Interactions between the gut microbiome and host gene regulation in cystic fibrosis

Gargi Dayama[1,*], Sambhawa Priya[1,*], David E. Niccum[2], Alexander Khoruts[2,3,&], Ran Blekhman[1,4,&]

[1] Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN

[2] Department of Medicine, University of Minnesota, Minneapolis, MN;

[3] Center for Immunology; BioTechnology Institute

[4] Department of Ecology, Evolution, and Behavior, University of Minnesota, Minneapolis, MN

* these authors contributed equally

[&] to whom correspondence should be addressed: khoru001@umn.edu (AK), blekhman@umn.edu (RB)

Abstract

Cystic Fibrosis (CF) is the most common autosomal recessive genetic disease in Caucasians. It is caused by mutations in the *CFTR* gene, leading to poor hydration of mucus and impairment of the respiratory, digestive, and reproductive organ functions. Advancements in medical care have lead to markedly increased longevity of patients with CF, but new complications have emerged, such as early onset of colorectal cancer (CRC). Although the pathogenesis of CRC in CF remains unclear, altered host-microbe interactions might play a critical role. Here, we characterize the changes in the gut microbiome and host gene expression in colonic mucosa of CF patients relative to healthy controls. We find that CF patients show decreased microbial diversity, decreased abundance of taxa such as *Butyricimonas, Sutterella*, and Ruminococcaceae, and increased abundance of other taxa, such as Actinobacteria and Firmicutes. We find that 1543 genes, including *CFTR*, show differential expression in the colon of CF patients compared to healthy controls. Interestingly, we find that these genes are enriched with functions related to gastrointestinal and colorectal cancer, such as metastasis of CRC, tumor suppression, cellular dysfunction, p53 and mTOR signaling pathways. Lastly, we modeled associations between relative abundances of specific bacterial taxa in the gut mucosa and host gene expression, and identified CRC-related genes, including *LCN2* and *DUOX2*, for which gene expression is correlated with the abundance of CRC-associated bacteria, such as Ruminococcaceae and *Veillonella*. Our results provide new insight into the role of host-microbe interactions in the etiology of CRC in CF.

Keywords: Cystic fibrosis, host-microbe interactions, gene regulation, microbiome, colorectal cancer.

Introduction

Cystic fibrosis (CF) is the most common autosomal recessive genetic disease in Caucasians, where it occurs with a frequency of 1 in 3,000 births, although it is also present at lower rates in populations of non-European descent [1]. CF is caused by mutations in the cystic fibrosis transmembrane conductor regulatory (*CFTR*) gene, which plays critical functions in epithelial ion transport and hydration of mucus. Absent or reduced *CFTR* activity results in thick, viscous secretions that impair functions of the respiratory, digestive, and reproductive organ systems.

Multiple advances in medical care in CF, once a fatal pediatric disease, have led to remarkable gains in patient life expectancy. However, increased longevity of CF patients into adulthood has led to new challenges, such as gastrointestinal cancer. The average onset of colorectal cancer (CRC) in CF patients is approximately 20-30 years earlier than in the general population [2,3]. Systematic data on colonoscopic screening and surveillance suggest that CF-associated CRC arises via the classical adenoma to cancer sequence, but adenomatous polyps develop at a younger age in CF and progress faster to more advanced neoplasms [4]. In fact, loss of *CFTR* expression in tumors of non-CF patients has been associated with a worse prognosis in early stage CRC [5]. Recently, specific recommendations for CRC screening were introduced in standard care of adult CF patients, which include earlier initiation of screening and shorter intervals for surveillance [6].

Although previous studies have identified *CFTR* as a tumor suppressor gene that may play a role in early onset of colon cancer [5,7], the pathogenesis of CRC in CF remains unclear. A number of factors can be considered. Thus, stagnant mucus in CF is associated with bacterial overgrowth at the mucosal surface [8,9], which might result in greater levels of tonic microbial stimulation of the epithelia and account for their increased rate of their turnover [10]. It is likely that the altered microbiota composition and microbiota-mucosal interface are also the reasons for a chronic state of low-grade mucosal inflammation in CF [11,12]. Notably, in the colon *CFTR* is hyper-expressed in the stem cell compartment of the intestinal crypt [13,14], which is the site of CRC origination [15].

Than and colleagues have shown altered expression of genes involved in immune cell homeostasis and inflammation, mucins, cell signaling and growth regulation, detoxification and stress response, lipid metabolism, and stem cell regulation in the intestines of *CFTR* mutant mice [5]. The intestinal microbiota of these animals is also distinguished by lower bacterial community richness, evenness, and diversity, consistent with a major impact of *CFTR* deficiency on gastrointestinal physiology [16]. Altered fecal microbiome has also been demonstrated in a number of clinical CF cohorts, where it was characterized by decreased microbial diversity, lower temporal microbial community stability, and decreased relative abundances of taxa associated with health, such as *Faecalibacterium*, *Roseburia*, *Bifidobacterium, Akkermansia, Clostridium cluster XIVa* [17–23]. Greater degrees of dysbiosis were noted to correlate with severity of CF disease phenotype, burden of antibiotics, and evidence for intestinal inflammation. Notably, most of these patient studies have been in diverse pediatric cohorts with varying degrees of fat malabsorption and extent of recent exposure to broad-spectrum antibiotics.

Here, we compare the mucosal microbiome (via 16S rRNA sequencing) and colonic gene expression (via RNA-seq) in adult patients with CF and healthy controls undergoing CRC screening by colonoscopy. We explore interactions between the gut microbiome and host gene regulation by integrative analysis, characterizing genes and microbes that may play a joint role in the development of CRC in CF patients.

Methods

Patients and mucosal biopsy samples

Mucosal biopsies were obtained from patients undergoing CRC screening and surveillance colonoscopies at the University of Minnesota. The majority of CF patients receiving care at the Minnesota Cystic Fibrosis Center participate in a systematic colonoscopic CRC screening program as described previously [4]. Control samples were obtained from non-CF patients undergoing routine colonoscopic CRC screening or surveillance. Pinch biopsies, four per patient, were obtained using the Radial Jaw 4 Jumbo w/Needle 240 (length) forceps for 3.2 mm working channel (Boston Scientific, Marlborough, MA; Catalog # M00513371) in the right colon and placed into RNAlater stabilization solution (ThermoFisher Scientific, Waltham, MA). The protocol was approved by the University of Minnesota Institutional Review Board (IRB protocol 1408M52889). Gene expression was analyzed by RNA-Seq from a total of 33 samples obtained from 18 CF patients and 15 non-CF control participants (Fig S1).

RNA extraction and sequencing

Biopsy tissue was kept in the RNAlater stabilization solution overnight at 4°C. RNA was prepared following tissue homogenization and lysis using the TRIzol Plus RNA Purification Kit (ThermoFisher Scientific; catalogue # 2183-555) following detailed manufacturer instructions. Total RNA samples were converted to Illumina sequencing libraries using Illumina's Truseq Stranded mRNA Sample Preparation Kit (Cat. # RS-122-2103). Total RNA is oligo-dT purified using oligo-dT coated magnetic beads, fragmented and then reverse transcribed into cDNA. The cDNA is adenylated and then ligated to dual-indexed (barcoded) adaptors and amplified using 15 cycles of PCR. Final library size distribution is validated using capillary electrophoresis and quantified using fluorimetry (PicoGreen). Indexed libraries are then normalized, pooled and then size selected to 320bp +/- 5% using Caliper's XT instrument. Truseq libraries are hybridized to a paired end flow cell and individual fragments are clonally amplified by bridge amplification on the Illumina cBot. Once clustering is complete, the flow cell is loaded on the HiSeq 2500 and sequenced using Illumina's SBS chemistry (Fig S1).

Host RNA-seq quality control, read mapping and filtering

We performed quality check on raw sequences from all 33 samples to assure better downstream analysis using FastQC [24]. This helped assess any biases due to parameters such as quality of the reads, GC content, number of reads, read length and species to which the majority of the reads mapped (Fig S2). The FASTQ files for forward and reverse (R1 and R2) reads were mapped to the reference genome using kallisto [25], where an index for the transcriptomes was generated to quantify estimated read counts and TPM values. Mean distribution for the TPM values was plotted using R to filter all the transcripts below a threshold value of log2[TPM] < 0. We generated PCA plots using sleuth [26] to examine sample clusters and visualization of expression patterns for genes using bar plots (Fig S3 and Fig S4). For further analysis of outlier samples box plots were generated using Cook's distance and heat map clustered by condition and mutation status was generated for the top 20 expressed genes (Fig S5 and Fig S6).

