Effect of the Tumor Microenvironment and Microbiome on the CMS of Colorectal Cancer

Shin Young Chang

Department of Medical Science

The Graduate School, Yonsei University

Effect of the Tumor Microenvironment and Microbiome on the CMS of Colorectal Cancer

Shin Young Chang

Department of Medical Science

The Graduate School, Yonsei University

Effect of the Tumor Microenvironment and Microbiome on the CMS of Colorectal Cancer

Directed by Professor Tae Il Kim

The Master's Thesis
submitted to the Department of Medical Science,
the Graduate School of Yonsei University
in partial fulfilment of the requirements for the degree of
Master of Medical Science

Shin Young Chang

December 2022

This certifies that the Master's Thesis
of Shin Young Chang is approved.

_____

Thesis Supervisor: Tae Il Kim

_____

Thesis Committee Chair: Ki Taek Nam

_____

Thesis Committee Member: Hyunki Kim

The Graduate School
Yonsei University

December 2022

ACKNOWLEDGEMENTS

I want to express my sincerest thanks to my colleagues, friends, and family that supported me along the journey. The road was not easy; it was only through their presence and encouragement that I could keep going. These past two years were some of the hardest times I have ever experienced, and I was challenged and tested in ways I never expected. Besides academic knowledge, I learned about life, relationships, and myself. Working as a full-time researcher with a dedicated project made me realize my love of researching, despite any failures or obstacles that propped up.

Professor Tae Il, you were the best mentor a student can have, and through your tutelage, I learned to see things from a new perspective and push myself to be a better scientist. Thank you for spending valuable time, effort, patience, and care to discuss any concerns I had, be they work-related or not. Your dedication to guiding me has never gone unnoticed or unappreciated.

Thank you to Dr. Dong Keon Kim for teaching me the basics of data analysis skills and other techniques vital to any thriving researcher. Our conversations on research always left me with new ideas to contemplate

on. Thank you to the other lab members as well for any help provided.

I want to thank Ms. Sang Mi Oh for sharing her expert lab techniques, tips, and tricks. I couldn't have learned from a more capable person. I am grateful that I was blessed with the best sunbae, Joyeon Kang Unni, an incredible person who I could always look up to and be inspired by. Thank you to Ms. Bomi Kwon, whose bright and sincere disposition always cheered me up and made our lunches something to look forward to. I am truly blessed to be able to get close to these three during my masters as they were the best company to be around and some of the greatest advisors and listeners one could hope for.

Thank you to Minwoo sunbae for sharing your insight on life and work in science. Working as an intern under you in undergraduate school was a joy, and I genuinely appreciated how much you cared for us. I also want to thank my dearest friends Sooyeon, Ayesha, Nayab, and Hashir for always being there to celebrate or sympathize with me. I both apologize, and thank you for listening to my rants, lol. You guys are the best. The biggest shout-out goes to my SO; I couldn't have done this without you. I am a stronger and better person because of you.

My final thanks go to my family for sticking up for me and supporting me. To my sister, Yun, although we have our moments, I know I can always count on you to have my back when the time comes. You inspire me, and

I admire you for everything you have achieved and overcome. Thank you to my mom, grandmother, and dad for supporting me financially and letting me achieve my dreams.

Thank you once again, everyone, for everything.

December 2022

Shin Young Chang

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Effect of the Tumor Microenvironment and Microbiome on the CMS of Colorectal Cancer

Shin Young Chang

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Tae Il Kim)

CMS4, a subtype of colorectal cancer (CRC), is a mesenchymal subtype with high stromal content and the worst survival. Its main causative factors are still unknown. Since the gut microbiome is an essential part of the tumor microenvironment (TME) of CRC, I hypothesized that gut bacteria played a significant role in inducing this cancer subtype. RNA sequencing and CMScaller were used on patient tumor tissues (n= 115) and patient-derived organoids (PDOs, n=115) for gene expression data and to identify their CMS. 16S-rRNA Microbiome Taxonomic Profiling (MTP) was used to find CMS4-specific bacteria in the tissues of the same patient group. Single-cell and coculture experiments were performed using CRC CMS2 cell line (LS1034) and PDOs.

For coculture, 18Co and THP1 were used to represent TME cells. Bacteria treatment was done in two combinations for coculture experiment: bacteria treatment of 1) both TME cells and PDO/cell line, 2) only the TME cells. PDOs were treated with bacteria for 1.5 hours, and cell lines (LS1034, TME) were treated for 2 hours in anaerobic conditions before they were cultured in aerobic conditions. Gene Set Enrichment Analysis (GSEA) was used for CMS4-specific gene expression analysis. MTP analysis revealed that *Bacteroides fragilis* had a large LDA effect size of 4.38 (p-value = 0.008) in CMS4 tissue microbiome relative to other subtypes (CMS2 and CMS3). CRC patients with CMS4 tissues and non-consensus subtype PDOs were selected and used for *in vitro* experiments. CMScaller results showed that LS1034 cultured without TME cells had constant subtype after both 24 hours and 7 days although for PDOs it changed to CMS3. TME-cocultured PDOs subtype did not change after 24 hours from seeding, but it changed to CMS4 after 7 days only in *B. fragilis*-treated conditions. GSEA analysis displayed significant upregulation of CMS4 genes in *B. fragilis*-treated TME-cocultured PDOs compared to when bacteria is treated to both PDO and TME cells and in untreated samples (Enrichment Score = 0.294, p-value<0.001). Furthermore, KEGG pathway enrichment analysis showed significant upregulation of cancer-associated pathways in bacteria-treated PDOs. We found that *B. fragilis* is highly enriched in CRC CMS4 tissues and demonstrated that *B. fragilis* induced

CMS4 genes significantly in cocultured PDO model.

Key words: colorectal cancer, consensus molecular subtype, microenvironment, microbiome, *Bacteroides fragilis*

Effect of the Tumor Microenvironment and Microbiome on the CMS of Colorectal Cancer

Shin Young Chang

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Tae Il Kim)

I. INTRODUCTION

Colorectal cancer (CRC) is the third most prevalent cancer worldwide, with almost 2 million cases in 2020, and it is only second to lung cancer in terms of lethality[1]. Although preventative measures such as CRC screenings help to diagnose cancer in its early stages and start prompt treatment, current diagnosis methods only have a 40 percent success rate in detecting early CRC[2]. Furthermore, CRC cases can recur after treatment or surgery, which can be attributed to the heterogeneous nature of CRC tumors.

In 2015, the CRC subtyping consortium formed the consensus molecular subtyping (CMS) system after extensive data analysis of core subtyping patterns in the gene expression

of the heterogenous cancer[3]. They categorized CRC into four major molecular subtypes that cover almost 87 percent of all cases and allow for targeted patient therapy based on their subtype. CMS1 is characterized by hypermethylated DNA, frequent BRAF mutations and high microsatellite instability. CMS2 is the differentiated epithelial subtype that most frequently ails patients and has upregulated WNT and MYC signal transduction pathways. CMS3 is signified by metabolic dysregulation and overrepresented KRAS mutations and CMS4 is a mesenchymal subtype characterized by stromal infiltration, high TGF-b signaling, and angiogenesis. Of the four subtypes, CMS4 has the worst survival rate in both relapse-free and overall survival conditions. Patients with this subtype are usually diagnosed at the later stages (III and IV) of cancer. The invasive and mesenchymal characteristics of CMS4 often lead to metastasis and thus indicate a poor prognosis for the patient[4]. As CMS4 is a mesenchymal subtype, the tumor microenvironment (TME) is undoubtedly significant to understanding it.

Since long ago, in vitro studies performed to research CRC and drug discovery widely involved using 2D cell line culture. The recent establishment of 3D organoid culture techniques with better clinical translation and representation of in vivo environment than 2D cultures have made them powerful alternatives for use in research[5]. However, there lies a major drawback to both as they lack the complex tumor microenvironment with multiple cell types interacting with each other, usually present naturally in CRC patients. Numerous scientific articles have been published that go in-depth on the significance of TME, such as cancer-associated fibroblasts (CAFs), in cancer progression[6]. The lack of TME components in cell

cultures makes them a poor model for CRC studies, especially for subtypes such as CMS4 that are stromal-rich and depend heavily on their microenvironment for tumor progression. The TME consists of multiple non-epithelial cell types, including vasculature, CAFs, inflammatory cells, extracellular matrix (ECM), and other molecules associated with the matrix. Transformed epithelial cells in tumor are known to manipulate their microenvironment to boost their growth, survival, and spread[7].

There is another component that is an essential part of CRC TME: the gut microbiome. The human gastrointestinal (GI) tract consists of trillions of microorganisms that have evolved over thousands of years and usually live in a mutually beneficial symbiotic relationship with their host. The microbiome comprises archaea, eukarya, and bacteria. 16S ribosomal RNA and whole-genome shotgun metagenomic studies found that 93.5% of the gut microbiome consisted of species belonging to the Proteobacteria, Firmicutes, Actinobacteria, and Bacteroidetes phyla. However, it is a dynamic composition that changes depending on diet, health, environment, and other intrinsic and extrinsic factors[8-10]. The microorganisms living in the GI tract have a variety of functions, including protecting from pathogenic invaders, immune response regulation, and maintaining the intestinal epithelium. They usually affect human health positively[11]. Consequently, imbalances in the microbiota makeup can have severe consequences, such as chronic inflammation, abnormal immune system and gut dysbiosis which has been linked to various human diseases such as obesity, metabolic, cardiovascular, inflammatory bowel disease (IBD), and CRC[12].

Several microbial species have been associated with being responsible for colorectal cancer initiation and progression. In vivo studies using AOM-treated, colitis-susceptible, IL10 deficient, germ-free mice found that +*pks Escherichia coli* NC101 induced invasive adenocarcinomas in mono-colonised mice whereas other bacterial strain without the *pks* island did not suggesting that bacterial strains with this gene could be carcinogenic. CRC and IBD patients were also found to have a high abundance of +*pks E. coli*[13]. Another study found *Fusobacterium nucleatum* to be in increased quantities in adenomas and colorectal cancer. When APC gene mutated mice were colonised with *F. nucleatum*, they had enhanced colonic tumor formation, infiltration of myeloid cells and pronounced NF-κB proinflammatory gene expression with upregulated PTGS2. The results suggested that the myeloid cell-mediated inflammatory signaling may support early colorectal neoplasia's development into cancer[14]. Enterotoxigenic *Bacteroides fragilis* (ETBF) is yet another species that has been directly associated as a promotor of colon tumorigenesis. ETBF produces a toxin (BFT) that causes inflammatory diarrhoea, and research done in 2009 found that ETBF contributed to colonic hyperplasia and tumorigenesis by using the STAT3/Th17 pathway and through chronic inflammation[15].

Furthermore, as the role of gut microbiota in CRC is becoming more prominent, the microbiota composition has even been found to differ depending on the subtype and whether it is an early or late-stage cancer. Certain species of bacteria were only significantly enriched in one specific subtype, such as *Fusobacterium hwasooki* and *Porphyromonas gingivalis* in

CMS1 and *Selenoma* and *Prevotella* species in CMS2[16]. The research did not include CMS4 tissues for microbiome analysis. Although research exists that comments on significant differences in microbiome bacteria depending on the molecular subtype, thus far, there has been no direct evidence of the effect of the tumor microbiome on the CRC CMS, especially in the TME-dependent CMS4. Therefore, CRC tissue microbiome was investigated for species enriched only in the CMS4 and its influence in inducing CMS4 was examined.

Table 1. Patient cohort characteristics. (n = number of patients)

		Patients (n)
Age	26-92 years	
	Mean = 63 years	
Gender	Males	66
	Females	49
Biopsy Method	Colonoscopy	52
	Surgery	63
Site	Right	26
	Left	47
	Rectum	42
Stage	0	2
	1	24
	2	29
	3	36
	4	24
Histology	Well Differentiated	22
	Moderately Differentiated	86
	Poorly Differentiated	3
	Mucinous	2
	Low Grade Dysplasia	1
	other	1

II. MATERIALS AND METHODS

1. Patient cohort and samples

Colorectal cancer tumor samples were collected from a total of 115 patients. Written informed consent was obtained from patients prior to the beginning of the project. The study was carried out with the approval of the Institutional Review Board of Yonsei University College of Medicine, Severance Hospital (IRB No. 4-2012-0859) and all experiments were performed according to the appropriate guidelines and regulations. The patients ranged between 26-92 years (mean age, 63 years) with 66 male patients and 49 female patients. Biopsy samples were obtained via either surgery or colonoscopy and the site of tumors samples included 26 from the right colon, 47 from the left colon and 42 from the rectum. Histologically, 22 tumors were described as well differentiated, 86 were moderately differentiated and 3 were poorly differentiated. Table 1 presents detailed information regarding patient cohort characteristics.

2. Cell culture and reagents

Human myofibroblast cell line (18Co) and monocytic myeloid cell line (THP-1) were purchased from Korean Cell Line Bank (KCBL, Seoul, South Korea). CMS2 human colorectal carcinoma cell line (LS1034) was purchased from the American Type Culture Collection (ATCC, Maryland, USA). All cells were cultured in High Glucose Dulbecco's modified Eagle's medium (DMEM) (Cytiva, Massachusetts, USA) supplemented with 10%

fetal bovine serum (FBS) (Gibco-Life Technologies, New York, USA), 100 units/ml penicillin, and 100 mg/ml streptomycin (P/S) (Invitrogen, California, USA). All cells were maintained in a 5% CO2 incubator at 37°C.

