Microbiome

RESEARCH

Open Access

Metatranscriptomic analysis of the microbiota of tumor tissue in colon cancer

Elena Buetas[1†], Kelly Conde-Pérez[2†], Ángel Concha[3], Catuxa Celeiro[3], José F. Noguera[4], Germán Bou[2,5], Simón Pardiñas-López[6,7,8], Juan A. Vallejo[2], Margarita Poza[2,5,9*], Miguel Carda-Diéguez[1*] and Alex Mira[1,10*]

Abstract

Background Colorectal cancer (CRC) is the second leading cause of cancer-related deaths, accounting for more than 900,000 deaths a year worldwide. Microbial dysbiosis, including the presence of oral bacteria in the gut, has been linked to CRC. Some mechanisms by which specific microorganisms potentially drive tumorigenesis have been described, but there is a lack of studies elucidating whole microbiota activity in the tumor and their implication for the development of the disease. Here, the metatranscriptomic data of tumor and control tissue-associated microbiota ($n = 18$ pairs), as well as from subgingival sulcus ($n = 15$) of CRC patients, was analyzed.

Results We confirmed that Fusobacterium nucleatum was more active in the tumor tissue than in the control gut mucosa. In addition, the activity of this species was positively correlated with other oral bacteria in the tumors, including Parvimonas micra, Peptostreptococcus stomatis, and Granulicatella adiacens, along with gut bacteria like Hungatella hathewayi, suggesting a potential relationship among them. Regarding bacterial gene expression, a change in the functional profile was observed, including a higher expression of genes associated with carbon metabolism in control in contrast to an increase of amino acid-related genes in tumor. Furthermore, genes implicated in the biosynthesis and transport of lipopolysaccharide were increased in tumors. Interestingly, a significantly higher expression in tumor than control tissue of potential virulence factors from F. nucleatum was found, supporting their relevance in niche colonization and tumorigenesis. Correlation analysis of the bacterial activity with the host transcriptional profile showed significant correlations of the Fusobacterium-Peptostreptotoccus-Hungatella cluster with human genes involved in inflammation and metastasis, confirming the association of this microbial consortium with tumor development. For the first time, the gene expression profiles of oral bacteria in the gut and the oral cavity were compared. The cluster of co-active bacteria identified in tumors was partially found in the oral samples, suggesting a stable interaction and potential synergy. Although there were thousands of differentially expressed genes between subgingival sulcus and tumor tissue, the expression of key virulence factors was not significantly different.

[†]Elena Buetas and Kelly Conde-Pérez made equal contributions to this work.

*Correspondence:
Margarita Poza
margarita.poza.dominguez@sergas.es
Miguel Carda-Diéguez
miguel.carda@fisabio.es
Alex Mira
alex.mira@fisabio.es
Full list of author information is available at the end of the article

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Conclusions In short, this study discovered new traits about tumor microbial-associated composition and activity and its connection with the oral composition that would be essential to unravel the translocation, colonization, and tumorigenesis of the CRC.

Keywords Metatranscriptome, Gut microbiome, Colorectal cancer, Oral microbiome, Fusobacterium nucleatum, Clusters, Synergy

Introduction

Colorectal cancer (CRC) is the third most common type of tumor worldwide and the second highest cause of cancer deaths according to WHO [1]. Most cases do not have a hereditary origin (>60% sporadic) so they are associated with environmental agents and risk factors like alcohol and tobacco consumption, obesity and sedentarism, red and processed meat consumption or aging [2, 3]. In the last years, the microbiome has been recognized as a potential hallmark of cancer [4].

In the case of CRC, individual microorganisms such as Streptococcus gallolyticus, Bacteroides fragilis, pks+Escherichia coli, and Fusobacterium nucleatum have been linked to CRC pathogenesis [5]. Although the mechanisms by which the individual bacteria might contribute to carcinogenesis (genotoxicity, host immune modulation, and activation of oncogenic signaling pathways) have been studied through in vitro and in vivo mouse models and have been recently reviewed [6], the impact of whole microbial communities (including those exogenous to the gut) and their metabolism on tumor progression is not fully understood.

To study the gut microbiota, most studies rely on fecal samples, which represent not only bacteria from the tumor but also microbiota associated with other non-affected tissues all along the colon, decreasing the sensitivity to detect on-site shifts in microbiota [7]. The meta-analysis of fecal metagenomes has nevertheless suggested an increase in oral bacteria (i.e., Fusobacterium or Porphyromonas) as a signature of the disease [8, 9]. Interestingly, in the oral cavity, these bacteria have been related to the development of periodontal disease [10]. Periodontitis is a microbially-driven chronic inflammatory disease of the gums that increases the risk of developing CRC [11, 12]. Using tumor tissue samples, the presence of Fusobacterium as well as other oral-related bacteria such as Porphyromonas and Parvimonas [13–15] had been confirmed through 16S rRNA gene sequencing. Nevertheless, to explore the mechanisms by which bacterial communities potentially promote tumors, metatranscriptomic analyses may provide a powerful tool.

Previous studies which performed RNA sequencing consistently detected active F. nucleatum on tumor tissue [16–19]. However, the active functional profile of the microbial community was not assessed in depth. There is therefore a lack of knowledge regarding the metabolic pathways and potential virulence factors expressed by the microbial community in tumor samples, as well as their potential interactions.

Understanding microbiome activity in tumors is crucial to elucidate their contribution in CRC development and progression. To achieve this, RNA sequencing of adenocarcinoma and paired non-affected tissue has been performed, providing transcriptomic profiles of the microbiome and the host. Considering the pivotal role that oral bacteria appear to have in CRC, special attention has been given to the activity of oral bacteria, and the RNA from the subgingival sulcus of CRC patients has been additionally sequenced. Thus, the current manuscript explores the active functional profile of the gut and its associated bacterial community in CRC tumor tissue, as well as the corresponding microbial metatranscriptome in the oral niche.

Methods

Participants and sample collection

Participants were recruited at the University Hospital of A Coruña (CHUAC) between 2019 and 2022. Ninety-three CRC patients provided informed consent and passed inclusion and exclusion criteria (see [13], Fig. 1). Each participant completed a questionnaire recording their personal data (age, sex, weight, height), oral health and lifestyle habits before the surgery. Colorectal tumor tissue and non-neoplastic (control) tissue from the surrounding areas were collected from 47 patients via sterile dissection during surgical resection. Tissue samples were immediately placed in 500 µL of RNAlater reagent (Thermo Fisher Scientific) and stored at −80 °C until nucleic acid extraction and sequencing.

A portion of the 93 patients ($n=23$), who voluntarily agreed, underwent a dental check-up (Pardiñas Medical Dental Clinic, A Coruña, Spain). Periodontal assessment for each patient was done according to the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions [20]. The gingival condition of each patient was recorded using the Silness-Loe gingival index [21]. During examinations, subgingival samples were collected. Eight sterile paper points (ISO 30, Henry Schein, Melville, NY, USA) were placed in the subgingival sulcus of different teeth for 30 s and kept in a

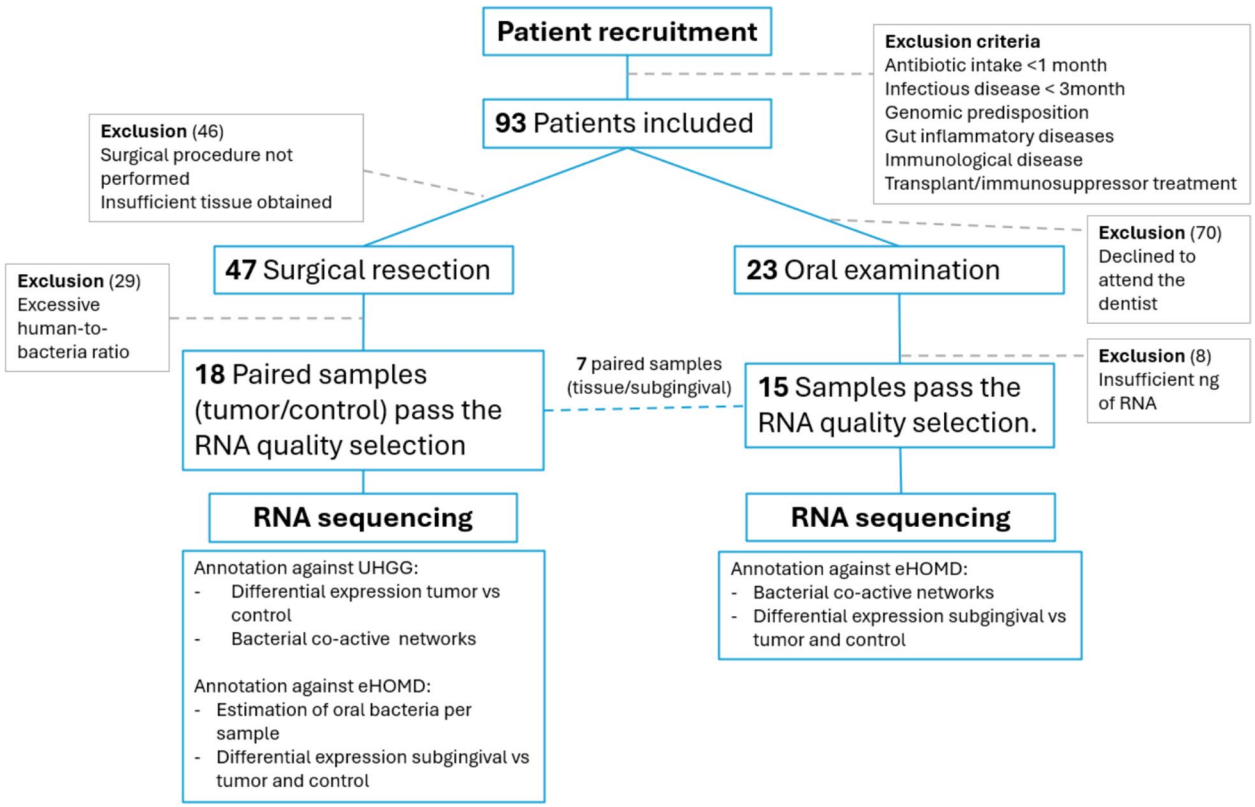

Fig. 1 Workflow of the study. Patients were recruited after CRC diagnosis at the University Hospital of A Coruña, Spain. Patients who fulfilled the inclusion criteria and signed informed consent participated in the study. From them, the tumor resection of 47 participants was obtained after surgery. Only 23 of them agreed to visit the dental clinic. Oral (subgingival plaque) and gut (tumor and unaffected tissue) samples were subject to RNA extraction and RNAseq, followed by bioinformatic analysis to assess active bacterial composition and transcriptional profiles in the microbial populations

tube containing 500 µL of RNAlater (Thermo Fisher Scientific). All samples were stored at − 80 °C until further analysis.

