5-2022

Investigation of Microbe and Host Tissue Interactions Contributing to the Pathogenesis of Colorectal Cancer

Ryan Chapman
rchapman@unomaha.edu

Dhundy Bastola
University of Nebraska at Omaha

Investigation of Microbe and Host Tissue Interactions Contributing to the Pathogenesis of Colorectal Cancer

Ryan Chapman, Dhundy Bastola

***Abstract*—Colorectal cancer (CRC) is one of the leading causes of cancer-related death worldwide. The pathogenesis of this disease can fall under broad categories; however, the specific precursory mechanism of CRC pathogenesis is still unknown. Dysregulations of the gut microbiome have been identified in the CRC tissue environment. Additionally, CRC tissue gene expression has been observed to differ from that of healthy tissue. Despite these noticeable changes, few studies have directly compared the microorganism composition to the gene expression of CRC tissue. Doing so may identify whether the differentially abundant microorganisms influence the changes in gene expression. The goal of this study is to utilize bioinformatics to identify enriched microorganism populations as well as differentially expressed genes within CRC tissue samples to explore the potential interactions leading to disease pathogenesis. The results of this study found that populations of toxin-producing bacteria as well as oral bacteria are present within the CRC tissue environment. Additionally, changes in gene expression suggest that these bacteria may be causing an inflammatory immune response, which could catalyze tumorigenesis.**

I. Introduction

Colorectal cancer (CRC) is a highly prevalent subtype of cancer and is the second and third most diagnosed subtype in women and men, respectively [1]. Each year, around 600,000 individuals die from CRC, making it the most common cause of cancer-related death worldwide [2]. CRC can arise through chronic inflammation of the colon, known as colitis-associated colorectal cancer (CAC), or through a spontaneous mutation in colon cells, known as sporadic colorectal cancer (SCC) [3]. A mutation in the APC gene, a known tumor suppressor gene, has been identified as the cause of SCC, and TP53, another tumor suppressor gene, has been found to be commonly mutated in CAC [4], [5]. Though the specific genetic mutations involved in the pathogenesis of CRC are known, precursory events that lead to these mutations are still being discovered. Numerous factors have been found to play a role in the development of CRC including diet and intrinsic factors dependent on the affected individual [6]. Further elucidating these factors can improve the ability to diagnose, treat, and prevent CRC. Additionally, these mechanisms may also provide insight into the pathogenesis of other cancer subtypes.

A. Microbiome Dysregulation in CRC

Recently, the effect of the gut microbiome on the tumorigenesis of CRC has been investigated. It is known that the human microbiome is important in maintaining a healthy physiology [6]. Additionally, dysregulation of the gut microbiome has been found to play a role in allergy, obesity, and type 2 diabetes as well as diseases such as irritable bowel syndrome (IBS) and gastric cancer [7], [8]. In the case of gastric cancer, a direct link between the microbe *Helicobacter pylori* and the development of the disease has been observed [8]. It is hypothesized that changes in the microbiome and the mucosa layer of the colon affect one another, likely playing a role in the development of CRC [3]. Several organisms have been found to be enriched in the CRC microbiome.

1) Fusobacterium Nucleatum: Studies by Kostic et al., Viljeon et al., Gao et al., Castellarin et al., Wang et al, Sakamoto et al., Périchon et al, and Yu et al. found higher levels of bacteria belonging to the genus *Fusobacterium* present in colorectal cancer samples, specifically the species *Fusobacterium nucleatum* [9]–[16]. Outside of these instances, *F. nucleatum* has been found to be associated with various forms of tissue inflammation, including in GI-related diseases [12]. Kostic et al., also found that *F. nucleatum* accelerated the growth of tumors in mice and caused a pro-inflammatory environment [9]. Additionally, Kostic et al. and Castellarin et al. note that *F. nucleatum* is most abundant in the oral cavity suggesting an association between oral and gut microbiome dysbiosis [9], [12]. Further supporting this hypothesis, Yu et al. identified several other oral pathogens, in addition to *F. nucleatum*, that were overrepresented in CRC samples including *Parvimonas micra*, *Peptostreptococcus stomatis*, and *Solobacterium moorei* [15].

2) Escherichia coli: Viljeon et al., Buc et al., Arthur et al., and Bonnet et al. found a higher prevalence of *Escherichia coli* in colorectal cancer patients [10], [17]–[19]. Particularly, cyclomodulin-positive *E. coli* was noted by Viljeon et al., Buc et al., and Bonnet et al. Cyclomodulins are toxins that have been attributed to tumorigenesis and genotoxicity [10], [19]. Specifically, it is thought that changes to the receptors of the mucosal lining allow for the colonization of cyclomodulin-positive *E. coli* [17].

3) Streptococcus bovis and galloyticus: Streptococcus bovis and its subspecies have also been found to be abundant in the microbiome of CRC patients by Wang et al., Boleij et al., Abdulamir et al., and Gupta et al. [13], [20]–[22] The subtype *Streptococcus galloyticus* is noted to be highly prevalent [16], [20], [21]. *S. bovis* is known to promote cell proliferation and carcinogenesis, pointing to its relevance in CRC progression [13].

4) Helicobacter pylori: As previously mentioned, *Helicobacter pylori* is a known cause of gastric cancer. However, it

has also shown a potential association with CRC [23]. Several pro-tumor effects are caused by infection with *H. pylori*, with changes to the mucosa of digestive organs being a primary factor in its contribution to tumorigenesis [24]. Though, the organism's role and overrepresentation in CRC samples have been disputed, and a definitive relationship still remains undetermined [25]. A meta-analysis of 14 studies, conducted by Zumkeller et al., reported that *H. pylori* infection and CRC had a 1.4% odds ratio [26]. If an association is found, changes in the mucosa caused by *H. pylori* could provide an explanation for other changes in the gut microbiome composition in CRC patients.

5) Bacteroides fragilis: The presence of *Bacteroides fragilis* was found to be upregulated in fecal samples of CRC patients by Viljoen et al., Wang et al., Périchon et al, and Toprak et al. [10], [13], [16], [27]. *B. fragilis* is a normal inhabitant of the intestinal microbiome, according to Hajishengalis et al. [28]. However, a subgroup of this bacteria, known as enterotoxigenic *B. fragilis* (ETBF), has been found to be associated with inflammatory disease of the digestive tract [29]. Toprak et al., specifically, found EBTF to be more abundant in CRC patient samples [27]. ETBF produces a toxin that has been suggested to lead to inflammation in the tissue. A study by Wu et al. found that EBTF colonization led to tumorigenesis in mice through a TH17-related immune response [30]. Specifically, CD4+ T-cells were related to tumor growth in the mice, suggesting an interaction between ETBF and the host immune system.

