

**Elucidating the role of the mucus-associated microbiota and  
mucin glycosylation in inflammatory bowel disease**

**Elizabeth Thursby, BSc (Hons)**

A thesis submitted in accordance with the requirements for the degree of Doctor of  
Philosophy (PhD)

To

**The University of East Anglia**

Institute of Food Research  
Gut Health and Food Safety  
Norwich Research Park  
Colney  
NR4 7UA

**September 2016**

## Abstract

The human gastrointestinal (GI) tract is host to a dynamic community of  $10^{13}$ - $10^{14}$  bacteria, which mainly reside in the colonic lumen and outer mucus layer covering the GI tract. Mucins are decorated in a diverse array of O-glycans, providing nutrients and attachment sites for microbes. Dysbiosis of the microbiota, and alterations in mucin glycosylation have been associated with Inflammatory Bowel Disease (IBD). However, the causal relationship between these two factors remains unclear. Here, we employed a multidisciplinary approach to address this relationship, and the molecular mechanisms mediating these changes. Mucosal lavages and biopsies obtained from the sigmoid and ascending colon of patients were used to assess alterations in the mucosal microbiota and the glycosylation of associated mucus in ulcerative colitis (UC) patients. Secondly, *in vitro* growth assays and gnotobiotic mouse experiments were performed to investigate the reciprocal role of mucin-degrading bacteria in the modulation of mucin O-glycosylation.

These analyses highlighted inter-patient variability, but a similar microbial composition between colonic sites. In contrast, mucin glycosylation and the expression of glycosyltransferases was regio-specific. In UC, changes in the abundances of bacterial groups, including a decrease in the *A. muciniphila* to *R. gnavus* ratio were apparent. UC mucins displayed a decrease in fucosylation, increase in sialylation, and a decrease in many complex glycan structures found in abundance in controls. *In vitro* growth assays suggested that UC-like mucin glycosylation impaired *A. muciniphila* growth, whilst *R. gnavus* remained unaffected, potentially explaining changes in these species in UC. Furthermore, gnotobiotic mouse experiments showed that *A. muciniphila* and *R. gnavus* were able to remodel mucin glycosylation.

Our findings suggest a multifactorial dysregulation at the epithelial interface in IBD, where mucus-associated microbiota and mucin glycosylation are interdependent. It is likely that an initial disruption in either of these components drives alterations in other mucosal constituents, propagating disease and exacerbating inflammation.

## Preface

### Oral presentations:

**Thursby E., Kavanaugh D., Crost E. H., Tailford L. E., Gunning P., Tremelling M., Watson A., Juge N.** *Exploring the relationship between mucolytic bacteria and mucin glycosylation in health and disease.* Institute of Food Research Science Symposium, John Innes Conference Centre, Norwich, UK, 28<sup>th</sup> April 2015.

**Thursby E., Kavanaugh D., Crost E. H., Tailford L. E., Gunning P., Tremelling M., Watson A., Juge N.** *Unravelling the relationship between mucins and mucolytic bacteria during intestinal inflammation Conference.* GUT and Liver Inflammation: A Translational Science Masterclass, British Society of Gastroenterology Annual Meeting, Manchester Central, Manchester, UK, 19<sup>th</sup> June 2014.

### Poster presentations:

**Thursby E., Gunning P., Kavanaugh D., Crost E. H., Tailford L. E., Tremelling M., Watson A., Juge N.** *Unravelling the relationship between mucolytic bacteria and mucin glycosylation in health and disease.* Annual IFR Student Science Showcase, Institute of Food Research, Norwich, UK, 16<sup>th</sup> September 2015.

**Thursby E., Gunning P., Kavanaugh D., Crost E. H., Tailford L. E., Tremelling M., Watson A., Juge N.** *Unravelling the relationship between mucolytic bacteria and mucin glycosylation in health and disease.* Mucins in Health and Disease, 13th International Workshop on Carcinoma Associated Mucins, Robinson College, Cambridge, UK, 19<sup>th</sup> July 2015.

**Kavanaugh D., Reyes Martinez J., Etzold S., Kober O., Latousakis D., Thursby E., MacKenzie D., Flitsch S., Juge N.** *Mechanistic Insights into Bacterial Adhesion to Gastrointestinal Mucus.* Society for General Microbiology. Irish Division Meeting: Microbial Interfaces, Galway, Ireland, 17<sup>th</sup> June 2015

**Gunning A. P., Kirby A. R., Fuell C., Pin C., Tailford L. E., Rigby N. M., Morris V. J., Barker G., Thursby E., Juge N.** *Cracking the mucin glycode by force.* Institute of Food Research Science Symposium, John Innes Conference Centre, Norwich, UK, 28<sup>th</sup> April 2015.

**Thursby E., Kavanaugh D., Crost E. H., Tailford L. E., Gunning P., Tremelling M., Watson A., Juge N.** *Unravelling the Relationship Between Mucins and Mucolytic Bacteria During Intestinal Inflammation.* International Scientific Conference on Probiotics and Prebiotics, Hotel InterContinental, Budapest, Hungary, 24<sup>th</sup> June 2014.

**Thursby E., Kavanaugh D., Crost E. H., Tailford L. E., Gunning P., Tremelling M., Watson A., Juge N.** *Investigating the impact of intestinal mucin changes in Inflammatory Bowel Disease.* Annual IFR Student Science Showcase, Institute of Food Research, Norwich, UK, 23<sup>rd</sup> June 2014

**Thursby E., Tailford L. E., Crost E. H., Kavanaugh D., Gunning A. P., Tremelling M., Watson A., Juge N.** *Investigating the impact of intestinal mucin changes in Inflammatory Bowel Disease.* Inaugural annual symposium, Gut Health and Food Safety ISP, John Innes Conference centre, Norwich, UK, 18<sup>th</sup> March 2014

**Thursby E., Tailford L. E., MacKenzie D., Watson A., Juge N.** *Deciphering the role of mucins and mucolytic bacteria during intestinal inflammation.* Norwich Biosciences Institutes Annual Meeting, John Innes Conference Centre, Norwich, UK, 16<sup>th</sup> October 2013

**Thursby E., Kober O., Tailford L. E., MacKenzie D., Watson A., Juge N.** *Deciphering the role of mucins and mucolytic bacteria in Inflammatory Bowel Disease.* Annual IFR Student Science Showcase, Institute of Food Research, Norwich, UK, 23<sup>rd</sup> May 2013.

## Acknowledgements

This PhD was supported by a four-year Norwich Research Park studentship award.

My biggest gratitude goes to my supervisory team Dr. Nathalie Juge, Prof. Alastair Watson, Dr. Arjan Narbad and Dr. David Swarbreck for their guidance, support and feedback. In particular, I would like to thank Nathalie for her never ending patience and trust in me throughout the project.

I would also like to thank Dr. Stuart Haslam and Dr. Gary Rowley for agreeing to read and discuss my work.

My acknowledgements go to the Gastroenterology department at the Norfolk and Norwich University Hospital, especially Dr. Mark Tremelling, Dr. Simon Chan and Dr. Jo Brooks for their dedication in assisting with my sample collection, even when they had extremely busy schedules themselves.

At IFR, my sincere thanks goes to many colleagues and collaborators who supported me with their expertise, knowledge and advice. In particular, my appreciation goes to Lindsay Hall, Jennifer Ketskemety, Suparna Mitra, Patrick Gunning, Neil Rigby, Isabelle Hautefort, Andrew Goldson, Arlaine Brion, Henri Tapp and Gwenaelle Le Gall. A big thank you also goes to the many past and present members of the Juge lab, of which a large number have assisted in the project, however particular recognition goes to Karine Lecointe, Emmanuelle Crost, Louise Tailford and Sandra Tribolo.

To my family and friends, from home, from Sheffield, and from Norwich, thank you for sticking by me and providing encouragement when I needed it most.

A special mention goes to Ben, because although at times it was hard for both of us, he provided laughter, reassured me, and allowed me to be myself. Thank you for being there through thick and thin.

Finally, I could not have arrived at where I am today without the ongoing love of my parents, grandparents and sister, who have endlessly supported me in every decision I have made. For this, I will be forever grateful.

## Contents

ABSTRACT .....	2
PREFACE .....	3
ACKNOWLEDGEMENTS .....	5
LIST OF FIGURES.....	10
LIST OF TABLES.....	13
ABBREVIATIONS.....	14
CHAPTER 1) INTRODUCTION .....	19
1.1 Overview of the human gastrointestinal tract .....	19
1.1.1 General structure and function .....	19
1.1.2 The intestinal epithelium .....	21
1.2 The GI mucus layer.....	22
1.2.1 Structure and organisation of GI mucus .....	22
1.2.2 Mucus composition .....	23
1.2.2.1 Mucin proteins .....	23
1.2.2.2 Biosynthesis and molecular properties of mucins .....	25
1.2.2.3 Mucin glycosylation .....	30
1.2.3 Other mucus components .....	35
1.2.4 Mechanisms of mucus regulation .....	36
1.2.5 Functional role of mucus in the GI tract .....	37
1.2.5.1 Mucus as a physical barrier.....	37
1.2.5.2 Mucus as a biological habitat .....	38
1.2.5.3 Mucus as an immune mediator .....	39
1.3 The GI microbiota .....	40
1.3.1 Development and composition of the human GI microbiota .....	40
1.3.1.1 Development of the GI microbiota from childhood .....	40
1.3.1.2 Composition and variability of the human GI microbiota.....	42
1.3.1.3 Factors shaping the GI microbiota.....	44
1.3.2 Role of the gut microbiota in health .....	47
1.3.2.1 Beneficial role of microbes .....	47
1.3.2.2 Dysbiosis and disease .....	48
1.4 Inflammatory Bowel Disease (IBD) .....	51
1.4.1 Definition and epidemiology of IBD .....	51
1.4.2 Aetiology of IBD.....	51
1.4.2.1 Role of genetics in IBD susceptibility.....	52
1.4.2.2 Role of the GI microbiota in IBD.....	54
1.4.2.3 Impact of the immune system in IBD.....	59
1.4.2.4 Involvement of environmental factors in IBD.....	61
1.4.3 Role of mucus and mucins in IBD .....	62
1.5 Aim and Objectives .....	66
CHAPTER 2) MATERIALS AND METHODS.....	67
2.1 Recruitment of patients and collection of samples.....	67
2.2 Initial processing and storage of human samples .....	68

<b>2.3 Methods for microbiota analysis from human mucosal lavage samples .....</b>	<b>68</b>
2.3.1 Extraction of bacterial DNA from lavage .....	68
2.3.2 qPCR assays of bacterial DNA from lavage .....	69
2.3.3 16S sequencing of bacterial DNA from human mucosal lavage .....	73
2.3.3.1 Preparation of 16S DNA libraries .....	73
2.3.3.2 Bioinformatics of 16S sequencing data .....	76
<b>2.4 Methods for mucin glycosylation analysis from human mucosal lavage samples and mucosal biopsies.....</b>	<b>77</b>
2.4.1 Purification of mucins from mucosal lavage samples .....	77
2.4.2 Composite AgPAGE of purified mucins .....	77
2.4.2.1 Composite AgPAGE gel casting and loading .....	77
2.4.2.2 Composite gel staining .....	78
2.4.2.3 Transfer of mucins to PVDF membrane.....	78
2.4.2.4 Alcian blue staining of PVDF membranes.....	78
2.4.3 Dot blot analysis of purified human mucins .....	79
2.4.3.1 Lectin probing of dot blots .....	79
2.4.4 AFM for visualisation of mucins .....	81
2.4.5 Force spectroscopy .....	81
2.4.5.1 Peak distance analysis .....	82
2.4.6 HPAEC-PAD for sialic acid quantification .....	82
2.4.7 MALDI-TOF analysis of mucin glycans from mucus samples .....	84
2.4.8 RNA extraction and cDNA synthesis from human mucosal biopsies.....	85
2.4.9 qRT-PCR of glycosyltransferase genes .....	85
<b>2.5 Methods for <i>in vitro</i> culture assays of <i>R. gnavus</i> CC55_001C and <i>A. muciniphila</i> ATCC BAA 835 .....</b>	<b>88</b>
2.5.1 Routine culture conditions .....	88
2.5.2 Growth assays in carbohydrate supplemented media.....	88
2.5.3 Purification of Sigma type III PGM.....	89
2.5.4 Purification of mucin from the LS174T cell line.....	89
2.5.5 Growth assays of <i>A. muciniphila</i> and <i>R. gnavus</i> in mucin medium .....	90
2.5.6 Monitoring <i>in vitro</i> growth of <i>A. muciniphila</i> and <i>R. gnavus</i> by qPCR.....	91
2.5.7 <sup>1</sup> H NMR metabolite analysis of spent media.....	92
<b>2.6 Methods for <i>In vivo</i> colonisation assays of germ-free mice with <i>R. gnavus</i> CC55_001C and <i>A. muciniphila</i> ATCC BAA 835 .....</b>	<b>93</b>
2.6.1 Gavaging of germ-free mice and collection of samples .....	93
2.6.2 Monitoring colonisation of <i>A. muciniphila</i> and <i>R. gnavus</i> by qPCR.....	93
2.6.3 Purification of mucins from mouse mucus scrapings and analysis of glycosylation .....	94
2.6.4 <sup>1</sup> H NMR metabolite analysis of caecal contents .....	94
<b>2.7 Statistical analyses .....</b>	<b>95</b>
<b>CHAPTER 3) CHARACTERISATION OF THE COLONIC MUCUS-ASSOCIATED MICROBIOTA.....</b>	<b>96</b>
<b>3.1 Introduction and objectives .....</b>	<b>96</b>
<b>3.2 Overall profiling of the mucus-associated microbiota by 16S sequencing.....</b>	<b>97</b>
3.2.1 Effect of confounding variables on taxonomic profile .....	97
3.2.2 Diversity, richness and evenness of the microbiota from mucosal lavage samples.....	100

3.3 Quantitation of bacterial groups of interest by qPCR .....	102
3.3.1 Effect of confounding variables on abundance of bacterial groups.....	102
3.3.2 Spatial variation in abundance of bacterial groups .....	104
3.3.3 Effect of disease state on abundance of bacterial groups.....	106
3.4 Discussion .....	110
<b>CHAPTER 4) CHARACTERISATION OF MUCIN GLYCOSYLATION FROM HUMAN COLONIC LAVAGES.....</b>	<b>115</b>
4.1 Introduction and objectives .....	115
4.2 Characterisation of mucus from human mucosal lavage samples.....	116
4.3 Profiling of glycans from human mucins.....	120
4.3.1 Lectin probing of mucins by dot blot and force spectroscopy .....	120
4.3.2 Quantification of mucin sialylation by HPAEC-PAD .....	127
4.3.3 Glycomic profiling by mass spectrometry.....	129
4.4 Quantitation of mucin glycosyltransferase expression by qRT-PCR.....	139
4.5 Discussion .....	145
<b>CHAPTER 5) ASSOCIATIONS BETWEEN MUCIN GLYCOSYLATION AND MUCIN DEGRADERS... 149</b>	
5.1 Introduction and objectives .....	149
5.2 Mechanistic studies of the association between <i>R. gnavus</i> CC55_001C, <i>A. muciniphila</i> ATCC BAA 835 and mucin glycosylation.....	151
5.2.1 Profiling of <i>R. gnavus</i> and <i>A. muciniphila</i> growth in monosaccharides and oligosaccharides .....	151
5.2.2 Profiling of <i>R. gnavus</i> and <i>A. muciniphila</i> growth in mucins .....	154
5.2.2.1 Characterisation of pPGM and LS174T mucin glycosylation .....	154
5.2.2.2 Initial characterisation of <i>A. muciniphila</i> and <i>R. gnavus</i> growth in pPGM .....	159
5.2.2.3 Growth assays of <i>R. gnavus</i> and <i>A. muciniphila</i> in pPGM and LS174T mucin .....	160
5.2.2.4 Metabolite production and pathways utilised by <i>A. muciniphila</i> and <i>R. gnavus</i> .....	162
5.2.3 Colonisation of germ-free mice with <i>R. gnavus</i> and <i>A. muciniphila</i> .....	168
5.2.3.1 Glycomic profiling of mucins from mice mono- and co-colonised with <i>A. muciniphila</i> and <i>R. gnavus</i> by mass spectrometry.....	170
5.2.3.2 Quantification of mucin sialylation from mice mono- and co-colonised with <i>A. muciniphila</i> and <i>R. gnavus</i> by HPAEC-PAD.....	177
5.2.3.3 Metabolite production in the caecal contents of mice mono- and co-colonised with <i>A. muciniphila</i> and <i>R. gnavus</i> .....	179
5.3 Discussion .....	182
<b>CHAPTER 6) CONCLUSIONS AND PERSPECTIVES.....</b>	<b>190</b>
<b>REFERENCES .....</b>	<b>197</b>
<b>APPENDIX 1.....</b>	<b>243</b>
Commercial suppliers of chemicals, reagents and equipment .....	243
<b>APPENDIX 2.....</b>	<b>247</b>
Patient metadata .....	247
<b>APPENDIX 3.....</b>	<b>251</b>
Electronic data.....	251
<b>APPENDIX 4.....</b>	<b>252</b>

Effect of confounding variables on microbiota composition (qPCR) .....	252
<b>APPENDIX 5</b> .....	<b>253</b>
Dot plots of glycan structures .....	253
Dot plots of glycosyltransferase expression .....	255
<b>APPENDIX 6</b> .....	<b>256</b>
BlastP results of <i>R. gnavus</i> CC55_001C homologous proteins to nan locus of <i>R. gnavus</i> ATCC 29149 .....	256

## List of figures

Figure 1   Anatomy of the human GI tract.....	19
Figure 2   Structure of the intestinal wall.....	20
Figure 3   Monomeric structures of MUC1 (transmembrane) and MUC2 (gel-forming) mucins.....	27
Figure 4   Biosynthesis and secretion of polymeric MUC2 .....	29
Figure 5   Glycan core structures and glycosyltransferases involved in GI mucin glycosylation.....	32
Figure 6   Diagram depicting transplantation of microbiota between conventionally and germ-free raised mice and zebrafish .....	34
Figure 7   Phylum level composition of the gut microbiota throughout life.....	42
Figure 8   Schematic diagram illustrating the effect of impaired mucin O-glycosylation in the GI tract .....	65
Figure 9   PCoA plot of 16S taxonomic profiles of mucus associated microbiota from control, UC and CD patients.....	98
Figure 10   Average family and genus level mucosal microbiota compositions of the ascending and sigmoid colon .....	99
Figure 11   Dot plots of diversity, evenness and richness indices of mucus associated microbiota from control and UC patients .....	101
Figure 12   PCA plots of samples based on abundance of 9 specific bacterial groups quantified by qPCR.....	103
Figure 13   Scatter plots of abundance of each bacterial group in sigmoid versus ascending colon .....	105
Figure 14   Dot plots showing abundances of bacterial groups measured by qPCR.....	107
Figure 15   Example of a mucin chain purified from human mucosal lavage and imaged by AFM.....	117
Figure 16   Example of a composite AgPAGE of mucins purified from human mucosal lavage.....	119
Figure 17   Dot blots of mucins purified from human mucosal lavage samples.....	122
Figure 18   Graphs showing the distribution of binding event separation distances in soluble mucins .....	123
Figure 19   Graphs showing the distribution of adhesion affinity measurements in soluble mucins.....	125

Figure 20  Abundance of sialic acid liberated from purified mucins, as determined in four patient groups by HPAEC-PAD .....	128
Figure 21  Example MALDI-TOF spectra of mucin glycans isolated from a human mucosal lavage sample from the ascending colon.....	131
Figure 22  Glycan structure profiles of sigmoid and ascending colon in control and UC patients .....	132
Figure 23  Dot plots of glycan structures determined to be significantly different in abundance between sigmoid and ascending colon of UC patients.....	135
Figure 24  Profile of fucosylated structures in sigmoid and ascending colon of control and UC patients .....	137
Figure 25  Profile of sialylated structures in sigmoid and ascending colon of control and UC patients .....	138
Figure 26  Sialyltransferases with significantly different expression patterns between sample groups.....	141
Figure 27  PCA plot of glycosyltransferase expression data .....	142
Figure 28  Schematic diagram demonstrating the various components which may impact on the development of IBD and how they may be interlinked .....	149
Figure 29  Growth of <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C in different carbohydrate sources.....	152
Figure 30  MALDI-TOF spectra of glycans liberated from pPGM and LS174T mucins.....	155
Figure 31  Growth of <i>R. gnavus</i> CC55_001C and <i>A. muciniphila</i> ATCC BAA 835 in YCFA basal medium supplemented with 1 % pPGM .....	159
Figure 32  Fold change of <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C in pPGM and LS174T mucins.....	161
Figure 33  Change in mM concentration of metabolites by <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C grown in pPGM supplemented CP media .....	164
Figure 34  Change in mM concentration of metabolites by <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C grown in LS174T mucin supplemented CP media .....	165
Figure 35  Metabolites produced by <i>R. gnavus</i> CC55_001C and <i>A. muciniphila</i> ATCC BAA 835 at 48 hours in the context of their metabolic pathways .....	167
Figure 36  Mono- and co-colonisation of germ-free C57BL/6J mice with <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C .....	169
Figure 37  Example MALDI-TOF spectra of colonic mucin glycans from mice.....	172

<b>Figure 38   Glycosylation profile in the colon of germ-free mice, and mice mono- and co-colonised with <i>R. gnavus</i> CC55_001C and <i>A. muciniphila</i> ATCC BAA 835 .....</b>	<b>173</b>
<b>Figure 39   Relative abundance of sialylated structures in colonic mucins from germ-free mice, and mice mono- and co- colonised with <i>R. gnavus</i> CC55_001C and <i>A. muciniphila</i> ATCC BAA 835 .....</b>	<b>175</b>
<b>Figure 40   Relative abundance of fucosylated structures in colonic mucins from germ-free mice, and mice mono- and co- colonised with <i>R. gnavus</i> CC55_001C and <i>A. muciniphila</i> ATCC BAA 835 .....</b>	<b>176</b>
<b>Figure 41   Abundance of sialic acid as determined by HPAEC-PAD in colonic mucins from mice mono- and co- colonised with <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C.....</b>	<b>178</b>
<b>Figure 42   Concentrations (mM) of the main metabolites in the caecal contents of germ-free mice .....</b>	<b>180</b>
<b>Figure 43   Change in mM concentration of metabolites in mice mono- and co-colonised with <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C when compared with germ-free controls .....</b>	<b>181</b>

## List of tables

Table 1   Human membrane-bound and secreted mucins in the GI tract.....	24
Table 2   Studies of mucosal associated microbiota in IBD .....	57
Table 3   Family, genus and species specific primers used for 16S targeted quantification .....	72
Table 4   Primer design and sequences used to amplify the V1-V2 region of bacterial 16S DNA for 16S sequencing .....	75
Table 5   Table of lectins and their specificities .....	80
Table 6   Primers used to quantify the expression of human glycosyltransferases .....	87
Table 7   P values for statistical tests carried out with different microbial parameters measured.....	109
Table 8   Percentage of total samples containing structure, and mean relative abundance of 24 most prevalent glycan structures in 4 patient groups .....	133
Table 9   P values for statistical tests comparing glycosylation in UC patients and controls .....	143
Table 10   P values for statistical tests comparing glycosyltransferase expression in UC patients and controls .....	144
Table 11   Changes in optical density of carbohydrate supplemented media after 48 hrs of inoculation with <i>A. muciniphila</i> ATCC BAA 835 or <i>R. gnavus</i> CC55_001C .....	153
Table 12   Relative abundance of oligosaccharides detected in pPGM by MALDI-TOF ...	156
Table 13   Relative abundance of oligosaccharides detected by MALDI-TOF in the MUC2 mucin fraction purified from LS174T adenocarcinoma cell line .....	157
Table 14   Relative abundance of oligosaccharides detected by MALDI-TOF in the mixed mucin fraction purified from LS174T adenocarcinoma cell line .....	158
Table 15   Change in mM concentration of metabolites after 48 hours growth in pPGM and LS174T mucin supplemented medium .....	163
Table 16   Occurrence of predicted enzymes involved in pathways of propionate metabolism in <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C.....	166
Table 17   Table of glycan structures identified in mucins of mice mono- and co-colonised with <i>A. muciniphila</i> ATCC BAA 835 and <i>R. gnavus</i> CC55_001C.....	174

## Abbreviations

<b>AC</b>	alternating current
<b>AFM</b>	atomic force microscopy
<b>Ag</b>	agarose
<b>AgPAGE</b>	composite agarose-polyacrylamide gel electrophoresis
<b>ANOVA</b>	analysis of variance
<b>APS</b>	ammonium persulfate
<b>BabA</b>	blood group Ag-binding adhesin
<b>BHI-YH</b>	brain-heart infusion media supplemented with yeast extract
<b>bp</b>	base pairs
<b>BSA</b>	bovine serum albumin
<b>CARD</b>	caspase recruitment domain
<b>CD</b>	Crohn's disease
<b>cDNA</b>	complementary DNA
<b>Coll.</b>	collaborators
<b>Ct</b>	cycle threshold
<b>DC</b>	direct current
<b>DCs</b>	dendritic cells
<b>DMB</b>	1,2-diamino-4,5-methylene-dioxybenzene
<b>DNA</b>	deoxyribonucleic acid
<b>DSS</b>	dextran sodium sulfate
<b>DTT</b>	dithiotreitol
<b>EDTA</b>	ethylenediaminetetraacetic acid
<b>ER</b>	endoplasmic reticulum
<b>Extr-GuHCl</b>	extraction guanidine hydrochloride
<b>Fcgbp</b>	fc-gamma binding protein
<b>FISH</b>	fluorescence <i>in situ</i> hybridisation
<b>FITC</b>	fluorescein isothiocyanate
<b>Fuc</b>	L-fucose
<b>Gal</b>	D-galactose
<b>GalNAc</b>	N-acetylgalactosamine
<b>GALT</b>	gut-associated lymphoid tissue
<b>GAPs</b>	goblet cell-associated antigen passages
<b>gDNA</b>	genomic DNA

<b>GI</b>	gastrointestinal
<b>Glc</b>	D-glucose
<b>GlcNAc</b>	N-acetylglucosamine
<b>GuCl</b>	guanidinium chloride
<b>GuHCl</b>	guanidine hydrochloride
<b>GWAs</b>	genome-wide association studies
<b>h</b>	hour(s)
<b>HCl</b>	hydrochloric acid
<b>HMP</b>	human microbiome project
<b>HPAEC</b>	high-performance anion-exchange chromatography
<b>HPLC</b>	high performance liquid chromatography
<b>HT29-MTX</b>	HT29-methotrexate
<b>Hz</b>	hertz
<b>IBD</b>	inflammatory bowel disease
<b>IBS</b>	irritable bowel syndrome
<b>IL</b>	interleukin
<b>IRGM</b>	immunity-related GTPase family M protein
<b>JAK</b>	janus kinase
<b>kDa</b>	kilo dalton
<b>KDN</b>	3-deoxy-d-glycerol-galacto-2-nonulosonic acid
<b>Lac</b>	D-lactose
<b>LacNAc</b>	N-Acetyllactosamine
<b>LC/MS</b>	liquid chromatography/mass spectrometry
<b>LP</b>	lamina propria
<b>LPS</b>	lipopolysaccharide
<b>LRR</b>	leucine-rich repeats
<b>MACs</b>	microbial accessible carbohydrates
<b>MALDI-TOF</b>	matrix-assisted laser desorption/ionization-time of flight
<b>MalII</b>	<i>Maackia amurensis</i> lectin II
<b>MAMPs</b>	microbial associated molecular patterns
<b>MDa</b>	mega dalton
<b>MDR</b>	multiple drug resistance gene
<b>min</b>	minute(s)
<b>MUC</b>	human mucin ( <i>MUC</i> , gene)

<b>Muc</b>	mouse mucin ( <i>Muc</i> , gene)
<b>MWCO</b>	molecular weight cut-off
<b>Myd88</b>	myeloid differentiation primary response gene 88
<b>NaCl</b>	sodium chloride
<b>NaOH</b>	sodium hydroxide
<b>Neu5Ac</b>	N-Acetylneuraminic acid
<b>Neu5Gc</b>	N-Glycolylneuraminic acid
<b>NF-<math>\kappa</math>B</b>	nuclear factor kappa-light-chain-enhancer of activated B cells
<b>NK-T</b>	natural killer T cells
<b>NLRP6</b>	NOD-like receptor family pyrin domain containing 6
<b>NMR</b>	Nuclear Magnetic Resonance
<b>NUH</b>	Norfolk and Norwich University Hospital
<b>NOD2</b>	nucleotide-binding oligomerization domain-containing protein 2
<b>O/N</b>	overnight
<b>OCTN</b>	organic cation transporter gene
<b>PA</b>	polyacrylamide
<b>PAGE</b>	polyacrylamide gel electrophoresis
<b>PAMPs</b>	pathogen associated molecular patterns
<b>PAD</b>	pulsed amperometric detection
<b>PAS</b>	periodic acid-schiff
<b>PBS</b>	phosphate buffered saline
<b>PCA</b>	principal component analysis
<b>PCoA</b>	principal coordinate analysis
<b>PCR</b>	polymerase chain reaction
<b>PGM</b>	porcine gastric mucin
<b>PMSF</b>	phenylmethanesulfonyl fluoride
<b>PNA</b>	peanut agglutinin
<b>pPGM</b>	purified porcine gastric mucin
<b>PRRs</b>	pattern recognition receptors
<b>PTS</b>	proline, threonine and serine repeats
<b>PVDF</b>	polyvinylidene fluoride
<b>qPCR</b>	quantitative PCR
<b>qRT-PCR</b>	quantitative reverse transcription polymerase chain reaction
<b>RCA</b>	<i>Ricinus communis</i> agglutinin

<b>Red-GuHCl</b>	reduction guanidine hydrochloride
<b>RNA</b>	ribonucleic acid
<b>rpm</b>	revolutions per minute
<b>rRNA</b>	ribosomal RNA
<b>s</b>	seconds
<b>SabA</b>	sialic acid-binding adhesion
<b>SCFA</b>	short chain fatty acid
<b>SCID</b>	severe combined immunodeficient
<b>SDS</b>	sodium dodecyl sulfate
<b>SEA</b>	sea-urchin-sperm protein-enterokinase-agrin
<b>Ser</b>	serine
<b>SFB</b>	segmented filamentous bacteria
<b>SI</b>	small intestine
<b>SIgA</b>	secretory immunoglobulin A
<b>SNA</b>	<i>Sambucus nigra</i> lectin
<b>STAT</b>	signal transducers and activators of transcription
<b>STn</b>	sialyl-Tn
<b>TEM</b>	transmission electron microscopy
<b>TEMED</b>	tetramethylethylenediamine
<b>TF</b>	Thomsen-Friedenreich
<b>Th</b>	t helper
<b>Thr</b>	threonine
<b>TLR</b>	toll-like receptor
<b>TMB</b>	3,3',5,5'-tetramethylbenzidine
<b>TNF</b>	tumour necrosis factor
<b>T<sub>reg</sub></b>	regulatory t cells
<b>T-RFLP</b>	terminal restriction fragment length polymorphism
<b>TSP</b>	sodium 3-(Trimethylsilyl)-propionate- <i>d</i> 4
<b>UC</b>	ulcerative colitis
<b>UEA</b>	University of East Anglia
<b>UEA</b>	<i>Ulex europaeus</i> agglutinin
<b>VNTR</b>	variable number tandem repeats
<b>Vol</b>	volume
<b>Vol/vol</b>	volume per volume

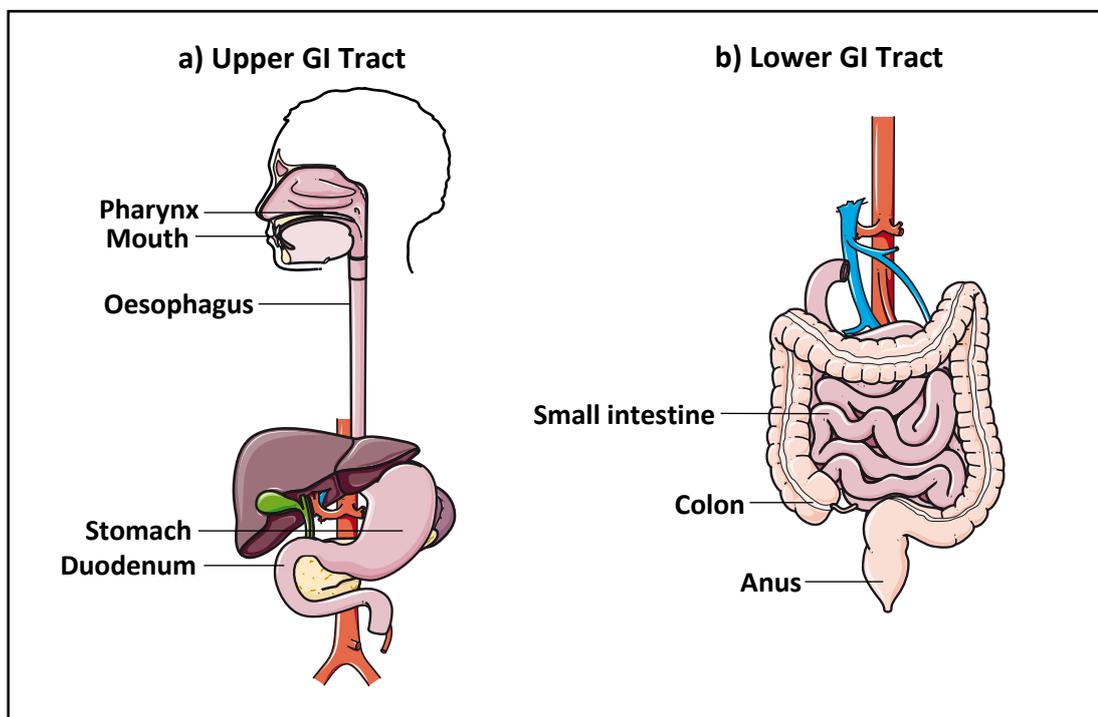
<b>vWF</b>	von willebrand factor
<b>WGA</b>	wheat germ agglutinin
<b>Wt/vol</b>	weight per volume

## Chapter 1) Introduction

### 1.1 Overview of the human gastrointestinal tract

#### 1.1.1 General structure and function

The gastrointestinal (GI) tract comprises a system of organs facilitating a number of crucial human functions, including the digestion of food and extraction of energy, absorption of nutrients, water and electrolyte balance, hormone production and excretion of remaining waste (Sansone, 2004). The GI tract is divided into two anatomically distinct parts; the upper tract (mouth, pharynx, oesophagus, stomach and duodenum) and the lower tract (small intestine, SI; large intestine or colon and anus) (Fig. 1). The morphology of each organ is highly specialised to carry out a unique function. For example, the stomach is important for its role in the chemical breakdown of the food bolus, whereas the SI is responsible for nutrient absorption, and the colon for absorption of salts and water from undigested food (Adibi, 1976; Sandle, 1998).

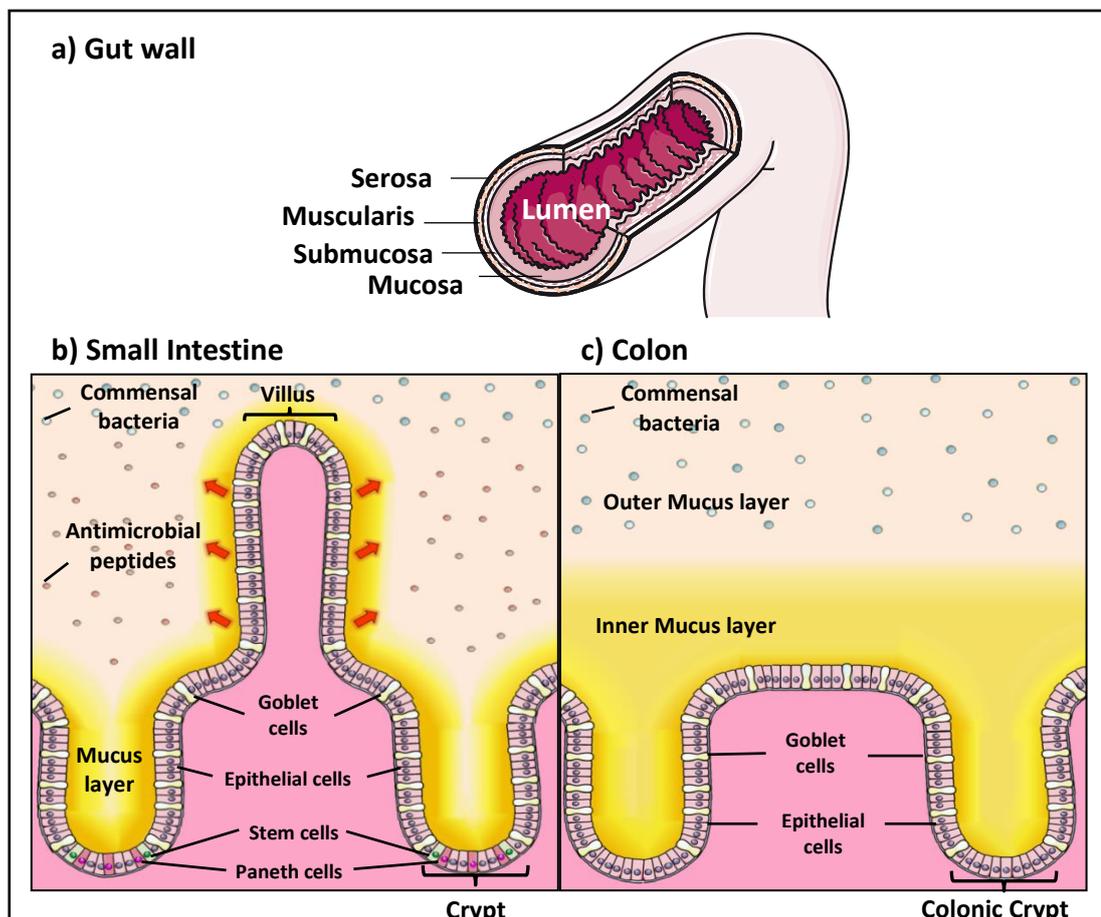


**Figure 1 | Anatomy of the human GI tract**

*Schematic representation of the upper (a) and lower (b) GI tract with its constituents*

*(Servier Medical Art, 2016)*

The many segments of the GI tract form a surface area of between 250 and 400 m<sup>2</sup>, representing one of the primary and largest interfaces between the host, environmental factors and antigens in the human body. Approximately 60 tonnes of food are expected to be ingested and pass through the GI tract in a life time, bringing with it a wealth of microorganisms from the environment which impose a huge threat on gut integrity (Bengmark, 1998). To protect from insult and preserve homeostasis, the GI tract limits exposure of the host immune system to luminal contents through the employment of a multifactorial and dynamic intestinal barrier (Fig. 2). This barrier is comprised of several integrated and interactive components that are physical (the epithelium and mucus), biochemical (enzymes, anti-microbial proteins), immunological (IgA and epithelia-associated immune cells), and microbial (commensal microbes; i.e. microbiota) in nature (Fig. 2b-c) (Hooper and Macpherson, 2010), as further described below.



**Figure 2 | Structure of the intestinal wall**

(a) Section of the GI tract with layers indicated. From the luminal compartment outwards: mucosa, submucosa, muscle layer, and serosa. Schematic diagrams of (b) Small intestinal, and (c) colonic cell walls showing elements of the intestinal barrier.

(Servier Medical Art, 2016)

### 1.1.2 The intestinal epithelium

A critical contributor to the physical barrier in the gut is the intestinal epithelium, which forms part of a layer, known as the mucosa, in the gut wall (Fig. 2a). At the innermost surface, the mucosa is a dynamic and rapidly self-renewed layer composed of the epithelium, lamina propria (LP) (connective tissue that is rich in capillaries and containing a central lymph vessel) and muscularis mucosae (smooth muscle that maintains gentle agitation of gut contents through the GI tract). The epithelium is comprised of a variety of cell types with different functional properties, which are sealed by tight junctions to form a relatively impenetrable shield to microbes (Fig. 2b-c) (Marchiando et al., 2011). The distribution and organisation of these cells helps to maintain a specific barrier function within each organ.

In the SI, the surface area is vastly increased by the formation of finger-like projections called villi and invaginations, known as the Crypts of Lieberkuhn (Fig. 2b). Proliferation and differentiation of multipotent stem cells that are concentrated near the base of the crypts provide a rapid and continual epithelial cell turnover in these structures. (Barker et al., 2008; Garrett et al., 2010; Marshman et al., 2002; Sancho et al., 2003). The process is similar in the mucosa of the colon, except for the absence of villi, which results in terminally differentiated cells being shed from the surfaces around the mouths of crypts (Crosnier et al., 2006). Migrating cells are destined to become one of three main cell types: absorptive enterocytes, mucus secreting goblet cells and hormone secreting enteroendocrine cells (Crosnier et al., 2006). The villi of the SI are mainly populated by absorptive enterocytes which facilitate the absorption of nutrients through the epithelium by virtue of their highly folded brush border and release of a cocktail of hydrolytic enzymes (Barker et al., 2008). In contrast, the colonic epithelium is populated by a high proportion of goblet cells (Barker et al., 2008). The process of migration upwards replaces terminally differentiated cells that are shed from the villus tips into the lumen. Hence cells are disseminated along the length of the crypt-villus axes, and consequently, the position along the length of the axes is an indication of cell age (Marshman et al., 2002). Exceptionally, a fourth cell type, Paneth cells, produced only in the SI, migrate downwards and reside at the base of the crypts, for 3-6 weeks (Barker et al., 2008; Sancho et al., 2003). The Paneth cells are responsible for the secretion of anti-microbial agents such as cryptdins and defensins, and are therefore important in shaping the microbial community and maintaining homeostasis (Crosnier et al., 2006; Sancho et al., 2003).

In contrast to Paneth cells, intestinal epithelial cells, having undergone up to six rounds of division while migrating up the villus, are replaced by cell shedding every 5-7 days in mice

(Marshman et al., 2002). Specifically in the SI, it is estimated that approximately 1400 epithelial cells are shed from a single villus tip every 24 hours (h) (Potten, 1990; Williams et al., 2015). In humans, it is estimated that  $10^{11}$  cells are lost from the SI per day (Williams et al., 2015). This rapid shedding and renewal of the epithelium leaves gaps where junctional integrity and paracellular permeability may be compromised, presenting a challenge for the maintenance of barrier function. Loss of barrier function has been associated with conditions such as inflammatory bowel disease (IBD), due to augmented epithelial permeability (Garrett et al., 2010; Schulzke et al., 2009; Watson et al., 2009). A delicate balance between cell shedding and cell proliferation is therefore vital to prevent disease.

## 1.2 The GI mucus layer

### 1.2.1 Structure and organisation of GI mucus

Mucus is produced as a lining in all bodily tissues that are exposed to the external environment, including the eyes, respiratory system and reproductive organs (Bansil and Turner, 2006). Throughout the GI tract, a mucus layer is secreted by epithelial goblet cells. The mucus is a water-insoluble, biochemically complex gel made up of large molecular weight proteins known as mucins, many of which are known to have been conserved from early metazoan species (Lang et al., 2007). In addition to mucins, the intestinal mucus layer comprises lipids, electrolytes, hydrolases, antimicrobial peptides, immunoglobulins and many other noxious agents (See section 1.2.3) (Antoni et al., 2013; Atuma et al., 2001; Juge, 2012).

Studies of mucus properties are hampered by its biological complexity and high water content (~98%). The mucus is easily lost during tissue sampling and handling, and the use of fixatives to perform histological techniques can cause dehydration resulting in the collapse of the mucus layer (Cohen et al., 2012; Hansson, 2012). However, in a pioneering study in 2001, Atuma and collaborators (coll.) founded an *in vivo* method allowing the thickness of rat intestinal mucus to be measured. This showed that the thickness of mucus is variable between regions of the GI tract. Of six regions studied, the measurements ranged from  $\approx 830$   $\mu\text{m}$  in the colon, where it was at its thickest, to  $\approx 123$   $\mu\text{m}$  in the jejunum, at its thinnest (Atuma et al., 2001). Similar variability has since been shown *ex vivo* in the mouse intestine (Ermund et al., 2013; Gustafsson et al., 2012b). In humans, the colonic mucus thickness is estimated to be  $450 \mu\text{m} \pm 70 \mu\text{m}$  (Gustafsson et al., 2012b).

In addition to regional variations in thickness, the organisation of GI mucus is also variable throughout the GI tract. In the SI there is a single, unattached and loose layer (Ermund et al., 2013). In contrast, in the colon, mucus is divided into two distinct layers, an inner, densely packed and impermeable layer and a loose outer coating (Atuma et al., 2001; Ermund et al., 2013). Various studies with different animal models and systems of measuring mucus have shown that the outer layer in the colon can be easily aspirated leaving just the inner mucus layer intact. Following removal, the outer layer was replenished over 60-90 minutes (min) (Atuma et al., 2001; Gustafsson et al., 2012b; Johansson et al., 2008). In contrast, the small intestinal mucus was patchier and in some places could be removed almost completely. The replenishment of this layer was slower or did not occur at all (Atuma et al., 2001; Gustafsson et al., 2012b; Johansson et al., 2008). Using tissue explants of mouse intestine, it was shown that whilst the mucus of the SI was fully penetrable by bacteria-sized beads, the inner layer of mucus in the colon was either completely or partially impenetrable (Ermund et al., 2013). This arrangement of mucus supports the individual functions of these GI regions, in the SI providing a permeable lining which permits the absorption of food derived nutrients and development of the immune system, and in the colon maintaining a physical separation of the epithelium from luminal threats such as microorganisms (Ermund et al., 2013; Johansson et al., 2013a). Another degree of variation is due to the composition of mucus along the GI tract (See below).

## 1.2.2 Mucus composition

### 1.2.2.1 *Mucin proteins*

The mucin family includes up to 20 known high molecular weight and heterogeneous glycoproteins which are comprised of around 70% carbohydrate (Corfield, 2015; Linden et al., 2008). Each mucin has a tissue or cell specific pattern and can be classified into one of two groups, secreted (gel-forming), or adherent to (transmembrane) the epithelium (Johansson et al., 2013b). Gel forming mucins in the GI tract include MUC2, MUC5AC, MUC5B and MUC6 which are encoded by the chromosome locus 11p15.5 (Table 1) (Derrien et al., 2010). They are secreted onto the mucosal surface where they form homo-oligomeric networks which are viscoelastic in nature and comprise the largest glycoproteins in the body. MUC2 is the predominant secreted mucin in the SI and colon and the main component of the inner and outer mucus layers (Corfield, 2015; Johansson et al., 2011b; Kesimer et al., 2010). The secreted mucins play an important role of mediating the host relationship with the gut microbiota (See sections 1.2.2.3 and 1.2.5.2). In contrast, transmembrane mucins are

typically monomeric proteins firmly tethered to the epithelial surface. These mucins are a major contributor to a layer known as the glycocalyx, a dense impermeable barrier residing below the two intestinal mucus layers (Johansson et al., 2011b; Sheng et al., 2012). The transmembrane mucins have a dual function, involved in cell recognition and adhesion as well as signal transduction to the underlying epithelia. Mucins belonging to this group in the GI tract include MUC1, MUC3A, MUC3B, MUC4, MUC12, MUC13, MUC15 and MUC17 (Table 1) (Derrien et al., 2010; Juge, 2012).

Type	MUC Gene	Tissue localization
Membrane	MUC1	Stomach, duodenum, colon
	MUC3A/B	Goblet and absorptive cells; Jejunum, ileum and colon; small intestinal columnar cells and surface colonic epithelium
	MUC4	Stomach and colon
	MUC12	Stomach, small intestine and colon
	MUC13	Small intestine and colon
	MUC15	Small intestine and colon
	MUC17	Stomach, small intestine (highest expression in duodenum) and colon (transverse)
	MUC20	Colon
	MUC21	Colon
	Secreted	MUC2
MUC5AC		Stomach
MUC5B		Colon (weakly expressed)
MUC6		Stomach (glands), duodenum (Brunner glands)

**Table 1 | Human membrane-bound and secreted mucins in the GI tract**

*Table adapted from Tailford et al., 2015*

### 1.2.2.2 Biosynthesis and molecular properties of mucins

#### - Transmembrane mucins

There are 11 known genes encoding the transmembrane mucins that are present on the surface of epithelial cells in a wide variety of mucosal sites (Corfield, 2015). The most ubiquitously expressed transmembrane mucin, MUC1, has been extensively studied, and its structure well characterised (Fig. 3). At the N-terminus is an extracellular subunit containing a signal peptide and a variable number tandem repeat (VNTR) domain, which contains varying numbers of a 20-amino acid motif. This repeat sequence is dominated by proline, threonine and serine (PTS) residues which are heavily O-glycosylated. Sometimes likened to the bristles on a bottle brush, the O-glycan side chains confer a rigid and extended structure upon the mucin molecule in two ways: i) by spatially blocking movement around the peptide bonds on the protein backbone and ii) by repulsion between the negative charges on the O-glycans themselves. As a result, the extracellular domain is expected to be around 200-500 nm in length, and towers above the cell surface. This extended structure also endows the mucin molecules with a high water holding capacity generating a large volume that acts as a barrier to luminal contents (Gendler et al., 1990; Perez-Vilar and Hill, 1999; Shirazi et al., 2000).

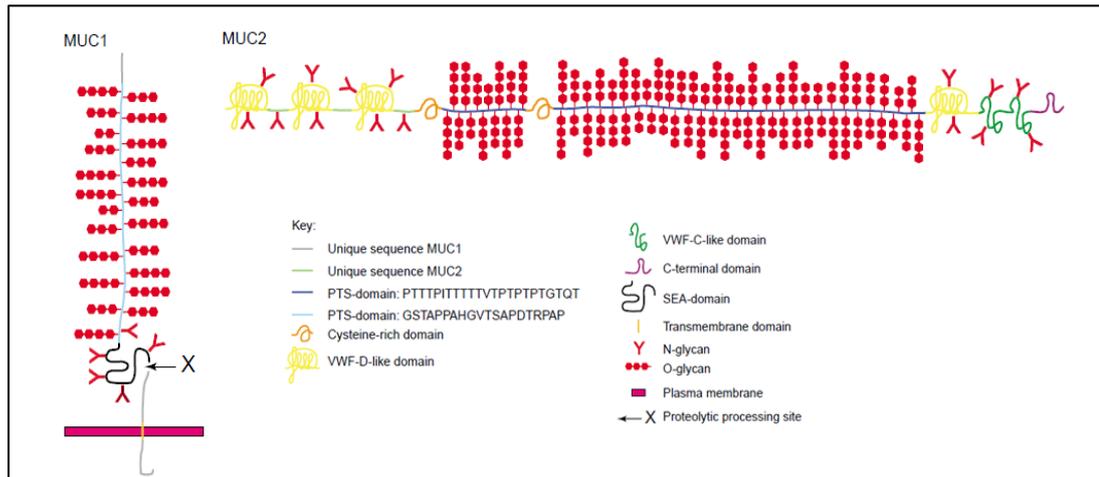
At the C terminal end of the extracellular domain close to the cell membrane is a sea-urchin-sperm protein-enterokinase-agrin (SEA) domain, which is shared by all transmembrane mucins, except for MUC4. The SEA domain contains a serine (Ser) residue which is auto-catalytically cleaved during synthesis to form a non-covalent association between the extracellular and transmembrane domain (Fig. 3). More importantly, the extracellular domain can be shed from the surface, a process which is proposed to occur either *via* a second distinct cleavage at this site or by shear forces physically separating the two domains (Linden et al., 2008; Macao et al., 2006; Palmi-Pallag et al., 2005). This shedding may allow the mucin proteins to act as detachable decoy ligands for bacterial adhesins, particularly as mucins are decorated with many of the same oligosaccharides that can be found on the epithelial surface. An example of this has been shown in the case of *Helicobacter pylori* infection, where interactions between MUC1 and blood group Ag-binding adhesin (BabA) or sialic acid-binding adhesin (SabA) are suggested to prevent attachment of the pathogen to the epithelial surface (Linden et al., 2009). In further support of this, the expression of cell-surface mucins is most diverse in regions of the body where the risk of infection is highest, including the eyes, respiratory tract and GI tract, providing an advantage in the protection against a range of potential pathogens (Linden et al., 2008). Following the SEA domain is a

membrane-spanning region that anchors the mucin to the cell and a cytoplasmic tail, which is rich in Ser and threonine (Thr) residues. These residues are implicated in signal transduction, acting as docking sites which can sense the environment through association with signalling factors, such as P53 tumour suppressor protein and  $\beta$ -catenin, a protein involved in the Wnt signalling pathway. A series of post-translational modifications (e.g. phosphorylation) then transmit these signals into the cell (Johansson et al., 2011b; Juge, 2012; Lillehoj et al., 2004; Singh and Hollingsworth, 2006).

- **Gel-forming mucins**

Studies of the biosynthesis of full length mucins are limited by the difficulty in generating full length recombinant mucin proteins. Alternative approaches have therefore concentrated on developing shorter domain constructs, leading to increased knowledge about the functional roles of the mucin domains (Backstrom et al., 2013). Studies of gel-forming mucin structure have primarily focussed on MUC2. Unlike cell-surface mucins, MUC2, as well as other secretory mucins such as MUC5AC and MUC6 do not contain SEA or transmembrane domains (Dekker et al., 2002). These mucins contain a number of cysteine rich regions dispersed throughout the protein, which are not found in MUC1 (Fig. 3). The remaining features of gel-forming mucins are remarkably similar to MUC1. For example, MUC2 contains a large protein backbone of ~200-500 kilo daltons (kDa), at the centre of which are two VNTR domains divided by a cysteine rich region. The sequence, number and length of the smaller VNTR domain appears to be less conserved than the larger domain, which contains a repeated 23-amino acid motif. Alike the cell-surface mucins more than 60% of the VNTR sequence is comprised of proline, Thr and Ser residues. The PTS domains are decorated in O-glycans, and the addition of sialic acid and sulphate residues to these glycans results in a highly negative surface charge generating a stiff 'bottle brush' conformation (Fig. 3) (Asker et al., 1995; Bansil and Turner, 2006; Perez-Vilar and Hill, 1999; Shirazi et al., 2000). At the N- and C-terminals of the backbone, and occasionally interspersed between the PTS repeats are the cysteine rich (>10%), sparsely glycosylated regions which possess sequence similarity to domains such as C-terminal cysteine knot domains, von Willebrand factor (vWF) C and D domains and CysD domains. The terminal cysteine regions are involved in the polymerisation of mucin *via* the formation of disulphide bonds, resulting in dimerisation and trimerisation of mucin monomers (Asker et al., 1995; Godl et al., 2002). This process has been confirmed by visualisation using transmission electron microscopy (TEM) and atomic force microscopy (AFM), showing that the mucins form long fibres of varying lengths (ocular mucins: 200-600 nm, sometimes reaching 1500 nm, bronchial mucins: 300-2500 nm) (Mikkelsen et

al., 1985; Round et al., 2002). Further confirmation of these bonds was obtained by the addition of a reducing agent resulting in chemical cleavage accompanied by a reduction in the molecular weight (Bansil and Turner, 2006; Perez-Vilar and Hill, 1999). On the other hand, the CysD domains, located amongst the PTS repeats, have been proposed to form additional non-covalent cross-links within the mucin gel thereby establishing the pore size of the mucus. These bonds are not affected by the addition of reducing agent (Ambort et al., 2011).



**Figure 3 | Monomeric structures of MUC1 (transmembrane) and MUC2 (gel-forming) mucins.**

The PTS domains belonging to MUC1 and MUC2 are indicated in light and dark blue respectively. MUC1 contains a SEA-domain, shown in black, which can be cleaved by autoproteolysis. In contrast, MUC2 contains C- and N-terminal and interspersed cysteine rich regions shown in green, yellow and orange respectively (Dekker et al., 2002; Linden et al., 2008).

Figure adapted from Dekker et al., 2002.

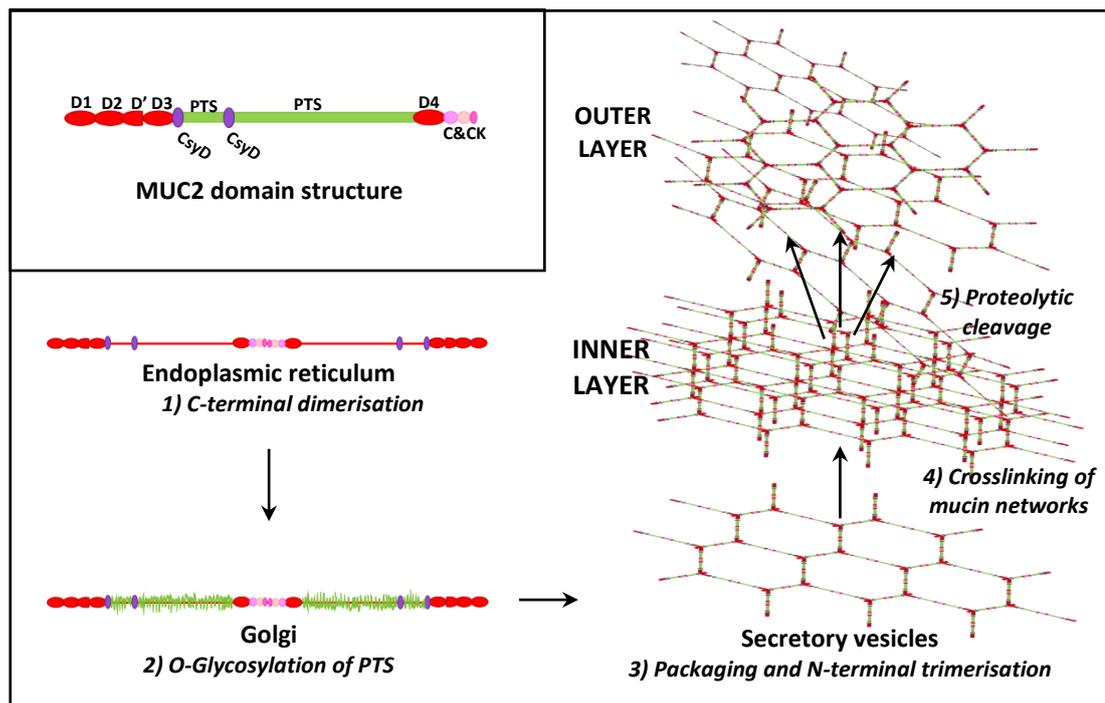
The process for MUC2 secretion is well characterized (Fig. 4). Unlike the other gel-forming mucins, MUC2 is produced and stored in goblet cells (Cornick et al., 2015). After translation, MUC2 peptide chains are translocated to the lumen of the endoplasmic reticulum (ER) where they fold and are dimerized at the C-terminus *via* the cysteine knot domains (Asker et al., 1995; Lidell et al., 2003). The mucin dimers are then heavily O-glycosylated in the Golgi and packaged into secretory vesicles ready for polymerization at the N-terminus D3 domain to form disulphide bonded trimers (Asker et al., 1998; Godl et al., 2002). Further cross-linkage of MUC2 polymers *via* both covalent and non-covalent mechanisms aids the formation of the well stratified polymeric network upon which both outer and inner mucus layers are organised. These cross-linkages include non-covalent bonds between the CysD domains, as

mentioned above, as well as covalent cross-linking between cleaved vWF D domains and a protein present in mucus known as immunoglobulin G Fc-gamma binding protein (Fcgbp) (Ambort et al., 2011; Johansson et al., 2009). Whilst fully glycosylated monomeric MUC2 is estimated to be approximately 2.5 mega daltons (MDa) in size, polymerized mucin may reach sizes in excess of 100 MDa (Johansson et al., 2011a).

Packaging of mucins occurs in the secretory vesicles at a pH of 6.2 and in the presence of calcium (Gustafsson et al., 2012a). Packaged MUC2 is formed of concatenated rings comprised of five or six repeating units. These rings are maintained in a condensed structure by noncovalent interactions between trimers at the N-terminal region (Ambort et al., 2012b). Each ring contains three C-terminal disulphide linked dimers, and a single N-terminal disulphide linked trimer (Fig. 4). Electron microscopy and 3D maps of the N-terminal D3 domains revealed cage-like structures with 2- and 3-fold symmetries, confirming that the MUC2 mucin forms branched net-like structures (Nilsson et al., 2014). Furthermore, MUC2 appears to be packaged as two N-terminal concatenated ring platforms which are turned by 180° against each other, suggesting that every other MUC2 in unfolded mature mucus is turned upside down (Nilsson et al., 2014). The structures are unfolded and flattened upon an increase in pH and removal of calcium ions generating a 1000-fold rapid expansion and a net-like structure (Fig. 4) (Ambort et al., 2012a; Ambort et al., 2012b). Appropriate unfolding of mucin is dependent on functional bicarbonate and chloride channels, the lack of which appears to link cystic fibrosis with its mucus phenotype (Gustafsson et al., 2012a). Furthermore, detachment of small intestinal mucus from the cell surface is dependent on cleavage at the N-terminal region of MUC2 by a host-encoded metalloprotease, meprin-β (Schutte et al., 2014). This process is dependent not only on exposure of the N-terminus *via* bicarbonate-mediated mucin unfolding, but also on microbial-induced cleavage of meprin-β from the cell membrane, as demonstrated by the lack of mucin detachment in germ-free mice (Schutte et al., 2014).

Similarly, in the colon, conversion of the inner mucus layer, to the outer loose layer is likely to be mediated by commensal bacterial proteases. It is proposed that this conversion occurs through continual proteolytic cleavage events within the MUC2 polymers (Johansson et al., 2008). Whilst these events are expected to have a large impact on the structure and properties of mucus, they do not necessarily disrupt the polymeric network (Ambort et al., 2012a). The cleavage events are responsible for a fourfold expansion in volume that converts the mucus from the dense, and insoluble inner layer (Fig. 4) (e.g. *in vitro* in chaotrophic salts,

such as guanidium chloride, GuCl) to the outer loose layer that is readily solubilized. This expansion also allows infiltration of the outer mucus layer by bacteria, whilst the inner layer remains impenetrable. This arrangement into two layers allows a physical separation of microbes from the underlying epithelia, which is critical to maintaining homeostasis in the gut (Ambort et al., 2012b; Johansson et al., 2011a; Johansson et al., 2011b).



**Figure 4 | Biosynthesis and secretion of polymeric MUC2**

Figure adapted from Johansson et al., 2011a.

### 1.2.2.3 Mucin glycosylation

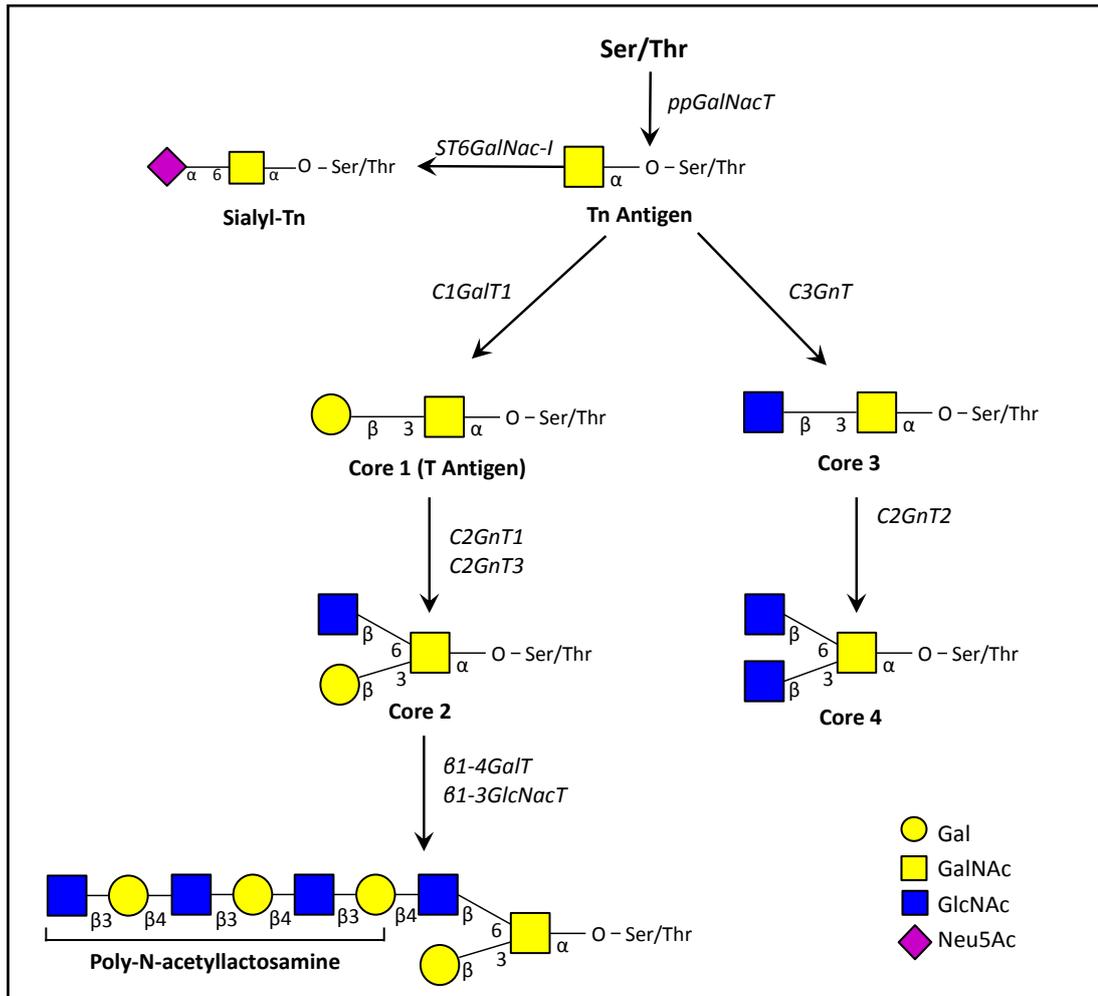
Mucins are decorated by a rich and diverse repertoire of O-glycans that are concentrated in the PTS regions. The glycans protect the mucin backbone from degradation by proteases and facilitate the gel-like structure of mucus by providing a high water holding capacity (Larsson et al., 2013). Aside from these physical functions, it is proposed that mucin O-glycosylation plays an essential biological role in the gut by providing a habitat for commensal organisms. Variation in mucin O-glycosylation occurs both between different species and along the GI tract (See below) (Mastrodonato et al., 2012; Robbe et al., 2004). It is believed that many bacteria can adapt to the host mucosal environment by expressing cell-surface adhesins, such as lectins which they use to anchor themselves to the mucus (Erdem et al., 2007; Etzold et al., 2014; Juge, 2012; Roos and Jonsson, 2002; Tasteyre et al., 2001). In addition, mucin glycans can provide nutrients to bacteria which have adapted to the mucosal environment (Tailford et al., 2015b).

Initiation of mucin O-glycosylation occurs through the addition of an N-acetylgalactosamine (GalNAc) residue to the hydroxyl group of Ser or Thr residues within the PTS domains, forming the Tn antigen (GalNAc-O-Ser/Thr) (Fig. 5). In humans, this reaction can be carried out by one of 20 different peptidyl-GalNAc-transferases, each of which has a unique acceptor peptide substrate specificity (Bennett et al., 2012; Johansson et al., 2011b). The Tn antigen acts as a substrate for downstream glycosylation events which are also catalysed by the sequential action of further glycosyltransferases. Firstly, the addition of a number of other residues occurs to form one of several core structures (Bergstrom and Xia, 2013). To date, a total of 8 core structures have been identified, with cores 1 to 4 being the most abundant in the human gut (Fig. 5). The core structure generated is directly influential in the downstream glycosylation pathway, and therefore the biological function of mucins in different tissues is reliant on the glycosyltransferases expressed (Bennett et al., 2012; Brockhausen, 1999).

Commonly, the Tn antigen is extended to form the core 3 structure (GlcNAc $\beta$ 1 $\rightarrow$ 3GalNAc $\alpha$ -O-Ser/Thr) by the addition of a N-acetylglucosamine (GlcNAc) residue, catalysed by the enzyme core 3  $\beta$ 1,3 N-acetylglucosaminyltransferase (C3GnT) (Fig. 5). Core 3 is expressed only in specific tissues, including the GI tract and the salivary glands, and is the major core structure found in human MUC2 (An et al., 2007; Robbe et al., 2004). Additionally, the Tn antigen can be extended by the addition of a Gal *via* the action of a core 1  $\beta$ 1,3-galactosyltransferase (C1GalT1) to form the core 1 structure (Gal $\beta$ 1  $\rightarrow$  3GalNAc $\alpha$ -O-Ser/Thr) (Fig. 5) (Fu et al., 2011). The action of C1GalT1 is dependent on the function of a

chaperone known as Cosmc (Ju and Cummings, 2002), which helps to promote C1GalT1 stability and is essential for correct glycosylation of mucins, demonstrated by the lethality of its deletion in mice (Wang et al., 2010). A lack of C1GalT1 or of functional Cosmc can lead to the increased expression of Tn antigen or of its sialylated version, Sialyl-Tn (STn), which is generated by the  $\alpha$ 2-6 linkage of sialic acid to Tn *via* the action of a sialyltransferase known as ST6GalNAc-I (Fig. 5) (Marcos et al., 2004). An over-expression of ST6GalNAc-I is associated with observed increases in STn expression in breast tumours (Sewell et al., 2006). Furthermore, ST6GalNAc-I is overexpressed in intestinal metaplasia and associated with STn (Marcos et al., 2011). It is suggested that the overexpression of ST6GalNAc-I facilitates the increased abundance of STn by outcompeting other glycosyltransferases (Julien et al., 2012). After the formation of the core structure, the glycan chains are further elongated through addition of repeating units of type 1 or type 2 N-acetylglucosamine (LacNAc) to variable lengths, forming a backbone region. Most commonly, type 2 repeated poly-N-acetylglucosamine chains are formed through the action of  $\beta$ 1,3-acetylglucosaminyltransferases and  $\beta$ 1,4-galactosyltransferases (Fig. 5) (Brockhausen, 1999).

Finally, the backbone is terminated by a variety of units, including D-galactose (Gal), sialic acid and L-fucose (Fuc) (Juge, 2012). These terminal epitopes show substantial variation and provide a large source of glycan diversity (Tailford et al., 2015a). The expression of terminal ABO histo-blood group antigens in mucin glycans is determined by the secretor status (expression of an  $\alpha$ 1,2-fucosyltransferase (FUT2) gene) of the host, and provides one source of variation between individuals (Rausch et al., 2011; Tailford et al., 2015b). Further factors adding to this diversity includes the ability of glycans to be added in a linear or branched manner, and in a structure from 1 up to 20 residues long. MUC2 of the human sigmoid colon presents over 100 different complex oligosaccharides including mono-, di- and tri-sialylated, accounting for 80% of total mucin mass (Larsson et al., 2009). This large repertoire of O-glycan diversity is thought to be critical in maintaining a high bacterial binding capacity (Arike and Hansson, 2016).

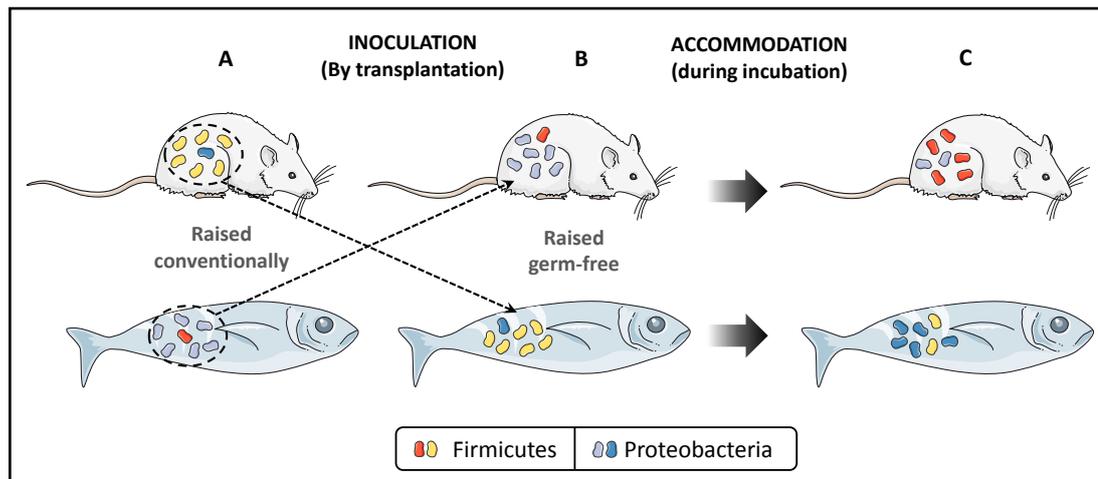


**Figure 5| Glycan core structures and glycosyltransferases involved in GI mucin glycosylation**

Mucin glycosylation is highly heterogeneous, with a regio-specific pattern at different mucosal surfaces throughout the body (Corfield, 2015; Linden et al., 2008). This phenomenon is also known to occur within different regions of the GI tract. In blood group matched individuals, it has been observed that there is a decreasing gradient of Fuc and an increasing acidic gradient in human intestinal mucins from ileum to rectum. The gradient of sialic acid was found to be accompanied by varying degrees of O-acetylation, with the level being especially high in the rectum (Robbe et al., 2003). The decreasing gradient of Fuc reflects the increased abundance of ABO histo-blood group antigens in the proximal intestine (Larsson et al., 2013). In addition, it was shown that the abundance of core structures was variable between different regions of the GI tract. For example, in the human SI, glycans were abundant in Fuc and these were predominantly formed from core 4 structures, whereas core 2 carrying sulpho-Lewis<sup>x</sup> were mainly found in the distal colon (Robbe et al., 2004).

Interestingly, in mice, the opposing gradients can be observed, with sialylation being dominant in ileum, jejunum and duodenum, but fucosylation being greater in the colon (Larsson et al., 2013). These species specific patterns in glycosylation may play an important role in selecting for the most functionally beneficial microbiota within specific regions of the GI tract (Larsson et al., 2013). Glycosylation gradients are not present in the foetus, suggesting they are acquired after birth due to alterations in the expression of glycosyltransferases (Robbe-Masselot et al., 2009). These changes in glycosylation patterns are accompanied by alterations in the abundance of gut microbes, suggesting that the glycans contribute to the selection of the GI mucus-associated microbiota (Rokhsfat et al., 2016).

Whilst data suggest heterogeneity between tissues, it appears that O-glycosylation patterns in the colon are relatively conserved between individuals. Using a sensitive nano-liquid chromatography/mass spectrometry (LC/MS) approach, it was shown that the structure and abundance of O-glycans from mucins purified from sigmoid colonic biopsies were relatively similar between 25 patients. This conserved pattern was characterised by predominantly sialylated and/or sulfated core 3 glycans (Larsson et al., 2009). In the gastric mucosa, however, it has been observed that mucin glycans always carry antigens relating to the blood group of the individual (Rossez et al., 2012). It is proposed that the homogeneity seen between colonic mucin glycans from different individuals allows for the selection of a 'core' microbiota. Organ to organ, and species to species variation in mucin glycosylation is suggested to provide a selective landscape that supports the most mutually beneficial relationship between host and bacteria (Johansson et al., 2011a). The type of mucin O-glycosylation is dependent on the glycosyltransferases expressed and where in the golgi apparatus they are located (Arike and Hansson, 2016). The contribution of host genetics to the selection of the gut microbiota is supported by the fact that, following transplantation with microbiota from a zebrafish, germ-free mice select for and adjust their colon microbiota to that more typical of wild type mice (Fig. 6) (Rawls et al., 2006).



**Figure 6 | Diagram depicting transplantation of microbiota between conventionally and germ-free raised mice and zebrafish**

Figure adapted from McFall-Ngai, 2006.  
(Servier Medical Art, 2016)

However, it is important to note that this is a bi-directional phenomenon as gut commensal bacteria have the ability to remodel the mucin glycan epitopes to their advantage. For example, *Ruminococcus gnavus* E1 was shown to be able to significantly induce the expression of glycosyltransferases and mucins in the mucosa of mono-associated mice. Specifically, *R. gnavus* colonisation led to a significant increase in the expression of *ST6Gal-I*, *C1GalT1*, *Muc1* and *Muc2* (Graziani et al., 2016). In the SI of germ-free mice, *Bacteroides thetaiotaomicron*, an abundant member of the commensal flora, is capable of increasing the expression of Fuc, and specifically Fuc $\alpha$ 1,2Gal- containing glycans (Hooper et al., 1999). Furthermore, a soluble factor secreted by *B. thetaiotaomicron* is able to increase the galactosylation of the mucus-producing intestinal cell line HT29-methotrexate (HT29-MTX) (Freitas et al., 2001). Varyukhina and coll. observed decreases in cell-surface sialic acid and  $\alpha$ -linked mannose/Fuc upon incubation of HT29-MTX cells with *B. thetaiotaomicron* spent culture media, as well as increased expression of galactosyltransferases. It was therefore suggested that the decrease in sialic acid, mannose and Fuc was due to masking of these residues by the increased expression of Gal (Varyukhina et al., 2012). The spent media of a specific strain of *Lactobacillus casei* DN-114 001 is also capable of increasing cell surface galactosylation as well as sialic acid, whilst decreasing GalNAc expression (Freitas et al., 2003). In addition, cultures of HT29-MTX cells which had been incubated with spent media with *B. thetaiotaomicron* and *L. casei* were resistant to infection by rotavirus (Varyukhina et

al., 2012). It is proposed that alterations to mucin glycosylation, such as those described above affect the ability of other commensals or pathogens to colonise, potentially giving some commensal species a competitive advantage in the gut (Freitas et al., 2001).