Host RNA-seq differential expression and enrichment analysis

To determine differentially expressed genes between CF and healthy samples we quantified and annotated the transcripts used DESeq2 [27]. The output from kallisto was imported into DESeq2 using the tximport package [28]. The transcripts were annotated against the ensemble database using bioMART to obtain gene symbols [29]. Transcripts below a threshold of row-sum of 1 were filtered and collapsed at gene symbol level. Prior to differentially expressed gene analysis, the read counts were normalized and the gene-wise estimates were shrunken towards the fitted estimates represented by the red line in the dispersion plot (Fig S7). The gene-wise estimates that are outliers are not shrunk and are flagged by the blue circles in the plot (Fig S7). DESeq2 applies the Wald test on estimated counts and uses a negative binomial generalized linear model determines differentially expressed genes and the log-fold changes (Fig S8). The log-fold change shrinkage (*lcfshrink()*) function was applied for ranking the genes and data visualization. For data smoothing, MA plots were generated before and after log2 fold shrinkage. We found no change in the MA plot (Fig S9) post smoothing, as there are no large log-fold changes in the current data (log2 fold change between -1 and 1) due to low counts. The data were further transformed and the normalized values were extracted using regularized logarithm (rlog) to remove the dependence of variance on mean. We used the Benjamini-Hochberg method for reducing the false discovery rate (FDR) with a cutoff of 0.05 for identifying differentially expressed genes for further analysis. Enrichment analysis was done using Ingenuity Pathway Analysis (IPA, QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis). The log-fold changes, p-values, and FDR values (for all the genes with FDR < 0.05) were fed into IPA for both up- and down-regulated differentially expressed genes between CF and healthy samples. Disease/functional pathways and gene networks were determined based on the gene enrichment. Furthermore, we looked at how many target upstream regulators were enriched based on our list of differentially expressed genes using IPA. We found 134 targets that passed the filter (p-value < 0.01) from a total of 492 targets, of which 96 were transcription regulators.

16S rRNA extraction and sequencing

Mucosal biopsies samples (~ 3 x 3 mm) from 13 CF and 12 healthy individuals were collected in 1 mL of RNAlater and stored for 24 hours at 4°C prior to freezing at -80°C. DNA was extracted using a MoBio PowerSoil DNA isolation kit according to the manufacturer instructions (QIAGEN, Carlsbad, USA). To look at the tissue associated microbiome, V5-V6 region of 16S rRNA gene was amplified as described by Huse et al. [30] using the following indexing primers (V5F_Nextera: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGRGG ATTAGATACCC, V6R_Nextera: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGACRRCCATGCANCACCT). Index and flowcell adaptors were added with this step. Forward indexing primer used is - **AATGATACGGCGACCACCGA**GATCTACAC[i5]TCGTCGGCAGCGTC and reverse indexing primers used is - **CAAGCAGAAGACGGCATACGA**GAT[i7]GTCTCGTGGGCTCGG. Post two rounds of PCR, pooled, size-selected samples were denatured with NaOH, diluted to 8 pM in Illumina's HT1 buffer, spiked with 15% PhiX, and heat denatured at 96°C for 2 minutes immediately prior to loading. A MiSeq 600 cycle v3 kit was used to sequence the sample.

Gut mucosal microbiome data processing, quality assessment, and diversity analysis

We processed the FASTQ files using FastQC [24] to perform quality control on the raw sequences. We then used SHI7 [31] for trimming Nextera adaptors, stitching paired-end reads and performing quality trimming at both ends of the stitched reads until a minimum Phred score of 32 was reached. Following quality control, we obtained an average of 217,500 high quality reads per sample (median 244,000; range 9551 - 373,900) with an average length of 281.9 bases and an average quality score of 37.19. These merged and filtered reads were used for closed reference OTU picking and taxonomy assignment against GreenGenes database with 97% similarity level using the NINJA-OPS program [32].

We performed alpha and beta-diversity analysis in R using the vegan [33] and phyloseq [34] packages. We used resampling-based computation of alpha-diversity, where the OTU table is subsampled 100 times at minimum read depth (9551 reads) across all samples, and computed average richness estimate for each alpha-diversity metric (chao1, observed-OTUs, and Shannon). Wilcoxon rank-sum test was used for testing the statistical significance of the associations between alpha-diversity of the CF and healthy conditions. For computing beta-diversity, we first rarefied the OTU table (using vegan's *rrarefy()* function) at minimum sequence depth (i.e. 9551 reads) across the samples and then computed Bray-Curtis dissimilarity, weighted UniFrac, and unweighted UniFrac metrics. The Adonis test was used for assessing if there is significant association between the beta-diversity of the CF/healthy condition and the diversity results are plotted using the ggplot2 package in R.

Gut mucosal microbiome differential abundance and functional analysis

We performed differential abundance testing using the phyloseq [34] package in R. We first created a phyloseq object from the OTU table (using the *phyloseq()* function) and filtered this object to only include OTUs with at least 0.1% relative abundance occurring in at least half of all samples (using the *filter_taxa()* function). The filtered phyloseq object was converted into a DESeqDataSet object (using *phyloseq_to_deseq2()*) and the *DESeq()* function was invoked. This performed dispersion estimations and Wald's test for identifying differentially abundant OTUs, with their corresponding log-fold change, p-value, and FDR-adjusted q-values between the CF and healthy conditions. We agglomerated the OTUs at different taxonomic ranks (using the *tax_glom()* function) and repeated the above steps to identify differentially abundant taxa at genus, family, order, class, and phylum levels.

We also tested for associations between taxonomic abundance and mutation status of CF samples. We first categorized samples into three genotype categories: (1) Healthy: Samples with no mutations; (2) CF_df508: CF samples with homozygous delta-F508 deletion, which is associated with more severe CF condition [35]; and (3) CF_other: CF samples with df508 heterozygous deletion or other mutation status. We used DESeq2's likelihood ratio test (LRT) to identify taxa that showed significant difference in abundance across the three categories.

We then generated the predicted functional profiles for the gut microbes using PICRUSt v1.0.0 pipeline, [36] where pathways and enzymes are assigned using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The KEGG level 3 pathways were filtered for rare pathways by only including pathways with relative abundance > 0.1% in at least half of the samples, normalized to relative abundance, and tested for association with CF/Healthy conditions using non-parametric Wilcoxon rank-sum test followed by FDR adjustment.

Integrated analysis of interactions between host gene dysregulations and changes in microbiome

For this analysis, differentially expressed genes from host and gut microbial OTUs from their respective overlapping samples were used (22 samples in total, with 12 healthy samples and 10 CF samples). We further subset differentially expressed genes between CF and healthy conditions (FDR < 0.05), specifically, enriched for gastrointestinal cancer disease pathways (524 genes). Using absolute expression log ratio greater than 0.35, we obtained a representative set of both up- and down-regulated genes from these pathways, leaving 250 genes for downstream analysis. The OTU table was collapsed at the genus level (or the last characterized level) and filtered for rare taxa by only including taxa with at least 0.1% relative abundance present in at least half of all samples, resulting in 35 taxa for further processing. Following this, centered log ratio transform was applied on the filtered table. We then performed correlation analysis between host gene expression data for 250 genes and gut microbiome abundance data for 35 taxa (genus level) defined above. Spearman correlation was used for this analysis as it performs better with normalized counts (gene expression) as well as compositional data (microbiome relative abundance) compared to other metrics, such as Pearson correlation (Weiss et al. 2016). We computed the

Spearman rank correlation coefficients and the corresponding p-values using the *cor.test()* function with two-sided alternative hypothesis. A total of 8750 (250 genes x 35 taxa) statistical tests were performed, and p-values were corrected for multiple comparisons using the qvalue package in R [37]. Representative gene-taxa correlations were visualized using corrplots [38] in R, where the strength of the correlation is indicated by the color and size of the visualization element (square) and the significance of the correlation is indicated via asterisk. We also computed the Sparse Correlation for Compositional Data (SparCC) [39] for the taxa found significantly correlated (q-value < 0.1) with the CRC genes. Pseudo p-values were computed using 100 randomized sets. Significant gene-microbe correlations (q-value < 0.1) and significant microbe-microbe correlations (SparCC |R| >=0.1 and p-value <0.05) were visualized as a network using Cytoscape v3.5.1 [40].

Results

Host RNA-Seq sample preprocessing and quality assessment

We first examined gene expression in colonic biopsies from 18 CF and 15 healthy individuals. Overall, CF and healthy samples had comparable number of reads (28,250,473 and 30,041,827 reads on average, respectively) with the average quality greater than 30 phred score across all samples (Fig S2). The sequences were annotated to generate estimated read counts and transcripts per kilobase million (TPM) using kallisto, [25] resulting in 173,259 total transcripts, of which 56,283 passed the filter of mean TPM greater than 1 (TPM>1). While the Principal component analysis (PCA) plots showed an overlap between the expression profile of most samples from CF and healthy individuals, it identified two possible outliers (samples 1096 and 1117) (Fig S3). In addition, the top five transcripts driving the PC were of mitochondrial origin (Fig S4). Hence, to reduce any bias in identifying differentially expressed genes, we filtered out all the mitochondrial transcripts from the data. We further investigated the outliers using the remaining transcripts by calculating cook's distance between the samples, and found that the two samples (1096 and 1117) were still outliers (Fig S5). This was further evident by the heatmap of the top 20 most highly expressed genes (Fig S6), where we found an alternate expression pattern for the two samples, compared to the rest. Therefore, the two outlier CF samples (1096 and 1117) were eliminated from further analysis.

