
Network biology approaches to study
the crosstalk between bifidobacteria,
immune cells and the intestinal
epithelium

Agatha Treveil

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia
Earlham Institute
Quadram Institute Bioscience

United Kingdom

September 2020

© This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived therefrom must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

The intestinal epithelial cell (IEC) barrier represents a key interface between host immune cells and commensal microbes. Communication between these compartments is crucial to maintenance of gut homeostasis by protecting against pathogens, maintaining a balance of commensal microbes and preventing overactivation of inflammation. A mechanistic understanding of how these compartments communicate with and respond to each other is crucial for developing preventative measures and treatments for complex gut dysbiosis, such as observed in inflammatory bowel disease (IBD). In this thesis I sought to study interactions between the health-promoting bacterial genus *Bifidobacterium*, the intestinal epithelium and the immune system to gain understanding about this complex system. To do this, I complemented experimental approaches with computational methods such as molecular interaction networks, to investigate inter- and intra-cellular molecular regulation at a systems level.

Using transcriptomics data from small intestinal organoid models enriched for specific epithelial cell types, I showed that Paneth cells and goblet cells exhibit shared and unique transcriptional and post-transcriptional regulation. Meanwhile, I highlighted a possible connection between IBD and IECs at the regulatory level. Further extending the study of IBD, I investigated the effect of IBD-relevant cytokines on IECs, shedding light on the causes of non-response to anti-cytokine treatments and presenting a potential new candidate for therapeutic targeting. Additionally, given the role of bifidobacteria in promoting gut health, myself and colleagues sought to define the impact of bifidobacteria on IECs of different aged mice and to study their interaction with macrophages at a molecular level.

Overall, this multidisciplinary work has increased mechanistic understanding of the interplay between IECs, immune cells and commensal microbes, while demonstrating the use of networks for such studies. This should ultimately lead to a better understanding of gut homeostasis and drive development of targeted approaches for prevention and treatment of gut dysbiosis related disorders.

Acknowledgements

I would like to thank all those who contributed to my PhD, scientifically, personally and financially (Figure 0.1). While too many cooks spoil the broth, successful completion of this PhD required many scientists and many independent supporters.

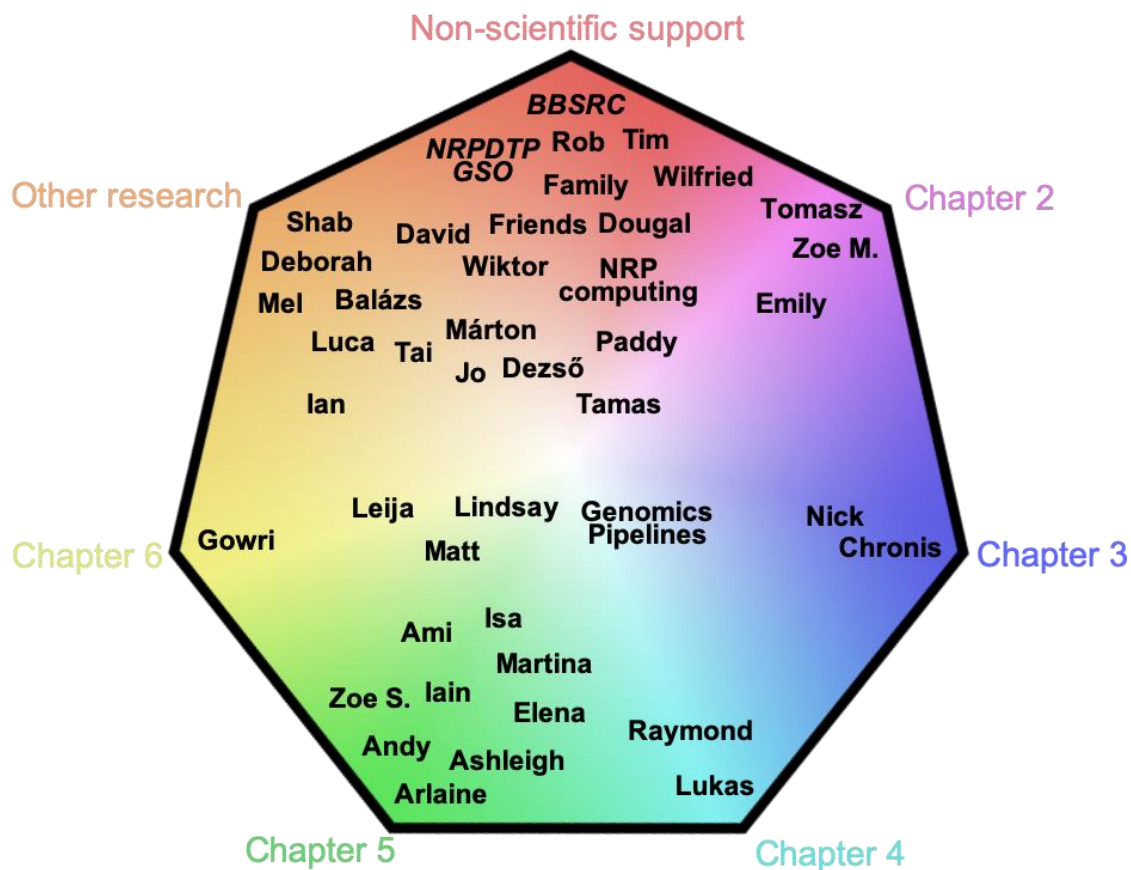


Figure 0.1. Thesis acknowledgements. All those people (and dog) whom helped me through my PhD journey – scientifically and non-scientifically. Friends – includes Erin, Charlie, Alicia, Becky, Beth. Family – Mum, Dad, Frances plus all the rest who shared their enthusiasm and encouraged me to keep going.

From a scientific perspective, I am thankful to all those who have contributed to the science within this thesis. In addition to the members of the Korcsmaros, Hall, Powell, Haerty and Macaulay groups mentioned in Figure 0.1, I would also like to thank all collaborators who have directly and indirectly contributed. Moreover, I thank those who influenced this thesis through advice and proof reading: Dezső Modos, Isabelle Hautefort, Leija Gul, Martina Poletti, Matthew Madgwick, John Thomas, Lindsay Hall, Sree Gowrinadh Javadi and Tamas Korcsmaros. Particularly, I extend my gratitude to Dezső for his thorough checks and great advice. Further

thanks to Raymond Kiu, whose thesis I used as a gold standard whilst creating mine. To Lindsay Hall, my secondary supervisor, thank you for being there when I needed you. Thank you for welcoming me into your group and for all the scientific advice and support.

From a personal perspective, I am grateful to all my colleagues at Oxford Gene Technology who introduced me to computational biology and inspired me to pursue a PhD. Even though my friends and family thought I was crazy (and I probably wouldn't have said this two years ago), I am very pleased I followed your advice. Thank you to my parents, whose own scientific endeavours inspired my interest in research, and whose enthusiasm, tenacity and ambition I strive to emulate. Your unwavering support is deeply appreciated. Of course, this PhD would not have been possible without the extensive emotional support provided by friends, including Charlie Davies, Alicia Russell, Becky Doherty and Erin Baggs for being there and for patiently listening to my rants (even from the other side of the world). Furthermore, I am appreciative of all members of the Korcsmaros and Hall groups, past and present, with whom I have shared lab meetings, journal clubs, lunches, coffee meetings and social events. To Rob, what can I say? Thank you for patiently listening, for encouraging me and for endlessly supporting all I do. Thank you for being there through the ups and the downs. To Dougal, you can't read, but thank you for the love and enthusiasm.

Otherwise, I am grateful to the Norwich Research Park Biosciences Doctoral Training Partnership (NRP DTP), to Wiktor Jurkowski, and to the Biotechnology and Biological Sciences Research Council (BBSRC) for giving me the opportunity and funding to undertake this PhD. Thank you to the Graduate Studies Office for all the administrative support and to the Earlham Institute and Quadram Institute Bioscience for welcoming me into your communities. Further, I am grateful to Tim Winters and Anna Crispe for organising and giving me a fantastic professional internship in the Norfolk County Council Knowledge and Intelligence team.

Finally, my deepest gratitude goes to Tamas Korcsmaros for his unfaltering enthusiasm, support and friendly advice. I could not have wished for a better supervisor. All those to come have a lot to live up to.

Table of contents

ABSTRACT.....	3
ACKNOWLEDGEMENTS.....	4
TABLE OF CONTENTS.....	6
LIST OF ABBREVIATIONS.....	11
LIST OF FIGURES.....	16
LIST OF TABLES.....	18
LIST OF ACCOMPANYING MATERIALS	19
LIST OF PEER REVIEWED PUBLICATIONS.....	20
CHAPTER 1: GENERAL INTRODUCTION	22
1.1 Introduction.....	22
1.2 Intestinal epithelium	24
1.2.1 Intestinal epithelial cells.....	25
1.2.2 Differentiation of intestinal epithelial cells.....	28
1.2.3 Epithelial barrier functions.....	29
1.2.4 Regulation of the epithelium	31
1.3 Immune cells of the gut	34
1.3.1 Intestinal T cells.....	35
1.3.2 Intestinal macrophages	36
1.4 Inflammatory bowel disease.....	37
1.4.1 The intestinal epithelium in IBD	38
1.4.2 Cytokines in IBD.....	38
1.4.3 IBD treatment strategy.....	42
1.4.4 IBD biologic treatments.....	43
1.5 Bifidobacteria.....	44
1.5.1 Health benefits of bifidobacteria	45
1.5.2 The effect of bifidobacteria on intestinal epithelial cells.....	46
1.5.3 The effect of bifidobacteria on gut immune cells	50
1.5.4 <i>Bifidobacterium breve</i> UCC2003	51
1.6 Organoids	52

1.7	Networks	55
1.7.1	Regulatory networks.....	57
1.7.2	Signalling networks.....	60
1.7.3	Host-microbe protein-protein interactions.....	60
1.7.4	Network contextualisation using ‘omics data	62
1.7.5	Network applications.....	63
1.8	Primary research aims	65
1.9	Structure of the thesis	65
CHAPTER 2: THE REGULATORY LANDSCAPE OF SMALL INTESTINAL EPITHELIAL CELLS		67
2.1	Introduction	67
2.2	Aims	71
2.3	Methods	72
2.3.1	Small intestinal organoid growth	72
2.3.2	RNA sequencing.....	73
2.3.3	Differentially expressed transcripts.....	73
2.3.4	Enrichment of marker genes	74
2.3.5	Reconstruction of regulatory networks.....	74
2.3.6	Paneth cell and goblet cell regulator prediction	76
2.3.7	Regulatory rewiring analysis.....	76
2.3.8	Evaluating disease relevance.....	77
2.3.9	Quantification and statistical analysis	77
2.3.10	Data and software availability	78
2.4	Results	79
2.4.1	Enteroids enriched for target cell type signatures	79
2.4.2	Regulatory networks are altered by enteroid differentiation skewing.....	83
2.4.3	Paneth cells and goblet cells have shared and unique regulators	86
2.4.4	Regulators are rewired between Paneth cells and goblet cells	90
2.4.5	Regulatory networks are relevant to study IBD	91
2.5	Discussion	97
2.6	Future research directions	102
CHAPTER 3: THE EFFECT OF CYTOKINES ON THE COLONIC EPITHELIUM		105
3.1	Introduction	105

3.2	Aims	108
3.3	Methods	109
3.3.1	Transcriptomics data from cytokine-treated organoids	109
3.3.2	Cytokine transcriptional signatures in IBD patient biopsies	110
3.3.3	Reconstructing cytokine causal networks in human colonoids	110
3.3.4	Functional analysis and visualisation	113
3.4	Results	114
3.4.1	Cytokines effect epithelial gene expression.....	114
3.4.2	Cytokine transcriptional signatures are enriched in IBD patient biopsies.....	115
3.4.3	Reconstructing cytokine causal networks in human colonoids	118
3.4.4	Cytokine-responsive signalling pathways converge at key transcription factors 122	
3.4.5	ETS1 is a major regulator of the cytokine signalling in intestinal inflammation	127
3.5	Discussion	128
3.6	Future research directions	132
CHAPTER 4: THE EFFECT OF BIFIDOBACTERIA ON THE SMALL INTESTINAL EPITHELIUM.....		134
4.1	Introduction	134
4.2	Aims	136
4.3	Methods	137
4.3.1	Mouse work.....	137
4.3.2	Bacterial culturing, inoculum preparation, mouse challenge with <i>B. breve</i> UCC2003 and CFU enumeration	137
4.3.3	Tissue harvesting and processing.....	138
4.3.4	RNA extraction, preparation and sequencing.....	138
4.3.5	Sequence pre-processing and differential expression analysis	138
4.3.6	Signalling network reconstruction and analysis.....	139
4.3.7	Cell type signature analysis	139
4.3.8	Functional analysis	140
4.4	Results	141
4.4.1	<i>B. breve</i> impacts the neonatal intestinal epithelial transcriptome.....	141
4.4.2	<i>B. breve</i> modulates neonatal cell maturation processes	142
4.4.3	Neonatal affected genes are enriched with epithelial stem cell markers	144
4.5	Discussion	146

4.6	Future research directions	149
CHAPTER 5: THE EFFECT OF BIFIDOBACTERIA ON SMALL INTESTINAL STEM CELLS AND PANETH CELLS..... 150		
5.1	Introduction	150
5.2	Aims	151
5.3	Methods	152
5.3.1	Mouse work	152
5.3.2	Bacterial culturing, inoculum preparation, mouse challenge with <i>B. breve</i> UCC2003 and CFU enumeration	152
5.3.3	Tissue harvesting and processing	153
5.3.4	Flow cytometry	153
5.3.5	RNA extraction, preparation and sequencing	155
5.3.6	Transcriptomics data processing	155
5.3.7	Computationally assessing cell identities	159
5.3.8	Differential expression analysis	160
5.3.9	Weighted gene co-expression network analysis	160
5.3.10	Functional analysis.....	161
5.4	Results	162
5.4.1	<i>B. breve</i> has a modest impact on juvenile intestinal epithelial transcriptome .	162
5.4.2	Intestinal stem cells express cell surface protein CD24a	164
5.4.3	Correlation analyses reveal multiple cellular functions marginally affected by <i>B. breve</i> in juvenile intestinal epithelial cells.....	166
5.5	Discussion	171
5.6	Future research directions	175
CHAPTER 6: INTERACTIONS BETWEEN BIFIDOBACTERIA AND MACROPHAGES..... 177		
6.1	Introduction	177
6.2	Aims	179
6.3	Methods	180
6.3.1	NF- κ B activation screen	180
6.3.2	Isolation of total proteome	181
6.3.3	Liquid chromatography–mass spectrometry	181
6.3.4	Host – microbe interaction predictions.....	182

6.3.5	Lipoprotein prediction.....	186
6.4	Results.....	187
6.4.1	<i>B. breve</i> UCC2003 activates NF-κB in THP-1 cells	187
6.4.2	Bifidobacterial protein domains can interact with macrophage protein motifs 187	
6.4.3	Predicted bifidobacterial proteins not present in mass spectrometry results .	189
6.4.4	NF-κB activating molecule likely to be a lipoprotein	191
6.5	Discussion	193
6.6	Future research directions	196
CHAPTER 7: INTEGRATED DISCUSSION		197
REFERENCES.....		202
APPENDIX 1: SUPPLEMENTARY DATA FOR CHAPTER 2.....		239
APPENDIX 2: SUPPLEMENTARY DATA FOR CHAPTER 3.....		246
APPENDIX 3: SUPPLEMENTARY DATA FOR CHAPTER 4.....		247
APPENDIX 4: SUPPLEMENTARY DATA FOR CHAPTER 5.....		254
APPENDIX 5: SUPPLEMENTARY DATA FOR CHAPTER 6.....		260
APPENDIX 6: PEER-REVIEWED PUBLICATIONS		261

List of abbreviations

7AAD	7-Aminoactinomycin D
AGC	Automatic gain control
AHR	Aryl hydrocarbon receptor
AMP	Antimicrobial peptide
Ang4	Angiogenin 4
APC	APC microbiome Ireland
Apoa-1	Apolipoprotein A-I
ARID2	AT-rich interaction domain 2
Ascl2	Achaete-scute family BHLH transcription factor 2
ASPA	Animals (Scientific Procedures) Act 1986
ATG16L1	Autophagy related protein 16-1
Atoh1	Atonal homolog 1
<i>B. breve</i>	<i>Bifidobacterium breve</i>
<i>B. longum</i>	<i>Bifidobacterium longum</i>
<i>B. bifidum</i>	<i>Bifidobacterium bifidum</i>
<i>B. dentium</i>	<i>Bifidobacterium dentium</i>
<i>B. infantis</i>	<i>Bifidobacterium infantis</i>
<i>B. adolescentis</i>	<i>Bifidobacterium adolescentis</i>
BHI	Brain Heart Infusion
BIK	BCL2 interacting killer
BMP	Bone morphogenetic protein
Cbfa2t2	CBFA2/RUNX1 partner transcriptional co-repressor 2
cCD	Colonic Crohn's disease
CD	Crohn's disease
CD14	Cluster of differentiation 14
Cd244a	CD244 molecule A
CD24a	Cluster of differentiation 24, Signal transducer 24a
CEBPA	CCAAT enhancer-binding protein alpha
ChgA	Chromogranin A
Clps	Colipase
CLR	C-type lectin receptor
Copz2	COPI coat complex subunit zeta 2
CRP	C-reactive protein
CTFR	Cystic fibrosis transmembrane conductance regulator
Cytb	Cytochrome b
Da	Daltons
DAPI	4',6-diamidino-2-phenylindole
Dbp	D-box binding PAR BZIP transcription factor
DCLK1	Doublecortin-like kinase 1
Defa2	Defensin alpha 2
DEG	Differentially expressed gene

Dll4	Delta-like ligand
DPBS	Dulbecco's phosphate-buffered saline
DSS	Dextran sulphate sodium
E2F1	E2F transcription factor 1
E2F2	E2F Transcription Factor 2
EDTA	Ethylenediaminetetraacetic acid
EEC	Enteroendocrine cell
EGF	Epidermal growth factor
EGFP	Enhanced green fluorescence protein
EI	Earlham Institute
ELM	Eukaryotic Linear Motif (database)
ENA	European Nucleotide Archive
EpCAM	Epithelial cell adhesion molecule
EPS	Exopolysaccharide
eQTL	Expression quantitative trait loci
ER	Endoplasmic reticulum
ESR	Erythrocyte sedimentation rate
ETS1	Protein C-ets-1
ETV4	ETS variant transcription factor 4
Fabp6	Gastrotropin
FACS	Fluorescence activated cell sorting
Fgfr1	Fibroblast growth factor receptor like
FITC	Fluorescein isothiocyanate
FOS	FOS proto-oncogene, AP-1 transcription factor subunit
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GABA	γ -aminobutyric acid
GalC	Solute-binding protein of ABC transporter system for sugars
GCeE	Goblet cell enriched enteroid
GF	Germ free
Gfi1	Growth factor independent 1
Gfi1b	Growth factor independent 1B transcriptional repressor
GlgP1	Alpha-1,4 glucan phosphorylase
GlnA	Glutamine synthetase
GO:BP	Gene Ontology: Biological Process
GPCR	G-protein coupled receptor
GSEA	Gene set enrichment analysis
GSK3 β	Glycogen synthase kinase three beta
Habp2	Hyaluronan binding protein 2
HATH1	Atonal homolog 1 (ATOH1)
HD	Human α -defensin
Hes1	Hairy and enhancer of split-1
HMP	Human neutrophil peptides
Hnf4a	Hepatocyte nuclear factor 4 alpha

HPLC	High-performance liquid chromatography
Hspb8	Heat shock protein family B member 8
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
IEC	Intestinal epithelial cell
IFN γ	Interferon-gamma
IL	Interleukin
ILC	Innate lymphoid cell
IPA	Ingenuity Pathway Analysis
iPSCs	Induced pluripotent stem cells
IRF1	Interferon regulatory factor 1
IspA	Lipoprotein signal peptidase protein
JAK inhibitor	Janus kinase inhibitor
JUN	Jun proto-oncogene AP-1 transcription factor subunit
KCL	Kings College London
KCNH2	Potassium voltage-gated channel subfamily H member 2
kDa	Kilo Daltons
Klf4	Krüppel-like factor 4
KMT2A	Lysine methyltransferase 2A
LC-MS	Liquid chromatography-mass spectrometry
LC3	Microtubule-associated proteins 1A/1B light chain 3B, MAP1LC3B
LFC	Log ₂ fold change
LGR5	Leucine-rich repeat-containing G-protein coupled receptor 5
Lgt	Preprolipoprotein diacylglycerol transferase
lncRNA	Long non-coding RNA
LPS	Lipopolysaccharide
LYZ1	Lysozyme
M cells	Microfold/membranous cells
MAFK	MAF BZIP transcription factor K
MAMP	Microbe-associated molecular pattern
MAP1LC3B	Microtubule-associated proteins 1A/1B light chain 3B, LC3
MAPK(14)	Mitogen-activated protein kinase (14)
Math1	Atonal homolog 1, aka Atoh1
MAZ	MYC associated zinc finger protein
MCODE	Molecular complex detection (tool)
miRNA	Micro RNA
MLCK	Myosin light chain kinase
mRNA	Messenger RNA
MRS	De Man, Rogosa and Sharpe
Muc2	Mucin 2
MYB	MYB proto-oncogene
MYBL2	MYB proto-oncogene like 2
MYC	MYC proto-oncogene

MyD88	Myeloid differentiation primary response 88
Mymx	Myomixer myoblast fusion factor
Nd1	NADH dehydrogenase 1
Neurog3	Neurogenin 3
NF- κ B	Nuclear factor kappa-light-chain-enhancer of activated B cells
NK cell	Natural killer cells
NLR	Nucleotide oligomerization domain-like receptor
NLS-SS	Natren Life Start super strain
NOTCH2,3	Notch receptor 2,3
NR1D1	Nuclear receptor subfamily 1 group D member 1
NR3C1	Glucocorticoid receptor (Nuclear receptor subfamily 3 group C member 1)
Nr5a2	Nuclear receptor subfamily 5 group A member 2
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCeE	Paneth cell enriched enteroid
PE	Phycoerythrin
PFA	Paraformaldehyde
PknA1	Serine/threonine protein kinase
Pla2g2a	Phospholipase A2 group IIA
PLAU	Plasminogen activator urokinase
PLXND1	Plexin D1
Pou2f3	POU class 2 homeobox 3
PPI	Protein-protein interaction
PRR	Pattern recognition receptor
QIB	Quadram Institute Bioscience
RCA	Reinforced Clostridial Agar
RCM	Reinforced Clostridial Medium
RegIII α , β , γ	Regenerating islet-derived protein 3 alpha, beta, gamma
Relm- β	Resistin-like molecule
RfbF	dTDP-rhamnosyl transferase
RIN	RNA Integrity Number
RLR	Retinoic acid inducible gene I-like receptor
RORA	RAR related orphan receptor A
RspA	SSU ribosomal protein S1P
RXRA	Retinoid X receptor alpha
scRNA-seq	Single cell RNA sequencing
SerA1	D-3-phosphoglycerate dehydrogenase
SerA2	D-3-phosphoglycerate dehydrogenase 2
Serpin	Serine protease inhibitor
SLiM	Short linear motif
SNP	Single nucleotide polymorphism
Sntb1	Syntrophin beta 1

SOCS	Suppressor of cytokine signalling
Sox9	SRY-box transcription factor 9
SPDEF	SAM pointed domain containing ETS transcription factor
SPF	Specific pathogen free / conventionalized
SpiB	Spi-B transcription factor
sPLA2	Secretory phospholipase A2
STAT3	Signal transducer and activator of transcription 3
TA(high/low)	Transit amplifying high/low (cells)
TCA	tricarboxylic acid
Tcf4	Transcription factor 4
TES	Total enrichment score
TF	Transcription factor
TF-TG	Transcription factor – target gene interactions
Tff3	Trefoil factor 3
TG	Target gene
TGF- β	Transforming growth factor- β
Th cell	T helper cell
Tig	Trigger factor
TLR	Toll-like receptor
TNBS	2,4,6-trinitrobenzene sulfonic acid
TNF α	Tumour necrosis factor-alpha
Tnfsf15	TNF superfamily member 15
TPM	Transcripts per million
UC	Ulcerative colitis
UEA	University of East Anglia
UMAP	Uniform Manifold Approximation and Projection
VDR	Vitamin D receptor
Vil1	Villin 1
WGCNA	Weighted-gene co-expression network analysis
Wnt3	Wnt family member 3
WT	Wild-type
Zg16	Zymogen granule protein 16

List of figures

Figure 1.1. Project summary: crosstalk between the intestinal epithelium, immune cells and bifidobacteria.....	23
Figure 1.2. Cross section of the small intestinal and colonic epithelium (human and mouse)..	24
Figure 1.3. Intestinal epithelial cell differentiation.....	28
Figure 1.4. Regulation of intestinal epithelial cells (IECs) by immune cells.	34
Figure 1.5. Pro- and anti-inflammatory cytokines are exposed to intestinal epithelial cells during chronic intestinal inflammation in inflammatory bowel disease.	39
Figure 1.6. Relative abundance of gut bifidobacteria during the human life cycle.	45
Figure 1.7. Summary of the different ways in which Bifidobacterial strains have been shown to affect the intestinal epithelial layer.....	47
Figure 1.8. Small intestinal organoids.	53
Figure 1.9. Graph diagram of a directed network.....	56
Figure 1.10. Transcriptional and post transcriptional regulation.	58
Figure 2.1. Schematic overview of study design and analysis workflow in Chapter 2.	70
Figure 2.2. Principle component analysis of transcript expression of enteroids.....	78
Figure 2.3. Small intestinal 3D organoid culture.	79
Figure 2.4. Differentially expressed genes in Paneth cell enriched enteroids (PCeEs) and goblet cell enriched enteroids (GCeEs) (compared to conventionally differentiated enteroids).....	81
Figure 2.5. Transcript abundances and differential expression of five major cell type markers.	82
Figure 2.6. Summary and cluster analysis of regulatory network for Paneth cell enriched enteroid (PCeE) and goblet cell enriched enteroid (GCeE) datasets.	85
Figure 2.7. Regulator-marker subnetworks for Paneth cell and goblet cell datasets.	87
Figure 2.8. Matrices of interactions between markers and their regulators in the Paneth cell and goblet cell subnetworks.....	89
Figure 2.9. Crohn's disease and ulcerative colitis associated SNP genes and their targets within the PCeE and GCeE networks.	95
Figure 3.1. Schematic overview of the primary data sources and analyses carried out in Chapter 3.	107
Figure 3.2. Schematic of node and interaction types included in the cytokine causal networks.	111

Figure 3.3. Differentially expressed genes upon cytokine treatment of colonoids.	114
Figure 3.4. Overlap between cytokine programme categories and cCD and UC biopsy DEGs.	116
Figure 3.5. Gradient of top cytokine programme activation in IBD.	118
Figure 3.6. Reconstructing causal networks of signalling and regulatory molecular interactions connecting cytokines to observed differentially expressed genes within the colonic organoids.....	119
Figure 3.7. Chord diagram visualisation of differentially expressed genes (DEGs) targeted by each cytokine in the causal networks, with functional associations.	121
Figure 3.8. Overlap of nodes and edges in the four cytokine causal networks.	123
Figure 3.9. Transcription factors (TFs) shared between multiple cytokine causal networks....	125
Figure 3.10. Signalling pathways linking cytokines to shared transcription factors.	126
Figure 4.1. Schematic overview of neonatal bulk epithelium study design and analysis workflow.	135
Figure 4.2. Signalling network analysis, IEC subtyping and key regulator analysis.....	143
Figure 5.1. Schematic overview of juvenile cell type-specific study design and analysis workflow.	151
Figure 5.2. Read and alignment quality control plots.	156
Figure 5.3. Normalised gene counts quality control plots.	157
Figure 5.4. Gene abundance density plots for determining expression cut offs.	159
Figure 5.5. UMAP plots of normalised counts data.....	163
Figure 5.6. Gene expression heatmap of top 50 variant genes among all specific pathogen free samples.	164
Figure 5.7. Differentially expressed genes for each cell population compared to intestinal epithelial cell marker gene sets.	165
Figure 5.8. WGCNA results of germ free/monocolonised mice Paneth cell expression data...	168
Figure 5.9. Expression of key transcription factors in intestinal epithelial cell differentiation across specific pathogen free mice samples.....	170
Figure 6.1. Schematic overview of the experiments and analyses carried out in Chapter 6.	179
Figure 6.2. Summary of bifidobacteria-macrophage interaction prediction pipeline.	183
Figure 6.3. Density plot of gene abundance across samples in project PRJNA163281.	184
Figure 6.4. Density plot of average gene abundance across macrophage cells.....	185
Figure 6.5. Summary of results from computational and experimental approaches.	191

List of tables

Table 2.1. A summary of the physical interactions compiled to generate the universal network.	75
Table 2.2. Literature associations relating to autophagy, inflammation and IBD for putative master regulators.	93
Table 3.1. All possible cytokine – receptor interactions for the causal networks, based on literature curation.....	111
Table 3.2. Number of differentially expressed genes upon cytokine treatment of colonoids and in the resulting causal networks.	120
Table 5.1. Antibody panel for sorting Paneth, stem and transit amplifying cells.	154
Table 5.2. Summary of overrepresented Gene Ontology biological processes.....	167
Table 6.1. Proteins of <i>B. breve</i> UCC2003 predicted to interact with macrophage surface proteins vis domain-motif interactions.....	188
Table 6.2. Potential NF- κ B activating proteins based on LC-MS.....	189

List of accompanying materials

Chapter 2:

- **File S2.1** - Differentially expressed genes from cell type enriched enteroids vs conventionally differentiated enteroids (electronic supplementary materials).
- **File S2.2** - Functional enrichment analysis of the top five most rewired (shared) marker regulators (electronic supplementary materials).

Chapter 5:

- **File S5.1** - Results of gene set enrichment analysis comparing bifidobacteria-exposed to control mice in each mouse type and cell type using Reactome and Gene Ontology Biological Process annotations (electronic supplementary materials).

List of peer reviewed publications

Peer-reviewed journal articles published during my PhD (2017-2020). Those covered in Chapters 2 and 4 are reproduced in Appendix 6.

Chapter 2:

- Treveil, A., Sudhakar, P., Matthews, Z.J., Wrzesiński, T., Jones, E.J., Brooks, J., Ölbei, M., Hautefort, I., Hall, L.J., Carding, S.R., et al. (2020). Regulatory network analysis of Paneth cell and goblet cell enriched gut organoids using transcriptomics approaches. *Mol. Omics*. *16(1)*, 39-58.

Chapter 3:

- Pavlidis, P., Tsakmaki, A., Treveil, A., Li, K., Cozzetto, D., Yang, E., Niazi, U., Hayee, BH., Saqi, M., Friedman, J., Korcsmaros, T., Bewick, G., Powell, N. (2020). Cytokine responsive networks in human colonic epithelial organoids unveils a novel molecular classification of inflammatory bowel disease. **Under Review.**

Chapter 4:

- Kiu, R., Treveil, A., Harnisch, L.C., Caim, S., Leclaire, C., van Sinderen, D., Korcsmaros, T., and Hall, L.J. (2020). *Bifidobacterium breve* UCC2003 Induces a Distinct Global Transcriptomic Program in Neonatal Murine Intestinal Epithelial Cells. *IScience* *23*, 101336.

Others (not presented in thesis):

- Treveil, A., Bohar, B., Sudhakar, P., Gul, L., Csabai, L., Olbei, M., Poletti, M., Madgwick, M., Andrighetti, T., Hautefort, I., Modos, D. and Korcsmaros, T. (2020). ViralLink: An integrated workflow to investigate the effect of SARS-CoV-2 on intracellular signalling and regulatory pathways. *BioRxiv*, <https://www.biorxiv.org/content/10.1101/2020.06.23.167254v2>, **Under Review.**
- Püngel, D., Treveil, A., Dalby, M.J., Caim, S., Colquhoun, I.J., Booth, C., Ketskemety, J., Korcsmaros, T., van Sinderen, D., Lawson, M.A. and Hall, L.J. (2020). *Bifidobacterium breve* UCC2003 Exopolysaccharide Modulates the Early Life Microbiota by Acting as a Potential Dietary Substrate. *Nutrients* *12(4)*, 948.

- Jones, E.J., Matthews, Z.J., Gul, L., Sudhakar, P., Treveil, A., Divekar, D., Buck, J., Wrzesinski, T., Jefferson, M., Armstrong, S.D., Hall, L.J., Watson, A.J.M., Carding, S.R., Haerty, W., Di Palma, F., Mayer, U., Powell, P.P., Hautefort, I., Wileman, T. and Korcsmaros, T. (2019). Integrative analysis of Paneth cell proteomic and transcriptomic data from intestinal organoids reveals functional processes dependent on autophagy. *Disease Models & Mechanisms* 12(3).

Chapter 1: General introduction

1.1 Introduction

Specialised epithelial cells lining the surface of the mammalian gastrointestinal tract form the primary interface between the body's internal tissues and the luminal content. Importantly, these cells work in cohort with the local immune cells and the gut microbiome to maintain homeostasis within the gut. Indeed, effective communication between these compartments is critical to health. In this thesis I complimented experimental methods with network biology methods to study the interactions between intestinal epithelial cells (IECs), commensal bacteria *Bifidobacterium* and immune cells (macrophages) and mediators (cytokines), to gain mechanistic understanding of interplay between them.

The intestinal epithelium is a single layer of epithelial cells consisting of multiple distinct cell types. Balance within the gut is maintained, in part, by dynamic functioning of IECs in response to diverse signals from the luminal content and the immune system. Disruption to this balance can result in increased susceptibility to microbial infections and is implicated in a number of autoimmune and inflammatory conditions, such as inflammatory bowel disease (IBD). IBD is defined as chronic gut inflammation due to an inappropriate immune response to the gut microbiome in genetically susceptible hosts (Fakhoury et al., 2014).

The gut microbiota is a complex composition of pathogenic and commensal bacteria, fungi and viruses which have co-evolved with the innate and adaptive immune systems of their host. Recognition of commensal gut microbes by IECs can play a role in balancing immune activation against tolerance, by altering mucus and antimicrobial peptide production, stem cell proliferation and increasing epithelial integrity (Kim et al., 2010). Meanwhile, immune cells such as phagocytes can detect exogenous bacteria and signal their activity to other immune cells and IECs, activating a response to defend against infection. Furthermore, commensal bacteria can interact directly with immune cells to dampen pro-inflammatory immune responses - for example the interaction between *Bifidobacterium longum* 35624 and dendritic cells results in repression of local T helper cell (Th)17 responses (Schiavi et al., 2016). Such interactions can occur in a healthy gut through recognition of luminal bacterial by intraepithelial lymphocytes,

lumen-sampling dendritic cells or through Payer's patches (see General Introduction section 1.3). Moreover, bacterial-immune cell interactions occur in large numbers when epithelial barrier functions are disrupted, for example in IBD, whereby lumen content is able to cross the epithelium into the lamina propria below. This creates a complex network of communication between IECs, immune populations and the gut microbiome which serves to maintain gut homeostasis and health (Figure 1.1).

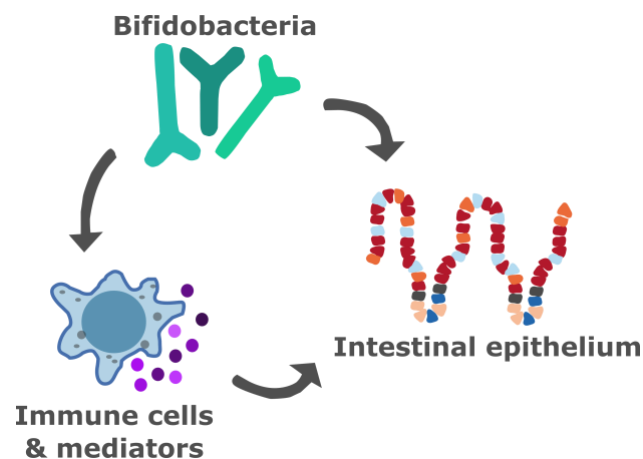


Figure 1.1. Project summary: crosstalk between the intestinal epithelium, immune cells and bifidobacteria. Each arrow represents a results chapter in this thesis.

Lack of fundamental understanding of the interplay between IECs, the immune system and gut microbes is a major barrier to progress in understanding and curing human diseases. To progress towards this comprehension of complex gut dysbiosis, it is important to study these interactions. Due to increasing appreciation of the role of beneficial commensal bacteria in immune system and IEC cross-talks, I have focused on the health-promoting genus *Bifidobacterium*. Accordingly, there are five overarching aims to this thesis:

1. Develop workflows and processes to analyse intracellular regulation in a cell type-specific manner to gain biological insights.
2. Apply these workflows to increase our understanding of how cytokines alter the regulation of epithelial cells.
3. Apply these workflows to increase our understanding of how *Bifidobacterium* alters the regulation of epithelial cells using bulk transcriptomics data.
4. Apply these workflows to increase our understanding of how *Bifidobacterium* alters the regulation of epithelial cells using cell type-specific transcriptomics data.
5. Study the interactions of *Bifidobacterium* with immune cell populations.

In this introductory chapter I explore background theory and literature relating to the structure and function of the epithelial lining and the impact of immune and microbiome signalling on regulation of IECs. I will introduce IBD (a clinical implication of epithelial and immune disruption in the gut) and discuss the gut commensal *Bifidobacterium*. Furthermore, I will cover *ex vivo* and *in silico* methods to study cellular function and regulation which have been used in this thesis.

1.2 Intestinal epithelium

The intestinal wall consists of four primary layers: the mucosa which contains epithelial cells (IECs), the lamina propria and smooth muscle (adjacent to the gut lumen), the submucosa which contains loose connective tissue and blood vessels, the muscularis which contains smooth muscle and neurons and the enveloping connective tissue of the serosa. As IECs form the interface between host tissues and the luminal environment, they are essential for absorption of water and nutrients and to provide a physical and biochemical barrier to protect the human body from foreign particles and microbial infections (Okumura and Takeda, 2017; Peterson and Artis, 2014). To increase the surface area for absorption, the intestinal epithelium is arranged in fold-like invaginations (Figure 1.2). In the small intestine, finger-like protrusions termed villi are surrounded by invaginations called the crypts of Lieberkühn, which were discovered and published by Jonathan Nathanael Lieberkühn in 1745 (Clevers, 2013; Lieberkühn, 1744). The colonic epithelium does not contain villous projections but has tubular pits termed crypts, which increase in depth towards the rectum. To carry out their diverse functions, IECs recognise and respond to a variety of signals, including those from immunological mediators and gut microbes.

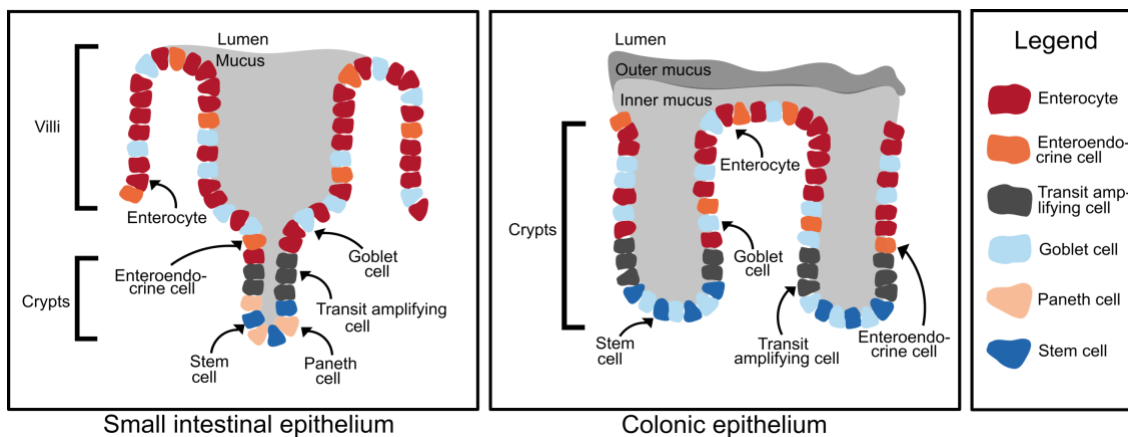


Figure 1.2. Cross section of the small intestinal and colonic epithelium (human and mouse). Major epithelial cell types shown.

Proliferative stem cells residing at the base of the intestinal crypts gradually differentiate as they migrate upwards away from the crypts. During the transit amplifying stage, cells undergo 4-5 rapid divisions whilst differentiating into their final cell types (Marshman et al., 2002). At the end of their life, mature differentiated IECs undergo apoptosis and are shed into the lumen. Through this process the intestinal epithelium undergoes a constant cycle of regeneration, with cell renewal on average every four to seven days (Clevers, 2013; van der Flier and Clevers, 2009; Zachos et al., 2016).

1.2.1 Intestinal epithelial cells

In addition to stem cells, there are six major recognised differentiated cell types of the intestinal epithelium; enterocytes, goblet cells, enteroendocrine cells, Paneth cells, microfold/membranous (M) cells and tuft cells (van der Flier and Clevers, 2009). All of the above cell types can be found in the small intestines and the colon, apart from Paneth cells which are typically only observed in the small intestine. Despite this similarity, functional and structural differences are observed between cell populations from different segments of the intestinal tract.

1.2.1.1 Stem cells

The primary role of stem cells in the gut is to facilitate the regeneration of the epithelial cells. Two distinct models exist to describe multipotent stem cells in the intestinal epithelium. Whilst there is consensus that the crypt contains four to six stem cells, the exact identity of the stem cells has been debated (Barker et al., 2008). The “+4 position” model describes stem cells in the +4 position of the crypt with Paneth cells occupying the first three positions at the base of the crypt (Potten et al., 1974). The “stem cell zone” model describes small undifferentiated cycling cells called crypt base columnar cells at the base of the crypts interspersed with Paneth cells (Cheng and Leblond, 1974). These fast cycling cells are marked by their high expression of Leucine-rich repeat-containing G-protein coupled receptor 5 (LGR5), as also seen in stomach and hair follicle stem cells (Barker et al., 2007, 2010; Jaks et al., 2008; Muñoz et al., 2012). Despite reports of molecular markers for the quiescent +4 stem cell population (Montgomery et al., 2011; Powell et al., 2012; Sangiorgi and Capecchi, 2008; Takeda et al., 2011), subsequent studies have not observed cell type-specific expression of these markers (van der Flier et al., 2009; Wong et al., 2012). However, evidence still exists for this population of cells, which have been shown to be insensitive to injury and to increase their stem cell activity upon damage to potentially

replenish LGR5+ cells (Montgomery et al., 2011; Tian et al., 2011). Clearly, further research is required to reconcile these two postulated populations of stem cells.

Throughout this thesis, any unspecified reference to stem cells of the intestinal epithelium will refer to the LGR5+ population. This population is currently deemed the primary population and is key to the generation of small intestinal and colonic organoids.

1.2.1.2 Enterocytes and colonocytes

The most populous cells of the intestinal epithelium are enterocytes of the small intestine and colonocytes of the colon (Ohno, 2016). With a lifespan of five to seven days (Zachos et al., 2016), their primary roles are to apically absorb nutrients for basal export and to maintain water/electrolyte homeostasis. These cells have a characteristic microvilli brush border to increase their surface area.

1.2.1.3 Goblet cells

Belonging to the secretory lineage of IECs, goblet cells make up 5-15% of the small intestinal epithelium and up to 50% of the colonic epithelium (Kim and Ho, 2010; Noah et al., 2011). Intestinal goblet cells play a key role in barrier protection through the secretion of mucus, anti-microbial proteins, chemokines and cytokines. However, mucin glycoprotein mucin (MUC) 2 is their primary secretion (Knoop and Newberry, 2018). These proteins have heavily glycosylated central tandem repeat domains flanked by the C-terminal cysteine knot domain and domains of von Willebrand factor. MUC2 is termed a gel-forming mucin because dimerisation and oligomerisation of these domains results in the viscoelastic properties of mucus (Godl et al., 2002; Kim and Ho, 2010). The mucus layer is further described in section 1.2.3.2.

1.2.1.4 Paneth cells

First identified by Austrian physician Joseph Paneth in the 1880s (Paneth, 1887), Paneth cells are secretory cells of the small intestine that reside amongst the stem cells at the base of crypts. Unlike other cells of the gut epithelium, Paneth cells are very long lived (>30 days) and do not migrate up the villi (Bjerknes and Cheng, 2006; Zachos et al., 2016). Paneth cells act as a major effector of the mucosal immune system through the release of a diverse repertoire of products. These products primarily consist of antimicrobial peptides such as defensins and lysozyme, but also include pro-inflammatory mediators and signal transduction proteins. For more information

about antimicrobial peptides see section 1.2.3.3. Furthermore they secrete factors required for stem cell maintenance including Wnt family member 3 (Wnt3), the Notch delta-like ligand (Dll4), and epidermal growth factor (EGF) (Wittkopf et al., 2014). Given the nature of these secreted products, it follows that Paneth cells play a key role in host defence against pathogens, modulating endogenous bacterial communities, immune regulation and intercellular communication. In addition, other unique specialised features of Paneth cells, such as their longevity, suggest further important functional roles are played by this population of cells.

1.2.1.5 Enteroendocrine cells

Making up 1% of the epithelium, enteroendocrine cells (EECs) are secretory cells which release peptide hormones, such as secretin and gastrin, in response to luminal nutrients. In addition, they are able to sense microbial metabolites and release cytokines in response (Worthington et al., 2018). Multiple subsets of EECs exist, which release specific cohorts of hormones (Haber et al., 2017). In addition, a recent study has identified diversity in function of EECs between villi and crypt locations based on bone morphogenetic protein (BMP) signalling gradient (Beumer et al., 2018).

1.2.1.6 Tuft cells

Despite occurring throughout the epithelium, tuft cells represent only 0.5% of the gut epithelial cells, depending on location in the gut (Banerjee et al., 2018). Tuft cells, named after their apical microvilli, are marked by doublecortin-like kinase 1 (DCLK1) (Gerbe et al., 2009). In the small intestine, tuft cells are believed to play a role in type 2 immunity against eukaryotic infections *via* chemosensory mechanisms. Colonic tuft cells have been poorly studied, but initial evidence suggests they may have different specification and function (Banerjee et al., 2018).

1.2.1.7 Microfold cells

Unlike the other aforementioned cell types, these cell are present only in follicle associated epithelium, where they make up 5-10% of cells (Nicoletti, 2000; Ohno, 2016). The primary role of M cells is to deliver microbial antigens to gut associated lymphoid tissue for efficient mucosal and systemic immune responses (Ohno, 2016).

1.2.2 Differentiation of intestinal epithelial cells

Intestinal epithelial cells arise from the LGR5⁺ stem cells of their associated crypt through a process of differentiation and migration. The favoured model of epithelial differentiation describes neutral competition between dividing stem cells which results in epithelial crypts drifting to clonality within a period of 1–6 months. Consequently, all cells of a crypt and its associated villus flanks originate clonally from one stem cell (Snippert et al., 2010).

Differentiation of epithelial stem cells is driven primarily by Notch, Wnt and BMP signalling pathways (Noah et al., 2011; Worthington et al., 2018). Primarily mediated by hairy and enhancer of split-1 (Hes1), Notch signalling drives differentiation of absorptive lineages such as enterocytes (Figure 1.3). Conversely, atonal homolog 1 (Atoh1/Math1) inhibits Notch signalling through Hes1 to drive secretory lineage differentiation. The Wnt/ β -catenin signalling pathway, which plays a role in maintenance of stemness, has also been shown to interact with Notch signalling and is therefore implicated in the secretory/absorptive lineage decision. BMP signalling has been shown to limit epithelial expansion through inhibiting self-renewal of stem cells (Qi et al., 2017). Following the secretory lineage decision, Growth factor independent 1 (Gfi1) is required for goblet cell and Paneth cell differentiation, while Neurog3 determines EEC differentiation (Figure 1.3). In addition, further lineage specific factors have been described for IECs, as shown in Figure 1.3.

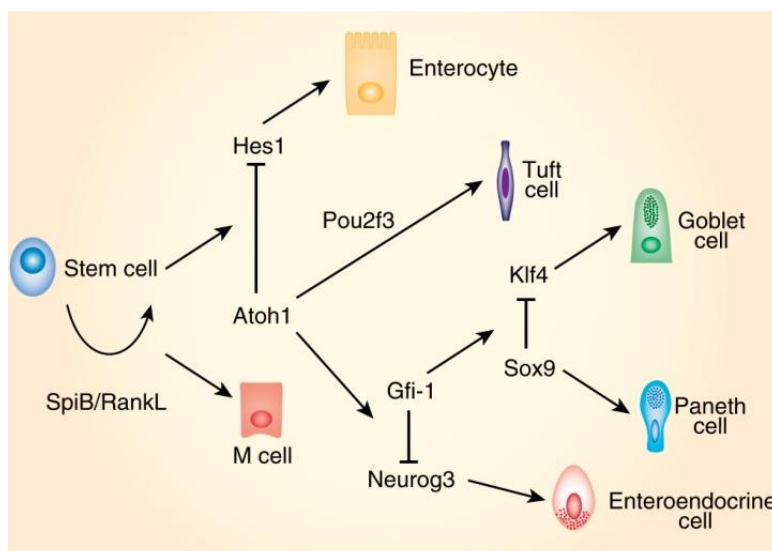


Figure 1.3. Intestinal epithelial cell differentiation. Image reproduced from Worthington et al. (2018) with permission of the rights holder, Springer Nature.

1.2.3 Epithelial barrier functions

The primary function of the epithelial layer is to provide a barrier between the gut lumen and the lamina propria to protect the host against microbes, toxins and other immunogenic molecules. This barrier consists of a number of components including the physical barrier of the IECs, the mucus layer and antimicrobial peptides. M-cells and intraepithelial immune cells are also implicated in barrier function, but are described in sections 1.2.1.7 and 1.3.

1.2.3.1 *Physical barrier function*

By forming a coherent monolayer of cells, connected by junctional complexes, IECs protect against harmful bacteria, antigens and toxins while permitting passage of nutrients and immune sensing functions (Vancamelbeke and Vermeire, 2017). Junctional complexes which join cells together include (apical) tight junctions, (central) adherens junctions and (basal) desmosomes (Groschwitz and Hogan, 2009; Williams et al., 2015). Disruption of the epithelial barrier can occur *via* dysregulation of junctional complexes or through excessive or dysregulated cellular shedding – a process of apoptotic extrusion of intestinal epithelial cells at the tip of the villi. Altered IEC physical barrier function is associated with a number of intestinal and extra-intestinal diseases such as IBD, coeliac disease and type I diabetes (Groschwitz and Hogan, 2009) and can be regulated by gut bacteria (Yu et al., 2012b).

1.2.3.2 *Mucus layer*

A further layer of defence in the gut is the mucus layer lining the epithelial cells – thin and loose in the small intestine while thick and dense in the colon. This layer consists of gel-forming glycoproteins called mucins which are secreted by goblet cells (see section 1.2.1.3). In the small and large intestines, MUC2 is the major constituent of mucus (Schroeder, 2019). The recognised functions of the intestinal mucus layer include: facilitating uptake of dietary molecules, a physical barrier for opportunistic pathogens in the outer layer of the colonic mucus, a site of long-term bacterial colonisation and a carbon and energy source for intestinal microbiota (Sicard et al., 2017; Vancamelbeke and Vermeire, 2017). Similarly to the cellular physical barrier, a reduction in mucus thickness and function has been observed in a number of health conditions including IBD (Johansson et al., 2014; Swidsinski et al., 2007) and is affected by gut microbes and their metabolites, including pathogens and commensals (Caballero-Franco et al., 2007; Sperandio et al., 2013; Wrzosek et al., 2013).

1.2.3.3 *Antimicrobial peptides*

Antimicrobial peptides (AMPs) are host defence peptides that play a major role in the innate immune protection of the intestines by protecting against pathogens and shaping the composition of the microbiome (Muniz et al., 2012). AMPs act primarily through direct antimicrobial activity, but are also capable of neutralising bacterial exotoxins, acting as chemoattractants for immune cells and modulating differentiation and maturation of immune cells (Grigat et al., 2007; Lehrer et al., 2009; Mahlapuu et al., 2016; Rodríguez-García et al., 2009; Yang et al., 1999). Within the gut, AMPs are primarily secreted by Paneth cells (see section 1.2.1.4) but are also secreted by other IECs such as enterocytes and by some immune cell populations such as neutrophils (Bevins and Salzman, 2011; Muniz et al., 2012). Three primary classes of AMP exist: defensins, cathelicidins and C-type lectins. Defensins are small cationic peptides which act primarily through disrupting bacteria cell walls or membranes. In human guts, there exists six α -defensins, expressed by neutrophils (human neutrophil peptides, HNPs 1–4) and Paneth cells (human α -defensins, HD-5 and HD-6). In mice α -defensins are known as cryptidins, of which there are 19, expressed primarily by Paneth cells (Muniz et al., 2012). Furthermore, there are numerous β -defensins expressed by different types of intestinal epithelial cells in both humans and mice. Similarly to defensins, cathelicidins are small cationic peptides with broad antibacterial activity. Only one cathelicidin has been identified in humans and mice: LL-37 and CRAMP respectively. C-type lectins, which consist of a carbohydrate recognition domain and an N-terminal signal peptide, exert antimicrobial activity against Gram-positive bacteria through binding to peptidoglycan. In humans, regenerating islet-derived protein 3 alpha (RegIII α) is the primary C-type lectin. The mouse ortholog of this gene is regenerating islet-derived protein 3 gamma (RegIII γ), and both are constitutively expressed in the intestinal epithelium but can be further induced by toll-like receptor (TLR) signalling (which is further described in section 1.2.4.1) (Cash et al., 2006). In addition to these classes, additional important AMPs are expressed in the intestine. For example, lysozyme C (lysozyme, LYZ1) is a glycoside hydrolase which cleaves peptidoglycan in Gram-positive bacterial cell wall. It is secreted by Paneth cells and in mice also by macrophages. Secretory phospholipase A2 (sPLA2) is a further Paneth cell expressed AMP, which affects microbial cell integrity by degrading bacterial phospholipids (Muniz et al., 2012).

1.2.4 Regulation of the epithelium

To carry out their diverse functions, IECs work in cohort to recognise and respond to a variety of signals, including those from immunological mediators, gut microbes and their metabolites. This plasticity in structure and function of the intestinal epithelium can result in increased protection against pathogens, tolerance of commensal bacteria and appropriate water and nutrient intake. However, dysregulation of this system or subversion by pathogens, can also lead to chronic inflammation, cancer and microbial invasion.

1.2.4.1 *Microbiome*

Pathogenic and commensal microbes play a role in the regulation of IECs through nucleic acids, small molecules, metabolites and proteins (Guyen-Maiorov et al., 2017). Pattern recognition receptors (PRRs) on the surface and endosomal membranes of epithelial cells are the primary mechanism through which microbes are detected. PRRs can recognise microbe-associated molecular patterns (MAMPs), including lipopolysaccharides, DNA and flagellin, and subsequently initiate protective inflammatory cascades (Coleman and Haller, 2017). As a result of activation, PRRs induce a number of responses including phagocytosis, inflammation and maturation of antigen-presenting cells (Hato and Dagher, 2015).

There are four main families of PRRs: toll-like receptors (TLRs), nucleotide oligomerization domain-like receptors (NLRs), C-type lectin receptors (CLRs) and retinoic acid inducible gene I-like receptors (RLR) (Gourbeyre et al., 2015). TLRs are a diverse class of membrane protein receptors which can activate mitogen-activated protein kinase (MAPK) and canonical nuclear factor kappa-light-chain-enhancer of activated B cells (NF- κ B) signalling cascades following binding of ligands and recruitment of adaptor proteins (Wells et al., 2011). Expression of TLRs increase under inflammatory conditions, indicating they play a key role in gut immune responses. One study showed that activation of TLR4 in human IEC cell lines (HT-29 and T84) enhanced production of pro-inflammatory cytokines by co-cultured peripheral blood mononuclear cells (PBMCs). In the same study it was found that apical IEC TLR9 activation resulted in a regulatory T helper cell 1 (Th1) effector immune response, indicating both pro- and anti-inflammatory roles for TLR signalling (de Kivit et al., 2011). Another class of plasma membrane receptors are CLRs which recognise carbohydrate structures in a calcium-dependent manner, playing a role in innate and adaptive immunity (Chiffolleau, 2018). On the other hand, NLRs are cytoplasmic receptors. The best studied NLRs are NOD1 and NOD2 which recognise a

diverse range of MAMPs and host damage molecular patterns. One example of NLR signalling is the activation of Paneth cell cytoplasmic NOD2 by bacterial muramyl dipeptide resulting in defensin production through NF- κ B signalling (Voss et al., 2006). RLRs are a family of cytoplasmic RNA helicases that recognise double stranded viral RNAs (Wells et al., 2011).

In addition to PRRs, bacteria can interact with host IECs directly via protein-protein interactions (PPIs) within or on the outside membrane of the host cells (Doxey and McConkey, 2013). Pathogens have been shown to use eukaryotic-like domains to mimic host proteins to hijack host processes for enhanced invasive abilities. One such example is the *Salmonella* protease, YhjJ, which interacts with the human selective autophagy receptor Microtubule-associated proteins 1A/1B light chain 3B (MAP1LC3B/LC3) resulting in cleavage of autophagy proteins and reduced intracellular clearance (Sudhakar et al., 2019). Commensal bacteria are also likely to exhibit molecular mimicry, but would be predominantly constrained to interact with the external surfaces of host cells. Direct interactions between commensal bacteria (and their metabolites) with IECs are less well studied, but could uncover mechanisms useful for developing new strategies to prevent and treat diseases. One study found that host-microbe interactions could be heavily driven by microbial metabolites (specifically N-acyl amides) which bind to host G-protein-coupled receptors (GPCRs) (Cohen et al., 2017). In addition to being common therapeutic targets for small-molecule treatments, GPCRs have been implicated in various diseases which also exhibit gut microbial changes e.g. inflammatory bowel disease and diabetes - highlighting the importance of studying these host-microbe interactions.

1.2.4.2 Immune system

As well as responding to microbial factors, IECs also respond to signals from adjacent immune cells. Immune cells of the gut are covered in more detail in the following section (1.3). The mediating factors are primarily cytokines: small secreted proteins which act in a paracrine fashion to induce wide ranging effects on their targets, including IECs. During inflammation, IECs are able to respond to secreted cytokines following binding to receptor proteins on their apical or basal membranes. For example, interleukin (IL)-2 can promote proliferation or apoptosis of IECs in a concentration dependent manner (Mishra et al., 2012) and IL-22 can induce expression of Paneth cell antimicrobial peptide regenerating islet-derived protein 3 gamma (RegIII γ) and initiate mucosal wound healing (Kinnebrew et al., 2012; Pickert et al., 2009). Further immune-IEC interactions are outlined in Figure 1.4. For example, IL-22 produced by innate lymphoid cell

(ILC) 2 is capable of inducing mucin production via STAT3 and increasing tight junction proteins, while IL-10 secreted by macrophages can induce epithelial repair (Soderholm and Pedicord, 2019). Further experiments have shown that, in addition to paracrine signalling, hetero-cellular communication exists between immune cells and IECs, for example, due to the formation of adjoining gap-junction channels (Al-Ghadban et al., 2016). However, Al-Ghadban et al. (2016) showed that gap-junction communication between macrophages and IECs results primarily in activation of macrophages rather than affecting IECs. Macrophage activation in turn causes release of inflammatory cytokines, which affect IECs through paracrine signalling. Together, these evidences highlight the complex and varied communications which occur between IECs and immune cells. Furthermore, these interactions are additionally influenced by signals such as microbial recognition, adding a level of complexity to the system. The interaction between immune cells and IECs is relevant to the pathogenesis of IBD - for more information about cytokines in IBD see Section 1.4.2.

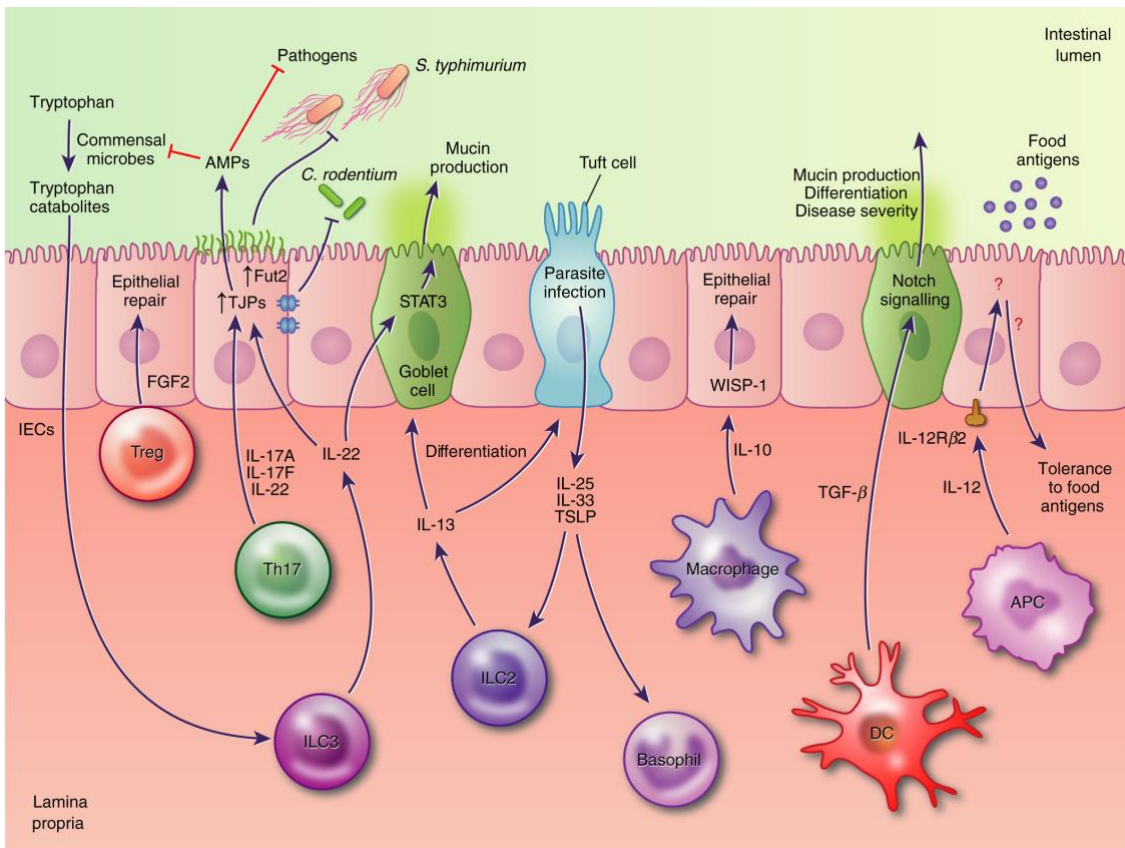


Figure 1.4. Regulation of intestinal epithelial cells (IECs) by immune cells. Communication is primarily driven by cytokines released by immune cells of the lamina propria. TJPs - tight junction proteins; FGF2 - fibroblast growth factor 2; IL - interleukin; TSLP - thymic stromal lymphopoietin; TGF- β – transforming growth factor beta; ILC - innate lymphoid cell; Th - T helper cell; DC - dendritic cell; APC - antigen presenting cell. Image reproduced from Soderholm et al. (2019) with permission of the rights holder, John Wiley and Sons.

1.3 Immune cells of the gut

In addition to immune functions of the epithelial layer, the gut immune system consists of phagocytes and lymphocytes. Phagocytes, including macrophages, neutrophils and dendritic cells, are cells of the innate immune system which primarily act through phagocytosis to engulf bacteria and other foreign particles, kill them and then present antigens to other immune cells. On the other hand, lymphocytes, which include T cells, B cells and natural killer cells, function primarily as part of an adaptive immune response following activation by foreign antigens.

The gut-associated lymphoid tissue contains approximately 70% of the body's immune cells (Heel et al., 1997). It has a complex organisation which can be divided into effector sites or organised tissues (Heel et al., 1997). Effector sites include intraepithelial lymphocytes within the

epithelium and immune cells within the lamina propria (Mowat, 2003). Organised tissues include Peyer's patches and mesenteric lymph nodes. Peyer's patches, found only in the small intestine, are lymphoid aggregates consisting of B cells and T cells separated from the lumen by epithelial cells known as follicle-associated epithelium. Differing from the standard epithelium, the follicle-associated epithelium also contains M cells (Section 1.2.1.7) and is infiltrated with B cells, T cells, macrophages and dendritic cells. Smaller individual lymphoid follicles also line the small intestine and colon (Brandtzaeg, 2017). In addition, other immune cells such as dendritic cells, macrophages and T cells are present in the lamina propria (Macdonald and Monteleone, 2005). Here I focus specifically on T cells, whose secreted cytokines are studied in Chapter 3 in the context of inflammatory bowel disease, and on macrophages, which are studied in Chapter 6 due to their key role in balancing immune activation and tolerance in response to gut microbes.

1.3.1 Intestinal T cells

As discussed in Section 1.2.4.2, cytokines are the primary mediators of communication from immune cells to IECs. Several different populations of immune cells secrete cytokines, including macrophages, B cells and T cells (Xue and Falcon, 2019). In particular, cytokines secreted by T helper (Th) cells have been implicated in the pathogenesis of chronic gut inflammatory diseases such as IBD. The effect of T cell secreted cytokines on colonic IECs is studied in Chapter 3 in the context of IBD.

T cells are lymphocyte cells which play a major role in controlling intestinal homeostasis through sophisticated mechanisms balancing immune activation and tolerance. All T cells originate from haemopoietic stem cells within the bone marrow which migrate to the thymus, differentiate into specialised T cell types, undergo priming by antigens in peripheral lymphoid organs and finally migrate to gut tissues. Two major subsets of T cells can be classified based on T cell receptor (TCR) and coreceptor expression. 'Type a' conventional mucosal T cells express TCR $\alpha\beta$ and TCR coreceptors CD4 or CD8 $\alpha\beta$. 'Type b' non-conventional mucosal T cells express either TCR $\alpha\beta$ or TCR $\gamma\delta$ and usually coreceptor CD8 $\alpha\alpha$. Non-conventional T cells are not primed within peripheral lymphoid tissues but migrate to the gut directly from the thymus. Within the gut conventional T cells reside primarily in the lamina propria whereas non-conventional T cells reside primarily in the epithelium as intraepithelial lymphoid cells (Ma et al., 2019; van Wijk and Cheroutre, 2010). Within the epithelium, non-conventional T cells are able to react to strong

stimulation in a cytolytic manner, but have limited pathogen specificity and are finely tuned to avoid uncontrolled immune reactions (van Wijk and Cheroutre, 2010). Further, evidence suggests that non-conventional T cells play a role in regulating epithelial turnover and repair (Komano et al., 1995)

Conventional T cells can be further classified based on expression of CD8 or CD4 on their surface. CD8+ T cells, also known as cytotoxic T cells, primarily act to destroy infected cells and tumour cells through cytotoxin secretions such as perforin. While also primed by antigen presenting cells, CD4+ T cells, act primarily to organise the immune response through secreting cytokines. CD4+ T cells include regulatory T cells and helper T cells (Th cells). Th cells secrete a variety of pro- and anti-inflammatory cytokines which assist an immune response through influencing maturation of B cells and activation of cytotoxic T cells and macrophages. For example, Th1 cells secrete interferon-gamma (IFN γ) which acts against intracellular bacteria, viruses and cancer. On the other hand, Th9 cells secrete interleukin-9 (IL-9) which can defend against helminths. Th cells and their secreted cytokines which are implicated in IBD are discussed in section 1.4.2. On the other hand, regulatory T cells act to inhibit Th cells through direct contact or by releasing anti-inflammatory cytokines such as IL-10 (Xue and Falcon, 2019).

1.3.2 Intestinal macrophages

In the lamina propria of a healthy gut, macrophages are the most abundant white blood cells (Mowat and Agace, 2014). In addition to phagocytosis, macrophages can produce mediators which drive epithelial cell regeneration and T cell differentiation as well as secreting anti-inflammatory cytokines. Interestingly, it has been found that most intestinal macrophages do not produce pro-inflammatory cytokines upon exposure to bacteria or their products, likely for the purpose of preventing inflammation in the mucosa (Bain et al., 2013; Smythies et al., 2005). It is currently believed that this is not driven by downregulation of bacterial recognition receptors (such as TLRs), but rather by blocking the downstream signals from these receptors within the macrophages. This blocking occurs by downregulation of signalling molecules such as cluster of differentiation 14 (CD14) and Myeloid differentiation primary response 88 (MyD88) (Bain and Mowat, 2014; Smith et al., 2011; Wang et al., 2019b). The underlying cause of this downregulation is unconfirmed, but might be related to high anti-inflammatory interleukin (IL)-10 production by macrophages themselves. Other potential mechanisms include the release of transforming growth factor- β (TGF- β) by stromal cells, which has been shown to block NF- κ B (a

central mediator of inflammatory processes) in response to TLR2, 4 and 5 activation in macrophages (Naiki et al., 2005; Smith et al., 2011; Smythies et al., 2010).

NF- κ B is a small family of inducible nuclear transcription factors which play a key role in almost all mammalian cells. In addition to inflammatory processes, NF- κ B is a central mediator of stress response and cell proliferation. In intestinal macrophages, NF- κ B is important for regulation of inflammatory response following activation of cell surface pattern recognition receptors and cytokine receptors such as TLRs (Dorrington and Fraser, 2019; Neurath et al., 1998). NF- κ B is of particular importance as it is overactivated in IBD patients resulting in increased pro-inflammatory cytokine production (Atreya et al., 2008; Schreiber et al., 1998).

1.4 Inflammatory bowel disease

Dysfunction of gut epithelial cell and immune cell functions, for example disruption of epithelial integrity, can predispose to microbial infections, food allergy and a number of gut diseases including IBD (König et al., 2016). IBD is a multi-systemic inflammatory disorder primarily characterised by chronic inflammation of the gastrointestinal tract, including dysfunction of the epithelial and immune cells (Levine et al., 2018). Chronic intestinal inflammation causes debilitating symptoms, such as abdominal pain and diarrhoea, and severe complications, such as cancer and intestinal failure (Mozdiak et al., 2015; Seyedian et al., 2019). The two major forms of IBD are Crohn's disease (CD) and ulcerative colitis (UC). In CD, inflammation can affect the entire bowel wall in any part of the small and/or large intestine. In UC, inflammation is contained to the epithelial lining (mucosa) of the large intestine. While described as idiopathic, IBD is believed to be caused by an inappropriate immune response to commensal bacteria in genetically susceptible hosts (Fakhoury et al., 2014). For example, it has been shown that altered mucus production and epithelial barrier dysfunction can result in increased translocation of toxins and microbes, which in turn causes a pro-inflammatory immune response and increased susceptibility to infection (Fakhoury et al., 2014). This disease has a global and accelerating incidence, particularly in industrialised communities. In Europe alone, approximately 2.5-3 million people are affected (Burisch et al., 2013; Ng et al., 2018).

1.4.1 The intestinal epithelium in IBD

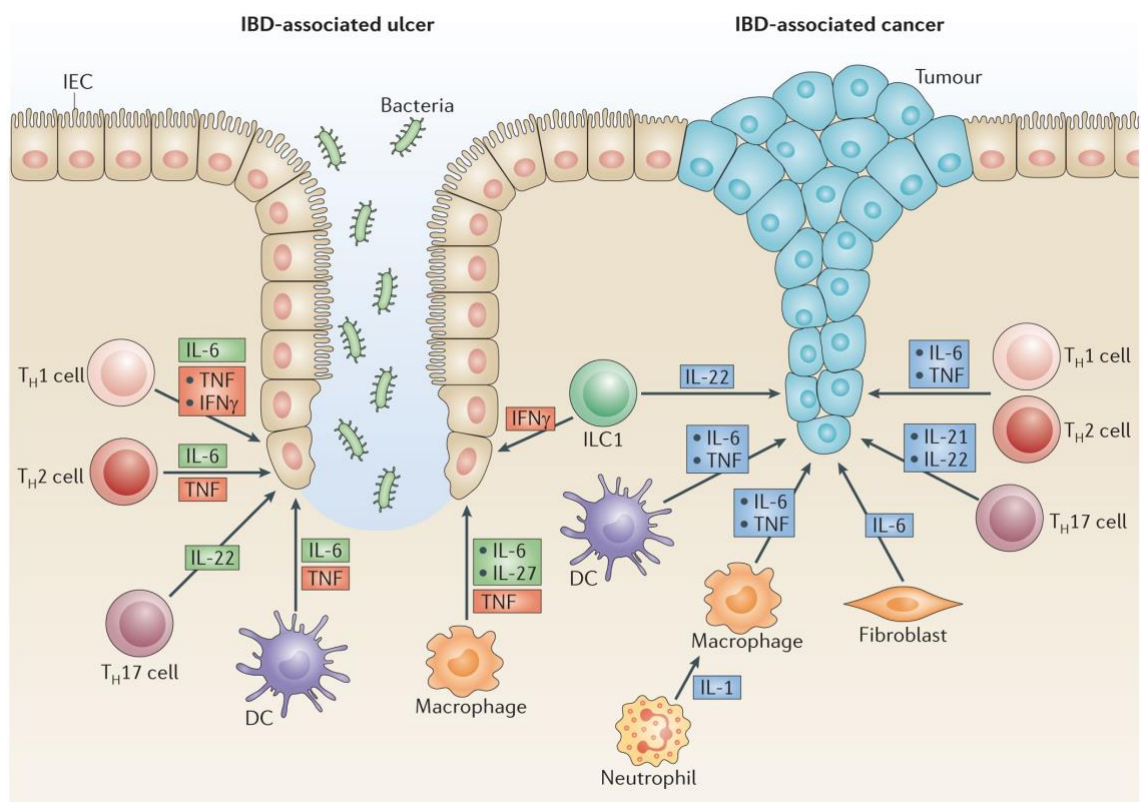
Although the aetiology of IBD is incompletely understood, the intestinal epithelium is increasingly regarded as a central player (Okamoto and Watanabe, 2016). In IBD, dysregulated epithelial processes, including microbial sensing, autophagy, and the unfolded protein response are mechanistically implicated in impaired barrier function and IBD aetiology (Kaser et al., 2008; Wehkamp et al., 2005). In turn, chronic inflammation in the gut, as observed in IBD, can result from an impaired barrier function which leads to greater translocation of luminal content through the epithelial layer, resulting in an overactivation of lamina propria immune responses.

Current understanding implicates specific IEC types in the dysregulation of homeostasis in IBD (Adolph et al., 2013). Whilst primary IECs all originate from Leucine-rich repeat-containing G-protein coupled receptor 5 (Lgr5)⁺ stem cells, differentiation results in differences in gene expression and signalling and regulatory wiring (Crosnier et al., 2006; Vanuytsel et al., 2013). These differences can result in altered phenotypic functions, responses to stress and susceptibilities to specific dysregulations. Specifically, dysfunctional Paneth cells with reduced secretion of anti-microbial peptides have been shown to contribute to the pathogenesis of CD (Liu et al., 2016). In contrast to CD, UC is not linked to reduced antimicrobial peptides (Fahlgren et al., 2003; Nuding et al., 2007; Wehkamp et al., 2003). However, where the mucus layer is of normal or thicker width in CD, it is thinner and more variable in UC (McCormick et al., 1990; Pullan et al., 1994). Furthermore, a reduction in goblet cell numbers and defective goblet cell function has been associated with UC (Gersemann et al., 2009; Kim and Ho, 2010). Moreover, genome-wide association studies and mechanistic studies have identified genes closely related to intestinal epithelial cell function which predispose patients to IBD (Franke et al., 2010; Rioux et al., 2007). For example, Cadwell et al. found that one Crohn's disease risk allele, autophagy related protein 16-1 (ATG16L1), results in Paneth cell granule abnormalities and, in mice, results in increased expression of lipid metabolism genes relating to intestinal injury response (Cadwell et al., 2008). Thus, uncovering patterns and mechanisms at a cell type-specific level is crucial to uncover the role of the intestinal epithelium in IBD.

1.4.2 Cytokines in IBD

In addition to its aforementioned roles, the intestinal epithelium can secrete cytokines and chemokines for recruitment and activation of immune cells (Allaire et al., 2018). Furthermore, epithelial function can be directly regulated by mucosal immune activity, and cytokines

produced by tissue resident lymphocytes, particularly T cells and innate lymphoid cells, profoundly impact epithelial phenotype (Dahan et al., 2007). Studies have shown that mucosal healing of the epithelium is dependent on cytokines produced by the intestinal epithelial cells (IECs) and by local immune populations (Figure 1.5) (Neurath, 2014). Furthermore, cytokines such as TNF α , interleukin (IL)-22 and IL-9, which are excessively produced in IBD, have been shown to drive epithelial-specific pathological processes, including endoplasmic reticulum (ER) stress, apoptosis and impaired barrier function, triggering colitis in preclinical models (Garrett et al., 2007; Gerlach et al., 2014).



Nature Reviews | Immunology

Figure 1.5. Pro- and anti-inflammatory cytokines are exposed to intestinal epithelial cells during chronic intestinal inflammation in inflammatory bowel disease. Green boxes show beneficial effects of cytokines, red boxes highlight pathogenic effects of cytokines and blue boxes indicate pro-tumour effects of cytokines. DC - dendritic cell; IFN - interferon; IL - interleukin; ILC - innate lymphoid cell; Th cell - T helper cell; TNF - tumour necrosis factor. Image reproduced from Neurath (2014) with permission of the rights holder, Springer Nature.

It is believed that CD4⁺ T helper cells (Th cells) (introduced in Section 1.3.1) play a major role in initiation of IBD (Imam et al., 2018). T helper cells are lymphoid cells of the adaptive immune system which are able to activate B cells, cytotoxic T cells and macrophages *via* surface molecules and release of cytokines. They are activated themselves by recognising an antigen

and a co-stimulatory molecule on an antigen-presenting cell (Alberts et al., 2002). Not only are these cells enriched in inflamed tissue from IBD patients, depleting and blocking actions of CD4+ Th cells has been effective in treating UC and CD (Emmrich et al., 1991; Stronkhorst et al., 1997; Imam et al., 2018). Five major populations of Th cells have been described in the context of IBD (Imam et al., 2018). A representative cytokine(s) from each category are described below:

- Th1 cells which secrete interferon-gamma (IFN γ) and tumour necrosis factor-alpha (TNF α)
- Th2 cells which secrete IL-4, IL-5 and IL-13
- Th9 cells which secrete IL-9
- Th17 cells which secrete IL-17A and IL-23
- Th22 cells which secrete IL-22

TNF α is one of the pro-inflammatory cytokines implicated in IBD. Secreted primarily by macrophages in response to IL-1 and bacterial products, TNF α is involved in a number of biological processes including lipid metabolism, cell proliferation and apoptosis (Adegbola et al., 2018). IBD patients have been shown to secrete large amounts of TNF α from adipocytes, CD14+ macrophages, T cells and fibroblasts (Atreya et al., 2011; Kamada et al., 2008). In IBD, high levels of TNF α affect a number of different cells including IECs. For example, TNF α can induce IEC damage via myosin light chain kinase (MLCK) and cause Paneth cell death by necrosis (Günther et al., 2011; Neurath, 2014; Su et al., 2013).

Similarly to TNF α , IFN γ is an IBD-implicated, pro-inflammatory cytokine which affects barrier properties and self-renewal of the intestinal epithelium (Nava et al., 2010). IFN γ is secreted by a number of cell types including T cells (most notably Th1 cells) and natural killer cells (NK cells) (Imam et al., 2018; Tau and Rothman, 1999). A number of studies have demonstrated a role for IFN γ in IBD. For example, in CD it has been shown that Th1 cells, which primarily secrete IFN γ , accumulate in the intestinal tract. Ito et al. showed that IFN γ deficient mice do not develop dextran sulphate sodium induced (DSS-induced) colitis and a meta-analysis of CD and UC genome-wide association scans by Jostins *et al.* identified IFN γ receptor gene *IFNGR2* within CD risk loci (Ito et al., 2006; Jostins et al., 2012). However, the role of IFN γ remains controversial due to contrasting studies. For example, one study showed that IFN γ deficient mice are more susceptible to DSS-induced colitis and another described neither a protective or detrimental

effect of IFN γ in a 2,4,6-trinitrobenzene sulfonic acid induced (TNBS) model of colitis (Jin et al., 2012; Muzaki et al., 2016).

IL-13 is classically described as Th2 cell cytokine which affects macrophages, epithelial cells, smooth muscle cells and neurons (Biancheri et al., 2014). IL-13 has been implicated in UC, where it is secreted by innate lymphoid cells and invariant natural killer T cells (Fuss and Strober, 2008). However, while some studies have reported greater IL-13 in UC patients compared to CD patients, others have reported the opposite or no difference at all (Biancheri et al., 2014; Fuss et al., 2004; Vainer et al., 2000). Similarly, the pro-inflammatory role of IL-13 is controversial. Heller *et al.* reported an increase in apoptosis and impaired tight junctions in IEC's treated with IL-13, whereas Wilson *et al.* report that in the absence of IL-13 decoy receptor, IL-13R α 2, IL-13 suppresses pro-inflammatory Th1 and Th17 responses (Heller et al., 2005; Wilson et al., 2011).

Traditionally, IBD was described in the context of the Th1/Th2 paradigm. Here Th cells are classified into two primary subsets based on their secreted cytokines: Th1 which broadly induced cell mediated immunity, or Th2 which broadly induces humeral immunity. Furthermore, the paradigm was used to classify an inflammatory response as primarily Th1 or Th2 mediated – given that the cell subpopulations reciprocally inhibit each other (Fuss, 2008; Kiely, 1998). In IBD, this paradigm is used to describe mucosal inflammation of the gut in CD as an excessive Th1 response and in UC an excessive Th2 response (Fuss, 2008). However, more recently the discovery of other IBD-relevant CD4+ Th cells such as Th17 cells has drawn focus away from this polarising paradigm. Th17 cells secrete cytokines IL-17A (also known as IL-17) and/or IL-17F which have been shown to alter the production of inflammatory chemokines and cytokines by target cells and affect the epithelial cell barrier (Imam et al., 2018). Similarly to other pro-inflammatory cytokines, IL-17A has been found to be increased in the mucosa and serum of IBD patients (Fujino et al., 2003). However, there is also evidence that blocking IL-17A can result in IBD deterioration (Smith et al., 2019) and that IL-17A can dampen the production of IFN γ (O'Connor et al., 2009).

In both CD and UD, endoscopic Mayo scores have been positively correlated with IL-9 production (Gerlach et al., 2014) and higher levels of serum and systemic IL-9 have been associated with worse symptoms and prognosis (Defendenti et al., 2015). However, IL-9 is secreted by different immune populations including Th9, Th17 and Treg cells. Studies have found that IL-9 produced

by Th17 is classically pro-inflammatory in a model of experimental autoimmune encephalomyelitis, whereas IL-9 secreted from Treg cells can mediate graft tolerance (Lu et al., 2006; Nowak et al., 2009). This suggests that context may play a role in the effect of IL-9.

Along with ILCs and neutrophils and dendritic cells, Th22 cells produce the IL-22 cytokine. While primarily seen as an anti-inflammatory cytokine, IL-22 has been shown to have pro-inflammatory roles in some contexts (Eken et al., 2014; Neurath, 2014). Anti-inflammatory roles include protecting against DSS or TNBS-induced colitis, increasing epithelial cell proliferation, wound healing and IEC production of antimicrobial peptides. A reduction in Th22 cells has been recorded for UC but not CD (Neurath, 2014).

In conclusion, there are many diverse cytokines produced by different CD4⁺ Th cells and other immune populations. The cellular targets of these cytokines are also varied, including other immune populations, IECs and muscle cells. What is clear is that considerable interplay exists between different cytokine-producing populations and that cytokine-mediated effects are highly contextual. However, little is known about how qualitatively different arms of host immunity differentially regulate epithelial function in IBD. Understanding these interactions has been hindered by inaccessibility of the human gastrointestinal tract and limitations of available experimental tools, including immortalised epithelial cell lines and primary cells.

1.4.3 IBD treatment strategy

Classification of IBD, as with other immune mediated inflammatory diseases, is typically based on descriptive clinical parameters, which are poor predictors of patient trajectories and are unhelpful as tools for treatment stratification. As such, most patients are treated with a step-up therapy approach whereby frontline therapies are given in a stepwise manner until suitable remission is achieved. Although the guidelines vary slightly between CD and UC, treatment often begins with corticosteroids or aminosalicylate drugs followed by immunomodulators such as thiopurines, then biologics including anti-cytokine and anti-integrin drugs. Finally for UC, a Janus Kinase (JAK) inhibitor (tofacitinib) can be used to block cytokine signalling (Lamb et al., 2019; Wang et al., 2019a). If all lifestyle changes and pharmacological treatments fail, patients may be recommended surgery to remove affected intestinal tissue. Such a procedure can eliminate UC but is usually only a temporary solution for CD as the disease often reoccurs in nearby tissue.

However, due to the heterogeneity of IBD, no single approach works for all, and adaptation of treatment approaches to each individual based on their disease will likely result in better clinical outcomes, reduced side effects and reduced healthcare costs. A number of molecular biomarkers exist currently to help diagnose and stratify patients. Examples of these include C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), albumin and platelet count; but their use is limited by poor sensitivity and specificity. Additionally, faecal calprotectin, which indicates the level of neutrophil-driven inflammation in the gut, is often used as a proxy for intestinal inflammation and to help the clinician differentiate between irritable bowel syndrome (IBS) and IBD (Wang et al., 2019a). However, raised calprotectin levels may also indicate intestinal inflammation secondary to other causes. Therefore, better biomarkers to diagnose and stratify patients based on disease severity, risk of relapse and treatment responses, are required to achieve an individualised treatment approach. In the future, it is hoped that an improved molecular understanding of inflammation biology will yield novel classification systems underpinned by the underlying immunopathology of the different diseases.

1.4.4 IBD biologic treatments

Despite remaining questions regarding the actions of cytokines in IBD, a number of cytokine-targeting biologic treatments have been developed to dampen the generalised inflammatory response in UC and CD. As described above, these treatments are often employed only when other treatment options have failed.

The most widely targeted cytokine is TNF α , with four biologics (infliximab, adalimumab, certolizumab and golimumab) and a number of biosimilars currently approved for use in CD and UC (Rawla et al., 2018). Despite demonstrable improvements in patient quality of life and disease burden, up to 30% of patients with IBD do not respond to this treatment, and up to 46% lose response over time (Roda et al., 2016). Neither the mechanism of action of anti-TNF treatments nor the reasons for non-response are fully understood. However, it is known that the effects of these anti-TNF drugs cannot be attributed solely to neutralisation of TNF α , as other anti-TNF drugs such as Etanercept, which is used primarily for rheumatoid arthritis, are not effective in IBD, and can even worsen the disease (Koelink et al., 2019; Korzenik et al., 2019). The only other approved biologic targeting interleukins is ustekinumab, an antagonist of IL-12 and IL-23 for the treatment of CD (Rawla et al., 2018). Clinical trials for IL-17A blockade treatment (bimekizumab) were terminated early due to adverse side effects and no clear

evidence of efficacy, despite success treating psoriasis (EU Clinical Trials Register, 2019). Similarly, for UC, anti-IL-13 antibodies (anrukinzumab and tralokinumab) have been trialled with little therapeutic benefit (Danese et al., 2015; Reinisch et al., 2015)

A greater understanding of cytokines in IBD, including the interplay between IECs, immune cells and cytokines, is required for development of more effective treatment and preventative strategies.

1.5 Bifidobacteria

In 1899 Henry Tissier was the first to describe the bacterial group *Lactobacillus bifidus*, which was re-classified after the 1960's as the genus *Bifidobacterium* (Tissier, 1899, 1900). Through observation of their predominance within the gut of breast-fed infants he conceived their use as probiotics, promoting oral administration as a therapeutic for infant diarrhoea. Since 1899 the genus has been well documented as a health-promoting commensal and is now one of the most heavily used probiotic taxa, alongside Lactobacilli (O'Neill et al., 2017; O'Toole et al., 2017).

As a genus of the *Actinobacteria* phyla, *Bifidobacterium* are saccharolytic and anaerobic Gram positive bacteria which are non-sporulating, non-gas producing and non-motile (Bottacini et al., 2014). At present 80 (sub)species have been classified, together occupying a range of ecological niches including sewage, water kefir, insect guts and the gastro-intestinal tracts and oral cavities of various mammals (Turrone et al., 2011, 2019). Some of these bifidobacterial species are considered pioneer species of the human gut acquired shortly after birth from ingestion of breast milk (Lewis and Mills, 2017). In particular, *Bifidobacterium infantis*, *Bifidobacterium longum*, *Bifidobacterium breve* and *Bifidobacterium bifidum* are the primary bifidobacterial species present within infant gastro-intestinal tracts, with increasing diversification of the genus seen with age (Di Gioia et al., 2014). Notably, the proportion of bifidobacteria observed within the human gut also varies across the life course (Figure 1.6). It is widely accepted that newly born infants host the highest proportions, with cited figures between 45 and 95% for breast fed babies (Arboleya et al., 2016; Bezirtzoglou et al., 2011; Fallani et al., 2010). Whilst still an important member of the microbiota, the proportion of bifidobacteria reduces gradually into adulthood where they remain relatively stable at around 3 to 10% until old age when they decrease further (Arboleya et al., 2016).

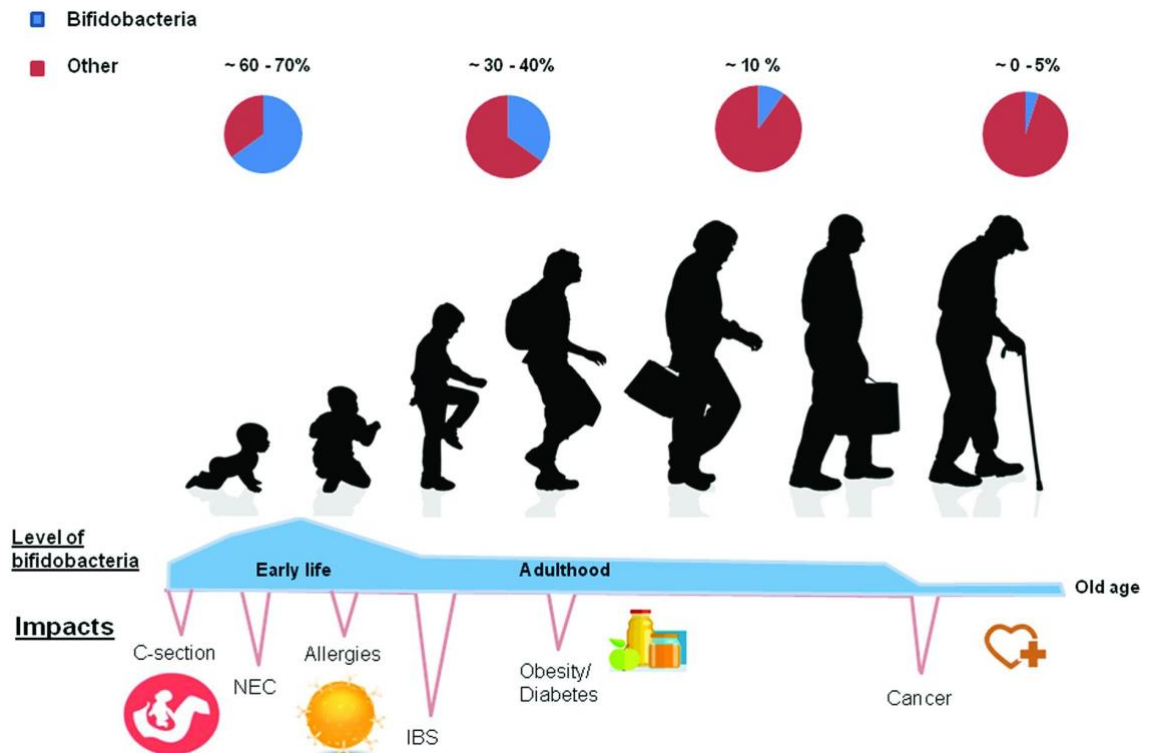


Figure 1.6. Relative abundance of gut bifidobacteria during the human life cycle. Environmental factors and conditions associated with low bifidobacterial counts are shown below. Figure reproduced from Arboleya et al. (2016) under the Creative Commons BY licence.

1.5.1 Health benefits of bifidobacteria

Many studies have identified an association with decreased bifidobacterial levels and reduced diversity of gut bifidobacterial species with increased disease symptoms (O’Callaghan and van Sinderen, 2016; O’Neill et al., 2017; Tojo et al., 2014). For example, a higher proportion of gut *B. longum* has been observed in healthy children compared to children with allergic disease (Akay et al., 2014; Ouwehand et al., 2001) and larger populations of bifidobacterial species have been observed in healthy children compared to those with active and non-active coeliac disease (Collado et al., 2008). It has also been observed in adults, that lower levels of gut *Bifidobacterium* and *Lactobacillus* are associated with diseases such as functional constipation and Irritable Bowel Syndrome (IBS) and ulcerative colitis (Khalif et al., 2005; Kim et al., 2015b; Macfarlane et al., 2004; Parkes et al., 2012). Additionally, the levels of mucosal *Bifidobacterium* have been negatively associated with the number of days patients with IBS experienced pain or discomfort (Parkes et al., 2012).

Furthermore, supplementation of the resident microbiota with additional bifidobacteria appears to have a variety of positive outcomes. This has been shown in healthy patients, for example supplementation with *Bifidobacterium animalis* DN-173 010 decreased transit time in healthy women, suggesting an association between the bacteria and gut motility (Marteau et al., 2002). Furthermore, there have been benefits of using bifidobacteria in intervention studies for gastro-intestinal disorders such as IBS, ulcerative colitis and lactose intolerance; although many of these studies are carried out using combinations of probiotics and a food starter, so it is hard to definitively attribute the effects to increases in bifidobacteria (Furrie et al., 2005; He et al., 2008; O'Mahony et al., 2005). More mechanistic studies have indicated that different bifidobacterial species are capable of modulating immune function, thus contributing to immune maturation, gut homeostasis, pathogen protection and anti-tumour immunity (Hart et al., 2004; Silva et al., 2004; Sivan et al., 2015).

Despite this significant body of evidence, the exact factors which modulate these protective effects are only beginning to be elucidated, with progress heavily hampered by the complexity (microbe genomics, impact of diet, host responses etc.) and inaccessibility of the gastro-intestinal ecosystem (O'Neill et al., 2017; Russell et al., 2011).

1.5.2 The effect of bifidobacteria on intestinal epithelial cells

IECs are a primary site of interaction between bifidobacteria and their host, thus, play a significant role in mediating the host response and beneficial effects of bifidobacteria. A number of *in vivo* and *in vitro* studies have evidenced the ability for bifidobacterial strains to modulate the function of IECs through a variety of molecular mechanisms including metabolites, proteinaceous pili and exopolysaccharides (EPS) (Castro-Bravo et al., 2019; Fanning et al., 2012a; Lee et al., 2018; O'Connell Motherway et al., 2019). These findings are explored below and summarised in Figure 1.7 - categorised based on the observed host functional changes.

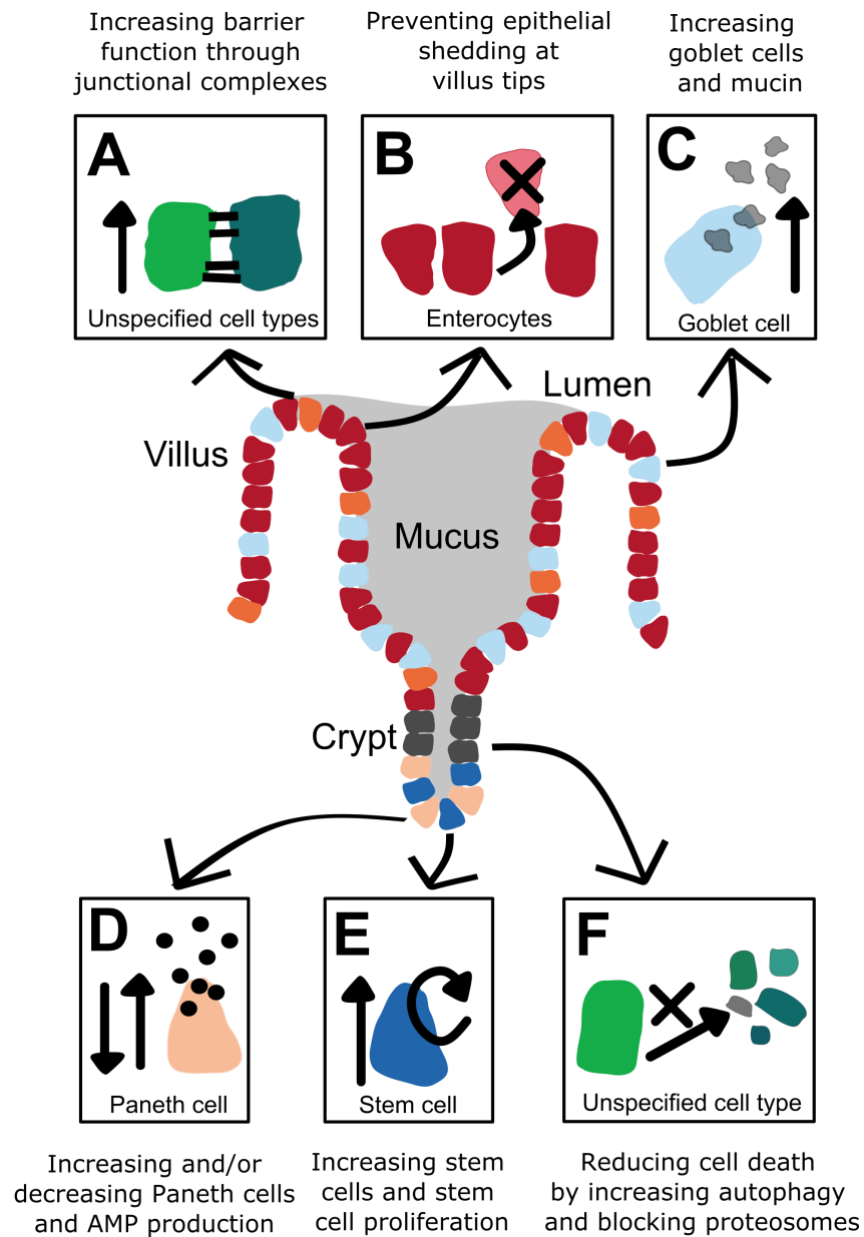


Figure 1.7. Summary of the different ways in which Bifidobacterial strains have been shown to affect the intestinal epithelial layer. **A.** Increasing barrier function through junctional complexes (Din et al., 2020; Hsieh et al., 2015; Srutkova et al., 2015; Yan et al., 2019; Yang et al., 2017). **B.** Preventing epithelial cell shedding of enterocytes at the villus tips (Hughes et al., 2017) **C.** Increasing numbers of goblet cells and production and expulsion of mucins (Becker et al., 2013; Engevik et al., 2019; Mangin et al., 2018; Schroeder et al., 2018) **D.** Different experiments have shown Increases and decreases in Paneth cell numbers and production of antimicrobial peptides (Lee et al., 2018; Natividad et al., 2013; Pinto-Sánchez et al., 2017; Underwood et al., 2012). **E.** Increasing numbers of stem cells and stem cell proliferation (Lee et al., 2018) **F.** Reduced cell death through increasing autophagy and blocking proteasomes (Inaba et al., 2016; Lin et al., 2014). Referenced evidences are from small intestinal cells, colonic cells and cell lines. AMP – antimicrobial peptide.

1.5.2.1 Epithelial physical barrier function

A primary function of IECs is the maintenance of an effective physical barrier between the gut lumen and the lamina propria, as described in section 1.2.3.1. Decreased expression of junctional complexes and a correlated increase in gut permeability is observed in mice treated with dextran sulphate sodium (DSS) – a commonly studied mouse colitis model. Studies have shown that specific strains of *Bifidobacterium* can protect against epithelial damage associated with DSS treatment (Figure 1.7A). For example, *Bifidobacterium longum* CCM 7952 promoted epithelial barrier function thorough preventing DSS-induced transcriptional decreases in tight junction associated genes *Occludin* and *Zonulin-1* (Srutkova et al., 2015). Similarly, in separate experiments, *B. longum subsp. longum* YS108R and *Bifidobacterium bifidum* ATCC 29521 were shown to maintain the expression of tight junction associated genes *Claudin-1*, *Claudin-3* and *Zonula occludens-1* and mucin gene *Muc2* following DSS-treatment (Din et al., 2020; Yan et al., 2019). A study of transepithelial electrical resistance in IEC18 cells showed that *Bifidobacterium infantis*, *Bifidobacterium youth*, *B. longum* and *B. bifidum* strains had only a modest effect on cellular permeability, likely abated by decreased expression of innate immune receptors Toll-like receptor (TLR) 2 and/or TLR4 (Yang et al., 2017). Further study of transepithelial electrical resistance in Caco-2 cell lines showed that *Bifidobacterium* species prevent intestinal epithelial barrier disruption induced by TNF- α , *via* metabolites such as acetate (Hsieh et al., 2015). Barrier function has also been studied in the context of epithelial cell shedding, where it was found that *Bifidobacterium breve* UCC2003 reduced small intestinal cell shedding in mice in an EPS-dependent manner (Figure 1.7B) (Hughes et al., 2017). Here, they observed that MyD88, a downstream effector molecule of TLR signalling, was required to modulate the protective effect.

1.5.2.2 Mucus layer

A further layer of defence in the gut is the mucus layer lining the epithelial cells, as described in section 1.2.3.2. Several papers have evidenced a role for Bifidobacteria in fortifying and protecting the intestinal mucus layer through alterations of goblet cell function (Figure 1.7C). For example, Becker *et al.* found that heat inactivated *B. breve* could increase mucin *Muc1* expression, but not *Muc2* expression in cell line LS174T (Becker et al., 2013). Further, *Bifidobacterium dentium* secreted acetate was able to increase MUC2 levels in T84 cells, while secreted γ -aminobutyric acid (GABA) stimulated mucin expulsion *via* an increase in autophagy (Engevik et al., 2019). This work also investigated *B. dentium* monoassociated mice compared to germ free mice, finding that live bifidobacteria, unlike heat-killed bifidobacteria, increase

colonic expression of goblet cell markers Krüppel-like factor 4 (*Klf4*), Trefoil factor 3 (*Tff3*), Resistin-like molecule β (*Relm- β*), Mucin-2 (*Muc2*), and several glycosyltransferases (Engevik et al., 2019). Similarly, in a rat model it was found that viable *Bifidobacterium pseudolongum patronus* increased colonic mucus layer thickness (Mangin et al., 2018), and in a mouse model *B. longum* NCC 2705 was able to ameliorate mucus growth defects but not repair penetrability, following damage from a western style diet (Schroeder et al., 2018).

Together these experiments highlight the complexity of mucus homeostasis; being dependent on constitution and thickness of mucus as well as goblet cell numbers, and goblet cell production and expulsion of mucins. Further, Bifidobacteria has strain specific effects on mucus, mediated by different factors associated with heat-killed and viable bacteria.

1.5.2.3 Antimicrobial peptides

As described in section 1.2.3.3, antimicrobial peptides (AMPs) are host defence peptides that play a major role in the innate immune protection of the intestines. A link has been found between bifidobacteria and the alteration of AMP production in the gut, however evidence is inconsistent between different AMPs, different experimental models and different bacterial strains (Figure 1.7D). Underwood et al. (2012) found that breast-fed premature rats and those fed with formula supplemented with *B. bifidum* showed reduced expression of the Paneth cell antimicrobials *sPla2* and *Lyz1* compared to those fed with milk formula only. Similarly, Pinto-Sánchez et al. (2017) found that oral administration of *B. infantis* Natren Life Start super strain (NLS-SS) resulted in reduced Paneth cell and macrophage counts as well as decreased HD-5 in duodenal biopsies from celiac patients. On the other hand, Natividad et al. (2013) noted that mono-colonisation with *B. breve* NCC2950, but not with *E. coli*, upregulated *RegIII γ* expression compared to germ free mice (in a MyD88 and Ticam1 dependent manner). Further, Lee et al. (2018) have shown that lactate derived from *Bifidobacterium* and *Lactobacillus* spp. increases the expansion of stem cells, Paneth cells and goblet cells in the small intestines of mice (and in organoids derived from mice), while increasing expression of *Lyz1* and Regenerating islet-derived protein 3 beta (*RegIII β*) and *RegIII γ* . Interestingly, it has been shown that human residential Bifidobacteria are tolerant to lysozyme, suggesting that increased AMP production by IECs is not likely an immune response specifically against bifidobacteria (Dan et al., 2018; Sakurai et al., 2017).

Despite conflicting evidence and a lack of identified mediating factors, previous findings support further investigation of the impact of bifidobacteria on AMP production by IECs.

1.5.2.4 *Stem cells and cell death*

Finally, bifidobacteria has been shown to affect the intestinal epithelial layer through promoting cell expansion and preventing cell death (Figure 1.7E). For example, pilin subunit TadE, of *B. breve* UCC2003 was shown to increase colonic epithelial cell proliferation in monoassociated mice compared to germ free mice (O'Connell Motherway et al., 2019). Further, lactate derived from *Bifidobacterium* and *Lactobacillus* spp. was shown to signal through G-protein-coupled receptor Gpr81 to elicit mouse intestinal stem cell proliferation, especially in new born mice (Lee et al., 2018). This effect required Wnt/ β -catenin signals of Paneth cells and intestinal stromal cells and protected against gut injury from combined radiation and chemotherapy treatment.

