

RESEARCH

Open Access



# Analysis of changes in microbiome compositions related to the prognosis of colorectal cancer patients based on tissue-derived 16S rRNA sequences

Sukjung Choi<sup>1</sup>, Jongsuk Chung<sup>2</sup>, Mi-La Cho<sup>3</sup>, Donghyun Park<sup>2</sup> and Sun Shim Choi<sup>1\*</sup> 

## Abstract

**Background:** Comparing the microbiome compositions obtained under different physiological conditions has frequently been attempted in recent years to understand the functional influence of microbiomes in the occurrence of various human diseases.

**Methods:** In the present work, we analyzed 102 microbiome datasets containing tumor- and normal tissue-derived microbiomes obtained from a total of 51 Korean colorectal cancer (CRC) patients using 16S rRNA amplicon sequencing. Two types of comparisons were used: 'normal versus (vs.) tumor' comparison and 'recurrent vs. nonrecurrent' comparison, for which the prognosis of patients was retrospectively determined.

**Results:** As a result, we observed that in the 'normal vs. tumor' comparison, three phyla, Firmicutes, Actinobacteria, and Bacteroidetes, were more abundant in normal tissues, whereas some pathogenic bacteria, including *Fusobacterium nucleatum* and *Bacteroides fragilis*, were more abundant in tumor tissues. We also found that bacteria with metabolic pathways related to the production of bacterial motility proteins or bile acid secretion were more enriched in tumor tissues. In addition, the amount of these two pathogenic bacteria was positively correlated with the expression levels of host genes involved in the cell cycle and cell proliferation, confirming the association of microbiomes with tumorigenic pathway genes in the host. Surprisingly, in the 'recurrent vs. nonrecurrent' comparison, we observed that these two pathogenic bacteria were more abundant in the patients without recurrence than in the patients with recurrence. The same conclusion was drawn in the analysis of both normal and tumor-derived microbiomes.

**Conclusions:** Taken together, it seems that understanding the composition of tissue microbiomes is useful for predicting the prognosis of CRC patients.

**Keywords:** Colorectal cancer, Microbiome, 16S rRNA, *F. nucleatum*, *B. fragilis*

## Introduction

Colorectal cancer (CRC), like many other cancers, is a malignant disease that occurs as a result of the accumulation of complex genetic and epigenetic changes. Although it has been reported that the majority of CRC cases (~80%) are due to nongenetic or epigenetic changes and less than 20% of CRC cases are caused by genetic mutations [1], these two types of risks are actually

\*Correspondence: [schoi@kangwon.ac.kr](mailto:schoi@kangwon.ac.kr)

<sup>1</sup> Division of Biomedical Convergence, College of Biomedical Science, Institute of Bioscience & Biotechnology, Kangwon National University, Chuncheon 24341, Republic of Korea  
Full list of author information is available at the end of the article



entangled in a very complicated way that makes it almost impossible to differentiate between the upstream driver risk and the downstream passenger risk in causing CRC. Genetically, alterations in *Wnt* signaling pathways initiated by *APC* mutation are known as one of the common causes of familial types of CRC [2, 3]. On the other hand, smoking cigarettes, diets rich in red meats and processed foods, and drinking alcohol have frequently been linked to nongenetic environmental risk factors for CRC [4, 5], and the microbiota has recently been added to that list. The microbiome has been proven in several studies to be a mediator between genetic mutations and harmful diets in the onset and progression of CRC.

It is known that abnormal changes in the composition of the gut microbiome can lead to disruption of epithelial barrier function, which increases inflammation and in turn leads to various gastrointestinal diseases, including CRC [6, 7]. In fact, it has been reported that the distribution of the microbiome differs significantly between normal and tumor tissues or between normal and cancer fecal samples, mainly due to dysbiosis of the microbiome under tumorigenic conditions. For instance, pathogenic bacteria such as *Bacteroides (B.) fragilis* and *Fusobacterium (F.) nucleatum* were significantly more enriched in the tumor tissue than in the normal tissue, and conversely, nonpathogenic members of the Bacteroidetes and Firmicutes phyla were more abundant under normal conditions than tumorigenic conditions for both tissue samples and fecal samples [8, 9].

Pathogenic bacteria are known to directly or indirectly cause enhanced inflammation and oxidative DNA damage and even stimulate cancer-causing signaling pathways inside the cell [10–13]. Particularly, according to Strauss et al. [13], *Fusobacteria* can invade colonic epithelial cells, destroying the epithelial barrier that allows CRC cells to survive or be maintained. In addition, some studies have shown that *F. nucleatum* can activate *Wnt/β-catenin* signaling, promoting cell proliferation and inflammation, through binding of its *FadA* adhesion protein to E-cadherin on the surface of colon cells [14–16], or through activating *TLR4* signaling to *NF-κB* [17]. Likewise, another pathogenic bacterium abundant in CRC, *B. fragilis*, also known as an enterotoxin-producing bacterium, can take part in multistep tumorigenesis by producing toxins. Toxins are known to induce E-cadherin degradation, causing downstream  $\beta$ -catenin signaling, and to stimulate the release of reactive oxygen species and the expression of inflammatory cytokines that cause DNA damage [18–20].

Tjalsma et al. [21] proposed a ‘driver-passenger’ model to explain how the microbiome can facilitate CRC tumorigenesis. According to the model, driver pathogenic bacteria induce DNA damage in colon epithelial cells,

leading to the initiation of tumorigenesis. Damaged epithelial cells in turn change the surrounding tumor microenvironment such that opportunistic bacteria (i.e., passenger bacteria) with a competitive advantage in this altered tumor microenvironment defeat and replace healthy gut bacteria, eventually worsening inflammation and accelerating cell proliferation, promoting tumorigenesis. It is, however, worth noting that thus far, no single bacterial species has universally been associated with all CRC patients because substantial variations are present in the compositions of microbiota associated with CRC [22, 23]. It seems that changes in both pathogenic and nonpathogenic microbiomes are responsible for the initiation and/or progression of CRC.

In the present work, using the microbiome information estimated from 16S rRNA amplicon sequencing data generated from matched samples of CRC patients, including tumors and adjacent normal tissues derived from the same patient, we investigated compositional changes in microbiomes related to the tumorigenesis of CRC. We also investigated the compositions of microbiomes between nonrecurrent CRC (named ‘crc\_nRC’) and recurrent CRC (named ‘crc\_RC’), revealing the bacterial population associated with the relapse of CRC.

## Materials and methods

### Sample collection and generation of 16S rRNA sequences

A total of 51 matched normal and tumor samples obtained from the same individuals with CRC (mostly at TNM stage 2 and 3, see Additional file 1: Table S1, aged 43–86, 51 males collected from the cecum to the rectum at the Samsung Medical Center in Seoul, Republic of Korea) who underwent resection surgery were used for producing the host RNA-seq data and 16S rRNA data to investigate gene expression patterns and the composition of microbiomes that the CRC tissues carry. The RNA-seq data and the V3–V4 amplicon sequencing data of 16S rRNAs were obtained with an Illumina MiSeq reagent kit v3 (2 × 300 bp, Illumina, USA). The PCR primers, i.e., forward (CCTACG GGNGGCWGCAG) and reverse (GACTACHVGGG TATCTAATCC), were designed from the hypervariable regions (V3–V4) of 16S rRNAs. PCR was conducted using 2 × KAPA HiFi HotStart ReadyMix (Roche) under the following conditions: 95 °C solution chain for 3 min, 25 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 45 s, followed by a 72 °C extension for 5 min. Sequencing libraries were then constructed using a TruSeq<sup>®</sup> DNA PCR-Free Sample Preparation Kit (Illumina, USA) and TruSeq<sup>®</sup> Nextera XT index primer (Illumina, USA), and 2 × KAPA HiFi HotStart ReadyMix (Roche) using the PCR products after purification. Subsequently, paired-end reads were generated

by sequencing on the MiSeq platform after determining the quality of the library with the TapeStation 4200 platform (Agilent Technologies) and a Qubit Fluorometer (Thermo Fisher Scientific).