Differentially expressed host genes between CF and healthy mucosal samples

To examine gene expression differences we used read counts from the remaining 16 CF and 15 healthy samples. Using DESeq2 we identified 1543 differentially expressed genes at q-value < 0.05 (Benjamini-Hochberg correction; see Fig S8 for a volcano plot). Of the 1543 differentially expressed genes, 919 (59%) were up-regulated and 624 (41%) were down-regulated in CF patients. Including sex as a covariate in the model did not susbstantially alter the results (only 43 additional differentially expressed genes were identified); therefore, we did not include sex in downstrem analyses. The full list of differentially expressed genes significant at q-value < 0.05 is available in Additional File 2.

We visualized the expression pattern of six representative genes, selected from genes included in the colorectal cancer disease pathway (**Figure 1A**). Consistent with the expectation of changes in mucosal immunity that could compensate for a diminished protective mucus function, we noted *LCN2* to be one of the top differentially expressed genes (q-value = 2.54E-08, Wald test). *LCN2* encodes for lipocalin 2, which limits bacterial growth by sequestering iron-laden bacterial siderophore [41]. However, a number of other top genes are involved in major cellular biology processes, and were previously related to cancer pathogenesis and colon cancer. Examples include *RRS1* (q-value = 6.16E-09), which encodes for the ribosomal biogenesis protein homolog that promotes angiogenesis and cellular proliferation, but suppresses apoptosis [42]; *KRTAP5-5* (q-value = 4.89E-08), which encodes for keratin-associated protein 5-5, a protein that plays important roles in cytoskeletal function and facilitates various malignant behaviors that include cellular motility and vascular invasion [43]; *ALDOB* (q-value = 2.64E-07), which encodes for aldolase B, an enzyme that promotes metastatic cancer-associated metabolic reprogramming [44]. Additional examples of differentially expressed genes (Log-fold change > 0.5 and q-value < 0.05), such as *CDH3, TP53INP2, E2F1, CCND2, and SERPINE1*, were also previously shown to have direct roles in colorectal and digestive cancers [45–47]. While, some of these genes participate in basic cancer-related cellular functions such as proliferation and invasion [45–52], others, e.g. *BEST2*, play important roles in gut barrier function and anion transport [53,54]. In addition to the genes visualized in Figure 1A, additional randomly selected differentially expressed genes are visualized in Additional File 1 (Fig S10), showing expression pattern difference between the CF and healthy samples.

We next performed an enrichment analysis to categorize functional and disease pathways among differentially expressed genes (q-value <0.05) in IPA. The top canonical pathways (Fig S11) are mostly responsible for signaling and regulatory functions, such as *EIF2* signaling (p-value = 3.32E-35), mTOR signaling (p-value = 3.83E-08) and regulation of chromosomal replication (p-value = 1.60E-06). Of the 39 significantly enriched disease and functional pathways (p-value < 1.00E-05; **Figure 1B**), 14 are related to cancer, including gastrointestinal cancer (p-value = 2.61E-06), abdominal cancer (p-value = 9.23E-03), large intestine cancer (p-value = 7.00E-05), and colorectal cancer (p-value = 8.63E-03). In addition, using the list of differentially expressed genes we found that the promoter sequences are enriched with binding sites of 96 potential transcription regulators (p-value < 0.01; see Methods). Among these transcription factors, many have been previously shown to control cancer related pathways. For example, *MYCN* and *KRAS* are prominently involved in neuroblastoma and colorectal cancer, respectively [55,56]. *NHF4A* is involved in transcriptional regulation of many aspects of epithelial cell morphogenesis and function, which has been linked to colorectal cancer [57]. *CST5*, which encodes cytostatin D, is a direct target of p53 and vitamin D receptor, promotes mesenchymal-epithelial transition to suppress tumor progression and metastasis [58]. *E2F3* is a potent regulator of the cell cycle and apoptosis that is commonly deregulated in oncogenesis [59,60].

A metabolic network for the gastrointestinal (GI) cancer-related differentially expressed genes is shown in **Figure 1C,** illustrating the interactions between genes that are up-regulated in CF (Eg. *TP53INP1, SERPINE1, NCOR1* and *CAPN2*) and down-regulated in CF (*E2F1, MED1, ECND2 and AS3MT*), highlighting the cellular location of these genes' product. Additional gene network for colorectal cancer can be found in Additional file 1 (Fig S12), where the genes are also positioned in the region of the cell where they are most active. We found that genes such as *BEST2* (involved in ion transport) and *RUVBL1* (involved in cell cycle, cell division, and cell damage) are down-regulated, while genes such as *TP53INP2* (involved in transcription regulation) and *CDH3* (involved in sensory transduction) are up-regulated. Given the predicted role of gene regulation in colorectal cancer and the dysregulation of CRC-related pathways, these results may help understand mechanisms controlling early onset of colon cancer in cystic fibrosis.

Difference in microbiome composition between CF and healthy gut mucosa

To further understand the potential of altered microbiota-host interaction in the CF colon, we next investigated differences in the composition of the mucosal microbiome between CF and healthy individuals. We found a significant difference between beta-diversity of gut mucosal microbiome in CF patients compared to healthy individuals with respect to unweighted UniFrac and non-phylogenetic Bray-Curtis metrics (Adonis p-value = 0.001). As observed in the PCoA plot (**Figure 2A**) the samples were clustered based on their disease condition (CF or healthy). The overall biodiversity of mucosal microbiome was depleted in CF compared to healthy samples, which is depicted by significant decrease in alpha diversity measured by Chao1 (p-value = 0.015, Wilcoxon rank sum test, **Figure 2A**) and Observed OTUs (p-value = 0.024, Wilcoxon rank sum test, in Additional File 1 (Fig S13)) metrics in CF relative to healthy controls.

We assessed the changes in abundance of microbes at various taxonomic levels between CF and healthy gut mucosal microbiome using phyloseq. We found 51 OTUs that were significantly differentially abundant between CF and healthy individuals (q-value < 0.1, Additional file 3). At different taxonomic ranks, we found 7 genera, 10 families, 4 orders, 4 classes, and 5 phyla differentially abundant between CF and healthy samples (q-value < 0.1 by Wald's test; Additional file 3). Overall, an increased abundance in taxa, predominantly belonging to Firmicutes and Fusobacteria, was observed in CF individuals compared to healthy controls, while taxa belonging to Bacteroidetes, Verrucomicrobia, and Proteobacteria phyla showed a marked decrease in patients with CF relative to healthy controls (**Figure 2B**). In particular, there was an increase in abundance of class Actinobacteria in individuals with CF compared to healthy controls (q-value = 0.079), while *Butyricimonas* (q-value = 0.009), Ruminococcaceae (q-value = 0.081), *Sutterella* (q-value=0.040) were found depleted in CF samples (**Figure 2C**). Additional examples of differentially abundant taxa between CF and healthy samples can be found in the Additional file 1 (Fig S14).

Next, we tested whether *CFTR* genotype, which affects disease severity, is associated with variation in the microbiome. Specifically, we hypothesized that variation in the microbiome is correlated with the number of alleles of the DF508 mutation, a deletion of an entire codon within *CFTR* that is the most common cause for CF. To test this, we performed likelihood ratio test to identify differentially abundant taxa between three genotype classes: CF-DF508 (homozygous for the DF508 mutation), CF-other (either one or zero copies of the DF508 mutation), and healthy (no known mutations in *CFTR*). We found a gradient-like trend in abundance for Actinobacteria (q-value = 0.081), showing increase in abundance with increasing severity of mutation status (**Figure 2D**).

To assess the potential functional changes in the microbiome, we predicted abundance of metabolic pathways and enzymes using the PICRUSt pipeline [36] and KEGG database, and compared them for differences between CF and healthy individuals. Seven predicted pathways (as defined by KEGG level 3) were found to be differentially abundant between CF and healthy: bacterial toxins were enriched in CF compared to healthy, while propanoate metabolism, restriction enzyme, pantothenate and CoA biosynthesis, thiamine metabolism, amino acid related enzymes, and aminoacyl-tRNA biosynthesis were depleted in CF compared to healthy (q-value < 0.2 using Wilcoxon rank sum test; in Additional File 1 (Fig S15)).

Interactions between gastrointestinal cancer-related host genes and gut microbes

In order to investigate the relationship between host genes and microbes in the colonic mucosa and their potential role in the pathogenesis of gastrointestinal cancers in CF patients, we considered correlations between 250 differentially expressed genes enriched for GI cancers and 35 microbial taxa (collapsed at genus or last characterized level, and filtered at 0.1% relative abundance, see Methods). Using Spearman correlations, we found 50 significant unique gene-microbe correlations in the gut (q-value < 0.1), where the magnitude of correlation (Spearman rho) ranged between (-0.77, 0.79) (Additional file 4). Interestingly, most of the taxa that significantly correlated with the genes also differed significantly in abundance between CF and healthy individuals. We visualized all the correlations between taxa abundance and host gene expression in **Figure 3A**. In particular, we found some significant positive gene-taxa correlations (q-value < 0.05), between *Butyricimonas* and *ZNHIT6* (Spearman rho = 0.76), Christensenellaceae and *MDN1* (Spearman rho = 0.78), and *Oscillospira* and *NUDT14* (Spearman rho = 0.79). A few significant negative correlations (q-value < 0.05), such as between Christensenellaceae and *TBX10* (Spearman rho = - 0.78), and Ruminococcaceae and *LCN2* (Spearman rho = -0.77) were also found.

