BSA in PBS and incubated for two hours at room temperature. Please refer to Key Resources Table for antibodies used. Slides were washed in PBS containing 0.05% Tween 20 (PBS-T) and incubated with ImmPRESS™ HRP Anti-Rabbit or Anti-Mouse IgG (Peroxidase) Polymer Detection Kit (Vector Laboratories, Peterborough,U.K) for thirty minutes at room temperature. The slides were washed in PBS-T and the peroxidase activity was visualised using ImmPACT DAB Peroxidase (HRP) Substrate (Vector Laboratories, Peterborough,U.K). Finally, sections were counter-stained with Mayer's haematoxylin, dehydrated and mounted. Primary antibody was excluded in the negative controls.

For double staining, sections were de-paraffinized and antigen retrieval was performed by incubation for 20 minutes in citrate buffer, pH 6 in a vegetable steamer. Slides were allowed to cool and endogenous peroxidase was blocked by incubation with 0.3% H_2O_2 in distilled water for thirty minutes. Slides were washed with Tris-Hcl-buffered saline (TBS) and blocked with 2.5% Normal Goat Serum for thirty minutes. Sections were incubated with the first primary antibody for two hours. Thereafter, the slides were washed with TBS and incubated for thirty minutes with ImmPRESS™-AP Anti-Mouse or Anti-Rabbit IgG (alkaline phosphatase) Polymer

Detection Kit (Vector Laboratories, Peterborough,U.K). Slides were washed in TBS containing 0.05% Tween 20 for five minutes. AP activity was visualised with the AP substrate kit I Vector Red (SK-5100, Vector laboratories). Following this, the slides were washed in tap water and antigen retrieval was performed again for fifteen minutes in citrate buffer, pH6 in a vegetable steamer. The slides were subsequently blocked with Lab Vision™ Ultra V Block (Thermo Fisher Scientific, Paisley, U.K.) for fifteen minutes and incubated with the second primary antibody for two hours at room temperature. The slides were washed with TBS, incubated for thirty minutes with ImmPRESS™-AP Anti-Mouse or Anti-Rabbit IgG (alkaline phosphatase) Polymer Detection Kit and AP activity visualised with AP substrate kit Vector Blue (SK-5300, Vector laboratories). The slides were washed in tap water and allowed to air dry before mounting in Vectamount (Vector laboratories).

2.6.3 Single molecule in-situ hybridisation (smISH)

smISH was carried out on tissues that were fixed in 10% neutral buffered formalin for at least 24 hours. smISH was performed on 5 µm sections. All probes and RNAscope 2.5 HD assay — brown (cat. 310035) were purchased from Advanced Cell Diagnostics (ACD, Milan, Italy) and used according to the manufacturer's instructions. Probes used are listed in the Key Resources Table, paraffin sections were pretreated with Pretreat 1, 2, and 3 (all purchased from ACD). Pre-warmed (40 °C) probes were added to the slides and incubated in the HybEZ oven (catalog 321461; ACD) for 2 hours at 40 °C. After a 6-step signal amplification, tissues were detected by DAB (all part of the RNAscope 2.5 HD assay — brown kit) and counter-stained with Mayer's haematoxylin. Slides were mounted with PERTEX® mounting medium (Göteborg, Sweden) and photographed.

2.6.4 qPCR

RNA extraction was performed using the RNeasy Mini Kit (Qiagen) according to the kit instructions. Samples were homogenised using a QIAshredder (Qiagen) and DNA removed by treatment with RNAse free DNAse (Qiagen). RNA was eluted and quantified using a spectrophotometer. Reverse transcription was performed

using High-Capacity RNA-to-cDNA Kit (Thermo Fisher). The resultant cDNA was diluted ten-fold in nuclease free water and either used immediately or stored at -20C.

TaqMan gene expression assays were diluted in Gene Expression Master Mix (Applied Biosystems) according to the manufacturers instructions. Sample cDNA was added to give a final reaction volume of 10uL. Assays were performed in technical triplicate using the default ΔΔCT protocol on a 7500 Fast RT-PCR system. Glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) was used as an internal control.

2.7 Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat polyclonal anti-mouse IgG	BioLegend	405308
Donkey polyclonal anti-rabbit IgG	BioLegend	406410
Rabbit polyclonal anti-human ABI3BP	Atlas Antibodies	HPA044127
Rabbit polyclonal anti-human BCAS1	Atlas Antibodies	HPA051816
Mouse monoclonal anti-human CCL19	Bio-Techne	MAB361-100
Mouse monoclonal anti-human CD24	BioLegend	311135
Mouse monoclonal anti-human CD31	BioLegend	303121
Mouse monoclonal anti-human CD45	BioLegend	304023
Mouse monoclonal anti-human CD45	Agilent	M070129-2
Mouse monoclonal anti-human CD74	eBioscience	11-0748-41
Mouse monoclonal anti-human CD90	BioLegend	328107
Rabbit polyclonal anti-human CLCA1	Sigma	HPA059301
Mouse monoclonal anti-human EpCAM	BioLegend	324205
Rabbit polyclonal anti-human FDCSP	Abcam	ab121420
Recombinant human IgG1 anti-human fibroblast antigen	Miltenyi Biotec	130-100-139
Mouse monoclonal anti-human IL-33	Bio-Techne	NBP2-27333
Rabbit monoclonal anti-human LEF1	Cell Signaling	#2230
Rat monoclonal anti-human PDPN	BioLegend	337011
Rabbit polyclonal anti-human REG4	Atlas Antibodies	HPA046555
Rabbit polyclonal anti-human SMAD7	Atlas Antibodies	HPA028897
Mouse monoclonal anti-human SOX6	Abcam	ab84880
Rabbit polyclonal anti-human ST6GALNAC1	Atlas Antibodies	HPA014975
Mouse monoclonal anti-human ZEB1	Atlas Antibodies	AMAb90510
Rabbit polyclonal anti-human ZEB2	Atlas Antibodies	HPA003456
Biological Samples		

Human intestinal tissue biopsies	This study	
C57BL/6 male mice	Envigo Laboratories	
RAW 264.7	ATCC	TIB71
Chemicals, Peptides, and Recombinant Proteins		
ACK Lysing Buffer	Thermo Fisher	A1049201
Bovine Serum Albumin	Sigma	A1933
CellTracker™ Green CMFDA Dye	Thermo Fisher	C7025
CellTracker™ Orange CMRA Dye	Thermo Fisher	C34551
Collagenase VIII	Sigma Aldrich	C2139
Deoxyribonuclease 1 (DNAse 1)	Sigma	DN-25
DMEM	Sigma	D5671
Dextran Sodium Sulphate (DSS)	MP Biomedicals	lot #M9147
DL-Dithiothreitol solution (DTT)	Sigma	43816
UltraPure™ 0.5M EDTA, pH 8.0	Thermo Fisher	15575020
Foetal Bovine Serum	Sigma	F9665
HBSS, no calcium, no magnesium	Thermo Fisher	14170112
HEPES solution	Sigma	H0887
LIVE/DEAD® Fixable Dead Cell Stain	Thermo Fisher	L10120
NucBlue® Live Reagent	Thermo Fisher	R37605
Penicillin-Streptomycin	Sigma	P0781
Percoll	GE Healthcare	17-0891-01
Propidium Iodide Solution	BioLegend	421301
RPMI 1640	Thermo Fisher	11875093
TrypLE™ Express	Thermo Fisher	12605010
Critical Commercial Assays		
CD45 Microbeads, human	Miltenyi Biotec	130-045-801
CD45 MicroBeads, mouse	Miltenyi Biotec	130-052-301
CD235a MicroBeads, human	Miltenyi Biotec	130-050-501
CD326 (EpCAM) MicroBeads, human	Miltenyi Biotec	130-061-101
CD326 (EpCAM) MicroBeads, mouse	Miltenyi Biotec	130-105-958
Chromium™ Single Cell 3′ Library & Gel Bead Kit v2	10X Genomics	120237
Cytofix/Cytoperm™	BD Biosciences	554714
Foxp3 / Transcription Factor Staining Buffer Set	eBioscience	00-5523-00
LD Columns	Miltenyi Biotec	130-042-901
Lightning-Link APC conjugation kit	Innova Biosciences	705-0030
Nextera XT DNA Library Preparation Kit	Illumina	FC-131-1096
SMARTer® Ultra Low RNA Kit	Clontech	634832
Umbilical Cord Dissociation Kit human	Miltenyi Biotec	130-105-737
Sequence-Based Reagents		
RNAscope® Probe - Hs-ADAMDEC1	ACD Europe SRL	Cat#478471
RNAscope® Probe - Hs-BMP2	ACD Europe SRL	Cat#430641
RNAscope® Probe - Hs-BMP5	ACD Europe SRL	Cat#472461

RNAscope® Probe - Hs-CXCL12	ACD Europe SRL	Cat#422991
RNAscope® Probe - Hs-DCN	ACD Europe SRL	Cat#589521
RNAscope® Probe - Hs-F3	ACD Europe SRL	Cat#407611
RNAscope® Probe - Hs-FDCSP	ACD Europe SRL	Cat#444231
RNAscope® Probe - Hs-FRZB	ACD Europe SRL	Cat#412391
RNAscope® Probe - Hs-HSD17B2	ACD Europe SRL	Cat#467271
RNAscope® Probe - Hs-POSTN	ACD Europe SRL	Cat#409181
RNAscope® Probe - Hs-SLIT2	ACD Europe SRL	Cat#466221
RNAscope® Probe - Hs-WNT5A	ACD Europe SRL	Cat#604921
RNAscope® Probe - Hs-WNT5B	ACD Europe SRL	Cat#423041
RNAscope® Probe - Mm-BMP2	ACD Europe SRL	Cat#406661
RNAscope® Probe - Mm-FRZB	ACD Europe SRL	Cat#404861
RNAscope® Probe - Mm-WNT5A	ACD Europe SRL	Cat#316791
RNAscope® Probe - Mm-WNT5B	ACD Europe SRL	Cat#405051
Deposited Data		
single-cell RNAseq data	This study	GEO: GSE95459

Software	**Source**
Cell Ranger	`https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest`
clusterProfiler (v3.2.6)[249]	`https://github.com/GuangchuangYu/clusterProfiler`
DESeq2 (v1.17.13)[238]	`http://bioconductor.org/packages/release/bioc/html/DESeq2.html`
destiny (v2.5.6)[257]	`http://www.helmholtz-muenchen.de/icb/destiny`
FastQC (v0.10.1)	`http://www.bioinformatics.babraham.ac.uk/projects/fastqc/`
featureCounts (v1.5.0)[233]	`http://subread.sourceforge.net`
HISAT (v2.0.3b)[232]	`https://ccb.jhu.edu/software/hisat/index.shtml`
MultiQC (v0.7)[236]	`http://multiqc.info`
NMF (v0.20.6)[245]	`https://cran.r-project.org/package=NMF`
Picard Tools (v2.3.0)	`http://broadinstitute.github.io/picard/`
RUVSeq (v1.8.0)[242]	`https://github.com/drisso/RUVSeq`
scater (v1.2.0)[237]	`https://github.com/davismcc/scater`
scde (v2.5.0)[239]	`http://pklab.med.harvard.edu/scde`
scran (v1.2.0)[240]	`https://github.com/MarioniLab/scran`
Seurat (v1.4.0.16)[255]	`http://www.satijalab.org/seurat`
Skewer[231]	`https://github.com/relipmoc/skewer`
t-SNE[251, 252]	`https://github.com/lvdmaaten/bhtsne/`
WGCNA (v1.51)[243]	`https://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/`

Chapter 3

Single cell RNAseq of human colonic mucosal cells

Contents

3.1 Introduction 59
3.1.1 Motivation 61
3.2 Results 61
3.2.1 Cell isolation 61
3.2.2 Single cell capture, imaging and cDNA library generation 63
3.2.3 Single cell transcriptome assembly 64
3.2.4 Cell and gene QC process 64
3.2.5 Differential gene expression 65
3.2.6 Epithelial mesenchymal reciprocal signalling 69
3.2.7 Innate immune function in the the mesenchymal compartment 71
3.2.8 Correlation network analysis 76
3.2.9 Cell clustering based on module expression 81
3.2.10 Evaluation of clustering in epithelial dataset 84
3.2.11 Comparison with other clustering methods 87
3.3 Discussion 89

3.1 Introduction

Tissues at barrier surfaces such as the intestinal mucosa are continuously exposed to microbial and environmental stimuli that have the potential to compromise their function. They must have the capacity both to mobilise inflammatory responses to eliminate harmful sources and to limit immune activation under homeostatic conditions. Cells of the haematopoietic immune system are the critical effectors in

producing inflammatory responses and are abundantly distributed in the lamina propria and overlying epithelial monolayer at intestinal mucosal surfaces. In these locations, they are in close proximity to to non-immune cells that regulate their function. In the colon, two broad cell types — epithelial cells and connective tissue mesenchymal cells — constitute the bulk of this population.

While there is some overlap, it is clear that epithelial and mesenchymal cells sense their environment and signal immunologically using distinct pathways. The high diversity and spatial compartmentalisation of immune cell types in the gut suggests that further specialisation exists, perhaps influenced by factors such as the micro-anatomical niche or cell subtype. For example, an upper-crypt 'sentinel' goblet cell population has recently been described that senses PAMPs via TLRs 2, 4 and 5 and activates the Nlrp6 inflammasome to trigger mucin exocytosis[259]. It is likely that immunologically specialised subsets exist within the mesenchymal compartment as well, given the heterogeneity in developmental origin, activation and differentiation states present in the resident cell populations.

Construction of a cohesive structural framework for the multiple epithelial and mesenchymal cell subtypes present in the colonic mucosa is a necessary pre-requisite for an integrative understanding of barrier immunity in this setting. Delineation of the hard-wired innate immune pathways expressed by cell sub-populations will provide insight into their contribution to immune homeostasis and a basis for future functional experiments. However, the complex mix of cell types, lack of specific surface markers, and dependence on cues from the microenvironment and microbiome have made this a formidable challenge using the established cell selection and expansion techniques required for bulk assays of gene and protein expression. The recent emergence of microfluidic technologies for cell handling together with sequencing chemistries compatible with extremely small amounts of starting material have made it possible to capture a cross-section of transcriptional activity from a single cell[260]. Single-cell RNA-seq (scRNAseq) has been used to classify cells in the murine adult and embryonic nervous system[222], spleen[223], murine lung and intestinal epithelial cells[224, 225] and immune cells such as CD4+ T cells

and ILCs[226–228]. As yet this technology has not been applied to classify rare human intestinal cells in health and inflammation.

3.1.1 Motivation

The aims of this section of my project were to utilise single-cell RNAseq to interrogate the heterogeneity and immune functions of human primary epithelial and mesenchymal cells. This required the optimisation and validation of a pipeline for the isolation and single-cell sequencing of cells derived from endoscopic forceps biopsies together with implementation of bioinformatic approaches to explore cellular heterogeneity and differential gene expression. In the absence of a clear consensus regarding the expected mesenchymal cell types within the colonic mucosa, and to maximise the potential for discovery of novel cell subsets, an unsupervised classification was employed in which cells were clustered according to an objective measure of their gene-expression similarity.

3.2 Results

3.2.1 Cell isolation

Samples used for single cell RNAseq and in optimisation of the isolation protocols consisted of endoscopic forceps biopsies collected from healthy donors undergoing clinically-indicated diagnostic colonoscopy or flexible sigmoidoscopy procedures. The mean age of participants was 51 years and no abnormalities were detected either during their endoscopy procedures or on histological examination of the biopsy tissue (see Appendix A, Tables A.1, A.2).

Epithelial cells

IECs were isolated using a previously published method[229] with some modifications (see Chapter 2). In order to minimize processing time prior to lysis and thus obtain transcriptional profiles of cells close to their native states, I omitted marker-based selection of cells and instead separated whole crypt units from mesenchymal elements by mechanical dissociation (Figure 3.1A), then further dissociated the crypts to

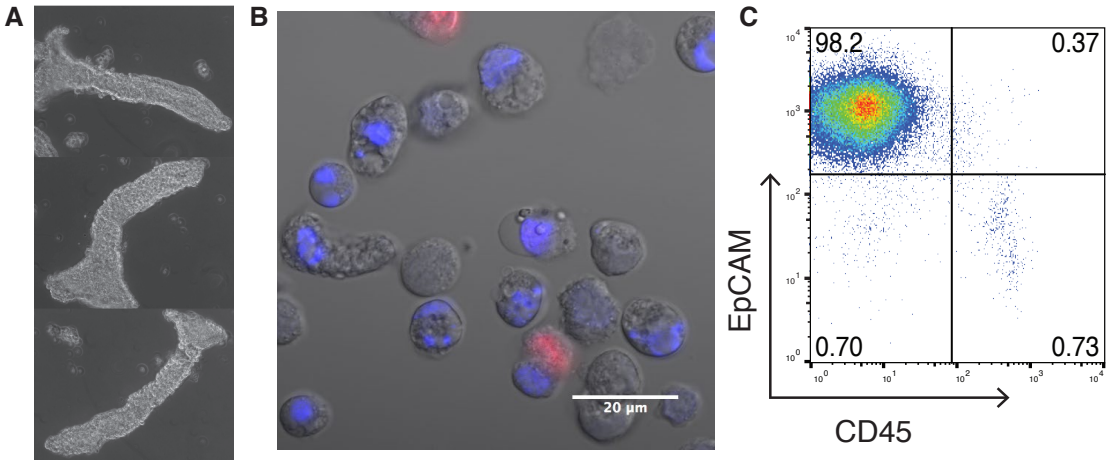


Figure 3.1: Isolation of human colonic epithelial cells. (A) Intact epithelial crypt units dissociated from colonoscopic biopsies. (B) Single cell suspension derived from A. The image is a composite of phase contrast, Hoechst 33342 (blue) and propidium iodide (red) channels. (C) Flow cytometric analysis of the indicated surface markers on disaggregated crypt suspensions.

a single cell suspension (Figure 3.1B). The composition of this suspension was assessed by flow cytometry (Figure 3.1C), which confirmed $> 98\%$ purity for the epithelial lineage marker EPCAM. Imaging of the resultant cell suspension revealed a mean cell diameter of 13.5 µm (SD 2.1). This was compatible with the 'medium' Fluidigm C1 IFC (cell tolerance 10 µm–17 µm). Attempts at further purification by FACS or density gradient centrifugation produced only incremental increases in cell purity at a cost of increased processing time and reduced cell yield.

Mesenchymal cells

To facilitate unbiased capture of a cross-section of mesenchymal cells in the context of a lack of specific markers and unknown heterogeneity within this group, a negative selection protocol was employed. Whole biopsies were dissociated to single cells by a combined mechanical and enzymatic approach (see Chapter 2) and cells expressing EPCAM, CD45, and CD235a were depleted by MACS, targeting the whole of the haematopoietic and epithelial compartments. Flow cytometric analysis of the products of this separation confirmed depletion of CD45 and EpCAM positive cells, together with enrichment for THY1 (CD90), a candidate stromal marker, in the depleted fraction (Figure 3.2). Microscopic examination of the cell suspension

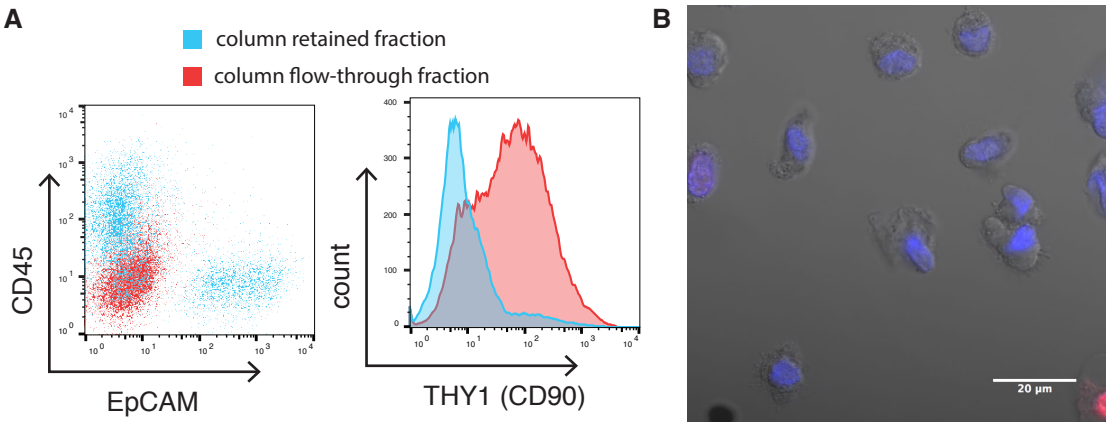


Figure 3.2: Isolation of human colonic mesenchymal cells. (A) Flow cytometry analysis of the indicated surface markers on dissociated to single cell suspensions obtained from colonic biopsies following removal of epithelial and haematopoietic cells by MACS. Column flow-through fraction is shown in red and column-retained fraction in blue. (B) Suspension of single mesenchymal cells. Image is a composite of phase contrast, Hoechst 33342 (blue) and propidium iodide (red) channels.

revealed a mean cell diameter of 10 µm (SD 2.3), compatible with the 'small' Fluidigm C1 IFC (cell tolerance 5 µm to 10 µm).

3.2.2 Single cell capture, imaging and cDNA library generation

Single cell capture was performed using the Fluidigm C1 system. C1 IFCs were removed after sample loading and imaged using an automated acquisition platform and a pre-registered template containing the capture site locations. Capture site images were manually inspected to annotate capture site outcomes (Appendix A, Figure A.1). In the epithelial experiment, a mean of 48 viable single cells (range 34–56) and were captured per donor compared to 61 (range 57–68) for the mesenchymal samples. Control samples consisting of empty capture site cDNA and bulk cDNA, derived from 1,000 cells lysed together, were generated to provided upper and lower limits against which the quality single cell transcriptomes could be assessed. 92 epithelial cell cDNA libraries were generated from 3 donors and sequenced together with 3 bulk controls (1 per donor) and 1 empty well control. For the mesenchymal experiment, 183 single cell libraries were sequenced together with 6 bulk controls and 3 empty capture site controls.

3.2.3 Single cell transcriptome assembly

Reads were trimmed to remove adaptor sequences and aligned to the Ensembl GRCh38 build of the human genome using a splice site and SNP-aware alignment tool (see Chapter 2). Median alignment rates were 92% and 89% for the epithelial and mesenchymal single cell libraries respectively. Aligned reads were summarised to gene count tables. Quantification outputs are shown in Appendix A, Figure A.2. Multiple mapping and ambiguously mapping reads were discarded, but reads flagged as possible duplicates were retained (see Chapter 2 for further information). The effect of duplicate removal on individual gene counts at the majority of biologically relevant expression levels was low (Appendix A, Figure A.3). A higher-than-expected proportion of aligned reads could not be assigned to any annotated genomic feature (i.e. they were aligned to intergenic or intronic sequences). This was observed in multiple datasets within the group where single cells were derived from digestion of primary tissue as opposed to suspension cells or cell lines, and may reflect a degree of DNA contamination.

3.2.4 Cell and gene QC process

A quality control (QC) process was employed to identify and remove low quality single cell libraries. This is detailed in Chapter 2. Cells were assessed using a panel of QC measures encompassing several domains including sequencing depth, library complexity and sequencing saturation (Appendix A, Figure A.4), the proportion of counts arising from ERCC spike-in RNA molecules (Appendix A, Figure A.5), and the proportion of counts arising from mitochondrial features (Appendix A, Figure A.6). An aggregate score was calculated for all measures (Appendix A, Figure A.7), and based on this a total of 70 epithelial and 122 mesenchymal single cell libraries passed the QC process. The median number of unique genes detected per cell after quality control was 5,800 in the epithelial dataset and 4,900 in the mesenchyme. Genes that showed low-expression over the dataset as a whole were also removed (Appendix A, Figure A.8).

3.2.5 Differential gene expression

Differential expression analysis was performed on the cells and genes passing QC to give an initial indication of the success of the single-cell isolation and sequencing approaches in capturing the intended cell types. Multiple methods are available for performing differential expression analyses, although only a small subset have been benchmarked in single-cell settings. In light of this, two methods were used in parallel and the results compared. DESeq2[238] is a widely used method developed for bulk RNAseq experiments in which overall expression strength and fold change between experimental conditions are estimated for each gene using negative binomial models. Differentially expressed genes are then identified based on increased dispersion compared to a fit curve that captures the overall trend of dispersion-mean dependence. Expression estimates from single-cell RNAseq experiments are, however, frequently bimodal[219], and this is cannot be adequately captured in a negative binomial model. SCDE[239] is a single-cell specific method that uses a zero-inflated negative binomial model to better characterise single cell expression distributions.

A small number of contaminating endothelial and plasma cell libraries were identified in the mesenchymal dataset based on high expression of well known markers for these cell types (Appendix A Figure A.9). Following removal of these, an expression matrix of all remaining cells that passed QC was constructed. Differential expression analysis was performed using both approaches. Reassuringly, there was considerable overlap in the differentially expressed genes identified, with 4,481 common to both methods (Figure 3.3).

The genes identified as showing the strongest differential expression between the epithelial and mesenchymal cells were inspected for the presence of known cell type markers (Figure 3.4A). Pan-fibroblast markers such as the intermediate filament vimentin (*VIM*) and collagen types 1 and 3 (*COL1A1*, *COL3A1*) were among the highest ranked genes in the mesenchyme, while well-known epithelial lineage markers such as epithelial cell adhesion molecule (*EPCAM*), epithelial cadherin (*CDH1*) and Keratin 19 (*KRT19*) showed specific expression in the



Figure 3.3: Identification of genes showing differential expression in the epithelial and mesenchymal compartments. Number of significantly differentially expressed genes detected using the methods indicated. A multiple testing correction was applied with a false-discovery rate threshold of 5%

epithelium as expected. Marker genes also provided insight into the differing biological roles of these cells, with the epithelial fraction showing enrichment for genes involved in maintenance of the intestinal barrier such as mucins (*MUC4*) and the polymeric immunoglobulin receptor (*PIGR*) which facilitates the secretion of immunoglobulin A into the colonic lumen. Mesenchymal markers included genes involved in extracellular matrix remodelling such as matrix metalloproteinase 2 (*MMP2*) and ADAM Like Decysin 1 (*ADAMDEC1*). Interestingly, *ADAMDEC1* has been linked to a number of inflammatory diseases including rheumatoid arthritis and Crohn's disease[261, 262]. It's expression in the steady state was thought to be restricted to gastrointestinal tract macrophages and dendritic cells[263], but this data suggests that it is also abundantly expressed in lamina propria fibroblast-like cells. Mouse knock-out models show increased susceptibility to infectious and chemical colitis challenges[264].

To better define the distinct activities of these cell subsets, biological process ontology enrichment was performed (Figure 3.4B). Mesenchymal cells showed enrichment for genes involved in vascularisation, extracellular matrix organisation and mesenchymal development. Epithelial cells enriched for digestive and absorptive functions and protein glycosylation — a critical step in the production of mucins that form the intestinal mucous layer. Both cell types showed enrichment for genes

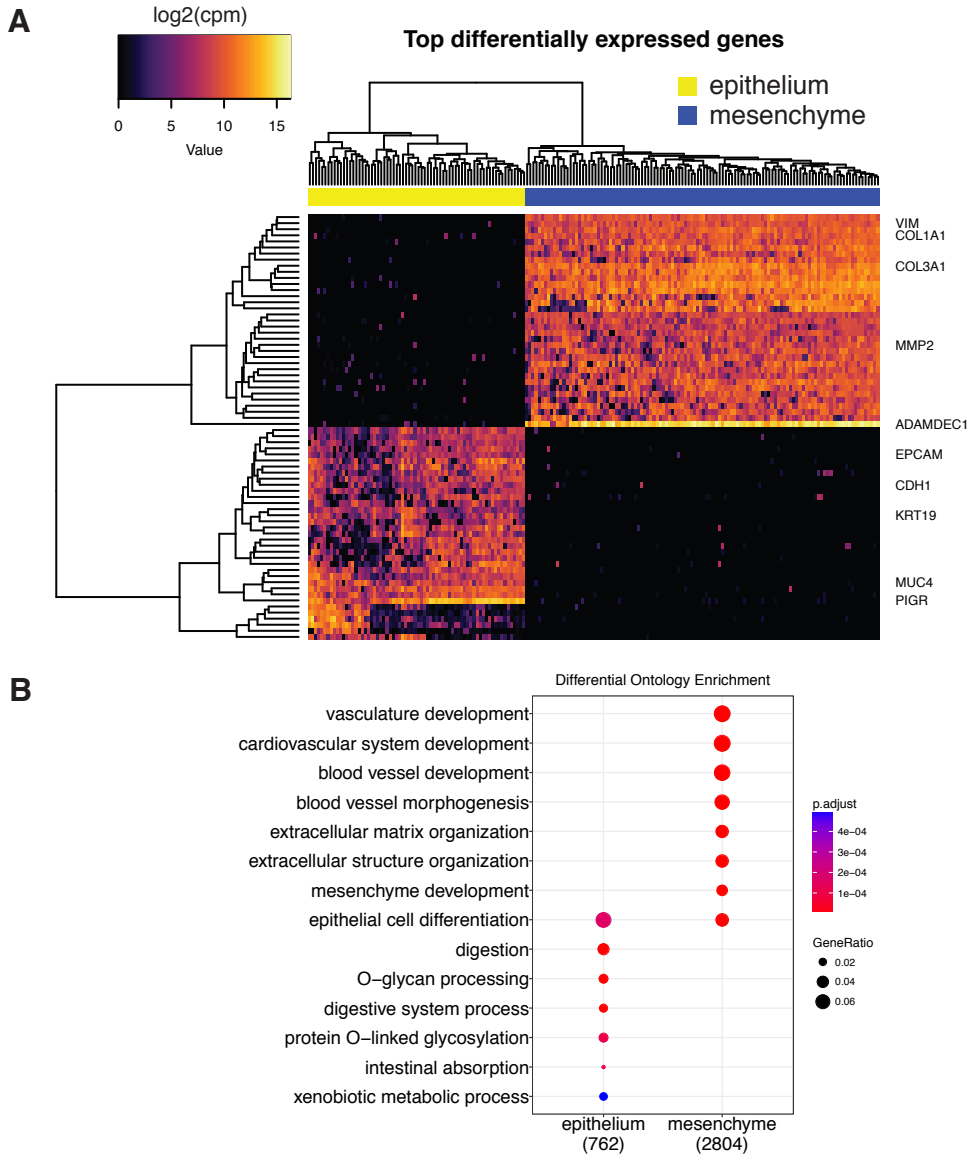


Figure 3.4: Differentially expressed marker genes and processes in the colonic epithelial and mesenchymal compartments. (A) Heat map showing differentially expressed genes (rows) in mesenchymal and epithelial cells (columns). The top and bottom 35 genes ranked by SCDE median log expression difference are included. Marker genes mentioned in the text are shown (row labels). (B) Enriched biological process ontology terms (rows). The background number of differentially expressed genes for each cell type (x-axis, parenthesised), the proportion of annotated genes per term (circle size) and the adjusted p-value (circle colour) are indicated.

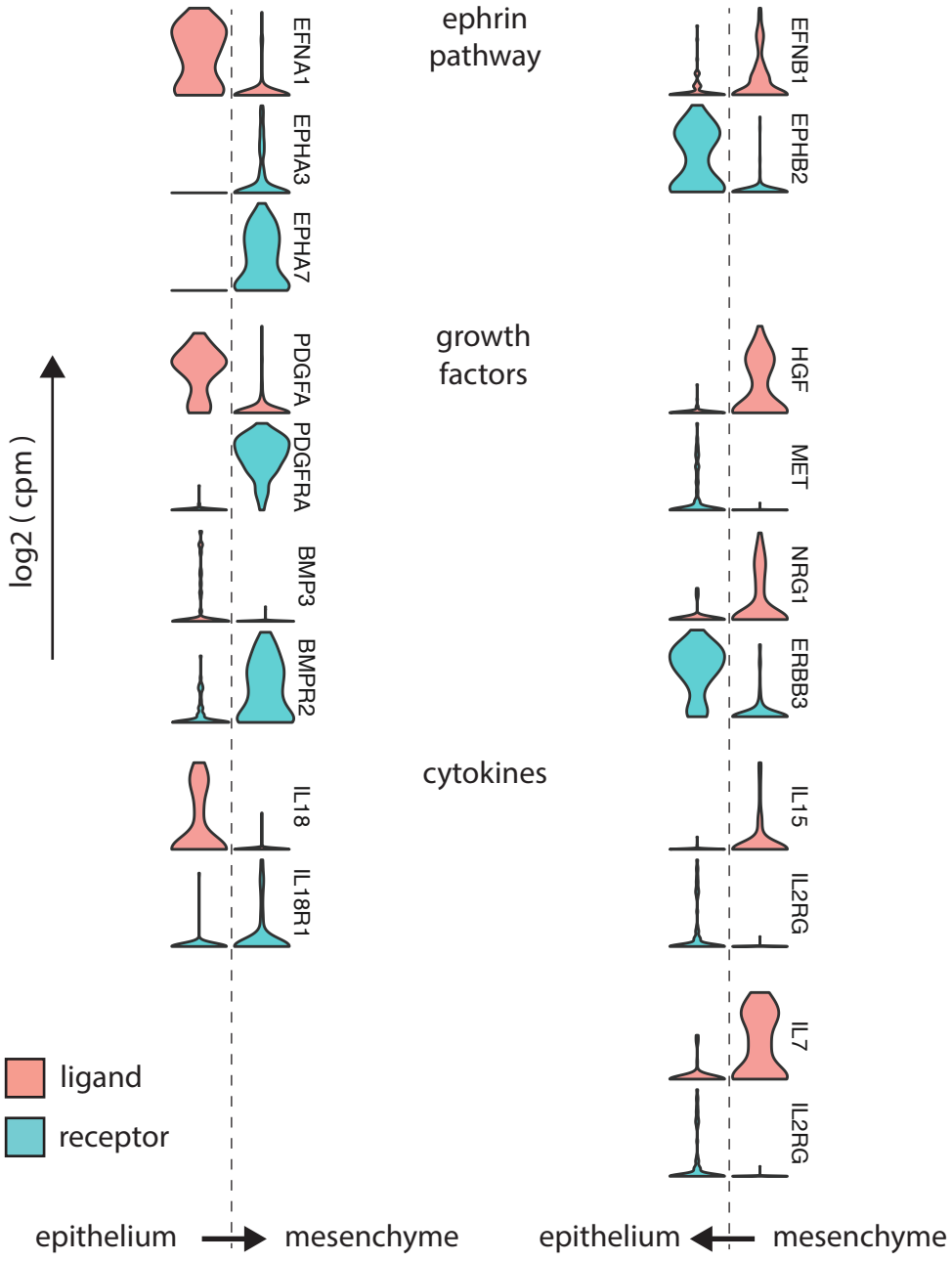


Figure 3.5: Reciprocal paracrine signalling pathways identified from single-cell RNAseq data. Selected ligand-receptor pairs showing significant differential expression between epithelial (left column facet) and mesenchymal (right column facet) compartments. Log-transformed read counts per million (cpm) are shown on the y-axis. Local cell density over the range of expression values is represented by violin width.

implicated in epithelial differentiation, perhaps reflecting the importance of crosstalk between the mesenchyme and epithelium in directing this process.

3.2.6 Epithelial mesenchymal reciprocal signalling

To explore the nature of epithelial-mesenchymal crosstalk, differentially expressed genes were screened for cognate ligand-receptor pairs where the ligand showed significant enrichment in one compartment and its receptors in the other. Candidate pairs were derived from a previously published list[265]. Among the identified hits were members of the ephrin family, growth factors and cytokines (Figure 3.5). Known important axes such as HGF/MET, BMP signalling and PDGF-A were recovered by this method in addition to novel hits.

Ephrins, the membrane-tethered ligands for the Eph-receptor tyrosine kinases, play an important role in the spatial organisation and patterning of multiple tissues. Mice mutant for the receptors EphB2, EpHB3 or their ligand Ephrin-B1 show defects in crypt compartmentalisation[266, 267]. The major source of Ephrin-B1 is thought to be differentiated epithelial cells at the crypt top. However, this data suggests significant mesenchymal expression. The role of the Ephrin-A subclass in the intestine is less well understood, although Ephrin-A has been shown to regulate apoptosis in other tissues[268] and to increase tumour burden in animal models of colorectal cancer[269]. The major source of Ephrin-A in this dataset is epithelial cells, while the *EPHA3* and *EPHA7* receptors are expressed predominantly in the mesenchyme.

Platelet-derived growth factor signalling follows a similar pattern, with the *PDGFA* ligand expressed in the epithelium and the *PDGFRA* receptor in the mesenchyme. This is concordant with experimental evidence from knockout models showing that epithelial PDGF-A maintains a functionally distinct PDGFRα^+ fibroblast population required for villus formation[78].

Signalling in the opposite direction is seen in the human epidermal growth factor pathway. The receptor *ERBB3* (or *HER3*) shows high expression in the epithelial compartment, whereas its ligand neuregulin 1 (*NRG1*) is expressed in the mesenchyme. *ERBB3* promotes cell survival and proliferation via interaction with *ERBB2*/*HER2* and activation of PI-3 K/Akt signaling[270]. This important pathway is targeted clinically by the monoclonal antibody tradilizumab (herceptin)

in the treatment of breast cancer. Neuregulin-1 has been shown to promote both epithelial organoid expansion in culture[271] and colorectal tumour progression in animal models[272], while selective epithelial Erbb3 knockout worsens and prolongs DSS colitis by reducing epithelial cell proliferation and survival[273].

The cytokines *IL18*, *IL15* and *IL7* also showed reciprocal ligand-receptor expression between the mesenchymal and epithelial compartments. *IL18* is an IL-1 family cytokine produced by epithelial cells that has been implicated both as a protective and exacerbating factor in different animal models of colitis[274, 275]. In the epithelial compartment, selective ablation of the IL-18 receptor protects against goblet cell depletion[276]. The physiological function of IL-18 signalling in lamina propria fibroblasts has not been defined, but this data suggests that the receptor is expressed more abundantly than in the epithelium in the steady state.

IL-7 has a well-defined role in lymph node biology, where it is produced by fibroblastic reticular cells and maintains the naive lymphocyte population[168]. In the intestinal mucosa, it is consumed by mature T-cells and acts as a survival factor. In lymphopenic animals, IL-7 accumulates and induces hyperplasia in the epithelium, suggesting a role maintaining a balance between inflammation and repair[277]. The major cellular source of IL-7 in the intestine is thought to be epithelial cells[278]. However, in this experiment *IL7* expression is significantly higher in the mesenchyme than the epithelium, which would be analogous to the situation in the lymph node.

IL-15, which is upregulated in the mesenchymal compartment, is a pleiotrophic cytokine that is induced following inflammatory stimuli and required for the maintenance of intestinal intra-epithelial lymphocytes (IELs)[279]. It also reduces epithelial cell apoptosis and ablation is protective in the AOM/DSS model of inflammation induced colonic carcinogenesis[280]. Delivery is mediated by a unique trans-presentation mechanism that requires direct contact with neighbouring cells expressing of the low-affinity IL-2Rβ and γC (*IL2RG*) receptor chains[281]. The cells providing IL-15 signals in the colonic mucosa have not been determined, but fibroblast-like cells would be promising candidates.

3.2.7 Innate immune function in the the mesenchymal compartment

A notable feature of the ontology analysis was the presence of multiple enriched terms related to immune functions in mesenchymal cells. A total of 93 offspring of 'immune system process' and 11 offspring of 'innate immune response' were significantly enriched among the differentially expressed genes showing increased expression in the mesenchyme (Figure 3.6). None of the offspring of either term was enriched in the epithelial compartment. Focussing on pattern recognition receptor signaling, upregulated genes were identified at multiple sequential stages of the signaling pathway including sensors such as *TLR4*, downstream kinases such as *TAK1* and *MAP3K1*, transcription factors such as *NFKB1* and *IRF3*, and a range of transcriptional targets including interferon stimulated genes and chemokines.

To further characterise the immune functions of mesenchymal cells, pathway enrichment was performed using the KEGG pathway database[282], a manually curated collection of pathways derived from experimental evidence. Among immune pathways, significant enrichment was found for 'antigen processing and presentation' and 'NOD-like receptor signalling' (Figures 3.7, 3.8). Several of the critical genes involved in antigen presentation, including both classes of MHC molecules, the peptide transport and loading machinery (*TAP1*, *TAPBP*), and components of the immunoproteosome (*PSME1*, *PSME2*), a modified proteosome that processes class I MHC peptides, were identified as differentially upregulated. The related pathway 'phagosome' also showed significant enrichment consistent with founding data for this project that colonic fibroblast-like cells are capable of phagocytosis and antigen presentation[116] (Figure 3.9).

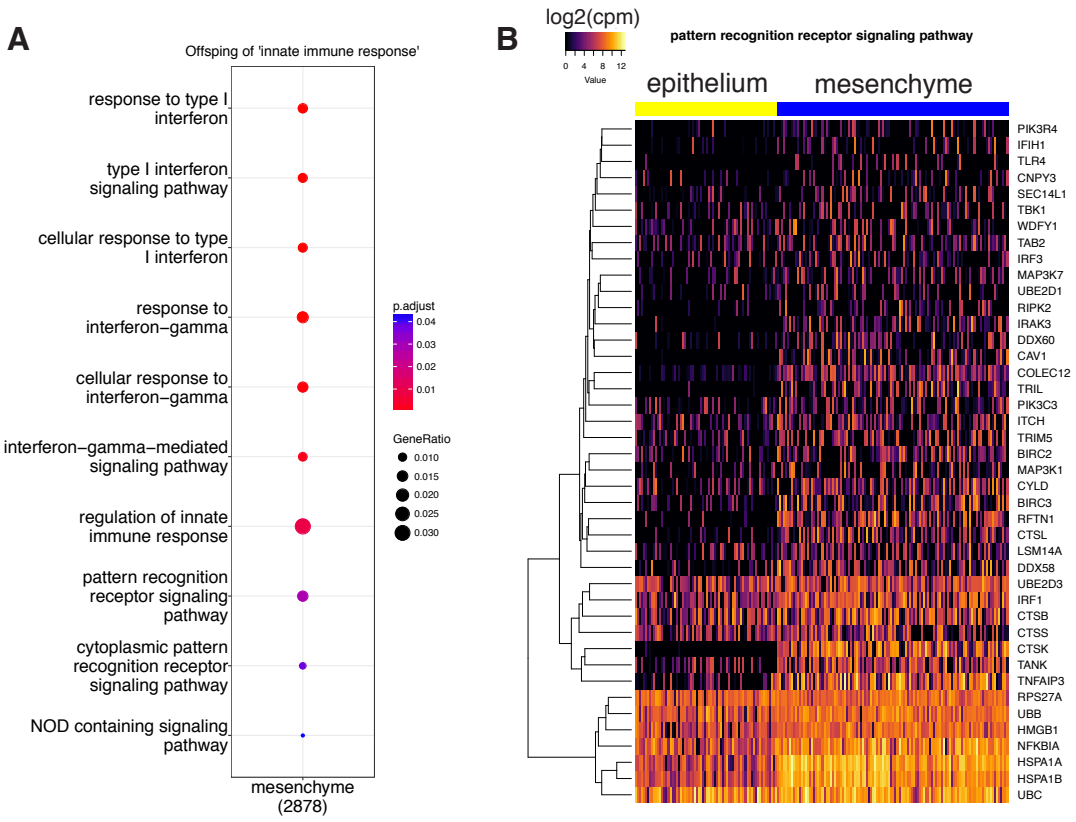


Figure 3.6: Enrichment for innate-immune processes in colonic mesenchymal cells. (A) Offspring of the ontology term 'innate immune response' showing statistically significant enrichment among genes identified as differentially expressed. The proportion of annotated genes per term (circle size) and the adjusted p-value (circle colour) are indicated. (B) Heat map showing differentially expressed genes (rows) annotated with the ontology term 'pattern recognition receptor signaling pathway' against cells (columns) ordered by isolation protocol (colour bar).

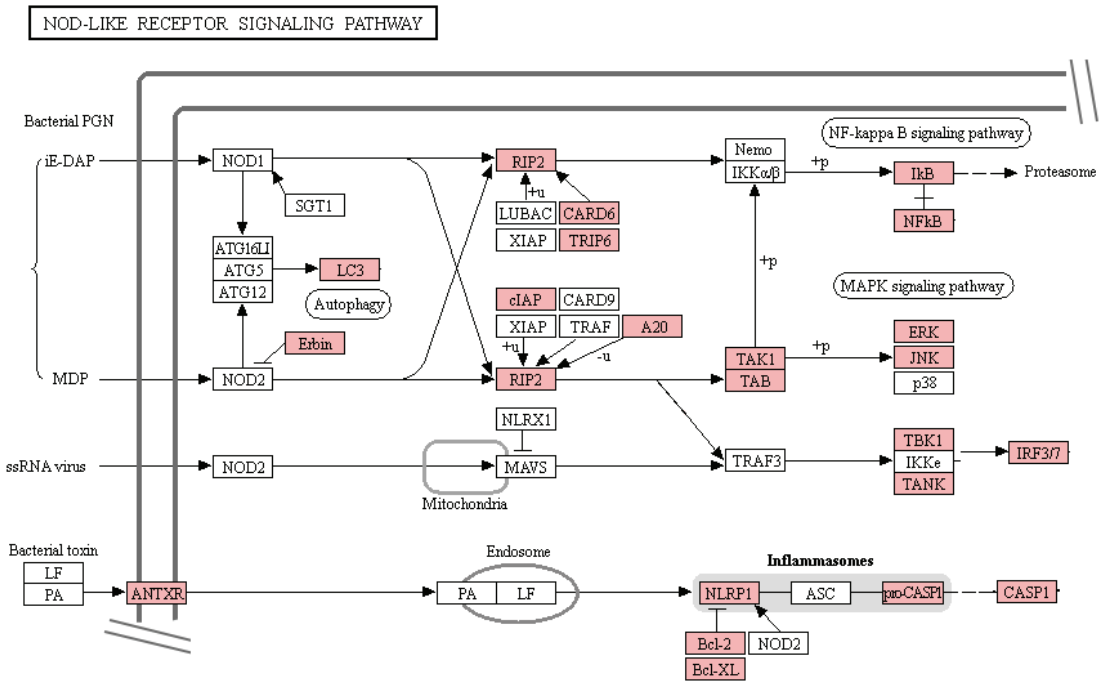


Figure 3.7: KEGG pathway enrichment: NOD-Like receptor signaling. A section of the indicated KEGG pathway map. Shaded nodes show significantly increased expression in colonic mesenchymal cells compared to epithelial cells. Abbreviations: peptidoglycan (PGN), protective antigen (PA), lethal factor (LF), muramyl dipeptide (MDP), D-glutamyl-meso-diaminopimelic acid (iE-DAP), phosphorylation (+p), ubiquitination (+u).

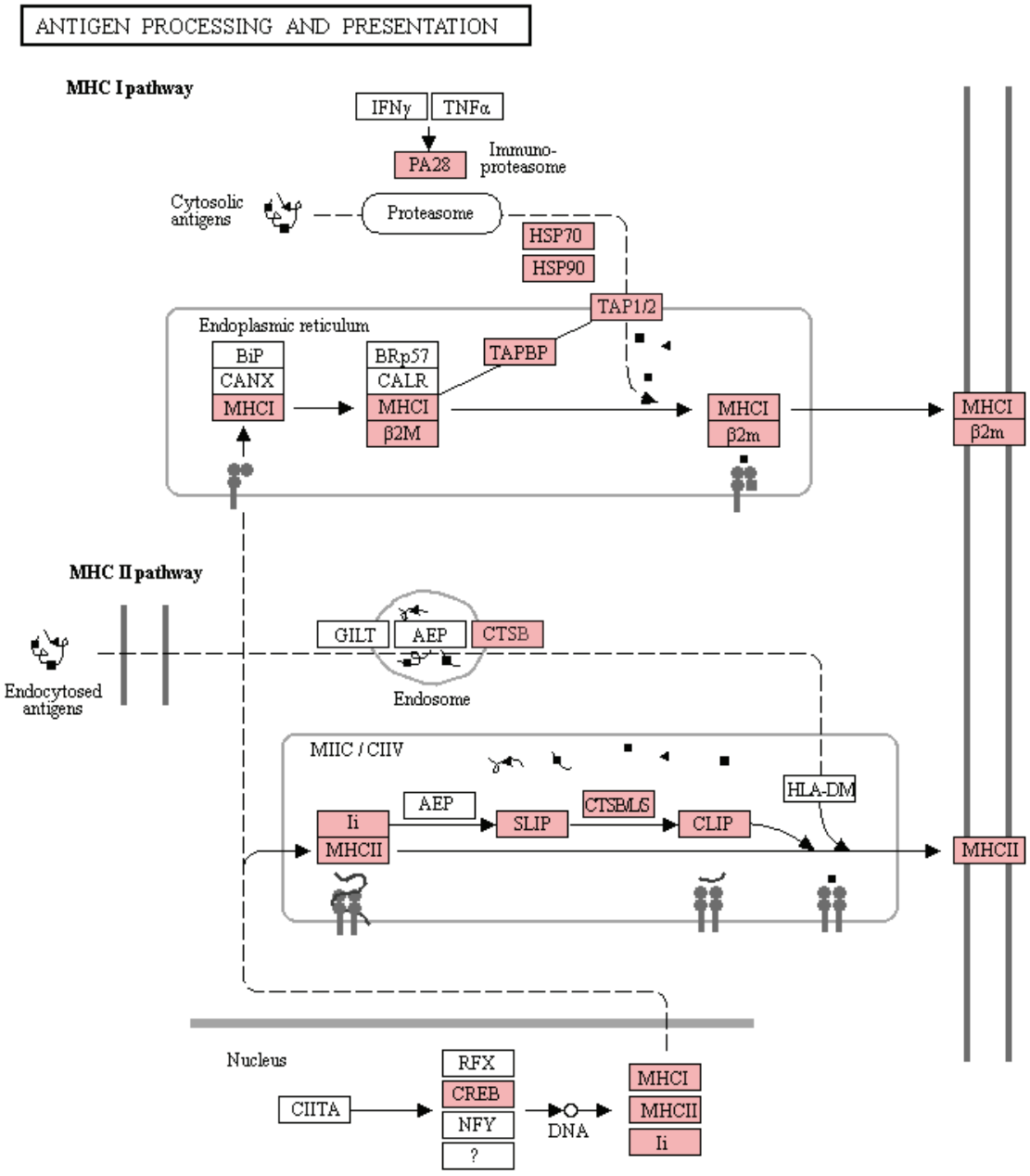


Figure 3.8: KEGG pathway enrichment: Antigen processing and presentation. KEGG pathway map as indicated. Shaded nodes show significantly increased expression in colonic mesenchymal cells compared to epithelial cells.

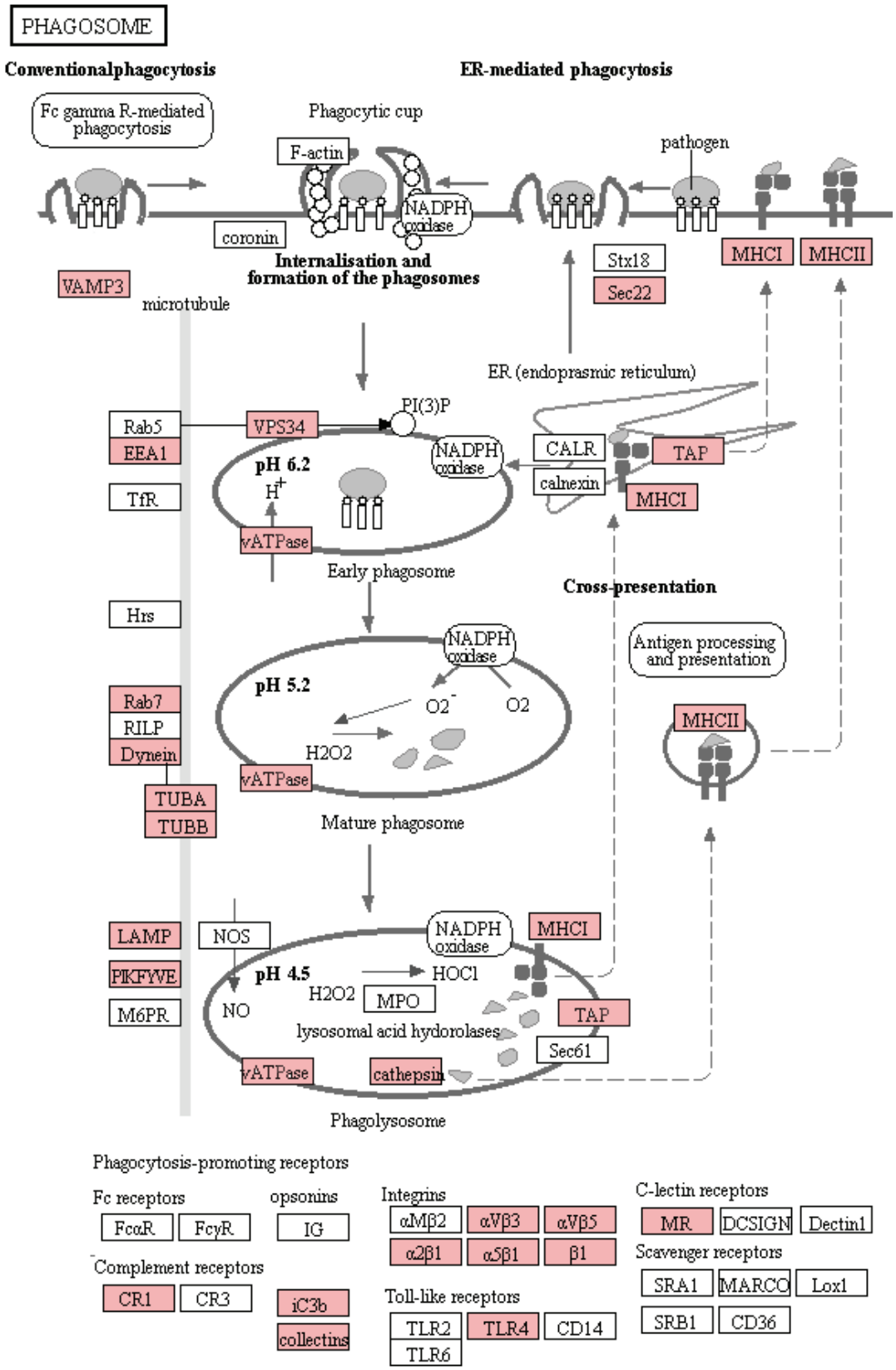


Figure 3.9: KEGG pathway enrichment: Phagosome. KEGG pathway map as indicated. Shaded nodes show significantly increased expression in mesenchymal cells.

3.2.8 Correlation network analysis

While differential expression analysis is useful in making a comparison of the relative importance of a set of genes or pathways between cell groups, it depends on prior knowledge of group assignment and its results must be interpreted in the context of the comparator used. For example, comparing mesenchymal cells to professional immune cells in the analysis above would likely reveal a different perspective on their biological activities. Correlation network analysis is an alternate approach to identifying structure in high-dimensional data that has been the subject of considerable interest in gene expression and systems biology applications. Instead of performing comparisons between groups, a correlation function is used to define relationships between the quantitative measured variables (gene expression values in this case) within the group of interest. Pairwise gene correlations can be expressed as an adjacency matrix, or more intuitively as a network with genes as vertices and edges representing the strength of their co-expression relationships. Network proximity measures can then be used to identify groups of genes showing high expression similarity, with the underlying assumption that this would be driven by important biological factors such as shared transcriptional regulators, participation in the same protein complexes or pathways, or specific expression in the same cell sub-populations. In this way, a picture of the heterogeneity and biological processes occurring in a given dataset is produced without reference to a control or external gene annotation.

Weighted gene co-expression network analysis (WGCNA) is an implementation of this approach that has been employed to extract biologically meaningful gene-expression signatures from genomic datasets in a number of settings, including single-cell RNAseq datasets[243, 283–285]. It employs hierarchical clustering to capture network proximity in the form of a clustering dendrogram. A branch 'pruning' algorithm then identifies promising groups of genes as 'modules' based on branch size, joining height and geometry. Unlike the methods used for differential expression analysis, which specifically address non-biological sources of variability in RNAseq data such as overall library size and batch effects in fitting their

expression models, WGCNA is a general method that requires prior adjustment of the expression matrix to account for these factors. Library size normalisation was therefore performed using the package 'scran'[240] and batch effect correction with the package 'RUVseq'[242] to generate an adjusted expression matrix. A discussion of alternate normalisation and batch correction approaches is included in Chapter 2, while measures of the effectiveness of these adjustments in reducing these sources of unwanted variation are shown in Appendix A (Figures A.10, A.11). Finally, a subset of 'biologically variable genes' was identified based on individual gene variances exceeding a fit of the mean-variance trend for the dataset as a whole (Figure A.12). A total of 7,076 biologically variable protein-coding genes were identified in the epithelial dataset and 7,095 in the mesenchyme. Adjusted expression values for these genes were used as a basis for network construction and module detection using WGCNA (see methods). Five modules of co-expressed genes were detected in the epithelial dataset and four modules in the mesenchyme (Figure 3.10).

The quality of detected modules can be assessed using inherent network characteristics or by reference to external annotation such as ontology databases[286]. Relevant network characteristics include density, connectivity, centralisation, and clustering coefficient. A gene's connectivity is the sum of its connection strengths to all other genes in the module. Mean connectivity in a module therefore relates to the overall degree of expression similarity observed but will naturally be higher for larger modules. The related density statistic, which is the mean pairwise intra-modular connection strength, does not suffer from this problem. Centralisation reflects the presence of central hub genes, which show co-expression relationships with many other genes within a module and may be biologically significant (Figure 3.11A). The clustering coefficient reflects the degree of 'cliquishness' in the module, which is the extent to which a gene's neighbours also have co-expression relationships with one another (Figure 3.11B). The performance in all of these measures was significantly higher in detected modules than in control modules of genes sampled at random (Figure 3.11C,D).

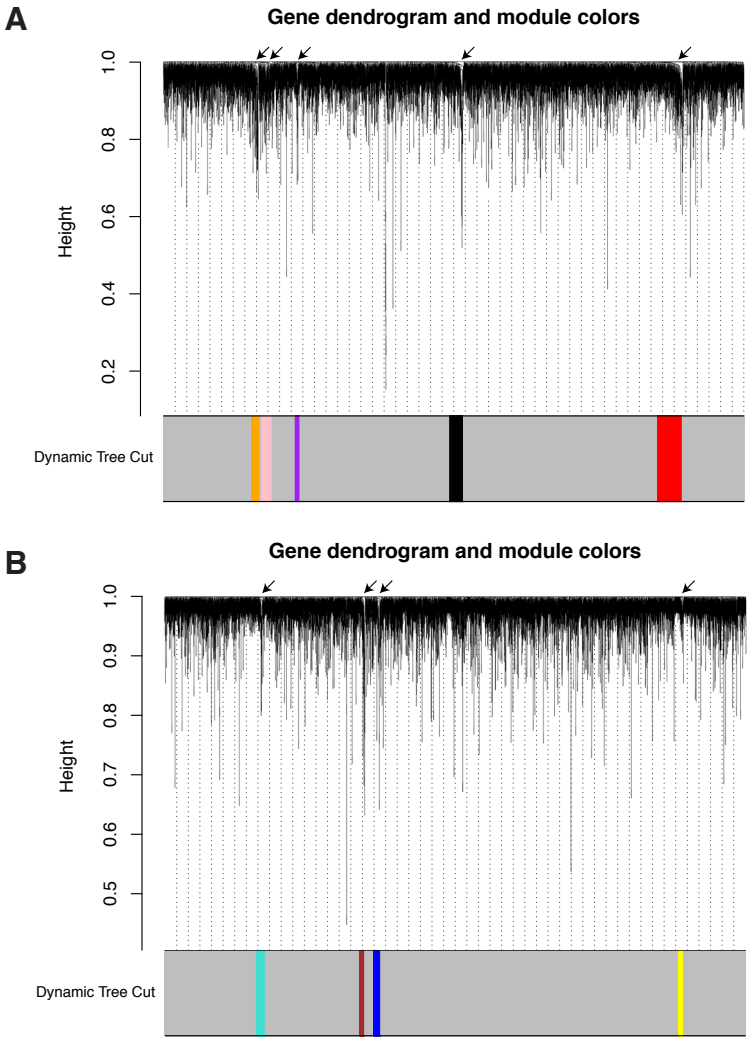


Figure 3.10: Detection of gene co-expression modules. WGCNA clustering dendrograms and detected modules for the epithelial (A) and mesenchymal (B) datasets. The y-axis, height, reflects expression dissimilarity from 0 (identical expression) to 1 (expression correlation ≤ 0). The dendrogram was cut to identify modules of genes showing similar expression profiles (arrows, see methods), which were assigned a unique colour identifier (colour bar).

Modules reflecting authentic biological processes might also be expected to show ontology enrichment, with the proviso that not all cellular processes are equally well annotated. In these datasets, all detected modules showed functional ontology enrichment for multiple terms (Figure 3.12). In the epithelial dataset, the red and black modules enriched for ion transport functions and protein glycosylation respectively, while the orange and pink modules enriched for processes associated with the G1/S and G2/M cell cycle phases. The remaining purple module enriched

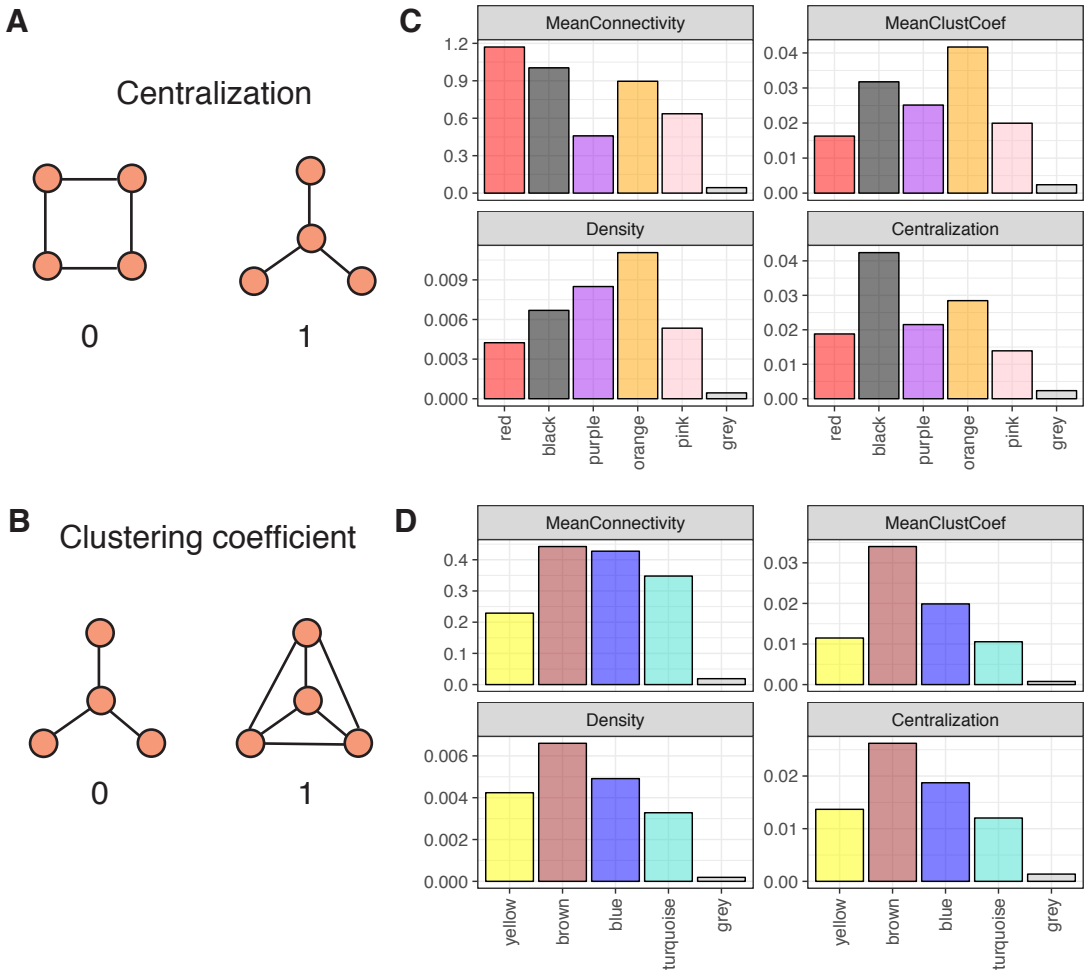


Figure 3.11: Fundamental network concepts in detected modules. (A,B) Graphs illustrating network concepts as labelled. (C,D) Network concepts calculated for modules detected in the epithelial (C) and mesenchymal (D) networks. The 'grey' module represents the 95th centile of 100 control modules each composed of a randomly permuted sample of 100 unassigned genes.

for a range of cellular metabolic functions including nucleotide synthesis, ribosome biogenesis and respiration. In the mesenchymal dataset, the blue module enriched for a number of terms related to the regulation of growth and differentiation, the brown module enriched for muscle contraction and its regulation, the turquoise module enriched for cellular stress related processes, and the yellow module enriched for extracellular matrix organisation and chemokine signalling.