For subculture, cells were seeded with a density of 1×10^6 cells/mL in 100 mm culture plates and cultured until they reached 80% confluency. Cells were washed using 1x PBS (GeneTech, Gyeonggi, South Korea). Adherent cells were detached from plate surface using 0.25% trypsin-EDTA (Invitrogen) and neutralized with High Glucose DMEM containing 10% FBS, 1 %P/S. Cells density was measured using cell counter (NanoEnTek, Seoul, South Korea).

For the monoculture experiments, $3x10^5$ LS1034 cells were seeded in 6 well plates and grown over 3 days. For coculture experiments, $1x10^5$ LS1034 cells were seeded in the upper well of 12-well transwell plate. For each coculture experiment, 1×10^5 cells of each cell line (THP-1, 18Co) were mixed and cultured in lower well of a 12-well transwell.

3. Tumor organoid culture

Patient-derived tumor organoids (PDOs) were established from tissue samples of the patient cohort in Table 1. Intestinal crypts were disassociated from tissue samples and embedded in growth factor-reduced Matrigel (Corning, New York, USA), and seeded in 48-well plates (20 µl of Matrigel per well). The Matrigel was left to polymerize for 15 min at 37°C after which 250 µl/well basal culture medium (Advanced DMEM/F12 medium

supplemented with 1% P/S , 1x Glutamax, 1x N2, B27 (Invitrogen), and 1mM N-acetyl-l-cysteine (Sigma-Aldrich, St. Louis, MO, USA) was added with the following optimized growth factor combinations: 2 mM l-glutamine, 10 mM nicotinamide, 10 nM gastrin I, 500 nM A-83-01, 10 μM SB202190, 50 ng/ml EGF, 100ng/ml hNoggin and 100ng/ml Wnt3a.

All organoids were maintained in a 5% CO2 incubator at 37°C. For passaging, PDOs were washed with PBS and disassociated using 1x TrypLE (Gibco-Life Technologies) after which it was seeded in 20 ul Matrigel and overlaid with PDO growth media. For monoculture experiments, $6x10^3$ cells were seeded in 24-well culture plate and grown for 7 days in PDO growth media. For coculture experiments, $3x10^3$ organoid cells were grown in the upper chambers of the transwell plates for 7 days. PDO cells in each condition were measured after a week of growth.

4. Bacteria culture

Study was performed using *Bacteroides fragilis*, *Fusobacterium nucleatum* and *Lactobacillus Acidophilus* from ATCC which were grown as suspension cultures on 5% sheep blood (MBcell, Seoul, South Korea), Wilkins Chalgren (Oxoid, Basingstoke, UK) or MRS Broth (MBcell) and 1g/L sodium thioglycolate (Sigma-Aldrich) in anaerobic conditions using 2.5 L anaerobe jar (Oxoid) and CO_2 gas pack. Bacterial cell density was calculated using ELISA plate reader using O.D. at 600 nm wavelength. Cell lines and PDOs were treated with bacteria in a 1:100 cell to bacteria ratio. Cell lines in serum free media were treated with bacteria for 2 hours in anaerobic conditions. PDOs suspended in serum free media were

treated with bacteria for 1.5 hours in anaerobic conditions. After bacteria treatment, growth media supplemented with 100 ug/ml Normocure (Invitrogen) was added to the cells/PDO and cultured in aerobic conditions.

Cell Tracker Red (CMTPX; Invitrogen) was used to label the bacteria to confirm cell infection. The bacteria were treated with 1 uM of product and stored in anaerobic conditions for 15 mins after which they were removed and washed with PBS before being used for cell line or PDO treatment.

5. RNA isolation and sequencing, Quantitative RT-PCR and RNA sequencing

Total RNA was isolated using Trizol reagent (Gibco-Life Technologies) and RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer protocol. RNA quantity was assessed using NanoDrop ND-1000 spectrophotometer and sent for RNA sequencing to an external company. cDNA synthesis from total RNA was performed using 5x Reverse Transcription Master Premix (HK Genomics, Daejeon, South Korea). Real-time qPCR was performed using SYBR Green Master mix (Enzynomics, Daejeon, South Korea) and primers for CDX2, FRMD6, HTR2B, ZEB1 with GAPDH used for normalization.

6. Consensus Molecular Subtype classification

Read counts from RNA sequencing data were processed in CMScaller[17] which is a R package. It was used to predict a molecular subtype (CMS) of CRC for each sample based

on their gene expression profiles. For samples with non-classifiable subtypes, nearest CMS prediction was used.

7. 16s rRNA MTP analysis

16s rRNA gene sequencing was used to perform comparative microbiome taxonomic profiling (MTP) analysis using EzBioCloud[18]. CMS4 16s rRNA samples were used as one MTP set and CMS2 and CMS3 as another MTP set for all microbiome analysis. Alpha diversity was found using Simpson Index and Alpha diversity richness was analyzed using Chao1 index. Beta diversity was calculated using Principal Coordinate Analysis (PCoA) using Unifrac as a measure. Taxonomic biomarker discovery was performed using LEfSe analysis to find meaningful microbiota abundances.

8. Gene-set enrichment analysis (GSEA)

GSEA software ver. 4.3.2. from Broad Institute was utilized to determine significantly upregulated CMS4 genes in bacteria-treated samples compared to untreated samples[19]. Fold change values obtained from logCPM of RNAseq data of samples were run on GSEA pre-ranked gene list using CMS4-specific gene set database which was obtained from CMScaller[32]. Sets within the size range of 15 and 236 were analyzed and 1000 permutations were done per analysis. The cut-off values were set to p value < 0.05 after which they were written in red.

9. Statistical analysis

Statistical analysis was performed using free statistics software Jamovi (website: https://www.jamovi.org). Wilcoxon Rank-sum test and Non-parametric, Kruskal Wallis tests were performed as appropriate. P value threshold was set to 5% for statistical significance. To assess similarity between tissue and PDO cocultured samples correlation matrix heatmap was drawn using Spearman's correlation coefficient after normalization of RNA expression data with patient tissue's data.

10. Bioinformatic analysis

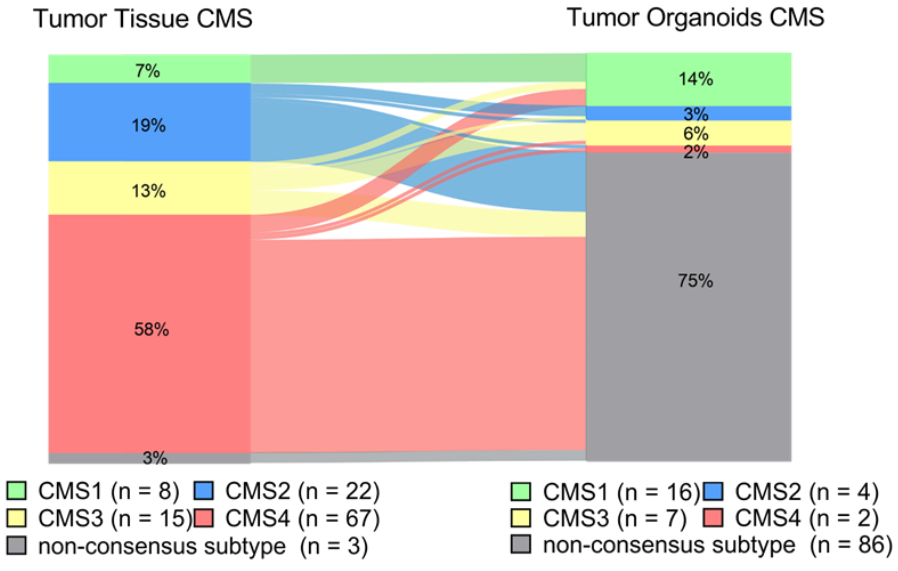
KEGG pathway enrichment analysis was performed to discover functional roles of DEGs in bacteria-treated samples. Database for Annotation, Visualization and Integrated Discovery (DAVID) (website: https://david.ncifcrf.gov/home.jsp) was used for KEGG pathway analysis using differentially expressed genes with fold change expression>2. Non-negative matrix (NMF) clustering was performed using RNAsequencing read counts data and using project R package.

III. RESULTS

1. Majority of CRC CMS4 tumor tissues change their subtype when cultured as PDOs

The difference in CMS between CRC patient tumor tissues and PDOs was found after the RNA of 115 colorectal cancer patients was collected and sequenced. This data was then processed into CMScaller to identify each patient's CMS. Organoids cultured from these patients were similarly analyzed and subtyped. The results showed that besides CMS1 tissue samples, other tissue samples all had a subset of tissues whose subtypes modified when cultured as PDO (Fig. 1A). This was especially pronounced in the case of CMS4 as only 1.5% of CMS4 tissues retained their subtype upon being cultured as organoids. In comparison, the majority (89.6%) did not fit into any major molecular subtype as PDOs (Fig. 1B).

A



B

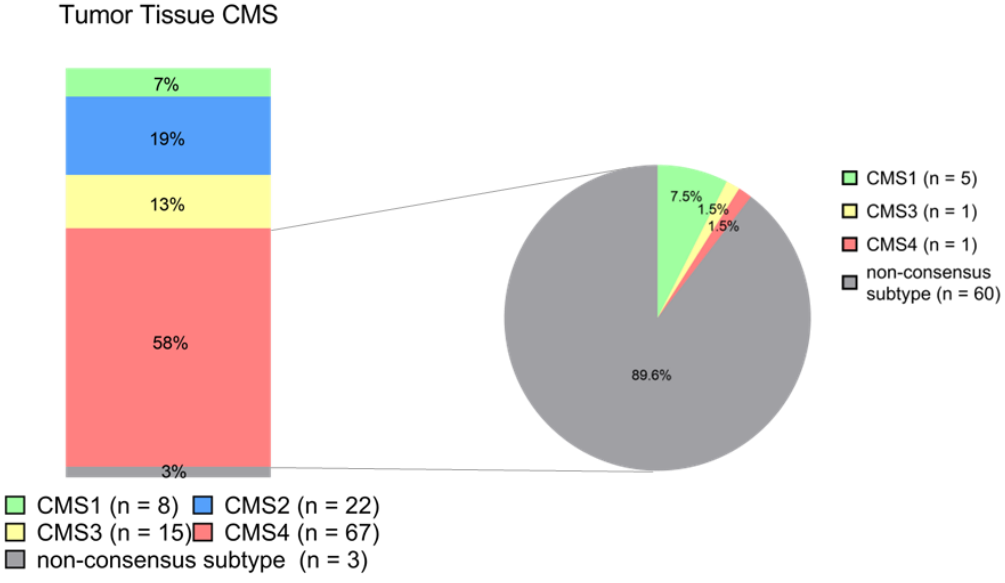


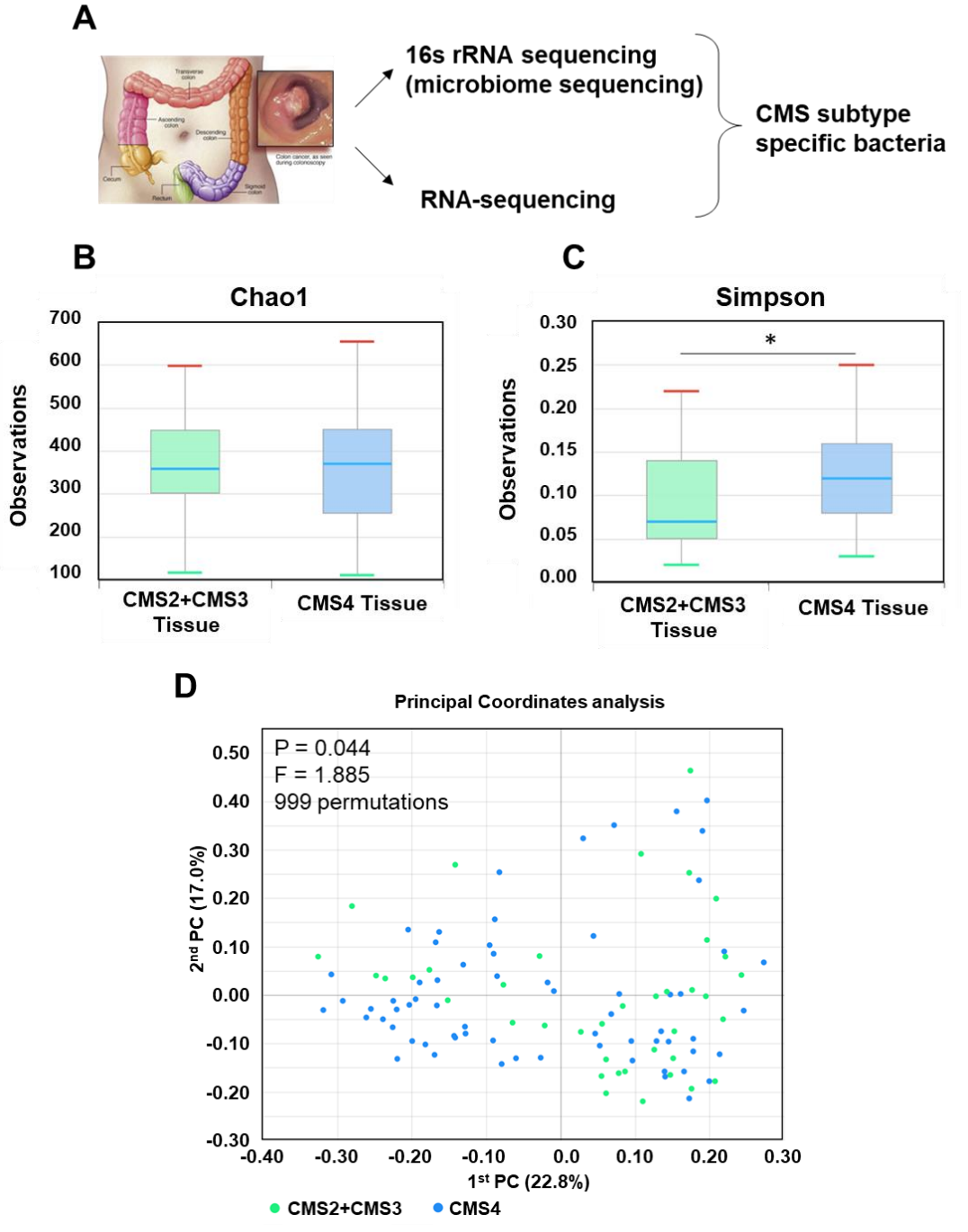
Figure 1. CRC tissue subtypes changed when cultured as PDOs.

