Fecal metatranscriptomics of macaques with idiopathic chronic diarrhea reveals altered mucin degradation and fucose utilization

Samuel T. Westreich[1], Amir Ardeshir[2], Zeynep Alkan[3], Mary E. Kable[3,4], Ian Korf[1] and Danielle G. Lemay[1,3,4*]

Abstract

Background: Idiopathic chronic diarrhea (ICD) is a common cause of morbidity and mortality among juvenile rhesus macaques. Characterized by chronic inflammation of the colon and repeated bouts of diarrhea, ICD is largely unresponsive to medical interventions, including corticosteroid, antiparasitic, and antibiotic treatments. Although ICD is accompanied by large disruptions in the composition of the commensal gut microbiome, no single pathogen has been concretely identified as responsible for the onset and continuation of the disease.

Results: Fecal samples were collected from 12 ICD-diagnosed macaques and 12 age- and sex-matched controls. RNA was extracted for metatranscriptomic analysis of organisms and functional annotations associated with the gut microbiome. Bacterial, fungal, archaeal, protozoan, and macaque (host) transcripts were simultaneously assessed. ICD-afflicted animals were characterized by increased expression of host-derived genes involved in inflammation and increased transcripts from bacterial pathogens such as *Campylobacter* and *Helicobacter* and the protozoan *Trichomonas*. Transcripts associated with known mucin-degrading organisms and mucin-degrading enzymes were elevated in the fecal microbiomes of ICD-afflicted animals. Assessment of colon sections using immunohistochemistry and of the host transcriptome suggests differential fucosylation of mucins between control and ICD-afflicted animals. Interrogation of the metatranscriptome for fucose utilization genes reveals possible mechanisms by which opportunists persist in ICD. *Bacteroides* sp. potentially cross-fed fucose to *Haemophilus* whereas *Campylobacter* expressed a mucosa-associated transcriptome with increased expression of adherence genes.

Conclusions: The simultaneous profiling of bacterial, fungal, archaeal, protozoan, and macaque transcripts from stool samples reveals that ICD of rhesus macaques is associated with increased gene expression by pathogens, increased mucin degradation, and altered fucose utilization. The data suggest that the ICD-afflicted host produces fucosylated mucins that are leveraged by potentially pathogenic microbes as a carbon source or as adhesion sites.

Keywords: Metatranscriptomics, ICD, Macaque, Microbiome, RNA-seq, Dysbiosis, Pathology, Gut microbiome, Fecal microbiome, Mucin, Gastrointestinal, Diarrhea, Inflammation, Mucosal immunity, Metatranscriptome, Colitis

* Correspondence: danielle.lemay@ars.usda.gov
[1]Genome Center, University of California, Davis, California, USA
[3]USDA ARS Western Human Nutrition Research Center, Davis, California, USA
Full list of author information is available at the end of the article

Background

Rhesus macaques (*Macaca mulatta*) provide a close analog for humans for the study of various conditions, including SIV/AIDS, and are maintained in captive colonies for research purposes [1]. Captive macaque juveniles often develop idiopathic chronic diarrhea (ICD), characterized by repeated episodes of intestinal distress and intestinal inflammation, and lack the presence of known intestinal pathogens detectable by culture-based methods [2, 3]. Successive episodes of ICD cause chronic dehydration. ICD is responsible for the majority of non-medical captive macaque culls [2]. Although pathogenic bacteria have been observed in a minority of cases [4, 5], their presence is not a consistent disease marker. Antibiotics alleviate acute outbreaks, but symptoms re-emerge weeks or months after treatment. Roughly 3–5% of all captive macaques in colonies at the California National Primate Research Center (CNPRC) develop ICD annually [2]. ICD in macaques phenotypically resembles a sub-category of ulcerative colitis in humans in which the ulcers are microscopic. ICD is histologically characterized by lymphocytic, plasmacytic inflammation of lamina propria, goblet cell depletion, crypt hyperplasia, and abscesses [2, 3, 6]. Prior work demonstrated that microbial diversity is decreased in macaques with colitis [7] and that therapeutic helminth infection ameliorates the clinical symptoms of ICD that are associated with the restoration of microbial diversity [3]. However, it is not yet known which features of a microbiome, beyond diversity, are protective against ICD.

Composition of the species dwelling in the gut microbiome varies with diet [8–10], host genetics [11], and a mixture of other genetic and environmental variables [12]. The majority of microbial species in the colon subsists on breakdown products of complex polysaccharides, derived either from the host diet or from host-derived mucins [8, 13, 14]. When adequate fiber is not provided through diet, or as a result of other triggers that are not yet fully understood, the microbiome may dynamically change activity and enzyme production may change when food sources are switched. Increased microbial consumption of host epithelial mucin degrades the barrier matrix between the microbiome and intestinal wall faster than it may be replenished by goblet cells, leading to increased contact between the microbiome and the epithelial wall, greater pathogen susceptibility, and potential invasion by opportunistic pathogens [8, 15–17]. Thus, although inflammation and disease may be triggered by opportunistic pathogens, their ability to gain access to the intestinal wall may be due to a weakening of the mucosal matrix from increased microbial degradation and consumption of the mucin layer [18].

Recent observations of the gut microbiome have focused less on individual species and more on the set of functional niches, each filled by one or more "specialist" capable of preferentially consuming a particular substrate [18]. Evaluation of these activities is possible through application of metagenomic or metatranscriptomic approaches. Although 16S rRNA studies of rhesus macaques have examined the makeup of the macaque gut microbiome [2, 3, 7, 19, 20], there is a lack of understanding about host-microbe and microbe-microbe relationships in ICD. Metatranscriptomics—the genes expressed by all organisms in a sample—can provide observability into which organisms are present and what they are doing [21]. We therefore conducted a metatranscriptomics study of fecal samples from macaques with ICD and their controls to test the hypothesis that ICD is associated with increased gene expression by mucin-degrading bacteria and/or pathogens. We first investigated host-derived transcripts to confirm that inflammation is increased in ICD. We next surveyed the community-wide bacterial metatranscriptome and screened the transcriptomes for specific known bacterial and non-bacterial pathogens. We then probed transcripts specifically from mucin-degrading bacteria and mucin-degrading enzymes. Upon discovering an increase in bacterial fucosidases, we investigated the differential fucosylation of colon sections via immunohistochemistry and further analyzed the metatranscriptome for fucose utilization. The metatranscriptomes suggest two mechanisms by which bacterial pathogens leverage fucosylated mucins in ICD.

Results

Host macaque transcripts confirm increased inflammation in ICD

Prior reports of ICD indicate that it is a form of inflammatory bowel disease of the colon [2, 3]. To verify this phenotype in our samples, we investigated the host transcriptome, which is known to include intestinal epithelial cells of the colon that are exfoliated in fecal samples [22]. RNA reads mapped to rhesus macaque (host) proteins were examined to compare host transcript abundance in control and ICD samples. There were 562 host transcripts that were robustly increased in ICD (log2 fold change ≥ 2 relative to control; adjusted $p \leq 0.05$) and 358 host transcripts increased in control animals (log2 fold change ≥ 2 relative to ICD; adjusted $p \leq 0.05$). Functional enrichment analyses yielded major clusters for transcripts increased in ICD, including "collagen/extracellular matrix(ECM)", "ECM-receptor interaction/PI3K-AKT signaling," "inflammatory bowel disease", and "Toll-like receptor signaling pathway." Upstream regulator analysis revealed well-known inflammatory regulators interferon gamma (IFNγ), tumor necrosis factor (TNF), interleukin-1 beta (IL1β), and interferon alpha (IFNα) as highly significant ($-\log(p\text{-value}) > 30$) in ICD animals relative to controls. For host transcripts that

were elevated in control relative to ICD, there were no significant upstream regulators.
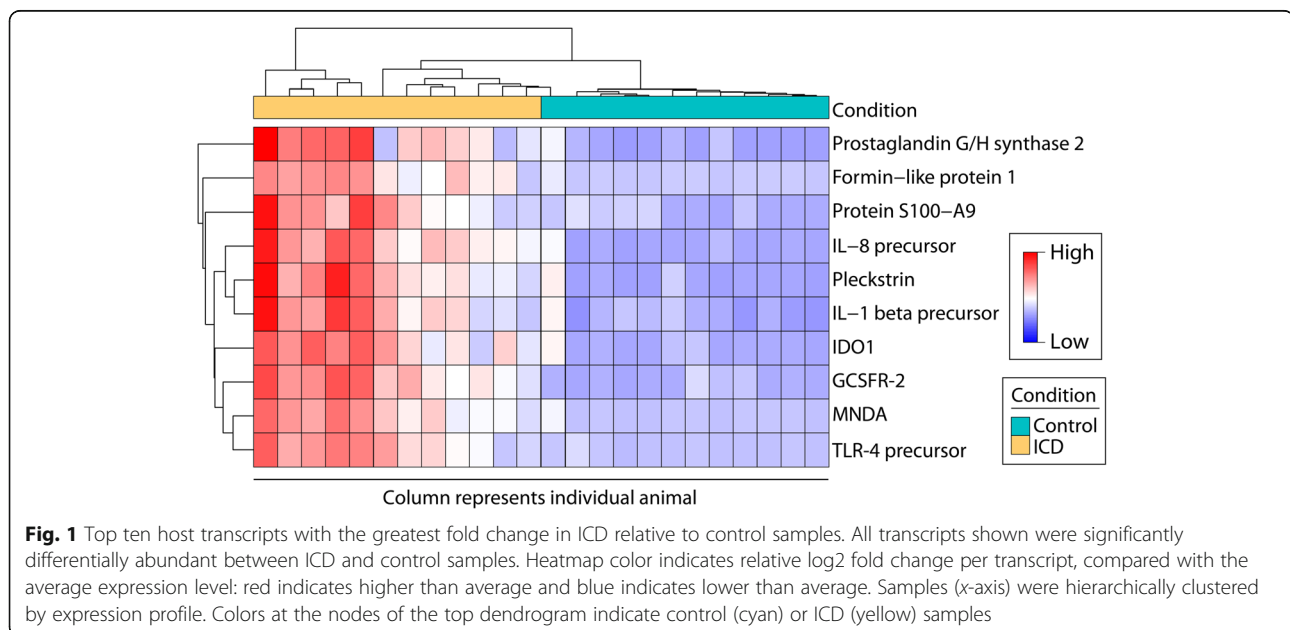
Ten host transcripts were 32× higher in ICD than controls (Fig. 1). Prostaglandin G/H synthase 2 (PTGS2, also known as cyclooxygenase-2 or COX-2) is a downstream target of innate immune signaling via the Toll-like receptor 4 (TLR4) signaling pathway. Formin-like protein 1 (FMNL1) is a major component of macrophage podosomes, which is an integrin-based structure that is unique to invasive cells [23]. Protein S100-A9 forms a heterodimer with S100-A8 (also increased in ICD, log2 fold change = 3.80, adjusted p = 0.0002). This heterodimer is known as calprotectin, which comprises more than half of the cytosolic protein in neutrophils [24]. Fecal calprotectin is frequently used as a marker of intestinal inflammation [25]. Interleukin-8 precursor (IL-8, also known as CXCL8) is expressed by intestinal epithelial cells to recruit neutrophils from the lamina propria to the epithelium [26]. Pleckstrin (PLEK) is most highly expressed in neutrophils and macrophages (using GeneVisible [27]). Interleukin-1 beta (IL1β) is a pro-inflammatory cytokine that is secreted by activated macrophages and also by colonocytes [28]. Indoleamine 2,3-dioxygenase 1 (IDO1) is an enzyme that is the rate-limiting step in tryptophan catabolism. IDO1 is highly expressed by antigen-presenting cells to reduce tryptophan levels in the local microenvironment, which reduces activated T cell responses and promotes regulatory T cell activity [29]. In addition to antigen-presenting cells, colonocytes also express IDO1 during inflammation [30]. Granulocyte colony-stimulating factor receptor (CSF3R, also known as CD114) is expressed by neutrophils [31]. Myeloid cell nuclear differentiation antigen (MNDA) can be expressed by innate immune cells or intestinal epithelial cells

to activate inflammasomes [32]. TLR4 is a receptor expressed by intestinal epithelial cells that recognize bacterial PAMPs, such as lipopolysaccharide [33]. Taken together, these most highly differentially expressed genes suggest that the stool samples from animals with ICD contain cells which express a pro-inflammatory profile.