The represented ontologies encompass a plausible cross-section of biological activities in their respective cell types. Interestingly for the epithelial dataset,

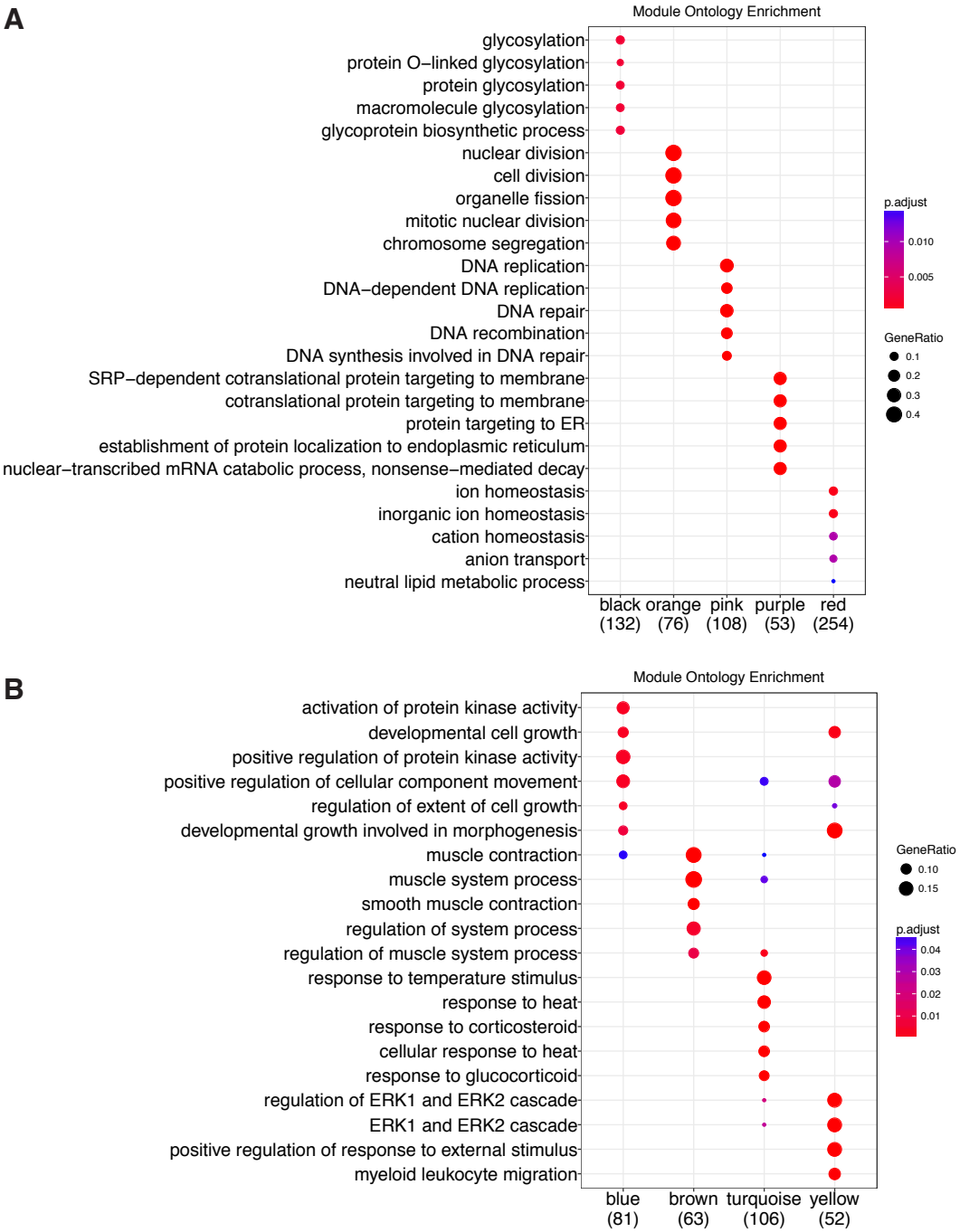


Figure 3.12: Ontology enrichment analysis in detected modules. Biological process ontology enrichment analysis was performed using a hypergeometric method for each of the detected modules in the epithelial (A) and mesenchymal (B) datasets (columns). The number of genes in each module is indicated (x-axis, parenthesised). The top 5 enriched ontology terms (by adjusted p-value) for each module are shown (rows). Circle size depicts the proportion of genes in the module annotated with the term, and color depicts the multiple-testing corrected p value.

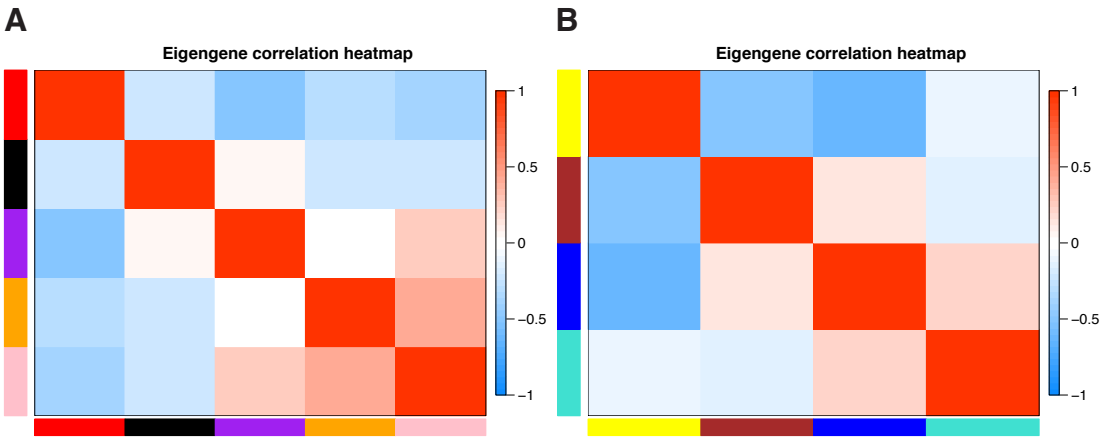


Figure 3.13: Module eigengene correlation. Heatmaps illustrating eigengene correlation for the detected modules (colour bars) in the epithelial (A) and mesenchymal (B) datasets. Off-diagonal tiles represent pairwise module eigengene correlation values.

where heterogeneity is better characterised, the modules might be expected to show differential expression at alternate stages of epithelial differentiation. For example, in rapidly dividing cells in the crypt base, cell cycle related processes might be expected to dominate, whereas at the crypt top differentiated enterocytes are specialised in ion transport functions and goblet cells rely on protein glycosylation in the production of mucins. To examine this further, eigengene correlation was calculated for the detected modules in both datasets (Figure 3.13). A module eigengene is mathematically the first principal component of the expression matrix for all module genes and can be used to quantify module expression as a whole on a cell-by-cell basis. Modules that are expressed in different cell subsets will show negative eigengene correlation. In the epithelial datasets, the red and black modules, which were interpreted as representing enterocyte and goblet cell processes, were negatively correlated with each other and with the remaining three modules of cell-cycle related genes. Strong negative correlations were also observed in the mesenchyme, particularly between the yellow, brown and blue modules.

3.2.9 Cell clustering based on module expression

Defining cell clusters using single-cell RNAseq data is a challenging problem, as evidenced by the multitude of disparate approaches that have been employed[222–225,

254, 287, 288]. A common initial step is the reduction of the dataset to a subset of genes considered to be most informative. This can be achieved in a supervised manner, using for example known marker-gene signatures[288], or objectively based on a statistical measure such as contribution to significant principal components or dispersion[224, 254, 287]. The former approach is restricted to cases where suitable annotation is available and is limited in its capacity to detect novel cell populations. The use of principal component analysis is complicated by the fact that technical factors such as cell library complexity often show strong correlation with high-ranked principal components[289], and the number of significant genes or components can be difficult to determine. The gene co-expression modules detected by WGCNA represent a reduced subset of genes that may be participating in coordinated biological processes. As such, they constitute an intuitively appealing basis for cell clustering.

Non-negative matrix factorisation (NMF) is a dimensionality reduction technique with an inherent clustering property that performs well with single-cell gene expression data[246, 290]. An initial expression matrix, containing all genes that had been assigned to modules, was factorised to basis and coefficient matrices. The choice of an optimal factorisation rank is discussed in Chapter 2, while surveys of rank estimation metrics for a range of candidate values are shown in Appendix A, Figure A.13. Cell-cell distances were calculated from the basis matrix, and hierarchical clustering was performed. The NMF method is non-deterministic, so a consensus clustering was derived from 50 randomly initiated runs (Figure 3.14). This revealed five cell clusters in the epithelium and four in the mesenchyme. Each module eigengene detected by WGCNA showed a positive correlation with one of the detected cell clusters, so the same colour identifiers were applied (Appendix A, Figure A.14). In addition to colour identifiers, the epithelial cell clusters were assigned labels based on module expression and inferred biological activity. Three clusters were labelled by cell cycle stage (G1, G1/S, G2/M), while the remaining two clusters were designated as enterocytes and secretory goblet cells (ENT, SEC).

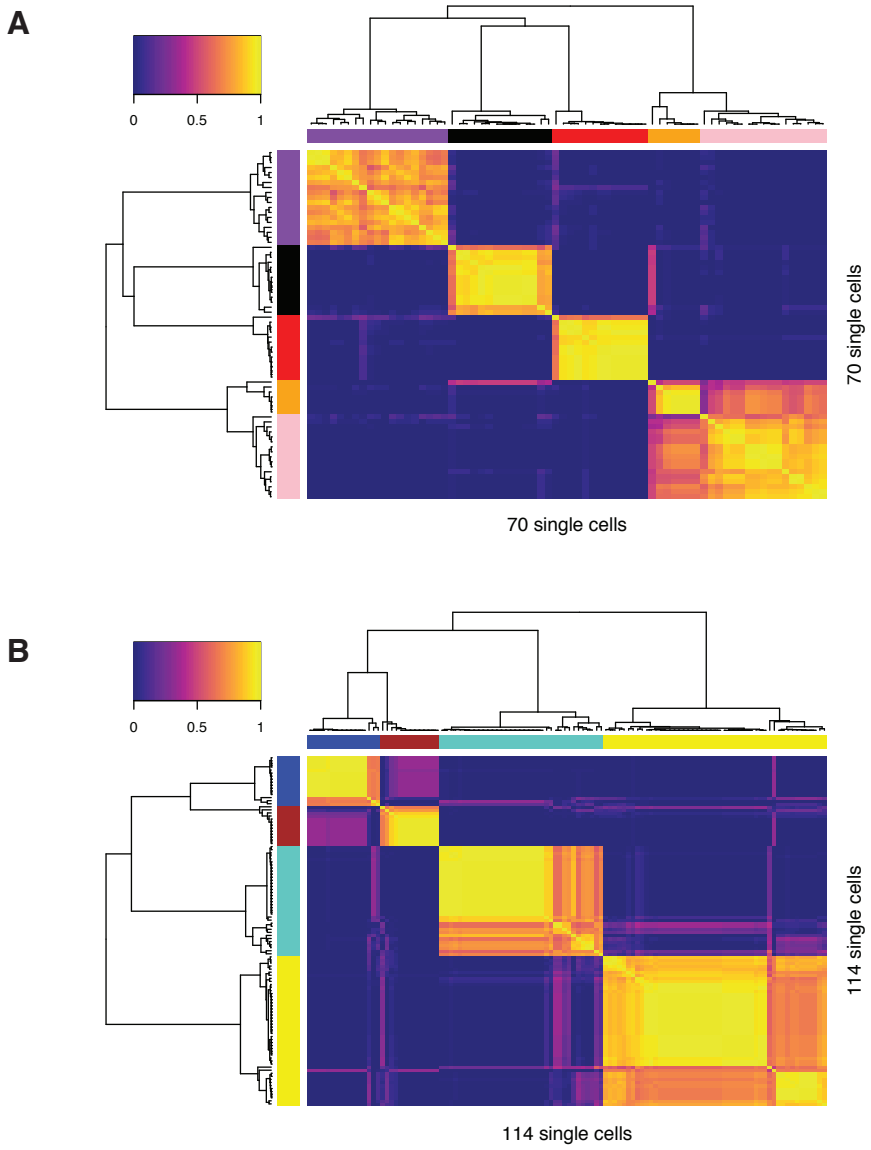


Figure 3.14: Consensus clustering of colonic single cells. Clustering of 70 epithelial single cells based on expression of 685 module genes (A) and 114 mesenchymal single cells based on 318 module genes (B). Heat colour depicts co-clustering probability over 50 randomly initiated clusterings, with values ranging from 1 (always in the same cluster) to 0 (never in the same cluster). Cell clusters and assigned the colour identifier of their highest-expressed gene module (colour bars).

3.2.10 Evaluation of clustering in epithelial dataset

To further explore the utility of the WCGNA and NMF annotation method in correctly segregating known cell subsets, we examined the distribution of known epithelial marker genes within the detected clusters using a panel of candidate makers identified from the literature and validated by single-cell qPCR [291]. Enterocyte and goblet cell markers showed strong enrichment within the expected cell clusters (Figure 3.15A, B). This, together with ontology enrichment compatible with the biological functions of these cell types, led us to conclude that the unsupervised approach had successfully identified these cells. Proliferative markers, such as Ki-67 (*MKI67*), showed enrichment within the G1/S and G2/M clusters, identifying these as the mid-crypt transit amplifying cells (Figure 3.15C). Progenitor marker genes (Figure 3.15D) showed enrichment in the G1/S cluster (*PTPRO*, *RGMB*) or no significant enrichment. The remaining G1 cluster may represent a mix of crypt base cells and post-mitotic cells at an early stage of differentiation.

Having defined cell types, candidate marker genes were identified and ranked by area under the receiver-operating characteristic (ROC) curve (see Chapter 2). Immunohistochemical staining of colonic tissue sections with antibodies to selected markers from the secretory cell cluster confirmed localization within goblet cells, including for genes with no previous known association with this cell type, such as *BCAS1*, a candidate oncogene of unknown function, and *ST6GALNAC1*, a sialic acid transferase (Figure 3.16). In summary, the unsupervised clustering approach outlined produces a structure consistent with the consensus model of colonic epithelial development and identifies novel subset-specific markers.

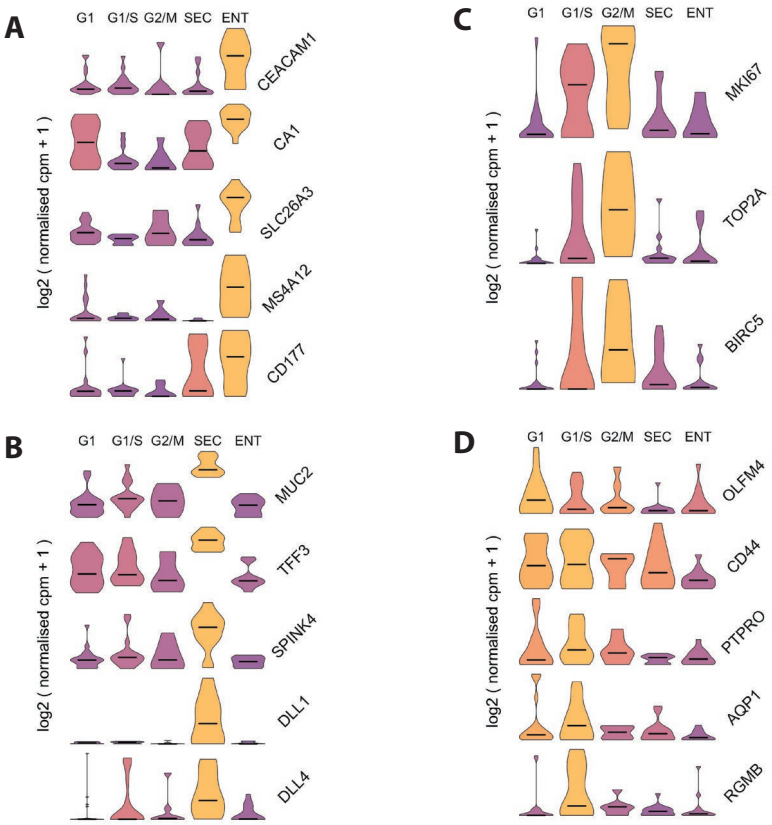


Figure 3.15: Epithelial marker gene expression in detected cell clusters. Expression distributions (violin plots) were examined for known marker genes associated with expected epithelial cell populations within a colonic crypt[291]. Marker gene families for enterocytes, goblet cells, proliferating cells, and progenitors are shown in panels A-D respectively. Cell clusters identified by the unsupervised approach are faceted into columns (G1, G1/S, G2/M, SECretory, ENTerocyte), while row facets are individual markers. Log-transformed normalised read counts per million (cpm) are shown on the y-axis, with a crossbar at the median. Local cell density over the range of expression values is represented by violin width, while the violin colour represents the row-scaled mean expression level for each cluster.

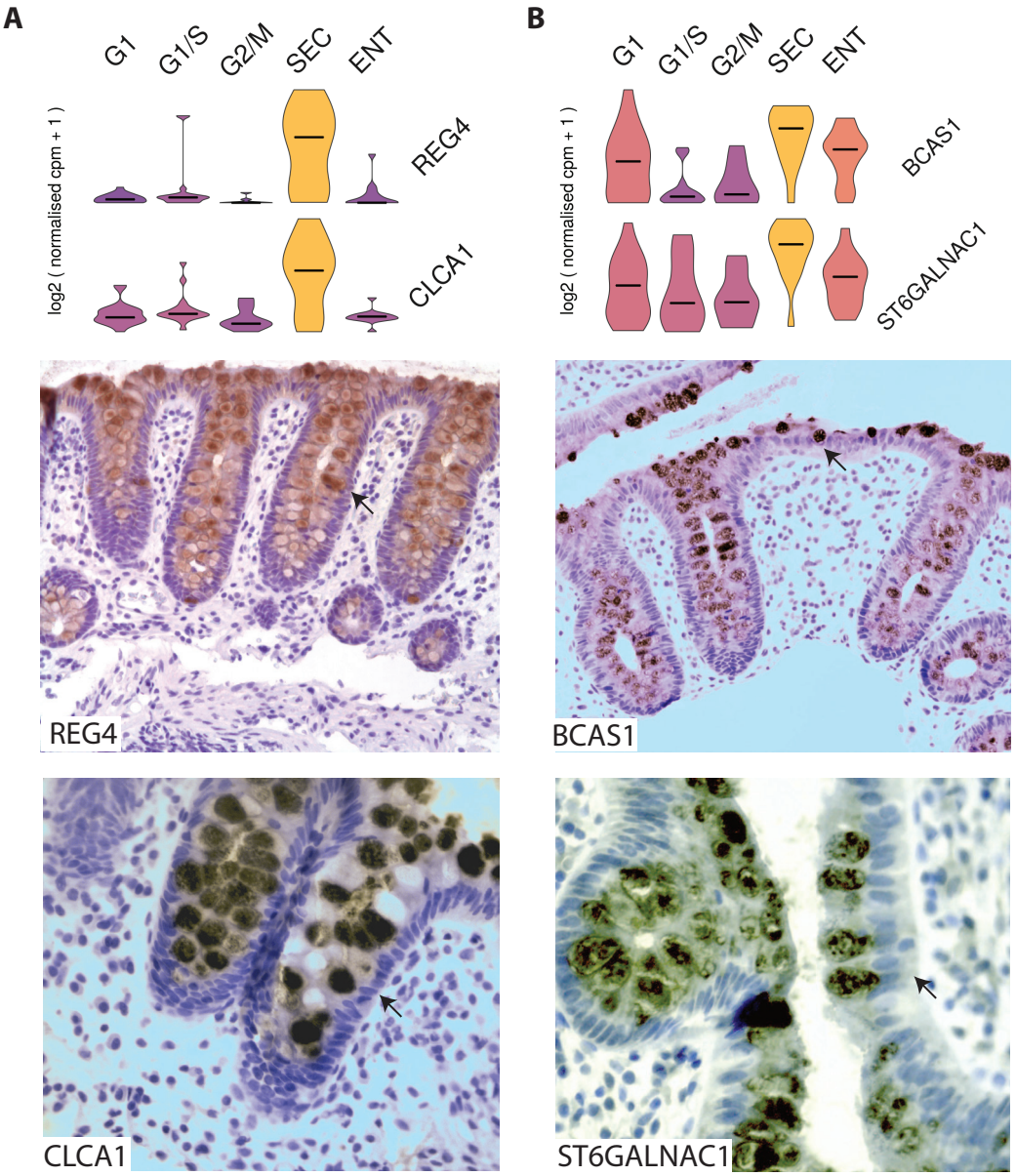


Figure 3.16: Validation of known and novel goblet cell markers by immunohistochemistry. Selected markers from the secretory cluster were validated by immunohistochemistry in normal human colonic tissue, including both (A) known goblet cell markers (REG4, CLCA1), and (B) novel markers predicted by the transcriptomic model (BCAS1, ST6GALNAC1). Positive signal (brown) was seen in goblet cells (arrows). Nuclear counterstaining was performed with haematoxylin (blue).

3.2.11 Comparison with other clustering methods

A comparison was made between the cell cluster assignment made by WGCNA and NMF and other clustering methods for single cell transcriptomic data. The package 'pcaReduce' utilises an iterative approach combining K-means clustering and reduced-dimension representations of the gene expression data generated by PCA to produce a cell clustering structure[292]. The 'SC3' package combines spectral transformations, K-means clustering, and complete-linkage hierarchical clustering[248]. Using pcaReduce, no stable clustering structure was found at any value of k in the epithelial dataset. The SC3 method, however, determined the optimal cluster number to be 3 and produced a clustering solution that was compatible with the structure generated using WGCNA and NMF albeit with the three cell-cycle related clusters (G1, G1/S, G2M) combined (Figure 3.17). Given that the effect of cell-cycle stage on overall transcriptional activity is large[293], and that multiple cell-cycle stages cannot be superposed in the same cell, the refined structure generated by WGCNA and NMF offers a biologically plausible solution to the problem of identifying meaningful cell states from primary human intestinal single-cell RNA-seq data.

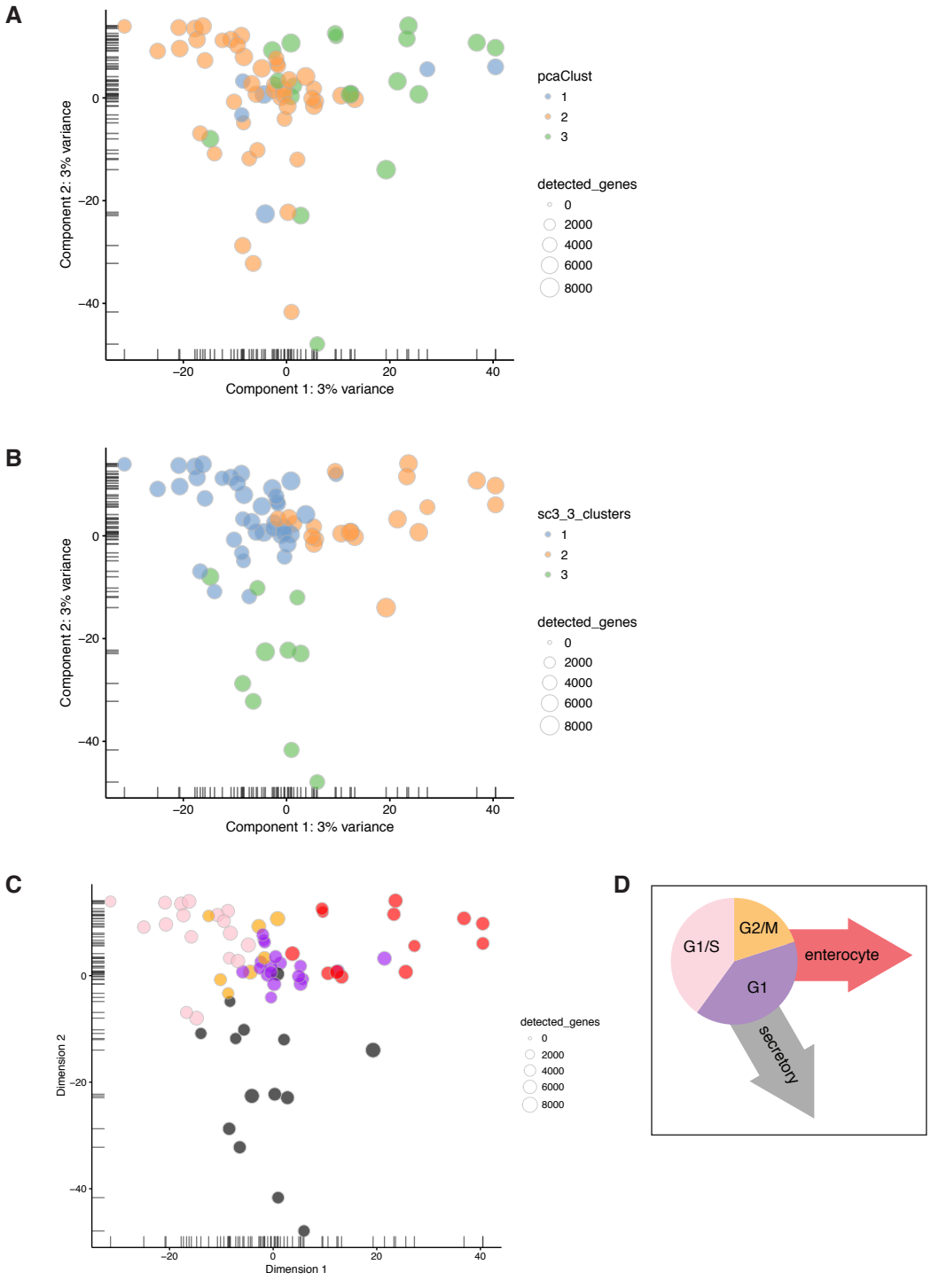


Figure 3.17: Comparison of clustering methods in the epithelial dataset. Projections of cells in the first two principle components based on expression of 6469 biologically variable protein-coding genes. Individual cells are coloured by cluster assignment in (A) pcaReduce, using an arbitrary k value of 3 for this method as no stable structure was found at any k, (B) SC3, and (C) WGCNA and NMF. A biological interpretation of the clusters is shown (D).

3.3 Discussion

This chapter describes an approach to the isolation, sequencing and analysis of single cells derived from human endoscopic forceps biopsies. Several methods have been proposed for the isolation of these cells but there is no accepted consensus. The optimised protocols presented here are both rapid and avoid positive marker-based selection so that unbiased cross-sections of the target cell populations are obtained. Furthermore, validation by FACS analysis and the generation of highly distinct and biologically plausible single cell transcriptomes suggest that they are specific in enriching for cells originating in the epithelial and mesenchymal compartments respectively. The detection of epithelial cells across the spectrum of stages in their differentiation hierarchy suggests that the transcriptomes generated are an accurate reflection of the native transcriptional diversity. Speed of isolation is critical to this as differentiated enterocytes rapidly undergo anoikis when detached from the basement membrane[294].

The need to move rapidly from cell isolation to RNA capture necessarily compromises some aspects of the experimental design. In particular, the need to process biological replicates as separate batches causes the biological variability between donors to be 100% confounded by technical batch effects, which are a significant problem in single-cell RNAseq in general and clearly present in these datasets (see Appendix A Figure A.11). An optimal design in this regard would be to divide biological replicates across sequencing batches[295]. The true technical batch effects could then be accurately quantified and adjusted. However, this would require both processing biological replicates in parallel, which may not be feasible for rare disease states, and the simultaneous use of multiple C1 platforms, which were not available at the time these experiments were performed. A promising alternative would be to fix cells after isolation and recover RNA for sequencing at a later date. While fixative protocols for single cell RNAseq exist[285], pilot experiments with these cell suspensions resulted in a marked decline in library quality[296]. The work-around presented here, adjusting batch effect based on

an *in silico* defined control gene set, produced interpretable results in the context of a sub-optimal design in this regard.

The expression profiles obtained provided insight into the immunological functions of epithelial and mesenchymal cell types. The IBD-linked genes *ADAMDEC1* and *IL7* were detected in the stromal compartment, having not previously been associated with this anatomical niche. Evidence of epithelial-mesenchymal crosstalk mediated by the growth factor neuregulin 1 together with the cytokines IL-15 and IL-18, was also of interest given that system-wide perturbation of these paracrine pathways modulates the colitis phenotype in animal models. In addition, the global enrichment seen in cellular pathways for phagocytosis, innate sensing and antigen presentation in stromal cells corroborates *in vitro* functional assays suggesting that these cells are capable of detecting perturbations in the local tissue environment and mobilising an immune response, and perhaps positions them as the principle local immune rheostats.

Also presented is an unsupervised clustering approach that recapitulates the known differentiation hierarchy in the epithelial compartment and predicts novel subset specific markers. In this challenging setting, with prominent batch effect and the absence of a uniform genetic background between replicates, it compares favourably to existing methods and provides a basis for further exploration of the heterogeneity within the stromal compartment.

Chapter 4

Intestinal stromal cell transcriptional subtypes in health and Inflammatory Bowel Disease

Contents

4.1 Introduction 92
4.1.1 Motivation 93
4.2 Results 93
4.2.1 WGCNA-based clustering captures an important source of heterogeneity in human colonic stromal cells 93
4.2.2 Characterization of human colonic stromal subtypes . . 95
4.2.3 Single-cell transcriptomes of UC-derived colonic stromal cells 104
4.2.4 Altered global expression of epithelial morphogens in UC stromal cells 105
4.2.5 An additional gene co-expression module and cell cluster is detected in the stroma in acute UC 107
4.2.6 Validation of the UC-specific Stromal 4 subset 111
4.2.7 Inflammation-associated changes in stromal composition in UC 113
4.2.8 Single-cell transcriptomes of colonic stromal cells from patients with UC-associated dysplasia 115
4.2.9 Transcriptional changes induced by TNF-alpha stimulation of colonic stromal cells 119
4.3 Discussion 121

4.1 Introduction

That there is heterogeneity among lamina propria fibroblast-like cells is clear both from ultrastructural studies and expression of marker proteins such as α-SMA[9]. This has given rise to a broad division of these cells into myofibroblasts and fibroblasts which, while undoubtedly helpful, does not fully address or align with the diverse functions attributed to this cellular compartment. Fibroblast-like cells are implicated in the maintenance of both the epithelial and mucosal immune cell niches in health, and are central players in the adaptive repair response to mucosal injury[12]. In animal models of IBD, lamina propria fibroblasts isolated prior to disease onset mediate extracellular matrix remodelling and promote adaptive immune responses suggesting a role in disease initiation[193]. They also contribute to a local axis of TNFα signalling that has full pathogenic potential even in the absence of TNFα signalling in the haematopoietic compartment[192]. It is unclear whether all fibroblast-like cells are equally capable of mediating these activities or whether a process of cell fate specification and commitment produces distinct functionally specialised sub-populations. Specific markers that might be used to generate genetic models capable of refining the contribution of fibroblast subsets to mucosal homeostasis and disease are also lacking.

The study of the stromal biology of lymphoid tissues has provided abundant evidence that fibroblasts are capable of functional specialisation and commitment. For example, immunologically specialised stromal cells are critical in initiating the lymph node anlage[297], recruiting and maintaining lymphocyte subsets[298], maintaining follicle and germinal centre boundaries[299], presenting antigen[169], and facilitating antibody production[300]. Multiple functionally distinct stromal subsets have been characterised directing these activities, and specific marker genes and key transcriptional regulators identified[42].

Where bulk genome-wide assays have been used to classify stromal fibroblasts from non-lymphoid tissues, these have revealed wide-ranging and persistent differences in gene expression correlating with the tissue of origin[44, 45]. However,

this approach has been limited by the need to first expand fibroblast populations derived from primary samples *in vitro* to generate sufficient starting material. Not all lymphoid fibroblast subtypes are equally amenable to cell culture[168], and it remains unclear to what extent growth in culture distorts their native gene expression profiles. The latter problem is of particular concern for fibroblast-like cells given their known transcriptional plasticity[301], and the dominant role of the local tissue microenvironment in shaping their transcription responses.

4.1.1 Motivation

Having obtained plausible transcriptional profiles from human primary colonic epithelial and mesenchymal cells and developed a bioinformatic approach using WGCNA and NMF that successfully identified known cell subtypes in the epithelial compartment, I next sought to define the transcriptional heterogeneity in the lamina propria stromal compartment at single-cell resolution and develop a classification system for colonic stromal cells based on a global perspective of their biological activities. This could then be extended using the same experimental approach to define changes in stromal biology occurring in important disease states such as UC and pre-malignant UC-associated dysplasia.

4.2 Results

4.2.1 WGCNA-based clustering captures an important source of heterogeneity in human colonic stromal cells

The use of WGCNA in a dataset consisting of transcriptional profiles of 114 single fibroblast-like cells derived from distal colonic biopsies in 3 healthy donors revealed 4 gene co-expression modules (Chapter 3, Figure 3.10). In contrast to the epithelial dataset, cell-cycle annotation was not enriched, consistent with a view of fibroblasts as a relatively long-lived, stable cell population in the steady state[66]. However, multiple ontology terms relevant to fibroblast and myofibroblast biology did show significant enrichment including extracellular matrix organisation and regulation

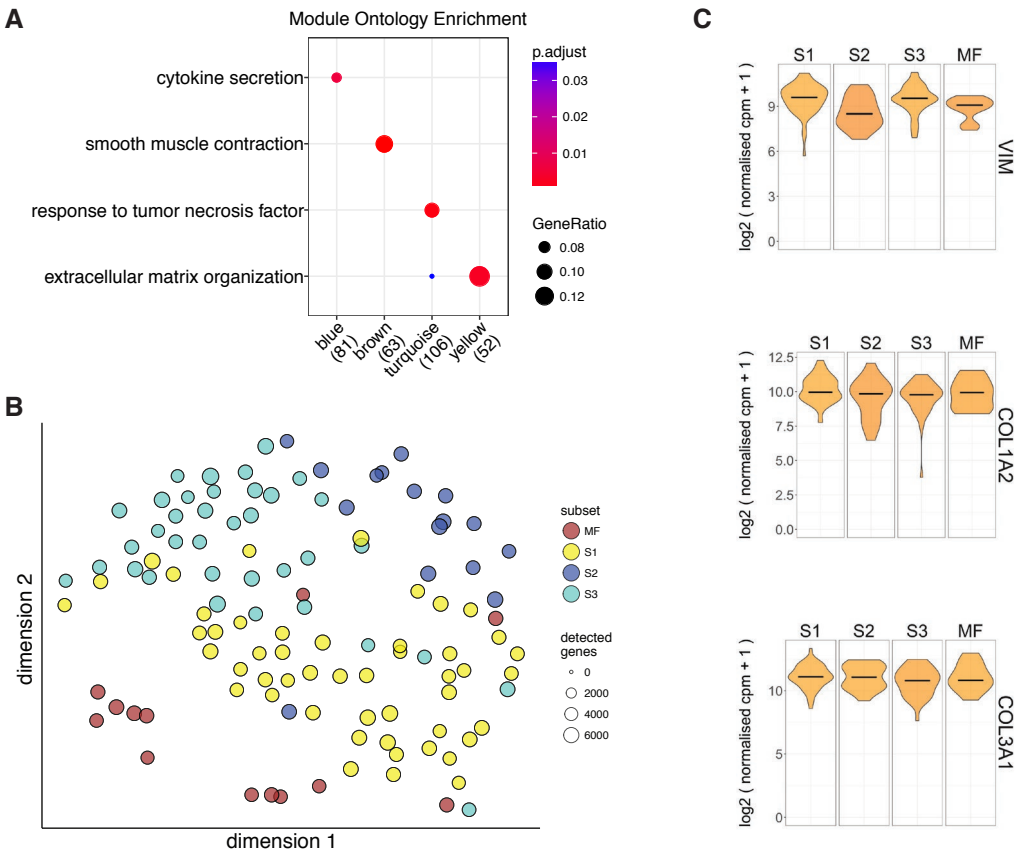


Figure 4.1: Four distinct fibroblast subsets are identified in healthy colonic samples. (A) Selected enriched ontology terms in detected modules (columns). The background number of genes in each module (x-axis, parenthesised), the proportion of annotated genes per term (circle size) and the adjusted p-value (circle colour) are indicated. (B) 2D visualisation of the detected cell clusters using t-SNE. Individual points correspond to single cells and are coloured by cluster assignment (S1–3, Stromal subsets 1–3; MF, myofibroblast) and sized according to the number of detected genes. (C) Violin plots for the pan-fibroblast marker genes vimentin (*VIM*) and collagen types 1 and 3 (*COL1A2*, *COL3A1*) across clusters detected.

of adhesion (yellow), smooth muscle contraction (brown), and cytokine secretion (blue), and the responses to TNFα (turquoise) (Figure 4.1A). Cell clustering by NMF using modular genes produced four stable clusters, which were assigned colour identifiers related to the highest expressed module (Chapter 3, Figure 3.14B, Appendix A, Figure A.14B). Although only 318 genes were identified as showing modular behaviour, two-dimensional t-SNE embedding incorporating all 7,000 biologically variable genes suggested that the approach had captured a dominant source of variability within the dataset (Figure 4.1B).

Candidate marker genes for the four mesenchymal cell clusters were ranked by area under the ROC curve, and biological process ontology enrichment was again performed for the top 200 marker genes within each subset using a hypergeometric test with adjustment for multiple testing (see Chapter 2). All four clusters show high and similar expression of pan-fibroblast markers such as the intermediate fibre vimentin and collagen types 1 and 3 (*VIM*, *COL1A2*, *COL3A1*) (Figure 4.1C). In summary single-cell RNA-seq identified four colonic stromal subtypes in health, possessing equivalent expression of pan-fibroblast markers but exhibiting distinct biological process ontology enrichment suggestive of distinct cell types.

4.2.2 Characterization of human colonic stromal subtypes

Further characterization the newly identified stromal subtypes was then undertaken. We termed one of the clusters myofibroblasts (MF). This enriched for processes relating to muscle contraction and its regulation and showed high expression of genes for contractile machinery such as myosin heavy chain 11 (*MYH11*) and enteric smooth muscle actin (*ACTG2*) (Figure 4.2B). A recently described specific colonic myofibroblast marker, *AOC3*[33], localized to this group (Appendix B, Figure B.1). Alpha smooth muscle actin (*ACTA2*), another marker that has been widely used to identify intestinal myofibroblasts, shows highest expression within the MF subset, although intermediate expression is also noted in an additional cell cluster (Stromal 2, S2). The MF subset expressed increased levels of the transcriptional regulators *MYCOD*, *RCOR1*, *TRERF1*, *SOX5*, and *STAT5B* and effector molecules such as *TGFBR3* (Figure 4.2C). *TGFBR3* is a SMAD1-induced gene associated with development of myofibroblasts from fibroblasts, whose inhibition aggravates fibrosis[302]. None of the cells sequenced showed expression of the abundant smooth muscle marker gene desmin (*DES*).

The remaining three cell clusters, expressing fibroblast-associated but not contractile genes were designated Stromal 1, 2 and 3 (S1–3). These were dissimilar both in their transcriptional profiles and ontology enrichment, suggesting distinct biological roles. Stromal subset 1 gene ontology enrichment terms include ‘extracellular matrix

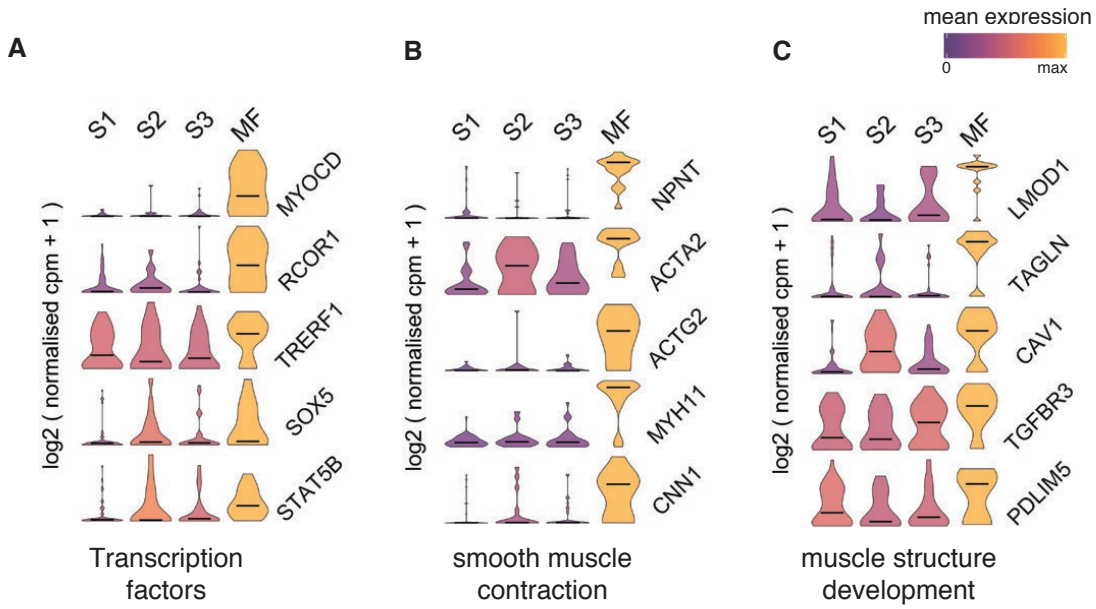


Figure 4.2: Transcription factor and ontology enrichment in the myofibroblast (MF) subset. (A) Violin plots for high-ranked transcriptional regulators (GO annotation 'Transcription factor — sequence-specific DNA binding'). (B, C) Violin plots for marker genes sharing ontology annotation for significantly enriched terms as indicated. Crossbar indicates median expression level.

organization', 'angiogenesis', 'cellular response to cytokine stimulus'. Transcriptional regulators enriched in this group include *NDN*, *SMAD1*, *GAS7*, *HCFC1*, *ARID1A* (Figure 4.3A). Immunological features of stromal group 1 include expression of the chemokines *CXCL12*, *CXCL1*, *CCL11*, *CCL13*, chemotactic for lymphocytes, neutrophils, eosinophils, and monocytes respectively[303] (Figure 4.3B). The localization of professional immune cells within the lamina propria is dependent on migration along chemokine gradients and Stromal 1 appears poised to direct this function. Stromal 1 cells also express *TNFSF13B* or B cell activating factor (BAFF). Inadequate level of BAFF will fail to activate B cells to produce immunoglobulin leading to immunodeficiency. Excessive level of BAFF causes abnormally high antibody production, results in systemic lupus erythematosus, rheumatoid arthritis, and many other autoimmune diseases[304].

Both Stromal subsets 1 and 2 show enrichment for genes annotated with ontology terms related to organisation of the extracellular matrix, a central fibroblast function

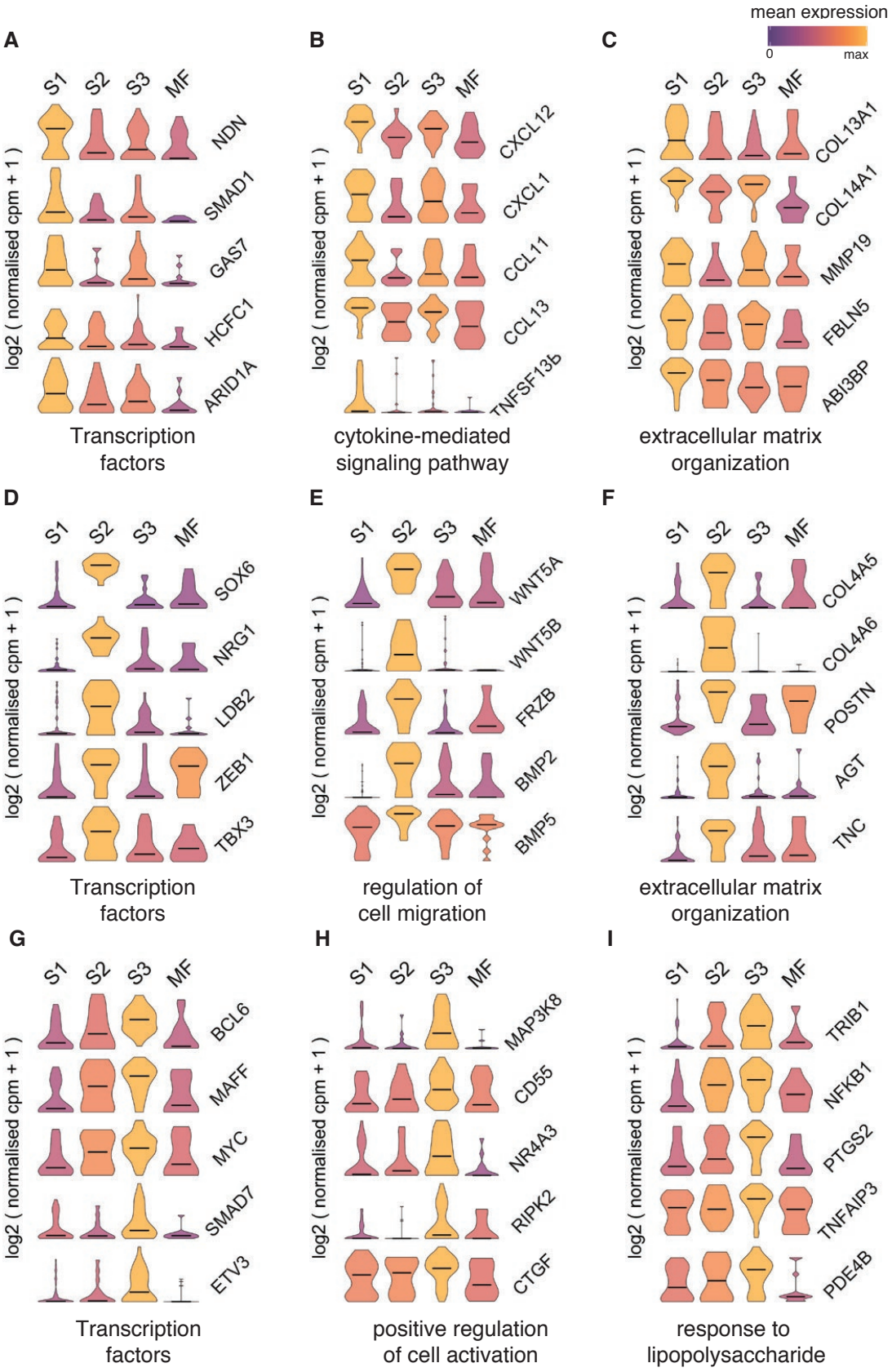


Figure 4.3: Transcription factor and ontology enrichment in stromal 1–3.

Figure 4.3 ***(previous page)*:** Violin plots for high-ranked marker genes in the Stromal 1 (A, B, C), Stromal 2 (D, E, F), and Stromal 3 (G, H, I) clusters. Transcriptional regulators (GO annotation 'Transcription factor — sequence-specific DNA binding') and genes annotated with the indicated enriched ontology terms are shown. Crossbars indicate median expression.

(Figure 4.3C, 4.3F, Appendix B Figure B.1). A key difference, however, is that the stromal 1 subset enriches for components of fibrillar collagen (*COL14A1*, *COL15A1*) and elastic fibres (*FBLN1*, *FBLN2*, *FBLN5*, *EFEMP1*, *FN1*), while the stromal 2 subset shows specific expression of sheet collagens (*COL4A5*, *COL4A6*) that are key constituents of the epithelial basement membrane. This is compatible with ultra-structural descriptions of the lamina propria showing a layered architecture, with spindle-shaped cells embedded in a dense reticular sheet adjacent to the basement membrane, and stellate interconnected cells forming a sieve-like network below[12]. In addition to generating the structural components of the extracellular matrix, stromal cells produce signaling molecules that influence its biological activities. Here again, differences between subsets were observed, with, for example, Stromal subset 2 expressing high levels of tenascin C and periostin (*TNC*, *POSTN*, Figure 4.3F), which have previously been shown to promote tumorigenesis[305, 306].

Stromal subset 2, which expressed high levels of transcriptional regulators *SOX6*, *NRG1*, *LDB2*, *ZEB1* and *TBX3* (Figure 4.3D), was also distinguished by high expression of a number of hedgehog-induced morphogens. These included TGF-beta superfamily ligands *BMP2* and *BMP5*, the non-canonical wnt ligands *WNT5A* and *WNT5B*, and the secreted wnt antagonist *FRZB* (Figure 4.3E). Hedgehog signals from differentiated enterocytes at the luminal surface induce stromal BMP production that, in turn, regulates the size of the epithelial proliferative compartment[81]. The combined effects of the morphogens expressed by Stromal 2 on epithelial development are unknown, however they have individually been shown to restrict epithelial proliferation *in vitro*[307, 308]. In addition, it has been demonstrated that WNT5A is essential for epithelial reconstitution after injury via a mechanism that involves potentiation of TGF-β signaling[309]. The platelet derived

growth factor receptor alpha (*PDGFRA*) also shows enrichment in the stromal 2 group and has previously been identified as a candidate marker for a distinct subset of intestinal stromal cells[31] (Appendix B Figure B.1). Stromal group 2 is enriched for genes directing epithelial proliferation, differentiation and homeostasis.

Stromal 3 is characterized by expression of transcriptional regulators *BCL6*, *MAFF*, *MYC*, *SMAD7* and *ETV3* (Figure 4.3G) and ontology enrichment for terms including 'response to molecule of bacterial origin', 'pattern recognition receptor signaling pathway', and 'response to interleukin-1'. Both *MAP3K8* and *PTGS2* show increased expression within the Stromal 3 subset (Figure 4.3H, 4.3I). Fibroblasts play a critical homeostatic role in the epithelial response to injury via a mechanism that is dependent on sensing the inflammatory milieu and bacterial components, activation of a molecular pathway involving *MAP3K8* and *PTGS2*, and synthesis of prostaglandin E_2 (PGE_2)[34, 35]. Stromal 3 likely plays a role in this function. PGE_2 has been shown to inhibit fibroblast to myofibroblast differentiation and this, together with increased expression of the senescence-inhibiting transcription factors *MYC* and *BCL6*, would be compatible with progenitor potential within the Stromal 3 group.

I next sought to define the tissue distribution of the stromal subsets predicted by the single cell model using immunohistochemistry (IHC) and single-molecule in-situ hybridization (sm-ISH) for selected marker genes in colonic tissue sections. Expression of stromal 1 markers (*ADAMDEC1*, *DCN*, *SLIT2*, *CXCL12*) could be detected by sm-ISH in mesenchymal cells distributed throughout the lamina propria, while stromal 2 markers (*F3*, *HSD17B2*, *POSTN*, *FRZB*, *WNT5A*, *WNT5B*, *BMP2*, *BMP5*) were restricted to a smaller sub-population in close approximation to the epithelial monolayer (Figure 4.4). Similar distinct staining patterns were observed at protein level by IHC comparing the stromal 1 markers ZEB2 and ABI3BP to the stromal 2 markers ZEB1 and SOX6 (Figure 4.5A, 4.5B).

To further validate the existence of distinct subsets of stromal cells based on data derived from the single-cell RNAseq analysis, co-staining was performed with

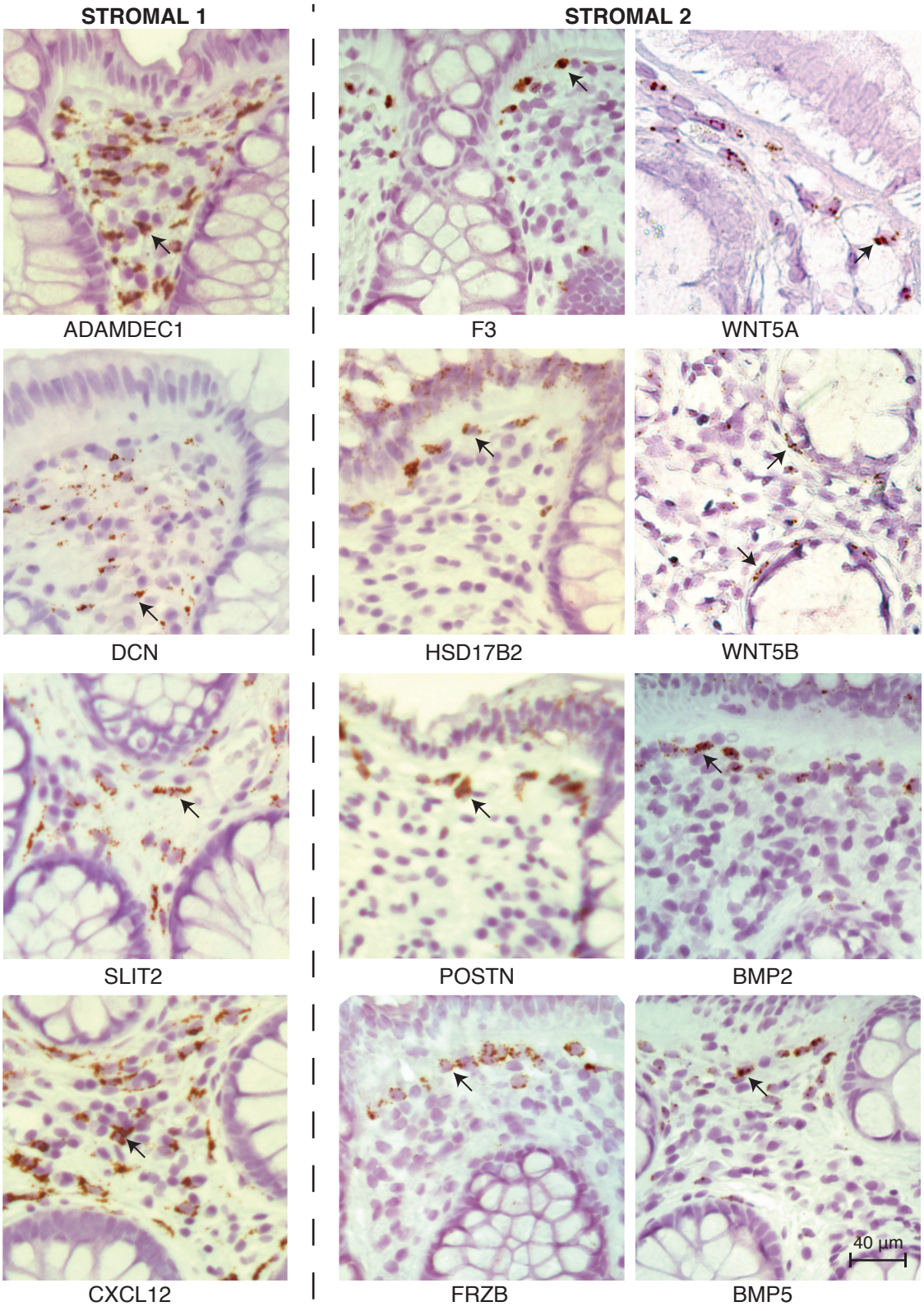


Figure 4.4: Tissue distribution of detected stromal subsets by sm-ISH.

Figure 4.4 ***(previous page)*:** Single-molecule ISH labelling using probes directed to the indicated Stromal 1 (left panel) and Stromal 2 (right panel) marker genes in healthy human colonic mucosal tissue sections counter-stained with haematoxylin (blue). Positive staining cells (brown) are indicated (arrows). Positive and negative ISH technical controls are shown in Appendix B, Figure B.2.

antibodies to transcription factors linked to either Stromal 1 or Stromal 2. Co-localization was observed as predicted when the stromal 2 markers *ZEB1* and *LEF1* were co-stained, but not when *ZEB1* was co-stained with the stromal 1 marker *ZEB2* (Figure 4.5C). Quadruple staining using candidate markers for the stromal 1, 2 and 3 subsets (ZEB2, F3, SMAD7 respectively) and the pan-leucocyte marker CD45 was also performed, confirming the stromal specificity of these markers and their contrasting distributions.

Separate $SOX6^{-}ZEB2^{+}$ stromal 1 and $SOX6^{+}ZEB2^{-}$ stromal 2 populations were seen when dissociated fibroblasts from healthy human colonic tissue were subjected to flow cytometric analysis (Figure 4.7).

Together, this data provides evidence for the existence of distinct colonic stromal cell states with specific functional properties that can be identified by dissimilar marker gene expression and anatomical location within the lamina propria.

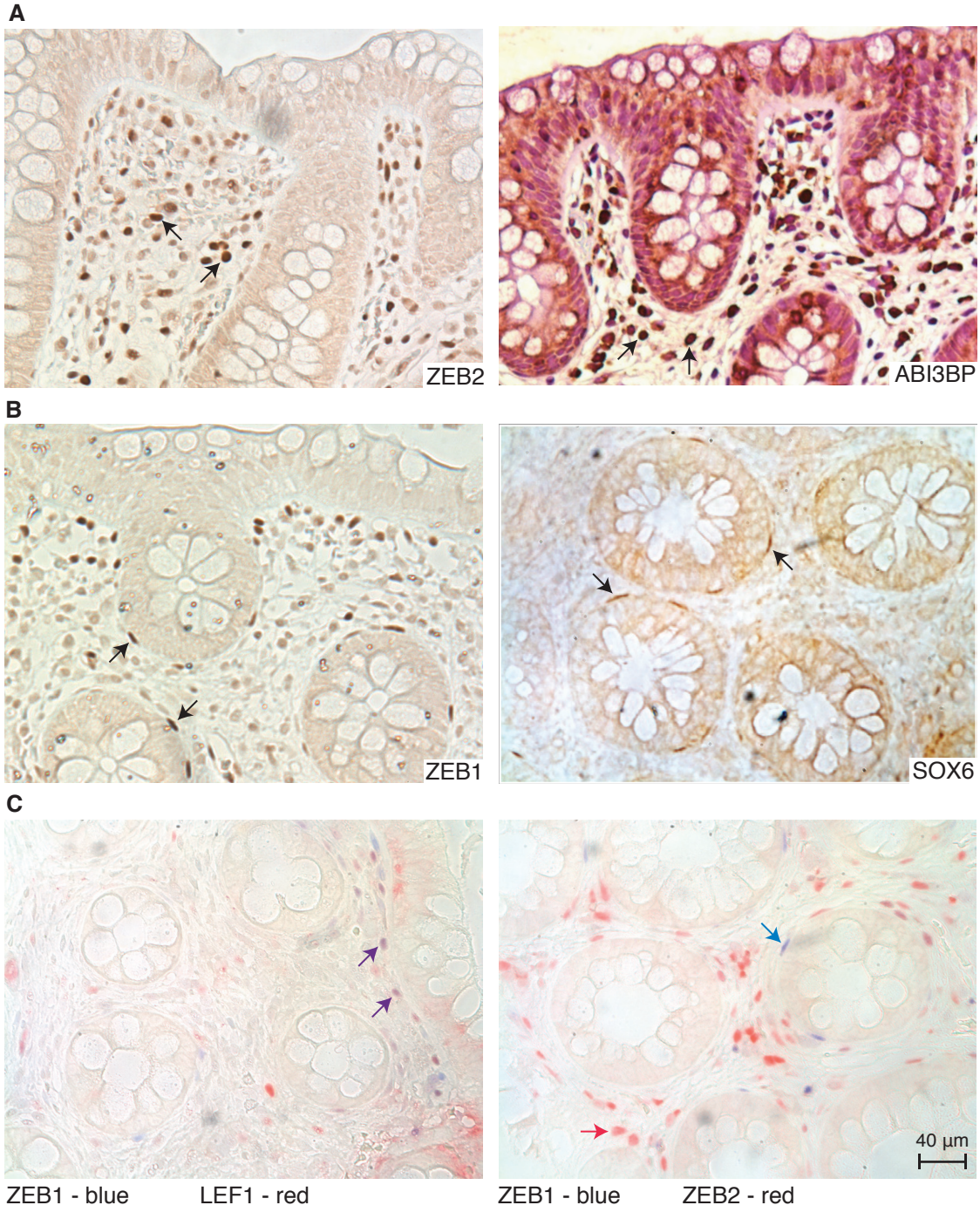


Figure 4.5: Tissue distribution of detected stromal subsets by IHC. (A,B) Healthy human colonic tissue sections stained with antibodies to the indicated markers for the Stromal 1 (A) and Stromal 2 (B) subsets. Positive staining cells (brown) are indicated (arrows). Counter-staining with haematoxylin and eosin was performed in the case of ABI3BP only. (C) Co-staining of ZEB1 and LEF1 (both stromal 2 markers, left panel) and ZEB1 and ZEB2 (stromal 2 and stromal 1, right panel). Arrows indicate positive-staining cells (ZEB1 — red, ZEB2 — blue, ZEB1 / LEF1 co-staining — purple).