#### **Analysis of the microbiomes using bioinformatics tools**

The sequencing reads were selected by filtering out low-quality sequences, including primer sequences, truncated sequences, and sequences that were classified into Eukarya and Archaea lineages, following a previously reported QIIME (v1.9.1) quality control process [24]. After finishing quality control procedures, an average of 188,342 high-quality reads per sample (median 190,643; range 128,130–256,314) were obtained, where the average length and quality score were 268.1 bp and 33.01, respectively. Then, the paired-end reads were assembled using the Fast Length Adjustment of SHort reads (FLASH) [25] tool, and chimeric sequences were also excluded by matching the clean tag sequences to the reference database using the Usearch software v6.1 algorithm [26]. Eventually, an average of 139,572 clean reads per sample (range 93,441–208,822) were obtained after filtering chimeric reads.

All the cleaned sequences were used for clustering analysis that led to the identification of operational taxonomic units (OTUs) after removing singleton OTUs. The taxonomic rank (i.e., phylum, class, order, family, genus, and species) of each sample was determined using the Ribosomal Database Project (RDP) classifier [27] by aligning the sequence to the GreenGene reference database (release 13.8) [28] at a 97% minimum similarity level. The final OTU table was used to generate a taxonomic profile graph by including only taxa with at least 0.1% relative abundance in each group. See Additional file 2: Table S2: the OTU table used in this study. The compositional characteristics of the microbiomes differentially enriched in normal and tumor tissues were investigated by linear discriminant analysis effect size (LEfSe) [29].

#### **Estimation of $\alpha$ - and $\beta$ -diversity**

The  $\alpha$ -diversity was evaluated by the Shannon index and observed OTUs with QIIME software, while  $\beta$ -diversity was estimated by principal coordinate analysis (PCoA) based on the Bray–Curtis distance [30]. Permutational multivariate analysis of variance (PERMANOVA) as implemented by the ‘*adonis*’ function in the R package ‘Vegan’ was applied to test the microbial composition between groups. The box plots and diagrams for these analyses were constructed with the ‘ggplot2’ package in R

(v3.6.2). All statistical significance tests were performed with the ‘Wilcoxon rank-sum’ test using the R package.

#### **Prediction of metabolic pathways based on the composition of microbiomes**

To predict the functions of bacteria, software called ‘PICRUSt’, i.e., an acronym for ‘phylogenetic investigation of communities by reconstructing of unobserved states’, was used, the main procedures for which were well described previously [31]. The metabolic functions were estimated by mapping the composition of the identified bacteria into the KEGG database. Statistical Analysis for Metagenomic Profiles (STAMP) [32] was used to identify different metabolic functional abundances between groups. A corrected  $P$ -value  $< 0.05$  was considered to be significant.

#### **Estimation of differentially expressed genes**

After the quality of sequencing reads was determined by FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), the low-quality (Phred score  $< 33$ ) and adaptor sequences were removed by Trimmomatic (v0.39) [33]. The reference genome (GRCh38/hg38) was then indexed by STAR (v2.7.6a) [34]. Subsequently, the cleaned reads were mapped to the indexed reference genome using STAR, following previously reported procedures [35, 36]. The count value for each gene was then estimated using ‘htseq-count’ [37] after gene names were assigned for the mapped reads by the ‘GTF’ file of the ‘GENCODE Gene Set’ (release 30) ([https://www.gencodegenes.org/human/release\\_30.html](https://www.gencodegenes.org/human/release_30.html)). Finally, differentially expressed genes (DEGs) estimated by comparing gene expression between normal and tumor conditions were identified using ‘DESeq2’ [38] after the read counts were normalized. Two thresholds, an adjusted  $P$ -value (i.e.,  $Q$ -value)  $< 0.01$  and  $|\log_2$ fold change (fc)|  $> 1$  (i.e.,  $\text{abs}(\log_2\text{fc})$ ), were applied to estimate DEGs by comparing gene expression levels between tumors and normal tissues. Principal component analysis (PCA) revealed that tumor and normal samples were clearly distinguished. However, one non-recurrent sample (10003704) was revealed to be an outlier and was removed from later analysis.

#### **Gene set enrichment analysis and cellular heterogeneity of host genes**

Two annotation methods were used for the analysis of DEGs. (i) The single-sample GSEA (ssGSEA) method, an extension of gene set enrichment analysis (GSEA), was used to calculate separate enrichment scores for each pairing of a sample and gene set. (ii) A cell type deconvolution tool, xCell, was used to analyze cellular

heterogeneity between tumors and normal tissues. Subsequently, the specific gene set enrichment score and deconvoluted cell composition information were used for correlation analyses with microbial compositions.

