Interaction between Host MicroRNAs and the Gut Microbiota in Colorectal Cancer

Ce Yuan,a,b Michael B. Burns,c Subbaya Subramanian,a,b,d Ran Blekhman,e,f

a Bioinformatics and Computational Biology Program, University of Minnesota, Rochester, Minnesota, USA
b Department of Surgery, University of Minnesota, Minneapolis, Minnesota, USA
c Department of Biology, Loyola University, Chicago, Illinois, USA
d Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota, USA
e Department of Genetics, Cell Biology, and Development, University of Minnesota, Minneapolis, Minnesota, USA
f Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, Minnesota, USA

ABSTRACT Although variation in gut microbiome composition has been linked with colorectal cancer (CRC), the factors that mediate the interactions between CRC tumors and the microbiome are poorly understood. MicroRNAs (miRNAs) are known to regulate CRC progression and are associated with patient survival outcomes. In addition, recent studies suggested that host miRNAs can also regulate bacterial growth and influence the composition of the gut microbiome. Here, we investigated the association between miRNA expression and microbiome composition in human CRC tumor and normal tissues. We identified 76 miRNAs as differentially expressed (DE) in tissue from CRC tumors and normal tissue, including the known oncogenic miRNAs miR-182, miR-503, and mir-17~92 cluster. These DE miRNAs were correlated with the relative abundances of several bacterial taxa, including Firmicutes, Bacteroidetes, and Proteobacteria. Bacteria correlated with DE miRNAs were enriched with distinct predicted metabolic categories. Additionally, we found that miRNAs that correlated with CRC-associated bacteria are predicted to regulate targets that are relevant for host-microbiome interactions and highlight a possible role for miRNA-driven glycan production in the recruitment of pathogenic microbial taxa. Our work characterized a global relationship between microbial community composition and miRNA expression in human CRC tissues.

IMPORTANCE Recent studies have found an association between colorectal cancer (CRC) and the gut microbiota. One potential mechanism by which the microbiota can influence host physiology is through affecting gene expression in host cells. MicroRNAs (miRNAs) are small noncoding RNA molecules that can regulate gene expression and have important roles in cancer development. Here, we investigated the link between the gut microbiota and the expression of miRNA in CRC. We found that dozens of miRNAs are differentially regulated in CRC tumors and adjacent normal colon and that these miRNAs are correlated with the abundance of microbes in the tumor microenvironment. Moreover, we found that microbes that have been previously associated with CRC are correlated with miRNAs that regulate genes related to interactions with microbes. Notably, these miRNAs likely regulate glycan production, which is important for the recruitment of pathogenic microbial taxa to the tumor. This work provides a first systems-level map of the association between microbes and host miRNAs in the context of CRC and provides targets for further experimental validation and potential interventions.

KEYWORDS colorectal cancer, gene regulation, microRNA, microbiome, tumor microenvironment
The colon microenvironment hosts trillions of microbes, known as the gut microbiome. A healthy microbiome helps maintain colon microenvironment homeostasis, immune system development, gut epithelial function, and other organ functions (1–5). Although many factors impact the composition of the gut microbiome, the overall functional profiles remain stable over time (6, 7). Nevertheless, changes in the taxonomic and functional compositions of the microbiome have been implicated in many diseases, including colorectal cancer (CRC) (8–11). Although the association between microbiome alterations and disease processes has been extensively demonstrated, the directionality, as well as the mediators of the host-microbiome interaction, remains unclear.

Diet has been independently associated with both the gut microbiome and CRC. For example, the Western diet (characterized by low fiber and high protein, fat, and sugar) affects gut microbiome composition in humanized mice, whereby mice fed a Western diet have an increased abundance of Firmicutes and a decreased abundance of Bacteroidetes (12, 13). The same Western diet has also long been considered a risk factor for developing CRC (14–16). Using an animal model of CRC, Schulz et al. demonstrated that the high-fat diet (HFD) exacerbates CRC progression; however, treating animals with antibiotics blocks HFD-induced CRC progression (17). This suggests that diet can drive microbiome composition change in the gut as a precursor to CRC development.

Recent studies have found that host genetic variation is correlated with microbiome composition. For example, a polymorphism near the LCT gene, which encodes the lactase enzyme, is associated with an abundance of Bifidobacterium in the gut microbiome, and microbes in the Christensenellaceae family were shown to be heritable, with a higher similarity between monozygotic than dizygotic twins (18–23). Another recent study investigated CRC tumors and identified a correlation between coding mutations in tumors and the composition of the microbial community in the tumor microenvironment (24). Interestingly, in a genetic mutation model of intestinal tumors, germfree animals developed significantly fewer tumors in the small intestine (25). Although the finding is limited to the small intestine, the trend shows that CRC development partially depends on the microbiome. In an animal model of colitis-associated CRC, Uronis et al. showed that germfree mice exhibit normal histology and do not develop tumors, compared to 62% of conventionalized mice that developed tumors (n = 13) (26). These results support an interaction between the microbiome and host genomics that may affect tumor development.