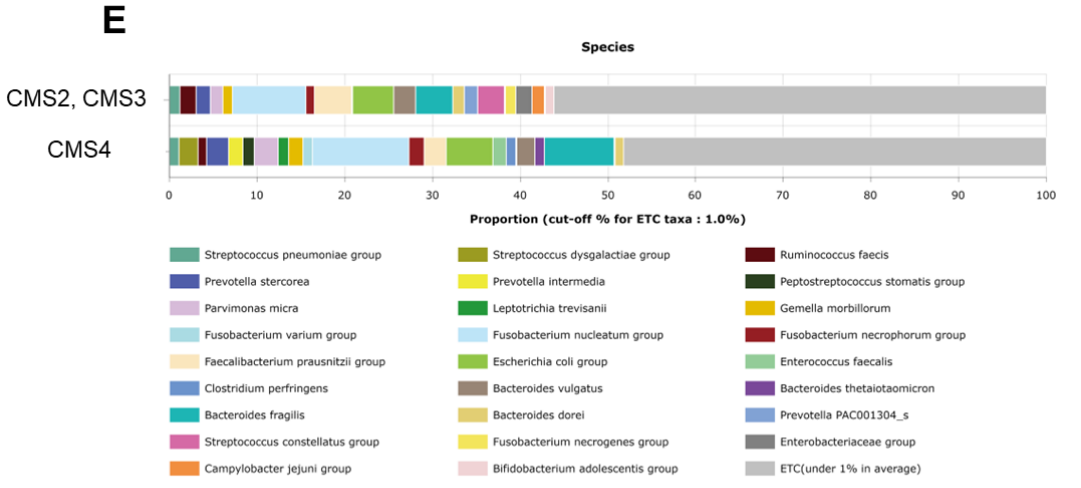
A. Alluvial plot shows CMS of 115 CRC patient tumor tissues compared with their subtype post-culture as patient-derived organoids (PDO). **B**. Relative CMS proportions of PDOs after being cultured from primary CMS4 CRC tumor tissue. n: number of each sample.

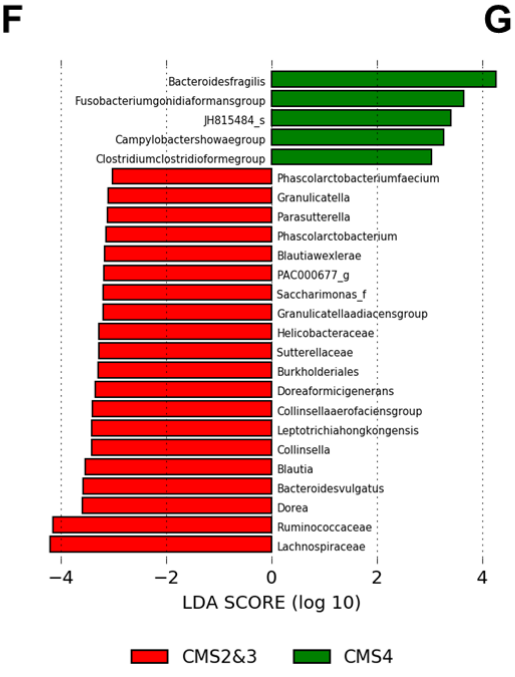
2. *Bacteroides fragilis* is substantially enriched in CMS4 tumor tissue microbiomes

The tissue microbiome of each subtype was analyzed to investigate the reason for the subtype difference between tumor tissue and its respective organoid. 16S rRNA sequencing data from 106 CRC patients of known subtypes (CMS2, CMS3, and CMS4) was processed in a comparative Microbiome Taxonomic Profiling (MTP) analyzer to compare their relative taxonomic compositions. 16s rRNA data of CMS2 and CMS3 were combined and compared against the 16s rRNA MTP data of CMS4 to discover significant taxonomic biomarkers (Fig. 2A). Alpha diversity (Simpson) analysis displayed a considerable variety of microorganisms in CMS4 tissue microbiome compared to CMS2 and CMS3 microbiome. However, Chao1 alpha diversity richness analysis did not show a significant difference in species richness of the two sets (Fig. 2B-C). Additionally, microbiome beta diversity analysis using PCoA between CMS4 and combined CMS2 and CMS3 data found meaningful difference between the two MTP sets species communities with pseudo-F = 1.885 (Fig. 2D).

Averaged species taxonomic compositions of the two MTP sets uncovered a notable difference in their *B. fragilis* carriage. CMS4 tissue microbiome had almost twice the abundance of *B. fragilis* species community (7.99%) in contrast to CMS2, CMS3 microbiome (4.19%) (Fig. 2E). LEfSe analysis of the 16s rRNA data also showed considerable LDA effect size (LDA = 4.3, p value = 0.006) for *B. fragilis* species in CMS4 tissues compared to CMS2, CMS3 tissues (Fig. 2F). A comparison of the taxonomic abundance of bacteria species having LDA score > 3 in CMS4 tissues showed that *B. fragilis* had the most significant abundance in CMS4 tissues (Fig. 2G).







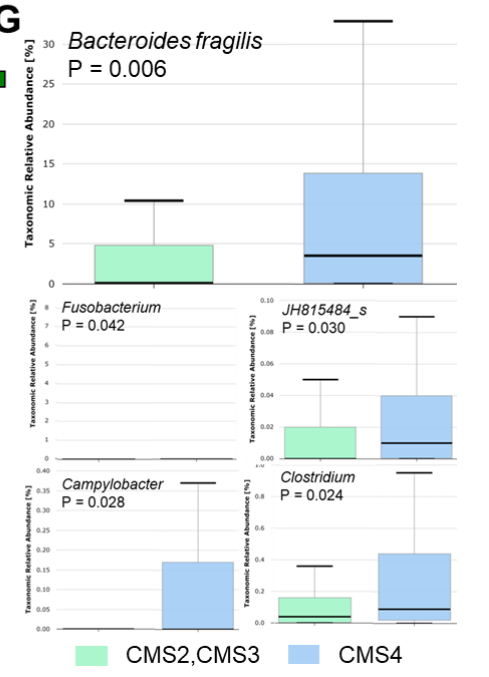


Figure 2. CMS4 tissue microbiome analysis compared to CMS2, CMS3 tissue microbiomes.

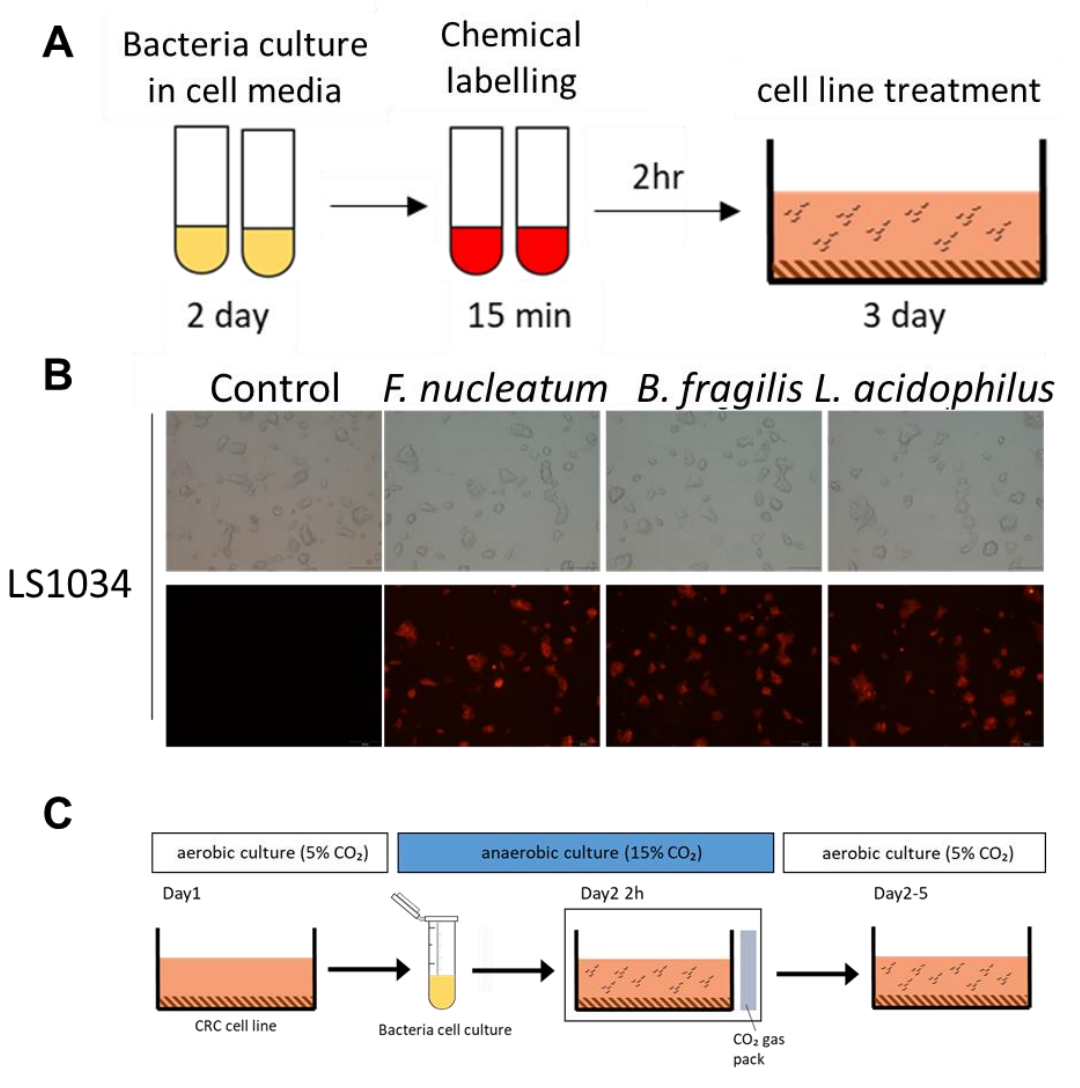
A. Simple outline of the process of identifying subtype specific microbiome. (**B-C**) Box plots show the alpha diversity richness (Chao1) and diversity index (Simpson) profiles of CMS4 tissue microbiome and CMS2 and CMS3 tissue microbiomes. **D**. Significance of beta diversity distance and similarity at species-level was analyzed using principal coordinates analysis (PCoA) with UniFrac set as the distance metric. **E**. Averaged species taxonomic compositions of the CMS2 and CMS3, and CMS4 MTP sets. **F**. Linear Discriminant Analysis (LDA) Effect Size (LEfSe) plot of differentially abundant taxonomic biomarkers identified in the gut microbiome of CMS2 and CMS3 tissues, and CMS4 tissue. The threshold of the logarithmic discriminant analysis (LDA) score was kept at 3. **G**. Relative taxonomic abundance of specific bacteria species with LDA score > 3 in CMS4 tissues. *p value < 0.05; P: p value; F: pseudo-F value; PC: principal coordinate.

3. CMS4 specific bacteria treatment of CMS2 CRC cell line resulted in increased expression of CMS4 markers

To determine whether *B. fragilis* enrichment in CMS4 tumor microbiome carried meaning, CMS2 colorectal carcinoma cell line, LS1034 was selected and treated with *B. fragilis*, *Fusobacterium nucleatum*, and *Lactobacillus acidophilus* bacteria. *F. nucleatum* and *L. acidophilus* were used as negative controls.

To confirm that the bacteria could successfully infect cells, they were labeled with a fluorescent probe and then added to the cell line for 2 hours in anaerobic conditions (Fig. 3A). Fluorescent imaging revealed that LS1034 had successfully infected with bacteria as red fluorescent signals could be seen from cells treated with bacteria. Cells not treated with bacteria did not express any fluorescence (Fig. 3B).

To verify CMS specific gene expressions in cells treated with bacteria, qPCR was carried out using known CMS2 and CMS4 markers. Cells were treated with bacteria for 2 hours in anaerobic conditions and then transferred to aerobic conditions, where they were cultured for three days. Their RNA was then isolated and tested for gene expression using CMS2/3 (CDX2) and CMS4 (FRMD6, HTR2B) markers using quantitative RT-PCR (Fig. 3C). qPCR results showed that the CMS2 cell line, LS1034, displayed significantly increased expression of CMS4 markers, FRMD6 and HTR2B, in *B. fragilis* treated cells compared to control conditions with no bacteria treatment in cells (Fig. 3D).