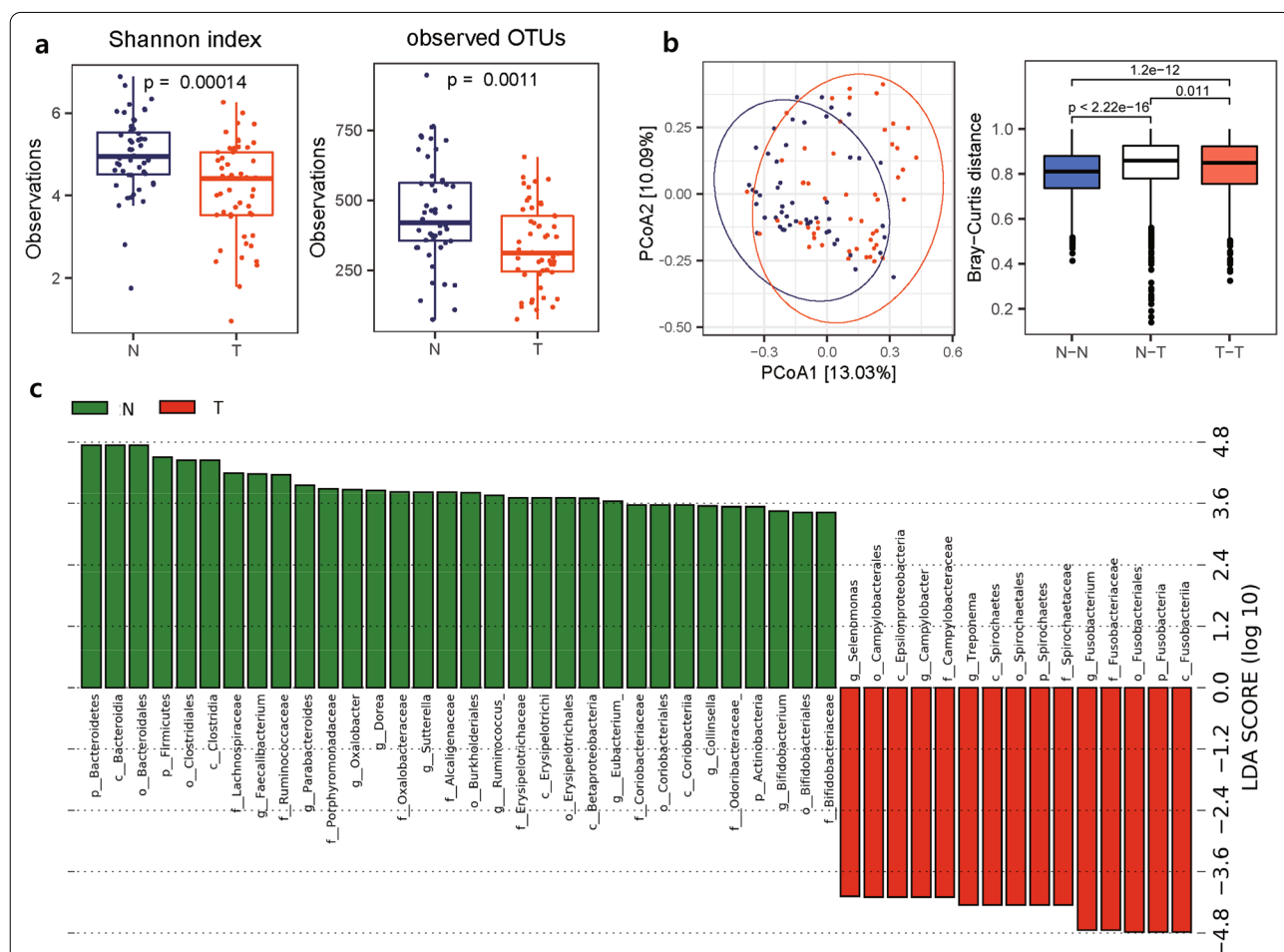
**Results**

**Tumors had a lower bacterial diversity than normal tissues**

From the 16S rRNA amplicon sequencing data obtained from a total of 51 CRC patients with matched normal and tumor tissues, we attempted to identify microbial communities and to estimate their diversity and abundance. We found that  $\alpha$ -diversity in tumor tissues was significantly lower than that in normal tissues (Fig. 1a), indicating that the number of inhabiting bacterial species is significantly reduced in the location where tumor formation and progression occur. A possible explanation for this observation is that the tissue environment affected by dysbiosis of the microbiota can be detrimental to some

healthy bacteria. The  $\beta$ -diversity estimated using PCoA plots also indicated that the bacterial population structure in tumor tissues was distinct from that in normal tissues, with a significant Bray–Curtis distance ( $R^2 = 0.039$ ,  $p = 0.001$ ) (Fig. 1b). Relatedly, using unsupervised hierarchical clustering, we examined whether the population structure of the microbiome was similar between normal tissues and tumors in the same patient or between normal tissues or tumors in different patients (Additional file 1: Fig. S1). Although the clustering pattern was complex in that the clusters in the dendrogram were mixed patterns supporting the former or the latter scenario, the similarity of microbial population structures seemed to be higher between normal or tumor tissues in different patients than between normal and tumor tissues in the same patient.

Subsequently, LefSe, i.e., linear discriminant analysis (LDA) effect size, was performed to investigate



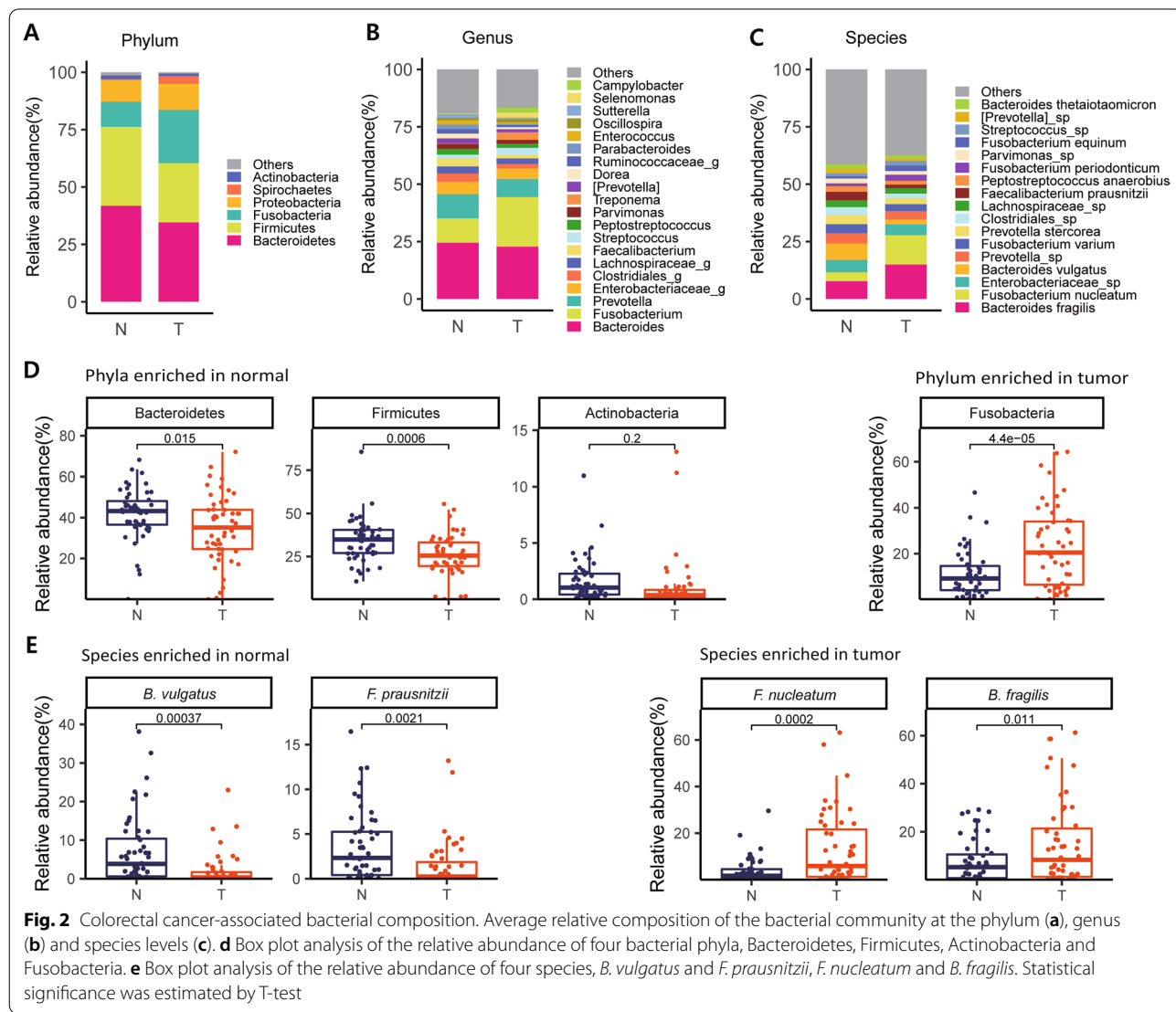
**Fig. 1** Bacterial diversity of normal and tumor tissues in CRC patients. **a** The  $\alpha$ -diversity estimated by the Shannon index and observed OTUs. **b** (left) The  $\beta$ -diversity estimated using PCoA of OTUs. (right) Distribution of Bray–Curtis distances of OTUs in normal samples (N–N), tumor samples (T–T), and normal and tumor samples (N–T). **c** LefSe plot illustrating microbial taxa enriched in normal compared with CRC tumor tissues. N: normal, T: tumor

differentially abundant microbiome features (clades, OTUs, etc.) in normal and tumor tissues, and it was confirmed that normal-enriched microbiome features were distinct from tumor-enriched microbiome features (Fig. 1c, Additional file 1: Fig. S2). Namely, in the differential microbiome features ranked by effect size, four genera, *Fusobacterium* (g\_Fusobacterium), *Treponema* (g\_Treponema), *Selenomonas* (g\_Selenomonas), and *Campylobacter* (g\_Campylobacter), were enriched in tumor tissues, whereas three phyla, Bacteroidetes (p\_Bacteroidetes), Actinobacteria (p\_Actinobacteria) and Firmicutes (p\_Firmicutes) (including Clostridia at its lower class level (c\_Clostridia)), were abundant in normal tissues (Fig. 1c). Similarly, hierarchical clustering analysis of bacterial proportions accompanied by a heatmap confirmed that pathogenic bacteria and healthy bacteria were separately grouped into subclusters, indicating that

bacteria with similar characteristics coevolved and co-occurred under the influence of an altered environment (Additional file 1: Fig. S3). It is notable that LefSe-based analysis also adheres to the idea that tumor tissues harbor fewer bacterial features than normal tissues.

