MANCHESTER
1824
The University of Manchester

The gut microbiota in homeostasis and inflammation

2018

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health

Gurdeep Singh

School of Biological Sciences, Division of Infection, Immunity and Respiratory Medicine

Contents

List of Figures and Supplementary Figures ... 8

List of Tables ... 10

Abstract ... 11

Declaration ... 12

Copyright Statement ... 13

Acknowledgements ... 14

List of Abbreviations ... 15

1.1 Overview ... 19

1.2 Gut Barrier Function ... 20

1.2.1 Eosinophils in gut homeostasis ... 24

1.3 The gastrointestinal tract and the resident microbiota ... 25

1.4 The microbiota in immune function ... 27

1.5 Acquisition of the microbiota ... 29

1.5.1 Initial acquisition ... 29

1.5.2 Shaping the infant microbiome ... 29

1.6. Development of the gut microbiota ... 31

1.6.1 Shaping the adult gut microbiota ... 31

1.6.1.1 Age ... 32

1.6.1.2 Sex ... 33

1.6.1.3 Species ... 34

1.6.1.4 Antibiotics ... 35

1.7 The Microbiota and Metabolites ... 36

1.7.1 SCFAs ... 37

1.7.2 Bile acids ... 38

1.7.3 Amino acids and tryptophan-derived metabolites ... 38

1.7.4 Analysis of metabolites ... 39

1.8 Inflammatory Bowel Disease...40
1.8.1 The *mdr1a*$^{-/-}$ Mouse Model...41
1.8.2 The Microbiota in IBD...42
1.8.3 Metabolites in IBD...43
1.9 Analysing the microbiome...44
1.9.1 Sample collection...44
1.9.2 Sample extraction...45
1.9.3 Sequencing-independent analysis...45
1.9.4 Sequencing...45
1.9.5 OTU's...47
1.9.6 Diversity Analysis...47
1.9.7 Visualising data...48
1.9.8 Decision Trees and Random Forests...49
1.9.9 Inferring function...52
1.10 Hypothesis, aims and objectives...53
1.11 References...56
2. Gut microbial taxa at distinct phylogenetic scales distinguish host age, social group and microbial niche, but not genotype, in a murine model of inflammatory bowel disease...82
2.1 Abstract...83
2.2 Introduction...84
2.3 Materials and Methods...87
2.3.1 Animal Maintenance...87
2.3.2 Isolation of genomic DNA...87
2.3.3 Histology...87
2.3.4 Fluoresence *in-situ* hybridisation (FISH)...88
2.3.5 16S rRNA gene sequencing processing...88
2.3.6 Phylogenetic Tree...89

2.3.7 Ordination........89

2.3.8 Machine learning........89

2.3.9 Validation of Model........89

2.3.10 Statistical Analysis........90

2.4 Results........91

2.4.1 Phylogenetic tree of 16S rRNA data derived from the gut microbiota........91

2.4.2 Strong separation of the gut microbiota by microbial niche, age and cage, but not host genotype........93

2.4.3 The gut microbiota is strongly associated with age and microbial niche........95

2.4.4 Abundant, low-level taxa distinguish cage microbiomes but not age or niche........98

2.5 Discussion........101

2.6 References........105

2.7 Supplementary Information........111

3. Mucus metabolite and transcriptional changes precede the onset of colitis-induced inflammation........117

3.1 Abstract........118

3.2 Introduction........119

3.3 Materials and Methods........121

3.3.1 Animal Maintenance........121

3.3.2 Collection of tissue........121

3.3.3 Histology and Staining........121

3.3.4 Albumin ELISA........123

3.3.5 RNA-seq........123

3.3.6 Metabolite profiling using LC-MS........123

3.3.7 Statistical Analysis........124

3.4 Results........126

3.4.1 No evidence of gut inflammation in *mdr1a*$^{-/-}$ mice........126

3.4.2 Alterations to angiogenesis and permeability-associated transcription in *mdr1a*$^{-/-}$ mice 128

3.4.3 Difference in mucus-resident metabolites at 6 weeks of age in *mdr1a*$^{-/-}$ mice 130

3.5 Discussion 134

3.6 References 137

3.7 Supplementary Information 141

4. Gut eosinophils and their impact on the mucus-dwelling microbiota 146

4.1 Abstract 147

4.2 Introduction 148

4.3 Materials and Methods 150

4.3.1 Animal Maintenance 150

4.3.2 Sample preparation 150

4.3.3 Histology and Staining 150

4.3.4 Fluorescence *in-situ* Hybridisation (FISH) 151

4.4.5 Flow Cytometry for IgA Plasma Cells 151

4.5.6 IgA ELISA 152

4.5.7 16S DNA extraction, amplification and purification 152

4.5.8 Denaturing gradient gel electrophoresis (DGGE) 152

4.5.9 qPCR Analysis of Bacteria 153

4.5.10 Statistical Analysis 153

4.4 Results 155

4.4.1 Absence of eosinophils does not affect colonic or small intestinal morphology under homeostasis in younger mice 155

4.4.2 Mucus barrier integrity 156

4.4.3 Eosinophils influence microbial diversity 160

4.5 Discussion 164

4.6 Supplementary Information 171

5. Summary Discussion 182

5.1 Summary of Findings 182
5.2 What is the impact of an altered mucus-microbiome on host-health? 184
5.3 How important are eosinophils for regulation of gut homeostasis? 188
5.4 How could changes in the microbiota drive inflammation and colitis? 193
5.5 Are current methods of microbiome analysis adequate? 198
5.6 Using littermate controls: Why is cage effect so important? 201
5.7 Conclusions 203
5.8 References 204
6. Supplementary Materials and Methods 215
6.1 Animal Maintenance 215
6.1.1 Genotyping of *ΔdblGATA-1*$^{-/-}$ mice and *mdr1a*$^{-/-}$ mice 215
6.2 Histology 218
6.2.1 Tissue Processing 218
6.2.2 Haematoxylin and Eosin (H+E) Staining 218
6.2.3 Goblet Cell Staining 219
6.2.4 Colitis Scoring 219
6.2.5 Fluorescence *in-situ* Hybridisation (FISH) 221
6.3 Enzyme-linked immunosorbent Assays (ELISAs) 222
6.3.1 IgA ELISA 222
6.3.2 Albumin ELISA 222
6.4 Analysis of Microbial Communities via DGGE 223
6.4.1 16S DNA extraction, amplification and purification 223
6.4.2 Denaturing gradient gel electrophoresis (DGGE) 225
6.4.3 qPCR Analysis of Bacteria 226
6.5 Analysis of Microbial Communities via 16S rRNA Sequencing 227
6.5.1 16S rRNA gene sequencing processing 227
6.5.2 Phylogenetic Tree 229

6.5.3 Ordination....229
6.5.4 Machine Learning....229
6.5.5 Validation of the Model....230
6.6 Metabolite Profiling Using LC-MS....230
6.7 RNA-seq....231
6.8 Statistical Analysis....232
6.9 References....233
7. Appendices....236
7.1 Appendix 1....236
7.2 Appendix 2....237
7.2.1 List of Author Publications....237
7.2.2 Submission Pending....237

Word Count: 72480

List of Figures and Supplementary Figures

Chapter 1

Figure 1: Schematic illustration of differences in the gastrointestinal tract, between the small and large intestine. 21

Figure 2: Schematic representation of a decision tree. 50

Chapter 2

Figure 1: Distinct localisation of stool and mucus microbiota. 92

Figure 2: Significant separation of microbiota via NMDS for age and microbial niche. 94

Figure 3: Random forest model identifies strong associations between the microbiota, niche, age and cage. 97

Figure 4: Abundant, low-level taxa distinguish cage microbiomes but not age or niche. 98

Figure 5: Clade importance between forests. 99

Supplementary Figure 1: Major phyla are represented on the phylogenetic tree. 111

Supplementary Figure 2: Wide distribution of sequences on a phylogenetic tree when coloured by treatment groups. 112

Supplementary Figure 3: Redistributing the relative abundance of clade Erysipelotrichaceae confers importance to genotype. 113

Supplementary Figure 4: Assessing Robustness of the RF model. 114

Chapter 3

Figure 1: Gut barrier integrity in *mdr1a*$^{-/-}$ mice. 127

Figure 2: Transcriptional differences between WT and *mdr1a*$^{-/-}$ mice. 129

Figure 3: LC-MS (Positive Ionisation Mode) reveals variation in mucus-residing metabolites. 132

Figure 4: LC-MS (Negative Ionisation Mode) reveals variation in mucus-residing metabolites. 133

Supplementary Figure 1: Random Forest highlights important metabolites. 141

Chapter 4

Figure 1: Loss of eosinophils did not impact gut architecture. 156

Figure 2: Mucus barrier integrity in *ΔdblGATA-1*$^{-/-}$ mice. 158

Figure 3: Loss of eosinophils does not impact upon IgA 159

Figure 4: Absence of eosinophils leads to a significantly altered microbiota. 162

Figure 5: Bacterial communities significantly different between microbial niches. 163
Supplementary Figure 1: Loss of eosinophils leads to altered gut morphology in old Δ*dblGATA-1*$^{-/-}$ mice...171
Supplementary Figure 2: Loss of eosinophils did not impact on morphological differences in small intestine structure. 172
Supplementary Figure 3: Lack of eosinophils does not impact upon small intestinal goblet cells.. 173
Supplementary Figure 4: Loss of eosinophils does not impact upon goblet cell number in young or old mice. 174
Supplementary Figure 5: Trend towards increased small intestinal goblet cells in Δ*dblGATA-1*$^{-/-}$ mice..175
Supplementary Figure 6: Inner mucus layer characterisation...176
Supplementary Figure 7: Loss of eosinophils leads to increased serum IgA in younger mice 177
Supplementary Figure 8: Differences in the bacterial communities and diversity in the stool of Heterozygous (Het) and eosinophil deficient (Δ*dblGATA-1*$^{-/-}$) mice. 178
Supplementary Figure 9: Differences in the bacterial communities and diversity in the colonic mucus of Heterozygous (Het) and eosinophil deficient (Δ*dblGATA-1*$^{-/-}$) mice....................179
Supplementary Figure 10**:** Expression of Enterrococcaceae and SFB in WT and *ΔdblGATA-1*$^{-/-}$ mice……………………………………………………………………………………180

Chapter 7

Figure 1: The microbiota of an individual can be linked back to its mother, irrespective of cage. 237

List of Tables

Chapter 1

Table 1: Example dataset used to inform a decision tree..51

Chapter 3

Table 1: Histological scoring system for mucosal inflammation 122

Chapter 4

Table 1: List of primers used 154

Chapter 6

Table 1: List of primers used 219

Table 2: Histological scoring system for mucosal inflammation 222

Table 3: Primers used to amplify bacteria for denaturing gel electrophoresis (DGGE)........226

Table 4: List of primers used for qPCR 228

Table 5: Primers used for next generation sequencing 230

Abstract

The gut has a substantial resident microbiota localised in two major niches: the lumen and the intestinal mucus layer. The microbiota is vital for host health, yet the gut must also prevent uncontrolled incursion by pathogens or commensal microbes. Integral to host-microbe interaction is a gut barrier; comprising the intestinal epithelial cells, mucus and antimicrobials that regulate microbial entry, composition and the transit of metabolites and other molecules. Disruption of the gut barrier and the microbiota are associated with inflammation such as the autoimmune disorder, inflammatory bowel disease. I hypothesised that host-health is associated with characteristic changes in the gut microbiota, with the mucus-resident bacteria having the most profound impact on the host, due to their closer proximity to host cells.

Common techniques to probe microbiome datasets are based purely on subjective analysis and biases. In this thesis, I developed a novel method for exploring microbiome data. By constructing a phylogenetic tree of 16S rRNA sequence data derived from the stools and mucus of wildtype mice, and those that develop spontaneous colitis (*mdr1a*$^{-/-}$ mice), I used a random forest model on phylogenetically-defined clades. I found that the gut microbiota could be used to distinguish our treatment groups, such as mouse age, and identified the microbial characteristics that facilitated these associations. Hence, this method can be used to provide informative information about the microbiota and its associations with other conditions of interest.

To further explore the importance of the mucus microbiome, I then sought to define the contribution of mucus-resident metabolites to host function. I explored the mucus metabolite profile in the *mdr1a*$^{-/-}$ mouse model. I show that although there were no overall differences in the metabolite profile, there was variability in individual metabolites between wildtype and *mdr1a*$^{-/-}$ mice. These differences were also concordant with significant intestinal transcriptional changes. These data would suggest that changes to the mucus microbiota coincide with metabolomic and transcriptional differences in *mdr1a*$^{-/-}$ mice that predispose them to colitis.

My data highlighted the importance of microbial niche in colitis. To further explore the mucus microbiota and its interaction with the host, I investigated the microbiome in eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) mice that had altered immune and barrier function. My data showed that there was a significant reduction in bacterial diversity in the mucus that was not seen in stool samples in Δ*dblGATA-1*$^{-/-}$ mice. Although I saw overall differences in the microbiota of mice that lack eosinophils, a focused qPCR panel revealed that the biggest differences in the microbiota lay between different microbial niches, i.e. stool, colonic and small intestinal mucus.

Collectively, my studies confirm that a focus on the stool microbiome alone is insufficient to capture the diversity of the gut microbiome. As a result, it is vital to explore all niches within the gut, wherever possible, in order to gain a comprehensive insight into the role of the gut microbiome in host health.

Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made

iii. The ownership of certain copyright, patents, designs, trademarks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420), in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University’s policy on Presentation of Theses.

Acknowledgements

Firstly, I want to thank my supervisors for their incredible support. Sheena has been amazing throughout my PhD and I feel very fortunate to have been a member of her lab. Chris has been pivotal throughout many aspects of my PhD and without his support, my coding never would have developed. Andy has been an excellent source of ideas and the fear of God that he instils when he turns up in the write up area has motivated me to work harder. I am grateful to have had such a fantastic mix of supervisors. Many thanks to Professor Andrew McBain for being my advisor throughout my PhD.

I would also like to acknowledge all the facilities that I have used: The Biological Sciences Facility, the Flow Cytometry facility, the Genomics Facility and the Histology Facility. Additionally, I am grateful for the help of my collaborators: Professor Roy Goodacre, Dr Katherine Hollywood, Dr Drupad Trivedi and Dr Yun Xu. Additionally, I acknowledge the funding that I have received as part of the BBSRC Doctoral Training Programme.

I also want to thank my colleagues. From Chris's Lab: Huw, Rok, Danna, Adriana and Sam have been great and we have had many entertaining lab meetings and pub outings together. Danna was my go to person for coding support when Chris was away.

From Jo Pennock's lab: Jeya has been a great person to have around and is often unintentionally hilarious. Charis has been an excellent source of support in the midst of writing up and is super cool. Helen W was my mentor during my Master's project and I'm extremely grateful for everything she has taught me. I would also like to thank Dr Jo Pennock for her support in the Green Impact team and kindly agreeing to be my internal examiner!

From Kathryn Else's lab: I would like to thank Ruth for her extensive help with culls, general advice and always knowing what to do in any given situation. Larisa has provided much technical advice throughout my PhD for which I am grateful for. I would also like to thank Hannah for being a supportive friend and introducing me to Overwatch. Dr Iris Mair has also been super helpful and supportive. I would also like to thank Professor Kathryn Else who has provided invaluable support throughout my PhD.

From Sheena's lab, I would start by thanking the past members. Szu-Wei for being an endless source of amusement, Michael for being an extremely supportive and helpful individual, Tarfa for being an ally in my second year and Becky for being a last minute saviour when all other options had failed.

For the current members: Sheila has been an uplifting and highly entertaining individual and has made being in the lab much more fun. Rachel, whom I have known the longest, has taught me a great deal experimentally and is a fantastic scientist. Jordie, in addition to possessing a master level flow cytometry skill and extensive Simpsons knowledge, is a great addition to the lab and great to have around. James, who is my best dude, made my final two years much more awesome and I appreciate all the pub trips and video gaming sessions that we have had. I am grateful to have worked with you all. *Well done* everyone :P

Finally, I would like to thank my family for their support throughout my many years in education and for putting up with my occasional grumpiness! I am eternally grateful for their support. Special thanks to my sister who bought a rug for herself, but thought it matched my room better and so gave it to me.

And many thanks to you, person reading this. Yes, **you**.

List of Abbreviations

AMP	Anti-microbial peptides
Angpt	Angiopoietin
AQP	Aquaporin
C57BL/6	C57 black 6
CD	Crohn's Disease
DAPI	4',6-diamidino-2-phenylindole
DGGE	Denaturing gradient gel electrophoresis
dH2O	Distilled water
diH2O	Deionised water
DSS	Dextran Sodium Sulphate
DTT	Dithiothreitol
ECP	Eosinophil cationic protein
EDN	Eosinophil-derived neurotoxin
ELISA	Enzyme-linked Immunosorbent Assay
EPX	Eosinophil peroxidase
FDR	False discovery rate
FISH	Fluorescent *in-situ* Hybridisation
FVB	Friend leukaemia virus B
GATA-1	GATA-binding factor 1
GI	Gastrointestinal
H&E	Haematoxylin and Eosin
Het	Heterozygous
HMDB	Human Metabolome Database
IBD	Inflammatory Bowel Disease
IEC	Intestinal epithelial cell
IL	Interleukin

IFN-γ	Interferon gamma
IgA	Immunoglobulin A
IMS	Industrial Methylated Spirits
KEGG	Kyoto Encyclopaedia of Genes and Genomes
KO	Knockout
LC-MS	Liquid-chromatography mass spectometry
MBP	Major Basic Protein
MDA	Mean Decrease Accuracy
Mdr1a	Multidrug resistant gene 1a
MUC2	Mucin-2
NLRP	NACHT, LRR and PYD domains-containing protein
NMDS	Non-metric multidimensional scaling
NOD2	Nucleotide Oligomerisation Domain 2
OTU	Operational Taxonomic Unit
PBS	Phosphate buffered saline
PCA	Principle Components Analysis
PCoA	Principle Coordinates Analysis
PERMANOVA	Permutational multivariate analysis of variance
PGF	Placental Growth Factor
Pgp	P-glycoprotein
PICRUSt	phylogenetic investigation of communities by reconstruction of unobserved states
PRR	Pattern recognition receptor
QIIME	Quantitative Insights into Microbial Ecology
qPCR	quantitative PCR
REG	Regenerating islet-derived protein
RF	Random forest

RT	Room temperature
SCFAs	Short-chained fatty acids
SFB	Segmented filamentous bacteria
SI	Small intestine
SPF	Specific pathogen-free
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TGF-β	Transforming growth factor beta
Th	T helper
TJ	Tight Junction
TLR	Toll-like receptor
TNF	Tumour necrosis factor
Treg	T Regulatory Cell
TRUC	*T-bet*$^{-/-}$ *x Rag2*$^{-/-}$ ulcerative colitis
UC	Ulcerative Colitis
VEGF	Vascular Endothelial Growth Factor
WT	Wildtype
ZO	Zonula occludens
γΔ T cell	Gamma delta T cell

Chapter One
Introduction

1.1 Overview

The human body is host to a vast number of microbial species collectively termed the microbiota. The term "microbiota" itself was first described by Lederberg and McCray (2001) and refers to all microbial species within the host, encompassing bacteria (bacteriome), fungi (mycobiome) and viruses (virome). However, the focus in this thesis will be on bacteria. The microbial burden varies depending upon anatomical location, with some of the key microbial sites being the mouth (Siqueira and Rocas, 2010), skin (Grice and Segre, 2011) and the gut which has the largest microbial burden at approximately 10^{13} bacteria (Sender et al., 2016). Due to their sheer number, it is reasonable to expect that the microbiota would play significant physiological roles within the host and many roles have been demonstrated experimentally, ranging from digestion of food substrates (Turnbaugh et al., 2006), drug metabolism (Yoo et al., 2014) and immune development (Bengmark, 2013). As a result, projects such as the Human Microbiome Project (HMP) conducted by the National Institute of Health, have aimed to characterise and understand this extensive microbial system (Peterson et al., 2009). Understanding what species inhabit the host and their various genetic/metabolic contributions has been a considerable area of research within recent years. As well as being beneficial, alterations in the microbial communities and/or altered immune reactivity to the gut microbiota has been implicated in inflammatory disease, particularly inflammatory bowel disease (IBD) (Li et al., 2014a; Glymenaki et al., 2017).

Inflammatory bowel disease (IBD) refers to a spectrum of chronic inflammatory disorders, including Crohn's disease and ulcerative colitis which are characterised by altered immune reactivity to the microbiota. Mouse models have commonly been used to investigate genetic changes that predispose to IBD which broadly comprise altered barrier permeability, altered recognition of bacterial components and reduced anti-inflammatory function. In addition, specific changes in the gut microbiota are associated with IBD (Kiesler et al., 2015). Much of the research focuses on analysis of microbes found within the stools. However, the microbiota varies along the length of the gastrointestinal (GI) tract and even within distinct niches such as the gut mucus layer versus the gut lumen. The microbes found in the gut mucus are within close proximity to our host cells and thus may have the biggest functional impact on the host, yet remain relatively uncharacterised. Furthermore, methods to analyse and characterise the microbiota are often purely correlative and biased by a focus on single taxonomic levels. One functional mechanism behind the gut microbiota and its ability to regulate host function could be via the production of metabolites. The microbiota produces

metabolites derived from the diet and a particular group of metabolites known as short-chained fatty acids (SCFAs) mediate anti-inflammatory effects on the host (Trompette et al., 2014). However, the majority of studies to date have focused on serum, urine and stool metabolites. Indeed, the mucus has been less well explored.

This thesis seeks to explore the hypothesis that host-health is associated with characteristic changes in gut bacterial communities, with the mucus-resident bacteria having the most profound impact on the host. To facilitate analysis of the microbiota, I developed a novel non-discriminatory method to find associations between the gut microbiota and various conditions of interest, such as microbial niche. I explored microbial diversity within the mucus of the small and large intestine versus stool and investigated the metabolite profile of the mucus. Overall the work showed that a focus on stool is insufficient to capture the full diversity of the gut microbiota.

In the introduction I will describe the organisational structure of the gastrointestinal (GI) tract and the immunological mechanisms that maintain gut homeostasis. I discuss the resident microbiota and its organisation in the GI tract. I then explain the acquisition and development of the gut microbiota, including strategies to analyse the gut microbiome, starting from sample collection to processing of the data. Following on from this, I discuss the importance of metabolites and the role of the microbiota in IBD. Finally, I detail the hypotheses, aims and objectives for this thesis.

1.2 Gut Barrier Function

The gastrointestinal tract (GI) is a major immunological site within the body that incorporates several components: the oesophagus, the stomach, the small intestine, the large intestine and rectum. The small intestine and large intestine are further subdivided into different regions. The small intestine comprises the duodenum, the jejunum and the ileum whereas the large intestine comprises the caecum, proximal, mid and distal colon. The microbiota are separated from the host by a ‘barrier’ of intestinal epithelial cells (IECs) and mucus. This gut barrier is vital to maintain healthy gut homeostasis. The IECs can act as a physical barrier that helps to maintain intestinal permeability- the ability of fluid and solutes to diffuse through the lumen and gut tissue. A more permeable gut is associated with inflammation (Collett et al., 2008) and therefore the ability to control the passage of fluid and small molecules is vital. There are two key routes through which products can be transported across the epithelium:

transcellularly (i.e. through cells, mediated by transporters) or paracellularly (between cells). Paracellular permeability is regulated by a complex of proteins, known as tight junction (TJ) proteins that interface with epithelial cells to form intercellular barriers. Key TJ proteins include junctional adhesion molecules (JAM), occludin, claudin and ZO-1 (Lee, 2015). Claudin comprises a family of proteins that act as the vital backbone for the TJ, where mice lacking claudin-1 expression died within 24 hours of birth due to dramatic fluid loss (Lee, 2015). TJ proteins are regulated by a number of factors, and immune cells are important including a subset of γδ intraepithelial lymphocytes (IELs) (Dalton et al., 2006). Eosinophils are specialised immune cells that have also been linked to epithelial integrity although the mechanism by which they do this has not been fully elucidated (Johnson et al., 2015). The role of eosinophils in the gut will be discussed in more detail in **Section 1.2.1** Fluid permeability is also maintained by a group of membrane proteins known as aquaporins (AQP's) (Ishibashi et al., 2009). Hence, epithelial cells play a vital role in controlling intestinal permeability.

The IECs are a heterogeneous population of cells, that can comprise goblet cells that secrete mucus, enteroendocrine cells involved in signalling and absorption, sensory tuft cells and paneth cells (van der Flier and Clevers, 2009; Tschurtschenthaler et al., 2014) (**Figure 1**). Goblet cells in the intestine secret the mucus protein MUC2 (Chang et al., 1994) and maintain the integrity of the mucus barrier (Johansson et al., 2011). Production of mucus is thought to be dependent upon the formation of an immune complex known as the NLRP6 inflammasome, where mice deficient in NLRP6 fail to clear *Citrobacter rodentium* infection (Wlodarska et al., 2014), which was linked to the impaired secretion of mucus by goblet cells. Hence, goblet cells are thought to be a key player in mucosal immunity (Johansson and Hansson, 2014; Knoop and Newberry, 2018). Enteroendocrine cells are specialised epithelial cells that release a variety of compounds in response to luminal signals. For example, they secrete the satiation hormone ghrelin, serotonin and gastrin (Sternini et al., 2008). Curiously, they are thought to be involved in the production of 'taste-signalling molecules' that sample the local mucosa for nutrients and potentially harmful compounds, to either prepare for absorption or a protective response (Sternini et al., 2008). Similarly, tuft cells are chemo-sensory epithelial cells that comprise a small fraction of the small intestinal epithelium and have been shown to contribute towards the type 2 immune response in parasite infections (Gerbe et al., 2011; Howitt et al., 2016).

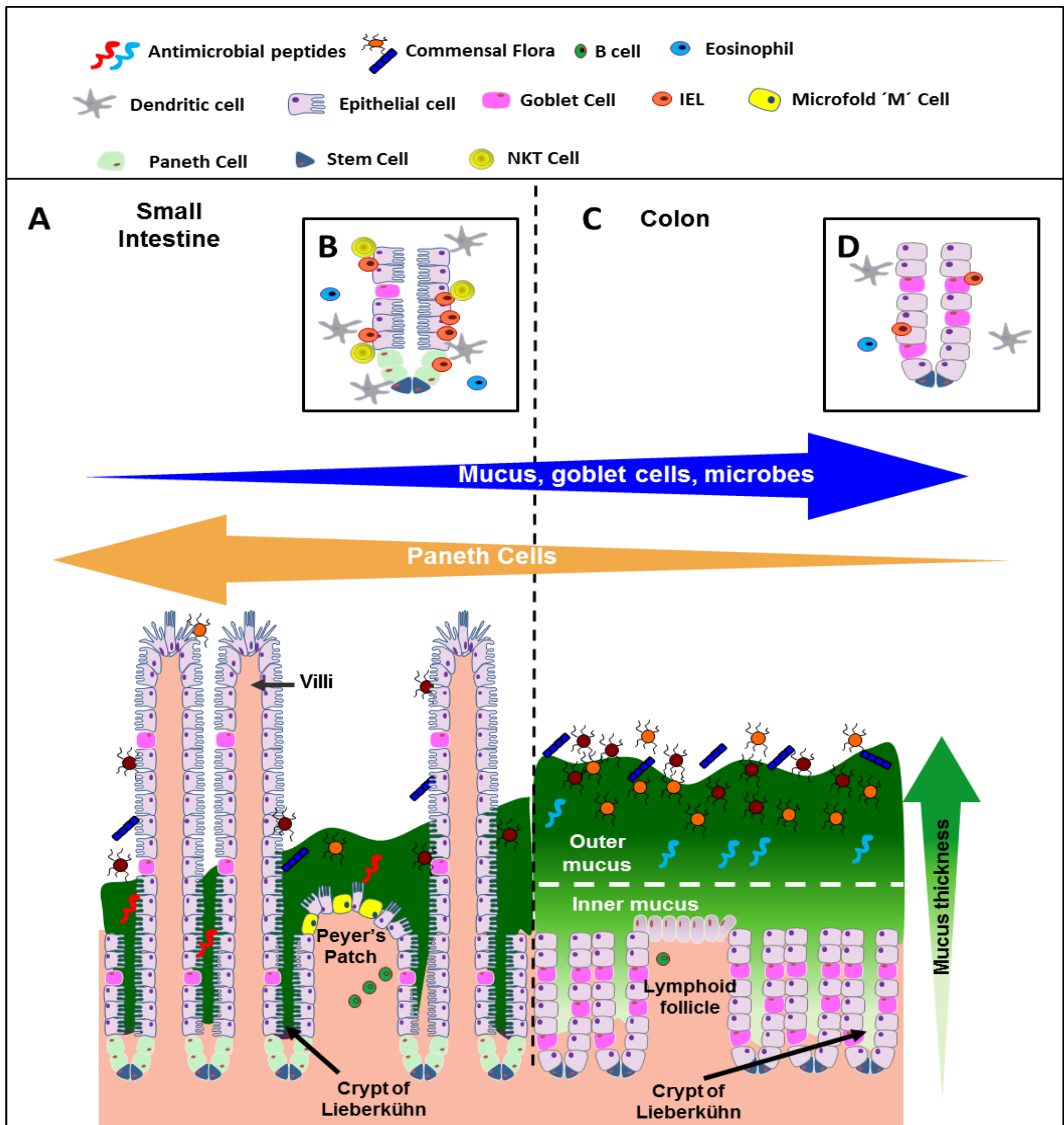


Figure 1: Schematic illustration of differences in the gastrointestinal tract, between the small and large intestine. (A) The small intestinal (SI) tissue folds giving rise to Crypts of Lieberkühn and projection-like villi. The SI also contains Peyer's patches, where resident microfold (M) cells transcytose luminal antigens. The SI immune population under homeostasis is also illustrated (B). (C) The colon folds giving rise to Crypts of Lieberkühn. It also possesses increased mucus thickness, goblet cells and bacterial numbers. The colonic immune population under homeostasis is illustrated (D). IEL: Intraepithelial lymphocyte; NKT: Natural Killer Cell. **Adapted from: Bowcutt et al. (2014).**

Collectively, the epithelial cells produce a slew of regulatory signalling molecules. Of relevance to the microbiota, there are a group of antimicrobial peptides known as defensins and cathelicidins that kill bacteria. In the small intestine, Paneth cells in the base of the villus, known as the crypt of Lieberkühn, contain inactive α-defensins that are cleaved into an active state by Paneth cell-derived trypsin (Ghosh et al., 2002; Nakamura et al., 2016). However, β-defensins, produced by Paneth cells, neutrophils and macrophages, also comprise a major arm of the defensins repertoire (Donnarumma et al., 2016). β-defensins have been shown to mediate a variety of effects, such as dampening the pathology of *Salmonella* infection in the gut (Fusco et al., 2017) and influencing wound repair (Williams et al., 2018). Additionally, there are a host of other antimicrobial factors such as trefoil peptides, a unique family of protease-resistant peptides that help to support intestinal epithelial barrier function (Kindon et al., 1995). There are also antimicrobial lectins, such as the C-type lectin RegIIIγ. RegIIIγ has been implicated in creating a bacteria-free region adjacent to the surface of the epithelial cells (Cash et al., 2006; Vaishnava et al., 2011).

The mucus itself also contributes to the barrier function. There is a wide array of literature discussing the protective properties of the mucus lining (Miller and Jarrett, 1971; Peterson and Artis, 2014). It acts as both a physical barrier and a biologically active barrier against commensal and pathogenic bacteria. The fundamental organisation of the mucus changes from the stomach to the colon. The small intestine has a discontinuous mucus layer and is less dense as opposed to the large intestine which is more substantial, comprising two mucus layers: a loose outer layer inhabited by various microbial species and an inner, dense 'sterile' layer that is firmly attached to the epithelial cells (Johansson et al., 2008; Johansson et al., 2011). The physical properties of this inner mucus layer prevents bacteria from breaching it. However, bacteria are also controlled by secretory immunoglobulin (Ig)A produced by plasma cells in the lamina propria. IgA is transported across the epithelium and into the mucus layer which aids in antimicrobial activity (Macpherson et al., 2008; Macpherson et al., 2018). Curiously, deficiency in IgA does not cause dramatic microbial perturbation, but is thought to impact on localisation of bacteria, for example, oropharynx bacteria colonising the gut which does not happen under normal conditions (Fadlallah et al., 2018). IgA adhesion to bacteria is also thought to be involved in positively or negatively selecting for bacteria, helping to shape the mucus-resident community (McLoughlin et al., 2016). These findings were confirmed, where IgA binding to the mucus-resident species *Bacteroides thetaiotaomicron* modulated its gene expression and function *in vivo* (Nakajima et al., 2018).

B. thetaiotaomicron-secreted molecules induced by the IgA-binding were thought to be vital for promoting healthy bacterial competition and protection against chemically-induced colitis. Binding of high affinity IgA to bacteria has also been shown to select for colitis-inducing bacteria (Palm et al., 2014). Therefore, IgA has vital roles beyond its antimicrobial function and so it is important that mechanisms are in place to regulate IgA-secreting plasma cells. It was previously shown that an immune cell type, the eosinophil, may play a role in the regulation of IgA-secreting plasma cells (Chu et al., 2014).

1.2.1 Eosinophils in gut homeostasis

Eosinophils are granule-containing lymphocytes and can microscopically be identified by their bi-lobed nucleus. Eosinophil-derived granules can mediate a variety of functions, the majority of these involving a cytotoxic effect (Acharya and Ackerman, 2014). These granules include major basic protein-1 and -2 (MBP-1 and MBP-2), eosinophil peroxidase (EPX), eosinophil cationic protein (ECP) and eosinophil-derived neurotoxin (EDN) (Acharya and Ackerman, 2014). Historically, eosinophils were implicated in allergy, allergic airway disease and asthma (Ramirez et al., 2018). Additionally, they contribute to anti-helminth immunity, although they have been shown to have variable effects on infection depending on the infection model (Klion and Nutman, 2004). Eosinophils have been implicated in remodelling and wound repair (Leitch et al., 2009) and the regulation of gut homeostasis. Specifically, eosinophils have been shown to contribute to the maintenance of intestinal permeability (Johnson et al., 2015). Mice fed a high fat diet were depleted of eosinophils in the gastrointestinal tract and developed increased intestinal permeability (Johnson et al., 2015) whereas obese mice with normal eosinophil levels had no changes in gut permeability. The mechanism behind the reduced eosinophil numbers was not fully elucidated, although Johnson et al. (2015) stipulated that it could be a consequence of a defect to eosinophil tracking to the gut. Interestingly, other studies have also reported a loss of eosinophils in obesity (Withers et al., 2017). In line with maintaining permeability, eosinophils may also play an important anti-inflammatory role. Indeed, eosinophils residing in white adipose tissue were necessary for polarising macrophages from a naïve to an anti-inflammatory state (Wu et al., 2011a). The lack of eosinophils led to a loss of glucose tolerance that was only restored upon restoration of eosinophil numbers.

Another aspect of gut barrier integrity that eosinophils may be involved in is the regulation of IgA-secreting plasma cells (Chu et al., 2014; Forman et al., 2016). However, studies have shown conflicting information with regards to the IgA effect. One study showed a decrease in

IgA levels under naïve conditions (Chu et al., 2014), whereas another study showed a decrease in the small intestine and only in the context of infection (Forman et al., 2016) and one study found no associations between eosinophils and IgA-secreting plasma cells (Haberland et al., 2018)- although only spleen and bone marrow were investigated. It is thus likely that the impact on eosinophils on IgA varies dependent on site and perturbations such as infection. The use of littermate controls and the strain of mice may also be a factor in inconsistencies in this literature (Bramhall et al., 2015).

A variety of mouse models have been used to start to dissect the contribution of eosinophils to host health as reviewed by Jacobsen et al. (2014). Previous studies used *IL-5*$^{-/-}$ mice as a proxy to indicate the impact of eosinophils on host function (Kopf et al., 1996) as IL-5 is an important cytokine that promotes the generation of eosinophils. However, this transgenic mouse does not specifically deplete eosinophils and will have other effects. The most commonly used eosinophil-deficient mouse models are PHIL mice (Lee et al., 2004) and Δ*dblGATA-1*$^{-/-}$ mice (Shivdasani et al., 1997), the latter of which have been used in this thesis. Δ*dblGATA-1*$^{-/-}$ mice have a specific mutation to the transcription factor GATA-1, which is vital for eosinophil development and maturation. Therefore, eosinophils are specifically ablated within these mice. Given the importance of eosinophils in gut homeostasis and their potential role in maintaining IgA producing plasma cells, it is likely that eosinophils play a significant role in the regulation of the gut microbiota. However, there has been no littermate-controlled study that has investigated the microbiota in eosinophil-deficient mice.

1.3 The gastrointestinal tract and the resident microbiota

The abundance and makeup of the microbiota varies and also increases down the length of the GI tract, with the largest concentration of bacteria being found within the large intestine (Sender et al., 2016). Although previous estimates have suggested that bacteria outnumber human cells 10-fold (Gill et al., 2006), more recent analysis suggests that bacterial numbers are approximately equal to our own cells (Sender et al., 2016).

Harsh physiological conditions within the oesophagus and the stomach prevent a diverse array of bacteria from inhabiting these locations. The oesophagus has a relatively small microbial concentration of 10 cells per g/ml (Di Pilato et al., 2016) and bacterial diversity is limited. The majority of bacteria found throughout the length of the oesophagus include

Streptococcus, Prevotella and *Veillonella* species (Gagliardi et al., 1998; Pei et al., 2004; Fillon et al., 2012). The stomach is similarly limited in terms of the resident bacteria, with concentrations ranging between 10^2-10^4 colony-forming units (CFUs)/ml (Delgado et al., 2013). The stomach is a particularly interesting environment, due to fluctuations in the levels of gastric acid that can influence pH, an environmental factor known to have a substantial effect on the gut microbiota (Jackson et al., 2016a). Major phyla within the stomach include *Firmicutes, Proteobacteria, Actinobacteria* and *Fusobacteria* (Bik et al., 2006). However, the presence of the potentially pathogenic bacterium, *Helicobacter pylori* can influence the gastric microbiota to some degree, with reduced number of *Proteobacteria* in *H. pylori* positive patients (Bik et al., 2006). A large proportion of individuals are colonised with *H. pylori,* and for many it remains a commensal bacteria (Cariani et al., 1992). However, *H. pylori* has been linked with conditions such as gastritis, inflammation of the stomach lining and peptic ulcers (Watari et al., 2014) and gastric cancer (Rhee et al., 2014). *H. pylori* is a dominant member of the gastric microbiota in gastric cancer patients (Yu et al., 2017) and as such, has been targeted therapeutically to try and reduce this cancer incidence (Hollister et al., 2014; Moyat and Velin, 2014).

Although acidic conditions within the stomach limit bacterial diversity, the bacterial diversity increases further down the gastrointestinal tract. This is likely to coincide with a change in the flow of digestive fluids through the lumen and alkalinisation of the environmental pH (Fallingborg, 1999). The microbial concentration within the small intestine is thought to range between 10^2-10^7 cells per gram of tissue (El Aidy et al., 2015). In human patients, characterising the small intestinal microbiota remains a problematic issue as the site is inaccessible via routine methods and typically requires invasive surgical intervention (Wang et al., 2003; El Aidy et al., 2015). However, one paper was able to overcome this issue by focusing on illeostomists, patients that have had their complete colon removed and thus the illeal end of the small intestine is accessible via a stoma (Zoetendal et al., 2012). They identified high levels of *Streptococcus and Clostridium sp.* within the small intestine and these findings are consistent with other studies. *Veillonella sp.* are also a common bacteria found within the small intestine (Gevers et al., 2014).

The large intestine possesses the most numerous and diverse array of bacteria within the human body with an estimated 3.8 x 10^{13} bacteria in the large intestine (Sender et al., 2016). The large intestinal microbiota is dominated by key bacterial phyla that include *Firmicutes, Bacteroidetes* and to a lesser extent *Proteobacteria, Actinobacteria and Verrucomicrobia* (Li

et al., 2014b; Hugon et al., 2015). However, it is important to consider the numerous microbial niches within the gut.

There are different microbial niches within both the large and the small intestine. A large proportion of the gut microbiota is present in the lumen, which also facilitates the transport of food material through the GI tract (Bengmark, 2013). However, the microbiota can also reside in a mucus layer that is situated above the IECs. Typically, stool samples are used as a proxy for the microbiota and this cannot fully capture the diversity along the GI tract or the diversity across mucus and luminal resident bacteria. The microbiota inhabiting the mucus is distinct from the community inhabiting the lumen (Li et al., 2015). One study examined mucosal bacteria across different regions of the large intestine and found that mucus resident bacterial species were generally consistent throughout the length of the large intestine mucus but differed significantly from that of faeces (Zoetendal et al., 2002). Specifically, one study showed that mucus samples had a greater relative abundance of *Proteobacteria* and *Fusobacteria* than stool, although the stool had a greater proportion of *Firmicutes and Bacteroidetes* (Tang et al., 2015). Our previous work has reflected the increased proportion of *Proteobacteria*, although the mucus had a greater proportion of *Firmicutes* than stool (Glymenaki et al., 2017). At lower taxonomic levels, *Acinetobacter, Mucispirillum* and *Flavobacteria* have been identified as mucus-associated, whereas *Roseburia* and the families Veillonellaceae and S24-7 have been identified as stool-associated (Hollister et al., 2014; Tang et al., 2015; Glymenaki et al., 2017). Understanding these differences is important because changes in one population of bacteria could have functional consequences for the host, for example, the mucus but not the stool microbiota changing before the onset of colitis (Glymenaki et al., 2017). Similarly, the mucus microbiota of cirrhosis patients was significantly altered compared to healthy patients, whereas stool bacteria were not changed (Bajaj et al., 2012). Such studies suggest the importance of understanding the mucus microbiome to better study disease processes. However, the relationship between the bacteria in the mucus and the lumen has largely remained unexplored. Understanding these relationships is important, because focusing specifically on the faecal or mucus bacteria in isolation would not give a full insight into the role of the microbiome.

1.4 The microbiota in immune function

The GI tract is a major immunological site and thus there are mechanisms in place to regulate both the immune system and the gut microbiota. However, the gut microbiota itself can also

regulate immune development. Indeed, germ-free mice develop reduced numbers of T cell subsets, including T helper (Th) cells such as Th1 and Th17 cells (Ivanov et al., 2008; Spasova and Surh, 2014). Additionally, the microbiota play a role in the development of T-regulatory (T-reg) cells (Ostman et al., 2006). Specifically, species such as *Bacteroides fragilis* have been associated with T-reg induction via its polysaccharide A (PSA) capsule and therefore promote health (Round and Mazmanian, 2010). However, other bacteria can regulate immunity in more harmful ways. For example, segmented-filamentous bacteria (SFB) are thought to induce T-cells to become Th17 cells, a cell type with mostly inflammatory roles (Ivanov et al., 2009; Wu et al., 2010). Th17 cells produce an array of cytokines that include IL-17, IL-17F, IL-21 and IL-22. These cytokines mediate inflammatory and protective functions, such as antibacterial and antifungal immunity (Hayes et al., 2014). SFBs are found in a variety of vertebrate species including humans and mice (Davis and Savage, 1974; Klaasen et al., 1993) and are primarily localised to the ileum, where they have been shown to mediate resistance against pathogen invasion (Garland et al., 1982). This resistance mechanism could be due to increased IL-17 and IL-22 expression, as when SFB were transplanted into Th17-cell deficient mice, CD4+ T cells were stimulated to produce IL-17 and IL-22 (Ivanov et al., 2009). Given that these are typically inflammatory cytokines, this induction can also be harmful, driving phenotypes such as arthritis (Wu et al., 2010).

It has been speculated that the development of the immune system is primed by early microbial factors, such as the presence of specific bacteria or bacterial DNA in the placenta. Indeed, Satokari et al. (2009) showed that the presence of bacterial species within the placenta may prime the neonate's immune system much earlier than previously thought, via the activation of the innate immune receptor toll-like receptor 9 (TLR9). Additionally, it was shown that a lack of commensal bacteria impaired the development of peripheral lymphoid organs in germ-free animal models (Smith et al., 2007). As such, establishment of the microbiota is a key stage in healthy immune development. From the gestational period, the gut microbiota can also influence permeability of the blood brain barrier and development of the specialised macrophage population in the brain- the microglia (Braniste et al., 2014; Foster et al., 2017; Moya-Perez et al., 2017). Therefore, it is important to consider the acquisition of our microbiota when considering immune development and disease.

1.5 Acquisition of the microbiota

1.5.1 Initial acquisition

The time of initial acquisition, the point at which bacteria first colonise the infant, was a point of contention. Early research suggested that microbial acquisition only occurred after delivery of the infant and that the GI tract of newborns was sterile (Escherich, 1988; Mackie et al., 1999). However, there is evidence that indicates the presence of bacteria within the meconium (earliest faeces from a neonate), albeit in low numbers and with low diversity (Jimenez et al., 2008). It has been debated whether there was a placental microbiota, however it is now known that bacterial products can cross the placenta to influence the offspring's immune development (Blaser and Dominguez-Bello, 2016; Gomez de Aguero et al., 2016). Bacterial DNA from *Bifidobacteria* and *Lactobacilli* have been identified in the placenta (Satokari et al., 2009) and so there could be a transition of maternal microbiota to the neonate as suggested by Thum et al. (2012). It was recently shown that maternally-derived gut bacteria found within the infant gut were considerably more stable than those bacteria derived from other sources (Ferretti et al., 2018). Given the strong role of transmission of microbiota from mother to offspring, littermate controls are vital in microbiota studies in order to account for potential differences that may arise due to mice having different mothers (Scholz et al., 2014).

1.5.2 Shaping the infant microbiome

After birth the method of delivery can have an immediate impact upon the composition of the infant's microflora. A caesarean delivery usually leads to colonisation of the infant by the maternal skin microbiota, including species such as *Staphylococcus aureus* and other potentially harmful, non-skin resident species such as *Clostridium difficile* (Dominguez-Bello et al., 2010; Nicholson et al., 2012). In contrast, if delivered naturally, the infant is typically colonised with *Bifidobacterium* species (Mikami et al., 2009). Indeed, transmission of a variety of *Bifidobacterium sp.* are thought to be inherited from the mother's intestine if the baby is delivered naturally (Makino et al., 2013). Additionally, the infant acquires bacteria via the vagina including species such as *Lactobacilli and Prevotella* (Dominguez-Bello et al., 2010). Natural births are thought to lead to healthy establishment of the gut microbiota which in turn, is important development of innate and adaptive immunity (Macpherson and Harris, 2004). Compared to caesarean deliveries, natural births are thought to result in fewer negative health consequences associated with the gut microbiota in later life (Moya-Perez et al., 2017).

The consequences of 'aberrant' colonisation are thought to be numerous (Moya-Perez et al., 2017). For example, postnatal acquisition at an early developmental stage was shown to be important for neural regulation (Sudo et al., 2004) such that caesarean birth could lead to defects in neural regulation that would be associated with the gut microbiota. Another study found an association between the levels of *S. aureus* in the gut at the time of infancy and the development of childhood obesity (Kalliomaki et al., 2008). Due to the potentially negative consequences that may be associated with aberrant microbial colonisation via caesarean, there has been a trend towards swabbing infants with vaginal fluids to restore the microbiota. One study addressed the efficacy of this practice and although they found that the microbiota could be partially restored, the long-term consequences were unknown (Dominguez-Bello et al., 2016).

Numerous other factors can affect the development or acquisition of the infant microbiome. For example, some cases of premature birth require treatment within intensive care units and this environment can select for higher levels of *C. difficile*. However, Nicholson et al. (2014) argue that there is little difference between premature and full-term infants in acquiring *C. difficile* in hospital, unless hospitalisation is prolonged. Similarly, a recent study was unable to identify a strong association between antibiotic usage and perturbations to the microbiota in preterm infants (Dardas et al., 2014). However, they did note an increase in *Actinobacteria* and *Proteobacteria* in infant faecal samples.

In addition to effects of birth on microbial colonisation, infant nutrition is vital in shaping the microbiota. Breastmilk contains numerous beneficial fats and proteins that are vital for infant development and may protect against necrotising enterocolitis, a major cause of infant death (Autran et al., 2018). Breastmilk also contains a variety of bacteria that includes species such as *Bacteroides*, *Clostridium*, *Faecalibacteria*, *Roseburia*, and *Bifidobacteria* (Collado et al., 2009; Jost et al., 2013; Wang et al., 2017). It is thought that the levels of lactoferrin, an immunoprotective component found within breastmilk, correlates with increased numbers of *Bifidobacterium* species (Mastromarino et al., 2014). The process of breastfeeding plays a vital role in seeding the infant's gut with a 'healthy' microbiota and contributes to a positive outcome for infant health (Pannaraj et al., 2017). For example, the protective benefits of breastfeeding are thought to extend to reduced incidence of neonatal fever (Netzer-Tomkins et al., 2016) and reduced incidence of obesity and diabetes (Binns et al., 2016). Following on from breastfeeding, the gut microbiota becomes more stable one year after birth which often

coincides with the introduction of solid food (Nicholson et al., 2012). How the microbiota develops from this point onwards will now be discussed.

1.6. Development of the gut microbiota

1.6.1 Shaping the adult gut microbiota

Although the infant microbiota appears malleable to change, the microbiota becomes more stable and less susceptible to change relatively early in life (Toole and Claesson, 2010). Development of the gut microbiota from infancy is relatively consistent across humans (Yatsunenko et al., 2012), where once the microbiota has established, it is dominated by the two major phyla, *Firmicutes* and *Bacteroidetes,* (**Section 1.3**). *Firmicutes* are gram positive rod-shaped bacteria, primarily located within the mucus layer (Harrell et al., 2012). *Bacteroidetes* are gram negative rod-shaped bacteria. The microbiota is then shaped by diet, which is one of the largest factors influencing the microbial composition. The 'Westernised' diet is thought to have negative consequences for the gut microbiota and host health, due to the high levels of fat and low levels of fibre (Zinocker and Lindseth, 2018). A study investigated the difference between a calorie-restricted diet, where recommended intake for nutrients was met, but was 50% less calories compared to the westernised, 'American diet' and showed that the American diet led to significantly reduced microbial diversity (Griffin et al., 2017). Reduced microbial diversity is typically associated with ill health in the gut, such as inflammation and obesity (Ott et al., 2004; Manichanh et al., 2006). These conditions are thought to arise, in part, due to diet-induced changes in the gut microbiota (Zinocker and Lindseth, 2018). Specifically, a high fat diet can induce metabolic endotoxaemia associated with an increase in endotoxin-producing bacteria (Pendyala et al., 2012). For example, a mouse study revealed that a high fat diet resulted in increased bacteria from the class *Erysipelotrichi,* which led to altered metabolic activity (Turnbaugh et al., 2009). The high fat diet often coincides with a reduction in the availability of dietary fibre (Chassaing et al., 2015) which may be the key factor in having negative health consequences for the host (Makki et al., 2018). Depletion of fibre caused increased abundance of mucus-degrading bacteria, leading to a reduction in mucus thickness and inflammation (Desai et al., 2016). Indeed, in response to a fibre depleted diet, the microbiota used mucus glycoproteins as a nutrient source which led to degradation of the mucus barrier.

A well reported example of a high fibre vs fat diet refers to a hospital feeding study, where two cohorts of adult patients were fed either a high fat/low fibre diet or a low fat/high fibre diet (Wu et al., 2011b). The findings revealed that those on a high fat/low fibre diet possessed a greater proportion of *Bacteroides* whereas the high fibre cohort possessed a greater proportion of *Prevotella* species. Individuals within this study were between 18-40 years old, where the microbiota should be more stable (Jeffery et al., 2012) and notably the altered proportions of bacteria were only stable for the duration of the feeding study, leading to the suggestion that long-term dietary changes are required for a sustained alteration of the microbiota. The effect of long-term dietary alterations can be seen geographically, within culturally distinct populations. Children in Burkina Faso, who predominately have a plant based diet, had a greater proportion of *Prevotella* species in comparison to children in Italy who had a greater proportion of *Bacteroides* (De Filippo et al., 2010). Dietary choice was important in determining bacterial diversity, with the children in Burkina Faso having a greater diversity of microbial species, including *Xylanibacter.* Similarly, patients who regularly consumed fresh fruit and vegetables also had greater bacterial diversity (Claesson et al., 2012; Hollister et al., 2014). Thus it may be that a high intake of fruit and vegetables that are fibre rich has a positive influence for the gut microbiota. One study of a strict vegetarian diet found that *Clostridium* dominated the gut microbiota (Hayashi et al., 2002). Another study investigated the faecal microbiota comparing vegetarian and omnivorous diets in south Indian women (Kabeerdoss et al., 2012). Curiously, omnivores seemed to have greater numbers of bacteria from the Clostridium cluster XIVa, which are butyrate-producing bacteria (Graf et al., 2015). Although this could have potential health benefits for omnivores, Graf et al. (2015) discuss the fact that the diet study was conducted in south Indian women, whose omnivorous and vegetarian diet is likely to differ substantially from the westernised diet. Although diet is vital for shaping the microbiota, there are several periods within adult life that can cause microbial instability. These periods include puberty, old age and in women, pregnancy and menopause (Cabrera-Rubio et al., 2012; Konkel, 2013).

1.6.1.1 Age

Within the elderly, bacterial diversity decreases and there is a great deal of variability between elderly individuals (>65 years old) than those of a younger age (Claesson et al., 2011). One study explored the microbiota in patients on the extreme end of the age spectrum, in centenarians (Santoro et al., 2018). This study stipulated that centenarians represented an excellent model for exploring the aged gut microbiota, as these patients typically have no

chronic disorders that may skew microbial analysis. However, many elderly subjects experience reduced chewing ability, salivation and low gut motility that may reduce the nutritional load supplied to the gut microbiota (Lovat, 1996). Additionally, the elderly are more predisposed to immunosenescence and low grade chronic inflammation that can alter the gut environment in such a way as to lead to more aerobic conditions (Claesson et al., 2011; Santoro et al., 2018). Consequently, it has been shown that the ratio of *Firmicutes* to *Bacteroidetes* decreases in old age, with *Firmicutes* being dominant in younger subjects and *Bacteroidetes* prevalent in older subjects (Mariat et al., 2009). A change to more aerobic conditions is thought to lead to this reduction in *Firmicutes*, but can also be followed by the increased incidence of pathobionts. Pathobionts are bacteria that are not necessarily pathogenic under homeostasis, but have the propensity to become pathogenic, for example, in the context of old age. Pathobionts include various aerobic bacterial families such as *Enterobacteriaceae*, *Enterococcaceae* and *Staphylococcaceae* (Santoro et al., 2018). Older and frailer patients have also been shown to have key marker bacteria that include *Eubacterium dolichum and Eggerthella lentha,* which coincides with a reduction in *Faecalibacterium prausnitzii,* a bacteria typically associated with positive health (Jackson et al., 2016b). Additionally, in a study of 161 elderly Irish individuals, *Bacteroides* was the predominant phylum, but in comparison to younger control subjects, there was a greater abundance of *Clostridium sp.* (Claesson et al., 2011). The significance of an altered microbiota in the elderly is that it could potentially mediate functional consequences for the host. One study implicated the age-associated changes of the gut microbiota as the causative factor of age-related inflammation (Thevaranjan et al., 2017). Thus, age is a powerful factor that influences the microbiota with great physiological effect.

1.6.1.2 Sex

It is known that sex differences can mediate biological differences within the host, for example in terms of altered immunity. Indeed, males are thought to be more susceptible to infection than females (vom Steeg and Klein, 2016). Thus it is likely that the gut microbiota also differs between males and females. One study investigated the microbiota between mice of different strains and different sexes (Elderman et al., 2018). They found several species that were enriched in one sex, but not the other, but this was dependent upon the strain of the mouse. For example, in BALB/c mice, *Bifidobacteria* were enriched in females compared to males, whereas in C57/BL6 mice, *Lactobacillus plantarum* was enriched in females compared to males. These findings are supported by another study, who also show-sex

differences in the microbiota, but that the extent of these differences are dependent on strain and host genotype (Org et al., 2016). Differences in the microbiota correlated with differences in immune gene expression, suggesting that sex differences in the gut microbiota could mediate functional differences for the host (Elderman et al., 2018). However, one study does reveal that there are immune differences between sexes that are independent of the gut microbiota (Fransen et al., 2017). Indeed, they showed that type I interferon signalling was increased in the small intestine of germfree female mice. As the male mice were less immunologically active, they noted an increase within these mice of *Alistipes, Rikenella*, and the family *Porphyromonadaceae*, bacteria that are thought to 'bloom' when innate immunity is reduced (Fransen et al., 2017). Therefore, sex differences are something that should be taken into account for most experimental studies and in the case of the microbiota, it is vital to ensure that appropriate comparisons are made.

1.6.1.3 Species

Although it may be expected that different species would have a different microbiota, mouse models are often used, as it facilitates a more practical approach to understand how the microbiota may be associated with a variety of conditions of interest. Thus, it is important to consider whether the mouse and the human microbiota can be compared, an issue that was reviewed by Nguyen et al. (2015). Anatomically, the intestinal structure is similar between mouse and humans, although there are regional differences, such as mouse villi being longer than human villi to provide a larger surface area for absorption (Nguyen et al., 2015). In terms of the gut microbiota, under naïve conditions, both humans and mice share similarities at higher taxonomic levels. Indeed, at the phylum level, both the human and the mouse gut microbiota are dominated by *Firmicutes and Bacteroidetes* (Eckburg et al., 2005). However, there are some phylum level differences with mice containing *Deferribacteres* in the gut, which are localised to the stomach and comprise a small fraction of the human microbiota (Bik et al., 2006; Hugenholtz and de Vos, 2018). At lower taxonomic levels, genera are typically different between humans and mice, with 85% of genera in mice not found in humans (Ley et al., 2005). In terms of similarities however, *Clostridium*, *Bacteroides* and *Blautia* are found in both mice and humans with similar levels of abundance (Nguyen et al., 2015). However, they also show that genera common between the two species can differ in abundance. For example, *Prevotella*, *Faecalibacteria* and *Ruminococcus* are more common in the human gut, whereas *Lactobacillus*, *Alistipes* and *Turicibacter* are more abundant in mouse gut.

Thus, there are both similarities and differences in the healthy gut microbiota between humans and mice, but changes in the microbiota are also associated with disease, such as IBD. Do changes seen in the human gut microbiota also correlate with changes seen in the mouse microbiota? IBD is typically characterised by a reduction in bacterial diversity in both humans and mice (Ott and Schreiber, 2006; Nagao-Kitamoto et al., 2016; Pascal et al., 2017). However, there are specific differences in how the microbiota changes, for example *F. prausnitzii* is often reduced in human IBD patients, but this reduction is not seen within mice (Vigsnaes et al., 2012; Nguyen et al., 2015). Crucially, mice have potentially many more considerations for factors that could influence their microbiota. Like humans, the role of maternal transmission of the gut microbiota is crucial thus it is important to use littermate controls **(Section 1.5.1)**. However, mice also partake in coprophagy, leading to the gradual homogenisation of the microbiota between mice within a cage (Soave and Brand, 1991). Thus, cage effect is a powerful factor that can influence the microbiota (Hildebrand et al., 2013). However, many animal studies do not report these vital factors (Bramhall et al., 2015; Florez-Vargas et al., 2016).

1.6.1.4 Antibiotics

Antibiotics have revolutionised healthcare and despite the increasing incidence of antibiotic resistance, they are still widely used to combat serious bacterial infection. However, antibiotics also have consequences for the gut microbiota especially during childhood (Langdon et al., 2016). Evidence suggests that childhood antibiotic usage reduces colonisation with beneficial species such as *Bifidobacteria* and *Lactobacilli* (Penders et al., 2006). Delayed colonisation of the gut by these species may be associated with a spectrum of disorders such as asthma, eczema, inflammatory bowel disease, obesity and diabetes (Dietert, 2014). One study found that usage of beta-lactam antibiotics such as penicillin led to an alteration in the gut microbiota and carbohydrate metabolism, similar to that seen in obese patients (Hernandez et al., 2013). Studies have also found associations between early antibiotic usage and asthma (Droste et al., 2000; Stensballe et al., 2013). Children who had taken antibiotics in the first year of life had significant associations with development of asthma, taking into account factors such as earlier childhood respiratory illness (Droste et al., 2000). Links with asthma were also attributed to maternal use of antibiotics during pregnancy (Stensballe et al., 2013). An integrative study that comprised genome, transcript, protein and metabolite analysis aimed to investigate how antibiotics could influence each of these factors (Perez-Cobas et al., 2013). Overall, they identified oscillatory shifts in bacterial composition

and metabolites at 6, 11 and 14 days from the introduction of antibiotics (Perez-Cobas et al., 2013). However the expense of conducting such an integrative approach into microbiome studies is often a limiting factor.

The adult microbiota however is less susceptible to long-term disruption than the childhood microbiota perhaps because it has fully established and is more stable. One study of five healthy volunteers who took amoxicillin for 30 days (De La Cochetiere et al., 2005) showed that although there was significant initial disruption to the microbiota, sequence data from faecal samples after 4 days of treatment suggested that the microbiota was 74% similar to the profile found at day 0, suggesting resilience of the adult microbiota. However, in one individual, the disruption lasted for a further two months and so they suggest that variability in resilience may be why some adults are more susceptible to outcomes such as antibiotic-induced diarrhoea (De La Cochetiere et al., 2005). Such diarrhoea is often associated with the presence of 'harmful' species including *C. difficile* and *C. perfringens,* which can bloom during antibiotic treatment (Ackermann et al., 2005). Other consequences of these perturbations to the normal gut microbiota could lead to an altered metabolite profile.

1.7 The Microbiota and Metabolites

One mechanism through which the microbiota facilitates health is via the production of metabolites- small molecules derived from the diet, drug metabolism and from xenobiotics (foreign, small molecules) (Li and Jia, 2013; Flint et al., 2014). The key metabolites that microbiota contribute include bile and amino acids, tryptophan-derived metabolites and short-chained fatty acids (SCFAs) (Aw and Fukuda, 2014). In germ-free mice, the concentrations of plasma metabolites were significantly reduced compared to conventionally-housed mice, highlighting the importance of the microbiota in metabolite production (Wikoff et al., 2009). The microbiota are vital for determining the host's capacity to harvest energy from the diet, as shown by the transfer of the microbiota from obese mice to germ-free mice, where germ-free mice acquiring this microbiota had dramatically increased energy intake (Turnbaugh et al., 2006). Specifically, the microbiota expands the metabolomic profile of the host, that the host would be unable to produce otherwise (Qin et al., 2010). The pathways involved in producing these metabolites has been reviewed recently by Rowland et al. (2018). However, specific metabolites will now be discussed.

1.7.1 SCFAs

The most commonly discussed metabolites in the context of the microbiota are the SCFAs, notably acetate, propionate and butyrate, known major contributory components to host health (Morrison and Preston, 2016). The highest levels of SCFAs are found within the colon, where they are absorbed by colonocytes, or can otherwise enter the liver and circulation via the hepatic and portal venous system (Bloemen et al., 2009; Layden et al., 2013). SCFAs are typically produced from the saccharolytic fermentation of carbohydrates, derived from cellulose and dietary fibres (Morrison and Preston, 2016). The carbohydrates that SCFAs are derived from are referred to as 'non-digestible' as the host is incapable of breaking them down and depends upon the large intestinal gut microbiota to metabolise them (Levy et al., 2016). Acetate is mainly produced by the phylum *Bacteroidetes* and that butyrate is produced by the phylum *Firmicutes* (Hoverstad and Midtvedt, 1986; Macfarlane and Macfarlane, 2003). With regards to propionate, its production appears to be spread across different phylum but with a limited range of bacteria, mostly comprising the mucin-degrading bacteria *Akkermansia muciniphilla* (Derrien et al., 2004; Morrison and Preston, 2016). The species *Ruminococcus bromii* plays a key role in butyrate production, as without this bacteria, starch degradation is diminished resulting in reduced levels of butyrate (Walker et al., 2011; Ze et al., 2012). *F. prausnitzii* is another butyrate-associated bacteria, and low levels of this bacteria correlates with IBD in human patients (Takahashi et al., 2016; Ferreira-Halder et al., 2017).

SCFAs have numerous functional impacts upon the host. In addition to being an energy source (accounting for up to 10% of caloric intake per day (Donohoe et al., 2011)), they are potent signalling molecules that bind to a wide array of receptors. For example, a diet containing non-digestible fibres are thought to be linked to reduced incidence of obesity, type 2 diabetes and cardiovascular disease (Cho et al., 2013). The mechanism for this health benefit could be due to SCFAs mediating a switch from lipogenesis to fatty acid oxidation, by binding to the peroxisome proliferator-activated receptor gamma receptor (PPARγ) (den Besten et al., 2015). This in turn can potentially lead to increased intestinal gluconeogenesis (De Vadder et al., 2014). SCFAs are also known to bind to fatty acid receptors, such as free fatty acid receptor (FFAR) 2 and 3, which can lead to modulation of adiposity and changes in gut transit time (Brown et al., 2003; Samuel et al., 2008; De Vadder et al., 2014).

SCFAs are also important to regulate immunity. Butyrate production by commensal bacteria was involved in modulation of T regulatory (T reg) cells, which are involved in suppressing

immune responses (Arpaia et al., 2013). Other protective SCFAs include propionate, which has been shown to have a protective benefit in allergic airway disease (Trompette et al., 2014). Microbiome-produced acetate mediates protection against enterohaemorrhagic *Escherichia coli* (EHEC) O157:H7 (Fukuda et al., 2011). *Bifidobacteria* strains that expressed ATP-binding cassette (ABC) carbohydrate transporters could produce acetate in the distal colon which was thought to mediate protection against EHEC-induced death, perhaps due to enhanced defence at the epithelial cell barrier. Additionally, acetate plays an important role in heart, muscle and brain cell development (Wong et al., 2006; Qin et al., 2010).

1.7.2 Bile acids

Bile acids are the end products of cholesterol metabolism, accounting for a large portion of converted cholesterol (Chiang, 1998; Chiang, 2013). Cholic acid (CA) and chenodeoxycholic acid (CDCA) are the major bile acids synthesised (Chiang, 2013) and act to solubilise dietary fat within the small intestine to facilitate the absorption of fat-derived vitamins (Brestoff and Artis, 2013). Bile acids have also been shown to be an important mediator of the gut microbiota and host health (Kakiyama et al., 2013; Ridlon et al., 2014). Indeed, in cirrhosis, it was shown that dysbiosis of the colonic mucus microbiota was linked to a reduction in bile acids in the intestine (Bajaj et al., 2012; Kakiyama et al., 2013). A reduction in bile acids led to an increase in the pathobiont family, *Enterrobacteriaceae* (Bajaj et al., 2012). Additionally, bile acids are thought to be involved in control of hepatic tumour development, where the gut microbiota metabolises bile acids to indirectly influence natural killer (NK) T cells within the liver, a cell type involved in killing host cells (Ma et al., 2018).

1.7.3 Amino acids and tryptophan-derived metabolites

Amino acids are defined as organic substances that contain both amino and acid groups and are the building blocks of proteins, promote synthesis of hormones as well as acting as key signalling molecules in a variety of pathways (Wu, 2009). The gut microbiota plays a vital role in metabolism and maintenance of amino acids within the host (Lin et al., 2017). For example, the microbiota is required for the metabolism of tyrosine and tryptophan to p-cresol and indole-3-aldehyde (De Preter et al., 2015). Indole-3-aldehyde mediates protection against chemically induced colitis and *Candida albicans* infection, via an IL-22 dependent mechanism (Zelante et al., 2013).

Tryptophan-derived metabolites are important in mediating host health and the gut microbiota. *Lactobacilli,* are responsible for producing many of these tryptophan-derived metabolites (Levy et al., 2016). The resulting compounds act as ligands for the aryl hydrocarbon receptor (AhR), a transcription factor with potent immune effects (Monteleone et al., 2011; Zelante et al., 2013) such as development of intestinal lymphoid follicles (Kiss et al., 2011). AhR is important for maintenance of intestinal barrier function and the absence of AhR is associated with reduced intestinal epithelial cell turnover (Li et al., 2011; Levy et al., 2016). AhR activity is strongly mediated by diet, with 'cruciferous vegetables' (vegetables such as broccoli and cauliflower) leading to increasing activity (Li et al., 2011). However, a high fat diet can reduce certain tryptophan-metabolites, including indole-3-acetate (I3A) that attenuate hepatic inflammation (Krishnan et al., 2018).

1.7.4 Analysis of metabolites

A variety of techniques can be used to analyse metabolites and are likely to be context specific. For example, studies typically analyse serum and urine to give an overall 'metabolomic footprint' of the host (Kell et al., 2005). However, this is unlikely to reflect on metabolite changes at the local level, for example in specific tissue sites. Additionally, the metabolites found within these samples are likely to be host-derived, which could be a confounding factor when bacteria-specific metabolites need to be investigated.