Nucleic acid extraction

Tissue samples (20 mg) were homogenized using Lysing Matrix E tubes (MP Biomedicals, St. Ana, CA, USA) and a 1600 MiniG system (Thermo Fisher Scientific). Subgingival samples were vortexed (2 min) to detach bacteria from gingival papers, which were carefully discarded. DNA and RNA were isolated using the AllPrep DNA/RNA mini kit (Qiagen) following the manufacturer's instructions with an additional lysis step with an enzymatic cocktail for bacteria and fungi [13]. Negative controls for each extraction were included.

Tissue sample selection

DNA from gut tissue samples was used to measure the ratio of human cells to bacterial cells. In short, a quantitative PCR (qPCR) for the human ß-actin gene (ActineF: 5′-TTGTTACAGGAAGTCCCTTGCC-3′; ActineR: 5′-ATGCTATCACCTCCCCTGTGTG-3′) and for the bacterial 16S rRNA gene (16S_U515F: 5′-GTGCCAGCMGCCGCGGTAA-3′; 16S_U789R: 5′-GCGTGGACTACCAGGGTATCT-3′) was performed in duplicate for each sample. The number of human cells and the number of bacteria were estimated through the respective standard curve [22]. To obtain higher bacterial reads per sample in the RNA whole sequencing, samples with a lower proportion of human cells (log[cel/cfu] < 2, based on previous in-house optimization Fig. S1) were selected, obtaining 18 paired (control and tumor from the same patient) samples. To assess the potential selection bias of this process, the differences in clinical parameters (sex, age, tumor size, tumor location) between groups (selected/non-selected) were tested using Fisher exact test and Wilcoxon test for categorical and continuous variables respectively. No significant differences were found (Table S1).

Library construction and sequencing

The V3–V4 regions of the 16S rRNA gene were sequenced using paired-end Illumina MiSeq v3 reagent kits (2×300) (Illumina) in CHUAC as previously detailed [13]. For the metatranscriptomic sequencing, the RNA concentration of the control, tumor, and subgingival samples was measured by the Qubit RNA HS Assay Kit (Thermo Fisher Scientific). Only 15 subgingival samples had enough concentration to perform the RNA sequencing (among these, 7 of them have paired tumor and control samples). At the FISABIO sequencing platform (Valencia, Spain), gut tissue samples were treated with the Illumina Ribo-Zero Plus rRNA Depletion Kit. Finally, libraries were prepared and sequenced using NextSeq Illumina Technology (single ends, high/mid-output 1×150 bp) following the manufacturer's instructions and negative controls were included.

Metatranscriptome analysis

Raw reads were trimmed to remove adapters with cutadapt 1.18 [23] and then by quality and length using the PRINSEQ program [24]. After that, sequences that aligned to the human genome (GRCh38.p9) or to the ribosomal database SILVA 132 [25] with Bowtie2 were discarded. Additionally, the remaining reads were aligned to the CHM13 v2.0 human reference genome using Bowtie2 [26]. We used Kraken2 to map the reads against a comprehensive database including GenBank, RefSeq, third party annotation (TPA), and protein data bank (PDB) data built on 28th December 2024 (https://benlangmead.github.io/aws-indexes/k2) (from now on, core_nt Database) [27]. We estimate the abundance of each kingdom, and in the cases where sequencing depth was enough to cover diversity, we studied the differences between tumor and control samples.

For the bacterial component, samples with more than 1×106 filtered reads (after human and ribosomal depletion) were mapped to the Unified Human Gastrointestinal Genome (UHGG) database [28] -gut samples- or the extended human oral microbiome database (eHOMD) [29] -subgingival samples-, using Bowtie2 (version 2.4.2) [30] (implementing the –very-sensitive option) and Samtools (version 1.12) [31] to convert from SAM to sorted BAM. Importantly, no comparisons were performed between samples annotated using different databases.

According to the alignments and gene coordinates, we counted (R package GenomicAlignments 1.16.0) [32] the number of hits mapping at each gene in order to build an abundance matrix of gene expression only considering primary alignments (in case of multi-mapping). An abundance matrix of genomes was also built counting the number of primary alignments on all the contigs of the same genome. Reads were aligned to the concatenated reference of F. nucleatum lineages using Bowtie2 with the –very-sensitive preset. Only the primary alignment per read was retained for downstream analyses. Mapping quality (MAPQ) distributions were examined to assess multimapping, and sensitivity analyses were performed by collapsing all lineages into a single F. nucleatum group, confirming that the tumor-associated signal was preserved. Negative controls report less than 4×103 annotated reads.

A KEGG (version 2016) [33] annotation was added for each gene. First, the genes were translated into amino acids by means of EMBOSS Transeq (version 6.6.0) [34], with translation Table 11 (bacterial) and frame 1. Next, each peptide was mapped against the KEGG database with HMMERsearch (version 3.3.2) [35], maximum e-value 1e–06. Finally, for each gene, the KEGG annotation corresponding to the best (highest domain score) alignment was selected, with the aid of a custom R script (version 3.6.0) [36]. To estimate the number of oral-related bacteria, reads from gut samples were mapped again to the eHOMD using the same approach.

Human transcripts

Total reads after quality filtering were mapped with Bowtie2 to the human reference genome (GRCh38.14). An abundance matrix of gene expression was built. Rarefaction curves were performed using the minimum number of annotated reads per sample (3.0×106). Differential expression analysis was performed using DESeq2. To establish enriched pathways, we used ShinyGo 0.82 [37]. Normalized gene counts (TPM) of enriched pathways were correlated with bacterial gene expression of all samples using sPLS canonical method from mixOmics package [38]. LASSO penalization was used to keep 40 variables per component, and filter of relevance and correlation (0.5 and 0.5, respectively) were applied.

Bacterial transcripts

Rarefaction curves and alpha diversity (Chao1 estimator and the Shannon index) were calculated using the minimum number of annotated reads to the UHGG database in a sample (99.800 reads). The differential active composition of the microbiome was studied with ANCOM-BC [39] adjusting for multiple comparisons using the false discovery rate (FDR) method. In addition, bacteria with less than 0.015% abundance and with prevalence below 30% in a group were filtered out. To infer if differentially active bacteria in tumors are over-expressed or only more abundant in the niche, normalized 16S rRNA gene data and transcript reads were compared [40]. As 16S rRNA gene sequencing data cannot achieve species resolution, the ratio (MTT/16S) at the genus level was performed for the estimation in each patient.

Oral-associated bacteria that pass the filter of prevalence and abundance in the samples were subset (Table S2), and correlations were performed using the sPLS-canonical method based on normalized data. LASSO penalization was used to keep 15 variables per component, and correlations were filtered by relevance and score (0.75 and 0.5, respectively) [38].

Differential gene expression in bacteria

Then differential gene expression analysis between tumor and control samples was performed to paired samples using DESeq2 [41] for assigned KEGG sequences. Enriched pathways based on differential expressed genes (p-value < 0.05) were identified through an hypergeometric test as in [42], and the corresponding p-value was corrected for multiple comparisons using the FDR method. To estimate the taxonomic contribution to the expression of a gene, the percentage of hits that belong to each bacterium in each sample was calculated and the average was performed by groups.

Oral-gut differences

Subgingival samples and gut samples annotated against the eHOMD database were normalized using RPKM (reads per kilobase per million mapped sequences). A principal coordinates analysis (PCoA) based on Bray–Curtis distance from the gene expression was performed. Differences on gene expression among oral bacteria from the three niches were computed with DESeq2 [41]. Boruta algorithm was used to confirm relevant differential features [43]. Figure 7 was created using BioRender.

Results

The main goal of this study was to elucidate the potential bacterial mechanisms that could be promoting tumorigenesis in the colon. To accomplish this, RNA sequencing from surgical sections of CRC tumors and matched unaffected control tissue was performed. After RNA quality control and sample selection based on the human cells per bacteria ratio, 18 paired samples (tumor and control samples from the same patient) were included. Considering the potential implication of oral microbes in the tumorigenesis process, the metatranscriptome of subgingival samples from 15 participants was also compared between oral environment and the gut (Fig. 1). After an evaluation, 73.3% of the donors of subgingival samples presented some degree of periodontitis (Table S3).

Active microbial composition in tumors and non-affected tissue

After human and ribosomal sequences removal, a mean of 1.1×106 (min 1.3×105, max 7.5×106) reads per sample were obtained. To explore the whole microbial composition of our samples, unassigned reads were annotated against a comprehensive database (core_nt Database). A mean of 0.1% and 1.4% of the reads correspond to archaea and virus, respectively (Fig. S2). Indeed, only 6 samples presented more than 1000 archaea reads and therefore did not provide sufficient sequencing depth to analyze these data with confidence. The most abundant archaea among these samples were Methanobrevibacter and Methonsphaera. In the case of viruses, a mean of 16,800 reads per sample was annotated. Interestingly, no differences were observed in viral diversity between tumor and control tissues. Furthermore, Orthopicobirnavirus hominis was increased in tumors whereas bacteriophages (Siphoviridae sp., Myoviridae sp) were higher in control tissue than in tumors. Regarding fungi, a single patient presented more than 1000 reads, which corresponded to Candida glabrata (currently Nakaseomyces glabratus).

