16S sequencing and functional analysis of the faecal microbiome during treatment of newly diagnosed paediatric inflammatory bowel disease

Running head: The longitudinal microbiome in PIBD

James J Ashton\* 1, 2, Catherine M Colquhoun\* 3, David W Cleary4, Tracy Coelho1, 2, Rachel Haggarty2, Imke Mulder3,5, Akshay Batra1, Nadeem A Afzal1, R Mark Beattie Ɨ1, Karen P Scottɨ3and Sarah Ennisɨ 2

\* Denotes joint 1st author

Ɨ Denotes joint final author

1. Department of Paediatric Gastroenterology, Southampton Children’s Hospital
2. Department of Human Genetics and Genomic Medicine, University of Southampton
3. Gut Health Division, Rowett Institute, University of Aberdeen
4. Academic Unit of Clinical and Experimental Sciences, University of Southampton
5. 4D Pharma PLC, Aberdeen

For all correspondence

James J Ashton

Academic clinical fellow

Paediatric gastroenterology

University Hospital Southampton

Tremona Road

Southampton

SO16 6YD, UK

+44 (0)23 8120 8614

j.ashton@soton.ac.uk

**Contributorship**

SE, RMB, JJA and KPS conceived the study. Patients were recruited by JJA, TC and RH. Data were gathered by CMC and KPS. Data were analysed by JJA and DC with input from CMC and KPS. The manuscript was written by JJA with assistance from all authors. All authors approved the manuscript prior to submission.

**Conflicts of interest and financial declaration**

At the time of this research being conducted I Mulder was not associated with 4D Pharma. At the point of manuscript preparation she was employed in this role. There are therefore no conflicts of interest or financial interests to declare.

**Abbreviations**

PIBD; paediatric inflammatory bowel disease

CD; Crohn’s disease

UC; Ulcerative colitis

IBDU; inflammatory bowel disease unclassified

SCFA; short chain fatty acids

QIIME; Quantitative-Insights-Into-Microbial-Ecology

PiCRUST; Phylogenetic Investigation of Communities by Reconstruction of Unobserved States

**Abstract**

**Objectives-** The human microbiome is of considerable interest to paediatric inflammatory bowel disease (PIBD) researchers with one potential mechanism for disease development being aberrant immune handling of the intestinal bacteria. This study analyses the faecal microbiome through treatment in newly diagnosed PIBD patients and compares to co-habiting siblings where possible.

**Methods-** Patientswere recruited on clinical suspicion of PIBD, prior to diagnosis. Treatment-naïve faecal samples were collected, with further samples at 2 and 6 weeks into treatment. Samples underwent 16S ribosomal RNA gene sequencing and short chain fatty acids (SCFAs) analysis, results were analysed using Quantitative-Insights-Into-Microbial-Ecology (QIIME).

**Results-** Six PIBD patients were included in the cohort; 4 Crohn’s disease, 1 ulcerative colitis, 1 IBD unclassified, median age 12.6 (range 10 - 15.1 years); 3 patients had an unaffected healthy sibling recruited. Microbial diversity (observed species/Chao1/Shannon diversity) was reduced in treatment-naïve patients compared to siblings and patients in remission. Principal coordinate analysis using Bray-Curtis dissimilarity and UniFrac revealed microbial shifts in Crohn’s disease over the treatment course. In treatment-naïve PIBD there was reduction in functional ability for amino acid metabolism and carbohydrate handling compared to controls (p= 0.038) and patients in remission (p= 0.027). Metabolic function returned to normal after remission was achieved. SCFA revealed consistent detection of lactate in treatment-naïve samples.

**Conclusion-** This study adds in-depth 16S rRNA sequencing analysis on a small longitudinal cohort to the literature and includes sibling controls and patients with ulcerative colitis/IBD unclassified. It highlights the initial dysbiosis, reduced diversity, altered functional potential and subsequent shifts in bacteria from diagnosis over time to remission.

**Key Words-** 16S,Crohn’s disease, inflammatory bowel disease, paediatrics, microbiome

**Introduction**

Paediatric inflammatory bowel disease (PIBD), comprising Crohn’s disease (CD), ulcerative colitis (UC) and inflammatory bowel disease unclassified (IBDU) are a group of conditions leading to significant morbidity in children. An interaction between dysfunctional host immunity in the genetically susceptible individual and an environmental trigger appears to underlie disease pathogenesis (1, 2). Recent studies have shown an increasing incidence of PIBD within the United Kingdom but reasons behind this remain unclear (3, 4).