Due to the complexity of the metabolome, one single method is insufficient to accurately profile the metabolites within the host (Dunn et al., 2005). The most common methods involve either gas or liquid chromatography, followed by mass spectrometry (GC or LC-MS) (Dunn et al., 2011). If the metabolites to be analysed are from tissue, the tissue is mechanically disaggregated resulting in separation of the sample into nonpolar and polar fractions. Both fractions are collected and vacuum concentrated separately. Metabolites in the polar fraction undergo a process known as chemical derivatisation, which makes the compounds more volatile (Begley et al., 2009). This process is necessary for GC-MS, where compounds are required to be volatile and thermally stable. GC-MS is a high resolution technique that allows the separation of structurally similar compounds (Begley et al., 2009). The remaining non-polar fraction goes forward for LC-MS. LC-MS can detect a broad range of non-polar metabolites and is used where compounds cannot be chemically derivatised (De Preter and Verbeke, 2013). MS is typically the same across any instrument, where samples are introduced, ions are formed and separated according to their mass:charge (m/z) ratio and spatially or temporally separated ions are detected and analysed (Ellis et al., 2007). The

identity of each metabolite can be determined by its m/z value and retention time of its ion peak, and its relative concentration estimated based on the height of the ion peak (Yoshida et al., 2012).

Analysis of metabolites is a particularly important step, especially if functional inferences are to be made from the data. As stated in this section, metabolites are important for maintenance of host health. Therefore, a growing area of research is how bacteria-derived metabolites may affect the development of IBD.

1.8 Inflammatory Bowel Disease

IBD refers to a spectrum of chronic inflammatory diseases namely ulcerative colitis (UC) and Crohn's Disease (CD). Morphologically, signs of IBD manifest in terms of loss of gut architecture, such as crypt or villi destruction (Geboes and Dalle, 2002), though the extent and nature of the damage differs in CD and UC (Nostrant et al., 1987). UC is characterised by a continuous inflammation of the colon and is specific to the large intestine, whereas CD can affect anywhere in the GI tract (from the mouth to the anus) and the inflammation is typically more severe leading to fistulas (Podolsky, 1991; Li et al., 2014a). CD is associated with high levels of inflammatory cytokines IFN-γ and IL-17 and thus often considered a Th1 type of disease, whereas UC tends to have higher levels of IL-13 and can be thought of as more Th2 biased (Bao et al., 2017).

The disease typically manifests in adulthood, although can occur at any time throughout childhood (Wehkamp et al., 2016). IBD is becoming increasingly more prevalent in developed countries and is also now being diagnosed in developing countries (Molodecky et al., 2012; Ng et al., 2018). IBD is associated with a variety of symptoms such as abdominal pain and diarrhoea, relapsing and remitting inflammation and a frequent need for surgery to remove inflamed gut (Kim and Cheon, 2017). The exact aetiology of the disease is unknown and is multifaceted, including increased barrier permeability, altered immune responses to commensal bacteria and immune dysregulation (Kang et al., 2013). The mechanisms behind these disruptions could be as a result of both genetic and environmental factors (Halfvarson, 2011; Ek et al., 2014). Twin and heritability studies have failed to shed light into how important genetic factors are for the development of IBD, but discuss that the interplay between genetics and environment appears to be the key factor for IBD development (Halfvarson, 2011; Gordon et al., 2015). However, there are some genetic predispositions for

IBD, particularly with regards to genes that alter the interaction of the host with the microbiota (Jostins et al., 2012). For example the PRR *NOD2* has classically been linked to the development of ileal forms of Crohn's Disease (Peyrin-Biroulet and Chamaillard, 2007; Vignal et al., 2007). *NOD2* is involved in microbial recognition (by binding a component of bacterial peptidoglycan) and hence contributes to immune and microbial homeostasis (Moreira and Zamboni, 2012). As such, *NOD2* deficiency has been implicated in increasing bacterial load within the gut and thus has the potential to mediate pathology (Petnicki-Ocwieja et al., 2009).

In many patients with IBD, there is often 'dysbiosis' of the gut microbiota, where the microbiota is different when compared to the healthy state of the host. However, whether inflammation occurs first which then impacts on the gut microbiota, or whether the microbiota change before the onset of inflammation is unknown. However, we have shown that the mucus-resident bacteria are altered in a spontaneous mouse model of colitis, before the onset of any inflammation (Glymenaki et al., 2017). In colitis generally, the mucus layer is thinner and goblet cell numbers decrease (Alipour et al., 2016; Glymenaki et al., 2017). The microbiome is normally separated from the host within the lumen and the mucus by the antimicrobial properties of both the mucus and the epithelial cell barrier (Kostic et al., 2014). Thus, the implications of a thinner mucus are that the microbiota could directly contact host cells and potentially drive inflammation (Glymenaki et al., 2017). In order to investigate the aetiology of IBD, various mouse models have been employed. These models can take different forms. There are chemically induced IBD models, such as the trinitrobenzene sulfonic acid (TNBS)-induced IBD mouse model (Ettreiki et al., 2012) and the dextran-sodium sulphate (DSS) mouse model (Okayasu et al., 1990; Chassaing et al., 2014). There are biologically-induced models, for example, the parasite *Trichuris muris* has been used to drive inflammation in the gut (Levison et al., 2011; Levison et al., 2013). There are also gene-knockout models that lead to the spontaneous development of colitis, such as Muc2-deficient mice (Van der Sluis et al., 2006), mice that have the regulatory cytokine IL-10 knocked out (IL-10$^{-/-}$ mice) (Rennick et al., 1997) and *mdr1a*$^{-/-}$ mice (Panwala et al., 1998). This thesis will focus on the multidrug resistance (*mdr1a*$^{-/-}$) mouse model.

1.8.1 The *mdr1a*$^{-/-}$ Mouse Model

The *mdr1a*$^{-/-}$ mouse model is a genetically engineered model that develops spontaneous colitis over time, while having a fully functioning immune system (Panwala et al., 1998). *Mdr1a* is a large (170 kDa) P-glycoprotein multidrug resistance pump expressed on the cell

surface (Wilk et al., 2005). It is part of a family of ATP binding cassette transporters that transport small hydrophobic molecules, such as xenobiotics, across the membrane of the intestinal epithelium. The efflux of xenobiotics is likely to have a protective effect on the intestinal epithelial barrier by reducing potential toxicity (Panwala et al., 1998) and has also been suggested to promote cell differentiation and survival (Johnstone et al., 2000). *Mdr1a* is encoded by two genes in humans (MDR1 and MDR3) and three genes in rodents (*mdr1a*, *mdr1b* and *mdr2*) (Chen et al., 1986; Hsu et al., 1989). The *mdr1a* gene is expressed on a variety of cell types, including CD4+ and CD8+ T cells, natural killer cells, dendritic cells, IECs and endothelial cells at the blood brain barrier (Chaudhary et al., 1992; Panwala et al., 1998; Wilk et al., 2005).

Having the *mdr1a* gene knocked out results in the eventual development of colitis in mice housed under SPF conditions, whereas germ-free *mdr1a*$^{-/-}$ do not develop colitis (Wilk et al., 2005). Colitis can be reversed by the administration of antibiotics, supporting the importance of the microbiota in colitis development (Panwala et al., 1998). There is a variation in when this colitis develops, with a range of 8 to 35 weeks (Wilk et al., 2005). In our hands however, the early signs of colitis occur around 18 weeks (Glymenaki et al., 2017). In humans, the MDR1 gene has been reported to be a potential IBD risk factor (Ardizzone et al., 2007; Mijac et al., 2018).

Mdr1a mice have a significantly altered gut microbiota, especially in the mucus-resident bacteria, compared to WT mice (Glymenaki et al., 2017). The ratio of gram positive to gram negative bacteria was increased in the mucus of pre-colitic *mdr1a*$^{-/-}$ mice, but not in the stool. An increase in the gram negative bacteria in the stool was associated with inflammation symptoms in the *mdr1a*$^{-/-}$ mice. However, the overall microbial burden was the same in both stool and mucus samples (Glymenaki et al., 2017).

1.8.2 The Microbiota in IBD

IBD is associated with aberrant immune reactivity to the microbiota. Antibiotic treated mice, or colitis-prone strains of mice such as the IL-2$^{-/-}$and *IL-10*$^{-/-}$ mice do not develop IBD symptoms if reared in germ-free conditions (Dove et al., 1997). Studies of IBD have consistently shown a reduction in gut bacterial diversity (Manichanh et al., 2006; Alipour et al., 2016; Glymenaki et al., 2017). Specifically, this means a reduction in the number of different types of species within the gut, which could correspond with the pathobiont bloom (**Section 1.6.1.1**). In human studies, it has been shown that *F. prausnitzii* is reduced in IBD

patients, from both stool samples (Sokol et al., 2009; Ferreira-Halder et al., 2017) and the colonic mucus (Lopez-Siles et al., 2015). *F. prausnitzii* is thought to mediate a protective effect, as *in vitro* application of its supernatant to peripheral blood mononuclear cells (PBMCs) led to reduced inflammatory gene expression of IFN-γ and an upregulation of regulatory cytokine IL-10 (Sokol et al., 2008). Additionally, in a chemically-induced mouse model of IBD (the TNBS model) they revealed that oral administration of live *F. prausnitzii* or its supernatant significantly reduced pathology (Sokol et al., 2008). The mechanism behind these protective effects could be linked to butyrate as *F. prausnitzii* and an additional genus *Roseburia,* are thought to be key SCFA producers (Morgan et al., 2012; Geirnaert et al., 2017; Imhann et al., 2018). *Roseburia hominis* is a known butyrate producing species and was found to be diminished in the gut microbiota of UC patients (Machiels et al., 2014).

Other important bacteria include Clostridia, particularly of the cluster XIVa and IV (Kim and Cheon, 2017) which are thought to induce the production of T-reg cells that are involved in anti-inflammatory responses (Atarashi et al., 2011). Treatment of mice with antibiotics that targeted gram positive or gram negative bacteria, demonstrated a role for gram positive bacteria and specifically spore-forming Clostridia in intestinal T-regs production (Atarashi et al., 2011). *Clostridia,* but not other bacteria, such as *Lactobacilli,* caused IECs to produce greater levels of the active form of TGF-β, an anti-inflammatory cytokine that induces T-reg cells. Depletion of *Clostridia* can be associated with IBD development as well as coincide with upregulation of invasive *E. coli* that have been identified in the mucosa of CD patients (Baumgart et al., 2007).

1.8.3 Metabolites in IBD

In IBD, both CD and UC patients have a disrupted metabolism, and quantitative metabolomic profiling can discriminate between healthy and IBD patients (Schicho et al., 2012). SCFAs have been used therapeutically to ameliorate the symptoms of IBD (Harig et al., 1989; Scheppach et al., 1992; Scheppach, 1996) and highlight the role of butyrate as an anti-inflammatory agent. A study that examined the faeces of both CD and UC, showed decreased levels of butyrate and acetate compared to control subjects and elevated levels of amino acids (Marchesi et al., 2007). Indeed, one study investigated the serum and plasma of IBD patients and found that isoleucine and one of its by-products, 3-methyl-2-oxovalerate is increased in the serum and plasma of patients with IBD (Schicho et al., 2012; Bao et al., 2017). There also appears to be region-specific differences, as levels of amino acids in the faeces and the colonic mucosa of IBD patients differ (Marchesi et al., 2007; Balasubramanian et al., 2009).

Specifically, amino acids, particularly isoleucine and leucine have reduced concentrations in the colonic mucus of IBD patients (Balasubramanian et al., 2009). Arginine, an important metabolite known to play a role in wound repair with anti-inflammatory effects (Alexander and Supp, 2014), ameliorated the symptoms of TNBS-induced colitis in rats (Al-Drees and Khalil, 2016).

A caveat of studies that look at both the microbiome and metabolites is how the microbiome has been analysed. Therefore, it is important to consider the overall procedures used, starting from sample collection to data analysis.

1.9 Analysing the microbiome

With an ever growing number of microbiome studies, it is important to consider how such studies are analysed as the analysis can have dramatic impacts on the quality of findings. In this section I will consider the methods for sampling, analysing and presenting data from these studies.

1.9.1 Sample collection

For studies involving the gut microbiome, it is common to take faecal samples. This can only give a partial insight into the intestinal microbiota as bacteria reside in the mucus and lumen and vary along the GI tract. However, extracting bacteria from the mucus is not as straightforward because samples may require surgical collection from the gut, as well as careful preservation.

Most patients undergoing surgery or biopsy will have undergone some form of colonic lavage (Hollister et al., 2014). *Mucispirillum*, a genus known for colonising the GI mucus layer in laboratory rodents, is lost after a bowel preparation (Robertson et al., 2005). In addition, lavage would likely alter the luminal bacteria and so is a practice that should possibly be avoided with regards to gut microbiome studies. One study examined tissue from patients undergoing emergency bowel operations, where no colonic preparation was performed (Ahmed et al., 2007) that confirmed differences in the microbiota between different microbial niches within the gut. Taking tissue in this instance may avoid the issue of bowel treatments altering the microbiome, but it likely to be impractical. A solution to this problem has been proposed by the company Origin Sciences, who have developed a novel mucus extraction method from human patients (Sciences, 2018).

1.9.2 Sample extraction

Following sample collection, the specimens must be processed for downstream analysis, the process of which will differ according to the microbial community to be analysed i.e. bacteria, fungi or viruses as well as the subsequent sequencing to be performed. Several comprehensive reviews have analysed the efficacy of these extraction methods (Kuczynski et al., 2012; Morgan and Huttenhower, 2014; Wesolowska-Andersen et al., 2014). Once the DNA has been extracted there are several approaches that can be taken to analyse the microbiome.

1.9.3 Sequencing-independent analysis

Denaturing gel gradient electrophoresis (DGGE) is a method that allows a prompt, qualitative visualisation of differences in bacterial diversity between samples. By exposing the microbiome sample to extreme denaturing conditions, the sample can be run on a gel and bands corresponding to different taxa separate depending upon their percentage GC content. This is commonly referred to as 'genetic fingerprinting' (Tzeneva et al., 2008). Individual bands that differ between treatment groups can subsequently be excised and sequenced. qPCR is another common method to analyse the gut microbiome and can ultimately provide greater taxonomic resolution than NGS, depending on the primers used. Other sequencing-independent techniques include fluorescence *in-situ* hybridisation (FISH), that utilises fixed tissue sections and fluorescent staining to visualise specific DNA sequences. It is used to identify and visualise specific or groups of taxa and is therefore more focused on identifying where bacteria are localised within the tissue. A relatively recent technique has involved exploiting IgA binding of bacteria (Palm et al., 2014). The rationale was that specific IgA types would develop to target potentially damaging bacteria in the microbiota and bind these with high affinity. Indeed, bacteria that possessed high levels of bound IgA could be isolated and discovered to induce colitis when transplanted into germ-free mouse models (Palm et al., 2014). However, specifically identifying the IgA-coated bacteria required the use of sequencing technology.

1.9.4 Sequencing

After acquisition and processing of appropriate samples, they must be analysed. The analysis involves the use of bacterial DNA sequence to determine bacterial identity and as a result, allows the identification of taxonomic diversity within a given environment (Morgan and Huttenhower, 2012). Sequencing bacteria can be done in various ways, the most common methods being 16S rRNA sequencing and whole genome sequencing. Whole genome

sequencing is becoming increasingly more feasible and has been shown to have numerous advantages over 16S rRNA sequencing, such as increased read length and greater resolution down to the species level (Ranjan et al., 2016). However, given that that this thesis has focused on 16S rRNA sequence data and that the technique remains one of the most commonly used methods used, this technique will be discussed.

Rather than sequence a whole genome, which was historically impractical, it was necessary to identify a suitable bacterial marker that is ubiquitously expressed within all species, but also allows the elucidation of different bacterial taxa. As such, the 16S rRNA gene was an excellent candidate (Tringe and Hugenholtz, 2008). The gene comprises nine hypervariable regions flanked by conserved sites. The conserved sites facilitate the use of universal sequencing primers and the variable regions allow taxonomic identification of bacteria (Tyler et al., 2014). The variable regions themselves are a key part of 16S rRNA sequencing. Specifically, primers can either target a specific variable region or can span several of the regions. Studies have investigated 'optimal' variable region and the consensus is that different regions are better at identifying certain types of taxa, for example V3 was better at identifying *Haemophillus* species (Chakravorty et al., 2007). Another study suggested that V4-V6 were the most optimal regions for taxonomic resolution (Yang et al., 2016). In reality, the region used is ultimately dependent upon the practices of the sequencing facility at a given institution.

The process of 16S rRNA sequencing must be facilitated by appropriate sequencing technology. Following capillary sequencing, advancing technology led to more high-throughput sequencing methods and this has broadened research into diverse microbial communities in a range of environments (DeLong et al., 2006; Qin et al., 2010). One of the earliest was Roche 454 pyrosequencing, although this has ultimately been superseded by Illumina sequencing. These technologies all sequence DNA on a 'massively-parallel' scale, albeit through different mechanisms. 454 sequencing measures light released upon the joining of a nucleotide to the DNA strand, which is problematic when sequencing homopolymeric regions (Luo et al., 2012). Illumina sequencing uses reversibly-terminating nucleotides containing fluorophores. This takes place on a flow chip, although sequencing discrepancies have been noted on different tiles of the flow chip (Luo et al., 2012). At comparable cost, Illumina has been shown to outperform 454 under numerous circumstances (Luo et al., 2012; Li et al., 2014c).

However, all next-generation technologies have relatively short read lengths, i.e. the maximum length of DNA that can be sequenced before artefacts begin to appear is quite short. Read length is particularly important because this allows improved resolution when identifying bacterial taxa (Tyler et al., 2014). Read lengths are often kept lower to minimise the risk of sequencing errors and artefacts within the reads, although as technology improves, the read length should also improve. The sequencing output then undergoes 'quality control'. This comprises the removal of chimeric sequences (where partial PCR products have been used as a template for amplification to generate recombinant sequences), low quality reads (where read length falls below a predetermined threshold in comparison to the mean read length), duplicate reads and other sequencing artefacts (Haas et al., 2011; Matsen, 2014). The sequences then require identification. For 16S rRNA sequences, data is often separated into operational taxonomic units (OTU's).

1.9.5 OTU's

Next generation sequencing generates a large amount of data that can be computationally difficult to process due to the size. OTU's are one way of condensing down the data while providing taxonomic identification of the microbiota. OTU's are formed by the clustering of similar sequences at a specified threshold (usually 95-97%) and typically, a representative sequence from this cluster is compared against a reference database in order to determine taxonomic identity (Ursell et al., 2012; Tyler et al., 2014). Any sequences that do not match against a database are either discarded for downstream analysis (closed-reference OTU picking) or clustered together into their own 'de novo' OTU (open-reference OTU picking) (Caporaso et al., 2010). From this point, the data can be processed in several ways, but it is common to determine relative abundance of an OTU. Specifically, this provides information about the abundance of each identified taxa relative to each sample. Other types of analysis include diversity analysis.

1.9.6 Diversity Analysis

Many studies calculate 'diversity' in order to describe the microbiome. Calculated parameters usually include 'alpha diversity', which refers to the richness of the sample (i.e. the number of different species within a sample) (Morgan and Huttenhower, 2014). Specifically, alpha diversity refers to the total number of different organisms in a sample. If one sample has a greater variety of organisms than another, it would be considered to have a higher alpha diversity. In order to calculate alpha diversity, specific statistical methods are required in order to estimate richness, with the most common estimators being the Shannon-Wiener

index, abundance-based coverage estimator (ACE) and Chao1 (Chao and Lee, 1991; (Hughes et al., 2001).

Beta diversity is also used, which assesses the degree of dissimilarity between the taxa of two different samples. Specifically, it addresses how different the microbiota is between different samples. There are many methods to calculate beta diversity, although Bray Curtis dissimilarity is perhaps the most commonly used. The idea is to construct a data table that contains samples as rows and the species as columns. The number of each species, in each sample is determined. A mathematical model is then applied to convert this table into a 'distance matrix'. These distances are used to work out how similar (or dissimilar) the microbial communities are from one another.

Despite wide usage of the aforementioned diversity metrics, they do have disadvantages. Diversity does not necessarily take into account relationships and diversity structure between species, such that two OTUs that are closely related and two OTUs that are distantly related will contribute equally to the diversity measure (Matsen, 2014). Indeed, this leads to a large loss of potentially useful information (Lozupone and Knight, 2005). Instead, phylogenetic diversity (PD) measures can be used which do take into account relatedness.

PD ultimately requires the construction of a phylogenetic tree and uses a combination of branch length and structure to quantify diversity (Matsen, 2014). UniFrac (unique fraction) is the main computational method used for calculating PD (Lozupone and Knight, 2005). After constructing a combined tree based on two samples, it measures the 'fraction of the branch length that leads to descendants in one environment, but not the other' (Lozupone and Knight, 2005). Regardless of method, the processing of sequence data is particularly important. For example, in terms of OTU binning, errors in base pair reading can lead to new OTUs being incorrectly assigned and this can affect downstream analysis (Morgan and Huttenhower, 2012).

1.9.7 Visualising data

An additional point to consider is what happens after the calculation of diversity metrics etc. The data produced is often presented in the form of ordination, such as non-metric multidimensional scaling (NMDS) or principle component analysis (PCA) (Morgan and Huttenhower, 2014). NMDS involves calculating pairwise distances between samples (using the beta diversity calculations) and transposing these graphically, so that similar samples cluster together on a graph and dissimilar samples are separated.

PCA is ultimately a method of transforming a larger set of complicated data into a smaller set of linearly uncorrelated variables, plotted against axes known as principle components (Tyler et al., 2014). The principle components refer to factors that explain the most variation within the data, with the first principle component (PC1) explaining the most variation, and subsequent factors explaining progressively less. Data is usually only plotted graphically against PC1 and PC2. Ultimately, PCA allows a qualitative visualisation of factors that are associated with variance of the data (Morgan and Huttenhower, 2014). Clustering of the data points is then examined in the context of phenotypic variables such as treatments used (Tyler et al., 2014). The overall advantage provided by PCA is that it allows complex, large data sets to be condensed and understood more clearly. It should be noted that PCA is distinct from Principle Coordinates Analysis (PCoA), which uses aforementioned distance calculations to determine position of points on the axis (Tyler et al., 2014). In the context of the microbiome, PCA and PCoA have been widely used to visualise the large, complex sequencing outputs to illustrate differences between gender (Markle et al., 2013), age (Claesson et al., 2011), as well as providing 'evidence' for enterotyping (Arumugam et al., 2011). However, Knights et al. (2014) argue that due to the visual nature of PCA, plots such as 'scatterbursts', in combination with colours, can cause the eye to perceive discrete clusters that aren't necessarily there. In concordance with Knights et al. (2014), Morgan and Huttenhower (2014) also suggest that such ordinations do not provide information about statistical significance or whether external factors outside of those phenotypically described influence the patterns displayed. Additionally, as PCA only tends to focus on extreme variation, more subtle effects, such as within-cluster variation, may be missed (Knights et al., 2014; Morgan and Huttenhower, 2014).

It should be noted that PCA is ultimately a form of unsupervised analysis, which is able to provide non-specific analysis based on similarities between microbiome samples (Cui and Zhang, 2013). Ultimately, this highlights major relationships within the data in unbiased terms. This is in contrast to supervised analysis, which is able to select for pre-specified patterns that may not be immediately apparent and can be a powerful tool for exploring large datasets (Cui and Zhang, 2013). Examples of supervised analysis include Decisions Trees and Random Forests.

1.9.8 Decision Trees and Random Forests

Decision trees are a type of supervised learning that are designed to use features in the data to categorise it into groups. It is first important to determine the question to be asked. For

instance, one question could be, using existing data: will a mouse develop colitis? An example dataset is illustrated in **Table 1** and **Figure 2**. The tree process involves starting at the root (which represents the entire dataset). This root gets split into smaller datasets based on a feature. For instance, the feature could be the abundance of *Lactobacilli* bacteria in a sample. Node 'purity' is then assessed, which is a measure of homogeneity of samples when the data is split. In the example data, there are 12 mice, 5 of which have colitis and 7 do not. If the abundance of *Lactobacilli* is used to split the samples, it can be seen that mice are fully separated into healthy mice and those that have colitis. However, if age is used to split samples, all mice that have colitis are identified, but only 1 out of 7 healthy mice is identified. Therefore, the node that fully separated all mice (the split based on *Lactobacilli*) has a higher purity value. If no feature can fully separate out the samples, more features can be added that help to make the data more homogenous. Once the decision tree is established, novel data can be added. For example, if an 18 week old mouse has been identified that has low abundance of *Lactobacilli,* the decision tree would suggest that the mouse is likely to develop colitis. However, there is a danger of overfitting, where the decision tree works well under testing conditions, but fares poorly against novel data. Another method, the Random Forest (RF) model, is less susceptible to overfitting.

Table 1: Example dataset used to inform a decision tree.

Age (weeks)	Genotype	Abundance of Lactobacilli	Colitis
6	WT	0.6	No
6	KO	0.7	No
6	WT	0.7	No
18	KO	0.2	Yes
18	KO	0.1	Yes
18	WT	0.6	No
6	WT	0.8	No
6	KO	0.8	No
6	KO	0.6	No
18	KO	0.1	Yes
18	KO	0.3	Yes
18	KO	0.4	Yes

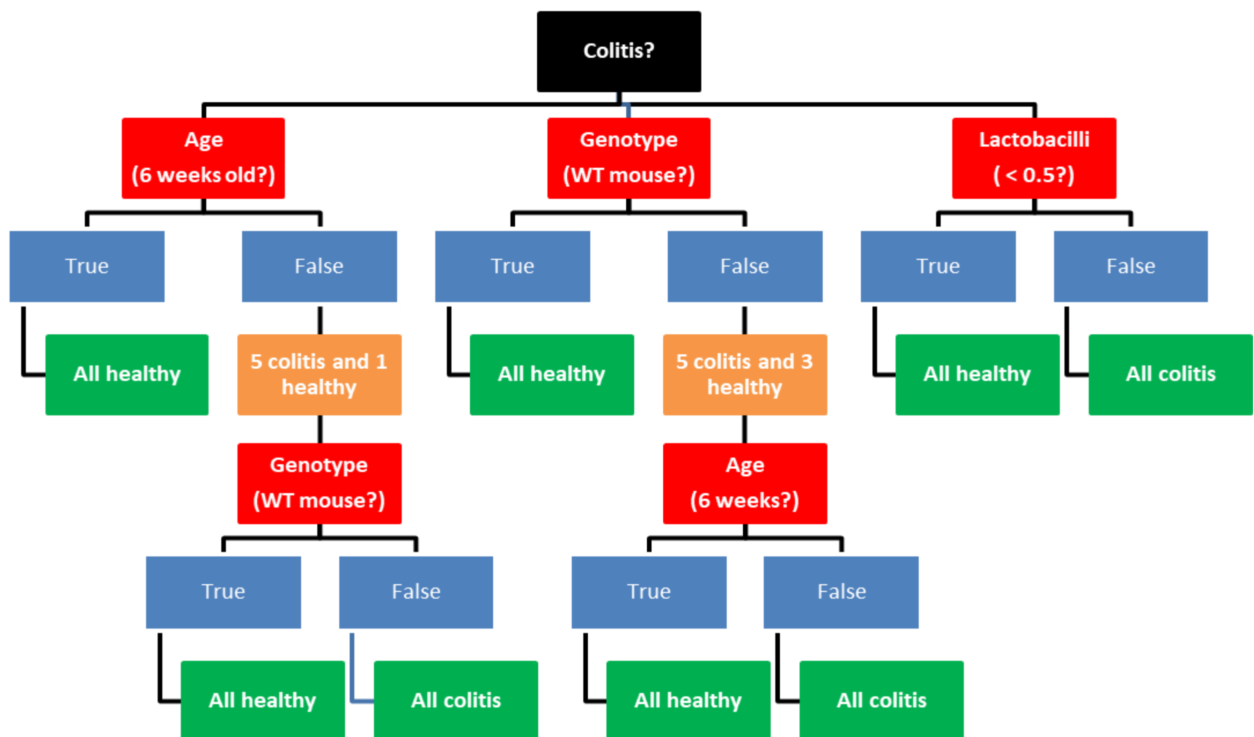


Figure 2: Schematic representation of a decision tree. The decision tree was used to determine whether the mice had colitis or not, based on the data provided. The abundance of *Lactobacilli* was the best for discriminating between healthy and colitic mice.

The RF model is ultimately an aggregated form of decision trees (Breiman 1996). The RF combines the results of the individual trees to make a consensus. The data is then randomly split into different subsets (with overlap) and has reduced bias and variance compared to decision trees alone. Thus, the random forest is less likely to 'overfit' data than using decision trees. It is also capable of handling bigger and higher dimensionality datasets. The RF is able to carry out both regression and classification tasks. Regression trees are used when the dependent variable is continuous and classification trees are used when the dependent variable is categorical. Generally, the more trees that were used in the RF, the more robust the model. The RF has several parameters that can be controlled. These include the number of trees, mtry (number of features selected when making the trees) and the sample size. Information resulting from the RF includes the out of bag (OOB) error rate, which states how accurate the RF was at discriminating between different treatment groups. There is the 'Mean Decrease Accuracy' value, which is the loss of accuracy that would be incurred when a certain feature is dropped from the RF. There is also the 'GINI index', which measures node purity when a variable is used to split groups. The RF model has been successfully applied to the microbiota in a human study of IBD, where remission was identified with 73% accuracy and active CD was identified with 78% accuracy (Tedjo et al., 2016).

1.9.9 Inferring function

So far, 16S rRNA analysis of the microbiome has been discussed. Although it is a popular and widely used analytical tool, it does not necessarily provide functional information on the bacterial community, although this can be inferred (Langille et al., 2013). Studies are beginning to move away from simply characterising bacteria within a given environment, to putting community structure into the context of transcriptomic, proteomic and metabolic profiles of the resident bacteria. In order to do this effectively, techniques such as whole-metagenome sequencing (WMS) are employed (Morgan and Huttenhower, 2014). In contrast to 16S rRNA sequencing, WMS involves sequencing all genetic material within the sample, as opposed to just a single gene used for taxonomic identification (Lim et al., 2014). This includes all microbial kingdoms, as well as increased resolution at a species and strain level (Vetrovsky and Baldrian, 2013). By contrast, 16S rRNA sequences can sometimes fail to differentiate between species, such as *C. bartletti* and *C. difficile* (Morgan and Huttenhower, 2014). This is an important distinction to make, because *C. bartletti* is comparatively benign in comparison to *C. difficile* which can be pathogenic (Rupnik et al., 2009). The usage of WMS is not always advised, as WMS on mucosa can result in extensive contamination with

host DNA (Morgan and Huttenhower, 2014). Despite this, WMS facilitates better studies with regards to functional inference in terms of metatranscriptomics (mRNA), metaproteomics (proteins) and metametabolomics (metabolites) by comparing sequence data to various reference databases such as Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000). This is still a major computational challenge and one that is still relatively expensive.

An alternative method is to make functional inferences using 16S rRNA sequence data. This is through the use of Phylogenetic Investigation of Communities by Reconstruction of Unobserved States PICRUSt (Langille et al., 2013). The overall idea is that although 16S rRNA focuses specifically on taxonomic identification, assumptions in function can be made by comparing the sequenced taxa against any cultured relative. For example, Langille et al. (2013) discuss how *Bacteroides* identified by 16S rRNA sequencing may have a similar gene expression profiles when compared against cultured strains of *Bacteroides.* Hence, PICRUSt compares 16S rRNA data against a reference genome database to make inferences into a predicted metagenome (Langille et al., 2013). An excellent illustration of this method was performed by McHardy et al. (2013), who analysed bacteria isolated from colonic mucus. They analysed both 16S rRNA datasets using aforementioned methods to define taxonomic identification and metabolic pathways, and mass spectrometry to analyse metabolites. Interestingly, their ability to make metagenomic inferences was relatively successful at identifying metabolites assayed via mass spectrometry. Additionally, they found a strong association between the presence of metabolites and community structure. This once again highlights that metabolites, alongside other microbial factors, can be used as possible biomarkers of community structure (McHardy et al., 2013). Indeed, the functional repertoire of the microbiota is considerably more stable between individuals than the exact bacteria present (Consortium., 2012). Therefore, investigating not only what species are present but also the factors that they produce is vital to understanding how the microbiome contributes to host health.

1.10 Hypothesis, aims and objectives

The microbiota exists within multiple niches within the host. However, many studies have focused almost exclusively on bacteria found within the stool (Arumugam et al., 2011; Wu et al., 2011b) which cannot fully represent the diversity of the microbiota in the GI tract and

specifically the mucus-resident bacteria. I hypothesised that host-health is associated with characteristic changes in gut bacterial communities, with the mucus-resident bacteria having the most profound impact on the host. This hypothesis will be investigated using the following aims:

Aim 1: Development of a data driven methodology to define important characteristics of the microbiota in disease

A major issue hindering microbiome research is that current analytical methods rely on *a priori* choices, for instance in similarity thresholds and external data sources such as taxonomies containing unknown biases. In order to explore my overall hypothesis further, I aimed to develop methodologies to explore the microbiota of the mucus and stool. Using microbiome data derived from the stools and mucus of littermate control *mdr1a*$^{-/-}$ mice, mice that spontaneously develop colitis over time, I used traditional methods to see whether the gut microbiome could be distinguished by treatment group: 6 and 18 week old mice, stool and mucus samples, wildtype (WT) and *mdr1a*$^{-/-}$ samples and social group (cage). I then used a machine-learning technique, Random Forest, to see whether the microbiome could be similarly associated to the different treatment groups. As bacteria work as communities, I identified what bacterial clades were most important in making these associations. I quantified the characteristics that allowed these clades to be associated with the conditions of interest with a hypothesis that abundant taxa within a treatment group, of particular taxonomic levels, would have the most importance. This aim was addressed in Chapter 2, utilising previous sequence data (Glymenaki et al., 2017).

Aim 2: Determining the role of mucus-resident metabolites in disease

In addition to considering the bacteria present within a community, it is also important to consider their functional profiles. Metabolites that the microbiota produce are vital to host health and several studies have tried to associate bacteria with their metabolomic profiles (Wikoff et al., 2009; McHardy et al., 2013). Metabolites in serum, urine and stool are typically investigated but the mucus could be a potential site of important metabolites from both the host and the microbiota. The mucus has a distinct population of bacteria and due to their closer proximity to host cells, are perhaps more likely to mediate functional effects on the host than stool bacteria. Given that there are changes in the mucus microbiota of *mdr1a*$^{-/-}$ mice that precede the onset of colitis, I investigated the metabolic profile within the mucus to

explore the hypothesis that the metabolites would also be significantly altered before the onset of colitis (Chapter 3).

Aim 3: Characterisation of the stool and mucus microbiota in a ∆dbl*GATA-1*$^{-/-}$ mouse model

My overall hypothesis aimed to address whether host-health was associated with changes in the gut microbiota, especially with regards to the mucus. To further explore the role of the mucus microbiota in homeostasis, I used a mouse model with a disrupted immune function. The mouse lacked eosinophils which are thought to play a role in the regulation of gut homeostasis and barrier function. Eosinophils have been implicated in maintaining epithelial gut barrier integrity and are thought to influence the survival of plasma cells that produce IgA, a critical antibody in regulating bacterial homeostasis in the gut. Thus, eosinophil-deficient *(*∆*dblGATA-1*$^{-/-}$) mice are likely to have disrupted mucosal immunity with an impact on the gut microbiota, that I hypothesised would be most profound in the mucus microbial communities in which the greatest density of IgA is associated (Chapter 4).

1.11 References

Acharya, K. R. & Ackerman, S. J. (2014). Eosinophil granule proteins: form and function. *J Biol Chem,* 289(25)**,** 17406-15.

Ackermann, G., Thomalla, S., Ackermann, F., Schaumann, R., Rodloff, A. C. & Ruf, B. R. (2005). Prevalence and characteristics of bacteria and host factors in an outbreak situation of antibiotic-associated diarrhoea. *J Med Microbiol,* 54**,** 149-53.

Ahmed, S., Macfarlane, G. T., Fite, A., McBain, A. J., Gilbert, P. & Macfarlane, S. (2007). Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol,* 73**,** 7435-42.

Al-Drees, A. & Khalil, M. S. (2016). Histological and immunohistochemical effects of L-arginine and silymarin on TNBS-induced inflammatory bowel disease in rats. *Histol Histopathol,* 31(11)**,** 1259-70.

Alexander, J. W. & Supp, D. M. (2014). Role of Arginine and Omega-3 Fatty Acids in Wound Healing and Infection. *Adv Wound Care (New Rochelle),* 3(11)**,** 682-690.

Alipour, M., Zaidi, D., Valcheva, R., Jovel, J., Martinez, I., Sergi, C., Walter, J., Mason, A. L., Wong, G. K., Dieleman, L. A., Carroll, M. W., Huynh, H. Q. & Wine, E. (2016). Mucosal Barrier Depletion and Loss of Bacterial Diversity are Primary Abnormalities in Paediatric Ulcerative Colitis. *J Crohns Colitis,* 10(4)**,** 462-71.

Ardizzone, S., Maconi, G., Bianchi, V., Russo, A., Colombo, E., Cassinotti, A., Penati, C., Tenchini, M. L. & Bianchi Porro, G. (2007). Multidrug resistance 1 gene polymorphism and susceptibility to inflammatory bowel disease. *Inflamm Bowel Dis,* 13(5)**,** 516-23.

Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., deRoos, P., Liu, H., Cross, J. R., Pfeffer, K., Coffer, P. J. & Rudensky, A. Y. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature,* 504**,** 451-5.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Dore, J., Antolin, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K. U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S. D. & Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature,* 473**,** 174-80.

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., Taniguchi, T., Takeda, K., Hori, S., Ivanov, II, Umesaki, Y., Itoh, K. & Honda, K. (2011). Induction of colonic regulatory T cells by indigenous Clostridium species. *Science,* 331(6015)**,** 337-41.

Autran, C. A., Kellman, B. P., Kim, J. H., Asztalos, E., Blood, A. B., Spence, E. C. H., Patel, A. L., Hou, J., Lewis, N. E. & Bode, L. (2018). Human milk oligosaccharide composition predicts risk of necrotising enterocolitis in preterm infants. *Gut,* 67(6)**,** 1064-1070.

Aw, W. & Fukuda, S. (2014). Toward the comprehensive understanding of the gut ecosystem via metabolomics-based integrated omics approach. *Semin Immunopathol*.

Bajaj, J. S., Hylemon, P. B., Ridlon, J. M., Heuman, D. M., Daita, K., White, M. B., Monteith, P., Noble, N. A., Sikaroodi, M. & Gillevet, P. M. (2012). Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. *Am J Physiol Gastrointest Liver Physiol,* 303(6)**,** G675-85.

Balasubramanian, K., Kumar, S., Singh, R. R., Sharma, U., Ahuja, V., Makharia, G. K. & Jagannathan, N. R. (2009). Metabolism of the colonic mucosa in patients with inflammatory bowel diseases: an in vitro proton magnetic resonance spectroscopy study. *Magn Reson Imaging,* 27(1)**,** 79-86.

Bao, X., Feng, Z., Yao, J., Li, T. & Yin, Y. (2017). Roles of Dietary Amino Acids and Their Metabolites in Pathogenesis of Inflammatory Bowel Disease. *Mediators Inflamm,* 2017**,** 6869259.

Baumgart, M., Dogan, B., Rishniw, M., Weitzman, G., Bosworth, B., Yantiss, R., Orsi, R. H., Wiedmann, M., McDonough, P., Kim, S. G., Berg, D., Schukken, Y., Scherl, E. & Simpson, K. W. (2007). Culture independent analysis of ileal mucosa reveals a selective increase in invasive Escherichia coli of novel phylogeny relative to depletion of Clostridiales in Crohn's disease involving the ileum. *Isme j,* 1(5)**,** 403-18.

Begley, P., Francis-McIntyre, S., Dunn, W. B., Broadhurst, D. I., Halsall, A., Tseng, A., Knowles, J., Goodacre, R. & Kell, D. B. (2009). Development and performance of a gas chromatography-time-of-flight mass spectrometry analysis for large-scale nontargeted metabolomic studies of human serum. *Anal Chem,* 81(16)**,** 7038-46.

Bengmark, S. (2013). Gut microbiota, immune development and function. *Pharmacol Res,* 69**,** 87-113.

Bik, E. M., Eckburg, P. B., Gill, S. R., Nelson, K. E., Purdom, E. A., Francois, F., Perez-Perez, G., Blaser, M. J. & Relman, D. A. (2006). Molecular analysis of the bacterial microbiota in the human stomach. *Proc Natl Acad Sci U S A,* 103(3)**,** 732-7.

Binns, C., Lee, M. & Low, W. Y. (2016). The Long-Term Public Health Benefits of Breastfeeding. *Asia Pac J Public Health,* 28(1)**,** 7-14.

Blaser, M. J. & Dominguez-Bello, M. G. (2016). The Human Microbiome before Birth. *Cell Host Microbe,* 20(5)**,** 558-560.

Bloemen, J. G., Venema, K., van de Poll, M. C., Olde Damink, S. W., Buurman, W. A. & Dejong, C. H. (2009). Short chain fatty acids exchange across the gut and liver in humans measured at surgery. *Clin Nutr,* 28(6)**,** 657-61.

Bowcutt, R., Forman, R., Glymenaki, M., Carding, S. R., Else, K. J. & Cruickshank, S. M. (2014). Heterogeneity across the murine small and large intestine. *World J Gastroenterol,* 20**,** 15216-15232.

Bramhall, M., Florez-Vargas, O., Stevens, R., Brass, A. & Cruickshank, S. (2015). Quality of methods reporting in animal models of colitis. *Inflamm Bowel Dis,* 21(6)**,** 1248-59.

Braniste, V., Al-Asmakh, M., Kowal, C., Anuar, F., Abbaspour, A., Toth, M., Korecka, A., Bakocevic, N., Ng, L. G., Kundu, P., Gulyas, B., Halldin, C., Hultenby, K., Nilsson, H., Hebert, H., Volpe, B. T., Diamond, B. & Pettersson, S. (2014). The gut microbiota influences blood-brain barrier permeability in mice. *Sci Transl Med,* 6(263)**,** 263ra158.

Brestoff, J. R. & Artis, D. (2013). Commensal bacteria at the interface of host metabolism and the immune system. *Nat Immunol,* 14(7)**,** 676-84.

Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., Muir, A. I., Wigglesworth, M. J., Kinghorn, I., Fraser, N. J., Pike, N. B., Strum, J. C., Steplewski, K. M., Murdock, P. R., Holder, J. C., Marshall, F. H., Szekeres, P. G., Wilson, S., Ignar, D. M., Foord, S. M., Wise, A. & Dowell, S. J. (2003). The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J Biol Chem,* 278(13)**,** 11312-9.

Cabrera-Rubio, R., Collado, M. C., Laitinen, K., Salminen, S., Isolauri, E. & Mira, A. (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *Am J Clin Nutr,* 96**,** 544-51.

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(5)**,** 335-6.

Cariani, G., Vandelli, A., Fontana, G., Bonora, G. & Mazzoleni, G. (1992). Autoimmune gastritis: is Helicobacter pylori a merely commensal or a pathogenic agent? *Dig Dis Sci,* 37**,** 1304-5.

Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. (2006). Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science,* 313(5790)**,** 1126-30.

Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods,* 69(2)**,** 330-9.

Chang, S. K., Dohrman, A. F., Basbaum, C. B., Ho, S. B., Tsuda, T., Toribara, N. W., Gum, J. R. & Kim, Y. S. (1994). Localization of mucin (MUC2 and MUC3) messenger RNA and peptide expression in human normal intestine and colon cancer. *Gastroenterology,* 107(1)**,** 28-36.

Chassaing, B., Aitken, J. D., Malleshappa, M. & Vijay-Kumar, M. (2014). Dextran sulfate sodium (DSS)-induced colitis in mice. *Curr Protoc Immunol,* 104**,** Unit 15.25.

Chassaing, B., Miles-Brown, J., Pellizzon, M., Ulman, E., Ricci, M., Zhang, L., Patterson, A. D., Vijay-Kumar, M. & Gewirtz, A. T. (2015). Lack of soluble fiber drives diet-induced adiposity in mice. *Am J Physiol Gastrointest Liver Physiol,* 309(7)**,** G528-41.

Chaudhary, P. M., Mechetner, E. B. & Roninson, I. B. (1992). Expression and activity of the multidrug resistance P-glycoprotein in human peripheral blood lymphocytes. *Blood,* 80(11)**,** 2735-9.

Chen, C. J., Chin, J. E., Ueda, K., Clark, D. P., Pastan, I., Gottesman, M. M. & Roninson, I. B. (1986). Internal duplication and homology with bacterial transport proteins in the mdr1 (P-glycoprotein) gene from multidrug-resistant human cells. *Cell,* 47(3)**,** 381-9.

Chiang, J. Y. (1998). Regulation of bile acid synthesis. *Front Biosci,* 3**,** d176-93.

Chiang, J. Y. (2013). Bile acid metabolism and signaling. *Compr Physiol,* 3(3)**,** 1191-212.

Cho, S. S., Qi, L., Fahey, G. C., Jr. & Klurfeld, D. M. (2013). Consumption of cereal fiber, mixtures of whole grains and bran, and whole grains and risk reduction in type 2 diabetes, obesity, and cardiovascular disease. *Am J Clin Nutr,* 98(2)**,** 594-619.

Chu, V. T., Beller, A., Rausch, S., Strandmark, J., Zanker, M., Arbach, O., Kruglov, A. & Berek, C. (2014). Eosinophils promote generation and maintenance of immunoglobulin-A-expressing plasma cells and contribute to gut immune homeostasis. *Immunity,* 40**,** 582-93.

Claesson, M. J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., Marchesi, J. R., Falush, D., Dinan, T., Fitzgerald, G., Stanton, C., van Sinderen, D., O'Connor, M., Harnedy, N., O'Connor, K., Henry, C., O'Mahony, D., Fitzgerald, A. P., Shanahan, F., Twomey, C., Hill, C., Ross, R. P. & O'Toole, P. W. (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A,* 108 Suppl 1**,** 4586-91.

Claesson, M. J., Jeffery, I. B., Conde, S., Power, S. E., O'Connor, E. M., Cusack, S., Harris, H. M., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., Fitzgerald, G. F., Deane, J., O'Connor, M., Harnedy, N., O'Connor, K., O'Mahony, D., van Sinderen, D., Wallace, M., Brennan, L., Stanton, C., Marchesi, J. R., Fitzgerald, A. P., Shanahan, F., Hill, C., Ross, R. P. & O'Toole, P. W. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature,* 488**,** 178-84.

Collado, M. C., Delgado, S., Maldonado, A. & Rodriguez, J. M. (2009). Assessment of the bacterial diversity of breast milk of healthy women by quantitative real-time PCR. *Lett Appl Microbiol,* 48(5)**,** 523-8.

Collett, A., Higgs, N. B., Gironella, M., Zeef, L. A., Hayes, A., Salmo, E., Haboubi, N., Iovanna, J. L., Carlson, G. L. & Warhurst, G. (2008). Early molecular and functional changes in colonic epithelium that precede increased gut permeability during colitis development in mdr1a(-/-) mice. *Inflamm Bowel Dis,* 14(5)**,** 620-31.

Consortium., H. M. P. (2012). Structure, function and diversity of the healthy human microbiome. *Nature,* 486**,** 207-14.

Cui, H. & Zhang, X. (2013). Alignment-free supervised classification of metagenomes by recursive SVM. *BMC Genomics,* 14**,** 641.

Dalton, J. E., Cruickshank, S. M., Egan, C. E., Mears, R., Newton, D. J., Andrew, E. M., Lawrence, B., Howell, G., Else, K. J., Gubbels, M. J., Striepen, B., Smith, J. E., White, S. J. & Carding, S. R. (2006). Intraepithelial gammadelta+ lymphocytes maintain the integrity of intestinal epithelial tight junctions in response to infection. *Gastroenterology,* 131(3)**,** 818-29.

Dardas, M., Gill, S. R., Grier, A., Pryhuber, G. S., Gill, A. L., Lee, Y. H. & Guillet, R. (2014). The impact of postnatal antibiotics on the preterm intestinal microbiome. *Pediatr Res,* 76**,** 150-8.

Davis, C. P. & Savage, D. C. (1974). Habitat, succession, attachment, and morphology of segmented, filamentous microbes indigenous to the murine gastrointestinal tract. *Infect Immun,* 10(4)**,** 948-56.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J. B., Massart, S., Collini, S., Pieraccini, G. & Lionetti, P. (2010). Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A,* 107**,** 14691-6.

De La Cochetiere, M. F., Durand, T., Lepage, P., Bourreille, A., Galmiche, J. P. & Dore, J. (2005). Resilience of the dominant human fecal microbiota upon short-course antibiotic challenge. *J Clin Microbiol,* 43**,** 5588-92.

De Preter, V., Machiels, K., Joossens, M., Arijs, I., Matthys, C., Vermeire, S., Rutgeerts, P. & Verbeke, K. (2015). Faecal metabolite profiling identifies medium-chain fatty acids as discriminating compounds in IBD. *Gut,* 64(3)**,** 447-58.

De Preter, V. & Verbeke, K. (2013). Metabolomics as a diagnostic tool in gastroenterology. *World J Gastrointest Pharmacol Ther,* 4(4)**,** 97-107.

De Vadder, F., Kovatcheva-Datchary, P., Goncalves, D., Vinera, J., Zitoun, C., Duchampt, A., Backhed, F. & Mithieux, G. (2014). Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell,* 156(1-2)**,** 84-96.

Delgado, S., Cabrera-Rubio, R., Mira, A., Suarez, A. & Mayo, B. (2013). Microbiological survey of the human gastric ecosystem using culturing and pyrosequencing methods. *Microb Ecol,* 65(3)**,** 763-72.

DeLong, E. F., Preston, C. M., Mincer, T., Rich, V., Hallam, S. J., Frigaard, N. U., Martinez, A., Sullivan, M. B., Edwards, R., Brito, B. R., Chisholm, S. W. & Karl, D. M. (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science,* 311**,** 496-503.

den Besten, G., Bleeker, A., Gerding, A., van Eunen, K., Havinga, R., van Dijk, T. H., Oosterveer, M. H., Jonker, J. W., Groen, A. K., Reijngoud, D. J. & Bakker, B. M. (2015). Short-Chain Fatty Acids Protect Against High-Fat Diet-Induced Obesity via a PPARgamma-Dependent Switch From Lipogenesis to Fat Oxidation. *Diabetes,* 64(7)**,** 2398-408.

Derrien, M., Vaughan, E. E., Plugge, C. M. & de Vos, W. M. (2004). Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol,* 54(Pt 5)**,** 1469-76.

Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Nunez, G. & Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell,* 167(5)**,** 1339-1353.e21.

Di Pilato, V., Freschi, G., Ringressi, M. N., Pallecchi, L., Rossolini, G. M. & Bechi, P. (2016). The esophageal microbiota in health and disease. *Ann N Y Acad Sci,* 1381(1)**,** 21-33.

Dietert, R. R. (2014). The microbiome in early life: self-completion and microbiota protection as health priorities. *Birth Defects Res B Dev Reprod Toxicol,* 101**,** 333-40.

Dominguez-Bello, M. G., Costello, E. K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N. & 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(26)**,** 11971-5.

Dominguez-Bello, M. G., De Jesus-Laboy, K. M., Shen, N., Cox, L. M., Amir, A., Gonzalez, A., Bokulich, N. A., Song, S. J., Hoashi, M., Rivera-Vinas, J. I., Mendez, K., Knight, R. & Clemente, J. C. (2016). Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. *Nat Med,* 22(3)**,** 250-3.

Donnarumma, G., Paoletti, I., Fusco, A., Perfetto, B., Buommino, E., de Gregorio, V. & Baroni, A. (2016). beta-Defensins: Work in Progress. *Adv Exp Med Biol,* 901**,** 59-76.

Donohoe, D. R., Garge, N., Zhang, X., Sun, W., O'Connell, T. M., Bunger, M. K. & Bultman, S. J. (2011). The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab,* 13(5)**,** 517-26.

Dove, W. F., Clipson, L., Gould, K. A., Luongo, C., Marshall, D. J., Moser, A. R., Newton, M. A. & Jacoby, R. F. (1997). Intestinal neoplasia in the ApcMin mouse: independence from the microbial and natural killer (beige locus) status. *Cancer Res,* 57**,** 812-4.

Droste, J. H., Wieringa, M. H., Weyler, J. J., Nelen, V. J., Vermeire, P. A. & Van Bever, H. P. (2000). Does the use of antibiotics in early childhood increase the risk of asthma and allergic disease? *Clin Exp Allergy,* 30**,** 1547-53.

Dunn, W. B., Bailey, N. J. & Johnson, H. E. (2005). Measuring the metabolome: current analytical technologies. *Analyst,* 130(5)**,** 606-25.

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B. & Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc,* 6(7)**,** 1060-83.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science,* 308(5728)**,** 1635-8.

Ek, W. E., D'Amato, M. & Halfvarson, J. (2014). The history of genetics in inflammatory bowel disease. *Ann Gastroenterol,* 27(4)**,** 294-303.

El Aidy, S., van den Bogert, B. & Kleerebezem, M. (2015). The small intestine microbiota, nutritional modulation and relevance for health. *Curr Opin Biotechnol,* 32**,** 14-20.

Elderman, M., Hugenholtz, F., Belzer, C., Boekschoten, M., van Beek, A., de Haan, B., Savelkoul, H., de Vos, P. & Faas, M. (2018). Sex and strain dependent differences in mucosal immunology and microbiota composition in mice. *Biol Sex Differ,* 9(1)**,** 26.

Ellis, D. I., Dunn, W. B., Griffin, J. L., Allwood, J. W. & Goodacre, R. (2007). Metabolic fingerprinting as a diagnostic tool. *Pharmacogenomics,* 8(9)**,** 1243-66.

Escherich, T. (1988). The intestinal bacteria of the neonate and breast-fed infant. 1884. *Rev Infect Dis,* 10(6)**,** 1220-5.

Ettreiki, C., Gadonna-Widehem, P., Mangin, I., Coëffier, M., Delayre-Orthez, C. & Anton, P. M. (2012). Juvenile ferric iron prevents microbiota dysbiosis and colitis in adult rodents. *World J Gastroenterol,* 18**,** 2619-29.

Fadlallah, J., El Kafsi, H., Sterlin, D., Juste, C., Parizot, C., Dorgham, K., Autaa, G., Gouas, D., Almeida, M., Lepage, P., Pons, N., Le Chatelier, E., Levenez, F., Kennedy, S., Galleron, N., de Barros, J. P., Malphettes, M., Galicier, L., Boutboul, D., Mathian, A., Miyara, M., Oksenhendler, E., Amoura, Z., Dore, J., Fieschi, C., Ehrlich, S. D., Larsen, M. & Gorochov, G. (2018). Microbial ecology perturbation in human IgA deficiency. *Sci Transl Med,* 10(439).

Fallingborg, J. (1999). Intraluminal pH of the human gastrointestinal tract. *Dan Med Bull,* 46**,** 183-96.

Ferreira-Halder, C. V., Faria, A. V. S. & Andrade, S. S. (2017). Action and function of Faecalibacterium prausnitzii in health and disease. *Best Pract Res Clin Gastroenterol,* 31(6)**,** 643-648.

Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D. T., Manara, S., Zolfo, M., Beghini, F., Bertorelli, R., De Sanctis, V., Bariletti, I., Canto, R., Clementi, R., Cologna, M., Crifo, T., Cusumano, G., Gottardi, S., Innamorati, C., Mase, C., Postai, D., Savoi, D., Duranti, S., Lugli, G. A., Mancabelli, L., Turroni, F., Ferrario, C., Milani, C., Mangifesta, M., Anzalone, R., Viappiani, A., Yassour, M., Vlamakis, H., Xavier, R., Collado, C. M., Koren, O., Tateo, S., Soffiati, M., Pedrotti, A., Ventura, M., Huttenhower, C., Bork, P. & Segata, N. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host Microbe,* 24(1)**,** 133-145.e5.

Fillon, S. A., Harris, J. K., Wagner, B. D., Kelly, C. J., Stevens, M. J., Moore, W., Fang, R., Schroeder, S., Masterson, J. C., Robertson, C. E., Pace, N. R., Ackerman, S. J. & Furuta, G. T. (2012). Novel device to sample the esophageal microbiome--the esophageal string test. *PLoS One,* 7(9)**,** e42938.

Flint, H. J., Duncan, S. H., Scott, K. P. & Louis, P. (2014). Links between diet, gut microbiota composition and gut metabolism. *Proc Nutr Soc***,** 1-10.

Florez-Vargas, O., Brass, A., Karystianis, G., Bramhall, M., Stevens, R., Cruickshank, S. & Nenadic, G. (2016). Bias in the reporting of sex and age in biomedical research on mouse models. *Elife,* 5.

Forman, R., Bramhall, M., Logunova, L., Svensson-Frej, M., Cruickshank, S. M. & Else, K. J. (2016). Eosinophils may play regionally disparate roles in influencing IgA(+) plasma cell numbers during large and small intestinal inflammation. *BMC Immunol,* 17(1)**,** 12.

Foster, J. A., Rinaman, L. & Cryan, J. F. (2017). Stress & the gut-brain axis: Regulation by the microbiome. *Neurobiol Stress,* 7**,** 124-136.

Fransen, F., van Beek, A. A., Borghuis, T., Meijer, B., Hugenholtz, F., van der Gaast-de Jongh, C., Savelkoul, H. F., de Jonge, M. I., Faas, M. M., Boekschoten, M. V., Smidt, H., El Aidy, S. & de Vos, P. (2017). The Impact of Gut Microbiota on Gender-Specific Differences in Immunity. *Front Immunol,* 8**,** 754.

Fusco, A., Savio, V., Cammarota, M., Alfano, A., Schiraldi, C. & Donnarumma, G. (2017). Beta-Defensin-2 and Beta-Defensin-3 Reduce Intestinal Damage Caused by Salmonella typhimurium Modulating the Expression of Cytokines and Enhancing the Probiotic Activity of Enterococcus faecium. *J Immunol Res,* 2017**,** 6976935.

Gagliardi, D., Makihara, S., Corsi, P. R., Viana Ade, T., Wiczer, M. V., Nakakubo, S. & Mimica, L. M. (1998). Microbial flora of the normal esophagus. *Dis Esophagus,* 11(4)**,** 248-50.

Garland, C. D., Lee, A. & Dickson, M. R. (1982). Segmented filamentous bacteria in the rodent small intestine: Their colonization of growing animals and possible role in host resistance toSalmonella. *Microb Ecol,* 8(2)**,** 181-90.

Geboes, K. & Dalle, I. (2002). Influence of treatment on morphological features of mucosal inflammation. *Gut,* 50 Suppl 3**,** Iii37-42.

Geirnaert, A., Calatayud, M., Grootaert, C., Laukens, D., Devriese, S., Smagghe, G., De Vos, M., Boon, N. & Van de Wiele, T. (2017). Butyrate-producing bacteria supplemented in vitro to Crohn's disease patient microbiota increased butyrate production and enhanced intestinal epithelial barrier integrity. *Sci Rep,* 7(1)**,** 11450.

Gerbe, F., van Es, J. H., Makrini, L., Brulin, B., Mellitzer, G., Robine, S., Romagnolo, B., Shroyer, N. F., Bourgaux, J. F., Pignodel, C., Clevers, H. & Jay, P. (2011). Distinct ATOH1 and Neurog3 requirements define tuft cells as a new secretory cell type in the intestinal epithelium. *J Cell Biol,* 192(5)**,** 767-80.

Gevers, D., Kugathasan, S., Denson, L. A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S. J., Yassour, M., Morgan, X. C., Kostic, A. D., Luo, C., Gonzalez, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., Stephens, M., Heyman, M., Markowitz, J., Baldassano, R., Griffiths, A., Sylvester, F., Mack, D., Kim, S., Crandall, W., Hyams, J., Huttenhower, C., Knight, R. & Xavier, R. J. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe,* 15(3)**,** 382-392.

Ghosh, D., Porter, E., Shen, B., Lee, S. K., Wilk, D., Drazba, J., Yadav, S. P., Crabb, J. W., Ganz, T. & Bevins, C. L. (2002). Paneth cell trypsin is the processing enzyme for human defensin-5. *Nat Immunol,* 3(6)**,** 583-90.

Gill, S. R., Pop, M., Deboy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M. & Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science,* 312(5778)**,** 1355-9.

Glymenaki, M., Singh, G., Brass, A., Warhurst, G., McBain, A. J., Else, K. J. & Cruickshank, S. M. (2017). Compositional Changes in the Gut Mucus Microbiota Precede the Onset of Colitis-Induced Inflammation. *Inflamm Bowel Dis,* 23(6)**,** 912-922.

Gomez de Aguero, M., Ganal-Vonarburg, S. C., Fuhrer, T., Rupp, S., Uchimura, Y., Li, H., Steinert, A., Heikenwalder, M., Hapfelmeier, S., Sauer, U., McCoy, K. D. & Macpherson, A.

J. (2016). The maternal microbiota drives early postnatal innate immune development. *Science,* 351(6279)**,** 1296-302.

Gordon, H., Trier Moller, F., Andersen, V. & Harbord, M. (2015). Heritability in inflammatory bowel disease: from the first twin study to genome-wide association studies. *Inflamm Bowel Dis,* 21(6)**,** 1428-34.

Graf, D., Di Cagno, R., Fak, F., Flint, H. J., Nyman, M., Saarela, M. & Watzl, B. (2015). Contribution of diet to the composition of the human gut microbiota. *Microb Ecol Health Dis,* 26**,** 26164.

Grice, E. A. & Segre, J. A. (2011). The skin microbiome. *Nat Rev Microbiol,* 9(4)**,** 244-53.

Griffin, N. W., Ahern, P. P., Cheng, J., Heath, A. C., Ilkayeva, O., Newgard, C. B., Fontana, L. & Gordon, J. I. (2017). Prior Dietary Practices and Connections to a Human Gut Microbial Metacommunity Alter Responses to Diet Interventions. *Cell Host Microbe,* 21(1)**,** 84-96.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methe, B., DeSantis, T. Z., Petrosino, J. F., Knight, R. & Birren, B. W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res,* 21**,** 494-504.

Haberland, K., Ackermann, J. A., Ipseiz, N., Culemann, S., Pracht, K., Englbrecht, M., Jack, H. M., Schett, G., Schuh, W. & Kronke, G. (2018). Eosinophils are not essential for maintenance of murine plasma cells in the bone marrow. *Eur J Immunol*.

Halfvarson, J. (2011). Genetics in twins with Crohn's disease: less pronounced than previously believed? *Inflamm Bowel Dis,* 17(1)**,** 6-12.

Harig, J. M., Soergel, K. H., Komorowski, R. A. & Wood, C. M. (1989). Treatment of diversion colitis with short-chain-fatty acid irrigation. *N Engl J Med,* 320(1)**,** 23-8.

Harrell, L., Wang, Y., Antonopoulos, D., Young, V., Lichtenstein, L., Huang, Y., Hanauer, S. & Chang, E. (2012). Standard colonic lavage alters the natural state of mucosal-associated microbiota in the human colon. *PLoS One,* 7**,** e32545.

Hayashi, H., Sakamoto, M. & Benno, Y. (2002). Fecal microbial diversity in a strict vegetarian as determined by molecular analysis and cultivation. *Microbiol Immunol,* 46(12)**,** 819-31.

Hayes, M. D., Ovcinnikovs, V., Smith, A. G., Kimber, I. & Dearman, R. J. (2014). The aryl hydrocarbon receptor: differential contribution to T helper 17 and T cytotoxic 17 cell development. *PLoS One,* 9(9)**,** e106955.

Hernandez, E., Bargiela, R., Diez, M. S., Friedrichs, A., Perez-Cobas, A. E., Gosalbes, M. J., Knecht, H., Martinez-Martinez, M., Seifert, J., von Bergen, M., Artacho, A., Ruiz, A., Campoy, C., Latorre, A., Ott, S. J., Moya, A., Suarez, A., Martins dos Santos, V. A. & Ferrer, M. (2013). Functional consequences of microbial shifts in the human gastrointestinal tract linked to antibiotic treatment and obesity. *Gut Microbes,* 4(4)**,** 306-15.

Hildebrand, F., Nguyen, T. L., Brinkman, B., Yunta, R. G., Cauwe, B., Vandenabeele, P., Liston, A. & Raes, J. (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol,* 14(1)**,** R4.

Hollister, E. B., Gao, C. & Versalovic, J. (2014). Compositional and functional features of the gastrointestinal microbiome and their effects on human health. *Gastroenterology,* 146**,** 1449-58.

Hoverstad, T. & Midtvedt, T. (1986). Short-chain fatty acids in germfree mice and rats. *J Nutr,* 116(9)**,** 1772-6.

Howitt, M. R., Lavoie, S., Michaud, M., Blum, A. M., Tran, S. V., Weinstock, J. V., Gallini, C. A., Redding, K., Margolskee, R. F., Osborne, L. C., Artis, D. & Garrett, W. S. (2016). Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in the gut. *Science,* 351(6279)**,** 1329-33.

Hsu, S. I., Lothstein, L. & Horwitz, S. B. (1989). Differential overexpression of three mdr gene family members in multidrug-resistant J774.2 mouse cells. Evidence that distinct P-glycoprotein precursors are encoded by unique mdr genes. *J Biol Chem,* 264(20)**,** 12053-62.

Hugenholtz, F. & de Vos, W. M. (2018). Mouse models for human intestinal microbiota research: a critical evaluation. *Cell Mol Life Sci,* 75(1)**,** 149-160.

Hughes, J. B., Hellmann, J. J., Ricketts, T. H. & Bohannan, B. J. (2001). Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl Environ Microbiol,* 67**,** 4399-406.

Hugon, P., Dufour, J. C., Colson, P., Fournier, P. E., Sallah, K. & Raoult, D. (2015). A comprehensive repertoire of prokaryotic species identified in human beings. *Lancet Infect Dis,* 15(10)**,** 1211-1219.

Imhann, F., Vich Vila, A., Bonder, M. J., Fu, J., Gevers, D., Visschedijk, M. C., Spekhorst, L. M., Alberts, R., Franke, L., van Dullemen, H. M., Ter Steege, R. W. F., Huttenhower, C., Dijkstra, G., Xavier, R. J., Festen, E. A. M., Wijmenga, C., Zhernakova, A. & Weersma, R. K. (2018). Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut,* 67(1)**,** 108-119.

Ishibashi, K., Hara, S. & Kondo, S. (2009). Aquaporin water channels in mammals. *Clin Exp Nephrol,* 13(2)**,** 107-117.

Ivanov, II, Atarashi, K., Manel, N., Brodie, E. L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K. C., Santee, C. A., Lynch, S. V., Tanoue, T., Imaoka, A., Itoh, K., Takeda, K., Umesaki, Y., Honda, K. & Littman, D. R. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell,* 139**,** 485-98.

Ivanov, II, Frutos Rde, L., Manel, N., Yoshinaga, K., Rifkin, D. B., Sartor, R. B., Finlay, B. B. & Littman, D. R. (2008). Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe,* 4**,** 337-49.

Jackson, M. A., Goodrich, J. K., Maxan, M. E., Freedberg, D. E., Abrams, J. A., Poole, A. C., Sutter, J. L., Welter, D., Ley, R. E., Bell, J. T., Spector, T. D. & Steves, C. J. (2016a). Proton pump inhibitors alter the composition of the gut microbiota. *Gut,* 65(5)**,** 749-56.

Jackson, M. A., Jeffery, I. B., Beaumont, M., Bell, J. T., Clark, A. G., Ley, R. E., O'Toole, P. W., Spector, T. D. & Steves, C. J. (2016b). Signatures of early frailty in the gut microbiota. *Genome Med,* 8(1)**,** 8.

Jacobsen, E. A., Lee, N. A. & Lee, J. J. (2014). Re-defining the Unique Roles for Eosinophils in Allergic Respiratory Inflammation. *Clin Exp Allergy,* 44(9)**,** 1119-36.

Jeffery, I. B., Claesson, M. J., O'Toole, P. W. & Shanahan, F. (2012). Categorization of the gut microbiota: enterotypes or gradients? *Nat Rev Microbiol,* 10**,** 591-2.

Jimenez, E., Marin, M. L., Martin, R., Odriozola, J. M., Olivares, M., Xaus, J., Fernandez, L. & Rodriguez, J. M. (2008). Is meconium from healthy newborns actually sterile? *Res Microbiol,* 159**,** 187-93.

Johansson, M. E. & Hansson, G. C. (2014). Is the intestinal goblet cell a major immune cell? *Cell Host Microbe,* 15(3)**,** 251-2.

Johansson, M. E., Larsson, J. M. & Hansson, G. C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc Natl Acad Sci U S A,* 108 Suppl 1**,** 4659-65.

Johansson, M. E., Phillipson, M., Petersson, J., Velcich, A., Holm, L. & Hansson, G. C. (2008). The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc Natl Acad Sci U S A,* 105**,** 15064-9.

Johnson, A. M., Costanzo, A., Gareau, M. G., Armando, A. M., Quehenberger, O., Jameson, J. M. & Olefsky, J. M. (2015). High fat diet causes depletion of intestinal eosinophils associated with intestinal permeability. *PLoS One,* 10(4)**,** e0122195.

