

RESEARCH

Open Access



Distinct microbes, metabolites, and the host genome define the multi-omics profiles in right-sided and left-sided colon cancer

Lei Liang^{1,2†}, Cheng Kong^{1,2†}, Jinming Li^{1,2†}, Guang Liu^{3†}, Jinwang Wei⁴, Guan Wang⁴, Qinying Wang^{1,2}, Yongzhi Yang^{1,2}, Debing Shi^{1,2}, Xinxiang Li^{1,2} and Yanlei Ma^{1,2*}

Abstract

Background Studies have reported clinical heterogeneity between right-sided colon cancer (RCC) and left-sided colon cancer (LCC). However, none of these studies used multi-omics analysis combining genetic regulation, microbiota, and metabolites to explain the site-specific difference.

Methods Here, 494 participants from a 16S rRNA gene sequencing cohort (50 RCC, 114 LCC, and 100 healthy controls) and a multi-omics cohort (63 RCC, 79 LCC, and 88 healthy controls) were analyzed. 16S rRNA gene, metagenomic sequencing, and metabolomics analyses of fecal samples were evaluated to identify tumor location-related bacteria and metabolites. Whole-exome sequencing (WES) and transcriptome sequencing (RNA-seq) were conducted to obtain the mutation burden and genomic expression pattern.

Results We found unique profiles of the intestinal microbiome, metabolome, and host genome between RCC and LCC. The bacteria *Flavonifractor plautii* (*Fp*) and *Fusobacterium nucleatum*, the metabolites L-phenylalanine, and the host genes *PHLDA1* and *WBP1* were the key omics features of RCC; whereas the bacteria *Bacteroides* sp. A1C1 (*B.A1C1*) and *Parvimonas micra*, the metabolites L-citrulline and D-ornithine, and the host genes *TCF25* and *HLA-DRB5* were considered the dominant omics features in LCC. Multi-omics correlation analysis indicated that RCC-enriched *Fp* was related to the accumulation of the metabolite L-phenylalanine and the suppressed *WBP1* signal in RCC patients. In addition, LCC-enriched *B.A1C1* was associated with the accumulation of the metabolites D-ornithine and L-citrulline as well as activation of the genes *TCF25*, *HLA-DRB5*, and *AC079354.1*.

Conclusion Our findings identify previously unknown links between intestinal microbiota alterations, metabolites, and host genomics in RCC vs. LCC, suggesting that it may be possible to treat colorectal cancer (CRC) by targeting the gut microbiota–host interaction.

Keywords Left-sided colon cancer, Right-sided colon cancer, Metagenomics, Metabolomics, Host genomics, Multi-omics analysis

[†]Lei Liang, Cheng Kong, Jinming Li, and Guang Liu contributed equally to this work.

*Correspondence:

Yanlei Ma

yanleima@fudan.edu.cn

Full list of author information is available at the end of the article



Background

The difference between left-sided colon cancer (LCC) and right-sided colon cancer (RCC) has gained global attention due to the heavy burden of this cancer in both developing countries and developed countries [1, 2]. Although colon cancer is often grouped as one disease, accumulating evidence has demonstrated biological and clinical differences between proximal and distal colon tumors. For this reason, colon cancer is topographically classified into RCC, originating from the colon proximal to the splenic flexure, including the cecum, ascending colon, and transverse colon, and LCC, arising from the colorectum distal to the splenic flexure, including the descending colon, sigmoid colon, and rectosigmoid [3–8]. This classification considers the embryonic development whereby the proximal colon originates from the embryonic midgut and is perfused by the superior mesenteric artery, whereas the distal colorectum is derived from the hindgut and is served by the inferior mesenteric artery. The discrepancies among RCC and LCC, including an older patient age at diagnosis, poorer differentiation, a more advanced stage, and a worse prognosis for RCC vs. LCC, corroborate the classification, which is potentially caused by the distinct carcinogenesis and pathological behaviors attributed to genetic differences [9–12]. Previous studies have sought to explain these differences by analyzing significantly mutated genes and RNA expression [13]. However, these studies merely analyze the association between clinicopathologic features and genetic profiles, and there is no integrated multi-omics analysis that illustrates the distinct molecular profiles and the corresponding patient prognosis among different locations of colon cancer.