6) Enterococcus faecalis: Wang et al. and Balamurugan et al. found the bacterium *Enterococcus faecalis* to be more prevalent in fecal samples of CRC patients [13], [31]. Though, the involvement of this bacteria in the regulation of the gut microbiome is somewhat controversial. *E. faecalis* has been reported to have some probiotic effects in the treatment of bronchitis or the regulation of GI-related dysmotility [32], [33]. However, *E. faecalis* has also been found to produce superoxide radicals, which can have damaging effects on cellular DNA, providing a mechanism for tumorigenesis [34]. A study by Williamson et al. found that coculturing of collagenase-producing *E. faecalis* and murine colorectal carcinoma cells leads to increased invasion and migration of the cells [35]. They hypothesize that the host tissue and bacteria interact to degrade the stroma and increase the invasiveness of the tumor.

B. Gene Expression changes in CRC

Similar to the wide variety of mechanistic possibilities found relating to microorganisms in the gut microbiome, investigation of host tissue gene expression has suggested a multitude of potential transcriptomic mechanisms and alterations in CRC. A study by Guo et al. utilized microarray data to identify differentially expressed genes and pathways in CRC tissue [36]. Notably, genes relating to both cell-cycle regulation and inflammation were identified. The genes *CDK1*, *CCNB1*, *CENPE*, *KIF20A*, *CCNA2*, and MAD2L1 were the top differentially expressed genes relating to the cell-cycle [36]. Additionally, the genes *CXCL1*, *CXCL2*, *CXCL6*, *CXCL8*, and *CXCL12*, relating to the CXCL family of ligands, were upregulated, implicating inflammatory responses [36]. Similarly, Yan et al. found *CDK1*, *CCNB1*, and *MAD2L1* to be upregulated in CRC tumor tissues [37]. Along with these genes, *BUB1B*, *SOX4*, *MYC*, and *CCND1* were also upregulated. A study by Kuo et al. identified the genes *FOS*, *FN1*, *PPP1CC*, and *CYP2B6* as being central to protein-protein interaction networks containing differentially expressed genes in CRC tumors [38]. These hub genes implicate processes such as cell proliferation, transcription regulation, cell cycle, cytoskeleton reorganization, and inflammation in the development of CRC, similar to the results of Guo et al. and Yan et al. [38].

While both the transcriptome and the microbiome of CRC have been studied previously, few studies have considered a connection between the two, despite implications in experimental findings. Specifically, alterations in genes relating to metabolism, inflammation, and cytoskeletal reorganization are particularly interesting as these processes may be directly affected by microbiome dysbiosis. Our hypothesis is that dysregulation of the colon microbiome leads to transcriptional changes in colon tissue, in turn leading to cancer. The goal of this study is to investigate the differences in both the colon microbiome and the colon transcriptome between CRC tumor tissue and adjacent healthy tissue samples. Based on data from transcriptomic and metagenomic analyses, the functions of the identified microbes can be compared to the expression of the CRC tissue to hopefully elucidate a mechanism for dysbiosis-related tumorigenesis. Characterization of this interaction will be important in developing newer diagnostic and preventative methods for CRC.

II. Methods

A. Data Availability

Metagenomic shotgun sequencing data published in a study by Debesa-Tur et al. was used to conduct a metagenomic analysis (EBI Accession PRJEB34333) [39]. The dataset consists of 98 tissue samples taken from 49 patients with CRC. Patients in the study were not consistent in disease stage. Each patient in the study had one sample of tissue removed from the tumor and another sample of tissue located in a distant region of the colon removed to serve as a negative control. DNA was extracted from the tissue samples and sequenced on an Illumina NextSeq500 using 150bp paired-end read technology.

RNA-sequencing data collected by Wu et al. (SRA Accession PRJNA387172) was used to assemble the transcriptomic profile of CRC tissue [40]. This study included 104 tissue samples from CRC patients at different disease stages. From each patient, RNA was taken from tumor tissue and adjacent healthy tissue with the adjacent healthy tissue serving as the negative control. RNA was sequenced on an Illumina NextSeq500.

B. Metagenomics Analysis

Quality control and trimming of the metagenomic samples was conducted using FastQC and Trimmomatic version 0.11.9 and 0.39, respectively [41], [42]. A sliding window approach was used to remove regions of the reads with a

PHRED quality of less than 20, and adapter sequences were removed from the reads. Finally, any reads that fell below a minimum length threshold of 50bp were removed. Next, human DNA contamination was removed by aligning of the reads to the Human genome using bowtie2 version 2.4.5 [43]. The GRCh38.p14 release of the human genome was used as the reference (NCBI RefSeq Accession: GCF_000001405.40). Taxonomic classification and metabolic pathway analysis of the metagenome was performed using the tools MetaPhlAn3 version 3.0.14 and HUMAnN3 version 3.0.0, respectively [44]. These tools utilize the ChocoPhlAn3 database which contains unique marker genes from over 90,000 different bacterial and archaeal organisms. Circular visualizations of the sample microbiome compositions were created using the export2graphlan and GraPhlAn tools, versions 0.22 and 1.1.3 respectively [45], [46].

C. Transcriptomic Analysis

To accommodate the large size of the RNA-sequencing dataset, given the time limitations for the project, 3 subsets containing 5 random sample pairings were analyzed. Sample pairings within each subset were randomly selected, and no sample pairing was analyzed twice. The published data had already been pre-trimmed; thus, no quality control was required. Transcript alignment and quantification was performed using Kallisto version 0.48.0, a combined pseudo-alignment and quantification tool with high accuracy and robust isoform identification [47]. The most updated ensemble release of the Human Genome was used as a reference (Ensembl Accession: GCA_000001405.28).

Next, differential expression analysis of each subset was conducted using DESeq2 version 1.30.1 [48]. To import the transcripts into the DESeq2 package, the tximport and Genomic Features packages (versions 1.18.0 and 1.42.3, respectively) were used to map transcripts to genes based on the ensembl assembly [49], [50]. Upregulated genes were denoted as genes that had a log2(Fold Change) greater than 0 and an adjusted p-value less than 0.05. Similarly, downregulated genes were denoted as genes that had log2(Fold Change) less than 0 and an adjusted p-value less than 0.05. The differential expression analysis was performed in relation to the tumor samples; thus, upregulated and downregulated genes refer to those in tumor samples. Functional annotation and pathway analyses were performed on the differentially upregulated and downregulated genes using the Biomartr and pathfindR packages, versions 2.46.3 and 1.6.3 respectively [51]–[53]. The EnhancedVolcano package, version 1.8.0, was used to generate volcano plots representing the distribution of differentially expressed genes within each subset [54]. Differentially upregulated and downregulated genes between the subsets were also used to identify a set of common upregulated and downregulated genes, overall.