### 1.2.3 Other mucus components

Mucus is also an important reservoir of other non-mucin proteins in the GI tract. Studies of GI mucus from mice revealed a proteome consisting of ~1,300 proteins, which whilst similar in repertoire throughout the GI tract, displays significant quantitative variation between intestinal sites (Rodriguez-Pineiro et al., 2013). Amongst the proteins associated with the GI mucus are antimicrobials, including angiogenin 4,  $\alpha$ -defensins, cathelicidins, collectins, histatins, lipopolysaccharide (LPS)-binding protein, lysozymes, secretory phospholipase A2, and lectins such as REGIII $\alpha/\gamma$ , which are produced primarily by Paneth cells (McGuckin et al., 2011). The importance of antimicrobials such as these has been shown in patients with ileal Crohn's disease (CD), who exhibit reduced mucosal  $\alpha$ -defensin expression (Wehkamp et al., 2004; Wehkamp et al., 2005). Antimicrobial proteins appear to be localised to the mucus layer, and are virtually absent from the lumen, probably either due to poor diffusion through the mucus layer or luminal degradation (Hooper and Macpherson, 2010; Meyer-Hoffert et al., 2008). Retention of antimicrobials can be mediated through direct interactions with the mucin proteins. For example, in the saliva both MUC5B and MUC7 form complexes with histatins and statherins (Bruno et al., 2005; Iontcheva et al., 1997). Many of the antimicrobial proteins secreted into mucus kill bacteria through direct interaction with, and disruption of the bacterial cell wall or inner membrane *via* enzymatic attack (Hooper and Macpherson, 2010).

Another component of the intestinal mucus is the secretory antibody, SIgA (McGuckin et al., 2011). SIgA has been shown to co-localise with gut bacteria in the outer mucus layer (Rogier et al., 2014). Whilst SIgA assists in limiting the exposure of the epithelial cell surface to bacteria, Rogier and coll. showed that it was not essential for excluding bacteria from the inner mucus layer, whilst Muc2 was (Rogier et al., 2014). Rather, SIgA may be important in shaping the composition of the gut microbiome (Rogier et al., 2014). In particular, in the outer mucus layer SIgA is proposed to mediate bacterial biofilm formation *via* binding to SIgA receptors on bacteria (Bollinger et al., 2003). In IgA-deficient individuals the expression of these receptors by bacteria is lower (Friman et al., 1996). Dysbiosis of the microbiota, in particular an overrepresentation of segmented filamentous bacteria (SFB) occurs in mice deficient in IgA, an effect that may be particularly damaging to the host due to the ability of

SFB to strongly adhere to the epithelium and activate the immune system (Suzuki et al., 2004).

#### 1.2.4 Mechanisms of mucus regulation

Mucins are constitutively expressed in order to renew the mucus layer, which is subject to constant degradation by luminal bacteria (McGuckin et al., 2011). However, a range of host, bacterial and environmental derived stimuli are capable of regulating the synthesis and release of mucin, generating a dynamic mucosal barrier (McGuckin et al., 2011).

At the host level, infection at the mucosal surface can stimulate the rapid exocytosis of mucins stored in granules to enhance the mucus barrier. The NOD-like receptor family pyrin domain containing 6 (NLRP6) inflammasome is essential in orchestrating this process, and deficiency in the NLRP6 protein results in an impaired mucin secretion, mucus layer formation, increased bacterial-epithelial contact and increased susceptibility to infection (Wlodarska et al., 2014). Furthermore, a large number of host cells, including immune and epithelial cells secrete bioactive molecules known as cytokines, upon stimulation by pathogen associated molecular patterns (PAMPs). T helper type 1 (Th1) cytokines, generally involved in immunity at the cellular level, and type 2 (Th2), which participate in humoral immune responses, have been implicated in the modulation of mucin production (Andrianifahanana et al., 2006; Smirnova et al., 2001). Amongst the Th1 cytokines, tumour necrosis factor (TNF)- $\alpha$  has been the main cytokine shown to regulate mucin production in intestinal models (induction of MUC5AC and MUC5B secretion). The main Th2 cytokines which regulate mucin production include interleukins IL-4 and IL-13 (upregulation of *MUC2* and *MUC5AC* gene expression), IL-6 (upregulation of *MUC2*, *MUC5B* and *MUC6* gene expression) and IL-22 (upregulation of *MUC1*, *MUC3*, *MUC10* and *MUC13* gene expression) (Enss et al., 2000; Iwashita et al., 2003; Sugimoto et al., 2008). IL-10, a further Th2 cytokine, is involved in the modulation and maintenance of the mucus layer. Knockdown of IL-10 results in ER stress, misfolding of Muc2 and spontaneous intestinal inflammation, characteristic features of IBD (Hasnain et al., 2012). Generally, the modulation of mucin expression by cytokines occurs *via* the janus kinase/signal transducers and activators of transcription (JAK/STAT) pathway through the dimerisation and translocation of STAT to the nucleus where it regulates transcription (Theodoropoulos and Carraway, 2007).

In addition to shaping mucin glycosylation (See section 1.2.2.3), commensal bacteria can regulate the expression of mucins. For example, the commensal species *Akkermansia muciniphila* stimulates mucin synthesis, facilitating turnover of mucins and enhancing barrier

function (Derrien et al., 2016). In contrast, pathogenic bacteria and their products can alter mucin expression in order to facilitate invasion of host tissue. The gastric ulcer-causing bacterium, *H. pylori* is capable of modulating not only mucin expression but also mucin properties to provide itself with an advantage in the gastric environment. In the colonic goblet cell line HT29-CL.16E, *H. pylori* resulted in significant inhibition of mucus exocytosis from granules (Micots et al., 1993). In addition, the bacterium can alter the pH and viscoelastic properties of the mucus layer allowing improved mobility through the mucus layer (Celli et al., 2009). Furthermore, purified bacterial products such as LPS and flagellin have been shown to regulate mucins by inducing expression through the Ras pathway, leading to enhanced mucin transcription (McNamara and Basbaum, 2001; Smirnova et al., 2003; Theodoropoulos and Carraway, 2007). The ability of bacteria to modulate mucus thickness has been demonstrated in germ-free mice, where the inner adherent mucus layer, which is usually thin in comparison to conventionalized mice can be restored upon stimulation by bacterial products such as LPS or peptidoglycan (Petersson et al., 2011). In addition, short chain fatty acids (SCFAs) produced by bacterial fermentation in the colon are proposed to influence mucin expression. In a number of different colonic cell lines, it has been demonstrated that the SCFA, butyrate, is capable of modifying the production of several different mucins including MUC2, MUC3, MUC5AC and MUC5B (Burger-van Paassen et al., 2009; Gaudier et al., 2004; Hatayama et al., 2007).

### 1.2.5 Functional role of mucus in the GI tract

#### 1.2.5.1 *Mucus as a physical barrier*

Mucus has traditionally been recognised for its physical function to act as a lubricant facilitating, for example, the low-friction passage of solids in the intestine. This property is primarily mediated by hydrophobic interactions between phospholipids at the mucus surface, and the luminal contents (Gibson and Muir, 2005).

Another important function for mucus is its role as a protective barrier. In the stomach and duodenum, the mucus layer is responsible for protecting the epithelium from damage through contact with the harsh and acidic luminal environment. This is proposed to occur through the secretion and trapping of bicarbonate by the stable, inner mucus layer at the cell surface (Allen and Flemstrom, 2005). This results in the generation of a neutral pH at the epithelium, and multiple lines of evidence support the presence of a pH gradient across the mucus layer (Flemstrom and Kivilaakso, 1983; Kivilaakso and Flemstrom, 1984; Quigley and Turnberg, 1987; Williams and Turnberg, 1981). In the SI and colon, mucus aids in the physical

separation of luminal contents from the epithelium, acting as a critical contributor to the innate immune system. The mucus in these regions acts as a reservoir for antimicrobial proteins, such as defensins, cathelicidins and lysozyme (See section 1.2.3) (Antoni et al., 2013). RegIII $\gamma$ , a C-type lectin that is produced by Paneth cells in the SI specifically targets gram-positive bacteria and is essential in maintaining a  $\sim 50 \mu\text{m}$  zone separating gut microbiota from the SI surface epithelium (Vaishnava et al., 2011). In the colon, the protective role of mucus is demonstrated by the impenetrability of the inner layer to beads the size of bacteria. In mice, fluorescently labelled beads were maintained at a distance of  $\sim 200 \mu\text{m}$  from the epithelium by this inner layer (Ermund et al., 2013). This stratification of the mucus was also observed in humans, where the fluorescent beads were maintained at a distance of  $\sim 400 \mu\text{m}$  from the epithelium in colonic biopsies (Johansson et al., 2013a). Mice deficient in Muc2 were observed to develop signs of colitis at 5 weeks of age (Van der Sluis et al., 2006). This suggests that Muc2 deficiency results in a breach in the gut barrier, exposing the epithelium to contact with bacteria (Van der Sluis et al., 2006). A similar pattern has been observed in ulcerative colitis (UC) patients, where the inner mucus layer is penetrable to beads the size of bacteria (Johansson et al., 2013a). Furthermore, in patients with inflamed mucosa, mucus thickness has been shown to be depleted, accompanied by bacterial-epithelial adherence and an increase in leucocytes (Swidsinski et al., 2007). However, it is becoming increasingly evident that mucus also plays a critical biological role (in addition to the physical properties) in shaping the gut environment.

#### 1.2.5.2 *Mucus as a biological habitat*

Mucus is vital in mediating a relationship with gut commensal bacteria, allowing the beneficial properties of the microbiota, such as protection against pathogens, to be harvested, and helping to maintain gut homeostasis. This biological role is less well understood and has only become a focus of research in the last ten years. As discussed previously (See sections 1.2.1 and 1.2.2.3), the colonic mucus is divided into two layers and the outer layer with its expanded volume allows bacteria to penetrate. The glycan rich mucus provides an energy source and preferential binding sites for the commensals (Johansson et al., 2011a; Juge, 2012; Tailford et al., 2015a). The presence of bacteria in the outer mucus layer was first supported in mice using FISH staining (Johansson et al., 2008). A recent study using 16S sequencing showed that the outer mucus layer harboured a unique niche in which bacterial species display different patterns of proliferation and utilisation of resources compared to their counterparts in the lumen (Li et al., 2015). This study highlighted the role of the mucus layer for providing binding sites for commensal bacteria. The composition of

the mucus is likely to be key in mediating this special relationship as it allows for the selection of the most optimal microbial species to mediate functions such as the competitive exclusion of pathogens and modulation of the immune system (See section 1.2.2.3) (Arpaia et al., 2013; Furusawa et al., 2013; Zarepour et al., 2013). It is believed that mucin-degrading bacteria are keystone species of this habitat (See section 1.4.2.2) (Tailford et al., 2015a).

#### 1.2.5.3 *Mucus as an immune mediator*

As well as the role of mucus in hosting immune components, there is evidence to suggest that mucus secreting goblet cells and mucins can also shape the immune system. Recently, goblet cell-associated antigen passages (GAPs) were identified as a novel mechanism by which small soluble luminal antigens (< 70 kDa) could be delivered to CD103+ LP dendritic cells (DCs) (McDole et al., 2012; Miller et al., 2014). In the SI, GAP formation is regulated by level of production of acetylcholine (Knoop et al., 2015). On the other hand, in the colon, GAP formation is regulated by responsiveness of goblet cells to acetylcholine, mediated through myeloid differentiation primary response gene 88 (Myd88)-dependent microbial sensing (Knoop et al., 2015). GAP formation also appears to be associated with the type of goblet cell mucus secretion. For example, GAP formation does not occur during cholera toxin-induced secretion, where goblet cells secrete *via* primary exocytosis (Knoop et al., 2015). This indicates that antigen delivery to the immune system may be prevented in unfavourable conditions, whilst still maintaining the mucus barrier (Knoop et al., 2015). In terms of mucin proteins themselves, in the SI, it has been demonstrated that MUC2 can deliver immunomodulatory signals to DCs and intestinal epithelial cells, conditioning their responses towards gut microbiota, and inducing tolerance (Shan et al., 2013). Furthermore, in *Muc2* <sup>-/-</sup> mice, gavage of MUC2 enhanced tolerance to dextran sulfate sodium (DSS) administration (Shan et al., 2013). Mucus and epithelial cell interaction with the immune system was recently reviewed by Pelaseyed and coll. (Pelaseyed et al., 2014). However, our understanding of the role of mucus in shaping the immune system is still in its infancy and is currently under further investigation.

## 1.3 The GI microbiota

Humans are colonised by a dynamic community of microorganisms, covering all three domains of life: bacteria, archaea and eukarya. This collection of organisms is termed the 'microbiota' and is present in many bodily sites, including the oral and nasal cavities, vagina, skin and GI tract (Backhed et al., 2005; Hattori and Taylor, 2009; McFall-Ngai, 2006). Whilst in some regions of the body, such as the brain endothelium (van Sorge and Doran, 2012), contact with prokaryotes would seriously compromise host health, in these locations host-microbial contact has coevolved over thousands of years to form a complex relationship (Neish, 2009). The number of organisms living in the gut alone is in excess of  $10^{14}$ , which encompasses ~10 times more bacterial cells than the number of human cells and over 100 times the amount genomic content (microbiome) as the human genome (Backhed et al., 2005; Gill et al., 2006). However, a recently revised estimate has suggested that the ratio of human : bacterial cells is closer to 1:1 (Sender et al., 2016). As a result of the vast number of bacterial cells in the body, the host and the microorganisms inhabiting it are often referred to as a 'superorganism' (Gill et al., 2006; Luckey, 1972).

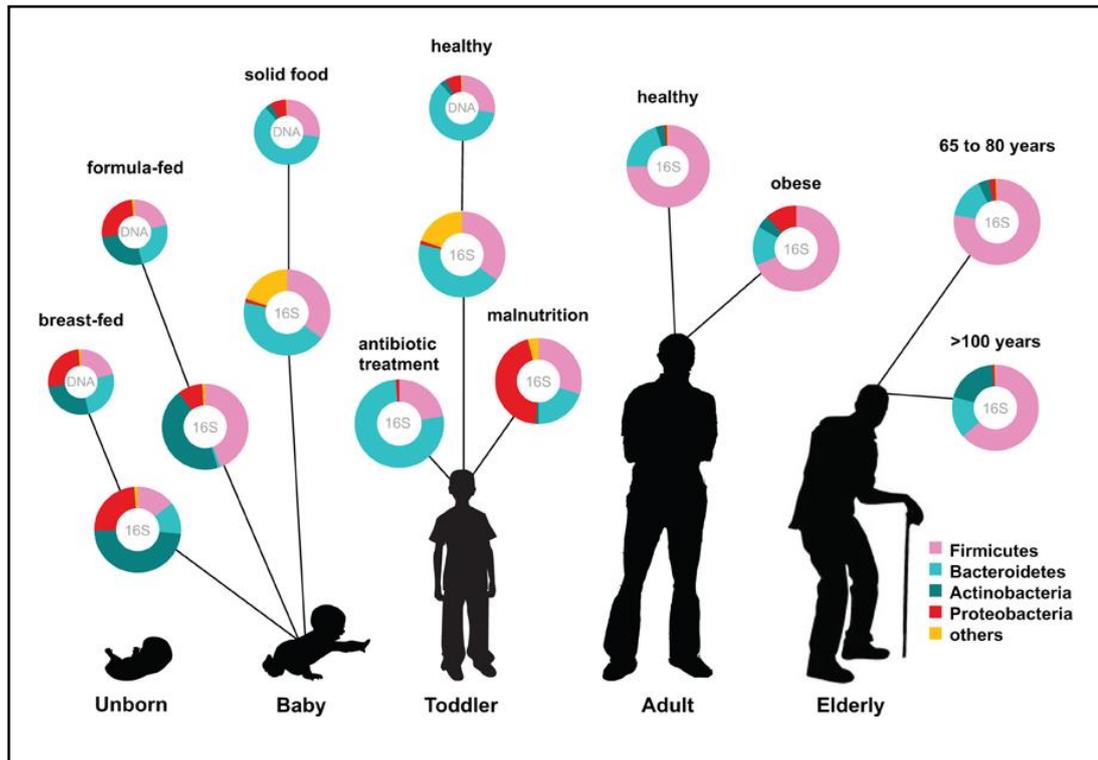
### 1.3.1 Development and composition of the human GI microbiota

#### 1.3.1.1 *Development of the GI microbiota from childhood*

In general, it is accepted that the development of the microbiota begins from birth, although there is some debate over whether microbiota colonisation begins in the uterus, since a limited number of studies have detected microbes in womb tissues, such as the placenta (Aagaard et al., 2014), as recently reviewed (Rodriguez et al., 2015).

Following birth, the GI tract is rapidly colonised, with seemingly chaotic shifts in the composition which can be attributed to life events such as illness, antibiotic treatment, mode of delivery and changes in diet (Fig. 7) (Koenig et al., 2011; Rodriguez et al., 2015). In vaginally delivered infants, during the first few days, the microbiota is colonised by a high abundance of lactobacilli, reflecting the high load of lactobacilli in the vaginal flora (Aagaard et al., 2012; Avershina et al., 2014). In infants delivered by C-section, the microbiota is depleted in, and delayed in the colonisation of the *Bacteroides* genus, and is colonized by facultative anaerobes such as *Clostridium* species (Jakobsson et al., 2014; Salminen et al., 2004). Whilst 72% of the early microbiota of vaginally delivered infants resemble that of their mothers faecal microbiota, this percentage is reduced to only 41% in babies delivered by C-section (Backhed et al., 2015).

During early stages in development, the microbiota is relatively low in diversity and dominated by two main phyla, Actinobacteria and Proteobacteria (Fig. 7) (Backhed, 2011; Rodriguez et al., 2015). Over the first year of life the microbial community converges towards a distinct adult-like profile. This transition is marked by an increased in microbial diversity and temporal patterns that are unique to each infant (Palmer et al., 2007). At around 2.5 years of age, the composition and diversity of the infant microbiota resembles that of an adult gut microbiota, with many of the same functional capabilities (Koenig et al., 2011; Rodriguez et al., 2015). In adulthood, although the gut microbiota composition is relatively stable, it is still subject to perturbation by life events (Dethlefsen and Relman, 2011). In the ageing population, there are apparent shifts in the microbial community, marked by an increased abundance of bacteria belonging to the phyla Bacteroidetes, and a significant relationship between microbial diversity and living arrangements, such as community dwelling or long term residential care (Claesson et al., 2011; Claesson et al., 2012). In contrast, a separate study observed few differences between in the elderly population (70 years), and a young cohort (Biagi et al., 2010). However, in centenarians, not only was the species diversity reduced, but the abundance of facultative anaerobes such as *Escherichia coli*, *Klebsiella pneumoniae* and *Pseudomonas* was increased (Biagi et al., 2010). It has also been shown that the capacity of the microbiota to carry out amylolysis and produce SCFAs is reduced in the elderly, whilst proteolytic activity is increased (Woodmansey et al., 2004).



**Figure 7 | Phylum level composition of the gut microbiota throughout life**

(Ottman et al., 2012)

### 1.3.1.2 Composition and variability of the human GI microbiota

Around a decade ago, most of our knowledge about the adult human gut microbiota stemmed from labour intensive culture based methods (Moore and Holdeman, 1974). More recently, the advent of culture-independent approaches such as high-throughput and low cost sequencing methods has greatly improved our ability to survey the breadth of the gut microbiota. A common approach involves specific targeting of the bacterial 16S ribosomal RNA (rRNA) gene (Poretsky et al., 2014). This gene is a landmark present in all bacteria and archaea, and can be used to distinguish between species due to its nine hypervariable regions (V1-V9), which demonstrate considerable diversity. Earlier studies utilising this technique focussed on sequencing the entire 16S rRNA gene. Using this method, Suau and coll. highlighted the extreme insensitivity and bias of culturing methods, observing that 76% of the rRNA sequences obtained from an adult male faecal sample belonged to novel and uncharacterised species. The output of this study gave a gut microbiota consisting of 82 distinct species (Suau et al., 1999). During further more comprehensive investigations, this number has gradually increased. In 2005, a study by Eckburg and coll., which examined mucosal and faecal samples from six major colonic subdivisions of three healthy adult subjects, identified 395 bacterial phlotypes and a single archaeal phlotype ( $\leq 99\%$  sequence

identity threshold) from an analysis of 13,335 16S sequences (Eckburg et al., 2005). Recently, the sequencing of 16S rRNA has shifted to focus on analysing shorter sub-regions of the gene at greater depth (Mizrahi-Man et al., 2013), however the utilisation of this method, which generates shorter read lengths can introduce errors (Poretsky et al., 2014). Whole genome shotgun (WGS) metagenomics approaches may provide more robust estimates of microbiota composition and diversity due to higher resolution and sensitivity (Poretsky et al., 2014). Recently data generated using culture based experiments from the Human Microbiome Project (HMP) and metagenomics data from MetaHit have provided the most comprehensive view of the entire repertoire of microbes associated with humans (Hugon et al., 2015; Li et al., 2014). Compilations of data from literature and culture databases identified 2172 species that have been isolated from human beings, classified in to 12 different phyla. Of these, 93.5% species belonged to Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes, and three of the 12 identified phyla contained only one species isolated from humans. This included *A. muciniphila*, a species isolated from the gut, as the only known representative of the Verrucomicrobia phyla. Of the identified species, 386 are strictly anaerobic and hence will only usually be found in sites such as the skin, and mucosal regions such as the oral cavity and GI tract (Hugon et al., 2015).

Although evidently comprised of a large number of different species, the gut microbiota is not as diverse as some other bodily microbial communities, such as the skin on the forearm, palm and back of the knee, suggesting specific selection of a population which provides the most beneficial range of functions to the host (Costello et al., 2009; Schluter and Foster, 2012). The functional capacity of the human gut microbiome was recently extensively catalogued by the MetaHit study, resulting in the identification of 9,879,896 genes, established through a combination of 249 newly sequenced and 1018 published samples (Li et al., 2014). The study identified the presence of country specific microbial signatures, suggesting that gut microbiota composition is shaped by environmental factors, such as diet, and possibly also by host genetics (Li et al., 2014). This information is crucial for developing therapeutic strategies to modify and shape the microbial community in disease.

Microbiota composition in the GI tract is reflective of the physiological properties in a given region. The density of bacteria is affected by chemical, nutritional, and immunological gradients along the gut. In the SI, bacterial colonisation is limited by higher levels of acids, oxygen and antimicrobials, in contrast to the colon, where there is a greater bacterial load (Donaldson et al., 2016). The composition between different colonic mucosal sites of the

same individual demonstrates spatial conservation in terms of composition and diversity, a property which is maintained despite regional inflammation in colitis (Eckburg et al., 2005; Lavelle et al., 2015).

In contrast, the faecal/luminal and mucosal compositions are significantly different (Eckburg et al., 2005; Lavelle et al., 2015). For example, the abundance of Bacteroidetes appears to be higher in faecal/luminal samples than in the mucosa (Eckburg et al., 2005). In contrast, Firmicutes, specifically *Clostridium* cluster XIVa are enriched in the mucus layer compared to the lumen (Van den Abbeele et al., 2013). Similar findings have also been made in mice (Nava et al., 2011; Swidsinski et al., 2005). These observations highlight the need for careful consideration in choosing a sampling method when studying and reporting the microbiota.

Differences in the species and subspecies arrangement of the microbiota from one non-related host to the next are proposed to outweigh differences in the community arrangement within an individual. In the study by Eckburg and coll., the greatest amount of variability between mucosal and faecal samples of three healthy subjects was explained by inter-subject difference, a finding which has been replicated by several other studies (Eckburg et al., 2005; Jakobsson et al., 2010; Turnbaugh et al., 2009). Previously, suggestions have been made of the presence of a 'core microbiota', proposed to be a set of the same abundant organisms present in all individuals. Turnbaugh and coll. found that this core structure is better defined at a functional level, since they showed that there was a higher degree of similarity at the gene rather than organismal level (Turnbaugh et al., 2009).

Furthermore, although the specific taxonomic arrangement of the microbiota differs between individuals and bodily sites, it appears that these arrangements can be classified into 'community types' that can be predictive of each other and are associated with background (Ding and Schloss, 2014). Specifically, multidimensional analysis of 33 samples from different nationalities revealed the presence of 3 enterotypes identifiable by variations in the level of one of three genera: *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2) and *Ruminococcus* (enterotype 3) (Arumugam et al., 2011). However, evidence surrounding what may drive the formation of these enterotypes is controversial and there is even controversy over the existence of the enterotypes themselves, as thoroughly reviewed (Jeffery et al., 2012).

#### 1.3.1.3 *Factors shaping the GI microbiota*

The microbiota is subject to shaping by host selective pressures. Host fitness depends upon the range of essential functions provided by the microbial community, and hence an

individual microbes longevity is determined by whether it is contributing to these functions. 'Cheater' organisms who do not assist in providing beneficial functions are proposed to be controlled by, and may occasionally be purged during, for example, transferral of the microbiota to a new host (Ley et al., 2006a; Travisano and Velicer, 2004). In addition, the microbiota could be directly shaped by the host immune system. However, due to a high amount of functional redundancy in the microbiota, it is unlikely that the immune system is capable of driving the microbial community at the species level, and there is little evidence to support this (Ley et al., 2006a).

Gut microbes must be adapted to a certain type of lifestyle due to the relatively fewer number of biochemical niches available in the gut, compared with other microbially rich environments. In the gut, energy can generally be derived through processes such as fermentation and sulfate reduction of dietary and host carbohydrates, such as the mucin O-glycans. The organisms that can survive in the gut are therefore limited by their phenotypic traits (Ley et al., 2006a). This is evident from a study by Seedorf and coll., reporting that the success of colonization of the germ-free mouse gut by microbiota from diverse habitats is determined by their ability to metabolize dietary and host carbohydrates and bile acids (Seedorf et al., 2014).

Current research suggests that the greatest impact on microbiota composition is exerted by the diet (Donaldson et al., 2016). In early infancy, the abundance of some bacterial groups in the gut microbiota are affected by the feeding method. *Bifidobacterium longum* and several species of *Bacteroides* are capable of utilizing fucosylated oligosaccharides present in human milk, allowing them to outcompete other bacteria such as *Escherichia coli* and *Clostridium perfringens* (Marcobal et al., 2011; Yu et al., 2013). The abundance of *Bifidobacteria* spp. in breast fed infant microbiota is typically high (Yu et al., 2013). In formula fed infants, the microbiota appears to be more diverse, and contains fewer *Bifidobacterium* (Bezirtzoglou et al., 2011). Furthermore there are altered levels of other bacterial groups such as *E. coli*, *Clostridium difficile*, *Bacteroides fragilis* and lactobacilli (Bezirtzoglou et al., 2011; Favier et al., 2002; Penders et al., 2006). Undernourished children harbour immature and dysbiotic microbial communities containing an increased number of enteropathogens, including an enrichment of *Enterobacteriaceae* (Kau et al., 2015). Transmission of faecal microbiota from monozygotic Malawian twin pairs with severe undernutrition to germ-free mice resulted in significant weight loss and increased gut barrier permeability (Kau et al., 2015). In contrast, germ-free mice colonised with the microbiota of a healthy donor gain significantly more

weight than mice colonised with an undernourished microbiota (Blanton et al., 2016). The administration of certain bacterial species can either ameliorate or augment these effects. Mice receiving doses of *Clostridium scindens* and *A. muciniphila* prior to gavage with an undernourished microbiota exhibited significantly less mortality than those with no intervention (Kau et al., 2015). Furthermore, administration of two invasive species, *R. gnavus* and *Clostridium symbiosum* to the microbiota of undernourished donors restored abnormalities in growth and metabolic function in animals receiving undernourished microbiota transplants (Blanton et al., 2016).

Shaping of the gut microbiota is subject to the availability of microbiota accessible carbohydrates (MACs) which are found in dietary fibre. In infants from rural Africa, with a diet dominated by starch, fibre and plant polysaccharides, the microbiota is abundant in microorganisms belonging to the Actinobacteria (10.1%) and Bacteroidetes (57.7%) phyla (De Filippo et al., 2010). In European children, whose diet is rich in sugar, starch and animal protein, these figures are just 6.7% and 22.4% (De Filippo et al., 2010). Some SCFA-producing bacteria, such as *Prevotella* were found exclusively in the African children, possibly contributing to protection against invasion by pathogenic microbes and inflammation, and allowing maximal energy extraction from the ingested plant polysaccharides (De Filippo et al., 2010). This trend was also reflected in individuals consuming high amounts of carbohydrates and simple sugars (Wu et al., 2011). The abundance of MACs are substantially reduced in the Western diet. A loss of microbial diversity is observed in mice fed a low MAC diet, when compared to a control group. It was recently shown that as new generations of low MAC diet mice were bred, there was increasing divergence from the microbiota of controls and a decreasing glycoside hydrolase capacity. Restoration of diversity required administration of MACs in combination with the bacterial taxa that were missing (Sonnenburg et al., 2016).

The ability of microbes to cooperate and share host resources allows colonisation by a more diverse set of organisms. Multiple studies suggest that gut microbes cooperate *via* cross-feeding mechanisms, as the metabolic products of some organisms provide substrate to support the growth of others. For example, lactate, a major product produced by lactic acid bacteria including lactobacilli and bifidobacteria, can be utilised and converted to butyrate by other microbiota members such as *Eubacterium hallii* and *Anaerostipes caccae* (Duncan et al., 2004b; Louis et al., 2007). Modification of an individual organisms genome can occur through, for example, mutation or lateral gene transfer, allowing expansion and diversification of the population and the exploitation of new niches (Bjedov et al., 2003; Xu

et al., 2007). In turn, diversification introduces new bacterial functions and promotes further niche variation, creating a positive feedback loop in which more diversification can occur (Emerson and Kolm, 2005; Svanback and Bolnick, 2007).

A further environmental factor which plays a key role in shaping the gut microbiota is the presence of antimicrobials. Treatment with antibiotics dramatically disrupts the microbial balance both short and long-term, including decreases in the richness and diversity of the community. Examples of antibiotic treatments which have been demonstrated to affect the community structure include clindamycin (long term impacts of up to two years) (Jernberg et al., 2007), clarithromycin and metronidazole (long term perturbations for up to four years) (Jakobsson et al., 2010), and ciproflaxin (effects still apparent after one week) (Dethlefsen and Relman, 2011). The exact effects and the length of time for recovery of the microbiota following administration of antibiotics appears to be individual dependent, a likely effect of the inter-individual variation in the microbiota prior to treatment (Dethlefsen and Relman, 2011; Jakobsson et al., 2010; Jernberg et al., 2007). Antimicrobials localised to the mucus layer, such as host derived  $\alpha$ -defensins and SIgA are also able to shape the microbiota (See section 1.2.3).

Several other environmental factors have also been found to be associated with microbiota composition including, but not limited to, geographical location, surgery, smoking, depression and living arrangements (urban or rural) (Biedermann et al., 2013; Jiang et al., 2015; Rodriguez et al., 2015; Tyakht et al., 2013).

### 1.3.2 Role of the gut microbiota in health

#### 1.3.2.1 *Beneficial role of microbes*

Along with its large amount of genomic content, the gut microbiota confers a range of beneficial properties to the host. These include extended metabolic and digestive abilities, such as digestion of complex plant polysaccharides, and production of SCFAs.

Three SCFAs dominate in the human GI tract, propionate, butyrate and acetate. The ratio in which these SCFAs are found in the GI tract is determined by the composition of the microbiota (Burger-van Paassen et al., 2009), however they are typically found in a ratio of 1:1:3 (propionate:butyrate:acetate) (Louis et al., 2014). All three metabolites are rapidly absorbed by epithelial cells in the GI tract where they are involved in the modulation of cellular processes such as gene expression, chemotaxis, differentiation, proliferation and

apoptosis (Correa-Oliveira et al., 2016). Butyrate appears to be a particularly important source of energy for colonocytes (Correa-Oliveira et al., 2016). It is proposed that butyrate mediates intestinal epithelial turnover and homeostasis due to its decreasing gradient of concentration from lumen-to-crypt, which promotes colonocyte proliferation at the bottom of crypts, whilst increasing apoptosis and exfoliation of cells closer to the lumen (Donohoe et al., 2012). Furthermore, SCFAs are known to modify the production of cytokines, including increases in IL-18, which is involved in maintaining and repairing epithelial integrity (Correa-Oliveira et al., 2016).

The GI microbiota is also crucial to the synthesis of essential vitamins such as vitamin B12 and folate (Albert et al., 1980; Hill, 1997; Hooper et al., 2002; Kripke et al., 1989; Whitehead et al., 1986). This enhanced metabolism has been demonstrated in mice raised with and without microbes, the latter having 40% less body fat despite an increased appetite (Backhed et al., 2004). In addition, the microbiota plays a crucial role in maintaining epithelial homeostasis through mechanisms such as pathogen exclusion, toll-like receptor (TLR) signalling, and in shaping and regulating the immune system *via*, for example, cytokine induction (Christensen et al., 2002; Rakoff-Nahoum et al., 2004; von der Weid et al., 2001).

The relationship of the microbiota with the host is termed 'mutualistic', as in return the host provides the microbial community with an immediate stable environment and plentiful supply of nutrients (Neish, 2009). Much like the microbiota shapes the host immune system, the immune system controls and shapes the microbiota through, for example, stratification and compartmentalization to the lumen. This is vital in order to avoid opportunistic invasion of host tissue resulting in disruption of intestinal homeostasis (Cash et al., 2006; Hooper et al., 2012; Macpherson et al., 2000; Macpherson and Uhr, 2004).

#### 1.3.2.2 *Dysbiosis and disease*

When studying the microbiota, a deviation away from normal intestinal homeostasis is generally referred to as 'dysbiosis'. The microbiota has been studied in a number of disease states where a divergence from the normal composition has been found. In obesity, for example, several studies involving obese Americans and *ob/ob* genotype mice have found that the microbiota contains a higher ratio of Firmicutes to Bacteroidetes phyla compared to lean individuals (Ley et al., 2005; Ley et al., 2006b; Turnbaugh et al., 2009; Turnbaugh et al., 2006). This ratio was shown to be diminished following diet and bariatric surgery induced weight loss (Furet et al., 2010; Ley et al., 2006b). Obesity-associated metabolic phenotypes were transmissible to germ-free mice from the faecal communities of twin pairs discordant

for obesity. However cohousing recipient mice with microbiota from lean and obese donors prevented the obesity phenotype, which was associated with an increase in Bacteroidetes phyla transmitted from the lean to the obese mice (Ridaura et al., 2013). These studies point toward a depletion of Bacteroidetes phylum in obesity, however, this association has not been shown in all studies of obese individuals (Duncan et al., 2008; Jumpertz et al., 2011; Schwiertz et al., 2010). Recently it was shown that in individuals with low bacterial richness, groups such as Bacteroides and members of the Lachnospiraceae family such as *R. gnavus* and *Ruminococcus torques* were more dominant. Low bacterial richness is generally associated with increased adiposity, insulin resistance and dyslipidaemia (Le Chatelier et al., 2013). In contrast, in those with high bacterial richness, Faecalibacterium, Bifidobacterium, Lactobacillus and Akkermansia were some of the genera that were more dominant (Le Chatelier et al., 2013). Despite *ob/ob* mice having a microbiome with an increased functional capacity to harvest energy from the diet, the abundance of Firmicutes and Bacteroidetes does not appear to be correlated long-term with calorie intake, as determined by measuring the energy and SCFA content in faecal samples (Turnbaugh et al., 2008; Turnbaugh et al., 2006). In mice fed with a high fat diet, the energy intake was not significantly different from mice fed a normal diet, despite observing a shift in abundance of Firmicutes and Bacteroidetes (Murphy et al., 2010).

In contrast to obesity, the abundance of the Firmicutes phyla appears to be decreased in type 2 diabetes when compared with non-diabetics, whilst the abundance of Bacteroidetes and Proteobacteria is increased (Larsen et al., 2010). Moreover, the ratio of Bacteroidetes to Firmicutes was found to be significantly correlated with a lack of tolerance to glucose (Larsen et al., 2010). However, once again there is contradictory evidence surrounding these trends. In a study of three groups of individuals classified based on tolerance of glucose it was shown that intolerance is associated with the abundance of several taxonomic groups, including a depletion in the abundance of Bacteroides in individuals with type 2 diabetes (Zhang et al., 2013). Furthermore, in individuals classified as 'pre-diabetic' based on their glucose intolerance status, a lower abundance of butyrate producing species, including *Faecalibacterium prausnitzii* could be observed when compared to a group of patients with normal glucose tolerance (Zhang et al., 2013). Together, these findings cast doubt on the validity of using the Firmicutes and Bacteroidetes ratio as a predictor of energy intake and obesity.

The microbiota also appears to be distorted in irritable bowel syndrome (IBS), however these findings are complicated by the multiple subtypes of disease (diarrhoea predominant, IBS-D, constipation predominant, IBS-C, or mixed bowel habit, IBS-M) (Kassinen et al., 2007). An initial investigation of IBS performed in 2005 found alterations in the abundance of *Clostridium coccooides* and *Bifidobacterium catenulatum*, and also a significantly lower amount of *Lactobacillus* spp. in IBS-D compared to IBS-C (Malinen et al., 2005). Furthermore a decrease of *Bacteroides* and *Actinobacteria* and increase in *Firmicutes* has also been observed in IBS (Krogus-Kurikka et al., 2009; Noor et al., 2010). However, as with the other disorders discussed here, the results are not always consistent in other studies (Carroll et al., 2010).

An altered microbiota has also been implicated in a plethora of other diseases, including vaginosis, circulatory disease, allergy, asthma, cardiovascular disease and autism (Carding et al., 2015; Kinross et al., 2011; Kuczynski et al., 2012). Dysbiosis of the microbiota in IBD has been well studied and documented, using a range of techniques and sampling methods, as detailed in section 1.4.2.2. Therefore, although the exact changes in microbiota composition in health and disease remain unclear, it is obviously an important factor that must be taken into consideration.

Our understanding of what constitutes a balanced or imbalanced microbiota is limited by the complexity of the community and the variation in the species arrangement from one individual to the next. A microbiota in dysbiosis may be better described on a person by person (profiled over periods of health and disease) or functional (microbiome) basis. At a functional level, a potential way to describe a 'dysbiotic' microbiota might be one which fails to provide the host with the full complement of beneficial properties it requires (Willing et al., 2010b). In addition we have little understanding of whether changes in the community structure of the microbiota are a causal or consequential factor in disease, and so emphasis must be made on studying the underlying molecular mechanisms driving these changes.

## 1.4 Inflammatory Bowel Disease (IBD)

### 1.4.1 Definition and epidemiology of IBD

IBD defines a complex range of disorders that broadly incorporates ulcerative colitis (UC) and Crohn's disease (CD). Symptoms include severe and repetitive inflammation of the GI tract and is associated with an alteration of the normal gut flora and activation of the immune system (Thompson-Chagoyán et al., 2005). Whilst UC is characterized by a continuously inflamed and non-granulomatous mucosa generally restricted to the colon, CD can affect the entire GI tract, and is clinically recognised by segmented and granulomatous transmural inflammation (Nell et al., 2010). If uncontrolled, IBD can result in a significantly increased risk of developing colorectal cancers (Sheng et al., 2012).

Growing rates of IBD occurrence are placing increasingly large burdens on healthcare systems. In 2013, it was estimated that 0.3% of the European population was affected by IBD, equating to around 2.5-3 million people (Burisch et al., 2013). Although currently, incidence and prevalence of IBD is highest in developed countries (in Europe, the United Kingdom and Scandinavia), the occurrence of these diseases is increasing all over the world (Burisch et al., 2013). The costs of providing healthcare to IBD patients is calculated at 4.6-5.6 billion euros per year (Burisch et al., 2013). These costs are exacerbated by extra-intestinal manifestations affecting the eye, joints and skin that relatively frequently accompany the disease. In 2003, in a cohort of over 800 patients, 21.3% of IBD sufferers reported incidences of extra-intestinal manifestations, including ocular, cutaneous, hepatobiliary and joint problems (Lakatos et al., 2003).

### 1.4.2 Aetiology of IBD

The exact factors leading to IBD remain elusive. Recent descriptions suggest that IBD is a result of dysregulation of the gut mucosal immune system to commensal bacteria in genetically predisposed hosts (Xu et al., 2014). Evidence for a genetic basis is provided through the observation of the familial nature of the disease- for some time the increased risk of developing IBD has been recognised in first degree relatives (Orholm et al., 1991). Several genetic risk loci have been identified in IBD by genome-wide association studies (GWAs) (See section 1.4.2.1). Due to the pattern of incidence (most significant number of cases affecting higher socioeconomic classes), it has also been suggested that environmental factors such as hygiene and diet also contribute to the development of IBD, resulting in an

imbalance of the microbiota. The hygiene hypothesis proposes that excessive cleanliness results in decreased exposure to immune stimulating antigens in childhood, hence predisposing the child to inflammatory conditions such as IBD (Colombel et al., 2008). The role of environmental factors is supported by the fact that in genetically identical twins the concordance rates of IBD are relatively low (~50% for CD and 10% for UC) (Halfvarson et al., 2003). Moreover, the lower concordance rate for UC in these studies suggests a greater involvement of environment in the development of this disease than in CD.

#### 1.4.2.1 *Role of genetics in IBD susceptibility*

Several years ago, the majority of our knowledge about the involvement of genetics in IBD stemmed from studies using single nucleotide polymorphism and transgenic/deletion techniques in mice. Caspase recruitment domain (CARD) family member 15, also known as nucleotide-binding oligomerization domain-containing protein 2 (NOD2), was the first gene to be associated with IBD (Hugot et al., 2001; Ogura et al., 2001a). CARD15 comprises a protein containing two N-terminal CARDS, a nucleotide-binding domain, and 10 C-terminal leucine-rich repeats (LRRs), which is mainly expressed in monocytes (Ogura et al., 2001b). In wild-type NOD2, whilst both CARD regions act in concert to induce signalling *via* the nuclear factor NF- $\kappa$ B, the LRR region acts as an intracellular receptor for bacterial LPS and regulates the activation of NF- $\kappa$ B (Hugot et al., 2001; Ogura et al., 2001a; Ogura et al., 2001b). In CD, mutations in the CARD15 gene have been observed, with a particularly high frequency (93%) in the distal third of the gene, containing the LRR region. It is therefore proposed that mutations in the LRR region lead to a disruption in LPS recognition, and NF- $\kappa$ B signalling, although the exact mechanisms are unclear (Lesage et al., 2002). Mutant CARD15 with truncated LRRs is inefficient at preventing invasion of epithelial cells by *Salmonella* Typhimurium (Hisamatsu et al., 2003). Furthermore, in *Nod2*<sup>-/-</sup> mice given an intragastric dose of *Listeria monocytogenes*, a higher number of viable bacteria were recovered from the liver and spleen than in wild type mice, suggesting a role of *Nod2* in protecting against infection (Kobayashi et al., 2005).

Advances in the techniques used to study genetics have provided a much broader view of the genetic risk loci involved in the aetiology of IBD. GWAs are increasingly used to profile the genomes of many individuals and identify genetic variations associated with disease. In 2010 and 2011, two large GWA studies increased the total number of loci found to be implicated in IBD onset to 99, including 28 shared between UC and CD (Anderson et al., 2011; Franke et al., 2010; Lees et al., 2011). In a further large study of individuals with European

ancestry this number was dramatically increased to 163 loci (Jostins et al., 2012). Recently, the addition of an extra cohort of non-European individuals increased the number of loci by a further 38. Some of these loci were heterogeneous when compared to their European counterparts, indicating the importance of studying individuals with diverse ancestry (Liu et al., 2015).

A gene commonly associated with CD through GWAs is ATG16L1, a risk gene involved in the process of autophagy, the self-directed lysosomal degradation of damaged cellular components of intracellular bacteria, such as *S. Typhimurium* (Rioux et al., 2007). As well as removing damaged cell components, autophagy is an important process in providing cells with nutrients and energy and is upregulated during periods of, for example, starvation and high energy demands (Levine and Kroemer, 2008). The importance of ATG16L1 in mediating homeostasis with gut bacteria has been shown in knockdown models. Clearance of *S. Typhimurium* is impaired in a knockdown HeLa cell model of ATG16L1 (Rioux et al., 2007). Additionally in ATG16L1 deficient mice, there are disruptions in the Paneth cell granule exocytosis pathway, resulting in the abnormal release of antimicrobial peptides and lysozyme into the mucus layer (Cadwell et al., 2008). Therefore, it is apparent that functional ATG16L1 is important for effective bacterial clearance in the gut.

Further loci implicated in IBD include IL-23R, immunity-related GTPase family M protein (IRGM), organic cation transporter genes 1 and 2 (OCTN1/OCTN2), and multiple drug resistance gene 1 (MDR1) (Liu et al., 2015; Noble et al., 2005; Panwala et al., 1998; Rioux et al., 2007). However, together the identified loci only account for 13.1% and 8.2% of the variance in susceptibility for CD and UC, respectively, highlighting the need for further work to account for the remaining IBD risk (Liu et al., 2015).

GWAs have also implicated some mucin genes, as well as genes involved in mucin glycosylation in IBD. For example, associations between rare alleles of the *MUC3* gene and UC have been indicated (Kyo et al., 1999). Null variants of the fucosyltransferase, *Fut2*, have also been linked to IBD (McGovern et al., 2010). This, as well as alterations in expression, secretion, maturation, and glycosylation of mucins in IBD will be discussed further in section 1.4.3.

#### 1.4.2.2 Role of the GI microbiota in IBD

During the past decade there has been increasing focus on gut microbiota as an influential factor in IBD development in humans.

IBD is characterised by a reduction in the mucosal microbiota diversity when compared with healthy controls (Ott et al., 2004). In a subset of UC and CD patients, the mucosal microbial composition is abnormal, characterized by a depletion in Firmicutes and Bacteroidetes phyla (Frank et al., 2007). Similarly, multiple other studies reported a decrease in the microbial diversity in both mucosal and faecal samples from UC and CD patients in relapse/remission (Dicksved et al., 2008; Kang et al., 2010; Lepage et al., 2011; Manichanh et al., 2006; Scanlan et al., 2006). Despite this decrease in diversity, in patients with both UC and CD, the number of total mucosal associated bacteria increases with increasing levels of inflammation (Swidsinski et al., 2002). In addition to these changes, the microbiota has also been shown to be subject to temporal instability during periods of intestinal inflammation. In a study by Martinez and coll., healthy controls had a relatively stable microbiota composition over a period of 24 months, whereas patients with UC in remission demonstrated considerable variability in their faecal microbiota (Martinez et al., 2008). Likewise, the temporal stability of the microbiota in CD has also been observed to be decreased compared to controls (Scanlan et al., 2006). A longitudinal study of DSS administered mice showed that alterations in the microbiota occurred concurrently with markers of inflammation (Schwab et al., 2014). In mice administered with successive rounds of DSS, the microbiota became increasingly divergent over time, with shifts in community composition, and bacterial biomarkers of inflammation could be detected in the stool (Berry et al., 2015).

Aside from these broad level observations, changes in the abundance of more specific bacterial groups have been found in both UC and CD. Increases in the prevalence of potentially pathogenic bacteria have been observed including a consistent increase in the abundance of Enterobacteriaceae, such as strains of adherent-invasive *E. coli* in CD (Baumgart et al., 2007; Darfeuille-Michaud et al., 2004; Liu et al., 1995; Swidsinski et al., 2002). The mucosa associated population of *E. coli* is augmented in the ileum of CD patients, and rectum and sigmoid colon of both UC and CD patients (de Souza et al., 2012). There have also been several observations of a depletion of *F. prausnitzii* in IBD patients with active disease. This decrease is associated with the rate of recurrence in these patients (Joossens et al., 2011; Machiels et al., 2014; Sokol et al., 2008; Sokol et al., 2009). It is proposed that

*F. prausnitzii* contributes towards intestinal homeostasis through the immuno-modulatory effects exerted by its production of butyrate (Inan et al., 2000; Segain et al., 2000). Controversially, an increase of *F. prausnitzii* has been observed in 11 *de novo* paediatric cases of CD, confirmed by both quantitative reverse transcription polymerase chain reaction (qRT-PCR) and pyrosequencing methods, which provides an interesting insight into the treatment naïve IBD microbiota (Hansen et al., 2012b). Furthermore, in a separate study, a decrease in *F. prausnitzii* was only reported in ileal CD, whereas an increase in *F. prausnitzii* was found in colonic CD (Willing et al., 2010a). Recently, no differences were observed in the abundance of *F. prausnitzii* in CD patients with different disease activity (Schaffler et al., 2016). In a study examining CD, UC, and IBS patients it was shown that mucosa-associated *F. prausnitzii* and *E. coli* co-abundance can distinguish IBS and IBD phenotypes (Lopez-Siles et al., 2014).

A 2010 study revealed a link between the abundance of mucolytic bacteria and IBD. This study reported a >4-fold increase in *R. gnavus* and a ~100-fold increase in *R. torques* (Png et al., 2010). An increase in *R. torques* have also been demonstrated in other diseases such as IBS (Malinen et al., 2010). Comparison of ileal mucosa samples from healthy individuals and patients suffering from ileal CD revealed an increased abundance of *R. gnavus* in CD patients (Willing et al., 2010a). Similar results were observed using fecal samples from CD patients compared to unaffected controls (Joossens et al., 2011; Sokol et al., 2009). An earlier study reported an increase in anaerobic bacteria in colonic biopsies of CD patients when compared with normal controls; in the SI, whilst *R. gnavus* was increased in CD patients, there was a decrease in the *Clostridium leptum* and *Prevotella nigrescens* subgroups (Prindiville et al., 2004). Conversely, in patients with UC, *R. gnavus* was found to be depleted in abundance in the colonic mucosa of patients with active disease (Nishikawa et al., 2009). Mucin-degrading bacteria are capable of removing terminal sugars such as sialic acid and sulphate residues from glycans (and in particular, from mucus), making them more vulnerable to further degradation (Png et al., 2010). It has been hypothesized that increased mucin-degrading bacteria in IBD provide increased substrate to sustain non-mucolytic mucosa-associated bacteria, which could explain the increased total mucosa-associated bacteria in IBD. For example, the liberation of sialic acid from mucins has been shown to enhance the growth of *E. coli*, potentially driving an exaggerated immune response (Huang et al., 2015). *R. gnavus* strains (ATCC 29149 and 35913) were recently demonstrated to express an intramolecular *trans*-sialidase releasing sialic acid from  $\alpha$ 2-3 sialylated structures in the form of 2,7-anhydro-Neu5Ac. It is proposed that this activity provides such bacteria with a competitive nutritional

advantage over other species within the gut, specifically in the IBD mucosal environment which is rich in short sialylated mucin glycans (See section 1.4.3) (Tailford et al., 2015b).

However, not all mucin-degraders follow the same pattern. The abundance of *A. muciniphila*, the most prevalent mucolytic bacteria in controls was reduced in both UC and CD (Png et al., 2010). Whilst in control individuals, *A. muciniphila* correlated with total mucosal associated bacteria and *R. gnavus/R. torques* did not, the inverse was true for IBD patients (Png et al., 2010). It is also worth noting that a high proportion of *A. muciniphila* has been correlated with protection against inflammation in other diseases including type 1 diabetes mellitus (Hansen et al., 2012a), atopic dermatitis (Candela et al., 2012), autism (Wang et al., 2011), type 2 diabetes mellitus (Ellekilde et al., 2014), and obesity (Everard et al., 2013; Le Chatelier et al., 2013), suggesting that this strain possesses anti-inflammatory properties. In contrast to studies in humans, the abundance of *Akkermansia spp.* has been demonstrated to increase during acute-phase colitis in mice (Berry et al., 2015; Schwab et al., 2014).

Despite some general trends, there are conflicting results regarding alterations in the gut microbiota in IBD, including mucolytic bacteria, as also reported with other intestinal diseases. This variability could arise for a number of different reasons, including the choice of sample type, such as faecal or mucosal samples, where the microbial communities are known to differ. Furthermore, findings may also be impacted by sample handling, choice of DNA extraction and analysis method, reflecting the need for standardisation when studying the microbiota. A summary of studies examining these differences and changes in the mucosal community in IBD is given in table 2. Based on these studies, it is evident that future work is required to gain a clearer understanding of the role of mucin-degraders in metabolic syndromes such as IBD.

Study	Subject	Samples	Method	Major findings
Swidsinski A., 2007	IBD (n = 20)	Mucosal biopsy	Fluorescence <i>in situ</i> hybridisation (FISH)	Biofilm formation of <i>B. fragilis</i> in IBD patients
	C (n = 20)			Density of the biofilm is increased in IBD patients
Frank DN., 2007	CD, UC, C (total n = 190)	Resected tissues	Sequencing of nearly full-length 16S rRNA	↓ Bacteroidales and Lachnospiraceae
			qPCR	↑ Alpha, Beta and Gammaproteobacteria, Actinobacteria Reduced Bacterial load
Nishikawa J., 2009	UCa (n= 9)		Terminal Restriction	Reduced diversity in UC
	UCi (n =7)	Mucosal biopsies from non-inflamed tissue	Fragment Length Polymorphism (T-RFLP)	Mucosal microbiota of active and inactive UC differs to non-IBD
	C (n= 11)			Mucosal microbiota of active UC differs to inactive UC
Phg CW., 2010	CD (n = 26)			Increased total bacteria in non-inflamed biopsies in IBD
	UC (n = 20)	Mucosal biopsies of inflamed and non-inflamed tissue	RT-PCR	Increased mucolytic bacteria in IBD
	C (n = 20)			↓ <i>A. muciniphila</i> in IBD ↑ <i>R. gnnavus</i> and <i>R. torques</i> in IBD
Willing BP., 2010	Concordant/discordant twin pairs (n = 9)	Mucosal samples	454-Pyrosequencing	Similar microbial composition throughout the colon in IBD patients ↑ <i>R. gnnavus</i> in CD
Walker AW., 2011	CD (n = 6)			Reduced diversity (predominantly CD)
	UC (n = 6)	Paired mucosal biopsies of inflamed and non-inflamed tissue	Sequencing of nearly full-length 16S rRNA	Significant difference between inflamed and non-inflamed
	C (n = 5)		qPCR	↓ Firmicutes, ↑ Bacteroides in IBD ↑ Enterobacteriaceae in CD

Table 2 | Studies of mucosal associated microbiota in IBD

Study	Subject	Samples	Method	Major findings
Morgan XC., 2012	CD (n= 121)			↓ Roseburia, Phascolarctobacterium, ↑ Clostridium in IBD
	UC (n= 75)	Stool and mucosal biopsies	16s sequencing of V3-V5 region	↓ Ruminococcaceae, ↑ Enterobacteriaceae in CD ↓ Leuconostocaceae in UC
	C (n= 27)			Stool and biopsies significantly different
Tong M., 2013	CD (n= 16)			Reduced diversity in IBD
	UC (n= 16)	Mucosal lavage from non-inflamed regions	16S Illumina HiSeq sequencing of V4 region	↓ Firmicutes
	C (n=32)			↑ Actinobacteria, Proteobacteria and Tenericutes ↓ <i>F. prausnitzii</i>
Gevers D., 2014	CD (n= 447)			Reduced richness in CD
	C (n= 221)	Mucosal biopsies and serum samples, stool from some patients	16S Illumina HiSeq sequencing of V4 region	↑ Enterobacteriaceae, Pasteurellaceae, Veillonellaceae, Fusobacteriaceae ↓ Bacteroidales, Clostridiales, Erysipelotrichaceae
				Stool samples do not reflect dysbiosis as well as biopsies
Kabeerdoss J., 2014	CD (n= 28)			↑ Bacteroidetes, <i>E. coli</i> , <i>Lactobacillus</i> in UC
	UC (n=32)	Mucosal biopsies	qRT-PCR	↓ Firmicutes, <i>Clostridium leptum</i> in CD
	C (n= 30)			No differences between non-inflamed and inflamed mucosa
Lavelle A., 2015	UC (n=5)			↓ Bacteroidaceae and Akkermansaceae in UC
	C (n=4)	Luminal brush, mucosal biopsy and laser capture of mucus gel	16S sequencing of V4 region	↑ Clostridiaceae, Enterobacteriaceae, Ruminococcaceae, Bifidobacteriaceae in UC Lumen and mucus gel microbiota different
				Spatial conservation of microbiota despite local inflammation
Schaffner H., 2016	CD (n=26)	Mucosal biopsies of inflamed and non-inflamed tissue	16S Illumina MiSeq sequencing of V2 region	No difference in diversity depending on CD severity ↓ Bacteroidetes in severe CD

Table 2 continued | Studies of mucosal associated microbiota in IBD

### 1.4.2.3 *Impact of the immune system in IBD*

Gut homeostasis is maintained through tolerogenic innate and adaptive immune responses to commensal bacteria (Hooper and Macpherson, 2010). Tolerance is a lack of an antigen-specific pro-inflammatory response. In IBD, the tolerance to commensal bacteria is disrupted (Jump and Levine, 2004).

As a first line of innate defence, gut intestinal epithelial cells secrete mucins and antimicrobials, such as  $\alpha$ -defensins, into the gut lumen (Ayabe et al., 2000; Hansson, 2012). A variety of cell types contribute towards tolerance by sampling environmental antigens *via* pathogen recognition receptors (PRRs) such as TLRs, and NOD2/CARD15. These receptors recognise a range of microbial associated molecular patterns (MAMPs). Information obtained *via* this sampling is processed by the organised gut-associated lymphoid tissue (GALT) (Bakhtiar et al., 2013). Numerous mechanisms have been proposed for the maintenance of tolerance to the microbiota in the normal human gut. One suggestion is that epithelial cells and tissue resident immune cells, such as macrophages maintain tolerance to gut commensals through the downregulation of bacterial recognition receptors, such as CD14, TLR4 and MD-2, resulting in a lack of stimulation by bacterial LPS. Despite this, macrophages retain the ability to phagocytose pathogenic bacteria such as *S. typhimurium* (Abreu et al., 2001; Smythies et al., 2005). Furthermore, non-responsiveness to flagellin appears to be maintained by expression of TLR5 only on the basolateral and not the apical side of intestinal epithelial cells. As a result, pro-inflammatory responses will only be initiated by invasive bacteria (Gewirtz et al., 2001).

Mechanistic studies have confirmed the role of the innate immune system in maintaining gut homeostasis. Knockout of TLRs results in the development of, or increased susceptibility to colitis. For example, knockout of TLR5 in mice causes clinical, serologic, and histopathologic indicators of colitis in 35-40% of mice (Vijay-Kumar et al., 2007). An increased bacterial load was found in the colon of TLR5<sup>-/-</sup> mice, suggested to cause excessive activation of other immune pathways which drive the development of colitis (Vijay-Kumar et al., 2007). This was supported by the fact that a double knockout of TLR5 and TLR4 resulted in the protection of mice from the development of colitis (Vijay-Kumar et al., 2007). TLR4 and TLR2 knockout mice are more susceptible to DSS-induced colitis, resulting in increased bacterial translocation and disruption of epithelial tight junctions (Cario et al., 2007; Fukata et al., 2005). In IBD, polymorphisms in each of these TLRs, and in others such as TLR1, TLR6 and

TLR9 have been reported either to influence disease onset, or disease phenotype (Franchimont et al., 2004; Gewirtz et al., 2006; Torok et al., 2004).

A loss of innate immune tolerance in IBD is apparent by aberrant levels of various immune components, including cytokines and chemokines. Loss of the anti-inflammatory cytokine, IL-10 has been implicated in new-onset infantile IBD (Kotlarz et al., 2012). In contrast, DCs from CD patients produced higher levels of the pro-inflammatory IL-12p40 and IL6 cytokines (Ng et al., 2011). The differential production of cytokines elicits an abnormal adaptive T cell immune response in IBD. CD is dominated by a Th1 type immune response, including increases in the cytokines IL-12 and IFN- $\gamma$  (Kosiewicz et al., 2001; Parronchi et al., 1997; Sakuraba et al., 2009). Additionally, the Th17 immune response is also thought to play a role in the development of CD. IL-17 and IL-23 cytokines, mediators of this response, are both increased in CD (Liu et al., 2011; Yen et al., 2006). In contrast, UC is considered to have a Th2 type cytokine profile (Neurath, 2014). In an experimental oxazalone induced colitis model, it was shown that UC is mediated *via* the secretion of IL-13 by natural killer (NK-T) cells, since NK-T deficient mice fail to elicit a colitic response to oxazalone administration (Heller et al., 2002). However, these findings have not been replicated elsewhere leading to controversy about the role of IL-13 in UC (Kadivar et al., 2004). The levels of cytokines correlate with the abundance of some bacterial groups. For example, IL-6 negatively correlated with *F. prausnitzii* (Ng et al., 2011). A thorough review of cytokine production in IBD emphasised the need for a balance between pro-inflammatory and anti-inflammatory cytokines in maintaining homeostasis (Neurath, 2014).

A further arm of T cell adaptive immune system which is dysregulated in IBD is the action of a subpopulation of T cells which are CD4<sup>+</sup> CD25<sup>+</sup>, otherwise known as regulatory T cells (T<sub>reg</sub>). In healthy individuals, upon stimulation with antigen, T<sub>reg</sub> cells exert a suppressive effect upon the immune system, by inhibiting naive T cell proliferation and producing anti-inflammatory cytokines such as IL-10, avoiding excessive immune responses (Asseman et al., 1999; Huber et al., 2011; Maloy et al., 2003; Takahashi et al., 1998; Thornton and Shevach, 1998). A role for T<sub>regs</sub> in preventing IBD was also demonstrated many years ago in mice severe combined immunodeficient (SCID) mice. Whilst SCID mice reconstituted with a population of CD4<sup>+</sup> CD45RB<sub>hi</sub> (naïve) T cells developed spontaneous colitis, this effect was reversed in mice co-populated with CD4<sup>+</sup> CD45RB<sub>hi</sub> and CD4<sup>+</sup> CD45RB<sub>lo</sub> T cells. It was later demonstrated that this effect was mediated by a subpopulation of CD4<sup>+</sup> CD45RB<sub>lo</sub> cells expressing CD25 (Powrie et al., 1993; Read et al., 2000). LP mononuclear cells from patients with IBD appear to be resistant to inhibition by T<sub>reg</sub> cells, suggested to be mediated *via* upregulation of Smad7, a

protein that inhibits T<sub>reg</sub> signalling (Fantini et al., 2009; Monteleone et al., 2001). In germ-free mice, administration of SCFAs *via* drinking water has been shown to increase colonic T<sub>reg</sub> populations. Furthermore, SCFAs protected against experimentally induced colitis (Smith et al., 2013).

#### 1.4.2.4 *Involvement of environmental factors in IBD*

Environmental factors are also considered to be influential in the development of IBD. Associations between environmental factors and IBD have been reported for many years. For example, an inverse correlation between UC and cigarette smoking was identified in the 1980's and has been replicated on several occasions (Logan et al., 1984). Lower rates of UC have been noted amongst smokers, suggesting that smoking confers protection against this disease (Logan et al., 1984). Furthermore, smoking appears to modulate the severity of UC, with non-smokers requiring more periods of hospitalisation and operations than smokers (Boyko et al., 1988; Odes et al., 2001). In contrast, evidence suggests that smoking results in an increased risk of developing CD (Franceschi et al., 1987). As with UC, smoking appears to be capable of influencing the development of CD. For example, in the most severe smokers CD appears to mostly affect the small bowel (Lindberg et al., 1992; Russel et al., 1998). Recently, these trends have been confirmed in a prospective study of women (Higuchi et al., 2012). Despite these correlative observations there is very little understanding about the mechanisms causing this relationship. Studies have been performed investigating the role of cigarette smoke, and its components such as nicotine and carbon monoxide in IBD (Verschuere et al., 2012). Interestingly, the work of one group showed that whilst nicotine has negative effects in the small bowel, there are mixed effects in the colon, possibly mediated through differential induction of cytokines in these regions (Eliakim and Karmeli, 2003). However findings such as these must be taken with caution due to the substantial differences between this experimental set up and real life IBD cases.

A further environmental influence proposed to be protective from the onset of UC is appendectomy. In an early study, only 0.6% of UC patients but 24.5% of controls had undergone surgery to remove the appendix (Rutgeerts et al., 1994). This suggests that appendectomy reduces the chances of developing UC and has been demonstrated in several other studies, including one large study with a cohort of over 200,000 Swedish individuals (Andersson et al., 2001). As with smoking, it has been shown that appendectomy may also affect the disease course, with one study finding that previously appendectomised patients

had active disease for only 48% of the time, whilst non-appendicectomised patients had active disease for 62% of the time over the course of three years (Cosnes et al., 2002).

Diet is also proposed to play a role in the aetiology of IBD. A diet high in sugar, animal fat and linoleic acid increases the risk of IBD onset, whilst a diet rich in fibre and citrus fruit is considered to be protective (Owczarek et al., 2016). In a Japanese case-controlled study evaluating the role of dietary factors in the development of IBD, a positive association was noted between the consumption of sweets and onset of UC whilst a negative association was found with vitamin C (Sakamoto et al., 2005). Furthermore, a positive association was identified between CD and the consumption of sugars, sweeteners, fats (of varying types) and oils, fish and shellfish, vitamin E, and implications have been found for the consumption of coca cola and chocolate in IBD (Russel et al., 1998; Sakamoto et al., 2005).

Further environmental factors proposed to play a role in the onset of IBD potentially include infections, stress and depression, oral contraceptives and events in early childhood such as domestic hygiene or perinatal infections (Ananthakrishnan, 2013; Loftus, 2004).

#### 1.4.3 Role of mucus and mucins in IBD

Since mucus acts as the first point of contact for microbes with the host, it is unsurprising that aberrations in mucin expression, secretion or glycosylation have been associated with a number of diseases such as colitis, colonic cancer and IBD in humans (Forni et al., 2014; Larsson et al., 2011; McGovern et al., 2010; Parmar et al., 2012; Rausch et al., 2011) and mouse models (An et al., 2007; Fu et al., 2011; Stone et al., 2009).

Alterations in the gene expression of mucins have been identified in both the terminal ileum and colon of CD and UC patients (Moehle et al., 2006). Moehle and coll. described strong downregulation in the mRNA level of *MUC1*, *MUC2*, *MUC4*, *MUC5B*, *MUC12*, *MUC13*, *MUC17* and *MUC20* (Moehle et al. 2006). Patients with active UC showed significantly higher expression of *MUC16*, whilst those in remission had significantly increased levels of *MUC20* when compared to active UC and normal controls (Yamamoto-Furusho et al., 2015). In inflamed biopsies from the terminal ileum of CD patients, lower mRNA levels of *MUC2* were observed when compared with healthy controls (Hensel et al., 2014). In contrast, significantly higher levels of *MUC1* and *MUC2* expression was observed in non-inflamed IBD tissue when compared with controls (Hensel et al., 2014). In CD, the expression of *MUC2* was inversely correlated with the expression of the inflammatory marker *IL-8*, suggesting that *MUC2* is a biomarker of mucosal healing and regeneration (Hensel et al., 2014). *Muc2* deficiency in mice leads to the development of spontaneous colitis, associated with increased bacterial contact

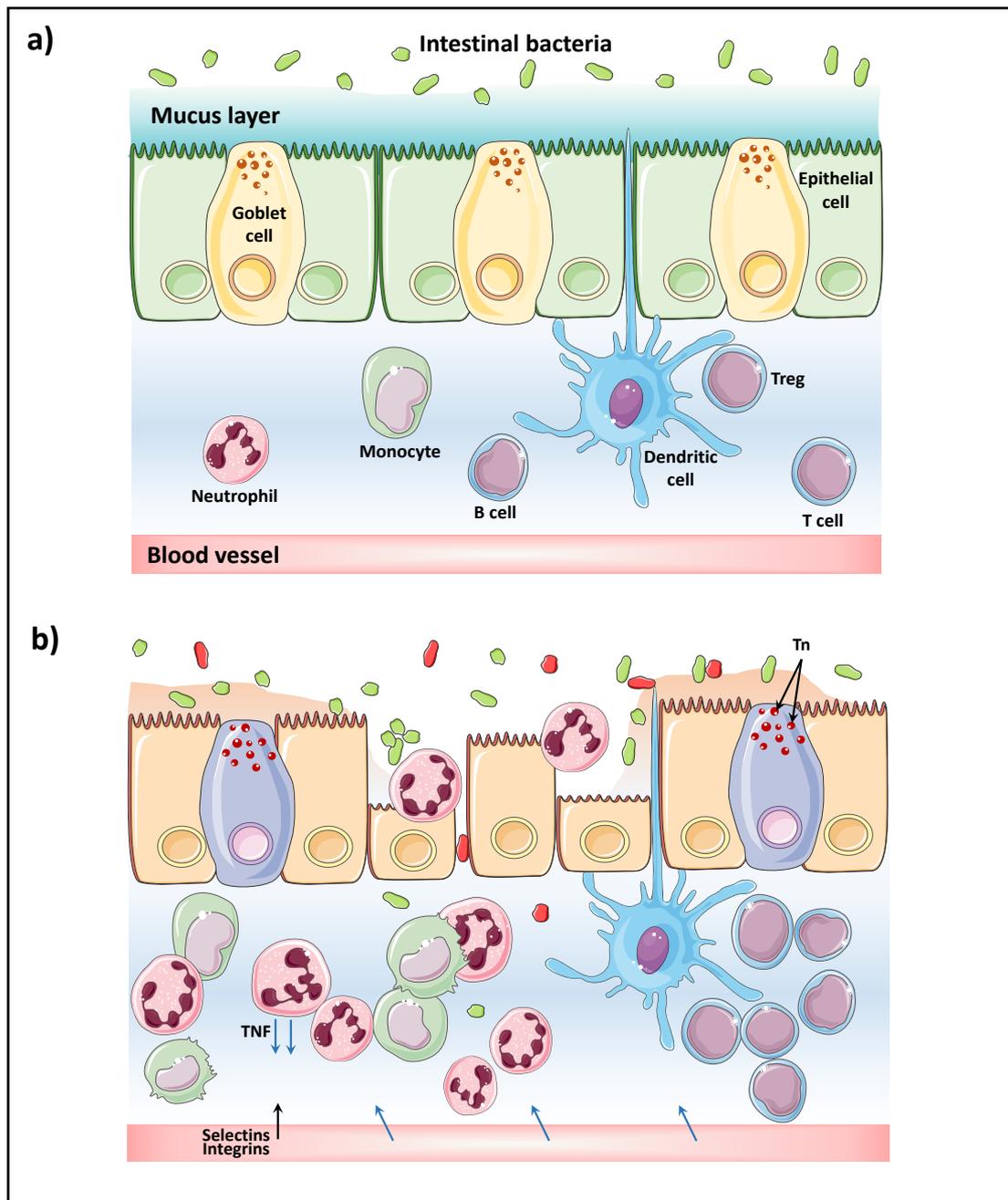
with the epithelium, increases in the expression of genes involved in immune activation and inflammatory responses, and enhanced production of immune cells, such as neutrophils, lymphocytes and DCs (Sovran et al., 2016; Wenzel et al., 2014). These changes correlated with alterations in gut microbiota diversity and composition (Sovran et al., 2016). These studies highlight the protective effect of mucus, which has been confirmed in the impaired resistance of *Muc2*<sup>-/-</sup> mice to *Trichuris muris* infection (Hasnain et al., 2010).

An alteration in mucin glycosylation is also associated with IBD. The protective function of mucin glycosylation has also been demonstrated across a number of studies. Knockout of core-1, core-2 and core-3 derived O-glycans in mice resulted in increased susceptibility and spontaneous development of colitis (Fig. 8), supporting a role of altered mucin glycosylation in driving the onset of inflammation, possibly *via* associated changes in the mucosal microbiota composition (An et al., 2007; Fu et al., 2011; Stone et al., 2009). Fu and coll. demonstrated that the colonic epithelium expressed Tn antigen in a subset of UC patients (Fig. 8) (Fu et al., 2011). In the core-1 and core-3 knockout mice increased bacterial translocation to the epithelium was detected (Fig. 8) (An et al., 2007; Fu et al., 2011). Furthermore, knockout of *C1GalT1* in mice resulted in the spontaneous development of colitis by 2 weeks of age, and *C3GnT*<sup>-/-</sup> mice were more susceptible to DSS induced colitis (An et al., 2007; Fu et al., 2011). Loss of *C1GalT1* led to subtle differences in the microbiota, including an increase in Bacteroidetes but a decrease in Firmicutes (Sommer et al., 2014). A recent study observed that *C1GalT1* deficiency primarily resulted in the onset of colitis in the distal colon whilst mice lacking both *C1GalT1* and *C3GnT* developed colitis both in the distal and proximal regions, suggesting that C3GnT plays a protective role in the proximal colon (Bergstrom et al., 2016). The mucins of these double knockout mice were more susceptible to proteolysis than those of wild type mice. Furthermore, antibiotic treatment reduced the faecal bacterial load, restored the mucus layer and reduced the severity of colitis (Bergstrom et al., 2016).

Alteration in mucin glycosylation have also been observed in IBD patients. A depletion in overall carbohydrate content in patients with active UC was first observed over 30 years ago (Campbell et al., 1995; Clamp et al., 1981). In 2011, Larsson and coll. reported a decrease in complex glycans and an increase in smaller glycans, such as the STn antigen in active UC, suggested to be due to increased activity of the ST6GalNAc-I sialyltransferase. These changes correlated with the severity of disease (Larsson et al., 2011). Increases in shorter glycans such as these may explain why bacteria such as *R. gnavus* are found in increased abundance in IBD. In contrast, *A. muciniphila*, which can grow on complex mucins with the synergistic

action of several enzymes, cannot utilise sialic acid as a sole carbon source, possibly explaining its decreased abundance in IBD (Tailford et al., 2015b). An overexpression of Thomsen-Friedenreich (TF) antigen has also been demonstrated in UC (Campbell et al., 1995). In normal histological mucosa, TF antigen is present but concealed by further glycosylation such as sialylation and fucosylation, again supporting the finding of a decrease in glycan complexity in UC (Campbell et al., 1995). Individuals with null variants of the *Fut2* allele, which determines secretor status i.e. the presence (secretor) or absence (non-secretor) of H and ABO antigens in GI mucosa, were found to be more susceptible to the onset of CD (McGovern et al., 2010). Furthermore in a cohort of healthy and CD patients, a number of bacterial groups were associated with the *Fut2* genotype (Rausch et al., 2011).

Despite the apparent link between mucins, mucin glycosylation and IBD, the exact molecular mechanisms that define the relationship between microbes and the mucus layer are poorly understood. Whilst the above studies in transgenic mice support a causal role for mucin glycosylation in IBD, there are other lines of evidence to suggest that aberrant mucin glycosylation may occur secondary to inflammation (Arike and Hansson, 2016). For example, the O-glycosylation of MUC2 from a UC patient analysed on two separate occasions was only altered when the disease was active, whilst in remission, glycosylation had returned to a normal pattern (Larsson et al., 2011). This suggests that aberrant glycosylation is associated with inflammation and not just with disease. It is evident that the role of mucin glycosylation and its interaction with the microbiota is unclear based on current findings. More work is therefore necessary to support the causal link between altered O-linked glycosylation and inflammation, as recently reviewed (Theodoratou et al., 2014).



**Figure 8 | Schematic diagram illustrating the effect of impaired mucin O-glycosylation in the GI tract**

a) Normal O-glycosylation; the mucus layer and intestinal epithelium are intact b) Aberrant O-glycosylation e.g. as a result of loss of core 1 or core 3 glycans; intestinal bacteria interact abnormally with the epithelium, resulting in inappropriate activation of immune cells.

Figure adapted from Fu et al., 2011 (Servier Medical Art, 2016)

## 1.5 Aim and Objectives

The intestinal mucus layer forms a protective barrier that provides a nutrient source and habitat for commensal organisms, and plays a key role in the maintenance of gut homeostasis. Dysbiosis of the microbiota, and alterations in mucin glycosylation have been associated with IBD. However further work is required to provide clarification on the causal relationship between these two factors. The overall aim of this PhD project is to use a multidisciplinary approach to help understand the relationship between the mucus-associated microbiota and mucin glycosylation, and the underpinning molecular mechanisms mediating these changes in IBD.

The specific objectives of this project are as follows:

1. To identify changes in mucin glycosylation in IBD patients.
2. To determine changes in mucosa-associated bacterial composition in IBD patients.
3. To investigate molecular mechanisms leading to IBD using *in vitro* growth assays and *in vivo* using mouse models.

## Chapter 2) Materials and methods

### 2.1 Recruitment of patients and collection of samples

Patients undergoing colonoscopy or sigmoidoscopy procedures to investigate a suspected flare of IBD, for a routine surveillance of IBD or for investigation of another possible bowel condition (but found to be macroscopically and microscopically normal) were recruited in the Gastroenterology department at the Norfolk and Norwich University Hospital (NNUH). Ethical approval was obtained through the Faculty of Medicine and Health Sciences Research Ethics Committee at the UEA, and NHS research and development permission was granted from the NNUH. On the day of the procedure, patients were fully briefed and asked for written informed consent *via* the Norwich Biorepository information sheet and consent form. All patients received a bowel preparation of either sodium picosulfate (Picolax) or Macrogol (Kleanprep) for colonoscopy procedures or enema for sigmoidoscopy procedures. To collect mucus, lavage samples were collected as previously described, with minor amendments (Li et al., 2011). Briefly, 50 ml sterile water was injected onto the surface of the colon. Vacuum suction was then used to collect the aspirated mucus into a mucus trap (Bard medicals, Crawley, UK). Two lavage samples (one from sigmoid colon and one from ascending colon) were collected during colonoscopy procedures, whilst only sigmoidal samples were collected during sigmoidoscopy procedures. The mucus trap was sealed and kept on ice for lavage processing.

Biopsies were also collected from a nearby location using 5 mm pinch biopsy forceps, and transferred immediately to dry ice. For colonoscopy procedures, four biopsy samples (two from sigmoid colon and two from ascending colon) were collected. As above, only sigmoidal samples were collected during sigmoidoscopy procedures.

Inflammation at each region was categorised as mild/moderate/severe according to assessment of separate clinical samples independently graded by a histologist (NNUH). Anonymised information about patient sex, age, and medication was recorded (See appendix 2).