In recent years, the advancement of sequencing techniques has enabled researchers to investigate and understand the involvement of microbes and metabolites in the development of colonic tumorigenesis. Previous studies have identified some site-specific bacteria [14–16] and distinct host genetic characteristics between RCC and LCC [9, 13, 17]. However, most of these studies only detected bacteria at the genus level and did not use multi-omics analysis combining host genetic regulation, microbiota, and metabolites to explain the site-specific difference of colon cancer. Therefore, the present study aimed to investigate the gut microbiome, metabolome, and host genome spectrums and their multi-omics association between patients with RCC vs. LCC.

Methods

Clinical specimens and study design

Feces and fresh tumor specimens were prospectively obtained from patients with primary sporadic colorectal cancer (CRC) and healthy controls at Fudan University Shanghai Cancer Center, Shanghai, China. For the

primary sporadic CRC group, fecal samples were collected preoperatively, and fresh tumor specimens were collected during surgery. For the healthy control group, volunteers confirmed as not having any gastrointestinal tumors after colonoscopy screening were recruited. Each fecal and tissue sample was collected in a sterile tube and then stored at -80°C prior to microbial and genomic analyses. None of the participants were treated with chemotherapy, antibiotics, or probiotics for one month before enrollment in this study. Ethical approval was obtained from the Institutional Review Board of Fudan University Shanghai Cancer Center, and written informed consent was provided by all participants before sampling.

This study consisted of two cohorts, including a 16S rRNA gene sequencing cohort (16S-seq) and a multi-omics cohort [metagenomics, metabolomics, tumor whole-exome sequencing, and transcriptome sequencing (Meta/liquid chromatography-mass spectrometry (LC-MS)/ WES/RNA-seq)]. Overall, 494 patients with RCC ($n=113$), LCC ($n=193$), or healthy controls ($n=188$) were analyzed. The 16S-seq cohort consisted of 264 CRC patients and healthy controls whose fecal samples were collected preoperatively and detected using 16S rRNA gene sequencing. Data were analyzed to search for bacteria that were associated with the tumor location: i.e., RCC ($n=50$), LCC ($n=114$), and healthy controls (control, $n=100$). In the multi-omics cohort, fecal samples from 230 preoperative patients with CRC and healthy controls were collected and detected by metagenomic sequencing and LC-MS to further obtain deeper insights into the gut microbiota species and metabolite identification. Among them, 59 matched fresh tumor samples (26 RCC patients and 33 LCC patients) were extracted from the multi-omics cohort (all existing samples are available from the tissue biobank), and WES and RNA-seq of the tumor samples were performed to obtain the genomic expression patterns related to the tumor locations. The multi-omics design and key results are summarized in Fig. 1.

Fecal DNA extraction for microbiome analysis

Genomic DNA of the fecal samples was extracted by using the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's guidelines. The DNA integrity and size were verified by 1.0% agarose gel electrophoresis, and DNA concentrations were determined by using a NanoDrop spectrophotometer (NanoDrop, Germany).

High-throughput 16S ribosomal RNA gene sequencing

16S rRNA gene amplification was performed using the primers (319F: 5'-ACTCCTACGGGAGGCAGCAG-3';