Additionally, it has been found that small organic molecules of *B. breve* prevent oxidant-induced IEC death through autophagy related genes (*Atg5* and *7*) and blockade of proteasomes (Figure 1.7F) (Inaba et al., 2016). Lin et al. (2014) also showed that bifidobacteria could activate autophagy in the IEC18 cell line.

1.5.3 The effect of bifidobacteria on gut immune cells

Clinical trials, *in vivo* experiments and *in vitro* experiments have indicated that bifidobacteria have immunomodulatory effects on their host (Ruiz et al., 2017). In addition to impacting intestinal epithelial cell functions, it has been shown that bifidobacteria can interact with immune cells to alter innate and adaptive immune processes (O'Neill et al., 2017; Ruiz et al., 2017). However, studies to date suggest that this interaction is complex and heavily context-dependent, with bifidobacteria exerting both pro- and anti-inflammatory effects. These effects have been studied in both phagocytes and lymphocytes, but here I focus only on macrophages as they are studied in Chapter 6 (O'Neill et al., 2017).

1.5.3.1 *The effect of bifidobacteria on macrophages*

A number of previous experiments have shown that different *Bifidobacterium* strains can affect macrophage function. He et al. (2002) showed that many strains of heat inactivated bifidobacteria, but particularly those associated with adult colonisation, can induce IL-12 and

Tumour necrosis factor α (TNF α) production from murine macrophage-like cell line, J774.1. They also demonstrated that most tested strains could induce anti-inflammatory cytokine IL-10 production. Using a different macrophage cell line, RAW 264.7, Lee et al. (2012) found that sonicated *Bifidobacterium adolescentis* SPM0308 and *Bifidobacterium longum* SPM1207, as well as their cell-free supernatant, activated production of TNF α and nitric oxide (required for cytotoxic activity of macrophages). Interestingly, the observed effects were reduced compared to stimulating the macrophage cell line with lipopolysaccharide (LPS) alone. A similar observation was noted by Okada et al. (2009) who found that *B. breve* and *B. longum* and *B. adolescentis* caused a reduced activation of IL-12p40, IL-1 β and TNF α expression levels (in a strain specific manner) compared to LPS alone in RAW 264.7 cells. Further, they identified that all three strains may act through reducing LPS-induced phosphorylation of NF- κ B inhibitor I κ B- α , and increasing expression of NF- κ B inhibitors, suppressor of cytokine signalling (SOCS) 1 and 3. Finally, to increase the relevance of the macrophages used, Mokrozub et al. (2015) cultivated macrophages from the peritoneal cavity of mice. They found that two different *B. adolescentis* strains could induce accumulation of reactive oxygen compounds and nitric oxide in macrophages, but neither significantly influenced production of IL-12 or interferon- γ (IFN γ).

Taken together, these studies show that bifidobacteria (in a strain specific manner) are capable of activating macrophages, albeit at a lesser magnitude than LPS. However, the molecular mechanisms of this interaction remain unclear.

1.5.4 *Bifidobacterium breve* UCC2003

To date, many of the *in vivo* studies of the role of *Bifidobacterium* in modulating immune cell and IEC function have focused on acute or chronic gut inflammation, often following pre-colonisation of the gut with *Bifidobacterium* strains (Din et al., 2020; Hsieh et al., 2015; Pinto-Sánchez et al., 2017; Schroeder et al., 2018; Srutkova et al., 2015; Yan et al., 2019). These studies suggest that initial priming by bifidobacteria during normal 'healthy' conditions may modulate subsequent protective responses. However, these studies were mostly performed in adult mice rather than during early developmental stages, where *Bifidobacterium* effects are expected to be most pronounced and long term. Indeed, where tested, the effect of bifidobacteria was more marked in new born mice than in adult mice (Lee et al., 2018).

Bifidobacterial species which are particularly abundant human infant intestines include *B. longum*, *B. breve*, and *B. bifidum* (Arboleya et al., 2016; Makino, 2018). Moreover, these species have the strongest evidence for IEC and gut barrier modulation. One such strain is *B. breve* UCC2003 which was originally isolated from a nursing stool (O’Connell Motherway et al., 2011). Various experiments have shown that *B. breve* UCC2003 plays an important role in host-commensal interactions through immune cell and IEC modulation as well as defence against pathogen infection. In addition to reducing epithelial cell shedding and increasing epithelial proliferation (as discussed previously; (Hughes et al., 2017; O’Connell Motherway et al., 2019)), *B. breve* UCC2003 has been shown to evade adaptive B cell host response, to protect mice against colonisation by gut pathogen *Citrobacter rodentium*, to protect *Caenorhabditis elegans* against *Salmonella* infection and to modulate gut microbiota through exopolysaccharide cross-feeding (Christiaen et al., 2014; Fanning et al., 2012b, 2012a; Püngel et al., 2020).

Whilst *B. breve* UCC2003 is a relevant bifidobacterial strain to study due to its observed health-benefits and presence in infant guts, it has further advantages which make it a practical model for developing methods and/or studying the effect of bifidobacteria in specific conditions. Specifically, *B. breve* UCC2003 has been shown to be highly efficient at colonising the murine gastrointestinal tract (small intestine, caecum and colon), and to stably persist at high levels for at least seven weeks (Cronin et al., 2008; O’Connell Motherway et al., 2011). Moreover, a Tn5 insertion library of nearly 20,000 transposon insertion mutants of *B. breve* UCC2003 has been developed using a random mutagenesis system, representing the first genome-wide random mutagenesis approach for bifidobacteria (Ruiz et al., 2013). Together with the fully sequenced genome, these factors make *B. breve* UCC2003 a particularly useful strain with which to study bifidobacterial-host interactions.

1.6 Organoids

In the past, a lack of *in vitro* systems to propagate cell lines of intestinal epithelium has hindered the study of mechanistic details relating to IEC function in healthy and diseased states (Chopra et al., 2010). However, advancements in understanding LGR5+ stem cells and their regulating pathways has led to the development of an *in vitro* culture system to grow three-dimensional (3D) intestinal epithelial organoids (Sato et al., 2009). Such systems have revolutionised the

study of IECs in development and disease and their interactions with immune, microbial and environmental signals (Fatehullah et al., 2016).

3D intestinal organoids, containing all the major cell types of the epithelium, are grown from LGR5+ stem cells – in isolation or as part of intestinal crypts. The stem cells/crypts are seeded onto a collagen- and laminin- rich matrix (Matrigel) and exposed to R-spondin, epidermal growth factor (EGF), Noggin and in the case of colonic organoids, Wnt (Sato and Clevers, 2013). R-spondin binds to LGR5 and acts as a Wnt signalling agonist to support crypt proliferation and EGF is additionally required for intestinal epithelial stemness. Noggin increases the number of crypts through blocking bone morphogenetic protein (BMP) signalling. In the presence of this minimal, essential stem cell maintenance factor cocktail, stem cells proliferate into a single layer of epithelial cells, forming a sealed round structure with crypt and villus architecture (Figure 1.8). The lumen of these 3D organoids represents the gut lumen, with basolateral cell surfaces facing outwards (Sato and Clevers, 2013; Sato et al., 2009). When replated each week, organoid cultures can be maintained for at least 1.5 years (Sato et al., 2009).

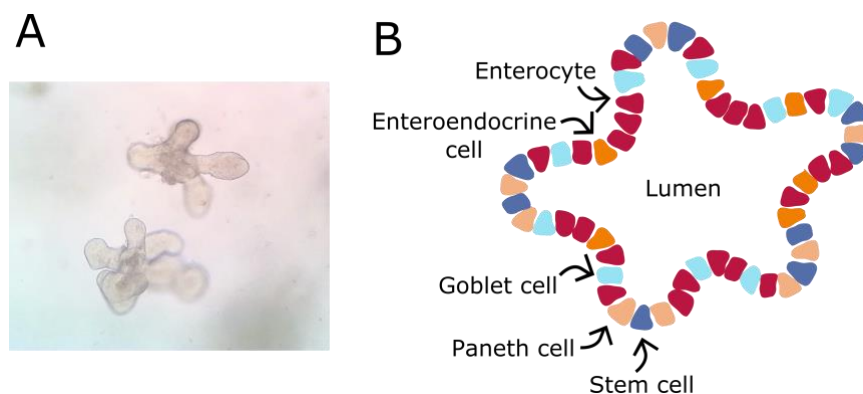


Figure 1.8. Small intestinal organoids. **A:** Mouse small intestinal organoids at day five of growth. Image generated by Isabelle Hautefort (Earlham Institute). **B:** Small intestinal organoid cross-section.

These long-term organoid cultures closely mimic *in vivo* conditions, and remain genetically and phenotypically stable over time, allowing great opportunity for experimental application (Bar-Ephraim et al., 2019). For example, intestinal organoids have been used to study epithelial receptors and cell function (Sodhi et al., 2012), to evaluate the effect of gene knock outs and mutations (Jones et al., 2019) and to study the interaction of the intestinal epithelium with bacteria and immune cells (Biton et al., 2018; Wilson et al., 2015; Zhang et al., 2014). In addition, intestinal organoids have been generated using patient-derived biopsies for patient- and

disease-specific models. One example is the use of primary intestinal organoids from cystic fibrosis patients to measure function of the cystic fibrosis transmembrane conductance regulator (CTFR) protein (Dekkers et al., 2013; de Poel et al., 2020). This model is currently used to test the efficacy of different treatments on cystic fibrosis patients, allowing access to drugs based on personalised medicine (Berkers et al., 2019).

Current limitations with the intestinal organoid model include cost, difficulties including non-epithelial cells in the design and loss of inflammatory phenotype when generated from inflamed tissue biopsies (Almeqdadi et al., 2019; Arnauts et al., 2019). In addition, exposing 3D organoids to apical-only signals (such as bacteria) is challenging due to the closed and difficult-to-access lumen environment. In response to these limitations, a number of organoid modifications and adaptations have been developed. To improve access to the apical side of the epithelium and decrease accumulation of dead cells, mucus and bacterial by-products, monolayer (2D) organoids cultures have been generated (Kozuka et al., 2017; Liu and Chen, 2018; Moon et al., 2014). These systems have been shown to form correctly polarised layers of epithelium with effective barrier functions, but are technically more challenging (Altay et al., 2019). Combined with micro-fluidics systems such as the HuMiX device, organoid monolayers can be co-cultured with microbes and basal immune populations to generate representative *in vitro* systems (Shah et al., 2016). More recently, a system has been developed to reverse the polarity of intestinal organoids (Co et al., 2019).

Furthermore, methods have been identified to increase the proportion and fidelity of secretory lineages in 3D organoids (Yin et al., 2014). Following an initial growth period using standard growth factors, the addition of a Notch inhibiting γ -secretase inhibitor, DAPT, and a Wnt inducing small molecule CHIR99021 causes small intestinal organoids to develop with a greater quantity of Paneth cells. Instead, by replacing CHIR99021 with Wnt inhibitor IWP-2, differentiation is skewed towards goblet cells and, to a lesser extent, enteroendocrine cells. Subsequent studies have shown that these drug treatments generate organoids which represent the *in vivo* condition (Luu et al., 2018; Mead et al., 2018). In particular, Mead et al. used transcriptomics, proteomics, cytometric and morphological characterisation to show that DAPT and CHIR99021 enrichment generates Paneth cells with greater fidelity and functional similarity to *in vivo* Paneth cells compared to conventional organoid Paneth cells (Mead et al., 2018). Furthermore, work from my group has shown that enteroids enriched for Paneth cells and

goblet cells recapitulate *in vivo* characteristics on the proteomics level (Jones et al., 2019; Luu et al., 2018). Moreover, in Jones et al. (2019), myself and colleagues showed that Paneth cell enriched enteroids contain greater Paneth cell marker gene expression than conventional enteroids. Furthermore, we showed that cell type enriched organoids are a useful tool for the investigation of health and disease related processes in specific intestinal cell types. Specifically, we generated Paneth cell enriched enteroids with epithelial-specific autophagy impairment. Using proteomics and transcriptomics data analysis, we identified several autophagy dependent cellular processes while mechanistically linking autophagy impairment to Paneth cell dysfunction, both of which are commonly observed in IBD (Jones et al., 2019). Therefore, whilst organoid enrichment methods do not present single cell type resolution, they provide useful tools to study Paneth cell and goblet cell populations in the context of the other major epithelial cell types (Mead and Karp, 2019).

Colleagues and I have employed organoid systems in Chapters 2 and 3 of this thesis to provide a controlled human IEC culture system and to enable detailed study of Paneth cell and goblet cell regulation.

1.7 Networks

In biology, networks are used to describe complex relationships between molecules, organisms, microorganisms, metabolic reactions, genetic interactions and many other entities. The study of networks in biology can be used for many purposes including to visualise systems, to trace signal flow, to understand functional relationships between entities and to study dynamics of systems (Han, 2008).

Molecular networks are a type of network used to capture direct interactions and/or functional associations between molecules (including genes, proteins and RNA) to better understand cellular mechanisms. In the graph representations of these networks, the entities (nodes) of the network represent the molecules and the connections between nodes (edges) represent a physical or functional interaction. These networks are vital for many aspects of cellular function including metabolism and transcriptional regulation. There are a number of features which these networks can have, for example they can be directed or undirected; indicating whether edges have a direction of interaction, and they can be signed, for example an inhibition or an

activation (Figure 1.9). Furthermore, these networks can exist in layered structures where different ‘versions’ of the networks represent different molecular levels e.g. RNA and protein, or instead different nodes can represent different types of molecule just within one network. The change in structure of networks over time and space is termed network dynamics (Han, 2008; Winterbach et al., 2013).

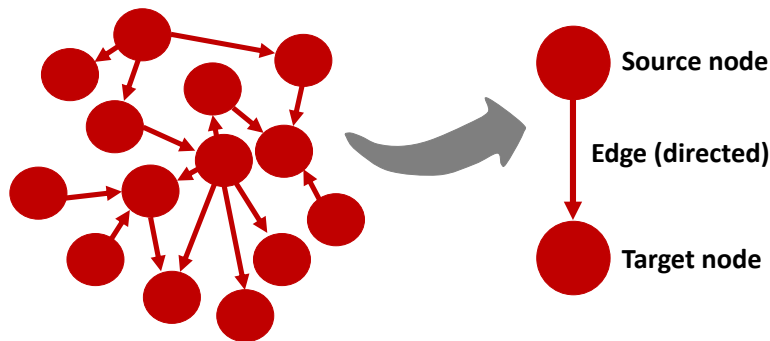


Figure 1.9. Graph diagram of a directed network.

Molecular interaction networks can be reconstructed using experimental or computational methods (Papin et al., 2005; Siahpirani and Roy, 2017). Experimental techniques include targeted and high-throughput methods to determine interacting molecules. Yeast-two-hybrid methods and tandem affinity purification are the primary technologies used to measure protein-protein interactions (PPIs) at scale (Brückner et al., 2009), whereas chromatin immunoprecipitation methods, ChIP-chip and ChIP-seq, are used to study interactions between DNA and proteins (histone and transcription factors, TFs) (Furey, 2012). Furthermore, interactions between microRNAs (miRNAs) and messenger RNAs (mRNA) and between long non coding RNAs (lncRNA) and miRNAs have been identified using HITS-CLIP methods (Chi et al., 2009). Experimentally determined molecular interactions are often stored in databases, which may be focused on a particular interaction type (e.g. TF – target interactions stored in TRRUST (Han et al., 2018)), a particular molecule, biological pathway or set of pathways (e.g. SignaLink2 (Fazekas et al., 2013)) or on a particular model organism (e.g. SignaFish (Csályi et al., 2016)). These databases are primarily second party, where multiple distinct experimental results are collected together, or third party, where multiple second party databases are combined. An example of a second party database is miRNA-lncRNA interaction collection LncBase (Paraskevopoulou et al., 2016), while molecular prior knowledge collection OmniPath and

transcription factor-target interaction collection DoRothEA are examples of third party databases (Garcia-Alonso et al., 2019; Türei et al., 2016).

The primary issue with all of these experimental data sources is that experimental methods and studies can be biased and do not uncover every possible interaction, leading to incomplete datasets (Kumar Bajpai et al., 2020). To avoid these issues, many different network inference algorithms have been developed that predict molecular interactions based on diverse data such as expression information, structural profiles and sequence homology (Chai et al., 2014; Chan et al., 2017; Huang et al., 2016; Siahpirani and Roy, 2017). However, network inference methods also have limitations due to under sampling and poor generalisation. They are often specific to one state and vary greatly based on the inference method and input datasets used. Furthermore, computationally predicted interactions often have little overlap with experimentally derived networks, which has led to development of ensemble learning approaches and the integration of computational and experimentally derived knowledge (Castro et al., 2019).

For the work outlined in this thesis, I have focused primarily on the interpretation and analysis of context-specific experimental data using experimentally determined molecular interactions. Such an approach permits the contextualisation of known molecular interactions in which we have greater confidence of existence, but I acknowledge probable incompleteness and a degree of experimental bias.

1.7.1 Regulatory networks

Regulatory networks are molecular networks detailing regulatory interactions within a system. Most frequently used are transcriptional regulatory networks which comprise interactions between TFs and their target genes (TGs). These networks are often presented as gene regulatory networks where both TFs and their target genes are represented as genes and the connections between them describe an indirect regulatory interaction (Winterbach et al., 2013). In addition to TF-TG interactions, regulatory networks may contain additional transcriptional and post-transcriptional regulatory interactions including: TF-lncRNA, TF-miRNA, miRNA-mRNA and lncRNA-miRNA interactions. MiRNAs and lncRNAs, together with TFs perform critical regulatory functions in maintaining intestinal homeostasis. Dysregulation of these functions has been associated with various gut pathologies (Chapman and Pekow, 2015; Mirza et al., 2015). For example a miRNA upregulated in the colon mucosa of CD patients, miR-106b, was found to

reduce expression of *ATG16L1*, causing a reduction in autophagy (Lu et al., 2014; Zhai et al., 2013). This in turn leads can lead to increased epithelial penetrance of CD-associated bacteria, indicating its role in CD pathogenesis. The study of regulatory interactions using networks can help to unravel the mechanisms through which these molecules act, which can in turn aid the study of disease and associated drug treatments. Regulatory actions of TFs and non-coding RNAs are described in the following sections.

1.7.1.1 Transcription factors

The best-studied eukaryotic gene regulatory mechanism is the control of gene expression by TFs. Here, TF proteins bind to *cis*-regulatory elements on the DNA via their DNA-binding region to promote or block RNA polymerase recruitment; depending on whether the TFs are acting as activators or repressors, respectively. Specifically, TFs can act to acetylate or deacetylate histone proteins, or to block or scaffold the binding of RNA polymerase to the DNA. Often these actions are carried out by TFs in cohort with coactivators and/or corepressors which are recruited to the TF-DNA complex (Figure 1.10). *Cis*-regulatory elements, termed enhancers, promoters or silencers, are usually present upstream of the regulated gene initiation site, but can also be found downstream, within gene introns or at a distance to the gene itself. The combined actions of multiple TFs, coactivators and corepressors at multiple *cis*-regulatory regions ultimately defines the transcription level of the gene in question (Campbell and Reece, 2008).

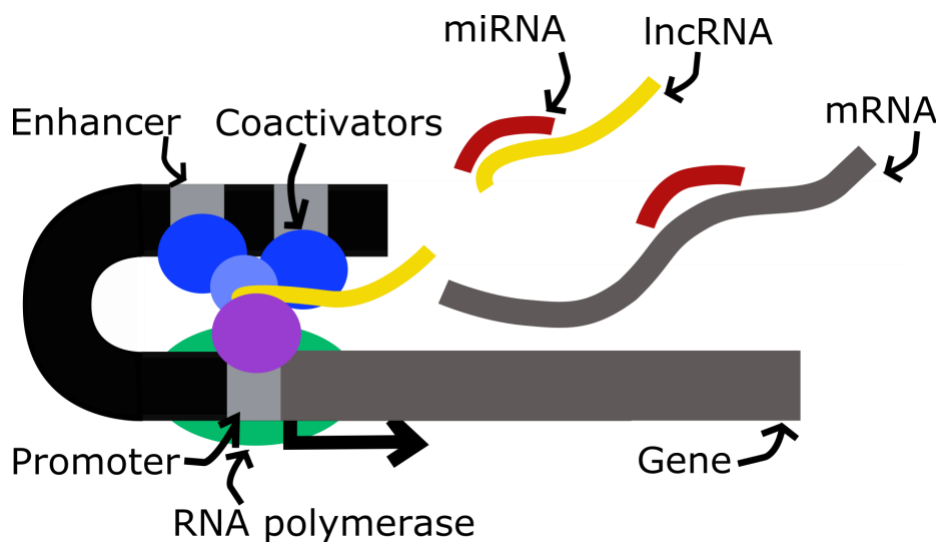


Figure 1.10. Transcriptional and post transcriptional regulation.

At any one time, only 10-50% of the genes of a typical human cell are expressed (Campbell and Reece, 2008; Marinov et al., 2014). This variation in expression allows the cells to alter their function in response to internal and external signals. TFs and histone modifying enzymes play an indispensable role in this process. As such, it has been shown that mutations in TF binding motifs of *cis*-regulatory elements are implicated in many diseases, including IBD (John et al., 2011; Landa et al., 2013; Pals et al., 2004).

1.7.1.2 *Non-coding RNAs*

In addition to regulation of expression, a number of post-transcriptional mechanisms exist to fine-tune levels of mRNAs. One example is miRNAs; ~22 nucleotide non-coding RNAs which post-transcriptionally regulate mRNA targets (Bartel, 2018). To date there are over 2,600 mature human miRNAs described in miRbase (version 22) – the primary archive of miRNA sequences and annotations (Kozomara et al., 2019). Canonical miRNAs are processed from stem-loop regions of longer RNAs called pri-miRNAs through the Drosha/Dicer pathway. Here, a heterotrimeric complex called microprocessor, containing a Drosha endonuclease and DGCR8 proteins, cleaves the pri-miRNA into pre-miRNAs (Nguyen et al., 2015). Once exported from the nucleus, pre-miRNAs are cleaved into 22 nucleotide duplexes by Dicer and then loaded (one or both strands) into the RNA-induced silencing complex (Bartel, 2018). The miRNAs are then guided to 3'UTRs (untranslated regions) of mRNAs where they bind to complementary (or nearly complimentary) sequences. The mRNA is subsequently degraded or translationally repressed (Chapman and Pekow, 2015). In the last decade many studies have shown that miRNA-associated regulation is associated with cell function in healthy and disease conditions (Ardekani and Naeini, 2010; Cao et al., 2017; Drury et al., 2017).

As well as miRNAs, the transcribed non-coding genome consists of thousands of other RNAs defined purely by their length: >200 nucleotides. These RNAs, termed lncRNAs, are detected across many cellular- and tissue-specific contexts, however their significance in cellular regulation is contentious. Evidence exists for a functional role of some lncRNAs, while many appear not to harbour any regulatory sequences or properties (Goff and Rinn, 2015; Long et al., 2017). Among those with identified regulatory functions, there are diverse mechanisms through which they can exert their effects. They can regulate gene expression through recruiting protein complexes to DNA or through inhibiting the binding of TFs to DNA. They can act as miRNA sponges, blocking the binding actions of miRNAs (Jalali et al., 2013; Paraskevopoulou and

Hatzigeorgiou, 2016) and they can interfere with post-transcriptional processing of mRNA, for example through binding to splicing factors and blocking alternative splicing (Bhat et al., 2016).

1.7.2 Signalling networks

Cells can receive and process signals from their environment, from surrounding cells and from different regions within themselves. External signals are primarily received through interactions with receptor proteins on the cell surface. In turn these receptors undergo a conformational change, initiating signal transduction through a chain of cellular events leading to a molecular response of the system to the initial trigger. Most often this signalling cascade involves the changes in the activity of effector enzymes which in turn regulate the function of other proteins through post-translational events. In the case of internal receptor binding, for example in steroid hormone systems, the ligand-receptor complex often acts more directly on gene levels by acting as a modulator of transcription (Buchanan et al., 2010). These signalling pathways are high complex systems of interacting molecules which can also form larger networks of interacting pathways. Therefore, signalling pathways are well represented through network approaches, which provide a clear data storage, visualisation and analysis method. Given the nature of signalling pathways, they are often represented by a type of protein-protein interaction network, where links are directed according to the flow of molecular signals (Winterbach et al., 2013).

1.7.3 Host-microbe protein-protein interactions

In addition to the aforementioned interaction types, networks can be used to study interactions between a host cell and a microbe. Such interactions can occur via various molecules ranging from proteins and metabolites to lipopolysaccharide (LPS). For example, for bifidobacteria, identified molecular effectors include, but are not limited to:

- Type IVb tight adherence (Tad) pili, which has been shown to promote colonic epithelial proliferation (O'Connell Motherway et al., 2011, 2019)
- The EPS capsule which can, for example, repress local Th17 responses and reduce pro-inflammatory cytokine production (Fanning et al., 2012b)
- The serine protease inhibitor (serpin) protein which inhibits pancreatic elastase and neutrophil elastase, potentially protecting the host against overactivation of neutrophils, and in turn the bacteria against phagocytosis (Ivanov et al., 2006)

Protein-protein interactions between host and microbe represent one of the major types of communication. Here proteins physically interact, often resulting in activation or inhibition of the host downstream signalling pathways. Such interactions are often divided into two major types: domain-domain interactions and domain-motif interactions. Domains are independent folding subunits of proteins with stable and distinct tertiary structures. Domains are not protein-specific, often occurring in many proteins of many different species. Moreover, a single protein can contain multiple different domains. Domains usually have specific functional roles, such as calcium-binding, and can be readily identified from nucleotide or amino-acid sequences (Bagowski et al., 2010). Domain-domain interactions occur when a domain of one protein physically interacts with a domain of another protein, resulting in one protein exerting its effect on the other. Experimentally, domain-domain interactions are primarily inferred using high-resolution three-dimensional structures (Raghavachari et al., 2008). Moreover, many domain-domain interactions have been predicted computationally using diverse feature sets and methods – such as sequence co-evolution, phylogenetic profiling, probabilistic frameworks and machine learning approaches (Yellaboina et al., 2011). Databases such as DOMINE exist to collate known and predicted domain-domain interactions (Raghavachari et al., 2008; Yellaboina et al., 2011).

Domain-motif interactions, on the other hand, involve the binding of protein domains to protein motifs, resulting in the domain-containing protein exerting its effect on the motif-containing protein. These interactions are mediated by 3-10 amino acid sequences called short linear motifs (SLiMs) which play a major role in cellular processes, such as signal transduction, through transient interactions with protein domains (Akiva et al., 2012; Brito and Pinney, 2017). These motifs occur in intrinsically disordered regions of proteins or in exposed flexible loops within folded domains making them accessible for binding to domains. Like domains, many protein motifs and domain-motif interactions have been previously identified by diverse small-scale experiments and by computational predictions (Gibson et al., 2015). Once identified, motifs can be typified by their sequence making them identifiable within other genomes using computational methods. Furthermore, databases of curated domain-motif interactions, such as the Eukaryotic Linear Motif (ELM) database, enable prediction of interactions between proteins of interest based on previously identified domain-motif interactions (Dinkel et al., 2016; Korcsmaros et al., 2013; Puntervoll et al., 2003). In the Chapter 5, I have applied such an

approach to build a bacterial-macrophage protein interaction network for prediction of possible bifidobacterial effector proteins which can affect macrophage activation.

1.7.4 Network contextualisation using 'omics data

As discussed above, network contextualisation is one method to reconstruct biological networks based on the integration of context specific 'omics data with experimentally determined molecular interactions (often called *a priori* interactions or prior knowledge network) (Dugourd and Saez-Rodriguez, 2019). The specific methods used in this approach vary depending on the available data and the biological question of interest: ranging from simple data overlaps to ensembles of mathematically complex models. Comprehensive networks of intra and inter-cellular interactions can be reconstructed using multi-omics datasets consisting of transcriptomic, phosphoproteomic and metabolomic information from the same biological context (Dugourd et al., 2020). However, the availability of such datasets is usually limited; due to cost and accessibility, transcriptomics is often the only obtainable 'omics data. Therefore, the following tools and methods (and the research of this thesis), will cover only transcriptomics data approaches.

Given that the activities of signalling pathways only partially correlate with levels of gene expression (Vogel and Marcotte, 2012), transcriptional level data is best used in studies of transcriptional regulation. For example, superimposing *a priori* regulatory interactions with differentially expressed genes of interest can identify possible regulators. Subsequently, regulators can be filtered for relevance using different approaches, such as:

- Identification of master or hub regulators based on the number of target genes.
- Hypergeometric significance tests to identify regulators whose targeted genes exhibit significantly large levels of differential expression, such as that implemented in the Cytoscape CHAT tool (Muetze et al., 2016).
- The VIPER approach which uses analytic rank-based enrichment analysis to compute changes in transcription of specific regulons (groups of similarly regulated genes) when projected on a rank-sorted gene expression signature. In addition, this approach integrates a metric based on the correlation of expression between a regulator and its target gene and measures of pleiotropy (Alvarez et al., 2016).

On the other hand, methods have been developed to predict activity of signalling pathways based on gene expression data, without directly inferring protein activity from gene expression. For example, PROGENy, developed by Schubert *et al.* uses a footprint based approach to predict pathway activity based on pathway responsive genes, which are obtained using a large compendium of perturbation experiments (Schubert et al., 2018). However, this approach is limited and biased by the perturbation datasets used, and may not accurately reflect specific contexts such as poorly studied tissue types.

Finally, causal networks can be generated by combining signalling pathway and regulatory interactions. Such networks can explain the flow of signal from a cellular perturbation, such as the recognition of a signalling molecule at the cell surface, through signalling pathways to transcription factors which regulate the expression of genes affected by the perturbation. Such networks can be reconstructed by identifying possible PPI paths (using a prior knowledge network) connecting an initial perturbation (e.g. a receptor) to transcription factors identified using approaches such as those previously discussed. Advancements to this method can utilise causal reasoning and/or diffusion algorithms to identify most likely signalling paths. For example, a recently published workflow by Liu et al. (2019) called CARNIVAL, can reconstruct causal networks with or without an upstream perturbation using VIPER and PROGENy combined with an integer linear programming optimization problem. Here the most optimal PPI paths are identified using interaction direction and sign alongside perturbation, pathway and TF constraints.

1.7.5 Network applications

The ultimate aim of a systems biologist is to understand a whole system using a unified framework – and networks are a key tool to achieve this. In addition to functioning as a data storage and visualisation method, analysis of molecular interaction networks can aid the understanding of cellular processes and functional organisation, can be used to predict key regulators and novel drug targets and can aid in annotating molecular functions (Miryala et al., 2018). Methods to analyse and interpret networks include topological approaches and functional enrichment. Topological approaches, which study the arrangement and structure of networks, can be used to both describe properties of a biological network and as a predictive tool (Winterbach et al., 2013). For example, centrality is a group of graph metrics which score nodes based on their importance in the network. The simplest is degree centrality, which states

that nodes with high degree (many connecting edges) are more important than those with low degree. Alternatively, betweenness centrality calculates the fraction of shortest paths passing through a node when every pair of nodes is connected. A higher betweenness centrality metric indicates that a node is more important for information flow in the network. Alternatively, molecular networks can be studied in terms of their modularity. For example, clustering approaches can be used to identify groups of nodes, termed modules or clusters, which are more densely connected than the rest of the network. In molecular interaction networks, identified modules often represent biological pathways or functions, aiding functional interpretation and prediction of key molecules within the network. Unfortunately, module detection is a complex problem. Whilst multiple different algorithms have been developed, results between them are not always comparable and selecting the ideal algorithm for a particular problem is especially challenging (Tripathi et al., 2016). Where module detection is carried out in this thesis, I have used the seed-based clustering tool Molecular Complex Detection (MCODE) which was designed specifically for protein interaction networks to detect protein complexes and functional modules (Bader and Hogue, 2003; Kaalia and Rajapakse, 2019).

Overall, a number of different network approaches have been used throughout this thesis to reconstruct, visualise, analyse and interpret biological data, each selected based on the data type and research question.

1.8 Primary research aims

The four primary research aims of this thesis are as follows:

1. Develop workflows and processes to analyse intracellular regulation in a cell type-specific manner to gain biological insights.
2. Apply these workflows to increase our understanding of how cytokines alter the regulation of epithelial cells.
3. Apply these workflows to increase our understanding of how *Bifidobacterium* alters the regulation of epithelial cells using bulk transcriptomics data.
4. Apply these workflows to increase our understanding of how *Bifidobacterium* alters the regulation of epithelial cells using cell type-specific transcriptomics data.
5. Study the interactions of *Bifidobacterium* with immune cell populations.

1.9 Structure of the thesis

This thesis is organised into 7 chapters:

Chapter 1 – An introductory chapter presenting the research background and summary of the aims of the thesis.

Chapter 2 – Presents an interdisciplinary workflow developed by myself and colleagues to study the regulatory landscape of small intestinal epithelial cells. The workflow is applied to investigate the effect of small molecule treatments on skewing differentiation of small intestinal organoids (enteroids) and to predict key regulators of Paneth cells and goblet cells.

Chapter 3 – Here, I study the effect of cytokines responsible for canonical mucosal immune responses on colonic organoids using causal networks. Further, myself and collaborators evaluated cytokine regulated transcription activity in biopsies from IBD patients.

Chapter 4 – Twinned with Chapter 5, this chapter contains the first of two distinct studies focusing on the impact of bifidobacteria on small intestinal epithelial cells. This chapter specifically focuses on the mouse neonatal epithelium and does not contain cell type-specific data.

Chapter 5 – Twinned with Chapter 4, this chapter contains the second of two distinct studies focusing on the impact of bifidobacteria on small intestinal epithelial cells. This chapter extends the previous work by focusing on stem cells and Paneth cells in SPF and GF mice who have recently weaned from their mother's breast milk.

Chapter 6 – This chapter describes a study of the impact of *Bifidobacterium* on macrophage activation, focusing primarily on inter-cellular interactions with the aim to identify the effector molecule/s of bifidobacteria.

Chapter 7 – This is the final chapter discussing future perspectives, impact and conclusions of the thesis.

Chapter 2: The regulatory landscape of small intestinal epithelial cells

2.1 Introduction

Gut barrier integrity is critically important for intestinal homeostasis and efficient nutrient absorption (Zhang et al., 2015). Disruption of the epithelial barrier along with dysregulated immune responses are some of the underlying reasons behind the development of inflammatory gut conditions such as inflammatory bowel disease (IBD) (Mokry et al., 2014). Therefore, a greater understanding of the functions of intestinal cells and their role in regulatory signalling will further our understanding of gut dysbioses, including IBD aetiology.

Whilst primary IECs all originate from Leucine-rich repeat-containing G-protein coupled receptor 5 (Lgr5)⁺ stem cells, differentiation results in differences in gene expression, signalling and regulatory wiring (Crosnier et al., 2006; Vanuytsel et al., 2013). These differences can result in altered phenotypic functions, responses to stress and susceptibilities to specific dysregulations. Thus, uncovering patterns and mechanisms at a cell type-specific level is crucial to uncover the role of the intestinal epithelium in homeostasis and disease. However, previously, many disease-focused studies have used biopsy samples to produce -omics read-outs from intestinal epithelial cells (IECs) (Balfe et al., 2018; Mirza et al., 2015). Due to the cellular heterogeneity of the biopsies, these readouts represent a combination of different cell types (including semi-differentiated cells), which can result in obscuration of signals when cell types are not acting in cohort.

On the other hand, recent studies have employed single cell transcriptomics sequencing of tissue samples to characterise the proportion and signatures of different epithelial cell types in the intestines of healthy and IBD patients (Haber et al., 2017; Parikh et al., 2019; Smillie et al., 2019). However, to provide deeper insights into the role of specific cell populations (such as Paneth cells and goblet cells) in IBD, *in vitro* models are required for in-depth testing and manipulation. Such models can be used to study specific mechanisms of action, host-microbe interactions, intercellular communication, patient specific therapeutic responses and to develop new diagnostic approaches. Due to ease of manipulation, observation and analysis, organoid models, including small intestinal models (enteroids), are increasingly used in the IBD field

(Aberle et al., 2018; Lindeboom et al., 2018; Noben et al., 2017). Therefore, the development of experimental and computational methods to improve the utility of organoids for detailed scientific investigation are highly valued, despite the growth of single cell technologies. For example, small molecule treatments have been developed that skew the differentiation of enteroids towards Paneth cell or goblet cell lineages, improving representation of these cells within the enteroid cell population (as described in the General Introduction section 1.6) (Farin et al., 2012; Yin et al., 2014).

Nevertheless, the effect of Paneth cell and goblet cell enrichment of enteroids on key regulatory landscapes has not been extensively characterised and few computational methods have been applied to study cell type-specific regulation in an organoid model (Mead et al., 2018; Qin et al., 2020). In this chapter, I use transcriptomics data combined with network methods to characterise the effect of small-molecule skewing on enteroids and to test the ability of this kind of approach for studying cell type-specific regulatory landscapes using organoids. Moreover, myself and colleagues developed an interdisciplinary workflow to investigate the regulatory landscape of Paneth cells and goblet cells (including transcriptional and post-transcriptional regulation) by comparing cell type enriched enteroids to control enteroids. The future utility of this approach was evidenced by predicting key regulators of Paneth cells and goblet cells and by exploring the relevance of the generated regulatory networks to the study of inflammatory bowel disease (IBD).

The study design and analysis workflow are described in Figure 2.1. Specifically, we used small intestinal crypts from healthy adult mice to grow three different types of enteroid cultures: conventionally differentiated, Paneth cell enriched and goblet cell enriched (based on the protocols in Yin et al. (2014)). RNA sequencing was carried out on each type of enteroid to quantify messenger RNA (mRNA), microRNA (miRNA) and long non-coding RNA (lncRNA) signatures. Differentially expressed genes (DEGs) were determined by comparing the cell type enriched enteroids to the conventionally differentiated organoids – generating a list of genes significantly more or less expressed as a result of Paneth cell enrichment, and a similar list as a result of goblet cell enrichment. I used published molecular interaction datasets to reconstruct regulatory interaction networks (one for each cell type) which link these DEGs by possible regulatory connections. Specifically, the nodes of the interaction networks represent the DEGs and the edges represent regulatory connections (molecular interactions) between the nodes inferred from databases of transcriptional and post-transcriptional interactions. Subsequently, I

carried out a number of different analyses on the networks. I clustered the networks to identify highly connected nodes which specific functions. I incorporated known Paneth cell and goblet cell marker genes to predict master regulators of Paneth cell and goblet cell differentiation and/or maintenance. Furthermore, I highlighted varying downstream actions of shared regulators between the cell types. This phenomenon, called regulatory rewiring, highlights the importance of changes in regulatory connections in the function and differentiation of specific cell types. Finally, I identified and analysed Crohn's disease (CD) and ulcerative colitis (UC) associated genes within the networks. Taken together, we show that cell type enriched enteroids combined with the presented network biology workflow have potential for application to the study of epithelial dysfunction and mechanisms of action of multifactorial diseases such as IBD in specific intestinal cell types.

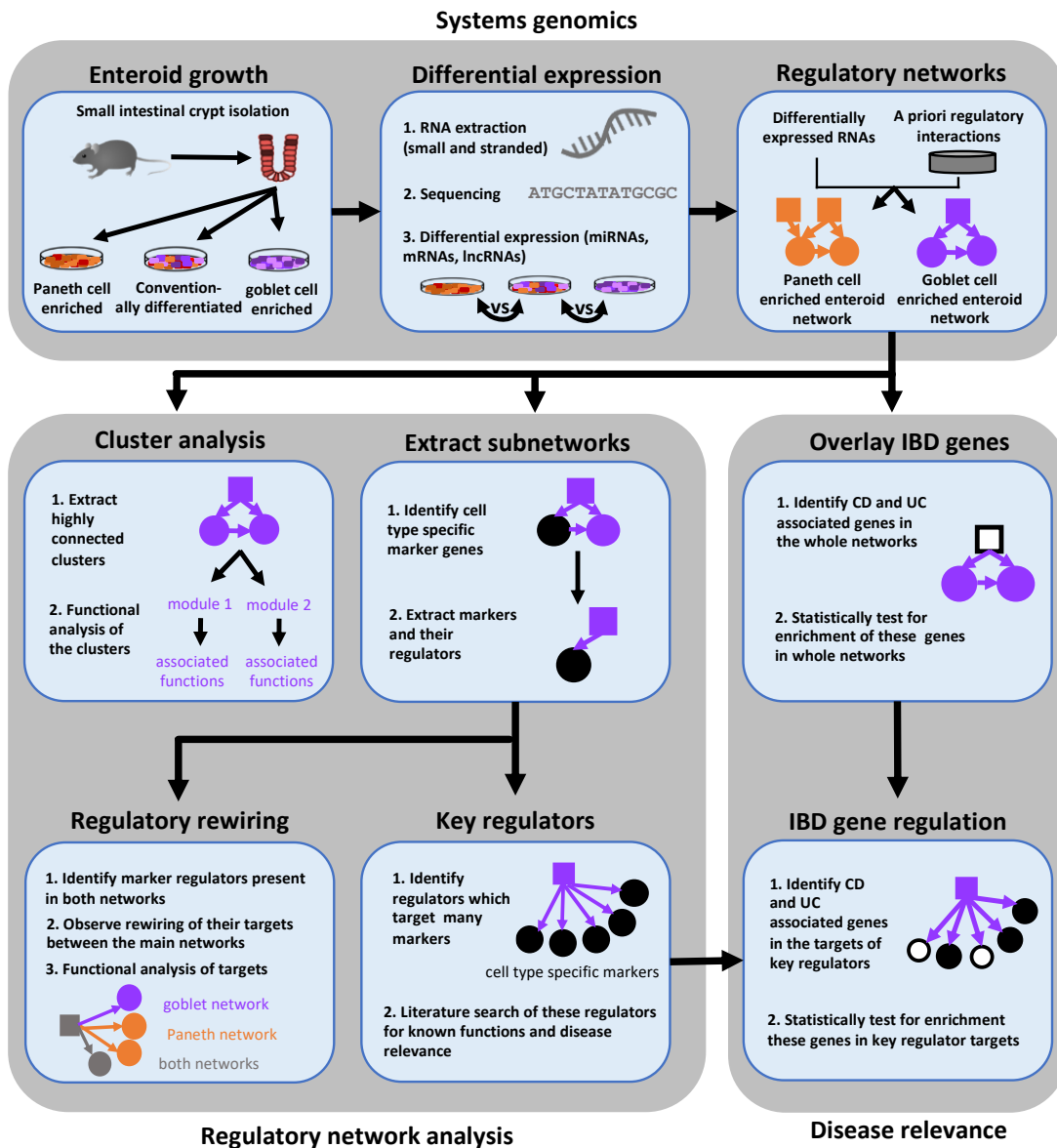


Figure 2.1. Schematic overview of study design and analysis workflow in Chapter 2. PCeE/GCeE network - Paneth cell enriched enteroid / goblet cell enriched enteroid network; TF - transcription factor; lncRNA - long non-coding RNA; miRNA - microRNA; mRNA - messenger RNA; UC - ulcerative colitis; CD - Crohn’s disease. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

The contents of this chapter are primarily based on (verbatim) the peer-reviewed article published in *Molecular Omics* that I am first author (Treveil et al., 2020). The published article is reproduced in Appendix 6.

2.2 Aims

The aims for this project were as follows:

- Develop a multidisciplinary pipeline to study cell type-specific regulatory landscapes using enteroids and RNA sequencing.
- Assess the value of cell type enriched organoids for studying cell type-specific regulatory landscapes.
- Characterise the effect of Paneth cell and goblet cell enrichment of enteroids on key regulatory landscapes.
- Predict key regulators of Paneth cells and goblet cells.
- Investigate relevance of generated networks to the study of inflammatory bowel disease (IBD).

2.3 Methods

All organoid work and RNA extraction was carried out by Zoe Matthews (UEA, Norwich Medical School), Emily Jones and Isabelle Hautefort (from our group). Next-generation sequencing and library construction was delivered via the BBSRC National Capability in Genomics and Single Cell (BB/CCG1720/1) at the Earlham Institute by the Genomics Pipelines Group. Initial data processing to obtain differentially expressed transcripts and collation of molecular interaction resources was carried out by Tomasz Wrzesinski (Haerty group, Earlham Institute, EI) and Padhmanand Sudhakar from our group. All other computational analysis and interpretation was carried out by myself, including network reconstruction, marker gene analysis and functional enrichment.

2.3.1 Small intestinal organoid growth

C57BL/6J mice of both sexes were used for enteroid generation as described previously (Jones et al., 2019; Sato and Clevers, 2013; Sato et al., 2009), from three separate animals for each condition. Briefly, 5mm pieces of small intestine were washed in Ethylenediaminetetraacetic acid (EDTA) and shaken in phosphate buffered saline (PBS) until five fractions had been generated. Crypt suspensions from the fractions were passed through a 70 μ m filter to remove any villus fragments, centrifuged at 300 \times g for 5 minutes before pellets were resuspended in 200 μ l phenol-red free Matrigel (Corning), seeded in 24-well plates and incubated at 37°C for 20 minutes to allow Matrigel to polymerise. Enteroid media containing Epidermal growth factor (EGF), Noggin and R-spondin (ENR media) was then overlaid. On days two, five and seven post-crypt isolation, additional factors were added to the ENR media to enrich specific cell types by chemically inducing differentiation: 3 μ M glycogen synthase kinase 3 β (GSK3 β) inhibitor CHIR99021 (Tocris) and 10 μ M Notch inhibitor DAPT (Tocris) [Paneth cells]; 2 μ M Wnt pathway inhibitor IWP-2 (Tocris) and 10 μ M DAPT [goblet and enteroendocrine cells] (Yin et al., 2014). On day eight post-crypt isolation, enteroids were fixed with 4% paraformaldehyde (PFA; Sigma-Aldrich), permeabilized with 0.1% Triton X-100 (Sigma-Aldrich) and incubated in blocking buffer containing 10% goat serum (Sigma-Aldrich). Immunostaining was performed overnight using primary antibodies for E-cadherin (BD Transduction Laboratories), Mucin-2 (Muc2) (Santa Cruz) and lysozyme (Lyz1) (Dako), followed by Alexa Fluor-488 and -594 conjugated secondary antibodies (ThermoFisher Scientific). DNA was stained with 4',6-diamidino-2-phenylindole

(DAPI) (Molecular Probes). Images were acquired using a fluorescence microscope (Axioimager.M2) and analysed using ImageJ/FIJI V1.51.

2.3.2 RNA sequencing

RNA was extracted from enteroids on day eight post crypt-isolation using miRCURY RNA Isolation Tissue Kit (Exiqon, 300115). Stranded RNA Libraries were constructed using the NEXTflex™ Rapid Directional RNA-Seq Kit (PerkinElmer, 5138-07) using the polyA pull down beads from Illumina TruSeq RNA v2 library construction kit (Illumina, RS-122-2001). Small RNA libraries were made using the TruSeq Small RNA Library Prep Kits (Illumina, 15004197). Stranded RNA was sequenced on the Illumina HiSeq2000 instrument to obtain 100 base paired-end reads. Small RNA was sequenced on the Illumina HiSeq2500 instrument to obtain 50 base paired-end reads.

2.3.3 Differentially expressed transcripts

The quality of stranded reads was assessed by FastQC software (v0.11.4) (Andrews, 2010). Reads were aligned using HISAT (v2.0.5) (Kim et al., 2015a) and a reference-based *de novo* transcriptome assembly was carried out for each biological repeat and merged together using StringTie (v1.3.2). Coding potential of each novel transcript was determined with CPC (v0.9.2) and CPAT (v1.2.2) (Kong et al., 2007; Wang et al., 2013b). From the novel transcripts, only non-coding transcripts (as predicted by both tools) were included in final GTF file. Gene and transcript abundances were estimated with Kallisto (v0.43.0) (Bray et al., 2016). Sleuth (v0.28.1) R library was used to perform differential gene expression, comparing Paneth cell enriched enteroids (PCeEs) to conventionally differentiated enteroids (CDEs), and goblet cell enriched enteroids (GCeEs) to CDEs (Pimentel et al., 2017).

The small RNA reads were analysed using the sRNAbench tool within the sRNAtoolbox suite (Rueda et al., 2015). Trimmed and length filtered reads were mapped to mature miRBase miRNAs (v21) (Kozomara and Griffiths-Jones, 2014) in addition to the annotated version of the mouse genome (mm10). Normalised read-counts from the corresponding cell type enriched enteroids were compared against the CDEs to identify differentially expressed miRNAs in a pair-wise manner using edgeR (Robinson et al., 2010).

mRNAs, lncRNAs and miRNAs with an absolute log₂ fold change ≥ 1 and q value ≤ 0.05 were considered to be differentially expressed. Differentially expressed genes were grouped by their presence in the PCeE dataset, the GCeE dataset or in both. Each group of differentially expressed genes was tested for functional enrichment (hypergeometric model, q value ≤ 0.1) based on Reactome and KEGG annotations using the *ReactomePA* R package (Fabregat et al., 2018a; Kanehisa et al., 2017; Ogata et al., 1999; Yu and He, 2016) following conversion from mouse to human identifiers using Inparanoid (v8) (O'Brien et al., 2005; Sonnhammer and Östlund, 2015).

2.3.4 Enrichment of marker genes

Cell type-specific signature genes were obtained from the droplet-based and the plate-based data of a mouse single cell sequencing survey (Haber et al., 2017). Gene symbols were converted to Ensembl gene IDs using bioDBnet db2db (Mudunuri et al., 2009).

Hypergeometric distribution testing was carried out using a bespoke R script to measure enrichment of cell type-specific marker genes in the differentially upregulated gene sets. To standardise the universal dataset, only markers which are present in the output of the Wald test (genes with variance greater than zero among samples) were used. All genes present in the output of the Wald test were used as the background. Similarly, to enable fair comparisons, only differentially expressed protein coding genes and documented lncRNAs were used from the DEG lists, as was surveyed in the cell type-specific marker paper. Bonferroni correction was applied and significance scores were calculated using $-\log_{10}(\text{adjusted p value})$ (Bland and Altman, 1995).

2.3.5 Reconstruction of regulatory networks

A priori mice regulatory networks containing directed regulatory layers were retrieved from multiple databases as described in Table 2.1. Only miRNA-mRNA and lncRNA-miRNA interactions determined using HITS-CLIP (Chi et al., 2009) experiments were considered. Bedtools (Quinlan and Hall, 2010) was used for the custom analyses to look for overlaps between coordinates. All the nodes in the collected interactions were represented by their Ensembl gene IDs for standardization.

To generate PCeE and GCeE regulatory networks, interactions in this collated universal network were filtered using the transcriptomics data (Figure 2.1). The assumption was made that if both nodes of a particular interaction were expressed (transcripts per million > 0 in at least one

replicate) in the RNAseq data, the interaction is possible. Furthermore, to filter for the interactions of prime interest, only nodes which were differentially expressed (PCeE vs CDE, GCeE vs CDE) and their associated interactors were included in the regulatory networks.

Interaction type	Source(s)	# Unique interactions	Quality control criteria
TF-TG (TFs regulating target genes)	TRRUST v2 (Han et al., 2015, 2018) GTRD (Yevshin et al., 2017) ORegAnno v3.0 (Lesurf et al., 2016)	1066383	<ul style="list-style-type: none"> ChIP-Seq peaks should not overlap any gene annotation; if peak on + strand, only the first gene downstream to the gene or if peak on - strand, only first gene upstream to the peak is considered. Genes attributed to the transcription factor which lie within a 10kb window on either side of the ChIP-seq peak (ORegAnno) or meta-cluster (in the case of GTRD).
TF-lncRNA (TFs regulating lncRNAs)	GTRD	159055	<ul style="list-style-type: none"> ChIP-Seq peaks should not overlap any gene annotation; if peak on + strand, only the first gene downstream to the gene or if peak on - strand, only first gene upstream to the peak is considered. Genes attributed to the transcription factor which lie within a 10kb window on either side of the meta-cluster. Only if the first annotation feature within a 10kb genomic window downstream to the ChIP-seq peak / meta-cluster was designated as an intergenic lncRNA, a regulatory interaction between the TF and the lncRNA was assigned - to avoid assigning false regulatory interactions due to the high number of instances where the lncRNAs overlap with protein-coding genes.
miRNA-mRNA (miRNAs regulating mRNAs)	TarBase v7.0 (Vlachos et al., 2015)	141892	<ul style="list-style-type: none"> Only HITS-CLIP based experimental evidence considered. Co-expression based inferences not considered.
TF-miRNA (TFs regulating miRNAs)	TransmiR v1.2 (Wang et al., 2010) TRRUST v2 GTRD	9204	<ul style="list-style-type: none"> ChIP-Seq peaks should not overlap any gene annotation; if peak on + strand, only the first gene downstream to the gene or if peak on - strand, only first gene upstream to the peak is considered. Co-expression based inferences not considered
lncRNA – miRNA (lncRNAs regulating miRNAs)	lncBase2 (Paraskevopoulou et al., 2016)	6637	<ul style="list-style-type: none"> Only HITS-CLIP based experimental evidence considered. Co-expression based inferences not considered.

Table 2.1. A summary of the physical interactions compiled to generate the universal network. Table reproduced from Treveil et al. (2020) under the Creative Commons BY licence. The 10kb window was chosen based on (Maclsaac et al., 2010).

Clusters of highly interconnected regions within the PCeE and GCeE regulatory networks were identified using the MCODE plugin within Cytoscape with default parameters (Bader and Hogue, 2003; Shannon et al., 2003). The nodes of each cluster were tested for functional enrichment (hypergeometric model, q value ≤ 0.05) based on Reactome annotations using the *ReactomePA* R package (Fabregat et al., 2018a; Kanehisa et al., 2017; Ogata et al., 1999; Yu and He, 2016) following conversion from mouse to human identifiers using Inparanoid (v8) (O'Brien et al., 2005; Sonnhammer and Östlund, 2015). Cases where the number of nodes associated with a pathway < 5 were considered not significant regardless of the q value. The top 5 significant Reactome pathways associated with each cluster were visualised using a heatplot generated in R. More than 5 pathways were visualised where multiple Reactome pathways had equal q values.

2.3.6 Paneth cell and goblet cell regulator prediction

To identify potential master regulators of the Paneth cell and the goblet cell types, the upstream regulators of cell type-specific markers (from (Haber et al., 2017)) were investigated. To do this, all markers were mapped to the relevant networks then subnetworks were extracted consisting of markers and their regulators.

2.3.7 Regulatory rewiring analysis

To calculate rewiring scores for regulators, sub-networks were extracted (from the PCeE and GCeE regulatory networks) containing just the regulator of interest and its downstream targets. For each regulator of interest, the subnetworks from the PCeE and GCeE networks were compared using the Cytoscape app DyNet (Goenawan et al., 2016; Shannon et al., 2003). The degree corrected D_n score was extracted for each regulator and used to quantify rewiring of the regulator's downstream targets between the PCeE and GCeE regulatory networks. Functional analysis was carried out on the targets of the top five most rewired regulators. For each regulator, the targets were classified based on whether they are present in only the PCeE network, only the GCeE network or in both networks. Each group of targets was tested for functional enrichment (hypergeometric model, q value ≤ 0.1) based on Reactome and KEGG annotations using the *ReactomePA* and *ClusterProfiler* R packages (Fabregat et al., 2018a; Kanehisa et al., 2017; Ogata et al., 1999; Yu and He, 2016) following conversion from mouse to human identifiers using Inparanoid (v8) (O'Brien et al., 2005; Sonnhammer and Östlund, 2015).

2.3.8 Evaluating disease relevance

Genes associated with UC and CD based on single nucleotide polymorphisms were obtained from two studies (Farh et al., 2015; Jostins et al., 2012). Additionally, the top 100 differentially expressed genes were obtained from goblet cell analysis of inflamed UC vs healthy human colonic tissue from (Smillie et al., 2019). Genes were converted to Mouse Ensembl identifiers using Inparanoid (v8) and bioDBnet db2db (Mudunuri et al., 2009; O'Brien et al., 2005; Sonnhammer and Östlund, 2015). Additionally, to enable hypergeometric significant testing with the universal network as the background, only UC and CD genes present in the universal network are included in the analyses. eQTL datasets for CD were retrieved from (Di Narzo et al., 2016) while the list of targets related to drug-interactions was downloaded from (Cotto et al., 2018).

2.3.9 Quantification and statistical analysis

Statistical parameters including the exact value of n and statistical significance are reported in the Figures and Figure Legends. n represents the number of enteroid biological replicates generated. Where relevant, data is judged to be statistically significant when Bonferroni adjusted p value ≤ 0.01 . Genes with absolute log₂ fold change of ≥ 1 and q value ≤ 0.05 were considered to be differentially expressed. Based on principal component analysis of transcript expression, one biological replicate from the Paneth cell enriched enteroids was identified as an outlier and removed (Figure 2.2). Where stated, the hypergeometric distribution model was used to calculate significance using R.

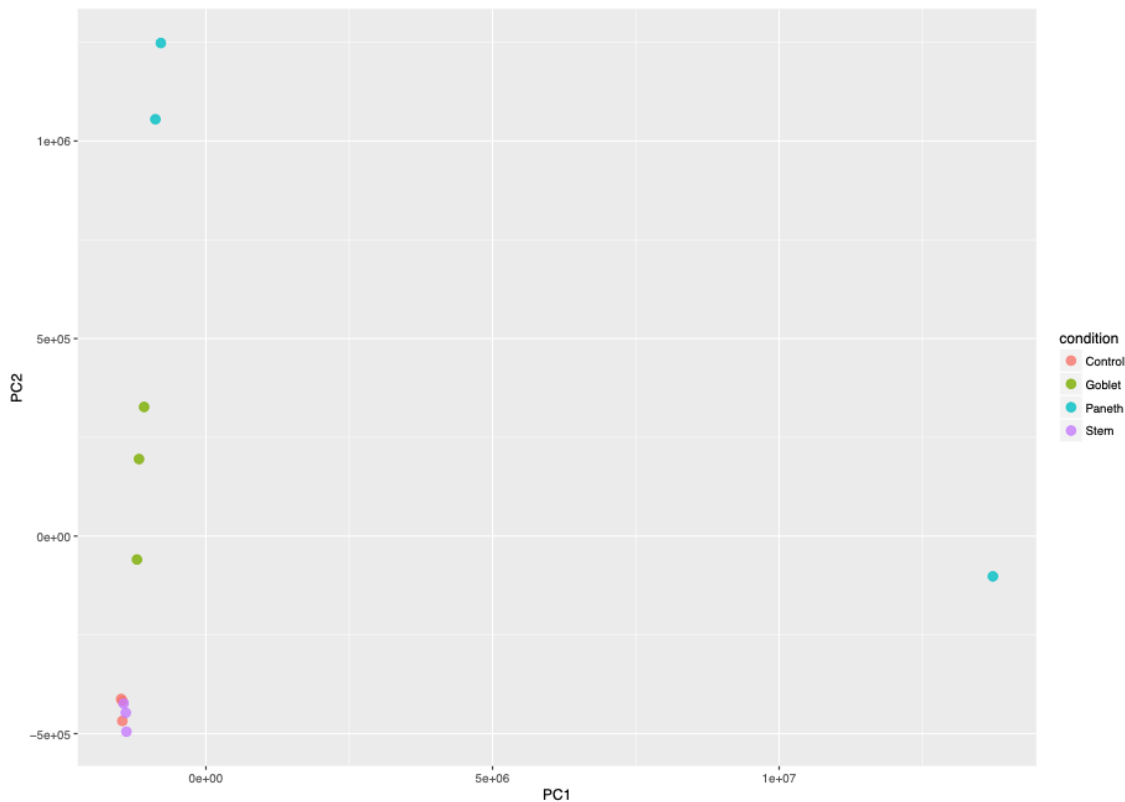


Figure 2.2. Principle component analysis of transcript expression of enteroids. Paneth cell enriched enteroid outlier to the left side of PC1 was removed from downstream analysis. Plot courtesy of Tomasz Wrzesinski (EI). Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

2.3.10 Data and software availability

Small and stranded RNA-seq data has been deposited in the European Nucleotide Archive (ENA) with accession numbers PRJEB32354 and PRJEB32366 respectively. Scripts to analyse the differentially expressed genes are available on GitHub: https://github.com/korcsmarosgroup/organoid_regulatory_networks.

2.4 Results

2.4.1 Enteroids enriched for target cell type signatures

Zoe Matthews (UEA, Norwich Medical School) generated 3D self-organising enteroid cultures *in vitro* from murine small intestinal crypts (Figure 2.3) (Sato and Clevers, 2013; Sato et al., 2009, 2011). In addition to conventionally differentiated enteroids (CDEs), she generated enteroids enriched for Paneth cells and goblet cells using well-established and published protocols, presented in detail in the Methods (Sato and Clevers, 2013; Yin et al., 2014).

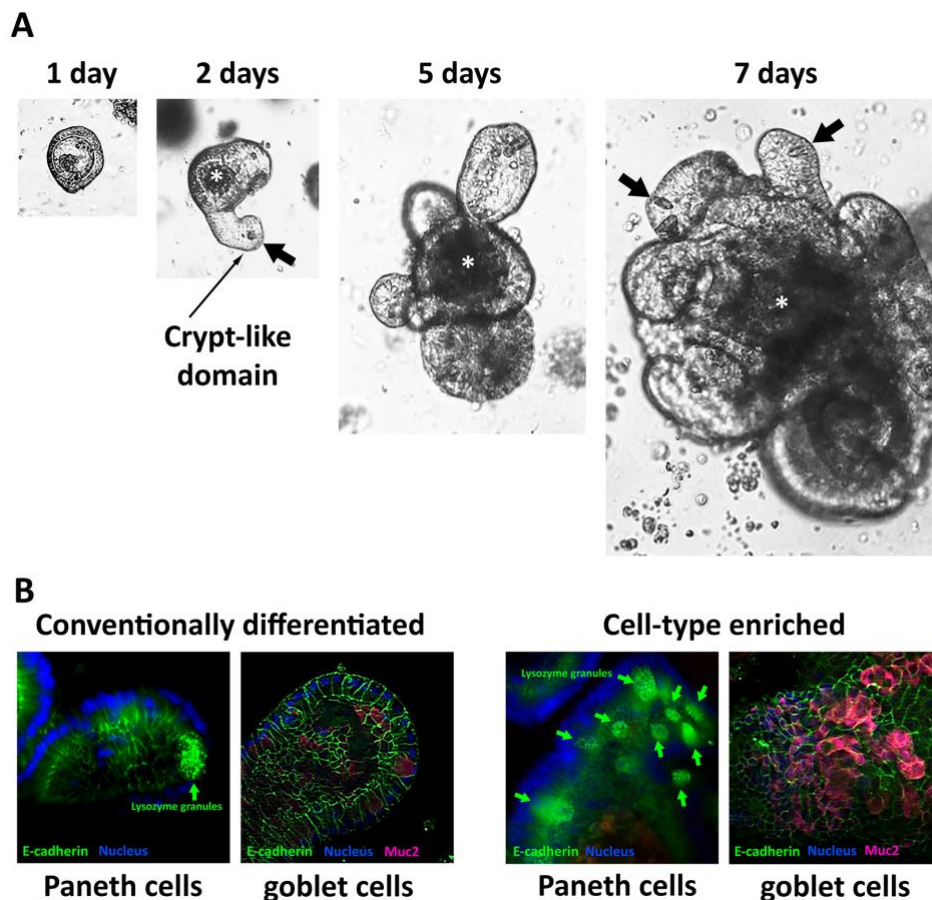


Figure 2.3. Small intestinal 3D organoid culture. **A.** Culture of isolated mouse small intestinal epithelial crypts in Matrigel matrix and ENR media (conventionally differentiated) for 7 days. Isolated crypts form 3D cysts which bud after 2 days of culture to form crypt- and villus-like domains. Paneth cells are clearly visible by light microscopy (Black arrows). Mucus and shedding cells accumulate in the central lumen of organoids (*). $n = 3$. **B.** cell type-specific enrichment illustrated by immunofluorescence labelling of cultured mouse 3D enteroids, conventionally differentiated (left) and enriched for either Paneth cells or goblet cells (right). Lysozyme granules characteristic of Paneth cells are indicated with a green arrow. Goblet cells were identified using a specific anti-Muc2 mucin antibody (pink). Images provided by Zoe Matthews, Emily Jones and Isabelle Hautefort. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Bulk transcriptomics data was obtained from each set of enteroids to determine genes with differential expression resulting from enteroid skewing protocols. Differentially expressed genes were obtained by comparing the RNA expression levels (including protein coding genes, lncRNAs and miRNAs) of enteroids enriched for Paneth cells or goblet cells to those of CDEs. 4,135 genes were differentially expressed (absolute log₂ fold change ≥ 1 and q value ≤ 0.05) in the PCeE dataset, and 2,889 were differentially expressed in the GCeE dataset (Figure 2.4A-C, File S2.1). The larger number of differentially expressed genes (DEGs) in the PCeE data could be attributed to the highly specialised nature of Paneth cells (Clevers and Bevins, 2013; Stappenbeck and McGovern, 2017). The majority of the DEGs were annotated as protein coding: 79% in the PCeE dataset and 84% in the GCeE dataset. In addition, we identified lncRNAs (PCeE, 11%; GCeE, 9%) and miRNAs (PCeE, 4%; GCeE, 2%) among the DEGs (Figure 2.4B). I identified some DEGs that were in both the PCeE and the GCeE datasets, exhibiting the same direction of change compared to the CDE data. In total, 1,363 genes were found upregulated in both the PCeE and the GCeE data, while 442 genes were found downregulated in both datasets (Figure 2.4C). This result highlights considerable overlap between the results of skewing enteroids towards Paneth cells and goblet cells and can be explained by the shared differentiation history and secretory function of both Paneth cells and goblet cells.

I employed pathway analysis to study functional associations of the DEGs (Figure 2.4D). The PCeE-specific DEGs were associated with a number of metabolic pathways, including Metabolism of vitamins and cofactors, Pyruvate metabolism and Citric Acid (TCA) cycle and Cholesterol biosynthesis. On the other hand, GCeE-specific DEGs were associated with the cell cycle through pathways such as Cell Cycle Checkpoints, DNA replication and G1/S transition. Pathways associated with the shared DEGs included *Transmission across Chemical Synapses*, *Integration of energy metabolism* and a number of pathways linked to hormones. As hormone functions are characteristic of enteroendocrine cells, this analysis suggests that enteroendocrine cells are enriched in both the PCeEs and the GCeEs.

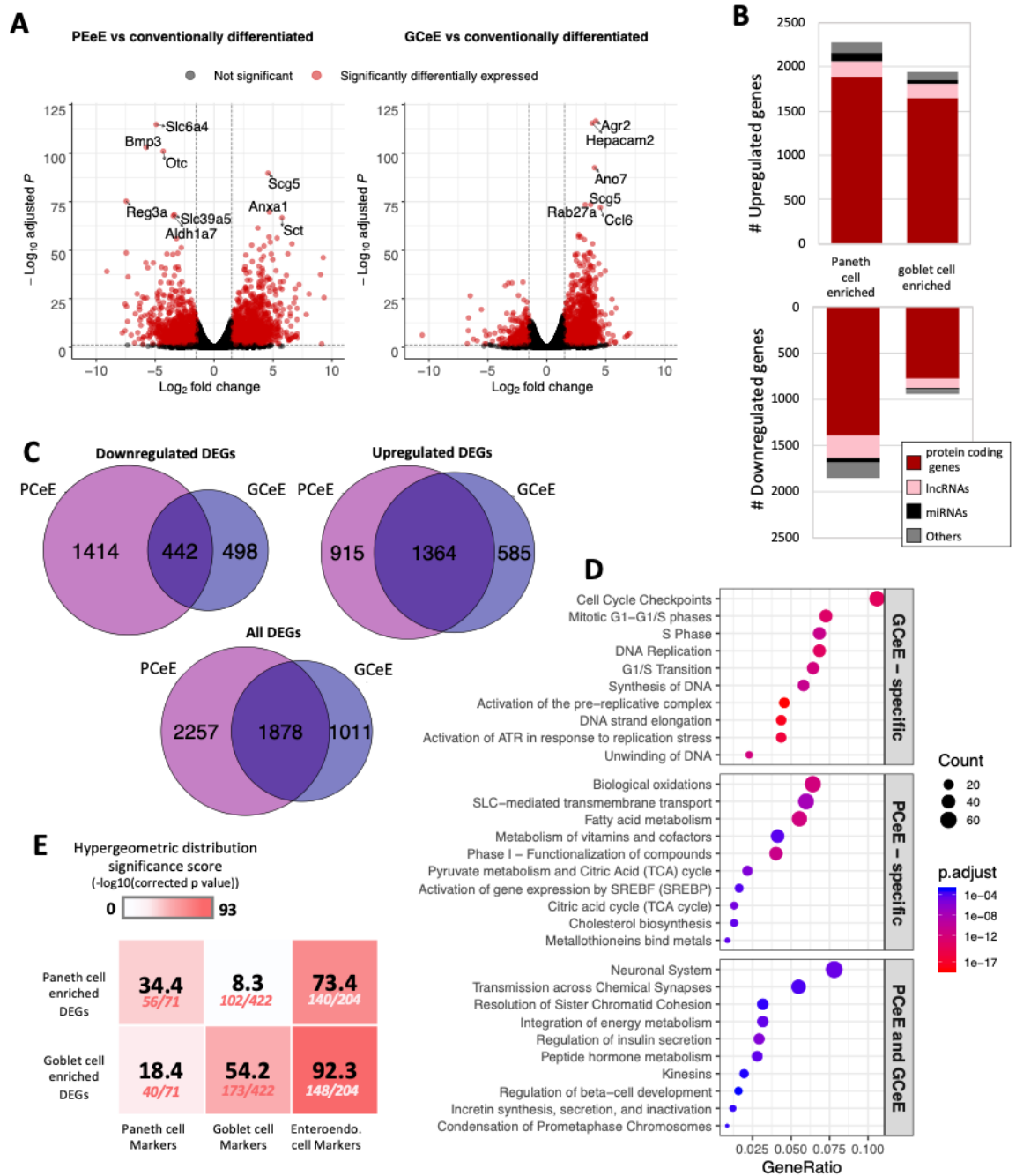


Figure 2.4. Differentially expressed genes in Paneth cell enriched enteroids (PCeEs) and goblet cell enriched enteroids (GCEEs) (compared to conventionally differentiated enteroids). **A:** Volcano plots showing log₂ fold change and adjusted p value for each gene following differential expression analysis of PCeEs (left) and GCEEs (right). Horizontal and vertical lines indicate the differential expression criteria cut offs (q value ≤ 0.05 and absolute log₂ fold change ≥ 1). **B:** Number of differentially expressed genes (DEGs). miRNA - microRNA; lncRNA - long non-coding RNA; Genes annotated as ‘other’ include pseudogenes and antisense genes. **C:** Venn diagrams indicating the number of DEGs (passing the cut off criteria). **D:** Top 10 Reactome pathways of the 50 most significant DEGs (by q value). **E:** Enrichment of cell type-specific marker genes in the DEG lists. Higher significance scores indicate greater enrichment. Number of markers in DEG list out of the total number of markers shown below significance score. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

To validate the cell types present in the enteroids, I investigated the expression of five previously reported major cell type-specific markers across the enteroids using transcript abundances and RNA differential expression results (Figure 2.5). The control enteroids and the cell type enriched enteroids expressed all five investigated markers: *Lgr5* (stem cells), Chromogranin A (*ChgA*) (enteroendocrine cells), *Muc2* (goblet cells), *Lyz1* (Paneth cells) and Villin 1 (*Vil1*) (epithelial cells). We observed an upregulation of *Muc2*, *Lyz1* and *ChgA* and a downregulation of *Lgr5* in PCeEs and GCeEs compared to the control enteroids, confirming the more pronounced differentiated status of the enteroids. In addition, a number of Paneth cell specific antimicrobial peptide genes were differentially expressed in the PCeE dataset, including Angiogenin 4 (*Ang4*), Regenerating islet-derived protein 3 gamma (*RegIIIγ*), Phospholipase A2 group IIA (*Pla2g2a*) and Defensin alpha 2 (*Defa2*) (Table S2.1). Some of these genes were also differentially expressed in the GCeE dataset but with smaller log fold change values, e.g. *Lyz1* and *Ang4*. Conversely, a number of goblet cell mucin related genes (including *Muc2* and trefoil factor 3 (*Tff3*)) were differentially expressed in both datasets although all genes exhibited a smaller increase in the PCeEs (Table S2.1). Therefore, using primary cell type-specific markers, antimicrobial peptide genes and mucin-related genes, I show that the enteroids contain all major cell types, and that Paneth cell are most upregulated in the PCeEs, while goblet cells are most upregulated in the GCeEs. We also note that both differentiation methods resulted in increases of other secretory cell types as well.

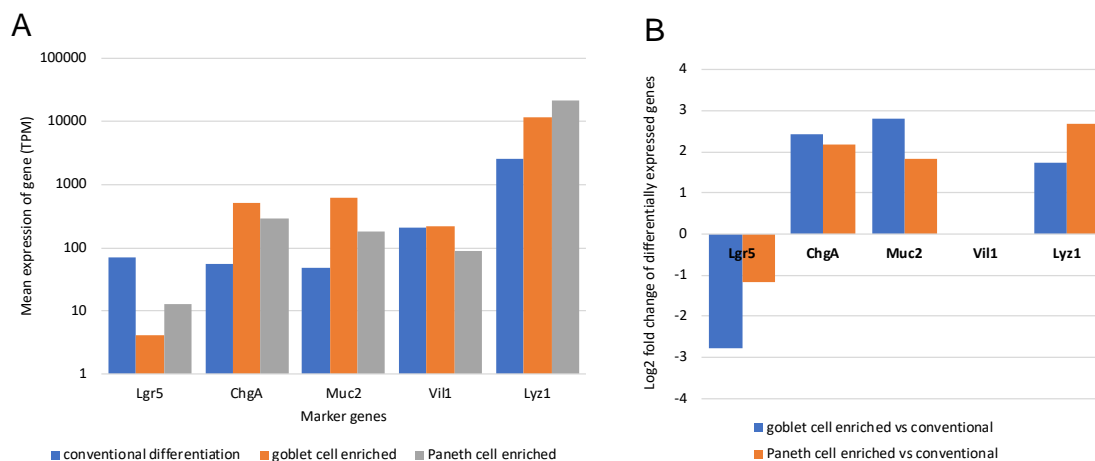


Figure 2.5. Transcript abundances and differential expression of five major cell type markers. **A:** Mean transcript abundances in the conventionally/normally differentiated, goblet cell enriched and Paneth cell enriched enteroids. TPM - transcripts per million. **B:** Log₂ fold change in the goblet cell enriched enteroid vs conventional enteroid analysis and the Paneth cell enriched enteroid vs conventional enteroid analysis. Data only presented where the differential expression criteria passed (q value ≤ 0.05 and absolute log₂ fold change ≥ 1). Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

To further investigate secretory cell type-specific signatures of the enteroids, I measured enrichment of IEC lineages in the upregulated DEG lists using additional marker genes of Paneth cells, goblet cells, enteroendocrine cells, tuft cells and enterocytes. These marker genes were obtained from a single cell study of mouse small intestinal epithelium by (Haber et al., 2017). All tests were significantly enriched for secretory cell types (hypergeometric model, q value ≤ 0.05), with greater enrichment of Paneth cell markers in the PCeE DEG list and goblet markers in the GCeE DEG list (Figure 2.4E, Table S2.2). This confirms that both enteroid enrichment protocols were successful in increasing the proportion of their target cell type, but also increased proportions of other secretory lineages, albeit to a lesser extent. This observation confirms previous studies that these enteroid differentiation protocols result in enteroendocrine enrichment in addition to Paneth cell and goblet cell enrichment (Luu et al., 2018; Yin et al., 2014).

In conclusion, we have used image-based validation, pathway analysis and marker gene investigation to show successful enrichment of target cell types in the PCeE's and GCeE's. I also highlighted an additional increase in other secretory lineages, particularly enteroendocrine cells, as a result of both enrichment protocols.

2.4.2 Regulatory networks are altered by enteroid differentiation skewing

To gain an understanding of the regulatory changes occurring when enteroid development is skewed, we applied a network biology approach to identify regulator-target relationships within the DEG lists. First, Tomasz Wrzesinski and Padhmanand Sudhakar generated a large network of non-specific molecular interactions known to occur in mice, by collating lists of published data (described in the methods Table 2.1). The resulting network (termed the universal network) consisted of 1,383,897 unique regulatory interactions connecting 23,801 molecular entities. All interactions within the network represent one of the following types of regulation, where every node is a DEG: TF-TG, TF-lncRNA, TF-miRNA, miRNA-mRNA or lncRNA-miRNA. TF-TGs and TF-lncRNAs make up the majority of the network at 77% and 11% of all interactions, respectively. Due to its non-specific nature, this universal network contains many interactions not relevant for the current biological context. In order to get a clearer and valid picture of regulatory interactions occurring in our enteroids, I used the universal network to annotate the PCeE DEGs and GCeE DEGs with regulatory connections. Combining these connections, I generated specific

regulatory networks for PCeEs and GCeEs, where every node is a DEG and every interaction has been observed in mice previously.

In total, the PCeE network, generated using differential expression data from the PCeEs compared to the CDEs, contained 37,062 interactions connecting 208 unique regulators with 3,023 unique targets (Figure 2.6A). The GCeE network, generated using differential expression data from the GCeEs compared to the CDEs, contained 19,171 interactions connecting 124 unique regulators with 2,095 unique targets (Figure 2.6A). 15.7% of all interactions (8,856 out of 56,234) were shared between the PCeE and GCeE networks, however the interacting molecular entities in these interactions (termed nodes) did not all exhibit the same direction of differential expression between the networks (comparing PCeE or the GCeE data to the CDE data). In each of the enriched enteroid regulatory networks, a particular gene was represented (as a node in the network) only once, but may have been involved in multiple different interactions. In different interactions, a single node could act either as a regulator or as a target and in different molecular forms, for example, as a lncRNA in one interaction and as a target gene in another.

To further investigate the makeup of these networks, I employed cluster analysis to identify highly interconnected regions (possible regulatory modules) in the PCeE and GCeE regulatory networks. Using the MCODE software (Bader and Hogue, 2003), I identified five distinct clusters in the PCeE network and seven distinct clusters in the GCeE network. A total of 1314 nodes are present in the PCeE network clusters and 698 in the GCeE network clusters. Functional analysis identified Reactome pathways (Fabregat et al., 2018a) associated with each of the modules. Significant pathways ($q \text{ val} \leq 0.05$) were identified only for the highest ranked three modules from each network, with a total of 12 pathways shared between the PCeE and GCeE associated clusters (out of 32 associated with the PCeE clusters and 42 with the GCeE clusters) (Figure 2.6B-D). Of particular note, the first cluster of the GCeE network has associations with the *endosomal/vacuolar pathway* and *antigen presentation*, the second cluster is associated with the *cell cycle*. Of the PCeE clusters, the first cluster is associated with a range of functions including *nuclear receptor transcription pathway*, *regulation of lipid metabolism* and *senescence*. The second is associated with *response to metal ions* and *endosomal/vacuolar pathway* and the third with *G alpha (i) signalling events*.

In conclusion, I have generated regulatory interaction networks, including transcriptional and post-transcriptional interactions, which illustrate the effect of skewing enteroid differentiation towards Paneth cell and goblet cell lineages.

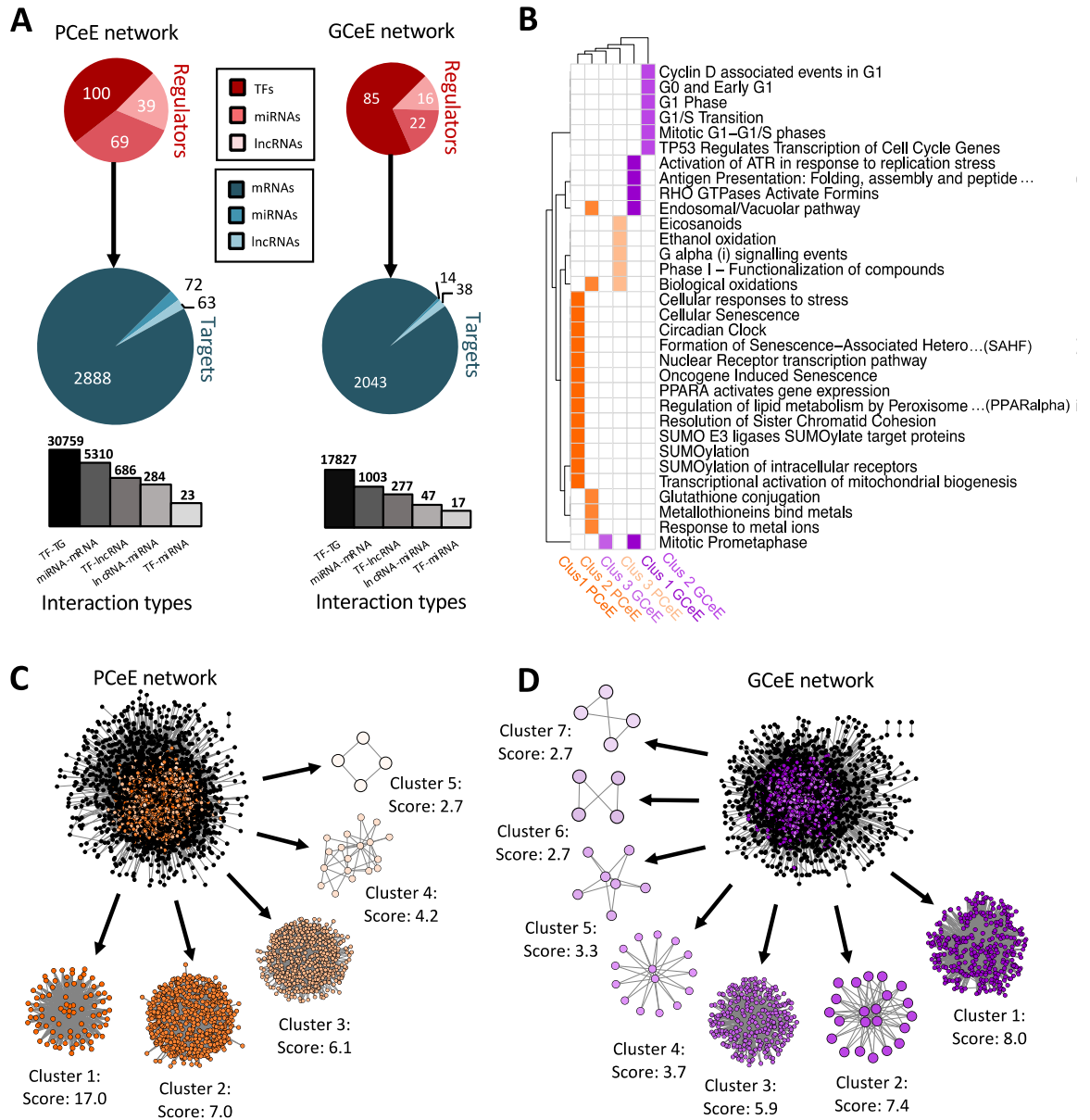


Figure 2.6. Summary and cluster analysis of regulatory network for Paneth cell enriched enteroid (PCeE) and goblet cell enriched enteroid (GCeE) datasets. **A:** Summary of number of nodes and interactions in the whole PCeE (left) and GCeE (right) networks. Total number of each regulator type shown in red, number of each target type shown in blue. In the targets pie-chart, mRNAs also represent protein coding genes and proteins, miRNAs also represent miRNAs genes and lncRNAs also represent lncRNA genes. Size of circles represents log10(total unique regulators/targets). Bar chart represents the distribution of interaction types in the networks (log10 scale). **B:** Heatplot of Reactome pathways significantly associated ($q \text{ val} \leq 0.05$) with each

cluster of the PCeE (orange) and GCeE (purple) networks. Only the top 5 pathways shown for each group (or more where equal q values). Only the top 3 clusters had significantly associated pathways. Clusters labelled with rank and cell type and colours match the colour of the cluster shown in C and D. **C, D:** Visualisation of the PCeE and GCeE regulatory networks with their associated clusters. The cluster rank and score is given next to each cluster. Black nodes in the whole networks represent nodes which were not found in any cluster, whereas coloured nodes represent the cluster which they are part of. TF - transcription factor; miRNA - microRNA; lncRNA - long non-coding RNA. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

2.4.3 Paneth cells and goblet cells have shared and unique regulators

Through pathway and marker analysis I predicted that our PCeE and GCeE datasets (i.e. DEG lists), and consequently our regulatory networks, contain signatures from the cell type of interest as well as additional noise from other secretory lineages. To focus specifically on the cell type-specific elements of the networks, I used previously identified cell type-specific markers to extract predicted Paneth cell and goblet cell regulators from our PCeE and GCeE networks (Haber et al., 2017). As cell type-specific markers represent genes performing functions specific to a particular cell type, I expect that the regulators of these marker genes will have an important role in determining the function of said cell type. To identify these regulators, I extracted from the PCeE and GCeE networks, all relevant cell type-specific markers and their direct regulators. These new networks were termed the Paneth cell subnetwork and goblet cell subnetwork respectively. The Paneth cell subnetwork contained 33 markers specific for Paneth cells with 62 possible regulators. The goblet cell subnetwork contained 150 markers with 63 possible regulators (Figure 2.7). Observing the ratio of regulators and markers, the Paneth cell subnetwork had, on average, 1.88 regulators for each marker. On the other hand, the goblet cell subnetwork exhibited only 0.42 regulators for each marker. The quantity of markers identified in each subnetwork (33 in the Paneth network and 150 in the goblet network) correlates with the number of marker genes identified by Haber *et al.* (Haber et al., 2017). However, far fewer regulators were identified in the goblet cell subnetwork per marker than for the Paneth cell subnetwork. Whilst the underlying reason for this discrepancy is unknown, it could potentially be evidence of the complex regulatory environment required to integrate and respond to the arsenal of signals recognised by Paneth cells in comparison to goblet cells (Stappenbeck and McGovern, 2017).

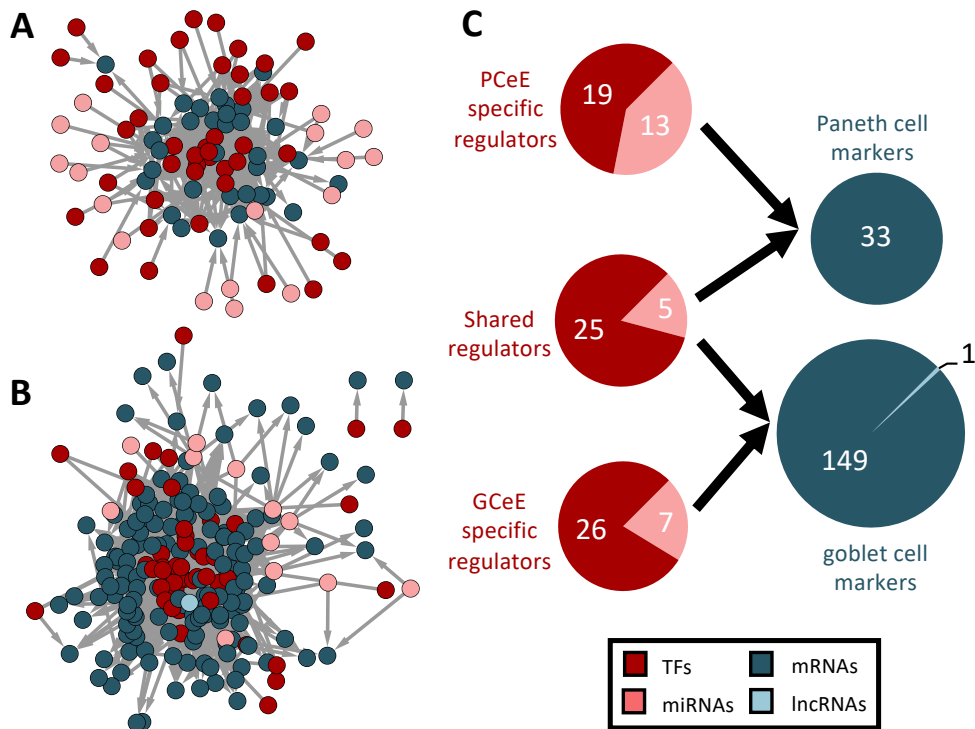


Figure 2.7. Regulator-marker subnetworks for Paneth cell and goblet cell datasets. **A, B:** Paneth cell (A) and goblet cell (B) subnetworks. Nodes represent genes, transcription factors or RNAs and edges represent directed physical regulatory connections. Regulators are shown in red and pink. Cell type-specific markers are shown in blue. **C:** Summary of the number of nodes present in both the subnetworks. Paneth cell data above and goblet cell data below. Total number of each regulator type shown in red, number of each target type shown in blue. Regulators have been categorised based on their membership in the two subnetworks - shared regulators are present in both networks. In the targets pie-chart, mRNAs also represent protein coding genes. Size of circles represents \log_{10} (total unique regulators/targets). TF - transcription factor; miRNA - microRNA; lncRNA - long non-coding RNA. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Of the 95 marker regulators, we identified approximately one-third (30/95) as present in both subnetworks (Figure 2.7C). Given that the markers are different between the cell types, a regulator shared between the Paneth cell and goblet cell subnetworks must show an altered pattern of regulatory targeting in the two cell types. This phenomenon, referred to as regulatory rewiring, often results in functional differences of shared regulators in different environments (Han et al., 2017) - for example, in this case, between the Paneth cells and goblet cells.

Further investigation of the distinct regulator-marker interactions highlighted a gradient of regulator specificity. We generated matrices to visualise the markers controlled by each regulator in the goblet cell (Figure 5A) and the Paneth cell (Figure 5B,C) subnetworks. Each

coloured square indicates that a marker (shown on the y-axis) is regulated by the corresponding regulator (shown on the x-axis). Squares are coloured blue if the associated regulator is shared between the Paneth cell and goblet cell subnetworks and orange if they are specific to one subnetwork. A collection of regulators (both subnetwork specific and shared) appear to regulate large proportions of the markers. For example, Protein C-ets-1 (ETS1), Glucocorticoid receptor (NR3C1) and Vitamin D receptor (VDR) regulate >50% of the markers in both the Paneth cell and the goblet cell subnetworks. Specific to the Paneth cell subnetwork, CCAAT enhancer-binding protein alpha (CEBPA), Jun proto-oncogene AP-1 transcription factor subunit (JUN), Nuclear receptor subfamily 1 group D member 1 (NR1D1) and Retinoid X receptor alpha (RXRA) regulate >50% of the markers. Specific to the goblet cell subnetwork, Growth factor independent 1B transcriptional repressor (GFIB1) and MYC proto-oncogene (MYC) regulate >50% of the markers. These regulators represent potential master regulators of differentiation or maintenance of the given cell types in the enriched enteroids. Referring back to the highly-interconnected clusters identified in the PCeE and GCeE networks (Figure 3C-D), we find these predicted master regulators in different clusters. In the PCeE network, CEPBA, NR1D1, NR3C1 and RXRA are in cluster 1, VDR is in cluster 2, JUN is in cluster 3 and ETS1 is unclustered. In the GCeE network, ETS1 and MYC are in cluster 1, NR3C1 and VDR are in cluster 2 and GF11B is in cluster 3. This suggests a wide range of central functions are carried out by this group of regulators, with possible divergence of roles between the Paneth cell and the goblet cell. In contrast to the predicted master regulators, regulators such as MAF BZIP transcription factor K (MAFK) in the Paneth cell subnetwork and SAM pointed domain containing ETS transcription factor (SPDEF) in the goblet cell subnetwork regulate only one marker. These regulators likely have more functionally specific roles.

Together, these results highlight potential regulators which likely play key roles in specification and maintenance of Paneth and goblet cells and their functions in cell type enriched enteroids.

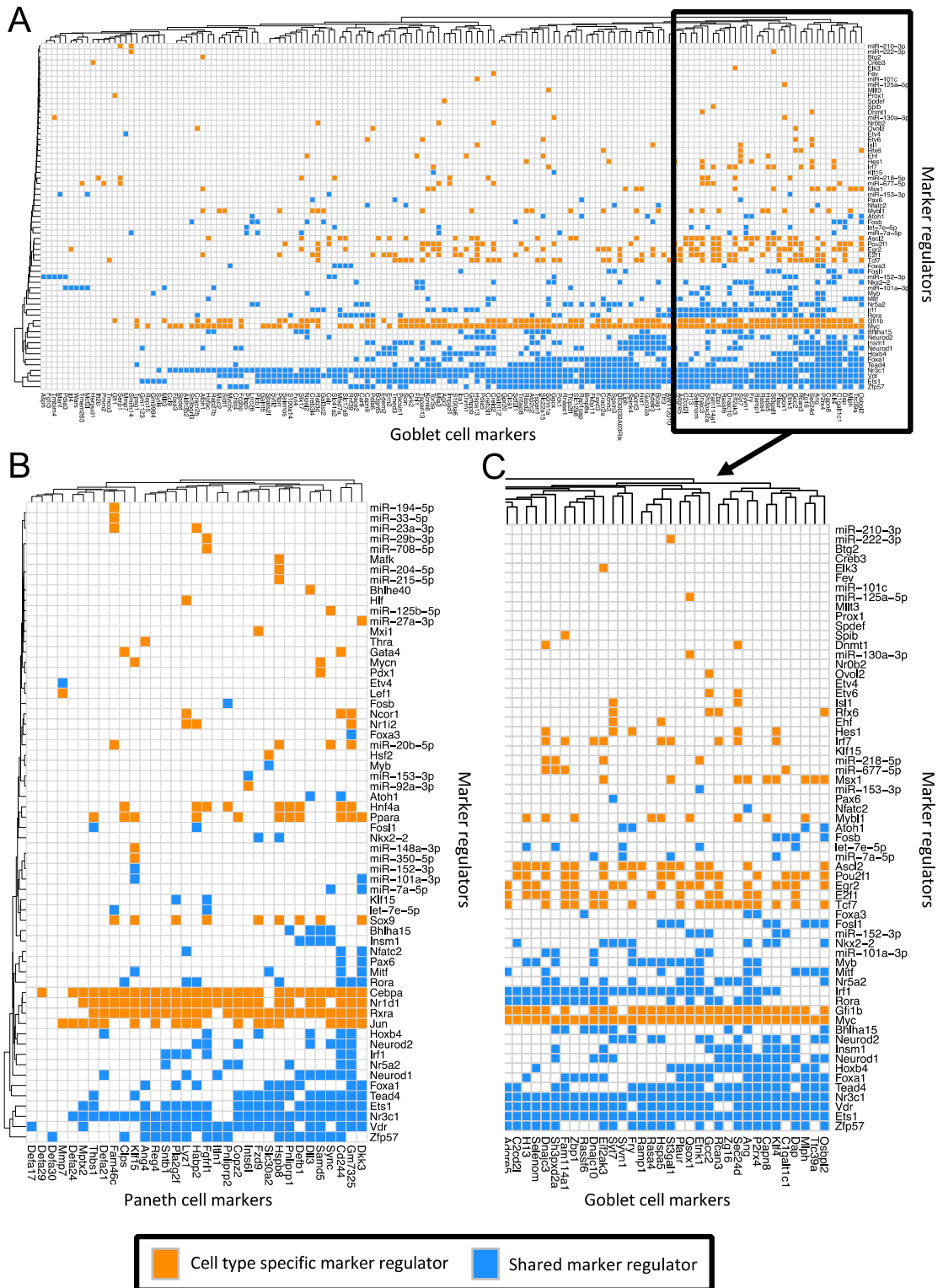


Figure 2.8. Matrices of interactions between markers and their regulators in the Paneth cell and goblet cell subnetworks. Regulators on y-axis, markers (regulator targets) on x-axis. Orange boxes indicate the interaction of a regulator and a marker where the regulator is only found in one of the two subnetworks. Blue boxes signify that the regulator is found in both the Paneth

cell and the goblet cell subnetworks. **A:** All goblet cell markers (Haber et al., 2017) and their regulators in the goblet cell subnetwork. **B:** All Paneth cell markers (Haber et al., 2017) and their regulators in the Paneth cell subnetwork. **C:** Sub-section of A showing the markers (and their regulators) which have the most regulatory connections. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

2.4.4 Regulators are rewired between Paneth cells and goblet cells

Cell type-specific markers, which carry out cell type-specific functions, are inherently different between the Paneth cell and goblet cell subnetworks (mutually exclusive). Therefore, the regulators observed in both Paneth cell and goblet cell subnetworks (shared regulators) are expected to target different marker genes. To do this, the regulators must have different regulatory connections in the different cell types, a phenomenon termed ‘rewiring’ (Park and Wang, 2018). I extended the analysis to the original regulatory networks (PCeE and GCeE networks) to investigate whether any of the 30 identified shared regulators are rewired between the whole PCeE and the GCeE networks, and thus are highly likely to have different functions in the two types of enriched enteroids as well as between Paneth and goblet cells. To quantify rewiring of each of these regulators, I observed their targets in the PCeE and GCeE networks using the Cytoscape application, DyNet (Goenawan et al., 2016). DyNet assigns each regulator a rewiring score depending on how different their targets are between the two regulatory networks (Table S2.3). Using these rewiring scores, I identified the five most rewired regulators (of 30) as ETS variant transcription factor 4 (ETV4), let-7e-5p, miR-151-3p, MYB proto-oncogene (MYB) and RAR related orphan receptor A (RORA). Functional enrichment analysis was carried out on the targets of these regulators to test whether the targets specific to the PCeE and GCeE networks have different functions (hypergeometric model, q value ≤ 0.1) (File S2.2). Across all five regulators the general trend indicated that targets specific to the PCeE network are associated with metabolism; targets specific to the GCeE network are associated with cell cycle and DNA repair. As pathway analysis carried out on the enteroid DEGs identified the same phenomenon (Figure 2.4D), this suggests that the rewired regulators could be key drivers of transcriptional changes during the skewing of enteroid differentiation towards Paneth cell or goblet cell lineages. In addition, given that the strongest signal of enriched enteroids represents their enriched cell type, I predict that these functions are key features of Paneth cells and goblet cells in the enteroids, and that the rewired regulators are important drivers of cell type-specific functions.

Looking at the regulators in more detail, the GCeE specific targets of miR-151-3p, for example, are significantly enriched in functions relating to antigen presentation, cell junction organisation, Notch signalling and the calnexin/calreticulin cycle. None of these functions are enriched in the shared or PCeE targets. Of particular interest is the calnexin/calreticulin cycle, which is known to play an important role in ensuring proteins in the endoplasmic reticulum are correctly folded and assembled (Leach and Williams, 2003). Dysfunction of protein folding and the presence of endoplasmic reticulum stress are both associated with IBD (Kaser and Blumberg, 2009; Kaser et al., 2008, 2011). Therefore, we predict that miR-151-3p plays a role in the secretory pathway of goblet cells and could be an interesting target for IBD research. In addition, different functional profiles were also observed for the targets of RORA in the PCeE and GCeE regulatory networks: targets present in both networks are significantly associated with mitosis, whereas those specific to the PCeE network are associated with metabolism, protein localisation, nuclear receptor transcription pathway, circadian clock and hypoxia induced signalling. GCeE specific targets of RORA are connected to Notch signalling, cell cycle and signalling by Rho GTPases (associated with cell migration, adhesion and membrane trafficking) and interferon.

Altogether these observations show that some of the regulators of both Paneth cell and goblet cell marker genes have different targets (with different associated functions) between the PCeE and the GCeE networks. This suggests that regulatory rewiring occurs between Paneth cell and goblet cell types.

2.4.5 Regulatory networks are relevant to study IBD

To investigate the function and relevance of the predicted master regulators in IBD, I carried out three analyses:

- 1) a literature search to check what is known about the identified master regulators.
- 2) an enrichment analysis to evaluate the disease relevant genes in the PCeE and GCeE networks and among the targets of the predicted master regulators (with help from Padhmanand Sudhakar)
- 3) a comparative analysis with human biopsy based single cell dataset to confirm the relevance of the PCeE and GCeE networks.

The literature search was carried out using the three groups of predicted master regulators: those specific to the Paneth cell markers (CEBPA, JUN, NR1D1, RXRA), those specific to the goblet cell markers (GFI1B, MYC) and those which appear to regulate many of the markers of both cell types (ETS1, NR3C1, VDR). We identified five genes (ETS1, NR1D1, RXRA, NR3C1, VDR) with associations to inflammation, autophagy and/or inflammatory bowel disease (IBD), as shown in Table 2.2. These genes correspond to 71% (5/7) of the Paneth cell associated master regulators and 60% (3/5) of the goblet cell associated master regulators. Interestingly, four of these genes (all apart from ETS1), encode nuclear hormone receptors.

Putative master regulator	Autophagy / inflammation / IBD associations	References
NR1D1 (Nuclear receptor subfamily 1 group D member 1)	Modulates autophagy and lysosome biogenesis in macrophages leading to antimycobacterial effects	(Chandra et al., 2015)
	SNP rs12946510 which has associations to IBD, acts as a cis-eQTL for NR1D1	(Mirza et al., 2015)
NR3C1 (Glucocorticoid receptor)	Associations with cellular proliferation and anti-inflammatory responses	(Oakley and Cidlowski, 2013)
	Exogenous glucocorticoids are heavily used as anti-inflammatory therapy for IBD	(Prantera and Marconi, 2013; Rutgeerts, 1998)
	ATG16L1, an autophagy related gene, was downregulated in patients who do not respond to glucocorticoid treatment	(De Iudicibus et al., 2011; Dubois-Camacho et al., 2017)
	Transcriptionally regulates NFK β 1, a SNP affected gene in ulcerative colitis	(Dinkel et al., 2016; Yemelyanov et al., 2007)
VDR (Vitamin D receptor)	Regulates autophagy in Paneth cells through ATG16L1 – dysfunction of autophagy in Paneth cells has been linked to Crohn’s disease	(Bakke et al., 2018; Wu et al., 2015)
	Induces antimicrobial gene expression in other cell lines	(Gombart et al., 2005; Wang et al., 2004)
	Specific polymorphisms in the VDR genes have been connected to increased susceptibility to IBD	(Pei et al., 2011)
	A study looking at colonic biopsies of IBD patients observed reduced VDR expression compared to healthy biopsies	(Abreu et al., 2004)
	Interacts with five SNP affected UC genes	(Bovolenta et al., 2012; Lesurf et al., 2016)
RXRA (Retinoid X receptor alpha)	Heterodimerizes with VDR (see above)	(Bettoun et al., 2003)
ETS1 (Protein C-ets-1)	Important role in the development of hematopoietic cells and Th1 inflammatory responses	(Grenningloh et al., 2005; Mouly et al., 2010)
	Angiogenic factors in the VEGF-Ets-1 cascades are upregulated in UC and downregulated in CD	(Konno et al., 2004)
	IBD susceptibility gene	(Li et al., 2018)

Table 2.2. Literature associations relating to autophagy, inflammation and IBD for putative master regulators. Table reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Given the possible relationship between the identified master regulators and IBD, I tested the potential of the PCeE and GCeE regulatory networks to study the pathomechanisms of CD or UC. I checked for the presence of known CD or UC associated genes in the networks, using data from two studies of single nucleotide polymorphisms (SNPs) (Farh et al., 2015; Jostins et al., 2012) and one study of CD expression quantitative trait loci (eQTLs) (Di Narzo et al., 2016). Using hypergeometric significance tests, I found that the PCeE network was significantly enriched in all tested lists: genes with UC associated SNPs (13/47, $p < 0.005$), genes with CD associated SNPs (22/97, $p < 0.005$) and genes with CD associated eQTLs (290/1607, $p < 0.0001$) (Table S2.4, Table S2.5). On the other hand, we found that the GCeE network was significantly enriched in genes with UC associated SNPs (10/47, $p < 0.005$) but regarding CD, the genes with SNP associations were not significantly enriched (12/97, $p = 0.11$) and the genes with eQTL associations were enriched with a larger p value ($p < 0.05$) (Table S2.4, Table S2.5).

Next, I investigated whether any of the genes with UC or CD associated SNPs act as regulators in the PCeE and GCeE networks. Of the genes with CD associated SNPs, one acts as a regulator in each network. Similarly, two different genes with UC associated SNPs act as regulators in the networks. A summary of these genes and their regulated targets is given in Figure 2.9. Specifically, regarding CD associated genes, in the PCeE network, the gene D-box binding PAR BZIP transcription factor (*Dbp*) regulates BIK, which encodes the BCL2 interacting killer, a pro-apoptotic, death promoting protein. In the GCeE network, Notch receptor 2 (NOTCH2) regulates Notch receptor 3 (*Notch3*) and Hairy and enhancer of split-1 (*Hes1*). Specifically, regarding UC associated genes, in the PCeE network, Hepatocyte Nuclear Factor 4 Alpha (HNF4A) regulates 994 genes/RNAs including nine Paneth cell markers and one other gene with UC associated SNPs (TNF superfamily member 15, *Tnfsf15*).