Using the intestinal based UHGG database, 78.92% of the reads were annotated on average [28]. The genera accounting for the largest proportion in the RNA pool were Bacteroides, followed by Blautia, Faecalicatena, Anaerostipes, and Parabacteroides in both sample types. In addition, oral-related genera like Streptococcus, Fusobacterium, Porphyromonas, and Veillonella were among the 25 most abundant bacteria detected in the active community (Fig. S3).

When comparing the diversity indexes using UHGG, no differences in alpha diversity were found between control- and tumor-associated microbiomes. However, evaluation of the changes in bacterial representation showed 247 species differentially active (adjusted p-value < 0.05) between tissue types. Blautia wexlerae, Blautia massiliensis, Bifidobacterium adolescentis and Bifidobacterium infantis were more active in control than in tumor. In the tumors, Roseburia intestinalis as well as oral-related species such as Fusobacterium nucleatum (Fn) subspecies animalis, Fn. vincentii and Streptococcus salivarius showed the highest log2FC values (Fig. 2A). Other oral species, Fn. nucleatum and Peptostreptococcus stomatis were more active in tumor (log2FC > 1.5) but the difference was not statistically significant when correcting for multiple testing (p-value < 0.05, adjusted p-value > 0.05). To establish if the higher activity detected corresponds to a higher abundance of the bacteria in the niche, we used the relative abundance data obtained using 16S rRNA gene sequencing described by Conde-Pérez et al. [13] (Fig. S4). Considering the ratio (mean of log2(MTT/16S)) at the genus level, Fusobacterium, Peptostreptococcus and Roseburia were overexpressed (higher expression than relative abundance), whereas Streptococcus was more abundant than active in both tissues (Fig. 2B, Table S4).

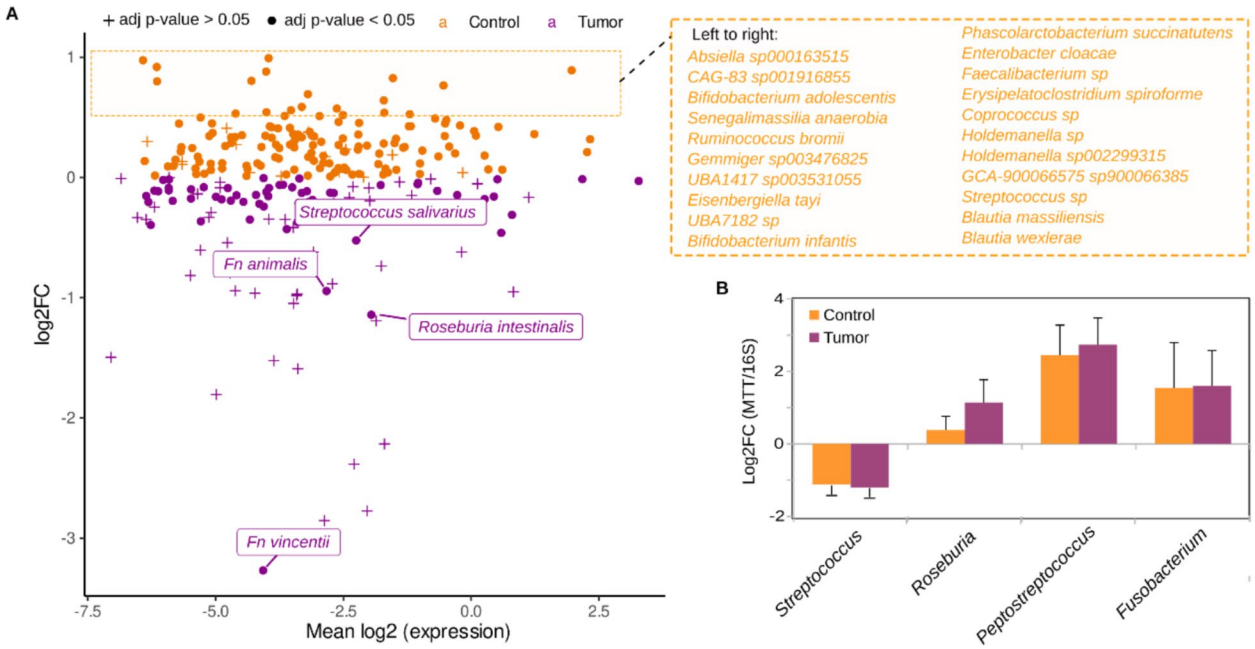

Fig. 2 Active bacterial composition in paired tumor and unaffected, control samples. **A** Bacterial species differentially active in control and tumor samples as assessed by the ANCOM-BC method. Species with log2FC >|0.5| are marked. [+ Not sig identifies cases with adjusted *p*-value > 0.05, ● Sig cases with an adjusted *p*-value < 0.05; FC: Fold-Change control/tumor; Mean log2 (expression): (log2(mean relative expression control) + log2(mean relative expression tumor))/2]. **B** Ratio of bacterial relative abundance based on the transcripts (RNA) vs 16S gene sequencing (DNA). Mean and standard error

Since oral bacteria seemed to account for the main difference between control and tumor samples and the database used (UHGG) is based on intestinal data, the filtered sequences (without human or ribosomal reads) were now mapped against eHOMD which harbored an up-to-date oral bacteria genomic catalogue. With this approach, a mean of 9% and 13% of reads were annotated in control and tumor, respectively (against 82.5% in control, 74.61% in tumor in UHGG database) (Fig. S5A), some of which (i.e., the periodontal pathogen Tanerella) were not annotated in the gut microbial database. Among the oral bacteria found, the periodontitis-associated bacteria tended to be more present in the tumor samples than in control, Fusobacterium being the most relevant genus (Fig. S5B).

Oral bacteria and their interactions in the gut

Given that the main differences in bacterial composition in the transcriptome corresponded to oral-related microbes, we studied the correlation of their activity with the rest of the community to detect co-activity patterns. In control samples, we observed that several Veillonella and Streptococcus species correlated positively with the gut bacteria Prevotella copri, B. adolescentis, and B. infantis. Other clusters of positively correlated bacteria include mainly oral pathogens such as Fusobacterium species (Fn. nucleatum, Fn. animalis, Fn. polymorphum, Fn. vincentii, and F. periodonticum), P. stomatis, and Granulicatella adiacens. Interestingly, Streptococcus anginosus positively correlated with both groups mentioned above (Fig. 3A). Although similar groups were found in tumor, a positive correlation of the pathogenic group with Parvimonas micra, another oral species involved in periodontal disease, was found. At the same time, this bacterium negatively correlated with Bacteroides dorei, Bacteroides sartorii, and Neobitarella massiliensis. In addition, the oral pathogenic consortia correlated with several gut bacteria like Hungatella hathewayi, Clostridium bolteae, Butyricimonas virosa, or Bilophila wadsworthia (Fig. 3B). Therefore, coactivation networks suggest that oral bacteria could interact with specific gut bacteria, or that they could benefit from the same environmental conditions.

Bacterial gene expression in gut samples

To further explore the differences between control and tumor we studied the differences in gene expression between these samples. Among the two niches, 319 genes were differentially expressed (*p*-value < 0.05), 104 genes with higher expression in tumor and 215 in control (Table S5). The categorization of these genes into pathways and the comparison of the representation of pathways between control and tumor demonstrated that