Johnstone, R. W., Ruefli, A. A. & Smyth, M. J. (2000). Multiple physiological functions for multidrug transporter P-glycoprotein? *Trends Biochem Sci,* 25(1)**,** 1-6.

Jost, T., Lacroix, C., Braegger, C. & Chassard, C. (2013). Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches. *Br J Nutr,* 110(7)**,** 1253-62.

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., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S. L., Raychaudhuri, S., Goyette, P., Wei, Z., Abraham, C., Achkar, J. P., Ahmad, T., Amininejad, L., Ananthakrishnan, A. N., Andersen, V., Andrews, J. M., Baidoo, L., Balschun, T., Bampton, P. A., Bitton, A., Boucher, G., Brand, S., Buning, C., Cohain, A., Cichon, S., D'Amato, M., De Jong, D., Devaney, K. L., Dubinsky, M., Edwards, C., Ellinghaus, D., Ferguson, L. R., Franchimont, D., Fransen, K., Gearry, R., Georges, M., Gieger, C., Glas, J., Haritunians, T., Hart, A., Hawkey, C., Hedl, M., Hu, X., Karlsen, T. H., Kupcinskas, L., Kugathasan, S., Latiano, A., Laukens, D., Lawrance, I. C., Lees, C. W., Louis, E., Mahy, G., Mansfield, J., Morgan, A. R., Mowat, C., Newman, W., Palmieri, O., Ponsioen, C. Y., Potocnik, U., Prescott, N. J., Regueiro, M., Rotter, J. I., Russell, R. K., Sanderson, J. D., Sans, M., Satsangi, J., Schreiber, S., Simms, L. A., Sventoraityte, J., Targan, S. R., Taylor, K. D., Tremelling, M., Verspaget, H. W., De Vos, M., Wijmenga, C., Wilson, D. C., Winkelmann, J., Xavier, R. J., Zeissig, S., Zhang, B., Zhang, C. K., Zhao, H., Silverberg, M. S., Annese, V., Hakonarson, H., Brant, S. R., Radford-Smith, G., Mathew, C. G., Rioux, J. D., Schadt, E. E., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature,* 491(7422)**,** 119-24.

Kabeerdoss, J., Devi, R. S., Mary, R. R. & Ramakrishna, B. S. (2012). Faecal microbiota composition in vegetarians: comparison with omnivores in a cohort of young women in southern India. *Br J Nutr,* 108(6)**,** 953-7.

Kakiyama, G., Pandak, W. M., Gillevet, P. M., Hylemon, P. B., Heuman, D. M., Daita, K., Takei, H., Muto, A., Nittono, H., Ridlon, J. M., White, M. B., Noble, N. A., Monteith, P., Fuchs, M., Thacker, L. R., Sikaroodi, M. & Bajaj, J. S. (2013). Modulation of the fecal bile acid profile by gut microbiota in cirrhosis. *J Hepatol,* 58(5)**,** 949-55.

Kalliomaki, M., Collado, M. C., Salminen, S. & Isolauri, E. (2008). Early differences in fecal microbiota composition in children may predict overweight. *Am J Clin Nutr,* 87(3)**,** 534-8.

Kanehisa, M. & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res,* 28**,** 27-30.

Kang, M. J., Kim, H. G., Kim, J. S., Oh do, G., Um, Y. J., Seo, C. S., Han, J. W., Cho, H. J., Kim, G. H., Jeong, T. C. & Jeong, H. G. (2013). The effect of gut microbiota on drug metabolism. *Expert Opin Drug Metab Toxicol,* 9**,** 1295-308.

Kell, D. B., Brown, M., Davey, H. M., Dunn, W. B., Spasic, I. & Oliver, S. G. (2005). Metabolic footprinting and systems biology: the medium is the message. *Nat Rev Microbiol,* 3(7)**,** 557-65.

Kiesler, P., Fuss, I. J. & Strober, W. (2015). Experimental Models of Inflammatory Bowel Diseases. *Cell Mol Gastroenterol Hepatol,* 1(2)**,** 154-170.

Kim, D. H. & Cheon, J. H. (2017). Pathogenesis of Inflammatory Bowel Disease and Recent Advances in Biologic Therapies. *Immune Netw,* 17(1)**,** 25-40.

Kindon, H., Pothoulakis, C., Thim, L., Lynch-Devaney, K. & Podolsky, D. K. (1995). Trefoil peptide protection of intestinal epithelial barrier function: cooperative interaction with mucin glycoprotein. *Gastroenterology,* 109**,** 516-23.

Kiss, E. A., Vonarbourg, C., Kopfmann, S., Hobeika, E., Finke, D., Esser, C. & Diefenbach, A. (2011). Natural aryl hydrocarbon receptor ligands control organogenesis of intestinal lymphoid follicles. *Science,* 334(6062)**,** 1561-5.

Klaasen, H. L., Koopman, J. P., Van den Brink, M. E., Bakker, M. H., Poelma, F. G. & Beynen, A. C. (1993). Intestinal, segmented, filamentous bacteria in a wide range of vertebrate species. *Lab Anim,* 27(2)**,** 141-50.

Klion, A. D. & Nutman, T. B. (2004). The role of eosinophils in host defense against helminth parasites. *J Allergy Clin Immunol,* 113(1)**,** 30-7.

Knights, D., Ward, T. L., McKinlay, C. E., Miller, H., Gonzalez, A., McDonald, D. & Knight, R. (2014). Rethinking "enterotypes". *Cell Host Microbe,* 16**,** 433-7.

Knoop, K. A. & Newberry, R. D. (2018). Goblet cells: multifaceted players in immunity at mucosal surfaces. *Mucosal Immunol.*

Konkel, L. (2013). The environment within: exploring the role of the gut microbiome in health and disease. *Environ Health Perspect,* 121**,** A276-81.

Kopf, M., Brombacher, F., Hodgkin, P. D., Ramsay, A. J., Milbourne, E. A., Dai, W. J., Ovington, K. S., Behm, C. A., Kohler, G., Young, I. G. & Matthaei, K. I. (1996). IL-5-deficient mice have a developmental defect in CD5+ B-1 cells and lack eosinophilia but have normal antibody and cytotoxic T cell responses. *Immunity,* 4(1)**,** 15-24.

Kostic, A. D., Xavier, R. J. & Gevers, D. (2014). The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology,* 146**,** 1489-99.

Krishnan, S., Ding, Y., Saedi, N., Choi, M., Sridharan, G. V., Sherr, D. H., Yarmush, M. L., Alaniz, R. C., Jayaraman, A. & Lee, K. (2018). Gut Microbiota-Derived Tryptophan Metabolites Modulate Inflammatory Response in Hepatocytes and Macrophages. *Cell Rep,* 23(4)**,** 1099-1111.

Kuczynski, J., Lauber, C. L., Walters, W. A., Parfrey, L. W., Clemente, J. C., Gevers, D. & Knight, R. (2012). Experimental and analytical tools for studying the human microbiome. *Nat Rev Genet,* 13**,** 47-58.

Langdon, A., Crook, N. & Dantas, G. (2016). The effects of antibiotics on the microbiome throughout development and alternative approaches for therapeutic modulation. *Genome Med,* 8(1)**,** 39.

Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Clemente, J. C., Burkepile, D. E., Vega Thurber, R. L., Knight, R., Beiko, R. G. & Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol,* 31**,** 814-21.

Layden, B. T., Angueira, A. R., Brodsky, M., Durai, V. & Lowe, W. L., Jr. (2013). Short chain fatty acids and their receptors: new metabolic targets. *Transl Res,* 161(3)**,** 131-40.

Lee, J. J., Dimina, D., Macias, M. P., Ochkur, S. I., McGarry, M. P., O'Neill, K. R., Protheroe, C., Pero, R., Nguyen, T., Cormier, S. A., Lenkiewicz, E., Colbert, D., Rinaldi, L., Ackerman, S. J., Irvin, C. G. & Lee, N. A. (2004). Defining a link with asthma in mice congenitally deficient in eosinophils. *Science,* 305(5691)**,** 1773-6.

Lee, S. H. (2015). Intestinal permeability regulation by tight junction: implication on inflammatory bowel diseases. *Intest Res,* 13(1)**,** 11-8.

Leitch, V. D., Strudwick, X. L., Matthaei, K. I., Dent, L. A. & Cowin, A. J. (2009). IL-5-overexpressing mice exhibit eosinophilia and altered wound healing through mechanisms involving prolonged inflammation. *Immunol Cell Biol,* 87(2)**,** 131-40.

Levison, S. E., Fisher, P., Hankinson, J., Zeef, L., Eyre, S., Ollier, W. E., McLaughlin, J. T., Brass, A., Grencis, R. K. & Pennock, J. L. (2013). Genetic analysis of the Trichuris muris-induced model of colitis reveals QTL overlap and a novel gene cluster for establishing colonic inflammation. *BMC Genomics,* 14**,** 127.

Levison, S. E., Fisher, P., Newman, S., McLaughlin, J., Grencis, R. K. & Pennock, J. L. (2011). Genetic susceptibility to chronic trichuris muris induced experimental colitis: translation and relevance to human Crohn's disease. *Gut,* 60(1)**,** A60.

Levy, M., Thaiss, C. A. & Elinav, E. (2016). Metabolites: messengers between the microbiota and the immune system. *Genes Dev,* 30(14)**,** 1589-97.

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D. & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A,* 102(31)**,** 11070-5.

Li, H. & Jia, W. (2013). Cometabolism of microbes and host: implications for drug metabolism and drug-induced toxicity. *Clin Pharmacol Ther,* 94**,** 574-81.

Li, H., Limenitakis, J. P., Fuhrer, T., Geuking, M. B., Lawson, M. A., Wyss, M., Brugiroux, S., Keller, I., Macpherson, J. A., Rupp, S., Stolp, B., Stein, J. V., Stecher, B., Sauer, U., McCoy, K. D. & Macpherson, A. J. (2015). The outer mucus layer hosts a distinct intestinal microbial niche. *Nat Commun,* 6**,** 8292.

Li, J., Butcher, J., Mack, D. & Stintzi, A. (2014a). Functional Impacts of the Intestinal Microbiome in the Pathogenesis of Inflammatory Bowel Disease. *Inflamm Bowel Dis*.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J. R., Prifti, E., Nielsen, T., Juncker, A. S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X., Xiao, L., Liang, S., Zhang, D., Zhang, Z., Chen, W., Zhao, H., Al-Aama, J. Y., Edris, S., Yang, H., Hansen, T., Nielsen, H. B., Brunak, S., Kristiansen, K., Guarner, F., Pedersen, O., Dore, J., Ehrlich, S. D. & Bork, P. (2014b). An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol,* 32(8)**,** 834-41.

Li, J. Z., Chapman, B., Charlebois, P., Hofmann, O., Weiner, B., Porter, A. J., Samuel, R., Vardhanabhuti, S., Zheng, L., Eron, J., Taiwo, B., Zody, M. C., Henn, M. R., Kuritzkes, D. R., Hide, W., Wilson, C. C., Berzins, B. I., Acosta, E. P., Bastow, B., Kim, P. S., Read, S. W., Janik, J., Meres, D. S., Lederman, M. M., Mong-Kryspin, L., Shaw, K. E., Zimmerman, L. G., Leavitt, R., De La Rosa, G. & Jennings, A. (2014c). Comparison of illumina and 454 deep sequencing in participants failing raltegravir-based antiretroviral therapy. *PLoS One,* 9**,** e90485.

Li, Y., Innocentin, S., Withers, D. R., Roberts, N. A., Gallagher, A. R., Grigorieva, E. F., Wilhelm, C. & Veldhoen, M. (2011). Exogenous stimuli maintain intraepithelial lymphocytes via aryl hydrocarbon receptor activation. *Cell,* 147(3)**,** 629-40.

Lim, Y. W., Evangelista, J. S., 3rd, Schmieder, R., Bailey, B., Haynes, M., Furlan, M., Maughan, H., Edwards, R., Rohwer, F. & Conrad, D. (2014). Clinical insights from metagenomic analysis of sputum samples from patients with cystic fibrosis. *J Clin Microbiol,* 52**,** 425-37.

Lin, R., Liu, W., Piao, M. & Zhu, H. (2017). A review of the relationship between the gut microbiota and amino acid metabolism. *Amino Acids,* 49(12)**,** 2083-2090.

Lopez-Siles, M., Martinez-Medina, M., Abella, C., Busquets, D., Sabat-Mir, M., Duncan, S. H., Aldeguer, X., Flint, H. J. & Garcia-Gil, L. J. (2015). Mucosa-associated Faecalibacterium prausnitzii phylotype richness is reduced in patients with inflammatory bowel disease. *Appl Environ Microbiol,* 81(21)**,** 7582-92.

Lovat, L. B. (1996). Age related changes in gut physiology and nutritional status. *Gut,* 38**,** 306-9.

Lozupone, C. & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol,* 71**,** 8228-35.

Luo, C., Tsementzi, D., Kyrpides, N., Read, T. & Konstantinidis, K. T. (2012). Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One,* 7**,** e30087.

Ma, C., Han, M., Heinrich, B., Fu, Q., Zhang, Q., Sandhu, M., Agdashian, D., Terabe, M., Berzofsky, J. A., Fako, V., Ritz, T., Longerich, T., Theriot, C. M., McCulloch, J. A., Roy, S., Yuan, W., Thovarai, V., Sen, S. K., Ruchirawat, M., Korangy, F., Wang, X. W., Trinchieri,

G. & Greten, T. F. (2018). Gut microbiome-mediated bile acid metabolism regulates liver cancer via NKT cells. *Science,* 360(6391).

Macfarlane, S. & Macfarlane, G. T. (2003). Regulation of short-chain fatty acid production. *Proc Nutr Soc,* 62(1)**,** 67-72.

Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., Ballet, V., Claes, K., Van Immerseel, F., Verbeke, K., Ferrante, M., Verhaegen, J., Rutgeerts, P. & Vermeire, S. (2014). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut,* 63(8)**,** 1275-83.

Mackie, R. I., Sghir, A. & Gaskins, H. R. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr,* 69**,** 1035s-1045s.

Macpherson, A. J. & Harris, N. L. (2004). Interactions between commensal intestinal bacteria and the immune system. *Nat Rev Immunol,* 4(6)**,** 478-85.

Macpherson, A. J., McCoy, K. D., Johansen, F. E. & Brandtzaeg, P. (2008). The immune geography of IgA induction and function. *Mucosal Immunol,* 1**,** 11-22.

Macpherson, A. J., Yilmaz, B., Limenitakis, J. P. & Ganal-Vonarburg, S. C. (2018). IgA Function in Relation to the Intestinal Microbiota. *Annu Rev Immunol,* 36**,** 359-381.

Makino, H., Kushiro, A., Ishikawa, E., Kubota, H., Gawad, A., Sakai, T., Oishi, K., Martin, R., Ben-Amor, K., Knol, J. & Tanaka, R. (2013). Mother-to-infant transmission of intestinal bifidobacterial strains has an impact on the early development of vaginally delivered infant's microbiota. *PLoS One,* 8**,** e78331.

Makki, K., Deehan, E. C., Walter, J. & Backhed, F. (2018). The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe,* 23(6)**,** 705-715.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J. & Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut,* 55(2)**,** 205-11.

Marchesi, J. R., Holmes, E., Khan, F., Kochhar, S., Scanlan, P., Shanahan, F., Wilson, I. D. & Wang, Y. (2007). Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *J Proteome Res,* 6(2)**,** 546-51.

Mariat, D., Firmesse, O., Levenez, F., Guimaraes, V., Sokol, H., Dore, J., Corthier, G. & Furet, J. P. (2009). The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol,* 9**,** 123.

Markle, J. G., Frank, D. N., Mortin-Toth, S., Robertson, C. E., Feazel, L. M., Rolle-Kampczyk, U., von Bergen, M., McCoy, K. D., Macpherson, A. J. & Danska, J. S. (2013). Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science,* 339(6123)**,** 1084-8.

Mastromarino, P., Capobianco, D., Campagna, G., Laforgia, N., Drimaco, P., Dileone, A. & Baldassarre, M. E. (2014). Correlation between lactoferrin and beneficial microbiota in breast milk and infant's feces. *Biometals,* 27**,** 1077-86.

Matsen, F. A. t. (2014). Phylogenetics and the Human Microbiome. *Syst Biol*.

McHardy, I. H., Goudarzi, M., Tong, M., Ruegger, P. M., Schwager, E., Weger, J. R., Graeber, T. G., Sonnenburg, J. L., Horvath, S., Huttenhower, C., McGovern, D. P., Fornace, A. J., Jr., Borneman, J. & Braun, J. (2013). Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome,* 1**,** 17.

McLoughlin, K., Schluter, J., Rakoff-Nahoum, S., Smith, A. L. & Foster, K. R. (2016). Host Selection of Microbiota via Differential Adhesion. *Cell Host Microbe,* 19(4)**,** 550-9.

Mijac, D., Vukovic-Petrovic, I., Mijac, V., Perovic, V., Milic, N., Djuranovic, S., Bojic, D., Popovic, D., Culafic, D., Krstic, M., Jankovic, G., Pravica, V. & Markovic, M. (2018). MDR1 gene polymorphisms are associated with ulcerative colitis in a cohort of Serbian patients with inflammatory bowel disease. *PLoS One,* 13(3)**,** e0194536.

Mikami, K., Takahashi, H., Kimura, M., Isozaki, M., Izuchi, K., Shibata, R., Sudo, N., Matsumoto, H. & Koga, Y. (2009). Influence of maternal bifidobacteria on the establishment of bifidobacteria colonizing the gut in infants. *Pediatr Res,* 65**,** 669-74.

Miller, H. R. P. & Jarrett, W. F. H. (1971). Immune reactions in mucous membranes: I. Intestinal mast cell response during helminth expulsion in the rat. *Immunology,* 20**,** 277-88.

Molodecky, N. A., Soon, I. S., Rabi, D. M., Ghali, W. A., Ferris, M., Chernoff, G., Benchimol, E. I., Panaccione, R., Ghosh, S., Barkema, H. W. & Kaplan, G. G. (2012). Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology,* 142(1)**,** 46-54.e42; quiz e30.

Monteleone, I., Rizzo, A., Sarra, M., Sica, G., Sileri, P., Biancone, L., MacDonald, T. T., Pallone, F. & Monteleone, G. (2011). Aryl hydrocarbon receptor-induced signals up-regulate IL-22 production and inhibit inflammation in the gastrointestinal tract. *Gastroenterology,* 141(1)**,** 237-48, 248 e1.

Moreira, L. O. & Zamboni, D. S. (2012). NOD1 and NOD2 Signaling in Infection and Inflammation. *Front Immunol,* 3**,** 328.

Morgan, X. C. & Huttenhower, C. (2012). Chapter 12: Human microbiome analysis. *PLoS Comput Biol,* 8**,** e1002808.

Morgan, X. C. & Huttenhower, C. (2014). Meta'omic analytic techniques for studying the intestinal microbiome. *Gastroenterology,* 146**,** 1437-1448.e1.

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., Reyes, J. A., Shah, S. A., LeLeiko, N., Snapper, S. B., Bousvaros, A., Korzenik, J., Sands, B. E., Xavier, R. J. & Huttenhower, C. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol,* 13(9)**,** R79.

Morrison, D. J. & Preston, T. (2016). Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes,* 7(3)**,** 189-200.

Moya-Perez, A., Luczynski, P., Renes, I. B., Wang, S., Borre, Y., Anthony Ryan, C., Knol, J., Stanton, C., Dinan, T. G. & Cryan, J. F. (2017). Intervention strategies for cesarean section-induced alterations in the microbiota-gut-brain axis. *Nutr Rev,* 75(4)**,** 225-240.

Moyat, M. & Velin, D. (2014). Immune responses to Helicobacter pylori infection. *World J Gastroenterol,* 20**,** 5583-93.

Nagao-Kitamoto, H., Shreiner, A. B., Gillilland, M. G., 3rd, Kitamoto, S., Ishii, C., Hirayama, A., Kuffa, P., El-Zaatari, M., Grasberger, H., Seekatz, A. M., Higgins, P. D., Young, V. B., Fukuda, S., Kao, J. Y. & Kamada, N. (2016). Functional Characterization of Inflammatory Bowel Disease-Associated Gut Dysbiosis in Gnotobiotic Mice. *Cell Mol Gastroenterol Hepatol,* 2(4)**,** 468-481.

Nakajima, A., Vogelzang, A., Maruya, M., Miyajima, M., Murata, M., Son, A., Kuwahara, T., Tsuruyama, T., Yamada, S., Matsuura, M., Nakase, H., Peterson, D. A., Fagarasan, S. & Suzuki, K. (2018). IgA regulates the composition and metabolic function of gut microbiota by promoting symbiosis between bacteria. *J Exp Med,* 215(8)**,** 2019-2034.

Nakamura, K., Sakuragi, N., Takakuwa, A. & Ayabe, T. (2016). Paneth cell alpha-defensins and enteric microbiota in health and disease. *Biosci Microbiota Food Health,* 35(2)**,** 57-67.

Netzer-Tomkins, H., Rubin, L. & Ephros, M. (2016). Breastfeeding Is Associated with Decreased Hospitalization for Neonatal Fever. *Breastfeed Med,* 11**,** 218-21.

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., Sung, J. J. Y. & Kaplan, G. G. (2018). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet,* 390(10114)**,** 2769-2778.

Nguyen, T. L., Vieira-Silva, S., Liston, A. & Raes, J. (2015). How informative is the mouse for human gut microbiota research? *Dis Model Mech,* 8(1)**,** 1-16.

Nicholson, J. K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W. & Pettersson, S. (2012). Host-gut microbiota metabolic interactions. *Science,* 336**,** 1262-7.

Nostrant, T. T., Kumar, N. B. & Appelman, H. D. (1987). Histopathology differentiates acute self-limited colitis from ulcerative colitis. *Gastroenterology,* 92(2)**,** 318-28.

Okayasu, I., Hatakeyama, S., Yamada, M., Ohkusa, T., Inagaki, Y. & Nakaya, R. (1990). A novel method in the induction of reliable experimental acute and chronic ulcerative colitis in mice. *Gastroenterology,* 98(3)**,** 694-702.

Org, E., Mehrabian, M., Parks, B. W., Shipkova, P., Liu, X., Drake, T. A. & Lusis, A. J. (2016). Sex differences and hormonal effects on gut microbiota composition in mice. *Gut Microbes.*

Ostman, S., Rask, C., Wold, A. E., Hultkrantz, S. & Telemo, E. (2006). Impaired regulatory T cell function in germ-free mice. *Eur J Immunol,* 36(9)**,** 2336-46.

Ott, S. J., Musfeldt, M., Wenderoth, D. F., Hampe, J., Brant, O., Folsch, U. R., Timmis, K. N. & Schreiber, S. (2004). Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut,* 53(5)**,** 685-93.

Ott, S. J. & Schreiber, S. (2006). Reduced microbial diversity in inflammatory bowel diseases. *Gut,* 55(8)**,** 1207.

Palm, N. W., de Zoete, M. R., Cullen, T. W., Barry, N. A., Stefanowski, J., Hao, L., Degnan, P. H., Hu, J., Peter, I., Zhang, W., Ruggiero, E., Cho, J. H., Goodman, A. L. & Flavell, R. A.

(2014). Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell,* 158**,** 1000-10.

Pannaraj, P. S., Li, F., Cerini, C., Bender, J. M., Yang, S., Rollie, A., Adisetiyo, H., Zabih, S., Lincez, P. J., Bittinger, K., Bailey, A., Bushman, F. D., Sleasman, J. W. & Aldrovandi, G. M. (2017). Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome. *JAMA Pediatr,* 171(7)**,** 647-654.

Panwala, C. M., Jones, J. C. & Viney, J. L. (1998). A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, mdr1a, spontaneously develop colitis. *J Immunol,* 161**,** 5733-44.

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., Martinez, X., Varela, E., Sarrabayrouse, G., Machiels, K., Vermeire, S., Sokol, H., Guarner, F. & Manichanh, C. (2017). A microbial signature for Crohn's disease. *Gut,* 66(5)**,** 813-822.

Pei, Z., Bini, E. J., Yang, L., Zhou, M., Francois, F. & Blaser, M. J. (2004). Bacterial biota in the human distal esophagus. *Proc Natl Acad Sci U S A,* 101(12)**,** 4250-5.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., van den Brandt, P. A. & Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics,* 118**,** 511-21.

Pendyala, S., Walker, J. M. & Holt, P. R. (2012). A high-fat diet is associated with endotoxemia that originates from the gut. *Gastroenterology,* 142(5)**,** 1100-1101.e2.

Perez-Cobas, A. E., Gosalbes, M. J., Friedrichs, A., Knecht, H., Artacho, A., Eismann, K., Otto, W., Rojo, D., Bargiela, R., von Bergen, M., Neulinger, S. C., Daumer, C., Heinsen, F. A., Latorre, A., Barbas, C., Seifert, J., dos Santos, V. M., Ott, S. J., Ferrer, M. & Moya, A. (2013). Gut microbiota disturbance during antibiotic therapy: a multi-omic approach. *Gut,* 62**,** 1591-601.

Peterson, J., Garges, S., Giovanni, M., McInnes, P., Wang, L., Schloss, J. A., Bonazzi, V., McEwen, J. E., Wetterstrand, K. A., Deal, C., Baker, C. C., Di Francesco, V., Howcroft, T. K., Karp, R. W., Lunsford, R. D., Wellington, C. R., Belachew, T., Wright, M., Giblin, C., David, H., Mills, M., Salomon, R., Mullins, C., Akolkar, B., Begg, L., Davis, C., Grandison, L., Humble, M., Khalsa, J., Little, A. R., Peavy, H., Pontzer, C., Portnoy, M., Sayre, M. H., Starke-Reed, P., Zakhari, S., Read, J., Watson, B. & Guyer, M. (2009). The NIH Human Microbiome Project. *Genome Res,* 19**,** 2317-23.

Peterson, L. W. & Artis, D. (2014). Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat Rev Immunol,* 14**,** 141-53.

Petnicki-Ocwieja, T., Hrncir, T., Liu, Y. J., Biswas, A., Hudcovic, T., Tlaskalova-Hogenova, H. & Kobayashi, K. S. (2009). Nod2 is required for the regulation of commensal microbiota in the intestine. *Proc Natl Acad Sci U S A,* 106**,** 15813-8.

Peyrin-Biroulet, L. & Chamaillard, M. (2007). NOD2 and defensins: translating innate to adaptive immunity in Crohn's disease. *J Endotoxin Res,* 13(3)**,** 135-9.

Podolsky, D. K. (1991). Inflammatory bowel disease (1). *N Engl J Med,* 325**,** 928-37.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D., Cao, J., Wang, B.,

Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J. M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H. B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Jian, M., Zhou, Y., Li, Y., Zhang, X., Qin, N., Yang, H., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P. & Ehrlich, S. D. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature,* 464**,** 59-65.

Ramirez, G. A., Yacoub, M. R., Ripa, M., Mannina, D., Cariddi, A., Saporiti, N., Ciceri, F., Castagna, A., Colombo, G. & Dagna, L. (2018). Eosinophils from Physiology to Disease: A Comprehensive Review. *Biomed Res Int,* 2018**,** 9095275.

Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun,* 469(4)**,** 967-77.

Rennick, D. M., Fort, M. M. & Davidson, N. J. (1997). Studies with IL-10-/- mice: an overview. *J Leukoc Biol,* 61(4)**,** 389-96.

Rhee, K. H., Park, J. S. & Cho, M. J. (2014). Helicobacter pylori: Bacterial Strategy for Incipient Stage and Persistent Colonization in Human Gastric Niches. *Yonsei Med J,* 55**,** 1453-1466.

Ridlon, J. M., Kang, D. J., Hylemon, P. B. & Bajaj, J. S. (2014). Bile acids and the gut microbiome. *Curr Opin Gastroenterol,* 30(3)**,** 332-8.

Robertson, B. R., O'Rourke, J. L., Neilan, B. A., Vandamme, P., On, S. L., Fox, J. G. & Lee, A. (2005). Mucispirillum schaedleri gen. nov., sp. nov., a spiral-shaped bacterium colonizing the mucus layer of the gastrointestinal tract of laboratory rodents. *Int J Syst Evol Microbiol,* 55**,** 1199-204.

Round, J. L. & Mazmanian, S. K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci U S A,* 107**,** 12204-9.

Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I. & Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. *Eur J Nutr,* 57(1)**,** 1-24.

Rupnik, M., Wilcox, M. H. & Gerding, D. N. (2009). Clostridium difficile infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol,* 7**,** 526-36.

Samuel, B. S., Shaito, A., Motoike, T., Rey, F. E., Backhed, F., Manchester, J. K., Hammer, R. E., Williams, S. C., Crowley, J., Yanagisawa, M. & Gordon, J. I. (2008). Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc Natl Acad Sci U S A,* 105(43)**,** 16767-72.

Santoro, A., Ostan, R., Candela, M., Biagi, E., Brigidi, P., Capri, M. & Franceschi, C. (2018). Gut microbiota changes in the extreme decades of human life: a focus on centenarians. *Cell Mol Life Sci,* 75(1)**,** 129-148.

Satokari, R., Gronroos, T., Laitinen, K., Salminen, S. & Isolauri, E. (2009). Bifidobacterium and Lactobacillus DNA in the human placenta. *Lett Appl Microbiol,* 48(1)**,** 8-12.

Scheppach, W. (1996). Treatment of distal ulcerative colitis with short-chain fatty acid enemas. A placebo-controlled trial. German-Austrian SCFA Study Group. *Dig Dis Sci,* 41(11)**,** 2254-9.

Scheppach, W., Sommer, H., Kirchner, T., Paganelli, G. M., Bartram, P., Christl, S., Richter, F., Dusel, G. & Kasper, H. (1992). Effect of butyrate enemas on the colonic mucosa in distal ulcerative colitis. *Gastroenterology,* 103(1)**,** 51-6.

Schicho, R., Shaykhutdinov, R., Ngo, J., Nazyrova, A., Schneider, C., Panaccione, R., Kaplan, G. G., Vogel, H. J. & Storr, M. (2012). Quantitative metabolomic profiling of serum, plasma, and urine by (1)H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. *J Proteome Res,* 11(6)**,** 3344-57.

Scholz, F., Badgley, B. D., Sadowsky, M. J. & Kaplan, D. H. (2014). Immune mediated shaping of microflora community composition depends on barrier site. *PLoS One,* 9**,** e84019.

Sciences, O. (2018). *The Oricol Sampling Kit* [Online]. Available: http://www.originsciences.com/index.php/the-oricol-sampling-kit [Accessed August 31st 2018].

Sender, R., Fuchs, S. & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biol,* 14(8)**,** e1002533.

Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A. & Orkin, S. H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *Embo j,* 16**,** 3965-73.

Siqueira, J. F., Jr. & Rocas, I. N. (2010). The oral microbiota: general overview, taxonomy, and nucleic acid techniques. *Methods Mol Biol,* 666**,** 55-69.

Smith, K., McCoy, K. D. & Macpherson, A. J. (2007). Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin Immunol,* 19**,** 59-69.

Soave, O. & Brand, C. D. (1991). Coprophagy in animals: a review. *Cornell Vet,* 81(4)**,** 357-64.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J. J., Blugeon, S., Bridonneau, C., Furet, J. P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottiere, H. M., Dore, J., Marteau, P., Seksik, P. & Langella, P. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A,* 105(43)**,** 16731-6.

Sokol, H., Seksik, P., Furet, J. P., Firmesse, O., Nion-Larmurier, I., Beaugerie, L., Cosnes, J., Corthier, G., Marteau, P. & Dore, J. (2009). Low counts of Faecalibacterium prausnitzii in colitis microbiota. *Inflamm Bowel Dis,* 15(8)**,** 1183-9.

Spasova, D. S. & Surh, C. D. (2014). Blowing on embers: commensal microbiota and our immune system. *Front Immunol,* 5**,** 318.

Stensballe, L. G., Simonsen, J., Jensen, S. M., Bonnelykke, K. & Bisgaard, H. (2013). Use of antibiotics during pregnancy increases the risk of asthma in early childhood. *J Pediatr,* 162**,** 832-838.e3.

Sternini, C., Anselmi, L. & Rozengurt, E. (2008). Enteroendocrine cells: a site of 'taste' in gastrointestinal chemosensing. *Curr Opin Endocrinol Diabetes Obes,* 15(1)**,** 73-8.

Sudo, N., Chida, Y., Aiba, Y., Sonoda, J., Oyama, N., Yu, X. N., Kubo, C. & Koga, Y. (2004). Postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice. *J Physiol,* 558(Pt 1)**,** 263-75.

Takahashi, K., Nishida, A., Fujimoto, T., Fujii, M., Shioya, M., Imaeda, H., Inatomi, O., Bamba, S., Sugimoto, M. & Andoh, A. (2016). Reduced Abundance of Butyrate-Producing Bacteria Species in the Fecal Microbial Community in Crohn's Disease. *Digestion,* 93(1)**,** 59-65.

Tang, M. S., Poles, J., Leung, J. M., Wolff, M. J., Davenport, M., Lee, S. C., Lim, Y. A., Chua, K. H., Loke, P. & Cho, I. (2015). Inferred metagenomic comparison of mucosal and fecal microbiota from individuals undergoing routine screening colonoscopy reveals similar differences observed during active inflammation. *Gut Microbes,* 6(1)**,** 48-56.

Tedjo, D. I., Smolinska, A., Savelkoul, P. H., Masclee, A. A., van Schooten, F. J., Pierik, M. J., Penders, J. & Jonkers, D. M. (2016). The fecal microbiota as a biomarker for disease activity in Crohn's disease. *Sci Rep,* 6**,** 35216.

Thevaranjan, N., Puchta, A., Schulz, C., Naidoo, A., Szamosi, J. C., Verschoor, C. P., Loukov, D., Schenck, L. P., Jury, J., Foley, K. P., Schertzer, J. D., Larche, M. J., Davidson, D. J., Verdu, E. F., Surette, M. G. & Bowdish, D. M. E. (2017). Age-Associated Microbial Dysbiosis Promotes Intestinal Permeability, Systemic Inflammation, and Macrophage Dysfunction. *Cell Host Microbe,* 21(4)**,** 455-466.e4.

Thum, C., Cookson, A. L., Otter, D. E., McNabb, W. C., Hodgkinson, A. J., Dyer, J. & Roy, N. C. (2012). Can nutritional modulation of maternal intestinal microbiota influence the development of the infant gastrointestinal tract? *J Nutr,* 142**,** 1921-8.

Tringe, S. G. & Hugenholtz, P. (2008). A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol,* 11**,** 442-6.

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., Blanchard, C., Junt, T., Nicod, L. P., Harris, N. L. & Marsland, B. J. (2014). Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat Med,* 20(2)**,** 159-66.

Tschurtschenthaler, M., Wang, J., Fricke, C., Fritz, T. M., Niederreiter, L., Adolph, T. E., Sarcevic, E., Kunzel, S., Offner, F. A., Kalinke, U., Baines, J. F., Tilg, H. & Kaser, A. (2014). Type I interferon signalling in the intestinal epithelium affects Paneth cells, microbial ecology and epithelial regeneration. *Gut*.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R. & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature,* 444**,** 1027-31.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R. & Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med,* 1(6)**,** 6ra14.

Tyler, A. D., Smith, M. I. & Silverberg, M. S. (2014). Analyzing the human microbiome: a "how to" guide for physicians. *Am J Gastroenterol,* 109**,** 983-93.

Tzeneva, V. A., Heilig, H. G., van Vliet, W. A., Akkermans, A. D., de Vos, W. M. & Smidt, H. (2008). 16S rRNA targeted DGGE fingerprinting of microbial communities. *Methods Mol Biol***,** 335-49.

Ursell, L. K., Metcalf, J. L., Parfrey, L. W. & Knight, R. (2012). Defining the human microbiome. *Nutr Rev,* 70 Suppl 1**,** S38-44.

Vaishnava, S., Yamamoto, M., Severson, K. M., Ruhn, K. A., Yu, X., Koren, O., Ley, R., Wakeland, E. K. & Hooper, L. V. (2011). The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science,* 334**,** 255-8.

van der Flier, L. G. & Clevers, H. (2009). Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu Rev Physiol,* 71**,** 241-60.

Van der Sluis, M., De Koning, B. A., De Bruijn, A. C., Velcich, A., Meijerink, J. P., Van Goudoever, J. B., Buller, H. A., Dekker, J., Van Seuningen, I., Renes, I. B. & Einerhand, A. W. (2006). Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology,* 131(1)**,** 117-29.

Vetrovsky, T. & Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One,* 8**,** e57923.

Vignal, C., Singer, E., Peyrin-Biroulet, L., Desreumaux, P. & Chamaillard, M. (2007). How NOD2 mutations predispose to Crohn's disease? *Microbes Infect,* 9**,** 658-63.

Vigsnaes, L. K., Brynskov, J., Steenholdt, C., Wilcks, A. & Licht, T. R. (2012). Gram-negative bacteria account for main differences between faecal microbiota from patients with ulcerative colitis and healthy controls. *Benef Microbes,* 3(4)**,** 287-97.

vom Steeg, L. G. & Klein, S. L. (2016). SeXX Matters in Infectious Disease Pathogenesis. *PLoS Pathog,* 12(2)**,** e1005374.

Walker, A. W., Ince, J., Duncan, S. H., Webster, L. M., Holtrop, G., Ze, X., Brown, D., Stares, M. D., Scott, P., Bergerat, A., Louis, P., McIntosh, F., Johnstone, A. M., Lobley, G. E., Parkhill, J. & Flint, H. J. (2011). Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *Isme j,* 5**,** 220-30.

Wang, X., Heazlewood, S. P., Krause, D. O. & Florin, T. H. (2003). Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. *J Appl Microbiol,* 95(3)**,** 508-20.

Wang, X., Lu, H., Feng, Z., Cao, J., Fang, C., Xu, X., Zhao, L. & Shen, J. (2017). Development of Human Breast Milk Microbiota-Associated Mice as a Method to Identify Breast Milk Bacteria Capable of Colonizing Gut. *Front Microbiol,* 8**,** 1242.

Watari, J., Chen, N., Amenta, P. S., Fukui, H., Oshima, T., Tomita, T., Miwa, H., Lim, K. J. & Das, K. M. (2014). Helicobacter pylori associated chronic gastritis, clinical syndromes, precancerous lesions, and pathogenesis of gastric cancer development. *World J Gastroenterol,* 20(18)**,** 5461-73.

Wehkamp, J., Gotz, M., Herrlinger, K., Steurer, W. & Stange, E. F. (2016). Inflammatory Bowel Disease. *Dtsch Arztebl Int,* 113(5)**,** 72-82.

Wesolowska-Andersen, A., Bahl, M. I., Carvalho, V., Kristiansen, K., Sicheritz-Ponten, T., Gupta, R. & Licht, T. R. (2014). Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome,* 2**,** 19.

Wikoff, W. R., Anfora, A. T., Liu, J., Schultz, P. G., Lesley, S. A., Peters, E. C. & Siuzdak, G. (2009). Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc Natl Acad Sci U S A,* 106**,** 3698-703.

Wilk, J. N., Bilsborough, J. & Viney, J. L. (2005). The mdr1a-/- mouse model of spontaneous colitis: a relevant and appropriate animal model to study inflammatory bowel disease. *Immunol Res,* 31**,** 151-9.

Williams, H., Campbell, L., Crompton, R. A., Singh, G., McHugh, B. J., Davidson, D. J., McBain, A. J., Cruickshank, S. M. & Hardman, M. J. (2018). Microbial Host Interactions and Impaired Wound Healing in Mice and Humans: Defining a Role for BD14 and NOD2. *J Invest Dermatol*.

Withers, S. B., Forman, R., Meza-Perez, S., Sorobetea, D., Sitnik, K., Hopwood, T., Lawrence, C. B., Agace, W. W., Else, K. J., Heagerty, A. M., Svensson-Frej, M. & Cruickshank, S. M. (2017). Eosinophils are key regulators of perivascular adipose tissue and vascular functionality. *Sci Rep,* 7**,** 44571.

Wlodarska, M., Thaiss, C. A., Nowarski, R., Henao-Mejia, J., Zhang, J. P., Brown, E. M., Frankel, G., Levy, M., Katz, M. N., Philbrick, W. M., Elinav, E., Finlay, B. B. & Flavell, R. A. (2014). NLRP6 inflammasome orchestrates the colonic host-microbial interface by regulating goblet cell mucus secretion. *Cell,* 156(5)**,** 1045-59.

Wong, J. M., de Souza, R., Kendall, C. W., Emam, A. & Jenkins, D. J. (2006). Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol,* 40**,** 235-43.

Wu, D., Molofsky, A. B., Liang, H. E., Ricardo-Gonzalez, R. R., Jouihan, H. A., Bando, J. K., Chawla, A. & Locksley, R. M. (2011a). Eosinophils sustain adipose alternatively activated macrophages associated with glucose homeostasis. *Science,* 332**,** 243-7.

Wu, G. (2009). Amino acids: metabolism, functions, and nutrition. *Amino Acids,* 37(1)**,** 1-17.

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F. D. & Lewis, J. D. (2011b). Linking long-term dietary patterns with gut microbial enterotypes. *Science,* 334**,** 105-8.

Wu, H. J., Ivanov, II, Darce, J., Hattori, K., Shima, T., Umesaki, Y., Littman, D. R., Benoist, C. & Mathis, D. (2010). Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity,* 32(6)**,** 815-27.

Yang, B., Wang, Y. & Qian, P. Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics,* 17**,** 135.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B.,

Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., Knight, R. & Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature,* 486**,** 222-7.

Yoo, D. H., Kim, I. S., Van Le, T. K., Jung, I. H., Yoo, H. H. & Kim, D. H. (2014). Gut microbiota-mediated drug interactions between lovastatin and antibiotics. *Drug Metab Dispos,* 42**,** 1508-13.

Yoshida, M., Hatano, N., Nishiumi, S., Irino, Y., Izumi, Y., Takenawa, T. & Azuma, T. (2012). Diagnosis of gastroenterological diseases by metabolome analysis using gas chromatography-mass spectrometry. *J Gastroenterol,* 47(1)**,** 9-20.

Yu, G., Torres, J., Hu, N., Medrano-Guzman, R., Herrera-Goepfert, R., Humphrys, M. S., Wang, L., Wang, C., Ding, T., Ravel, J., Taylor, P. R., Abnet, C. C. & Goldstein, A. M. (2017). Molecular Characterization of the Human Stomach Microbiota in Gastric Cancer Patients. *Front Cell Infect Microbiol,* 7**,** 302.

Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *Isme j,* 6(8)**,** 1535-43.

Zelante, T., Iannitti, R. G., Cunha, C., De Luca, A., Giovannini, G., Pieraccini, G., Zecchi, R., D'Angelo, C., Massi-Benedetti, C., Fallarino, F., Carvalho, A., Puccetti, P. & Romani, L. (2013). Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity,* 39(2)**,** 372-85.

Zinocker, M. K. & Lindseth, I. A. (2018). The Western Diet-Microbiome-Host Interaction and Its Role in Metabolic Disease. *Nutrients,* 10(3).

Zoetendal, E. G., Raes, J., van den Bogert, B., Arumugam, M., Booijink, C. C., Troost, F. J., Bork, P., Wels, M., de Vos, W. M. & Kleerebezem, M. (2012). The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. *Isme j,* 6**,** 1415-26.

Zoetendal, E. G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A. D. & de Vos, W. M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol,* 68**,** 3401-7.

Chapter Two

Phylogenetic mouse gut microbiomes

Author Contributions

Gurdeep Singh: Designed and performed experiments, analysed data and wrote the manuscript.

Andrew Brass: Designed and supervised the project, critically reviewed the manuscript.

Sheena Cruickshank: Designed and supervised the project, critically reviewed and edited the manuscript.

Christopher Knight: Designed and supervised the project, supervised analysis of the sequencing data, critically reviewed and edited the manuscript.

2. Gut microbial taxa at distinct phylogenetic scales distinguish host age, social group and microbial niche, but not genotype, in a murine model of inflammatory bowel disease

Gurdeep Singh[1], Andrew Brass[2], Sheena M. Cruickshank[1], and Christopher G. Knight[3].

[1] Faculty of Biology, Medicine and Health, Lydia Becker Institute of Immunology and Inflammation, Manchester Academic Health Science Centre, A.V. Hill Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom.

[2] Faculty of Biology, Medicine and Health, Division of Informatics, Imaging and Data Sciences, Stopford Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom.

[3] Faculty of Science and Engineering, School of Earth and Environmental Sciences, Michael Smith Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom

Corresponding author:

Professor Sheena Cruickshank[1]

A.V. Hill Building

The University of Manchester

Oxford Road

Manchester

M13 9PT

Sheena.Cruickshank@manchester.ac.uk

Phone +44 (0) 161 275 1582

Short title: **Phylogenetic mouse gut microbiomes**

2.1 Abstract

Gut microbial communities vary with factors such as the niche they occupy, host age or home environment, and determining the microbial taxa that are important in these distinctions is challenging. Current approaches to amplicon-sequencing experiments frequently do not use the full power of the data, either focusing on summary statistics or taxonomic scales (e.g. operational taxonomic units) that can bias analyses. Analyses tend not to consider particular taxa, whether low or high abundance, on an equal footing. To address these issues, we have developed a novel community-orientated approach, analysing the relative abundance of taxa at all phylogenetic scales, and used it to explore the role of the gut microbiome in a mouse model of colitis.

We constructed a phylogenetic tree using 16S ribosomal RNA (rRNA) sequence data derived from the stools and colonic mucus of cohoused healthy wildtype mice and littermates that spontaneously develop colitis over time (*mdr1a*$^{-/-}$). Machine learning (random forest) models incorporating all possible clades within that phylogenetic tree were used individually to test whether aspects such as age or genotype could be distinguished. These models discriminated between the age, microbial niche and cage of samples with >90% accuracy but found no consistent distinction between *mdr1a* genotypes. This analysis enabled us to identify distinct taxa specific to discrete features such as mucus versus stool and enabled identification of patterns in the phylogenetic positioning and abundance of taxa distinguishing microbiomes. Cage-specific microbiomes were distinguished by abundant low level taxa, such as groups within the genus *Bacteroides*, whereas mucus and stool microbiomes were distinguished by intermediate level, intermediate abundance taxa, such as the order Pseudomonadales.

This random forest approach is applicable to a wide range of questions around the role of the microbiome in health and disease and may help understand the balance between homeostasis and divergence as well as identifying specific evolutionary groups of bacteria important in disease pathogenesis.

2.2 Introduction

The gut microbiota has been linked to numerous aspects of host health such as immune development (Bouskra et al., 2008), drug metabolism (Clayton et al., 2009) and digestion (Macfarlane and Macfarlane, 2012). Perturbations of the bacterial communities are associated with a variety of diseases (Imhann et al., 2018). However there is considerable variance in the data and a lack of consensus as to which core microbiomes determine key features of health or disease. Determining how gut microbiome stability and differences resolve is challenging, illustrated by the problems of attempts to identify enterotypes, i.e. being able to group people by a set of distinct microbial signatures (Arumugam et al., 2011). However, interpreting these signatures is highly subject to bias (Knights et al., 2014). What is required, are rigorous approaches to both experiment and analysis, capable of identifying discriminating bacteria that are truly associated with health or disease.

Gut microbiota research has tended to focus on stool samples as a representation of the overall gut microbial community. Using stool samples, several diseases, most notably inflammatory bowel disease (IBD), have been causally associated with aberrant immune responses to components of the microbiota (Kalliomaki et al., 2008; Papa et al., 2012). However, there is a lack of consensus around causality for the roles of particular microbial taxa, where, for instance, *Bifidobacteria* and *Lactobacilli*, in stools have been associated with healthy subjects when compared with IBD-affected subjects (Giaffer et al., 1991; Favier et al., 1997). However, stool samples alone do not fully reflect the gut microbiota. Indeed, bacteria within the gut inhabit different niches, notably the mucus layer that lines the intestinal epithelial cells (Li et al., 2015) and we and others have shown that the mucus microbiome is discrete from that detected in stools (Ahmed et al., 2007; Glymenaki et al., 2017). We have also shown that changes in the mucus microbiota precede both changes in the stool microbiota and the onset of disease (Glymenaki et al., 2017). These findings underlie the potential importance of the mucus microbiome for understanding disease, with IBD representing a test case for understanding microbiome associations across stool and mucus.

A further issue of mouse studies of microbiota is that few still record how the mice are caged and whether they are littermates despite this being a known confounder in microbiota studies (Hildebrand et al., 2013; Song et al., 2013; Bramhall et al., 2015). Coprophagy among a co-housed social group facilitates this sharing for gut communities. Experimentally, this presents both a challenge and opportunities. The challenge is the necessity to design experiments in ways that avoid confounding group effects with treatment effects of interest. If this is not

done, distinctions in the microbiota will be found that are falsely assigned to the treatment of interest. The opportunities include using coprophagy to look at faecal-oral invasion (Ridaura et al., 2013); but there is also the opportunity to identify change, or homeostasis, in the microbiota that is robust to repeated re-inoculation with any particular group's stool microbiome.

Microbial populations can also differ at widely varying phylogenetic scales. Large-scale shifts among phyla, for instance the Firmicutes to Bacteroidetes ratio, have been implicated in obesity (Ley et al., 2005). At the same time, individual species have been associated with gut health (Machiels et al., 2014). However, such analyses does not reflect the fact that the bacteria work together as communities and taxonomically similar species may compensate functionally for each other. Therefore, it is crucial that analyses consider bacteria as communities. However, the power of individual measurements or statistics, at whatever level, to distinguish the complexities of communities is extremely limited. The value of such metrics in the face of technical variation among amplicon sequencing protocols and their implementation in different laboratories is also questionable (Hiergeist et al., 2016). A range of taxonomic levels are also considered individually and although a much more complete phylogenetic tree may be included, it is often relegated to diversity statistics calculated by UniFrac (Lozupone and Knight, 2005). Any such summary statistics are poor tools with which to identify specific taxa, for instance those differing consistently between healthy and diseased patient groups. To date, explicitly considering such taxa at different levels, in a single analytical framework that would enable the relative importance of differences at different phylogenetic scales, to be assessed, is rare.

Even where multiple phylogenetic levels are considered together (Ramirez et al., 2018), analyses are highly constrained by available taxonomies – even the addition of operational taxonomic units (OTUs) to the traditional Linnean hierarchy from domains to species does not come close to capturing the diverse intricacy of microbial phylogeny (Hug et al., 2016). Such taxonomies are inadequate, not only because all taxonomies are incomplete and different taxonomies disagree (Balvociute and Huson, 2017; Breitwieser et al., 2017), but because only a small amount of the evolutionary history of any group of microbes can be captured in a taxonomy with a limited number of levels. Any such taxonomy also introduces bias due to a wide variety of factors, from the lack of clear equivalence between the same taxonomic levels in different regions of the tree to drastic variation in the scientific attention paid to different parts of the bacterial phylogeny. A similar risk of bias also applies to the

widely-used, but necessarily arbitrary, 97% similarity threshold for grouping sequences into OTUs. While it may represent a biologically-based compromise, no such threshold can work equally well across the bacterial tree. What is needed are approaches that both capture more of the evolutionary history of microbes in a single analysis and those that are less biased. This could be achieved by more data-driven approaches. At the limit, this would mean using only information available from the sequences collected in a particular experiment. Here we approach this limit for an amplicon-sequencing experiment by using the sequences themselves to estimate the phylogenetic relationships among them and thereby abundances at different phylogenetic scales, only afterwards returning to the wider understanding of microbial taxonomy to interpret the resulting findings.

We have developed a community-focused approach to microbiome analysis in the context of an IBD mouse model (colitis-prone, *mdr1a*$^{-/-}$), incorporating taxa at all phylogenetic levels. We use machine learning models to interrogate the microbiota of stools and colonic mucus of co-housed healthy controls and *mdr1a*$^{-/-}$ mice. We find that the microbiomes of each cage are clearly distinguishable, being separated by differences in common, small-scale taxa (notably particular genera within the order Bacteroidales). Despite this variability, microbiomes from the different niches (stool and mucus) are clearly distinct, as are microbiomes from young (6-week) versus old (18 week) mice, each distinguished by taxa at intermediate phylogenetic and abundance scales. Surprisingly, despite the fact that our analyses are very effective at distinguishing these groupings and identifying the taxa involved, we were unable to find any consistent distinction between the microbiomes (stool or mucus), of *mdr1a*$^{-/-}$ and wildtype (WT) mice at the stages tested. This indicates a striking degree of homeostatic robustness to host genetic perturbation in the gut microbiome.

2.3 Materials and Methods

2.3.1 Animal Maintenance

Mdr1a$^{-/-}$ mice (FVB.129P2-Abcb1atm1Bor N7) (Schinkel et al., 1994) were bred with control FVB mice purchased from Taconic Farms (Albany, NY), to produce the F2 generation. Male mice from each litter were co-housed. Thus, WT and *mdr1a*$^{-/-}$ mice from the same litters were used for all subsequent experiments. Male mice at 6 and 18 weeks of age were used for experiments. Food (Beekay Rat and Mouse Diet No1 pellets; B&K Universal, UK) and water were available *ad libitum*. Ambient temperature was maintained at 21 (+/- 2°C) and the relative humidity was 55 (+/- 10%) with a 12h light/dark cycle. All animals were kept under specific, pathogen-free (SPF) conditions at the University of Manchester and experiments were performed according to the regulations issued by the Home Office under amended ASPA, 2012.

2.3.2 Isolation of genomic DNA

Sample collection and processing was performed as described by Glymenaki et al. (2017). In brief, samples were harvested from mice at two time points, 6 and 18 weeks of age. Stool samples were collected from mice in individual autoclaved cages into sterile tubes and snap frozen on dry ice. Mice were sacrificed via CO_2 inhalation, the proximal colon was cut open and the colonic mucus scraped using cell scrapers and Inhibitex buffer (QIAGEN, Manchester, UK) and snap frozen until use. Genomic DNA was extracted using QIAamp Fast Stool Mini-Kits according to the manufacturer’s instructions (QIAGEN). Additionally, snips of the proximal colon were fixed in Carnoy’s solution (60% methanol, 30% chloroform, 10% glacial acetic acid) and embedded in paraffin for histological analysis.

2.3.3 Histology

Carnoy’s fixed colon samples were incubated in two changes of dry methanol (Sigma-Aldrich, Dorset, UK) for 30 minutes each, followed by absolute ethanol (ThermoFisher Scientific, Paisley, UK) for two incubations at 30 minutes each. Finally, tissue cassettes were processed in a Micro-spin Tissue Processor STP120 (ThermoFisher Scientific) and immersed in paraffin. Colon snips were embedded in paraffin blocks using a Leica Biosystems embedding station (Leica Biosystems, Milton Keynes, UK), with the luminal surface of the colon exposed for tissue sectioning. 5µm tissue sections were cut using a Leica Biosystems microtome and adhered to uncoated microscope slides (ThermoFisher Scientific). Slides were dried for 48 hours at 50ºC before use.

2.3.4 Fluoresence *in-situ* hybridisation (FISH)

FISH was performed as described previously (Glymenaki et al., 2017). In brief, FISH staining was performed using the universal bacterial probe-EUB338 (5′-Cy3-GCTGCCTCCCGTAGGAGT-3′), followed by immunostaining with a rabbit polyclonal MUC2 antibody and goat anti-rabbit Alexa-Fluor 488 antibody (Life Technologies, Paisley, United Kingdom). Slides were imaged using a BX51 upright microscope and a Coolsnap EZ camera (OLYMPUS, Tokyo, Japan) and images were processed using Image J.

2.3.5 16S rRNA gene sequencing processing

16S amplicon sequencing targeting the V3 and V4 variable regions of the 16S rRNA (341F: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3' and 805R: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3') was performed on the Illumina MiSeq platform (Illumina, California, USA) according to manufacturer's guidelines and generated paired-end reads of 300bp in each direction. Illumina reads were demultiplexed to remove adapter sequences and trim primers. Illumina paired-end reads were merged together using SeqPrep (St John, 2018) and submitted to MG-RAST's metagenomics pipeline (Meyer et al., 2008). Reads were pre-processed to remove low-quality and uninformative reads using SolexQA (Cox et al., 2010). The quality-filtering process included removal of reads with low quality ends (i.e. ambiguous leading/trailing bases) and the removal of reads with a read length two standard deviations below the mean. Artificial duplicate reads were then removed based on MG-RAST's pipeline.

The resulting FASTQ files for every sample were merged into a single file of 590822 sequences to simplify processing, manually adding 3 known Archaeal 16S rRNA sequences from *Acidilobus saccharovorans*, *Sulfolobus tokodaii* and *Methanobrevibacter smithii*. Sequences were aligned using a specialist 16S RNA aligner using the Infernal algorithm (Nawrocki and Eddy, 2013), via a web-based interface provided by the Ribosomal Database Project (Cole et al., 2014). This file was then manually curated in R (Team, 2016). Unless otherwise stated, all analyses were performed using custom scripts in R. The number of aligned bases in each sequence was recorded and the distribution of continuously aligned bases was examined. Any sequence that had less than 437 continuously aligned bases was discarded. The remaining 496550 sequences were taken forward for analysis. All sequences were identified using BLAST+ and the top hit for each sequence was recorded (Camacho et al., 2009). The 'classification' function in the 'taxize' R package (Chamberlain and Szocs,

2013) was then used to assign full taxonomic information to each identified taxa where possible.

2.3.6 Phylogenetic Tree

A phylogenetic tree of all sequences was generated using FastTree (Price et al., 2010), using the general time reversible (GTR) + CAT model and default parameters. The tree was rooted using the archaeal sequences as an outgroup. Phylogenetic clades were obtained using the 'Ancestor' function in the 'phangorn' R package (Schliep, 2011). A relative abundance matrix, with abundance based on how many times sequences belonging to a phylogenetic clade appeared in a sample, was calculated.

2.3.7 Ordination

Bray-Curtis dissimilarity values were calculated among all samples (based on the relative abundance matrix) and used for non-metric multidimensional scaling (NMDS) via the 'MASS' (Venables and Ripley, 2002) and 'ecodist' R packages (Goslee and Urban, 2007).

2.3.8 Machine learning

Random forest (RF) models were run using the 'randomForest' package (Liaw and Wiener, 2002) in R. Specifically, the relative abundance matrix was used as an input for the RF, using a forest of 100000 trees and the mtry value was left at default settings. Separate forests were run to predict whether a sample was 6 or 18 weeks old, whether a sample was stool or mucus, whether it was a WT or an *mdr1a*$^{-/-}$ sample, what cage the sample was taken from, the mother of the offspring and whether it could discriminate between combinations of these treatment groups. The 'MeanDecreaseAccuracy' (MDA) value was used as a measure of how important each node was at predicting treatment information and the out-of-bag (OOB) error rate was used to determine the predictive accuracy of the model. Nodes were ranked based on MDA value, taking the five most important nodes, determining the descendant tips and confirming the identity of the tip sequences via the BLAST+ results (Camacho et al., 2009). Additionally, the depth of each node was determined using the 'distances' function in the igraph R package (Csard and Nepusz, 2006). A phylogenetic tree annotated with the resulting information was plotted using the 'plot.phylo' function in the 'ape' package (Paradis et al., 2004).

2.3.9 Validation of Model

In order to validate each model, we included a 'randomised' negative control RF where relative abundances of each node were permuted with respect to each sample and the

predictive accuracy was assessed. In addition, we took the relative abundances of an important node for age and redistributed the abundance to only WT samples. The RF was repeated to investigate whether this node would appear as important for genotype. We also ran RF's with an increasing number of trees, using three different random seeds and performed Spearman's Rank correlation on the MDA values obtained among each set of three RFs of the same size. The Monod/Michaelis-Menten model was fitted, to determine how an increasing number of trees affected correlation of the MDA values.

2.3.10 Statistical Analysis

Analysis of the real vs null RF models predictive accuracy was performed using 2-way ANOVA, with a Sidak's post hoc test in GraphPad Prism 7 (GraphPad Software, La Jolla, USA). Significance of clustering in the NMDS was determined using permutational multivariate analysis of variance (PERMANOVA) via the 'Adonis' function in the 'vegan' R package (Oksanen et al., 2016).

2.4 Results

2.4.1 Phylogenetic tree of 16S rRNA data derived from the gut microbiota

We collected microbiota samples from the stools and colonic mucus of 40 male mice at two different time points (6 vs 18 weeks of age) and two genotypes (WT vs *mdr1a*$^{-/-}$) in an experiment designed to have co-housed siblings of different genotypes. Thus, littermates, irrespective of genotype were co-housed in 7 cages. The colonic mucus resident bacteria was visualised by FISH (Figure 1A). Pathology was assessed by Glymenaki et al. (2017), where 5, 18 week-old *mdr1a*$^{-/-}$ mice had indications of colitis with a loss of healthy gut architecture (data not shown) and the remaining mice were healthy. On average, 10442 16S sequences (range 1892-25681) were obtained per sample. All sequences were used to create a phylogenetic tree, in which the major phyla were separated (Supplementary Figure 1). We then assessed whether the sequences on the phylogenetic tree had separated by treatment group. Sequences derived from stool and mucus were clearly separated in the tree (Figure 1B) and other factors (age, genotype and cage) were more widely distributed across the tree (Supplementary Figure 2).

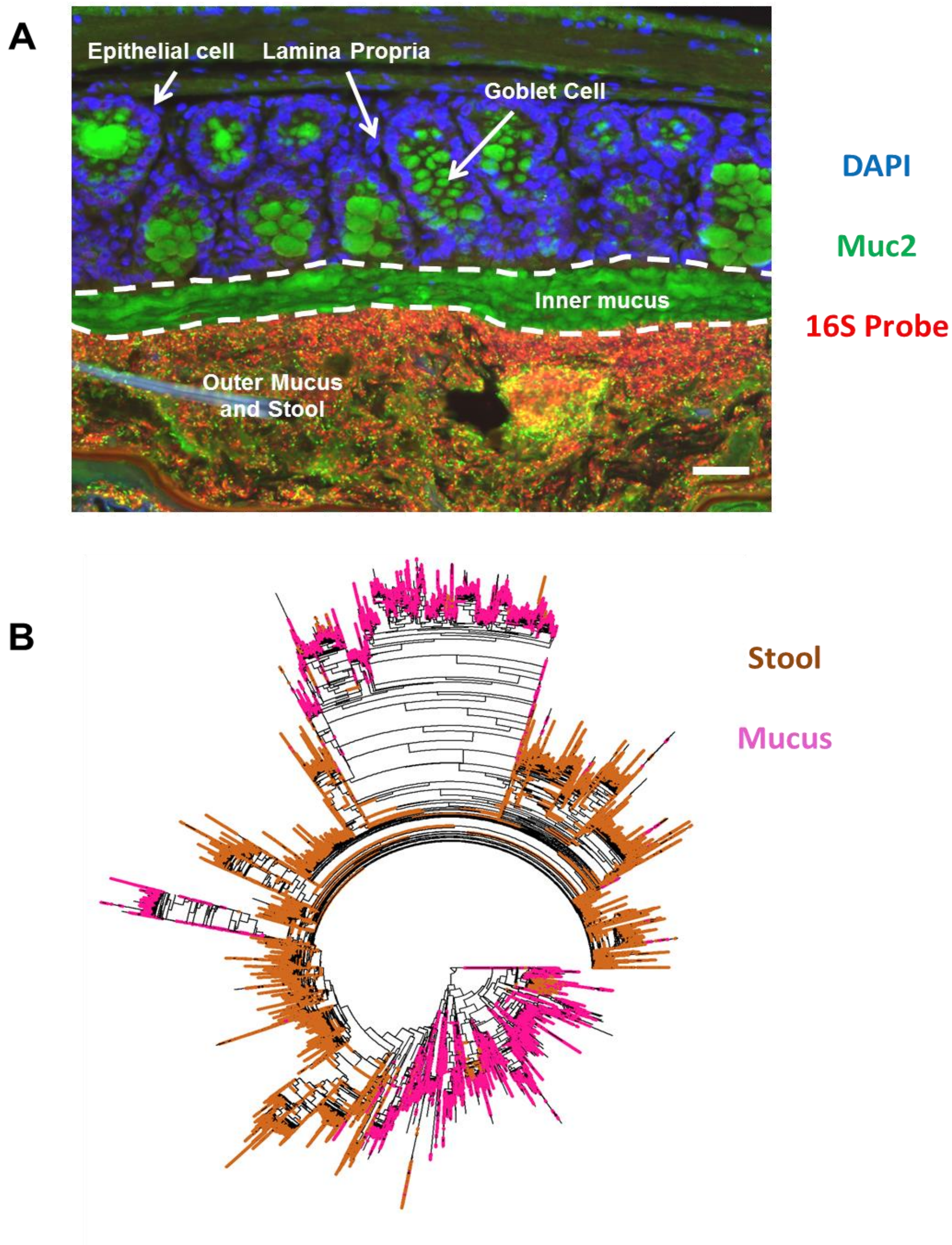


Figure 1: Distinct localisation of stool and mucus microbiota. Colonic tissue sections from a male wildtype (WT) mouse was stained with a fluorescent DNA probe specific for the 16S rRNA gene to identify bacteria (red), a Muc2 antibody (green) to identify mucus and counterstained with DAPI (blue) (A). A phylogenetic tree of 16S rRNA sequences derived from the gut microbiota of 6 and 18 week old, FVB background, WT mice and *mdr1a*$^{-/-}$ mice was plotted (B). Colours indicate stool (brown) and mucus (pink) -derived sequences, taken from n = 10-11 mice per genotype.

2.4.2 Strong separation of the gut microbiota by microbial niche, age and cage, but not host genotype.

To avoid bias by taxonomic level, we constructed a data matrix comprising the relative abundance ((number of tips in clade in sample) / (total number of tips in sample)) of all 428234 internal clades of the phylogenetic tree in each of our samples. This avoided assigning OTUs, or using a reference database. To visualise the major differences in the microbial communities, Bray-Curtis dissimilarity values were calculated between all samples and used as an input for a 2-dimensional non-metric multidimensional scaling (NMDS) ordination (Figure 2). As expected from their separation in the tree (Figure 1B) there was significant separation of samples by niche (Figure 2A) (PERMANOVA: $P < 0.0001$). There was also significant separation by mouse age (6 vs 18 weeks; PERMANOVA: $P < 0.0001$) (Figure 2B) which is in concordance with our previous work (Glymenaki et al., 2017). Samples from the same cage localised closely in the ordination, often with clear separation among cages (PERMANOVA: $P < 0.0001$). Nonetheless, cages containing different litters from the same mother were adjacent or overlapping, suggesting maternal effects influencing, but not fully explaining, cage-specific microbiomes (Figure 2C). No significant clustering was found when comparing genotypes (Figure 2D).

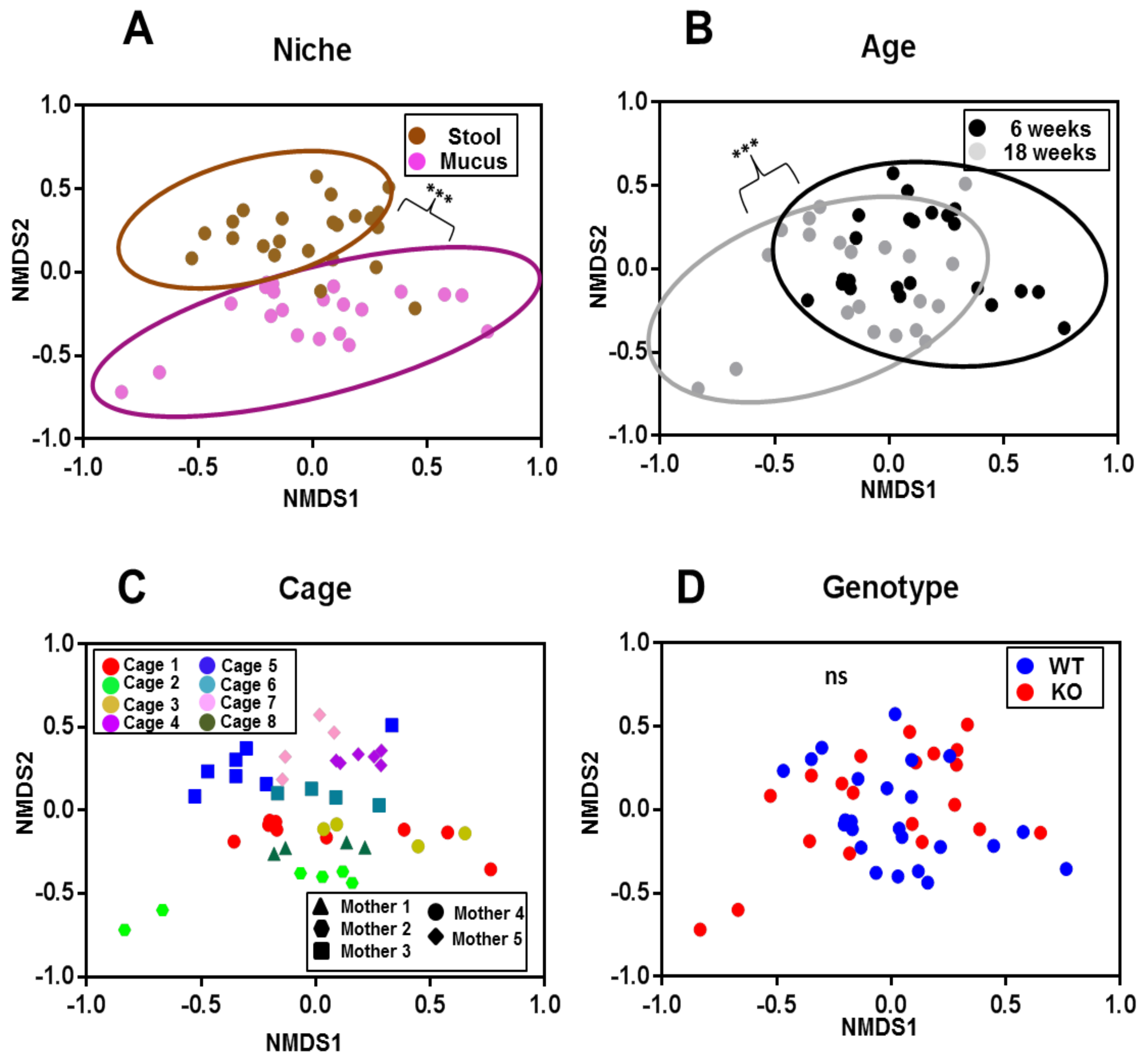


Figure 2: Significant separation of microbiota via NMDS for age and microbial niche. 16S rRNA sequences derived from the stools and mucus of 6 and 18 week old male mice (FVB background, wildtype vs *mdr1a*$^{-/-}$ mice) were used to construct a phylogenetic tree. A data matrix was established comparing samples and the relative abundance of taxonomic clades within that sample. Non-metric multidimensional scaling was performed using a Bray Curtis dissimilarity matrix based on the relative abundance of the clades. Plots highlighting 6 and 18 week samples (A), stool and mucus samples (B), different cages and mothers (C) and WT and KO samples (D) are illustrated. Asterisks represent significance determined by permutational multivariate analysis of variance: P = < 0.0001 (***), ns = non-significant. Each point corresponds to a stool or mucus sample taken from n = 10-11 mice per genotype.