- CD244 molecule, *Cd244a*
- Fibroblast growth factor receptor like, *Fgfr1*
- Colipase, *Clps*
- Hyaluronan binding protein 2, *Habp2*
- Heat shock protein family B member 8, *Hspb8*
- Pancreatic lipase related protein 1/2, *Pnliprp1/2*
- Defensin beta 1, *Defb1*
- Myomixer myoblast fusion factor, *Mymx*

Additionally, a gene with UC associated SNPs, Nuclear receptor subfamily 5 group A member 2 (*Nr5a2*), was found in both the PCeE and GCeE networks regulating 389 and 276 genes/RNAs respectively. In the PCeE network *Nr5a2* targets include 6 Paneth cell markers (*Cd244a*, COPI coat complex subunit zeta 2 (*Copz2*), *Pnliplr1/2*, Syntrophin beta 1 (*Sntb1*), *Mymx*). Ultimately, the large number of targets of these regulatory UC associated genes suggests they have wide ranging effects on the regulatory network of Paneth and goblet cells. To further establish the relevance of the inferred PCeE and GCeE networks, Padhmanand Sudhakar also found an over-representation of drug target associated genes in both the PCeE and GCeE networks (2683/16223 and 1918/16223 respectively, $p < 0.0001$), highlighting their potential for the study of therapeutic implications.

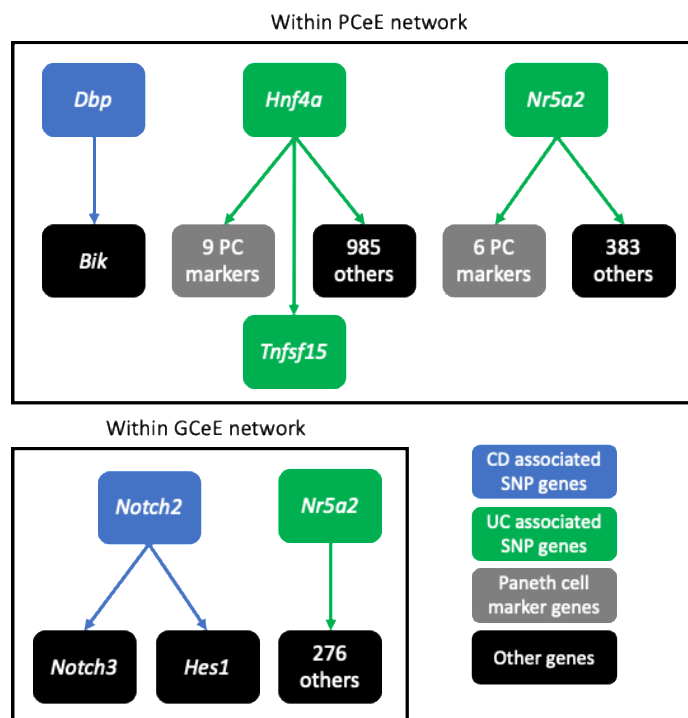


Figure 2.9. Crohn's disease and ulcerative colitis associated SNP genes and their targets within the PCeE and GCeE networks. CD - Crohn's disease; UC - ulcerative colitis; PCeE - Paneth cell enriched enteroid; GCeE - goblet cell enriched enteroid network; PC - Paneth cell.

To investigate the link between predicted master regulators and IBD, I observed whether the genes with UC and CD associated SNPs are regulated by the predicted master regulators in the PCeE and GCeE networks (Table S2.6). Given that Paneth cell dysregulation is classically associated with CD and goblet cell dysregulation/depletion with UC (Cader and Kaser, 2013), I focused this analysis only on these pairings, examining CD genes amongst targets of Paneth cell predicted master regulators, and UC genes amongst targets of goblet cell predicted master

regulators. In the PCeE network, I found 21 (of 22) of the CD genes were regulated by at least one of the seven Paneth cell predicted master regulators, while the targets of these master regulators were significantly enriched with CD genes in the PCeE network ($p < 0.001$). Similarly, I observed that all 10 UC genes in the GCeE network were regulated by at least one of the five goblet cell predicted master regulators, while the targets of these master regulators were significantly enriched with UC genes ($p < 0.005$).

To confirm the relevance of these predicted master regulators in a human system, a similar analysis was carried out using goblet cell differentially expressed genes from a recent single cell study of human inflamed UC colon biopsies (Smillie et al., 2019). Using the top 100 differentially expressed genes, following conversion to mouse Ensembl identifiers, 20 were found to be targeted by the predicted goblet cell master regulators in the GCeE network. This represents a significant enrichment amongst all master regulator targets ($p < 0.005$) (Table S2.6).

Ultimately, by integrating functional annotations obtained through literature searches, we show that the Paneth cell and goblet cell regulatory networks contain genes with direct and indirect associations with IBD. Furthermore, we find that the PCeE and GCeE networks and the targets of predicted master regulators are enriched with IBD associated genes - this finding is corroborated using human single cell data from UC colon biopsies. Consequently, these networks and the workflow to reconstruct and analyse them have great potential for the study of IBD pathomechanisms in specific intestinal cell types.

2.5 Discussion

By comparing the transcriptional signature of small and stranded RNAs in the CDEs to the cell type enriched enteroids, we identified genes with altered expression due to enteroid enrichment. I hypothesise that the larger quantity of DEGs observed in the Paneth cell enriched data represents the more diverse selection of environmental signals integrated by Paneth cells and the lesser role of the goblet cell in the small intestine, given that they are primarily located in the colon (Clevers and Bevins, 2013; Stappenbeck and McGovern, 2017). Functional overrepresentation analysis on the top 50 DEGs identified metabolic pathways associated with the PCeE-specific DEGs and cell cycle pathways associated with the GCeE-specific DEGs. The addition of further functional analysis methods such as Gene Set Enrichment Analysis and network-aware functional analysis, could enable the study of all identified DEGs and their direction of change, contributing further biological insight (Castresana-Aguirre and Sonnhammer, 2020; Subramanian et al., 2005).

Using cell type marker genes from a single cell sequencing paper (Haber et al., 2017), I showed that enriching enteroids for Paneth cells and goblet cells results in an increase in transcriptomics signatures from Paneth cells and goblet cells, respectively. The observation of additional enrichment of other secretory cell types, particularly enteroendocrine cells, likely reflects the shared differentiation pathways of these cells. In the future, the generation and application of cell signature lists from multiple different studies could reduce bias associated with using only one dataset (Haber et al., 2017). Such biases could have arisen, for example, from the age and gut microbiome of the mice and from the experimental and computational methods. In future analysis, evaluation of changes occurring in other cell types (e.g. enterocytes) as a result of differentiation skewing could further evidence cell type specificity of the altered differentiation protocols. Single cell sequencing of enteroids could be used for such an investigation, allowing comparison of gene expression of enterocytes from Paneth cell or goblet cell enriched enteroids to those from CDEs.

The enrichment of enteroendocrine cell signatures correlates with previous investigations of cell type enriched enteroids at both the transcriptomic and proteomic levels (Jones et al., 2019; Luu et al., 2018; Mead et al., 2018; Yin et al., 2014) and is important to consider when using these organoid models. Regardless, as enteroids contain a mixed population of cell types by nature

and because intercellular communication is key to a functioning epithelium (Sato et al., 2011; Thorne et al., 2018), the increased proportion of non-targeted secretory lineages should not be an issue for the application of these models to research. In fact, the enrichment of specific cell types is beneficial for enteroid-based research to increase the signal originating from a specific population of cells and to provide a larger population of cells of interest for downstream single cell analysis of enteroids, which is particularly beneficial when studying rare populations such as Paneth cells. The comparison of 'omics data from a cell type enriched enteroid to a CDE enables generation of cell type signatures with more specificity than can be obtained otherwise (e.g. from whole tissue biopsy samples) - except through single cell sequencing. It is possible to carry out single cell sequencing on enteroids, however this comes at a greater financial cost and provides lower coverage which can be problematic for rare cell types and lowly expressed RNAs (Brazovskaja et al., 2019; Jung and Jung, 2016). Organoid models are particularly valuable given the lack of *in vitro* models for long-term culture of non-self-renewing small intestinal epithelial cells (Chopra et al., 2010; Lukovac and Roeselers, 2015). A number of previous studies have shown that these cell type enriched enteroid models, which offer a simplified and manipulatable version of the intestinal environment, are useful for the investigation of health and disease related processes (Jones et al., 2019; Luu et al., 2018; Mead and Karp, 2019). Furthermore, Mead *et al.* recently showed that within cell type enriched enteroids, transcriptomic changes are well correlated to *in vivo* gene expression (Mead et al., 2018), supporting their use for *in vitro* studies. Through this presented work, I showed that cell type enriched organoids can be used to study cell type-specific regulation by comparing enriched to control organoids – although the specificity of the signature is not as great as would be achieved using a single cell sequencing approach.

Using an *a priori* universal network of non-specific molecular interactions, I annotated the DEGs with transcriptional and post-transcriptional regulatory connections. MiRNAs and lncRNAs were included in these networks as they have been shown to perform critical regulatory and mediatory functions in maintaining intestinal homeostasis (Chapman and Pekow, 2015; Farh et al., 2015; Mirza et al., 2015). For example, lncRNAs have been found to be differentially expressed in IBD and often co-localised with IBD-associated single nucleotide polymorphisms (Mirza et al., 2015). At least three separate studies have previously captured RNA profiles of healthy and/or diseased intestinal cell types; capturing mRNA, miRNA and lncRNA signatures (Haber et al., 2017; Mirza et al., 2015; Peck et al., 2017). However, we believe this is the first

comprehensive analysis of miRNAs, lncRNAs and TFs in conjunction with the genes and proteins they regulate which has been performed on a systems-level in a standardized manner. Unfortunately, only small proportions of the generated PCeE and the GCeE networks contained miRNA and lncRNA interactions, due to lack of published interaction information, particularly from murine studies. The addition of further 'omics data-types to the described approach could generate a more holistic view of cellular molecular mechanisms, including the ability to observe post-translational regulation. These networks will not contain every possible regulatory interaction within the cell type of interest but will contain interactions which are likely relevant to cell type-specific functions. For example, whilst regulators do not necessarily show strong co-expression with their targets, where co-expression exists, there is a greater chance that the association is functionally interesting. Therefore, we can use these networks to represent and analyse current biological knowledge as well as to generate hypotheses and guide further research. The *a priori* universal network approach to collating networks (regulatory or otherwise) has been used for a wide variety of research aims, such as the identification of genes functioning in a variety of diseases (Huang et al., 2018; Novarino et al., 2014), the prioritisation of therapeutic targets (Wachi et al., 2005) and for a more general understanding of gene regulation in biological systems (Kubisch et al., 2013; Yu et al., 2003). The application of prior knowledge avoids the need for reverse engineering / inference of regulatory network connections, which is time consuming, computationally expensive and requires large quantities of high quality data (Vijesh et al., 2013).

To investigate the substructure and functional associations of the generated PCeE and GCeE regulatory networks I applied a clustering approach. The identified clusters represent collections of highly interconnected nodes, which likely form regulatory modules. Functional analysis confirmed distinct functional associations between the clusters as well as between the networks. The observation that less than half of the network nodes exist in clusters is consistent with the view that regulatory networks are hierarchical and scale free with most genes exhibiting low pleiotropy (Barabási and Oltvai, 2004; Wagner and Zhang, 2011).

Given the observed additional increase in secretory lineages based on the DEGs of the enriched organoids, I chose to use cell type-specific markers to extract interactions specific to Paneth cells and goblet cells from the generated regulatory networks. This enables further enrichment of Paneth cell and goblet cell signatures and reduction in noise in the networks due to the presence of other cell types in the enteroids. Using this approach, I identified possible regulators of cell

type-specific functions in Paneth cells and goblet cells. Some of these regulators were predicted to be important in both cell types but exhibited differential targeting patterns between the PCeE and the GCeE networks, indicating rewiring of regulators between the cell types. Functional analysis of the targets of the most rewired regulators (ETV4, let-7e-5p, miR-151-3p, MYB and RORA) highlights an overrepresentation of metabolism associated targets in the PCeE network and cell cycle associated targets in the GCeE network. A similar result was observed when functional analysis was carried out on genes with significantly different expression levels between the cell type enriched enteroids and the CDEs (Figure 2.4B). This suggests that transcriptional changes during the skewing of enteroid differentiation could be driven by rewired regulators and that these functions are key features of Paneth cells and goblet cells in the enteroids. The latter is supported by current understanding that Paneth cells rely on high levels of protein and lipid biosynthesis for secretory functions (Cadwell et al., 2008) and play an important role in metabolically supporting stem cells (Rodríguez-Colman et al., 2017). Additionally, as terminally-differentiated cells do not undergo cell division, this result suggests that enteroid goblet cell signatures are derived from a large population of semi-differentiated goblet-like cells, a phenomenon previously observed in tissue sample based studies (Paulus et al., 1993; Smillie et al., 2019). This analysis highlights apparent redundancy and/or cooperation of regulators which control similar cell type-specific functions and shows the potential importance of regulatory rewiring in the evolution of cell type-specific pathways and functions, something which has been shown previously to occur (Davis and Rebay, 2017; Mendoza-Parra et al., 2016).

As an extension of the cell type-specific marker analysis, I identified putative cell type master regulators which control at least 50% of the cell type-specific markers in the PCeE and GCeE regulatory networks. Literature investigation highlighted that many of these regulators, particularly those associated with Paneth cells, have connections to autophagy, inflammation and IBD (Table 2.2). It is known that Paneth and goblet cell immune-associated secretory functions, which play a major role in gut homeostasis, are highly dependent on the cellular process autophagy (Cadwell et al., 2008; Jones et al., 2019; Patel et al., 2013; Stappenbeck, 2010). For example, it has recently been shown that the release of lysozyme from Paneth cells requires a form of autophagy called secretory autophagy, in which LC3+ vesicles containing lysozyme are routed to the apical surface of the cell for secretion (Bel et al., 2017). This secretory autophagy was shown to be disrupted in mice harbouring a Crohn's disease risk allele in the ATG16L1 gene, thus providing a link between secretion, inflammation, autophagy and IBD. In

addition, dysfunction of secretion and autophagy in Paneth cells and goblet cells has been associated with IBD through mechanisms such as reduced intracellular bacterial killing, increased endoplasmic reticulum stress and impaired secretion of mucus from goblet cells (Cadwell et al., 2009; Gersemann et al., 2009; Kaser and Blumberg, 2009; Patel et al., 2013). Furthermore, many IBD susceptibility genes are associated with autophagy function, further highlighting the link between autophagy, secretion and IBD (Lassen and Xavier, 2017). Ultimately, the literature associations with the predicted cell type-specific master regulators, particularly those relating to the Paneth cell, highlight the importance of autophagy, secretion and inflammation in Paneth cell and goblet cell function and suggest that dysregulation of key cell master regulators could lead to IBD.

To further investigate the IBD link, I identified Crohn's disease (CD) and ulcerative colitis (UC) genes in the PCeE and GCeE networks. I found that CD associated genes are more strongly associated with the PCeE network than the GCeE network. Given that Paneth cell dysfunction is classically associated with CD, this finding highlights the relevance of the generated networks to the *in vivo* situation. In the PCeE network one SNP associated CD gene, *Dbp*, acts as a regulator. *Dbp*, encoding the D site binding protein, regulates *Bik*, which encodes the BCL2 interacting killer, a pro-apoptotic, death promoting protein. Interestingly, rate of apoptosis has been implicated in IBD disease mechanisms (Nunes et al., 2014) and has been associated with IBD drug response (Aghdaei et al., 2018). Therefore, this finding highlights a possible regulatory connection between CD susceptibility genes and IBD pathology on a Paneth cell specific level. In the GCeE network, the SNP associated CD gene NOTCH2 acts as a regulator for *Notch3* and *Hes1*. It has been previously demonstrated that this pathway can block glucocorticoid resistance in T-cell acute lymphoblastic leukaemia via NR3C1 (predicted master regulator) (Real et al., 2009). This is relevant to IBD given that glucocorticoids are a common treatment for IBD patients (Prantera and Marconi, 2013). Additionally, the repression of *HATH1* by HES1 via Notch signalling has been previously associated with goblet cell depletion in humans (Zheng et al., 2011).

Furthermore, we identified a significant enrichment of UC associated genes in both the PCeE and GCeE networks. The majority of UC associated genes identified in the networks (9/14) were present in both, suggesting that genetic susceptibilities of UC do not have a Paneth cell or goblet cell specific effect. Two of the identified UC associated genes act as regulators in the networks (*Nr5a2* and *Hnf4a*), targeting hundreds of genes and thus suggesting a broad ranging effect on

the networks. Building on the identified literature associations of predicted master regulators, we found that the targets of Paneth cell master regulators are enriched with CD associated genes, and the targets of the goblet cell master regulators are enriched with UC associated genes. This finding was further illustrated using UC associated goblet cell genes from a human biopsy study (Smillie et al., 2019), highlighting the relevance of these findings in a human system. Ultimately, the observation of IBD susceptibility genes in the regulatory networks of these enteroids highlights possible application of this model system to the study disease regulation in specific intestinal cell types, through understanding specific mechanistic pathways. In addition, combined with patient genetic profiles, this approach could help to understand patient specific drug responses and identify new targets for drug actions.

2.6 Future research directions

We have shown how network biology techniques can be successfully applied to generate interaction networks representing the change in regulatory environments between two sets of enteroids, evidencing the value of organoids for cell type-specific studies. The described workflow could be applied to a variety of 'omics datasets and enteroid conditions. For example, to test the response of enteroids to external stimuli, such as bacteria, and on enteroids grown from human-derived biopsies, enabling patient-specific experiments. The application of further 'omics data-types to the described approach could generate a more holistic view of cellular molecular mechanisms, including the ability to observe post-translational regulation. In this study, we integrated miRNA and lncRNA expression datasets, in addition to mRNA data. However, only small proportions of the generated PCeE and the GCeE networks contained miRNA and lncRNA interactions, due to lack of published interaction information, particularly from murine studies. Both the application of human enteroid data and the future publication of high-throughput interaction studies involving miRNAs and lncRNAs will improve the ability to study such interactions. Instead, regulatory connections could be predicted using computational approaches, rather than relying on *a priori* resources. For example, one approach can draw insight about lncRNA-protein interactions using lncRNA-miRNA interactions and miRNA-protein interactions (Zhou et al., 2019). Another approach uses multilayer convolutional neural networks to predict miRNA interactions based on experimentally validated interactions (Zheng et al., 2020).

Extension of our workflow to single cell sequencing of cell type enriched enteroid cells would provide increased cell type-specificity. This could enable generation of cell type-specific regulatory networks (without noise from other secretory cell types) and, if applied without gating (e.g. 10x Genomics droplet-based systems (Zheng et al., 2017)), could prove or disprove that goblet cell signatures in enriched enteroids are derived from a large population of semi-differentiated goblet-like cells. Gating-based single cell sequencing has already been carried out by Mead et al. on Paneth cells from enriched enteroids, but network analysis was not applied to contextualise this data (Mead et al., 2018). Droplet-based single cell sequencing could add further understanding of the cell populations within enriched and conventional enteroids, but would require a large number of organoids to mitigate cellular complexity and batch heterogeneity and a powerful, reproducible and accurate computational pipeline to analyse the data (Chen et al., 2019).

Further work using knock-out mice or organoids could attempt to validate the importance of predicted master regulators in Paneth cells and goblet cells. However, this poses significant challenges due to their wide expression and broad function range. If the master regulators are controlling differentiation as opposed to cell function maintenance, evaluating lineage arrest or delay could be carried out using a gene knockout or knock down. However, the effects of pleiotropy will significantly hamper the results and such a study would require significant follow-up studies. On the other hand, if key regulators were predicted by applying the presented computational workflow to condition-specific organoids compared to control organoids (e.g. drug treated organoids vs non-treated organoids), the validation would be much simpler.

Given the functional associations, I predict that one of the most rewired regulators identified in this study, miR-151-3p, plays a role in the secretory pathway of goblet cells. This could be further investigated using knockout or knock down studies in enteroids. It's relevance in IBD could be studied using enteroids generated from IBD patient biopsies. Furthermore, investigation of other IBD-related predictions could be carried out using enteroids. For example, enteroids containing a knockout or knock down of one of the UC associated genes acting as regulators in the networks (*Nr5a2* and *Hnf4a*), could be used to evaluate the phenotypic effect on Paneth cells and goblet cells – for example using 'omics data and/or microscopy approaches. Alternatively, mice containing an epithelial knockout of the investigated gene could be studied for their susceptibility to intestinal colitis. Moreover, enteroids can be grown from IBD patient biopsies of patients harbouring the UC associated gene variant and Paneth cell or goblet cell

phenotypes could be observed compared to wildtype enteroids – alternatively biopsies themselves could be investigated. Similar studies could be carried out using the predicted Paneth cell and goblet master regulators. Such investigations would contribute mechanistic understanding of connections between Paneth cell and goblet cell function and IBD phenotypes.

In conclusion, we developed an integrative systems biology workflow to compare regulatory landscapes between enteroids from different conditions, incorporating information on transcriptional and post-transcriptional regulation. We applied the workflow to compare Paneth cell and goblet cell enriched enteroids to CDEs and predicted Paneth cell and goblet cell specific regulators, which could provide potential targets for further study of IBD mechanisms. Application of this workflow to patient derived organoids, genetic knockout and/or microbially challenged enteroids, alongside appropriate validation and single cell sequencing if available, will aid discovery of key regulators and signalling pathways of healthy and disease associated intestinal cell types.

Chapter 3: The effect of cytokines on the colonic epithelium

3.1 Introduction

Interactions between the immune system and the intestinal epithelium play an important role in the pathogenesis of chronic immune mediated inflammatory diseases, including inflammatory bowel disease (IBD). In IBD, debilitating symptoms and complications including abscesses and cancer, are associated with aberrant cytokine production and resulting intestinal epithelial damage. Despite the advent of biological therapies targeting key pathogenic cytokines, like tumour necrosis alpha (TNF α), fewer than 40% of IBD patients achieve complete disease control and mucosal healing (Cholapranee et al., 2017). Furthermore, classification of IBD and patient stratification is typically based on descriptive clinical parameters, which are poor predictors of patient trajectories. Therefore, new molecular insights are needed to understand the role of cytokines and intestinal epithelial cells (IECs) in IBD. In turn this knowledge should inform new treatment strategies and may aid stratification of patients for best use of existing treatment options.

This chapter describes an interdisciplinary and collaborative project which investigates the impact of cytokines on IECs using transcriptomics data. Using this data, we studied the signalling and regulatory responses of IECs to cytokines and compared cytokine-induced transcriptional signatures to patient biopsy transcriptional signatures, identifying a novel molecular classification of IBD. The data sources and analyses for this project are outlined in Figure 3.1. Specifically, we generated transcriptomics data from healthy colonic organoids (colonoids) which were treated with five IBD-relevant cytokines representing different T helper (Th) cell responses: interferon-gamma (IFN γ) and TNF α (Th1), interleukin (IL)-9 (Th9), IL-13 (Th2) and IL-17A (Th17). Differentially expressed genes were obtained by comparing each cytokine-treated colonoid to an untreated control colonoid dataset. This data was used in two main ways. First it was compared to transcriptional signatures of IBD patient colon biopsies, finding that IBD patient cohorts exhibit a gradient of cytokine-induced transcriptional changes which is significantly correlated with subsequent patient response to anti-cytokine treatments – providing a new method to stratify patients. Second, the colonoid data was used to generate

causal networks (termed cytokine-responsive networks) modelling the signal flow from recognition of a cytokine at the outer membrane of IECs through intracellular signalling pathways to the resulting transcriptional changes observed in the colonoids. This analysis also integrated molecular interactions from the databases OmniPath (protein-protein interactions, PPIs) and DoRothEA (transcription factor – target gene interactions, TF-TGs) and colon expression data from the Human Protein Atlas (Garcia-Alonso et al., 2019; Türei et al., 2016; Uhlén et al., 2015). Investigation and analysis of these networks revealed previously unrecognised levels of shared and distinct regulation by different cytokines. Further, I identified downstream regulatory bottlenecks, including Protein C-ets-1 (ETS1), in cytokine-responsive networks. Using IBD-patient transcriptomics data we demonstrated that expression of ETS1 in diseased colonic tissue is significantly associated with response to anti-cytokine treatments, thus providing a novel molecular biomarker and a potential therapeutic target.

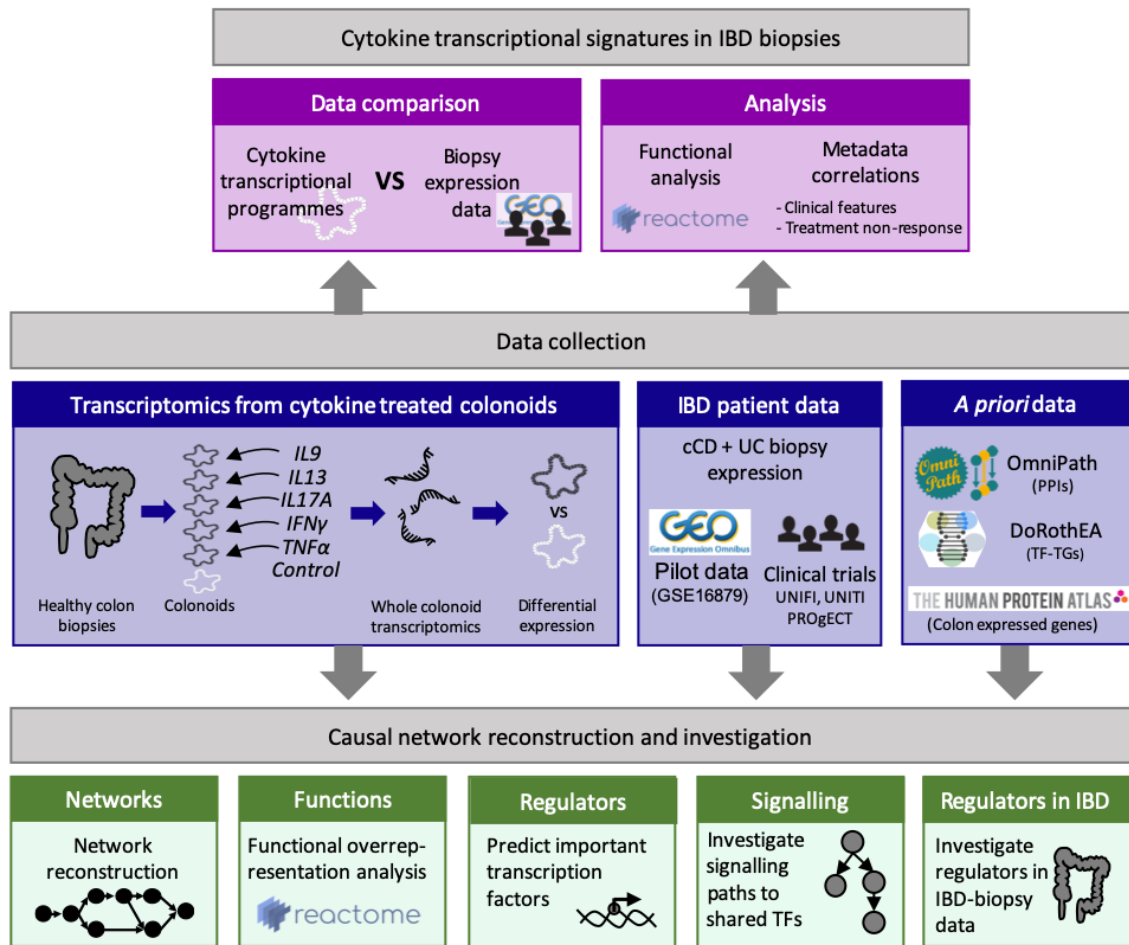


Figure 3.1. Schematic overview of the primary data sources and analyses carried out in Chapter 3. Initial analysis of colonoid transcriptomics data is covered in Results section 3.4.1 (Cytokines effect epithelial gene expression). Cytokine transcriptional signatures in IBD biopsies is covered in Results section 3.4.2 (Cytokine transcriptional signatures are enriched in IBD patient biopsies). Network reconstruction is in section 3.4.3 (Reconstructing cytokine causal networks in human colonoids). Network regulators and signalling is in Results section 3.4.4 (Cytokine-responsive signalling pathways converge at key transcription factors) and finally network regulators in IBD is in section 3.4.5 (ETS1 is a major regulator of the cytokine signalling in intestinal inflammation).

My role in this project was generation and analysis of cytokine-responsive networks, overlap of IBD signatures with cytokine profiles and the generation of text and figures to accompany this work. The following chapter has a greater focus on these aspects of the project. Organoid work, RNA sequencing, initial transcriptomics data processing and all other bioinformatics work was carried out by collaborators from Nick Powell’s research group in Kings College London. The contents of this chapter are primarily based on an article which is currently (September 2020) in review for publication (Pavlidis et al., 2020).

3.2 Aims

The aims of this project were as follows:

- Define the transcriptional effect of IBD-linked cytokines $\text{TNF}\alpha$, $\text{IFN}\gamma$, IL-9, IL-13 and IL-17A on colonic organoids from healthy human donors.
- Compare cytokine transcriptional signatures to transcriptional signatures from IBD colonic biopsies.
- Generate and investigate causal networks linking cytokines to their transcriptional effect, to identify similarities and differences between epithelial responses to different cytokines.

3.3 Methods

The following work was carried out in collaboration with clinical researcher Polychronis Pavlidis and other members of Nick Powell's research group at Kings College London (KCL). Generation and pre-processing of organoid data was carried out by them, as was the work comparing cytokine transcriptional programmes to a large collection of IBD patient biopsy data. I carried out the pilot analysis to identify cytokine transcriptional signatures in IBD patient biopsies using a small microarray dataset. Furthermore, all other work to reconstruct and investigate the causal networks was carried out by myself.

3.3.1 Transcriptomics data from cytokine-treated organoids

3.3.1.1 *Colonoid culture*

Isolation of human colonic crypts and subsequent establishment of human colonoids was performed as previously described previously by Fujii *et al.* (Fujii et al., 2015). During the last 24h of differentiation, human colonoids were treated with human recombinant IL-17A (50ng/mL), TNF α (10ng/mL), IFN γ (20ng/mL), IL-13(10 ng/mL) or IL-9 (10ng/mL).

3.3.1.2 *Next generation sequencing and analysis*

Harvested colonoids were put in Qiazol and then RNA was extracted with the RNeasy kit (Qiagen) as per manufacturer's guidelines. cDNA was created using the Revertaid cDNA synthesis kit (ThermoFisher). Sequencing libraries were generated using NEBNext[®] UltraTM RNA Library Prep Kit for Illumina[®] (NEB, USA) following manufacturer's recommendations. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using HiSeq PE Cluster Kit cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the paired end libraries were sequenced on an Illumina HiSeq platform.

Fastq files were preprocessed with in-house Perl scripts to carry out adapter and quality trimming. Read pairs were aligned to the human genome (GRCh37/hg19) using TopHat2 (v2.0.12) (Kim et al., 2013). HTSeq (v0.6.1) was used to count the read pairs mapped uniquely and concordantly to each gene (Anders et al., 2015). The raw count matrix was screened for genes with low expression levels across all samples (i.e. average count less than 3), and then normalized (Anders et al., 2015). Differentially expressed genes (DEGs) were identified through a varying intercepts hierarchical modelling approach (Gelman et al., 2014) implemented in R (R,

2018) and Stan (Carpenter et al., 2017). Gene counts were modelled as a negative binomial variable dependent on cytokine treatment as well as covariates accounting for repeated measurements from the same donor and additional sample similarities detected by PCA and hierarchical clustering. The quality of the estimated statistical model was assessed through posterior predictive simulations that compare replicated datasets to the actual data. The output p-values were corrected for multiple testing with the Benjamini-Hochberg method (Benjamini et al., 2001). Any gene with adjusted p value ≤ 0.1 was considered differentially expressed – no log fold change cut off was applied. Initial pathway analysis of DEG lists was performed with Ingenuity Pathway Analysis (IPA, Qiagen) (Krämer et al., 2014).

3.3.2 Cytokine transcriptional signatures in IBD patient biopsies

In the pilot work, I observed cytokine signatures in colonic CD (cCD, n=19) and UC (n=24) biopsies using microarray mucosal expression profiles from IBD patients prior to infliximab treatment from the Gene Expression Omnibus (accession GSE16879, n=6 controls) (Arijs et al., 2009). The microarray expression data was processed to obtain differentially expressed genes in cCD or in UC compared to healthy controls using GEO2R (Barrett et al., 2013). The differential expression analysis was carried out using patients as replicates. All genes with adjusted p value ≤ 0.01 were differentially expressed. The direction of fold change was not considered. Functional overrepresentation analysis and visualisation was carried out as described in section 3.3.4.

For the primary analysis, transcriptomics data and accompanying metadata was obtained from large clinical trials UNIFI, UNITI and PROgECT for active UC (n=702) and active cCD (n=126) patients prior to biologic treatment (Feagan et al., 2016; Sands et al., 2019; Telesco et al., 2018). Data was analysed per patient.

3.3.3 Reconstructing cytokine causal networks in human colonoids

The transcriptomics data from cytokine-exposed colonoids was used to generate cytokine causal networks through bespoke Python scripts. The networks consist of five separate node types and three separate interaction types, as shown in Figure 3.2.

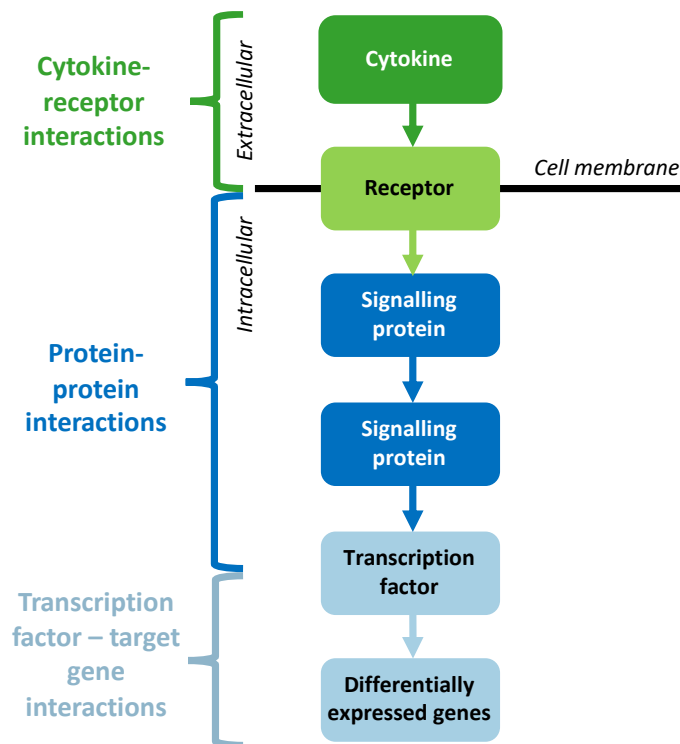


Figure 3.2. Schematic of node and interaction types included in the cytokine causal networks.

3.3.3.1 Cytokine-receptor interactions

Cytokine-receptor interactions were constructed using manually curated information on published cytokine receptors, as given in Table 3.1. To better represent the known biology of IL-13 signalling, connections between IL-13RA1 and IL-2RG or IL-4R were not included. Whilst these proteins do form complexes together, the downstream signalling from IL-2RG and IL-4R is not initiated by IL-13 binding (Hershey, 2003; Minton, 2008).

Cytokine symbol	Cytokine Uniprot ID	Receptor symbol	Receptor Uniprot ID
IFN γ	P01579	IFNGR1, IFNGR2	P15260, P38484
IL-22	Q9GZX6	IL-22RA1, IL-22RA2, IL-10RB	Q8N6P7, Q969J5, Q08334
IL-17A	Q16552	IL-17RA, IL-17RB, IL-17RC, IL-17RD, IL-17RE	Q96F46, Q9NRM6, Q8NAC3, Q8NFM7, Q8NFR9
TNF α	P01375	TNFRSF1A, TNFRSF1B	P19438, P20333
IL-13	P35225	IL-13RA1, IL-13RA2	P78552, Q14627

Table 3.1. All possible cytokine – receptor interactions for the causal networks, based on literature curation.

3.3.3.2 *Protein–protein interactions*

The signalling parts of the networks connect cytokine receptors to transcription factors through the shortest possible paths of protein-protein interactions (PPIs). Signalling connections between the proteins were obtained from the OmniPath database (v0.7.111) (Türei et al., 2016). OmniPath is a literature curated collection of human and rodent signalling pathways from a number of different sources, such as SignalLink, Signor and the Autophagy Regulatory Network (Fazekas et al., 2013; Korcsmáros et al., 2010; Licata et al., 2020; Perfetto et al., 2016; Türei et al., 2015). All paths consist of three or less intermediary signalling mediators making a total of six or less steps between the cytokine and the differentially expressed genes. All proteins of the signalling level, including the cytokine receptors, are expressed in the transcriptomics data for the relevant cytokine (FPKM > 0 in ≥ 2 replicates) and in the colon dataset from the Human Protein Atlas (v18.1) (Uhlén et al., 2015). The Human Protein Atlas contains a Tissue Atlas of expression profiles of human genes based on deep sequencing of RNA and antibody-based protein profiling using immunohistochemistry.

3.3.3.3 *Transcription factor–target gene interactions*

The transcriptional regulation level of the causal networks was generated by identifying and filtering transcription factors known to regulate the genes which were differentially expressed upon cytokine treatment of colonoids (adjusted $p \leq 0.01$). These transcription factor–target gene interactions were obtained from the published database DoRothEA v2 (Garcia-Alonso, 2018). This is a collection of signed transcription factor (TF) - target gene (TF-TG) interactions inferred from literature curated resources, ChIP-seq peaks, TF binding site motifs and from gene expression data. Interactions of confidence level A-D were used, which includes interactions from all sources except those derived only from computational predictions – as they have the lowest confidence level.

Nodes of the TF-TG interactions (TFs and DEGs) were filtered to remove any not expressed in the colon dataset from the Human Protein Atlas (Uhlén et al. 2015). In addition, to reduce the size of the network and to focus on the most important regulatory interactions, transcription factors were filtered using two further criteria. Firstly, only transcription factors exhibiting differential expression were included. Secondly, we applied an internal tool written by Matthew Madgwick (EI, QIB, Korcsmaros group; unpublished data) based on the Cytoscape app CHAT (Muetze et al., 2016). Here TFs were filtered for their influence in the network using the

transcriptomics data adjusted p values and the degree of the nodes. Specifically, the tool inputted the DoRothEA TF-TG network filtered for expressed nodes, alongside the list of all DEGs. Then, a hypergeometric significance test was carried out on any node with degree > 5 to determine if the proportion of connected nodes which are differentially expressed is higher than in the whole network. Any differentially expressed TF with adjusted p value ≤ 0.1 following Benjamini-Hochberg correction were deemed significant and used to filter the causal networks for the most influential TFs. ID conversion was carried out using Uniprot Swissprot to Ensembl id (GRCh38.p12) conversion, downloaded from Biomart on 11/4/19 (Smedley et al., 2015).

3.3.4 Functional analysis and visualisation

Functional enrichment of gene lists was carried out against Reactome pathways using the R package *ReactomePA* (Yu and He, 2016). Pathways reported when at least three genes of interested were identified in the pathway and when q value ≤ 0.1 .

All networks were visualised using Cytoscape (v3.7.1) (Shannon et al., 2003). Chord diagrams and circus plots were generated using custom Python and R scripts and the R library *circlize* (v0.4.6) (Gu et al., 2014). Venn diagrams were generated using the python package *venn* (v0.1.3).

3.4 Results

3.4.1 Cytokines effect epithelial gene expression

To measure the impact of cytokines on epithelial cells, human colonic organoids (colonoids) were treated with IBD-linked cytokines IL-9, IL-13, IL-17A, IFN γ and TNF α (as a pro-inflammatory control). Transcriptomics data was generated from each colonoid set and compared to an untreated control to obtain differentially expressed genes (DEGs) (determined by adjusted p value ≤ 0.1). In total 3,431 genes were differentially expressed upon treatment with the cytokines, some distinct and some shared between the different cytokines (Figure 3.3A). IL-13 had the greatest and most distinct impact on colonoid expression with fewer than a third of IL-13 induced DEGs shared with other cytokines. On the other hand, IL-9-treated colonoids were excluded from the analysis due to having a negligible impact on the colonoids (only 8 differentially expressed genes). Relatively few genes were affected by IL-17A expression, with >75% additionally targeted by other cytokines, suggesting redundancy in IL-17A control of epithelial function.

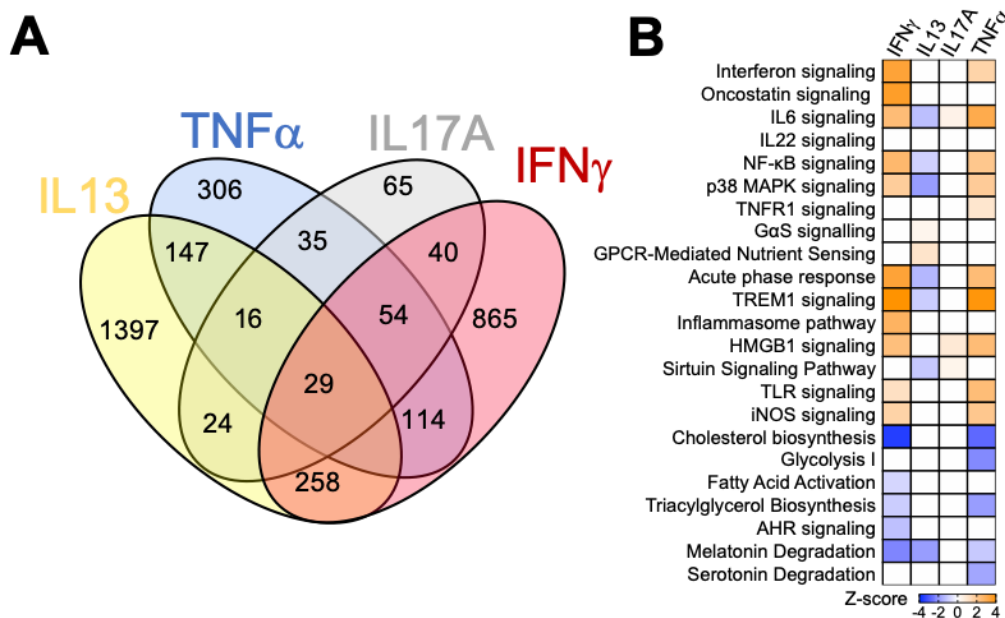


Figure 3.3. Differentially expressed genes upon cytokine treatment of colonoids. **A.** Overlap of differentially expressed genes across canonical cytokine transcriptional programmes. **B.** Regulation of key immunological and metabolic pathways by canonical cytokines in human colonoids. All differentially expressed genes determined as adjusted $p \leq 0.01$. Part A generated by myself. Part B generated by Polychronis Pavlidis (KCL) using Ingenuity Pathway Analysis software (Krämer et al., 2014).

Functional analysis of differentially expressed genes identified a number of cytokine-, immune- and inflammation-associated pathways affected by cytokine treatment (Figure 3.3B). Specifically, we found a similar pattern of pathway activation among IFN γ and TNF α while IL-13 appears to regulate these pathways in the opposite direction.

3.4.2 Cytokine transcriptional signatures are enriched in IBD patient biopsies

Having defined the transcriptional programmes of canonical cytokines in colonoids, we explored whether these signatures were enriched in diseased tissue of patients with IBD. To address this, we carried out two separate analyses. The pilot analysis, carried out by myself, compared cytokine-induced DEGs (cytokine programmes) to DEGs from colonic CD (cCD, n=19) and UC (n=24) biopsies using a publicly available microarray expression dataset (GSE16879, n=6 normal controls) (Arijs et al., 2009). Following this analysis, a further analysis was carried out using a much larger dataset with per-patient metadata available. This primary analysis, carried out by members of Nick Powell's research group (KCL), used transcriptomics data from colonic biopsies of active cCD (n=126) and UC (n=702) patients prior to biologic treatment with anti-cytokine drugs. This data was obtained from large clinical trials UNIFI, UNITI and PROgECT, and analysed per-patient (Feagan et al., 2016; Sands et al., 2019; Telesco et al., 2018).

For the pilot analysis, all DEGs previously observed using cytokine-treated colonoids were categorised based on the combination of cytokines which were shown to affect them. For example, the IL-13,IL-17A programme category contained genes which were differentially expressed upon colonoid treatment with IL-13 and IL-17A, but not IFN γ or TNF α . Between 24% and 50% of the genes in each cytokine programme category were also DEGs in the cCD and/or the UC biopsies (compared to healthy control) from the publicly available microarray dataset (Figure 3.4A, Table 3.1). In addition, I observed that all cytokine programme categories were activated at a similar level in the cCD and UC biopsies, with a large proportion shared between the two IBDs. For example, comparing the cytokine-specific programme categories (DEGs affected by only one cytokine) in cCD to those in UC, I found that roughly half of the genes were shared between cCD and UC (Figure 3.4). These results indicate a similar pattern of canonical cytokine programme activation between the IBDs.

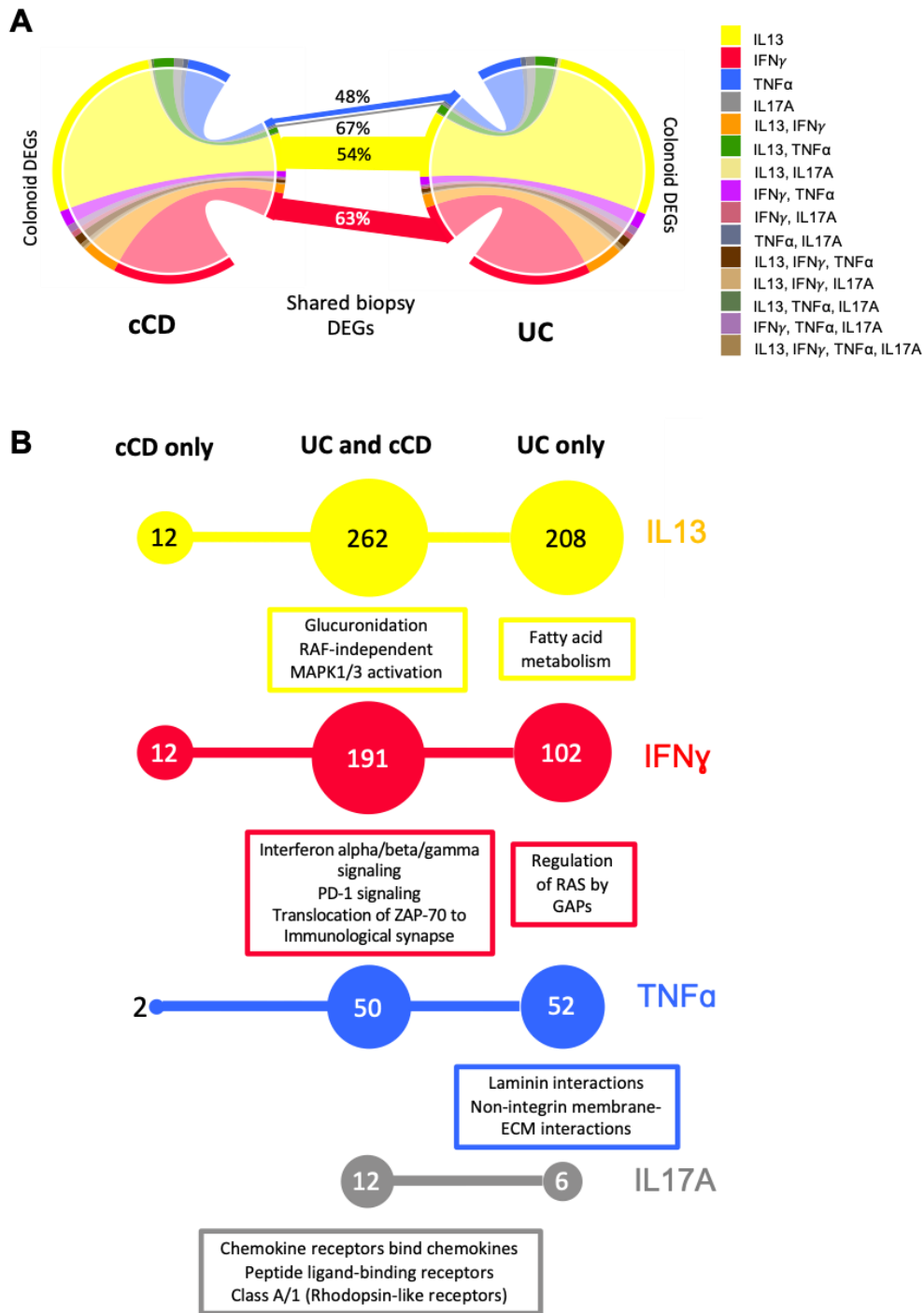


Figure 3.4. Overlap between cytokine programme categories and cCD and UC biopsy DEGs (from GSE16879). **A.** Cytokine programmes categories (e.g. IL-13, IL-17A) contain colonoid DEGs categorised by the combination of cytokines which affect them (q value ≤ 0.1). Width of bars on the circos plots represent the quantity of DEGs in the programme category. Width of horizontal lines joining the circos plots represent the proportion of the cytokine programme of cCD and UC which is shared between the two IBDs – only shown for the cytokine-specific programme categories. **B.** Cytokine-specific programme categories present in the cCD and UC biopsy data. Numbers represent the total number of differentially expressed genes (DEGs), size of circles proportional to $\log_{10}(\text{number of DEGs})$. Top three Reactome pathways shown where significant (q value ≤ 0.1) and where at least three DEGs are in the pathway. Similar pathways merged.

Next, I used pathway overrepresentation analysis to investigate the cytokine-specific programmes which were distinct and shared between cCD and UC (Figure 3.4B). I found that DEGs shared between colonoids, cCD biopsies and UC biopsies were associated with glucuronidation and MAPK activation (for IL-13), PD-1 and interferon signalling (for IFN γ) and chemokine and G-protein coupled receptors (for IL-17A). DEGs shared between colonoids and UC biopsies (but not cCD biopsies) were associated with fatty acid metabolism (for IL-13), RAS GTPases (for IFN γ) and extra-cellular matrix interactions (for TNF α). However, there were too few cCD-specific DEGs to determine functional associations.

Following the interesting results of the pilot study, members of Nick Powell's group (KCL), carried out a different analysis of cytokine programmes to identify patient-specific patterns of cytokine programmes, using a large (>1000) IBD patient dataset of diseased tissue expression data (with associated clinical metadata). Here they employed gene set variation analysis to test enrichment of the top 50 upregulated colonic organoid DEGs (top cytokine programmes for each treated cytokine) in the tissue expression data. Note that the top cytokine programmes were not categorised as in the pilot study. They identified significant enrichment of IFN γ and IL-17A top cytokine programmes in cCD patient biopsies compared with healthy control subjects. In active UC, they saw significant enrichment of IFN γ , IL-13, IL-17A and TNF α top cytokine programmes. Furthermore, using unsupervised hierarchical clustering they demonstrated a gradient of enrichment of top cytokine programmes, where some patients had simultaneous activation of all programmes, whereas others had minimal enrichment across all of the datasets investigated (Figure 3.5). This observation was consistent across multiple independent cohorts of cCD and UC patients. Interestingly, they observed weak correlations between the biopsy cytokine enrichment score (sum of all top cytokine enrichment scores) and clinical features, including total Mayo score, C-reactive protein, faecal calprotectin or disease duration.

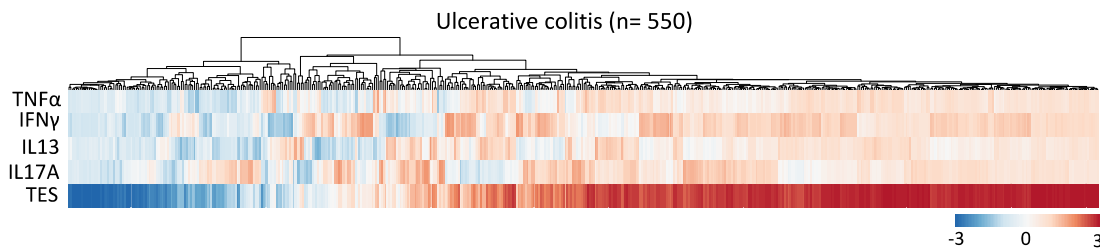


Figure 3.5. Gradient of top cytokine programme activation in IBD. Each column represents a single patient. The sum of all four scores per subject is also depicted as the total enrichment score (TES). Columns have been clustered by Euclidean distance (method: average, tree ordering: original, figure generated with ClustVis). Data from cohorts: UNITI (n=126), UNIFI (n=550), PROgECT (n=152). The top cytokine transcriptional programme is the top 50 upregulated colonic organoid DEGs for each cytokine-treated colonoid dataset. Figure generated by Polychronis Pavlidis (KCL).

In addition, my collaborators used further biopsy transcriptomics data to investigate whether simultaneous activation of multiple cytokine response pathways in individual cCD patients could explain resistance to biological therapies, where individual cytokines are selectively targeted. Here, they found that stratification of the cohorts according to cytokine enrichment score (in biopsies sampled at baseline), predicted their subsequent response to either anti-TNF α treatment, infliximab, or to anti-IL-12p40 treatment, ustekinumab. Patients with high total enrichment scores were very unlikely to respond to treatment, whereas patients with low total enrichment scores were highly unlikely to respond. A similar observation was found for active UC patients, although the predictive power was less differentiating than in cCD.

Taken together, our analyses show that diseased colonic tissues of IBD patients are enriched with cytokine transcriptional programmes - with some similarities and differences observed between cCD and UC. Further, our data highlights that gradients of cytokine programmes in IBD patients provides a previously unrecognised molecular classification of IBD, which can be harnessed to predict response to anti-cytokine therapy.

3.4.3 Reconstructing cytokine causal networks in human colonoids

Next, I used network biology to further explore the impact of cytokine cues on colonic epithelial cells using the colonoid transcriptomics data. Given that we found that non-response to anti-cytokine treatments targeting a single cytokine is associated with the simultaneous activation of multiple cytokine transcriptional programmes, we aimed to identify shared and distinct wiring connections in epithelial cells as a result of cytokine cues. Here, I generated causal networks for

each cytokine, predicting molecular signal flow from the recognition of a cytokine at the IEC surface, through intra-cellular signalling pathways to the observed changes in DEGs (genes differentially expressed in cytokine-treated colonoids compared to untreated colonoids) (Figure 3.6A). These networks integrated observed transcriptional changes in the cytokine-treated organoids with using *a priori* knowledge of PPIs and TF-TG interactions available in published databases of experimentally-verified molecular interactions. These networks provide a systems-level, mechanistic understanding of cytokine-mediated regulation of epithelial function (Figure 3.6B).

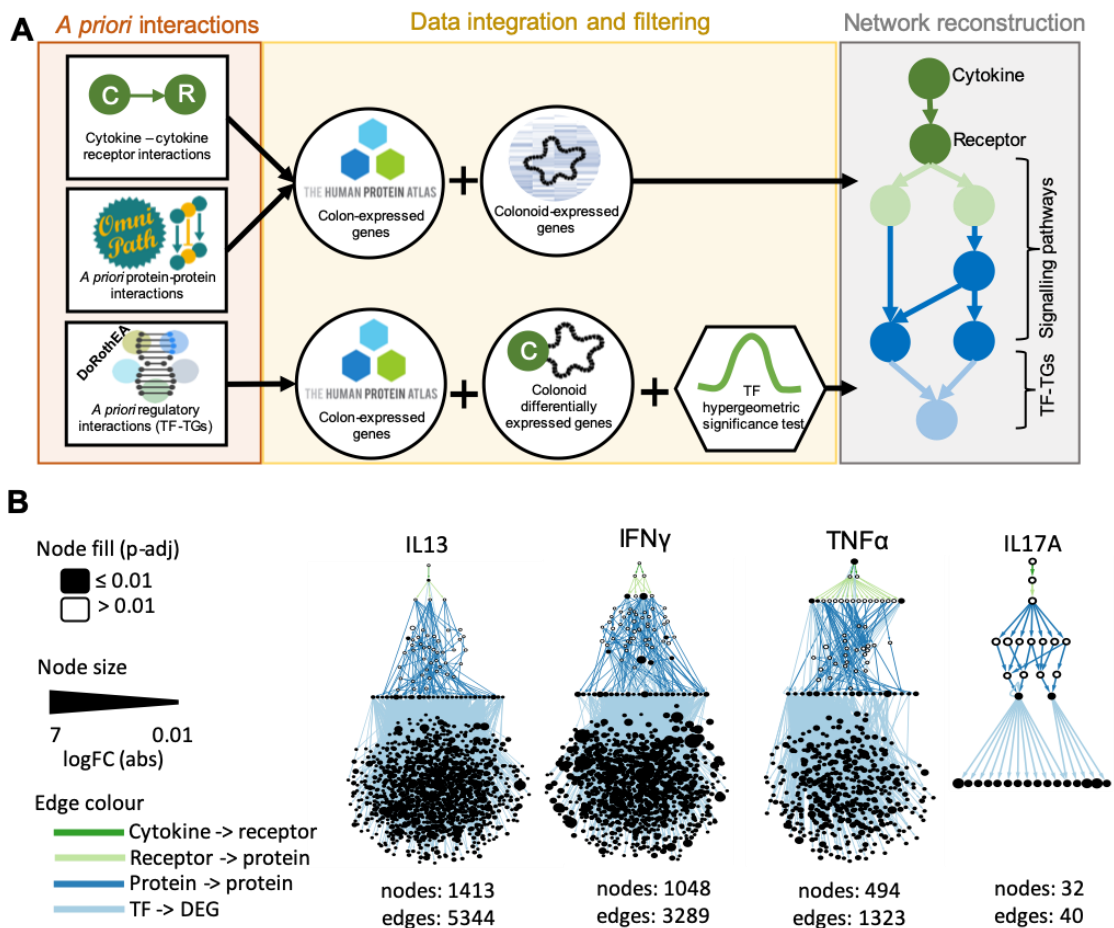


Figure 3.6. Reconstructing causal networks of signalling and regulatory molecular interactions connecting cytokines to observed differentially expressed genes within the colonic organoids. **A.** Workflow of causal network reconstruction for each cytokine. See Methods section 3.3.3 for more detail. **B.** Generated causal networks. One directed network for each cytokine tested. Colours represent the levels/type of interactions of the networks. Arrows represent direction of signal transduction. Although some nodes are present in multiple levels of the network, they are only displayed once in first occurring level (nearest to the cytokine). All differentially expressed genes determined as adjusted p value ≤ 0.01 . TF - transcription factor; DEG - differentially expressed gene.

A number of filtering criteria were applied to select only the most likely signalling or regulatory paths (see Methods section 3.3.3). Due to these criteria and to reliance on published molecular interactions, some DEGs had no identified upstream paths to their associated cytokine, and were thus excluded (Table 3.2). Differences in network sizes represent variation in total number of DEGs and the variation in number of direct pathways identified between the cytokine and the DEGs (Figure 3.6B). For example, the IL-17A network is smaller, in part because fewer DEGs were observed upon organoid treatment, but also because very few DEG targeting TFs were identified which are also differentially expressed (one of the filtering criteria applied). This shows that IL-17A has a lesser effect on the transcriptomic landscape of organoids than the other cytokines, including at the TF level. On the other hand, IL-13 treatment of organoids resulted in the largest observed number of DEGs, many of which are not targeted by any of the other cytokines (Figure 3.7). Correspondingly, the IL-13 network contains the largest number of nodes and edges with 1413 and 5344 respectively. Only 27.8% of these nodes are also present in one or more of the other cytokine networks.

Cytokine	# DEGs (p adj ≤ 0.01)	# DEGs in network	% of all DEGs in network
IL-13	1952	1371	70.24
IFN γ	1441	1032	71.62
TNF α	766	452	59.01
IL-17A	279	17	6.09

Table 3.2. Number of differentially expressed genes upon cytokine treatment of colonoids and in the resulting causal networks. DEG = differentially expressed gene.

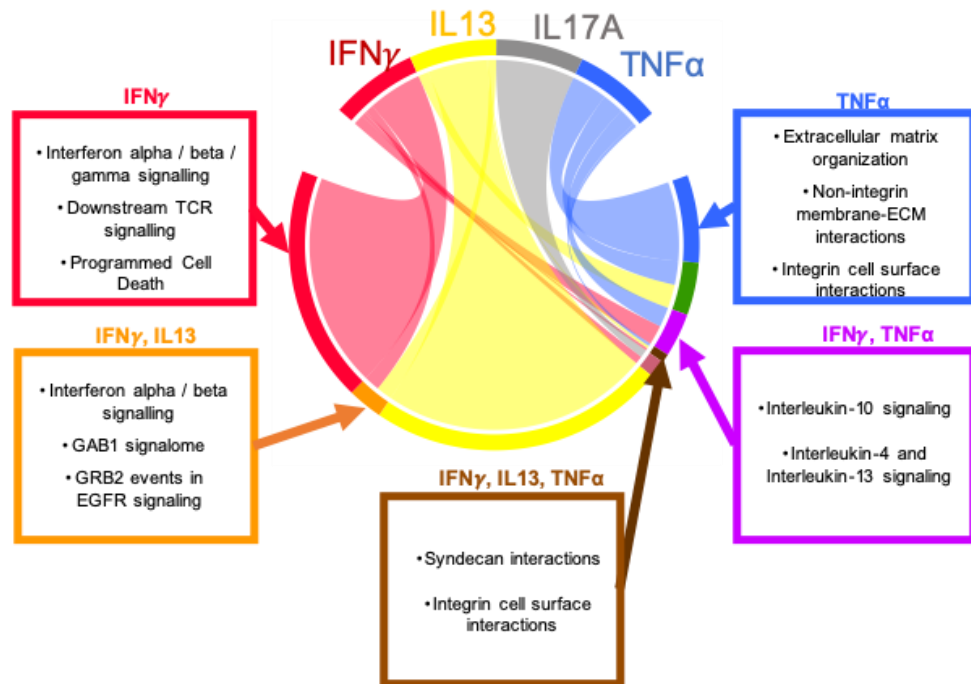


Figure 3.7. Chord diagram visualisation of differentially expressed genes (DEGs) targeted by each cytokine in the causal networks, with functional associations. The cytokines are given in the upper part of the figure. The targeted DEGs are given in the lower part of the figure. The length of the lower bars and the width of the connecting lines represents the number of DEGs. The colours represent which networks each DEG/cytokine is in. The boxes show the top three enriched Reactome pathways associated with each group of DEGs. Reactome pathways only shown if total DEG group size is 3 or more and where adjusted p value ≤ 0.1 (Fabregat et al., 2018a). Where pathways are very similar, they have been reported together.

Next, I carried out functional overrepresentation analysis on the DEGs targeted by each collection of cytokines in the causal networks. Only five categories of DEGs were significantly associated to any Reactome pathways, with results shown in Figure 3.7. The identified pathways were mostly associated with cytokine signalling, for example DEGs targeted solely by IFN γ and those targeted by IFN γ and IL-13 were associated most strongly with interferon signalling. DEGs targeted by IFN γ and TNF α were associated with interleukin signalling, indicating overlap between effects of different types of cytokines. Additionally, TNF α -specific DEGs were associated with extracellular matrix organisation and cell surface interactions. DEGs shared between by IFN γ , IL-13 and TNF α were also associated with cell surface interactions, suggesting a role for the extracellular matrix in the response of epithelial cells to cytokines.

3.4.4 Cytokine-responsive signalling pathways converge at key transcription factors

Next, I sought to compare the nodes (proteins, TFs and DEGs) and interactions (node interactions) of the causal networks to identify shared and unique features of IEC responses to different cytokines. Comparing the nodes of each network, 20.4% of all network nodes (total 499) were shared between at least two of the cytokine causal networks (Figure 3.8A). This is comparable to the proportion of DEGs shared by two or more cytokines (23.2%, Figure 3.3A). Only five nodes (<0.3%) are present in all of the networks (Figure 3.8A). However, none of these nodes have a strong link to IBD:

- Potassium voltage-gated channel subfamily H member 2 (KCNH2) - present as a DEG in every network.
- Plasminogen activator urokinase (PLAU) - present as a DEG in every network.
- Plexin D1 (PLXND1) - present as a DEG in every network.
- E2F Transcription Factor 2 (E2F2) - present as a DEG in every network, except the IL-17A network where E2F2 acts only as a TF.
- Mitogen-activated protein kinase 14 (MAPK14) - not differentially expressed in any cytokine treatment but present as a signalling protein in every network.

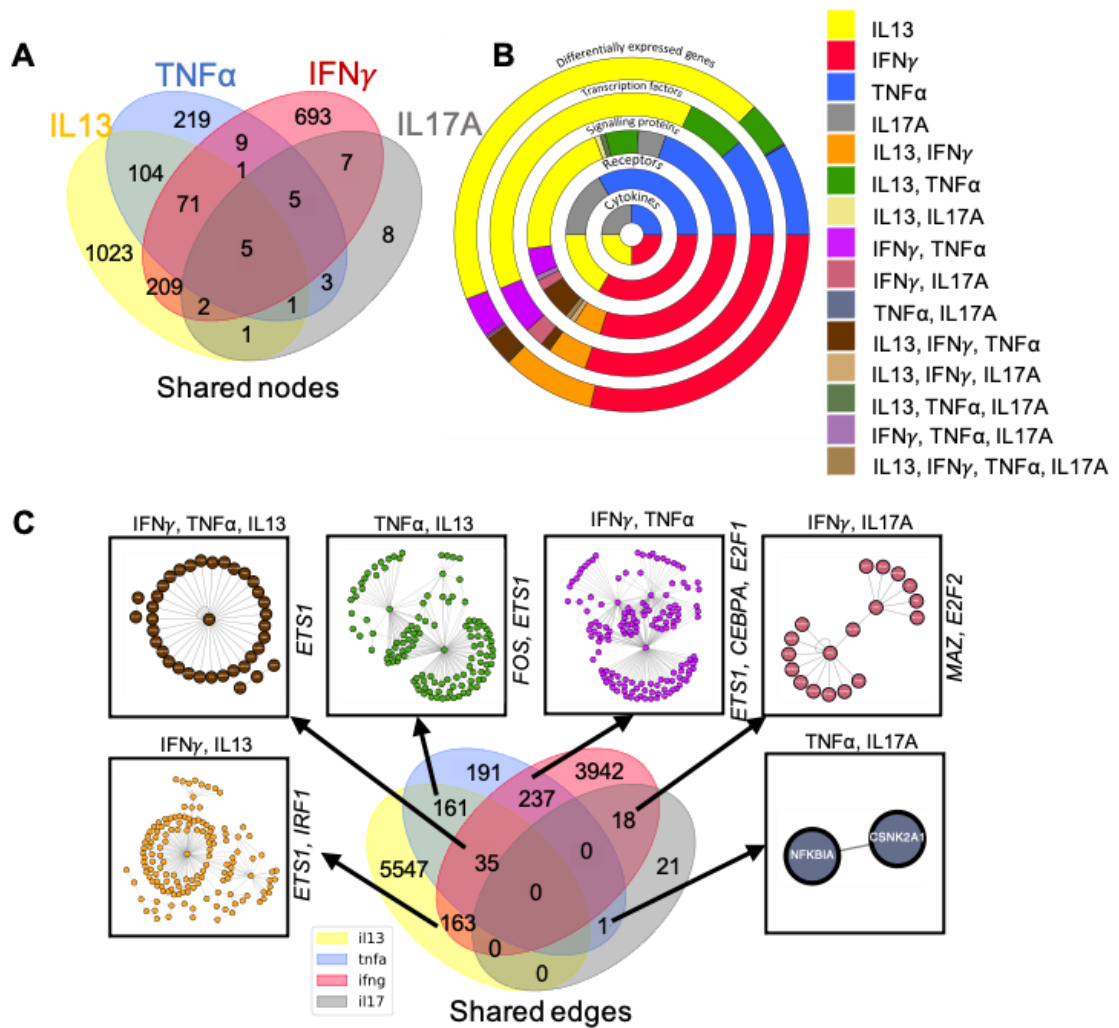


Figure 3.8. Overlap of nodes and edges in the four cytokine causal networks. **A:** Venn diagram showing overlap of nodes in the networks (not to scale). **B:** Circus plot of all network nodes, where each ring represents one level of the causal networks, with cytokines in the middle and the differentially expressed genes at the edge. The coloured bands represent the proportion of all nodes in the level which belong to a specific category. The categories, identified by colour, are defined by which causal networks the given node is present in. **C:** Venn diagram showing overlap of edges (interactions) in the networks (not to scale) with network visualisations of all edges shared between two or more of the networks. Where relevant, the transcription factor responsible for the majority of the shared interactions is given at the right side of the network images.

After categorising nodes by their network layer (signalling proteins, TFs and DEGs), the proportion of nodes shared with more than one cytokine network is fairly consistent (Figure 3.8B). This is could be due to an overlap of epithelial signalling and regulatory pathways initiated by different cytokines. Furthermore, I investigated the edges (interactions) shared between the

cytokine causal networks (Figure 3.8C). I found that the majority of interactions shared between networks are involving seven different transcription factors and their DEG targets:

- Protein C-ets-1 (ETS1)
- Interferon regulatory factor 1 (IRF1)
- CCAAT enhancer binding protein alpha (CEBPA)
- MYC associated zinc finger protein (MAZ)
- FOS proto-oncogene (FOS)
- E2F transcription factor 1 (E2F1)
- E2F transcription factor 2 (E2F2)

In particular, TF ETS1 plays the largest role, with interactions shared between the IL-13, TNF α and IFN γ causal networks. Most of these TFs (6/7) were found in the IFN γ network, suggesting that IFN γ shares regulatory mechanisms with many other cytokines. Taken together these networks highlight that the cytokines have shared and unique effects on signalling within the colonic organoids and that TFs could play a key role in the intersection between different cytokine-induced signal flows.

To further investigate the importance of TFs cytokine causal networks, I visualised the TFs targeted by each cytokine (Figure 3.9A). I found that the majority of transcription factors are specific to a cytokine, while 12 TFs are targeted by more than one cytokine (present as a TF in more than one network). Specifically, these 12 shared TFs constitute 21%, 37.5%, and 42% of all TFs affected by IL-13, IFN γ , or TNF α , respectively, indicating a substantial overlap among the downstream regulators of these cytokines. Meanwhile the IL-17A causal network only contains 2 TFs which are both also targeted by IFN γ .

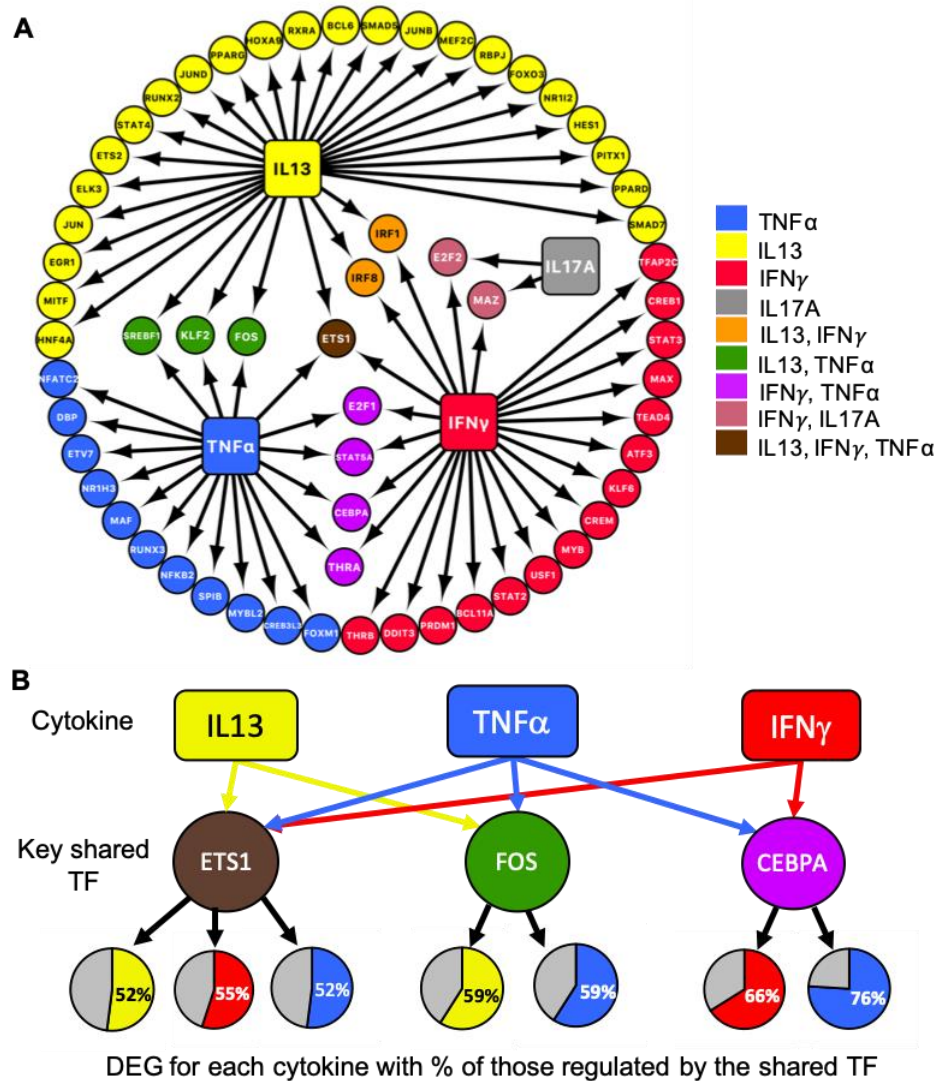


Figure 3.9. Transcription factors (TFs) shared between multiple cytokine causal networks. A. Network visualisation of TFs targeted by multiple cytokines (shared TFs). Cytokines are larger square nodes, transcription factors are smaller circular nodes. **B.** ETS1, FOS and CEBPA TFs regulate the majority of DEGs in the IL-13, IFN γ and TNF α cytokine causal networks.

I investigated DEGs regulated by the shared TFs in each network to understand which TFs could interfere with the transcriptional program of another cytokine. I found that the majority of DEGs in each network were regulated by at least one shared TF (IL-13, 81.8%; IFN γ , 93.8%; TNF α , 99.5%; IL-17A, 100%), indicating that signals from one cytokine could have a significant impact on another cytokine’s transcriptional programme, through shared TFs. Moreover, I quantified the network DEGs targeted by each of the shared TFs separately, and found that three of the 12 shared TFs (ETS1, CEBPA, FOS) regulated the majority of all the DEGs in colonoids (Figure 3.9B).

This signifies a particularly important role for these three TFs in the regulation of multiple concurrent cytokine responses in IECs.

Given the likely importance of shared TFs, I explored whether they are activated through the same pathways in the different causal networks. I found that cytokine receptors do not act through shared signalling pathways, but instead have (mostly) unique signal transduction paths which converge on the key shared TFs. As an example, Figure 3.10 depicts the signalling pathways which lead to ETS1, IRF8 and IRF1. Furthermore, many of the predicted signalling proteins connecting cytokines to their targeted genes are not differentially expressed (Figure 3.6B). This suggests that signals are propagated through the networks by post-translational modifications, such as phosphorylation of JAK/STAT proteins.

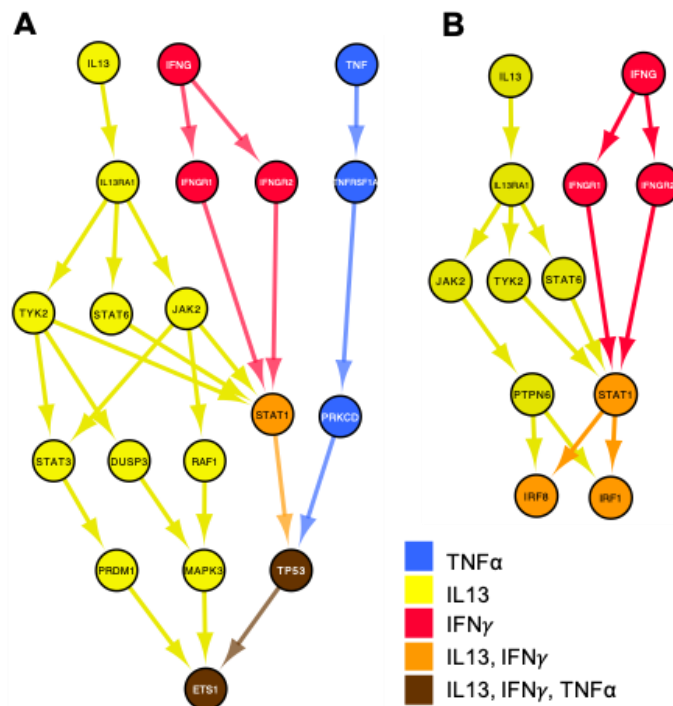


Figure 3.10. Signalling pathways linking cytokines to shared transcription factors. From causal networks. Node and edge colour indicate which cytokine causal networks the node/edge is present in. **A:** IL-13, TNF α and IFN γ signalling pathways to ETS1. **B:** IL-13 and IFN γ signalling pathways to IRF8 and IRF1.

Taken together, these findings show that cytokines have independent signalling mechanisms which affect cytokine-specific TFs as well as converge on key shared TFs. The shared TFs, particularly ETS1, FOS and CEBPA, affect large proportions of the cytokine transcriptional

programmes, and could partly explain non-response to single cytokine-targeting therapies in patients with high activation of multiple cytokine programmes.

3.4.5 ETS1 is a major regulator of the cytokine signalling in intestinal inflammation

Finally, Polychronis Pavlidis (KCL) tested the functional relevance of the 12 shared TFs by exploring their expression in the colon of patients with and without IBD. Using principal component analysis, he showed that IBD patients with active colonic inflammation are completely differentiated from control patients based on the expression of these shared TFs. Unsupervised hierarchical clustering based on expression levels demonstrated that these TFs tended to cluster with the cytokines TNF α , IFN γ , IL-17A, and their receptors. Furthermore, using expression of *ETS1*, *FOS* and *CEBPA* in a multivariate logistic regression model to predict response to anti-TNF treatment (infliximab) in IBD, identified ETS1 as the single predictive TF. Specifically, he found that non-responders had higher expression of *ETS1* prior to commencement of infliximab, with an area under the curve predicting response of 0.82, 95%CI (0.69, 0.96), $p=0.0003$.

These observations further evidence that activation of multiple concurrent cytokine transcriptional programmes drives resistance to therapy, whilst further highlighting ETS1 as a potential therapeutic target in IBD and as a biomarker to stratify patient response to treatment.

3.5 Discussion

To our knowledge this is the first study describing the transcriptomic landscape of the immune-epithelial interactome in the gut. By treating healthy colonoids with canonical cytokines we identified shared and distinct patterns of regulated gene expression in the epithelium. Combining cytokine-mediated transcriptional data with network biology methods and diseased tissue transcriptomics data, we identified a novel molecular classification of IBD based on gradients of cytokine transcriptional programmes and predicted the TF ETS1 as a potential biomarker and therapeutic target in intestinal inflammation.

The large epithelial response to IL-13 in comparison to other tested cytokines is unexplained but indicates the importance of IL-13-epithelium interactions. Conversely, IL-9 showed almost no impact on the colonoids, suggesting that IL-9 does not significantly affect epithelial cells at the given concentration, or that it requires co-stimulatory molecules or alternatively, that it acts only on specific cell types whose signature is masked by whole colonoid sequencing (Gerlach et al., 2014).

Pathway analysis revealed many inflammatory pathways associated with the cytokine-induced DEGs, which is expected as a cellular response to cytokines. IFN γ and TNF α show similar patterns of pathway activation/inactivation indicating significant overlap in colonic epithelial response to these cytokines. Given only 12% of the DEGs upon IFN γ or TNF α treatment are shared between the two conditions, the observed pathway overlap suggests synergism between the cytokines. This synergism has been previously described in the context of epithelial cell proliferation and apoptosis (Nava et al., 2010). Many pathways affected by IL-13 are activated/inactivated in the opposite direction to the other cytokines, suggesting an anti-inflammatory role of IL-13 on epithelial cells. However, it has been shown previously that IL-13 exposure of HT-29/B6 colonic epithelial cells results in increased apoptosis and reduced transepithelial resistance, suggesting that IL-13 plays both pro- and anti-inflammatory roles (Heller et al., 2005).

When mapping cytokine transcriptional programmes identified in the colonoids to pilot colonic biopsy transcriptomics data from a small number of IBD patients, I found a highly conserved pattern of expression in patients with UC and cCD. This indicates that the immune landscape of colonic inflammation of both these diseases, at least at a transcriptional level, is very similar. It

also shows that no particular cytokine profile characterises UC or cCD, challenging the Th1/Th2 depiction of IBD (Fuss, 2008; Imam et al., 2018). However, thorough investigation of the cytokine-specific programme categories identified in the biopsy datasets highlighted a functional difference between UC and cCD. Specifically, I found that IL-13-responsive genes associated with fatty acid metabolism were present in the UC dataset but not the cCD dataset, suggesting a role of IL-13 in fatty acid metabolism in UC. Both UC and cCD have been previously connected to fatty acids and levels of fatty acids have been shown to regulate cytokine levels (Heimerl et al., 2006; Scoville et al., 2018), but as far as I am aware, no publication has yet linked IL-13 signalling to downstream altered fatty acid metabolism in UC or cCD. On the other hand, a direct link between TNF α and the extra-cellular matrix in UC has been previously published (Wang, 2007). Further investigation into the other functional associations in connection with their targeted cytokine could reveal mechanistic differences between UC and cCD. It should be noted however, that in this small biopsy dataset most of the cCD-associated genes were also present in the UC dataset, and very few were specific to cCD. Therefore, application of further biopsy datasets would strengthen this functional analysis.

Additional investigation of cytokine transcriptional programmes was carried out using independent cohorts making up the biggest dataset of IBD tissue ever analysed (>1000 samples). This analysis further challenges the outdated T-cell lineage identity paradigms in IBD patient classification, by demonstrating that IBD patients can be stratified based on combined enrichment of cytokine-induced transcriptional programmes. This finding corroborates recent understanding that IBD involves 'polyfunctional' T-cell responses, in which individual T-cells co-produce different cytokines, such as TNF α and IL-17A (Langrish et al., 2005; Tang and Iwakura, 2012). Importantly, this finding might explain the success or failure of some of commonly used anti-cytokine therapies which blockade a single cytokine and thus would be unlikely to antagonize multiple concurrent pathways. Moreover, this novel classification system using cytokine transcriptional programmes could be harnessed as a precision medicine tool to stratify patients according to their likelihood of responding to anti-cytokine therapies. Those patients with high cytokine enrichment scores could be offered alternative therapies such as accelerated surgery or dual administration of biologics targeting different cytokines. Indeed, clinical trials evaluating blockade of TNF α and IL-23 simultaneously have already begun (NCT03662542).

Causal networks were constructed to further explore the effect of cytokines on epithelial cells at a systems-level. These networks represent the most likely pathways of action but will not include every possible molecular link and may contain false positive interactions. Moreover, the network reconstruction method employed here does not correct for bias due to common hubs and assumes that pathways are linear and short, which is often not true. Hubs are present in *a priori* protein-protein interaction resources due to study bias based on research interest and on bias of experimental techniques (Schaefer et al., 2015). Correction of hub bias can occur during network investigation, for example by using the Contextual Hub Analysis Tool (CHAT) tool to identify important hubs using 'omics data rather than only by degree (Muetze et al., 2016). Alternatively, hub bias can be mitigated during network reconstruction through the use of heat diffusion algorithms such as HotNet2 and TieDIE (Leiserson et al., 2015; Paull et al., 2013; Vandin et al., 2012). A diffusion algorithm approach could also improve on the presented method by avoiding the assumption that signalling pathways are short and linear.

In addition to possible missing interactions in the *a priori* interaction sources, network nodes, particularly TFs, were heavily filtered to prioritise those of most importance. Such a filtering method may bias the networks by including only TFs which are themselves transcriptionally regulated while excluding those regulated by posttranslational modifications and protein-protein interactions (Tootle and Rebay, 2005). Moreover, protein activation levels cannot be directly inferred from expression levels. A number of tools have been developed which could improve the presented method by applying a different TF filtering method. For example, algorithms such as VIPER could be used to infer TF activity based on regulon expression and pleiotropy (Alvarez et al., 2016).

We observed that most signalling proteins in the networks were not differentially expressed, indicating that signal transduction occurs *via* post-translational modifications. This also implies a possible lack of negative feedback loop in the system, resulting in cytokine pathways being prone to overactivation. Such loops are often seen in stress response signalling, whereby pathway protein members are downregulated as a result of the signalling flow (Hua et al., 2009). As activation levels of proteins cannot be directly inferred from expression levels, the addition of proteomics data to the study could uncover important post-translational regulation and improve the prediction of signalling pathways. However, overall, the size and overlap of causal networks is well correlated with the DEGs, indicating that biological signals have been retained.

Comparison of the cytokine causal networks highlighted previously unrecognised levels of shared and distinct characteristics, demonstrating redundancy and complementarity in cytokine-epithelial interactions. Specifically, I identified a collection of seven TFs which drive the majority of interactions shared between multiple networks, forming potential molecular bottlenecks of cytokine signalling. Literature searches reveal that all of these TFs have documented associations with immunomodulation, with many of them known to regulate and/or be regulated by cytokines. For example, IRF1 has a documented role in response to cytokine signalling, particularly IFN signalling, leading to immunomodulation (Honda and Taniguchi, 2006; Kröger et al., 2002). A recent study found that IL-13 can act through IRF1 to induce apoptosis of neonatal Th1 cells (Miller et al., 2019). E2F1 and E2F2 are regulators of cell cycle and have been shown to act in feedback loops with cytokines and drive T cell proliferation (DeRyckere and DeGregori, 2005; Ertosun et al., 2016; Zhang et al., 2018). MAZ, which is highly expressed in UC and colon cancer, has been shown to be a critical driver of inflammation through Signal transducer and activator of transcription 3 (STAT3) signalling (Triner et al., 2018).

Furthermore, three of these TFs were shown to regulate more than 50% of the cytokine-induced programme for each affecting cytokine, indicating their centrality and importance in cytokine responses. ETS1 (targeted by IL-13, TNF α and IFN γ) has a known role in cytokine and chemokine gene regulation in addition to other functions such as stem cell development and proliferation of lymphoid cells (Dittmer, 2003; Russell and Garrett-Sinha, 2010). Studies have shown that ETS1 is linked to the regulation of Jak-Stat signalling, one of the foremost pathways downstream of cytokine receptors (Murray, 2007). *ETS1* has also been identified as a susceptibility gene for IBD and is known to affect development of T-cells (Li et al., 2018). In addition, *Ets1* knockout mice show decreased expression of cytokines including IL-13, TNF α and IFN γ in T cells, indicating a role in positive feedback loops of cytokine expression (Grenningloh et al., 2005; Russell and Garrett-Sinha, 2010). Interestingly, IL-17A showed increased expression in a similar experiment, suggesting a different mechanism of signalling - possibly explaining the lack of ETS1 in the IL-17A causal network (Moisan et al., 2007). Furthermore, both CEBPA and FOS have documented links to immunomodulation. For example, CEBPA is involved with cellular differentiation and response to inflammatory insult and a genetic region involving the *Cebpa* gene was identified as susceptibility locus for early onset IBD (Imielinski et al., 2009; Lekstrom-Himes and Xanthopoulos, 1998). FOS has been shown to regulate degranulation and cytokine production in mast cells (Lee et al., 2004).

Our novel analytical workflow allowed us to discover potential molecular bottlenecks downstream of the canonical cytokines' engagement with their relevant receptor. Ultimately, the strong literature links between the identified TF bottlenecks and cytokines evidences the biological relevance of the generated causal networks and highlights their potential for identifying key drivers of cytokine mediated inflammation.

Finally, the importance of *ETS1* in regulating immune-epithelial interactions was evidenced through validation against whole tissue transcriptomics data from IBD patient colonic biopsies. Specifically, Polychronis Pavlidis (KCL) found that a higher expression of *ETS1* was significantly associated with subsequent non-response to anti-TNF α drug infliximab. Additionally, he found that those patients with high *ETS1* expression also had high *IFN γ* expression, which further evidences why anti-cytokine treatments targeting single cytokines may not be effective in some patients. Importantly, this finding offers a novel patient stratification method whereby *ETS1* expression within whole biopsies samples could be used to predict patient response to infliximab. In turn, this could allow clinicians to personalise treatment options, offering different or more advanced treatments to those predicted not to respond. Further, these findings propose that a novel therapeutic approach targeting *ETS1* could potentially block downstream effects of multiple cytokines concurrently, and thus be beneficial for patients with high enrichment of multiple cytokine programmes. However, significant challenges and unknowns would need to be addressed to use *ETS1* as a therapeutic target, including possible off target effects and subcellular location difficulties.

3.6 Future research directions

Through the generation of causal networks, I made a number of predictions regarding key TFs, post-translational modifications and cytokine-specific functional associations to IBD. For translation of these predictions to actionable results, further research and investigation is required. For example, targeted knock out or inhibition studies could be used to validate the importance of the key TFs in cytokine signalling. If found to result in few off-target effects, these TFs, especially *ETS1*, could serve as therapeutic targets for inhibiting cytokine-driven immune response in IBD – although this could be challenging due to its intracellular location. Furthermore, validation of the specific cytokine-function associations predicted in UC and CD

could enhance mechanistic understanding of these diseases, potentially leading to disease-specific approaches to cytokine inhibiting treatments.

For future work, the causal network reconstruction methodology should be improved. The addition of further molecular interaction resources and/or changes in DEG and interaction filtering criteria would likely result in more realistic causal networks. For example, the application of more advanced modelling tools to predict causal networks, such as CARNIVAL, or diffusion methods such as TieDIE, could provide more detailed networks with a reduced bias towards hub proteins (Liu et al., 2019; Paull et al., 2013). Moreover, different approaches to TF filtering, such as using the VIPER algorithm, would reduce the bias due to including only differentially expressed TFs (Alvarez et al., 2016). However, one further challenge is to determine which methods lead to the most realistic models, which would require significant experimental validation. Nevertheless, one would expect major findings, such as the importance of ETS1, to be replicated using similar approaches.

Finally, the work described could be extended to study cytokine effects in a cell type-specific manner or in different tissue types. Such experiments could utilise different organoid models, single cell sequencing and/or cell sorting technologies to obtain -omics data from different tissue and cell types. Such an approach may link known alterations in Paneth cells and goblet cells in IBD with cytokine aberrations, and lead to more targeted therapeutic approaches. For example, in Chapter 2 I predicted that ETS1 is also an important TF in Paneth and goblet cell function. As Paneth cells are known to be disrupted in CD, it would be interesting to investigate possible links between ETS1, Paneth cells, cytokine responses and CD (Liu et al., 2016; Treveil et al., 2020). However, phosphoproteomic studies should also be carried out to confirm activation levels of ETS1.

In conclusion, we investigated the transcriptomic landscape of the immune-epithelial interactome, finding previously unrecognised levels of shared and distinct transcriptional regulation of epithelial function by different cytokines. The complex and multifaceted nature of downstream signalling and regulation from cytokines ques illustrates why many questions remain unanswered regarding the action of cytokines in IBD. Nevertheless, we feel that our primary findings are of significant translational importance and pave the way for personalized medicine approaches in IBD.

Chapter 4: The effect of bifidobacteria on the small intestinal epithelium

4.1 Introduction

The *Bifidobacterium* genus are health-promoting commensal bacteria found in the human gastrointestinal tract. As described in the General Introduction (section 1.5), bifidobacteria are primary colonizers and dominant members of the early life gut microbiota, and consequently play a critical role in immune maturation and programming (Figure 1.6) (Arboleya et al., 2016; Ruiz et al., 2017; Stewart et al., 2018). IECs are a primary site of interaction between bifidobacteria and their host, thus, play a significant role in mediating the host response and beneficial effects of bifidobacteria. Therefore, a greater understanding of the effect of bifidobacteria on IECs will build knowledge of the mechanisms through which bifidobacteria exerts their beneficial effects. In turn, this knowledge can aid development of prevention and treatment options for gut and systemic inflammatory diseases. To date, a number of studies have investigated the effect of different bifidobacteria strains on IECs (see General Introduction section 1.5.2), however these works were mostly performed on adult mice in the context of acute or chronic gut inflammation and often focused on specific IEC functions such as mucus production and barrier function (Hsieh et al. 2015; Srutkova et al. 2015; Pinto-Sánchez et al. 2017; Schroeder et al. 2018; Yan et al. 2019; Din et al. 2020). However, very little work has focused on the role of *Bifidobacterium* on IECs in early life or in healthy conditions. Such studies could uncover the importance of bifidobacteria in gut development within babies and for maintenance of a healthy gut – in the absence of inflammation. Furthermore, most studies to date have focused only on specific functions of the epithelium, potentially overlooking other important effects of bifidobacteria.

Therefore, we aimed to explore how *Bifidobacterium* can modulate IEC homeostasis within the early life developmental window using healthy specific pathogen free (SPF, conventionalised) 2-week-old mice, as shown in in Figure 4.1. We compared this response to 10-12-week-old adult mice to uncover the influence of age in bifidobacterial effects (data not shown). Specifically, we investigated the effects of human infant associated strain *Bifidobacterium breve* UCC2003 which has been shown previously to have health beneficial effects (as described in General introduction section 1.5.4). Selective culturing and 16S rRNA microbiota profiling were used to

confirm the presence of *B. breve* within the mice gastrointestinal tracts and to quantify other microbiota alterations. Following isolation of IECs, we employed global transcriptomics analysis to compare bifidobacteria-treated mice to control mice – without limiting scope to specific IEC functions. We found that *B. breve* administration extensively alters the murine neonatal IEC transcriptome (~4,000 significantly upregulated genes), but has no significant effect on the transcriptomics of adult mice IECs (data not shown). Combining functional overrepresentation analysis with protein-protein interaction network reconstruction and clustering approaches, we identified a number of IEC functions affected in the neonatal mice, particularly epithelial barrier function, cell differentiation and proliferation. Furthermore, using cell type marker genes we identified an overrepresentation of stem cell marker genes among the differentially expressed genes (DEGs), indicating that bifidobacteria can cause an increase in the regenerative potential of the epithelial layer in neonatal mice. Regulatory network reconstruction was used to predict key regulators of the differentially expressed stem cell marker genes through which *B. breve* UCC2003 may be acting. Together these findings evidence the significant role of *Bifidobacterium* in early life modulation and development of IEC function.

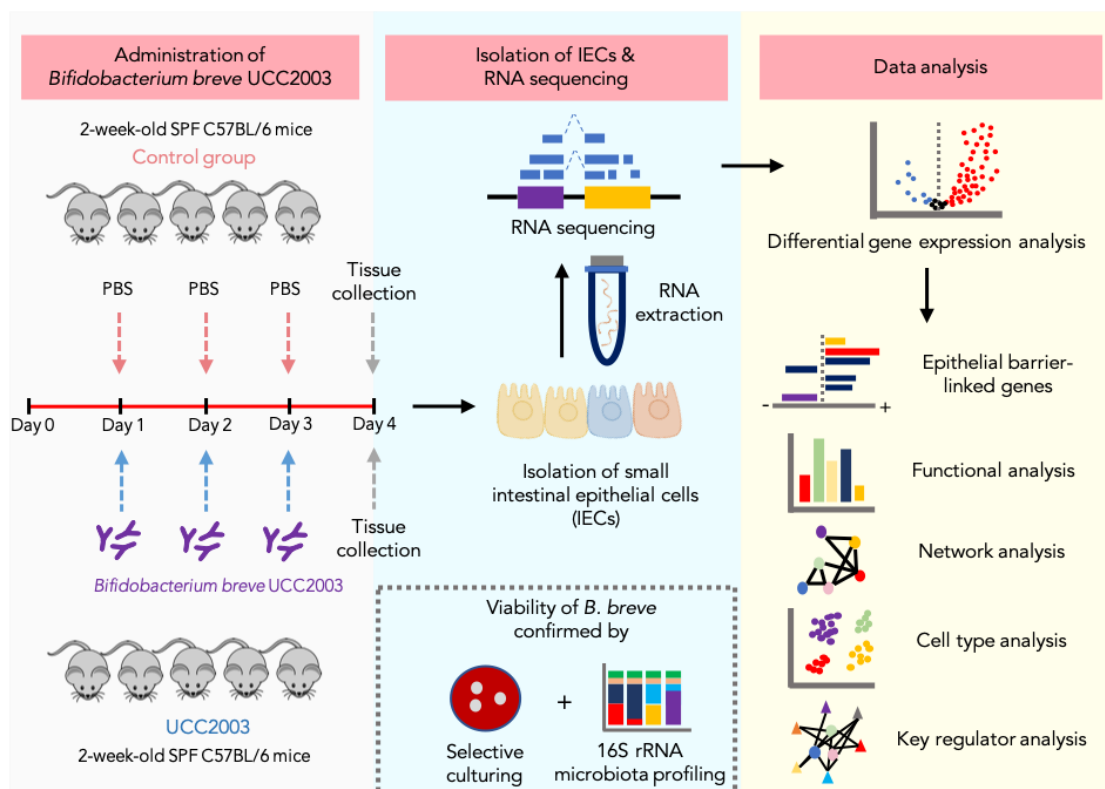


Figure 4.1. Schematic overview of neonatal bulk epithelium study design and analysis workflow. Figure adapted from Kiu et al. (2020) under the Creative Commons BY licence.

Carried out in collaboration with Lindsay Hall's research group (Quadram Institute Bioscience, QIB), my primary role in this project was the reconstruction and analysis of networks and cell type marker genes, while also supporting and training Raymond Kiu (QIB) in bioinformatics techniques for data processing and differential expression analysis. This study is subsequently termed the "neonatal bulk epithelium" study to differentiate it from the study described in Chapter 5. This chapter is based on (verbatim) the peer-reviewed article published in *iScience* in which I am second author (Kiu et al., 2020). The published article is reproduced in Appendix 6.

4.2 Aims

The aims for this project were as follows:

- Evaluate the global transcriptional response to *B. breve* UCC2003 in small intestinal IECs from young mice
- Predict which IEC types, if any, are particularly affected by *B. breve* UCC2003

4.3 Methods

All experimental work and the sequencing data analysis were carried out by Lukas Harnisch during his PhD project and Raymond Kiu (QIB). I have advised Raymond Kiu in RNA sequencing data processing and trained and supported him to carry out the signalling network analysis in section 4.3.6. I performed all the analysis and interpretation described thereafter. Further details of the experimental methods are described in Kiu et al. (2020).

4.3.1 Mouse work

All animal experiments and related protocols were performed in accordance with the Animals (Scientific Procedures) Act 1986 (ASPA). C57BL6/J female mice were housed within UEA Disease Modelling Unit. Two weeks old SPF mice (n=10) and 10-12 weeks old SPF mice (n=10) were fed autoclaved chow diet *ad libitum*.

4.3.2 Bacterial culturing, inoculum preparation, mouse challenge with *B. breve* UCC2003 and CFU enumeration

B. breve UCC2003 (also known as NCIMB 8807) was streaked from frozen glycerol stocks onto autoclaved Reinforced Clostridial Agar (RCA) plates (Oxoid, UK) and incubated in an anaerobic chamber (miniMACS, Don Whitley Scientific) at 37°C for 48 h prior to picking single colonies for inoculation in prewarmed liquid sterilised Reinforced Clostridial Medium (RCM) (Oxoid, UK). For preparation of mice gavage inocula, RCM strain pre-cultures were sub-cultured into De Man, Rogosa and Sharpe (MRS) medium (Oxoid, UK), incubated, centrifuged, washed and diluted in Phosphate Buffered Saline (PBS) (De Man et al., 1960). Bacterial concentration of inoculum was enumerated by plating serial dilutions in sterile PBS on RCA plates and enumerating colonies following two-day incubation to calculate CFU/ml. Mice received 50µl oral gavages with bacterial inoculations of 10⁸ CFU/ml or control samples of sterile PBS every 24 h for 3 consecutive days. Colonisation was confirmed by serial dilution and plating of fresh faeces on RCA supplemented with 50 mg/L mupirocin, and counting of colonies following 2-day incubation. Gut microbiota profiling was carried out by 16S rRNA amplicon sequencing of caecal samples on day 4 using the FastDNA Spin Kit for Soil (MP Biomedicals) the Illumina MiSeq platform and the QIIME analysis (Caporaso et al., 2010) software as described in Kiu et al. (2020).

4.3.3 Tissue harvesting and processing

On day 4 post oral gavage, the mice were humanely culled and 10cm sections of small intestine were immediately harvested and dissected into 0.5cm² pieces and IECs were isolated using adapted Weisser method (Hughes et al., 2017) as described in Kiu et al. (2020).

4.3.4 RNA extraction, preparation and sequencing

RNA was extracted from IEC isolations using QIAshredder spin columns (QIAGEN) followed by centrifugation. Follow-through was added to RLT lysis buffer and 70% ethanol and mixed by pipetting. Samples were then loaded to a RNeasy spin column until all of sample had gone through the filter. Buffer RW1 was added to column and centrifuged to remove carbohydrates, proteins and fatty acids. Flow through was discarded and filter placed in a new collection tube with wash buffer RPE and spun again, followed by discarding of flow through. Additional wash buffer RPE was pipetted into column and centrifuged again. Columns were transferred to a RNA low-bind Eppendorf tube with RNase free water and incubated for 1 min at room temperature. Sample was centrifuged and flow through containing RNA stored at -80°C prior to processing. Isolated RNA was processed by poly-A selection and/or Ribo-depletion. Purified RNA was quantified, and quality controlled using RNA 6000 Nano kit on a 2100 Bioanalyser (Agilent). Only samples with RNA Integrity Number (RIN) values above 8 were sequenced. RNA sequencing was performed at the Wellcome Trust Sanger Institute (Hinxton, UK) on paired-end 75 bp inserts on an Illumina HiSeq 2000 platform. All samples were sequenced using non-stranded, paired-end protocol.

4.3.5 Sequence pre-processing and differential expression analysis

Quality of sequencing reads was assessed using FastQC (v0.11.8). and fastp (v0.20.0) with options -q 10 (phred quality < 10 was discarded). rRNA sequences were removed using SortMeRNA (v2.1) based on SILVA rRNA database (Chen et al., 2018; Kopylova et al., 2012; Quast et al., 2013). Transcript mapping and quantification were performed using Kallisto (v0.44.0) (Bray et al., 2016). Briefly, *Mus musculus* (C57BL/6 mouse) cDNA sequences (GRCm38.release-98_k31) were retrieved from Ensembl database and built into an index database with Kallisto utility index at default parameter that was used for following transcript mapping and abundance quantification via Kallisto utility quant at 100 bootstrap replicates (Zerbino et al., 2018). Differential gene expression analysis was performed using R library Sleuth (v0.30.0) (Pimentel et al., 2017). Gene transcripts were mapped to individual genes using Ensembl BioMart database

with Sleuth function `sleuth_prep` with option `gene mode = TRUE` (Kinsella et al., 2011). Genes with an absolute $\log_2(\text{fold change}) > 1$ and q value < 0.05 were considered to be differentially expressed (or, significantly regulated).

4.3.6 Signalling network reconstruction and analysis

A signalling network of all upregulated DEGs and their first neighbours was built using all available biological signalling databases in the Cytoscape (v3.7.2) OmniPath app (v1, *Mus musculus*) (Türei et al., 2016). Modules of highly connected genes within the signalling network were identified using the MCODE plug-in within Cytoscape (Bader and Hogue, 2003). The nodes of each individual module were tested for functional enrichment based on both Reactome and PANTHER annotations using PANTHER Classification System as described in section 4.3.8 (Croft et al., 2011; Mi et al., 2019; Thomas et al., 2003).

4.3.7 Cell type signature analysis

Cell type signature gene sets for murine intestinal epithelial cells were obtained from Haber *et al.* (Haber et al., 2017). From this dataset, both droplet and plate-based results were used. Gene symbols were converted to Ensembl IDs using db2db (Mudunuri et al., 2009). Hypergeometric significance calculations were applied to test the presence of cell type-specific signatures in the list of differentially expressed genes using all expressed genes as the statistical background (normalised counts ≥ 1 in ≥ 1 sample). Bonferroni multiple correction was applied and any adjusted $p < 0.05$ was deemed significant. Genes with normalised counts ≥ 1 in ≥ 1 sample per condition (*B. breve* UCC2003 treated or control) were used to identify cell type signature genes expressed per condition. Where calculated, adjusted p values ≤ 0.05 were considered significant.

Differentially expressed stem cell signature genes were contextualised using regulatory networks. Mouse directed transcription factor - target gene (TF-TG) interactions were obtained from DoRothEA (Garcia-Alonso et al., 2019) using confidence levels A-D (all except predicted interactions) via the OmniPath Cytoscape app (Shannon et al., 2003; Türei et al., 2016). After filtering for nodes which are expressed in the dataset (normalised counts ≥ 1 in ≥ 1 sample). TFs were further filtered for relevance in the network using a Python script written by Matthew Madgwick (EI, QIB, Korcsmaros group; unpublished data) based on the Cytoscape app CHAT (Muetze et al., 2016). This tool inputted the DoRothEA TF-TG network filtered for expressed nodes, alongside the list of all DEGs. Then, a hypergeometric significance test was carried out

on any node with degree > 5 to determine if the proportion of connected nodes which are differentially expressed is higher than in the whole network. Any TF with adjusted p value ≤ 0.05 following Benjamini-Hochberg correction were deemed significant and used to filter the stem cell signature gene subnetwork. Network visualisation was carried out in Cytoscape (Shannon et al., 2003; Su et al., 2014).