To determine which host cell types might be differentially present in the stool samples, the expression of cell marker-associated genes were examined. Keratins KRT8, KRT18, KRT19, and KRT20 were all more highly expressed in ICD, confirming that exfoliated colonocytes were present in the stool and more abundant in ICD. Genes CD14, CD44, CD53, and CD97 were more highly expressed in ICD, suggesting that a heterogeneous population of leukocytes may be present in ICD. The high expression of CSF3R, S100-A8, and S100-A9 was supportive of the presence of neutrophils. Thus, the host transcriptome provided some evidence that there were more exfoliated colonocytes and leukocytes in the stool of ICD-afflicted animals compared with controls. Collectively, the host cells in stool from animals with ICD included colonocytes as well as leukocytes, and these cells exhibited a pro-inflammatory transcriptional program.

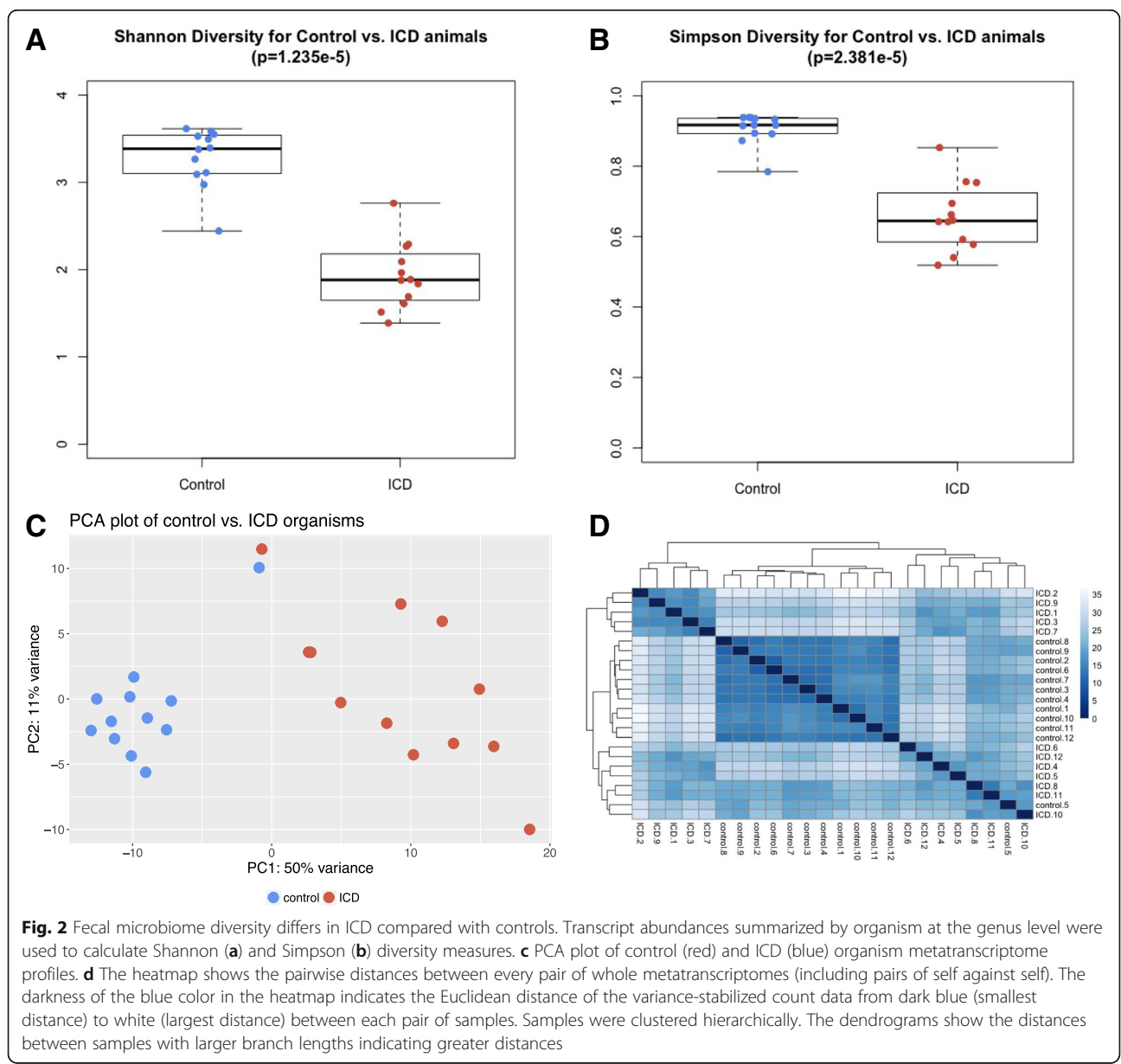## Fecal metatranscriptomes had decreased diversity in ICD samples and clustered by phenotype

Using all RNA-Seq reads that mapped to bacterial proteins, diversity among the microbial populations was calculated based on transcript abundances, with a 0.1% abundance cutoff threshold for organism identification at the genus level. Comparison of the calculated diversity metrics between samples from control and ICD-afflicted animals showed a significant decrease in diversity in

Fig. 1 Top ten host transcripts with the greatest fold change in ICD relative to control samples. All transcripts shown were significantly differentially abundant between ICD and control samples. Heatmap color indicates relative log2 fold change per transcript, compared with the average expression level: red indicates higher than average and blue indicates lower than average. Samples (*x*-axis) were hierarchically clustered by expression profile. Colors at the nodes of the top dendrogram indicate control (cyan) or ICD (yellow) samples

ICD. Using the Shannon diversity calculation, controls had an average diversity score of 3.28, compared with the ICD-associated average diversity score of 1.93 ($p = 1.245e$ −5) (Fig. 2a). Similarly, by the Simpson diversity calculation, there was a higher average diversity among control samples when compared with ICD (0.906 vs. 0.656, $p = 2.380e-5$) (Fig. 2b).

Using the same organism transcript abundance measurements with the 0.1% threshold, a principal component analysis (PCA) (Fig. 2c) demonstrated that control and ICD phenotypes clearly segregated, with greater variation seen across ICD samples when compared to the control samples. Hierarchical clustering of whole transcriptomes also showed clustering of control samples (Fig. 2d). However,

ICD-associated metatranscriptomes split into two clusters, with some ICD samples (ICD.8, ICD.11, ICD.10) more similar to controls than to other ICD samples. These data suggest greater heterogeneity among the ICD samples. In both the PCA and clustering analyses, control animal 5's metatranscriptome was more similar to the ICD animals than other controls. Upon re-inspection of life histories, individual macaque control 5 was in the hospital for more days than any other control animal, so it is possible that this animal was not truly a healthy control. However, for the purposes of the remaining analyses, this sample was retained among the controls. In summary, microbial transcriptome diversity was decreased in ICD samples, but heterogeneity between ICD samples was increased.

Fig. 2 Fecal microbiome diversity differs in ICD compared with controls. Transcript abundances summarized by organism at the genus level were used to calculate Shannon (**a**) and Simpson (**b**) diversity measures. **c** PCA plot of control (red) and ICD (blue) organism metatranscriptome profiles. **d** The heatmap shows the pairwise distances between every pair of whole metatranscriptomes (including pairs of self against self). The darkness of the blue color in the heatmap indicates the Euclidean distance of the variance-stabilized count data from dark blue (smallest distance) to white (largest distance) between each pair of samples. Samples were clustered hierarchically. The dendrograms show the distances between samples with larger branch lengths indicating greater distances

Control and ICD samples harbored significant differences in organism-specific gene expression

Using all mRNA reads mapped to bacterial organisms, transcript counts were summarized at the genus level. The genera with the highest transcript abundances in control samples included *Prevotella, Bacteroides, Treponema, Ruminococcus, Clostridium, Streptococcus, Oscillibacter, Megasphaera, Bacteroidales*, and *Eubacterium* (Fig. 3). The genera with the highest transcript abundances in ICD samples included *Prevotella, Bacteroides, Megasphaera, Eubacterium, Clostridium, Ruminococcus, Roseburia, Faecalibacterium, Parabacteroides*, and *Lactobacillus*. Although many genera

were common between both control and ICD samples, their relative transcript abundances differed; importantly, the "Other" category (all other genera not represented by the ten most abundant) was significantly greater in control samples (Fig. 3). Examination of differential gene expression revealed many differences between ICD and controls with 732 bacterial genera, out of 1511 total detected, showing significantly different expression levels (Additional file 1). Among the genera with the highest transcript abundances, the largest difference was observed in *Prevotella*, which were greatly increased (2.7-log2 fold change, adjusted $p = 2.68e{-}16$) in the ICD-afflicted animals relative to controls. There were also
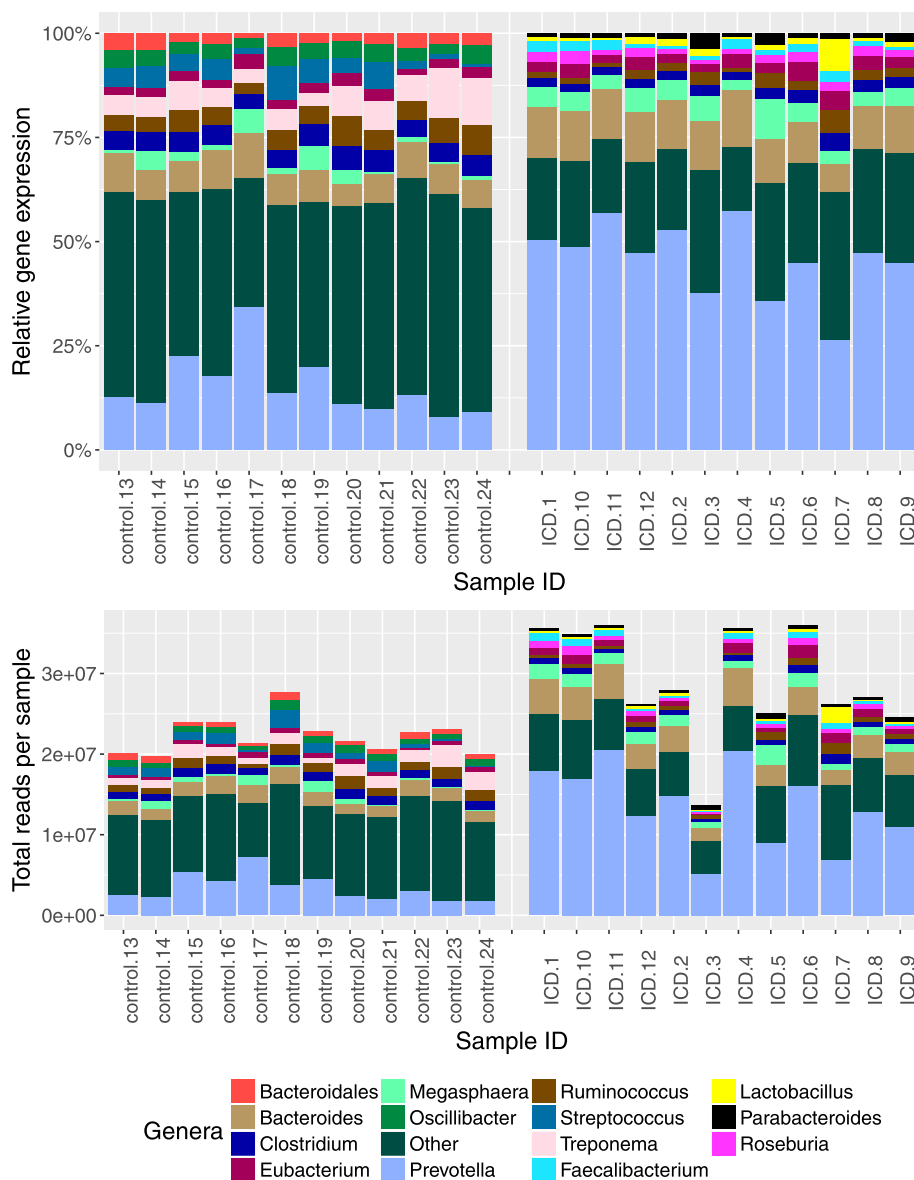