2.4.3 The gut microbiota is strongly associated with age and microbial niche

To determine the taxa driving the observed differences in community structure, the relative abundance matrix was used to construct machine learning models (random forests, RFs). Separate RF models were created to identify age, genotype, niche and cage based on the relative abundance of the taxa (as defined by the phylogenetic tree, Figure 1B) in each sample. These models were compared against a null (negative control) model where relative abundances were permuted among taxa within samples to remove true associations. Niche could be determined from the microbiota with 90% accuracy, age with 98% accuracy and cage with ~80% accuracy, in all instances substantially higher than the negative control model (Figure 3A) (Two Way ANOVA- Sidak's post hoc test: $P < 0.0001$). Associations with genotype were comparable to that of the negative control i.e. no better than random with approximately 50% accuracy (Figure 3A), and will not be considered further.

The RFs give an importance value for each clade (internal nodes) in discriminating between groups. To identify which bacteria the clades encompassed, we used BLAST+ on all sequences, recording the taxonomic identity of the top hit (hits that had a percentage coverage <100% were discarded). The finest-scale taxonomic grouping containing each sequence descending from the five most important clades is shown in Figure 3B, C and D for niche, age and cage RFs respectively.

For microbial niche, the most important distinguishing clades were all Gram negative and mostly comprised Proteobacteria: the order Pseudomonadales and two clades in the families Burkholderiaceae and Deferribacteraceae (all these clades were prevalent in the mucus samples). The Deferribacteraceae containing clade mostly comprised *Mucispirillum*, a known mucus-associated bacteria (Rodriguez-Pineiro and Johansson, 2015). The genus *Porphyromonas* was associated with stool samples.

The most important clades separating ages were the families Erysipelotrichaceae and Lachnospiraceae within the Firmicutes phylum (which have each been specifically associated with young mouse microbiomes before (Kim et al., 2017)) plus three genera: *Natranaerovirga Desulfovibrio*, and *Vampirovibrio* in the Firmicutes, Proteobacteria and Cyanobacteria phylum respectively. With the exception of *Natranaerovirga*, all these bacteria were prevalent in the 18 week old mice.

The most important clades separating cages were *Natranaerovirga* (a different clade from that important for separating ages) plus four clades within the order Bacteroidales – two each in the genera *Bacteroides* and *Barnesiella*.

To validate the method for identifying important taxa, we took the most important clade identified for age (within the Erysipelotrichaceae family) and redistributed the abundances of this clade to only be present in all WT samples. Upon repeating the random forest separating on genotype, the Erysipelotrichaceae clade became the most important and this change was sufficient for the genotype RF to be significantly better than the null model (Two Way ANOVA- Sidak's post hoc test: $P < 0.05$) (Supplementary Figure 3). Similarly, repeating the random forest separating on age using the redistributed dataset, the Erysipelotrichaceae clade ceased to be the most important, leaving most of the other important clades relatively unaffected (Supplementary Figure 3). This data would suggest that important clades are robust and maintain their importance, even when other clades are altered.

We also used the Michaelis-Menten model to predict how an increasing number of trees affected correlation of the MDA values, with a maximum possible correlation of 0.82 (Supplementary Figure 4). With a forest of 100000 trees, we predicted that we had achieved 92% of this correlation, suggesting that 100000 trees is close to the optimal number of trees required for minimising data variability.

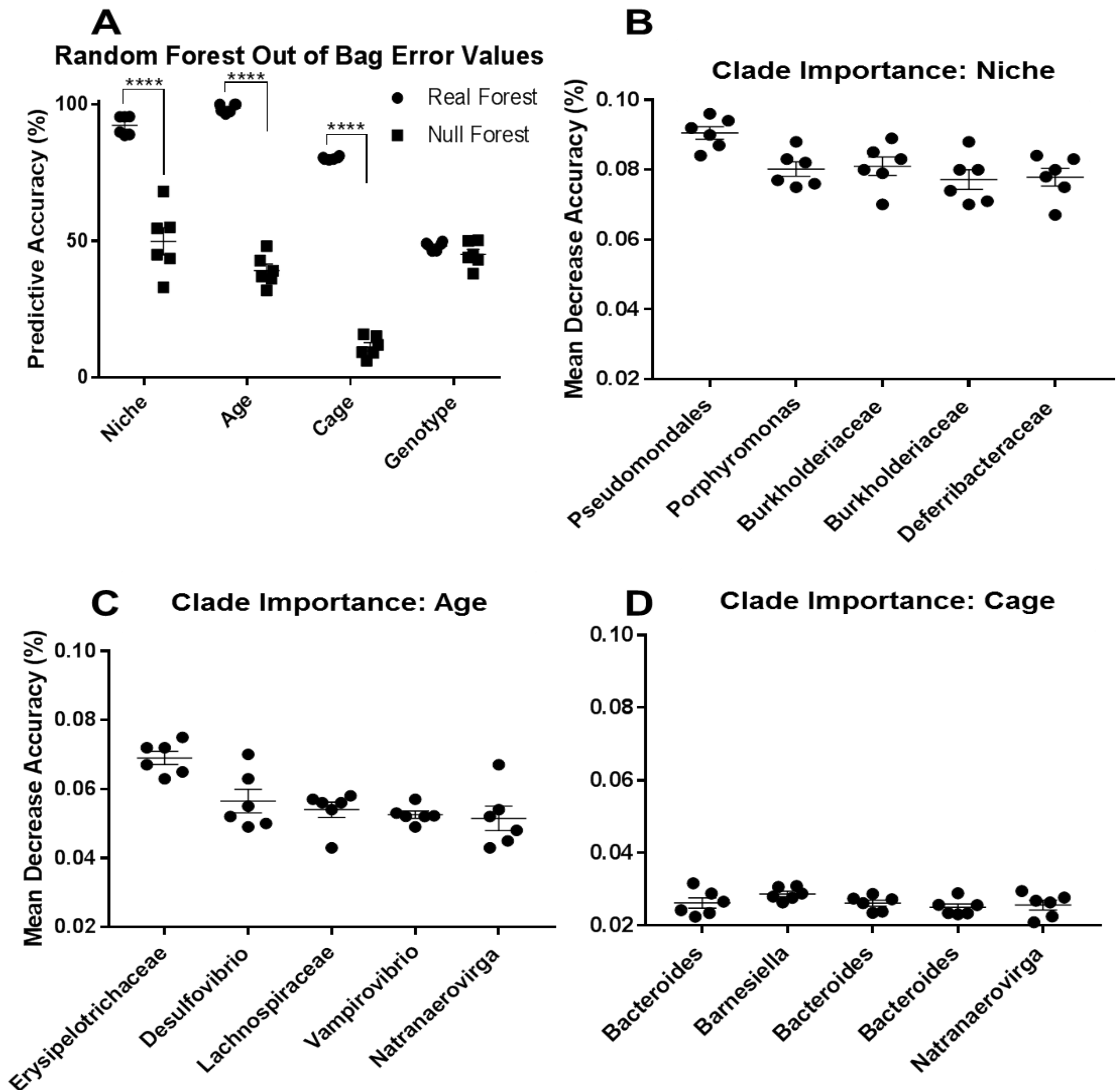


Figure 3: Random forest model identifies strong associations between the microbiota, niche, age and cage. 16S rRNA sequence data derived from the stools and mucus of 6 and 18 week old mice (FVB background, wildtype vs *mdr1a*$^{-/-}$ mice) were used to construct a phylogenetic tree (Figure 1B). A Random forest (RF) model of the relative abundance of phylogenetic clades was used to find associations between the gut microbiota and different experimental treatments. The predictive accuracy of the RF model at taking a sample and discriminating between the different treatment groups is shown (A). The five most important nodes associated with age (B) and microbial niche (C) are illustrated. Asterisks represent significance determined using Two Way ANOVA- Sidak's post hoc: p =< 0.00001 (****). n = 6.

2.4.4 Abundant, low-level taxa distinguish cage microbiomes but not age or niche.

Above we identified taxa at different taxonomic levels as particularly important for separating microbiomes. We therefore looked systematically at the phylogenetic scales that are important for separating different microbiomes. Clade importance is shown as a function of the distance from each clade to the root of the phylogenetic tree, referred to as 'taxonomic subdivision' (Figure 4). Clades with greater taxonomic subdivision (deeper clades) correlate with taxonomic levels such as genera, whereas clades with lower taxonomic subdivision (shallower clades) correlate with taxonomic levels such as phyla. For both age and niche, neither the deepest nor the shallowest clades are important (little separation from the null model), but the important clades are at intermediate depths in the tree (Figure 4 A-B). However, for differences among cages, while intermediate level clades are important, the most important groups are at the extreme of low level taxa, i.e. differences in sub-specific groupings (Figure 4C).

The number of sequences within a clade of the tree that are present in a particular set of samples is an estimate of its abundance in that microbiome. We therefore asked how abundance of bacterial taxa correlates with its importance in distinguishing microbiomes. We find that, for separating age, niche or cage microbiomes, moderately abundant taxa are important, whereas the rarest taxa are never important (Figure 4D-F). More interestingly, the most abundant taxa are important for distinguishing cage microbiomes, but are much closer to the null expectation for distinguishing microbiomes from different ages or niches.

For each forest (age, niche and cage), we compared the clade importance (MDA value) to determine whether clades in one forest were as important for another (Figure 5). Each treatment has clades that are exclusively important for it alone. There are also clades that have moderate importance for all treatment groups (Figure 5A, B and C). However, no clade has particularly high importance for more than one treatment group. The null forest had a similar spread of importances to the real forest (Figure 5D, E and F). Therefore, the data would suggest that only clades with the greatest MDA values can be reliably associated with their respective treatment group (for age > 0.02, for niche > 0.03, for cage > 0.0007).

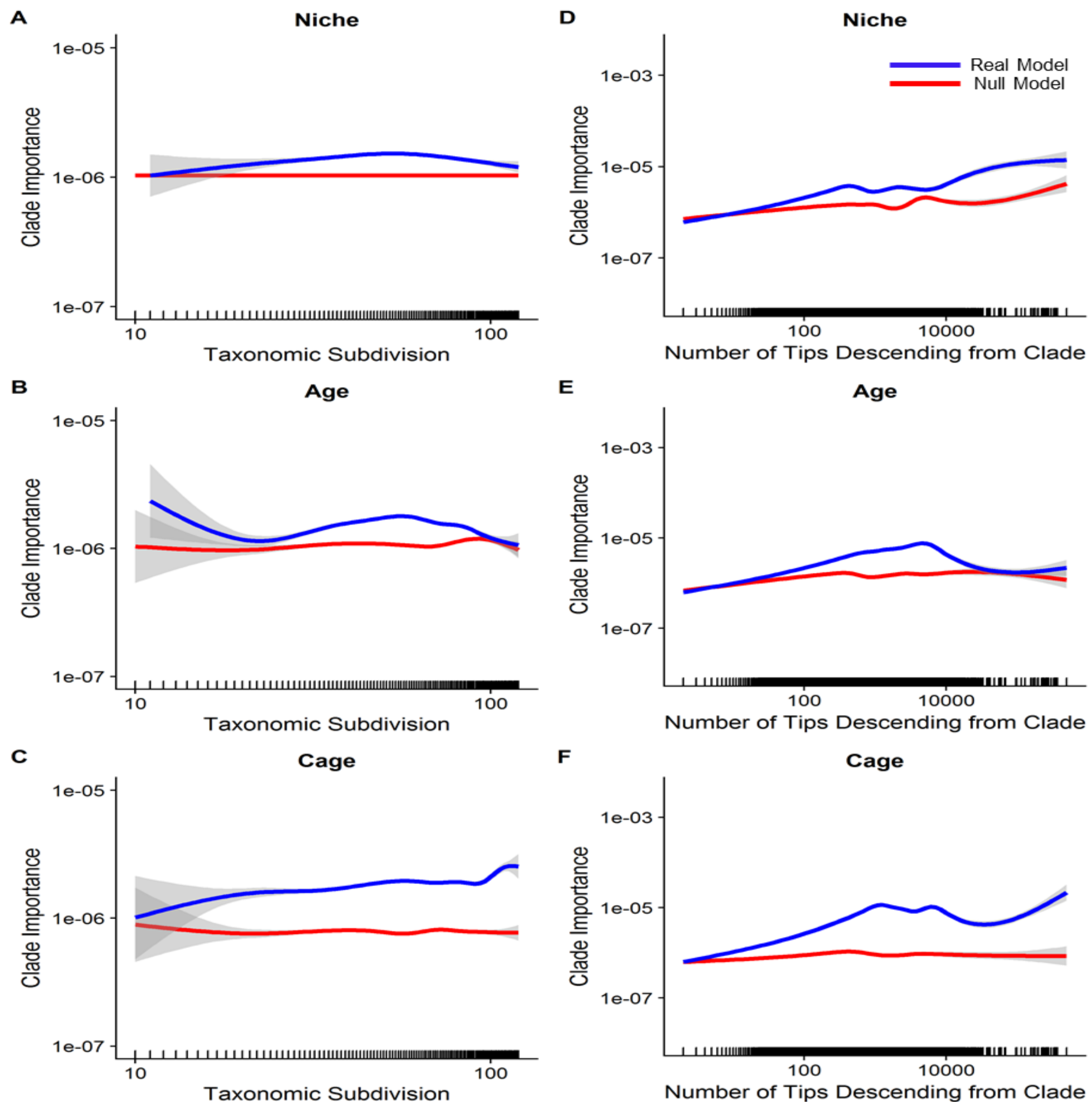


Figure 4: Abundant, low-level taxa distinguish cage microbiomes but not age or niche. 16S rRNA sequence data derived from the stools and mucus of 6 and 18 week old mice (FVB background, wildtype vs *mdr1a*$^{-/-}$ mice) were used to construct a phylogenetic tree A data matrix was established comparing samples and the relative abundance of taxonomic clades within that sample. This matrix was used as an input for a Random Forest (RF) model to find associations between the gut microbiota and different experimental treatments. The depth of a clade was compared against the 'mean decrease accuracy' (MDA) value when running a forest that compared the niche (A), age (B) and cage (C). The number of tips descending from each clade was plotted against its MDA value, when running a forest that compared the niche (D), age (E) and cage (F). The 'real' random forest model is illustrated in blue and a null (negative control) random forest model is illustrated in red. n = 6.

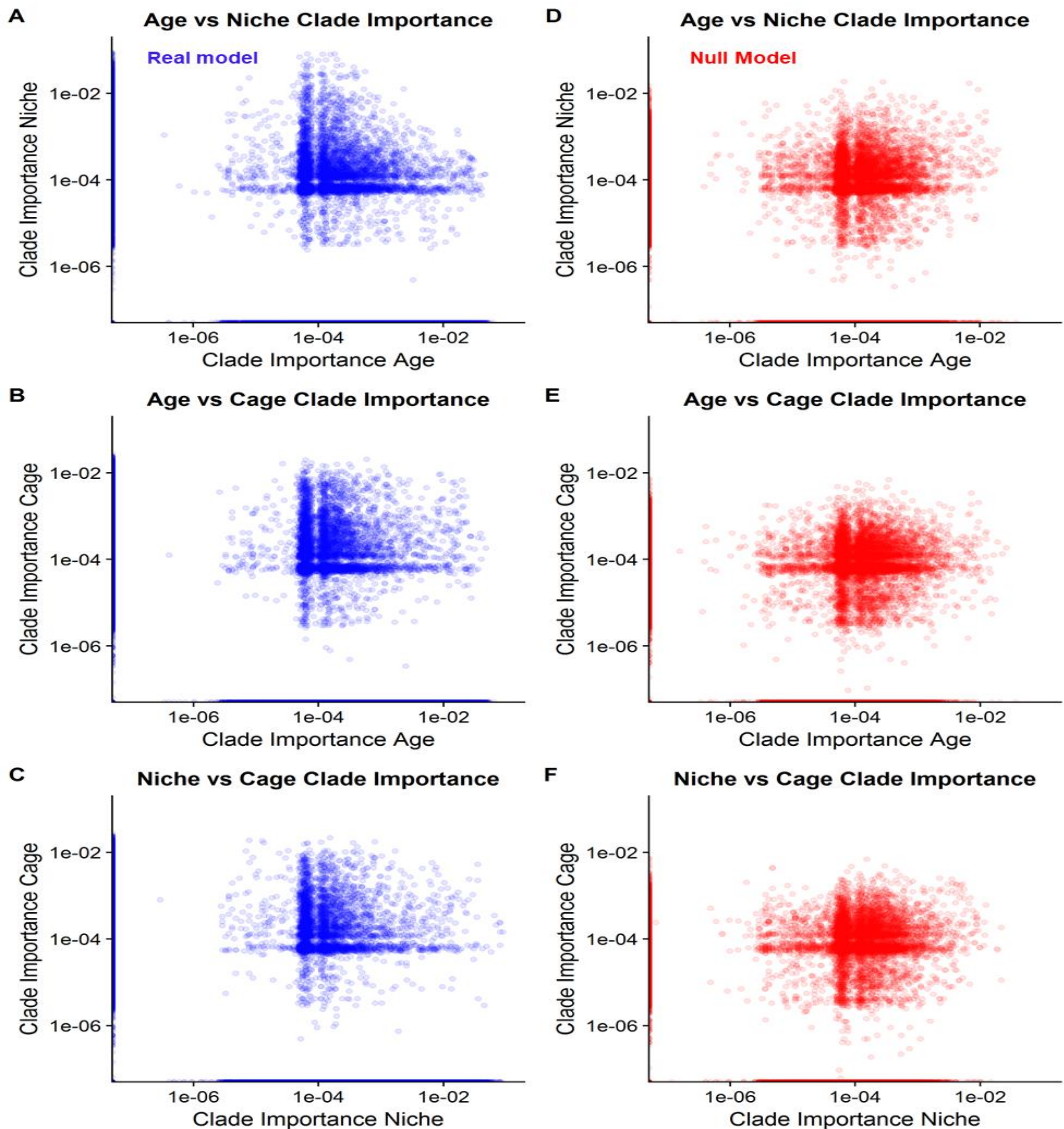


Figure 5: Clade importance between forests. 16S rRNA sequence data derived from the stools and mucus of 6 and 18 week old mice (FVB background, wildtype vs *mdr1a*$^{-/-}$ mice) were used to construct a phylogenetic tree A data matrix was established comparing samples and the relative abundance of taxonomic clades within that sample. This matrix was used as an input for a random forest (RF) model to find associations between the gut microbiota and different experimental treatments. The importance (Mean Decrease Accuracy value) of a clade in one forest was plotted against its importance in another forest for the real model (blue): Age vs Niche (A), Age vs Cage (B), Niche vs Cage (C) and the null (negative control, red) model: Age vs Niche (D), Age vs Cage (E) and Niche vs Cage (F). n = 6

2.5 Discussion

Our modelling approach was able to discriminate clearly between the microbiomes of 6 and 18 week old mice and between mucus and stool samples. This is consistent with others' work, for instance showing substantial microbiome changes with age in both humans and mice (Langille et al., 2014; Odamaki et al., 2016) and work identifying microbial niche as the strongest factor for separation of the microbiota (Eckburg et al., 2005; Glymenaki et al., 2017). Our mice are still relatively young, initial samples taken only ~2 weeks after weaning and at 18 weeks old. Therefore the difference may be due to the microbiota adjusting due to changed diet. Nonetheless, microbial changes in mice can happen within as little as two weeks (Kozik, 2017). The fact that we can see these differences clearly and identify the key taxa responsible validates our approach to modelling these microbiome changes using machine-learning tools (random forests) in conjunction with clades of microbes defined, phylogenetically, without reference to external taxonomies or OTU thresholds. It is therefore striking that even this apparently powerful approach can find no consistent differences between the gut microbiomes of co-housed wildtype and colitis-prone (*mdr1a*$^{-/-}$) genotypes.

Differences in the microbiota of WT and *mdr1a*$^{-/-}$ mice *have* been reported (Nones et al., 2009; Glymenaki et al., 2017). The fact that we do not see them here is therefore puzzling. Discrepancies in sample size between treatment groups can be a problem for RFs applied to such data (Walters et al., 2014). However, here sample sizes are well balanced (20 wildtype and 20 *mdr1a*$^{-/-}$ mice). Some, but not all, of the older *mdr1a*$^{-/-}$ mice were starting to develop colitis. Therefore, changes in the microbiome with the onset of colitis may have obscured any consistent differences among genotypes. Alternatively, any machine learning approach used on such a dataset could be under-powered with too small a sample size to identify subtle differences, or that previous analyses were compounded by large cage effects (Figure 2D) into erroneously attributing some of that variation to differences among genotypes. Whatever the explanation, it is clear that, in comparison to differences between gut age and niche, the mouse gut microbiome is relatively robust to this host genotype change affecting the gut.

The majority of the gut microbiota fall within a small range of phyla, with Firmicutes and Bacteroidetes making up the largest proportion and Actinobacteria, Proteobacteria, Fusobacteria and Verrucomicrobia a smaller proportion (Candela et al., 2010). Shifts in the proportions of phyla can have substantial physiological effects on the host, for example increased Firmicutes are associated with increased incidence of obesity (Turnbaugh et al., 2006; Turnbaugh et al., 2009). In the context of IBD, numerous phylum level changes have

been implicated in the progression of inflammation. A reduction in the abundance and diversity of Firmicutes is associated with IBD in human patients (Frank et al., 2007; Ott et al., 2004; Manichanh et al., 2006) and Bacteroidetes has been shown to be both increased (Walker et al., 2011) and decreased (Frank et al., 2007). However, our data did not find such large-scale taxa to show consistent differences in any of our microbiome comparisons. These high-level taxa are also abundant taxa and, *a priori*, it might have been reasonable to expect that the more abundant taxa would have the most important functional consequences for the host and therefore be most likely to differ between different circumstances. However, the only microbiome comparison in which we find the most abundant taxa to be important is in distinguishing among cages. Even here, it is low-level taxonomic groupings (e.g. clades within the abundant genus *Bacteroides*), not phyla that distinguished among cage-specific microbiomes.

Conversely, rare species have been suggested to play a large role a range of ecosystems, including the host and the environment (Shade et al., 2014; Jousset et al., 2017) and rare taxa can lead to gut inflammation (Powell et al., 2012). We have previously shown that lower level taxonomic changes can also have functional significance, with the genus *Pseudomonas* leading to a delay in wound repair (Williams et al., 2017). Our analysis here does not find the rarest taxa to be important in discriminating between microbiomes. This however is likely to be an artefact of the fact that, almost by definition in a complex microbiome, rare taxa are likely to be missed from at least a subset of samples, through random sampling, and therefore not show consistent differences among the factors considered (age, niche or cage).

Several specific taxa we highlighted as being important in distinguishing microbiomes included taxa that have been associated with differences in gut microbiomes before. The role of the family Erysipelotrichaceae, which we implicate in distinguishing older mice from younger, has been associated with colorectal cancer (Chen et al., 2012) and IBD in an infection-induced mouse model of colitis (Craven et al., 2012) supporting the idea that Erysipelotrichaceae are involved in inflammation development. However, this bacterial family was significantly decreased in a murine model of colitis driven by tumour necrosis factor (TNF) (Schaubeck et al., 2016) and therefore it may not be that clear cut. It may be that the presence or absence of certain taxa allows other families/species to flourish or be inhibited, thus altering the host/microbial homeostasis and driving inflammation. Our RF models can account for such interactions among taxa, and the ‘importance’ assigned to a

taxon takes these into account. However, unpicking these relationships is a challenge and doing so would require more direct experimental manipulation of the microbiome.

In distinguishing between microbial stool versus mucus niches, we did not identify typical lumen/faecal associated bacteria such as *Ruminococcus,* as important (Jandhyala et al., 2015). The most important clade identified was one that encompassed bacteria within the order Pseudomonadales. This order includes genera such as *Pseudomonas and Acinetobacter*. *Acinetobacter* species are associated with the colonic mucus (Pedron et al., 2012) and therefore these taxa could be good markers of the mucus microbiota. However, it notable that another common mucus bacteria, *Akkermansia muciniphila,* (Berry et al., 2013) which was identified in this study, did not appear as one of the important taxa, suggesting its occurrence was more variable among samples. Indeed our previous work suggested this bacteria only became prevalent in inflamed guts (Glymenaki et al., 2017). The family Deferribactericeae also distinguished mucus and stool. This family contains the genus *Mucispirillum* which are known to be mucus-associated bacteria (Robertson et al., 2005). Together these findings validate our approach, showing it capable of identifying particular taxa that distinguish gut niches based on their importance, in a more nuanced way than traditional correlative methods.

RF models can be used to address clear questions about the microbiome, while also taking account of its complexity (Walters et al., 2014), showing discrimination between lean and obese subjects, where simple summary statistics such as the ratio of Firmicutes to Bacteroidetes was not. Similarly, RF was used to discriminate between patients with active Crohn's disease and those in remission with ~70% accuracy (Tedjo et al., 2016). By building a phylogenetic tree of the sequences and using the full range of clades in that tree as explanatory variables in the RF model, we can now develop this idea further. We are able to identify particular clades as important whatever taxonomic level they occur at, in an equivalent, data-driven way and we can ask what big-scale patterns exist in the relative importance of clades at different scales and abundances. A risk of our approach is losing the connection to specific microbial taxa. However, simple post-hoc similarity searches of the sequences involved were effective at naming key taxa involved, demonstrating both the similarities (e.g. the importance of *Natranaerovirga* in distinguishing both cages and ages) and the differences (e.g. that despite the importance of similar phylogenetic levels and taxon abundances in distinguishing niches and ages, Proteobacteria clades were most important for the former and Firmicutes clades for the latter).

Building phylogenetic trees only using the information in amplicons from subsets of the 16S rRNA is very restrictive. Even full 16S-based trees of bacteria do not fully capture their evolutionary history (Vetrovsky and Baldrian, 2013), and our tree does not fully capture the topology of more thorough 16S-based trees (Woese et al., 1990). Nonetheless, taking this restrictive approach ensures that the power of the data is fully used without attempting to shoe-horn it into a pre-existing framework. The approach here avoids the risks both of over-stretching the data (e.g. assigning a sequence read to one taxon rather than another when it is in fact similarly very close or very distant to both) and of losing power that is in the data (e.g. clear phylogenetic structure among sequences that are closer than a given threshold, typically 97% identity). While the tree constructed is undoubtedly an imperfect representation of the evolutionary history of these sequences, it is much more nuanced than a simple taxonomy and, as with other forms of analysis, acknowledging the evolutionary relationships between organisms, even with an inadequate tree, is much better than ignoring that structure.

In conclusion, taking a carefully designed factorial experiment involving co-housing of different mice with genotypes that affect their susceptibility to IBD, we have been able to identify major changes in the gut microbiome with age, niches and cages, but not genotype. Our machine learning approach, focused on phylogenetically related groups at all evolutionary scales, proved effective, not only in identifying distinct versus homeostatic microbiomes, but in identifying the phylogenetic groupings important in making distinctions. Furthermore, this approach revealed differences in the patterns of phylogenetic groupings (high or low level, rare or abundant taxa) that distinguish different microbiome features. Together, this work reveals the subtlety of the balance between homeostasis and difference in the gut microbiome that can be used to help better define the host interaction with the microbiome in a range of conditions.

2.6 References

Ahmed, S., Macfarlane, G. T., Fite, A., McBain, A. J., Gilbert, P. & Macfarlane, S. (2007). Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol,* 73**,** 7435-42.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., de Vos, W. M., Brunak, S., Dore, J., Antolin, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K. U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S. D. & Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature,* 473**,** 174-80.

Balvociute, M. & Huson, D. H. (2017). SILVA, RDP, Greengenes, NCBI and OTT - how do these taxonomies compare? *BMC Genomics,* 18(Suppl 2)**,** 114.

Berry, D., Stecher, B., Schintlmeister, A., Reichert, J., Brugiroux, S., Wild, B., Wanek, W., Richter, A., Rauch, I., Decker, T., Loy, A. & Wagner, M. (2013). Host-compound foraging by intestinal microbiota revealed by single-cell stable isotope probing. *Proc Natl Acad Sci U S A,* 110(12)**,** 4720-5.

Bouskra, D., Brezillon, C., Berard, M., Werts, C., Varona, R., Boneca, I. G. & Eberl, G. (2008). Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature,* 456(7221)**,** 507-10.

Bramhall, M., Florez-Vargas, O., Stevens, R., Brass, A. & Cruickshank, S. (2015). Quality of methods reporting in animal models of colitis. *Inflamm Bowel Dis,* 21(6)**,** 1248-59.

Breitwieser, F. P., Lu, J. & Salzberg, S. L. (2017). A review of methods and databases for metagenomic classification and assembly. *Brief Bioinform*.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.

Candela, M., Consolandi, C., Severgnini, M., Biagi, E., Castiglioni, B., Vitali, B., De Bellis, G. & Brigidi, P. (2010). High taxonomic level fingerprint of the human intestinal microbiota by ligase detection reaction--universal array approach. *BMC Microbiol,* 10**,** 116.

Chamberlain, S. A. & Szocs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Res,* 2**,** 191.

Chen, W., Liu, F., Ling, Z., Tong, X. & Xiang, C. (2012). Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS One,* 7(6)**,** e39743.

Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R. & Nicholson, J. K. (2009). Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc Natl Acad Sci U S A,* 106**,** 14728-33.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R. & Tiedje, J. M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res,* 42(Database issue)**,** D633-42.

Cox, M. P., Peterson, D. A. & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics,* 11**,** 485.

Craven, M., Egan, C. E., Dowd, S. E., McDonough, S. P., Dogan, B., Denkers, E. Y., Bowman, D., Scherl, E. J. & Simpson, K. W. (2012). Inflammation drives dysbiosis and bacterial invasion in murine models of ileal Crohn's disease. *PLoS One,* 7(7)**,** e41594.

Csard, G. & Nepusz, T. (2006). *The igraph software package for complex network research* [Online]. Available: http://igraph.org [Accessed 31st August 2018].

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E. & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science,* 308(5728)**,** 1635-8.

Favier, C., Neut, C., Mizon, C., Cortot, A., Colombel, J. F. & Mizon, J. (1997). Fecal beta-D-galactosidase production and Bifidobacteria are decreased in Crohn's disease. *Dig Dis Sci,* 42(4)**,** 817-22.

Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N. & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A,* 104(34)**,** 13780-5.

Giaffer, M. H., Holdsworth, C. D. & Duerden, B. I. (1991). The assessment of faecal flora in patients with inflammatory bowel disease by a simplified bacteriological technique. *J Med Microbiol,* 35(4)**,** 238-43.

Glymenaki, M., Singh, G., Brass, A., Warhurst, G., McBain, A. J., Else, K. J. & Cruickshank, S. M. (2017). Compositional Changes in the Gut Mucus Microbiota Precede the Onset of Colitis-Induced Inflammation. *Inflamm Bowel Dis,* 23(6)**,** 912-922.

Goslee, S. C. & Urban, D. L. (2007). The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *22,* 22(7)**,** 1-19.

Hiergeist, A., Reischl, U. & Gessner, A. (2016). Multicenter quality assessment of 16S ribosomal DNA-sequencing for microbiome analyses reveals high inter-center variability. *Int J Med Microbiol,* 306(5)**,** 334-342.

Hildebrand, F., Nguyen, T. L., Brinkman, B., Yunta, R. G., Cauwe, B., Vandenabeele, P., Liston, A. & Raes, J. (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol,* 14(1)**,** R4.

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C. & Banfield, J. F. (2016). A new view of the tree of life. *Nat Microbiol,* 1**,** 16048.

Imhann, F., Vich Vila, A., Bonder, M. J., Fu, J., Gevers, D., Visschedijk, M. C., Spekhorst, L. M., Alberts, R., Franke, L., van Dullemen, H. M., Ter Steege, R. W. F., Huttenhower, C., Dijkstra, G., Xavier, R. J., Festen, E. A. M., Wijmenga, C., Zhernakova, A. & Weersma, R. K. (2018). Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut,* 67(1)**,** 108-119.

Jandhyala, S. M., Talukdar, R., Subramanyam, C., Vuyyuru, H., Sasikala, M. & Nageshwar Reddy, D. (2015). Role of the normal gut microbiota. *World J Gastroenterol,* 21(29)**,** 8787-803.

Jousset, A., Bienhold, C., Chatzinotas, A., Gallien, L., Gobet, A., Kurm, V., Kusel, K., Rillig, M. C., Rivett, D. W., Salles, J. F., van der Heijden, M. G., Youssef, N. H., Zhang, X., Wei, Z. & Hol, W. H. (2017). Where less may be more: how the rare biosphere pulls ecosystems strings. *Isme j,* 11(4)**,** 853-862.

Kalliomaki, M., Collado, M. C., Salminen, S. & Isolauri, E. (2008). Early differences in fecal microbiota composition in children may predict overweight. *Am J Clin Nutr,* 87(3)**,** 534-8.

Kim, Y. G., Sakamoto, K., Seo, S. U., Pickard, J. M., Gillilland, M. G., 3rd, Pudlo, N. A., Hoostal, M., Li, X., Wang, T. D., Feehley, T., Stefka, A. T., Schmidt, T. M., Martens, E. C., Fukuda, S., Inohara, N., Nagler, C. R. & Nunez, G. (2017). Neonatal acquisition of Clostridia species protects against colonization by bacterial pathogens. *Science,* 356(6335)**,** 315-319.

Knights, D., Ward, T. L., McKinlay, C. E., Miller, H., Gonzalez, A., McDonald, D. & Knight, R. (2014). Rethinking "enterotypes". *Cell Host Microbe,* 16(4)**,** 433-7.

Kozik, A. J. (2017). Sex, Age, and TNF Influence the Gut Microbiota in a Mouse Model of TNBS Colitis. *FASEB J.*

Langille, M. G., Meehan, C. J., Koenig, J. E., Dhanani, A. S., Rose, R. A., Howlett, S. E. & Beiko, R. G. (2014). Microbial shifts in the aging mouse gut. *Microbiome,* 2**,** 50.

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D. & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A,* 102(31)**,** 11070-5.

Li, H., Limenitakis, J. P., Fuhrer, T., Geuking, M. B., Lawson, M. A., Wyss, M., Brugiroux, S., Keller, I., Macpherson, J. A., Rupp, S., Stolp, B., Stein, J. V., Stecher, B., Sauer, U., McCoy, K. D. & Macpherson, A. J. (2015). The outer mucus layer hosts a distinct intestinal microbial niche. *Nat Commun,* 6**,** 8292.

Liaw, A. & Wiener, M. (2002). Classification and Regression by randomForest. *R News,* 2(3)**,** 18-22.

Lozupone, C. & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol,* 71**,** 8228-35.

Macfarlane, G. T. & Macfarlane, S. (2012). Bacteria, colonic fermentation, and gastrointestinal health. *J AOAC Int,* 95**,** 50-60.

Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., Ballet, V., Claes, K., Van Immerseel, F., Verbeke, K., Ferrante, M., Verhaegen, J., Rutgeerts, P. & Vermeire, S. (2014). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut,* 63(8)**,** 1275-83.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J. & Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut,* 55(2)**,** 205-11.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. & Edwards, R. A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics,* 9**,** 386.

Nawrocki, E. P. & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics,* 29(22)**,** 2933-5.

Nones, K., Knoch, B., Dommels, Y. E., Paturi, G., Butts, C., McNabb, W. C. & Roy, N. C. (2009). Multidrug resistance gene deficient (mdr1a-/-) mice have an altered caecal microbiota that precedes the onset of intestinal inflammation. *J Appl Microbiol,* 107**,** 557-66.

Odamaki, T., Kato, K., Sugahara, H., Hashikura, N., Takahashi, S., Xiao, J. Z., Abe, F. & Osawa, R. (2016). Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiol,* 16**,** 90.

Oksanen, J., Guillaume-Blanchet, F., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H. & Wagner, H. (2016). *vegan: Community Ecology Package* [Online]. Available: https://CRAN.R-project.org/package=vegan [Accessed 27th October 2017].

Ott, S. J., Musfeldt, M., Wenderoth, D. F., Hampe, J., Brant, O., Folsch, U. R., Timmis, K. N. & Schreiber, S. (2004). Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut,* 53(5)**,** 685-93.

Papa, E., Docktor, M., Smillie, C., Weber, S., Preheim, S. P., Gevers, D., Giannoukos, G., Ciulla, D., Tabbaa, D., Ingram, J., Schauer, D. B., Ward, D. V., Korzenik, J. R., Xavier, R. J., Bousvaros, A. & Alm, E. J. (2012). Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS One,* 7(6)**,** e39242.

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics,* 20**,** 289-290.

Pedron, T., Mulet, C., Dauga, C., Frangeul, L., Chervaux, C., Grompone, G. & Sansonetti, P. J. (2012). A crypt-specific core microbiota resides in the mouse colon. *MBio,* 3(3).

Powell, N., Walker, A. W., Stolarczyk, E., Canavan, J. B., Gokmen, M. R., Marks, E., Jackson, I., Hashim, A., Curtis, M. A., Jenner, R. G., Howard, J. K., Parkhill, J., MacDonald, T. T. & Lord, G. M. (2012). The transcription factor T-bet regulates intestinal inflammation mediated by interleukin-7 receptor+ innate lymphoid cells. *Immunity,* 37(4)**,** 674-84.

Price, M. N., Dehal, P. S. & Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One,* 5(3)**,** e9490.

Ramirez, K. S., Knight, C. G., de Hollander, M., Brearley, F. Q., Constantinides, B., Cotton, A., Creer, S., Crowther, T. W., Davison, J., Delgado-Baquerizo, M., Dorrepaal, E., Elliott, D. R., Fox, G., Griffiths, R. I., Hale, C., Hartman, K., Houlden, A., Jones, D. L., Krab, E. J., Maestre, F. T., McGuire, K. L., Monteux, S., Orr, C. H., van der Putten, W. H., Roberts, I. S.,

Robinson, D. A., Rocca, J. D., Rowntree, J., Schlaeppi, K., Shepherd, M., Singh, B. K., Straathof, A. L., Bhatnagar, J. M., Thion, C., van der Heijden, M. G. A. & de Vries, F. T. (2018). Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nature Microbiology,* 3(2)**,** 189-196.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing* [Online].Available: https://www.r-project.org/ [Accessed 16th January 2017].

Ridaura, V. K., Faith, J. J., Rey, F. E., Cheng, J., Duncan, A. E., Kau, A. L., Griffin, N. W., Lombard, V., Henrissat, B., Bain, J. R., Muehlbauer, M. J., Ilkayeva, O., Semenkovich, C. F., Funai, K., Hayashi, D. K., Lyle, B. J., Martini, M. C., Ursell, L. K., Clemente, J. C., Van Treuren, W., Walters, W. A., Knight, R., Newgard, C. B., Heath, A. C. & Gordon, J. I. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science,* 341(6150)**,** 1241214.

Robertson, B. R., O'Rourke, J. L., Neilan, B. A., Vandamme, P., On, S. L., Fox, J. G. & Lee, A. (2005). Mucispirillum schaedleri gen. nov., sp. nov., a spiral-shaped bacterium colonizing the mucus layer of the gastrointestinal tract of laboratory rodents. *Int J Syst Evol Microbiol,* 55**,** 1199-204.

Rodriguez-Pineiro, A. M. & Johansson, M. E. (2015). The colonic mucus protection depends on the microbiota. *Gut Microbes,* 6(5)**,** 326-30.

Schaubeck, M., Clavel, T., Calasan, J., Lagkouvardos, I., Haange, S. B., Jehmlich, N., Basic, M., Dupont, A., Hornef, M., von Bergen, M., Bleich, A. & Haller, D. (2016). Dysbiotic gut microbiota causes transmissible Crohn's disease-like ileitis independent of failure in antimicrobial defence. *Gut,* 65(2)**,** 225-37.

Schinkel, A. H., Smit, J. J., van Tellingen, O., Beijnen, J. H., Wagenaar, E., van Deemter, L., Mol, C. A., van der Valk, M. A., Robanus-Maandag, E. C., te Riele, H. P. & et al. (1994). Disruption of the mouse mdr1a P-glycoprotein gene leads to a deficiency in the blood-brain barrier and to increased sensitivity to drugs. *Cell,* 77**,** 491-502.

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics,* 27(4)**,** 592-593.

Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N. & Gilbert, J. A. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio,* 5(4)**,** e01371-14.

Song, S. J., Lauber, C., Costello, E. K., Lozupone, C. A., Humphrey, G., Berg-Lyons, D., Caporaso, J. G., Knights, D., Clemente, J. C., Nakielny, S., Gordon, J. I., Fierer, N. & Knight, R. (2013). Cohabiting family members share microbiota with one another and with their dogs. *Elife,* 2**,** e00458.

St John, J. (2018). *SeqPrep* [Online]. Available: https://github.com/jstjohn/SeqPrep [Accessed 1st August 2018].

Tedjo, D. I., Smolinska, A., Savelkoul, P. H., Masclee, A. A., van Schooten, F. J., Pierik, M. J., Penders, J. & Jonkers, D. M. (2016). The fecal microbiota as a biomarker for disease activity in Crohn's disease. *Sci Rep,* 6**,** 35216.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A.

C., Knight, R. & Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature,* 457(7228)**,** 480-4.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., Magrini, V., Mardis, E. R. & Gordon, J. I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature,* 444**,** 1027-31.

Venables, W. & Ripley, N. (2002). *Modern Applied Statistics with S.* New York: Fourth Edition.

Vetrovsky, T. & Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One,* 8**,** e57923.

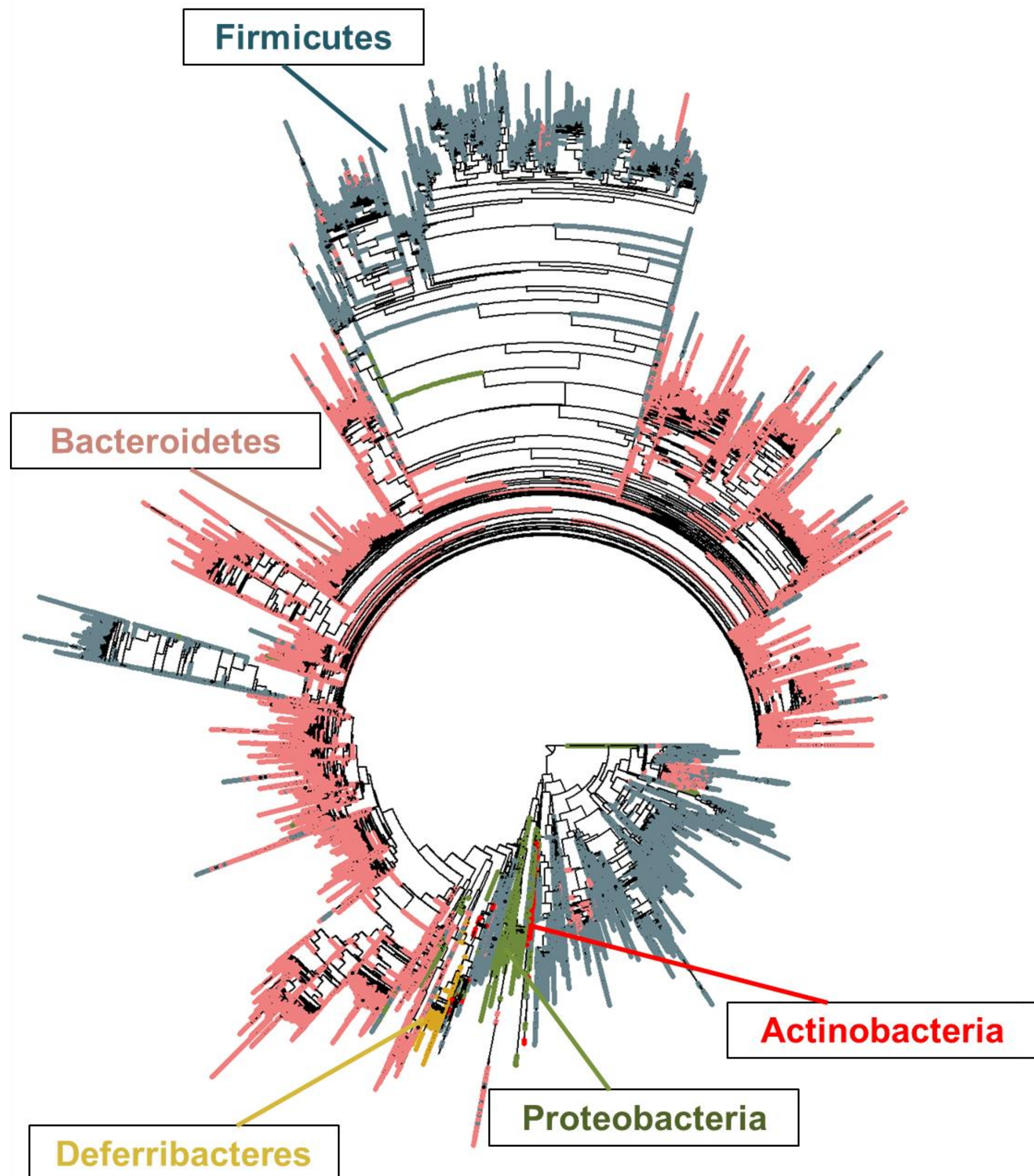
Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., Brostoff, J., Parkhill, J., Dougan, G. & Petrovska, L. (2011). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol,* 11**,** 7.

Walters, W. A., Xu, Z. & Knight, R. (2014). Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett,* 588(22)**,** 4223-33.

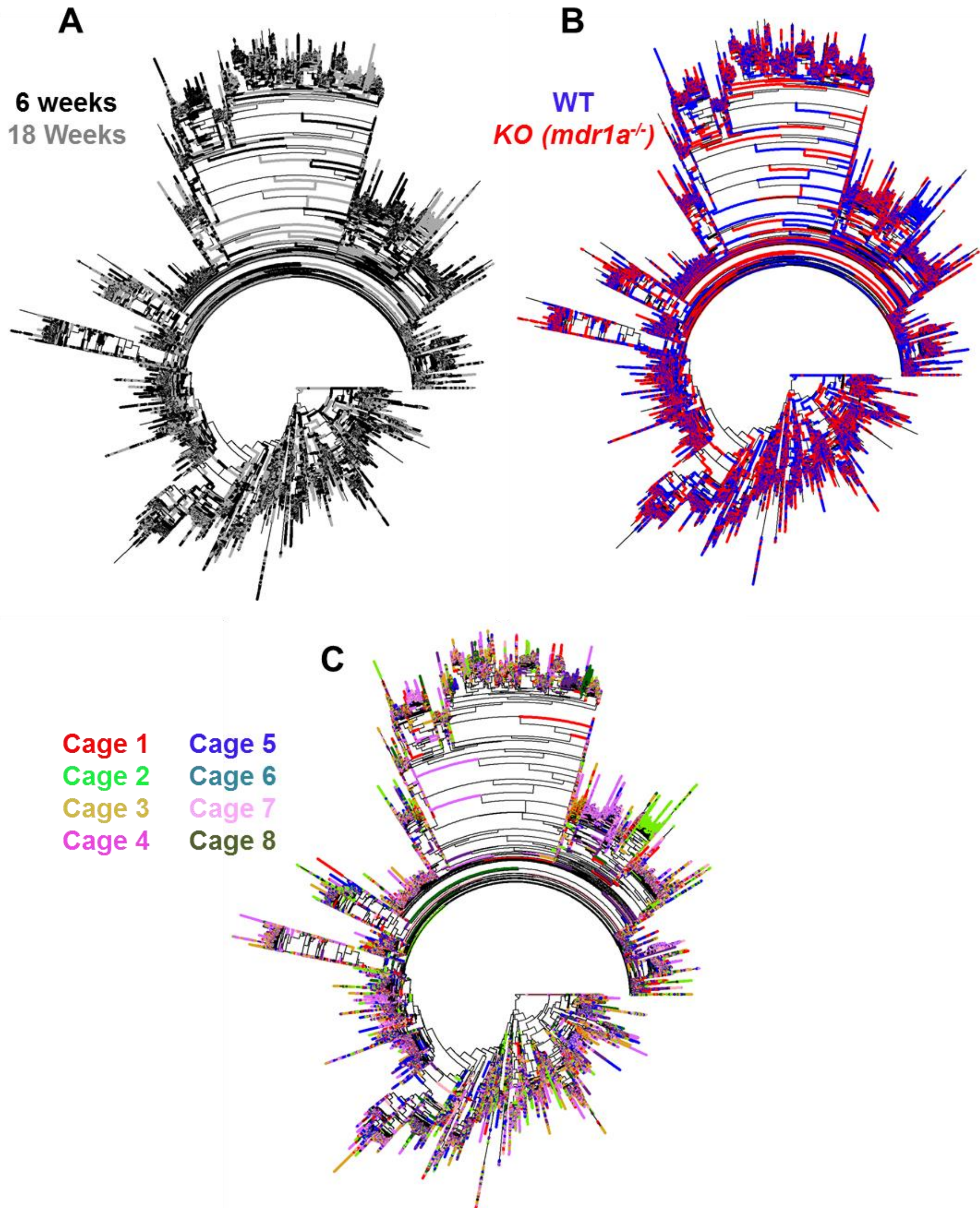
Williams, H., Crompton, R. A., Thomason, H. A., Campbell, L., Singh, G., McBain, A. J., Cruickshank, S. M. & Hardman, M. J. (2017). Cutaneous Nod2 Expression Regulates the Skin Microbiome and Wound Healing in a Murine Model. *J Invest Dermatol,* 137(11)**,** 2427-2436.

Woese, C. R., Kandler, O. & Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A,* 87(12)**,** 4576-9.

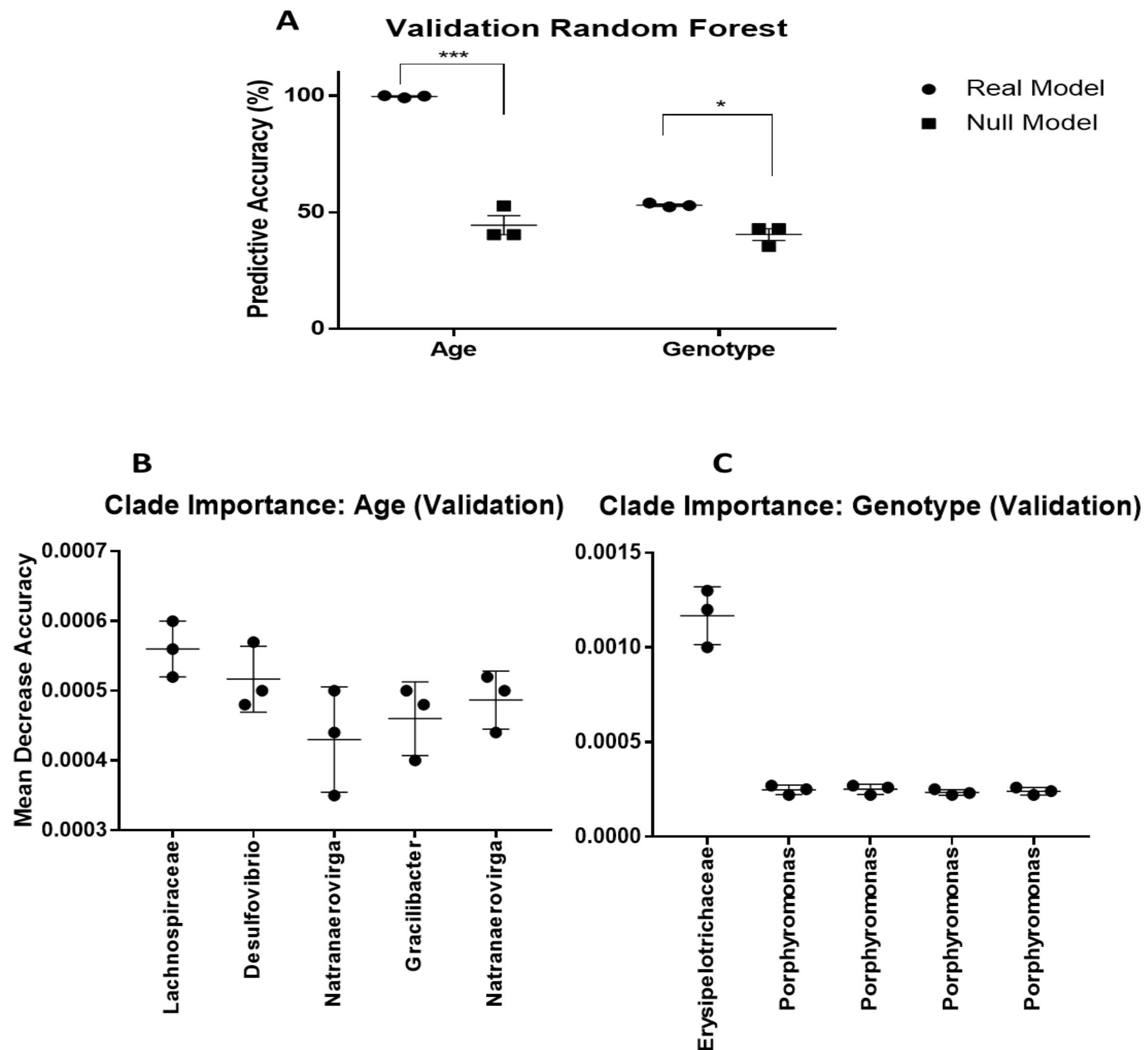
2.7 Supplementary Information



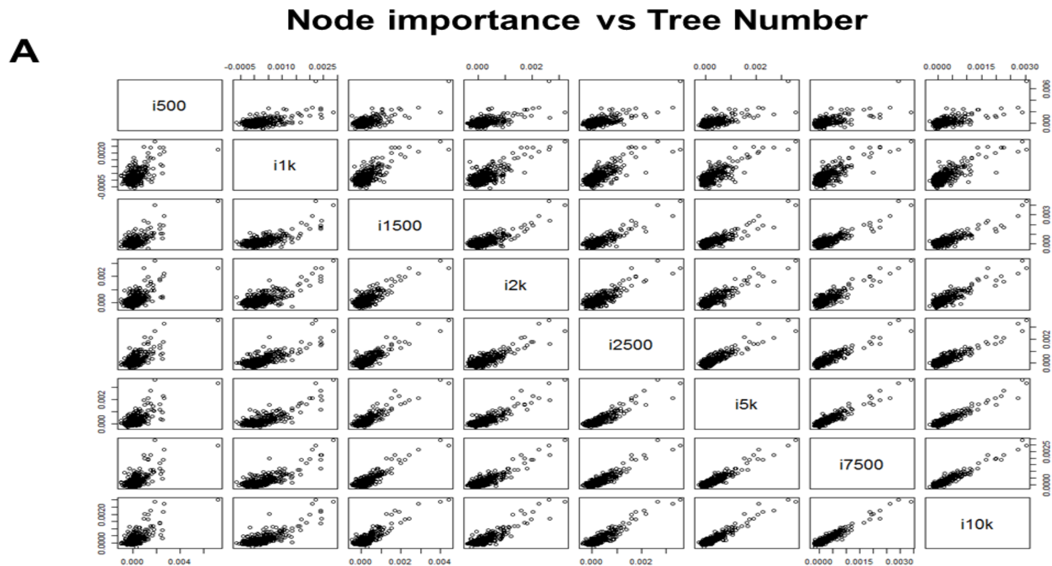
Supplementary Figure 1: Major phyla are represented on the phylogenetic tree. A phylogenetic tree of 16S rRNA sequences derived from the gut microbiota of 6 and 18 week old, FVB wildtype (WT) mice and *mdr1a*$^{-/-}$ mice. The distribution of major gut phyla are highlighted on the tree: Firmicutes (grey), Bacteroidetes (pink), Proteobacteria (olive), Actinobacteria (red, and Deferribacteres (gold).

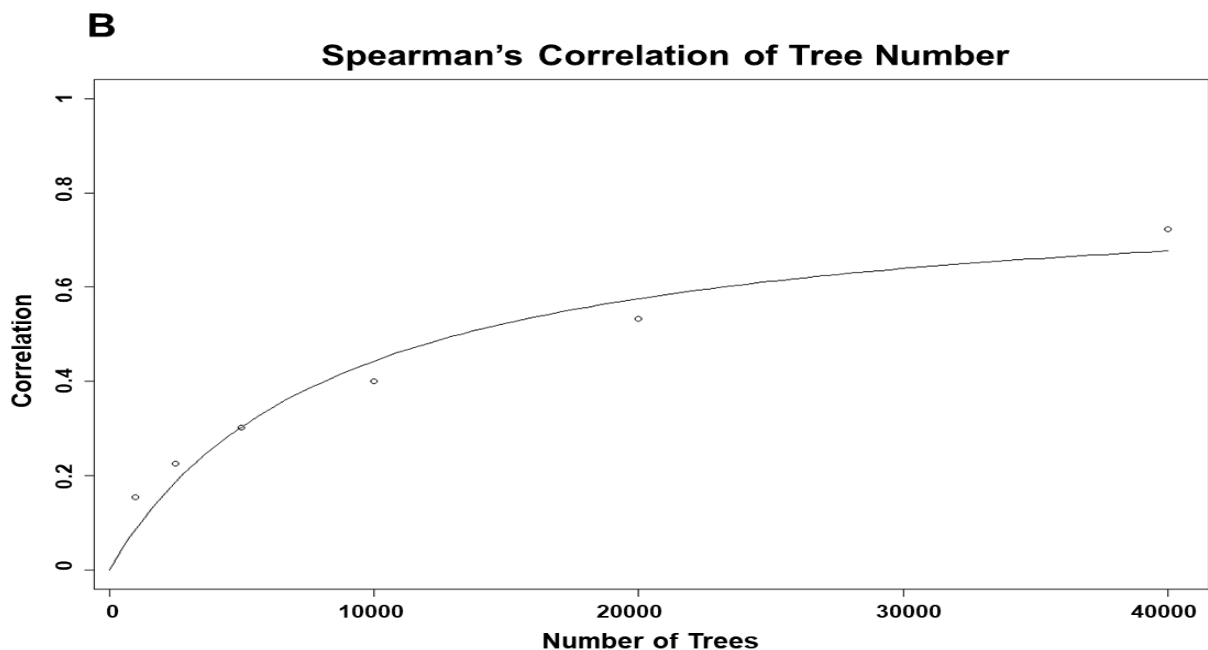


Supplementary Figure 2: Wide distribution of sequences on a phylogenetic tree when coloured by treatment groups. A phylogenetic tree of 16S rRNA sequences derived from the gut microbiota of FVB wildtype (WT) mice and *mdr1a*$^{-/-}$ mice was plotted and coloured by: Age (A), genotype (B) and cage (C). Colours indicate 6 weeks of age (black), 18 weeks of age (grey), WT mice (blue), *mdr1a*$^{-/-}$ mice (red) and different cages (red, green, goldenrod, purple, dark blue, steel blue, pink and dark green).



Supplementary Figure 3: Redistributing the relative abundance of clade Erysipelotrichaceae confers importance to genotype. 16S rRNA sequence data derived from the stools and mucus of 6 and 18 week old mice (FVB background, wildtype vs *mdr1a*$^{-/-}$ mice) were used to construct a phylogenetic tree. A data matrix was established comparing samples and the relative abundance of taxonomic clades within that sample. This matrix was used as an input for a random forest (RF) model to find associations between the gut microbiota and different experimental treatments. The most important clade for age was identified and its relative abundances were redistributed into WT samples only and the RF was repeated. The predictive accuracy of the RF model at taking a sample and discriminating between age and genotype are displayed (A). The five most important nodes when comparing age (B) and genotype via the RF are illustrated. Data shown as mean +/- standard error mean (SEM). Asterisks represent significance determined using Two Way ANOVA: p =< 0.00001 (****). n = 3.





Supplementary Figure 4: Assessing Robustness of the RF model. 16S rRNA sequence data derived from the stools and mucus of 6 and 18 week old mice (FVB background, wildtype vs *mdr1a*$^{-/-}$ mice) were used to construct a phylogenetic tree. A data matrix was established comparing samples and the relative abundance of taxonomic clades within that sample. This matrix was used as an input for a random forest (RF) model to find associations between the gut microbiota and different experimental treatments. The RF was run with an increasing number of trees in R and the 'MeanDecreaseAccuracy (MDA) value was plotted for all clades (A). Spearman's correlation was used to investigate the number of trees needed for an accurate estimation of importance and plotted with a Michaelis-Menten Model (Vmax = 0.82) (B). n = 1.

Chapter Three

Mucus-residing metabolites and colitis

Author Contributions

Gurdeep Singh: Designed and performed experiments, analysed data and wrote the manuscript.

Andrew Brass: Designed and supervised the project, critically reviewed the manuscript.

Sheena Cruickshank: Designed and supervised the project, critically reviewed and edited the manuscript

Roy Goodacre: Supervised LC-MS aspect of the project

Katherine Hollywood: Performed the LC-MS, critically reviewed and edited the manuscript.

Christopher Knight: Designed and supervised the project, critically reviewed the manuscript

Drupad Trivedi: Processed the LC-MS data.

Yun Xu: Analysed the LC-MS data, critically reviewed the manuscript.

3. Mucus metabolite and transcriptional changes precede the onset of colitis-induced inflammation

Gurdeep Singh[1], Andrew Brass[2], Sheena M. Cruickshank[1], Roy Goodacre[3], Katherine Hollywood[3], Christopher G. Knight[4], Drupad Trivedi[3] and Yun Xu[3]

[1] Faculty of Biology, Medicine and Health, Lydia Becker Institute of Immunology and Inflammation, Manchester Academic Health Science Centre, A.V. Hill Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom.

[2] Faculty of Biology, Medicine and Health, Division of Informatics, Imaging and Data Sciences, Stopford Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom.

[3] Faculty of Science and Engineering, Manchester Synthetic Biology Research Centre, Manchester Institute of Biotechnology, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom

[4] Faculty of Science and Engineering, School of Earth and Environmental Sciences, Michael Smith Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom

Corresponding authors:

Professor Sheena Cruickshank

Faculty of Biology, Medicine and Health

The University of Manchester

AV Hill Building

Oxford Road

Manchester

M13 9PT

sheena.cruickshank@manchester.ac.uk

Phone +44 (0) 161 275 1582

Short title: **Mucus-residing metabolites and colitis**

3.1 Abstract

The gut is home to the largest concentration of our resident microbiota. Thus, there are many mechanisms in place to regulate contact between our resident bacteria and our host cells, notably the gut barrier. Inflammatory bowel diseases are a group of autoimmune disorders of the gut that are associated with a loss of gut barrier function and altered bacterial communities. Alterations in bacterial communities will in turn impact on both bacteria and host-derived metabolites that are known to play a crucial role in the maintenance of host health. The gut microbiota exists in two compartments in the large intestine – the mucus and the gut lumen. We have previously shown in a model of colitis, that perturbations in the mucus-resident microbiota are evident before stool changes can be detected. However, the mucus has largely been ignored in metabolomic studies. We therefore hypothesise that a perturbation to the mucus-resident bacteria will lead to an altered metabolomic profile within the mucus, with a consequential impact on host barrier function and health.

We did a littermate-controlled, comparative analysis between the gut mucus-residing metabolites in wildtype (WT) mice and mice that develop spontaneous colitis over time (*mdr1a*$^{-/-}$ mice) and developed methodologies to profile the mucus-resident metabolites. Despite no evidence of inflammation based on histological analysis, there was genotypic variability in individual mucus-resident metabolites between WT and *mdr1a*$^{-/-}$ mice. This was concordant with significant alterations to host transcription within the gut, notably genes associated with angiogenesis and intestinal permeability. Alterations to factors associated with intestinal permeability may predispose the *mdr1a*$^{-/-}$ mice to the development of colitis.

Our data support the importance of profiling the mucus resident bacteria and metabolome in the study of IBD.

3.2 Introduction

The microbiota is known to be important in various aspects of host health and a trialogue of contact between immune cells, gut epithelial cells and the microbiota contribute to the maintenance of gut homeostasis (Shulzhenko et al., 2011). Gut homeostasis is in part maintained by the 'barrier function', where the gut epithelial cells can act as a physical barrier to maintain intestinal permeability, and produce antimicrobial factors to regulate our microbiota. A loss of barrier function can lead to increased bacterial invasion, driving inflammation and the onset of inflammatory bowel disease (IBD) (Collett et al., 2008). However, the microbiota can also modulate gut homeostasis via the production of metabolites (Yamada et al., 2016). Metabolites are small molecules primarily derived from the diet and produced by the gut microbiota, from substrates that escape absorption in the upper gastrointestinal tract (Nyangale et al., 2012). These small chemical molecules mediate functional effects on the host including immune development and the regulation of inflammation (Trompette et al., 2014). Notably, a group of metabolites collectively referred to as short-chained fatty acids (SCFAs) have been shown to be vital in the maintenance of host health, such as regulating adiposity and glucose tolerance (Brown et al., 2003; Samuel et al., 2008; De Vadder et al., 2014).

Alterations in both the gut microbiota and metabolites are associated with negative health consequences, such as the development of inflammatory bowel disease (IBD) (Wu, 2014). Although the microbiota has been strongly implicated in IBD, previous work has generally focused on stool samples as a representation of the bacterial communities living within the gut. However, it is equally important to consider the mucus microbiota, as our previous work indicates that there are changes in the mucus microbiota of colitis-prone mice long before the onset of detectable inflammation (Glymenaki et al., 2017b). It is therefore likely that the altered microbiota will impact on the metabolomic profile within the mucus. Given the closer proximity of the mucus to our host cells, these changes in both the mucus microbiota and their metabolites could contribute to increased barrier permeability and development of inflammation seen in IBD. Furthermore, alterations to metabolites could result in host-transcriptional changes within the intestine that may impact on barrier integrity and further exacerbate the development of IBD.

We developed methodologies to profile the mucus-residing metabolites and investigated them in mice that spontaneously develop colitis over time ($mdr1a^{-/-}$), along with littermate controls. We specifically looked early before the onset of colitis (6 weeks of age) to investigate the

hypothesis that early changes in microbiota impact on metabolite production. In parallel, we investigated the host transcription in these mice. We demonstrate that these mice have changes to specific mucus-resident metabolites compared to WT mice, before the onset of inflammation. We also see significant transcriptional changes within the gut, including genes involved in intestinal permeability and angiogenesis. Our data suggest an early disruption to gut barrier function and that metabolites may play a role in the development of colitis in *mdr1a*$^{-/-}$ mice, supporting the importance for mucus profiling in understanding disease aetiology.

3.3 Materials and Methods

3.3.1 Animal Maintenance

Mdr1a$^{-/-}$ mice (FVB.129P2-Abcb1atm1Bor N7) (Schinkel et al., 1994) were bred with control FVB mice purchased from Taconic Farms (Albany, NY), to produce the F2 generation. All mice were maintained under specific, pathogen-free conditions and co-housed. Thus, 6 week old WT and *mdr1a*$^{-/-}$ male mice from the same litters were used for all subsequent experiments. Food (Beekay Rat and Mouse Diet No1 pellets; B&K Universal, UK) and water were available *ad libitum*. Ambient temperature was maintained at 21 (+/- 2°C) and the relative humidity was 55 (+/- 10%) with a 12h light/dark cycle. All experiments were performed according to the regulations issued by the Home Office under amended ASPA, 2012.