**Identification of pathogenic bacteria associated with tumor progression**

Using the taxonomic level information for each sample obtained by QIIME analysis, we constructed stacked graphs of the microbiome proportions for three selected levels, phylum, genus, and species. As shown in Fig. 2a–c, at the phylum level, a majority of the bacterial population (>75%) in normal tissues consisted of two OTUs, Bacteroidetes and Firmicutes, consistent with what was previously reported [39]. In tumor tissues, these two abundant bacterial OTUs were also highly abundant but



**Fig. 2** Colorectal cancer-associated bacterial composition. Average relative composition of the bacterial community at the phylum (a), genus (b) and species levels (c). d Box plot analysis of the relative abundance of four bacterial phyla, Bacteroidetes, Firmicutes, Actinobacteria and Fusobacteria. e Box plot analysis of the relative abundance of four species, *B. vulgatus* and *F. prausnitzii*, *F. nucleatum* and *B. fragilis*. Statistical significance was estimated by T-test

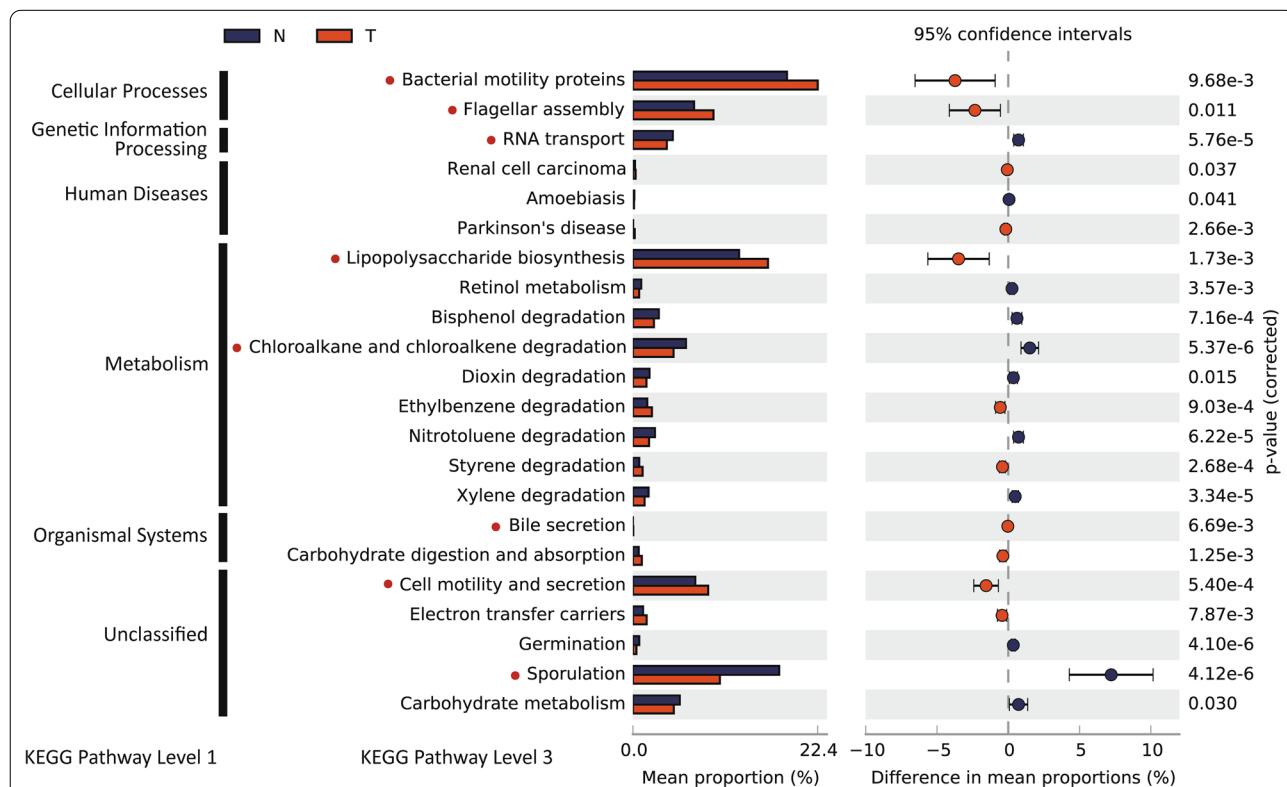
in significantly decreased proportions (~60%). In contrast, *Fusobacterium* showed a significantly increased proportion in tumor tissues (~23%) compared to that in normal tissues (~12%) (Fig. 2b). Actinobacteria was found to be more abundant in normal tissue than in tumor tissues, while Spirochaetes was the opposite, although the proportions of these two bacteria were very low under each condition (Fig. 2a).

Most OTUs were similarly proportionated between normal and tumor tissues at the genus and species levels, but a few bacterial compositions were significantly different between the two tissue conditions (Fig. 2b, c). For instance, at the genus level, *Bacteroides*, *Clostridiales* and *Prevotella* were more abundant in normal tissues; in contrast, *Fusobacterium* and *Treponema* were more enriched in tumor tissues (Fig. 2b). At the species level, *B. vulgatus* and *Faecalibacterium (F.) prausnitzii* were more abundant in normal tissues, whereas *F. nucleatum* and *B. fragilis* were significantly increased in tumor tissue (Fig. 2c). Statistical analysis of differences in bacterial compositions in normal and tumor tissues was performed for some selected bacteria at the phylum and species levels (Fig. 2d, e). Of particular interest are significantly higher amounts of two bacterial species, i.e., *F. nucleatum* and *B.*

*fragilis*, in tumor tissues than in normal tissues because these bacterial species have been repeatedly identified as pathogenic bacteria associated with intestinal inflammatory diseases and even with CRC [13–15, 18, 20]. Additional taxa with significant differences in OTUs between tumor and normal conditions are shown in Additional file 1: Fig. S4.

**Prediction of metabolic pathways exerted by microbiomes in normal and tumor tissues**

It has been suggested that the cross-talk between the microbiota and the host tissue may be mediated by short-chain fatty acids (SCFAs) produced by the microbiomes. Therefore, to predict microbiome-driven metabolic functions, we used a tool named ‘PICRUSt, i.e., a tool for making inferences by mapping marker genes to known sequenced genomes with information about the identified bacteria and their compositions. In particular, we found that the pathways of production and assembly of bacterial motility proteins (such as flagella) and of lipopolysaccharide biosynthesis were significantly enriched in tumors compared to normal tissues (Fig. 3) ( $P < 0.05$ ), which is consistent with a previous report based on the gut microbiome of Moroccan CRCs [40].



**Fig. 3** Functional pathways predicted with tumor- and normal tissue-enriched bacteria. KEGG pathways of OTUs enriched differentially between normal and tumor tissues were analyzed using PICRUSt (see “Materials and methods”). P-values were estimated by Welch’s t-test

In relation to this observation, it has been reported that overexpression of *flhDC*, a bacterial motility regulator, produced from *Salmonella* is associated with increased tumor cell mass [41]. Another notable compositional difference enriched in tumor tissue was bile acid secretion (Fig. 3) ( $P < 0.01$ ) because it was reported that some of the gut microbiome secretes bile acid, by which the microbiome can provoke a proinflammatory response in hepatic stellate cells [42].