Fig. 3 Composition of genera with the highest number of transcripts in stool samples from control and ICD-afflicted macaques. Stacked bar graphs are relative (top) and absolute (bottom) transcript counts per genus in each sample. Only the top ten genera from either control or ICD samples are depicted; all remaining genera are summarized as the "Other" category (dark green)

increases in ICD in *Bacteroides* (1.8-log2 fold change, adjusted $p = 1.09\text{e}{-}13$), *Megasphaera* (2.0-log2 fold change, adjusted $p = 4.82\text{e}{-}6$), *Selenomonas* (2.2-log2 fold change, adjusted $p = 3.56\text{e}{-}14$), and *Campylobacter* (2.7-log2 fold change, adjusted $p = 2.54\text{e}{-}9$), whereas there was a decrease in *Treponema* (1.7-log2 fold decrease, adjusted $p = 7.2\text{e}{-}14$). Thus, while many genera expressed genes in both control and ICD samples, their relative expression differed substantially between the two phenotypes, with the largest significant fold changes being among abundant organisms in *Prevotella* and *Campylobacter* genera.

Community-wide and *Prevotella*-specific transcriptomes suggest differences in nutrient availability between control and ICD samples

To further explore microbial community-wide differences in the metatranscriptomes between control and ICD samples, reads were mapped to a database derived from the SEED Subsystems hierarchy [34]. Comparison of the ICD cases to controls revealed changes even at the broadest (level 1) SEED hierarchy categories. The largest increases were in membrane transport (1.23-log2 fold increase, adjusted $p = 1.76\text{e}{-}29$), sulfur metabolism (0.75-log2 fold increase, adjusted $p = 1.20\text{e}{-}13$), and carbohydrate-linked functions (0.58-log2 fold increase, adjusted $p = 7.72\text{e}{-}23$). Notable decreases in ICD-afflicted microbiome functions were seen in the categories of secondary metabolism (2.02-log2 fold decrease, adjusted $p = 2.82\text{e}{-}8$), dormancy and sporulation (1.16-log2 fold decrease, adjusted $p = 3.59\text{e}{-}7$), nitrogen metabolism (0.99-log2 fold decrease, adjusted $p = 7.1\text{e}{-}6$), and iron acquisition and metabolism (0.98-log2 fold decrease, adjusted $p = 9.84\text{e}{-}13$). (Fig. 4). Within the sulfur metabolism category, one significantly increased enzyme was sulfatase (1.8-log2 fold increase, adjusted $p = 9.92\text{e}{-}7$), which may be one of the inflammation-contributing factors [35]. The SEED level 2 annotations "Central carbohydrate metabolism," "Monosaccharides," "Polysaccharides," and "Di- and oligosaccharides" were all increased in ICD (Additional file 2), suggesting ample access to carbon sources [35]. Thus, the overall functional annotation of the bacterial community-wide transcriptome suggests that the fecal microbiome from ICD had better access to nutrients than the control fecal microbiome. This may possibly be due to the faster transit time of diarrhea, which has been shown to alter microbial composition, diversity, and activity [36].
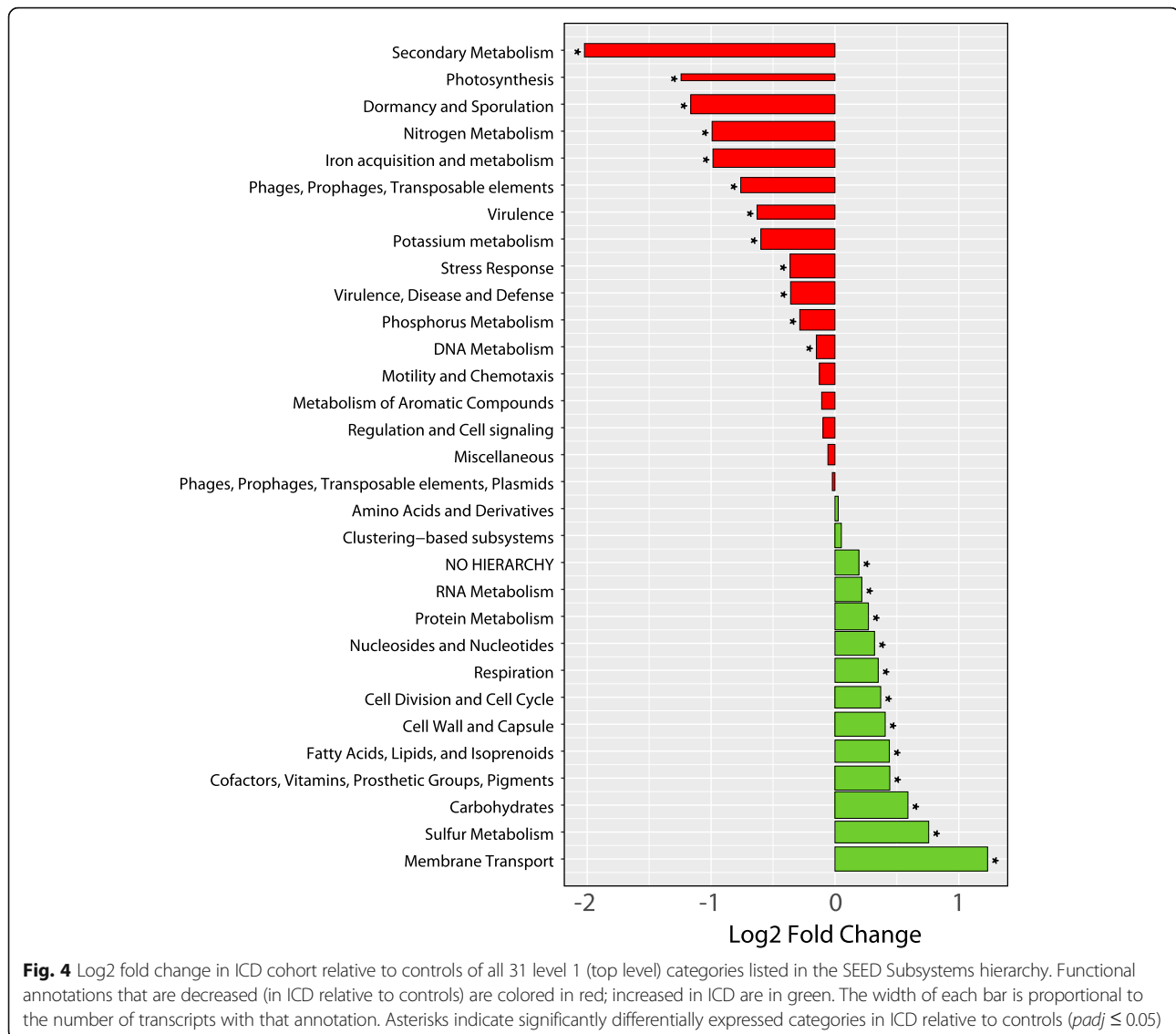
To determine whether the functional differences of the ICD metatranscriptome may be due to faster-growing microbes in the fecal samples as would be typical of the small intestine, we examined annotations related to growth rate. Epithelial cells of the small intestine have a higher turnover rate than cells in the colon [37], and shorter transit times (e.g., diarrhea) have been linked to faster growth of some microbial species [38]. At the community level, "DNA replication" was not significantly different between control and ICD, but the level 3 annotation, "RNA_polymerase_bacterial" was increased in ICD (log2 fold change 0.55, $p = 0.0001$). The concentration of RNA polymerase is known to increase during exponential growth [39]; thus, it is possible that fecal microbes in ICD were growing faster than fecal microbes in controls, which would be consistent with more simple carbohydrate sources reaching the distal colon.

Given that *Prevotella* had greater relative abundance of transcripts in ICD, we investigated the functional annotations of genes specifically expressed by this genus. The largest significant increases in ICD relative to control samples were in the categories of respiration (0.74-log2 fold increase, adjusted $p = 5.35\text{e}{-}25$) and membrane transport (0.68-log2 fold increase, adjusted $p = 9.98\text{e}{-}24$). The largest significant decreases were in the categories of stress response (0.89-log2 fold decrease, adjusted $p = 1.32\text{e}{-}13$), potassium metabolism (0.70-log2 fold decrease, adjusted $p = 3.03\text{e}{-}5$), and phage and transposable elements (0.59-log2 fold decrease, adjusted $p = 9.82\text{e}{-}7$) (Additional file 3). Like the community-wide metatranscriptome, the *Prevotella*-specific transcriptome also had increased "Central carbohydrate metabolism," "Monosaccharides," "Polysaccharides," and "Di- and oligosaccharides" in ICD relative to controls. We next asked whether *Prevotella* had greater "DNA replication" in ICD, but there was actually a slight decrease (-0.14-log2 fold, adjusted $p = 0.0041$). "RNA_polymerase_bacterial" was statistically unchanged. In summary, the *Prevotella*-specific transcriptome suggests that *Prevotella* in fecal samples of animals with ICD may have had greater access to nutrients without being in an exponential stage of growth.

Transcripts from bacterial pathogens were more abundant in ICD animals

Several pathogenic species have been previously linked with ICD [2–4], although none has definitively been proven to be responsible for the condition. We therefore examined the metatranscriptome data for transcripts that mapped to these known pathogens: *Camplyobacter coli* (*C. coli*), *Camplyobacter jejuni* (*C. jejuni*), *Helicobacter pylori* (*H. pylori*), *Shigella flexneri* (*S. flexneri*), and *Yersinia enterocolitica* (*Y. enterocolitica*). Interestingly, some gene expression for all of these pathogens was detectable even in control samples. However, relative transcriptional abundances of all of these known bacterial pathogens were significantly higher in the ICD samples (Fig. 5). The largest differences between control and ICD were in *Campylobacter* species.

Fig. 4 Log2 fold change in ICD cohort relative to controls of all 31 level 1 (top level) categories listed in the SEED Subsystems hierarchy. Functional annotations that are decreased (in ICD relative to controls) are colored in red; increased in ICD are in green. The width of each bar is proportional to the number of transcripts with that annotation. Asterisks indicate significantly differentially expressed categories in ICD relative to controls (*padj* ≤ 0.05)

Screening of fungi, protozoa, and archaea yielded potential non-bacterial opportunists

To determine whether non-bacterial opportunists may potentially be active in ICD, fecal metatranscriptome reads were mapped to fungal, protozoan, and archaeal databases. The fungus, *Sordaria macrospora*, had higher transcript abundances in ICD (3.18-log2 fold increase, adjusted $p = 6.33e-07$). Another species from this genus, *Sordaria fimicola*, is known to be present in mammalian dung [40]. *Sordaria fimicola* was not in the reference data set, nor was any other species of *Sordaria*, so it is possible that the active species in macaques was *Sordaria fimicola* due to the limitations of the data. Interestingly, transcripts associated with the known fungal opportunist, *Candida albicans*, were not differentially abundant between control and ICD animals.