All Code is publicly available at https://github.com/rchapman2022/BIOI4980-Senior-Project-Chapman

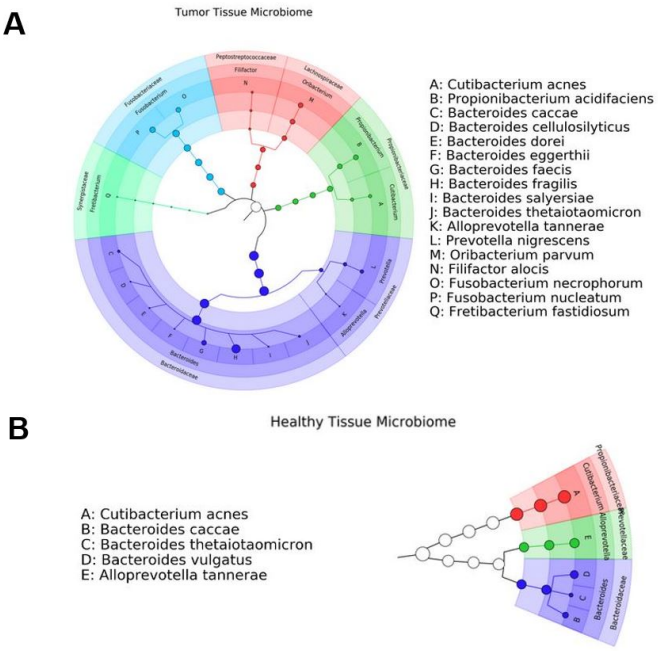

Fig. 1. Circular phylogenetic trees showing the abundance and relationship of microorganisms present in **A.** Tumor tissue **B.** Adjacent Healthy Tissue.

TABLE I
METABOLIC PATHWAYS IDENTIFIED IN CRC TUMOR TISSUE METAGENOMIC SAMPLES

Metabolic Pathway	**Species Associated**
coenzyme A biosynthesis I	None
dTDP-β-L-rhamnose biosynthesis	*Bacteroides cellulosilyticus*, *Bacteroides fragilis*
adenosine ribonucleotides de novo biosynthesis	*Bacteroides fragilis*, *Bacteroides thetaiotaomicron*
dTDP-N-acetylthomosamine biosynthesis	*Bacteroides fragilis*

III. RESULTS

A. Metagenomic Taxonomic Classification

Taxonomic classification using the MetaPhlAn3 software identified a diverse population of microorganisms present in the CRC tumor tissue compared to adjacent healthy tissue (**Figure 1A and 1B**). A larger population of organisms in the genus *Bacteroides* are observed in the tumor tissue microbiome, including *B. fragilis*. Additionally, organisms of the genus *Fusobacterium*, including *F. nucleatum*, are found exclusively in the tumor tissue microbiome.

B. Microbiome Metabolic Pathway Analysis

Several metabolic pathways were identified by HUMAnN3 to be enriched in CRC tumor tissue compared to healthy tissue (**Table 1**). Pathways identified were primarily associated with species in the genus *Bacteroides*.

C. Host Tissue Differential Expression Analysis

Differential expression was performed separately for each subset analyzed, and the top differentially upregulated genes (**Figure 2A-C**) and downregulated genes (**Figure 3A-C**) for each subset were identified. Lists of common differentially upregulated and downregulated genes were then compiled through the overlap of these three subsets (**Tables 2 and 3**).

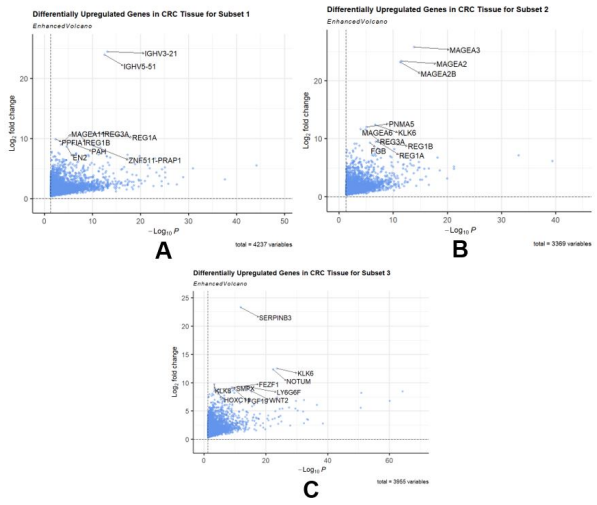

Fig. 2. Distribution of differentially upregulated genes in CRC tissue for **A.** Subset 1, **B.** Subset 2, and **C.** Subset 3. The top 10 differentially upregulated genes based on Log2(Fold Change) are labeled for each subset.

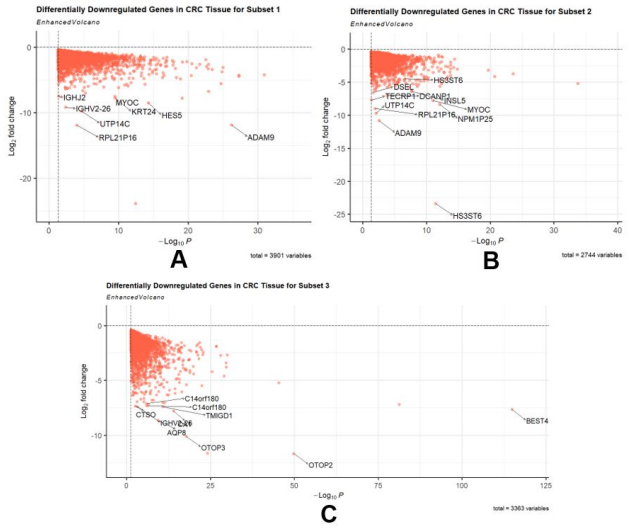

Fig. 3. Distribution of differentially downregulated genes in CRC tissue for **A.** Subset 1, **B.** Subset 2, and **C.** Subset 3. The top 10 differentially downregulated genes based on Log2(Fold Change) are labeled for each subset.

None of the subsets have any top 10 differentially upregulated or downregulated genes in common.

D. Host Tissue Pathway Analysis

Based on the differentially expressed genes in each subset, a pathway analysis was performed. The top 15 enriched pathways were identified for each subset (**Figure 4A-C**). Between the three subsets, there is diversity in the pathways enriched, though there are common pathways. These include Oxidative Phosphorylation, Cell Cycle, Chemical Carcinogenesis – reactive oxygen species, Nucleocytoplasmic Transport, Prion disease, Ribosome biogenesis, Thermogenesis, Ubiquitin mediated proteolysis, and others.