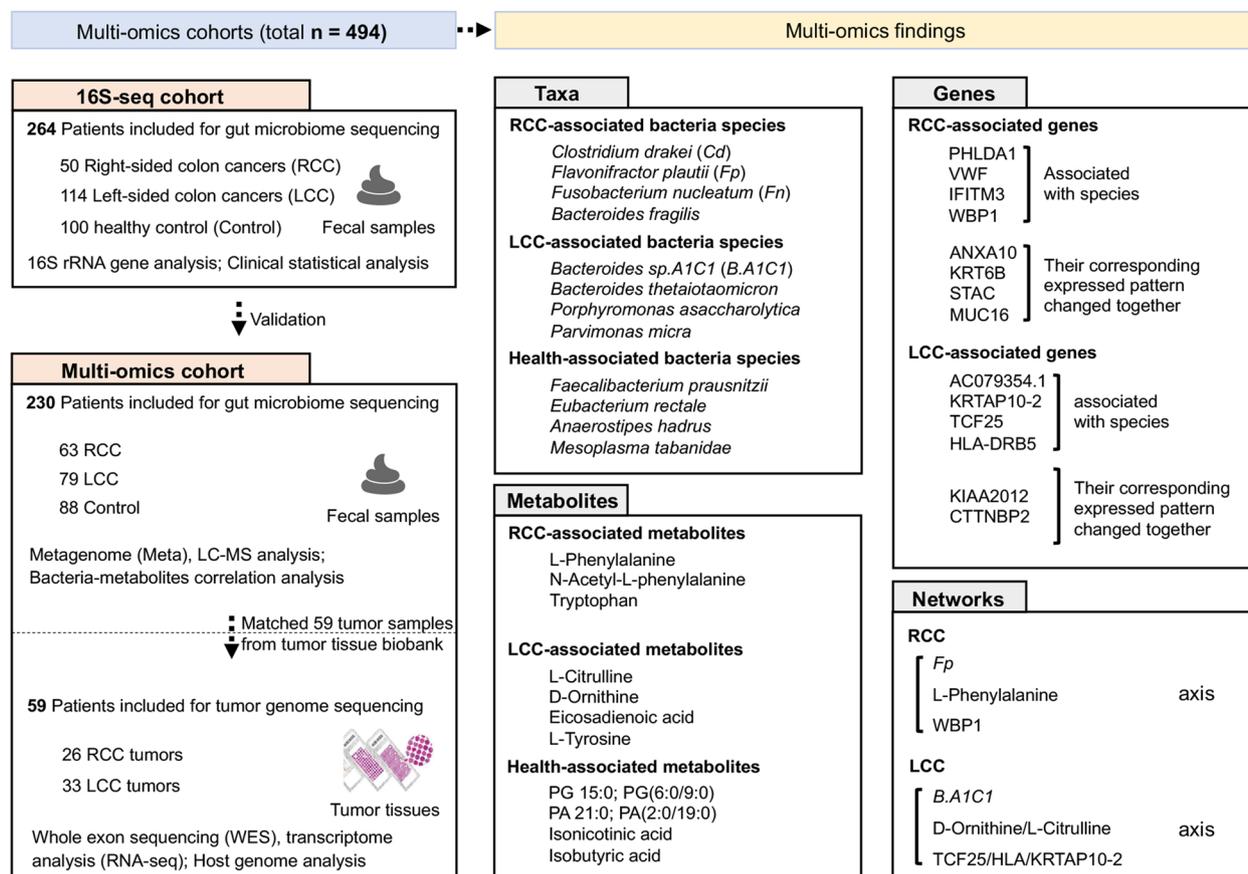


Fig. 1 A schematic flow diagram of the study design. This study consisted of two cohorts, including a 16S rRNA gene sequencing cohort (16S-seq) and a multi-omics cohort (metagenomics, metabolomics, tumor whole-exome sequencing, and transcriptome sequencing (Meta/LC-MS/WES/RNA-seq)). Overall, 494 patients with right-sided colon cancer (RCC, $n = 113$), left-sided colon cancer (LCC, $n = 193$), or healthy controls ($n = 188$) were analyzed. The 16S-seq cohort consisted of 264 colorectal cancer (CRC) patients and healthy controls whose fecal samples were collected preoperatively and detected using 16S rRNA gene sequencing. Data were analyzed to search for bacteria that were associated with the tumor location: i.e., RCC ($n = 50$), LCC ($n = 114$), and healthy controls (control, $n = 100$). In the multi-omics cohort, the fecal samples from 230 preoperative patients with CRC and healthy controls were collected and detected by metagenomic sequencing and LC-MS to further obtain deeper insights into the gut microbiota species and metabolite identification. Among them, 59 matched fresh tumor samples (26 RCC patients and 33 LCC patients) were extracted from the multi-omics cohort in our tissue biobank, and WES and RNA-seq of the tumor samples were performed to obtain genomic expression patterns related to each tumor location. The multi-omics results are summarized in the Taxa, Metabolites, Genes, and Network boxes