To characterize potential microbe-microbe interactions in our dataset, we computed correlations between the microbes significantly correlated (q-value <0.1) with the genes using sparCC (see Methods and Additional file 4) [39]. The notable aspects of the significant gene-microbe correlations (q-value < 0.1) and significant microbe-microbe correlations (SparCC |R| >=0.1 and pseudo-P value <0.05) are graphically represented in **Figure 3B**, where solid edges denote gene-microbe correlations and dashed edges represent microbe-microbe correlations. This subnetwork of microbe-microbe correlations depicts correlated abundance changes in the microbiome as a function of their presence (**Figure 3B**, dashed edges). For instance, *Bilophila* and *Butyricimonas* are both depleted in CF (q-value < 0.05), and the abundance of the two genera is also correlated across individuals (SparCC R = 0.5, pseudo-P value = 0.04). On the other hand, Ruminococcaceae was found depleted in CF (q-value = 0.081), while *Clostridium* was enriched in CF (q-value = 0.0004), and this inverse co-occurrence pattern leads to a negative correlation between the two taxa across study participants (SparCC R = -0.66, pseudo-P value = 0). Furthermore, in the gene-microbe subnetwork (**Figure 3B**, solid edges), microbial nodes have more edges on average compared to genes, where Christensenellaceae and *Clostridium* formed distinct hubs in the network. This potentially implies that these microbes and their pathways are shared across multiple GI cancer-associated genes. Of note, *Bilophila, Clostridium*, and *Pseudomonas* are mostly negatively correlated with GI cancer genes, while *Haemophilus, Oscillospira, Veillonella, Fusobacterium and Acidaminococcus* are only positively correlated with GI cancer genes (q-value < 0.1).

In addition to the overall network, **Figure 3C** depicts pairwise correlations between host gene expression and microbial taxa where both have been previously linked to CRC, and thus may be of interest. For example, *LCN2*, known to be overexpressed in human CRC and other cancers (Maier et al. 2014), is negatively correlated with Ruminococcaceae (Spearman rho = -0.77, q-value = 0.040), which is found depleted in CRC [61,62]. Both *DUOX2* and *DUOXA2* are found to be negatively correlated with Christensenellaceae (Spearman rho < -0.65, q-value < 0.1), while *DUOXA2* is positively correlated with *Veillonella* (Spearman rho = 0.70, q-value = 0.082). *DUOX2* and its maturation factor *DUOXA2* are responsible for H2O2 production in human colon and are known to be up-regulated in gastrointestinal inflammation [63,64]. Christensenellaceae, a heritable taxon [65], has been shown to decrease in abundance in conventional adenoma [62], a precursor of CRC, whereas *Veillonella*, which is known to be pro-inflammatory, is found to be represented in human CRC [66]. Thus, the pattern of grouping by CF and healthy samples in these representative correlations are found to be similar to known associations in CRC and other gastrointestinal malignancies.

Discussion

Recent advances in the treatment have significantly prolonged the lives of CF patients [67,68]. However, this led to new challenges, such as an elevated risk for gastrointestinal cancer [2,69]. Thus, CF patients show 5-10-fold increased risk of CRC compared to healthy individuals, and that increases even further with immunosuppressive drugs [3,6]. Understanding the molecular mechanisms that control the increased risk is key for early detection and the development of tailored treatments[6]. The importance of interactions between host and microbiome in the pathogenesis of colorectal cancer has become increasingly clear [61,70,71]. To understand the role of these interactions in CF, we jointly profiled host colon gene expression and mucosal microbiome composition data in CF patients and healthy controls. We observed an enrichment of cancer-associated dysregulated genes — specifically colon cancer — in CF patients compared to healthy controls. We also observed a shift in the microbiome and identified strains previously linked to colon cancer that varied in their abundance between CF and healthy individuals. We further found relevant correlations between these cancer enriched genes and microbes that may illuminate the mechanisms of CRC development in CF patients.

Several previous studies have studied the role of host gene regulation in CF patients [5,72,73]. While results from previous studies are based on either phenotypic observations, examining candidate genes such as *CFTR*, or an exploration of gene expression data from respiratory or blood samples [5,72,73], our work is the first, as far as we know, that focused on a comprehensive transcriptomic analysis of colon biopsies. This allowed us to characterize patterns of host gene regulation specific to the CF colon epithelium. In addition to an enrichment of cancer-related pathways among genes that are differentially expressed in CF, we also observed an enrichment for immune response pathways, including signal transduction, cell adhesion, and viral infection. Interestingly, one of the most significant pathways enriched in our current data, the eIF2 signaling pathway, has been previously shown to play an important role in immune response, and cells with defective eIF2 signaling pathway were more susceptible to bacterial infections [74]. In addition, our analysis revealed that tumor suppressor genes are differentially regulated in the colon of CF patients. In addition to *CFTR*, we found other tumor suppressor genes, such as *HPGD*, to be down-regulated in CF patients colon. *HPGD* was previously shown to be down-regulated in lungs of CF patients [5,75]. Down-regulation of these tumor suppressor genes can lead to predisposition of colon cancer [42,76,77], suggesting a potential mechanism underlying the reported increased risk and early development of colon cancer in CF patients [5,69].

In addition to host gene regulation, the microbiome has also been implicated in the development of many diseases, including CRC [61,78]. In the context of CF, previous studies have focused on characterizing shifts in the fecal or airway microbiome [20,79]. Here, we profiled the colonic mucosal microbiome, with the goal of understanding its role in the development of CRC in CF patients. We found a clear distinction between microbiome populations from CF compared to healthy mucosa. Overall, similar to several other GI diseases, we also observed a reduced microbial biodiversity in the CF population [80]. We also found a depletion in butyrate producing bacteria, such as Ruminococcaceae and *Butyricimonas, s*imilar to previously reported depletion in butyrate producing microbes by Manor et al [20] in their study comparing CF fecal samples from children on varying degree of fat intake. Butyrate helps promote growth and can also act as an anti-inflammatory agent, and is therefore an important compound for colon health [20]. Interestingly, mice with compromised GI defense system also had a reduced number of butyrate producing bacteria, similar to our observations in the CF patients, who generally consume a high fat diet [81]. In addition to decrease in Ruminococcaceae, we also observed a depletion in another butyrate producing bacteria *Sutterella;* loss of abundance in both of these strains have been previously observed in CRC [61,82]. We also found an increase in Actinobacteria, one of the most predominant genera found in the sputum of CF patients [79,83], but decreased in colon cancer gut microbiome [84]. Furthermore, our observation of significant decrease in the abundance of Verrucomicrobia, and increase in abundance of Firmicutes and Actinobacteria in CF patients, is consistent with findings from the fecal microbiome of CF patients [21]. Lastly, we found an increase in predicted bacterial toxins in the CF population, which might be explained by the increase in pathogenic bacteria such as *Pseudomonas* and *Veillonella.* This can potentially damage epithelial cells or induce mutations leading to unfavorable clinical outcome [85,86].

Integrating mucosal microbiome and host gene expression profiles, we observed several correlations between differentially expressed colon epithelial genes and gut mucosal bacteria in CF. Co-culture and obligate cross-feeding studies have shown an increased virulence of a pathogen in the presence of other bacteria, thus triggering an immune response that can determine the clinical outcome [87,88]. One such example is the increased virulence of *Pseudomonas* in presence of *Veillonella* as seen in mice tumor model resulting in host clinical deterioration [88]. Interestingly, we found both of these microbes (*Veillonella* and *Pseudomonas*) in higher abundance in CF patients. Furthermore, we also found a strong correlation between *Veillonella* and *DUOXA2*, a highly expressed gene causing inflammation in ulcerative colitis [89]. Another such correlation that we observed was between highly expressed *LNC2* gene, which plays a role in innate immunity and has been previously found to be up-regulated in human colon cancers [90], and depletion of Ruminococcaceae, a butyrate producing bacteria that helps maintain colon health [20].

Our study has several limitations. First, CF patients have a high burden of antibiotic exposure. Since antibiotics affects the gut microbiome [91–93], this may impact the differences we observe between CF and healthy mucosal microbiome. It is challenging to account for this potential bias; although matched healthy controls that are also on antibiotics can be used, the effects of long-term antibiotics usage may be impossible to match in a non-CF control population. In addition, although we report potential host gene-microbe and microbe-microbe interactions, our study focused on correlations, and causality is not inferred. Although studying causality in host-microbe is challenging in humans, future studies using animal or cell models can be useful to disentangle the direction of interaction [94].

Conclusions

To summarize, we report an analysis of the mucosal microbiome and host gene expression in the gut of CF patients and healthy controls. We find down-regulation of tumor suppressor genes, as well as up-regulation of genes that play a role in immune response and cause inflammation. Furthermore, we observe a shift in microbiome with depletion in butyrate producing bacteria that may help maintain colon health and increase in pathogenic strains in individuals with CF. Lastly, our study provides a set of candidate interactions between gut microbes and host genes in the CF gut. Our work sheds light on the role of host-microbiome interactions and their relevance for the early development of CRC in CF patients. Our results can provide clinicians and researchers with biomarkers that may potentially serve as targets for stratifying risk of CRC in patients with CF.