Non-human primates are vulnerable to several enteric protozoa, including *Entamoeba histolytica*, *Giardia lamblia*, *Cryptosporidium parvum*, *Cyclospora cayetanensis*, and *Balantidium coli* [41]. *E. histolytica* and *G. lamblia* were among the protozoans with the highest gene expression in this cohort, but their transcripts were not differentially abundant between control and ICD animals. RNA-Seq reads mapping to *Cryptosporidium parvum* were relatively lower in this cohort, compared with *E. histolytica* and *G. lamblia*, and reads mapping to *Cyclospora cayetanensis* were exceedingly rare and likely to be off-target hits. There were no species from the *Balantidium* genus in the data set.

Protozoans with the most abundant transcripts in the fecal samples from the macaques were *Blastocystis* sp. and *Trichomonas vaginalis*. *Blastocystis* sp. was higher in the control animals (1.96-log2 fold decrease, adjusted
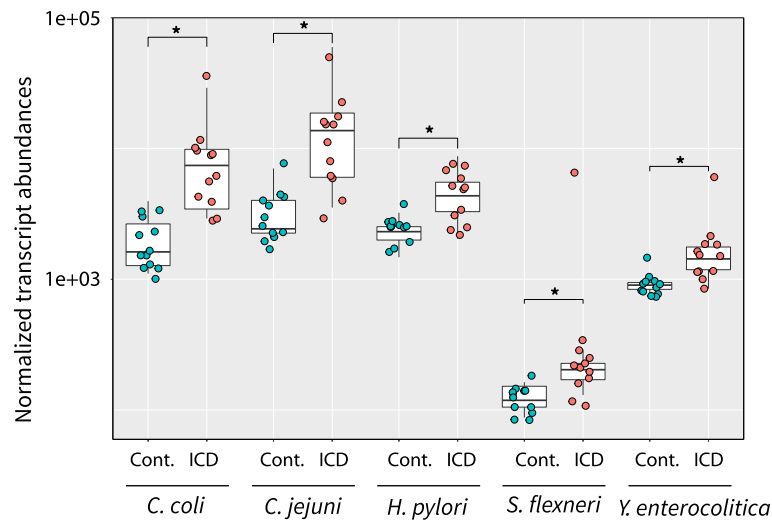
Fig. 5 Normalized transcript abundance levels associated with known pathogens in control and ICD samples. Boxplots summarize the distributions of transcript abundances for each pathogen in each control (blue) or ICD (red) samples with individual samples denoted by dots. Stars indicate statistical significance of differences between control and ICD samples (*padj* ≤ 0.05). Normalized transcript abundances are shown on a log scale

$p = 0.03$). *T. vaginalis* was higher in the animals with ICD (2.24-log2 fold increase, adjusted $p = 0.002$). *T. vaginalis* was the only species in the reference data set representative of the *Trichomonas* genus, so it is possible that the particular species with increased gene expression in macaques with ICD was not *T. vaginalis*.

Although there are no known pathogens among archaea, some archaeal genera are known to be gut-associated. Among the ten most abundant archaeal genera detected in the fecal samples, six were significantly different between control and ICD animals and all of these were lower in ICD. These include *Methanosarcina* (0.16-log2 fold increase, adjusted $p = 0.004$), *Methanococcus* (0.16-log2 fold decrease, adjusted $p = 0.018$), *Methanobacterium* (0.13-log2 fold decrease, adjusted $p = 0.009$), *Methanocorpusculum* (0.21-log2 fold decrease, adjusted $p = 0.044$), and *Methanocella* (0.18-log2 fold decrease, adjusted $p = 0.003$). These genera are predominately methane producers. Methane-producing archaea may be depleted in ICD; however, this observation may be confounded by gut transit time, which is reduced by the presence of methane [42]. In summary, ICD was associated with increased gene expression by a fungus (*Sordaria*) and a protozoan (*Trichomonas*) and decreased gene expression by archaeal methanogens.

Campylobacter expresses mucosa-associated transcripts in ICD

Of all potential bacterial and non-bacterial pathogens (excluding viruses, which are not represented in the metatranscriptome), *Campylobacter* transcripts were the most differentially abundant between ICD and controls. We therefore investigated the functional annotations of

genes specifically expressed by this genus. Functional annotation differences between control and ICD animals (Additional file 4) indicate that organisms within this genus experienced more oxidative stress and were more adherent and likely more virulent in ICD relative to control animals. Fermentation, and specifically the abundance of anaerobic respiratory reductases, was reduced in ICD relative to control, suggesting that *Campylobacter* was exposed to a more oxygen-rich environment in the ICD intestine compared with control. Abundance of genes related to oxidative stress and the stringent response ((p)ppGpp_metabolism) were higher in ICD-afflicted animals, whereas genes related to iron acquisition and metabolism were lower in ICD relative to control. This is expected given the inflammatory environment of the colon during ICD. Control of iron acquisition is a common microbial mechanism for reducing exposure to damage via the Fenton reaction.

The oxidative stress that was clearly experienced by *Campylobacter* could potentially be due to a more adherent population in ICD. Genes involved in adhesion were significantly more abundant in ICD animals as were genes for motility and chemotaxis, including flagellum (Additional file 4). Flagellar motility is required for *Campylobacter* colonization [43] and invasion of cultured intestinal epithelial cells [44–46]. This tighter affiliation with the host likely facilitates both increased oxidative stress for the *Campylobacter* present in ICD animals, as described above, and increased virulence. The stringent response, which aids in survival of oxygen stress conditions, has also been shown to be necessary for attachment and invasion of intestinal epithelial cells

[47]. Further evidence of a more virulent phenotype in ICD animals was that the pVir plasmid was more nearly four times as abundant (1.8-log2 fold, $p = 4.66e{-}04$) in ICD animals as in control animals (Additional file 4). Although the definitive role of this plasmid in the virulence of *Campylobacter* has yet to be uncovered, it contains genes with homology to type four secretion system genes in *Helicobacter pylori* and *Agrobacterium tumefaciens*. The mutation of several of these genes affects *C. jejuni* adhesion and invasion in an intestinal epithelial cell culture model [48]. Overall, the *Campylobacter*-specific transcriptome suggests that *Campylobacter* were more closely associated with the mucosa in ICD than in controls.

Transcripts from mucin-degrading bacteria and mucin-degrading enzymes were increased in ICD

Having screened the metatranscriptomes for potential pathogens, we next investigated our hypothesis that ICD is associated with increased mucin degradation. First, we subsetted the data to compare abundances of transcripts from bacterial species known to be linked with inflammatory bowel conditions [49], although they are also considered to be commensals. Significantly higher transcript levels were observed in ICD for *Ruminococcus gnavus*, *Ruminococcus torques*, *Bacteroides thetaiotaomicron*, *Bacteroides caccae*, and *Bacteroides fragilis* (Fig. 6a). Abundances of transcripts associated with several *Bifidobacterium* species (*B. bifidum*, *B. breve*, *B. longum*) were slightly, but significantly, higher as well. In short, transcripts associated with all known mucin-degrading gut microbes except for *Akkermasia muciniphilia* were significantly more abundant in the feces of animals with ICD. These organisms are associated with both degradation of the protective mucin layer that forms on the inside of the gut epithelium and with increased presence in human irritable bowel disease [49–51].

Given that mucin-degrading species (Fig. 6a) are also known to be commensals that can thrive on dietary carbohydrates, we next examined the abundances of transcripts arising from the expression of specific enzymes known to be involved in mucin degradation. All of these mucin-degrading enzymes were differentially expressed between ICD and controls (Fig. 6b). Transcript abundances from six of the seven mucin-degrading enzymes—alpha- and beta-galactosidase, alpha-fucosidase, alpha-mannosidase, *N*-acetylglucosaminidase, and *O*-acetylesterase—were all higher in ICD. In general, transcripts from both mucin-degrading bacteria and mucin-degrading enzymes were higher in ICD.

Intestinal cells synthesize fucosylated mucins in ICD

Given the differential expression of mucus-degrading enzymes in ICD compared with controls, we hypothesized that control and ICD-afflicted animals may have different mucus composition. Colon sections from control and ICD animals were stained with FITC-conjugated lectins for mannose and fucose (Fig. 7a). All images were selected and analyzed in a blinded manner (animal status unknown). Animals with ICD had markedly higher fucose relative to controls ($p = 0.008$) (Fig. 7b). No changes in intensity were observed for mannose levels between control and ICD animals. These observations suggest that the colons of animals with ICD have greater amounts of fucose than the controls.

Given that transcripts for microbial fucosidases were higher in ICD-associated metatranscriptomes and fucose was more abundant in the ICD-associated colon sections, we asked whether the host cell transcriptome was consistent with what would be expected of increased synthesis of fucosylated mucins. The most highly expressed mucin transcripts were mucin-12 (MUC12), mucin-2 (MUC2), mucin-17 (MUC17), and mucin-13 (MUC13). All of these mucins are known to be expressed by enterocytes in the colon. They were all significantly increased in ICD relative to controls: mucin-12, 2.1-log2 fold change, adjusted $p = 2.01e{-}5$; mucin-2, 2.3-log2 fold change, adjusted $p = 3.98e{-}6$; mucin-17, 1.3-log2 fold change, adjusted $p = 0.005$; and mucin-13, 1.6-log2 fold change, adjusted $p = 1.77e{-}5$. Overall, host fucosyltransferase gene expression was low across all samples. However, the most abundant fucosyltransferase, galactoside 3(4)-L-fucosyltransferase (FUT3), was more highly expressed in ICD (1.5-log2 fold change, adjusted $p = 0.02$) than the controls. Interestingly, the second most abundant fucosyltransferase, galactoside 2-alpha-L-fucosyltransferase 2 (FUT2), which is known to confer secretor status in humans [52], was not differentially expressed between ICD and control. In summary, the immunohistochemistry of colon sections and the examination of the host cell transcriptome suggest that ICD is associated with increased fucosylated mucins.

Fucose utilization in ICD

Given the increased abundance of transcripts from mucin-degrading enzymes and the increased availability of fucose in the colon of animals with ICD, we next investigated how fucose might be utilized. First, the metatranscriptome was subsetted by the annotation "alpha-L-fucosidase" to determine which microbes were liberating fucose. The top five species expressing this enzyme were all members of the *Bacteroides* genera: *B. fragilis*, *B. xylanisolvens*, *B. vulgatus*, *B. thetaiotaomicron*, and *B. helcogenes*. Expression of alpha-L-fucosidase by these species was significantly increased in ICD relative to controls.