A recent report demonstrated that fecal microRNAs (miRNAs) can shape the composition of the gut microbiome (27), indicating a mechanism by which host cells can regulate the microbial community. In CRC, several miRNAs, such as miR-182, miR-503, and mir-17–92 cluster, can regulate multiple genes and pathways and have been found to promote malignant transformation and disease progression (28–30). Interestingly, studies have also found that microbiome-derived metabolites can change host gene expression, including expression of miRNAs, in the colon (31, 32). Taken together, these results suggest a bi-directional interaction between host cells and microbes, potentially mediated through miRNA activity. However, we still know very little about the role of miRNAs in host-microbiome interactions, especially in the context of CRC. With thousands of unique miRNAs and microbial taxa present in the CRC microenvironment, it is challenging to experimentally study all possible pairwise interactions. Nevertheless, genomic characterization of both miRNA expression and microbial composition in patients with CRC can identify potential interactions between miRNAs and microbes, which can then be used as candidates for functional inspection.

Here, we establish the relationships between miRNA expression and microbiome composition in CRC patients. We sequenced small RNAs and integrated 16S rRNA gene sequencing data from both tumor and normal colon tissues from 44 patients (88 samples total). We explored the correlation between miRNAs and the microbiome through imputing the miRNA functional pathways and microbiome metabolic pathways in silico (see Fig. S1 in the supplemental material). To our knowledge, this is the
first analysis to establish a global relationship between miRNA expression and the microbiome in CRC.

**RESULTS**

**MicroRNAs differentially expressed in tumor tissues.** Before performing differential expression (DE) analysis, we performed extensive quality control of the miRNA data. Our results indicate that miRNA expression is not strongly affected by tumor location, patient gender, patient age, or read coverage and show a clear clustering of miRNA data by tumor and normal samples (Fig. 1; see also Materials and Methods below). To identify small RNAs that are DE between tumor and normal samples, we performed DE analysis using DESeq2 (see Materials and Methods). A total of 76 DE miRNAs were identified, with 55 upregulated and 21 downregulated in tumor tissues compared to normal tissues ($P$ value $< 0.05$ after false-discovery rate [FDR] correction). A full list of DE miRNAs is available in Table S1 in the supplemental material. DE miRNAs with higher expression levels in tumor tissues include miR-182, miR-183, miR-503, and the miR-17–92 cluster miRNAs (Fig. 2; Table S1), all consistent with our previous reports (28, 33). These miRNAs have all been previously shown to contribute to CRC disease progression; for example, miR-182 and miR-503 were found to cooperatively target FBXW7 and contribute to CRC malignant transformation and progression and were also predictive of patient survival (28). The miR-17–92 cluster regulates multiple tumor-suppressive genes in CRC and other cancers (34). In addition, miR-1, miR-133a, and miR-448 (Table S1) were observed at higher levels in normal tissues than in matched tumor tissues, also in agreement with previous reports (33, 35).

**Predicted functions of microbiome taxa correlated with DE miRNAs in tumor samples.** To investigate the relationship between individual miRNAs and the microbiome in CRC tumor samples, we performed correlation analysis using Sparse Correlations for Compositional Data (SparCC). SparCC is developed specifically to analyze compositional genomic survey data, such as 16S rRNA gene sequencing and other types of high-throughput sequencing data (36). Hierarchical clustering revealed several clusters of significantly correlated miRNAs and bacterial taxa (Fig. S5). To further investigate the relationship between miRNAs and the microbiome in CRC, we selected bacteria significantly correlated with the DE miRNAs (Fig. 3A). The correlations clearly show a distinct pattern based on the enrichment of miRNAs, even though the correlation analysis is performed only with tumor samples. We then built a network visualizing the relationship between the top 9 DE miRNAs and their significantly correlated bacteria (Fig. 3B and C). The correlation network shows a highly interconnected relationship between these miRNAs and bacteria. Interestingly, *Blautia*, a genus previously found to have lower abundance in tumor samples, is negatively correlated with miR-20a, miR-21, miR-96, miR-182, miR-183, and miR-7974, which are all miRNAs with high expression levels in tumor tissues. *Blautia* is also positively correlated with the expression level of miR-139, which is an miRNA with high expression levels in normal tissues. Experimental validations are required to investigate the correlations.

We then analyzed the predicted functional composition of the microbiome data and investigated correlations with miRNAs (Fig. S6). We hypothesized that if miRNAs selectively affect the growth of certain bacteria, then bacteria correlated with DE miRNAs are likely to represent functional differences between tumor and normal tissues, while the uncorrelated bacteria would not. Using the PICRUSt v.1.0.0 software, we generated the predicted functional profiles of the correlated and uncorrelated bacteria by assigning pathways and enzymes using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. A total of 25 pathways have significantly altered enrichment (two-sided Wilcoxon signed-rank test with an FDR-corrected $P$ value of $<0.05$) (Fig. S6). Interestingly, several metabolic pathways and signaling pathways, including signal transduction, amino acid metabolism, energy metabolism, and linoleic acid metabolism, were all enriched in the uncorrelated group, suggesting increased metabolic processes in this group. For bacteria significantly correlated with DE miRNAs, however, pathways related to transporters,
peptidoglycan, and terpenoid backbone biosynthesis have significant enrichment. It is worth noting that the predicted metagenome may not accurately represent the function of the microbiome; further validation using quantitative PCR or high-throughput sequencing is required.
Predicted functions of miRNAs correlated with CRC-associated bacteria. To investigate the function of miRNAs correlated with CRC-associated bacteria, we focused on bacterial genera previously associated with CRC, including *Fusobacterium*, *Providencia*, *Bacteroides*, *Akkermansia*, *Roseburia*, *Porphyromonas*, and *Peptostreptococcus* (8, 37–40). We hypothesized that if these bacteria affect CRC through modulating miRNA expression, then miRNAs that are significantly correlated with the bacteria should show enrichment in cancer-related genes and pathways. A list of miRNAs significantly correlated with these bacteria is available in Table S3. We separated these miRNAs into groups with positive correlation and negative correlation with each bacterium independently. Then, using the miRPath v.3 software, we predicted the functions of miRNAs by assigning pathways to the miRNA targets using the KEGG database (Table S4). We visualized the pathways with a q value of \(< 0.01\) (modified Fisher exact test; FDR corrected) in Fig. 4.