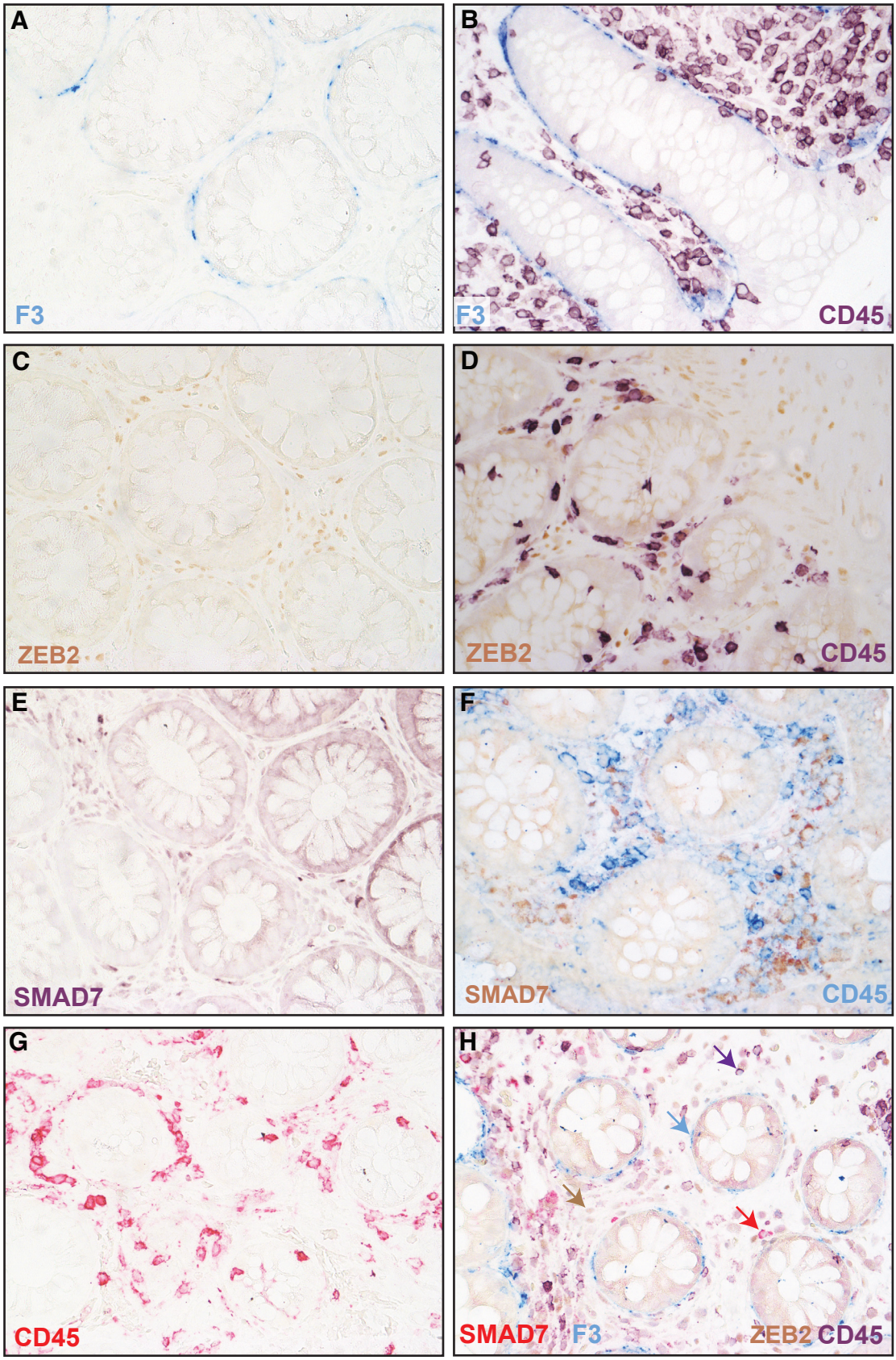


Figure 4.6: CD45⁻ stromal subsets are anatomically separated.

Figure 4.6 ***(previous page)*****:** Healthy human colonic tissue sections stained with antibodies to stromal subset markers. Co-stainings with the pan-leucocyte marker CD45 were also performed (right column). (A,B) The stromal 2 marker F3 (light blue). (C,D) The stromal 1 marker ZEB2 (tan). (E,F) The stromal 3 marker SMAD7 (left panel, purple; right panel, tan). (G) CD45 single stain (red). (H) Quadruple stain (SMAD7 red, F3 light blue, ZEB2 tan, CD45 purple). Arrows indicate positively staining cells.

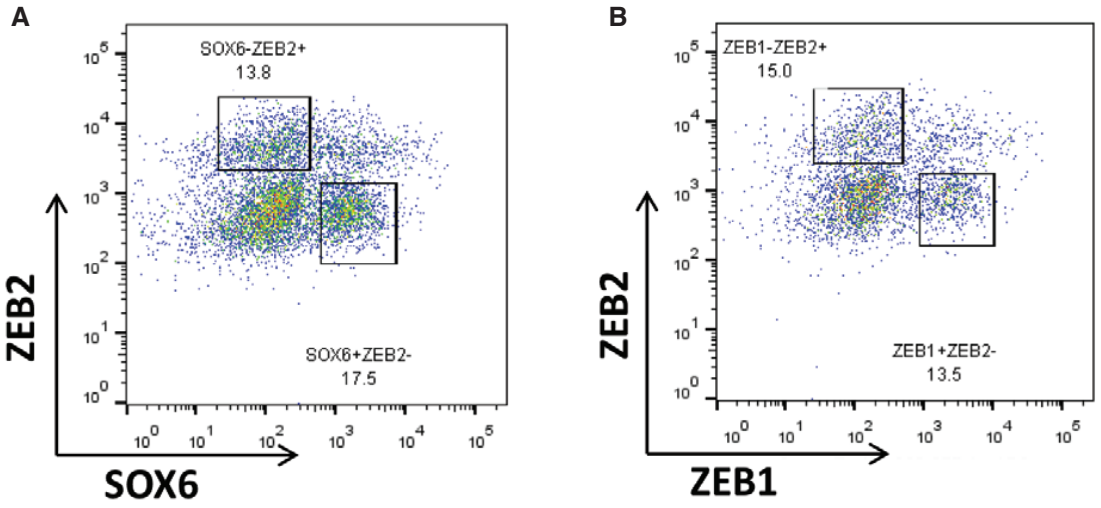


Figure 4.7: Distinct fibroblast sub-populations identified by FACS. (A) Identification of SOX6$^-$ZEB2$^+$ stromal 1 and SOX6$^+$ZEB2$^-$ stromal 2 subsets in healthy human colonic tissue by flow cytometry. (B) Identification of ZEB1$^-$ZEB2$^+$ stromal 1 and ZEB1$^+$ZEB2$^-$ stromal 2 subsets in healthy human colonic tissue by FACS. Representative gating strategies are shown in Appendix B Figure B.3.

4.2.3 Single-cell transcriptomes of UC-derived colonic stromal cells

The changes in stromal composition and activity associated with IBD remain poorly understood. Having obtained a baseline cross-section of stromal transcriptional states in health, I next applied the same method to samples derived from patients with active UC, whose phenotype is of chronic inflammation extending proximally from the rectum. The whole colon may be involved but limited disease characterised by an abrupt transition between inflamed and healthy mucosa usually in the left colon is more common. Colonic biopsies were obtained from the inflamed and proximal non-inflamed mucosa of patients with acute limited UC. To avoid effects of drug treatment on the molecular signatures observed, patients were selected on the basis of being early diagnosis and off immunotherapies at the time of sampling

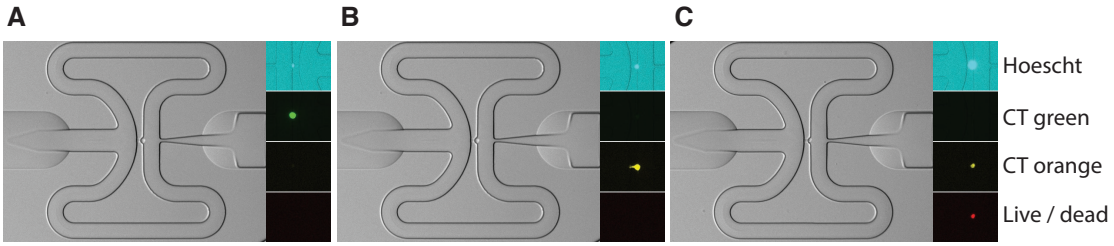


Figure 4.8: Sample multiplexing on Fluidigm C1 IFCs. (A) Fluidigm C1 capture site imaging with cell labelling by site of origin. Biopsies from inflamed and non-inflamed sites were processed separately and the resultant cell suspensions were labelled with either (A) CellTracker Green (CT green) or (B) CellTracker Orange (CT orange). Non viable cells were identified using Live/Dead Far Red (C).

(Appendix B Table B.1). Clinical Biopsies were processed to enrich for stromal cells as previously prior to single-cell RNAseq analysis. 127 single cells from 3 donors with UC were sequenced. Both inflamed and non-inflamed sites were sampled and a cell staining and on-chip imaging protocol was employed so that single cell libraries could be annotated with the endoscopic and histologic grading of inflammatory activity at their site of origin (see Chapter 2, Figure 4.8). Following quality control and removal of non-fibroblast cell types, 76 single cells remained in the analysis.

4.2.4 Altered global expression of epithelial morphogens in UC stromal cells

An initial differential expression analysis was performed comparing mesenchymal cells from UC samples to the analogous cell population from healthy controls. As before, the DESeq2 and SCDE methods were applied independently and the results compared. The overall number of genes showing significant differential expression was lower than in the epithelial to mesenchymal comparison as expected, and the SCDE method identified fewer targets than DESeq2 (Figure 4.9). A total of 75 genes were identified as differentially expressed by both methods. Of these 70 showed reduced expression in UC versus 5 that were upregulated. Library complexity was lower in the UC dataset than in the healthy experiment (median 2,974 versus 4,924 unique genes detected per cell) and this may have contributed to the imbalance. The highest magnitude upregulated gene was *GREM1*. This encodes the BMP antagonist gremlin 1, which regulates the size of the epithelial stem

cell compartment and is expressed only in a small sub-population of stromal cells located in the muscularis propria adjacent to the crypt bases in health[310]. Another upregulated target was the extracellular matrix component biglycan (*BGN*). Like gremlin 1, biglycan antagonises BMP signalling but also signals as a DAMP via TLR4 and TLR2 to induce a pro-inflammatory response in myeloid cells[311, 312]. The top down-regulated gene was dipeptidyl peptidase 4 (*DPP4*), an enzyme that degrades glucagon-like peptides 1 and 2 (GLP1 and 2) and an established therapeutic target in Type 2 diabetes. In the gut, GLP-2 secreted by enteroendocrine cells supports epithelial proliferation, and inhibition of DPP4 ameliorates disease severity in murine colitis models[313]. The trans-membrane C-type lectin CD69 was also significantly down-regulated. This receptor has an established function regulating the egress of lymphocytes from lymphoid organs[314]. Its role in stromal cells is unknown, however engagement of CD69 *in vitro* has been shown to stimulate TGF-β production[315]. Taken together, these findings suggest a co-ordinate shift in stromal signalling to the epithelial compartment in the context of inflammation from TGF-β and BMP-driven differentiation to proliferation mediated by gremlin 1 and GLP-2.

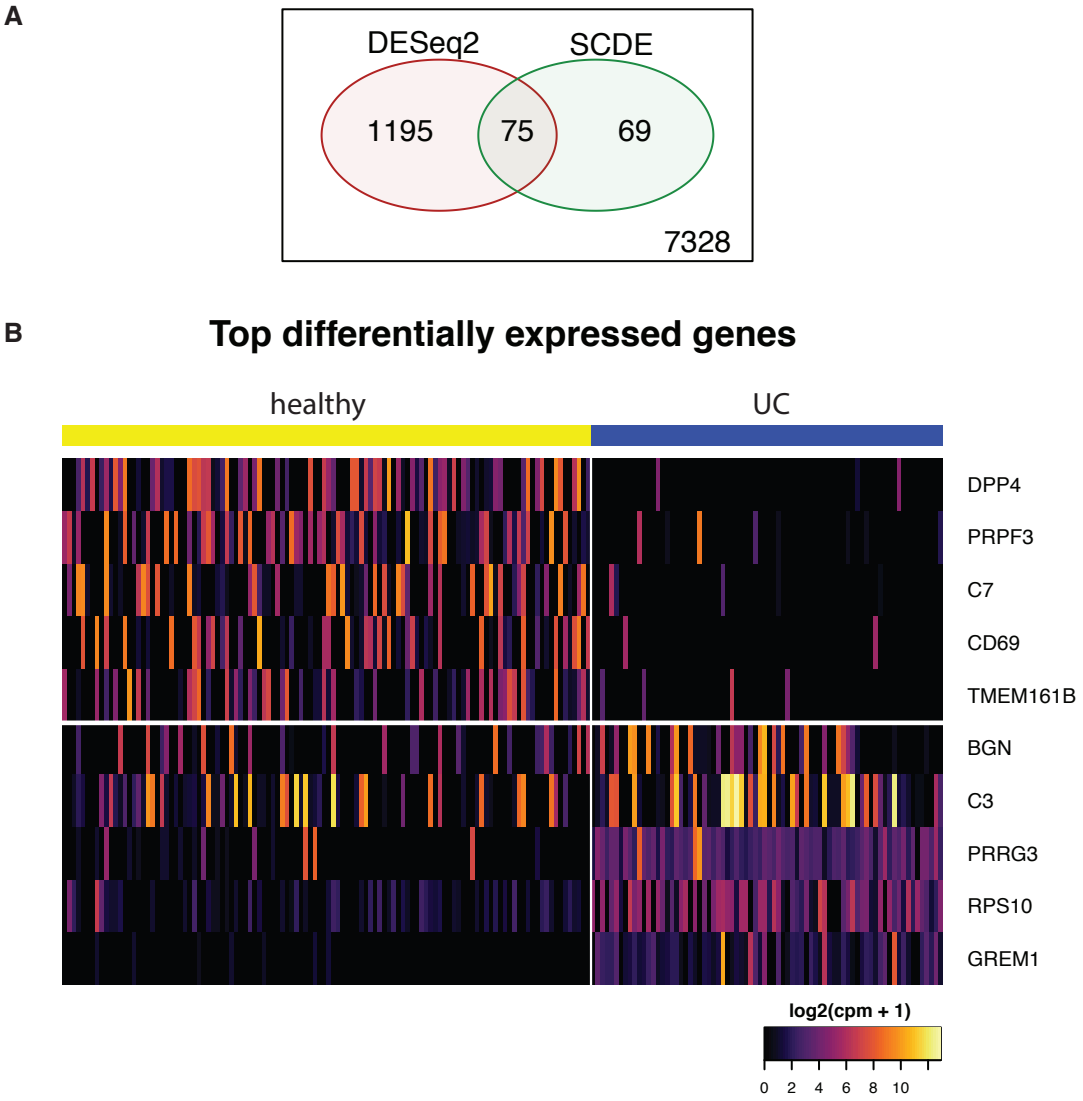


Figure 4.9: Differential gene expression between healthy and UC lamina propria stromal cells. (A) Numbers of significantly differentially expressed genes identified by the DESeq2 and SCDE methods (5% FDR). (B) Heat map showing the top 5 up and down regulated targets identified by both methods and ranked by median expression fold change calculated by the SCDE approach.

4.2.5 An additional gene co-expression module and cell cluster is detected in the stroma in acute UC

WGCNA was employed to detect co-expressed gene modules as before (Figure 4.10A). Of the five modules detected, four showed strong preservation when compared to the four modules detected in the healthy dataset (Figure 4.10B, 4.10 C). The fifth module (green), which showed no significant overlap with any module detected in

the healthy dataset, enriched for genes annotated with ontology terms including 'innate immune response', 'T cell activation', 'antigen processing and presentation' and 'regulation of leucocyte migration' (Figure 4.11). Unsupervised clustering of cells by NMF using modular genes identified an additional cell cluster, which was characterized by high expression of the green module (Figure 4.12, Appendix B Figure B.4). This additional cell subset was termed Stromal 4 (S4).

Highly ranked Stromal 4 markers include a number of fibroblastic reticular cell (FRC)-associated genes. FRCs are a major mesenchymal component of secondary lymphoid structures and FRC-like stromal cells have been described in discrete lymphoid aggregates within the healthy colon[42, 166]. Markers identified included the lymphocyte homing cytokines *CCL19* and *CCL21*, co-stimulatory TNF-superfamily ligand *TNFSF14*, retinoic acid signaling components *RBP5* and *RARB*, the MHC class II invariant chain (*CD74*), and the cell aggregation-inducing molecular chaperone clusterin (*CLU*) (Figure 4.13, Appendix B Figure B.5). In contrast to lymphoid FRCs, however, stromal 4 subset cells lack contractile machinery or significant expression of the lymphotoxin beta receptor (*LTBR*). Furthermore, markers associated with tolerogenic FRC functions, such as *PDL1* and *DEAF1*, are not enriched. This module also bears resemblance to a follicular dendritic cell (FDC) in terms of expression of follicular dendritic cell secreted protein (*FDCSP*) and *CXCL13*. A number of other high-ranked marker genes for this subset, including *C3*, *CD24*, *IL32* and the alarmin *IL33*, have been mechanistically linked to human inflammatory diseases including IBD[316, 317]. In UC single-cell RNAseq identifies expansion of a novel stromal cell population enriched for pro-inflammatory and FRC and FDC genes.

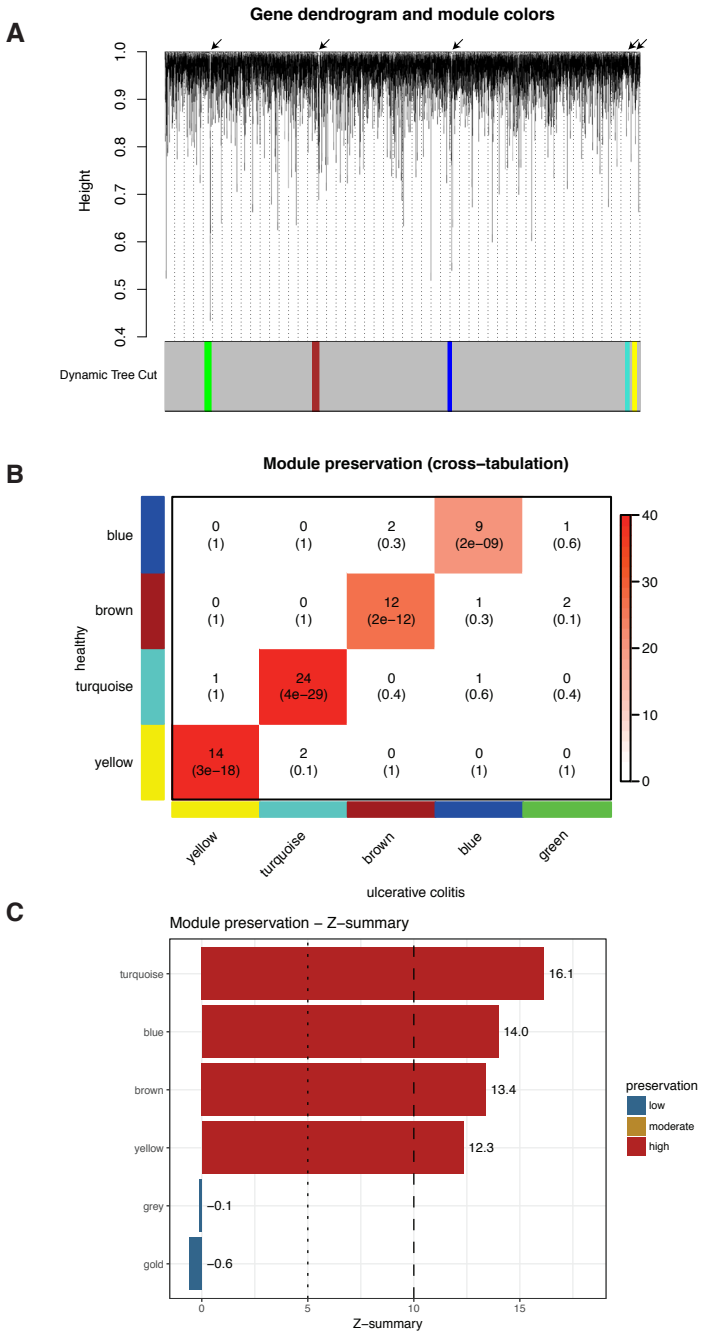


Figure 4.10: WGCNA module detection and preservation in UC stromal cells. (A) Gene dendrogram showing modules detected following WGCNA analysis of single cell RNAseq data derived from colonic mesenchymal cells from patients with UC. Modules detected are shown (arrows and colour bar). (B) Plot showing shared genes between the UC module (column) and healthy module (row). Fisher's Exact Test was used to calculate a p-value (parenthesised) and a colour scale applied ($-\log$(p-value)). (C) 'Z-summary' module preservation score for the healthy colon modules within the gene co-expression network generated from the UC dataset. The grey and gold control modules represent random samples genes (taken from all unassigned genes and all genes respectively). A score of > 10 indicates strong preservation (see Chapter 2).

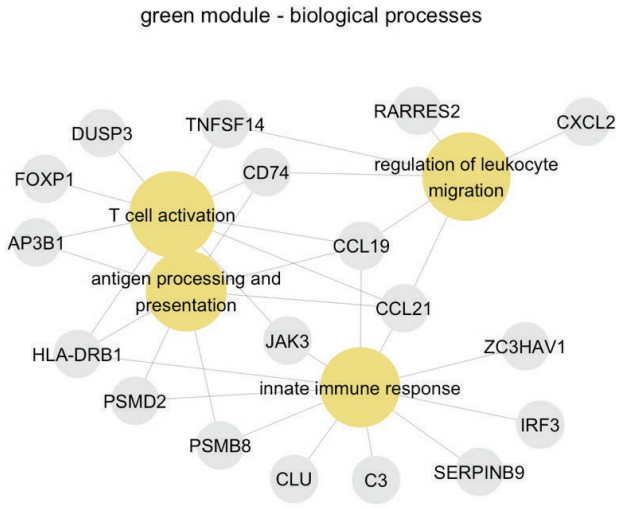


Figure 4.11: Ontology enrichment for UC-specific 'green' module. Network plot of enriched ontology terms (yellow) and linked module genes (grey) for the green module, which was not detected in the healthy dataset.

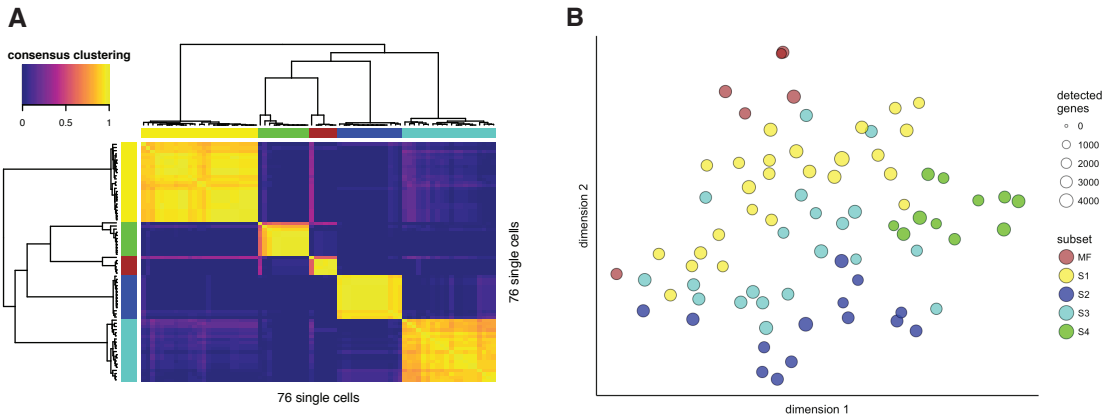


Figure 4.12: NMF clustering of UC stromal cells. (A) Consensus clustering over 50 randomly initiated NMF runs for the UC dataset. Heat colour represents co-clustering probability. Cluster assignment is indicated by the colour bar. (B) A 2D t-SNE visualisation of the UC dataset incorporating all biologically variable genes. Points represent single cells and are coloured by cluster assignment. The UC-specific stromal 4 (S4) cluster characterised by high expression of the green module.

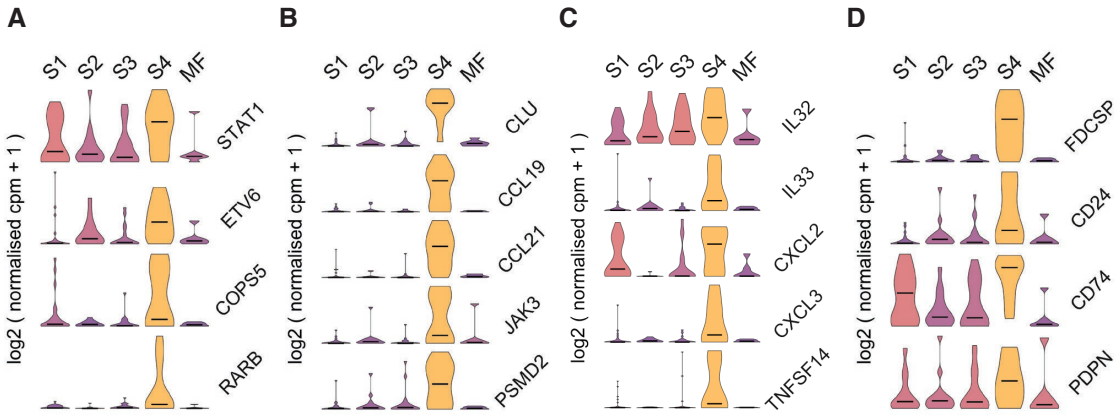


Figure 4.13: Stromal subset 4 marker genes. Violin plots for stromal 4 marker genes, including transcriptional regulators (A), genes annotated with the enriched ontology term 'innate immune response' (B), additional cytokines and chemokines (C), and candidate markers selected for validation (D). Crossbar indicates median expression.

4.2.6 Validation of the UC-specific Stromal 4 subset

We next investigated whether the pro-inflammatory Stromal 4 group observed in the single-cell RNA-seq data could be detected at a protein level in colonic tissue samples from UC patients. We obtained biopsy samples from inflamed distal colon from patients with a recent diagnosis of UC. We stained dissociated cell suspensions derived from patients with UC and healthy controls with antibodies to predicted Stromal 4 subset markers. Figure 4.14A confirms the existence of an FDCSP^{high}, CD24^{high} population of stromal cells in inflamed UC tissue but not in healthy donors by FACS analysis. Increased FDCSP staining is also seen within the lamina propria of inflamed UC tissue sections by IHC (Figure 4.14B). Additional Stromal 4 marker proteins CD74 and PDPN are also enriched in colonic stromal cells in active UC (Figure 4.15). CCL19 and IL33 expression, assessed by intracellular FACS staining, was found to be increased in the fibroblast subset expressing the stromal 4 markers CD74, CD24 and PDPN (Figure 4.16).

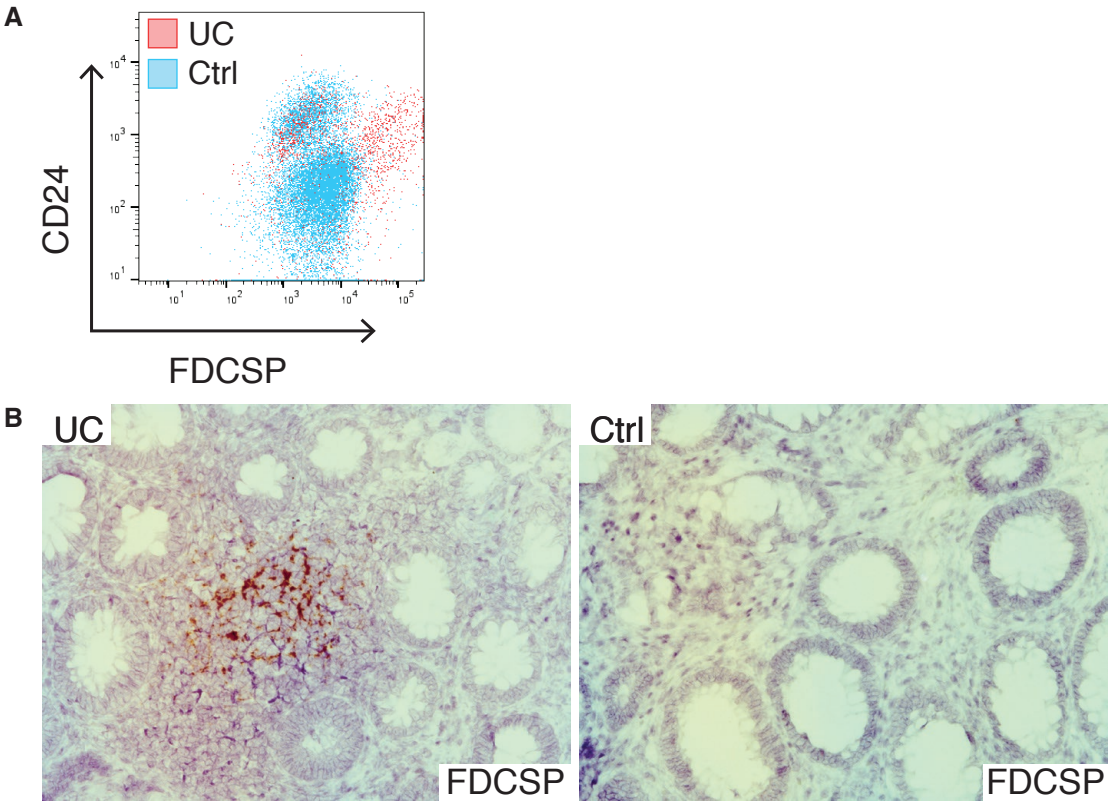


Figure 4.14: Stromal FDCSP expression in healthy and UC specimens. (A) Flow cytometry analysis of FDCSP^{high} and CD24^{high} colonic stromal cells from healthy controls (blue) or patients with active UC (red). (B) Single molecule ISH staining of FDCSP in Ctrl or UC colonic tissue sections (positive signal — brown, haematoxylin nuclear counterstain — blue).

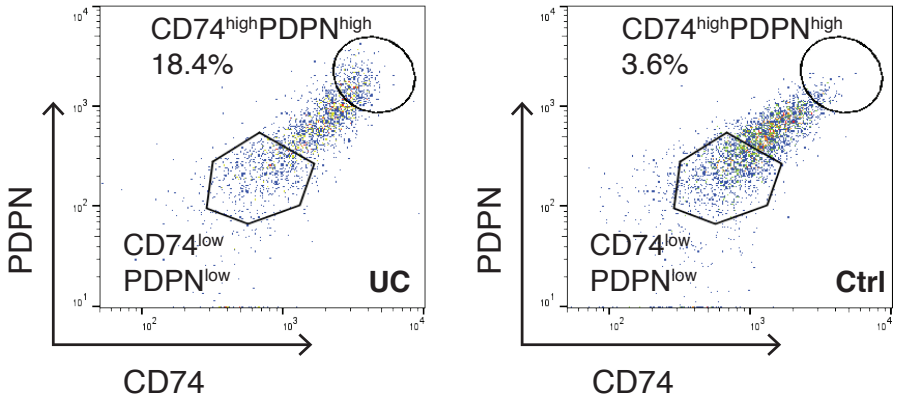


Figure 4.15: Stromal PDPN and CD74 expression in healthy and UC specimens. (A) Flow cytometry analysis of CD74 and PDPN expression on colonic stromal cells from healthy controls (right panel) or patients with active UC (left panel). The percentages of cells in the CD74^{high} PDPN^{high} gate are indicated.

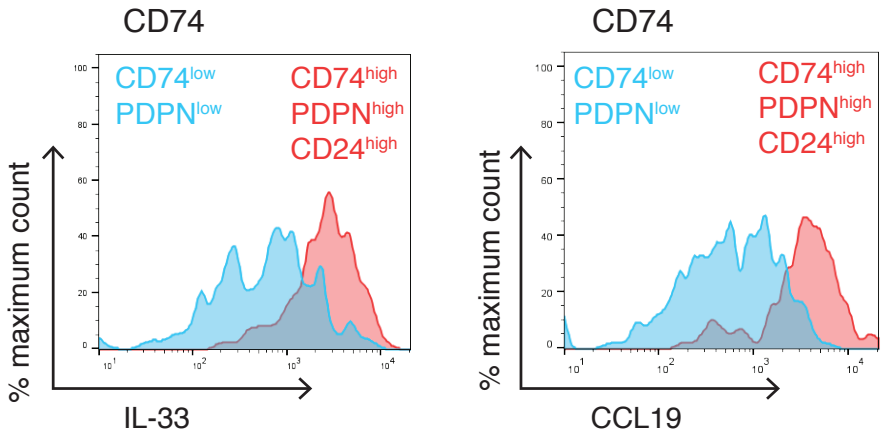


Figure 4.16: Intracellular cytokine staining in colonic stromal cells from patients with active UC. Comparison of intracellular CCL19 and IL-33 levels in $CD74^{high}PDPN^{high}CD24^{high}$ cells (red) versus the corresponding $CD74^{low}PDPN^{low}$ subset (blue) in inflamed UC colonic tissue.

4.2.7 Inflammation-associated changes in stromal composition in UC

I next examined changes in stromal gene expression and the proportions of various stromal cell subsets comparing inflamed to non-inflamed biopsy sites in patients with UC. Global differential expression analysis with DESeq2 identified 66 differentially expressed genes, while SCDE identified only 3. The two genes identified as up-regulated at inflamed sites by both methods were *IDO1* and *CHI3L1* (Figure 4.17A). Indolemine 2,3 dioxygenase (*IDO1*) acts as the first and rate-limiting step in tryptophan catabolism along the kyneurine pathway, which mediates immune tolerance via suppression of effector T-cell responses[318]. Proteomic and gene expression studies have identified it to be among the most highly over-expressed genes in human IBD and IBD animal models[319, 320]. Inhibition of IDO1 worsens colitis, while induction limits severity in animal models[321, 322]. Induction of IDO1 has previously been localised to the epithelial cell fraction[323]. However, expression in the stroma as observed here may be of relevance in the context of severe inflammation or ulceration when the overlying epithelium is depleted. Chitinase 3-like 1 (*CHI3L1*) lacks chitinase activity but is secreted by activated stromal and immune cells and is thought to be involved in the processes of inflammation and

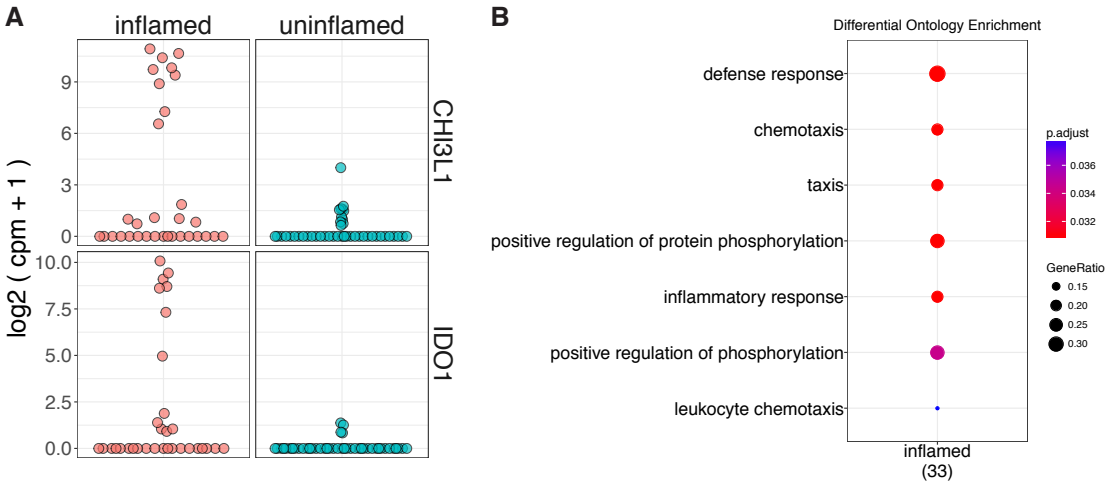


Figure 4.17: Differentially expression analysis comparing sites of inflammation in UC to adjacent uninvolved mucosa. (A) Expression plots for the indicated targets identified as differentially expressed by both SCDE and DESeq2. Points represent single cell expression values. (B) Enriched ontology terms among 37 genes identified as up-regulated at inflamed sites by DESeq2.

tissue remodelling. It is undetectable in healthy colonic tissue but up-regulated in colitis and colitis-associated cancer. In these settings it has been shown to induce proliferation of colonic epithelial cells[324]. Ontology enrichment using the longer list of differentially expressed genes identified by DESeq2 revealed multiple terms related to mobilisation of the inflammatory response, suggesting that stromal fibroblast-like cells are active participants in this process (Figure 4.17B).

Following WGCNA clustering, each sequenced single cell was annotated both with a stromal subset and whether it originated in a biopsy from an inflamed or an uninflamed site. From this, I derived an estimate of the change in stromal composition with inflammation in UC. The proportion of stromal 1 and 2 cells fell at inflamed sites with compensatory increases in the stromal 3 and 4 populations, with the most marked changes seen in stromal 2 and stromal 4 (Figure 4.18A). The loss of stromal subset 2 from 25% to 9% of the total population was borne out by FACS analysis. We observed a decrease in SOX6^{+} stromal cells in biopsies from inflamed UC colonic tissue when compared to healthy controls (Figure 4.18B). Therefore, changes in proportions of differing stromal cell states are observed in UC, with expansion of a pro-inflammatory (Stromal 4) and depletion of an

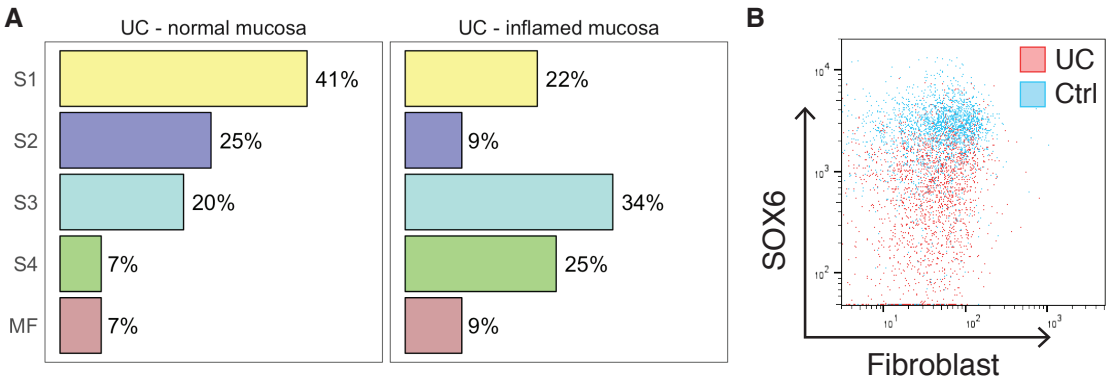


Figure 4.18: Alterations in stromal composition at sites of active inflammation compared to adjacent uninvolved mucosa. (A) Comparison of the cluster assignment for single cells derived from biopsies taken from inflamed sites compared to adjacent non-inflamed mucosa in UC patients. (B) Flow cytometric analysis of SOX6 (a Stromal 2 marker gene) expression in control (blue) or UC (red) colonic stromal cells.

epithelial-interacting compartment (Stromal 2) reflecting a shift in the overall activity of the stromal compartment from regulating epithelial differentiation to mobilizing an immune response.

4.2.8 Single-cell transcriptomes of colonic stromal cells from patients with UC-associated dysplasia

Individuals with ulcerative colitis are at increased risk of developing colorectal carcinoma compared to the general population, even when symptom-based and endoscopic indices of inflammatory activity are low[325]. The mechanisms by which cancers arise are IBD is different from sporadic CRC with diffuse, regional dysplastic change in the epithelium often preceding cancer onset[326]. The process of epithelial differentiation, ultimately leading to apoptosis and shedding of differentiated cells into the gut lumen, is an important protective mechanism, as harmful mutations generated in the proliferative compartment are typically eradicated by this route. Given the important role of stromal cells in maintaining the balance between epithelial differentiation and proliferation, I next sought to define the changes in stromal gene expression in individuals with UC-associated dysplasia, with a view to identifying dysregulated genes that might influence this pathway.

Single mesenchymal cells from 2 donors with UC-associated dysplasia were

captured and sequenced using the C1 platform as previously (Appendix B Table B.2). Both dysplastic and adjacent non-dysplastic sites were sampled, and counter-staining was performed so that the single cell libraries could be traced to their site of origin. A total of 90 single cell transcriptomes were generated, of which 81 passed the QC process. Analysis with SCDE and DESeq2 identified 558 genes showing differential expression by both methods. Interestingly, the genes *DPP4*, *CD69*, *GREM1*, and *BGN*, involved in TGF-β signalling and the regulation of cell growth, showed the same pattern of altered expression in this dataset as in acute UC despite the absence of histological active inflammation in these subjects (Figure 4.19A, Appendix B Table B.2). If the observed changes were, as previously hypothesised, supportive of an expanded epithelial proliferative compartment, this could predispose to the generation and retention of deleterious mutations.

Highly ranked up-regulated genes were enriched for markers that had previously been identified as central 'hubs' in the blue module and corresponding Stromal 2 cell cluster (Figure 4.19B). In contrast, the Stromal 1 chemokines *CCL8* and *CCL13*, chemotactic for multiple leucocyte subsets, were among the highest ranked down-regulated genes. In light of this, clustering was performed using WGCNA and NMF to identify changes in stromal composition.

The gene co-expression modules identified in healthy donors again showed high preservation in this independent dataset (Appendix B Figure B.6) and clustering using modular genes again captured a dominant source of variability within the dataset that is clearly observable in the first two principle components (Figure 4.20A). As expected, the Stromal 2 fraction was increased, constituting 50% of cells from dysplastic regions and 38% in adjacent non-dysplastic mucosa. The increase in Stromal 2 cells, which sm-ISH and IHC had suggested were in close approximation to the epithelial monolayer, would be consistent with expansion of the epithelial compartment in this setting. Together with the reduction in Stromal 1, chemotactic for various leukocyte subsets, these changes in stromal composition and gene expression offer insight in to the mechanisms by which the local tissue microenvironment is altered to favour the development of dysplasia.

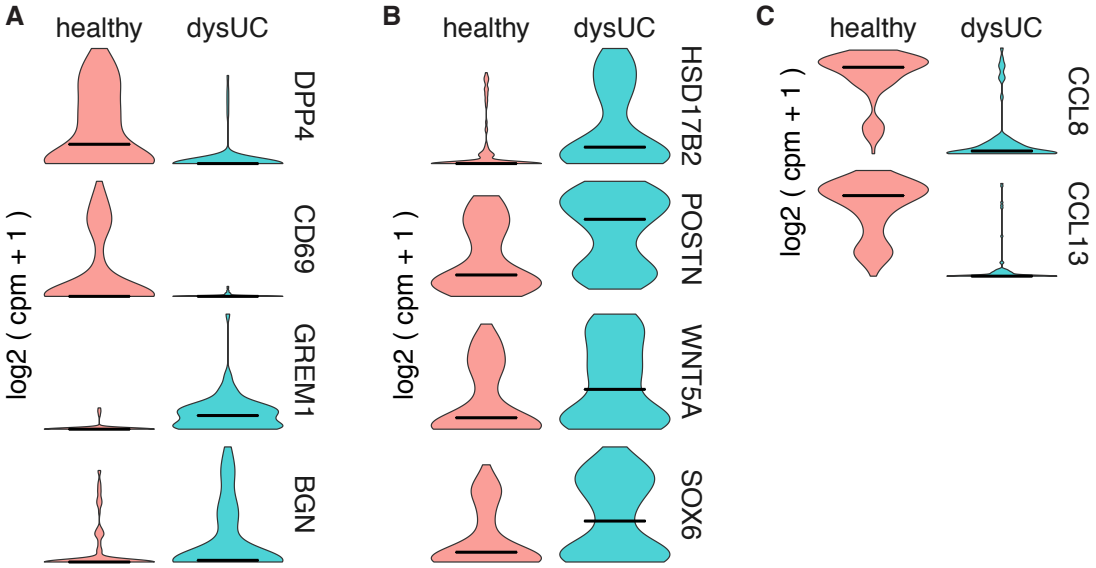


Figure 4.19: Differential stromal cell gene expression in dysplastic UC. (A) Gene expression violin plots for regulators of epithelial proliferation (A), Stromal subset 2 hub genes (B), and Stromal subset 1 chemokines (C) in lamina propria fibroblast-like cells from healthy donors and patients with dysplastic UC (dysUC). Crossbar indicates median expression.

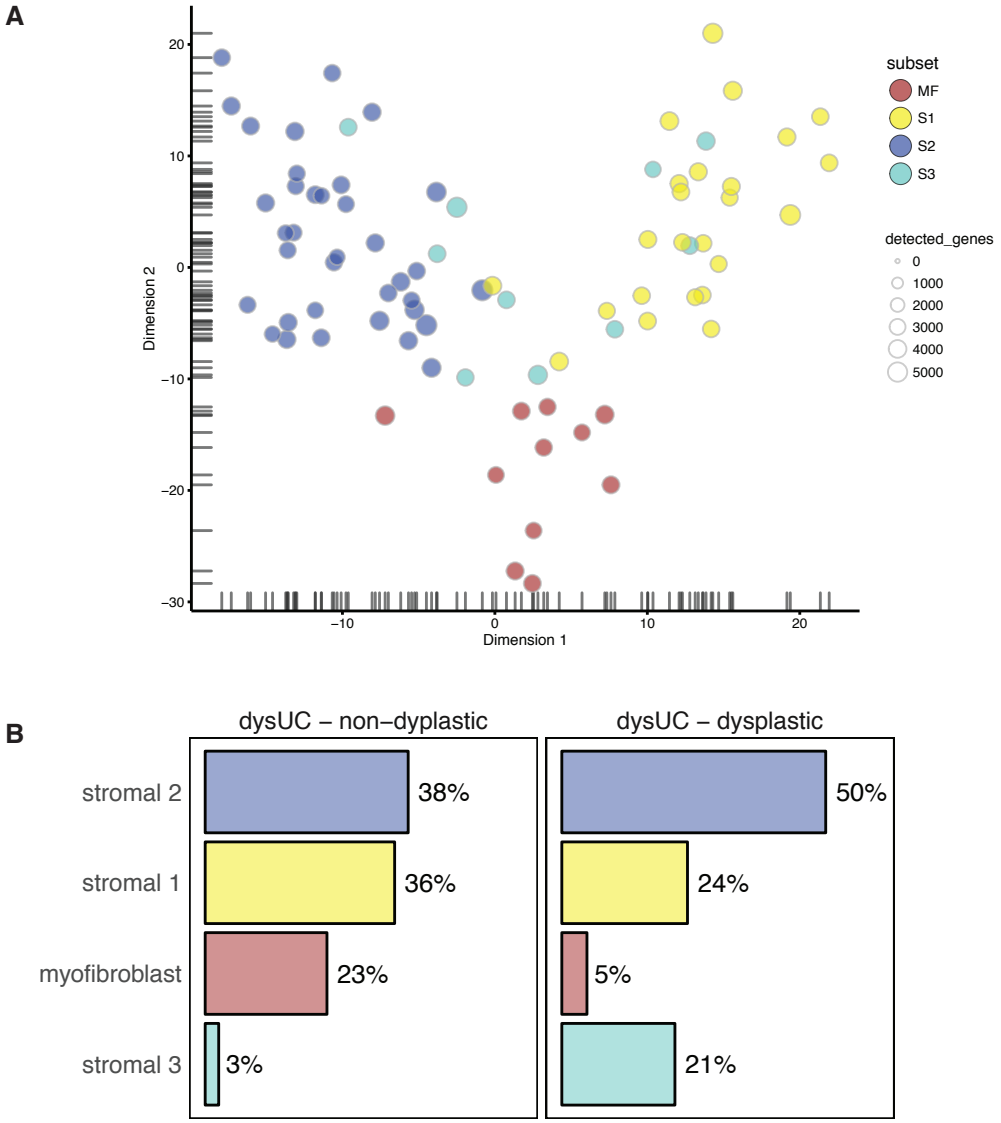


Figure 4.20: Stromal composition by cluster assignment in dysplastic UC. (A) Projection of cells (circles) in the first two principle components incorporating all biologically variable genes. Cells are coloured by cluster assignment. (B) Stromal composition by cluster assignment in cell suspensions derived from dysplastic sites and adjacent non-dysplastic mucosa in patients undergoing surveillance for known UC-associated dysplasia.

4.2.9 Transcriptional changes induced by TNF-alpha stimulation of colonic stromal cells

The gene expression profiles of lamina propria fibroblast-like cells showed enrichment both for for markers associated with pattern-recognition receptor signalling (Chapter 3 Figures 3.7) and the response to TNFα (Figure 4.21) compared to colonic epithelial cells. I next sought to determine whether these pathways could be activated *in vitro* and to define the early transcriptional responses in this context.

Freshly isolated colonic mesenchymal cells from three healthy donors were incubated for 2 hours with either TNFα or one of a panel of PRR ligands targeting NOD2 (MDP / PAM), TLR4 (LPS), or TLR5 (Flagellin). The expression of known downstream targets of these pathways such as IL-1β, IL-6, and CXCL8 (IL-8) was then quantified by RT-qPCR, internally normalised to the housekeeping gene *GAPDH* and expressed as a fold change versus an unstimulated control sample from the same donor. A consistent induction of CXCL8 expression was observed on TNFα stimulation (Figure 4.22A). The remaining targets were expressed both basally and on stimulation below the threshold for reliable detection in these cells.

Having established that isolated lamina propria fibroblasts were responsive to TNFα stimulation, single-cell RNAseq was again performed. Single cells were isolated from a further three healthy donors (Appendix B Table B.3) and processed as separate batches using the C1 platform. A total of 90 cells were sequenced, of which 61 passed the QC process. The DESeq2 method identified 1143 genes as showing significant differential expression between the TNFαstimulated and control conditions. However, no genes reached the significance threshold for the more conservative SCDE method. The highest ranked up-regulated gene on TNFα stimulation was Cholesterol-25-Hydroxylase (*CH25H*), the critical enzyme involved in the synthesis of the leucocyte chemoattractant 7α, 25-dihydroxycholesterol (Figure 4.22B). Mice deficient in CH25H show defects in lymphocyte trafficking and attenuated immune responses[327]. The subset of upregulated genes also showed moderate enrichment for a members of a curated list of predominantly apoptosis-regulating genes known to be induced by TNFα independently of p38

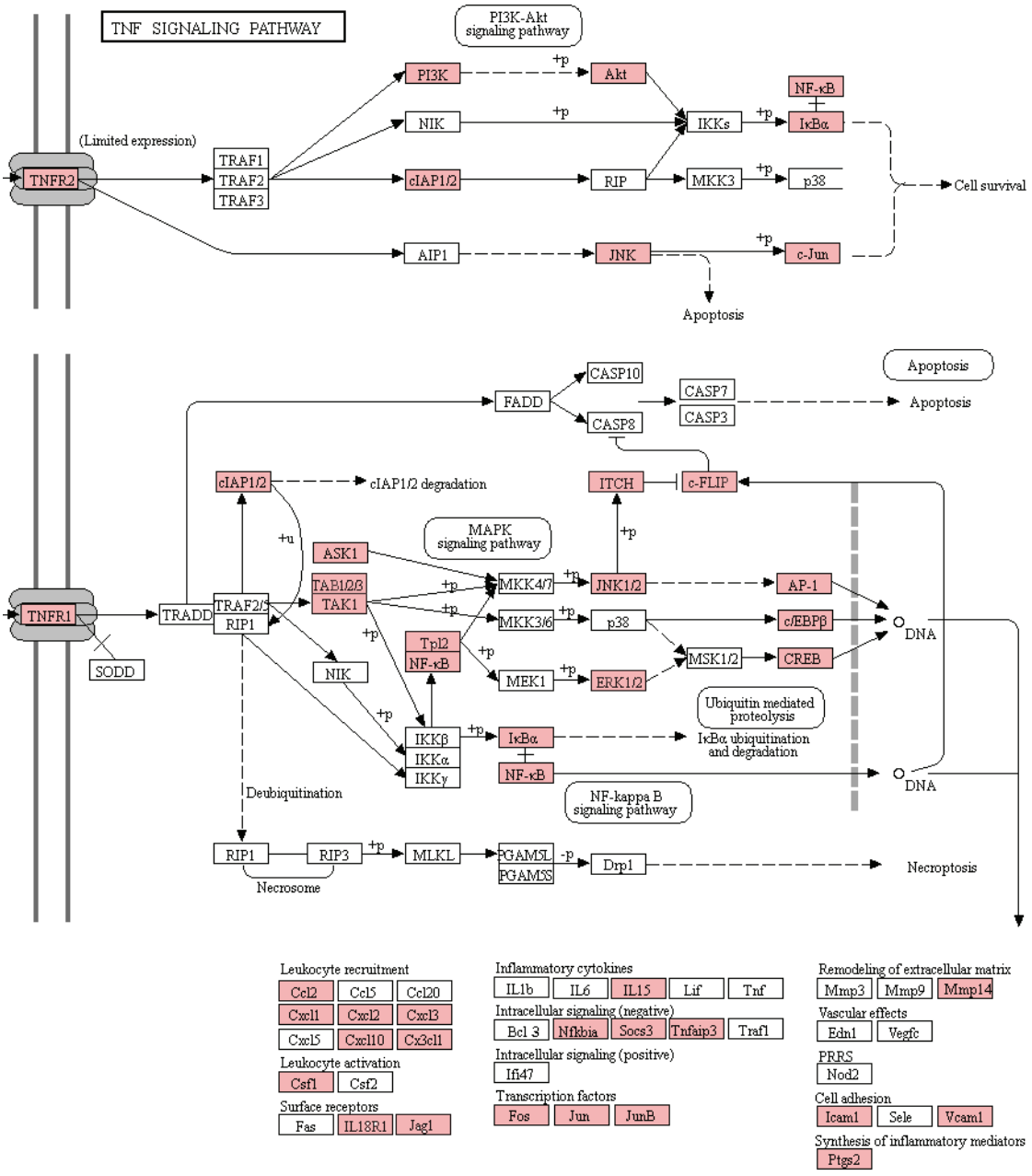


Figure 4.21: KEGG pathway enrichment: TNFα signaling pathway. KEGG pathway map as indicated. Shaded nodes show significantly increased expression in colonic mesenchymal cells compared to epithelial cells.

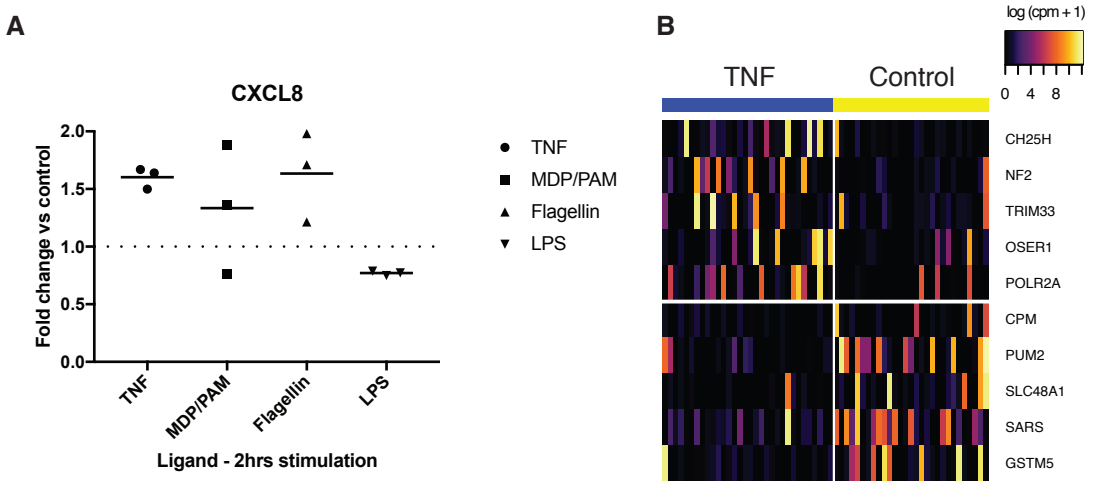


Figure 4.22: Transcriptional response to TNFα stimulation in freshly-isolated human stromal cells. (A) Fold-change in CXCL8 (IL8) expression following stimulation with the indicated ligands. Data points represent biological replicates. (B) Heat map showing the top 5 up and down regulated targets on TNFα stimulation of fibroblast-like cells from healthy donors identified by the DESeq2 and ranked by p-value.

(p-value 0.03)[328]. There was, however, no further ontology enrichment among either the up-regulated or down-regulated gene sets identified by DESeq2.

4.3 Discussion

Here I used single-cell RNA-seq to shed light on the molecular architecture of human colonic stromal cells and in health and IBD. An unsupervised bioinformatic approach was used to analyse single-cell RNA-seq data derived from human colonic stromal cells and grouped these cells into four distinct clusters that resembled myofibroblasts and distinct subtypes of fibroblasts with divergent functions. Clusters designated Stromal 1, 2 and 3 all expressed known fibroblast markers such as vimentin, but appeared functionally and anatomically distinct.

True contractile myofibroblasts made up a relatively small proportion of cells sampled by this method. The label ‘myofibroblast’ has been applied somewhat inconsistently under the existing classification in part based on discretising intermediate levels of α-SMA expression. While numerous stromal functions have been attributed to ‘myofibroblasts’ in the healthy colon, several investigators have

suggested that they are a small population, distinct from the α-SMA positive pericryptal fibroblasts[17, 66].

The most abundant fibroblast subset was Stromal 1. Consistent with a role in maintenance of the extracellular matrix, they exhibit increased expression of components of fibrillar collagen, and elastic fibres. Cells expressing Stromal 1 genes are localized deep within the lamina propria and not directly adjacent to epithelia. The increased expression of a number of chemokines by this group of cells suggest that they have a role in directing the positioning of immune cells within the lamina propria.

This work suggests that a specialized subgroup of stromal cells exists to direct epithelial cell function in the intestine. The Stromal 2 group showed increased expression of non-canonical WNT ligands and members of the TGF-β superfamily, suggesting an interaction with epithelial cells and maintenance of the epithelial stem cell niche[329]. These cells were located in proximity to epithelial cells in colonic tissue sections.

Stromal 3 enriched for genes involved in the synthesis of prostaglandin E_2 and innate sensing. Intestinal fibroblast-like cells play an important homeostatic role upon epithelial injury by sensing inflammatory cues generated in damaged tissue or by microbes and activating a Tpl2-Cox2-PGE_2 pathway promoting epithelial proliferation[34]. Stromal 3 cells appear transcriptionally primed to fulfil this function.

Examining stromal cells derived from patients with distal ulcerative colitis (UC) reveals not only the same four subsets found in healthy tissue but also an additional subset (Stromal 4). Stromal 4 demonstrated ontology enrichment for immune processes such as antigen presentation, leukocyte activation and chemotaxis. This subset shared a number of transcriptional hallmarks with fibroblastic reticular cells (FRCs), a critical mesenchymal component of secondary lymphoid structures[168]. It has been observed in colonic inflammation there is a correlation between the degree of epithelial damage and the number and size of sub-epithelial lymphoid follicles[330]. In this regard, it was interesting to detect the presence of Stromal 4, which may

represent the formation of lymphoid follicles, as expanded in inflamed UC but as also detected (albeit in lower numbers) in adjacent non-involved colonic tissue.

The relative abundance of fibroblasts subtypes was altered in disease states. In acute UC, there was depletion of the sub-epithelial stromal 2 subset in samples from inflamed sites, consistent with a breakdown in epithelial barrier function, and expansion in the stromal 3 and stromal 4 subsets, consistent with roles in response to injury and mobilization of the immune response. In contrast, the stromal 2 compartment was expanded at dysplastic sites in patients with long-standing quiescent colitis. Whether the alterations in the abundance of Stromal 2 cells are driving or reacting to the changing size and proliferative activity of the epithelial compartment is unknown. However, the increased expression of critical stromal-derived epithelial morphogens such as *GREM1* in both disease states hints at an active role.

The extent to which the detected transcriptional states reflect committed cell type specifications as opposed to reversible polarisations of transcriptional activity determined by local environmental cues remains to be determined. However, the identification and validation of a number of 'hub' marker genes for each subset in these experiments will provide a basis for future work using lineage tracing and selective ablation in animal models to elaborate their ontogeny and functions.

Primary human colonic fibroblast-like cells are a challenging substrate for *in vitro* experiments. While a there was a detectable response to TNFα stimulation by RT-qPCR, biological interpretation of the resultant single-cell RNAseq data was limited. Factors contributing to this include the necessity for short stimulation times due to declining cell viability, the fact that the starting populations are inherently heterogenous and subject to random variation when sampled in small numbers, and the confounding presence of a cellular stress response associated with isolation even in the control condition. Furthermore, it has been shown that even in well defined model systems, such as the stimulation of dendritic cells with LPS, the concordance between observations made at bulk and single-cell level is low[289].

The possibility of sampling greater numbers of single cells using high-throughput sequencing platforms will address some of these issues.

Chapter 5

Single-cell RNAseq of colonic stromal cells in a murine DSS model of Inflammatory Bowel Disease

Contents

5.1 Introduction 126
5.1.1 Motivation 128
5.2 Results 129
5.2.1 DSS challenge 129
5.2.2 Single-cell dissociation and capture 130
5.2.3 Sequencing results and assessment of batch effect 132
5.2.4 Unsupervised clustering of stromal cells from the control group 132
5.2.5 Conservation of human Stromal 1–3 transcriptional signatures in mouse 136
5.2.6 DSS-induced changes in stromal composition 141
5.2.7 Expansion of a Stromal 4-like fibroblast subset on DSS challenge 144
5.2.8 Biological activities of murine stromal subsets in health and DSS colitis 149
5.2.9 Ontogeny of stromal fibroblast subsets 151
5.2.10 Subset-specific transcriptional responses to DSS challenge 154
5.2.11 Specificities of existing fibroblast markers and novel marker candidates 156
5.3 Discussion 157

5.1 Introduction

Transcriptional profiling of freshly isolated human colonic stromal cells revealed multiple distinct fibroblast subtypes. These comprised myofibroblasts, showing expression both of contractile machinery and fibroblast marker genes, and the fibroblast subtypes Stromal 1–4. Ontology enrichment for differentially expressed genes within these groups suggests functional specialisation. The Stromal 1 group enriches for ECM turnover and chemokine signalling, the Stromal 2 group for pathways implicated in epithelial-mesenchymal crosstalk, the Stromal 3 group for repair and proliferation, and the Stromal 4 group for interaction with immune cell subsets. Validation by immunohistochemistry, smISH and FACS supports the classification structure derived from the transcriptomic data, which is also highly conserved across experiments.

While the single-cell transcriptional data provides a valuable resource for defining fibroblast heterogeneity, functional assays are required to gain definitive insight into the properties of the stromal subsets and their distinct contributions to homeostasis and disease. The use of freshly isolated cells from endoscopic forceps biopsies is hampered by the small numbers of stromal cells that can be generated and the cellular stress associated with tissue dissociation and further marker based selection to purify subsets. Colonic stromal cells can be expanded in culture, and this approach has been used to define several important aspects of their biology including their ability to respond to innate immune stimuli, direct T-lymphocyte differentiation, and support epithelial organoid growth[107, 116, 156]. However, the extent to which the fibroblast heterogeneity and specialisation seen *in vivo* is maintained in culture remains unclear.

The tissue architecture of the colonic lamina propria is highly conserved in the mouse, and murine models are the basis of much of the current understanding of fibroblast biology in this area. Key signalling axes are common to both species. For example, similar epithelial phenotypes are observed in human Juvenile polyposis syndrome, an autosomal dominant condition characterised by mutations

in BMP signalling intermediaries and murine knockout models targeting the same pathway[79]. There are also numerous murine models of IBD, which show pathological similarity to human disease and are amenable to the same spectrum of therapeutic interventions, such as neutralising antibodies to TNFα.

Among the most widely used murine IBD models is the DSS colitis challenge[331]. Dextran sodium sulphate (DSS) is a negatively charged sulphated polysaccharide with a highly variable molecular weight. Administration of 40–50 kDa DSS to mice in drinking water results in an acute colitis similar to that seen in acute UC. The precise mechanism underlying colitis induction has not been fully determined but likely involves disruption of the intestinal mucous layer and direct toxicity to enterocytes causing a loss of barrier integrity. This allows translocation of pro-inflammatory intestinal contents, such as bacteria and their products, into the underlying tissue. Colitis induction is dependent on the microbiota but, unlike human disease, does not require B- or T-lymphocytes. Despite this limitation, the DSS challenge model is very widely used due to its simplicity, reproducibility and controllability.

The intense interest in single-cell RNAseq across multiple disciplines has driven technological innovation at great pace. Initial studies using FACS sorting and plate-based assays were first complemented by microfluidic platforms, such as Fluidigm C1, that allow reactions to take place in smaller volumes for greater capture efficiency and cost-effectiveness, and more recently by high-throughput methods that increase the number of cells captured by orders of magnitude[254]. A common feature of high-throughput methods is early barcoding, the process of appending a unique oligonucleotide barcode sequence to all transcripts originating from a particular cell as the cell is lysed. Following this, amplification steps can be performed in aggregate, greatly simplifying the protocol and reducing workload.