The liberation of fucose from fucosylated mucins by these species is known to increase the amount of available fucose in the lumen [53]. We therefore investigated the expression of "fucose permease" to determine which microbes were importing fucose. The species with the
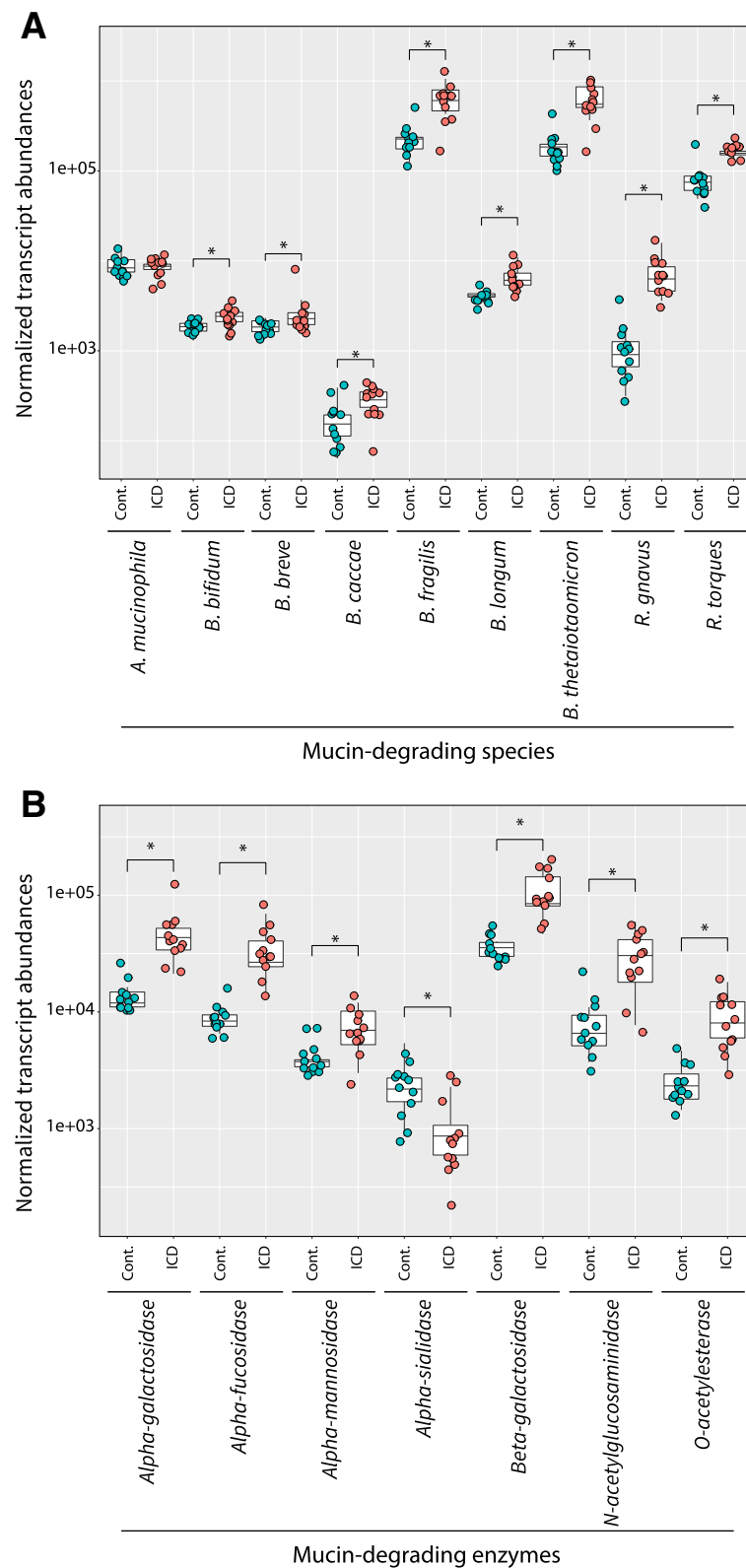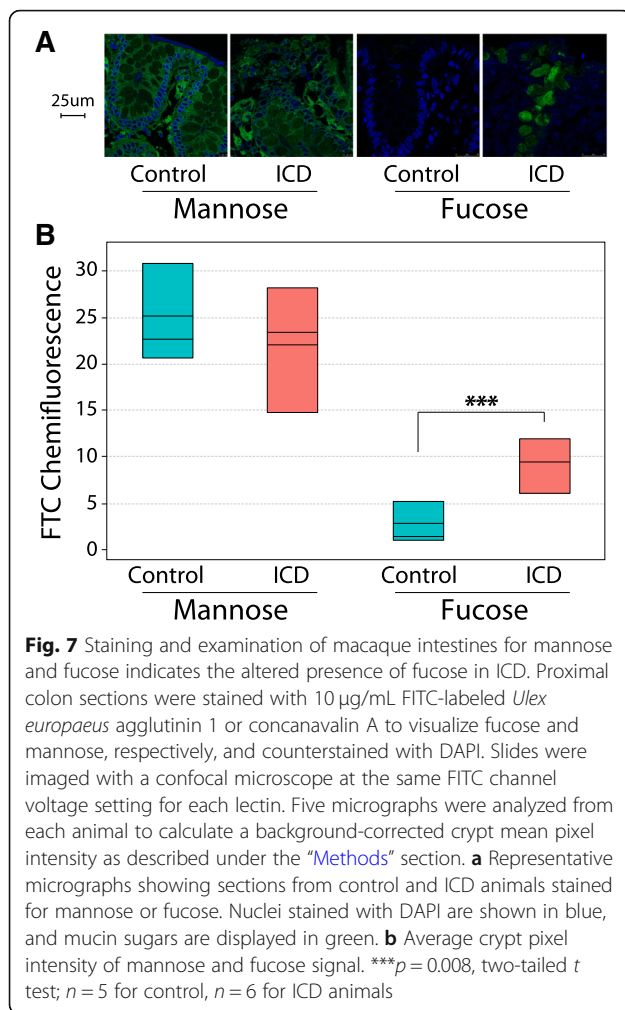
Fig. 6 Normalized transcript abundance associated with mucin-degrading species (**a**) and mucin-degrading enzymes (**b**). Boxplots summarize the distributions of transcript abundances for the each species(**a**) or enzyme (**b**) within each group of control (blue) or ICD (pink) samples with individual samples denoted by dots. Stars indicate statistical significance of differences between control and ICD samples (*padj* ≤ 0.05). Normalized transcript abundances (*y*-axis) are shown on a log scale

Fig. 7 Staining and examination of macaque intestines for mannose and fucose indicates the altered presence of fucose in ICD. Proximal colon sections were stained with 10 μg/mL FITC-labeled *Ulex europaeus* agglutinin 1 or concanavalin A to visualize fucose and mannose, respectively, and counterstained with DAPI. Slides were imaged with a confocal microscope at the same FITC channel voltage setting for each lectin. Five micrographs were analyzed from each animal to calculate a background-corrected crypt mean pixel intensity as described under the "Methods" section. **a** Representative micrographs showing sections from control and ICD animals stained for mannose or fucose. Nuclei stained with DAPI are shown in blue, and mucin sugars are displayed in green. **b** Average crypt pixel intensity of mannose and fucose signal. ***$p = 0.008$, two-tailed t test; $n = 5$ for control, $n = 6$ for ICD animals

most significantly increased abundance of "fucose permease" transcripts in ICD were *Bacteroides salanitronis*, *Prevotella dentalis*, *Bacteroides fragilis*, *Bacteroides thetaiotaomicron*, *Bacteroides vulgatus*, and *Haemophilus influenzae*. These results suggest that *Bacteroides* could potentially be cross-feeding fucose to *Prevotella* and *Haemophilus* in ICD.

An overview of mechanisms proposed by the results in this study is given in Fig. 8. *Bacteroides* sp. is known to stimulate host production of fucosylated mucins [54]. We found that enterocyte mucins and fucosyltransferase-3 transcripts were elevated in the host transcriptome. Immunohistochemistry of colon sections also illustrated the much higher presence of fucose in ICD relative to controls. Multiple *Bacteroides* sp. in ICD expressed more alpha-L-fucosidase, which liberates fucose from fucosylated mucins in the lumen. *Bacteroides* sp. as well as other microbes, such as *Prevotella* and *Haemophilus*, expressed fucose permeases which would have enabled them to import fucose, which likely improves their persistence in ICD. Interestingly, *Campylobacter* did not express more transcri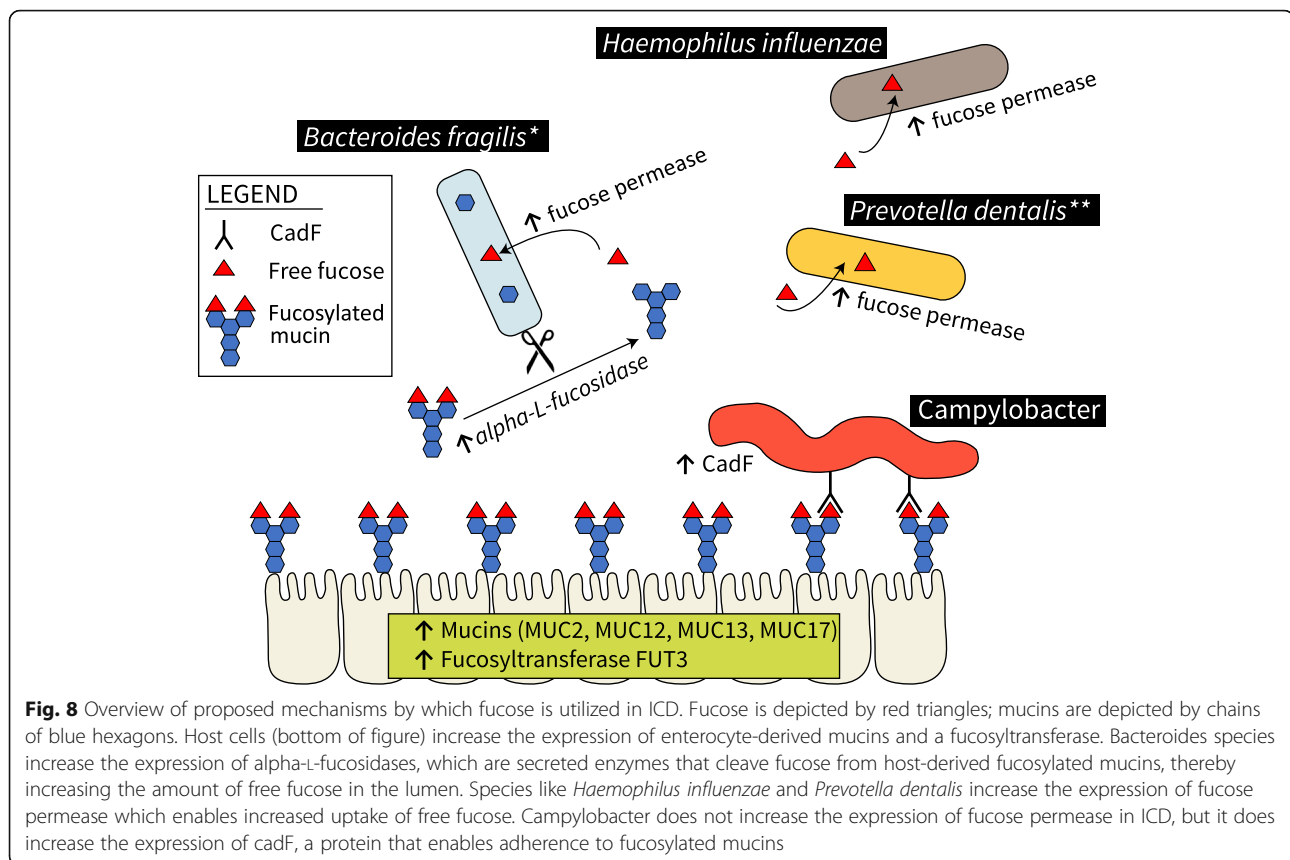pts that give rise to proteins which import fucose; instead, it expressed more transcripts for adherence proteins, like cadF, that would have enabled it to adhere to fucosylated mucins.