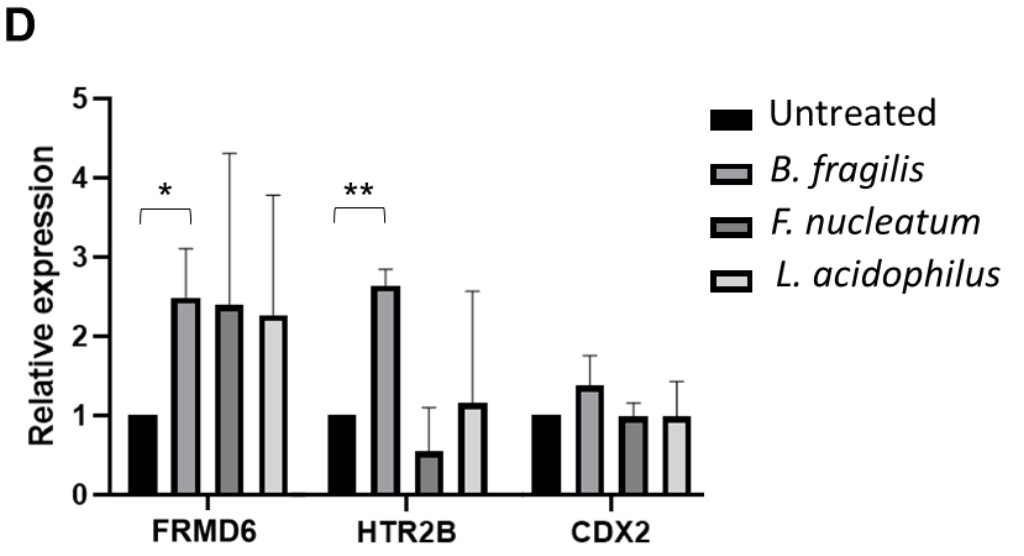
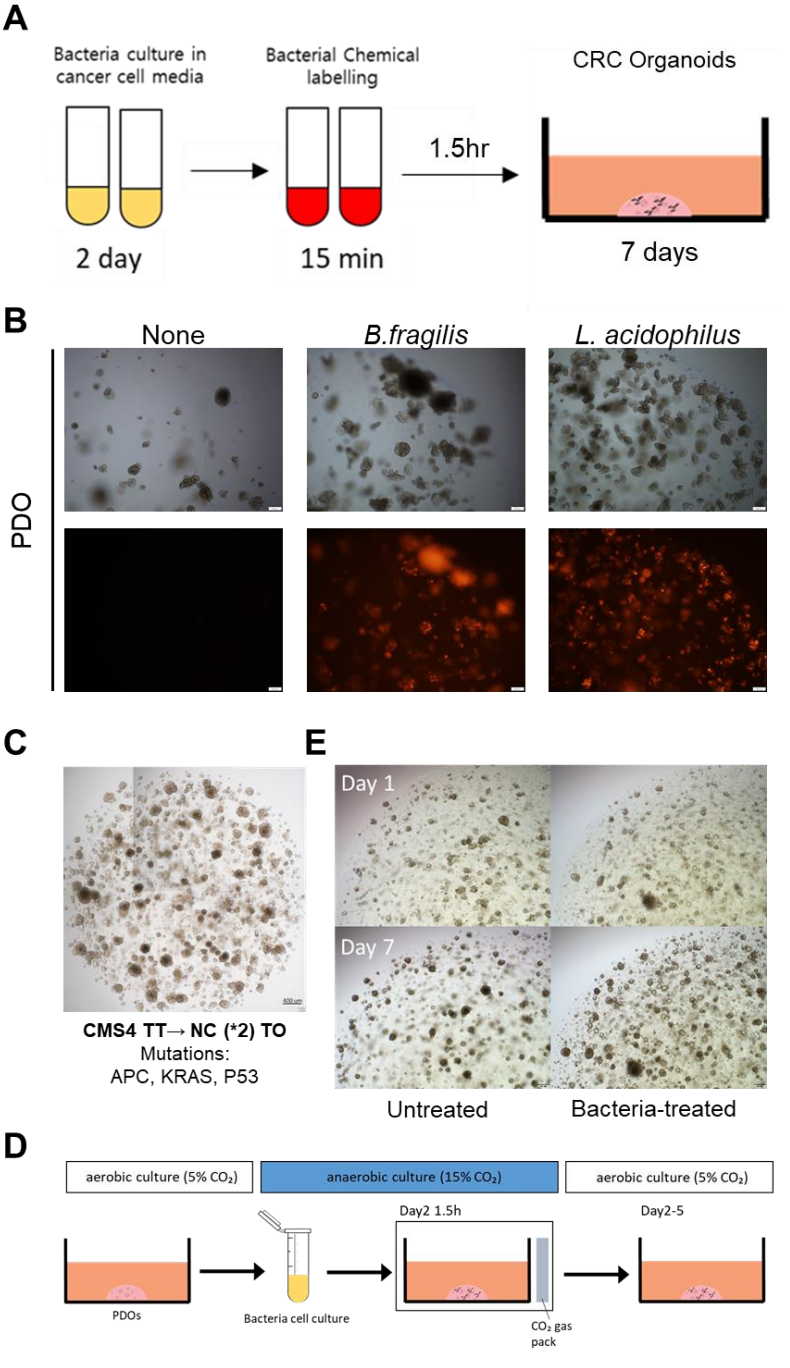


Figure 3. *B. fragilis*-treatment of LS1034 cells significantly increased their gene expression of CMS4 markers HTR2B and FRMD6.

(**A-B**) CMS2 cell line LS1034 was treated with fluorescent-labeled *F. nucleatum*, *B. fragilis* and *L. acidophilus* and imaged live under fluorescent microscope at 40x magnification. Cells not treated with bacteria were set as negative control. (**C-D**) LS1034 cells was treated with bacteria for 2 hours in anaerobic conditions and then grown in aerobic conditions for 3 days after which qPCR was performed using CMS2/CMS3 marker CDX2, CMS4 markers HTR2B and FRMD6 and housekeeping gene GAPDH. The resultant $2^{\Delta\Delta CT}$ values were plotted on a bar graph. Cells untreated with bacteria were set as negative control. Data is expressed as the mean ± standard error of three different experiments. *$p<0.05$; **$p<0.01$

4. Patient-derived organoids (PDOs) tend to have enhanced growth when treated with *B. fragilis*

To see the effect of bacteria on PDOs, PDOs (nearest predicted CMS2) derived from a CMS4 primary patient tissue were treated with *B. fragilis* and compared with untreated PDOs (control). Fluorescent-labeled *B. fragilis* verified that bacteria successfully infected PDOs as only fluorescent-labeled, bacteria-treated, cells showed fluorescence (Fig. 4A-B). PDOs treated with the bacteria for 1.5 hours and grown for 7 days showed a trend of increased cell growth and proliferation compared to untreated PDOs (Fig. 4C-E). qPCR results also showed a trend of increased CMS4 marker, FRMD6, and ZEB1, expression in 24-hour cultured PDOs, which followed the LS1034 expression pattern, although the marker expression decreased on day 7 (Fig. 4F-G)



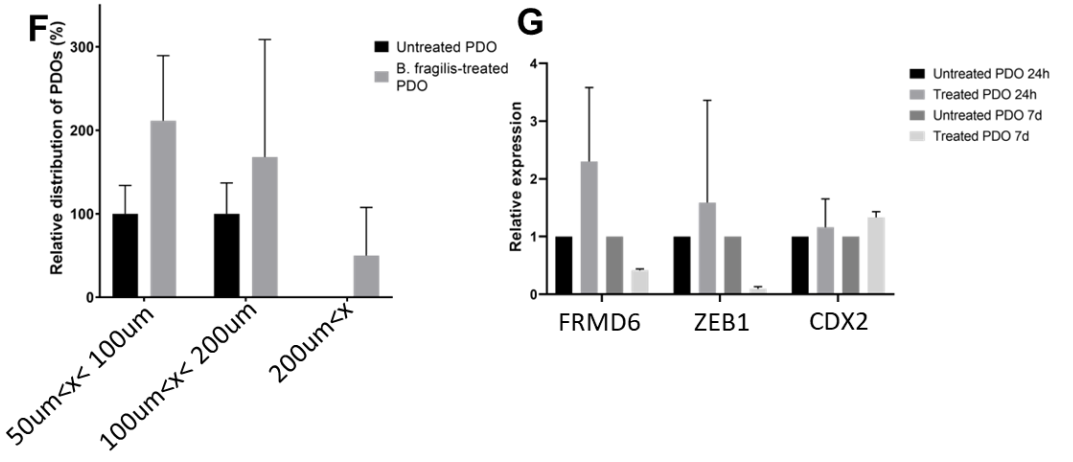


Figure 4. *B. fragilis*-treated PDOs induced change of proliferation and size of PDO.

(**A-B**) PDOs treated with fluorescent labeled *B. fragilis* and *L. acidophilus* were imaged live under a fluorescent microscope at 40x. PDOs without bacteria-treatment were set as negative control, and *L. acidophilus* was set as the positive control. (**C-E**) CMS4 tumor tissue (TT) from a patient with mutations in APC, KRAS and P53 was cultured as tumor organoids (TO) with a non-consensus (nearest predicted CMS2) subtype. They were treated with *B. fragilis* and imaged live on the microscope on different days. **F**. Relative distribution of PDOs of various sizes on day 7. **G**. Relative gene expression in day 7 PDOs was found using qPCR markers CDX2, ZEB1, and FRMD6. The bar graph shows the resultant $2^{\Delta\Delta CT}$ values that were plotted on a bar chart. All images were taken at 40x magnitude. NC: non-consensus subtype, x: PDO size.

5. CMS of *B. fragilis*-treated cells, compared to untreated cells, did not change in LS1034 and PDOs

The RNA from LS1034 and PDOs that were treated with or without *B. fragilis* was extracted and sequenced. The gene expression data from the RNA sequencing was used to find their CMS using the classify function in the R package, CMScaller. The LS1034 subtype remained constant with or without *B. fragilis* treatment and did not change in either the samples cultured for 24 hours or in samples cultured over 3 days. *B. fragilis*-treated and untreated PDOs cultured for 24 hours had the same subtype outcome when analyzed, although after 7 days, the subtype converted to CMS3. This change in CMS in day 7 PDOs was present in both bacteria-treated and control conditions. (Fig. 5).

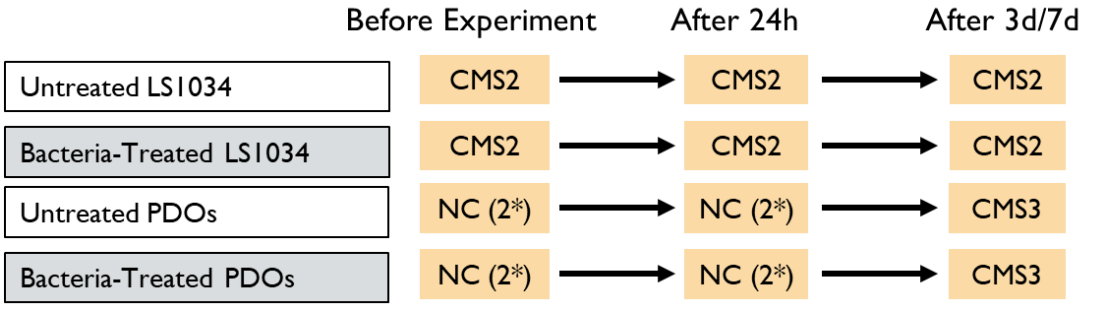


Figure 5. Change in the CMS in *B. fragilis*-treated cell lines and PDOs.

RNA sequencing data from LS1034 and PDOs was used to find their subtype. CMS of bacteria-treated and untreated samples are displayed according to predicted -or *nearest

predicted- CMS as analyzed by CMScaller. CMS was identified for all samples after 24 hours and after 3 days (LS1034) or 7 days (PDOs). NC: non-consensus subtype.

6. *B. fragilis* induced CMS4 genes in LS1034 cells but not in PDOs

Knowing the CMS of the cells and the CMS marker gene expression results using qPCR, the effect of *B. fragilis* on CMS4-related genes in a cell line or PDOs compared to untreated cells was verified using gene set enrichment analysis (GSEA). Fold change values from RNA sequencing data were input into pre-ranked GSEA, and expression levels were analyzed based on the signature CMS4 gene set obtained from CMScaller. Although LS1034 did not show any changes in its subtype after bacteria treatment, GSEA results indicated significant enrichment of CMS4 genes in 3-day bacteria-treated cells with an enrichment score (ES) of 0.169 (p value = 0.028) compared to untreated cells (Fig. 6A). Although, PDO showed a change in CMS on day 3 of culture, it was to CMS3, and this outcome was alike for both bacteria treated and non-treated PDOs. GSEA results also revealed that both 24-hour and 7-day cultured *B. fragilis*-treated PDOs were not significantly enriched in CMS4-specific genes compared to untreated PDOs (Fig. 6B).

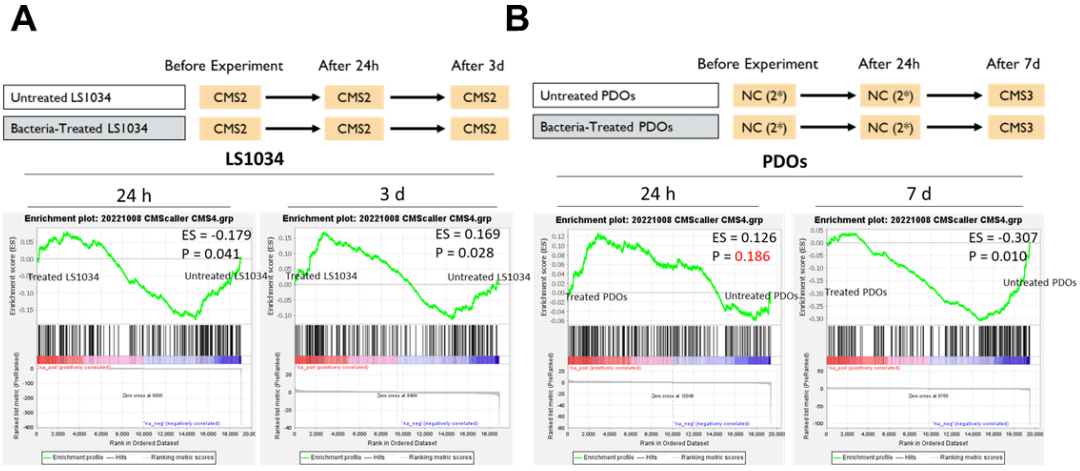


Figure 6. Change of CMS and GSEA results for CMS4 specific gene set after treatment of *B. fragilis* in LS1034 cells and PDOs.