Our results show that *Akkermansia* is the only taxon correlated with miRNAs associated with the colorectal cancer pathway. *Fusobacterium*, *Providencia*, and *Roseburia* correlate with miRNAs associated with cancer-related pathways, including the glioma, pancreatic cancer, and renal cell carcinoma pathways and pathways in cancer. Interestingly, glycan-related pathways, including the pathways mucin-type O-glycan biosynthesis, other O-glycan biosynthesis, glycosaminoglycan biosynthesis–heparan sulfate/heparin, and proteoglycans in cancer, have correlations with all bacterial genera analyzed, except for *Akkermansia*. This finding corresponds to a previous study showing that *Fusobacterium nucleatum* infection stimulates mucin secretion in vitro (41). Additionally, *Fusobacterium nucleatum* binds to specific Gal-GalNAc, which is expressed by CRC tumors, through the Fap2 protein (42). *Porphyromonas gingivalis* was shown to induce shedding of a proteoglycan, syndecan-1, in oral epithelial cells (43). However, the role of the bacteria and glycan interaction is not clear in the context of CRC. Cell
signaling pathways previously implicated in CRC, such as the Ras, PI3K/Akt, ErbB, and Hippo pathways, are also correlated with these bacteria (44–47).

**DISCUSSION**

Although there is a known association between gut microbiome composition change and CRC (8–11), the potential mediators of this relationship remain unclear. One potential mediator is host genetics and, specifically, CRC tumor mutational profiles (25, 26). Additional evidence indicates that miRNAs can mediate host-microbiome interactions in patients with CRC (27). Here, we presented the first integrated analysis of miRNA expression and gut microbiome profiles in CRC patients. Our data show a highly interconnected correlation network between miRNA expression and the composition of the microbiome and support the role for miRNAs in mediating host-microbiome interactions.

Active interactions between host and the microbiome in CRC have been previously observed, leading to the proposition that pathogenic “passenger” bacteria colonizing tumor tissue might lead to exacerbated tumor progression (48). In our analysis, we focused on potential passenger bacteria, including *Fusobacterium*, *Providencia*, *Bacteroides*, *Akermansia*, *Roseburia*, *Porphyromonas*, and *Peptostreptococcus*. *Fusobacterium*
includes several pathogenic species and is implicated in dental disease, infections, and CRC (49–51). Similarly, *Providencia* has also been implicated in gastrointestinal infections (8, 52–54). The mechanism of *Fusobacterium* in promoting CRC tumorigenesis and progression has been investigated. It activates the Wnt/β-catenin signaling pathway through FadA protein, which binds to the E-cadherin protein on intestinal epithelial cells (IECs), thus promoting cell proliferation (49). Several mechanisms might explain this observation. One possibility is that bacteria can infiltrate the intestinal epithelial barrier after certain pathogenic bacteria, cleaving the E-cadherin (49, 55). This might lead to an increased inflammatory response in the colon microenvironment, and the inflammation can lead to DNA damage and contribute to disease progression (48, 49). Another potential mechanism is that bacteria can directly cause mutations in IECs through virulence proteins. Several of these virulence proteins were found in *Escherichia coli* and *Helicobacter pylori* (56, 57), and results indicate that these virulence factors may be enriched in the CRC microbiome, especially in *Fusobacterium* and...
Providencia (8). However, it is unclear whether these bacteria produce virulence proteins that can directly cause DNA damage, and further investigation is required to elucidate this mechanism.

The Wnt/β-catenin pathway activation by Fusobacterium can lead to upregulation of numerous genes related to CRC (58–60). One such gene, MYC, is a transcription factor that targets multiple genes related to cell proliferation, the cell cycle, and apoptosis. The miR-17–92 cluster is a known transcriptional target of MYC and has oncogenic properties in several cancer types (30, 34, 61, 62). Interestingly, butyrate, a short-chain fatty acid produced by members of the microbiome, diminishes MYC-induced miR-17–92 overexpression in CRC in vitro through its function as a histone deacetylase inhibitor (31). Studies of CRC have consistently found low fecal butyrate levels as well as a reduced relative abundance of butyrate-producing bacteria, such as members of the Firmicutes phylum (31, 37, 63). One potential explanation is that, in CRC, the DE miRNAs can affect the growth of certain microbes, which eventually outcompete other species and form a biofilm on tumor tissues (27). Indeed, our data show several enriched bacterial nutrient biosynthesis and metabolism pathways in the microbes uncorrelated with DE miRNAs, but not in the correlated group. Interestingly, pathways in bacterial cell motility and secretion are also enriched among uncorrelated bacteria, suggesting that, in addition to promoting bacterial growth, certain miRNAs may be involved in recruiting bacteria to tumor tissues. This may also provide a possible explanation for the observed difference in alpha diversity of tumor microbiomes (8, 64, 65).