TABLE II
TOP 10 COMMON UPREGULATED GENES AMONG THE THREE SUBSETS

	Log2(Fold Change)			
Gene Name	**Subset 1**	**Subset 2**	**Subset 3**	**Average**
KLK6	8.437379	12.39904	12.50286	11.11309
REG1B	10.68007	9.459354	8.442904	9.527442
REG1A	10.94704	9.544892	8.024177	9.505371
CLDN2	7.279824	7.20779	8.32882	7.605478
FOXQ1	6.806735	7.142049	8.21668	7.388488
WNT2	7.187942	5.361155	8.861862	7.136986
PAH	9.208053	5.367758	6.700507	7.092106
EN2	8.695967	5.813924	6.573458	7.027783
FGB	5.131777	9.302612	6.57919	7.004526
DPEP1	6.680126	6.792719	7.410364	6.96107

Genes were sorted by Average Log2(Fold Change) between the three subsets. Genes without a gene name were not included. All individual Log2(Fold Change) values are statistically significant (** adjusted p-value < 0.05).

TABLE III
TOP 10 COMMON DOWNREGULATED GENES AMONG THE THREE SUBSETS

	Log2(Fold Change)			
Gene Name	**Subset 1**	**Subset 2**	**Subset 3**	**Average**
OTOP2	-6.62933	-5.82574	-11.6584	-8.03782
MYOC	-7.49499	-7.77593	-4.78106	-6.68399
AQP8	-5.93522	-5.25344	-8.63064	-6.60643
NPM1P25	-6.19772	-8.33165	-4.82793	-6.45243
BEST4	-5.67734	-5.61269	-7.65452	-6.31485
CA7	-5.67219	-5.66319	-7.20415	-6.17984
CA1	-4.80864	-4.94489	-7.78172	-5.84508
MT1M	-5.74548	-5.00503	-6.30886	-5.68645
PI16	-5.54374	-5.18076	-5.75746	-5.49398
GUCA2B	-4.93464	-4.3947	-7.05959	-5.46297

Genes were sorted by Average Log2(Fold Change) between the three subsets. Genes without a gene name were not included. All individual Log2(Fold Change) values are statistically significant (** adjusted p-value < 0.05).

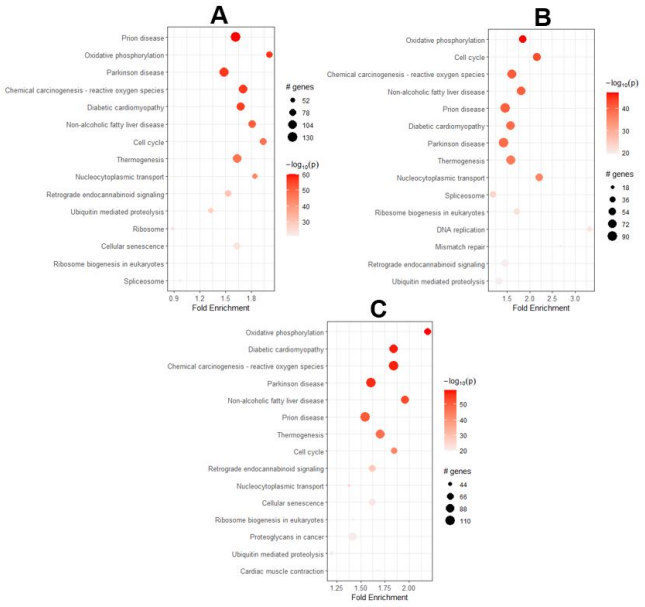

Fig. 4. Enrichment plots depicting pathways found to be enriched based on differentially expressed genes in **A.** Subset 1 **B.** Subset 2 **C.** Subset 3. Differentially expressed genes with an adjusted p-value less than 0.05 were included in the pathway analysis.

IV. Discussion

Consistent with the literature, this study identifies differences in both the microbiome composition and host tissue gene expression between CRC tumor and normal tissue. Moreover, the CRC tissue and healthy tissue samples analyzed are paired suggesting localized changes in microbiome composition within the same patient. Regional differences in microbiome composition have been reported in the literature, as well. A study by Amos et al. found that, amongst different inflammatory bowel diseases, the microbiome changed depending on disease location [55]. As well, a study by Lkhagva et al. observed diversity throughout the entire GI tract of mice [56]. It may be the case that GI-related diseases, including CRC, differ in microbiome-related pathogenesis based on disease location, and future studies could benefit from investigating location-specific microbiome variations in CRC tissues.

The CRC tissue samples were found to have a much greater diversity of microorganisms than the adjacent healthy tissue. Amongst these microorganisms are several which have been found to be associated with CRC as well as other diseases. *F. nucleatum* has been found previously to be associated with CRC as well as other diseases such as periodontitis [57]. The bacterium does not produce any known toxins, although, in the case of periodontitis, it is known to aid other organisms in exacerbating the disease [57]. Additionally, it is known to be highly invasive to host cells [57]. Perhaps, *F. nucleatum* is involved in recruiting and developing the microorganism population capable of tumorigenesis.

The idea of *F. nucleatum* recruiting pathogenic organisms is further supported by the presence of additional oral microorganisms. *Propionibacterium acidifaciens*, another oral pathogen in the CRC tumor samples, has been found in one study to cause tooth decay [58]. Another bacterium, *Oribacterium parvum*, was present in dental plaques and is known to produce different lipids [59]. *Provotella nigrescens* was also identified in the CRC tissue samples. In addition to being found in the oral cavity, experiments have suggested that *P. nigrescens* is associated with an inflammatory response in cystic fibrosis tissue [60], [61]. *Filifactor alocis* is a bacterium associated with periodontal disease and was also found in the CRC tissue samples [62]. Interestingly, this bacterium is able to survive in environments with oxidative stress and was found to elicit an inflammatory response from macrophages [63]. Taken together, it appears to *F. nucleatum* may recruit a variety of microorganisms from the oral microbiome that interact with the host tissue to promote tumorigenesis.

B. fragilis was also identified, along with several other organisms of the genus *Bacteroides*, in the CRC tumor tissue sample. Three of these species were identified in the adjacent healthy tissue. This finding is not surprising, as the *Bacteroides* genus is abundant in the gut microbiome and provides many beneficial functions [64]. Though, *B. fragilis* is known to be associated with CRC, specifically Enterotoxic *B. fragilis* (EBTF) [64]. This study was not able to differentiate EBTF, though it is notable that *B. fragilis* was only found in the CRC tissue samples. Additionally, two of the microorganism metabolic pathways identified in the CRC tissue samples, dTDP-β-L-rhamnose biosynthesis and dTDP-N-acetylthomosamine biosynthesis, were related to *B. fragilis*. Both molecules are carbohydrates, which may allow *B. fragilis* to evade the host immune system, as *Bacteroides* are able to modify their surface polysaccharides [65].