4.3.8 Functional analysis

Functional overrepresentation analysis was carried out using the Panther web tool (Thomas et al., 2003, 2006). Gene lists were tested against Panther Gene Ontology-Slim Biological pathways and against Reactome pathways (Fabregat et al., 2018a) with default settings. All pathways with q value ≤ 0.05 are considered significantly overrepresented. When testing the functional associations of all stem cell DEGs and their regulators (total 64 genes), all nodes in the unfiltered DoRothEA network are used as a background for the statistical test.

4.4 Results

4.4.1 *B. breve* impacts the neonatal intestinal epithelial transcriptome

To examine the effects of bifidobacteria on host IECs, neonatal (two weeks old) and young adult (10-12 weeks old) mice were gavaged with *B. breve* UCC2003 for three consecutive days ($n=5$ per group) prior to whole IEC RNA sequencing. Culture and 16S rRNA microbiota profiling approaches were used to confirm colonisation and determine the impact of *B. breve* UCC2003 on the wider microbiota. Raymond Kiu and Shabhonam Caim demonstrated that while increasing the proportion of *B. breve* UCC2003 itself, colonisation of neonatal mice with this bacteria had minimal impact on overall microbiota profiles - although very low relative abundance (<2%) microbiota members *Streptococcus*, *Ruminococcus*, *Prevotella* and *Coprococcus* were significantly reduced in the *B. breve* UCC2003 group (Kiu et al., 2020).

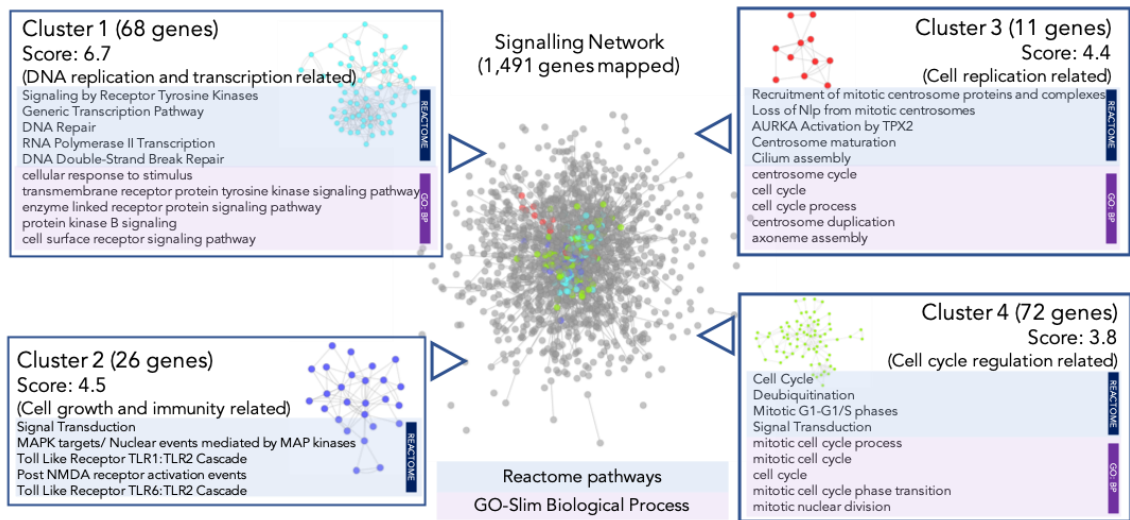
Whole transcriptome analysis of small intestinal IECs identified a significant impact of *B. breve* UCC2003 on neonatal intestinal epithelium, while no change was observed in young adult mice, based on differentially expressed genes (DEGs). This suggests *B. breve* UCC2003 modulation of IECs is strongest within the early life window under homeostatic conditions. In the neonatal group, a total of 3,996 DEGs were significantly upregulated, while 465 genes were significantly downregulated in *B. breve* UCC2003 supplemented animals when compared to controls (absolute $\log_2(\text{fold change}) > 1$ and adjusted $p < 0.05$).

As *B. breve* strains have been previously shown to modulate tight junction and other barrier-related proteins, Raymond Kiu investigated DEGs associated with intestinal epithelial barrier development and intestinal structural organisation. He observed a number of upregulated genes associated with tight junctions, adherens junctions and gap junctions (Figure S4.1). Furthermore, a number of genes involved in mucus layer generation were also upregulated (Figure S4.1). Surprisingly, anti-microbial peptide genes, including defensins were not differentially expressed. Functional overrepresentation analysis identified a variety of enriched functions including DNA repair, cell cycle, transcription and chromatin organisation (Table S4.1). Together these results suggest *B. breve* UCC2003 induces extensive transcriptional changes in neonatal IECs, including genes relating to enhanced epithelial barrier development.

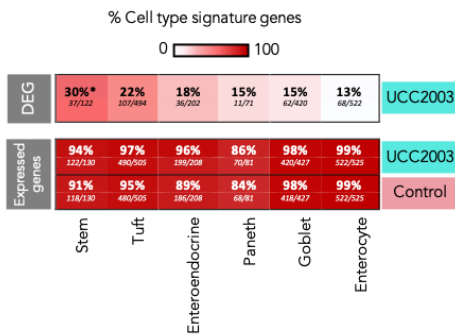
4.4.2 *B. breve* modulates neonatal cell maturation processes

To delve further into the data, Raymond Kiu and myself constructed a signalling network based on upregulated DEGs (n=3,996) with the aim of identifying signalling pathways potentially modulated by bifidobacteria in IECs (Figure 4.2A). To do this, we used a prior knowledge protein-protein interaction (PPI) network to determine possible PPI connections between the (translated) DEGs and their direct protein interactors (protein first-neighbours). Identifying connections between (translated) genes permits grouping of genes into clusters which work closely together and are therefore likely to have similar functions. Similarly, the first neighbour approach can aid the annotation of functional associations through identifying additional biologically relevant proteins (Módos et al., 2017).

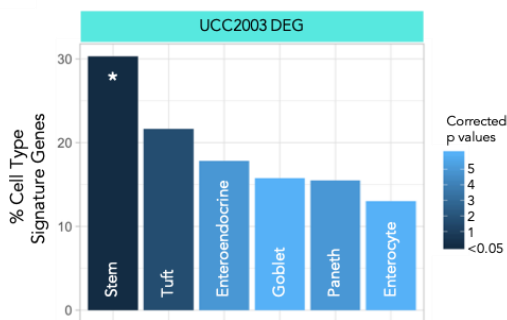
A



B



C



D

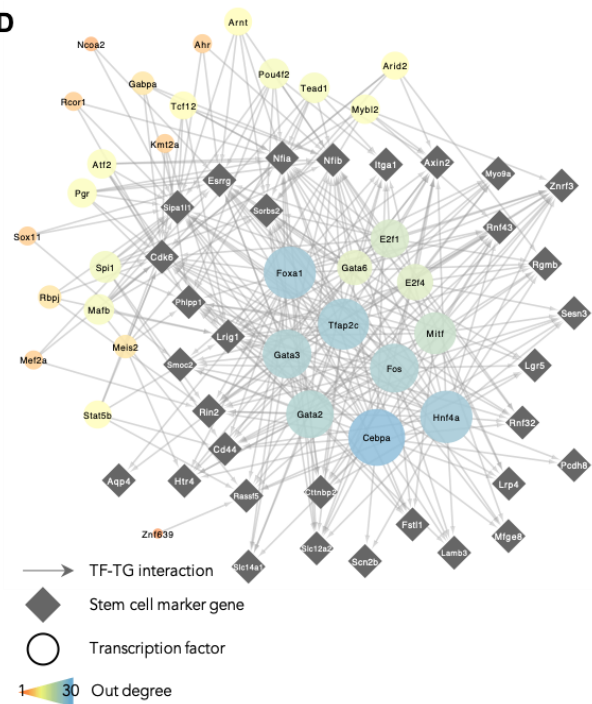


Figure 4.2. Signalling network analysis, IEC subtyping and key regulator analysis. **A.** Cluster analysis of signalling network for significantly upregulated genes (n=3,996). Representative enriched pathways (Reactome) and GO terms (Biological Process) identified in each individual cluster were listed alongside. **B.** Heat plot showing percentage of cell type signature genes in DEG and expressed genes (both control and UCC2003 groups). All expressed genes are well represented in IEC cell type signature genes. **C.** Cell type analysis on IEC DEGs using known cell-specific signature genes. Stem cells were statistically overrepresented in DEGs. * p < 0.05. **D.** Key regulators of stem cell DEGs. Figure reproduced from Kiu et al. (2020) under the Creative Commons BY licence.

Overall, 1,491 DEGs were successfully mapped (37.3%) to a signalling network that comprised 8,180 genes. Four individual clusters of highly connected genes were detected, with functional assignment and pathway analysis implemented on these clusters (Figure 4.2A, Table S4.2). All gene clusters were associated with cell differentiation and maturation, with cluster 1 (68 genes) linked specifically with DNA replication and transcription, cluster 2 (26 genes) with cell growth and immunity (including toll like receptor 2, TLR2 cascades), cluster 3 (11 genes) with cell replication, and cluster 4 (72 genes) related to cell cycle and cell division. Whilst only representing a third of the upregulated DEGs, this analysis highlights the importance of cell maturation processes as a response to *B. breve*.

4.4.3 Neonatal affected genes are enriched with epithelial stem cell markers

IECs include several absorptive and secretory cell types, namely enterocytes, Paneth cells, goblet cells, enteroendocrine cells, tuft cells and stem cells. Each of these cells perform different functions in the gut (as described in General Introduction section 1.2.1) and could be differentially affected by bacterial signals. Therefore, I investigated whether *B. breve* UCC2003 had a cell type-specific effect on the intestinal epithelium, using known cell type-specific gene markers (Haber et al., 2017).

First, I validated the presence of all IEC types in our study data by identifying cell type markers within the genes expressed in the transcriptomics data of the control and UCC2003 groups (Figure 4.2B, C). I found that all the cell type markers were well represented in both datasets (between 84% and 99% markers) and that there was little difference between the control and UCC2003 groups. Next, I investigated whether any genes differentially expressed after *B. breve* UCC2003 supplementation were cell type markers. This analysis revealed that stem cell marker genes were significantly enriched (30%; $P < 0.05$) among the six IEC types after *B. breve* UCC2003 supplementation (Table S4.3). Signatures of other cell types were also present but not significantly over-represented: Tuft cells (22%), enteroendocrine cells (18%), goblet cells (15%), Paneth cells (15%) and enterocytes (13%). These data indicated that intestinal epithelial stem cells, cells primarily involved in cell differentiation, were the primary cell type whose numbers and transcriptomic programme were regulated by *B. breve* UCC2003.

Further investigation of this stem cell signature revealed that of the 37 differentially expressed marker genes, 35 are upregulated in the presence of *B. breve* UCC2003 (96%) – whereas only 54% of the enterocyte markers were upregulated (37/68). This indicates an increase in the quantity of stem cells or semi-differentiated cells in the epithelium, consistent with the overrepresentation of cell cycle and DNA replication associated genes observed in the whole differential expression dataset. Functional analysis of the 37 stem cell signature genes revealed only one overrepresented process - Regulation of Frizzled by ubiquitination ($p \text{ val} < 0.05$), which is a subprocess of WNT signalling. WNT signalling is important in maintaining the undifferentiated state of stem cells (Nusse, 2008).

Moreover, I employed a network approach to predict key transcription factor (TF) regulators of these 37 genes, through which *B. breve* UCC2003 could be acting. Using the TF-target gene database, DoRothEA, we identified expressed TFs known to regulate the stem cell signature genes (Garcia-Alonso et al., 2019). Five genes had no known and expressed regulator thus were excluded. Hypergeometric significance testing was used to identify which of these TFs were most influential based on out-degree (the number of genes regulated by the TF) and the transcriptomics data (see Methods for detail, section 4.3.7). This analysis identified 32 TF regulators, some which target up to 30 of the stem cell signature genes (Figure 4.2D). Of the regulators, 12 are differentially expressed in the IEC dataset (all upregulated): *Fos*, *Gabpa*, *Rcor1*, *Arid2*, *Tead1*, *Mybl2*, *Mef2a*, *Ahr*, *Pgr*, *Kmt2a*, *Ncoa2* and *Tcf12*. Functional analysis of the 32 regulators and their 32 targeted stem cell signature genes together revealed overrepresented pathways involved with WNT signalling, histone methylation for self-renewal and proliferation of hematopoietic stem cells, nuclear receptor (incl. oestrogen) signalling, signal transduction and gene expression (Table S4.4). These data provide evidence that *B. breve* UCC2003 directly affects key transcriptomic programmes which regulate specific signalling processes, particularly within stem cells.

4.5 Discussion

Bifidobacteria are predominant in the guts of healthy human infants during a critical stage of immune development and programming, where microbe-host interactions impact health in the short and longer term (Arboleya et al., 2016; Isolauri, 2012). As such, studies have shown that bifidobacteria play a key role in modulating and priming specific immune populations and protecting against immune linked diseases (O'Neill et al., 2017). However as far as I am aware, no study has yet investigated the age-specific effects of bifidobacteria. We showed that *B. breve* had a large impact on the IEC transcriptome of neonatal mice, but no significant impact on the transcriptome of young adult mice (data not shown). The striking differences in DEGs between these two life points indicate that, in a healthy gut, *B. breve* modulation of IECs is limited to the early life window.

Given the limited change in the gut microbiota of bifidobacteria-exposed mice compared to control SPF mice, we assume that the observed IEC expression changes were driven by bifidobacterial surface molecules, metabolites or secreted products. The interactions of these molecules and IECs (directly or indirectly) resulted in a large-scale upregulation of IEC genes and impacted many processes previously associated with bifidobacteria (Engevik et al., 2019; Ewaschuk et al., 2008; Srutkova et al., 2015). For example, we revealed that expressions of key genes associated with formation of epithelial barrier components were up-regulated, including major cell junction protein-encoding genes (75%; 42/56 genes). One such group was integrins, which facilitate cell-cell and cell-extracellular matrix adhesion that is pivotal for cell migration and cell differentiation (Harburger and Calderwood, 2009). Integrins also play an important role in downstream intracellular signalling that controls cell differentiation, proliferation, and cell survival, including the Raf-MEK-ERK signalling pathway (we also observed enrichment of genes involved in this pathway) (Chernyavsky et al., 2005; Li et al., 2016). Tight junction proteins were also found to be upregulated, indicating that bifidobacteria is capable of increasing barrier function in a healthy condition, in addition to previous evidence which showed that bifidobacteria can prevent barrier function decline due to damaging agents or disease (Din et al., 2020; Srutkova et al., 2015; Yan et al., 2019). Dysfunctional epithelial barrier function may lead to a “leaky” gut, which is characteristic of numerous intestinal disorders including inflammatory bowel diseases (Krug et al., 2014). Notably, previous work has suggested early life microbiota disruptions (*via* antibiotic usage) and reductions in *Bifidobacterium* are correlated

with increased risk and/or symptoms of ulcerative colitis and Crohn's disease (Duranti et al., 2016; Favier et al., 1997; Giaffer et al., 1991; Kronman et al., 2012), although the opposite has also been reported (Wang et al., 2014). Furthermore, several clinical studies have indicated that supplementation with certain *Bifidobacterium* strains positively modulate gastrointestinal symptoms of patients, which is corrected with reductions of inflammatory markers in colonic IEC-containing biopsies; however, *B. breve* UCC2003 has not been used clinically in this patient setting (Furrie et al., 2005; Steed et al., 2010). Further clinical studies would be required to probe these findings in detail to determine their importance during healthy infant development.

Furthermore, in line with previous evidence, we found that *B. breve* UCC2003 upregulated mucin producing genes, which play a crucial role in intestinal protection *via* formation of a physical barrier between the gut lumen and IECs (Caballero-Franco et al., 2007; Engevik et al., 2019; Mangin et al., 2018). Deficiencies in mucins have been linked with experimental colitis and increased inflammation in patients with inflammatory bowel disease, further evidencing a link between bifidobacteria and inflammatory diseases (Shirazi et al., 2000; Van der Sluis et al., 2006). On the other hand, despite some literature evidence that bifidobacteria can impact the expression or secretion of anti-microbial peptides, we did not observe any significant changes in these genes (Lee et al., 2018; Natividad et al., 2013; Underwood et al., 2012).

Although we observed a substantial IEC transcriptional response induced by bifidobacteria, we cannot rule out the possibility that these changes would occur upon introduction of any new microbiota member. However, the similarity between observed transcriptional responses and previous literature evidence, alongside the lack of a generalised immune response in our bifidobacteria-exposed mice, indicates that our results represent bifidobacteria-specific effects on IECs. Further work would be required to extrapolate the observed results to humans and to test whether bifidobacteria-induced changes are transient or have long-term effects on IECs and immune maturation.

Using *a priori* information on protein-protein interactions, we were able to identify proteins which can theoretically interact with more than 1/3rd of the (translated) upregulated genes. This approach permitted generation of a PPI network, in which connections between DEGs and their first neighbours indicate similarity and cooperation in biological functions. In turn, this permitted clustering of genes/proteins into four distinct highly inter-connected groups. Functional analysis revealed that all four clusters were associated with cell maturation and cell

differentiation, suggesting that neonatal *B. breve* exposure positively modulates IEC cell differentiation, growth, and maturation. Similar findings have been previously reported in colonic epithelial cells of bifidobacteria-monoassociated mice (O'Connell Motherway et al., 2019). Further, Lee *et al.* (2018) observed that *Bifidobacterium*-derived lactate supports IEC differentiation through Paneth cell and stromal cell Wnt3 secretions – highlighting the possibility of indirect effects of bifidobacteria based on communication between IECs.

Among the cluster associated functions, we also identified the TLR2 pathway. This may link to previous work indicating that the *B. breve* UCC2003 EPS signals *via* TLR2 to induce MyD88 signalling cascades to protect IECs during intestinal inflammation (Hughes et al., 2017). *B. breve* M-16V was also shown to interact with TLR2 to up-regulate ubiquitin-editing enzyme A20 expression that correlated with increased tolerance to a TLR4 cascade in porcine IECs, further supporting the involvement of *B. breve* in programming key host immunoregulation receptors (Tomosada et al., 2013). In the future, the use of additional *a priori* PPI resources could help to annotate a larger proportion of the DEGs, for identification of further clusters.

Different cell types of the small intestinal epithelia have different functions, such as mucin production by goblet cells and AMP production by Paneth cells. Our findings in this experiment, as well as other published works, indicate that bifidobacteria can affect these cell type-specific functions. While we did not obtain cell type-specific data from this experiment, I predicted which cell types were most affected by *B. breve* UCC2003 in neonatal IECs by observing which DEGs also occur in a published collection of IEC marker genes (Haber et al., 2017). This analysis revealed stem cells as the IEC type most affected by *B. breve*, with absorptive enterocytes least affected despite being most accessible to bacteria in the gut. It could be hypothesised that *B. breve* or their secreted metabolites may reach the crypts of the small intestinal epithelium. Previous evidence based on *in situ* hybridization histology *in vivo* suggests that *B. breve* can reach small intestinal crypts, while it is also possible that inter-cellular communication can transmit signals to the bottom of the intestinal crypts, such as the communication observed between stromal cells, Paneth cells and stem cells following lactate administration (Hughes et al., 2017; Lee et al., 2018). Specifically, I found that 37 (of 122) stem cell marker genes were differentially expressed upon exposure to bifidobacteria. This represented a significant overlap (adjusted $p \leq 0.05$, hypergeometric significance test). All but two of the 37 were upregulated in the presence of *B. breve* UCC2003, indicating an activating effect resulting in increased

pluripotency of stem cells, increased quantity of stem cells, and/or an increased quantity of semi-differentiated cells. This finding also corroborates with the overrepresentation of cell cycle associated genes amongst the DEGs. Further, using *a priori* regulatory interactions, I predicted 32 TFs associated with these differentially expressed stem cell marker genes - providing possible targets for future investigation of the mechanisms underlying these responses. While they have not been previously associated with bifidobacteria, many of the TFs are involved in cell proliferation, differentiation, survival and cell cycle. For example, FOS Proto-Oncogene, AP-1 transcription factor subunit (FOS), MYB proto-oncogene like 2 (MYBL2) and Aryl hydrocarbon receptor (AHR) (Bar Hoover et al., 2010; Brown et al., 1998; Musa et al., 2017). Further a number of the TFs are associated with histone and chromatin remodelling, such as AT-rich interaction domain 2 (ARID2) and Lysine methyltransferase 2A (KMT2A), indicating a role for *B. breve* in epigenetic modification (Duan et al., 2016; Huang et al., 2017). Functional analysis of the differentially expressed stem cell marker genes and their regulators suggests that *B. breve* increases pluripotency of stem cells and/or semi-differentiated epithelial cells through WNT signalling and nuclear hormone signalling (Jeong and Mangelsdorf, 2009). Substantiating this finding, WNT signals of Paneth cells and stromal cells have been previously implicated in mediating bifidobacteria effects on stem cell proliferation (Lee et al., 2018).

4.6 Future research directions

In conclusion, we have shown that *B. breve* UCC2003 plays a central role in orchestrating global neonatal IEC gene responses in a distinct manner as shown in our murine model, modulating genes involved in epithelial barrier development, and driving universal transcriptomic alteration that facilitates cell replication, differentiation, and growth, particularly within the stem cell compartment. Further work is required to investigate the cell type-specific effects of bifidobacteria, in particular to validate their impact on stem cells. In Chapter 5, myself and colleagues employ fluorescence activated cell sorting (FACS) to isolate stem cells and Paneth cells from IECs. Moreover, single-cell sequencing of IECs could be used to further investigate cell type-specific effects. Additionally, further work could investigate the impact of bifidobacteria on the colonic epithelium, the age at which bifidobacteria ceases to significantly impact IECs, or determine host and bacterial metabolome and proteome after *B. breve* exposure to investigate specific underlying molecular mechanisms of interaction (Guo et al., 2015).

Chapter 5: The effect of Bifidobacteria on small intestinal stem cells and Paneth cells

5.1 Introduction

In Chapter 4, we found that *B. breve* UCC2003 extensively regulates the transcriptome of the healthy neonatal small intestinal epithelium. Furthermore, our analysis indicated a larger effect on stem cells than other IEC types, evidencing a cell type-specific action of *B. breve*. To further investigate the effect of bifidobacteria at a cell type-specific level, we carried out a second study, presented in Chapter 5 of this thesis, focusing specifically on stem cells and Paneth cells of the small intestine. Here, stem cells were chosen due to the findings outlined within Chapter 4, while Paneth cells were selected to confirm that *B. breve* UCC2003 does not affect antimicrobial peptide (AMP) release as seen in Chapter 4, despite contention within the literature (see General Introduction section 1.5.2.3). Furthermore, we extended this investigation to germ free (GF) mice in addition to specific pathogen free (SPF) mice, to evaluate the effect of bifidobacteria as the initial colonisers of the gut. The mice investigated were four weeks old (two weeks older than in the neonatal bulk epithelium study), and thus will have recently weaned from their mother's breast milk. The reason for this was to enable the study of GF mice and to investigate the effect of bifidobacteria on young yet mature epithelial cells. The study design and analysis workflow for this study, termed the "juvenile cell type-specific" study, is outlined in Figure 5.1.

This study was carried out by myself and others members of Tamas Korcsmaros's research group (Earlham Institute, EI, QIB), with support from the UEA Disease Modelling Unit, the QIB germfree mice facility and the Genomics Pipelines Group at the EI. A single gavage of *Bifidobacterium breve* UCC2003 was administered to juvenile GF and SPF mice (Figure 5.1). Subsequently, small intestinal Paneth cells, stem cells and two intermediary cell populations were isolated by Fluorescence-activated cell sorting (FACS) before low-input RNA sequencing. Contrasting the first study, we observed only very modest changes in the transcriptional profile of the intestinal epithelial cells upon exposure to bifidobacteria, despite evidence of successful colonisation of GF mice based on selective culturing. However, correlation analyses, including co-expression network analysis and gene set enrichment analysis, identified a number of cellular functions marginally affected by bifidobacteria, including those identified in the previous study: cell cycle, autophagy, apoptosis and cell-cell junctions. Further analysis was carried out to evaluate the

FACS protocol used in this study. My roles were: experimental design and planning, bacterial work and most of the data processing/analysis.

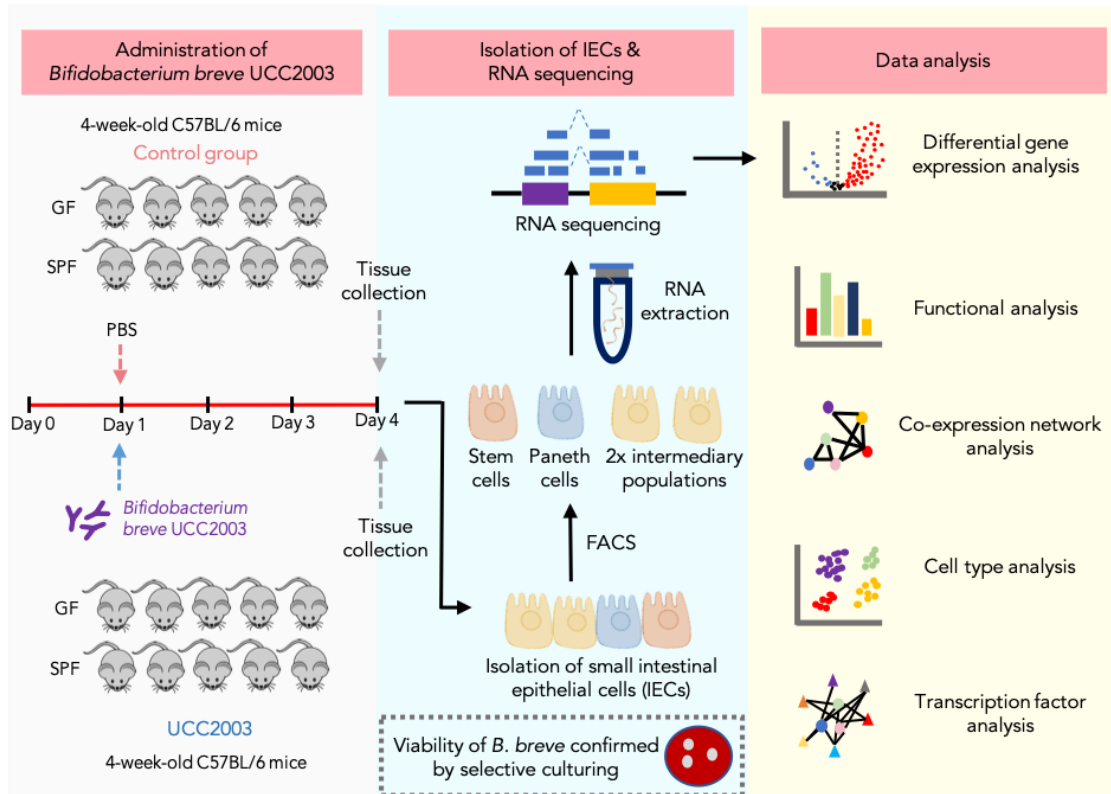


Figure 5.1. Schematic overview of juvenile cell type-specific study design and analysis workflow. GF - germ free; SPF - specific pathogen free/conventionalised; FACS - fluorescence-activated cell sorting. Figure adapted from Kiu et al. (2020) under the Creative Commons BY licence.

5.2 Aims

The aims for this project were as follows:

- Evaluate the global transcriptional response to *B. breve* UCC2003 in specific small intestinal IEC types (using fluorescence-activated cell sorting (FACS) and low cell input RNA sequencing)
- Investigate differences in the effect of *B. breve* UCC2003 between mono-associated and conventionalised mice

5.3 Methods

Due to the complexity of the methods and the number of people involved, the following experimental work was managed by Isabelle Hautefort (from our group). Mouse work including gavage was carried out by Arlaine Brion and Andrew Goldson (QIB). All bacterial work was carried out by myself. Tissue harvesting and processing were carried out by Andrew Goldson as well as Martina Poletti (from our group). FACS was carried out by Amanda Demeter and Elena Rodriguez from our group with support from Iain Macaulay (EI). RNA extraction, preparation and sequencing were carried out by Amanda Demeter, Ashleigh Lister (EI), Elena Rodriguez and Earlham Institute Genomic Pipelines. Processing of raw sequencing reads was carried out by myself with help from Matthew Madgwick from our group. All other computational analysis and interpretation was carried out by myself.

5.3.1 Mouse work

All animal experiments and related protocols were performed in accordance with the Animals (Scientific Procedures) Act 1986 (ASPA). C57BL6/J female mice were housed within UEA Disease Modelling Unit. 10 germ free (GF) and 10 conventionalised (SPF) mice of four weeks old (n=5 for each condition; GF+UCC2003, GF control, SPF+UCC2003, SPF control) were kept in sterile individually ventilated cages receiving sterile food and sterile water *ad libitum*.

5.3.2 Bacterial culturing, inoculum preparation, mouse challenge with *B. breve* UCC2003 and CFU enumeration

B. breve UCC2003 (also known as NCIMB 8807) was streaked from frozen glycerol stocks onto autoclaved Reinforced Clostridial Agar (RCA) plates and incubated in an anaerobic chamber at 37°C for 48h prior to picking single colonies for inoculation in prewarmed sterilised MRS medium. All media and agar were supplemented with 50mg/L cysteine.

Only two mice for each group were challenged with *B. breve* UCC2003 or phosphate buffered saline (PBS) control at a time as tissue samples and cell isolation and sorting were greatly time-consuming. The complete series of *B. breve* UCC2003/PBS challenge was therefore staggered over a period of five weeks. To standardise gavage inoculums over that experiment period, *B. breve* UCC2003 inoculum batch was freeze-dried in 500 µl aliquots with PBS + 10% skimmed milk

(lyoprotectant) (MODULYO D freeze drier) and stored at -70°C . On day of gavage, cultures were resuspended in sterile PBS at 10^9 CFU/ml. Mice received a single $100\mu\text{l}$ oral gavage of *B. breve* UCC2003 in PBS + 10% milk (approximately 10^8 cells/mouse) or PBS+ 10% milk (control). Inoculum viability/consistency was controlled by plating 10-fold serial dilutions made in sterile PBS on RCA plates, incubation at 37°C in anaerobiosis for 48h, and colony counting to calculate CFU/ml. Contamination of inoculum was checked by plating serial dilutions on Brain Heart Infusion (BHI) agar plates in aerobic incubation for 24-48h. Pre-gavage and 24h post-gavage colonisation levels of bifidobacteria were checked through serial dilutions and plating of fresh faeces and caecal content on RCA and De Man Rogosa and Sharpe (MRS) agar (De Man et al., 1960) supplemented with 50 mg/L cysteine, in addition to BHI plates for aerobic microorganisms. Colonies were counted following 24-48h incubation. At the end of the experiments, caecum content was collected for serial dilution plating on RCA and MRS agar supplemented with 50 mg/L cysteine. The remaining content was immediately snap-frozen for future metagenomics and 16S rRNA profiling.

5.3.3 Tissue harvesting and processing

Seventy-two hours post oral gavage, animals were humanely culled and 10cm-long segment of flushed ileal tissues were harvested, longitudinally opened and cut into 5-8mm long pieces. Following the removal of fat, mesenteric tissue, mucus and debris, fragments were washed up to 6 times in Dulbecco's phosphate buffered saline (DPBS) without Mg^{2+} and Ca^{2+} . For isolation of crypt epithelial cells, the fragments underwent incubation at room temperature in gentle dissociation reagent (StemCell Technologies, 07174) to remove large villi debris. For single cell isolation, the crypt fraction was spun down and the pellet was resuspended in a solution containing 1ml TryPLE Express Enzyme (Fisher Scientific, 12605036) and $120\mu\text{L}$ DNase I (Roche, 04536282001). The samples were placed in a water-bath at 37°C for 1-2 minutes. The pellet was resuspended in cold Dulbecco's Modified Eagle Medium and run through a $40\mu\text{m}$ cell strainer to remove large cell clumps. The cells were then counted in a haemocytometer, to calculate the appropriate number to be used for the antibody staining procedure.

5.3.4 Flow cytometry

The following protocol was optimised from Yilmaz et al. (Yilmaz et al., 2012). In that study, the authors sorted stem cells from *Lgr5-EGFP-IRES-creERT2* knock-in reporter mice, allowing isolation by flow cytometry of Leucine-rich repeat-containing G-protein coupled receptor 5

(Lgr5) – Enhanced green fluorescence protein (EGFP)^{high} cells which also have low Cluster of differentiation 24 (CD24 / CD24a) marker fluorescence. In our study we used a similar protocol but employed wildtype mice and measured surface Lgr5 marker fluorescence, targeting the presence of LGR5 protein instead of its gene expression. Furthermore, in addition to sorting stem cells and Paneth cells based on the Yilmaz protocol, we sorted intermediate cells with high fluorescence of both LGR5 and CD24a (TA^{high}) or low fluorescence of both (TA^{low}).

Ileal stem cells, Paneth cells and two populations of intermediate cells were labelled with a pre-optimised antibody panel and sorted in batches of 50 cells (4 replicates for each cell population and for each mouse) for RNA sequencing. For antibody labelling, samples were centrifuged and resuspended in FACS Buffer with the antibody cocktails. Antibody panel given in Table 5.1. Samples were incubated for 15 minutes, washed twice and resuspended in FACS Buffer with 7-Aminoactinomycin D (7AAD) viability stain (Biolegend, 420403).

Antibodies	Fluorescent colour	Cell population	Manufacturer
αCD45, αCD31, αTer-119	Phycoerythrin (PE)	Immune cells	Biolegend (103105, 102507, 116207)
αCD326 (Epithelial cell adhesion molecule, EpCAM)	VioBlue	Epithelial cells	Miltenyi (130-102-421)
αCD24	Fluorescein isothiocyanate (FITC)	Paneth cells	Miltenyi (130-102-731)
αLgr5	APC-Vio770	Stem cells	Miltenyi (130-111-392)

Table 5.1. Antibody panel for sorting Paneth, stem and transit amplifying cells. Using FACS Melody machine. Based on Yilmaz et al. (2012).

FACS was carried out on the BD Bioscience FACSMelodyTM Cell Sorter (Becton, Dickinson Company) following optimisation of gating using fluorescence minus one control. Representative gating strategy is shown in Figure S5.1. Live single cells with high granularity were selected using forward scatter, side scatter and 7AAD signals. In addition to the cells sorted based on the Yilmaz *et al.* panel, we also sorted intermediary populations with high LGR5 high CD24 (termed transit amplifying high cells) and low LGR5 low CD24 (termed transit amplifying low cells) (von Furstenberg et al., 2011; Gracz et al., 2010; King et al., 2012; Yilmaz et al., 2012). Later I assessed cell identities using the transcriptomics data as described in section 5.3.7. To conclude, four cell populations were sorted which were labelled as follows: Paneth cells (high CD24, low LGR5), Stem cells (low CD24, high LGR5), transit amplifying (TA) high cells (high CD24, high LGR5) and TA low cells (low CD24, low LGR5).

5.3.5 RNA extraction, preparation and sequencing

The RNA from specific sorted cell types was extracted and amplified using a SMART-seq 2 method for low input samples, based on a protocol adapted from Picelli et al. (2013). Bead cleaning was carried out using a Biomek NXP robot (Beckman Coulter) with Angecourt AMPure Beads (Beckman Coulter). Samples were quality checked using an Agilent Bioanalyser 2100, a LabChip GX Touch™ (PerkinElmer). Next, libraries were prepared using a NEXTERA XT Library preparation protocol. All samples were pooled and quality checks were performed using the Qubit and the Bioanalyzer following 0.8X Ampure bead clean-up. Illumina Sequencing was carried out by the Genomics Pipelines Group at the Earlham Institute on 1 lane of an Illumina NovaSeq 6000 S2 flow cell with 100PE reads.

5.3.6 Transcriptomics data processing

Sequencing reads from the flow sorted cells were trimmed using Trimmomatic (v0.38) (Bolger et al., 2014) using Standard NEXTERA adapters plus SmartSeq2 adapters on paired end mode with all other settings as default.

Read and sequencing quality was checked using FastQC (v0.11.7), MultiQC (v1.5) and custom R scripts (Andrews, 2010; Ewels et al., 2016). Trimmomatic (v0.38) was used to trim adapters (Nextera and Smartseq2 adapters) and remove low-quality reads. Sequence length, sequence quality and number of reads was used to check for outlier samples (Figure 5.2A). Following an initial round of data analysis, two samples were removed due to aberrant normalised counts distribution (Figure 5.3).

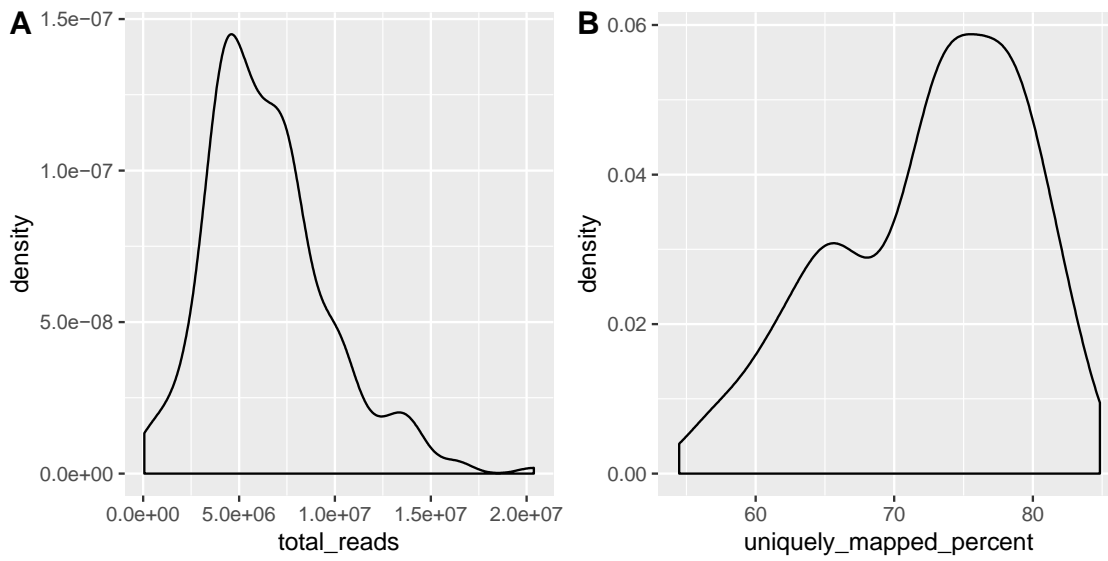


Figure 5.2. Read and alignment quality control plots. A: Density plot of number of reads across all samples after trimming step. **B:** Density plot of percentage of reads uniquely mapped to the genome across all samples.

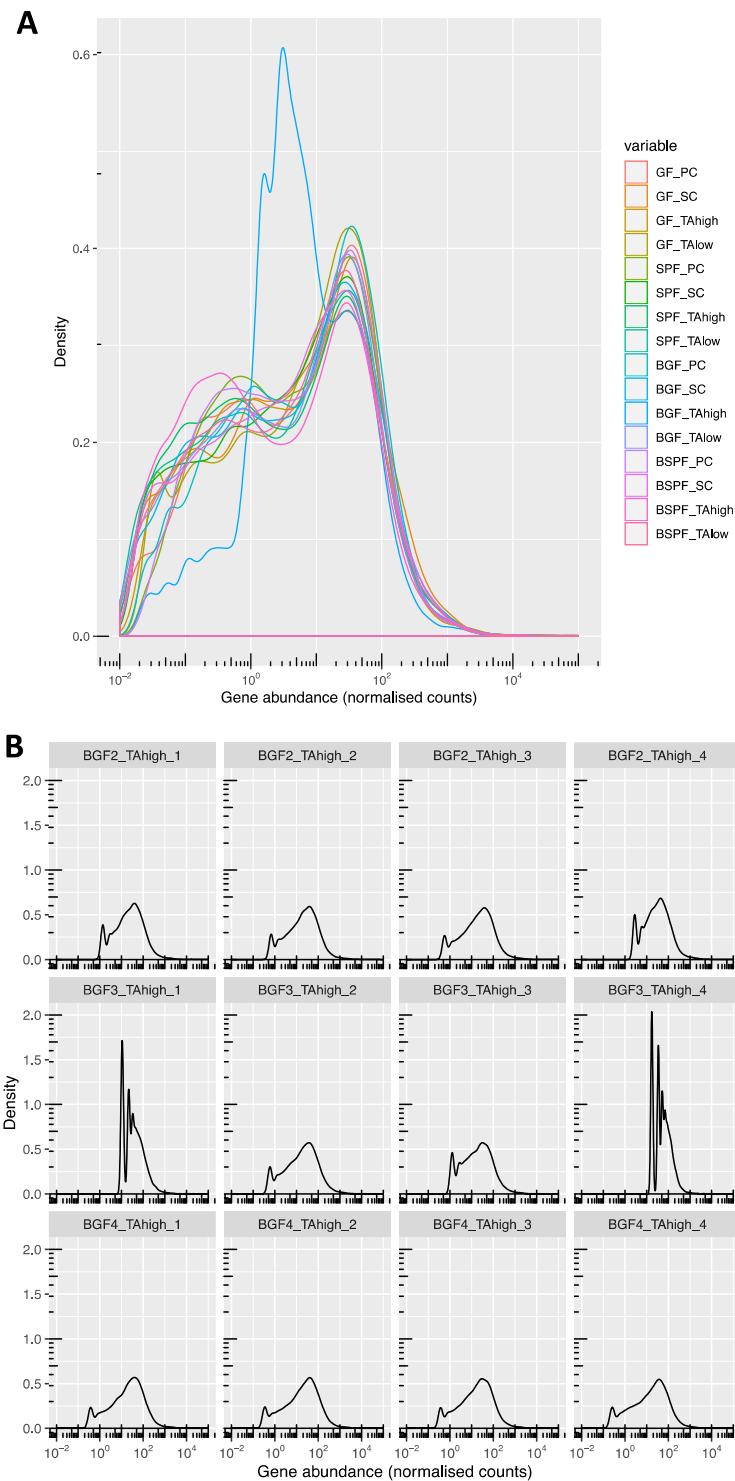


Figure 5.3. Normalised gene counts quality control plots. A: Density of gene abundance across each biological condition. **B:** Density of gene abundance across each replicate of the bifidobacteria treated germ free mice condition. BGF3_TAhigh_1 and BGF3_TAhigh_4 samples removed due to overrepresentation of genes with very low counts. GF - control germ free; SPF - control specific pathogen free; BGF - bifidobacteria treated germ free; BSPF - bifidobacteria treated specific pathogen free; PC - Paneth cell; SC - stem cell; TAhigh - transit amplifying high cell; TAlow - transit amplifying low cell.

Alignment and quantification was carried out using a traditional alignment approach instead of using a pseudo-aligner, for increased accuracy and gene abundance in addition to gained information regarding non-coding regions of the genome (Du et al., 2020). Specifically, reads were aligned to the Ensembl *Mus musculus* primary assembly (GRCm38.98) using STAR (v2.6.0c) (Dobin et al., 2013; Zerbino et al., 2018). The pipelining tool Snakemake (Köster and Rahmann, 2012) and the SLURM workflow manager (Yoo et al., 2003) were used to run these tools on the Earlham Institute High Performance Cluster.

Processing of aligned reads was carried out using the R package *Seurat* (v3.0) (Butler et al., 2018; Stuart et al., 2019). Normalisation of counts was performed using standard log-normalisation and z-score transformation, followed by removal of any samples with >15% mitochondrial reads. Cell cycle genes were not regressed out of the data due to well mixing of the S and G2M cell cycle scores among the samples based on Uniform Manifold Approximation and Projection (UMAP) visualisations. Each sequenced plate was treated as a separate dataset. A feature selection stage was used on every plate dataset to find the top 2000 highly variable genes using variance stabilising transformation. Then each plate dataset was integrated using Seurat's standard integration method to remove batch effects associated with sequencing by finding 'anchors' between the plate datasets. This produced an integrated expression matrix for each cell within dataset that enabled them to be analysed together. Visualisation of integrated normalised data was carried out using dimensionality reduction methods principal component analysis (PCA) and UMAP using R (Becht et al., 2018; McInnes et al., 2018).

Normalised counts data was plotted in R to determine the cut offs for expressed genes. The cut off for a gene to be expressed in a sample was determined as mean scaled normalised counts ≥ 0.02 (Figure 5.4A). The cut off for a gene to be expressed in a condition or cell type was determined as scaled normalised counts ≥ 0.02 in ≥ 3 samples per condition (Figure 5.4B, C), based on density plots. The 50 top variant genes across SPF and GF samples were identified and visualised using R.

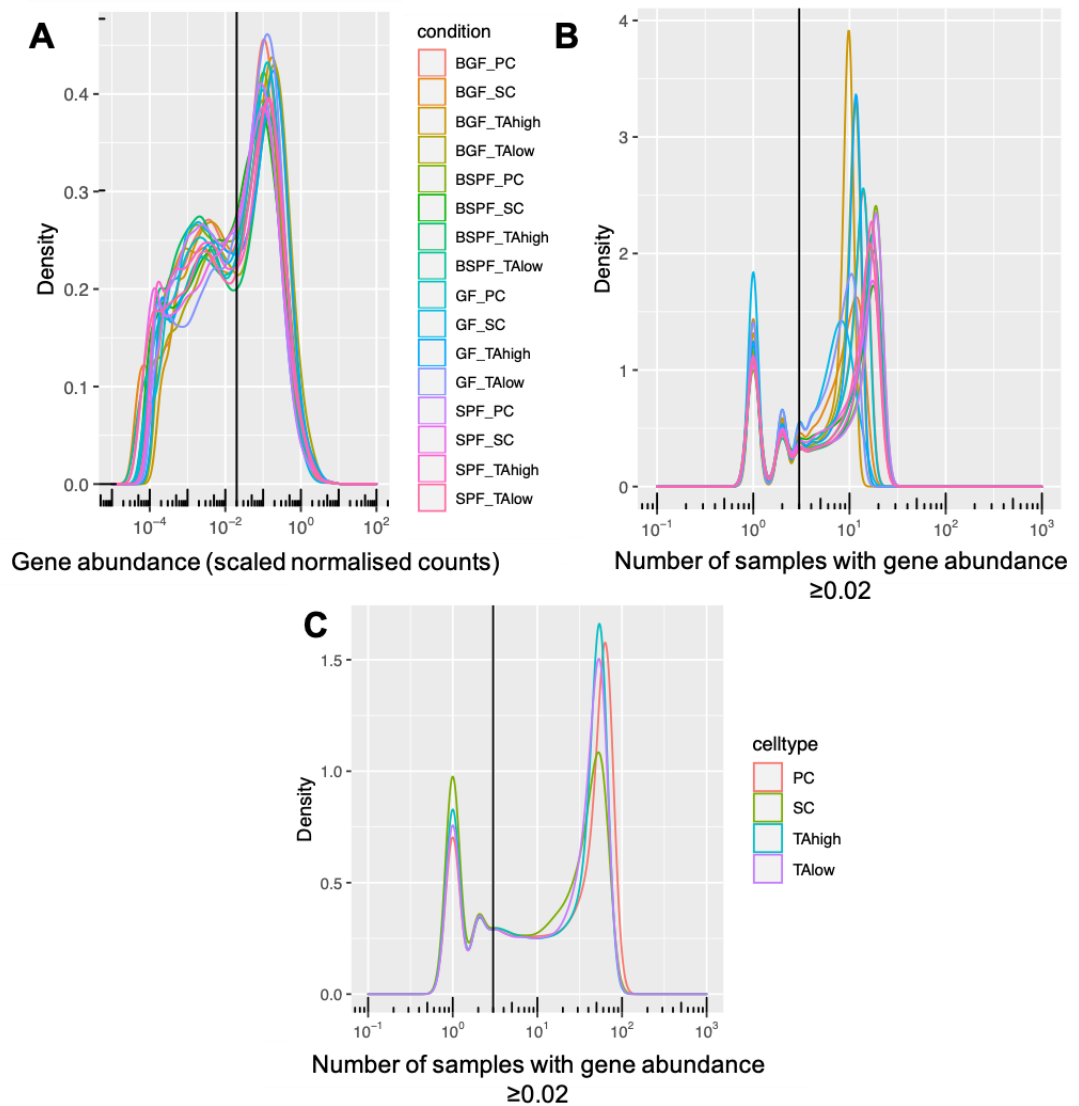


Figure 5.4. Gene abundance density plots for determining expression cut offs. **A:** Density of mean gene abundance (scaled normalised counts) across each biological condition. **B:** Density plot of number of samples of each biological condition where gene is expressed (scaled normalised counts ≥ 0.02). Legend as in A. **C:** Density plot of number of samples of each cell type where gene is expressed (normalised count ≥ 0.02). Vertical lines indicate expression cut offs. GF - control germ free, SPF - control specific pathogen free, BGF - bifidobacteria treated germ free; BSPF - bifidobacteria treated specific pathogen free; PC - Paneth cell; SC - stem cell; TAhhigh - transit amplifying high cell; TAlow - transit amplifying low cell.

5.3.7 Computationally assessing cell identities

MAST (v3.11) was used to identify differentially expressed genes (cluster biomarkers) relating to each cell type based on *a priori* labels (Finak et al., 2015). Here the gene expression of samples of the tested cell type were compared with expression of all the other samples. All genes with

adjusted p value ≤ 0.05 and average $\log_2(\text{fold change}) \geq 0.25$ were considered significant. These markers were compared to marker gene sets for murine intestinal epithelial cells from Haber *et al.* (Haber *et al.*, 2017) using hypergeometric significance calculations with all genes expressed in at least one sample (mean scaled normalised counts ≥ 0.02) as the statistical background. From Haber *et al.*, both droplet and plate-based results were used and gene symbols were converted to Ensembl IDs using db2db (Mudunuri *et al.*, 2009). Further, the same analysis was carried out using intestinal epithelial cell signature genes from Zhao *et al.* (Zhao *et al.*, 2020). Adjusted p values ≤ 0.05 (Bonferroni multiple correction) were considered significant. Data was plotted in R using *gplots*.

5.3.8 Differential expression analysis

Differential expression was carried out using MAST (v3.11) (Finak *et al.*, 2015). Differentially expressed genes were determined by comparing *Bifidobacterium* – exposed mice to control mice in each condition (cell type + mouse type). Any genes with absolute $\log_2(\text{fold change}) \geq 1$ and adjusted p value ≤ 0.05 were considered differentially expressed.

5.3.9 Weighted gene co-expression network analysis

Weighted-gene co-expression network analysis (WGCNA) was carried out using the automatic one-step method of the *WGCNA* R package (Langfelder and Horvath, 2008). The top 25% genes based on variance of normalised counts were pre-selected before being passed into the R package *ComBat* to remove batch effects due to the sequencing plate. Network topology analysis was carried out and scale-free fit and mean connectivity were plot against soft-thresholding powers between 1 and 20. Subsequently, the soft threshold was selected as the lowest power for which the scale-free topology fit index curve flattens out upon reaching a high value. Hierarchical clustering was used to identify and remove outlier samples. Signed networks were reconstructed in one block using default parameters. Module - trait associations, module membership and gene significant analyses were carried out using default parameters. Modules with p value ≤ 0.05 with respect to trait associations were considered significant. Functional analysis of modules was carried out against Gene Ontology Biological processes (GO:BP) and Reactome Pathways as described in section 5.3.10 using all genes expressed in at least one sample (mean scaled normalised counts ≥ 0.02) as the statistical background.

5.3.10 Functional analysis

Functional overrepresentation analysis and gene set enrichment analysis (GSEA) were carried out against Reactome and GO:BPs using R packages *ReactomePA* and *ClusterProfiler* respectively (Subramanian et al., 2005; Yu and He, 2016; Yu et al., 2012a). Where necessary, analysis was carried out following conversion to ENTREZ IDs using the R package *org.Mm.eg.db*. All expressed genes were used as the statistical background and redundant GO:BP pathways were removed using the *simplify* command. Any functions/pathways with q value ≤ 0.1 were deemed significantly enriched.

5.4 Results

5.4.1 *B. breve* has a modest impact on juvenile intestinal epithelial transcriptome

To determine the impact of bifidobacteria on Paneth cells and stem cells of the small intestine, we gavaged juvenile (4-week-old) SPF and GF mice with *B. breve* UCC2003. The correct monocolonisation of GF mice by *B. breve* was shown by the plating of caecal content, which showed high population levels across all repeated experiments ($\sim 5 \times 10^9$ CFU/mL/g) post-gavage (Figure S5.2A). Conversely, *B. breve* UCC2003 numbers in SPF mice faecal samples pre- and post-gavage (24h) were not significantly different, indicating that i) bifidobacteria were already present in the complex SPF microbiota and that ii) *B. breve* either transits through the gut and is rapidly undetectable, or has replaced (at least partially) the normal resident bifidobacteria population (Figure 5.1B). Of note, *B. breve* UCC2003 could colonise the gastrointestinal tract of GF mice at significantly higher levels than the levels at which resident bifidobacteria were in SPF mice before/after gavage, illustrating the highly controlled balance of abundance and functions between all components of a healthy gut complex microbiota.

Following FACS sorting and low input RNA sequencing of Paneth cells and stem cells, I carried out visualisation of the transcriptomics data using a non-linear dimensionality reduction approach called uniform manifold approximation and projection (UMAP) (Becht et al., 2018; McInnes et al., 2018). The results demonstrated a difference in transcriptomic profiles based on cell type: with a continuous pattern of variation across the cell types, in line with knowledge of crypt differentiation (Figure 5.5A) (Lueschow and McElroy, 2020). However, no distinct separation was observed between bifidobacteria-exposed and control samples or between GF and SPF samples (Figure 5.5B, Figure S5.3). This indicated that, based on all genes in the genome, there was no observable effect of *B. breve* UCC2003 on any of the four cell populations tested. The same finding was replicated using other dimensionality-reduction visualisation methods (Figure S5.4).

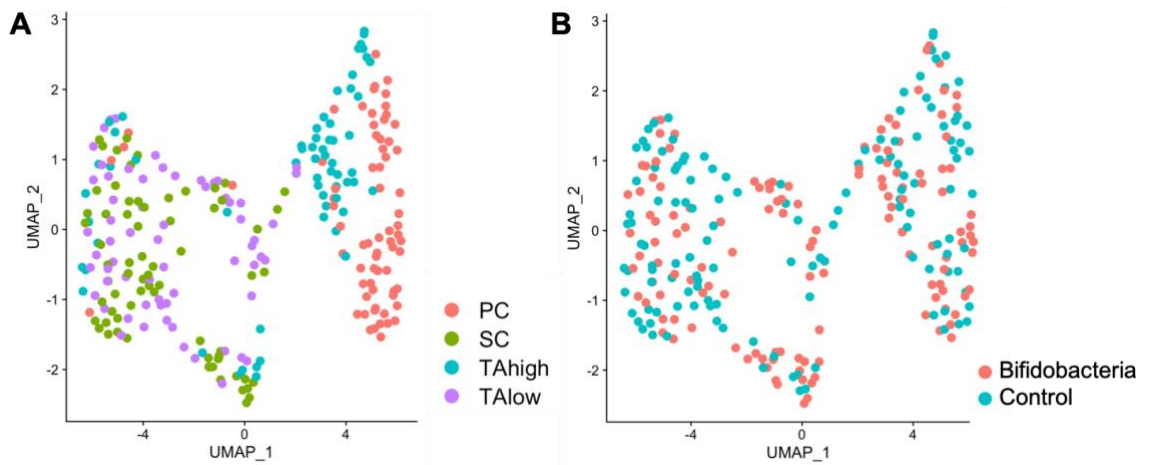


Figure 5.5. UMAP plots of normalised counts data. A. Data coloured by cell type. PC - Paneth cell; SC - stem cell; TAhigh - transit amplifying high; TAlow - transit amplifying low. **B.** Data coloured by condition: bifidobacteria-treated or control.

Next, the top variant genes across SPF and GF samples were identified and visualised (Figure 5.6, Figure S5.5). Based on hierarchical clustering of the top 50 variant genes, I confirmed that samples do not separate based on presence of bifidobacteria but instead based on cell type. More than 1/3rd of the variant genes among the SPF and GF samples were encoding antimicrobial proteins, in particular Paneth cell-associated defensins. Other variant genes were associated with epithelial cell types such as enterocytes (apolipoprotein A-I, *Apoa-1* and gastrotropin, *Fabp6*) and goblet cells (trefoil factor 3, *Tff3* and zymogen granule protein 16, *Zg16*) (Haber et al., 2017). Visual analysis indicated idiosyncrasies within the cell populations. For example, a subpopulation of SPF TAhigh cells expressed higher levels of enterocyte-associated genes and a subpopulation of SPF TAlow cells expressed higher levels of goblet-associated genes. Furthermore, a few mitochondrial associated genes are among the top variant genes (e.g. cytochrome b, *Cytb* and NADH dehydrogenase 1, *Nd1*). This could be evidence of apoptotic or lysing cells caused by experimental procedures, or a genuine biological signal relating to cell type-specific expression levels.

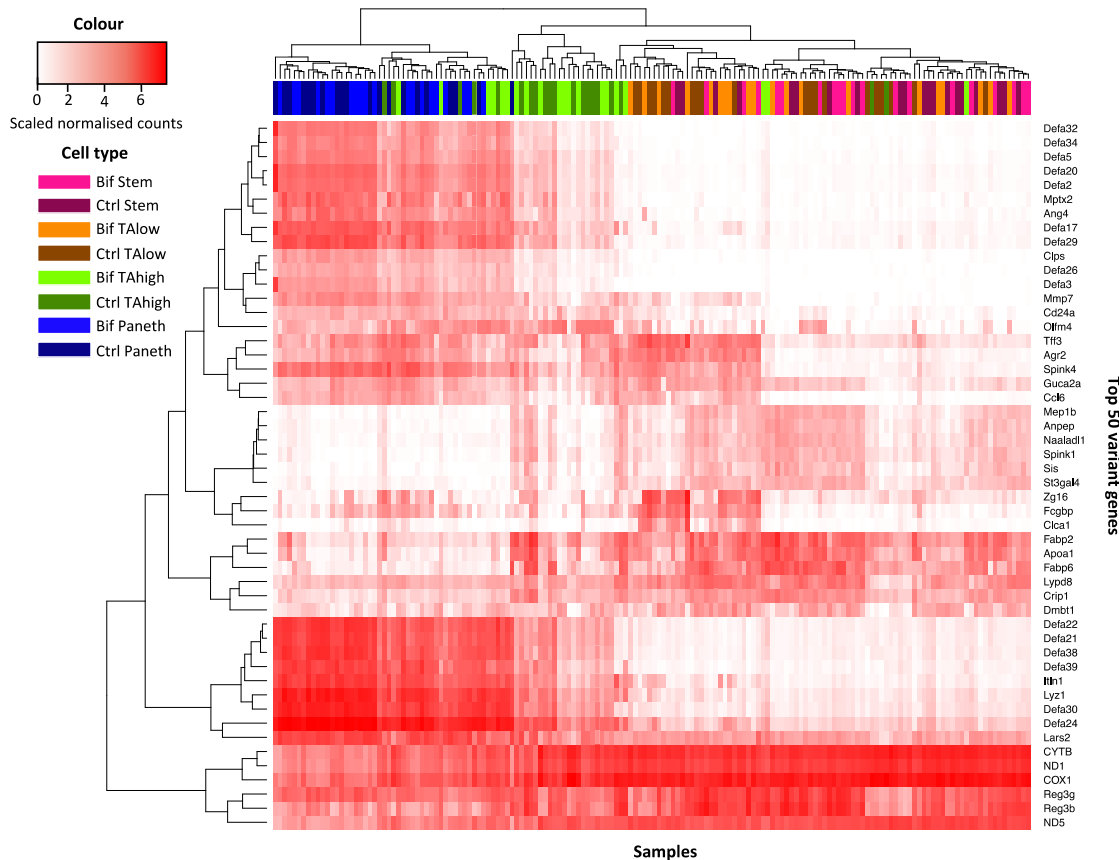


Figure 5.6. Gene expression heatmap of top 50 variant genes among all specific pathogen free samples. Samples and genes are clustered. Bif - bifidobacteria-treated samples; Ctrl - control samples.

5.4.2 Intestinal stem cells express cell surface protein CD24a

To further investigate and validate the identity of the flow-sorted cells, I determined differentially expressed genes (also known as cluster biomarkers) for each cell population. I compared these genes to two different published collections of IEC marker genes obtained through extensive single cell sequencing of mouse small intestinal epithelia (Haber et al., 2017; Zhao et al., 2020) using hypergeometric significance test (see Methods section 5.3.7). Using this approach, I found that the Paneth cell samples, which were sorted based on of $LGR5^{low}CD24a^{high}$ marker fluorescence (see Methods section 5.3.4), were significantly similar to Paneth cells from the two single cell studies. In addition, they were significantly similar to other closely related secretory lineages and immature enterocytes (Figure 5.7). TAlow cells, which were sorted based on having of $LGR5^{low}CD24a^{low}$ marker fluorescence, were most similar to enterocytes. Unexpectedly, stem cells, sorted based on of $LGR5^{high}CD24a^{low}$ marker fluorescence, were also

found to be significantly associated with enterocyte markers, but not with expected stem cell markers. Conversely, the TA^{high} cells, which had both of LGR5^{high}CD24a^{high} marker fluorescence, were significantly overlapping with stem cell markers, in addition to transit amplifying cells, tuft cells and enteroendocrine cells. Furthermore, the levels of *Lgr5* expression across the different samples does not match the expectation based on the markers used for the FACS. Where we would expect to see high expression in the samples labelled stem and the samples labelled TA^{high}, instead it is the Paneth cell and TA^{high} samples which have high *Lgr5* expression (Figure S5.6). For clarity, the remainder of this chapter continues to use the preassigned cell type labels, despite this uncovered uncertainty regarding their identify.

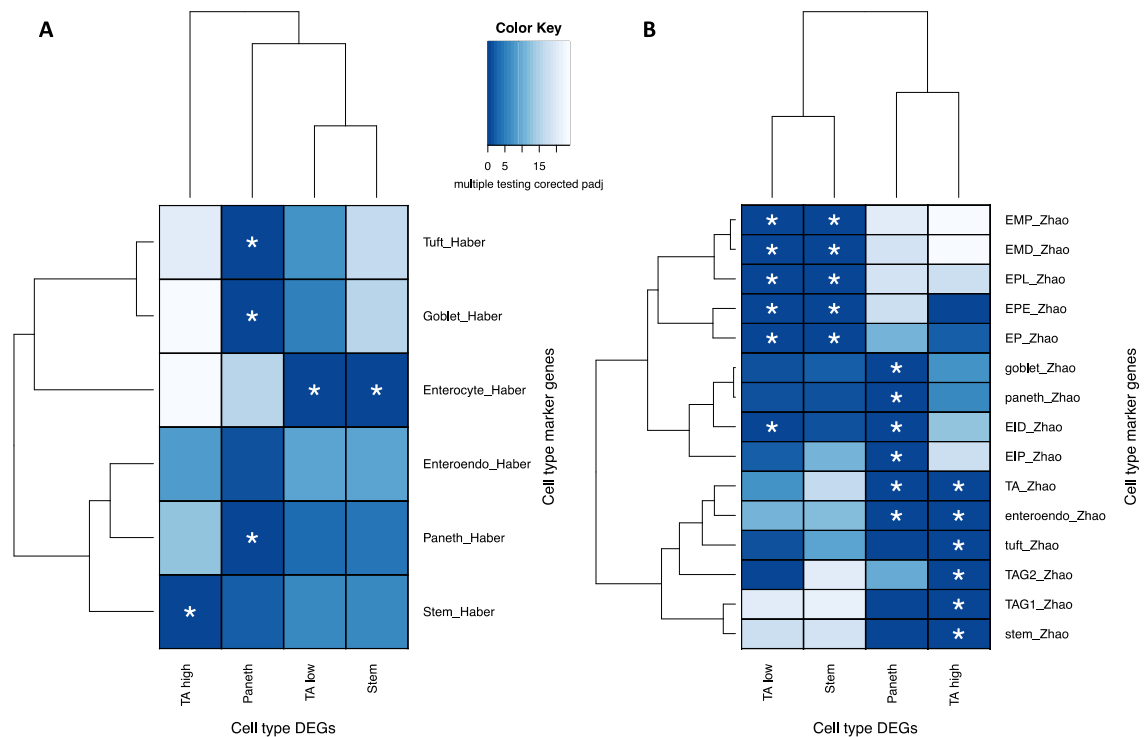


Figure 5.7. Differentially expressed genes for each cell population compared to intestinal epithelial cell marker gene sets. White asterisks indicate statistically significant overlaps (adjusted p value ≤ 0.05 , hypergeometric significant test, Bonferroni multiple correction) **A.** Marker gene set from Haber et al. (2017). **B.** Marker gene set from Zhao et al. (2020). DEGs - differentially expressed genes; TA^{high} - transit amplifying high; TA^{low} - transit amplifying low. enteroendo - enteroendocrine; EMP - enterocyte mature proximal; EMD - enterocyte mature distal; EPL - enterocyte progenitor late; EPE - enterocyte progenitor early; EP - enterocyte progenitor; EID - enterocyte immature distal; EIP - enterocyte immature proximal; TA, TAG1, TAG2 - transit-amplifying progenitors.

5.4.3 Correlation analyses reveal multiple cellular functions marginally affected by *B. breve* in juvenile intestinal epithelial cells

Next, I carried out differential expression analysis to specifically compare *B. breve* exposed mice to control mice for each cell type/mouse type. Contrasting the neonatal bulk epithelium study, no genes were identified as differentially expressed upon bifidobacteria exposure in any of the conditions (absolute Log₂FC ≥ 1, adjusted p value ≤ 0.05). Further, I employed a gene set enrichment analysis (GSEA) to investigate functional associations relating to modest changes in expression of genes based on Reactome and Gene Ontology Biological Process (GO:BP) annotations (Ashburner et al., 2000; Fabregat et al., 2018b; Subramanian et al., 2005; The Gene Ontology Consortium, 2017). This analysis employs a rank-based approach using the log₂ fold change of all tested genes (see Method section 5.3.10). A large number and range of pathways and functions were found significantly enriched (q value ≤ 0.1) (File S5.1). The functional results were similar between the different mouse and cell types, indicating a common response to the bifidobacteria. Based on GO:BP annotations, this response includes an innate and humeral immune response, ribosome biogenesis and protein and lipid localisation (summarised in Table 5.2). Interestingly, cell-cell adhesion was identified in the SPF stem cells, reflecting previous evidence of bifidobacteria effect on epithelial barrier function. Furthermore, Reactome results included the following pathways and functional groups (File S5.1):

- Cell cycle, translation and metabolism
- Immune functions such as defensins, cytokines and neutrophils
- Apoptosis and autophagy
- Cell-cell junctions and membrane trafficking

Mouse type	Cell type	Functions summary (GO:BP)
GF	Paneth	ribosome biogenesis; membrane disruption in other organism; innate immune response
GF	Stem	anion transport; humoral immune response; organic hydroxy compound metabolic process
GF	TAhigh	ribosome biogenesis; membrane disruption in other organism; antibacterial humoral response; lipid localisation
GF	TALow	mitochondrion organization; ATP metabolic process; membrane disruption in other organism; protein localisation; humoral immune response; mitotic nuclear division
SPF	Paneth	ribosome biogenesis; innate immune response; membrane disruption in other organism; protein localisation
SPF	Stem	ribosome biogenesis; anion transport; cell-cell adhesion; inflammatory response; protein localisation
SPF	TAhigh	monocarboxylic acid metabolic process; membrane disruption in other organism; mucosal immune response
SPF	TALow	ribosome biogenesis; inflammatory response

Table 5.2. Summary of overrepresented Gene Ontology biological processes. Data based on gene set enrichment analysis comparing bifidobacteria-exposed to control mice in each mouse type and cell type. q value ≤ 0.1 . GO:BP - Gene Ontology biological process; GF - germ free/monocolonised mouse; SPF - specific pathogen free/conventionalised mouse; TAhigh - transit amplifying high cell; TALow - transit amplifying low cell. For full dataset see File S5.1.

Weighted gene co-expression network analysis (WGCNA) is a systems biology method to identify patterns of correlation within genes based on expression data (Zhang and Horvath, 2005; Langfelder and Horvath, 2008). I applied WGCNA to predict clusters of significantly co-expressed genes whose expression is correlated to the presence or absence of bifidobacteria (see Methods section 5.3.9). This analysis was carried out on each condition (mouse type + cell type) separately to permit identification of condition-specific features. However, only two conditions contained gene clusters which were significantly associated with presence of bifidobacteria (adjusted p value ≤ 0.05) – GF Paneth cells and SPF stem cells. The GF Paneth cell data contained 12 separate modules which were significantly associated with presence of bifidobacteria, totalling 3891 genes (Figure 5.8A). A functional overrepresentation analysis (using GO:BP) was carried out on all the module genes together as a separate analysis yet it did not provide significant results. A number of enriched functions were identified including mitotic nuclear division, chromatin organisation, autophagic mechanisms and membrane permeability (Figure 5.8B). On the other hand, the SPF stem cell data contained only one significant module of 51 genes, which was

functionally related to plasma membrane bounded cell projection assembly and ruffle organisation.

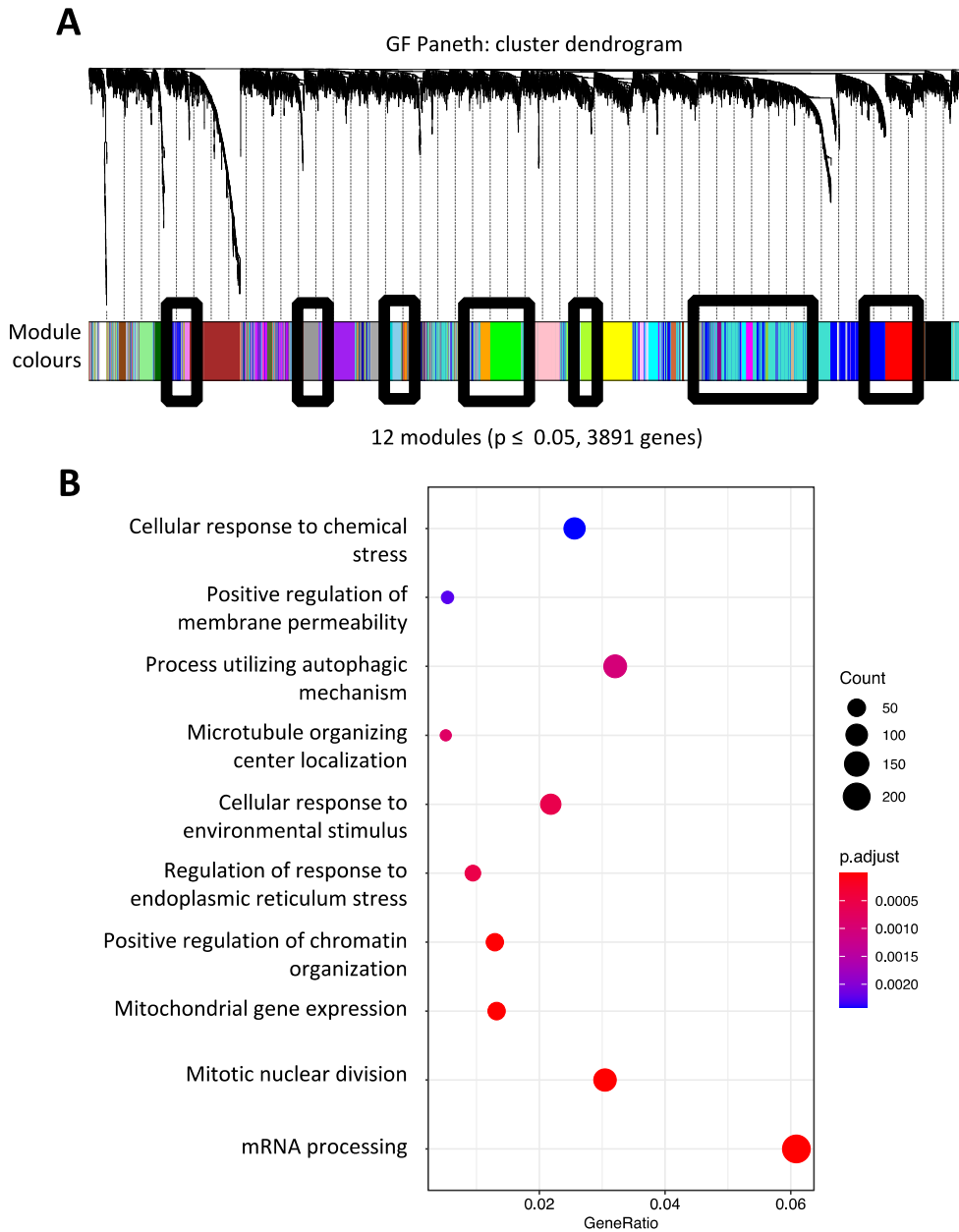


Figure 5.8. WGCNA results of germ free/monocolonised mice Paneth cell expression data (top 25% variant genes). **A.** WGCNA cluster dendrogram highlighting modules which are significantly associated (adjusted $p \leq 0.05$) with the presence of bifidobacteria - making up a total of 3891 genes. **B.** Gene ontology biological processes overrepresented among the 3891 significant module genes (q value ≤ 0.01).

Finally, I investigated the possibility that bifidobacteria could be affecting the direction of stem cell differentiation by observing expression levels of key differentiation-related transcription factors (TFs) (as described in General Introduction section 1.2.2). For this analysis I focused on stem cells and secretory lineages, thus excluding the T_{ALow} cells. Based on the SPF samples, I observed that Paneth cell and T_{Ahigh} cell samples expressed the highest stem cell-associated TFs: Achaete-scute family BHLH transcription factor 2 (*Ascl2*), Transcription factor 4 (*Tcf4*), Jun proto-oncogene AP-1 transcription factor subunit (*Jun*) (Figure 5.9A) (van der Flier et al., 2009; Sancho et al., 2009; Schuijers et al., 2015). This finding concurs with previous finding that the samples sorted as T_{Ahigh} cells are likely to be stem cells, whereas the cells sorted as stem cells are not. Further, I observed that SRY-box transcription factor 9 (*Sox9*), a TF associated with differentiation of Paneth cells (Bastide et al., 2007), was also highly expressed in the T_{Ahigh} cells in addition to Paneth cells. However, literature suggests that *Sox9* is also expressed in stem cells (Bastide et al., 2007; Haber et al., 2017; Jo et al., 2014). Other genes investigated were:

- Enterocyte-associated TF gene, Hairy and enhancer of split-1 (*Hes1*) (Worthington et al., 2018)
- Secretory lineage TF genes, Atonal homolog 1 (*Atoh1*) and CBFA2/RUNX1 partner transcriptional co-repressor 2 (*Cbfa2t2*)
- Paneth and goblet TF gene, Growth factor independent 1 (*Gfi1*)
- Goblet cell TF gene, Krüppel-like factor 4 (*Klf4*)
- Tuft cell TF gene, POU class 2 homeobox 3 (*Pou2f3*)
- M cell TF gene, Spi-B transcription factor (*Spib*)
- Enteroendocrine TF gene, Neurogenin 3 (*Neurog3*) (Amann et al., 2005; Worthington et al., 2018).

However, none of the observed TF genes were differentially expressed between the bifidobacteria treated and control samples, in the SPF mice or GF mice, indicating that bifidobacteria does not affect cellular differentiation in this context (Figure 5.9, Figure S5.7).

Taken together, these findings indicate a number of cellular functions which may be affected by bifidobacteria in Paneth cells and stem cells. However, the lack of DEGs between bifidobacteria treated and control samples indicates that any effect is very small.

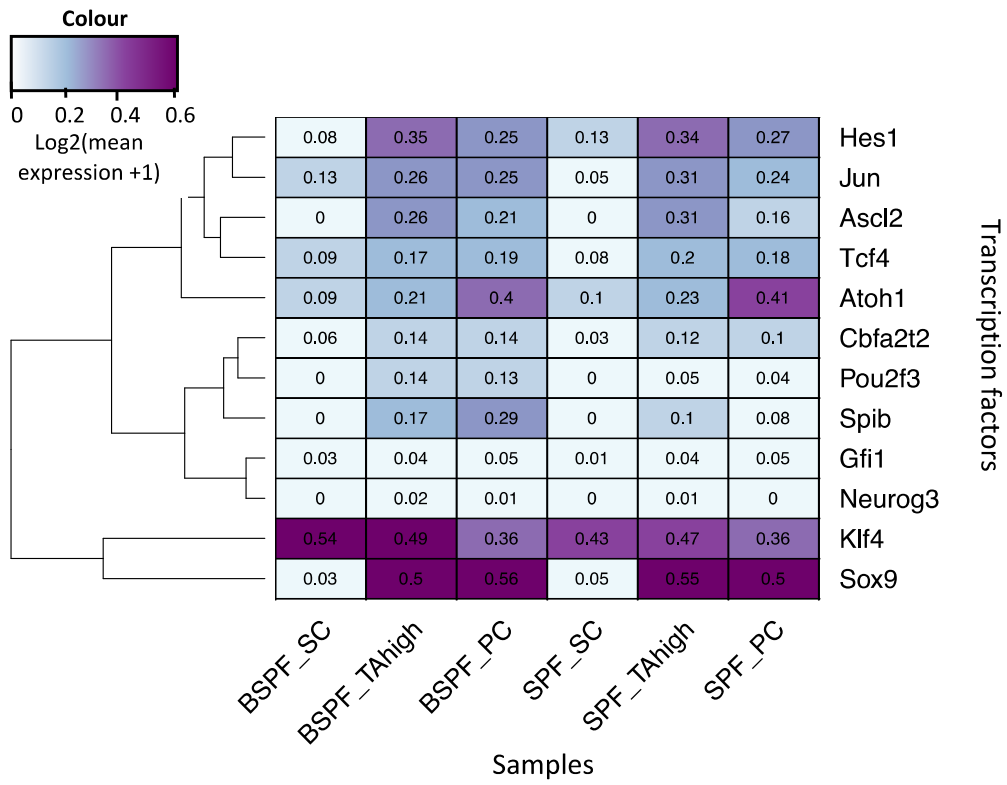


Figure 5.9. Expression of key transcription factors in intestinal epithelial cell differentiation across specific pathogen free mice samples. Mean scaled expression for each SPF cell type. PC - Paneth cell; SC - stem cell, TAhigh - transit amplifying high cells.

5.5 Discussion

In 1900, *Bifidobacterium* was first identified as a health promoting bacterial genus (Tissier, 1899, 1900). However, despite plentiful research into bifidobacteria and their beneficial effects, much remains unknown about their mechanisms of action. Furthermore, many studies have focused on the interplay between bifidobacteria, the immune system and disease, while interactions with IECs (the gut's frontline physical barrier) have been largely understudied (O'Neill et al., 2017). As previous studies have mostly targeted specific pathways and functions using *in vitro* or diseased conditions, we aimed to investigate the global transcriptomics effect of *B. breve* UCC2003 on healthy mouse small intestinal epithelial cells *in vivo* (Boesten et al., 2011; Hsieh et al., 2015; Yang et al., 2017). In Chapter 4, myself and colleagues observed that *B. breve* had a global impact on the IEC transcriptome. Based on these results and previous literature evidence, we investigated the effect of *B. breve* UCC2003 on stem cells and Paneth cells of the murine small intestine (Lee et al., 2018). Here, we studied both SPF and GF mice to determine if bifidobacteria plays a different role - given that the microbiota is key to development and priming the gut and its associated immune system (Jiao et al., 2020). Due to technical challenges relating to GF mice, it was necessary to study four-week-old mice. While these mice are considered juvenile, they were recently weaned onto solid food, whereas the two-week-old mice from the neonatal study were still nursing. One possible explanation for the lack of transcriptional differences in the juvenile mice exposed to bifidobacteria compared to the neonatal mice, would be that *B. breve* is only modulatory during this very early life period. Additionally, it is also possible that the reduced gavage protocol and increased time between gavage and sample collection could have affected the observed impact of bifidobacteria. However, through plating we confirmed that the GF mice were successfully colonised with bifidobacteria on the day of sample collection, suggesting that the gavage protocol was sufficient in these mice. Additional metagenomic or 16s RNA sequencing of the caecal or faecal contents (samples collected but not analysed) would be required to confirm whether *B. breve* successfully colonised the SPF mice at 72 hours, or whether their effect was transient. Two further possible explanations for the lack of a strong response to bifidobacteria in the juvenile mice is that the cell populations investigated are not the affected cells or that the cell isolation and FACS protocol perturbed gene expression. However previous literature evidence suggests that these scenarios are unlikely (Lee et al., 2018; Richardson et al., 2015).

Data visualisation and analysis of top variant genes confirmed that, at the whole transcriptome level, the samples separated primarily by cell population. This confirmed that the FACS protocol successfully selected four different cell populations and highlighted the continuous nature of cellular differentiation within the intestinal crypts. The top variant genes were primarily Paneth cell associated genes, further supporting success of the FACS. However, a number of enterocyte and goblet cell associated genes were among the top variant genes, indicating that the FACS markers employed are not fully selective for the populations of interest and further highlighting the continuous nature of differentiation in the epithelium, where many semi-differentiated cells express multiple different cell type markers simultaneously. The addition of further marker proteins to the FACS protocol could potentially exclude progenitor cells destined for other cell populations. Although, adding FACS markers rapidly increases the complexity of the sort protocol and requires extensive testing procedures. Alternatively, microfluidic or droplet based single cell sequencing, such as Chromium from 10X Genomics, could be used to sequence all cells of the intestinal crypts, avoiding issues with cell sorting. The disadvantages of this type of approach include cost, sequencing depth and the large number of input cells required, which can be particularly challenging when the primary research focus is on small cell populations such as stem cells (See et al., 2018).

The FACS protocol employed in this study for sorting cells was optimised from Yilmaz *et al.* (Yilmaz et al., 2012). Here the authors sorted stem cells from *Lgr5-EGFP-IRES-creERT2* knock-in reporter mice, allowing isolation by flow cytometry of *Lgr5*^{high}-EGFP^{high} cells which also have low CD24a marker fluorescence. In our study we used a similar protocol but employed wildtype mice and measured surface LGR5 marker fluorescence, targeting the presence of LGR5 protein instead of its gene expression. Furthermore, in addition to sorting stem cells and Paneth cells based on the Yilmaz protocol, we sorted intermediate cells with high fluorescence of both LGR5 and CD24a (TA^{high}) or low fluorescence of both (TA^{low}) (see Methods section 5.3.4, Table 5.1).

Using IEC marker genes obtained from two independent mouse single cell studies, I found that all sorted cells were significantly similar to at least one IEC cell type based on global expression levels. This confirmed success in sorting live IECs. However, the cells sorted using the Yilmaz method of LGR5^{high}CD24a^{low} were most similar to enterocytes, rather than stem cells as expected (Yilmaz et al., 2012). The reason for this discrepancy is currently unknown and thus requires further investigation. It is possible that this result is in part due to differences between *Lgr5*-

EGFP reporter expression which is used in the Yilmaz protocol, and LGR5 cell surface marker fluorescence which was applied here - especially as these cells do not appear to be expressing high levels of *Lgr5* based on our transcriptomics results. Potentially we extracted cells which retained LGR5 surface proteins after the cells had stopped expressing *Lgr5*. Given that all the cells sorted were obtained from intestinal crypts, this cell population are likely to be semi-differentiated/transit amplifying enterocytes. In fact, enterocyte precursors are among the most abundant semi-differentiated cells within the crypt (Bankaitis et al., 2018), and it has been shown that enterocyte precursors can dedifferentiate to regain stemness upon loss of LGR5⁺ stem cells (Jones and Dempsey, 2016; Tetteh et al., 2016). However, this justification does not explain why we did not obtain cells with high LGR5 and low CD24a fluorescence which were also expressing high *Lgr5*. Based on the FACS gating plots (Figure S5.1E), following extraction of cells with low CD24a and high epithelial cell adhesion molecule (EpCAM), no population was excluded which had higher LGR5 fluorescence, suggesting this population did not exist in our samples. However, this could be due to the rarity of these cells within in crypt.

On the other hand, the TA^{high} cells sorted based on LGR5^{high}CD24a^{high} were significantly similar to stem cells, suggesting that intestinal stem cells express high *Cd24a*. CD24a, also known as signal transducer 24, is a small glycosylated cell surface protein expressed in many different cell types (Liu et al., 1992; Stutte et al., 2008). It has been shown to act as a costimulatory molecule for T-cells and dendritic cells, but its function in the intestines is currently not known. Previous work has shown that intestinal stem cells express *Cd24a* and can be sorted based on presence of CD24a cell surface markers (King et al., 2012). However other experiments have shown that LGR5⁺ stem cells have very low or no CD24 marker expression, whereas other 'reserve' stem cell populations in the 4⁺ position of the crypts have more CD24a (Gracz et al., 2010, 2013). Investigation of CD24a levels by von Furstenberg et al. (2011) found that the epithelial crypt population with present but low CD24a contained the most actively cycling cells. Authors carried out further investigation of the CD24a^{low} population confirming that the cells have an intestinal stem cell or Paneth cell phenotype, however they did not carry out further characterisation of the CD24a^{high} population. Wang et al. (2013) used a larger panel of markers to find that CD44⁺CD24^{low}CD166⁺ cells express many stem cell associated genes, whereas CD44⁺CD24^{high}CD166⁺ express secretory cell marker genes. Yilmaz et al. (2012) also used low CD24a to sort stem cells from mouse small intestinal crypts, but extensive validation of stem cell identity was not carried out in this experiment. Taken together, literature evidence suggests

that low CD24a should identify intestinal stem cells, but questions remain regarding exactly which populations are sorted by which marker quantities. As far as we are aware, no previous work has described a cell population with high LGR5 and high CD24a, as we identified here. Therefore, more research is required to determine the cell identities sorted here and to evaluate the marker panel more thoroughly. Potentially the expression of CD24a is dependent on outside factors such as the age or strain of mouse or, given its role in immune activation, expression could depend on the gut microbiota. Alternatively, different populations of stem cells may express different levels of *CD24a* within one intestinal crypt, as suggested by Gracz et al. (2010, 2013). Based on this theory, the T^A_{high} cells we isolated in this experiment could be ‘reserve’ 4+ stem cells, although we found they also expressed high levels of *Lgr5*. An alternative explanation is that the T^A_{high} cells are not fully differentiated Paneth cells that can revert to a stem cell phenotype, or early transit amplifying cells that have not completely reprogrammed their gene expression profile and retain markers of stem cells (Buczacki et al., 2013; Roth et al., 2012). In comparison, the Paneth cells which we sorted based on the Yilmaz method of LGR5^{low}CD24a^{high} were found to be significantly similar to Paneth cells and the T^A_{low} cells (LGR5^{low}CD24a^{low}) were similar to enterocytes, confirming the success sorting these populations. In the future, *in situ* hybridisation and electron microscopy approaches should be used to confirm the cell identities and their localisation within the crypt. Moreover, different cell marker panels could be employed to isolate intestinal stem cells.

Whilst no genes were differentially expressed between the bifidobacteria treated mice and the control mice, correlation-based analyses revealed some significant differences in functional profiles. Many of the identified functions (such as cell-cell junctions and cell cycle) agree with published research and/or the findings of our neonatal study (Din et al., 2020; Yan et al., 2019). However, the lack of differentially expressed genes indicated that these findings are minor in magnitude and further experiments would be required to confirm their validity.

WGCNA analysis identified a collection of 3891 genes whose expression was correlated with presence of bifidobacteria in the GF Paneth cells, and whose associated functions closely compared to the findings of the previous neonatal study (Chapter 4). However, removal of batch effects for the WGCNA analysis was carried out following selection of features based on variance. Given that batch normalisation alters variance of data and that low variance samples can aid batch normalisation, this method may have slightly altered the results. In future, batch

normalisation should be carried out before feature selection. Amongst the results, I discovered cell projection assembly-associated genes correlating to bifidobacteria presence in the SPF stem cell data. Whilst it has been reported that bifidobacteria can inhibit membrane ruffles in gliadin-treated Caco-2 cells, further investigation revealed that the relevant genes were marginally upregulated in the bifidobacteria-exposed mice of our experiment (Lindfors et al., 2008). It is possible that another microbiota member is responsible for this signal, or it could be a false positive result due to overinterpretation of minor variation in the data. Regardless, this uncertainty combined with the lack of results in the other datasets suggests that any potential transcriptional change of four-week-old IECs due to bifidobacteria exposure is minimal in homeostatic conditions. However, it would be interesting to repeat such studies in disease model mice where the effect of bifidobacteria might be greater, or instead with a longer exposure to bifidobacteria.

5.6 Future research directions

In conclusion, we have shown that *B. breve* UCC2003 plays a central role in orchestrating global neonatal IEC gene responses. However, our experiment in juvenile mice demonstrates how much remains unknown about the effect of *B. breve* UCC2003 on IECs. Further study is required to determine whether the contrasting results from this study were due to the age of the mice, the reduced gavage schedule, the cell types investigated, or whether a real biological signal was masked by expression changes induced by the cell isolation and sorting protocol.

Additionally, further work should investigate the age at which *B. breve* ceases to significantly impact IECs in a healthy condition, as well as the length of time during which bifidobacteria can exert its effects following gavage. However, such experiments are challenging as they require a large quantity of mice with many distinct experimental conditions. In part, these challenges could be avoided by employing alternative experimental models such as 3D organoids – microinjected, grown with inverted polarity or cultured as 2D monolayers (Bartfeld, 2016; Co et al., 2019; Sato and Clevers, 2013). While these models can provide a simpler and higher throughput approach, they are further removed from physiological accuracy. For example, they cannot fully account for the epithelial-immune system interactions or the impact of mouse age or microbiota. Furthermore, strict anaerobiosis cannot be applied to the apical side of the organoid IECs without requiring highly complex microfluidics systems, such as HuMiX, which require extensive optimisation for this type of experiment (Shah et al., 2016). On the other hand,

organoid models could be valuable for testing a basic response of IECs to bifidobacteria, to identify affected cell types, to investigate strain specificity and to explore the mechanisms of interaction. For example, the use of mutants and transcriptionally active strains as positive controls, in tandem with metabolomic and proteomic approaches, is required to advance our understanding on the key host pathways and bifidobacterial molecules governing development and maturation of the intestinal barrier during the early life window. In the future, based on results from organoid research, more targeted mouse experiments could be carried out, followed by clinical studies to explore the application of findings to human health. Ultimately, understanding how specific microbiota members modulate host responses in pre-clinical models may help the design and development of next-stage targeted microbiota therapies in humans.

Chapter 6: Interactions between bifidobacteria and macrophages

6.1 Introduction

Clinical trials, *in vivo* experiments and *in vitro* experiments have indicated that bifidobacteria have immunomodulatory effects on their host (Ruiz et al. 2017). Indeed, a number of previous experiments have shown that different *Bifidobacterium* strains can affect macrophage function (He et al. 2002; Lee et al. 2012; Mokrozub et al. 2015). For example, Okada et al. (2009) showed that exposure of RAW264.7 macrophage-like cells to *Bifidobacterium* and lipopolysaccharide (LPS) significantly reduced proinflammatory cytokine production compared to with LPS alone. Macrophages are the most abundant white blood cells in the lamina propria of a healthy gut, playing a key role in bacterial recognition and phagocytosis as well as impacting epithelial cell regeneration, T cell differentiation and secreting anti-inflammatory cytokines. Moreover, macrophages are important for maintaining a balance between tolerance to commensal bacteria and attack against foreign antigens (Wang et al., 2019b). Dysregulation of this balance can contribute to gut pathologies such as inflammatory bowel disease, where the immune system is known to be over-activated by commensal bacteria in the gut (Zhang et al., 2017). Therefore, the effect of commensal bacteria on macrophages is important for elucidation of the anti-inflammatory mechanisms of probiotics. Whilst the effect of bifidobacteria strains on macrophages has been studied, many questions remain regarding the context-specific effect and the precise molecular mechanisms of interaction between macrophages and bifidobacteria (He et al., 2002; Lee et al., 2012; Mokrozub et al., 2015; Okada et al., 2009). Such information will help to uncover how the host immune system interacts with bifidobacteria to support the balance of inflammation in the gut, whilst also recognising aberrant bifidobacteria which have breached the cell epithelium. This knowledge may also lead to novel treatment methods to rebalance an aberrant intestinal immune response.

Bifidobacterium breve UCC2003 is a strain of bifidobacteria isolated from the stool of a breast fed infant which has been shown to confer health benefits, including protecting the murine host against bacterial infections and improving gut barrier function in neonatal mice (Fanning et al., 2012b, 2012a; Kiu et al., 2020) (Chapter 4). Some of the mechanistic factors driving the impact

of *B. breve* UCC2003 have been identified, including the type IVb tight adherence pili which can promote colonic epithelial proliferation, and the exopolysaccharide (EPS) capsule which can repress local T helper cell (Th)17 responses and reduce pro-inflammatory cytokine release (Fanning et al., 2012a; Hughes et al., 2017; O'Connell Motherway et al., 2011; Püngel et al., 2020). However, much remains unknown about the effect of *B. breve* on different cell types and the specific *B. breve* genes and molecules required for crosstalk with the host.

This final results chapter presents a collaborative project, carried out with members of Lindsay Hall's research group (Quadram Institute Bioscience, QIB), studying the interaction of *B. breve* UCC2003 and macrophages to gain a more thorough understanding of their role in human health. Here we have used macrophage nuclear factor (NF)- κ B activation as a proxy for macrophage activation, as NF- κ B is important for regulation of inflammatory response following activation of cell surface pattern recognition receptors (as described in General Introduction section 1.3.2). The experiments and data analyses covered in Chapter 6 are outlined in Figure 6.1. Specifically, using a mutant library of *B. breve* UCC2003 and a macrophage-like cell line (THP-1 cell) containing a NF- κ B reporter, Ian O'Neill (QIB, APC microbiome Ireland, APC) identified mutant strains with significantly greater or reduced NF- κ B activating ability. Further experimentation carried out by Sree Gowrinadh Javvadi (QIB) identified the effector molecule/s as secreted proteins (or protein-containing), while mass spectrometry and whole genome sequencing were employed to further characterise the molecule/s. Meanwhile, I applied a computational pipeline to predict possible protein-protein interactions (PPIs) between secreted bifidobacterial proteins and macrophage cell surface proteins based on domain-motif interactions. Unfortunately, no computational PPI predictions were made which correlated with experimental results. In fact, the experimental results, combined with further computational analysis indicate that the molecular interaction likely occurs *via* bifidobacterial secreted lipoproteins which often interact with host cells *via* their lipid moieties.

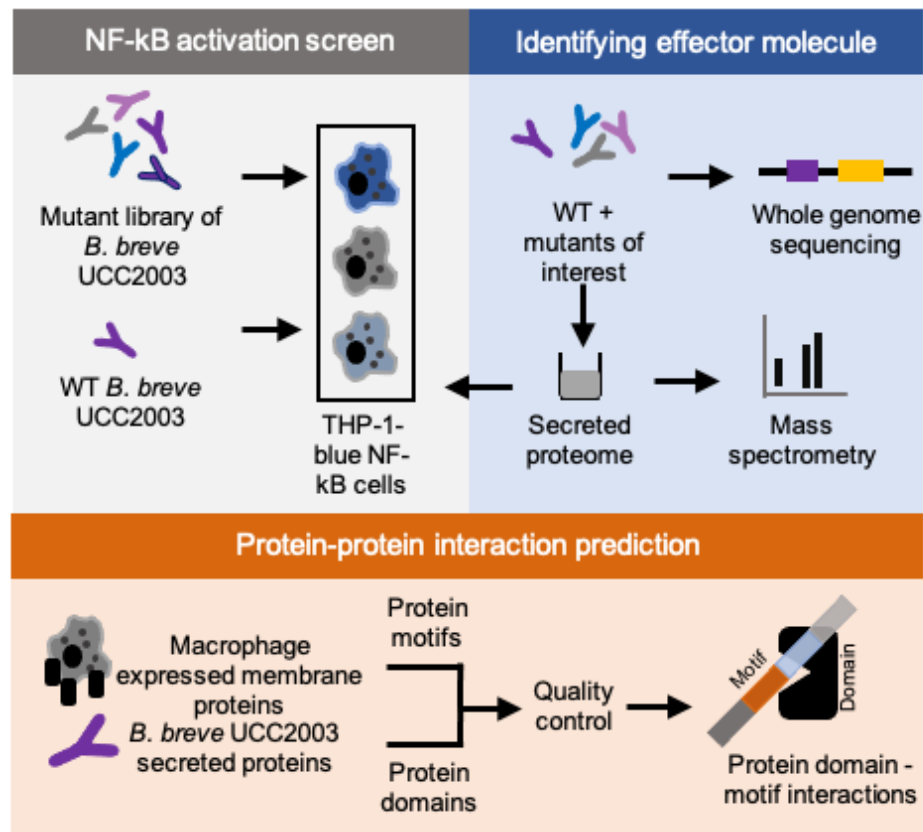


Figure 6.1. Schematic overview of the experiments and analyses carried out in Chapter 6.

6.2 Aims

The aims of this project were as follows:

- Identify *B. breve* UCC2003 mutants which have a significantly greater or lesser effect on macrophage NF-κB activation compared to the wild-type strain.
- Identify the effector molecules responsible using a combined experimental and computational approach.

6.3 Methods

All experimental work was carried out by Ian O'Neill and Sree Gowrinadh Javvadi of Lindsay Hall's research group in QIB. Liquid chromatography-mass spectrometry (LC-MS) was processed by the University of Bristol proteomics facility. Processing and analysis of whole genome sequencing data was carried out by Ian O'Neill (QIB, APC). All subsequent computational analysis and interpretation was carried out by myself.

6.3.1 NF- κ B activation screen

The effect of *B. breve* UCC2003 on NF- κ B activation in macrophages was assessed using a THP-1-blue NF- κ B cell reporter line (Invivogen, UK) (Zuliani-Alvarez et al., 2017). Cells were revived in growth medium RPMI 1640 with 2mM L-glutamine, 25mM HEPES (Merck), 10% heat-inactivated foetal bovine serum (Gibco), 100 μ g/ml Normocin (Invivogen), and Pen-Strep (100U/ml-100 μ g/ml) (Gibco). The Quanti-blue assay was conducted as per manufacturer instructions (Invivogen). In brief, once 70% confluence was reached, Blasticidin (Invivogen) treated THP1-Blue NF- κ B cells were transferred from cell culture flasks and seeded at 1×10^5 per well in a 96-well tissue culture plate. THP-1 cells were then treated with live bifidobacteria cells or with secreted proteome samples, as described below.

Briefly, *B. breve* UCC2003 mutants (n=2592) from a Tn5 insertion library were grown overnight in 96-well plates before 10 μ l of each bacterial culture was added to 96-well plates containing THP-1 cells (Ruiz et al., 2013). 24-hour incubations were carried out before NF- κ B activity was determined by colorimetric assay. Phosphate buffered saline was used as a negative control. Mutants which induced or inhibited NF- κ B levels more than twice the standard deviation away from the plate average for the wild-type (WT) *B. breve* UCC2003 were selected for additional incubation experiments in six-well plates to verify the phenotype. Subsequently, selected mutants were subjected to inverse PCR analysis to identify the genomic location of the mini-Tn5 insertion site. Whole genome sequencing was used to confirm the mutation site in the top selected *B. breve* UCC2003 mutants.

Following isolation of total proteome (Methods section 6.3.2) THP-1 cells were stimulated with 5 μ l (3.6mg/mL) total proteome of WT and mutants and incubated for 18 hours in a humidified incubator at 37 $^{\circ}$ C with 5% CO₂. LPS provided by manufacturer was used as a positive control.

NF- κ B activity was determined by colorimetric assay as per manufacturer's instructions. Different molecular weight fragments of total proteome were subsequently tested for NF- κ B activation using the same protocol.

6.3.2 Isolation of total proteome

To test the effect of *B. breve* secreted proteins on NF- κ B activation, THP-1 cells were incubated with extracellular proteome extracted from the WT and top mutant *B. breve* strains. In brief, the top selected mutants (20, 24 and 48) and WT *B. breve* UC20003 were cultured in De Man, Rogosa and Sharpe (MRS) medium up to saturation (De Man et al., 1960). Cultures were centrifuged and the supernatants were passed through a 0.22 μ m pore filter (Millipore) before precipitation with 10% v/v trichloroacetic acid (Sigma-Aldrich) at 4 °C for 3 h and harvesting by centrifugation. The precipitate was washed twice with ice-cold acetone and dried in a centrifugal vacuum concentrator (Vacufuge 5301, Eppendorf). Protein pellets were solubilized in ammonium bicarbonate buffer pH 8.5 (AmBic) and protein concentrations were determined using Protein assay kit (Invitrogen, UK) as per manufacturer instructions and stored for further purifications and in-vitro assays at -20°C (Vazquez-Gutierrez et al., 2017). Different molecular weight fractions of the proteome were obtained using cut off columns and preparative high-performance liquid chromatography (HPLC) and tested for NF- κ B activity on THP-1 cells as described previously.

6.3.3 Liquid chromatography–mass spectrometry

Each sample was separated by SDS-PAGE prior to in-gel tryptic digestion using a DigestPro automated digestion unit (Intavis Ltd.). The resulting peptides were fractionated using an Ultimate 3000 nano-LC system in line with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). All spectra were acquired from the mass spectrometer by Xcalibur 2.1 software (Thermo Scientific) operated in data-dependent acquisition mode. FTMS1 spectra were collected at a resolution of 120 000 over a scan range (m/z) of 350-1550, with an automatic gain control (AGC) target of 400 000 and a max injection time of 100ms. Precursors were filtered according to charge state (to include charge states 2-7), with monoisotopic peak determination set to peptide and using an intensity range from 5E3 to 1E20. Previously interrogated precursors were excluded using a dynamic window (40s +/-10ppm). The MS2 precursors were isolated with a quadrupole mass filter set to a width of 1.6m/z. ITMS2 spectra were collected with an AGC target of 5000, max injection time of 50ms and HCD collision energy of 35%.

The raw data files were processed and quantified using Proteome Discoverer software v2.1 (Thermo Scientific) and searched against the UniProt *Bifidobacterium breve* database (downloaded March 2020; 1827 sequences), the Uniprot *Bos taurus* database (downloaded June 2019; 46309 sequences), the Uniprot *Saccharomyces cerevisiae* database (downloaded January 2019; 6645 sequences) and an in-house common contaminants database using the SEQUEST algorithm. Peptide precursor mass tolerance was set at 10ppm, and MS/MS tolerance was set at 0.6Da. Search criteria included 'oxidation of methionine' (+15.995Da), 'acetylation of the protein N-terminus'(+42.011Da) and 'methionine loss plus acetylation of the protein N-terminus' (-89.03Da) as variable modifications and 'carbamidomethylation of cysteine' (+57.021Da) as a fixed modification. Searches were performed with full tryptic digestion and a maximum of 2 missed cleavages were allowed. The reverse database search option was enabled and all data was filtered to satisfy q value ≤ 0.05 .

6.3.4 Host – microbe interaction predictions

The following analysis workflow, as described in Figure 6.2, was based on pipelines initially developed in our research group (Korcsmaros et al., 2013; Sudhakar et al., 2019) and recently modified by Leila Gul.

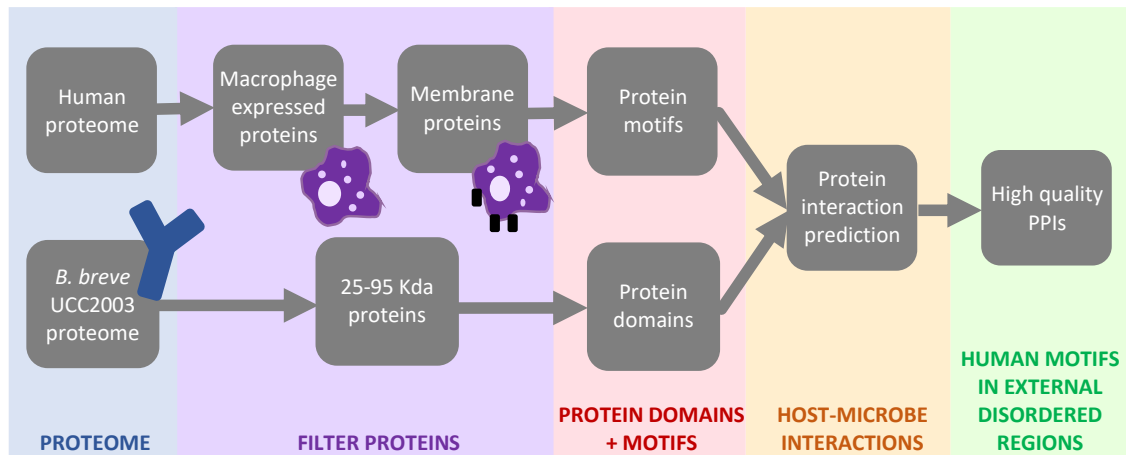


Figure 6.2. Summary of bifidobacteria-macrophage interaction prediction pipeline. Human and bacterial proteome data, including protein size, was obtained from UniProt (UniProt Consortium, 2019). Macrophage expressed proteins were determined using data from single cell experiment (Smillie et al., 2019) and THP-1 cell line expression experiment (Mullochandov et al., 2012) (ENA accession PRJNA163281). MatrixDB, ComPPI, Human protein atlas and Locate databases used to identify macrophage membrane proteins (Launay et al., 2015; Sprenger et al., 2008; Thul et al., 2017; Veres et al., 2015). InterProScan was used to identify bacterial protein motifs and ELM for human protein motifs and domain-motif interactions (Jones et al., 2014). Finally, IUPred and Phobius were used to filter human motifs in external disordered regions (Käll et al., 2007; Mészáros et al., 2018). PPIs - protein-protein interactions.