In our analysis of the functions of miRNAs correlated with selected bacteria known to have associations with CRC, prion disease, and morphine addiction pathways found to be enriched in our analysis do not immediately seem related to cancer (Fig. 4). Upon further investigation of miRNA target genes in these pathways, we found that several genes included in the pathways may have relevant functions in cancer. For example, mitogen-activated protein kinase (MAPK) is central to cell proliferation and survival, interleukin-6 (IL6) and interleukin-beta (IL1β) are cytokines involved in inflammation, protein kinase A (PKA) is important in regulating nutrient metabolism, Bcl-2-associated X protein (BAX) is a tumor suppressor gene; and prion protein (PRNP) are known to have a significant role in regulating immune cell function (66–68).

A recent study has suggested an additional mechanism affecting host-microbiome interactions that may promote CRC tumorigenesis and progression (42, 69). Abed et al. showed that Fap2 produced by Fusobacterium binds to glycan produced by CRC to attach to the tumor tissue (42). Interestingly, glycan biosynthesis pathways were enriched in targets of the miRNAs correlated with CRC-associated bacteria. The increased glycan production may increase recruitment of certain bacteria, such as Fusobacterium, to the tumor location. This result highlights a novel potential mechanism for miRNAs, through regulating glycan biosynthesis, to attract specific microbes to the tumor microenvironment and thus impact tumor development. Interestingly, the mucin-type O-glycan biosynthesis pathway is enriched in miRNAs positively correlated with Fusobacterium but negatively correlated with Bacteroides and Porphyromonas. This suggests that these bacteria may have different mechanisms of attachment to the mucosal surface due to different abilities to bind to O-glycan (70). Additional studies are required to test the association between Fusobacterium, tumorigenesis, and miRNA-driven glycan production.

It is important to note that our study uses 16S rRNA gene sequencing to characterize microbiome taxonomic composition and computationally predicted pathway composition using PICRUSt v1.0.0 (71). Although this method is widely used, metagenomic shotgun sequencing can be more accurate and informative in understanding the functional makeup of a microbial community. Similarly, to impute miRNA functional profiles, we used an in silico prediction method, miRPath (71, 72). While both of these methods have been rigorously tested and validated with experimental data, the results remain predictions and may not represent the real biological system (71, 72). Another limitation of our approach is that it identifies correlations and not causal relation-
ships. Nevertheless, this approach allows us to generate a microbiome- and miRNA transcriptome-wide characterization of potential interactions, which shed light on potential new mechanisms of host-microbiome interactions.

In addition, we highlight candidates for potentially interacting host miRNAs and microbial taxa, which can be directly validated and explored in model systems (73). For example, mouse models have been extensively used to study host-microbiome interactions in the gut (74), and studies have quantified how microbiome colonization can modulate gene expression in the host gut (75, 76). In addition, in vitro approaches can be useful in dissecting the regulatory effects of the microbiome and in characterizing the effects of variation in individual taxon abundances on gene expression in host cells (77, 78). These studies can validate interactions identified in our current study and shed light on the directionality and causality.

Conclusions. Our analysis, together with evidence from previous studies, suggests that miRNAs likely mediate host-microbiome interaction in CRC. We identify potential novel mechanisms that mediate this interaction and may have a role in CRC tumorigenesis, including a possible role for miRNA-driven glycan production in the recruitment of pathogenic microbial taxa. The interactions identified here might be a direct target for developing therapeutic strategies that can benefit CRC patients. Follow-up studies using model systems are warranted to assess the causal role of individual microbes and miRNAs in CRC.

MATERIALS AND METHODS

Tissue samples. A total of 88 matched tumor and adjacent normal tissues were collected from 44 patients by the University of Minnesota Biological Materials Procurement Network. A detailed description of sample collection was previously published (8). Briefly, all patients provided written, informed consent. All research conform to the Helsinki Declaration and was approved by the University of Minnesota Institutional Review Board, protocol 1310E44403. Tissue pairs were resected concurrently, rinsed with sterile water, flash frozen in liquid nitrogen, and characterized by staff pathologists. Detailed deidentified sample metadata, including age, gender, tumor location, tumor stage, and microsatellite stability (MSS) status, are available in Table S1 in the supplemental material.