Results of the gene expression analysis suggest that the CRC tumor tissue is reacting to some sort of physical and immunological stress, potentially caused by the microbiome. Among the differentially upregulated genes, several have been found in literature to be upregulated in different types of cancer and inflammatory bowel diseases including *KLK1*, *REG1A*, *REG1B*, *WNT2*, and *EN2* [66]–[70]. Similarly, the top downregulated genes appear to be general markers associated with the prevention of cancer progression and inflammatory bowel disease, making their downregulation significant. These genes include *OTOP2*, *APQ8*, *CA7*, *MT1M*, and *GUCA2B* [71]–[75]. Among these genes, *GUCA2B* has been recently identified as a hub-gene for CRC pathogenesis, with miRNA targeting this gene found to be upregulated in CRC [75]. The gene *BEST4* was downregulated in these results, but identified in a study by He et al. to be upregulated in CRC [76]. The study notes that *BEST4* is important to the activation of the *Akt* signaling pathway involved in cell proliferation [76]. Thus, the downregulation of *BEST4* suggests unregulated cellular proliferation. Another gene, *CA1*, was identified in a study by Zheng et al. to be associated with the calcification of breast cancer [77]. Although it was noted as a potential oncogene in breast cancer, its downregulation may suggest that the tumor tissue is trying to prevent additional environmental stress.

Several genes, in addition to being associated with cancer progression, were related to immune cell recruitment and response. One of these genes, *FOXQ1*, is associated with the recruitment of tumor-associated macrophages [78]. While this type of macrophage has been associated with several types of cancers, including gastric cancer, the mechanism that initiates the promotion these cells is unknown [78]. Perhaps, microbiome dysbiosis plays a role in initializing the recruitment of these immune cells. Additionally, one of the previously mentioned bacterium in CRC tissue, *F. alocis*, was found to elicit a macrophage-related immune response [63]. Additionally, *DPEP1*, found to be differentially upregulated in all three samples, has been associated with neutrophil recruitment in both the liver and the lungs [79]. Like the macrophage responses related to *FOXQ1*, dysbiosis may lead to the recruitment of neutrophils, in turn, leading to the inflammatory and tumorigenic environment indicated by gene expression.

Most interesting is the upregulation of the *FGB* gene, which has been associated with the presence of oral pathogens in the aorta of mice [80]. *F. nucleatum* was among the pathogens identified in the aorta, however, *FGB* was found to be downregulated in this instance [80]. *FGB* is related to blood clotting, which suggests that the tissue may be directly reacting to damage caused by pathogens or general inflammation. The idea of tissue damage is further suggested by the upregulation of the *CLDN2* gene, which is associated with epithelial tight junctions [81]. It is possible that the tissue is responding

to the invasiveness of the microorganisms in the area by strengthening its epithelial barrier.

Though, among the upregulated and downregulated genes between subsets, the top genes in each subset were noticeably absent in the others. While this may appear troubling, this does support the fact that CRC tumors are heterogeneous in their gene expression, making a single unified picture difficult to obtain. Another explanation for this heterogeneity is the fact that the tumor samples came from patients in various stages of the disease. The heterogeneity of the tumors does, to a certain extent, elicit some support for the genes identified between all three subsets as being common to tumorigenesis. Similarly, the enriched pathways showed some degree of homogeneity as well as a relation to the general progression of cancer. However, enrichment of the chemical carcinogenesis - reactive oxygen species pathway supports the overall suggested mechanisms. Reactive oxygen species are an important factor in inflammatory diseases and are not effective against some oral pathogens found in the CRC tissue samples including, *F. alocis* [82].

V. Conclusion

Altogether, the results of this analysis suggest microbiome dysbiosis is occurring in CRC tissue. This dysregulation was characterized as having an increase in oral pathogens, possibly induced by the bacterium *F. nucleatum*, as well as an increase in a previously identified bacterium related to CRC, *B. fragilis*. Investigation into gene expression provided potential mechanisms by which these oral pathogens can influence tumorigenesis: through the triggering of an inflammatory immune response as well as tumor tissue damage response. Although no discrete connections could be made, these suggestions do hold some merit and should be further investigated in subsequent studies.

VI. Limitations

There are several limitations to this study that contribute to the inconclusively of the data. These factors would have been controlled for or corrected had this study not been limited by the timeline of an undergraduate thesis. One major limitation is the fact that the metagenomic and transcriptomic datasets were not taken from the same patient and the fact that disease stage was not controlled for. Ideally, both metagenomic DNA and RNA samples should have been taken from the same patient. However, for the purposes of this study, the fact that both studies consisted of paired tumor and healthy tissue samples was sufficient to make a connection.

Another limitation of this study was the inability to analyze the entire transcriptomic dataset. The server available for this analysis is utilized by the UNO Bioinformatics department and was not large enough to store all of the intermediate data and results. To make up for this, three subsets of 5 paired samples were analyzed and the results were compared.

One final limitation is the quality of the metagenomics dataset. In each sample, $> 90\%$ of the reads were human contamination, and many samples showed no organism classification. The publication that the data was derived from corroborates this. Newer metagenomic analysis tools were employed to see if the results could have been improved, but this was not the case. Thus, the metagenomic data is likely only a small portion of a larger population of organisms.

VII. References

[1] E. Dekker, P. J. Tanis, J. L. A. Vleugels, P. M. Kasi, and M. B. Wallace, "Colorectal cancer," The Lancet, vol. 394, no. 10207, pp. 1467–1480, Oct. 2019, doi: 10.1016/S0140-6736(19)32319-0.

[2] H. Brenner, M. Kloor, and C. P. Pox, "Colorectal cancer," The Lancet, vol. 383, no. 9927, pp. 1490–1502, Apr. 2014, doi: 10.1016/S0140-6736(13)61649-9.

[3] R. Gao, Z. Gao, L. Huang, and H. Qin, "Gut microbiota and colorectal cancer," Eur. J. Clin. Microbiol. Infect. Dis., vol. 36, no. 5, pp. 757–769, 2017, doi: 10.1007/s10096-016-2881-8.

[4] R. Fodde, "The APC gene in colorectal cancer," Eur. J. Cancer, vol. 38, no. 7, pp. 867–871, May 2002, doi: 10.1016/S0959-8049(02)00040-0.

[5] D. Hirsch et al., "Molecular characterization of ulcerative colitis-associated colorectal carcinomas," Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc, vol. 34, no. 6, pp. 1153–1166, Jun. 2021, doi: 10.1038/s41379-020-00722-5.

[6] C. Cueva, M. Silva, I. Pinillos, B. Bartolomé, and M. V. Moreno-Arribas, "Interplay between Dietary Polyphenols and Oral and Gut Microbiota in the Development of Colorectal Cancer," Nutrients, vol. 12, no. 3, p. 625, Feb. 2020, doi: 10.3390/nu12030625.

[7] S. Schippa and M. P. Conte, "Dysbiotic Events in Gut Microbiota: Impact on Human Health," Nutrients, vol. 6, no. 12, pp. 5786–5805, Dec. 2014, doi: 10.3390/nu6125786.