3.3.2 Collection of tissue

Sample collection and processing was performed as described by Glymenaki et al. (2017b). In brief, samples were harvested from mice at 6 weeks of age. Faecal samples were collected into sterile Eppendorf tubes and snap-frozen on dry ice. Colon snips were fixed in Carnoy's fixative (60% ethanol absolute, 30% chloroform and 10% glacial acetic acid) to preserve the mucus and embedded in paraffin for histological analysis. The remaining colon was then bisected to expose the inner surface and any remaining faecal matter was removed and gently washed away with PBS (Sigma, Poole, UK). The inner surface of the colon was scraped using a cell scraper to remove mucus from the mucus lining which was snap-frozen on dry ice.

3.3.3 Histology and Staining

Carnoy's fixed samples were incubated in two changes of dry methanol (Sigma, UK) for 30 minutes each, followed by absolute ethanol (ThermoFisher Scientific, Crawley, UK) for two incubations at 30 minutes each. Tissue cassettes were processed in a Micro-spin Tissue Processor STP120 (ThermoFisher Scientific) and immersed in paraffin using a Leica Biosystems embedding station (Leica Biosystems, Milton Keynes, UK), with the luminal surface of the colon exposed for tissue sectioning. 5µm tissue sections were cut using a Leica Biosystems microtome and adhered to uncoated microscope slides (ThermoFisher Scientific). Slides were dried for 48 hours at 50ºC before use. Haematoxylin and Eosin (H+E) and goblet cell staining were performed and analysed as described previously (Glymenaki et al., 2017b). In brief, sections were scored out of 9 according to parameters in **Table 1**.

Table 1: Histological scoring system for mucosal inflammation

Active inflammation	Lamina propria cellularity	Surface ulceration
0: Normal	0: Normal	0: No ulceration with intact surface epithelium
1: Mild crypt distortion and loss and/or mild cryptitis (<5% of crypts infiltrated by neutrophils) with mild crypt abscess formation	1: Mild but unequivocal increase in mixed inflammatory cells	1: Probable erosion with focally stripped epithelium
2: Moderate crypt distortion and/or moderate cryptitis (<50% of crypts infiltrated by neutrophils) with mild crypt abscess formation	2: Moderate increase mixed inflammatory cells	2. Unequivocal erosion
3: Severe crypt distortion and loss with widespread and diffuse cryptitis (>50% of crypts involved) and diffuse goblet cell depletion	3. Severe and diffuse increase in inflammatory cells	3. Surface ulceration and granulation tissue formation

3.3.4 Albumin ELISA

Faecal samples were mixed with PBS (Sigma) at 100µL per 10mg of faeces and homogenised. Samples were centrifuged at 200xg for 5 minutes to pellet debris, the supernatant was collected and centrifuged at 8000xg for 5 minutes to pellet the bacteria. The supernatant was collected and stored at -80ºC. Albumin levels in serum diluted 1/100 were assessed using a Mouse Albumin ELISA Kit (Universal Biologicals Ltd, Cambridge, UK).

3.3.5 RNA-seq

Proximal colon snips were taken from 6 week old WT and *mdr1a*$^{-/-}$ mice, stored in RNA-Later (ThermoFisher Scientific) and stored at -80ºC until use. RNA was then extracted using the RNeasy Mini Kit (Qiagen, Manchester, UK). Samples were diluted to 1µg/µL concentration and rRNA depleted using a Ribo-Zero Gold rRNA Removal Kit (Illumina, Essex, UK). Samples were diluted to 1ng/µL concentration for RNA-seq. RNA-seq was performed using the HiSeq4000 (Illumina, California, USA). Data was analysed using DESEQ2 in R. Specifically, unmapped paired-end sequences from an Illumina HiSeq4000 sequencer were tested by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequence adapters were removed and reads were quality trimmed using Trimmomatic_0.36 (Bolger et al., 2014). The reads were mapped against the reference mouse genome (mm10/GRCm38) and counts per gene were calculated using annotation from GENCODE M14 (http://www.gencodegenes.org/) using STAR_2.4.2 (Dobin et al., 2013). Normalisation, Principal Components Analysis, and differential expression was calculated with DESeq2_1.16.1 (Love et al., 2014). A heatmap was produced in R using the Gplots package (Gregory et al., 2016).

3.3.6 Metabolite profiling using LC-MS

Mucus homogenisation and subsequent metabolite extraction was undertaken using a Tissuelyser II (Qiagen). The homogenisation solvent (1:1 chloroform:methanol, 800 µL pre-chilled to -20 °C) was added to each sample, a steel bead added and then homogenised for 20 min at 25 Hz. Once homogenised, 400 µL of water was added and the sample vortex mixed for 15 s. To initiate phase separation the samples were then centrifuged (8000 xg for 10 min) before the organic fraction was collected for LC-MS analysis. The organic fraction was subsequently vacuum concentrated (Eppendorf Vacuum concentrator, RT, 8 h) and stored at -20 °C until analysis. Upon analysis, the metabolite extracts were resuspended in methanol (normalised to dry-weight) and analysed by LC-MS. A portion of each sample was pooled to

give a representative pooled biological 'quality control' (QC) which can be used to assess instrumental variation during the analytical run (Broadhurst et al., 2018).

All LC-MS analysis was conducted on a QExactive Plus equipped with an Ultimate 3000 UHPLC (ThermoFisher Scientific). The UHPLC was equipped with a Hypersil Gold reverse phase column (C18 -2.1 mm x 150 mm; 1.9 µm particle size). The solvents employed were (A) water + 0.1% formic acid and (B) methanol + 0.1 % formic acid. The flow gradient was programmed to equilibrate at 95% A for 2 min followed by a linear gradient to 95% B over 8 min and held at 95% B for 2 min before returning to 95% A for 2 min. The column was maintained at 40 °C and the samples chilled in the autosampler at 4 °C. A sample volume of 5 µL was injected onto the column with a constant flow rate of 400 µL/min. Blank injections were analysed at the start and end of the analytical batch to assess the background and carryover. In addition pooled QC samples (as above) were analysed at every 6th injection to assess for analytical drift over time. Data acquisition was conducted in full MS mode in the scan range of 70-1050 m/z with a resolution of 70,000, an AGC target of $3e^6$ and a maximum integration time of 200 ms. The samples were analysed in positive and negative mode in separate acquisitions.

The raw data was converted to mzXML file format (Proteowizard) and XCMS was used to deconvolve the data. The data underwent QC-based filtering, where any feature that was missing in more than 50% of QCs was removed. All features within a given sample were normalised to account for variation that may arise between sample injections. PCA's were plotted from the processed data using R (R Core Team, 2018) and the ggfortify package (Tang et al., 2016). A Random Forest model was run using the randomForest package in R (Liaw and Wiener, 2002), with the default mtry value and 100000 trees. The relative peak area distribution for the top 4 metabolites identified by the Random Forest, based on their 'MeanDecreaseAccuracy' value, were plotted. The m/z ratio for each metabolite feature was used to assign putative identity based on comparisons to databases such as the Kyoto Encyclopaedia for Genes and Genomes (KEGG) and the Human Metabolome Database (HMDB).

3.3.7 Statistical Analysis

All statistical analysis was performed using either GraphPad Prism 7 (GraphPad Software, La Jolla, USA) or R. Student's T Test was used to compare inflammation score, faecal albumin levels, crypt length, muscle wall thickness and goblet cell number. Man Whitney U Test was

used to compare the relative peak area distribution for metabolites between genotype, with false discovery rate correction for multiple testing. Permutational multivariate analysis of variance was used to calculate differences in overall metabolites and transcription using the ‘adonis’ function in R, in the VEGAN package (Oksanen et al., 2016).

3.4 Results

3.4.1 No evidence of gut inflammation in *mdr1a*$^{-/-}$ mice

Young (6 week old) *mdr1a*$^{-/-}$ and WT littermate controls were assessed for indications of inflammation by histology, assessing gut morphology and numbers of mucus-secreting goblet cells (Figure 1). In concordance with our previous work (Glymenaki et al., 2017b), we found no differences in inflammation scores between WT and *mdr1a*$^{-/-}$ mice at this timepoint, based on colonic crypt length, muscle wall thickness, goblet cell number and cellular infiltration. We also saw no differences in intestinal permeability based on levels of faecal albumin (Figure 1F), although one out of 6 *mdr1a*$^{-/-}$ mice had high levels of faecal albumin. However, this mouse had normal gut architecture and no overall indications of inflammation.

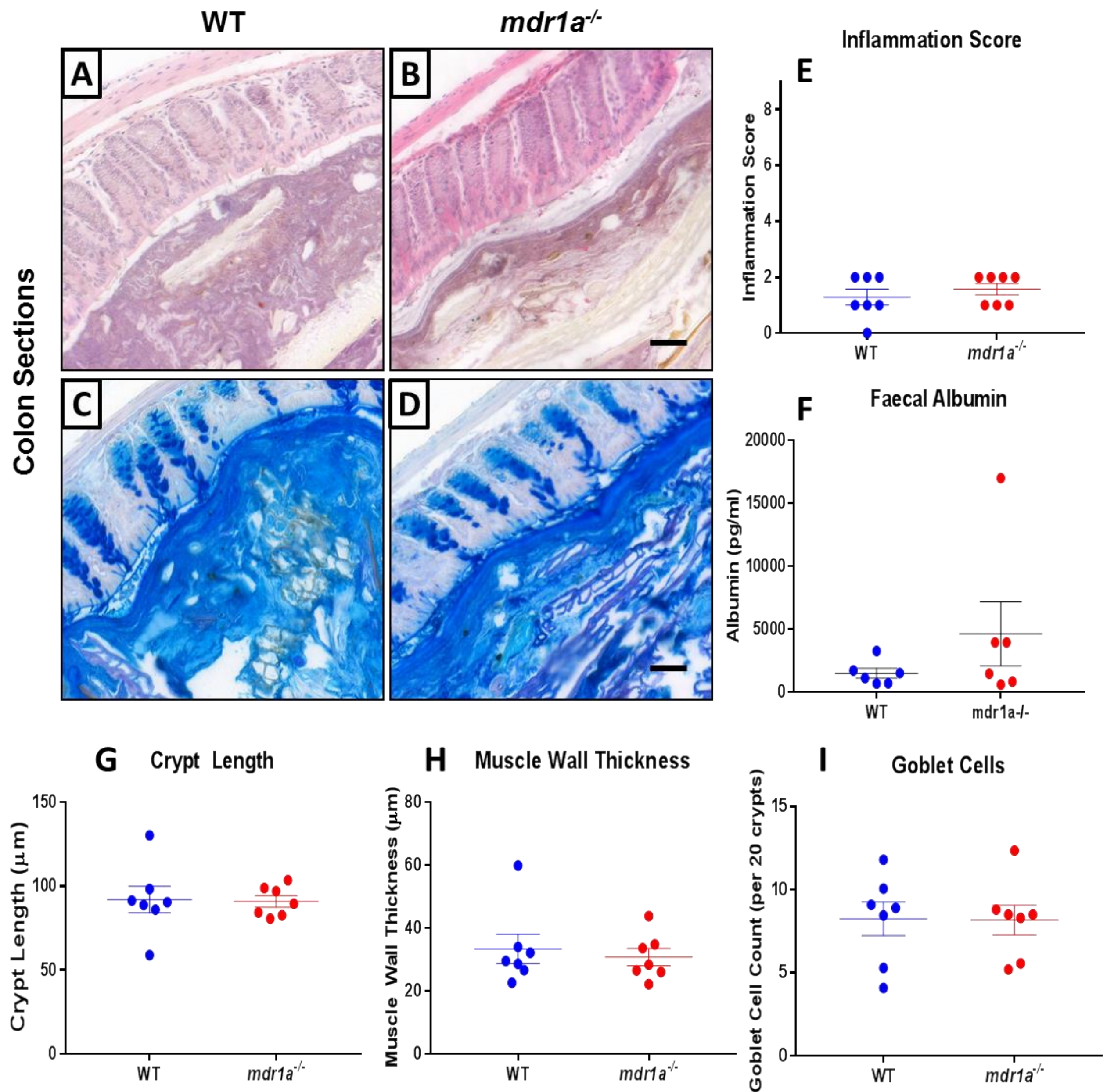


Figure 1: Gut barrier integrity in *mdr1a*$^{-/-}$ mice. Colonic tissue sections from 6 week old, male, FVB background wildtype (WT) and *mdr1a*$^{-/-}$ littermate mice were investigated for evidence of inflammation. Tissue was stained with haematoxylin and eosin to observe the gut morphology and representative images were taken for WT mice (A) and *mdr1a*$^{-/-}$ mice (B). Goblet cells were stained using periodic acid, Alcian blue and Schiff's reagent and representative images for WT mice (C) and *mdr1a*$^{-/-}$ mice (D) are displayed. Sections were scored based on previously described parameters for signs of colitis (maximum possible score is 9) (E). Faecal albumin levels were measured to assess gut permeability (F). Colonic crypt length (G), and muscle wall thickness (H) and goblet cells were quantified (I). Data is shown as mean +/- standard error of the mean (SEM). n = 6-7. Scale bar = 50µm.

3.4.2 Alterations to angiogenesis and permeability-associated transcription in *mdr1a*$^{-/-}$ mice

In order to investigate barrier function, we performed RNA-seq to examine changes in gene expression in the proximal colon. We revealed a significant difference in overall host transcription between WT and *mdr1a*$^{-/-}$ mice (PEMANOVA, $P < 0.05$) (Figure 2A). In terms of clustering, *mdr1a*$^{-/-}$ samples were clustered closer together indicating reduced variability in expression of the *mdr1a*$^{-/-}$ transcriptional profile. Conversely WT samples were spread far apart, suggesting that transcription was variable.

Upon investigating where these transcriptional differences lay, we found that mice clustered by treatment group, based on differences in several specific pathways including those associated with angiogenesis and intestinal permeability (Figure 2B). Angiogensis-associated genes include those associated with the vascular endothelial growth factor (VEGF)-signalling pathway, such as VEGF-A, B, D and placental growth factor (PGF). Specifically, VEGF-B had higher expression in the *mdr1a*$^{-/-}$ mice and VEGF-D had higher expression in the WT mice. We also saw differences in angiopoetin 1 and 2 (Angpt1 and 2), where Angpt1 had higher expression in WT mice and Angpt2 had higher expression in the *mdr1a*$^{-/-}$ mice. With respect to intestinal permeability, tight junction genes including occludin (Ocln) and claudin-2 (cldn2), and aquaporins 7 and 9 (AQP7 and 9) were also affected. Claudin-2 generally had higher expression in the WT mice, whereas occludin had higher expression in the *mdr1a*$^{-/-}$ mice. Aquaporin expression was more variable between treatment groups. Two additional genes were identified, the gene encoding neuronal acetylcholine receptor subunit alpha-7 (Chrna7) and Actin alpha 3 (Actn3). For both these genes, WT mice had consistently reduced levels of expression whereas *mdr1a*$^{-/-}$ expression of these genes was more variable. Notably, actin is involved in the formation of tight junctions.

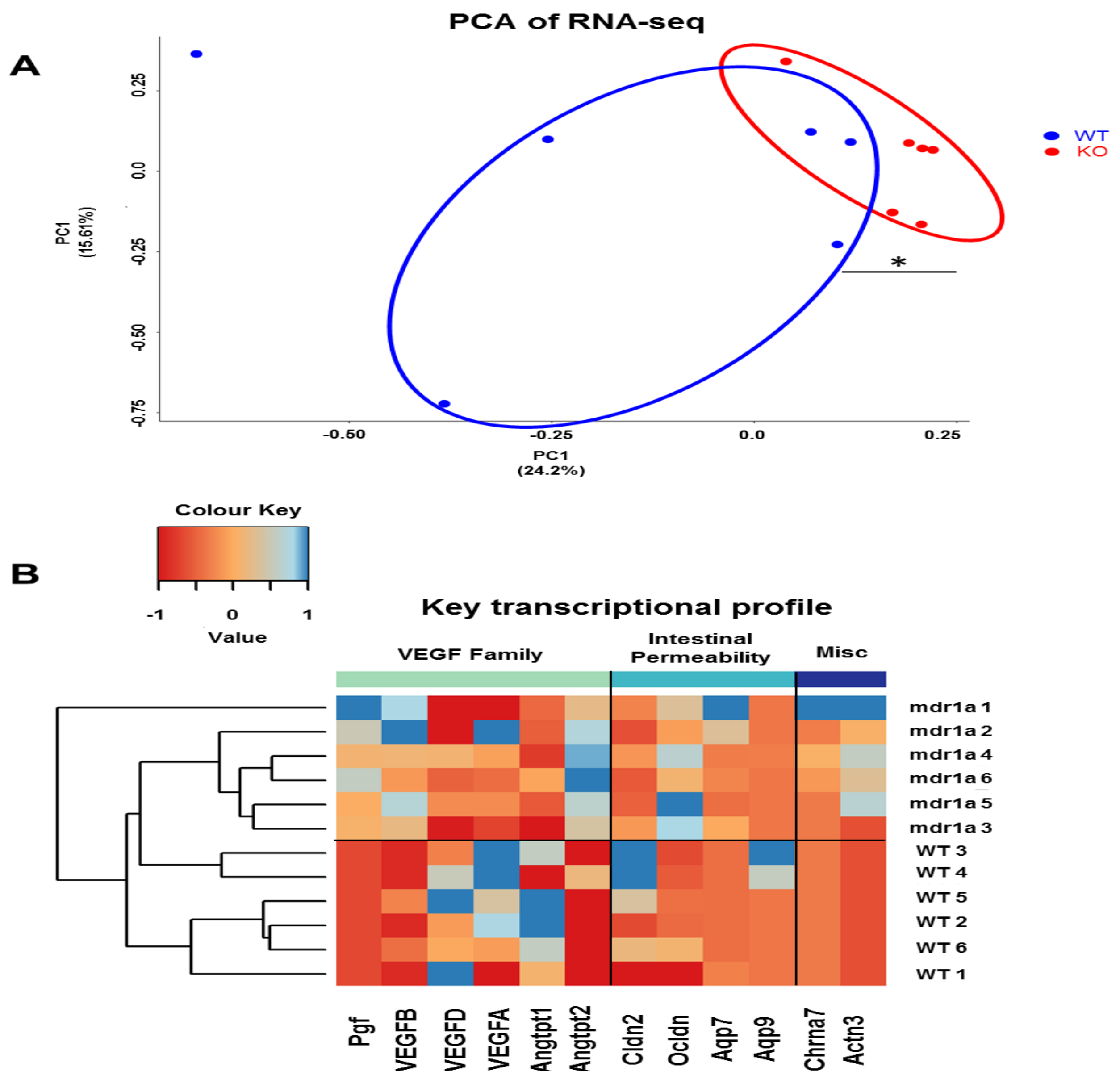


Figure 2: Transcriptional differences between WT and *mdr1a*$^{-/-}$ mice. Proximal colon tissue was taken from 6 week old male, FVB background wildtype (WT) and *mdr1a*$^{-/-}$ littermate mice and the RNA was extracted. Transcription was assessed via RNA-seq. Principle component analysis was used to assess overall transcriptional differences (A). A heatmap was used to illustrate differences in key signalling pathways (B). Pgf (placental growth factor), VEGF (Vascular Endothelial Growth Factor), Angpt1 and 2 (Angiopoietin 1 and 2), Cldn2 (Claudin 2), Ocldn (Occludin), Aqp7 and 9 (Aquaporin 7 and 9), Chrna7 (Neuronal acetylcholine receptor subunit alpha-7) and Actn3 (Actin alpha 3). Asterisks represent significance determined by a permutational multivariate analysis of variance (*) $P < 0.05$. n = 6.

3.4.3 Difference in mucus-resident metabolites at 6 weeks of age in *mdr1a*$^{-/-}$ mice

Given the differences identified in terms of intestinal permeability-associated transcription, these changes could be mediated by metabolites. We previously demonstrated that the mucus microbiota in *mdr1a*$^{-/-}$ mice was significantly different compared to that of healthy mice (Glymenaki et al., 2017a). It was therefore possible that the mucus metabolite profile in these mice could also be altered. Metabolites were determined based on their relative peak area distribution obtained during LC-MS analysis. Therefore, every reference to a 'metabolite' is a reference to a 'metabolite feature' based on the peaks obtained. Our previous analysis based on 16S rRNA sequence data and the predictive tool PICRUSt revealed that the mucus-residing metabolites were relatively stable between WT and *mdr1a*$^{-/-}$ mice (Glymenaki et al., 2017a). In this study, PCA analysis of mucus-residing metabolites, in both positive (Figure 3A) and negative (Figure 4A) ionisation mode did not clearly seperate all WT and *mdr1a*$^{-/-}$ mice. However, samples did segregate into two distinct clusters. A random forest was used to determine whether the metabolite profile could discriminate between different treatment groups.

The predictive accuracy for the random forest was lower than 50%, suggesting that overall, the metabolite features cannot be used to discriminate between different treatment groups (Supplementary Figure 1A). Nevertheless, there were metabolites that had great bearing on the accuracy of the RF and their importance values have been reported (Supplementary Figure 1B and C). For positive ionisation mode, metabolites 1451, 239, 967 and 1450 were the four most important metabolites highlighted by the random forest and their relative peak area distribution have been plotted (Figure 3B, C D and E). For negative ionisation mode, four different metabolites were highlighted: 175, 619, 81 and 1 (Figure 4B, C, D and E).

The m/z ratios and retention times of each metabolite peak was used to assign putative identities for both positive ionisation mode (Supplementary Table 1) and negative ionisation mode (Supplementary Table 2). It should be noted that due to differences in retention times between chromatography runs and different platforms, identities can only be truly confirmed through validation of known standards on the in-house machinery. Therefore, the feature identities proposed are speculative.

In the negative ionisation run, metabolites identified included 3,5-Dichlorophenylcarboximide and 6-Bromo-7H-Purine which have not previously been identified in the context of the gut. Additionally, the RF model identified methylglyoxal as an

important metabolite, a known ‘toxic’ metabolite that can contribute to irritable bowel syndrome development (Zhang et al., 2014). It was also shown to be a potent inducer of calcium channels on bacteria, controlling growth of bacteria such as *E. coli* (Campbell et al., 2007). Conversely, the positive ionisation mode had numerous metabolites identified, but none have previously been associated with the gut.

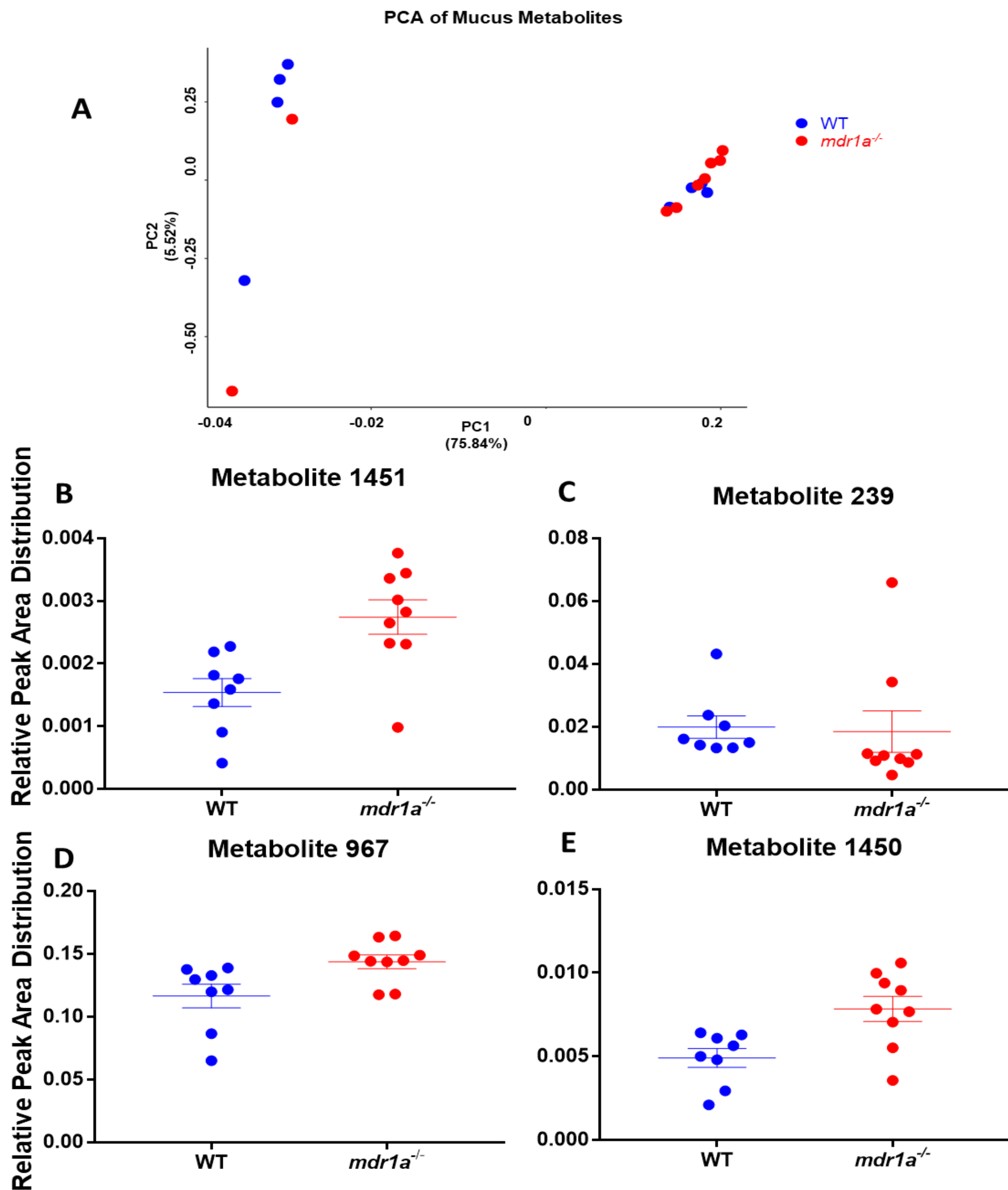


Figure 3: LC-MS (Positive Ionisation Mode) reveals variation in mucus-residing metabolites. Colonic mucus was extracted from 6 week old male, FVB background wildtype (WT) and *mdr1a*$^{-/-}$ littermate mice. Metabolites were assayed via LC-MS and a PCA of the overall metabolites is illustrated (A). A random forest model was used to identify metabolites that were important to discriminate between the different treatment groups and the top five have been presented with their relative peak area distribution (B, C, D and E). n = 8-9.

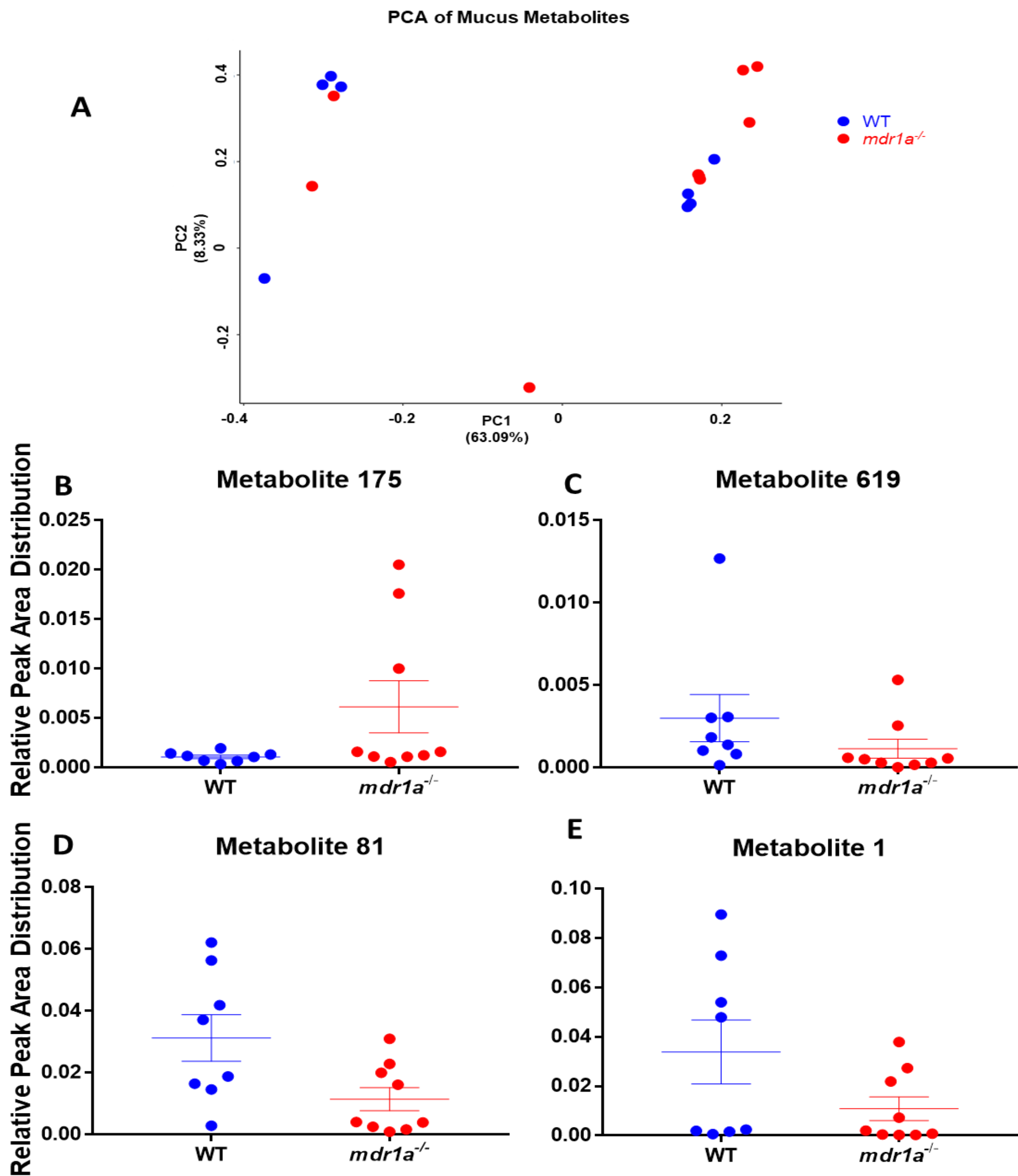


Figure 4: LC-MS (Negative Ionisation Mode) reveals variation in mucus-residing metabolites. Colonic mucus was extracted from 6 week old male, FVB background wildtype (WT) and *mdr1a*$^{-/-}$ littermate mice. Metabolites were assayed via LC-MS and a PCA of the overall metabolites is illustrated (A). A random forest model was used to identify metabolites that were important to discriminate between the different treatment groups and the top five have been presented with their relative peak area distribution (B, C, D and E). n = 8-9.

3.5 Discussion

Mdr1a$^{-/-}$ mice are known to develop spontaneous colitis over time and a previous study identified that there were changes in the caecal microbiota that preceded the onset of colitis (Nones et al., 2009). We subsequently showed that there were changes in the mucus microbiota but not the stool that preceded colitis (Glymenaki et al., 2017b). It has long been established that metabolites are important to maintain healthy host function, with SCFAs implicated in early forms of IBD therapy (Harig et al., 1989). However, much work has focused on serum and urine, which although give an overall metabolomic 'fingerprint', would not highlight local changes within a tissue environment (Kell et al., 2005). Given the changes that we see in the mucus bacteria, we investigated the mucus for changes in metabolites.

We saw no overall changes in our mucus metabolite features between genotype, however the samples appeared to cluster into two distinct groups. Indeed, the PCA highlighted that the biggest variation in the data was due to samples clustering by cage (data not shown), suggesting that cage effect has a large impact on metabolites. Our random forest model was unable to discriminate between genotype using the metabolite profile. However, random forests often require larger datasets in order to train the model accurately. Additionally, our mice were co-housed, littermate controls and it is known that cage effect can have powerful effects on the gut microbiota. Co-housing would prevent 'artificial' genotype-associated microbiota and metabolite changes and is more reflective of a real environment. Cage effect could be one reason for why we do not see genotypic differences. It is also possible that metabolite changes are more likely to happen as a consequence of severe inflammation and thus at a later timepoint in our mice. Indeed, it is known that butyrate metabolism in ulcerative colitis patients is diminished but only in patients with severe, but not mild colitis (Duffy et al., 1998). In our mouse model, the 6 week timepoint is before the onset of any inflammation, thus there may not be large overall changes to the metabolomic profile until the onset of colitis.

However, we did see variability in individual metabolite features between genotypes, despite seeing no significant differences. The potency of individual metabolites should not be underestimated. For example, although it was not detected in this study, butyrate has the ability to regulate inflammation through T regulatory cell induction (Furusawa et al., 2013) and has been shown to be more effective than propionate and acetate in mediating anti-inflammatory effects (Sakata, 1987). Depletion of butyrate-producing Clostridia led to increased colonisation of Salmonella in the gut of mice (Rivera-Chavez et al., 2016).

Therefore, even though only a small number of metabolites show variation between genotype in our study, this could still be enough to mediate large physiological effects. In positive ionisation mode, none of the metabolites provisionally identified were associated with the gut and thus there could still be potentially important metabolites within this fraction of samples. It should be noted that retention times vary between chromatography runs and that the identity of metabolites cannot always be extrapolated. Hence, the IDs discussed are putative.

In concert with differences in the individual metabolites, we also saw differences in gene expression in the gut. Gene expression in young, non-inflamed *mdr1a*$^{-/-}$ mice has been explored previously, but not compared against littermate controls (Collett et al., 2008). That study revealed that *mdr1a*$^{-/-}$ mice had alterations to genes associated with antigen presenting, whereas we primarily identified genes associated with intestinal permeability- suggesting an impact on barrier function- and angiogenesis.

With regards to intestinal permeability, we saw differences in the expression of two tight-junction proteins claudin-2 and occludin. Of note, claudin-2 expression was generally higher in WT mice than *mdr1a*$^{-/-}$ mice, whereas the level of occludin expression was higher in the *mdr1a*$^{-/-}$ mice. A previous study in human patients found that claudin-2 was downregulated in Crohn's Disease and ulcerative colitis patients and this led to redistribution of claudin-5 and 8 (Zeissig et al., 2007), although these two claudins were not significantly different based on the RNA-seq. Occludin was also downregulated and redistributed in the human study, despite our *mdr1a*$^{-/-}$ mice showing increased expression of it. Therefore, even though occludin had increased expression, its localisation could be such that this has no physiological effect on barrier function. Metabolites have been previously shown to influence tight junction proteins (Shimada et al., 2013; Venkatesh et al., 2014), although we could not establish a link in this study. Perhaps there are transcriptional alterations happening at the 6 week timepoint that may not have any direct phenotypic consequence in the short term, but could potentially predispose the *mdr1a*$^{-/-}$ mice to colitis. Additionally, there were differences in expression of two members from a family of water channel proteins known as aquaporins (Agre et al., 1993). Specifically, Aqp7 and Aqp9 are described as aquaglyceroporins which are permeable to water and various small solutes such as urea (Ricanek et al., 2015). Reduced aquaporin expression has been associated with early-stage IBD (Ricanek et al., 2015).

In addition to genes associated with permeability, we also identified a cluster of genes involved in angiogenesis that were different between treatment groups. Angiogenesis has

previously been reported as a contributing factor to IBD (Salem and Wadie, 2017). Our RNA-seq identified multiple targets involved in angiogenesis including various VEGF family members and Angpts. With regards to VEGF, VEGF-A appears to be the best described and has been associated with IBD (Scaldaferri et al., 2009). However, our WT mice appear to have increased expression of VEGF-A compared to the *mdr1a*$^{-/-}$ mice. The study by Scaladaferri et al. (2009) investigated DSS-treated mice and clinical samples from IBD patients, thus, VEGF-A could be important in more aggressive or inflammatory forms of IBD.

It is also possible that other VEGF family members may be important in the development of colitis. Specifically VEGF-B had increased expression in *mdr1a*$^{-/-}$ mice. However, there is a current lack of studies that have investigated VEGF-B in the context of IBD. Notably, another VEGF family member, PGF, had significantly elevated expression in *mdr1a*$^{-/-}$ mice. In other studies, PGF has been linked to increased angiogenesis in human microvascular epithelial cells (Zhou et al., 2016). The interplay between Angpt1 and Angpt2 has also been identified as a factor that may control inflammation. Indeed, Angpt1 was lower in patients in remission of IBD and Angpt2 was shown to be the best predictor of histological activity (Algaba et al., 2013). As we saw almost no signs of inflammation, these findings emphasise the fact that although there are transcriptional changes at this time point, any resulting phenotypic effects may only be apparent later in life, if at all.

In conclusion, we confirm that *mdr1a*$^{-/-}$ mice exhibit almost no signs of inflammation at 6 weeks of age. However, we do identify variation in specific metabolites and significant transcriptional differences in *mdr1a*$^{-/-}$ mice that may impair gut barrier function. These changes could be as a consequence of the altered microbiota at this timepoint. However, we have yet to fully correlate changes in the metabolites and transcription with the onset of colitis in these mice.

3.6 References

Agre, P., Sasaki, S. & Chrispeels, M. J. (1993). Aquaporins: a family of water channel proteins. *Am J Physiol,* 265(3 Pt 2)**,** F461.

Algaba, A., Linares, P. M., Fernandez-Contreras, M. E., Ordonez, A., Trapaga, J., Guerra, I., Chaparro, M., de la Poza, G., Gisbert, J. P. & Bermejo, F. (2013). Relationship between levels of angiogenic and lymphangiogenic factors and the endoscopic, histological and clinical activity, and acute-phase reactants in patients with inflammatory bowel disease. *J Crohns Colitis,* 7(11)**,** e569-79.

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30(15)**,** 2114-20.

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R. & Dunn, W. B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics,* 14(6)**,** 72.

Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., Muir, A. I., Wigglesworth, M. J., Kinghorn, I., Fraser, N. J., Pike, N. B., Strum, J. C., Steplewski, K. M., Murdock, P. R., Holder, J. C., Marshall, F. H., Szekeres, P. G., Wilson, S., Ignar, D. M., Foord, S. M., Wise, A. & Dowell, S. J. (2003). The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J Biol Chem,* 278(13)**,** 11312-9.

Campbell, A. K., Naseem, R., Holland, I. B., Matthews, S. B. & Wann, K. T. (2007). Methylglyoxal and other carbohydrate metabolites induce lanthanum-sensitive Ca2+ transients and inhibit growth in E. coli. *Arch Biochem Biophys,* 468(1)**,** 107-13.

Collett, A., Higgs, N. B., Gironella, M., Zeef, L. A., Hayes, A., Salmo, E., Haboubi, N., Iovanna, J. L., Carlson, G. L. & Warhurst, G. (2008). Early molecular and functional changes in colonic epithelium that precede increased gut permeability during colitis development in mdr1a(-/-) mice. *Inflamm Bowel Dis,* 14(5)**,** 620-31.

De Vadder, F., Kovatcheva-Datchary, P., Goncalves, D., Vinera, J., Zitoun, C., Duchampt, A., Backhed, F. & Mithieux, G. (2014). Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell,* 156(1-2)**,** 84-96.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29(1)**,** 15-21.

Duffy, M. M., Regan, M. C., Ravichandran, P., O'Keane, C., Harrington, M. G., Fitzpatrick, J. M. & O'Connell, P. R. (1998). Mucosal metabolism in ulcerative colitis and Crohn's disease. *Dis Colon Rectum,* 41(11)**,** 1399-405.

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T. A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., Takahashi, M., Fukuda, N. N., Murakami, S., Miyauchi, E., Hino, S., Atarashi, K., Onawa, S., Fujimura, Y., Lockett, T., Clarke, J. M., Topping, D. L., Tomita, M., Hori, S., Ohara, O., Morita, T., Koseki, H., Kikuchi, J., Honda, K., Hase, K. & Ohno, H. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature,* 504**,** 446-50.

Glymenaki, M., Barnes, A., Hagan, S. O., Warhurst, G., McBain, A. J., Wilson, I. D., Kell, D. B., Else, K. J. & Cruickshank, S. M. (2017a). Stability in metabolic phenotypes and inferred metagenome profiles before the onset of colitis-induced inflammation. *Sci Rep,* 7(1)**,** 8836.

Glymenaki, M., Singh, G., Brass, A., Warhurst, G., McBain, A. J., Else, K. J. & Cruickshank, S. M. (2017b). Compositional Changes in the Gut Mucus Microbiota Precede the Onset of Colitis-Induced Inflammation. *Inflamm Bowel Dis,* 23(6)**,** 912-922.

Gregory, R., Bolker, B., Warnes, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M. & Venables, B. (2016). *gplots: Various R Programming Tools for Plotting Data* [Online]. Available: https://CRAN.R-project.org/package=gplots [Accessed 14th August 2018].

Harig, J. M., Soergel, K. H., Komorowski, R. A. & Wood, C. M. (1989). Treatment of diversion colitis with short-chain-fatty acid irrigation. *N Engl J Med,* 320(1)**,** 23-8.

Kell, D. B., Brown, M., Davey, H. M., Dunn, W. B., Spasic, I. & Oliver, S. G. (2005). Metabolic footprinting and systems biology: the medium is the message. *Nat Rev Microbiol,* 3(7)**,** 557-65.

Liaw, A. & Wiener, M. (2002). Classification and Regression by randomForest. *R News,* 2(3)**,** 18-22.

Love, M. I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15(12)**,** 550.

Nones, K., Knoch, B., Dommels, Y. E., Paturi, G., Butts, C., McNabb, W. C. & Roy, N. C. (2009). Multidrug resistance gene deficient (mdr1a-/-) mice have an altered caecal microbiota that precedes the onset of intestinal inflammation. *J Appl Microbiol,* 107**,** 557-66.

Nyangale, E. P., Mottram, D. S. & Gibson, G. R. (2012). Gut microbial activity, implications for health and disease: the potential role of metabolite analysis. *J Proteome Res,* 11(12)**,** 5573-85.

Oksanen, J., Guillaume-Blanchet, F., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H. & Wagner, H. (2016). *vegan: Community Ecology Package* [Online]. Available: https://CRAN.R-project.org/package=vegan [Accessed 27th October 2017].

R Core Team. (2016). *R: A Language and Environment for Statistical Computing* [Online].Available: https://www.r-project.org/ [Accessed 16th January 2017].

Ricanek, P., Lunde, L. K., Frye, S. A., Stoen, M., Nygard, S., Morth, J. P., Rydning, A., Vatn, M. H., Amiry-Moghaddam, M. & Tonjum, T. (2015). Reduced expression of aquaporins in human intestinal mucosa in early stage inflammatory bowel disease. *Clin Exp Gastroenterol,* 8**,** 49-67.

Rivera-Chavez, F., Zhang, L. F., Faber, F., Lopez, C. A., Byndloss, M. X., Olsan, E. E., Xu, G., Velazquez, E. M., Lebrilla, C. B., Winter, S. E. & Baumler, A. J. (2016). Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. *Cell Host Microbe,* 19(4)**,** 443-54.

Sakata, T. (1987). Stimulatory effect of short-chain fatty acids on epithelial cell proliferation in the rat intestine: a possible explanation for trophic effects of fermentable fibre, gut microbes and luminal trophic factors. *Br J Nutr,* 58(1)**,** 95-103.

Salem, H. A. & Wadie, W. (2017). Effect of Niacin on Inflammation and Angiogenesis in a Murine Model of Ulcerative Colitis. *Sci Rep,* 7(1)**,** 7139.

Samuel, B. S., Shaito, A., Motoike, T., Rey, F. E., Backhed, F., Manchester, J. K., Hammer, R. E., Williams, S. C., Crowley, J., Yanagisawa, M. & Gordon, J. I. (2008). Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc Natl Acad Sci U S A,* 105(43)**,** 16767-72.

Scaldaferri, F., Vetrano, S., Sans, M., Arena, V., Straface, G., Stigliano, E., Repici, A., Sturm, A., Malesci, A., Panes, J., Yla-Herttuala, S., Fiocchi, C. & Danese, S. (2009). VEGF-A links angiogenesis and inflammation in inflammatory bowel disease pathogenesis. *Gastroenterology,* 136(2)**,** 585-95.e5.

Schinkel, A. H., Smit, J. J., van Tellingen, O., Beijnen, J. H., Wagenaar, E., van Deemter, L., Mol, C. A., van der Valk, M. A., Robanus-Maandag, E. C., te Riele, H. P. & et al. (1994). Disruption of the mouse mdr1a P-glycoprotein gene leads to a deficiency in the blood-brain barrier and to increased sensitivity to drugs. *Cell,* 77**,** 491-502.

Shimada, Y., Kinoshita, M., Harada, K., Mizutani, M., Masahata, K., Kayama, H. & Takeda, K. (2013). Commensal bacteria-dependent indole production enhances epithelial barrier function in the colon. *PLoS One,* 8(11)**,** e80604.

Shulzhenko, N., Morgun, A., Hsiao, W., Battle, M., Yao, M., Gavrilova, O., Orandle, M., Mayer, L., Macpherson, A. J., McCoy, K. D., Fraser-Liggett, C. & Matzinger, P. (2011). Crosstalk between B lymphocytes, microbiota and the intestinal epithelium governs immunity versus metabolism in the gut. *Nat Med,* 17(12)**,** 1585-93.

Tang, Y., Horikoshi, M. & Li, W. (2016). ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. *The R Journal,* 8.

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., Blanchard, C., Junt, T., Nicod, L. P., Harris, N. L. & Marsland, B. J. (2014). Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat Med,* 20(2)**,** 159-66.

Venkatesh, M., Mukherjee, S., Wang, H., Li, H., Sun, K., Benechet, A. P., Qiu, Z., Maher, L., Redinbo, M. R., Phillips, R. S., Fleet, J. C., Kortagere, S., Mukherjee, P., Fasano, A., Le Ven, J., Nicholson, J. K., Dumas, M. E., Khanna, K. M. & Mani, S. (2014). Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4. *Immunity,* 41(2)**,** 296-310.

Wu, G. D. (2014). Diet, the Gut Microbiome and the Metabolome in IBD. *Nestle Nutr Inst Workshop Ser,* 79**,** 73-82.

Yamada, T., Takahashi, D. & Hase, K. (2016). The diet-microbiota-metabolite axis regulates the host physiology. *J Biochem,* 160(1)**,** 1-10.

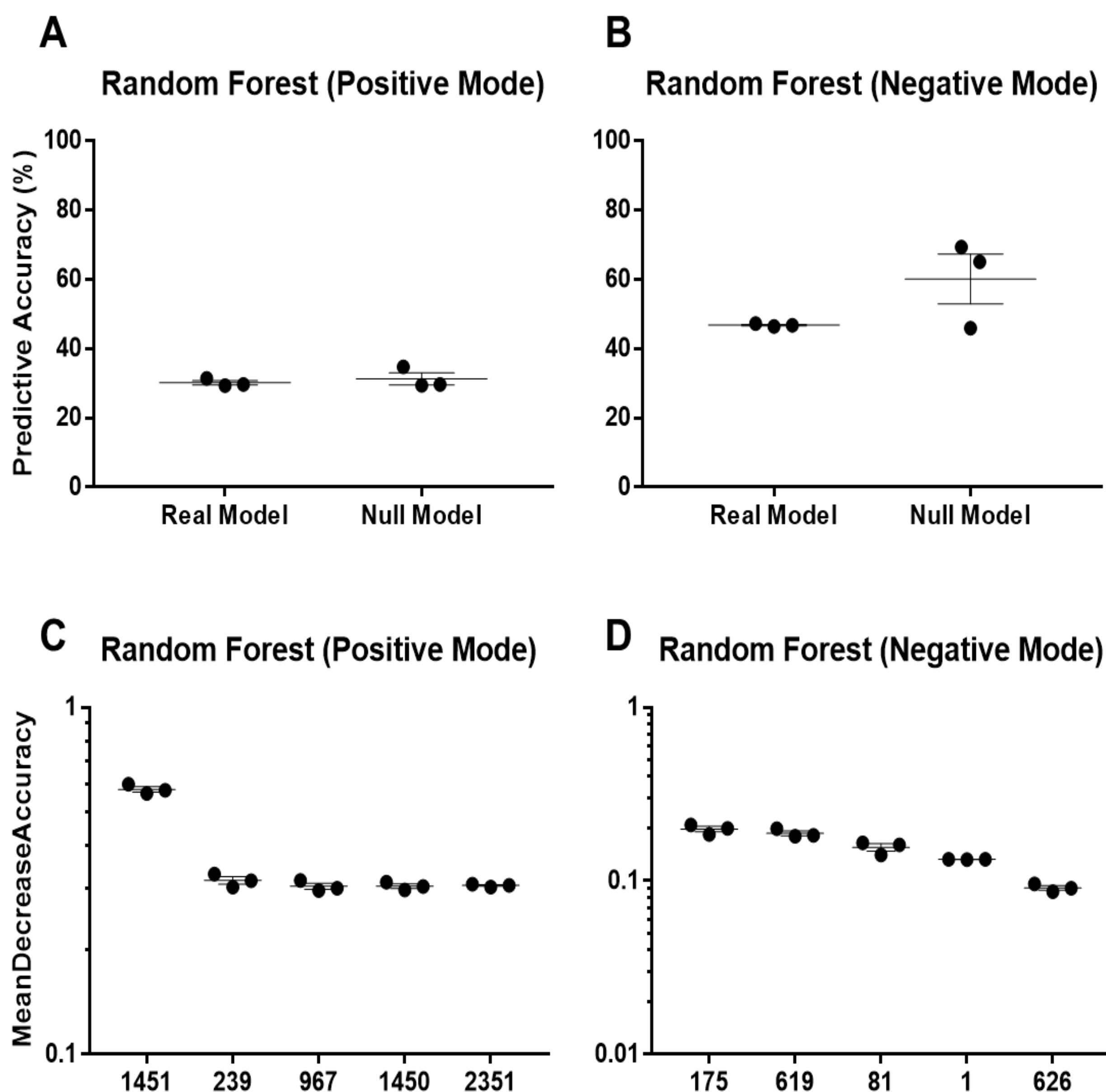
Zeissig, S., Burgel, N., Gunzel, D., Richter, J., Mankertz, J., Wahnschaffe, U., Kroesen, A. J., Zeitz, M., Fromm, M. & Schulzke, J. D. (2007). Changes in expression and distribution of

claudin 2, 5 and 8 lead to discontinuous tight junctions and barrier dysfunction in active Crohn's disease. *Gut,* 56(1)**,** 61-72.

Zhang, S., Jiao, T., Chen, Y., Gao, N., Zhang, L. & Jiang, M. (2014). Methylglyoxal induces systemic symptoms of irritable bowel syndrome. *PLoS One,* 9(8)**,** e105307.

Zhou, Y., Tu, C., Zhao, Y., Liu, H. & Zhang, S. (2016). Placental growth factor enhances angiogenesis in human intestinal microvascular endothelial cells via PI3K/Akt pathway: Potential implications of inflammation bowel disease. *Biochem Biophys Res Commun,* 470(4)**,** 967-74.

3.7 Supplementary Information



Supplementary Figure 1: Random Forest highlights important metabolites. Colonic mucus was extracted from 6 week old male, FVB background wildtype (WT) and *mdr1a*$^{-/-}$ mice. Metabolites were assayed via LC- and a matrix of the M/Z ratios was used as an input for a random forest model. The predictive accuracy of the model is shown for positive ionisation mode (A) and negative ionisation mode (B). The five most important metabolites identified via the Mean Decrease Accuracy value were then selected for positive ionisation mode (C) and negative ionisation mode (D). n = 3.

Supplementary Table 1: Metabolites identified in positive ionisation mode.

Peak ID	Query m/z	Name	Formula	Exact Mass
1451*	320.1993622	-	-	-
1459	326.2153168	Farnesylcysteine	C18H31NO2S	325.2075499
134	101.0596434	Methyl methacrylate;MMA;Methyl 2-methylpropenoate	C5H8O2	100.0524295
1450*	319.6975019	-	-	-
2133	550.3955128	PC(P-18:0/2:0)	C28H56NO7P	549.3794
1847	439.326948	(17Z)-1alpha,25-dihydroxy-26,27-dimethyl-17,20,22,22,23,23-hexadehydrovitamin D3 / (17Z)-1alpha,25-dihydroxy-26,27-dimethyl-17,20,22,22,23,23-hexadehydrocholecalciferol	C29H42O3	438.3134
2351*	632.737244	-	-	-
127	100.0755659	N-Methyl-2-pyrrolidinone;1-Methyl-2-pyrrolidinone	C5H9NO	99.06841392
967*	194.1148305	3,4-Methylenedioxymethamphetamine;N-Methyl-3,4-methylenedioxyamphetamine	C11H15NO2	193.1102787
2412	662.4650193	PC(15:1(9Z)/12:0)	C35H68NO8P	661.4683
1998	503.3248693	Kolanone	C33H42O4	502.3083098
2436	675.4319896	N-Acetyl-leu-leu-leu-leu-tyr-amide;NAc-L4Y-amide	C35H58N6O7	674.4366982
229	114.0912525	epsilon-Caprolactam	C6H11NO	113.084064
1467	329.262853	MG(0:0/16:1(9Z)/0:0)	C19H36O4	328.2613596
1630	379.24161	Cornudentanone	C22H34O5	378.2406
239*	115.0943557	-	-	-
1404	305.1830154	Dolichol phosphate	C15H29O4P	304.1803459
2172	562.3607859	N-Acetyl-leu-leu-leu-tyr-amide;NAc-L3Y-amide	C29H47N5O6	561.3526343
2276	612.3837117	Zizyphine A	C33H49N5O6	611.3682843
2358	633.0717273	Vescalin	C27H20O18	632.0649638
2373	652.4185759	PS(12:0/14:0)	C32H62NO10P	651.4111
2456	747.47431	Myxol glycoside/ (Myxoxanthophyll)	C46H66O8	746.4758
2470	772.5324081	-	C45H74NO7P	771.5203

** Identified as important in the Random Forest*

Supplementary Table 2: Metabolites identified in negative ionisation mode.

Peak ID	Query m/z	Name	Formula	Exact Mass
626*	242.9760248	3,5-Dichlorophenylcarboximide	C9H6Cl2N2O2	243.9806329
411	179.9803284	Unknown	-	-
503	196.9432906	6-BROMO-7H-PURINE	C5H3BrN4	197.9541087
81*	103.9299527	Unknown	-	-
619*	240.9108433	Unknown	-	-
1*	71.01383028	Methylglyoxal;Pyruvaldehyde;Pyruvic-aldehyde;2-Ketopropionaldehyde;2-Oxopropanal	C3H4O2	72.02112937
175*	129.0544771	3-Methyl-2-oxopentanoate;2-Oxo-3-methylvalerate	C6H10O3	130.0629942

** Identified as important in the Random Forest*

Chapter Four

The gut microbiota in eosinophil-deficient mice

Author Contributions

Gurdeep Singh: Designed and performed experiments, analysed data and wrote the manuscript.

Andrew Brass: Designed and supervised the project.

Sheena Cruickshank: Designed and supervised the project, critically reviewed and edited the manuscript

Christopher Knight: Designed and supervised the project.

4. Gut eosinophils and their impact on the mucus-dwelling microbiota

Gurdeep Singh[1], Andrew Brass[2], Christopher G. Knight[3] and Sheena M. Cruickshank[1].

[1] Faculty of Biology, Medicine and Health, Lydia Becker Institute of Immunology and Inflammation, Manchester Academic Health Science Centre, A.V. Hill Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom.

[2] Faculty of Biology, Medicine and Health, Division of Informatics, Imaging and Data Sciences, Stopford Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom.

[3] Faculty of Science and Engineering, School of Earth and Environmental SciencesMichael Smith Building, The University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom

Corresponding authors:

Professor Sheena Cruickshank

Faculty of Biology, Medicine and Health

The University of Manchester

AV Hill Building

Oxford Road

Manchester

M13 9PT

sheena.cruickshank@manchester.ac.uk

Phone +44 (0) 161 275 1582

Short title: **The gut microbiota in eosinophil-deficient mice**

4.1 Abstract

The gut has the largest commensal bacterial population in the body and its composition can be impacted by host factors such as production of IgA. Eosinophils in the gut have been implicated in the production of anti-bacterial factors and maintenance of IgA-secreting plasma cells. We used an eosinophil-deficient mouse (Δ*dblGATA-1*$^{-/-}$) to investigate the role of eosinophils in the regulation of the microbiota, with a particular emphasis on mucus-resident species in the small and large intestine.

We found no differences in IgA production or IgA expressing plasma cells between naive littermates in the small or large intestine. However, there were clear differences in the bacterial communities of the mucus and stools between WT vs Δ*dblGATA-1*$^{-/-}$ mice with the greatest separation between the mucus microbial communities. Mucus-resident bacteria in Δ*dblGATA-1*$^{-/-}$ mice had reduced diversity in the mucus compared with the stools. A qPCR panel of selected bacteria showed that that the most significant differences in the microbiota were between mucus-resident bacteria such as the abundance of Clostridiales and Bacteroides of the small and large intestine versus the stools.

Our data implicates eosinophils in the regulation of the microbiota, especially the bacteria most hyperlocal to the gut barrier. However, we propose that this is not via regulation of IgA plasma cells. Importantly, the data revealed that the mucus and stool microbiota are discrete communities and that stool analysis alone may be insufficient to comprehensively explore and define the role of the gut microbiota in health and disease.

4.2 Introduction

An effective gut barrier is crucial for our health by helping prevent entry of pathogens whilst enabling the entry of helpful nutrients and commensal bacterial products. The gut barrier comprises a mucus layer that is situated on top of a monolayer of intestinal epithelial cells. Intestinal epithelial cells and immune cells secrete and transport an array of anti-microbial factors that include anti-bacterial peptides and secreted IgA, which are particularly concentrated within the mucus layer (O'Neil et al., 1999). Another crucial component of the gut barrier is the gut microbiota which also serves to protect against pathogens by competing for resources or production of anti-microbial factors (Garcia-Gutierrez et al., 2018). A variety of host cells are involved in the maintenance of the gut barrier with immune cells such as eosinophils thought to play a key role (Johnson et al., 2015).

Eosinophils are resident immune cells found all along the length of the gut with the highest proportion in the small intestine (Forman et al., 2016). Eosinophils have been linked to the regulation of IgA-secreting plasma cells and the gut mucus layer (Chu et al., 2014), both major components of the gut barrier. Additionally, eosinophils have been linked to the maintenance of intestinal permeability (Johnson et al., 2015). Therefore, understanding how eosinophils contribute to barrier integrity and their impact on the gut microbiota is an important area of investigation.

The microbiota varies along the length of the gastrointestinal tract with fewer bacteria found at the top of the small intestine and the greatest number and most varied communities found in the large intestine (Ahmed et al., 2007). The gut microbiota then exists within two major niches within the gut, the lumen and the gut mucus layer. Given these distinctions among gut microbial communities, it is perhaps surprising that much of the work to date has focused on characterising stool bacteria that are unlikely to fully recapitulate this diversity. The importance of investigating different microbial niches is crucial, as it is likely that the microbiota within these niches have different functional impacts for the host. Indeed, our previous work has demonstrated that the colonic mucus microbiota, but not the stool bacteria, changes before the onset of colitis (Glymenaki et al., 2017).

Here, we investigate how eosinophil-deficiency impacts upon both IgA-secreting plasma cells and the gut microbiota, with a particular focus on exploring small versus large intestine and mucus versus stools. Using littermate-controlled WT and eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) mice, we saw no differences in the numbers of IgA-secreting plasma cells

between genotype. However, we show that both the stool and the mucus microbiota are significantly different between strains, with stronger separation in our mucus samples. Specific analysis of which types of bacteria were different indicated that the microbial niche (i.e. stool vs mucus) was a powerful influencing factor on the gut microbiota.

4.3 Materials and Methods

4.3.1 Animal Maintenance

ΔdblGATA-1 mice (Shivdasani et al., 1997) (kindly provided by Professor Avery August, Pennsylvania State University- USA) were bred with C57BL/6 mice to produce the F2 generation, which were used for all experiments. Mice from each litter were co-housed but separated by gender. Thus, WT male mice and heterozygous (Het) female mice, and *ΔdblGATA-1*$^{-/-}$ mice from the same litters were used for all subsequent experiments. It should be noted that due to the nature of the mutation, female mice could only be Het in this study but function as WTs, given the presence of eosinophils in these mice. Food (Beekay Rat and Mouse Diet No1 pellets; B&K Universal, UK) and water were available *ad libitum*. Ambient temperature was maintained at 21 (+/- 2°C) and the relative humidity was 55 (+/- 10%) with a 12h light/dark cycle. Male mice at 12 weeks old and female mice at 12 weeks and 1 year of age, were used for experiments. All animals were kept under specific, pathogen-free (SPF) conditions at the University of Manchester and experiments were performed according to the regulations issued by the Home Office under amended ASPA, 2012.

4.3.2 Sample preparation

Faecal samples were collected into sterile Eppendorf tubes and snap-frozen on dry ice. Small intestinal and distal colon snips were fixed in Carnoy's fixative (60% ethanol absolute, 30% chloroform and 10% glacial acetic acid) to preserve the mucus. The remaining colon was then bisected and any remaining faecal matter was removed and gently washed away with PBS (Sigma, Poole, UK). The inner surface of the colon was scraped using a cell scraper to remove mucus from the mucus lining which was snap-frozen. Serum was incubated at 37ºC for 2 hours, before centrifugation at 7000xg for 10 minutes to collect the supernatant. The supernatant was stored at -80ºC.

4.3.3 Histology and Staining

Carnoy's fixed samples were incubated in two changes of dry methanol (Sigma, UK) for 30 minutes each, followed by absolute ethanol (ThermoFisher Scientific, Crawley, UK) for two incubations at 30 minutes each. Tissue cassettes were processed in a Micro-spin Tissue Processor STP120 (ThermoFisher Scientic) and immersed in paraffin using a Leica Biosystems embedding station (Leica Biosystems, Milton Keynes, UK), with the luminal surface of the colon exposed for tissue sectioning. 5µm tissue sections were cut using a Leica Biosystems microtome and adhered to uncoated microscope slides (ThermoFisher Scientific).

Slides were dried for 48 hours at 50ºC before use. Haematoxylin and Eosin (H+E) and goblet cell staining were performed and analysed as described previously (Glymenaki et al., 2017).

4.3.4 Fluorescence *in-situ* Hybridisation (FISH)

FISH was performed as described previously (Glymenaki et al., 2017). In brief, FISH staining was performed using the universal bacterial probe-EUB338 (5′-Cy3-GCTGCCTCCCGTAGGAGT-3′), followed by immunostaining with a rabbit polyclonal MUC2 antibody and goat anti-rabbit Alexa-Fluor 488 antibody (Life technologies, Paisley, United Kingdom). The thickness of the inner mucus barrier was quantified by measuring the distance between the epithelium and the outer mucus layer. Bacterial penetrance was also investigated by assigning values from 0-4 depending on where bacteria were localised: 0 = bacteria in the lumen and outer mucus layer, 1 = bacteria in the inner mucus layer, 2 = bacteria in contact with the epithelium, 3 = bacteria in the crypts, 4 = bacteria in the lamina propria. All slides were scored in a blind manner.

4.4.5 Flow Cytometry for IgA Plasma Cells

Whole colon and small intestine were harvested and fat was removed from the tissue and washed thoroughly in PBS Tissue was placed into warm strip buffer (PBS containing 1% Foetal Bovine Serum, 0.5M EDTA and 0.2mM DTT (all Sigma), and incubated for 10 minutes at 37ºC on a shaking incubator (Cole Palmer, Staffordshire, UK) at 205rpm. The supernatant was removed, and the process repeated before incubation in digest buffer (RPMI, 10% FCS, 1% penicillin-streptomycin, 0.025mg/ml Deoxyribonuclease I from bovine pancreas (all Sigma) and 0.01% collagenase-dispase (Roche, West Sussex, UK) for 45 minutes at 37ºC on a shaking incubator at 205rpm. Cells were strained through a 70µm sieve and pelleted at 450xg for 5 minutes. Cells were resuspended in FACS buffer (0.5% BSA in PBS) and stained with B220-APCCy7, CD3-PERCP, CD19-AmCyan, MHCII-PacBlue (ThermoFisher Scientific), IgA-PE (eBioscience- ThermoFisher Scientific) and CD45-BV650 (BioLegend, California, USA). After staining, cells were fixed in IC Fixation Buffer (ThermoFisher Scientific), resuspended in FACS buffer and then analysed using a BD Biosciences LSR Fortessa (BD Biosciences, Berkshire, UK). Doublets and dead cells were excluded from analysis. IgA+ plasma cells were reported as a proportion of total live CD45+ cells. IgA+ plasma

4.5.6 IgA ELISA

An ELISA was performed using the Invitrogen™ eBioscience™ Mouse IgA ELISA Ready-SET-Go!™ Kit (Fisher Scientific, UK), according to manufacturer's instructions. Serum was used at a 1/5000 dilution. Stool samples were prepared using IDK Extract® stool preparation tubes according to manufacturer's instructions (Immundiagnostik, Bensheim, Germany). The resulting faecal extract was used at a 1/100 dilution. The ELISA was read using a VersaMax microplate reader (Molecular Devices, San Jose, USA).

4.5.7 16S DNA extraction, amplification and purification

DNA extraction was performed using a Qiamp® Fast Stool DNA Mini Kit (Qiagen, Manchester, UK), using a modified version of the manufacturer's instructions. Faecal samples were incubated in Inhibitex buffer (Qiagen) and mechanically disaggregated, before incubation at 95°C for 30 minutes. Mucus samples were centrifuged (13000xg for 10 minutes) and the mucus pellets incubated in Inhibitex buffer, mechanically disaggregated and incubated at 95°C for 30 minutes. 300uL of the resulting lysate was used for the subsequent steps, which were then performed according to manufacturer's instructions. Optical density at 260nm was recorded using a UV1101 spectrophotometer (Biochrom Ltd., Cambridge, UK) to measure DNA concentration.

For the identification of different bacterial species, the 16S rRNA gene was amplified using the universal primer pairs P3_GC-341F and P2_518 (**Table 1**), as described previously (Glymenaki et al., 2017).

4.5.8 Denaturing gradient gel electrophoresis (DGGE)

A denaturing gradient gel was prepared according to the methods initially developed by Fischer and Lerman (Fischer and Lerman, 1983). The gel was run as described previously (Glymenaki et al., 2017).

The gel was then analysed using Phoretix Software (TOTALLAB, Newcastle Upon Tyne, UK). Lane boundaries were defined to correct for any potential distortions during the gel run and manually curated to ensure that the bands detected were not artefacts. Reference bands were selected to align bands across the gel and 'Rf values' generated to measure the bands migration. A binary matrix was then generated based on the Rf values, with 0's and 1's indicating the absence or the presence of a bacterial 'species' in a sample. This matrix was processed in R (R Core Team, 2016). Specifically, the 'VEGAN' (Oksanen et al., 2016), 'ecodist' (Goslee and Urban, 2007) and 'MASS' (Venables and Ripley, 2002) packages were

used to statistically compare the presence or absence of species (bands) between groups, i.e. WT lumen, WT mucus, *ΔdblGATA-1*$^{-/-}$ lumen and *ΔdblGATA-1*$^{-/-}$ mucus. The same packages were used to perform non-metric multidimensional scaling to examine clustering between groups. Finally, bacterial diversity (total number of bands) was calculated using the Shannon-Wiener Diversity Index.

4.5.9 qPCR Analysis of Bacteria

Stool, colonic mucus and small intestinal mucus DNA was used as a template for a qPCR reaction that consisted of 100ng/µL DNA, 10uM primers, 2x KAPA SYBR® FAST qPCR Mastermix (Sigma) and nuclease free water to a total volume of 20µL. Primers used are illustrated in **Table 1**. A universal 16S rRNA gene was used as a housekeeping control. Results were calculated using the ΔΔCT method. Cycling steps for all primers were: Denaturing at 95 ºC for 10 minutes, followed by 35 cycles of 95 ºC for 15 seconds and 60 ºC for 1 minute.

4.5.10 Statistical Analysis

All statistical analysis was performed using either GraphPad Prism 7 (GraphPad Software, La Jolla, USA) or R. Student's *t* Test was used to compare crypt and villi length, muscle wall thickness, goblet cell number, inner mucus thickness and bacterial localisation between genotype in male mice. Permutational multivariate analysis of variance (PERMANOVA) was used to calculate differences in species presence between samples using the 'adonis' function in the vegan R package (Oksanen et al., 2016). Relative expression of quantified bacteria was compared between niche and genotype using 2 Way Analysis of Variance (ANOVA) with a Tukey's Post Hoc Test. Additionally, when comparing young and old female mice, 2 Way Analysis of Variance (ANOVA) with a Tukey's Post Hoc Test was used to compare crypt and villi length, muscle wall thickness, goblet cell number, inner mucus thickness and bacterial localisation between genotype.