16S rRNA sequencing and sequence analysis. The 16S rRNA gene sequencing data were previously published (8). Raw sequences were deposited in the NCBI Sequence Read Archive under project accession number PRJNA284355, and processed data files are available in the work of Burns et al. (8). Briefly, total DNA was extracted from approximately 100 mg of tissue. Tissues were first physically disrupted by placing the tissue in 1 ml of Qiazol lysis solution in a 65°C ultrasonic water bath for 1 to 2 h. The efficiency of this approach was verified by observing high abundances of Gram-positive bacteria across all samples, including those from the phylum Firmicutes. DNA was then purified using an AllPrep nucleic acid extraction kit (Qiagen, Valencia, CA). The V5-V6 region of the 16S rRNA gene was PCR amplified with multiplexing barcodes (79). The bar-coded amplicons were pooled and ligated to Illumina adaptors. Sequencing was performed on a single lane on an Illumina MiSeq instrument (paired-end reads). The forward and reverse read pairs were merged using the USEARCH v7 program fastq_mergepairs, allowing stagger but no mismatches (80). Operational taxonomic units (OTUs) were picked using the closedreference picking script in QIIME v1.7.0 and the Greengenes database (August 2013 release) (81–83). The similarity threshold was set at 97%, reverse read matching was enabled, and reference-based chimera calling was disabled. The unfiltered OTU table used for the analysis is available in Table S2.

MicroRNA sequencing. To prepare samples for small-RNA sequencing, total RNA was extracted using an AllPrep nucleic acid extraction kit (Qiagen, Valencia, CA). RNA was quantified using the RiboGreen fluorometric assay (Thermo Fisher, Waltham, WA). RNA integrity was then measured using a model 2100 Bioanalyzer (Agilent, Santa Clara, CA). Library creation and sequencing were performed by the Mayo Clinic Genome Analysis Core. Briefly, small-RNA libraries were prepared using 1 μg of total RNA per the manufacturer’s instructions for the NEBNext multiplex small-RNA kit (New England Biolabs, Ipswich, MA). After purification of the amplified cDNA constructs, the concentration and size distribution of the PCR products were determined using an Agilent (Santa Clara, CA) Bioanalyzer DNA 1000 chip and Qubit fluorometry (Invitrogen, Carlsbad, CA). Four of the cDNA constructs are pooled, and the 120- to 160-bp miRNA library fraction is selected using Pippin Prep (Sage Science, Beverly, MA). The concentration and size distribution of the completed libraries were determined using an Agilent Bioanalyzer DNA 1000 chip and Qubit fluorometry. Sequencing was performed across 4 lanes on an Illumina HiSeq 2000 instrument (paired end).

MicroRNA sequence data processing and QC. See Fig. S1 for an overview of the data analysis steps. Briefly, quality control (QC) of miRNA sequencing data was performed using FastQC before and after adaptor trimming with Trimmomatic (84). Then, the paired-end reads were assembled using PANDAseq and aligned to the hg38 genome assembly using bowtie2 (85, 86). Finally, the total mature miRNA counts were generated with HTSeq (87). We removed 7 samples (S01, S02, S93, S36, S40, S41, S43) due to a low number of total raw reads (fewer than 500,000 raw reads) from the analysis (Table S1). A previous study showed that a number of miRNA sequencing reads as low as 500,000 provides sufficient coverage for
analysis (88). The remaining 81 samples have between 519,373 and 17,048,093 (median, 6,010,361) reads per library, with an average quality score of c[tex]\geq 37\] in all libraries. Between 66.79\% and 96.14\% (median, 83.53\%) of reads passed adapter trimming (Fig. S2). Of all the reads passing adapter trimming, between 287,356 and 11,102,869 (median, 3,701,487) reads were identified as concordant pairs by PANDAseq. After being mapped to the hg38 genome, between 18,947 and 4,499,805 (median, 859,546) reads were assigned to a total of 2,588 mature miRNAs (Fig. S2). Principal-component analysis (PCA) visually shows a clear separation between tumor and normal samples (Fig. 1A), while tumor location, gender, age, total raw reads, and total mature miRNA reads do not appear to have an impact on the data (Fig. 1B to F). Similarly, PCA plots, including an additional principal component, did not detect clustering based on these factors (Fig. S3). We further performed discriminant analysis of principal components (DAPC) using the adegenet package in R, and it confirmed the existence of separate clusters for tumor and normal samples (\(P < 2.2 \times 10^{-16}\)) (see Fig. S4) but not for gender and tumor locations (\(P > 0.2\) for all comparisons) (Fig. S4) (89). Between 283 and 1,000 (median, 670) miRNAs had coverage over 1 read, and between 134 and 599 (median, 367) miRNAs had coverage over 5 reads (Fig. 1G). Overall, the quality of our sequencing results is on par with those of previous studies and our previous observations (90).

**MicroRNA differential expression and correlation analysis.** We identified differentially expressed (DE) miRNAs between tumor and normal samples using the DESeq2 package (1.10.1) in R (version 3.2.3) (91). Raw miRNA counts were filtered to include miRNAs with \(\geq 1\) read in \(\geq 80\%\) of the samples. The remaining 392 miRNAs were then used for DESeq2 analysis. We define DE miRNAs as showing a fold change of over 1.5, with a false-discovery rate (FDR)-adjusted \(P\) value (\(q\) value) of \(<0.05\). We performed correlation analysis for the tumor samples using Sparse Correlations for Compositional Data (SparCC) at the genus level for bacteria and at the species level (L7). The predicted metagenomes are then generated by following the standard PICRUSt metagenome prediction pipeline. We included miRNAs with significant correlations with CRC-associated genera (Fusobacterium, Providencia, Bacteroides, Akkermansia, Roseburia, Porphyromonas, and Peptostreptococcus) to perform pathway enrichment analysis using miRPath v.3 (72, 92). We generated network visualization of miRNA-microbe using Cytoscape v3.5.1.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00205-17.