Discussion

Previous work showed that the gut microbiome of ICD had reduced diversity [55], but the salient features of that microbiome remained unknown. In the current study, we confirmed the reduced diversity at the transcriptional level and identified numerous novel features, such as relatively higher expression of mucin-degrading enzymes and of transcripts from known pathogens, such as *Campylobacter*, and altered fucose utilization.

Some of the novel features of ICD could be due to the capturing of fast-transit material from the small intestine, as might be expected of a diarrhea sample. For example, transcriptional features such as decreased diversity, lower stress, better access to simple sugars, and lower methane production are all consistent with what might be expected of a small intestine microbiome [56] or a faster transit time [36, 38]. However, mucin foraging is a primary feature in the colon; thus, our findings of increased gene expression by mucin-degrading bacteria, increased expression of mucin-degrading enzymes, increased expression of mucin and fucosylation by host cells, and increased fucose utilization by specific bacteria suggest these strong signals of the metatranscriptome in ICD do not arise purely from the proximal intestine. Furthermore, the colon sections, which were extracted from the same part of the colon in both ICD and control animals, showed differential presence of fucose.

Transcript matches to organisms at the genus level revealed domination of the ICD-associated fecal microbiome by multiple *Prevotella* sp. and a smaller but still significant increase by *Bacteroides*. Elevated *Prevotella* presence is also noted in other dysbiotic conditions, such as HIV [57] and intestinal and systemic inflammation in humans [58]. It has additionally been shown, using a dextran sodium sulfate (DSS)-treated mouse model, that the human intestinal *Prevotella* isolate, *P. copri*, is capable of exacerbating externally induced inflammation [59]. On the other hand, *Prevotella* is known to be increased in the microbiomes of healthy individuals habitually consuming a long-term vegetarian diet [60]. Therefore, it is possible that increased transcripts from *Prevotella* in ICD are merely a marker of increased carbohydrate availability. Relative to control samples, central carbohydrate metabolism was highly increased in the *Prevotella*-specific transcriptome during ICD. Annotations also suggest that *Prevotella* was not under oxidative stress. *Prevotella* may have filled nutritional niches for food carbohydrates after antibiotic exposure, which is common for macaques with repeated episodes of diarrhea.

Fig. 8 Overview of proposed mechanisms by which fucose is utilized in ICD. Fucose is depicted by red triangles; mucins are depicted by chains of blue hexagons. Host cells (bottom of figure) increase the expression of enterocyte-derived mucins and a fucosyltransferase. Bacteroides species increase the expression of alpha-L-fucosidases, which are secreted enzymes that cleave fucose from host-derived fucosylated mucins, thereby increasing the amount of free fucose in the lumen. Species like *Haemophilus influenzae* and *Prevotella dentalis* increase the expression of fucose permease which enables increased uptake of free fucose. Campylobacter does not increase the expression of fucose permease in ICD, but it does increase the expression of cadF, a protein that enables adherence to fucosylated mucins

Prevotella and *Bacteroides* are often considered to be antagonistic; gut microbiome communities are usually abundant in *Prevotella* or *Bacteroides* but not both [61]. One explanation for the prevalence of both genera in ICD could be related to cross-feeding. The annotation "Sialic_acid_Metabolism" was increased in the *Prevotella*-specific transcriptome during ICD, but not "sialidase" (Additional file 3). Meanwhile, *B. thetaiotaomicron* has sialidases and can free sialic acid, but lacks the genetic capability to consume free sialic acid [62]. It is therefore possible that *Bacteroides* cross-feeds *Prevotella* in ICD.

One of the drawbacks of our metatranscriptomic approach is that we were unable to distinguish the difference between many transcripts from fewer bacteria compared with fewer transcripts from many bacteria. For this purpose, metagenomics combined with metatranscriptomics would have yielded a more powerful data set and should be considered for future studies. Nevertheless, the metatranscriptomic analysis was immensely useful to reveal microbe-microbe and host-microbe interactions.

We investigated the hypothesis that ICD was associated with pathogens that were missed via traditional culture-based methods. We observed increased gene expression by numerous bacterial pathogens, primarily *Campylobacter* and *Helicobacter*. Observed increases in

gene expression by *Campylobacter* and *Helicobacter* in the ICD samples are consistent with previous observations of increased abundance in gut dysbiosis [4, 19, 63]. *Helicobacter* was previously positively linked with macaque ICD [4], and *Campylobacteraceae* was more commonly found in macaques with colitis [7]. Interestingly, our data suggests that these organisms are present in both ICD and control samples, but their transcripts are more abundant in ICD.

We also found elevated gene expression by a fungal genus and a protozoan genus. Increased transcripts were observed from the *Sordaria* genus of fungi in ICD; this genus is commonly found in the intestine of herbivores [40]. Annotation against a protozoan-specific database revealed elevated gene expression by *Trichomonas*, a common protozoan flagellate present in macaques both with and without active ICD [64]. It is unclear whether these are innocent bystanders or causative agents.

Perhaps just as important as the positive associations of potential pathogens with ICD is what can be safely ruled out as causative agents. We did not find evidence of increased gene expression by any of the usual suspects among protozoans for enteric disease in rhesus macaques. These include *Entamoeba histolytica*, *Giardia lamblia*, *Cryptosporidium parvum*, *Cyclospora cayetanensis*, and *Balantidium coli* [41]. It is noteworthy that

we did not see an expansion of the bacterial family *Enterobacteriaceae*, which has been observed in other dysbiotic intestinal environments. This suggests that chronic ICD is not indicative of an environment with a high level of oxygen-related stressors [65].

Ordinarily, the colonic epithelium is protected against pathogens by layers of mucus. Biopsies from humans with ulcerative colitis have revealed dysfunctional mucus that can be penetrated by bacteria [66]. We investigated the hypothesis that ICD is associated with increased mucin degradation. We found relatively more transcripts from mucin-degrading bacteria and mucin-degrading enyzmes in ICD than in controls. It was previously postulated that disruption of the mucus barrier leads to persistent irritation and inflammation of the gut epithelium during ulcerative colitis in humans and ICD in macaques [2, 3, 50]. Further support for this theory comes from the success of helminthic therapy, in which the addition of helminth whipworms stimulates new mucin production, reduces pro-inflammatory markers, and alleviates ICD symptoms [3]. The relative increase in transcripts associated with known mucin-degrading species posits that these microbes may be at least in part responsible for continued degradation of the protective mucin layer, although further investigation is needed to determine the magnitude of their influence and the level of mucosal degradation.

In the current study, both host transcript abundances and the immunohistochemistry of colon sections provided evidence of increased fucosylation of mucins in ICD. Although fucosylation is usually beneficial to the host in support of commensal microbes, it can also be leveraged by pathogens, particularly after insults with antibiotics have removed commensals [67]. Pickard and Chervonsky [67] summarize mechanisms used by pathogens to exploit fucose that falls into two categories: use as a carbon source or use as an adhesion site. In our data, we found evidence of both mechanisms that were used by different pathogens. *Haemophilus influenzae* had increased fucose permease transcripts in ICD. While *Haemophilus* did not express a fucosidase transcript, *Bacteroides* sp. increased transcripts for alpha-L-fucosidases in ICD; these are secreted enzymes that are known to increase the amount of free fucose in the lumen [53], thus making fucose available to species like *H. influenzae*. Evidence for the second mechanism was seen with *Campylobacter*, which did not have increased expression of fucose permease transcripts but did have elevated transcripts for cadF, a protein that enables *Campylobacter* to adhere to fucosylated glycans. It is known that *Campylobacter* chemotaxes toward fucose and is able to bind fucosylated mucins. The overall transcriptional profile of *Campylobacter* in ICD further supports the finding that it adheres to the mucus because its transcriptional

profile is consistent with what would classically be expected of a pathogen in close contact with the slightly oxygenated intestinal epithelial layer: increased oxidative stress and increased secretion systems.

Tropini et al. [68] postulate that the spatial reprogramming of commensals is an important feature of chronic inflammatory bowel disease, which is associated with inappropriate immune activation of mucus-degrading commensals mislocalized to the mucosa. Our data suggest that this is also true of ICD in macaques. We suggest a model in which commensals like *Bacteroides* stimulate the host to produce fucosylated mucins, which then sustain those commensals as well as other cross-feeding species. The fucosylated mucins also serve as adherence sites for microbes like *Campylobacter*, which has the capacity to stimulate inflammation and continue the viscous cycle.

Conclusions

Although the cause of ICD remains unclear, the current study provides several clues. The data confirm several previous observations, such as decreased microbial diversity and increased host inflammation among animals with ICD. More importantly, the data dramatically extend prior observations to include new associations with ICD: increased transcripts associated with *Prevotella*, with known pathogens such as *Campylobacter*, *Helicobacter*, and *Trichomonas*, and with mucin-degrading bacteria. Relatively higher expression of microbial genes in ICD that give rise to mucin-degrading enzymes is suggestive of a dysfunctional microbiome that degrades the mucus barrier instead of catabolizing dietary polysaccharides. During ICD, the host appears to produce fucosylated mucins that are leveraged by microbes as a carbon source or as adhesion sites, enabling the sustained presence of potentially pathogenic species.

Methods

Collection of macaque fecal samples

Fecal samples were collected from a cohort of 12 healthy controls and 12 ICD-diagnosed juvenile macaques. ICD animals were diagnosed clinically being non-responsive to previous antibiotic treatment and having had three consecutive negative culture results for enteric pathogenic bacteria (*Shigella flexneri*, *Yersinia enterocolitica*, *Campylobacter jejuni*, and *Salmonella*) and negative IFA results for parasites (*Cryptosporidium* and *Giardia*) within the 30 days prior to sample collection. In addition to age, ICD and control animals were approximately matched for early-life environment and most recent environment: all animals had spent the first 90 days of life in an outdoor environment and the last 30 days before sample collection in an indoor environment. None of

the animals received antibiotics within 30 days of sample collection.

To collect stool samples, all animals were intermittently pair-housed (ICDs.

and controls separately) indoors. In the afternoon, a clean cage pan (metal tray) was placed under the cage after separating the paired animals. Stool samples were collected from the cage pan of each animal the next morning. When possible, stool was collected within approximately 15 min after the morning meal. Stool samples were collected using disposable sterile transfer pipets into 15-mL conical tubes containing approximately 5 volumes of RNAlater and archived in a − 80 °C freezer after collection.

RNA extraction from macaque fecal samples

Each RNAlater-preserved stool sample was vortexed and homogenized after thawing for 10 min. Approximately 0.6 g of stool slurry was incubated with lysis buffer and periodically vortexed. The sample was homogenized using a bead beater with 0.1-mL beads, followed by QIAshredder treatment. Extraction was performed on the homogenized sample using the Qiagen RNeasy RNA Isolation kit, with additional Turbo DNAse treatment to remove lingering DNA contamination. RNA concentration and integrity were verified using NanoDrop and Bioanalyzer traces.

Metatranscriptome sequencing

For each sample, RNA-Seq libraries were prepared using > 2 μg total extracted RNA at the DNA Sequencing Core of the UC Davis Genome Center. All extracted RNA samples were first ribodepleted using the RiboZero Magnetic Gold Kit (Epidemiology version), catalog number MRZE706. The Illumina TruSeq protocol, without poly(A) selection, was used to prepare the RNA-Seq libraries. The 24 samples were run on 6 lanes of Illumina HiSeq 3000 at 100 bp, paired-end, with indexing to allocate ~ 25% of the lane to each sample metatranscriptome. This approach yielded approximately 95 million raw reads per sample.