806R: 5'-GGACTACHVGGGTWTCTAAT-3') directionally targeting the V3 and V4 hypervariable regions of the 16S rRNA gene. To differentiate each sample and yield accurate phylogenetic and taxonomic information, the gene products were attached with forward and reverse error-correcting barcodes. The amplicons were quantified after purification and sequenced on the MiSeq PE300 sequencing instrument (Illumina, San Diego, CA, USA) by using 2×300 bp chemistry, according to the manufacturer's specifications.

DNA library construction and metagenomic sequencing

Sequencing libraries were constructed by a TruSeq Nano DNA LT Library Preparation Kit (Illumina). DNA

was fragmented by dsDNA Fragmentase (New England Biolabs, Ipswich, MA, USA) and incubated at 37 °C for 30 min. Blunt-end DNA fragments were generated by fill-in reactions and exonuclease activity. The provided sample purification beads were used for size selection. An A-base was then added to the blunt ends of each strand, preparing them for ligation to the indexed adapters. Each adapter contained a T-base overhang for ligating the adapter to the A-tailed fragmented DNA. These adapters contained the full complement of sequencing primer hybridization sites for single, paired-end, and indexed reads. Single- or dual-index adapters were ligated to the fragments, and the ligated products were amplified by the polymerase chain reaction. After library purification,

quantification, and quality control, high-throughput sequencing was carried out on the NovaSeq6000 platform (Illumina), according to the manufacturer's specifications.

Sequencing data analysis

16S rRNA gene sequencing data

The raw data of 16S rRNA gene sequencing were analyzed by using the QIIME2 platform (v2020.2) [18, 19]. Briefly, the DADA2 plugin was used to filter the sequencing reads and to obtain the feature tables and feature sequences. The taxonomy information of the sequencing reads was assigned against the Silva Database (v138.1). Alpha and beta diversity analyses were conducted by using the diversity plugin. Principal component analysis was conducted to display the distance among samples. To identify differentially abundant taxa among groups, the Wilcoxon rank-sum test with the Benjamini–Hochberg method was performed on the fecal microbiota composition between groups.

Metagenomic sequencing data

The raw sequencing reads were processed to obtain valid reads for further analysis. First, the sequencing adapters were removed from the sequencing reads by using cutadapt (v1.9). Second, the low-quality reads were trimmed by using fqtrim (v0.94) with a sliding-window algorithm. Third, the reads were aligned to the host genome by using bowtie2 (v2.2.0) to remove host contamination. In order to obtain the species-level information, the clean reads were aligned to the suggested database (v202003) by using Kraken2 software (v2.1.1) [20] and Braken software (v2.5) [20]. The database can be freely downloaded from the Kraken2 website (ftp://ftp.ccb.jhu.edu/pub/data/kraken2_dbs/).

Fecal metabolite extraction and LC–MS analysis for the Meta/LC–MS cohort

The fecal metabolite extraction method was adapted from a previously published protocol [21]. First, all chromatographic separations were performed by using an ultra-performance liquid chromatography (UPLC) system (SCIEX, Macclesfield, UK). An ACQUITY UPLC HSS T3 column (100 mm × 2.1 mm, 1.8 μm, Waters, UK) was used for the reversed-phase separation. The column oven was maintained at 35 °C, and the flow rate was 0.4 mL/min. The mobile phase consisted of solvent A (water, 0.1% formic acid) and solvent B (acetonitrile, 0.1% formic acid).