In contrast, the pathways associated with sporulation, RNA transport, and chloroalkane and chloroalkene degradation were more significantly enriched in normal tissues than in tumor tissues (Fig. 3) ( $P < 0.0001$ ), although no good explanation for the cause and effect of this enrichment has yet been provided. Sporulation, i.e., a metabolic event that is expected to occur in gram-positive bacteria such as *Clostridia* [43], which was relatively more enriched in normal tissues, as shown in Fig. 1c, may partially explain this observation.

#### Correlation analysis between mRNA expression levels and proportions of bacteria

We then attempted to investigate the functional classes of genes that are differentially expressed in tumor tissues compared to normal tissues, which were conjectured to have been affected by changes in the composition of the microbiota. Briefly, DEGs were estimated by comparing mRNA gene expression in tumor tissues to that in normal tissues, in which functional classes thought to have altered expression together were identified by ssGSEA. Second, a correlation analysis was performed between each of the functional classes in ssGSEA and the proportions of the five bacterial features we selected in Fig. 2 (1 genus; *Fustobacterium*, 4 species; *B. fragilis*, *F. nucleatum*, *F. prausnitzii*, *B. vulgatus*). As a result, we found that the expression levels of host genes had a significant positive or negative correlation with the proportions of tumor-enriched bacteria or normal tissue-enriched bacteria (Fig. 4a), as expected. Interestingly, genes involved in tumor formation, including the cell cycle, cell adhesion, and the *Wnt* signaling pathway, were positively correlated with pathogenic bacteria, including *F. nucleatum* and *B. fragilis*, whereas normal tissue-enriched bacteria including *F. prausnitzii* and *B. vulgatus* were positively correlated with genes involved in starch and sucrose metabolism, the intestinal immune network and ABC transporters.

We also tried to investigate which cell types are likely to interact with microbial communities during tumor formation. After the cell types expected to contribute to bulk RNA-sequencing data were deconvoluted using a program called xCell [44], a correlation analysis was performed in the same way as was done for the DEGs

described above. Interestingly, we found that lymphoid cells, including both B and T cells, were positively correlated with normal tissue enriched bacteria, whereas myeloid cells, including monocytes and pericytes were positively associated with tumor-enriched bacteria (Fig. 4b).

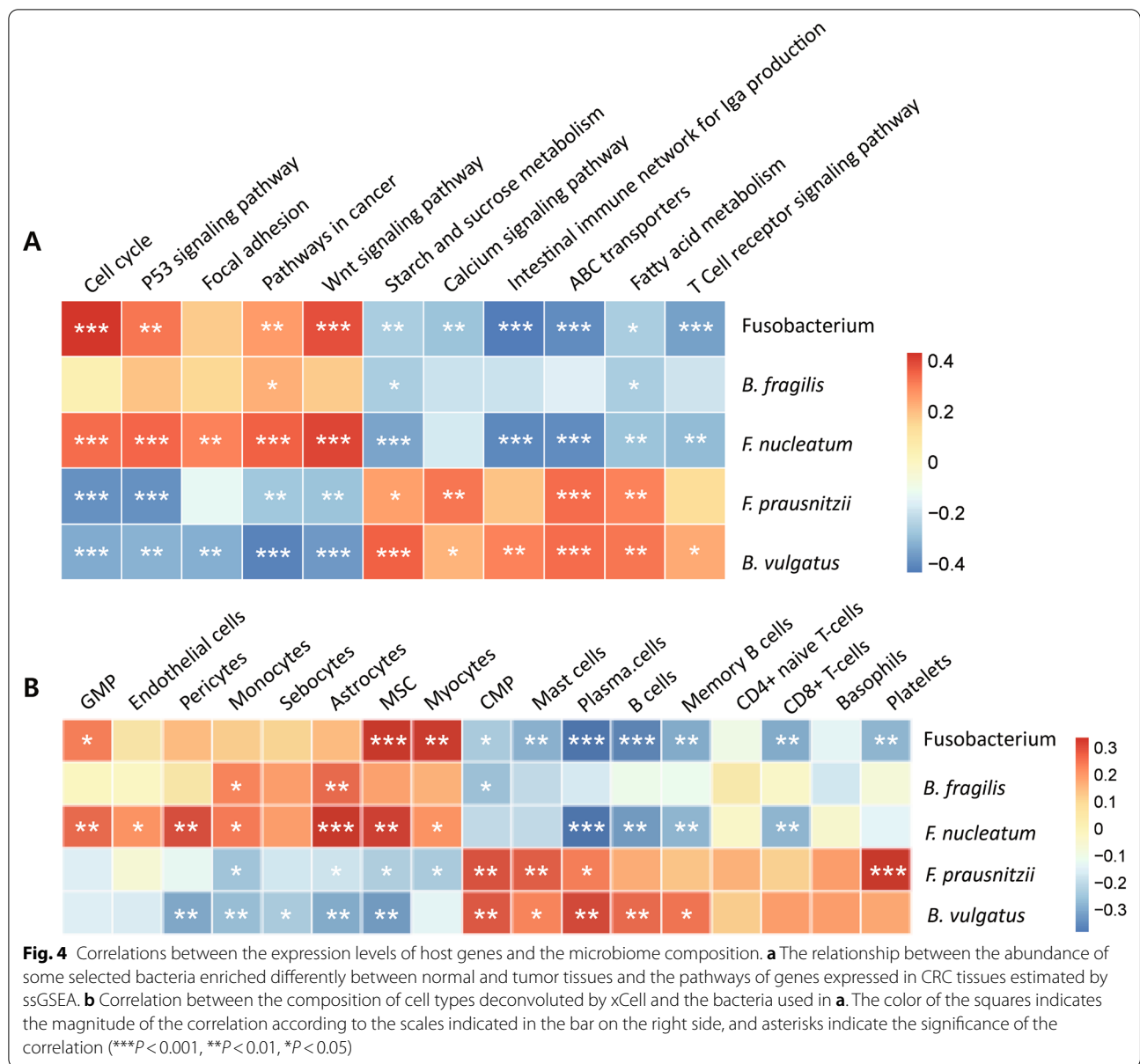
#### Enrichment of pathogenic bacteria can be a biomarker for better CRC prognosis

We next wondered whether the microbial composition could predict the prognosis of CRC patients. The proportions of tumor-enriched bacteria and normal tissue-enriched bacteria were compared between the 'crc\_RC' (patients with recurrence) and 'crc\_nRC' groups (patients without recurrence). Since normal and tumor tissues were collected from the same individuals, comparisons ('crc\_RC' vs. 'crc\_nRC') were performed separately for the normal tissue-derived microbiome (i.e., 'N\_crc\_RC' vs. 'N\_crc\_nRC') and for the tumor-derived microbiome (i.e., 'T\_crc\_RC' vs. 'T\_crc\_nRC'). Surprisingly, we found that tumor-enriched bacteria, including *B. fragilis* and *F. nucleatum*, were significantly more abundant in 'crc\_nRC' than in 'crc\_RC', indicating that the enrichment of these well-known pathogenic bacteria was associated with better prognosis of CRC patients (Fig. 5), which is contradictory with what has been previously reported [8, 18, 45–47]. To exclude the possibility that these unexpected observations were due to the nonrandom distribution of patients with early or late TNM stages into 'crc\_RC' and 'crc\_nRC', the comparison was conducted again after the TNM stages were controlled; comparison of 'crc\_RC' vs. 'crc\_nRC' was performed only for patients with TNM stage 2 (Fig. 6) and similarly only for patients with TNM stage 3. The samples from TNM stages 1 and 4 were removed because the number of samples was too small ( $< 9$ ) (Additional file 1: Table S1). Nevertheless, even in the stage-fixed sets, the conclusion was consistent, in that enrichment of pathogenic bacteria was associated with better prognosis of CRC patients.