The accuracy of transcript quantification has also been improved by the use of unique molecular identifiers (UMIs)[332]. These degenerate oligonucleotide tags are appended to cellular mRNA molecules alongside the cellular barcodes and offer a method to trace sequencing reads back to unique molecules present at the point of lysis that is resilient to biases introduced during the multiple amplification steps.

Both cellular barcodes and UMIs are appended to the 3′ poly-A tail of mRNA molecules. Paired end sequencing is then employed, with one end of the read pair revealing the cell and molecular identifiers and the other, via alignment to a reference transcriptome, the identity of the transcript. Although the full-length transcript is not sequenced and therefore information regarding, for example, splice variants is not captured, accurate gene expression estimates can be generated from around 50,000 reads per cell, greatly increasing the number of cells that can be multiplexed and reducing sequencing costs.

Platforms such as drop-seq and the 10X Genomics Chromium™ Single Cell 3′ Solution use microfluidic devices to co-encapsulate single cells with nanoparticles (gels or beads) that deliver barcoded primers. Cell lysis and barcoding then occurs within the nanoliter-scale aqueous droplets in a reverse emulsion, before the emulsion is broken down to allow the remainder of the protocol to be performed in bulk. These approaches have been used to generate single cell datasets much larger than was previously possible[106, 333, 334]. Applied to a cell classification problem, increasing the number of cells sequenced improves both the ability to resolve similar subtypes and power to detect rare transcriptional states. The chance of capturing cells in transition between states is also increased and this has been used to refine developmental hierarchies both in the haematopoietic compartment and the intestinal epithelium[106, 335].

5.1.1 Motivation

In this section, I aimed to determine the extent to which the transcriptional subtypes seen in human lamina propria stromal cells were conserved in mouse. This was achieved by generation of a large single-cell dataset using the 10X Genomics Chromium™ Single Cell 3′ Solution. Both healthy controls and DSS challenged animals were utilised to provide a basis for comparison with human healthy and UC-derived biopsies. The dataset generated offers potential to refine the classification structure proposed in the previous chapters, to examine the specificities of existing fibroblast Cre recombinase systems and to propose new

markers for use in future functional studies including knockout models, subset ablation approaches and lineage tracing.

5.2 Results

5.2.1 DSS challenge

Male C57BL/6 mice were divided into treatment and control groups and a DSS challenge administered as described in Chapter 2. A robust and consistent induction of colitis was observed in the treatment group. DSS treated animals had increased in-life disease activity scores, incorporating parameters such as diarrhoea and rectal bleeding, increased colon mass and reduced colon length (Figure 5.1, Appendix C Figure C.1).

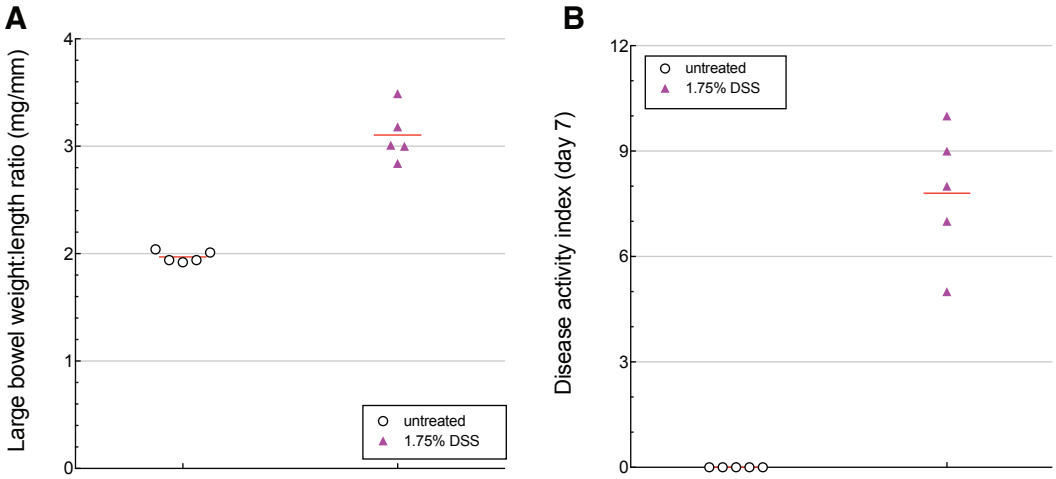


Figure 5.1: DSS colitis and control mice, in-life and post-mortem observations. (A) Ratio of large bowel weight to length by treatment group. Measurements were made post-mortem on study day 7. (B) A composite score of in-life disease activity measures (comprising weight loss, diarrhoea and rectal bleeding) for all treatment groups. Group means are indicated (cross-bars).

5.2.2 Single-cell dissociation and capture

Following euthanasia on study day 7, colons were isolated and single cell dissociation and capture performed the same day. A previously published protocol was employed with some modifications to generate single-cell suspensions[253]. MACS separation was used to deplete epithelial and immune cells as for the human samples (Appendix C, Figure C.2). Three colons (biological replicates) from each treatment group were processed for single cell RNAseq with the remainder fixed and retained for smISH validation. Cell viability post isolation determined by Trypan blue staining was approximately 90%. The median viable stromal cell yield was 3.1×10^5 for healthy controls and 4.7×10^5 for DSS samples. 5% Control cells (RAW 264.7 murine macrophages expressing a blasticidin resistance transgene) were added to each primary cell suspension immediately prior to capture. The pooled samples were sequenced across 3 HiSeq 4000 lanes. A summary of the experimental design is show in Figure 5.2.

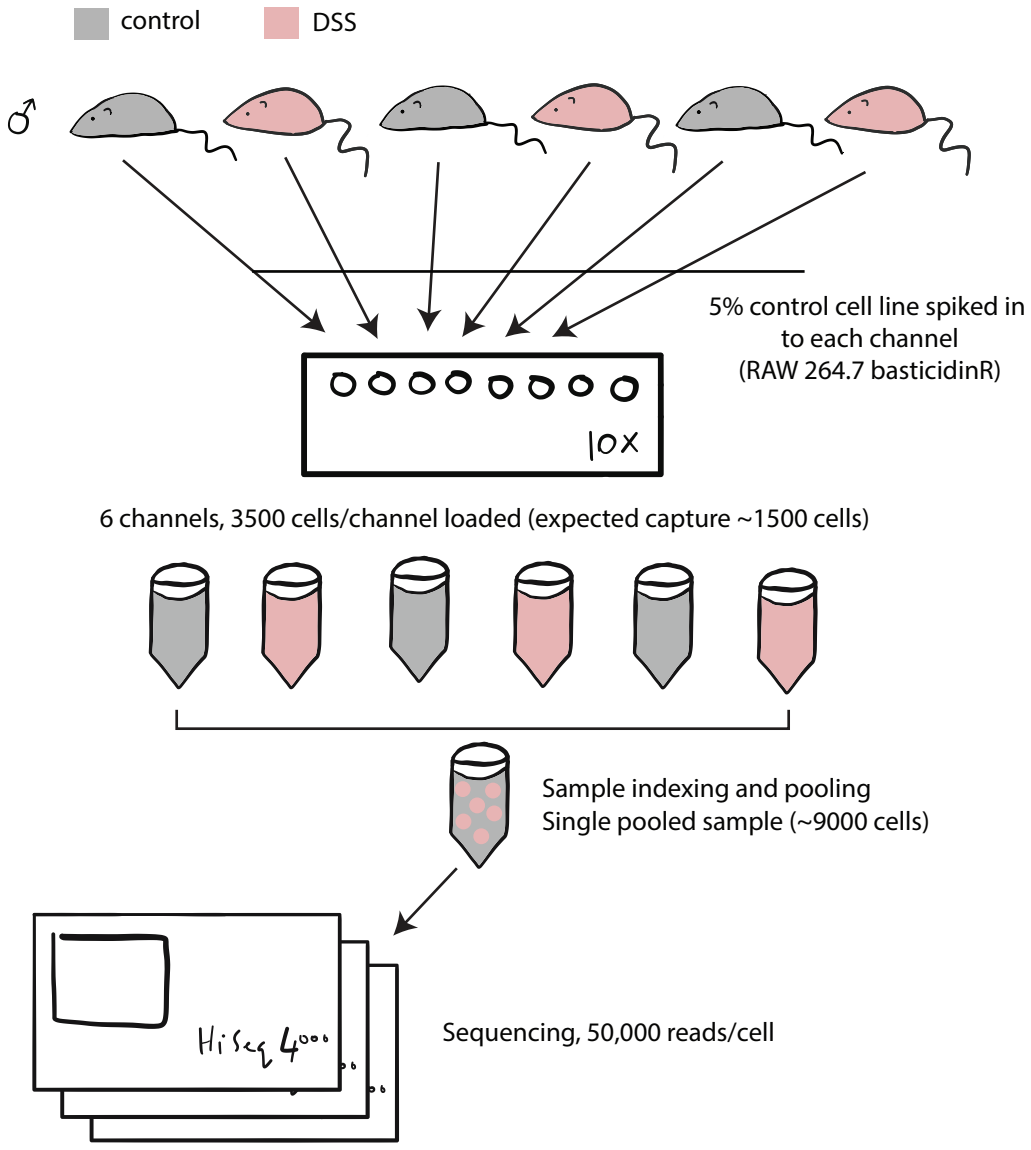


Figure 5.2: Murine DSS challenge: summary of experimental design. Colonic stromal cells isolated from 3 mice following a 7-day DSS challenge and 3 age and sex-matched controls were processed as indicated.

5.2.3 Sequencing results and assessment of batch effect

Median confidently mapped reads per cellular barcode was 48,081 for the healthy dataset and 49,606 for the DSS samples. The median UMI counts per cell were 7,931 and 11,839 respectively. As the UMI counts observed for control cells were similar across all channels (Appendix C, Figure C.6A,B), the increased library complexity seen in stromal cells from DSS samples is likely biological in origin and reflects increased transcriptional activity in the context of the DSS inflammatory milieu.

Principal component analysis of the control cell libraries was performed to investigate channel related batch effects within the dataset. A cell-cycle classifier was used to annotate cell cycle stage (see Chapter 2) and control cells were distributed throughout the cell cycle as expected for a cell line under conditions of exponential growth (Appendix C, Figure C.6C). A strong correlation was observed between cell-cycle stage and PC1 suggesting this was the major factor driving control cell heterogeneity (Appendix C, Figure C.6D). Neither channel nor treatment group showed significant correlation with the first three PCs, but a small treatment group effect was noted in PCs 4 and 5 (Appendix C, Figure C.7A). The genes with highest absolute weighting in PC4 were epiregulin (*Ereg*), serum amyloid A3 (*Saa3*) and C-X-C Motif Chemokine Ligand 5 (*Cxcl5*). These were all noted to be strongly upregulated in the primary stromal cells derived from DSS samples (Appendix C, Figure C.7B), and their detection in control cell libraries probably reflects a degree of contamination from cell-free mRNA or cellular debris arising from cells that had been damaged or ruptured during the isolation steps.

5.2.4 Unsupervised clustering of stromal cells from the control group

Following control cell removal and QC, a total of 7,171 single cells remained in the analysis (3,817 healthy, 3,354 DSS). Single cell clustering was performed in principal component space using the community detection algorithm from the R package 'Seurat' (see Chapter 2). Initially, 15 clusters were detected. A hierarchical phylogenetic tree was constructed to relate clusters (Appendix C, Figure C.10A),

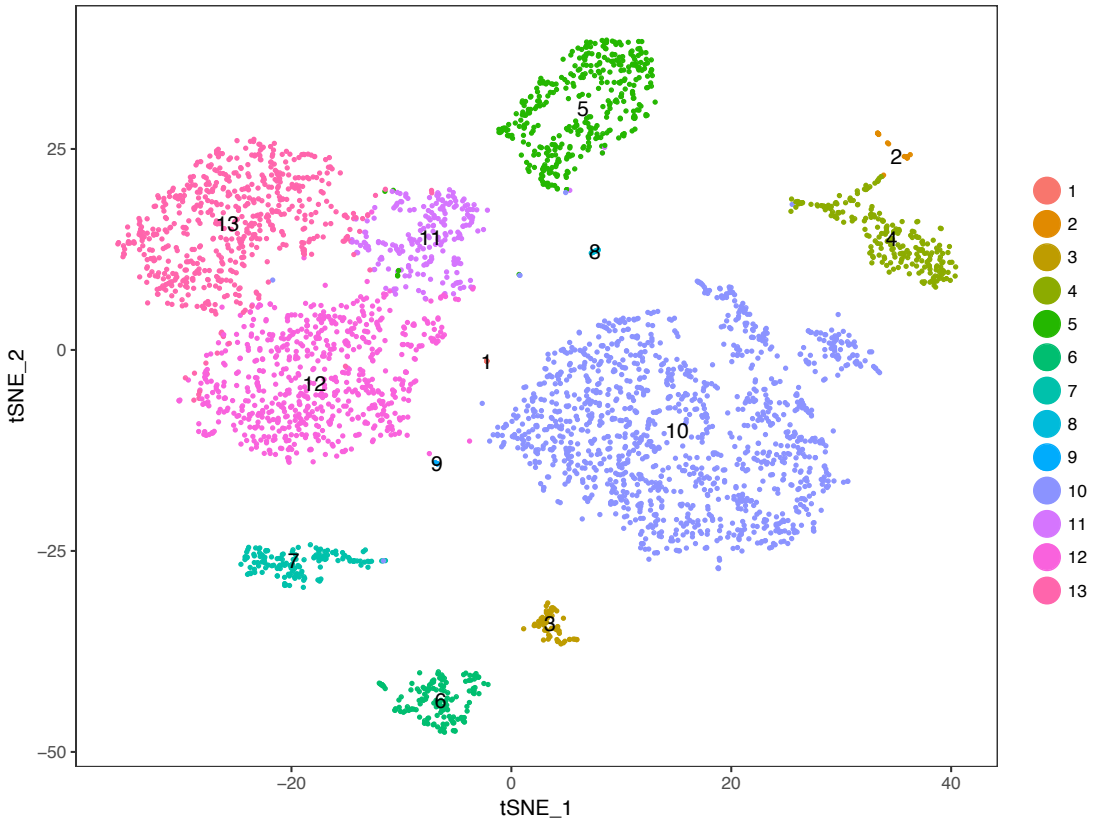


Figure 5.3: Detection of cell clusters in healthy control mice. (A) A t-SNE projection of 3,817 single cells derived from 3 healthy control mice. Cells are coloured by cluster annotation, derived using a graph-based method (Seurat). Each cluster is assigned a numeric identifier as indicated.

and cluster confidence was quantified in the form of the Out of Box Error (OOBE) in a random forest classifier trained on each internal node (Appendix C, Figure C.10B). Three similar clusters were merged (Appendix C, Figure C.10C) leaving a final clustering solution consisting of 13 clusters, which were assigned numeric identifiers. A t-SNE plot was generated for visualisation (Figure 5.3).

Multiple approaches were employed to assign cluster identity. Cluster marker genes were first ranked according to the area under the ROC curve in a binary classifier comparing cluster to non-cluster cells. The highest-ranked markers were then examined for known cell-type specific genes. Clusters showing specific expression of epithelial (*Epcam* and *Krt19*), pericyte (*Rgs5* and *Pdgfrb*)[20], vascular endothelial (*Pecam1 / Cd31*), lymphatic endothelial (*Lyve1*)[336], glial (*S100b* and *Gfap*)[337] and haematopoietic cell markers (*Cd52* and *Ptprc / Cd45*) were readily

identified (Figure 5.4A). Ontology enrichment was also performed using cluster marker genes (genes with AUC > 0.75). Enriched ontology terms were consistent with the known functions of the predicted cell types (Figure 5.4B).

Cluster 2, a small cluster of 32 cells, showed expression of markers associated both with enteric smooth muscle (*Myh11* and *Des*) and ICCs (*Kit* and *Ano1*). Further examination of this cluster revealed that it was composed of two distinct subclusters consistent with ICCs and smooth muscle cells respectively (Appendix C, Figure C.11). As the Seurat algorithm assigns cluster membership based on a cell's 30 nearest neighbours, clusters of less than 30 cells can become merged as is the case here.

Only 5 immune cells and 6 epithelial cells were detected, demonstrating the effectiveness of the MACS depletion of these cell types. Other low-abundance clusters included enteric glial cells (14 cells) and pericytes (67 cells). The ratio of pericytes to vascular endothelial cells was 1:2.3, which is within the range of previous estimates of pericyte abundance from multiple tissues[20]. The relative rarity of smooth muscle cells, ICCs and glial cells was not expected, as these cell types were not actively depleted. It is likely that some aspect of the isolation protocol, which included enzymatic digestion, density selection and passage through a 35 µm filter, resulted in the depletion of these populations.

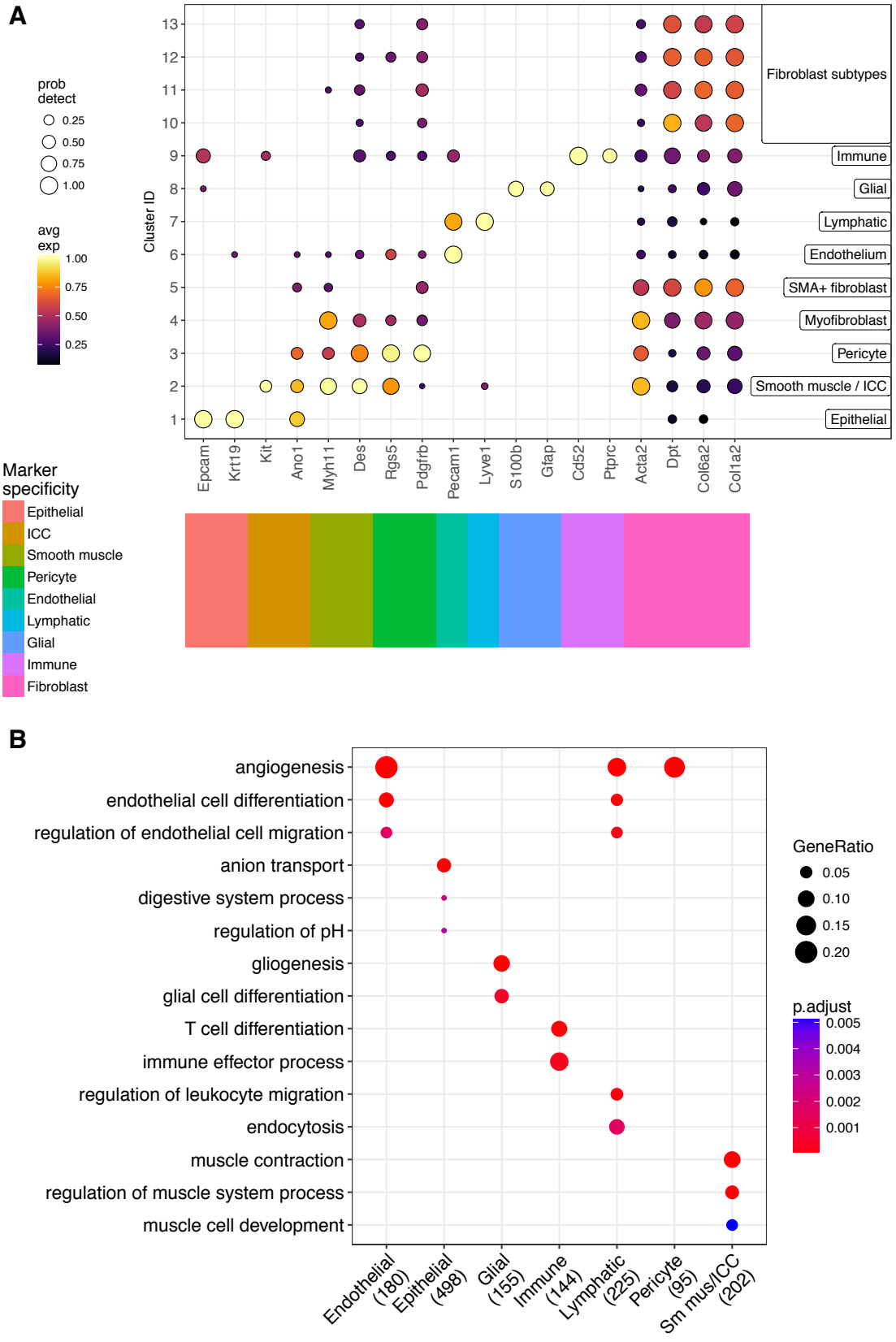


Figure 5.4: Correlation of clusters to expected cell types in healthy controls.

Figure 5.4 ***(previous page)*****:** (A) A dot plot showing expression of canonical marker genes for expected colonic cell types against detected cell clusters. Circle size represents the within-cluster probability of gene detection, while the fill colour represents the normalised mean expression level. The cell-type specificity of each marker is indicated (colour bar). The numeric cluster identifiers and corresponding inferred cell types are shown (left and right y axis labels). (B) Selected ontology terms showing significant enrichment among top marker genes for the non-fibroblast clusters. Marker genes were identified by ranking all genes by AUROC in a binary classifier comparing cells within the cluster to all other cells. The number of markers identified for each cluster is indicated (x-axis, parenthesised). Circle size corresponds to the proportion of markers annotated to a given term, while the fill colour indicates the adjusted p-value.

5.2.5 Conservation of human Stromal 1–3 transcriptional signatures in mouse

The remaining 6 cell clusters (4–5, 10–13), comprising 3,391 cells or 89% of the dataset were fibroblast-like cells characterised by expression of the pan-fibroblast markers such as *Dpt*, *Col6a2* and *Col1a2* (Figure 5.4A). Clusters 4 and 5 showed α-SMA expression, while only cluster 4 showed significant expression of smooth muscle myosin (*Myh11*). Marker gene cross tabulation was used to determine whether the murine fibroblast clusters could be related to human stromal subsets (Figure 5.5B). Marker genes were identified using the AUROC method in both datasets and the human genes lists were converted to murine orthologs. AUROC thresholds were set at 0.67 for the murine clusters and 0.65 for human, so that the numbers of markers cross-tabulated was broadly similar. A Fisher's Exact test was used to determine the significance of marker overlap. Highly significant overlaps were detected between the human myofibroblast and stromal 2 groups and murine clusters 4 and 5. The human Stromal 1 group showed overlap with clusters 11–13, which was unambiguous in the latter two clusters. Stromal 3 showed significant overlap with clusters 10 and 11, although overall significance was relatively weak in comparison to the other groups.

Murine clusters were assigned identifiers based on the human cluster to which they were most similar. In the case of Stromal 1, where a one-to-many relationship was observed, decimal point identifiers were appended to denote subclusters. The ability to further subdivide this cluster likely results from the far greater number of cells captured in the murine dataset, although the fact that whole colons were

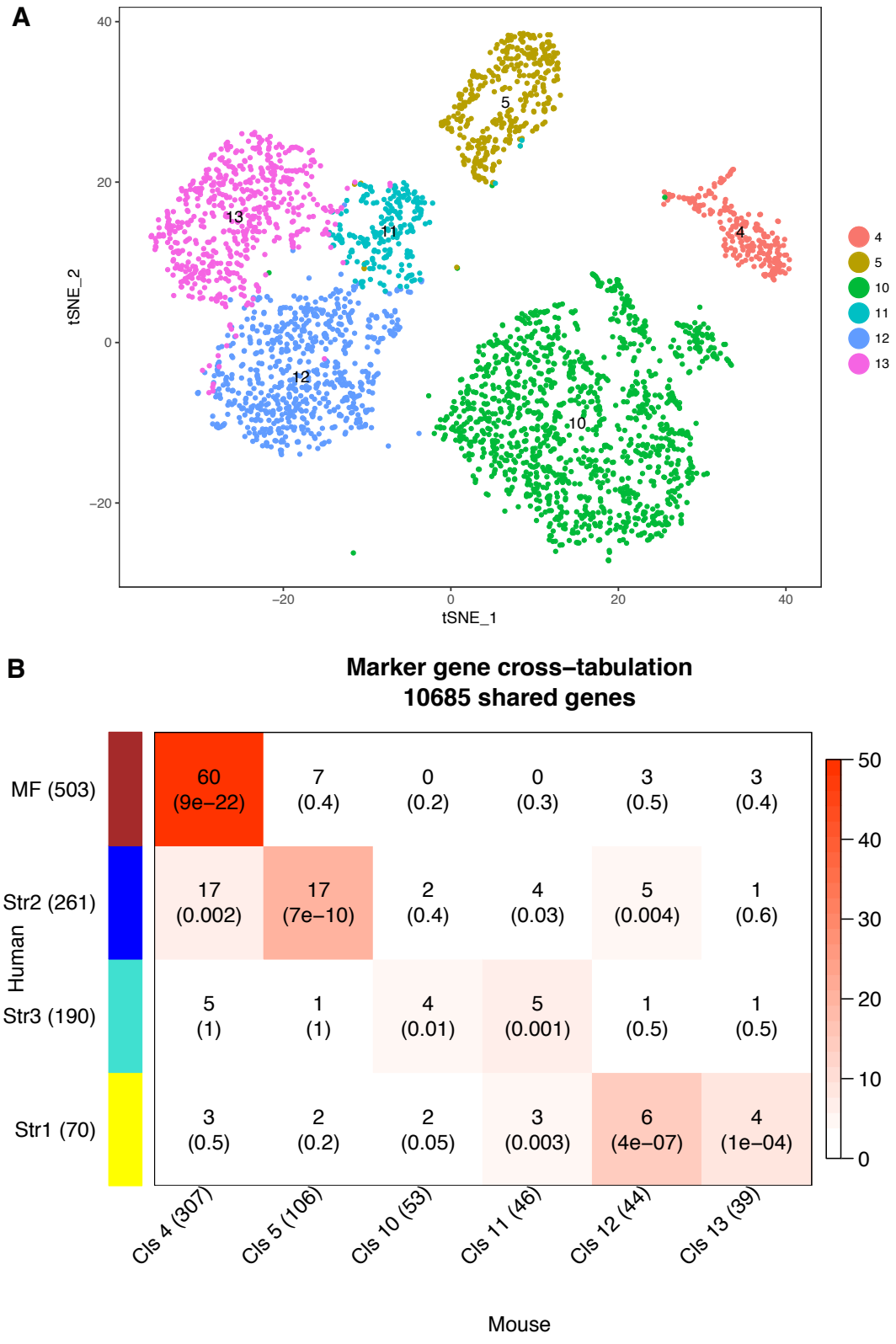


Figure 5.5: Significant overlap exists between murine and human fibroblast subsets.

Figure 5.5 ***(previous page)*****:** (A) TSNE projection of the 6 clusters of fibroblast-like cells identified in the healthy murine dataset. Numeric cluster identifiers are shown. (B) Cross-tabulation of human (rows) and murine (columns) fibroblast subset marker genes. The number of markers identified (using an AUROC threshold) for each cluster is indicated (axis text, parenthesised). The number of shared markers and the corresponding p-value (Fisher's Exact Test, parenthesised) are indicated in the text matrix, and a colour scale is applied ($-\log$(p-value)).

digested in the case of mouse versus mucosal forceps biopsies in human may also have contributed. Murine cluster 11 overlapped with human Stromal subsets 1 and 3, but showed a close phylogenetic relationship to the other Stromal 1 clusters, so was termed Stromal 1.3.

While the overall number of shared markers was low, subset specific marker pairs identified from the human data again demonstrated correlated expression in the mouse (Figure 5.6). Stomal 2 markers localised to the same anatomical niche, adjacent to the epithelial layer surrounding the upper crypts (Figure 5.7). These observations support the view that fibroblast heterogeneity is at least partially conserved between species. A fully annotated t-SNE representation of the dataset was now generated and is shown in Figure 5.8B.

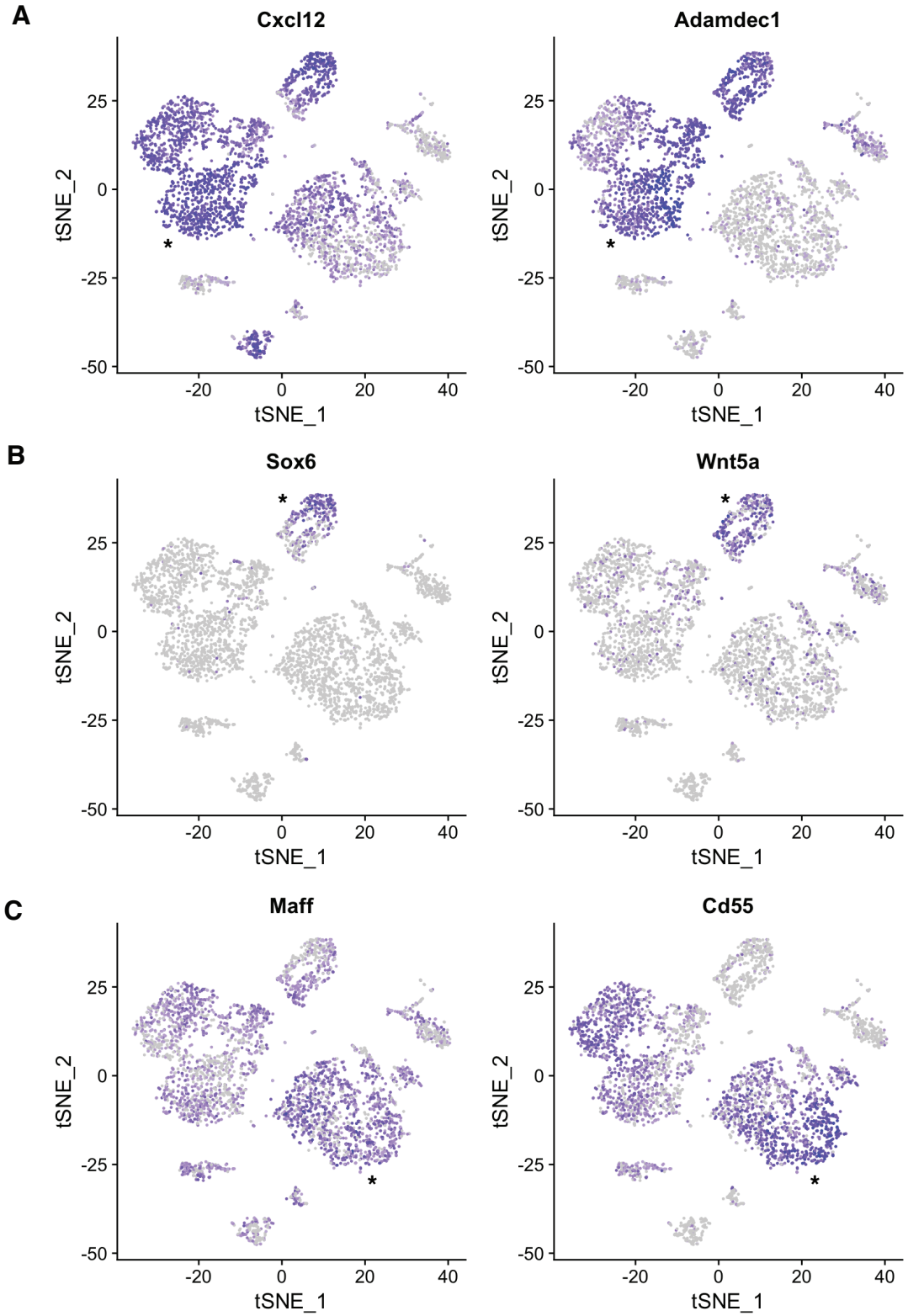


Figure 5.6: Expression plots of human fibroblast subset markers in the murine dataset.

Figure 5.6 ***(previous page)*****:** t-SNE projections of the murine healthy dataset. Cells are coloured by normalised expression levels of the indicated marker genes. The murine cluster with the highest mean expression is also indicated (*). (A) Human stromal 1 markers, (B) human stromal 2 markers, (C) human stromal 3 markers.

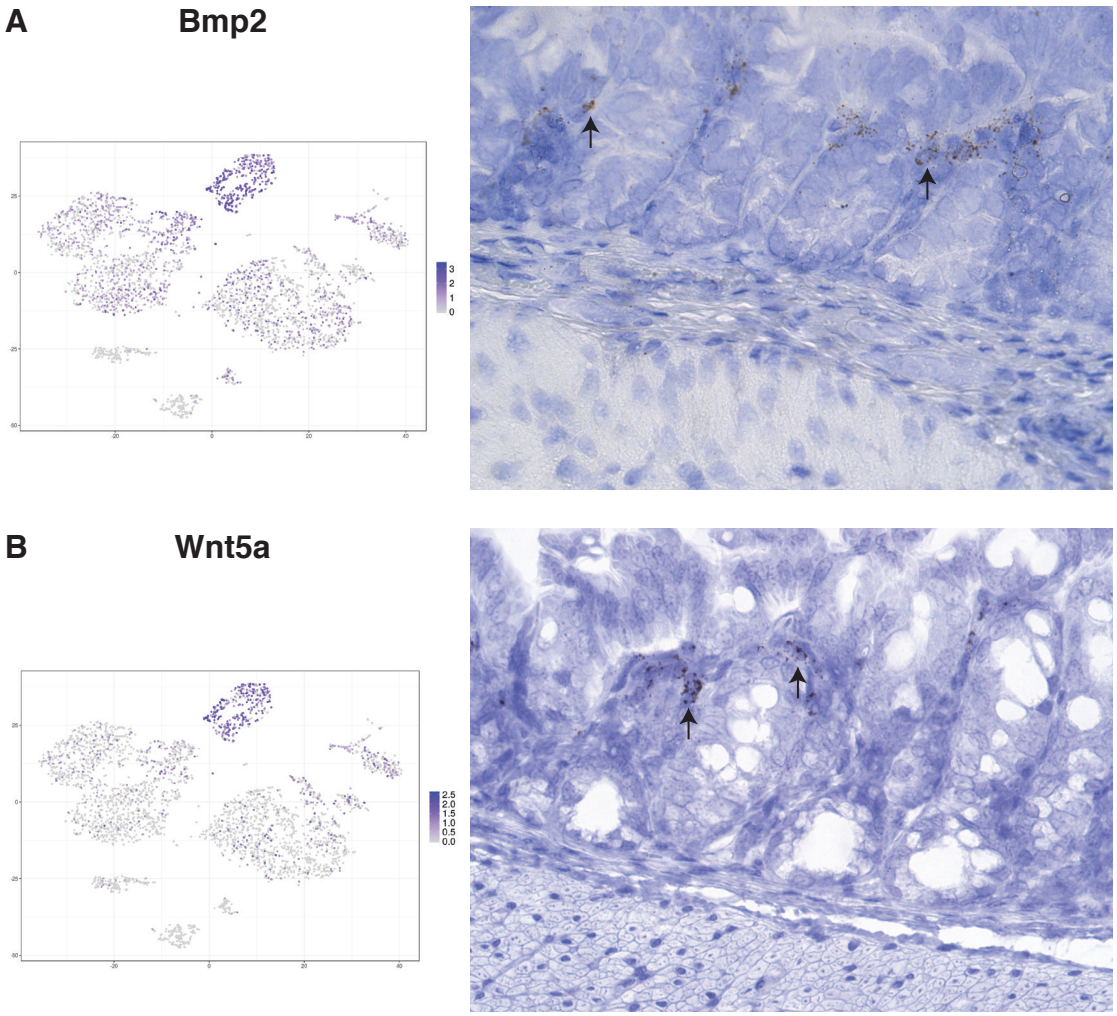


Figure 5.7: smISH localisation of Stromal 2 marker genes. t-SNE projections (left panels) of the murine healthy dataset. Cells are coloured by normalised expression levels of the indicated marker genes. Single-molecule ISH labelling (brown, right panels) using probes directed to the same markers genes in healthy murine colonic tissue sections counter-stained with haematoxylin. Positively staining cells are indicated (arrows).

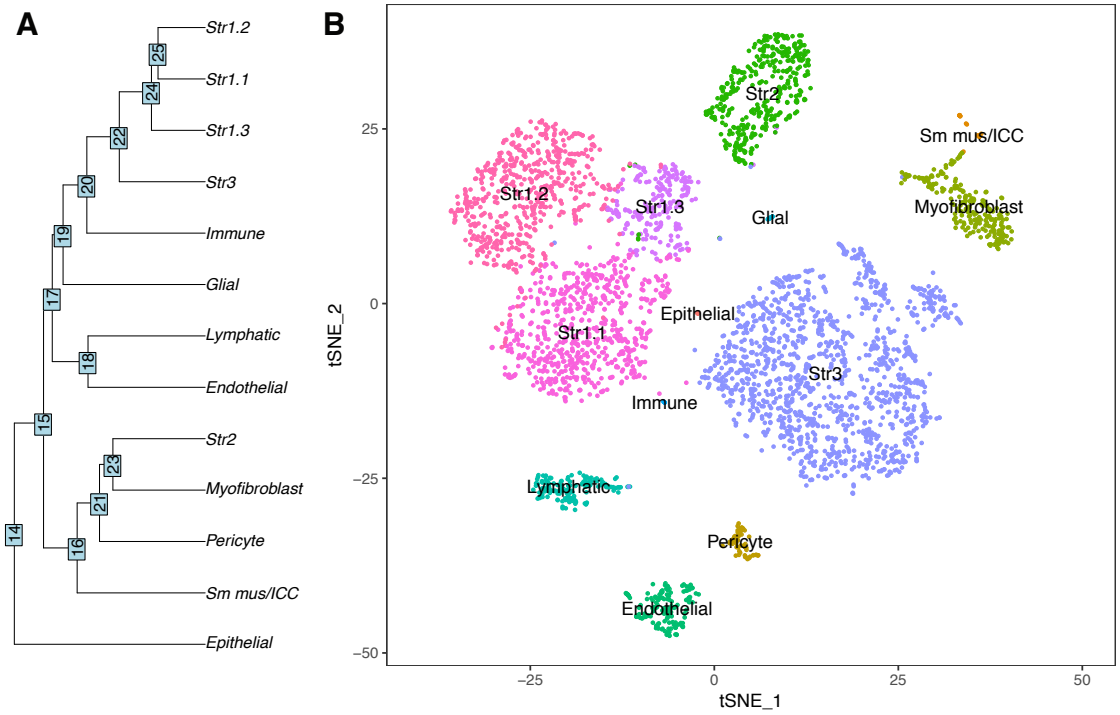


Figure 5.8: Phylogenetic tree and identities of murine stromal cell clusters in health. (A) Phylogenetic tree of murine clusters representing inter-cell distances between the average cells for each cluster in gene expression space. (B) t-SNE plot of murine healthy dataset. Single cells are coloured by cluster annotation. Descriptive cluster labels are shown replacing numeric identifiers.

5.2.6 DSS-induced changes in stromal composition

The clustering (Appendix C, Figure C.12A) and cluster merging (Appendix C, Figure C.12B,C) steps were repeated for the 3,354 cells sequenced from the three mice that were subjected to DSS colitis challenge. 11 clusters were identified and assigned numeric identifiers as before (Figure 5.9A). A random forest classifier was then trained using the healthy dataset and used to classify cells from the DSS animals (Figure 5.9B). The results of clustering and random-forest classification were cross-tabulated to determine the identities of the DSS clusters with reference to the clusters seen in health (Appendix C, Figure C.13). The DSS endothelial, lymphatic, pericyte, myofibroblast, Stromal 1.1, and Stromal 2 clusters were unambiguously identified using this approach. The cells comprising DSS cluster 11 were mostly classified as Stromal 1.2, albeit with a minority classified as Stromal 1.1. DSS clusters 4–7 were all classified as Stromal 3, suggesting that heterogeneity within

the Stromal 3 group is increased in the presence of DSS-driven inflammation.

Using the random forest classifier described above, the changes in stromal subtype composition associated with DSS challenge could be quantified across biological replicates. There was a significant increase in the relative abundance of Stromal 3 cells, from 34% to 47% of the dataset (Figure 5.10). No significant changes in abundance were detected for the remaining subsets.

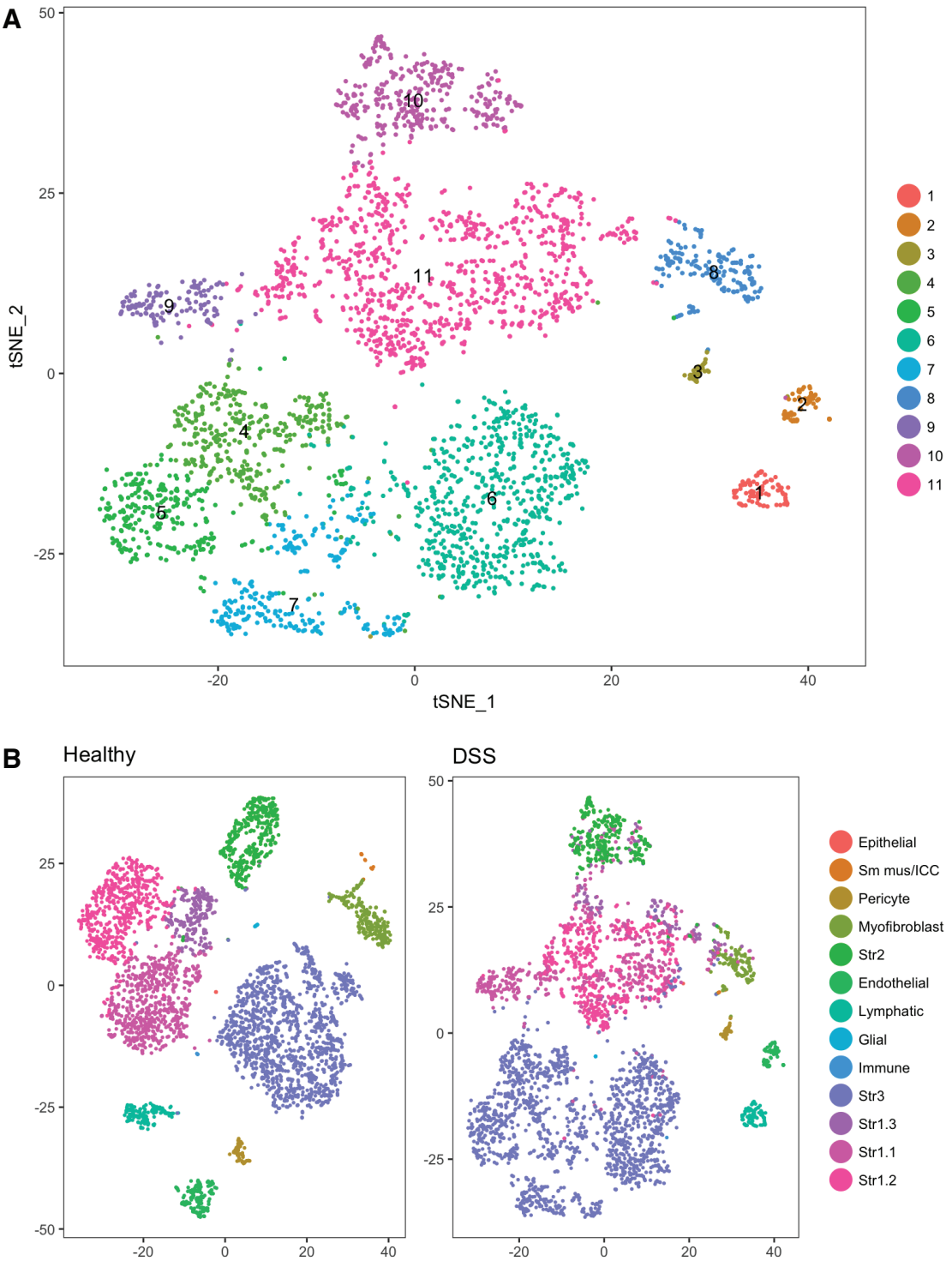


Figure 5.9: Cell clustering and cluster identity in DSS colitis.

Figure 5.9 ***(previous page)*****:** (A) A t-SNE projection of 3,354 single cells derived from 3 mice following a DSS colitis challenge. Cells are coloured by cluster annotation, and clusters are assigned numeric identifiers as before. (B) A random forest classifier trained using the healthy dataset (left panel) was used to classify cells from DSS-challenged mice (right panel). From this, the identities of clusters in the DSS dataset were inferred (see Appendix C Figure C.13).

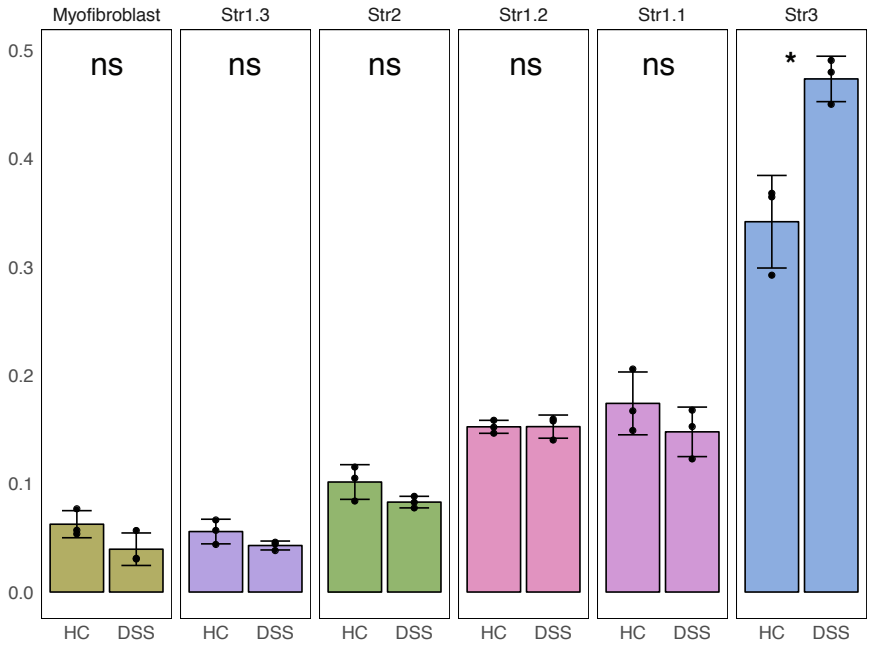


Figure 5.10: An increase in the relative abundance of the Stromal 3 subset is seen in DSS colitis. The size of each fibroblast cluster (column facets) expressed as a proportion of the total number of cells was compared across the three biological replicates for healthy controls (HC) and DSS-challenged mice (DSS). Individual data points, mean and standard deviation are shown. DSS challenge resulted in a significant increase in the fraction of Stromal 3 cells (Student's t-test p value = 0.02).

5.2.7 Expansion of a Stromal 4-like fibroblast subset on DSS challenge

The UC-specific human Stromal 4 subset showed expression of a number of marker genes associated with interaction with immune cell subsets and transcriptional similarity to the fibroblastic reticular cells found in lymphoid tissues. To determine whether an analogous subset was present in the context of a murine DSS challenge, a marker cross-tabulation approach was employed as before (Figure 5.11A). There was significant overlap between the murine orthologs of Stromal 4 markers and

DSS cluster 7. Among the 13 shared marker genes identified were both the FRC-associated chemokine *Ccl19* and the IL-1 family alarmin *Il33* (Figure 5.11B,C). The forest classifier had previously shown cluster 7 to be one of the Stromal 3 subclusters, suggesting that Stromal 4 cells may arise from a Stromal 3 progenitor in the context of inflammation.

An annotated t-SNE representation and phylogenetic hierarchy were now constructed for the DSS dataset. Stromal 3 sub-clusters were assigned decimal suffixes as before with the exception of cluster 7 which was labelled as Stromal 4 on the basis of its overlap with that human subset (Figure 5.12).

The relative increase in the size of the Stromal 3 cluster (comprising subclusters Stromal 3.1, 3.2, 3.3 and 4) on DSS challenge could be due either to differential rates of proliferation or cell loss among stromal subtypes. To investigate the former, a cell-cycle classifier was utilised to annotate the predicted cell cycle stage of each cell in the dataset. Cell cycle scores for G1 and G2/M phases are determined for each cell using a large panel of gene pairs known to exhibit cell-cycle stage specific expression in murine cell lines. This method has been shown to be robust in multiple datasets, although its limitations include a lower accuracy in discriminating S-phase and an inability to distinguish resting G0 cells from cycling cells in G1 phase[256]. As expected, the majority of colonic fibroblast-like cells in both treatment groups were determined to be in the G0/G1 phase (Figure 5.13A, B). There was, however, and increase in G2/M annotated cells on DSS challenge. The pericytes, vascular and lymphatic endothelial cells showed the highest proliferative activity. Among fibroblast-like clusters, the highest G2M proportions were observed in the Stromal 3 and 4 subsets, suggesting that proliferation accounts at least in part for their increased abundance (Figure 5.13C).

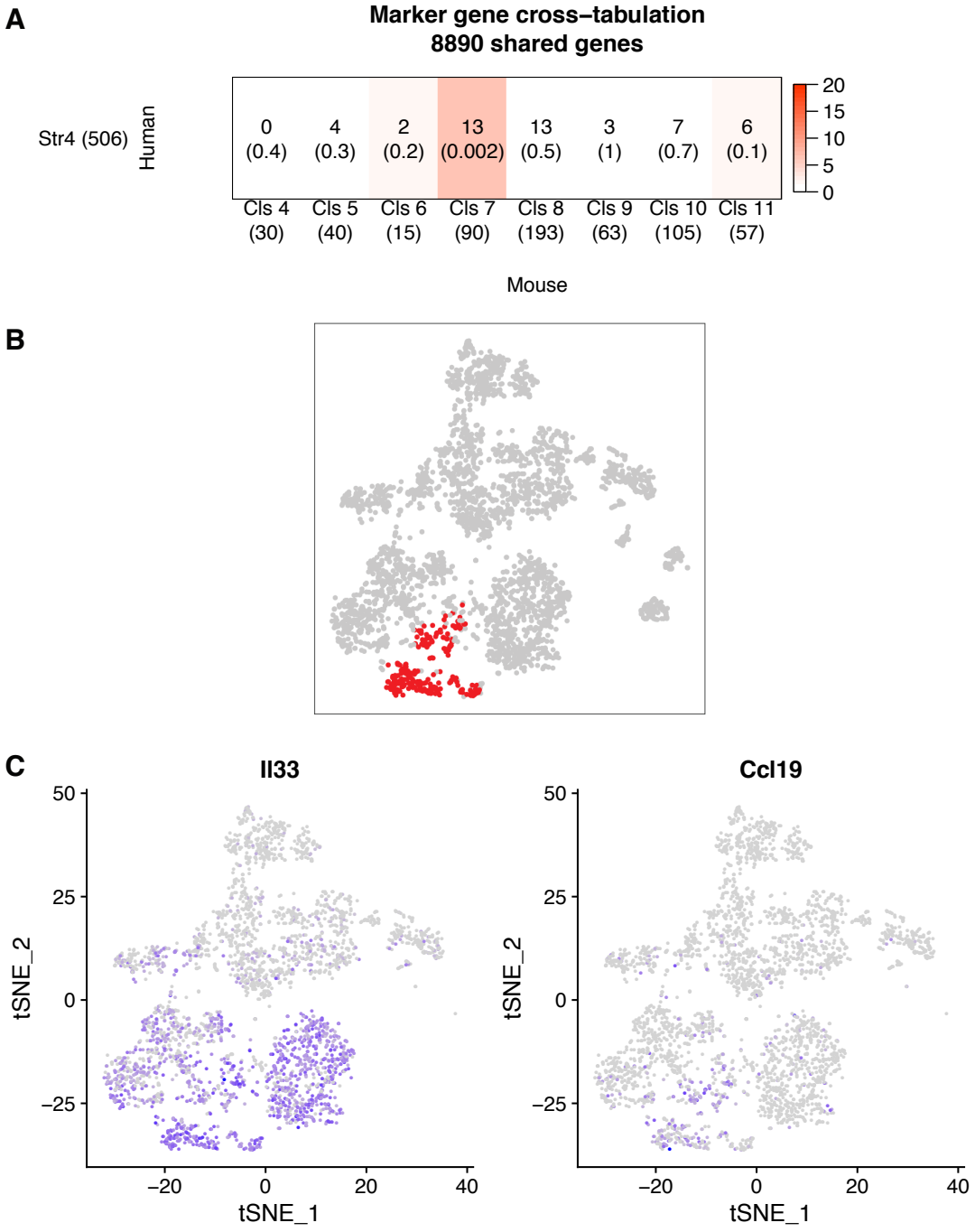


Figure 5.11: An immunologically specialised fibroblast subset analogous to human Stromal 4 is identified in the murine DSS model. (A) Cross-tabulation of human Stromal 4 marker genes against marker genes for the 8 clusters of fibroblast-like cells identified in the DSS dataset. The number of shared markers and p-value (Fisher's Exact Test) are shown. Colour scale $-\log$(p-value). (B) Location of Cluster 7 (red) within the TSNE representation of the DSS dataset. (C) Expression of Stromal 4 marker genes *Il33* and *Ccl19* within the murine DSS dataset.

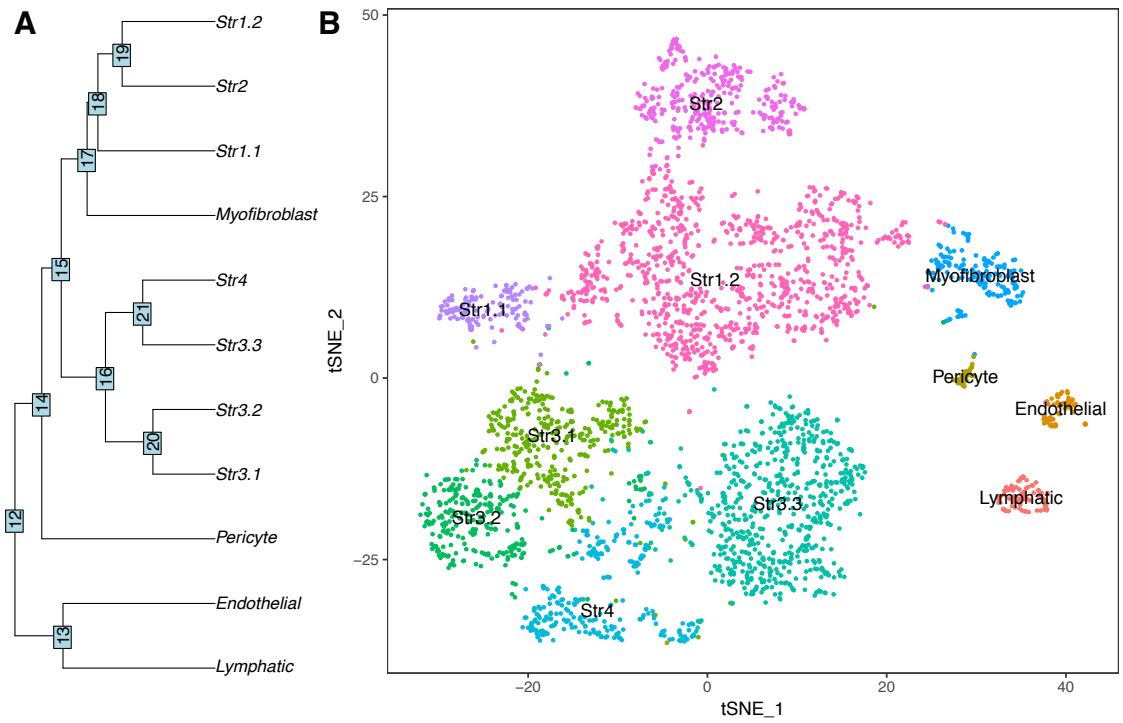


Figure 5.12: Phylogenetic tree and identities of murine stromal cell clusters in DSS. (A) Phylogenetic tree of murine clusters representing inter-cell distances between the average cells for each cluster in gene expression space. (B) TSNE plot of murine DSS dataset. Single cells are coloured by cluster annotation. Descriptive cluster labels are shown replacing numeric identifiers.

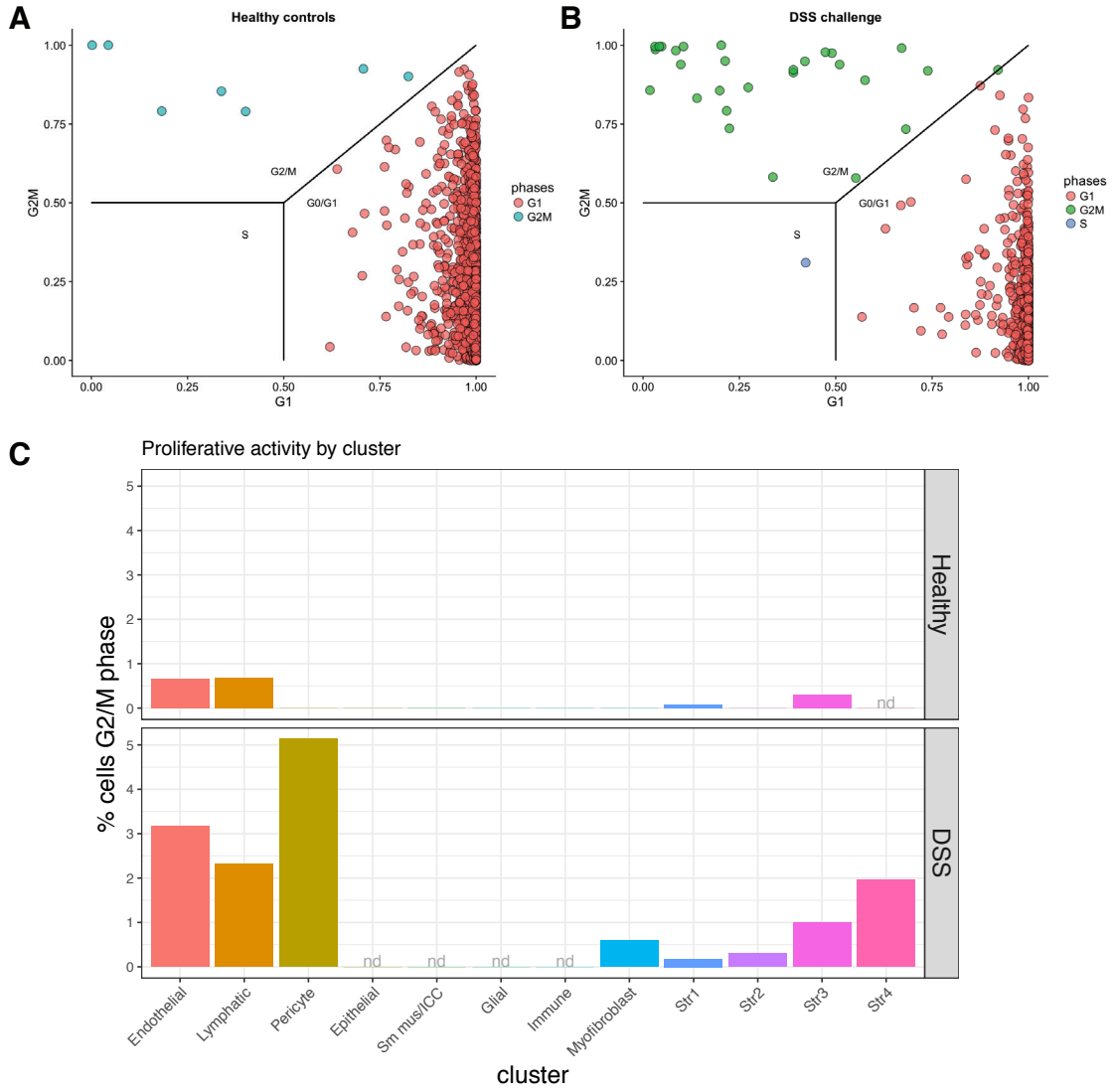


Figure 5.13: Fibroblast subsets show differential proliferative activity on DSS challenge. (A,B) Cell cycle phase annotation for the healthy and DSS datasets using a pre-trained murine cell cycle classifier (cyclone, 'pairs' method). (C) Percentages of cells in G2M phase by cluster (nd = no equivalent cluster detected in dataset).

5.2.8 Biological activities of murine stromal subsets in health and DSS colitis

Functional ontology enrichment was performed using the top 100 marker genes (ranked by the AUROC method) for each fibroblast cluster detected in the control treatment group. Selected enrichment results are shown in Figure 5.14. Enrichment for terms related to established fibroblast functions, such as 'extracellular matrix organisation' and 'cell-substrate adhesion' was detected in multiple subsets (Figure 5.14A). Other shared terms aligned with specific fibroblast activities in the colonic mucosa, such as regulation of the epithelial compartment and cell migration, which is of relevance to the positioning of immune cells.

Differential ontology enrichment was also observed, supporting the hypothesis that the transcriptomic clustering structure has captured important functional specialisation. As in the human datasets, myofibroblasts enriched predominantly for contractile processes, while the stromal 2 group enriched for TGFβ responses and BMP signalling, which is a major driver of epithelial differentiation (Figure 5.14B)[80]. Interestingly, the Stromal 1.2 and 2 groups showed enrichment for positive and negative regulation of Wnt signalling respectively. Wnt signalling is a major determinant of epithelial proliferative activity[1], and the opposing actions of these two subsets may be a mechanism for maintaining homeostasis. The enrichment for 'ameboidal-type cell migration' seen in Stromal subsets 1.1 and 1.3 is also interesting. Fibroblast migration has been described by multiple methods[35, 57], but a lack of reproducibility has been problematic[59].

The large Stromal 3 cluster shows increased heterogeneity on DSS challenge, leading to the detection of distinct sub-clusters Stromal 3.1, 3.2, 3.3 and 4. Functional ontology enrichment was performed using marker genes for these sub-clusters to identify candidate biological processes driving this diversification (Figure 5.15). A overlapping progression of ontology enrichment was observed. The stromal 3.1 group showed no unique enrichment, while Stromal 3.2–4 enriched sequentially for synthesis of nitric oxide and reactive oxygen species (ROS), inflammatory cytokine production, and proliferation and lymphocyte activation. This represents a plausible

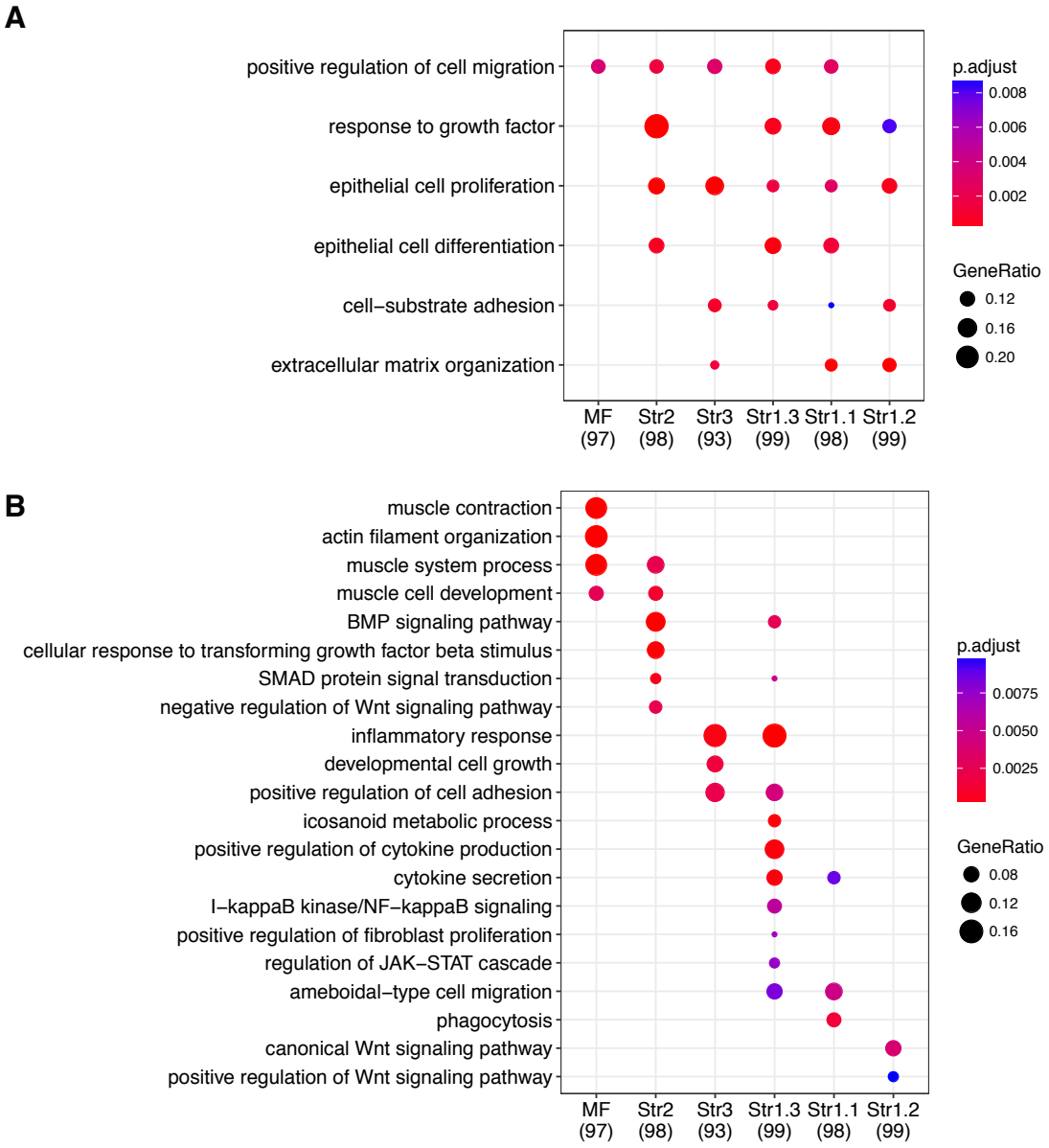


Figure 5.14: Common and differential functional ontology enrichment among fibroblast subsets seen in health. (A) Selected biological process ontology terms showing significant enrichment in multiple fibroblast subsets. (B) Selected ontology terms showing subset-specific enrichment. Circle size indicates the proportion of marker genes annotated, colour scale reflects adjusted p-value.