Alongside CMS results that were found previously using CMScaller, pre-ranked GSEA analysis was performed using CMS4-specific gene set obtained from CMScaller. GSEA was performed on LS1034 and PDOs after 24 hours or 3 days (LS1034) /7 days (PDOs). The maximum and minimum sizes for gene sets were 237 and 15 and significant threshold was p value < 5%. The y-axis represents enrichment score (ES) and the x-axis shows CMS4-specific genes included in the gene set. The curve connects points of ES and genes where ES is the maximum deviation from zero and displays the degree of over-representation of a gene set at the top or the bottom of the gene list. The coloured band shows the degree of correlation of genes with *B. fragilis*-treated samples (red for positive and blue for negative correlation). Non-significant (n.s.) p values for GSEA results are written in red; NC: non-consensus subtype; *nearest predicted CMS

7. Cocultured LS1034 cell line showed increased CMS4 marker expression in *B. fragilis*-treated conditions

To better represent in vivo CRC tumors, LS1034 cell line was cocultured with TME cells (18Co: fibroblastic cell line; THP1: monocytic cell line). This coculture was treated with *B. fragilis*, *F. nucleatum*, and *L. acidophilus* in two conditions: bacteria-treated to only TME cells and bacteria-treated to LS1034 and TME cells (Fig. 7A). qPCR using RNA from samples on day 3 found that in cell cultures with bacteria-treated TME cells only, there was a significant increase in ZEB1 (CMS4 marker) expression in the *B. fragilis*-treated group compared to the control group (Fig. 7B). In the second condition, where both LS1034 and TME cells were treated with bacteria, there was a general trend of increase in gene expression of all markers to be seen for *B. fragilis*-treated cells compared to other bacteria and control. (Fig. 7C).

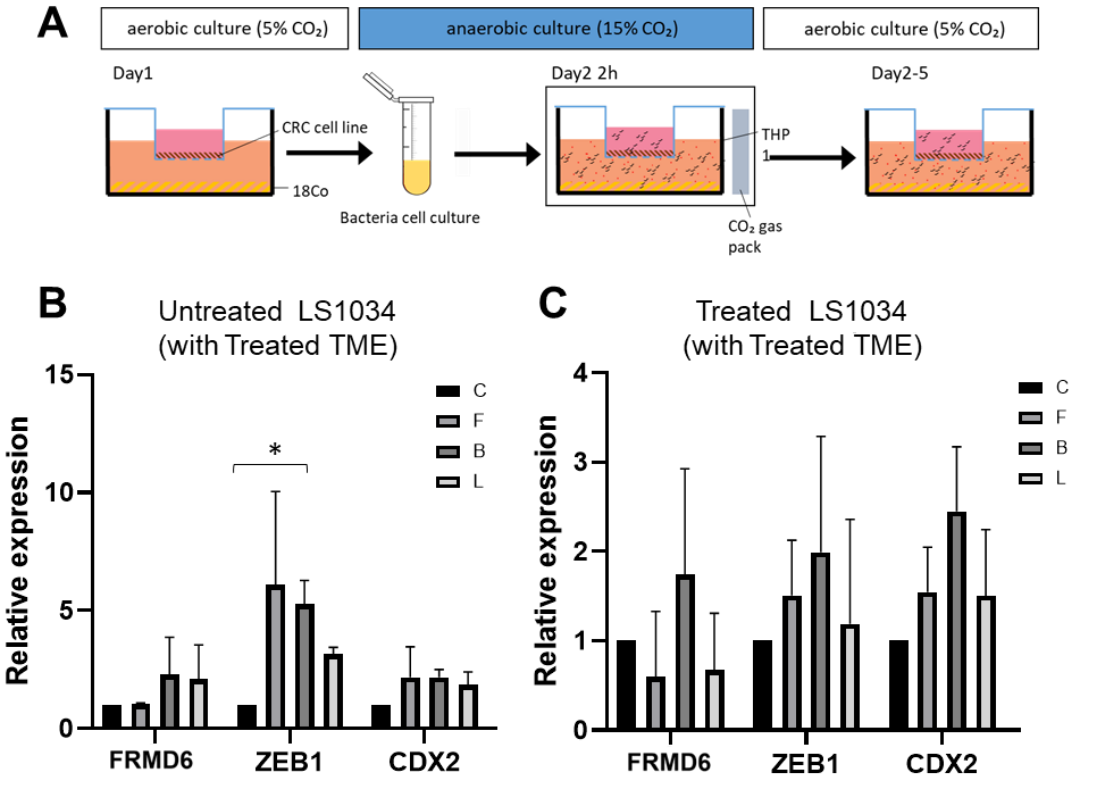


Figure 7. Change of CMS4 marker expression in LS1034 cells cocultured with 18Co and THP1 cells in *B. fragilis*-treated conditions.

LS1034 was cocultured with 18Co and THP1 in a transwell and treated with *B. fragilis*, *F. nucleatum* and *L. acidophilus* in two combinations (Untreated LS1034, Bacteria-Treated TME and Treated LS1034, Bacteria-Treated TME) and after 3 days, the RNA was harvested and qPCR was performed using biomarkers CDX2, ZEB1 and FRMD6. The results were plotted on a bar graph was plotted using $2^{\Delta\Delta CT}$ values to quantify genes expressed. Data is expressed

as the mean ± standard error of three different experiments. TME: 18Co and THP1; C: control (no bacteria treatment); B: *Bacteroides fragilis*; F: *Fusobacteria nucleatum*; L: *Lactobacillus acidophilus*.

8. *B. fragilis* induced CMS4 genes in cocultured LS1034 in *B. fragilis*-treated conditions

RNA sequencing data of cocultured LS1034 in the two bacteria treatment conditions (bacteria-treated LS1034 and stromal cells, and only stromal cells bacteria-treated) at 24 hours and on day 3 were input into the CMScaller R package to analyze their CMS. The subtype of LS1034 remained CMS2 after 24 hours in all conditions, but 3 days later, the subtype changed to CMS3 in the control condition, where both LS1034 and TME were not treated with bacteria. Cocultured LS1034 with *B. fragilis*-treated conditions had non-consensus subtype after 3 days (nearest predicted subtype: CMS1). (Fig. 8A).

After subtyping the cocultured LS1034, GSEA analysis was conducted to determine whether *B. fragilis* impacted CMS4 genes expressed in the cocultured system. The analysis showed that although the CMS of 24-hour LS1034 may not have come out as CMS4 after bacteria treatment, CMS4 genes were significantly enriched in both *B. fragilis*-treated conditions in the cocultured samples. Untreated LS1034 cells cocultured with bacteria-treated stromal cells had an ES of 0.219, and in the case of bacteria-treatment in both LS1034 and

stromal cells, there was an ES of 0.230. When CMS4 enrichment analysis was carried out to compare the two bacteria treatment conditions in 24-hour cocultured LS1034 cells, the results showed no notable difference in their CMS4 gene expressions.

For 3-day cocultured samples, untreated LS1034 cocultured with bacteria-treated stromal cells had an overrepresentation of CMS4 genes with a high ES value of 0.226. However, when *B. fragilis* was used to treat both LS1034 and stromal cells, the cells did not show an enrichment of CMS4 genes. A comparison of CMS4 gene expressions of the two *B. fragilis*-treated cocultured LS1034 conditions revealed insignificant differences (Fig. 8B).

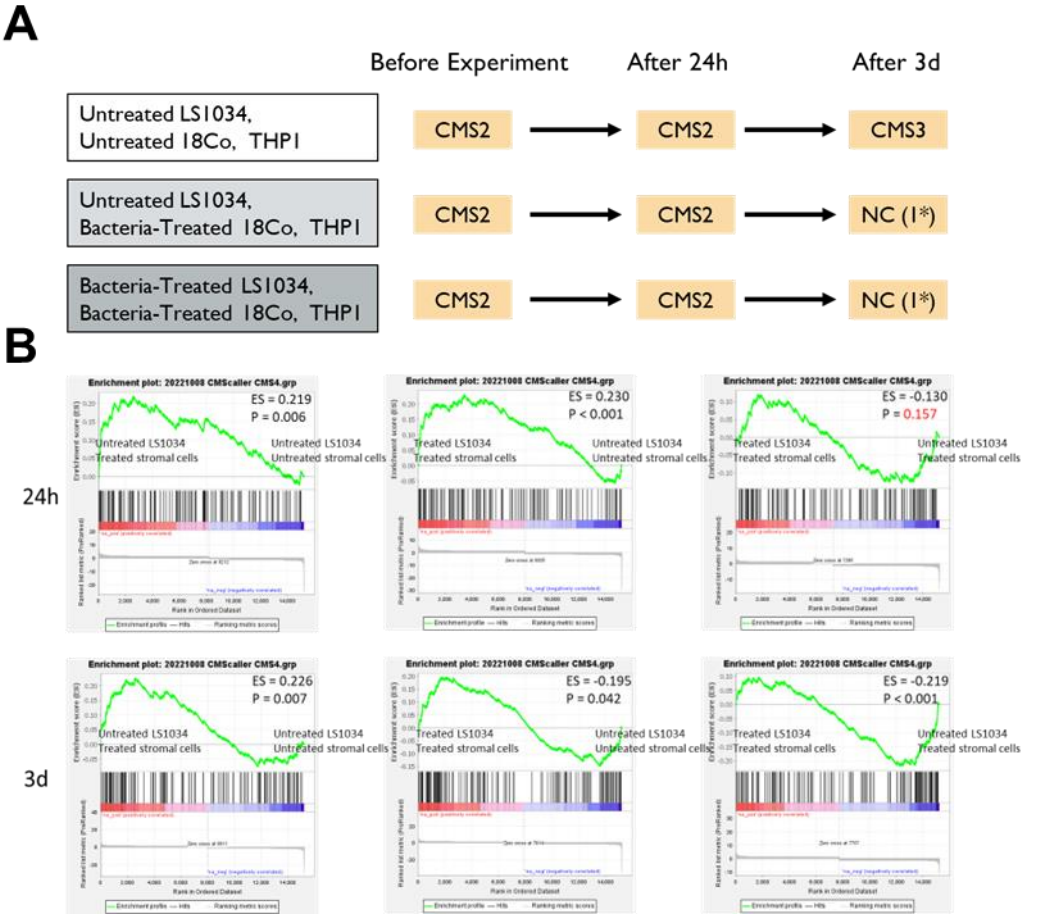


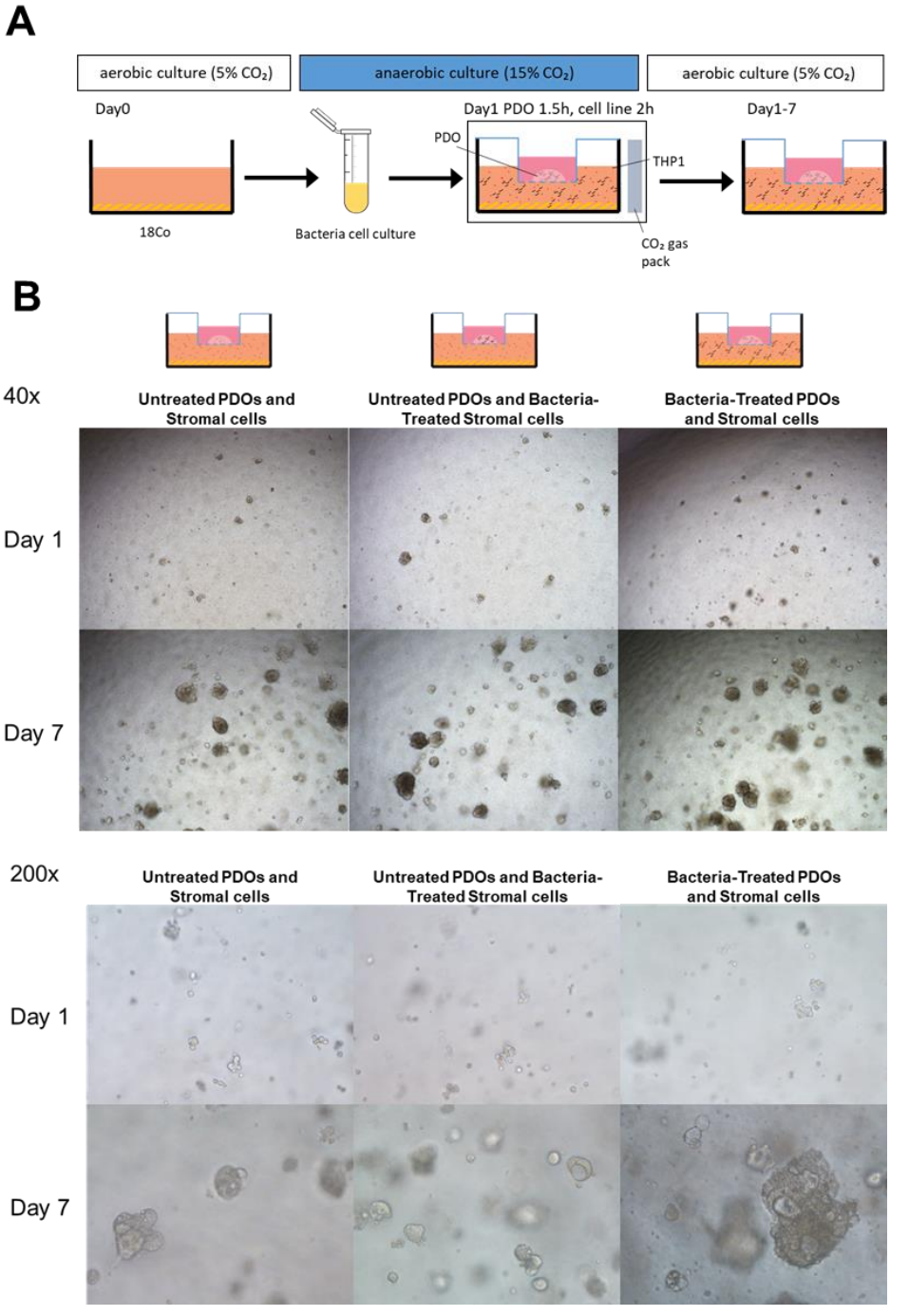
Figure 8. Change of CMS and GSEA for CMS4 specific geneset after treatment of *B. fragilis* in cocultured LS1034 cells.