Annotation and analysis of metatranscriptome reads

The raw metatranscriptome files were preprocessed, annotated, and analyzed using a modified version of the SAMSA2 pipeline [21, 69]. Preprocessing involved removal of low-quality sequences and adaptor contamination with Trimmomatic [70] and paired-end alignment with PEAR [71], removing approximately 40% of raw reads. Cleaned sequences were annotated by DIAMOND [72] with an *e* value cutoff of 0.001; reads were mapped against NCBI's most recent RefSeq non-redundant protein release [73] and against SEED Subsystems [34]; approximately 45% of cleaned reads met quality score cutoffs to receive an annotation. Specific versions of the SEED and NCBI databases used in this project are available via a bioshare link on GitHub (https://github.com/transcript/macaqueICD, see the "Databases Used" section of the README). On average, roughly 25 million transcripts per sample received an annotation against the RefSeq NR Bacteria database. The minimum number of per-sample annotations was 11 million; the maximum was 36.5 million. For downstream analysis, the resulting annotation files were aggregated and merged using custom Python and R scripting [21]. Statistical computations were performed using the DESeq2 package for R [74]. Individual organism and functional identifications were drawn from NCBI's RefSeq database and the SEED [34, 73]. Subsequent subsetting of data to profile specific organisms in regard to their transcript abundances was performed using Python scripting included with the SAMSA2 pipeline [21, 69].

Functional enrichment analyses

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 was used to conduct gene set enrichment analyses [75]. For each differentially expressed gene set, statistically significant genes were sorted by fold change, and only those with log2 fold change ≥ 2 and adjusted p ≤ 0.05 were used for enrichment analyses. For each analysis, *Macaca mulatta* was selected as the background gene list and the Benjamini and Hochberg adjusted p value ≤ 0.05 was taken to be significant. Upstream regulator analysis was conducted on the same differentially expressed gene sets using Ingenuity Pathways Analysis [76].

Identification and annotation of host organism reads

A DIAMOND custom database was created using all rhesus macaque protein sequences, accessed through NCBI's RefSeq repository (release 102, created 22 December 2015). Metatranscriptome raw reads were annotated against this host sequence database, with an e value cutoff of 0.001, to identify macaque RNA transcripts. There were between 1.5 million and 2 million host-associated annotations per individual macaque.

Statistical analyses

All statistical analyses of the metatranscriptomics data were conducted using scripts adapted from SAMSA2 [21, 69]. R scripts used in the current project are available at https://github.com/transcript/macaqueICD/tree/master/R_scripts. Specifically, PCA and clustering analyses were conducted using make_DESeq_PCA.R and make_DESeq_heatmap.R, respectively. Differential abundance testing of transcripts was conducted using DESeq2 [74]. By default, DESeq2 computes a Benjamini and Hochberg adjusted p value (*padj*) to correct for multiple hypothesis testing. We used *padj* ≤ 0.05 as statistically significant.

Histology staining and analysis

Proximal colon tissues were collected at the time of necropsy and immediately submerged in 10% neutral buffered formalin after collection for 72 h or until processed. To acquire formalin-fixed paraffin-embedded (FFPE) tissues, formalin-fixed samples were submerged in a new jar of formalin at 37 °C for 1 h, then incubated in a series of 70% ethanol (EtOH) at 37 °C for 30 min three times, 95% EtOH at 37 °C for 1 h twice, 100% EtOH at 37 °C for 1 h three times, toluene at 37 °C for 1 h twice, and paraffin at 62 °C for 1 h twice, and embedded during the last stage in paraffin at 62 °C for 1 h. Tissue blocks were cut by Anatomic Pathology Service at UC Davis Veterinary Medicine. Five-micron-thick FFPE proximal colon sections were de-paraffinized and rehydrated with xylene and a graded series of alcohol. Sections were blocked with 10% BSA in PBS for 1 h, washed with PBS, and then incubated with 10 μg/mL FITC-labeled lectin (*Ulex europaeus* agglutinin 1 and concanavalin A for fucose and mannose, respectively; Vector Laboratories, Burlingame, CA) in PBS for 1 h in a dark, humidified container at room temperature. Slides were washed with PBS, mounted with ProLong Diamond Antifade Mountant with DAPI (ThermoFisher Scientific) and a 170-μm-thick glass coverslip, and cured for 24 h at room temperature. Slides were imaged at UC Davis Health Sciences District Advanced Imaging Facility with a Leica TCS SP8 STED 3X confocal microscope at the same FITC channel voltage setting. Five micrographs were selected and analyzed from each animal while blinded to animals' status as ICD or control. Average FITC channel pixel intensity was determined from a polygon selection of single crypt and a blank area from the same image with ImageJ 1.47v software to calculate a background-corrected crypt mean pixel intensity.

Additional files

Additional file 1: Macaque_bacteria_genus_DESeq_results. This tab-separated file contains the DESeq2 output of differential bacterial gene expression at the genus level between ICD and control samples, with adjusted *p* values. The controlMean refers to mean counts of the control samples. ExperimentalMean refers to mean counts of the ICD samples. Log2 fold changes refer to ICD relative to control. The *p* values are adjusted for multiple hypothesis testing. (TSV 149 kb)

Additional file 2: All_bacteria_Subsystems_DESeq_results. This Excel spreadsheet contains the differential expression by SEED Subsystem hierarchies between ICD and control samples for community-wide metatranscriptomes. The first tab contains a description of all hierarchies included in the analyses. The remaining tabs contain results corresponding to the level 1, level 2, and level 3 hierarchies. The controlMean refers to mean counts of the control samples. ExperimentalMean refers to mean counts of the ICD samples. Log2 fold changes refer to ICD relative to control. The *p* values are adjusted for multiple hypothesis testing. (XLSX 444 kb)

Additional file 3: Prevotella_Subsystems_DESeq_results. This Excel spreadsheet contains the differential expression by SEED Subsystem hierarchies between ICD and control samples for *Prevotella*-specific transcriptome (only those reads mapping to *Prevotella*). The first tab contains a description of all hierarchies included in the analyses. The remaining tabs contain results corresponding to the level 1, level 2, and level 3 hierarchies. The controlMean refers to mean counts of the control samples. ExperimentalMean refers to mean counts of the ICD samples. Log2 fold changes refer to ICD relative to control. The *p* values are adjusted for multiple hypothesis testing. (XLSX 365 kb)

Additional file 4: Campylobacter_Subsystems_DESeq_results. This Excel spreadsheet contains the differential expression by SEED Subsystems hierarchies between ICD and control samples for *Campylobacter*-specific transcriptome (only those reads mapping to *Campylobacter*). The first tab contains a description of all hierarchies included in the analyses. The remaining tabs contain results corresponding to the level 1, level 2, and level 3 hierarchies. The controlMean refers to mean counts of the control samples. ExperimentalMean refers to mean counts of the ICD samples. Log2 fold changes refer to ICD relative to control. The *p* values are adjusted for multiple hypothesis testing. (XLSX 409 kb)

Availability of data and materials

The raw sequencing data has been deposited in the NCBI SRA repository, accession numbers SRX3517701-SRX3517724, and analyzed data are available in NCBI GEO, accession number GSE108572. The version of SAMSA used in this paper was version 2.0.0 [21, 69]. All components and tools used in the SAMSA2 pipeline, as well as documentation files, are freely available from GitHub at http://github.com/transcript/SAMSA2. All workflows used to generate the data are available through Github: http://github.com/transcript/macaqueICD.

Authors' contributions

DGL and AA conceived and designed the experiments. AA contributed to the sample collection. STW contributed to the RNA extraction and protocol optimization. ZA contributed to the immunohistochemistry. STW and DGL contributed to the data analysis and bioinformatics work. STW, MEK, AA, IK, and DGL contributed to the data interpretation. All authors contributed in the writing of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All care and use of the animals was approved prior to experiments by the Institutional Animal Care and Use Committee at UC Davis, and were consistent with the requirements of the Animal Welfare Act and with the recommendations of the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. All procedures related to animal care and use were conducted as per standard operating procedures at the California National Primate Research Center (CNPRC). The CNPRC houses over 4000 nonhuman primates and is accredited by the Association for the Assessment and Accreditation of Laboratory Animal Care, International (AAALAC).

Consent for publication

Not applicable.

Author details
[1]Genome Center, University of California, Davis, California, USA. [2]California National Primate Research Center, University of California, Davis, California, USA. [3]USDA ARS Western Human Nutrition Research Center, Davis, California, USA. [4]Department of Nutrition, University of California, Davis, California, USA.