- **FIG S1**, PDF file, 0.1 MB.
- **FIG S2**, PDF file, 0.2 MB.
- **FIG S3**, PDF file, 0.1 MB.
- **FIG S4**, PDF file, 0.1 MB.
- **FIG S5**, JPG file, 0.2 MB.
- **FIG S6**, PDF file, 0.1 MB.
- **TABLE S1**, XLSX file, 0.02 MB.
- **TABLE S2**, TXT file, 0.5 MB.
- **TABLE S3**, XLSX file, 0.04 MB.
- **TABLE S4**, XLSX file, 0.02 MB.

**ACKNOWLEDGMENTS**

We thank members of the Blekhman and the Subramanian lab for helpful discussions. We thank Anne Sarver of the Subramanian lab for developing the miRNA analysis pipeline. We also thank the Medical Genome Facility Genome Analysis Core at the Mayo Clinic (Rochester, MN) for performing library prep and sequencing. This work was carried out, in part, using computing resources at the Minnesota Supercomputing Institute.

C.Y. is supported by a Norman Wells Memorial Colorectal Cancer fellowship and a Healthy Foods, Healthy Lives Institute graduate and professional student research grant. This work is supported by the Randy Shaver Cancer Research and Community Fund (R.B.), an institutional research grant (124166-IRG-58-001-55-IRG53) from the American Cancer Society (R.B.), a research fellowship from the Alfred P. Sloan Foundation (R.B.), and award MNP 17.26 from the Minnesota Partnership for Biotechnology and Medical Genomics.
C.Y., S.S., and R.B. developed the concept, C.Y. carried out data analysis with assistance from M.B. C.Y., S.S., and R.B. wrote the manuscript.

REFERENCES

1. Abdollahi-Roodsaz S, Abramson SB, Scher JL. 2016. The metabolic role of the gut microbiota in health and rheumatic disease: mechanisms and interventions. Nat Rev Rheumatol 12:446--455. https://doi.org/10.1038/nrrheum.2016.68.

2. Belkaid Y, Hand TW. 2014. Role of the microbiota in immunity and inflammation. Cell 157:121--141. https://doi.org/10.1016/j.cell.2014.03.011.

3. Candon S, Perez-Arroyo A, Marquet C, Valette F, Foray AP, Pelletier B, Milani C, Ventura M, Bach JF, Chatenoud L. 2015. Antibiotics in early life alter the gut microbiome and increase disease incidence in a spontaneous mouse model of autoimmune insulin-dependent diabetes. PLoS One 10:e0125448. https://doi.org/10.1371/journal.pone.0125448.

4. Vangay P, Ward T, Gerber JS, Knights D. 2015. Antibiotics, pediatric dysbiosis, and disease. Cell Host Microbe 17:553--564. https://doi.org/10.1016/j.chom.2015.04.004.

5. Tremaroli V, Backhed F. 2012. Functional interactions between the gut microbiota and host metabolism. Nature 489:242--249. https://doi.org/10.1038/nature11552.

6. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559--563. https://doi.org/10.1038/nature12820.

7. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Sedor JR, Good- man AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. 2013. The long-term stability of the human gut microbiota. Science 341:1237439. https://doi.org/10.1126/science.1237439.

8. Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. 2015. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Med 7:55. https://doi.org/10.1186/s13059-015-0177-8.

9. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, Ng SC, Tsoi H, Dong Y, Zhang N, He Y, Kang Q, Cao L, Wang K, Zhang J, Liang Q, Yu J, Sung JYJ. 2015. Gut mucosal microbiome across stages of colorectal carcino- genesis. Nat Commun 6:8727. https://doi.org/10.1038/ncomms9727.

10. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, Jia W, Cai S, Zhao L. 2012. Structural segregation of gut microbiota between colorectal patients and healthy volunteers. ISME J 6:320--329. https://doi.org/10.1038/ismej.2011.109.

11. Shen XJ, Raiwls RF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Woodward MJ, and healthy volunteers. JEM 6:320--329. https://doi.org/10.1038/ismej.2011.109.

12. Shen XJ, Raiwls RF, Randall T, Burcal L, Mpande CN, Jenkins N, Jovov B, Woodward MJ, and healthy volunteers. JEM 6:320--329. https://doi.org/10.1038/ismej.2011.109.

13. Ridaura VK, Faith JJ, Rey FE, Cheng JQ, Duncan AE, Kau AL, Griffin NW, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome and increases disease incidence in a spontaneous mouse model of autoimmune insulin-dependent diabetes. PLoS One 10:e0125448. https://doi.org/10.1371/journal.pone.0125448.

14. Vangay P, Ward T, Gerber JS, Knights D. 2015. Antibiotics, pediatric dysbiosis, and disease. Cell Host Microbe 17:553--564. https://doi.org/10.1016/j.chom.2015.04.004.