[8] S. F. Moss, "The Clinical Evidence Linking Helicobacter pylori to Gastric Cancer," Cell. Mol. Gastroenterol. Hepatol., vol. 3, no. 2, pp. 183–191, Dec. 2016, doi: 10.1016/j.jcmgh.2016.12.001.

[9] A. D. Kostic et al., "Genomic analysis identifies association of Fusobacterium with colorectal carcinoma," Genome Res., vol. 22, no. 2, pp. 292–298, Feb. 2012, doi: 10.1101/gr.126573.111.

[10] K. S. Viljoen, A. Dakshinamurthy, P. Goldberg, and J. M. Blackburn, "Quantitative Profiling of Colorectal Cancer-Associated Bacteria Reveals Associations between Fusobacterium spp., Enterotoxigenic Bacteroides fragilis (ETBF) and Clinicopathological Features of Colorectal Cancer," PLoS ONE, vol. 10, no. 3, p. e0119462, Mar. 2015, doi: 10.1371/journal.pone.0119462.

[11] Z. Gao, B. Guo, R. Gao, Q. Zhu, and H. Qin, "Microbiota disbiosis is associated with colorectal cancer," Front. Microbiol., vol. 6, p. 20, Feb. 2015, doi: 10.3389/fmicb.2015.00020.

[12] M. Castellarin et al., "Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma," Genome Res., vol. 22, no. 2, pp. 299–306, Feb. 2012, doi: 10.1101/gr.126516.111.

[13] T. Wang et al., "Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers,"

ISME J., vol. 6, no. 2, pp. 320–329, Feb. 2012, doi: 10.1038/ismej.2011.109.

[14] Y. Sakamoto et al., "The relationship of Fusobacterium nucleatum with antitumour immunity in colorectal cancer liver metastasis," Cancer Sci., Aug. 2021, doi: 10.1111/cas.15126.

[15] J. Yu et al., "Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer," Gut, vol. 66, no. 1, pp. 70–78, Jan. 2017, doi: 10.1136/gutjnl-2015-309800.

[16] B. Périchon et al., "Detection of Streptococcus gallolyticus and Four Other CRC-Associated Bacteria in Patient Stools Reveals a Potential 'Driver' Role for Enterotoxigenic Bacteroides fragilis," Front. Cell. Infect. Microbiol., vol. 12, p. 794391, Mar. 2022, doi: 10.3389/fcimb.2022.794391.

[17] E. Buc et al., "High Prevalence of Mucosa-Associated E. coli Producing Cyclomodulin and Genotoxin in Colon Cancer," PLoS ONE, vol. 8, no. 2, p. e56964, Feb. 2013, doi: 10.1371/journal.pone.0056964.

[18] J. C. Arthur et al., "Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota," Science, vol. 338, no. 6103, pp. 120–123, Oct. 2012, doi: 10.1126/science.1224820.

[19] M. Bonnet et al., "Colonization of the Human Gut by E. coli and Colorectal Cancer Risk," Clin. Cancer Res., vol. 20, no. 4, pp. 859–867, Feb. 2014, doi: 10.1158/1078-0432.CCR-13-1343.

[20] A. Boleij, M. M. H. J. van Gelder, D. W. Swinkels, and H. Tjalsma, "Clinical Importance of Streptococcus gallolyticus Infection Among Colorectal Cancer Patients: Systematic Review and Meta-analysis," Clin. Infect. Dis., vol. 53, no. 9, pp. 870–878, Nov. 2011, doi: 10.1093/cid/cir609.

[21] A. S. Abdulamir, R. R. Hafidh, and F. A. Bakar, "Molecular detection, quantification, and isolation of Streptococcus gallolyticus bacteria colonizing colorectal tumors: inflammation-driven potential of carcinogenesis via IL-1, COX-2, and IL-8," Mol. Cancer, vol. 9, p. 249, Sep. 2010, doi: 10.1186/1476-4598-9-249.

[22] A. Gupta, R. Madani, and H. Mukhtar, "Streptococcus bovis endocarditis, a silent sign for colonic tumour," Colorectal Dis., vol. 12, no. 3, pp. 164–171, 2010, doi: 10.1111/j.1463-1318.2009.01814.x.

[23] M. Venerito, R. Vasapolli, T. Rokkas, J.-C. Delchier, and P. Malfertheiner, "Helicobacter pylori, gastric cancer and other gastrointestinal malignancies," Helicobacter, vol. 22, no. S1, p. e12413, 2017, doi: 10.1111/hel.12413.

[24] Y.-J. Kong, H.-G. Yi, J.-C. Dai, and M.-X. Wei, "Histological changes of gastric mucosa after Helicobacter pylori eradication: A systematic review and meta-analysis," World J. Gastroenterol. WJG, vol. 20, no. 19, pp. 5903–5911, May 2014, doi: 10.3748/wjg.v20.i19.5903.

[25] J. Gagnière et al., "Gut microbiota imbalance and colorectal cancer," World J. Gastroenterol., vol. 22, no. 2, pp. 501–518, Jan. 2016, doi: 10.3748/wjg.v22.i2.501.

[26] N. Zumkeller, H. Brenner, M. Zwahlen, and D. Rothenbacher, "Helicobacter pylori Infection and Colorectal Cancer Risk: A Meta-Analysis," Helicobacter, vol. 11, no. 2, pp. 75–80, 2006, doi: 10.1111/j.1523-5378.2006.00381.x.

[27] N. U. Toprak et al., "A possible role of Bacteroides fragilis enterotoxin in the aetiology of colorectal cancer," Clin. Microbiol. Infect., vol. 12, no. 8, pp. 782–786, Aug. 2006, doi: 10.1111/j.1469-0691.2006.01494.x.

[28] G. Hajishengallis, R. P. Darveau, and M. A. Curtis, "The Keystone Pathogen Hypothesis," Nat. Rev. Microbiol., vol. 10, no. 10, pp. 717–725, Oct. 2012, doi: 10.1038/nrmicro2873.

[29] C. L. Sears, "Enterotoxigenic Bacteroides fragilis: a Rogue among Symbiotes," Clin. Microbiol. Rev., vol. 22, no. 2, pp. 349–369, Apr. 2009, doi: 10.1128/CMR.00053-08.

[30] S. Wu et al., "A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses," Nat. Med., vol. 15, no. 9, pp. 1016–1022, Sep. 2009, doi: 10.1038/nm.2015.

[31] R. Balamurugan, E. Rajendiran, S. George, G. V. Samuel, and B. S. Ramakrishna, "Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, Desulfovibrio and Enterococcus faecalis in the feces of patients with colorectal cancer," J. Gastroenterol. Hepatol., vol. 23, no. 8pt1, pp. 1298–1303, 2008, doi: 10.1111/j.1440-1746.2008.05490.x.