The microbiome is of considerable interest to IBD researchers. The intestinal microbiome consists of luminal and mucosal communities (5); a possible mechanism for disease development in PIBD is aberrant immune handling of the microbiome resulting in inflammation (6); alternatively a dysbiosis in the microbiota may trigger disease in immunologically normal individuals (7). Several studies focussing on the treatment naïve paediatric CD patient have made observations of significant mucosal microbial dysbiosis compared to healthy controls (8, 9) and this dysbiosis has also been observed in historically diagnosed, active PIBD (10, 11). Some studies comparing the faecal and mucosal microbiome have demonstrated differences in the faecal organisms seen in PIBD cases and controls (10-13), whilst others have shown little variation (8). It is widely accepted that alterations in the mucosal microbiome are more pronounced than those observed in the faecal microbiome (8, 12).

Despite the intense interest in the role of the microbiome in PIBD there have been few longitudinal studies (14). Dysbiosis of the gut microbiota in paediatric CD patients treated with exclusive enteral nutrition (EEN) has been broadly demonstrated in patients using denaturing gradient gel electrophoresis (15) and more recently through next generation sequencing including metagenomic analysis (14, 16).

With the advent of high throughput next-generation 16S rRNA gene sequencing, the microbiome can now be characterised in depth and detail that was not previously available (8). This study analyses the faecal microbiome from pre-diagnosis to 6 weeks of treatment in newly diagnosed, treatment naïve paediatric IBD patients, and additionally compares individuals to co-habiting sibling controls where available.

**Materials and Methods**

All patients were recruited following referral to Southampton Children’s Hospital. The paediatric gastroenterology service at Southampton diagnoses approximately 60-70 new cases of PIBD every year. Diagnosis is based on the Porto criteria (17).

Six PIBD patients were recruited, of whom two were male. Subsequent diagnosis indicated that 4 had Crohn’s disease, 1 ulcerative colitis and 1 IBD unclassified (See Table 1). The median age at diagnosis was 12.7 years. Patients with a suspected diagnosis of PIBD were recruited after the initial clinic visit, prior to investigation or treatment. Specific histological diagnosis was obtained within 2 weeks for all patients. Faecal samples were obtained prior to the commencement of treatment, and at 2-4 weeks and 6-8 weeks post diagnostic endoscopy. Clinical data on disease activity including faecal calprotectin were collected prospectively at each visit.

Single control samples were obtained from siblings living within the same household at the time of their sibling’s diagnosis.

Ethical approval

The study has ethical approval from Southampton & South West Hampshire Research Ethics Committee (09/H0504/125). This study was conducted in accordance with relevant guidelines, all protocols were approved by the ethics committee above and informed consent was obtained from all subjects.

Sample collection

Samples were collected using a commode specimen collection kit (Ardinal). All samples were processed within 24 hours. Prior to processing and during transit all samples were in sealed containers with some air and kept on ice in an insulated cool-bag at a temperature estimated to be less than 8°C.

Sample Processing

Samples were homogenised. Two grams of faeces was immediately frozen at -80°C for Short Chain Fatty Acid (SCFA) analysis. Five grams of faeces was separated for further processing. 10ml of PBS/30% glycerol solution was added and samples were further homogenised using sterile glass beads and vortex, these samples were then frozen immediately at -80°C in 450μl aliquots.

DNA extraction and 16S rRNA gene sequencing

DNA was extracted using MP Biomedical Faeces Extraction kit following the manufacturer’s instructions. 20ng of extracted DNA underwent processing at Source Bioscience (Nottingham, UK). The V4 area was targeted by amplification using the degenerate primers (forward TATGGTAATTGTGTGCCAGCMGCCGCGGTAA, reverse AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT). PCR products were between 300-350bp in length with an individual sample barcode incorporated into the reverse primer. PCR fragments were then sequenced on an Illumina MiSeq with two 150bp paired end reads. Steps in sequencing were as follows; 1st stage PCR, PCR clean-up, 2nd stage PCR, PCR clean-up 2, library quantification and normalization, library denaturing and MiSeq sample loading at 10pM.