A. CMS of cocultured LS1034 in the varying *B. fragilis* treatment conditions before experiment, at 24h and at 3d. **B**. Results from pre-ranked GSEA analysis using CMS4-specific gene set obtained from CMScaller. The maximum and minimum sizes for gene sets were 237 and 15 and significant threshold was p value < 0.05. NC: non-consensus subtype; *Nearest

predicted CMS; N.s. p values for GSEA are marked in red.

9. PDOs treated with *B. fragilis* and cocultured with *B. fragilis*-treated 18Co and THP1 cells showed significantly high proliferation and PDO size compared to untreated conditions

PDOs were cocultured with stromal cells (18Co, THP1) similarly to LS1034 and treated with CMS4-specific bacteria *B. fragilis* in two conditions: only stromal cells treated with bacteria and both PDO and stromal cells treated with bacteria. PDO images at 40x and 200x magnification showed larger and more number of organoids in coculture setting with bacteria-treated to both stromal and cocultured PDOs on day 7. When the size and number of organoids were manually counted on day 7, there were significantly more organoids observed in the condition where *B. fragilis* was treated to both PDOs and stromal cells compared to others. PDOs also had greater numbers for all size ranges than untreated samples ($p < 0.01$) and even had more cells in the 100 um to 200 um and >200 um range than cocultured untreated PDOs and bacteria-treated stromal cells ($p < 0.05$). Although the coculture condition where bacteria was treated to only stromal cells showed slightly larger number of organoids in the 50 um to 100 um range and 100 um to 200 um range, this difference was not significant (Fig. 9).



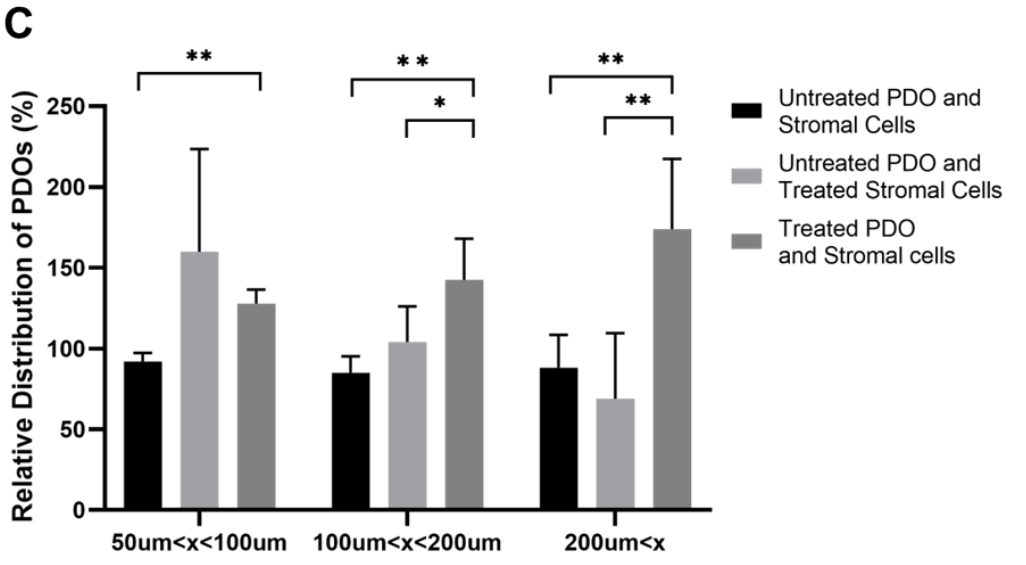


Figure 9. Cocultured PDOs in *B. fragilis*-treated conditions induced change of proliferation and size of PDOs.

PDO was cocultured with TME cells in a transwell and treated with *B. fragilis* in two combinations (Untreated PDO, Treated TME; Treated PDO, Treated TME), and after 7 days, the organoids were harvested, its distributions per bacteria-treatment condition were plotted after being individually counted and classified into three sizes: 50um to 100um, 100um to 200 um and greater than 200 um. Data is expressed as the mean ± standard error of three different experiments. Significance was calculated using non-parametric Kruskal Wallis test *p value < 0.05; **p value < 0.01

10. CMS of cocultured PDO in of *B. fragilis*-treated conditions changed to original tumor tissue subtype, CMS4, on day 7

RNA sequencing data of cocultured PDOs in the two bacteria treatment conditions (both PDO and stromal cells treated, only stromal cells treated) at 24 hours and 7 days were utilized to find their subtype at various time stages. PDOs initially of nearest predicted subtype CMS2, were cocultured with TME cells, and in the control condition with both PDO and TME cells untreated with bacteria, the CMS after 24h was unchanged. However, in *B. fragilis*-treated conditions, the subtype changed to the nearest predicted CMS4 after 24h. After 7 days of culture, control cocultured PDOs that had a subtype of CMS3, whereas cocultured PDOs in bacteria-treated conditions had CMS4 - the same subtype as the original patient tumor tissue (Fig. 10).

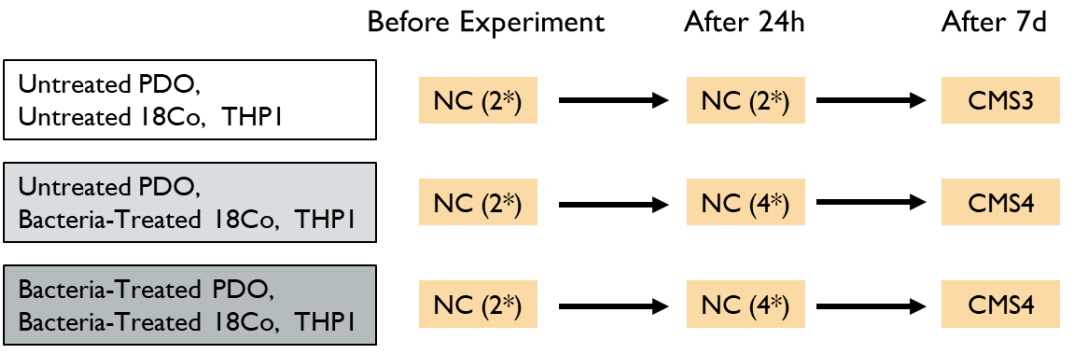


Figure 10. Change of CMS in *B. fragilis*-treated cocultured PDOs.

CMS analysis of *B. fragilis*-treated and untreated cocultured PDOs before the experiment, after 24 hours and after 7 days. NC: non-consensus subtype; *Nearest predicted CMS.

11. *B. fragilis* induced CMS4 genes in cocultured PDOs in *B. fragilis*-treated conditions

After obtaining the CMS subtyping of stromal cocultured PDOs at various time intervals, the expression of CMS4 genes in *B. fragilis*-treated coculture PDOs were examined via pre-ranked GSEA. All *B. fragilis*-treated conditions exhibited a clear overrepresentation of CMS4 genes compared to control cocultured PDOs. *B. fragilis*-treated to both PDOs and stromal cells had a higher CMS4 ES score of 0.419 than when only the stromal cells were treated with bacteria in coculture settings (ES = 0.321). In 7-day cocultured PDO samples, though, bacteria treatment of only stromal cells and not the PDOs produced a better enrichment score (ES =

0.441) of CMS4 genes than when both stromal and PDOs were treated with *B. fragilis* (ES = 0.322). When the two bacteria-treated conditions were compared against each other for CMS4 gene expression at 24 hours, the condition with bacteria-treated PDO and bacteria-treated stromal cells induced significantly more CMS4 genes. However, for 7-day cocultured samples, CMS4 enrichment was not significantly different between the two conditions (Fig. 11).

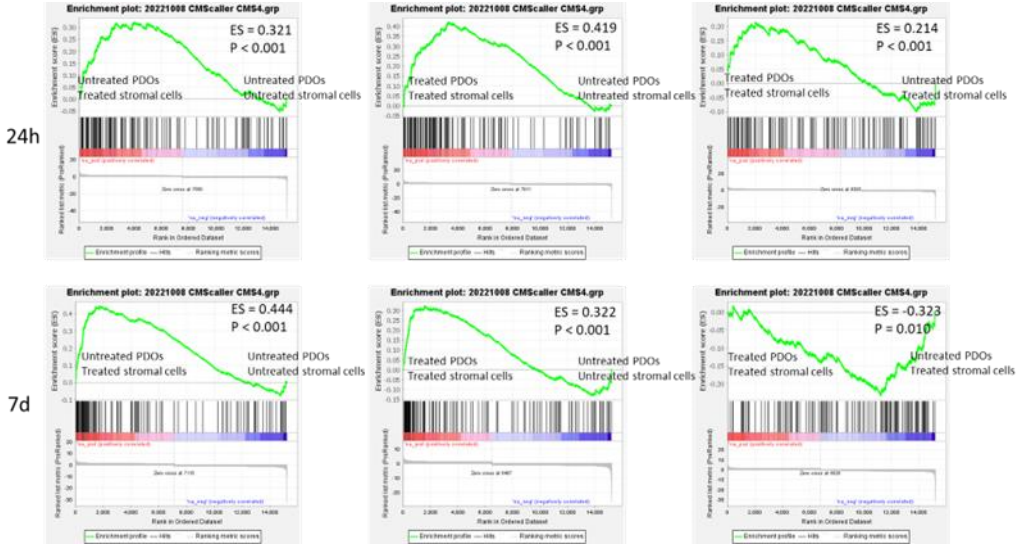


Figure 11. Gene set enrichment analysis using CMS4 specific gene set in PDOs after treatment of *B. fragilis*.

Pre-ranked GSEA analysis performed using CMS4-specific gene set on stromal cocultured PDOs at 24h and on the 7th day after bacteria treatment. Left to right: bacteria treated to only stromal cells compared to control (no bacteria treatment); bacteria treated to both PDOs and stromal cells compared to control; bacteria treated to both PDOs and stromal cells compared

to bacteria treated to only stromal cells. The maximum and minimum sizes for gene sets were 237 and 15 and significant threshold was p value < 5%.

12. Cocultured PDOs with *B. fragilis* treatment conditions are more alike original tumor tissue than untreated or mono-cultured PDOs

NMF clustering analysis that clusters samples based on their similar gene expression pattens showed that only the *B. fragilis*-treated 7-day stromal cocultured samples were in the same cluster as the original patient tissue. The heatmap made using spearman's correlation matrix and coefficient values also exhibited similarity between the 7-day *B. fragilis*-treated cocultured conditions and the original tumor tissue. 7-day cocultured PDOs with bacteria-treated stromal cells had the highest coefficient value to the original patient tumor tissue and bacteria treated to both PDO and cocultured stromal cells had the second highest coefficient value (Fig. 12).