## 2.2 Initial processing and storage of human samples

Lavage samples were centrifuged at 1000 x g for 10 min to separate bacteria (pellet) and mucus (supernatant). Bacterial pellets were re-suspended in 5 ml 1x phosphate buffered saline (PBS), and stored at -80°C until further use. The supernatants were then centrifuged at 4000 x g for 30 min to remove solid components. The aqueous phase was transferred in to a fresh tube and four volumes (vol) of acetone added to precipitate proteins for a minimum of 1 night at -80°C. Samples were then thawed and the precipitated protein pelleted by centrifugation at 4000 x g for 30 min. The pellet was re-suspended and stored in 1-2 ml of extraction guanidine hydrochloride (extr-GuHCl) buffer (4 M with 3.3 mM EDTA, 0.006 M NaH<sub>2</sub>PO<sub>4</sub>) at -20°C until mucin extraction.

Human mucosal biopsies were transferred to -80°C until further use.

## 2.3 Methods for microbiota analysis from human mucosal lavage samples

### 2.3.1 Extraction of bacterial DNA from lavage

Bacterial samples prepared in PBS (See section 2.2) were thawed and transferred to a tared tube. Following centrifugation at 14,000 revolutions per minute (rpm) for 10 min at 4°C, the supernatant was discarded and the mass of the bacterial pellet quantified by weighing the tube. Pellets were re-suspended in 1x PBS to a concentration of 0.3125 mg/μl. An aliquot (80 μl), equating to 25 mg material, was taken for DNA extraction using the QiaAmp DNA mini kit (Qiagen, Hilden, Germany) and the remaining suspension frozen at -20°C.

Aliquots were first homogenised by adding 100 μl ATL buffer (QiaAmp DNA mini kit, Qiagen, Hilden, Germany), and bead beating for 2 x 45 seconds (s) at 5.0 M/s in a FastPrep-24 Instrument (MP Biomedicals, Santa Ana, CA) using autoclaved 0.1 mm silica/zirconium beads (B. Braun Melsungen AG, Germany). Bacterial DNA extraction was then completed according to the manufacturer's instructions, following the 'DNA Purification from Tissues' protocol from step 3 in the QiaAmp DNA mini kit handbook (Qiagen, Hilden, Germany). At the elution step, two elutions were performed with 200 μl buffer AE from the QiaAmp DNA mini kit.

Pellet Paint Co-Precipitant (Merck Millipore, Darmstadt, Germany) was used to precipitate and concentrate extracted bacterial DNA. Briefly, a 2 μl aliquot of pellet paint equilibrated to room temperature was added to the DNA followed by the addition of a 0.1 vol of 3 M sodium acetate and 1 vol of isopropanol. Samples were briefly vortexed and then incubated

at room temperature for 2 min to allow the DNA to precipitate. Co-precipitated pellet paint and DNA were centrifuged at 14,000-16,000 x g for 5 min. The supernatant was carefully removed with a pipette and the pellet washed twice with 1 ml of 70% ethanol, vortexed and centrifuged as above. A final wash was carried out with 1 ml of 100% ethanol, centrifugation as above, the supernatant removed and the pellet air dried before re-suspending in 20 µl buffer AE (QiaAmp DNA mini kit, Qiagen, Hilden, Germany).

### 2.3.2 qPCR assays of bacterial DNA from lavage

Primers targeting bacterial groups of interest, including mucin-degraders were identified from the literature (Table 3). The validity of primers was confirmed by the production of standard curves using amplicons from representative organisms of each bacterial group.

Genomic DNA (gDNA) was first extracted from the strains listed in table 3 using a GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific, Waltham, MA), according to the manufacturer's instructions. Briefly, for gram-negative bacteria, pellets from bacterial cultures were re-suspended in 180 µl of digestion solution, 20 µl of proteinase K solution and samples incubated at 56°C for 30 min. A 20 µl aliquot of RNase A was added, samples mixed and incubated for 10 min at room temperature. Then, 200 µl of lysis solution was added and samples were mixed by vortexing for 15 s. For gram-positive bacteria, pellets were re-suspended in 180 µl of gram-positive lysis buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 1.2% Triton X-100, 20 mg/ml lysozyme), and incubated for 30 min at 37°C. A 200 µl aliquot of lysis solution and 20 µl aliquot of proteinase K was added and mixed thoroughly. Samples were incubated at 56°C for 30 min, following which a 20 µl aliquot of RNase A solution was added.

To the bacterial lysates, a 400 µl aliquot of 50% ethanol was added, samples mixed by vortexing, and the lysates transferred to a GeneJET gDNA purification column. Columns were centrifuged 1 min at 6000 xg, the flow-through discarded, and 500 µl of wash buffer I, supplied with the GeneJET kit, added. Columns were then centrifuged for 1 min at 8000 xg, the flow-through discarded, and 500 µl of wash buffer II added. Columns were centrifuged for 3 min at maximum speed, the collection tubes containing the flow-through solution discarded and the columns placed into a sterile 1.5 ml eppendorf tube. To elute gDNA, two successive elutions with 200 µl of elution buffer were performed, which were then pooled.

The 16S region of purified DNA was then amplified using universal primers 27F (5'-AGAGTTTGATCMTGGCTCAG- 3') and RP2 (5'-ACGGCTACCTGTTACGACTT-3') with the reaction mixture and amplification conditions as follows:

**16S PCR reaction mix (per reaction)**

Reagent	Vol	<b><u>Reaction conditions</u></b>		
		Temp	Duration	Cycles
Hotstar 10 x buffer (Qiagen, Hilden, Germany)	5 µl	95°C	15 min	
dNTPs (10 mM)	1 µl	94°C	45 sec	} x 30
forward primer (27F- 10 µM)	2.5 µl	52°C	45 sec	
reverse primer (RP2- 10 µM)	2.5 µl	72°C	2 min	
Hotstar Taq (Qiagen, Hilden, Germany)	0.3 µl	72°C	10 min	
template DNA (150 ng/µl)	1 µl			
dH <sub>2</sub> O	to 50 µl			

Amplicons were purified using a QIAquick PCR purification kit (Qiagen, Hilden, Germany) and eluted with sterile dH<sub>2</sub>O. The PCR products were diluted to 10<sup>10</sup> copies/µl (~16.3 ng/µl). A template dilution series was used to produce standard curves by qPCR, for validation of group specific primers and use in subsequent analysis of human mucosal associated bacteria.

The qPCR reaction mixture and amplification conditions were as follows:

**qPCR reaction mix (per reaction)**

Reagent	Vol	<b><u>Reaction conditions</u></b>		
		Temp	Duration	Cycles
SYBR master mix (Qiagen, Hilden, Germany)	5 µl	95°C	5 min	
Forward primer (10 µM)	0.5 µl	95°C	10 sec	} x 40
Reverse primer (10 µM)	0.5 µl	60°C	35 sec	
dH <sub>2</sub> O	2 µl	95°C	15 sec	
DNA template	2 µl	60°C	1 min	
		95°C	15 sec	
		60°C	15 sec	

To quantify bacterial groups of interest in samples, the concentration of bacterial DNA prepared in section 2.3.1. was measured using a broad range fluorimetric double stranded DNA assay on a Qubit® 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA), according to the manufacturer's instructions. The DNA was diluted to a concentration of 5 ng/μl using sterile dH<sub>2</sub>O, and used in a reaction mixture for relative quantitative analysis by qPCR, as described above. Each sample was run in triplicate. Following qPCR, data were analysed by measuring the Ct values for individual bacterial groups at a threshold of 0.05 Rn. The relative abundance for each targeted group was expressed as a % of total 16S using the following calculation:

$$\text{Relative \% of bacterial group of interest} = (E_o^{-\text{Ct}(o)} / E_u^{-\text{Ct}(u)}) \times 100$$

E<sub>o</sub>- efficiency of primers for target groups (Table 3)

E<sub>u</sub>- efficiency of universal 16S qPCR primers (Table 3)

Primer names	Sequence	Target	Product length (bp)	Reference
UniF	GTGSTGCAYGGYGTGCTCA	Universal	147-148	(Fuller et al., 2007)
UniR	ACGTCRTCCMCNCCTTCCTC			
RrecF	GCGGTRCGGCAAGTCTGA	Roseburia spp. genus & <i>Eubacterium rectale</i> species	81	(Ramirez-Farias et al., 2009)
Rrec630mR	CCTCCGACACTCTAGTMCGAC			
FPR-2F	GGAGGAAGAAGGTCTTCGG	Faecalibacterium spp. genus	248	(Ramirez-Farias et al., 2009)
Fprau645R	AATTCGGCCTACCTCTGCAC			
Bac303F	GAAGGTCCCCCACATTG	Bacteroides spp. genus & Prevotella spp. genus	103	(Ramirez-Farias et al., 2009)
Bfr-Fmrev	CGCKACTTGGCTGGTTCAG			
Biff	TGCGTCYGGTGTGAAAG	Bifidobacterium spp. genus	128	(Belenguer et al., 2006)
g-Bifid-R	GGTGTCTTCCCGATATCTACA			
Lac-1	AGCAGTAGGGAATCTCCA	Lactobacillaceae family & Weissella spp. genus	341	(Rinttila et al., 2004)
Lab0677	CACCGCTACACATGGAG			
16S-BF-F	TCAGGAAGAAAGCTTGCT	<i>Bacteroides fragilis</i> species	161	(Tong et al., 2011)
16S-BF-R	CATCCTTACCGGAATCCT			
16S-AM-F	CAGCAGTGAAGGTGGGAC	<i>Akkermansia muciniphila</i> species	327	(Collado et al., 2007)
16S-AM-R	CCTTGGGTTGGCTTCAGAT			
16SRg5F	TGGCGGCGTGCTTAACA	<i>Ruminococcus gnavus</i> species	57	This study
16SRg5R	TCCGAAGAAATCCGTCAAGGT			
LachnoF	AGTAACGCCCCGAAGTCAGTG	Lachnospiraceae family	76	This study
LachnoR	CGACTTCACCCCGATTATCG			

Table 3 | Family, genus and species specific primers used for 16S targeted quantification

### 2.3.3 16S sequencing of bacterial DNA from human mucosal lavage

#### 2.3.3.1 Preparation of 16S DNA libraries

The concentration of samples prepared in section 2.3.1. was measured using a broad range fluorimetric double stranded DNA assay on a Qubit® 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA), according to the manufacturer's instructions. DNA was diluted to a concentration of 5 ng/μl using sterile dH<sub>2</sub>O and used in the PCR reaction described below. For 16S sequencing, the V1-V2 region of 16S DNA was amplified using V1-V2 specific primers. Forward primers consisted of a 5' illumina adaptor sequence, a 10 nucleotide 'pad' sequence designed to boost the sequencing primer melting temperatures, 2-nucleotide 'link' sequence that is anti-complementary to the known sequences, and a V1 16S gene specific primer sequence (27F). Reverse primers consisted of the same adaptor, pad and link sequences plus a V2 16S gene specific primer sequence (338R). Additionally, each forward and reverse primer had a unique 8 nucleotide index region, allowing samples to be barcoded (Table 4).

PCR was carried out using the Q5 high-fidelity polymerase kit (New England Biolabs, Ipswich, MA) with the following reaction mixture and conditions:

<b><u>16S V1-V2 reaction mix (per reaction)</u></b>		<b><u>Reaction Conditions</u></b>		
<b>Reagent</b>	<b>Vol</b>	<b>Temp</b>	<b>Duration</b>	<b>Cycles</b>
Nuclease free water (Merck Millipore)	14.25 μl	98°C	2 min	x20
5x Q5 PCR Buffer (New England Biolabs)	5 μl	98°C	30 sec	
10 mM dNTPs (Thermo Fisher Scientific)	0.5 μl	50°C	30 sec	
Forward primer (10 μM)	1.25 μl	72°C	90 sec	
Reverse primer (10 μM)	1.25 μl	72°C	5 min	
Q5 Taq PCR enzyme (New England Biolabs)	0.25 μl			
DNA template	2.5 μl			

PCR was carried out in a 96-well plate, in triplicate per sample.

The size of each PCR product was estimated by electrophoresis on a 1% agarose gel. Only samples yielding the expected V1-V2 product size of approximately 435 bp were selected for sequencing.

Following amplification and electrophoresis, triplicate PCR products were pooled (75 μl total volume) into a 96-well PCR plate and purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA). Briefly, an equal volume of AMPure XP beads was added to each sample

(75  $\mu$ l), gently mixed by pipetting up and down 10 times, and incubated at room temperature for 5 min. The PCR plate was placed on a magnetic stand for 2 min until the supernatant had cleared. Whilst still on the magnetic plate the supernatant from each well was removed and discarded. A 200  $\mu$ l aliquot of freshly prepared 80% ethanol was then added for 30 s. The washing step was repeated before removing the ethanol and allowing the beads to air-dry for 15 min. Then, the PCR plate was removed from the magnetic stand and 52.5  $\mu$ l of 10 mM Tris pH 8.5 added to each well, gently mixed by pipetting up and down 10 times, and incubated at room temperature for 2 min to elute DNA from the beads. The beads were collected by placing the amplicon PCR plate onto the magnetic stand for 2 min, and 50  $\mu$ l of the supernatant transferred into a new 96-well PCR plate.

Purified samples were quantified using the broad range or high sensitivity fluorimetric double stranded DNA assay on a Qubit<sup>®</sup> 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA), as above. Based on the measured concentrations, a DNA library was constructed using an equimolar mix of samples. The library was purified to remove primer dimers and contaminants by electrophoresis on a 1% agarose gel. Prior to sequencing, the library DNA band was excised and the DNA purified using the Wizard SV gel and PCR clean-up kit (Promega, Madison, WI), according to the manufacturer's instructions. The library DNA was sequenced on an Illumina MiSeq platform at the Sanger Institute (Cambridge, UK).

<b>Forward primers</b>	<b>Forward primer sequences</b>
Primer design	5'illumina adaptor – 8nt index sequence – 10nt pad sequence – 2nt linker sequence – V1 gene specific region
V1FW_SD501	aatgatacggcgaccaccaggatctacac – aagcagca – tatgtaatt – gt – agmgttygatymtggctcag
V1FW_SD502	aatgatacggcgaccaccaggatctacac – acgcgtga – tatgtaatt – gt – agmgttygatymtggctcag
V1FW_SD503	aatgatacggcgaccaccaggatctacac – cgatctac – tatgtaatt – gt – agmgttygatymtggctcag
V1FW_SD504	aatgatacggcgaccaccaggatctacac – tgcgtcac – tatgtaatt – gt – agmgttygatymtggctcag
V1FW_SD505	aatgatacggcgaccaccaggatctacac – gtctagt – tatgtaatt – gt – agmgttygatymtggctcag
V1FW_SD506	aatgatacggcgaccaccaggatctacac – ctagtatg – tatgtaatt – gt – agmgttygatymtggctcag
<b>Reverse primers</b>	<b>Reverse primer sequences</b>
Primer design	5'illumina adaptor – 8nt index sequence – 10nt pad sequence – 2nt linker sequence – V2 gene specific region
V2RV_SD701	caagcagaagaaggcaccacagat – acctagta – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD702	caagcagaagaaggcaccacagat – agctacgt – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD703	caagcagaagaaggcaccacagat – atatcgcg – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD704	caagcagaagaaggcaccacagat – caccagatg – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD705	caagcagaagaaggcaccacagat – cgtatcgc – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD706	caagcagaagaaggcaccacagat – ctgcgact – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD707	caagcagaagaaggcaccacagat – gctgtaac – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD708	caagcagaagaaggcaccacagat – ggacgtta – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD709	caagcagaagaaggcaccacagat – ggtcgtag – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD710	caagcagaagaaggcaccacagat – taagtctc – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD711	caagcagaagaaggcaccacagat – tacacagt – agtcagtcag – cc – gctgcctcccgtaggag
V2RV_SD712	caagcagaagaaggcaccacagat – ttgacgca – agtcagtcag – cc – gctgcctcccgtaggag

**Table 4 | Primer design and sequences used to amplify the V1-V2 region of bacterial 16S DNA for 16S sequencing**

### 2.3.3.2 *Bioinformatics of 16S sequencing data*

Processing of 16S sequencing data was carried out by an in-house bio-informatician.

Raw sequence reads were processed through quality control using the FASTX-Toolkit ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)), keeping a minimum quality threshold of 33 for at least 50% of the bases. Reads that passed the threshold were aligned against the SILVA database (version SILVA\_119\_SSURef\_tax\_silva, <http://www.arb-silva.de/>) using BLASTN (ncbi-blast-2.2.25+; Max e-value  $10^{-3}$ ) (Altschul et al., 1990). After performing BLASTX alignment, all output files of paired read sequences were imported and analysed using the paired-end protocol of MEGAN (Huson et al., 2011).

For processing the BLAST files by MEGAN5, parameter settings of “Min Score = 50”, “Top Percent = 10” were used. Reads which had no matches to the respective database were placed under a “No hit” node, and reads that were originally assigned to a taxon but did not meet the selected threshold criterion were pushed back using the lowest common ancestor (LCA) algorithm to higher nodes where the threshold was met. After importing datasets in MEGAN, MEGAN-own “rma files” were obtained for each data mapped onto NCBI taxonomy based on the selected threshold. Furthermore, all the files were compared and analysed within MEGAN.

## 2.4 Methods for mucin glycosylation analysis from human mucosal lavage samples and mucosal biopsies

### 2.4.1 Purification of mucins from mucosal lavage samples

Following initial processing of samples (See section 2.2), a 200  $\mu$ l aliquot of protein re-suspended in extr-GuHCl was centrifuged at 13,200 rpm for 30 min, and the GuHCl soluble (supernatant) and insoluble (pellet) fractions were separated. The soluble fraction was stored at -20°C. To the insoluble fraction, a further vol of extr-GuHCl (100  $\mu$ l, 6 M with 5 mM EDTA, 0.01 M NaH<sub>2</sub>PO<sub>4</sub>) was added, and the solution stirred overnight (O/N) at 4°C. The sample was then centrifuged as above and the soluble fractions pooled. To the insoluble fraction, reduction guanidine hydrochloride (red-GuHCl) was added (120  $\mu$ l, 6 M with 5 mM EDTA, 0.1 M Tris). Dithiotreitol (DTT) was added to both insoluble and soluble fractions to a final concentration of 50 mM and the samples stirred for 1.5 h at 37°C. Alkylating agent 4-vinylpyridine was added to a final concentration of 125 mM and the samples stirred for 1 h at room temperature in the dark. The samples were then transferred to 10 kDa molecular weight cut-off (MWCO) dialysis cups (Slide-A-Lyzer MINI dialysis cups, Thermo Fisher Scientific, Waltham, MA) and dialysed against 4 L water O/N. Dialysed samples were dried in a SpeedVac-concentrator (Eppendorf, Hamburg, Germany) and re-suspended in fresh water to the desired concentration.

### 2.4.2 Composite AgPAGE of purified mucins

#### 2.4.2.1 *Composite AgPAGE gel casting and loading*

All composite agarose- polyacrylamide gel electrophoresis (AgPAGE) gels were cast at 60°C in an oven.

Following the method of Schulz and coll. (Schulz et al., 2002), with minor amendments, an agarose (Ag- 0.06 g agarose, 2.4 ml 5x Tris-hydrochloric acid (HCl) pH 8.1, 9.6 ml dH<sub>2</sub>O) and polyacrylamide (PA- 0.12 g agarose, 2.4 ml 5x Tris-HCl pH 8.1, 2.4 ml 50% glycerol and 5.4 ml dH<sub>2</sub>O) gel solution was prepared. Both solutions were melted in a microwave and placed in a water bath at 60°C to equilibrate. At this point 1.8 ml of 40% polyacrylamide was added to the PA mixture. A 5 ml aliquot of the Ag solution was then pipetted into the right hand chamber of the gradient mixer (Discontinued, MSE, London, UK), and 2  $\mu$ l 40% ammonium persulfate (APS), followed by 2  $\mu$ l tetramethylethylenediamine (TEMED) added. The same volume of PA (5 ml) was then pipetted into the left hand chamber and APS and TEMED added as above. The peristaltic pump (GE Healthcare, Little Chalfont, UK) was switched on at a flow

rate of  $150 \text{ ml}\cdot\text{h}^{-1}$  and the valve between the gradient mixer chambers opened. Once cast to the top of the mould, a 10-well comb was inserted and the pump flushed with  $\text{dH}_2\text{O}$ . The gel was allowed to set at room temperature for approximately 1 h, before being transferred to a humid environment at  $4^\circ\text{C}$ .

Following concentration using a SpeedVac-concentrator, mucus samples prepared in section 2.4.1 were re-suspended in a 1:2:1 ratio of  $\text{dH}_2\text{O}$ , 2X loading buffer (0.75 M Tris-HCl pH 8.1, 60% (vol/vol) glycerol, 0.01% (wt/vol) bromophenyl blue, 2% (wt/vol) sodium dodecyl sulfate, SDS) and 1 M DTT. The samples were incubated at  $37^\circ\text{C}$  for 3 h, and stored at  $-20^\circ\text{C}$  until further use. Before loading, the samples were boiled at  $95^\circ\text{C}$  for 5 min. An appropriate volume of each sample (15-30  $\mu\text{l}$ ) was loaded alongside a high molecular weight protein marker (HiMark™ pre-stained protein standard, Thermo Fisher Scientific, Waltham, MA). The gel was run with composite gel running buffer (192 mM boric acid, 1 mM ethylenediaminetetraacetic acid (EDTA), 0.1% (wt/vol) SDS, pH 7.6) at 300 V, 12 mA and 50 W O/N at  $4^\circ\text{C}$ .

#### 2.4.2.2 Composite gel staining

To identify mucin proteins, the composite gels were washed briefly in  $\text{dH}_2\text{O}$  followed by staining in GelCode blue stain reagent (Thermo Fisher Scientific, Waltham, MA) for 30 min to 1 h. The gels were then washed 2 x with  $\text{dH}_2\text{O}$  and imaged.

To stain for mucin glycans, GelCode stained composite gels were first de-stained in 50% methanol for 30 min to O/N, rehydrated with 3% acetic acid (2 x 10 min) and then stained using the Glycoprotein staining kit (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's instructions (from step 3).

#### 2.4.2.3 Transfer of mucins to PVDF membrane

Composite AgPAGE transfer buffer was prepared prior to transfer (250 mM Tris-HCl, 1.92 M glycine, 0.4% (wt/vol) SDS, 20% (vol/vol) methanol). Unstained gels were washed thoroughly with  $\text{dH}_2\text{O}$  before equilibration in transfer buffer for 10 min. Blotting was carried out on 0.45  $\mu\text{M}$  Hybond-P PVDF membrane (GE Healthcare, Little Chalfont, UK) using an XCell II blot module (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's instructions. Gels were blotted at  $4^\circ\text{C}$  for 4 h at 150 V, 300 mA and 40 W.

#### 2.4.2.4 Alcian blue staining of PVDF membranes

Blots were rinsed with  $\text{dH}_2\text{O}$  and transferred into an Alcian Blue solution (25% (vol/vol) ethanol, 10% (vol/vol) acetic acid, 0.125% (wt/vol) Alcian Blue 8GX) for 30 min to O/N. The

blots were destained with 100% methanol, rinsed with dH<sub>2</sub>O, air-dried and imaged using a FluorChem E system (Protein Simple, San Jose, CA).

### 2.4.3 Dot blot analysis of purified human mucins

Purified mucins (See section 2.4.1) were normalised to a concentration of 2.5 mg/ml protein, as measured using the protein A280 function on a Nanodrop instrument (Thermo Fisher Scientific, Waltham, MA). For dot blotting, 1 µl of each soluble and insoluble fraction was spotted onto a Hybond-C nitrocellulose membrane (GE Healthcare, Little Chalfont, UK). The membranes were dried in air and then incubated in a protein-free blocking buffer (Thermo Fisher Scientific, Waltham, MA) for 1 h to O/N.

Blots were visualised with fluorescein isothiocyanate (FITC)- or biotinylated-lectins (See section 2.4.3.1), or with alcian blue staining as above.

#### 2.4.3.1 *Lectin probing of dot blots*

Following O/N blocking, blots were rinsed 3 x with PBS and then incubated with lectins for probing.

For FITC labelled lectins (PNA, RCA, UEA, WGA, SNA, Table 5), a 15 µg/ml solution was added to the blots, which were then incubated for 1 h at room temperature. The blots were rinsed 3 x with PBS and imaged using a Pharos FX Plus imager (Bio Rad, Hercules, CA).

For biotinylated lectins (MALII, Table 5), a 1 µg/ml solution was added to the blots which were incubated for 24-48 h at 4°C. Blots were then probed with a 1:4000 dilution of extravidin-peroxidase for 1 h at room temperature, and binding visualised with 3,3',5,5'-tetramethylbenzidine (TMB) substrate (Sigma-Aldrich, St Louis, MO). Probed blots were washed 3 x with PBS and imaged using a GS-800™ Calibrated Densitometer (Bio Rad, Hercules, CA).

Lectin	Abbreviation	Company	Specificity
Peanut Agglutinin	PNA	Vector laboratories	Gal $\beta$ 1-3GalNAc $\alpha$ 1-Ser/Thr (T-antigen)
<i>Ricinus communis</i> Agglutinin	RCA	Vector laboratories	Gal $\beta$ 1-4GlcNAc $\beta$ 1-R (Terminal $\beta$ Gal)
<i>Ulex europaeus</i> Agglutinin	UEA	Vector laboratories	Fuc $\alpha$ 1-2Gal-R (Fuc)
Wheat germ Agglutinin	WGA	Vector laboratories	GlcNAc $\beta$ 1-4GlcNAc $\beta$ 1-4GlcNAc (N-acetylglucosamine), Neu5Ac (Sialic acid)
Sambucus Nigra	SNA	Vector laboratories	Neu5Ac $\alpha$ 2-6Gal $\beta$ 1-4GalNAc
Maackia Amurensis II	MALII	Vector laboratories	Neu5Ac $\alpha$ 2-3Gal $\beta$ 1-4GalNAc

Table 5 | Table of lectins and their specificities

#### 2.4.4 AFM for visualisation of mucins

Purified mucins, re-suspended in H<sub>2</sub>O were visualised using AFM as described previously (Gunning et al., 2013). Briefly, 4 µl of prepared mucins (See section 2.4.1) were deposited onto freshly cleaved mica, incubating for 1 min to allow adsorption to take place, and blowing the excess liquid off using argon. Imaging was carried out using a MFP-3D BIO atomic force microscope (Asylum Research, Goleta, CA) in alternating current (AC) mode in air, at a scan rate of 1 Hz. The cantilevers used were Olympus AC160TS (Olympus, Tokyo, Japan). Imaging was carried out both before and after treatment of mucins with 1 Kunitz unit of DNase enzyme at 37°C O/N.

#### 2.4.5 Force spectroscopy

For force spectroscopy, purified mucins (See section 2.4.1) were first dialysed against sodium chloride (NaCl). Briefly, 100 µl of mucin sample at a concentration of 1 mg/ml was prepared and dialysed against at least 10x the sample volume of 1 M NaCl solution for 16 h at room temperature, using 100 kDa Micro Float-A-Lyzer dialysis devices (VWR International, Radnor, Pennsylvania). Force spectroscopy and analysis was performed as previously described (Gunning et al., 2013). Briefly, cleaned glass slides were functionalized with prepared mucins by silanization with 3-mercaptopropyltrimethoxysilane (MTS, Sigma-Aldrich, St Louis, MO), and attachment of a heterobifunctional PEG-3000 linker molecule. To link the mucin chains *via* their N-termini to the derivatized slides, a 100 µl aliquot of the 1 M NaCl dialyzed mucin was dispensed onto the slides for 1 h, followed by washing in PBS. To amine-cap any unreacted succinimide groups on to the glass, the surfaces were incubated in 10 mg/ml glycine, washed in PBS to remove any unbound glycine, and inserted into the liquid cell of the AFM. Measurements were carried out with a range of immobilized lectins (Table 5) on functionalized silicon nitride AFM tips in direct current (DC) mode under liquid and at a scan rate of 1 hertz (Hz). A 3 µm<sup>2</sup> area of the mucin coated glass slide was probed for interactions with each lectin of interest. The data were captured by 'force-volume' (FV) (at a rate of 2 µm.s<sup>-1</sup> in the Z direction and a pixel density of 32 x 32), a combination of imaging and carrying out force-distance measurements at each sample point (i.e. pixel). The piezo scanner is moved in the Z direction over the selected scan area and records the subsequent deflection of the cantilever as it is pushed into (maximum load force 300 pN), then retracted away from the sample surface. This produces a matrix of 1024 force-distance curves relating to the image coordinates. The spring constant, k, of the cantilevers was determined by fitting

the thermal noise spectra (Hutter and Bechhoefer, 1993), yielding typical values in the range 0.03-0.06 N.m-1.

#### 2.4.5.1 *Peak distance analysis*

The separation distances between individual adhesive events in the force spectra were determined by a bespoke Excel macro (Gunning et al., 2008; Gunning et al., 2013), by quantifying the distances of the turning points in the retraction curves. Discrimination of the adhesive events from noise was carried out by setting a negative peak threshold of 6x the amplitude of the noise level in the data.

#### 2.4.6 HPAEC-PAD for sialic acid quantification

For quantification of sialic acid on mucins purified from human mucosal lavage or mouse mucus scrapings (See sections 2.4.1 and 2.6.1), glycans were first liberated from purified mucins by acid hydrolysis following the method of Rohrer and coll. (Rohrer et al., 1998), and analysed by high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Briefly, a reaction mixture was prepared containing 0.5 mg of the sample of interest, 0.1 M HCl and dH<sub>2</sub>O to a final volume of 2 ml. The mixture was incubated at 80°C for 1 h, then dried using a SpeedVac-concentrator. The dried material was re-suspended in 225 µl dH<sub>2</sub>O plus 25 µl of a 1 mg/ml solution of 3-deoxy-d-glycerol-galacto-2-nonulosonic acid (KDN final concentration 0.01 mg/ml, Sigma-Aldrich, St Louis, MO). A 10 µl vol of each sample was injected into an amino trap and CarboPac PA10 carbohydrate column installed in a Dionex ICS-5000 chromatography system (Thermo Fisher Scientific, Waltham, MA). A standard mixture of 0.1 mg/ml sialic acid (Sigma-Aldrich, St Louis, MO) containing 0.1 mg/ml KDN, sialic acid at a range of concentrations (from 70 µg/ml to 0.14 µg/ml) and 0.5 mg fetuin (Sigma-Aldrich, St Louis, MO) were used as a positive controls, and 0.5 mg bovine serum albumin (BSA) as a negative control. Eluents were comprised of 100 mM sodium hydroxide (NaOH) for eluent A and 100mM NaOH containing 1 M sodium acetate (Thermo Fisher Scientific, Waltham, MA) for eluent B. Using a flow rate of 1 ml/min and column temperature of 30°C, separation of sialic acid was achieved with the following gradient protocol:

<u>Time</u>	<u>Eluent Composition</u>
0 – 10 min	7-30 % Eluent B
10 – 11 min	20 % Eluent B
11 – 12 min	30-7% Eluent B
12 - 28 min	7 % Eluent B

Samples were monitored by pulsed amperometric detection (PAD) in the carbohydrate 4-potential waveform.

As the detection of the sialic acid standard curve was not linear, to quantify the amount of sialic acid in samples, the raw sialic acid peak area values (nC\*min) were first corrected. Briefly, the data were multiplied by a constant obtained by measuring the ratio between the nonlinear standard curve and an extrapolation of the linear region. The concentration of sialic acid was then determined according to the following calculation, where F is the ratio of responses between a known concentration of the internal standard, KDN and the analyte, sialic acid:

$$\frac{\text{Area internal standard peak}}{\text{Area of sialic acid peak}} = F \frac{\text{Concentration of internal standard}}{\text{Concentration of sialic acid}}$$

#### 2.4.7 MALDI-TOF analysis of mucin glycans from mucus samples

Glycans were first liberated from purified mucins (See sections 2.4.1 and 2.6.1) by alkaline borohydride treatment and analysed using matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (MS). Briefly, the mucins were submitted to  $\beta$ -elimination for 20 h at 45°C under reductive conditions (0.1 M NaOH, 1 M NaBH<sub>4</sub>). The reaction was stopped by adding Dowex 50W X8, hydrogen form, strongly acidic 16-50 mesh (Sigma-Aldrich, St Louis, MO) and filtered before 3 rounds of co-evaporation with methanol. Remaining salts were then removed by Carbograph (Grace, Columbia, MD).

Released O-glycans were permethylated by solubilisation in 200  $\mu$ l dimethyl sulfoxide, followed by addition of NaOH and 300  $\mu$ l iodomethane in anhydrous conditions. The mixture was vigorously shaken at room temperature for 90 min. The permethylation reaction was stopped by addition of 1 ml 5% acetic acid. Purification of permethylated O-glycans was carried out using HLB Oasis cartridges. Briefly, cartridges were activated by methanol, equilibrated with a 1:19 ratio of methanol to water (vol:vol), and samples were added to the cartridges. Cartridges were washed with a 1:19 ratio of methanol to water (vol:vol) and the permethylated O-glycans eluted with methanol.

MALDI-TOF and TOF/TOF-MS data were acquired using the Bruker Autoflex analyser mass spectrometer (Applied Biosystems, Foster City, CA) in the positive-ion and reflectron mode by using 2,5-dihydroxybenzoic acid (DHB; 10 mg/ml in 7:3 ratio of methanol to water, Sigma-Aldrich, St Louis, MO) as the matrix.

#### 2.4.8 RNA extraction and cDNA synthesis from human mucosal biopsies

Biopsies stored at -80°C were removed and placed on dry ice. To the frozen whole biopsies, 200 µl of RNAlater-ICE solution (Thermo Fisher Scientific, Waltham, MA) was added, and the biopsies transferred to -20°C O/N. The following day, biopsies were removed using sterile forceps, transferred into a fresh Eppendorf containing 350 µl buffer RLT (RNeasy Mini Kit, Qiagen, Hilden, Germany) and 3.5 µl β-mercaptoethanol (Sigma-Aldrich, St Louis, MO), and incubated for 5 min at room temperature. Disruption was carried out with a disposable, autoclaved pellet pestle (Sigma-Aldrich, St Louis, MO), pipetting and vortexing the sample until it was fully lysed. Ribonucleic acid (RNA) extraction was then completed by following the RNeasy Mini Kit 'Purification of Total RNA from Animal Tissues' protocol from step 5, following steps D1-D4 in replacement of step 6, and carrying out two elution steps with 30 µl RNase free water. Extracted RNA was stored at -80°C until cDNA synthesis.

Prior to the reverse transcription (RT) reaction, template RNA was thawed on ice. The quality of RNA was assessed by using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA), according to the manufacturer's instructions. A measurement by Nanodrop was also taken to assess the 260/280 and 260/230 purity ratios, as well as to quantify the RNA. Elimination of gDNA (0.5 µg RNA per reaction, scaled up depending on volume of cDNA required) and cDNA synthesis was carried out using the Quantitect reverse transcription kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

#### 2.4.9 qRT-PCR of glycosyltransferase genes

For qRT-PCR, primers targeting glycosyltransferases were validated against cDNA derived from the mucus producing colorectal adenocarcinoma HT29-MTX cell line (Table 6). Briefly, RNA was extracted and cDNA synthesised (1 µg RNA per reaction) as above and the resulting cDNA used in a two-fold template dilution series to generate standard curves. The RT-PCR reaction mixture and amplification conditions for samples and standards were as follows:

<u>qPCR reaction mix (per reaction)</u>		<u>Reaction conditions</u>		
Reagent	Vol	Temp	Duration	Cycles
SYBR master mix (Qiagen, Hilden, Germany)	5 $\mu$ l	95°C	5 min	x 40
Forward primer (10 $\mu$ M)	0.5 $\mu$ l	95°C	10 sec	
Reverse primer (10 $\mu$ M)	0.5 $\mu$ l	58°C	45 sec	
dH <sub>2</sub> O	2 $\mu$ l	95°C	15 sec	
DNA template	2 $\mu$ l	60°C	1 min	
		95°C	15 sec	
		60°C	15 sec	

For quantification of glycosyltransferase gene expression in the samples, qRT-PCR was run in triplicate under the same conditions described above, using a four-fold dilution of template cDNA. Following qRT-PCR, data were analysed by measuring the cycle threshold (Ct) values for individual bacterial groups at a threshold of 0.05 Rn. Data were expressed as  $-D_{Ct}$  ( $-(Ct_{18S \text{ reference}} - (Ct_{\text{gene of interest}}))$ ) for downstream statistics and analysis.

Primer names	Sequence	Reference
C1GalT1-3F	GAGGTGGCTTCTTTCAAATACGACCC	(Blixt et al., 2011)
C1GalT1-3R	CATCTCCCCAGTGCTAAGTCTTCAATG	
C3GnT-3F	CGGCTAGACTATCTCTTCATCCTC	(Blixt et al., 2011)
C3GnT-3R	CCACTCACTTGTAACAGTGGAAA	
ST3Gal-I-F	TGGTCCTGGAGCTCTCCGAGAA	(Sakuma et al., 2012)
ST3Gal-I-R	GACTGTCTATCTCAGGCCATAAGAAGA	
ST3Gal-II-F	GATGATGCTGCAGCCCCAGTTC	(Sakuma et al., 2012)
ST3Gal-II-R	ACATCCTGCTCAAAGCCCACGGTT	
ST3Gal-III-F	TGTTCCCTGGATGACTCCTTTTCGCA	(Chachadi et al., 2013)
ST3Gal-III-R	CTTGTTGGCAAGAACGCCTCCATT	
ST3Gal-IV-F	ATGAGCAGATCACGCTCAAGTCCA	(Chachadi et al., 2013)
ST3Gal-IV-R	TCCCATCTCCAGCATCCGCTTAAT	
ST3Gal-VI-2F	TCTATTGGGTGGCACCTGTGGAAA	(Chachadi et al., 2013)
ST3Gal-VI-2R	TGATGAAACCTCAGCAGAGAGGCA	
ST6GalNAc-I-F	CAGAGGCACAATCATGGAAG	(Chik et al., 2014)
ST6GalNAc-I-R	GCTGACTTTTGGGAATGAGC	
ST6GalNAc-II-2F	AAGCTGCTACATCCGGACTTCA	(Ding et al., 2009)
ST6GalNAc-II-2R	GGGACAGATCGTGGTTTGCATA	
Fut2-2F	CCTTCAGCAGGACCAGGTGAGA	(Sakuma et al., 2012)
Fut2-2R	GGTCCCAGTGCCTTTGATGTTGAG	
18S-F	CACGGGAAACCTCACCCGGC	This study
18S-R	CGGGTGGCTGAACGCCACTT	

**Table 6 | Primers used to quantify the expression of human glycosyltransferases**

## 2.5 Methods for *in vitro* culture assays of *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835

### 2.5.1 Routine culture conditions

The growth of *R. gnavus* and *A. muciniphila* followed strict anaerobic conditions.

Equipment and media were pre-reduced in an anaerobic cabinet (Don Whitley, Shipley, UK) for at least 48 h. All cultures were grown at 37°C with a gas mixture of 5% carbon dioxide (CO<sub>2</sub>), 10% hydrogen, 85% nitrogen.

Starter cultures of *A. muciniphila* ATCC BAA 835 (ATCC, Manassas, VA) and *R. gnavus* CC55\_001C (kindly provided by Emma Allen-Vercoe's lab, University of Guelph, Ontario, Canada) were routinely grown from glycerol stocks in Brain-heart infusion media supplemented with 5 g/l of Bacto™ yeast extract (Becton, Dickinson and Company, Franklin Lakes, NJ) and 5 mg/l of hemin (Sigma-Aldrich, St Louis, MO) (BHI-YH).

Depending on the type of growth assay, yeast extract-casein hydrolysate-fatty acid (YCFA) or CP media were used to further assess the ability of *R. gnavus* and *A. muciniphila* to grow on specific carbohydrate sources (See sections 2.5.2 and 2.5.5).

YCFA medium consisted of (per 1 L): 10 g casitone, 2.5 g yeast extract, 4 g NaHCO<sub>3</sub>, 1 g L-cysteine hydrochloride, 450 mg K<sub>2</sub>HPO<sub>4</sub>, 450 mg KH<sub>2</sub>PO<sub>4</sub>, 900 mg NaCl, 90 mg MgSO<sub>4</sub>·7H<sub>2</sub>O, 90 mg CaCl<sub>2</sub>, 1 mg resazurine, 10 mg hemin, 10 mg biotin, 10 mg cobalamin, 30 mg p-aminobenzoic acid, 50 mg folic acid and 150 mg pyridoxamine (Crosthair et al., 2013).

CP medium consisted of (per 1 L): 0.4 g KH<sub>2</sub>PO<sub>4</sub>, 0.53 g Na<sub>2</sub>HPO<sub>4</sub>, 0.3 g NH<sub>4</sub>Cl, 0.3 g NaCl, 4 g NaHCO<sub>3</sub> and 0.25 g Na<sub>2</sub>S<sub>7</sub>·9H<sub>2</sub>O, 0.1 g MgCl<sub>2</sub>·6H<sub>2</sub>O, 0.11 g CaCl<sub>2</sub>, 1 ml alkaline trace element solution (0.1 mM Na<sub>2</sub>SeO<sub>3</sub>, 0.1 mM Na<sub>2</sub>WO<sub>4</sub>, 0.1 mM Na<sub>2</sub>MoO<sub>4</sub>, 10 mM NaOH), 1 ml acid trace element solution (7.5 mM FeCl<sub>2</sub>, 1 mM H<sub>3</sub>BO<sub>3</sub>, 0.5 mM ZnCl<sub>2</sub>, 0.1 mM CuCl<sub>2</sub>, 0.5 mM MnCl<sub>2</sub>, 0.5 mM CoCl<sub>2</sub>, 0.1 mM NiCl<sub>2</sub>, 50 mM HCl), 0.5 mg resazurine and 3 g cysteine (Derrien et al., 2004).

### 2.5.2 Growth assays in carbohydrate supplemented media

The monosaccharides, D-glucose (Glc), Gal, N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), Fuc, lactose (Lac), N-acetylneuraminic acid (Neu5Ac), N-glycolylneuraminic acid (Neu5Gc) and type III pig gastric mucin (PGM) were purchased from Sigma-Aldrich (St Louis, MO).

Oligosaccharides, 2'-fucosyllactose (2'FL), 3-fucosyllactose (3'FL), lacto-N-neo-tetraose (LNnT) lacto-N-tetraose (LNT) and 6'-O-sialyllactose (6'SL) were provided by Glycom A/S (Lyngby, Denmark). 3'-sialyllactose (3'SL) and LacNAc were purchased from Carbosynth Limited (Campton, UK).

Growth assays were carried out using anaerobic basal YCFA medium supplemented with 27.7 mM of mono- or oligosaccharides, as previously described (Crosthwaite et al., 2013). Growth was determined spectrophotometrically by measuring changes in optical density at 48 h, using a wavelength of 595 nm and comparing to the same medium without bacterial inoculum.

### 2.5.3 Purification of Sigma type III PGM

Mucin for *in vitro* culturing assays was purified from commercially available porcine gastric mucin (Sigma-Aldrich, St Louis, MO), following a method previously described (Gunning et al., 2013). Briefly, 5 g mucin was dissolved in 200 mL of buffer containing 0.1 M NaCl, 0.02 M sodium phosphate pH 7.8 and 3 drops of toluene. The solution was stirred for 1 h at room temperature and then 2 M NaOH added to adjust the pH to 7-7.4. The adjusted solution was stirred O/N at room temperature. Following stirring, the solution was centrifuged for 1 h at 10,000 xg and 4°C. To the supernatant, pre-chilled absolute ethanol was added to a concentration of 60 %. The solution was stirred for 30 min at 4°C and then centrifuged at 10,000 xg for 1 h at 4°C. Following centrifugation, the pellet was dissolved in 150 ml 0.1 M NaCl and stirred O/N at 4°C. Pre-chilled ethanol was added to the solution as above, stirred for 30 min and centrifuged for 30 min at 10,000 xg. The pellet was then allowed to air dry before being dissolved in 100 ml H<sub>2</sub>O with O/N stirring at 4°C. The solution was dialysed O/N against H<sub>2</sub>O using a 100 kDa MWCO membrane before being freeze dried ready for use.

### 2.5.4 Purification of mucin from the LS174T cell line

Mucin for *in vitro* culturing assays was purified from the LS174T colon adenocarcinoma cells (ATCC® CL-188) (Tom et al., 1976) by isopycnic ultracentrifugation. Briefly, guanidinium chloride (GuCl) buffer was prepared (100 ml, containing 6 M GuCl, 6.36 mM EDTA, 9.8 mM benzamidine, 5 mM N-ethylmaleimide, 10 mg soy bean trypsin inhibitor, 3 mM sodium azide and 1 mM phenylmethanesulfonyl fluoride (PMSF) dissolved in 500 µl isopropanol).

Mucus was harvested from cell lines by collecting 15 ml of spent medium from cells cultured in a 75 cm<sup>2</sup> culture flask (Starstedt, Nümbrecht, Germany) every 3-4 days. The medium was freeze dried and the dried material re-suspended in GuCl buffer (15-20 ml per 10 tubes of spent medium) to a final concentration of 4 M and stirred for 2 h at a minimum of 4°C. The

mixture was centrifuged at 12,000 rpm for 30 min at 4°C, and the pellet (containing MUC2) and supernatant (containing other mixed mucins) separated.

The supernatant, containing mixed mucins, was diluted to 4 M with PBS, and DTT added to a concentration of 9.33 mM. The density of the solution was adjusted to 1.4 g/ml with caesium chloride. The samples were ultracentrifuged at 58,000 rpm for 48-72 h at 20°C. Using a syringe, 1 ml fractions were collected from the top of the tubes and weighed. Fractions with a density of 1.35 to 1.50 g/ml were stained with Periodic Acid Schiff (PAS) staining to confirm the presence of mucins. Briefly, 20 µl of each fraction was slot blotted onto a nitrocellulose membrane. Membranes were rinsed with PBS. The membrane was incubated for 30 min in a solution containing 1% periodic acid and 3% acetic acid. The membrane was washed with PBS and stained with Schiff's fuschin-sulfite reagent (Sigma-Aldrich, St Louis, MO) for 15 min in the dark. Finally, the membrane was washed in 0.1% sodium metabisulfite in 1 mM HCl. PAS positive fractions were dialysed against pure water containing 50 mM ammonium bicarbonate in 12-14 kDa dialysis tubing (SpectraPor, Spectrum labs, Rancho Dominguez, CA) for 2 days, changing the buffer twice daily. Dialysed samples were freeze-dried and stored at 4°C until further use.

The pellet, containing MUC2, was re-suspended in 5-10 ml reduction GuCl buffer (6M containing 0.6 M Tris pH 8.3 without protease inhibitors). DTT was added to a concentration of 10 mM. The mixture was incubated for 4 h at 45°C in a water bath. After cooling, iodoacetamide was added to a concentration of 25 mM and incubated O/N at room temperature in the dark. Samples were dialysed as above. Samples were freeze-dried and stored at 4°C until further use.

#### 2.5.5 Growth assays of *A. muciniphila* and *R. gnavus* in mucin medium

CP media containing 0.5% pPGM (Gunning et al., 2013) or 0.5% purified mucin from LS174T cells (See sections 2.5.3 and 2.5.4), was prepared.

Starter BHI-YH cultures (See section 2.5.1) of *A. muciniphila* ATCC BAA 835 grown to ~48 h and *R. gnavus* CC55\_001C grown to ~24 h were subcultured 1/25 into mucin CP medium. To monitor bacterial growth, an aliquot of each culture was taken at the desired time point and centrifuged at 2,000 xg for 3 min. The supernatant and pellet were separated and individually frozen at -20°C until further use.

### 2.5.6 Monitoring *in vitro* growth of *A. muciniphila* and *R. gnavus* by qPCR

Bacterial pellets were thawed and the gDNA extracted using the GeneJET Genomic DNA purification kit (Thermo Fisher Scientific, Waltham, MA), following the gram negative or gram positive protocol, according to the manufacturer's instructions and as described in section 2.3.2. The concentration of bacterial gDNA extracted was determined using the high sensitivity fluorimetric double stranded DNA assay on a Qubit® 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA), according to the manufacturer's instructions.

For quantification of bacterial cells in cultures of *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C, standard curves of gDNA were first prepared. The theoretical genome mass of *A. muciniphila* (Genbank CP001071.1) and *R. gnavus* (Genbank AZJF00000000.1) was determined using the following equation:

$$\text{Genome mass (kg)} = ((N_a \times MW_a) + (N_c \times MW_c) + (N_g \times MW_g) + (N_t \times MW_t) + 157.9) \times \text{Atomic mass constant (1.66} \times 10^{-27})$$

$N_a$ = Number of adenine base pairs in double stranded DNA

$N_c$ = Number of cytosine base pairs in double stranded DNA

$N_g$ = Number of guanine base pairs in double stranded DNA

$N_t$ = Number of thymine base pairs in double stranded DNA

$MW_a$ = Molecular weight adenine

$MW_c$ = Molecular weight cytosine

$MW_g$ = Molecular weight guanine

$MW_t$ = Molecular weight thymine

The gDNA extracted from *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C grown in BHI-YH media was diluted to a concentration equating to  $10^7$  cells, based on the calculated theoretical genome mass. Serial dilutions ranging from  $10^6$  to  $10^1$  cells were then generated in H<sub>2</sub>O containing 5 µg/ml salmon testes DNA (Sigma-Aldrich, St Louis, MO). qPCR of standards prepared from cells grown in BHI-YH, and samples of bacterial cultures grown in mucin media were carried out as in section 2.3.2, using a 2 µl aliquot of each serial dilution template or sample template. For samples of bacterial cultures grown in mucin media, the reaction mixture was modified by replacing the 2 µl H<sub>2</sub>O with 2 µl salmon testes DNA at a concentration of 5 µg/ml. Quantification of *A. muciniphila* and *R. gnavus* was carried out by comparison of each reactions Ct value against the prepared gDNA standard curves.

### 2.5.7 $^1\text{H}$ NMR metabolite analysis of spent media

$^1\text{H}$  NMR was used to identify the presence and concentration of several metabolites produced by *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C grown in pPGM and LS174T mucin containing medium.

Sample supernatants collected in section 2.5.5 were thawed at room temperature and prepared for  $^1\text{H}$  NMR spectroscopy by mixing 400  $\mu\text{l}$  of extracts with 200  $\mu\text{l}$  NMR buffer (0.26 g  $\text{NaH}_2\text{PO}_4$  and 1.41 g  $\text{K}_2\text{HPO}_4$ ) made up in 100%  $\text{D}_2\text{O}$  (100 ml), containing 0.05%  $\text{NaN}_3$  (50 mg), and 1 mM sodium 3-(Trimethylsilyl)-propionate- $d_4$ , (TSP) (17 mg) as a chemical shift reference. The samples were mixed, and 500  $\mu\text{l}$  was transferred into a 5 mm NMR tube for spectral acquisition. The  $^1\text{H}$  NMR spectra were recorded at 600 MHz on a Bruker Avance spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) running Topspin 2.0 software and fitted with a 5 mm TCI cryoprobe. Sample temperature was controlled at 300 K. Each spectrum consisted of 64-1032 scans of 65,536 complex data points with a spectral width of 12.3 ppm (acquisition time 2.66 s). The “noesypr1d” pre-saturation sequence was used to suppress the residual water signal with low power selective irradiation at the water frequency during the recycle delay ( $D1 = 3$  s) and mixing time ( $D8 = 0.01$  s). Spectra were transformed with a 0.3 Hz line broadening, manually phased, baseline corrected, and referenced by setting the TSP methyl signal to 0 ppm. The metabolites were quantified using the software Chenomx NMR suite 7.6™. Metabolites were identified using information found in the literature or on the web (Human Metabolome Database, <http://www.hmdb.ca/>) and by use of the 2D-NMR methods, COSY, HSQC, and HMBC.

## 2.6 Methods for *In vivo* colonisation assays of germ-free mice with *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835

### 2.6.1 Gavaging of germ-free mice and collection of samples

*A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C were grown from glycerol stocks in BHI-YH media as described above (See section 2.5.1). At 48 h growth for *A. muciniphila* and 24 h growth for *R. gnavus*, a 1 ml aliquot was taken to quantify the number of cells in the culture by gDNA extraction and qPCR. Prior to gavaging, the remaining culture was concentrated 50x, by centrifugation for 5 mins at 10,000 xg, removal of the supernatant and resuspension of the pellet in a 0.02 vol of pre-reduced sterile PBS. Germ-free C57BL/6J mice aged between 11 and 14 weeks were each gavaged with 100 µl of bacterial culture. Mice of the same age were kept germ-free as controls in a separate isolator.

Faecal samples were taken from the mice at day 0 (pre-gavaging), and at intermittent time points throughout the experiment. At the endpoint of the experiment, the mice were euthanized *via* exposure to carbon dioxide. Colon mucus scrapings were collected by cutting longitudinally along the length of the intestines and gently scraping the mucosa with a glass slide. Collected mucus was placed into 6 M GuCl buffer (See section 2.5.4) and stored at 4°C until further use. In addition, caecal contents were collected and placed immediately on dry ice, followed by storage at -80°C until further use.

### 2.6.2 Monitoring colonisation of *A. muciniphila* and *R. gnavus* by qPCR

gDNA was extracted from faecal samples (See section 2.6.1) using the FastDNA™ Spin kit for soil (MP Biomedicals, Santa Ana, CA). The extraction was carried out following the manufacturer's instructions, with minor amendments. Briefly, the weight of faecal material was measured in tared tubes. To each faecal sample, a 978 µl aliquot of sodium phosphate buffer and 122 µl aliquot of MT buffer (provided in the kit) was added and the sample incubated at 4°C for 1 h. The sample was added to a lysing matrix E tube, and homogenised in a FastPrep-24 instrument (MP Biomedicals, Santa Ana, CA) for 3 x 40 s at speed setting of 6.0 with 5 min on ice between each beating step. The lysed sample was centrifuged at 14,000 x g for 10 min to pellet debris, and the supernatant transferred to a clean 2 ml microcentrifuge tube. A 250 µl aliquot of protein precipitation solution was added and mixed by shaking the tube 10 times by hand. The sample was centrifuged at 14,000 x g for 5 min and the supernatant transferred to a clean 15 ml tube, to which 1 ml of binding matrix was added. The tube was inverted by hand for 2 min and then placed in a rack for 3 min to allow

settling of silica matrix. A 800 µl aliquot of supernatant was carefully removed and discarded, and the remainder of the supernatant re-suspended with the binding matrix. A 600 µl aliquot of the mixture was transferred into a SPIN filter column and centrifuged for 1 min at 14,000 x g. The flow through was discarded and the process repeated with the remainder of the re-suspended binding matrix. To the column, 500 µl of SEWS-M buffer was added and the pellet re-suspended by gentle pipetting. The mixture was centrifuged at 14,000 x g for 1 min, the flow through discarded and the pellet washed again with SEWS-M. Following removal of the flow through, the column was centrifuged again at 14,000 x g for 2 min to remove any residual liquid. The column was placed into a fresh catch tube, and the SPIN filter air dried for 5 min at room temperature before re-suspending the binding matrix in 100 µl DES, and centrifuging at 14,000 x g for 1 min, to elute bound DNA.

The concentration of extracted gDNA was determined using a broad range fluorimetric double stranded DNA assay on a Qubit® 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA). Each sample was diluted to 2.5 ng/µl according to this measurement, and qPCR was carried out as in section 2.3.2, using 2 µl of this template. Quantification of *A. muciniphila* and *R. gnavus* was carried out by comparison of reaction Ct values against gDNA standard curves, prepared as in section 2.5.6, but with minor modifications to the reaction mixture to include 1 µl of template DNA (rather than 2 µl), 2 µl of H<sub>2</sub>O and 1 µl of DNA extracted from the faecal samples of germ-free mice.

### 2.6.3 Purification of mucins from mouse mucus scrapings and analysis of glycosylation

Mucin purification was carried out from mucus scrapings of the colon (See section 2.6.1) as previously described (See section 2.5.4). Samples were freeze-dried and 1 mg used in preparation and analysis of mucin glycosylation by MALDI-TOF as described above (See section 2.4.7). Further samples were re-suspended to a concentration of 0.5 mg/ml, as determined by a Nanodrop measurement, and the relative abundance of sialic acid quantified by HPAEC-PAD as previously described (See section 2.4.6).

### 2.6.4 <sup>1</sup>H NMR metabolite analysis of caecal contents

<sup>1</sup>H NMR was used to identify the presence and concentration of metabolites produced in the caecal contents of gnotobiotic mice mono- and co-colonised with *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C.

Briefly, 50 mg of caecal contents was transferred to 600  $\mu$ l of NMR buffer as described in section 2.5.7. The mixture was homogenised using a sterile pellet pestle (Sigma-Aldrich, St Louis, MO) with a cordless motor device. Following homogenisation, samples were centrifuged for 5 min at 15,000  $\times$ g at 4°C, and the supernatant transferred into a 5 mm NMR tube for spectral acquisition as described in section 2.5.7.

## 2.7 Statistical analyses

Data for all experiments carried out in this study were performed using MATLAB, version 8.5.0.197613 (R2015a) and GraphPad Prism, version 5.04. One-way analysis of variance (ANOVA), Pearson correlation, Mann-Whitney test, Wilcoxon matched-pairs signed-rank test, and principal component analysis (PCA) were performed in this study. Degrees of significance are represented as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  and \*\*\*\* $p < 0.0001$ . The false discovery rate was controlled for at a threshold of  $q = 0.05$ .

## Chapter 3) Characterisation of the colonic mucus-associated microbiota

### 3.1 Introduction and objectives

The gut microbiota has been well studied in quiescent and active IBD. Several studies have observed quantitative and qualitative alterations in the composition of the gut microbiota in UC and CD. Broadly, IBD has been characterized by a depletion in Firmicutes and Bacteroidetes, an increase in Proteobacteria and Actinobacteria, and a decrease in microbial diversity (Becker et al., 2015). Specific alterations in the abundance of mucin degrading species are also evident. The most prevalent mucolytic bacterium, *A. muciniphila* is depleted in both UC and CD, whilst *R. gnavus* shows a more than fourfold increase in abundance (Png et al., 2010). Dysbiosis induced during inflammation, including changes in the prevalence of mucolytic bacteria, may impact on mucin glycosylation pattern by acting on the expression of glycosyltransferases and/or degrading specific mucin O-glycans. A large majority of studies which characterised the microbiota in disease have been carried out on faecal samples and biopsies (Joossens et al., 2011; Machiels et al., 2014; Png et al., 2010; Sokol et al., 2008; Sokol et al., 2009; Willing et al., 2010a). Faecal samples may not provide the most accurate representation of the microbiota composition that would be found in association the intestinal mucus. Sampling of the microbiota by luminal brush, whole mucosal biopsy and laser captured mucus revealed that the luminal and mucosal compartments do indeed host distinct microbiota populations (Lavelle et al., 2015). Mucosal microbiota populations can be analysed using biopsies and other sampling techniques, as summarised in table 2. However, neither faecal nor biopsy samples provide an easy method of sampling mucus, either due to its loss from the tissue or difficulty in its separation from stool (Hansson, 2012; Loktionov et al., 2016). As a result, elucidation of the mucin-microbiota relationship is troublesome.

In order to better understand the interaction between mucin and mucus-associated microbiota, it is necessary to obtain samples from the mucosal surface, and allowing both the microbiota and the associated mucus to be examined. We therefore utilised mucosal lavage, allowing us to assess both of these components from the same sample and from different colon sites (Li et al., 2011).

Two approaches were used to analyse the microbiota composition. Firstly, 16S sequencing, allowing us to obtain a broad overview of the mucus-associated microbiota composition, and

secondly, quantitative PCR (qPCR), allowing us to determine the abundance of more specific bacterial groups of interest including mucin degrading species.

### 3.2 Overall profiling of the mucus-associated microbiota by 16S sequencing

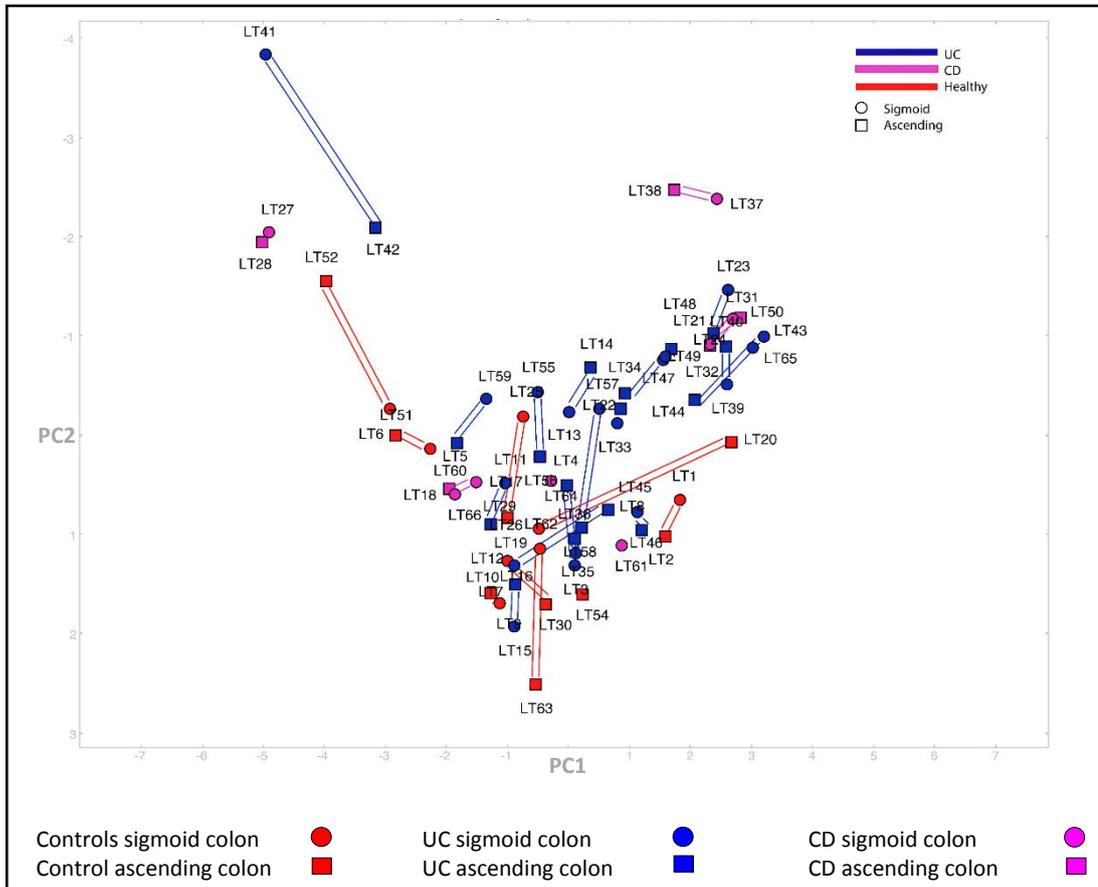
Bacterial DNA was extracted from 74 mucosal lavage samples from the sigmoid and ascending colon of 41 patients, representing a range of controls, UC and CD patients. Briefly, samples were homogenized and DNA extraction carried out according to a DNA mini kit protocol. To obtain a broad level of analysis we first performed 16S sequencing targeting the V2 region of the 16S gene. Sequencing was carried out on 65 colonic samples from 9 controls, 18 UC and 8 CD patients (See appendix 2 for patient metadata). Following normalisation, the number of reads per sample ranged from a minimum of 42,183 to a maximum of 107,8026. Out of a total of 65 samples, 55 samples displayed a normalised read number greater than 100,000 reads.

Analysis of the 16S sequencing data at the bacterial family level resulted in the detection of 65 families across samples analysed. Across individuals, the most common families detected in the microbial community were Lachnospiraceae, Ruminococcaceae and Bacteroidaceae (Sigmoid; Bacteroidaceae, 22.53%, Lachnospiraceae, 28.26%, Ruminococcaceae, 21.06%, Ascending; Bacteroidaceae, 19.62%, Lachnospiraceae, 29.00%, Ruminococcaceae, 22.10%). At the genus level, 138 genera were detected across the samples analysed, which mainly consisted of the *Bacteroides* genus and *Faecalibacterium* genus (Sigmoid; *Bacteroides*, 32.65%, *Faecalibacterium*, 20.02%, Ascending; *Bacteroides*, 28.83%, *Faecalibacterium*, 20.83%) across individuals (Fig. 9). Interestingly, no sequences were detected that classified under the genera to which the mucin degraders *R. gnavus* or *A. muciniphila* belong (See appendix 3).

#### 3.2.1 Effect of confounding variables on taxonomic profile

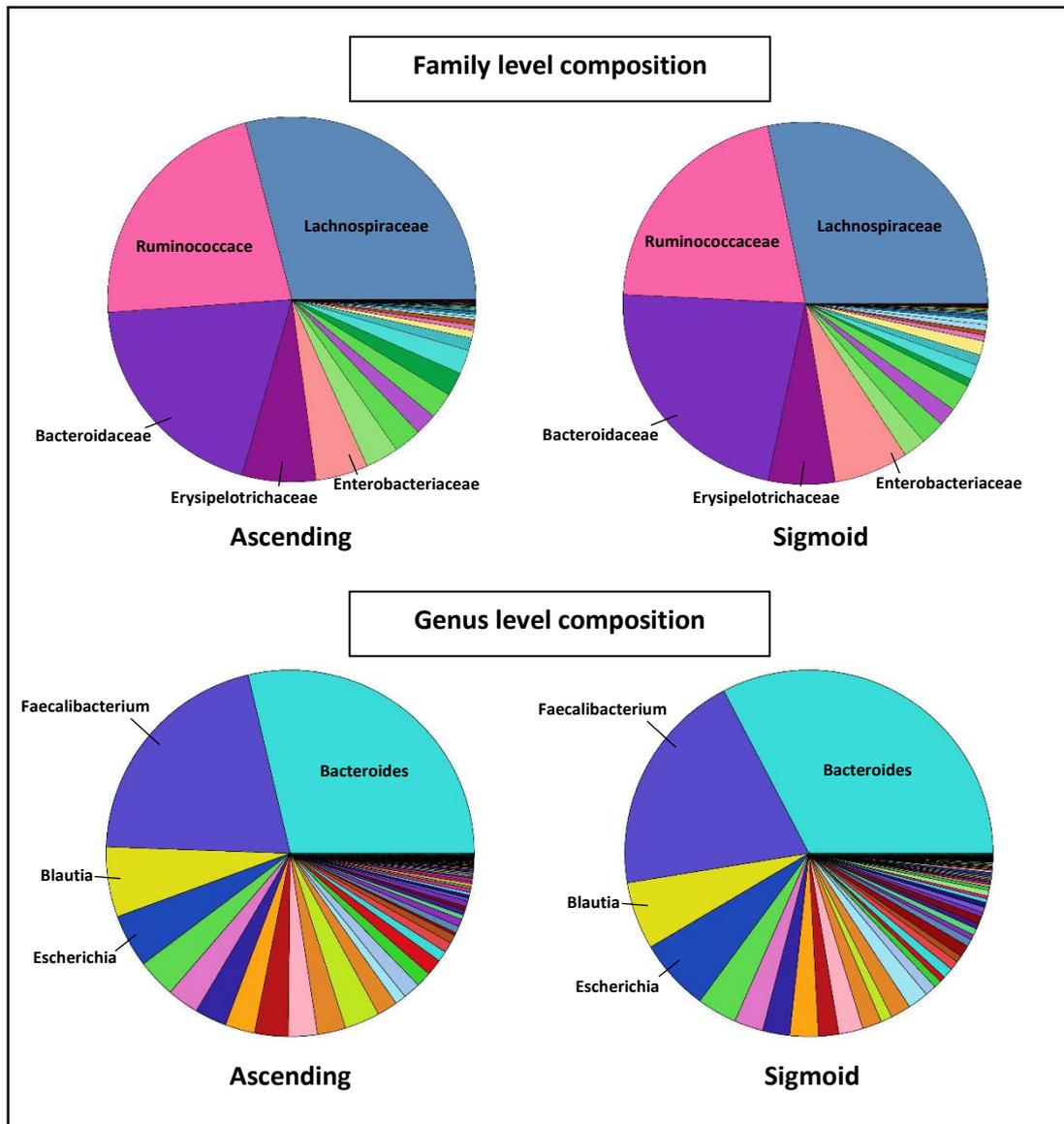
Based on the overall taxonomic profile established from the processed 16S sequencing data, the sigmoid and ascending samples from each individual were plotted on a principal coordinate analysis (PCoA) plot. Using this method of analysis, it was apparent that there was no specific clustering of the data based on disease state, but a high degree of inter-individual variation. Furthermore, age and other confounding factors did not contribute substantially to this variation. In contrast, many of the paired sigmoid and ascending samples from the same patient appeared to be plotted within a similar region on the PCoA plot,

suggesting that the microbial composition in both regions was similar (Fig. 9). Spatial conservation was further evident by the similarity of pie charts of the microbiota composition in the sigmoid and ascending colon, at both the family and genus level (Fig. 10).



**Figure 9 | PCoA plot of 16S taxonomic profiles of mucus associated microbiota from control, UC and CD patients**

*Taxonomic profiles of the mucus associated microbiota in control, UC and CD patients were determined by sequencing the V2 region of the 16S gene. Samples are plotted following decomposition of the variable matrix using the Bray Curtis dissimilarity index. Samples positioned closest on the PCoA plot are more similar in taxonomic composition, whilst those positioned further apart are more dissimilar. Paired samples from the same patient are indicated by connecting lines. Principal component 1 (PC1) summarises the largest amount of the total variance in the data (25.7%) whilst principal component 2 (PC2) summarises the second largest amount (15.7%).*



**Figure 10| Average family and genus level mucosal microbiota compositions of the ascending and sigmoid colon**

*Taxonomic profiles of the mucus associated microbiota in control, UC and CD patients were determined by sequencing the V2 region of the 16S gene. The pie charts represent the average family and genus level compositions across all mucosal microbiota samples sequenced. Comparison of the sigmoid and ascending regions shows a highly similar microbiota composition in these regions.*

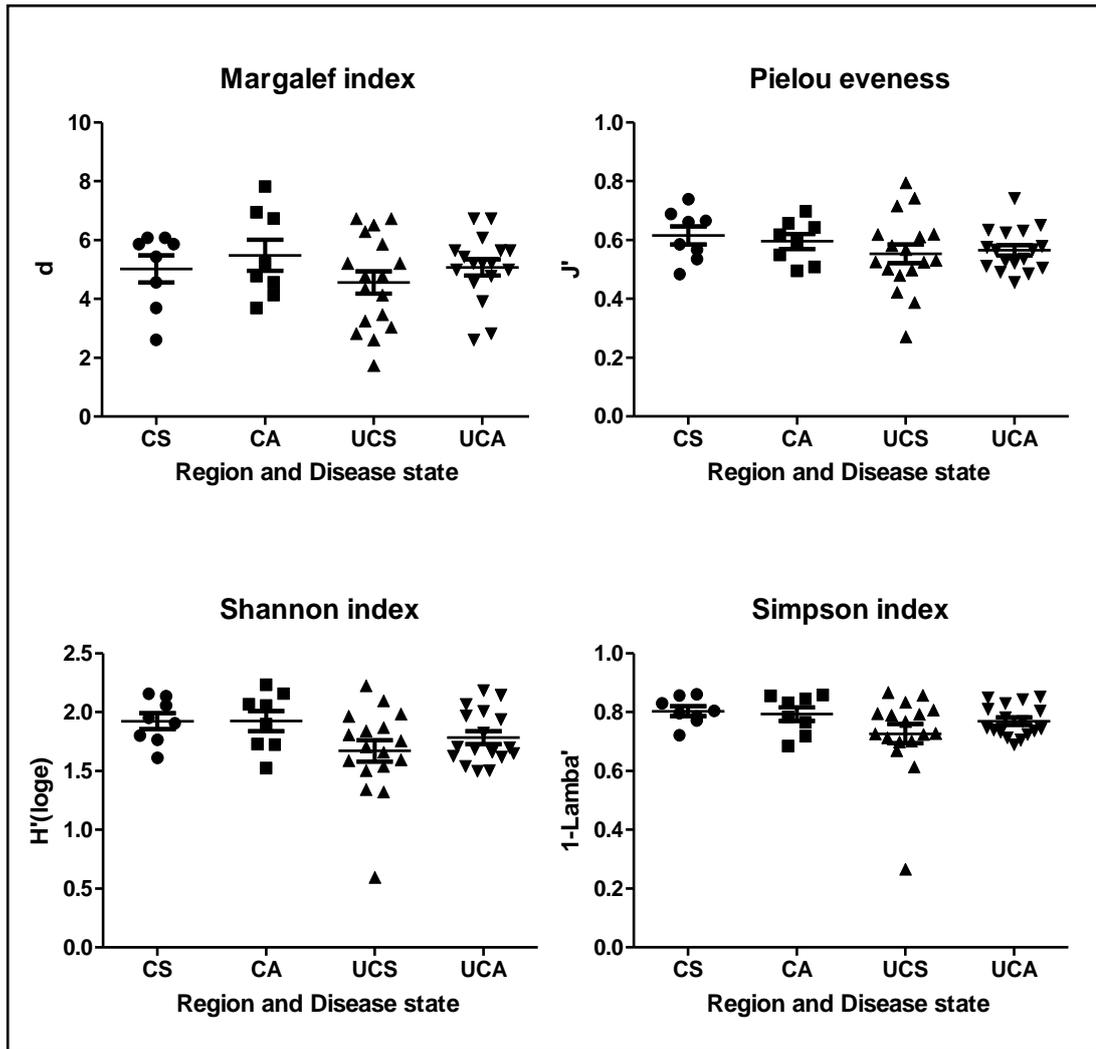
### 3.2.2 Diversity, richness and evenness of the microbiota from mucosal lavage samples

Several indices were measured to obtain a view of the diversity, richness and evenness of the microbiota within each sample. In order to carry out statistical analyses to assess for differences in these indices depending on location and disease state, it was necessary to sort the data. To prevent skewing of data, patients with measurements made only in one of the two colon locations were removed. This resulted in the selection of samples from 8 control patients, 17 UC patients and 5 CD patients with both sigmoidal and ascending colon locations measured. We also decided to exclude CD patient data at this stage given the small sample number in this group.

Margalef's richness scores ranged between 1.7 in the sigmoid colon of a UC patient and 7.8 in the ascending colon of a control patient. Pielou's evenness scores ranged between 0.27 in the sigmoid colon of a UC patient to 0.79, also in the sigmoid colon of a UC patient. Shannon diversity scores, combined measurements of evenness and richness, ranged between 1.3 and 2.2 apart from one UC sample from the sigmoid colon (14TB0097) which had an exceptionally low Shannon score, and therefore a low diversity, of 0.59. A similar observation was made for the Simpson index, also a measure of diversity. Simpson scores ranged between 0.6-0.8 for most samples except for 14TB0097 s, for which the score was exceptionally low at 0.26.

For statistical analyses of the data, Wilcoxon matched pairs signed rank nonparametric tests were first performed to determine if there were any statistically significant differences in diversity, richness or evenness between the ascending and sigmoid colon of either controls or UC patients. Using this method of analysis, no statistically significant differences could be observed in either the control patients or the UC group, with any of the indices measured. This observation was not affected by the presence or absence of outliers such as 14TB0097 S (Fig. 11) (Table 7).

To assess if there were any differences in richness, diversity and evenness depending on disease state, Mann Whitney tests were performed between the sigmoid colon of UC patients and controls, and likewise for the ascending colon. No significant differences could be observed between the mucus-associated microbiota of controls and UC patients using any of the measured indices (Fig. 11) (Table 7).