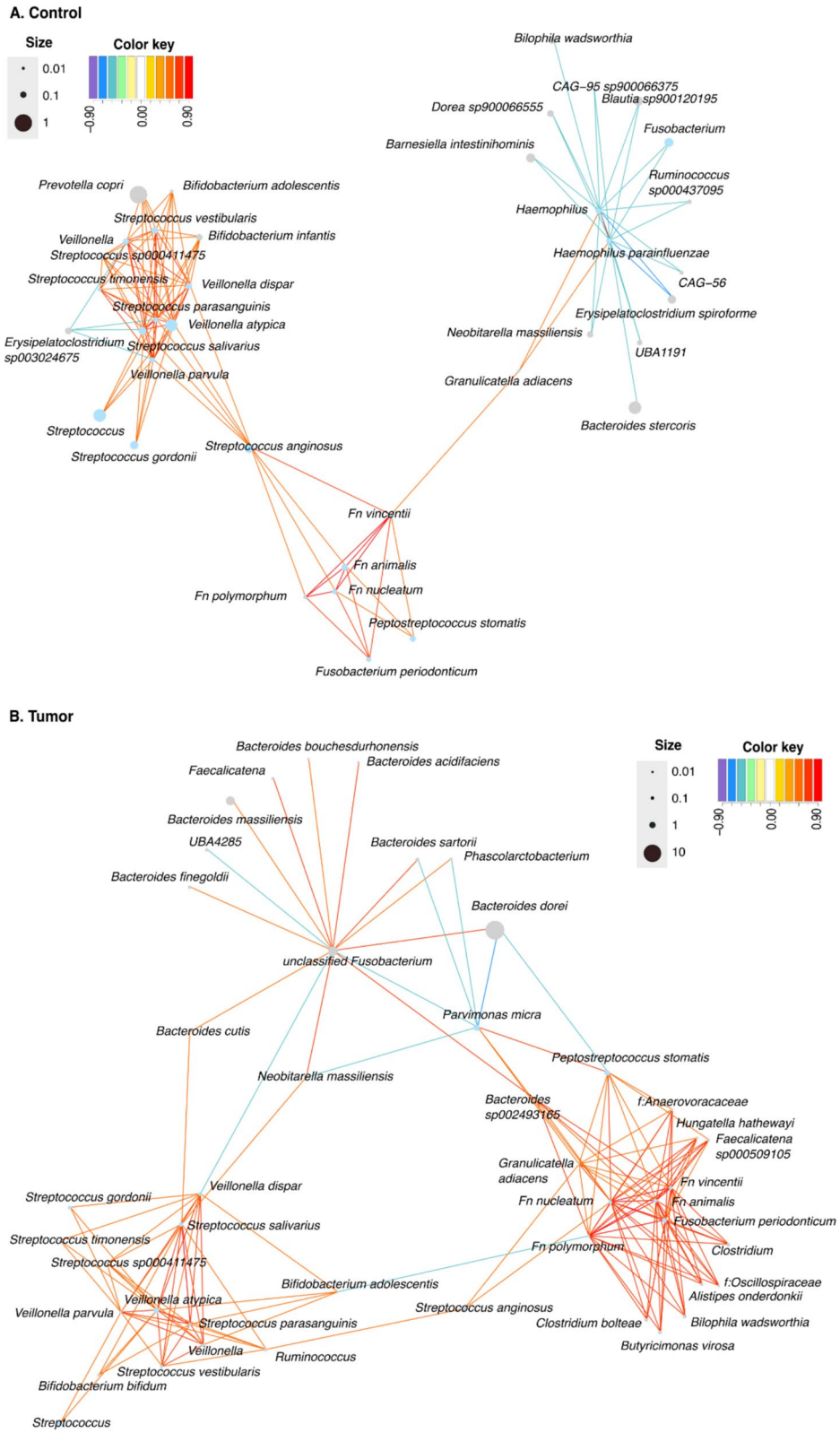

Fig. 3 Potential bacterial interactions in the gut. **A** Correlation networks of oral bacteria based on their transcriptional profile in the gut (unaffected) control tissue, as assessed by the sPLS method. **B** Correlation networks of oral bacteria based on their transcriptional profile in the gut tumor tissue by the sPLS method. Blue nodes indicate oral species, grey nodes indicate gut species. Lines are marked in different colors as specified in the color key according to their positive (red) or negative (blue) correlations. Node size corresponds to relative abundance based on transcripts

the carbon fixation pathway, ABC transporters, methane metabolism, and citrate cycle were more expressed in control than in tumor (adj. $p < 0.05$), whereas in tumor, there was a significant (adj. $p < 0.05$) enrichment in ribosome-related genes, translation factors, aminoacyl-tRNA biosynthesis, and amino acid-related enzymes (Fig. 4A, B).

Among the genes related to ABC-transporters that were found more abundant in C (29 genes), there were transporters for lactose, raffinose, and fructose, oligogalacturonide, or N-acetylglucosamine. In addition, four genes for PTS transporters were found significantly higher in control, namely a phosphotransferase system enzyme I PTS-EI, the enzyme II ascorbate-specific ulaA, and the N-acetilgalactosamine-specific agaW and agaE. In contrast, only 10 genes related to transporters were significantly more detected in tumor. Among them, there were ABC-type dipeptide/oligopeptide/nickel transport systems and a lipopolysaccharide (LPS) export system. In addition to LPS transporters, other genes related to the biosynthesis of LPS were more expressed in tumor than in control. Gene-level analyses are presented as exploratory and are not corrected for multiple testing; these results are provided to facilitate biological interpretation of the adjusted pathway-level findings and should be interpreted with appropriate caution.

Whereas no tRNA synthetases were increased in control, five were found more expressed in tumor, as well as translation factors. Genes involved in bacterial motility (i.e., flagellar) were also found increased in tumor (Fig. S7A). In the case of genes related with antimicrobial resistance, four were detected at higher expression levels in tumorand five, mainly related with β-lactam resistance in control samples.

A significantly higher expression of described virulence factors like bigA (putative surface-exposed virulence protein), fhaB, and fhaC (filamentous hemagglutinin and hemolysin activator/secretion protein) was detected in the tumor tissue. Interestingly, these genes were expressed mainly by Fusobacterium species (Fig. S7B). In contrast, hemolysin III as well as other iron-related transporters were found more expressed in control and mainly by Faecalicatena gnavus, Anaerostipes hadrus, and B. wexlerae. To confirm the annotation of these relevant genes expressed by oral Fusobacterium, Fusobacterium bigA sequences were mapped against the Fusoportal database [44]. The genes annotated as bigA in our metatranscriptome correspond to 5 different autotransporters secretion system type Va (T5aSS) of Fn. nucleatum (23726_Gene_665, 23726_Gene_362 Aim1, 23726_Gene_868, 23726_Gene_2068 Fap2, and 23726_Gene_32 RadD) which are known virulence factors of this bacterium [45]. Therefore, the tumor-microbiome had less expression of carbon metabolism-related genes and higher expression of virulence factors than the control microbiome, being Fn a key contributor.

Gene expression of oral associated bacteria in oral cavity vs gut

As oral species such as F. nucleatum were previously reported to be involved in CRC and the potential implication of other oral species was suggested by our dataset, we explored differences in gene expression between oral and tumor samples. For that, the transcriptional profile of subgingival samples was compared with control and tumor samples annotated against eHOMD database. The use of an oral database limits the annotation of gut samples but facilitates the comparison between niches. Clear differences in the composition of active oral microbes between subgingival and gut tissue samples were observed. Regarding co-activation patterns, Veillonella correlated positively with Streptococcus and negatively with Fusobacterium in subgingival samples, similar to what we observed in tumor samples. However, only in subgingival samples Fusobacterium correlated positively with a diverse group of bacteria including

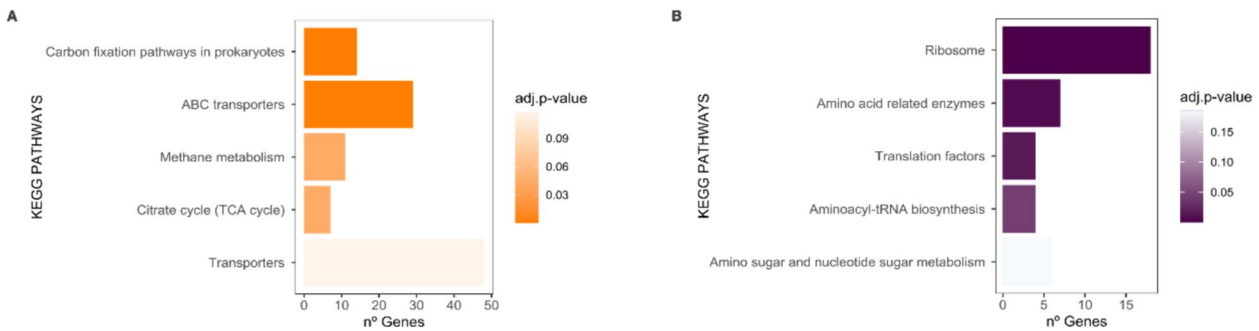

Fig. 4 Bacterial gene expression profiles in the gut tissue. **A** Enriched pathways in control (unaffected) gut tissue, *p*-value adjusted by FDR method. **B** Enriched pathways in colorectal cancer tumor tissue, *p*-value adjusted by FDR method

different species of Haemophilus, Capnocytophaga and Prevotella (Fig. S8). Interestingly, Granulicatella adiacens correlated positively with the Veillonella group in subgingival in opposition to in tumor where it correlated with Fusobacterium. The clustering of samples in the PCoA plot based on their transcriptional profile suggested that gut samples had high interindividual variation (Fig. 5A). Differential expression analysis using DESeq2 showed 1751 genes significantly different between subgingival-control samples and 1921 comparing subgingival-tumor samples, 354 and 369 of which were confirmed by the Boruta algorithm, respectively (Fig. 5B). Genes related with ribosome and transcription machinery were more expressed in control or tumor samples than in subgingival ones whereas bacterial motility and secretion system genes presented the reverse pattern. Although most of the differential confirmed genes were shared in both comparisons (209) there were genes which were significantly different between control and subgingival samples but not between tumor and subgingival (Fig. 5B). Among them we found peptidases (proline aminopeptidase pepP, glutamyl endopeptidase sspA), toxins (toxin A/B tcdAB, toxin complex tccC), chemotaxis proteins (cheX, cheD, cheZ, cheV) and virulence factors as internalin A (inlA). In addition, the main virulence genes overexpressed in tumor (bigA, fhaB, fhaC) by Fn are not significantly different between subgingival and tumor tissue.