Table 1: List of primers used

Gene	Primers	Sequence
GATA-1	GATA 1 S	5’-CCCAATCCTCTGGACTCCCA-3’
	GATA 1 AS	5’-CCTACTGTGTACCAGGCTAT-3’
16S rRNA	P3_GC-341F	5’-CGCCCGCCGCGCGCGGCGGGCGGGGCG GGGGCACGGGGGGCCTACGGGAGGCAGCAG-3’
	P2_518R	5’-ATTACCGCGGCTGCTGG-3’
Universal	UniF340	5’-ACTCCTACGGGAGGCAGCAGT-3’
	UniR514	5’-ATTACCGCGGCTGCTGGC-3’
Actinobacteria	Act664F	TGTAGCGGTGGAATGCGC
	Act941R	AATTAAGCCACATGCTCCGCT
Bacteriodetes	Bac960F	5’-GTTTAATTCGATGATACGCGAG-3’
	Bac1100R	TTAASCCGACACCTCACGG
Deferribacteres	Defer1115F	CTATTTCCAGTTGCTAACGG
	Defer1265R	GAGHTGCTTCCCTCTGATTATG
Firmicutes	Firm934F	5’-GGAGYATGTGGTTTAATTCGAAGCA-3’
	Firm934R	5’-AGCTGACGACAACCATGCAC-3’
Verrucomicrobia	Ver1165F	5’-TCAKGTCAGTATGGCCCTTAT-3’
	Ver1263R	5’-GAGHTGCTTCCCTCTGATTATG-3’
Bacteroides	BacteroidesF	5’-GGTTCTGAGAGGAGGTCCC-3’
	BacteroidesR	5’-GCCTCAAGGGCACAACCTCCAAG-3’
Clostridiales	ClostF	ACTCCTACGGGAGGCAGC
	ClostR	GCTTCTTAGTCAGGTACCGTCAT
Enterrobacteriaceae	Uni515F	GTGCCAGCAGCCGCGGTAA
	Ent826R	GCCTCAAGGGCACAACCTCCAAG
Lachnospiraceae/Rumminococceae	LachnoRumF	CGGTACCTGACTAAGAAGC
	LachnoRumR	AGTTTCATTCTTGCGAACG
Akkermansia muciniphila	Amuc_1599F	GACCGGCATGTTCAAGCAGACT
	Amuc_1599R	AAGCCGCATTGGGATTATTTGTT
Segmented Filamentous Bacteria	SFBF	GACGCTGAGGCATGAGAGCAT
	SFBR	GACGGCACGGATTGTTATTCA

4.4 Results

4.4.1 Absence of eosinophils does not affect colonic or small intestinal morphology under homeostasis in younger mice

Our overall aim was to investigate how the absence of eosinophils could influence the gut microbiota. However, we first needed to establish whether there were any changes to overall gut architecture in eosinophil-deficient mice, specifically changes in intestinal crypt length and muscle wall thickness.

H+E staining revealed that there were no significant differences in either crypt length or muscle wall thickness of young male mice (12 weeks of age, Figure 1). Additionally, we investigated the morphology of the small intestine and found no significant differences in villus length or muscle wall thickness between genotypes (Figure 1). Thus, the absence of eosinophils did not alter gross gut morphology in male Δ*dblGATA-1*$^{-/-}$ mice.

When female mice were investigated, there were also no significant differences in muscle wall thickness between Het and Δ*dblGATA-1*$^{-/-}$ mice, in 12 week old mice (young) or mice aged to one year (old) (Supplementary Figure 1). Young Δ*dblGATA-1*$^{-/-}$ mice tended to have longer colonic crypts than their Het counterparts (2 way ANOVA with Tukey's Post Hoc Test: $P = 0.011$), but this was restored to levels akin to Het in older mice (Supplementary Figure 1). There were no significant morphological differences in the small intestine between Het and Δ*dblGATA-1*$^{-/-}$ mice of any age (Supplementary Figure 2).

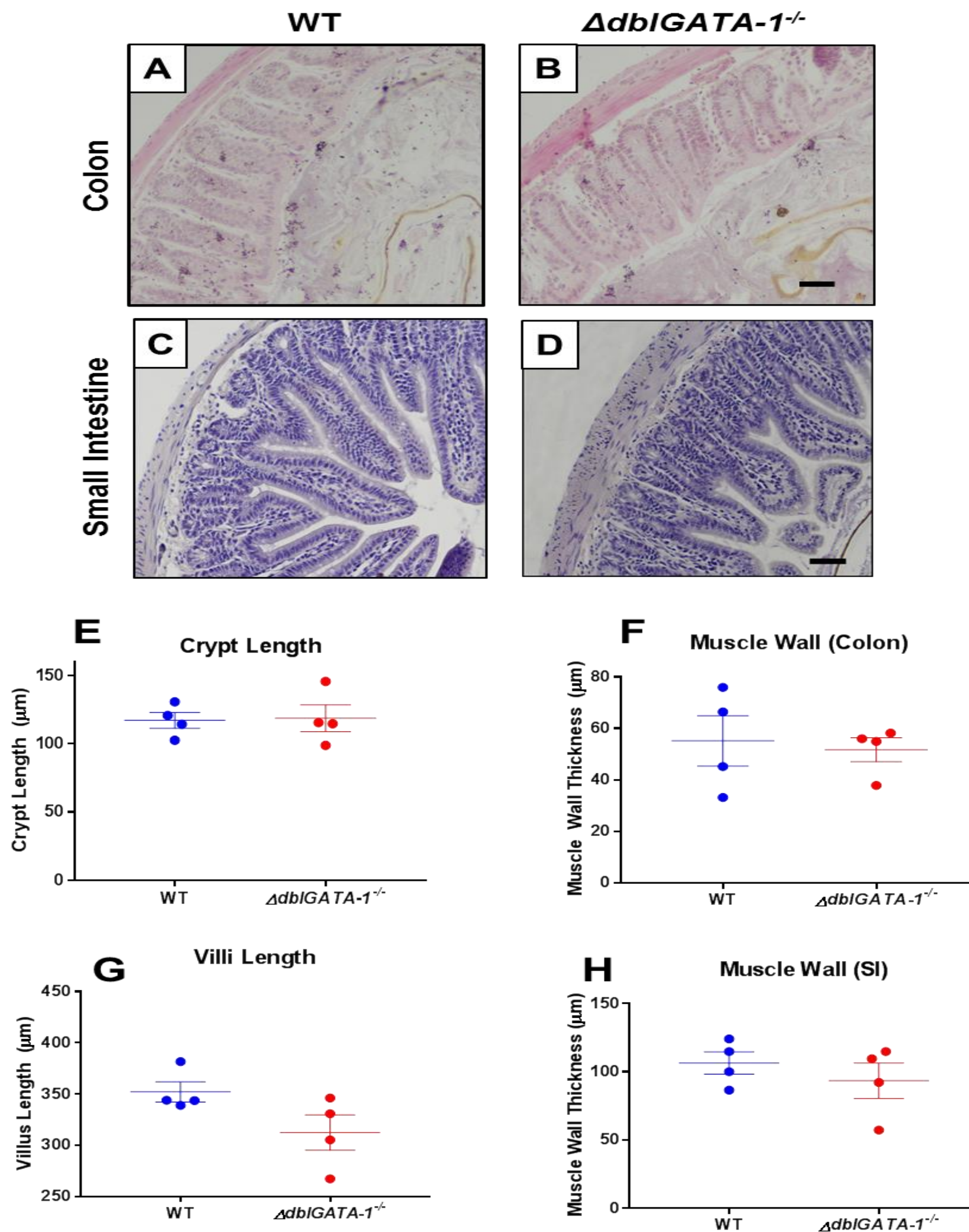


Figure 1: Loss of eosinophils does not alter gut architecture. Colonic and small intestinal tissue sections from 12-week old male, C57BL/6 background wildtype (WT) and eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) littermate mice, were stained with haematoxylin and eosin to observe the gut morphology. Representative images were taken for the colon of WT mice (A) and Δ*dblGATA-1*$^{-/-}$ mice (B), and the small intestine of WT (C) and Δ*dblGATA-1*$^{-/-}$ mice (D). Colonic crypt length (E) and muscle wall thickness (F) and small intestinal villus length (G) and muscle wall thickness (H) were measured. Data is shown as mean +/- standard error of the mean (SEM). n = 4. Scale bar = 50μm.**4.4.2 Mucus barrier integrity**

Eosinophils may play a role in the regulation of the mucus barrier (Chu et al., 2014) and so we quantified the number of goblet cells and the thickness of the colonic mucus layer (Figure 2). Goblet cell numbers in the colon were equivalent between genotype (Figure 2E). There were also no significant differences in small intestinal goblet cell number (Supplementary Figure 3). We also investigated the thickness of the inner mucus layer and saw a similar inner mucus thickness between genotypes in male mice (Figure 2F). Similar trends were seen in female mice (Supplementary Figure 4 and 5) although younger, female eosinophil-deficient mice had more variable mucus thickness when compared to Het mice, which was significantly different between genotype (2 way ANOVA with Tukey's Post Hoc Test: P = 0.04, Supplementary Figure 6).

Irrespective of mucus thickness, the quality of the mucus barrier may be altered such that bacterial localisation could be affected. We therefore stained colon sections with a fluorescent universal probe for bacteria, and a Muc2 antibody (Figure 2). Sections were scored based on how far bacteria had travelled from the lumen to the lamina propria. Two mice that lacked eosinophils had evidence of increased bacterial localisation further into the gut, however most of the Δ*dblGATA-1*$^{-/-}$ mice were equivalent to WT counterparts (Figure 2G). Furthermore, bacterial localisation was similar between WT and Δ*dblGATA-1*$^{-/-}$ female mice irrespective of age (Supplementary Figure 6).

As IgA is a key component of the mucus barrier, we determined whether eosinophils play a role in the regulation of IgA-secreting plasma cells (Figure 3). Flow cytometry was used to define plasma cells which were gated as CD45+, MHCII+, CD3-, IgA+ and B220-. Analysis of colonic and small intestinal cells isolated from male, WT and Δ*dblGATA-1*$^{-/-}$ mice revealed no striking differences in the number of IgA-expressing cells between genotype (Figure 3A-B). This was concordant with no differences in serum or faecal IgA levels between young male WT and Δ*dblGATA-1*$^{-/-}$ mice (Figure 3C-D). However, in female mice there was a trend towards higher levels of serum IgA in Δ*dblGATA-1*$^{-/-}$ mice compared with Het, with significant differences in younger Δ*dblGATA-1*$^{-/-}$ mice compared with Het (2 way ANOVA with Tukey's Post Hoc Test: P = 0.002) (Supplementary Figure 7).

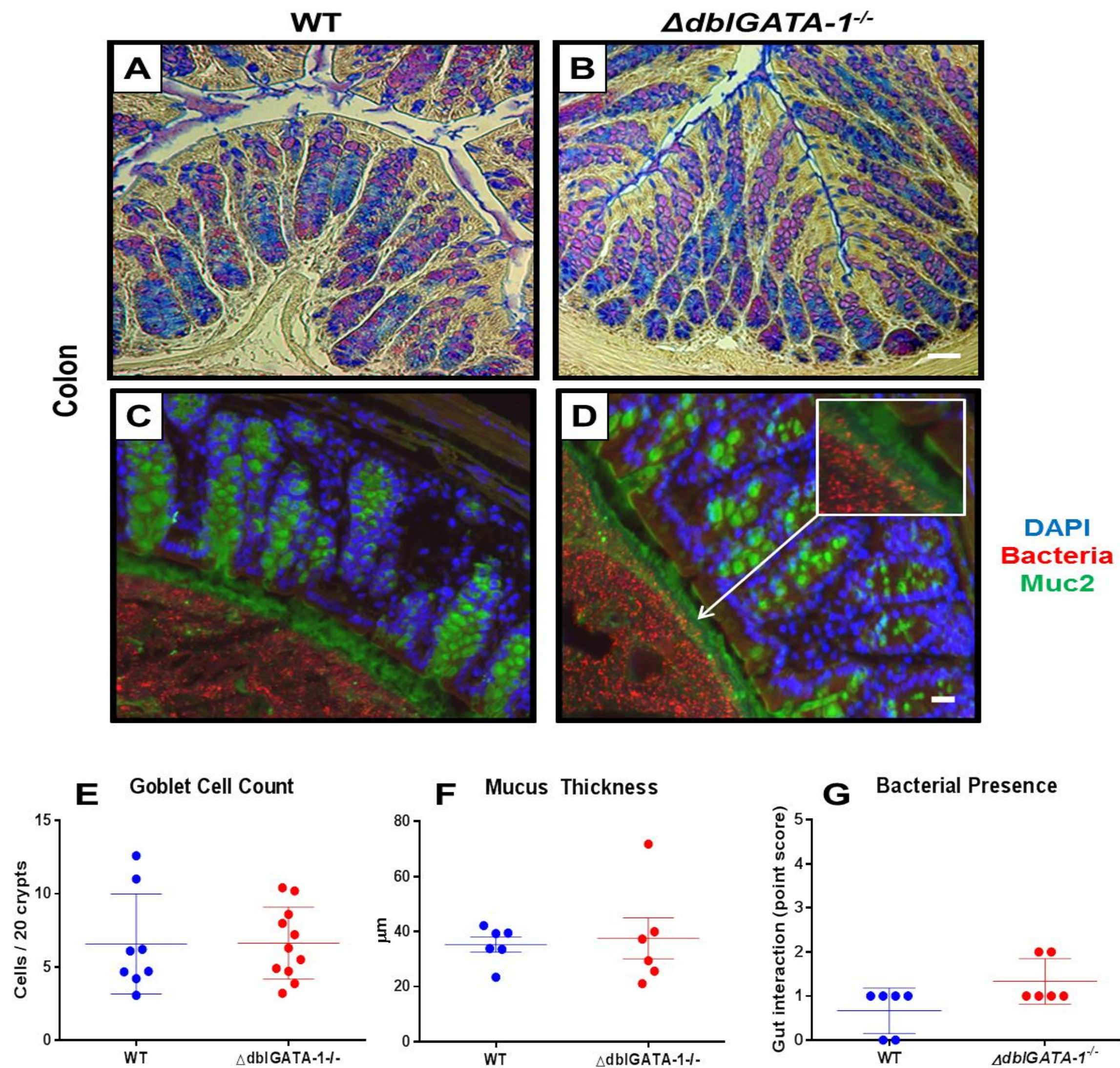


Figure 2: Mucus barrier integrity in *ΔdblGATA-1*$^{-/-}$ mice. Colonic tissue sections were taken from 12 week old male, C57BL/6 background wildtype (WT) and eosinophil-deficient (*ΔdblGATA-1*$^{-/-}$) littermate mice. Goblet cells were stained using periodic acid, Alcian blue and Schiff's reagent and representative images for WT mice (A) and Δ*dblGATA-1*$^{-/-}$ mice (B) are displayed. Sections were also stained with a fluorescent DNA probe specific for the 16S rRNA gene to identify bacteria (red), a Muc2 antibody (green) to identify mucus and counterstained with DAPI (blue). Representative images for WT (C) and Δ*dblGATA-1*$^{-/-}$ mice (D) are displayed. Goblet cells (E), inner mucus thickness (F) and bacterial localisation (G) were measured. Bacteria were scored based on their location within the gut: 0 = bacteria in the lumen and outer mucus layer, 1 = bacteria in the inner mucus layer, 2 = bacteria in contact with the epithelium, 3 = bacteria in the crypts, 4 = bacteria in the lamina propria. Data shown as mean +/-SEM. n = 6-11. Scale bars = 50µm.

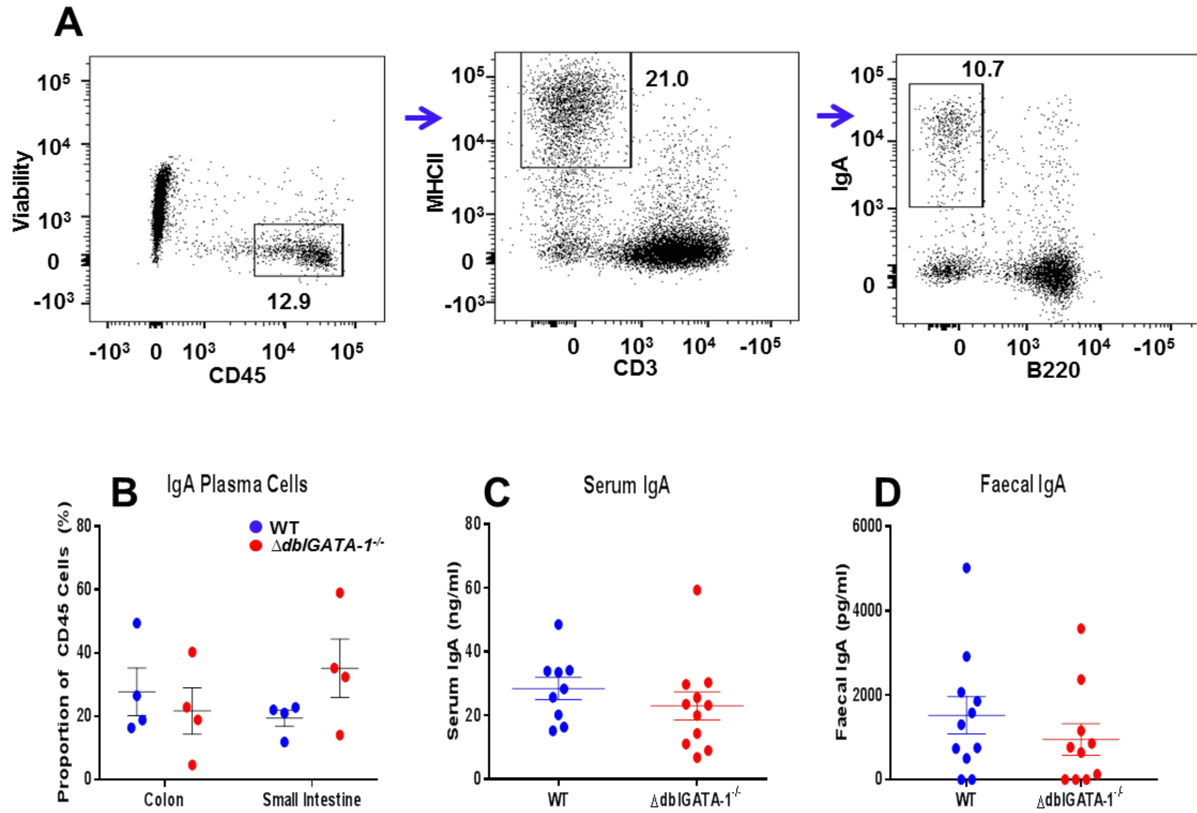


Figure 3: Loss of eosinophils does not impact IgA. Cells isolated from the colon and small intestine, from C57BL/6 background wildtype (WT) and Δ*dblGATA-1* littermate mice were gated to calculate the frequency of IgA+ plasma cells, gated as CD45+, MHCII+, CD3-, IgA+ and B220-.(A). Numbers of IgA+ cells were reported as a total proportion of live CD45+ cells (B). Serum (C) and stool homogenates (D) were analysed via ELISA to determine secreted levels of IgA. Data shown as mean +/- standard error of the mean (SEM). n = 4-11

4.4.3 Eosinophils influence microbial diversity

We next investigated whether the bacterial communities themselves could be affected by eosinophil-deficiency. We used a gel-fingerprinting technique (DGGE) to provide an overview of the bacterial communities in our WT and Δ*dblGATA-1*$^{-/-}$ mice (Figure 4). Although the stool communities were significantly different between genotypes (PERMANOVA: P = 0.001), the mucus microbiota had a clearer and more significant separation between genotypes (PERMANOVA: P = 0.00003). Alpha-diversity, a measure of the range of different bacteria within samples, revealed that eosinophil-deficient mice had significantly lower diversity in the mucus compared with WT mucus (2 Way ANOVA with Tukey's Post Hoc Test: P = 0.0018). In contrast, there was no difference in stool alpha diversity between genotypes.

In aged female mice, there was no clear difference between Het and Δ*dblGATA-1*$^{-/-}$ stool although the stool microbiota was significantly different when comparing young and old female Δ*dblGATA-1*$^{-/-}$ mice (PERMANOVA: P = 0.004) (Supplementary Figure 8). We also saw no differences in stool diversity between Het and Δ*dblGATA-1*$^{-/-}$ mice (Supplementary Figure 8). When the mucus was examined in female mice, we saw significant differences between young Het and young Δ*dblGATA-1*$^{-/-}$ mice (PERMANOVA: P = 0.02) but not the older mice. However, there was an age effect with young and old Het mice and young and old Δ*dblGATA-1*$^{-/-}$ mice having significantly separated bacterial fingerprints (PERMANOVA: P = 0.01 and P = 0.004 respectively) (Supplementary Figure 9).

We also saw a significant reduction in mucus diversity in old relative to young Δ*dblGATA-1*$^{-/-}$ mice. However, we did not see a reduction in mucus diversity with age in Het genotypes. Taken together, the data from male and female mice would suggest that the lack of eosinophils does impact on the microbiome with a stronger influence on the mucus microbiota compared with the stool communities.

We then quantified the differences in the microbiota via qPCR (Figure 5). A panel of the most common bacterial phyla, orders, families and species that comprise the gut microbiota were selected. Unexpectedly, we saw no significant differences in overall microbial burden or in the levels of common gut phyla, orders and species between genotype. However, there were striking differences in the bacteria among microbial niches, with stool versus colonic mucus or small intestinal mucus containing significantly different bacteria. For example, at the phylum level in WT mice, we saw a significant reduction in Firmicutes (2 way ANOVA

with Tukey's Post Hoc Test: P = 0.04) and a significant increase in small intestinal Actinobacteria (2 way ANOVA with Tukey's Post Hoc Test: P = 0.004) when comparing stool and small intestinal mucus. In Δ*dblGATA-1*$^{-/-}$ mice, we saw differences between the stool and small intestinal mucus, with a significant reduction in Firmicutes (2 way ANOVA with Tukey's Post Hoc Test: P = 0.01) and a significant increase in Actinobacteria (2 way ANOVA with Tukey's Post Hoc Test: P = 0.0006) in the small intestine. We saw no differences in the level of the phylum Bacteroidetes. At the order level, we saw a significant reduction in Bacteroides when comparing WT stool and small intestinal mucus (2 way ANOVA with Tukey's Post Hoc Test: P = 0.01). In Δ*dblGATA-1*$^{-/-}$ mice, we saw a significant increase in Clostridiales when comparing stool and colonic mucus (2 way ANOVA with Tukey's Post Hoc Test: P = 0.0007) and there was a significant reduction in Clostridiales in the small intestinal mucus compared to the colonic mucus (2 way ANOVA with Tukey's Post Hoc Test: P = 0.0001). At the family and genus level, we saw no significant differences in Lachnospiraceae, *Lactobacilli*, Ruminococcaceae and segmented filamentous bacteria (SFB) (Supplementary Figure 9).

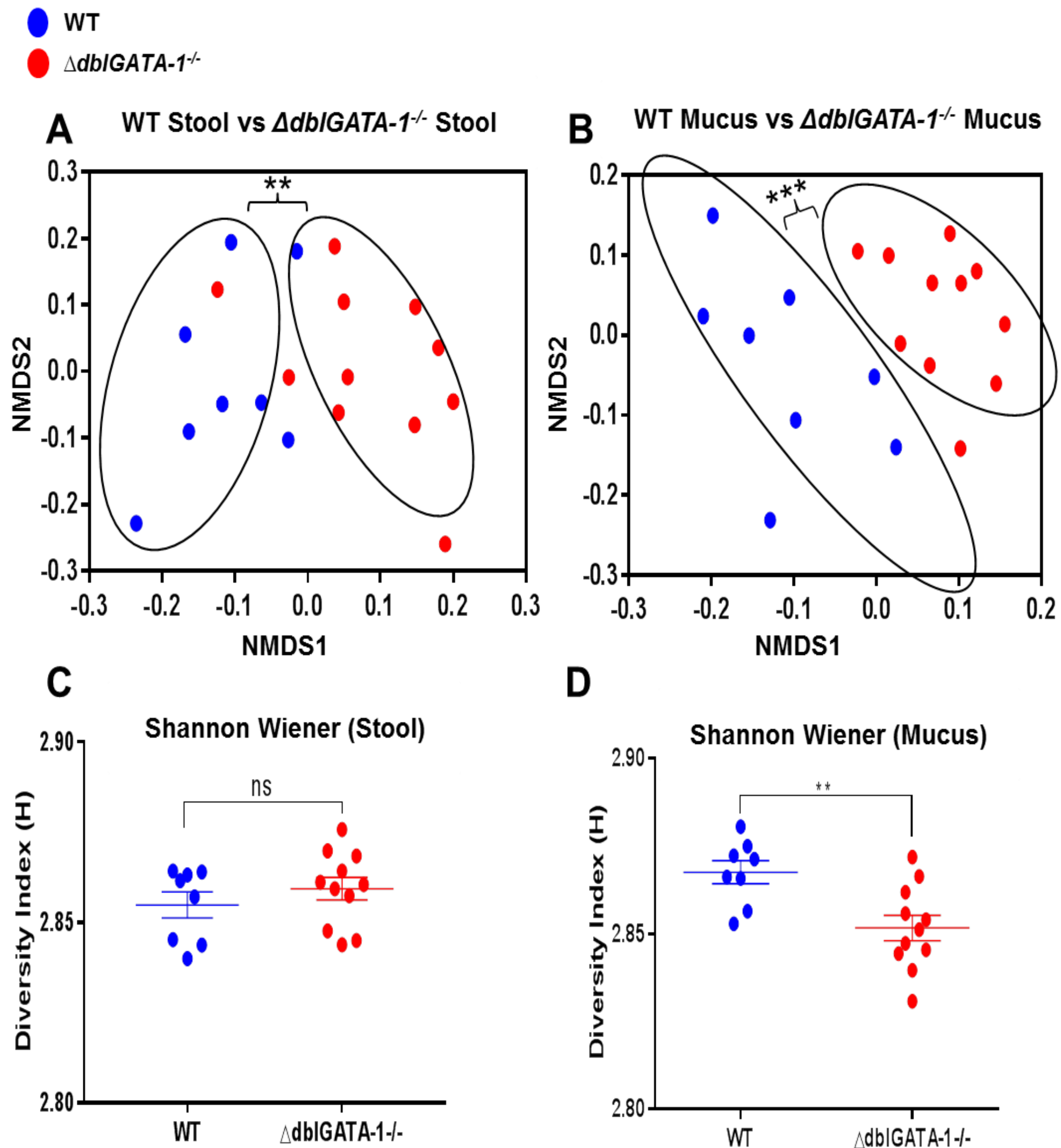


Figure 4: Absence of eosinophils leads to significantly altered gut microbiota. Differences in bacterial species composition and diversity between the stools and mucus of C57BL/6 background WT and *ΔdblGATA-1*$^{-/-}$ littermate mice, were analysed by DGGE. Differences in the bacterial communities were plotted using non-metric multidimensional scaling (NMDS), for the stool (A) and mucus communities (B). Rings indicate a significant difference between the bacterial communities of the respective treatment groups, as determined by permutational multivariate analysis of variance (** p = <0.01, *** p = <0.001). Subsequent diversity analysis was then performed on stool (C) and mucus (D). Data shown as mean +/- standard error of the mean (SEM). n = 8-11.

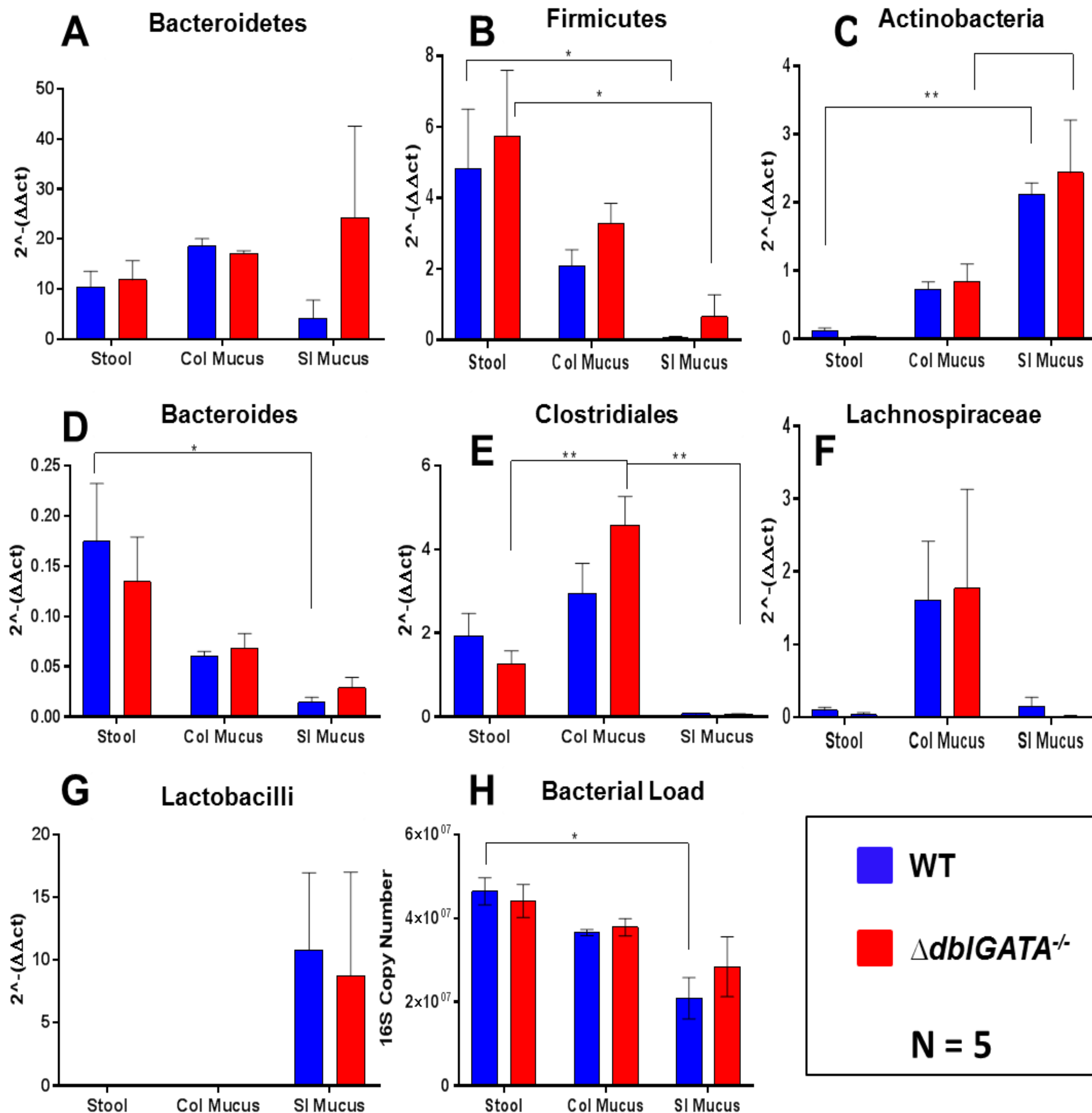


Figure 5: Bacterial communities significantly different between microbial niches. Real-time quantitative PCR was used to assess expression of gut bacteria in C57BL/6 background WT and *ΔdblGATA-1*$^{-/-}$ littermate mice, in stool, colonic (col) mucus and small intestinal (SI) mucus. The relative expression of Bacteriodetes (A), Firmicutes (B), Actinobacteria (C), Bacteroides (D), Clostridiales (E), Lachnospiraceae and Ruminococcaceae (F), Enterrococcaeae (G) and Lactobacilli (H) are illustrated. Overall bacterial load is also displayed (I). Data shown as mean +/- standard error of the mean (SEM). Asterisks represent significant as determined by 2 Way Analysis of Variance with a Sidak's Post Hoc Test (* p = <0.05, ** p = <0.01). n = 5.

4.5 Discussion

Eosinophils have previously been implicated in harmful inflammatory contexts such as allergy (Ramirez et al., 2018). However, more recent studies suggest that eosinophils play an active role in the maintenance of gut homeostasis, for example, in the regulation of IgA-secreting plasma cells (Chu et al., 2014; Forman et al., 2016) and the mucus barrier. However, we found no differences in IgA levels when comparing young WT and eosinophil-deficient mice, but showed a significant increase in IgA cells in older eosinophil-deficient mice compared to younger mice. Previous studies have shown conflicting results with regards to the impact of eosinophils on IgA plasma cells with data implicating both positive and negative effects (Chu et al., 2014; Forman et al., 2016; Haberland et al., 2018). One difference between our present study and the aforementioned studies is strain, where we use C57BL/6 mice as opposed to mice on a BALB/c background. It was previously shown that BALB/c mice natively have significantly higher levels of IgA production compared to C57BL/6 mice (Fransen et al., 2015). It is also possible that differences in IgA may be niche specific and become apparent under the context of infection or inflammation. Our previous work showed that IgA differences were most apparent in infection-induced inflammation and only in the small intestine (Forman et al., 2016). There may also be an effect of sex influencing IgA levels, as within our study only young, female eosinophil-deficient mice had significantly higher levels of IgA in the serum compared to Het. However, this could be a consequence of female control mice being heterozygous, as opposed to true wildtypes.

Given the location of eosinophils within the gut, there could be cross-talk between eosinophils and intestinal epithelial cells that maintain gut homeostasis that do not involve IgA. Indeed, eosinophils have been shown to influence mucus thickness (Chu et al., 2014). However, in our hands we saw no changes in goblet cell numbers or mucus thickness. Physical changes in the properties of the mucus, for example glycosylation and viscosity however could still impact on the microbiome. Our finding that there was not consistent altered gut bacterial penetrance suggests that it did not impact on bacterial localisation at least in homeostasis. However eosinophils could still directly impact on the makeup of the microbiota. For example, eosinophils contain a variety of cytotoxic granules that can have a notable physiological impact on the host (Acharya and Ackerman, 2014). Indeed eosinophils interact with the gut microbiota to limit *Clostridium dificile* infection in an IL-25-dependent manner (Buonomo et al., 2016). It was notable in our study that mucus bacteria were markedly different from the stool bacteria as revealed by qPCR, fingerprint profiling and

alpha diversity analysis. Given our observation that there were no striking differences in IgA production, we would hypothesise that the impact on mucus resident bacteria is either via production of epithelial derived anti-bacterial peptides or secretion of anti-bacterial factors from the eosinophils themselves.

With regards to the microbiota, it is known that cage effect has a marked impact (Campbell et al., 2012). The microbiota is hugely dependent on the mother, food and the environment that the mouse is reared in (Cabrera-Rubio et al., 2012; Lukens et al., 2014) and so robust studies into the microbiota should control for the environment. Indeed, cage effect has been reported to have a stronger impact on gut microbiota variation than genetics (Hildebrand et al., 2013). Thus, it is important to consider whether littermate controls have been used in a study. Many microbiota studies do not report the use of littermate controls or crucial information about animal housing that could influence results (Bramhall et al., 2015; Florez-Vargas et al., 2016). Although previous work (Chu et al., 2014) showed that there was a significantly altered microbial community in the stools of Δ*dblGATA-1*$^{-/-}$ mice, littermate controls were not reported as being used. Importantly, we controlled for littermates and saw that although there was an effect on stool microbiota, the most striking differences were in the mucus-resident bacteria of littermate controls. Indeed, we demonstrated a reduction in mucus-resident alpha diversity with no changes to stool alpha diversity in mice that lacked eosinophils. Reduced gut microbial diversity is often associated with diseases such as inflammatory bowel disease (Menni et al., 2017).

Although we saw overall differences in the gut microbiota by DGGE, our specific analysis of common bacterial phyla, orders and families revealed no differences between genotypes in our mice. It is possible that the bacteria significantly different between genotype were not encompassed in our qPCR panel of common gut bacteria. However, our qPCR and DGGE analysis of stool, colonic and small intestinal mucus highlighted differences in the microbial population among those niches. It is known that the microbial composition between stool and mucus is significantly different (Li et al., 2015; Tang et al., 2015; Glymenaki et al., 2017). For example, one study showed that mucus samples had a greater relative abundance of *Proteobacteria* and *Fusobacteria* than stool, although the stool had a greater proportion of *Firmicutes and Bacteroidetes* (Tang et al., 2015). Taken together, these data emphasise the need to investigate the different microbial niches within the gut, to comprehensively explore the gut microbiota and the impact of genotype or environment.

We demonstrate that eosinophils did not influence IgA production, goblet cell number or mucus thickness. However, eosinophils did play a role in the regulation of the gut microbiota, with a notably greater impact on the mucus bacteria, rather than the stool bacteria. Our data highlights that a focus on stool is insufficient to capture the overall complexity of the gut microbiota.

Given the location of eosinophils within the gut, there could be cross-talk between eosinophils and intestinal epithelial cells that maintain gut homeostasis that do not involve IgA. Indeed, eosinophils have been shown to influence mucus thickness (Chu et al., 2014). However, in our hands we saw no changes in goblet cell numbers or mucus thickness. Physical changes in the properties of the mucus, for example glycosylation and viscosity however could still impact on the microbiome. Our finding that there was not altered gut bacterial penetrance suggests that it did not impact on bacterial localisation; however eosinophils could still impact on the makeup of the microbiota. Indeed eosinophils have also been shown to interact with the gut microbiota in imparing *Clostridium dificile* infection in an IL-25-dependent manner (Buonomo et al., 2016). It was notable in our study that our mucus bacteria were markedly different from the stool bacteria as revealed by fingerprint profiling and alpha diversity analysis. Given our observation that there were no striking differences in IgA production, we would hypothesise that the impact on mucus resident bacteria is either via production of epithelial derived anti-bacterial peptides or secretion of anti-bacterial factors from the eosinophils themselves.

With regards to the microbiota, it is known that cage effect has a marked impact (Campbell et al., 2012). The microbiota is hugely dependent on the mother, food and the environment that the mouse is reared in (Cabrera-Rubio et al., 2012; Lukens et al., 2014) and so robust studies into the microbiota should control for the environment. Indeed, cage effect has been reported to have a stronger impact on gut microbiota variation than genetics (Hildebrand et al., 2013). Thus, it is important to consider whether littermate controls have been used in a study. Many microbiota studies do not report the use of littermate controls or crucial information about animal housing that could influence results (Bramhall et al., 2015; Florez-Vargas et al., 2016). Although previous work (Chu et al., 2014) showed that there was a significantly altered microbial community in the stools of $\Delta dblGATA\text{-}1^{-/-}$ mice, littermate controls were not used. Importantly, we controlled for littermates and saw that although there was an effect on stool microbiota, the most striking differences were in the mucus-resident bacteria of littermate controls. Indeed, we demonstrated a reduction in mucus-resident alpha diversity

with no changes to stool alpha diversity in mice that lacked eosinophils. Within the gut environment, reduced microbial diversity is often associated with diseases such as inflammatory bowel disease and high diversity is associated with a healthy phenotype (Menni et al., 2017) and thus it may be that this makes the mice more susceptible to inflammatory disease.

Although we saw overall differences in the gut microbiota by DGGE, it is curious to note that our specific analysis of common bacterial phylum, orders and families revealed no differences between genotypes in our mice. One study identified a significant reduction in Bacteroidetes in the stools of *ΔdblGATA-1*$^{-/-}$ mice compared with WT; however it is not clear whether this study used littermate controls (Jung et al., 2015). We analysed stool, colonic and small intestinal mucus and highlighted differences in microbial population within those niches. It is known that the microbial composition between stool and mucus is significantly different (Li et al., 2015; Tang et al., 2015; Glymenaki et al., 2017). The fact that many significant differences in bacteria between colonic mucus and small intestinal mucus were found in mice lacking eosinophils, could support the suggestion by Forman et al. (2016) that eosinophils may play different roles in the colon and small intestine. Although the qPCR screen of bacteria was unable to highlight genotypic differences in the gut microbiota, highthroughput sequencing may allow the elucidation of the specific differences in the gut microbiota that were not demonstrated in the present study. Taken together, these data emphasise the need to investigate the different microbial niches within the gut, to comprehensively explore the gut microbiota and the impact of genotype or environment.

We demonstrate that eosinophils did not influence IgA production, goblet cell number or mucus thickness. However, eosinophils did play a role in the regulation of the gut microbiota, with a notably greater impact on the mucus bacteria, rather than the stool bacteria. Our data highlights that a focus on stool is insufficient to capture the overall diversity of the gut microbiota.

References

Ahmed, S., Macfarlane, G. T., Fite, A., McBain, A. J., Gilbert, P. & Macfarlane, S. (2007). Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol,* 73**,** 7435-42.

Bramhall, M., Florez-Vargas, O., Stevens, R., Brass, A. & Cruickshank, S. (2015). Quality of methods reporting in animal models of colitis. *Inflamm Bowel Dis,* 21(6)**,** 1248-59.

Buonomo, E. L., Cowardin, C. A., Wilson, M. G., Saleh, M. M., Pramoonjago, P. & Petri, W. A., Jr. (2016). Microbiota-Regulated IL-25 Increases Eosinophil Number to Provide Protection during Clostridium difficile Infection. *Cell Rep,* 16(2)**,** 432-443.

Cabrera-Rubio, R., Collado, M. C., Laitinen, K., Salminen, S., Isolauri, E. & Mira, A. (2012). The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *Am J Clin Nutr,* 96**,** 544-51.

Campbell, J. H., Foster, C. M., Vishnivetskaya, T., Campbell, A. G., Yang, Z. K., Wymore, A., Palumbo, A. V., Chesler, E. J. & Podar, M. (2012). Host genetic and environmental effects on mouse intestinal microbiota. *Isme j,* 6(11)**,** 2033-44.

Chu, V. T., Beller, A., Rausch, S., Strandmark, J., Zanker, M., Arbach, O., Kruglov, A. & Berek, C. (2014). Eosinophils promote generation and maintenance of immunoglobulin-A-expressing plasma cells and contribute to gut immune homeostasis. *Immunity,* 40**,** 582-93.

Fischer, S. G. & Lerman, L. S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proc Natl Acad Sci U S A,* 80(6)**,** 1579-83.

Florez-Vargas, O., Brass, A., Karystianis, G., Bramhall, M., Stevens, R., Cruickshank, S. & Nenadic, G. (2016). Bias in the reporting of sex and age in biomedical research on mouse models. *Elife,* 5.

Forman, R., Bramhall, M., Logunova, L., Svensson-Frej, M., Cruickshank, S. M. & Else, K. J. (2016). Eosinophils may play regionally disparate roles in influencing IgA(+) plasma cell numbers during large and small intestinal inflammation. *BMC Immunol,* 17(1)**,** 12.

Fransen, F., Zagato, E., Mazzini, E., Fosso, B., Manzari, C., El Aidy, S., Chiavelli, A., D'Erchia, A. M., Sethi, M. K., Pabst, O., Marzano, M., Moretti, S., Romani, L., Penna, G., Pesole, G. & Rescigno, M. (2015). BALB/c and C57BL/6 Mice Differ in Polyreactive IgA Abundance, which Impacts the Generation of Antigen-Specific IgA and Microbiota Diversity. *Immunity,* 43(3)**,** 527-40.

Garcia-Gutierrez, E., Mayer, M. J., Cotter, P. D. & Narbad, A. (2018). Gut microbiota as a source of novel antimicrobials. *Gut Microbes***,** 1-21.

Glymenaki, M., Singh, G., Brass, A., Warhurst, G., McBain, A. J., Else, K. J. & Cruickshank, S. M. (2017). Compositional Changes in the Gut Mucus Microbiota Precede the Onset of Colitis-Induced Inflammation. *Inflamm Bowel Dis,* 23(6)**,** 912-922.

Goslee, S. C. & Urban, D. L. (2007). The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *22,* 22(7)**,** 1-19.

Haberland, K., Ackermann, J. A., Ipseiz, N., Culemann, S., Pracht, K., Englbrecht, M., Jack, H. M., Schett, G., Schuh, W. & Kronke, G. (2018). Eosinophils are not essential for maintenance of murine plasma cells in the bone marrow. *Eur J Immunol*.

Hildebrand, F., Nguyen, T. L., Brinkman, B., Yunta, R. G., Cauwe, B., Vandenabeele, P., Liston, A. & Raes, J. (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol,* 14(1)**,** R4.

Johnson, A. M., Costanzo, A., Gareau, M. G., Armando, A. M., Quehenberger, O., Jameson, J. M. & Olefsky, J. M. (2015). High fat diet causes depletion of intestinal eosinophils associated with intestinal permeability. *PLoS One,* 10(4)**,** e0122195.

Jung, Y., Wen, T., Mingler, M. K., Caldwell, J. M., Wang, Y. H., Chaplin, D. D., Lee, E. H., Jang, M. H., Woo, S. Y., Seoh, J. Y., Miyasaka, M. & Rothenberg, M. E. (2015). IL-1beta in eosinophil-mediated small intestinal homeostasis and IgA production. *Mucosal Immunol,* 8(4)**,** 930-42.

Li, H., Limenitakis, J. P., Fuhrer, T., Geuking, M. B., Lawson, M. A., Wyss, M., Brugiroux, S., Keller, I., Macpherson, J. A., Rupp, S., Stolp, B., Stein, J. V., Stecher, B., Sauer, U., McCoy, K. D. & Macpherson, A. J. (2015). The outer mucus layer hosts a distinct intestinal microbial niche. *Nat Commun,* 6**,** 8292.

Lukens, J. R., Gurung, P., Vogel, P., Johnson, G. R., Carter, R. A., McGoldrick, D. J., Bandi, S. R., Calabrese, C. R., Walle, L. V., Lamkanfi, M. & Kanneganti, T. D. (2014). Dietary modulation of the microbiome affects autoinflammatory disease. *Nature*.

Menni, C., Jackson, M. A., Pallister, T., Steves, C. J., Spector, T. D. & Valdes, A. M. (2017). Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *Int J Obes (Lond),* 41(7)**,** 1099-1105.

O'Neil, D. A., Porter, E. M., Elewaut, D., Anderson, G. M., Eckmann, L., Ganz, T. & Kagnoff, M. F. (1999). Expression and regulation of the human beta-defensins hBD-1 and hBD-2 in intestinal epithelium. *J Immunol,* 163(12)**,** 6718-24.

Oksanen, J., Guillaume-Blanchet, F., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H. & Wagner, H. (2016). *vegan: Community Ecology Package* [Online]. Available: https://CRAN.R-project.org/package=vegan [Accessed 27th October 2017].

Ramirez, G. A., Yacoub, M. R., Ripa, M., Mannina, D., Cariddi, A., Saporiti, N., Ciceri, F., Castagna, A., Colombo, G. & Dagna, L. (2018). Eosinophils from Physiology to Disease: A Comprehensive Review. *Biomed Res Int,* 2018**,** 9095275.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing* [Online].Available: https://www.r-project.org/ [Accessed 16th January 2017].

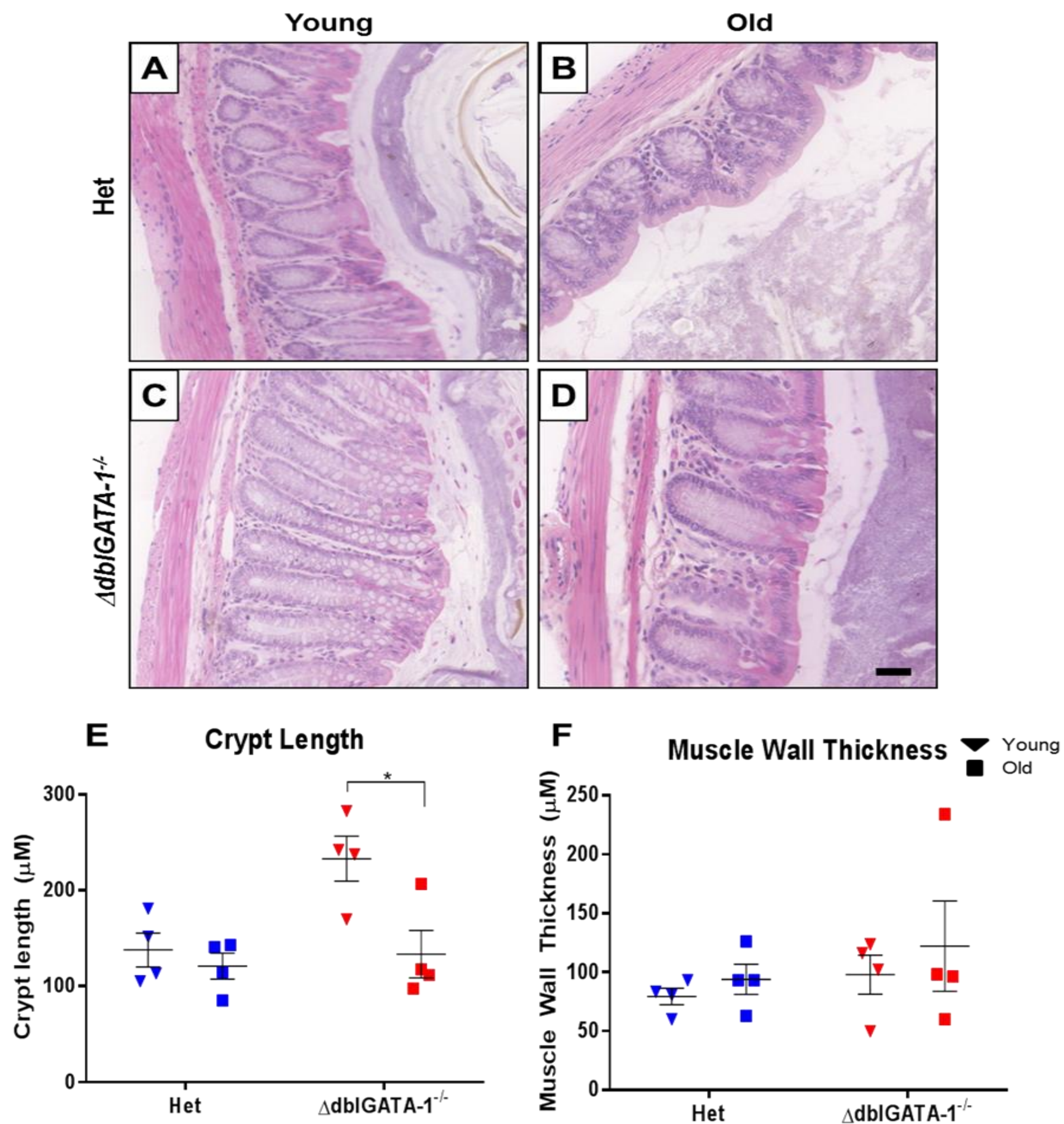
Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A. & Orkin, S. H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *Embo j,* 16**,** 3965-73.

Tang, M. S., Poles, J., Leung, J. M., Wolff, M. J., Davenport, M., Lee, S. C., Lim, Y. A., Chua, K. H., Loke, P. & Cho, I. (2015). Inferred metagenomic comparison of mucosal and

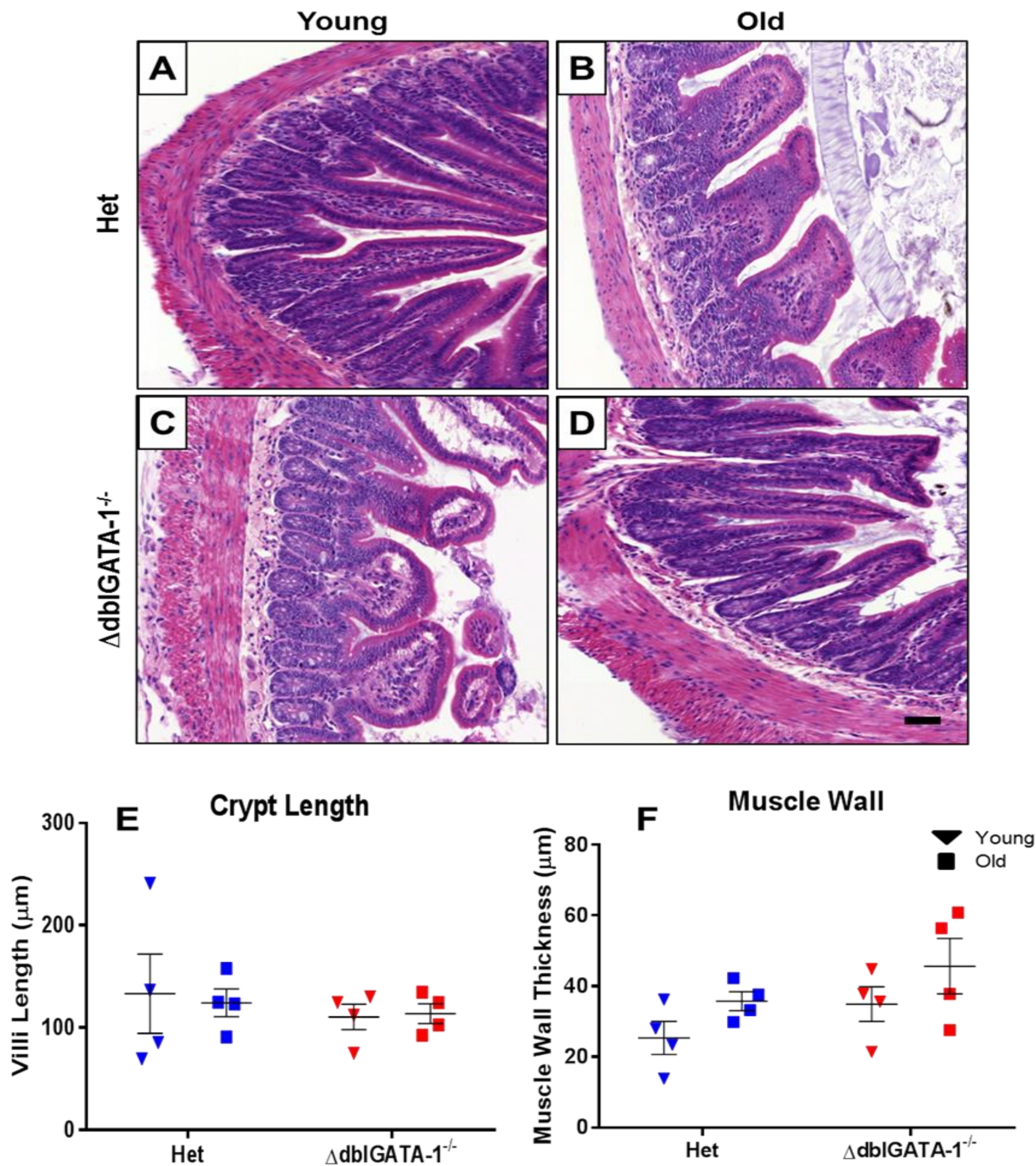
fecal microbiota from individuals undergoing routine screening colonoscopy reveals similar differences observed during active inflammation. *Gut Microbes,* 6(1)**,** 48-56.

Venables, W. & Ripley, N. (2002). *Modern Applied Statistics with S.* New York: Fourth Edition.

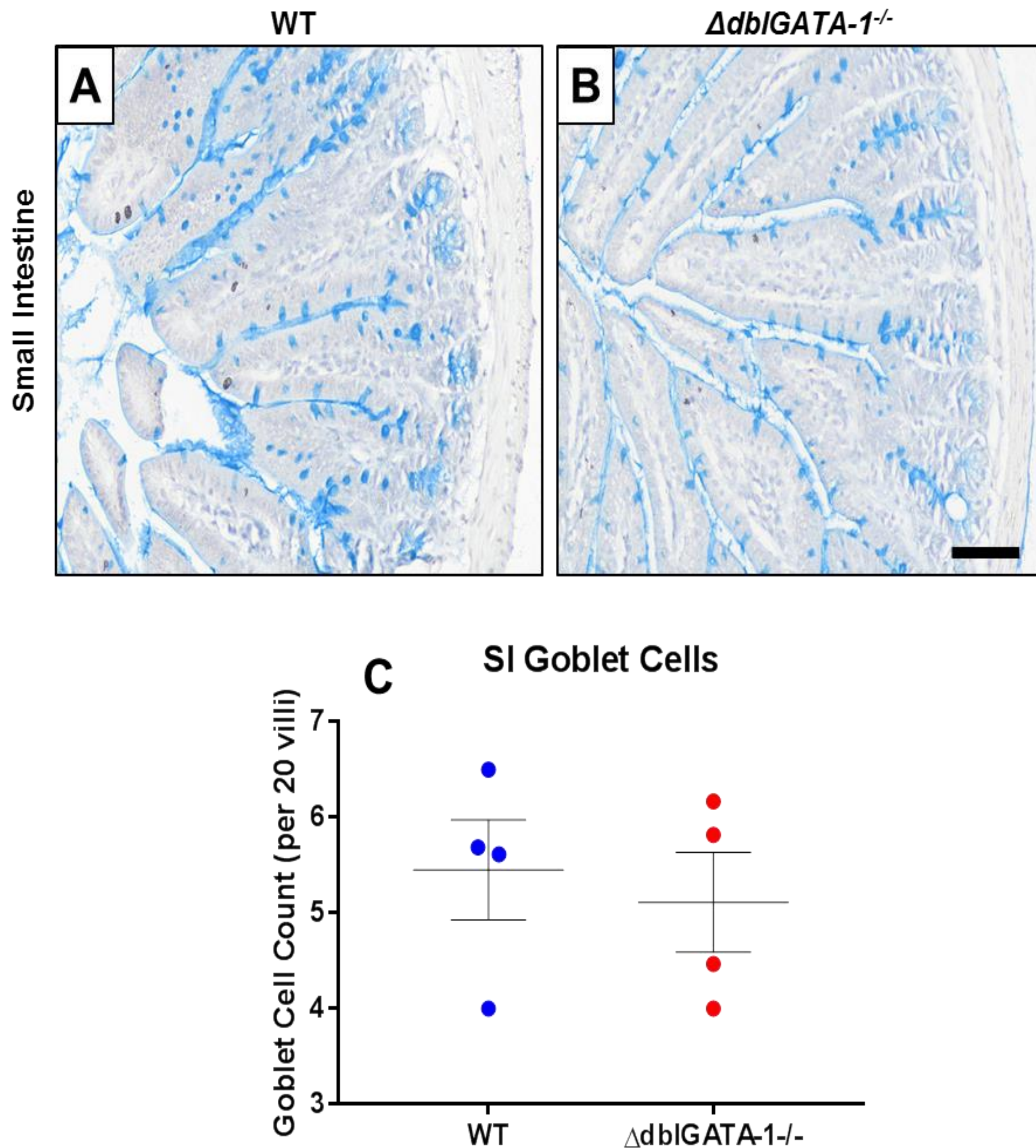
4.6 Supplementary Information



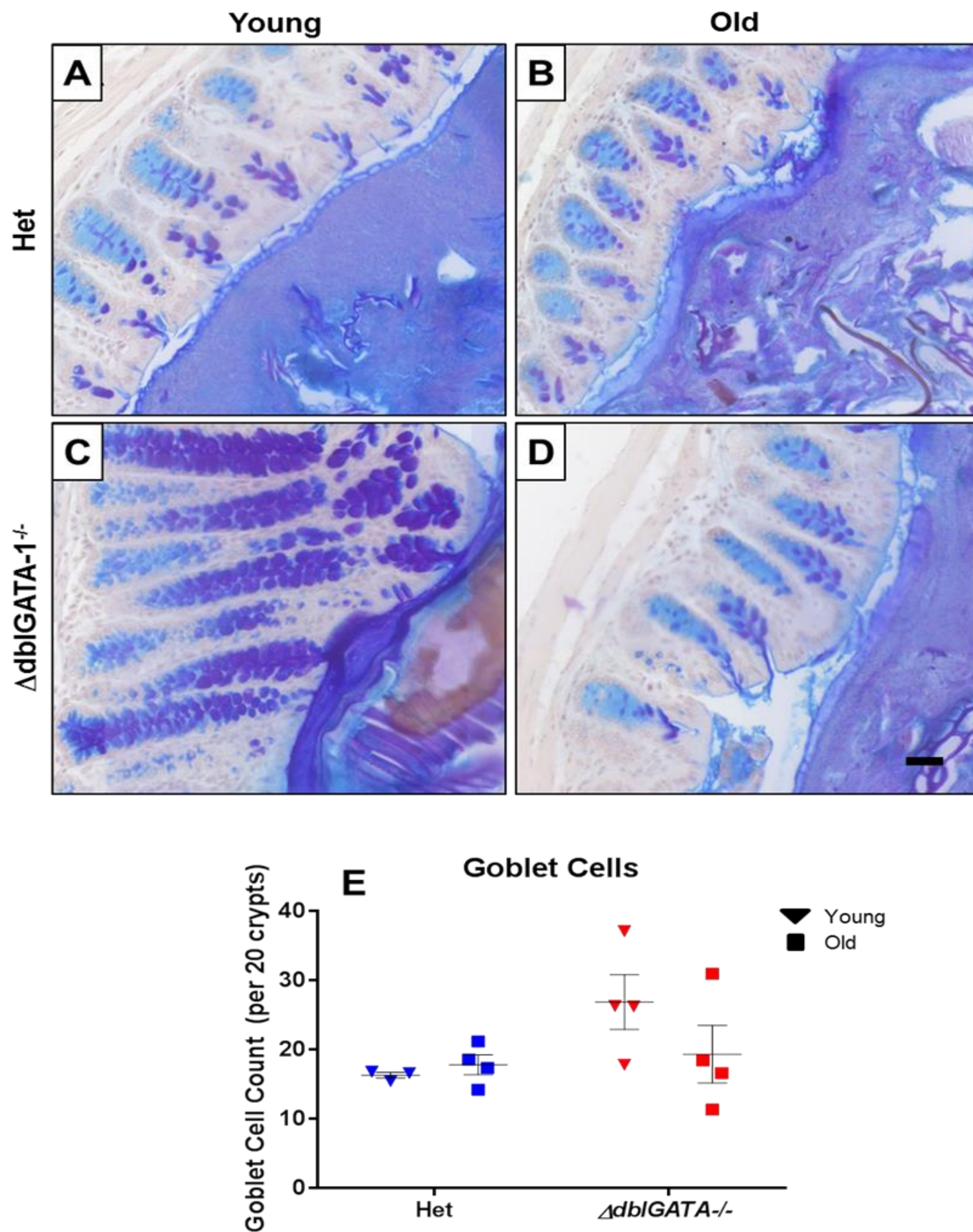
Supplementary Figure 1: Loss of eosinophils leads to altered gut morphology in old Δ*dblGATA-1*$^{-/-}$ mice. Colonic tissue sections from female 12 week (young) and 12 month (old) old, C57BL/6 background heterozygous (Het) and eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) mice, were stained with haematoxylin and eosin to observe the gut morphology. Representative images of young Het mice (A); old Het mice (B); young Δ*dblGATA-1*$^{-/-}$ mice (C) and old Δ*dblGATA-1*$^{-/-}$ mice (D). Colonic crypt length (E) and muscle wall thickness (F) were measured. Data is shown as mean +/- standard error of the mean (SEM). Asterisks represent significant as determined by 2 Way Analysis of Variance with a Tukey's Post Hoc Test ($p^* = <0.05$). n = 4. Scale bar = 50μm.



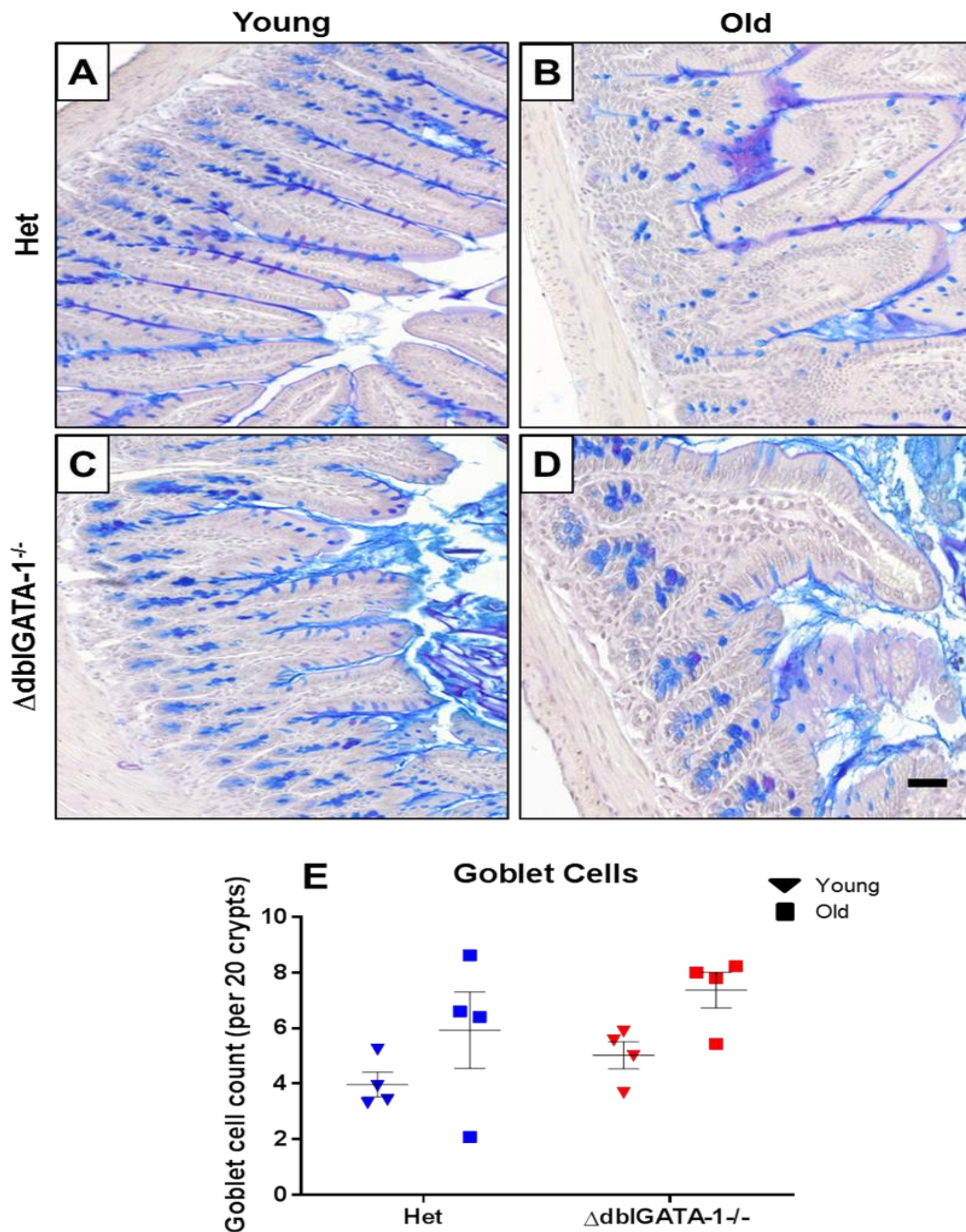
Supplementary Figure 2: Loss of eosinophils did not impact on morphological differences in small intestine structure. Small intestinal tissue sections from 12 week (young) and 12 month (old) old, C57BL/6 background heterozygous (Het) and eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) mice, were stained with haematoxylin and eosin. Representative images were taken from young mice (A) and old (B) Het mice; young (C) and old (D) Δ*dblGATA-1*$^{-/-}$ mice and small intestinal crypt length (E) and muscle wall thickness (F) were measured. Data is shown as mean +/- standard error of the mean (SEM). Asterisks represent significant as determined by 2 Way Analysis of Variance with a Tukey's Post Hoc Test ($p^* = <0.05$). n = 4. Scale bar = 50μm.



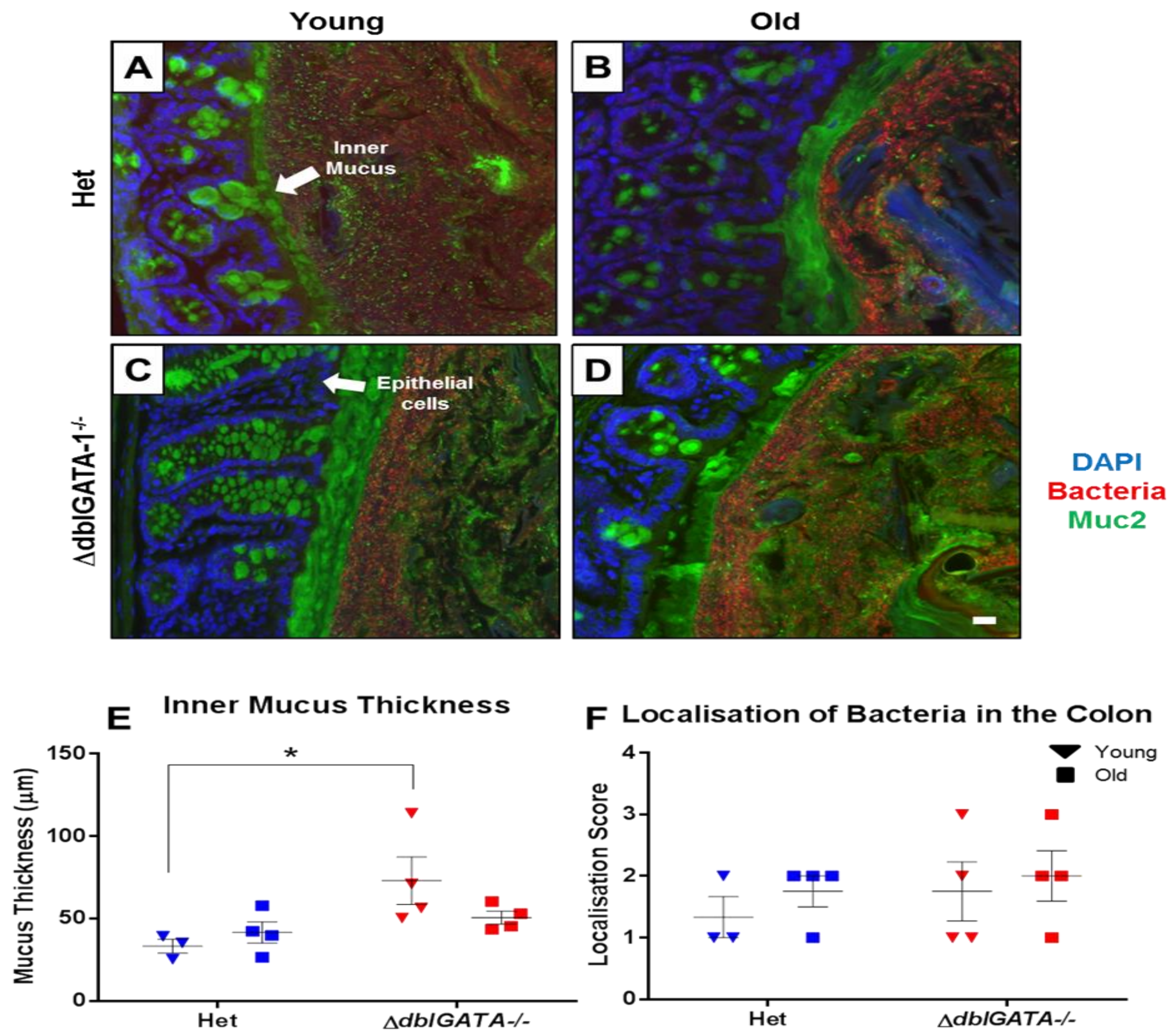
Supplementary Figure 3: Lack of eosinophils does not impact upon small intestinal goblet cells. Small intestinal (SI) tissue sections were taken from 12-week old male, C57BL/6 background wildtype (WT) and eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) littermate mice. Goblet cells were stained using periodic acid, Alcian blue and Schiff's reagent and representative images for WT mice (A) and Δ*dblGATA-1*$^{-/-}$ mice (B) are displayed. Goblet cells were quantified (C). Data is shown as mean +/- standard error of the mean (SEM). n = 4. Scale bar = 100µm.