References
1. Vallender EJ, Miller GM. Nonhuman primate models in the genomic era: a paradigm shift. ILAR J. 2013;54(2):154–65.
2. Ardeshir A, Oslund KL, Ventimiglia F, Yee J, Lerche NW, Hyde DM. Idiopathic microscopic colitis of rhesus macaques: quantitative assessment of colonic mucosa. Anat Rec (Hoboken). 2013;296(8):1169–79.
3. Broadhurst MJ, Ardeshir A, Kanwar B, Mirpuri J, Gundra UM, Leung JM, Wiens KE, Vujkovic-Cvijin I, Kim CC, Yarovinsky F, et al. Therapeutic helminth infection of macaques with idiopathic chronic diarrhea alters the inflammatory signature and mucosal microbiota of the colon. PLoS Pathog. 2012;8(11):e1003000.
4. Fox JG, Handt L, Xu S, Shen Z, DeWhirst FE, Paster BJ, Dangler CA, Lodge K, Motzel S, Klein H. Novel Helicobacter species isolated from rhesus monkeys with chronic idiopathic colitis. J Med Microbiol. 2001;50:8.
5. Farkas T, Falkenstein KP, Bohm RP, Pecotte J, Sestak K. High incidence of rhesus enteric calicivirus infections and diarrhea in captive juvenile macaques: a likely association. J Med Primatol. 2012;41(5):325–8.
6. Blackwood RS, Tarara RP, Christe KL, Spinner A, Lerche NW. Effects of the macrolide drug tylosin on chronic diarrhea in rhesus macaques (Macaca mulatta). Comparative medicine. 2008;58(1):81–7.
7. McKenna P, Hoffmann C, Minkah N, Aye PP, Lackner A, Liu Z, Lozupone CA, Hamady M, Knight R, Bushman FD. The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. PLoS Pathog. 2008;4(2):e20.
8. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, Pudlo NA, Kitamoto S, Terrapon N, Muller A, et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. Cell. 2016;167(5):1339–53 e1321.
9. Chen T, Long W, Zhang C, Liu S, Zhao L, Hamaker BR. Fiber-utilizing capacity varies in Prevotella- versus Bacteroides-dominated gut microbiota. Sci Rep. 2017;7(1):2594.
10. Sonnenburg ED, Sonnenburg JL. Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. Cell Metab. 2014;20(5):779–86.
11. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, Mujagic Z, Vila AV, Falony G, Vieira-Silva S, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science. 2016;352(6285):565–9.
12. Sonnenburg ED, Smits SA, Tikhonov M, Higginbottom SK, Wingreen NS, Sonnenburg JL. Diet-induced extinctions in the gut microbiota compound over generations. Nature. 2016;529(7585):212–5.
13. Parthasarathy G, Chen J, Chen X, Chia N, O'Connor HM, Wolf PG, Gaskins HR, Bharucha AE. Relationship between microbiota of the colonic mucosa vs feces and symptoms, colonic transit, and methane production in female patients with chronic constipation. Gastroenterology. 2016;150(2):367–79 e361.
14. El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. Nat Rev Microbiol. 2013;11(7):497–504.
15. Engevik MA, Yacyshyn MB, Engevik KA, Wang J, Darien B, Hassett DJ, Yacyshyn BR, Worrell RT. Human Clostridium difficile infection: altered mucus production and composition. Am J Physiol Gastrointest Liver Physiol. 2015;308(6):G510–24.
16. Conway T, Cohen PS. Commensal and pathogenic Escherichia coli metabolism in the gut. Microbiol Spectr. 2015;3(3).
17. Levy M, Kolodziejczyk AA, Thaiss CA, Elinav E. Dysbiosis and the immune system. Nat Rev Immunol. 2017;17(4):219–32.
18. Pereira FC, Berry D. Microbial nutrient niches in the gut. Environ Microbiol. 2017;19(4):1366–78.
19. Marini RP, Muthupalani S, Shen Z, Buckley EM, Alvarado C, Taylor NS, Dewhirst FE, Whary MT, Patterson MM, Fox JG. Persistent infection of rhesus monkeys with 'Helicobacter macacae' and its isolation from an animal with intestinal adenocarcinoma. J Med Microbiol. 2010;59(Pt 8):961–9.
20. Sonnenberg A, Melton SD, Genta RM, Lewis AD. Absence of focally enhanced gastritis in macaques with idiopathic colitis. Inflamm Bowel Dis. 2011;17(12):2456–61.
21. Westreich ST, Korf I, Mills DA, Lemay DG. SAMSA: a comprehensive metatranscriptome analysis pipeline. BMC Bioinformatics. 2016;17(1):399.
22. Loktionov A. Cell exfoliation in the human colon: myth, reality and implications for colorectal cancer screening. Int J Cancer. 2007;120(11):2281–9.
23. Mersich AT, Miller MR, Chkourko H, Blystone SD. The formin FRL1 (FMNL1) is an essential component of macrophage podosomes. Cytoskeleton (Hoboken). 2010;67(9):573–85.
24. Moein S, Qujeq D, Vaghari Tabari M, Kashifard M, Hajian-Tilaki K. Diagnostic accuracy of fecal calprotectin in assessing the severity of inflammatory bowel disease: from laboratory to clinic. Caspian J Intern Med. 2017;8(3):178–82.
25. Sipponen T, Kolho KL. Fecal calprotectin in diagnosis and clinical assessment of inflammatory bowel disease. Scand J Gastroenterol. 2015;50(1):74–80.
26. Kucharzik T, Hudson JT 3rd, Lugering A, Abbas JA, Bettini M, Lake JG, Evans ME, Ziegler TR, Merlin D, Madara JL, et al. Acute induction of human IL-8 production by intestinal epithelium triggers neutrophil infiltration without mucosal injury. Gut. 2005;54(11):1565–72.
27. Grennan AK. Genevestigator. Facilitating web-based gene-expression analysis. Plant Physiol. 2006;141(4):1164–6.
28. Jarry A, Vallette G, Cassagnau E, Moreau A, Bou-Hanna C, Lemarre P, Letessier E, Le Neel J-C, Galmiche J-P, Laboisse CL. Interleukin 1 and interleukin 1β converting enzyme (caspase 1) expression in the human colonic epithelial barrier. Caspase 1 downregulation in colon cancer. Gut. 1999;45:246–51.
29. Ciorba MA. Indoleamine 2,3 dioxygenase in intestinal disease. Curr Opin Gastroenterol. 2013;29(2):146–52.
30. Ferdinande L, Demetter P, Perez-Novo C, Waeytens A, Taildeman J, Rottiers I, Rottiers P, De Vos M, Cuvelier CA. Inflamed intestinal mucosa features a specific epithelial expression pattern of Indoleamine 2,3-dioxygenase. Int J Immunopathol Pharmacol. 2008;21(2):289–95.
31. Gombart AF, Krug U, O'Kelly J, An E, Vegesna V, Koeffler HP. Aberrant expression of neutrophil and macrophage-related genes in a murine model for human neutrophil-specific granule deficiency. J Leukoc Biol. 2005;78(5):1153–65.
32. Peeters PM, Wouters EF, Reynaert NL. Immune homeostasis in epithelial cells: evidence and role of inflammasome signaling reviewed. J Immunol Res. 2015;2015:828264.
33. Kumar A, Yu FS. Toll-like receptors and corneal innate immunity. Curr Mol Med. 2006;6(3):327–37.
34. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 2014;42(Database issue):D206–14.
35. Chatzidaki-Livanis M, Comstock Laurie E. Friend turned foe: a role for bacterial sulfatases in colitis. Cell Host Microbe. 2015;17(5):540–1.
36. Roager HM, Hansen LB, Bahl MI, Frandsen HL, Carvalho V, Gobel RJ, Dalgaard MD, Plichta DR, Sparholt MH, Vestergaard H, et al. Colonic transit time is related to bacterial metabolism and mucosal turnover in the gut. Nat Microbiol. 2016;1(9):16093.
37. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. Nat Rev Microbiol. 2016;14(1):20–32.
38. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. Gut. 2016;65(1):57–62.
39. Bremer H, Dennis PP. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. EcoSal Plus. 2008;3(1).
40. Newcombe G, Campbell J, Griffith D, Baynes M, Launchbaugh K, Pendleton R. Revisiting the life cycle of dung fungi, including Sordaria fimicola. PLoS One. 2016;11(2):e0147425.
41. Li J, Dong H, Wang R, Yu F, Wu Y, Chang Y, Wang C, Qi M, Zhang L. An investigation of parasitic infections and review of molecular characterization

of the intestinal protozoa in nonhuman primates in China from 2009 to 2015. Int J Parasitol Parasites Wildl. 2017;6(1):8–15.

42. Pimentel M, Lin HC, Enayati P, van den Burg B, Lee HR, Chen JH, Park S, Kong Y, Conklin J. Methane, a gas produced by enteric bacteria, slows intestinal transit and augments small intestinal contractile activity. Am J Physiol Gastrointest Liver Physiol. 2006;290(6):G1089–95.

43. Pavlovskis OR, Rollins DM, Haberberger RL, Green AE, Habash L, Strocko S, Walker RI. Significance of flagella in colonization resistance of rabbits immunized with Campylobacter spp. Infect Immun. 1991;59(7):2259–64.

44. Grant CC, Konkel ME, Cieplak W Jr, Tompkins LS. Role of flagella in adherence, internalization, and translocation of Campylobacter jejuni in nonpolarized and polarized epithelial cell cultures. Infect Immun. 1993;61(5):1764–71.

45. Wassenaar TM, Bleumink-Pluym NM, van der Zeijst BA. Inactivation of Campylobacter jejuni flagellin genes by homologous recombination demonstrates that flaA but not flaB is required for invasion. EMBO J. 1991; 10(8):2055–61.

46. Yao R, Burr DH, Doig P, Trust TJ, Niu H, Guerry P. Isolation of motile and non-motile insertional mutants of Campylobacter jejuni: the role of motility in adherence and invasion of eukaryotic cells. Mol Microbiol. 1994;14(5): 883–93.

47. Gaynor EC, Wells DH, MacKichan JK, Falkow S. The Campylobacter jejuni stringent response controls specific stress survival and virulence-associated phenotypes. Mol Microbiol. 2005;56(1):8–27.

48. Bacon DJ, Alm RA, Hu L, Hickey TE, Ewing CP, Batchelor RA, Trust TJ, Guerry P. DNA sequence and mutational analyses of the pVir plasmid of Campylobacter jejuni 81-176. Infect Immun. 2002;70(11):6242–50.

49. Tailford LE, Crost EH, Kavanaugh D, Juge N. Mucin glycan foraging in the human gut microbiome. Front Genet. 2015;6:81.

50. Lepage P, Häsler R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, Ott S, Kupcinskas L, Doré J, Raedler A, et al. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. Gastroenterology. 2011;141(1):227–36.

51. Maloy KJ, Powrie F. Intestinal homeostasis and its breakdown in inflammatory bowel disease. Nature. 2011;474(7351):298–306.

52. Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB. Sequence and expression of a candidate for the human secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J Biol Chem. 1995;270(9):4640–9.

53. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI. A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. Science. 2003;299(5615):2074–6.

54. Hooper LV, Xu J, Falk PG, Midtvedt T, Gordon JI. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. Proc Natl Acad Sci U S A. 1999;96(17):9833–8.

55. Shin NR, Whon TW, Bae JW. Proteobacteria: microbial signature of dysbiosis in gut microbiota. Trends Biotechnol. 2015;33(9):496–503.

56. Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Booijink CC, Troost FJ, Bork P, Wels M, de Vos WM, Kleerebezem M. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. Isme j. 2012;6(7):1415–26.

57. Noguera-Julian M, Rocafort M, Guillen Y, Rivera J, Casadella M, Nowak P, Hildebrand F, Zeller G, Parera M, Bellido R, et al. Gut microbiota linked to sexual preference and HIV infection. EBioMedicine. 2016;5:135–46.

58. Dillon SM, Lee EJ, Kotter CV, Austin GL, Gianella S, Siewe B, Smith DM, Landay AL, McManus MC, Robertson CE, et al. Gut dendritic cell activation links an altered colonic microbiome to mucosal and systemic T-cell activation in untreated HIV-1 infection. Mucosal Immunol. 2016;9(1):24–37.

59. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, Bielski C, Rostron T, Cerundolo V, Pamer EG, Abramson SB, et al. Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. Elife. 2013;2:e01202.

60. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011;334(6052):105–8.

61. Ley RE. Gut microbiota in 2015: Prevotella in the gut: choose carefully. Nat Rev Gastroenterol Hepatol. 2016;13(2):69–70.

62. Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, Desai P, Lebrilla CB, Weimer BC, Mills DA, German JB, et al. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. Cell Host Microbe. 2011;10(5):507–14.

63. Ohkawara S, Furuya H, Nagashima K, Asanuma N, Hino T. Effect of oral administration of Butyrivibrio fibrisolvens MDT-1 on experimental enterocolitis in mice. Clin Vaccine Immunol. 2006;13(11):1231–6.

64. Sestak K, Merritt CK, Borda J, Saylor E, Schwamberger SR, Cogswell F, Didier ES, Didier PJ, Plauche G, Bohm RP, et al. Infectious agent and immune response characteristics of chronic enterocolitis in captive rhesus macaques. Infect Immun. 2003;71(7):4079–86.

65. Rivera-Chavez F, Lopez CA, Baumler AJ. Oxygen as a driver of gut dysbiosis. Free Radic Biol Med. 2017;105:93–101.

66. Johansson ME, Gustafsson JK, Holmen-Larsson J, Jabbar KS, Xia L, Xu H, Ghishan FK, Carvalho FA, Gewirtz AT, Sjovall H, et al. Bacteria penetrate the normally impenetrable inner colon mucus layer in both murine colitis models and patients with ulcerative colitis. Gut. 2014;63(2):281–91.

67. Pickard JM, Chervonsky AV. Intestinal fucose as a mediator of host-microbe symbiosis. J Immunol. 2015;194(12):5588–93.

68. Tropini C, Earle KA, Huang KC, Sonnenburg JL. The gut microbiome: connecting spatial organization to function. Cell Host Microbe. 2017;21(4):433–42.

69. Westreich ST, Treiber ML, Mills DA, Korf I, Lemay DG. SAMSA2: a standalone metatranscriptome analysis pipeline. BMC Bioinformatics. 2018;19(1):175.

70. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.

71. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30(5):614–20.

72. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, et al. The metagenomics RAST server-a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;9:386.

73. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. RefSeq microbial genomes database: new representation and annotation strategy. Nucleic Acids Res. 2014;42(Database issue):D553–9.

74. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

75. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4(1):44–57.

76. Kramer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523–30.