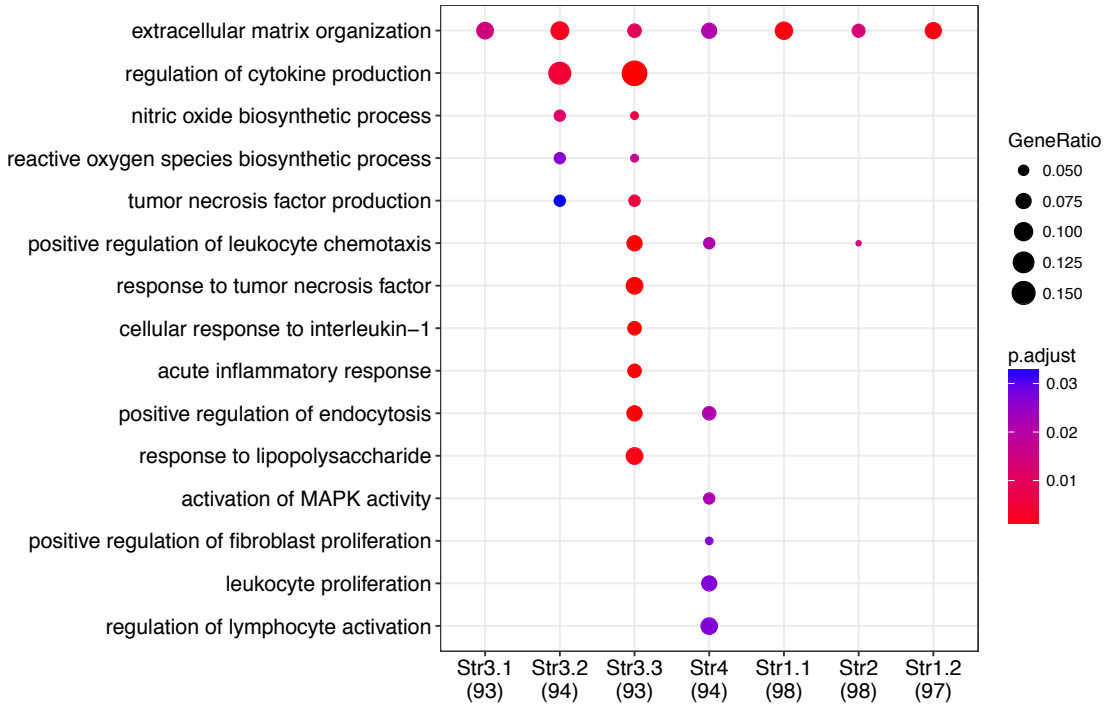


Figure 5.15: Progressive enrichment for stress, activation and proliferation seen within Stromal 3 subset in DSS. (A) Selected biological process ontology terms showing significant enrichment within the Stromal 3.1, 3.2, 3.3 and 4 subsets in the DSS dataset. Circle size indicates the proportion of marker genes annotated, colour scale reflects adjusted p-value.

fibroblast activation sequence. Cellular ROS production is required for TGFβ induced myofibroblast differentiation in a murine model of idiopathic pulmonary fibrosis[338], while stromal nitric oxide produced in response to inflammatory cues regulates lymphocyte activation in lymphoid tissues[339–341].

5.2.9 Ontogeny of stromal fibroblast subsets

Phylogenetic trees and overlaps in ontology enrichment provide some insight into the relatedness of the detected fibroblast subsets, but cannot be used to construct a differentiation hierarchy. Neither is it possible to take serial RNAseq measurements from the same cell and directly observe changes in transcriptional state over time. Bioinformatic approaches have been developed to address this problem. These are based on the assumption that if enough cells distributed along a developmental trajectory are sequenced at a single time point, cells in transition between stable states will be captured. The trajectory can then be reconstructed by ordering cells

along a path that minimises the transcriptional differences between neighbours. Diffusion pseudotemporal ordering in one implementation of this approach that can recover branched developmental trajectories[258]. It first requires construction of a diffusion map, a non-linear weighted nearest neighbour graph of the dataset based on the probabilities of transitioning between pairs of cells over several paths of a random walk[257].

Non-fibroblast cell types were first removed from the control dataset as these were considered unlikely to form part of the same developmental hierarchy. The myofibroblast cluster was also removed, as this clustered separately from the remaining fibroblasts in diffusion map space and intermediate forms were not observed at this sampling density. The remaining clusters (Str1.1, Str1.2, Str1.3, Str2 and Str3) produced the branched structure shown in Figure 5.16A. While any of the three vertices (populated by Str1.1, Str2 and Str3 cells) could represent the origin, Stromal 3 was considered the most likely candidate given its expansion and proliferative activity on DSS challenge. Diffusion pseudotime was therefore calculated from this point, with Stromal 1.2 and 1.3 representing intermediates, and Stromal 1.1 and Stromal 2 the fully differentiated states. Using this model, it is possible to identify genes such as *Ebf1*, *Thy1* and *Adamdec1* predicted to show sequential activation as well as genes specific to one or other developmental trajectory such as *Ptgs2* and *Il33* (Figure 5.16B).

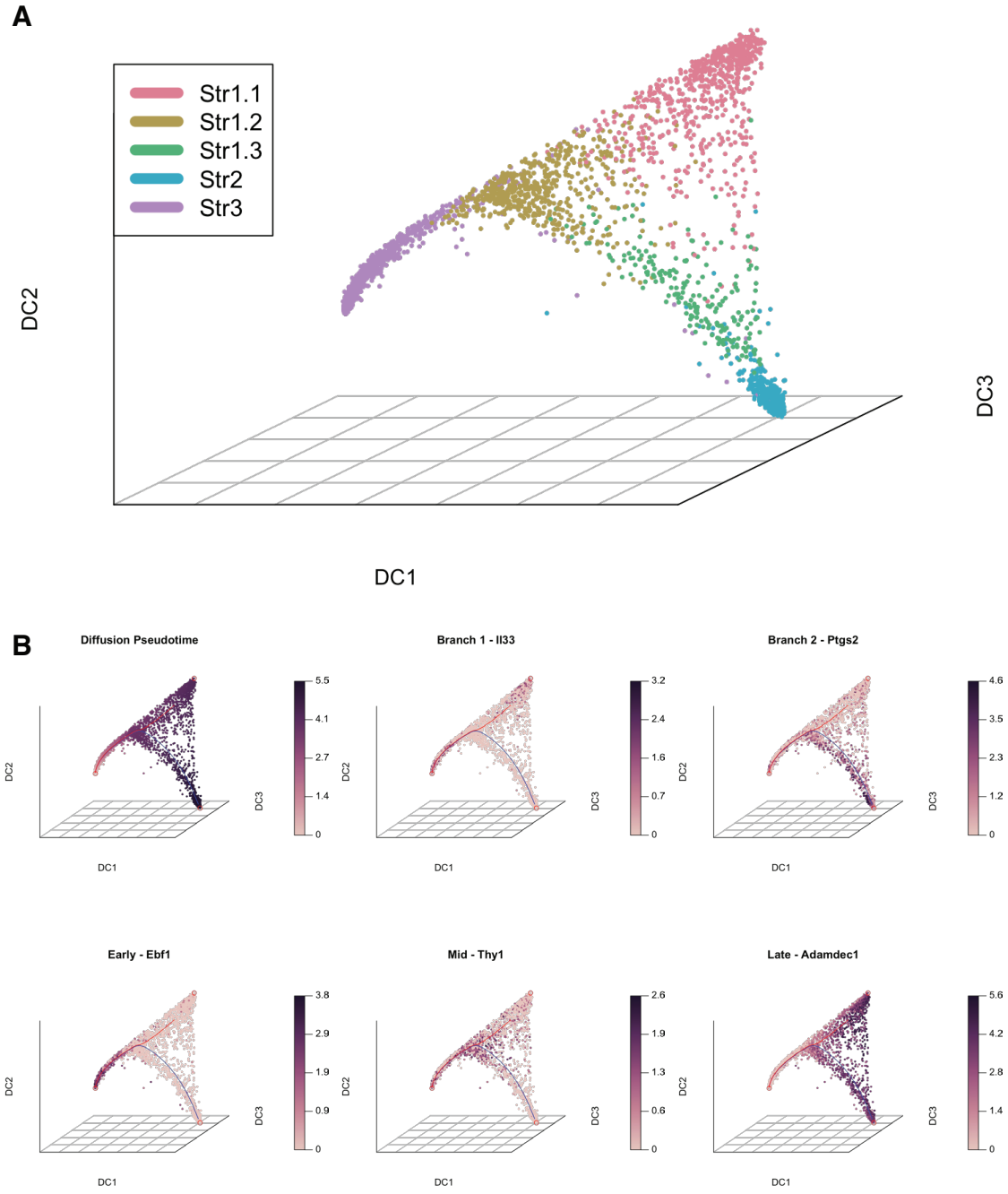


Figure 5.16: Diffusion pseudotime provides a framework to relate observed fibroblast clusters. (A) Diffusion component plot for colonic fibroblast-like cells from healthy mice. Individual points represent single cells coloured by cluster annotation. (B) Application of diffusion pseudotime to diffusion component plot. Cells are coloured by pseudotemporal order. Alternate differentiation paths are indicated (red and blue lines). Expression dynamics of selected markers are shown as indicated.

5.2.10 Subset-specific transcriptional responses to DSS challenge

Having identified corresponding clusters of fibroblast-like cells in the healthy and DSS datasets (Appendix C, Figure C.14), differential expression analysis was performed using a negative binomial differential expression test. The transcriptional responses to DSS challenge were markedly dissimilar between fibroblast subsets, supporting the hypothesis that the detected clusters are functionally distinct. Stromal 1 upregulated genes included *Grem1*, a BMP-antagonist implicated in promoting epithelial proliferation and crypt fission[310], *Il11*, an anti-inflammatory cytokine that supports epithelial proliferation and is protective in an experimental model of erosive colitis[342], and the matrix metalloproteinase *Mmp3* (Figure 5.17A). The Stromal 2 cluster showed strong induction of the chemokine *Cxcl13*, chemotactic for B-lymphocytes[343], lipocalin 2 (*Lcn2*), an anti-microbial peptide that enhances phagocytic bacterial clearance and is protective in colitis[344], and the metalloproteinase inhibitor *Timp3* (Figure 5.17B). The Stromal 3 group showed strong induction of the pro-inflammatory cytokine *Il6*, together with *Il33* and the soluble pattern recognition receptor Pentraxin 3 *Ptx3*, which has a protective role in innate immunity (Figure 5.17C)[345].

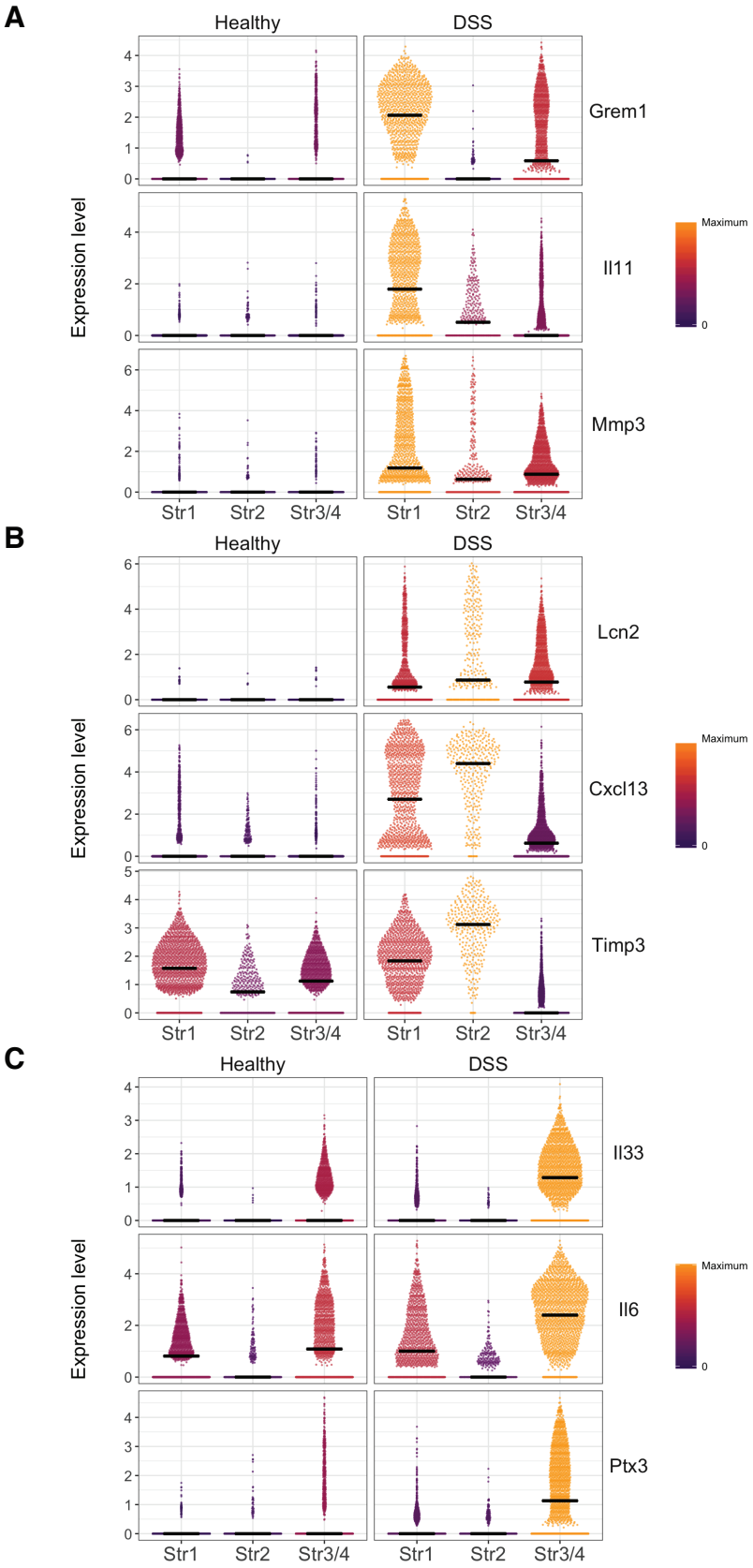


Figure 5.17: Fibroblast subsets show differential responses to DSS challenge.

Figure 5.17 ***(previous page)*:** Violin plots for the indicated genes, which were significantly induced on DSS challenge in stromal subsets 1–3 (A—C). Individual cells are represented as points. A colour scale reflecting the row-normalised mean expression is applied. Crossbars are plotted to indicate cluster median expression.

5.2.11 Specificities of existing fibroblast markers and novel marker candidates

The expression of previously proposed fibroblast subset markers was examined within the control dataset (Appendix C, Figure C.15). The mesothelial marker *Wt1* showed expression within a small, localised subpopulation of Stromal 3 cells[56]. The proposed myofibroblast marker *Aoc3* was detected predominantly in the myofibroblast and pericyte clusters[33]. *Ptgs2* and *Cd44* showed similar patterns of expression, as has been previously suggested[35]. Maximal expression localised to the interface of the Stromal 1.3 and Stromal 2 groups. *Pdgfra* expression was maximal in Stromal 2. α-SMA (*Acta2*) was maximally expressed in smooth muscle, with progressive reductions in expression in the myofibroblast, pericyte and Stromal 2 groups. This is consistent with the experimental observation of distinct populations of $Pdgfra^{+}$ and $\alpha\text{-SMA}^{+}$ cells, presumably Stromal 2 and myofibroblasts, in the pericryptal sheath[31]. *Foxl1* expression, proposed to identify stromal cells contributing the the epithelial stem cell niche, was localised to myofibroblasts and Stromal 2[40]. *Thy1* (CD90) was maximally expressed in lymphatic endothelial cells, with the Stromal 1.2 group showing intermediate expression of this marker[30].

The specificities of existing stromal Cre recombinase systems were examined (Figure 5.18A). *Myh11* targeted smooth muscle and myofibroblasts, *Cspg4* (NG2) pericytes and *Fap* Stromal 3. The remaining genes examined showed broad expression across multiple subsets. By ranking subset marker genes by specificity, novel candidates for construction of selective models targeting individual stromal subsets could be proposed (Figure 5.18B).

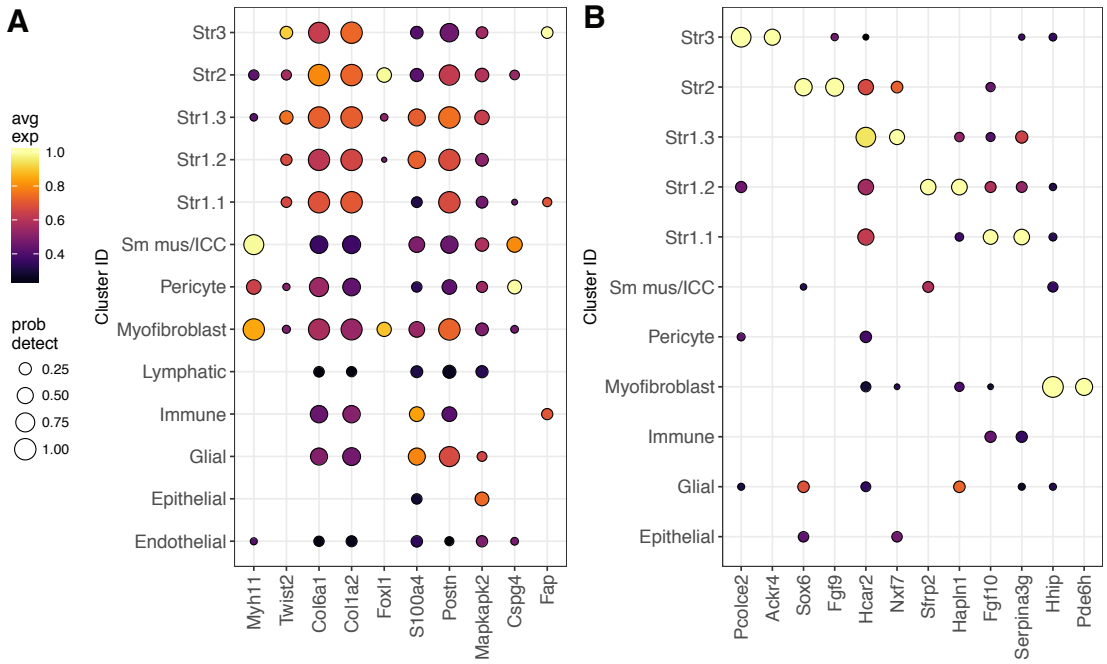


Figure 5.18: Candidate molecular markers for further subset characterisation and functional study. Specificity of candidate marker genes (x-axis) for detected fibroblast subsets. (A) Existing or previously proposed markers. (B) Markers showing high subset specificity in this dataset. Circle size represents the within-cluster probability of gene detection, while the fill colour represents the normalised mean expression level.

5.3 Discussion

The application of high-throughput single cell RNA sequencing to colonic stromal cells derived from healthy mice and mice with DSS-induced colitis corroborates and extends the classification structure derived from human samples. The transcriptomes of over 7,000 single cells were used to generate a comprehensive inventory of non-haematopoietic, non-epithelial cell types. These were predominantly fibroblast-like cells, with a smaller contribution from the endothelial cells and pericytes associated with the colonic vasculature and lymphatic system. Non-fibroblast cell types, including rare cells such as glial cells and cells that are not fully characterised in this setting such as pericytes, can be confidently identified by reference to specific marker genes previously published in the literature and consistent ontology enrichment. This provides a valuable internal control, demonstrating that the isolation and bioinformatic methods are successful in capturing the aspects of

cellular transcriptional variation that are relevant to determining cell identity. It follows that the structure seen in the fibroblast clusters could also be used as a basis for assignment of distinct cellular identities.

There was significant marker gene overlap between all the stromal subsets detected in human colonic tissue and corresponding clusters in the murine dataset. Preservation was strongest in the Stromal 1, Stromal 2 and myofibroblast groups. Validation experiments using smISH demonstrated localisation of Stromal 2 cells adjacent to the upper crypt epithelium as in human tissue. A number of technical factors may have contributed to the weaker preservation observed in the Stromal 3 and 4 groups, including the fact that the starting tissue was not equivalent (mucosal biopsies versus whole colons) and the isolation protocols, sequencing chemistry, transcript quantification method, and sequencing depth all varied between experiments.

The changes in stromal composition in the context of inflammation were partially reproduced in the murine dataset. A significant increase in Stromal 3 abundance and the emergence of a Stromal 4-like cluster were again observed. The UC-associated loss of Stromal 2 cells was not reproduced in the murine DSS model. The fact that an acute DSS challenge was used, in contrast to the chronic inflammation associated with UC may go some way to explaining this.

With increased cell number, additional substructure could be resolved. Diffusion pseudotime was used to project cells along a hypothetical developmental trajectory and showed strong agreement with the clustering structure generated using the Seurat community detection algorithm. Fibroblast subsets could be annotated as progenitors (Stromal 3), intermediates (Stromal 1.2, Stromal 1.3) or differentiated states (Stromal 1.1, Stromal 2) using this approach.

While the diffusion pseudotime model requires experimental validation, a number of external observations support it as a plausible framework for fibroblast differentiation in the colon. Early expression of the transcription factor Early B-cell factor 1 (Ebf1) is predicted. Ebf1 has been identified using murine knockout models as a critical regulator of osteocyte and adipocyte differentiation[346]. In addition, the

small population of fibroblast-like cells expressing the mesothelial marker *Wt1* are located close to the temporal origin, potentially connecting the observed fibroblast subsets to the mesothelium, which has been shown by lineage tracing to be the likely source of colonic fibroblasts both in development and in the adult[55].

The transcriptomic landscape revealed in these experiments offers a basis to reconcile some of the fibroblast subsets proposed in the literature, with *Pdgfra*, *Ptgs2*, *Aoc3*, *Foxl1* and *Thy1* all showing differing subset specificities. The need for caution in interpreting fibroblast α-SMA positivity is apparent, with three distinct subsets (myofibroblasts, pericytes and Stromal 2) all showing intermediate levels of expression of this widely used marker. Finally, specific marker genes are proposed, which could be used in the construction of selective ablation or lineage tracing models to further define the distinct functions and developmental relationships between fibroblast subsets.

Chapter 6

Concluding Discussion

Contents

6.1 Introduction 160
6.2 Generation of interpretable stromal single-cell transcriptomes from primary human and murine tissue samples 162
6.3 Development of an unsupervised cell classification method for single cell RNAseq data using WGCNA and NMF 163
6.4 Single cell RNAseq identifies aspects of colonic stromal gene expression not accessible to bulk methods . 166
6.5 Multiple transcriptionally and spatially distinct fibroblast subsets are detected within the human colonic lamina propria 167
6.6 Changes in stromal composition and gene expression associated with ulcerative colitis 169
6.7 Similar fibroblast transcriptional heterogeneity is observed in a murine model 171
6.8 Defining the stromal response to DSS-induced acute colitis 173
6.9 Specificities of existing murine Cre models and identification of new candidates 175
6.10 Wider significance — stromal biology 175
6.11 Wider significance — Inflammatory Bowel Disease . . 176
6.12 Future directions 177

6.1 Introduction

From their original description as a structural sheath surrounding the epithelial crypt, there has been considerable progress in describing the specialisation and diverse

functions of the fibroblast-like cells of the intestinal lamina propria. Ultrastructural and immunohistochemical studies have demonstrated surprising heterogeneity among these cells[6, 35, 36]. In parallel, functional studies have defined important roles in the regulation of epithelial development and the maintenance of immune cell populations at the intestinal barrier[107, 108, 116, 138]. Their importance in the development of multiple disease states including fibrosis, IBD and cancer is increasingly appreciated. The focus of this work has been to apply recent technological advances in single-cell RNA sequencing to interrogate stromal biology in the intestinal lamina propria. The data generated provides a detailed insight into the identity and activities of these cells that is not accessible using bulk approaches. It provides a basis for a transcriptional classification of fibroblast like cells that complements existing work using electron microscopy and antibodies to candidate marker proteins.

There are a number of challenges associated with the study of stromal cells that single cell RNAseq is well suited to address. Perhaps foremost among these is a lack of specific protein markers that can be used to reliably identify these cells. Much of the existing literature relies on non-specific markers such as α-SMA or negative staining for the markers for other cell types. Important exceptions exist in the form of elegant studies showing, for example that PDGFRA, PTGS2 and AOC3 can all be used to identify subsets of fibroblast-like cells with specific distributions and properties[31, 33, 35]. Similarly, functional aspects of fibroblast biology such as the roles of fibroblasts in directing epithelial development and inflammatory responses have been demonstrated in part through the use of a small number of fibroblast-selective Cre drivers[76, 192]. The specificities of these within the fibroblast compartment are not fully determined. This work proposes a model of stromal transcriptional subsets. Candidate marker genes are identified, and a selection of these have been validated by independent methods in human and murine tissue. It also provides a comprehensive overview of the complement of the stromal compartment and with it a basis for comparing the specificities of current Cre models and relating new and previously proposed fibroblast subsets to one another.

In addition to revealing fibroblast heterogeneity and providing a basis for classification, the broad perspective of cellular activity offered by single-cell RNAseq provides a basis for further study of important areas of fibroblast function in this locale. Key areas of interest include the signalling pathways mediating cross-talk with the epithelial and immune compartments, the mechanisms by which fibroblasts sense inflammatory stimuli, and their repertoire of effector responses. Specific controversies within the field are also be addressed, including the identification and characterisation of distinct sub-epithelial fibroblast and contractile myofibroblast populations as has been suggested by some authors[13, 31]. Finally, through the use of pseudotemporal ordering, a model of fibroblast differentiation is proposed linking the various fibroblast subtypes to a likely progenitor population with mesothelial characteristics[55, 56].

6.2 Generation of interpretable stromal single-cell transcriptomes from primary human and murine tissue samples

The intestinal stromal cell network and epithelial compartment represent a challenging substrate for single cell RNA sequencing. In their native states, both cell types are firmly attached to their neighbours and ECM components. Isolation protocols for use in other applications such as flow cytometry and cell culture have been published, but the efficacy of these methods in generating cell suspensions suitable for single cell RNAseq has not been demonstrated[229, 253]. Difficulties encountered include the small amount of starting material available from endoscopic forceps biopsies which are the only practical source of primary human colonic tissue, the need for rapid isolation of viable single cells and separation of these from non-cellular debris, and enrichment of the generated cell suspensions for the target populations in absence of specific surface markers. The protocols described here resulted in acceptable capture efficiency using both the Fluidigm C1 (viable cells captured versus possible capture sites) and 10X Genomics (actual versus predicted cellular GEM number) platforms. Efficiency ranged from approximately 30% for human primary colonic

epithelial cells to greater than 60% for stromal cells. The quality of the resultant single cell libraries was comparable to other published datasets in terms of numbers of genes or molecules detected[106, 224, 287, 289, 333]. While there are no absolute quality parameters that can be applied across dissimilar experiments, the quality thresholds selected are of a similar or greater stringency to those used elsewhere.

There is considerable interest in applying single-cell RNAseq to phenotype human disease states at a single cell level[347]. The protocols outlined here offer a useful starting point for this undertaking using clinical samples derived from the colon. The main barriers that remain to be overcome before datasets can be generated from large clinical cohorts are the need to preserve samples while retaining RNA integrity to allow sample pooling and batch processing, and the per-sample costs associated with single cell capture. The use of RNA fixatives such as dithiobis(succinimidyl propionate) (DSP) offers a possible solution to the problem of cell preservation, and pilot experiments were performed using this agent in combination with the Fluidigm C1 platform[296]. While some loss of library complexity and increased RNA degradation was observed, sufficient information remained to clearly distinguish primary stromal cells from a human cell line. Further optimisation of this aspect of the project is ongoing. The per-sample capture cost can be significantly reduced by employing a recently described bioinformatic approach that utilises expressed SNPs as 'biological barcodes' to determine the donor of origin for single cells from multiple human donors captured in a single pooled 10X reaction[348]. Pilot experiments using this method with human colonic samples are also in progress.

6.3 Development of an unsupervised cell classification method for single cell RNAseq data using WGCNA and NMF

One of the most common applications of single cell RNAseq in the literature is in cell classification problems — specifically the resolution of complex cell mixtures into distinct clusters with defined transcriptomic characteristics. RNAseq provides a far more information about individual cells than alternate methods,

such as IHC, flow cytometry and CyTOF, when applied in this way. However, the data generated is challenging from a statistical perspective. Specific difficulties encountered include the high levels of noise in single cell data due to phenomena such as transcriptional bursting, the prevalence of drop-out events, where even relatively abundant transcripts may evade capture and amplification due to the inefficiency of reverse transcriptase, and biases introduced during multiple rounds of amplification[349]. These aspects exacerbate the general statistical problems associated with high-dimensional datasets, the so-called 'curse of dimensionality', characterised by the tendency as dimensionality increases for all objects to become dissimilar in many ways[350].

Multiple approaches to clustering single cell RNAseq data have been described[222–225, 254, 287, 288]. A general principle is the identification of a small subset of genes that are most informative in ascribing cell identities. These might take the form of genes with higher-than-expected heterogeneity in expression, genes with high weightings in important principle components, or genes that have been previously identified as specific markers of particular cell types[224, 254, 287, 288]. There is no consensus regarding the best approach and factors such as the total number of cells sequenced, the quality of the resultant libraries, the extent of biological heterogeneity present in the sample, and the genetic baseline similarity between biological replicates (if present) all influence the chance of success of a particular method.

The datasets generated in this study are challenging because cell numbers achievable with the C1 platform are relatively low, library quality for human primary cells digested from clinical tissue samples is inevitably lower than for suspension cells or cell lines, fibroblast-like cells show greater transcriptional similarity than is present in, for example, the epithelial or haematopoietic compartments, and human donors have far higher background heterogeneity than congenic mouse strains. Neither variance nor PCA-based dimensionality reduction methods were successful in these datasets. Instead a combination of WGCNA and NMF was employed to first identify modules of genes showing correlated expression and then classify cells based on the combination of modules they expressed. While they are general methods, both

WGCNA and NMF have been used with success in single cell RNAseq datasets[246, 283–285, 290]. However, this is the first example of their application in combination to a cell classification problem. WGCNA has been used widely in gene expression datasets and experimental evidence suggests that genes forming co-expression modules often share important biological characteristics such as participation in the same pathways or protein complexes[351, 352]. Classification of cells based on the sum of their important biological processes is intuitively appealing. Furthermore, this approach successfully recapitulated the expected heterogeneity in colonic epithelial cells, identifying cell clusters characterised by expression of well-known markers of epithelial progenitors, differentiated enterocytes, and secretory cells.

While the combination of hierarchical clustering, spectral decomposition and rank correlation implemented in the R package SC3[248] produced a similar clustering structure in the epithelial dataset, it should be noted that this method was not available at the time these experiments were performed and that it fails to identify subgroups within the epithelial progenitor cluster. Similarly, the PCA-space neighbourhood detection algorithm implemented in Seurat[254] was also not available until the latter stages of this project and is in any case of limited use in small datasets due to the difficulty in identifying clusters with fewer than 30 cells. This problem is of less significance when working with larger datasets, such as those generated here using the 10X genomics platform, and Seurat was preferred in these cases due to its close integration with the 10X Genomics pipeline, computational efficiency, and comprehensive array of analysis and visualisation functions. Together, these offer considerable time savings over a bespoke approach. Nonetheless, a niche exists for the WGCNA / NMF method developed here even as single cell datasets increase in size. Very rare cells, such as adult stem cells, may not be accessible in large numbers from clinical samples even with high throughput methods. In addition, WGCNA can be used identify important biological processes occurring across multiple cell types and the 'hub' genes that co-ordinate them, complementing cell classification approaches in the construction of an integrated understanding of the biological functions of heterogenous tissues.

6.4 Single cell RNAseq identifies aspects of colonic stromal gene expression not accessible to bulk methods

Using single cell RNAseq, it has been possible to generate a detailed gene expression profiles of colonic mucosal epithelial cells and fibroblast-like cells. The technique is more sensitive than previous approaches based on bulk gene expression analysis using microarrays or RNAseq followed by screening of 'hits' by IHC or ISH to determine whether they localise to particular cell types[92]. A number of interesting facets of fibroblast activity have been uncovered. Firstly, it is clear that fibroblast-like cells show global increased expression of genes associated with innate immune sensing, antigen processing and presentation relative to the epithelial compartment. This is consistent with earlier work within the Simmons group and by others showing that cultured mucosal fibroblasts possess functional innate sensors and are able to present antigen[116, 152]. The significance of these pathways *in vivo* in the maintenance of immune homeostasis at the intestinal barrier remains to be determined. However, this could be explored using fibroblast-specific knockout models for key pathway intermediates. The datasets generated here provide candidates genes for use both as fibroblast-specific drivers and targets for abrogation. For example, the innate sensor TLR4 and the NOD pathway signalling intermediary RIP2 were identified as showing significantly increased expression in the mesenchymal compartment. Global murine knockout models for both of these genes have been generated and show defects in pathogen handling[353, 354]. Conditional knockouts for use with the Cre-loxP system are available in commercial repositories as are Cre-drivers such as Col1a2-CreER, predicted to possess broad specificity for mucosal fibroblasts.

Exploration of epithelial-mesenchymal crosstalk was achieved by screening the respective single-cell datasets for cognate ligand-receptor pairs. This produced both well known axes such as BMP and HGF signalling and new hits such as the Ephrin A pathway and interleukins 7, 15 and 18. The biological functions of these axes could be further explored using cell-type specific knockout models as above or by exposing

epithelial organoids to activating ligands, blocking antibodies, or co-culture with transgenic fibroblasts. Initial experiments using the latter approach are in progress.

High expression of certain disease-relevant genes was also detected in mucosal fibroblast-like cells. The secreted metalloproteinase ADAMDEC1 is SNP-associated with IBD, and murine knockout models show increased susceptibility to colitis[264]. Its expression has previously been localised to mononuclear cells within the lamina propria, presumed to be macrophages. This work demonstrates that it is abundantly expressed in a large proportion of mucosal fibroblast-like cells. 'Conditional-ready' murine knockout strains for this gene are again available and could form the basis with an appropriate driver to explore the functional significance of ADAMDEC1 expression in the fibroblast compartment.

6.5 Multiple transcriptionally and spatially distinct fibroblast subsets are detected within the human colonic lamina propria

Using the unsupervised WGCNA / NMF approach to analyse single-cell RNA-seq data derived from human colonic stromal cells, these cells formed four distinct clusters that resembled myofibroblasts and distinct subtypes of fibroblasts with divergent functions. The myofibroblast cluster was distinguished by expression of contractile proteins. The candidate myofibroblast marker AOC3 showed specificity for this group as did TGFBR3, a SMAD1-induced gene associated with the development of myofibroblasts from fibroblasts, whose inhibition aggravates fibrosis[302]. The stromal 1, 2 and 3 groups all expressed known fibroblast markers such as vimentin, but appeared functionally and anatomically distinct.

Cells expressing stromal 1 genes were localized deep within the lamina propria, not directly adjacent to epithelia. Aside from expressing genes with a role in maintenance of the extracellular matrix, Stromal 1 cells showed the highest expression of the transcription factor ZEB2, which has been shown to direct epithelial to mesenchymal transition (EMT) and to act as a cell fate switch in haematopoietic cells. Multi-lineage differentiation and progenitor production is perturbed upon

ZEB2 deletion[355]. Other markers for this subset included SLIT2, ABI3BP and CXCL12. SLIT2 is secreted by fibroblasts to prevent monocytes from transitioning to a pro-fibrotic state in tissue termed fibrocytes[356]. ABI3BP is an extracellular matrix protein necessary for the switch from proliferation to differentiation in mesenchymal stem cells[357]. In the gastrointestinal tract, blocking the CXCL12-CXCR4 axis attenuates colitis[358]. Further functional studies are required to define the dominant role of Stromal Group 1 cells in the lamina propria.

This work suggests that a specialised subgroup of stromal cells exists to direct epithelial cell function in the intestine. The stromal 2 group is a dominant candidate for this role, as it shows increased expression of non-canonical WNT ligands and of members of the TGFβ superfamily, suggesting an interaction with epithelial cells. Stromal 2 cells expressed SOX6, ZEB1, FRZB, WNT5A and were located proximal to epithelial cells in colonic tissue sections. The controlled turnover of the intestinal epithelial layer is important in avoiding aggregation of damaging mutations induced by the intraluminal microenvironment and is maintained in balance by opposing proliferative and differentiation-inducing pathways. Many of the stromal 2-associated morphogens, including WNT5A, BMP2 and BMP4, have been individually shown to promote differentiation of epithelial cells[79, 309]. It is therefore likely that the net effect of Stromal 2 activity in health is in constraining the size of the epithelial proliferative compartment.

Stromal 2 cells showed the highest expression of PDGFRA and are highly compatible both in location and marker expression to the proposed PDGFRA$^+$ stromal subset that has previously been identified[31]. An interesting relationship between PDGFRA$^+$ fibroblasts, α-SMA$^+$ myofibroblasts and the colonic crypt was described in the same paper, with PDGFRA$^+$ cells adjacent to the upper crypt, and myofibroblasts occupying the equivalent position at the crypt base. Previous work has also demonstrated that hedgehog signalling from the epithelium to the mesenchyme is essential to the maintenance of distinct proliferative and differentiating zones and acts by reciprocal induction of TGFβ superfamily ligands from the mesenchyme[93]. This work suggests a refinement to this proposed

mechanism, with upper-crypt stromal 2 cells expressing a repertoire of differentiation-inducing morphogens, including members of the TGFβ superfamily, and lower-crypt myofibroblasts expressing high levels of the hedgehog inhibitor HHIP.

The stromal 3 population expressed high levels of genes involved in innate sensing and in the synthesis of prostaglandin E_2, a critical regulator of intestinal barrier repair following injury[359]. This population also expressed inhibitors of signaling cascades that would induce pro-inflammatory genes; for instance, the MAPK inhibitor TRIB1[360] and TNFAIP3, which inhibits NF-κB activation as well as TNF-mediated apoptosis[361], suggesting that this group plays an immunoregulatory role. High expression of PTGS2 and MAP3K8 in this group together with enrichment for innate signalling make it somewhat compatible with a mesenchymal $PTGS2^+$ cell population that has been previously identified and implicated in the preservation of epithelial proliferation during tissue injury[35].

6.6 Changes in stromal composition and gene expression associated with ulcerative colitis

Differential expression analysis was performed comparing the single cell transcriptomes of colonic fibroblast-like cells derived from patients with ulcerative colitis to those from healthy donors. Genes showing the highest magnitude differential expression were enriched for known epithelial morphogens. The greatest fold-change increase in expression was seen in the case of the BMP antagonist GREM1, which has previously been shown to induce epithelial crypt budding and fission[310]. This process is required for epithelial restitution following mucosal injury. Expression of GREM1 is restricted in the steady state to a small subset of mesenchymal cells in close proximity to the crypt base, and ectopic expression in the epithelial compartment leads to Hereditary Mixed Polyposis Syndrome (HMPS) characterised by the development of mixed-morphology colorectal tumours. The generalised upregulation of GREM1 in the mesenchyme in the context of acute UC may represent a physiological repair mechanism. Given the known association between dysregulated GREM1 expression and carcinogenesis, it was interesting also to

observe elevated mesenchymal GREM1 expression in patients with longstanding UC and UC-associated dysplasia even in the absence of active inflammation.

Patients with UC shared the same four colonic stromal subsets seen in healthy individuals. However, they also displayed a new subset that was more highly represented in inflamed tissue compared to adjacent non-inflamed tissue (stromal 4). This subset shared some transcriptional hallmarks with fibroblastic reticular cells (FRCs), a critical mesenchymal component of secondary lymphoid structures[168]. The degree of epithelial damage in colonic inflammation correlates with the number and size of sub-epithelial lymphoid follicles[330]. Thus, the appearance of stromal 4 may correspond to the formation and expansion of lymphoid follicles in non-inflamed and inflamed colonic tissue, respectively.

Stromal 4 showed increased levels of transcriptional regulators STAT1, ATV6 and TP53, previously implicated in inflammation and cancer. Expression of STAT1 in cancer-associated fibroblasts promotes tumour progression[362] and is associated with upregulation of PD-L1[363]. Stromal 4 was also enriched for immune effector genes including CCL19 and CCL21, ligands for CCR7 that help attract immune cells[364], and for genes previously implicated in colitis, such as IL-33, IL-32, JAK3, TNFSF14 and TNFSF15. IL-33-deficient mice are highly susceptible to colitis and colitis-associated cancer[365], and IL-33 promotes Treg function in colitis[146]. TNFSF14 (LIGHT) signals through the lymphotoxin β receptor in the colon to regulate the innate immune response and mediate recovery from intestinal inflammation, and *Tnfsf14*$^{-/-}$ mice develop more severe colitis than control mice[366]. TNFSF14 has also recently emerged as a potential target for therapeutic interventions aiming to halt tissue fibrosis. The induction of genes involved in mucosal protective responses hints that expansion of stromal 4 in UC may be important in attenuating tissue inflammation observed in acute stages of the disease.

While single-cell RNAseq provides an overview of heterogeneity within the stromal compartment and indirect evidence of functional specialisation, there is clearly much work to be done in defining the true natures of these stromal subsets and their ontogenies. It should also be noted that the structure presented here

is based on relatively small numbers of sampled cells. The ability to discriminate transcriptional clusters may be greatly improved by increasing the sample size, an undertaking made ever more accessible by technological progress in this area. The true transcriptional heterogeneity within the stromal compartment may be an order of magnitude greater than that presented here.

A number of extensions to this work using human primary cells are ongoing. These include the use of subset-specific surface markers to FACS separate freshly isolated colonic stromal cells for co-culture with human immune cell subsets and colonic epithelial organoids and the development of a 'stromal classifier' CyTOF panel. This will comprise markers for each subset and is being developed to quantify changes in stromal composition in larger cohorts of clinical samples from healthy individuals and those with IBD, as has recently been performed for resident immune cell types[367].

6.7 Similar fibroblast transcriptional heterogeneity is observed in a murine model

Recapitulating the complex tissue environment of the colonic mucosa with human primary cells in tissue culture systems is inherently challenging, and the generation of an equivalent murine dataset to guide construction of subset-specific ablation and lineage tracing models provides an alternative approach. The emergence of high-throughput single cell sequencing platforms, such as the 10X Genomics® Chromium™ Single Cell 3′ Solution, allows the numbers of cells sampled to be increased by 1–2 orders of magnitude[333], and a decision was made to switch to this approach rather than continuing with Fluidigm C1. This was not entirely without compromises as some aspects of the Fluidigm pipeline including the ability to inspect capture sites for multiple capture and cell viability are no longer possible. In addition, transcripts are quantified by 3′-end counting as opposed to full length sequencing, limiting the power to detect splice variants or distinguish gene isoforms that share a common 3′ end. The sequencing depth and library complexity per

cell were also reduced compared to the earlier approach, although it was hoped that greater cell numbers would compensate for this.

The positive identification of multiple expected cell types based on expression of previously validated marker genes provides a useful internal control for these experiments. Pericytes, smooth muscle cells, lymphatic and vascular endothelial cells were all readily identified as distinct clusters together with small numbers of contaminating epithelial and immune cells. The same unsupervised approach again revealed multiple separate fibroblast and myofibroblast cell clusters, supporting the idea that there is greater heterogeneity among these cells than has been previously appreciated. In the case of the stromal 2 and myofibroblast subsets, highly significant overlap was observed by marker gene cross-tabulation between the murine and human datasets suggesting that these groups are conserved between species. While it remained statistically significant, the overlap in marker genes was less convincing for the stromal 1 and 3 subsets. This may be due to interspecies differences, but it should be noted that several important aspects of the experimental design were altered between the human and murine experiments including the samples used (forceps biopsies versus whole colons), the isolation protocol, and the bioinformatic approach. Experiments analysing human samples using the 10X genomics pipeline have been performed and analysis is in progress at the time of writing. This will give a better idea as to the extent to which fibroblast diversity is conserved.

As expected, increased cell number provides a parallel increase in the power to distinguish fibroblast transcriptional states[349]. The stromal 1 and 3 groups, numerically the largest groups in the human samples, showed overlap with multiple murine subsets with differing ontology enrichment profiles. For example, the murine stromal 1.2 group showed unique enrichment for the positive regulation of canonical Wnt signalling, a major pathway promoting the maintenance of epithelial stem cells[1, 101]. The small subset of stromal cells that contribute to the epithelial stem cell niche may lie within this group. Establishing the tissue distribution of the stromal 1.2 subset using candidate surface markers would be a useful initial step in testing this.

The increased sampling density afforded by the 10X platform also allowed modelling of the relationships between fibroblast clusters. Using the unsupervised diffusion mapping and diffusion pseudotime methods[257, 258], a three-branched structure was generated with stromal groups 3, 1.1 and 2 as its vertices. Although no inference regarding which vertex constituted the temporal origin could be derived from the model itself, a number of observations support the hypothesis that it is the stromal 3 group. These include the observed increases in the proportion of stromal 3 cells and cell cycle activity on DSS challenge, the expression of transcription factors such as Ebf1 that are associated with progenitor potential in other cell types[346], and the presence of a subgroup of stromal 3 cells expressing the mesothelial marker Wt1. This latter observation is of particular interest as there is compelling evidence that stromal fibroblasts have their developmental origin in the mesothelium and are replaced in the adult by the same route[55, 56]. Both mesothelin and Wt1 lineage tracing mouse models exist and in combination with single cell RNAseq would provide an excellent basis for future work defining the precise relationships between the differentiated fibroblast subtypes found in the intestinal mucosa.

6.8 Defining the stromal response to DSS-induced acute colitis

The DSS challenge is a well-characterised model of acute colitis and provides a useful background against which stromal transcriptional changes can be captured using single cell RNAseq[331]. An initial notable finding was of a global increase in library complexity among stromal cells following DSS administration. The median number of unique RNA molecules captured per cell increased by 50% from approximately 8,000 to 12,000. The contribution of technical factors such as RT efficiency and sequencing saturation to this increase was small as evidenced by the similar complexities observed for the control RAW 264.7 murine macrophages that were spiked into each reaction. Globally upregulated genes included epiregulin (Ereg), serum amyloid A3 (SAA3), and CXCL5. Epiregulin is an epidermal growth factor family

member that stimulates fibroblast and smooth muscle proliferation. Epregulin-null mice develop normally but exhibit greatly increased mortality on exposure to DSS[368]. SAA3 is an acute phase protein with possible antimicrobial activity that has been previously evaluated as a biomarker of enteric inflammation[369, 370]. The chemokine CXCL5 is implicated in angiogenesis and neutrophil recruitment and its dysregulation is associated with paediatric onset UC[371]. These observations support the view that fibroblast-like cells are active participants in the generation and resolution of acute inflammatory insults in the colon.

Changes in the structure of the fibroblast population associated with DSS challenge were observed. Increased heterogeneity was seen in the stromal 3 group, with several sub-clusters emerging and enriching sequentially for generation of free radicals, inflammatory cytokine production and interaction with immune cells. Included within this sequence was one cluster sharing significant transcriptional overlap with the stromal 4 group that had previously been observed in samples derived from human donors with active UC. The overall number of shared genes was small and, in addition to the factors outlined above, this may be due to the limitations of acute DSS colitis as a model of human disease. In particular, DSS colitis lacks the chronic phase of inflammation seen in human IBD and is not dependent on lymphocytes[331]. If as has been suggested above, the stromal 4 group represents fibroblasts recruited to nascent lymphoid aggregates, then repeating these experiments with a murine model that recapitulates this aspect of UC pathogenesis would be worthwhile.

Differential expression analysis revealed markedly different transcriptional responses to DSS challenge among different fibroblast subsets. For example, the sub-epithelial stromal 2 group showed strong induction of the matrix metalloproteinase inhibitor Timp3, and only very limited induction of the matrix metalloproteinase Mmp3, which was increased in the other stromal subsets. Precise micro-anatomical control over tissue remodelling could be beneficial in, for example, limiting access to luminal pathogens while enabling migration of immune cells. It may also explain

the apparently contradictory finding that knockout of either of these genes, with their opposing activities, increases susceptibility to colitis[372, 373].

6.9 Specificities of existing murine Cre models and identification of new candidates

Of the mesenchymal-targeting Cre drivers examined in this dataset, few showed high specificities for any particular fibroblast subset as defined by the unsupervised model. The widely used Col1a2 and Col6a1 genes showed similar expression across all stromal subsets, with slightly lower expression in myofibroblasts. The Twist2-Cre model is proposed to target mesenchymal and osteochondral progenitor cells[374], and consistent with this Twist2 was most highly expressed in the Stromal 3 group. Fibroblast activation protein α (Fap) was also selective for Stromal 3. FAP^+ stromal cells appear to suppress anti-tumour immunity in murine pancreatic and lung carcinoma models[5]. Myh11-Cre, which has been used to target myofibroblasts[190], was selective for this group albeit with high expression also noted in smooth muscle cells. The pericyte marker Cspg4 was also selective for its intended target population[20]. Using the single cell gene expression data, it is possible to identify candidate genes for construction of genetic models with superior performance characteristics. One shortcoming of this dataset in this regard, however, is that only very small numbers of epithelial and mucosal immune cells were sequenced. Therefore, the specificity of the suggested targets has not been assessed against a comprehensive catalogue of mucosal cell types. Single-cell RNAseq experiments targeting the epithelial and haematopoietic compartments are underway to address this issue.

6.10 Wider significance — stromal biology

The work presented here supports the view that there is considerable diversity and functional specialisation among fibroblast-like cells in the colonic mucosa. The existence of differentiated fibroblast subsets specialised in interaction with epithelial and immune cells together with progenitors and intermediate states is supported by

an unsupervised cross-sectional analysis of single-cell transcriptional activity. The decision to explore this area owes much to pioneering studies demonstrating similar specialisation among stromal cells from lymphoid tissues[42]. As for the lymph node, functional follow up studies are now required to dissect the precise roles and ontogeny of the stromal subsets proposed here. The identification and validation of specific marker genes will provide a basis for this work. It is highly likely that a similar degree of specialisation is present in the fibroblast networks of many tissues, as evidenced by the recent publication of two studies using similar approaches to examine fibroblast-like cells from the airways[375, 376]. Unravelling the molecular principles of fibroblast differentiation and how these become subverted in diseases such as fibrosis and the fibrotic reactions that occur around some tumours will be a priority for future work in this area.

6.11 Wider significance — Inflammatory Bowel Disease

While IBD can generally be broadly classified as Crohn's disease or ulcerative colitis, these conditions are themselves heterogenous both in their pathological features and clinical course. Biological treatments targeting TNFα, $\alpha_4\beta_7$-integrin, and IL23 are in clinical use, but the remission rates achieved with these agents remain suboptimal even in the context of carefully selected populations in randomised controlled trials[377, 378]. Clinical, endoscopic and histological tools have all been applied to the problem of phenotyping disease, but success has been limited. This project has identified changes in stromal composition and activity associated with both acute UC and longstanding UC with associated dysplasia. An inflammation specific stromal subtype has been described and validated. The tools to extend this approach to a larger clinical cohort either directly using highly multiplexed single cell RNAseq or by surveying a subset of discriminatory markers using an alternate method with higher throughput such as CyTOF are largely in place. The aims of such a survey would be to correlate measures of stromal activation with clinical outcomes with a view to identifying both novel therapeutic targets and disease biomarkers that

could be used to individualise therapy. The stroma is a promising substrate for such an undertaking given the longevity of stromal cells, their capacity for imprinting and memory, and the known utility of existing stromal-derived markers such as oncostatin M in predicting the likelihood of treatment failure[4].

6.12 Future directions

The information provided here provides a basis for the comparative analysis of intestinal stromal cell plasticity in other digestive diseases. It will be of interest to define changes in colonic stromal cell populations that drive precancerous changes in steady state or in the context of chronic inflammation of the intestine, such as in poorly controlled IBD or UC associated with primary sclerosing cholangitis. Excessive Wnt activity is a hallmark of most sporadic colorectal cancers and numerous pathway specific drugs are currently in clinical evaluation for this disease[379, 380]. Despite increasing evidence of the role for tumour-associated stromal cells in cancer progression, there is still much ambiguity with respect to the identification of cancer-associated fibroblasts, myofibroblasts and peri-tumoural fibroblasts in cancer tissue. This work forms a foundation for the comparative analysis of these subsets in model systems and human disease states.

In addition to highlighting novel subsets of intestinal mesenchymal cells and their role in IBD this work has provided molecular information about existing populations that will enable lineage tracing or diphtheria toxin receptor (DTR) based ablation of relevant subtypes to facilitate definition of the contribution of each subset to homeostasis, infection, inflammation and cancer. Future studies of Cre-expressing reporter and fate-mapping mouse lines will illuminate lineage relationships and functions of these different subtypes *in vivo*. Rinkovich et al. previously identified a mesothelial precursor lineage for colonic stromal cells[55]. This work identified Wt1, a mesothelial marker segregating within Stromal 3, a likely progenitor population. Understanding pathways of trans-differentiation will pinpoint mechanisms to enhance specific functional features of these cells and restore tissue homeostasis in diseases such as IBD.

IBD affects 1 in 250 western populations and is increasing in incidence in relation to urbanization. Up to 40% of IBD patients fail to respond to or lose response to conventional immunotherapies. While many new therapies are entering the clinic for IBD a challenge remains in development of models that are truly predictive of their efficacy, as existing approaches more often lead to failure in phase II/III clinical trials. Currently it is mandatory to test novel immunotherapies *in vivo* however often model systems lack direct alignment with human disease leading to failure of phase II/III trials and contributing to the cost of development of new treatments. During the course of this work on human colonic mesenchymal cells, the single cell transcriptomes of cultured primary stromal cells were compared with subtypes isolated fresh from tissue samples, and a gradual loss of key subset specific attributes was observed over time in culture. This emphasizes the shortfalls in current culture systems to accurately recapitulate human physiology but highlights the utility of single-cell phenotyping in design of more representative models. The *in vivo* work demonstrated the commonly used DSS colitis model differs from human UC by showing no significant depletion of the crypt niche mesenchyme and is therefore poorly representative of this aspect of human IBD pathogenesis. This project demonstrates the potential of a non-supervised approach to map differences between murine and human physiology more intricately as a basis to develop improved *in vivo* and *in vitro* 3D tissue models capable of recapitulating aspects of human disease and providing greater predictive value in drug discovery for IBD.

Follow up work arising from this project has demonstrated a correlation in the magnitude of activated Stromal 4 expansion and Stromal 2 depletion with clinical disease severity in UC by CyTOF and illustrates the utility of a combined genomic and proteomic single cell approach in capturing the behaviour of rare mucosal cells in a clinically relevant manner. In future work, it will be possible to develop the CyTOF panel to incorporate key functional attributes of colonic stromal subsets required for immune monitoring to enable more in depth stratification, detection of pre-cancerous states, detection of pro-fibrotic states and prediction of likelihood of response to current immunotherapies in IBD.

Appendices

Appendix A

Supplementary Material for Chapter 3

Contents

A.1 Supplementary figures 180
A.2 Supplementary tables 194

A.1 Supplementary figures

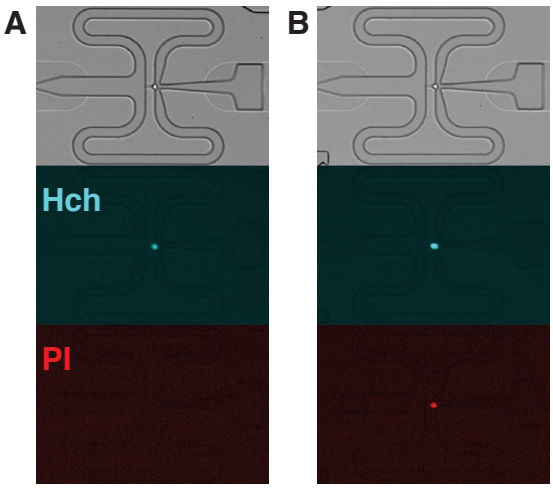


Figure A.1: IFC imaging to detect viable single cell capture. Bright field images (top panel) of Fluidigm C1 capture sites together with Hoechst 33342 (Hch) and propidium iodide (PI) fluorescence demonstrating discrimination of viable (A) from non-viable (B) single cells.

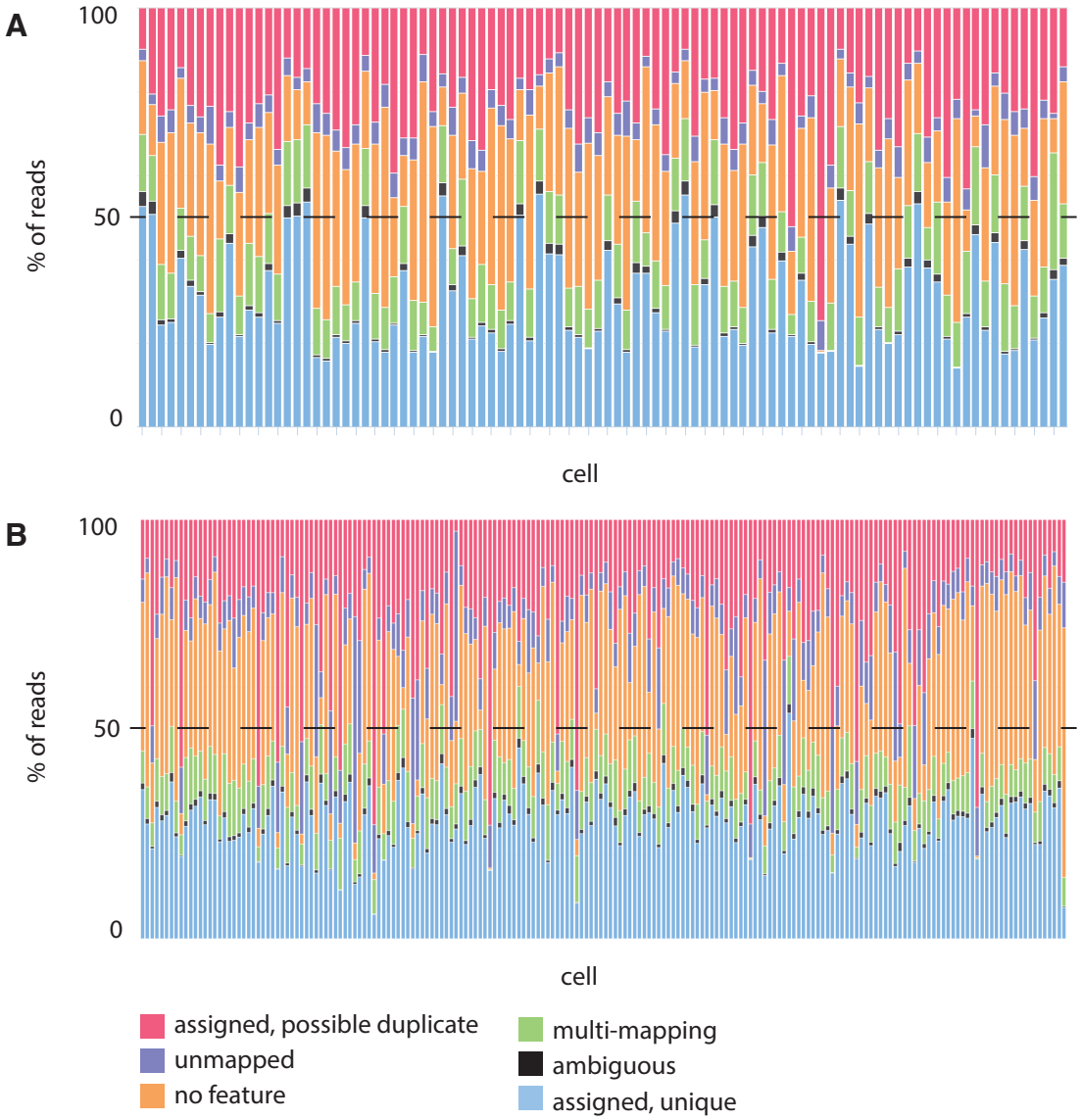


Figure A.2: Assignment of reads to genomic features. Percentages of adaptor-trimmed reads aligning to annotated exonic sequences (assigned), genomic sequences with no associated annotation (no feature), multiple locations, ambiguous locations, or failing to align (unmapped). The assigned reads are further divided into reads with unique start and end positions and reads flagged as possible duplicates. Data for the epithelial (A) and mesenchymal (B) datasets are shown.

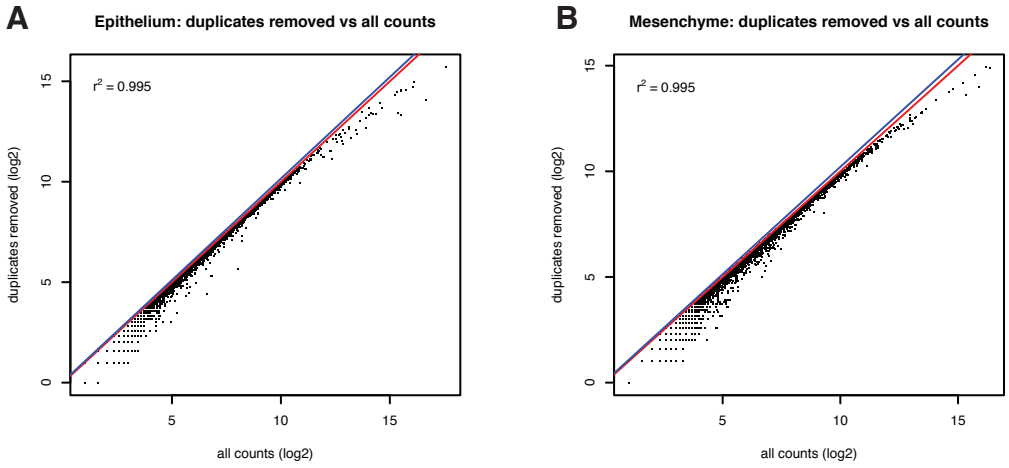


Figure A.3: Effect of duplicate removal on read counts. Random samples of 10,000 non-zero gene expression estimates for the epithelial (A) and mesenchymal (B) datasets. Count values are plotted against the matching value with duplicates removed. The identity line (blue) and line of best fit (red) are shown.