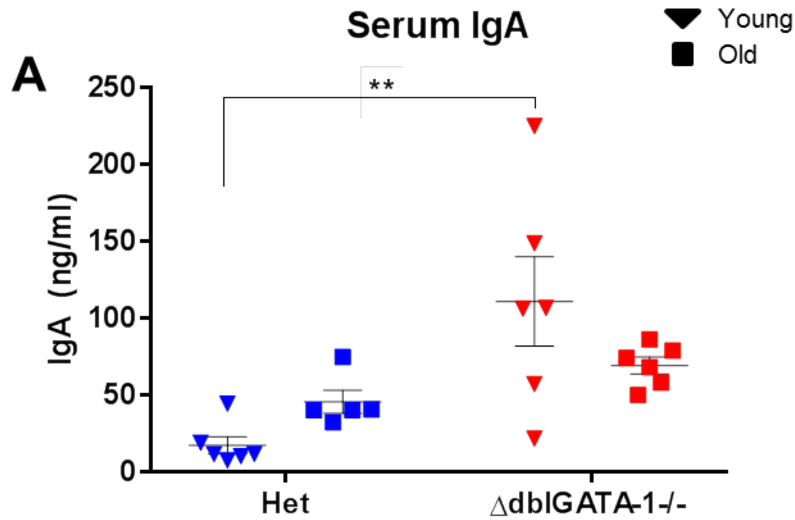
Supplementary Figure 4: Loss of eosinophils does not impact upon goblet cell number in young or old mice. Colonic tissue sections from 12 week (young) and 12 month (old) old, C57BL/6 background heterozygous (Het) and eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) mice, were stained for mucopolysaccharides using alcian blue, periodic acid and Schiff's reagent. Representative images were taken for each treatment group: young Het mice (A); old Het mice (B); young Δ*dblGATA-1*$^{-/-}$ mice (C) and old Δ*dblGATA-1*$^{-/-}$ mice (D). Goblet cells were measured (E). Data is shown as mean +/- standard error of the mean (SEM). n = 3-4. Scale bar = 50μm.



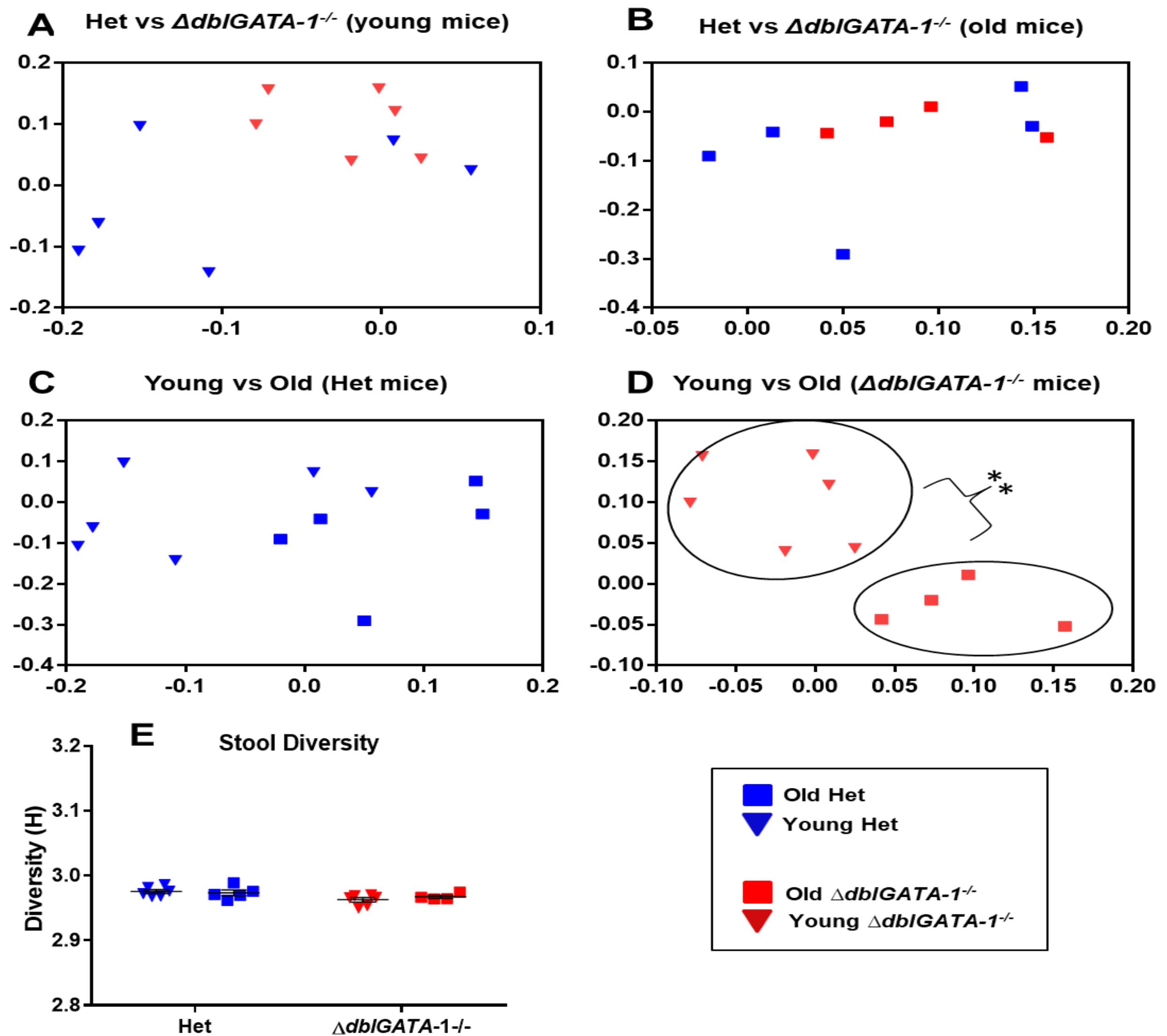
Supplementary Figure 5: Trend towards increased small intestinal goblet cells in Δ*dblGATA-1*$^{-/-}$ mice. Small intestinal tissue were obtained, from 12 week (young) and 12 month (old) old, C57BL/6 background heterozygous (Het) and eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) mice. Sections were stained for mucopolysaccharides using alcian blue, periodic acid and Schiff's reagent. Representative images were taken for each treatment group: young Het mice (A); old Het mice (B); young Δ*dblGATA-1*$^{-/-}$ mice (C) and old Δ*dblGATA-1*$^{-/-}$ mice (D). Goblet cells were measured (E). Data is shown as mean +/- standard error of the mean (SEM). n = 4. Scale bar = 50μm.



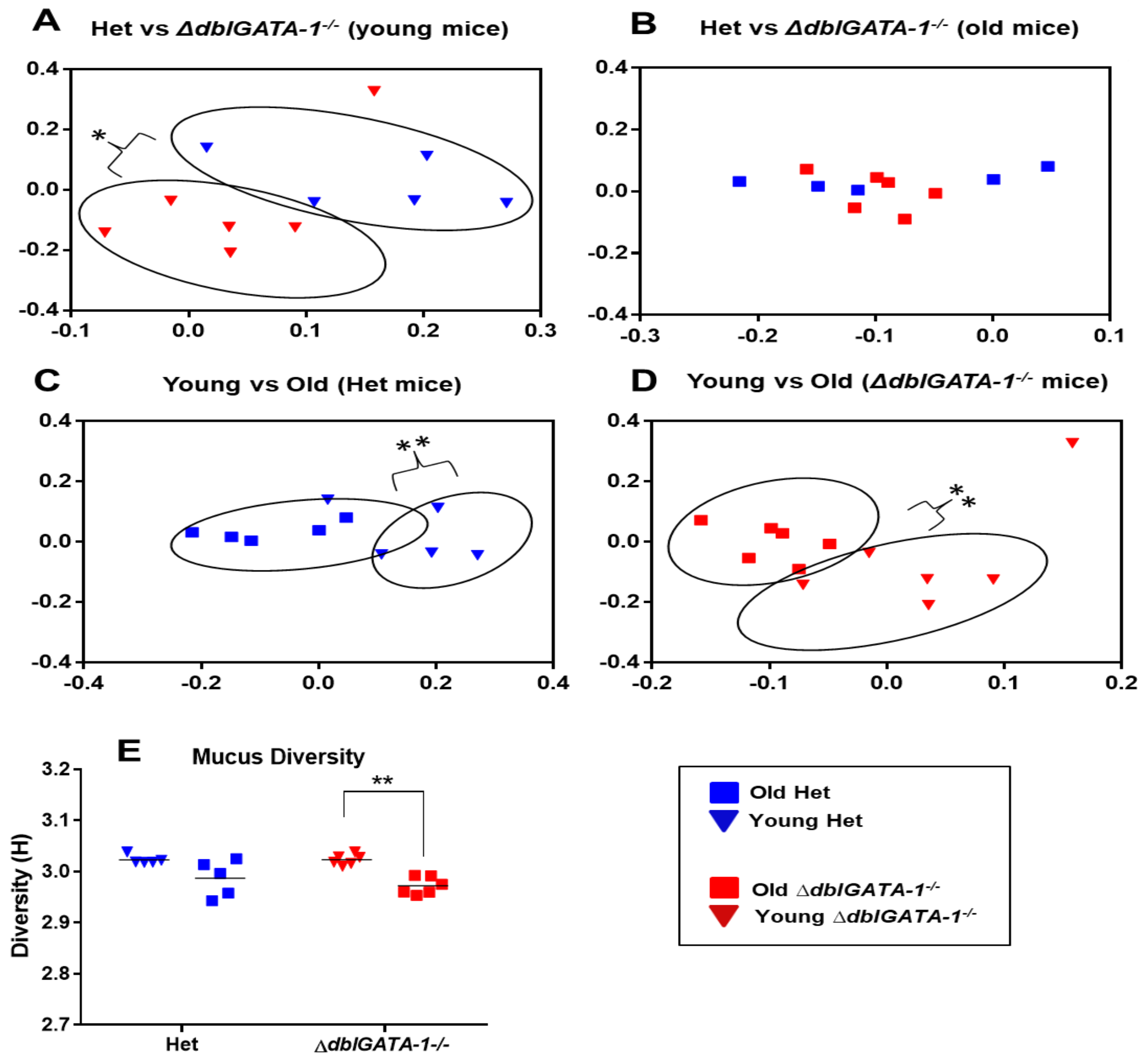
Supplementary Figure 6: Inner mucus layer characterisation. Colonic tissue sections from 12 week (young) and 12 month (old) old, C57BL/6 background heterozygous (Het) and eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) mice were stained with a fluorescent DNA probe specific for the 16S rRNA gene to identify bacteria (red), Muc2 antibody (green) to identify mucus and counterstained with DAPI (blue). Representative images were taken for each treatment group: young Het mice (A); old Het mice (B), young Δ*dblGATA-1*$^{-/-}$ mice (C) and Δ*dblGATA-1*$^{-/-}$ mice (D). Inner mucus thickness (E) and bacterial localisation (F) were measured. Bacteria were scored based on their location within the gut: 0 = bacteria in the lumen and outer mucus layer, 1 = bacteria in the inner mucus layer, 2 = bacteria in contact with the epithelium, 3 = bacteria in the crypts, 4 = bacteria in the lamina propria. Data is shown as mean +/- standard error of the mean (SEM). Asterisks represent significant as determined by 2 Way Analysis of Variance with a Tukey's Post Hoc Test (p* = <0.05). n = 3-4. Scale bar = 50µm.



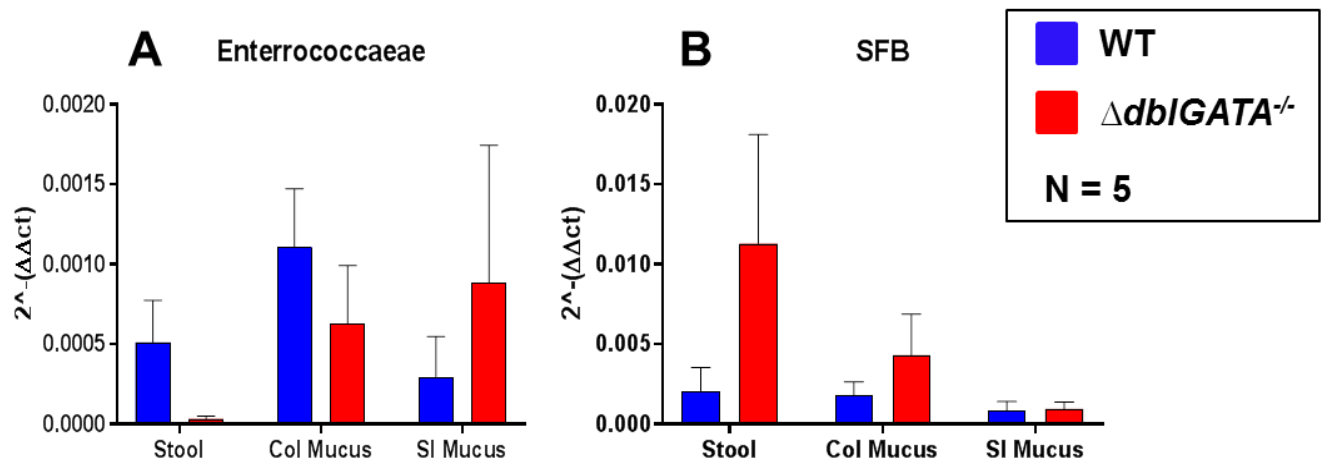
Supplementary Figure 7: Loss of eosinophils leads to increased serum IgA in younger mice. Serum was analysed via ELISA to determine levels of IgA in 12 week old and 1 year old C57BL/6 background heterozygous (Het) and Δ*dblGATA-1*$^{-/-}$ mice. Data shown as mean +/- standard error of the mean (SEM). Asterisks represent significant as determined by 2 Way Analysis of Variance with a Tukey's Post Hoc Test (p* = <0.05). n = 4-6.



Supplementary Figure 8: Differences in the bacterial communities and diversity in the stool of Heterozygous (Het) and eosinophil deficient (Δ*dblGATA-1*$^{-/-}$) mice. Differences in bacterial species composition and diversity between the stools of 12 week (young) and 12 month (old) old, C57BL/6 background heterozygous (Het) and Δ*dblGATA-1*$^{-/-}$ mice, were analysed by denaturing gel-gradient electrophoresis. Differences in the bacterial communities were plotted using non-metric multidimensional scaling (NMDS) for Young Het mice versus young Δ*dblGATA-1*$^{-/-}$ mice (A); Old Het mice versus old Δ*dblGATA-1*$^{-/-}$ mice; Young Het mice versus old Het mice (C); young Δ*dblGATA-1*$^{-/-}$ mice versus old Δ*dblGATA-1*$^{-/-}$ mice (D). Rings indicate significant differences between the bacterial communities of the respective treatment groups, as determined by permutational multivariate analysis of variance (p** = <0.01). Subsequent diversity analysis was then performed on stools (E). Data is shown as mean +/- standard error of the mean (SEM). n = 4-6.



Supplementary Figure 9: Differences in the bacterial communities and diversity in the colonic mucus of Heterozygous (Het) and eosinophil deficient (Δ*dblGATA-1*$^{-/-}$) mice. Differences in bacterial species composition and diversity between the colonic mucus of 12 week (young) and 12 month (old), C57BL/6 background heterozygous (Het) and Δ*dblGATA-1*$^{-/-}$ mice, were analysed by denaturing gel-gradient electrophoresis. Differences in the bacterial communities were plotted using non-metric multidimensional scaling (NMDS) for Young Het mice versus young Δ*dblGATA-1*$^{-/-}$ mice (A); Old Het mice versus old Δ*dblGATA-1*$^{-/-}$ mice; Young Het mice versus old Het mice (C); young Δ*dblGATA-1*$^{-/-}$ mice versus old Δ*dblGATA-1*$^{-/-}$ mice (D). Rings indicate significant differences between the bacterial communities of the respective treatment groups, as determined by permutational multivariate analysis of variance (p** = <0.01). Subsequent diversity analysis was then performed on mucus (E). Data is shown as mean +/- standard error of the mean (SEM). n = 4-6.



Supplementary Figure 10: Expression of Enterrococcaceae and SFB in WT and *ΔdblGATA-1*$^{-/-}$ mice. Real-time quantitative PCR was used to assess expression of gut bacteria in 12 week old male WT and *ΔdblGATA-1*$^{-/-}$ mice. DNA was extracted from stool, colonic mucus and small intestinal (SI) mucus and used as a template for qPCR, using a 16S rRNA gene as a housekeeping control. The relative expression of Enterrococcaeae (A) and Segmented Filamentous Bacteria (SFB) are illustrated. Data shown as mean +/- standard error of the mean (SEM).

Chapter Five

Summary Discussion

5. Summary Discussion

5.1 Summary of Findings

This thesis has overall aimed to explore the gut microbiota, with a particular focus on the mucus microbiota and its functional contributions to the host. In the introductory chapter, I hypothesised that host-health is associated with characteristic changes in gut bacterial communities, with the mucus-resident bacteria having the most profound impact on the host. I have investigated the gut microbiota in a wide array of contexts, such as in mice of different ages, genotypes and social groups, as well as the microbial niche. In every case, I found that the microbiota can be associated with these different conditions of interest. These associations were most evident in **Chapter 2**.

Chapter 2 involved the development of a method to characterise microbiota associated with different treatment groups: age (6 and 18 week old mice), genotype (WT and *mdr1a*$^{-/-}$ mice- mice that spontaneously develop colitis over time), microbial niche (stool vs mucus) and social group (cage). We employed traditional methods, beta diversity and NMDS, and then a machine-learning approach (random forest- RF) in order to distinguish the microbiome in different conditions. The traditional method of analysis revealed clear separation of the microbiome by microbial niche, age and cage, but not by genotype. When an RF model was employed, we found that the microbiome could be distinguished by treatment group to an even further extent than the traditional method, although, genotype could still not be distinguished. However, this method allowed the elucidation of characteristics that were important to find microbial associations with treatment groups. I hypothesised that the most abundant taxa of any phylogenetic scale would be important for discriminating between treatment groups. Taxa of intermediate phylogenetic scales were important for distinguishing niche, age and cage. However, taxa on the lowest end of the taxonomic scale (sub-specific groupings) were also important to distinguish cage. Contrary to my hypothesis, it was bacteria of moderate abundance that were most important for distinguishing niche, age and cage, although highly abundant taxa were also important for cage. This work highlights the intricacies of how the microbiota can differ between treatment groups that would otherwise be undetectable through traditional methods alone. Notably, I also confirm that cage has a significant impact on the gut microbiota and emphasises the need for littermate controls, to avoid confounding effects of different environments with treatment effects. After establishing methodology to characterise the microbiome, I then focused on more functional analysis.

In **Chapter 3**, I investigated the metabolite profile of the colonic mucus. The mucus layer within the colon is host to a distinct population of bacteria and it was previously demonstrated that *mdr1a*$^{-/-}$ mice have an altered mucus-resident microbiota that precedes the onset of colitis. Given that the microbiota makes substantial metabolomic contributions, I hypothesised that these mice would have an altered metabolite profile in the mucus when compared to healthy WT mice. I found that there were no overall differences in the mucus metabolite profile between WT and *mdr1a*$^{-/-}$ mice. However, there was variability in the relative concentration of individual metabolites, although the identities assigned to these metabolites were putative. In concordance with these putative, individually altered metabolites, I found significant differences in the transcriptional profile between genotype, particularly with regards to genes associated with intestinal permeability and angiogenesis. These transcriptional and metabolomic changes, in concert with an alteration to the mucus microbiota, may predispose *mdr1a*$^{-/-}$ mice to the development of colitis. Overall, this work suggests that the mucus could be a key source of host and bacterial-derived metabolites that could impact on host health, although the methodology for metabolite extraction and analysis requires further development. Given my focus on the mucus, I then investigated the microbiota in a mouse model of disrupted mucosal immunity.

In **Chapter 4**, I investigated the gut microbiota within eosinophil-deficient (Δ*dblGATA-1*$^{-/-}$) mice. My overall hypothesis stipulated that host-health was associated with changes in different gut bacterial populations, with mucus-resident bacteria having the most profound impact on the host. Eosinophils have been shown to play a role in gut barrier function and the maintenance of IgA-secreting plasma cells. Thus, I hypothesised that there were would be a significantly altered gut microbiota in mice lacking eosinophils, with a notable impact on the mucus-resident bacteria. The most substantial changes in the gut microbiota are within the mucus of Δ*dblGATA-1*$^{-/-}$ mice when compared to WT mice. However, these changes did not lead to an obvious functional impact on host-health, with no evidence of inflammation in Δ*dblGATA-1*$^{-/-}$ mice. When quantifying the microbial differences selecting key gut bacterial taxa or families via qPCR, I could no longer detect a genotype effect, but instead highlighted how taxa can vary between stool, colonic and small intestinal mucus. These data support findings in **Chapter 2** and to an extent **Chapter 3**, that a comprehensive study into the gut microbiome should consider microbial niche, as a focus on stool is insufficient to capture the diversity of the gut microbiota.

5.2 What is the impact of an altered mucus-microbiome on host-health?

Much of this thesis has addressed the mucus in various contexts. I developed an RF model to investigate the stool and mucus microbiota and found significant associations between the microbiota and microbial niche i.e. stool vs mucus, with over 90% accuracy. The RF model therefore confirms the distinct microbial communities within the stool and mucus (**Chapter 2**). Additionally, further work using a qPCR panel of bacteria highlighted that there were significant differences between different microbial niches; stool, colonic mucus and small intestinal mucus (**Chapter 4**). For instance, the small intestinal mucus is dominated by the phylum Actinobacteria, whereas Bacteroidetes and Firmicutes dominate in the stool and colonic mucus. Given that this work builds on previous findings showing that the microbiome varies by niche (Zoetendal et al., 2002; Ahmed et al., 2007; **Section 1.3**), an important question to ask would be how an altered mucus-microbiome impacts upon host health.

Several studies of disease have shown that there can be niche-specific changes, where the mucus but not the stool microbiota is altered. For example, the colonic mucus microbiota was significantly different in cirrhosis patients, whereas the stool microbiota was unaffected (Bajaj et al., 2012), suggesting that the mucus-resident bacteria could potentially mediate more systemic effects beyond the gut. In human patients, the mucus-associated bacteria *F. prausnitzii* is thought to be decreased in the context of IBD, specifically Crohn's Disease (CD) (Frank et al., 2007). Additionally, we established that the mucus-resident bacteria changed before the onset of colitis in *mdr1a*$^{-/-}$ mice (Glymenaki et al., 2017b). However, how these changes in the mucus-resident bacteria may functionally impact on the host have not fully been elucidated.

In **Chapter 4**, we reported that Δ*dblGATA-1*$^{-/-}$ mice had a significantly different microbiota, with reduced diversity in mucus-resident bacteria that was not seen in the stool. Lower microbial diversity in the gut is typically associated with negative consequences, in diseases such as IBD or infirm ageing (Manichanh et al., 2006; Pascal et al., 2017). Despite this impact on the mucus diversity, there was no evidence of inflammation in the Δ*dblGATA-1*$^{-/-}$ mice. However, there could be an impact on the physical properties of the mucus, for example glycosylation and viscosity, which was not investigated in my study. This in turn, could impact on mucus permeability and therefore impaired barrier function, driving inflammation in the host. A thinner mucus barrier is associated with inflammation in the gut (Desai et al., 2016). If the mucus barrier were less viscous in the Δ*dblGATA-1*$^{-/-}$ mouse, it could make the mouse more vulnerable to pathogen challenge or barrier disruption. However,

various studies do not detect defects in the ability of eosinophil-deficient mice to combat parasite infection (Swartz et al., 2006; O'Connell et al., 2011; Fabre et al., 2009). Regardless, decreased viscosity could lead to increased permeability that could be specifically assessed by the use of FITC-dextran labelled conjugates (Johansson et al., 2010). Although I saw no evidence of increased bacterial localisation or changes to mucus thickness, in the event it is less viscous, small molecules such as metabolites could diffuse more readily which could have impacts on inflammation and barrier function. Indeed, several immune cell types are known to respond to metabolites, such as B cells, where SCFAs can facilitate the production of healthy antibody responses (Kim et al., 2016). Metabolite localisation could be investigated via RAMAN-spectroscopy (Carter et al., 2009). Understanding this localisation could give a key as to how metabolites are interacting with the host cells and thus mediating their function.

Mucus degradation can also be mediated by the microbiota. When the mucus-microbiota population is altered, there can be a detrimental impact on barrier function, leading to inflammation. When mice were fed a high fat (low fibre) diet, it lead to an increase in mucin-degrading bacteria *A. muciniphila* and *Bacteroides caccae* that resulted in a thinner mucus barrier and consequential inflammation for the host (Desai et al., 2016). In **Chapter 2**, the RF model identified the sulphate-reducing genera *Desulfovibrio*, as important to discriminate between the ages of mice. Curiously, this genus was found almost exclusively in 18 week old mice, the time point in which *mdr1a*$^{-/-}$ mice begin to develop colitis. *Desulfovibrio* bacteria are typically involved in sulphur reduction which produces hydrogen sulphide. Hydrogen sulphide has been shown to facilitate the degradation of the mucus barrier and impact on host health (Carbonero et al., 2012; Ijssennagger et al., 2016). Sulphate reducing bacteria and increased levels of hydrogen sulphide have also been associated with IBD (Loubinoux et al., 2002; Khalil et al., 2014). This could be one example of how a change in the mucus-resident microbiota could predispose mice to developing colitis and thus impact on host-health.

One hypothesis was that the change in mucus-resident microbiota in *mdr1a*$^{-/-}$ mice would lead to an altered metabolomic profile within the mucus (**Chapter 3**). My study expands upon previous stool, serum and urine-focused work by investigating a niche containing potentially novel metabolites. One study did investigate metabolites within the mucus of human patients, although it was by taking whole caecal or sigmoid colon explants (McHardy et al., 2013). They found that genera comprising *Faecalibacteria, Phascolarcobacteria* and *Roseburia* were associated with genes linked to fatty acid synthesis. Additionally, the PICRUSt tool was

used to make functional inferences based on the stool and mucus microbiota identified in WT and *mdr1a*$^{-/-}$ mice with regards to metabolites (Glymenaki et al., 2017a), and the primary metabolic functions identified involved fatty acid synthesis. However, in my study, we did not see overall metabolomic changes and instead saw variability in individual metabolites between WT and *mdr1a*$^{-/-}$ mice. One limitation of the method used is that the identities assigned were putative. Therefore, it is difficult to make functional inferences based on my current analysis. The urinary metabolite profile can separate out *mdr1a*$^{-/-}$ mice by genotype (Glymenaki et al., 2017a), so the mucus may not give the best representation of the overall metabolomic profile, being likely to be restricted to local host or microbial constituents. Urinary metabolite profiling is a routine procedure, whereas we are still developing and optimising the methodology for working with mucus samples. As a result, my current data may only be partially reflective of the mucus metabolomic profile. The methodology could be improved through further optimisation. As SCFAs are thought to be particularly important for the gut barrier, it would be interesting to do a more focused analysis. Specifically, this would require the running of known SCFAs such as butyrate to act as standards, through the LC-MS pipeline in order to confirm the m/z ratio and retention times of the SCFAs on the in-house machinery. The metabolite features arising from these standards could then be compared against my samples.

An alternative explanation for the limited changes in metabolites, is that these changes may only become apparent in the context of progressed inflammation. Indeed, even when urinary metabolites were investigated in *mdr1a*$^{-/-}$ mice, the metabolite profile remained relatively unchanged, even at the onset of inflammation (Glymenaki et al., 2017a). Studies of human IBD, DSS-induced colitis or those using *IL-10*$^{-/-}$ mice that developed severe inflammation and colitis showed differences in the metabolite profiles when comparing inflammation to healthy controls (Murdoch et al., 2008; Le Gall et al., 2011; Schicho et al., 2010; Schicho et al., 2012). Other than differences in the sampling approach to analyse metabolites, a key difference between my study and these models is that my mice did not develop colitis at the 6 week time point chosen (i.e. the age of the *mdr1a*$^{-/-}$mice). This time point was chosen because the mucus-microbiota are altered, whereas the stool microbiota remains unaffected and pathology is normal, making it a good time point to study initiating events in the inflammation pathway (Glymenaki et al., 2017b). Thus, there is potential to explore older *mdr1a*$^{-/-}$ mice where their colitis has progressed, however this would not reveal whether metabolite changes were causative or as a consequence of inflammation. As my work in

Chapter 3 was a pilot study, I focused solely on the mucus. However, future work could investigate stool, urine and serum in order gain a broader understanding of how metabolites change in the context of the development of severe inflammation. One limitation with the *mdr1a*$^{-/-}$ mouse model is that colitis development can be unreliable, however, unlike methods such as DSS-treatment, the mucus remains intact and facilitates our investigation of the mucus. As a compromise, alternative mouse models could be investigated, such as NEMO-deficient mice (Nenci et al., 2007) that are known to more reliably develop severe chronic intestinal inflammation, but the deficiency itself does not directly impact on the mucus.

I have chosen to investigate metabolites within the mucus because the microbiota are responsible for a large portion of dietary-derived metabolites. However, the mucus also presents an opportunity to look for novel, non-metabolite biomarkers of inflammation. Current studies have primarily focused on stool and serum. For example, the calcium-binding protein, S100A12, has been identified as a biomarker of IBD in human serum and stool (de Jong et al., 2006; Manolakis et al., 2010). Levels of faecal calprotectin, a protein in neutrophils and macrophages, are also commonly used as a biomarker for IBD (Smith and Gaya, 2012). There are a wealth of other IBD biomarkers that have been identified in patients as reviewed by Norouzinia et al. (2017). Some of the issues of these biomarkers can be specificity, for example can the marker discriminate between IBD and irritable bowel syndrome, and sensitivity, whether the marker has a high rate of false positives/negatives. Perhaps biomarkers in the mucus could be more robust. For example, increased levels of hydrogen sulphide can lead to an increase in mucus-degradation, so the levels of hydrogen sulphide within the mucus could be investigated to see whether this correlates with inflammation. Other candidates to explore in the mucus could be the levels of antimicrobial peptides (AMPs). Faecal beta-defensin has been used a biomarker for IBD in children (Kapel et al., 2009) and so if there is a significant change/increase in the mucus-resident bacteria, there could be increased levels of AMPs in the mucus. The AMP RegIIIγ was shown to impact more on the mucus-resident microbiota than the stool microbiota and could therefore be another potential target for analysis. Machine learning approaches developed in **Chapter 2** could be used to try and correlate what communities of bacteria drive levels of AMPs and could therefore help in the identification of the most influential bacteria.

Skin homeostasis can also be regulated by both the microbiota and the production of antimicrobial peptides (Williams et al., 2017; Williams et al., 2018). For example, wound repair was significantly decreased when mice were administered muramyl dipeptide (MDP),

an agonist of the NOD2 pattern recognition receptor (Williams et al., 2018). This stimulation also led to the upregulation of beta-defensins, notably 1, 3 and 14. In mice deficient in beta-defensin 14, wound healing was impaired with a delay in neutrophil recruitment, but an increase in the localisation of pro-inflammatory macrophages. These findings highlight the immunomodulatory capabilities of AMPs. Therefore, if AMPs are altered in the gut as a consequence of a perturbed mucus-resident microbiota, this highlights a potential mechanism for how the microbiota can have an indirect immunomodulatory function on the host and thus influence host health.

It is known that the microbiota is altered at the onset of IBD and it is now becoming more apparent that faecal samples alone are insufficient to capture these microbial alterations (Gevers et al., 2014; Glymenaki et al., 2017b; Zhang et al., 2018). Notably, one recent study discussed that probiotics are recommended for patients that have received a course of antibiotics in order to restore the microbiota (Suez et al., 2018). However, they found that probiotics impeded the reconstitution of the mucus microbiome. This could therefore impact on patient treatment, where patients may be more susceptible to opportunistic pathogen infection until their mucus microbiome is restored. However, there is one major limitation to investigating the mucus. Stool is commonly used because it is practical and easily accessible via routine procedures. Even if mucus is a valid source of biomarkers and metabolites, extracting mucus from human patients is not routinely practical and typically even for biopsies, patients will have had a bowel prep treatment that could damage their mucus lining (Hollister et al., 2014). However, a novel method has been developed by Origin Sciences (Origin Sciences, 2018) that extracts rectal mucus with limited invasiveness and can potentially facilitate the investigation of the mucus in human patients.

Overall, changes in the mucus-specific microbiota could potentially mediate a multitude of effects for gut health and susceptibility to inflammation. Another important area to consider is how cells interact to mediate barrier protection. This thesis focused on eosinophils and their potential role in gut homeostasis will be discussed next.

5.3 How important are eosinophils for regulation of gut homeostasis?

There has been considerable debate as to how or whether eosinophils contribute to gut homeostasis and they are best understood currently for the roles in Th2-associated inflammation (Lee et al., 2004; Rothenberg, 2004; Jacobsen et al., 2014). However, eosinophil recruitment is also a hallmark of Th1 inflammatory conditions suggesting their

role is not restricted to Th2 immunity and anti-helminth or allergic responses (Spencer et al., 2009). Eosinophils are armed with a plethora of preformed mediators as well as being able to make cytokines and mediators *de novo* and it is now known they can mediate both inflammatory and anti-inflammatory effects. Indeed some studies to suggest that eosinophils may be beneficial, in terms of glucose tolerance, intestinal permeability and the regulation of IgA-secreting plasma cells (Wu et al., 2011; Chu et al., 2014; Johnson et al., 2015; Masterson et al., 2015). Thus, although eosinophilia is a trademark of several inflammatory conditions, eosinophils are resident cells in a healthy gut. It is therefore important to consider how eosinophils contribute to the regulation of gut homeostasis.

Gut barrier function is a crucial aspect of gut homeostasis (**Section 1.2**). Eosinophils have been shown to impact on barrier function negatively. The eosinophil-derived granule protein eosinophil peroxidase (EPX) downregulated expression of the epithelial TJ protein occludin *in vitro*, which is important for epithelial barrier integrity (Furuta et al., 2005). It is therefore perhaps curious that Johnson et al. (2015) show that mice fed a high fat diet had increased intestinal permeability, that was associated with a depletion in eosinophil numbers. However these disparate data may be due to differences in the function of the eosinophils. In the study by Furuta et al. (2005), eosinophils are known to respond to environmental cues that will dictate their function. For example, it was recently shown that eosinophils recruited into the gut activate a specific transcriptional profile in response to live bacteria that suppresses inflammation (Arnold et al., 2018). Without the environmental cues found *in vivo*, perhaps the *in vitro* eosinophils are more pathogenic. Additionally, the study by Johnson et al. (2015) would suggest that eosinophils may play a role in the maintenance of permeability. However, this was a correlation and they did not specifically delineate the mechanism behind the increased permeability. Intestinal permeability is maintained in part by tight junction proteins. Other cell types, such as γΔ T cells are known to regulate tight junction proteins (Dalton et al., 2006) and so it is conceivable that eosinophils could also be involved. Perhaps eosinophils produce factors that maintain barrier integrity as well as deplete it. As suggested above, eosinophils respond to environmental cues, so in one context eosinophils could be protective (such as obesity) and in another they could be harmful (allergy). Therefore, expression and localisation of TJ proteins could be investigated and such experiments could be applied to my own work in Δ*dblGATA-1*$^{-/-}$ mice. Specifically, TJ protein expression and distribution could be analysed by immunohistochemistry *in situ.* Eosinophils and intestinal epithelial cells could also be co-cultured to investigate specifically how eosinophils can

impact on barrier integrity and TJ protein expression, either through directly quantifying TJ proteins or through measuring transepithelial electrical resistance (TEER) (Srinivasan et al., 2015). FITC-dextran absorption *in vivo* or assessment of faecal albumin could give measures of intestinal permeability. If these experiments are combined with eosinophils in different contexts, such as health, inflammation and obesity, these data would further our understanding as to how eosinophils specifically interact with host epithelial cells and influence barrier integrity.

The mucus layer of the gut is also an important part of the gut barrier. Eosinophils have been shown to influence mucus thickness (Chu et al., 2014). However, in our hands we saw no changes in goblet cell numbers or mucus thickness. A key difference between our study and the previous work is the use of littermate controls. Therefore, it is possible that the differences in environment could have mediated the differences in mucus thickness between studies. Even within the same animal house, the mucus thickness is dependent upon the microbiota and affected in mice housed in different rooms (Jakobsson et al., 2015). These data highlight the need to control for environment.

The microbiota both contributes to and is regulated by gut homeostasis. In our own hands, we saw a significant difference in the microbiota in the stools and mucus of Δ*dblGATA-1*$^{-/-}$ mice via DGGE. The data I showed suggested that loss of eosinophils led to a significant alteration to the gut microbiota and this could have functional consequences for the host, such as driving inflammation. However, I saw no evidence of inflammation in the Δ*dblGATA-1*$^{-/-}$ mice, which could suggest that the microbiota and the lack of eosinophils does not impact the host in the steady state. Curiously, when the microbiota were quantified via qPCR, I found no genotypic differences in stool, colonic or small intestinal mucus. This could simply be a limitation of my qPCR panel, which although screened for the most common types of gut bacteria, may have missed where differences in the bacterial communities lay. Therefore, it would be of interest to perform high throughput sequencing in order to fully characterise the microbiome in this mouse model. This sequencing would be necessary to quantify specifically how eosinophils impact on the gut microbiome.

All the data in my thesis was in naive mice and it may be that eosinophils may play different roles in a perturbed state, such as under the context of a high fat diet or parasite infection. One study infected WT and Δ*dblGATA-1*$^{-/-}$ littermates with *Trichuris muris* to drive inflammation in the colon and *Toxoplasma gondii*, to drive inflammation in the small

intestine (Forman et al., 2016). The lack of eosinophils had no impact on inflammation or parasite burden. However, there were differences in the ability of eosinophils to regulate IgA plasma cells in the small intestine, with a significant decrease in small intestinal IgA+ plasma cells post-infection and in the large intestine, with a significant increase in large intestinal IgA+ plasma cells post-infection. These data would suggest that eosinophils can play functionally different roles depending on their location or dependent on inflammation stimuli. For example, eosinophils in the adipose tissue can polarise macrophages to become anti-inflammatory and this function was required to maintain glucose tolerance (Wu et al., 2011). The significance of eosinophils having different phenotypes based on their location, could be one reason for the discrepancy between eosinophils positively and negatively contributing to host health in different studies. For example, there may be different signals from the host or even the microbiota that influence eosinophil function in the different anatomical regions The suggestion to co-culture eosinophils with intestinal epithelial cells could therefore be combined with eosinophils isolated from different regions of the host (for example colon and small intestine) and would help to illustrate the phenotypic differences that may arise. For example colonic eosinophils may positively contribute to barrier integrity, whereas small intestinal eosinophils may impact negatively. The factors that eosinophils produce in these cultures could be analysed via ELISA and the cellular RNA could be extracted for qPCR analysis.

Eosinophils were previously thought to play a role in the regulation of IgA-secreting plasma cells under naïve conditions, where eosinophil-deficient mice had reduced levels of IgA (Chu et al., 2014). However, this study did not use littermate controls and the WT control mice were bought in, so it cannot be ruled out that the microbiota was responsible for an induction of greater IgA levels in the control mice. Indeed, subsequent littermate-controlled studies could not confirm an impact on IgA in naïve eosinophil-deficient mice (Forman et al., 2016; Haberland et al., 2018). In my study, I could not see differences in the numbers of IgA plasma cells between WT and Δ*dblGATA-1*$^{-/-}$ mice. My work showed that contrary to the study by Chu et al. (2014), which suggested a reduction in IgA levels in naive mice, I saw no difference in naive young mice. However, there were slightly higher levels of serum IgA in older Δ*dblGATA-1*$^{-/-}$ mice when compared to WT littermates. Although this is not the same as IgA plasma cell numbers, it may potentially suggest that eosinophils play more important roles under perturbed conditions, in this case ageing. The consequences of altered IgA could be numerous. For example, one study showed that high-affinity IgA bound to bacteria could

select for colitis-inducing species (Palm et al., 2014). Therefore, if eosinophils do indeed regulate IgA, it could perhaps predispose mice to the development of colitis by altering the amount of IgA available to bind colitis-inducing microbes or enhance susceptibility to infection by reduced ability to bind pathogenic microbes. The methodology developed by Palm et al. (2014) could be applied to my own work, where IgA-bound bacteria can be isolated by flow cytometry and sequenced. The resulting data would help to illustrate whether eosinophils can impact on the microbiota, IgA-binding and also whether this functionally matters for the host.

Eosinophils have been implicated in a wide array of gastrointestinal conditions such as IBD, reviewed by Rothenberg (2004). In a model of experimental colitis induced by the chemical DSS, there was prominent colonic eosinophilia and high levels of the eosinophil granule, EPX (Forbes et al., 2004). They also revealed that eotaxin was a key driver of eosinophil localisation to the gut after DSS administration. In human patients, eosinophil-associated cytokines were shown to be linked to both CD and ulcerative colitis (UC) patients (Neubauer et al., 2018). Therefore, it is possible that eosinophils are key drivers of inflammation in these contexts and could in fact play a detrimental role in gut homeostasis. To test this idea, eosinophil-deficient mice could be treated with DSS to experimentally induce colitis and such experiments have been carried out before. One study showed that Δ*dblGATA-1*$^{-/-}$ mice developed greater clinical pathology and weight loss when compared to WT mice upon treatment with DSS (Vieira et al., 2009). This is in contrast to an alternative study, where PHIL mice, showed that wildtype mice developed significantly worse colitis than those that were eosinophil-deficient (Masterson et al., 2015). The PHIL mice study used littermate-controls whereas there Δ*dblGATA-1*$^{-/-}$ mice used bought in WT controls. Although there are obvious differences in the genetic mechanism behind eosinophil-depletion in these models, it is possible that sharing of the gut microbiota between WT and PHIL mice led to a more benign microbial profile. Conversely, single housing of Δ*dblGATA-1*$^{-/-}$ mice may have fostered a more pathogenic microbiota. This hypothesis could be tested by comparing the effects of DSS-treatment on co-housed WT and Δ*dblGATA-1*$^{-/-}$ mice and those separated by genotype.

In summary, eosinophils have often been associated with inflammation. However, they have also been associated with more positive roles, such as the maintenance of glucose tolerance and intestinal permeability. Under naïve conditions, the role that eosinophils play in gut homeostasis may not be apparent from my own experiments. However further experiments

involving perturbations to the host could allow the elucidation of how eosinophils contribute to gut homeostasis. It is also known that the microbiota regulates aspects of homeostasis. How the microbiota can drive inflammation will now be discussed.

5.4 How could changes in the microbiota drive inflammation and colitis?

In **Section 5.2**, I focused specifically on how changes in the mucus microbiota could impact on host health. I will now discuss how general changes in the microbiota could drive inflammation and colitis, another focus of my thesis. A change in the microbiota is likely to coincide with a change in the metabolite profile and the metabolites could drive inflammation. Various studies have investigated metabolites in the context of IBD (Garner et al., 2007; Machiels et al., 2014; De Preter et al., 2015). One study investigated the faecal metabolite profile in CD and UC patients (De Preter et al., 2015) and showed that CD patients had significantly increased benzene, phenol and carbon disulphide, compared to healthy patients. The increase in carbon disulphide is particularly intriguing, given that sulphide products are thought to mediate degradation of the mucus barrier as discussed in **Section 5.2**. UC patients had significantly increased cyclohexane, 3-methyl butanal and pyrrole that were not present in CD or healthy patients. The aforementioned studies focused on faecal samples to give a more accurate representation of the microbiota-associated metabolite profile, compared to serum and urine alone. The SCFA butyrate is thought to provide anti-inflammatory effects (Furusawa et al., 2013; Trompette et al., 2014) and decreased faecal butyrate is associated with IBD (Machiels et al., 2014). Additionally, butyrate has been shown to regulate barrier integrity *in vitro* in Caco-2 cells (a human gut epithelial cell line), where the cells were cultured with supernatants from a batch culture of CD patient-derived faecal microbiota (Geirnaert et al., 2017). Specifically, one portion of the culture supernatant was mixed with a cocktail of butyrate-producing species, whereas the other was not. The cells had decreased barrier integrity when treated with the plain portion of supernatant, but showed higher barrier integrity when cultured with the mixed supernatant, as determined by TEER measurements.

One mechanism for butyrate's ability to promote barrier integrity is that butyrate stimulates mucin production (Canani et al., 2011), that may contribute to increased mucus thickness. It is thus interesting to note that we have previously shown the mucus barrier to be thinner in *mdr1a*$^{-/-}$ mice (Glymenaki et al., 2017b). Although we did not specifically identify SCFAs in our metabolite analysis, many of our metabolites could not be identified when compared to various databases. Given that there are differences in retention time and peak intensity

between different chromatography runs and different machines, it is still possible there was an alteration to SCFA metabolism that we were unable to detect (Dunn et al., 2011). Therefore, as stated in **Section 5.2**, metabolites of interest are required to act as standards to run through the LC-MS pipeline, before thorough metabolomic interpretation can be made.

Beyond metabolites, we identified transcriptional differences in the colon of *mdr1a*$^{-/-}$ mice at 6 weeks of age. Given that microbial dysbiosis is linked with IBD development, the change in the microbiota could have an impact on host-transcription. A previous study in young (4-5 week old) non-inflamed *mdr1a*$^{-/-}$ mice performed a microarray examining intestinal transcription, although littermate controls were not used (Collett et al., 2008). This study identified changes in genes associated with antigen recognition, which were not significant in our hands. Notably, we identified changes in genes associated with intestinal permeability and angiogenesis. This discrepancy could be driven by the microbiota, as different animal facilities can have a functionally different microbiota (Jakobsson et al., 2015).

In IBD, given that there is continuous damage to the gut structure, repair mechanisms are needed to alleviate the tissue damage caused. Angiogenesis is the process of developing new blood vessels and plays a role in a variety of contexts such as wound repair, cancer and inflammation (Dvorak et al., 1995; Scaldaferri et al., 2009). Thus, angiogenesis is an important process in IBD. Factors that promote angiogenesis are those associated with the VEGF-family. As discussed in **Chapter 3**, there are various members of the VEGF-family. A previous IBD study has associated one factor, VEGF-A with gut inflammation in humans and mice (Scaldaferri et al., 2009). Specifically, they showed that blockade of VEGF-A led to a decrease in colitis severity in mice. The microbiota has been shown to play a role in the regulation of angiogenesis, where a high fat diet led to gut microbial dysbiosis that drove low-grade chronic inflammation associated with cytokines such as IL-1β, TNF-α and VEGF-A (Andriessen et al., 2016). The combination of these factors contributed to age-related macular degeneration in the eye. The microbiota may also drive intestinal angiogenesis, where a study used human intestinal microvascular cells and treated them with bacterial ligands (Schirbel et al., 2013). They showed that TLR and NOD receptor stimulation by bacterial ligands drove angiogenesis. This could potentially happen *in vivo* and therefore, there is potential to investigate the levels of VEGF-associated factors, either by qPCR or by immunohistochemistry in the *mdr1a*$^{-/-}$ mice, in order to validate the RNA-seq in my study. The machine learning method developed in **Chapter 2** could be used to associate certain communities of bacteria with the levels of VEGF family members. Additionally, the

intestinal epithelial cells could be co-cultured with specific bacteria of interest or metabolites to investigate whether bacteria (or their products) could drive expression of angiogenic factors. These experiments have the potential to highlight whether the microbiota can influence angiogenesis and therefore provide a relatively novel mechanism for the microbiota driving inflammation.

Although not VEGF family members, we noted differences in angiopoietins (Angpts) that also mediate angiogenesis. In our hands, we saw reduced levels of Angpt1 expression in *mdr1a*$^{-/-}$ mice, whereas we saw higher expression of Angpt2. Angpt1 and Angpt2 are ligands of the Tie-2 receptor expressed on endothelial cells and their binding helps to control vascular stability and remodelling (Linares et al., 2014). Specifically Angpt1 acts as a regulator of blood vessel maturation and is involved in anti-inflammatory responses whereas Angpt2 is involved in the development of lymphatic vasculature (Linares et al., 2014). DSS-treated mice deficient in Angpt2 had significantly reduced leukocyte infiltration and inflammation when compared to WT mice (Ganta et al., 2010). Additionally, a study in human patients showed significantly higher levels of Angpt2 in patients with active IBD when compared to healthy patients (Koutroubakis et al., 2006). No study has yet linked the gut microbiota to the expression of these proteins. However, the gut microbiota has been shown to mediate impaired glucose tolerance in mice deficient in another Angpt family member, Angpt4 (Janssen et al., 2018). Therefore, it is possible that there is a link between the microbiota and other members of the Angpt family. If there is a link, it could implicate the microbiota in regulating vascular permeability and thus a mechanism for the microbiota driving inflammation.

Inflammation could also be driven by an increase in intestinal permeability, the ability of fluid and solutes to diffuse through the lumen and gut tissue. *Mdr1a*$^{-/-}$ mice already have a defect in transport of small molecules across the gut barrier, due to the loss of the p-glycoprotein transporter. However, permeability is regulated by a variety of other factors, including tight junction (TJ) proteins. We specifically found a reduction in claudin-2 expression and an increase in occludin expression in our *mdr1a*$^{-/-}$ mice. Altered TJ protein expression is a hallmark of IBD. Occludin has been reported to be significantly decreased in IBD (Gassler et al., 2001), however, occludin is just one protein in a complex of tight junction proteins that are important for the barrier. Occludin-deficient mice could still form intact tight junctions (Saitou et al., 1998) and selective knockdown of both *in vitro* and *in vivo* occludin did not affect the tight junction structure per se, although there was increased

permeability (Al-Sadi et al., 2011). Specifically, disruption of occludin enabled larger macromolecules to pass through the barrier more readily. The claudin family of TJ proteins appears to be more important for tight junction formation. Mice deficient in claudin-1 die within 24 hours of birth due to large electrolyte and fluid loss as a result of the impaired barrier function (Lee, 2015). Therefore, the increase in occludin transcription may not necessary have any physiological impact, whereas the decrease in claudin expression may lead to more dramatic physiological consequences, for example, increased bacterial translocation.

Although studies have implicated the microbiota in the degradation of the mucus-barrier, leading to increased intestinal permeability and increased inflammation (Rey et al., 2013; Desai et al., 2016), far fewer studies have directly examined the association of the microbiota and tight junction protein expression. One study indirectly investigated the role of the gut microbiota in tight junction protein expression, through studying how dietary protein impacted upon the microbiota and consequently, the gut barrier (Fan et al., 2017). The microbiota produces beneficial metabolites such as butyrate, that are known to mediate improved barrier function through improved TJ protein expression (Ma et al., 2012).

Beyond the gut barrier, the microbiota has been shown to be necessary for blood brain barrier development (Braniste et al., 2014). Specifically, they revealed that germ-free mice had reduced expression of claudin-5 and occludin, leading to altered endothelial tight junctions in the brain. Functionally, they showed that metabolites produced by bacteria could be important in developing these tight junctions. They colonised mice with either butyrate producing *Clostridium tyrobutyricum* or *Bacteroides thetaiotaomicron* that produces acetate and propionate, or supplemented germ-free mice with butyrate. In all cases, permeability of the blood brain barrier was restored to levels akin to pathogen free mice. In the gut, it has been shown that the mucus-degrading bacteria *Akkermansia* can regulate tight junctions through the production of extracellular vesicles (Chelakkot et al., 2018). Therefore, it is possible that the altered microbiota in *mdr1a*$^{-/-}$ mice could contribute to a decline in barrier integrity before the onset of colitis.

Although they are not TJ proteins, I saw a decrease in aquaporin (AQP) expression in our mice. AQP7 was highly variable in *mdr1a*$^{-/-}$ mice, but consistently low in WT mice. Conversely, AQP9 had consistently low expression in *mdr1a*$^{-/-}$ mice, but two WT mice had high expression. Aquaporins act as water channels across biological membranes, but AQP7 is

thought to also transport glycerol, while AQP9 can transport other small solutes such as urea (Ishibashi et al., 2009). Therefore, a defect in the presence of these aquaporins could potentially lead to altered epithelial cell polarity and nutritional defects from impaired fluid transport. AQP9 may play a role in driving colitis. It was shown to be expressed on leukocytes (Ishibashi et al., 1998) and therefore an alteration to AQP9 expression on these cells could lead to functional changes, for example, predisposing cells to become more inflammatory or regulatory. One study showed that AQP7 expression was reduced in the ileum and colon of patients with UC, in addition to other family members including AQP3 and 8 (Ricanek et al., 2015). Notably, although it was not identified in our experiments, knockdown of AQP3 was shown to lead to a dramatic increase in bacterial translocation *in vitro*, as well as a reduction in claudin-1 and occludin expression (Zhang et al., 2011). This data would suggest that aquaporins are able to interact with TJ proteins and can potentially mediate bacterial translocation. This could potentially influence the changes in the mucus-resident microbiota that are seen at the 6 week time point in the *mdr1a*$^{-/-}$ mice.

What is surprising is that very few studies have focused specifically on how both the stool and the mucus microbiota interact with gut epithelial cells. The interaction of microbes is thought to be crucial for epithelial turnover, where germ-free mice had impaired epithelial cell turnover that could be restored with either oral administration of chloroform-treated faeces from SPF mice, or the administration of SCFAs (Park et al., 2016). Thus, the metabolites produced by bacteria were vital to epithelial cell turnover. Given that barrier function is vital to maintain healthy gut homeostasis, this represents an important area to investigate how the microbiota can impact upon host health. There is potential to identify bacteria that are significantly altered between the healthy and inflamed state and co-culture these bacteria with intestinal epithelial cells, to see whether the microbiota can influence aspects of barrier function, such as tight junction protein expression. Germ-free mice could be colonised with the identified bacteria to investigate their impact on barrier integrity such as TJ formation. However, germ-free mice have disrupted mucosal immunity with less gut associated lymphoid tissue (Round and Mazmanian, 2009). Given that certain immune cell types such as Δγ T cells can regulate TJ's (Dalton et al., 2006), there could be confounding factors in germ-free mice that impede barrier function. However, there are numerous other study possibilities, such as sampling the microbiota and metabolites (stool and mucus), isolating the intestinal epithelial cells and performing RNA-seq for example, under different conditions (e.g. young and old, normal and high fat diet, WT and *mdr1a*$^{-/-}$ mice etc.). This

will allow holistic analysis of how changes in microbial communities and metabolites can impact specifically on epithelial cell transcription. Given that I have already developed methodology to find associations between the microbiota and different conditions of interest (**Chapter 1**), my machine learning approach can be adapted to facilitate the proposed analysis of linking changes in the microbiota (cause) to changes in host metabolites and transcription (effect).

5.5 Are current methods of microbiome analysis adequate?

The microbiome has been studied in a variety of environments such as on the skin (Grice and Segre, 2011), in the lung (Moffatt and Cookson, 2017) and in the gut. The microbiota has also been explored beyond a host, such as in the soil (Ramirez et al., 2017). Analysis of the microbiome however, is relatively consistent, irrespective of the environment that is sampled. However, there are both sequencing and sequencing-independent methods to consider.

In **Chapter 4**, I used sequencing-independent methods, i.e. DGGE and qPCR, to explore the gut microbial communities (stool and mucus) in Δ*dblGATA-1*$^{-/-}$mice. One obvious discrepancy in my work is that DGGE showed a genotypic difference between WT and Δ*dblGATA-1*$^{-/-}$mice, whereas qPCR did not. DGGE was performed using 16S rRNA sequencing primers and gives an overall microbial fingerprint, where bands could correspond to any taxonomic level and any number of taxa. qPCR is more specific and although I performed a screen for the most common gut taxa, it is not as high throughput as other methods and thus I may simply not have screened the specific groups of bacteria that are different between WT and Δ*dblGATA-1*$^{-/-}$mice. Ultimately, it is important to consider what information is required with regards to microbiome data. If a microbial-fingerprint is required, then DGGE is suitable to provide a broad and inexpensive scan of community differences. If the focus is on a specific group of bacteria, then qPCR is a viable method to quantify such changes. However, if more high-throughput quantification is required, then a form of NGS (16S or Whole Genome Sequencing- WGS) would be the preferred method. qPCR could then be a way of validating the findings from the sequencing experiments. Many microbiome studies have focused on 16S rRNA sequencing to investigate the gut microbial communities and this could perhaps be a limitation of my own work. Indeed, one of the limitations of 16S rRNA sequencing is a loss of taxonomic resolution at the species level, whereas WGS can more accurately define species (Ranjan et al., 2016).

Almost all sequence data undergoes a form of quality control before analysis and this is a necessary and standardised step, depending on the sequencing platform. It is the processing of sequences subsequent to these quality control steps that can influence a researcher's findings. The pipeline implicated by QIIME has become the gold standard for 16S rRNA analysis, often implicating a 97% cut off for sequence similarity (Caporaso et al., 2010). The purpose of this cut off is to account for differences that may arise due to sequencing error (Schloss, 2010) and is a reasonable consideration to make. QIIME analysis allows a broad and well established snapshot of the gut microbiota and can certainly be used effectively. Changes in the microbiota at individual taxonomic levels, for example at the phylum level with regards to obesity (Ley et al., 2005) and at the species level with regards to wound healing (Williams et al., 2017) have been identified. The issue is that various studies will use these individual measurements to distinguish the gut microbiota and their conditions of interest. However, given that the microbiome is a diverse community of microorganisms, a loss of individually defined-taxa may have no bearing on host function if there are other bacteria that can perform a similar role. Microbiome analyses that consider such a possibility, i.e. taxa at all phylogenetic scales and the importance of changes at different taxonomic levels, are rare. The novelty of the RF method proposed in my thesis (**Chapter 2**) is that bias towards individual taxa is avoided as much as possible, by focusing on phylogenetically-defined clades that are more reflective of the true microbial community. Given that numerous bacteria fall within a clade, we have focused on bacteria at all levels and are therefore not biased by single taxonomic levels.

It should be noted that other microbiome studies have used RF models before, although they have used OTU's as the input. For example, the RF technique was used to find associations between changes in brain structure and diet-dependent changes on the gut microbiota (Ong et al., 2018). Additionally, the microbiota was used to determine the post-mortem period in samples derived from grave soil and skin using the RF method (Belk et al., 2018). These studies would suggest that the RF technique can be used to identify particularly complex associations between the microbiota and conditions of interest (Ong et al., 2018). In a more relevant context, the RF has been used in the study of IBD (Papa et al., 2012; Tedjo et al., 2016). One study investigated whether RFs could identify important OTUs associated with remission or active CD in human patients (Tedjo et al., 2016) and identified remission with 73% accuracy and active CD with 78% accuracy. An additional study was able to discriminate between UC and CD in child patients using a RF model (Papa et al., 2012). As

stated, the limitation in these studies was that OTUs were assigned using QIIME and so potentially novel information could be missed. Specifically, by using phylogenetic clades, I could determine important characteristics about bacteria that distinguished our treatment groups that would not be possible with the OTU method. For example, I identified that moderately abundant bacteria of intermediate taxonomic levels were most important to distinguish between stool and mucus samples. This information could be obtained without assigning any identity to the 16S rRNA sequences, which again, would not be possible via the traditional OTU method. Once the important clades have been identified, taxonomic information can then be provided to identify the bacterial communities that do matter when being associated with a condition, such as inflammation.

One limitation of my study was that only a small number of mice developed colitis and the RF was unable to discriminate between healthy and colitic mice based on this small subset (**Chapter 2**). Indeed, RF's can be susceptible to variation in sample size between treatment groups and this is one of the limitations of the RF method (Walters et al., 2014). Therefore, it would be necessary to increase the sample size in order to accurately study whether the microbiota can be associated with inflammation. The model could then be applied to different datasets and different conditions of interest for further validation. For example, in the context of wound healing, to identify bacteria associated with healing and non-healing wounds. We already possess such data from human patients (Williams et al., 2018) and it would be interesting to see whether my method would be as effective in human data as compared to mouse data. My method could also be directly compared to the results of QIIME analysis. For example, can differences in the relative abundance of certain taxa identified via QIIME also be identified in our proposed method? Additionally, does the RF detect these bacteria as important? Such comparisons could help to further validate and strengthen our proposed method, which could be used as a supplement to QIIME analysis.

Overall, the gold-standard QIIME method has been used by almost every microbiome study to date. It provides a broad, accessible understanding of the gut microbiome and how it changes from one state to another. However, limitations in the method are such that data is biased to available taxonomies and focused on single level taxonomic changes. I have avoided this bias, by focusing on phylogenetically-defined clades and without the need to assign taxonomic information. It is therefore hoped that my method could be used to identify what bacterial communities do matter when distinguishing a condition of interest.

5.6 Using littermate controls: Why is cage effect so important?

One of the most important factors in mouse microbiota research is the use of littermate controls. Indeed it has been shown that cage effect has a powerful influence on the microbiota (Deloris Alexander et al., 2006; Hildebrand et al., 2013). However, many studies do not report on these crucial factors (Bramhall et al., 2015; Florez-Vargas et al., 2016). Using the RF model, I was able to illustrate this cage effect, where the microbiota could be associated with around 80% accuracy to the cage a mouse was reared in (**Chapter 2**), suggesting that each cage can have a distinct microbial signature. Cage effect may be due to the different environments as well as a result of influence of the maternal microbiome. Indeed my analysis showed it was possible to track back to the mother with ~98% accuracy based on the microbiota (**Appendix 1**). Therefore, mice split into different cages in different environments or with different mothers could have different microbiotas (Hildebrand et al., 2013). One study found that mice deficient in the caspase-1 activating protein, apoptosis-associated speck-like protein (ASC), developed severe DSS-induced colitis symptoms that were not seen in separately housed WT mice (Elinav et al., 2011). In isolation, this data would suggest that the ASC mutation increases the susceptibility of mice to colitis. However, the study then co-housed the $ASC^{-/-}$ mice with either WT mice that were bought in externally, or WT mice bred in-house. The externally purchased WT's developed severe colitis symptoms akin to the $ASC^{-/-}$ mice, whereas the WT's bred in-house developed milder colitis symptoms. Therefore it is possible that the bought in WT mice could have more pathogenic microbiota that facilitates colitis development, whereas the in-house WT mice could have a microbiota more robust to influence by the $ASC^{-/-}$ mice. Another study showed that the microbiota of mice housed in two separate rooms within the same animal facility had divergent microbial profiles, and mice from one room had greater inner mucus permeability compared to mice from the other room (Jakobsson et al., 2015; Rodriguez-Pineiro and Johansson, 2015). Collectively, these data highlight that the environment, be it a cage or a different location within the same animal facility, can lead to a significant and (most likely) unintended phenotypic outcome.

There also appears to be a strong maternal impact on the gut microbiota. Indeed, the mother could be associated with her offspring based on the gut microbiota alone with around 98% accuracy, irrespective of cage (**Appendix 1**). From my data, it might be inferred that the maternal effect is stronger than the environmental cage effect and homogenisation of the microbiota via coprophagy. This is in contrast to another study, which showed that cage effect was stronger than the effect of litter (Deloris Alexander et al., 2006). The focus of that

study was on the Altered Schaedler Flora (ASF), a 'concoction' of bacteria commonly found within the gut (Schaedler et al., 1965; Dewhirst et al., 1999). Germ-free mice were colonised with the ASF and thus maternal transmission of the microbiota could not be investigated. A study did attempt to quantify the effects of genetics, cage and inter-individual variation (Hildebrand et al., 2013) and found that strain had the biggest effect on the microbiota, followed by cage effect and then genetics. Notably, they co-housed mice from different strains, so although they could assess strain and cage effect, they could not assess genetic effect within an individual strain.

Cross-fostering mice (removing mouse pups from their original mother and transferring them to a surrogate) can also have a significant impact on the gut microbiota and host function. In a skin microbiota study, we showed that when separated, *Nod2*$^{-/-}$ mice had a significantly different microbiota compared to WT mice, concordant with a delayed wound-healing phenotype (Williams et al., 2017). However, when *Nod2*$^{-/-}$ mice were cross-fostered with WT mice, the WT mice acquired the defective skin microbiota and consequently the delayed wound-healing phenotype. Perhaps the most famous example of cross-fostering having significant physiological implications was that of the *T-bet*$^{-/-}$*x RAG2*$^{-/-}$ Ulcerative Colitis (TRUC) mice (Garrett et al., 2007). TRUC mice have an impairment to adaptive immunity that results in the development of colitis. However, when WT mice were cross-fostered with a TRUC mother, the WT mice developed colitis which suggested that the mother and cage effect can have a substantial physiological impact.

In human studies, the microbiota is also vertically transmissible (from mother to offspring) and horizontally transmissible (between family members) (Korpela et al., 2018). A recent study found that populations from the phylum Actinobacteria and the class Bacteroidia were almost always transferred to vaginally-delivered babies and that these bacteria were stable throughout adulthood (Korpela et al., 2018). However, Bacteroidia was the strain least likely to be replaced through diet and the ageing process. Sharing of strains between family members was significantly more frequent than non-related individuals (Korpela et al., 2018; Rothschild et al., 2018). However, the incidence of person to person transmission peaked between the ages of 2 and 10. The significance of these findings could have implications for treatments that involve altering the gut microbiota, such as faecal microbiota transplants (FMTs). Indeed, a mother with an especially 'robust' microbiota may have children that are resistant to such transplants, whereas a different mother may have a microbiota that is malleable to change. This could explain the discrepancy in the effectiveness of FMTs in some

patients. Similarly, the efficacy of probiotics could be influenced by the robustness of the maternally-inherited microbiota. Indeed, one study found that some individuals are 'permissive' to the effects of probiotics on the microbiota, whereas others are more resistant (Zmora et al., 2018). Notably, the study also showed that that probiotics can have a significant impact on the faecal microbiota, but almost no impact on the mucosal microbiota. Given the functionality of the mucus microbiota as discussed in **Section 5.2**, probiotics that can target and affect the mucus microbiota would be of great interest.

Overall, there are many factors in animal studies that can influence the microbiota that can be reasonably controlled for, i.e. cage and mother. I have illustrated that there can be functional consequences that result from separating, co-housing and using littermate controls. If the effect of a condition such as genotype is to be investigated, co-housed, littermate controls should be used to account for environmental effects. Given that maternal and person to person transmission of the microbiota also occurs in humans, there could be relevant functional impacts to consider that may be of clinical relevance for human health.

5.7 Conclusions

My thesis has overall aimed to explore how changes in the gut microbiota, especially the mucus-resident bacteria, can impact on host-health. Our machine learning model confirmed that the communities in the stool and mucus are distinct and highlighted the effect of the environment and genetics. I showed what characteristics were important for associating bacteria with certain conditions, for example, moderately abundant taxa of intermediate taxonomic levels could be used to discriminate between microbial niches. I subsequently showed that the presence of several metabolites was variable between genotype within the mucus of *mdr1a*$^{-/-}$ mice and in concert, identified significant transcriptional changes in these mice, particularly with regards to intestinal permeability. Therefore, metabolite and transcriptional changes, in addition to the altered mucus-resident microbiota could predispose these mice to colitis. In a mouse model with perturbations to mucosal immunity, I revealed that the biggest differences in the microbiota were found in the mucus, rather than stool. Collectively, I show that a focus on stool alone is insufficient to fully capture the diversity microbiota or its functional profile. Therefore, it is important to explore all niches within the gut in order to truly understand how the microbiota can impact on host function.