**Figure 11| Dot plots of diversity, evenness and richness indices of mucus associated microbiota from control and UC patients**

*Taxonomic profiles of the mucus associated microbiota in control and UC patients were determined by sequencing the V2 region of the 16S gene. Richness (Margalef index), evenness (Pielou index) and a combined measure of richness and evenness (Shannon and Simpson indices) of the microbiota was measured in each patient sample and plotted in dot plots. The mean and standard error of the values in each patient group is shown.*

*Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending*

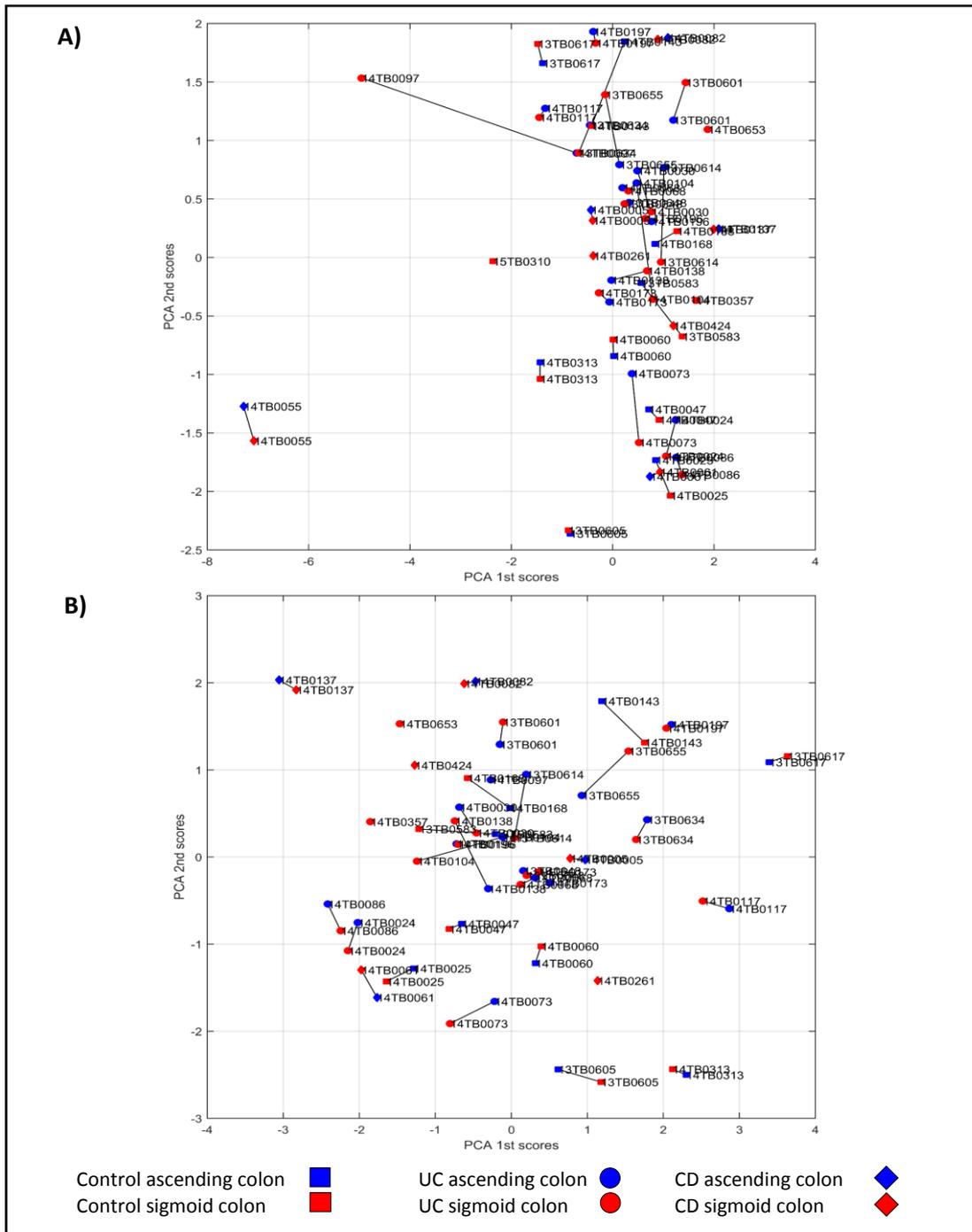
### 3.3 Quantitation of bacterial groups of interest by qPCR

Next, we used quantitative PCR (qPCR) with 16S specific primers to target and quantify bacterial groups of interest, including mucin degrading species, such as *A. muciniphila* and *R. gnavus*. The relative abundance of 9 bacterial groups was measured by qPCR as follows: Roseburia genus and *Eubacterium rectale* species, Faecalibacterium genus, Bacteroides and Prevotella genera, Bifidobacterium genus, Lactobacillaceae family and Weissella genus, *B. fragilis* species, *A. muciniphila* species, *R. gnavus* species. The latter three are of interest due to their mucin degrading properties, with *R. gnavus* and *B. fragilis* capable of catabolizing terminal sialic acid, and *A. muciniphila* degrading mucin with the synergistic action of several glycoside hydrolases (Tailford et al. 2015). Each qPCR reaction was carried out using bacterial DNA derived from human mucosal lavage samples. The Ct value for which the amplification signal of each bacterial group crossed a threshold level of 0.05 Rn was then used to quantify the group by expressing it as a percentage of total 16S, as measured using universal primers. qPCR was successfully performed with 68 of the available colonic samples (See appendix 2 for patient metadata).

#### 3.3.1 Effect of confounding variables on abundance of bacterial groups

As with the 16S sequencing, a PCA plot of all samples analysed by qPCR revealed a large amount of inter-individual variability. There was no specific clustering of samples based on disease, age, gender, or treatment. However, as previously observed, the PCA plot revealed that the sigmoid and ascending samples taken from the same patient mostly tended to cluster together (Fig. 12).

To further test the effect of continuous and categorical confounding variables, statistical analysis was performed separately on the sigmoid and ascending colon. The effect of the continuous variable, age, on microbiota composition was tested by Pearson correlation and for categorical variables, gender and treatment, by ANOVA. The results confirmed the observations of the PCA plot, as none of the significance levels obtained were lower than the threshold value after controlling the false discovery rate to  $q = 0.05$ , suggesting that there were no associations between any of these variables and the abundance of the 9 bacterial groups analysed in this region (See appendix 4).



**Figure 12| PCA plots of samples based on abundance of 9 specific bacterial groups quantified by qPCR**

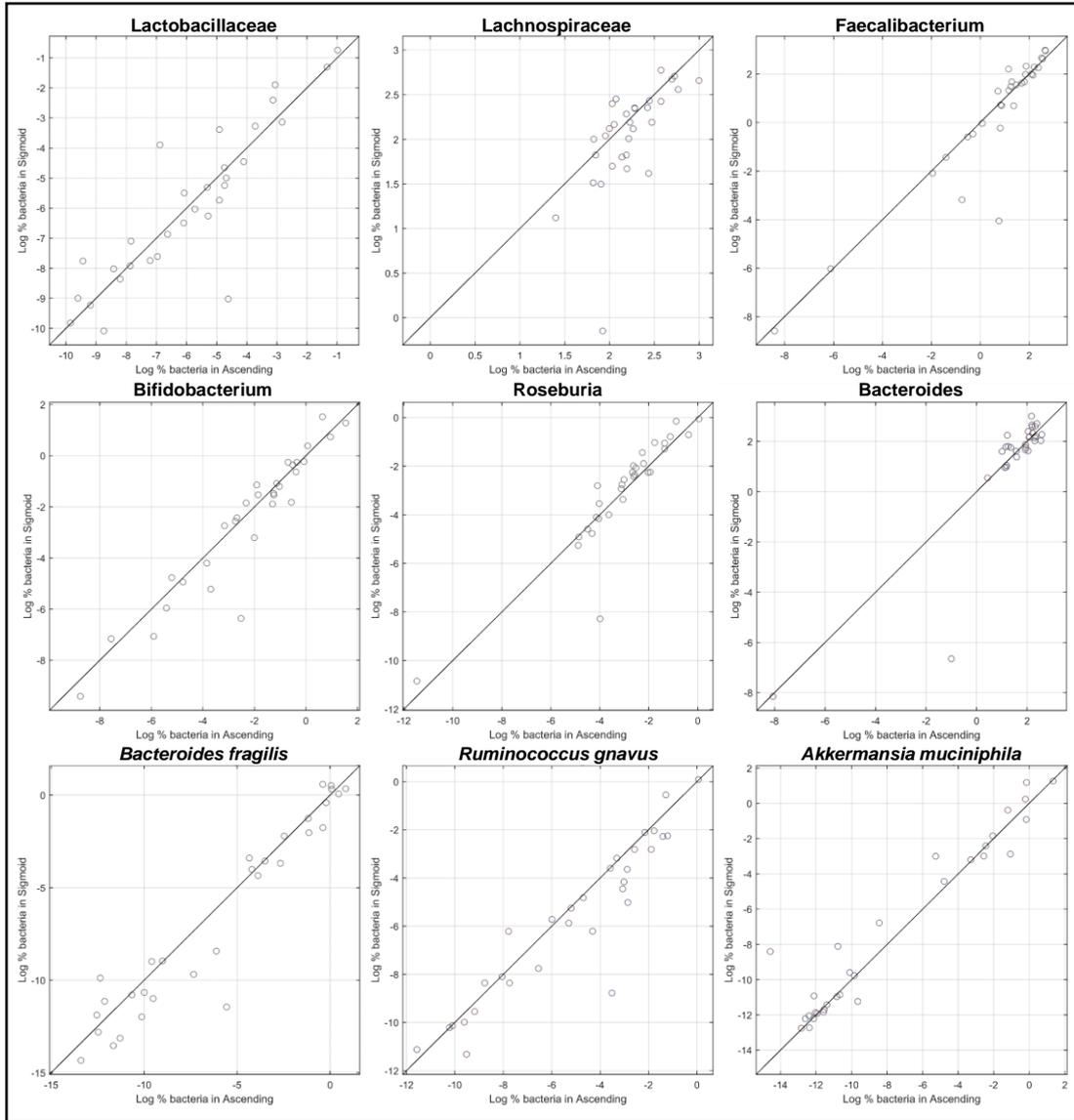
A) PCA plots of all control, UC and CD samples included in qPCR analysis.

B) PCA following removal of 4 patient outliers; 14TB0055, 14TB0097, 14TB0313, 15TB0310

The abundance of specific bacterial groups was determined by qPCR relative to the universal 16S gene. Samples are plotted following Eigen-value decomposition of the variable matrix. Samples positioned closest on the PCA plot are more similar in terms of bacterial abundance, whilst those further apart are more dissimilar. Paired samples from the same patient are indicated with connecting lines. The PCA 1<sup>st</sup> score summarises the largest amount of the total variance in the data whilst the PCA 2<sup>nd</sup> score summarises the second largest amount. In A), the PCA 1<sup>st</sup> and 2<sup>nd</sup> scores explain 32.3% and 16.1% of the variance respectively, whilst in B), they explain 25.2% and 16.2% of the variance, respectively.

### 3.3.2 Spatial variation in abundance of bacterial groups

For statistical analysis, data were sorted as in section 3.2.2; 9 control patients, 17 UC patients and 5 CD patients with both sigmoidal and ascending colon locations were selected. As previously, CD patients' data were excluded from the analysis due to the low number of samples in this group. The resulting data consisted of 52 samples in total. With the remaining control and UC patients, univariate analysis was carried out to determine if any statistically significant differences (Table 7) could be observed in the abundance of the nine bacterial groups of interest between the sigmoid and ascending colon. Wilcoxon matched pairs signed rank nonparametric tests were performed to determine the significance of the differences. Using this method of analysis, only one significant P value could be observed, explained by a higher amount of Lachnospiraceae in the ascending colon of UC patients when compared with sigmoid colon ( $P=0.0490$ ). However, this significance level was above that required to control the false discovery rate to below  $q=0.05$ . No significant differences were observed in any of the other possible comparisons, suggesting that the sigmoid and ascending colon are very similar in microbiota composition, consistent with the clustering of sigmoid and ascending colon on a PCA plot (Fig. 12). This spatial conservation was further confirmed by the remarkable line of unity observed with all bacterial groups analysed when plotting the sigmoid versus ascending data on scatter plots, irrespective of disease state (Fig. 13).

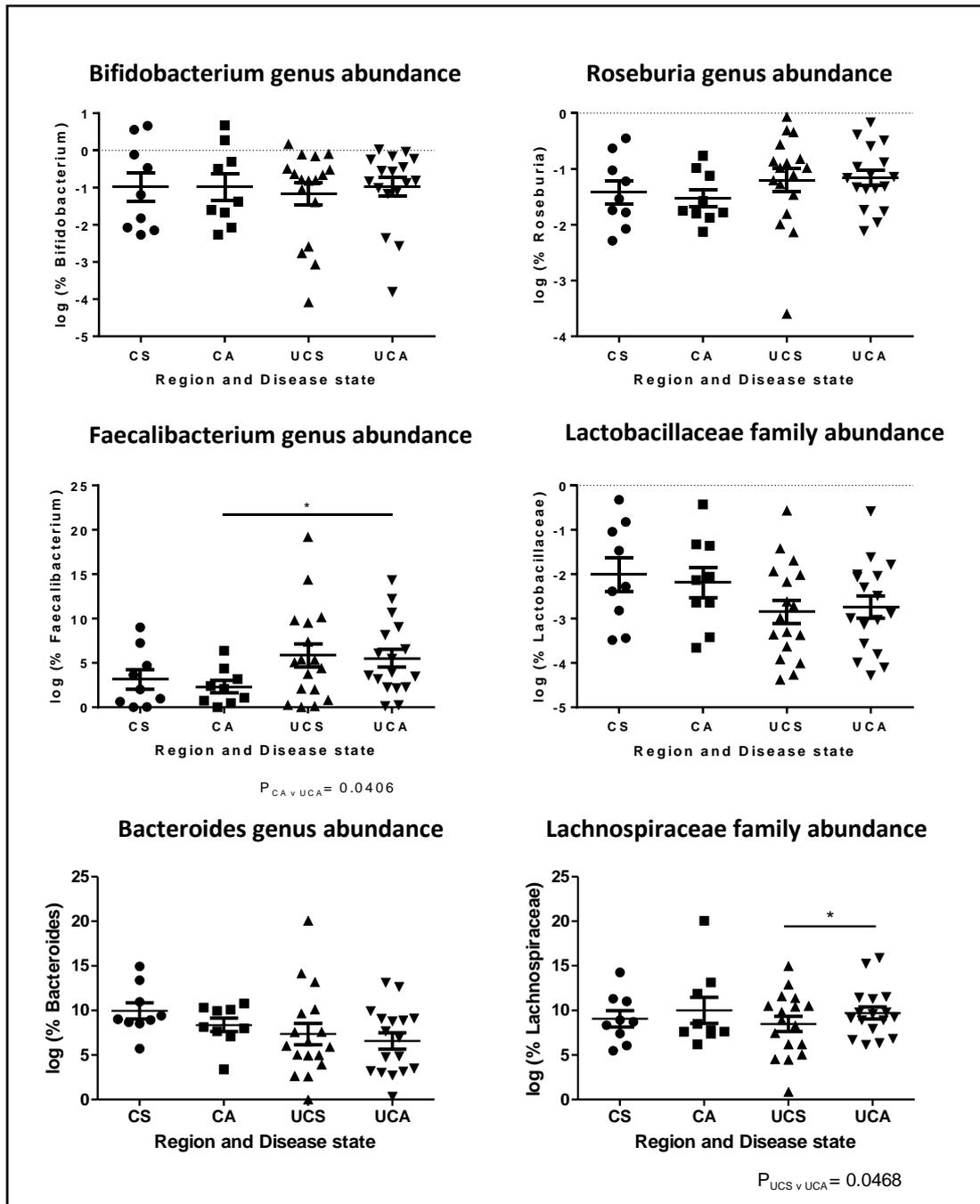


**Figure 13 | Scatter plots of abundance of each bacterial group in sigmoid versus ascending colon**

*The abundance of specific bacterial groups in control and UC patient microbiotas was determined by qPCR relative to the universal 16S gene. Analysis of each bacterial group was performed in triplicate in each patient sample. A log transformed average of the abundance of bacterial groups in the sigmoid colon is plotted against the abundance in the corresponding ascending colonic sample from the same patient. Spatial conservation in the abundance of these bacterial groups is indicated by the positive correlation between the sigmoid and ascending colon in all scatter plots.*

### 3.3.3 Effect of disease state on abundance of bacterial groups

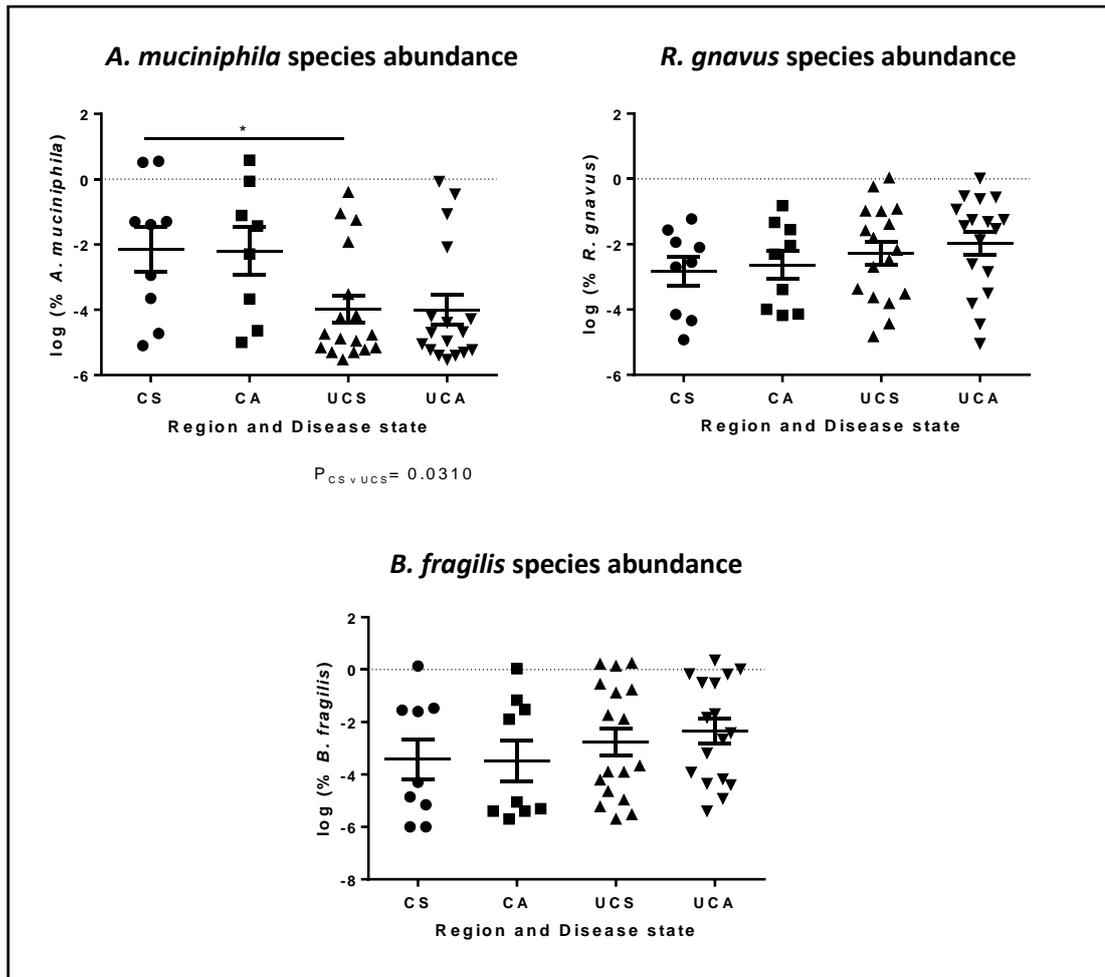
Next, we conducted an analysis to determine if there were any significant differences in bacterial abundance between controls and UC patients in each of the colonic regions sampled, sigmoid and ascending. A total of 18 comparisons were made between controls and UC, using Mann-Whitney nonparametric tests (Table 7). Only two of the returned P values were statistically significant. In the ascending colon, there was an apparent increase in the abundance of *Faecalibacterium* in the UC patients when compared controls ( $P=0.0406$ ) (Fig. 14). On the other hand, in the sigmoid colon, the abundance of *A. muciniphila* appeared to be significantly lower in UC versus controls ( $P=0.0310$ ) (Fig. 14). However, neither of these P values were lower than the threshold required to control the false discovery rate to below  $q = 0.05$ . The same trends could also be observed in the alternate region sampled by plotting the data on a dot plot, although failed to reach statistical significance (*Faecalibacterium*; sigmoid colon, UC>controls,  $P=0.1611$ , *A. muciniphila*; ascending colon, UC<controls,  $P=0.1311$ ). Furthermore, dot plots of the remaining data revealed apparent trends towards alterations in some of the other bacterial groups tested, but these also failed to reach statistical significance. These included a decrease in *Bacteroides* and *Lactobacillaceae* in UC, and an increase in *Roseburia* in UC (Fig. 14) (Table 7).



**Figure 14|** Dot plots showing abundances of bacterial groups measured by qPCR

The abundance of specific bacterial groups in control and UC patient microbiotas was determined by qPCR relative to the universal 16S gene. Analysis of each bacterial group was performed in triplicate in each patient sample. A log transformed average of the abundance of bacterial groups in each patient sample is plotted. The mean and standard error of the values in each patient group is shown. Statistically significant differences between patient groups are indicated. None of the calculated *P* values fell below the false discovery threshold.

Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending



**Figure 14 continued | Dot plots showing abundances of bacterial groups measured by qPCR**

The abundance of specific bacterial groups in control and UC patient microbiotas was determined by qPCR relative to the universal 16S gene. Analysis of each bacterial group was performed in triplicate in each patient sample. A log transformed average of the abundance of bacterial groups in each patient sample is plotted. The mean and standard error of the values in each patient group is shown. Statistically significant differences between patient groups are indicated. None of the calculated P values fell below the false discovery threshold. Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending

	Wilcoxon matched pairs signed rank test			Mann-Whitney test		
	CS v CA	UCS v UCA	CA v UCA	CS v UCS	CA v UCA	CA v UCA
<i>Bifidobacterium</i>	0.5703	0.8129	0.8293	0.7876	0.8293	0.8293
<i>Roseburia</i>	0.1289	0.1298	0.1181	0.2811	0.1181	0.1181
<i>Faecalibacterium</i>	0.4961	0.9246	<b>0.0406</b>	0.1611	<b>0.0406</b>	<b>0.0406</b>
Lactobacillaceae	0.1289	0.5700	0.2357	0.1313	0.2357	0.2357
Bacteroides	0.2500	0.7049	0.1611	0.0592	0.1611	0.1611
Lachnospiraceae	0.3594	<b>0.0468</b>	0.7060	0.8716	0.7060	0.7060
<i>A. muciniphila</i>	0.5703	0.6291	0.1311	<b>0.0310</b>	0.1311	0.1311
<i>R. gnavus</i>	0.0547	0.2977	0.1779	0.3320	0.1779	0.1779
<i>B. fragilis</i>	0.6523	0.2012	0.1608	0.3595	0.1608	0.1608
Margalef index	0.6236	0.3273	0.8607	0.5595	0.8607	0.8607
Pielou evenness	0.1953	0.6026	0.3080	0.2104	0.3080	0.3080
Shannon index	0.9453	0.1183	0.1374	0.0665	0.1374	0.1374
Simpson index	0.5469	0.0720	0.3080	0.0969	0.3080	0.3080

Table 7 | P values for statistical tests carried out with different microbial parameters measured

*P* values are shown for the comparison of sigmoid vs. ascending and control vs. UC patients. Statistical tests were carried out to assess for significant differences in the abundance of 9 specific bacterial groups quantified in human samples by qPCR and in the diversity indices measured by 16S sequencing. *P* values  $\leq 0.05$  are considered significant (highlighted in bold).

Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending

### 3.4 Discussion

16S sequencing provided an insight into the overall composition of the gut microbiota associated with mucus in the colon. This identified a family level composition with similarities to that identified in previous studies, with the greatest abundance of sequences belonging to the Lachnospiraceae family, followed by Ruminococcaceae and Bacteroidaceae (Goodrich et al., 2014; Gosalbes et al., 2011). The overarching theme notable within the 16S sequencing dataset, which was also observed by qPCR, was the spatial conservation of the microbiota in the two GI tract regions analysed, the sigmoid and ascending colon. A conserved spatial structure has been demonstrated in a recent study that sampled the luminal and mucosal microbiota from caecum, transverse colon, descending colon and rectum (Lavelle et al., 2015). In agreement with our findings, no evidence could be demonstrated of longitudinal variability in the microbiota. Moreover, the conserved spatial structure could still be observed in areas of regional inflammation (Lavelle et al. 2015). The regional stability but inter-individual variation of the microbiota underscores the importance of studying changes in the microbial composition during disease in the context of an individual's signature microbiota.

There were no differences in the diversity, richness or evenness of the microbiota between controls and UC patients. A reduction in diversity of the gut microbiota in IBD was previously reported, however the majority of these studies have been carried out using faecal or biopsy samples (Nemoto et al., 2012; Nishikawa et al., 2009; Ott et al., 2004). The utilisation of endoscopic mucosal lavage to sample the microbiota may result in the acquisition of a different microbial subpopulation than that obtained through biopsy/faecal sampling. At the mucosal surface, microbes are in closer proximity to the epithelial surface and are likely to be more directly affected by host pressures, whilst luminal bacteria are likely to be shaped by environmental pressures. An example of this is the differential effects of innate immune effectors in the gut. Whilst human  $\alpha$ -defensin-5 plays a role in shaping the overall community structure, the antibacterial lectin RegIII only controls bacteria at the mucosal surface and does not penetrate into the lumen (Salzman et al., 2010; Vaishnava et al., 2011). In the luminal compartment of the gut, the nutrient availability may be in constant flux due to dietary variability, whilst at the mucosal surface there is a more stable supply of nutrients providing preferential niches for mucin-degrading organisms (Donaldson et al., 2016). Indeed, using laser capture of the mucus gel, it has recently been shown that the abundance of some bacterial families is different between the luminal and mucosal microbiota (Lavelle et al., 2015). Since the mucosal microbiota accommodate a separate niche to faecal or biopsy

communities, it is possible that this different subpopulation of organisms does not undergo the same alterations as reported from studies performed on faecal samples.

However, in a recent study by Tong and coll., endoscopic lavage sampling was used, and 16S sequencing revealed decreases in diversity in both UC and CD, although this observation was more significant in CD (Tong et al., 2013). A number of variances in study design may explain the differences between this study and our own observations. Whilst Tong and coll. analysed a similar number of UC patients to us, their cohort included a larger number of control patients. Differences in cohort sizes may influence the ability to observe subtle changes in microbiota composition and in part may explain the differences with our findings. Furthermore, Tong and coll. analysed samples from non-inflamed regions of the colon from IBD patients, whilst samples from this study were derived from both non-inflamed and active regions. A lack of consensus over alterations in diversity in UC is apparent, since some studies have found decreases in gut microbiota diversity in UC during relapse and remission, regardless of disease activity (Nemoto et al., 2012), whilst others have observed this decrease solely in active disease (Nishikawa et al., 2009; Ott et al., 2004). In line with our findings, a more recent study showed no differences in the diversity in the microbiota of *de novo* paediatric UC patients (Hansen et al., 2012b).

In addition, the approach used for 16S sequencing, including the choice of region to target and data processing/bioinformatics differed between our study and Tong and coll. (2013). We chose to target the V2 hypervariable region, whilst Tong and coll. used the V4 region. The region chosen may result in bias when investigating the diversity of a bacterial population (Nguyen et al., 2016). However, there is a lack of agreement over the best hypervariable region to target, since alternative methods result in differently favoured regions in terms of classification efficiency (Claesson et al., 2010). Furthermore, the algorithms used by Tong and coll. for taxonomy assignment differed to our approach. Whilst Tong and coll. used Greengenes database for taxonomic assignment, we used the SSU Ref release of the SILVA database (SSU Ref version 119), which contains only high quality and nearly full length sequences (Pruesse et al., 2007). In addition, whilst Tong and coll. used a version of Greengenes from February 2011, the SILVA database used in our study was dated July 2014. Therefore as well as performing alignment with high quality sequences, we used a more up-to-date database for classification, which may have contributed to the discrepancies between the diversity measurements of the two studies.

Obtaining reliable taxonomic classification of organisms *via* high-throughput 16S sequencing, which yields shorter reads than traditional sequencing methods which are more time-consuming and expensive, is challenging (Mizrahi-Man et al., 2013). As a result of this, it is inherently difficult to monitor changes in the abundance of bacteria at the species and strain level by 16S sequencing, which are likely to be the most relevant groups in terms of mucin degradation and therefore, in the context of this study. Indeed, mucin degradation has been shown to be strain-dependent with *R. gnavus* strains ATCC 29149 and 35913 but not *R. gnavus* E1 of capable utilising mucin glycans (Crosthair et al., 2013; Crosthair et al., 2016). Using 16S sequencing, we were unable to detect sequences corresponding to the genera to which the most prevalent mucolytic bacteria, *A. muciniphila* and mucin degrading strains of *R. gnavus* belong. For this reason, a more targeted approach was used to quantify specific mucin degrading species/strains.

Accordingly, using targeted qPCR we were able to detect differences in the abundance of specific mucin degrading groups between controls and UC patients. One notable trend was the depletion of *A. muciniphila*, which was significantly different between the sigmoid colon of control and UC patients. Alterations in *A. muciniphila* have also been documented in a previous study investigating these changes in mucosal biopsies (Png et al. 2010). Interestingly, in the same study, the abundance of *R. gnavus*, was observed to be increased in IBD (Png et al., 2010). Although we did not detect a significant increase in the abundance of *R. gnavus*, there was no notable decrease. These findings suggest that the mucosal environment in IBD patients may be altered, resulting in unfavourable conditions for *A. muciniphila*, whilst *R. gnavus* remains unaffected.

In addition, we observed an increase in the abundance of the Faecalibacterium genus in our study, which was significantly different between the ascending colon of control and UC patients. This finding is particularly interesting as *F. prausnitzii*, the only cultured representative of the Faecalibacterium genus is proposed to have anti-inflammatory properties due to its production of butyrate (Morgan et al., 2012; Sokol et al., 2008). Butyrate exerts a number of effects, including the stimulation of mucin synthesis, and down regulation of pro-inflammatory cytokines (Hatayama et al., 2007; Segain et al., 2000). Indeed, a number of studies have observed a depletion of *F. prausnitzii* belonging to the Faecalibacterium genus in IBD (Joossens et al., 2011; Machiels et al., 2014; Sokol et al., 2009; Tong et al., 2013). However, there are controversial findings of *F. prausnitzii* abundance in IBD. In a recent study, no differences were observed in the abundance of *F. prausnitzii* in CD patients with different disease activity (Schaffler et al., 2016). Furthermore, in treatment naïve cases of CD

and CD involving the colon, like our study, *F. prausnitzii* has been shown to be over-represented (Hansen et al., 2012b; Willing et al., 2010a). Growth of *F. prausnitzii* is proposed to be stimulated by acetate (Duncan et al., 2002b). In germ-free rats, *F. prausnitzii* was unable to colonise in the absence of the acetate producer *B. thetaiotaomicron* (Wrzosek et al., 2013). Recently, *R. gnavus* was also shown to produce acetate in similar quantities to *B. thetaiotaomicron* in the caecal contents of mono-colonised mice (Hoffmann et al., 2016). The increase of *F. prausnitzii* in our study may therefore be explained by increases in *R. gnavus* and/or other acetate producing bacteria in IBD. It is also possible that *F. prausnitzii* is subject to transient changes in abundance in IBD, which could be dependent on changes in disease status and/or the host mucosal environment. In our study, the abundance of Faecalibacterium appeared to be subject to much more inter-individual variation in UC patients than in controls, supporting this idea. Conflicting results may therefore arise due to the study design, for example, the point during relapse at which samples are taken from the patient. It is evident from discrepancies between studies that the role for *F. prausnitzii* and the Faecalibacterium genus in IBD is more complex than first thought, and that hypotheses such as these must be investigated more thoroughly.

Further trends towards alterations of specific groups were also apparent, although failed to reach significance, probably due to the low number of samples in our study. We observed a possible decrease in the abundance of the Bacteroides genus. This is in accordance with previous studies showing a 50-fold decrease in the Bacteroidetes phylum in colonic biopsies (Frank et al., 2007), as well as the absence of *B. vulgatus*, *B. ovatus*, *B. uniformis*, and Parabacteroides sp. in the faecal microbiota of UC patients from healthy controls (Noor et al., 2010). Controversially, colonic biopsies from UC patients contained a higher abundance of Bacteroidetes phyla when compared with those from non-IBD patients (Lucke et al., 2006). Whilst in our study there was a possible decrease in the abundance of the Lactobacillaceae family, both reductions and increases in Lactobacillus species have been demonstrated in colonic biopsies and faecal samples from UC and CD patients (Bullock et al., 2004; Ott et al., 2004; Wang et al., 2014). Furthermore, there appeared to be a trend towards an increase in the Roseburia genus. Whilst there was no alteration of Roseburia in the faecal samples of patients with UC or colonic CD, Roseburia was reduced in those with ileal CD (Willing et al., 2010a). Another study found that *Roseburia hominis*, belonging to the Roseburia genus was depleted in faecal samples from UC patients (Machiels et al., 2014). Alike *F. prausnitzii*, *R. hominis* is a main butyrate producing bacteria from the Firmicutes phylum, and therefore is likely to exert similar immuno-modulatory effects in the GI tract

(Machiels et al., 2014). The conflicting findings in our study may occur for similar reasons to those mentioned above.

In order to further understand the role of bacterial groups such as *F. prausnitzii*, *Bacteroides* and *Roseburia* in IBD, longitudinal studies of individual IBD patients would be required. This would allow more robust control of variation caused by inter-individual differences, such as diet, antibiotic treatment and medication. However, it is also worth considering that other experimental variables could influence the results, such as bowel preparation, sampling methods (faecal samples, biopsies or lavage), sample handling and processing (e.g. DNA extraction), and analysis method. Introduction of cross-study standardisation would be beneficial in order to reduce variation caused by factors such as these.

Together, these data highlight the lack of consistent information regarding a 'signature' microbiota of IBD. Most of the changes associated with IBD appear to only occur sporadically within individual patients (Tong et al., 2013). The lack of consistency between studies reflects the need for caution over whether we can classify certain bacterial groups as 'biomarkers of health' or 'biomarkers of disease'. It is likely that the observed changes in the microbiota in a particular study is affected by a number of factors, including the patient-specific disease aetiology, and time during disease progression at which the microbiota is sampled.

## Chapter 4) Characterisation of mucin glycosylation from human colonic lavages

### 4.1 Introduction and objectives

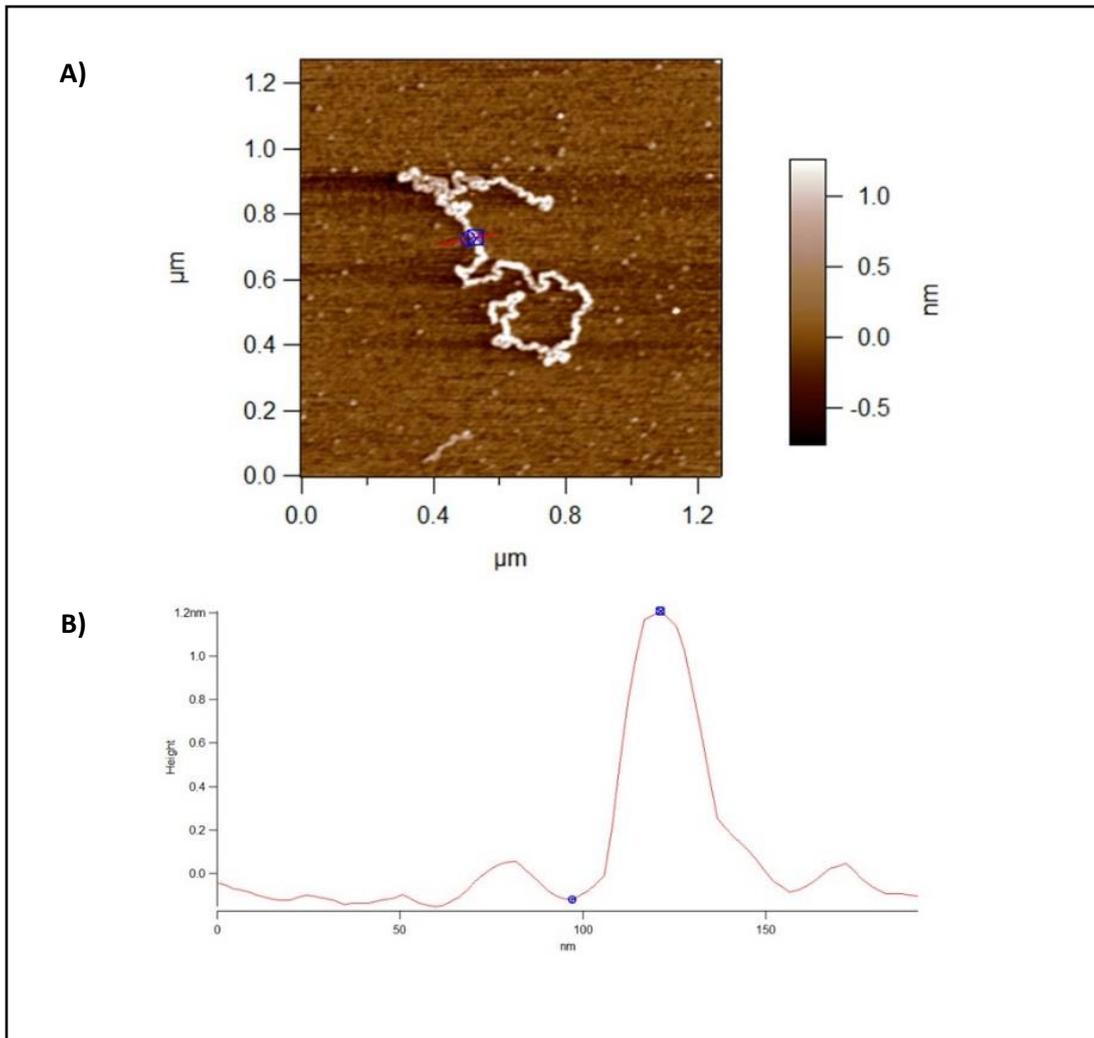
Mucins are a crucial part of the intestinal barrier, not only because of their physical function but also due to the relationship they mediate between the host and the microbiota. The mucin glycan profile is determined by the expression of glycosyltransferases in different tissues and can be modulated by the local environment. The glycans act as nutrients and attachment sites for commensal organisms and may therefore shape the microbiota composition to the advantage of the host. In disease, alterations in the mucin glycan landscape could be driven by changes in glycosyltransferase expression and result in the dysbiosis of the microbiota. As mentioned previously (See section 3.1) the relationship between mucin glycosylation, glycosyltransferase expression and the mucus-associated microbiota in IBD is unclear. Recently, it was shown that alterations in core glycosyltransferase expression resulted in a compromised mucus layer and subtle alterations in the microbiota composition in a murine model (Bergstrom et al., 2016; Sommer et al., 2014). However, there are very few studies investigating the relationship between these in humans. Using mucosal biopsies, alterations of mucin glycosylation have been demonstrated in IBD patients, including decreases in more complex glycans and an increase in shorter and sialylated glycans (Larsson et al., 2011). Although it was suggested that this could be due to alterations in the activity of the sialyltransferase ST6GalNAc-I, the expression of this gene was not analysed. Furthermore the microbiota composition was not investigated (Larsson et al., 2011). In a separate study the influence of the Fut2 genotype on the GI microbiota was investigated, showing substantial influence on the composition and diversity of the microbiota. However in this study the glycosylation profile of GI tract mucins was not investigated (Rausch et al., 2011).

One factor contributing to the lack of information about all three components is the difficulty in obtaining mucus, associated microbiota and tissue from the same sample site in humans. Furthermore, the mucus from biopsies most likely originates from the inner mucus layer which is not in direct contact with the microbiota. Here, we used mucosal lavage allowing us to sample both mucus and the associated microbiota, in addition to biopsies to investigate the expression of glycosyltransferases in the same patient, and at different colonic sites.

## 4.2 Characterisation of mucus from human mucosal lavage samples

Mucins were purified from 73 human mucosal lavage samples, corresponding to the sigmoid and/or ascending colon of 40 patients. Purification was carried out following an adaptation of Larsson and coll. (Larsson et al., 2009). Briefly, soluble and insoluble fractions were obtained by extraction in GuHCl. MUC2 of the inner firm mucus layer is densely packed and therefore insoluble in GuHCl, whilst the outer loose mucus layer has expanded to approximately four times the volume and can be readily solubilized (Johansson et al., 2011a). This method therefore enables mucus originating from each layer to be extracted. Purification was completed by reduction and alkylation of mucin disulphide bonds, allowing resolution of these large molecules.

In order to verify the quality of the purified mucins, samples were first imaged using AFM. In most samples, chains with contour lengths ranging from tens of nanometers to several microns could be observed (Fig. 15). The chains were particularly notable in the GuHCl soluble fractions. The presence of mucin chains of the expected size and conformation, including heights of  $\sim 0.5\text{-}1.5$  nm (Fig. 15) was confirmed following DNase treatment.



**Figure 15| Example of a mucin chain purified from human mucosal lavage and imaged by AFM**

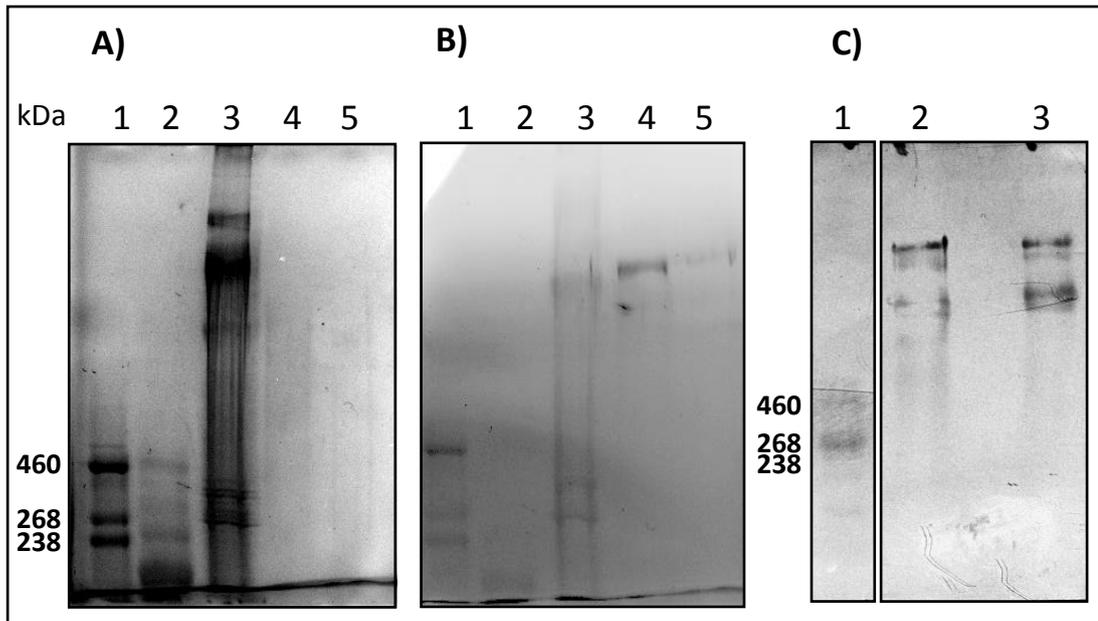
*A) Purified mucin sample imaged following overnight DNase treatment*

*B) Graph depicting typical mucin height*

*Images were acquired in alternating current (AC) mode. The height of material detected on the AFM slide is depicted using pixelation brightness, according to the scale bar shown. B) represents a measurement of mucin height at a cross section of the mucin chain (as shown in A))*

In order to further validate the extraction method and characterise the mucins from lavage, we performed electrophoretic separation and western blot of the purified samples. Due to the high molecular weight of soluble and insoluble mucins (monomeric GuHCl insoluble MUC2 ~2.5 MDa), and the high proportion of carbohydrates (up to 80% of the weight), analysis of mucins by electrophoresis requires the use of a specific type of AgPAGE gradient gel, or 'composite' gel (Fig. 16). This technique allows analytical separation and resolution of glycoforms of very high molecular weight that can be monitored by lectin, immunochemical or histochemical staining by western blot (Schulz et al., 2002). Mucins separated by AgPAGE were first visualised using GelCode blue protein staining followed by a specific glycoprotein stain (Fig. 16). Using these two methods, only faint staining could be observed with GelCode blue. In contrast, more intense signals could be obtained using the glycoprotein staining kit. Furthermore, mucin bands could be visualised upon western blotting and alcian blue staining (Fig. 16).

In both soluble and insoluble fractions, the apparent molecular mass of all bands exceeded 500 kDa and were of a double to quadruple nature. The GuHCl soluble fractions, corresponding to the outer mucus layer often stained most intensely, whereas the insoluble fractions, corresponding to the inner mucus layer appeared almost devoid of glycoprotein (Larsson et al., 2009).



**Figure 16 | Example of a composite AgPAGE of mucins purified from human mucosal lavage**

*A) Gel stained using a colloidal coomassie based dye, GelCode (Thermo Fisher Scientific, Waltham, MA); Lane 1, HiMark pre-stained protein ladder (Thermo Fisher Scientific, Waltham, MA), Lane 2, Fetuin (2 mg/ml), Lane 3, Soluble mucin fraction from LS174T cell line, Lane 4, Soluble mucin fraction from a sigmoid human mucosal lavage sample, Lane 5, Insoluble mucin fraction from a sigmoid human mucosal lavage sample.*

*B) Gel stained using a Periodic acid-schiff based Glycoprotein staining kit (Thermo Fisher Scientific, Waltham, MA); Lanes as in A).*

*C) Western blot stained using Alcian blue; Lane 1, HiMark pre-stained protein ladder, Lane 2, Soluble mucin fraction from an ascending human mucosal lavage sample, Lane 3, Soluble mucin fraction from an ascending human mucosal lavage sample*

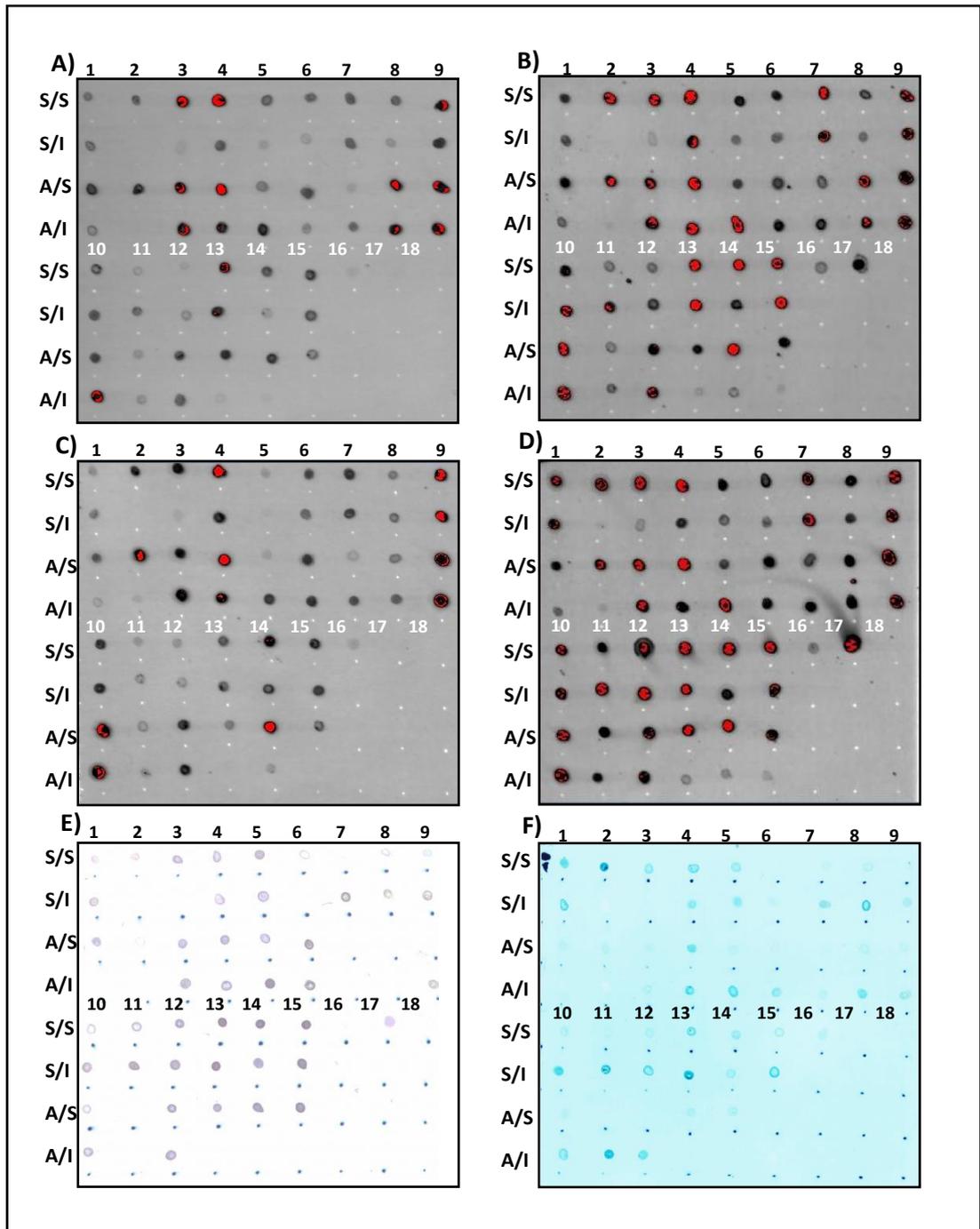
## 4.3 Profiling of glycans from human mucins

### 4.3.1 Lectin probing of mucins by dot blot and force spectroscopy

Profiling of mucin glycans was first performed by dot blotting purified mucins onto PVDF membranes, and probing with a range of different fluorescently labelled and biotinylated lectins with specific affinities for glycan structures. The soluble and insoluble fractions of 30 samples, corresponding to the sigmoid and ascending colon of 15 patients were probed with 5 different lectins (Table 5), and alcian blue stained. Using this method, it was observed that lectin binding to  $\beta$ -gal (FITC-RCA) and GlcNAc (FITC-WGA) residues was fairly consistent across patients (Fig. 17B and D), whilst binding to Fuc (FITC-UEA),  $\alpha$ 2-3 linked sialic acid (Biotinylated MALII) and T-antigen (FITC-PNA) was much more heterogeneous between individuals (Fig. 17A, C and E). Furthermore, staining of acidic mucin glycans with alcian blue was also highly variable between individuals (Fig. 17F). Mucins from the soluble and insoluble fractions of the same colonic region in the same patients appeared to show no trend in glycosylation. Whilst in some patients, staining of both fractions appeared almost identical with the same lectin, in others staining intensity was very different.

For further characterisation of mucin glycans, analysis was only carried out on the soluble fractions, due to the apparent lack of mucin isolated in the GuHCl insoluble fractions. The purified GuHCl soluble mucins from a subset of samples were probed using lectin functionalised force spectroscopy tips. A total of 14 mucin samples corresponding to the sigmoid and ascending colon of 4 UC patients and 3 controls were probed using the lectins, RCA, UEA and SNA (specificity for  $\alpha$ 2-6 linked sialic acid). In total, 1024 unique force spectra were obtained for each sample/lectin probing experiment. For each specific binding event within a given force spectrum, data were obtained about the force required to break the lectin-glycan interaction (adhesion affinity), and when two or more specific binding events were observed within one spectrum, about the distance between these glycans along the mucin chain (adhesion event separation). These data were collected from all 1024 spectra per probing experiment and frequency distribution graphs plotted giving information about the properties of glycans within that mucin sample. The distributions for the majority of samples in both sigmoid and ascending colon had remarkably similar profiles when probed with RCA lectin, in that each profile had a similar modal value and reasonably tight distribution. This was the case for both the adhesion event separation (Fig. 18 A and D) and adhesion affinity (Fig. 19 A and D). In contrast, the profiles for adhesion affinity (Fig. 19 B, C, E and F) and adhesion event separation (Fig. 18 B, C, E and F) were much more variable when

probing with SNA and UEA lectins. There were no apparent patterns in adhesion affinity or adhesion event separation based on disease state, control or UC.



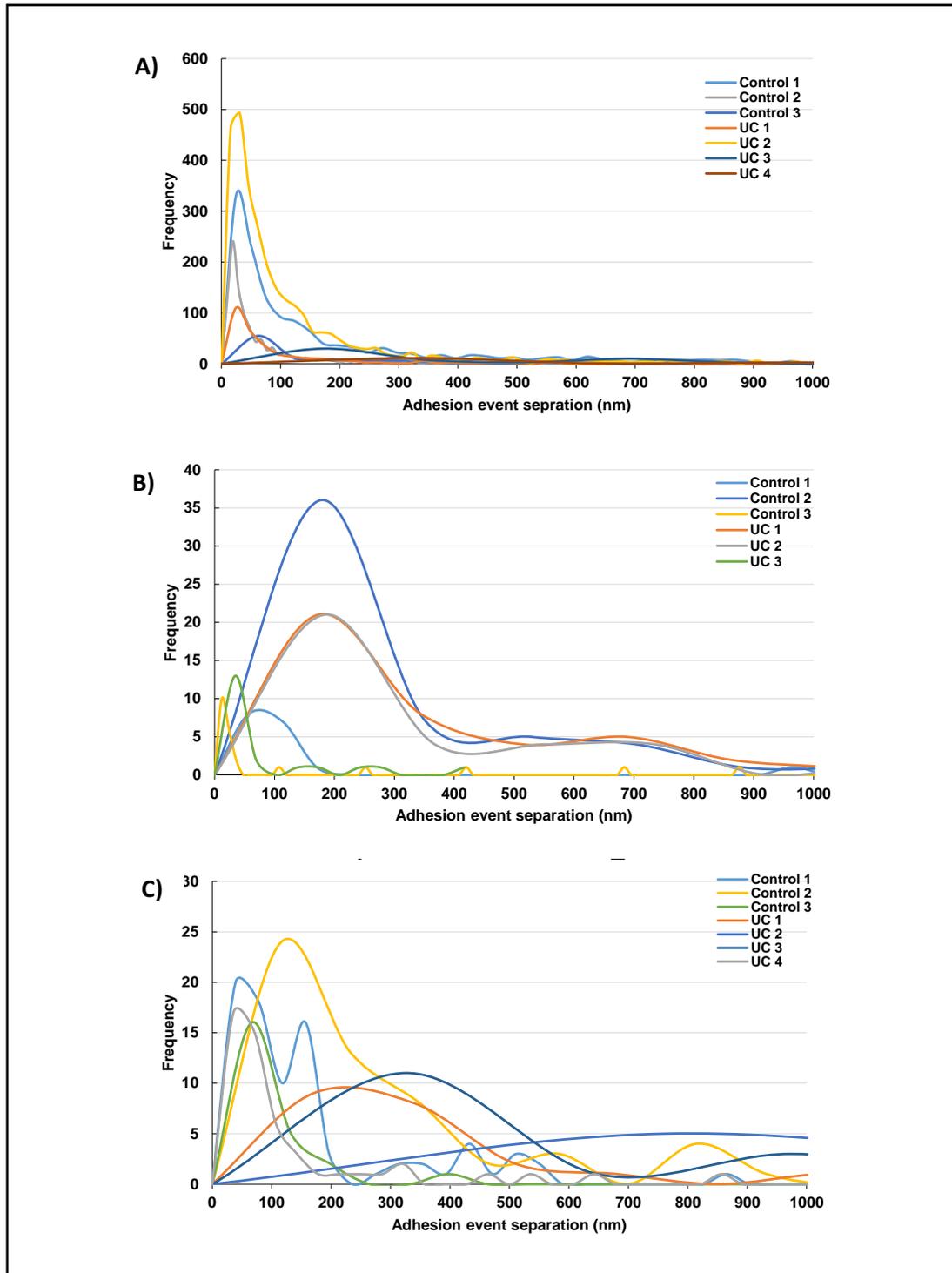
**Figure 17 | Dot blots of mucins purified from human mucosal lavage samples**

Mucins purified from mucosal lavage were normalised to a concentration of 2.5 mg/ml prior to blotting. Blots were stained using a range of FITC- labelled and biotinylated lectins and dyes as follows:

A) FITC-PNA, B) FITC-RCA, C) FITC-UEA, D) FITC-WGA, E) Biotinylated MALII, F) Alcian blue

Patient samples as follows; 1, 13TB0634, 2, 13TB0667, 3, 13TB0648, 4, 14TB0005, 5, 14TB0024, 6, 14TB0060, 7, 14TB0055, 8, 14TB0025, 9, 14TB0047, 10, 14TB0061, 11, 13TB0617, 12, 13TB0614, 13, 13TB0605, 14, 13TB0583, 15, 13TB0601, 16, LS174T soluble fraction, 17, 4 mg/ml Fetuin, 18, H<sub>2</sub>O

Key; S/S, Sigmoid soluble fractions, S/I, Sigmoid insoluble fractions, A/S, Ascending soluble, A/I, Ascending insoluble. Red coverage indicates a saturated signal.



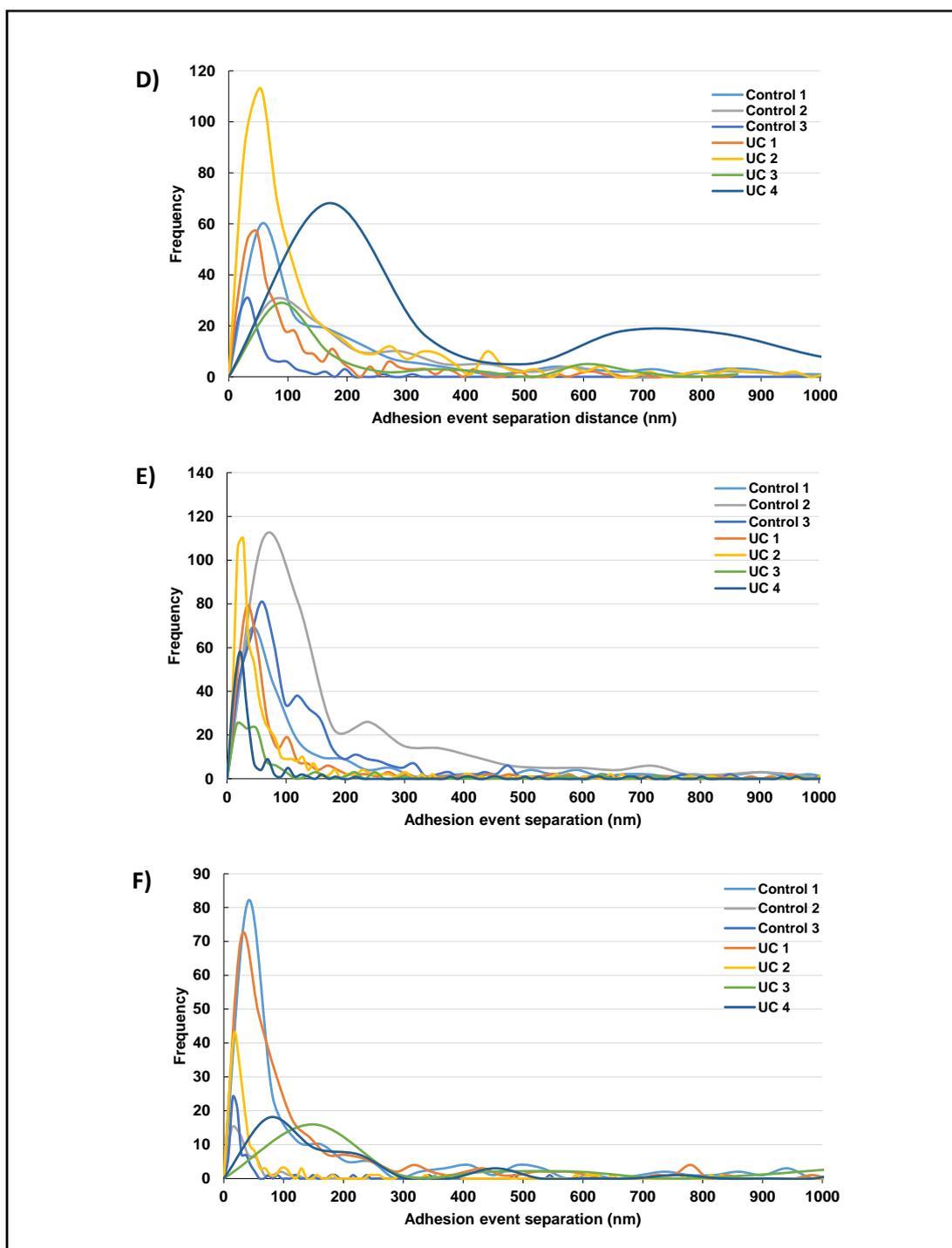
**Figure 18|** Graphs showing the distribution of binding event separation distances in soluble mucins

A) Force-volume experiments against sigmoid colonic mucins using RCA lectin

B) Force-volume experiments against sigmoid colonic mucins using SNA lectin

C) Force-volume experiments against sigmoid colonic mucins using UEA lectin

Measurements were made of the separation between the adhesion peaks in each force spectrum. This was applied to a whole set of curves for a given force-volume experiment (n=1024) and the frequency of adhesion event separation distances plotted in distribution curves.



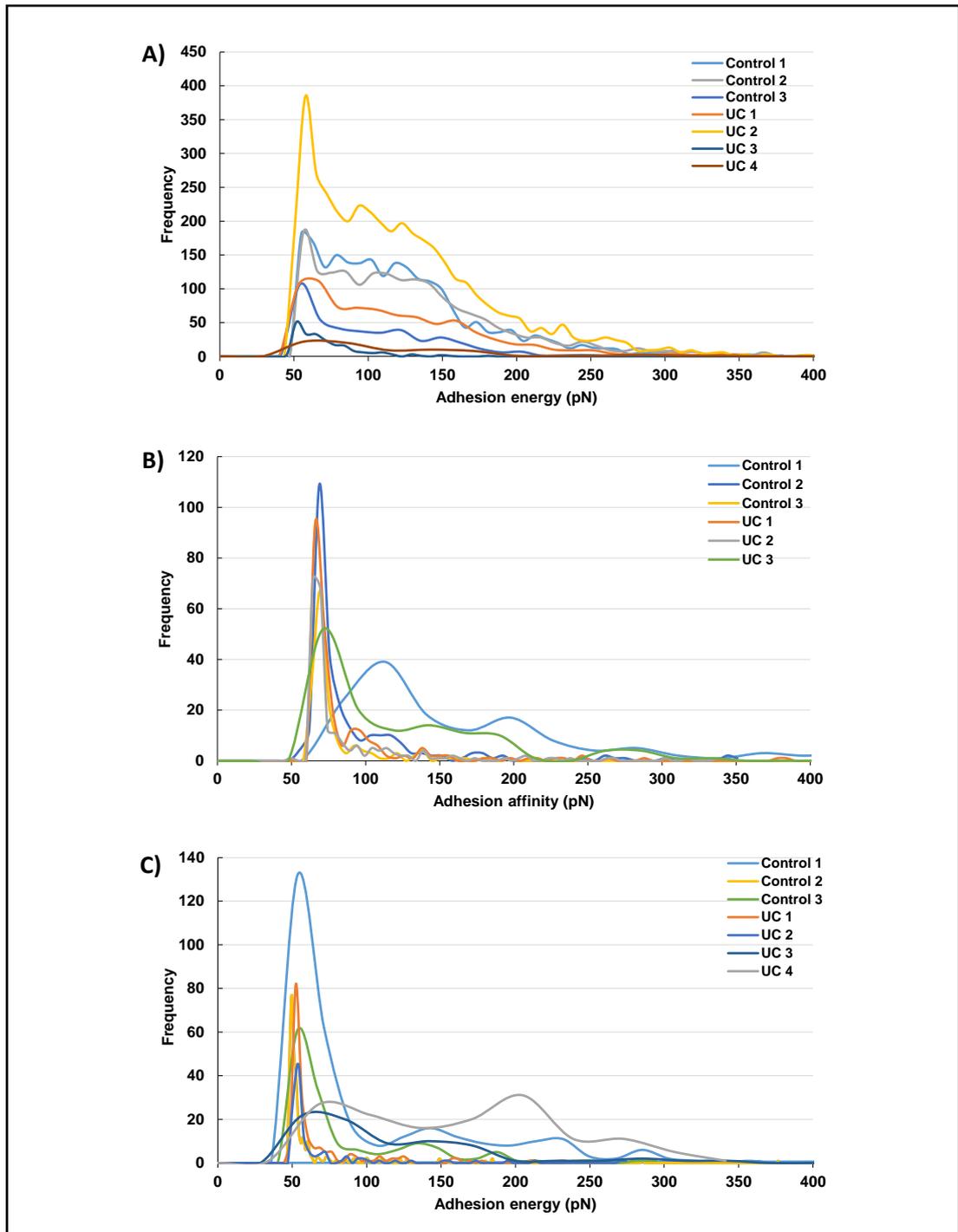
**Figure 18 continued | Graphs showing the distribution of binding event separation distances in soluble mucins**

*D) Force-volume experiments against ascending colonic mucins using RCA lectin*

*E) Force-volume experiments against ascending colonic mucins using SNA lectin*

*F) Force-volume experiments against ascending colonic mucins using UEA lectin*

*Measurements were made of the separation between the adhesion peaks in each force spectrum. This was applied to a whole set of curves for a given force-volume experiment (n=1024) and the frequency of adhesion event separation distances plotted in distribution curves.*



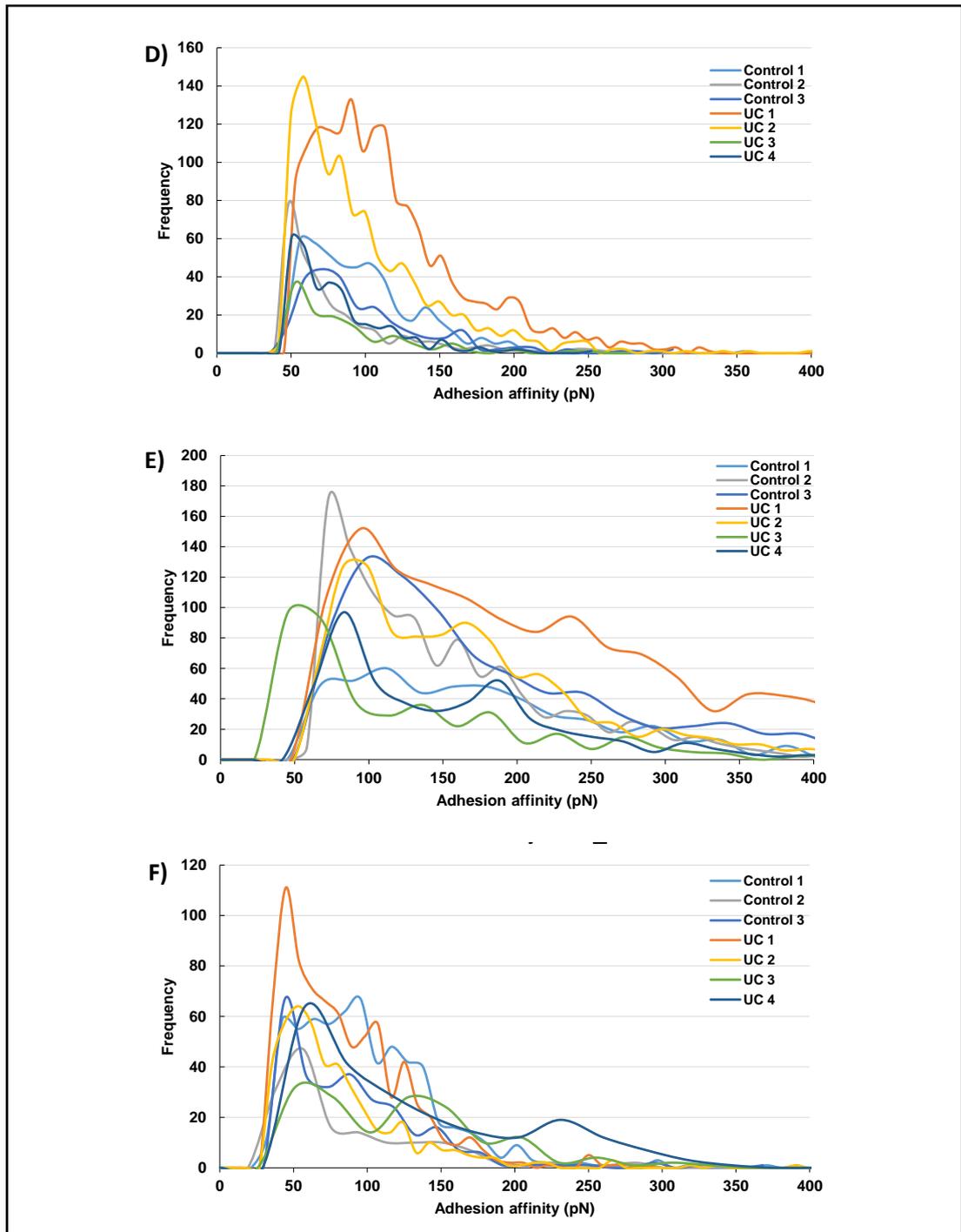
**Figure 19|** Graphs showing the distribution of adhesion affinity measurements in soluble mucins

A) Force-volume experiments against sigmoid colonic mucins using RCA lectin

B) Force-volume experiments against sigmoid colonic mucins using SNA lectin

C) Force-volume experiments against sigmoid colonic mucins using UEA lectin

For each specific binding event within a given force spectra, measurements of the force required to break the lectin-glycan interaction were obtained. This was applied to a whole set of curves for a given force-volume experiment (n=1024) and the frequency of adhesion affinity measurements plotted in distribution curves



**Figure 19 continued | Graphs showing the distribution of adhesion affinity measurements in soluble mucins**

*D) Force-volume experiments against ascending colonic mucins using RCA lectin*

*E) Force-volume experiments against ascending colonic mucins using SNA lectin*

*F) Force-volume experiments against ascending colonic mucins using UEA lectin*

*For each specific binding event within a given force spectra, measurements of the force required to break the lectin-glycan interaction were obtained. This was applied to a whole set of curves for a given force-volume experiment (n=1024) and the frequency of adhesion affinity measurements plotted in distribution curves*

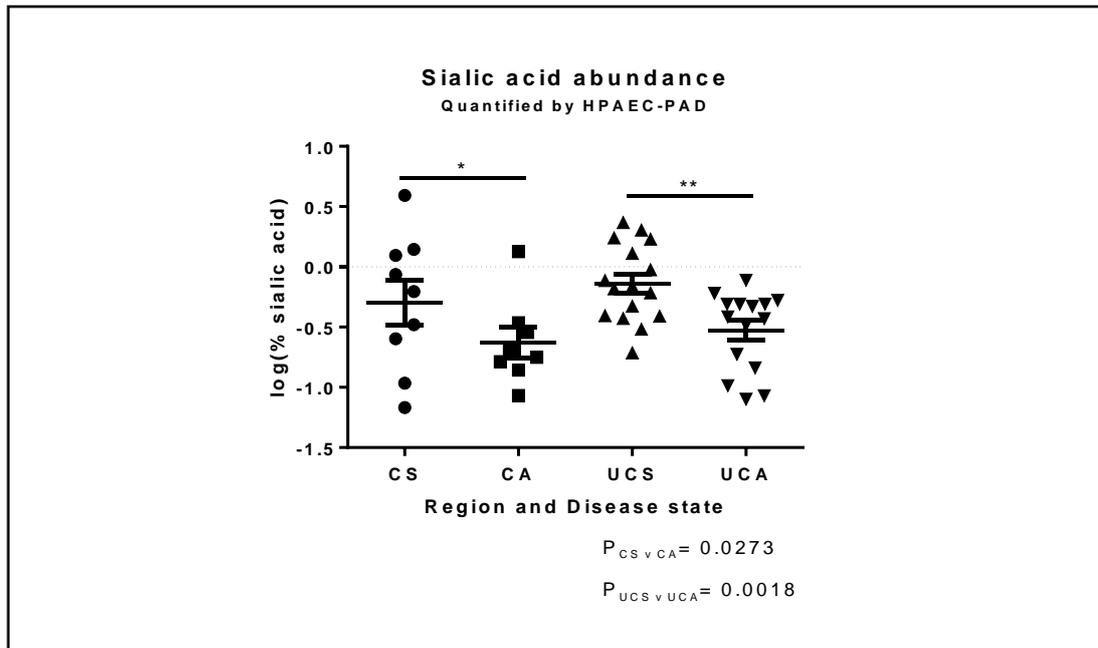
#### 4.3.2 Quantification of mucin sialylation by HPAEC-PAD

To analyse the level of sialic acid in mucins from lavage samples, high performance anionic exchange chromatography with pulsed amperometric detection (HPAEC-PAD) was performed following acid hydrolysis of GuHCl soluble fractions from lavage samples. Separation of sialic acid was achieved using a gradient method with 7-30% 1 M sodium acetate plus 100 mM NaOH as an eluent. Using this method, sialic acid eluted at around 5.7 min, whilst the internal standard, KDN eluted at around 6.9 min. An acid hydrolysed sample of fetuin, a protein control containing sialylated glycans confirmed the validity of the method, with a peak eluting in the expected region on the chromatogram. On the other hand, the spectra from the non-sialylated protein, BSA showed no peaks in the expected regions of the chromatogram.

A total of 68 soluble mucin samples were analysed by HPAEC-PAD from 37 patients. Across all HPAEC-PAD runs performed on samples, the mean abundance of sialic acid measured in the positive control, fetuin, was 3.68%. The abundance of sialic acid in samples ranged from 3.9% to 0.05% and was detected in all but 5 samples.

In order to make meaningful comparisons with the microbiota composition of the samples, data were selected as in section 3.2.2 (exclusion of patients with only one colonic region measured, and of CD patients). The resulting data set consisted of 20 samples from 10 control patients and 32 samples from 16 UC patients with both sigmoid and ascending samples measured (total of 52 samples). The mean abundances of sialic acid in control ascending, control sigmoid, UC ascending and UC sigmoid groups were 0.27%, 0.54%, 0.35% and 0.93% respectively. A wilcoxon matched pairs signed rank test identified that in both control and UC patients the abundance of sialic acid measured by HPAEC-PAD was significantly higher in the sigmoid colon when compared with ascending (controls,  $P=0.0273$ , UC,  $P=0.0018$ ) (Fig. 20) (Table 9). Both of these  $P$  values were lower than the threshold value set when controlling for the false discovery rate below  $q=0.05$ .

Next, a Mann-Whitney test was performed to determine if there were significant differences between the abundance of sialic acid in the sigmoid colon of control patients versus UC as well as in the ascending colon. The  $P$  values showed that there was no significant difference in the sialylation between the two groups (Fig. 20) (Table 9).



**Figure 20 | Abundance of sialic acid liberated from purified mucins, as determined in four patient groups by HPAEC-PAD**

*Sialic acid was liberated from mucins purified from human mucosal lavage using mild acid hydrolysis. HPAEC-PAD analysis of each sample was performed in triplicate. Each data point represents a log transformed average of technical replicates per sample. The mean and standard error of the values in each patient group is shown. Statistically significant differences between patient groups are indicated. Both of the statistically significant P values fell below the false discovery threshold.*

*Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending*

### 4.3.3 Glycomic profiling by mass spectrometry

To complement the approaches used so far assessing the distribution of terminal sugar moieties, a more detailed analysis of the mucin glycan structures was performed by mass spectrometry (MS). Glycans were liberated from mucins by  $\beta$ -elimination, purified using HLB cartridges and subjected to MALDI-TOF.

MS analysis was successfully performed on 39 samples from 21 patients; 5 controls, 11 UC patients and 4 CD patients. The number of samples which could be analysed was limited by the quantity of mucin that could be successfully purified from lavage. The relative abundance of each glycan structure identified in the MALDI-TOF spectra was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra (Fig. 21). As previously, data were sorted so that only patients with both sigmoid and ascending regions sampled were included in the statistical analysis (See section 3.2.2). Furthermore, CD patients were excluded from the analysis. This resulted in the inclusion of 10 samples from 5 control patients and 22 samples from 11 UC patients.

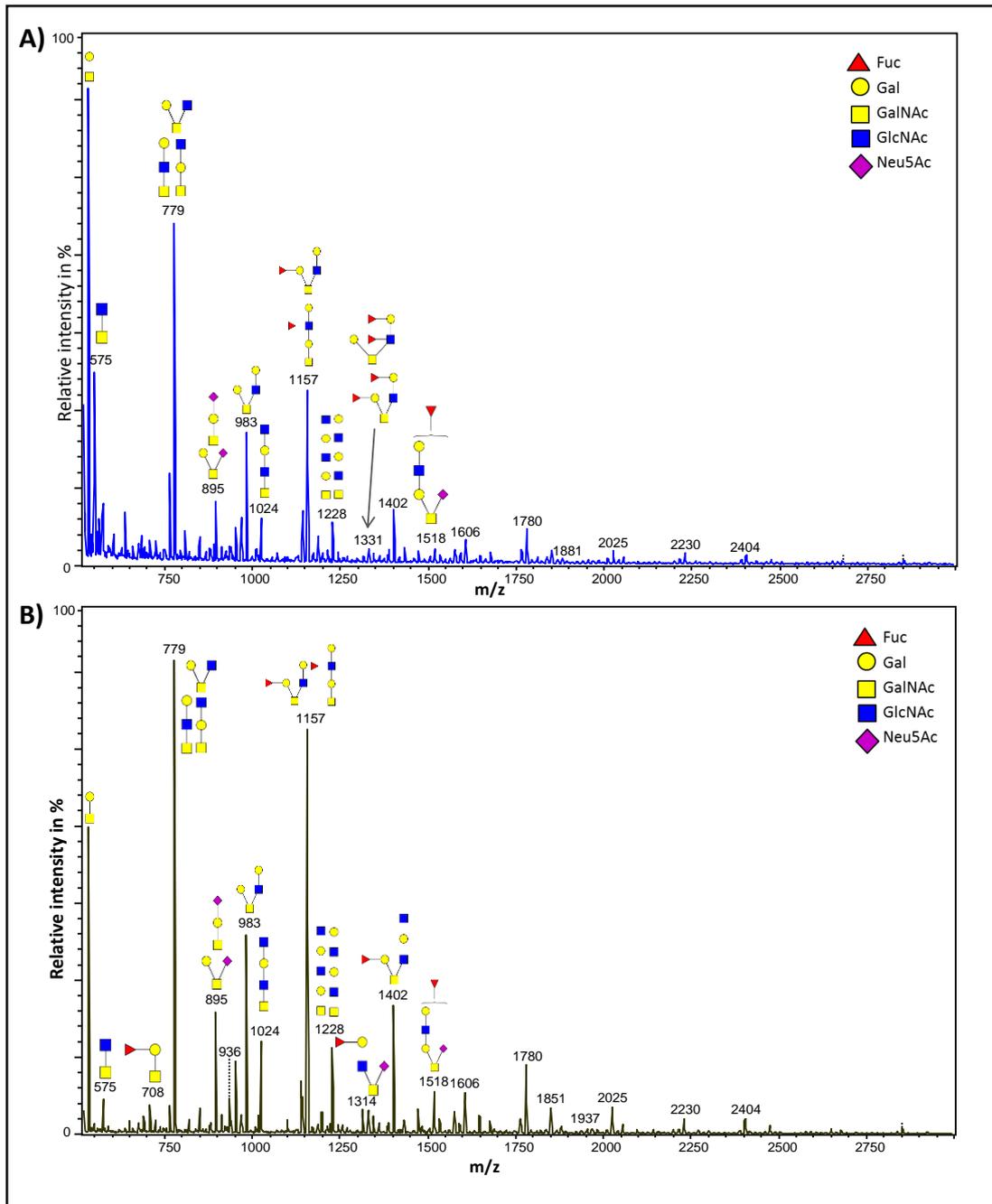
Overall, 110 unique glycan structures were detected across the different samples analysed (See appendix 3). Inter-individual heterogeneity was apparent in the abundance of individual glycan structures, with some structures undetected in a number of samples.