#### Discussion

In the present work, we showed that alterations in the compositions of the microbiome were significantly associated with changes in the host tissue states from normal tissue to tumors, coupled with changes in the levels of some genes expressed in host tissues.

The method of comparing the microbiome compositions obtained under two different physiological conditions is basically similar to many other genomic, transcriptomic, and proteomic data analyses performed in the control-case design. However, finding biologically meaningful associations between the composition of the microbiome and human disease is not easy for



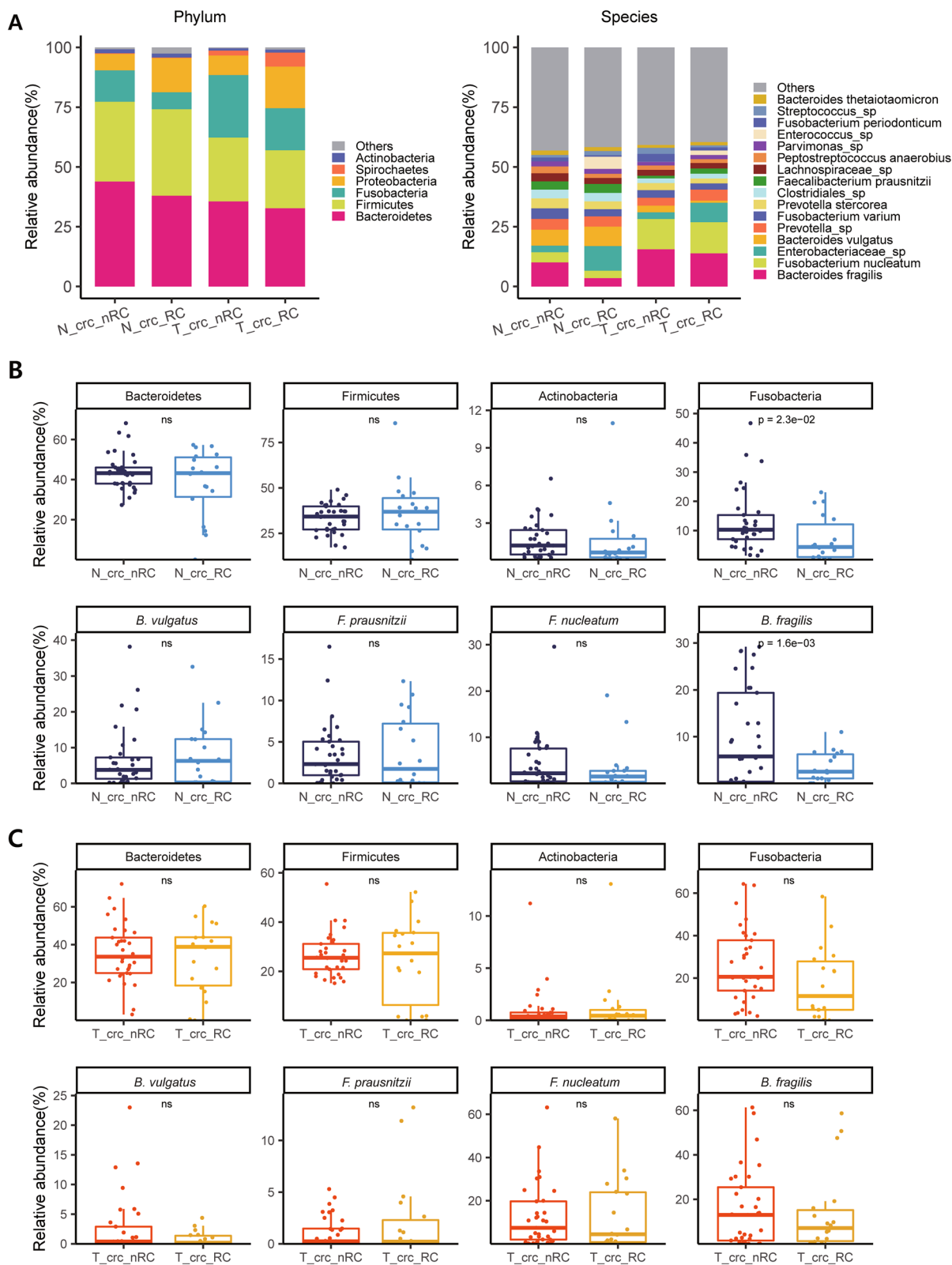
several reasons. First, microbiomes are highly heterogeneous, to the extent to which even the same individual can carry varied microbiomes depending on diets or physiological states, not to mention that different

individuals have different microbial compositions in the same tissue type. Second, various sample sources, such as fecal samples, mucus samples or tissue samples of patients, are used to isolate the microbiome.

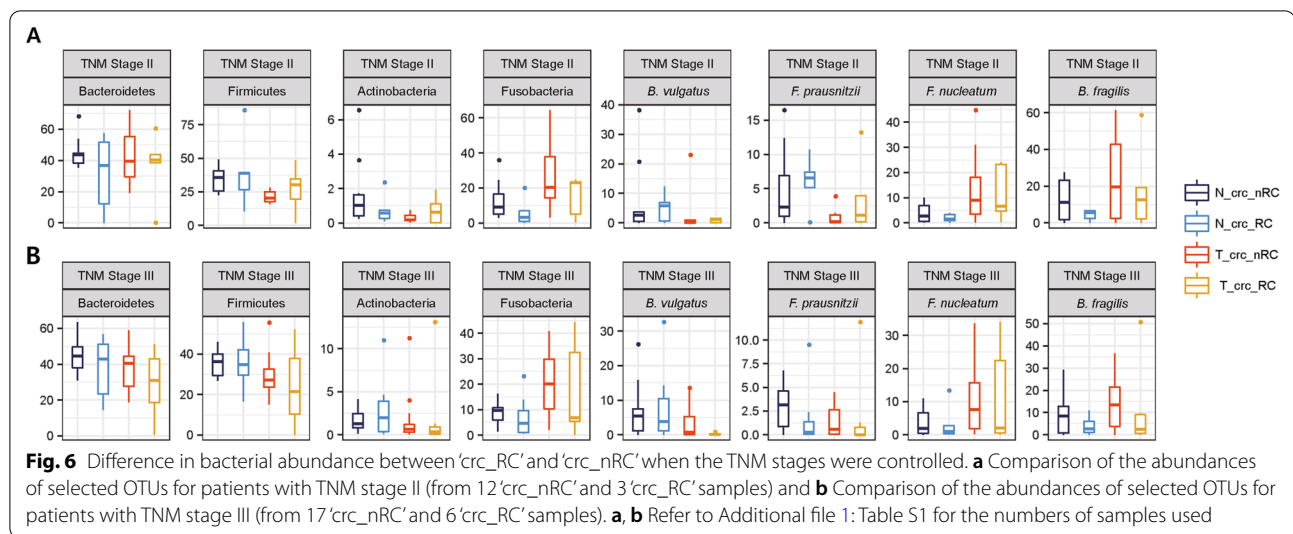
(See figure on next page.)