15. Tremaroli V, Backhed F. 2012. Functional interactions between the gut microbiota and host metabolism. Nature 489:242--249. https://doi.org/10.1038/nature11552.

16. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559--563. https://doi.org/10.1038/nature12820.

17. Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Sedor JR, Good- man AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. 2013. The long-term stability of the human gut microbiota. Science 341:1237439. https://doi.org/10.1126/science.1237439.

18. Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. 2015. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Med 7:55. https://doi.org/10.1186/s13059-015-0177-8.

19. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WKK, Ng SC, Tsoi H, Dong Y, Zhang N, He Y, Kang Q, Cao L, Wang K, Zhang J, Liang Q, Yu J, Sung JYJ. 2015. Gut mucosal microbiome across stages of colorectal carcino- genesis. Nat Commun 6:8727. https://doi.org/10.1038/ncomms9727.
33. Server AL, French AJ, Borrillho PM, Thaynathiy V, Oberg AL, Silverstein KAT, Morlan BW, Riska SM, Boardman LA, Cunningham JM, Subramanian S, Wang L, Smyrk TC, Rodrigues CM, Thibodeau SN, Steer CJ. 2009. Human colon cancer profiles show differential microRNA expression depending on mismatch repair status and are characteristic of undifferentiated proliferative states. BMC Cancer 9:401. https://doi.org/10.1186/1471-2407-9-401.

34. Mogilyansky E, Rigoutsos I. 2013. The miR-17-92 cluster: a comprehensive update on its genetics, functions and increasingly important and numerous roles in health and disease. Cell Death Differ 20:1603–1614. https://doi.org/10.1038/cdd.2013.125.

35. Oberg AL, French AJ, Server AL, Subramanian S, Morlan BW, Riska SM, Borrillho PM, Cunningham JM, Boardman LA, Wang L, Smyrk TC, Asmann Y, Steer CJ, Thibodeau SN. 2011. miRNA expression in colon polyps provides evidence for a multihit model of colon cancer. PLoS One 6:e20465. https://doi.org/10.1371/journal.pone.0020465.

36. Friedman J, Alm JE. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687. https://doi.org/10.1371/journal.pcbi.1002687.

37. Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. 2013. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. PLoS One 8:e70803. https://doi.org/10.1371/journal.pone.0070803.

38. Geng J, Fan H, Tang X, Zhai H, Zhang Z. 2013. Diversified pattern of the human colorectal cancer microbiome. Gut Pathog 5:2. https://doi.org/10.1186/1757-4749-5-2.

39. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, Goedert R, Kalali B, Gerhard M, Sartori AA, Lopes M, Müller A. 2011. Carcinogenic bacterial pathogen Helicobacter pylori triggers DNA double-strand breaks and a DNA damage response in its host cells. Proc Natl Acad Sci U S A 108:14944–14949. https://doi.org/10.1073/pnas.100959108.

40. Yu J, Feng Q, Wong SH, Zhang D, Liang Q, Zhang M, Franco A, Sears CL. 2007. Bacteroides fragilis regulates epithelial cell shedding and gamma-secretase-dependent E-cadherin cleavage. J Cell Sci 120:1944–1952. https://doi.org/10.1242/jcs.03455.

41. Noelens KJ, Steger M, Hartung ML, Yau TO, Lan Z, Jia H, Jiang J, Wang J. 2017. Metagenomic analysis of fecal microbiome as a tool towards targeted molecular biomarkers for colorectal cancer. Gut 66:70–77. https://doi.org/10.1136/gutjnl-2015-309800.

42. Dharmiani P, Strauss J, Ambrose C, Allen-Vercoe E, Chadee K. 2011. Fusobacterium nucleatum infection of colonic cells stimulates MUC2 mucin and tumor necrosis factor alpha. Infect Immun 79:2597–2607. https://doi.org/10.1128/IAI.05118-11.

43. Abed J, Emmar DJ, Zamir G, Faroja M, Almogy G, Grenov A, Sol A, Naor P, Gepstein E, Bockron J, Meul A, Chassu S, Manson AL, Earl AM, Ou N, Brennan CA, Garrett WS, Bachrach G. 2016. Fap2 mediates Fusobacterium nucleatum infection of colonic cells stimulates MUC2 mucin and tumor necrosis factor alpha. Infect Immun 84:1379–1386. https://doi.org/10.1128/IAI.00137-16.

44. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Zhang M, Franco A, Sears CL. 2007. Bacteroides fragilis colonizes the gut microbiota of patients with colorectal cancer and promotes colorectal carcinogenesis. Science 313:848–851. https://doi.org/10.1126/science.1127059.

45. Yokoyama T, Hori H, Kato T, Katoh S, Yamada M, Kurosawa Y, Ito Y, Shimizu S, Ito T, Kiyota T, Nakamura T, et al. 2014. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. Cell Host Microbe 14:195–206. https://doi.org/10.1016/j.chom.2013.07.012.

46. Tjalsma H, Boleij A, Marchesi JR, Duthile B. 2012. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol 10:575–582. https://doi.org/10.1038/nrmicro2819.