Relative quantification was performed to obtain the total percentage of sialylated, fucosylated and sulphated structures within each patient. Whilst a large amount of inter-individual variability was apparent, trends were notable in the data. The majority of glycan structures in all patient groups were fucosylated. The mean abundances of fucosylated structures in control ascending, control sigmoid, UC ascending and UC sigmoid groups was 68.93 %, 41.32 %, 49.99 % and 28.91 % respectively, revealing the presence of a proximal to distal decreasing gradient of Fuc in both controls and UC patients. A wilcoxon matched pairs signed rank test revealed that the difference in fucosylation was significant between the sigmoid and ascending colon of UC patients ( $P=0.0029$ ). This gradient was also apparent on an individual basis, with 14 of 16 patients displaying a decrease in Fuc from ascending to sigmoid colon. The mean abundances of fucosylated structures also indicated an overall trend for a reduction of fucosylation in UC when compared to controls (decrease of fucosylation by 18.94 % in ascending and 12.41 % in sigmoid colon of UC patients compared to controls). However, Mann-Whitney tests calculated that this difference was not significant in either the sigmoid ( $P=0.5096$ ) or ascending ( $P=0.1149$ ) colon. The mean abundances of

sialylated structures were 9.01 %, 14.04 %, 14.35 % and 15.13 % in control ascending, control sigmoid, UC ascending and UC sigmoid groups respectively. Whilst these means were in accordance with the increasing gradient of sialic acid observed by HPAEC-PAD (See section 4.3.2), this trend was much more subtle and was not determined to be significant in either control or UC patients by statistical analysis. On an individual basis, 10 out of the 16 patients displayed the increasing proximal to distal gradient of sialic acid. The mean abundances of sialylated structures also indicated an overall trend for an increase of sialylation in the ascending colon of UC patients compared to controls (increase of 5.34 %). Mann-Whitney tests calculated that this difference was not significant ( $P= 0.0893$ ), however upon removal of one UC patient with an exceptionally low level of sialylation in the ascending colon, the difference became significant ( $P= 0.0280$ ). Sulphated structures represented a much smaller proportion of the overall structures detected, with mean abundances of 0.39 %, 0.92 %, 1.07 % and 5.23 % in control ascending, control sigmoid, UC ascending and UC sigmoid groups respectively.

In order to perform statistical analyses, 24 of the most prevalent glycan structures across patients were selected. Of these 24 structures, the most common glycan epitopes were glycans  $m/z$  534 (TF antigen/Core 1), present in 96.9% (31/32) of samples,  $m/z$  779 (two potential isomers, one of which is core 2) and  $m/z$  983, both present in 100% (32/32) of samples (Table 8).

Interestingly in both the sigmoid and ascending colon, the mean abundance of the TF antigen, the shortest glycan detected by MS was higher in UC patients than in the sigmoid and ascending colon of controls (Control Sigmoid, 8.56 %, UC Sigmoid, 12.03 %, Control Ascending, 5.5 %, UC Ascending, 9.99 %) (Fig. 22) (Table 8). This was the case for a number of the shorter glycans detected, including glycan  $m/z$  691 (STn) and  $m/z$  779 (core 2 isomer when containing a GlcNAc). In contrast, the longest glycan out of the set of 24, glycan  $m/z$  2025 was more abundant in both the sigmoid and ascending colon of controls when compared to UC patients (Control Sigmoid, 1.32 %, UC Sigmoid, 0.51 %, Control Ascending, 1.95 %, UC Ascending, 1.17 %) (Fig. 22) (Table 8).

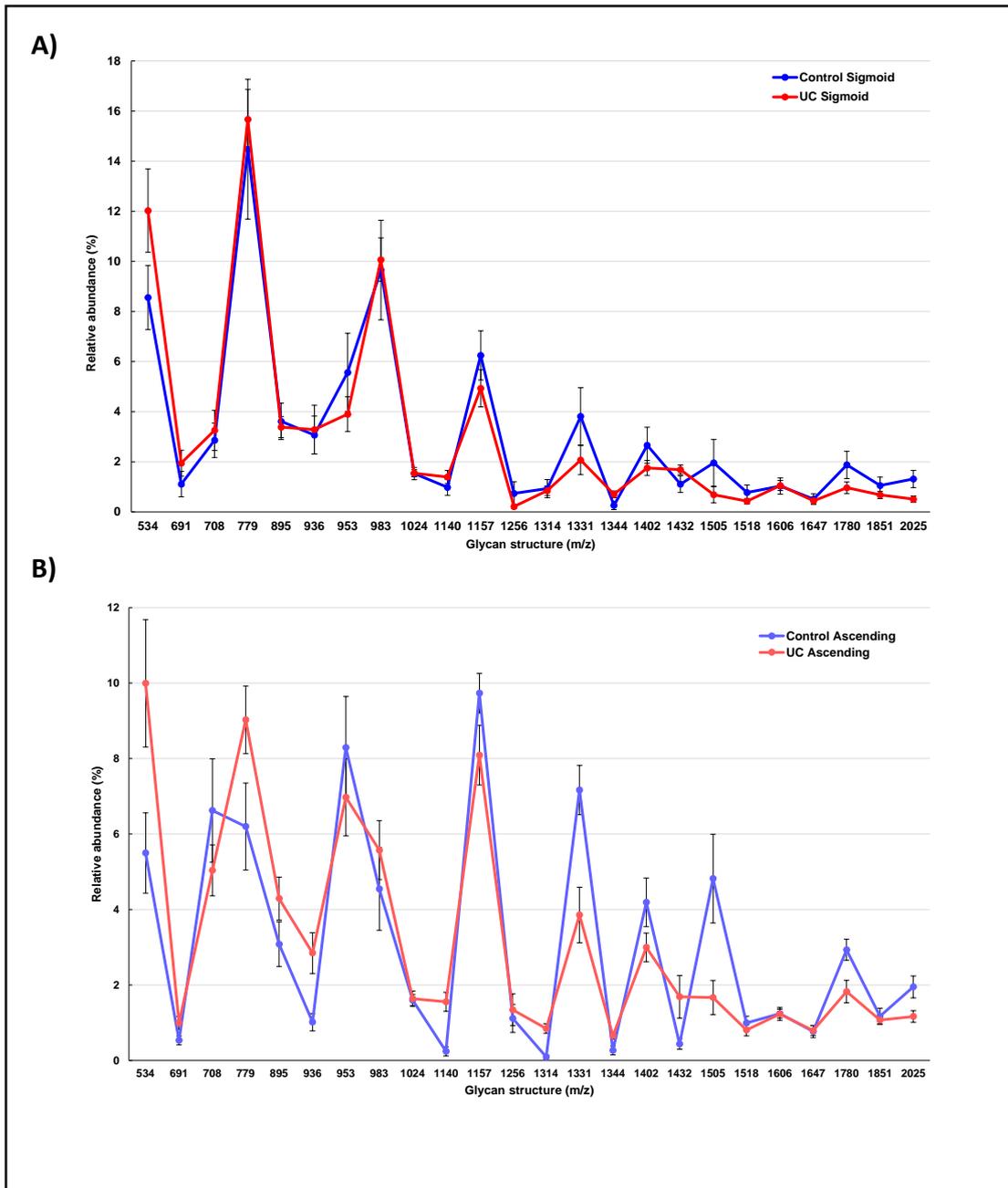


**Figure 21 | Example MALDI-TOF spectra of mucin glycans isolated from a human mucosal lavage sample from the ascending colon**

*A) Example spectra of mucin glycans purified from the sigmoid colon*

*B) Example spectra of mucin glycans purified from the ascending colon*

*Glycans were liberated from purified mucins by reductive  $\beta$ -elimination, and analysed by MALDI-TOF using a DHB matrix. The relative abundance of each glycan structure was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The most abundant glycan peaks are annotated with mass to charge ( $m/z$ ) ratio and structure(s).*



**Figure 22| Glycan structure profiles of sigmoid and ascending colon in control and UC patients**

*A) Mucin glycosylation profile in sigmoid colon*

*B) Mucin glycosylation profile in ascending colon*

*The relative abundance of glycan structures in each patient sample was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The 24 most prevalent glycan structures across patients are shown in the graph. Each data point represents the mean relative abundance of the glycan of interest in the patient group of interest. The standard error of the mean is also shown for glycan structures in each patient group.*

Structure	m/z	% total samples containing structure	Mean relative abundance			
			CS	CA	UCS	UCA
1 Gal, GalNAcol	534	96.9	8.56	5.5	12.03	9.99
1 NeuAc, GalNAcol	691 (S)	71.9	1.11	0.53	1.95	0.99
1 Gal, 1 Fuc, GalNAcol	708 (F)	84.4	2.86	6.62	3.26	5.04
1 Gal, 1 HexNAc, GalNAcol	779	100	14.48	6.2	15.67	9.02
1 Gal, 1 NeuAc, GalNAcol	895 (S)	87.5	3.62	3.08	3.38	4.29
1 HexNAc, 1 NeuAc, GalNAcol	936 (S)	81.3	3.07	1.02	3.29	2.85
1 Gal, 1 HexNAc, 1 Fuc, GalNAcol	953 (F)	87.5	5.57	8.29	3.90	6.97
2 Gal, 1 HexNAc, GalNAcol	983	100	9.66	4.54	10.07	5.57
1 Gal, 2 HexNAc, GalNAcol	1024	90.6	1.53	1.6	1.55	1.64
1 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1140 (S)	71.9	0.98	0.24	1.39	1.55
2 Gal, 1 HexNAc, 1 Fuc, GalNAcol	1157 (F)	90.6	6.25	9.73	4.93	8.09
1 Gal, 2 NeuAc, GalNAcol	1256 (S)	43.8	0.74	1.12	0.21	1.34
1 Gal, 1 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	1314 (S)	56.3	0.93	0.09	0.85	0.85
2 Gal, 1 HexNAc, 2 Fuc, GalNAcol	1331 (F)	81.3	3.81	7.16	2.07	3.86
2 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1344 (S)	56.3	0.26	0.27	0.70	0.66
2 Gal, 2 HexNAc, 1 Fuc, GalNAcol	1402 (F)	81.3	2.66	4.19	1.75	3.00
3 Gal, 2 HexNAc, GalNAcol	1432	81.3	1.11	0.44	1.68	1.69
2 Gal, 1 HexNAc, 3 Fuc, GalNAcol	1505 (F)	50.0	1.96	4.82	0.69	1.67
2 Gal, 1 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	1518 (F)(S)	56.3	0.77	0.99	0.43	0.81
3 Gal, 2 HexNAc, 1 Fuc, GalNAcol	1606 (F)	78.1	1.03	1.24	1.06	1.23
2 Gal, 3 HexNAc, 1 Fuc, GalNAcol	1647 (F)	59.4	0.51	0.77	0.44	0.79
3 Gal, 2 HexNAc, 2 Fuc, GalNAcol	1780 (F)	68.8	1.88	2.93	0.97	1.83
3 Gal, 3 HexNAc, 1 Fuc, GalNAcol	1851 (F)	75.0	1.05	1.17	0.68	1.07
3 Gal, 3 HexNAc, 2 Fuc, GalNAcol	2025 (F)	68.8	1.32	1.95	0.51	1.17

**Table 8 | Percentage of total samples containing structure, and mean relative abundance of 24 most prevalent glycan structures in 4 patient groups**

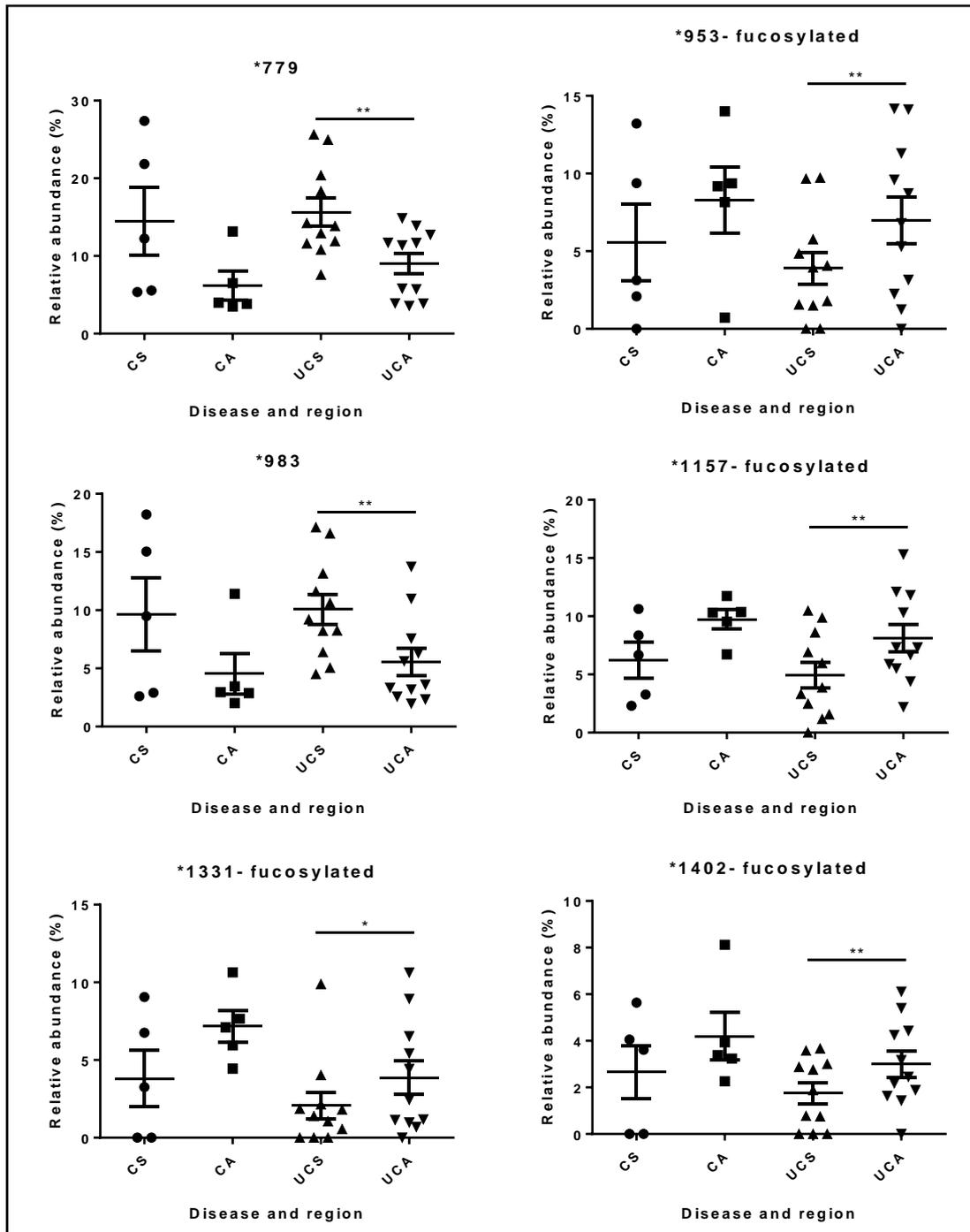
*The relative abundance of glycan structures in each patient sample was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The mean relative abundance is shown for a subset of 24 prevalent glycans across 4 patient sample groups. Full data including all glycan structures and their relative abundance in each patient sample can be found in appendix 3.*

*Key; (F), Fucosylated, (S), Sialylated, CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending*

Wilcoxon matched pairs signed rank tests were performed to assess for statistically significant differences in the abundance of the 24 most common glycans between sigmoid and ascending colon of controls and of UC patients. A number of structures were found to be significantly different in abundance between the sigmoid and ascending colon of UC patients; m/z 779 (P= 0.0098), m/z 953 (P= 0.0039), m/z 983 (P= 0.0098), m/z 1157 (P= 0.0098), m/z 1331 (P= 0.0273), m/z 1402 (P=0.0098), m/z 1505 (P=0.0313), m/z 1647 (P=0.0391), m/z 1780 (P= 0.0156), m/z 2025 (P= 0.0078) (Fig. 23) (Table 9). However, none of these P values fell below the threshold value once controlling for the false discovery rate  $q= 0.05$ . Of the 10 structures, 8 were fucosylated (m/z 953, m/z 1157, m/z 1331, m/z 1402, m/z 1505, m/z 1647, m/z 1780, m/z 2025), and the significant difference was in agreement with the reduced abundance of Fuc in the sigmoid colon compared with the ascending region.

Next, Mann-Whitney tests were performed to determine if there were significant differences between the abundance of glycan structures between the sigmoid colon of controls and UC patients, and similarly for the ascending colon. Two significant differences were observed between the ascending colon regions of controls and UC patients. These corresponded to an increased abundance of the sialylated structure, m/z 1140 in UC patients (P= 0.0193) compared with controls, as well as an increased abundance of the sialylated and fucosylated structure m/z 1314 in UC (P= 0.0169). Neither of these P values fell below the threshold value when controlling for the false discovery rate (Table 9).

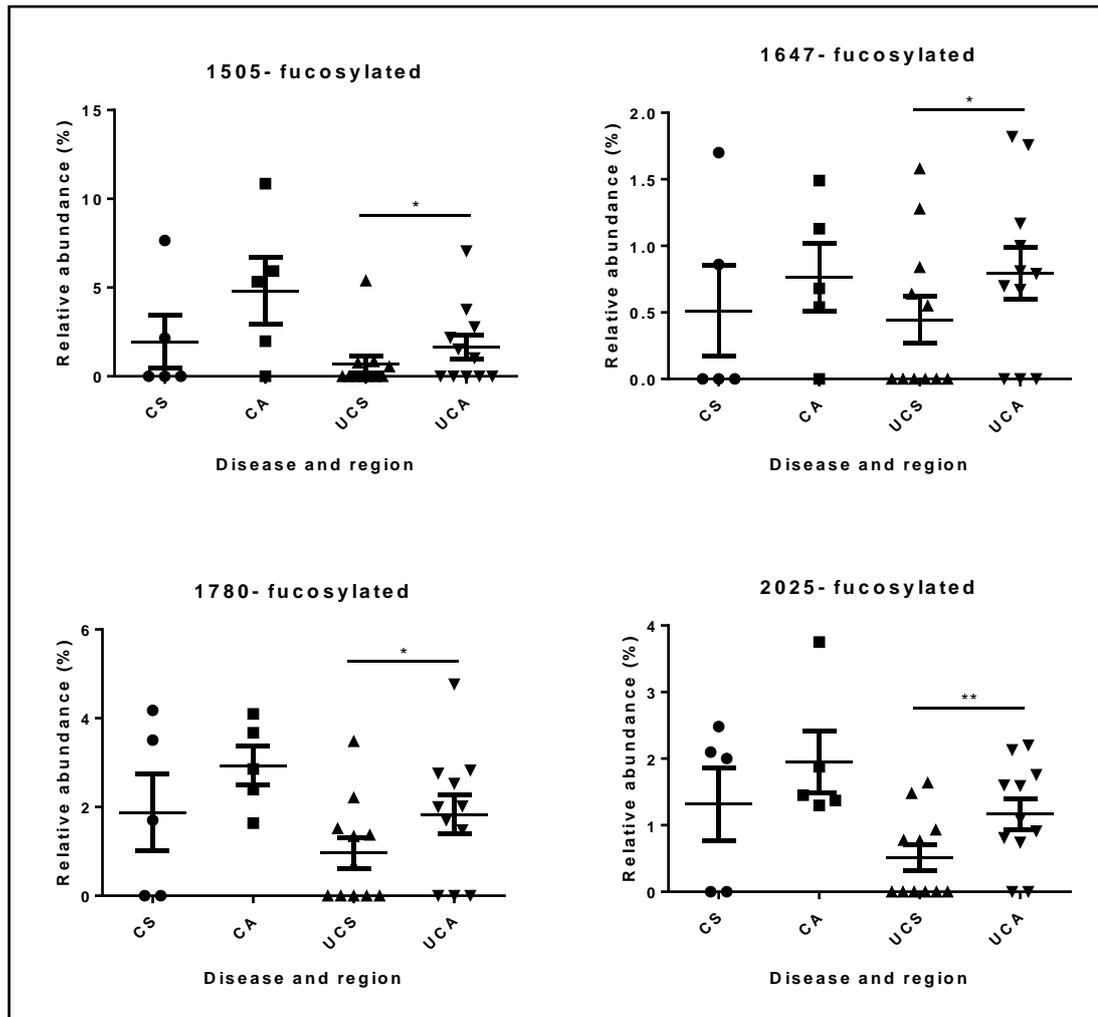
However, the profiles of sialylated and fucosylated structures out of the 24 most abundant glycans revealed apparent trends for differences between the two groups. In particular, the abundances of most fucosylated structures was higher in control patients when compared to UC in both the sigmoid and ascending colon (Fig. 24). In addition, in the ascending colon, the abundance of most sialylated structures was higher in UC patients (Fig. 25).



**Figure 23| Dot plots of glycan structures determined to be significantly different in abundance between sigmoid and ascending colon of UC patients**

The relative abundance of glycan structures in each patient sample was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. Statistical tests were carried out on the 24 most prevalent glycans identified across patients. The mean and standard error is shown for the relative abundance of glycans in each patient group. Statistically significant differences between patient groups are indicated although none of the P values fell below the false discovery threshold. See appendix 5 for dot plots of glycan structures where no significant differences were calculated between patients.

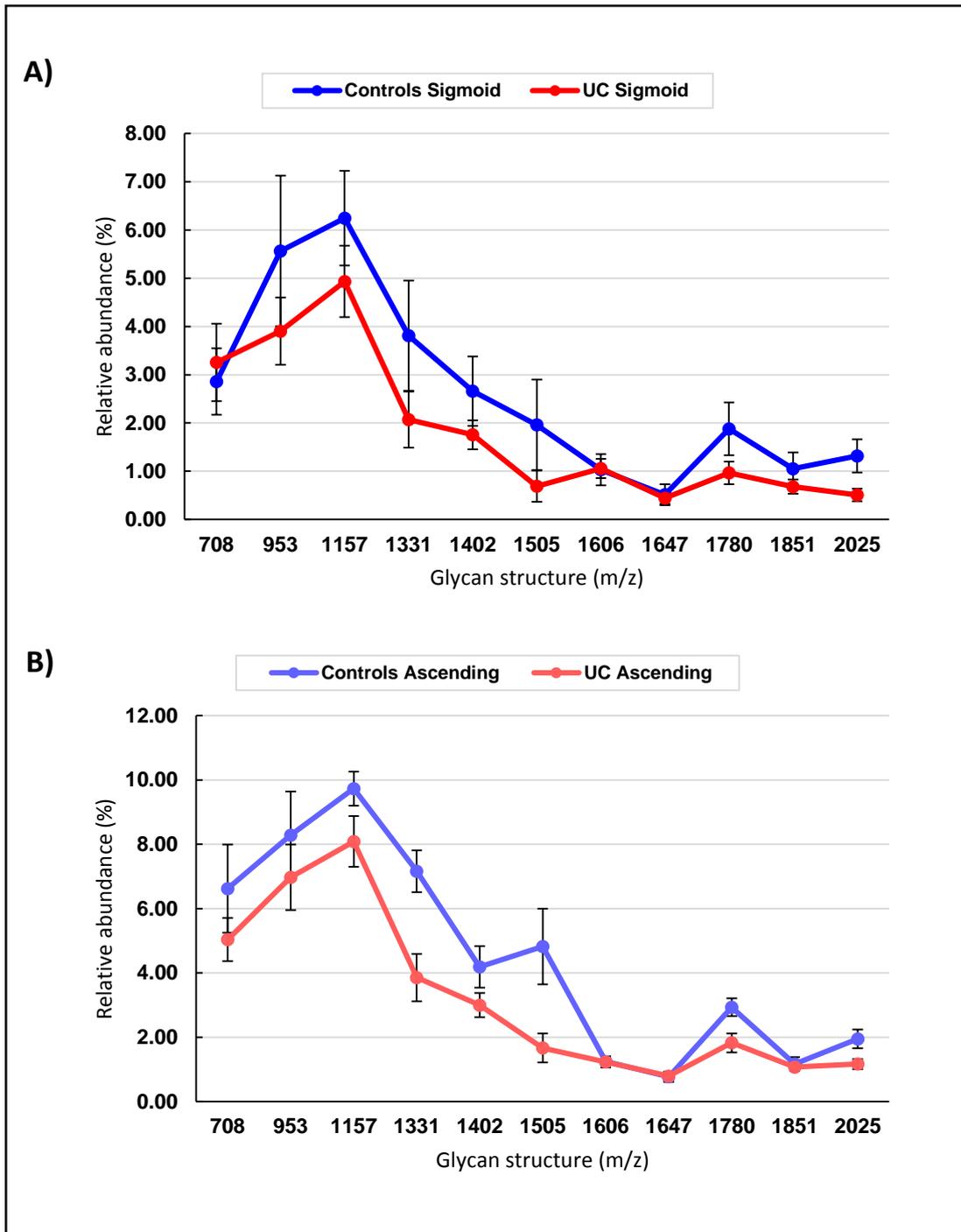
Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending



**Figure 23 continued | Dot plots of glycan structures determined to be significantly different in abundance between sigmoid and ascending colon of UC patients**

*The relative abundance of glycan structures in each patient sample was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. Statistical tests were carried out on the 24 most prevalent glycans identified across patients. The mean and standard error is shown for the relative abundance of glycans in each patient group. Statistically significant differences between patient groups are indicated although none of the P values fell below the false discovery threshold. See appendix 5 for dot plots of glycan structures where no significant differences were calculated between patients.*

*Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending*

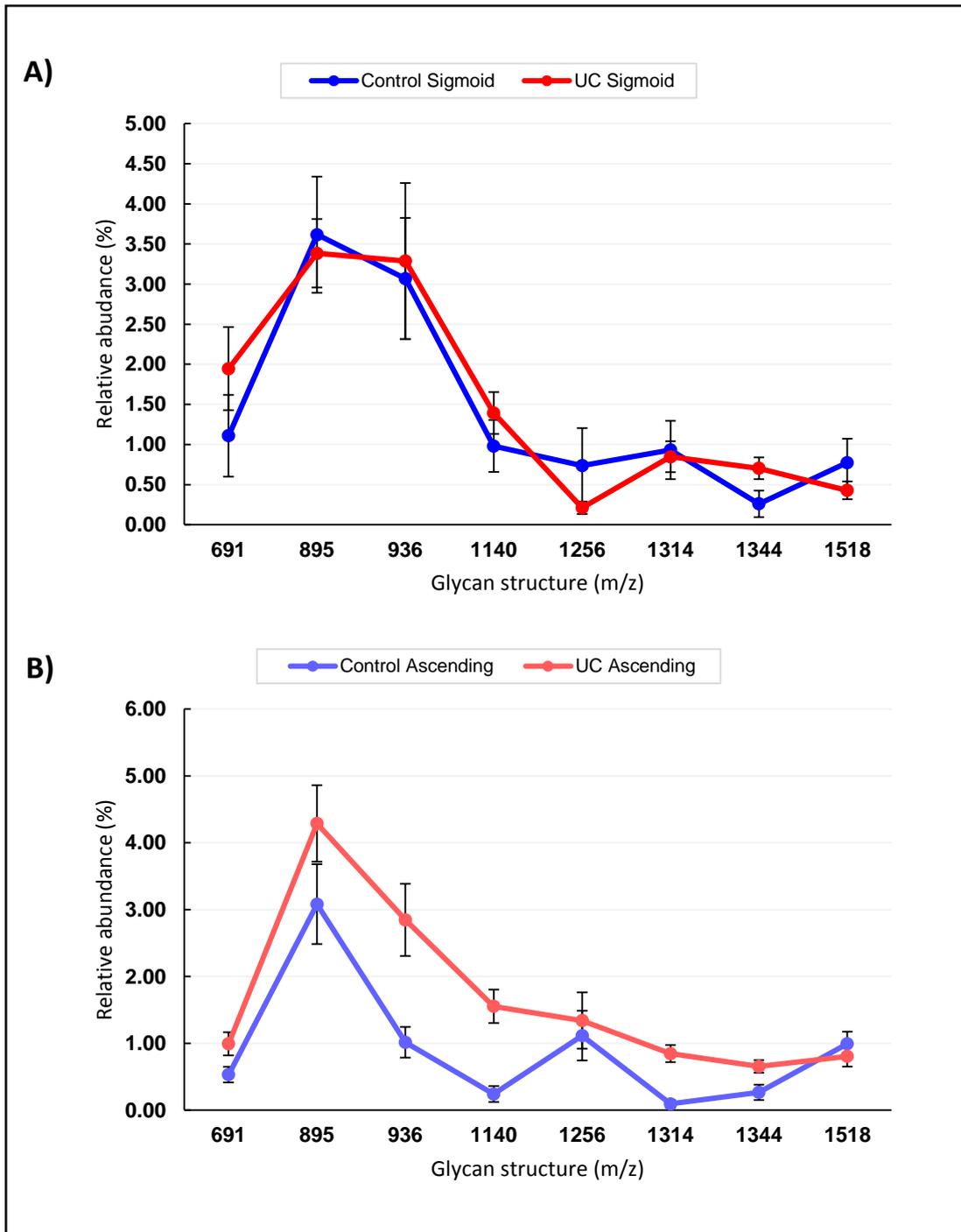


**Figure 24| Profile of fucosylated structures in sigmoid and ascending colon of control and UC patients**

*A) Profile of fucosylated glycan structures in the sigmoid colon*

*B) Profile of fucosylated glycan structures in the ascending colon*

*The relative abundance of fucosylated glycan structures in each patient sample was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The mean and standard error is shown for the relative abundance of fucosylated glycans, selected out of a subset of the 24 most prevalent structures across patients.*



**Figure 25 | Profile of sialylated structures in sigmoid and ascending colon of control and UC patients**

A) Profile of sialylated glycan structures in the sigmoid colon

B) Profile of sialylated glycan structures in the ascending colon

The relative abundance of sialylated glycan structures in each patient sample was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The mean and standard error is shown for the relative abundance of sialylated glycans, selected out of a subset of the 24 most prevalent structures across patients.

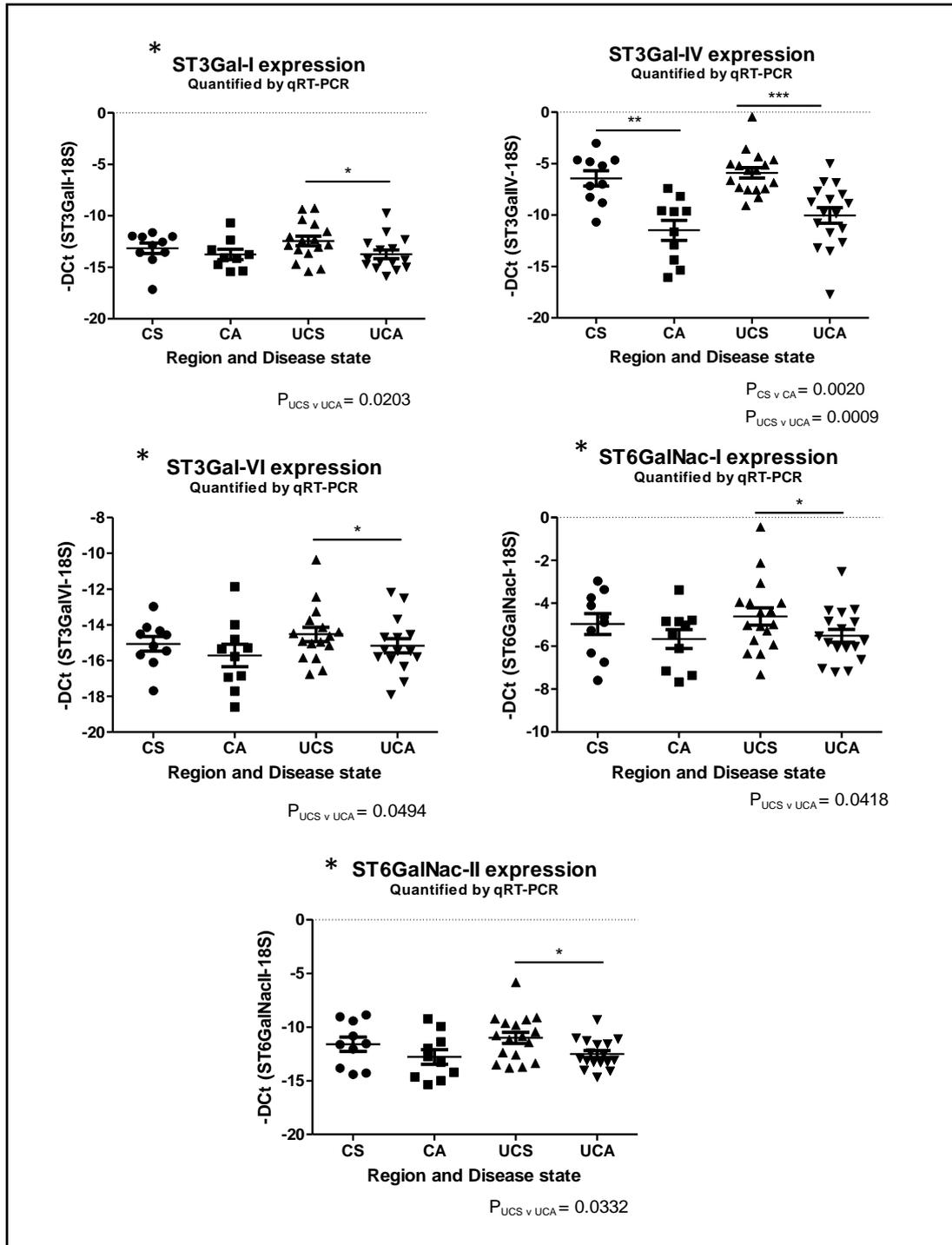
#### 4.4 Quantitation of mucin glycosyltransferase expression by qRT-PCR

To analyse the expression of the main mucin glycosyltransferases involved in the generation of core, sialylated or fucosylated structures, total RNA was extracted from human mucosal biopsies taken from the sigmoid and ascending colon and reverse transcribed to cDNA. The expression of 10 glycosyltransferases, *C1GalT1*, *C3GnT* and *Fut2*, selected based on previous links to IBD and, *ST3Gal-I*, *ST3Gal-II*, *ST3Gal-III*, *ST3Gal-IV*, *ST3Gal-VI*, *ST6GalNAc-I* and *ST6GalNAc-II* for their role in the sialylation of mucins, was analysed by qRT-PCR. Relative expression was determined in comparison to the reference 18S gene (-DcT). qRT-PCR was successfully carried out on 69 of the available colonic samples from 37 patients. The most abundantly expressed glycosyltransferases out of those analysed were *C1GalT1*, *C3GnT*, *ST3Gal-IV* and *ST6GalNAc-I*.

As with the analysis of the microbiota composition and sialic acid abundance, for statistical analysis, the data from four groups of samples were analysed: sigmoid colon of controls, ascending colon of controls, sigmoid colon of UC patients, ascending colon of UC patients (See section 3.2.2). The resulting data consisted of 20 samples from 10 control patients and 34 samples from 17 UC patients. Wilcoxon matched pairs signed rank tests were performed to assess for significant differences between the expression of glycosyltransferases in the sigmoid and ascending colon in each patient group. These revealed significant differences in the expression of glycosyltransferases; namely sialyltransferases. In particular, the expression of the sialyltransferase *ST3Gal-IV* was significantly higher in the sigmoid colon in both controls and UC patients ( $P= 0.0020$  and  $P= 0.0009$ , respectively) (Fig. 26) (Table 10). These  $P$  values were lower than the threshold value set when controlling the false discovery rate to below  $q= 0.05$ . Furthermore, the expression of several other sialyltransferases were significantly enhanced in the sigmoid colon of UC patients; *ST3Gal-I* ( $P= 0.0203$ ), *ST3Gal-VI* ( $P= 0.0494$ ), *ST6GalNAc-I* ( $P=0.0418$ ), *ST6GalNAc-II* ( $P= 0.0332$ ), however these  $P$  values did not fall below the false discovery threshold (Fig. 26) (Table 10). Furthermore, whilst the same trend was apparent in control patients, the differences between the ascending and sigmoid colon failed to reach statistical significance (Table 10). In contrast, using Mann-Whitney tests, no significant differences or trends could be observed in the expression of glycosyltransferases between control and UC patients (Table 10).

A PCA plot of the gene expression data confirmed the above findings. Distinct clustering of samples was apparent based on colon location, suggesting different expression profiles in the sigmoid and ascending regions (Fig. 27). In contrast, there was no obvious clustering

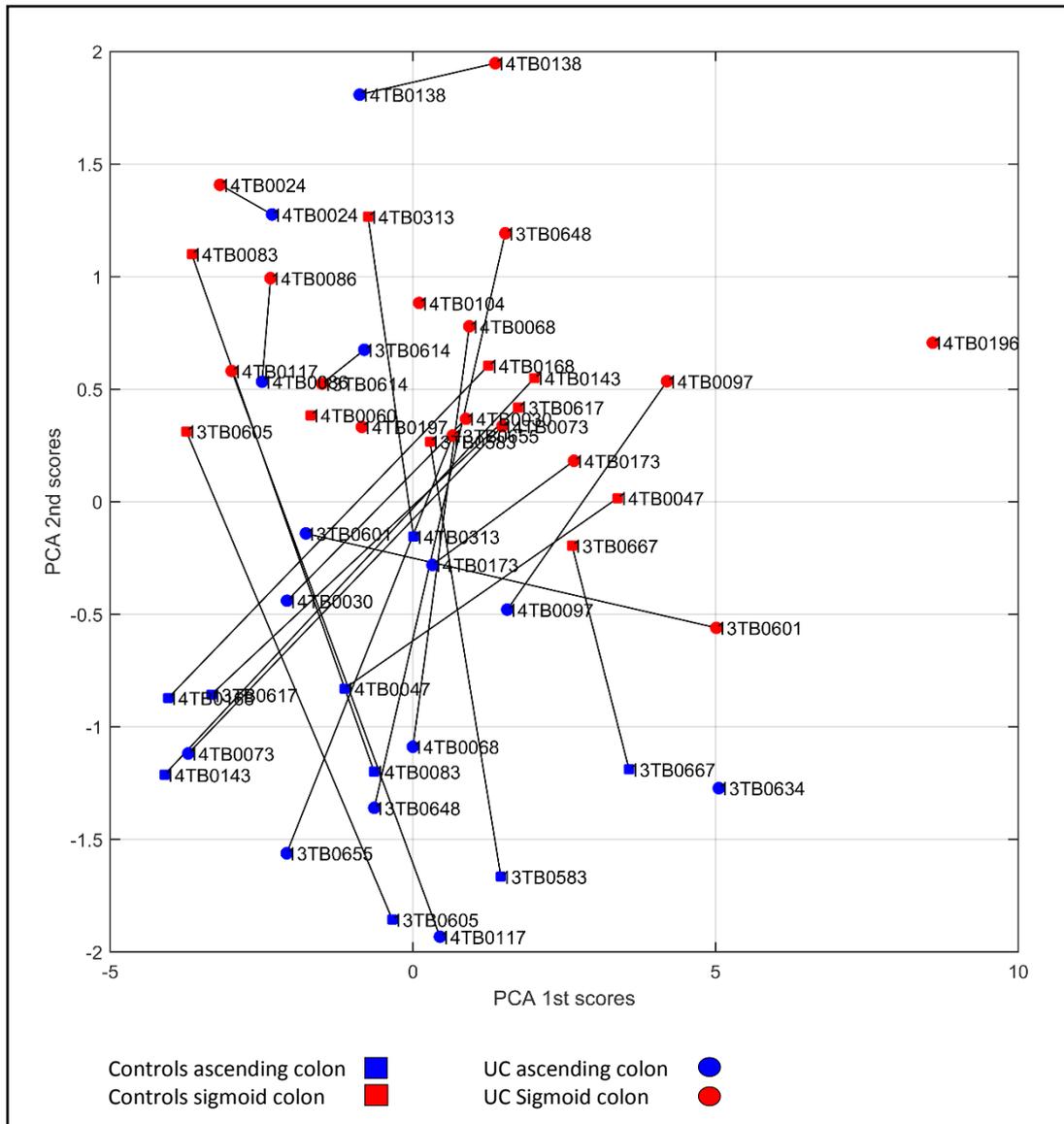
based on disease state, suggesting no difference between control patients and patients with UC (Fig. 27).



**Figure 26| Sialyltransferases with significantly different expression patterns between sample groups**

Glycosyltransferase expression data was generated by normalisation of raw qRT-PCR Ct values to the Ct of an 18S reference gene. Three technical replicates were carried out for each qRT-PCR reaction. The mean and standard error is shown for the gene expression values in each patient group. Statistically significant differences between patient groups are indicated. \*P values for these glycosyltransferases do not fall below the false discovery rate. See appendix 5 for dot plots of glycosyltransferases where no significant differences were calculated between the expression in different patient groups.

Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending



**Figure 27 | PCA plot of glycosyltransferase expression data**

*Glycosyltransferase expression data was generated by normalisation of raw qRT-PCR Ct values to the Ct of an 18S reference gene. Samples are plotted following Eigen-value decomposition of the variable matrix. Samples positioned closest on the PCA plot are more similar in terms of glycosyltransferase gene expression, whilst those positioned further apart are more dissimilar. Paired samples from the same patient are indicated with connecting lines. The PCA 1<sup>st</sup> score summarises the largest amount of the total variance in the data (72.1%), whilst the PCA 2<sup>nd</sup> score summarises the second largest amount (9.2%).*

	Wilcoxon matched pairs signed rank test		Mann-Whitney test	
	CS v CA	UCS v UCA	CS v UCS	CA v UCA
534 (TF antigen/Core 1)	0.3125	0.1309	0.4278	0.7340
691 (Sialyl-Tn)	0.8125	0.3223	0.4496	0.4264
708 (F)	0.4375	0.0645	0.8641	0.6504
779 (Core 2 with GlcNAc)	0.1250	<b>0.0098</b>	0.7340	0.3079
895 (Sialyl-TF)	0.4375	0.6953	1.0000	0.4278
936 (S)	0.3125	0.7646	0.7754	0.3352
953 (F)	0.4375	<b>0.0039</b>	0.7332	0.8208
983	0.4375	<b>0.0098</b>	0.9098	0.5711
1024	1.0000	0.8311	0.4612	0.9095
1140 (S)	0.2500	0.8457	0.5650	<b>0.0193</b>
1157 (F)	0.0625	<b>0.0098</b>	0.5711	0.4278
1256 (Disialyl-TF)	0.7500	0.0547	1.0000	0.9074
1314 (S)(F)	0.5000	0.7695	0.9518	<b>0.0169</b>
1331 (F)	0.1250	<b>0.0273</b>	0.6455	0.0699
1344 (S)	1.0000	0.2500	0.2761	0.1456
1402 (F)	0.1875	<b>0.0098</b>	0.4209	0.3079
1432	0.2500	0.1971	0.4257	0.2114
1505 (F)	0.2500	<b>0.0313</b>	0.6487	0.1459
1518 (S)(F)	0.6250	0.0925	0.7080	0.7299
1606 (F)	0.8125	0.6523	0.9541	1.0000
1647 (F)	0.6250	<b>0.0391</b>	0.9007	0.8641
1780 (F)	0.3125	<b>0.0156</b>	0.3441	0.1397
1851 (F)	0.8750	0.1055	0.7269	0.6097
2025 (F)	0.3125	<b>0.0078</b>	0.1832	0.2569
Sialic Acid	<b>0.0273</b>	<b>0.0018</b>	0.2799	0.1622

**Table 9 | P values for statistical tests comparing glycosylation in UC patients and controls**

*P values are shown for the comparison of sigmoid vs. ascending and control vs. UC patients. Statistical tests were carried out to assess for significant differences in the relative abundance of glycan structures quantified in human samples by MALDI-TOF and in the abundance of sialic acid measured by HPAEC-PAD. P values  $\leq 0.05$  are considered significant (highlighted in bold).*

*Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending*

	Wilcoxon matched pairs signed rank test		Mann-Whitney test	
	CS v CA	UCS v UCA	CS v UCS	CA v UCA
<b>C1GalT1</b>	0.7695	0.8871	0.4666	0.8212
<b>C3GnT</b>	0.3223	0.8129	0.9001	0.4979
<b>ST3Gal-I</b>	0.3594	<b>0.0203</b>	0.5800	0.9525
<b>ST3Gal-II</b>	0.6953	0.2012	0.6334	0.3033
<b>ST3Gal-III</b>	0.5566	0.0929	0.6926	0.4074
<b>ST3Gal-IV</b>	<b>0.0020</b>	<b>0.0009</b>	0.7824	0.3033
<b>ST3Gal-VI</b>	0.5566	<b>0.0494</b>	0.5444	0.5982
<b>ST6GalNAc-I</b>	0.3750	<b>0.0418</b>	0.9001	0.8212
<b>ST6GalNAc-II</b>	0.3750	<b>0.0332</b>	0.5303	0.5981
<b>Fut2</b>	0.2754	0.4777	0.8605	0.9800

**Table 10| P values for statistical tests comparing glycosyltransferase expression in UC patients and controls**

*P values are shown for the comparison of sigmoid vs. ascending and control vs. UC patients. Statistical tests were carried out to assess for significant differences in expression of glycosyltransferase enzymes quantified in human samples by qRT-PCR. P values  $\leq 0.05$  are considered significant (highlighted in bold).*

*Key; CS, Control Sigmoid, CA, Control Ascending, UCS, UC Sigmoid, UCA, UC Ascending*

## 4.5 Discussion

Our data showed that it was possible to isolate mucins from human mucosal lavage in a form and amount enabling future structural characterisation. To our knowledge, this is the first study reporting the purification of mucin and imaging by AFM from lavage samples. Human colonic mucins from lavage formed a wide variety of mucin chain lengths, similar to PGM, (Gunning et al., 2013; Znamenskaya et al., 2012), as also observed with bovine submaxillary mucin, human ocular mucins, and human respiratory mucins (Baos et al., 2012; Kesimer et al., 2013; Znamenskaya et al., 2012). Visualisation of mucins purified from human mucosal lavage suggested the increased abundance of mucin in the GuHCl soluble fractions as compared to the insoluble fraction, in contrast to mucins isolated from human biopsies (Larsson et al., 2009). The solubility of the mucin fractions suggest that the mucus sampled *via* the lavage technique mainly originates from the loose outer mucus layer, where it is in direct contact with the gut microbiota (Johansson et al., 2010; Johansson et al., 2008).

In mucins purified from mucosal biopsies, it has been shown by AgPAGE that GuHCl insoluble mucins form monomers (monomeric MUC2 ~2.5 MDa) as well as non-reducible dimers and other oligomers (Larsson et al., 2009). Here, purified mucins from mucosal lavage formed double to quadruple bands in AgPAGE. Additional experiments are required to further characterise the molecular weight and proteome of mucins from mucosal lavage, e.g. using analytical ultracentrifugation and shotgun proteomic approaches (Herrmann et al., 1999; Kesimer and Sheehan, 2012). Here, we focussed on the glycosylation component of the mucins. The heavily glycosylated nature of the mucins was apparent on AgPAGE by the lack of protein specific staining but positive glycan staining, suggesting that the protein backbone was concealed and made inaccessible by the glycan side chains.

Inter-individual variation in terminal mucin glycosylation was observed by lectin screening and force spectroscopy. Heterogeneity of Fuc and sialic acid was particularly of note *via* lectin probing. Both of these epitopes are frequently found terminally in mucin glycans in the colon and therefore are the initial sugars that must be cleaved and degraded to allow access to the rest of the glycan (Tailford et al., 2015a). Detailed MS analysis also revealed many sialylated and fucosylated structures and confirmed a large degree of inter-individual variability. In line with our observations, in a previous study on mucins purified from biopsies over 100 complex O-linked oligosaccharides were identified (Larsson et al., 2009). However, it is important to note that in our study, this number is probably underestimated since MALDI-TOF cannot distinguish between isomer oligosaccharides. Whilst a subset of 24 structures were present

in most individuals, the abundance of most of these structures tended to be quite variable from one individual to the next. In the study by Larsson and coll. the repertoire and relative abundance of glycans on mucins purified from biopsies appeared to be relatively constant in healthy human subjects (n= 25) (Larsson et al., 2009). It is possible that mucus from lavage samples has been subjected to increased degradation by bacterial glycoside hydrolases than mucus from biopsies, due to its closer proximity to the microbiota, thereby explaining the higher level of inter-individual heterogeneity.

As previously reported with biopsies (Robbe et al., 2003), by MALDI-TOF, we were able to detect a significant decrease in the overall fucosylation of mucins, as well as in the abundance of many individual fucosylated glycan structures from proximal to distal colon. Furthermore, there was a trend towards an increase in sialic acid abundance towards the distal end of the colon, mainly detectable in control patients, but also apparent in UC, although not significant. It is important to note that the analysis performed by MALDI-TOF in our study is only semi-quantitative, since it did not account for glycan structures possessing two or more sialic acid or Fuc epitopes. In accordance with this, quantification by HPAEC-PAD, where the total amount of sialic acid is measured, revealed that the increase in sialic acid from proximal to distal colon was significant in both patient groups. This was reflected in a distinct increasing gradient of sialyltransferase expression from ascending to sigmoid colon in both groups. *ST6GalNAc-I* was amongst the sialyltransferases to display this trend in both controls and UC patients. The sialyltransferase with the most significant difference between proximal and distal colon was the *ST3Gal-IV* sialyltransferase, which was also one of the most highly expressed glycosyltransferases out of those studied, in accordance with Affymetrix GeneChip assays in healthy colonic tissue (Kemmer et al., 2003). It is important to note that changes in gene expression do not necessarily correlate with protein expression levels or enzyme activity, however an increasing protein abundance of both *ST6GalNAc-I* and *ST3Gal-IV* from ascending to sigmoid colon has been demonstrated previously by proteomic analysis (van der Post and Hansson, 2014). It has been suggested that *ST3Gal-IV* is involved in the generation of the Sda/Cad epitope, of which there is also an increasing gradient from proximal to distal colon (van der Post and Hansson, 2014). However, we were unable to decipher the abundance of this epitope in our patients because, as previously mentioned MALDI-TOF data only gives the molecular ions and does not distinguish between isomer oligosaccharides.

Comparisons between control and UC patients revealed that there were no significant alterations in the overall abundance of sialic acid by HPAEC-PAD. However, the profile of 24

most common, and abundant mucin glycans detected by MS displayed subtle alterations in UC patients compared with controls. For example, the abundance of shorter glycan structures tended to be greater in UC patients than in controls, whilst the reverse was true for longer, more complex glycans. A tendency towards shorter glycan structures in UC has been observed previously in mucins from biopsies (Larsson et al., 2011). In particular, the abundance of the short sialylated structure, STn was found to be significantly increased in UC patients (Larsson et al., 2011). The abundance of STn (glycan structure m/z 691) and indeed many of the other sialylated structures from the subset of 24 appeared to be higher in UC than in controls in our study, particularly in the ascending colon. In addition, in both sigmoid and ascending colon, the abundance of fucosylated structures in the subset of 24 was lower in UC patients. However, none of the observed differences were significant enough to fall below the false discovery rate. The lack of significance in our data may again be attributed to inter-individual variation associated with the host genetics and/or the composition of the mucus-associated microbiota (in particular, the abundance of mucin degraders), and reflects the need for a higher number of samples in both patient groups.

It has been proposed that the increased prevalence of STn antigen in UC patients may be due to an increased expression of ST6GalNAc-I enzyme (Larsson et al., 2011). Furthermore, alterations in core 3 and core 1 glycosyltransferase expression were recently demonstrated to compromise the mucus layer and cause subtle alterations in the microbiota in mice (Bergstrom et al. 2016, Sommer et al. 2014). However, we could not detect significant differences in the expression of these enzymes or the other glycosyltransferases measured in UC patients in our study, when compared with controls. It is possible that the expression of other glycosyltransferases (not listed in this study), responsible for generating the precursors of the glycans modified in UC, are altered. However, the lack of significant alterations in glycosyltransferase expression between controls and UC patients may also suggest that other factors, such as the mucus-associated microbiota composition can drive the observed changes in mucin glycosylation in UC.

Overall, we observed an increasing gradient of sialic acid and sialyltransferase expression, and a decreasing gradient of Fuc from proximal to distal colon, supported by MS, HPAEC-PAD and qRT-PCR. In contrast, only subtle alterations were apparent in the glycosylation profile of mucins from mucosal lavage of UC patients when compared with controls, following a trend similar to that previously described from biopsies. However, the alterations were less significant than previously described. The glycosylation of mucins from the outer layer is probably less uniform between individuals due to the relative increased complexity in the

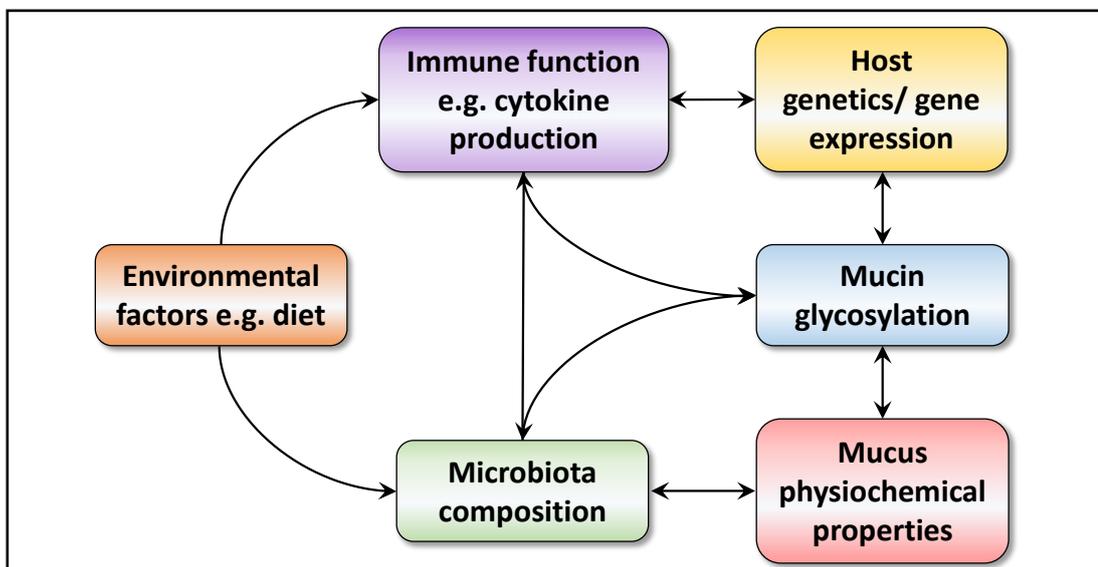
outer mucus layer when compared to the sterile inner layer, caused by additional factors, such as bacterial enzymes remodelling the mucosal surface. As a result, changes in the outer mucus layer in UC are likely to be more subtle, requiring larger study cohorts to increase significance. In order to further elucidate the mucus-microbiota relationship in IBD, additional studies addressing the composition of these components in both the inner and outer mucus layer would be beneficial. Furthermore, studies of glycosyltransferase protein expression and activity, in addition to gene expression would help to clarify the role of these enzymes in the altered glycosylation observed in UC.

## Chapter 5) Associations between mucin glycosylation and mucin degraders

### 5.1 Introduction and objectives

Evidence from previous studies suggests a deviation of the mucosal barrier from its homeostatic state in IBD. This includes changes in the GI microbiota composition and host genetics. These components are interlinked and an initial alteration in one of these may drive alterations in the others and exacerbate disease (Fig. 28).

Changes in mucin glycosylation driven by host genetics may directly impact on mucus-associated microbiota composition (including mucin-degraders) and/or physiochemical properties of the mucus layer. As a result, the mucus may become more penetrable to pathogenic bacteria, resulting in increased epithelial contact and inflammation. Alternatively, changes in the mucus-associated microbiota may be driven by alterations in environmental factors such as the diet, and in turn lead to a remodelling of the mucin glycan profile *via* direct mucin degradation, or indirectly through changes in glycosyltransferase expression (Fig. 28). In order to thoroughly investigate this relationship, mechanistic studies in mice and *in vitro* must be performed in parallel with association studies in humans.



**Figure 28| Schematic diagram demonstrating the various components which may impact on the development of IBD and how they may be interlinked**

Here, we decided to focus our attention on two mucin degrading strains of *A. muciniphila* and *R. gnavus*, and their association with mucin glycosylation. In IBD, these species have been found to be differentially altered; whilst the abundance of *A. muciniphila* decreased in IBD, the abundance of *R. gnavus* increased (Png et al., 2010), resulting in a decreased *A. muciniphila*/*R. gnavus* ratio. Our findings by qPCR recapitulated the depletion of *A. muciniphila* in UC, and whilst we did not observe a significant increase of *R. gnavus*, it did not display the same trend as *A. muciniphila*, however it was not possible to obtain information at the strain level. In IBD, an increase in shorter sialylated glycans has also been observed (Larsson et al., 2009). We hypothesised that the difference in the abundance of these bacteria in disease may be closely associated to changes in the glycosylation profile in IBD. Indeed, *A. muciniphila* and *R. gnavus* show distinct mucin degrading strategies reflecting the glycoside hydrolase repertoire of these strains. *A. muciniphila* is more adapted to the utilisation of complex mucin glycans, whilst *R. gnavus* specifically targets sialic acid (Crost et al., 2016; Tailford et al., 2015b). We therefore hypothesised that the glycosylation profile observed in IBD would favour the mucin degradation strategy of *R. gnavus*, leading to an altered ratio of *A. muciniphila* and *R. gnavus*. Mechanistic investigations were performed *in vitro* to help elucidate if the glycan landscape differentially affects the growth of these two bacteria, and *in vivo* to assess whether these bacteria can modify the glycosylation profile.

## 5.2 Mechanistic studies of the association between *R. gnavus* CC55\_001C, *A. muciniphila* ATCC BAA 835 and mucin glycosylation

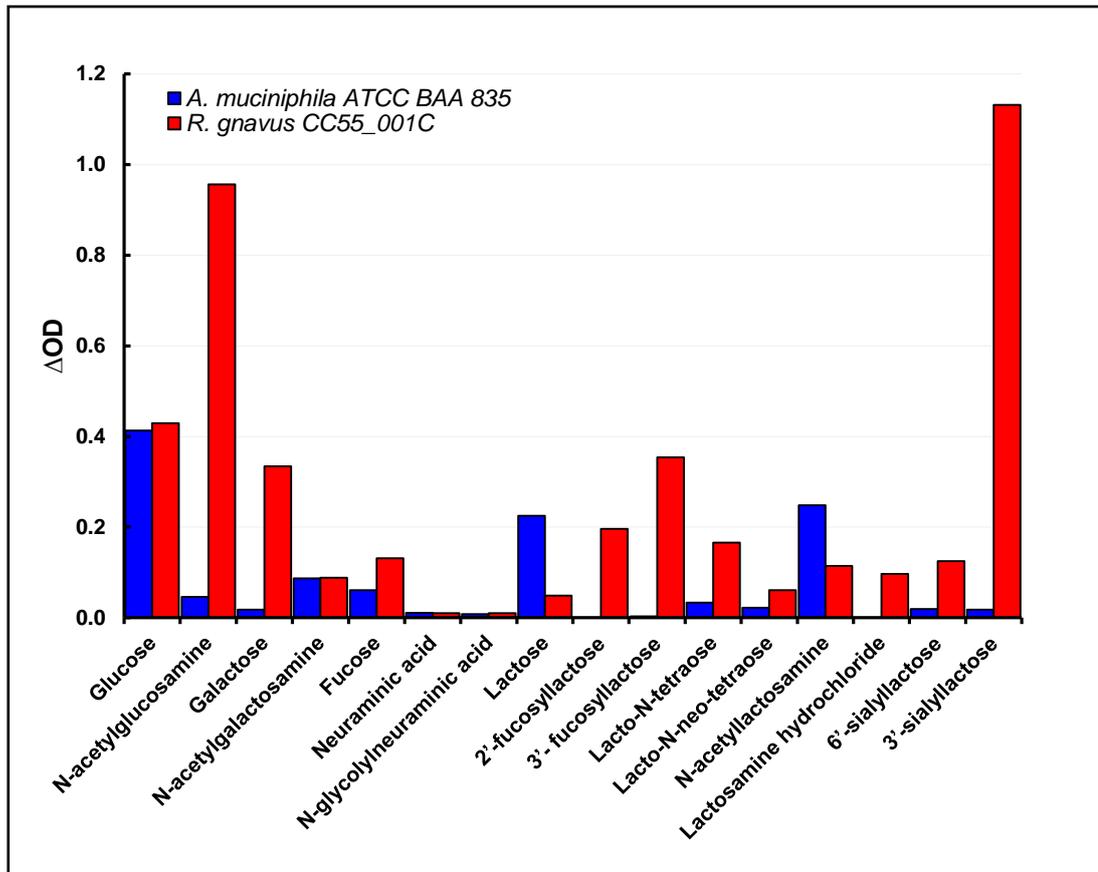
*In vitro* and *in vivo* mechanistic investigations of the association of *R. gnavus* and *A. muciniphila* abundance with mucin glycosylation were carried out using the following representative organisms; *R. gnavus* CC55\_001C, isolated from colonic biopsy tissue of a human subject as part of the HMP, and *A. muciniphila* ATCC BAA 835, isolated from the faeces of a healthy adult (Derrien et al., 2004).

### 5.2.1 Profiling of *R. gnavus* and *A. muciniphila* growth in monosaccharides and oligosaccharides

To obtain an initial overview of the carbohydrate utilisation abilities of *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835, the anaerobic growth of both bacteria were monitored on basal medium supplemented with diverse host monosaccharides and oligosaccharides as sole carbon sources. The growth assays were performed under strict anaerobic conditions. After 48 h the optical density of the control media was compared with that of the media inoculated with each bacteria to determine any increases in OD, and hence growth. A  $\Delta OD$  of 0.1 or greater was considered as an indication of growth.

*R. gnavus* CC55\_001C grew well on 3' SL ( $\Delta OD = 1.13$ ), and GlcNAc ( $\Delta OD = 0.96$ ). In addition growth was detected on Glc ( $\Delta OD = 0.43$ ), Gal ( $\Delta OD = 0.33$ ), Fuc ( $\Delta OD = 0.13$ ), 2' FL ( $\Delta OD = 0.20$ ), 3' FL ( $\Delta OD = 0.35$ ), LNT ( $\Delta OD = 0.17$ ), LacNAc ( $\Delta OD = 0.11$ ), Lactosamine hydrochloride ( $\Delta OD = 0.10$ ) and 6' SL ( $\Delta OD = 0.13$ ) (Fig. 29) (Table 11). *R. gnavus* could not grow on GalNAc, Neu5Ac, Neu5Gc, Lac or LNnT.

In marked contrast, growth of *A. muciniphila* ATCC BAA 835 was only observed on 3 of the substrates tested. *A. muciniphila* growth was greatest on Glc ( $\Delta OD = 0.41$ ), followed by Lac ( $\Delta OD = 0.23$ ) and LacNAc ( $\Delta OD = 0.25$ ) (Fig. 29) (Table 11). No growth was observed on GlcNAc, Gal, GalNAc, Fuc, Neu5Ac, Neu5Gc, 2' FL, 3' FL, LNT, LNnT, Lactosamine hydrochloride, 6' SL or 3' SL. These data showed that *R. gnavus* is more adapted to utilising host mono- and di-saccharides as compared to *A. muciniphila*.



**Figure 29 | Growth of *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C in different carbohydrate sources**

*Carbohydrate supplemented YCFA media was inoculated with A. muciniphila or R. gnavus cells from BHI-YH starter cultures. Growth was determined by measuring the change in optical density after 48 hours of inoculation of the carbohydrate supplemented media, as compared to control media without inoculum. Measurements were taken at a wavelength of 595 nm. A  $\Delta OD$  of  $\geq 0.1$  was considered as bacterial growth.*

Carbohydrate source	Abbreviation	<i>Ruminococcus gnavus</i> CC55_001C		<i>Akkermansia muciniphila</i> ATCC BAA 835	
		$\Delta$ OD	Growth?	$\Delta$ OD	Growth?
D-glucose	Glc	0.43	YES	0.41	YES
N-acetylglucosamine	GlcNAc	0.96	YES	0.05	NO
D-galactose	Gal	0.33	YES	0.02	NO
N-acetylgalactosamine	GalNAc	0.09	NO	0.09	NO
L-fucose	Fuc	0.13	YES	0.06	NO
N-acetylneuraminic acid	Neu5Ac	0.01	NO	0.01	NO
N-glycolylneuraminic acid	Neu5Gc	0.01	NO	0.01	NO
D-lactose	Lac	0.05	NO	0.23	YES
2'-fucosyllactose	2' FL	0.20	YES	0.00	NO
3'-fucosyllactose	3' FL	0.35	YES	0.00	NO
Lacto-N-tetraose	LNT	0.17	YES	0.03	NO
Lacto-N-neo-tetraose	LNT	0.06	NO	0.02	NO
N-acetyllactosamine	LacNAc	0.11	YES	0.25	YES
Lactosamine hydrochloride		0.10	YES	0.00	NO
6'-sialyllactose	6' SL	0.13	YES	0.02	NO
3'-sialyllactose	3' SL	1.13	YES	0.02	NO

**Table 11 | Changes in optical density of carbohydrate supplemented media after 48 hrs of inoculation with *A. muciniphila* ATCC BAA 835 or *R. gnavus* CC55\_001C**

*Growth was determined in a range of different carbohydrates by measuring the change in optical density at a wavelength of 595 nm. A  $\Delta$ OD of  $\geq 0.1$  was considered as bacterial growth.*

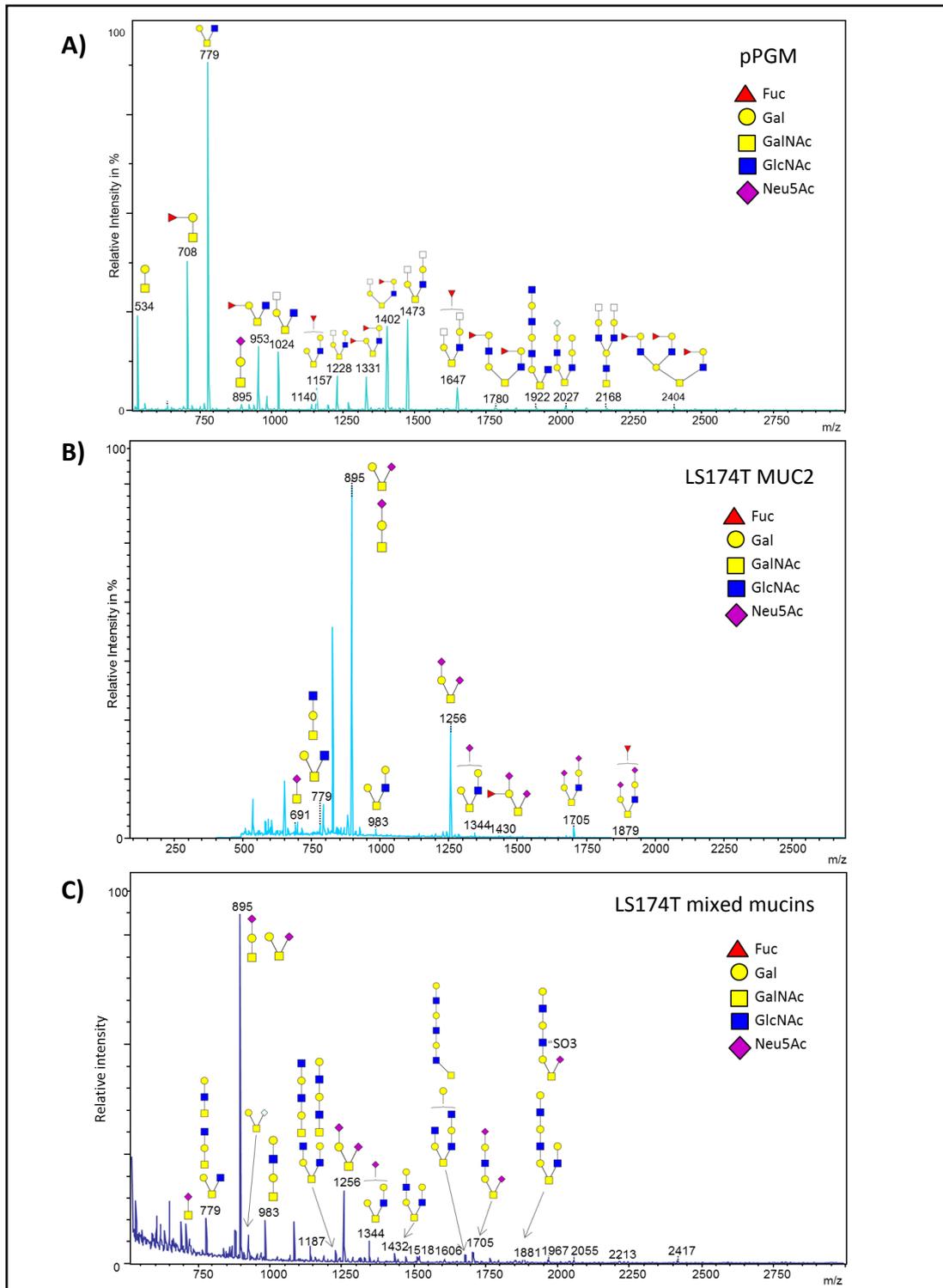
## 5.2.2 Profiling of *R. gnavus* and *A. muciniphila* growth in mucins

### 5.2.2.1 Characterisation of pPGM and LS174T mucin glycosylation

In order to compare the growth of *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C on mucins and assess how it may be affected by the mucin glycosylation profile, two different types of mucins were used for bacterial growth assays; pPGM which is commercially available and was purified as described in section 2.5.3, and mucin purified from the LS174T colon adenocarcinoma cell line (See section 2.5.4). Prior to testing the growth of *R. gnavus* and *A. muciniphila* on pPGM and LS174T mucin, the glycosylation profile of both types of mucin was analysed by liberation of the glycans and MALDI-TOF analysis (Fig. 30).

Analysis of pPGM glycans revealed short glycan structures as well as longer more complex glycans containing LacNAc (Gal $\beta$ 1-4GlcNAc) motifs. Relative quantification revealed that nearly half of the glycans were fucosylated whilst only around 2 % were sialylated (Table 12).

Two different mucin fractions were purified from the LS174T mucin which were obtained in an approximate 1:1 ratio from LS174T cells. The first fraction consisted of MUC2 mucins. The glycans from these mucins were mainly short and sialylated structures. These structures were dominated by sialyl-TF (~60 %), disialyl-TF (~23 %) and STn (~2 %). Approximately 3 % of structures on MUC2 mucin were also sulphated (Table 13). The other mucin fraction, containing a mixture of non-MUC2 mucins contained glycans with 70 % sialylation and 4.8 % sulphation (Table 14).



**Figure 30 | MALDI-TOF spectra of glycans liberated from pPGM and LS174T mucins**

A) Spectra obtained from pPGM

B) Spectra obtained from LS174T MUC2

C) Spectra obtained from LS174T mixed mucins

Glycans were liberated from purified mucins by reductive  $\beta$ -elimination, and analysed by MALDI-TOF using a DHB matrix. The relative abundance of each glycan structure was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The most abundant glycan peaks are annotated with mass to charge ( $m/z$ ) ratio and structure(s).

Sequence of Oligosaccharides	m/z	Relative %
1 Gal, GalNAcol	534	3.18
1 Gal, 1 Fuc, GalNAcol	708	8.51
1 Gal, 1 HexNAc, GalNAcol	779	19.99
1 Gal, 1 NeuAc, GalNAcol	895	0.55
2 Gal, 1 Fuc, GalNAcol	912	0.08
1 Gal, 1 HexNAc, 1 Fuc, GalNAcol	953	4.87
1 NeuGc, HexNAc, GalNAcol	967	0.09
2 Gal, 1 HexNAc, GalNAcol	983	1.06
1 Gal, 2 HexNAc, GalNAcol	1024	4.42
1 Gal, 1 Fuc, 1 NeuAc, GalNAcol	1069	0.14
1 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1140	0.91
2 Gal, 1 HexNAc, 1 Fuc, GalNAcol	1157	2.49
1 Gal, 2 HexNAc, 1 Fuc, GalNAcol	1198	1.14
2 Gal, 2 HexNAc, GalNAcol	1228	3.19
1 Gal, 2 NeuAc, GalNAcol	1256	0.10
1 Gal, 3 HexNAc, GalNAcol	1269	1.36
2 Gal, 1 HexNAc, 2 Fuc, GalNAcol	1331	6.11
2 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1344	0.12
3 Gal, 1 HexNAc, 1 Fuc, GalNAcol	1361	0.11
2 Gal, 2 HexNAc, 1 Fuc, GalNAcol	1402	14.14
3 Gal, 2 HexNAc, GalNAcol	1432	0.19
2 Gal, 3 HexNAc, GalNAcol	1473	14.38
2 Gal, 1 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	1518	0.11
3 Gal, 1 HexNAc, 2 Fuc, GalNAcol	1535	0.21
2 Gal, 2 HexNAc, 2 Fuc, GalNAcol	1576	1.00
3 Gal, 2 HexNAc, 1 Fuc, GalNAcol	1606	0.24
2 Gal, 3 HexNAc, 1 Fuc, GalNAcol	1647	5.94
3 Gal, 3 HexNAc, GalNAcol	1677	0.49
2 Gal, 4 HexNAc, GalNAcol	1718	0.14
3 Gal, 2 HexNAc, 2 Fuc, GalNAcol	1780	0.54
2 Gal, 3 HexNAc, 2 Fuc, GalNAcol	1821	0.22
3 Gal, 3 HexNAc, 1 Fuc, GalNAcol	1851	0.35
3 Gal, 4 HexNAc, GalNAcol	1922	0.51
3 Gal, 3 HexNAc, 2 Fuc, GalNAcol	2025	1.26
3 Gal, 4 HexNAc, 1 Fuc, GalNAcol	2096	0.58
4 Gal, 3 HexNAc, 3 Fuc, GalNAcol	2404	0.27
4 Gal, 4 HexNAc, 2 Fuc, GalNAcol	2475	0.26

**Table 12 | Relative abundance of oligosaccharides detected in pPGM by MALDI-TOF**

*The relative abundance of each glycan structure was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The structure corresponding to each m/z value is shown.*

Sequence of Oligosaccharides	m/z	Relative %
1 NeuAc, GalNAcol	691	2.15
1 Gal, 1 HexNAc, GalNAcol	779	1.94
1 Gal, 1 HexNAc, 1 SO <sub>3</sub> , GalNAcol	867	1.75
1 Gal, 1 NeuAc, GalNAcol	895	59.14
1 Gal, 1 NeuGc, GalNAcol	925	2.07
2 Gal, HexNAc, GalNAcol	983	1.64
2 Gal, 2 HexNAc, GalNAcol	1228	1.38
1 Gal, 2 NeuAc, GalNAcol	1256	22.61
1 Gal, 2 HexNAc, 1 Fuc, 1 SO <sub>3</sub> , GalNAcol	1286	1.01
2 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1344	1.46
1 Gal, 1 Fuc, 2 NeuAc, GalNAcol	1430	0.70
2 Gal, 3 HexNAc, GalNAcol	1473	0.71
3 Gal, 3 HexNAc, GalNAcol	1677	0.29
2 Gal, HexNAc, 2 NeuAc, GalNAcol	1705	2.88
2 Gal, 1 HexNAc, 1 Fuc, 2 NeuAc, GalNAcol	1879	0.27

**Table 13| Relative abundance of oligosaccharides detected by MALDI-TOF in the MUC2 mucin fraction purified from LS174T adenocarcinoma cell line**

*The relative abundance of each glycan structure was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The structure corresponding to each m/z value is shown.*

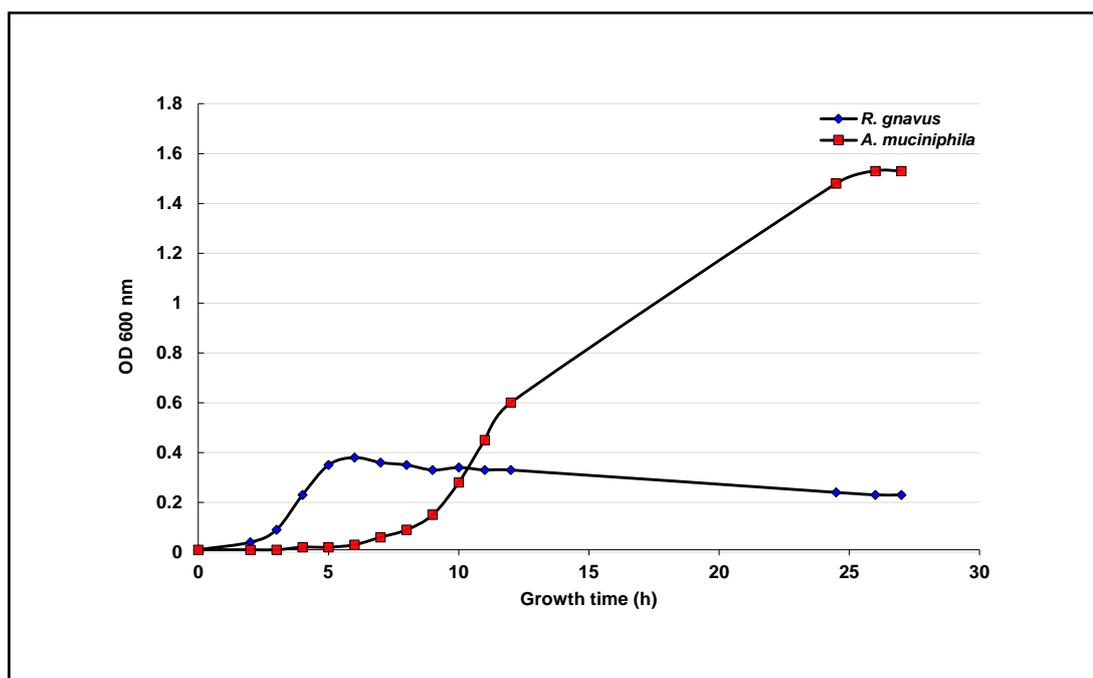
Sequence of Oligosaccharides	m/z	Relative %
1 Gal, GalNAcol	534	4.59
1 NeuAc, GalNAcol	691	4.70
1 Gal, 1 Fuc, GalNAcol	708	3.78
1 Gal, 1 HexNAc, GalNAcol	779	4.08
1 Gal, 1 NeuAc, GalNAcol	895	38.20
1 Gal, 1 NeuGc, GalNAcol	925	2.97
2 Gal, 1 HexNAc, GalNAcol	983	4.43
1 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1140	2.04
3 Gal, 1 HexNAc, GalNAcol	1187	0.75
2 Gal, 2 HexNAc, GalNAcol	1228	1.54
1 Gal, 2 NeuAc, GalNAcol	1256	9.76
1 Gal, 2 HexNAc, 1 Fuc, 1 SO <sub>3</sub> , GalNAcol	1286	0.80
2 Gal, 2 HexNAc, 1 SO <sub>3</sub> , GalNAcol	1316	0.73
2 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1344	3.50
2 Gal, 1 HexNAc, 1 NeuGc, GalNAcol	1374	0.78
3 Gal, 2 HexNAc, GalNAcol	1432	1.53
2 Gal, 3 HexNAc, GalNAcol	1473	0.95
2 Gal, 1 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	1518	1.26
3 Gal, 2 HexNAc, 1 Fuc, GalNAcol	1606	0.79
3 Gal, 3 HexNAc, GalNAcol	1677	1.78
2 Gal, 1 HexNAc, 2 NeuAc, GalNAcol	1705	2.81
3 Gal, 3 HexNAc, 1 SO <sub>3</sub> , GalNAcol	1766	0.97
3 Gal, 2 HexNAc, 1 NeuAc, GalNAcol	1793	0.75
3 Gal, 3 HexNAc, 1 Fuc, GalNAcol	1851	0.68
4 Gal, 3 HexNAc, GalNAcol	1881	1.89
3 Gal, 2 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	1967	0.85
4 Gal, 3 HexNAc, 1 Fuc, GalNAcol	2055	0.90
3 Gal, 3 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	2213	0.79
4 Gal, 3 HexNAc, 1 NeuAc, GalNAcol	2243	0.63
4 Gal, 3 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	2417	0.78

**Table 14 | Relative abundance of oligosaccharides detected by MALDI-TOF in the mixed mucin fraction purified from LS174T adenocarcinoma cell line**

*The relative abundance of each glycan structure was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The structure corresponding to each m/z value is shown.*

### 5.2.2.2 Initial characterisation of *A. muciniphila* and *R. gnavus* growth in pPGM

The growth pattern of *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C in pPGM was initially characterised by monitoring the OD change of both bacteria in YCFA basal media supplemented with the mucin over 27 h. In these conditions, *R. gnavus* had no apparent lag phase. Rather, a short exponential phase was observed during which the OD increased, reaching a maximum at 5-6 h. During the following 21 h, the OD gradually started to decrease again, indicating cell death. In contrast, following inoculation, *A. muciniphila* had a lag phase of around 5 h, following which an exponential phase was observed until around 25 h after which the OD appeared to start to plateau. Growth beyond 27 h was not monitored, however the overall  $\Delta$ OD was more than 4-fold greater than that of *R. gnavus* (Fig. 31).