Abbreviations: CF: cystic fibrosis; CRC: colorectal cancer; GI: gastrointestinal; FDR: false discovery rate; OTU: operational taxonomic unit; PICRUSt: Phylogenetic Investigation of Communities by Reconstruction of Unobserved States; KEGG: Kyoto Encyclopedia of Genes and Genomes.

Acknowledgements

We thank the members of the Blekhman Lab for helpful discussions. This work was carried out, in part, by resources provided by the Minnesota Supercomputing Institute. This work is supported by award MNP 17.26 from the Minnesota Partnership for Biotechnology and Medical Genomics (R.B.), a McKnight Land-Grant Professorship from the University of Minnesota (R.B.), the Chainbreaker Breakthrough Cancer Research Grant from the Masonic Cancer Center at the University of Minnesota (A.K. and R.B), With One Breath (A.K.), and the Cystic Fibrosis Foundation (D.E.N.). Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health Award Number

UL1-TR002494. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. O'Sullivan BP, Freedman SD. Cystic fibrosis. Lancet. 2009;373:1891–904.

2. Maisonneuve P, Marshall BC, Knapp EA, Lowenfels AB. Cancer risk in cystic fibrosis: a 20-year nationwide study from the United States. J Natl Cancer Inst. 2013;105:122–9.

3. Yamada A, Komaki Y, Komaki F, Micic D, Zullow S, Sakuraba A. Risk of gastrointestinal cancers in patients with cystic fibrosis: a systematic review and meta-analysis. Lancet Oncol. 2018;19:758–67.

4. Niccum DE, Billings JL, Dunitz JM, Khoruts A. Colonoscopic screening shows increased early incidence and progression of adenomas in cystic fibrosis. J Cyst Fibros. 2016;15:548–53.

5. Than BLN, Linnekamp JF, Starr TK, Largaespada DA, Rod A, Zhang Y, et al. CFTR is a tumor suppressor gene in murine and human intestinal cancer. Oncogene. 2016;35:4179–87.

6. Hadjiliadis D, Khoruts A, Zauber AG, Hempstead SE, Maisonneuve P, Lowenfels AB, et al. Cystic Fibrosis Colorectal Cancer Screening Consensus Recommendations. Gastroenterology. 2018;154:736–45.e14.

7. Starr TK, Allaei R, Silverstein KAT, Staggs RA, Sarver AL, Bergemann TL, et al. A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science. 2009;323:1747–50.

8. Norkina O, Burnett TG, De Lisle RC. Bacterial overgrowth in the cystic fibrosis transmembrane conductance regulator null mouse small intestine. Infect Immun. 2004;72:6040–9.

9. Fridge JL, Conrad C, Gerson L, Castillo RO, Cox K. Risk factors for small bowel bacterial overgrowth in cystic fibrosis. J Pediatr Gastroenterol Nutr. 2007;44:212–8.

10. Gallagher AM, Gottlieb RA. Proliferation, not apoptosis, alters epithelial cell migration in small intestine of CFTR null mice. Am J Physiol Gastrointest Liver Physiol. 2001;281:G681–7.

11. Norkina O, Kaur S, Ziemer D, De Lisle RC. Inflammation of the cystic fibrosis mouse small intestine. Am J Physiol Gastrointest Liver Physiol. 2004;286:G1032–41.

12. Bruzzese E, Raia V, Gaudiello G, Polito G, Buccigrossi V, Formicola V, et al. Intestinal inflammation is a frequent feature of cystic fibrosis and is reduced by probiotic administration. Aliment Pharmacol Ther. 2004;20:813–9.

13. Jakab RL, Collaco AM, Ameen NA. Physiological relevance of cell-specific distribution patterns of CFTR, NKCC1, NBCe1, and NHE3 along the crypt-villus axis in the intestine. Am J Physiol Gastrointest Liver Physiol. 2011;300:G82–98.

14. Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat AA, et al. Single-cell

dissection of transcriptional heterogeneity in human colon tumors. Nat Biotechnol. 2011;29:1120–7.

15. Barker N, Ridgway RA, van Es JH, van de Wetering M, Begthel H, van den Born M, et al. Crypt stem cells as the cells-of-origin of intestinal cancer. Nature. 2009;457:608–11.

16. Lynch SV, Goldfarb KC, Wild YK, Kong W, De Lisle RC, Brodie EL. Cystic fibrosis transmembrane conductance regulator knockout mice exhibit aberrant gastrointestinal microbiota. Gut Microbes. 2013;4:41–7.

17. Duytschaever G, Huys G, Bekaert M, Boulanger L, De Boeck K, Vandamme P. Dysbiosis of bifidobacteria and Clostridium cluster XIVa in the cystic fibrosis fecal microbiota. J Cyst Fibros. 2013;12:206–15.

18. Schippa S, Iebba V, Santangelo F, Gagliardi A, De Biase RV, Stamato A, et al. Cystic fibrosis transmembrane conductance regulator (CFTR) allelic variants relate to shifts in faecal microbiota of cystic fibrosis patients. PLoS One. 2013;8:e61176.

19. Flass T, Tong S, Frank DN, Wagner BD, Robertson CE, Kotter CV, et al. Intestinal lesions are associated with altered intestinal microbiome and are more frequent in children and young adults with cystic fibrosis and cirrhosis. PLoS One. 2015;10:e0116967.

20. Manor O, Levy R, Pope CE, Hayden HS, Brittnacher MJ, Carr R, et al. Metagenomic evidence for taxonomic dysbiosis and functional imbalance in the gastrointestinal tracts of children with cystic fibrosis. Sci Rep. 2016;6:22493.

21. Burke DG, Fouhy F, Harrison MJ, Rea MC, Cotter PD, O'Sullivan O, et al. The altered gut microbiota in adults with cystic fibrosis. BMC Microbiol. 2017;17:58.

22. Miragoli F, Federici S, Ferrari S, Minuti A, Rebecchi A, Bruzzese E, et al. Impact of cystic fibrosis disease on archaea and bacteria composition of gut microbiota. FEMS Microbiol Ecol [Internet]. 2017;93. Available from: http://dx.doi.org/10.1093/femsec/fiw230

23. de Freitas MB, Moreira EAM, Tomio C, Moreno YMF, Daltoe FP, Barbosa E, et al. Altered intestinal microbiota composition, antibiotic therapy and intestinal inflammation in children and adolescents with cystic fibrosis. PLoS One. 2018;13:e0198457.

24. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

25. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7.

26. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017;14:687–90.

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

28. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4:1521.

29. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4:1184–91.

30. Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet. 2008;4:e1000255.

31. Al-Ghalith GA, Hillmann B, Ang K, Shields-Cutler R, Knights D. SHI7 is a Self-Learning pipeline for multipurpose Short-Read DNA quality control. mSystems 3: e00202-17. DOI: https://doi org/10 1128/mSystems. 2018;00202–17.

32. Al-Ghalith GA, Montassier E, Ward HN, Knights D. NINJA-OPS: Fast Accurate Marker Gene Alignment Using Concatenated Ribosomes. PLoS Comput Biol. 2016;12:e1004658.

33. Oksanen J, Kindt R, Legendre P, O'Hara B, Simpson GL, Solymos P, et al. The vegan package. Community ecology package [http://r-forge r-project org/projects/vegan/] [Internet]. 2008; Available from: https://www.researchgate.net/profile/Gavin_Simpson/publication/228339454_The_vegan_Package/links/0912f50be86bc29a7f000000/The-vegan-Package.pdf

34. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013;8:e61217.

35. Johansen HK, Nir M, Koch C, Schwartz M, Høiby N. Severity of cystic fibrosis in patients homozygous and heterozygous for ΔF508 mutation. Lancet. 1991;337:631–4.

36. Langille MGI, 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;31:814–21.

37. Dabney A, Storey JD, Warnes GR. qvalue: Q-value estimation for false discovery rate control. R package version [Internet]. 2010;1. Available from: ftp://ftp.uni-bayreuth.de/pub/math/statlib/R/CRAN/src/contrib/Descriptions/qvalue.html

38. Wei T, Simko V. corrplot: Visualization of a correlation matrix. R package version 0 73. 2013;230:11.

39. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8:e1002687.

40. Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 2011;27:431–2.

41. Flo TH, Smith KD, Sato S, Rodriguez DJ, Holmes MA, Strong RK, et al. Lipocalin 2 mediates an innate immune response to bacterial infection by sequestrating iron. Nature. 2004;432:917–21.

42. Wu X-L, Yang Z-W, He L, Dong P-D, Hou M-X, Meng X-K, et al. RRS1 silencing suppresses colorectal cancer cell proliferation and tumorigenesis by inhibiting G2/M progression and angiogenesis. Oncotarget. 2017;8:82968–80.

43. Berens EB, Sharif GM, Schmidt MO, Yan G, Shuptrine CW, Weiner LM, et al. Keratin-associated protein 5-5 controls cytoskeletal function and cancer cell vascular invasion. Oncogene. 2017;36:593–605.