Interestingly, although not confirmed by Boruta, genes related to the prokaryotic defense system as a modulating expression of haemolysin hha, a stress resistance protein

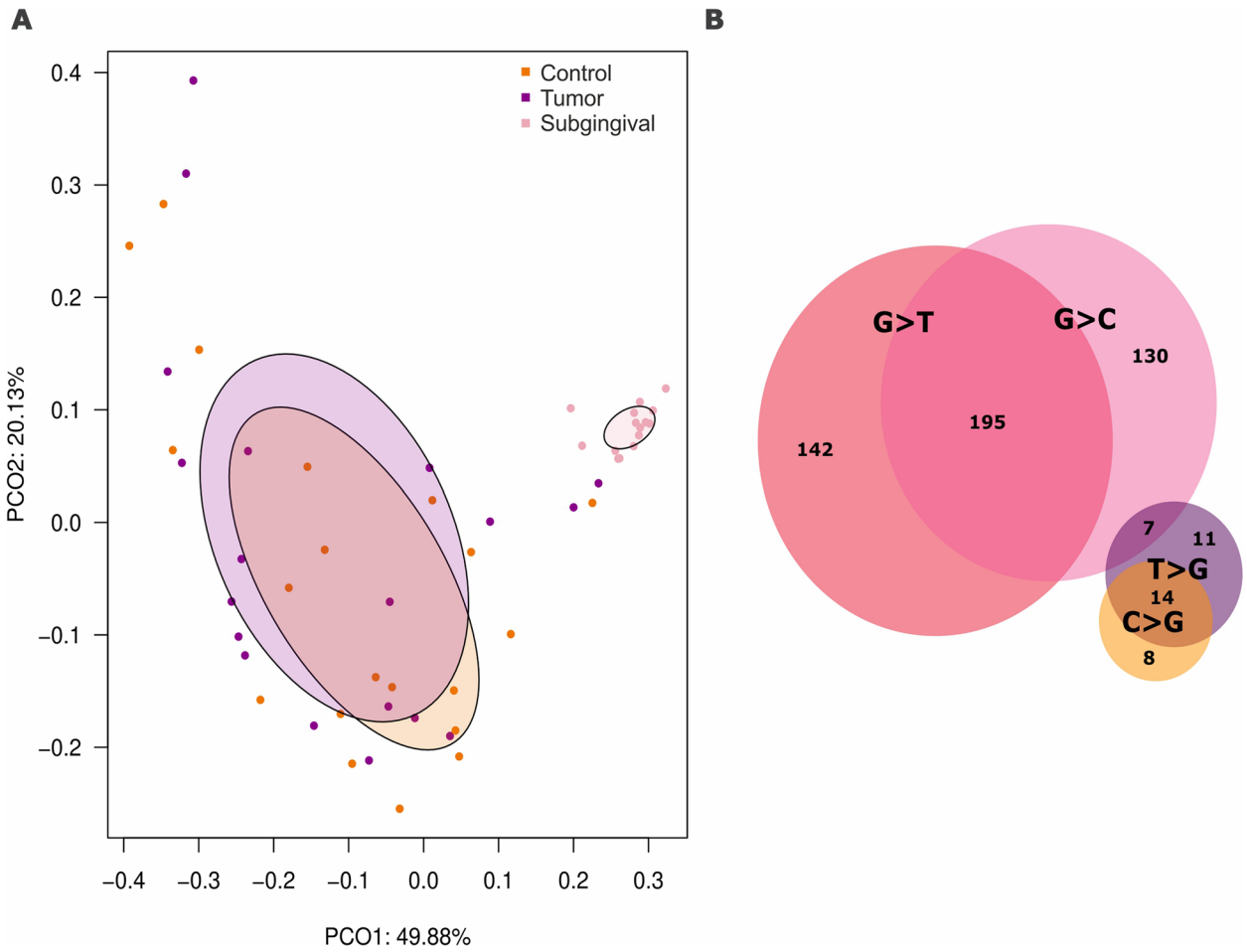

Fig. 5 Oral bacteria gene expression in the gut and the oral cavity. **A** PCoA based on Bray Curtis distance from normalized counts (RPKM) of subgingival, gut control and gut tumor samples annotated against the Extended Human Oral Microbiome Database (eHOMD). **B** Venn's diagram showing the number of differentially expressed genes confirmed by the Boruta algorithm between the oral and gut environments. In the orange circle, genes more expressed in control (**C**) than in subgingival samples (G) are shown; in the purple circle, more expressed in tumor (T) than in subgingival samples, pink circles genes more expressed in subgingival samples than in control or tumor tissue are shown; the middle overlap area depicts the number of differentially expressed genes shared between comparisons

bhsA, genes of outer membrane pore subunits (ompC, ompF), and putrescine importer puuP were significantly more expressed in the tumor than in the subgingival samples. These results suggest that although the subgingival and tumor microbiotas differ significantly, the bacterial peptidases, toxins, and virulence factors were similarly expressed in both niches and were not exclusive to the tumor microenvironment.

Microbial-host interaction

The host transcriptome of control and tumor tissue samples was analyzed based on min. 3×106 annotated reads per sample covering the gene expression diversity. The differential expression analysis showed 2744 genes significantly more expressed in tumor than in control (adjusted p-value < 0.05) and 2648 more expressed in control than in tumor. Significantly different genes were used to establish enriched pathways in control and tumor tissue. Among the enriched pathways in control, we found pyruvate and fatty acid metabolism whereas in tumor there were pathways related to cell cycle, immune system and microRNAs in cancer (Fig. 6). This confirms that the RNAseq data of the samples reflects the health status of gut tissues. Finally, we performed correlations of the expression of those host genes with bacterial activity (Fig. 6C and Table S6). We observed that bacteria from the oral consortium identified in tumors (F. nucleatum, F. vincetii, F. polymorphum, F. periodonticum, P. micra, P. stomatis, plus H. hathewayi and C. bolteae) positively correlated with genes related to IL-17 signaling such as IL-1β, CXCL8 and FOSL1. Moreover, P. micra and P. stomatis positively correlated with the expression of specific metalloproteinases (MMP9, MMP14) and together with F. nucleatum, they exhibit a positive correlation with TGF-β2 PTGS2 and ICAM-1. Therefore, the activity of the Fusobacterium's cluster is linked with an increased expression of genes involved in inflammation and metastasis by the host, confirming the association of this microbial consortium with tumor development. Although Bacteroides fragilis and B. nordii form part of the same cluster, a list of multiple Bacteroides species shows a positive correlation with host genes over-expressed in control tissue. Two of them (i.e., B. dorei and B. sartorii) have also a negative correlation with host genes over-expressed in the tumor tissue. Similarly, a large block in the heatmap shows negative correlations between a cluster of cancer-associated host genes and a set of bacteria, that could therefore be considered health-associated. These included short-chain fatty acid (SCFA) producers like Roseburia, Dorea or Blautia, whose potential protective role should be further studied. A table with the normalized counts for each human gene has been publicly deposited (https://doi.org/10.6084/m9.figshare.29574308) for reference and further analyses.

Discussion

The present study explores the gene expression profile of CRC-associated microbiota on tumors and adjacent non-affected tissue to elucidate their contribution to carcinogenesis and disease development. As previously reported, Fn was found to be more active in tumors [16–18]. In addition, a higher activity of S. salivarius and R. intestinalis was detected in the tumor tissue. S. salivarius was found to be more present than active (higher proportion in the DNA than in the RNA pool) in the tissue in agreement with a previous report in a fecal metatranscriptome of CRC patients [40]. Interestingly, R. intestinalis has been consistently detected in CRC tissue and fecal samples [17] but it was proposed to have anti-carcinogenic effects [46].

Our correlation networks showed that Fusobacterium species were positively correlated in tumors with other periodontal disease-associated bacteria such as P. micra, P. stomatis, or other oral bacteria such as G. adiacens, as well as with the gut bacteria H. hathewayi. All of them were found more active in tumor than in control tissue in our dataset even though they were not statistically significant. We previously reported the correlation between these bacteria based on 16S rRNA gene analysis [13], and our RNAseq data confirm that these bacteria were not only co-occurrent but also co-active. Individually, these oral microbes (Fn, P. micra, P. stomatis) had been suggested as potential etiological agents of CRC, and their carcinogenic capacity has been confirmed through mouse models [6, 46]. However, our results suggest that they could be cooperating from the mouth through the gut, and future research should investigate any potential synergy between these organisms to increase their survival, capacity to invade, and enhance tumor progression (Fig. 7). Taking into consideration that H. hathewayi has also been associated with CRC [47–50], our data suggest that a potential interaction between the oral consortium and H. hathewayi could be occurring and could potentially be implicated in CRC development. It has to be borne in mind that, because of the high genomic similarity among F. nucleatum lineages, some degree of read multimapping is unavoidable. While a fraction of reads could be uniquely assigned, species-level assignments should be interpreted with caution, and our results are most robust when considered at the level of F. nucleatum as a group.

Concerning gene expression, the results showed a change in functional profile between control- and tumor-associated microbiota. While in control tissue there was a higher expression of carbon associated genes and genes

related to simple sugar and carbohydrate transport suggesting a sugar-based metabolism, tumors showed an over-expression of protein-based metabolism due to upregulation of the transcription machinery and amino acid related genes. This is in agreement with the taxonomic change observed between the niches, as Bacteroides, Blautia, Faecalicatena, and Anerostipes, the most active in control tissue, are known saccharolytic bacteria [51–54], whereas Fusobacterium and Peptostreptococcus are highly proteolytic organisms. This switch was recently reported by Lamaudière et al. who performed a metatranscriptome of fecal samples of 10 CRC patients [55]. Further studies are needed to determine whether CRC-associated microbiota is taking advantage of a proteolytic environment or if they promote it. In the gingival pockets, the oral consortium identified here is considered inflammophilic, and its members are known to benefit from the high-protein content associated with periodontitis, as a consequence of tissue breakdown and secretion of the gingival crevicular fluid [56].

Interestingly, a higher abundance of genes related to the biosynthesis and transport of LPS was found in tumor tissues confirming the observation of Lamaudière et al. [52]. LPS is an important inflammatory mediator which has been shown to have a modulatory effect on the immune response in CRC tumors and it has been proposed as a metastasis promotor [57, 58]. Accordingly, we found a positive correlation between the oral consortium activity and the host expression of Il-1β (the principal downstream effector cytokine of the inflammasomes) and CXCL-8 which promote the activation and recruitment of inflammatory innate immune cells [59, 60]. Moreover, the secretion of Il-1β induces cyclooxygenase-2 (PTGS2) and ICAM-1 which facilitate inflammatory cell recruitment and have been linked with metastasis [59, 61]. The expression of those genes and metalloproteinases was also correlated with the consortium activity, supporting the link oral consortium-inflammasome-metastasis. In addition to the Fusobacterium consortium, Bacteroides fragilis appears in the same correlation cluster, supporting previous reports that associate this species with carcinogenesis [19, 62] (Fig. 7). It is also notable that over 10 different species of Bacteroides have also been found in our analysis to correlate with genes overexpressed in control tissue (Fig. 7) and deserve further investigation. Accordingly, several species of Bacteroides, particularly B. uniformis, B. ovatus, and sometimes B. vulgatus, have previously been shown to have a negative correlation with colorectal tumor tissue, suggesting a protective or homeostatic role [9]. This underlines the importance of species-level analysis in omics studies and future experimental work will be crucial to elucidate potential causal relationships of these disease- and health-related correlations.