A high-resolution tandem mass spectrometer (TripleTOF5600plus, SCIEX) was used to detect the metabolites that eluted from the column. The Quadrupole Time-of-Flight was operated in both positive- and

negative-ion modes. The curtain gas, ion source gas 1, and ion source gas 2 were set at 30 psi, 60 psi, and 60 psi, respectively, and the interface heater temperature was 650°C. For positive- and negative-ion modes, the ion-spray voltage floating was set at 5000 V and –4500 V, respectively. The mass spectrometry data were acquired in information-dependent acquisition mode. The time-of-flight mass range was 60–1200 Da. During the acquisition, the mass accuracy was calibrated every 20 samples. Furthermore, a quality control sample (pool of all samples) was acquired after every 10 samples in order to evaluate the stability of the LC–MS during the whole acquisition. The detailed methods for LC–MS raw data file processing, calculation of the exact molecular mass data (m/z), online Kyoto Encyclopedia of Genes and Genomes (KEGG) and Human Metabolome Database (HMDB) database annotation, and quality control have been described previously [22].

After processing the information with eXtensible Computational Mass Spectrometry (XCMS) software, we initially matched the m/z values with the KEGG database using the open-source software metaX to obtain the primary metabolite identification results. Due to the possibility of multiple metabolites having similar mass-to-charge ratios, the primary identification results may have included cases where a single m/z corresponds to multiple metabolites. For more reliable metabolite identification, we utilized an in-house secondary mass spectrometry database to match the metabolite data, which resulted in a higher confidence of metabolite identification. In terms of ion annotation, we converted the raw mass spectrometry data into a readable format (mzXML) using Proteowizard's MSConvert software. Subsequently, we conducted peak extraction and quality control using XCMS, and the substances extracted were subjected to ion annotation using CAMERA. Our metabolite identification process can be summarized as follows: (1) Conversion of the raw data to the mzXML format using MSConvert. (2) Peak extraction and quality control using XCMS, with ion annotation performed using CAMERA. (3) Metabolite identification using metaX, which involves primary identification through database matching (based on the MS1 information) and secondary identification through matching with an in-house standard compound database (based on the MS2 information). (4) Annotation of candidate-identified substances using databases such as HMDB and KEGG to explain their physicochemical properties and biological functions. (5) Quantification of metabolites and screening for differential metabolites using metaX. The software versions used for this analysis are as follows: metaX (version 1.4.19); xcms (version 3.9.3). Metabolite annotation was based on the KEGG results for secondary metabolites (MS2Metabolite).

For quality tolerance and feature filtering, we utilized XCMS to extract the signal intensity information of each feature across different samples, followed by quality control using metaX. First, we removed the low-quality peaks (those missing in more than 50% of the QC samples or more than 80% of the actual samples). Subsequently, we employed the K-nearest neighbors (KNN) method to impute missing values and then performed data normalization using probabilistic quotient normalization (PQN) and QC-robust spline batch correction (QC-RSC). Following data correction, we applied a filtering step, removing all ions with a coefficient of variation (CV) greater than 30% in the QC samples (i.e., ions with $CV > 30\%$ exhibit significant fluctuations during the experimental process and are not suitable for differential quantification analysis).

RNA-seq

RNA was extracted using an RNeasy Plus Mini Kit and quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop, Wilmington, DE, USA). The RNA integrity was assessed by a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) and confirmed by electrophoresis with denaturing agarose gel electrophoresis. Poly(A) RNA was purified from 1 μg of total RNA using Dynabeads Oligo (Thermo Fisher Scientific, Waltham, MA, USA) using two rounds of purification. Cleaved RNA fragments were copied into double-stranded cDNA using reverse transcriptase, random primers, DNA polymerase, and RNase H, in sequence. Following ligation of the sequencing adapter, the cDNA products were purified. Finally, we performed the 2×150 -bp paired-end sequencing (PE150) on an Illumina Novaseq™ 6000 instrument (LC-Bio Technology CO., Ltd., Hangzhou, China). After removing the low-quality bases and undetermined bases, HISAT2 software was used to map the reads to the genome. After the final transcriptome was generated, StringTie and ballgown were used to estimate the expression levels of all transcripts and to calculate the fragments per kilobase of transcript per million mapped reads.