44. Bu P, Chen K-Y, Xiang K, Johnson C, Crown SB, Rakhilin N, et al. Aldolase B-Mediated

Fructose Metabolism Drives Metabolic Reprogramming of Colon Cancer Liver Metastasis. Cell Metab. 2018;27:1249–62.e4.

45. Kumara HMCS, Bellini GA, Caballero OL, Herath SAC, Su T, Ahmed A, et al. P-Cadherin (CDH3) is overexpressed in colorectal tumors and has potential as a serum marker for colorectal cancer monitoring. Oncoscience. 2017;4:139–47.

46. Zhu H, Dougherty U, Robinson V, Mustafi R, Pekow J, Kupfer S, et al. EGFR Signals Downregulate Tumor Suppressors miR-143 and miR-145 in Western Diet–Promoted Murine Colon Cancer: Role of G1 Regulators. Mol Cancer Res [Internet]. American Association for Cancer Research; 2011 [cited 2019 Jan 10]; Available from: http://mcr.aacrjournals.org/content/early/2011/07/01/1541-7786.MCR-10-0531.short

47. Romero M, Sabaté-Pérez A, Francis VA, Castrillón-Rodriguez I, Díaz-Ramos Á, Sánchez-Feutrie M, et al. TP53INP2 regulates adiposity by activating β-catenin through autophagy-dependent sequestration of GSK3β. Nat Cell Biol. 2018;20:443–54.

48. Yao L, Tak YG, Berman BP, Farnham PJ. Functional annotation of colon cancer risk SNPs. Nat Commun. 2014;5:5114.

49. Langenskiöld M, Holmdahl L, Angenete E, Falk P, Nordgren S, Ivarsson M-L. Differential prognostic impact of uPA and PAI-1 in colon and rectal cancer. Tumour Biol. 2009;30:210–20.

50. Dong Q, Meng P, Wang T, Qin W, Qin W, Wang F, et al. MicroRNA let-7a inhibits proliferation of human prostate cancer cells in vitro and in vivo by targeting E2F2 and CCND2. PLoS One. 2010;5:e10147.

51. Mazzoccoli G, Pazienza V, Panza A, Valvano MR, Benegiamo G, Vinciguerra M, et al. ARNTL2 and SERPINE1: potential biomarkers for tumor aggressiveness in colorectal cancer. J Cancer Res Clin Oncol. 2012;138:501–11.

52. Tazawa H, Tsuchiya N, Izumiya M, Nakagama H. Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. Proc Natl Acad Sci U S A. 2007;104:15472–7.

53. Yu K, Lujan R, Marmorstein A, Gabriel S, Hartzell HC. Bestrophin-2 mediates bicarbonate transport by goblet cells in mouse colon. J Clin Invest. 2010;120:1722–35.

54. Qu Z, Hartzell HC. Bestrophin Cl− channels are highly permeable to HCO3−. American Journal of Physiology-Cell Physiology. American Physiological Society; 2008;294:C1371–7.

55. Ham J, Costa C, Sano R, Lochmann TL, Sennott EM, Patel NU, et al. Exploitation of the Apoptosis-Primed State of MYCN-Amplified Neuroblastoma to Develop a Potent and Specific Targeted Therapy Combination. Cancer Cell. 2016;29:159–72.

56. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. N Engl J Med. 1988;319:525–32.

57. Chellappa K, Robertson GR, Sladek FM. HNF4α: a new biomarker in colon cancer? Biomark Med. 2012;6:297–300.

58. Hünten S, Hermeking H. p53 directly activates cystatin D/CST5 to mediate mesenchymal-epithelial transition: a possible link to tumor suppression by vitamin D3. Oncotarget.

2015;6:15842–56.

59. Feber A, Clark J, Goodwin G, Dodson AR, Smith PH, Fletcher A, et al. Amplification and overexpression of E2F3 in human bladder cancer. Oncogene. 2004;23:1627–30.

60. Miles WO, Tschöp K, Herr A, Ji J-Y, Dyson NJ. Pumilio facilitates miRNA regulation of the E2F3 oncogene. Genes Dev. 2012;26:356–68.

61. Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Med. 2015;7:55.

62. Peters BA, Dominianni C, Shapiro JA, Church TR, Wu J, Miller G, et al. The gut microbiota in conventional and serrated precursors of colorectal cancer. Microbiome. 2016;4:69.

63. Wu Y, Antony S, Juhasz A, Lu J, Ge Y, Jiang G, et al. Up-regulation and sustained activation of Stat1 are essential for interferon-gamma (IFN-gamma)-induced dual oxidase 2 (Duox2) and dual oxidase A2 (DuoxA2) expression in human pancreatic cancer cell lines. J Biol Chem. 2011;286:12245–56.

64. Wu Y, Antony S, Hewitt SM, Jiang G, Yang SX, Meitzler JL, et al. Functional activity and tumor-specific expression of dual oxidase 2 in pancreatic cancer cells and human malignancies characterized with a novel monoclonal antibody. Int J Oncol. 2013;42:1229–38.

65. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014;159:789–99.

66. Geng J, Song Q, Tang X, Liang X, Fan H, Peng H, et al. Co-occurrence of driver and passenger bacteria in human colorectal cancer. Gut Pathog. 2014;6:26.

67. Elborn JS, Flume PA, Loutit J, Cohen F. WS7.5 Prolonged improvement in lung function and quality of life in cystic fibrosis: a 24-week extension study of levofloxacin nebulization solution (APT-1026) versus tobramycin nebulization solution in stable CF patients with chronic Pseudomonas aeruginosa infection. J Cyst Fibros. 2014;13:S16.

68. Cohen-Cymberknoh M, Shoseyov D, Kerem E. Managing cystic fibrosis: strategies that increase life expectancy and improve quality of life. Am J Respir Crit Care Med. 2011;183:1463–71.

69. Hegagi M, Aaron SD, James P, Goel R, Chatterjee A. Increased prevalence of colonic adenomas in patients with cystic fibrosis. J Cyst Fibros. 2017;16:759–62.

70. Hurwitz BL. 28 The Relationship of Host Genetics and the Microbiome in Colon Cancer. J Anim Sci. 2018;96:15–15.

71. Bhutia YD, Ogura J, Sivaprakasam S, Ganapathy V. Gut Microbiome and Colon Cancer: Role of Bacterial Metabolites and Their Molecular Targets in the Host. Curr Colorectal Cancer Rep. 2017;13:111–8.

72. Tata M, Wolfinger MT, Amman F, Roschanski N, Dötsch A, Sonnleitner E, et al. RNASeq Based Transcriptional Profiling of Pseudomonas aeruginosa PA14 after Short- and Long-Term Anoxic Cultivation in Synthetic Cystic Fibrosis Sputum Medium. PLoS One. 2016;11:e0147811.

73. Kormann MSD, Dewerth A, Eichner F, Baskaran P, Hector A, Regamey N, et al.

Transcriptomic profile of cystic fibrosis patients identifies type I interferon response and ribosomal stalk proteins as potential modifiers of disease severity. PLoS One. 2017;12:e0183526.

74. Shrestha N, Bahnan W, Wiley DJ, Barber G, Fields KA, Schesser K. Eukaryotic initiation factor 2 (eIF2) signaling regulates proinflammatory cytokine expression and bacterial invasion. J Biol Chem. 2012;287:28738–44.

75. Wu Y 'an, Wang X, Wu F, Huang R, Xue F, Liang G, et al. Transcriptome profiling of the cancer, adjacent non-tumor and distant normal tissues from a colorectal cancer patient by deep sequencing. PLoS One. 2012;7:e41001.

76. Kapitanovic S. APC tumor suppressor gene in sporadic colon cancer. Eur J Cancer. 2002;38:S51.

77. Myung S-J, Rerko RM, Yan M, Platzer P, Guda K, Dotson A, et al. 15-Hydroxyprostaglandin dehydrogenase is an in vivo suppressor of colon tumorigenesis. Proc Natl Acad Sci U S A. 2006;103:12098–102.

78. O'Keefe SJ. Abstract SS01-01: The microbiome and colon cancer risk. Cancer Epidemiol Biomarkers Prev. American Association for Cancer Research; 2014;23:SS01–01 – SS01–01.

79. Moran Losada P, Chouvarine P, Dorda M, Hedtfeld S, Mielke S, Schulz A, et al. The cystic fibrosis lower airways microbial metagenome. ERJ Open Res [Internet]. 2016;2. Available from: http://dx.doi.org/10.1183/23120541.00096-2015

80. Tilg H, Kaser A. Gut microbiome, obesity, and metabolic dysfunction. J Clin Invest. 2011;121:2126–32.

81. Hildebrandt MA, Hoffmann C, Hamady M, Chen Y-Y, Knight R, Bushman FD, et al. 662 High Fat Diet Determines the Composition of the Gut Microbiome Independent of Host Genotype and Phenotype. Gastroenterology. 2009;136:A – 102.

82. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. Nat Commun. 2015;6:8727.

83. Rudkjøbing VB, Thomsen TR, Alhede M, Kragh KN, Nielsen PH, Johansen UR, et al. The microorganisms in chronically infected end-stage and non-end-stage cystic fibrosis patients. FEMS Immunol Med Microbiol. 2012;65:236–44.