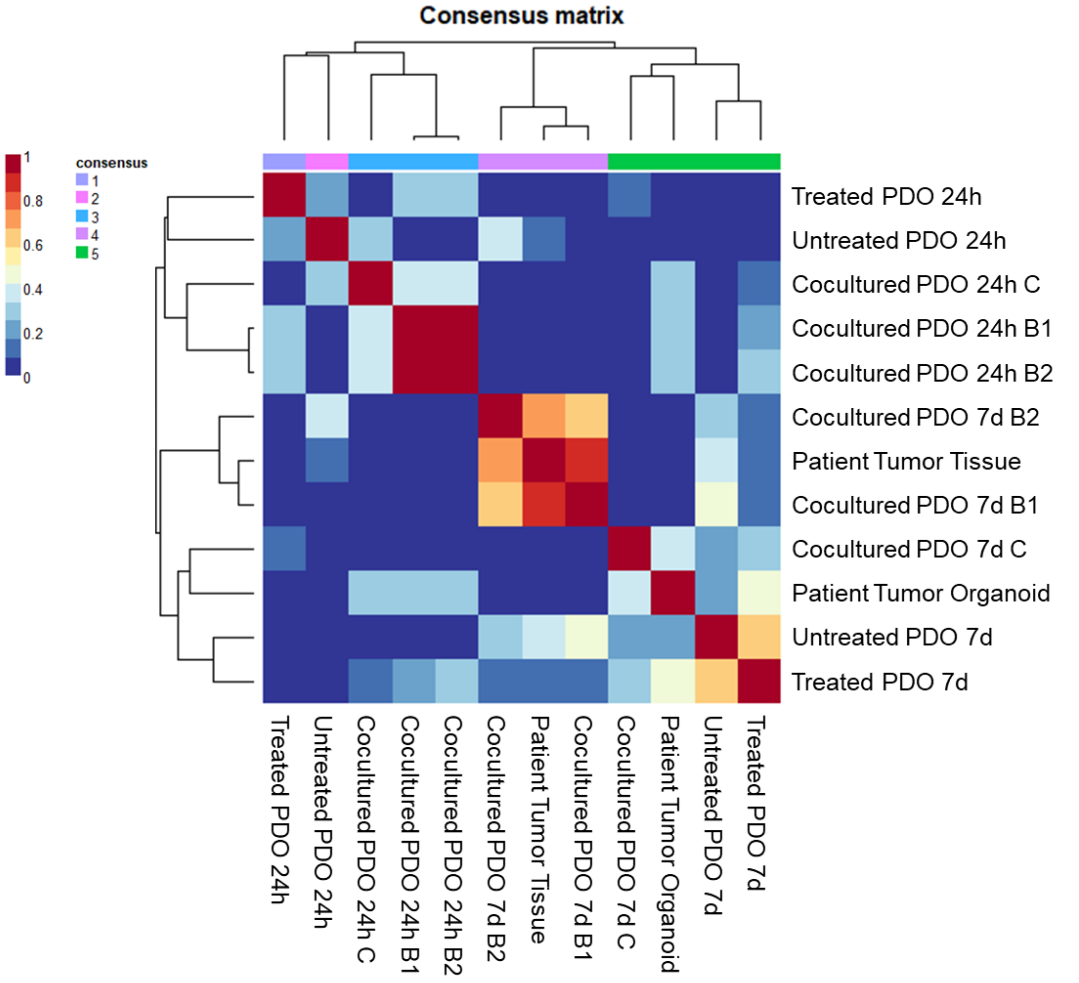


Figure 12. NMF clustering and Spearman correlation matrix heatmap of gene expression in original patient tumor tissue, monocultured and cocultured PDOs in *B. fragilis*-treated conditions.

The heat map shows Spearman's correlation matrix to visualize the correlation between the gene expression of original tumor tissue, organoids derived from this tumor before the

experiment, and after culture in various bacteria-treated conditions. The coloured bands above the heatmap show the NMF clusters (k=5) that were formed based on the sample's gene expression patterns. C: untreated PDOs, untreated stromal cells; B1: untreated PDOs, *B. fragilis*-treated stromal cells; B2: *B. fragilis*-treated PDOs, *B. fragilis*-treated stromal cells.

13. Pathways related to colorectal cancer progression are upregulated in cocultured PDO in *B. fragilis*-treated conditions

KEGG pathway analysis on PDO cocultured in the two bacteria-treated conditions showed that colorectal cancer-associated pathways such as IL-17 signaling pathway, TNF pathway, and NF-kB signaling pathways were all upregulated in the *B. fragilis*-treated conditions (Fig. 13A). Some of the top differentially expressed genes (DEGs) in *B. fragilis*-treated conditions included complement component C3, interleukin17C (IL17C) and solute carrier family 6 member 14 (SLC6A14) (Fig. 13B-C).

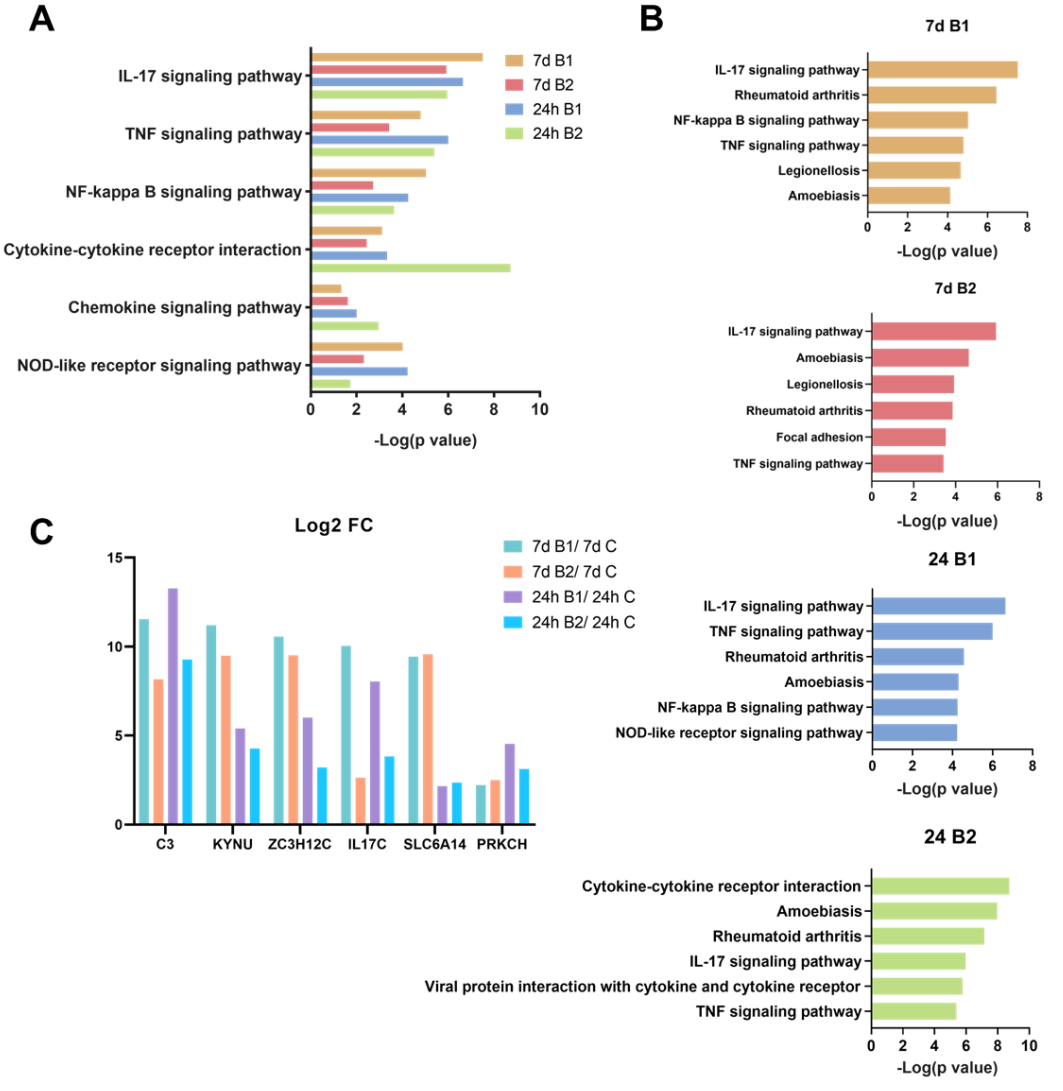


Figure 13. Pathways and genes significantly upregulated in *B. fragilis*-treated cocultured PDOs.

A. KEGG pathway enrichment analysis of top six pathways mutually upregulated in PDO cocultured *B. fragilis*-treated conditions pathway. **B**. KEGG pathway enrichment analysis of top six pathways upregulated in individual *B. fragilis*-treated conditions in cocultured PDOs. Upregulated pathways were found using DAVID database and plotted using -log (p value). **C**. Chief differentially expressed genes (DEGs) commonly expressed in *B. fragilis*-treated cocultured samples plotted using log2 fold change values. (C: untreated PDO and stromal cells; B1: untreated PDO and treated stromal cells; B2: treated PDO and treated stromal cells).

IV. DISCUSSION

Colorectal cancer is still one of the most prevalent cancers globally. Despite increasing research, diagnosing and treating CRC is still challenging due to its complex heterogenous nature as it has different consensus molecular subtypes, each responding differently to therapy. Research has come a long way in the last 20 years as we now have a better understanding of the importance of the tumor microenvironment, which has immune and stromal cells that can communicate with transformed epithelial colorectal cells for tumorigenesis. The gut microbiome has also been the centre of great interest as the dichotomy of benefit versus harm the microorganisms living in our GI tract provide to their host is explored. However, most in vitro studies still employ 2D cell line cultures or the slightly superior 3D patient-derived organoid cultures for research and drug testing. These tools of research lack TME factors such as immune cells, stromal cells and gut microbiome components and thus do not provide accurate representations of CRC subtypes such as CMS4 mesenchymal subtype that rely heavily on their TME partners for survival and spread[4,5].

Burkhardt Flemer's research investigating the difference in tumor and non-tumor microbiota revealed significant differences in the microbiome of healthy and colorectal cancer patients and found that even within cancer patients, the microbiome could differ depending on its location[20]. Purcell's paper in 2017 established that specific bacteria in the human gut microbiome have specificity for different CRC subtypes[16]. It would explain why the tumor microbiome composition in cancer patients depends on its localization. It may be due to

intratumor heterogeneity often present in advanced colorectal cancer, where the subtype is site-specific. Therefore, the microbiome is also different depending on where it is situated[21,22]. Multiple gut bacterial species have been associated with initiating or progressing CRC tumors, such as *F. nucleatum* and *B. fragilis*. Yet research regarding subtype-specific bacteria and how these bacteria contribute to CMS development has not yet been explored.

This study established the importance of CMS4-specific bacteria in colorectal cancer by identifying *Bacteroides fragilis* species, which has a clear enrichment in the tumor microbiome of CMS4 patient tissues. It further provided compelling evidence that *B. fragilis* can change the CMS of cocultured PDOs from CMS2 to CMS4 by significantly inducing CMS4-specific genes.

The patient tissue cohort used in this study consisted of similar proportions of CMS2 and CMS3, although there was a higher number of CMS4 (58%) tissues than was used in Guinney's published research on CMS (23%)[3]. CMS1 proportion in the cohort was significantly low (7%), which led to its exclusion from the microbiome studies. Our patient cohort was vastly small (115 patients) compared to that used in Guinney's paper (3104 patients), which may explain the difference in proportions. Molecular subtype classification of CRC cohort and their respective PDOs highlighted the lack of proper translation of CMS from patient tissue to tissue-derived organoids as almost all tissue subtypes had cases of CMS converting. Notably, there was a striking difference in the initial number of CMS4 tissues and the final record of CMS4 in PDOs. Approximately 90% of CMS4 tissues became

unclassifiable as a major subtype upon culture as organoids. CMS4 is a known mesenchymal subtype with signature upregulated genes, such as EMT genes, that are only upregulated in the tumor fibroblast cells[23]. PDOs grown in the absence of such TME cells experience a loss of signature signals vital for maintaining the CMS4 characteristics in tumor cells. This absence of essential signals would explain the discrepancy between the CRC patient tissue and PDO subtypes. This result also stresses the importance of TME components (immune, fibroblastic, microbial) in the CMS4 tumor and how its absence can produce drastic effects, such as changing the subtype entirely.

16s rRNA sequencing of CMS4, CMS2, and CMS3 tissues was used to find differences in the tumor microbiome of each subtype. CMS4 tumor tissue microbiome had considerably more alpha diversity than CMS2 and CMS3 microbiomes. This difference may be due to external or internal factors such as diet and the tumor's location, which can affect microbiome composition[24,25]. Examination of species-level taxonomic composition in the tissue microbiome revealed distinct differences in the bacterial populations in each MTP set. Although several bacteria were specifically enriched in CMS4 tissue microbiomes, *Bacteroides fragilis* was the most significantly enriched species in CMS4 tissue microbiome compared to CMS2/CMS3 microbiome.

Several research articles have already associated *Bacteroides fragilis* with cancer initiation and progression. It is a non-spore-forming, rod-like, aerotolerant anaerobe having two classes: non-toxigenic *B. fragilis* (NTBF) and enterotoxigenic *B. fragilis* (ETBF)[26]. The

NTBF is thought to be benign and commonly present in healthy people. On the other hand, ETBF carries the metalloprotease bft gene in its DNA that encodes for *Bacteroides fragilis* toxin (BFT). This toxin irritates the intestinal walls and causes chronic inflammation in CRC, leading to toxin-mediated and ROS-mediated DNA damage and, eventually, tumorigenesis[27]. Another mechanism by which *B. fragilis* promotes CRC development involves cleavage of the extracellular domain of E-cadherin, resulting in disruption of cell-cell adhesion, B-catenin activation, and transcription of protooncogene c-Myc[28].

I investigated this bacteria's effect on the proliferation rate, CMS, and gene expression in CRC cells by treating LS1034 cells and PDOs with it. Although the CMS of *B. fragilis*-treated LS1034 and PDO did not change to the original patient tumor subtype, CMS4, *B. fragilis*-treated, 3-day cultured, LS1034 showed increased expression of CMS4 genes in qPCR compared to untreated cells. This result was further supported by GSEA results that showed overexpression of CMS4-related genes in bacteria-treated cells compared to untreated cells. Although bacteria-treated PDOs showed a slight increase in growth and size compared to untreated PDOs, they did not show CMS4 gene enrichment at any time point when analyzed by qPCR or by GSEA. The combined results on LS1034 and PDO could be interpreted to mean that *B. fragilis* needs more than 24 hours for CMS4 gene induction in cells as none of the 24-hour treated experiments showed CMS4 enrichment, although the 3-day cultured LS1034 cells did show an effect. The lack of change in CMS also suggested that treatment of monoculture cells might not be sufficient to change the subtype of cells to CMS4.