Bioinformatics -microbiome analysis

16S rRNA gene sequencing data were processed using the quantitative insights into microbial ecology (QIIME) analysis pipeline. Bacterial community composition of a sample is achieved by clustering reads into operational taxonomic units (OTUs) and assigning taxonomy using the Greengenes (gg\_13\_8) database of 16S rRNA sequences (18). Clustering was performed using a 97% similarity index thereby allowing identification of bacteria to the species level. Further analysis of alpha and beta diversity was performed using QIIME.

Functional analysis

Metagenomic inference and functional analysis was performed using phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) (19). PICRUSt utilises a computational approach and predicts a metagenome using 16S data, from this it predicts which genes are present and thus predicts functional composition of the metagenome. Pathway analysis was conducted using Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways (20).

Output data from PICRUSt were filtered for metabolic pathways as these are assumed to be the most relevant for IBD (8). Median gene counts from patients at each time point (treatment naïve, control and remission) were compared. Differences were calculated between time points for each metabolic pathway to estimate functional differences in the microbiome.

Faecal calprotectin measurements

Estimation of Calprotectin concentrations were carried out using the CALPROLAB™ Calprotectin ELISA (ALP) kit and Calpro Faecal Extraction Device. 100mg of homogenised faeces was dispersed in 4.9ml of extraction buffer by vortexing for 3 min. The faecal extract and standards were assayed by sandwich ELISA following the manufacturer’s instructions. Sample values were calculated by comparison with standard curve plots prepared using the calprotectin standards provided in the kit.

SCFA analysis

Aliquots of homogenised faecal samples (~ 1 g) were mixed with sterile H2O (1:3 w/v), vortexed for 2 min, centrifuged for 10 min at 1000 × g and the supernatant collected and frozen. Faecal acetate, propionate, butyrate, lactate and total SCFAs were subsequently quantified in duplicate using capillary gas chromatography with 0.1 M 2-ethyl butyric acid as an internal standard, as described previously (21).

Statistics

Statistical analysis was with Shapiro-Wilk normality test (in order to test the distribution of data), Student’s T-test (alpha diversity) and Mann Whitney U test (comparison of groups in functional analysis) and was performed using SPSS v22 (IBM).

**Results**

Six PIBD patients were recruited; 4 Crohn’s disease, 1 ulcerative colitis and 1 IBD unclassified (See Table 1). Two were male. Median age at diagnosis was 12.7 years. All patients were diagnosed according to the modified Porto criteria (17). Three of these patients had a healthy sibling control recruited simultaneously. (faecal calprotectin values were 120, 32 and 24 mg/kg for sibling controls of patients 1-3, respectively). All recruited siblings were reported to be well at the time of investigation and have not developed PIBD subsequently. All patients provided faecal samples prior to commencement of bowel preparation for diagnostic endoscopy and further faecal samples at 2 weeks (Median 14 days, Range 12-25 days) and 6 weeks (Median 41 days, Range 36-49 days) after commencement of treatment. These 21 samples were characterised for microbiome composition, faecal calprotectin and SCFA concentrations (18 IBD, 3 sibling controls).

Disease activity and treatment

Disease activity and remission was assessed by physician assessment and subsequently in conjunction with faecal calprotectin levels (figure 1). Faecal calprotectin is a marker protein for colonic inflammation which can be stably detected in faeces.

Patients were treated with either exclusive enteral nutrition (EEN) (Modulen, Nestle) or corticosteroid therapy, depending on disease type and site, in accordance with national guidelines (22, 23). In all cases treatment was started within 24 hours of diagnosis (based on endoscopy findings).

Microbial composition analysis

A heatmap of common genera produced by Bray-Curtis dissimilarity (see below) can be seen in figure 2. This details the most common species seen across all samples and the relationship between samples and bacteria.

There was significant variation between all samples with significant intra-individual variation. Several patients had distinct microbiota at diagnosis. For example patient 1, and their sibling control had large numbers of *Prevotella* species (66% and 60% respectively). These siblings are of Nigerian ethnicity but were born in the UK, and these bacteria are known to be associated with a high fibre ‘rural African diet’ (24). With EEN this genus rapidly decreased over 2 weeks to below 0.0006%. Patient 2 (white British) had 35% *Prevotella* at diagnosis, again declining with EEN treatment. No other patient had greater than 13.5% prevalence of *Prevotella* at diagnosis. Patient 5’s (UC) treatment-naive sample was also distinctive with a high proportion of *Streptococcus* (and *Proteobacteria)* in comparison to the other patients; at a genus level this patient had 37% *Streptococcus* and 18% *Sutturella*.

