Microbe-host associations in IBD biopsies revealed by cDNA amplicon sequencing and RNAseq

Nina Konstantinidou,[1,2] Christine Tataru,[2] Sidney Walker,[1] Fergus Shanahan,[1] Karim Dabbagh,[2] Todd DeSantis,[2] Marcus Claesson,[1] Shoko Iwai[2]

[1]APC Microbiome Ireland, University College Cork, Cork, Ireland; [2]Second Genome, Inc., South San Francisco, CA 94080, USA

Abstract: Gut microbiome associations with Inflammatory Bowel Disease (IBD) have been reported in multiple studies. However, the interaction between the active bacteria and the host gene expression has not yet been well understood. In this study, we analysed 185 biopsies of intestinal mucosa from subjects with 58 Crohn's Disease (CD), 88 Ulcerative Colitis (UC) and 39 Controls. 16S rRNA gene amplicon sequencing from inflamed and Control biopsies was performed on extracted DNA and RNA, hereafter termed DNA-16S and RNA-16S, respectively. Host gene expression analysis was performed via RNAseq. DADA2 de-noised sequences were used for microbial differential abundance and activity analysis using metagenomeSeq package in R statistical environment. DESeq2 package was used for the host RNAseq analysis. DNA-16S demonstrated one and four dynamically depleted Ribosomal Sequence Variants (RSVs) in CD and UC respectively, compared to the Controls. Whereas, five and six RSVs from RNA-16S were dynamic and differentially expressed in CD and UC respectively. Of those, an *Anaerostipes hadrus* RSV was significantly depleted in UC in DNA-16S,in both CD and UC in RNA-16S. Following on these findings, we compared RNA-16S levels of *A. hadrus* to host RNAseq gene expressions. Termination of O-glycan biosynthesis pathway was significantly enriched, which could be an advantage of this organism to survive and prosper in this environment. These results demonstrated differences between abundant and metabolically active bacteria, as well as specific microbiota association with the host gene expressions.

Methods

Study Design Total of 185 biopsies of intestinal mucosa were collected (Table 1). Biopsy samples were obtained from inflamed tissues of adult IBD patients or healthy tissues of control subjects. Exclusion criteria was the use of antibiotics one month prior or during the GI investigation.

Microbiota analysis DNA and RNA were extracted using AllPrep DNA/RNA/Protein Mini kit (Qiagen). Total RNA was reverse transcribed to cDNA. 16S rRNA V3-V4 hypervariable region was amplified using 341F and 805R primer set. Amplified DNA (DNA-16S) and cDNA (RNA-16S) were sequenced using Illumina MiSeq for 2x300 bp reads. DADA2 (Callahan et al., 2016) was used to denoise DNA-16S and RNA-16S sequences. Strain level taxonomy was assigned to Ribosomal Sequence Variants (RSVs) using StrainSelect 2016 (http://secondgenome.com/strainselect). Differential test was conducted using R package, metagenomeSeq.

Host RNAseq analysis Aliquots of RNA samples were used for host transcriptome RNAseq using TruSeq Stranded mRNA Sample Prep Kit (Illumina) with Illumina HiSeq 4000 2x100bp reads. Quality filtered reads were aligned to the human genome (GRCh38) using HiSat2 and a count table was generated using SUBREAD. Differential expression analysis on host RNAseq was performed using DESeq2 package (Love et al., 2014). Pathway enrichment analysis was conducted using R package, ReactomePA.

Study cohort

Table 1. Study cohort demographics.