6.3.4.1 Filtering human and bacterial proteins

6.3.4.1.1 Bifidobacterial proteins

All *B. breve* UCC2003 protein sequences and their corresponding mass were downloaded from UniProt accession UP000000297 (based on genome assembly ASM22013v1, November 2019, 1826 sequences) (UniProt Consortium, 2019). R scripting was used to filter the protein sequence FASTA file for only proteins with mass 25-95 kDa – based on the results from testing different molecular weight fractions of the proteome on THP-1 cells (Methods section 6.3.2). 1218 bifidobacterial proteins were taken forward for subsequent analysis.

6.3.4.1.2 Human macrophage proteins

All human protein sequences were downloaded from UniProt accession UP000005640 (November 2019, 75004 sequences) (UniProt Consortium, 2019). Human proteins were filtered for those expressed in macrophages using two different datasets; a transcriptomics dataset from a human THP-1 monocyte cell line (Mullochandov et al., 2012) and a single cell RNA-sequencing (scRNA-seq) dataset from colon mucosal macrophages (Smillie et al., 2019). All proteins

expressed in both datasets were used to filter predicted host-microbe interactions to obtain only those involving a human protein expressed in macrophages.

THP-1 monocyte expression data was obtained from the European Nucleotide Archive project named *mRNA profiling of THP1 cell line*, with accession PRJNA163281, using REST URLs (Harrison et al., 2019; Mullokandov et al., 2012). Single end FASTA files were downloaded in triplicate from Illumina HiSeq 2000 sequencing. The data was processed to obtain gene counts data using the Snakemake pipelining tool (Köster and Rahmann, 2012). This pipeline applied Trimalore (v 0.5.0) to trim adapters from the sequences and Kallisto (v 0.44.0) to quantify gene abundance based on protein coding sequences downloaded from Gencode (GRCh38.p13) (Bray et al., 2016; Frankish et al., 2019; Krueger, 2019). Trimalore parameters were: quality=20 and length=50. Kallisto parameters were: bootstrap=100, mean fragment length=100 and standard distribution=20. Custom R scripts employing the packages *Tximport* and *ggplot2* were used to process Kallisto output into counts tables and to plot the density (Soneson et al., 2015; Wickham, 2016). Based on the gene abundance density plot, genes with expression ≥ 2 in all samples are considered expressed. In total 11,586 gene were expressed above this cut off.

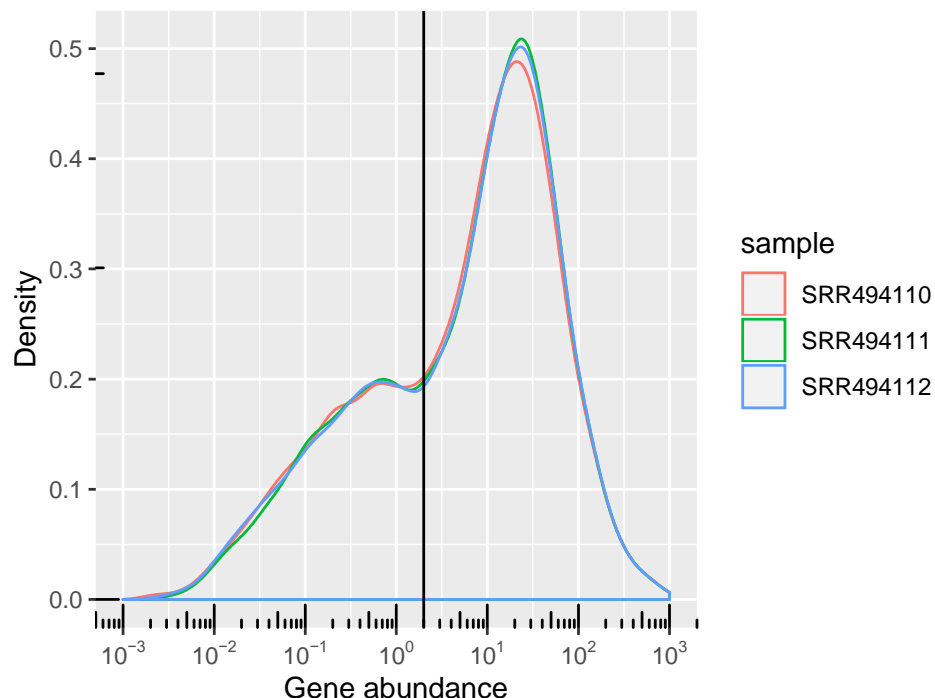


Figure 6.3. Density plot of gene abundance across samples in project PRJNA163281. Cut off for expression indicated as vertical line.

Chromium 10X single cell expression data from colon mucosal macrophages (healthy individuals) was obtained from Smillie et al. (2019) *via* the Broad Institute Single Cell Portal (accession SCP259). Average expression across all identified macrophage cells was converted to transcripts per million (TPM) from transcripts per 10,000. The data was processed and plotted using custom R scripts. To determine a gene as expressed, the cut off of $\text{TPM} \geq 2$ was chosen based on this density plot (Figure 6.4) and the THP-1 density plot (Figure 6.3). In total 10,681 genes were expressed above this cut off, of which 9,286 were also expressed in the THP-1 data and thus taken forwards for the subsequent analysis.

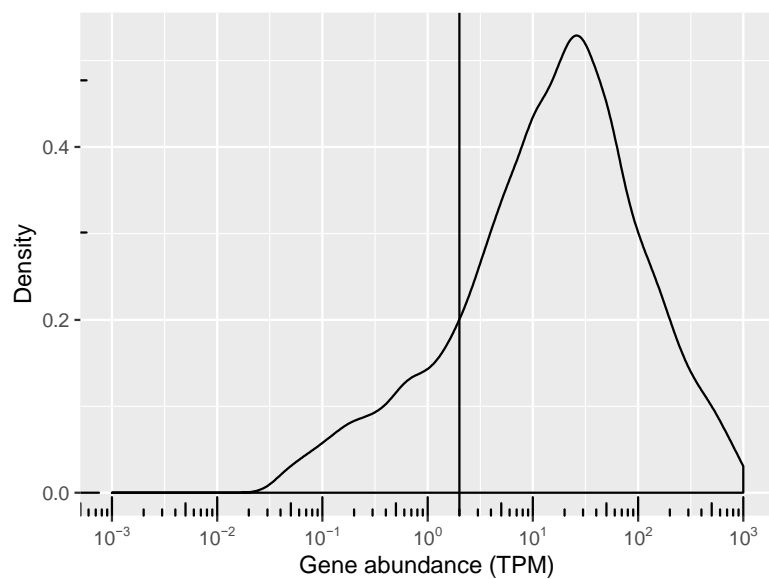


Figure 6.4. Density plot of average gene abundance across macrophage cells in Smillie et al. (2019). Cut off for expression indicated as vertical line.

6.3.4.2 Human membrane proteins

A list of predicted human membrane proteins was obtained by Leila Gul (EI). Information from four databases was used to collate this list:

- Locate v1 (Sprenger et al., 2008)
- MatrixDB v1 (Launay et al., 2015)
- ComPPI v2.1.1 (Veres et al., 2015)
- Human Protein Atlas v19.1 (Thul et al., 2017)

Any gene predicted to encode a membrane protein in any of these data resources was used for subsequent analyses. Python was used to obtain a list of proteins predicted to be membrane based and expressed in macrophages. The Uniprot ID Mapping Service was used to convert protein and gene IDs (UniProt Consortium, 2019).

6.3.4.3 *Domain and motif predictions*

Pfam domains were predicted in all 25-95 kDa bifidobacterial proteins using InterPro scan (v5) (El-Gebali et al., 2019; Jones et al., 2014). It was not necessary to predict motifs for the human proteins as the information is present in the Eukaryotic Linear Motif (ELM) database (Dinkel et al., 2016).

6.3.4.4 *Domain – motif interaction prediction*

Human motifs and domain-motif interactions were downloaded from the ELM database (v32.0) (Dinkel et al., 2016). Python and R scripts written by myself and Leila Gul were used to filter ELM domain-motif interactions to contain only the predicted bacterial domains and macrophage membrane proteins.

6.3.4.5 *Membrane protein filtering*

Predicted human-bacterial interactions were further filtered to ensure that human domains were predicted to appear in the extracellular region of the membrane protein. FASTA sequences of all human proteins in the predicted interactions were passed into the Phobius tool to predict cytoplasmic, transmembrane and non-cytoplasmic regions (Käll et al., 2004, 2007). Python was used to filter the human-bacterial interactions based on the Phobius output, so that all human domains entirely span a non-cytoplasmic region of the protein (not-including signal peptides).

6.3.4.6 *Disordered region filtering*

Leila Gul applied IUPred2A and ANCHOR2 to identify disordered residues and disordered binding regions in the human proteins, respectively (Mészáros et al., 2018). Predicted domain-motif interactions were discarded where the human protein motif was not considered disordered – where more than one amino acid in the domain had an IUPred2A score < 0.5 and an ANCHOR2 score < 0.4 .

6.3.5 **Lipoprotein prediction**

To predict which proteins are lipoprotein precursors, I used hidden Markov model and neural network tools PRED_LIPO (v1) and SignalP (v5) (Bagos et al., 2008; Nielsen, 2017). PRED_LIPO predicts secretory signal peptides in gram positive bacteria while SignalP can identify signal peptides in any archaea, bacteria or eukaryote. The results from the two tools were in agreement. Input protein sequences were obtained from UniProt (UniProt Consortium, 2019).

6.4 Results

6.4.1 *B. breve* UCC2003 activates NF- κ B in THP-1 cells

Using an NF- κ B reporter macrophage cell line (THP-1 cells), Ian O'Neill found that WT *B. breve* UCC2003 has an activatory effect on NF- κ B. Additionally, he identified 20 Tn5 insertion mutants with a significantly increased or decreased effect compared to the WT. Three top mutants were selected for further investigation which display the most increased NF- κ B activation (mutants 20 and 24) and the most decreased NF- κ B activation (mutant 48) compared to the WT.

Subsequently, the total proteome was extracted from the WT and top mutant strains by Sree Gowrinadh Javvadi. Testing different molecular weight fractions of the proteomes on THP-1 cells confirmed that the immunogenic molecule was likely a protein with molecular weight between 50 and 70 kDa which has an activatory effect on NF- κ B.

6.4.2 Bifidobacterial protein domains can interact with macrophage protein motifs

Knowing that the activation of NF- κ B in macrophages is likely driven by a secreted protein of bifidobacteria, I aimed to identify which bifidobacterial proteins can theoretically interact with macrophage surface proteins to induce downstream effects in the macrophage. Here I adapted a previously applied workflow to predict domain-motif interactions between secreted bacterial proteins and macrophage surface proteins (Sudhakar et al., 2019). As protein domains and motifs are cross-species and typified by their sequences, I could identify known domains and motifs in the bacterial and macrophage proteins (respectively) and use a database of previous predicted and/or identified eukaryotic domain-motif interactions (the ELM database) to predict interactions between them (Akiva et al., 2012; Dinkel et al., 2016; Puntervoll et al., 2003).

In total, I identified 23,218 possible interactions between 24 bifidobacterial proteins and 1,302 macrophage surface proteins. Specifically, these proteins contained 29 different motifs (human proteins) and 6 different domains (bacterial proteins). Upon further investigation of the bifidobacterial proteins I found that four of them had active domains which were predicted to be cytoplasmic in location based on InterPro annotations (Hunter et al., 2009). These were subsequently disregarded. The final list of bifidobacterial proteins potentially interacting with

human macrophage proteins *via* domain-motif interactions is given in Table 6.1. The most common type of proteins were glycosyltransferases, of which there were nine, and serine/threonine protein kinases, of which there were three. In addition, there were three dehydrogenases, two different transferases, two FHA domain-containing proteins and a phosphoesterase.

UniProt protein ID	Gene name	Protein name (symbol where known)	Mass (Da)	Pfam Domain/s
F9XYD8	Bbr_1653	D-3-phosphoglycerate dehydrogenase (SerA2)	35,261	PF00389
F9XYF7	Bbr_1674	FHA domain-containing protein	48,802	PF00498
F9XZ54	Bbr_1769	Phosphoesterase	48,563	PF00149
F9XZ72	Bbr_1788	Glycosyltransferase involved in cell wall biogenesis	39,283	PF00535
F9XZ80	Bbr_1796	Glycosyltransferase	49,959	PF00535
F9XZI1	Bbr_0068	Serine/threonine protein kinase (PknA1)	34,650	PF00069
F9XZI6	Bbr_0073	FHA domain-containing protein	25,371	PF00498
F9XZX6	Bbr_1895	Glycosyltransferase	43,621	PF00535
F9XZY9	Bbr_1908	Serine/threonine protein kinase	76,574	PF00069
F9Y002	Bbr_0084	Glycosyltransferase	36,887	PF00535
F9Y0N5	Bbr_0238	Glycosyltransferase	39,152	PF00535
F9Y1Q6	Bbr_1268	D-3-phosphoglycerate dehydrogenase (SerA1)	43,229	PF00389
F9Y1V9	Bbr_0435	Beta-1,6-N-acetylglucosaminyltransferase	34,174	PF02485
F9Y1W2	Bbr_0438	Glycosyltransferase	37,920	PF00535
F9Y1W9	Bbr_0445	Glycosyltransferase	39,673	PF00535
F9Y1X2	Bbr_0448	Glycosyltransferase	36,641	PF00535
F9Y240	Bbr_1327	dTDP-rhamnosyl transferase (RfbF)	38,142	PF00535
F9Y246	Bbr_1333	Conserved protein with hydroxyacid dehydrogenase catalytic domain	35,848	PF00389
F9Y2A1	Bbr_0504	Serine/threonine protein kinase	41,889	PF00069
F9Y2U2	Bbr_1504	Glycosyltransferase	35,135	PF00535

Table 6.1. Proteins of *B. breve* UCC2003 predicted to interact with macrophage surface proteins vis domain-motif interactions.

Because the database used to predict interactions covers only eukaryotic interactions, the six bacterial Pfam domains are all known in eukaryotic species (Pfam domain IDs):

- D-isomer specific 2-hydroxyacid dehydrogenase catalytic domain is involved in oxidation-reduction interactions (PF00389)

- Forkhead-associated domain is a phosphopeptide recognition domain (PF00498)
- Calcineurin-like phosphoesterase domain (PF00149)
- Glycosyl transferase family 2 domain which is known to transfer sugars to a range of substrates including cellulose, dolichol phosphate and teichoic acids (PF00535)
- Protein kinase domain which phosphorylates proteins (PF00069)
- Core-2/I-Branching enzyme domain which has glucuronosyltransferase activity (PF02485)

6.4.3 Predicted bifidobacterial proteins are not present in mass spectrometry results

For experimental identification of the immunogenic protein/s secreted by *B. breve* UCC2003, Sree Gowrinadh Javad carried out LC-MS on total proteomes of the WT and top mutant strains. He aimed to identify differences between NF- κ B activating and non-NF- κ B activating strains, and thus predict which proteins could be responsible for the observed effects. Comparing the results from the NF- κ B activating strains (WT, 20 and 24) to the non-NF- κ B activating strain (48), we identified 11 bifidobacterial proteins present in the activating strains but not the inactivating strains (Table 6.2).

UniProt protein ID	Gene name	Protein name (symbol where known)	Mass (Da)
F9XZH3	Bbr_0060	Alpha-1,4 glucan phosphorylase (GlgP1)	91,088
F9XY36	Bbr_0670	Glutamine synthetase (GlnA)	53,254
F9XZF9*	Bbr_0046	Conserved hypothetical secreted protein	68,877
F9Y1U3*	Bbr_0417	Solute-binding protein of ABC transporter system for sugars (GalC)	48,864
F9XZ02*	Bbr_0843	Conserved hypothetical secreted protein with exalibur domain	25,334
F9Y2P4	Bbr_1454	Conserved hypothetical membrane spanning protein with Endonuclease/Exonuclease/phosphatase family domain	41,467
F9XYM5	Bbr_0791	Trigger factor (Tig)	49,413
F9XZB8*	Bbr_1836	Sugar-binding protein of ABC transporter system	41,473
F9XZ70*	Bbr_1785	Hypothetical secreted protein	33,891
F9XYI8	Bbr_0753	SSU ribosomal protein S1P (RspA)	54,615
F9Y109	Bbr_0288	Conserved hypothetical secreted protein	67,705

Table 6.2. Potential NF- κ B activating proteins based on LC-MS. *B. breve* UCC2003 proteins identified by LC-MS in the wild-type strain and two NF- κ B activating mutants 20 and 24, but not in the non-NF- κ B activating mutant 48. * Predicted prelipoproteins.

Unfortunately, there were no proteins present in the LC-MS results (Table 6.2) which were also predicted to take part in domain-motif interactions with macrophages (Table 6.1, Figure 6.5). For a better understanding of the reason behind this disparity, I investigated the reasons why the proteins predicted by LC-MS were not predicted by the computational predictions. I found that two of the proteins (hypothetical secreted proteins, F9XZ70 and F9Y109) do not contain any known eukaryotic domains based on PFAM (Table S6.1) (El-Gebali et al., 2019). Relating to the remaining nine proteins, none of their PFAM annotated domains were involved in domain-motif interactions in the *a priori* collection which I used (from the ELM database) (Puntervoll et al., 2003). Therefore, I assume that the LC-MS predicted proteins might contain prokaryotic-specific domains of which we do not have high quality *a priori* interaction information or alternatively these proteins don't interact with macrophages through domain-motif interactions between proteins. Examples of both are present. For example, the protein GalC (F9Y1U3) contains the domain 'Bacterial extracellular solute-binding protein', which is a prokaryotic-specific domain and protein RspA (F9XYI8) has the domain 'S1 RNA binding domain', which interacts with RNAs and not proteins (Table S6.1). In conclusion, the immunogenic protein of interest most likely does not interact with macrophages *via* eukaryotic domain-motif interactions and therefore cannot be identified using the bioinformatic pipeline applied.

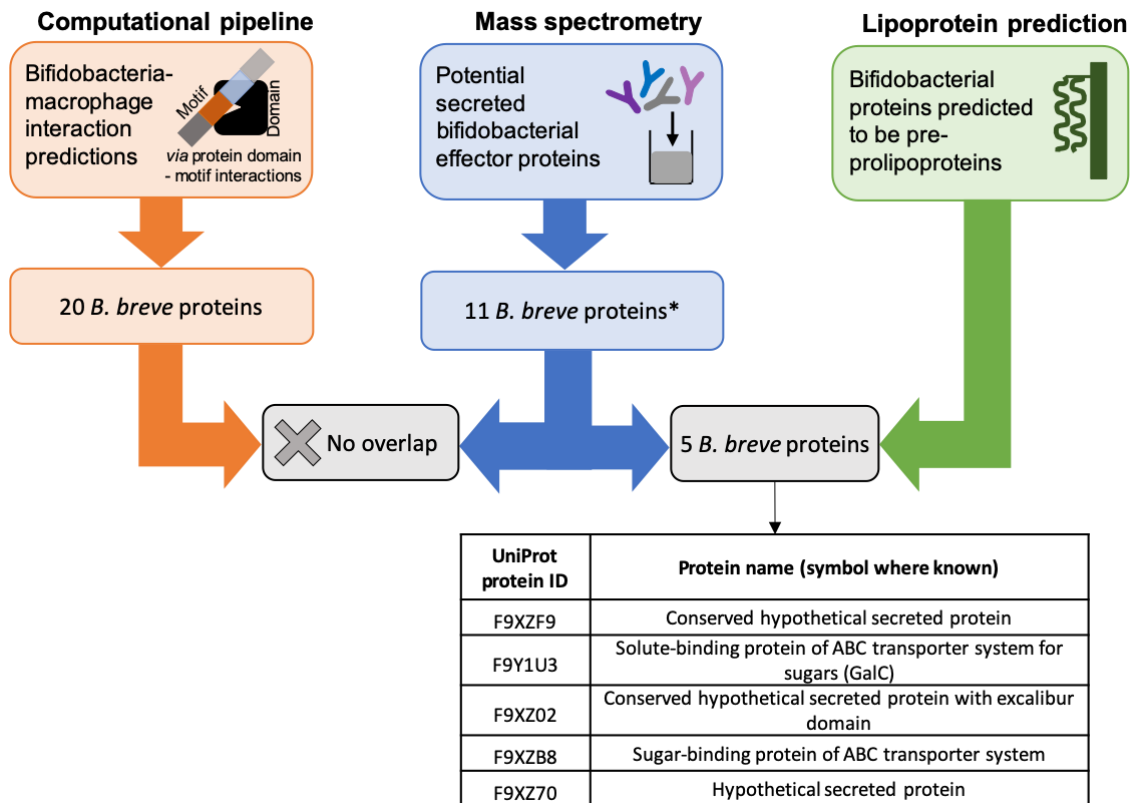


Figure 6.5. Summary of results from computational and experimental approaches. Using the domain-motif prediction pipeline I predicted 20 *B. breve* UCC2003 proteins (with mass 25-95 kDa) which could potentially interact with macrophage surface proteins. Liquid chromatography mass spectrometry (LC-MS) of secreted proteome of wild-type and mutant strains identified 11 *B. breve* UCC2003 proteins which would potentially interact with macrophages to activate NF- κ B. None of the LC-MS predicted proteins overlapped with the 20 computationally predicted proteins. Preprolipoprotein prediction tools PRED_LIPO (v1) and SignalP (v5) were applied to the 11 LC-MS proteins, identifying five preprolipoproteins (Bagos et al., 2008; Nielsen, 2017). * None of these 11 proteins contained eukaryotic domains which were involved in domain-domain interactions in the DOMINE database (Raghavachari et al. 2008).

6.4.4 NF- κ B activating molecule likely to be a lipoprotein

Whole genome sequencing results revealed that the non-NF- κ B activating mutant (48) contains its Tn5 insertion site within the lipoprotein signal peptidase gene Bbr_1299 (contains signal peptidase II domain, Pfam PF01252). This gene is also known as *ispA* (F9Y212) and is part of a multistep pathway to process lipoproteins. First preprolipoproteins are secreted through the bacterial membrane, usually *via* the general secretion (Sec) pathway (Zückert, 2014). Next preprolipoproteins are acylated by a lipoprotein diacylglycerol transferase protein (Igt), anchoring the protein to the lipid membrane. In Gram-positive bacteria such as bifidobacteria, the final step in the process involves a lipoprotein signal peptidase protein (IspA) which catalyses

the removal of signal peptides from prelipoproteins creating mature lipoproteins. In Gram-negative bacteria a further stage occurs in which a third acyl chain is attached by a lipoprotein N-acyl transferase (Zückert, 2014). Given the role of lipoprotein signal peptidases in lipoprotein processing, combined with the knowledge that lipoproteins are well recognised pathogen-associated molecular patterns, it follows that inactivation of *ispA* likely prevents maturation of lipoproteins which in turn could prevent recognition of *B. breve* by macrophages – perhaps through a structure change of the lipoprotein or through alterations in lipoprotein release from the membrane.

To further investigate this finding, I used the tools SignalP and PRED-LIPO to predict which of the 11 LC-MS identified proteins, could be secreted prelipoproteins (the protein precursors to lipoproteins) (Bagos et al., 2008; Nielsen, 2017). Here I identified five proteins which are predicted prelipoproteins containing lipoprotein signal peptides, as shown in Figure 6.5: Conserved hypothetical secreted protein (F9XZF9), GalC (F9Y1U3), Conserved hypothetical secreted protein with excalibur domain (F9XZ02), Sugar-binding protein of ABC transporter system (F9XZB8), Hypothetical secreted protein (F9XZ70). I conclude that one or multiple of these five proteins is likely the cause of NF- κ B activation in macrophages. However, it remains unknown whether the protein or the lipid part of the lipoprotein is responsible for interacting with the macrophages.

As a final check of possible protein-protein interactions, I looked at whether domain-domain computational predictions would yield possible interactions between the five LC-MS predicted prelipoproteins and the macrophage surface proteins. To do this I searched for the domains of these proteins (annotated in Table S6.1) within the domain-domain interaction database DOMINE (Raghavachari et al., 2008). None of the domains were involved in any known interactions within DOMINE. Therefore, through computational prediction tools available, I was unable to identify possible mechanisms of protein-protein interactions which any of these proteins and macrophage surface proteins. Nevertheless, it remains possible that these prelipoproteins can directly bind to macrophage proteins through bacterial-specific domains. On the other hand, based on the function of the mutated gene *IspA*, it is perhaps more likely that the lipid which is attached to the prelipoproteins is responsible for binding with macrophage proteins.

6.5 Discussion

As macrophages are the most abundant white blood cell in the healthy gut, they provide a key interface between gut microbes and the host immune system (Mowat and Agace 2014). Bifidobacteria are a major constituent of the juvenile intestinal microbiota playing a key role in development of the intestinal epithelium and the immune system. Furthermore, bifidobacteria have been shown to protect the host from gut pathogens and are often used as probiotics to improve and maintain general gut health (O'Neill et al., 2017). As such, a thorough understanding of the interactions between bifidobacteria and gut macrophages, and the effector molecules associated, is required for a better understanding of gut health in children and adults.

By combining a mutant library of *B. breve* UCC2003 with a reporter cell line, Ian O'Neill demonstrated the applicability and practicality of such a high throughput approach. In addition to providing real-time readouts of NF- κ B activation, this approach offers great potential for the identification of bacterial effector molecules. One limitation of the presented approach is the questionable applicability of THP-1 cell lines to the intestinal situation. As described in the background (section 1.3.2), intestinal macrophages have different phenotypes to standard macrophages, exhibiting hyporesponsiveness to TLR activation. THP-1 cells are only able to mimic some of the characteristics of intestinal macrophages, and are tested in isolation, preventing any possible phenotype modifications due to neighbouring cell secretions, such as TGF- β (which can block NF- κ B activation) (Naiki et al., 2005; Smith et al., 2011; Smythies et al., 2010). However, due to ease of use application of this cell line is beneficial when used as an initial screen prior to further investigatory and validatory experiments in more relevant models. Such experiments are also required to confirm the effect of bifidobacterial-initiated NF- κ B activation in THP-1 cells on other phenotypes such as cytokine release and nitric oxide production.

Using a bioinformatic approach I combined *a priori* knowledge on gene expression, protein sequence, protein subcellular location, domain and motif annotations and domain-motif interactions to predict 20 bifidobacterial secreted proteins which could theoretically bind to macrophage surface proteins. Such an approach can be very powerful when combined with experimental validations, but is heavily limited by the *a priori* knowledge available. For example,

here the predicted domain-motif interactions were based on the ELM database of eukaryotic interactions (Dinkel et al., 2016; Puntervoll et al., 2003). Therefore, any protein-protein interactions involving bacterial domains would not be included – which is highly restrictive when researching prokaryotic-based interactions. As far as I am aware, no resource yet exists for bacterial domain-eukaryotic motif interactions, likely due to vast numbers of bacterial domains which have not been fully characterised or studied. On the other hand, computational tools exist which can predict interactions based on protein sequence.

Here I predicted domain-motif interactions due to the identified importance of these kinds of interactions in signalling cascades combined with the directionality of the signal (Akiva et al., 2012; Diella et al., 2008). However, extensions to this approach could also predict domain-domain interactions using databases such as DOMINE (Raghavachari et al., 2008).

Based on mass spectrometry results, we identified 11 possible proteins of interest. However, unfortunately none of these 11 were present among the 20 computationally predicted proteins. This result could be due on limitations of the computational method. Specifically, the interaction may occur *via* prokaryotic domains for which we do not have prior information, or through eukaryotic domains where the interaction of interest is missing from the database due to not yet having been identified. In fact, none of the 11 LC-MS predicted proteins contained eukaryotic domains which had predicted domain-motif interactions in the ELM database. Additionally, the interaction may occur through bacterial proteins binding to a non-protein molecule of the macrophage. For example, one of the LC-MS predicted bacterial proteins contained an RNA binding domain. Such interactions are out of scope for this computational method. More likely, the interaction may not occur *via* bacterial protein domains. For example, the whole genome sequencing of mutants indicated a role for the lipoprotein processing gene *ispA*, implicating lipoproteins in the interaction between bifidobacteria and macrophages. Such an interaction could occur *via* the lipid moieties of the lipoprotein and is therefore also out of the scope of the computational method applied. Nevertheless, it must also be considered that the result is due to an error in the mass spectrometry or the computational predictions, although less likely. Additional experiments are required to identify and confirm the bacterial molecules responsible for macrophage activation, after which the interaction type can be more easily identified.

Whole genome sequencing of the non-NF- κ B activating mutant identified a mutation in a lipoprotein signal peptidase gene, *ispA*. In prokaryotic organisms, this transmembrane enzyme is responsible for cleaving the signal sequence of the lipoproteins following transport across the cytoplasmic membrane and addition of diacylglycerol moieties by the prelipoprotein diacylglycerol transferase (Lgt) protein (Zückert, 2014). This protein itself is not likely to be the secreted effector molecule influencing NF- κ B activation in macrophages, but it probably causes downstream changes in lipoprotein processing which affect NF- κ B activation: highlighting the complexities of using mutant strains to identify causative molecules. Possibility, the absence of correct lipoprotein processing in the non-NF- κ B mutant resulted in semi-processed lipoproteins which could not activate macrophages. Indeed, a previous experiment has shown that the *ispA* gene is required in *Mycobacterium tuberculosis* for stimulation of TLR2 reporter cells by lipoproteins (Banaiee et al., 2006). It is well known that TLR2 on the surface of macrophages (and other cells) recognises bacterial lipoproteins resulting in initiation of NF- κ B signalling (Aliprantis et al., 1999). Furthermore, previous experiments have found that *B. breve* UCC2003 can activate intestinal epithelial cell TLR2, albeit through their EPS (Hughes et al., 2017). Whilst it is well accepted that NF- κ B activation in macrophages leads to pro-inflammatory responses, research has shown that other gut commensals exploit TLR2 activation of T-cells to suppress immunity through increased IL-10 production (Round et al., 2011). It has also been found that an aggregated lipoprotein from another strain of *B. breve* can interact with dendritic cell TLR2 to increase IL-10 and prolong survival of the dendritic cells (Scuotto et al., 2014). Macrophages also secrete IL-10, suggesting that bifidobacterial activation of macrophages may have anti-inflammatory and/or pro-inflammatory effects. Further experimentation is required to evaluate this.

To further investigate the possibility that bifidobacterial lipoproteins activate macrophage NF- κ B, I used computational tools to identify lipoproteins among the 11 LC-MS predicted proteins. I found that a large proportion of the predicted proteins were lipoproteins (5/11), providing further evidence of their importance. Further investigation is required to determine whether any of these lipoproteins are the effector molecule influencing NF- κ B activation, *via* their lipid or protein parts, and whether they act *via* TLR or through another mechanism. Additionally, due to the complexity of microbial-host interactions, we must consider that multiple proteins/lipoproteins are acting in cohort. It is possible that the aberrant processing of all bifidobacterial secreted lipoproteins is required for significant reductions in NF- κ B activation.

6.6 Future research directions

Following the presented work, many questions remain regarding the interaction of *B. breve* UCC2003 and macrophages. First, it is necessary to confirm the relevance of lipoproteins by isolating the LC-MS predicted lipoproteins and testing them on THP-1 cells and/or a more relevant macrophage population. In addition, further investigation should test the involvement of TLR2 using knockout macrophages or by blocking the receptors (Flores, 2018). Examining pro- and anti-inflammatory cytokine release as a result of bifidobacterial activation will add context to the findings, highlighting the relevance of this interaction to gut health. Furthermore, it would be useful to determine whether the lipid or protein parts of the lipoprotein are interacting with the macrophages. Unfortunately, lipid-protein interactions are inherently difficult to predict computationally, especially without prior knowledge of the exact lipid structure of these lipoproteins (Corradi et al., 2019).

In this study we predicted the effector molecules in WT *B. breve* UCC2003 missing from the non-NF- κ B activating strain. However, we did not identify possible reasons why the strains 20 and 24 increased NF- κ B activation compared to the WT strain. Such investigation will uncover more details on how bifidobacteria interact with the host immune system. On the other hand, the predicted domain-motif interactions between *B. breve* and macrophages provide further avenues of research.

Together our experiments have evidenced the value of high throughput screens and highlighted the potential benefits and drawbacks of a combined experimental and computational approach to research. Following subsequent validation of the effect of bifidobacterial lipoproteins on macrophages, this work will contribute to a greater understanding of the role of bifidobacteria in immune cross talks with their host.

Chapter 7: Integrated discussion

The intestinal epithelial cell (IEC) barrier represents a key interface between host immune cells and commensal microbes. Understanding communication between these compartments and its role in gut homeostasis will aid the study of intestinal diseases and the development of methods to prevent and treat them. In this thesis I aimed to develop workflows to study intra- and inter-cellular regulation in a cell type-specific manner, and to apply these to investigate interplay between commensal bacteria *Bifidobacterium*, immune cells and the intestinal epithelium.

Inflammatory bowel disease (IBD) is one particular context in which this interplay is fundamental. While the exact causes of IBD have not been identified, it is understood that IBD occurs in genetically susceptible individuals where an environmental trigger results in a defective mucosal immune response to the gut microbiota (Yue et al., 2019). Dysregulation of the immune response results in epithelial damage, infiltration of commensal flora into the lamina propria and a generalised inflammatory response resulting in high levels of pro-inflammatory cytokines (Guan, 2019). In Chapter 3, I showed that colonic IECs recognise and respond to these cytokines by activating inflammatory responses *via* signalling pathways with no apparent negative feedback loop. Anti-cytokine therapies represent one of the primary treatment options for IBD patients, however response rates are poor (Roda et al., 2016). By combining *in vivo* organoid models and network analysis approaches I identified similarities in the IEC response to different cytokines, whereby intracellular signals converge on a few key transcription factors (TFs). Notably, the colonic expression levels in IBD patients of one transcription factor, Protein C-ets-1 (ETS1), was significantly associated with subsequent non-response to anti-TNF α drug infliximab. These findings shed light on the causes of non-response to anti-cytokine treatments, present a novel classification system to predict patient response to therapy and offer a potential new candidate for therapeutic targeting. Furthermore, through this work I have highlighted the benefits of studying the detailed molecular response of cells to separate components within a complex system.

Development and maintenance of gut homeostasis relies on a complex interplay between the gut microbiota, the gut lining and the immune system. Therefore, in addition to studying the communication between host cellular compartments, one must also consider the impact of commensal bacteria on host cells. Bifidobacteria are health-promoting bacterial genus which

are widely used as probiotic supplements - despite many unanswered questions relating to their mechanisms of action (Sarkar and Mandal, 2016). In Chapter 4, I showed that gut commensal *Bifidobacterium breve* UCC2003 has a significant effect on the transcriptional program of IECs of two-week-old mice but not adult mice. Specifically, transcriptional changes related to cell differentiation, cell proliferation and epithelial barrier function, particularly within the stem cell compartment, thus providing evidence that bifidobacteria plays a role in development of the gut lining in early life. Whilst the benefits of administration of probiotic supplements to healthy infants is contentious, perhaps due to natural levels of bifidobacteria colonisation, these findings uncover greater mechanistic explanations for the beneficial effects of supplementation in preterm infants (Braga et al., 2011; Quin et al., 2018; Stratiki et al., 2007; Szajewska et al., 2010). Furthermore, these results indicated a change in epigenetic modification due to bifidobacteria, which provides an interesting avenue for further research. Although methylation of enterocyte cell lines by *Bifidobacterium*-containing probiotics has previously been noted (Cortese et al., 2016), little research has followed up this finding, particularly in relation to the stem cell compartment in which epigenetic changes could affect the whole epithelial layer. Defining epigenetic changes may shed light on the long-term effects of *Bifidobacterium* exposure in early life. The experimental methods we established to isolate stem cells from bifidobacteria-treated mice in Chapter 5 would be well suited to studying epigenetic changes at a cell type-specific level. Application of cell-type proportion inference algorithms such as PLIER, may contribute to the analysis of affected cell types in this dataset (Mao et al., 2019). Unfortunately, in Chapter 5 we did not observe any significant effect of bifidobacteria on sorted IECs. Improvements to this experiment would have seen more similarities between the experimental set up of Chapter 4 and Chapter 5, especially relating to the age of the mice, as well as more thorough evaluation of fluorescence activation cell sorting protocols prior to carrying out the primary experiment.

Unexpectedly, in Chapter 4 we observed very little change in the innate immune functions of two-week-old mouse IECs due to bifidobacteria exposure, indicating a lack of generalised immune response to bifidobacteria in the gut lumen. On the other hand, in Chapter 6 we showed that *B. breve* UCC2003, likely via secreted lipoproteins, activates nuclear factor (NF)- κ B in macrophage-like cells. While the downstream effects of NF- κ B activation have not yet been confirmed in this setting, this finding is likely to be a pro-inflammatory response to aberrant bifidobacteria identified in the blood or gut lamina propria, indicating a context-specific host response. However, it is also possible that activation of macrophages by *B. breve* results in an

concurrent anti-inflammatory response, as has been shown by Scutto et al. (2014) to occur in dendritic cells upon exposure to a *B. breve* C50 secreted lipoprotein. Further experimental work will be necessary to shed light on the downstream effects of NF- κ B activation and determine the relevance of this model to the gut compartment. However, through this experiment we evidenced the value of high throughput screens to aid identification of the effector molecules responsible for the effect of bifidobacteria. Such detailed mechanistic studies are required to develop a holistic understanding of how bifidobacteria affects its hosts cells, which will lead to development of safer and more effective next-stage targeted microbiota therapies for humans. Nevertheless, dedicated clinical studies would be required to determine if our findings relating to the impact of bifidobacteria on IECs and macrophages extrapolate to the human setting.

Moreover, the work carried out in Chapter 4 evidences the cell type-specificity of IEC response to stimulus, by predicting that bifidobacteria particularly affects stem cells in the neonatal mice epithelium. Whilst further investigation is required to validate this finding, this study and others presented in this thesis, strongly indicate that mechanistic research should focus on specific cell types rather than treating the epithelium as a single homogenous unit. Indeed, there is ample evidence that dysregulation of specific IEC types, such as Paneth cells and goblet cells, has a key role in the pathogenesis of IBD (Liu et al., 2016; Okamoto and Watanabe, 2016; Zheng et al., 2011). Therefore, additional future work investigating the differences between cytokine-driven responses between types of IECs, could uncover further details about epithelial function in IBD. Interestingly, I predicted in Chapter 2 that ETS1 is a key regulator of Paneth cells and goblet cells, targeting a large proportion of the known cell type markers. A possible connection between these findings can be explored in future studies, with the potential to lead to more targeted therapeutic approaches. Of note, the work presented in the 2nd chapter and the 5th chapter highlight the continuous nature of cellular differentiation in the epithelium and the inherent challenges assigning a singular identify to a cell. This should be considered when discussing and studying IEC cell types. Throughout the thesis I have evidenced different ways in which specific cell types can be studied (cell type enriched organoids, *in vivo* cell lines, fluorescence activated cell sorting), while rapid developments in single cell technologies also make these approaches increasingly applicable and affordable (Tang et al., 2019).

In addition to the biological findings of this thesis outlined above, through these projects I have demonstrated many different ways in which networks can be used to extract biological insights

from transcriptomics datasets. In Chapter 2, I reconstructed regulatory networks combining a variety of different regulatory interaction types, for the purpose of predicting and comparing key molecular regulators of Paneth cells and goblet cells. Possible improvements to this approach would include computational prediction of lncRNA and miRNA interactions, preventing loss of data during the network reconstruction stage, and improved scoring of TF activity. In Chapter 3, I extended regulatory interactions to reconstruct causal networks linking the recognition of a cytokine by cellular receptors to a transcriptional change in the cell. Here, I used the predicted upstream signalling pathways to contextualise the predicted regulatory interactions, maintaining the primary focus on the TFs. Possible improvements to this approach would be to employ a heat diffusion method to connect cytokine receptors to TFs, lessening biases due to hub nodes and avoiding the assumption of short, linear signalling pathways (Paull et al., 2013). Moreover, the use of a TF activation prediction tool such as VIPER would improve TF predictions, avoiding the need to filter based on transcription level changes of the TF (Alvarez et al., 2016). Further additions could include the use of the PROGENy tool to predict signalling pathway activations based on transcriptional footprints, and the incorporation of proteomics data to improve signalling pathway reconstruction (Schubert et al., 2018). In Chapter 4 and 5, two further kinds of networks were employed: a protein-protein interaction network was used to add first neighbours to a list of proteins and weighted gene co-expression networks (WGCNA) were used to link genes together which have correlated expression values (Langfelder and Horvath, 2008; Módos et al., 2017). Finally, in Chapter 6, I used inter-cellular protein-protein interaction networks to study possible communication between bifidobacteria and macrophages. While this analysis did not successfully identify the effector molecule of interest, it demonstrated an additional network method which can be employed to study cellular regulation. Unfortunately, no methods exist to reliably predict molecular interactions with lipids, which was ultimately a shortcoming of this approach when applied to bifidobacteria-macrophage interactions. On the other hand, when considering communication between two eukaryotic cells this method can be extended to map ligand-receptor interactions between cells using *a priori* knowledge contained in databases such as OmniPath2 (Türei et al., 2016; Turei et al., 2020). Such a method could be used to study other inter-cellular communications between macrophages and IECs, in addition to those through secreted cytokines. Ultimately a number of different network reconstruction and analysis methods can be used to aid biological interpretation of 'omics data. The ideal methods to use depend not only on the biological question but also the dataset/s available and the accessible experimental validation options.

Apart from WGCNA, all network approaches used in this thesis are based on contextualisation of *a priori* knowledge. Network inference approaches form a large collection of additional methods in which biological networks can be reconstructed and analysed, without reliance on *a priori* interactions. These methods use mathematical models and algorithms, such as Bayesian network and ordinary differential equations, to reconstruct networks from high-throughput data (Chai et al., 2014; De Smet and Marchal, 2010). Application of these methods can be favourable where *a priori* knowledge is lacking but high-quality experimental data is plentiful.

Finally, through work covered in this thesis I have developed workflows and pipelines for contextualisation and analysis of transcriptomics data which have and will continue to be used within my research group for future projects. Moreover, following the publication of the research contained in Chapters 2, 3 and 4, I hope that other research groups will recognise the value of integrating experimental approaches with network analysis and proceed to employ such approaches in their own work. Specifically, the research presented here highlights the benefits of mechanistic studies of cellular regulation while presenting accessible methods, which can be applied to many different contexts in many different fields of research.

In conclusion, this PhD research has contributed to the mechanistic understanding of interplay between the gut commensal bifidobacteria, the intestinal epithelium and the immune system, including in the context of IBD. Meanwhile, the presented projects have promoted and aided the use of networks for interpreting transcriptomics data and studying cellular regulation. The findings outlined in this thesis will pave the way for future in-depth and validity research. This should ultimately lead to a better understanding of gut homeostasis and drive development of targeted approaches for prevention and treatment of gut dysbiosis related disorders such as IBD.

References

- Aberle, M.R., Burkhart, R.A., Tiriach, H., Olde Damink, S.W.M., Dejong, C.H.C., Tuveson, D.A., and van Dam, R.M. (2018). Patient-derived organoid models help define personalized management of gastrointestinal cancer. *Br. J. Surg.* *105*, e48–e60.
- Abreu, M.T., Kantorovich, V., Vasiliauskas, E.A., Gruntmanis, U., Matuk, R., Daigle, K., Chen, S., Zehnder, D., Lin, Y.C., Yang, H., et al. (2004). Measurement of vitamin D levels in inflammatory bowel disease patients reveals a subset of Crohn's disease patients with elevated 1,25-dihydroxyvitamin D and low bone mineral density. *Gut* *53*, 1129–1136.
- Adegbola, S.O., Sahnan, K., Warusavitarne, J., Hart, A., and Tozer, P. (2018). Anti-TNF Therapy in Crohn's Disease. *Int. J. Mol. Sci.* *19*.
- Adolph, T.E., Tomczak, M.F., Niederreiter, L., Ko, H.-J., Böck, J., Martinez-Naves, E., Glickman, J.N., Tschurtschenthaler, M., Hartwig, J., Hosomi, S., et al. (2013). Paneth cells as a site of origin for intestinal inflammation. *Nature* *503*, 272–276.
- Aghdaei, H.A., Kadijani, A.A., Sorrentino, D., Mirzaei, A., Shahrokh, S., Balaii, H., Geraci, M., and Zali, M.R. (2018). An increased Bax/Bcl-2 ratio in circulating inflammatory cells predicts primary response to infliximab in inflammatory bowel disease patients. *United European Gastroenterol. J.* *6*, 1074–1081.
- Akay, H.K., Bahar Tokman, H., Hatipoglu, N., Hatipoglu, H., Siraneci, R., Demirci, M., Borsa, B.A., Yuksel, P., Karakullukcu, A., Kangaba, A.A., et al. (2014). The relationship between bifidobacteria and allergic asthma and/or allergic dermatitis: a prospective study of 0-3 years-old children in Turkey. *Anaerobe* *28*, 98–103.
- Akiva, E., Friedlander, G., Itzhaki, Z., and Margalit, H. (2012). A dynamic view of domain-motif interactions. *PLoS Comput. Biol.* *8*, e1002341.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Helper T Cells and Lymphocyte Activation*.
- Al-Ghadban, S., Kaissi, S., Homaidan, F.R., Naim, H.Y., and El-Sabban, M.E. (2016). Cross-talk between intestinal epithelial cells and immune cells in inflammatory bowel disease. *Sci. Rep.* *6*, 29783.
- Aliprantis, A.O., Yang, R.B., Mark, M.R., Suggett, S., Devaux, B., Radolf, J.D., Klimpel, G.R., Godowski, P., and Zychlinsky, A. (1999). Cell activation and apoptosis by bacterial lipoproteins through toll-like receptor-2. *Science* *285*, 736–739.
- Allaire, J.M., Crowley, S.M., Law, H.T., Chang, S.-Y., Ko, H.-J., and Vallance, B.A. (2018). The intestinal epithelium: central coordinator of mucosal immunity. *Trends Immunol.* *39*, 677–696.
- Almeqdadi, M., Mana, M.D., Roper, J., and Yilmaz, Ö.H. (2019). Gut organoids: mini-tissues in culture to study intestinal physiology and disease. *Am. J. Physiol. Cell Physiol.* *317*, C405–C419.
- Altay, G., Larrañaga, E., Tosi, S., Barriga, F.M., Batlle, E., Fernández-Majada, V., and Martínez, E. (2019). Self-organized intestinal epithelial monolayers in crypt and villus-like domains show effective barrier function. *Sci. Rep.* *9*, 10140.
- Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* *48*, 838–847.
- Amann, J.M., Chyla, B.J.I., Ellis, T.C., Martinez, A., Moore, A.C., Franklin, J.L., McGhee, L., Meyers,

- S., Ohm, J.E., Luce, K.S., et al. (2005). Mtgr1 is a transcriptional corepressor that is required for maintenance of the secretory cell lineage in the small intestine. *Mol. Cell. Biol.* *25*, 9576–9585.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq — a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.
- Andrews, S. (2010). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.
- Arboleya, S., Watkins, C., Stanton, C., and Ross, R.P. (2016). Gut bifidobacteria populations in human health and aging. *Front. Microbiol.* *7*, 1204.
- Ardekani, A.M., and Naeini, M.M. (2010). The role of micrnas in human diseases. *Avicenna J Med Biotechnol* *2*, 161–179.
- Arijs, I., De Hertogh, G., Lemaire, K., Quintens, R., Van Lommel, L., Van Steen, K., Leemans, P., Cleyne, I., Van Assche, G., Vermeire, S., et al. (2009). Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. *PLoS One* *4*, e7984.
- Arnauts, K., Verstockt, B., Vancamelbeke, M., Vermeire, S., Verfaillie, C., and Ferrante, M. (2019). OP11 Organoids derived from inflamed intestinal biopsies of patients with ulcerative colitis lose their inflammatory phenotype during *ex vivo* culture. *Journal of Crohn's and Colitis* *13*, S007–S007.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.* *25*, 25–29.
- Atreya, I., Atreya, R., and Neurath, M.F. (2008). NF-kappaB in inflammatory bowel disease. *J. Intern. Med.* *263*, 591–596.
- Atreya, R., Zimmer, M., Bartsch, B., Waldner, M.J., Atreya, I., Neumann, H., Hildner, K., Hoffman, A., Kiesslich, R., Rink, A.D., et al. (2011). Antibodies against tumor necrosis factor (TNF) induce T-cell apoptosis in patients with inflammatory bowel diseases via TNF receptor 2 and intestinal CD14⁺ macrophages. *Gastroenterology* *141*, 2026–2038.
- Bader, G.D., and Hogue, C.W.V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* *4*, 2.
- Bagos, P.G., Tsigos, K.D., Liakopoulos, T.D., and Hamodrakas, S.J. (2008). Prediction of lipoprotein signal peptides in Gram-positive bacteria with a Hidden Markov Model. *J. Proteome Res.* *7*, 5082–5093.
- Bagowski, C.P., Bruins, W., and Te Velthuis, A.J.W. (2010). The nature of protein domain evolution: shaping the interaction network. *Curr. Genomics* *11*, 368–376.
- Bain, C.C., and Mowat, A.M. (2014). Macrophages in intestinal homeostasis and inflammation. *Immunol. Rev.* *260*, 102–117.
- Bain, C.C., Scott, C.L., Uronen-Hansson, H., Gudjonsson, S., Jansson, O., Grip, O., Williams, M., Malissen, B., Agace, W.W., and Mowat, A.M. (2013). Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6Chi monocyte precursors. *Mucosal Immunol.* *6*, 498–510.
- Bakke, D., Lu, R., Agrawal, A., Zhang, Y., and Sun, J. (2018). 18 myeloid vitamin d receptor signaling regulates paneth cell function and intestinal homeostasis. *Gastroenterology* *154*, S41.
- Balfe, A., Lennon, G., Lavelle, A., Docherty, N.G., Coffey, J.C., Sheahan, K., Winter, D.C., and

- O'Connell, P.R. (2018). Isolation and gene expression profiling of intestinal epithelial cells: crypt isolation by calcium chelation from in vivo samples. *Clin Exp Gastroenterol* 11, 29–37.
- Banaiee, N., Kincaid, E.Z., Buchwald, U., Jacobs, W.R., and Ernst, J.D. (2006). Potent inhibition of macrophage responses to IFN-gamma by live virulent *Mycobacterium tuberculosis* is independent of mature mycobacterial lipoproteins but dependent on TLR2. *J. Immunol.* 176, 3019–3027.
- Banerjee, A., McKinley, E.T., von Moltke, J., Coffey, R.J., and Lau, K.S. (2018). Interpreting heterogeneity in intestinal tuft cell structure and function. *J. Clin. Invest.*
- Bankaitis, E.D., Ha, A., Kuo, C.J., and Magness, S.T. (2018). Reserve stem cells in intestinal homeostasis and injury. *Gastroenterology* 155, 1348–1361.
- Barabási, A.-L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113.
- Bar-Ephraim, Y.E., Kretzschmar, K., and Clevers, H. (2019). Organoids in immunological research. *Nat. Rev. Immunol.*
- Bar Hoover, M.A., Hall, J.M., Greenlee, W.F., and Thomas, R.S. (2010). Aryl hydrocarbon receptor regulates cell cycle progression in human breast cancer cells via a functional interaction with cyclin-dependent kinase 4. *Mol. Pharmacol.* 77, 195–201.
- Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., et al. (2007). Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* 449, 1003–1007.
- Barker, N., van de Wetering, M., and Clevers, H. (2008). The intestinal stem cell. *Genes Dev.* 22, 1856–1864.
- Barker, N., Huch, M., Kujala, P., van de Wetering, M., Snippert, H.J., van Es, J.H., Sato, T., Stange, D.E., Begthel, H., van den Born, M., et al. (2010). *Lgr5*(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro. *Cell Stem Cell* 6, 25–36.
- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41, D991-5.
- Bartel, D.P. (2018). Metazoan MicroRNAs. *Cell* 173, 20–51.
- Bartfeld, S. (2016). Modeling infectious diseases and host-microbe interactions in gastrointestinal organoids. *Dev. Biol.* 420, 262–270.
- Bastide, P., Darido, C., Pannequin, J., Kist, R., Robine, S., Marty-Double, C., Bibeau, F., Scherer, G., Joubert, D., Hollande, F., et al. (2007). *Sox9* regulates cell proliferation and is required for Paneth cell differentiation in the intestinal epithelium. *J. Cell Biol.* 178, 635–648.
- Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.*
- Becker, S., Oelschlaeger, T.A., Wullaert, A., Vlantis, K., Pasparakis, M., Wehkamp, J., Stange, E.F., and Gersemann, M. (2013). Bacteria regulate intestinal epithelial cell differentiation factors both in vitro and in vivo. *PLoS One* 8, e55620.
- Bel, S., Pendse, M., Wang, Y., Li, Y., Ruhn, K.A., Hassell, B., Leal, T., Winter, S.E., Xavier, R.J., and Hooper, L.V. (2017). Paneth cells secrete lysozyme via secretory autophagy during bacterial infection of the intestine. *Science* 357, 1047–1052.

- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* *125*, 279–284.
- Berkers, G., van Mourik, P., Vonk, A.M., Kruisselbrink, E., Dekkers, J.F., de Winter-de Groot, K.M., Arets, H.G.M., Marck-van der Wilt, R.E.P., Dijkema, J.S., Vanderschuren, M.M., et al. (2019). Rectal organoids enable personalized treatment of cystic fibrosis. *Cell Rep.* *26*, 1701–1708.e3.
- Bettoun, D.J., Burris, T.P., Houck, K.A., Buck, D.W., Stayrook, K.R., Khalifa, B., Lu, J., Chin, W.W., and Nagpal, S. (2003). Retinoid X receptor is a nonsilent major contributor to vitamin D receptor-mediated transcriptional activation. *Mol. Endocrinol.* *17*, 2320–2328.
- Beumer, J., Artegiani, B., Post, Y., Reimann, F., Gribble, F., Nguyen, T.N., Zeng, H., Van den Born, M., Van Es, J.H., and Clevers, H. (2018). Enteroendocrine cells switch hormone expression along the crypt-to-villus BMP signalling gradient. *Nat. Cell Biol.* *20*, 909–916.
- Bevins, C.L., and Salzman, N.H. (2011). Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat. Rev. Microbiol.* *9*, 356–368.
- Bezirtzoglou, E., Tsiotsias, A., and Welling, G.W. (2011). Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe* *17*, 478–482.
- Bhat, S.A., Ahmad, S.M., Mumtaz, P.T., Malik, A.A., Dar, M.A., Urwat, U., Shah, R.A., and Ganai, N.A. (2016). Long non-coding RNAs: Mechanism of action and functional utility. *Noncoding RNA Res* *1*, 43–50.
- Biancheri, P., Di Sabatino, A., Ammoscato, F., Facciotti, F., Caprioli, F., Curciarello, R., Hoque, S.S., Ghanbari, A., Joe-Njoku, I., Giuffrida, P., et al. (2014). Absence of a role for interleukin-13 in inflammatory bowel disease. *Eur. J. Immunol.* *44*, 370–385.
- Biton, M., Haber, A.L., Rogel, N., Burgin, G., Beyaz, S., Schnell, A., Ashenberg, O., Su, C.-W., Smillie, C., Shekhar, K., et al. (2018). T helper cell cytokines modulate intestinal stem cell renewal and differentiation. *Cell* *175*, 1307–1320.e22.
- Bjerknes, M., and Cheng, H. (2006). Intestinal epithelial stem cells and progenitors. *Meth. Enzymol.* *419*, 337–383.
- Bland, J.M., and Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. *BMJ* *310*, 170.
- Boesten, R.J., Schuren, F.H.J., Willemsen, L.E.M., Vriesema, A., Knol, J., and De Vos, W.M. (2011). *Bifidobacterium breve* - HT-29 cell line interaction: modulation of TNF- α induced gene expression. *Benef. Microbes* *2*, 115–128.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* *30*, 2114–2120.
- Bottacini, F., O’Connell Motherway, M., Kuczynski, J., O’Connell, K.J., Serafini, F., Duranti, S., Milani, C., Turroni, F., Lugli, G.A., Zomer, A., et al. (2014). Comparative genomics of the *Bifidobacterium breve* taxon. *BMC Genomics* *15*, 170.
- Bovolenta, L.A., Acencio, M.L., and Lemke, N. (2012). HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC Genomics* *13*, 405.
- Braga, T.D., da Silva, G.A.P., de Lira, P.I.C., and de Carvalho Lima, M. (2011). Efficacy of *Bifidobacterium breve* and *Lactobacillus casei* oral supplementation on necrotizing enterocolitis in very-low-birth-weight preterm infants: a double-blind, randomized, controlled trial. *Am. J. Clin. Nutr.* *93*, 81–86.
- Brandtzaeg, P. (2017). Role of the intestinal immune system in health. In *Crohn’s Disease and*

- Ulcerative Colitis, D.C. Baumgart, ed. (Cham: Springer International Publishing), pp. 23–56.
- Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* *34*, 525–527.
- Brazovskaja, A., Treutlein, B., and Camp, J.G. (2019). High-throughput single-cell transcriptomics on organoids. *Curr. Opin. Biotechnol.* *55*, 167–171.
- Brito, A.F., and Pinney, J.W. (2017). Protein-Protein Interactions in Virus-Host Systems. *Front. Microbiol.* *8*, 1557.
- Brown, J.R., Nigh, E., Lee, R.J., Ye, H., Thompson, M.A., Saudou, F., Pestell, R.G., and Greenberg, M.E. (1998). Fos family members induce cell cycle entry by activating cyclin D1. *Mol. Cell. Biol.* *18*, 5609–5619.
- Brückner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* *10*, 2763–2788.
- Buchanan, M., Caldarelli, G., De Los Rios, P., Rao, F., and Vendruscolo, M. (2010). *Networks in Cell Biology* (Cambridge: Cambridge University Press).
- Buczacki, S.J.A., Zecchini, H.I., Nicholson, A.M., Russell, R., Vermeulen, L., Kemp, R., and Winton, D.J. (2013). Intestinal label-retaining cells are secretory precursors expressing Lgr5. *Nature* *495*, 65–69.
- Burisch, J., Jess, T., Martinato, M., Lakatos, P.L., and ECCO -EpiCom (2013). The burden of inflammatory bowel disease in Europe. *J. Crohns Colitis* *7*, 322–337.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* *36*, 411–420.
- Caballero-Franco, C., Keller, K., De Simone, C., and Chadee, K. (2007). The VSL#3 probiotic formula induces mucin gene expression and secretion in colonic epithelial cells. *Am. J. Physiol. Gastrointest. Liver Physiol.* *292*, G315–22.
- Cader, M.Z., and Kaser, A. (2013). Recent advances in inflammatory bowel disease: mucosal immune cells in intestinal inflammation. *Gut* *62*, 1653–1664.
- Cadwell, K., Liu, J.Y., Brown, S.L., Miyoshi, H., Loh, J., Lennerz, J.K., Kishi, C., Kc, W., Carrero, J.A., Hunt, S., et al. (2008). A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. *Nature* *456*, 259–263.
- Cadwell, K., Patel, K.K., Komatsu, M., Virgin, H.W., and Stappenbeck, T.S. (2009). A common role for Atg16L1, Atg5 and Atg7 in small intestinal Paneth cells and Crohn disease. *Autophagy* *5*, 250–252.
- Campbell, N.A., and Reece, J.B. (2008). *Biology (biology)* (Benjamin Cummings).
- Cao, B., Zhou, X., Ma, J., Zhou, W., Yang, W., Fan, D., and Hong, L. (2017). Role of mirnas in inflammatory bowel disease. *Dig. Dis. Sci.* *62*, 1426–1438.
- Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* *7*, 335–336.
- Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). *stan*: A probabilistic programming language. *J Stat Softw* *76*, 1–32.

- Cash, H.L., Whitham, C.V., Behrendt, C.L., and Hooper, L.V. (2006). Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* *313*, 1126–1130.
- Castresana-Aguirre, M., and Sonnhammer, E.L.L. (2020). Pathway-specific model estimation for improved pathway annotation by network crosstalk. *Sci. Rep.* *10*, 13585.
- Castro, D.M., de Veaux, N.R., Miraldi, E.R., and Bonneau, R. (2019). Multi-study inference of regulatory networks for more accurate models of gene regulation. *PLoS Comput. Biol.* *15*, e1006591.
- Castro-Bravo, N., Margolles, A., Wells, J.M., and Ruas-Madiedo, P. (2019). Exopolysaccharides synthesized by *Bifidobacterium animalis* subsp. *lactis* interact with TLR4 in intestinal epithelial cells. *Anaerobe* *56*, 98–101.
- Chai, L.E., Loh, S.K., Low, S.T., Mohamad, M.S., Deris, S., and Zakaria, Z. (2014). A review on the computational approaches for gene regulatory network construction. *Comput Biol Med* *48*, 55–65.
- Chan, T.E., Stumpf, M.P.H., and Babbie, A.C. (2017). Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. *Cell Syst.* *5*, 251–267.e3.
- Chandra, V., Bhagyaraj, E., Nanduri, R., Ahuja, N., and Gupta, P. (2015). NR1D1 ameliorates *Mycobacterium tuberculosis* clearance through regulation of autophagy. *Autophagy* *11*, 1987–1997.
- Chapman, C.G., and Pekow, J. (2015). The emerging role of miRNAs in inflammatory bowel disease: a review. *Therap. Adv. Gastroenterol.* *8*, 4–22.
- Chen, G., Ning, B., and Shi, T. (2019). Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front. Genet.* *10*, 317.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* *34*, i884–i890.
- Cheng, H., and Leblond, C.P. (1974). Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. *Am J. Anat.* *141*, 537–561.
- Chernyavsky, A.I., Arredondo, J., Karlsson, E., Wessler, I., and Grando, S.A. (2005). The Ras/Raf-1/MEK1/ERK signaling pathway coupled to integrin expression mediates cholinergic regulation of keratinocyte directional migration. *J. Biol. Chem.* *280*, 39220–39228.
- Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* *460*, 479–486.
- Chiffolleau, E. (2018). C-Type Lectin-Like Receptors As Emerging Orchestrators of Sterile Inflammation Represent Potential Therapeutic Targets. *Front. Immunol.* *9*, 227.
- Cholapranee, A., Hazlewood, G.S., Kaplan, G.G., Peyrin-Biroulet, L., and Ananthkrishnan, A.N. (2017). Systematic review with meta-analysis: comparative efficacy of biologics for induction and maintenance of mucosal healing in Crohn's disease and ulcerative colitis controlled trials. *Aliment. Pharmacol. Ther.* *45*, 1291–1302.
- Chopra, D.P., Dombkowski, A.A., Stemmer, P.M., and Parker, G.C. (2010). Intestinal epithelial cells in vitro. *Stem Cells Dev.* *19*, 131–142.
- Christiaen, S.E.A., O'Connell Motherway, M., Bottacini, F., Lanigan, N., Casey, P.G., Huys, G., Nelis, H.J., van Sinderen, D., and Coenye, T. (2014). Autoinducer-2 plays a crucial role in gut colonization and probiotic functionality of *Bifidobacterium breve* UCC2003. *PLoS One* *9*, e98111.

- Clevers, H. (2013). The intestinal crypt, a prototype stem cell compartment. *Cell* *154*, 274–284.
- Clevers, H.C., and Bevins, C.L. (2013). Paneth cells: maestros of the small intestinal crypts. *Annu. Rev. Physiol.* *75*, 289–311.
- Co, J.Y., Margalef-Català, M., Li, X., Mah, A.T., Kuo, C.J., Monack, D.M., and Amieva, M.R. (2019). Controlling Epithelial Polarity: A Human Enteroid Model for Host-Pathogen Interactions. *Cell Rep.* *26*, 2509–2520.e4.
- Cohen, L.J., Esterhazy, D., Kim, S.-H., Lemetre, C., Aguilar, R.R., Gordon, E.A., Pickard, A.J., Cross, J.R., Emiliano, A.B., Han, S.M., et al. (2017). Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* *549*, 48–53.
- Coleman, O.I., and Haller, D. (2017). Bacterial signaling at the intestinal epithelial interface in inflammation and cancer. *Front. Immunol.* *8*, 1927.
- Collado, M.C., Donat, E., Ribes-Koninckx, C., Calabuig, M., and Sanz, Y. (2008). Imbalances in faecal and duodenal Bifidobacterium species composition in active and non-active coeliac disease. *BMC Microbiol.* *8*, 232.
- Corradi, V., Sejdiu, B.I., Mesa-Galoso, H., Abdizadeh, H., Noskov, S.Y., Marrink, S.J., and Tieleman, D.P. (2019). Emerging Diversity in Lipid-Protein Interactions. *Chem. Rev.* *119*, 5775–5848.
- Cortese, R., Lu, L., Yu, Y., Ruden, D., and Claud, E.C. (2016). Epigenome-Microbiome crosstalk: A potential new paradigm influencing neonatal susceptibility to disease. *Epigenetics* *11*, 205–215.
- Cotto, K.C., Wagner, A.H., Feng, Y.-Y., Kiwala, S., Coffman, A.C., Spies, G., Wollam, A., Spies, N.C., Griffith, O.L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. *Nucleic Acids Res.* *46*, D1068–D1073.
- Croft, D., O’Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* *39*, D691-7.
- Cronin, M., Sleator, R.D., Hill, C., Fitzgerald, G.F., and van Sinderen, D. (2008). Development of a luciferase-based reporter system to monitor Bifidobacterium breve UCC2003 persistence in mice. *BMC Microbiol.* *8*, 161.
- Crosnier, C., Stamatakis, D., and Lewis, J. (2006). Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nat. Rev. Genet.* *7*, 349–359.
- Csályi, K., Fazekas, D., Kadlecsek, T., Türei, D., Gul, L., Horváth, B., Módos, D., Demeter, A., Pápai, N., Lenti, K., et al. (2016). SignaFish: A Zebrafish-Specific Signaling Pathway Resource. *Zebrafish* *13*, 541–544.
- Dahan, S., Roth-Walter, F., Arnaboldi, P., Agarwal, S., and Mayer, L. (2007). Epithelia: lymphocyte interactions in the gut. *Immunol. Rev.* *215*, 243–253.
- Dan, L., Liu, S., Shang, S., Zhang, H., Zhang, R., and Li, N. (2018). Expression of recombinant human lysozyme in bacterial artificial chromosome transgenic mice promotes the growth of Bifidobacterium and inhibits the growth of Salmonella in the intestine. *J. Biotechnol.* *272–273*, 33–39.
- Danese, S., Rudziński, J., Brandt, W., Dupas, J.-L., Peyrin-Biroulet, L., Bouhnik, Y., Kleczkowski, D., Uebel, P., Lukas, M., Knutsson, M., et al. (2015). Tralokinumab for moderate-to-severe UC: a randomised, double-blind, placebo-controlled, phase IIa study. *Gut* *64*, 243–249.
- Davis, T.L., and Rebay, I. (2017). Master regulators in development: Views from the Drosophila

- retinal determination and mammalian pluripotency gene networks. *Dev. Biol.* *421*, 93–107.
- De Iudibus, S., Franca, R., Martellosi, S., Ventura, A., and Decorti, G. (2011). Molecular mechanism of glucocorticoid resistance in inflammatory bowel disease. *World J. Gastroenterol.* *17*, 1095–1108.
- De Man, J.C., Rogosa, M., and Sharpe, M.E. (1960). A medium for the cultivation of lactobacilli. *Journal of Applied Bacteriology* *23*, 130–135.
- De Smet, R., and Marchal, K. (2010). Advantages and limitations of current network inference methods. *Nat. Rev. Microbiol.* *8*, 717–729.
- Defendenti, C., Sarzi-Puttini, P., Saibeni, S., Bollani, S., Bruno, S., Almasio, P.L., Declich, P., and Atzeni, F. (2015). Significance of serum IL-9 levels in inflammatory bowel disease. *Int J Immunopathol Pharmacol* *28*, 569–575.
- Dekkers, J.F., Wiegerinck, C.L., de Jonge, H.R., Bronsveld, I., Janssens, H.M., de Winter-de Groot, K.M., Brandsma, A.M., de Jong, N.W.M., Bijvelds, M.J.C., Scholte, B.J., et al. (2013). A functional CFTR assay using primary cystic fibrosis intestinal organoids. *Nat. Med.* *19*, 939–945.
- DeRyckere, D., and DeGregori, J. (2005). E2F1 and E2F2 are differentially required for homeostasis-driven and antigen-induced T cell proliferation in vivo. *J. Immunol.* *175*, 647–655.
- Di Gioia, D., Aloisio, I., Mazzola, G., and Biavati, B. (2014). Bifidobacteria: their impact on gut microbiota composition and their applications as probiotics in infants. *Appl. Microbiol. Biotechnol.* *98*, 563–577.
- Di Narzo, A.F., Peters, L.A., Argmann, C., Stojmirovic, A., Perrigoue, J., Li, K., Telesco, S., Kidd, B., Walker, J., Dudley, J., et al. (2016). Blood and Intestine eQTLs from an Anti-TNF-Resistant Crohn's Disease Cohort Inform IBD Genetic Association Loci. *Clin Transl Gastroenterol* *7*, e177.
- Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N.P., Trave, G., and Gibson, T.J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front. Biosci.* *13*, 6580–6603.
- Din, A.U., Hassan, A., Zhu, Y., Zhang, K., Wang, Y., Li, T., Wang, Y., and Wang, G. (2020). Inhibitory effect of *Bifidobacterium bifidum* ATCC 29521 on colitis and its mechanism. *J. Nutr. Biochem.* *79*, 108353.
- Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kühn, H., Behrendt, A., et al. (2016). ELM 2016--data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.* *44*, D294-300.
- Dittmer, J. (2003). The biology of the Ets1 proto-oncogene. *Mol. Cancer* *2*, 29.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.
- Dorrington, M.G., and Fraser, I.D.C. (2019). NF- κ B Signaling in Macrophages: Dynamics, Crosstalk, and Signal Integration. *Front. Immunol.* *10*, 705.
- Doxey, A.C., and McConkey, B.J. (2013). Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence* *4*, 453–466.
- Drury, R.E., O'Connor, D., and Pollard, A.J. (2017). The clinical application of miRNAs in infectious disease. *Front. Immunol.* *8*, 1182.
- Du, Y., Huang, Q., Arisdakessian, C., and Garmire, L.X. (2020). Evaluation of STAR and Kallisto on Single Cell RNA-Seq Data Alignment. *G3 (Bethesda)* *10*, 1775–1783.

- Duan, Y., Tian, L., Gao, Q., Liang, L., Zhang, W., Yang, Y., Zheng, Y., Pan, E., Li, S., and Tang, N. (2016). Chromatin remodeling gene ARID2 targets cyclin D1 and cyclin E1 to suppress hepatoma cell progression. *Oncotarget* 7, 45863–45875.
- Dubois-Camacho, K., Ottum, P.A., Franco-Muñoz, D., De la Fuente, M., Torres-Riquelme, A., Díaz-Jiménez, D., Olivares-Morales, M., Astudillo, G., Quera, R., and Hermoso, M.A. (2017). Glucocorticosteroid therapy in inflammatory bowel diseases: From clinical practice to molecular biology. *World J. Gastroenterol.* 23, 6628–6638.
- Dugourd, A., and Saez-Rodriguez, J. (2019). Footprint-based functional analysis of multi-omic data. *Current Opinion in Systems Biology* 15, 82–90.
- Dugourd, A., Kuppe, C., Sciacovelli, M., Gjerga, E., Emdal, K.B., Bekker-Jensen, D.B., Kranz, J., Bindels, E.J.M., Costa, S., Olsen, J.V., et al. (2020). Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *BioRxiv*.
- Duranti, S., Gaiani, F., Mancabelli, L., Milani, C., Grandi, A., Bolchi, A., Santoni, A., Lugli, G.A., Ferrario, C., Mangifesta, M., et al. (2016). Elucidating the gut microbiome of ulcerative colitis: bifidobacteria as novel microbial biomarkers. *FEMS Microbiol. Ecol.* 92.
- Eken, A., Singh, A.K., Treuting, P.M., and Oukka, M. (2014). IL-23R+ innate lymphoid cells induce colitis via interleukin-22-dependent mechanism. *Mucosal Immunol.* 7, 143–154.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47, D427–D432.
- Emmrich, J., Seyfarth, M., Fleig, W.E., and Emmrich, F. (1991). Treatment of inflammatory bowel disease with anti-CD4 monoclonal antibody. *Lancet* 338, 570–571.
- Engevik, M.A., Luk, B., Chang-Graham, A.L., Hall, A., Herrmann, B., Ruan, W., Endres, B.T., Shi, Z., Garey, K.W., Hyser, J.M., et al. (2019). Bifidobacterium dentium Fortifies the Intestinal Mucus Layer via Autophagy and Calcium Signaling Pathways. *MBio* 10.
- Ertosun, M.G., Hapil, F.Z., and Osman Nidai, O. (2016). E2F1 transcription factor and its impact on growth factor and cytokine signaling. *Cytokine Growth Factor Rev.* 31, 17–25.
- EU Clinical Trials Register (2019). EudraCT Number 2016-000420-26 - Clinical trial results - EU Clinical Trials Register.
- Ewaschuk, J.B., Diaz, H., Meddings, L., Diederichs, B., Dmytrash, A., Backer, J., Looijer-van Langen, M., and Madsen, K.L. (2008). Secreted bioactive factors from Bifidobacterium infantis enhance epithelial cell barrier function. *Am. J. Physiol. Gastrointest. Liver Physiol.* 295, G1025-34.
- Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048.
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018a). The reactome pathway knowledgebase. *Nucleic Acids Res.* 46, D649–D655.
- Fabregat, A., Korninger, F., Viteri, G., Sidiropoulos, K., Marin-Garcia, P., Ping, P., Wu, G., Stein, L., D'Eustachio, P., and Hermjakob, H. (2018b). Reactome graph database: Efficient access to complex pathway data. *PLoS Comput. Biol.* 14, e1005968.
- Fahlgren, A., Hammarström, S., Danielsson, A., and Hammarström, M.L. (2003). Increased expression of antimicrobial peptides and lysozyme in colonic epithelial cells of patients with

ulcerative colitis. *Clin. Exp. Immunol.* *131*, 90–101.

Fakhoury, M., Negrulj, R., Mooranian, A., and Al-Salami, H. (2014). Inflammatory bowel disease: clinical aspects and treatments. *J Inflamm Res* *7*, 113–120.

Fallani, M., Young, D., Scott, J., Norin, E., Amarri, S., Adam, R., Aguilera, M., Khanna, S., Gil, A., Edwards, C.A., et al. (2010). Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *J. Pediatr. Gastroenterol. Nutr.* *51*, 77–84.

Fanning, S., Hall, L.J., and van Sinderen, D. (2012a). Bifidobacterium breve UCC2003 surface exopolysaccharide production is a beneficial trait mediating commensal-host interaction through immune modulation and pathogen protection. *Gut Microbes* *3*, 420–425.

Fanning, S., Hall, L.J., Cronin, M., Zomer, A., MacSharry, J., Goulding, D., Motherway, M.O., Shanahan, F., Nally, K., Dougan, G., et al. (2012b). Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc. Natl. Acad. Sci. USA* *109*, 2108–2113.

Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shores, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* *518*, 337–343.

Farin, H.F., Van Es, J.H., and Clevers, H. (2012). Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. *Gastroenterology* *143*, 1518–1529.e7.

Fatehullah, A., Tan, S.H., and Barker, N. (2016). Organoids as an in vitro model of human development and disease. *Nat. Cell Biol.* *18*, 246–254.

Favier, C., Neut, C., Mizon, C., Cortot, A., Colombel, J.F., and Mizon, J. (1997). Fecal beta-D-galactosidase production and Bifidobacteria are decreased in Crohn's disease. *Dig. Dis. Sci.* *42*, 817–822.

Fazekas, D., Koltai, M., Türei, D., Módos, D., Pálffy, M., Dúl, Z., Zsákai, L., Szalay-Bekó, M., Lenti, K., Farkas, I.J., et al. (2013). Signalink 2 - a signaling pathway resource with multi-layered regulatory networks. *BMC Syst. Biol.* *7*, 7.

Feagan, B.G., Sandborn, W.J., Gasink, C., Jacobstein, D., Lang, Y., Friedman, J.R., Blank, M.A., Johanns, J., Gao, L.-L., Miao, Y., et al. (2016). Ustekinumab as induction and maintenance therapy for crohn's disease. *N. Engl. J. Med.* *375*, 1946–1960.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* *16*, 278.

van der Flier, L.G., and Clevers, H. (2009). Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu. Rev. Physiol.* *71*, 241–260.

van der Flier, L.G., van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Begthel, H., van den Born, M., Guryev, V., Oving, I., et al. (2009). Transcription factor achaete scute-like 2 controls intestinal stem cell fate. *Cell* *136*, 903–912.

Flores, A.Y.L. (2018). Effect of the blocking by monoclonal antibodies of TLR2 and CD14 receptors on the expression of intracellular signaling molecules and the production of cytokines in THP-1 macrophages infected with Mycobacterium tuberculosis. *The Journal of Immunology* *200*, 109.18-109.18.

- Franke, A., McGovern, D.P.B., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., et al. (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* *42*, 1118–1125.
- Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* *47*, D766–D773.
- Fujii, M., Matano, M., Nanki, K., and Sato, T. (2015). Efficient genetic engineering of human intestinal organoids using electroporation. *Nat. Protoc.* *10*, 1474–1485.
- Fujino, S., Andoh, A., Bamba, S., Ogawa, A., Hata, K., Araki, Y., Bamba, T., and Fujiyama, Y. (2003). Increased expression of interleukin 17 in inflammatory bowel disease. *Gut* *52*, 65–70.
- Furey, T.S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* *13*, 840–852.
- Furrie, E., Macfarlane, S., Kennedy, A., Cummings, J.H., Walsh, S.V., O'neil, D.A., and Macfarlane, G.T. (2005). Synbiotic therapy (Bifidobacterium longum/Synergy 1) initiates resolution of inflammation in patients with active ulcerative colitis: a randomised controlled pilot trial. *Gut* *54*, 242–249.
- von Furstenberg, R.J., Gulati, A.S., Baxi, A., Doherty, J.M., Stappenbeck, T.S., Gracz, A.D., Magness, S.T., and Henning, S.J. (2011). Sorting mouse jejunal epithelial cells with CD24 yields a population with characteristics of intestinal stem cells. *Am. J. Physiol. Gastrointest. Liver Physiol.* *300*, G409-17.
- Fuss, I.J. (2008). Is the Th1/Th2 paradigm of immune regulation applicable to IBD? *Inflamm. Bowel Dis.* *14 Suppl 2*, S110-2.
- Fuss, I.J., and Strober, W. (2008). The role of IL-13 and NK T cells in experimental and human ulcerative colitis. *Mucosal Immunol.* *1 Suppl 1*, S31-3.
- Fuss, I.J., Heller, F., Boirivant, M., Leon, F., Yoshida, M., Fichtner-Feigl, S., Yang, Z., Exley, M., Kitani, A., Blumberg, R.S., et al. (2004). Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. *J. Clin. Invest.* *113*, 1490–1497.
- Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* *29*, 1363–1375.
- Garrett, W.S., Lord, G.M., Punit, S., Lugo-Villarino, G., Mazmanian, S.K., Ito, S., Glickman, J.N., and Glimcher, L.H. (2007). Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell* *131*, 33–45.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2014). *Bayesian Data Analysis (Chapman & Hall/CRC Texts in Statistical Science)* (Boca Raton: Chapman and Hall/CRC).
- Gerbe, F., Brulin, B., Makrini, L., Legraverend, C., and Jay, P. (2009). DCAMKL-1 expression identifies Tuft cells rather than stem cells in the adult mouse intestinal epithelium. *Gastroenterology* *137*, 2179–80; author reply 2180.
- Gerlach, K., Hwang, Y., Nikolaev, A., Atreya, R., Dornhoff, H., Steiner, S., Lehr, H.-A., Wirtz, S., Vieth, M., Waisman, A., et al. (2014). TH9 cells that express the transcription factor PU.1 drive T cell-mediated colitis via IL-9 receptor signaling in intestinal epithelial cells. *Nat. Immunol.* *15*, 676–686.

- Gersemann, M., Becker, S., Kübler, I., Koslowski, M., Wang, G., Herrlinger, K.R., Griger, J., Fritz, P., Fellermann, K., Schwab, M., et al. (2009). Differences in goblet cell differentiation between Crohn's disease and ulcerative colitis. *Differentiation*. 77, 84–94.
- Giaffer, M.H., Holdsworth, C.D., and Duerden, B.I. (1991). The assessment of faecal flora in patients with inflammatory bowel disease by a simplified bacteriological technique. *J. Med. Microbiol.* 35, 238–243.
- Gibson, T.J., Dinkel, H., Van Roey, K., and Diella, F. (2015). Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. *Cell Commun. Signal.* 13, 42.
- Godl, K., Johansson, M.E.V., Lidell, M.E., Mörgelin, M., Karlsson, H., Olson, F.J., Gum, J.R., Kim, Y.S., and Hansson, G.C. (2002). The N terminus of the MUC2 mucin forms trimers that are held together within a trypsin-resistant core fragment. *J. Biol. Chem.* 277, 47248–47256.
- Goenawan, I.H., Bryan, K., and Lynn, D.J. (2016). DyNet: visualization and analysis of dynamic molecular interaction networks. *Bioinformatics* 32, 2713–2715.
- Goff, L.A., and Rinn, J.L. (2015). Linking RNA biology to lncRNAs. *Genome Res.* 25, 1456–1465.
- Gombart, A.F., Borregaard, N., and Koeffler, H.P. (2005). Human cathelicidin antimicrobial peptide (CAMP) gene is a direct target of the vitamin D receptor and is strongly up-regulated in myeloid cells by 1,25-dihydroxyvitamin D₃. *FASEB J.* 19, 1067–1077.
- Gourbeyre, P., Berri, M., Lippi, Y., Meurens, F., Vincent-Naulleau, S., Laffitte, J., Rogel-Gaillard, C., Pinton, P., and Oswald, I.P. (2015). Pattern recognition receptors in the gut: analysis of their expression along the intestinal tract and the crypt/villus axis. *Physiol. Rep.* 3.
- Gracz, A.D., Ramalingam, S., and Magness, S.T. (2010). Sox9 expression marks a subset of CD24-expressing small intestine epithelial stem cells that form organoids in vitro. *Am. J. Physiol. Gastrointest. Liver Physiol.* 298, G590-600.
- Gracz, A.D., Fuller, M.K., Wang, F., Li, L., Stelzner, M., Dunn, J.C.Y., Martin, M.G., and Magness, S.T. (2013). Brief report: CD24 and CD44 mark human intestinal epithelial cell populations with characteristics of active and facultative stem cells. *Stem Cells* 31, 2024–2030.
- Grenningloh, R., Kang, B.Y., and Ho, I.-C. (2005). Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses. *J. Exp. Med.* 201, 615–626.
- Grigat, J., Soruri, A., Forssmann, U., Riggert, J., and Zwirner, J. (2007). Chemoattraction of macrophages, T lymphocytes, and mast cells is evolutionarily conserved within the human alpha-defensin family. *J. Immunol.* 179, 3958–3965.
- Groschwitz, K.R., and Hogan, S.P. (2009). Intestinal barrier function: molecular regulation and disease pathogenesis. *J. Allergy Clin. Immunol.* 124, 3–20; quiz 21.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.
- Guan, Q. (2019). A comprehensive review and update on the pathogenesis of inflammatory bowel disease. *J. Immunol. Res.* 2019, 7247238.
- Günther, C., Martini, E., Wittkopf, N., Amann, K., Weigmann, B., Neumann, H., Waldner, M.J., Hedrick, S.M., Tenzer, S., Neurath, M.F., et al. (2011). Caspase-8 regulates TNF- α -induced epithelial necroptosis and terminal ileitis. *Nature* 477, 335–339.
- Guo, S., Guo, Y., Ergun, A., Lu, L., Walker, W.A., and Ganguli, K. (2015). Secreted Metabolites of *Bifidobacterium infantis* and *Lactobacillus acidophilus* Protect Immature Human Enterocytes

- from IL-1 β -Induced Inflammation: A Transcription Profiling Analysis. *PLoS One* 10, e0124549.
- Guven-Maiorov, E., Tsai, C.-J., and Nussinov, R. (2017). Structural host-microbiota interaction networks. *PLoS Comput. Biol.* 13, e1005579.
- Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal epithelium. *Nature* 551, 333–339.
- Han, J.-D.J. (2008). Understanding biological functions through molecular networks. *Cell Res.* 18, 224–237.
- Han, H., Shim, H., Shin, D., Shim, J.E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., et al. (2015). TRRUST: a reference database of human transcriptional regulatory interactions. *Sci. Rep.* 5, 11432.
- Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., et al. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res.* 46, D380–D386.
- Han, P., Gopalakrishnan, C., Yu, H., and Wang, E. (2017). Gene Regulatory Network Rewiring in the Immune Cells Associated with Cancer. *Genes (Basel)* 8.
- Harburger, D.S., and Calderwood, D.A. (2009). Integrin signalling at a glance. *J. Cell Sci.* 122, 159–163.
- Harrison, P.W., Alako, B., Amid, C., Cerdeño-Tárraga, A., Cleland, I., Holt, S., Hussein, A., Jayathilaka, S., Kay, S., Keane, T., et al. (2019). The european nucleotide archive in 2018. *Nucleic Acids Res.* 47, D84–D88.
- Hart, A.L., Lammers, K., Brigidi, P., Vitali, B., Rizzello, F., Gionchetti, P., Campieri, M., Kamm, M.A., Knight, S.C., and Stagg, A.J. (2004). Modulation of human dendritic cell phenotype and function by probiotic bacteria. *Gut* 53, 1602–1609.
- Hato, T., and Dagher, P.C. (2015). How the innate immune system senses trouble and causes trouble. *Clin. J. Am. Soc. Nephrol.* 10, 1459–1469.
- He, F., Morita, H., Ouwehand, A.C., Hosoda, M., Hiramatsu, M., Kurisaki, J., Isolauri, E., Benno, Y., and Salminen, S. (2002). Stimulation of the secretion of pro-inflammatory cytokines by Bifidobacterium strains. *Microbiol. Immunol.* 46, 781–785.
- He, T., Priebe, M.G., Zhong, Y., Huang, C., Harmsen, H.J.M., Raangs, G.C., Antoine, J.M., Welling, G.W., and Vonk, R.J. (2008). Effects of yogurt and bifidobacteria supplementation on the colonic microbiota in lactose-intolerant subjects. *J. Appl. Microbiol.* 104, 595–604.
- Heel, K.A., McCauley, R.D., Papadimitriou, J.M., and Hall, J.C. (1997). Peyer's patches. *J. Gastroenterol. Hepatol.* 12, 122–136.
- Heimerl, S., Moehle, C., Zahn, A., Boettcher, A., Stremmel, W., Langmann, T., and Schmitz, G. (2006). Alterations in intestinal fatty acid metabolism in inflammatory bowel disease. *Biochim. Biophys. Acta* 1762, 341–350.
- Heller, F., Florian, P., Bojarski, C., Richter, J., Christ, M., Hillenbrand, B., Mankertz, J., Gitter, A.H., Bürgel, N., Fromm, M., et al. (2005). Interleukin-13 Is the Key Effector Th2 Cytokine in Ulcerative Colitis That Affects Epithelial Tight Junctions, Apoptosis, and Cell Restitution. *Gastroenterology* 129, 550–564.
- Hershey, G.K.K. (2003). IL-13 receptors and signaling pathways: an evolving web. *J. Allergy Clin. Immunol.* 111, 677–90; quiz 691.

- Honda, K., and Taniguchi, T. (2006). IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat. Rev. Immunol.* 6, 644–658.
- Hsieh, C.-Y., Osaka, T., Moriyama, E., Date, Y., Kikuchi, J., and Tsuneda, S. (2015). Strengthening of the intestinal epithelial tight junction by *Bifidobacterium bifidum*. *Physiol. Rep.* 3.
- Hua, G., Zhu, B., Rosa, F., Deblon, N., Adélaïde, J., Kahn-Perlès, B., Birnbaum, D., and Imbert, J. (2009). A negative feedback regulatory loop associates the tyrosine kinase receptor ERBB2 and the transcription factor GATA4 in breast cancer cells. *Mol. Cancer Res.* 7, 402–414.
- Huang, J.K., Carlin, D.E., Yu, M.K., Zhang, W., Kreisberg, J.F., Tamayo, P., and Ideker, T. (2018). Systematic evaluation of molecular networks for discovery of disease genes. *Cell Syst.* 6, 484–495.e5.
- Huang, L., Liao, L., and Wu, C.H. (2016). Inference of protein-protein interaction networks from multiple heterogeneous data. *EURASIP J. Bioinform. Syst. Biol.* 2016, 8.
- Huang, Y.-C., Lin, S.-J., Shih, H.-Y., Chou, C.-H., Chu, H.-H., Chiu, C.-C., Yuh, C.-H., Yeh, T.-H., and Cheng, Y.-C. (2017). Epigenetic regulation of NOTCH1 and NOTCH3 by KMT2A inhibits glioma proliferation. *Oncotarget* 8, 63110–63120.
- Hughes, K.R., Harnisch, L.C., Alcon-Giner, C., Mitra, S., Wright, C.J., Ketskemety, J., van Sinderen, D., Watson, A.J.M., and Hall, L.J. (2017). *Bifidobacterium breve* reduces apoptotic epithelial cell shedding in an exopolysaccharide and MyD88-dependent manner. *Open Biol* 7.
- Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., et al. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res.* 37, D211-5.
- Imam, T., Park, S., Kaplan, M.H., and Olson, M.R. (2018). Effector T helper cell subsets in inflammatory bowel diseases. *Front. Immunol.* 9, 1212.
- Imielinski, M., Baldassano, R.N., Griffiths, A., Russell, R.K., Annese, V., Dubinsky, M., Kugathasan, S., Bradfield, J.P., Walters, T.D., Sleiman, P., et al. (2009). Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat. Genet.* 41, 1335–1340.
- Inaba, Y., Ueno, N., Numata, M., Zhu, X., Messer, J.S., Boone, D.L., Fujiya, M., Kohgo, Y., Musch, M.W., and Chang, E.B. (2016). Soluble bioactive microbial mediators regulate proteasomal degradation and autophagy to protect against inflammation-induced stress. *Am. J. Physiol. Gastrointest. Liver Physiol.* 311, G634–G647.
- Isolauri, E. (2012). Development of healthy gut microbiota early in life. *J. Paediatr. Child Health* 48 Suppl 3, 1–6.
- Ito, R., Shin-Ya, M., Kishida, T., Urano, A., Takada, R., Sakagami, J., Imanishi, J., Kita, M., Ueda, Y., Iwakura, Y., et al. (2006). Interferon-gamma is causatively involved in experimental inflammatory bowel disease in mice. *Clin. Exp. Immunol.* 146, 330–338.
- Ivanov, D., Emonet, C., Foata, F., Affolter, M., Delley, M., Fisseha, M., Blum-Sperisen, S., Kochhar, S., and Arigoni, F. (2006). A serpin from the gut bacterium *Bifidobacterium longum* inhibits eukaryotic elastase-like serine proteases. *J. Biol. Chem.* 281, 17246–17252.
- Jaks, V., Barker, N., Kasper, M., van Es, J.H., Snippert, H.J., Clevers, H., and Toftgård, R. (2008). Lgr5 marks cycling, yet long-lived, hair follicle stem cells. *Nat. Genet.* 40, 1291–1299.
- Jalali, S., Bhartiya, D., Lalwani, M.K., Sivasubbu, S., and Scaria, V. (2013). Systematic transcriptome wide analysis of lncRNA-miRNA interactions. *PLoS One* 8, e53823.
- Jiao, Y., Wu, L., Huntington, N.D., and Zhang, X. (2020). Crosstalk between gut microbiota and

- innate immunity and its implication in autoimmune diseases. *Front. Immunol.* **11**, 282.
- Jin, Y., Lin, Y., Lin, L., and Zheng, C. (2012). IL-17/IFN- γ interactions regulate intestinal inflammation in TNBS-induced acute colitis. *J. Interferon Cytokine Res.* **32**, 548–556.
- Jo, A., Denduluri, S., Zhang, B., Wang, Z., Yin, L., Yan, Z., Kang, R., Shi, L.L., Mok, J., Lee, M.J., et al. (2014). The versatile functions of Sox9 in development, stem cells, and human diseases. *Genes Dis.* **1**, 149–161.
- Johansson, M.E.V., Gustafsson, J.K., Holmén-Larsson, J., Jabbar, K.S., Xia, L., Xu, H., Ghishan, F.K., Carvalho, F.A., Gewirtz, A.T., Sjövall, H., et al. (2014). Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut* **63**, 281–291.
- John, G., Hegarty, J.P., Yu, W., Berg, A., Pastor, D.M., Kelly, A.A., Wang, Y., Poritz, L.S., Schreiber, S., Koltun, W.A., et al. (2011). NKX2-3 variant rs11190140 is associated with IBD and alters binding of NFAT. *Mol. Genet. Metab.* **104**, 174–179.
- Jones, J.C., and Dempsey, P.J. (2016). Enterocyte progenitors can dedifferentiate to replace lost Lgr5+ intestinal stem cells revealing that many different progenitor populations can regain stemness. *Stem Cell Investig.* **3**, 61.
- Jones, E.J., Matthews, Z.J., Gul, L., Sudhakar, P., Treveil, A., Divekar, D., Buck, J., Wrzesinski, T., Jefferson, M., Armstrong, S.D., et al. (2019). Integrative analysis of Paneth cell proteomic and transcriptomic data from intestinal organoids reveals functional processes dependent on autophagy. *Dis. Model. Mech.* **12**.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240.
- Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124.
- Jung, J., and Jung, H. (2016). Methods to analyze cell type-specific gene expression profiles from heterogeneous cell populations. *Animal Cells Syst (Seoul)* **20**, 113–117.
- Kaalia, R., and Rajapakse, J.C. (2019). Refining modules to determine functionally significant clusters in molecular networks. *BMC Genomics* **20**, 901.
- Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036.
- Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2007). Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. *Nucleic Acids Res.* **35**, W429–32.
- Kamada, N., Hisamatsu, T., Okamoto, S., Chinen, H., Kobayashi, T., Sato, T., Sakuraba, A., Kitazume, M.T., Sugita, A., Koganei, K., et al. (2008). Unique CD14 intestinal macrophages contribute to the pathogenesis of Crohn disease via IL-23/IFN-gamma axis. *J. Clin. Invest.* **118**, 2269–2280.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361.
- Kaser, A., and Blumberg, R.S. (2009). Endoplasmic reticulum stress in the intestinal epithelium and inflammatory bowel disease. *Semin. Immunol.* **21**, 156–163.

- Kaser, A., Lee, A.-H., Franke, A., Glickman, J.N., Zeissig, S., Tilg, H., Nieuwenhuis, E.E.S., Higgins, D.E., Schreiber, S., Glimcher, L.H., et al. (2008). XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. *Cell* *134*, 743–756.
- Kaser, A., Flak, M.B., Tomczak, M.F., and Blumberg, R.S. (2011). The unfolded protein response and its role in intestinal homeostasis and inflammation. *Exp. Cell Res.* *317*, 2772–2779.
- Khalif, I.L., Quigley, E.M.M., Konovitch, E.A., and Maximova, I.D. (2005). Alterations in the colonic flora and intestinal permeability and evidence of immune activation in chronic constipation. *Dig. Liver Dis.* *37*, 838–849.
- Kiely, P.D. (1998). The Th1-Th2 model--what relevance to inflammatory arthritis? *Ann. Rheum. Dis.* *57*, 328–330.
- Kim, Y.S., and Ho, S.B. (2010). Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Curr. Gastroenterol. Rep.* *12*, 319–330.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* *14*, R36.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015a). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* *12*, 357–360.
- Kim, M., Ashida, H., Ogawa, M., Yoshikawa, Y., Mimuro, H., and Sasakawa, C. (2010). Bacterial interactions with the host epithelium. *Cell Host Microbe* *8*, 20–35.
- Kim, S.-E., Choi, S.C., Park, K.S., Park, M.I., Shin, J.E., Lee, T.H., Jung, K.W., Koo, H.S., Myung, S.-J., and Constipation Research group of Korean Society of Neurogastroenterology and Motility (2015b). Change of Fecal Flora and Effectiveness of the Short-term VSL#3 Probiotic Treatment in Patients With Functional Constipation. *J Neurogastroenterol Motil* *21*, 111–120.
- King, J.B., von Furstenberg, R.J., Smith, B.J., McNaughton, K.K., Galanko, J.A., and Henning, S.J. (2012). CD24 can be used to isolate Lgr5+ putative colonic epithelial stem cells in mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* *303*, G443-52.
- Kinnebrew, M.A., Buffie, C.G., Diehl, G.E., Zenewicz, L.A., Leiner, I., Hohl, T.M., Flavell, R.A., Littman, D.R., and Pamer, E.G. (2012). Interleukin 23 production by intestinal CD103(+)CD11b(+) dendritic cells in response to bacterial flagellin enhances mucosal innate immune defense. *Immunity* *36*, 276–287.
- Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., et al. (2011). Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* *2011*, bar030.
- Kiu, R., Treveil, A., Harnisch, L.C., Caim, S., Leclaire, C., van Sinderen, D., Korcsmaros, T., and Hall, L.J. (2020). *Bifidobacterium breve* UCC2003 Induces a Distinct Global Transcriptomic Program in Neonatal Murine Intestinal Epithelial Cells. *IScience* *23*, 101336.
- de Kivit, S., van Hoffen, E., Korthagen, N., Garssen, J., and Willemsen, L.E.M. (2011). Apical TLR ligation of intestinal epithelial cells drives a Th1-polarized regulatory or inflammatory type effector response in vitro. *Immunobiology* *216*, 518–527.
- Knoop, K.A., and Newberry, R.D. (2018). Goblet cells: multifaceted players in immunity at mucosal surfaces. *Mucosal Immunol.* *11*, 1551–1557.
- Koelink, P.J., Bloemendaal, F.M., Li, B., Westera, L., Vogels, E.W.M., van Roest, M., Gloudemans, A.K., van 't Wout, A.B., Korf, H., Vermeire, S., et al. (2019). Anti-TNF therapy in IBD exerts its

therapeutic effect through macrophage IL-10 signalling. *Gut*.

Komano, H., Fujiura, Y., Kawaguchi, M., Matsumoto, S., Hashimoto, Y., Obana, S., Mombaerts, P., Tonegawa, S., Yamamoto, H., and Itohara, S. (1995). Homeostatic regulation of intestinal epithelia by intraepithelial gamma delta T cells. *Proc. Natl. Acad. Sci. USA* *92*, 6147–6151.

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* *35*, W345-9.

König, J., Wells, J., Cani, P.D., García-Ródenas, C.L., MacDonald, T., Mercenier, A., Whyte, J., Troost, F., and Brummer, R.-J. (2016). Human intestinal barrier function in health and disease. *Clin Transl Gastroenterol* *7*, e196.

Konno, S., Iizuka, M., Yukawa, M., Sasaki, K., Sato, A., Horie, Y., Nanjo, H., Fukushima, T., and Watanabe, S. (2004). Altered expression of angiogenic factors in the VEGF-Ets-1 cascades in inflammatory bowel disease. *J. Gastroenterol.* *39*, 931–939.

Kopylova, E., Noé, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* *28*, 3211–3217.

Korcsmáros, T., Farkas, I.J., Szalay, M.S., Rovó, P., Fazekas, D., Spiró, Z., Böde, C., Lenti, K., Vellai, T., and Csermely, P. (2010). Uniformly curated signaling pathways reveal tissue-specific cross-talks and support drug target discovery. *Bioinformatics* *26*, 2042–2050.

Korcsmaros, T., Dunai, Z.A., Vellai, T., and Csermely, P. (2013). Teaching the bioinformatics of signaling networks: an integrated approach to facilitate multi-disciplinary learning. *Brief. Bioinformatics* *14*, 618–632.

Korzenik, J., Larsen, M.D., Nielsen, J., Kjeldsen, J., and Nørgård, B.M. (2019). Increased risk of developing Crohn's disease or ulcerative colitis in 17 018 patients while under treatment with anti-TNF α agents, particularly etanercept, for autoimmune diseases other than inflammatory bowel disease. *Aliment. Pharmacol. Ther.* *50*, 289–294.

Köster, J., and Rahmann, S. (2012). Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics* *28*, 2520–2522.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* *42*, D68-73.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* *47*, D155–D162.

Kozuka, K., He, Y., Koo-McCoy, S., Kumaraswamy, P., Nie, B., Shaw, K., Chan, P., Leadbetter, M., He, L., Lewis, J.G., et al. (2017). Development and characterization of a human and mouse intestinal epithelial cell monolayer platform. *Stem Cell Rep.* *9*, 1976–1990.

Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* *30*, 523–530.

Kröger, A., Köster, M., Schroeder, K., Hauser, H., and Mueller, P.P. (2002). Activities of IRF-1. *J. Interferon Cytokine Res.* *22*, 5–14.

Kronman, M.P., Zaoutis, T.E., Haynes, K., Feng, R., and Coffin, S.E. (2012). Antibiotic exposure and IBD development among children: a population-based cohort study. *Pediatrics* *130*, e794-803.

Krueger, F. (2019). FelixKrueger/TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS

data.

Krug, S.M., Schulzke, J.D., and Fromm, M. (2014). Tight junction, selective permeability, and related diseases. *Semin. Cell Dev. Biol.* *36*, 166–176.

Kubisch, J., Türei, D., Földvári-Nagy, L., Dunai, Z.A., Zsákai, L., Varga, M., Vellai, T., Csermely, P., and Korcsmáros, T. (2013). Complex regulation of autophagy in cancer - integrated approaches to discover the networks that hold a double-edged sword. *Semin. Cancer Biol.* *23*, 252–261.

Kumar Bajpai, A., Davuluri, S., Tiwary, K., Narayanan, S., Oguru, S., Basavaraju, K., Dayalan, D., Thirumurugan, K., and Acharya, K.K. (2020). Systematic comparison of the protein-protein interaction databases from a user's perspective. *J. Biomed. Inform.* 103380.

Lamb, C.A., Kennedy, N.A., Raine, T., Hendy, P.A., Smith, P.J., Limdi, J.K., Hayee, B., Lomer, M.C.E., Parkes, G.C., Selinger, C., et al. (2019). British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut* *68*, s1–s106.

Landa, I., Ganly, I., Chan, T.A., Mitsutake, N., Matsuse, M., Ibrahimasic, T., Ghossein, R.A., and Fagin, J.A. (2013). Frequent somatic TERT promoter mutations in thyroid cancer: higher prevalence in advanced forms of the disease. *J. Clin. Endocrinol. Metab.* *98*, E1562-6.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* *9*, 559.

Langrish, C.L., Chen, Y., Blumenschein, W.M., Mattson, J., Basham, B., Sedgwick, J.D., McClanahan, T., Kastelein, R.A., and Cua, D.J. (2005). IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J. Exp. Med.* *201*, 233–240.

Lassen, K.G., and Xavier, R.J. (2017). Genetic control of autophagy underlies pathogenesis of inflammatory bowel disease. *Mucosal Immunol.* *10*, 589–597.

Launay, G., Salza, R., Multedo, D., Thierry-Mieg, N., and Ricard-Blum, S. (2015). MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. *Nucleic Acids Res.* *43*, D321-7.

Leach, M.R., and Williams, D.B. (2003). Calnexin and calreticulin, molecular chaperones of the endoplasmic reticulum. In *Calreticulin*, P. Eggleton, and M. Michalak, eds. (Boston, MA: Springer US), pp. 49–62.

Lee, D.-K., Kim, M.-J., Ham, J.-W., An, H.-M., Cha, M.-K., Lee, S.-W., Park, C.-I., Shin, S.-H., Lee, K.-O., Kim, K.-J., et al. (2012). In vitro evaluation of antibacterial activities and anti-inflammatory effects of *Bifidobacterium* spp. addressing acne vulgaris. *Arch. Pharm. Res.* *35*, 1065–1071.

Lee, Y.-N., Tuckerman, J., Nechushtan, H., Schutz, G., Razin, E., and Angel, P. (2004). c-Fos as a regulator of degranulation and cytokine production in FcεpsilonRI-activated mast cells. *J. Immunol.* *173*, 2571–2577.

Lee, Y.-S., Kim, T.-Y., Kim, Y., Lee, S.-H., Kim, S., Kang, S.W., Yang, J.-Y., Baek, I.-J., Sung, Y.H., Park, Y.-Y., et al. (2018). Microbiota-Derived Lactate Accelerates Intestinal Stem-Cell-Mediated Epithelial Development. *Cell Host Microbe* *24*, 833–846.e6.

Lehrer, R.I., Jung, G., Ruchala, P., Andre, S., Gabius, H.J., and Lu, W. (2009). Multivalent binding of carbohydrates by the human alpha-defensin, HD5. *J. Immunol.* *183*, 480–490.

Leiserson, M.D.M., Vandin, F., Wu, H.-T., Dobson, J.R., Eldridge, J.V., Thomas, J.L., Papoutsaki, A., Kim, Y., Niu, B., McLellan, M., et al. (2015). Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* *47*, 106–114.