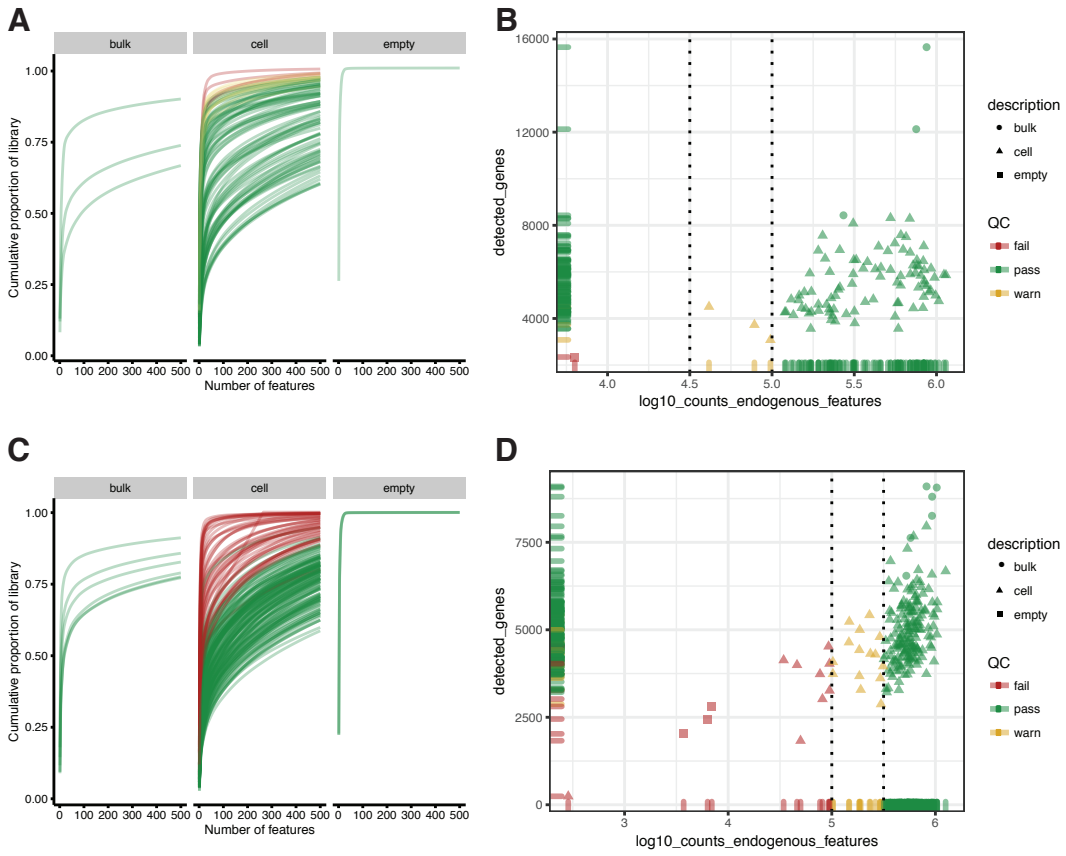


Figure A.4: Single cell library quality measures: depth, coverage and saturation. Quality metrics as indicated for the epithelial (A,B) and mesenchymal (C,D) datasets. Metrics were calculated for single cells (cell), bulk controls (bulk) and empty capture site controls (empty). Single cell libraries are coloured according to whether they passed, failed or generated warnings in the metrics shown. Dotted lines indicate selected thresholds for these outcomes.

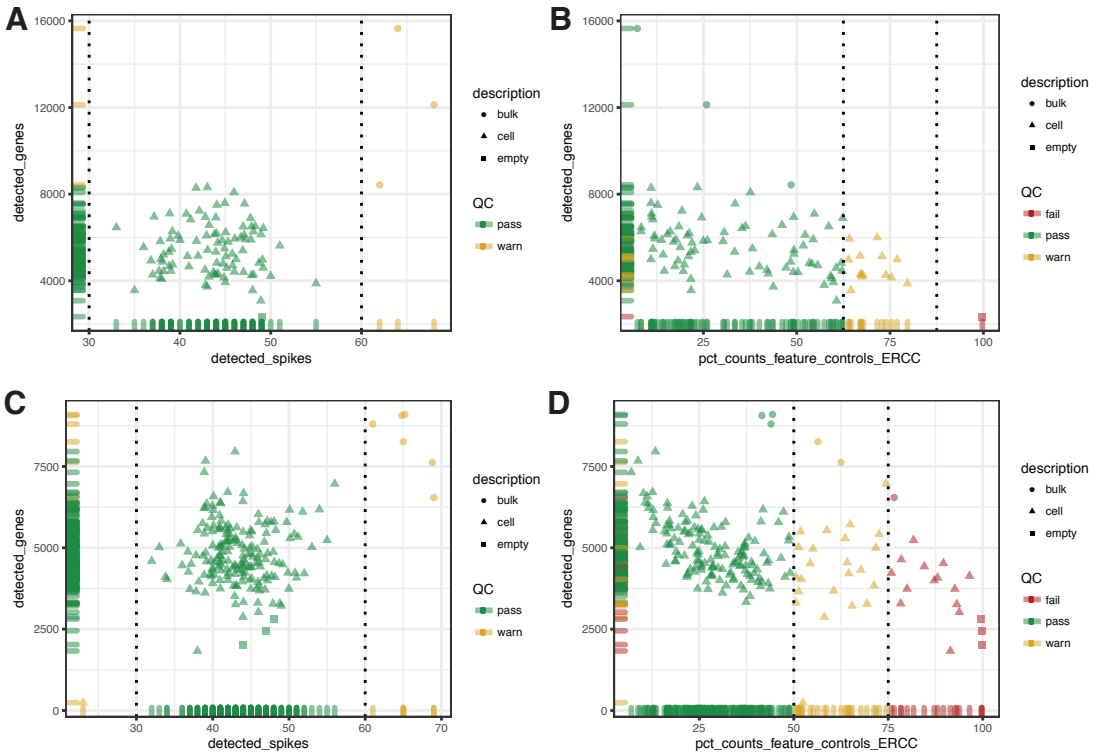


Figure A.5: Single cell library quality measures: spike-in RNAs. Single cell ERCC-based library quality measures for the epithelial (A,B) and mesenchymal (C,D) datasets. Scoring thresholds for the number of unique ERCC molecules detected (detected_spikes) and the percentage of total counts derived from ERCC molecules are indicated (dotted lines). The ERCC concentration used was 4-fold lower in the mesenchymal experiment.

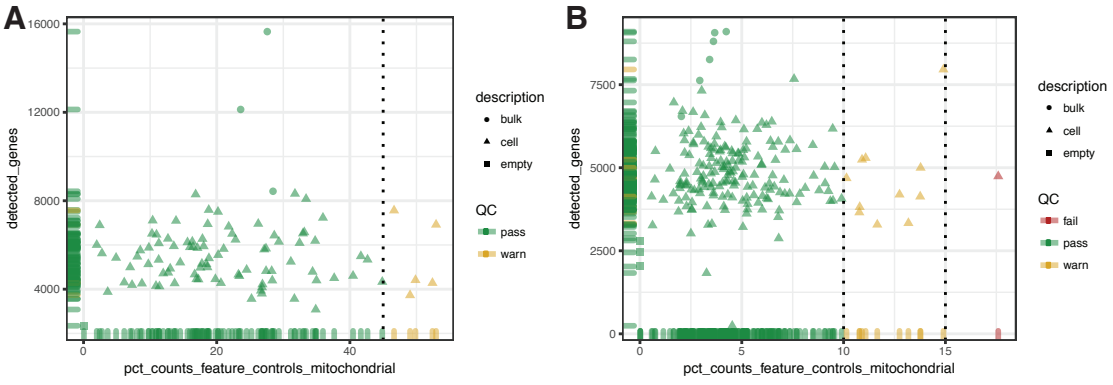


Figure A.6: Single cell library quality measures: mitochondrial genes. Percentages of total counts derived from genes located on the mitochondrial chromosome for the epithelial (A) and mesenchymal (B) datasets. Thresholds used for QC scoring are indicated (dotted lines)

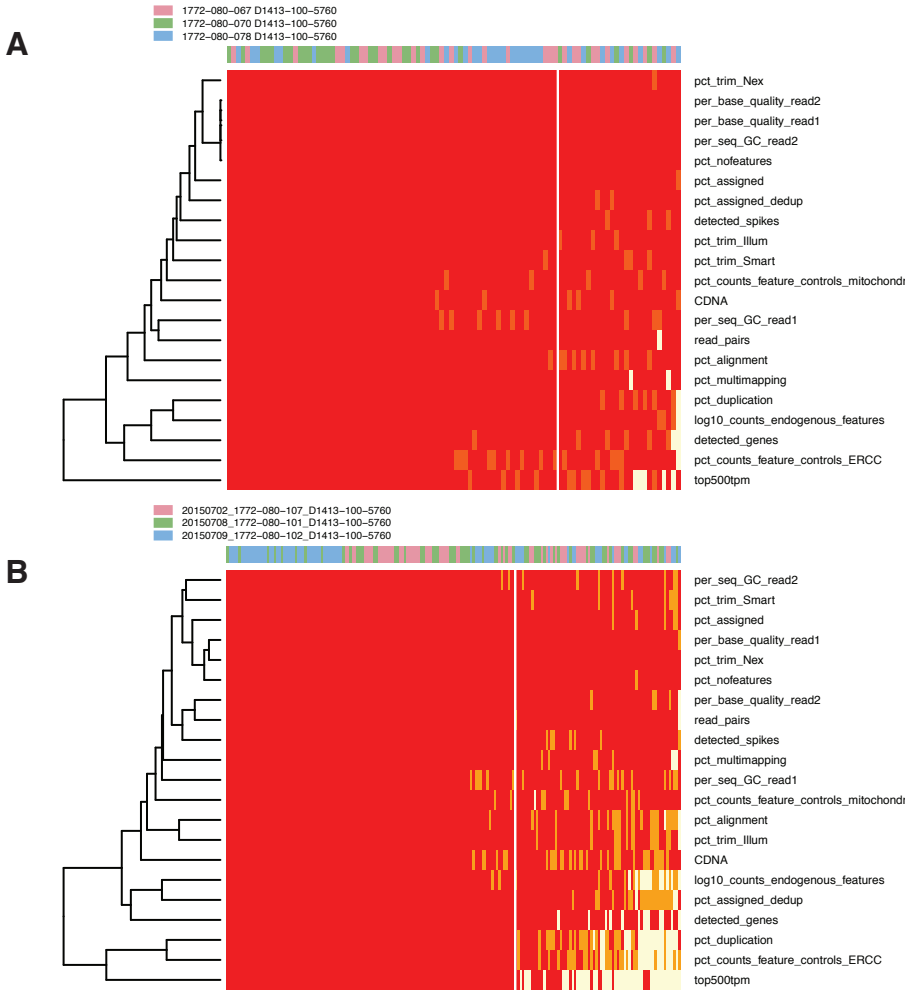


Figure A.7: Cell quality control summary. Scoring for all cells (columns) in all QC metrics evaluated (rows) in the epithelial (A) and mesenchymal (B) datasets. In any given measure, cells could pass (score 0, red), generate a warning (score 1, amber), or fail (score 2, white). Cells are ordered left to right by total score. Cells with a total score > 1 (separator bar) were removed from the dataset. Colour bar denotes biological replicate.

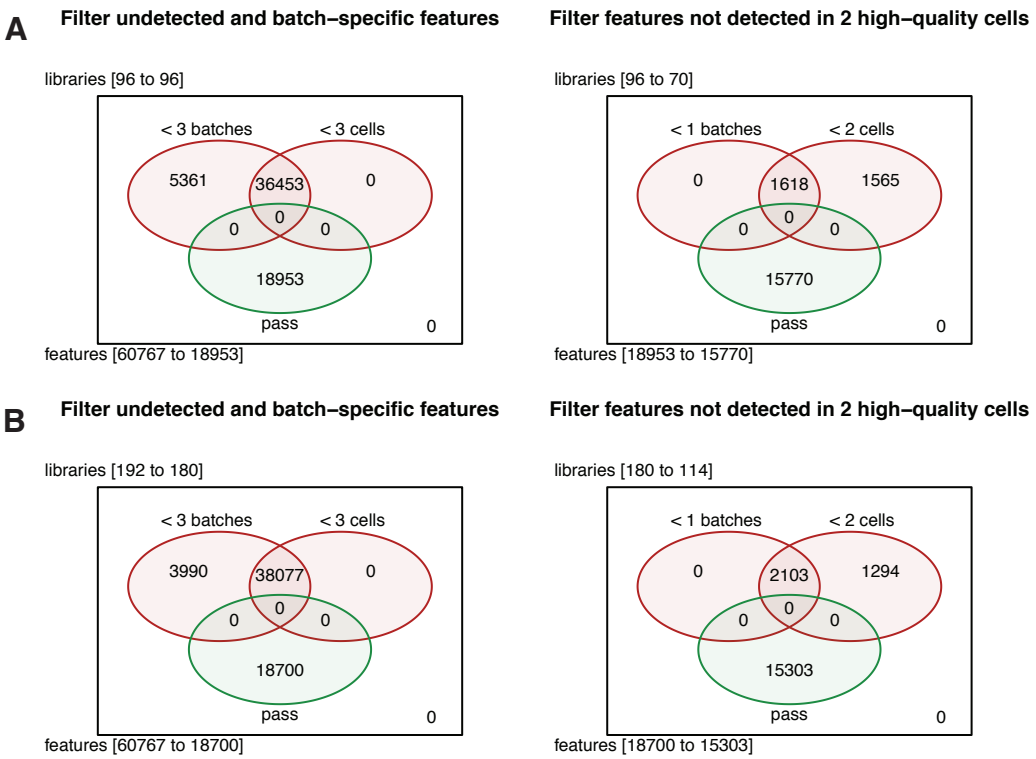


Figure A.8: Gene quality control summary. Removal of low-quality genes was performed in two steps. First, genes not detected in all 3 replicates were removed (left panel). Next, genes not detected in at least 2 cells passing QC were removed (right panel). Data for the epithelial (A) and mesenchymal (B) datasets are shown.

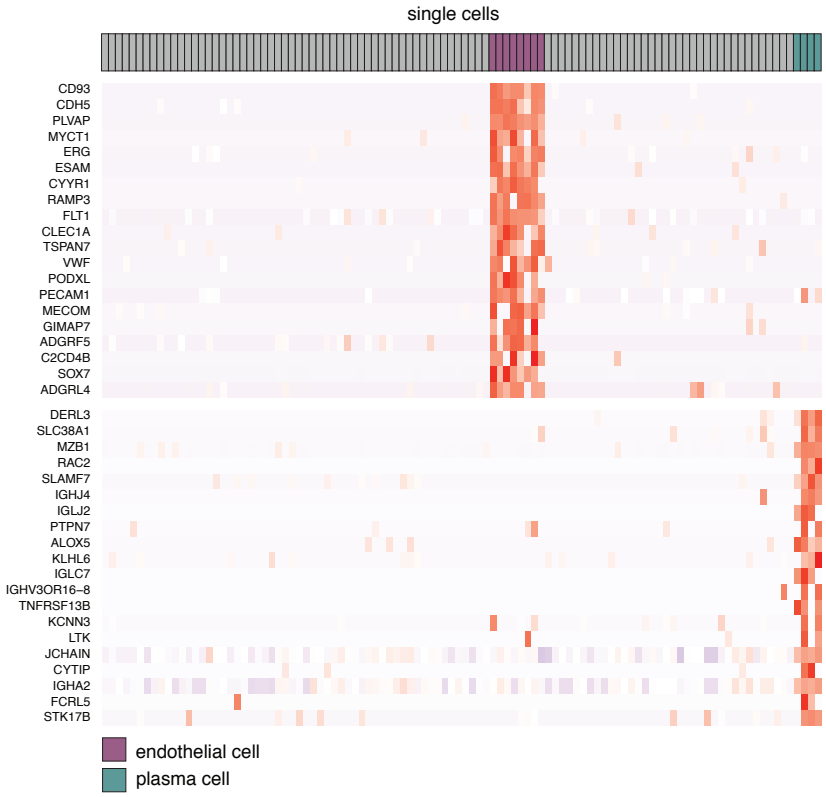


Figure A.9: Removal of contaminating non-fibroblast-like cells. Modules enriching for 'angiogenesis' (upper panel) and 'adaptive immune response' (lower panel) detected on initial WGCNA analysis. Central hub genes in these modules (rows) include *PECAM1* (CD31) and *SLAMF7*, robust markers for endothelial and plasma cells respectively. Single cells expressing these modules were removed (colour bar). A further module enriching for neural functions containing the hub gene *S100B* was also detected in some datasets and cells expressing this module, interpreted as enteric glial or neural cells, were also removed.

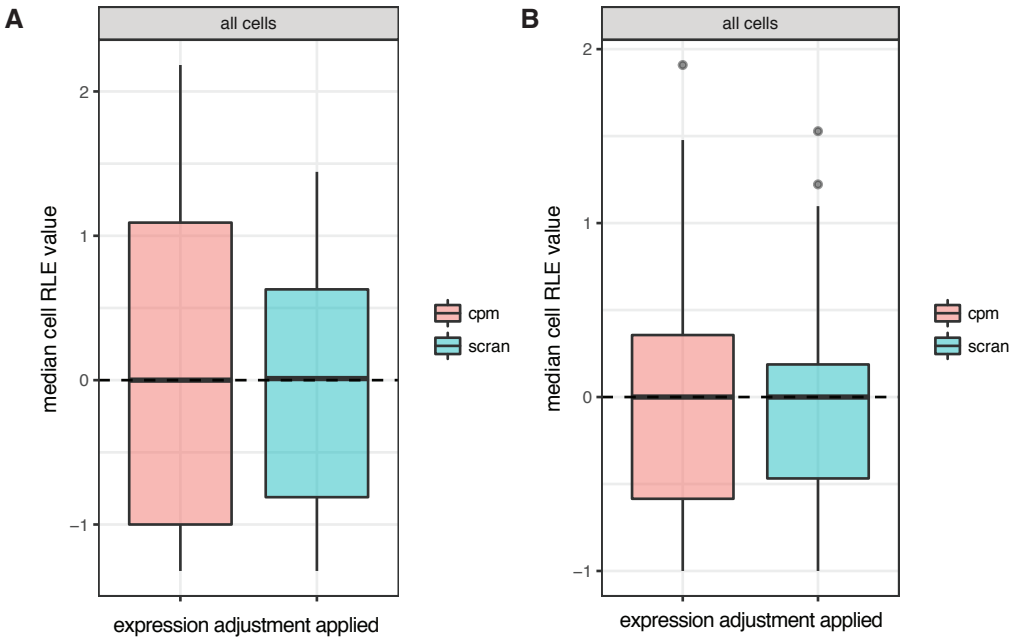


Figure A.10: Effect of library size normalisation on per-cell median relative log expression value. Distributions of per-cell median relative log expression values plotted for the epithelial (A) and mesenchymal (B) datasets using naive cpm normalisation (cpm) and using size factors derived by cell pooling (scran). Each data point represents a single cell. The dataset median (crossbar), 25^{th} and 75^{th} percentiles (upper and lower box hinges) are indicated. The box whiskers extend to the highest and lowest data points within 1.5 times the interquartile range. Outliers are plotted as points. Assuming the majority of genes are not differentially expressed, the median RLE value for appropriately normalised cells would approach zero (see Chapter 2).

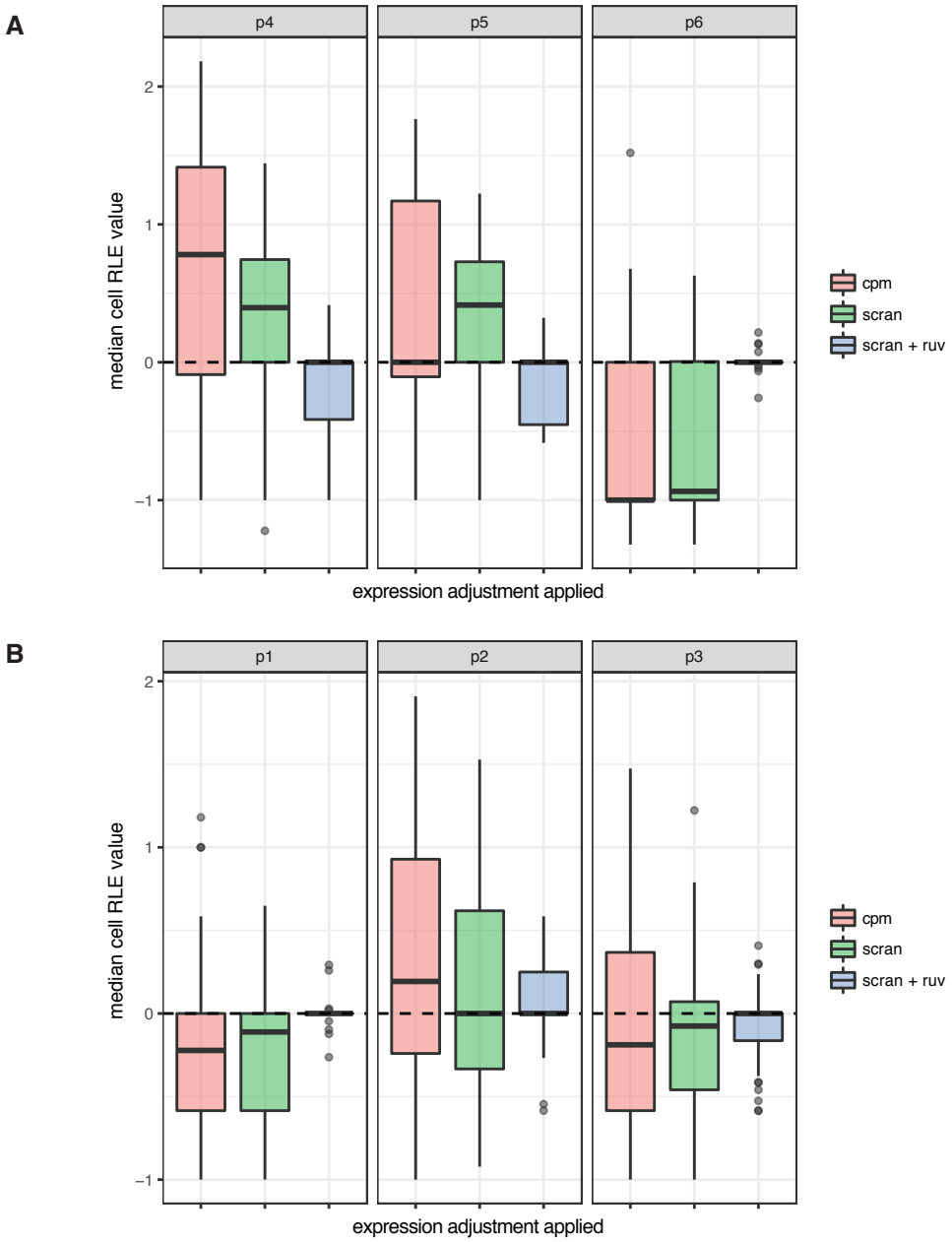


Figure A.11: Adjustment of expression values to reduce unwanted batch effects. Distribution of per-cell median relative log expression values plotted for the epithelial (A) and mesenchymal (B) datasets. Individual biological replicates (batches) are shown as column facets. Each data point represents the median relative log expression value for a single cell. The alternate adjustment approaches of naive cpm normalisation (cpm), scran library size normalisation (scran), and scran normalisation plus batch effect correction with the RUVSeq RUVg method using an *in-silico* defined set of control genes (scran + ruv) are shown (see Chapter 2). The median (crossbar), 25^{th} and 75^{th} percentiles (upper and lower box hinges) are shown. Outliers are plotted as points. Optimally normalised cells would have a median RLE value approaching zero assuming the majority of genes are not differentially expressed.

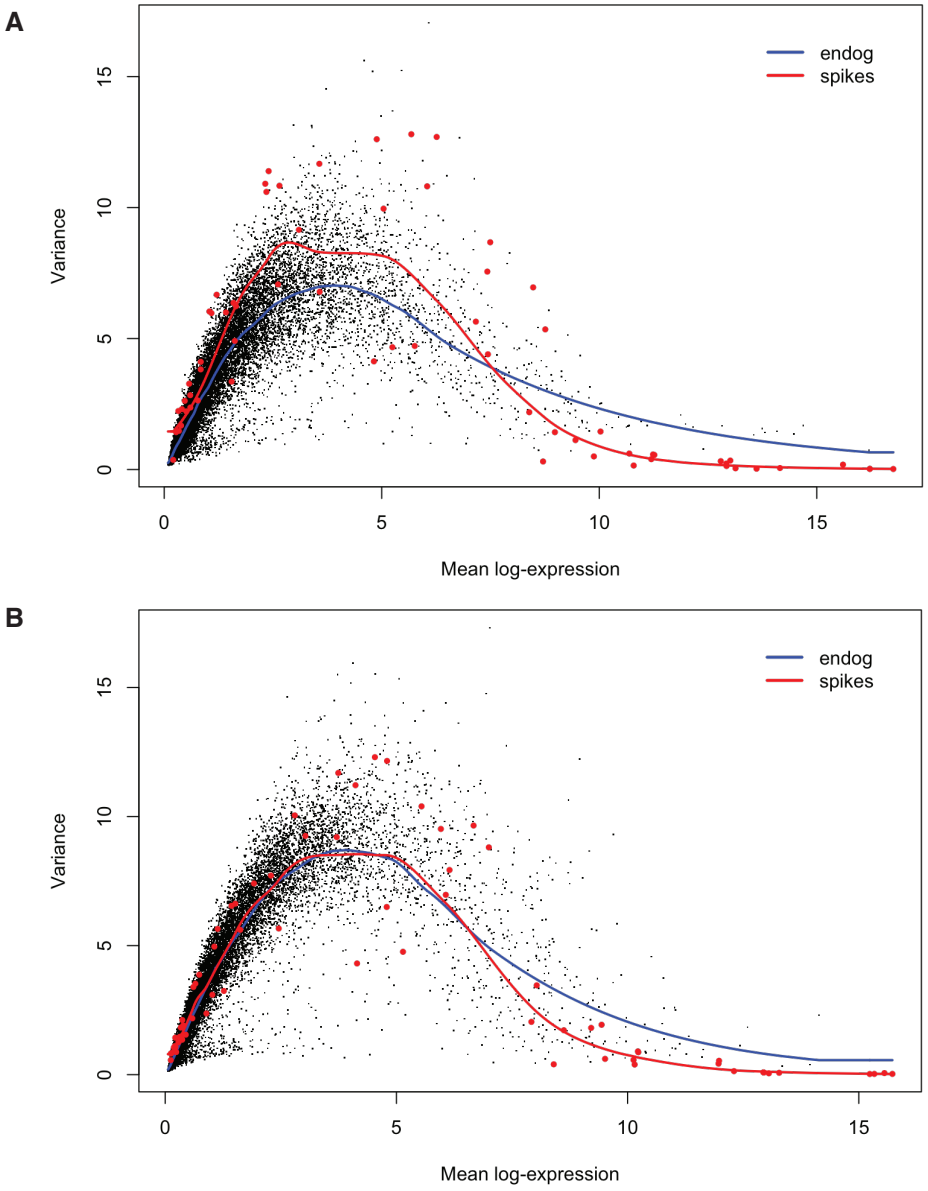


Figure A.12: Decomposition of gene variances into biological and technical components. Plots of mean log expression versus variance for the epithelial (A) and mesenchymal (B) datasets. Individual gene values and fitted trend lines are shown for endogenous genes (black points, blue line) and ERCC controls (red points and line). The endogenous trend line was taken as the upper bound of purely technical variation in gene expression, and genes showing greater total variability than this threshold were considered biologically variable.

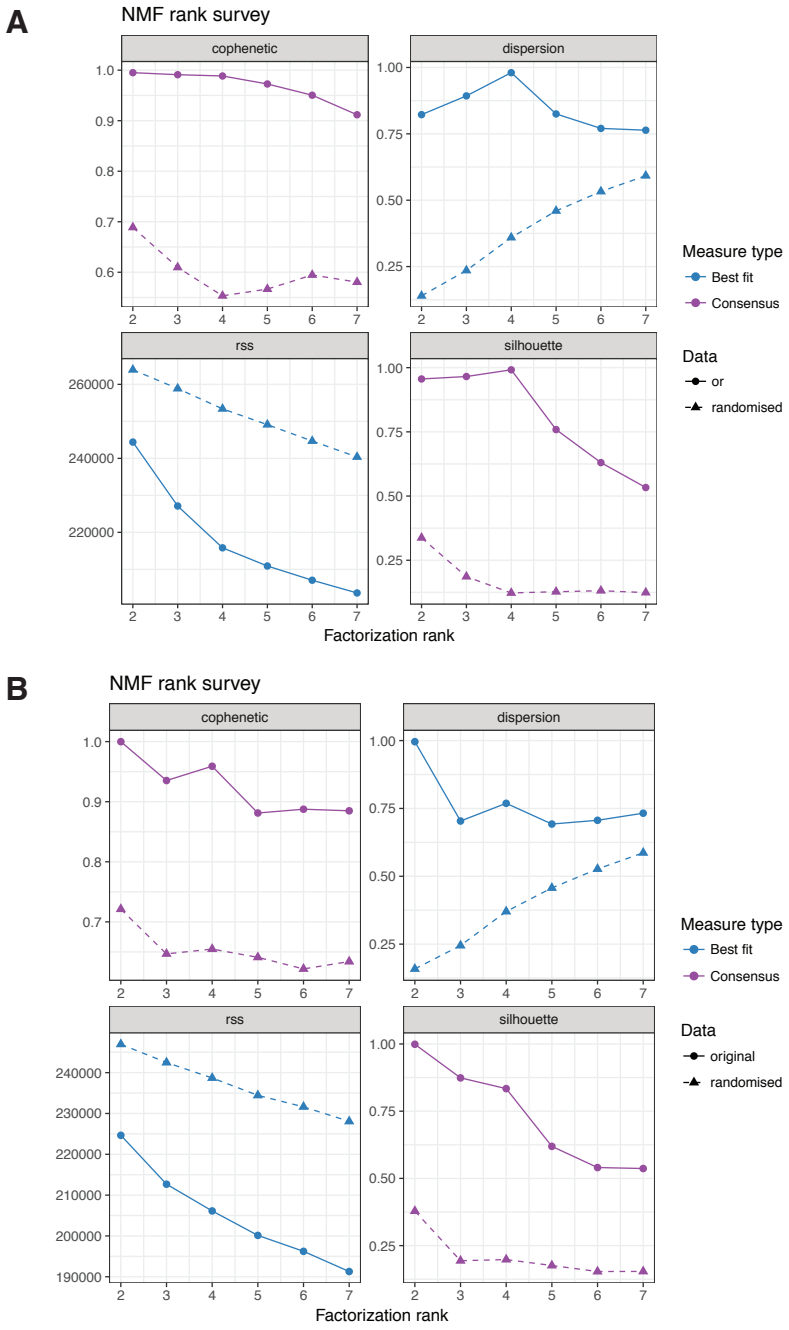


Figure A.13: NMF factorisation rank surveys. The cophentic correlation, factor dispersion, residual sum of squares (rss) and average silhouette width calculated for a range of candidate factorisation rank values (x-axis) in the epithelial (A) and mesenchymal (B) datasets. For each value, 50 randomly initiated runs were performed. The survey was repeated with randomly permuted expression data as a control (randomised). See Chapter 2 for interpretation.

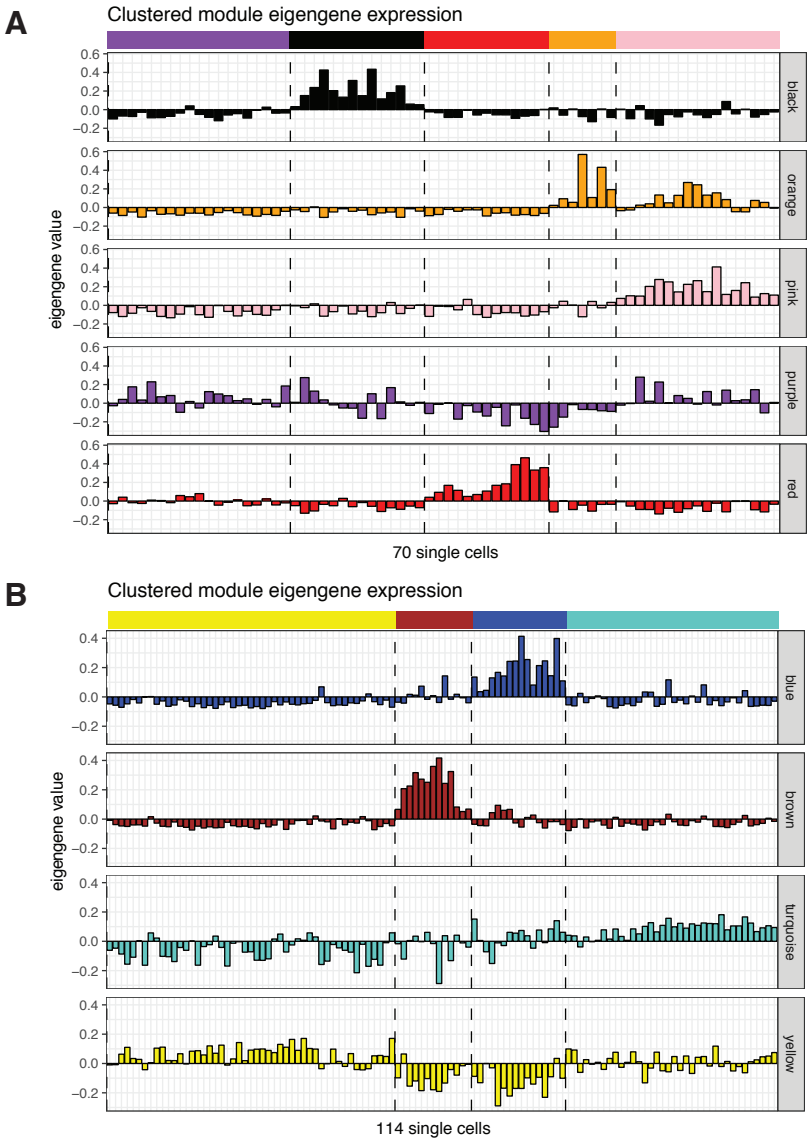


Figure A.14: Healthy stroma module eigengene expression by cell cluster. Module eigengene expression (row facets) by cell cluster (dashed lines) for (A) epithelial cells and (B) mesenchymal cells. Colour identifiers for cell clusters, selected based on the highest expressed module, are shown (colour bars).

A.2 Supplementary tables

Table A.1: Sample acquisition and processing — healthy left colon epithelium

Participants			
Age	45	29	56
Gender	F	M	F
Indication	change in bowel habit	change in bowel habit	change in bowel habit
Medications	none	venlafaxine, fluoxetine	none
Endoscopy findings	normal	normal	normal
Histology findings	normal	normal	normal
Biopsy sites	Sigmoid colon x 8	Sigmoid colon x 8	Sigmoid colon x 8
Sample processing			
Viability %	88	71	71
Live count (cells x 10^3)	950	180	370
C1 chip capture			
Single viable cells	54	34	56
Non-viable cells (high PI)	13	12	7
Other failure	13	12	28
Empty	16	38	5
Total	96	96	96
Sequencing			
Single cells sequenced	30	31	31
Bulks sequenced	1	1	1
Empty wells sequenced	0	0	1

Table A.2: Sample acquisition and processing — healthy left colon mesenchyme

Participants			
Age	69	57	50
Gender	F	F	F
Indication	follow up (diverticulitis)	change in bowel habit	change in bowel habit
Medications	lisinopril	calcium	none
Endoscopy findings	normal	normal	normal
Histology findings	normal	normal	normal
Biopsy sites	Sigmoid colon x 8	Sigmoid colon x 8	Sigmoid colon x 8
Sample processing			
Viability %	88	66	79
Live count (cells x 10^3)	350	170	300
C1 chip capture			
Single viable cells	57	58	68
Non-viable cells (high PI)	22	17	14
Other failure	16	16	13
Empty	1	5	1
Total	96	96	96
Sequencing			
Single cells sequenced	57	58	68
Bulks sequenced	2	2	2
Empty wells sequenced	0	3	0

Appendix B

Supplementary Material for Chapter 4

Contents

B.1 Supplementary figures 196
B.2 Supplementary tables 201

B.1 Supplementary figures

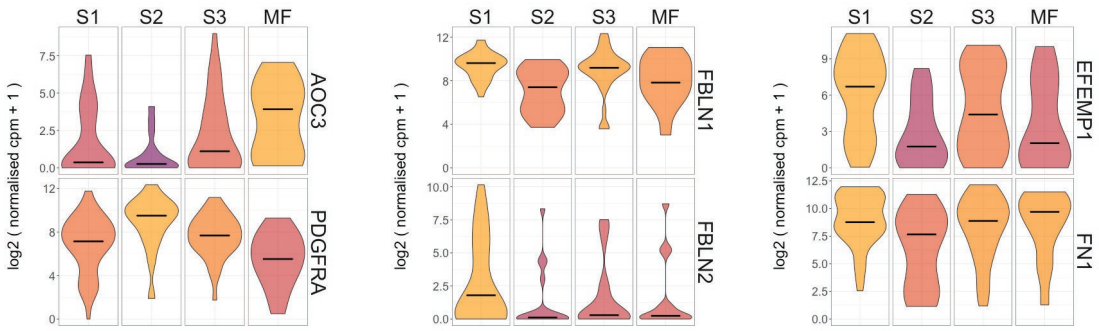


Figure B.1: Additional subset marker violin plots (healthy stroma). Violin plots for selected subset marker genes mentioned in the text.

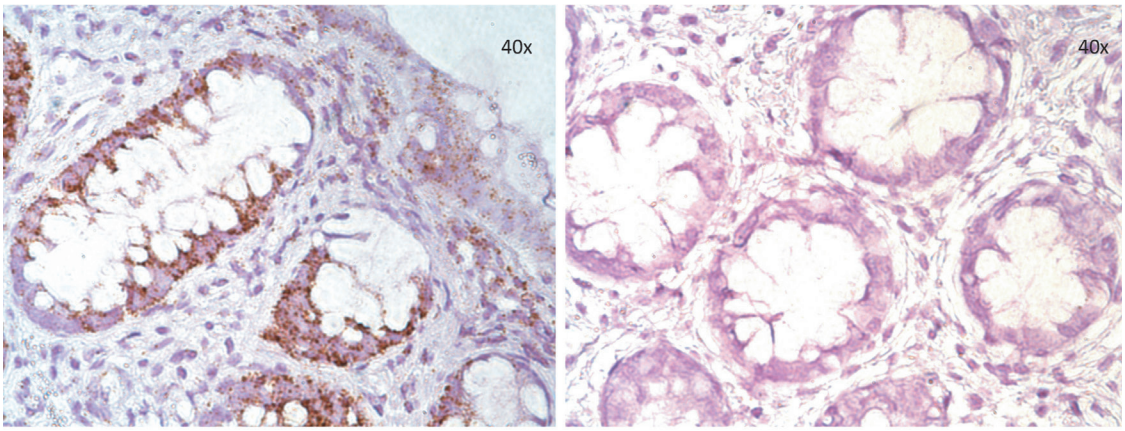


Figure B.2: Single molecule ISH technical controls. Healthy human colonic tissue stained by single molecule ISH with probes for the human housekeeping gene POL2RA (left panel) and the bacterial DapB gene (right panel). Nuclear counterstaining was performed with haematoxylin (blue). Positive signal appears brown.

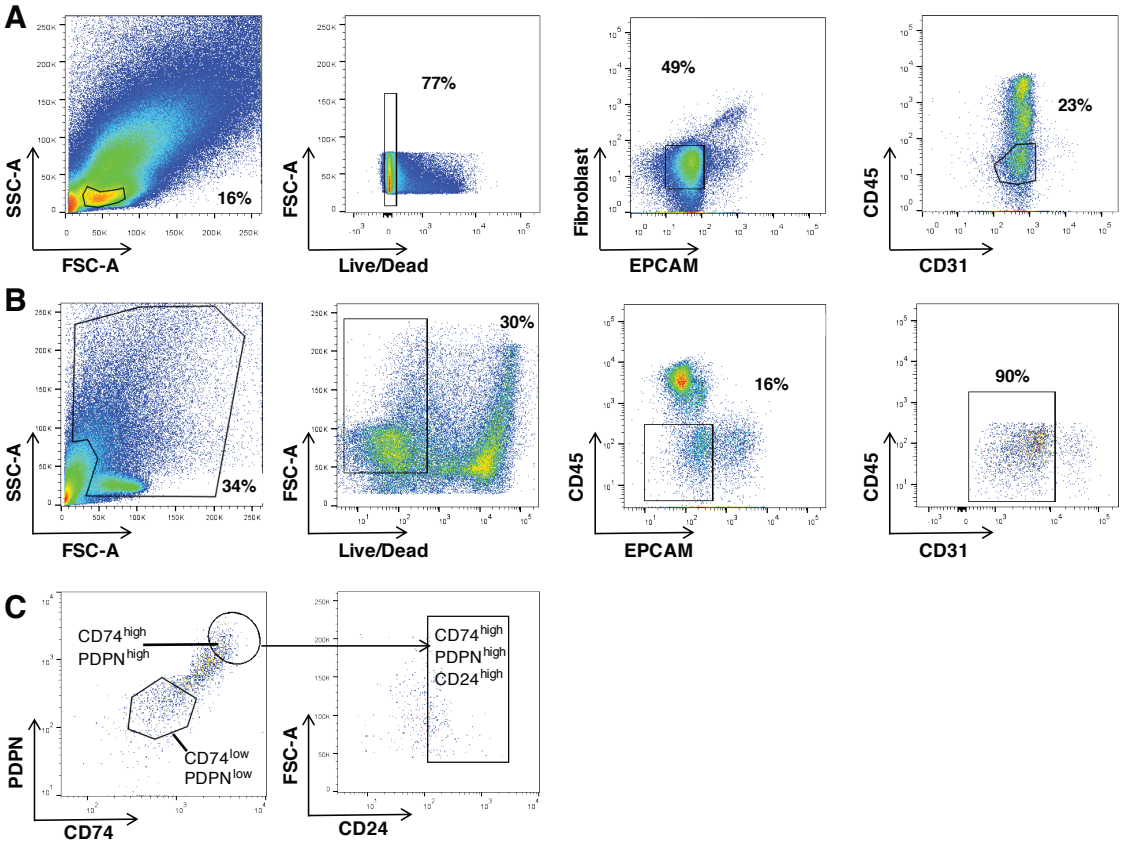


Figure B.3: Representative FACS gating strategies. (A) Representative gating strategies for the detection of nuclear targets. (B) Representative gating strategies for the detection of cytoplasmic targets. (C) Gating strategy for identifying the $CD74^{high}PDPN^{high}CD24^{high}$ Stromal 4 population in inflamed UC biopsies.

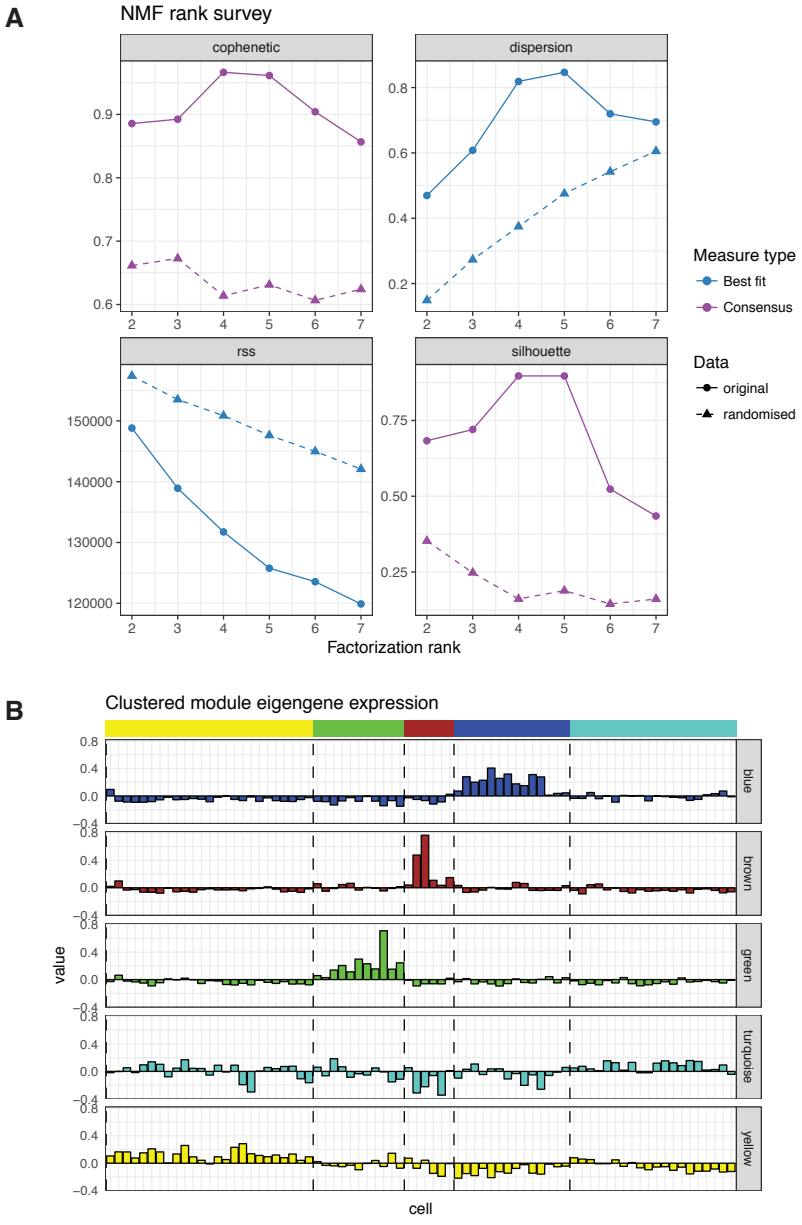


Figure B.4: UC dataset NMF factorisation rank survey and eigengene expression. (A) NMF factorisation rank survey for the UC dataset (see Chapter 2). (B) Module eigengene expression (row facets) by cluster annotation for 76 single mesenchymal cells from UC samples.

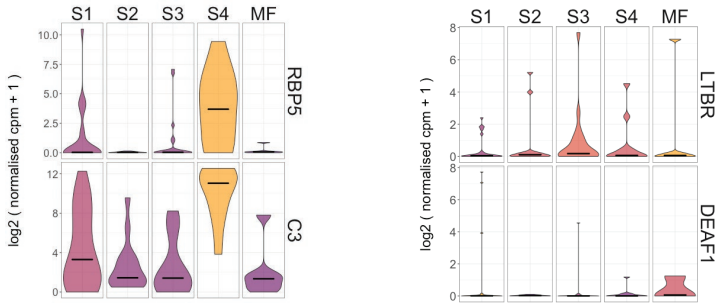


Figure B.5: Expression profiles of selected FDC-related genes in the UC dataset. Violin plots for the indicated genes associated with FDC function mentioned in the text. Crossbar indicates median expression.

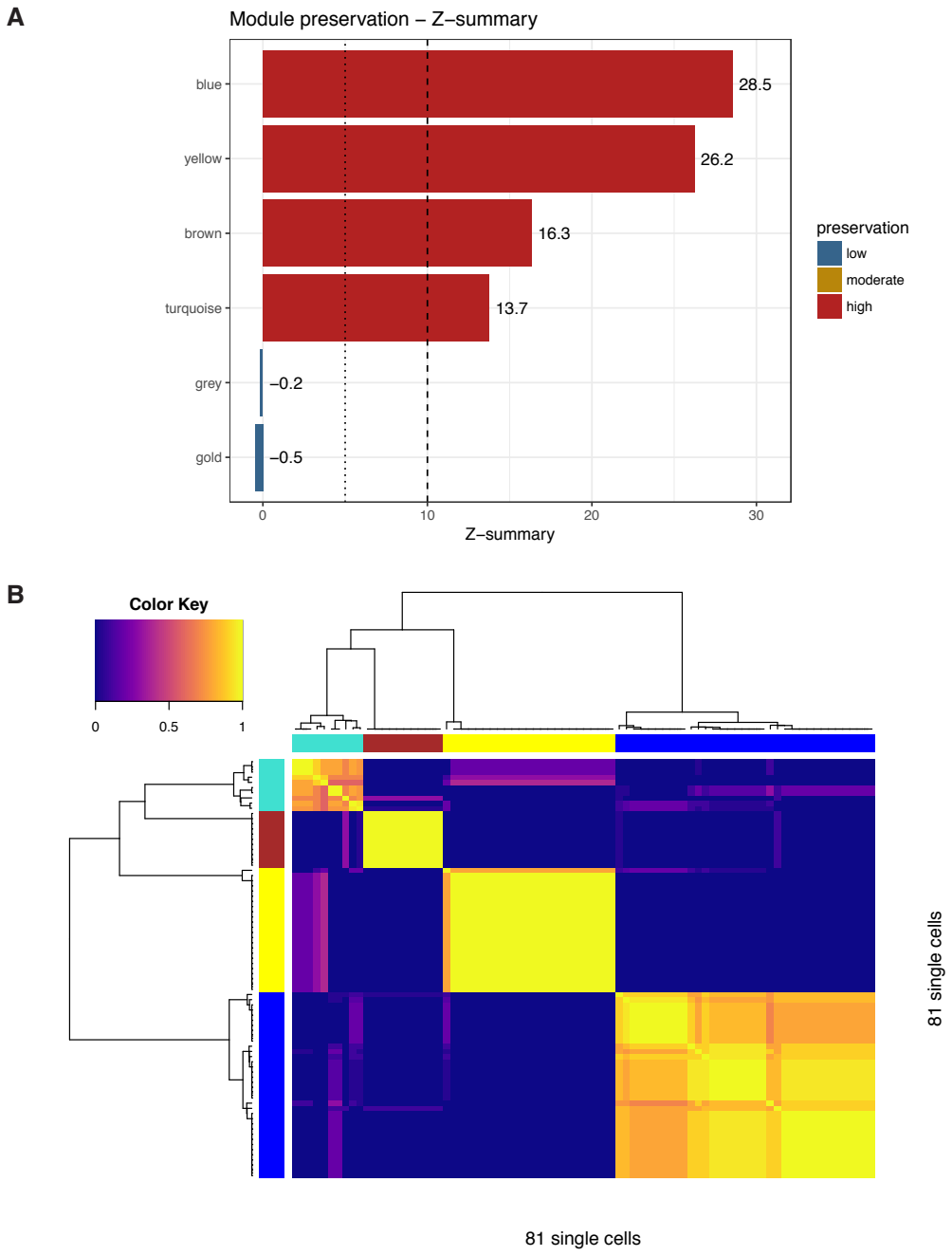


Figure B.6: Module preservation and NMF clustering in dysplastic UC. (A) 'Z-summary' module preservation score for the healthy colon modules within the gene co-expression network generated from the dysplastic UC dataset. The grey and gold control modules represent random samples genes (taken from all unassigned genes and all genes respectively). A score of > 10 indicates strong preservation (see Chapter 2). (B) NMF clustering of 81 single cells from donors with dysplastic UC. Heat colour represents co-clustering probability in 50 randomly initiated runs. Cluster assignments are indicated (colour bars). A factorisation rank of 4 was used.

B.2 Supplementary tables

Table B.1: Sample acquisition and processing — UC left colon mesenchyme

Participants			
Age	29	45	30
Gender	M	M	M
Indication	Diarrhoea, rectal bleeding	Known UC, symptomatic	Diarrhoea, rectal bleeding
Medications	None	Mesalazine	None
Endoscopy findings	Distal colitis	Distal colitis	Distal colitis
UCEIS[1] score	5/8	6/8	2/8
Histology, inflamed	Active chronic UC	Active chronic UC	Active chronic UC
Histology, non-inflamed	Normal	No active inflammation	No active inflammation
C1 chip capture			
Single cells, inflamed	13	31	9
Single cells, non-inflamed	24	19	31
Sequencing			
Single cells sequenced	37	50	40
Bulks sequenced	1	1	1
Empty wells sequenced	1	1	1

[1] Ulcerative Colitis Endoscopic Index of Severity[381]

Table B.2: Sample acquisition and processing — dysplastic UC left colon mesenchyme

	Participants	
Age	71	74
Gender	M	M
Indication	Surveillance	Surveillance
Co-morbidity	NIDDM[1], COPD[2]	NIDDM[1], hypertension
Endoscopy findings	Quiescent colitis	Quiescent colitis
Histology, dysplastic site	TVA[3] with low grade dysplasia	Low grade dysplasia
Histology, non-dysplastic site	Quiescent colitis	Quiescent colitis
	C1 chip capture	
Single cells, dysplastic site	23	23
Single cells, non-dysplastic site	23	23
	Sequencing	
Single cells sequenced	46	46
Bulks sequenced	1	1
Empty wells sequenced	1	1

[1] Non-insulin dependent diabetes mellitus
[2] Chronic obstructive pulmonary disease
[3] Tubulovillous adenoma

Table B.3: Sample acquisition and processing — TNF stimulation in healthy donors

Participants			
Age	59	45	31
Gender	F	M	M
Indication	Change in bowel habit	Rectal bleeding	Rectal bleeding
Medications	None	None	None
Endoscopy findings	Normal	Normal	Normal
Histology	Not indicated	Not indicated	Not indicated
C1 chip capture			
Single cells, TNF	12	16	18
Single cells, control	10	16	18
Sequencing			
Single cells sequenced	22	32	36
Bulks sequenced	1	1	1
Empty wells sequenced	1	1	1

Appendix C

Supplementary Material for Chapter 5

Contents

C.1 Supplementary figures 204

C.1 Supplementary figures

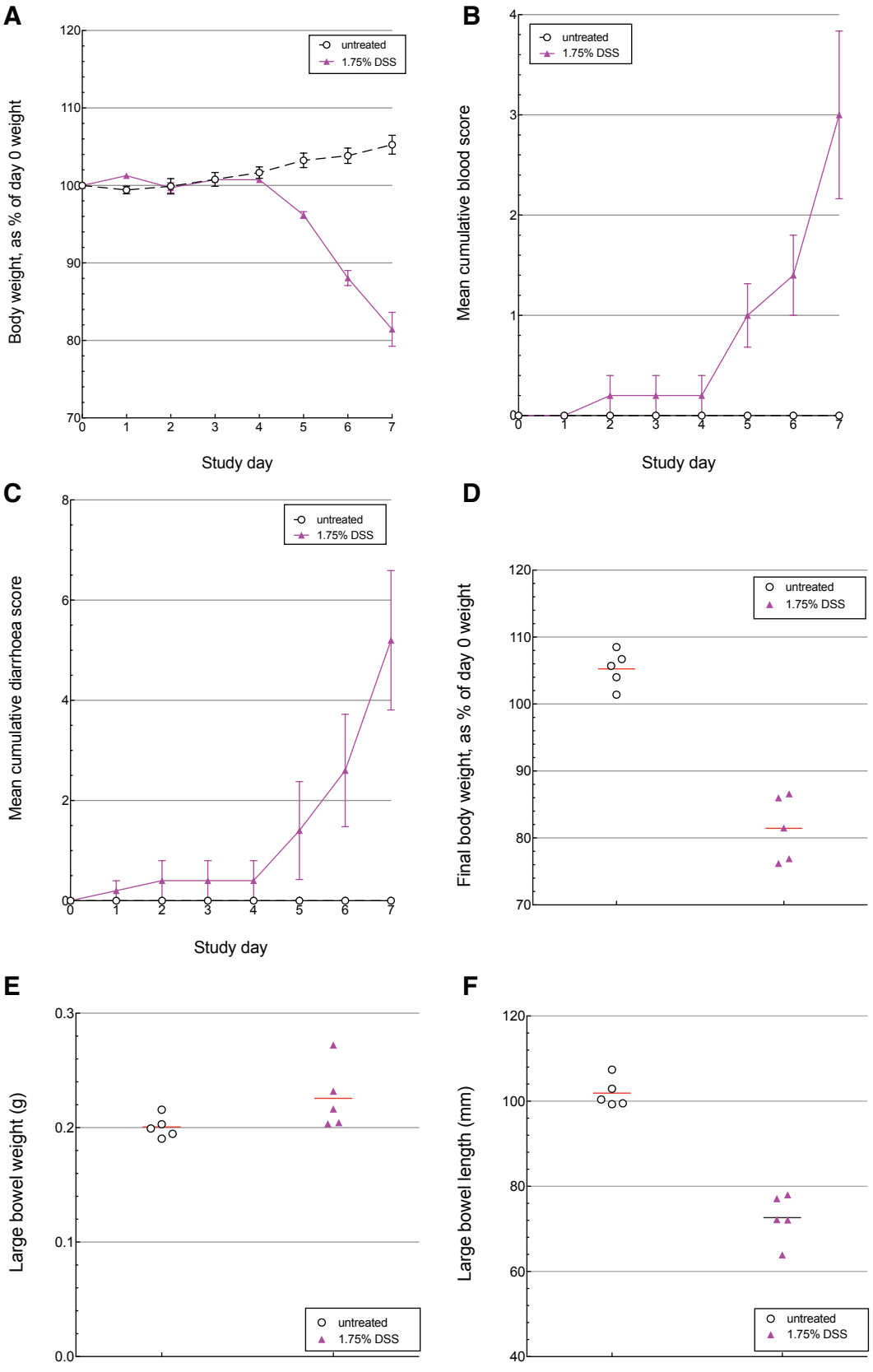


Figure C.1: Further in-life and post-mortem outcomes of DSS challenge.

Figure C.1 ***(previous page)*****:** (A—F) Scores for measures indicated across treatment groups. (A—C) Mean and SEM. (D—F) Individual data points and group mean (crossbar).

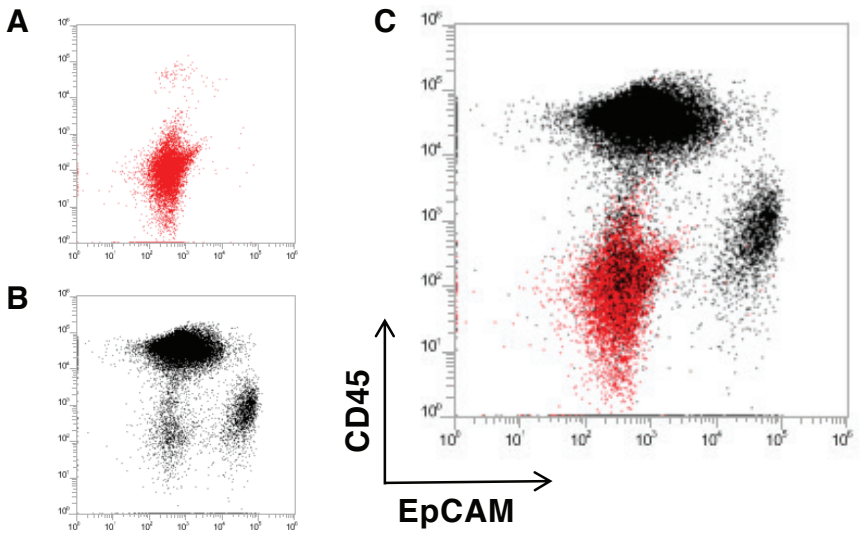


Figure C.2: FACS assessment of murine colonic stromal cell isolation. Expression levels of pan-leukocyte marker CD45 versus epithelial marker EpCAM for a mesenchyme-enriched single cell suspension derived from murine healthy colon and fractionated using MACS. The column flow-through (A) and retained (B) fractions are shown, together with an overlay of both populations (C).

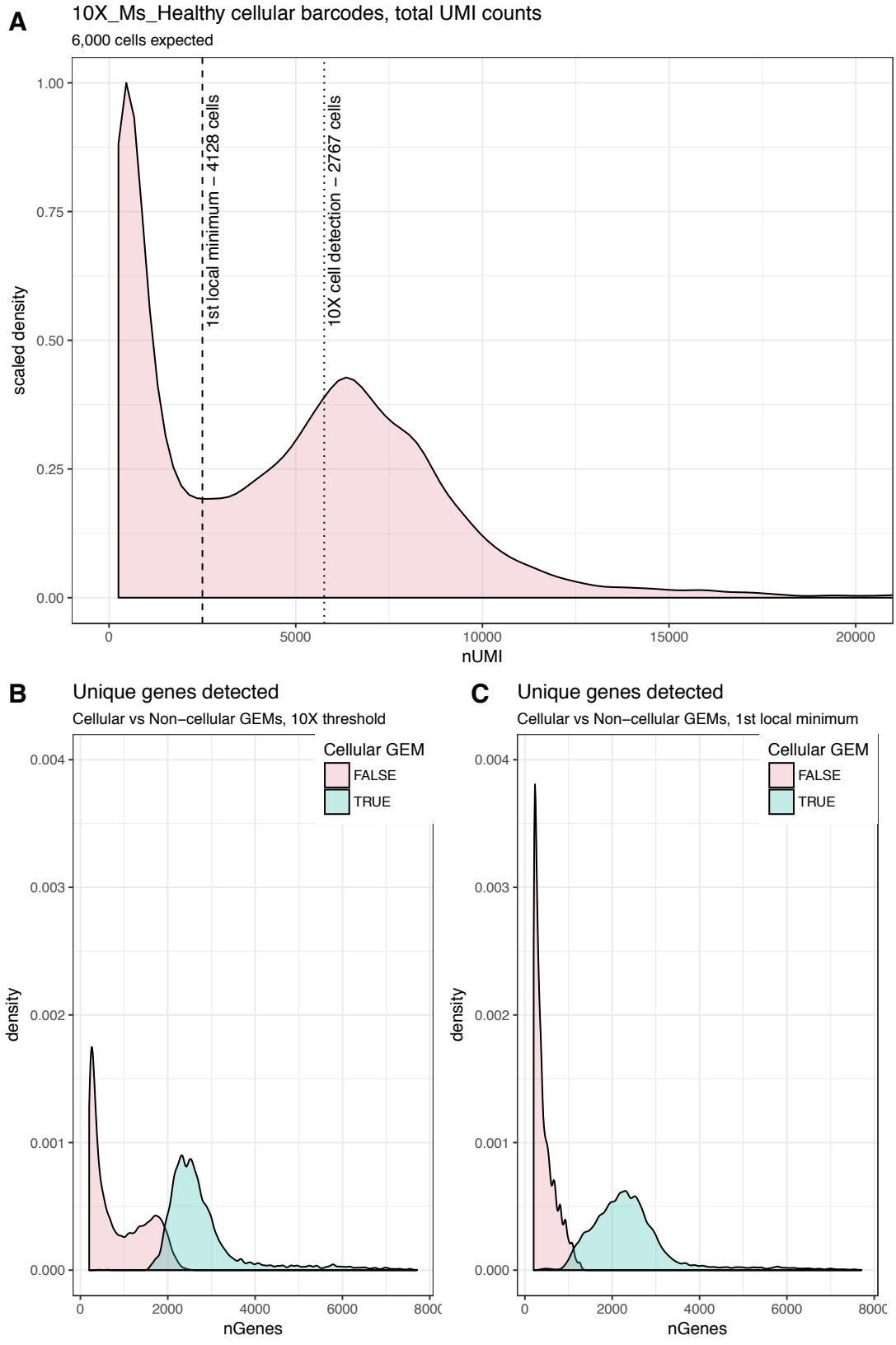


Figure C.3: Identification of cellular barcodes — healthy controls.