		CD	UC	Control	
Number of total subjects		58	88	39	
	DNA-16S	42	61	29	
	RNA-16S	46	75	25	
	Host RNAseq	52	83	27	
Gender (% Female)		45	48	41	Chi-Square p=0.74
Age (y) Mean ± SD		41.6 ± 12.3	47.7 ± 13.0	58.4 ± 11.6	ANOVA p<0.001
Colon location					
	Ascending	4	10	3	
	Cecum	1	13	1	
	Descending	3	2	7	
	Rectum	2	1	11	
	Sigmoid	5	1	12	
	Transverse	3	1	0	
Diagnosis period (y) Mean ± SD		11.3 ± 8.68	10.7 ± 8.14	NA	ANOVA p=0.67

Age was a potential confounding factor with ANOVA p<0.001. However, the highest correlation coefficients were rho=0.29 and 0.3 for DNA-16S and RNA-16S RSVs, respectively, and none of them were significant (adjusted p>0.05).

Microbial composition of RNA-16S samples demonstrated a shift from DNA-16S samples

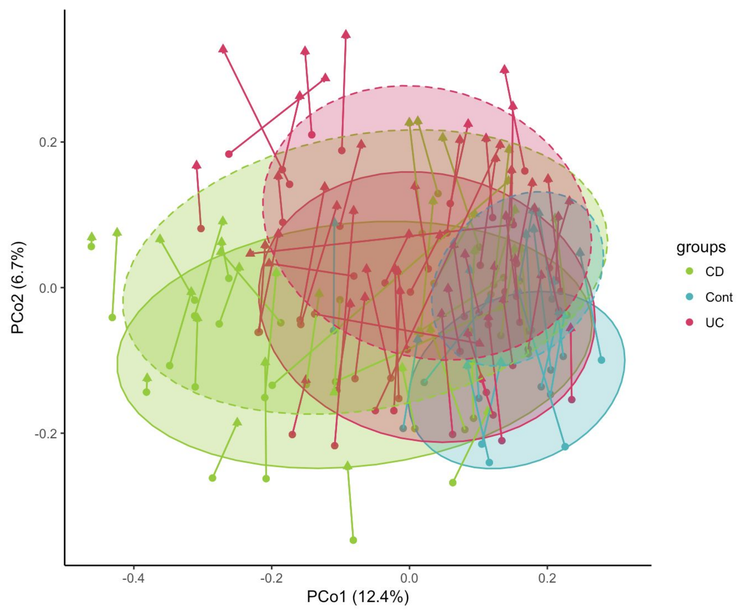

Figure 1. DNA-16S and RNA-16S data were compared using 102 samples which have both datasets. From the total of 486 RSVs, 262 were shared between two datasets, and 51 and 173 RSVs were unique to DNA-16S and RNA-16S, respectively. Principal Coordinate Analysis (PCoA) on Bray-Curtis dissimilarity matrix indicated a marked shift from DNA-16S (circles) to RNA-16S (triangles), confirmed by PERMANOVA test (p=0.001).