After remission was reached (at 2-6 weeks, see table 1), the most commonly occurring bacteria were generally altered. There were increased numbers of *Bacteroides* in all patients, *Clostridium* classes were still prevalent but included a greater diversity of species. There was also an expansion of *Bifidobacterium* in 3 patients.

Alpha diversity

Three measures were used to analyse species diversity within samples. Data were normally distributed. The mean results can be seen in table 2. All 3 measures (observed species, chao1 and Shannon diversity) show an increase in diversity from diagnosis throughout treatment, although none reach statistical significance (see table 2). Controls have increased alpha diversity compared to patients at any time point but this is not statistically significant. Samples at 6 weeks post diagnosis, when all patients were in clinical remission, have the most similar diversity metrics to controls.

*Individuals during treatment*

Alpha diversity for all PIBD patients across the 3 time points can be seen in figure 3a-c. There is significant variation within the mean values with a tendency towards increased diversity across treatment.

*Treatment type*

Results were collated by treatment type. Increases in diversity from diagnosis to remission were observed for those treated with both exclusive enteral nutrition and steroid therapy (table 2).

*Antibiotics*

There were 3 patients who had been treated with antibiotics in the 6 weeks prior to initial sample (2 with metronidazole, 1 with amoxicillin). These were patients 2, 4 and 5. Comparing alpha diversity measures at diagnosis between these patients and those not treated with antibiotics showed decreases in the average number of observed species (antibiotics 9096 v no antibiotics 10333), Chao1 (26110 v 31150) and Shannon diversity (5.15 v 5.57).

Compositional changes in Microbiota

Principal coordinate analysis (PCoA) was undertaken for all samples following analysis through application of Bray-Curtis dissimilarity (figure 4A, BC, compositional dissimilarity between samples), unweighted UniFrac (figure 4B, UU, diversity between samples incorporating relative relatedness) and weighted UniFrac (figure 4C, WU, diversity between samples incorporating relative relatedness and relative abundance).

Both Bray-Curtis and UniFrac PCoA revealed high similarity between patient 1 (at diagnosis) and their sibling control, and patient 3 (at diagnosis) and their sibling control.

Patients with Crohn’s disease or UC showed a distinct shift on PCoA (BC, UU and WU) from diagnosis to treatment. The single patient with IBDU (patient 6) did not show any discernible shift with all 3 time points clustering on PCoA (figure 4A).

Metagenomic and Functional analysis

Output data from QIIME was analysed using PICRUSt enabling the metagenomic make-up of samples to be inferred from the 16S data. 147 metabolic pathways were identified by PICRUSt (see supplementary table 1). Subsequent analysis looked at the functional impact of microbial shifts. Comparison of treatment naïve patients at diagnosis with control samples and with samples from the patients in remission showed significantly different metabolic function across the 147 KEGG pathways identified, as assessed by gene copy number (p= 0.038 and p= 0.027 respectively) (figure 5). Patients in remission and controls showed no differences in metabolic function as assessed by gene copy number across the 147 KEGG pathways (p= 0.86). There were large differences (predicted functional reduction > 30%) in 31 pathways between treatment naïve patients and both controls and patients in remission. An additional 36 pathways displayed a predicted functional reduction of between 10-30%. Specifically there were large reductions in amino acid synthesis and metabolism pathways (particularly in those associated with arginine, proline, lysine, valine, leucine and isoleucine) and in carbohydrate handling, particularly in pathways associated with fructose, mannose, galactose, starch and sucrose metabolism. There was also a predicted reduction in nucleotide (purine and pyrimidine) metabolism and decreased nitrogen metabolism in treatment naive patients.