47. Raymond H, Bhatia K, Wang HG. 2003. Bax plays a pivotal role in thapsigargin-induced apoptosis of human colon cancer HCT116 cells by...
controlling Smac/Diablo and Omi/HtrA2 release from mitochondria. Cancer Res 63:1483–1489.

69. Wyndaele E, Verbeke F, D’Hondt M, Hendrix A, Van De Wiele C, Burvenich C, Peremans K, De Wever O, Bracke M, De Spiegeleer B. 2015. Crosstalk between the microbiome and cancer cells by quorum sensing peptides. Peptides 64:40–48. https://doi.org/10.1016/j.peptides.2014.12.009.

70. Tailford LE, Crost EH, Kavanaugh D, Juge N. 2015. Mucin glycan foraging in the human gut microbiome. Front Genet 6:811. https://doi.org/10.3389/fgene.2015.00081.

71. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Cai L, Ye L, Tong AHY, Lok S, Zhang T. 2013. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. mSystems 1:e00067-16. https://doi.org/10.1128/mSystems.00067-16.

72. Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, Dalamagas T, Hatzigeorgiou AG. 2015. Diana-miRPath v3.0: deciphering microRNA function with experimental support. Nucleic Acids Res 43:W460–W466. https://doi.org/10.1093/nar/gkv403.

73. Luca F, Kupfer SS, Knights D, Khoruts A, Blekhman R. 2018. Functional genomics of host-microbiome interactions in humans. Trends Genet 34:30–40. https://doi.org/10.1016/j.tig.2017.10.001.

74. Leamy LJ, Kelly SA, Nietfeldt J, Legge RM, Ma F, Hua K, Sinha R, Peterson DA, Walter J, Benson AK, Pomp D. 2014. Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. Genome Biol 15:552. https://doi.org/10.1186/s13059-014-0552-6.

75. Davison JM, Lickwar CR, Song L, Breton G, Crawford GE, Rawls JF. 2017. Microbiota regulate intestinal epithelial gene expression by suppressing the transcription factor hepatocyte nuclear factor 4 alpha. Genome Res 27:1195–1206. https://doi.org/10.1101/gr.220111.116.

76. Camp JG, Frank CL, Lickwar CR, Guturu H, Rube T, Wenger AM, Chen J, Bejerano G, Crawford GE, Rawls JF. 2014. Microbiota modulate transcription in the intestinal epithelium without remodeling the accessible chromatin landscape. Genome Res 24:1504–1516. https://doi.org/10.1101/gr.163945.113.

77. Richards AL, Burns MB, Alazizi A, Barreiro LB, Pique-Regi R, Blekhman R, Luca F. 2016. Genetic and transcriptional analysis of human host response to healthy gut microbiota. mSystems 1:e00067-16. https://doi.org/10.1128/mSystems.00067-16.

78. Richards A, Muehlbauer AL, Alazizi A, Burns MB, Gould TJ, Cascardo C, Pique-Regi R, Blekhman R, Luca F. 2017. Variation in gut microbiota composition impacts host gene expression. BioRxiv https://www.biorxiv.org/content/early/2017/10/27/210294.

79. Cai L, Ye L, Tong AHY, Lok S, Zhang T. 2013. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS One 8:e53649. https://doi.org/10.1371/journal.pone.0053649.

80. Edgar RC. 2010. Search and clustering of Read faster than BLAST. Bioinformatics 26:2460 –2461. https://doi.org/10.1093/bioinformatics/btq461.

81. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072. https://doi.org/10.1128/AEM.03006-05.

82. Navas-Molina JA, Peralta-Sánchez JM, González A, McMurdie PJ, Vázquez-Baeza Y, Xu Z, Ursell WK, Lauber C, Zhou H, Song SJ, Hutyra L, Ackermann GL, Berg-Lyons D, Holmes S, Caporaso JG, Knights R. 2013. Advancing our understanding of the human microbiome using QIIME. Methods Enzymol 513:371–444. https://doi.org/10.1016/B978-0-12-407863-5.00019-8.

83. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EC, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303.

84. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

85. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.

86. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. 2012. Functional genomics of host-microbiome interactions in humans. Trends Genet 27:1195–1206. https://doi.org/10.1016/j.tig.2011.11.016.

87. Anders S, Pyl PT, Huber W. 2015. HTSeq—a python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169. https://doi.org/10.1093/bioinformatics/btu638.

88. Metpally RPR, Nasser S, Malenica I, Courtright A, Carlson E, Ghaffari L, Villa S, Tembe W, Van Keuren-Jensen K. 2013. Comparison of analysis tools for miRNA high throughput sequencing using nerve crush as a model. Front Genet 4:20. https://doi.org/10.3389/fgene.2013.00020.

89. Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94.

90. Lopez JP, DiIorio A, Crouzeau C, Fiori LM, Laboissiere S, Guillet I, Fontaine J, Ragoussis J, Benes V, Turecki G, Ernst C. 2015. Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing. BMC Med Genomics 8:35. https://doi.org/10.1186/s12920-015-0109-x.

91. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550–550. https://doi.org/10.1186/s13059-014-0550-8.

92. Lewis BP, Burge CB, Bartel DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120:15–20. https://doi.org/10.1016/j.cell.2004.12.035.