Host RNAseq confirmed inflammation in CD and UC compared to Controls

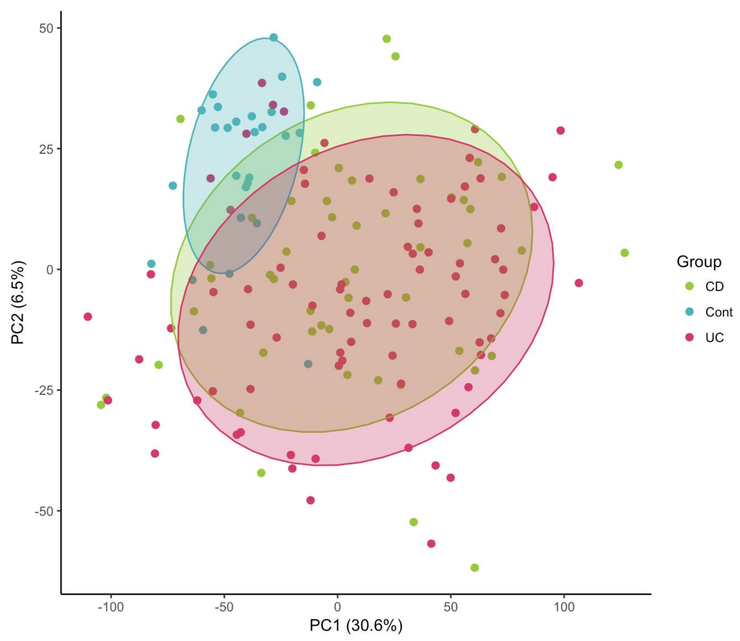

Figure 2. Disease status (CD, UC or Control) was significantly associated with host gene expression profile (PERMANOVA p=0.001). Pathway enrichment analysis demonstrated significant change in IL-4/IL-13, Extracellular matrix organization, Collagen degradation, and IL-10 signaling pathways in both CD and UC patients compared to controls.

Anaerostipes hadrus metabolic activity was significantly depleted in both UC and CD