**Figure 31 | Growth of *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835 in YCFA basal medium supplemented with 1 % pPGM**

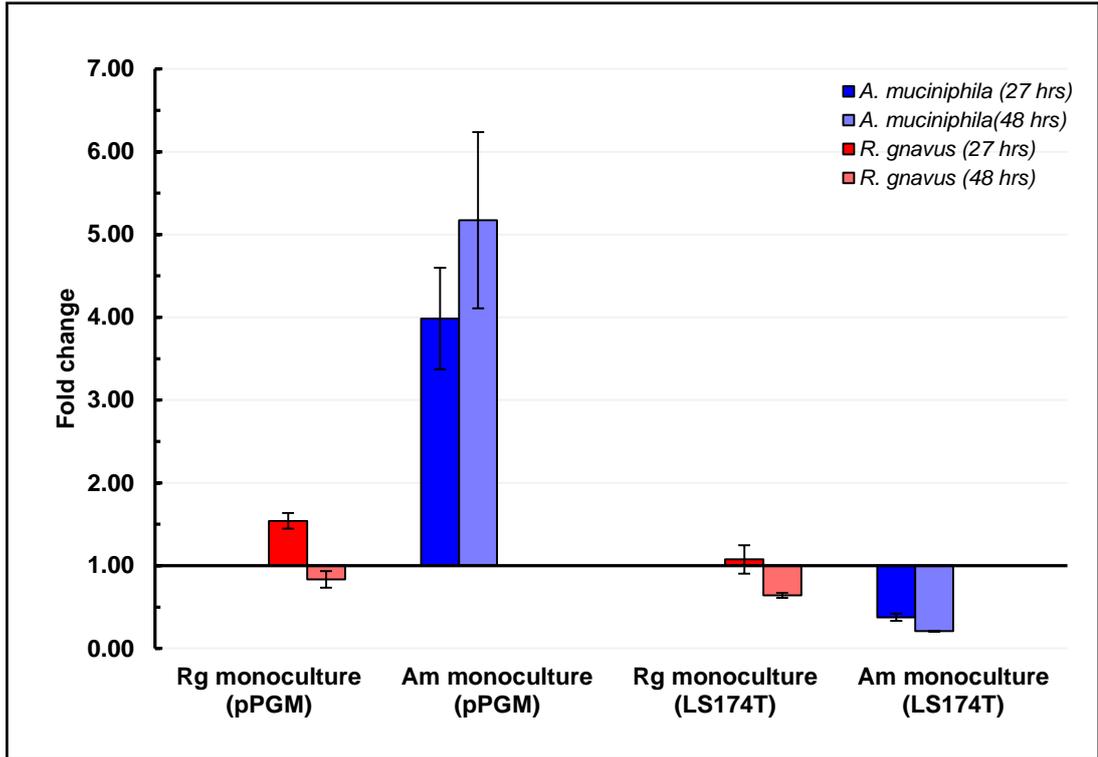
*pPGM* supplemented YCFA media was inoculated with *A. muciniphila* or *R. gnavus* cells from BHI-YH starter cultures. Growth was determined by measuring the change in optical density at a wavelength of 600 nm at various intervals over a 27 h period.

### 5.2.2.3 Growth assays of *R. gnavus* and *A. muciniphila* in pPGM and LS174T mucin

For further characterisation of *R. gnavus* and *A. muciniphila* growth in mucins containing difference glycosylation profiles, additional growth assays were performed using qPCR rather than OD to provide a more accurate quantification of cell growth. However, due to issues with aggregation of bacteria in YCFA basal media, creating difficulties in pelleting the bacteria by centrifugation, an alternative bicarbonate buffered basal media recipe (CP media) was used (Derrien et al., 2004).

*R. gnavus* and *A. muciniphila* were inoculated from starter cultures in BHI media into pPGM or LS174T mucin supplemented CP media and were cultured up to 48 h. Cells were enumerated in the inoculum and in the mucin cultures at 27 and 48 h post-inoculation by qPCR, using 16S sequence specific primers as described previously (See section 2.5.6). The assay was carried out in duplicate and technical replicates of each qPCR reaction were performed.

Each *A. muciniphila* mucin culture was inoculated with  $1.3 \times 10^8$  cells, whilst each *R. gnavus* mucin culture was inoculated with  $2.0 \times 10^8$  cells. When pPGM was used as a carbon source, *A. muciniphila* showed an average 4-fold increase in growth by 27 h and 5.2-fold increase by 48 h. In contrast, *R. gnavus* showed on average a 1.5-fold increase in growth by 27 h, but by 48 h cell numbers had reduced to 0.8-fold of the original inoculum. In the medium containing LS174T mucin as a sole carbon source, by 27 h, *A. muciniphila* cell numbers were decreased from the original inoculum with an average fold change of 0.38, and even more so at 48 h with a fold change of 0.21. In contrast, at 27 h *R. gnavus* cell numbers were sustained from the original inoculum, with an average fold change of 1.08. However, the number of *R. gnavus* cells decreased at 48 h with a fold change of 0.64 from the inoculum (Fig. 32).



**Figure 32 | Fold change of *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C in pPGM and LS174T mucins**

*pPGM* and *LS174T* mucin supplemented CP media were inoculated with *A. muciniphila* or *R. gnavus* cells from starter cultures in BHI-YH media. *A. muciniphila* assays were inoculated with  $1.3 \times 10^8$  cells whilst *R. gnavus* assays were inoculated with  $2.0 \times 10^8$  cells. The graph shows the fold change in cell numbers from the initial inoculum after 27 and 48 hrs of growth. Cell numbers in the initial inoculum and 27 and 48 hr assay samples were determined by qPCR quantification against a standard curve of known cell numbers. Assays were performed in duplicate.

Key; Rg, *R. gnavus*, Am, *A. muciniphila*

#### 5.2.2.4 Metabolite production and pathways utilised by *A. muciniphila* and *R. gnavus*

The utilisation and production of metabolites was measured in the spent media of *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C cultures grown on LS174T mucins or pPGM. The concentration of metabolites was measured by NMR after 48 h of growth and compared to that of the control media at the same time point, allowing metabolite production (mM) to be analysed.

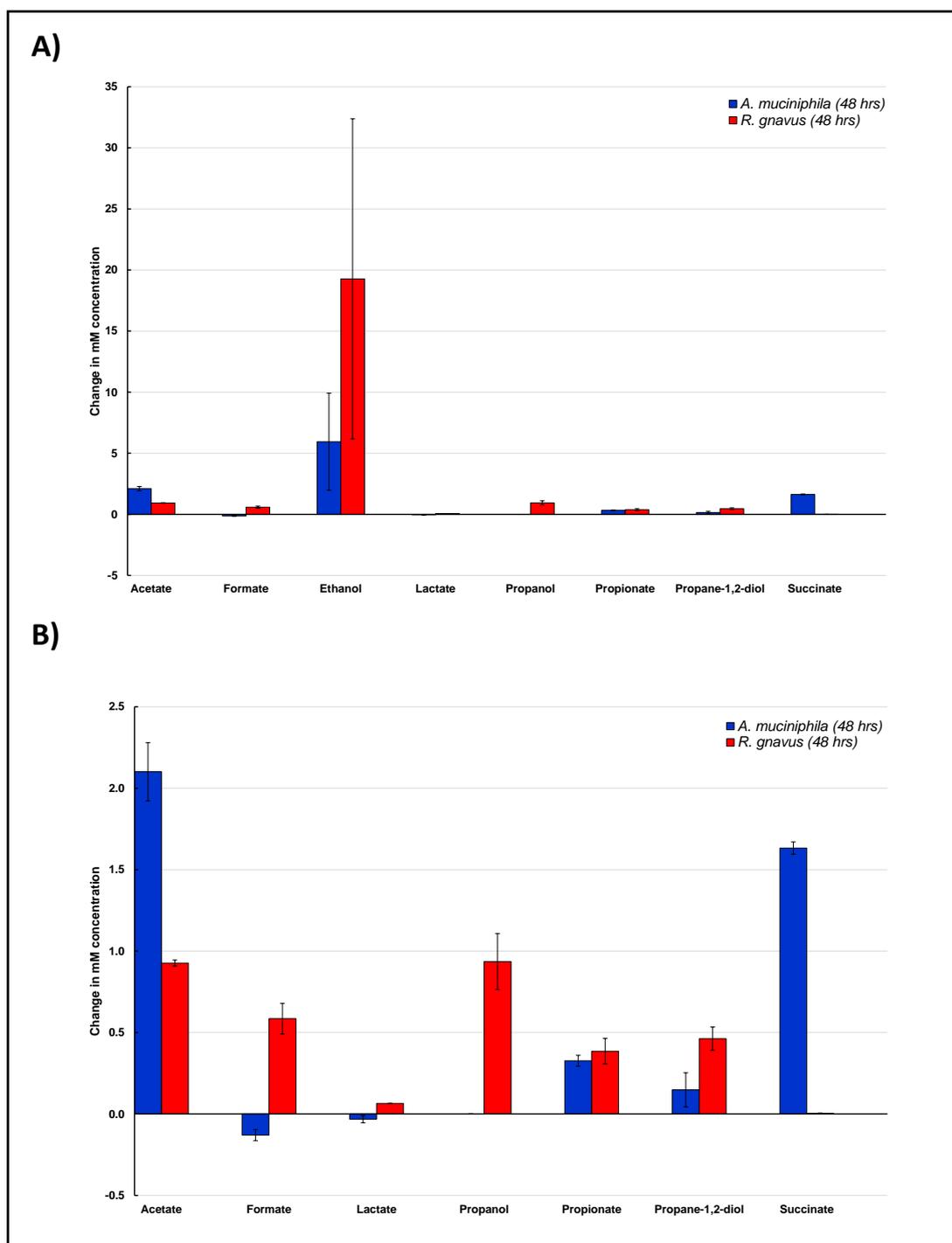
In pPGM supplemented medium, after 48 h, the main metabolite produced by both *R. gnavus* and *A. muciniphila* was ethanol (~20 mM by *R. gnavus* and ~6 mM by *A. muciniphila*). In addition, *R. gnavus* produced nearly 1 mM of both acetate and propanol, whilst formate, propionate and propane-1,2-diol were also produced, although at a lower amount (~0.6 mM, ~0.4 mM and ~0.5 mM, respectively). As well as ethanol, *A. muciniphila* also produced acetate (~2.1 mM), succinate (~1.6 mM), propionate (~0.3 mM) and propane-1,2-diol (~0.2 mM). Changes in concentration of metabolites following growth of bacteria, as compared to the control media are given in table 15 (Fig. 33).

In LS174T mucin, after 48 h, ethanol was still the major metabolite produced by *R. gnavus* (~23 mM), whilst acetate and formate were produced in similar levels to those observed in pPGM (~1.3 mM and ~0.8 mM, respectively). Propanol, propionate and propane-1,2-diol were produced in very low quantities, in contrast with growth in pPGM (~0.1 mM, ~0.03 mM, and ~0.04 mM, respectively). Production of metabolites by *A. muciniphila* in LS174T mucin was completely depleted, with only minimal amounts of acetate, formate and propionate produced, in stark contrast to growth in pPGM (Fig. 34) (Table 15).

	pPGM supplemented media		LS174T mucin supplemented media	
	<i>A. muciniphila</i>	<i>R. gnavus</i>	<i>A. muciniphila</i>	<i>R. gnavus</i>
Acetate	2.10	0.93	0.03	1.25
Formate	-0.13	0.59	0.08	0.78
Ethanol	5.95	19.27	-6.30	23.42
Lactate	-0.03	0.06	-0.01	0.08
Propanol	0.00	0.94	0.00	0.12
Propionate	0.33	0.39	0.05	0.03
Propane-1,2-diol	0.15	0.46	0.00	0.04
Succinate	1.63	0.00	0.01	0.02

**Table 15 | Change in mM concentration of metabolites after 48 hours growth in pPGM and LS174T mucin supplemented medium**

*Metabolites were quantified using <sup>1</sup>H NMR. The table shows the main metabolites with their change in mM concentration in mucin supplemented CP media inoculated with *R. gnavus* or *A. muciniphila*, as compared to media without inoculation.*

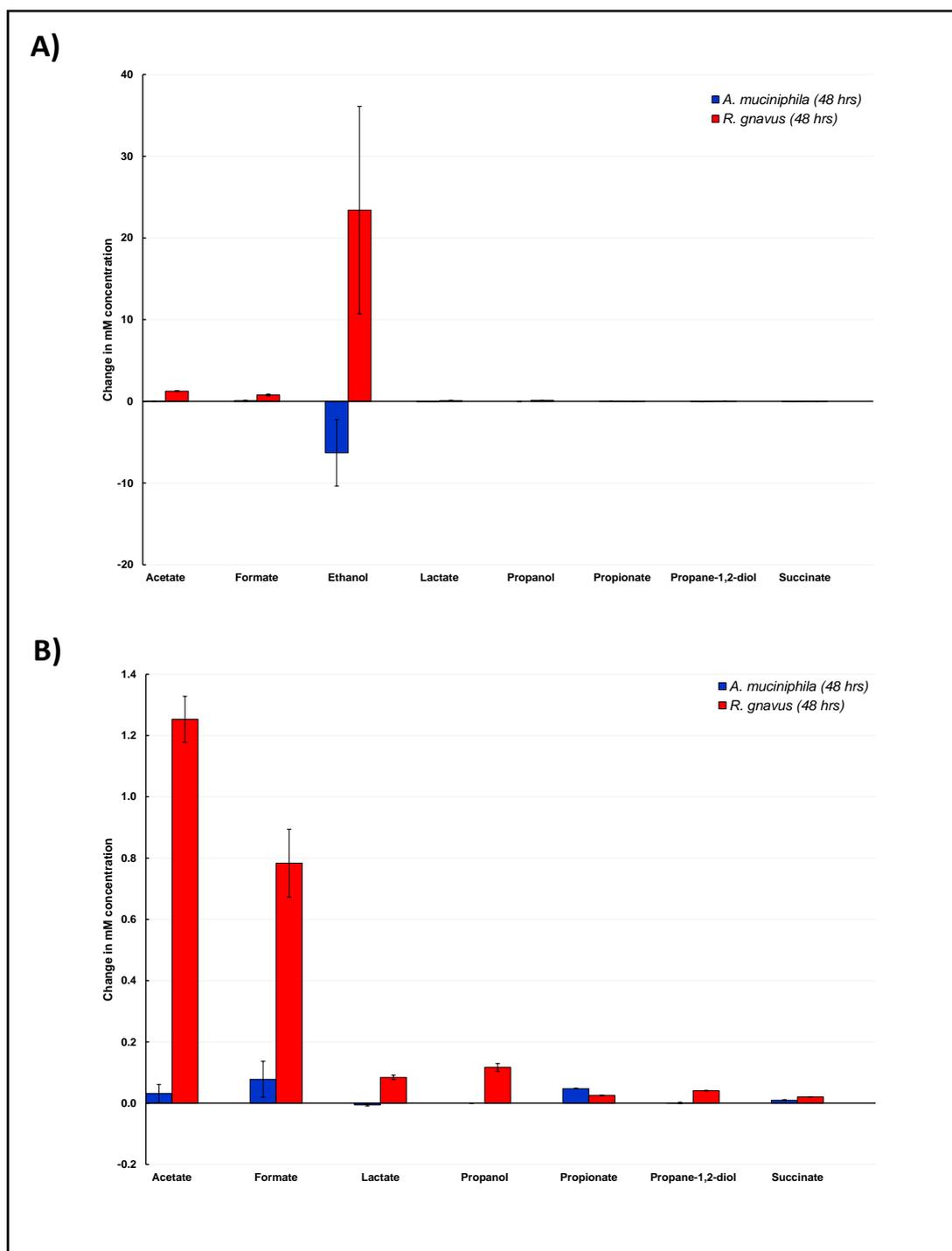


**Figure 33| Change in mM concentration of metabolites by *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C grown in pPGM supplemented CP media**

*A) Change in mM concentration of all metabolites detected*

*B) y axis scaled up to view fold changes of acetate, formate, lactate, propanol, propionate, propane-1,2-diol and succinate*

*pPGM supplemented CP media was inoculated with *A. muciniphila* or *R. gnavus* cells from starter cultures in BHI-YH media. Graphs show the change in mM concentration of metabolites in inoculated CP cultures after 48 hrs, as compared to media without inoculation. Metabolites were quantified using  $^1\text{H}$  NMR. The metabolites shown are the main metabolites detected in the spent media. Assays were performed in duplicate.*



**Figure 34 | Change in mM concentration of metabolites by *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C grown in LS174T mucin supplemented CP media**

*A) Change in mM concentration of all metabolites detected*

*B) y axis scaled up to view fold changes of acetate, formate, lactate, propanol, propionate, propane-1,2-diol and succinate*

*LS174T mucin supplemented CP media was inoculated with *A. muciniphila* or *R. gnavus* cells from starter cultures in BHI-YH media. Graphs show the change in mM concentration of metabolites in inoculated CP cultures after 48 hrs, as compared to media without inoculation. Metabolites were quantified using <sup>1</sup>H NMR. The metabolites shown are the main metabolites detected in the spent media. Assays were performed in duplicate.*

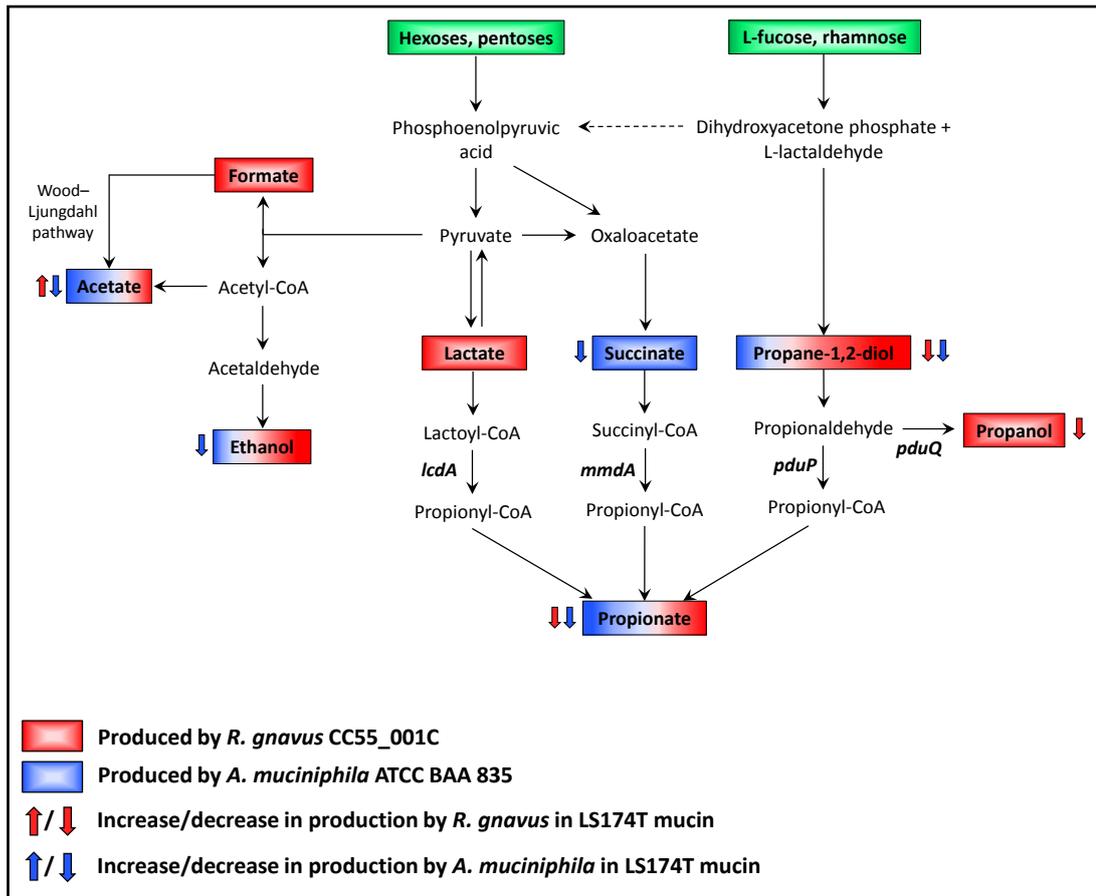
To help identify the metabolic pathways of SCFA production being carried out by *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835, a blastp search was performed against deduced sequences of genes diagnostic of these pathways in other organisms, as described by Reichardt and coll. (Reichardt et al., 2014). With regards to propionate production, *A. muciniphila* ATCC BAA 835 possessed a gene with 52% homology to that of *mmdA*; an enzyme found in the succinate pathway of *Veillonella parvula* and Bacteroidetes (Table 16). *R. gnavus* CC55\_001C also possessed a gene encoding a protein with 39% homology to *mmdA*, as well as another predicted protein with 77% homology to *PduP*; a propionaldehyde dehydrogenase found in the propanediol pathway of related species *Ruminococcus obeum* and *Roseburia inulinivorans* (Table 16). Furthermore, both organisms possess a gene encoding a protein homologous (*A. muciniphila*; 58% and *R. gnavus*; 81%) to *PduQ*; a propanol dehydrogenase, involved in the by-production of propanol in the propanediol pathway (Table 16). This predicted protein is also found in *R. inulinivorans* ABC25529 (Reichardt et al., 2014). These enzymes and the metabolites produced by *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835, as well as the enzymatic functions of these proteins are shown in the context of their metabolic pathways in figure 35.

		Protein sequence identity (Query coverage)	
		<i>Akkermansia muciniphila</i> ATCC BAA 835	<i>Ruminococcus gnavus</i> CC55_001C
<b>Succinate pathway</b>	<i>mmdA</i>	<b>52%</b> (100%)	<b>39%</b> (92%)
<b>Acrylate pathway</b>	<i>LcdA</i>	<b>None</b>	<b>None</b>
<b>Propanediol pathway</b>	<i>PduP</i>	<b>None</b>	<b>77%</b> (100%)
	<i>PduQ</i>	<b>58%</b> (90%)	<b>81%</b> (100%)

**Table 16| Occurrence of predicted enzymes involved in pathways of propionate metabolism in *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C**

*A blastp search was performed on the A. muciniphila and R. gnavus genomes against deduced sequences of genes diagnostic of the succinate, acrylate and propanediol pathways in other organisms (Reichardt et al., 2014). % protein identity is given for all matches >30% with >80% query coverage*

*MmdA, Methylmalonyl-CoA decarboxylase  $\alpha$ -subunit, LcdA, Lactoyl-CoA dehydratase  $\alpha$ -subunit, PduP, CoA-dependent propionaldehyde dehydrogenase, PduP, Propanol dehydrogenase*



**Figure 35 | Metabolites produced by *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835 at 48 hours in the context of their metabolic pathways**

*Production in pPGM at 48 hours is shown by coloured boxes. Alterations in metabolite production in LS174T mucin at 48 hours is shown by adjacent arrows. Predicted enzymes involved in these pathways are shown in italics.*

*R. gnavus and A. muciniphila both produce a number of metabolites, including SCFAs propionate and acetate when grown in pPGM. In LS174T mucin, R. gnavus no longer appears to utilise the propanediol pathway (propionate production via propane-1,2-diol) whilst A. muciniphila metabolite production is almost completely abolished.*

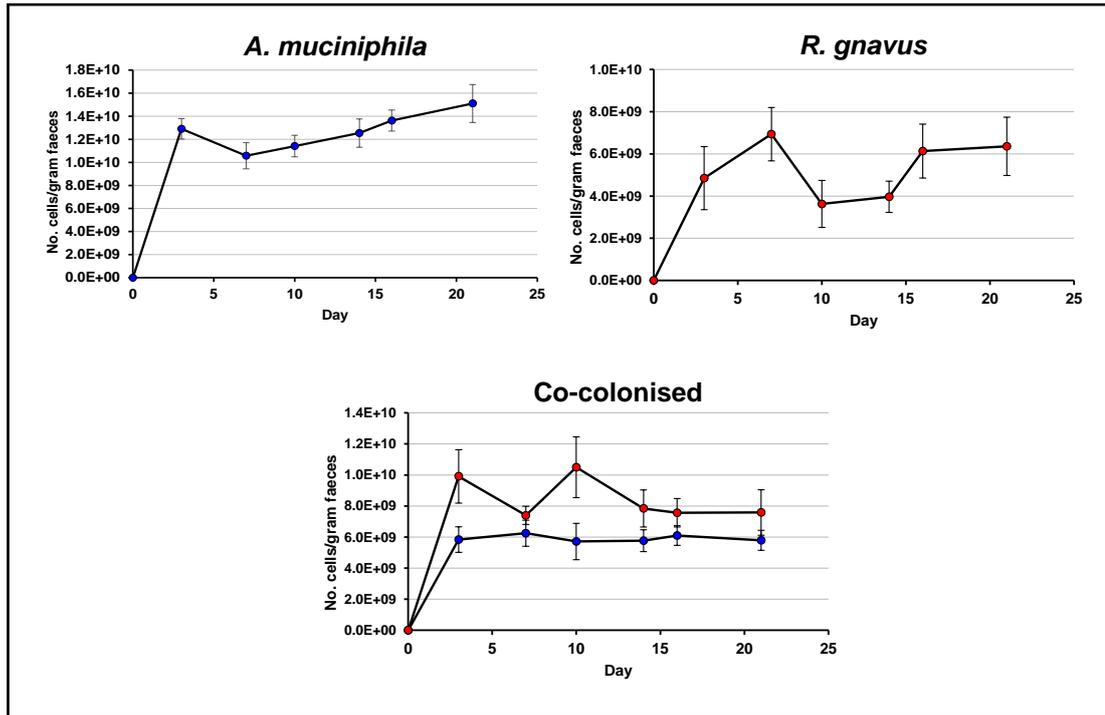
### 5.2.3 Colonisation of germ-free mice with *R. gnavus* and *A. muciniphila*

To establish if the mucin degrading strains of *R. gnavus* or *A. muciniphila* could alter the glycosylation profile of mucins, *in vivo* colonisation experiments were carried out in germ-free mice. Mice were mono-colonised with either *R. gnavus* CC55\_001C, *A. muciniphila* ATCC BAA 835, co-colonised with both, or were kept germ-free, allowing the glycosylation profile of each group to be compared to a 'baseline' profile. Each group was comprised of 5 female C57BL/6J mice between the ages of 11-14 weeks. Mice were fed a standard sterilised diet.

Mice were gavaged with bacterial cells from cultures grown in BHI-YH, enumerated using qPCR with 16S specific bacterial primers. Each mouse received a 100  $\mu$ l dose intragastrically corresponding to  $5.79 \times 10^8$  cells for *R. gnavus* mono-associated mice,  $6.01 \times 10^9$  cells for *A. muciniphila* mono-associated mice, and  $3.85 \times 10^8$  *R. gnavus* cells plus  $3.31 \times 10^8$  *A. muciniphila* cells for the co-colonised mice.

Mice were kept in gnotobiotic conditions for 21 days and faecal samples collected every 3 or 4 days to carry out gDNA extractions. Between days 0 and 3, the concentration of DNA recovered per gram of faeces increased dramatically for both mono- and co-associated mice. For *R. gnavus* mono-associated mice, there was a 21-fold increase, for *A. muciniphila* mono-associated mice, there was a 20-fold increase, and for co-associated mice, there was a 31-fold increase in gDNA concentration. In contrast, there was no change in the concentration of gDNA recovered from the faeces of the germ-free control group. For the rest of the study, the gDNA concentration recovered from faecal samples remained stable in all groups.

The number of bacterial cells colonising mice was determined by qPCR using 16S group specific primers. At least two technical replicates were performed for each qPCR reaction. Ct values were compared against a standard curve with known numbers of cells, based on the theoretical molecular weight of each bacteria. Cell enumeration showed that both *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835 could rapidly and stably colonise the germ-free GI tract. In line with the increasing gDNA concentrations, the number of cells in each of the mono- and co-associated mouse groups increased rapidly between days 0 and 3, and remained quite stable throughout the rest of the experiment (Fig. 36). *A. muciniphila* mono-associated mice were colonised to a mean density of  $1.3 \times 10^{10}$  cells per gram of faecal sample. *R. gnavus* mono-associated mice were colonised to a mean density of around  $5.3 \times 10^9$  cells per gram of faecal sample. In the co-associated group, mice were colonised with a mean density of  $5.9 \times 10^9$  *A. muciniphila* cells and  $8.5 \times 10^9$  *R. gnavus* cells per gram of faecal sample.



**Figure 36| Mono- and co-colonisation of germ-free C57BL/6J mice with *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C**

Germ-free mice were gavaged with *A. muciniphila*, *R. gnavus* or a combination of both bacteria, initially grown in BHI-YH media. Each mouse received an intragastric dose of  $5.79 \times 10^8$  cells for *R. gnavus* mono-associated mice,  $6.01 \times 10^9$  cells for *A. muciniphila* mono-associated mice, and  $3.85 \times 10^8$  *R. gnavus* cells plus  $3.31 \times 10^8$  *A. muciniphila* cells for the co-colonised mice. Colonisation was determined by measuring the number of bacterial cells recovered in faecal samples from the mice at intervals following mono- and co-colonisation. The number of bacterial cells recovered per gram of faeces at intervals following gavage is shown. Cell numbers in the initial gavage and in faecal samples during the experiment were determined by qPCR quantification against a standard curve of known cell numbers.

Key; red circles, number of *R. gnavus* cells, blue circles, number of *A. muciniphila* cells

### 5.2.3.1 Glycomic profiling of mucins from mice mono- and co-colonised with *A. muciniphila* and *R. gnavus* by mass spectrometry

To assess for differences in mucin glycosylation between controls and mice mono- and co-colonised, scrapings of mucus were taken from the colon of mice sacrificed on day 21 of the experiment. Mucus from 5 mice in each group were pooled and the mucins purified by caesium chloride ultracentrifugation, as reported in section 2.5.4. The glycans were released by  $\beta$ -elimination and analysed by MS, as in section 2.4.7. The relative abundance of each glycan structure was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all main structure peaks in the spectra (Fig. 37). Alterations in glycosylation of mucins from mice mono-colonised with *R. gnavus* CC55\_001C, *A. muciniphila* ATCC BAA 835 or co-colonised with both was determined by comparing the glycosylation profile of the pooled mucins of each group to that of a pooled sample from the 5 control germ-free mice.

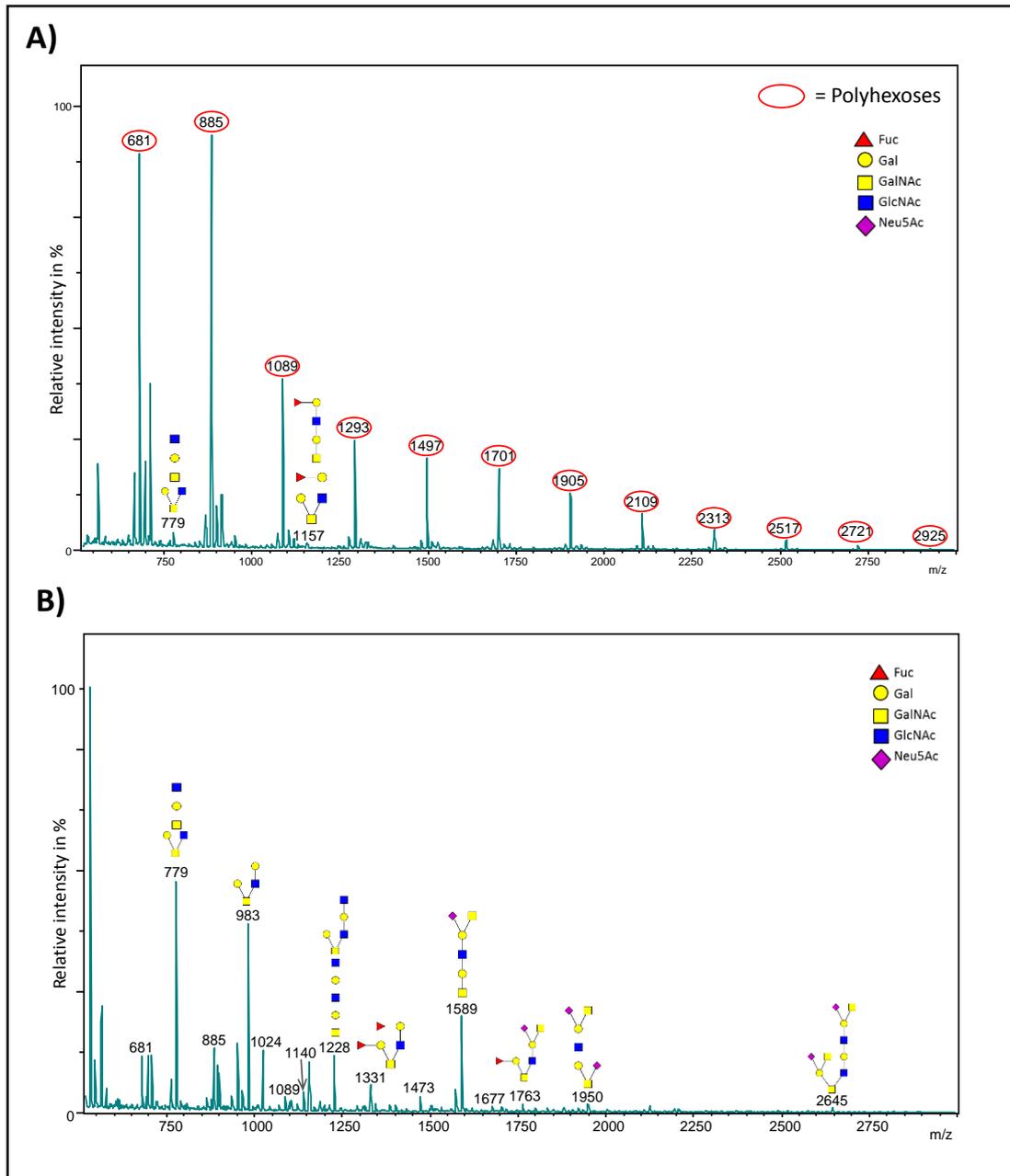
Two separate mucin fractions were analysed from the colon, corresponding to gel-forming/secreted Muc2, and other mixed mucins. Overall, in control mice, a greater number of structures were found in the colon mixed mucin fraction (34 structures) when compared with the Muc2 fraction (15 structures) (See appendix 3). The predominant structures in both mixed mucin and Muc2 fractions included the sialyl-TF antigen with an m/z of 895, and the structure m/z 1589, comprising two potential mono-sialylated core 2 isomers (Fig. 38). Interestingly, however, a higher percentage of structures on mixed mucins, as compared to Muc2, were sialylated (mixed mucins, 75.22%, Muc2, 49.30%) (Fig. 39), whilst the reverse was true for fucosylation (mixed mucins, 8.97%, Muc2, 35.73%) (Fig. 40).

A comparison between mice mono-colonised with *R. gnavus* CC55\_001C, *A. muciniphila* ATCC BAA 835 or co-colonised with both strains revealed a number of differences in the glycosylation of Muc2 and mixed mucin fractions between the groups. In the Muc2 mucin fraction, the overall number of detected structures increased following colonisation with *R. gnavus* CC55\_001C (23 structures) or with a mixture of *R. gnavus* and *A. muciniphila* ATCC BAA 835 (22 structures), as compared to germ-free mice. In contrast, the overall number of structures was decreased on mixed mucins from these groups (*R. gnavus* mono-colonised, 29 structures, co-colonised, 18 structures). Glycosylation of both Muc2 and mixed mucins in mice mono-colonised with *A. muciniphila* ATCC BAA 835 could not be determined, due to these samples seemingly containing a large amount of polyhexose contaminants (Fig. 37).

The overall percentage of sialylated structures in the Muc2 fraction was reduced when mice were mono-colonised with *R. gnavus* or co-colonised with both bacteria as compared to controls (GF control, 49.30%, *R. gnavus* mono-colonised, 25.60%, co-colonised, 26.57%) (Fig. 39). This trend was also observed in the mixed mucin fraction (germ-free control, 75.22%, *R. gnavus* mono-colonised, 59.21%, co-colonised, 33.66%) (Fig. 39). Furthermore, in the Muc2 fraction, the percentage of fucosylated structures was lower in mice mono-colonised with *R. gnavus* (21.20%) but slightly increased in the co-colonised mice (38.96%) as compared to germ-free controls (35.73%) (Fig. 40). In the mixed mucin fraction, the same trend was apparent, however the decrease of fucosylation in mono-colonised mice was more subtle (7.73%), whilst the increase in fucosylation in co-colonised mice was more pronounced (18.82%), as compared to germ-free controls (8.97%) (Fig. 40).

There were also some obvious differences in the levels of individual glycan structures between germ-free and colonised mice (Fig. 38). In the Muc2 fractions, the abundance of the TF antigen (m/z 534) in *R. gnavus* colonised mice was almost 5-fold higher than in germ-free mice, and 2-fold higher in co-colonised mice (Fig. 38) (See table 17 for structures). Although not as pronounced, an increase of TF antigen was also evident in the mixed mucin population. A depletion of sialyl-TF antigen was apparent in *R. gnavus* and co-colonised mice in both mucin fractions (Fig. 38). In contrast, the abundance of structure m/z 1589 was differentially altered in the mucin fractions of *R. gnavus* colonised mice; decreased in Muc2 but increased in mixed mucins, whilst it was depleted in both fractions of co-colonised mice when compared with germ-free mice (Fig. 38).

Together these findings indicate that colonisation of germ-free mice with either *R. gnavus*, *A. muciniphila* or both results in remodelling of the mucin glycan profile.

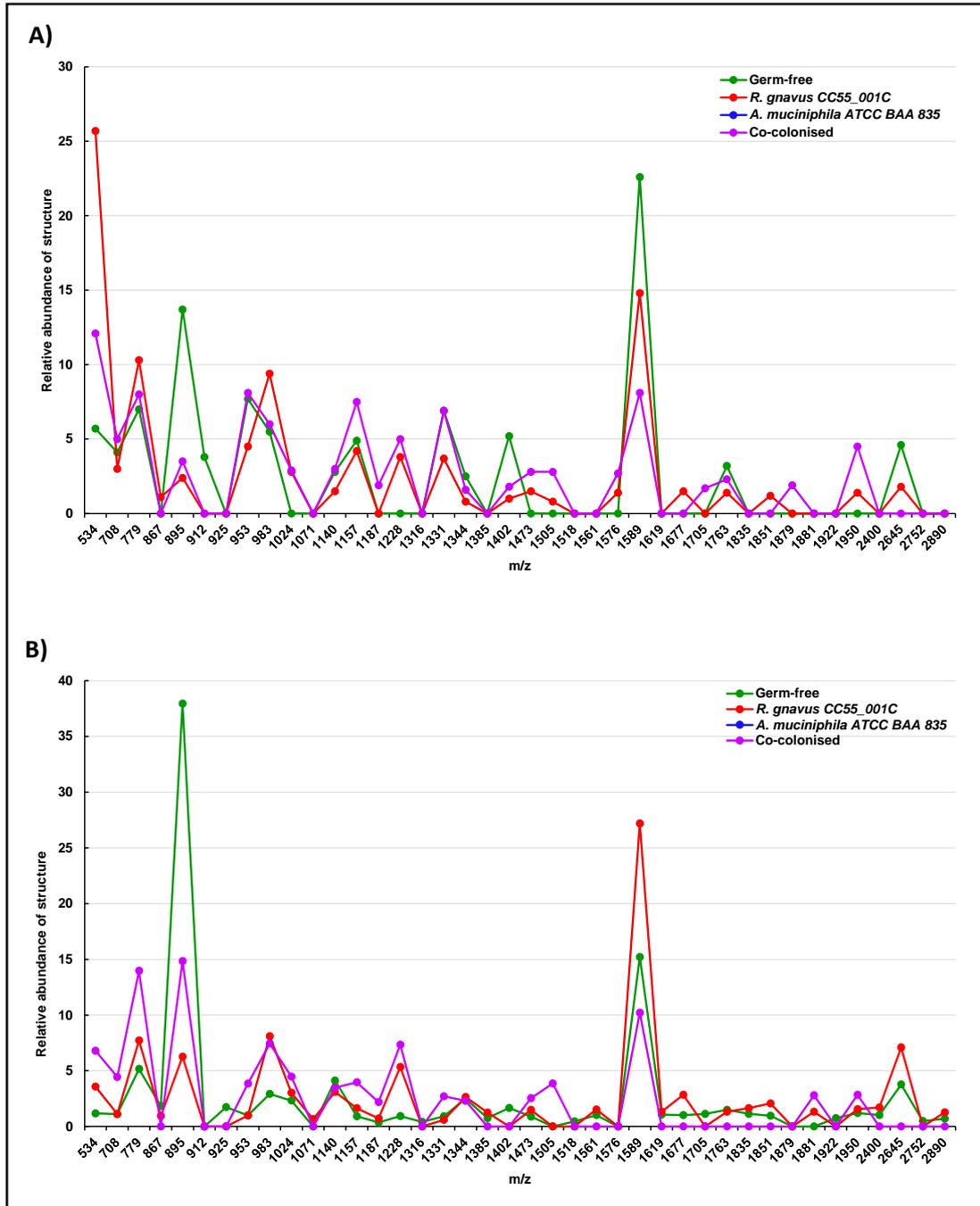


**Figure 37 | Example MALDI-TOF spectra of colonic mucin glycans from mice**

A) MUC2 mucin glycosylation of *A. muciniphila* ATCC BAA 835 mono-colonised mice

B) MUC2 mucin glycosylation of *R. gnavus* CC55\_001C mono-colonised mice

Glycans were liberated from purified mucins by reductive  $\beta$ -elimination, and analysed by MALDI-TOF using a DHB matrix. The relative abundance of each glycan structure was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all main structure peaks in the spectra. The most abundant glycan peaks are annotated with mass to charge ( $m/z$ ) ratio and structure(s). A) shows a typical spectra obtained from *A. muciniphila* mono-colonised mice where polyhexose contaminants are evident, whilst B) shows a typical spectra obtained from *R. gnavus* mono-colonised mice where this contamination is not apparent (as with co-colonised mice).



**Figure 38| Glycosylation profile in the colon of germ-free mice, and mice mono- and co-colonised with *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835**

A) Glycan profile in Muc2 mucin fraction

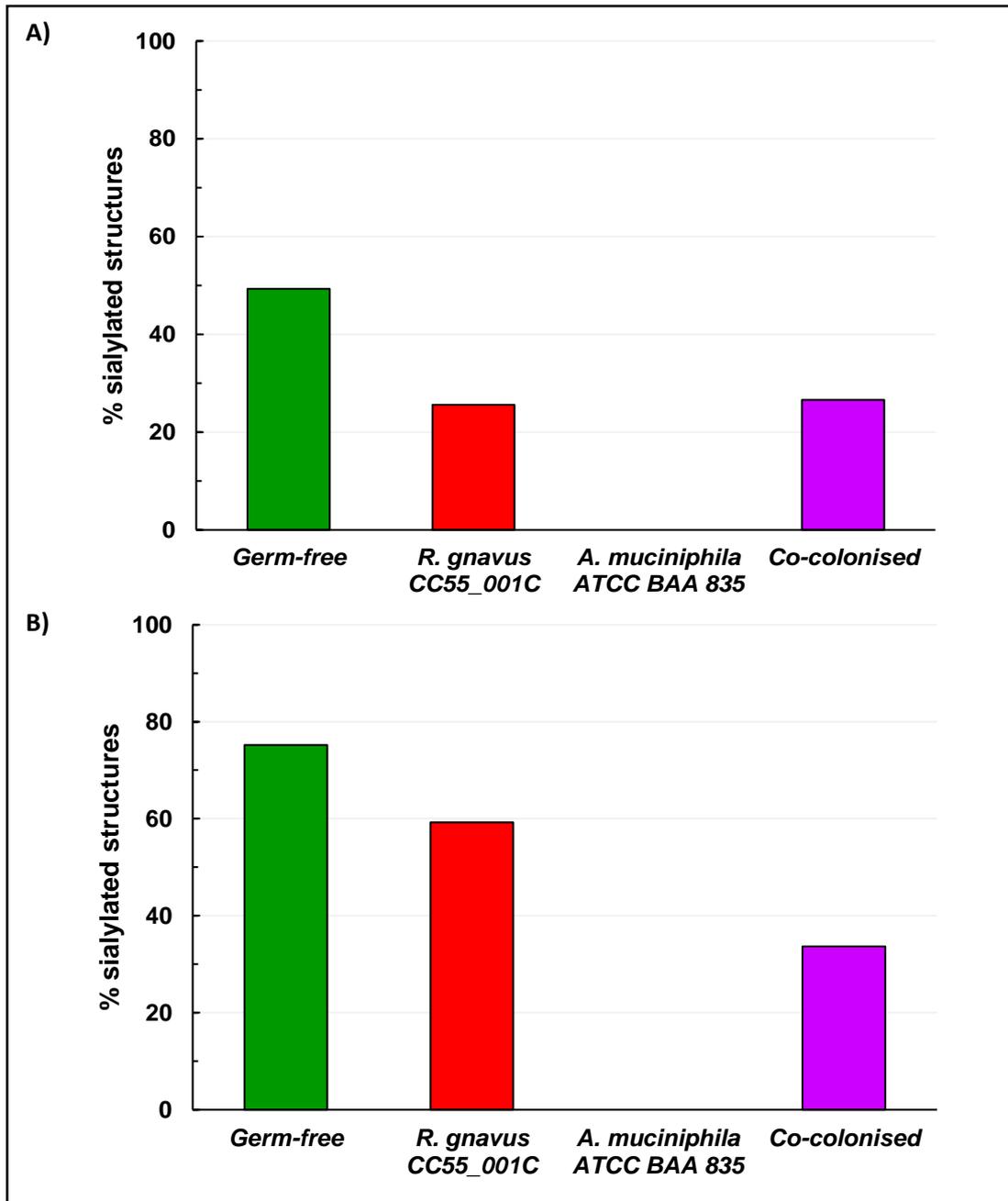
B) Glycan profile in mixed mucin fraction

Mucus was harvested from the colon of mice mono- or co-colonised with *A. muciniphila* and/or *R. gnavus* for 21 days. Mucus from 5 mice in each colonisation group was pooled. The mucins were purified, separating them into two mucin fractions (Muc2 and mixed mucins). Glycans were liberated by reductive  $\beta$ -elimination, and analysed by MALDI-TOF using a DHB matrix. The relative abundance of glycan structures from mouse samples was determined by measuring the area of the main peak corresponding to the mass of the glycan and comparing it to the total area of all peaks in the spectra. The graphs show the glycan profile in the two different mucin fractions from 4 mouse groups.

Structure	m/z
1 Gal, GalNAcol	534
1 Gal, 1 Fuc, GalNAcol	708 (F)
1 Gal, 1 HexNAc, GalNAcol	779
1 Gal, 1 HexNAc, 1 SO <sub>3</sub> , GalNAcol	867
1 Gal, 1 NeuAc, GalNAcol	895 (S)
2 Gal, 1 Fuc, GalNAcol	912 (F)
1 Gal, 1 NeuGc, GalNAcol	925 (S)
1 Gal, 1 HexNAc, 1 Fuc, GalNAcol	953 (F)
2 Gal, 1 HexNAc, GalNAcol	983
1 Gal, 2 HexNAc, GalNAcol	1024
2 Gal, 1 HexNAc, 1 SO <sub>3</sub> , GalNAcol	1071
1 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1140 (S)
2 Gal, 1 HexNAc, 1 Fuc, GalNAcol	1157 (F)
3 Gal, 1 HexNAc, GalNAcol	1187
2 Gal, 2 HexNAc, GalNAcol	1228
2 Gal, 2 HexNAc, 1 SO <sub>3</sub> , GalNAcol	1316
2 Gal, 1 HexNAc, 2 Fuc, GalNAcol	1331 (F)
2 Gal, 1 HexNAc, 1 NeuAc, GalNAcol	1344 (S)
1 Gal, 2 HexNAc, 1 NeuAc, GalNAcol	1385 (S)
2 Gal, 2 HexNAc, 1 Fuc, GalNAcol	1402 (F)
2 Gal, 3 HexNAc, GalNAcol	1473 (F)
2 Gal, 1 HexNAc, 3 Fuc, GalNAcol	1505 (F)
2 Gal, 1 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	1518 (F)(S)
2 Gal, 3 HexNAc, 1 SO <sub>3</sub> , GalNAcol	1561
2 Gal, 2 HexNAc, 2 Fuc, GalNAcol	1576 (F)
2 Gal, 2 HexNAc, 1 NeuAc, GalNAcol	1589 (S)
2 Gal, 2 HexNAc, 1 NeuGc, GalNAcol	1619 (S)
3 Gal, 3 HexNAc, GalNAcol/2 Gal, 2 HexNAc, 1 NeuAc, 1 SO <sub>3</sub> , GalNAcol	1677 (S)
2 Gal, 1 HexNAc, 2 NeuAc, GalNAcol	1705 (S)
2 Gal, 2 HexNAc, 1 Fuc, 1 NeuAc, GalNAcol	1763 (F)(S)
2 Gal, 3 HexNAc, 1 NeuAc, GalNAcol	1835 (S)
3 Gal, 3 HexNAc, 1 Fuc, GalNAcol	1851 (F)
2 Gal, 1 HexNAc, 1 Fuc, 2 NeuAc, GalNAcol	1879 (F)(S)
4 Gal, 3 HexNAc, GalNAcol	1881
3 Gal, 4 HexNAc, GalNAcol	1922
2 Gal, 2 HexNAc, 2 NeuAc, GalNAcol	1950 (S)
3 Gal, 3 HexNAc, 2 NeuAc, GalNAcol	2400 (S)
3 Gal, 4 HexNAc, 2 NeuAc, GalNAcol	2645 (S)
4 Gal, 3 HexNAc, 5 Fuc, GalNAcol	2752 (F)
3 Gal, 5 HexNAc, 2 NeuAc, GalNAcol	2890 (S)

**Table 17 | Table of glycan structures identified in mucins of mice mono- and co-colonised with *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C**

Structures corresponding to m/z values are shown in the table. Full data including relative abundances of glycans structures in each mouse group can be found in appendix 3

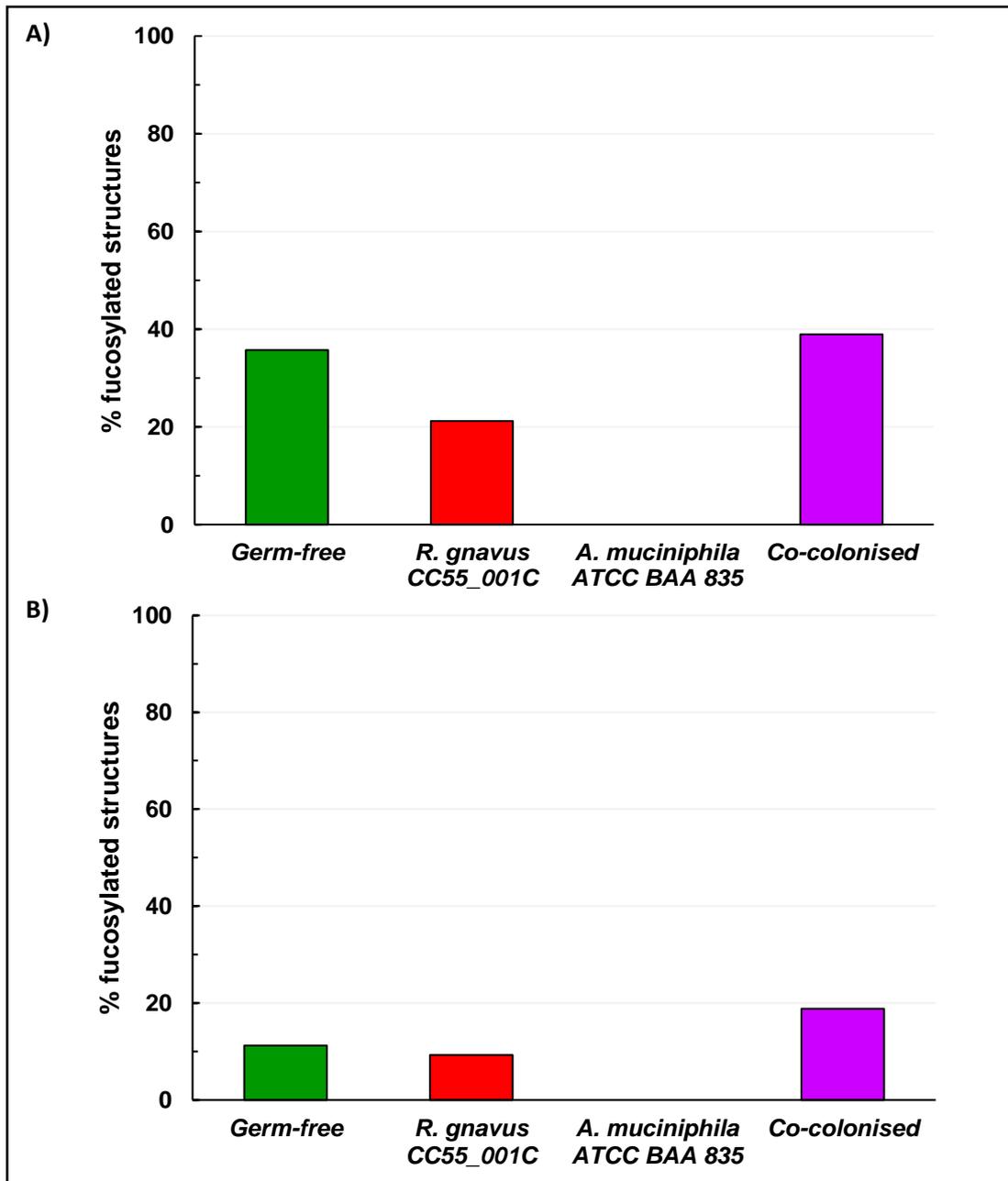


**Figure 39|** Relative abundance of sialylated structures in colonic mucins from germ-free mice, and mice mono- and co- colonised with *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835

A) Abundance of sialylated structures in Muc2 mucin fraction

B) Abundance of sialylated structures in mixed mucin fraction

Colonic mucus was harvested from mice mono- or co-colonised with *A. muciniphila* and/or *R. gnavus* for 21 days. Mucus from 5 mice in each colonisation group was pooled. The mucins were purified, separating them into two mucin fractions (Muc2 and mixed mucins). Glycans were liberated by reductive  $\beta$ -elimination, and analysed by MALDI-TOF using a DHB matrix. The relative abundance of sialylated glycan structures from mouse samples was determined by measuring the area of the main peaks corresponding to the mass of sialylated glycans and comparing it to the total area of all peaks in the spectra. The graphs show the relative abundance of sialylated structures in the two different mucin fractions from 4 mouse groups.



**Figure 40|** Relative abundance of fucosylated structures in colonic mucins from germ-free mice, and mice mono- and co- colonised with *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835

A) Abundance of fucosylated structures in Muc2 mucin fraction

B) Abundance of fucosylated structures in mixed mucin fraction

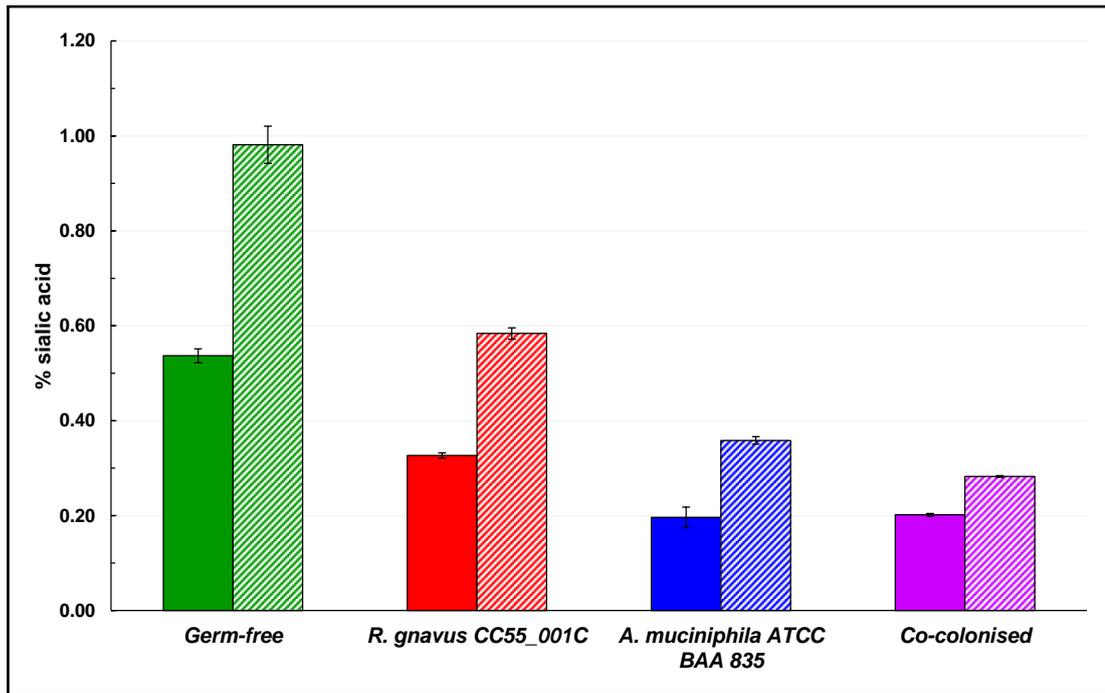
Colonic mucus was harvested from mice mono- or co-colonised with *A. muciniphila* and/or *R. gnavus* for 21 days. Mucus from 5 mice in each colonisation group was pooled. The mucins were purified, separating them into two mucin fractions (Muc2 and mixed mucins). Glycans were liberated by reductive  $\beta$ -elimination, and analysed by MALDI-TOF using a DHB matrix. The relative abundance of fucosylated glycan structures from mouse samples was determined by measuring the area of the main peaks corresponding to the mass of fucosylated glycans and comparing it to the total area of all peaks in the spectra. The graphs show the relative abundance of fucosylated structures in the two different mucin fractions from 4 mouse groups.

#### 5.2.3.2 Quantification of mucin sialylation from mice mono- and co-colonised with *A. muciniphila* and *R. gnavus* by HPAEC-PAD

To quantify the level of sialic acid in mucins purified from colonic mucus scrapings of mice mono- and co-colonised with *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C, pooled samples of mucins purified from the 5 mice in each group were analysed in triplicate. HPAEC-PAD was performed following acid hydrolysis of both the Muc2 and mixed mucin fractions, as described in section 2.4.6.

In general, as observed by MS, mixed mucins from germ-free mice were more highly sialylated (0.98% sialic acid) than those in the Muc2 fraction (0.54% sialic acid) (Fig. 41).

A comparison between mice mono-colonised with *R. gnavus* CC55\_001C, *A. muciniphila* ATCC BAA 835 or co-colonised with both revealed that the degree of mucin sialylation was reduced in all colonised mice as compared to germ-free controls (Fig. 41). In mice mono-colonised with *R. gnavus*, sialylation was reduced to 0.33% in Muc2 (compared to 0.54% in controls) and 0.58% in mixed mucins (compared to 0.98% in controls). This trend was also apparent in mucins purified from *A. muciniphila* mono-colonised and co-colonised mice, and was also more pronounced in these mice, compared to *R. gnavus* mice (Muc2; *A. muciniphila* mono-colonised, 0.20%, co-colonised, 0.20%, mixed mucins; *A. muciniphila* mono-colonised, 0.36%, co-colonised, 0.28%) (Fig. 41), suggesting that colonisation with these bacteria results in the degradation/release of sialic acid from mucins.



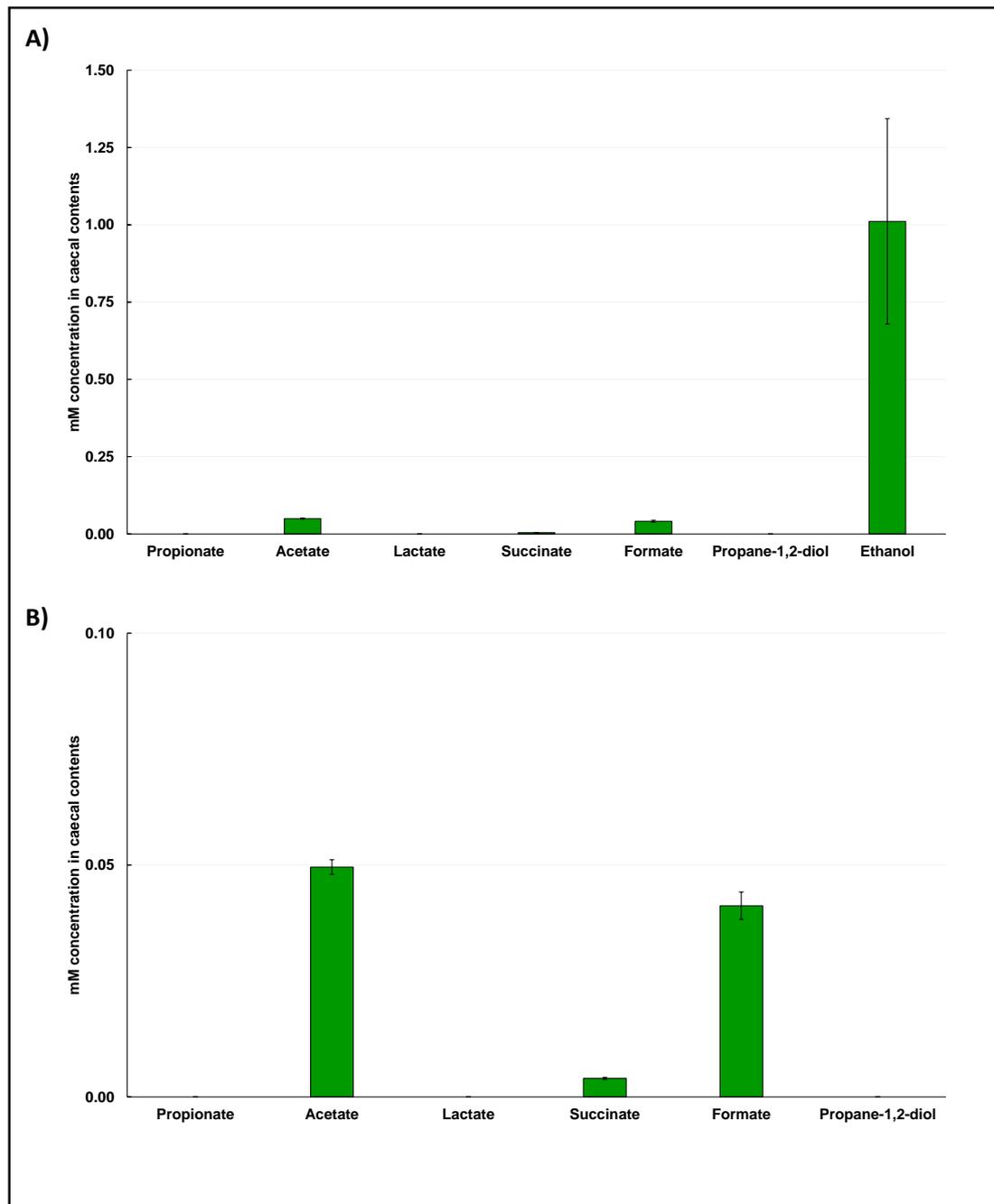
**Figure 41 | Abundance of sialic acid as determined by HPAEC-PAD in colonic mucins from mice mono- and co- colonised with *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C**

*Mucus was harvested from the colon of mice mono- or co-colonised with A. muciniphila and/or R. gnavus for 21 days. Mucus from 5 mice in each colonisation group was pooled. The mucins were purified, separating them into two mucin fractions (Muc2 and mixed mucins). Sialic acid was liberated from purified mucins using mild acid hydrolysis. HPAEC-PAD analysis of each sample was performed in triplicate. The graph shows the relative abundance of sialic acid in the two different mucin fractions from 4 mouse groups. The standard error is shown for each group.*

*Key; Block colours, sialylation of Muc2 fractions, hashed colours, sialylation of mixed mucin fractions*

#### 5.2.3.3 *Metabolite production in the caecal contents of mice mono- and co-colonised with A. muciniphila and R. gnavus*

Very few metabolites were detected in the caecal contents of germ-free mice. The main metabolite detected was ethanol (~1 mM) with very low concentrations of acetate (~ 0.05 mM) and formate (~0.04 mM) (Fig. 42). Upon colonisation with *R. gnavus* CC55\_001C the concentrations of ethanol, acetate and lactate increased when compared with germ-free controls (~2 mM, ~0.4mM and ~0.6 mM, respectively) (Fig. 43). There was also a small increase in the concentration of formate in these mice (~0.1 mM) (Fig. 43). Very few metabolites were observed to change in concentration when mice were mono-colonised with *A. muciniphila* ATCC BAA 835. The main difference was an increase in ethanol (~1.1 mM) (Fig. 43). In co-colonised mice, the pattern of metabolite production mainly reflected the combined effect of *R. gnavus* and *A. muciniphila* colonisation (production of acetate, lactate, formate and ethanol) (Fig. 43).

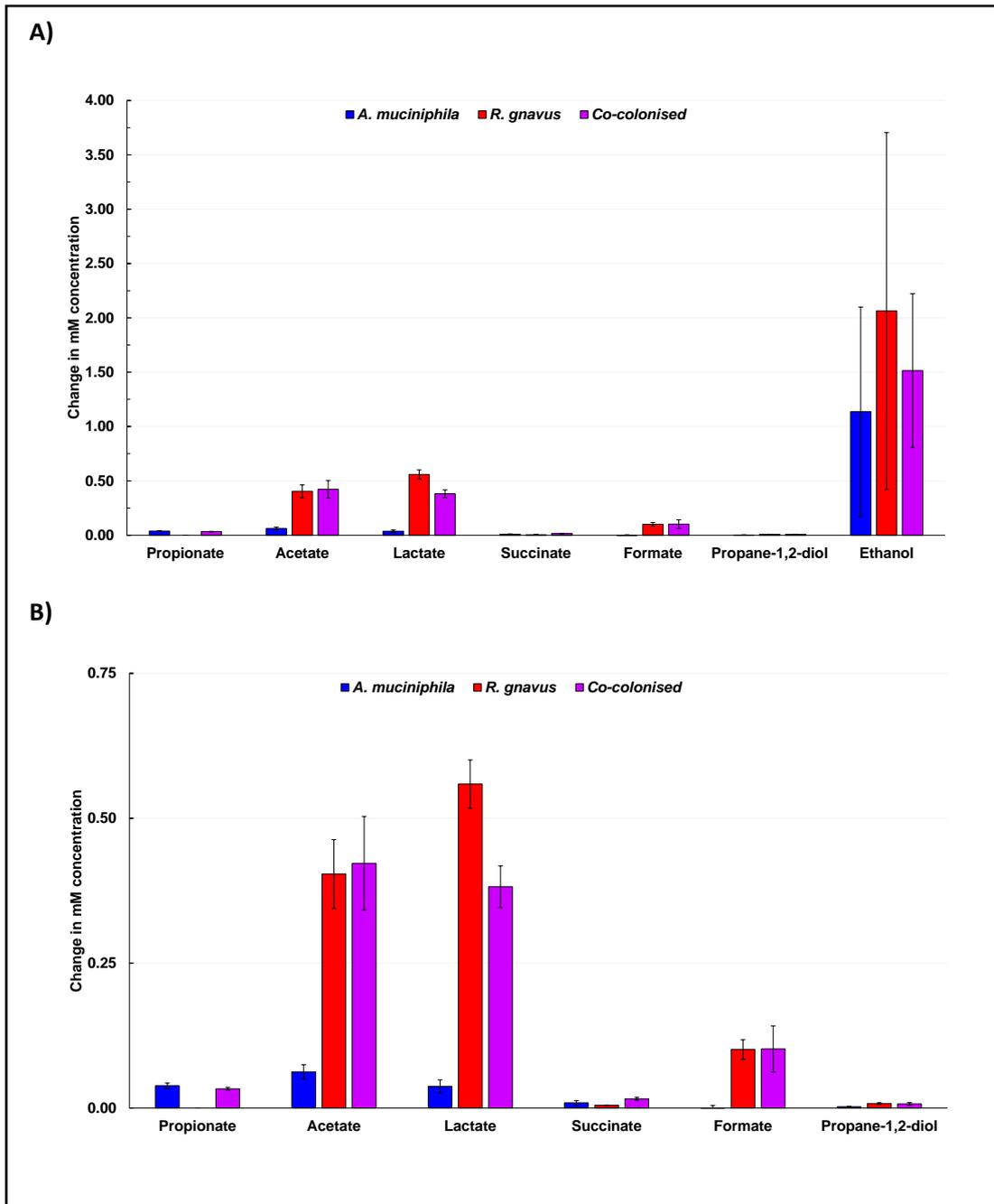


**Figure 42 | Concentrations (mM) of the main metabolites in the caecal contents of germ-free mice**

*A) mM concentration of all selected metabolites*

*B) y axis scaled up to view mM concentrations of propionate, acetate, lactate, succinate, formate and propane-1,2-diol*

*Caecal contents were harvested from germ-free mice. Metabolites were quantified in the supernatant of caecal contents from germ-free mice using  $^1\text{H}$  NMR. The mM concentration of the main metabolites involved in SCFA metabolism are shown in the graphs.*



**Figure 43 | Change in mM concentration of metabolites in mice mono- and co-colonised with *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C when compared with germ-free controls**

*A) Change in mM concentration of all selected metabolites*

*B) y axis scaled up to view change in mM concentrations of propionate, acetate, lactate, succinate, formate and propane-1,2-diol*

*Caecal contents were harvested from mice mono- and co-colonised with *A. muciniphila* and/or *R. gnavus* for 21 days. Metabolites were quantified in the supernatant of caecal contents using  $^1\text{H}$  NMR. Graphs show the change in mM concentration of the main metabolites involved in SCFA metabolism in mono- and co-colonised mice when compared with germ-free controls.*

### 5.3 Discussion

Mechanistic investigations were performed *in vitro* and in germ-free mice to investigate the relationship between the mucin degrading strains of *R. gnavus* and *A. muciniphila* and mucin glycosylation.

*In vitro* experiments revealed that *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835 utilise different sets of mono- and small oligosaccharides. *R. gnavus* CC55\_001C was able to grow on Fuc, 2' FL, 3' FL and 3' SL, as previously reported for *R. gnavus* type strain ATCC 29149 (Crost et al., 2013). However, no detectable growth was observed when the sialic acids, Neu5Ac or Neu5Gc or Lac were used as a sole carbon source, also in agreement with previous observations with the ATCC 29149 strain (Crost et al., 2013). *R. gnavus* CC55\_001C encodes the full complement of genes in the cluster dedicated to sialic acid metabolism (See appendix 6 for blastp homology), including an intramolecular sialidase, which in *R. gnavus* ATCC 291459 and *R. gnavus* ATCC 35193 was shown to cleave off and release sialic acid from  $\alpha$ 2-3 linked sialylated glycoproteins in a 2,7-anhydro sialic acid form, supporting the growth of these bacteria (Crost et al., 2016; Tailford et al., 2015b). It has been suggested that the presence of the intramolecular *trans*-sialidase confers to these strains a competitive nutritional advantage by allowing them to release sialic acid from mucins in a form they can preferentially access (a 'selfish' mechanism) (Crost et al., 2013; Crost et al., 2016). It is very likely that the same mechanism of sialic acid metabolism applies to *R. gnavus* CC55\_001C. This was further supported by our *in vitro* growth assays where we showed that, in stark contrast to *A. muciniphila*, *R. gnavus* CC55\_001C cell numbers were sustained from the initial inoculum for 27 h before any detectable depletion was observed in the highly sialylated LS174T mucin. However, a higher concentration of LS174T mucin may be required in order to provide *R. gnavus* with enough substrate for growth, since mucins from colorectal cancers contain a lower number of carbohydrate chains as identified previously (Mihalache et al., 2015) and confirmed in this study. In addition to growth on 3' SL, here, we also observed minimal growth of *R. gnavus* CC55\_001C on 6' SL, a substrate on which *R. gnavus* ATCC 29149 could not grow (Crost et al., 2013). Growth of *R. gnavus* CC55\_001C on this substrate may indicate a broader substrate specificity of the predicted sialidase encoded by *R. gnavus* CC55\_001C, although this would need to be verified by a biochemical approach.

*A. muciniphila* ATCC BAA 835 growth was observed in very few simple sugars. Unlike *R. gnavus*, *A. muciniphila* ATCC BAA 835 could not grow on Fuc, sialic acid, fucosylated or sialylated mono- and disaccharides. Although *A. muciniphila* encodes two sialidases,

AkmNan0625 and AkmNan1835 which are active against both 3' SL and 6' SL, this species is unable to utilise sialic acid (Tailford et al., 2015b; van Passel et al., 2011). This may be due to the *A. muciniphila* genome lacking the full complement of genes in the nan cluster (as identified by an in-house bioinformatician). In our study, the highest growth was apparent in the media containing pPGM. Since pPGM contains a complex O-glycan profile, this would suggest that *A. muciniphila* favours the utilisation of more complex glycan chains for growth. It has been proposed that in order to grow, *A. muciniphila* requires a combination of oligosaccharides as well as amino acids, provided by the complex mucin structure (Derrien et al., 2004). Furthermore *A. muciniphila* is the predominant mucin-degrading organism in human faeces (Derrien et al., 2004). Our *in vitro* culturing assays suggested that the pPGM, containing complex mucin glycans was more favourable to the growth of *A. muciniphila*, which ultimately had a higher fold growth in medium containing this mucin than *R. gnavus*. In contrast, in LS174T mucin, which possesses smaller and mostly sialylated oligosaccharide chains, *A. muciniphila* cell numbers reduced dramatically from the original inoculum, with no detectable growth. Overall, the decrease in growth between pPGM and LS174T mucin was much more dramatic for *A. muciniphila* than *R. gnavus*, suggesting that the LS174T mucin glycan profile is more favourable to *R. gnavus* than to *A. muciniphila*. This may partly contribute to the observed decrease in *A. muciniphila* abundance in UC, where an altered O-glycan profile similar to that of LS174T mucins has been demonstrated to occur (Larsson et al., 2011; Larsson et al., 2009).

In parallel with the growth assays, we measured the abundance of metabolites, including SCFAs produced by *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835 grown in mucin supplemented medium. In the colon, there are typically three main bacterial SCFAs produced; propionate, butyrate and acetate, which are usually found in a 1:1:3 ratio (Louis et al., 2014). The SCFAs link the microbiota and the host immune system as they are rapidly absorbed by the intestinal epithelium, and exert a number of immuno-modulatory effects. In general, butyrate appears to exert the greatest effects on host immunity and is a major source of energy for colonocytes, whilst propionate, and acetate in particular appear to be less potent (Correa-Oliveira et al., 2016; Tedelind et al., 2007).