**Fig. 5** Comparison of the microbiome composition in four different tissue types. **a** Patients were divided into four subgroups by adding prognostic information that was retrospectively determined (RC: recurrence; nRC: nonrecurrence), accompanied by normal tissue- and tumor-derived microbiomes, i.e., 'N\_crc\_RC', 'N\_crc\_nRC', 'T\_crc\_RC', and 'T\_crc\_nRC', as described in the main text. Differences in microbial composition are displayed at the phylum (left) and species levels (right). **b** Box plot analysis of the relative abundance of selected OTUs at the phylum level (top) and the species level (bottom) between 'N\_crc\_RC' and 'N\_crc\_nRC'. **c** Box plot analysis of the relative abundance of selected OTUs at the phylum level (top) and the species level (bottom) between 'T\_crc\_RC' and 'T\_crc\_nRC'. **a, b** The samples used were derived from a total of 33 'crc\_nRC' and 18 'crc\_RC' samples for both normal and tumor samples (Additional file 1: Table S1)





**Fig. 5** (See legend on previous page.)



Third, various sequencing methodologies, either whole genome shotgun sequencing or 16S rRNA amplicon sequencing methods, are chosen to generate source sequencing data to identify microbiome components. Therefore, conclusions are often inconsistent regarding the increase or decrease in certain bacterial species associated with a given human disease, and the microbiome related to CRC is no exception. However, two bacterial species, *F. nucleatum* and *B. fragilis*, are consistently reported to increase in feces or mucus from CRC patients compared to healthy individuals or at increased levels in tumor tissue compared to normal tissue of the same CRC patient [8, 48, 49]. It seems that *F. nucleatum* is the most studied bacterial species related to the onset or progression of CRC. Our present study also drew the same conclusion. Taken together, the two bacteria, *F. nucleatum* and *B. fragilis*, may have a causal relationship in provoking inflammatory diseases and cancers.

As expected, these pathogenic bacteria have been reported to be associated with poor prognosis in CRC patients. For instance, patients with a high amount of *F. nucleatum* tended to have shorter survival times than patients with a low amount of *F. nucleatum* [45–47]. Yu et al. [50] showed that *F. nucleatum* was more enriched in chemoresistant recurrent CRC patients than in chemosensitive nonrecurrent patients by triggering the autophagy pathway via the TLR4/MYD88 pathway, which is consistent with the results of Zhang et al. [51]. However, we observed otherwise in the present work, showing that these two pathogenic bacteria were more enriched in CRC patients without recurrence (i.e., 'crc\_nRC') than in CRC patients with recurrence (i.e., 'crc\_RC'). As shown in Figs. 5 and 6, patients with higher

levels of pathogenic bacteria in their tissues had a consistently better prognosis, regardless of the sources of microbiomes (i.e., tumor- or normal tissue-derived microbiomes).

Interestingly, some studies have reported a good prognostic association of *F. nucleatum* in CRC. For instance, according to Oh et al. [52], the survival of *F. nucleatum*-high CRC patients was better than that of *F. nucleatum*-low CRC patients, when only a subgroup of microsatellite-stable CRC patients with nonsigmoid colon cancers treated with oxaliplatin-based chemotherapy were separately investigated. Notably, both Oh et al.'s samples and ours are based on microbiome data generated by the 16S rRNA amplicon sequencing method for tissue samples of homogenous Korean-only CRC patients. Saito et al. [53] showed that *F. nucleatum* could be associated with a good prognosis in a subgroup of CRC patients with *FOXP3*<sup>lo</sup> non-Treg cell infiltration.

Unfortunately, no good explanation has yet been proposed for this unexpected link between pathogenic bacteria and a good prognosis, unlike the case for the association with a poor prognosis. It is possible that there are strain-to-strain differences in the bacterial species present in different ethnic populations or that differences in the genetic makeup or local diet can cause the same pathogenic bacteria to have a different effect in the individuals tested. In addition, another possibility can be clued from the relationship between the density of *F. nucleatum* and the density of tumor-infiltrating lymphocytes (TILs); the density of *F. nucleatum* was reported to be positively correlated with the density of TILs in some CRCs [54], and the high density of TILs was shown to be associated with a better prognosis in CRC [55]. Therefore, it will be a great opportunity to develop a

microbiome-based prognostic marker in the future if we can determine how these pathogenic bacteria can inhibit the recurrence of cancer after surgical treatment and chemotherapy.

## Conclusions

We investigated whether alterations in the compositions of the microbiome were significantly associated with changes in the host tissue states from normal tissue to tumors, coupled with changes in the levels of some genes expressed in host tissues. We showed that the two pathogenic bacteria, *F. nucleatum* and *B. fragilis*, that were more abundant in tumor tissues than normal tissues were surprisingly more abundant in the patients without recurrence than in the patients with recurrence. We believe that our study will contribute to exploring the composition of tissue microbiomes that is critical in predicting the prognosis of CRC patients.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-021-03154-0>.

**Additional file 1.** Additional figures and table.

**Additional file 2.** Taxa at species level.

## Acknowledgements

Not applicable.

## Authors' contributions

SSC and DP conceived and designed the experiments; SJC performed the data analyses; DP, JC and MRJ contributed reagents and materials; SSC and SJC wrote the paper. All authors read and approved the manuscript.

## Funding

This research was supported by National Research Foundation of Korea (NRF) grants funded by the Ministry of Education, Science, and Technology (MEST, 2019R1A2C1002350) and by the Ministry of Science and ICT (MSIT, 2018M3C9A6017315).

## Availability of data and materials

The raw data is publicly available on NCBI portal at Sequence Read Archive (SRA) BioProject ID: PRJNA743150. 16S rRNA microbiome data—Submission ID: SUB9930275 and RNA-Seq data—Submission ID: SUB9954281.

## Declarations

### Ethics approval and consent to participate

This study was performed under the principles of the Declaration of Helsinki and was approved by the ethics committee of Samsung Medical Center in South Korea (No. SMC 2018-04-074-004). Informed written consent was obtained from all participants.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Division of Biomedical Convergence, College of Biomedical Science, Institute of Bioscience & Biotechnology, Kangwon National University, Chuncheon 24341, Republic of Korea. <sup>2</sup>GENINUS Inc., Seoul 05836, Republic of Korea. <sup>3</sup>Department of Medical Life Science, College of Medicine, Catholic University of Korea, Seoul 06591, Republic of Korea.