Sample preparation and WES

The patients in the WES cohort were included for the evaluation of the genomic expression pattern related to the *Fn* levels and tumor locations. WES and analysis were performed at the Genomics Laboratory of GenomicCare Biotechnology (Shanghai, China). For frozen tissue or blood, DNA was extracted from thawed materials by using a Maxwell RSC Blood DNA Kit (cat# AS1400, Promega, Madison, WI, USA) on a Maxwell RSC system (cat# AS4500, Promega). For formalin-fixed, paraffin-embedded (FFPE) tissue, DNA was extracted using a MagMAX FFPE DNA/RNA Ultra Kit (cat# A31881,

Thermo Fisher Scientific) on a KingFisher Flex system (Thermo Fisher Scientific). The extracted DNA was sheared by using a Covaris L220 sonicator, and then the exomic DNA was captured by using a SureSelect Human All Exon V7 kit (cat# 5991-9039EN, Agilent), prepared to the library by using the SureSelectXT Low Input Target Enrichment and Library Preparation system (cat# G9703-90000, Agilent), and sequenced on an Illumina NovaSeq-6000 sequencer (Illumina) to generate 2×150 -bp paired-end reads. Image analysis and base calling were performed by using RTA3 software (Illumina).

Somatic variant identification

The running environment with default parameters was implemented to process the following steps sequentially: read alignment to the National Center for Biotechnology Information human genome reference assembly hg19 by using the Burrows–Wheeler Aligner algorithm, duplication sorting, realignment and recalibration, and somatic mutation calling including single nucleotide variations (SNVs) and short insertion/deletions. During the mutation calling stage, the reads from the tumor samples were compared with the paired blood from the same patient to generate the somatic mutation list. The called somatic mutations were then filtered, meaning that only the mutations with a variant allele frequency ≥ 0.05 and supported by at least three reads were retained, and annotated using the Variant Effect Predictor package [23].

The most frequently mutated genes and their distribution across different clinical features, the significantly different mutated genes between groups, and the mutually exclusive or co-occurring mutated genes were calculated and visualized by the MAFtools package in R [24]. Gene set enrichment analysis (GSEA) was performed by using GSEA software (<http://software.broadinstitute.org/gsea/index.jsp>) [25]. The most frequently mutated genes and their distribution across different clinical features as well as the significantly different mutated genes between groups were calculated and visualized by the MAFtools package (v2.6.05) in R. The “OncogenicPathways” module in MAFtools was used to examine the enrichment of ten canonical oncogenic signaling pathways derived from The Cancer Genome Atlas cohorts, including the cell cycle, Hippo, Myc, Notch, Nrf2, PI3K/Akt, RTK-RAS, TGF β , TP53, and Wnt signaling pathways. Next, the mutated genes were assigned to oncogenic signaling pathways, and the fractions of genes affected were calculated for each pathway.

As is well known, the intestinal microbial communities are closely related to human health, which can be influenced by many factors, such as lifestyle, host disease, and host genetics. Two different methods were used to test the associations between host genetics and the

abundance of gut bacteria as well as metabolites. First, Spearman correlation analysis was used to calculate the correlation. Correlations resulting from the Spearman analysis were further analyzed using a generalized linear model (GLM) to adjust for age and sex [26]. The count abundance was submitted to logistic regression using GLM with a negative binomial distribution using the number of SNVs as a predictor, including the covariates age and sex. The `Manyglm` function from the `mvabund` package (v 4.2.1) in R (v4.0.0) was used in this process.

Statistical analysis

Comparison of quantitative data between groups was conducted by using the unpaired Student's *t*-test, Mann–Whitney *U*-test, or Dunnett's *t*-test, where appropriate. The associations between or among the clinical characteristics were determined by Pearson's chi-squared test or Fisher's exact test. The linear discriminant analysis effect size (LEfSe, <https://huttenhower.sph.harvard.edu/galaxy/>) algorithm was used to identify the taxonomic features and metabolite features that are differentially abundant between groups. DESeq2 [27] (R Package v1.30.1) was used to perform the differential gene expression analysis based on the negative binomial distribution with the RNA-seq data. Spearman correlation analysis was performed to analyze the correlation between the gut microbiota, metabolites, and RNA-seq data. All *p*-values were two-tailed, and *p*-values ≤ 0.05 were considered statistically significant. All data were analyzed by using Graph Pad Prism 8.0 software (Graph Pad software Inc., San Diego, CA, USA), R version 4.0.0 (R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>), and Microsoft Excel (Microsoft Corporation, Seattle, WA, USA).