[32] W. Habermann, K. Zimmermann, H. Skarabis, R. Kunze, and V. Rusch, "[The effect of a bacterial immunostimulant (human Enterococcus faecalis bacteria) on the occurrence of relapse in patients with]," Arzneimittelforschung., vol. 51, no. 11, pp. 931–937, Nov. 2001, doi: 10.1055/s-0031-1300140.

[33] J. Gong, T. Bai, L. Zhang, W. Qian, J. Song, and X. Hou, "Inhibition effect of Bifidobacterium longum, Lactobacillus acidophilus, Streptococcus thermophilus and Enterococcus faecalis and their related products on human colonic smooth muscle in vitro," PLoS ONE, vol. 12, no. 12, p. e0189257, Dec. 2017, doi: 10.1371/journal.pone.0189257.

[34] M. M. Huycke, V. Abrams, and D. R. Moore, "Enterococcus faecalis produces extracellular superoxide and hydrogen peroxide that damages colonic epithelial cell DNA," Carcinogenesis, vol. 23, no. 3, pp. 529–536, Mar. 2002, doi: 10.1093/carcin/23.3.529.

[35] A. J. Williamson et al., "Enterococcus faecalis promotes a migratory and invasive phenotype in colon cancer cells," Neoplasia N. Y. N, vol. 27, p. 100787, Mar. 2022, doi: 10.1016/j.neo.2022.100787.

[36] Y. Guo, Y. Bao, M. Ma, and W. Yang, "Identification of Key Candidate Genes and Pathways in Colorectal Cancer by Integrated Bioinformatical Analysis," Int. J. Mol. Sci., vol. 18, no. 4, p. 722, Mar. 2017, doi: 10.3390/ijms18040722.

[37] M. Yan, M. Song, R. Bai, S. Cheng, and W. Yan, "Identification of potential therapeutic targets for colorectal cancer by bioinformatics analysis," Oncol. Lett., vol. 12, no. 6, pp. 5092–5098, Dec. 2016, doi: 10.3892/ol.2016.5328.

[38] Y. Kou, S. Zhang, X. Chen, and S. Hu, "Gene expression profile analysis of colorectal cancer to investigate potential mechanisms using bioinformatics," OncoTargets Ther., vol. 8, pp. 745–752, Apr. 2015, doi: 10.2147/OTT.S78974.

[39] G. Debesa-Tur et al., "Metagenomic analysis of formalin-fixed paraffin-embedded tumor and normal mucosa reveals differences in the microbiome of colorectal cancer

patients," Sci. Rep., vol. 11, no. 1, p. 391, Jan. 2021, doi: 10.1038/s41598-020-79874-y.

[40] S.-M. Wu et al., "Comprehensive transcriptome profiling of Taiwanese colorectal cancer implicates an ethnic basis for pathogenesis," Sci. Rep., vol. 10, no. 1, p. 4526, Mar. 2020, doi: 10.1038/s41598-020-61273-y.

[41] "Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data." https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed Nov. 07, 2021).

[42] "Trimmomatic: a flexible trimmer for Illumina sequence data — Bioinformatics — Oxford Academic." https://academic.oup.com/bioinformatics/article/30/15/2114/2390096 (accessed Nov. 07, 2021).

[43] B. Langmead and S. L. Salzberg, "Fast gapped-read alignment with Bowtie 2," Nat. Methods, vol. 9, no. 4, pp. 357–359, Apr. 2012, doi: 10.1038/nmeth.1923.

[44] F. Beghini et al., "Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3," eLife, vol. 10, p. e65088, May 2021, doi: 10.7554/eLife.65088.

[45] F. Asnicar, G. Weingart, T. L. Tickle, C. Huttenhower, and N. Segata, "Compact graphical representation of phylogenetic data and metadata with GraPhlAn," PeerJ, vol. 3, p. e1029, Jun. 2015, doi: 10.7717/peerj.1029.

[46] Segata Lab, export2graphlan. Segata Lab, 2022. Accessed: May 04, 2022. [Online]. Available: https://github.com/SegataLab/export2graphlan

[47] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter, "Near-optimal probabilistic RNA-seq quantification," Nat. Biotechnol., vol. 34, no. 5, pp. 525–527, May 2016, doi: 10.1038/nbt.3519.

[48] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," Genome Biol., vol. 15, no. 12, p. 550, Dec. 2014, doi: 10.1186/s13059-014-0550-8.

[49] C. Soneson, M. I. Love, and M. D. Robinson, "Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences." Feb. 29, 2016. doi: 10.12688/f1000research.7563.2.

[50] M. Lawrence et al., "Software for Computing and Annotating Genomic Ranges," PLOS Comput. Biol., vol. 9, no. 8, p. e1003118, Aug. 2013, doi: 10.1371/journal.pcbi.1003118.

[51] D. Smedley et al., "BioMart – biological queries made easy," BMC Genomics, vol. 10, no. 1, p. 22, Jan. 2009, doi: 10.1186/1471-2164-10-22.

[52] G. Yu and Q.-Y. He, "ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization," Mol. Biosyst., vol. 12, no. 2, pp. 477–479, Jan. 2016, doi: 10.1039/C5MB00663E.

[53] "Frontiers — pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks — Genetics." https://www.frontiersin.org/articles/10.3389/fgene.2019.00858/full (accessed Apr. 16, 2022).

[54] K. Blighe, EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. 2022. Accessed: May 04, 2022. [Online]. Available: https://github.com/kevinblighe/EnhancedVolcano

[55] G. C. A. Amos et al., "Exploring how microbiome signatures change across inflammatory bowel disease conditions and disease locations," Sci. Rep., vol. 11, p. 18699, Sep. 2021, doi: 10.1038/s41598-021-96942-z.

[56] E. Lkhagva et al., "The regional diversity of gut microbiome along the GI tract of male C57BL/6 mice," BMC Microbiol., vol. 21, p. 44, Feb. 2021, doi: 10.1186/s12866-021-02099-0.

[57] C. A. Brennan and W. S. Garrett, "Fusobacterium nucleatum — symbiont, opportunist and oncobacterium," Nat. Rev. Microbiol., vol. 17, no. 3, pp. 156–166, Mar. 2019, doi: 10.1038/s41579-018-0129-6.

[58] J. Obata, K. Fujishima, E. Nagata, and T. Oho, "Pathogenic mechanisms of cariogenic Propionibacterium acidifaciens," Arch. Oral Biol., vol. 105, pp. 46–51, Sep. 2019, doi: 10.1016/j.archoralbio.2019.06.005.

[59] M. V. Sizova et al., "Oribacterium parvum sp. nov. and Oribacterium asaccharolyticum sp. nov., obligately anaerobic bacteria from the human oral cavity, and emended description of the genus Oribacterium," Int. J. Syst. Evol. Microbiol., vol. 64, no. Pt 8, pp. 2642–2649, Aug. 2014, doi: 10.1099/ijs.0.060988-0.