- Lekstrom-Himes, J., and Xanthopoulos, K.G. (1998). Biological role of the CCAAT/enhancer-binding protein family of transcription factors. *J. Biol. Chem.* *273*, 28545–28548.
- Lesurf, R., Cotto, K.C., Wang, G., Griffith, M., Kasaian, K., Jones, S.J.M., Montgomery, S.B., Griffith, O.L., and Open Regulatory Annotation Consortium (2016). ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res.* *44*, D126–32.
- Levine, A., Sigall Boneh, R., and Wine, E. (2018). Evolving role of diet in the pathogenesis and treatment of inflammatory bowel diseases. *Gut* *67*, 1726–1738.
- Lewis, Z.T., and Mills, D.A. (2017). Differential establishment of bifidobacteria in the breastfed infant gut. *Nestle Nutr. Inst. Workshop Ser.* *88*, 149–159.
- Li, D., Haritunians, T., Potdar, A., and McGovern, D.P.B. (2018). 16 Genetic analysis identified novel loci associated with IBD. *Inflamm. Bowel Dis.* *24*, S14–S14.
- Li, L., Zhao, G.-D., Shi, Z., Qi, L.-L., Zhou, L.-Y., and Fu, Z.-X. (2016). The Ras/Raf/MEK/ERK signaling pathway and its role in the occurrence and development of HCC. *Oncol. Lett.* *12*, 3045–3050.
- Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L., Peluso, D., Calderone, A., Castagnoli, L., and Cesareni, G. (2020). SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. *Nucleic Acids Res.* *48*, D504–D510.
- Lieberkühn, J.N. (1744). *Dissertatio anatomico-physiologica de fabrica et actione villorum intestinorum tenuium hominis. Iconibus aeri incisus illustrata* (apud Conrad et Georg. Jac. Wishof).
- Lin, R., Jiang, Y., Zhao, X.Y., Guan, Y., Qian, W., Fu, X.C., Ren, H.Y., and Hou, X.H. (2014). Four types of Bifidobacteria trigger autophagy response in intestinal epithelial cells. *J Dig Dis* *15*, 597–605.
- Lindeboom, R.G., van Voorthuijsen, L., Oost, K.C., Rodríguez-Colman, M.J., Luna-Velez, M.V., Furlan, C., Baraille, F., Jansen, P.W., Ribeiro, A., Burgering, B.M., et al. (2018). Integrative multi-omics analysis of intestinal organoid differentiation. *Mol. Syst. Biol.* *14*, e8227.
- Lindfors, K., Blomqvist, T., Juuti-Uusitalo, K., Stenman, S., Venäläinen, J., Mäki, M., and Kaukinen, K. (2008). Live probiotic *Bifidobacterium lactis* bacteria inhibit the toxic effects induced by wheat gliadin in epithelial cell culture. *Clin. Exp. Immunol.* *152*, 552–558.
- Liu, Y., and Chen, Y.-G. (2018). 2D- and 3D-Based Intestinal Stem Cell Cultures for Personalized Medicine. *Cells* *7*.
- Liu, A., Trairatphisan, P., Gjerga, E., Didangelos, A., Barratt, J., and Saez-Rodriguez, J. (2019). From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. *NPJ Syst. Biol. Appl.* *5*, 40.
- Liu, T.-C., Gurram, B., Baldridge, M.T., Head, R., Lam, V., Luo, C., Cao, Y., Simpson, P., Hayward, M., Holtz, M.L., et al. (2016). Paneth cell defects in Crohn's disease patients promote dysbiosis. *JCI Insight* *1*, e86907.
- Liu, Y., Jones, B., Aruffo, A., Sullivan, K.M., Linsley, P.S., and Janeway, C.A. (1992). Heat-stable antigen is a costimulatory molecule for CD4 T cell growth. *J. Exp. Med.* *175*, 437–445.
- Long, Y., Wang, X., Youmans, D.T., and Cech, T.R. (2017). How do lncRNAs regulate transcription? *Sci. Adv.* *3*, eaao2110.
- Lu, C., Chen, J., Xu, H.-G., Zhou, X., He, Q., Li, Y.-L., Jiang, G., Shan, Y., Xue, B., Zhao, R.-X., et al. (2014). MIR106B and MIR93 prevent removal of bacteria from epithelial cells by disrupting

ATG16L1-mediated autophagy. *Gastroenterology* *146*, 188–199.

Lu, L.-F., Lind, E.F., Gondek, D.C., Bennett, K.A., Gleeson, M.W., Pino-Lagos, K., Scott, Z.A., Coyle, A.J., Reed, J.L., Van Snick, J., et al. (2006). Mast cells are essential intermediaries in regulatory T-cell tolerance. *Nature* *442*, 997–1002.

Lueschow, S.R., and McElroy, S.J. (2020). The paneth cell: the curator and defender of the immature small intestine. *Front. Immunol.* *11*, 587.

Lukovac, S., and Roeselers, G. (2015). Intestinal crypt organoids as experimental models. In *The Impact of Food Bioactives on Health: In Vitro and Ex Vivo Models*, K. Verhoeckx, P. Cotter, I. López-Expósito, C. Kleiveland, T. Lea, A. Mackie, T. Requena, D. Swiatecka, and H. Wichers, eds. (Cham (CH): Springer), p.

Luu, L., Matthews, Z.J., Armstrong, S.D., Powell, P.P., Wileman, T., Wastling, J.M., and Coombes, J.L. (2018). Proteomic Profiling of Enteroid Cultures Skewed toward Development of Specific Epithelial Lineages. *Proteomics* *18*, e1800132.

Ma, H., Tao, W., and Zhu, S. (2019). T lymphocytes in the intestinal mucosa: defense and tolerance. *Cell Mol Immunol* *16*, 216–224.

Macdonald, T.T., and Monteleone, G. (2005). Immunity, inflammation, and allergy in the gut. *Science* *307*, 1920–1925.

Macfarlane, S., Furrie, E., Cummings, J.H., and Macfarlane, G.T. (2004). Chemotaxonomic analysis of bacterial populations colonizing the rectal mucosa in patients with ulcerative colitis. *Clin. Infect. Dis.* *38*, 1690–1699.

Maclsaac, K.D., Lo, K.A., Gordon, W., Motola, S., Mazor, T., and Fraenkel, E. (2010). A quantitative model of transcriptional regulation reveals the influence of binding location on expression. *PLoS Comput. Biol.* *6*, e1000773.

Mahlapuu, M., Håkansson, J., Ringstad, L., and Björn, C. (2016). Antimicrobial peptides: an emerging category of therapeutic agents. *Front. Cell Infect. Microbiol.* *6*, 194.

Makino, H. (2018). Bifidobacterial strains in the intestines of newborns originate from their mothers. *Biosci. Microbiota Food Health* *37*, 79–85.

Mangin, I., Dossou-Yovo, F., Lévêque, C., Dessoy, M.-V., Sawoo, O., Suau, A., and Pochart, P. (2018). Oral administration of viable *Bifidobacterium pseudolongum* strain Patronus modified colonic microbiota and increased mucus layer thickness in rat. *FEMS Microbiol. Ecol.* *94*.

Mao, W., Zaslavsky, E., Hartmann, B.M., Sealfon, S.C., and Chikina, M. (2019). Pathway-level information extractor (PLIER) for gene expression data. *Nat. Methods* *16*, 607–610.

Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M., and Wold, B.J. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* *24*, 496–510.

Marshman, E., Booth, C., and Potten, C.S. (2002). The intestinal epithelial stem cell. *Bioessays* *24*, 91–98.

Marteau, P., Cuillerier, E., Meance, S., Gerhardt, M.F., Myara, A., Bouvier, M., Bouley, C., Tondu, F., Bommelaer, G., and Grimaud, J.C. (2002). *Bifidobacterium animalis* strain DN-173 010 shortens the colonic transit time in healthy women: a double-blind, randomized, controlled study. *Aliment. Pharmacol. Ther.* *16*, 587–593.

McCormick, D.A., Horton, L.W., and Mee, A.S. (1990). Mucin depletion in inflammatory bowel disease. *J. Clin. Pathol.* *43*, 143–146.

- McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv.
- Mead, B.E., and Karp, J.M. (2019). All models are wrong, but some organoids may be useful. *Genome Biol.* *20*, 66.
- Mead, B.E., Ordovas-Montanes, J., Braun, A.P., Levy, L.E., Bhargava, P., Szucs, M.J., Ammendolia, D.A., MacMullan, M.A., Yin, X., Hughes, T.K., et al. (2018). Harnessing single-cell genomics to improve the physiological fidelity of organoid-derived cell types. *BMC Biol.* *16*, 62.
- Mendoza-Parra, M.-A., Malysheva, V., Mohamed Saleem, M.A., Lieb, M., Godel, A., and Gronemeyer, H. (2016). Reconstructed cell fate-regulatory programs in stem cells reveal hierarchies and key factors of neurogenesis. *Genome Res.* *26*, 1505–1519.
- Mészáros, B., Erdos, G., and Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* *46*, W329–W337.
- Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P.D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* *47*, D419–D426.
- Miller, M.M., Barik, S., Cattin-Roy, A.N., Ukah, T.K., Hoeman, C.M., and Zaghoulani, H. (2019). A New IRF-1-Driven Apoptotic Pathway Triggered by IL-4/IL-13 Kills Neonatal Th1 Cells and Weakens Protection against Viral Infection. *J. Immunol.* *202*, 3173–3186.
- Minton, K. (2008). What “drives” IL-4 versus IL-13 signalling? *Nat. Rev. Immunol.* *8*, 167–167.
- Miryala, S.K., Anbarasu, A., and Ramaiah, S. (2018). Discerning molecular interactions: A comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* *642*, 84–94.
- Mirza, A.H., Berthelsen, C.H., Seemann, S.E., Pan, X., Frederiksen, K.S., Vilien, M., Gorodkin, J., and Pociot, F. (2015). Transcriptomic landscape of lncRNAs in inflammatory bowel disease. *Genome Med.* *7*, 39.
- Mishra, J., Waters, C.M., and Kumar, N. (2012). Molecular mechanism of interleukin-2-induced mucosal homeostasis. *Am. J. Physiol. Cell Physiol.* *302*, C735–47.
- Módos, D., Bulusu, K.C., Fazekas, D., Kubisch, J., Brooks, J., Marczell, I., Szabó, P.M., Vellai, T., Csermely, P., Lenti, K., et al. (2017). Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anticancer therapies. *NPJ Syst. Biol. Appl.* *3*, 2.
- Moisan, J., Grenningloh, R., Bettelli, E., Oukka, M., and Ho, I.-C. (2007). Ets-1 is a negative regulator of Th17 differentiation. *J. Exp. Med.* *204*, 2825–2835.
- Mokrozub, V.V., Lazarenko, L.M., Sichel, L.M., Babenko, L.P., Lytvyn, P.M., Demchenko, O.M., Melnichenko, Y.O., Boyko, N.V., Biavati, B., DiGioia, D., et al. (2015). The role of beneficial bacteria wall elasticity in regulating innate immune response. *EPMA J* *6*, 13.
- Mokry, M., Middendorp, S., Wiegerinck, C.L., Witte, M., Teunissen, H., Meddens, C.A., Cuppen, E., Clevers, H., and Nieuwenhuis, E.E.S. (2014). Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium. *Gastroenterology* *146*, 1040–1047.
- Montgomery, R.K., Carlone, D.L., Richmond, C.A., Farilla, L., Kranendonk, M.E.G., Henderson, D.E., Baffour-Awuah, N.Y., Ambruzs, D.M., Fogli, L.K., Algra, S., et al. (2011). Mouse telomerase

- reverse transcriptase (mTert) expression marks slowly cycling intestinal stem cells. *Proc. Natl. Acad. Sci. USA* *108*, 179–184.
- Moon, C., VanDussen, K.L., Miyoshi, H., and Stappenbeck, T.S. (2014). Development of a primary mouse intestinal epithelial cell monolayer culture system to evaluate factors that modulate IgA transcytosis. *Mucosal Immunol.* *7*, 818–828.
- Mouly, E., Chemin, K., Nguyen, H.V., Chopin, M., Mesnard, L., Leite-de-Moraes, M., Burlen-defranoux, O., Bandeira, A., and Bories, J.-C. (2010). The Ets-1 transcription factor controls the development and function of natural regulatory T cells. *J. Exp. Med.* *207*, 2113–2125.
- Mowat, A.M. (2003). Anatomical basis of tolerance and immunity to intestinal antigens. *Nat. Rev. Immunol.* *3*, 331–341.
- Mowat, A.M., and Agace, W.W. (2014). Regional specialization within the intestinal immune system. *Nat. Rev. Immunol.* *14*, 667–685.
- Mozdiak, E., O'Malley, J., and Arasaradnam, R. (2015). Inflammatory bowel disease. *BMJ* *351*, h4416.
- Mudunuri, U., Che, A., Yi, M., and Stephens, R.M. (2009). bioDBnet: the biological database network. *Bioinformatics* *25*, 555–556.
- Muetze, T., Goenawan, I.H., Wiencko, H.L., Bernal-Llinares, M., Bryan, K., and Lynn, D.J. (2016). Contextual Hub Analysis Tool (CHAT): A Cytoscape app for identifying contextually relevant hubs in biological networks [version 1; peer review: 1 approved]. *F1000Res.* *5*, 1745.
- Mulloikandov, G., Baccarini, A., Ruzo, A., Jayaprakash, A.D., Tung, N., Israelow, B., Evans, M.J., Sachidanandam, R., and Brown, B.D. (2012). High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. *Nat. Methods* *9*, 840–846.
- Muniz, L.R., Knosp, C., and Yeretssian, G. (2012). Intestinal antimicrobial peptides during homeostasis, infection, and disease. *Front. Immunol.* *3*, 310.
- Muñoz, J., Stange, D.E., Schepers, A.G., van de Wetering, M., Koo, B.-K., Itzkovitz, S., Volckmann, R., Kung, K.S., Koster, J., Radulescu, S., et al. (2012). The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent “+4” cell markers. *EMBO J.* *31*, 3079–3091.
- Murray, P.J. (2007). The JAK-STAT signaling pathway: input and output integration. *J. Immunol.* *178*, 2623–2629.
- Musa, J., Aynaud, M.-M., Mirabeau, O., Delattre, O., and Grünewald, T.G. (2017). MYBL2 (B-Myb): a central regulator of cell proliferation, cell survival and differentiation involved in tumorigenesis. *Cell Death Dis.* *8*, e2895.
- Muzaki, A.R.B.M., Tetlak, P., Sheng, J., Loh, S.C., Setiagani, Y.A., Poidinger, M., Zolezzi, F., Karjalainen, K., and Ruedl, C. (2016). Intestinal CD103(+)CD11b(-) dendritic cells restrain colitis via IFN- γ -induced anti-inflammatory response in epithelial cells. *Mucosal Immunol.* *9*, 336–351.
- Naiki, Y., Michelsen, K.S., Zhang, W., Chen, S., Doherty, T.M., and Arditi, M. (2005). Transforming growth factor-beta differentially inhibits MyD88-dependent, but not TRAM- and TRIF-dependent, lipopolysaccharide-induced TLR4 signaling. *J. Biol. Chem.* *280*, 5491–5495.
- Natividad, J.M.M., Hayes, C.L., Motta, J.-P., Jury, J., Galipeau, H.J., Philip, V., Garcia-Rodenas, C.L., Kiyama, H., Bercik, P., and Verdu, E.F. (2013). Differential induction of antimicrobial REGIII by the intestinal microbiota and *Bifidobacterium breve* NCC2950. *Appl. Environ. Microbiol.* *79*, 7745–7754.
- Nava, P., Koch, S., Laukoetter, M.G., Lee, W.Y., Kolegraff, K., Capaldo, C.T., Beeman, N., Addis,

- C., Gerner-Smidt, K., Neumaier, I., et al. (2010). Interferon-gamma regulates intestinal epithelial homeostasis through converging beta-catenin signaling pathways. *Immunity* 32, 392–402.
- Neurath, M.F. (2014). Cytokines in inflammatory bowel disease. *Nat. Rev. Immunol.* 14, 329–342.
- Neurath, M.F., Becker, C., and Barbulescu, K. (1998). Role of NF-kappaB in immune and inflammatory responses in the gut. *Gut* 43, 856–860.
- Ng, S.C., Shi, H.Y., Hamidi, N., Underwood, F.E., Tang, W., Benchimol, E.I., Panaccione, R., Ghosh, S., Wu, J.C.Y., Chan, F.K.L., et al. (2018). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 390, 2769–2778.
- Nguyen, T.A., Jo, M.H., Choi, Y.-G., Park, J., Kwon, S.C., Hohng, S., Kim, V.N., and Woo, J.-S. (2015). Functional anatomy of the human microprocessor. *Cell* 161, 1374–1387.
- Nicoletti, C. (2000). Unsolved mysteries of intestinal M cells. *Gut* 47, 735–739.
- Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. *Methods Mol. Biol.* 1611, 59–73.
- Noah, T.K., Donahue, B., and Shroyer, N.F. (2011). Intestinal development and differentiation. *Exp. Cell Res.* 317, 2702–2710.
- Noben, M., Vanhove, W., Arnauts, K., Santo Ramalho, A., Van Assche, G., Vermeire, S., Verfaillie, C., and Ferrante, M. (2017). Human intestinal epithelium in a dish: Current models for research into gastrointestinal pathophysiology. *United European Gastroenterol. J.* 5, 1073–1081.
- Novarino, G., Fenstermaker, A.G., Zaki, M.S., Hofree, M., Silhavy, J.L., Heiberg, A.D., Abdellateef, M., Rosti, B., Scott, E., Mansour, L., et al. (2014). Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders. *Science* 343, 506–511.
- Nowak, E.C., Weaver, C.T., Turner, H., Begum-Haque, S., Becher, B., Schreiner, B., Coyle, A.J., Kasper, L.H., and Noelle, R.J. (2009). IL-9 as a mediator of Th17-driven inflammatory disease. *J. Exp. Med.* 206, 1653–1660.
- Nuding, S., Fellermann, K., Wehkamp, J., and Stange, E.F. (2007). Reduced mucosal antimicrobial activity in Crohn's disease of the colon. *Gut* 56, 1240–1247.
- Nunes, T., Bernardazzi, C., and de Souza, H.S. (2014). Cell death and inflammatory bowel diseases: apoptosis, necrosis, and autophagy in the intestinal epithelium. *Biomed Res. Int.* 2014, 218493.
- Nusse, R. (2008). Wnt signaling and stem cell control. *Cell Res.* 18, 523–527.
- Oakley, R.H., and Cidlowski, J.A. (2013). The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. *J. Allergy Clin. Immunol.* 132, 1033–1044.
- O'Brien, K.P., Remm, M., and Sonnhammer, E.L.L. (2005). Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* 33, D476–80.
- O'Callaghan, A., and van Sinderen, D. (2016). Bifidobacteria and their role as members of the human gut microbiota. *Front. Microbiol.* 7, 925.
- O'Connell Motherway, M., Zomer, A., Leahy, S.C., Reunanen, J., Bottacini, F., Claesson, M.J., O'Brien, F., Flynn, K., Casey, P.G., Munoz, J.A.M., et al. (2011). Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc. Natl. Acad. Sci. USA* 108, 11217–11222.
- O'Connell Motherway, M., Houston, A., O'Callaghan, G., Reunanen, J., O'Brien, F., O'Driscoll, T.,

- Casey, P.G., de Vos, W.M., van Sinderen, D., and Shanahan, F. (2019). A Bifidobacterial pilus-associated protein promotes colonic epithelial proliferation. *Mol. Microbiol.* *111*, 287–301.
- O'Connor, W., Kamanaka, M., Booth, C.J., Town, T., Nakae, S., Iwakura, Y., Kolls, J.K., and Flavell, R.A. (2009). A protective function for interleukin 17A in T cell-mediated intestinal inflammation. *Nat. Immunol.* *10*, 603–609.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* *27*, 29–34.
- Ohno, H. (2016). Intestinal M cells. *J. Biochem.* *159*, 151–160.
- Okada, Y., Tsuzuki, Y., Hokari, R., Komoto, S., Kurihara, C., Kawaguchi, A., Nagao, S., and Miura, S. (2009). Anti-inflammatory effects of the genus *Bifidobacterium* on macrophages by modification of phospho-I kappaB and SOCS gene expression. *Int J Exp Pathol* *90*, 131–140.
- Okamoto, R., and Watanabe, M. (2016). Role of epithelial cells in the pathogenesis and treatment of inflammatory bowel disease. *J. Gastroenterol.* *51*, 11–21.
- Okumura, R., and Takeda, K. (2017). Roles of intestinal epithelial cells in the maintenance of gut homeostasis. *Exp Mol Med* *49*, e338.
- O'Mahony, L., McCarthy, J., Kelly, P., Hurley, G., Luo, F., Chen, K., O'Sullivan, G.C., Kiely, B., Collins, J.K., Shanahan, F., et al. (2005). *Lactobacillus* and *bifidobacterium* in irritable bowel syndrome: symptom responses and relationship to cytokine profiles. *Gastroenterology* *128*, 541–551.
- O'Neill, I., Schofield, Z., and Hall, L.J. (2017). Exploring the role of the microbiota member *Bifidobacterium* in modulating immune-linked diseases. *Emerg. Top. Life Sci.* *1*, 333–349.
- O'Toole, P.W., Marchesi, J.R., and Hill, C. (2017). Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. *Nat. Microbiol.* *2*, 17057.
- Ouwehand, A.C., Isolauri, E., He, F., Hashimoto, H., Benno, Y., and Salminen, S. (2001). Differences in *Bifidobacterium* flora composition in allergic and healthy infants. *Journal of Allergy and Clinical Immunology* *108*, 144–145.
- Pals, P., Lincoln, S., Manning, J., Heckman, M., Skipper, L., Hulihan, M., Van den Broeck, M., De Pooter, T., Cras, P., Crook, J., et al. (2004). alpha-Synuclein promoter confers susceptibility to Parkinson's disease. *Ann. Neurol.* *56*, 591–595.
- Paneth, J. (1887). Ueber die secernirenden Zellen des Dünndarm-Epithels. *Archiv f. Mikrosk. Anatomie* *31*, 113–191.
- Papin, J.A., Hunter, T., Palsson, B.O., and Subramaniam, S. (2005). Reconstruction of cellular signalling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.* *6*, 99–111.
- Paraskevopoulou, M.D., and Hatzigeorgiou, A.G. (2016). Analyzing MiRNA-LncRNA Interactions. *Methods Mol. Biol.* *1402*, 271–286.
- Paraskevopoulou, M.D., Vlachos, I.S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., Zagganas, K., Tsanakas, P., Floros, E., Dalamagas, T., et al. (2016). DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.* *44*, D231-8.
- Parikh, K., Antanaviciute, A., Fawcner-Corbett, D., Jagielowicz, M., Aulicino, A., Lagerholm, C., Davis, S., Kinchen, J., Chen, H.H., Alham, N.K., et al. (2019). Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature* *567*, 49–55.

- Park, J., and Wang, H.H. (2018). Systematic and synthetic approaches to rewire regulatory networks. *Current Opinion in Systems Biology* 8, 90–96.
- Parkes, G.C., Rayment, N.B., Hudspith, B.N., Petrovska, L., Lomer, M.C., Brostoff, J., Whelan, K., and Sanderson, J.D. (2012). Distinct microbial populations exist in the mucosa-associated microbiota of sub-groups of irritable bowel syndrome. *Neurogastroenterol. Motil.* 24, 31–39.
- Patel, K.K., Miyoshi, H., Beatty, W.L., Head, R.D., Malvin, N.P., Cadwell, K., Guan, J.-L., Saitoh, T., Akira, S., Seglen, P.O., et al. (2013). Autophagy proteins control goblet cell function by potentiating reactive oxygen species production. *EMBO J.* 32, 3130–3144.
- Paull, E.O., Carlin, D.E., Niepel, M., Sorger, P.K., Haussler, D., and Stuart, J.M. (2013). Discovering causal pathways linking genomic events to transcriptional states using Tied Diffusion Through Interacting Events (TieDIE). *Bioinformatics* 29, 2757–2764.
- Paulus, U., Loeffler, M., Zeidler, J., Owen, G., and Potten, C.S. (1993). The differentiation and lineage development of goblet cells in the murine small intestinal crypt: experimental and modelling studies. *J. Cell Sci.* 106 (Pt 2), 473–483.
- Peck, B.C.E., Mah, A.T., Pitman, W.A., Ding, S., Lund, P.K., and Sethupathy, P. (2017). Functional transcriptomics in diverse intestinal epithelial cell types reveals robust microRNA sensitivity in intestinal stem cells to microbial status. *J. Biol. Chem.* 292, 2586–2600.
- Pei, F.H., Wang, Y.J., Gao, S.L., Liu, B.R., Du, Y.J., Liu, W., Yu, H.Y., Zhao, L.X., and Chi, B.R. (2011). Vitamin D receptor gene polymorphism and ulcerative colitis susceptibility in Han Chinese. *J Dig Dis* 12, 90–98.
- Perfetto, L., Briganti, L., Calderone, A., Cerquone Perpetuini, A., Iannuccelli, M., Langone, F., Licata, L., Marinkovic, M., Mattioni, A., Pavlidou, T., et al. (2016). SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.* 44, D548-54.
- Peterson, L.W., and Artis, D. (2014). Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat. Rev. Immunol.* 14, 141–153.
- Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098.
- Pickert, G., Neufert, C., Leppkes, M., Zheng, Y., Wittkopf, N., Warntjen, M., Lehr, H.-A., Hirth, S., Weigmann, B., Wirtz, S., et al. (2009). STAT3 links IL-22 signaling in intestinal epithelial cells to mucosal wound healing. *J. Exp. Med.* 206, 1465–1472.
- Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* 14, 687–690.
- Pinto-Sánchez, M.I., Smecuol, E.C., Temprano, M.P., Sugai, E., González, A., Moreno, M.L., Huang, X., Bercik, P., Cabanne, A., Vázquez, H., et al. (2017). Bifidobacterium infantis NLS Super Strain Reduces the Expression of α -Defensin-5, a Marker of Innate Immunity, in the Mucosa of Active Celiac Disease Patients. *J. Clin. Gastroenterol.* 51, 814–817.
- de Poel, E., Lefferts, J.W., and Beekman, J.M. (2020). Intestinal organoids for Cystic Fibrosis research. *J Cyst Fibros* 19 Suppl 1, S60–S64.
- Potten, C.S., Kovacs, L., and Hamilton, E. (1974). Continuous labelling studies on mouse skin and intestine. *Cell Tissue Kinet.* 7, 271–283.
- Powell, A.E., Wang, Y., Li, Y., Poulin, E.J., Means, A.L., Washington, M.K., Higginbotham, J.N., Juchheim, A., Prasad, N., Levy, S.E., et al. (2012). The pan-ErbB negative regulator Lrig1 is an

intestinal stem cell marker that functions as a tumor suppressor. *Cell* *149*, 146–158.

Prantera, C., and Marconi, S. (2013). Glucocorticosteroids in the treatment of inflammatory bowel disease and approaches to minimizing systemic activity. *Therap. Adv. Gastroenterol.* *6*, 137–156.

Pullan, R.D., Thomas, G.A., Rhodes, M., Newcombe, R.G., Williams, G.T., Allen, A., and Rhodes, J. (1994). Thickness of adherent mucus gel on colonic mucosa in humans and its relevance to colitis. *Gut* *35*, 353–359.

Püngel, D., Treveil, A., Dalby, M.J., Caim, S., Colquhoun, I.J., Booth, C., Ketskemety, J., Korcsmaros, T., van Sinderen, D., Lawson, M.A., et al. (2020). Bifidobacterium breve UCC2003 Exopolysaccharide Modulates the Early Life Microbiota by Acting as a Potential Dietary Substrate. *Nutrients* *12*.

Punternvoll, P., Linding, R., Gemünd, C., Chabanis-Davidson, S., Mattingsdal, M., Cameron, S., Martin, D.M.A., Ausiello, G., Brannetti, B., Costantini, A., et al. (2003). ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* *31*, 3625–3630.

Qi, Z., Li, Y., Zhao, B., Xu, C., Liu, Y., Li, H., Zhang, B., Wang, X., Yang, X., Xie, W., et al. (2017). BMP restricts stemness of intestinal Lgr5+ stem cells by directly suppressing their signature genes. *Nat. Commun.* *8*, 13824.

Qin, X., Sufi, J., Vlckova, P., Kyriakidou, P., Acton, S.E., Li, V.S.W., Nitz, M., and Tape, C.J. (2020). Cell-type-specific signaling networks in heterocellular organoids. *Nat. Methods* *17*, 335–342.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* *41*, D590-6.

Quin, C., Estaki, M., Vollman, D.M., Barnett, J.A., Gill, S.K., and Gibson, D.L. (2018). Probiotic supplementation and associated infant gut microbiome and health: a cautionary retrospective clinical comparison. *Sci. Rep.* *8*, 8283.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841–842.

R, C.T. (2018). R: A language and environment for statistical computing. (Vienna, Austria: R Foundation for Statistical Computing).

Raghavachari, B., Tasneem, A., Przytycka, T.M., and Jothi, R. (2008). DOMINE: a database of protein domain interactions. *Nucleic Acids Res.* *36*, D656-61.

Rawla, P., Sunkara, T., and Raj, J.P. (2018). Role of biologics and biosimilars in inflammatory bowel disease: current trends and future perspectives. *J Inflamm Res* *11*, 215–226.

Real, P.J., Tosello, V., Palomero, T., Castillo, M., Hernando, E., de Stanchina, E., Sulis, M.L., Barnes, K., Sawai, C., Homminga, I., et al. (2009). Gamma-secretase inhibitors reverse glucocorticoid resistance in T cell acute lymphoblastic leukemia. *Nat. Med.* *15*, 50–58.

Reinisch, W., Panés, J., Khurana, S., Toth, G., Hua, F., Comer, G.M., Hinz, M., Page, K., O’Toole, M., Moorehead, T.M., et al. (2015). Anrukizumab, an anti-interleukin 13 monoclonal antibody, in active UC: efficacy and safety from a phase IIa randomised multicentre study. *Gut* *64*, 894–900.

Richardson, G.M., Lannigan, J., and Macara, I.G. (2015). Does FACS perturb gene expression? *Cytometry A* *87*, 166–175.

- Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., et al. (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat. Genet.* *39*, 596–604.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* *26*, 139–140.
- Roda, G., Jharap, B., Neeraj, N., and Colombel, J.-F. (2016). Loss of Response to Anti-TNFs: Definition, Epidemiology, and Management. *Clin Transl Gastroenterol* *7*, e135.
- Rodríguez-Colman, M.J., Schewe, M., Meerlo, M., Stigter, E., Gerrits, J., Pras-Raves, M., Sacchetti, A., Hornsveld, M., Oost, K.C., Snippert, H.J., et al. (2017). Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature* *543*, 424–427.
- Rodríguez-García, M., Oliva, H., Climent, N., Escribese, M.M., García, F., Moran, T.M., Gatell, J.M., and Gallart, T. (2009). Impact of alpha-defensins1-3 on the maturation and differentiation of human monocyte-derived DCs. Concentration-dependent opposite dual effects. *Clin. Immunol.* *131*, 374–384.
- Roth, S., Franken, P., Sacchetti, A., Kremer, A., Anderson, K., Sansom, O., and Fodde, R. (2012). Paneth cells in intestinal homeostasis and tissue injury. *PLoS One* *7*, e38965.
- Round, J.L., Lee, S.M., Li, J., Tran, G., Jabri, B., Chatila, T.A., and Mazmanian, S.K. (2011). The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* *332*, 974–977.
- Rueda, A., Barturen, G., Lebrón, R., Gómez-Martín, C., Alganza, Á., Oliver, J.L., and Hackenberg, M. (2015). sRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic Acids Res.* *43*, W467-73.
- Ruiz, L., Motherway, M.O., Lanigan, N., and van Sinderen, D. (2013). Transposon mutagenesis in *Bifidobacterium breve*: construction and characterization of a Tn5 transposon mutant library for *Bifidobacterium breve* UCC2003. *PLoS One* *8*, e64699.
- Ruiz, L., Delgado, S., Ruas-Madiedo, P., Sánchez, B., and Margolles, A. (2017). *Bifidobacteria and Their Molecular Communication with the Immune System*. *Front. Microbiol.* *8*, 2345.
- Russell, L., and Garrett-Sinha, L.A. (2010). Transcription factor Ets-1 in cytokine and chemokine gene regulation. *Cytokine* *51*, 217–226.
- Russell, D.A., Ross, R.P., Fitzgerald, G.F., and Stanton, C. (2011). Metabolic activities and probiotic potential of bifidobacteria. *Int. J. Food Microbiol.* *149*, 88–105.
- Rutgeerts, P. (1998). The use of oral topically acting glucocorticosteroids in the treatment of inflammatory bowel disease. *Mediators Inflamm.* *7*, 137–140.
- Sakurai, T., Hashikura, N., Minami, J., Yamada, A., Odamaki, T., and Xiao, J.-Z. (2017). Tolerance mechanisms of human-residential bifidobacteria against lysozyme. *Anaerobe* *47*, 104–110.
- Sancho, R., Nateri, A.S., de Vinuesa, A.G., Aguilera, C., Nye, E., Spencer-Dene, B., and Behrens, A. (2009). JNK signalling modulates intestinal homeostasis and tumourigenesis in mice. *EMBO J.* *28*, 1843–1854.
- Sands, B.E., Sandborn, W.J., Panaccione, R., O'Brien, C.D., Zhang, H., Johanns, J., Adedokun, O.J., Li, K., Peyrin-Biroulet, L., Van Assche, G., et al. (2019). Ustekinumab as induction and maintenance therapy for ulcerative colitis. *N. Engl. J. Med.* *381*, 1201–1214.
- Sangiorgi, E., and Capecchi, M.R. (2008). Bmi1 is expressed in vivo in intestinal stem cells. *Nat.*

Genet. *40*, 915–920.

Sarkar, A., and Mandal, S. (2016). Bifidobacteria-Insight into clinical outcomes and mechanisms of its probiotic action. *Microbiol Res* *192*, 159–171.

Sato, T., and Clevers, H. (2013). Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. *Science* *340*, 1190–1194.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., et al. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* *459*, 262–265.

Sato, T., van Es, J.H., Snippert, H.J., Stange, D.E., Vries, R.G., van den Born, M., Barker, N., Shroyer, N.F., van de Wetering, M., and Clevers, H. (2011). Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* *469*, 415–418.

Schaefer, M.H., Serrano, L., and Andrade-Navarro, M.A. (2015). Correcting for the study bias associated with protein-protein interaction measurements reveals differences between protein degree distributions from different cancer types. *Front. Genet.* *6*, 260.

Schiavi, E., Gleinser, M., Molloy, E., Groeger, D., Frei, R., Ferstl, R., Rodriguez-Perez, N., Ziegler, M., Grant, R., Moriarty, T.F., et al. (2016). The Surface-Associated Exopolysaccharide of *Bifidobacterium longum* 35624 Plays an Essential Role in Dampening Host Proinflammatory Responses and Repressing Local TH17 Responses. *Appl. Environ. Microbiol.* *82*, 7185–7196.

Schreiber, S., Nikolaus, S., and Hampe, J. (1998). Activation of nuclear factor kappa B inflammatory bowel disease. *Gut* *42*, 477–484.

Schroeder, B.O. (2019). Fight them or feed them: how the intestinal mucus layer manages the gut microbiota. *Gastroenterol Rep (Oxf)* *7*, 3–12.

Schroeder, B.O., Birchenough, G.M.H., Ståhlman, M., Arike, L., Johansson, M.E.V., Hansson, G.C., and Bäckhed, F. (2018). Bifidobacteria or Fiber Protects against Diet-Induced Microbiota-Mediated Colonic Mucus Deterioration. *Cell Host Microbe* *23*, 27–40.e7.

Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M.J., Blüthgen, N., and Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat. Commun.* *9*, 20.

Schuijers, J., Junker, J.P., Mokry, M., Hatzis, P., Koo, B.-K., Sasselli, V., van der Flier, L.G., Cuppen, E., van Oudenaarden, A., and Clevers, H. (2015). Ascl2 acts as an R-spondin/Wnt-responsive switch to control stemness in intestinal crypts. *Cell Stem Cell* *16*, 158–170.

Scoville, E.A., Allaman, M.M., Brown, C.T., Motley, A.K., Horst, S.N., Williams, C.S., Koyama, T., Zhao, Z., Adams, D.W., Beaulieu, D.B., et al. (2018). Alterations in Lipid, Amino Acid, and Energy Metabolism Distinguish Crohn's Disease from Ulcerative Colitis and Control Subjects by Serum Metabolomic Profiling. *Metabolomics* *14*, 17.

Scutto, A., Djorie, S., Colavizza, M., Romond, P.-C., and Romond, M.-B. (2014). *Bifidobacterium breve* C50 secretes lipoprotein with CHAP domain recognized in aggregated form by TLR2. *Biochimie* *107 Pt B*, 367–375.

See, P., Lum, J., Chen, J., and Ginhoux, F. (2018). A Single-Cell Sequencing Guide for Immunologists. *Front. Immunol.* *9*, 2425.

Seyedian, S.S., Nokhostin, F., and Malamir, M.D. (2019). A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. *J. Med. Life* *12*, 113–122.

Shah, P., Fritz, J.V., Glaab, E., Desai, M.S., Greenhalgh, K., Frchet, A., Niegowska, M., Estes, M.,

- Jäger, C., Seguin-Devaux, C., et al. (2016). A microfluidics-based in vitro model of the gastrointestinal human-microbe interface. *Nat. Commun.* *7*, 11535.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* *13*, 2498–2504.
- Shirazi, T., Longman, R.J., Corfield, A.P., and Probert, C.S. (2000). Mucins and inflammatory bowel disease. *Postgrad. Med. J.* *76*, 473–478.
- Siahpirani, A.F., and Roy, S. (2017). A prior-based integrative framework for functional transcriptional regulatory network inference. *Nucleic Acids Res.* *45*, e21.
- Sicard, J.-F., Le Bihan, G., Vogeleer, P., Jacques, M., and Harel, J. (2017). Interactions of Intestinal Bacteria with Components of the Intestinal Mucus. *Front. Cell Infect. Microbiol.* *7*, 387.
- Sigrist, C.J.A., de Castro, E., Cerutti, L., Cucho, B.A., Hulo, N., Bridge, A., Bougueleret, L., and Xenarios, I. (2013). New and continuing developments at PROSITE. *Nucleic Acids Res.* *41*, D344–7.
- Silva, A.M., Barbosa, F.H.F., Duarte, R., Vieira, L.Q., Arantes, R.M.E., and Nicoli, J.R. (2004). Effect of *Bifidobacterium longum* ingestion on experimental salmonellosis in mice. *J. Appl. Microbiol.* *97*, 29–37.
- Sivan, A., Corrales, L., Hubert, N., Williams, J.B., Aquino-Michaels, K., Earley, Z.M., Benyamin, F.W., Lei, Y.M., Jabri, B., Alegre, M.-L., et al. (2015). Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* *350*, 1084–1089.
- Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M.H., Baldock, R., Barbiera, G., et al. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* *43*, W589–98.
- Smillie, C.S., Biton, M., Ordovas-Montanes, J., Sullivan, K.M., Burgin, G., Graham, D.B., Herbst, R.H., Rogel, N., Slyper, M., Waldman, J., et al. (2019). Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* *178*, 714–730.e22.
- Smith, M.K., Pai, J., Panaccione, R., Beck, P., Ferraz, J.G., and Jijon, H. (2019). Crohn's-like disease in a patient exposed to anti-Interleukin-17 blockade (Ixekizumab) for the treatment of chronic plaque psoriasis: a case report. *BMC Gastroenterol.* *19*, 162.
- Smith, P.D., Smythies, L.E., Shen, R., Greenwell-Wild, T., Gliozzi, M., and Wahl, S.M. (2011). Intestinal macrophages and response to microbial encroachment. *Mucosal Immunol.* *4*, 31–42.
- Smythies, L.E., Sellers, M., Clements, R.H., Mosteller-Barnum, M., Meng, G., Benjamin, W.H., Orenstein, J.M., and Smith, P.D. (2005). Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J. Clin. Invest.* *115*, 66–75.
- Smythies, L.E., Shen, R., Bimczok, D., Novak, L., Clements, R.H., Eckhoff, D.E., Bouchard, P., George, M.D., Hu, W.K., Dandekar, S., et al. (2010). Inflammation anergy in human intestinal macrophages is due to Smad-induced I κ B α expression and NF- κ B inactivation. *J. Biol. Chem.* *285*, 19593–19604.
- Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., et al. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* *143*, 134–144.

- Soderholm, A.T., and Pedicord, V.A. (2019). Intestinal epithelial cells: at the interface of the microbiota and mucosal immunity. *Immunology* *158*, 267–280.
- Sodhi, C.P., Neal, M.D., Siggers, R., Sho, S., Ma, C., Branca, M.F., Prindle, T., Russo, A.M., Afrazi, A., Good, M., et al. (2012). Intestinal epithelial Toll-like receptor 4 regulates goblet cell development and is required for necrotizing enterocolitis in mice. *Gastroenterology* *143*, 708–18.e1.
- Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. [version 2; peer review: 2 approved]. *F1000Res.* *4*, 1521.
- Sonnhammer, E.L.L., and Östlund, G. (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. *Nucleic Acids Res.* *43*, D234–9.
- Sperandio, B., Fischer, N., Joncquel Chevalier-Curt, M., Rossez, Y., Roux, P., Robbe Masselot, C., and Sansonetti, P.J. (2013). Virulent *Shigella flexneri* affects secretion, expression, and glycosylation of gel-forming mucins in mucus-producing cells. *Infect. Immun.* *81*, 3632–3643.
- Sprenger, J., Lynn Fink, J., Karunaratne, S., Hanson, K., Hamilton, N.A., and Teasdale, R.D. (2008). LOCATE: a mammalian protein subcellular localization database. *Nucleic Acids Res.* *36*, D230–3.
- Srutkova, D., Schwarzer, M., Hudcovic, T., Zakostelska, Z., Drab, V., Spanova, A., Rittich, B., Kozakova, H., and Schabussova, I. (2015). *Bifidobacterium longum* CCM 7952 Promotes Epithelial Barrier Function and Prevents Acute DSS-Induced Colitis in Strictly Strain-Specific Manner. *PLoS One* *10*, e0134050.
- Stappenbeck, T.S. (2010). The role of autophagy in Paneth cell differentiation and secretion. *Mucosal Immunol.* *3*, 8–10.
- Stappenbeck, T.S., and McGovern, D.P.B. (2017). Paneth Cell Alterations in the Development and Phenotype of Crohn’s Disease. *Gastroenterology* *152*, 322–326.
- Steed, H., Macfarlane, G.T., Blackett, K.L., Bahrami, B., Reynolds, N., Walsh, S.V., Cummings, J.H., and Macfarlane, S. (2010). Clinical trial: the microbiological and immunological effects of synbiotic consumption - a randomized double-blind placebo-controlled study in active Crohn’s disease. *Aliment. Pharmacol. Ther.* *32*, 872–883.
- Stewart, C.J., Ajami, N.J., O’Brien, J.L., Hutchinson, D.S., Smith, D.P., Wong, M.C., Ross, M.C., Lloyd, R.E., Doddapaneni, H., Metcalf, G.A., et al. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* *562*, 583–588.
- Stratiki, Z., Costalos, C., Sevastiadou, S., Kastanidou, O., Skouroliakou, M., Giakoumatou, A., and Petrohilou, V. (2007). The effect of a bifidobacter supplemented bovine milk on intestinal permeability of preterm infants. *Early Hum. Dev.* *83*, 575–579.
- Stronkhorst, A., Radema, S., Yong, S.L., Bijl, H., ten Berge, I.J., Tytgat, G.N., and van Deventer, S.J. (1997). CD4 antibody treatment in patients with active Crohn’s disease: a phase 1 dose finding study. *Gut* *40*, 320–327.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell* *177*, 1888–1902.e21.
- Stutte, S., Jux, B., Esser, C., and Förster, I. (2008). CD24a expression levels discriminate Langerhans cells from dermal dendritic cells in murine skin and lymph nodes. *J. Invest. Dermatol.* *128*, 1470–1475.

- Su, G., Morris, J.H., Demchak, B., and Bader, G.D. (2014). Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics* 47, 8.13.1-24.
- Su, L., Nalle, S.C., Shen, L., Turner, E.S., Singh, G., Breskin, L.A., Khramtsova, E.A., Khramtsova, G., Tsai, P.-Y., Fu, Y.-X., et al. (2013). TNFR2 activates MLCK-dependent tight junction dysregulation to cause apoptosis-mediated barrier loss and experimental colitis. *Gastroenterology* 145, 407–415.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Sudhakar, P., Jacomin, A.-C., Hautefort, I., Samavedam, S., Fatemian, K., Ari, E., Gul, L., Demeter, A., Jones, E., Korcsmaros, T., et al. (2019). Targeted interplay between bacterial pathogens and host autophagy. *Autophagy* 15, 1620–1633.
- Swidsinski, A., Loening-Baucke, V., Theissig, F., Engelhardt, H., Bengmark, S., Koch, S., Lochs, H., and Dörffel, Y. (2007). Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut* 56, 343–350.
- Szajewska, H., Guandalini, S., Morelli, L., Van Goudoever, J.B., and Walker, A. (2010). Effect of *Bifidobacterium animalis* subsp *lactis* supplementation in preterm infants: a systematic review of randomized controlled trials. *J. Pediatr. Gastroenterol. Nutr.* 51, 203–209.
- Takeda, N., Jain, R., LeBoeuf, M.R., Wang, Q., Lu, M.M., and Epstein, J.A. (2011). Interconversion between intestinal stem cell populations in distinct niches. *Science* 334, 1420–1424.
- Tang, C., and Iwakura, Y. (2012). IL-23 in colitis: targeting the progenitors. *Immunity* 37, 957–959.
- Tang, X., Huang, Y., Lei, J., Luo, H., and Zhu, X. (2019). The single-cell sequencing: new developments and medical applications. *Cell Biosci.* 9, 53.
- Tau, G., and Rothman, P. (1999). Biologic functions of the IFN-gamma receptors. *Allergy* 54, 1233–1251.
- Telesco, S.E., Brodmerkel, C., Zhang, H., Kim, L.L.-L., Johanns, J., Mazumder, A., Li, K., Baribaud, F., Curran, M., Strauss, R., et al. (2018). Gene Expression Signature for Prediction of Golimumab Response in a Phase 2a Open-Label Trial of Patients With Ulcerative Colitis. *Gastroenterology* 155, 1008–1011.e8.
- Tetteh, P.W., Basak, O., Farin, H.F., Wiebrands, K., Kretzschmar, K., Begthel, H., van den Born, M., Korving, J., de Sauvage, F., van Es, J.H., et al. (2016). Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. *Cell Stem Cell* 18, 203–213.
- The Gene Ontology Consortium (2017). Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* 45, D331–D338.
- Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 13, 2129–2141.
- Thomas, P.D., Kejariwal, A., Guo, N., Mi, H., Campbell, M.J., Muruganujan, A., and Lazareva-Ulitsky, B. (2006). Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Res.* 34, W645-50.
- Thorne, C.A., Chen, I.W., Sanman, L.E., Cobb, M.H., Wu, L.F., and Altschuler, S.J. (2018). Enteroid

- monolayers reveal an autonomous WNT and BMP circuit controlling intestinal epithelial growth and organization. *Dev. Cell* *44*, 624–633.e4.
- Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. *Science* *356*.
- Tian, H., Biehs, B., Warming, S., Leong, K.G., Rangell, L., Klein, O.D., and de Sauvage, F.J. (2011). A reserve stem cell population in small intestine renders Lgr5-positive cells dispensable. *Nature* *478*, 255–259.
- Tissier, H. (1899). Le Bacterium coli et la Reaction chromophile d'Escherich. *C R Soc Biol* 943–945.
- Tissier, H. (1900). Recherchers sur la flora intestinale normale et pathologique du nourisson. Doctoral dissertation.
- Tojo, R., Suárez, A., Clemente, M.G., de los Reyes-Gavilán, C.G., Margolles, A., Gueimonde, M., and Ruas-Madiedo, P. (2014). Intestinal microbiota in health and disease: role of bifidobacteria in gut homeostasis. *World J. Gastroenterol.* *20*, 15163–15176.
- Tomosada, Y., Villena, J., Murata, K., Chiba, E., Shimazu, T., Aso, H., Iwabuchi, N., Xiao, J., Saito, T., and Kitazawa, H. (2013). Immunoregulatory effect of bifidobacteria strains in porcine intestinal epithelial cells through modulation of ubiquitin-editing enzyme A20 expression. *PLoS One* *8*, e59259.
- Tootle, T.L., and Rebay, I. (2005). Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays* *27*, 285–298.
- Treveil, A., Sudhakar, P., Matthews, Z.J., Wrzesiński, T., Jones, E.J., Brooks, J., Ölbei, M., Hautefort, I., Hall, L.J., Carding, S.R., et al. (2020). Regulatory network analysis of Paneth cell and goblet cell enriched gut organoids using transcriptomics approaches. *Mol. Omics* *16*, 39–58.
- Triner, D., Castillo, C., Hakim, J.B., Xue, X., Greenson, J.K., Nuñez, G., Chen, G.Y., Colacino, J.A., and Shah, Y.M. (2018). Myc-Associated Zinc Finger Protein Regulates the Proinflammatory Response in Colitis and Colon Cancer via STAT3 Signaling. *Mol. Cell. Biol.* *38*.
- Tripathi, S., Moutari, S., Dehmer, M., and Emmert-Streib, F. (2016). Comparison of module detection algorithms in protein networks and investigation of the biological meaning of predicted modules. *BMC Bioinformatics* *17*, 129.
- Türei, D., Földvári-Nagy, L., Fazekas, D., Módos, D., Kubisch, J., Kadlecsek, T., Demeter, A., Lenti, K., Csermely, P., Vellai, T., et al. (2015). Autophagy Regulatory Network - a systems-level bioinformatics resource for studying the mechanism and regulation of autophagy. *Autophagy* *11*, 155–165.
- Türei, D., Korcsmáros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* *13*, 966–967.
- Turei, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Ivanova, O., Gabor, A., Modos, D., Korcsmaros, T., and Saez-Rodriguez, J. (2020). Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *BioRxiv*.
- Turróni, F., van Sinderen, D., and Ventura, M. (2011). Genomics and ecological overview of the genus *Bifidobacterium*. *Int. J. Food Microbiol.* *149*, 37–44.
- Turróni, F., Duranti, S., Milani, C., Lugli, G.A., van Sinderen, D., and Ventura, M. (2019). *Bifidobacterium bifidum*: A Key Member of the Early Human Gut Microbiota. *Microorganisms* *7*.
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human

proteome. *Science* **347**, 1260419.

Underwood, M.A., Kananurak, A., Coursodon, C.F., Adkins-Reick, C.K., Chu, H., Bennett, S.H., Wehkamp, J., Castillo, P.A., Leonard, B.C., Tancredi, D.J., et al. (2012). Bifidobacterium bifidum in a rat model of necrotizing enterocolitis: antimicrobial peptide and protein responses. *Pediatr. Res.* **71**, 546–551.

UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515.

Vainer, B., Nielsen, O.H., Hendel, J., Horn, T., and Kirman, I. (2000). Colonic expression and synthesis of interleukin 13 and interleukin 15 in inflammatory bowel disease. *Cytokine* **12**, 1531–1536.

Van der Sluis, M., De Koning, B.A.E., De Bruijn, A.C.J.M., Velcich, A., Meijerink, J.P.P., Van Goudoever, J.B., Büller, H.A., Dekker, J., Van Seuningen, I., Renes, I.B., et al. (2006). Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131**, 117–129.

Vancamelbeke, M., and Vermeire, S. (2017). The intestinal barrier: a fundamental role in health and disease. *Expert Rev Gastroenterol Hepatol* **11**, 821–834.

Vandin, F., Clay, P., Upfal, E., and Raphael, B.J. (2012). Discovery of mutated subnetworks associated with clinical data in cancer. *Pac Symp Biocomput* 55–66.

Vanuytsel, T., Senger, S., Fasano, A., and Shea-Donohue, T. (2013). Major signaling pathways in intestinal stem cells. *Biochim. Biophys. Acta* **1830**, 2410–2426.

Vazquez-Gutierrez, P., Stevens, M.J.A., Gehrig, P., Barkow-Oesterreicher, S., Lacroix, C., and Chassard, C. (2017). The extracellular proteome of two Bifidobacterium species reveals different adaptation strategies to low iron conditions. *BMC Genomics* **18**, 41.

Veres, D.V., Gyurkó, D.M., Thaler, B., Szalay, K.Z., Fazekas, D., Korcsmáros, T., and Csermely, P. (2015). ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.* **43**, D485–93.

Vijesh, N., Chakrabarti, S.K., and Sreekumar, J. (2013). Modeling of gene regulatory networks: A review. *JBiSE* **06**, 223–231.

Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos, I.-L., Maniou, S., Karathanou, K., Kalfakakou, D., et al. (2015). DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.* **43**, D153–9.

Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232.

Voss, E., Wehkamp, J., Wehkamp, K., Stange, E.F., Schröder, J.M., and Harder, J. (2006). NOD2/CARD15 mediates induction of the antimicrobial peptide human beta-defensin-2. *J. Biol. Chem.* **281**, 2005–2011.

Wachi, S., Yoneda, K., and Wu, R. (2005). Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics* **21**, 4205–4208.

Wagner, G.P., and Zhang, J. (2011). The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12**, 204–213.

Wang, Y.-D. (2007). Expression of matrix metalloproteinase-1 and tumor necrosis factor- α in ulcerative colitis. *WJG* **13**, 5926.

- Wang, C., Baer, H.M., Gaya, D.R., Nibbs, R.J.B., and Milling, S. (2019a). Can molecular stratification improve the treatment of inflammatory bowel disease? *Pharmacol. Res.* *148*, 104442.
- Wang, F., Scoville, D., He, X.C., Mahe, M.M., Box, A., Perry, J.M., Smith, N.R., Lei, N.Y., Davies, P.S., Fuller, M.K., et al. (2013a). Isolation and characterization of intestinal stem cells based on surface marker combinations and colony-formation assay. *Gastroenterology* *145*, 383–95.e1.
- Wang, J., Lu, M., Qiu, C., and Cui, Q. (2010). TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Res.* *38*, D119–22.
- Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013b). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* *41*, e74.
- Wang, S., Ye, Q., Zeng, X., and Qiao, S. (2019b). Functions of macrophages in the maintenance of intestinal homeostasis. *J. Immunol. Res.* *2019*, 1512969.
- Wang, T.-T., Nestel, F.P., Bourdeau, V., Nagai, Y., Wang, Q., Liao, J., Tavera-Mendoza, L., Lin, R., Hanrahan, J.W., Mader, S., et al. (2004). Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression. *J. Immunol.* *173*, 2909–2912.
- Wang, W., Chen, L., Zhou, R., Wang, X., Song, L., Huang, S., Wang, G., and Xia, B. (2014). Increased proportions of Bifidobacterium and the Lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. *J. Clin. Microbiol.* *52*, 398–406.
- Wehkamp, J., Harder, J., Weichenthal, M., Mueller, O., Herrlinger, K.R., Fellermann, K., Schroeder, J.M., and Stange, E.F. (2003). Inducible and constitutive beta-defensins are differentially expressed in Crohn's disease and ulcerative colitis. *Inflamm. Bowel Dis.* *9*, 215–223.
- Wehkamp, J., Salzman, N.H., Porter, E., Nuding, S., Weichenthal, M., Petras, R.E., Shen, B., Schaeffeler, E., Schwab, M., Linzmeier, R., et al. (2005). Reduced Paneth cell alpha-defensins in ileal Crohn's disease. *Proc. Natl. Acad. Sci. USA* *102*, 18129–18134.
- Wells, J.M., Rossi, O., Meijerink, M., and van Baarlen, P. (2011). Epithelial crosstalk at the microbiota-mucosal interface. *Proc. Natl. Acad. Sci. USA* *108 Suppl 1*, 4607–4614.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. (New York: Springer-Verlag).
- van Wijk, F., and Cheroutre, H. (2010). Mucosal T cells in gut homeostasis and inflammation. *Expert Rev Clin Immunol* *6*, 559–566.
- Williams, J.M., Duckworth, C.A., Burkitt, M.D., Watson, A.J.M., Campbell, B.J., and Pritchard, D.M. (2015). Epithelial cell shedding and barrier function: a matter of life and death at the small intestinal villus tip. *Vet. Pathol.* *52*, 445–455.
- Wilson, M.S., Ramalingam, T.R., Rivollier, A., Shenderov, K., Mentink-Kane, M.M., Madala, S.K., Cheever, A.W., Artis, D., Kelsall, B.L., and Wynn, T.A. (2011). Colitis and intestinal inflammation in IL10^{-/-} mice results from IL-13R α 2-mediated attenuation of IL-13 activity. *Gastroenterology* *140*, 254–264.
- Wilson, S.S., Tocchi, A., Holly, M.K., Parks, W.C., and Smith, J.G. (2015). A small intestinal organoid model of non-invasive enteric pathogen-epithelial cell interactions. *Mucosal Immunol.* *8*, 352–361.
- Winterbach, W., Van Mieghem, P., Reinders, M., Wang, H., and de Ridder, D. (2013). Topology of molecular interaction networks. *BMC Syst. Biol.* *7*, 90.
- Wittkopf, N., Neurath, M.F., and Becker, C. (2014). Immune-epithelial crosstalk at the intestinal

surface. *J. Gastroenterol.* *49*, 375–387.

Wong, V.W.Y., Stange, D.E., Page, M.E., Buczacki, S., Wabik, A., Itami, S., van de Wetering, M., Poulosom, R., Wright, N.A., Trotter, M.W.B., et al. (2012). Lrig1 controls intestinal stem-cell homeostasis by negative regulation of ErbB signalling. *Nat. Cell Biol.* *14*, 401–408.

Worthington, J.J., Reimann, F., and Gribble, F.M. (2018). Enteroendocrine cells-sensory sentinels of the intestinal environment and orchestrators of mucosal immunity. *Mucosal Immunol.* *11*, 3–20.

Wrzosek, L., Miquel, S., Noordine, M.-L., Bouet, S., Joncquel Chevalier-Curt, M., Robert, V., Philippe, C., Bridonneau, C., Cherbuy, C., Robbe-Masselot, C., et al. (2013). *Bacteroides thetaiotaomicron* and *Faecalibacterium prausnitzii* influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. *BMC Biol.* *11*, 61.

Wu, S., Zhang, Y.-G., Lu, R., Xia, Y., Zhou, D., Petrof, E.O., Claud, E.C., Chen, D., Chang, E.B., Carmeliet, G., et al. (2015). Intestinal epithelial vitamin D receptor deletion leads to defective autophagy in colitis. *Gut* *64*, 1082–1094.

Xue, X., and Falcon, D.M. (2019). The role of immune cells and cytokines in intestinal wound healing. *Int. J. Mol. Sci.* *20*.

Yan, S., Yang, B., Zhao, J., Zhao, J., Stanton, C., Ross, R.P., Zhang, H., and Chen, W. (2019). A rropy exopolysaccharide producing strain *Bifidobacterium longum* subsp. *longum* YS108R alleviates DSS-induced colitis by maintenance of the mucosal barrier and gut microbiota modulation. *Food Funct* *10*, 1595–1608.

Yang, D., Chertov, O., Bykovskaia, S.N., Chen, Q., Buffo, M.J., Shogan, J., Anderson, M., Schröder, J.M., Wang, J.M., Howard, O.M., et al. (1999). Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* *286*, 525–528.

Yang, X., Gao, X.-C., Liu, J., and Ren, H.-Y. (2017). Effect of EPEC endotoxin and bifidobacteria on intestinal barrier function through modulation of toll-like receptor 2 and toll-like receptor 4 expression in intestinal epithelial cell-18. *World J. Gastroenterol.* *23*, 4744–4751.

Yellaboina, S., Tasneem, A., Zaykin, D.V., Raghavachari, B., and Jothi, R. (2011). DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.* *39*, D730-5.

Yemelyanov, A., Czwarnog, J., Chebotaev, D., Karseladze, A., Kulevitch, E., Yang, X., and Budunova, I. (2007). Tumor suppressor activity of glucocorticoid receptor in the prostate. *Oncogene* *26*, 1885–1896.

Yevshin, I., Sharipov, R., Valeev, T., Kel, A., and Kolpakov, F. (2017). GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. *Nucleic Acids Res.* *45*, D61–D67.

Yilmaz, Ö.H., Katajisto, P., Lamming, D.W., Gültekin, Y., Bauer-Rowe, K.E., Sengupta, S., Birsoy, K., Dursun, A., Yilmaz, V.O., Selig, M., et al. (2012). mTORC1 in the Paneth cell niche couples intestinal stem-cell function to calorie intake. *Nature* *486*, 490–495.

Yin, X., Farin, H.F., van Es, J.H., Clevers, H., Langer, R., and Karp, J.M. (2014). Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny. *Nat. Methods* *11*, 106–112.

Yoo, A.B., Jette, M.A., and Grondona, M. (2003). SLURM: simple linux utility for resource management. In *Job Scheduling Strategies for Parallel Processing*, D. Feitelson, L. Rudolph, and U. Schwiegelshohn, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 44–60.

- Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* *12*, 477–479.
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012a). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* *16*, 284–287.
- Yu, H., Luscombe, N.M., Qian, J., and Gerstein, M. (2003). Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet.* *19*, 422–427.
- Yu, L.C.-H., Wang, J.-T., Wei, S.-C., and Ni, Y.-H. (2012b). Host-microbial interactions and regulation of intestinal epithelial barrier function: From physiology to pathology. *World J Gastrointest Pathophysiol* *3*, 27–43.
- Yue, B., Luo, X., Yu, Z., Mani, S., Wang, Z., and Dou, W. (2019). Inflammatory Bowel Disease: A Potential Result from the Collusion between Gut Microbiota and Mucosal Immune System. *Microorganisms* *7*.
- Zachos, N.C., Kovbasnjuk, O., Foulke-Abel, J., In, J., Blutt, S.E., de Jonge, H.R., Estes, M.K., and Donowitz, M. (2016). Human enteroids/colonoids and intestinal organoids functionally recapitulate normal intestinal physiology and pathophysiology. *J. Biol. Chem.* *291*, 3759–3766.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* *46*, D754–D761.
- Zhai, Z., Wu, F., Chuang, A.Y., and Kwon, J.H. (2013). miR-106b fine tunes ATG16L1 expression and autophagic activity in intestinal epithelial HCT116 cells. *Inflamm. Bowel Dis.* *19*, 2295–2301.
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* *4*, Article17.
- Zhang, K., Hornef, M.W., and Dupont, A. (2015). The intestinal epithelium as guardian of gut barrier integrity. *Cell Microbiol.* *17*, 1561–1569.
- Zhang, M., Sun, K., Wu, Y., Yang, Y., Tso, P., and Wu, Z. (2017). Interactions between Intestinal Microbiota and Host Immune Response in Inflammatory Bowel Disease. *Front. Immunol.* *8*, 942.
- Zhang, R., Wang, L., Pan, J.-H., and Han, J. (2018). A critical role of E2F transcription factor 2 in proinflammatory cytokines-dependent proliferation and invasiveness of fibroblast-like synoviocytes in rheumatoid Arthritis. *Sci. Rep.* *8*, 2623.
- Zhang, Y.-G., Wu, S., Xia, Y., and Sun, J. (2014). Salmonella-infected crypt-derived intestinal organoid culture system for host-bacterial interactions. *Physiol. Rep.* *2*.
- Zhao, Y., Feng, Y., Liu, M., Chen, L., Meng, Q., Tang, X., Wang, S., Liu, L., Li, L., Shen, W., et al. (2020). Single-cell RNA sequencing analysis reveals alginate oligosaccharides preventing chemotherapy-induced mucositis. *Mucosal Immunol.* *13*, 437–448.
- Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* *8*, 14049.
- Zheng, X., Tsuchiya, K., Okamoto, R., Iwasaki, M., Kano, Y., Sakamoto, N., Nakamura, T., and Watanabe, M. (2011). Suppression of *hath1* gene expression directly regulated by *hes1* via notch signaling is associated with goblet cell depletion in ulcerative colitis. *Inflamm. Bowel Dis.* *17*, 2251–2260.
- Zheng, X., Chen, L., Li, X., Zhang, Y., Xu, S., and Huang, X. (2020). Prediction of miRNA targets by learning from interaction sequences. *PLoS One* *15*, e0232578.

Zhou, Y.-K., Shen, Z.-A., Yu, H., Luo, T., Gao, Y., and Du, P.-F. (2019). Predicting lncRNA-Protein Interactions With miRNAs as Mediators in a Heterogeneous Network Model. *Front. Genet.* *10*, 1341.

Zückert, W.R. (2014). Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim. Biophys. Acta* *1843*, 1509–1516.

Zuliani-Alvarez, L., Piccinini, A.M., and Midwood, K.S. (2017). Screening for Novel Endogenous Inflammatory Stimuli Using the Secreted Embryonic Alkaline Phosphatase NF- κ B Reporter Assay. *Bio Protoc* *7*.

Appendix 1: Supplementary data for Chapter 2

This appendix contains all the supplementary materials for Chapter 2.

- **File S2.1** - Differentially expressed genes from cell type enriched enteroids vs conventionally differentiated enteroids (electronic supplementary materials).
- **File S2.2** - Functional enrichment analysis of the top five most rewired (shared) marker regulators (electronic supplementary materials).
- **Table S2.1**- Differentially expressed antimicrobial peptide (AMP) and mucin related genes in Paneth cell enriched enteroids and goblet cell enriched enteroids (compared to conventionally/normally differentiated enteroids).
- **Table S2.2**- Hypergeometric significance testing of cell type-specific marker enrichment in upregulated differentially expressed gene lists.
- **Table S2.3** - Rewiring analysis results for the marker regulators present in the Paneth and the goblet subnetworks.
- **Table S2.4** - Crohn's disease SNP associated genes in the enriched enteroid regulatory networks.
- **Table S2.5** - Ulcerative colitis disease SNP associated genes in the enriched enteroid regulatory networks.

Gene Type	Gene Name	Ensembl ID	Paneth DEG LFC	Paneth DEG q val	Goblet DEG LFC	Goblet DEG q val
AMP	<i>Ang4</i>	ENSMUSG00000060615	4.64	8.61E-36	2.45	8.75E-13
AMP	<i>Defa17</i>	ENSMUSG00000060208	4.02	6.40E-46	2.95	5.56E-31
AMP	<i>Defa2</i>	ENSMUSG00000096295	4.58	1.70E-18	NA	NA
AMP	<i>Defa20</i>	ENSMUSG00000095066	4.40	1.19E-17	NA	NA
AMP	<i>Defa21</i>	ENSMUSG00000074447	5.18	2.22E-22	NA	NA
AMP	<i>Defa22</i>	ENSMUSG00000074443	5.40	3.09E-20	NA	NA
AMP	<i>Defa23</i>	ENSMUSG00000074446	3.69	1.94E-12	NA	NA

Table S2.1. Part 1/2. Differentially expressed antimicrobial peptide (AMP) and mucin related genes in Paneth cell enriched enteroids and goblet cell enriched enteroids (compared to conventionally/normally differentiated enteroids). Only genes which are differentially expressed ($\log_2\text{fc} \geq 1$ and $q \text{ value} \leq 0.05$) in at least one of the datasets was included. LFC - \log_2 fold change; q val - q value; DEG - differentially expressed gene; Paneth - Paneth enriched enteroid; goblet - goblet enriched enteroid. Table reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Gene Type	Gene Name	Ensembl ID	Paneth DEG LFC	Paneth DEG q val	Goblet DEG LFC	Goblet DEG q val
AMP	<i>Defa24</i>	ENSMUSG00000064213	3.91	8.51E-38	3.24	1.92E-32
AMP	<i>Defa26</i>	ENSMUSG00000060070	3.02	1.86E-39	2.22	8.61E-27
AMP	<i>Defa28</i>	ENSMUSG00000074434	2.84	3.21E-17	1.66	1.20E-07
AMP	<i>Defa29</i>	ENSMUSG00000074437	1.89	3.20E-06	NA	NA
AMP	<i>Defa3</i>	ENSMUSG00000074440	4.01	6.69E-26	2.90	4.28E-17
AMP	<i>Defa30</i>	ENSMUSG00000074444	3.89	1.94E-21	1.44	3.15E-4
AMP	<i>Defa32</i>	ENSMUSG00000094818	5.54	1.06E-13	NA	NA
AMP	<i>Defa33</i>	ENSMUSG00000094362	5.35	9.86E-13	NA	NA
AMP	<i>Defa34</i>	ENSMUSG00000063206	5.34	5.97E-57	2.32	1.00E-13
AMP	<i>Defa35</i>	ENSMUSG00000061845	5.85	3.04E-20	1.43	0.03
AMP	<i>Defa36</i>	ENSMUSG00000094662	4.52	7.61E-37	2.16	7.04E-11
AMP	<i>Defa5</i>	ENSMUSG00000074439	4.97	7.91E-33	NA	NA
AMP	<i>Lyz1</i>	ENSMUSG00000069515	3.86	6.73E-27	2.49	3.67E-14
AMP	<i>Pla2g2a</i>	ENSMUSG00000058908	3.23	6.78E-45	NA	NA
AMP	<i>Reg3g</i>	ENSMUSG00000074447	5.18	2.22E-22	NA	NA
Mucin related	<i>Fcgbp</i>	ENSMUSG00000047730	2.02	5.22E-08	4.40	8.45E-44
Mucin related	<i>Muc1</i>	ENSMUSG00000042784	NA	NA	NA	NA
Mucin related	<i>Muc13</i>	ENSMUSG00000022824	NA	NA	1.03	2.48E-4
Mucin related	<i>Muc2</i>	ENSMUSG00000025515	2.64	8.57E-10	4.06	2.24E-27
Mucin related	<i>Muc3</i>	ENSMUSG00000037390	-2.35	0.02	NA	NA
Mucin related	<i>Muc3a</i>	ENSMUSG00000094840	1.59	3.46E-06	2.46	1.09E-16
Mucin related	<i>Retnlb</i>	ENSMUSG00000022650	NA	NA	NA	NA
Mucin related	<i>Tff3</i>	ENSMUSG00000024029	3.27	1.89E-26	3.69	6.74E-42

Table S2.1. Part 2/2. Differentially expressed antimicrobial peptide (AMP) and mucin related genes in Paneth cell enriched enteroids and goblet cell enriched enteroids (compared to conventionally/normally differentiated enteroids). Only genes which are differentially expressed ($\log_2\text{fc} \geq 1$ and $q \text{ value} \leq 0.05$) in at least one of the datasets was included. LFC - \log_2 fold change; q val - q value; DEG - differentially expressed gene; Paneth - Paneth enriched enteroid; goblet - goblet enriched enteroid. Table reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Marker list cell type	DEG list	# Markers	# DEGs (lncRNAs & protein coding)	# DEG Markers	Hypergeometric significance test p value	Multiple testing adjusted p value (1)	-Log10 (qval) (1)	Multiple testing adjusted p value (2)
Paneth	Paneth	71	2077	56	6.07E-36	3.64E-35	34.44	6.07E-35
goblet	Paneth	422	2077	102	7.69E-10	4.61E-09	8.34	7.69E-09
enteroendocrine	Paneth	204	2077	140	7.32E-75	4.39E-74	73.36	7.32E-74
tuft	Paneth	490	2077	100	6.54E-06	NA	NA	6.54E-05
enterocyte	Paneth	518	2077	8	1	NA	NA	10
Paneth	Goblet	71	1797	40	6.97E-20	4.18E-19	18.38	6.97E-19
goblet	Goblet	422	1797	173	1.01E-55	6.08E-55	54.21	1.01E-54
enteroendocrine	Goblet	204	1797	148	8.14E-94	4.89E-93	92.31	8.14E-93
tuft	Goblet	490	1797	116	1.29E-14	NA	NA	1.29E-13
enterocyte	Goblet	518	1797	45	0.98	NA	NA	9.86

Table S2.2. Hypergeometric significance testing of cell type-specific marker enrichment in upregulated differentially expressed gene lists. Marker lists obtained from (Haber et al., 2017). DEG - differentially expressed gene. (1) = statistical test across only the Paneth, goblet and enteroendocrine cell type markers. (2) = statistical test across all 5 cell type markers. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Shared regulator	Shared regulator id	D _n -Score (corrected)	# Shared targets	# Pan only targets	# Gob only targets
Etv4	ENSMUSG00000017724	0.4	1	3	1
mmu-let-7e-5p	mmu-let-7e-5p	0.370370	49	102	38
mmu-miR-152-3p	mmu-miR-152-3p	0.3521127	63	104	46
Myb	ENSMUSG00000019982	0.3391473	83	108	67
Rora	ENSMUSG00000032238	0.3309266	281	386	164
Mitf	ENSMUSG00000035158	0.3302469	not calc	not calc	not calc
Hoxb4	ENSMUSG00000038692	0.3144654	not calc	not calc	not calc
Nr5a2	ENSMUSG00000026398	0.3176230	not calc	not calc	not calc
Irf1	ENSMUSG00000018899	0.3170732	not calc	not calc	not calc
mmu-miR-7a-5p	mmu-miR-7a-5p	0.3125	not calc	not calc	not calc
Foxa1	ENSMUSG00000035451	0.3114144	not calc	not calc	not calc
Tead4	ENSMUSG00000030353	0.3097313	not calc	not calc	not calc
Nkx2-2	ENSMUSG00000027434	0.3069544	not calc	not calc	not calc
Vdr	ENSMUSG00000022479	0.3029627	not calc	not calc	not calc
Ets1	ENSMUSG00000032035	0.3025706	not calc	not calc	not calc
Nr3c1	ENSMUSG00000024431	0.3022903	not calc	not calc	not calc
Foxa3	ENSMUSG00000040891	0.3010753	not calc	not calc	not calc
Bhlha15	ENSMUSG00000052271	0.2970660	not calc	not calc	not calc
mmu-miR-101a-3p	mmu-miR-101a-3p	0.2907895	not calc	not calc	not calc
Zfp57	ENSMUSG00000036036	0.287578	not calc	not calc	not calc
Fosl1	ENSMUSG00000024912	0.2925170	not calc	not calc	not calc
Pax6	ENSMUSG00000027168	0.2880184	not calc	not calc	not calc
Nfatc2	ENSMUSG00000027544	0.2954545	not calc	not calc	not calc
Neurod1	ENSMUSG00000034701	0.2808552	not calc	not calc	not calc
Insm1	ENSMUSG00000068154	0.2811218	not calc	not calc	not calc
mmu-miR-153-3p	mmu-miR-153-3p	0.2727273	not calc	not calc	not calc
Neurod2	ENSMUSG00000038255	0.2696850	not calc	not calc	not calc
Fosb	ENSMUSG00000003545	0.2652174	not calc	not calc	not calc
Klf15	ENSMUSG00000030087	0.2857143	not calc	not calc	not calc
Atoh1	ENSMUSG00000073043	0.2449495	not calc	not calc	not calc

Table S2.3. Rewiring analysis results for the marker regulators present in the Paneth and the goblet subnetworks. D_n score generated using Cytoscape app DyNet (Goenawan et al., 2016). Not calc - not calculated. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Cell type-specific regulatory network	Crohn's susceptibility gene	Direction of differential expression
Paneth	9430076C15Rik	Upregulated
Paneth	Atg16l2	Upregulated
Paneth	Fut2	Upregulated
Paneth	Hmha1	Upregulated
Paneth	Itln1	Upregulated
Paneth	Izumo1	Upregulated
Paneth	Jazf1	Upregulated
Paneth	Plcl1	Upregulated
Paneth	Tnfsf15	Upregulated
Paneth	Ccdc88b	Downregulated
Paneth	Dbp	Downregulated
Paneth	Fads1	Downregulated
Paneth	Fads2	Downregulated
Paneth	H2-Q1	Downregulated
Paneth	H2-Q10	Downregulated
Paneth	H2-Q2	Downregulated
Paneth	H2-Q6	Downregulated
Paneth	H2-Q7	Downregulated
Paneth	Kif21b	Downregulated
Paneth	Ksr1	Downregulated
Paneth	Ptpn22	Downregulated
Paneth	Zpbp2	Downregulated
Goblet	Fut2	Upregulated
Goblet	Hmha1	Upregulated
Goblet	Inpp5d	Upregulated
Goblet	Itln1	Upregulated
Goblet	Izumo1	Upregulated
Goblet	Jazf1	Upregulated
Goblet	Plcl1	Upregulated
Goblet	Tnfsf15	Upregulated
Goblet	Gart	Downregulated
Goblet	H2-Q7	Downregulated
Goblet	H2-Q6	Downregulated
Goblet	Notch2	Downregulated

Table S2.4. Crohn's disease SNP associated genes in the enriched enteroid regulatory networks. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Cell type-specific regulatory network	Ulcerative Colitis susceptibility gene	Direction of differential expression
Paneth	Dap	Upregulated
Paneth	Edem2	Upregulated
Paneth	Itgal	Upregulated
Paneth	Maml2	Upregulated
Paneth	Mmp24	Upregulated
Paneth	Nr5a2	Downregulated
Paneth	Plcl1	Upregulated
Paneth	Tnfsf15	Upregulated
Paneth	Zpbp2	Downregulated
Paneth	Card11	Downregulated
Paneth	Hnf4A	Downregulated
Paneth	Nusap1	Downregulated
Paneth	Procr	Upregulated
Goblet	Dap	Upregulated
Goblet	Edem2	Upregulated
Goblet	Itgal	Upregulated
Goblet	Mmp24	Upregulated
Goblet	Nr5a2	Downregulated
Goblet	Plcl1	Upregulated
Goblet	Tnfsf15	Upregulated
Goblet	Card11	Downregulated
Goblet	Cep250	Downregulated
Goblet	Procr	Upregulated

Table S2.5. Ulcerative colitis disease SNP associated genes in the enriched enteroid regulatory networks. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

UC SNP associated genes	CD SNP associated genes	Goblet differentially expressed genes
ENSMUSG00000026398	ENSMUSG00000053007	ENSMUSG00000013523
ENSMUSG00000027611	ENSMUSG00000010663	ENSMUSG00000017057
ENSMUSG00000027612	ENSMUSG00000017195	ENSMUSG00000024597
ENSMUSG00000030830	ENSMUSG00000018334	ENSMUSG00000027006
ENSMUSG00000036526	ENSMUSG00000024665	ENSMUSG00000027346
ENSMUSG00000038241	ENSMUSG00000027843	ENSMUSG00000027513
ENSMUSG00000038349	ENSMUSG00000038349	ENSMUSG00000027876
ENSMUSG00000039168	ENSMUSG00000047767	ENSMUSG00000028236
ENSMUSG00000050395	ENSMUSG00000047810	ENSMUSG00000031844
ENSMUSG00000038312	ENSMUSG00000050395	ENSMUSG00000032322
	ENSMUSG00000055978	ENSMUSG00000032978
	ENSMUSG00000059824	ENSMUSG00000034472
	ENSMUSG00000060550	ENSMUSG00000038039
	ENSMUSG00000063568	ENSMUSG00000039234
	ENSMUSG00000067235	ENSMUSG00000046841
	ENSMUSG00000073409	ENSMUSG00000055976
	ENSMUSG00000079507	ENSMUSG00000074004
	ENSMUSG00000038209	ENSMUSG00000075610
	ENSMUSG00000091705	ENSMUSG00000055963
	ENSMUSG00000035697	ENSMUSG00000036764
	ENSMUSG00000064158	

Table S2.6. IBD associated genes targeted by predicted master regulators in the enriched enteroid regulatory networks. Ulcerative colitis (UC) and Crohn's disease (CD) associated genes (from SNP data) targeted by at least one of the master regulators in the relevant networks; list of top 100 CD differentially expressed genes in human colonic biopsies (CD inflamed vs healthy) which are targeted by at least one of the predicted goblet cell master regulators in the GCeE network. Figure reproduced from Treveil et al. (2020) under the Creative Commons BY licence.

Appendix 2: Supplementary data for Chapter 3

This appendix contains all the supplementary materials for Chapter 3.

- **Table S3.1** – Number of differentially expressed genes from the cytokine-treated colonoids also differentially expressed in UC and cCD biopsy data (GSE16879) (Arijs et al., 2009).

Cytokine programme category	# colonoid DEGs	# colonoid DEGs in biopsies	% colonoid DEGs in biopsies	# colonoid DEGs in UC + cCD	# colonoid DEGs in UC not cCD	# colonoid DEGs in cCD not UC
IFN γ	871	305	35.02%	191	102	12
IFN γ ; IL-13	252	97	38.49%	63	31	3
IFN γ ; IL-13; IL-17a	16	4	25.00%	3	1	0
IFN γ ; IL-13; IL-17a; TNF α	29	10	34.48%	7	3	0
IFN γ ; IL-13; TNF α	65	31	47.69%	20	8	3
IFN γ ; IL-17a	40	15	37.50%	9	6	0
IFN γ ; IL-17a; TNF α	54	13	24.07%	11	2	0
IFN γ ; TNF α	114	58	50.88%	43	15	0
IL-13	1408	482	34.23%	262	208	12
IL-13; IL-17a	23	6	26.09%	3	3	0
IL-13; IL-17a; TNF α	16	8	50.00%	8	0	0
IL-13; TNF α	143	61	42.66%	34	25	2
IL-17a	66	18	27.27%	12	6	0
IL-17a; TNF α	35	14	40.00%	10	4	0
TNF α	310	104	33.55%	50	52	2

Table S3.1. Number of differentially expressed genes from the cytokine-treated colonoids also differentially expressed in UC and cCD biopsy data (GSE16879) (Arijs et al., 2009). Differential expression when adjusted p value ≤ 0.01 .

Appendix 3: Supplementary data for Chapter 4

This appendix contains all the supplementary materials for Chapter 4.

- **Figure S4.1** - Neonatal differentially expressed genes.
- **Table S4.1** - Reactome pathway enrichment analysis for neonatal upregulated differentially expressed genes.
- **Table S4.2** - Cluster analysis on neonatal protein-protein interaction network.
- **Table S4.3** - Overlap between cell type marker genes and differentially expressed genes.
- **Table S4.4** - Reactome pathway enrichment analysis of differentially expressed stem cell signature genes and their expressed regulators.

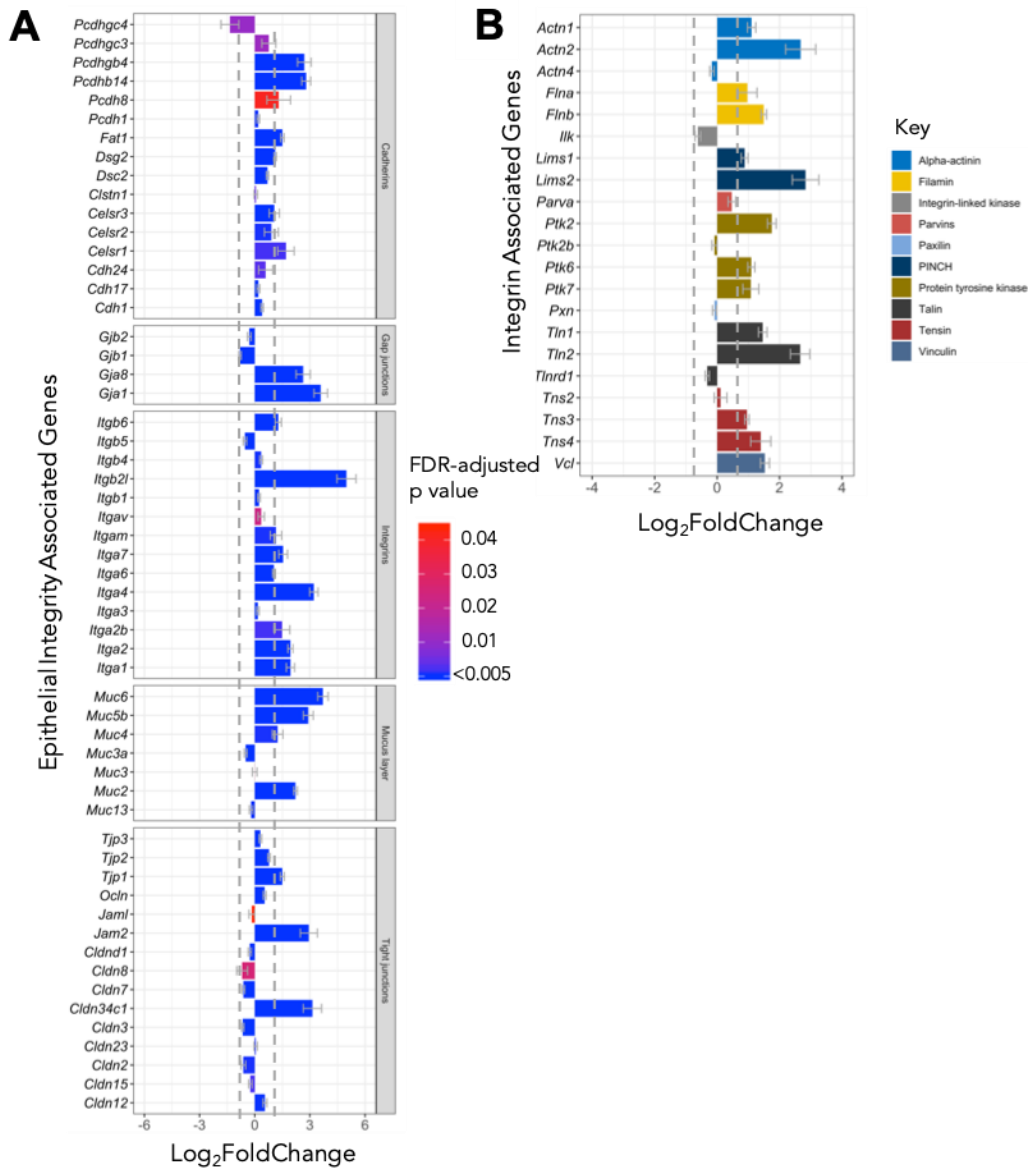


Figure S4.1. Neonatal differentially expressed genes. A. Epithelial integrity-associated DEGs **B.** Integrin-associated DEGs. All DEGs have q value < 0.05. Dotted line indicates the threshold of significance (absolute Log₂FC > 1.0). Data are represented as Mean ± SE. Figure created by Raymond Kiu (QIB). Figure reproduced from Kiu et al. (2020) under the Creative Commons BY licence.

Term	Gene count	Fold enrichment	Q value
Olfactory Signalling Pathway	108	1.88	3.20E-05
Cell Cycle	154	1.63	1.13E-04
Gene expression (Transcription)	238	1.45	1.18E-04
Chromatin organization	67	2.08	2.54E-04
Chromatin modifying enzymes	67	2.08	2.82E-04
DNA Double-Strand Break Repair	51	2.3	4.19E-04
G alpha (s) signalling events	122	1.62	6.50E-04
RNA Polymerase II Transcription	202	1.39	3.06E-03
Generic Transcription Pathway	173	1.39	8.50E-03
Signal Transduction	461	1.21	8.98E-03
G2/M DNA damage checkpoint	29	2.48	9.07E-03
DNA Repair	77	1.66	1.06E-02
Cell Cycle, Mitotic	122	1.48	1.06E-02
HATs acetylate histones	23	2.64	1.85E-02
Epigenetic regulation of gene expression	28	2.28	2.38E-02
Homology Directed Repair	30	2.23	2.39E-02
Nonhomologous End-Joining	19	2.82	2.46E-02
Mitotic Prometaphase	58	1.73	2.48E-02
HDR through Homologous Recombination	29	2.26	2.63E-02
Recruitment and ATM-mediated phosphorylation of repair	22	2.53	2.78E-02

Table S4.1. Reactome pathway enrichment analysis for neonatal upregulated differentially expressed genes. q value \leq 0.05. Table reproduced from Kiu et al. (2020) under the Creative Commons BY licence.

Cluster	Score (Density*# Nodes)	Nodes	Edges	Gene UNIPROT IDs
1	6.716	68	276	P70365, O89090, Q8BUR4, P25425, Q3TTA7, Q9JKF1, P13405, O35618, P20444, O70494, Q925J9, Q62190, P21803, P01132, Q80YR6, Q80UV9, Q9JHD1, Q61026, Q64143, Q9JL70, O35608, Q01279, Q8CBW3, P62806, Q00731, P34056, Q8K4J0, Q923E4, P97929, Q4U2R1, Q9QZR5, Q60760, Q62077, P50652, P16092, Q08297, P97313, Q6PDQ2, P84228, Q5PSV9, B2RWS6, Q6VNS1, Q61521, Q60751, Q03145, Q3TYD6, Q6NZM9, B2RQC6, Q6ZQF0, P15208, O08852, O70445, Q64455, Q01705, Q9Z265, Q9Z0Z3, Q8BGE5, P22682, O09053, P07901, Q6QI06, P34152, P26450, Q61084, P14234, P51943, Q62245, Q64701
2	4.48	26	68	Q62108, Q7TS75, Q8BSK8, O08586, Q8CHE4, Q8CIS0, O88572, Q02111, P23242, P12813, Q9EQD0, Q9EQY0, Q01147, Q9WVG5, Q570Y9, P16054, P01101, Q91Y86, Q8BTH8, Q9Z2A0, P39447, Q80YE7, Q8C050, P10637, P18654, Q9WV60
3	4.4	11	23	Q60952, A2AUM9, Q9D3R3, Q569L8, Q80Z25, Q64702, O35942, Q9ROL6, Q6P5D4, Q0VEJ0, Q6A078
4	3.775	72	167	Q60665, P36895, Q05909, P27512, A2A5Z6, Q9WV30, Q8BWW9, P10417, P70347, Q8BFP9, Q7TSG3, O55033, Q61214, Q6PDM2, Q9CR14, P70191, Q6PEE3, Q9WUL6, Q62210, O08863, P68433, P26041, Q62469, Q9WUN2, P48754, Q04750, Q9WVF7, Q9CQ37, Q8K368, P26039, O08553, Q03173, O08901, P15379, Q8BPZ8, Q91YM2, Q64261, P33609, Q62417, Q62448, Q60803, Q9QUM0, Q91WJ8, P23804, P00520, P22518, Q62167, Q8C863, Q80TQ2, P62484, Q0VBD2, Q5I043, O35607, Q8VHL1, P70335, Q9Z1S0, O35732, P11276, Q04736, O54781, Q64727, P25322, Q64700, O35516, Q62388, P53995, P01108, O70551, Q6A4J8, P30280, Q8BUN5, P13864

Table S4.2. Cluster analysis on neonatal protein-protein interaction network. Clusters identified using MCODE and must have ≥ 10 nodes (Bader and Hogue, 2003). Table reproduced from Kiu et al. (2020) under the Creative Commons BY licence.

Cell type	# Marker genes	# DEG marker genes	Q value	% Marker genes which are DEGs	Genes (Symbol)
Stem	122	37	0.026	30.33%	H2-Eb1, Mfge8, Lgr5, Rgmb, Fstl1, Scn2b, Sorbs2, Slc14a1, Slc14a1, Slc12a2, Rassf5, Rnf43, Lamb3, Cd44, Axin2, Lrig1, Cdk6, Rnf32, Smoc2, Esrrg, Znrf3, Aqp4, BC064078, Zbtb38, Myo9a, Lrp4, Arhgef26, Cttntp2, Htr4, Sipa1l1, Rin2, Sesn3, Phlpp1, Itga1, Pcdh8, Nfib, Nfia
Tuft	494	107	1.669	21.66%	Ackr4, Eppk1, Afdn, Jade1, Zfp512b, Klf6, Folr1, Nudt14, Ltc4s, Tmem9, Tmem141, S100a11, Espn, 1110008P14Rik, Gga2, Tmem51, Snrnp25, S100a1, Scand1, Ndufaf3, Rhog, Vamp8, Nradd, Pde6d, Galk1, Lect2, Bicc1, Prox1, Ehf, Wdfy2, Hivep2, Myo1b, Tmem160, Cpeb4, Fam49a, Klhl28, Aoep, Arhgef28, Jmy, Plk2, Slc4a7, Txndc16, Dpysl2, Cblb, Adcy5, Cpne5, Sik1, Pdpk1, Svil, Atat1, Ptptrj, Avil, Hspa4l, Map4k4, Fn1, Pkp1, Rabgap1, Gpcpd1, Pik3r3, Tas1r3, Ttl10, Klf3, Osbp13, Nfe2l3, Atf7ip, Itpr2, Psd3, Pgm2l1, Trak1, Zbtb41, Slc26a2, Exph5, Tiparp, Ankrd12, Rabgap1, Zdhhc17, Gmip, Atp8a1, Wnk2, Arap2, Otud7b, Jarid2, Zfhx3, Slco4a1, Lmtk2, Suco, Adam22, Madd, Dmxl2, Rgl2, Myzap, Atxn1, Samd9l, Omd, Bnip5, Mical3, Nav2, Sirt5, Dsp, Tmem245, Tead1, Pla2g4a, Msi2, Fryl, Fnbp1, Usp49, Vmn2r26
Entero-endocrine	202	36	4.942	17.82%	Cdkn1c, Gpx3, Bex3, Pcsk1n, Resp18, Sct, Adora3, Cacna1a, Cbfa2t2, Chd7, Chst11, Cnot6l, Dock4, Fam135a, Gng4, Kcnh6, Klhl31, Maml3, Map1b, Map3k15, Mapkbp1, Ncald, Pcsk1, Pde11a, Peg3, Phldb2, Plxnb1, Prnp, Rargef4, Rbfox2, Rimp2, Ruffy2, Syt13, Tox3, Trpm2, Unc13a

Table S4.3. Part 1/2. Overlap between cell type marker genes and differentially expressed genes in neonatal mice. Q value calculated using hypergeometric significance test with Benjamini-Hochberg correction for multiple testing. Table reproduced from Kiu et al. (2020) under the Creative Commons BY licence.

Cell type	# Marker genes	# DEG marker genes	Q value	% Marker genes which are DEGs	Genes (Symbol)
Paneth	71	11	4.955	15.49%	Defa42, Mptx1, Apoc2, Samd5, Thbs1, Angpt2, Acvr1c, C4bp, Slc16a7, Dll3, Lamb1
Goblet	420	62	5.993	14.76%	Naga, Guca2a, Lrrc26, S100a11, Sh3bgrl3, Scnn1a, Ccnd3, Cmtm8, Krtcap2, Selenom, Sec61g, Tmsb10, Ggcx, Rasa4, Eif2ak4, Igf2bp1, Itga2, Arfgef3, Appl2, Tc2n, Kif13a, Golm1, Sybu, Syt7, Muc2, Tfcp2l1, Galnt5, Ncoa3, Odf2l, Rfc1, Dipk1a, Aacs, Sytl2, Ern2, Sytl4, Mcf2l, Galnt7, Slc10a7, Myo5c, Golgb1, Tulp4, Cog3, Dcbld2, Ugg1, Zc3h7a, Gcc2, Capn8, Nipal2, Sgsm3, Rasef, Edem3, Smim5, Plcb1, Tbc1d30, Ggcx, Sh3pxd2a, Fgfr3, Clca3a1, Fry, Cracr2a, Ggcx, Gcc2
Enterocyte	522	68	6.000	13.03%	Spink1, Sult6b2, Smim24, Acp5, Apoc2, Fcgrt, Gstm3, Crip1, Tmem86a, Ndufa1, Fabp6, Amn, Cideb, Cst6, Prap1, Akr7a5, 2200002D01Rik, Apoc3, Apoa1, Rbp2, Reep6, Tmem37, Gstm1, Tmem253, Sco2, Gstp2, Epb41l3, Btl5, Dgkq, D130043K22Rik, Mertk, Hnf4g, Cobl, Snx13, Ccdc88c, Cast, Plec, Mylk, Pcsk5, Abcc2, Cubn, Slc23a2, Edn3, Pld1, Mme, Plb1, Ppargc1a, Rufy3, Bmp3, Slc13a1, Slc25a36, Farp2, Slc25a37, C530008M17Rik, Arhgap26, Tbc1d24, Slc18b1, Mical2, Txlng, Ptk6, Apol10b, Gm11437, Ccdc134, Mgam, Ahnak, Acnat1, H2-BI

Table S4.3. Part 2/2. Overlap between cell type marker genes and differentially expressed genes in neonatal mice. Q value calculated using hypergeometric significance test with Benjamini-Hochberg correction for multiple testing. Table reproduced from Kiu et al. (2020) under the Creative Commons BY licence.

Term	Gene count	Fold enrichment	Q value
Regulation of FZD by ubiquitination (R-MMU-4641263)	3	35.6	2.39E-02
Endogenous sterols (R-MMU-211976)	3	27.06	3.63E-02
Estrogen-dependent gene expression (R-MMU-9018519)	6	19.61	4.95E-04
ESR-mediated signalling (R-MMU-8939211)	6	18.04	4.69E-04
RUNX1 regulates transcription of genes involved in differentiation of HSCs (R-MMU-8939236)	4	14.78	3.41E-02
Signalling by Nuclear Receptors (R-MMU-9006931)	6	11.66	4.23E-03
Transcriptional regulation by RUNX1 (R-MMU-8878171)	6	9.73	9.64E-03
Generic Transcription Pathway (R-MMU-212436)	15	5.88	4.71E-05
RNA Polymerase II Transcription (R-MMU-73857)	15	5	1.89E-04
Gene expression (Transcription) (R-MMU-74160)	15	4.38	4.93E-04
Signal Transduction (R-MMU-162582)	19	2.37	3.65E-02

Table S4.4. Reactome pathway enrichment analysis of differentially expressed stem cell signature genes and their expressed regulators in neonatal mice. q value \leq 0.05. Table reproduced from Kiu et al. (2020) under the Creative Commons BY licence.

Appendix 4: Supplementary data for Chapter 5

This appendix contains all the supplementary materials for Chapter 5.

- **File S5.1** - Results of gene set enrichment analysis comparing bifidobacteria-exposed to control mice in each mouse type and cell type using Reactome and Gene Ontology Biological Process annotations (electronic supplementary materials).
- **Figure S5.1** - Representative gating strategy to sort Paneth, stem and transit amplifying (TA) cells from single cell suspension of mouse crypt of Lieberkühn cells.
- **Figure S5.2** - Lactic acid bacterial concentration in faecal and caecal content from germ free (GF) and specific pathogen free (SPF) juvenile mice.
- **Figure S5.3** - Germ free vs specific pathogen free UMAP plot of normalised counts data.
- **Figure S5.4** - Principle component analysis (PC1 v PC2) of juvenile mice samples.
- **Figure S5.5** - Gene expression heatmap of top 50 variant genes among all germ free/monocolonised juvenile samples.
- **Figure S5.6** - Violin plot of all samples showing levels of gene expression of *Lgr5*, *CD24a* and *Lyz1*.
- **Figure S5.7** - Expression of key transcription factors in intestinal epithelial cell differentiation across germ free/monocolonised juvenile mice samples.

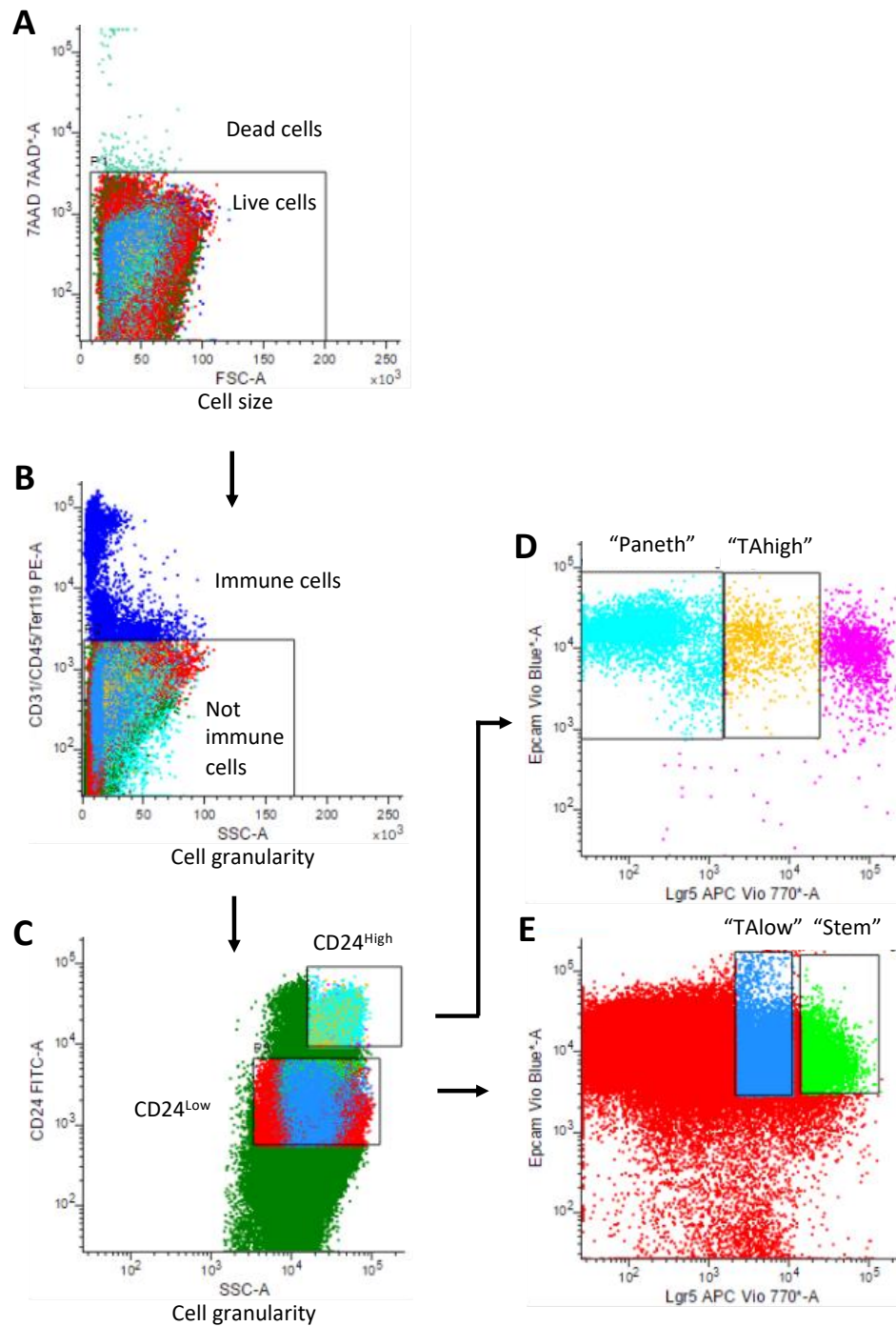


Figure S5.1. Representative gating strategy to sort Paneth, stem and transit amplifying (TA) cells from single cell suspension of mouse crypt of Lieberkühn cells. **A.** 7-Aminoactinomycin D (7AAD) fluorescence against forward scatter (FSC) to gate live and dead cells. **B.** Cluster of differentiation (CD)31/CD45/Ter119 against side scatter to gate immune cells and non-immune cells. **C.** CD24 against side scatter to obtain CD24^{high} and CD24^{low} cells. **D.** Epithelial cell adhesion molecule (EpCAM) against LGR5 fluorescence for CD24^{high} cells to sort Paneth cells and TA^{high} cells. **E.** EpCAM against LGR5 fluorescence for CD24^{low} cells to sort TA^{low} cells and stem cells. Data from specific pathogen free (SPF) mouse 1. Figures from the software BD FACSCorus of the FACSMelody sorting machine.

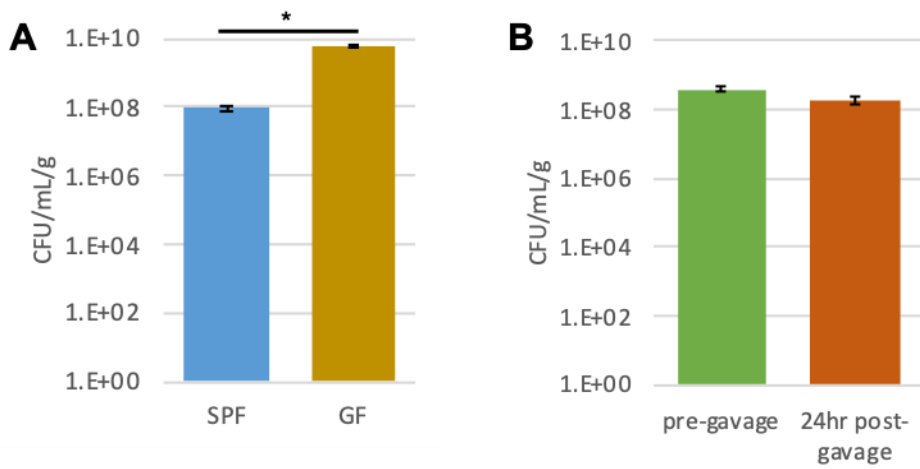


Figure S5.2. Lactic acid bacterial concentration in faecal and caecal content from germ free (GF) and specific pathogen free (SPF) juvenile mice. A. Bacterial concentration (CFU/mL/g) in SPF and GF caecal content 72h after gavage of *B. breve* UCC2003 * = $p < 0.05$ (paired t-test). **B.** Bacterial concentration (CFU/mL/g) in SPF mouse faecal content before gavage and 24h after gavage of *B. breve* UCC2003. Data not significant (paired t-test, $p > 0.05$). Bacteria grown anaerobically on De Man, Rogosa and Sharpe plates at 37° for 48h. Standard Error shown.