Short chain fatty acid analysis (SCFAs)

The faecal samples were analysed for SCFA content (patient five’s 6 week sample was unsuccessful) (figure 6). In all samples the predominant SCFAs were acetate, propionate and butyrate, in that order of abundance. Over treatment there were no dramatic shifts in relative concentrations of SCFA, and no consistent effects of either EEN or steroid treatment. Levels of succinate increased in patient 2 and patient 3 when remission was reached. Lactate was detected in all the samples (median concentration = 2.45mM), while it is often undetectable in studies of healthy individuals. In patient 5 there was a comparatively large amount of lactate prior to diagnosis (29.3mmol) which disappeared after 2 weeks of treatment. Patient 1 also had high lactate (15.3mmol) concentrations at diagnosis which again decreased during treatment, this time with an associated rise in butyrate concentrations.

SCFA and Functional analysis

Of the 147 metabolic pathways identified, those implicated in SCFA were examined. Two pathways (Butyrate metabolism and Propionate metabolism) were seen to be reduced in treatment naïve patients compared to controls. The relative abundance of these SCFA was observed; there were no statistically significant differences between treatment naïve (TN) patients and controls/patients in remission. Propionate relative abundance was slightly decreased (TN 20.4% vs Control/remission 18.1%), Butyrate was slightly increased (TN 11.8% vs Control/remission 13.6%) and isobutyrate was also slightly increased (TN 2.1% vs Control/remission 2.3%) (figure 6).

**Discussion**

This study details sequential analysis of the faecal microbiome of six PIBD (including CD, UC and IBDU) patients treated with standard therapies over a 6 week period. Our study provides in-depth analysis of the faecal microbiome in newly diagnosed treatment naïve PIBD patients and follows these through to remission. Theseresults including clear microbial shifts and detailed functional analysis suggests similar examination of larger numbers of patients could give insight into the aetiology of IBD. The utilisation of sibling controls is novel and it was hoped this would reduce the natural variation seen in the faecal microbiome caused by diet and environment (25). The value of this is apparent with the very high incidence of Prevotella observed in both patient 1 and their sibling control. Overall, the data generated in this study are valuable but further patient numbers are required to produce firm conclusions.

Previous studies have demonstrated a reduction in bacterial diversity in treatment naïve children with Crohn’s disease (14-16). We have shown similar profiles for all PIBD patients, there was reduced alpha diversity (consistently observed by 3 measures- observed species, Chao1 and Shannon index) in treatment naive patients compared to the same patients when remission was achieved. Both patient treatment groups, EEN and corticosteroids, had an increase in diversity for all measures (other than Shannon index in those treated with corticosteroids) during the treatment course.

Beta diversity was assessed through Bray-Curtis dissimilarity index and through weighted and unweighted UniFrac analysis. For all patients, other than patient 5 (UC), there are clear shifts on all analyses from prior to diagnosis to after treatment. Generally there was clustering of patients 1, 2, 3 and 4 (all Crohn’s disease) after treatment (samples at 2 weeks and 6 weeks) for both Bray-Curtis and UniFrac PCoA, although this was not ubiquitous for all principal coordinates. Patient 5 (UC) showed similar shifts in PCoA analysis to the patients with CD whereas there were no shifts for patient 6 (IBDU) with clustering of samples from all 3 time periods in all analyses. Interestingly controls were generally closely related to their affected sibling; with the closest clustering occurring between PIBD samples prior to diagnosis and their sibling control. This was especially evident for patient 1 and patient 3 who were closely related on both Bray-Curtis and unweighted UniFrac PCoA.

The predicted metabolic function of the microbiome was assessed. Despite low patient numbers, it appears that there is a clear difference in functional capability between the microbiome in treatment naïve patients and the same patients after they have gone into remission. Patients in remission resemble controls in terms of functional profile. Previous metagenomic analysis has characterised the functional changes of mucosal microbiome samples in CD, displaying clear evidence that there is reduced amino acid biosynthesis and carbohydrate metabolism (12).

This study has, for the first time, compared the functional ability of the microbiome prior to treatment to the functional ability in the same patients once they have achieved remission through inference of the metagenome by PICRUSt. Genes involved in both amino acid biosynthesis and metabolism, and carbohydrate metabolism (specifically starch, sucrose, fructose and mannose metabolism) were less abundant in the pre-treatment samples when the disease was active than in the later samples taken during remission. Interestingly the functional ability of the microbiome returns to a state very similar to that of controls when remission is achieved. Where there is significant inter- and intra-patient heterogeneity in the bacterial species present in IBD before and after treatment these results indicate that there is relative functional homogeneity.