Received: 24 September 2021 Accepted: 18 November 2021

Published online: 29 November 2021

## References

1. Nguyen HT, Duong H. The molecular characteristics of colorectal cancer: implications for diagnosis and therapy. *Oncol Lett.* 2018;16(1):9–18.
2. Schatoff EM, Leach BI, Dow LE. Wnt signaling and colorectal cancer. *Curr Colorectal Cancer Rep.* 2017;13(2):101–10.
3. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res.* 2012;5(1):19–27.
4. Macrae FA. Colorectal cancer: epidemiology, risk factors, and protective factors. *Uptodate.com* [ažurirano 9.lipnja 2017]. 2016.
5. Sandler RS. Epidemiology and risk factors for colorectal cancer. *Gastroenterol Clin N Am.* 1996;25(4):717–35.
6. Francescone R, Hou V, Grivennikov SI. Microbiome, inflammation, and cancer. *Cancer J.* 2014;20(3):181–9.
7. Kho ZY, Lal SK. The human gut microbiome—a potential controller of wellness and disease. *Front Microbiol.* 2018;9:1835.
8. Dahmus JD, Kotler DL, Kastenber DM, Kistler CA. The gut microbiome and colorectal cancer: a review of bacterial pathogenesis. *J Gastrointest Oncol.* 2018;9(4):769–77.
9. Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, et al. Mucosal microbiome dysbiosis in gastric carcinogenesis. *Gut.* 2018;67(6):1024–32.
10. De Martel C, Franceschi S. Infections and cancer: established associations and new hypotheses. *Crit Rev Oncol.* 2009;70(3):183–94.
11. Grivennikov SI, Wang K, Mucida D, Stewart CA, Schnabl B, Jauch D, et al. Adenoma-linked barrier defects and microbial products drive IL-23/IL-17-mediated tumour growth. *Nature.* 2012;491(7423):254–8.
12. Gagnaire A, Nadel B, Raoult D, Neefjes J, Gorvel J. Collateral damage: insights into bacterial mechanisms that predispose host cells to cancer. *Nat Rev Microbiol.* 2017;15(2):109–28.
13. Strauss J, Kaplan GG, Beck PL, Rioux K, Panaccione R, DeVinney R, et al. Invasive potential of gut mucosa-derived *Fusobacterium nucleatum* positively correlates with IBD status of the host. *Inflamm Bowel Dis.* 2011;17(9):1971–8.
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. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. *Fusobacterium nucleatum* promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe.* 2013;14(2):195–206.
16. Wu Y, Wu J, Chen T, Li Q, Peng W, Li H, et al. *Fusobacterium nucleatum* potentiates intestinal tumorigenesis in mice via a Toll-like receptor 4/p21-activated kinase 1 cascade. *Dig Dis Sci.* 2018;63(5):1210–8.
17. Yang Y, Weng W, Peng J, Hong L, Yang L, Toyama Y, et al. *Fusobacterium nucleatum* increases proliferation of colorectal cancer cells and tumor development in mice by activating Toll-like receptor 4 signaling to nuclear factor-κB, and up-regulating expression of microRNA-21. *Gastroenterology.* 2017;152(4):851–866.e24.
18. Ulger Toprak N, Yagci A, Gulluoglu B, Akin M, Demirkalem P, Celenk T, et al. A possible role of *Bacteroides fragilis* enterotoxin in the aetiology of colorectal cancer. *Clin Microbiol Infect.* 2006;12(8):782–6.
19. Viljoen KS, Dakshinamurthy A, Goldberg P, Blackburn JM. Quantitative profiling of colorectal cancer-associated bacteria reveals associations between *Fusobacterium* spp., enterotoxigenic *Bacteroides fragilis* (ETBF) and clinicopathological features of colorectal cancer. *PLoS ONE.* 2015;10(3):e0119462.
20. Haghi F, Goli E, Mirzaei B, Zeighami H. The association between fecal enterotoxigenic *B. fragilis* with colorectal cancer. *BMC Cancer.* 2019;19(1):1–4.

21. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver–passenger model for colorectal cancer: beyond the usual suspects. *Nat Rev Microbiol.* 2012;10(8):575–82.
22. Saus E, Iraola-Guzmán S, Willis JR, Brunet-Vega A, Gabaldón T. Microbiome and colorectal cancer: roles in carcinogenesis and clinical potential. *Mol Aspects Med.* 2019;69:93–106.
23. Gagniere J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E, et al. Gut microbiota imbalance and colorectal cancer. *World J Gastroenterol.* 2016;22(2):501–18.
24. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* 2010;7(5):335–6.
25. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics.* 2011;27(21):2957–63.
26. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26(19):2460–1.
27. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007;73(16):5261–7.
28. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol.* 2006;72(7):5069–72.
29. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol.* 2011;12(6):1–18.
30. Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr.* 1957;27(4):325–49.
31. Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol.* 2013;31(9):814–21.
32. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics.* 2014;30(21):3123–4.
33. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15–21.
35. Lee S, Choi N, Koh I, Kim B, Lee S, Kim S, et al. Putative positive role of inflammatory genes in fat deposition supported by altered gene expression in purified human adipocytes and preadipocytes from lean and obese adipose tissues. *J Transl Med.* 2020;18(1):1–14.
36. Choi H, Lee S, Lee M, Park D, Choi SS. Investigation of the putative role of antisense transcripts as regulators of sense transcripts by correlation analysis of sense-antisense pairs in colorectal cancers. *FASEB J.* 2021;35(4):e21482.
37. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166–9.
38. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):1–21.
39. Liu W, Zhang R, Shu R, Yu J, Li H, Long H, et al. Study of the relationship between microbiome and colorectal cancer susceptibility using 16SrRNA sequencing. *BioMed Res Int.* 2020;2020:1–17.
40. Allali I, Boukhatem N, Bouguenouch L, Hardi H, Boudouaya HA, Cadenas MB, et al. Gut microbiome of Moroccan colorectal cancer patients. *Med Microbiol Immunol.* 2018;207(3):211–25.
41. Raman V, Van Dessel N, O'Connor OM, Forbes NS. The motility regulator flhDC drives intracellular accumulation and tumor colonization of *Salmonella*. *J Immunother Cancer.* 2019;7(1):1–16.
42. Nguyen PT, Kanno K, Pham QT, Kikuchi Y, Kakimoto M, Kobayashi T, et al. Senescent hepatic stellate cells caused by deoxycholic acid modulates malignant behavior of hepatocellular carcinoma. *J Cancer Res Clin Oncol.* 2020;146(12):3255–68.
43. Dürre P. Physiology and sporulation in *Clostridium*. The bacterial spore: from molecules to systems. 2016. p. 313–329.
44. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):1–14.
45. Mima K, Nishihara R, Qian ZR, Cao Y, Sukawa Y, Nowak JA, et al. *Fusobacterium nucleatum* in colorectal carcinoma tissue and patient prognosis. *Gut.* 2016;65(12):1973–80.
46. Chen Y, Lu Y, Ke Y, Li Y. Prognostic impact of the *Fusobacterium nucleatum* status in colorectal cancers. *Medicine.* 2019;98(39):e17221.
47. Kunzmann AT, Proença MA, Jordao HW, Jiraskova K, Schneiderova M, Levy M, et al. *Fusobacterium nucleatum* tumor DNA levels are associated with survival in colorectal cancer patients. *Eur J Clin Microbiol Infect Dis.* 2019;38(10):1891–9.
48. Flanagan L, Schmid J, Ebert M, Soucek P, Kunicka T, Liska V, et al. *Fusobacterium nucleatum* associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. *Eur J Clin Microbiol Infect Dis.* 2014;33(8):1381–90.
49. Shariati A, Razavi S, Ghaznavi-Rad E, Jahanbin B, Akbari A, Norzaee S, et al. Association between colorectal cancer and *Fusobacterium nucleatum* and *Bacteroides fragilis* bacteria in Iranian patients: a preliminary study. *Infect Agents Cancer.* 2021;16(1):1–10.
50. Yu T, Guo F, Yu Y, Sun T, Ma D, Han J, et al. *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. *Cell.* 2017;170(3):548–563.e16.
51. Zhang S, Yang Y, Weng W, Guo B, Cai G, Ma Y, et al. *Fusobacterium nucleatum* promotes chemoresistance to 5-fluorouracil by upregulation of BIRC3 expression in colorectal cancer. *J Exp Clin Cancer Res.* 2019;38(1):1–13.
52. Oh HJ, Kim JH, Bae JM, Kim HJ, Cho NY, Kang GH. Prognostic impact of *Fusobacterium nucleatum* depends on combined tumor location and microsatellite instability status in stage II/III Colorectal cancers treated with adjuvant chemotherapy. *J Pathol Transl Med.* 2019;53(1):40–9.
53. Saito T, Nishikawa H, Wada H, Nagano Y, Sugiyama D, Atarashi K, et al. Two FOXP3 CD4 T cell subpopulations distinctly control the prognosis of colorectal cancers. *Nat Med.* 2016;22(6):679–84.
54. Hamada T, Zhang X, Mima K, Bullman S, Sukawa Y, Nowak JA, et al. *Fusobacterium nucleatum* in colorectal cancer relates to immune response differentially by tumor microsatellite instability status. *Cancer Immunol Res.* 2018;6(11):1327–36.
55. Pagès F, Mlecnik B, Marliot F, Bindea G, Ou F, Bifulco C, et al. International validation of the consensus immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet.* 2018;391(10135):2128–39.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