Therefore, bacteria treatment experiments using LS1034 and PDOs cocultured with 18Co and THP1 cells (TME cells) were carried out to find the effect of bacteria in a more in vivo-like setting. 3-day cocultured LS1034 showed increased CMS4 marker ZEB1 expressions when cocultured with *B. fragilis*-treated TME cells. This bacteria treatment condition also showed a higher CMS4 enrichment result in GSEA analysis than when both the TME cells and LS1034 were treated with *B. fragilis*. 7-day cocultured PDOs in bacteria-treated conditions showed enhanced proliferation rates and significantly increased cell numbers. GSEA results of day 7 cocultured PDO samples also revealed a high CMS4 gene expression result in bacteria-treated cocultured conditions.

Among day 7 samples, CMS4 genes were most overrepresented and had the highest ES score in cocultured PDO condition where PDOs were cultured with *B. fragilis*-treated TME cells. This result was similar to 3-day cocultured LS1034 GSEA results. The notion that cocultured cells with bacteria-treated TME cells had greater CMS4 enrichment and were more affected by bacteria treatment than other conditions was further supported when the GSEA was run on upper well (cell line/PDO), bottom well (TME cells) both bacteria-treated condition against only the bottom well bacteria treated. The ES for CMS4 specific genes was found to be insignificant for both cocultured PDO (7 day) and cocultured LS1034 (3 day) cases. Therefore, although there is an enrichment of CMS4 genes in bacteria-treated PDOs cocultured with bacteria-treated stromal cells, it is not to the extent that is exhibited in cocultured untreated PDOs and bacteria-treated stromal cells.

Interestingly, in the 24-hour cocultured cells, bacteria-treated LS1034 or PDOs coupled with bacteria-treated TME cells showed higher enrichment scores for CMS4 genes. However, as there is a lack of significant CMS change after just 24 hours in both monoculture and cocultured cells, this indicated that 24 hours might not be enough for bacteria-induced CMS4 change in cocultured cells as well as monoculture cells. This may be because cells are still recovering from the hypoxic effects of being in anaerobic conditions, briefly for bacteria treatment. However, chronic inflammation that induces hypoxic conditions in vivo in CRC patients can also lead to activation of cell survival signals such as angiogenesis that perpetuate inflammation which needs to be kept in mind[29].

More significantly, only the 7-day cocultured PDOs treated with *B. fragilis* achieved CMS4 that matched the original patient tumor subtype. These PDOs had a non-consensus subtype (nearest predicted CMS2) when first cultured as organoids from CMS4 tumor tissues and only in stromal coculture settings with *B. fragilis* treated conditions did they undergo subtype change from NC to CMS4. Furthermore, NMF consensus clustering showed that 7-day stromal cocultured PDOs in *B. fragilis*-treated conditions were in the same cluster as the original patient tumor tissue. These conditions also had the highest correlation coefficient to the original patient tumor tissue. PDOs that were cocultured with bacteria-treated stromal cells specifically displayed the most similarity to the patient tissue, which also supported the earlier statement regarding this condition having the greatest effect due to *B. fragilis* treatment. This finding highlighted the importance of CMS4-specific bacteria, *B. fragilis*, in CMS4 tumors.

LS1034 cocultured cells' lack of translation to CMS4 may be due to LS1034 being a 2D cell line and unable to represent and interact with the bacteria and stromal cells like 3D PDOs that are more alike in vivo cells would[5].

KEGG pathway enrichment analysis of cocultured *B. fragilis*-treated PDO found that IL-17 signaling pathway, TNF signaling pathway, and NF-kB signaling pathway were significantly enriched, all of which are widely implicated in colorectal cancer[30,31,32]. IL-17C, one of the top upregulated DEGs, is also associated with *B. fragilis* and CRC development and studies focusing on this may provide insights into the mechanism of *B. fragilis*' induction of CMS4 genes in CRC.

The tumor microbiome and tumor microenvironment's role in tumorigenesis are well established and are being used to study tumor progression mechanisms and paths in colorectal cancer. Through detailed microbiome analysis of each subtype, I discovered that *Bacteroides fragilis* is enriched explicitly in CMS4 tumor microbiomes. In vitro experimentations using PDOs and cell lines demonstrated that *B. fragilis* induced CMS4 genes significantly in CRC cells. Cocultured PDOs that were treated with this specific bacteria changed their subtype to CMS4. Research into specific molecular markers produced by *B. fragilis* that induce these CMS4-specific genes can be used to further our understanding of the role and mechanism by which this CMS4-specific tumor bacteria develops CMS4 in CRC.

V. CONCLUSION

Bacteroides fragilis is a CMS4 specific bacteria that is significantly enriched in CRC CMS4 tumor tissue microbiomes. This species causes CMS change to CMS4 by inducing CMS4-specific genes in cocultured PDOs.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021 May;71(3):209-249.
2. Hossain MS, Karuniawati H, Jairoun AA, Urbi Z, Ooi J, John A, et al. Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies. Cancers (Basel). 2022;14(7).
3. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. Nat Med. 2015;21(11):1350-6.
4. Ten Hoorn S, de Back TR, Sommeijer DW, Vermeulen L. Clinical Value of Consensus Molecular Subtypes in Colorectal Cancer: A Systematic Review and Meta-Analysis. J Natl Cancer Inst. 2022;114(4):503-16.
5. Kim S, Choung S, Sun RX, Ung N, Hashemi N, Fong EJ, et al. Comparison of Cell and Organoid-Level Analysis of Patient-Derived 3D Organoids to Evaluate Tumor Cell Growth Dynamics and Drug Response. SLAS Discov. 2020;25(7):744-54.
6. Koh B, Jeon H, Kim D, Kang D, Kim KR. Effect of fibroblast co-culture on the proliferation, viability and drug response of colon cancer cells. Oncol Lett. 2019;17(2):2409-17.
7. Peddareddigari VG, Wang D, Dubois RN. The tumor microenvironment in colorectal carcinogenesis. Cancer Microenviron. 2010;3(1):149-66.
8. Thursby E, Juge N. Introduction to the human gut microbiota. Biochem J. 2017;474(11):1823-36.
9. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. Nat Biotechnol. 2014;32(8):834-41.

10. Hugon P, Dufour JC, Colson P, Fournier PE, Sallah K, Raoult D. A comprehensive repertoire of prokaryotic species identified in human beings. Lancet Infect Dis. 2015;15(10):1211-9.
11. Cheng Y, Ling Z, Li L. The Intestinal Microbiota and Colorectal Cancer. Front Immunol. 2020;11:615056.
12. Pickard JM, Zeng MY, Caruso R, Núñez G. Gut microbiota: Role in pathogen colonization, immune responses, and inflammatory disease. Immunological Reviews. 2017;279(1):70-89.
13. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science. 2012;338(6103):120-3.
14. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell host & microbe. 2013;14(2):207-15.
15. Wu S, Rhee KJ, Albesiano E, Rabizadeh S, Wu X, Yen HR, et al. A human colonic commensal promotes colon tumorigenesis via activation of T helper type 17 T cell responses. Nat Med. 2009;15(9):1016-22.
16. Purcell RV, Visnovska M, Biggs PJ, Schmeier S, Frizelle FA. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. Sci Rep. 2017;7(1):11590.
17. Eide PW, Bruun J, Lothe RA, Sveen A. CMScaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models. Scientific Reports. 2017;7(1):16618.
18. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67(5):1613-7.

19. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005;102(43):15545-50.
20. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumor-associated and non-tumor-associated microbiota in colorectal cancer. Gut. 2017;66(4):633.
21. Marisa L, Blum Y, Taieb J, Ayadi M, Pilati C, Le Malicot K, et al. Intratumor CMS Heterogeneity Impacts Patient Prognosis in Localized Colon Cancer. Clin Cancer Res. 2021;27(17):4768-80.
22. Saito T, Niida A, Uchi R, Hirata H, Komatsu H, Sakimura S, et al. A temporal shift of the evolutionary principle shaping intratumor heterogeneity in colorectal cancer. Nature Communications. 2018;9(1):2884.
23. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nature Genetics. 2017;49(5):708-18.
24. Ott SJ, Musfeldt M, Wenderoth DF, Hampe J, Brant O, Fölsch UR, et al. Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. Gut. 2004;53(5):685-93.
25. Hoang T, Kim MJ, Park JW, Jeong S-Y, Lee J, Shin A. Nutrition-wide association study of microbiome diversity and composition in colorectal cancer patients. BMC Cancer. 2022;22(1):656.
26. Sears CL, Geis AL, Housseau F. *Bacteroides fragilis* subverts mucosal biology: from symbiont to colon carcinogenesis. J Clin Invest. 2014;124(10):4166-72.
27. Purcell RV, Pearson J, Aitchison A, Dixon L, Frizelle FA, Keenan JI. Colonization with enterotoxigenic *Bacteroides fragilis* is associated with early-stage colorectal neoplasia. PLoS One. 2017;12(2):e0171602.

28. Wu S, Morin PJ, Maouyo D, Sears CL. *Bacteroides fragilis* enterotoxin induces c-Myc expression and cellular proliferation. Gastroenterology. 2003;124(2):392-400.
29. Zeitouni NE, Chotikatum S, von Köckritz-Blickwede M, Naim HY. The impact of hypoxia on intestinal epithelial cell functions: consequences for invasion by bacterial pathogens. Molecular and Cellular Pediatrics. 2016;3(1):14.
30. Razi S, Baradaran Noveiry B, Keshavarz-Fathi M, Rezaei N. IL-17 and colorectal cancer: From carcinogenesis to treatment. (1096-0023 (Electronic)).
31. Zhao P, Zhang Z. TNF-α promotes colon cancer cell migration and invasion by upregulating TROP-2. Oncol Lett. 2018;15(3):3820-7.
32. Patel M, Horgan PG, McMillan DC, Edwards J. NF-κB pathways in the development and progression of colorectal cancer. Translational Research. 2018;197:43-56.

ABSTRACT (IN KOREAN)

대장암 분자 아형에 대한 종양미세환경과 세균의 영향

<지도교수 김태일>

연세대학교 대학원 의과학과

장신영

대장암(colorectal cancer, CRC)의 CMS(consensus molecular subtype) 4 아형은 간질 함량이 높고 생존율이 가장 낮은 중간엽 간질세포 특성의 유형이다. 그러나 그 주요 인자는 아직 밝혀지지 않고 있다. 장내 미생물군집은 대장암의 종양 미세환경(tumor microenvironment, TME)의 필수적인 부분이기 때문에 장내 세균이 CMS4 아형을 유도하는 데 중요한 역할을 한다고 가정하였다.

환자 종양 조직(n=115) 및 해당 종양의 환자유래 종양 오가노이드(patient-derived organoid, PDO)를 이용하여 유전자 발현 및 CMS 아형을 식별하기 위해 RNA 시퀀싱과 CMScaller 분석을 수행하였다.

또한 해당 대장암 조직의 16S-rRNA RNA 시퀀싱과 Microbiome Taxonomic Profiling(MTP) 분석을 통하여 동일한 대장암 환자의 조직에서 CMS2 또는 CMS3 아형에 비교하여 CMS4 아형에 특이적인 박테리아를 찾았다. 발굴된 박테리아의 CMS 아형 결정에 대한 역할 증명을 위해 CMS2 아형의 대장암 세포주(LS1034)와 대장암 조직에서는 CMS4 아형으로 분석되었지만 해당 환자의 PDO 에서는 CMS2 아형을 보인 대장암 환자의 PDO 를 *in vitro* 실험에 사용하였으며, 이 들 종양세포와 함께 섬유세포(18Co) 및 염주세포(THP1)의 Transwell 공배양(co-culture)을 이용하여 TME 환경을 조성하였다. 조합 1 은 TME 세포와 PDO/세포주 모두에 해당 세균 감염을 시켰고, 조합 2 는 TME 세포에만 세균 감염을 시켰다. PDO 는 1.5 시간 동안 혐기성 조건에서 세균 감염을 시켰고, LS1034 세포 및 TME 세포들은 2 시간 동안 같은 조건에서 세균 감염 후 호기성 조건에서 배양하였다. CMS4 특이적 유전자군의 발현 변화 분석을 위해 유전자 세트 농축 분석(GSEA)은 이용하였다.

대장암 조직의 MTP 분석에서 다른 CMS 아형(CMS2 및 CMS3)에 비해 CMS4 아형 조직의 미생물군집에서 *Bacteroides fragilis* 가 4.30(p value= 0.008)의 가장 큰 LDA 값을 보였다. TME 세포와 공배양 없이 배양된 LS1034

및 PDO 는 CMScaller 결과에서 변함없이 CMS2 아형을 그대로 보였다. TME 세포와 공배양된 PDO 는 24 시간 배양 후 CMS2 아형으로 변함이 없었고, 7 일 후에는 CMS3 아형으로 변화하였다. *B. fragilis* 가 처리된 TME 공배양 PDO 는 배양 7 일 후에 CMS4 아형으로 변화하였다. GSEA 분석에서는 *B. fragilis* 를 처리하지 않은 TME 공배양 PDO 에 비교하여, *B. fragilis* 를 처리한 TME 공배양 PDO 에서 CMS4 특이 유전자군의 발현이 유의하게 증가함을 확인하였다 (ES = 0.294, p <0.001). 또한, KEGG 경로분석은 *B. fragilis* 이 감염된 TME 공배양 PDO 에서 많은 암 관련 염증신호전달 경로가 증가함을 발견하였다. 이상의 결과로 *B. fragilis* 가 대장암의 CMS4 아형에서 유의하게 증가됨을 발견하였고, *B. fragilis* 가 TME 공배양 PDO 모델에서 CMS4 유전자군의 발현을 유의하게 증가시킴을 확인하였다.

핵심되는 말: colorectal cancer, consensus molecular subtype, microenvironment, microbiome, *Bacteroides fragilis*