This study also looked at SCFA concentrations in faeces over time. Previous work has suggested a role for specific SCFA as anti-inflammatory agents within the bowel (26, 27). The most common metabolic products of anaerobic bacteria are acetate, propionate and butyrate and these have been suggested to impact on the interaction that the immune system has with bacteria producing these SCFAs, with the overall effect of reducing inflammation through reduction of chemokine production and adhesion molecule expression (27). The predominant SCFAs in our samples were acetate, propionate and butyrate, including in controls. The relative concentrations of these SCFA over time did not show a consistent pattern with some patients seeing increases and some seeing decreases in both overall and relative concentrations. High concentrations of lactate, which has been implicated in ulcerative colitis were observed in the pre-treatment samples for this patient only (28). It is worth noting that SCFA detection in faecal samples is affected by both production and absorption of SCFAs in the large intestine. During disease remission the intestinal epithelial surface is less inflamed and may absorb more SCFA, so the absence of detectable changes in concentrations may not necessarily mean that bacterial fermentation has not changed.

The microbiome has previously been assessed in large cohorts of treatment naïve CD patients (8, 11). These have concluded that there is significant dysbiosis in ileal and rectal mucosal biopsy samples which is more weakly reflected in the faecal microbiome (8, 12). The same studies have also described shifts in the predominant bacterial species in diseased patients, with a general change to aerobic species in the faeces as inflammation increases (29). Our results indicate a similar shift with an increase in the anaerobic *Clostridia* and *Bacteroides* species from diagnosis to remission. It should be noted that the microbiota of faecal samples may be influenced by bowel preparation medication up to 2-4 weeks after administration.

Some studies have followed children with CD from diagnosis through treatment with EEN (14, 16, 30). Most of these studies have reported a decrease in diversity during treatment with EEN; our results demonstrate a different pattern over the period of EEN, with an increase in diversity, peaking at 2 weeks. Interestingly children treated with steroids also showed increased diversity from treatment to remission, with a peak in diversity at 6 weeks. The reasons behind this are unclear. There were 3 patients treated with antibiotics in the 2 months prior to recruitment (patients 2, 4 and 5); antibiotics are known have a significant impact on intestinal bacteria and may amplify intestinal dysbiosis in some cases (8, 31).

The strengths of this study lie in the robust and consistent data and sample collection by a single centre, and the linked assessment of both microbial composition and activity. The main limitation is the small patient number and heterogeneity of the population. Despite this we present detailed analysis of 16S rRNA sequencing data to a level not previously described in patients followed across time during treatment. There is little published data on the microbiome in Ulcerative Colitis or IBDU, Gevers *et al* included a small number of UC patients in their extended PCoA analysis in 2014 without obvious trends (8). Here we demonstrate a single patient with IBDU over time, whose microbiome appears to behave differently to patients with CD and UC in this study. The PCoA shift for the patient with UC resembles CD; whilst a previous study indicated substantial dysbiosis in faecal samples from children with severe colitis there have been no longitudinal studies of UC to date (10).

**Conclusion**

These data add in-depth 16S rRNA gene sequencing analysis on a longitudinal PIBD cohort, including patients with UC/IBDU to the literature. The inclusion of sibling controls helps to address differences that may be diet related. It highlights the initial dysbiosis, reduced diversity with dominance of single bacterial genera, and subsequent shifts in bacteria from prior to diagnosis over time to remission. Further characterisation of the SCFA composition of faeces would aid understanding of the roles SCFA may play in intestinal inflammation and microbe-host interaction. The functional ability of bacteria at presentation is significantly reduced in several key pathways however this returns to normal when remission is achieved, echoing previous findings targeting the mucosal microbiota. Further longitudinal studies on larger paediatric cohorts is required within this area.

**Acknowledgements**

JJA is funded by a National Institute of Health Research Academic Clinical Fellowship and has received an Action Medical Research training fellowship. TC is funded by a Crohn’s in Childhood research association fellowship. CMC received a PhD studentship from SULSA Spirit industrial studentship. The NGS analysis was made possible by the award of a grant from the Source Bioscience 110th year anniversary promotion to CMC. The Rowett Institute receives funding from the Scottish Government (RESAS).

**Tables and Figures**

**Figure 1- Changes in faecal calprotectin across time.** Solid line demonstrates the threshold for active intestinal inflammation (>200mg/kg), with the normal range 5-50 mg/kg. Patient numbers correspond to those described in Table 1.