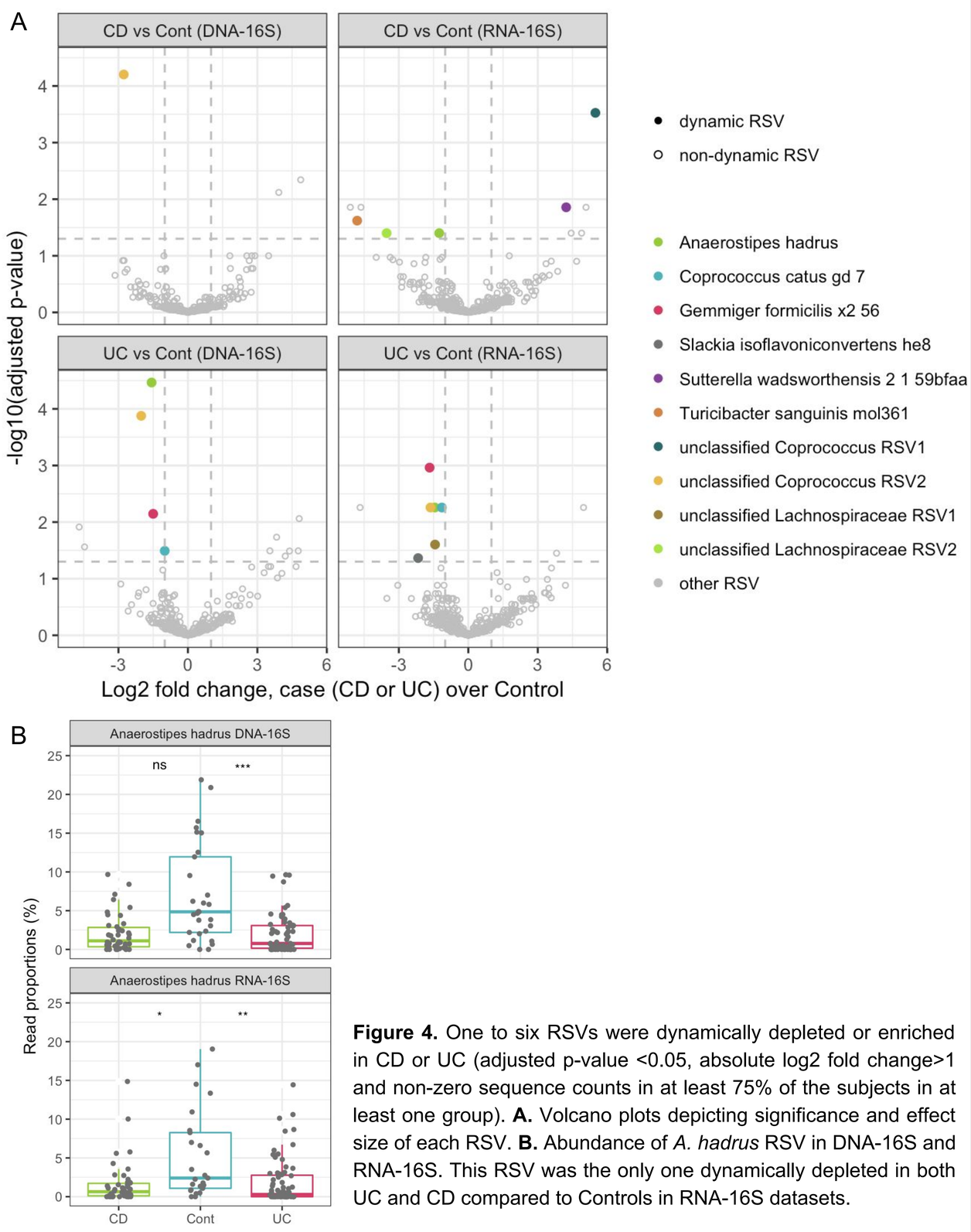

Figure 4. One to six RSVs were dynamically depleted or enriched in CD or UC (adjusted p-value <0.05, absolute log2 fold change>1 and non-zero sequence counts in at least 75% of the subjects in at least one group). **A.** Volcano plots depicting significance and effect size of each RSV. **B.** Abundance of *A. hadrus* RSV in DNA-16S and RNA-16S. This RSV was the only one dynamically depleted in both UC and CD compared to Controls in RNA-16S datasets.

A. hadrus metabolic activity was associated with termination of O-glycan biosynthesis pathway

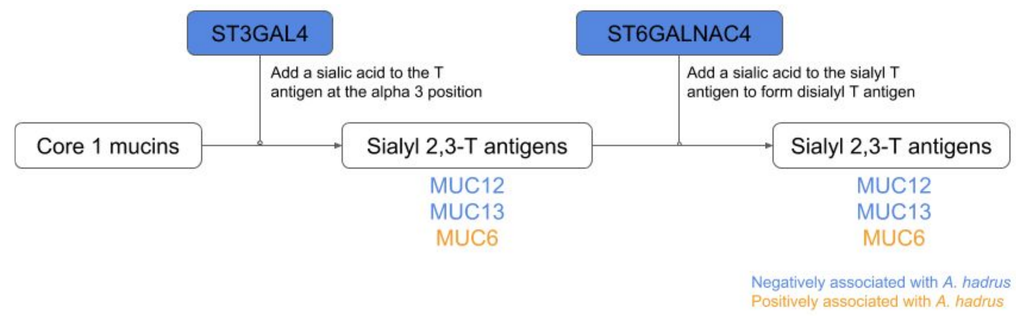

Figure 5. Termination of O-glycan biosynthesis pathway was significantly enriched by pathway enrichment analysis (adjusted p=0.002). Terminal modification of glycans are known to affect bacterial adhesion (Baos et al., 2012) and could serve as substrates for specific microbiota to provide nutritional advantage (Pacheco et al. 2012).

Conclusions

- We analysed one of the largest IBD mucosal cohorts published so far by DNA and RNA amplicon sequences for microbiota and RNAseq for host gene expressions.
- Shift was observed between microbiota composition of DNA-16S and RNA-16S.
- Host RNAseq confirmed significant enrichment of inflammatory pathways in UC or CD.
- Depleted metabolic activity of *Anaerostipes hadrus* was identified in both UC and CD. The activity was associated with termination of O-glycan biosynthesis pathway, which could be an advantage of this organism to survive and prosper in this environment.

References

Callahan et al. (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581–583.
Love et al. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.
Baos et al. (2012) Distribution of sialic acids on mucins and gels: A defense mechanism. Biophys J 102:176-184.
Pacheco et al. (2012) Fucose sensing regulates bacterial intestinal colonization. Nature 492:113-117.

Acknowledgements

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (grant numbers 11/SIRG/B2162 and SFI/12/RC/2273), the Irish Health Research Board (grant number PD/2009/30), and Second Genome, Inc., South San Francisco, California, USA. We also thank research nurses Margot Hurley, Catherine O'Riordan and Mary Looby for their assistance.

SECOND GENOME
THE MICROBIOME COMPANY