5.8 References

Ahmed, S., Macfarlane, G. T., Fite, A., McBain, A. J., Gilbert, P. & Macfarlane, S. (2007). Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. *Appl Environ Microbiol,* 73**,** 7435-42.

Al-Sadi, R., Khatib, K., Guo, S., Ye, D., Youssef, M. & Ma, T. (2011). Occludin regulates macromolecule flux across the intestinal epithelial tight junction barrier. *Am J Physiol Gastrointest Liver Physiol,* 300(6)**,** G1054-64.

Andriessen, E. M., Wilson, A. M., Mawambo, G., Dejda, A., Miloudi, K., Sennlaub, F. & Sapieha, P. (2016). Gut microbiota influences pathological angiogenesis in obesity-driven choroidal neovascularization. *EMBO Mol Med,* 8(12)**,** 1366-1379.

Arnold, I. C., Artola-Boran, M., Tallon de Lara, P., Kyburz, A., Taube, C., Ottemann, K., van den Broek, M., Yousefi, S., Simon, H. U. & Muller, A. (2018). Eosinophils suppress Th1 responses and restrict bacterially induced gastrointestinal inflammation. *J Exp Med,* 215(8)**,** 2055-2072.

Bajaj, J. S., Hylemon, P. B., Ridlon, J. M., Heuman, D. M., Daita, K., White, M. B., Monteith, P., Noble, N. A., Sikaroodi, M. & Gillevet, P. M. (2012). Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. *Am J Physiol Gastrointest Liver Physiol,* 303(6)**,** G675-85.

Belk, A., Xu, Z. Z., Carter, D. O., Lynne, A., Bucheli, S., Knight, R. & Metcalf, J. L. (2018). Microbiome Data Accurately Predicts the Postmortem Interval Using Random Forest Regression Models. *Genes (Basel),* 9(2).

Bramhall, M., Florez-Vargas, O., Stevens, R., Brass, A. & Cruickshank, S. (2015). Quality of methods reporting in animal models of colitis. *Inflamm Bowel Dis,* 21(6)**,** 1248-59.

Braniste, V., Al-Asmakh, M., Kowal, C., Anuar, F., Abbaspour, A., Toth, M., Korecka, A., Bakocevic, N., Ng, L. G., Kundu, P., Gulyas, B., Halldin, C., Hultenby, K., Nilsson, H., Hebert, H., Volpe, B. T., Diamond, B. & Pettersson, S. (2014). The gut microbiota influences blood-brain barrier permeability in mice. *Sci Transl Med,* 6(263)**,** 263ra158.

Canani, R. B., Costanzo, M. D., Leone, L., Pedata, M., Meli, R. & Calignano, A. (2011). Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World J Gastroenterol,* 17(12)**,** 1519-28.

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(5)**,** 335-6.

Carbonero, F., Benefiel, A. C., Alizadeh-Ghamsari, A. H. & Gaskins, H. R. (2012). Microbial pathways in colonic sulfur metabolism and links with health and disease. *Front Physiol,* 3**,** 448.

Carter, E. A., Tam, K. K., Armstrong, R. S. & Lay, P. A. (2009). Vibrational spectroscopic mapping and imaging of tissues and cells. *Biophys Rev,* 1(2)**,** 95-103.

Chelakkot, C., Choi, Y., Kim, D. K., Park, H. T., Ghim, J., Kwon, Y., Jeon, J., Kim, M. S., Jee, Y. K., Gho, Y. S., Park, H. S., Kim, Y. K. & Ryu, S. H. (2018). Akkermansia muciniphila-derived extracellular vesicles influence gut permeability through the regulation of tight junctions. *Exp Mol Med,* 50(2)**,** e450.

Chu, V. T., Beller, A., Rausch, S., Strandmark, J., Zanker, M., Arbach, O., Kruglov, A. & Berek, C. (2014). Eosinophils promote generation and maintenance of immunoglobulin-A-expressing plasma cells and contribute to gut immune homeostasis. *Immunity,* 40**,** 582-93.

Collett, A., Higgs, N. B., Gironella, M., Zeef, L. A., Hayes, A., Salmo, E., Haboubi, N., Iovanna, J. L., Carlson, G. L. & Warhurst, G. (2008). Early molecular and functional changes in colonic epithelium that precede increased gut permeability during colitis development in mdr1a(-/-) mice. *Inflamm Bowel Dis,* 14(5)**,** 620-31.

Dalton, J. E., Cruickshank, S. M., Egan, C. E., Mears, R., Newton, D. J., Andrew, E. M., Lawrence, B., Howell, G., Else, K. J., Gubbels, M. J., Striepen, B., Smith, J. E., White, S. J. & Carding, S. R. (2006). Intraepithelial gammadelta+ lymphocytes maintain the integrity of intestinal epithelial tight junctions in response to infection. *Gastroenterology,* 131(3)**,** 818-29.

de Jong, N. S., Leach, S. T. & Day, A. S. (2006). Fecal S100A12: a novel noninvasive marker in children with Crohn's disease. *Inflamm Bowel Dis,* 12(7)**,** 566-72.

De Preter, V., Machiels, K., Joossens, M., Arijs, I., Matthys, C., Vermeire, S., Rutgeerts, P. & Verbeke, K. (2015). Faecal metabolite profiling identifies medium-chain fatty acids as discriminating compounds in IBD. *Gut,* 64(3)**,** 447-58.

Deloris Alexander, A., Orcutt, R. P., Henry, J. C., Baker, J., Jr., Bissahoyo, A. C. & Threadgill, D. W. (2006). Quantitative PCR assays for mouse enteric flora reveal strain-dependent differences in composition that are influenced by the microenvironment. *Mamm Genome,* 17(11)**,** 1093-104.

Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., Pudlo, N. A., Kitamoto, S., Terrapon, N., Muller, A., Young, V. B., Henrissat, B., Wilmes, P., Stappenbeck, T. S., Nunez, G. & Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell,* 167(5)**,** 1339-1353.e21.

Dewhirst, F. E., Chien, C. C., Paster, B. J., Ericson, R. L., Orcutt, R. P., Schauer, D. B. & Fox, J. G. (1999). Phylogeny of the defined murine microbiota: altered Schaedler flora. *Appl Environ Microbiol,* 65(8)**,** 3287-92.

Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E., Francis-McIntyre, S., Anderson, N., Brown, M., Knowles, J. D., Halsall, A., Haselden, J. N., Nicholls, A. W., Wilson, I. D., Kell, D. B. & Goodacre, R. (2011). Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc,* 6(7)**,** 1060-83.

Dvorak, H. F., Detmar, M., Claffey, K. P., Nagy, J. A., van de Water, L. & Senger, D. R. (1995). Vascular permeability factor/vascular endothelial growth factor: an important mediator of angiogenesis in malignancy and inflammation. *Int Arch Allergy Immunol,* 107(1-3)**,** 233-5.

Elinav, E., Strowig, T., Kau, A. L., Henao-Mejia, J., Thaiss, C. A., Booth, C. J., Peaper, D. R., Bertin, J., Eisenbarth, S. C., Gordon, J. I. & Flavell, R. A. (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell,* 145(5)**,** 745-57.

Fabre, V., Beiting, D. P., Bliss, S. K., Gebreselassie, N. G., Gagliardo, L. F., Lee, N. A., Lee, J. J. & Appleton, J. A. (2009). Eosinophil deficiency compromises parasite survival in chronic nematode infection. *J Immunol,* 182(3)**,** 1577-83.

Fan, P., Liu, P., Song, P., Chen, X. & Ma, X. (2017). Moderate dietary protein restriction alters the composition of gut microbiota and improves ileal barrier function in adult pig model. *Sci Rep,* 7**,** 43412.

Florez-Vargas, O., Brass, A., Karystianis, G., Bramhall, M., Stevens, R., Cruickshank, S. & Nenadic, G. (2016). Bias in the reporting of sex and age in biomedical research on mouse models. *Elife,* 5.

Forbes, E., Murase, T., Yang, M., Matthaei, K. I., Lee, J. J., Lee, N. A., Foster, P. S. & Hogan, S. P. (2004). Immunopathogenesis of experimental ulcerative colitis is mediated by eosinophil peroxidase. *J Immunol,* 172(9)**,** 5664-75.

Forman, R., Bramhall, M., Logunova, L., Svensson-Frej, M., Cruickshank, S. M. & Else, K. J. (2016). Eosinophils may play regionally disparate roles in influencing IgA(+) plasma cell numbers during large and small intestinal inflammation. *BMC Immunol,* 17(1)**,** 12.

Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N. & Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A,* 104(34)**,** 13780-5.

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T. A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., Takahashi, M., Fukuda, N. N., Murakami, S., Miyauchi, E., Hino, S., Atarashi, K., Onawa, S., Fujimura, Y., Lockett, T., Clarke, J. M., Topping, D. L., Tomita, M., Hori, S., Ohara, O., Morita, T., Koseki, H., Kikuchi, J., Honda, K., Hase, K. & Ohno, H. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature,* 504**,** 446-50.

Furuta, G. T., Nieuwenhuis, E. E., Karhausen, J., Gleich, G., Blumberg, R. S., Lee, J. J. & Ackerman, S. J. (2005). Eosinophils alter colonic epithelial barrier function: role for major basic protein. *Am J Physiol Gastrointest Liver Physiol,* 289(5)**,** G890-7.

Ganta, V. C., Cromer, W., Mills, G. L., Traylor, J., Jennings, M., Daley, S., Clark, B., Mathis, J. M., Bernas, M., Boktor, M., Jordan, P., Witte, M. & Alexander, J. S. (2010). Angiopoietin-2 in experimental colitis. *Inflamm Bowel Dis,* 16(6)**,** 1029-39.

Garner, C. E., Smith, S., de Lacy Costello, B., White, P., Spencer, R., Probert, C. S. & Ratcliffe, N. M. (2007). Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. *Faseb j,* 21(8)**,** 1675-88.

Garrett, W. S., Lord, G. M., Punit, S., Lugo-Villarino, G., Mazmanian, S. K., Ito, S., Glickman, J. N. & Glimcher, L. H. (2007). Communicable ulcerative colitis induced by T-bet deficiency in the innate immune system. *Cell,* 131(1)**,** 33-45.

Gassler, N., Rohr, C., Schneider, A., Kartenbeck, J., Bach, A., Obermuller, N., Otto, H. F. & Autschbach, F. (2001). Inflammatory bowel disease is associated with changes of enterocytic junctions. *Am J Physiol Gastrointest Liver Physiol,* 281(1)**,** G216-28.

Geirnaert, A., Calatayud, M., Grootaert, C., Laukens, D., Devriese, S., Smagghe, G., De Vos, M., Boon, N. & Van de Wiele, T. (2017). Butyrate-producing bacteria supplemented in vitro to Crohn's disease patient microbiota increased butyrate production and enhanced intestinal epithelial barrier integrity. *Sci Rep,* 7(1)**,** 11450.

Gevers, D., Kugathasan, S., Denson, L. A., Vazquez-Baeza, Y., Van Treuren, W., Ren, B., Schwager, E., Knights, D., Song, S. J., Yassour, M., Morgan, X. C., Kostic, A. D., Luo, C., Gonzalez, A., McDonald, D., Haberman, Y., Walters, T., Baker, S., Rosh, J., Stephens, M., Heyman, M., Markowitz, J., Baldassano, R., Griffiths, A., Sylvester, F., Mack, D., Kim, S., Crandall, W., Hyams, J., Huttenhower, C., Knight, R. & Xavier, R. J. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe,* 15(3)**,** 382-392.

Glymenaki, M., Barnes, A., Hagan, S. O., Warhurst, G., McBain, A. J., Wilson, I. D., Kell, D. B., Else, K. J. & Cruickshank, S. M. (2017a). Stability in metabolic phenotypes and inferred metagenome profiles before the onset of colitis-induced inflammation. *Sci Rep,* 7(1)**,** 8836.

Glymenaki, M., Singh, G., Brass, A., Warhurst, G., McBain, A. J., Else, K. J. & Cruickshank, S. M. (2017b). Compositional Changes in the Gut Mucus Microbiota Precede the Onset of Colitis-Induced Inflammation. *Inflamm Bowel Dis,* 23(6)**,** 912-922.

Grice, E. A. & Segre, J. A. (2011). The skin microbiome. *Nat Rev Microbiol,* 9(4)**,** 244-53.

Haberland, K., Ackermann, J. A., Ipseiz, N., Culemann, S., Pracht, K., Englbrecht, M., Jack, H. M., Schett, G., Schuh, W. & Kronke, G. (2018). Eosinophils are not essential for maintenance of murine plasma cells in the bone marrow. *Eur J Immunol*.

Hildebrand, F., Nguyen, T. L., Brinkman, B., Yunta, R. G., Cauwe, B., Vandenabeele, P., Liston, A. & Raes, J. (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol,* 14(1)**,** R4.

Hollister, E. B., Gao, C. & Versalovic, J. (2014). Compositional and functional features of the gastrointestinal microbiome and their effects on human health. *Gastroenterology,* 146**,** 1449-58.

Ijssennagger, N., van der Meer, R. & van Mil, S. W. C. (2016). Sulfide as a Mucus Barrier-Breaker in Inflammatory Bowel Disease? *Trends Mol Med,* 22(3)**,** 190-199.

Ishibashi, K., Hara, S. & Kondo, S. (2009). Aquaporin water channels in mammals. *Clin Exp Nephrol,* 13(2)**,** 107-117.

Ishibashi, K., Kuwahara, M., Gu, Y., Tanaka, Y., Marumo, F. & Sasaki, S. (1998). Cloning and functional expression of a new aquaporin (AQP9) abundantly expressed in the peripheral leukocytes permeable to water and urea, but not to glycerol. *Biochem Biophys Res Commun,* 244(1)**,** 268-74.

Jacobsen, E. A., Lee, N. A. & Lee, J. J. (2014). Re-defining the Unique Roles for Eosinophils in Allergic Respiratory Inflammation. *Clin Exp Allergy,* 44(9)**,** 1119-36.

Jakobsson, H. E., Rodriguez-Pineiro, A. M., Schutte, A., Ermund, A., Boysen, P., Bemark, M., Sommer, F., Backhed, F., Hansson, G. C. & Johansson, M. E. (2015). The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep,* 16**,** 164-77.

Janssen, A. W. F., Katiraei, S., Bartosinska, B., Eberhard, D., Willems van Dijk, K. & Kersten, S. (2018). Loss of angiopoietin-like 4 (ANGPTL4) in mice with diet-induced obesity uncouples visceral obesity from glucose intolerance partly via the gut microbiota. *Diabetologia,* 61(6)**,** 1447-1458.

Johansson, M. E., Gustafsson, J. K., Sjoberg, K. E., Petersson, J., Holm, L., Sjovall, H. & Hansson, G. C. (2010). Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLoS One,* 5(8)**,** e12238.

Johnson, A. M., Costanzo, A., Gareau, M. G., Armando, A. M., Quehenberger, O., Jameson, J. M. & Olefsky, J. M. (2015). High fat diet causes depletion of intestinal eosinophils associated with intestinal permeability. *PLoS One,* 10(4)**,** e0122195.

Kapel, N., Benahmed, N., Morali, A., Svahn, J., Canioni, D., Goulet, O. & Ruemmele, F. M. (2009). Fecal beta-defensin-2 in children with inflammatory bowel diseases. *J Pediatr Gastroenterol Nutr,* 48(1)**,** 117-20.

Khalil, N. A., Walton, G. E., Gibson, G. R., Tuohy, K. M. & Andrews, S. C. (2014). In vitro batch cultures of gut microbiota from healthy and ulcerative colitis (UC) subjects suggest that sulphate-reducing bacteria levels are raised in UC and by a protein-rich diet. *Int J Food Sci Nutr,* 65(1)**,** 79-88.

Kim, M., Qie, Y., Park, J. & Kim, C. H. (2016). Gut Microbial Metabolites Fuel Host Antibody Responses. *Cell Host Microbe,* 20(2)**,** 202-14.

Korpela, K., Costea, P., Coelho, L. P., Kandels-Lewis, S., Willemsen, G., Boomsma, D. I., Segata, N. & Bork, P. (2018). Selective maternal seeding and environment shape the human gut microbiome. *Genome Res,* 28(4)**,** 561-568.

Koutroubakis, I. E., Xidakis, C., Karmiris, K., Sfiridaki, A., Kandidaki, E. & Kouroumalis, E. A. (2006). Potential role of soluble angiopoietin-2 and Tie-2 in patients with inflammatory bowel disease. *Eur J Clin Invest,* 36(2)**,** 127-32.

Le Gall, G., Noor, S. O., Ridgway, K., Scovell, L., Jamieson, C., Johnson, I. T., Colquhoun, I. J., Kemsley, E. K. & Narbad, A. (2011). Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. *J Proteome Res,* 10(9)**,** 4208-18.

Lee, J. J., Dimina, D., Macias, M. P., Ochkur, S. I., McGarry, M. P., O'Neill, K. R., Protheroe, C., Pero, R., Nguyen, T., Cormier, S. A., Lenkiewicz, E., Colbert, D., Rinaldi, L., Ackerman, S. J., Irvin, C. G. & Lee, N. A. (2004). Defining a link with asthma in mice congenitally deficient in eosinophils. *Science,* 305(5691)**,** 1773-6.

Lee, S. H. (2015). Intestinal permeability regulation by tight junction: implication on inflammatory bowel diseases. *Intest Res,* 13(1)**,** 11-8.

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D. & Gordon, J. I. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A,* 102(31)**,** 11070-5.

Linares, P. M., Chaparro, M. & Gisbert, J. P. (2014). Angiopoietins in inflammation and their implication in the development of inflammatory bowel disease. A review. *J Crohns Colitis,* 8(3)**,** 183-90.

Loubinoux, J., Bronowicki, J. P., Pereira, I. A., Mougenel, J. L. & Faou, A. E. (2002). Sulfate-reducing bacteria in human feces and their association with inflammatory bowel diseases. *FEMS Microbiol Ecol,* 40(2)**,** 107-12.

Ma, X., Fan, P. X., Li, L. S., Qiao, S. Y., Zhang, G. L. & Li, D. F. (2012). Butyrate promotes the recovering of intestinal wound healing through its positive effect on the tight junctions. *J Anim Sci,* 90 Suppl 4**,** 266-8.

Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., Ballet, V., Claes, K., Van Immerseel, F., Verbeke, K., Ferrante, M., Verhaegen, J., Rutgeerts, P. & Vermeire, S. (2014). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut,* 63(8)**,** 1275-83.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., Roca, J. & Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut,* 55(2)**,** 205-11.

Manolakis, A. C., Kapsoritakis, A. N., Georgoulias, P., Tzavara, C., Valotassiou, V., Kapsoritaki, A. & Potamianos, S. P. (2010). Moderate performance of serum S100A12, in distinguishing inflammatory bowel disease from irritable bowel syndrome. *BMC Gastroenterol,* 10**,** 118.

Masterson, J. C., McNamee, E. N., Fillon, S. A., Hosford, L., Harris, R., Fernando, S. D., Jedlicka, P., Iwamoto, R., Jacobsen, E., Protheroe, C., Eltzschig, H. K., Colgan, S. P., Arita, M., Lee, J. J. & Furuta, G. T. (2015). Eosinophil-mediated signalling attenuates inflammatory responses in experimental colitis. *Gut,* 64(8)**,** 1236-47.

McHardy, I. H., Goudarzi, M., Tong, M., Ruegger, P. M., Schwager, E., Weger, J. R., Graeber, T. G., Sonnenburg, J. L., Horvath, S., Huttenhower, C., McGovern, D. P., Fornace, A. J., Jr., Borneman, J. & Braun, J. (2013). Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome,* 1**,** 17.

Moffatt, M. F. & Cookson, W. O. (2017). The lung microbiome in health and disease. *Clin Med (Lond),* 17(6)**,** 525-529.

Murdoch, T. B., Fu, H., MacFarlane, S., Sydora, B. C., Fedorak, R. N. & Slupsky, C. M. (2008). Urinary metabolic profiles of inflammatory bowel disease in interleukin-10 gene-deficient mice. *Anal Chem,* 80(14)**,** 5524-31.

Nenci, A., Becker, C., Wullaert, A., Gareus, R., van Loo, G., Danese, S., Huth, M., Nikolaev, A., Neufert, C., Madison, B., Gumucio, D., Neurath, M. F. & Pasparakis, M. (2007). Epithelial NEMO links innate immunity to chronic intestinal inflammation. *Nature,* 446(7135)**,** 557-61.

Neubauer, K., Matusiewicz, M., Bednarz-Misa, I., Gorska, S., Gamian, A. & Krzystek-Korpacka, M. (2018). Diagnostic Potential of Systemic Eosinophil-Associated Cytokines and Growth Factors in IBD. *Gastroenterol Res Pract,* 2018**,** 7265812.

Norouzinia, M., Chaleshi, V., Alizadeh, A. H. M. & Zali, M. R. (2017). Biomarkers in inflammatory bowel diseases: insight into diagnosis, prognosis and treatment. *Gastroenterol Hepatol Bed Bench,* 10(3)**,** 155-167.

O'Connell, A. E., Hess, J. A., Santiago, G. A., Nolan, T. J., Lok, J. B., Lee, J. J. & Abraham, D. (2011). Major basic protein from eosinophils and myeloperoxidase from neutrophils are required for protective immunity to Strongyloides stercoralis in mice. *Infect Immun,* 79(7)**,** 2770-8.

Ong, I. M., Gonzalez, J. G., McIlwain, S. J., Sawin, E. A., Schoen, A. J., Adluru, N., Alexander, A. L. & Yu, J. J. (2018). Gut microbiome populations are associated with structure-specific changes in white matter architecture. *Transl Psychiatry,* 8(1)**,** 6.

Palm, N. W., de Zoete, M. R., Cullen, T. W., Barry, N. A., Stefanowski, J., Hao, L., Degnan, P. H., Hu, J., Peter, I., Zhang, W., Ruggiero, E., Cho, J. H., Goodman, A. L. & Flavell, R. A. (2014). Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell,* 158**,** 1000-10.

Papa, E., Docktor, M., Smillie, C., Weber, S., Preheim, S. P., Gevers, D., Giannoukos, G., Ciulla, D., Tabbaa, D., Ingram, J., Schauer, D. B., Ward, D. V., Korzenik, J. R., Xavier, R. J., Bousvaros, A. & Alm, E. J. (2012). Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. *PLoS One,* 7(6)**,** e39242.

Park, J. H., Kotani, T., Konno, T., Setiawan, J., Kitamura, Y., Imada, S., Usui, Y., Hatano, N., Shinohara, M., Saito, Y., Murata, Y. & Matozaki, T. (2016). Promotion of Intestinal Epithelial Cell Turnover by Commensal Bacteria: Role of Short-Chain Fatty Acids. *PLoS One,* 11(5)**,** e0156334.

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., Martinez, X., Varela, E., Sarrabayrouse, G., Machiels, K., Vermeire, S., Sokol, H., Guarner, F. & Manichanh, C. (2017). A microbial signature for Crohn's disease. *Gut,* 66(5)**,** 813-822.

Ramirez, K. S., Knight, C. G., de Hollander, M., Brearley, F. Q., Constantinides, B., Cotton, A., Creer, S., Crowther, T. W., Davison, J., Delgado-Baquerizo, M., Dorrepaal, E., Elliott, D. R., Fox, G., Griffiths, R. I., Hale, C., Hartman, K., Houlden, A., Jones, D. L., Krab, E. J., Maestre, F. T., McGuire, K. L., Monteux, S., Orr, C. H., van der Putten, W. H., Roberts, I. S., Robinson, D. A., Rocca, J. D., Rowntree, J., Schlaeppi, K., Shepherd, M., Singh, B. K., Straathof, A. L., Bhatnagar, J. M., Thion, C., van der Heijden, M. G. A. & de Vries, F. T. (2017). Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat Microbiol*.

Ranjan, R., Rani, A., Metwally, A., McGee, H. S. & Perkins, D. L. (2016). Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun,* 469(4)**,** 967-77.

Rey, F. E., Gonzalez, M. D., Cheng, J., Wu, M., Ahern, P. P. & Gordon, J. I. (2013). Metabolic niche of a prominent sulfate-reducing human gut bacterium. *Proc Natl Acad Sci U S A,* 110(33)**,** 13582-7.

Ricanek, P., Lunde, L. K., Frye, S. A., Stoen, M., Nygard, S., Morth, J. P., Rydning, A., Vatn, M. H., Amiry-Moghaddam, M. & Tonjum, T. (2015). Reduced expression of aquaporins in human intestinal mucosa in early stage inflammatory bowel disease. *Clin Exp Gastroenterol,* 8**,** 49-67.

Rodriguez-Pineiro, A. M. & Johansson, M. E. (2015). The colonic mucus protection depends on the microbiota. *Gut Microbes,* 6(5)**,** 326-30.

Rothenberg, M. E. (2004). Eosinophilic gastrointestinal disorders (EGID). *J Allergy Clin Immunol,* 113(1)**,** 11-28; quiz 29.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P. I., Godneva, A., Kalka, I. N., Bar, N., Shilo, S., Lador, D., Vila, A. V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B. C., Avnit-Sagi, T., Lotan-Pompan, M., Weinberger, A., Halpern, Z., Carmi, S., Fu, J., Wijmenga, C., Zhernakova, A., Elinav, E. & Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature,* 555(7695)**,** 210-215.

Round, J. L. & Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol,* 9(5)**,** 313-23.

Saitou, M., Fujimoto, K., Doi, Y., Itoh, M., Fujimoto, T., Furuse, M., Takano, H., Noda, T. & Tsukita, S. (1998). Occludin-deficient embryonic stem cells can differentiate into polarized epithelial cells bearing tight junctions. *J Cell Biol,* 141(2)**,** 397-408.

Scaldaferri, F., Vetrano, S., Sans, M., Arena, V., Straface, G., Stigliano, E., Repici, A., Sturm, A., Malesci, A., Panes, J., Yla-Herttuala, S., Fiocchi, C. & Danese, S. (2009). VEGF-A links angiogenesis and inflammation in inflammatory bowel disease pathogenesis. *Gastroenterology,* 136(2)**,** 585-95.e5.

Schaedler, R. W., Dubs, R. & Costello, R. (1965). ASSOCIATION OF GERMFREE MICE WITH BACTERIA ISOLATED FROM NORMAL MICE. *J Exp Med,* 122**,** 77-82.

Schicho, R., Nazyrova, A., Shaykhutdinov, R., Duggan, G., Vogel, H. J. & Storr, M. (2010). Quantitative metabolomic profiling of serum and urine in DSS-induced ulcerative colitis of mice by (1)H NMR spectroscopy. *J Proteome Res,* 9(12)**,** 6265-73.

Schicho, R., Shaykhutdinov, R., Ngo, J., Nazyrova, A., Schneider, C., Panaccione, R., Kaplan, G. G., Vogel, H. J. & Storr, M. (2012). Quantitative metabolomic profiling of serum, plasma, and urine by (1)H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. *J Proteome Res,* 11(6)**,** 3344-57.

Schirbel, A., Kessler, S., Rieder, F., West, G., Rebert, N., Asosingh, K., McDonald, C. & Fiocchi, C. (2013). Pro-angiogenic activity of TLRs and NLRs: a novel link between gut microbiota and intestinal angiogenesis. *Gastroenterology,* 144(3)**,** 613-623.e9.

Schloss, P. D. (2010). The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Comput Biol,* 6**,** e1000844.

Sciences, O. (2018). *The Oricol Sampling Kit* [Online]. Available: http://www.originsciences.com/index.php/the-oricol-sampling-kit [Accessed August 31st 2018].

Smith, L. A. & Gaya, D. R. (2012). Utility of faecal calprotectin analysis in adult inflammatory bowel disease. *World J Gastroenterol,* 18(46)**,** 6782-9.

Spencer, L. A., Szela, C. T., Perez, S. A., Kirchhoffer, C. L., Neves, J. S., Radke, A. L. & Weller, P. F. (2009). Human eosinophils constitutively express multiple Th1, Th2, and immunoregulatory cytokines that are secreted rapidly and differentially. *J Leukoc Biol,* 85(1)**,** 117-23.

Srinivasan, B., Kolli, A. R., Esch, M. B., Abaci, H. E., Shuler, M. L. & Hickman, J. J. (2015). TEER measurement techniques for in vitro barrier model systems. *J Lab Autom,* 20(2)**,** 107-26.

Suez, J., Zmora, N., Zilberman-Schapira, G., Mor, U., Dori-Bachash, M., Bashiardes, S., Zur, M., Regev-Lehavi, D., Ben-Zeev Brik, R., Federici, S., Horn, M., Cohen, Y., Moor, A. E., Zeevi, D., Korem, T., Kotler, E., Harmelin, A., Itzkovitz, S., Maharshak, N., Shibolet, O., Pevsner-Fischer, M., Shapiro, H., Sharon, I., Halpern, Z., Segal, E. & Elinav, E. (2018). Post-Antibiotic Gut Mucosal Microbiome Reconstitution Is Impaired by Probiotics and Improved by Autologous FMT. *Cell,* 174(6)**,** 1406-1423.e16.

Swartz, J. M., Dyer, K. D., Cheever, A. W., Ramalingam, T., Pesnicak, L., Domachowske, J. B., Lee, J. J., Lee, N. A., Foster, P. S., Wynn, T. A. & Rosenberg, H. F. (2006). Schistosoma mansoni infection in eosinophil lineage-ablated mice. *Blood,* 108(7)**,** 2420-7.

Tedjo, D. I., Smolinska, A., Savelkoul, P. H., Masclee, A. A., van Schooten, F. J., Pierik, M. J., Penders, J. & Jonkers, D. M. (2016). The fecal microbiota as a biomarker for disease activity in Crohn's disease. *Sci Rep,* 6**,** 35216.

Trompette, A., Gollwitzer, E. S., Yadava, K., Sichelstiel, A. K., Sprenger, N., Ngom-Bru, C., Blanchard, C., Junt, T., Nicod, L. P., Harris, N. L. & Marsland, B. J. (2014). Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat Med,* 20(2)**,** 159-66.

Vieira, A. T., Fagundes, C. T., Alessandri, A. L., Castor, M. G., Guabiraba, R., Borges, V. O., Silveira, K. D., Vieira, E. L., Goncalves, J. L., Silva, T. A., Deruaz, M., Proudfoot, A. E., Sousa, L. P. & Teixeira, M. M. (2009). Treatment with a novel chemokine-binding protein or eosinophil lineage-ablation protects mice from experimental colitis. *Am J Pathol,* 175(6)**,** 2382-91.

Walters, W. A., Xu, Z. & Knight, R. (2014). Meta-analyses of human gut microbes associated with obesity and IBD. *FEBS Lett,* 588(22)**,** 4223-33.

Williams, H., Campbell, L., Crompton, R. A., Singh, G., McHugh, B. J., Davidson, D. J., McBain, A. J., Cruickshank, S. M. & Hardman, M. J. (2018). Microbial Host Interactions and Impaired Wound Healing in Mice and Humans: Defining a Role for BD14 and NOD2. *J Invest Dermatol*.

Williams, H., Crompton, R. A., Thomason, H. A., Campbell, L., Singh, G., McBain, A. J., Cruickshank, S. M. & Hardman, M. J. (2017). Cutaneous Nod2 Expression Regulates the Skin Microbiome and Wound Healing in a Murine Model. *J Invest Dermatol,* 137(11)**,** 2427-2436.

Wu, D., Molofsky, A. B., Liang, H. E., Ricardo-Gonzalez, R. R., Jouihan, H. A., Bando, J. K., Chawla, A. & Locksley, R. M. (2011). Eosinophils sustain adipose alternatively activated macrophages associated with glucose homeostasis. *Science,* 332**,** 243-7.

Zhang, W., Xu, Y., Chen, Z., Xu, Z. & Xu, H. (2011). Knockdown of aquaporin 3 is involved in intestinal barrier integrity impairment. *FEBS Lett,* 585(19)**,** 3113-9.

Zhang, X., Deeke, S. A., Ning, Z., Starr, A. E., Butcher, J., Li, J., Mayne, J., Cheng, K., Liao, B., Li, L., Singleton, R., Mack, D., Stintzi, A. & Figeys, D. (2018). Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat Commun,* 9(1)**,** 2873.

Zmora, N., Zilberman-Schapira, G., Suez, J., Mor, U., Dori-Bachash, M., Bashiardes, S., Kotler, E., Zur, M., Regev-Lehavi, D., Brik, R. B., Federici, S., Cohen, Y., Linevsky, R., Rothschild, D., Moor, A. E., Ben-Moshe, S., Harmelin, A., Itzkovitz, S., Maharshak, N., Shibolet, O., Shapiro, H., Pevsner-Fischer, M., Sharon, I., Halpern, Z., Segal, E. & Elinav, E. (2018). Personalized Gut Mucosal Colonization Resistance to Empiric Probiotics Is Associated with Unique Host and Microbiome Features. *Cell,* 174(6)**,** 1388-1405.e21.

Zoetendal, E. G., von Wright, A., Vilpponen-Salmela, T., Ben-Amor, K., Akkermans, A. D. & de Vos, W. M. (2002). Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. *Appl Environ Microbiol,* 68**,** 3401-7.

Chapter Six

Supplementary Materials and Methods

6. Supplementary Materials and Methods

6.1 Animal Maintenance

All procedures were performed in accordance with the Home Office project licence (7018127).

ΔdblGATA-1$^{-/-}$ mice (Shivdasani et al., 1997), a kind gift from Professor Avery August (Pennsylvania State University, USA) were bred with C57BL/6 mice to produce the F2 generation, which were used for all experiments. Mice from each litter were co-housed but separated by gender. Thus, WT male mice and heterozygous (Het) female mice, and *ΔdblGATA-1*$^{-/-}$ mice from the same litters were used for all subsequent experiments. It should be noted that due to the nature of the mutation, female mice could only be Het in this study but function as WTs, given that their eosinophils remain intact. Male mice at 12 weeks old and female mice at 12 weeks and 1 year of age were used for experiments.

Mdr1a$^{-/-}$ mice (FVB.129P2-Abcb1atm1Bor N7) (Schinkel et al., 1994) were bred with control FVB mice purchased from Taconic Farms (Albany, NY), to produce the F2 generation. Thus, WT and *mdr1a*$^{-/-}$ mice from the same litters were used for all subsequent experiments. Male mice at 6 and 18 weeks of age were used for experiments.

Food (Beekay Rat and Mouse Diet No1 pellets; B&K Universal, UK) and water were available *ad libitum*. Ambient temperature was maintained at 21 (+/- 2°C) and the relative humidity was 55 (+/- 10%) with a 12h light/dark cycle. All animals were kept under specific, pathogen-free (SPF) conditions at the University of Manchester and experiments were performed according to the regulations issued by the Home Office under amended ASPA, 2012.

6.1.1 Genotyping of *ΔdblGATA-1*$^{-/-}$ mice and *mdr1a*$^{-/-}$ mice

The genotype of the *ΔdblGATA-1*$^{-/-}$ mice and *mdr1a*$^{-/-}$ mice was determined by PCR and gel electrophoresis. Ear snips were collected and placed into buffer (100mM Tris-HCl, pH 8.5; 5mM EDTA, pH 8.0; 200mM NaCl and 0.2% SDS). Samples were incubated at 54ºC under agitation for one hour in buffer containing 1µg/ml Proteinase K (Promega, Wisconsin, USA) After centrifugation (13000xg for 10 minutes) the resultant supernatant was mixed with isopropanol and incubated on ice for 40 minutes, washed in 70% ethanol and re-suspended in nuclease free water (Promega). PCR was then performed using a G-Storm Thermocycler (Labtech, Sussex, UK) and appropriate primers (Table 1). 1µL of DNA was added to a PCR

‘mastermix,’ consisting of Ranger buffer (Bioline, London, UK), Ranger polymerase (Bioline), 5µM of each primer and made up in nuclease free water

The PCR conditions used for the *ΔdblGATA-1*$^{-/-}$ mice are summarised as follows: denaturation/ polymerase activation at 95˚C for 5 minutes, followed by 30 cycles of denaturation at 95˚C for 1 minute, annealing at 61˚C for 1 minute, elongation at 68˚C for 1 minute, and a final elongation step at 68˚ for 1 minute. Samples were run on a 1.5% agarose gel and imaged with a Bio-Doc-It® Imaging System (UVP, Cambridge, UK). A band at 509 base pairs (bp) indicated a homozygous mutation, a band at 459bp indicated a wild type copy of the *ΔdblGATA-1* gene and two bands indicated a heterozygous mutant.

The PCR conditions used for the *mdr1a*$^{-/-}$ mice are summarised as follows: denaturation/ activation at 95°C for 8 minutes followed by 30 cycles of denaturation at 94°C for 45 seconds, annealing at 60°C for 1 minute and elongation 72°C for 1 minute, and a final elongation step at 72°C for 5 minutes. A band at 500bp corresponded to a homozygous mutation and a band at 270bp corresponded to a wildtype copy of the mdr1a gene. Two bands indicated a heterozygous mutant.

Table 1: List of primers used

Gene	Primers	Sequence
GATA-1	GATA 1 S	5’-CCCAATCCTCTGGACTCCCA-3’
	GATA 1 AS	5’-CCTACTGTGTACCAGGCTAT-3’
mdr1a	AS2	5’–CTCCTCCAAGGTGCATAGACC–3’
	AW2	5’–CCCAGCTCTTCATCTAACTACCCTG-3’
	AKO2	5’-CTTCCCAGCCTCTGAGCCCAG-3’

6.2 Histology

6.2.1 Tissue Processing

Ileum, proximal and distal colon snips were fixed in Carnoy's fixative (60% ethanol absolute, 30% chloroform and 10% glacial acetic acid) to preserve the mucus. Carnoy's fixed samples were incubated in two changes of dry methanol (Sigma-Aldrich, Dorset, UK) for 30 minutes each, followed by absolute ethanol (ThermoFisher Scientific, Paisley, UK) for two incubations at 30 minutes each. Tissue cassettes were processed in a Micro-spin Tissue Processor STP120 (ThermoFisher Scientific) and were immersed in 70% ethanol for 15 minutes, 90% ethanol for 30 minutes, 95% ethanol for 30 minutes, two changes of 100% ethanol for 60 minutes, 100% ethanol for 20 minutes, xylene for 15 minutes, two changes of xylene for 30 minutes (all at 40°C) and twice in fibrowax pastillated wax (Becton Dickinson, Oxford, UK) for 1 hour at 60°C. Tissue samples were then mounted in paraffin blocks using a Leica Biosystems Embedding Station (Leica Biosystems, Milton Keynes, UK). 5μm tissue sections were cut from these blocks using a Leica Biosystems microtome and adhered to uncoated microscope slides (ThermoFisher Scientific). Slides were dried for 48 hours at 50ºC before use.

6.2.2 Haematoxylin and Eosin (H+E) Staining

Tissue sections were dewaxed using Citroclear Solution (Sigma-Aldrich) for 15 minutes and rehydrated through a series of decreasing ethanol concentrations (100%, 90%, 70% and 50%) for 2 minutes and placed into distilled water (dH_2O). Following this rehydration, slides were stained with Harris Haematoxylin (Sigma-Aldrich, Dorset, UK), cleaned in running tap water for 1 minute and differentiated in acid alcohol solution (1% HCL in 70% ethanol) for 10 seconds. The slides were placed into running tap water for 5 minutes before staining with Eosin solution (Sigma-Aldrich) for 30 seconds. Sections were washed in tap water for 30 seconds and dehydrated through a series of increasing alcohol concentrations (50%, 70%, 90% and 100%) for 30 seconds each. Sections were then placed into two batches of fresh Citroclear for 2.5 minutes each. Slides were then mounted using DPX mounting medium (ThermoFisher Scientific). Tissue sections were analysed for length of intestinal crypts, villi and muscle wall thickness. 20 measurements were made per sample, for each parameter and averaged. Slides were analysed in a blinded manner. Images were captured using a Nikon Eclipse E600/SPOT camera (Image solutions Inc., Preston, UK) and were processed using MetaVue (Molecular Devices, Wokingham, UK) software.

6.2.3 Goblet Cell Staining

Goblet cells were visualised with a combination of mucin stains. Sections were dewaxed and rehydrated as described above. Sections were then stained with 1% Alcian blue (Sigma-Aldrich) in 3% acetic acid pH = 2.5, for 5 minutes and then washed in dH_2O for 1 minute. Slides were then treated with 1% periodic acid (in dH_2O) for 5 minutes and then rinsed in dH_2O for 1 minute. Slides were then transferred to running tap water for 5 minutes and rinsed in dH_2O for 1 minute. Sections were then treated with Schiff's reagent (Scientific Laboratory Supplies) for 15 minutes and washed with running tap water for 5 minutes. Slides were counterstained with Mayer's Haematoxylin (Sigma-Aldrich) for 1 minute and 'blued' in running tap water. Sections were dehydrated and mounted as described in **Section 6.2.2**. Goblet cells appear as blue or magenta stained cells and at least 20 crypts or villi per mouse were counted for goblet cells and averaged. Slides were analysed in a blinded manner. Images were captured using a Nikon Eclipse E600/SPOT camera (Image solutions Inc) and were processed using MetaVue (Molecular Devices) software.

6.2.4 Colitis Scoring

Slides were scored out of 9 for signs of inflammation, as detailed in Table 2**.**

Table 2: Histological scoring system for mucosal inflammation

Active inflammation	Lamina propria cellularity	Surface ulceration
0: Normal	0: Normal	0: No ulceration with intact surface epithelium
1: Mild crypt distortion and loss and/or mild cryptitis (<5% of crypts infiltrated by neutrophils) with mild crypt abscess formation	1: Mild but unequivocal increase in mixed inflammatory cells	1: Probable erosion with focally stripped epithelium
2: Moderate crypt distortion and/or moderate cryptitis (<50% of crypts infiltrated by neutrophils) with mild crypt abscess formation	2: Moderate increase mixed inflammatory cells	2. Unequivocal erosion
3: Severe crypt distortion and loss with widespread and diffuse cryptitis (>50% of crypts involved) and diffuse goblet cell depletion	3. Severe and diffuse increase in inflammatory cells	3. Surface ulceration and granulation tissue formation

6.2.5 Fluorescence *in-situ* Hybridisation (FISH)

Sections were pre-warmed at 60ºC for 10 minutes, before incubation in two batches of xylene substitute solution for 10 minutes each (Sigma-Aldrich). Slides were then air dried until all liquid had evaporated from the sections. A Cy3-conjugated DNA probe specific for the 16S rRNA gene *(EUB338, 5'CY3-GCTGCCTCCCGTAGGAGT-3'*) (Eurofin Genomics, Ebersberg, Germany) was mixed with hybridisation solution (Tris HCL 20mM, pH 7.4; NaCl 0.9M; and 10% SDS) and applied to each section. From this point, slides were kept in the dark and incubated at 48ºC for one hour. Following hybridisation, slides were immediately placed into wash buffer (Tris HCL 20mM, pH 7.4; NaCl 0.9M) pre-warmed at 46ºC. After 15 minutes, slides were immersed in cold dH_2O (4ºC), and then washed in 3 batches of fresh PBS for two minutes each. Sections were then reduced using 10mM DTT (in Tris HCL 0.1M solution, Sigma-Aldrich for 30 minutes at room temperature. This was followed by an alkylation step, where 25mM iodacetamide (in Tris HCL 0.1M solution) was added to each section (30 minutes, RT). Following 3 washes in PBS, sections were blocked using 10% Goat Serum (Vector Laboratories, UK) in PBS, containing 1% BSA for 30 minutes. After blocking, MUC2 antibody was added to each section and incubated overnight at 4ºC.

Following overnight incubation, slides were washed in cold PBS (4ºC) 3 times before incubation with Alexa-488-conjugated goat anti-rabbit antibody (Life Technologies- supplied by ThermoFisher Scientific) at 2.5µg/ml for one hour at 4ºC. Slides were washed 3 times in cold PBS (4ºC), before mounting with VECTASHIELD Hardset Mounting Medium with DAPI (Vector Laboratories). Slides were imaged using a BX51 upright microscope and a Coolsnap EZ camera (Olympus, Tokyo, Japan) and images were processed using Image J. The thickness of the inner mucus barrier was quantified by measuring the distance between the epithelium and the outer mucus layer. Bacterial penetrance was also investigated by assigning values from 0-4 depending on where bacteria were localised: 0 = bacteria in the lumen and outer mucus layer, 1 = bacteria in the inner mucus layer, 2 = bacteria in contact with the epithelium, 3 = bacteria in the crypts, 4 = bacteria in the lamina propria. All slides were scored blinded.

6.3 Enzyme-linked immunosorbent Assays (ELISAs)

6.3.1 IgA ELISA

An ELISA was performed using the Invitrogen™ eBioscience™ Mouse IgA ELISA Ready-SET-Go!™ Kit (Invitrogen, UK), according to manufacturer's instructions. In brief, Corning

Costar 9018 ELISA plates (Corning, Amsterdam, The Netherlands) were coated with capture IgA antibody in 1x PBS at a 1:250 dilution overnight. Wells were washed twice with PBS containing 0.5% Tween (Sigma-Aldrich) and blocked with the blocking buffer provided for 2 hours at room temperature. The provided standard at 50ng/ml was serially diluted 2-fold until a minimum concentration of 1.56ng/ml was achieved. Samples were diluted 1:1000 in the provided Assay buffer and 100µL of sample was added to the plate in duplicate. Samples were incubated on a shaker at 400rpm for 2 hours. Wells were washed 4 times before addition of the provided detection antibody. TMB provided in the kit was added and incubated for 15 minutes before quenching with H_2SO_4. The ELISA was read using a VersaMax microplate reader (Molecular Devices) at 450nm, with wavelength subtraction at 570nm.

6.3.2 Albumin ELISA

Faecal samples were mixed with PBS (Sigma-Aldrich) at 100µL per 10mg of faeces and homogenised. Samples were centrifuged at 200xg for 5 minutes to pellet debris, the supernatant was collected and centrifuged at 8000xg for 5 minutes to pellet the bacteria. The supernatant was collected and stored at -80ºC. Albumin levels in serum diluted 1/100 were assessed using a Mouse Albumin ELISA Kit (Universal Biologicals Ltd, Cambridge, UK) according to manufacturer's instructions as described in **Section 6.3.1**.

6.4 Analysis of Microbial Communities via DGGE

6.4.1 16S DNA extraction, amplification and purification

DNA extraction was performed using a Qiamp® Fast Stool DNA Mini Kit (Qiagen, Manchester, UK), using a modified version of the manufacturer's instructions. Faecal samples were incubated in Inhibitex buffer (Qiagen) and mechanically disaggregated, before incubation at 95ºC for 30 minutes. Mucus samples were centrifuged (13000xg for 10 minutes) and the mucus pellets incubated in Inhibitex buffer, mechanically disaggregated and incubated at 95ºC for 30 minutes. 300uL of the resulting lysate was used for the subsequent steps, which were then performed according to manufacturer's instructions. Optical density at 260nm was recorded using a UV1101 spectrophotometer (Biochrom Ltd., Cambridge, UK) to measure DNA concentration.

For the identification of different bacterial species, the 16S rRNA gene was amplified using the universal primer pairs P3_GC-341F and P2_518 (Table 2). For faecal samples, 100ng of DNA was added into each PCR reaction. For mucus samples, 200ng of DNA was added into

each PCR reaction. The PCR reaction contained: 2mM $MgCl_2$, 200μM of dNTP mix, 0.4μM of each primer pair and 2U of FastStart Taq DNA polymerase (All provided by Roche, UK). Bovine Serum Albumin (BSA) (0.1μg/μL, New England Biolabs) was also added to the PCR mixture. The PCR conditions used are summarised as follows: denaturation/ polymerase activation at 95˚C for 5 minutes, followed by 30 cycles of denaturation at 95˚C for 1 minute, annealing at 55˚C for 1 minute, elongation at 72˚C for 1 minute, and a final elongation step at 72˚ for 10 minutes.

Following successful amplification, DNA was purified using a MinElute Purification Kit (Qiagen) according to manufacturer's instructions and DNA yield was determined through measurement of optical density at 260nm using a UV1101 spectrophotometer. DNA was then stored at -20ºC until further use.

Table 3: Primers used to amplify bacteria for denaturing gel electrophoresis (DGGE)

Gene	Primers	Sequence
16S rRNA	P3_GC-341F	5'-CGCCCGCCGCGCGCGGCGGGCGGGGCG GGGGCACGGGGGGCCTACGGGAGGCAGCAG-3'
	P2_518R	5'-ATTACCGCGGCTGCTGG-3'

6.4.2 Denaturing gradient gel electrophoresis (DGGE)

A denaturing gradient gel was prepared according to the methods initially developed by Fischer and Lerman (Fischer and Lerman, 1983). A non-denaturing solution that contained 40% acrylamide/bis-acrylamide solution (Fischer Scientific), 50x TAE (Bio-Rad, Hartfordshire, UK) and dH_2O (0% denaturing solution) was prepared. A second solution that contained the same reagents, but also included 40% formamide (Sigma-Aldrich) and 7M urea (Sigma-Aldrich) as denaturing agents (100% denaturing solution) was also prepared. Both solutions were degassed under a vacuum, mixed and then 10% ammonium persulfate (Bio-Rad) and *N*,*N*,*N′*,*N′*-Tetramethylethylenediamine (TEMED) (Bio-Rad) were added to facilitate polymerisation of the gel. A gradient maker (Hoefer SG-50) was used to generate the gel gradient, following the addition of the mixed denaturing solutions. 150ng of purified DNA per sample (faecal and mucus) was loaded onto the gel. The DGGE gel was run at 60˚C for 16 hours and at 63V in 1x TAE buffer using the DCODE Universal Mutation Detection System (Bio-Rad). Gels were stained with SYBR Gold nucleic acid stain (ThermoFisher Scientific UK) for 30 min and gel bands were visualised and imaged using a Bio-Doc-It® Imaging System.

The gel was then analysed using Phoretix Software (TOTALLAB, Newcastle Upon Tyne, UK). Lane boundaries were defined to correct for any potential distortions during the gel run and manually curated to ensure that the bands detected were not artefacts. Reference bands were selected to align bands across the gel and 'Rf values' generated to measure the bands migration. A binary matrix was then generated based on the Rf values, with 0's and 1's indicating the absence or the presence of a bacterial 'species' in a sample. This matrix was processed in R. Specifically, the 'VEGAN' (Oksanen et al., 2016), 'ecodist' (Goslee and Urban, 2007) and 'MASS' (Venables and Ripley, 2002) packages were used to statistically compare the presence or absence of species (bands) between groups, e.g. WT lumen, WT mucus, *ΔdblGATA-1*$^{-/-}$ lumen and *ΔdblGATA-1*$^{-/-}$ mucus. The same packages were used to perform non-metric multidimensional scaling to examine clustering between groups. Finally, bacterial diversity (total number of bands) was calculated using the Shannon-Wiener Diversity Index.

6.4.3 qPCR Analysis of Bacteria

Stool, colonic mucus and small intestinal mucus DNA was used as a template for a qPCR reaction that consisted of 100ng/µL DNA, 10uM primers, 2x KAPA SYBR® FAST qPCR Mastermix and nuclease free water to a total volume of 20µL. Primers used are illustrated in Table 3. A universal 16S rRNA gene was used as a housekeeping control. Results were calculated using the ΔΔCT method. Cycling steps for all primers were: Denaturing at 95 ℃ for 10 minutes, followed by 35 cycles of 95 ℃ for 15 seconds and 60 ℃ for 1 minute.

Table 4: List of primers used for qPCR

Gene	Primers	Sequence
Universal	UniF340	5’-ACTCCTACGGGAGGCAGCAGT-3’
	UniR514	5’-ATTACCGCGGCTGCTGGC-3’
Actinobacteria	Act664F	TGTAGCGGTGGAATGCGC
	Act941R	AATTAAGCCACATGCTCCGCT
Bacteriodetes	Bac960F	5’-GTTTAATTCGATGATACGCGAG-3’
	Bac1100R	TTAASCCGACACCTCACGG
Deferribacteres	Defer1115F	CTATTTCCAGTTGCTAACGG
	Defer1265R	GAGHTGCTTCCCTCTGATTATG
Firmicutes	Firm934F	5’-GGAGYATGTGGTTTAATTCGAAGCA-3’
	Firm934R	5’-AGCTGACGACAACCATGCAC-3’
Verrucomicrobia	Ver1165F	5’-TCAKGTCAGTATGGCCCTTAT-3’
	Ver1263R	5’-GAGHTGCTTCCCTCTGATTATG-3’
Bacteroides	BacteroidesF	5’-GGTTCTGAGAGGAGGTCCC-3’
	BacteroidesR	5’-GCCTCAAGGGCACAACCTCCAAG-3’
Clostridiales	ClostF	ACTCCTACGGGAGGCAGC
	ClostR	GCTTCTTAGTCAGGTACCGTCAT
Enterrobacteriaceae	Uni515F	GTGCCAGCAGCCGCGGTAA
	Ent826R	GCCTCAAGGGCACAACCTCCAAG
Lachnospiraceae/Rumminococceae	LachnoRumF	CGGTACCTGACTAAGAAGC
	LachnoRumR	AGTTTCATTCTTGCGAACG
Akkermansia muciniphila	Amuc_1599F	GACCGGCATGTTCAAGCAGACT
	Amuc_1599R	AAGCCGCATTGGGATTATTTGTT
Segmented Filamentous Bacteria	SFBF	GACGCTGAGGCATGAGAGCAT
	SFBR	GACGGCACGGATTGTTATTCA

6.5 Analysis of Microbial Communities via 16S rRNA Sequencing

6.5.1 16S rRNA gene sequencing processing

16S amplicon sequencing targeting the V3 and V4 variable regions of the 16S rRNA (Table 5) was performed on the Illumina MiSeq platform (Illumina, California, USA) according to manufacturer's guidelines and generated paired-end reads of 300bp in each direction. Illumina reads were demultiplexed to remove adapter sequences and trim primers. Illumina paired-end reads were merged together using SeqPrep (StJohn, 2018) and submitted to MG-RAST's metagenomics pipeline (Meyer et al., 2008). Reads were pre-processed to remove low-quality and uninformative reads using SolexQA (Cox et al., 2010). The quality-filtering process included removal of reads with low quality ends (i.e. ambiguous leading/trailing bases) and the removal of reads with a read length two standard deviations below the mean. Artificial duplicate reads were then removed based on MG-RAST's pipeline.

The resulting FASTQ files for every sample were merged into a single file of 590822 sequences to simplify processing, manually adding 3 known Archaeal 16S rRNA sequences from *Acidilobus saccharovorans*, *Sulfolobus tokodaii* and *Methanobrevibacter smithii*. Sequences were aligned using a specialist 16S RNA aligner using the Infernal algorithm (Nawrocki and Eddy, 2013), via a web-based interface provided by the Ribosomal Database Project (Cole et al., 2014). This file was then manually curated in R (Team, 2016). . Unless otherwise stated, all analyses were performed using custom scripts in R. The number of aligned bases in each sequence was recorded and the distribution of continuously aligned bases was examined. Any sequence that had less than 437 continuously aligned bases was discarded. The remaining 496550 sequences were taken forward for analysis. All sequences were identified using BLAST+ and the top hit for each sequence was recorded (Camacho et al., 2009). The 'classification' function in the 'taxize' R package (Chamberlain and Szocs, 2013) was then used to assign full taxonomic information to each identified taxa where possible.

Table 5: Primers used for next generation sequencing

16S rRNA Sequencing Primers: V3-V4 (containing adapters)	F	5’-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3’
	R	5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3'

6.5.2 Phylogenetic Tree

A phylogenetic tree of all sequences was generated using FastTree (Price et al., 2010), using the general time reversible (GTR) + CAT model and default parameters. The tree was rooted using the archaeal sequences as an outgroup. Phylogenetic clades were obtained using the 'Ancestor' function in the 'phangorn' R package (Schliep, 2011). A relative abundance matrix, with abundance based on how many times sequences belonging to a phylogenetic clade appeared in a sample, was calculated.

6.5.3 Ordination

Bray-Curtis dissimilarity values were calculated among all samples (based on the relative abundance matrix) and used for non-metric multidimensional scaling (NMDS) via the 'MASS' (Venables and Ripley, 2002) and 'ecodist' R packages (Goslee and Urban, 2007). Significance of clustering was determined using permutational multivariate analysis via the 'Adonis' function in the 'vegan' R package (Oksanen et al., 2016).

6.5.4 Machine Learning

Random Forest (RF) models were run using the 'randomForest' package (Liaw and Wiener, 2002) in R. Specifically, the relative abundance matrix was used as an input for the RF, using a forest of 100,000 trees and the mtry value was left at default settings. 100000 trees gave close to a maximally accurate association of clades and importance (Supplementary Figure 2), minimising variability in importance values estimated. Separate forests were run to predict whether a sample was 6 or 18 weeks old, whether a sample was stool or mucus, whether it was a WT or an *mdr1a*$^{-/-}$ sample, what cage the sample was taken from, the mother of the offspring and whether it could discriminate between combinations of these treatment groups. The 'MeanDecreaseAccuracy' (MDA) value was used as a measure of how important each node was at predicting treatment information and the out-of-bag (OOB) error rate was used to determine the predictive accuracy of the model. Nodes were ranked based on MDA value, taking the five most important nodes, determining the descendant tips and confirming the identity of the tip sequences via the BLAST+ results (Camacho et al., 2009). Additionally, the depth of each node was determined using the 'distances' function in the igraph R package (Csard and Nepusz, 2006). A phylogenetic tree annotated with the resulting information was plotted using the 'plot.phylo' function in the 'ape' package (Paradis et al., 2004).

6.5.5 Validation of the Model

In order to validate each model, we included a 'randomised' negative control RF where relative abundances of each node were permuted with respect to each sample and the predictive accuracy was assessed. In addition, we also took the relative abundances of an important node for age and redistributed the abundance to only WT samples. The RF was repeated to investigate whether this node would appear as important for genotype. We also ran RF's with an increasing number of trees and performed Spearman's Rank correlation on the MDA values obtained between each RF. The Michaelis-Menten model was used to elucidate how an increasing number of trees affected correlation of the MDA values.

6.6 Metabolite Profiling Using LC-MS

Mucus homogenisation and subsequent metabolite extraction was undertaken using a Tissuelyser II (Qiagen). The homogenisation solvent (1:1 chloroform:methanol, 800 µL pre-chilled to -20 °C) was added to each sample, a steel bead added and then homogenised for 20 min at 25 Hz. Once homogenised, 400 µL of water was added and the sample vortex mixed for 15 s. To initiate phase separation the samples were then centrifuged (8000 xg for 10 min) before the organic fraction was collected LC-MS analysis. The organic fraction was subsequently vacuum concentrated (Eppendorf Vacuum concentrator, RT, 8 h) and stored at -20 °C until analysis. Upon analysis, the metabolite extracts were resuspended in methanol (normalised to dry-weight) and analysed by LC-MS. A portion of each sample was pooled to give a representative pooled biological 'quality control' (QC) which can be used to assess instrumental variation during the analytical run (Broadhurst et al., 2018).

All LC-MS analysis was conducted on a QExactive Plus equipped with an Ultimate 3000 UHPLC (Thermo, UK). The UHPLC was equipped with a Hypersil Gold reverse phase column (C18 -2.1 mm x 150 mm; 1.9 µm particle size). The solvents employed were (A) water + 0.1% formic acid and (B) methanol + 0.1 % formic acid. The flow gradient was programmed to equilibrate at 95% A for 2 min followed by a linear gradient to 95% B over 8 min and held at 95% B for 2 min before returning to 95% A for 2 min. The column was maintained at 40 °C and the samples chilled in the autosampler at 4 °C. A sample volume of 5 µL was injected onto the column with a constant flow rate of 400 µL/min. Blank injections were analysed at the start and end of the analytical batch to assess the background and carryover. In addition pooled QC samples (as above) were analysed at every 6th injection to assess for analytical drift over time. Data acquisition was conducted in full MS mode in the

scan range of 70-1050 m/z with a resolution of 70,000, an AGC target of $3e^6$ and a maximum integration time of 200 ms. The samples were analysed in positive and negative mode in separate acquisitions.

The raw data was converted to mzXML file format (Proteowizard) and XCMS was used to deconvolve the data. The data underwent QC-based filtering, where any feature that was missing in more than 50% of QCs was removed. All features within a given sample were normalised to account for variation that may arise between sample injections. PCA's were plotted from the processed data using R (R Core Team, 2018) and the ggfortify package (Tang et al., 2016). An RF model was run using the randomForest package in R (Liaw and Wiener, 2002), with the default mtry value and 100000 trees. The relative peak area distribution for the top 4 metabolites identified by the RF, based on their MDA value, were plotted. The m/z ratio for each metabolite feature was used to assign putative identity based on comparisons to databases such as the Kyoto Encyclopaedia for Genes and Genomes (KEGG) and the Human Metabolome Database (HMDB).

6.7 RNA-seq

Proximal colon snips were taken from 6 week old WT and *mdr1a*$^{-/-}$ mice, stored in RNA-Later (ThermoFisher Scientic) and stored at -80ºC until use. RNA was then extracted using the RNeasy Mini Kit (Qiagen). Samples were diluted to 1µg/µL concentration and rRNA depleted using a Ribo-Zero Gold rRNA Removal Kit (Illumina). Samples were diluted to 1ng/µL concentration for RNA-seq. RNA-seq was performed using the HiSeq4000 (Illumina, California, USA). Data was analysed using DESEQ2 in R. Specifically, unmapped paired-end sequences from an Illumina HiSeq4000 sequencer were tested by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequence adapters were removed and reads were quality trimmed using Trimmomatic_0.36 (Bolger et al., 2014). The reads were mapped against the reference mouse genome (mm10/GRCm38) and counts per gene were calculated using annotation from GENCODE M14 (http://www.gencodegenes.org/) using STAR_2.4.2 (Dobin et al., 2013). Normalisation, Principal Components Analysis, and differential expression was calculated with DESeq2_1.16.1 (Love et al., 2014). A heatmap was produced in R using the Gplots package (Gregory et al., 2016).

6.8 Statistical Analysis

All statistical analysis was performed using either GraphPad Prism 7 (GraphPad Software, La Jolla, USA) or R. Student's T Test was used to compare goblet cell number, inner mucus thickness and bacterial localisation between genotype. Man Whitney U Tests with false discovery rate correction for multiple testing was used to analyse relative peak area distribution between genotype. Permutational multivariate analysis of variance (PERMANOVA) was used to calculate differences in either the relative abundance of species between samples, or presence/absence of species in samples using the 'adonis' function in R. It was also used to calculate differences in overall metabolites and transcription. Relative expression of bacteria quantified via qPCR were assessed for significant differences using a 2 Way ANOVA with a Sidak's Post Hoc Test.

6.9 References

Bolger, A. M., Lohse, M. & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30(15)**,** 2114-20.

Broadhurst, D., Goodacre, R., Reinke, S. N., Kuligowski, J., Wilson, I. D., Lewis, M. R. & Dunn, W. B. (2018). Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. *Metabolomics,* 14(6)**,** 72.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.

Chamberlain, S. A. & Szocs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Res,* 2**,** 191.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R. & Tiedje, J. M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res,* 42(Database issue)**,** D633-42.

Cox, M. P., Peterson, D. A. & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics,* 11**,** 485.

Csard, G. & Nepusz, T. (2006). *The igraph software package for complex network research* [Online]. Available: http://igraph.org [Accessed 31st August 2018].

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* 29(1)**,** 15-21.

Fischer, S. G. & Lerman, L. S. (1983). DNA fragments differing by single base-pair substitutions are separated in denaturing gradient gels: correspondence with melting theory. *Proc Natl Acad Sci U S A,* 80(6)**,** 1579-83.

Goslee, S. C. & Urban, D. L. (2007). The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *22,* 22(7)**,** 1-19.

Gregory, R., Bolker, B., Warnes, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M. & Venables, B. (2016). *gplots: Various R Programming Tools for Plotting Data* [Online]. Available: https://CRAN.R-project.org/package=gplots [Accessed 14th August 2018].

Liaw, A. & Wiener, M. (2002). Classification and Regression by randomForest. *R News,* 2(3)**,** 18-22.

Love, M. I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15(12)**,** 550.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J. & Edwards, R. A. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics,* 9**,** 386.

Nawrocki, E. P. & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics,* 29(22)**,** 2933-5.

Oksanen, J., Guillaume-Blanchet, F., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H. & Wagner, H. (2016). *vegan: Community Ecology Package* [Online]. Available: https://CRAN.R-project.org/package=vegan [Accessed 27th October 2017].

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics,* 20**,** 289-290.

Price, M. N., Dehal, P. S. & Arkin, A. P. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One,* 5(3)**,** e9490.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing* [Online].Available: https://www.r-project.org/ [Accessed 16th January 2017].

Schinkel, A. H., Smit, J. J., van Tellingen, O., Beijnen, J. H., Wagenaar, E., van Deemter, L., Mol, C. A., van der Valk, M. A., Robanus-Maandag, E. C., te Riele, H. P. & et al. (1994). Disruption of the mouse mdr1a P-glycoprotein gene leads to a deficiency in the blood-brain barrier and to increased sensitivity to drugs. *Cell,* 77**,** 491-502.

Schliep, K. P. (2011). phangorn: phylogenetic analysis in R. *Bioinformatics,* 27(4)**,** 592-593.

Shivdasani, R. A., Fujiwara, Y., McDevitt, M. A. & Orkin, S. H. (1997). A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *Embo j,* 16**,** 3965-73.

StJohn, J. (2018). *SeqPrep* [Online]. Available: https://github.com/jstjohn/SeqPrep [Accessed 1st August 2018].

Tang, Y., Horikoshi, M. & Li, W. (2016). ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. *The R Journal,* 8.

Venables, W. & Ripley, N. (2002). *Modern Applied Statistics with S.* New York: Fourth Edition.

Chapter Seven

Appendices

7. Appendices

7.1 Appendix 1

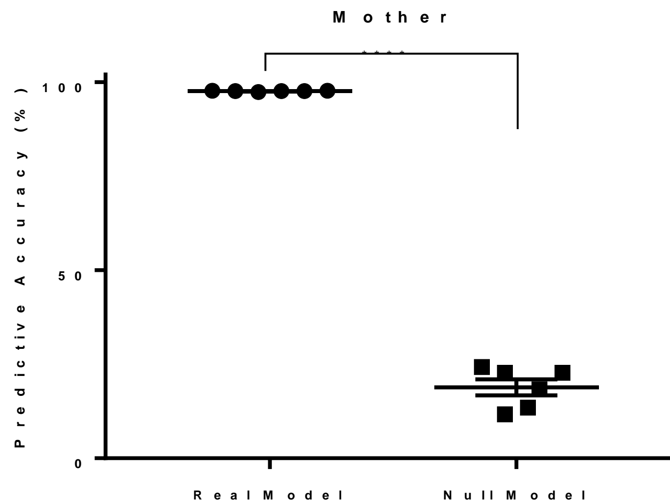


Figure 1: The microbiota of an individual can be linked back to its mother, irrespective of cage. 16S rRNA sequence data derived from the stools and mucus of 6 and 18 week old mice (FVB background, wildtype vs *mdr1a*$^{-/-}$ mice) were used to construct a phylogenetic tree. A Random Forest (RF) model of the relative abundance of phylogenetic clades was used to find associations between the gut microbiota and different experimental treatments. The predictive accuracy of the RF model at taking a sample and discriminating between the mother of an offspring is illustrated. Asterisks represent significance determined using Student's T Test: p =< 0.00001 (****).

7.2 Appendix 2

7.2.1 List of Author Publications

Glymenaki, M., **Singh, G**., Brass, A., Warhurst, G., McBain, A. J., Else, K. J. & Cruickshank, S. M. (2017b). Compositional Changes in the Gut Mucus Microbiota Precede the Onset of Colitis-Induced Inflammation. Inflamm Bowel Dis, 23(6), 912-922.

Williams, H., Campbell, L., Crompton, R. A., **Singh, G**., McHugh, B. J., Davidson, D. J., McBain, A. J., Cruickshank, S. M. & Hardman, M. J. (2018). Microbial Host Interactions and Impaired Wound Healing in Mice and Humans: Defining a Role for BD14 and NOD2. J Invest Dermatol.

Williams, H., Crompton, R. A., Thomason, H. A., Campbell, L., **Singh, G**., McBain, A. J., Cruickshank, S. M. & Hardman, M. J. (2017). Cutaneous Nod2 Expression Regulates the Skin Microbiome and Wound Healing in a Murine Model. J Invest Dermatol, 137(11), 2427-2436.

7.2.2 Submission Pending

Singh, G., Brass, A., Cruickshank, S.M. and Knight, C.G. Gut microbial taxa at distinct phylogenetic scales distinguish host age, social group and microbial niche, but not genotype, in a murine model of inflammatory bowel disease. *(To be submitted to Microbiome, pending formatting)*

Singh, G., Brass, A., Knight, C.G. and Cruickshank, S.M. Gut eosinophils and their impact on the mucus-dwelling microbiota *(Submitted to Immunology)*