In addition, Wu et al. suggested that LPS could promote glucose consumption and lactate production by human cancer cells (Warburg effect) [58]. This would decrease the bioavailability of glucose in the microenvironment and could contribute to the bacterial metabolic shift (saccharolytic to proteolytic) observed in our study. On the other hand, the oligopeptide/nickel transport system, which was found increased in the tumor, has been previously related to the infection of the urinary tract by Staphylococcus aureus and has been previously reported to be upregulated during Fn invasion of CRC cells in vitro [63, 64].

Our analysis showed that among the genes more expressed in tumor, there are genes mainly expressed by Fn which correspond to known virulence factors of these species, such as Fap2 and RadD. These genes codify to well-known adhesins, which are involved in co-aggregation with other bacteria and invasion capacity [65] as well as lymphocyte death induction [66]. In addition, Fap2 is a galactose-sensitive hemagglutinin that binds acetylgalactosamine (Gal-GalNAc), which is overexpressed in colorectal adenocarcinoma cells [67] and facilitates Fn colonization in tumors. In addition, it interacts with TIGIT (an inhibitory receptor present on NK cells), decreasing NK cytotoxicity and promoting tumor-immune evasion [68] (Fig. 7). In agreement with our data, a recent study reports a high expression of FadA, Fap2, FomA, and RadD in Fn-positive CRC tissues [19]. Interestingly, a gene that codifies for another hemagglutinin (fhaB), as well as the gene for secretory protein/hemolysin activator (fhaC), was enriched in tumor tissue in our study. The activation of hemolysins facilitates a source of iron, and indeed both genes were previously suggested as main virulence candidate proteins of Fn by Zanzoni et al.

(See figure on next page.)

Fig. 6 Host transcriptome. Enriched analysis showing the pathways with upregulated human genes in tumor (**A**) and control (**B**) tissue. The size of the node represents the number of genes on the pathway. Significance based on false discovery rate is represented by color scale. **C** The expression levels of genes within the enriched pathways were correlated with bacterial activity across all samples using the sparse partial least squares (sPLS) canonical method. The heatmap displays positive (red) and negative (blue) correlations. The upper rectangles indicate whether each pathway was upregulated in tumor (purple) or control (orange) samples. Additionally, specific pathways are color-coded according to the legend. The name of the genes (x axis) is detailed in Table S6

Fig. 6 (See legend on previous page.)

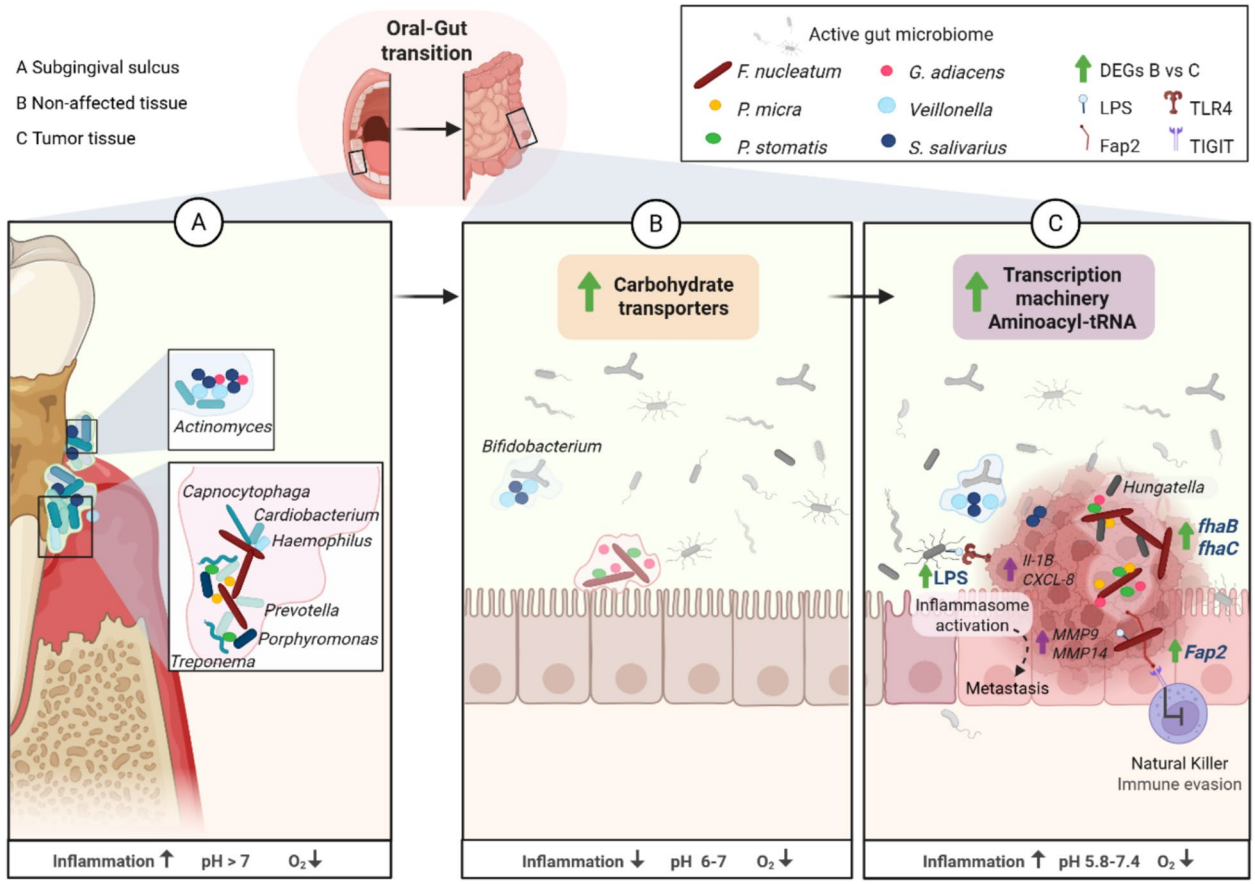

Fig. 7 Oral–gut transition model. Graphical representation of the major changes found between subgingival sulcus, gut non-affected tissue, and tumor based on metatranscriptomic analyses. **A** represents the two clusters of active bacteria found on subgingival pockets. Remaining oral clusters on gut non-affected tissue (**B**) and tumor tissue (**C**). The global enriched pathways in B vs. C were marked. In addition, the main differentially expressed genes in tumor were marked (DEGs). The bacteria LPS can interact with TLR4 and activate the inflammasome, which has been associated with the promotion of metastasis. It has been reported that Fap2 adhesins of F. nucleatum bind TIGIT receptors, inactivating natural killer and T cells, facilitating immunoevasion. The environment in parts of the tumor could be resemblant to the subgingival sulcus as it is inflamed and has hypoxic areas

[69] (fhaB corresponds to FN1817 and FN0291 genes; fhaC corresponds to FN131). Although Fn overexpresses the gene that codes for hemolysin-III during cell infection in in vitro assays [63], in our analysis, this gene was found enriched in non-affected tissue and mainly expressed by gut commensal bacteria.

In the current study, the gene expression pattern of oral bacteria in the gut tissue was compared for the first time with their expression in the oral cavity. This was a unique opportunity to test whether a microbial community undergoes transcriptomic shifts outside of their natural niche that could underlie its carcinogenetic potential. However, transcriptional changes observed could be shaped for other reasons, namely: (a) not all the oral bacteria reached the gut in all patients, thereby influencing the gene functions available for comparison; (b) bacteria could have suffered genomic rearrangements or accumulated mutations, especially in cases where the oral–gut translocation event took place longer time ago [70]; and (c) interactions and competition with other types of gut bacteria can modify their behavior. Interestingly, we did not find any of the main virulence factors from Fn differentially expressed in tumors compared to subgingival sulcus. This could be due to the subgingival samples mainly coming from patients with some degree of periodontitis, where virulence factors can already be expressed [71, 72]. As the prevalence of periodontitis in adults (> 55 years) reached 65% in Spain [73], a higher number of patients should be included to increase the number of orally healthy subjects in the analysis.

Conclusions

In short, our results confirm that Fn is more active in CRC tissue, where it is surrounded by an oral bacteria consortium (P. micra, P. stomatis, G. adiacens). In the tumor tissue, the bacterial community switched to protein as a nutrient source over carbohydrate metabolism and increased the production of LPS, increasing inflammation and metastasis. Specific virulence factors related to the adhesion of Fn were expressed in tumors, in a similar manner as observed in subgingival samples, suggesting that oral pathogens, especially periodontal bacteria, may be preadapted to the tumor environment.

Abbreviations

CRC	Colorectal caner
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
T	Tumor samples
C	Control samples
G	Subgingival samples
CHUAC	University Hospital of A Coruña
KEGG	Kyoto Encyclopedia of Genes and Genomes
ANCOM-BC	Analysis of Compositions of Microbiomes with Bias Correction
NMDS	Non-metric multidimensional scaling
sPLS	Sparse partial least squares
UHGG	Unified Human Gastrointestinal Genome
HOMD	Human oral microbiome database
RPKM	Reads per kilobase per million mapped reads
PCoA	Principal coordinates analysis
MTT	Metatranscriptome
FC	Fold Change
Fn	Fusobacterium nucleatum
LPS	Lipopolysaccharide
PTS	Phosphotransferase system
NK	Natural Killer
DEGs	Differentially expressed genes
TPM	Transcript per million
FDR	False discovery rate

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-026-02372-2.