**Figure 2- Heatmap of common genera in all samples across treatment produced using Bray-Curtis dissimilarity-** lighter colours indicates more abundance of that genera (white is most abundant). The degree of similarity of microbiota can be observed using the dendrogram on the x-axis. The degree of similarity of sample can be observed using the dendrogram on the y-axis.

**Figure 3- A- Observed species for individual patients across treatment, B- Chao1 Diversity for individual patients across treatment, C- Shannon Diversity for individual patients across treatment.** (○ indicates patients who went into remission at 2 weeks, □ indicates patients who went into remission at 6 weeks).

**Figure 4A-** **PCoA Bray-Curtis Dissimilarity PC1 v PC2,** D= prior to diagnosis, 2= 2 weeks into treatment, 6= 6 weeks into treatment, C= control. For patients details please see table 1

**Figure 4B-** **PCoA weighted UniFrac PC1 v PC2,** D= prior to diagnosis, 2= 2 weeks into treatment, 6= 6 weeks into treatment, C= control. For patients details please see table 1

**Figure 4C- PCoA unweighted UniFrac PC1 v PC2,** D= prior to diagnosis, 2= 2 weeks into treatment, 6= 6 weeks into treatment, C= control. For patients details please see table 1

**Figure 5- Comparison of 147 common metabolic pathways in IBD faecal microbiome samples-** markers indicate absolute difference in predicted function (gene copy number) between samples, a value of 0 indicates no difference in predicted function (gene copy number). Negative difference in gene copy numbers indicate loss of function in treatment naïve patients in that metabolic pathway. Differences are calculated by subtracting the predicted gene copy number in one group from another (see figure 5). For example the metabolic pathway associated with purine metabolism (1st pathway on x axis) displayed the greatest loss of function in treatment naïve samples compared to both control samples and samples from patients in remission. There was a median reduction of -3391471 (treatment naïve – control) and -3566025 (treatment naïve – remission) gene copies between samples indicating significantly reduced functional capability in this metabolic pathway in patients with treatment naïve IBD.

Controls and patients in remission were not statistically different (p=0.86). There are statistically significant differences in function (gene copy number) between both treatment naïve and controls (p=0.038), and treatment naïve and patients in remission (p=0.027) across all 147 KEGG pathways. Statistical analysis was with Shapiro-Wilk normality test (distribution of data) and Mann Whitney U test (functional analysis). Please see supplementary table 1 for clarification of KEGG pathways and exact gene copy number differences.

**Figure 6- Faecal SCFA relative concentration over time in PIBD patients and controls-** for absolute abundance see supplementary table 2.

**References**

1. Henderson P, van Limbergen JE, Wilson DC, Satsangi J, Russell RK. Genetics of childhood-onset inflammatory bowel disease. Inflamm Bowel Dis. 2011; 17: 346-61.

2. Kaser A, Zeissig S, Blumberg RS. Inflammatory bowel disease. Annu Rev Immunol. 2010; 28: 573-621.

3. Ashton JJ, Wiskin AE, Ennis S, Batra A, Afzal NA, Beattie RM. Rising incidence of paediatric inflammatory bowel disease (PIBD) in Wessex, Southern England. Arch Dis Child. 2014; 99(7):659-64.

4. Henderson P, Hansen R, Cameron FL, Gerasimidis K, Rogers P, Bisset WM, et al. Rising incidence of pediatric inflammatory bowel disease in Scotland. Inflamm Bowel Dis. 2012; 18: 999-1005.

5. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. Nature. 2007; 449(7164): 804-10.

6. Hisamatsu T, Kanai T, Mikami Y, Yoneno K, Matsuoka K, Hibi T. Immune aspects of the pathogenesis of inflammatory bowel disease. Pharmacology & therapeutics. 2013; 137(3): 283-97.

7. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. Nature. 2011; 474(7351): 307-17.

8. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The treatment-naive microbiome in new-onset Crohn's disease. Cell Host Microbe. 2014; 15(3): 382-92.

9. Haberman Y, Tickle TL, Dexheimer PJ, Kim M, Tang D, Karns R, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. J Clin Invest. 2015; 125(3): 1363.