[60] S. s. Socransky, A. d. Haffajee, M. a. Cugini, C. Smith, and R. L. Kent Jr., "Microbial complexes in subgingival plaque," J. Clin. Periodontol., vol. 25, no. 2, pp. 134–144, 1998, doi: 10.1111/j.1600-051X.1998.tb02419.x.

[61] A. Bertelsen, J. S. Elborn, and B. C. Schock, "Infection with Prevotella nigrescens induces TLR2 signalling and low levels of p65 mediated inflammation in Cystic Fibrosis bronchial epithelial cells," J. Cyst. Fibros., vol. 19, no. 2, pp. 211–218, Mar. 2020, doi: 10.1016/j.jcf.2019.09.005.

[62] E. Aja, M. Mangar, H. M. Fletcher, and A. Mishra, "Filifactor alocis: Recent Insights and Advances," J. Dent. Res., vol. 100, no. 8, pp. 790–797, Jul. 2021, doi: 10.1177/00220345211000656.

[63] A. V. B. Nogueira et al., "Filifactor alocis and Tumor Necrosis Factor-Alpha Stimulate Synthesis of Visfatin by Human Macrophages," Int. J. Mol. Sci., vol. 22, no. 3, p. 1235, Jan. 2021, doi: 10.3390/ijms22031235.

[64] S. Kim, A. Covington, and E. G. Pamer, "The intestinal microbiota: Antibiotics, colonization resistance, and enteric pathogens," Immunol. Rev., vol. 279, no. 1, pp. 90–105, Sep. 2017, doi: 10.1111/imr.12563.

[65] H. M. Wexler, "Bacteroides: the Good, the Bad, and the Nitty-Gritty," Clin. Microbiol. Rev., vol. 20, no. 4, p. 593, Oct. 2007, doi: 10.1128/CMR.00008-07.

[66] M. Avgeris, K. Mavridis, and A. Scorilas, "Kallikrein-related peptidase genes as promising biomarkers for prognosis and monitoring of human malignancies," Biol. Chem., vol. 391, no. 5, pp. 505–511, May 2010, doi: 10.1515/BC.2010.056.

[67] S. Makawita et al., "Validation of four candidate pancreatic cancer serological biomarkers that improve the performance of CA19.9," BMC Cancer, vol. 13, p. 404, Sep. 2013, doi: 10.1186/1471-2407-13-404.

[68] H. Mao et al., "Protective and anti-inflammatory role of REG1A in inflammatory bowel disease induced by JAK/STAT3 signaling axis," Int. Immunopharmacol., vol. 92, p. 107304, Mar. 2021, doi: 10.1016/j.intimp.2020.107304.

[69] Y.-S. Jung, S. Jun, S. H. Lee, A. Sharma, and J.-I. Park, "Wnt2 complements Wnt/-catenin signaling in colorectal cancer," Oncotarget, vol. 6, no. 35, pp. 37257–37268, Nov. 2015, doi: 10.18632/oncotarget.6133.

[70] Y. Li et al., "EN2 as an oncogene promotes tumor progression via regulating CCL20 in colorectal cancer," Cell Death Dis., vol. 11, no. 7, p. 604, Jul. 2020, doi: 10.1038/s41419-020-02804-3.

[71] K. Parikh et al., "Colonic epithelial cell diversity in health and inflammatory bowel disease," Nature, vol. 567, no. 7746, pp. 49–55, Mar. 2019, doi: 10.1038/s41586-019-0992-y.

[72] D. Q. Wu et al., "AQP8 inhibits colorectal cancer growth and metastasis by down-regulating PI3K/AKT signaling and PCDH7 expression," Am. J. Cancer Res., vol. 8, no. 2, pp. 266–279, 2018.

[73] G.-Z. Yang et al., "Prognostic value of carbonic anhydrase VII expression in colorectal carcinoma," BMC Cancer, vol. 15, p. 209, Apr. 2015, doi: 10.1186/s12885-015-1216-y.

[74] J. Mao et al., "Metallothionein MT1M is a tumor suppressor of human hepatocellular carcinomas," Carcinogenesis, vol. 33, no. 12, pp. 2568–2577, Dec. 2012, doi: 10.1093/carcin/bgs287.

[75] S. Nomiri et al., "Prediction and validation of GUCA2B as the hub-gene in colorectal cancer based on co-expression network analysis: In-silico and in-vivo study," Biomed. Pharmacother. Biomedecine Pharmacother., vol. 147, p. 112691, Mar. 2022, doi: 10.1016/j.biopha.2022.112691.

[76] X.-S. He et al., "Oncogenic potential of BEST4 in colorectal cancer via activation of PI3K/Akt signaling," Oncogene, vol. 41, no. 8, pp. 1166–1177, Feb. 2022, doi: 10.1038/s41388-021-02160-2.

[77] Y. Zheng et al., "CA1 contributes to microcalcification and tumourigenesis in breast cancer," BMC Cancer, vol. 15, p. 679, Oct. 2015, doi: 10.1186/s12885-015-1707-x.

[78] C. Wei et al., "Crosstalk between cancer cells and tumor associated macrophages is required for mesenchymal circulating tumor cell-mediated colorectal cancer metastasis," Mol. Cancer, vol. 18, p. 64, Mar. 2019, doi: 10.1186/s12943-019-0976-4.

[79] S. R. Choudhury et al., "Dipeptidase-1 Is an Adhesion Receptor for Neutrophil Recruitment in Lungs and Liver," Cell, vol. 178, no. 5, pp. 1205-1221.e17, Aug. 2019, doi: 10.1016/j.cell.2019.07.017.

[80] S. S. Chukkapalli, I. M. Velsko, M. F. Rivera-Kweh, D. Zheng, A. R. Lucas, and L. Kesavalu, "Polymicrobial Oral Infection with Four Periodontal Bacteria Orchestrates a Distinct Inflammatory Response and Atherosclerosis in ApoEnull Mice," PLoS ONE, vol. 10, no. 11, p. e0143291, Nov. 2015, doi: 10.1371/journal.pone.0143291.

[81] A. Larabi, N. Barnich, and H. T. T. Nguyen, "New insights into the interplay between autophagy, gut microbiota and inflammatory responses in IBD," Autophagy, vol. 16, no. 1, pp. 38–51, Jul. 2019, doi: 10.1080/15548627.2019.1635384.

[82] M. Mittal, M. R. Siddiqui, K. Tran, S. P. Reddy, and A. B. Malik, "Reactive Oxygen Species in Inflammation and Tissue Injury," Antioxid. Redox Signal., vol. 20, no. 7, pp. 1126–1167, Mar. 2014, doi: 10.1089/ars.2012.5149.