Supplementary Material 1. Table S1. Evaluation of potential selection bias. Table S2. Oral bacteria present in tissues ANCOM normalization. Table S3. Clinical parameters of patients included. Table S4. Wilcoxon test comparing control vs. tumor ratio of relativa abundance (RNA/DNA). Table S5. Differential genes tumor-control tissue. Table S6. Correlation between human genes and bacterial species.

Supplementary Material 2. Figure S1. Percentage of human reads after sequencing against the ratio of human cells per bacteria. The number of bacteria and human cells was estimated by performing qPCR of the 16S rRNA gene and beta actin gene in the DNA sample, respectively. Figure S2. Archaea and virus in gut tissue. A) Percentage of reads annotated to each kingdom. B) Distribution of annotated archaea reads per sample, rarefaction curves and top 10 most abundant species by sample type. C) Distribution of annotated virus reads per sample, rarefaction curves, top 15 most abundant species by sample type, diversity indexes and differentially abundant species between control and tumor samples by ANCOM-BC. Figure S3. Active microbiome composition as assessed by metagenomic RNA sequencing. A) Richness and diversity estimates (left to right: rarefaction curves, Chao1 richness and Shannon diversity indexes) of control and tumor tissue at the genus taxonomic level. B) Percentage of transcripts assigned to each bacterial genera. * adjusted *p*-value < 0.05 by the ANCOM-BC method. Mean and standard deviation. Figure S4. Microbiome composition as assessed by DNA extraction and 16S gene sequencing. A) Richness and diversity estimates (left to right: rarefaction curves, Chao1 and Shannon indexes) of control and tumor tissue at the genus taxonomic level. B) Relative abundance of top-30 genera. * adjusted p-value < 0.05 by the ANCOM-BC method. Mean and standard deviation. Figure S5. Oral bacteria in gut tissue. A) Percentage of reads annotated to the Extended Human Oral Microbiome Database (EHOMD) of paired tumor:control samples. B) Mean percentage of transcripts corresponding to oral bacteria (SE, standard error). * adjusted p-value < 0.05 by Wilcoxon paired test. Figure S6. Differences at gene level between tumor and control. A) Differentially expressed genes between tumor and control paired samples. Genes with *p*-value < 0.05 and log2FC > 0.5 were colored by KEGG pathway. B) Taxonomic contributions (%) to the gene expression profile of marked genes. Figure S7. Potential bacterial interactions in the subgingival samples. Correlation networks of the oral bacteria based on their transcriptional profile in the oral cavity by the sPLS method. Lines are marked in different colors according to their positive (red) or negative (blue) correlations. Node size corresponds to relative abundance based on transcripts.

Acknowledgements

We thank Dr Nuria Jimenez for Illumina library preparation and sequencing.

Authors' contributions

KC-P, AC, JFN and CC performed patient selection and gut tissue sample collection. SP performed dental examinations and oral sample collection. KC-P performed RNA and DNA gut tissue extraction and quantification. EB performed RNA and DNA oral sample extraction, as well as qPCR. EB and MC performed bioinformatic and biostatistical analyses and drafted the figures. AM and MP conceived the study and secured the funding. EB, MC and AM wrote a draft of the manuscript. All authors revised and approved the final version of the text.

Funding

This work has been supported by the project RTI2018-102032-B-I00 from the Spanish Ministry of Science and Innovation to AM. EB was supported by a grant from the Spanish Ministry of Science and Innovation PRE2019-088126. In addition, the present study received funding from the Instituto de Salud Carlos III (ISCIII), Spain, through the projects PI20/00413 and PI23/00696, co-financed by the European Union (EU) to MP. KC was financially supported by the Spanish Association against Cancer (AECC).

Data availability

Raw data of RNA-seq available at SRA bioproject PRJNA1004440. Additionally, the code used for analysis can be found at (https://github.com/ebuetas/Metatranscriptome-CRC).

Declarations

Ethics approval and consent to participate

This study was adhered to the standards of clinical practice and research regulations (Law of Biomedical Research 14/2007), in agreement with the Declaration of Helsinki and the Convention on Human Rights and Biomedicine. Compliance with the protection of non-public personal data of all those involved within the RGPD–UE 2016/679, LOPDGDD 3/2018, Law 41/2002 and its implementing regulations, Royal Decree 1720/2007, was enforced. This project (PI20/00413), granted by Carlos III Health Institute (ISCIII; Spain), was supervised by the local ethics committee, the Research Ethical Committee of Galicia (code CEIm-G 2018/609, Galicia, Spain), and by the Spanish Agency for Medicines and Healthcare Products (AEMPS) for the use of CRC patients' samples from CHUAC (A Coruña, Galicia, Spain). Informed consent for Biobank (CHUAC, A Coruña, Galicia, Spain, UNE-EN ISO 9001-2015) was signed by all individuals grouped in this study. Anonymized clinical data used during the study for CRC patients was obtained from the Galician Health Service (SERGAS).

Consent for publication

All individuals recruited in this project (CRC and non-CRC) signed a formal consent form for the publication of scientific and clinical results in scientific articles.

Competing interests
The authors declare no competing interests.

Author details
[1]Genomic and Health Department, Center for Advanced Research in Public Health, FISABIO Foundation, Valencia, Spain. [2]Grupo de Investigación en Microbiología, Instituto de Investigación Biomédica de A Coruña (INIBIC), Complejo Hospitalario Universitario de A Coruña (CHUAC), Sergas, Universidade de A Coruña (UDC), A Coruña, Galicia, Spain. [3]Servicio de Anatomía Patológica, Complejo Hospitalario Universitario de A Coruña (CHUAC), Sergas, A Coruña, Galicia, Spain. [4]Servicio de Cirugía General, Hospital Universitario Obispo Polanco, Teruel, Teruel, Spain. [5]CIBER de Enfermedades Infecciosas (CIBERINFEC-ISCIII), Madrid, Spain. [6]Periodontology and Oral Surgery, Clínica Médico Dental Pardiñas, A Coruña, Spain. [7]Grupo de Terapia Celular y Medicina Regenerativa, Instituto de Investigación Biomédica de A Coruña (INIBIC), Servizo Galego de Saúde (SERGAS), Complexo Hospitalario Universitario de A Coruña (CHUAC), A Coruña, Spain. [8]Department of Oral and Maxillofacial Surgery, College of Dentistry, New York University, New York, NY, USA. [9]Grupo de Microbioma y Salud, Centro Interdisciplinar de Química E Bioloxía (CICA), Facultad de Ciencias, Universidade da Coruña (UDC), Campus da Zapateira, Galicia, A Coruña, Spain. [10]CIBER Center for Epidemiology and Public Health, Madrid, Spain.

Received: 31 July 2025 Accepted: 4 February 2026
Published online: 12 May 2026

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Keum NN, Giovannucci E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. Nat Rev Gastroenterol Hepatol. 2019;16:713–32.
3. Abu-Ghazaleh N, Chua WJ, Gopalan V. Intestinal microbiota and its association with colon cancer and red/processed meat consumption. J Gastroenterol Hepatol. 2020. Available from: https://doi.org/10.1111/jgh.15042.
4. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12:31–46.
5. Alhinai EA, Walton GE, Commane DM. The role of the gut microbiota in colorectal cancer causation. Int J Mol Sci. 2019. Available from: https://doi.org/10.3390/ijms20215295. MDPI AG.
6. White MT, Sears CL. The microbial landscape of colorectal cancer. Nat Rev Microbiol. 2024;22:240–54.
7. Mira-Pascual L, Cabrera-Rubio R, Ocon S, Costales P, Parra A, Suarez A, et al. Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. J Gastroenterol. 2015;50:167–79.
8. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med. 2019;25:667–78.
9. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019;25:679–89.
10. Chen T, Marsh PD, Al-Hebshi NN. SMDI: an index for measuring subgingival microbial dysbiosis. J Dent Res. 2022;101:331–8.
11. Li W, Xu J, Zhang R, Li Y, Wang J, Zhang X, et al. Is periodontal disease a risk indicator for colorectal cancer? A systematic review and meta-analysis. J Clin Periodontol. 2021;48:336–47.
12. Xuan K, Jha AR, Zhao T, Uy JP, Sun C. Is periodontal disease associated with increased risk of colorectal cancer? A meta-analysis. Int J Dent Hyg. 2021;19:50–61.
13. Conde-Pérez K, Aja-Macaya P, Buetas E, Trigo-Tasende N, Nasser-Ali M, Rumbo-Feal S, et al. The multispecies microbial cluster of *Fusobacterium*, *Parvimonas*, *Bacteroides* and *Faecalibacterium* as a precision biomarker for colorectal cancer diagnosis. Mol Oncol. 2024;18:1093–122.
14. Debelius JW, Engstrand L, Matussek A, Brusselaers N, Morton JT, Stenmarker M, et al. The local tumor microbiome is associated with survival in late-stage colorectal cancer patients. Microbiol Spectr. 2023;11:e0506622.
15. Zhao L, Cho WC, Nicolls MR. Colorectal Cancer-associated microbiome patterns and signatures. Front Genet. 2021;12:1–15.
16. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. 2012;22:299–306.
17. Zhao L, Grimes SM, Greer SU, Kubit M, Lee H, Nadauld LD, et al. Characterization of the consensus mucosal microbiome of colorectal cancer. NAR Cancer. 2021;3. Available from: https://doi.org/10.1093/NARCAN/ZCAB049.
18. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas. Microbiome. 2013;1:16.
19. Kvich L, Fritz BG, Zschach H, Terkelsen T, Raskov H, Høst-Rasmussen K, et al. Biofilms and core pathogens shape the tumor microenvironment and immune phenotype in colorectal cancer. Gut Microbes. 2024;16:2350156.
20. Papapanou PN, Sanz M, Buduneli N, Dietrich T, Feres M, Fine DH, et al. Periodontitis: consensus report of workgroup 2 of the 2017 World Workshop on the Classification of Periodontal and Peri-Implant Diseases and Conditions: classification and case definitions for periodontitis. J Periodontol. 2018;89(Suppl 1):S173–82.
21. Löe H. The gingival index, the plaque index and the retention index systems. J Periodontol. 1967;38(Suppl):610–6.
22. Bauset C, Carda-Diéguez M, Cejudo-Garcés A, Buetas E, Seco-Cervera M, Macias-Ceja DC, et al. A disturbed metabolite-GPCR axis is associated with microbial dysbiosis in IBD patients: potential role of GPR109A in macrophages. Biochim Biophys Acta Mol Basis Dis. 2024;1870:167489.
23. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17:10.
24. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863–4.
25. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590.
26. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376:44–53.
27. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:257.
28. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat Biotechnol. 2021;39:105–14.
29. Carda-Diéguez M, Rosier BT, Lloret S, Llena C, Mira A. The tongue biofilm metatranscriptome identifies metabolic pathways associated with the presence or absence of halitosis. NPJ Biofilms Microbiomes. 2022;8:100.
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
32. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. Prlic A, editor. PLoS Comput Biol. 2013;9:e1003118.
33. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic acids research. Oxford University Press. 2000:27–30. Available from: https://doi.org/10.1093/nar/28.1.27.
34. Rice P, Longden L, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends in genetics. Elsevier Ltd. 2000:276–7. Available from: https://doi.org/10.1016/S0168-9525(00)02024-2.
35. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011;7:e1002195. https://doi.org/10.1371/journal.pcbi.1002195.
36. R Core Team. A language and environment for statistical computing. R Foundation Stat Comput. 2012;10:11–8.
37. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36:2628–9.
38. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Comput Biol. 2017;13:e1005752.

39. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. Nat Commun. 2020;11:1–11.
40. Lamaudière MTF, Arasaradnam R, Weedall GD, Morozov IY. The colorectal cancer gut environment regulates activity of the microbiome and promotes the multidrug resistant phenotype of ESKAPE and other pathogens. mSphere. 2023;8:e0062622.
41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
42. Falcon S, Gentleman R. Hypergeometric testing used for gene set enrichment analysis. Bioconductor case studies. New York, NY: Springer New York. 2008:207–20.
43. Kursa MB, Rudnicki WR. Feature selection with the Boruta package. J Stat Softw. 2010;36:1–13.
44. Sanders BE, Umana A, Lemkul JA, Slade DJ. FusoPortal: an interactive repository of hybrid MinION-sequenced Fusobacterium genomes improves gene identification and characterization. mSphere. 2018;3. Available from: https://doi.org/10.1128/msphere.00228-18.
45. Umaña A, Sanders BE, Yoo CC, Casasanta MA, Udayasuryan B, Verbridge SS, et al. Utilizing whole fusobacterium genomes to identify, correct, and characterize potential virulence protein families. J Bacteriol. 2019;201. Available from: https://doi.org/10.1128/JB.00273-19.
46. Montalban-Arques A, Katkeviciute E, Busenhart P, Bircher A, Wirbel J, Zeller G, et al. Commensal Clostridiales strains mediate effective anti-cancer immune response against solid tumors. Cell Host Microbe. 2021;29:1573-1588.e7.
47. Gutierrez-Angulo M, Ayala-Madrigal MdelaL, Moreno-Ortiz JM, Peregrina-Sandoval J, Garcia-Ayala FD. Microbiota composition and its impact on DNA methylation in colorectal cancer. Front Genet. 2023;14:1037406.
48. Huang Z, Wang C, Huang Q, Yan Z, Yin Z. *Hungatella hathewayi* impairs the sensitivity of colorectal cancer cells to 5-FU through decreasing CDX2 expression. Hum Cell. 2023;36:2055–65.
49. Zhang J, He Y, Xia L, Yi J, Wang Z, Zhao Y, et al. Expansion of colorectal cancer biomarkers based on gut bacteria and viruses. Cancers. 2022;14:4662.
50. Xia X, Wu WKK, Wong SH, Liu D, Kwong TNY, Nakatsu G, et al. Bacteria pathogens drive host colonic epithelial cell promoter hypermethylation of tumor suppressor genes in colorectal cancer. Microbiome. 2020;8:108.
51. Liu X, Mao B, Gu J, Wu J, Cui S, Wang G, et al. Blautia-a new functional genus with potential probiotic properties? Gut Microbes. 2021;13:1–21.
52. Sakamoto M, Iino T, Ohkuma M. *Faecalimonas umbilicata* gen. nov., sp. nov., isolated from human faeces, and reclassification of *Eubacterium contortum*, *Eubacterium fissicatena* and *Clostridium oroticum* as *Faecalicatena contorta* gen. nov., comb. nov., *Faecalicatena fissicatena* comb. nov. and *Faecalicatena orotica* comb. nov. Int J Syst Evol Microbiol. 2017;67:1219–27.
53. Schwiertz A, Hold GL, Duncan SH, Gruhl B, Collins MD, Lawson PA, et al. *Anaerostipes caccae* gen. nov., sp. nov., a new saccharolytic, acetate-utilising, butyrate-producing bacterium from human faeces. Syst Appl Microbiol. 2002;25:46–51.
54. Zafar H, Saier MH Jr. Gut Bacteroides species in health and disease. Gut Microbes. 2021;13:1–20.
55. Lamaudière MTF, Arasaradnam R, Weedall GD, Morozov IY. The colorectal cancer microbiota alter their transcriptome to adapt to the acidity, reactive oxygen species, and metabolite availability of gut microenvironments. mSphere. 2023;8:e0062722.
56. Mira A, Simon-Soro A, Curtis MA. Role of microbial communities in the pathogenesis of periodontal diseases and caries. J Clin Periodontol. 2017;44(Suppl 18):S23-38.
57. Sulit AK, Daigneault M, Allen-Vercoe E, Silander OK, Hock B, McKenzie J, et al. Bacterial lipopolysaccharide modulates immune response in the colorectal tumor microenvironment. NPJ Biofilms Microbiomes. 2023;9:59.
58. Wu X, Qian S, Zhang J, Feng J, Luo K, Sun L, et al. Lipopolysaccharide promotes metastasis via acceleration of glycolysis by the nuclear factor-κB/snail/hexokinase3 signaling axis in colorectal cancer. Cancer Metab. 2021;9:23.
59. Jang JH, Kim DH, Surh YJ. Dynamic roles of inflammasomes in inflammatory tumor microenvironment. NPJ Precis Oncol. 2021;5:18.
60. Dawson RE, Jenkins BJ. The role of inflammasome-associated innate immune receptors in cancer. Immune Netw. 2024;24:e38.
61. Lim EJ, Kang JH, Kim YJ, Kim S, Lee SJ. ICAM-1 promotes cancer progression by regulating SRC activity as an adapter protein in colorectal cancer. Cell Death Dis. 2022;13:417.
62. 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;15:1016–22.
63. Cochrane K, Robinson AV, Holt RA, Allen-Vercoe E. A survey of Fusobacterium nucleatum genes modulated by host cell infection. Microb Genom. 2020;6. Available from: https://doi.org/10.1099/mgen.0.000300.
64. Hiron A, Posteraro B, Carrière M, Remy L, Delporte C, La Sorda M, et al. A nickel ABC-transporter of Staphylococcus aureus is involved in urinary tract infection. Mol Microbiol. 2010;77:1246–60.
65. Groeger S, Zhou Y, Ruf S, Meyle J. Pathogenic mechanisms of *Fusobacterium nucleatum* on oral epithelial cells. Frontiers in Oral Health. 2022;3:831607.
66. Kaplan CW, Ma X, Paranjpe A, Jewett A, Lux R, Kinder-Haake S, et al. *Fusobacterium nucleatum* outer membrane proteins Fap2 and RadD induce cell death in human lymphocytes. Infect Immun. 2010;78:4773–8.
67. Abed J, Emgård JEM, Zamir G, Faroja M, Almogy G, Grenov A, et al. Fap2 mediates Fusobacterium nucleatum colorectal adenocarcinoma enrichment by binding to tumor-expressed Gal-GalNAc. Cell Host Microbe. 2016;20:215–25.
68. Wu J, Li Q, Fu X. Fusobacterium nucleatum Contributes to the Carcinogenesis of Colorectal Cancer by Inducing Inflammation and Suppressing Host Immunity. Transl Oncol. 2019;12:846–51.
69. Zanzoni A, Spinelli L, Braham S, Brun C. Perturbed human sub-networks by *Fusobacterium nucleatum* candidate virulence proteins. Microbiome. 2017;5:89.
70. Conde-Pérez K, Buetas E, Aja-Macaya P, Martin-De Arribas E, Iglesias-Corrás I, Trigo-Tasende N, et al. Parvimonas micra can translocate from the subgingival sulcus of the human oral cavity to colorectal adenocarcinoma. Mol Oncol. 2024;18:1143–73.
71. Ram-Mohan N, Meyer MM. Comparative metatranscriptomics of periodontitis supports a common polymicrobial shift in metabolic function and identifies novel putative disease-associated ncRNAs. Front Microbiol. 2020;11:482.
72. Yost S, Duran-Pinedo AE, Teles R, Krishnan K, Frias-Lopez J. Functional signatures of oral dysbiosis during periodontitis progression revealed by microbial metatranscriptome analysis. Genome Med. 2015;7:27.
73. Carasol M, Llodra JC, Fernández-Meseguer A, Bravo M, García-Margallo MT, Calvo-Bonacho E, et al. Periodontal conditions among employed adults in Spain. J Clin Periodontol. 2016;43:548–56.

Publisher's Note

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