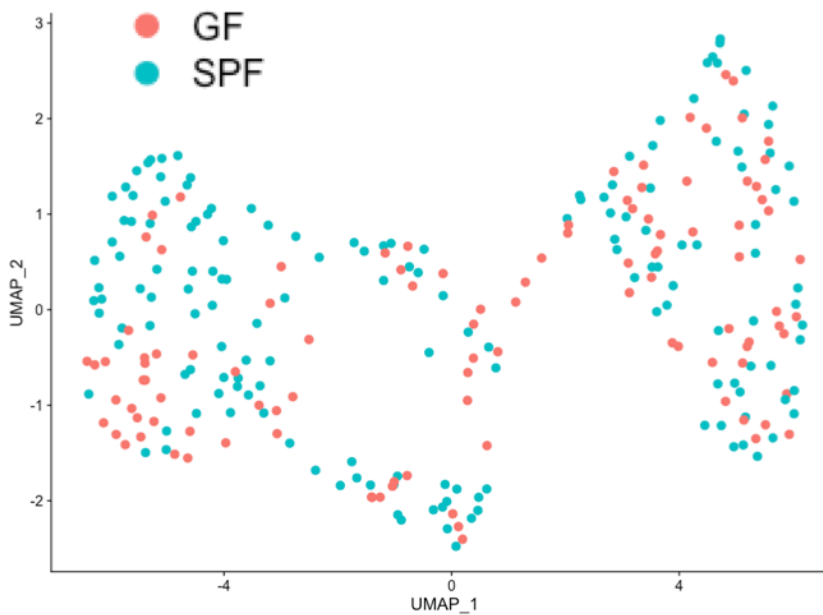


Figure S5.3. Germ free vs specific pathogen free UMAP plot of normalised counts data. Data coloured by mouse type. SPF - specific pathogen free mice; GF - germ free mice. See Figure 5.5 for matching IDs.

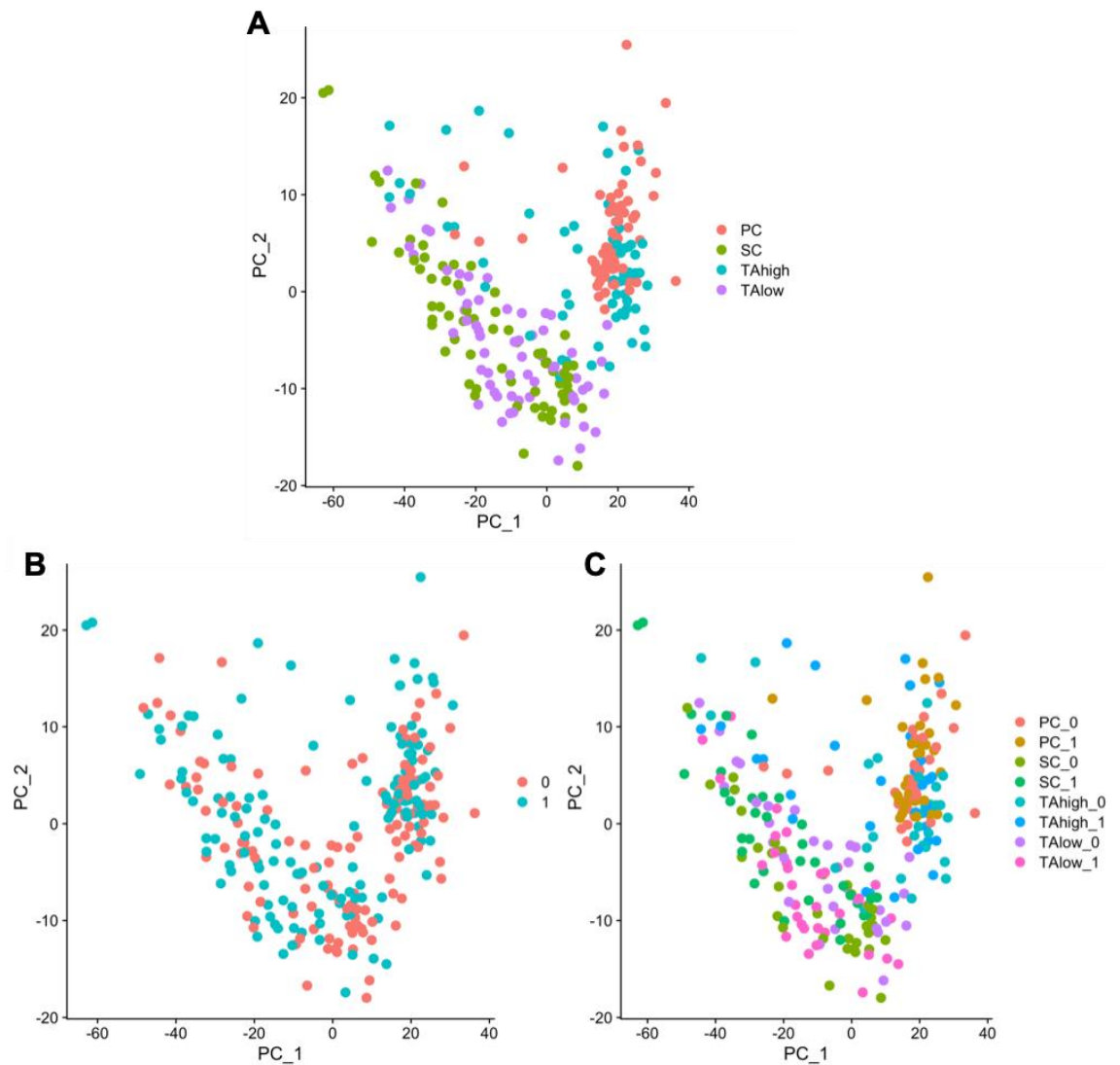


Figure S5.4. Principle component analysis (PC1 v PC2) of juvenile mice samples. A. Samples identified by pre-labelled cell type. B. Samples labelled by bifidobacteria-treated or control condition. C. Samples labelled by cell type and bifidobacteria-treated or control condition. PC - Paneth cell; SC - stem cell, TAhigh - transit amplifying high cells; TAlow - transit amplifying low cells; 0 - control condition; 1 - bifidobacteria-treated.

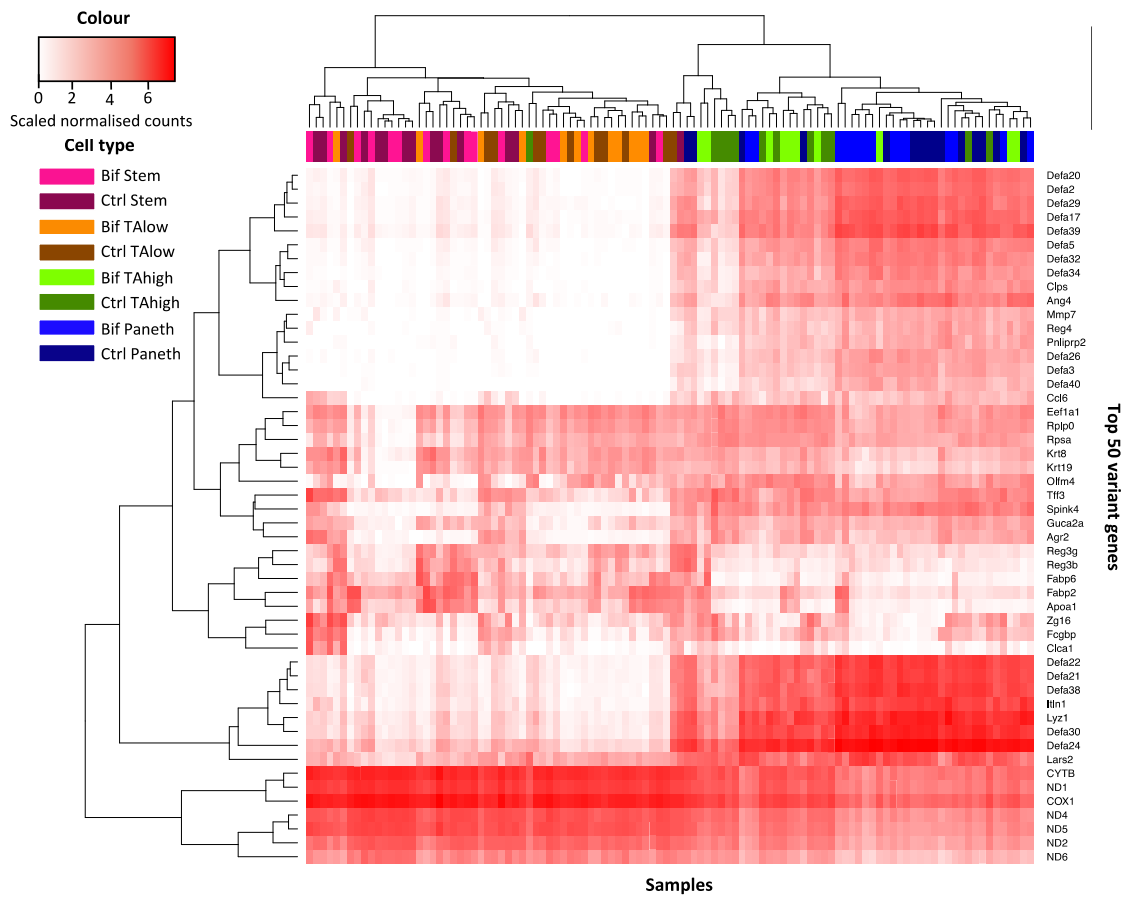


Figure S5.5. Gene expression heatplot of top 50 variant genes among all germ free/monocolonised juvenile samples. Samples and genes are clustered. Bif - bifidobacteria-treated samples; Ctrl - control samples.

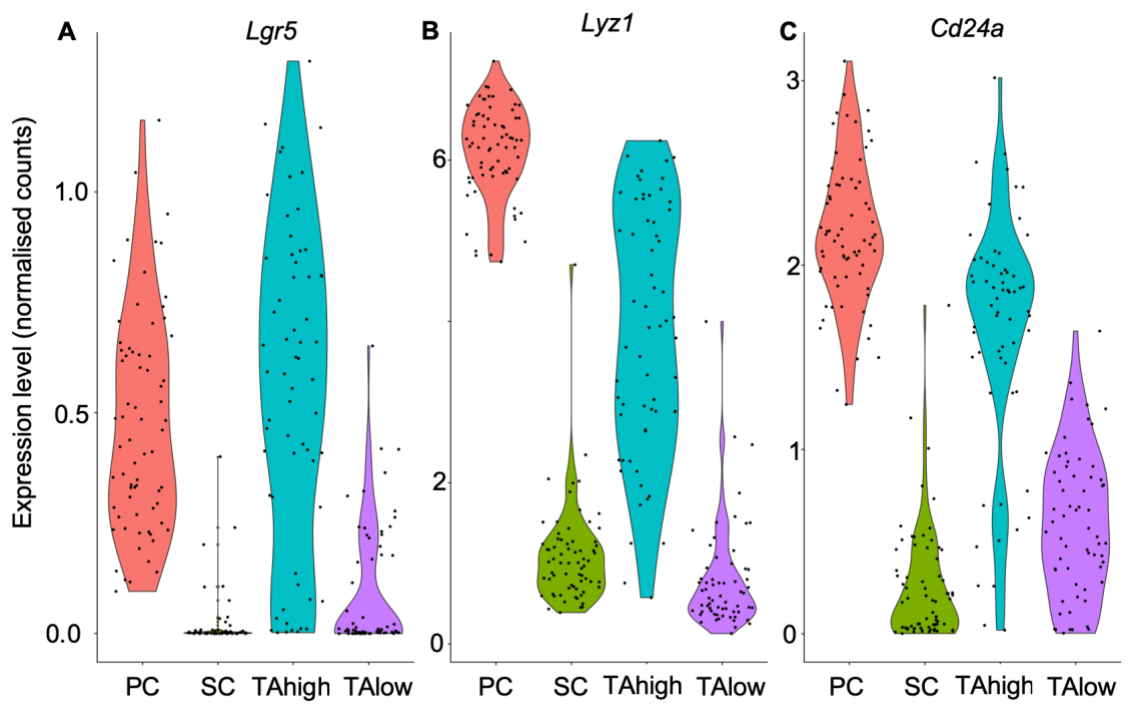


Figure S5.6. Violin plot of all samples showing levels of gene expression of *Lgr5*, *CD24a* and *Lyz1*. PC – Paneth cells; SC – stem cells; TAhigh – TAhigh cells; TALow – TALow cells.

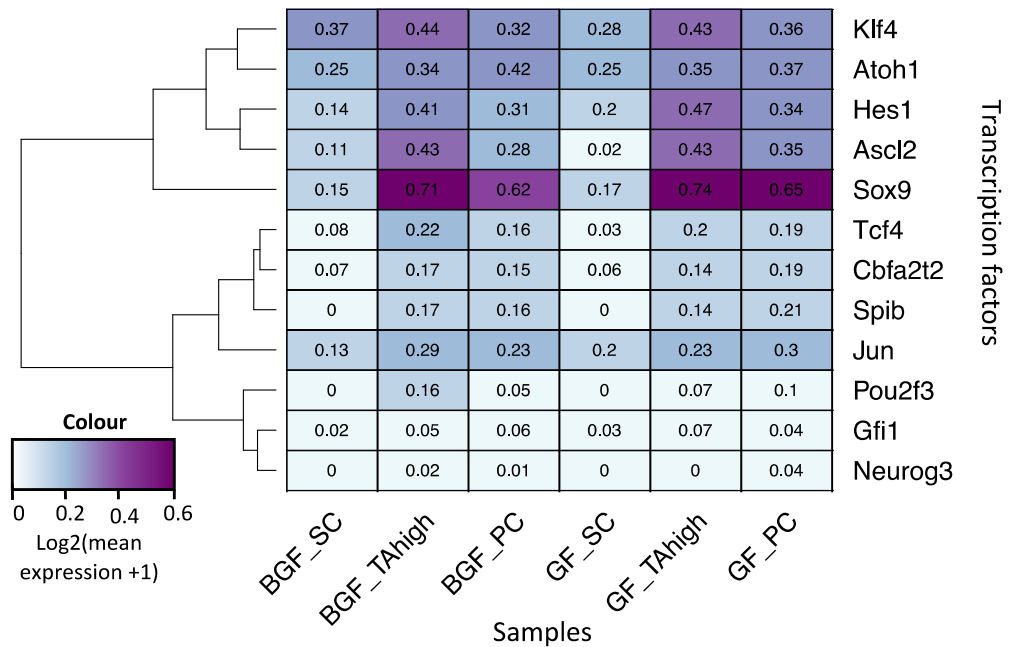


Figure S5.7. Expression of key transcription factors in intestinal epithelial cell differentiation across germ free/monocolonised juvenile mice samples. Mean scaled expression for each GF cell type. PC - Paneth cell; SC - stem cell, TAhigh - transit amplifying high cell.

Appendix 5: Supplementary data for Chapter 6

This appendix contains all the supplementary materials for Chapter 6.

- **Table S6.1** - PFAM domain annotations for potential NF- κ B activating proteins based on LC-MS.

UniProt protein ID	Protein name (and symbol where known)	Domain/s	Domain description
F9XZH3	Alpha-1,4 glucan phosphorylase (GlgP1)	PF00343	Carbohydrate phosphorylase
F9XY36	Glutamine synthetase (glnA)	PF00120; PF03951	Glutamine synthetase, catalytic domain; Glutamine synthetase, beta-Grasp domain
F9XZF9	Conserved hypothetical secreted protein	PF08757	CotH protein
F9Y1U3	Solute-binding protein of ABC transporter system for sugars (GalC)	PF01547	Bacterial extracellular solute-binding protein
F9XZ02	Conserved hypothetical secreted protein with Excalibur domain	PF05901	Excalibur calcium-binding domain
F9Y2P4	Conserved hypothetical membrane spanning protein with Endonuclease/Exonuclease/phosphatase family domain	PF03372	Endonuclease/Exonuclease/phosphatase family
F9XYM5	Trigger factor (Tig)	PF05698; PF05697; PF00254	Bacterial trigger factor protein (TF) C-terminus; Bacterial trigger factor protein (TF); FKBP-type peptidyl-prolyl cis-trans isomerase
F9XZB8	Sugar-binding protein of ABC transporter system	PF13407	Periplasmic binding protein domain
F9XZ70	Hypothetical secreted protein	(PS51257) *	(Prokaryotic membrane lipoprotein lipid attachment site profile)*
F9XY18	SSU ribosomal protein S1P (RspA)	PF00575	S1 RNA binding domain
F9Y109	Conserved hypothetical secreted protein	-	-

Table S6.1. PFAM domain annotations for potential NF- κ B activating proteins based on LC-MS.

*Annotation obtained from PROSITE rather than PFAM (El-Gebali et al., 2019; Sigrist et al., 2013).

Appendix 6: Peer-reviewed publications

Molecular
Omics



RESEARCH ARTICLE

View Article Online
View Journal | View Issue



Cite this: *Mol. Omics*, 2020,
16, 39

Regulatory network analysis of Paneth cell and goblet cell enriched gut organoids using transcriptomics approaches†

A. Treveil,[‡] P. Sudhakar,^{‡,ac} Z. J. Matthews,^{‡,d} T. Wrzesiński,^a E. J. Jones,^{ab} J. Brooks,^{abde} M. Ölbei,^{ab} I. Hautefort,^a L. J. Hall,^b S. R. Carding,^{bd} U. Mayer,^f P. P. Powell,^d T. Wileman,^{bd} F. Di Palma,^a W. Haerty^{*a} and T. Korcsmáros[‡]

The epithelial lining of the small intestine consists of multiple cell types, including Paneth cells and goblet cells, that work in cohort to maintain gut health. 3D *in vitro* cultures of human primary epithelial cells, called organoids, have become a key model to study the functions of Paneth cells and goblet cells in normal and diseased conditions. Advances in these models include the ability to skew differentiation to particular lineages, providing a useful tool to study cell type specific function/dysfunction in the context of the epithelium. Here, we use comprehensive profiling of mRNA, microRNA and long non-coding RNA expression to confirm that Paneth cell and goblet cell enrichment of murine small intestinal organoids (enteroids) establishes a physiologically accurate model. We employ network analysis to infer the regulatory landscape altered by skewing differentiation, and using knowledge of cell type specific markers, we predict key regulators of cell type specific functions: *Cebpa*, *Jun*, *Nr1d1* and *Rxra* specific to Paneth cells, *Gfi1b* and *Myc* specific for goblet cells and *Ets1*, *Nr3c1* and *Vdr* shared between them. Links identified between these regulators and cellular phenotypes of inflammatory bowel disease (IBD) suggest that global regulatory rewiring during or after differentiation of Paneth cells and goblet cells could contribute to IBD aetiology. Future application of cell type enriched enteroids combined with the presented computational workflow can be used to disentangle multifactorial mechanisms of these cell types and propose regulators whose pharmacological targeting could be advantageous in treating IBD patients with Crohn's disease or ulcerative colitis.

Received 16th August 2019.
Accepted 23rd October 2019

DOI: 10.1039/c9mo00130a

rsc.li/molomics

Introduction

Gut barrier integrity is critically important for efficient nutrient absorption and maintenance of intestinal homeostasis¹ and is maintained by the combined action of the various cell types lining the intestinal epithelium.² These intestinal epithelial cells serve to mediate signals between the gut microbiota and the host innate/adaptive immune systems.^{3,4} Disruption of

epithelial homeostasis along with dysregulated immune responses are some of the underlying reasons behind the development of inflammatory bowel disease (IBD) such as Crohn's disease (CD) and ulcerative colitis (UC).⁵ Therefore, a greater insight into the functions of intestinal cells will further our understanding of the aetiology of inflammatory gut conditions.

To date, various cell types have been identified in the intestinal epithelium based on specific functional and gene expression signatures. Paneth cells residing in the small intestinal crypts of Lieberkühn help to maintain the balance of the gut microbiota by secreting anti-microbial peptides, cytokines and other trophic factors.⁶ Located further up the intestinal crypts, goblet cells secrete mucin, which aggregates to form the mucus layer, which acts as a chemical and physical barrier between the intestinal lumen and the epithelial lining.⁷ Both of these cell populations have documented roles in gut-related diseases.^{8,9} Dysfunctional Paneth cells with reduced secretion of anti-microbial peptides have been shown to contribute to the pathogenesis of CD,¹⁰ while reduction in goblet cell numbers

^a Earlham Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UZ, UK.
E-mail: Tamas.Korcsmaros@earlham.ac.uk; Fax: +44-(0)1603450021;
Tel: +44-(0)1603450961

^b Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk, NR4 7UA, UK

^c Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, Belgium

^d Norwich Medical School, University of East Anglia, Norwich, Norfolk, NR4 7TJ, UK

^e Department of Gastroenterology, Norfolk and Norwich University Hospitals, Norwich, UK

^f School of Biological Sciences, University of East Anglia, Norwich, Norfolk, NR4 7TJ, UK

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9mo00130a

‡ Equal contributions.



View Article Online

Research Article

Molecular Omics

and defective goblet cell function has been associated with UC in humans.¹¹

Recent studies have employed single cell transcriptomics sequencing of tissue samples to characterise the proportion and signatures of different epithelial cell types in the intestines of healthy and IBD patients.^{12–14} However, to provide deeper insights into the role of specific cell populations (such as Paneth cells and goblet cells) in IBD, *in vitro* models are required for in depth testing and manipulation. Such models can be used to study specific mechanisms of action, host-microbe interactions, intercellular communication, patient specific therapeutic responses and to develop new diagnostic approaches. Due to ease of manipulation, observation and analysis, organoid models, including small intestinal models (enteroids), are increasingly used in the IBD field.^{15–17} Enteroid cell culture systems employ growth factors to expand and differentiate Lgr5+ stem cells into spherical models of the small intestinal epithelia which recapitulate features of the *in vivo* intestinal tissue.^{18–20} It has been shown that these enteroids contain all the major cell types of the intestinal epithelium, and exhibit normal *in vivo* functions.²¹ These models, generated using intestinal tissue from mice, from human patient biopsies or from induced pluripotent stem cells (iPSCs), have proven particularly valuable for the study of complex diseases which lack other realistic models and exhibit large patient variability, such as IBD.^{22,23} Small molecule treatments have been developed that skew the differentiation of enteroids towards Paneth cell or goblet cell lineages, improving representation of these cells within the enteroid cell population.^{24,25} Specifically, differentiation can be directed towards the Paneth cell lineage through the addition of DAPT, which inhibits notch signaling, and CHIR99021, which inhibits GSK3 β -mediated β -catenin degradation. Enteroid cultures enriched in goblet cells can be generated through the addition of DAPT and IWP-2, an inhibitor of Wnt signaling.²⁵ Whilst these methods do not present single cell type resolution, they provide useful tools to study Paneth cell and goblet cell populations in the context of the other major epithelial cell types.²⁶ A recent study by Mead *et al.* found that Paneth cells from enriched enteroids more closely represent their *in vivo* counterparts than those isolated from conventionally differentiated enteroids, based on transcriptomics, proteomics and morphologic data.²⁷ Furthermore, we have shown that enteroids enriched for Paneth cells and goblet cells recapitulate *in vivo* characteristics on the proteomics level^{28,29} and that they are a useful tool for the investigation of health and disease related processes in specific intestinal cell types.²⁹

Nevertheless, the effect of Paneth cell and goblet cell enrichment of enteroids on key regulatory landscapes has not been extensively characterised. In this study, we comprehensively profiled mRNAs, miRNAs and lncRNAs from mouse derived, 3D conventionally differentiated enteroids (control), Paneth cell enriched enteroids (PCeE) and goblet cell enriched enteroids (GCeE) to determine the extent to which these enteroids display increased Paneth cell and goblet cell signatures. We applied a systems-level analysis of regulatory interactions within the PCeEs and GCeEs to further characterise the effect of cell type

enrichment and to predict key molecular regulators involved with Paneth cell and goblet cell specific functions. This analysis was carried out using interactions networks, which are a primary method to collate, visualise and analyse biological systems. These networks are a type of systems biology data representation, which aids the interpretation of -omics read-outs by contextualising genes/molecules of interest and identifying relevant signalling and regulatory pathways.^{30,31} In the presented analysis, the nodes of the interaction networks represent genes/molecules of interest from the transcriptomics data and the edges represent regulatory connections (molecular interactions) between the nodes inferred from databases. Studying regulatory interactions using interaction networks has proven useful to uncover how cells respond to changing environments at a transcriptional level, to prioritise drug targets and to investigate the downstream effects of gene mutations and knockouts.^{32–34}

We used network approaches to interpret the PCeE and GCeE transcriptomic data by integrating directed regulatory connections from resources containing transcriptional and post-transcriptional interactions. This integrative strategy led us to define regulatory network landscapes altered by Paneth cell and goblet cell enrichment of enteroids, termed the network and GCeE network, respectively (Fig. 1). By incorporating known Paneth cell and goblet cell markers, we used these networks to predict master regulators of Paneth cell and goblet cell differentiation and/or maintenance in the enriched enteroids. Furthermore, we highlighted varying downstream actions of shared regulators between the cell types. This phenomenon, called regulatory rewiring, highlights the importance of changes in regulatory connections in the function and differentiation of specific cell types. Taken together, we show that cell type enriched enteroids combined with the presented network biology workflow have potential for application to the study of epithelial dysfunction and mechanisms of action of multifactorial diseases such as IBD.

Results

Secretory lineages are over-represented in cell type enriched enteroids compared to conventionally differentiated controls

We generated 3D self-organising enteroid cultures *in vitro* from murine small intestinal crypts (Fig. S1, ESI[†]).^{18–20} In addition to conventionally differentiated enteroids, we generated enteroids enriched for Paneth cells and goblet cells using well-established and published protocols, presented in detail in the Methods.^{20,25} Bulk transcriptomics data was obtained from each set of enteroids to determine genes with differential expression resulting from enteroid skewing protocols. Differentially expressed genes were calculated by comparing the RNA expression levels (including protein coding genes, lncRNAs and miRNAs) of enteroids enriched for Paneth cells or goblet cells to those of conventionally differentiated enteroids. 4135 genes were differentially expressed (absolute log₂ fold change ≥ 1 and false discovery rate ≤ 0.05) in the PCeE dataset, and 2889 were



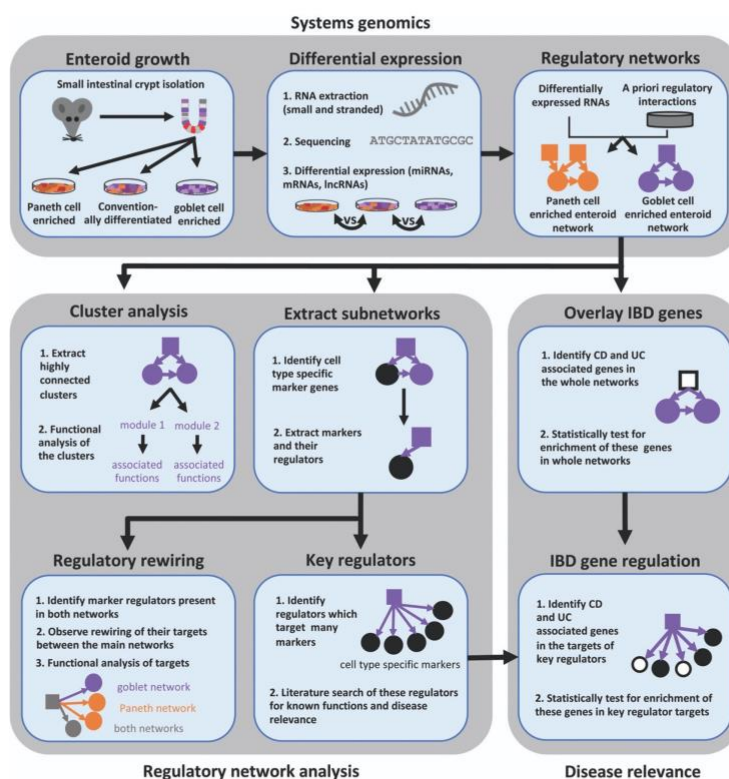


Fig. 1 Schematic representation of the workflow used to infer and analyse regulatory network landscapes altered by Paneth cell and goblet cell enrichment of enteroids. PCeE/GCeE network – Paneth cell enriched enteroid/goblet cell enriched enteroid network; TF – transcription factor; lncRNA – long non-coding RNA; miRNA – microRNA; mRNA – messenger RNA; UC – ulcerative colitis; CD – Crohn's disease.

differentially expressed in the GCeE dataset (Fig. 2A–C and Table S1, ESI†). The larger number of differentially expressed genes (DEGs) in the PCeE data could be attributed to the highly specialised nature of Paneth cells.^{35,36} The majority of the DEGs were annotated as protein coding: 79% in the PCeE dataset and 84% in the GCeE dataset. In addition, we identified lncRNAs (PCeE, 11%; GCeE, 9%) and miRNAs (PCeE, 4%; GCeE, 2%) among the DEGs (Fig. 2B). Some of these DEGs were identified in both the PCeE and the GCeE datasets, exhibiting the same direction of change compared to the conventionally differentiated enteroid data. In total, 1363 genes were found upregulated in both the PCeE and the GCeE data, while 442 genes were found downregulated in both datasets (Fig. 2C). This result highlights considerable overlap between the results of skewing enteroids towards Paneth cells and goblet cells, and can be explained by the shared differentiation history and secretory function of both Paneth cells and goblet cells.

Pathway analysis was employed to study functional associations of the DEGs (Fig. 2D). The PCeE-specific DEGs were associated with a number of metabolic pathways, including metabolism of vitamins and cofactors, pyruvate metabolism and citric acid (TCA) cycle and cholesterol biosynthesis. On the other hand, GCeE-specific DEGs were associated with the cell cycle through pathways such as cell cycle checkpoints, DNA replication and G1/S transition. Pathways associated with the shared DEGs included transmission across chemical synapses, integration of energy metabolism and a number of pathways linked to hormones. As hormone functions are characteristic of enteroendocrine cells, this analysis suggests that enteroendocrine cells are enriched in both the PCeEs and the GCeEs.

To validate the cell types present in the enteroids, the expression of five previously reported major cell type specific markers were investigated across the enteroids using transcript



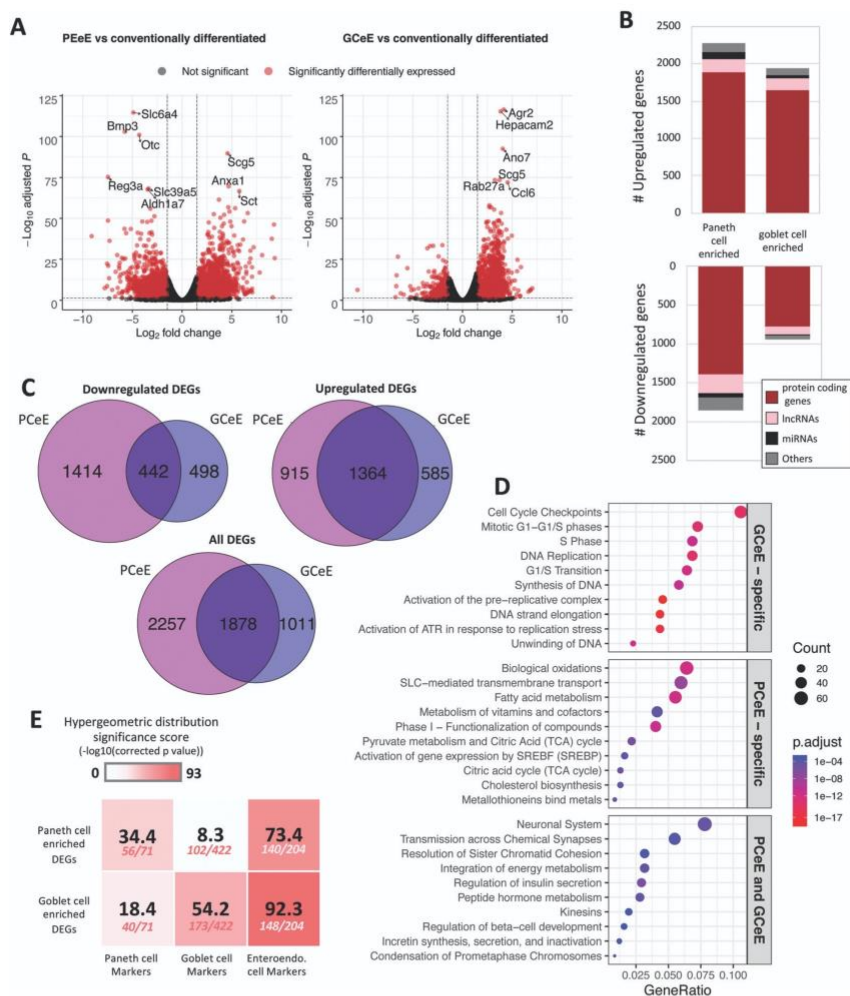


Fig. 2 Differentially expressed genes in Paneth cell enriched enteroids (PCEEs) and goblet cell enriched enteroids (GCEEs) (compared to conventionally differentiated enteroids). (A) Volcano plots showing log₂ fold change and adjusted *p* value for each gene following differential expression analysis of PCEEs (left) and GCEEs (right). Horizontal and vertical lines indicate the differential expression criteria cut offs (*q* value ≤ 0.05 and log₂ fold change ≥ |1|). (B) Number of differentially expressed genes (DEGs). miRNA = microRNA; lncRNA = long non-coding RNA; genes annotated as 'other' include pseudogenes and antisense genes. (C) Venn diagrams indicating the number of DEGs (passing the cut off criteria). (D) Top 10 Reactome pathways of the 50 most significant DEGs (by *q* value). (E) Enrichment of cell type specific marker genes in the DEG lists. Higher significance scores indicate greater enrichment. Number of markers in DEG list out of the total number of markers shown below significance score. Also see Tables S1 and S4 (ESI[†]).

abundances and RNA differential expression results (Table S2 and Fig. S2, ESI[†]). The control enteroids and the cell type enriched enteroids expressed all five investigated markers: *Lgr5* (stem cells), *Chga* (enteroendocrine cells), *Muc2* (goblet cells),

Lyz1 (Paneth cells) and *Vil1* (epithelial cells). We observed an upregulation of *Muc2*, *Lyz1* and *Chga* and a downregulation of *Lgr5* in PCEEs and GCEEs compared to the control enteroids, confirming the more pronounced differentiated status of the enteroids.

In addition, a number of Paneth cell specific antimicrobial peptide genes were differentially expressed in the PCeE dataset, including *Ang4*, *Reg3γ*, *Pla2g2a* and *Defa2* (Table S3, ESI†). Some of these genes were also differentially expressed in the GCeE dataset but with smaller log fold change values, e.g. *Lyz1* and *Ang4*. Conversely, a number of goblet cell mucin related genes

(including *Muc2* and *Tff3*) were differentially expressed in both datasets although all genes exhibited a smaller increase in the PCeEs (Table S2, ESI†). Therefore, using primary cell type specific markers, antimicrobial peptide genes and mucin-related genes, we show that the enteroids contain all major cell types, and that Paneth cells are most upregulated in the

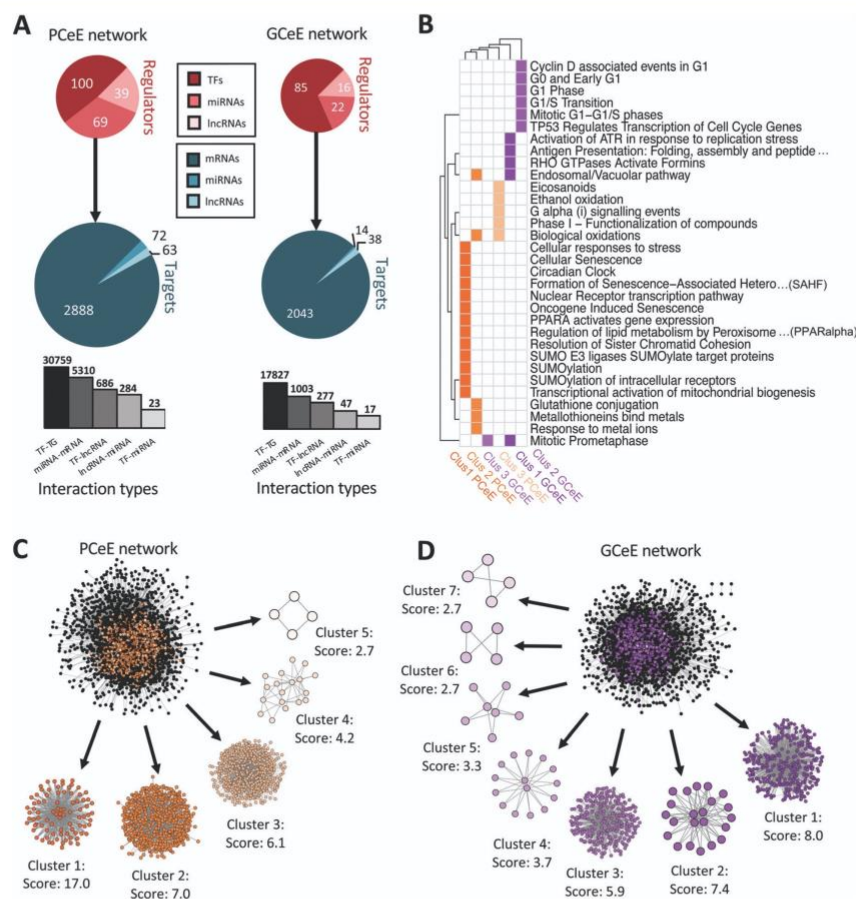


Fig. 3 Summary and cluster analysis of regulatory network for Paneth cell enriched enteroid (PCeE) and goblet cell enriched enteroid (GCeE) datasets. (A) Summary of number of nodes and interactions in the whole PCeE (left) and GCeE (right) networks. Total number of each regulator type shown in red, number of each target type shown in blue. In the targets pie-chart, mRNAs represent protein coding genes and proteins, miRNAs represent miRNAs genes and lncRNAs represent lncRNA genes. Size of circles represents log₁₀ (total unique regulators/targets). Bar chart represents the distribution of interaction types in the networks (log₁₀ scale). (B) Heatplot of Reactome pathways significantly associated (q value ≤ 0.05) with each cluster of the PCeE (orange) and GCeE (purple) networks. Only the top 5 pathways shown for each group (or more where equal q values). Only the top 3 clusters had significantly associated pathways. Clusters labelled with rank and cell type and colours match the colour of the cluster shown in (C and D). (C and D) Visualisation of the PCeE and GCeE regulatory networks with their associated clusters. The cluster rank and score is given next to each cluster. Black nodes in the whole networks represent nodes which were not found in any cluster, whereas coloured nodes represent the cluster which they are part of. TF – transcription factor; miRNA – microRNA; lncRNA – long non-coding RNA. See Items S1, S2 and Table S5 (ESI†).



PCeEs, while goblet cells are most upregulated in the GCeEs. We also note that both differentiation methods resulted in increases of other secretory cell types as well.

To further investigate secretory cell type specific signatures of the enteroids, we measured enrichment of secretory lineages in the upregulated DEG lists using additional marker genes of Paneth cells, goblet cells and enteroendocrine cells. While all tests were significant (hypergeometric model, q value ≤ 0.05), we identified greater enrichment of Paneth cell markers in the PCeE DEG list and goblet markers in the GCeE DEG list (Fig. 2E). This confirms that both enteroid enrichment protocols were successful in increasing the proportion of their target cell type, but also increased proportions of other secretory lineages, albeit to a lesser extent (Table S4, ESI[†]). This observation confirms previous studies that these enteroid differentiation protocols result in enteroendocrine enrichment in addition to Paneth cell and goblet cell enrichment.^{25,28}

In conclusion, we have used image-based validation, pathway analysis and marker gene investigation to show successful enrichment of target cell types in the PCeE's and GCeE's. We also highlighted an additional increase in other secretory lineages, particularly enteroendocrine cells, as a result of both enrichment protocols.

Reconstruction of regulatory networks altered by enteroid differentiation skewing

To gain an understanding of the regulatory changes occurring when enteroids development is skewed, we applied a network biology approach, and identified regulator–target relationships within the DEG lists. First, we generated a large network of non-specific molecular interactions known to occur in mice, by collating lists of published data (Table S6, ESI[†]). The resulting network (termed the universal network) consisted of 1 383 897 unique regulatory interactions connecting 23 801 molecular entities. All interactions within the network represent one of the following types of regulation: TF–TG, TF–lncRNA, TF–miRNA, miRNA–mRNA or lncRNA–miRNA. TF–TGs and TF–lncRNAs make up the majority of the network at 77% and 11% of all interactions, respectively. Due to its non-specific nature, this universal network contains many interactions not relevant for the current biological context. In order to get a clearer and valid picture of regulatory interactions occurring in our enteroids, we used the universal network to annotate the PCeE DEGs and GCeE DEGs with regulatory connections. Combining these connections, we generated specific regulatory networks for PCeEs and GCeEs, where every node is a DEG and every interaction has been observed in mice previously.

In total, the PCeE network, generated using differential expression data from the PCeEs compared to the conventionally differentiated enteroids, contained 37 062 interactions connecting 208 unique regulators with 3023 unique targets (Fig. 3A and Item S1, ESI[†]). The GCeE network, generated using differential expression data from the GCeEs compared to the conventionally differentiated enteroids, contained 19 171 interactions connecting 124 unique regulators with 2095 unique targets (Fig. 3A and Item S1, ESI[†]). 15.7% of all interactions (8856 out of 56 234)

were shared between the PCeE and GCeE networks, however the interacting molecular entities in these interactions (termed nodes) did not all exhibit the same direction of differential expression between the networks (comparing PCeE or the GCeE data to the conventionally differentiated enteroid data). In each of the enriched enteroid regulatory networks, a particular gene was represented (as a node in the network) only once, but may have been involved in multiple different interactions. In different interactions, a single node could act either as a regulator or as a target and in different molecular forms, for example, as a lncRNA in one interaction and as a target gene in another.

To further investigate the makeup of these networks, we employed cluster analysis to identify highly interconnected regions (possible regulatory modules) in the PCeE and GCeE regulatory networks. Using the MCODE software,³⁷ we identified five distinct clusters in the PCeE network and seven distinct clusters in the GCeE networks. A total of 1314 nodes are present in the PCeE network clusters and 698 in the GCeE network clusters. Functional analysis identified Reactome pathways³⁸ associated with each of the modules. Significant pathways (q value ≤ 0.05) were identified only for the highest ranked three modules from each network, with a total of 12 pathways shared between the PCeE and GCeE associated clusters (out of 32 associated with the PCeE clusters and 42 with the GCeE clusters) (Fig. 3B–D). Of particular note, the first cluster of the GCeE network has associations with the endosomal/vacuolar pathway and antigen presentation, the second cluster is associated with the cell cycle. Of the PCeE clusters, the first cluster is associated with a range of functions including nuclear receptor transcription pathway, regulation of lipid metabolism and senescence. The second is associated with response to metal ions and endosomal/vacuolar pathway and the third with G alpha (i) signalling events (Table S5, ESI[†]).

In conclusion, we have generated regulatory interaction networks, including transcriptional and post-transcriptional interactions, which illustrate the effect of skewing enteroid differentiation towards Paneth cell and goblet cell lineages.

Identification of potential cell type specific master regulators

Through pathway and marker analysis we predicted that our PCeE and GCeE datasets (*i.e.* DEG lists), and consequently our regulatory networks, contain signatures from the cell type of interest as well as additional noise from other secretory lineages. To focus specifically on the cell type specific elements of the networks, we used previously identified cell type specific markers to extract predicted Paneth cell and goblet cell regulators from our PCeE and GCeE networks. As cell type specific markers represent genes performing functions specific to a particular cell type, we expected that the regulators of these marker genes will have an important role in determining the function of said cell type. To identify these regulators, we extracted from the PCeE and GCeE networks, all relevant cell type specific markers and their direct regulators. These new networks were termed the Paneth cell subnetwork and goblet cell subnetwork respectively. The Paneth cell subnetwork



Molecular Omics

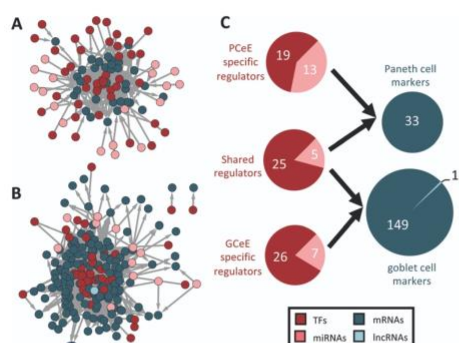


Fig. 4 Regulator–marker subnetworks for Paneth cell and goblet cell datasets. (A and B) Paneth cell (A) and goblet cell (B) subnetworks. Nodes represent genes, transcription factors or RNAs and edges represent directed physical regulatory connections. Regulators are shown in red and pink. Cell type specific markers are shown in blue. (C) Summary of the number of nodes present in both the subnetworks. Paneth cell data above and goblet cell data below. Total number of each regulator type shown in red, number of each target type shown in blue. Regulators have been categorised based on their membership in the two subnetworks – shared regulators are present in both networks. In the targets pie-chart, mRNAs also represent protein coding genes. Size of circles represents log₁₀ (total unique regulator/targets). TF – transcription factor; miRNA – microRNA; lncRNA – long non-coding RNA.

contained 33 markers specific for Paneth cells with 62 possible regulators. The goblet cell subnetwork contained 150 markers with 63 possible regulators (Fig. 4 and Table S7, ESI†). Observing the ratio of regulators and markers, the Paneth cell subnetwork had, on average, 1.88 regulators for each marker. On the other hand, the goblet cell subnetwork exhibited only 0.42 regulators for each marker. The quantity of markers identified in each subnetwork (33 in the Paneth network and 150 in the goblet network) correlates with the number of marker genes identified by Haber *et al.*¹² However, far fewer regulators were identified in the goblet cell subnetwork per marker than for the Paneth cell subnetwork. Whilst the underlying reason for this discrepancy is unknown, it could potentially be evidence of the complex regulatory environment required to integrate and respond to the arsenal of signals recognised by Paneth cells in comparison to goblet cells.³⁵

Of the 95 marker regulators, we identified approximately one-third (30/95) as present in both subnetworks (Fig. 4C). Given that the markers are different between the cell types, a regulator shared between the Paneth cell and goblet cell subnetworks must show an altered pattern of regulatory targeting in the two cell types. This phenomenon, referred to as regulatory rewiring, often results in functional differences of shared regulators in different environments³⁹ – for example, in this case, between the Paneth cells and goblet cells.

Further investigation of the distinct regulator–marker interactions highlighted a gradient of regulator specificity. We generated matrices to visualise the markers controlled by

each regulator in the goblet cell (Fig. 5A and C) and the Paneth cell (Fig. 5B) subnetworks. Each coloured square indicates that a marker (shown on the y-axis) is regulated by the corresponding regulator (shown on the x-axis). Squares are coloured blue if the associated regulator is shared between the Paneth cell and goblet cell subnetworks and orange if they are specific to one subnetwork. A collection of regulators (both subnetwork specific and shared) appear to regulate large proportions of the markers. For example, Ets1, Nr3c1 and Vdr regulate > 50% of the markers in both the Paneth cell and the goblet cell subnetworks. Specific to the Paneth cell subnetwork, Cebpa, Jun, Nr1d1 and Rxra regulate > 50% of the markers. Specific to the goblet cell subnetwork, Gfi1b and Myc regulate > 50% of the markers. These regulators represent potential master regulators of differentiation or maintenance of the given cell types in the enriched enteroids. Referring back to the highly-interconnected clusters identified in the PCeE and GCeE networks (Fig. 3C and D), we find these predicted master regulators in different clusters. In the PCeE network, Cebpa, Nr1d1, Nr3c1 and Rxra are in cluster 1, Vdr is in cluster 2, Jun is in cluster 3 and Ets1 is unclustered. In the GCeE network, Ets1 and Myc are in cluster 1, Nr3c1 and Vdr are in cluster 2 and Gfi1b is in cluster 3. This suggests a wide range of central functions are carried out by this group of regulators, with possible divergence of roles between the Paneth cell and the goblet cell. In contrast to the predicted master regulators, regulators such as Mafk in the Paneth cell subnetwork and Spdef in the goblet cell subnetwork regulate only one marker. These regulators likely have more functionally specific roles.

Together, these results highlight potential regulators which likely play key roles in specification and maintenance of Paneth and goblet cells and their functions in cell type enriched enteroids.

Regulators of cell type specific markers exhibit rewiring between Paneth cells and goblet cells

Cell type specific markers, which carry out cell type specific functions, are inherently different between the Paneth cell and goblet cell subnetworks (mutually exclusive). Therefore, the regulators observed in both Paneth cell and goblet cell subnetworks (shared regulators) are expected to target different marker genes. To do this, the regulators must have different regulatory connections in the different cell types, a phenomenon termed ‘rewiring’.⁴⁰ We extended the analysis to the original regulatory networks (PCeE and GCeE networks) to investigate whether any of the 30 identified shared regulators are rewired between the whole PCeE and the GCeE networks, and thus are highly likely to have different functions in the two types of enriched enteroids as well as between Paneth and goblet cells. To quantify rewiring of each of these regulators, we observed their targets in the PCeE and GCeE networks using the Cytoscape application, DyNet. DyNet assigns each regulator a rewiring score depending on how different their targets are between the two regulatory networks (Table S8, ESI†). Using these rewiring scores, we identified the five most rewired regulators (of 30) as Etv4, let-7e-5p, miR-151-3p, Myb and Rora.



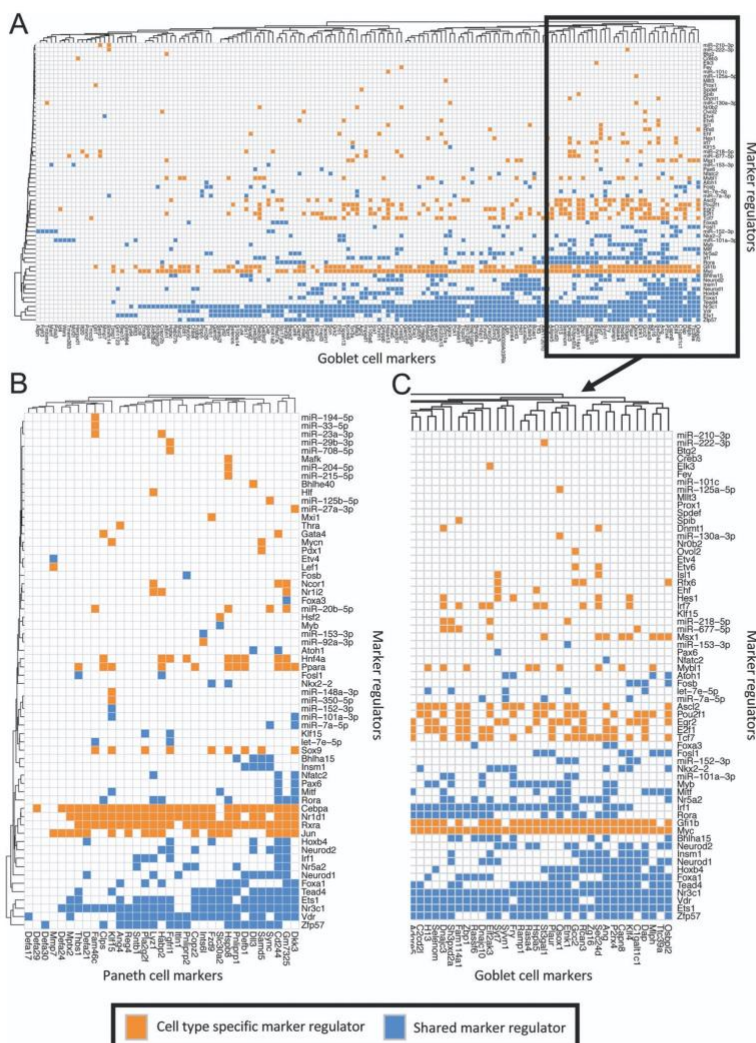


Fig. 5 Matrices of interactions between markers and their regulators in the Paneth cell and goblet cell subnetworks. Regulators (regulator targets) on y-axis, markers (regulator targets) on x-axis. Orange boxes indicate the interaction of a regulator and a marker where the regulator is only found in one of the two subnetworks. Blue boxes signify that the regulator is found in both the Paneth cell and the goblet cell subnetworks. (A) All goblet cell markers¹² and their regulators in the goblet cell subnetwork. (B) All Paneth cell markers¹² and their regulators in the Paneth cell subnetwork. (C) Sub-section of A showing the markers (and their regulators) which have the most regulatory connections. Interactions in Table S7 (ESI[†]).

Functional enrichment analysis was carried out on the targets of these regulators to test whether the targets specific to the PCeE and GCeE networks have different functions (hypergeometric model, q value ≤ 0.1) (Table S9, ESI[†]). Across all five

regulators the general trend indicated that targets specific to the PCeE network are associated with metabolism; targets specific to the GCeE network are associated with cell cycle and DNA repair. As pathway analysis carried out on the enteroid

Open Access Article. Published on 10 December 2019. Downloaded on 9/2/2020 3:33:23 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.



DEGs (Fig. 2D) identified the same phenomenon, this suggests that the rewired regulators could be key drivers of transcriptional changes during the skewing of enteroid differentiation towards Paneth cell or goblet cell lineages. In addition, given that the strongest signal of enriched enteroids represents their enriched cell type, we predict that these functions are key features of Paneth cells and goblet cells in the enteroids, and that the rewired regulators are important drivers of cell type specific functions.

Looking at the regulators in more detail, the GCeE specific targets of miR-151-3p, for example, are significantly enriched in functions relating to antigen presentation, cell junction organisation, Notch signalling and the calnexin/calreticulin cycle. None of these functions are enriched in the shared or PCeE targets. Of particular interest is the calnexin/calreticulin cycle, which is known to play an important role in ensuring proteins in the endoplasmic reticulum are correctly folded and assembled.⁴¹ Dysfunction of protein folding and the presence of endoplasmic reticulum stress are both associated with IBD.^{42–44} Therefore, we predict that miR-151-3p plays a role in the secretory pathway of goblet cells and could be an interesting target for IBD research. In addition, different functional profiles were also observed for the targets of Rora in the PCeE and GCeE regulatory networks: targets present in both networks are significantly associated with mitosis, whereas those specific to the PCeE network are associated with metabolism, protein localisation, nuclear receptor transcription pathway, circadian clock and hypoxia induced signalling. GCeE specific targets of Rora are connected to Notch signalling, cell cycle and signalling by Rho GTPases (associated with cell migration, adhesion and membrane trafficking) and interferon.

Altogether these observations show that some of the regulators of both Paneth cell and goblet cell marker genes have different targets (with different associated functions) between

the PCeE and the GCeE networks. This suggests that regulatory rewiring occurs between Paneth cell and goblet cell types.

Evaluating the disease relevance of the subnetwork specific master regulators

To investigate the function and relevance of the predicted master regulators in IBD, we carried out three analyses: (1) a literature search to check what is known about the identified master regulators; (2) an enrichment analysis to evaluate the disease relevant genes in the PCeE and GCeE networks and among the targets of the predicted master regulators; and finally, (3) a comparative analysis with human biopsy based single cell dataset to confirm the relevance of the data we identified PCeE and GCeE networks.

The literature search was carried out using the three groups of predicted master regulators: those specific to the Paneth cell markers (*Cebpa*, *Jun*, *Nr1d1* and *Rxra*), those specific to the goblet cell markers (*Gfi1b* and *Myc*) and those which appear to regulate many of the markers of both cell types (*Ets1*, *Nr3c1* and *Vdr*). We identified five genes (*Ets1*, *Nr1d1*, *Rxra*, *Nr3c1* and *Vdr*) with associations to inflammation, autophagy and/or inflammatory bowel disease (IBD), as shown in Table 1. These genes correspond to 71% (5/7) of the Paneth cell associated master regulators and 60% (3/5) of the goblet cell associated master regulators. Interestingly, four of these genes (all apart from *Ets1*), encode nuclear hormone receptors.

Given the possible relationship between the identified master regulators and IBD, we tested the potential of the PCeE and GCeE regulatory networks to study the pathomechanisms of CD or UC. We checked for the presence of known CD or UC associated genes in the networks, using data from two studies of single nucleotide polymorphisms (SNPs)^{67,68} and one study of CD expression quantitative trait loci (eQTLs).⁶⁹ Using hypergeometric significance tests, we found that the PCeE network was significantly enriched in all tested lists: genes with UC

Table 1 Literature associations relating to autophagy, inflammation and IBD for putative master regulators

Putative master regulator	Autophagy/inflammation/IBD associations	Ref.
NR1D1 (REV-ERBa)	Modulates autophagy and lysosome biogenesis in macrophages leading to antimycobacterial effects	45
	SNP rs12946510 which has associations to IBD, acts as a <i>cis</i> -eQTL for NR1D1	46
NR3C1 (glucocorticoid receptor)	Associations with cellular proliferation and anti-inflammatory responses	47
	Exogenous glucocorticoids are heavily used as anti-inflammatory therapy for IBD	48 and 49
	ATG16L1, an autophagy related gene, was down-regulated in patients who do not respond to glucocorticoid treatment	50 and 51
VDR (vitamin D receptor)	Transcriptionally regulates NFKβ1, a SNP affected gene in ulcerative colitis	52 and 53
	Regulates autophagy in Paneth cells through ATG16L1 – dysfunction of autophagy in Paneth cells has been linked to Crohn's disease	54 and 55
	Induces antimicrobial gene expression in other cell lines	56 and 57
	Specific polymorphisms in the VDR genes have been connected to increased susceptibility to IBD	58
	A study looking at colonic biopsies of IBD patients observed reduced VDR expression compared to healthy biopsies	59
RXRa (retinoid X receptor alpha)	Interacts with five SNP affected UC genes	60 and 61
	Heterodimerizes with VDR (see above)	62
ETS1 (ETS proto-oncogene 1)	Important role in the development of hematopoietic cells and Th1 inflammatory responses	63 and 64
	Angiogenic factors in the VEGF-Ets-1 cascades are upregulated in UC and downregulated in CD	65
	IBD susceptibility gene	66



associated SNPs (13/47, $p < 0.005$), genes with CD associated SNPs (22/97, $p < 0.005$) and genes with CD associated eQTLs (290/1607, $p < 0.0001$) (Tables S10–S12, ESI†). On the other hand, we found that the GCeE network was significantly enriched in genes with UC associated SNPs (10/47, $p < 0.005$) but regarding CD, the genes with SNP associations were not significantly enriched (12/97, $p = 0.11$) and the genes with qQTL associations were enriched with a larger p value ($p < 0.05$) (Tables S10–S12, ESI†).

Next we investigated whether any of the genes with UC or CD associated SNPs acts as regulators in the PCeE and GCeE networks. Of the genes with CD associated SNPs, one acts as a regulator in each network. Similarly, two genes with UC associated SNPs act as regulators in the networks. Specifically regarding CD associated genes, in the PCeE network, the gene *Dbbp* regulates *Bik*, which encodes the BCL2 interacting killer, a pro-apoptotic, death promoting protein. In the GCeE network, *Notch2* regulates *Notch3* and *Hes1*. Specifically, regarding UC associated genes, in the PCeE network, *Hnf4a* regulates 994 genes/RNAs including nine Paneth cell markers (*Cd244a*, *Fgfr1*, *Clps*, *Habp2*, *Hspb8*, *Pnliprp1/2*, *Defb1*, *Mymx*) and one other gene with UC associated SNPs (*Tnfsf15*). Additionally, a gene with UC associated SNPs, *Nr5a2*, was found in both the PCeE and GCeE networks regulating 389 and 276 genes/RNAs respectively. In the PCeE network *Nr5a2* targets include 6 Paneth cell markers (*Cd244a*, *Copz2*, *Pnliprp1/2*, *Sntb1*, *Mymx*). Ultimately, the large number of targets of these regulatory UC associated genes suggests they have wide ranging effects on the regulatory network of Paneth and goblet cells. To further establish the relevance of the inferred PCeE and GCeE networks, we also found an over-representation of drug target associated genes in both the PCeE and GCeE networks (2683/16223 and 1918/16223 respectively, $p < 0.0001$), highlighting their potential for the study of therapeutic implications (Table S12, ESI†).

To investigate the link between predicted master regulators and IBD, we observed whether the genes with UC and CD associated SNPs are regulated by the predicted master regulators in the PCeE and GCeE networks (Table S13, ESI†). Given that Paneth cell dysregulation is classically associated with CD and goblet cell dysregulation/depletion with UC,⁷⁰ we focused this analysis only on these pairings, examining CD genes amongst targets of Paneth cell predicted master regulators, and UC genes amongst targets of goblet cell predicted master regulators. In the PCeE network, we found 21 (of 22) of the CD genes were regulated by at least one of the seven Paneth cell predicted master regulators, while the targets of these master regulators were significantly enriched with CD genes in the PCeE network ($p < 0.001$). Similarly, we observed that all 10 UC genes in the GCeE network were regulated by at least one of the five goblet cell predicted master regulators, while the targets of these master regulators were significantly enriched with UC genes ($p < 0.005$).

To confirm the relevance of these predicted master regulators in a human system, a similar analysis was carried out using goblet cell differentially expressed genes from a recent single cell study of human inflamed UC colon biopsies.¹³ Using the

top 100 differentially expressed genes, following conversion to mouse Ensembl identifiers, 20 were found to be targeted by the predicted goblet cell master regulators in the GCeE network. This represents a significant enrichment amongst all master regulator targets ($p < 0.005$) (Table S13, ESI†).

Ultimately, by integrating functional annotations obtained through literature searches, we show that the Paneth cell and goblet cell regulatory networks contain genes with direct and indirect associations with IBD. Furthermore, we find that the PCeE and GCeE networks and the targets of predicted master regulators are enriched with IBD associated genes – this finding is corroborated using human single cell data from UC colon biopsies. Consequently, these networks and the workflow to reconstruct and analyse them have great potential for the study of IBD pathomechanisms in specific intestinal cell types.

Discussion

By generating and integrating cell type enriched enteroid RNAseq datasets with regulatory networks, we characterised the regulatory environment altered by differentiation skewing. Through focusing on cell type specific markers, we used the regulatory networks to predict master regulators of Paneth cells and goblet cells and to highlight the role of regulatory rewiring in cell differentiation. Given the relevance of Paneth cell and goblet cell dysfunction in IBD, future application of cell type enriched enteroids combined with our network analysis workflow can be used to disentangle multifactorial mechanisms of IBD.

Analysis using known cell type specific markers confirmed that skewing differentiation of enteroids towards Paneth cells or goblet cells results in an increase in the target cell signatures at the transcriptomics level. In addition, signatures of entero-endocrine cells were increased in both differentiation protocols, albeit at a lower amount than the target cell type. This finding correlates with previous investigations at both the transcriptomic and proteomic levels^{25,27–29} and is likely due to the shared differentiation pathways of these secretory cells. Nevertheless, as enteroids contain a mixed population of cell types by nature and because intercellular communication is key to a functioning epithelium,^{19,71} the increased proportion of non-targeted secretory lineages should not be an issue for the application of these models to research. In fact, the enrichment of specific cell types is beneficial for enteroid-based research to increase the signal originating from a specific population of cells and to provide a larger population of cells of interest for downstream single cell analysis of enteroids, which is particularly beneficial when studying rare populations such as Paneth cells. This is valuable due to the lack of *in vitro* models for long-term culture of non-self-renewing small intestinal epithelial cells.^{72,73} Specifically, the comparison of ‘omics data from a cell type enriched enteroid to a conventionally differentiated enteroid enables generation of cell type signatures with more specificity than can be obtained otherwise – except through single cell sequencing. Single cell sequencing, however, comes



at a greater financial cost and provides lower coverage which can be problematic for rare cell types and lowly expressed RNAs. Furthermore, a number of previous studies have shown that these cell type enriched enteroid models, which offer a simplified and manipulatable version of the intestinal environment, are useful for the investigation of health and disease related processes.^{26,28,29} It must be considered, however, that through applying chemical inhibitors to enteroids to enrich particular cell types, we may observe changes in gene expression which are related to the direct effects of the inhibitor but not to differentiation. Due to the nature of the inhibitors, it would be challenging to separate these effects. For example, CHIR99021, which is used for enriching Paneth cells, is a direct inhibitor of glycogen synthase kinase 3 (GSK-3). GSK-3 has a role in many cellular pathways including cellular proliferation and glucose regulation.⁷⁴

Using *a priori* molecular interaction knowledge, we annotated differentially expressed genes (cell type enriched enteroids vs. conventionally differentiated enteroids) with regulatory connections. Collecting all connections together generated regulatory networks for the Paneth cell enriched enteroid (PCeE) and goblet cell enriched enteroid (GCeE) datasets. This approach to collating networks (regulatory or otherwise) has been used for a wide variety of research aims, such as the identification of genes functioning in a variety of diseases,^{75,76} the prioritisation of therapeutic targets⁷⁷ and for a more general understanding of gene regulation in biological systems.^{78,79} The application of prior knowledge avoids the need for reverse engineering/inference of regulatory network connections, which is time consuming, computationally expensive and requires large quantities of high quality data.⁸⁰ To investigate the substructure and functional associations of the generated PCeE and GCeE regulatory networks we applied a clustering approach. The identified clusters represent collections of highly interconnected nodes, which likely form regulatory modules. Functional analysis confirmed distinct functional associations between the clusters as well as between the networks. The observation that less than half of the network nodes exist in clusters is consistent with the view that regulatory networks are hierarchical and scale free with most genes exhibiting low pleiotropy.^{81,82}

Given the observed additional increase in secretory lineages based on the DEGs of the enriched organoids, we chose to use cell type specific markers to extract interactions specific to Paneth cells and goblet cells from the generated regulatory networks. This enables further enrichment of Paneth cell and goblet cell signatures and reduction in noise in the networks due to the presence of other cell types in the enteroids. Using this approach, we identified possible regulators of cell type specific functions in Paneth cells and goblet cells. Some of these regulators were predicted to be important in both cell types, but exhibited differential targeting patterns between the PCeE and the GCeE networks, indicating rewiring of regulators between the cell types. This highlights apparent redundancy and/or cooperation of regulators which control similar cell type specific functions and shows the potential importance of regulatory rewiring in the evolution of cell type specific

pathways and functions, something which has been shown previously to occur.^{83,84} Functional analysis of the targets of the most rewired regulators (Etv4, let-7e-5p, miR-151-3p, Myb and Rora) highlights an overrepresentation of metabolism associated targets in the PCeE network and cell cycle associated targets in the GCeE network. A similar result was observed when functional analysis was carried out on genes with significantly different expression levels between the cell type enriched enteroids and the conventionally differentiated enteroids (Fig. 2B). This suggests that transcriptional changes during the skewing of enteroid differentiation could be driven by rewired regulators and that these functions are key features of Paneth cells and goblet cells in the enteroids. The latter is supported by current understanding that Paneth cells rely on high levels of protein and lipid biosynthesis for secretory functions,⁸⁵ and they play an important role in metabolically supporting stem cells.⁸⁶ Additionally, as terminally-differentiated cells do not undergo cell division, this result suggests that enteroid goblet cell signatures are derived from a large population of semi-differentiated goblet-like cells, a phenomenon previously observed in tissue sample based studies.^{13,87} Extension of our workflow to single cell sequencing of enteroid cells could validate these findings by providing greater cell type specificity. However, these techniques pose further technical and economic challenges.^{88,89} Specifically, a large number of organoids must be sequenced to mitigate cellular complexity and batch heterogeneity and powerful, reproducible and accurate computational pipelines are required to analyse such data.⁹⁰

We predicted key regulators involved in differentiation or maintenance of Paneth cells and goblet cells in the enteroids: Cebpa, Jun, Nr1d1 and Rxra specific to Paneth cells, Gfi1b and Myc specific for goblet cells and Ets1, Nr3c1 and Vdr shared between them. Validation of these regulators poses significant challenges due to their wide expression and broad function range. If the master regulators are controlling differentiation as opposed to cell function maintenance, evaluating lineage arrest or delay can be carried out using a gene knockout or knock down. However the effects of pleiotropy will significantly hamper the results and such a study would require significant follow-up studies. On the other hand, if key regulators were predicted by applying the presented computational workflow to condition-specific organoids compared to control organoids (*e.g.* drug treated organoids vs. non-treated organoids), the validation would be much simpler. Literature investigation highlighted that many of the predicted master regulators, particularly those associated with Paneth cells, have connections to autophagy, inflammation and IBD (Table 1). While some of these associations are related to different cell types, we assume that if a gene can contribute to a specific function in one cell type it may also contribute in another. This finding suggests that dysregulation of key cell master regulators could contribute to IBD.

To further investigate this finding, we identified Crohn's disease (CD) and ulcerative colitis (UC) genes in the PCeE and GCeE networks. We found that CD associated genes are more strongly associated with the PCeE network than the GCeE



network. Given that Paneth cell dysfunction is classically associated with CD, this finding highlights the relevance of the generated networks to the *in vivo* situation. In the PCeE network one SNP associated CD gene, *Dbp*, acts as a regulator. *Dbp*, encoding the D site binding protein, regulates *Bik*, which encodes the BCL2 interacting killer, a pro-apoptotic, death promoting protein. Interestingly, rate of apoptosis has been implicated in IBD disease mechanisms⁹¹ and has been associated with IBD drug response.⁹² Therefore, this finding highlights a possible regulatory connection between CD susceptibility genes and IBD pathology on a Paneth cell specific level. In the GCeE network, the SNP associated CD gene *Notch2* acts as a regulator for *Notch3* and *Hes1*. It has been previously demonstrated that this pathway can block glucocorticoid resistance in T-cell acute lymphoblastic leukaemia via NR3C1 (predicted master regulator).⁹³ This is relevant to IBD given that glucocorticoids are a common treatment for IBD patients.⁴⁹ Furthermore, this pathway has been previously associated with goblet cell depletion in humans⁹⁴ commonly observed in IBD patients. Furthermore, we identified a significant enrichment of UC associated genes in both the PCeE and GCeE networks. The majority of UC associated genes identified in the networks (9/14) were present in both, suggesting that genetic susceptibilities of UC do not have a Paneth cell or goblet cell specific effect. Two of the identified UC associated genes act as regulators in the networks (*Nr5a2* and *Hnf4a*), targeting hundreds of genes and thus suggesting a broad ranging effect on the networks. Building on the identified literature associations of predicted master regulators, we found that the targets of Paneth cell master regulators are enriched with CD associated genes, and the targets of the goblet cell master regulators are enriched with UC associated genes. This finding was further illustrated using UC associated goblet cell genes from a human biopsy study,¹³ highlighting the relevance of these findings in a human system. Ultimately, the observation of IBD susceptibility genes in the regulatory networks of these enteroids highlights possible application of this model system to study disease regulation in specific intestinal cell types, through understanding specific mechanistic pathways.

We have shown how network biology techniques can be applied to generate interaction networks representing the change in regulatory environments between two sets of enteroids. Here we presented this workflow, and by using transcriptomics data we characterised the effect of Paneth cell and goblet cell differentiation skewing protocols on enteroids. However, the described workflow could be applied to a variety of 'omics datasets and enteroid conditions. For example, to test the response of enteroids to external stimuli, such as bacteria, and on enteroids grown from human-derived biopsies, enabling patient-specific experiments. The application of further 'omics data-types to the described approach could generate a more holistic view of cellular molecular mechanisms, including the ability to observe post-translational regulation. In this study, we integrated miRNA and lncRNA expression datasets, in addition to mRNA data, as these molecules have been shown to perform critical regulatory and mediatory functions in maintaining intestinal homeostasis.^{46,68,95} However, only small proportions of the generated PCeE and the GCeE

networks contained miRNA and lncRNA interactions (Fig. 2B), due to lack of published interaction information, particularly from murine studies. Both the application of human enteroid data and the future publication of high-throughput interaction studies involving miRNAs and lncRNAs will improve the ability to study such interactions. Nevertheless, these network approaches are beneficial for contextualising gene lists through annotating relevant signalling and regulatory pathways³¹ and we can use them to represent and analyse current biological knowledge, to generate hypotheses and to guide further research.

In conclusion, we described an integrative systems biology workflow to compare regulatory landscapes between enteroids from different conditions, incorporating information on transcriptional and post-transcriptional regulation. We applied the workflow to compare Paneth cell and goblet cell enriched enteroids to conventionally differentiated enteroids and predicted Paneth cell and goblet cell specific regulators, which could provide potential targets for further study of IBD mechanisms. Application of this workflow to patient derived organoids, genetic knockout and/or microbially challenged enteroids, alongside appropriate validation and single cell sequencing if available, will aid discovery of key regulators and signalling pathways of healthy and disease associated intestinal cell types.

Methods

Animal handling

C57BL/6J mice of both sexes were used for enteroid generation. All animals were maintained in accordance with the Animals (Scientific Procedures) Act 1986 (ASPA).

Small intestinal organoid culture

Murine enteroids were generated as described previously.^{18,20,29} Briefly, the entire small intestine was opened longitudinally, washed in cold PBS then cut into ~5 mm pieces. The intestinal fragments were incubated in 30 mM EDTA/PBS for 5 minutes, transferred to PBS for shaking, then returned to EDTA for 5 minutes. This process was repeated until five fractions had been generated. The PBS supernatant fractions were inspected for released crypts. The crypt suspensions were passed through a 70 µm filter to remove any villus fragments, then centrifuged at 300 × *g* for 5 minutes. Pellets were resuspended in 200 µl phenol-red free Matrigel (Corning), seeded in small domes in 24-well plates and incubated at 37 °C for 20 minutes to allow Matrigel to polymerise. Enteroid media containing EGF, Noggin and R-spondin (ENR; (18)) was then overlaid. Enteroids were generated from three separate animals for each condition, generating three biological replicates.

To chemically induce differentiation, on days two, five and seven post-crypt isolation, ENR media was changed to include additional factors for each cell type specific condition: 3 µM CHIR99021 (Tocris) and 10 µM DAPT (Tocris) [Paneth cells]; 2 µM IWP-2 (Tocris) and 10 µM DAPT [goblet and entero-endocrine cells].²⁵



Small intestinal organoid immunofluorescence

On day eight post-crypt isolation, enteroids were fixed with 4% paraformaldehyde (PFA; Sigma-Aldrich) for 1 hour at 4 °C prior to permeabilization with 0.1% Triton X-100 (Sigma-Aldrich) and incubated in blocking buffer containing 10% goat serum (Sigma-Aldrich). Immunostaining was performed overnight at 4 °C using primary antibodies: mouse anti-E-cadherin (BD Transduction Laboratories), rabbit anti-muc2 (Santa Cruz) and rabbit anti-lysozyme (Dako), followed by Alexa Fluor-488 and -594 conjugated secondary antibodies (ThermoFisher Scientific). DNA was stained with DAPI (Molecular Probes). Images were acquired using a fluorescence microscope (Axioimager.M2, equipped with a Plan-Apochromat 63×/1.4 oil immersion objective) and analysed using ImageJ/FIJI V1.51.

RNA extraction

On day eight post-crypt isolation (allowing optimal cell type-enrichment as previously shown),²⁵ enteroids were extracted from Matrigel (Corning, 356237) using Cell Recovery Solution (BD Bioscience, 354253), rinsed in PBS and RNA was extracted using miRCURY RNA Isolation Tissue Kit (Exiqon, 300115) according to the manufacturer's protocol.

Stranded RNA library preparation

The enteroid transcriptomics libraries were constructed using the NEXTflex™ Rapid Directional RNA-Seq Kit (PerkinElmer, 5138-07) using the polyA pull down beads from Illumina TruSeq RNA v2 library construction kit (Illumina, RS-122-2001) with the NEXTflex™ DNA Barcodes – 48 (PerkinElmer, 514104) diluted to 6 μM. The library preparation involved an initial QC of the RNA using Qubit DNA (Life technologies, Q32854) and RNA (Life technologies, Q32852) assays as well as a quality check using the PerkinElmer GX with the RNA assay (CLS960010). Ligated products were subjected to a bead-based size selection using Beckman Coulter XP beads (A63880). As well as performing a size selection this process removed the majority of unligated adapters. Prior to hybridisation to the flow cell the samples were amplified to enrich for DNA fragments with adapter molecules on both ends and to amplify the amount of DNA in the library. The strand that was sequenced is the cDNA strand. The insert size of the libraries was verified by running an aliquot of the DNA library on a PerkinElmer GX using the High Sensitivity DNA chip (PerkinElmer, CLS760672) and the concentration was determined by using a High Sensitivity Qubit assay and q-PCR. Libraries were then equimolar pooled and checked by qPCR to ensure the libraries had the necessary sequencing adapters ligated.

Small RNA library preparation

The small RNA libraries were made using the TruSeq Small RNA Library Prep Kits (Illumina, 15004197), six-base indexes distinguish samples and allow multiplexed sequencing and analysis using 48 unique indexes (Set A: indexes 1–12 (Illumina, RS-200-0012)), Set B: indexes 13–24 (Illumina, RS-200-0024), Set C: indexes 25–36 (Illumina, RS-200-0036), Set D: indices 37–48

(Illumina, RS-200-0048) (TruSeq Small RNA Library Prep Kit Reference Guide (Illumina, 15004197 Rev.G)). The TruSeq Small RNA Library Prep Kit protocol is optimised for an input of 1 μg of total RNA in 5 μl nuclease-free water or previously isolated microRNA may be used as starting material (minimum of 10–50 ng of purified small RNA). Total RNA is quantified using the Qubit RNA HS Assay kit (ThermoFisher, Q32852) and quality of the RNA is established using the Bioanalyzer RNA Nano kit (Agilent Technologies, 5067-1511), it is recommended that RNA with an RNA Integrity Number (RIN) value ≥ 8 is used for these libraries as samples with degraded mRNA are also likely to contain degraded small RNA.

Library purification combines using BluePippin cassettes (Sage Science Pippin Prep 3% Cassettes Dye-Free (CDF3010), set to collection mode range 125–160 bp) to extract the library molecules followed by a concentration step (Qiagen MinElute PCR Purification, 28004) to produce libraries ready for sequencing. Library concentration and size are established using HS DNA Qubit and HS DNA Bioanalyzer. The resulting libraries were then equimolar pooled and qPCR was performed on the pool prior to clustering.

Stranded RNA sequencing on HiSeq 100PE

The final pool was quantified using a KAPA Library Quant Kit (Roche, 07960140001), denatured in NaOH and combined with HT1 plus a 1% PhiX spike at a final running concentration of 10 pM. The flow-cell was clustered using HiSeq PE Cluster Kit v3 (Illumina, PE-401-3001) utilising the Illumina PE HiSeq Cluster Kit V3 cBot recipe V8.0 method on the Illumina cBot. Following the clustering procedure, the flow-cell was loaded onto the Illumina HiSeq2000 instrument following the manufacturer's instructions with a 101 cycle paired reads and a 7-cycle index read. The sequencing chemistry used was HiSeq SBS Kit v3 (Illumina, FC-401-3001) with HiSeq Control Software 2.2.68 and RTA 1.18.66.3. Reads in bcl format were demultiplexed based on the 6 bp Illumina index by CASAVA 1.8, allowing for a one base-pair mismatch per library, and converted to FASTQ format by bcl2fastq.

Small RNA sequencing on HiSeq rapid 50SE

The final pool was quantified using a KAPA Library Quant Kit (Roche, 07960140001), denatured in NaOH and combined with HT1 plus a PhiX spike at a final running concentration of 20 pM. The libraries were hybridized to the flow-cell using TruSeq Rapid Duo cBot Sample Loading Kit (Illumina, CT-403-2001), utilising the Illumina RR_TemplateHyb_FirstExt_VR method on the Illumina cBot. The flow-cell was loaded onto the Illumina HiSeq2500 instrument following the manufacturer's instructions with a 51-cycle single read and a 7 cycle index read. The sequencing chemistry used was HiSeq SBS Rapid Kit v2 (Illumina, FC-402-4022) with a single read cluster kit (Illumina, GD-402-4002), HiSeq Control Software 2.2.68 and RTA 1.18.66.3. Reads in bcl format were demultiplexed based on the 6 bp Illumina index by CASAVA 1.8, allowing for a one base-pair mismatch per library, and converted to FASTQ format by bcl2fastq.



Differentially expressed transcripts

The quality of stranded reads was assessed by FastQC software (version 0.11.4).⁹⁶ All reads coming from technical repeats were concatenated together and aligned (in stranded mode, *i.e.* with ‘-rna-strandness RF’ flag) using HISAT aligner (version 2.0.5).⁹⁷ Subsequently, a reference-based *de novo* transcriptome assembly was carried out for each biological repeat and merged together using StringTie (version 1.3.2) with following parameters: minimum transcript length of 200 nucleotides, minimum FPKM of 0.1 and minimum TPM of 0.1.^{98,99} Coding potential of each novel transcript was determined with CPC (version 0.9.2) and CPAT (version 1.2.2).^{100,101} From the novel transcripts, only non-coding transcripts (as predicted by both tools) were included in final GTF file. Gene and transcript abundances were estimated with kallisto (version 0.43.0).¹⁰² Sleuth (version 0.28.1) R library was used to perform differential gene expression.¹⁰³ mRNAs and lncRNAs with an absolute log₂ fold change of 1 and *q* value ≤ 0.05 were considered to be differentially expressed.

The small RNA reads were analysed using the sRNAbench tool within the sRNAtoolbox suite of tools.¹⁰⁴ The barcodes from the 5′ end and adapter sequences from the 3′ end were removed respectively. Zero mismatches were allowed in detecting the adapter sequences with a minimum adapter length set at 10. Only reads with a minimum length of 16 and a read-count of 2 were considered for further analysis. The mice miRNA collection was downloaded from miRBase version 21.¹⁰⁵ The trimmed and length filtered reads were then mapped to the mature version of the miRBase miRNAs in addition to the annotated version of the mouse genome (version mm10). No mismatches were allowed for the mapping. A seed length of 20 was used for the alignment with a maximum number of multiple mappings set at 10. Read-counts normalised after multiple-mapping were calculated for all the libraries. The multiple-mapping normalised read-counts from the corresponding cell type enriched enteroids were compared against the conventionally differentiated enteroids to identify differentially expressed miRNAs in a pair-wise manner using edgeR.¹⁰⁶ miRNAs with an absolute log₂ fold change of 1 and false discovery rate ≤ 0.05 were considered to be differentially expressed.

Differentially expressed genes were grouped by their presence in the PCeE dataset, the GCeE dataset or in both. Each group of differentially expressed genes was tested for functional enrichment (hypergeometric model, *q* value ≤ 0.1) based on Reactome and KEGG annotations using the ReactomePA R package^{38,107–109} following conversion from mouse to human identifiers using Inparanoid (v8).^{110,111}

Enrichment of marker genes

Cell type specific marker gene lists were obtained a mouse single cell sequencing survey.¹² The cell type specific signature genes for Paneth, goblet and enteroendocrine cell types were obtained from the droplet-based and the plate-based methods. Gene symbols were converted to Ensembl gene IDs using bioDBnet db2db.¹¹²

Hypergeometric distribution testing was carried out using a custom R script to measure enrichment of cell type specific marker genes in the differentially upregulated gene sets. To standardise the universal dataset, only markers which are present in the output of the Wald test (genes with variance greater than zero among samples) were used. Similarly, to enable fair comparisons, only differentially expressed protein coding genes and documented lncRNAs were used from the DEG lists, as was surveyed in the cell type specific marker paper. Bonferroni correction was applied to the hypergeometric distribution *p* values to account for multiple testing and significance scores were calculated using $-\log_{10}$ (corrected *p* value). For the mapping of marker genes to the interaction networks, no filters were applied.

Reconstruction of molecular networks

Mice regulatory networks containing directed regulatory layers were retrieved from multiple databases (Table S6, ESI†): miRNA–mRNA (*i.e.*, miRNAs regulating mRNAs) and lncRNA–miRNA (*i.e.*, lncRNAs regulating miRNAs) interactions were downloaded from TarBase v7.0¹¹³ and LncBase v2.0,¹¹⁴ respectively. Only miRNA–mRNA and lncRNA–miRNA interactions determined using HITS-CLIP¹¹⁵ experiments were considered. Regulatory interactions between transcription factors (TFs) and miRNAs (*i.e.* TFs regulating miRNAs) were retrieved from TransmiR v1.2,¹¹⁶ GTRD¹¹⁷ and TRRUST v2.^{118,119} Co-expression based inferences were ignored from all the above resources. Transcriptional regulatory interactions (*i.e.*, TFs regulating target genes) were inferred using data from ORegAnno 3.0,⁶¹ GTRD and TRUSST. In cases where transcriptional regulatory interactions are derived from high-throughput datasets such as ChIP-seq, we attributed the regulatory interaction elicited by the bound transcription factor to genes which lie within a 10 kb window on either side of the ChIP-seq peak (ORegAnno) or meta-cluster (in the case of GTRD). TF–lncRNA interactions (*i.e.*, TFs regulating lncRNAs) were also inferred based on the ChIP-seq binding profiles represented by meta-clusters in GTRD. We used only TF–lncRNA interactions within intergenic lncRNAs to avoid assigning false regulatory interactions due to the high number of instances where the lncRNAs overlap with protein-coding genes. In addition, no overlaps were allowed between the coordinates of the ChIP-seq peaks/meta-clusters and any gene annotation. Only if the first annotation feature within a 10 kb genomic window downstream to the ChIP-seq peak/meta-cluster was designated as an intergenic lncRNA, a regulatory interaction between the TF and the lncRNA was assigned. Bedtools¹²⁰ was used for the custom analyses to look for overlaps between coordinates. All the nodes in the collected interactions were represented by their Ensembl gene IDs for standardization. A summary of the interactions collected from each resource and the quality control criteria applied is given in Table S6 (ESI†).

To generate PCeE and GCeE regulatory networks, interactions in this collated universal network were filtered using the differential expression data (Fig. 1). The assumption was made that if both nodes of a particular interaction were



expressed in the RNAseq data, the interaction is possible. Furthermore, to filter for the interactions of prime interest, only nodes which were differentially expressed and their associated interactors were included in the regulatory networks.

Cluster analysis

Clusters of highly interconnected regions within the PCeE and GCeE regulatory networks were identified using the MCODE plugin within Cytoscape.^{37,121} Default settings were applied: degree cutoff = 2, haircut = true, node score cutoff = 0.2, k-core = 2 and max depth = 100. Clusters were visualised in Cytoscape.

The nodes of each cluster were tested for functional enrichment (hypergeometric model, q value ≤ 0.05) based on Reactome annotations using the ReactomePA R package^{38,107–109} following conversion from mouse to human identifiers using Inparanoid (v8).^{110,111} Cases where the number of nodes associated with a pathway < 5 were considered not significant regardless of the q value. The top 5 significant Reactome pathways associated with each cluster were visualised using a heatmap generated in R (Fig. 3B). More than 5 pathways were visualised, where multiple Reactome pathways had equal q values.

Master regulator analysis

To identify potential master regulators of the Paneth cell and the goblet cell types, the upstream regulators of cell type specific markers (from ref. 12) were investigated. To do this, all markers were mapped to the relevant networks then sub-networks were extracted consisting of markers and their regulators (Table S7, ESI[†]).

Regulatory rewiring analysis

To calculate rewiring scores for regulators, sub-networks were extracted (from the PCeE and GCeE regulatory networks) containing just the regulator of interest and its downstream targets. For each regulator of interest, the subnetworks from the PCeE and GCeE networks were compared using the Cytoscape app DyNet.^{121,122} The degree corrected D_n score was extracted for each regulator and used to quantify rewiring of the regulator's downstream targets between the PCeE and GCeE regulatory networks. Functional analysis was carried out on the targets of the top five most rewired regulators. For each regulator, the targets were classified based on whether they are present in only the PCeE network, only the GCeE network or in both networks. Each group of targets was tested for functional enrichment (hypergeometric model, q value ≤ 0.1) based on Reactome and KEGG annotations using the ReactomePA R package^{38,107–109} following conversion from mouse to human identifiers using Inparanoid (v8).^{110,111}

IBD and drug target associated genes

Genes associated with UC and CD based on single nucleotide polymorphisms were obtained from two studies.^{67,68} Additionally, the top 100 differentially expressed genes were obtained from goblet cell analysis of inflamed UC vs. healthy human colonic tissue from ref. 13. Genes were converted to Mouse Ensembl identifiers using Inparanoid (v8) and bioDBnet db2db.^{110–112}

Additionally, to enable hypergeometric significant testing with the universal network as the background, only UC and CD genes present in the universal network are included in the analyses. eQTL datasets for CD were retrieved from ref. 69 while the list of targets related to drug-interactions was downloaded from ref. 123.

Quantification and statistical analysis

Statistical parameters including the exact value of n and statistical significance are reported in the figures and figure legends. n represents the number of enteroid biological replicates generated. Where relevant, data is judged to be statistically significant when Bonferroni corrected p value ≤ 0.01 . Genes with absolute log₂ fold change of $\geq |1|$ and false discovery rate ≤ 0.05 were considered to be differentially expressed. Based on principal component analysis of transcript expression, one biological replicate from the Paneth cell enriched enteroids was identified as an outlier and removed (Fig. S3, ESI[†]). Where stated, the hypergeometric distribution model was used to calculate significance using R.

Data and software availability

Small and stranded RNA-seq data has been deposited in the European Nucleotide Archive (ENA) with accession numbers PRJEB32354 and PRJEB32366 respectively. Scripts to analyse the differentially expressed genes are available on GitHub: https://github.com/korcsmarosgroup/organoid_regulatory_networks.

Funding

This work was supported by a fellowship to TK in computational biology at the Earlham Institute (Norwich, UK) in partnership with the Quadram Institute (Norwich, UK), and strategically supported by the Biotechnological and Biosciences Research Council, UK grants (BB/J004529/1, BB/P016774/1 and BB/CSP17270/1). SRC, LH, TWi and TK were also funded by a BBSRC ISP grant for Gut Microbes and Health BB/R012490/1 and its constituent project(s), BBS/E/F/000PR10353 and BBS/E/F/000PR10355. ZM was supported by a PhD studentship from Norwich Medical School. PP was supported by the BBSRC grant BB/J01754X/1. AT and MO were supported by the BBSRC Norwich Research Park Biosciences Doctoral Training Partnership (grant BB/M011216/1). TWr and WH were supported by an MRC award (MR/P026028/1). Next-generation sequencing and library construction was delivered *via* the BBSRC National Capability in Genomics and Single Cell (BB/CCG1720/1) at Earlham Institute by members of the Genomics Pipelines Group.

Author contributions

EJJ, ZJM and IH worked on the transcriptomics pipeline development; AT, PS, TWr, JB and MO designed and carried out specific network analysis tasks; SRC, UM, PPP and TWi established and supervised the organoid enrichment work; LJH and



FDP supervised the experimental and transcriptomics analysis work; WH, TWi and TK designed the study, supervised the experiments and the analysis. AT, PS, ZM, EJ, IH and TK wrote the manuscript.

Conflicts of interest

None.

Acknowledgements

The authors are grateful for the helpful discussions to the past and present members and visitors of the Wileman, Haerty and Korcsmaros groups.

References

- 1 K. Zhang, M. W. Horne and A. Dupont, The intestinal epithelium as guardian of gut barrier integrity, *Cell. Microbiol.*, 2015, **17**(11), 1561–1569.
- 2 R. Okumura and K. Takeda, Roles of intestinal epithelial cells in the maintenance of gut homeostasis, *Exp. Mol. Med.*, 2017, **49**(5), e338.
- 3 F. Gerbe and P. Jay, Intestinal tuft cells: epithelial sentinels linking luminal cues to the immune system, *Mucosal Immunol.*, 2016, **9**(6), 1353–1359.
- 4 B. A. Duerkop, S. Vaishnava and L. V. Hooper, Immune responses to the microbiota at the intestinal mucosal surface, *Immunity*, 2009, **31**(3), 368–376.
- 5 M. Mokry, S. Middendorp, C. L. Wiegnerinck, M. Witte, H. Teunissen and C. A. Meddens, *et al.*, Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium, *Gastroenterology*, 2014, **146**(4), 1040–1047.
- 6 C. L. Bevins and N. H. Salzman, Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis, *Nat. Rev. Microbiol.*, 2011, **9**(5), 356–368.
- 7 Y. S. Kim and S. B. Ho, Intestinal goblet cells and mucins in health and disease: recent insights and progress, *Curr. Gastroenterol. Rep.*, 2010, **12**(5), 319–330.
- 8 M. Gersemann, E. F. Stange and J. Wehkamp, From intestinal stem cells to inflammatory bowel diseases, *World J. Gastroenterol.*, 2011, **17**(27), 3198–3203.
- 9 R. Okamoto and M. Watanabe, Role of epithelial cells in the pathogenesis and treatment of inflammatory bowel disease, *J. Gastroenterol.*, 2016, **51**(1), 11–21.
- 10 T.-C. Liu, B. Gurram, M. T. Baldrige, R. Head, V. Lam and C. Luo, *et al.*, Paneth cell defects in Crohn's disease patients promote dysbiosis, *JCI Insight.*, 2016, **1**(8), e86907.
- 11 M. Gersemann, S. Becker, I. Kübler, M. Koslowski, G. Wang and K. R. Herrlinger, *et al.*, Differences in goblet cell differentiation between Crohn's disease and ulcerative colitis, *Differentiation*, 2009, **77**(1), 84–94.
- 12 A. L. Haber, M. Biton, N. Rogel, R. H. Herbst, K. Shekhar and C. Smillie, *et al.*, A single-cell survey of the small intestinal epithelium, *Nature*, 2017, **551**(7680), 333–339.
- 13 C. S. Smillie, M. Biton, J. Ordovas-Montanes, K. M. Sullivan, G. Burgin and D. B. Graham, *et al.*, Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis, *Cell*, 2019, **178**(3), 714–730.
- 14 K. Parikh, A. Antanaviciute, D. Fawcner-Corbett, M. Jagielowicz, A. Aulicino and C. Lagerholm, *et al.*, Colonic epithelial cell diversity in health and inflammatory bowel disease, *Nature*, 2019, **567**(7746), 49–55.
- 15 R. G. Lindeboom, L. van Voorthuijsen, K. C. Oost, M. J. Rodríguez-Colman, M. V. Luna-Velez and C. Furlan, *et al.*, Integrative multi-omics analysis of intestinal organoid differentiation, *Mol. Syst. Biol.*, 2018, **14**(6), e8227.
- 16 M. Noben, W. Vanhove, K. Arnauts, A. Santo Ramalho, G. Van Assche and S. Vermeire, *et al.*, Human intestinal epithelium in a dish: Current models for research into gastrointestinal pathophysiology, *United Eur. Gastroenterol. J.*, 2017, **5**(8), 1073–1081.
- 17 M. R. Aberle, R. A. Burkhart, H. Tiriak, S. W. M. Olde Damink, C. H. C. Dejong and D. A. Tuveson, *et al.*, Patient-derived organoid models help define personalized management of gastrointestinal cancer, *Br. J. Surg.*, 2018, **105**(2), e48–e60.
- 18 T. Sato, R. G. Vries, H. J. Snippert, M. van de Wetering, N. Barker and D. E. Stange, *et al.*, Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche, *Nature*, 2009, **459**(7244), 262–265.
- 19 T. Sato, J. H. van Es, H. J. Snippert, D. E. Stange, R. G. Vries and M. van den Born, *et al.*, Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts, *Nature*, 2011, **469**(7330), 415–418.
- 20 T. Sato and H. Clevers, Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications, *Science*, 2013, **340**(6137), 1190–1194.
- 21 N. C. Zachos, O. Kovbasnjuk, J. Foulke-Abel, J. In, S. E. Blutt and H. R. de Jonge, *et al.*, Human enteroids/colonoids and intestinal organoids functionally recapitulate normal intestinal physiology and pathophysiology, *J. Biol. Chem.*, 2016, **291**(8), 3759–3766.
- 22 L. Schulte, M. Hohwieler, M. Müller and J. Klaus, Intestinal organoids as a novel complementary model to dissect inflammatory bowel disease, *Stem Cells Int.*, 2019, 8010645.
- 23 I. Dotti and A. Salas, Potential Use of Human Stem Cell-Derived Intestinal Organoids to Study Inflammatory Bowel Diseases, *Inflammatory Bowel Dis.*, 2018, **24**(12), 2501–2509.
- 24 H. F. Farin, J. H. Van Es and H. Clevers, Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells, *Gastroenterology*, 2012, **143**(6), 1518–1529.
- 25 X. Yin, H. F. Farin, J. H. van Es, H. Clevers, R. Langer and J. M. Karp, Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny, *Nat. Methods*, 2014, **11**(1), 106–112.
- 26 B. E. Mead and J. M. Karp, All models are wrong, but some organoids may be useful, *Genome Biol.*, 2019, **20**(1), 66.



- 27 B. E. Mead, J. Ordovas-Montanes, A. P. Braun, L. E. Levy, P. Bhargava and M. J. Szucs, *et al.*, Harnessing single-cell genomics to improve the physiological fidelity of organoid-derived cell types, *BMC Biol.*, 2018, **16**(1), 62.
- 28 L. Luu, Z. J. Matthews, S. D. Armstrong, P. P. Powell, T. Wileman and J. M. Wastling, *et al.*, Proteomic Profiling of Enteroid Cultures Skewed toward Development of Specific Epithelial Lineages, *Proteomics*, 2018, **18**(16), e1800132.
- 29 E. J. Jones, Z. J. Matthews, L. Gul, P. Sudhakar, A. Treveil and D. Divekar, *et al.*, Integrative analysis of Paneth cell proteomic and transcriptomic data from intestinal organoids reveals functional processes dependent on autophagy, *Dis. Models Mech.*, 2019, **12**, 3.
- 30 C. K. Kwok and P. Y. Ng, Network analysis approach for biology, *Cell. Mol. Life Sci.*, 2007, **64**(14), 1739–1751.
- 31 T. Ideker and R. Nussinov, Network approaches and applications in biology, *PLoS Comput. Biol.*, 2017, **13**(10), e1005771.
- 32 N. M. Luscombe, M. M. Babu, H. Yu, M. Snyder, S. A. Teichmann and M. Gerstein, Genomic analysis of regulatory network dynamics reveals large topological changes, *Nature*, 2004, **431**(7006), 308–312.
- 33 M. G. P. van der Wijst, D. H. de Vries, H. Brugge, H.-J. Westra and L. Franke, An integrative approach for building personalized gene regulatory networks for precision medicine, *Genome Med.*, 2018, **10**(1), 96.
- 34 D. Módos, K. C. Bulusu, D. Fazekas, J. Kubisch, J. Brooks and I. Marczell, *et al.*, Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anticancer therapies, *NPJ Syst. Biol. Appl.*, 2017, **3**, 2.
- 35 T. S. Stappenbeck and D. P. B. McGovern, Paneth Cell Alterations in the Development and Phenotype of Crohn's Disease, *Gastroenterology*, 2017, **152**(2), 322–326.
- 36 H. C. Clevers and C. L. Bevins, Paneth cells: maestros of the small intestinal crypts, *Annu. Rev. Physiol.*, 2013, **75**, 289–311.
- 37 G. D. Bader and C. W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, *BMC Bioinf.*, 2003, **4**, 2.
- 38 A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie and P. Garapati, *et al.*, The reactome pathway knowledgebase, *Nucleic Acids Res.*, 2018, **46**(D1), D649–D655.
- 39 P. Han, C. Gopalakrishnan, H. Yu and E. Wang, Gene Regulatory Network Rewiring in the Immune Cells Associated with Cancer, *Genes*, 2017, **8**, 11.
- 40 J. Park and H. H. Wang, Systematic and synthetic approaches to rewire regulatory networks, *Curr. Opin. Syst. Biol.*, 2018, **8**, 90–96.
- 41 M. R. Leach and D. B. Williams, *Calnexin and Calreticulin, Molecular Chaperones of the Endoplasmic Reticulum – Madame Curie Bioscience Database – NCBI Bookshelf*, 2013.
- 42 A. Kaser, A.-H. Lee, A. Franke, J. N. Glickman, S. Zeissig and H. Tilg, *et al.*, XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease, *Cell*, 2008, **134**(5), 743–756.
- 43 A. Kaser and R. S. Blumberg, Endoplasmic reticulum stress in the intestinal epithelium and inflammatory bowel disease, *Semin. Immunol.*, 2009, **21**(3), 156–163.
- 44 A. Kaser, M. B. Flak, M. F. Tomczak and R. S. Blumberg, The unfolded protein response and its role in intestinal homeostasis and inflammation, *Exp. Cell Res.*, 2011, **317**(19), 2772–2779.
- 45 V. Chandra, E. Bhagyaraj, R. Nanduri, N. Ahuja and P. Gupta, NR1D1 ameliorates Mycobacterium tuberculosis clearance through regulation of autophagy, *Autophagy*, 2015, **11**(11), 1987–1997.
- 46 A. H. Mirza, C. H. Berthelsen, S. E. Seemann, X. Pan, K. S. Frederiksen and M. Vilien, *et al.*, Transcriptomic landscape of lncRNAs in inflammatory bowel disease, *Genome Med.*, 2015, **7**(1), 39.
- 47 R. H. Oakley and J. A. Cidlowski, The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease, *J. Allergy Clin. Immunol.*, 2013, **132**(5), 1033–1044.
- 48 P. Rutgeerts, The use of oral topically acting glucocorticosteroids in the treatment of inflammatory bowel disease, *Mediators Inflammation*, 1998, **7**(3), 137–140.
- 49 C. Prantera and S. Marconi, Glucocorticosteroids in the treatment of inflammatory bowel disease and approaches to minimizing systemic activity, *Ther. Adv. Gastroenterol.*, 2013, **6**(2), 137–156.
- 50 S. De Iudicibus, R. Franca, S. Martelossi, A. Ventura and G. Decorti, Molecular mechanism of glucocorticoid resistance in inflammatory bowel disease, *World J. Gastroenterol.*, 2011, **17**(9), 1095–1108.
- 51 K. Dubois-Camacho, P. A. Ottum, D. Franco-Muñoz, M. De la Fuente, A. Torres-Riquelme and D. Díaz-Jiménez, *et al.*, Glucocorticosteroid therapy in inflammatory bowel diseases: From clinical practice to molecular biology, *World J. Gastroenterol.*, 2017, **23**(36), 6628–6638.
- 52 A. Yemelyanov, J. Czornog, D. Chebotov, A. Karseladze, E. Kulevitch and X. Yang, *et al.*, Tumor suppressor activity of glucocorticoid receptor in the prostate, *Oncogene*, 2007, **26**(13), 1885–1896.
- 53 H. Dinkel, K. Van Roey, S. Michael, M. Kumar, B. Uyar and B. Altenberg, *et al.*, ELM2016 – data update and new functionality of the eukaryotic linear motif resource, *Nucleic Acids Res.*, 2016, **44**(D1), D294–D300.
- 54 S. Wu, Y.-G. Zhang, R. Lu, Y. Xia, D. Zhou and E. O. Petrof, *et al.*, Intestinal epithelial vitamin D receptor deletion leads to defective autophagy in colitis, *Gut*, 2015, **64**(7), 1082–1094.
- 55 D. Bakke, R. Lu, A. Agrawal, Y. Zhang and J. Sun, 18 myeloid vitamin d receptor signaling regulates Paneth cell function and intestinal homeostasis, *Gastroenterology*, 2018, **154**(1), S41.
- 56 T.-T. Wang, F. P. Nestel, V. Bourdeau, Y. Nagai, Q. Wang and J. Liao, *et al.*, Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression, *J. Immunol.*, 2004, **173**(5), 2909–2912.
- 57 A. F. Gombart, N. Borregaard and H. P. Koeffler, Human cathelicidin antimicrobial peptide (CAMP) gene is a direct target of the vitamin D receptor and is strongly up-regulated in myeloid cells by 1,25-dihydroxyvitamin D3, *FASEB J.*, 2005, **19**(9), 1067–1077.