Both bacteria produced acetate and propionate when grown in pPGM. Propionate production can occur *via* a number of molecular pathways (Reichardt et al., 2014). Here, bioinformatics and metabolite analyses suggest that *R. gnavus* mainly utilises the propanediol pathway, due to the presence of genes homologous to PduP and PduQ and the increase in concentration of propane-1,2-diol and propanol in the spent media, which are

intermediates and by-products of this pathway, respectively. This pathway has also been demonstrated to occur in the closely related species, *Roseburia inulinivorans* and *Ruminococcus obeum* (Reichardt et al., 2014). Furthermore, *R. gnavus* ATCC 29149 has also been suggested to produce propionate through this pathway, mainly through the utilisation of Fuc from fucosylated substrates (Croft et al., 2013). *A. muciniphila* ATCC BAA 835 also produced propionate, however in accordance with previous publications, no homology was identified with PduP, a propanol dehydrogenase enzyme involved in this pathway, and therefore *A. muciniphila* does not appear to utilise the propanediol pathway (Reichardt et al., 2014). It is possible that *A. muciniphila* utilises the succinate pathway to produce propionate, as suggested by the low level protein homology with MmdA, an  $\alpha$ -subunit of a methylmalonyl-CoA decarboxylase enzyme involved in this pathway, and by the large amount of succinate detected in the spent media. This pathway is widely used by Bacteroidetes to generate propionate from carbohydrates and by some Firmicutes to produce propionate from lactate or succinate (Reichardt et al., 2014).

When cells were grown in LS174T mucin, the amount of propionate produced by *R. gnavus* CC55\_001C, as well as intermediates of the propanediol pathway was reduced when compared to growth of *R. gnavus* in pPGM. However, the increase in acetate concentration, as well as formate by *R. gnavus* was similar in both pPGM and LS174T mucin. This suggests that during growth of *R. gnavus* CC55\_001C in LS174T mucin, the propanediol pathway is not active, whilst other metabolic pathways are still utilised. These observations are consistent with previous reports of *R. obeum* A2-162, where it was shown that the propanediol pathway was only active during growth on fucosylated substrates, producing propionate, whereas only acetate, formate and lactate were produced when the bacteria were grown on glucose (Reichardt et al., 2014). In contrast, in our study, production of all metabolites by *A. muciniphila* were reduced when this bacteria was grown in LS174T mucin as a carbon source when compared to pPGM. This suggests a sustained metabolism of *R. gnavus* CC55\_001C on LS174T mucin, and provides further evidence for the better ability of *R. gnavus* to survive on this source of carbon as compared to *A. muciniphila*.

The balance of metabolites in the gut lumen is delicately controlled by production, use and mucosal uptake. According to our results, this balance is potentially disrupted when the nutritional reservoir provided by the mucin glycans is altered. Interestingly, in a study performed in 1988, the ratio of acetate to total SCFAs was increased in colon cancer patients compared with controls, however not in patients with IBD or diverticulitis (Weaver et al., 1988). However, this study was performed using enema samples, which may not directly

represent the mucosal concentrations of SCFAs. Although *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835 were not observed to directly produce butyrate, it is understood that some gut bacteria, particularly Firmicutes, can utilise acetate and convert it to butyrate (Duncan et al., 2004a; Louis et al., 2014). In the healthy gut, several of the most abundant species including *F. prausnitzii* and *Roseburia* spp. are generally net users of acetate, and butyrate producers (Duncan et al., 2002a; Louis and Flint, 2009). A number of studies have reported a decrease in *F. prausnitzii* and *Roseburia* spp. in IBD (Machiels et al., 2014; Sokol et al., 2009). A reduction of these bacteria may therefore result in a net increase in acetate as well as a net decrease in butyrate, potentially affecting gut homeostasis. Machiels and coll. reported that whilst the butyrate concentration in UC was lower, this failed to reach significance (Machiels et al., 2014). In contrast, some studies have reported increases in butyrate producers such as *F. prausnitzii* in CD (Hansen et al., 2012b; Willing et al., 2010a). It is possible that alterations in mucin glycosylation favouring the growth of acetate producers such as *R. gnavus* results in increased metabolic substrate for acetate consumers such as *F. prausnitzii*, explaining these results. In support of this, *F. prausnitzii* was unable to colonise the GI tract of germ-free rats without the presence of the acetate producer *B. thetaiotaomicron* (Wrzosek et al., 2013). In active UC patients it has been shown that the intestinal epithelium has a diminished capacity to oxidize butyrate (Kato et al., 2007). Therefore the abundance of *F. prausnitzii* and its production of butyrate may not be the important factor in determining gut homeostasis, but rather the utilisation of butyrate by the host. It is evident that more comprehensive studies are required to fully elucidate this relationship and the factors driving changes in the abundance of butyrate producers in IBD, in relation to mucin glycosylation.

To further characterise the relationship between mucin degraders and mucin glycosylation, germ-free mice were mono- and co-colonised with *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835 to determine if these bacteria were able to modify the mucin glycan profile. Both bacteria could rapidly and stably colonise the GI tract, in agreement with previous studies where germ-free mice have been successfully mono-colonised with *R. gnavus* E1 and *A. muciniphila* (Crost et al., 2010; Derrien et al., 2011; Graziani et al., 2016). However, the density to which *R. gnavus* and *A. muciniphila* were able to colonise mice in our study appeared to be dependent on whether mice were mono- or co-associated with the bacteria. Whilst *A. muciniphila* survival appeared to be better when mono-associated with mice, *R. gnavus* appeared to thrive better in the presence of *A. muciniphila*. It is possible that in co-colonised mice, *A. muciniphila*, due to its cocktail of glycoside hydrolases and apparent

preference for complex mucin glycans, generates more accessible nutrients for *R. gnavus*. Furthermore, *R. gnavus* may be capable of utilising the glycans more rapidly than *A. muciniphila*, explaining the decreased density of *A. muciniphila* in these mice when compared to those that have been mono-colonised. Indeed, from our *in vitro* assays, *A. muciniphila* displayed a prolonged lag phase on pPGM compared with *R. gnavus*, suggesting that *R. gnavus* is able to access the nutrient supply provided by pPGM more quickly. In addition, transcriptome analysis following colonisation of mice with *A. muciniphila* demonstrated that this bacteria modulates genes involved in establishing homeostasis and immune tolerance toward commensal microbiota, with the largest transcriptional changes in the colon (Derrien et al., 2011). It is possible, therefore, that increased tolerance towards *R. gnavus*, as a result of the presence of *A. muciniphila* allowed *R. gnavus* to colonise to a higher density in these mice.

By mass spectrometry, we detected many of the same glycan structures in the colon of germ-free C57BL/6J mice as has been identified previously in this strain and in germ-free rats by MALDI-TOF and LC/MS techniques (See appendix 3) (Johansson et al., 2015; Wrzosek et al., 2013). It is of note that polyhexoses were also detected amongst the glycan structures, particularly in *A. muciniphila* mono-colonised mice where we could not detect any mucin glycan structures. It is unclear whether these polyhexoses are an inherent factor introduced during sample processing and analysis, which are normally undetected when high levels of glycosylation mask the polyhexose peaks. The polyhexose contamination in mucins from germ-free mice may be more visible due to a lower glycosylation. In mice colonised by *A. muciniphila*, bacterial degradation of mucin glycans may have caused the concentrations of most structures to fall below the detection limit, exacerbating the relative abundance of the polyhexose contaminants. Indeed, by HPAEC-PAD, although we could observe sialic acid, its abundance on mucins was depleted in *A. muciniphila* mono-colonised mice in comparison to germ-free controls.

In *R. gnavus* mono-colonised mice, colonic mucin glycosylation was still observed, but changes in the profile of glycan structures were apparent when compared to germ-free mice. In particular, both Muc2 and mixed mucins displayed a lower degree of sialylation, in accordance with the ability of *R. gnavus* to liberate and utilise sialic acid transglycosylation product due to the presence of an intramolecular *trans*-sialidase and the nan operon (Crost et al., 2013; Crost et al., 2016; Tailford et al., 2015b). A nearly 2-fold decrease in sialylation by *R. gnavus* in both Muc2 and mixed mucin fractions was confirmed by HPAEC-PAD. By MS, the decrease in sialic acid could be attributed to a depletion in two main structures in Muc2,

the sialyl-TF antigen (m/z 895) and a mono-sialylated core 2 structure of which there are two potential isomers (m/z 1589). In a previous study, this latter structure was also noted to be decreased in the Muc2 population of mice following conventionalisation with a normal C57BL/6J microbiota (Johansson et al., 2015). These results therefore suggest m/z 1589 and sialyl-TF may act as major nutrient sources for *R. gnavus* CC55\_001C. These findings may explain the prolonged survival of *R. gnavus* on LS174T mucins, in which we observed the MUC2 fraction to be rich in sialyl-TF (59.14% of structures), however m/z 1589 was undetected. It is important to note that differences between the O-glycosylation of Muc2 isolated from the distal colon of mouse models and humans have also been identified previously, and therefore interpretation of results from mouse studies must be approached with caution (Thomsson et al., 2012). In contrast to our findings in the Muc2 fraction, in the mixed mucin fraction from mice, structure m/z 1589 was increased in mice mono-colonised with *R. gnavus* when compared to germ-free controls. It is possible that in the mixed mucin fraction sialic acid is predominantly  $\alpha$ 2,6-linked to this structure, whilst in the Muc2 fraction it is  $\alpha$ 2,3-linked, and therefore it is not as readily cleaved in the mixed fraction due to the predicted  $\alpha$ 2,3 linkage specificity of the sialidase encoded by *R. gnavus*. Further structural analysis of mucin glycans in both fractions *via* MS/MS would allow the validation of this hypothesis.

Interestingly, although there was an apparent depletion in glycan structures in *A. muciniphila* mono-colonised mice, in co-colonised mice, we were still able to detect glycan structures in colonic mucins despite the presence of *A. muciniphila*. A possible explanation for this might be that *R. gnavus* stimulates the expression of glycosyltransferases, as demonstrated in previous studies (Graziani et al., 2016), helping to restore the glycan structures lost by *A. muciniphila* degradation. In particular, Graziani and coll. showed that colonization of germ-free mice with *R. gnavus* E1 resulted in an increase in GlcNAc and  $\alpha$ -1,2-Fuc residues on glycoproteins (Graziani et al. 2016). Consistent with this, we observed an increase in the relative amount of mucin fucosylation in co-colonised mice compared with germ-free controls. However, as we were unable to distinguish between GlcNAc and GalNAc ions using MALDI-TOF, we could not assess the impact of *R. gnavus* colonisation on GlcNAc abundance. A trend towards a decrease in sialylation was also evident in co-colonised mice by MS and HPAEC-PAD, particularly in the mixed mucin fraction. Alike the *R. gnavus* mono-colonised mice, this trend appeared to be mainly due to a decrease in the sialyl-TF and m/z 1589 structures. The relative abundance of m/z 1589 in particular was reduced to a greater extent in the Muc2 fraction of co-colonised mice compared with *R. gnavus* mice, and in contrast to

*R. gnavus* mono-colonised mice was also reduced in the mixed mucin fraction of co-colonised mice. This is consistent with the ability of *A. muciniphila* to also liberate sialic acid due to the two sialidases it encodes, which have specificity for both  $\alpha$ 2,3- and  $\alpha$ 2,6- linked structures, and provides further evidence that in mixed mucins, m/z 1589 may be sialylated *via* an  $\alpha$ 2-6 linkage.

In order to gain some insight into the metabolites produced by *R. gnavus* and *A. muciniphila* during colonisation of germ-free mice, we measured these in the caecal contents by NMR. Germ-free mice caecal contents contained acetate as the main SCFA, but lacked of propionate and butyrate, consistent with previous observations in germ-free animals (Wrzosek et al. 2013). However, the concentrations we detected were much lower than previously observed, possibly due to the difference animal model of germ-free mice versus rats. Furthermore, Wrzosek and coll. used gas liquid chromatography, which may differ in sensitivity to the NMR used in our analysis. In our study, the main metabolites produced in the caecum of mice mono-colonised with *R. gnavus* CC55\_001C were ethanol, acetate, lactate and formate. Interestingly, production of propionate and intermediates of the propanediol pathway were not detected. This pattern of metabolite production is consistent with the fact that mucin fucosylation is generally considered to be low in proximal regions of the mouse GI tract, and is similar to our *in vitro* observations of *R. gnavus* metabolite production when grown in LS174T mucin, which also has a low level of mucin fucosylation. Metabolite production in the caecum of mice mono-colonised with *A. muciniphila* ATCC BAA 835 was mainly restricted to ethanol, with very low production of other metabolites, unlike our *in vitro* observations of growth on mucins such as pPGM. This suggests that whilst *A. muciniphila* may be carrying out fermentation in caecum, it is utilising different substrates to those during growth on pPGM, probably as a result of a different glycosylation pattern on the mucins from this region. In co-colonised mice, the metabolite profile reflected that of the combined effect of *R. gnavus* and *A. muciniphila* growth, suggesting that metabolism of each bacteria was unaffected by the presence of the other, although more work is warranted to validate these hypotheses.

Overall, based on evidence from our *in vitro* assays, it is probable that the microbiota and particularly the abundance of mucin degraders is, at least in part, shaped by mucin glycosylation. However, it is important to note that the purified mucin used in these assays lack many of the physiological constituents of mucus such as antimicrobials and cytokines that may also modulate the microbiota. Furthermore, the ability of individual species to colonise is probably also dependent on the composition of the microbiota, as supported by

the increased colonisation of germ-free mice by *R. gnavus* in the presence of *A. muciniphila*. The mechanisms by which this occurs are numerous, but probably include cross-feeding between species and remodelling of mucin glycosylation. Indeed, we showed that colonisation of mice with both *A. muciniphila* and *R. gnavus* resulted in a glycan profile shifted from that of a germ-free mouse. The synergistic action of these bacteria may help to maintain a homeostatic mucosal barrier *via* the coordinated degradation and stimulation of mucin glycosylation, which may be disrupted during inflammatory conditions. In order to understand these interactions further, these organisms would need to be studied in the context of a conventional microbiota.

## Chapter 6) Conclusions and perspectives

The focus of this PhD project was to investigate the relationship between the mucus-associated microbiota and mucin glycosylation, and the mechanisms underpinning changes in these factors in IBD. To encompass the above, the first part of this thesis explored alterations in the human gut mucosa-associated microbiota composition and mucin glycosylation/the expression of glycosyltransferases in UC patients. The second part used *in vitro* growth assays and investigations in mice with the aim of elucidating the mechanisms by which two specific mucin degrading bacteria, *R. gnavus* CC55\_001C and *A. muciniphila* ATCC BAA 835 interact with mucin and mucin glycans. This chapter summarises the findings from these investigations, their impact in wider field, and future directions of work.

Previously, studies have identified a breakdown of gut homeostasis in IBD, incorporating 1) dysbiosis of the microbiota, and 2) alterations in mucin barrier function, including mucin glycosylation. However, both of these factors have rarely been studied in the same model, and particularly not in humans. As a result, the relationship between both components remains poorly understood. In our study, the collection of mucosal lavage samples from controls and IBD patients allowed us to study the composition of the microbial community in direct contact with the mucins, a major advantage of using this method of sampling.

Overall, our profiling of the mucus-associated microbiota and mucin glycosylation profile in controls and UC patients revealed general trends in accordance with the main properties of these components identified in previous studies, were they have been characterised separately. By both 16S sequencing and qPCR, the microbiota was highly similar in proximal and distal colon, which is consistent with previous observations (Lavelle et al., 2015). In stark contrast, mucins and mucin glycosyltransferases displayed gradients from proximal to distal colon, including increasing gradients of sialic acid and sialyltransferase expression, but decreasing gradients of Fuc in both controls and UC patients, again in accordance with previous observations (Robbe et al., 2003; van der Post and Hansson, 2014).

However, differences in both mucin glycosylation and the microbiota were apparent in UC patients when compared to controls. It is important to note that these differences were only evident following detailed analyses such as quantification of specific bacterial groups by qPCR, and structural characterisation of mucin glycans by mass spectrometry. As a consequence, it is worth noting that broader analysis methods, such as 16S sequencing platforms or lectin screening may not be sufficient to characterise microbiota and host glycan changes in IBD. For example, we observed no alterations in the diversity, richness or

evenness of the microbial community whereas distinct changes in the abundances of bacterial groups were apparent in UC, including a decrease in the ratio between *A. muciniphila* and *R. gnavus* in both sigmoid and ascending colon locations. In parallel, UC mucins displayed alterations in glycosylation including a site-wide decrease in many of the abundant complex glycan structures found in controls, as well as a site-wide decrease in the main fucosylated structures. Furthermore, we observed an increase in the main sialylated structures in UC, as previously observed with sialylated structures such as STn (Larsson et al., 2011), but in our study this was mainly a feature in the ascending colon. It is important to note that many of the trends we detected in the human lavage samples did not reach statistical significance due to the limited number of samples we were able to obtain. Another limitation was the variability between IBD patients which included those at different stages of disease (remission and relapse), on different medications and with variable lifestyles. Indeed the stage of disease, particularly the presence of inflammation, has been implicated as an important factor associated with the presence of aberrant mucin glycosylation (Larsson et al., 2011). Therefore, future investigations should focus on using larger patient numbers and better defined patient groups. Longitudinal studies of individual IBD patients would be greatly helpful, to take into account inter-individual variation.

Despite these limitations, this first part of the work allowed us to generate the hypothesis that alterations in mucin glycosylation in UC patients may result in a mucosal environment that is less favourable to the growth of *A. muciniphila*, explaining its decrease in UC. To test this hypothesis, the growth of *A. muciniphila* and *R. gnavus* was assayed using two different types of mucin; pPGM containing a complex glycosylation profile with a high proportion of fucosylation, or mucins from the LS174T colon adenocarcinoma human cell line, with a glycosylation profile more similar to UC, i.e. less complex, with decreased fucosylation and increased sialylation. The assay validated the trends we observed in humans, showing that *A. muciniphila* growth was drastically affected in media supplemented with the LS174T mucin, containing a 'UC-like' glycosylation profile. In contrast, *R. gnavus* was able to tolerate LS174T mucin, sustaining cell numbers for a longer period of time. The effect of a lack of fucose in these mucins on *R. gnavus* metabolism was evident through the decrease in the production of intermediates of the propanediol pathway, however production of acetate was maintained, suggesting that *R. gnavus* utilises other glycan nutrients from LS174T mucin. These observations support a causal role for mucin glycosylation in shaping the composition of the gut microbiota and the balance of SCFAs in UC. In accordance with this, recent findings showed that the availability of mucosal carbohydrates, in particular fucose and sialic acid,

has the potential to impact on the composition of microbial species including the expansion of pathogenic organisms such as *Salmonella enterica* and *C. difficile* (Ng et al., 2013; Tong et al., 2014). In addition, in humans, individuals appear to have a different microbiota composition depending on the expression ABO blood group antigens and secretor (FUT2 genotype) status (Makivuokko et al., 2012; Wacklin et al., 2014). As well as a lower species richness, the abundance of specific bacterial genera appears to be modified in non-secretors, including a lower abundance of Akkermansia spp. (Wacklin et al., 2014). For example, individuals harbouring the B blood group antigen have significantly different profiles of *E. rectale*, *C. coccoides* and *C. leptum* groups (Makivuokko et al., 2012).

Although our data provide an interesting insight into the impact of mucin glycosylation on *R. gnavus* and *A. muciniphila* growth and metabolism, additional experiments would be beneficial to help validate our findings. A main limiting factor in our *in vitro* assays was the time-consuming nature of LS174T mucin purification, from which only small yields could be obtained. Since we expect that LS174T mucin contains a lower carbohydrate content than pPGM, given more time, future experiments could investigate whether *R. gnavus* can grow in media containing higher percentages of this mucin. It may also be valuable to perform glycosylation analysis of mucins in the spent media following growth of these organisms, particularly to elucidate the major nutrients utilised by *R. gnavus* in the absence of fucosylated glycans. In addition, co-culture experiments with both bacteria would provide an insight into whether these bacteria can outcompete each other, or benefit from one other *via* cross-feeding on these mucin glycans. In the long term, and since we showed that mucin degradation properties can be strain specific, further experiments could also investigate the ability of other strains of *R. gnavus* and *A. muciniphila* to utilise differentially glycosylated mucins.

In addition to understanding the impact of mucin glycosylation on the microbiota, we tested whether *A. muciniphila* ATCC BAA 835 and *R. gnavus* CC55\_001C could shape mucin glycans by mono- and co-colonising germ-free mice with each bacterial species. Following 3 weeks of colonisation, distinct changes in the glycosylation of mucins from these mice were evident. *A. muciniphila* appeared to deplete mucin glycosylation, possibly due to the expression of a wide range of glycoside hydrolases. This is an interesting observation since *A. muciniphila* is generally regarded as a 'biomarker' of health. However, previous studies suggest that *A. muciniphila* may stimulate mucin synthesis, facilitating turnover of mucins and enhancing barrier function (Derrien et al., 2016). It is also possible that *A. muciniphila* contributes to

host health in other ways. For example, a recent study demonstrated the ability of *A. muciniphila* to bind to intestinal epithelial cells and strengthen the integrity of the cell monolayer (Reunanen et al., 2015). However, Shin and coll. found that serum LPS levels and gut permeability were not significantly different in mice administered with *A. muciniphila* when compared with control mice (Shin et al., 2014). Evidence also suggests that *A. muciniphila* alters the host transcriptome profile and balances immune responses (Derrien et al., 2011). Therefore, a role for *A. muciniphila* in maintaining host health may involve facilitating tolerance towards commensals, whilst other bacteria help to replenish mucin glycosylation through mechanisms such as stimulating glycosyltransferase expression. In accordance with this, in co-colonised mice, we observed that some mucin glycans are still intact despite the presence of *A. muciniphila*, suggesting that *R. gnavus* may help to restore mucin glycans degraded by *A. muciniphila*. *In vivo* experiments performed by Graziani and coll. showed that colonisation with *R. gnavus* E1 resulted in an increase in GlcNAc and  $\alpha$ -1,2-fuc residues on glycoproteins, and a significant induction in the expression of glycosyltransferase and mucin genes, including *ST6Gal1*, *C1GalT1*, *Muc1*, *Muc2* and *Muc3* (as well as *C3GnT* and *Fut3*, although not significant) (Graziani et al., 2016). A similar function has previously also been shown of *B. thetaiotaomicron* which is able to modulate host production of hydrolysable fucosylated glycans to its advantage (Hooper et al., 1999). Moreover, different microbiota can influence the mucus properties of mice of the same genetic background, leading to differences in the penetrability of the inner mucus layer to bacterial sized beads (Jakobsson et al., 2015). This provides further evidence to suggest that the mucus barrier is shaped by the microbiota.

In future, it will be of interest to investigate the mechanisms *via* which *R. gnavus* CC55\_001C alters mucin glycosylation. For example, analysis of the mouse transcriptome would allow us to establish if *R. gnavus* mediates these effects through upregulation of glycosyltransferases. Furthermore, in order to understand the biological impact of mucin glycan remodelling by *A. muciniphila* and *R. gnavus* in IBD, DSS treatment of these mice would allow us to evaluate their susceptibility to induced colitis. However, it is also important to acknowledge that the conclusions drawn from gnotobiotic experiments are limited as the normal behaviour of a bacterium is difficult to mimic in a germ-free mouse model, where competition for niche and substrates from other commensal organisms and pathogenic bacteria is lacking. Therefore, experiments could explore the effects of manipulating the microbiota composition and abundance of *R. gnavus*/*A. muciniphila* in a conventionalized microbiota context.

Together, our findings point towards a multifactorial dysregulation at the epithelial interface in IBD, encompassing (although not limited to) changes in mucosal barrier function including altered mucin glycosylation, dysbiosis of the microbiota and aberrations in metabolic function.

This study has provided evidence to suggest that mucin glycosylation may be a factor which contributes towards shaping the microbiota composition. On the other hand, the data also support a role for bacteria in remodeling mucin glycosylation and facilitating mucin turnover, a process which could be disrupted should the microbiota composition be altered. In IBD, therefore, it may be more appropriate to consider that the intestinal barrier system is in a feedback loop, where all components are interdependent and an initial disruption in any of these components drives further alterations in other mucosal constituents, propagating disease and exacerbating inflammation.

Such mechanisms would also explain the large amounts of seemingly contradictory data arising from the literature, since such alterations are likely to be impacted by the inter-individual variation of the patient cohorts studied and also by stage of IBD at which samples are taken. In addition to studying larger patient cohorts, as mentioned above, it is evident that attempts should be made to limit the effect of biases introduced as a result of study design, such as the method of sample collection, how the samples are processed, and how they are analysed. An important consideration to take into account during choice of cohort size is the impact bowel preparation may have on the parameters being analysed, since evidence already exists to suggest the microbiota composition and diversity is affected, however little information exists surrounding the effect on the mucus layer (Shobar et al., 2016).

A question that still remains is how either mucin glycosylation or the microbiota composition are initially driven into an altered state. Previous studies have suggested that altered expression of glycosyltransferases, such as an increase in ST6GalNAc-I may be responsible for aberrant mucin glycosylation in UC, in particular an increase in the STn antigen (Larsson et al., 2011). In our study, we did not observe differences between the gene expression of glycosyltransferases in controls and UC patients, however it is possible that alterations occur at the level of protein expression, activity of glycosyltransferases or in the expression of glycosyltransferases not covered in this study. Furthermore, disruption of the Golgi glycosylation machinery, such as the localisation of glycosyltransferases, may result in altered mucin glycosylation (Theodoratou et al., 2014). For example, an increase in Golgi pH

results in increases in the TF antigen, a characteristic alteration in mucin glycosylation in IBD (Campbell et al., 2001a). It may be worthwhile to consider possibilities such as these in future studies.

Furthermore, here we did not investigate the role of the immune system, which may be another factor contributing to inflammation. For example, mucin glycosylation may be initially altered *via* a dysregulation of immune components resulting in the onset of IBD. Knockout of  $\gamma\delta$  T-cells (TCR $\delta^{-/-}$ ) in mice results in increased *Muc2*, *Muc13* and *Muc17* gene expression in the small intestine, associated with a reduction in the sialylation of mucins. These mice are more susceptible to DSS induced colitis, supporting a role for  $\gamma\delta$  T-cells in maintaining mucosal homeostasis by regulating small intestinal mucin expression and glycosylation (Kober et al., 2014). Pro-inflammatory cytokines such as TNF and IFN- $\gamma$  have also been linked to changes in glycosylation, including increased TF antigen and sialyl-Lewis<sup>x</sup> although how these effects are mediated is unknown (Campbell et al., 2001b; Theodoratou et al., 2014). Alternatively, immune abnormalities or environmental factors could drive changes in the microbiota, and specifically mucin degraders, resulting in an altered mucin glycosylation pattern. Deficiency in the NLRP6 inflammasome in mice results in an alteration in faecal microbiota, including expansion of sulfatase producing Prevotellaceae (Elinav et al., 2011). DSS administration resulted in alterations in the gut microbiota composition, including increases in the abundance of the mucin degrader *A. muciniphila*. These alterations were associated with immunological changes (Hakansson et al., 2015).

Elucidating the role of the various host and environmental components in IBD is complicated with samples derived from the mucosa, a dynamic surface where constituents are in a constant state of flux. Although increasing numbers of techniques are being developed to study the components contributing to the intestinal barrier, limitations still exist and interpretation of data should be approached with caution. Recent evolution in glycoprofiling techniques, which have previously been lagging, means that high-throughput glycan analyses are now feasible and affordable, and effort should be made to undertake large and high-quality studies to replace previous smaller powered experiments (Theodoratou et al., 2014). Furthermore, with the advent of improved bioinformatics tools, broad level analysis of the microbiota by 16S rRNA gene and metagenomic/metatranscriptome analysis should also be readdressed to allow the composition of the microbiota to be resolved at the strain and genome level (Miyoshi and Chang, 2016). Furthermore, these analytical tools are of little use

without the development of user-friendly software to probe/model potential interactions between multiple parameters, a main limitation we recognised in our study.

In light of the apparent multifactorial nature of IBD, it is likely that the strategies to treat the disease will also need to be readdressed. Therapeutics are currently focussed on 'correcting' only one component of the mucosal barrier, such as the microbiota through the administration of probiotics, prebiotics or faecal microbiota transplant, or the immune system *via* immunotherapy. However, these therapeutics are often prescribed with little success. Given our current understanding of IBD, it may be better to design dual therapeutic strategies that target more than one of the components known to drive a deviation from intestinal homeostasis, e.g. providing immunotherapy in addition to faecal microbiota transplant to restore the gut microbiota. In addition, given our knowledge regarding inter-individual variation, it may be beneficial to modify these therapeutics to suit individual disease aetiology (Miyoshi and Chang, 2016). In the long term, focus should shift to providing therapies targeted at curing disease by restoring the genetic abnormalities pre-disposing individuals to IBD, whilst also providing treatment to normalise the microbiota, immune system and intestinal mucus at the mucosal interface.

## References

- Aagaard, K., Ma, J., Antony, K.M., Ganu, R., Petrosino, J., and Versalovic, J. (2014). The placenta harbors a unique microbiome. *Science translational medicine* 6.
- Aagaard, K., Riehle, K., Ma, J., Segata, N., Mistretta, T.A., Coarfa, C., Raza, S., Rosenbaum, S., Van den Veyver, I., Milosavljevic, A., *et al.* (2012). A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. *Plos One* 7, e36466.
- Abreu, M.T., Vora, P., Faure, E., Thomas, L.S., Arnold, E.T., and Arditi, M. (2001). Decreased expression of toll-like receptor-4 and MD-2 correlates with intestinal epithelial cell protection against dysregulated proinflammatory gene expression in response to bacterial lipopolysaccharide. *J Immunol* 167, 1609-1616.
- Adibi, S.A. (1976). Intestinal phase of protein assimilation in man. *The American journal of clinical nutrition* 29, 205-215.
- Albert, M.J., Mathan, V.I., and Baker, S.J. (1980). Vitamin-B12 synthesis by human small intestinal bacteria. *Nature* 283, 781-782.
- Allen, A., and Flemstrom, G. (2005). Gastroduodenal mucus bicarbonate barrier: protection against acid and pepsin. *Am J Physiol Cell Physiol* 288, C1-19.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-410.
- Ambort, D., Johansson, M.E., Gustafsson, J.K., Ermund, A., and Hansson, G.C. (2012a). Perspectives on mucus properties and formation-lessons from the biochemical world. *Cold Spring Harbor perspectives in medicine* 2, a014159.
- Ambort, D., Johansson, M.E.V., Gustafsson, J.K., Nilsson, H.E., Ermund, A., Johansson, B.R., Koeck, P.J.B., Hebert, H., and Hansson, G.C. (2012b). Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. *Proc Natl Acad Sci U S A* 109, 5645-5650.
- Ambort, D., van der Post, S., Johansson, M.E., Mackenzie, J., Thomsson, E., Kregel, U., and Hansson, G.C. (2011). Function of the CysD domain of the gel-forming MUC2 mucin. *The Biochemical journal* 436, 61-70.

An, G.Y., Wei, B., Xia, B., McDaniel, J.M., Ju, T.Z., Cummings, R.D., Braun, J., and Xia, L.J. (2007). Increased susceptibility to colitis and colorectal tumors in mice lacking core 3-derived O-glycans. *J Exp Med* 204, 1417-1429.

Ananthakrishnan, A.N. (2013). Environmental risk factors for inflammatory bowel disease. *Gastroenterology & Hepatology* 9, 367-374.

Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A., *et al.* (2011). Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet* 43, 246-252.

Andersson, R.E., Olaison, G., Tysk, C., and Ekblom, A. (2001). Appendectomy and protection against ulcerative colitis. *New Engl J Med* 344, 808-814.

Andrianifahanana, M., Moniaux, N., and Batra, S.K. (2006). Regulation of mucin expression: mechanistic aspects and implications for cancer and inflammatory diseases. *Biochim Biophys Acta* 1765, 189-222.

Antoni, L., Nuding, S., Weller, D., Gersemann, M., Ott, G., Wehkamp, J., and Stange, E.F. (2013). Human colonic mucus is a reservoir for antimicrobial peptides. *Journal of Crohn's & colitis* 7, e652-664.

Arike, L., and Hansson, G.C. (2016). The densely O-glycosylated MUC2 mucin protects the intestine and provides food for the commensal bacteria. *J Mol Biol* 428, 3221-3229.

Arpaia, N., Campbell, C., Fan, X., Dikiy, S., van der Veeken, J., deRoos, P., Liu, H., Cross, J.R., Pfeffer, K., Coffey, P.J., *et al.* (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 504, 451-455.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., *et al.* (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174-180.

Asker, N., Axelsson, M.A., Olofsson, S.O., and Hansson, G.C. (1998). Dimerization of the human MUC2 mucin in the endoplasmic reticulum is followed by a N-glycosylation-dependent transfer of the mono- and dimers to the Golgi apparatus. *J Biol Chem* 273, 18857-18863.

Asker, N., Baekstrom, D., Axelsson, M.A., Carlstedt, I., and Hansson, G.C. (1995). The human MUC2 mucin apoprotein appears to dimerize before O-glycosylation and shares epitopes with the 'insoluble' mucin of rat small intestine. *The Biochemical journal* *308*, 873-880.

Asseman, C., Mauze, S., Leach, M.W., Coffman, R.L., and Powrie, F. (1999). An essential role for interleukin 10 in the function of regulatory T cells that inhibit intestinal inflammation. *J Exp Med* *190*, 995-1003.

Atuma, C., Strugala, V., Allen, A., and Holm, L. (2001). The adherent gastrointestinal mucus gel layer: thickness and physical state in vivo. *American journal of physiology Gastrointestinal and liver physiology* *280*, G922-929.

Avershina, E., Storro, O., Oien, T., Johnsen, R., Pope, P., and Rudi, K. (2014). Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiol Ecol* *87*, 280-290.

Ayabe, T., Satchell, D.P., Wilson, C.L., Parks, W.C., Selsted, M.E., and Ouellette, A.J. (2000). Secretion of microbicidal alpha-defensins by intestinal Paneth cells in response to bacteria. *Nat Immunol* *1*, 113-118.

Backhed, F. (2011). Programming of host metabolism by the gut microbiota. *Ann Nutr Metab* *58* 44-52.

Backhed, F., Ding, H., Wang, T., Hooper, L.V., Koh, G.Y., Nagy, A., Semenkovich, C.F., and Gordon, J.I. (2004). The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci U S A* *101*, 15718-15723.

Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A., and Gordon, J.I. (2005). Host-bacterial mutualism in the human intestine. *Science* *307*, 1915-1920.

Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., Li, Y., Xia, Y., Xie, H., Zhong, H., *et al.* (2015). Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell host & microbe* *17*, 690-703.

Backstrom, M., Ambort, D., Thomsson, E., Johansson, M.E.V., and Hansson, G.C. (2013). Increased understanding of the biochemistry and biosynthesis of MUC2 and other gel-forming mucins through the recombinant expression of their protein domains. *Mol Biotechnol* *54*, 250-256.

Bakhtiar, S.M., Leblanc, J.G., Salvucci, E., Ali, A., Martin, R., Langella, P., Chatel, J.M., Miyoshi, A., Bermúdez-Humarán, L.G., and Azevedo, V. (2013). Implications of the human microbiome in Inflammatory Bowel Diseases. *FEMS Microbiol Lett* 342, 10-17.

Bansil, R., and Turner, B.S. (2006). Mucin structure, aggregation, physiological functions and biomedical applications. *Curr Opin Colloid Interf Sci* 11, 164-170.

Baos, S.C., Phillips, D.B., Wildling, L., McMaster, T.J., and Berry, M. (2012). Distribution of sialic acids on mucins and gels: A defense mechanism. *Biophys J* 102, 176-184.

Barker, N., van de Wetering, M., and Clevers, H. (2008). The intestinal stem cell. *Genes Dev* 22, 1856-1864.

Baumgart, M., Dogan, B., Rishniw, M., Weitzman, G., Bosworth, B., Yantiss, R., Orsi, R.H., Wiedmann, M., McDonough, P., Kim, S.G., *et al.* (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. *The ISME journal* 1, 403-418.

Becker, C., Neurath, M.F., and Wirtz, S. (2015). The intestinal microbiota in Inflammatory Bowel Disease. *ILAR J* 56, 192-204.

Belenguer, A., Duncan, S.H., Calder, A.G., Holtrop, G., Louis, P., Lobley, G.E., and Flint, H.J. (2006). Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. *Appl Environ Microbiol* 72, 3593-3599.

Bengmark, S. (1998). Ecological control of the gastrointestinal tract. The role of probiotic flora. *Gut* 42, 2-7.

Bennett, E.P., Mandel, U., Clausen, H., Gerken, T.A., Fritz, T.A., and Tabak, L.A. (2012). Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* 22, 736-756.

Bergstrom, K., Fu, J., Johansson, M.E., Liu, X., Gao, N., Wu, Q., Song, J., McDaniel, J.M., McGee, S., Chen, W., *et al.* (2016). Core 1- and 3-derived O-glycans collectively maintain the colonic mucus barrier and protect against spontaneous colitis in mice. *Mucosal immunology*.

Bergstrom, K.S.B., and Xia, L.J. (2013). Mucin-type O-glycans and their roles in intestinal homeostasis. *Glycobiology* 23, 1026-1037.

Berry, D., Kuzyk, O., Rauch, I., Heider, S., Schwab, C., Hainzl, E., Decker, T., Muller, M., Strobl, B., Schleper, C., *et al.* (2015). Intestinal microbiota signatures associated with inflammation history in mice experiencing recurring colitis. *Frontiers in microbiology* 6, 1408.

Bezirtzoglou, E., Tsiotsias, A., and Welling, G.W. (2011). Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe* 17, 478-482.

Biagi, E., Nylund, L., Candela, M., Ostan, R., Bucci, L., Pini, E., Nikkila, J., Monti, D., Satokari, R., Franceschi, C., *et al.* (2010). Through ageing, and beyond: Gut microbiota and inflammatory status in seniors and centenarians. *Plos One* 5, e10667.

Biedermann, L., Zeitz, J., Mwinyi, J., Sutter-Minder, E., Rehman, A., Ott, S.J., Steurer-Stey, C., Frei, A., Frei, P., Scharl, M., *et al.* (2013). Smoking cessation induces profound changes in the composition of the intestinal microbiota in humans. *Plos One* 8, e59260.

Bjedov, I., Tenaillon, O., Gerard, B., Souza, V., Denamur, E., Radman, M., Taddei, F., and Matic, I. (2003). Stress-induced mutagenesis in bacteria. *Science* 300, 1404-1409.

Blanton, L.V., Charbonneau, M.R., Salih, T., Barratt, M.J., Venkatesh, S., Ilkaveya, O., Subramanian, S., Manary, M.J., Trehan, I., Jorgensen, J.M., *et al.* (2016). Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science* 351.

Blixt, O., Bueti, D., Burford, B., Allen, D., Julien, S., Hollingsworth, M., Gammerman, A., Fentiman, I., Taylor-Papadimitriou, J., and Burchell, J.M. (2011). Autoantibodies to aberrantly glycosylated MUC1 in early stage breast cancer are associated with a better prognosis. *Breast Cancer Res* 13, R25.

Bollinger, R.R., Everett, M.L., Palestrant, D., Love, S.D., Lin, S.S., and Parker, W. (2003). Human secretory immunoglobulin A may contribute to biofilm formation in the gut. *Immunology* 109, 580-587.

Boyko, E.J., Perera, D.R., Koepsell, T.D., Keane, E.M., and Inui, T.S. (1988). Effects of cigarette smoking on the clinical course of ulcerative colitis. *Scandinavian journal of gastroenterology* 23, 1147-1152.

Brockhausen, I. (1999). Pathways of O-glycan biosynthesis in cancer cells. *Biochim Biophys Acta* 1473, 67-95.

Bruno, L.S., Li, X., Wang, L., Soares, R.V., Siqueira, C.C., Oppenheim, F.G., Troxler, R.F., and Offner, G.D. (2005). Two-hybrid analysis of human salivary mucin MUC7 interactions. *Biochim Biophys Acta* 1746, 65-72.

Bullock, N.R., Booth, J.C., and Gibson, G.R. (2004). Comparative composition of bacteria in the human intestinal microflora during remission and active ulcerative colitis. *Curr Issues Intest Microbiol* 5, 59-64.

Burger-van Paassen, N., Vincent, A., Puiman, P.J., van der Sluis, M., Bouma, J., Boehm, G., van Goudoever, J.B., Van Seuningen, I., and Renes, I.B. (2009). The regulation of intestinal mucin MUC2 expression by short-chain fatty acids: implications for epithelial protection. *The Biochemical journal* 420, 211-219.

Burisch, J., Jess, T., Martinato, M., Lakatos, P.L., and EpiCom, E. (2013). The burden of inflammatory bowel disease in Europe. *Journal of Crohns & Colitis* 7, 322-337.

Cadwell, K., Liu, J.Y., Brown, S.L., Miyoshi, H., Loh, J., Lennerz, J.K., Kishi, C., Kc, W., Carrero, J.A., Hunt, S., *et al.* (2008). A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. *Nature* 456, 259-263.

Campbell, B.J., Finnie, I.A., Hounsell, E.F., and Rhodes, J.M. (1995). Direct demonstration of increased expression of Thomsen-Friedenreich (TF) antigen in colonic adenocarcinoma and ulcerative-colitis mucin and its concealment in normal mucin. *J Clin Invest* 95, 571-576.

Campbell, B.J., Rowe, G.E., Leiper, K., and Rhodes, J.M. (2001a). Increasing the intra-Golgi pH of cultured LS174T goblet-differentiated cells mimics the decreased mucin sulfation and increased Thomsen-Friedenreich antigen (Gal beta 1-3GalNac alpha-) expression seen in colon cancer. *Glycobiology* 11, 385-393.

Campbell, B.J., Yu, L.G., and Rhodes, J.M. (2001b). Altered glycosylation in inflammatory bowel disease: A possible role in cancer development. *Glycoconjugate J* 18, 851-858.

Candela, M., Biagi, E., Maccaferri, S., Turrone, S., and Brigidi, P. (2012). Intestinal microbiota is a plastic factor responding to environmental changes. *Trends Microbiol* 20, 385-391.

Carding, S., Verbeke, K., Vipond, D.T., Corfe, B.M., and Owen, L.J. (2015). Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis* 26, 26191.

Cario, E., Gerken, G., and Podolsky, D.K. (2007). Toll-like receptor 2 controls mucosal inflammation by regulating epithelial barrier function. *Gastroenterology* 132, 1359-1374.

Carroll, I.M., Chang, Y.H., Park, J., Sartor, R.B., and Ringel, Y. (2010). Luminal and mucosal-associated intestinal microbiota in patients with diarrhea-predominant irritable bowel syndrome. *Gut Pathog* 2.

Cash, H.L., Whitham, C.V., Behrendt, C.L., and Hooper, L.V. (2006). Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* 313, 1126-1130.

Celli, J.P., Turner, B.S., Afdhal, N.H., Keates, S., Ghiran, I., Kelly, C.P., Ewoldt, R.H., McKinley, G.H., So, P., Erramilli, S., *et al.* (2009). *Helicobacter pylori* moves through mucus by reducing mucin viscoelasticity. *Proc Natl Acad Sci U S A* 106, 14321-14326.

Chachadi, V.B., Ali, M.F., and Cheng, P.W. (2013). Prostatic cell-specific regulation of the synthesis of MUC1-associated sialyl Lewis a. *Plos One* 8, e57416.

Chik, J.H.L., Zhou, J., Moh, E.S.X., Christopherson, R., Clarke, S.J., Molloy, M.P., and Packer, N.H. (2014). Comprehensive glycomics comparison between colon cancer cell cultures and tumours: Implications for biomarker studies. *J Proteomics* 108, 146-162.

Christensen, H.R., Frokiaer, H., and Pestka, J.J. (2002). Lactobacilli differentially modulate expression of cytokines and maturation surface markers in murine dendritic cells. *J Immunol* 168, 171-178.

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., *et al.* (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci U S A* 108, 4586-4591.

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., Harris, H.M.B., Coakley, M., Lakshminarayanan, B., O'Sullivan, O., *et al.* (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488, 178-184.

Claesson, M.J., Wang, Q., O'Sullivan, O., Greene-Diniz, R., Cole, J.R., Ross, R.P., and O'Toole, P.W. (2010). Comparison of two next-generation sequencing technologies for resolving

highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* 38, e200.

Clamp, J.R., Fraser, G., and Read, A.E. (1981). Study of the carbohydrate content of mucus glycoproteins from normal and diseased colons. *Clinical science* 61, 229-234.

Cohen, M., Varki, N.M., Jankowski, M.D., and Gagneux, P. (2012). Using unfixed, frozen tissues to study natural mucin distribution. *Journal of visualized experiments : JoVE*, e3928.

Collado, M.C., Derrien, M., Isolauri, E., de Vos, W.M., and Salminen, S. (2007). Intestinal integrity and *Akkermansia muciniphila*, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl Environ Microbiol* 73, 7767-7770.

Colombel, J.F., Watson, A.J., and Neurath, M.F. (2008). The 10 remaining mysteries of inflammatory bowel disease. *Gut* 57, 429-433.

Corfield, A.P. (2015). Mucins: a biologically relevant glycan barrier in mucosal protection. *Biochim Biophys Acta* 1850, 236-252.

Cornick, S., Tawiah, A., and Chadee, K. (2015). Roles and regulation of the mucus barrier in the gut. *Tissue Barriers* 3, e982426.

Correa-Oliveira, R., Fachi, J.L., Vieira, A., Sato, F.T., and Vinolo, M.A.R. (2016). Regulation of immune cell function by short-chain fatty acids. *Clin Transl Immunol* 5, e73.

Cosnes, J., Carbonnel, F., Beaugerie, L., Blain, A., Reijasse, D., and Gendre, J.P. (2002). Effects of appendicectomy on the course of ulcerative colitis. *Gut* 51, 803-807.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science* 326, 1694-1697.

Crosnier, C., Stamatakis, D., and Lewis, J. (2006). Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nat Rev Genet* 7, 349-359.

Crost, E.H., Pujol, A., Ladire, M., Dabard, J., Raibaud, P., Carlier, J.P., and Fons, M. (2010). Production of an antibacterial substance in the digestive tract involved in colonization-resistance against *Clostridium perfringens*. *Anaerobe* 16, 597-603.

Crost, E.H., Tailford, L.E., Le Gall, G., Fons, M., Henrissat, B., and Juge, N. (2013). Utilisation of mucin glycans by the human gut symbiont *Ruminococcus gnavus* is strain-dependent. *Plos One* 8, e76341.

Crost, E.H., Tailford, L.E., Monestier, M., Swarbreck, D., Henrissat, B., Crossman, L.C., and Juge, N. (2016). The mucin-degradation strategy of *Ruminococcus gnavus*: The importance of intramolecular trans-sialidases. *Gut microbes* 7, 1-11.

Darfeuille-Michaud, A., Boudeau, J., Bulois, P., Neut, C., Glasser, A.L., Barnich, N., Bringer, M.A., Swidsinski, A., Beaugerie, L., and Colombel, J.F. (2004). High prevalence of adherent-invasive *Escherichia coli* associated with ileal mucosa in Crohn's disease. *Gastroenterology* 127, 412-421.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G., and 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, 14691-14696.

de Souza, H.L., de Carvalho, V.R., Romeiro, F.G., Sasaki, L.Y., Keller, R., and Rodrigues, J. (2012). Mucosa-associated but not luminal *Escherichia coli* is augmented in Crohn's disease and ulcerative colitis. *Gut Pathog* 4, 21.

Dekker, J., Rossen, J.W., Buller, H.A., and Einerhand, A.W. (2002). The MUC family: an obituary. *Trends Biochem Sci* 27, 126-131.

Derrien, M., Belzer, C., and de Vos, W.M. (2016). *Akkermansia muciniphila* and its role in regulating host functions. *Microb Pathog*, 30178-30179.

Derrien, M., Van Baarlen, P., Hooiveld, G., Norin, E., Muller, M., and de Vos, W.M. (2011). Modulation of mucosal immune response, tolerance, and proliferation in mice colonized by the mucin-degrader *Akkermansia muciniphila*. *Frontiers in microbiology* 2, 166.

Derrien, M., van Passel, M.W., van de Bovenkamp, J.H., Schipper, R.G., de Vos, W.M., and Dekker, J. (2010). Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut microbes* 1, 254-268.

Derrien, M., Vaughan, E.E., Plugge, C.M., and de Vos, W.M. (2004). *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol* 54, 1469-1476.

Dethlefsen, L., and Relman, D.A. (2011). Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci U S A* 108, 4554-4561.

Dicksved, J., Halfvarson, J., Rosenquist, M., Jarnerot, G., Tysk, C., Apajalahti, J., Engstrand, L., and Jansson, J.K. (2008). Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *The ISME journal* 2, 716-727.

Ding, J.X., Xu, L.X., Zhu, L., Lv, J.C., Zhao, M.H., Zhang, H., and Wang, H.Y. (2009). Activity of alpha2,6-sialyltransferase and its gene expression in peripheral B lymphocytes in patients with IgA nephropathy. *Scand J Immunol* 69, 174-180.

Ding, T., and Schloss, P.D. (2014). Dynamics and associations of microbial community types across the human body. *Nature* 509, 357-360.

Donaldson, G.P., Lee, S.M., and Mazmanian, S.K. (2016). Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* 14, 20-32.

Donohoe, D.R., Collins, L.B., Wali, A., Bigler, R., Sun, W., and Bultman, S.J. (2012). The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. *Mol Cell* 48, 612-626.

Duncan, S.H., Barcenilla, A., Stewart, C.S., Pryde, S.E., and Flint, H.J. (2002a). Acetate utilization and butyryl coenzyme A (CoA):acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl Environ Microbiol* 68, 5186-5190.

Duncan, S.H., Hold, G.L., Harmsen, H.J., Stewart, C.S., and Flint, H.J. (2002b). Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 52, 2141-2146.

Duncan, S.H., Holtrop, G., Lobley, G.E., Calder, A.G., Stewart, C.S., and Flint, H.J. (2004a). Contribution of acetate to butyrate formation by human faecal bacteria. *Br J Nutr* 91, 915-923.

Duncan, S.H., Lobley, G.E., Holtrop, G., Ince, J., Johnstone, A.M., Louis, P., and Flint, H.J. (2008). Human colonic microbiota associated with diet, obesity and weight loss. *Int J Obesity* 32, 1720-1724.

Duncan, S.H., Louis, P., and Flint, H.J. (2004b). Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl Environ Microbiol* 70, 5810-5817.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., and Relman, D.A. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635-1638.

Eliakim, R., and Karmeli, F. (2003). Divergent effects of nicotine administration on cytokine levels in rat small bowel mucosa, colonic mucosa, and blood. *Israel Medical Association Journal* 5, 178-180.

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., *et al.* (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* 145, 745-757.

Ellekilde, M., Krych, L., Hansen, C.H., Hufeldt, M.R., Dahl, K., Hansen, L.H., Sorensen, S.J., Vogensen, F.K., Nielsen, D.S., and Hansen, A.K. (2014). Characterization of the gut microbiota in leptin deficient obese mice - Correlation to inflammatory and diabetic parameters. *Res Vet Sci* 96, 241-250.

Emerson, B.C., and Kolm, N. (2005). Species diversity can drive speciation. *Nature* 434, 1015-1017.

Enss, M.L., Cornberg, M., Wagner, S., Gebert, A., Henrichs, M., Eisenblatter, R., Beil, W., Kownatzki, R., and Hedrich, H.J. (2000). Proinflammatory cytokines trigger MUC gene expression and mucin release in the intestinal cancer cell line LS180. *Inflammation Res* 49, 162-169.

Erdem, A.L., Avelino, F., Xicohtencatl-Cortes, J., and Giron, J.A. (2007). Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing *Escherichia coli*. *J Bacteriol* 189, 7426-7435.

Ermund, A., Schutte, A., Johansson, M.E.V., Gustafsson, J.K., and Hansson, G.C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. I. Gastrointestinal mucus layers have different properties depending on location as well as over the Peyer's patches. *American journal of physiology Gastrointestinal and liver physiology* 305, G341-G347.

Etzold, S., Kober, O.I., Mackenzie, D.A., Tailford, L.E., Gunning, A.P., Walshaw, J., Hemmings, A.M., and Juge, N. (2014). Structural basis for adaptation of lactobacilli to gastrointestinal mucus. *Environ Microbiol* 16, 888-903.

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J.P., Druart, C., Bindels, L.B., Guiot, Y., Derrien, M., Muccioli, G.G., Delzenne, N.M., *et al.* (2013). Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci U S A* 110, 9066-9071.

Fantini, M.C., Rio, A., Fina, D., Caruso, R., Sarra, M., Stolfi, C., Becker, C., Macdonald, T.T., Pallone, F., Neurath, M.F., *et al.* (2009). Smad7 controls resistance of colitogenic T cells to regulatory T cell-mediated suppression. *Gastroenterology* 136, 1308-1316.

Favier, C.F., Vaughan, E.E., De Vos, W.M., and Akkermans, A.D.L. (2002). Molecular monitoring of succession of bacterial communities in human neonates. *Appl Environ Microbiol* 68, 219-226.

Flemstrom, G., and Kivilaakso, E. (1983). Demonstration of a pH gradient at the luminal surface of rat duodenum in vivo and its dependence on mucosal alkaline secretion. *Gastroenterology* 84, 787-794.

Forni, D., Cleyne, I., Ferrante, M., Cassinotti, A., Cagliani, R., Ardizzone, S., Vermeire, S., Fichera, M., Lombardini, M., Maconi, G., *et al.* (2014). ABO histo-blood group might modulate predisposition to Crohn's disease and affect disease behavior. *Journal of Crohns & Colitis* 8, 489-494.

Franceschi, S., Panza, E., Lavecchia, C., Parazzini, F., Decarli, A., and Porro, G.B. (1987). Nonspecific inflammatory bowel-disease and smoking. *Am J Epidem* 125, 445-452.

Franchimont, D., Vermeire, S., El Housni, H., Pierik, M., Van Steen, K., Gustot, T., Quertinmont, E., Abramowicz, M., Van Gossum, A., Deviere, J., *et al.* (2004). Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis. *Gut* 53, 987-992.

Frank, D.N., Amand, A.L.S., Feldman, R.A., Boedeker, E.C., Harpaz, N., and Pace, N.R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A* *104*, 13780-13785.

Franke, A., McGovern, D.P.B., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R., *et al.* (2010). Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* *42*, 1118-1125.

Freitas, M., Cayuela, C., Antoine, J.M., Piller, F., Sapin, C., and Trugnan, G. (2001). A heat labile soluble factor from *Bacteroides thetaiotaomicron* VPI-5482 specifically increases the galactosylation pattern of HT29-MTX cells. *Cell Microbiol* *3*, 289-300.

Freitas, M., Tavan, E., Cayuela, C., Diop, L., Sapin, C., and Trugnan, G. (2003). Host-pathogens cross-talk. Indigenous bacteria and probiotics also play the game. *Biol Cell* *95*, 503-506.

Friman, V., Adlerberth, I., Connell, H., Svanborg, C., Hanson, L.A., and Wold, A.E. (1996). Decreased expression of mannose-specific adhesins by *Escherichia coli* in the colonic microflora of immunoglobulin A-deficient individuals. *Infect Immun* *64*, 2794-2798.

Fu, J., Wei, B., Wen, T., Johansson, M.E.V., Liu, X., Bradford, E., Thomsson, K.A., McGee, S., Mansour, L., Tong, M., *et al.* (2011). Loss of intestinal core 1-derived O-glycans causes spontaneous colitis in mice. *J Clin Invest* *121*, 1657-1666.

Fukata, M., Michelsen, K.S., Eri, R., Thomas, L.S., Hu, B., Lukasek, K., Nast, C.C., Lechago, J., Xu, R.L., Naiki, Y., *et al.* (2005). Toll-like receptor-4 is required for intestinal response to epithelial injury and limiting bacterial translocation in a murine model of acute colitis. *American journal of physiology Gastrointestinal and liver physiology* *288*, G1055-G1065.

Fuller, Z., Louis, P., Mihajlovski, A., Rungapamestry, V., Ratcliffe, B., and Duncan, A.J. (2007). Influence of cabbage processing methods and prebiotic manipulation of colonic microflora on glucosinolate breakdown in man. *Br J Nutr* *98*, 364-372.

Furet, J.P., Kong, L.C., Tap, J., Poitou, C., Basdevant, A., Bouilliot, J.L., Mariat, D., Corthier, G., Dore, J., Henegar, C., *et al.* (2010). Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss links with metabolic and low-grade inflammation markers. *Diabetes* *59*, 3049-3057.

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T.A., Nakato, G., Takahashi, D., Nakanishi, Y., Uetake, C., Kato, K., Kato, T., *et al.* (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* 504, 446-450.

Garrett, W.S., Gordon, J.I., and Glimcher, L.H. (2010). Homeostasis and inflammation in the intestine. *Cell* 140, 859-870.

Gaudier, E., Jarry, A., Blottiere, H.M., de Coppet, P., Buisine, M.P., Aubert, J.P., Laboisse, C., Cherbut, C., and Hoebler, C. (2004). Butyrate specifically modulates MUC gene expression in intestinal epithelial goblet cells deprived of glucose. *American journal of physiology Gastrointestinal and liver physiology* 287, G1168-G1174.

Gendler, S.J., Lancaster, C.A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E.N., and Wilson, D. (1990). Molecular cloning and expression of human tumor-associated polymorphic epithelial mucin. *J Biol Chem* 265, 15286-15293.

Gewirtz, A.T., Navas, T.A., Lyons, S., Godowski, P.J., and Madara, J.L. (2001). Cutting edge: Bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. *J Immunol* 167, 1882-1885.

Gewirtz, A.T., Vijay-Kumar, M., Brant, S.R., Duerr, R.H., Nicolae, D.L., and Cho, J.H. (2006). Dominant-negative TLR5 polymorphism reduces adaptive immune response to flagellin and negatively associates with Crohn's disease. *American journal of physiology Gastrointestinal and liver physiology* 290, G1157-G1163.

Gibson, P.R., and Muir, J.G. (2005). Reinforcing the mucus: a new therapeutic approach for ulcerative colitis? *Gut* 54, 900-903.

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., and Nelson, K.E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science* 312, 1355-1359.

Godl, K., Johansson, M.E., Lidell, M.E., Morgelin, M., Karlsson, H., Olson, F.J., Gum, J.R., Jr., Kim, Y.S., and Hansson, G.C. (2002). The N terminus of the MUC2 mucin forms trimers that are held together within a trypsin-resistant core fragment. *J Biol Chem* 277, 47248-47256.

Goodrich, J.K., Waters, J.L., Poole, A.C., Sutter, J.L., Koren, O., Blehman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J.T., *et al.* (2014). Human genetics shape the gut microbiome. *Cell* 159, 789-799.

Gosalbes, M.J., Durban, A., Pignatelli, M., Abellan, J.J., Jimenez-Hernandez, N., Perez-Cobas, A.E., Latorre, A., and Moya, A. (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *Plos One* 6, e17447.

Graziani, F., Pujol, A., Nicoletti, C., Dou, S., Maresca, M., Giardina, T., Fons, M., and Perrier, J. (2016). *Ruminococcus gnavus* E1 modulates mucin expression and intestinal glycosylation. *J Appl Microbiol* 120, 1403-1417.

Gunning, A.P., Chambers, S., Pin, C., Man, A.L., Morris, V.J., and Nicoletti, C. (2008). Mapping specific adhesive interactions on living human intestinal epithelial cells with atomic force microscopy. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 22, 2331-2339.

Gunning, A.P., Kirby, A.R., Fuell, C., Pin, C., Tailford, L.E., and Juge, N. (2013). Mining the "glycocode"--exploring the spatial distribution of glycans in gastrointestinal mucin using force spectroscopy. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 27, 2342-2354.

Gustafsson, J.K., Ermund, A., Ambort, D., Johansson, M.E., Nilsson, H.E., Thorell, K., Hebert, H., Sjovall, H., and Hansson, G.C. (2012a). Bicarbonate and functional CFTR channel are required for proper mucin secretion and link cystic fibrosis with its mucus phenotype. *J Exp Med* 209, 1263-1272.

Gustafsson, J.K., Ermund, A., Johansson, M.E.V., Schutte, A., Hansson, G.C., and Sjovall, H. (2012b). An *ex vivo* method for studying mucus formation, properties, and thickness in human colonic biopsies and mouse small and large intestinal explants. *American journal of physiology Gastrointestinal and liver physiology* 302, G430-G438.

Hakansson, A., Tormo-Badia, N., Baridi, A., Xu, J., Molin, G., Hagslatt, M.L., Karlsson, C., Jeppsson, B., Cilio, C.M., and Ahrne, S. (2015). Immunological alteration and changes of gut microbiota after dextran sulfate sodium (DSS) administration in mice. *Clin Exp Med* 15, 107-120.

Halfvarson, J., Bodin, L., Tysk, C., Lindberg, E., and Jarnerot, G. (2003). Inflammatory bowel disease in a Swedish twin cohort: A long-term follow-up of concordance and clinical characteristics. *Gastroenterology* *124*, 1767-1773.

Hansen, C.H., Krych, L., Nielsen, D.S., Vogensen, F.K., Hansen, L.H., Sorensen, S.J., Buschard, K., and Hansen, A.K. (2012a). Early life treatment with vancomycin propagates *Akkermansia muciniphila* and reduces diabetes incidence in the NOD mouse. *Diabetologia* *55*, 2285-2294.

Hansen, R., Russell, R.K., Reiff, C., Louis, P., McIntosh, F., Berry, S.H., Mukhopadhyay, I., Bisset, W.M., Barclay, A.R., Bishop, J., *et al.* (2012b). Microbiota of de-novo pediatric IBD: Increased *Faecalibacterium prausnitzii* and reduced bacterial diversity in crohn's but not in ulcerative colitis. *Am J Gastroenterol* *107*, 1913-1922.

Hansson, G.C. (2012). Role of mucus layers in gut infection and inflammation. *Curr Opin Microbiol* *15*, 57-62.

Hasnain, S.Z., Tauro, S., Das, I., Tong, H., Chen, A.C.H., Jeffery, P.L., McDonald, V., Florin, T.H.J., and McGuckin, M.A. (2012). IL-10 Promotes Production of Intestinal Mucus by Suppressing Protein Misfolding and Endoplasmic Reticulum Stress in Goblet Cells. *Gastroenterology* *144*, 357-368.

Hasnain, S.Z., Wang, H., Ghia, J.-E., Haq, N., Deng, Y., Velcich, A., Grecis, R.K., Thornton, D.J., and Khan, W.I. (2010). Mucin gene deficiency in mice impairs host resistance to an enteric parasitic infection. *Gastroenterology* *138*, 1763-1771.

Hatayama, H., Washita, J., Kuwajima, A., and Abe, T. (2007). The short chain fatty acid, butyrate, stimulates MUC2 mucin production in the human colon cancer cell line, LS174T. *Biochem Biophys Res Commun* *356*, 599-603.

Hattori, M., and Taylor, T.D. (2009). The human intestinal microbiome: A new frontier of human biology. *DNA Res* *16*, 1-12.

Heller, F., Fuss, I.J., Nieuwenhuis, E.E., Blumberg, R.S., and Strober, W. (2002). Oxazolone colitis, a Th2 colitis model resembling ulcerative colitis, is mediated by IL-13-producing NK-T cells. *Immunity* *17*, 629-638.

Hensel, K.O., Boland, V., Postberg, J., Zilbauer, M., Heuschkel, R., Vogel, S., Godde, D., Wirth, S., and Jenke, A.C. (2014). Differential expression of mucosal trefoil factors and mucins in pediatric inflammatory bowel diseases. *Scientific reports* 4.

Herrmann, A., Davies, J.R., Lindell, G., Martensson, S., Packer, N.H., Swallow, D.M., and Carlstedt, I. (1999). Studies on the "insoluble" glycoprotein complex from human colon. Identification of reduction-insensitive MUC2 oligomers and C-terminal cleavage. *J Biol Chem* 274, 15828-15836.

Higuchi, L.M., Khalili, H., Chan, A.T., Richter, J.M., Bousvaros, A., and Fuchs, C.S. (2012). A prospective study of cigarette smoking and the risk of inflammatory bowel disease in women. *Am J Gastroenterol* 107, 1399-1406.

Hill, M.J. (1997). Intestinal flora and endogenous vitamin synthesis. *European Journal of Cancer Prevention* 6, S43-S45.

Hisamatsu, T., Suzuki, M., Reinecker, H.C., Nadeau, W.J., McCormick, B.A., and Podolsky, D.K. (2003). CARD15/NOD2 functions as an antibacterial factor in human intestinal epithelial cells. *Gastroenterology* 124, 993-1000.

Hoffmann, T.W., Pham, H.P., Bridonneau, C., Aubry, C., Lamas, B., Martin-Gallausiaux, C., Moroldo, M., Rainteau, D., Lapaque, N., Six, A., *et al.* (2016). Microorganisms linked to inflammatory bowel disease-associated dysbiosis differentially impact host physiology in gnotobiotic mice. *The ISME journal* 10, 460-477.

Hooper, L.V., Littman, D.R., and Macpherson, A.J. (2012). Interactions between the microbiota and the immune system. *Science* 336, 1268-1273.

Hooper, L.V., and Macpherson, A.J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nature reviews Immunology* 10, 159-169.

Hooper, L.V., Midtvedt, T., and Gordon, J.I. (2002). How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr* 22, 283-307.

Hooper, L.V., Xu, J., Falk, P.G., Midtvedt, T., and Gordon, J.I. (1999). A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc Natl Acad Sci U S A* 96, 9833-9838.

Huang, Y.L., Chassard, C., Hausmann, M., von Itzstein, M., and Hennet, T. (2015). Sialic acid catabolism drives intestinal inflammation and microbial dysbiosis in mice. *Nature communications* 6.

Huber, S., Gagliani, N., Esplugues, E., O'Connor, W., Jr., Huber, F.J., Chaudhry, A., Kamanaka, M., Kobayashi, Y., Booth, C.J., Rudensky, A.Y., *et al.* (2011). Th17 Cells Express Interleukin-10 Receptor and Are Controlled by Foxp3(-) and Foxp3(+) Regulatory CD4(+) T Cells in an Interleukin-10-Dependent Manner. *Immunity* 34, 554-565.

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

Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M., *et al.* (2001). Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411, 599-603.

Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res* 21, 1552-1560.

Hutter, J.L., and Bechhoefer, J. (1993). Calibration of atomic-force microscope tips. *Review of Scientific Instruments* 64, 1868-1873.

Inan, M.S., Rasoulpour, R.J., Yin, L., Hubbard, A.K., Rosenberg, D.W., and Giardina, C. (2000). The luminal short-chain fatty acid butyrate modulates NF-kappa B activity in a human colonic epithelial cell line. *Gastroenterology* 118, 724-734.

Iontcheva, I., Oppenheim, F.G., and Troxler, R.F. (1997). Human salivary mucin MG1 selectively forms heterotypic complexes with amylase, proline-rich proteins, statherin, and histatins. *J Dent Res* 76, 734-743.

Iwashita, J., Sato, Y., Sugaya, H., Takahashi, N., Sasaki, H., and Abe, T. (2003). mRNA of MUC2 is stimulated by IL-4, IL-13 or TNF-alpha through a mitogen-activated protein kinase pathway in human colon cancer cells. *Immunol Cell Biol* 81, 275-282.

Jakobsson, H.E., Abrahamsson, T.R., Jenmalm, M.C., Harris, K., Quince, C., Jernberg, C., Bjorksten, B., Engstrand, L., and Andersson, A.F. (2014). Decreased gut microbiota diversity,

delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut* *63*, 559-566.

Jakobsson, H.E., Jernberg, C., Andersson, A.F., Sjolund-Karlsson, M., Jansson, J.K., and Engstrand, L. (2010). Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *Plos One* *5*, e9836.

Jakobsson, H.E., Rodriguez-Pineiro, A.M., Schutte, A., Ermund, A., Boysen, P., Bemark, M., Sommer, F., Backhed, F., Hansson, G.C., and Johansson, M.E. (2015). The composition of the gut microbiota shapes the colon mucus barrier. *EMBO Rep* *16*, 164-177.

Jeffery, I.B., Claesson, M.J., O'Toole, P.W., and Shanahan, F. (2012). Categorization of the gut microbiota: enterotypes or gradients? *Nat Rev Microbiol* *10*, 591-592.

Jernberg, C., Lofmark, S., Edlund, C., and Jansson, J.K. (2007). Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *The ISME journal* *1*, 56-66.

Jiang, H., Ling, Z., Zhang, Y., Mao, H., Ma, Z., Yin, Y., Wang, W., Tang, W., Tan, Z., Shi, J., *et al.* (2015). Altered fecal microbiota composition in patients with major depressive disorder. *Brain Behavior and Immunity* *48*, 186-194.

Johansson, M.E., Gustafsson, J.K., Holmen-Larsson, J., Jabbar, K.S., Xia, L., Xu, H., Ghishan, F.K., Carvalho, F.A., Gewirtz, A.T., Sjovall, H., *et al.* (2013a). Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. *Gut* *63*, 281-291.

Johansson, M.E., Jakobsson, H.E., Holmen-Larsson, J., Schutte, A., Ermund, A., Rodriguez-Pineiro, A.M., Arike, L., Wising, C., Svensson, F., Backhed, F., *et al.* (2015). Normalization of host intestinal mucus layers requires long-term microbial colonization. *Cell host & microbe* *18*, 582-592.

Johansson, M.E., Larsson, J.M., and Hansson, G.C. (2011a). 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*, 4659-4665.

Johansson, M.E., Sjovall, H., and Hansson, G.C. (2013b). The gastrointestinal mucus system in health and disease. *Nature reviews Gastroenterology & hepatology* *10*, 352-361.

Johansson, M.E., Thomsson, K.A., and Hansson, G.C. (2009). Proteomic analyses of the two mucus layers of the colon barrier reveal that their main component, the Muc2 mucin, is strongly bound to the Fcgbp protein. *J Proteome Res* *8*, 3549-3557.

Johansson, M.E.V., Ambort, D., Pelaseyed, T., Schutte, A., Gustafsson, J.K., Ermund, A., Subramani, D.B., Holmen-Larsson, J.M., Thomsson, K.A., Bergstrom, J.H., *et al.* (2011b). Composition and functional role of the mucus layers in the intestine. *Cell Mol Life Sci* *68*, 3635-3641.

Johansson, M.E.V., Gustafsson, J.K., Sjoberg, K.E., Petersson, J., Holm, L., Sjovall, H., and Hansson, G.C. (2010). Bacteria Penetrate the Inner Mucus Layer before Inflammation in the Dextran Sulfate Colitis Model. *Plos One* *5*.

Johansson, M.E.V., Phillipson, M., Petersson, J., Velcich, A., Holm, L., and 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*, 15064-15069.

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., Vandamme, P., and Vermeire, S. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* *60*, 631-637.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., and Lee, J.C. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nat Cell Biol* *491*, 119-124.

Ju, T., and Cummings, R.D. (2002). A unique molecular chaperone Cosmc required for activity of the mammalian core 1 beta 3-galactosyltransferase. *Proc Natl Acad Sci U S A* *99*, 16613-16618.

Juge, N. (2012). Microbial adhesins to gastrointestinal mucus. *Trends Microbiol* *20*, 30-39.

Julien, S., Videira, P.A., and Delannoy, P. (2012). Sialyl-tn in cancer: (how) did we miss the target? *Biomolecules* *2*, 435-466.

Jump, R.L., and Levine, A.D. (2004). Mechanisms of natural tolerance in the intestine: implications for inflammatory bowel disease. *Inflamm Bowel Dis* *10*, 462-478.

Jumpertz, R., Duc Son, L., Turnbaugh, P.J., Trinidad, C., Bogardus, C., Gordon, J.I., and Krakoff, J. (2011). Energy-balance studies reveal associations between gut microbes, caloric load, and nutrient absorption in humans. *The American journal of clinical nutrition* *94*, 58-65.

Kadivar, K., Ruchelli, E.D., Markowitz, J.E., DeFelice, M.L., Strogatz, M.L., Kanzaria, M.M., Reddy, K.P., Baldassano, R.N., von Allmen, D., and Brown, K.A. (2004). Intestinal interleukin-13 in pediatric inflammatory bowel disease patients. *Inflamm Bowel Dis* *10*, 593-598.

Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and McSweeney, C.S. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* *16*, 2034-2042.

Kassinen, A., Krogus-Kurikka, L., Makivuokko, H., Rinttila, T., Paulin, L., Corander, J., Malinen, E., Apajalahti, J., and Palva, A. (2007). The fecal microbiota of irritable bowel syndrome patients differs significantly from that of healthy subjects. *Gastroenterology* *133*, 24-33.

Kato, K., Ishii, Y., Mizuno, S., Sugitani, M., Asai, S., Kohno, T., Takahashi, K., Komuro, S., Iwamoto, M., Miyamoto, S., *et al.* (2007). Usefulness of rectally administering [1-(13)C]-butyrate for breath test in patients with active and quiescent ulcerative colitis. *Scandinavian journal of gastroenterology* *42*, 207-214.

Kau, A.L., Planer, J.D., Liu, J., Rao, S., Yatsunencko, T., Trehan, I., Manary, M.J., Liu, T.C., Stappenbeck, T.S., Maleta, K.M., *et al.* (2015). Functional characterization of IgA-targeted bacterial taxa from undernourished Malawian children that produce diet-dependent enteropathy. *Science translational medicine* *7*, 276ra224.

Kemmner, W., Roefzaad, C., Haensch, W., and Schlag, P.M. (2003). Glycosyltransferase expression in human colonic tissue examined by oligonucleotide arrays. *Biochim Biophys Acta* *1621*, 272-279.

Kesimer, M., Ehre, C., Burns, K.A., Davis, C.W., Sheehan, J.K., and Pickles, R.J. (2013). Molecular organization of the mucins and glycocalyx underlying mucus transport over mucosal surfaces of the airways. *Mucosal immunology* *6*, 379-392.

Kesimer, M., Makhov, A.M., Griffith, J.D., Verdugo, P., and Sheehan, J.K. (2010). Unpacking a gel-forming mucin: a view of MUC5B organization after granular release. *American journal of physiology Lung cellular and molecular physiology* *298*, L15-22.

Kesimer, M., and Sheehan, J.K. (2012). Mass spectrometric analysis of mucin core proteins. *Methods in molecular biology* 842, 67-79.

Kinross, J.M., Darzi, A.W., and Nicholson, J.K. (2011). Gut microbiome-host interactions in health and disease. *Genome medicine* 3, 14.

Kivilaakso, E., and Flemstrom, G. (1984). Surface pH gradient in gastroduodenal mucosa. *Scand J Gastroenterol Suppl* 105, 50-52.

Knoop, K.A., McDonald, K.G., McCrate, S., McDole, J.R., and Newberry, R.D. (2015). Microbial sensing by goblet cells controls immune surveillance of luminal antigens in the colon. *Mucosal immunology* 8, 198-210.

Kobayashi, K.S., Chamaillard, M., Ogura, Y., Henegariu, O., Inohara, N., Nunez, G., and Flavell, R.A. (2005). Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* 307, 731-734.

Kober, O.I., Ahl, D., Pin, C., Holm, L., Carding, S.R., and Juge, N. (2014).  $\gamma\delta$  T-cell-deficient mice show alterations in mucin expression, glycosylation, and goblet cells but maintain an intact mucus layer. *American journal of physiology Gastrointestinal and liver physiology* 306, G582-G593.

Koenig, J.E., Spor, A., Scalfone, N., Fricker, A.D., Stombaugh, J., Knight, R., Angenent, L.T., and Ley, R.E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* 108 4578-4585.

Kosiewicz, M.M., Nast, C.C., Krishnan, A., Rivera-Nieves, J., Moskaluk, C.A., Matsumoto, S., Kozaiwa, K., and Cominelli, F. (2001). Th1-type responses mediate spontaneous ileitis in a novel murine model of Crohn's disease. *J Clin Invest* 107, 695-702.

Kotlarz, D., Beier, R., Murugan, D., Diestelhorst, J., Jensen, O., Boztug, K., Pfeifer, D., Kreipe, H., Pfister, E.D., Baumann, U., *et al.* (2012). Loss of interleukin-10 signaling and infantile inflammatory bowel disease: implications for diagnosis and therapy. *Gastroenterology* 143, 347-355.

Kripke, S.A., Fox, A.D., Berman, J.M., Settle, R.G., and Rombeau, J.L. (1989). Stimulation of intestinal mucosal growth with intracolonic infusion of short-chain fatty-acids. *Journal of Parenteral and Enteral Nutrition* 13, 109-116.

Krogus-Kurikka, L., Lyra, A., Malinen, E., Aarnikunnas, J., Tuimala, J., Paulin, L., Makivuokko, H., Kajander, K., and Palva, A. (2009). Microbial community analysis reveals high level phylogenetic alterations in the overall gastrointestinal microbiota of diarrhoea-predominant irritable bowel syndrome sufferers. *BMC Gastroenterol* 9, 11.

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

Kyo, K., Parkes, M., Takei, Y., Nishimori, H., Vyas, P., Satsangi, J., Simmons, J., Nagawa, H., Baba, S., Jewell, D., *et al.* (1999). Association of ulcerative colitis with rare VNTR alleles of the human intestinal mucin gene, MUC3. *Hum Mol Genet* 8, 307-311.

Lakatos, L., Pandur, T., David, G., Balogh, Z., Kuronya, P., Tollas, A., and Lakatos, P.L. (2003). Association of extraintestinal manifestations of inflammatory bowel disease in a province of western Hungary with disease phenotype: Results of a 25-year follow-up study. *World J Gastroenterol* 9, 2300-2307.

Lang, T., Hansson, G.C., and Samuelsson, T. (2007). Gel-forming mucins appeared early in metazoan evolution. *Proc Natl Acad Sci U S A* 104, 16209-16214.