Figure C.3 ***(previous page)*:** (A) Scaled cellular barcode density versus number of barcode UMIs. Barcodes with fewer than 250 UMIs have been removed to focus on the relevant area of the plot. The automated 10X detection threshold ($0.1 \times 99^{\text{th}}$ percentile of top n barcodes by UMI count, where n = number of cells expected) and first local minimum are shown, together with the total cells captured using these thresholds. (B,C) Barcode density versus number of unique genes detected for cellular and non-cellular barcodes using the 10X and first local minimum thresholds respectively.

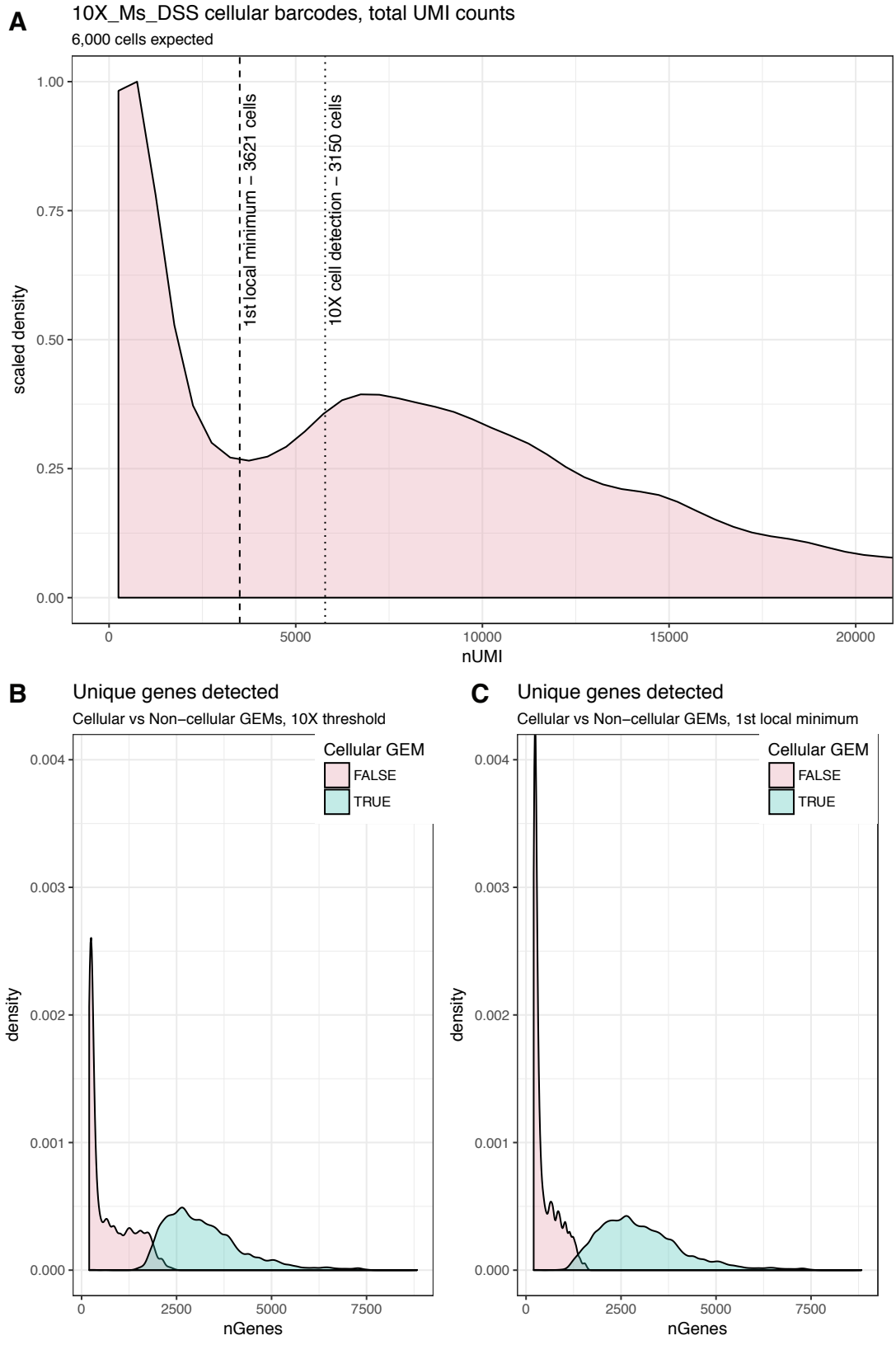


Figure C.4: Identification of cellular barcodes — DSS colitis.

Figure C.4 ***(previous page)*:** (A) Scaled cellular barcode density versus number of barcode UMIs. Barcodes with fewer than 250 UMIs have been removed to focus on the relevant area of the plot. The automated 10X detection threshold ($0.1 \times 99^{\text{th}}$ percentile of top n barcodes by UMI count, where n = number of cells expected) and first local minimum are shown, together with the total cells captured using these thresholds. (B,C) Barcode density versus number of unique genes detected for cellular and non-cellular barcodes using the 10X and first local minimum thresholds respectively.

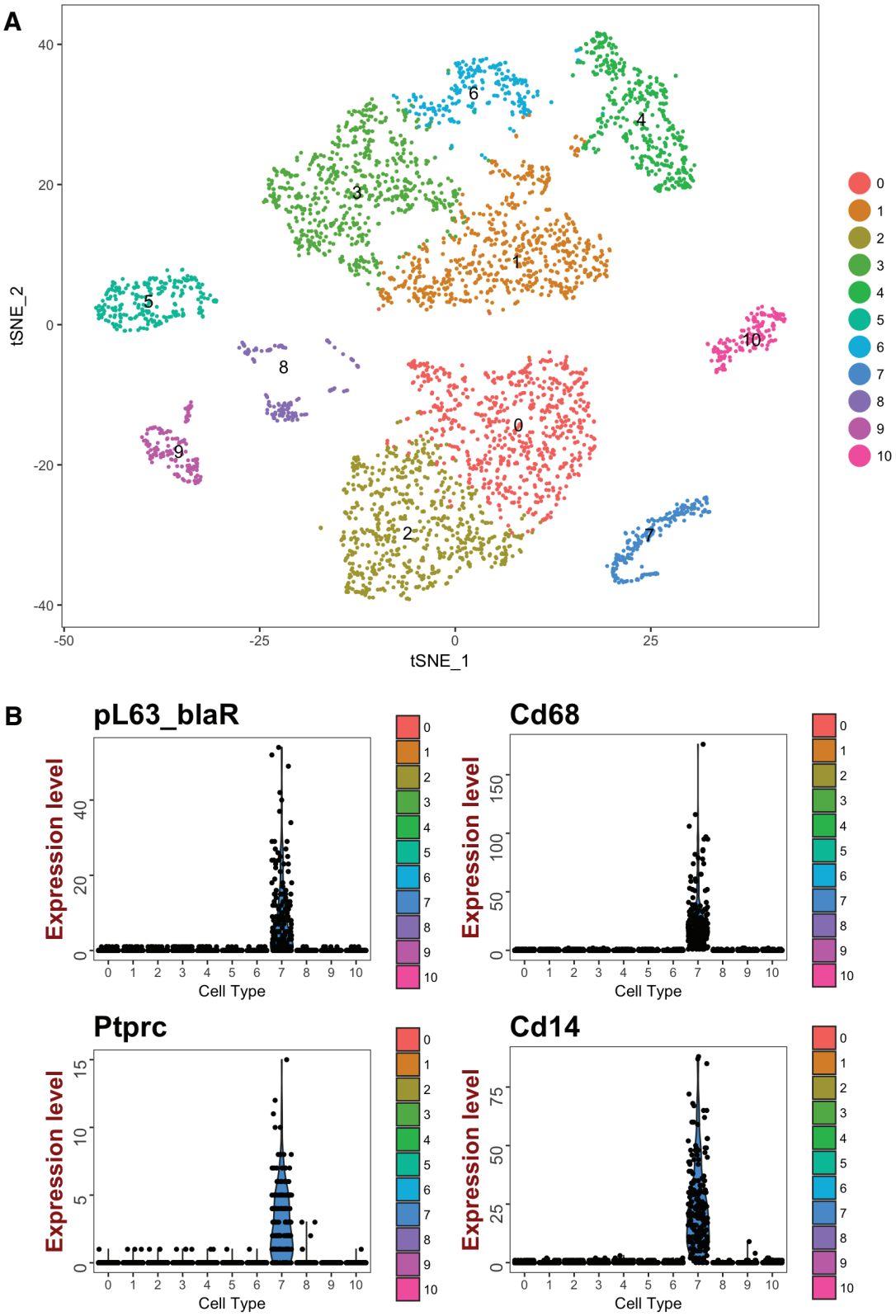


Figure C.5: Identification of spike-in control RAW 264.7 cells.

Figure C.5 ***(previous page)*:** (A) TSNE projection of single cells from the three healthy control channels. Cells are coloured by cluster annotation using following graph-based clustering on the first 10 principle components. Clusters are assigned numeric identifiers as indicated. (B) Expression of markers expected to show high specificity for control cells including including *Aspergillus fumigatus* blasticidin resistance transgene (pL63_blaR), macrophage marker Cd68, pan-leukocyte marker Ptprc (Cd45) and myeloid marker Cd14.

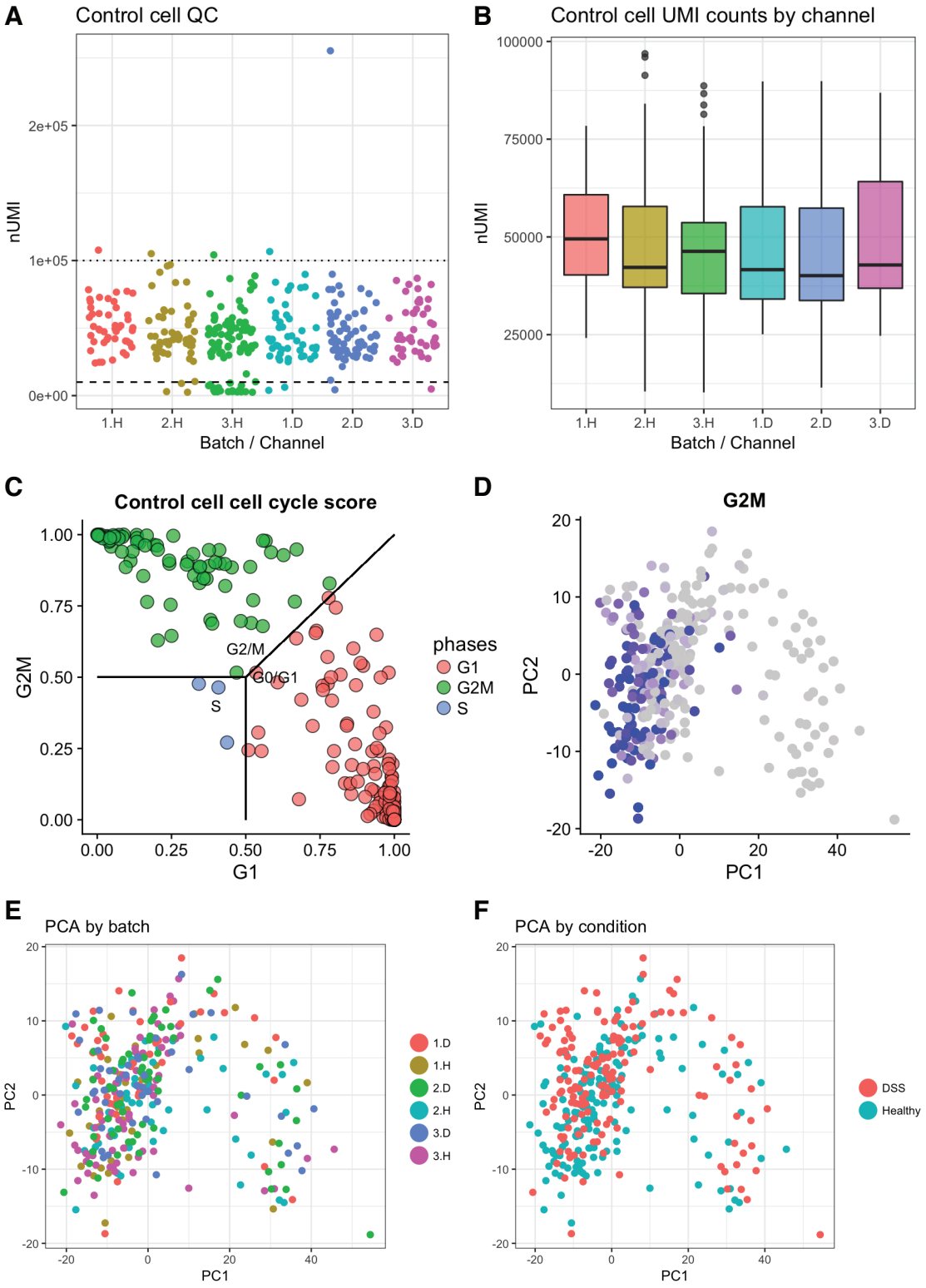


Figure C.6: Assessment of channel-associated batch effect using control cells.

Figure C.6 ***(previous page)*:** (A) Number of UMIs detected (nUMI) versus channel for control cells. The upper and lower QC thresholds spanning one order of magnitude are indicated (dotted and dashed lines). (B) Box plot of number of UMIs versus channel for control cells passing QC. Crossbar indicates group median. Upper and lower box 'hinges' are the 25^{th} and 75^{th} percentiles. Whiskers extend to $1.5 \times$ interquartile range. Outliers plotted as points. (C) Cell cycle annotation for control cells derived using the cyclone 'pairs' method. Thresholds for phase annotation are indicated (plot lines). (D—F) Principle components 1–2 for control cells considering all variable genes. Cells are coloured by G2M cell cycle score (D), batch (E), and experimental condition (F).

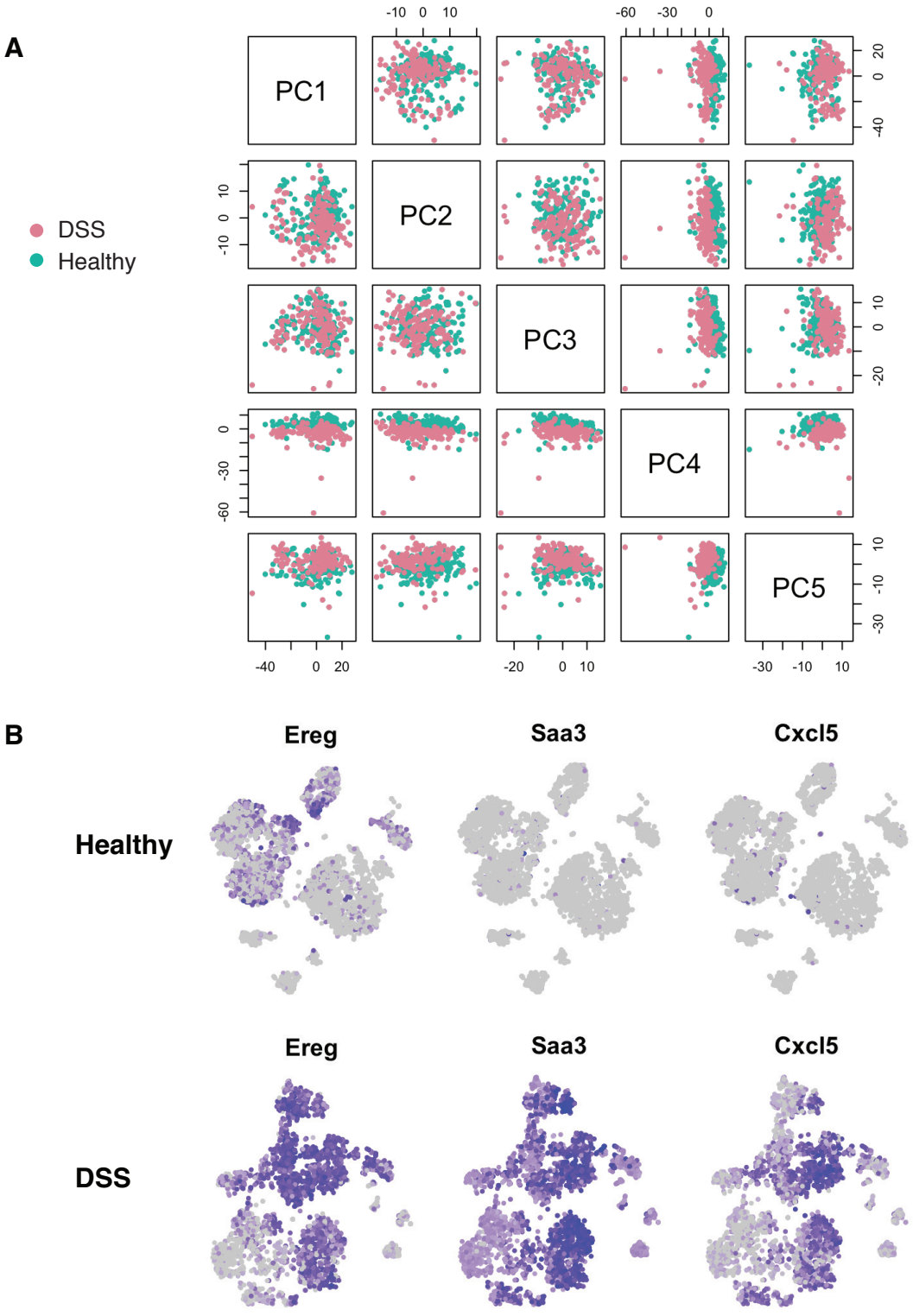


Figure C.7: Cell-free RNA produces a small batch effect between conditions. (A) Control cell PCA following adjustment for number of UMIs detected and cell cycle scores. Control cells are coloured according to whether the originated in a DSS or control channel. (B) TSNE projections of the healthy and DSS datasets coloured by expression of the indicated genes, which were the top-ranked genes in terms of PC4 loading from (A).

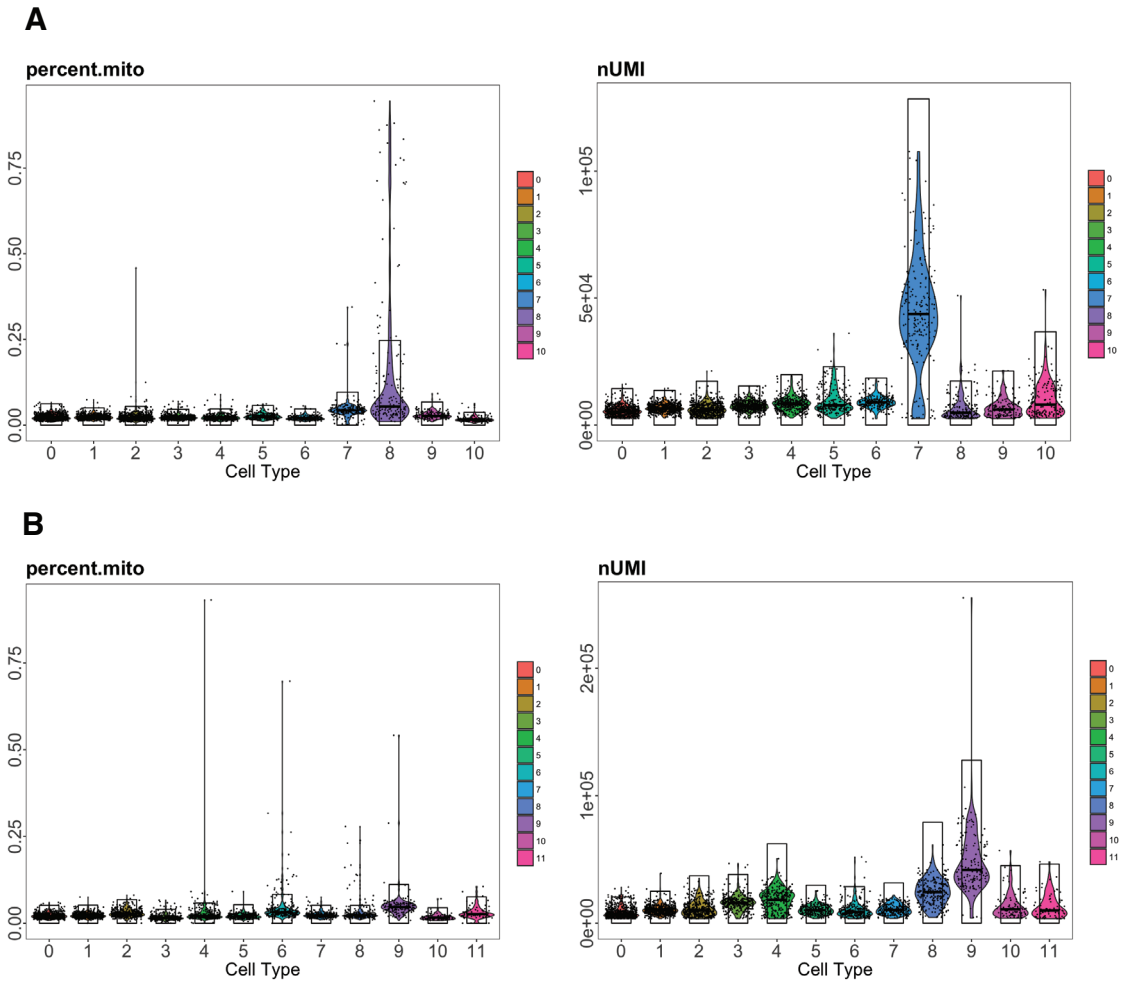


Figure C.8: Cellular QC. Violin plots showing percentage of UMI counts originating from mitochondrial features and the total UMI counts per cell for the healthy (A) and DSS (B) datasets. Cells are grouped by cluster annotation derived from and initial graph-based clustering using the first 10 principle components (x-axis). The group median (crossbar) and upper and lower QC thresholds (box hinges, median $\pm 5 \times$ median absolute deviation) are shown.

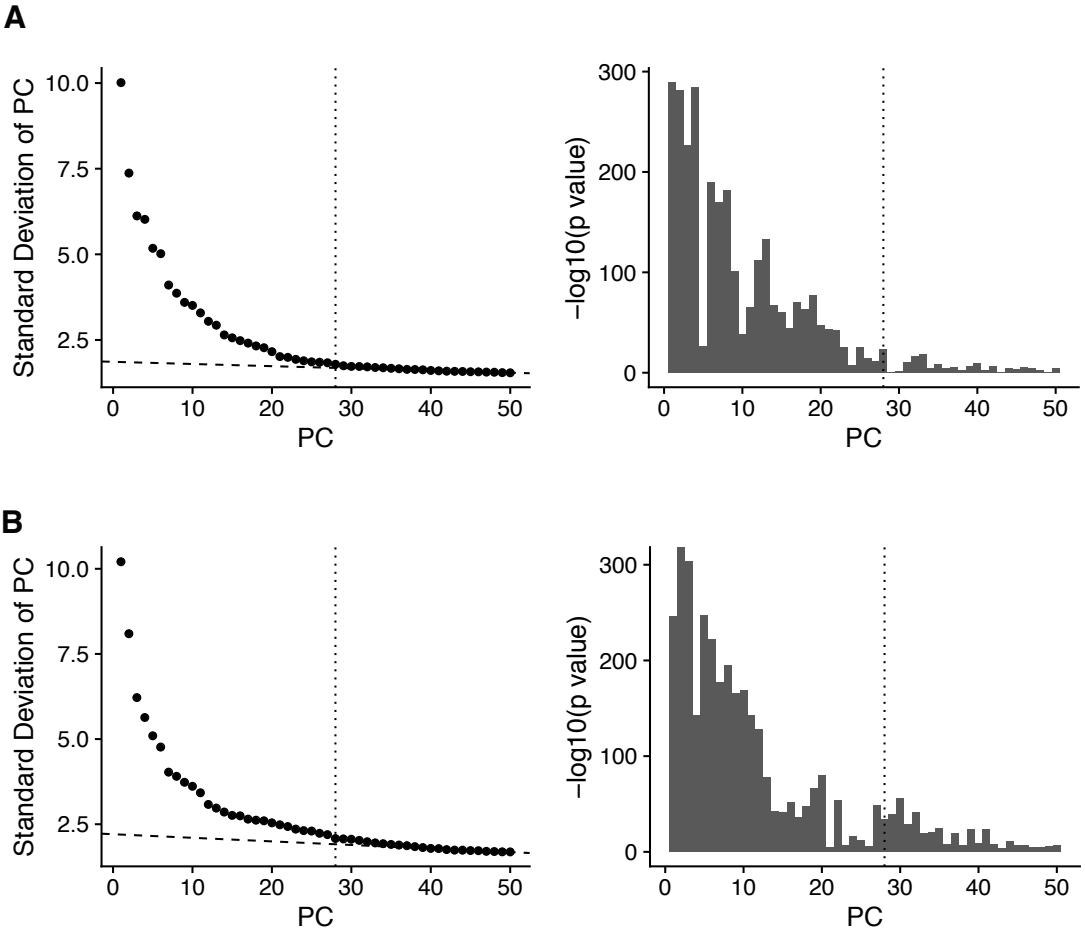


Figure C.9: Determination of important principle components for cell clustering. Principle component 'scree' plots (left panel) and significance (Jackstraws test) for the healthy (A) and DSS (B) datasets. A best fit line for principle components 40 to 50 is shown (dashed), as is a line indicating the threshold selected for significant PCs (dotted).

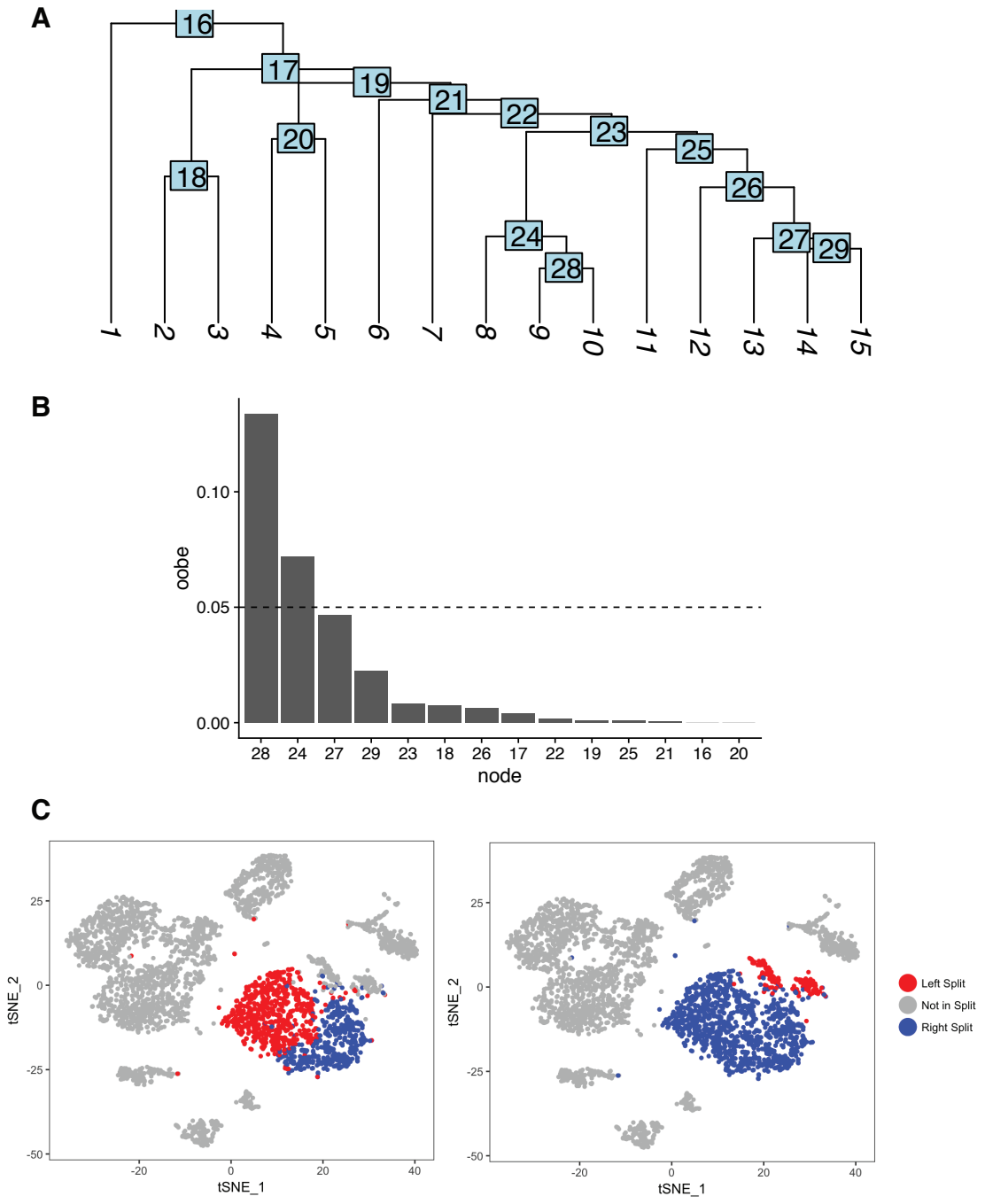


Figure C.10: Merging of low-confidence clusters, healthy dataset. (A) Phylogenetic tree representing inter-cell distances for the average cell in each detected cluster in the healthy control dataset. Branch splitting points (nodes) are assigned numeric identifiers. (B) Out of bag error (oobe) for a random forest classifier trained on each internal node (x-axis), reflecting confidence in assigning cells to the left or right branch. A threshold confidence of 0.05 is indicated (dashed line). (C) Sub-clusters defined by low confidence nodes 28 (left panel) and 24 (right panel). These are merged in the final clustering structure.

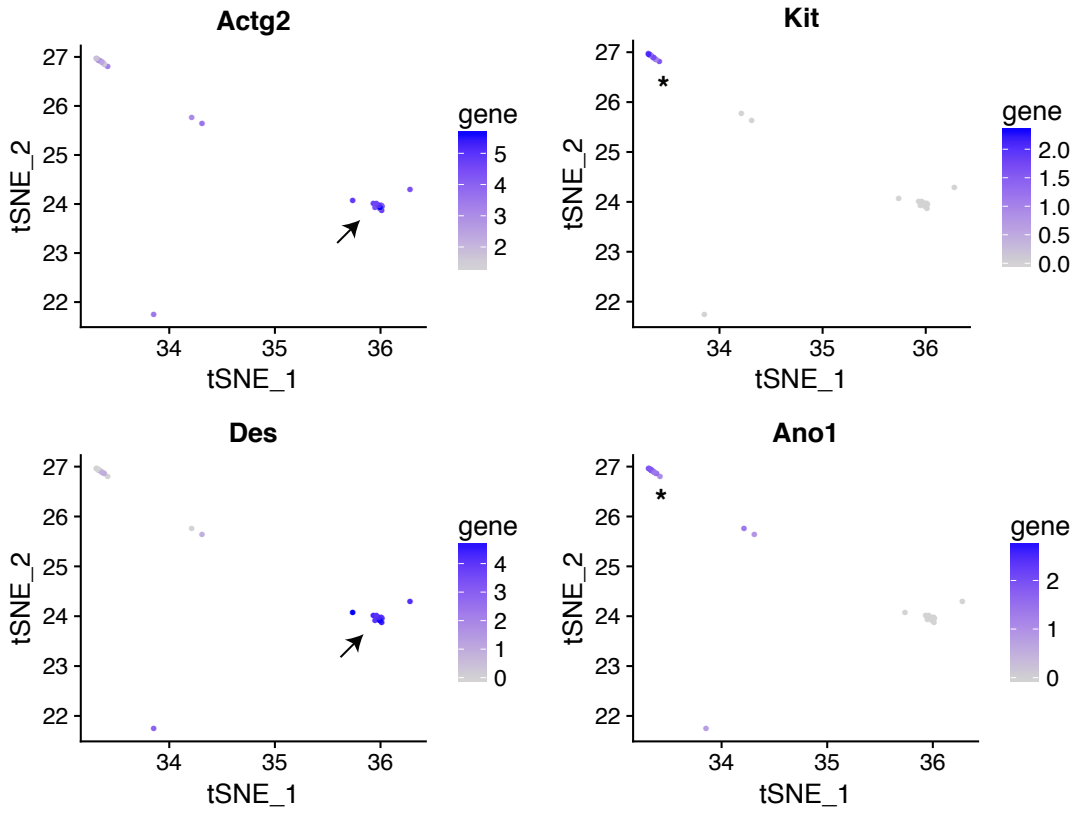


Figure C.11: Heterogeneity within the smooth muscle / ICC cell cluster. (A) TSNE projections of the Sm/ICC cluster coloured by expression of smooth muscle markers (Actg2, Des) and ICC markers (Kit and Ano1). Ano1^{high}, Kit^{high} and Actg2^{high}, Des^{high} cell groups are indicated (asterisks and arrows respectively).

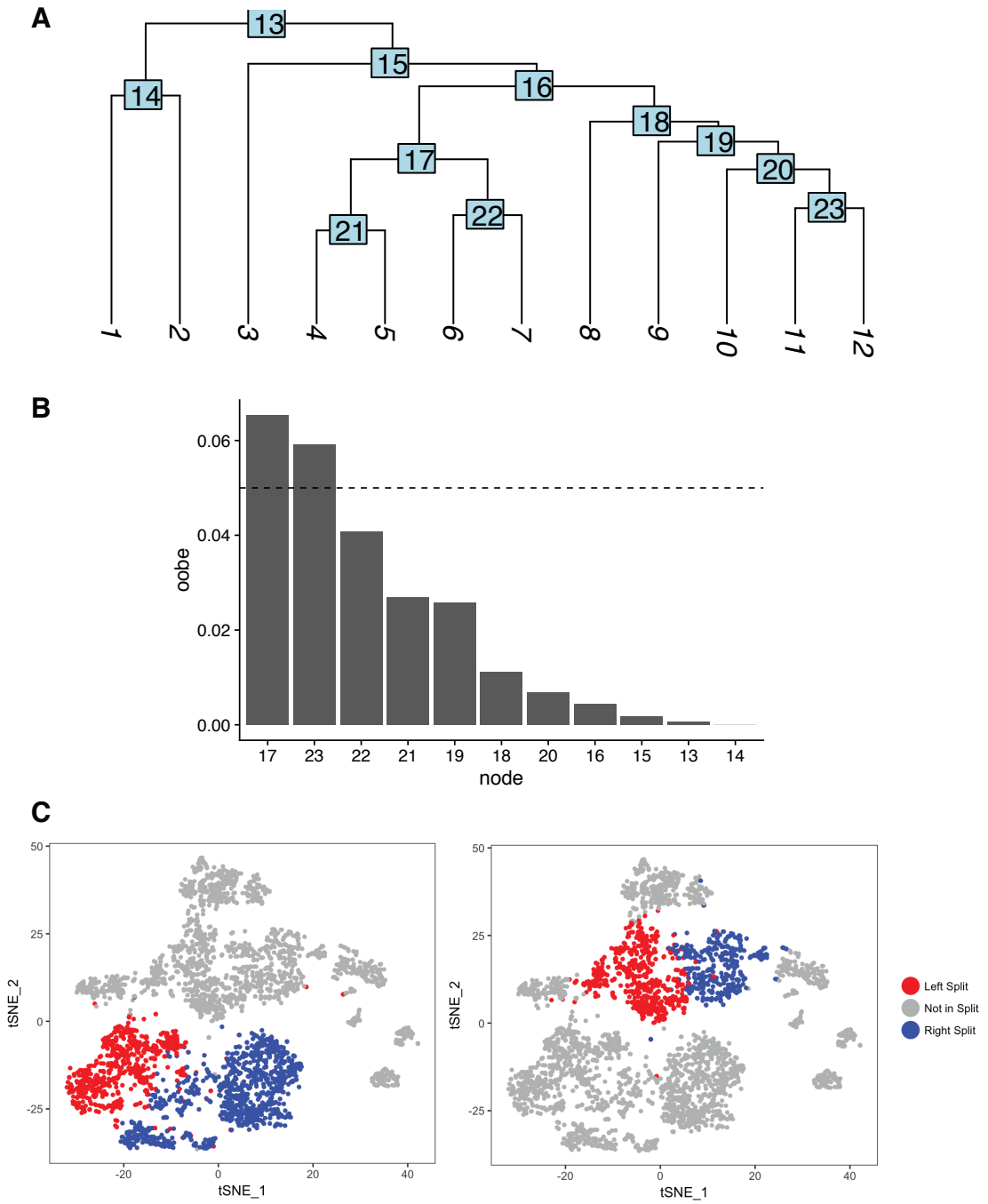


Figure C.12: Merging of similar clusters, DSS dataset. (A) Phylogenetic tree representing inter-cell distances for the average cell in each detected cluster in the DSS dataset. Branch splitting points (nodes) are assigned numeric identifiers. (B) Out of bag error (oobe) for a random forest classifier trained on each internal node (x-axis), reflecting confidence in assigning cells to the left or right branch. A threshold confidence of 0.05 is indicated (dashed line). (C) Sub-clusters defined by low confidence nodes 17 (left panel) and 23 (right panel). Merging was performed for node 23, but not node 17 due to the presence of high-confidence child nodes within the overall structure.

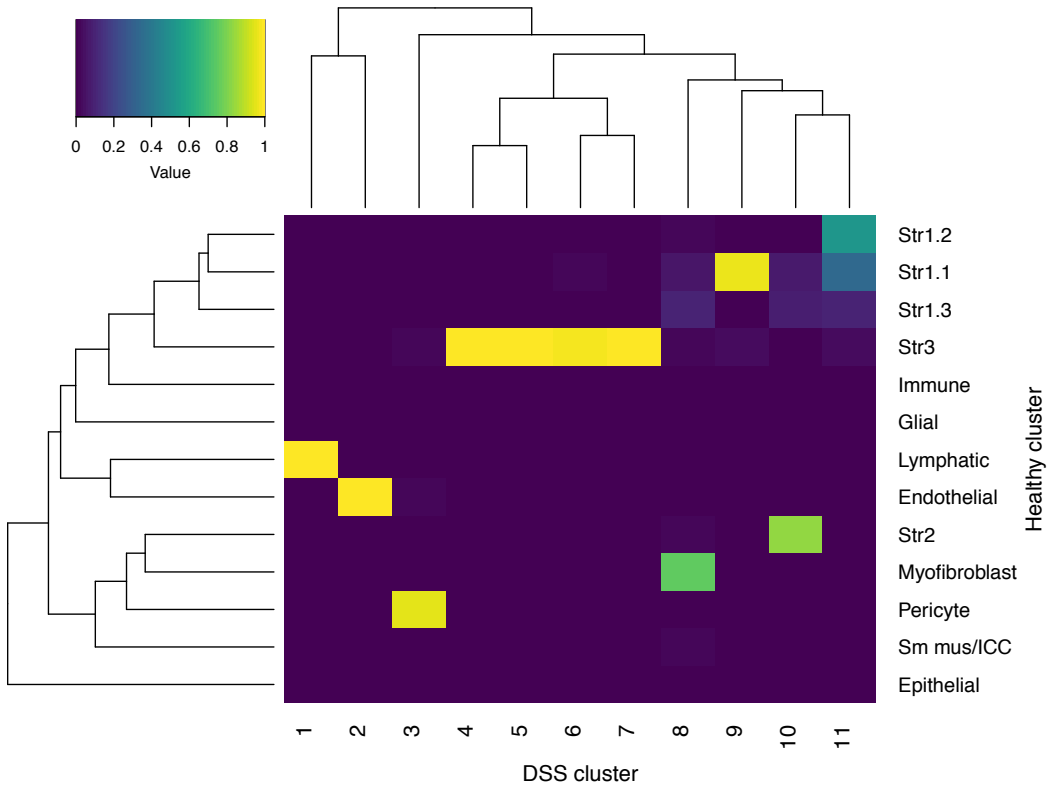


Figure C.13: Random-forest-based cluster identification in the DSS dataset. Cells from the DSS colitis dataset were classified using a graph based clustering approach (clusters 1 to 11, columns) and using a random-forest classifier trained with the healthy dataset (rows). Heat colour indicates the proportion of shared cells. Clusters with the greatest overlap were assigned the same identifier. A one-to-many relationship is observed between Str3 versus clusters 4 to 7.

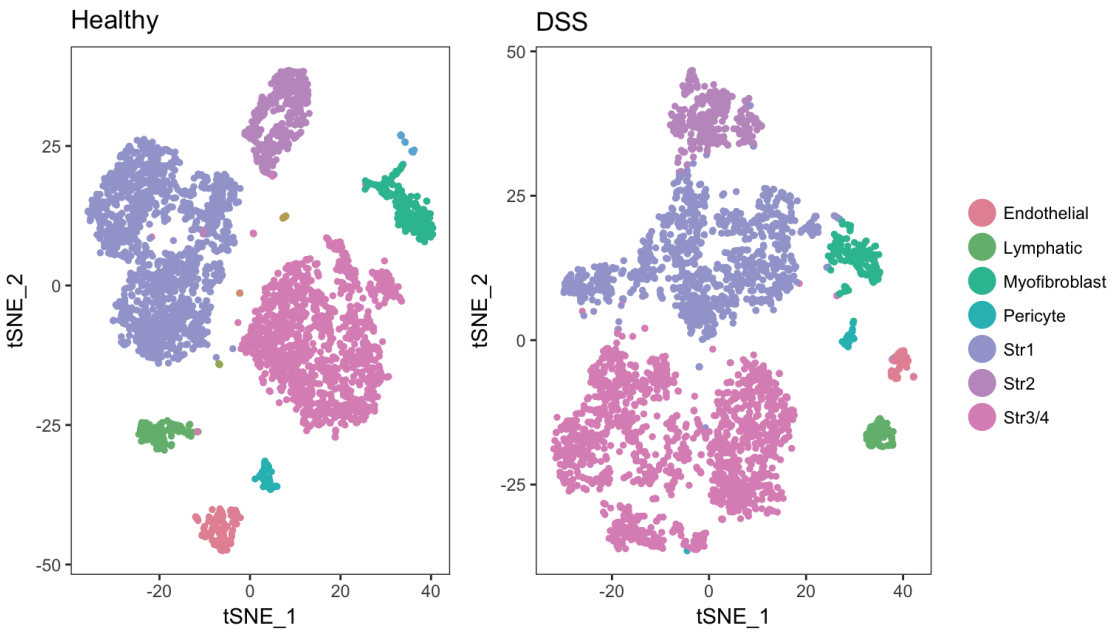


Figure C.14: Major fibroblast subsets as defined for differential expression analysis. TSNE representations of the healthy and DSS datasets with cells coloured by the cluster annotation used for differential expression analysis. Stromal 1 subclusters (Str1.1, Str1.2, Str1.3) and Stromal 3 subclusters (Str3.1, Str3.2, Str3.3, Str4) have been merged.

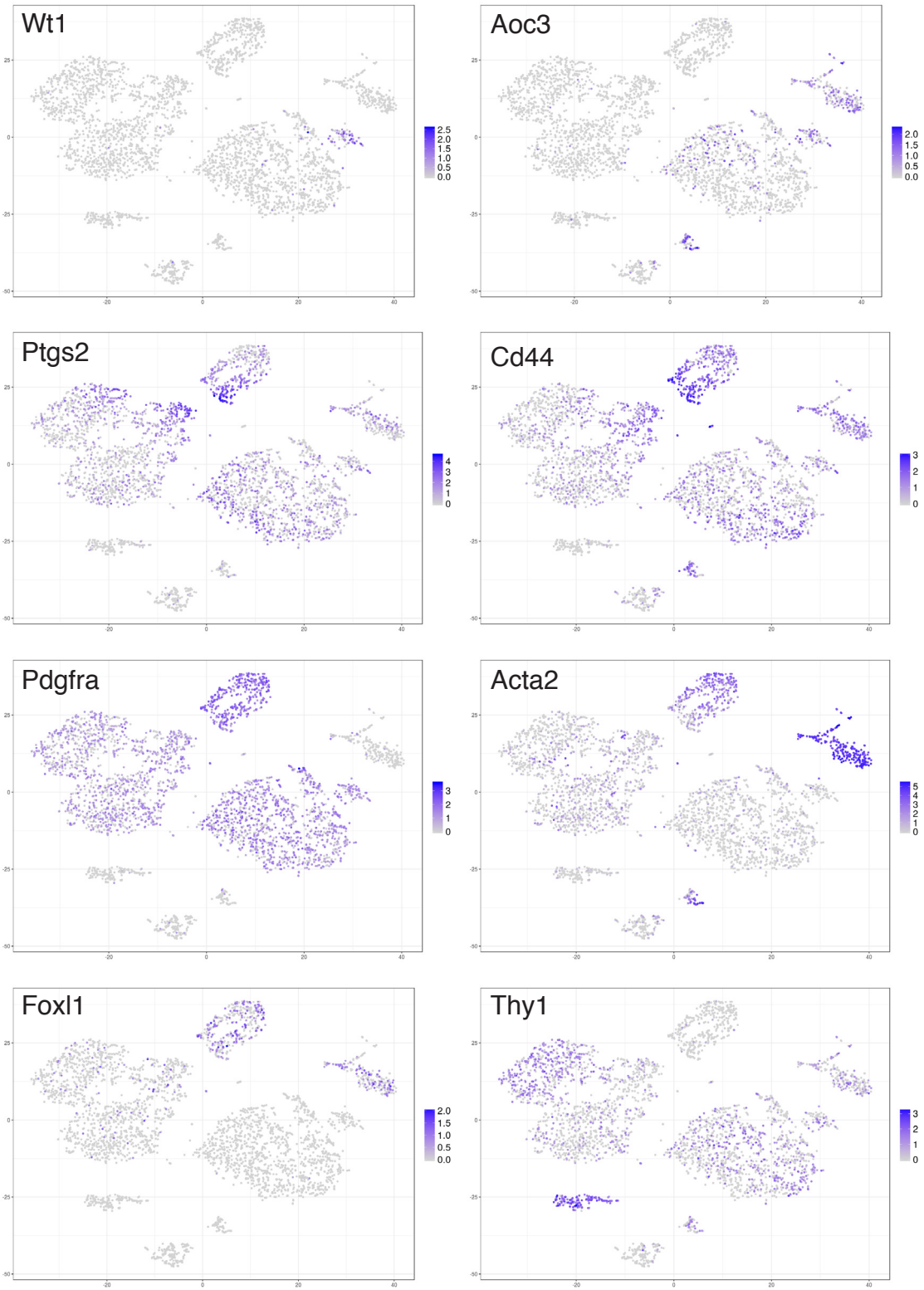


Figure C.15: Additional marker t-SNEs. TSNE representation of the healthy dataset with cells coloured by expression of the markers indicated.

References

[1] Hans Clevers. “The Intestinal Crypt, A Prototype Stem Cell Compartment”. In: *Cell* 154.2 (2013), pp. 274–284.

[2] Lora V Hooper, Dan R Littman, and Andrew J Macpherson. “Interactions Between the Microbiota and the Immune System”. In: *Science* 336.6086 (June 2012), pp. 1268–1273.

[3] Allan M Mowat and William W Agace. “Regional specialization within the intestinal immune system”. In: *Nat. Rev. Immunol.* 14.10 (Oct. 2014), pp. 667–685.

[4] Nathaniel R West et al. “Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease”. In: *Nat. Med.* 7 (Apr. 2017), p. 359.

[5] Matthew Kraman et al. “Suppression of antitumor immunity by stromal cells expressing fibroblast activation protein-alpha.” In: *Science* 330.6005 (Nov. 2010), pp. 827–830.

[6] H Toyoda et al. “Organization of the Lamina propria mucosae of Rat Intestinal Mucosa, with Special Reference to the Subepithelial Connective Tissue”. In: *CTO* 158.3 (1997), pp. 172–184.

[7] Robert R Pascal, Gordon I Kaye, and M D Nathan Lane. “Colonic Pericryptal Fibroblast Sheath: Replication, Migration, and Cytodifferentiation of a Mesenchymal Cell System in Adult Tissue”. In: *Gastroenterology* 54.5 (May 1968), pp. 835–851.

[8] Nancy C Joyce, Marcy F Haire, and George E Palade. “Morphologic and biochemical evidence for a contractile cell network within the rat intestinal mucosa”. In: *Gastroenterology* 92.1 (Jan. 1987), pp. 68–81.

[9] D W Powell et al. “Mesenchymal Cells of the Intestinal Lamina Propria”. In: *Annu. Rev. Physiol.* 73.1 (Mar. 2011), pp. 213–237.

[10] B M J Owens and A Simmons. “Intestinal stromal cells in mucosal immunity and homeostasis”. In: *Mucosal Immunology* 6.2 (Mar. 2013), pp. 224–234.

[11] Roni Nowarski, Ruaidhri Jackson, and Richard A Flavell. “The Stromal Intervention: Regulation of Immunity and Inflammation at the Epithelial-Mesenchymal Barrier”. In: *Cell* 168.3 (Jan. 2017), pp. 362–375.

[12] Manolis Roulis and Richard A Flavell. “Fibroblasts and myofibroblasts of the intestinal lamina propria in physiology and disease”. In: *Differentiation* 92.3 (2016), pp. 116–131.

[13] Brian Eyden, Alan Curry, and Guofeng Wang. "Stromal cells in the human gut show ultrastructural features of fibroblasts and smooth muscle cells but not myofibroblasts". In: *J. Cell. Mol. Med.* 15.7 (July 2010), pp. 1483–1491.

[14] I V Pinchuk et al. "Intestinal mesenchymal cells". In: *Curr Gastroenterol Rep* 12.5 (2010), pp. 310–318.

[15] G Gabbiani, G B Ryan, and G Majne. "Presence of modified fibroblasts in granulation tissue and their possible role in wound contraction." In: *Experientia* 27.5 (May 1971), pp. 549–550.

[16] Alexis Desmoulière, Christine Chaponnier, and Giulio Gabbiani. "Tissue repair, contraction, and the myofibroblast". In: *Wound Repair and Regeneration* 13.1 (2005), pp. 7–12.

[17] B Eyden. "The myofibroblast: phenotypic characterization as a prerequisite to understanding its functions in translational medicine." In: *J. Cell. Mol. Med.* 12.1 (Jan. 2008), pp. 22–37.

[18] A N Stratman et al. "Pericyte recruitment during vasculogenic tube assembly stimulates endothelial basement membrane matrix formation." In: *Blood* 114.24 (Dec. 2009), pp. 5091–5101.

[19] L Diaz-Flores et al. "Pericytes. Morphofunction, interactions and pathology in a quiescent and activated mesenchymal cell niche." In: *Histol. Histopathol.* 24.7 (July 2009), pp. 909–969.

[20] Annika Armulik, Guillem Genové, and Christer Betsholtz. "Pericytes: Developmental, Physiological, and Pathological Perspectives, Problems, and Promises". In: *Developmental Cell* 21.2 (Aug. 2011), pp. 193–215.

[21] Annika Armulik, Alexandra Abramsson, and Christer Betsholtz. "Endothelial/Pericyte Interactions". In: *Circulation Research* 97.6 (Sept. 2005), pp. 512–523.

[22] Kohsuke Uchida and Yuichiro Kamikawa. "Muscularis mucosae - the forgotten sibling". In: *Journal of Smooth Muscle Research* 43.5 (2007), pp. 157–177.

[23] Patrick A Adegboyega et al. "Immunohistochemical study of myofibroblasts in normal colonic mucosa, hyperplastic polyps, and adenomatous colorectal polyps." In: *Arch. Pathol. Lab. Med.* 126.7 (July 2002), pp. 829–836.

[24] W H Percy, T H Fromm, and C E Wangsness. "Muscularis mucosae contraction evokes colonic secretion via prostaglandin synthesis and nerve stimulation". In: *AJP: Gastrointestinal and Liver Physiology* 284.2 (Feb. 2003), G213–G220.

[25] B Ma et al. "Immmunohistochemical study of the blood and lymphatic vasculature and the innervation of mouse gut and gut-associated lymphoid tissue." In: *Anatom Histol Embryol* 36.1 (Feb. 2007), pp. 62–74.

[26] B B Madison. "Epithelial hedgehog signals pattern the intestinal crypt-villus axis." In: *Development* 132.2 (Jan. 2005), pp. 279–289.

[27] M Brittan et al. "Bone marrow derivation of pericryptal myofibroblasts in the mouse and human small intestine and colon." In: *Gut* 50.6 (June 2002), pp. 752–757.

[28] Laura Koumas et al. “Thy-1 expression in human fibroblast subsets defines myofibroblastic or lipofibroblastic phenotypes.” In: *The American Journal of Pathology* 163.4 (Oct. 2003), pp. 1291–1300.

[29] Laura Koumas, Terry J Smith, and Richard P Phipps. “Fibroblast subsets in the human orbit: Thy-1+ and Thy-1- subpopulations exhibit distinct phenotypes”. In: *Eur. J. Immunol.* 32.2 (Feb. 2002), pp. 477–485.

[30] Irina V Pinchuk et al. “PD-1 ligand expression by human colonic myofibroblasts/fibroblasts regulates CD4+ T-cell activity.” In: *Gastroenterology* 135.4 (Oct. 2008), 1228–1237.e2.

[31] Masaaki Kurahashi et al. “A novel population of subepithelial platelet-derived growth factor receptor α-positive cells in the mouse and human colon”. In: *AJP: Gastrointestinal and Liver Physiology* 304.9 (May 2013), G823–G834.

[32] P I Richman et al. “Colonic pericrypt sheath cells: characterisation of cell type with new monoclonal antibody.” In: *J Clin Pathol* 40.6 (June 1987), pp. 593–600.

[33] Lin-Ting Hsia et al. “Myofibroblasts are distinguished from activated skin fibroblasts by the expression of AOC3 and other associated markers”. In: *Proceedings of the National Academy of Sciences* 113.15 (2016), E2162–E2171.

[34] M Roulis et al. “Intestinal myofibroblast-specific Tpl2-Cox-2-PGE2 pathway links innate sensing to epithelial homeostasis”. In: *Proceedings of the National Academy of Sciences* 111.43 (2014), E4658–E4667.

[35] Sarah L Brown et al. “Myd88-dependent positioning of Ptgs2-expressing stromal cells maintains colonic epithelial proliferation during injury”. In: *J. Clin. Invest.* 117.1 (Jan. 2007), pp. 258–269.

[36] Nicholas A Manieri et al. “Igf2bp1 Is Required for Full Induction of Ptgs2 mRNA in Colonic Mesenchymal Stem Cells in Mice”. In: *Gastroenterology* 143.1 (July 2012), 110–121.e10.

[37] L da Silva Meirelles. “Mesenchymal stem cells reside in virtually all post-natal organs and tissues.” In: *J. Cell. Sci.* 119.11 (May 2006), pp. 2204–2213.

[38] Don W Powell and Jamal I Saada. “Mesenchymal Stem Cells and Prostaglandins May Be Critical for Intestinal Wound Repair”. In: *Gastroenterology* 143.1 (July 2012), pp. 19–22.

[39] K H Kaestner et al. “The mesenchymal winged helix transcription factor Fkh6 is required for the control of gastrointestinal proliferation and differentiation.” In: *Genes & Development* 11.12 (June 1997), pp. 1583–1595.

[40] Reina Aoki et al. “Foxl1-Expressing Mesenchymal Cells Constitute the Intestinal Stem Cell Niche”. In: *Cellular and Molecular Gastroenterology and Hepatology* 2.2 (Mar. 2016), pp. 175–188.

[41] Daniel L Worthley et al. “Gremlin 1 Identifies a Skeletal Stem Cell with Bone, Cartilage, and Reticular Stromal Potential”. In: *Cell* 160.1-2 (Jan. 2015), pp. 269–284.

[42] Anne L Fletcher, Sophie E Acton, and Konstantin Knoblich. “Lymph node fibroblastic reticular cells in health and disease”. In: *Nat. Rev. Immunol.* 15.6 (June 2015), pp. 350–361.

[43] Ryan R Driskell and Fiona M Watt. "Understanding fibroblast heterogeneity in the skin." In: *Trends in Cell Biology* 25.2 (Feb. 2015), pp. 92–99.

[44] Howard Y Chang et al. "Diversity, topographic differentiation, and positional memory in human fibroblasts." In: *Proceedings of the National Academy of Sciences* 99.20 (Oct. 2002), pp. 12877–12882.

[45] Youichi Higuchi et al. "Gastrointestinal Fibroblasts Have Specialized, Diverse Transcriptional Phenotypes: A Comprehensive Gene Expression Analysis of Human Fibroblasts". In: *PLoS ONE* 10.6 (June 2015), e0129241.

[46] L A Schildmeyer et al. "Impaired vascular contractility and blood pressure homeostasis in the smooth muscle alpha-actin null mouse." In: *The FASEB Journal* 14.14 (Nov. 2000), pp. 2213–2220.

[47] Tiegang Liu et al. "Lack of MK2 inhibits myofibroblast formation and exacerbates pulmonary fibrosis." In: *Am J Respir Cell Mol Biol* 37.5 (Nov. 2007), pp. 507–517.

[48] Nobuo Terada et al. "Immunohistochemical study of NG2 chondroitin sulfate proteoglycan expression in the small and large intestines." In: *Histochem Cell Biol* 126.4 (Apr. 2006), pp. 483–490.

[49] Yunchao Chang et al. "Ablation of NG2 Proteoglycan Leads to Deficits in Brown Fat Function and to Adult Onset Obesity". In: *PLoS ONE* 7.1 (Jan. 2012), e30637.

[50] Hirokazu Okada et al. "Selective depletion of fibroblasts preserves morphology and the functional integrity of peritoneum in transgenic mice with peritoneal fibrosing syndrome". In: *Kidney International* 64.5 (Nov. 2003), pp. 1722–1732.

[51] Yoshinao Kikuchi et al. "Periostin Is Expressed in Pericryptal Fibroblasts and Cancer-associated Fibroblasts in the Colon". In: *Journal of Histochemistry & Cytochemistry* 56.8 (Aug. 2008), pp. 753–764.

[52] Valérie A McLin, Susan J Henning, and Milan Jamrich. "The Role of the Visceral Mesoderm in the Development of the Gastrointestinal Tract". In: *Gastroenterology* 136.7 (June 2009), pp. 2074–2091.

[53] Rosa Artells et al. "Ultrastructural and Immunohistochemical Analysis of Intestinal Myofibroblasts During the Early Organogenesis of the Human Small Intestine". In: *The Anatomical Record* 294.3 (Mar. 2011), pp. 462–471.

[54] Florian Obermayr et al. "Development and developmental disorders of the enteric nervous system". In: *Nat Rev Gastroenterol Hepatol* 10.1 (Dec. 2012), pp. 43–57.

[55] Yuval Rinkevich et al. "Identification and prospective isolation of a mesothelial precursor lineage giving rise to smooth muscle cells and fibroblasts for mammalian internal organs, and their vasculature". In: *Nat. Cell Biol.* 14.12 (Nov. 2012), pp. 1251–1260.

[56] Bettina Wilm et al. "The serosal mesothelium is a major source of smooth muscle cells of the gut vasculature". In: *Development* 132.23 (Dec. 2005), pp. 5317–5328.

[57] Gordon I Kaye, Nathan Lane, and Robert R Pascal. "Colonic Pericryptal Fibroblast Sheath: Replication, Migration, and Cytodifferentiation of a Mesenchymal Cell System in Adult Tissue". In: *Gastroenterology* 54.5 (May 1968), pp. 852–865.

[58] Gordon I Kaye, Robert R Pascal, and Nathan Lane. "The Colonic Pericryptal Fibroblast Sheath: Replication, Migration, and Cytodifferentiation of a Mesenchymal Cell System in Adult Tissue". In: *Gastroenterology* 60.4 (Apr. 1971), pp. 515–536.

[59] J V Neal and C S Potten. "Description and basic cell kinetics of the murine pericryptal fibroblast sheath." In: *Gut* 22.1 (Jan. 1981), pp. 19–24.

[60] Samy Lamouille, Jian Xu, and Rik Derynck. "Molecular mechanisms of epithelial–mesenchymal transition". In: *Nat. Rev. Mol. Cell Biol.* 15.3 (Mar. 2014), pp. 178–196.

[61] Florian Rieder, Claudio Fiocchi, and Gerhard Rogler. "Mechanisms, Management, and Treatment of Fibrosis in Patients With Inflammatory Bowel Diseases". In: *Gastroenterology* 152.2 (2017), 340–350.e6.

[62] Sarah N Flier et al. "Identification of epithelial to mesenchymal transition as a novel source of fibroblasts in intestinal fibrosis." In: *J. Biol. Chem.* 285.26 (June 2010), pp. 20202–20212.

[63] Benjamin D Humphreys et al. "Fate Tracing Reveals the Pericyte and Not Epithelial Origin of Myofibroblasts in Kidney Fibrosis". In: *The American Journal of Pathology* 176.1 (Jan. 2010), pp. 85–97.

[64] Kojiro Taura et al. "Hepatocytes do not undergo epithelial-mesenchymal transition in liver fibrosis in mice". In: *Hepatology* 51.3 (Mar. 2010), pp. 1027–1036.

[65] Valerie S LeBleu et al. "Origin and function of myofibroblasts in kidney fibrosis". In: *Nat. Med.* 19.8 (Aug. 2013), pp. 1047–1053.

[66] R C Mifflin et al. "Intestinal myofibroblasts: targets for stem cell therapy". In: *AJP: Gastrointestinal and Liver Physiology* 300.5 (2011), G684–G696.

[67] Makio Ogawa, Amanda C LaRue, and Christopher J Drake. "Hematopoietic origin of fibroblasts/myofibroblasts: its pathophysiologic implications". In: *Blood* 108.9 (Nov. 2006), pp. 2893–2896.

[68] Ronald A Reilkoff, Richard Bucala, and Erica L Herzog. "Fibrocytes: emerging effector cells in chronic inflammation". In: *Nat. Rev. Immunol.* 11.6 (June 2011), pp. 427–435.

[69] R Abe et al. "Peripheral blood fibrocytes: differentiation pathway and migration to wound sites." In: *J Immunol* 166.12 (June 2001), pp. 7556–7562.

[70] Matthias Schmidt et al. "Identification of circulating fibrocytes as precursors of bronchial myofibroblasts in asthma." In: *J Immunol* 171.1 (July 2003), pp. 380–389.

[71] Nora G Singer and Arnold I Caplan. "Mesenchymal stem cells: mechanisms of inflammation." In: *Annu Rev Pathol* 6 (2011), pp. 457–478.

[72] Nicholas A Manieri and Thaddeus S Stappenbeck. "Mesenchymal stem cell therapy of intestinal disease: are their effects systemic or localized?" In: *Curr. Opin. Gastroenterol.* 27.2 (Mar. 2011), pp. 119–124.

[73] Jan Voswinkel et al. "Use of Mesenchymal Stem Cells (MSC) in Chronic Inflammatory Fistulizing and Fibrotic Diseases: a Comprehensive Review". In: *Clinic Rev Allerg Immunol* 45.2 (Jan. 2013), pp. 180–192.

[74] Cynthia Kosinski et al. “Indian Hedgehog Regulates Intestinal Stem Cell Fate Through Epithelial-Mesenchymal Interactions During Development”. In: *Gastroenterology* 139.3 (Sept. 2010), pp. 893–903.

[75] J Mao et al. “Hedgehog signaling controls mesenchymal growth in the developing mammalian digestive tract”. In: *Development* 137.10 (Apr. 2010), pp. 1721–1729.

[76] Tae-Hee Kim et al. “Endodermal Hedgehog signals modulate Notch pathway activity in the developing digestive tract mesenchyme”. In: *Development* 138.15 (Aug. 2011), pp. 3225–3233.

[77] Michael J Geske et al. “Fgf9 signaling regulates small intestinal elongation and mesenchymal development”. In: *Development* 135.17 (Sept. 2008), pp. 2959–2968.

[78] L Karlsson et al. “Abnormal gastrointestinal development in PDGF-A and PDGFR-alpha deficient mice implicates a novel mesenchymal structure with putative instructive properties in villus morphogenesis.” In: *Development* 127.16 (Aug. 2000), pp. 3457–3466.