10. Michail S, Durbin M, Turner D, Griffiths AM, Mack DR, Hyams J, et al. Alterations in the gut microbiome of children with severe ulcerative colitis. Inflammatory Bowel Diseases. 2012; 18(10): 1799-808.

11. Papa E, Docktor M, Smillie C, Weber S, Preheim SP, Gevers D, et al. Non-invasive mapping of the gastrointestinal microbiota identifies children with inflammatory bowel disease. PLoS ONE. 2012; 7(6).

12. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012; 13(9): R79.

13. Tjellstrom B, Hogberg L, Stenhammar L, Magnusson KE, Midtvedt T, Norin E, et al. Effect of exclusive enteral nutrition on gut microflora function in children with Crohn's disease. Scandinavian Journal of Gastroenterology. 2012; 47(12): 1454-9.

14. Quince C, Ijaz UZ, Loman N, Eren AM, Saulnier D, Russell J, et al. Extensive Modulation of the Fecal Metagenome in Children With Crohn's Disease During Exclusive Enteral Nutrition. Am J Gastroenterol. 2015; 110(12): 1718-29.

15. Leach ST, Mitchell HM, Eng WR, Zhang L, Day AS. Sustained modulation of intestinal bacteria by exclusive enteral nutrition used to treat children with Crohn's disease. Aliment Pharmacol Ther. 2008; 28(6): 724-33.

16. Kaakoush NO, Day AS, Leach ST, Lemberg DA, Nielsen S, Mitchell HM. Effect of exclusive enteral nutrition on the microbiota of children with newly diagnosed Crohn's disease. Clin Transl Gastroenterol. 2015; 6: e71-9.

17. Levine A, Koletzko S, Turner D, Escher JC, Cucchiara S, de Ridder L, et al. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. J Pediatr Gastroenterol Nutr. 2014; 58(6): 795-806.

18. http://greengenes.secondgenome.com/downloads/database/13\_5. Greengenes 16S rRNA gene database and tools 2015. Last accessed 13/12/2016

19. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol. 2013; 31(9): 814-21.

20. Kanehisa M, Goto M. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. 2000; 28(1): 27-30.

21. Richardson AJ, Calder AG, Stewart CS, Smith A. Simultaneous determination of volatile and non-volatile acidic fermentation products of anaerobes by capillary gas chromatography. Letters in App Micro. 2008; 9(1): 5-8.

22. Turner D, Levine A, Escher JC, Griffiths AM, Russell RK, Dignass A, et al. Management of pediatric ulcerative colitis: joint ECCO and ESPGHAN evidence-based consensus guidelines. J Pediatr Gastroenterol Nutr. 2012; 55(3): 340-61.

23. Ruemmele FM, Veres G, Kolho KL, Griffiths A, Levine A, Escher JC, et al. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. J Crohns Colitis. 2014; 8(10): 1179-207.

24. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. 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. 2010; 107(33): 14691-6.

25. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature. 2014; 505(7484): 559-63.

26. Tedelind S, Westberg F, Kjerrulf M, Vidal A. Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease. World J Gastroenterol. 2007; 13(20): 2826-32.

27. Vinolo MA, Rodrigues HG, Nachbar RT, Curi R. Regulation of inflammation by short chain fatty acids. Nutrients. 2011; 3(10): 858-76.

28. Hove H, Holtug K, Jeppesen PB, Mortensen PB. Butyrate absorption and lactate secretion in ulcerative colitis. Dis Colon Rectum. 1995; 38(5): 519-25.

29. Mimouna S, Gonçalvès D, Barnich N, Darfeuille-Michaud A, Hofman P, Vouret-Craviari V. Crohn disease-associated Escherichia coli promote gastrointestinal inflammatory disorders by activation of HIF-dependent responses. Gut Microbes. 2011; 2(6): 335-46.

30. Gerasimidis K, Bertz M, Hanske L, Junick J, Biskou O, Aguilera M, et al. Decline in presumptively protective gut bacterial species and metabolites are paradoxically associated with disease improvement in pediatric Crohn's disease during enteral nutrition. Inflamm Bowel Dis. 2014; 20(5): 861-71.

31. Jernberg C, Löfmark S, Edlund C, Jansson JK. Long-term impacts of antibiotic exposure on the human intestinal microbiota. Microbiology 2010; 156(11):3216-23.