Larsen, N., Vogensen, F.K., van den Berg, F.W.J., Nielsen, D.S., Andreasen, A.S., Pedersen, B.K., Abu Al-Soud, W., Sorensen, S.J., Hansen, L.H., and Jakobsen, M. (2010). Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *Plos One* 5, e9085.

Larsson, J.M.H., Karlsson, H., Crespo, J.G., Johansson, M.E.V., Eklund, L., Sjoval, H., and Hansson, G.C. (2011). Altered o-glycosylation profile of MUC2 mucin occurs in active ulcerative colitis and is associated with increased inflammation. *Inflamm Bowel Dis* 17, 2299-2307.

Larsson, J.M.H., Karlsson, H., Sjoval, H., and Hansson, G.C. (2009). A complex, but uniform O-glycosylation of the human MUC2 mucin from colonic biopsies analyzed by nanoLC/MSn. *Glycobiology* 19, 756-766.

Larsson, J.M.H., Thomsson, K.A., Rodriguez-Pineiro, A.M., Karlsson, H., and Hansson, G.C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. *American journal of physiology Gastrointestinal and liver physiology* 305, G357-G363.

Lavelle, A., Lennon, G., O'Sullivan, O., Docherty, N., Balfe, A., Maguire, A., Mulcahy, H.E., Doherty, G., O'Donoghue, D., Hyland, J., *et al.* (2015). Spatial variation of the colonic microbiota in patients with ulcerative colitis and control volunteers. *Gut* *64*, 1553-1561.

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J.M., Kennedy, S., *et al.* (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature* *500*, 541-546.

Lees, C.W., Barrett, J.C., Parkes, M., and Satsangi, J. (2011). New IBD genetics: common pathways with other diseases. *Gut* *60*, 1739-1753.

Lepage, P., Haesler, R., Spehlmann, M.E., Rehman, A., Zvirbliene, A., Begun, A., Ott, S., Kupcinskis, L., Dore, J., Raedler, A., *et al.* (2011). Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* *141*, 227-236.

Lesage, S., Zouali, H., Cezard, J.P., Colombel, J.F., Belaiche, J., Almer, S., Tysk, C., O'Morain, C., Gassull, M., Binder, V., *et al.* (2002). CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Gen* *70*, 845-857.

Levine, B., and Kroemer, G. (2008). Autophagy in the pathogenesis of disease. *Cell* *132*, 27-42.

Ley, R.E., Backhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D., and Gordon, J.I. (2005). Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* *102*, 11070-11075.

Ley, R.E., Peterson, D.A., and Gordon, J.I. (2006a). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* *124*, 837-848.

Ley, R.E., Turnbaugh, P.J., Klein, S., and Gordon, J.I. (2006b). Microbial ecology - Human gut microbes associated with obesity. *Nature* *444*, 1022-1023.

Li, H., Limenitakis, J.P., Fuhrer, T., Geuking, M.B., Lawson, M.A., Wyss, M., Brugiroux, S., Keller, I., Macpherson, J.A., Rupp, S., *et al.* (2015). The outer mucus layer hosts a distinct intestinal microbial niche. *Nature Commun* *6*.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J.R., Prifti, E., Nielsen, T., *et al.* (2014). An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 32, 834-841.

Li, X., LeBlanc, J., Truong, A., Vuthoori, R., Chen, S.S., Lustgarten, J.L., Roth, B., Allard, J., Ippoliti, A., Presley, L.L., *et al.* (2011). A metaproteomic approach to study human-microbial ecosystems at the mucosal luminal interface. *Plos One* 6, e26542.

Lidell, M.E., Johansson, M.E., Morgelin, M., Asker, N., Gum, J.R., Jr., Kim, Y.S., and Hansson, G.C. (2003). The recombinant C-terminus of the human MUC2 mucin forms dimers in Chinese-hamster ovary cells and heterodimers with full-length MUC2 in LS 174T cells. *The Biochemical journal* 372, 335-345.

Lillehoj, E.P., Kim, H., Chun, E.Y., and Kim, K.C. (2004). *Pseudomonas aeruginosa* stimulates phosphorylation of the airway epithelial membrane glycoprotein Muc1 and activates MAP kinase. *American journal of physiology Lung cellular and molecular physiology* 287, L809-815.

Lindberg, E., Jarnerot, G., and Huitfeldt, B. (1992). Smoking in Crohns-disease - effect on localization and clinical course. *Gut* 33, 779-782.

Linden, S.K., Sheng, Y.H., Every, A.L., Miles, K.M., Skoog, E.C., Florin, T.H., Sutton, P., and McGuckin, M.A. (2009). MUC1 limits *Helicobacter pylori* infection both by steric hindrance and by acting as a releasable decoy. *PLoS Pathog* 5, e1000617.

Linden, S.K., Sutton, P., Karlsson, N.G., Korolik, V., and McGuckin, M.A. (2008). Mucins in the mucosal barrier to infection. *Mucosal immunology* 1, 183-197.

Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., *et al.* (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 47, 979-986.

Liu, Y., Vankruiningen, H.J., West, A.B., Cartun, R.W., Cortot, A., and Colombel, J.F. (1995). Immunocytochemical evidence of *Listeria*, *Escherichia coli*, and *Streptococcus* antigens in Crohns-disease. *Gastroenterology* 108, 1396-1404.

Liu, Z., Yadav, P.K., Xu, X., Su, J., Chen, C., Tang, M., Lin, H., Yu, J., Qian, J., Yang, P.-C., *et al.* (2011). The increased expression of IL-23 in inflammatory bowel disease promotes

intraepithelial and lamina propria lymphocyte inflammatory responses and cytotoxicity. *J Leukocyte Biol* 89, 597-606.

Loftus, E.V., Jr. (2004). Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 126, 1504-1517.

Logan, R.F.A., Edmond, M., Somerville, K.W., and Langman, M.J.S. (1984). Smoking and ulcerative-colitis. *Br Med J* 288, 751-753.

Loktionov, A., Chhaya, V., Bandaletova, T., and Poullis, A. (2016). Assessment of cytology and mucin 2 in colorectal mucus collected from patients with inflammatory bowel disease: Results of a pilot trial. *J Gastroen Hepatol* 31, 326-333.

Lopez-Siles, M., Martinez-Medina, M., Busquets, D., Sabat-Mir, M., Duncan, S.H., Flint, H.J., Aldeguer, X., and Jesus Garcia-Gil, L. (2014). Mucosa-associated Faecalibacterium prausnitzii and Escherichia coli co-abundance can distinguish Irritable Bowel Syndrome and Inflammatory Bowel Disease phenotypes. *Int J Med Microbiol* 304, 464-475.

Louis, P., and Flint, H.J. (2009). Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol Lett* 294, 1-8.

Louis, P., Hold, G.L., and Flint, H.J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* 12, 661-672.

Louis, P., Scott, K.P., Duncan, S.H., and Flint, H.J. (2007). Understanding the effects of diet on bacterial metabolism in the large intestine. *J Appl Microbiol* 102, 1197-1208.

Lucke, K., Miehke, S., Jacobs, E., and Schuppler, M. (2006). Prevalence of Bacteroides and Prevotella spp. in ulcerative colitis. *J Med Microbiol* 55, 617-624.

Luckey, T.D. (1972). Introduction to intestinal microecology. *The American journal of clinical nutrition* 25, 1292-1294.

Macao, B., Johansson, D.G., Hansson, G.C., and Hard, T. (2006). Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nat Struct Mol Biol* 13, 71-76.

Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijis, I., Eeckhaut, V., Ballet, V., Claes, K., Van Immerseel, F., Verbeke, K., *et al.* (2014). A decrease of the butyrate-producing species

Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* 63, 1275-1283.

Macpherson, A.J., Gatto, D., Sainsbury, E., Harriman, G.R., Hengartner, H., and Zinkernagel, R.M. (2000). A primitive T cell-independent mechanism of intestinal mucosal IgA responses to commensal bacteria. *Science* 288, 2222-2226.

Macpherson, A.J., and Uhr, T. (2004). Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science* 303, 1662-1665.

Makivuokko, H., Lahtinen, S.J., Wacklin, P., Tuovinen, E., Tenkanen, H., Nikkila, J., Bjorklund, M., Aranko, K., Ouwehand, A.C., and Matto, J. (2012). Association between the ABO blood group and the human intestinal microbiota composition. *BMC Microbiol* 12, 94.

Malinen, E., Krogius-Kurikka, L., Lyra, A., Nikkila, J., Jaaskelainen, A., Rinttila, T., Vilpponen-Salmela, T., von Wright, A.J., and Palva, A. (2010). Association of symptoms with gastrointestinal microbiota in irritable bowel syndrome. *World J Gastroenterol* 16, 4532-4540.

Malinen, E., Rinttila, T., Kajander, K., Matto, J., Kassinen, A., Krogius, L., Saarela, M., Korpela, R., and Palva, A. (2005). Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. *Am J Gastroenterol* 100, 373-382.

Maloy, K.J., Salaun, L., Cahill, R., Dougan, G., Saunders, N.J., and Powrie, F. (2003). CD4(+)CD25(+) T-R cells suppress innate immune pathology through cytokine-dependent mechanisms. *J Exp Med* 197, 111-119.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., *et al.* (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55, 205-211.

Marchiando, A.M., Shen, L., Graham, W.V., Edelblum, K.L., Duckworth, C.A., Guan, Y., Montrose, M.H., Turner, J.R., and Watson, A.J. (2011). The epithelial barrier is maintained by in vivo tight junction expansion during pathologic intestinal epithelial shedding. *Gastroenterology* 140, 1208-1218.

Marcobal, A., Barboza, M., Sonnenburg, E.D., Pudlo, N., Martens, E.C., Desai, P., Lebrilla, C.B., Weimer, B.C., Mills, D.A., German, J.B., *et al.* (2011). Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell host & microbe* *10*, 507-514.

Marcos, N.T., Bennett, E.P., Gomes, J., Magalhaes, A., Gomes, C., David, L., Dar, I., Jeanneau, C., DeFrees, S., Krusturup, D., *et al.* (2011). ST6GalNAc-I controls expression of sialyl-Tn antigen in gastrointestinal tissues. *Front Biosci (Elite Ed)* *3*, 1443-1455.

Marcos, N.T., Pinho, S., Grandela, C., Cruz, A., Samyn-Petit, B., Harduin-Lepers, A., Almeida, R., Silva, F., Morais, V., Costa, J., *et al.* (2004). Role of the human ST6GalNAc-I and ST6GalNAc-II in the synthesis of the cancer-associated sialyl-Tn antigen. *Cancer Res* *64*, 7050-7057.

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

Martinez, C., Antolin, M., Santos, J., Torrejon, A., Casellas, F., Borruel, N., Guarner, F., and Malagelada, J.R. (2008). Unstable composition of the fecal microbiota in ulcerative colitis during clinical remission. *Am J Gastroenterol* *103*, 643-648.

Mastrodonato, M., Mentino, D., Liquori, G.E., and Ferri, D. (2012). Histochemical characterization of the sialic acid residues in mouse colon mucins. *Microsc Res Tech* *76*, 156-162.

McDole, J.R., Wheeler, L.W., McDonald, K.G., Wang, B., Konjufca, V., Knoop, K.A., Newberry, R.D., and Miller, M.J. (2012). Goblet cells deliver luminal antigen to CD103+ dendritic cells in the small intestine. *Nature* *483*, 345-349.

McFall-Ngai, M. (2006). Love the one you're with: vertebrate guts shape their microbiota. *Cell* *127*, 247-249.

McGovern, D.P., Jones, M.R., Taylor, K.D., Marciante, K., Yan, X., Dubinsky, M., Ippoliti, A., Vasilias, E., Berel, D., Derkowski, C., *et al.* (2010). Fucosyltransferase 2 (FUT2) non-secretor status is associated with Crohn's disease. *Hum Mol Genet* *19*, 3468-3476.

McGuckin, M.A., Linden, S.K., Sutton, P., and Florin, T.H. (2011). Mucin dynamics and enteric pathogens. *Nat Rev Microbiol* *9*, 265-278.

McNamara, N., and Basbaum, C. (2001). Signaling networks controlling mucin production in response to Gram-positive and Gram-negative bacteria. *Glycoconjugate J* *18*, 715-722.

Meyer-Hoffert, U., Hornef, M.W., Henriques-Normark, B., Axelsson, L.G., Midtvedt, T., Putsep, K., and Andersson, M. (2008). Secreted enteric antimicrobial activity localises to the mucus surface layer. *Gut* 57, 764-771.

Micots, I., Augeron, C., Laboisse, C.L., Muzeau, F., and Megraud, F. (1993). Mucin exocytosis - a major target for *Helicobacter pylori*. *J Clin Pathol* 46, 241-245.

Mihalache, A., Delplanque, J.F., Ringot-Destrez, B., Wavelet, C., Gosset, P., Nunes, B., Groux-Degroote, S., Léonard, R., and Robbe-Masselot, C. (2015). Structural characterization of mucin O-glycosylation may provide important information to help prevent colorectal tumor recurrence. *Front Oncol* 5, 217.

Mikkelsen, A., Stokke, B.T., Christensen, B.E., and Elgsaeter, A. (1985). Flexibility and length of human bronchial mucin studied using low-shear viscometry, birefringence relaxation analysis, and electron microscopy. *Biopolymers* 24, 1683-1704.

Miller, M.J., Knoop, K.A., and Newberry, R.D. (2014). Mind the GAPS: insights into intestinal epithelial barrier maintenance and luminal antigen delivery. *Mucosal immunology* 7, 452-454.

Miyoshi, J., and Chang, E.B. (2016). The gut microbiota and inflammatory bowel diseases. *Transl Res* 37, 47-55.

Mizrahi-Man, O., Davenport, E.R., and Gilad, Y. (2013). Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: Evaluation of effective study designs. *Plos One* 8, e53608.

Moehle, C., Ackermann, N., Langmann, T., Aslanidis, C., Kel, A., Kel-Margoullis, O., Schmitz-Madry, A., Zahn, A., Stremmel, W., and Schmitz, G. (2006). Aberrant intestinal expression and allelic variants of mucin genes associated with inflammatory bowel disease. *J Mol Med-Jmm* 84, 1055-1066.

Monteleone, G., Kumberova, A., Croft, N.M., McKenzie, C., Steer, H.W., and MacDonald, T.T. (2001). Blocking Smad7 restores TGF-beta 1 signaling in chronic inflammatory bowel disease. *J Clin Invest* 108, 601-609.

Moore, W.E.C., and Holdeman, L.V. (1974). Human fecal flora - normal flora of 20 Japanese-Hawaiians. *Applied Microbiology* 27, 961-979.

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., *et al.* (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 13.

Murphy, E.F., Cotter, P.D., Healy, S., Marques, T.M., O'Sullivan, O., Fouhy, F., Clarke, S.F., O'Toole, P.W., Quigley, E.M., Stanton, C., *et al.* (2010). Composition and energy harvesting capacity of the gut microbiota: relationship to diet, obesity and time in mouse models. *Gut* 59, 1635-1642.

Nava, G.M., Friedrichsen, H.J., and Stappenbeck, T.S. (2011). Spatial organization of intestinal microbiota in the mouse ascending colon. *The ISME journal* 5, 627-638.

Neish, A.S. (2009). Microbes in gastrointestinal health and disease. *Gastroenterology* 136, 65-80.

Nell, S., Suerbaum, S., and Josenhans, C. (2010). The impact of the microbiota on the pathogenesis of IBD: lessons from mouse infection models. *Nat Rev Microbiol* 8, 564-577.

Nemoto, H., Kataoka, K., Ishikawa, H., Ikata, K., Arimochi, H., Iwasaki, T., Ohnishi, Y., Kuwahara, T., and Yasutomo, K. (2012). Reduced diversity and imbalance of fecal microbiota in patients with ulcerative colitis. *Dig Dis Sci* 57, 2955-2964.

Neurath, M.F. (2014). Cytokines in inflammatory bowel disease. *Nature reviews Immunology* 14, 329-342.

Ng, K.M., Ferreyra, J.A., Higginbottom, S.K., Lynch, J.B., Kashyap, P.C., Gopinath, S., Naidu, N., Choudhury, B., Weimer, B.C., Monack, D.M., *et al.* (2013). Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* 502, 96-99.

Ng, S.C., Benjamin, J.L., McCarthy, N.E., Hedin, C.R.H., Koutsoumpas, A., Plamondon, S., Price, C.L., Hart, A.L., Kamm, M.A., Forbes, A., *et al.* (2011). Relationship between human intestinal dendritic cells, gut microbiota, and disease activity in crohn's disease. *Inflamm Bowel Dis* 17, 2027-2037.

Nguyen, N.-P., Warnow, T., Pop, M., and White, B. (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *Npj Biofilms And Microbiomes* 2.

Nilsson, H.E., Ambort, D., Backstrom, M., Thomsson, E., Koeck, P.J.B., Hansson, G.C., and Hebert, H. (2014). Intestinal MUC2 mucin supramolecular topology by packing and release resting on D3 domain assembly. *J Mol Biol* 426, 2567-2579.

Nishikawa, J., Kudo, T., Sakata, S., Benno, Y., and Sugiyama, T. (2009). Diversity of mucosa-associated microbiota in active and inactive ulcerative colitis. *Scandinavian journal of gastroenterology* 44, 180-186.

Noble, C.L., Nimmo, E.R., Drummond, H., Ho, G.T., Tenesa, A., Smith, L., Anderson, N., Arnott, I.D.R., and Satsangi, J. (2005). The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* 129, 1854-1864.

Noor, S.O., Ridgway, K., Scovell, L., Kemsley, E.K., Lund, E.K., Jamieson, C., Johnson, I.T., and Narbad, A. (2010). Ulcerative colitis and irritable bowel patients exhibit distinct abnormalities of the gut microbiota. *BMC Gastroenterol* 10, 134.

Odes, H.S., Fich, A., Reif, S., Halak, A., Lavy, A., Keter, D., Eliakim, R., Paz, J., Broide, E., Niv, Y., *et al.* (2001). Effects of current cigarette smoking on clinical course of Crohn's disease and ulcerative colitis. *Dig Dis Sci* 46, 1717-1721.

Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H., *et al.* (2001a). A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411, 603-606.

Ogura, Y., Inohara, N., Benito, A., Chen, F.F., Yamaoka, S., and Nunez, G. (2001b). Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappa B. *J Biol Chem* 276, 4812-4818.

Orholm, M., Munkholm, P., Langholz, E., Nielsen, O.H., Sorensen, T.I.A., and Binder, V. (1991). Familial occurrence of inflammatory bowel disease. *New Engl J Med* 324, 84-88.

Ott, S.J., Musfeldt, M., Wenderoth, D.F., Hampe, J., Brant, O., Folsch, U.R., Timmis, K.N., and Schreiber, S. (2004). Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 53, 685-693.

Ottman, N., Smidt, H., de Vos, W.M., and Belzer, C. (2012). The function of our microbiota: who is out there and what do they do? *Front Cell Infect Microbiol* 2, 104.

Owczarek, D., Rodacki, T., Domagala-Rodacka, R., Cibor, D., and Mach, T. (2016). Diet and nutritional factors in inflammatory bowel diseases. *World J Gastroenterol* 22, 895-905.

Palmai-Pallag, T., Khodabukus, N., Kinarsky, L., Leir, S.H., Sherman, S., Hollingsworth, M.A., and Harris, A. (2005). The role of the SEA (sea urchin sperm protein, enterokinase and agrin) module in cleavage of membrane-tethered mucins. *The FEBS journal* 272, 2901-2911.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* 5, e177.

Panwala, C.M., Jones, J.C., and 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, 5733-5744.

Parmar, A.S., Alakulppi, N., Paavola-Sakki, P., Kurppa, K., Halme, L., Farkkila, M., Turunen, U., Lappalainen, M., Kontula, K., Kaukinen, K., *et al.* (2012). Association study of FUT2 (rs601338) with celiac disease and inflammatory bowel disease in the Finnish population. *Tissue Antigens* 80, 488-493.

Parronchi, P., Romagnani, P., Annunziato, F., Sampognaro, S., Becchio, A., Giannarini, L., Maggi, E., Pupilli, C., Tonelli, F., and Romagnani, S. (1997). Type 1 T-helper cell predominance and interleukin-12 expression in the gut of patients with Crohn's disease. *Am J Pathol* 150, 823-832.

Pelaseyed, T., Bergstrom, J.H., Gustafsson, J.K., Ermund, A., Birchenough, G.M.H., Schutte, A., van der Post, S., Svensson, F., Rodriguez-Pineiro, A.M., Nystrom, E.E.L., *et al.* (2014). The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol Rev* 260, 8-20.

Penders, J., Thijs, C., Vink, C., Stelma, F.F., Snijders, B., Kummeling, I., van den Brandt, P.A., and Stobberingh, E.E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 118, 511-521.

Perez-Vilar, J., and Hill, R.L. (1999). The structure and assembly of secreted mucins. *J Biol Chem* 274, 31751-31754.

Petersson, J., Schreiber, O., Hansson, G.C., Gendler, S.J., Velcich, A., Lundberg, J.O., Roos, S., Holm, L., and Phillipson, M. (2011). Importance and regulation of the colonic mucus barrier

in a mouse model of colitis. *American journal of physiology Gastrointestinal and liver physiology* *300*, G327-333.

Png, C.W., Linden, S.K., Gilshenan, K.S., Zoetendal, E.G., McSweeney, C.S., Sly, L.I., McGuckin, M.A., and Florin, T.H.J. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol* *105*, 2420-2428.

Poretzky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *Plos One* *9*, e93827.

Potten, C.S. (1990). A comprehensive study of the radiobiological response of the murine (BDF1) small intestine. *Int J Radiat Biol* *58*, 925-973.

Powrie, F., Leach, M.W., Mauze, S., Caddle, L.B., and Coffman, R.L. (1993). Phenotypically distinct subsets of CD4<sup>+</sup> T cells induce or protect from chronic intestinal inflammation in C. B-17 scid mice. *Int Immunol* *5*, 1461-1471.

Prindiville, T., Cantrell, M., and Wilson, K.H. (2004). Ribosomal DNA sequence analysis of mucosa-associated bacteria in Crohn's disease. *Inflamm Bowel Dis* *10*, 824-833.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Gloeckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* *35*, 7188-7196.

Quigley, E.M., and Turnberg, L.A. (1987). pH of the microclimate lining human gastric and duodenal mucosa in vivo. Studies in control subjects and in duodenal ulcer patients. *Gastroenterology* *92*, 1876-1884.

Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R. (2004). Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* *118*, 229-241.

Ramirez-Farias, C., Slezak, K., Fuller, Z., Duncan, A., Holtrop, G., and Louis, P. (2009). Effect of inulin on the human gut microbiota: stimulation of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*. *Br J Nutr* *101*, 541-550.

Rausch, P., Rehman, A., Kunzel, S., Hasler, R., Ott, S.J., Schreiber, S., Rosenstiel, P., Franke, A., and Baines, J.F. (2011). Colonic mucosa-associated microbiota is influenced by an

interaction of Crohn disease and FUT2 (Secretor) genotype. *Proc Natl Acad Sci U S A* *108*, 19030-19035.

Rawls, J.F., Mahowald, M.A., Ley, R.E., and Gordon, J.I. (2006). Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* *127*, 423-433.

Read, S., Malmstrom, V., and Powrie, F. (2000). Cytotoxic T lymphocyte-associated antigen 4 plays an essential role in the function of CD25(+)CD4(+) regulatory cells that control intestinal inflammation. *J Exp Med* *192*, 295-302.

Reichardt, N., Duncan, S.H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K.P., Flint, H.J., and Louis, P. (2014). Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME journal* *8*, 1323-1335.

Reunanen, J., Kainulainen, V., Huuskonen, L., Ottman, N., Belzer, C., Huhtinen, H., de Vos, W.M., and Satokari, R. (2015). *Akkermansia muciniphila* adheres to enterocytes and strengthens the integrity of the epithelial cell layer. *Appl Environ Microbiol* *81*, 3655-3662.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J.Y., Duncan, A.E., Kau, A.L., Griffin, N.W., Lombard, V., Henrissat, B., Bain, J.R., *et al.* (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* *341*.

Rinttila, T., Kassinen, A., Malinen, E., Krogius, L., and Palva, A. (2004). Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J Appl Microbiol* *97*, 1166-1177.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., *et al.* (2007). Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* *39*, 596-604.

Robbe-Masselot, C., Maes, E., Rousset, M., Michalski, J.C., and Capon, C. (2009). Glycosylation of human fetal mucins: a similar repertoire of O-glycans along the intestinal tract. *Glycoconj J* *26*, 397-413.

Robbe, C., Capon, C., Coddeville, B., and Michalski, J.C. (2004). Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. *The Biochemical journal* 384, 307-316.

Robbe, C., Capon, C., Maes, E., Rousset, M., Zweibaum, A., Zanetta, J.P., and Michalski, J.C. (2003). Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. *J Biol Chem* 278, 46337-46348.

Rodriguez-Pineiro, A.M., Bergstrom, J.H., Ermund, A., Gustafsson, J.K., Schutte, A., Johansson, M.E.V., and Hansson, G.C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. II. Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of core proteins. *American journal of physiology Gastrointestinal and liver physiology* 305, G348-G356.

Rodriguez, J.M., Murphy, K., Stanton, C., Ross, R.P., Kober, O.I., Juge, N., Avershina, E., Rudi, K., Narbad, A., Jenmalm, M.C., *et al.* (2015). The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis* 26.

Rogier, E.W., Frantz, A.L., Bruno, M.E., and Kaetzel, C.S. (2014). Secretory IgA is concentrated in the outer layer of colonic mucus along with gut bacteria. *Pathogens* 3, 390-403.

Rohrer, J.S., Thayer, J., Weitzhandler, M., and Avdalovic, N. (1998). Analysis of the N-acetylneuraminic acid and N-glycolylneuraminic acid contents of glycoproteins by high-pH anion-exchange chromatography with pulsed amperometric detection (HPAEC/PAD). *Glycobiology* 8, 35-43.

Rokhsafat, S., Lin, A., and Comelli, E.M. (2016). Mucin-microbiota interaction during postnatal maturation of the intestinal ecosystem: Clinical implications. *Dig Dis Sci* 61, 1473-1486.

Roos, S., and Jonsson, H. (2002). A high-molecular-mass cell-surface protein from *Lactobacillus reuteri* 1063 adheres to mucus components. *Microbiology* 148, 433-442.

Rossez, Y., Maes, E., Lefebvre Darroman, T., Gosset, P., Ecobichon, C., Joncquel Chevalier Curt, M., Boneca, I.G., Michalski, J.C., and Robbe-Masselot, C. (2012). Almost all human gastric mucin O-glycans harbor blood group A, B or H antigens and are potential binding sites for *Helicobacter pylori*. *Glycobiology* 22, 1193-1206.

- Round, A.N., Berry, M., McMaster, T.J., Stoll, S., Gowers, D., Corfield, A.P., and Miles, M.J. (2002). Heterogeneity and persistence length in human ocular mucins. *Biophys J* *83*, 1661-1670.
- Russel, M.G., Engels, L.G., Muris, J.W., Limonard, C.B., Volovics, A., Brummer, R.J.M., and Stockbrugger, R.W. (1998). 'Modern life' in the epidemiology of inflammatory bowel disease: A case-control study with special emphasis on nutritional factors. *European Journal of Gastroenterology & Hepatology* *10*, 243-249.
- Rutgeerts, P., Dhaens, G., Hiele, M., Geboes, K., and Vantrappen, G. (1994). Appendectomy protects against ulcerative-colitis. *Gastroenterology* *106*, 1251-1253.
- Sakamoto, N., Kono, S., Wakai, K., Fukuda, Y., Satomi, A., Shimoyama, T., Inaba, Y., Miyake, Y., Sasaki, S., Okamoto, K., *et al.* (2005). Dietary risk factors for inflammatory bowel disease - A multicenter case-control study in Japan. *Inflamm Bowel Dis* *11*, 154-163.
- Sakuma, K., Aoki, M., and Kannagi, R. (2012). Transcription factors c-Myc and CDX2 mediate E-selectin ligand expression in colon cancer cells undergoing EGF/bFGF-induced epithelial-mesenchymal transition. *Proc Natl Acad Sci U S A* *109*, 7776-7781.
- Sakuraba, A., Sato, T., Kamada, N., Kitazume, M., Sugita, A., and Hibi, T. (2009). Th1/Th17 immune response is induced by mesenteric lymph node dendritic cells in crohn's disease. *Gastroenterology* *137*, 1736-1745.
- Salminen, S., Gibson, G.R., McCartney, A.L., and Isolauri, E. (2004). Influence of mode of delivery on gut microbiota composition in seven year old children. *Gut* *53*, 1388-1389.
- Salzman, N.H., Hung, K., Haribhai, D., Chu, H., Karlsson-Sjoeberg, J., Amir, E., Tegatz, P., Barman, M., Hayward, M., Eastwood, D., *et al.* (2010). Enteric defensins are essential regulators of intestinal microbial ecology. *Nat Immunol* *11*, 76-83.
- Sancho, E., Batlle, E., and Clevers, H. (2003). Live and let die in the intestinal epithelium. *Curr Opin Cell Biol* *15*, 763-770.
- Sandle, G.I. (1998). Salt and water absorption in the human colon: a modern appraisal. *Gut* *43*, 294-299.
- Sansonetti, P.J. (2004). War and peace at mucosal surfaces. *Nature reviews Immunology* *4*, 953-964.

Scanlan, P.D., Shanahan, F., O'Mahony, C., and Marchesi, J.R. (2006). Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *J Clin Microbiol* *44*, 3980-3988.

Schaffler, H., Herlemann, D.P., Alberts, C., Kaschitzki, A., Bodammer, P., Bannert, K., Koller, T., Warnke, P., Kreikemeyer, B., and Lamprecht, G. (2016). Mucosa-attached bacterial community in Crohn's Disease coheres with the Clinical Disease Activity Index. *Environ Microbiol Rep* *8*, 614-621.

Schluter, J., and Foster, K.R. (2012). The evolution of mutualism in gut microbiota via host epithelial selection. *PLoS Biol* *10*, e1001424.

Schulz, B.L., Packer, N.H., and Karlsson, N.G. (2002). Small-scale analysis of O-linked oligosaccharides from glycoproteins and mucins separated by gel electrophoresis. *Anal Chem* *74*, 6088-6097.

Schulzke, J.D., Ploeger, S., Amasheh, M., Fromm, A., Zeissig, S., Troeger, H., Richter, J., Bojarski, C., Schumann, M., and Fromm, M. (2009). Epithelial tight junctions in intestinal inflammation. *Ann N Y Acad Sci* *1165*, 294-300.

Schutte, A., Ermund, A., Becker-Pauly, C., Johansson, M.E.V., Rodriguez-Pineiro, A.M., Backhed, F., Muller, S., Lottaz, D., Bond, J.S., and Hansson, G.C. (2014). Microbial-induced meprin beta cleavage in MUC2 mucin and a functional CFTR channel are required to release anchored small intestinal mucus. *Proc Natl Acad Sci U S A* *111*, 12396-12401.

Schwab, C., Berry, D., Rauch, I., Rennisch, I., Ramesmayer, J., Hainzl, E., Heider, S., Decker, T., Kenner, L., Muller, M., *et al.* (2014). Longitudinal study of murine microbiota activity and interactions with the host during acute inflammation and recovery. *The ISME journal* *8*, 1101-1114.

Schwiertz, A., Taras, D., Schaefer, K., Beijer, S., Bos, N.A., Donus, C., and Hardt, P.D. (2010). Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* *18*, 190-195.

Seedorf, H., Griffin, N.W., Ridaura, V.K., Reyes, A., Cheng, J., Rey, F.E., Smith, M.I., Simon, G.M., Scheffrahn, R.H., Woebken, D., *et al.* (2014). Bacteria from diverse habitats colonize and compete in the mouse gut. *Cell* *159*, 253-266.

Segain, J.P., de la Bletiere, D.R., Bourreille, A., Leray, V., Gervois, N., Rosales, C., Ferrier, L., Bonnet, C., Blottiere, H.M., and Galmiche, J.P. (2000). Butyrate inhibits inflammatory responses through NF kappa B inhibition: implications for Crohn's disease. *Gut* 47, 397-403.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol* 14, e1002533.

Servier Medical Art (2016). Digestive system- Exploded view of Digestive apparatus (1).

Sewell, R., Backstrom, M., Dalziel, M., Gschmeissner, S., Karlsson, H., Noll, T., Gatgens, J., Clausen, H., Hansson, G.C., Burchell, J., *et al.* (2006). The ST6GalNAc-I sialyltransferase localizes throughout the Golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer. *J Biol Chem* 281, 3586-3594.

Shan, M.M., Gentile, M., Yeiser, J.R., Walland, A.C., Bornstein, V.U., Chen, K., He, B., Cassis, L., Bigas, A., Cols, M., *et al.* (2013). Mucus enhances gut homeostasis and oral tolerance by delivering immunoregulatory signals. *Science* 342, 447-453.

Sheng, Y.H., Hasnain, S.Z., Florin, T.H., and McGuckin, M.A. (2012). Mucins in inflammatory bowel diseases and colorectal cancer. *J Gastroenterol Hepatol* 27, 28-38.

Shin, N.R., Lee, J.C., Lee, H.Y., Kim, M.S., Whon, T.W., Lee, M.S., and Bae, J.W. (2014). An increase in the *Akkermansia* spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. *Gut* 63, 727-735.

Shirazi, T., Longman, R.J., Corfield, A.P., and Probert, C.S. (2000). Mucins and inflammatory bowel disease. *Postgrad Med J* 76, 473-478.

Shobar, R.M., Velineni, S., Keshavarzian, A., Swanson, G., DeMeo, M.T., Melson, J.E., Losurdo, J., Engen, P.A., Sun, Y., Koenig, L., *et al.* (2016). The effects of bowel preparation on microbiota-related metrics differ in health and in inflammatory bowel disease and for the mucosal and luminal microbiota compartments. *Clin Transl Gastroenterol* 7, e143.

Singh, P.K., and Hollingsworth, M.A. (2006). Cell surface-associated mucins in signal transduction. *Trends Cell Biol* 16, 467-476.

Smirnova, M.G., Guo, L., Birchall, J.P., and Pearson, J.P. (2003). LPS up-regulates mucin and cytokine mRNA expression and stimulates mucin and cytokine secretion in goblet cells. *Cell Immunol* 221, 42-49.

Smirnova, M.G., Kiselev, S.L., Birchall, J.P., and Pearson, J.P. (2001). Up-regulation of mucin secretion in HT29-MTX cells by the pro-inflammatory cytokines tumor necrosis factor-alpha and interleukin-6. *European Cytokine Network* 12, 119-125.

Smith, P.M., Howitt, M.R., Panikov, N., Michaud, M., Gallini, C.A., Bohlooly-Y, M., Glickman, J.N., and Garrett, W.S. (2013). The microbial metabolites, short-chain fatty acids, regulate colonic T-reg cell homeostasis. *Science* 341, 569-573.

Smythies, L.E., Sellers, M., Clements, R.H., Mosteller-Barnum, M., Meng, G., Benjamin, W.H., Orenstein, J.M., and Smith, P.D. (2005). Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J Clin Invest* 115, 66-75.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L.G., Gratadoux, J.J., Blugeon, S., Bridonneau, C., Furet, J.P., Corthier, G., *et al.* (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, 16731-16736.

Sokol, H., Seksik, P., Furet, J.P., Firmesse, O., Nion-Larmurier, L., Beaugerie, L., Cosnes, J., Corthier, G., Marteau, P., and Dore, J. (2009). Low counts of *Faecalibacterium prausnitzii* in colitis microbiota. *Inflamm Bowel Dis* 15, 1183-1189.

Sommer, F., Adam, N., Johansson, M.E., Xia, L., Hansson, G.C., and Backhed, F. (2014). Altered mucus glycosylation in core 1 O-glycan-deficient mice affects microbiota composition and intestinal architecture. *Plos One* 9, e85254.

Sonnenburg, E.D., Smits, S.A., Tikhonov, M., Higginbottom, S.K., Wingreen, N.S., and Sonnenburg, J.L. (2016). Diet-induced extinctions in the gut microbiota compound over generations. *Nature* 529, 212-215.

Sovran, B., Lu, P., Loonen, L.M., Hugenholtz, F., Belzer, C., Stolte, E.H., Boekschoten, M.V., van Baarlen, P., Smidt, H., Kleerebezem, M., *et al.* (2016). Identification of commensal species positively correlated with early stress responses to a compromised mucus barrier. *Inflamm Bowel Dis* 22, 826-840.

Stone, E.L., Ismail, M.N., Lee, S.H., Luu, Y., Ramirez, K., Haslam, S.M., Ho, S.B., Dell, A., Fukuda, M., and Marth, J.D. (2009). Glycosyltransferase function in core 2-type protein O glycosylation. *Mol Cell Biol* 29, 3770-3782.

Suau, A., Bonnet, R., Sutren, M., Godon, J.J., Gibson, G.R., Collins, M.D., and Dore, J. (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. *Appl Environ Microbiol* *65*, 4799-4807.

Sugimoto, K., Ogawa, A., Mizoguchi, E., Shimomura, Y., Andoh, A., Bhan, A.K., Blumberg, R.S., Xavier, R.J., and Mizoguchi, A. (2008). IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. *J Clin Invest* *118*, 534-544.

Suzuki, K., Meek, B., Doi, Y., Muramatsu, M., Chiba, T., Honjo, T., and Fagarasan, S. (2004). Aberrant expansion of segmented filamentous bacteria in IgA-deficient gut. *Proc Natl Acad Sci U S A* *101*, 1981-1986.

Svanback, R., and Bolnick, D.I. (2007). Intraspecific competition drives increased resource use diversity within a natural population. *Proceedings of the Royal Society B-Biological Sciences* *274*, 839-844.

Swidsinski, A., Ladhoff, A., Pernthaler, A., Swidsinski, S., Loening-Baucke, V., Ortner, M., Weber, J., Hoffmann, U., Schreiber, S., Dietel, M., *et al.* (2002). Mucosal flora in inflammatory bowel disease. *Gastroenterology* *122*, 44-54.

Swidsinski, A., Loening-Baucke, V., Lochs, H., and Hale, L.P. (2005). Spatial organization of bacterial flora in normal and inflamed intestine: a fluorescence in situ hybridization study in mice. *World J Gastroenterol* *11*, 1131-1140.

Swidsinski, A., Loening-Baucke, V., Theissig, F., Engelhardt, H., Bengmark, S., Koch, S., Lochs, H., and Dorffel, Y. (2007). Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut* *56*, 343-350.

Tailford, L., Crost, E., Kavanaugh, D., and Juge, N. (2015a). Mucin glycan foraging in the human gut microbiome. *Frontiers in Genetics* *6*.

Tailford, L.E., Owen, C.D., Walshaw, J., Crost, E.H., Hardy-Goddard, J., Le Gall, G., de Vos, W.M., Taylor, G.L., and Juge, N. (2015b). Discovery of intramolecular trans-sialidases in human gut microbiota suggests novel mechanisms of mucosal adaptation. *Nature communications* *6*, 7624.

Takahashi, T., Kuniyasu, Y., Toda, M., Sakaguchi, N., Itoh, M., Iwata, M., Shimizu, J., and Sakaguchi, S. (1998). Immunologic self-tolerance maintained by CD25(+)CD4(+) naturally

anergic and suppressive T cells: induction of autoimmune disease by breaking their anergic/suppressive state. *Int Immunol* *10*, 1969-1980.

Tasteyre, A., Barc, M.C., Collignon, A., Boureau, H., and Karjalainen, T. (2001). Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect Immun* *69*, 7937-7940.

Tedelind, S., Westberg, F., Kjerrulf, M., and Vidal, A. (2007). Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease. *World J Gastroenterol* *13*, 2826-2832.

Theodoratou, E., Campbell, H., Ventham, N.T., Kolarich, D., Pucic-Bakovic, M., Zoldos, V., Fernandes, D., Pemberton, I.K., Rudan, I., Kennedy, N.A., *et al.* (2014). The role of glycosylation in IBD. *Nature reviews Gastroenterology & hepatology* *11*, 588-600.

Theodoropoulos, G., and Carraway, K.L. (2007). Molecular signaling in the regulation of mucins. *J Cell Biochem* *102*, 1103-1116.

Thompson-Chagoyán, O.C., Maldonado, J., and Gil, A. (2005). Aetiology of inflammatory bowel disease (IBD): role of intestinal microbiota and gut-associated lymphoid tissue immune response. *Clin Nutr* *24*, 339-352.

Thomsson, K.A., Holmén-Larsson, J.M., Ångström, J., Johansson, M.E., Xia, L., and Hansson, G.C. (2012). Detailed O-glycomics of the Muc2 mucin from colon of wild-type, core 1-and core 3-transferase-deficient mice highlights differences compared with human MUC2. *Glycobiology* *22*, 1128-1139.

Thornton, A.M., and Shevach, E.M. (1998). CD4(+)CD25(+) immunoregulatory T cells suppress polyclonal T cell activation in vitro by inhibiting interleukin 2 production. *J Exp Med* *188*, 287-296.

Tom, B.H., Rutzky, L.P., Jakstys, M.M., Oyasu, R., Kaye, C.I., and Kahan, B.D. (1976). Human colonic adenocarcinoma cells. I. Establishment and description of a new line. *In vitro* *12*, 180-191.

Tong, J., Liu, C., Summanen, P., Xu, H., and Finegold, S.M. (2011). Application of quantitative real-time PCR for rapid identification of *Bacteroides fragilis* group and related organisms in human wound samples. *Anaerobe* *17*, 64-68.

Tong, M., Li, X., Parfrey, L.W., Roth, B., Ippoliti, A., Wei, B., Borneman, J., McGovern, D.P.B., Frank, D.N., Li, E., *et al.* (2013). A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. *Plos One* 8, e80702.

Tong, M., McHardy, I., Ruegger, P., Goudarzi, M., Kashyap, P.C., Haritunians, T., Li, X., Graeber, T.G., Schwager, E., Huttenhower, C., *et al.* (2014). Reprogramming of gut microbiome energy metabolism by the FUT2 Crohn's disease risk polymorphism. *The ISME journal* 8, 2193-2206.

Torok, H.P., Glas, J., Tonenchi, L., Bruennler, G., Folwaczny, M., and Folwaczny, C. (2004). Crohn's disease is associated with a toll-like receptor-9 polymorphism. *Gastroenterology* 127, 365-366.

Travisano, M., and Velicer, G.J. (2004). Strategies of microbial cheater control. *Trends Microbiol* 12, 72-78.

Turnbaugh, P.J., Backhed, F., Fulton, L., and Gordon, J.I. (2008). Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell host & microbe* 3, 213-223.

Turnbaugh, P.J., Hamady, M., Yatsunenkov, T., Cantarel, B.L., Duncan, A., Ley, R.E., Sogin, M.L., Jones, W.J., Roe, B.A., Affourtit, J.P., *et al.* (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480-484.

Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R., and Gordon, J.I. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 444, 1027-1031.

Tyakht, A.V., Kostryukova, E.S., Popenko, A.S., Belenikin, M.S., Pavlenko, A.V., Larin, A.K., Karpova, I.Y., Selezneva, O.V., Semashko, T.A., Ospanova, E.A., *et al.* (2013). Human gut microbiota community structures in urban and rural populations in Russia. *Nature Commun* 4.

Vaishnava, S., Yamamoto, M., Severson, K.M., Ruhn, K.A., Yu, X., Koren, O., Ley, R., Wakeland, E.K., and Hooper, L.V. (2011). The Antibacterial Lectin RegIII gamma Promotes the Spatial Segregation of Microbiota and Host in the Intestine. *Science* 334, 255-258.

Van den Abbeele, P., Belzer, C., Goossens, M., Kleerebezem, M., De Vos, W.M., Thas, O., De Weirdt, R., Kerckhof, F.-M., and Van de Wiele, T. (2013). Butyrate-producing Clostridium cluster XIVa species specifically colonize mucins in an in vitro gut model. *The ISME journal* 7, 949-961.

van der Post, S., and Hansson, G.C. (2014). Membrane protein profiling of human colon reveals distinct regional differences. *Molecular & Cellular Proteomics* 13, 2277-2287.

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., *et al.* (2006). Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* 131, 117-129.

van Passel, M.W., Kant, R., Zoetendal, E.G., Plugge, C.M., Derrien, M., Malfatti, S.A., Chain, P.S., Woyke, T., Palva, A., de Vos, W.M., *et al.* (2011). The genome of Akkermansia muciniphila, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *Plos One* 6, e16876.

van Sorge, N.M., and Doran, K.S. (2012). Defense at the border: the blood-brain barrier versus bacterial foreigners. *Future Microbiol* 7, 383-394.

Varyukhina, S., Freitas, M., Bardin, S., Robillard, E., Tavan, E., Sapin, C., Grill, J.P., and Trugnan, G. (2012). Glycan-modifying bacteria-derived soluble factors from Bacteroides thetaiotaomicron and Lactobacillus casei inhibit rotavirus infection in human intestinal cells. *Microbes Infect* 14, 273-278.

Verschuere, S., De Smet, R., Allais, L., and Cuvelier, C.A. (2012). The effect of smoking on intestinal inflammation: What can be learned from animal models? *Journal of Crohns & Colitis* 6, 1-12.

Vijay-Kumar, M., Sanders, C.J., Taylor, R.T., Kumar, A., Aitken, J.D., Sitaraman, S.V., Neish, A.S., Uematsu, S., Akira, S., Williams, I.R., *et al.* (2007). Deletion of TLR5 results in spontaneous colitis in mice. *J Clin Invest* 117, 3909-3921.

von der Weid, T., Bulliard, C., and Schiffrin, E.J. (2001). Induction by a lactic acid bacterium of a population of CD4(+) T cells with low proliferative capacity that produce transforming growth factor beta and interleukin-10. *Clin Diagn Lab Immunol* 8, 695-701.

Wacklin, P., Tuimala, J., Nikkila, J., Sebastian, T., Makivuokko, H., Alakulppi, N., Laine, P., Rajilic-Stojanovic, M., Paulin, L., de Vos, W.M., *et al.* (2014). Faecal microbiota composition in adults is associated with the FUT2 gene determining the secretor status. *Plos One* 9, e94863.

Wang, L., Christophersen, C.T., Sorich, M.J., Gerber, J.P., Angley, M.T., and Conlon, M.A. (2011). Low relative abundances of the mucolytic bacterium *Akkermansia muciniphila* and *Bifidobacterium* spp. in feces of children with autism. *Appl Environ Microbiol* 77, 6718-6721.

Wang, W., Chen, L., Zhou, R., Wang, X., Song, L., Huang, S., Wang, G., and Xia, B. (2014). Increased proportions of *Bifidobacterium* and the *Lactobacillus* group and loss of butyrate-producing bacteria in inflammatory bowel disease. *J Clin Microbiol* 52, 398-406.

Wang, Y., Ju, T., Ding, X., Xia, B., Wang, W., Xia, L., He, M., and Cummings, R.D. (2010). Cosmc is an essential chaperone for correct protein O-glycosylation. *Proc Natl Acad Sci U S A* 107, 9228-9233.

Watson, A.J., Duckworth, C.A., Guan, Y., and Montrose, M.H. (2009). Mechanisms of epithelial cell shedding in the mammalian intestine and maintenance of barrier function. *Ann N Y Acad Sci* 1165, 135-142.

Weaver, G.A., Krause, J.A., Miller, T.L., and Wolin, M.J. (1988). Short chain fatty acid distributions of enema samples from a sigmoidoscopy population: an association of high acetate and low butyrate ratios with adenomatous polyps and colon cancer. *Gut* 29, 1539-1543.

Wehkamp, J., Harder, J., Weichenthal, M., Schwab, M., Schaffeler, E., Schlee, M., Herrlinger, K.R., Stallmach, A., Noack, F., Fritz, P., *et al.* (2004). NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut* 53, 1658-1664.

Wehkamp, J., Salzman, N.H., Porter, E., Nuding, S., Weichenthal, M., Petras, R.E., Shen, B., Schaeffeler, E., Schwab, M., Linzmeier, R., *et al.* (2005). Reduced Paneth cell alpha-defensins in ileal Crohn's disease. *Proc Natl Acad Sci U S A* 102, 18129-18134.

Wenzel, U.A., Magnusson, M.K., Rydstrom, A., Jonstrand, C., Hengst, J., Johansson, M.E.V., Velcich, A., Ohman, L., Strid, H., Sjovall, H., *et al.* (2014). Spontaneous colitis in MUC2-

deficient mice reflects clinical and cellular features of active ulcerative colitis. *Plos One* 9, e100217.

Whitehead, R.H., Young, G.P., and Bhathal, P.S. (1986). Effects of short chain fatty-acids on a new human-colon carcinoma cell-line (Iim1215). *Gut* 27, 1457-1463.

Williams, J.M., Duckworth, C.A., Burkitt, M.D., Watson, A.J., Campbell, B.J., and Pritchard, D.M. (2015). Epithelial cell shedding and barrier function: a matter of life and death at the small intestinal villus tip. *Vet Pathol* 52, 445-455.

Williams, S.E., and Turnberg, L.A. (1981). Demonstration of a pH gradient across mucus adherent to rabbit gastric mucosa: evidence for a 'mucus-bicarbonate' barrier. *Gut* 22, 94-96.

Willing, B.P., Dicksved, J., Halfvarson, J., Andersson, A.F., Lucio, M., Zheng, Z., Jarnerot, G., Tysk, C., Jansson, J.K., and Engstrand, L. (2010a). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139, 1844-1854.

Willing, B.P., Gill, N., and Finlay, B.B. (2010b). The role of the immune system in regulating the microbiota. *Gut microbes* 1, 213-223.

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., *et al.* (2014). NLRP6 inflammasome orchestrates the colonic host-microbial interface by regulating goblet cell mucus secretion. *Cell* 156, 1045-1059.

Woodmansey, E.J., McMurdo, M.E.T., Macfarlane, G.T., and Macfarlane, S. (2004). Comparison of compositions and metabolic activities of fecal microbiotas in young adults and in antibiotic-treated and non-antibiotic-treated elderly subjects. *Appl Environ Microbiol* 70, 6113-6122.

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

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.-Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., *et al.* (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105-108.

Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., *et al.* (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol* 5, 1574-1586.

Xu, X.R., Liu, C.Q., Feng, B.S., and Liu, Z.J. (2014). Dysregulation of mucosal immune response in pathogenesis of inflammatory bowel disease. *World J Gastroenterol* 20, 3255-3264.

Yamamoto-Furusho, J.K., Ascano-Gutierrez, I., Furuzawa-Carballeda, J., and Fonseca-Camarillo, G. (2015). Differential expression of MUC12, MUC16, and MUC20 in patients with active and remission ulcerative colitis. *Mediators Inflamm* 2015.

Yen, D., Cheung, J., Scheerens, H., Poulet, F., McClanahan, T., McKenzie, B., Kleinschek, M.A., Owyang, A., Mattson, J., Blumenschein, W., *et al.* (2006). IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. *J Clin Invest* 116, 1310-1316.

Yu, Z.T., Chen, C., Kling, D.E., Liu, B., McCoy, J.M., Merighi, M., Heidtman, M., and Newburg, D.S. (2013). The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. *Glycobiology* 23, 169-177.

Zarepour, M., Bhullar, K., Montero, M., Ma, C., Huang, T., Velcich, A., Xia, L., and Vallance, B.A. (2013). The mucin MUC2 limits pathogen burdens and epithelial barrier dysfunction during *Salmonella enterica* serovar Typhimurium colitis. *Infect Immun* 81, 3672-3683.

Zhang, X., Shen, D., Fang, Z., Jie, Z., Qiu, X., Zhang, C., Chen, Y., and Ji, L. (2013). Human gut microbiota changes reveal the progression of glucose intolerance. *Plos One* 8, e71108.

Znamenskaya, Y., Sotres, J., Engblom, J., Arnebrant, T., and Kocherbitov, V. (2012). Effect of hydration on structural and thermodynamic properties of pig gastric and bovine submaxillary gland mucins. *J Phys Chem B* 116, 5047-5055.

## Appendix 1

Commercial suppliers of chemicals, reagents and equipment

### **Agilent Technologies**

5301 Stevens Creek Boulevard,  
Santa Clara, CA 95051

### **Asylum Research**

6310 Hollister Avenue,  
Goleta, CA 93117

### **ATCC**

0801 University Boulevard,  
Manassas, VA 20110-2209

### **Bard Medical**

Tilgate Forest Business Centre, Brighton Road,  
Crawley, RH11 9BP

### **B. Braun Melsungen AG**

Carl-Braun-Straße 1,  
34212 Melsungen

### **Beckman Coulter**

250 S Kraemer Boulevard,  
Brea, CA 92821

### **Becton, Dickinson and Company**

1 Becton Drive, Franklin Lakes,  
New Jersey, 07417-1880

### **Bio Rad**

1000 Alfred Nobel Drive,  
Hercules, CA 94547

**Bruker BioSpin GmbH**

Silberstreifen 4,  
76287 Rheinstetten

**Carbosynth**

9 Old Station Business Park,  
Compton, RG20 6NE

**Don Whitley**

14 Otley Road,  
Shipley, BD17 7SE

**Eppendorf**

Barkhausenweg 1,  
22339 Hamburg

**GE Healthcare**

Amersham Place,  
Little Chalfont, HP7 9NA

**Glycom**

Diplomvej 373,  
2800 Kongens, Lyngby

**Grace**

7500 Grace Drive,  
Columbia, MD 21044

**Ludger**

Culham Science Centre,  
Abingdon, OX14 3EB

**Merck Millipore**

Frankfurter Straße 250,

64293 Darmstadt

**MP Biomedicals**

3 Hutton Center Drive Suite 100,

Santa Ana, CA 92707

**New England Biolabs**

240 County Road,

Ipswich, MA 01938

**Olympus**

4-29-16 Owada-cho, Hachioji-shi,

Tokyo, 192-0045

**Pall Corporation**

25 Harbor Park Drive,

Port Washington, NY 11050

**Phenomenex**

411 Madrid Ave,

Torrance, CA 90501

**Promega**

2800 Woods Hollow Road,

Madison, WI 53711

**Protein Simple**

3001 Orchard Parkway,

San Jose, CA 95134

**Prozyme**

832 Bay Center Pl,

Hayward, CA 94545

**Qiagen**

Straße 1

Hilden 40724

**Sigma-Aldrich**

3050 Spruce Street,

St. Louis, MO 63103

**Spectrum labs**

18617 S Broadwick Street,

Rancho Dominguez, CA 90220

**Starstedt**

Sarstedtstraße 1,

51582 Nümbrecht

**Thermo Fisher Scientific**

81 Wyman Street,

Waltham, MA 02451

**VWR International**

100 Matsonford Road,

Radnor, PA 19087

## Appendix 2

### Patient metadata

#### **Key**

#### ***Regions;***

S, Sigmoid

A, Ascending

#### ***Grouping;***

UC, Ulcerative colitis

CD, Crohn's disease

#### ***Disease activity;***

Disease activity is based on histological scoring of extra biopsies taken from the same region

Quies, Quiescent

Quies\*, Quiescent in given region but activity notable elsewhere (as specified in \* column)

Mod, Moderate

#### ***Experimental analysis;***

Y, the relevant experimental technique was carried out with the sample listed

Patient no.	Region	Group	Referral reason	Disease Activity	*	Treatment (IBD)	Age	Gender	qPCR	16S Seq	qRT-PCR	HPAEC-MALDI-PAD	TOF
13TB0583	S	Control	Surveillance (previous cancer)	NA			64	F	Y	Y	Y	Y	Y
13TB0583	A	Control	Surveillance (previous cancer)	NA			64	F	Y	Y	Y	Y	Y
13TB0601	S	UC		Quies		5ASA	55	M	Y	Y	Y	Y	Y
13TB0601	A	UC		Mild		5ASA	55	M	Y	Y	Y	Y	Y
13TB0605	S	Control	Surveillance (previous cancer)	NA			75	M	Y	Y	Y	Y	Y
13TB0605	A	Control	Surveillance (previous cancer)	NA			75	M	Y	Y	Y	Y	Y
13TB0614	S	UC		Mod		5ASA/Thiopurines	68	F	Y	Y	Y	Y	Y
13TB0614	A	UC		Mod		5ASA/Thiopurines	68	F	Y	Y	Y	Y	Y
13TB0617	S	Control	Gorlin's syndrome	NA			72	M	Y	Y	Y	Y	Y
13TB0617	A	Control	Gorlin's syndrome	NA			72	M	Y	Y	Y	Y	Y
13TB0634	S	UC		Quies		5ASA	51	M	Y	Y	Y	Y	Y
13TB0634	A	UC		Quies		5ASA	51	M	Y	Y	Y	Y	Y
13TB0648	S	UC		Quies		Steroids	75	M	Y	Y	Y	Y	Y
13TB0648	A	UC		Quies		Steroids	75	M	Y	Y	Y	Y	Y
13TB0655	S	UC		Quies		5ASA	48	F	Y	Y	Y	Y	Y
13TB0655	A	UC		Quies		5ASA	48	F	Y	Y	Y	Y	Y
13TB0667	S	Control	Surveillance (previous cancer)	NA			62	M		Y	Y	Y	Y
13TB0667	A	Control	Surveillance (previous cancer)	NA			62	M		Y	Y	Y	Y
14TB0005	S	CD		Quies		5ASA	55	M	Y	Y	Y	Y	Y
14TB0005	A	CD		Quies		5ASA	55	M	Y	Y	Y	Y	Y
14TB0025	S	Control	Iron deficiency anaemia	NA			74	M	Y	Y	Y	Y	Y
14TB0025	A	Control	Iron deficiency anaemia	NA			74	M	Y	Y	Y	Y	Y
14TB0024	S	UC		Quies		5ASA	61	M	Y	Y	Y	Y	Y
14TB0024	A	UC		Quies		5ASA	61	M	Y	Y	Y	Y	Y
14TB0030	S	UC		Quies		5ASA	56	F	Y	Y	Y	Y	Y
14TB0030	A	UC		Quies		5ASA	56	F	Y	Y	Y	Y	Y
14TB0047	S	Control	Diarrhoea and bloating	NA			47	M	Y	Y	Y	Y	Y
14TB0047	A	Control	Diarrhoea and bloating	NA			47	M	Y	Y	Y	Y	Y
14TB0055	S	CD		Quies*	Mild ileal CD	Steroids	30	F	Y	Y	Y	Y	Y
14TB0055	A	CD		Quies*	Mild ileal CD	Steroids	30	F	Y	Y	Y	Y	Y
14TB0060	S	Control	Change in bowel habit	NA			47	F	Y	Y	Y	Y	Y
14TB0060	A	Control	Change in bowel habit	NA			47	F	Y	Y	Y	Y	Y
14TB0061	S	CD		Quies*	Focal active chronic colitis	5ASA/Thiopurines	50	M	Y	Y	Y	Y	Y
14TB0061	A	CD		Quies*	Focal active chronic colitis	5ASA/Thiopurines	50	M	Y	Y	Y	Y	Y

Patient no.	Region	Group	Referral reason	Disease Activity	*	Treatment (IBD)	Age	Gender	qPCR	16S Seq	qRT-PCR	HPAEC-MALDI-PAD	TOF
14TB0068	S	UC		Quies		5ASA	45	M	Y	Y	Y	Y	Y
14TB0068	A	UC		Quies		5ASA	45	M	Y	Y	Y	Y	Y
14TB0073	S	UC		Quies		5ASA	56	M	Y	Y	Y	Y	Y
14TB0073	A	UC		Quies		5ASA	56	M	Y	Y	Y	Y	Y
14TB0082	S	CD		Quies		5ASA/Thiopurines	49	M	Y	Y	Y	Y	Y
14TB0082	A	CD		Quies		5ASA/Thiopurines	49	M	Y	Y	Y	Y	Y
14TB0083	S	Control	Surveillance (previous polyps)	NA			73	M			Y		
14TB0083	A	Control	Surveillance (previous polyps)	NA			73	M			Y		
14TB0086	S	UC		Mod		5ASA	57	F	Y	Y	Y	Y	Y
14TB0086	A	UC		Mod		5ASA	57	F	Y	Y	Y	Y	Y
14TB0097	S	UC		Quies		5ASA	73	M	Y	Y	Y	Y	Y
14TB0097	A	UC		Quies		5ASA	73	M	Y	Y	Y	Y	Y
14TB0104	S	UC		Quies		5ASA	68	F	Y	Y	Y	Y	Y
14TB0104	A	UC		Quies		5ASA	68	F	Y	Y	Y	Y	Y
14TB0117	S	UC		Quies		5ASA	63	F	Y	Y	Y	Y	Y
14TB0117	A	UC		Quies		5ASA	63	F	Y	Y	Y	Y	Y
14TB0138	S	UC		Quies*	Focal active chronic colitis	None	42	F	Y	Y	Y	Y	Y
14TB0138	A	UC		Quies*	Focal active chronic colitis	None	42	F	Y	Y	Y	Y	Y
14TB0137	S	CD		Quies		5ASA	65	F	Y	Y	Y	Y	Y
14TB0137	A	CD		Quies		5ASA	65	F	Y	Y	Y	Y	Y
14TB0143	S	Control	Diverticular disease	NA			61	F	Y	Y	Y	Y	Y
14TB0143	A	Control	Diverticular disease	NA			61	F	Y	Y	Y	Y	Y
14TB0168	S	Control	Barrett's oesophagus	NA			54	M	Y	Y	Y	Y	Y
14TB0168	A	Control	Barrett's oesophagus	NA			54	M	Y	Y	Y	Y	Y
14TB0173	S	UC		Quies		None	67	F	Y	Y	Y	Y	Y
14TB0173	A	UC		Quies		None	67	F	Y	Y	Y	Y	Y
14TB0196	S	UC		Quies		5ASA	73	F	Y	Y	Y	Y	Y
14TB0196	A	UC		Quies		5ASA	73	F	Y	Y	Y	Y	Y
14TB0197	S	UC		Quies*	Moderate proctitis	Steroids	51	M	Y	Y	Y	Y	Y
14TB0197	A	UC		Quies*	Moderate proctitis	Steroids	51	M	Y	Y	Y	Y	Y
14TB0261	S	CD		Mild		Thiopurines/biologics	57	F	Y	Y	Y	Y	Y
14TB0313	S	Control	Surveillance (previous polyps)	NA			60	M	Y	Y	Y	Y	Y
14TB0313	A	Control	Surveillance (previous polyps)	NA			60	M	Y	Y	Y	Y	Y

Patient no.	Region	Group	Referral reason	Disease Activity	*	Treatment (IBD)	Age	Gender	qPCR	16S Seq	qRT-PCR	HPAEC-MALDI-PAD	TOF
14TB0354	S	CD		Mild		IV steroids, 5ASA, thiopurines	30	F	Y	Y	Y	Y	Y
14TB0357	S	UC		Mod		5ASA/thiopurines	38	F	Y	Y	Y	Y	Y
14TB0424	S	CD		Quies		5ASA	68	F	Y	Y	Y		
14TB0653	S	UC		Quies		5ASA	17	M	Y			Y	
15TB0058	S	CD				None	17	F			Y	Y	
15TB0163	S	CD		Mild		Steroids	24	M					
15TB0310	S	Control	Vomiting and diarrhoea	NA			64	F	Y	Y	Y	Y	Y

## Appendix 3

### Electronic data

*Normalised sequencing reads of DNA extracted from human lavage samples*

*MALDI-TOF data of mucins purified from human lavage samples*

*MALDI-TOF data of mucins purified from mice mono- and co- colonised with R. gnavus CC55\_001C and A. muciniphila ATCC BAA 835*

## Appendix 4

Effect of confounding variables on microbiota composition (qPCR)

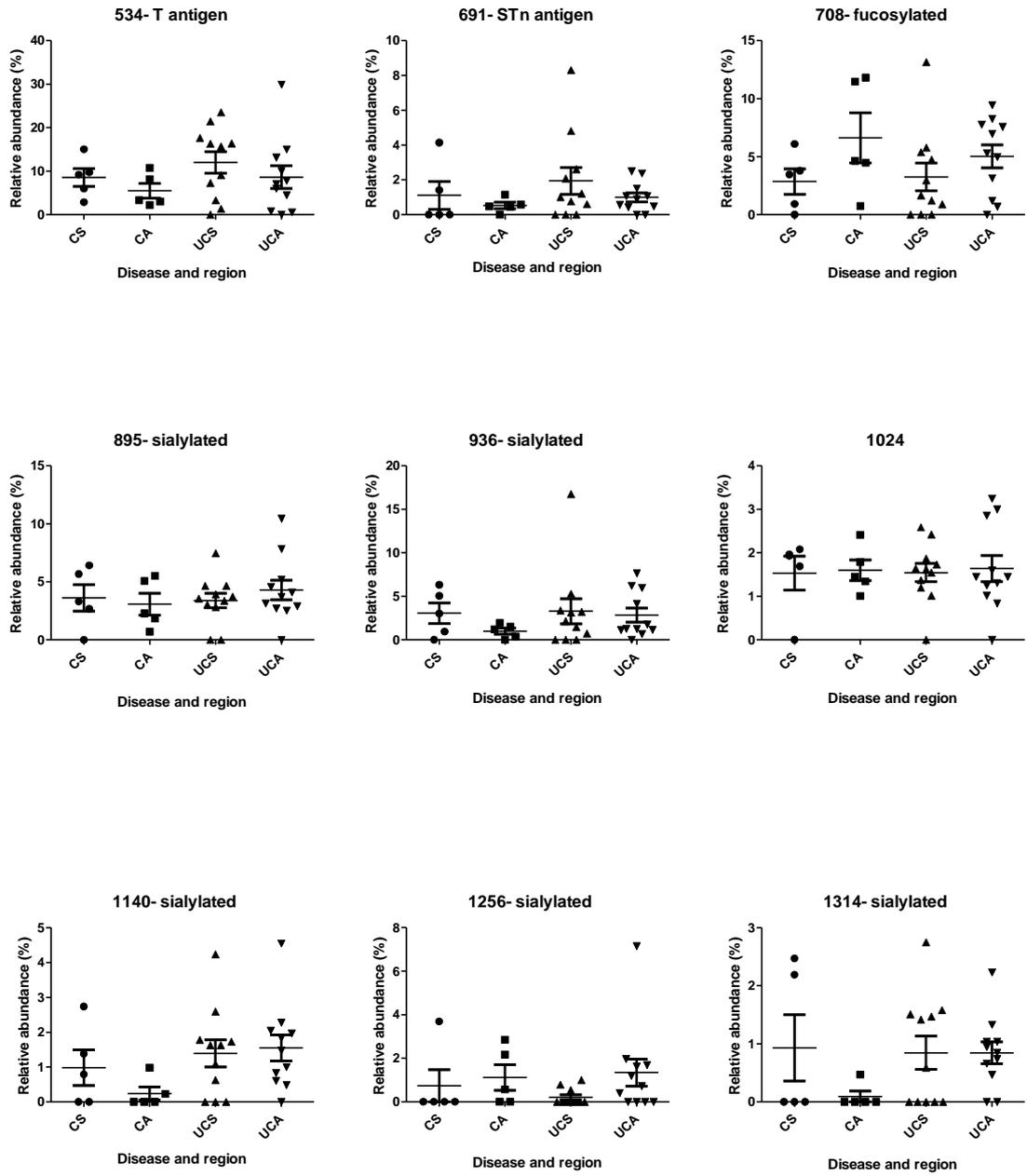
*P values of ANOVA and Pearson statistical tests in the Sigmoid and Ascending colon*

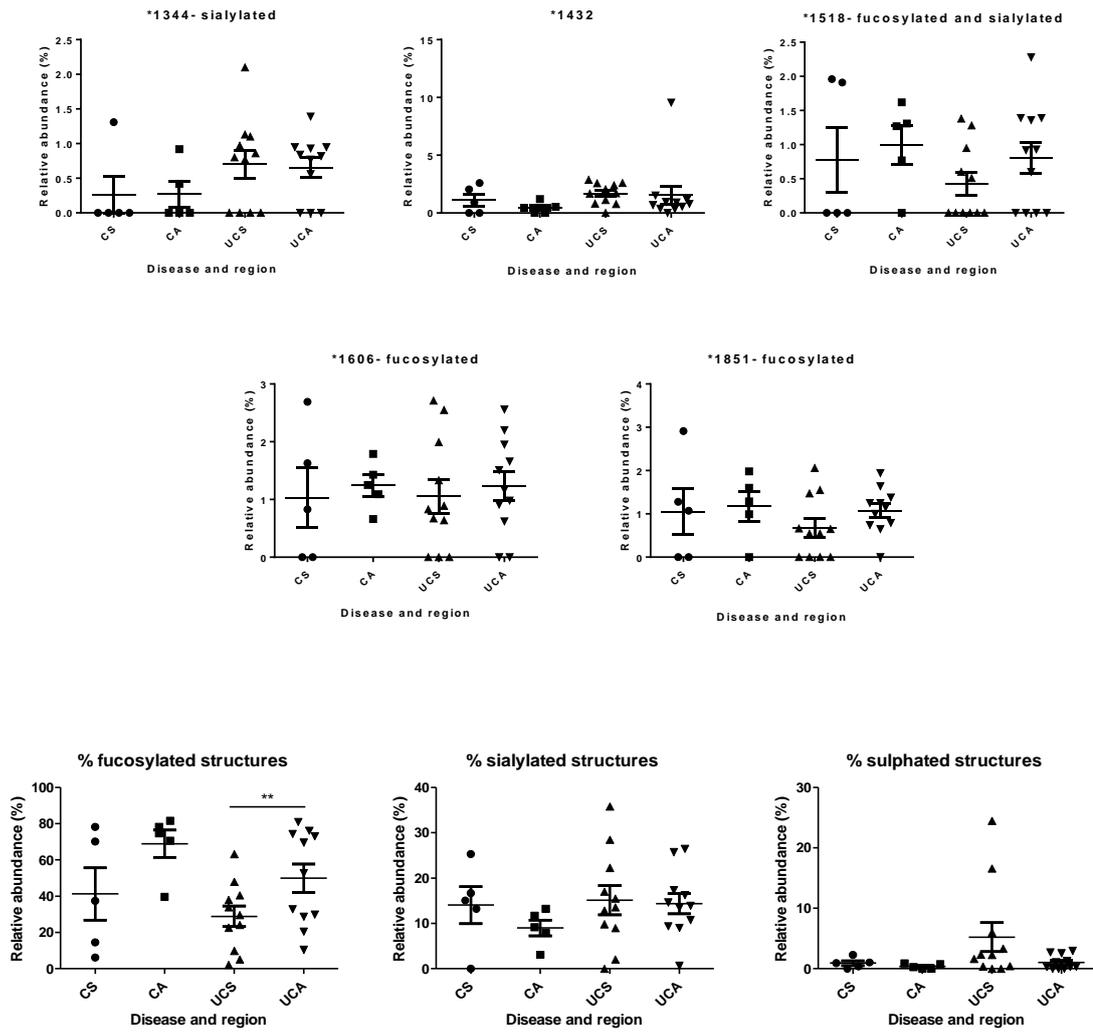
Sigmoid colon	ANOVA		Pearson
	Gender	Treatment	Age
<b>Bifidobacterium</b>	0.4475	0.5987	0.3690
<b>Roseburia</b>	0.9340	0.1214	0.7933
<b>Faecalibacterium</b>	0.3015	0.3003	0.5186
<b>Lactobacillaceae</b>	0.6473	0.1148	0.4466
<b>Bacteroides</b>	0.7944	0.0776	0.8340
<b>Lachnospiraceae</b>	0.1823	0.0870	0.3506
<b><i>A. muciniphila</i></b>	0.6659	0.0230	0.7071
<b><i>R. gnavus</i></b>	0.0532	0.4092	0.1137
<b><i>B. fragilis</i></b>	0.0764	0.9312	0.3686

Ascending colon	ANOVA		Pearson
	Gender	Treatment	Age
<b>Bifidobacterium</b>	0.5960	0.4221	0.4405
<b>Roseburia</b>	0.9756	0.0562	0.9874
<b>Faecalibacterium</b>	0.1116	0.2503	0.4620
<b>Lactobacillaceae</b>	0.9698	0.3104	0.8634
<b>Bacteroides</b>	0.3845	0.2085	0.8431
<b>Lachnospiraceae</b>	0.2722	0.1705	0.3482
<b><i>A. muciniphila</i></b>	0.6065	0.0650	0.6238
<b><i>R. gnavus</i></b>	0.1803	0.4108	0.2697
<b><i>B. fragilis</i></b>	0.0250	0.7201	0.5110

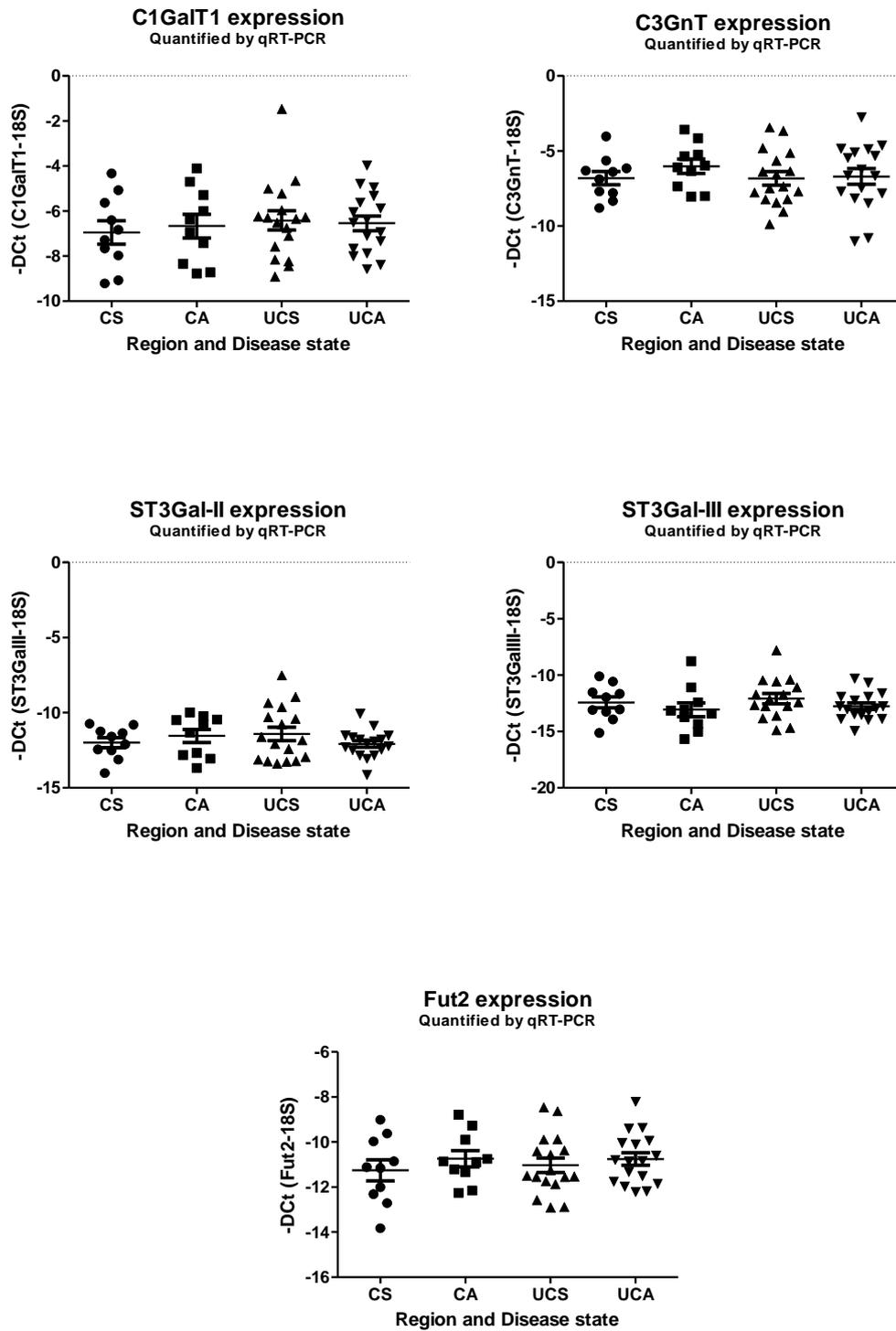
## Appendix 5

## Dot plots of glycan structures





## Dot plots of glycosyltransferase expression



## Appendix 6

BlastP results of *R. gnavus* CC55\_001C homologous proteins to nan locus of *R. gnavus*

ATCC 29149

Protein in <i>R. gnavus</i> ATCC 29149	Type of protein	Accession number	Homologous protein in CC55_001C	Blastp % homology	Accession number
RUMGNA_02701	Putative GDSL-like protein	EDN77087.1	Hypothetical protein	100	WP_004843642.1
RUMGNA_02700	Putative sugar isomerase	EDN77086.1	Hypothetical protein	100	WP_004843641.1
RUMGNA_02699	Protein with homology to transcriptional regulators of AraC family	EDN77085.1	AraC family transcriptional regulator	100	WP_004843639.1
RUMGNA_02698	Predicted solute-binding protein (ABC transporter)	EDN77084.1	Sugar ABC transporter substrate-binding protein	100	WP_004843638.1
RUMGNA_02697	Putative permeases (ABC transporter)	EDN77083.1	Hypothetical protein	100	ETD19280.1
RUMGNA_02696		EDN77082.1	Hypothetical protein	99	WP_023923961.1
RUMGNA_02695	Homology with oxidoreductase from the Gfo/ldh/MocA family	EDN77081.1	Oxidoreductase	100	WP_004843633.1
RUMGNA_02694	GH33 enzyme (nanH)	EDN77080.1	Anhydrosialidase	100	WP_004843631.1
RUMGNA_02693	Predicted ManNac-6-P (nanE)	EDN77079.1	Hypothetical protein	100	ETD19276.1
RUMGNA_02692	Putative Neu5Ac lyase (nanA)	EDN77078.1	N-acetylneuraminase lyase	99	WP_023923959.1
RUMGNA_02691	Predicted ManNac kinase (nanK)	EDN77077.1	Hypothetical protein	100	WP_004843625.1