- 58 F. H. Pei, Y. J. Wang, S. L. Gao, B. R. Liu, Y. J. Du and W. Liu, *et al.*, Vitamin D receptor gene polymorphism and ulcerative colitis susceptibility in Han Chinese, *J. Dig. Dis.*, 2011, **12**(2), 90–98.
- 59 M. T. Abreu, V. Kantorovich, E. A. Vasiliasukas, U. Gruntmanis, R. Matuk and K. Daigle, *et al.*, Measurement of vitamin D levels in inflammatory bowel disease patients reveals a subset of Crohn's disease patients with elevated 1,25-dihydroxyvitamin D and low bone mineral density, *Gut*, 2004, **53**(8), 1129–1136.
- 60 L. A. Bovolenta, M. L. Acencio and N. Lemke, HTRIDb: an open-access database for experimentally verified human transcriptional regulation interactions, *BMC Genomics*, 2012, **13**, 405.
- 61 R. Lesurf, K. C. Cotto, G. Wang, M. Griffith, K. Kasaian and S. J. M. Jones, *et al.*, ORegAnno 3.0: a community-driven resource for curated regulatory annotation, *Nucleic Acids Res.*, 2016, **44**(D1), D126–D132.
- 62 D. J. Bettoun, T. P. Burris, K. A. Houck, D. W. Buck, K. R. Stayrook and B. Khalifa, *et al.*, Retinoid X receptor is a nonsilent major contributor to vitamin D receptor-mediated transcriptional activation, *Mol. Endocrinol.*, 2003, **17**(11), 2320–2328.
- 63 R. Grenningloh, B. Y. Kang and I.-C. Ho, Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses, *J. Exp. Med.*, 2005, **201**(4), 615–626.
- 64 E. Mouly, K. Chemin, H. V. Nguyen, M. Chopin, L. Mesnard and M. Leite-de-Moraes, *et al.*, The Ets-1 transcription factor controls the development and function of natural regulatory T cells, *J. Exp. Med.*, 2010, **207**(10), 2113–2125.
- 65 S. Konno, M. Iizuka, M. Yukawa, K. Sasaki, A. Sato and Y. Horie, *et al.*, Altered expression of angiogenic factors in the VEGF-Ets-1 cascades in inflammatory bowel disease, *J. Gastroenterol.*, 2004, **39**(10), 931–939.
- 66 D. Li, T. Haritunians, A. Potdar and D. P. B. McGovern, 16 Genetic analysis identified novel loci associated with IBD, *Inflammatory Bowel Dis.*, 2018, **24**(suppl_1), S14–S14.
- 67 L. Jostins, S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern and K. Y. Hui, *et al.*, Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease, *Nature*, 2012, **491**(7422), 119–124.
- 68 K. K.-H. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley and S. Beik, *et al.*, Genetic and epigenetic fine mapping of causal autoimmune disease variants, *Nature*, 2015, **518**(7539), 337–343.
- 69 A. F. Di Narzo, L. A. Peters, C. Arghmann, A. Stojmirovic, J. Perrigoue and K. Li, *et al.*, Blood and Intestine eQTLs from an Anti-TNF-Resistant Crohn's Disease Cohort Inform IBD Genetic Association Loci, *Clin. Transl. Gastroenterol.*, 2016, **7**(6), e177.
- 70 M. Z. Cader and A. Kaser, Recent advances in inflammatory bowel disease: mucosal immune cells in intestinal inflammation, *Gut*, 2013, **62**(11), 1653–1664.
- 71 C. A. Thorne, I. W. Chen, L. E. Sanman, M. H. Cobb, L. F. Wu and S. J. Altschuler, Enteroid monolayers reveal an autonomous WNT and BMP circuit controlling intestinal epithelial growth and organization, *Dev. Cell*, 2018, **44**(5), 624–633.
- 72 D. P. Chopra, A. A. Dombkowski, P. M. Stemmer and G. C. Parker, Intestinal epithelial cells *in vitro*, *Stem Cells Dev.*, 2010, **19**(1), 131–142.
- 73 S. Lukovac and G. Roeselers, Intestinal crypt organoids as experimental models, in *The Impact of Food Bioactives on Health: in vitro and ex vivo models*. Cham (CH), ed. K. Verhoeckx, P. Cotter, I. López-Expósito, C. Kleiveland, T. Lea and A. Mackie *et al.*, Springer, 2015.
- 74 B. W. Doble and J. R. Woodgett, GSK-3: tricks of the trade for a multi-tasking kinase, *J. Cell Sci.*, 2003, **116**(Pt 7), 1175–1186.
- 75 G. Novarino, A. G. Fenstermaker, M. S. Zaki, M. Hofree, J. L. Silhavy and A. D. Heiberger, *et al.*, Exome sequencing links corticospinal motor neuron disease to common neurodegenerative disorders, *Science*, 2014, **343**(6170), 506–511.
- 76 J. K. Huang, D. E. Carlin, M. K. Yu, W. Zhang, J. F. Kreisberg and P. Tamayo, *et al.*, Systematic evaluation of molecular networks for discovery of disease genes, *Cell Syst.*, 2018, **6**(4), 484–495.
- 77 S. Wachi, K. Yoneda and R. Wu, Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues, *Bioinformatics*, 2005, **21**(23), 4205–4208.
- 78 H. Yu, N. M. Luscombe, J. Qian and M. Gerstein, Genomic analysis of gene expression relationships in transcriptional regulatory networks, *Trends Genet.*, 2003, **19**(8), 422–427.
- 79 J. Kubisch, D. Türei, L. Földvári-Nagy, Z. A. Dunai, L. Zsakai and M. Varga, *et al.*, Complex regulation of autophagy in cancer - integrated approaches to discover the networks that hold a double-edged sword, *Semin. Cancer Biol.*, 2013, **23**(4), 252–261.
- 80 N. Vijesh, S. K. Chakrabarti and J. Sreekumar, Modeling of gene regulatory networks: a review, *JBISE*, 2013, **06**(02), 223–231.
- 81 A.-L. Barabási and Z. N. Oltvai, Network biology: understanding the cell's functional organization, *Nat. Rev. Genet.*, 2004, **5**(2), 101–113.
- 82 G. P. Wagner and J. Zhang, The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms, *Nat. Rev. Genet.*, 2011, **12**(3), 204–213.
- 83 T. L. Davis and I. Rebay, Master regulators in development: Views from the *Drosophila* retinal determination and mammalian pluripotency gene networks, *Dev. Biol.*, 2017, **421**(2), 93–107.
- 84 M.-A. Mendoza-Parra, V. Malysheva, M. A. Mohamed Saleem, M. Lieb, A. Godel and H. Gronemeyer, Reconstructed cell fate-regulatory programs in stem cells reveal hierarchies and key factors of neurogenesis, *Genome Res.*, 2016, **26**(11), 1505–1519.
- 85 K. Cadwell, J. Y. Liu, S. L. Brown, H. Miyoshi, J. Loh and J. K. Lennerz, *et al.*, A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells, *Nature*, 2008, **456**(7219), 259–263.

- 86 M. J. Rodríguez-Colman, M. Schewe, M. Meerlo, E. Stigter, J. Gerrits and M. Pras-Raves, *et al.*, Interplay between metabolic identities in the intestinal crypt supports stem cell function, *Nature*, 2017, **543**(7645), 424–427.
- 87 U. Paulus, M. Loeffler, J. Zeidler, G. Owen and C. S. Potten, The differentiation and lineage development of goblet cells in the murine small intestinal crypt: experimental and modelling studies, *J. Cell Sci.*, 1993, **106**(Pt 2), 473–483.
- 88 J. Jung and H. Jung, Methods to analyze cell type-specific gene expression profiles from heterogeneous cell populations, *Anim. Cells Syst.*, 2016, **20**(3), 113–117.
- 89 A. Brazovskaja, B. Treutlein and J. G. Camp, High-throughput single-cell transcriptomics on organoids, *Curr. Opin. Biotechnol.*, 2019, **55**, 167–171.
- 90 G. Chen, B. Ning and T. Shi, Single-Cell RNA-Seq Technologies and Related Computational Data Analysis, *Front. Genet.*, 2019, **10**, 317.
- 91 T. Nunes, C. Bernardazzi and H. S. de Souza, Cell death and inflammatory bowel diseases: apoptosis, necrosis, and autophagy in the intestinal epithelium, *BioMed Res. Int.*, 2014, 218493.
- 92 H. A. Aghdaei, A. A. Kadijani, D. Sorrentino, A. Mirzaei, S. Shahrokh and H. Balaii, *et al.*, An increased Bax/Bcl-2 ratio in circulating inflammatory cells predicts primary response to infliximab in inflammatory bowel disease patients, *United Eur. Gastroenterol. J.*, 2018, **6**(7), 1074–1081.
- 93 P. J. Real, V. Tosello, T. Palomero, M. Castillo, E. Hernando and E. de Stanchina, *et al.*, Gamma-secretase inhibitors reverse glucocorticoid resistance in T cell acute lymphoblastic leukemia, *Nat. Med.*, 2009, **15**(1), 50–58.
- 94 X. Zheng, K. Tsuchiya, R. Okamoto, M. Iwasaki, Y. Kano and N. Sakamoto, *et al.*, Suppression of *hath1* gene expression directly regulated by *hes1* via notch signaling is associated with goblet cell depletion in ulcerative colitis, *Inflammatory Bowel Dis.*, 2011, **17**(11), 2251–2260.
- 95 C. G. Chapman and J. Pekow, The emerging role of miRNAs in inflammatory bowel disease: a review, *Ther. Adv. Gastroenterol.*, 2015, **8**(1), 4–22.
- 96 S. Andrews, Babraham Bioinformatics – FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. 2010 [cited 2018 Sep 28], available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- 97 D. Kim, B. Langmead and S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements, *Nat. Methods*, 2015, **12**(4), 357–360.
- 98 M. Pertea, G. M. Pertea, C. M. Antonescu, T.-C. Chang, J. T. Mendell and S. L. Salzberg, StringTie enables improved reconstruction of a transcriptome from RNA-seq reads, *Nat. Biotechnol.*, 2015, **33**(3), 290–295.
- 99 M. Pertea, D. Kim, G. M. Pertea, J. T. Leek and S. L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown, *Nat. Protoc.*, 2016, **11**(9), 1650–1667.
- 100 L. Kong, Y. Zhang, Z.-Q. Ye, X.-Q. Liu, S.-Q. Zhao and L. Wei, *et al.*, CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine, *Nucleic Acids Res.*, 2007, **35**(Web Server issue), W345–W349.
- 101 L. Wang, H. J. Park, S. Dasari, S. Wang, J.-P. Kocher and W. Li, CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model, *Nucleic Acids Res.*, 2013, **41**(6), e74.
- 102 N. L. Bray, H. Pimentel, P. Melsted and L. Pachter, Near-optimal probabilistic RNA-seq quantification, *Nat. Biotechnol.*, 2016, **34**(5), 525–527.
- 103 H. Pimentel, N. L. Bray, S. Puente, P. Melsted and L. Pachter, Differential analysis of RNA-seq incorporating quantification uncertainty, *Nat. Methods*, 2017, **14**(7), 687–690.
- 104 A. Rueda, G. Barturen, R. Lebrón, C. Gómez-Martín, Á. Alganza and J. L. Oliver, *et al.*, sRNAtoolbox: an integrated collection of small RNA research tools, *Nucleic Acids Res.*, 2015, **43**(W1), W467–W473.
- 105 A. Kozomara and S. Griffiths-Jones, miRBase: annotating high confidence microRNAs using deep sequencing data, *Nucleic Acids Res.*, 2014, **42**(Database issue), D68–D73.
- 106 M. D. Robinson, D. J. McCarthy and G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, 2010, **26**(1), 139–140.
- 107 H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono and M. Kanehisa, KEGG: kyoto encyclopedia of genes and genomes, *Nucleic Acids Res.*, 1999, **27**(1), 29–34.
- 108 G. Yu and Q.-Y. He, ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization, *Mol. BioSyst.*, 2016, **12**(2), 477–479.
- 109 M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato and K. Morishima, KEGG: new perspectives on genomes, pathways, diseases and drugs, *Nucleic Acids Res.*, 2017, **45**(D1), D353–D361.
- 110 K. P. O'Brien, M. Remm and E. L. L. Sonnhammer, InParanoid: a comprehensive database of eukaryotic orthologs, *Nucleic Acids Res.*, 2005, **33**(Database issue), D476–D480.
- 111 E. L. L. Sonnhammer and G. Östlund, InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic, *Nucleic Acids Res.*, 2015, **43**(Database issue), D234–D239.
- 112 U. Mudunuri, A. Che, M. Yi and R. M. Stephens, bioDBnet: the biological database network, *Bioinformatics*, 2009, **25**(4), 555–556.
- 113 I. S. Vlachos, M. D. Paraskevopoulou, D. Karagkouni, G. Georgakilas, T. Vergoulis and I. Kanellos, *et al.*, DIANA-TarBasev7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions, *Nucleic Acids Res.*, 2015, **43**(Database issue), D153–D159.
- 114 M. D. Paraskevopoulou, I. S. Vlachos, D. Karagkouni, G. Georgakilas, I. Kanellos and T. Vergoulis, *et al.*, DIANA-LncBasev2: indexing microRNA targets on non-coding transcripts, *Nucleic Acids Res.*, 2016, **44**(D1), D231–D238.
- 115 S. W. Chi, J. B. Zang, A. Mele and R. B. Darnell, Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps, *Nature*, 2009, **460**(7254), 479–486.



[View Article Online](#)

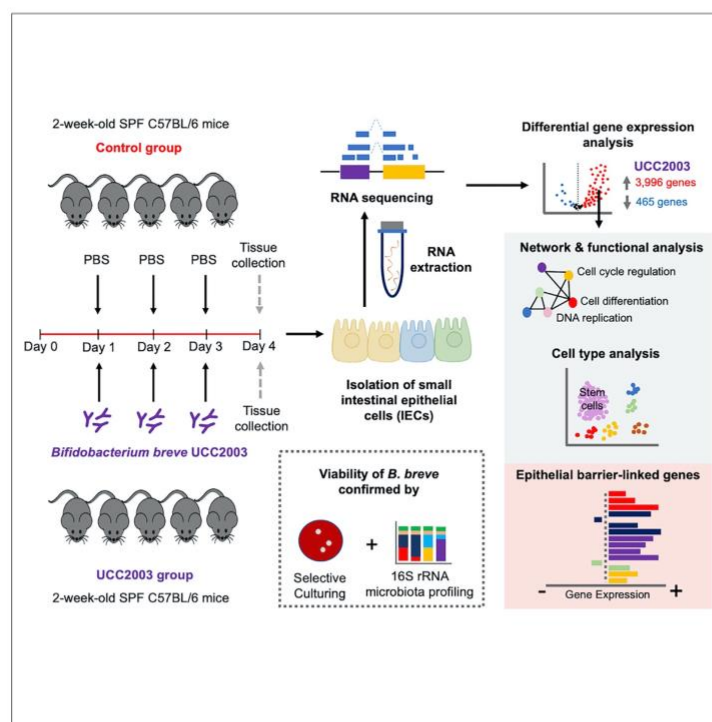
Research Article

Molecular Omics

- 116 J. Wang, M. Lu, C. Qiu and Q. Cui, TransmiR: a transcription factor-microRNA regulation database, *Nucleic Acids Res.*, 2010, **38**(Database issue), D119–D122.
- 117 I. Yevshin, R. Sharipov, T. Valeev, A. Kel and F. Kolpakov, GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments, *Nucleic Acids Res.*, 2017, **45**(D1), D61–D67.
- 118 H. Han, H. Shim, D. Shin, J. E. Shim, Y. Ko and J. Shin, *et al.*, TRRUST: a reference database of human transcriptional regulatory interactions, *Sci. Rep.*, 2015, **5**, 11432.
- 119 H. Han, J.-W. Cho, S. Lee, A. Yun, H. Kim and D. Bae, *et al.*, TRRUSTv2: an expanded reference database of human and mouse transcriptional regulatory interactions, *Nucleic Acids Res.*, 2018, **46**(D1), D380–D386.
- 120 A. R. Quinlan and I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics*, 2010, **26**(6), 841–842.
- 121 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang and D. Ramage, *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.*, 2003, **13**(11), 2498–2504.
- 122 I. H. Goenawan, K. Bryan and D. J. Lynn, DyNet: visualization and analysis of dynamic molecular interaction networks, *Bioinformatics*, 2016, **32**(17), 2713–2715.
- 123 K. C. Cotto, A. H. Wagner, Y.-Y. Feng, S. Kiwala, A. C. Coffman and G. Spies, *et al.*, DGIdb 3.0: a redesign and expansion of the drug-gene interaction database, *Nucleic Acids Res.*, 2018, **46**(D1), D1068–D1073.



Article

Bifidobacterium breve UCC2003 Induces a Distinct Global Transcriptomic Program in Neonatal Murine Intestinal Epithelial Cells

Raymond Kiu,
Agatha Treveil,
Lukas C.
Harnisch, ...,
Douwe van
Sinderen, Tamas
Korcsmaros,
Lindsay J. Hall

lindsay.hall@quadram.ac.uk

HIGHLIGHTS

B. breve administration significantly alters the murine neonatal IEC transcriptome

Genes/pathways involved in epithelial barrier function are particularly impacted

Bifidobacterium may target the IEC stem cell compartment to induce regeneration

Kiu et al., iScience 23, 101336
July 24, 2020 © 2020 The
Author(s).
[https://doi.org/10.1016/
j.isci.2020.101336](https://doi.org/10.1016/j.isci.2020.101336)





Article

Bifidobacterium breve UCC2003 Induces a Distinct Global Transcriptomic Program in Neonatal Murine Intestinal Epithelial Cells

Raymond Kiu,¹ Agatha Treveil,^{1,2} Lukas C. Harnisch,¹ Shabhonam Caim,¹ Charlotte Leclaire,¹ Douwe van Sinderen,³ Tamas Korcsmaros,^{1,2} and Lindsay J. Hall^{1,4,5,6,7,*}

SUMMARY

The underlying health-driving mechanisms of *Bifidobacterium* during early life are not well understood, particularly how this microbiota member may modulate the intestinal barrier via programming of intestinal epithelial cells (IECs). We investigated the impact of *Bifidobacterium breve* UCC2003 on the transcriptome of neonatal murine IECs. Small IECs from two-week-old neonatal mice administered *B. breve* UCC2003 or PBS (control) were subjected to global RNA sequencing, and differentially expressed genes, pathways, and affected cell types were determined. We observed extensive regulation of the IEC transcriptome with ~4,000 genes significantly up-regulated, including key genes linked with epithelial barrier function. Enrichment of cell differentiation pathways was observed, along with an overrepresentation of stem cell marker genes, indicating an increase in the regenerative potential of the epithelial layer. In conclusion, *B. breve* UCC2003 plays a central role in driving intestinal epithelium homeostatic development during early life and suggests future avenues for next-stage clinical studies.

INTRODUCTION

Bifidobacterium represents a keystone member of the early life gut microbiota (Arrieta et al., 2014; O'Neill et al., 2017; Derrien et al., 2019). Certain species and strains are found at high levels in vaginally delivered breast-fed infants including *Bifidobacterium longum* subsp. *infantis*, *B. longum* subsp. *longum*, *Bifidobacterium bifidum*, *Bifidobacterium pseudocatenulatum*, and *Bifidobacterium breve* (Dominguez-Bello et al., 2010; Mikami et al., 2012; Nagpal et al., 2017; Stewart et al., 2018). As a dominant member of the neonatal gut microbiota, *Bifidobacterium* is associated with metabolism of breast milk, modulation of host immune responses, and protection against infectious diseases (Fukuda et al., 2012; Ling et al., 2016; Robertson et al., 2020; Lawson et al., 2020; Patole et al., 2016; Baucells et al., 2016; Jacobs et al., 2013; Plummer et al., 2018). However, the mechanisms driving improved health outcomes during early life are largely underexplored and are likely strain dependent.

A key interface between *Bifidobacterium* and the host is the intestinal epithelial cell (IEC) barrier (Thoo et al., 2019; Groschwitz and Hogan, 2009). Previous studies have indicated that certain strains of *Bifidobacterium* specifically modulate IEC responses during inflammatory insults, which may help protect from certain gut disorders (Hsieh et al., 2015; Srutkova et al., 2015; Grimm et al., 2015). In murine experimental models, previous work by our group has shown that infant-associated *B. breve* UCC2003 modulates cell death-related signaling molecules, which in turn reduces the number of apoptotic IECs (Hughes et al., 2017). This protection from pathological IEC shedding appeared to be via the *B. breve* exopolysaccharide (EPS) capsule and the host-immune adaptor protein MyD88. Another strain of *B. breve*, NumRes 204 (commercial strain) has also been shown to up-regulate the tight junction (TJ) proteins Claudin 4 and Occludin in a mouse colitis model (Zheng et al., 2014; Plantinga et al., 2011).

Many of the studies to date have focused on the role of *Bifidobacterium* and modulation of IECs in the context of acute or chronic gut inflammation, with expression profiling limited to specific immune or apoptosis signaling targets (Plaza-Diaz et al., 2014; Riedel et al., 2006; Liu et al., 2010; Hsieh et al., 2015). As many of these studies have involved pre-colonization of the gut with *Bifidobacterium* strains, followed

¹Gut Microbes & Health, Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, UK

²Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK

³APC Microbiome Ireland & School of Microbiology, University College Cork, Cork T12YT20, Ireland

⁴Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

⁵Chair of Intestinal Microbiome, School of Life Sciences, Technical University of Munich, 85354 Freising, Germany

⁶ZIEL – Institute for Food & Health, Technical University of Munich, 85354 Freising, Germany

⁷Lead Contact

*Correspondence: lindsay.hall@quadram.ac.uk
<https://doi.org/10.1016/j.isci.2020.101336>



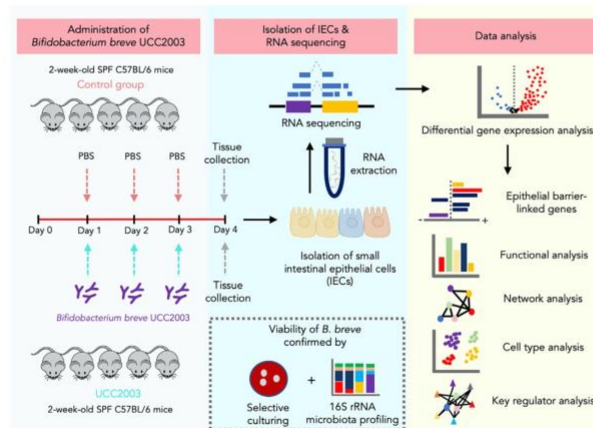


Figure 1. Schematic Representation of the Study Design and *In Silico* Analysis Workflow

by inflammatory insult, this suggests that initial priming during normal “healthy” conditions may modulate subsequent protective responses. Furthermore, these studies have often been performed in adult mice rather than exploring effects during the early life developmental window, where *Bifidobacterium* effects are expected to be most pronounced. Previous work has indicated that there is significant modulation of the neonatal IEC transcriptome in response to gut microbiota colonization, but to date no studies have probed how particular early life-associated microbiota members, like *Bifidobacterium*, may modulate neonatal IEC responses (Pan et al., 2018). Thus, to understand if and how *Bifidobacterium* may modulate IEC homeostasis during the early life developmental window, we administered *B. breve* UCC2003 to neonatal mice and profiled transcriptional responses in isolated small intestine IECs using global RNA sequencing (RNA-seq). Our analysis indicated whole-scale changes in the transcriptional program of IECs (~4,000 significantly up-regulated genes) that appear to be linked to cell differentiation/proliferation and immune development. Notably the stem cell compartment of IECs seemed to elicit the strongest gene signature. These data highlight the role of *B. breve* UCC2003 in driving early life epithelial cell differentiation and maturation, impacting intestinal integrity and immune functions, which provides a mechanistic basis for understanding associated health-promoting effects.

RESULTS

To examine the effects of *B. breve* UCC2003 on the transcriptional profiles of host IECs under homeostatic conditions, we extracted RNA from isolated IECs of healthy 2-week-old neonatal mice (control group) and mice gavaged with *B. breve* UCC2003 for three consecutive days ($n = 5$ per group). Isolated RNAs from IECs were subjected to RNA-seq to determine global mRNA expression (Figure 1). Subsequently, Differential Gene Expression analysis was performed to understand *B. breve*-associated gene regulation.

Minimal Impact of *B. breve* UCC2003 on the Wider Neonatal Gut Microbiota

Initially, we examined for the presence of *B. breve* UCC2003 in the gut microbiome and impact on the wider microbiota using culture and 16S rRNA microbiota profiling approaches (Figures 2A and 2B). We observed high levels of *B. breve* UCC2003 across the 4 days in fecal samples, with higher levels of viable *B. breve* UCC2003 within the colon ($\sim 10^8$ CFU/g [colony-forming unit]), when compared with the small intestine ($\sim 10^5$ CFU/g; Figure 2B). Based on 16S rRNA analysis, relative abundance of *Bifidobacterium* increased significantly in the UCC2003 group ($p = 0.012$) following bacterial administration, whereas the control group displayed very low relative *Bifidobacterium* abundance ($\sim 0.01\%$; Figure 2C). Principal-component analysis (PCA) on gut microbiota profiles (control versus UCC2003) showed a distinct change in microbial

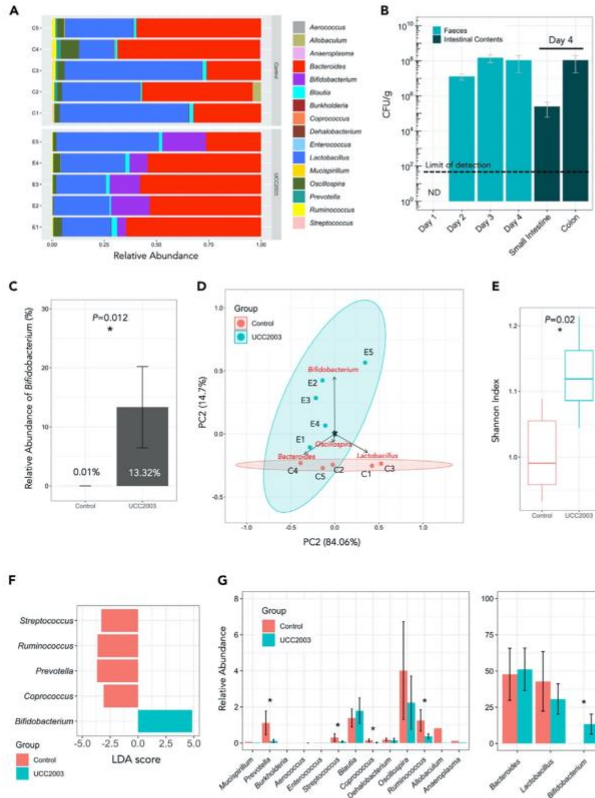


Figure 2. 16S rRNA Amplicon Sequencing Analysis of Murine Intestinal Microbiota
 (A) Genus-level 16S rRNA gene profiling of mice gut microbiota on day 4 (control versus UCC2003).
 (B) Dynamics of *B. breve* UCC2003 load (CFU/g) from day 1 (before *B. breve* administration) through day 4. *B. breve* was present in intestines throughout (small intestines and colon; on day 4). ND, non-detectable. Data are represented as mean \pm SD.
 (C) Relative abundance of genus *Bifidobacterium* in UCC2003 group is significantly increased.
 (D) Principal-component analysis on mice gut microbiota (control versus UCC2003 based on genus-level metatransomics).
 (E) Shannon diversity index on mice gut microbiota (control versus UCC2003). Data are represented as mean \pm SD. Significance test: t test (* $p < 0.05$; two-sided).
 (F) Linear Discriminant Analysis (LDA) showing enriched taxa in each group (control versus UCC2003).
 (G) Relative abundance comparison of all genera. * $p < 0.05$ (LDA).

community composition in the UCC2003 group primarily driven by increased relative abundance of *Bifidobacterium*, which may also correlate with increased overall microbial diversity in the UCC2003 group (Figures 2D and 2E). Linear Discriminant Analysis also indicated that *Bifidobacterium* was uniquely enriched in UCC2003 group, and that microbiota members with low relative abundance (<2%) such as *Streptococcus*, *Ruminococcus*, *Prevotella*, and *Coprococcus* were significantly lower (Figures 2F and 2G). Overall, administration of *B. breve* UCC2003 appeared to minimally impact the wider gut microbiota, without significantly

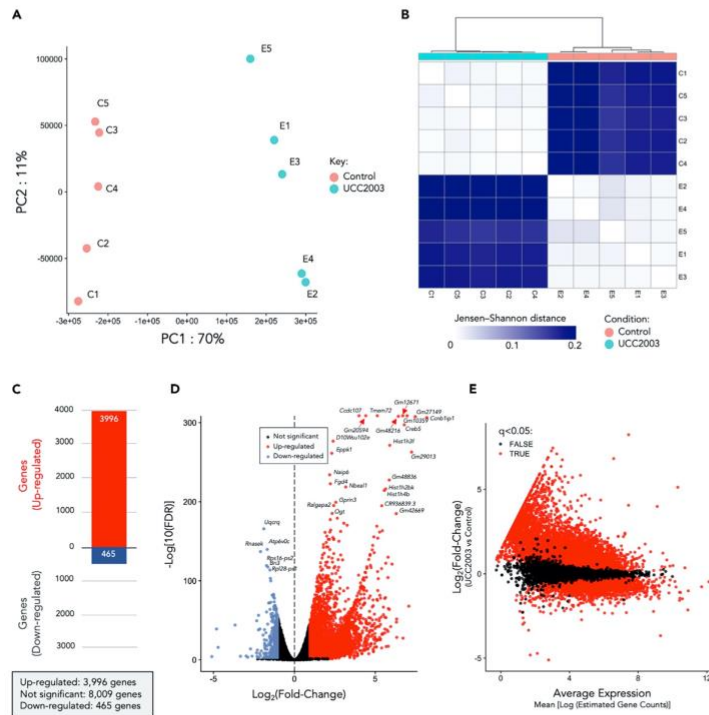


Figure 3. RNA-Seq Analysis and Statistics

(A) Principal-component analysis showing distinct overall gene expression profiles across all individual samples based on 12,965 highly expressed genes. See also [Table S1](#).

(B) Clustering of individual RNA-seq samples based on Jensen-Shannon distance. Distinct gene expression profiles were demonstrated between these two groups of samples (control versus UCC2003).

(C) Total number of differentially expressed genes (DEGs) in UCC2003 group.

(D) Volcano plot of global gene expression. Up-regulated DEGs are labeled as red dots, whereas down-regulated DEGs are labeled in blue.

(E) MA plot of global gene expression (plot of log-intensity ratios [M-values] versus log-intensity averages [A-values]).

altering relative abundance of other major resident taxa including *Lactobacillus*, *Bacteroides*, and *Blautia* compared with the control group.

Impact of *B. breve* UCC2003 on the Neonatal Intestinal Epithelial Transcriptome

To understand the distribution of samples based on IEC gene expression profiles we performed PCA analysis ([Figure 3A](#); [Table S1](#)). All samples clustered according to group (control versus UCC2003), suggesting a significant impact of *B. breve* UCC2003 on gene expression profiles, with distance-wise clustering (Jensen-Shannon) also supporting separation of experimental groups ([Figure 3B](#)). To define differentially expressed genes (DEGs), we employed a filter of absolute \log_2 fold change (LFC) > 1.0 (with adjusted $p < 0.05$), which equates to a minimum 2-fold change in gene expression ([Figures 3C–3E](#); [Table S2](#)). After analysis, a total of 3,996 DEGs were significantly up-regulated, whereas 465 genes were significantly down-regulated in *B. breve* UCC2003-supplemented animals when compared with controls ([Figures 3C and 4A](#)). Notably,



To determine the functional role of DEGs, we examined the most significantly regulated genes ranked by false discovery rate-adjusted p values (or, q values). We first looked at the top 20 up-regulated DEGs in the *B. breve* UCC2003 experimental group (Figure 4B). Most genes annotated with known biological processes had cell differentiation and cell component organization functions including *Ccnb1ip1*, *Hist1h4b*, *Vps13b*, and *Fgd4* (annotated in the PANTHER Gene Ontology [GO] Slim resource). Two genes were involved in cell death and immune system processes, namely, *Naip6* and *Gm20594* (Table S3). When we ranked the top-regulated genes using LFC, we observed increased expression of *Creb5*, which is involved in the regulation of neuropeptide transcription (cAMP response element-binding protein; CREB) (Figure 4C). CREB is also known to regulate circadian rhythm, and we also identified additional circadian-clock-related genes that were significantly up-regulated including *Per2* and *Per3*. We noted that several top down-regulated DEGs were annotated as genes involved in metal binding, or metal-related genes including *Mt1*, *Mt2*, *Hba-a1*, *Hbb-bt*, and *Ft1-ps1* (Figure 4D; Table S4).

Regulation of Intestinal Epithelial Barrier-Associated Genes

As *B. breve* strains have been previously shown to modulate certain TJ/barrier-related proteins, we next investigated DEGs associated with intestinal epithelial barrier development/intestinal structural organization (Figure 4E). Several TJ structural-associated DEGs were observed, including Claudin-encoding gene *Cldn34c1* (LFC 3.14), Junction Adhesion Molecules-encoding genes *Jam2* (LFC 2.9), and TJ protein (also called Zonula Occludens protein; ZO)-encoding gene *Tjp1* (LFC 1.49). Genes that encode integrins (involved in regulation of intracellular cytoskeleton) also exhibited a trend of increased expression (13/14; 92.8%). Both Piezo genes, which assist in TJ organization, *Piezo1* (LFC 1.25) and *Piezo2* (LFC 1.9), were significantly up-regulated in the *B. breve* UCC2003-treated group.

Over 90% cadherins, proteins associated with the assembly of adherens junctions (Figure 4E), were up-regulated, including *Pcdhb14* (LFC 2.8), *Pcdhgb4* (LFC 2.7), *Pcdh8* (LFC 1.3), *Fat1* (LFC 1.5), and *Dsg2* (LFC 1.1). Interestingly, several genes (4/7; 57.1%) involved in mucous layer generation were significantly up-regulated in the UCC2003 experimental group, including *Muc2* (LFC 2.2), *Muc6* (LFC 3.7), *Muc5b* (LFC 2.9), and *Muc4* (LFC 1.24). Genes *Gja1* (LFC 3.59) and *Gjb8* (LFC 2.63) that encode gap junction proteins were also up-regulated. In addition, we also investigated the differential expression of genes associated with integrin assembly and downstream integrin signaling pathways (Figure 4F). Over 70% (16/21) of these genes were up-regulated, with 52.3% (11/21) significantly increased in gene expression in the UCC2003 group (LFC >1.0).

We observed increased expression of genes associated with IEC barrier development including cadherins, gap junctions, integrins, mucous layer-associated genes, and several key TJ proteins. These strongly induced gene expression profiles suggest that *B. breve* UCC2003 is involved in enhancing epithelial barrier development in neonates.

Modulation of Cell Maturation Processes

We next sought to understand the biological functions of up-regulated DEGs by employing PANTHER GO-Slim functional assignment and process/pathway enrichment analysis (see Figure S1; Tables S5 and S6). DEGs were predominantly involved in general biological processes including cellular process (901 genes) and metabolic process (597 genes; Table S7). At the molecular function level, DEGs were primarily assigned to binding (868 genes) and catalytic activity (671 genes; Table S8), with Olfactory Signaling Pathway and Cell Cycle (biological) pathways also found to be enriched (Table S9).

To delve further into the data, we constructed a signaling network based on up-regulated DEGs ($n = 3,996$) with the aim of identifying specific gene networks involved in important signaling pathways (Figure 5A). Overall, 1,491 DEGs were successfully mapped (37.3%) to a signaling network that comprised 8,180 genes. Four individual clusters of genes were detected, with functional assignment and pathway analysis implemented on these clusters (Figure 5A). All gene clusters were associated with cell differentiation and maturation, with cluster 1 (68 genes) linked specifically with DNA replication and transcription, cluster 2 (26 genes) with cell growth and immunity, cluster 3 (11 genes) with cell replication, and cluster 4 (72 genes) related to cell cycle and cell division (Table S10).

Intestinal Cell Type Analysis on DEGs Identifies Significant Enrichment of Epithelial Stem Cells

IECs include several absorptive and secretory cell types, namely, enterocytes, Paneth cells, goblet cells, enteroendocrine cells, tuft cells, and stem cells. As these cells perform different functions in the gut, it was

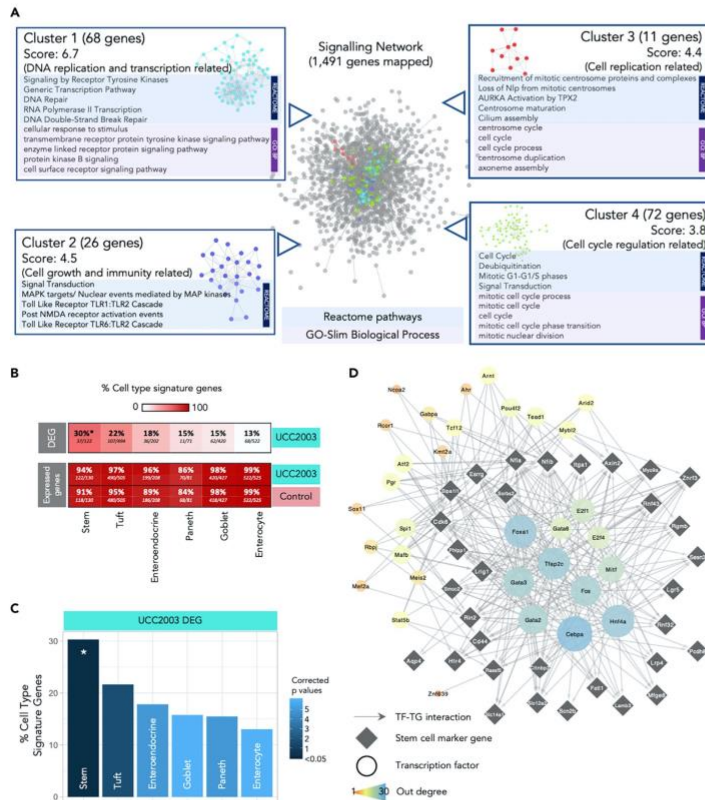


Figure 5. Signaling Network Analysis, IEC Subtyping, and Key Regulator Analysis
(A) Cluster analysis of signaling network for significantly up-regulated genes (n = 3,996). Representative enriched pathways (Reactome) and GO terms (Biological Process) identified in each individual cluster were listed alongside. See also Table S10.
(B) Heat plot showing percentage of cell type signature genes in DEG and expressed genes (both control and UCC2003 groups). All expressed genes are well represented in IEC cell type signature genes.
(C) Cell type analysis on IEC DEGs using known cell-specific signature genes. Stem cells were statistically over-represented in DEGs. *p < 0.05. See also Table S11.
(D) Key regulators of stem cell DEGs.

important to understand whether *B. breve* UCC2003 had a cell type-specific effect on the intestinal epithelium. Using known cell type-specific gene markers (Haber et al., 2017), we identified cell type gene signatures modulated within the UCC2003 group (Figures 5B and 5C). Importantly, all cell type markers were well represented in the expressed genes of the whole IEC transcriptomics data from both groups (control + UCC2003), thus validating the presence of all IEC types in our study data (Figure 5B). Cell type analysis of genes differentially expressed after *B. breve* UCC2003 supplementation revealed that stem cell marker genes were significantly enriched (30%; p < 0.05) among the six IEC types (Table S11). Signatures of other cell types were also present (linking to marker genes in the DEG list) but not significantly overrepresented:



tuft cells (22%), enteroendocrine cells (18%), goblet cells (15%), Paneth cells (15%), and enterocytes (13%; Figure 5C). These data indicated that intestinal epithelial stem cells, cells primarily involved in cell differentiation, were the primary cell type whose numbers and transcriptomic program were regulated by *B. breve* UCC2003.

Further investigation of this stem cell signature revealed that of the 37 differentially expressed marker genes, 35 are up-regulated in the presence of *B. breve* UCC2003. This indicates an increase in the quantity of stem cells or semi-differentiated cells in the epithelium, consistent with the overrepresentation of cell cycle- and DNA replication-associated genes observed in the whole differential expression dataset. Functional analysis of the 37 stem cell signature genes revealed only one overrepresented process—Regulation of Frizzled by ubiquitination ($p < 0.05$), which is a subprocess of WNT signaling. WNT signaling is important in maintaining the undifferentiated state of stem cells (Nusse, 2008).

Finally, we employed a network approach to predict key transcription factor (TF) regulators of the differentially expressed stem cell marker genes, through which *B. breve* UCC2003 may be acting (Figure 5D). Using the TF-target gene database, DoRothEA, we identified expressed TFs known to regulate these genes (Garcia-Alonso et al., 2019; Holland et al., 2019). Five genes had no known and expressed regulator, and thus were excluded. Hypergeometric significance testing was used to identify which of these TFs are the most influential (see Methods for details). This analysis identified 32 TF regulators (Figure 5D). Of these regulators, 12 were differentially expressed in the IEC dataset (all up-regulated): *Fos*, *Gabpa*, *Rcor1*, *Arid2*, *Tead1*, *Mybl2*, *Mef2a*, *Ahr*, *Pgr*, *Kmt2a*, *Ncoa2*, and *Tcf12*. Functional analysis of all the TF regulators and their targeted genes together, revealed overrepresented functions relating to WNT signaling, histone methylation for self-renewal and proliferation of hematopoietic stem cells, and nuclear receptor (incl. estrogen) signaling (Table S12). These data indicate that *B. breve* UCC2003 directly affects key transcriptomic programs that regulate specific signaling processes, particularly within stem cells.

DISCUSSION

The early life developmental window represents a crucial time for microbe-host interactions that impacts health both in the short and longer term. Understanding how specific microbiota members modulate host responses in pre-clinical models may help the design and development of next-stage targeted microbiota therapies in humans. Here we investigated how *B. breve* UCC2003 induces genome-wide transcriptomic changes in small intestine IECs of neonatal mice. We observed that *B. breve* had a global impact on the IEC transcriptome, evidenced by the large number of significantly up-regulated genes and pathways related to cell differentiation and cell proliferation, including genes associated with epithelial barrier function. We propose that *B. breve* may act as a key early life microbiota member driving fundamental cellular responses in murine IECs, particularly within the stem cell compartment, and thus drives epithelial barrier development and maintenance during neonatal life stages. However, further clinical studies would be required to determine if our findings extrapolate to the human setting.

B. breve is known to confer beneficial effect on gut health; however, our knowledge related to the mechanisms underlying these responses is limited. Most studies have focused on targeted immune cells or pathways (during disease and/or inflammation), and to our knowledge no studies have probed global transcriptomic changes within IECs, the frontline physical barrier between bacteria and host (Turroni et al., 2014; Gann, 2010). Our presented findings in a pre-clinical model, namely, ~4,000 up-regulated DEGs and ~450 down-regulated DEGs within the *B. breve* group, indicate that this *Bifidobacterium* strain modulates whole-scale changes within this critical single-cell layer. Notably, we also examined how *B. breve* modulates adult IEC responses; however, we did not observe any significantly differentially regulated genes when compared with control animals. The striking differences in DEGs between these two life points indicate that *B. breve* modulation of IECs is limited to the neonatal window. Dominance of *Bifidobacterium* in early life (including strains of *B. breve*) overlaps with the development and maturation of many host responses, including epithelial barrier integrity. Therefore, presence of these strains would be expected to play an over-sized role in this initial homeostatic priming, which may afford protection against inflammatory insults in later life, as has been shown previously in a mouse model of pathological epithelial cell shedding (Hughes et al., 2017). Further clinical studies would be required to probe these findings in detail to determine their importance during healthy infant development.

Exploring the murine transcriptional responses in more detail revealed that expressions of key genes associated with formation of epithelial barrier components were up-regulated, including major cell junction



protein-encoding genes (75%; 42/56 genes). More specifically, several integrin-associated genes were up-regulated in the presence of UCC2003. Integrins facilitate cell-cell and cell-extracellular matrix adhesion and binding and assembly of the fibronectin matrix that is pivotal for cell migration and cell differentiation (Harburger and Calderwood, 2009; Qin et al., 2004; Mosher et al., 1991). Integrins also play an important role in downstream intracellular signaling that controls cell differentiation, proliferation, and cell survival, including the Raf-MEK-ERK signaling pathway (we also observed enrichment of genes involved in this pathway) (Chernyavsky et al., 2005; Li et al., 2016). Another key intestinal barrier component is represented by TJs, linking complexes between intercellular spaces, and comprise transmembrane proteins including occludins, claudins, zona occludens, and junctional adhesion molecules (Edelblum and Turner, 2009; Groschwitz and Hogan, 2009). Dysfunctional TJ may lead to a “leaky” gut, which is characteristic of numerous intestinal disorders including inflammatory bowel diseases (Krug et al., 2014). Notably, previous work has suggested early life microbiota disruptions (via antibiotic usage) and reductions in *Bifidobacterium* are correlated with increased risk and/or symptoms of ulcerative colitis and Crohn’s disease (Kronman et al., 2012; Hildebrand et al., 2008; Favier et al., 1997; Shaw et al., 2010; Ng et al., 2011). Several clinical studies have indicated that supplementation with certain *Bifidobacterium* strains positively modulate gastrointestinal symptoms of patients, which is corrected with reductions of inflammatory markers in colonic IEC-containing biopsies; however, *B. breve* UCC2003 has not been used clinically in this patient setting (Furrie et al., 2005; Steed et al., 2010). Similar findings have also been reported in different animal models of intestinal inflammation (Philippe et al., 2011; Grimm et al., 2015; Zuo et al., 2014). A wide range of TJ-related genes were up-regulated after UCC2003 supplementation, particularly *Tjp1* (that encodes ZO-1), *Jam2*, and *Claudin34c1*, with a previous study indicating that other *Bifidobacterium* species (i.e., *B. bifidum*) also modulate TJ expression via ZO-1 (Din et al., 2020). These data indicated that specific strains of *Bifidobacterium* may modulate key barrier integrity systems during the neonatal period, and therefore absence of this key initial bacterial-host cross talk may correlate with increased risk of chronic intestinal disorders in later life (Shaw et al., 2010). Intestinal mucus, encoded by *Muc* genes (up-regulated due to *B. breve* UCC2003 in this study), plays a crucial role in colonic protection via formation of a physical barrier between the gut lumen and IECs, and deficiencies in MUC-2 have been linked with experimental colitis and increased inflammation in patients with inflammatory bowel disease (Shirazi et al., 2000; Van der Sluis et al., 2006). We have also observed that *B. breve* UCC2003 significantly increases goblet cell numbers and mucus production (in gnotobiotic and SPF mice; data not shown). Although the mucus layer may impact direct *Bifidobacterium*-IEC interactions, previous studies have indicated that *B. breve* UCC2003 surface molecules, such as EPS and the Tad pilus, may modulate IEC function via signaling through Toll-like receptors (TLRs) (O’Connell Motherway et al., 2019; Hughes et al., 2017). Moreover, bifidobacterial metabolites, such as short-chain fatty acids may also act to modulate the IEC transcriptome, with previous studies indicating enhanced expression of TJs and cadherins via acetate (Hsieh et al., 2015; Ling et al., 2016; Ewaschuk et al., 2008; Lewis et al., 2017).

Further network and functional analysis indicated that clusters of up-regulated DEGs were associated with cell maturation and cell differentiation (as confirmed by cell type-specific analysis), suggesting that neonatal *B. breve* exposure positively modulates IEC cell differentiation, growth, and maturation. Somewhat surprisingly, we did not observe the same type of striking responses in immune pathways, which may be masked by the sheer number of DEGs involved in cellular differentiation and processes. However, pathways such as TLR1 and TLR2 pathways do appear to be enriched (cluster 2 of signaling network analysis). This may link to previous work indicating that the UCC2003 EPS signals via TLR2 to induce MyD88 signaling cascades to protect IECs during intestinal inflammation (Hughes et al., 2017). *B. breve* M-16V was also shown to interact with TLR2 to up-regulate ubiquitin-editing enzyme A20 expression that correlated with increased tolerance to a TLR4 cascade in porcine IECs, further supporting the involvement of *B. breve* in programming key host immunoregulation receptors (Tomosada et al., 2013).

Cell type-specific analysis of DEGs revealed stem cells as the IEC type most affected by *B. breve*, with absorptive enterocytes least affected despite being most accessible to bacteria in the gut. It could be hypothesized that *B. breve* or their secreted metabolites may reach the crypts of the small intestinal epithelium. This has been previously suggested by *in situ* hybridization histology *in vivo* and by *Bifidobacterium*-conditioned media altering the expression of hundreds of host epithelial genes linked to immune response, cell adhesion, cell cycle, and development in IECs *in vitro* (Hughes et al., 2017; Guo et al., 2015). However, the direct impact of bifidobacterial-associated metabolites on these responses would require further studies to confirm metabolic activity of *B. breve* within the small intestine (via transcriptomics and metabolomics), although daily



supplementation with live bacteria may also provide a source of these metabolites in our model. Interestingly, certain *Bifidobacterium* and *Lactobacillus* strains that have been heat killed have also been shown to induce host responses, indicating that surface structures alone may play a role in downstream effects (Pique et al., 2019). All but two of the 37 differentially expressed stem cell marker genes were up-regulated in the presence of *B. breve* UCC2003, indicating an activating effect resulting in increased pluripotency of stem cells, increased quantity of stem cells, and/or an increased quantity of semi-differentiated cells. Single-cell sequencing of IECs could be used to further investigate this finding. Thirty-two TFs were predicted to regulate these stem cell signature genes, providing possible targets for future investigation of the mechanisms underlying these responses. Functional analysis of the stem cell signature genes and their regulators suggests that *B. breve* increases pluripotency of stem cells and/or semi-differentiated epithelial cells through WNT signaling and nuclear hormone signaling (Jeong and Mangelsdorf, 2009). Furthermore, the overrepresentation of the process "RUNX1 regulates transcription of genes involved in differentiation of HSCs" indicates a possible role for histone methylation in response to *B. breve* UCC2003 (Imperato et al., 2015). Further determination of host and bacterial metabolome and proteome after *B. breve* exposure may allow identification of the specific underlying molecular mechanisms (Guo et al., 2015).

In conclusion, *B. breve* UCC2003 plays a central role in orchestrating global neonatal IEC gene responses in a distinct manner as shown in our murine model, modulating genes involved in epithelial barrier development, and driving universal transcriptomic alteration that facilitates cell replication, differentiation, and growth, particularly within the stem cell compartment. This study enhances our overall understanding of the benefits of specific early life microbiota members in intestinal epithelium development, with prospective avenues to probe further health-promoting mechanisms of *Bifidobacterium* in humans. Further work exploring time-dependent transcriptional responses and impact of other *Bifidobacterium* species and strains (and use of mutants and transcriptionally active strains as positive controls), in tandem with metabolomic and proteomic approaches, is required to advance our understanding on the key host pathways and bifidobacterial molecules governing development and maturation of the intestinal barrier during the early life window. Nevertheless, further clinical studies would be essential to explore if these responses and findings are similar to those observed in humans.

Limitations of the Study

As we only observed low relative abundance of *Bifidobacterium* in our control neonatal animals this may suggest that induction of responses may be linked to the introduction of a new microbiota member (i.e., *B. breve* UCC2003), therefore results should be carefully interpreted. However, we did not observe associated global transcriptional inflammatory immune changes that would be expected if this was the case, but rather global changes in barrier function transcripts and pathways. Furthermore, *Bifidobacterium* has previously been isolated from C57BL/6 mice (including from our mouse colony), and therefore appears to be a resident rodent gut microbiota member, although it is found at varying abundances in different animal units and suppliers (Grimm et al., 2015; Hughes et al., 2020). Indeed, one particular study has shown that high levels of resident *Bifidobacterium* in mice directly correlated with improved immune responses to cancer immunotherapies (Sivan et al., 2015). In addition, we did not explore if *B. breve* UCC2003 is potentially driving more nuanced microbe-microbe interactions, and that, indirectly, these may also be stimulating IEC responses. Therefore, further studies probing these aspects in more detail, and comparing other *Bifidobacterium* strains, to compare and contrast responses, would be of interest.

B. breve UCC2003 is a model strain that was previously isolated from the stool of a breast-fed infant (National Collection of Industrial Food and Marine Bacteria, 2020; Sheehan et al., 2007). Although a human-associated strain, it has not been used in clinical studies, so directly extrapolating to human-specific settings should be cautiously considered. Further large-scale clinical studies would be required to confirm any positive strain-level impacts; however, in-depth analysis of, e.g., small IECs would be unethical in a healthy infant cohort, which emphasizes the importance of preclinical models.

Previous studies have shown that this strain can efficiently colonize (long term) the mouse gastrointestinal tract; however, we could not confirm this in our short-term, daily supplementation study (Cronin et al., 2008; O'Connell Motherway et al., 2011). Therefore the IEC responses observed may occur as a result of transient interactions with *B. breve* UCC2003 as it passes through the small intestine. Nevertheless, although at lower levels ($\sim 10^5$ CFU/g), we did observe viable *B. breve* UCC2003 in the small intestine, linking to our subsequent observations of significant impacts on the IEC transcriptome from this intestinal region.



Very-low-abundance microbiota members (<2% relative abundance), including *Streptococcus*, *Ruminococcus*, *Prevotella*, and *Coprococcus*, were significantly reduced in relative abundance compared with controls, raising the question whether supplementation of *Bifidobacterium* could have reduced these taxa. Regrettably, we could not determine if this is a bifidobacterial effect due to the lack of longitudinal samples, and we did not quantify bacterial titers, which is an important consideration for future work. We also did not profile microbial community composition within the small intestines, which is known to differ from fecal samples.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Lindsay J. Hall (Lindsay.Hall@quadram.ac.uk).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The code generated for RNA-seq analysis during this study is available at GitHub <https://github.com/raymondkiu/Bifidobacterium-IEC-transcriptomics>. The accession number for the raw sequencing reads (both RNA-seq and 16S rRNA amplicon sequencing) reported in this paper is European Nucleotide Archive (ENA): PRJEB36661.

METHODS

All methods can be found in the accompanying [Transparent Methods](#) supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101336>.

ACKNOWLEDGMENTS

This research was supported in part by the Norwich Bioscience Institutes (NBI) Computing infrastructure for Science (CiS) group through the provision of a High-Performance Computing (HPC) Cluster. L.J.H. is funded by a Wellcome Trust Investigator award (100974/C/13/Z) and together with T.K. by a BBSRC ISP grant for Gut Microbes and Health BB/R012490/1 and its constituent project(s), BBS/E/F/000PR10353 and BBS/E/F/000PR10355. T.K. is also funded by the Genomics for Food security CSP grant from the BBSRC (BB/CSP17270/1). A.T. is supported by the BBSRC Norwich Research Park Biosciences Doctoral Training Partnership (grant BB/M011216/1). D.v.S. is supported by Science Foundation Ireland (SFI/12/RC/2273-P1 and SFI/12/RC/2273-P2).

AUTHOR CONTRIBUTIONS

Conceptualization, R.K., L.C.H., and L.J.H.; Methodology, R.K., A.T., L.C.H., and L.J.H.; Software, R.K., A.T., and S.C.; Validation, R.K., A.T., T.K., and L.J.H.; Formal analysis, R.K. and A.T.; Investigation, R.K., A.T., L.C.H., and C.L.; Resources, S.C.; Data Curation, R.K.; Writing – Original Draft Preparation, R.K., A.T., and L.J.H.; Writing – Review and Editing, R.K., A.T., D.v.S., T.K., and L.J.H.; Visualization, R.K. and A.T.; Supervision, T.K. and L.J.H.; Project Administration, R.K.; Funding Acquisition, T.K., D.v.S., and L.J.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 30, 2020

Revised: June 5, 2020

Accepted: June 30, 2020

Published: July 24, 2020



REFERENCES

- Arrieta, M.C., Stiersma, L.T., Amenogbe, N., Brown, E.M., and Finlay, B. (2014). The intestinal microbiome in early life: health and disease. *Front. Immunol.* 5, 427.
- Baucells, B.J., Mercadal Hally, M., Alvarez Sanchez, A.T., and Figueras Aloy, J. (2016). Probiotic associations in the prevention of necrotizing enterocolitis and the reduction of late-onset sepsis and neonatal mortality in preterm infants under 1,500g: a systematic review. *An. Pediatr. (Barc.)* 85, 247–255.
- Chernyavsky, A.I., Arredondo, J., Karlsson, E., Wessler, I., and Grando, S.A. (2005). The Ras/Raf-1/MEK1/ERK signaling pathway coupled to integrin expression mediates cholinergic regulation of keratinocyte directional migration. *J. Biol. Chem.* 280, 39220–39228.
- Cronin, M., Sleator, R.D., Hill, C., Fitzgerald, G.F., and van Sinderen, D. (2008). Development of a luciferase-based reporter system to monitor *Bifidobacterium breve* UCC2003 persistence in mice. *BMC Microbiol.* 8, 161.
- Derrien, M., Alvarez, A.S., and de Vos, W.M. (2019). The gut microbiota in the first decade of life. *Trends Microbiol.* 27, 997–1010.
- Din, A.U., Hassan, A., and Zhu, Y. (2020). Inhibitory effect of *Bifidobacterium Bifidum* ATCC29521 on colitis and its mechanism. *J. Nutr. Biochem.* 79, 108353.
- Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl. Acad. Sci. U S A* 107, 11971–11975.
- Edelblum, K.L., and Turner, J.R. (2009). The tight junction in inflammatory disease: communication breakdown. *Curr. Opin. Pharmacol.* 9, 715–720.
- Ewaschuk, J.B., Diaz, H., Meddings, L., Dienerichs, B., Dmytrash, A., Backer, J., Looijer-van Langen, M., and Madsen, K.L. (2008). Secreted bioactive factors from *Bifidobacterium infantis* enhance epithelial cell barrier function. *Am. J. Physiol. Gastrointest. Liver Physiol.* 295, G1025–G1034.
- Favier, C., Neut, C., Mizon, C., Cortot, A., Colombel, J.F., and Mizon, J. (1997). Fecal beta-D-galactosidase production and *Bifidobacteria* are decreased in Crohn's disease. *Dig. Dis. Sci.* 42, 817–822.
- Fukuda, S., Toh, H., Taylor, T.D., Ohno, H., and Hattori, M. (2012). Acetate-producing *bifidobacteria* protect the host from enteropathogenic infection via carbohydrate transporters. *Gut Microbes* 3, 449–454.
- Furrie, E., Macfarlane, S., Kennedy, A., Cummings, J.H., Walsh, S.V., O'neil, D.A., and Macfarlane, G.T. (2005). Synbiotic therapy (*Bifidobacterium longum*/Synergy 1) initiates resolution of inflammation in patients with active ulcerative colitis: a randomised controlled pilot trial. *Gut* 54, 242–249.
- Gann, R.N. (2010). Host Signaling Response to Adhesion of *Bifidobacterium infantis*, All Graduate Theses and Dissertations Thesis (Utah State University).
- García-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* 29, 1363–1375.
- Grimm, V., Radulovic, K., and Riedel, C.U. (2015). Colonization of C57BL/6 mice by a potential probiotic *Bifidobacterium bifidum* strain under germ-free and specific pathogen-free conditions and during experimental colitis. *PLoS One* 10, e0139935.
- Groschwitz, K.R., and Hogan, S.P. (2009). Intestinal barrier function: molecular regulation and disease pathogenesis. *J. Allergy Clin. Immunol.* 124, 3–20, quiz 21–2.
- Guo, S., Guo, Y., Ergun, A., Lu, L., Walker, W.A., and Ganguli, K. (2015). Secreted metabolites of *Bifidobacterium infantis* and *Lactobacillus acidophilus* protect immature human enterocytes from IL-1 β -induced inflammation: a transcription profiling analysis. *PLoS One* 10, e0124549.
- Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., Tirosh, I., Beyaz, S., Dionne, D., and Zhang, M. (2017). A single-cell survey of the small intestinal epithelium. *Nature* 551, 333–339.
- Harburger, D.S., and Calderwood, D.A. (2009). Integrin signalling at a glance. *J. Cell Sci.* 122, 159–163.
- Hildebrand, H., Malmberg, P., Askling, J., Ekblom, A., and Montgomery, S.M. (2008). Early-life exposures associated with antibiotic use and risk of subsequent Crohn's disease. *Scand. J. Gastroenterol.* 43, 961–966.
- Holland, C.H., Szalai, B., and Saez-Rodriguez, J. (2019). Transfer of regulatory knowledge from human to mouse for functional genomics analysis. *Biochim. Biophys. Acta Gene Regul. Mech.* 1863, 194431.
- Hsieh, C.Y., Osaka, T., Moriyama, E., Date, Y., Kikuchi, J., and Tsuneda, S. (2015). Strengthening of the intestinal epithelial tight junction by *Bifidobacterium bifidum*. *Physiol. Rep.* 3, e12327.
- Hughes, K.R., Harnisch, L.C., Alcon-Giner, C., Mitra, S., Wright, C.J., Ketskemeti, J., van Sinderen, D., Watson, A.J., and Hall, L.J. (2017). *Bifidobacterium breve* reduces apoptotic epithelial cell shedding in an exopolysaccharide and MyD88-dependent manner. *Open Biol.* 7, 160155.
- Hughes, K.R., Schofield, Z., Dalby, M.J., Caim, S., Chalklen, L., Bernuzzi, F., Alcon-Giner, C., Le Gall, G., Watson, A.J.M., and Hall, L.J. (2020). The early life microbiota protects neonatal mice from pathological small intestinal epithelial cell shedding. *FASEB J.* 34, 7075–7088.
- Imperato, M.R., Cauchy, P., Obier, N., and Bonifer, C. (2015). The RUNX1-PU.1 axis in the control of hematopoiesis. *Int. J. Hematol.* 101, 319–329.
- Jacobs, S.E., Tobin, J.M., Opie, G.F., Donath, S., Tabrizi, S.N., Pirota, M., Morley, C.J., Garland, S.M., and Proprems Study, G. (2013). Probiotic effects on late-onset sepsis in very preterm infants: a randomized controlled trial. *Pediatrics* 132, 1055–1062.
- Jeong, Y., and Mangelsdorf, D.J. (2009). Nuclear receptor regulation of stemness and stem cell differentiation. *Exp. Mol. Med.* 41, 525–537.
- Kronman, M.P., Zaoutis, T.E., Haynes, K., Feng, R., and Coffin, S.E. (2012). Antibiotic exposure and IBD development among children: a population-based cohort study. *Pediatrics* 130, e794–803.
- Krug, S.M., Schulzke, J.D., and Fromm, M. (2014). Tight junction, selective permeability, and related diseases. *Semin. Cell Dev. Biol.* 30, 166–176.
- Lawson, M.A.E., O'neill, L.J., Kujawska, M., Gowriadh Javadi, S., Wijeyesekera, A., Flegg, Z., Chalklen, L., and Hall, L.J. (2020). Breast milk-derived human milk oligosaccharides promote *Bifidobacterium* interactions within a single ecosystem. *ISME J.* 14, 635–648.
- Lewis, M.C., Merrifield, C.A., Berger, B., Cloarec, O., Duncker, S., Mercenier, A., Nicholson, J.K., Holmes, E., and Bailey, M. (2017). Early intervention with *Bifidobacterium lactis* NCC2818 modulates the host-microbe interface independent of the sustained changes induced by the neonatal environment. *Sci. Rep.* 7, 5310.
- Li, L., Zhao, G.D., Shi, Z., Qi, L.L., Zhou, L.Y., and Fu, Z.X. (2016). The Ras/Raf/MEK/ERK signaling pathway and its role in the occurrence and development of HCC. *Oncol. Lett.* 12, 3045–3050.
- Ling, X., Linglong, P., Weixia, D., and Hong, W. (2016). Protective effects of *Bifidobacterium* on intestinal barrier function in LPS-induced enterocyte barrier injury of caco-2 monolayers and in a rat NEC model. *PLoS One* 11, e0161635.
- Liu, C., Zhang, Z.Y., Dong, K., and Guo, X.K. (2010). Adhesion and immunomodulatory effects of *Bifidobacterium lactis* HN019 on intestinal epithelial cells INT-407. *World J. Gastroenterol.* 16, 2283–2290.
- Mikami, K., Kimura, M., and Takahashi, H. (2012). Influence of maternal *bifidobacteria* on the development of gut *bifidobacteria* in infants. *Pharmaceuticals (Basel)* 5, 629–642.
- Mosher, D.F., Fogerty, F.J., Chernousov, M.A., and Barry, E.L. (1991). Assembly of fibronectin into extracellular matrix. *Ann. N. Y. Acad. Sci.* 614, 167–180.
- Nagpal, R., Kurakawa, T., Tsuji, H., Takahashi, T., Kawashima, K., Nagata, S., Nomoto, K., and Yamashiro, Y. (2017). Evolution of gut *Bifidobacterium* population in healthy Japanese infants over the first three years of life: a quantitative assessment. *Sci. Rep.* 7, 10097.
- National Collection of Industrial Food and Marine Bacteria (NCIMB) (2020). NCIMB 8807 General Info (Aberdeen).
- Ng, S.C., Benjamin, J.L., McCarthy, N.E., Hedlin, C.R., Koutsoumpas, A., Plamondon, S., Price, C.L., Hart, A.L., Kamm, M.A., Forbes, A., Knight, S.C., Lindsay, J.O., Whelan, K., and Stagg, A.J.



(2011). Relationship between human intestinal dendritic cells, gut microbiota, and disease activity in Crohn's disease. *Inflamm. Bowel Dis.* 17, 2027–2037.

Nusse, R. (2008). Wnt signaling and stem cell control. *Cell Res.* 18, 523–527.

O'Connell Motherway, M., Houston, A., O'callaghan, G., Reunanen, J., O'brien, F., O'driscoll, T., Casey, P.G., de Vos, W.M., van Sinderen, D., and Shanahan, F. (2019). A Bifidobacterial pilus-associated protein promotes colonic epithelial proliferation. *Mol. Microbiol.* 111, 287–301.

O'Connell Motherway, M., Zomer, A., Leahy, S.C., Reunanen, J., Bottacini, F., Claesson, M.J., O'brien, F., Flynn, K., Casey, P.G., Munoz, J.A., Kearney, B., Houston, A.M., and O'mahony, C. (2011). Functional genome analysis of Bifidobacterium breve UCC2003 reveals type IVb tight adherence (Tad) pili as an essential and conserved host-colonization factor. *Proc. Natl. Acad. Sci. U S A* 108, 11217–11222.

O'Neill, I., Schofield, Z., and Hall, L.J. (2017). Exploring the role of the microbiota member Bifidobacterium in modulating immune-linked diseases. *Emerg. Top. Life Sci.* 1, 333–349.

Pan, W.H., Sommer, F., Falk-Paulsen, M., Ulas, T., Best, P., Fazio, A., Kachroo, P., Luzzi, A., Jentsch, M., Rehman, A., Muller, F., Lengauer, T., Walter, J., Kunzel, S., and Baines, J.F. (2018). Exposure to the gut microbiota drives distinct methylome and transcriptome changes in intestinal epithelial cells during postnatal development. *Genome Med.* 10, 27.

Patole, S.K., Rao, S.C., Keil, A.D., Nathan, E.A., Doherty, D.A., and Simmer, K.N. (2016). Benefits of Bifidobacterium breve M-16V supplementation in preterm neonates - a retrospective cohort study. *PLoS One* 11, e0150775.

Philippe, D., Heupel, E., Blum-Sperisen, S., and Riedel, C.U. (2011). Treatment with Bifidobacterium bifidum 17 partially protects mice from Th1-driven inflammation in a chemically induced model of colitis. *Int. J. Food Microbiol.* 149, 45–49.

Pique, N., Berlanga, M., and Minana-Galbis, D. (2019). Health benefits of heat-killed (Tyndallized) probiotics: an overview. *Int. J. Mol. Sci.* 20, 2534.

Plantinga, T.S., van Maren, W.W., van Bergenhenegouwen, J., Hameetman, M., Nierkens, S., Jacobs, C., de Jong, D.J., Joosten, L.A., Van't Land, B., Garssen, J., Adema, G.J., and Netea, M.G. (2011). Differential Toll-like receptor

recognition and induction of cytokine profile by Bifidobacterium breve and Lactobacillus strains of probiotics. *Clin. Vaccin. Immunol.* 18, 621–628.

Plaza-Diaz, J., Gomez-Llorente, C., Fontana, L., and Gil, A. (2014). Modulation of immunity and inflammatory gene expression in the gut, in inflammatory diseases of the gut and in the liver by probiotics. *World J. Gastroenterol.* 20, 15632–15649.

Plummer, E.L., Bulach, D.M., Murray, G.L., Jacobs, S.E., Tabrizi, S.N., Garland, S.M., and Propremis Study, G. (2018). Gut microbiota of preterm infants supplemented with probiotics: sub-study of the ProPremis trial. *BMC Microbiol.* 18, 184.

Qin, J., Vinogradova, O., and Plov, E.F. (2004). Integrin bidirectional signaling: a molecular view. *PLoS Biol.* 2, e169.

Riedel, C.U., Fosta, F., Philippe, D., Adolffson, O., Eikmanns, B.J., and Blum, S. (2006). Anti-inflammatory effects of bifidobacteria by inhibition of LPS-induced NF-kappaB activation. *World J. Gastroenterol.* 12, 3729–3735.

Robertson, C., Sava, G.M., Clapuci, R., Jones, J., Maimouni, H., Brown, E., Minocha, A., Hall, L.J., and Clarke, P. (2020). Incidence of necrotising enterocolitis before and after introducing routine prophylactic Lactobacillus and Bifidobacterium probiotics. *Arch. Dis. Child. Fetal Neonatal Ed.* 105, 380–386.

Shaw, S.Y., Blanchard, J.F., and Bernstein, C.N. (2010). Association between the use of antibiotics in the first year of life and pediatric inflammatory bowel disease. *Am. J. Gastroenterol.* 105, 2687–2692.

Sheehan, V.M., Sleator, R.D., Hill, C., and Fitzgerald, G.F. (2007). Improving gastric transit, gastrointestinal persistence and therapeutic efficacy of the probiotic strain Bifidobacterium breve UCC2003. *Microbiology* 153, 3563–3571.

Shirazi, T., Longman, R.J., Corfield, A.P., and Probert, C.S. (2000). Mucins and inflammatory bowel disease. *Postgrad. Med. J.* 76, 473–478.

Sivan, A., Corrales, L., Hubert, N., Williams, J.B., Aquino-Michaels, K., Earley, Z.M., Benyamin, F.W., Lei, Y.M., Jabri, B., Alegre, M.L., Chang, E.B., and Gajewski, T.F. (2015). Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science* 350, 1084–1089.

Srutkova, D., Schwarzer, M., Hudcovic, T., Zakostelska, Z., Drab, V., Spanova, A., Rittich, B., Kozakova, H., and Schabussova, I. (2015).

Bifidobacterium longum CCM 7952 promotes epithelial barrier function and prevents acute DSS-induced colitis in strictly strain-specific manner. *PLoS One* 10, e0134050.

Steed, H., Macfarlane, G.T., Blackett, K.L., Bahrami, B., Reynolds, N., Walsh, S.V., Cummings, J.H., and Macfarlane, S. (2010). Clinical trial: the microbiological and immunological effects of synbiotic consumption - a randomized double-blind placebo-controlled study in active Crohn's disease. *Aliment. Pharmacol. Ther.* 32, 872–883.

Stewart, C.J., Ajami, N.J., O'brien, J.L., Hutchinson, D.S., Smith, D.P., Wong, M.C., Ross, M.C., Lloyd, R.E., Doddapaneni, H., Metcalf, G.A., Muzny, D., Gibbs, R.A., Vatanen, T., Huttenhower, C., Xavier, R.J., and Rewers, M. (2018). Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562, 583–588.

Thoo, L., Noti, M., and Krebs, P. (2019). Keep calm: the intestinal barrier at the interface of peace and war. *Cell Death Dis.* 10, 849.

Tomosada, Y., Villena, J., Murata, K., Chiba, E., Shimazu, T., Aso, H., Iwabuchi, N., Xiao, J.Z., Saito, T., and Kitazawa, H. (2013). Immunoregulatory effect of bifidobacteria strains in porcine intestinal epithelial cells through modulation of ubiquitin-editing enzyme A20 expression. *PLoS One* 8, e59259.

Turrioni, F., Taverniti, V., Ruas-Madiedo, P., Duranti, S., Guglielmetti, S., Lugli, G.A., Gioiosa, L., Palanza, P., Margolles, A., van Sinderen, D., and Ventura, M. (2014). Bifidobacterium bifidum PRL2010 modulates the host innate immune response. *Appl. Environ. Microbiol.* 80, 730–740.

Van der Sluis, M., de Koning, B.A., de Bruijn, A.C., Velich, A., Meijerink, J.P., van Goudoever, J.B., Buller, H.A., Dekker, J., van Seuningen, I., Renes, I.B., and Einerhand, A.W. (2006). Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* 131, 117–129.

Zheng, B., van Bergenhenegouwen, J., Overbeek, S., van de Kant, H.J., Garssen, J., Folkerts, G., Vos, P., Morgan, M.E., and Kraneveld, A.D. (2014). Bifidobacterium breve attenuates murine dextran sodium sulfate-induced colitis and increases regulatory T cell responses. *PLoS One* 9, e95441.

Zuo, L., Yuan, K.T., Yu, L., Meng, Q.H., Chung, P.C., and Yang, D.H. (2014). Bifidobacterium infantis attenuates colitis by regulating T cell subset responses. *World J. Gastroenterol.* 20, 18316–18329.

iScience, Volume 23

Supplemental Information

***Bifidobacterium breve* UCC2003 Induces
a Distinct Global Transcriptomic Program
in Neonatal Murine Intestinal Epithelial Cells**

**Raymond Kiu, Agatha Treveil, Lukas C. Harnisch, Shabbonam Caim, Charlotte
Leclaire, Douwe van Sinderen, Tamas Korcsmaros, and Lindsay J. Hall**

SUPPLEMENTAL FIGURE

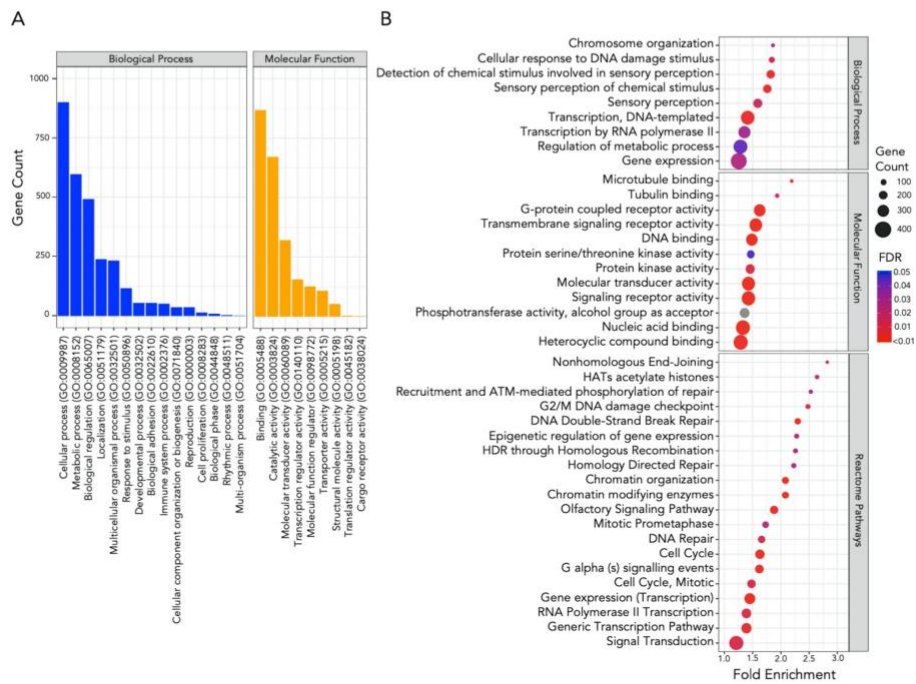


Figure S1. Functional analysis on differentially expressed genes. Related to Figure 4.

(A) Panther Slim GO-term major categories of significantly up-regulated genes ($n=3,996$). Related to Table S7 and Table S8.

(B) Functional and pathway enrichment analysis on significantly up-regulated genes (Panther Slim GO-term). Only top 20 FDR-ranked enriched pathways (Reactome pathways) are shown. Statistical significance cut-offs: $FDR < 0.05$. Statistical significance: Fisher's Exact Test. Fold Enrichment was calculated against all expressed genes in IECs as the background ($n=21,537$). Related to Table S5, Table S6 and Table S9.

SUPPLEMENTAL TABLES

Gene	Ensembl ID	Chromosome	Description/ Putative function	Biological process (GO)
<i>Gm27149</i>	ENSMUSG00000098426	7	Unknown	Unknown
<i>Ccnb1ip1</i>	ENSMUSG00000071470	14	Cyclin B1 interacting protein 1	Cell differentiation, Cellular component organisation, Protein metabolic process
<i>Gm10359</i>	ENSMUSG00000094708	5	glyceraldehyde-3-phosphate dehydrogenase pseudogene	Unknown
<i>Gm12671</i>	ENSMUSG00000095937	4	glyceraldehyde-3-phosphate dehydrogenase pseudogene	Unknown
<i>Gm48216</i>	ENSMUSG000000114367	13	Unknown	Unknown
<i>Gm17131</i>	ENSMUSG00000085328	5	Unknown	Unknown
<i>Tmem72</i>	ENSMUSG00000048108	6	transmembrane protein 72	Unknown
<i>Ccdc107</i>	ENSMUSG00000028461	4	coiled-coil domain containing 107	Unknown
<i>CR936839.3</i>	ENSMUSG000000111855	Unknown	Unknown	Unknown
<i>Hist1h4b</i>	ENSMUSG00000069266	13	H4 clustered histone 2	Cell differentiation, Cellular component organisation, Immune system process, system development
<i>Gm48836</i>	ENSMUSG000000113523	12	Unknown	Unknown
<i>Gm42669</i>	ENSMUSG000000106631	5	Unknown	Unknown
<i>Gprin3</i>	ENSMUSG00000045441	6	GPRIN family member 3	Unknown
<i>Fgd4</i>	ENSMUSG00000022788	16	FYVE, RhoGEF and PH domain containing 4	Cellular component of organisation, Protein metabolic process, Response to stimulus, Signaling
<i>D10Wsu102e</i>	ENSMUSG00000020255	10	DNA segment, Chr 10, Wayne State University 102, expressed	Unknown
<i>Gm20594</i>	ENSMUSG00000096887	6	Unknown	Cell death, Response to stimulus, Signaling
<i>Vps13b</i>	ENSMUSG00000037646	15	vacuolar protein sorting 13B	Establishment of localisation
<i>Gm48366</i>	ENSMUSG000000113523	12	Unknown	Unknown
<i>Gm48054</i>	ENSMUSG000000113921	13	Unknown	Unknown
<i>Naip6</i>	ENSMUSG00000078942	13	NLR family, apoptosis inhibitory protein 6	Cell death, Immune system process, Response to stimulus

Table S3. Annotation of top 20 significantly up-regulated genes. Related to Figure 4.

Gene	Ensembl ID	Chromosome	Description/ Putative function	Biological process (GO)
<i>Gm7849</i>	ENSMUSG00000079114	8	Defensin	Immune system process, response to stimulus
<i>Rps27rt</i>	ENSMUSG00000050621	9	ribosomal protein S27, retrogene	Unknown
<i>Gm6158</i>	ENSMUSG00000090381	14	Unknown	Unknown
<i>mt-Co2</i>	ENSMUSG00000064354	MT	mitochondrially encoded cytochrome c oxidase II	Carbohydrate derivative metabolism, cell death
<i>Gm3650</i>	ENSMUSG00000097891	18	Unknown	Unknown
<i>Gm5831</i>	ENSMUSG00000111133	9	Unknown	Unknown
<i>Gm7331</i>	ENSMUSG00000059461	X	Unknown	Unknown
<i>Polr3h</i>	ENSMUSG00000022476	15	polymerase (RNA) III (DNA directed) polypeptide H	Immune system process, response to stimulus
<i>Gm14698</i>	ENSMUSG00000071748	X	Unknown	Unknown
<i>Ftl1-ps1</i>	ENSMUSG00000062382	13	erritin light polypeptide 1	Unknown
<i>Rps16</i>	ENSMUSG00000037563	7	ribosomal protein S16	Protein metabolic process, response to stimulus
<i>Mt2</i>	ENSMUSG00000031762	8	metallothionein 2	Homeostatic process, response to stimulus, signaling
<i>Mt1</i>	ENSMUSG00000031765	8	metallothionein 1	Cell death, homeostatic process, response to stimulus, signaling
<i>Hba-a1</i>	ENSMUSG00000069919	11	hemoglobin alpha	Cell differentiation, homeostatic process, immune system process, response to stimulus, signaling, system development
<i>Hbb-bt</i>	ENSMUSG00000073940	7	hemoglobin, beta adult t chain	Unknown
<i>Rps2-ps10</i>	ENSMUSG00000091957	18	ribosomal protein S2	Unknown
<i>Gpx3</i>	ENSMUSG00000018339	11	glutathione peroxidase 3	Response to stimulus
<i>Smcp</i>	ENSMUSG00000074435	3	sperm mitochondria-associated cysteine-rich protei	Unknown
<i>Gm49384</i>	ENSMUSG00000113786	12	Predicted gene	Unknown
<i>Gm7336</i>	ENSMUSG00000078636	7	Predicted gene	Unknown

Table S4. Annotation of top 20 significantly down-regulated genes. Related to Figure 4.

GO Biological Process	Gene Count	Fold Enrichment	FDR
Transcription, DNA-templated	252	1.41	1.15E-03
Detection of chemical stimulus involved in sensory perception	97	1.82	1.92E-03
Sensory perception of chemical stimulus	98	1.76	1.97E-03
Cellular response to DNA damage stimulus	66	1.84	1.54E-02
Sensory perception	111	1.59	1.57E-02
Gene expression	350	1.25	2.66E-02
Chromosome organization	57	1.86	2.66E-02
Transcription by RNA polymerase II	193	1.35	3.09E-02
Regulation of metabolic process	262	1.28	4.48E-02

Table S5. GO Biological Process enrichment analysis in up-regulated DEGs. Related to Figure 4 and Figure S1.

GO Molecular Functions	Gene Count	Fold Enrichment	FDR
Transmembrane signaling receptor activity	279	1.56	5.07E-08
G-protein coupled receptor activity	224	1.63	8.61E-08
Molecular transducer activity	321	1.43	1.76E-06
Signaling receptor activity	302	1.43	4.06E-06
DNA binding	225	1.49	1.20E-05
Nucleic acid binding	344	1.33	1.14E-04
Heterocyclic compound binding	352	1.29	7.08E-04
Microtubule binding	43	2.2	3.01E-03
Protein kinase activity	133	1.46	7.97E-03
Tubulin binding	45	1.94	1.24E-02
Protein serine/threonine kinase activity	94	1.47	4.74E-02
Phosphotransferase activity, alcohol group as acceptor	144	1.36	5.04E-02

Table S6. GO Molecular Functions enrichment analysis in up-regulated DEGs. Related to Figure 4 and Figure S1.

Appendix 6: Peer-reviewed publications 300

Biological Process	Genes	Percentage
cellular process (GO:0009987)	901	31.70%
metabolic process (GO:0008152)	597	21.00%
biological regulation (GO:0065007)	492	17.30%
localization (GO:0051179)	239	8.40%
multicellular organismal process (GO:0032501)	233	8.20%
response to stimulus (GO:0050896)	117	4.10%
developmental process (GO:0032502)	55	1.90%
biological adhesion (GO:0022610)	55	1.90%
immune system process (GO:0002376)	52	1.80%
cellular component organization or biogenesis (GO:0071840)	37	1.30%
reproduction (GO:0000003)	37	1.30%
cell proliferation (GO:0008283)	15	0.50%
biological phase (GO:0044848)	10	0.40%
rhythmic process (GO:0048511)	4	0.10%
multi-organism process (GO:0051704)	2	0.10%

Table S7. GO Biological Process functional assignment to 3,996 upregulated genes. Related to Figure 4 and Figure S1.

Molecular Function	Genes	Percentage
binding (GO:0005488)	868	37.60%
catalytic activity (GO:0003824)	671	29.00%
molecular transducer activity (GO:0060089)	321	13.90%
transcription regulator activity (GO:0140110)	156	6.80%
molecular function regulator (GO:0098772)	127	5.50%
transporter activity (GO:0005215)	108	4.70%
structural molecule activity (GO:0005198)	53	2.30%
translation regulator activity (GO:0045182)	4	0.20%
cargo receptor activity (GO:0038024)	2	0.10%

Table S8. GO Molecular Function functional assignment to 3,996 upregulated genes. Related to Figure 4 and Figure S1.

Reactome Pathway	Gene Count	Fold Enrichment	FDR
Olfactory Signaling Pathway	108	1.88	3.20E-05
Cell Cycle	154	1.63	1.13E-04
Gene expression (Transcription)	238	1.45	1.18E-04
Chromatin organization	67	2.08	2.54E-04
Chromatin modifying enzymeS	67	2.08	2.82E-04
DNA Double-Strand Break Repair	51	2.3	4.19E-04
G alpha (s) signalling events	122	1.62	6.50E-04
RNA Polymerase II Transcription	202	1.39	3.06E-03
Generic Transcription Pathway	173	1.39	8.50E-03
Signal Transduction	461	1.21	8.98E-03
G2/M DNA damage checkpoint	29	2.48	9.07E-03
DNA Repair	77	1.66	1.06E-02
Cell Cycle, Mitotic	122	1.48	1.06E-02
HATs acetylate histones	23	2.64	1.85E-02
Epigenetic regulation of gene expression	28	2.28	2.38E-02
Homology Directed Repair	30	2.23	2.39E-02
Nonhomologous End-Joining	19	2.82	2.46E-02
Mitotic Prometaphase	58	1.73	2.48E-02
HDR through Homologous Recombination	29	2.26	2.63E-02
Recruitment and ATM-mediated phosphorylation of repair	22	2.53	2.78E-02

Table S9. Reactome pathway enrichment analysis in up-regulated DEGs. Related to Figure 4 and Figure S1.

Reactome Pathway	Gene Count	Fold Enrichment	FDR
Regulation of FZD by ubiquitination (R-MMU-4641263)	3	35.6	2.39E-02
Endogenous sterols (R-MMU-211976)	3	27.06	3.63E-02
Estrogen-dependent gene expression (R-MMU-9018519)	6	19.61	4.95E-04
ESR-mediated signaling (R-MMU-8939211)	6	18.04	4.69E-04
RUNX1 regulates transcription of genes involved in differentiation of HSCs (R-MMU-8939236)	4	14.78	3.41E-02
Signaling by Nuclear Receptors (R-MMU-9006931)	6	11.66	4.23E-03
Transcriptional regulation by RUNX1 (R-MMU-8878171)	6	9.73	9.64E-03
Generic Transcription Pathway (R-MMU-212436)	15	5.88	4.71E-05
RNA Polymerase II Transcription (R-MMU-73857)	15	5	1.89E-04
Gene expression (Transcription) (R-MMU-74160)	15	4.38	4.93E-04
Signal Transduction (R-MMU-162582)	19	2.37	3.65E-02

Table S12. Reactome pathway enrichment analysis of differentially expressed stem cell signature genes and their expressed regulators. Related to Figure 5.

TRANSPARENT METHODS

Animals

All animal experiments and related protocols were performed in accordance with the Animals (Scientific Procedures) Act 1986 (ASPA) under project licence (PPL: 80/2545) and personal licence (PIL: I68D4DCCF), approved by UK Home Office and University of East Anglia (UEA) FMH Research Ethics Committee. C57BL/6J two-week-old neonatal female mice ($n=10$) were housed in two separate cages with their mothers within UEA Disease Modelling Unit. Mice were euthanised via ASPA Schedule 1 protocol (CO_2 and cervical dislocation).

Bacterial culturing, inoculum preparation and CFU enumeration

B. breve UCC2003 (also known as NCIMB 8807) was streaked from frozen glycerol stocks onto autoclaved Reinforced Clostridial Agar (RCA) plates (Oxoid, UK) and incubated in an anaerobic chamber (miniMACS, Don Whitley Scientific) at 37°C for 48 h prior to picking single colonies for inoculation in prewarmed sterilised Reinforced Clostridial Medium (Oxoid, UK).

For preparation of gavage inoculums, 5 ml of inoculated broth was incubated overnight followed by sub-culturing into 5 ml De Man, Rogosa and Sharpe (MRS) medium (Oxoid). After an additional overnight incubation, another sub-culturing into 40 ml RCM was performed. Inoculums were prepared from cultures by 3 rounds of centrifugation at 3220 g for 10 min followed by three PBS washes before dilution in 4 ml (adult mice) or 2 ml (neonatal mice) sterile PBS. Bacterial concentration of inoculum was enumerated by plating serial dilutions in sterile PBS on RCA plates and enumerating colonies following two-day incubation to calculate CFU/ml.

Bacterial treatment and administration

Neonatal mice were orally gaged with *B. breve* UCC2003 inoculations of 10^8 CFU/ml in 50 μl every 24 h for 3 consecutive days. Control mice received oral gavages of sterile PBS. *B. breve* UCC2003 viable presence/transition through the gut was confirmed by collection of fresh faeces or intestinal content homogenised with 1 ml sterile PBS followed by serial-dilution plating in sterile PBS on RCA supplemented with 50 mg/L mupirocin and counting of colonies following two-day incubation to calculate CFU/mg.

Gut microbiota profiling by 16S rRNA amplicon sequencing and analysis

Genomic DNA extraction of mouse faecal samples on day 4 was performed with FastDNA Spin Kit for Soil (MP Biomedicals) following manufacturer's instructions and extending the bead-beating step to 3 min as described previously (Alcon-Giner et al., 2019). Extracted DNA was quantified, normalised and sequenced on Illumina MiSeq platform using a read length of 2×300 bp. After quality pre-filtering and removals of chimeras, sequencing reads were analysed using open-reference OTU clustering strategy (QIIME v1.9.1) to assign bacterial taxonomy based on SILVA_132 database (Quast et al., 2013). OTU tables in BIOM format was converted to genus counts in MEGAN6 and visualised using R library *ggplot2* as described previously (Kiu et al., 2019, Caporaso et al., 2010, Huson et al., 2016).

Tissue collection and isolation of small intestinal epithelial cells (IECs)

Upon tissue harvesting, 0.5 cm^2 sections of small intestines were collected in 200 μl RNeasy Lysis Buffer (Thermo Fisher Scientific) at the animal unit prior to IECs isolation (from fresh samples) via an adapted Weisser method as described below (Hughes et al., 2017). Sections of small intestines were placed in ice-cold PBS in 200 ml Duran bottles. Faecal matter was washed off by inverting 10 times in 0.154M NaCl and 1mM DTT. Liquid was drained and mucus layer removed through incubation of samples in 1.5mM KCl, 96mM NaCl, 27 mM Tri-sodium citrate, 8mM NaH_2PO_4 and 5.6mM Na_2HPO_4 for 15 min at 220 rpm and 37°C . IECs were separated from basal membrane by incubation in 1.5 mM EDTA and 0.5 mM DTT for 15 min at 200 rpm and 37°C followed by shaking vigorously 20 times. IECs were collected from solution by centrifugation at 500 g for 10 min at 4°C . Supernatant was then discarded and cell pellet resuspended in 3 ml of ice-cold PBS. Cell concentrations of

isolated IEC samples calculated by labelling dead cell with trypan blue at a 1:1 v/v ratio and enumeration of viable cells using a Neubauer haemocytometer on an inverted microscope (ID03, Zeiss).

RNA extraction and sequencing

RNA was extracted from IECs by adding a volume containing 2×10^6 cells in PBS to QIAshredder spin columns (QIAGEN) followed by centrifugation at 9,300 g for 1 min. Flow-through was mixed with 600 μ l RLT lysis buffer and used for subsequent RNA isolation. Homogenised samples in RLT buffer from IECs were processed by adding 700 μ l of 70% ethanol and mixing by pipetting. Samples were then added into RNeasy spin column and spun at 8,000 g for 15 s. Flow-through was discarded and process repeated until all of sample was filtered through column. Then 700 μ l of buffer RW1 was added to column and centrifuge at 8,000 g for 30 s. Again, flow through was discarded and filter placed in a new collection tube. To the filter, 500 μ l RPE was added and spun at 8,000 g for 30 s followed by discarding of flow through. An additional 500 μ l RPE was pipetted into column and centrifuged at 8,000 g for 2 min. Spin column was then placed in a new collection tube and centrifuged at 8,000 g for 2 min. Columns were transferred to a RNA low-bind Eppendorf tube and 30 μ l of RNase free water added to directly to the filter. After an incubation of 1 min at RT, sample was centrifuged at 8,000 g for 1 min and flow through containing RNA stored at -80°C .

Purified RNA was quantified, and quality controlled using RNA 6000 Nano kit on a 2100 Bioanalyser (Agilent). Only samples with RIN values above 8 were sequenced. RNA sequencing was performed at the Wellcome Trust Sanger Institute (Hinxton, UK) on paired-end 75 bp inserts on an Illumina HiSeq 2000 platform. Isolated RNA was processed by poly-A selection and/or Ribo-depletion.

Sequence pre-processing and Differential Gene Expression (DGE) analysis

Sequencing quality of raw FASTQ reads were assessed by FastQC software (v0.11.8). FASTQ reads were subsequently quality-filtered using fastp v0.20.0 with options `-q 10` (phred quality <10 was discarded) followed by merging reads into single read file for each sample (`merge-paired-reads.sh`) and rRNA sequence filtering via SortMeRNA v2.1 based on SILVA rRNA database optimised for SortMeRNA software (Chen et al., 2018, Kopylova et al., 2012). Filtered reads were then unmerged (`unmerge-paired-reads.sh`) and ready for transcript quantification.

Transcript mapping and quantification were performed using Kallisto v0.44.0 (Bray et al., 2016). Briefly, *Mus musculus* (C57BL/6 mouse) cDNA sequences (GRCm38.release-98_k31) were retrieved from Ensembl database and built into an index database with Kallisto utility `index` at default parameter that was used for following transcript mapping and abundance quantification via Kallisto utility `quant` at 100 bootstrap replicates (`-b 100`) (Zerbino et al., 2018).

DGE analysis was performed using R library Sleuth (v0.30.0) (Pimentel et al., 2017). Gene transcripts were mapped to individual genes using Ensembl BioMart database with Sleuth function `sleuth_prep` with option `gene_mode = TRUE`. Genes with an absolute $\log_2(\text{fold change}) > 1.0$ (based on Wald test statistics) and q value < 0.05 (or, FDR-adjusted p value; based on likelihood ratio test) were considered to be significantly regulated (Kinsella et al., 2011).

Functional annotation and enrichment analysis

Functional assignment and enrichment analysis was performed using PANTHER Classification System (Mi et al., 2019a). Briefly, for functional assignment analysis, a list of genes of interest in Ensembl IDs were supplied to the webserver to be mapped to the Mouse Genome Database (MGD) to generate functional classification on those genes of interest (Bult et al., 2019). For functional enrichment analysis, a gene list was supplied together with a background gene list in Ensembl IDs to Panther web server, then selected 'functional overrepresentation test' and chose a particular function class in the drop-down menu. Recommended by the database developers, Fisher's exact test and False Discovery Rate (FDR) correction were used to perform enrichment analysis (Mi et

al., 2019b). FDR <0.05 was used as the default cut-off for significant enrichment. Functional annotation of top 20 up/down-regulated genes was assigned manually via Ensembl and/or MGI (Mouse Genome Informatics) databases (Bult et al., 2019, Cunningham et al., 2019).

Network, cluster and signalling pathway analysis

A signalling network of all up-regulated DEGs and their first neighbours was built using all available biological signalling databases in the Cytoscape (v3.7.2) OmniPath app (v1, *Mus musculus*) (Turei et al., 2016, Shannon et al., 2003). Modules of highly connected genes within the signalling network were identified using the MCODE plug-in within Cytoscape (Bader and Hogue, 2003). Settings below were applied to obtain clusters in the network: degree cutoff = 3, haircut = true, fluff = false, node score cutoff = 0.5, k-core = 3 and max depth = 100.

The nodes of each individual module were tested for functional enrichment based on both Reactome and PANTHER annotations using PANTHER Classification System as described in previous sub-section 'Functional annotation and enrichment analysis' (Mi and Thomas, 2009, Croft et al., 2011, Mi et al., 2019a).

Enrichment of cell type specific marker genes

Cell type signature gene sets for murine intestinal epithelial cells were obtained from Haber et al. (Haber et al., 2017). Both droplet and plate-based results were used. Gene symbols were converted to Ensembl IDs using db2db (Mudunuri et al., 2009). Hypergeometric significance calculations were applied to test the presence of cell type specific signatures in the list of differentially expressed genes using all expressed genes as the statistical background (normalised counts > 1 in ≥ 1 sample). Bonferroni multiple correction was applied and any corrected $p < 0.05$ was deemed significant. Genes with normalised counts > 1 in ≥ 1 sample per condition (*B. breve* UCC2003 treated or control) were used to identify cell type signature genes expressed per condition.

Key regulator analysis

All mouse transcription factor - target gene interactions with quality scores A-D were obtained from DoRothEA v2 via the OmniPath Cytoscape app (Garcia-Alonso et al., 2019, Shannon et al., 2003, Turei et al., 2016). A subnetwork was generated consisting of differentially expressed stem cell signature genes and all their upstream TFs which were expressed in the transcriptomics dataset (normalised counts > 1 in ≥ 1 sample). These TFs were further filtered for their relevance in the network. Here all expressed genes and their upstream expressed TFs were extracted from the DoRothEA network. A hypergeometric significance test was carried out on any node with degree ≥ 5 to determine if the proportion of connected nodes which are differentially expressed is higher than in the whole network. Any TF with $p < 0.05$ following Benjamini-Hochberg correction were deemed significant and used to filter the stem cell signature gene subnetwork. Network visualisation was carried out in Cytoscape (Shannon et al., 2003). Functional enrichment carried out against Reactome pathways as described in previous sub-sections.

Statistical analyses and graphing

Student t-tests were performed using Rv.3.6.0, details of which were provided in the results and figure legends (R Development Core Team, 2010). LDA statistical tests for microbiome analysis was performed via LEfSe on Galaxy platform using default parameters (Segata et al., 2011, Jalili et al., 2020). PCA was performed via R library *ggfortify* function *autoplot* and *prcomp*, while Shannon diversity index was computed via R library *vegan* (Dixon, 2003, Tang et al., 2016, R Development Core Team, 2010). All other relevant statistical analyses (including enrichment analysis) were performed within specific software and details were provided in figure legends or as described in the previous sections.

All statistical graphs were either plotted using R library *ggplot2* or *Sleuth* (Wickham, 2016, Pimentel et al., 2017). Heatmaps were graphed using R library *gplots* function *heatmap.2* (Warnes et al., 2016).

Ethics approval

Animal experiments were performed under the UK Regulation of Animals (Scientific Procedures) Act of 1986. The project licence (PPL 80/2545) under which these studies were carried out was approved by the UK Home Office and the UEA Ethical Review Committee. Mice were sacrificed by CO₂ and cervical dislocation.

SUPPLEMENTAL REFERENCES

- ALCON-GINER, C., DALBY, M. J., CAIM, S., KETSKEMETY, J., SHAW, A., SIM, K., LAWSON, M., KIU, R., LECLAIRE, C., CHALKLEN, L., KUJAWSKA, M., MITRA, S., FARDUS-REID, F., BELTEKI, G., MCCOLL, K., SWANN, J. R., KROLL, J. S., CLARKE, P. & HALL, L. J. 2019. Microbiota supplementation with *Bifidobacterium* and *Lactobacillus* modifies the preterm infant gut microbiota and metabolome. *bioRxiv*, 698092.
- BADER, G. D. & HOGUE, C. W. 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4, 2.
- BRAY, N. L., PIMENTEL, H., MELSTED, P. & PACHTER, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*, 34, 525-7.
- BULT, C. J., BLAKE, J. A., SMITH, C. L., KADIN, J. A., RICHARDSON, J. E. & MOUSE GENOME DATABASE, G. 2019. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*, 47, D801-D806.
- CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PENA, A. G., GOODRICH, J. K., GORDON, J. I., HUTTLEY, G. A., KELLEY, S. T., KNIGHTS, D., KOENIG, J. E., LEY, R. E., LOZUPONE, C. A., MCDONALD, D., MUEGGE, B. D., PIRRUNG, M., REEDER, J., SEVINSKY, J. R., TURNBAUGH, P. J., WALTERS, W. A., WIDMANN, J., YATSUNENKO, T., ZANEVELD, J. & KNIGHT, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 7, 335-6.
- CHEN, S., ZHOU, Y., CHEN, Y. & GU, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34, i884-i890.
- CROFT, D., O'KELLY, G., WU, G., HAW, R., GILLESPIE, M., MATTHEWS, L., CAUDY, M., GARAPATI, P., GOPINATH, G., JASSAL, B., JUPE, S., KALATSKAYA, I., MAHAJAN, S., MAY, B., NDEGWA, N., SCHMIDT, E., SHAMOVSKY, V., YUNG, C., BIRNEY, E., HERMIAKOB, H., D'EUSTACHIO, P. & STEIN, L. 2011. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res*, 39, D691-7.
- CUNNINGHAM, F., ACHUTHAN, P., AKANNI, W., ALLEN, J., AMODE, M. R., ARMEAN, I. M., BENNETT, R., BHAI, J., BILLIS, K., BODDU, S., CUMMINS, C., DAVIDSON, C., DODIYA, K. J., GALL, A., GIRON, C. G., GIL, L., GREGO, T., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., KAY, M., LAIRD, M. R., LAVIDAS, I., LIU, Z., LOVELAND, J. E., MARUGAN, J. C., MAUREL, T., MCMAHON, A. C., MOORE, B., MORALES, J., MUDGE, J. M., NUHN, M., OGEH, D., PARKER, A., PARTON, A., PATRICIO, M., ABDUL SALAM, A. I., SCHMITT, B. M., SCHUILENBURG, H., SHEPPARD, D., SPARROW, H., STAPLETON, E., SZUBA, M., TAYLOR, K., THREADGOLD, G., THORMANN, A., VULLO, A., WALT, B., WINTERBOTTOM, A., ZADISSA, A., CHAKIACHVILI, M., FRANKISH, A., HUNT, S. E., KOSTADIMA, M., LANGRIDGE, N., MARTIN, F. J., MUFFATO, M., PERRY, E., RUFFIER, M., STAINES, D. M., TREVANION, S. J., AKEN, B. L., YATES, A. D., ZERBINO, D. R. & FLICEK, P. 2019. Ensembl 2019. *Nucleic Acids Res*, 47, D745-D751.
- DIXON, P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci*, 14, 927-930.
- HUSON, D. H., BEIER, S., FLADE, I., GORSKA, A., EL-HADIDI, M., MITRA, S., RUSCHEWEYH, H. J. & TAPPU, R. 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Comput Biol*, 12, e1004957.
- JALILI, V., AFGAN, E., GU, Q., CLEMENTS, D., BLANKENBERG, D., GOECKS, J., TAYLOR, J. & NEKRUTENKO, A. 2020. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res*.

- KINSELLA, R. J., KAHARI, A., HAIDER, S., ZAMORA, J., PROCTOR, G., SPUDICH, G., ALMEIDA-KING, J., STAINES, D., DERWENT, P., KERHORNOU, A., KERSEY, P. & FLICEK, P. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)*, 2011, bar030.
- KIU, R., BROWN, J., BEDWELL, H., LECLAIRE, C., CAIM, S., PICKARD, D., DOUGAN, G., DIXON, R. A. & HALL, L. J. 2019. Genomic analysis on broiler-associated *Clostridium perfringens* strains and exploratory caecal microbiome investigation reveals key factors linked to poultry necrotic enteritis. *Animal Microbiome*, 1, 12.
- KOPYLOVA, E., NOE, L. & TOUZET, H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28, 3211-7.
- MI, H., MURUGANUJAN, A., EBERT, D., HUANG, X. & THOMAS, P. D. 2019a. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res*, 47, D419-D426.
- MI, H., MURUGANUJAN, A., HUANG, X., EBERT, D., MILLS, C., GUO, X. & THOMAS, P. D. 2019b. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc*, 14, 703-721.
- MI, H. & THOMAS, P. 2009. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol*, 563, 123-40.
- MUDUNURI, U., CHE, A., YI, M. & STEPHENS, R. M. 2009. bioDBnet: the biological database network. *Bioinformatics*, 25, 555-6.
- PIMENTEL, H., BRAY, N. L., PUENTE, S., MELSTED, P. & PACTHER, L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods*, 14, 687-690.
- QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN, J., SCHWEER, T., YARZA, P., PEPLIES, J. & GLOCKNER, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*, 41, D590-6.
- R DEVELOPMENT CORE TEAM 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
- SEGATA, N., IZARD, J., WALDRON, L., GEVERS, D., MIROPOLSKY, L., GARRETT, W. S. & HUTTENHOWER, C. 2011. Metagenomic biomarker discovery and explanation. *Genome Biol*, 12, R60.
- SHANNON, P., MARKIEL, A., OZIER, O., BALIGA, N. S., WANG, J. T., RAMAGE, D., AMIN, N., SCHWIKOWSKI, B. & IDEKER, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13, 2498-504.
- TANG, Y., HORIKOSHI, M. & LI, W. X. 2016. ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. *The R Journal*, 8.
- TUREI, D., KORCSMAROS, T. & SAEZ-RODRIGUEZ, J. 2016. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods*, 13, 966-967.
- WARNES, G. R., BOLKER, B., BONEBAKKER, L., GENTLEMAN, R., HUBER, W., LIAW, A., LUMLEY, T., MAECHLER, M., MAGNUSSON, A., MOELLER, S., SCHWARTZ, M. & VENABLES, B. 2016. ggplots: Various R Programming Tools for Plotting Data. R package version 3.0.1 ed.
- WICKHAM, H. 2016. ggplot2: *Elegant Graphics for Data Analysis*, New York, Springer-Verlag.
- ZERBINO, D. R., ACHUTHAN, P., AKANNI, W., AMODE, M. R., BARRELL, D., BHAI, J., BILLIS, K., CUMMINS, C., GALL, A., GIRON, C. G., GIL, L., GORDON, L., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., TO, J. K., LAIRD, M. R., LAVIDAS, I., LIU, Z., LOVELAND, J. E., MAUREL, T., MCLAREN, W., MOORE, B., MUDGE, J., MURPHY, D. N., NEWMAN, V., NUHN, M., OGEH, D., ONG, C. K., PARKER, A., PATRICIO, M., RIAT, H. S., SCHUILENBURG, H., SHEPPARD, D., SPARROW, H., TAYLOR, K., THORMANN, A., VULLO, A., WALT, B., ZADISSA, A., FRANKISH, A., HUNT, S. E., KOSTADIMA, M., LANGRIDGE, N., MARTIN, F. J., MUFFATO, M., PERRY, E., RUFFIER, M., STAINES, D. M., TREVANION, S. J., AKEN, B. L., CUNNINGHAM, F., YATES, A. & FLICEK, P. 2018. Ensembl 2018. *Nucleic Acids Res*, 46, D754-D761.