[79] Anna-Pavlina G Haramis et al. “De Novo Crypt Formation and Juvenile Polyposis on BMP Inhibition in Mouse Intestine”. In: *Science* 303.5664 (Mar. 2004), pp. 1684–1686.

[80] Nikè V J A Büller et al. “Hedgehog signaling and maintenance of homeostasis in the intestinal epithelium.” In: *Physiology (Bethesda)* 27.3 (June 2012), pp. 148–155.

[81] Gijs R van den Brink et al. “Indian Hedgehog is an antagonist of Wnt signaling in colonic epithelial cell differentiation”. In: *Nat Genet* 36.3 (Feb. 2004), pp. 277–282.

[82] Taeko K Noah and Noah F Shroyer. “Notch in the Intestine: Regulation of Homeostasis and Pathogenesis”. In: *Annu. Rev. Physiol.* 75.1 (Feb. 2013), pp. 263–288.

[83] Rocio Sancho, Catherine A Cremona, and Axel Behrens. “Stem cell and progenitor fate in the mammalian intestine: Notch and lateral inhibition in homeostasis and disease”. In: *EMBO Reports* 16.5 (May 2015), pp. 571–581.

[84] David M Ornitz and Nobuyuki Itoh. “The Fibroblast Growth Factor signaling pathway”. In: *Wiley Interdisciplinary Reviews: Developmental Biology* 4.3 (May 2015), pp. 215–266.

[85] Stacey J Coleman et al. “The ins and outs of fibroblast growth factor receptor signalling”. In: *Clinical Science* 127.4 (July 2014), pp. 217–231.

[86] Xiuqin Zhang et al. “Reciprocal epithelial-mesenchymal FGF signaling is required for cecal development”. In: *Development* 133.1 (Jan. 2006), pp. 173–180.

[87] Johanna Andrae, Radiosa Gallini, and Christer Betsholtz. “Role of platelet-derived growth factors in physiology and medicine.” In: *Genes & Development* 22.10 (May 2008), pp. 1276–1312.

[88] Michael C Heinrich et al. “PDGFRA Activating Mutations in Gastrointestinal Stromal Tumors”. In: *Science* 299.5607 (Jan. 2003), pp. 708–710.

[89] Kristian Pietras et al. “Functions of Paracrine PDGF Signaling in the Proangiogenic Tumor Stroma Revealed by Pharmacological Targeting”. In: *PLOS Medicine* 5.1 (Jan. 2008), e19.

[90] Carl-Henrik Heldin. "Targeting the PDGF signaling pathway in tumor treatment". In: *Cell Communication and Signaling : CCS* 11.1 (2013), p. 97.

[91] Kohei Miyazono, Yuto Kamiya, and Masato Morikawa. "Bone morphogenetic protein receptors and signal transduction". In: *J Biochem* 147.1 (Jan. 2010), pp. 35–51.

[92] C Kosinski et al. "Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors". In: *Proceedings of the National Academy of Sciences* 104.39 (2007), pp. 15418–15423.

[93] Willemijn A van Dop et al. "Depletion of the Colonic Epithelial Precursor Cell Compartment Upon Conditional Activation of the Hedgehog Pathway". In: *Gastroenterology* 136.7 (June 2009), 2195–2203.e7.

[94] Xi C He et al. "BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt–β-catenin signaling". In: *Nat Genet* 36.10 (Sept. 2004), pp. 1117–1121.

[95] Lorene E Batts et al. "Bmp signaling is required for intestinal growth and morphogenesis". In: *Dev. Dyn.* 235.6 (June 2006), pp. 1563–1570.

[96] Christine Fritsch et al. "Epimorphin expression in intestinal myofibroblasts induces epithelial morphogenesis". In: *J. Clin. Invest.* 110.11 (Dec. 2002), pp. 1629–1641.

[97] Yuan Wang et al. "Epimorphin–/– mice have increased intestinal growth, decreased susceptibility to dextran sodium sulfate colitis, and impaired spermatogenesis". In: *Journal of Clinical Investigation* 116.6 (June 2006), pp. 1535–1546.

[98] Anisa Shaker et al. "Epimorphin deletion protects mice from inflammation-induced colon carcinogenesis and alters stem cell niche myofibroblast secretion". In: *J. Clin. Invest.* 120.6 (June 2010), pp. 2081–2093.

[99] Toshiro Sato et al. "Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche." In: *Nature* 459.7244 (May 2009), pp. 262–265.

[100] David Wei-Min Tan and Nick Barker. "Intestinal Stem Cells and Their Defining Niche". In: *Stem Cells in Development and Disease*. Elsevier, 2014, pp. 77–107.

[101] Tea Fevr et al. "Wnt beta-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells." In: *Mol. Cell. Biol.* 27.21 (Nov. 2007), pp. 7551–7559.

[102] Laurens G van der Flier et al. "The Intestinal Wnt/TCF Signature". In: *Gastroenterology* 132.2 (Feb. 2007), pp. 628–632.

[103] Nick Barker. "Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration". In: *Nat. Rev. Mol. Cell Biol.* 15.1 (Jan. 2014), pp. 19–33.

[104] Toshiro Sato et al. "Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium". In: *Gastroenterology* 141.5 (2011), pp. 1762–1772.

[105] Yong-Ri Jin and Jeong Kyo Yoon. "The R-spondin family of proteins: emerging regulators of WNT signaling." In: *The International Journal of Biochemistry & Cell Biology* 44.12 (Dec. 2012), pp. 2278–2287.

[106] Kelley S Yan et al. “Non-equivalence of Wnt and R-spondin ligands during Lgr5+ intestinal stem-cell self-renewal.” In: *Nature* 545.7653 (May 2017), pp. 238–242.

[107] Nicholas Lahar et al. “Intestinal Subepithelial Myofibroblasts Support in vitro and in vivo Growth of Human Small Intestinal Epithelium”. In: *PLoS ONE* 6.11 (Nov. 2011), e26898.

[108] Nan Ye Lei et al. “Intestinal Subepithelial Myofibroblasts Support the Growth of Intestinal Epithelial Stem Cells”. In: *PLoS ONE* 9.1 (2014), e84651.

[109] Henner F Farin, Johan H Van Es, and Hans Clevers. “Redundant Sources of Wnt Regulate Intestinal Stem Cells and Promote Formation of Paneth Cells”. In: *Gastroenterology* 143.6 (Dec. 2012), 1518–1529.e7.

[110] Zahra Kabiri et al. “Stroma provides an intestinal stem cell niche in the absence of epithelial Wnts”. In: *Development* 141.11 (June 2014), pp. 2206–2215.

[111] Xing Li et al. “Deconvoluting the intestine: molecular evidence for a major role of the mesenchyme in the modulation of signaling cross talk”. In: *Physiol. Genomics* 29.3 (May 2007), pp. 290–301.

[112] A Gregorieff et al. “Expression pattern of Wnt signaling components in the adult intestine”. In: *Gastroenterology* 129.2 (2005), pp. 626–638.

[113] Michael J Nemeth et al. “Wnt5a inhibits canonical Wnt signaling in hematopoietic stem cells and enhances repopulation.” In: *Proceedings of the National Academy of Sciences* 104.39 (Sept. 2007), pp. 15436–15441.

[114] Taro Kawai and Shizuo Akira. “The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors”. In: *Nature Immunology* 11.5 (May 2010), pp. 373–384.

[115] Jan-Michel Otte, Ian M Rosenberg, and Daniel K Podolsky. “Intestinal myofibroblasts in innate immune responses of the intestine”. In: *Gastroenterology* 124.7 (June 2003), pp. 1866–1878.

[116] Benjamin M J Owens et al. “CD90+ stromal cells are non-professional innate immune effectors of the human colonic mucosa”. In: *Front. Immunol.* 4 (2013), p. 307.

[117] Yun-Gi Kim et al. “The Nod2 Sensor Promotes Intestinal Pathogen Eradication via the Chemokine CCL2-Dependent Recruitment of Inflammatory Monocytes”. In: *Immunity* 34.5 (May 2011), pp. 769–780.

[118] Kristen L W Walton, Lisa Holt, and R Balfour Sartor. “Lipopolysaccharide activates innate immune responses in murine intestinal myofibroblasts through multiple signaling pathways”. In: *AJP: Gastrointestinal and Liver Physiology* 296.3 (Mar. 2009), G601–G611.

[119] G Pang et al. “GM-CSF, IL-1 alpha, IL-1 beta, IL-6, IL-8, IL-10, ICAM-1 and VCAM-1 gene expression and cytokine production in human duodenal fibroblasts stimulated with lipopolysaccharide, IL-1 alpha and TNF-alpha.” In: *Clin. Exp. Immunol.* 96.3 (June 1994), pp. 437–443.

[120] Zhoubin Zhang et al. “Interleukin-17 and lipopolysaccharides synergistically induce cyclooxygenase-2 expression in human intestinal myofibroblasts”. In: *J Gastroenterol Hepatol* 20.4 (Apr. 2005), pp. 619–627.

[121] Irina V Pinchuk et al. "Monocyte Chemoattractant Protein-1 Production by Intestinal Myofibroblasts in Response to Staphylococcal Enterotoxin A: Relevance to Staphylococcal Enterotoxigenic Disease". In: *J Immunol* 178.12 (June 2007), pp. 8097–8106.

[122] Sylvain Normand et al. "Nod-like receptor pyrin domain-containing protein 6 (NLRP6) controls epithelial self-renewal and colorectal carcinogenesis upon injury." In: *Proc. Natl. Acad. Sci. U.S.A.* 108.23 (June 2011), pp. 9601–9606.

[123] Kristina Scheibe et al. "IL-36R signalling activates intestinal epithelial cells and fibroblasts and promotes mucosal healing in vivo". In: *Gut* 66.5 (May 2017), pp. 823–838.

[124] E Lippert et al. "Soluble galectin-3 is a strong, colonic epithelial-cell-derived, lamina propria fibroblast-stimulating factor". In: *Gut* 56.1 (Jan. 2007), pp. 43–51.

[125] Akira Andoh et al. "Expression of IL-24, an Activator of the JAK1/STAT3/SOCS3 Cascade, Is Enhanced in Inflammatory Bowel Disease". In: *J Immunol* 183.1 (July 2009), pp. 687–695.

[126] Laurens Kruidenier et al. "Myofibroblast Matrix Metalloproteinases Activate the Neutrophil Chemoattractant CXCL7 From Intestinal Epithelial Cells". In: *Gastroenterology* 130.1 (Jan. 2006), pp. 127–136.

[127] T Okuno et al. "Interleukin-1beta and tumor necrosis factor-alpha induce chemokine and matrix metalloproteinase gene expression in human colonic subepithelial myofibroblasts." In: *Scand J Gastroenterol* 37.3 (Mar. 2002), pp. 317–324.

[128] John F Di Mari et al. "IL-1alpha-induced COX-2 expression in human intestinal myofibroblasts is dependent on a PKCzeta-ROS pathway." In: *Gastroenterology* 124.7 (June 2003), pp. 1855–1865.

[129] T De L Karlson, C V Whiting, and P W Bland. "Proinflammatory cytokine synthesis by mucosal fibroblasts from mouse colitis is enhanced by interferon-gamma-mediated up-regulation of CD40 signalling." In: *Clin. Exp. Immunol.* 147.2 (Feb. 2007), pp. 313–323.

[130] Shigeki Bamba et al. "Regulation of IL-11 expression in intestinal myofibroblasts: role of c-Jun AP-1- and MAPK-dependent pathways". In: *AJP: Gastrointestinal and Liver Physiology* 285.3 (Sept. 2003), G529–G538.

[131] Kazunori Hata et al. "IL-17 stimulates inflammatory responses via NF-κB and MAP kinase pathways in human colonic myofibroblasts". In: *AJP: Gastrointestinal and Liver Physiology* 282.6 (June 2002), G1035–G1044.

[132] William G C Horsnell et al. "Delayed Goblet Cell Hyperplasia, Acetylcholine Receptor Expression, and Worm Expulsion in SMC-Specific IL-4Rα–Deficient Mice". In: *PLOS Pathog* 3.1 (2007), e1.

[133] Reece G Marillier et al. "IL-4R-alpha-responsive smooth muscle cells increase intestinal hypercontractility and contribute to resistance during acute Schistosomiasis." In: *American Journal of Physiology - Gastrointestinal and Liver Physiology* 298.6 (June 2010), G943–51.

[134] G T Furuta et al. “Eosinophil granule-derived major basic protein induces IL-8 expression in human intestinal myofibroblasts”. In: *Clin. Exp. Immunol.* 122.1 (Oct. 2000), pp. 35–40.

[135] Yves Montier et al. “Central role of IL-6 and MMP-1 for cross talk between human intestinal mast cells and human intestinal fibroblasts”. In: *Immunobiology* 217.9 (Sept. 2012), pp. 912–919.

[136] G Rogler et al. “Differential activation of cytokine secretion in primary human colonic fibroblast/myofibroblast cultures.” In: *Scand J Gastroenterol* 36.4 (Apr. 2001), pp. 389–398.

[137] Akira Andoh et al. “Interleukin-17 augments tumor necrosis factor-alpha-induced granulocyte and granulocyte/macrophage colony-stimulating factor release from human colonic myofibroblasts.” In: *J. Gastroenterol.* 40.8 (Aug. 2005), pp. 802–810.

[138] I Vicente-Suarez et al. “Unique lamina propria stromal cells imprint the functional phenotype of mucosal dendritic cells”. In: *Mucosal Immunology* 8.1 (June 2014), pp. 141–151.

[139] Grazia Maria Spaggiari et al. “MSCs inhibit monocyte-derived DC maturation and function by selectively interfering with the generation of immature DCs: central role of MSC-derived prostaglandin E2”. In: *Blood* 113.26 (June 2009), pp. 6576–6583.

[140] Nikolas T Martin and Michael U Martin. “Interleukin 33 is a guardian of barriers and a local alarmin”. In: *Nature Immunology* 17.2 (Jan. 2016), pp. 122–131.

[141] Jon Sponheim et al. “Inflammatory Bowel Disease-Associated Interleukin-33 Is Preferentially Expressed in Ulceration-Associated Myofibroblasts”. In: *The American Journal of Pathology* 177.6 (Dec. 2010), pp. 2804–2815.

[142] Chun-Kwok Wong et al. “Activation of Eosinophils Interacting with Dermal Fibroblasts by Pruritogenic Cytokine IL-31 and Alarmin IL-33: Implications in Atopic Dermatitis”. In: *PLoS ONE* 7.1 (Jan. 2012), e29815.

[143] Damo Xu et al. “IL-33 exacerbates antigen-induced arthritis by activating mast cells.” In: *Proc. Natl. Acad. Sci. U.S.A.* 105.31 (Aug. 2008), pp. 10913–10918.

[144] Keisuke Oboki et al. “IL-33 is a crucial amplifier of innate rather than acquired immunity.” In: *Proc. Natl. Acad. Sci. U.S.A.* 107.43 (Oct. 2010), pp. 18581–18586.

[145] Lihua Duan et al. “Interleukin-33 ameliorates experimental colitis through promoting Th2/Foxp3+ regulatory T-cell responses in mice.” In: *Mol. Med.* 18 (July 2012), pp. 753–761.

[146] Chris Schiering et al. “The alarmin IL-33 promotes regulatory T-cell function in the intestine”. In: *Nature* 513.7519 (2014), pp. 564–568.

[147] N W Palm et al. “Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease”. In: *Cell* 158.5 (2014), pp. 1000–1010.

[148] Jörg H Fritz et al. “Acquisition of a multifunctional IgA+ plasma cell phenotype in the gut.” In: *Nature* 481.7380 (Dec. 2011), pp. 199–203.

[149] Sidonia Fagarasan et al. “In situ class switching and differentiation to IgA-producing cells in the gut lamina propria”. In: *Nature* 413.6856 (Oct. 2001), pp. 639–643.

[150] Hyung-Sik Kang et al. “Signaling via LTbetaR on the lamina propria stromal cells of the gut is required for IgA production.” In: *Nature Immunology* 3.6 (June 2002), pp. 576–582.

[151] Kenji Ina et al. “Intestinal Fibroblast-Derived IL-10 Increases Survival of Mucosal T Cells by Inhibiting Growth Factor Deprivation- and Fas-Mediated Apoptosis”. In: *J Immunol* 175.3 (Aug. 2005), pp. 2000–2009.

[152] Jamal I Saada et al. “Subepithelial Myofibroblasts are Novel Nonprofessional APCs in the Human Colonic Mucosa”. In: *J Immunol* 177.9 (Nov. 2006), pp. 5968–5979.

[153] Motoko Koyama et al. “Recipient nonhematopoietic antigen-presenting cells are sufficient to induce lethal acute graft-versus-host disease”. In: *Nat. Med.* 18.1 (Nov. 2011), pp. 135–142.

[154] J I Saada et al. “Intestinal Myofibroblasts and Immune Tolerance”. In: *Annals of the New York Academy of Sciences* 1029.1 (Dec. 2004), pp. 379–381.

[155] Ellen J Beswick et al. “TLR4 Activation Enhances the PD-L1–Mediated Tolerogenic Capacity of Colonic CD90+ Stromal Cells”. In: *J Immunol* 193.5 (Sept. 2014), pp. 2218–2229.

[156] Irina V Pinchuk et al. “Human Colonic Myofibroblasts Promote Expansion of CD4+ CD25high Foxp3+ Regulatory T Cells”. In: *Gastroenterology* 140.7 (June 2011), pp. 2019–2030.

[157] Zhoubin Zhang et al. “Interleukin-1beta and tumor necrosis factor-alpha upregulate interleukin-23 subunit p19 gene expression in human colonic subepithelial myofibroblasts.” In: *Int. J. Mol. Med.* 15.1 (Jan. 2005), pp. 79–83.

[158] Oliver Pabst et al. “Cryptopatches and isolated lymphoid follicles: dynamic lymphoid tissues dispensable for the generation of intraepithelial lymphocytes”. In: *Eur. J. Immunol.* 35.1 (Jan. 2005), pp. 98–107.

[159] A D O’Leary and E C Sweeney. “Lymphoglandular complexes of the colon: structure and distribution”. In: *Histopathology* 10.3 (Mar. 1986), pp. 267–283.

[160] Serge A van de Pavert and Reina E Mebius. “New insights into the development of lymphoid tissues”. In: *Nat. Rev. Immunol.* 10.9 (Sept. 2010), pp. 664–674.

[161] Claire Pearson, Holm H Uhlig, and Fiona Powrie. “Lymphoid microenvironments and innate lymphoid cells in the gut”. In: *Trends in Immunology* 33.6 (June 2012), pp. 289–296.

[162] A Iwasaki and B L Kelsall. “Localization of distinct Peyer’s patch dendritic cell subsets and their recruitment by chemokines macrophage inflammatory protein (MIP)-3alpha, MIP-3beta, and secondary lymphoid organ chemokine.” In: *Journal of Experimental Medicine* 191.8 (Apr. 2000), pp. 1381–1394.

[163] Kazuki Nagashima et al. “Identification of subepithelial mesenchymal cells that induce IgA and diversify gut microbiota”. In: *Nature Immunology* 9 (Apr. 2017), p. 618.

[164] Masayuki Tsuji et al. “Requirement for Lymphoid Tissue-Inducer Cells in Isolated Follicle Formation and T Cell-Independent Immunoglobulin A Generation in the Gut”. In: *Immunity* 29.2 (Aug. 2008), pp. 261–271.

[165] Djahida Bouskra et al. “Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis”. In: *Nature* 456.7221 (Nov. 2008), pp. 507–510.

[166] A P Baptista et al. “Colonic Patch and colonic SILT development are independent and differentially-regulated events”. In: *Mucosal Immunology* 6.3 (May 2013), pp. 511–521.

[167] Kathryn A Knoop et al. “Distinct Developmental Requirements for Isolated Lymphoid Follicle Formation in the Small and Large Intestine”. In: *The American Journal of Pathology* 179.4 (Oct. 2011), pp. 1861–1871.

[168] Deepali Malhotra et al. “Transcriptional profiling of stroma from inflamed and resting lymph nodes defines immunological hallmarks”. In: *Nature Immunology* 13.5 (Apr. 2012), pp. 499–510.

[169] Anne L Fletcher et al. “Lymph node fibroblastic reticular cells directly present peripheral tissue antigen under steady-state and inflammatory conditions”. In: *Journal of Experimental Medicine* 207.4 (Apr. 2010), pp. 689–697.

[170] Lucie Peduto et al. “Inflammation Recapitulates the Ontogeny of Lymphoid Stromal Cells”. In: *J Immunol* 182.9 (May 2009), pp. 5789–5799.

[171] Johan Van Limbergen, Graham Radford-Smith, and Jack Satsangi. “Advances in IBD genetics”. In: *Nat Rev Gastroenterol Hepatol* 11.6 (Mar. 2014), pp. 372–385.

[172] Alessandra Geremia et al. “Innate and adaptive immunity in inflammatory bowel disease”. In: *Autoimmunity Reviews* 13.1 (Jan. 2014), pp. 3–10.

[173] T J Creed and C S J Probert. “Review article: steroid resistance in inflammatory bowel disease - mechanisms and therapeutic strategies.” In: *Aliment Pharmacol Ther* 25.2 (Jan. 2007), pp. 111–122.

[174] Kevin J Maloy and Fiona Powrie. “Intestinal homeostasis and its breakdown in inflammatory bowel disease”. In: *Nature* 474.7351 (June 2011), pp. 298–306.

[175] Elena Neumann et al. “Rheumatoid arthritis progression mediated by activated synovial fibroblasts”. In: *Trends in Molecular Medicine* 16.10 (Oct. 2010), pp. 458–468.

[176] Christopher D Buckley et al. “Fibroblasts regulate the switch from acute resolving to chronic persistent inflammation”. In: *Trends in Immunology* 22.4 (Apr. 2001), pp. 199–204.

[177] Sonsoles Piera-Velazquez et al. “Persistent activation of dermal fibroblasts from patients with gadolinium-associated nephrogenic systemic fibrosis”. In: *Annals of the Rheumatic Diseases* 69.11 (Nov. 2010), pp. 2017–2023.

[178] M Li et al. “Emergence of Fibroblasts with a Proinflammatory Epigenetically Altered Phenotype in Severe Hypoxic Pulmonary Hypertension”. In: *J Immunol* 187.5 (Aug. 2011), pp. 2711–2722.

[179] Wibke Bechtel et al. “Methylation determines fibroblast activation and fibrogenesis in the kidney”. In: *Nat. Med.* 16.5 (May 2010), pp. 544–550.

[180] Emmanuel Karouzakis et al. “Increased recycling of polyamines is associated with global DNA hypomethylation in rheumatoid arthritis synovial fibroblasts”. In: *Arthritis Rheum.* 64.6 (June 2012), pp. 1809–1817.

[181] Shirley B Russell et al. “Epigenetically altered wound healing in keloid fibroblasts”. In: *J. Invest. Dermatol.* 130.10 (Oct. 2010), pp. 2489–2496.

[182] Brian C McKaig et al. “Expression and Regulation of Tissue Inhibitor of Metalloproteinase-1 and Matrix Metalloproteinases by Intestinal Myofibroblasts in Inflammatory Bowel Disease”. In: *The American Journal of Pathology* 162.4 (Apr. 2003), pp. 1355–1360.

[183] Markus Lang et al. “Gene expression profiles of mucosal fibroblasts from strictured and nonstrictured areas of patients with Crohn?s disease”. In: *Inflamm Bowel Dis* 15.2 (Feb. 2009), pp. 212–223.

[184] Serena Catarzi et al. “Oxidative state and IL-6 production in intestinal myofibroblasts of Crohn’s disease patients”. In: *Inflamm Bowel Dis* 17.8 (Aug. 2011), pp. 1674–1684.

[185] Filippo Fontani et al. “Redox regulation of MMP-3/TIMP-1 ratio in intestinal myofibroblasts: effect of N-acetylcysteine and curcumin.” In: *Experimental Cell Research* 323.1 (Apr. 2014), pp. 77–86.

[186] Martin J Meijer et al. “Effect of the anti-tumor necrosis factor-alpha antibody infliximab on the ex vivo mucosal matrix metalloproteinase-proteolytic phenotype in inflammatory bowel disease.” In: *Inflamm Bowel Dis* 13.2 (Feb. 2007), pp. 200–210.

[187] Michael Karin and Hans Clevers. “Reparative inflammation takes charge of tissue regeneration”. In: *Nature* 529.7586 (Jan. 2016), pp. 307–315.

[188] Qingzhu Sun et al. “PRMT1 Upregulated by Epithelial Proinflammatory Cytokines Participates in COX2 Expression in Fibroblasts and Chronic Antigen-Induced Pulmonary Inflammation”. In: *J Immunol* 195.1 (July 2015), pp. 298–306.

[189] B C McKaig et al. “Differential expression of TGF-beta isoforms by normal and inflammatory bowel disease intestinal myofibroblasts.” In: *American Journal of Physiology - Cell Physiology* 282.1 (Sept. 2001), pp. C172–C182.

[190] Raghu R Chivukula et al. “An Essential Mesenchymal Function for miR-143/145 in Intestinal Epithelial Regeneration”. In: *Cell* 157.5 (May 2014), pp. 1104–1116.

[191] Dimitris Kontoyiannis et al. “Genetic Dissection of the Cellular Pathways and Signaling Mechanisms in Modeled Tumor Necrosis Factor–induced Crohn’s-like Inflammatory Bowel Disease”. In: *Journal of Experimental Medicine* 196.12 (Dec. 2002), pp. 1563–1574.

[192] Maria Armaka et al. “Mesenchymal cell targeting by TNF as a common pathogenic principle in chronic inflammatory joint and intestinal diseases”. In: *Journal of Experimental Medicine* 205.2 (Feb. 2008), pp. 331–337.

[193] Manolis Roulis et al. “Intestinal epithelial cells as producers but not targets of chronic TNF suffice to cause murine Crohn-like pathology.” In: *Proc. Natl. Acad. Sci. U.S.A.* 108.13 (Mar. 2011), pp. 5396–5401.

[194] Caroline Francoeur et al. “Degeneration of the Pericryptal Myofibroblast Sheath by Proinflammatory Cytokines in Inflammatory Bowel Diseases”. In: *Gastroenterology* 136.1 (Jan. 2009), 268–277.e3.

[195] Hermann Brenner, Matthias Kloor, and Christian Peter Pox. "Colorectal cancer". In: *The Lancet* 383.9927 (Apr. 2014), pp. 1490–1502.

[196] Rona Yaeger et al. "Genomic Alterations Observed in Colitis-Associated Cancers Are Distinct From Those Found in Sporadic Colorectal Cancers and Vary by Type of Inflammatory Bowel Disease". In: *Gastroenterology* 151.2 (Aug. 2016), 278–287.e6.

[197] Vasiliki Koliaraki et al. "Mesenchymal Cells in Colon Cancer". In: *Gastroenterology* 152.5 (Apr. 2017), pp. 964–979.

[198] Sandro Goruppi and G Paolo Dotto. "Mesenchymal stroma: primary determinant and therapeutic target for epithelial cancer". In: *Trends in Cell Biology* 23.12 (Dec. 2013), pp. 593–602.

[199] Eliane Cortez, Pernilla Roswall, and Kristian Pietras. "Functional subsets of mesenchymal cell types in the tumor microenvironment". In: *Seminars in Cancer Biology* 25 (Apr. 2014), pp. 3–9.

[200] Alexandre Calon et al. "Dependency of colorectal cancer on a TGF-beta-driven program in stromal cells for metastasis initiation." In: *Cancer Cell* 22.5 (Nov. 2012), pp. 571–584.

[201] Mina J Bissell and William C Hines. "Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression". In: *Nat. Med.* 17.3 (Mar. 2011), pp. 320–329.

[202] Alexandre Calon et al. "Stromal gene expression defines poor-prognosis subtypes in colorectal cancer". In: *Nat Genet* 47.4 (Apr. 2015), pp. 320–329.

[203] Claudio Isella et al. "Stromal contribution to the colorectal cancer transcriptome". In: *Nat Genet* 47.4 (Feb. 2015), pp. 312–319.

[204] Sofia Torres et al. "Proteome Profiling of Cancer-Associated Fibroblasts Identifies Novel Proinflammatory Signatures and Prognostic Markers for Colorectal Cancer". In: *Clinical Cancer Research* 19.21 (Nov. 2013), pp. 6006–6019.

[205] Elena Sierra-Filardi et al. "CCL2 Shapes Macrophage Polarization by GM-CSF and M-CSF: Identification of CCL2/CCR2-Dependent Gene Expression Profile". In: *J Immunol* 192.8 (Apr. 2014), pp. 3858–3867.

[206] Mercedes Herrera et al. "Cancer-associated fibroblast and M2 macrophage markers together predict outcome in colorectal cancer patients". In: *Cancer Science* 104.4 (Apr. 2013), pp. 437–444.

[207] Jacqueline D Shields et al. "Induction of lymphoidlike stroma and immune escape by tumors that express the chemokine CCL21." In: *Science* 328.5979 (May 2010), pp. 749–752.

[208] Vasiliki Koliaraki, Manolis Pasparakis, and George Kollias. "IKKbeta in intestinal mesenchymal cells promotes initiation of colitis-associated cancer". In: *J. Exp. Med.* 212.13 (May 2015), pp. 2235–2251.

[209] Charles K Pallangyo, Paul K Ziegler, and Florian R Greten. "IKKbeta acts as a tumor suppressor in cancer-associated fibroblasts during intestinal tumorigenesis". In: *J. Exp. Med.* 212.13 (May 2015), pp. 2253–2266.

[210] Louis Vermeulen et al. “Wnt activity defines colon cancer stem cells and is regulated by the microenvironment”. In: *Nat. Cell Biol.* 12.5 (Apr. 2010), pp. 468–476.

[211] Vasiliki Koliaraki, Manolis Roulis, and George Kollias. “Tpl2 regulates intestinal myofibroblast HGF release to suppress colitis-associated tumorigenesis”. In: *J. Clin. Invest.* 122.11 (Oct. 2012), pp. 4231–4242.

[212] Kathryn E Hamilton et al. “Loss of Stromal IMP1 Promotes a Tumorigenic Microenvironment in the Colon”. In: *Mol. Cancer Res.* 13.11 (Nov. 2015), pp. 1478–1486.

[213] Neil A Bhowmick et al. “TGF-beta Signaling in Fibroblasts Modulates the Oncogenic Potential of Adjacent Epithelia”. In: *Science* 303.5659 (Feb. 2004), pp. 848–851.

[214] Elzbieta A Swietlicki et al. “Epimorphin deletion inhibits polyposis in the Apc min/+ mouse model of colon carcinogenesis via decreased myofibroblast HGF secretion”. In: *AJP: Gastrointestinal and Liver Physiology* 305.8 (Oct. 2013), G564–G572.

[215] Iftikhar Ali Khawar, Jung Ho Kim, and Hyo-Jeong Kuh. “Improving drug delivery to solid tumors: Priming the tumor microenvironment”. In: *Journal of Controlled Release* 201 (Mar. 2015), pp. 78–89.

[216] Paolo M Comoglio, Silvia Giordano, and Livio Trusolino. “Drug development of MET inhibitors: targeting oncogene addiction and expedience”. In: *Nat Rev Drug Discov* 7.6 (June 2008), pp. 504–516.

[217] Rajnish A Gupta and Raymond N Dubois. “Colorectal cancer prevention and treatment by inhibition of cyclooxygenase-2”. In: *Nat Rev Cancer* 1.1 (Oct. 2001), pp. 11–21.

[218] Ehud Shapiro, Tamir Biezuner, and Sten Linnarsson. “Single-cell sequencing-based technologies will revolutionize whole-organism science”. In: *Nature Reviews Genetics* 14.9 (2013), pp. 618–630.

[219] Alex K Shalek et al. “Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells”. In: *Nature* 498.7453 (2013), pp. 236–240.

[220] Antoine-Emmanuel Saliba et al. “Single-cell RNA-seq: advances and future challenges”. In: *Nucleic Acids Research* 42.14 (Aug. 2014), pp. 8845–8860.

[221] Nico Battich, Thomas Stoeger, and Lucas Pelkmans. “Control of Transcript Variability in Single Mammalian Cells”. In: *Cell* 163.7 (Dec. 2015), pp. 1596–1610.

[222] A Zeisel et al. “Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq”. In: *Science* 347.6226 (2015), pp. 1138–1142.

[223] D A Jaitin et al. “Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types”. In: *Science* 343.6172 (2014), pp. 776–779.

[224] Barbara Treutlein et al. “Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq”. In: *Nature* 509.7500 (2014), pp. 371–375.

[225] Dominic Grün et al. “Single-cell messenger RNA sequencing reveals rare intestinal cell types”. In: *Nature* 525.7568 (2015), pp. 251–255.

[226] Nir Yosef et al. “Dynamic regulatory network controlling TH17 cell differentiation”. In: *Nature* 496.7446 (2013), pp. 461–468.

[227] Valentina Proserpio et al. “Single-cell analysis of CD4 T-cell differentiation reveals three major cell states and progressive acceleration of proliferation”. In: *Genome Biology 2015 16:1* 17.1 (2016), p. 83.

[228] Åsa K Björklund et al. “The heterogeneity of human CD127 innate lymphoid cells revealed by single-cell RNA sequencing”. In: *Nature Immunology* 17.4 (2016), pp. 451–460.

[229] J Grossmann. “Progress on isolation and short-term ex-vivo culture of highly purified non-apoptotic human intestinal epithelial cells (IEC)”. In: *European Journal of Cell Biology* 82.5 (2003), pp. 262–270.

[230] Sarah Lamble et al. “Improved workflows for high throughput library preparation using the transposome-based nextera system”. In: *BMC Biotechnology* 13.1 (2013), p. 104.

[231] Hongshan Jiang et al. “Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads”. In: *BMC Bioinformatics* 15.1 (2014), p. 182.

[232] Daehwan Kim, Ben Langmead, and Steven L Salzberg. “HISAT: a fast spliced aligner with low memory requirements”. In: *Nature Methods* 12.4 (Apr. 2015), pp. 357–360.

[233] Y Liao, G K Smyth, and W Shi. “featureCounts: an efficient general purpose program for assigning sequence reads to genomic features”. In: *Bioinformatics* 30.7 (2013), pp. 923–930.

[234] Andrew Yates et al. “Ensembl 2016”. In: *Nucleic Acids Research* 44.D1 (2015), pp. D710–D716.

[235] Swati Parekh et al. “The impact of amplification on differential expression analyses by RNA-seq”. In: *Scientific Reports* 6.1 (May 2016), srep25533.

[236] Philip Ewels et al. “MultiQC: summarize analysis results for multiple tools and samples in a single report”. In: *Bioinformatics* 32.19 (Oct. 2016), pp. 3047–3048.

[237] Davis J McCarthy et al. “Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R.” In: *Bioinformatics* (Jan. 2017), btw777.

[238] Michael I Love, Wolfgang Huber, and Simon Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology 2015 16:1* 15.12 (Dec. 2014), p. 550.

[239] Peter V Kharchenko, Lev Silberstein, and David T Scadden. “Bayesian approach to single-cell differential expression analysis”. In: *Nature Methods* 11.7 (May 2014), pp. 740–742.

[240] Aaron T L Lun, Karsten Bach, and John C Marioni. “Pooling across cells to normalize single-cell RNA sequencing data with many zero counts”. In: *Genome Biology 2015 16:1* 17.1 (2016), p. 133.

[241] Nicolas Viphakone, Florence Voisinet-Hakil, and Lionel Minvielle-Sebastia. “Molecular dissection of mRNA poly(A) tail length control in yeast.” In: *Nucleic Acids Research* 36.7 (Apr. 2008), pp. 2418–2433.

[242] Davide Risso et al. “Normalization of RNA-seq data using factor analysis of control genes or samples”. In: *Nature Biotechnology* 32.9 (Sept. 2014), pp. 896–902.

[243] Peter Langfelder and Steve Horvath. “WGCNA: an R package for weighted correlation network analysis”. In: *BMC Bioinformatics* 9.1 (2008), p. 559.

[244] Peter Langfelder and Steve Horvath. “ Fast R Functions for Robust Correlations and Hierarchical Clustering ”. In: *Journal of Statistical Software* 46.11 (2012).

[245] Renaud Gaujoux and Cathal Seoighe. “A flexible R package for nonnegative matrix factorization”. In: *BMC Bioinformatics* 11.1 (2010), p. 367.

[246] J P Brunet et al. “Metagenes and molecular pattern discovery using matrix factorization”. In: *Proceedings of the National Academy of Sciences* 101.12 (2004), pp. 4164–4169.

[247] Lucie N Hutchins et al. “Position-dependent motif characterization using non-negative matrix factorization.” In: *Bioinformatics* 24.23 (Dec. 2008), pp. 2684–2690.

[248] Vladimir Yu Kiselev et al. “SC3: consensus clustering of single-cell RNA-seq data”. In: *Nature Methods* 14.5 (Mar. 2017), pp. 483–486.

[249] Guangchuang Yu et al. “clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters”. In: *OMICS: A Journal of Integrative Biology* 16.5 (2012), pp. 284–287.

[250] Peter Langfelder et al. “Is My Network Module Preserved and Reproducible?” In: *PLoS Computational Biology* 7.1 (2011), e1001057.

[251] L Van Der Maaten. “Accelerating t-SNE using tree-based algorithms”. In: *Journal of Machine Learning Research* 15 (Jan. 2015), pp. 3221–3245.

[252] L Van Der Maaten. “Visualizing data using t-SNE”. In: *Journal of Machine Learning Research* 9 (Nov. 2008), pp. 2579–2625.

[253] Alessandra Geremia et al. “IL-23-responsive innate lymphoid cells are increased in inflammatory bowel disease”. In: *Journal of Experimental Medicine* 208.6 (2011), pp. 1127–1133.

[254] Evan Z Macosko et al. “Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets”. In: *Cell* 161.5 (May 2015), pp. 1202–1214.

[255] Rahul Satija et al. “Spatial reconstruction of single-cell gene expression data”. In: *Nature Biotechnology* 33.5 (Apr. 2015), pp. 495–502.

[256] Antonio Scialdone et al. “Computational assignment of cell-cycle stage from single-cell transcriptome data”. In: *Methods* 85 (Sept. 2015), pp. 54–61.

[257] Philipp Angerer et al. “destiny: diffusion maps for large-scale single-cell data in R”. In: *Bioinformatics* 32.8 (Apr. 2016), pp. 1241–1243.

[258] Laleh Haghverdi et al. “Diffusion pseudotime robustly reconstructs lineage branching.” In: *Nature Methods* 13.10 (Oct. 2016), pp. 845–848.

[259] George M H Birchenough et al. “A sentinel goblet cell guards the colonic crypt by triggering Nlrp6-dependent Muc2 secretion”. In: *Science* 352.6293 (June 2016), pp. 1535–1542.

[260] Fuchou Tang et al. “mRNA-Seq whole-transcriptome analysis of a single cell”. In: *Nature Methods* 6.5 (May 2009), pp. 377–382.

[261] C L Galligan et al. “Distinctive gene expression signatures in rheumatoid arthritis synovial tissue fibroblast cells: correlates with disease activity”. In: *Genes Immun* 8.6 (June 2007), pp. 480–491.

[262] Magali de Bruyn et al. “Infliximab restores the dysfunctional matrix remodeling protein and growth factor gene expression in patients with inflammatory bowel disease.” In: *Inflamm Bowel Dis* 20.2 (Feb. 2014), pp. 339–352.

[263] J Lund et al. “ADAMDEC1 Is a Metzincin Metalloprotease with Dampened Proteolytic Activity”. In: *J. Biol. Chem.* 288.29 (July 2013), pp. 21367–21375.

[264] Nuala R O’Shea et al. “Critical Role of the Disintegrin Metalloprotease ADAM-like Decysin-1 (ADAMDEC1) for Intestinal Immunity and Inflammation”. In: *J Crohns Colitis* (May 2016), jjw111.

[265] T G Graeber and D Eisenberg. “Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles.” In: *Nat Genet* 29.3 (Nov. 2001), pp. 295–300.

[266] Eduard Batlle et al. “Beta-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB.” In: *Cell* 111.2 (Oct. 2002), pp. 251–263.

[267] Carme Cortina et al. “EphB-ephrin-B interactions suppress colorectal cancer progression by compartmentalizing tumor cells.” In: *Nat Genet* 39.11 (Nov. 2007), pp. 1376–1383.

[268] Vanessa Depaepe et al. “Ephrin signalling controls brain size by regulating apoptosis of neural progenitors”. In: *Nature* 435.7046 (May 2005), pp. 1244–1250.

[269] L Shi et al. “Ephrin-A1 promotes the malignant progression of intestinal tumors in Apcmin/+ mice”. In: *Oncogene* 27.23 (Feb. 2008), pp. 3265–3273.

[270] Jian Ma et al. “Targeting of erbB3 receptor to overcome resistance in cancer treatment”. In: *Molecular Cancer* 13.1 (2014), p. 105.

[271] Thierry Jardé et al. “Wnt and Neuregulin1/ErbB signalling extends 3D culture of hormone responsive mammary organoids”. In: *Nat Comms* 7 (Oct. 2016), p. 13207.

[272] Astrid De Boeck et al. “Bone marrow-derived mesenchymal stem cells promote colorectal cancer progression through paracrine neuregulin 1/HER3 signalling”. In: *Gut* 62.4 (Apr. 2013), pp. 550–560.

[273] Yongqin Zhang et al. “ErbB2 and ErbB3 regulate recovery from dextran sulfate sodium-induced colitis by promoting mouse colon epithelial cell survival”. In: *Lab Invest* 92.3 (Dec. 2011), pp. 437–450.

[274] Rosalba Salcedo et al. “MyD88-mediated signaling prevents development of adenocarcinomas of the colon: role of interleukin 18.” In: *J. Exp. Med.* 207.8 (Aug. 2010), pp. 1625–1636.

[275] Takanori Kanai et al. “Macrophage-derived IL-18–mediated intestinal inflammation in the murine model of Crohn’s disease”. In: *Gastroenterology* 121.4 (Oct. 2001), pp. 875–888.

[276] Roni Nowarski et al. “Epithelial IL-18 Equilibrium Controls Barrier Function in Colitis”. In: *Cell* 163.6 (Dec. 2015), pp. 1444–1456.

[277] Shabnam Shalapour et al. “Interleukin-7 Links T Lymphocyte and Intestinal Epithelial Cell Homeostasis”. In: *PLoS ONE* 7.2 (Feb. 2012), e31939.

[278] Shabnam Shalapour et al. “Commensal microflora and interferon-gamma promote steady-state interleukin-7 production in vivo.” In: *Eur. J. Immunol.* 40.9 (Sept. 2010), pp. 2391–2400.

[279] M K Kennedy et al. “Reversible defects in natural killer and memory CD8 T cell lineages in interleukin 15-deficient mice.” In: *Journal of Experimental Medicine* 191.5 (Mar. 2000), pp. 771–780.

[280] Rajia Bahri et al. “IL-15 suppresses colitis-associated colon carcinogenesis by inducing antitumor immunity.” In: *OncoImmunology* 4.9 (Sept. 2015), e1002721.

[281] L J Ma et al. “Trans-Presentation of IL-15 by Intestinal Epithelial Cells Drives Development of CD8 IELs”. In: *J Immunol* 183.2 (July 2009), pp. 1044–1054.

[282] Minoru Kanehisa et al. “KEGG: new perspectives on genomes, pathways, diseases and drugs.” In: *Nucleic Acids Research* 45.D1 (Jan. 2017), pp. D353–D361.

[283] Zhigang Xue et al. “Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing”. In: *Nature* 500.7464 (2013), pp. 593–597.

[284] Yuping Luo et al. “Single-Cell Transcriptome Analyses Reveal Signals to Activate Dormant Neural Stem Cells”. In: *Cell* 161.5 (May 2015), pp. 1175–1186.

[285] Elliot R Thomsen et al. “Fixed single-cell transcriptomic characterization of human radial glial diversity”. In: *Nature Methods* 13.1 (Jan. 2016), pp. 87–93.

[286] Jun Dong and Steve Horvath. “Understanding network concepts in modules.” In: *BMC Syst Biol* 1.1 (2007), p. 24.

[287] Alex A Pollen et al. “Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex”. In: *Nature Biotechnology* 32.10 (Aug. 2014), pp. 1053–1058.

[288] Anoop P Patel et al. “Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma.” In: *Science* 344.6190 (June 2014), pp. 1396–1401.

[289] A K Shalek et al. “Single-cell RNA-seq reveals dynamic paracrine control of cellular variation”. In: *Nature* 510 (2014), pp. 363–369.

[290] Chunxuan Shao and Thomas Höfer. “Robust classification of single-cell transcriptome data by nonnegative matrix factorization”. In: *Bioinformatics* 33.2 (2016), pp. 235–242.

[291] Piero Dalerba et al. “Single-cell dissection of transcriptional heterogeneity in human colon tumors”. In: *Nature Biotechnology* 29.12 (2011), pp. 1120–1127.

[292] Justina žurauskienė and Christopher Yau. “pcaReduce: hierarchical clustering of single cell transcriptional profiles”. In: *BMC Bioinformatics* 17.1 (Mar. 2016), p. 140.

[293] Florian Buettner et al. “Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells”. In: *Nature Biotechnology* 33.2 (Feb. 2015), pp. 155–160.

[294] Claudia Hofmann et al. “Cell-cell contacts prevent anoikis in primary human colonic epithelial cells.” In: *Gastroenterology* 132.2 (Feb. 2007), pp. 587–600.

[295] Po-Yuan Tung et al. “Batch effects and the effective design of single-cell gene expression studies.” In: *Scientific Reports* 7 (Jan. 2017), p. 39921.

[296] Moustafa Attar et al. “A practical solution for preserving single cells for RNA sequencing”. In: *bioRxiv* (2017). preprint, `https://doi.org/10.1101/160804`.

[297] Cécile Bénézech et al. “Ontogeny of Stromal Organizer Cells during Lymph Node Development”. In: *J Immunol* 184.8 (Apr. 2010), pp. 4521–4530.

[298] Alexander Link et al. “Fibroblastic reticular cells in lymph nodes regulate the homeostasis of naive T cells”. In: *Nature Immunology* 8.11 (Sept. 2007), pp. 1255–1265.

[299] Cyrille Mionnet et al. “Identification of a New Stromal Cell Type Involved in the Regulation of Inflamed B Cell Follicles”. In: *PLOS Biology* 11.10 (Oct. 2013), e1001672.

[300] Juan Dubrot et al. “Lymph node stromal cells acquire peptide–MHCII complexes from dendritic cells and induce antigen-specific CD4+ T cell tolerance”. In: *J. Exp. Med.* 211.6 (June 2014), pp. 1153–1166.

[301] Astrid Slany et al. “Plasticity of fibroblasts demonstrated by tissue-specific and function-related proteome profiling”. In: *Clinical Proteomics* 11.1 (Nov. 2014), p. 41.

[302] Keren Tazat et al. “TβRIII independently binds type I and type II TGF-β receptors to inhibit TGF-β signaling.” In: *Mol. Biol. Cell* 26.19 (Oct. 2015), pp. 3535–3545.

[303] Jason W Griffith, Caroline L Sokol, and Andrew D Luster. “Chemokines and Chemokine Receptors: Positioning Cells for Host Defense and Immunity”. In: *Annu. Rev. Immunol.* 32.1 (2014), pp. 659–702.

[304] Fabienne Mackay and Pascal Schneider. “Cracking the BAFF code”. In: *Nat. Rev. Immunol.* 9.7 (2009), pp. 491–502.

[305] Gertraud Orend. “Potential oncogenic action of tenascin-C in tumorigenesis”. In: *The International Journal of Biochemistry & Cell Biology* 37.5 (2005), pp. 1066–1083.

[306] Shideng Bao et al. “Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway”. In: *Cancer Cell* 5.4 (2004), pp. 329–339.

[307] Stayce E Beck et al. “Bone morphogenetic protein signaling and growth suppression in colon cancer”. In: *AJP: Gastrointestinal and Liver Physiology* 291.1 (July 2006), G135–G145.

[308] Ying Qu et al. “Over-expression of FRZB in gastric cancer cell suppresses proliferation and induces differentiation”. In: *J Cancer Res Clin Oncol* 134.3 (Aug. 2007), pp. 353–364.

[309] Hiroyuki Miyoshi et al. “Wnt5a potentiates TGF-β signaling to promote colonic crypt regeneration after tissue injury.” In: *Science* 338.6103 (Oct. 2012), pp. 108–113.

[310] Hayley Davis et al. “Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche”. In: *Nat. Med.* 21.1 (Dec. 2014), pp. 62–70.

[311] Mauricio Moreno et al. “Biglycan is a new extracellular component of the Chordin–BMP4 signaling pathway”. In: *The EMBO Journal* 24.7 (Apr. 2005), pp. 1397–1405.

[312] Liliana Schaefer et al. “The matrix component biglycan is proinflammatory and signals through Toll-like receptors 4 and 2 in macrophages”. In: *J. Clin. Invest.* 115.8 (Aug. 2005), pp. 2223–2233.

[313] Shunya Mimura et al. “Dipeptidyl peptidase-4 inhibitor anagliptin facilitates restoration of dextran sulfate sodium-induced colitis.” In: *Scand J Gastroenterol* 48.10 (Oct. 2013), pp. 1152–1159.

[314] Lawrence R Shiow et al. “CD69 acts downstream of interferon-alpha/beta to inhibit S1P1 and lymphocyte egress from lymphoid organs”. In: *Nature* 440.7083 (Mar. 2006), pp. 540–544.

[315] David Sancho et al. “CD69 downregulates autoimmune reactivity through active transforming growth factor-β production in collagen-induced arthritis”. In: *J. Clin. Invest.* 112.6 (Sept. 2003), pp. 872–882.

[316] T Sugihara et al. “The increased mucosal mRNA expressions of complement C3 and interleukin-17 in inflammatory bowel disease.” In: *Clin. Exp. Immunol.* 160.3 (June 2010), pp. 386–393.

[317] Luca Pastorelli et al. “Epithelial-derived IL-33 and its receptor ST2 are dysregulated in ulcerative colitis and in experimental Th1/Th2 driven enteritis”. In: *Proceedings of the National Academy of Sciences* 107.17 (2010), pp. 8017–8022.

[318] Andrew L Mellor and David H Munn. “Ido expression by dendritic cells: tolerance and tryptophan catabolism”. In: *Nat. Rev. Immunol.* 4.10 (2004), pp. 762–774.

[319] Silvia Barceló-Batllori et al. “Proteomic analysis of cytokine induced proteins in human intestinal epithelial cells: implications for inflammatory bowel diseases.” In: *Proteomics* 2.5 (May 2002), pp. 551–560.

[320] Jonathan J Hansen, Lisa Holt, and Balfour R Sartor. “Gene expression patterns in experimental colitis in IL-10-deficient mice”. In: *Inflamm Bowel Dis* 15.6 (2009), pp. 890–899.

[321] M A Ciorba et al. “Induction of IDO-1 by Immunostimulatory DNA Limits Severity of Experimental Colitis”. In: *The Journal of Immunology* 184.7 (2010), pp. 3907–3916.

[322] G Matteoli et al. “Gut CD103 dendritic cells express indoleamine 2,3-dioxygenase which influences T regulatory/T effector cell balance and oral tolerance induction”. In: *Gut* 59.5 (2010), pp. 595–604.

[323] A R B M Muzaki et al. “Intestinal CD103 CD11b- dendritic cells restrain colitis via IFN-gamma-induced anti-inflammatory response in epithelial cells”. In: *Mucosal Immunology* 9.2 (2015), pp. 336–351.

[324] Daren Low et al. “Chitinase 3-like 1 induces survival and proliferation of intestinal epithelial cells during chronic inflammation and colitis-associated cancer by regulating S100A9.” In: *Oncotarget* 6.34 (Nov. 2015), pp. 36535–36550.

[325] Steven H Itzkowitz and Xianyang Yio. "Inflammation and Cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation". In: *AJP: Gastrointestinal and Liver Physiology* 287.1 (July 2004), G7–G17.

[326] Audrey Lasry, Adar Zinger, and Yinon Ben-Neriah. "Inflammatory networks underlying colorectal cancer". In: *Nature Immunology* 17.3 (Mar. 2016), pp. 230–240.

[327] Sébastien Hannedouche et al. "Oxysterols direct immune cell migration via EBI2". In: *Nature* 475.7357 (July 2011), pp. 524–527.

[328] M S Phong et al. "p38 Mitogen-Activated Protein Kinase Promotes Cell Survival in Response to DNA Damage but Is Not Required for the G2 DNA Damage Checkpoint in Human Cancer Cells". In: *Mol. Cell. Biol.* 30.15 (2010), pp. 3816–3826.

[329] Zhen Qi et al. "BMP restricts stemness of intestinal Lgr5 stem cells by directly suppressing their signature genes". In: *Nat Comms* 8 (2017), p. 13824.

[330] Ferenc Sipos et al. "Regeneration associated growth factor receptor and epithelial marker expression in lymphoid aggregates of ulcerative colitis". In: *Scand J Gastroenterol* 45.4 (2010), pp. 440–448.

[331] Benoit Chassaing et al. "Dextran Sulfate Sodium (DSS)-Induced Colitis in Mice". In: *Current Protocols in Immunology* 40 (2014), pp. 15.25.1–15.25.14.

[332] Saiful Islam et al. "Quantitative single-cell RNA-seq with unique molecular identifiers". In: *Nature Methods* 11.2 (Feb. 2014), pp. 163–166.

[333] Grace X Y Zheng et al. "Massively parallel digital transcriptional profiling of single cells". In: *Nat Comms* 8 (Jan. 2017), p. 14049.

[334] Karthik Shekhar et al. "Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics". In: *Cell* 166.5 (Aug. 2016), 1308–1323.e30.

[335] Sonia Nestorowa et al. "A single cell resolution map of mouse haematopoietic stem and progenitor cell differentiation". In: *Blood* 128.8 (Jan. 2016), blood–2016–05–716480–e31.

[336] Suneale Banerji et al. "LYVE-1, a New Homologue of the CD44 Glycoprotein, Is a Lymph-specific Receptor for Hyaluronan". In: *The Journal of Cell Biology* 144.4 (Feb. 1999), pp. 789–801.

[337] Fernando Ochoa-Cortes et al. "Enteric Glial Cells: A New Frontier in Neurogastroenterology and Clinical Target for Inflammatory Bowel Diseases". In: *Inflamm Bowel Dis* 22.2 (Feb. 2016), pp. 433–449.

[338] Louise Hecker et al. "NADPH oxidase-4 mediates myofibroblast activation and fibrogenic responses to lung injury". In: *Nat. Med.* 15.9 (Aug. 2009), pp. 1077–1081.

[339] Veronika Lukacs-Kornek et al. "Regulated release of nitric oxide by nonhematopoietic stroma controls expansion of the activated T cell pool in lymph nodes". In: *Nature Immunology* 12.11 (Sept. 2011), pp. 1096–1104.

[340] Stefanie Siegert et al. "Fibroblastic Reticular Cells From Lymph Nodes Attenuate T Cell Expansion by Producing Nitric Oxide". In: *PLoS ONE* 6.11 (Nov. 2011), e27618.

[341] Omar Khan et al. “Regulation of T Cell Priming by Lymphoid Stroma”. In: *PLoS ONE* 6.11 (Nov. 2011), e26138.

[342] C J Pfeiffer et al. “Cellular pathology of experimental colitis induced by trinitrobenzenesulphonic acid (TNBS): Protective effects of recombinant human interleukin-11”. In: *Inflammopharmacol* 5.4 (Dec. 1997), pp. 363–381.

[343] Jason G Cyster et al. “A B-cell-homing chemokine made in lymphoid follicles activates Burkitt’s lymphoma receptor-1”. In: *Nature* 391.6669 (Feb. 1998), pp. 799–803.

[344] Takahiko Toyonaga et al. “Lipocalin 2 prevents intestinal inflammation by enhancing phagocytic bacterial clearance in macrophages”. In: *Scientific Reports* 6.1 (Oct. 2016), srep35014.

[345] Kenji Daigo and Takao Hamakubo. “Host-protective effect of circulating pentraxin 3 (PTX3) and complex formation with neutrophil extracellular traps”. In: *Front. Immunol.* 3 (2012), p. 378.

[346] David G T Hesslein et al. “Ebf1-dependent control of the osteoblast and adipocyte lineages”. In: *Bone* 44.4 (Apr. 2009), pp. 537–546.

[347] Alex K Shalek and Mikael Benson. “Single-cell analyses to tailor treatments”. In: *Sci Transl Med* 9.408 (Sept. 2017), eaan4730.

[348] Hyun Min Kang et al. “Multiplexing droplet-based single cell RNA-sequencing using natural genetic barcodes”. In: *bioRxiv* (2017). preprint, `https://doi.org/10.1101/118778`.

[349] Dominic Grün and Alexander van Oudenaarden. “Design and Analysis of Single-Cell Sequencing Experiments”. In: *Cell* 163.4 (2015), pp. 799–810.

[350] Tom Ronan, Zhijie Qi, and Kristen M Naegle. “Avoiding common pitfalls when clustering biological data”. In: *Science Signaling* 9.432 (June 2016), re6.

[351] Jeremy A Miller, Michael C Oldham, and Daniel H Geschwind. “A Systems Level Analysis of Transcriptional Changes in Alzheimer’s Disease and Normal Aging”. In: *J. Neurosci.* 28.6 (Feb. 2008), pp. 1410–1420.

[352] Tova F Fuller et al. “Weighted gene coexpression network analysis strategies applied to mouse weight”. In: *Mamm Genome* 18.6-7 (Aug. 2007), pp. 463–472.

[353] Masayuki Fukata et al. “Toll-like receptor-4 is required for intestinal response to epithelial injury and limiting bacterial translocation in a murine model of acute colitis”. In: *AJP: Gastrointestinal and Liver Physiology* 288.5 (May 2005), G1055–G1065.

[354] Koichi Kobayashi et al. “RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems”. In: *Nature* 416.6877 (2002), pp. 194–199.

[355] Jin Li et al. “The EMT transcription factor Zeb2 controls adult murine hematopoietic differentiation by regulating cytokine signaling.” In: *Blood* 129.4 (Jan. 2017), pp. 460–472.

[356] Darrell Pilling et al. “Fibroblasts secrete Slit2 to inhibit fibrocyte differentiation and fibrosis”. In: *Proceedings of the National Academy of Sciences* 111.51 (Dec. 2014), pp. 18291–18296.

[357] Conrad P Hodgkinson et al. “Abi3bp Is a Multifunctional Autocrine/Paracrine Factor that Regulates Mesenchymal Stem Cell Biology”. In: *Stem Cells* 31.8 (2013), pp. 1669–1682.

[358] S Mikami et al. “Blockade of CXCL12/CXCR4 Axis Ameliorates Murine Experimental Colitis”. In: *Journal of Pharmacology and Experimental Therapeutics* 327.2 (2008), pp. 383–392.

[359] Hiroyuki Miyoshi et al. “Prostaglandin E2 promotes intestinal repair through an adaptive cellular response of the epithelium”. In: *The EMBO Journal* 36.1 (2016), pp. 5–24.

[360] J Johnston et al. “Tribbles in inflammation”. In: *Biochemical Society Transactions* 43.5 (2015), pp. 1069–1074.

[361] Qing Zhou et al. “Loss-of-function mutations in TNFAIP3 leading to A20 haploinsufficiency cause an early-onset autoinflammatory disease”. In: *Nat Genet* 48.1 (2015), pp. 67–73.

[362] Victoria R Zellmer et al. “Tumor-induced Stromal STAT1 Accelerates Breast Cancer via Deregulating Tissue Homeostasis.” In: *Mol. Cancer Res.* (Jan. 2017), molcanres.0312.2016.

[363] Toshifumi Doi et al. “The JAK/STAT pathway is involved in the upregulation of PD-L1 expression in pancreatic cancer cell lines”. In: *Oncology Reports* 37.3 (2017), pp. 1545–1554.

[364] Reinhold Förster, Ana Clara Davalos-Misslitz, and Antal Rot. “CCR7 and its ligands: balancing immunity and tolerance”. In: *Nat. Rev. Immunol.* 8.5 (May 2008), pp. 362–371.

[365] Laurel A Monticelli et al. “IL-33 promotes an innate immune pathway of intestinal tissue protection dependent on amphiregulin–EGFR interactions”. In: *Proceedings of the National Academy of Sciences* 112.34 (2015), pp. 10762–10767.

[366] Petra Krause et al. “The Tumor Necrosis Factor Family Member TNFSF14 (LIGHT) Is Required for Resolution of Intestinal Inflammation in Mice”. In: *Gastroenterology* 146.7 (2014), 1752–1762.e4.

[367] Vincent van Unen et al. “Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets”. In: *Immunity* 44.5 (May 2016), pp. 1227–1239.

[368] Daekee Lee et al. “Epiregulin is not essential for development of intestinal tumors but is required for protection from intestinal damage.” In: *Mol. Cell. Biol.* 24.20 (Oct. 2004), pp. 8907–8916.

[369] Willem J S de Villiers et al. “Increased serum amyloid A levels reflect colitis severity and precede amyloid formation in Il-2 knockout mice”. In: *Cytokine* 12.9 (Sept. 2000), pp. 1337–1347.

[370] Christopher S Reigstad and Fredrik Bäckhed. “Microbial regulation of SAA3 expression in mouse colon and adipose tissue”. In: *Gut Microbes* 1.1 (Oct. 2014), pp. 55–57.

[371] Georgios Koukos et al. “A MicroRNA Signature in Pediatric Ulcerative Colitis”. In: *Inflamm Bowel Dis* 21.5 (May 2015), pp. 996–1005.

[372] C K F Li et al. “Impaired Immunity to Intestinal Bacterial Infection in Stromelysin-1 (Matrix Metalloproteinase-3)-Deficient Mice”. In: *The Journal of Immunology* 173.8 (2004), pp. 5171–5179.

[373] Ivan Monteleone et al. “Tissue Inhibitor of Metalloproteinase-3 Regulates Inflammation in Human and Mouse Intestine”. In: *Gastroenterology* 143.5 (Nov. 2012), 1277–1287.e4.

[374] K Yu. “Conditional inactivation of FGF receptor 2 reveals an essential role for FGF signaling in the regulation of osteoblast function and bone growth”. In: *Development* 130.13 (2003), pp. 3063–3074.

[375] Jarod A Zepp et al. “Distinct Mesenchymal Lineages and Niches Promote Epithelial Self-Renewal and Myofibrogenesis in the Lung”. In: *Cell* 170.6 (Sept. 2017), 1134–1148.e10.

[376] Joo-Hyeon Lee et al. “Anatomically and Functionally Distinct Lung Mesenchymal Populations Marked by Lgr5 and Lgr6”. In: *Cell* 170.6 (Sept. 2017), 1149–1163.e12.

[377] Brian G Feagan et al. “Vedolizumab as Induction and Maintenance Therapy for Ulcerative Colitis”. In: *New England Journal of Medicine* 369.8 (Aug. 2013), pp. 699–710.

[378] Brian G Feagan et al. “Ustekinumab as Induction and Maintenance Therapy for Crohn’s Disease”. In: *New England Journal of Medicine* 375.20 (Nov. 2016), pp. 1946–1960.

[379] Michael Kahn. “Can we safely target the WNT pathway?” In: *Nat Rev Drug Discov* 13.7 (July 2014), pp. 513–532.

[380] Benjamin Lu et al. “Wnt Drug Discovery: Weaving Through the Screens, Patents and Clinical Trials”. In: *Cancers 2016, Vol. 8, Page 82* 8.9 (Sept. 2016), p. 82.

[381] Simon P L Travis et al. “Developing an instrument to assess the endoscopic severity of ulcerative colitis: the Ulcerative Colitis Endoscopic Index of Severity (UCEIS)”. In: *Gut* 61.4 (Apr. 2012), pp. 535–542.