84. Gao Z, Guo B, Gao R, Zhu Q, Qin H. Microbiota disbiosis is associated with colorectal cancer. Front Microbiol. 2015;6:20.

85. Barbieri JT. Bacterial toxins that modify the epithelial cell barrier. Bacterial-Epithelial Cell Cross-Talk: Molecular Mechanisms in Pathogenesis. Cambridge University Press; 2006. p. 184–210.

86. Hamon MA, Batsché E, Régnault B, Tham TN, Seveau S, Muchardt C, et al. Histone modifications induced by a family of bacterial toxins. Proc Natl Acad Sci U S A. 2007;104:13467–72.

87. Adamowicz EM, Flynn J, Hunter RC, Harcombe WR. Cross-feeding modulates antibiotic tolerance in bacterial communities. ISME J. 2018;12:2723–35.

88. Pustelny C, Komor U, Pawar V, Lorenz A, Bielecka A, Moter A, et al. Contribution of Veillonella parvula to Pseudomonas aeruginosa-mediated pathogenicity in a murine tumor model system. Infect Immun. 2015;83:417–29.

89. MacFie TS, Poulsom R, Parker A. DUOX2 and DUOXA2 Form the Predominant Enzyme System Capable of Producing the Reactive Oxygen Species H2O2 in Active Ulcerative Colitis and are …. Inflamm Bowel Dis [Internet]. academic.oup.com; 2014; Available from: https://academic.oup.com/ibdjournal/article-abstract/20/3/514/4579005

90. Maier HT, Aigner F, Trenkwalder B, Zitt M, Vallant N, Perathoner A, et al. Up-regulation of neutrophil gelatinase-associated lipocalin in colorectal cancer predicts poor patient survival. World J Surg. 2014;38:2160–7.

91. Francino MP. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. Front Microbiol. 2015;6:1543.

92. Jakobsson HE, Jernberg C, Andersson AF, Sjölund-Karlsson M, Jansson JK, Engstrand L. Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. PLoS One. 2010;5:e9836.

93. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, et al. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn’s Disease. Cell Host Microbe. 2015;18:489–500.

94. Luca F, Kupfer SS, Knights D, Khoruts A, Blekhman R. Functional Genomics of Host–Microbiome Interactions in Humans. Trends Genet. 2018;34:30–40.

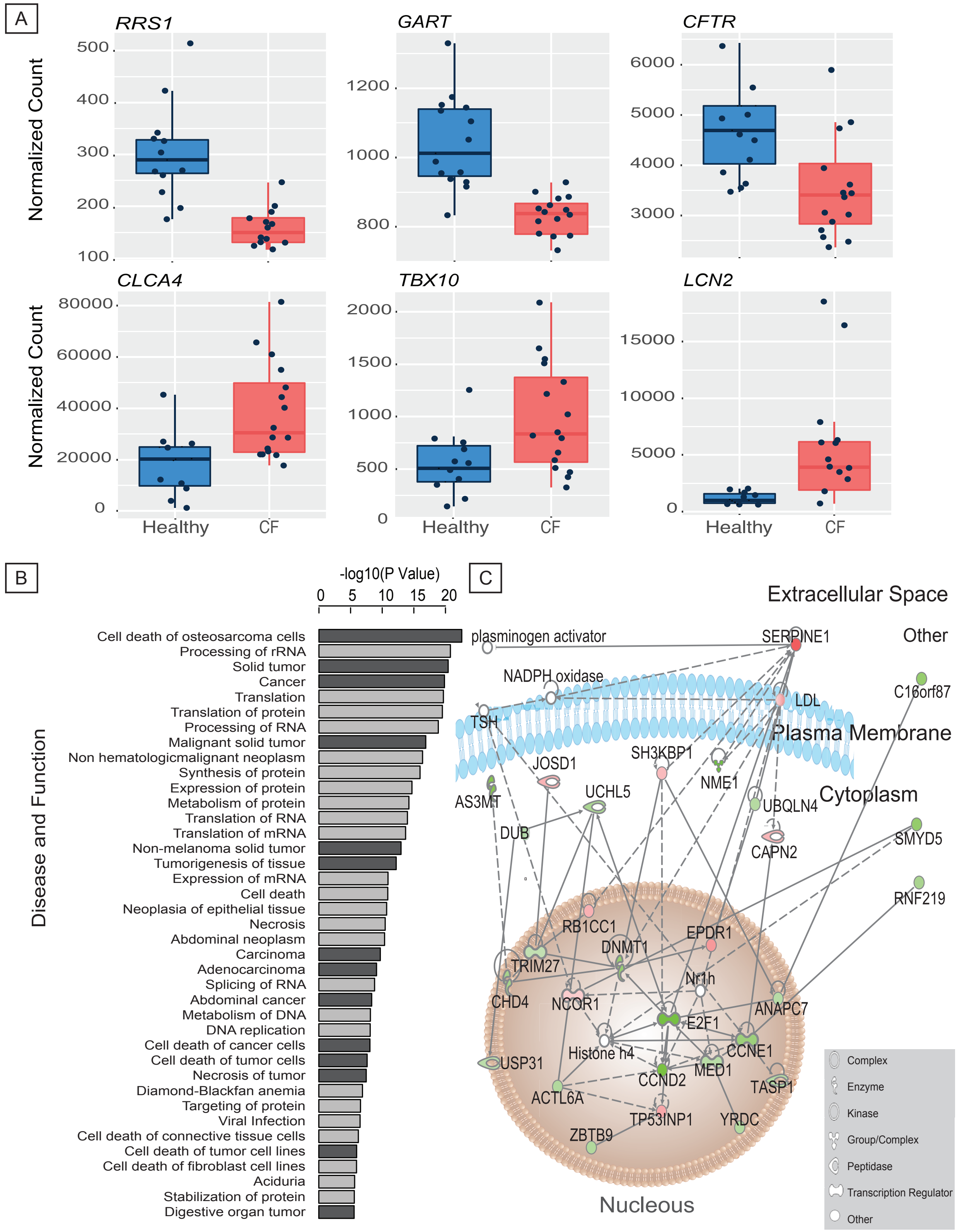
A
Normalized Count
RRS1
500
400
300
200
100
GART
1200
1000
800
CFTR
6000
5000
4000
3000
Normalized Count
CLCA4
80000
60000
40000
20000
0
Healthy
CF
TBX10
2000
1500
1000
500
0
Healthy
CF
LCN2
15000
10000
5000
0
Healthy
CF
B
-log10(P Value)
0
5
10
15
20
Disease and Function
Cell death of osteosarcoma cells
Processing of rRNA
Solid tumor
Cancer
Translation
Translation of protein
Processing of RNA
Malignant solid tumor
Non hematologicmalignant neoplasm
Synthesis of protein
Expression of protein
Metabolism of protein
Translation of RNA
Translation of mRNA
Non-melanoma solid tumor
Tumorigenesis of tissue
Expression of mRNA
Cell death
Neoplasia of epithelial tissue
Necrosis
Abdominal neoplasm
Carcinoma
Adenocarcinoma
Splicing of RNA
Abdominal cancer
Metabolism of DNA
DNA replication
Cell death of cancer cells
Cell death of tumor cells
Necrosis of tumor
Diamond-Blackfan anemia
Targeting of protein
Viral Infection
Cell death of connective tissue cells
Cell death of tumor cell lines
Cell death of fibroblast cell lines
Aciduria
Stabilization of protein
Digestive organ tumor
C
Extracellular Space
Other
plasminogen activator
SERPINE1
NADPH oxidase
C16orf87
LDL
TSH
Plasma Membrane
JOSD1
SH3KBP1
NME1
AS3MT
UCHL5
UBQLN4
Cytoplasm
DUB
CAPN2
SMYD5
RNF219
RB1CC1
EPDR1
DNMT1
TRIM27
Nr1h
CHD4
NCOR1
ANAPC7
E2F1
Histone h4
CCNE1
USP31
MED1
CCND2
TASP1
ACTL6A
TP53INP1
YRDC
ZBTB9
Nucleous
Complex
Enzyme
Kinase
Group/Complex
Peptidase
Transcription Regulator
Other

Figure 1. Differentially expressed (DE) host genes between CF patients and healthy individuals. (A) Box plots showing the expression level of six significantly DE genes that are a part of the gastrointestinal cancer pathway (B) Functional and disease categories that are enriched among DE genes sorted by the p-value (y-axis)), with darker bars indicating cancer-related categories. (C) Gene-gene interaction network showing genes in the gastrointestinal cancer pathway with green representing up-regulated genes in CF and red representing down-regulated genes in CF. The intensity of the color is indicative of more (brighter) extreme or less (duller) log-fold change measurement in the dataset. The shapes represent the protein function, illustrated in the legend at the bottom right. The figure is structured to show the cellular location in which each gene is active.

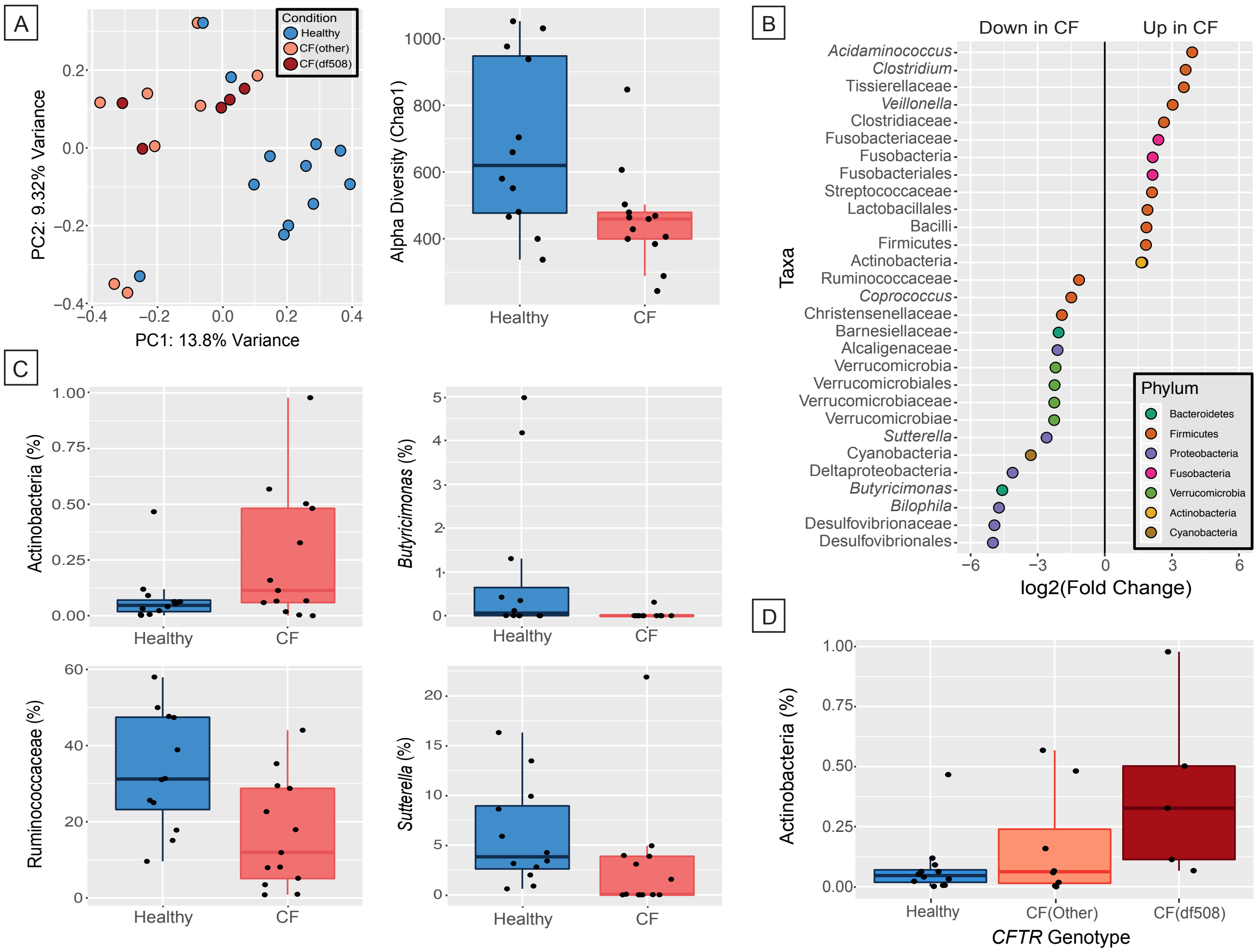

Figure 2. Differences between CF and healthy gut mucosal microbiota. (A) (left) Principal coordinate analysis plot based on Bray-Curtis distance indicating difference in beta-diversity between CF and healthy gut mucosal microbiome. The axes represent the percentage variance along the first two principal components, and the color of samples indicates their mutation status, i.e. healthy (i.e. no known mutation in CFTR), CF(df508) (homozygous for the DF508 mutation), and CF(other) (either one or zero alleles of the DF508 mutation); (right) Boxplot depicting difference in alpha-diversity for Chao1 metric between CF and healthy gut microbiome. (B) Dotplot showing significantly differentially abundant taxa (q-value < 0.1) between CF and healthy samples. The taxa are listed along the y-axis and are colored by their phylum,, and the x-axis indicates the log2 fold-change in CF compared to healthy as baseline. (C) Boxplots indicating the percentage relative abundance of taxa showing differential abundance between CF and healthy gut microbiome (q-value < 0.1). (D) Boxplot depicting the abundance of Actinobacteria for three mutation levels - Healthy, CF(other) and CF(df508).

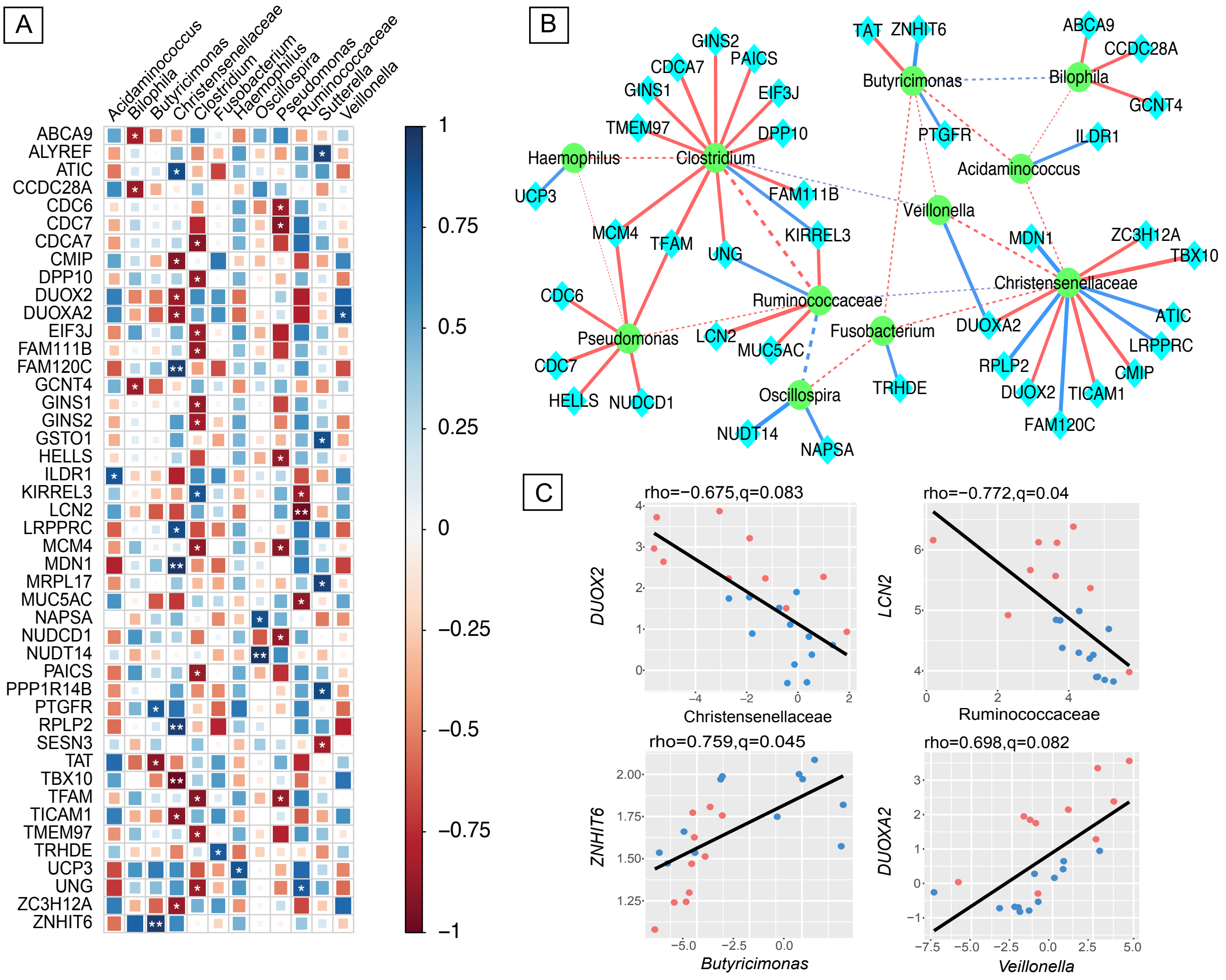

Figure 3. Interactions between gastrointestinal cancer-related host genes and gut mucosal microbes. (A) Correlation plot depicting gene-microbe correlations. Color and size of the squares indicate the magnitude of the correlation, asterisks indicate significance of correlation (** indicates q-value < 0.05 and * indicates q-value < 0.1). (B) Network visualizing the significant gene-microbe correlations (solid edges, q-value < 0.1) and significant microbe-microbe correlations (dashed edges, SparCC |R| >=0.1 and p-value < 0.05). Blue edges indicate positive correlation and red edges indicate negative correlation. Edge thickness represents the strength of the correlation. (C) Scatterplots depicting pattern of grouping by CF (red) and healthy (blue) samples in a few representative gene-microbe correlations, where the strength of correlation (Spearman rho) and significance (q) are indicated on the top.