As for the method to identify the taxonomic biomarkers that are differentially abundant between groups, the linear discriminant analysis effect size (LEfSe, <https://huttenhower.sph.harvard.edu/galaxy/>) algorithm was performed, and the criteria were as follows: (1) Kruskal–Wallis *p*-value < 0.05 ; (2) LDA (linear discriminant analysis) > 2 . LEfSe analysis was also used to identify the metabolomics biomarkers with the criterion: (1) Kruskal–Wallis *p*-value < 0.05 ; (2) LDA (linear discriminant analysis) > 1.5 . MaAsLin2 (MicrobiomeMultivariableAssociation with Linear Models, <https://huttenhower.sph.harvard.edu/maaslin/>) was used to adjust the effects of age and gender in taxonomic and metabolomics analysis with Maaslin2 package (v1.5.1) in R software (v4.0.0). Simple Spearman analysis was performed using `corr.test()` function from `psych` packages (v2.3.12) in R software, while partial spearman analysis was performed using `pcor.test()` function from `ppcor` packages (v1.1) in R software (v4.0.0). DESeq2 (R Package v1.30.1) was used

to perform the differential gene expression analysis based on the negative binomial distribution with the RNA-seq data in R software (v4.0.0), and gene expressions with *p*.*fdr* of < 0.05 and \log_2 Fold Change < -1 or > 1 were considered significant. The correlations between SNVs and taxa/metabolites were assessed using simple Spearman correlation analysis and the generalized linear model (GLM) analysis, respectively. First, simple Spearman correlation analysis was used to calculate the correlation with `corr.test()` function from `psych` packages (v2.3.12) in R (v4.0.0). Second, the generalized linear model (GLM) was used to calculate the correlation. The count abundance was submitted to logistic regression using GLM with a negative binomial distribution using the number of SNVs as a predictor, including the covariates age and sex. The `Manyglm()` function from the `mvabund` package (v 4.2.1) in R (v4.0.0) was used in this process, and *p* < 0.05 was considered significant.

Results

Association of gut microbiota with tumor location in colon cancer

To search for differences in the relative abundance of fecal gut microbiota among RCC, LCC, and healthy controls, we analyzed the gut microbiota profiles of 264 healthy participants and colon cancer patients in the 16S-seq cohort (Fig. 2A). The clinical characteristics of age, sex, differentiation, serum carcinoembryonic antigen (CEA), serum cancer antigen 19–9 (CA 19–9), tumor–lymph node–metastasis (TNM) stage, lymphatic invasion, nerve invasion, and vascular invasion were similar between the RCC and LCC groups (Table S1). Bacterial diversity was assessed by the Chao1, Shannon, and Simpson indexes, and the Chao1 index in the RCC and LCC groups was determined to be significantly reduced compared with the control, while there was no significant difference between the RCC and LCC groups in terms of the alpha diversity (Table S2). Beta diversity was calculated by using unweighted Unifrac measurements, and principal coordinate analysis was performed. There was a significantly separated distribution among the three groups as well as between the RCC and LCC groups (PERMANOVA, *p* < 0.05 , Fig. 2B). The differentially abundant genus signatures among the RCC, LCC, and control groups were assessed by the LEfSe algorithm, and a total of 34 genera among the three groups were identified as differentially abundant bacterial species with a linear discriminant analysis (LDA) score > 2.0 , and *p*-value < 0.05 (Fig. 2C, Table S3). An increased abundance of the cancer-promoting bacteria *Fusobacterium* was observed in RCC, while the genera *Parvimonas* and *Gemella* were identified as key microbiota in LCC. Furthermore, the genera *Faecalibacterium*, *Eubacterium*, and *Blautia* were

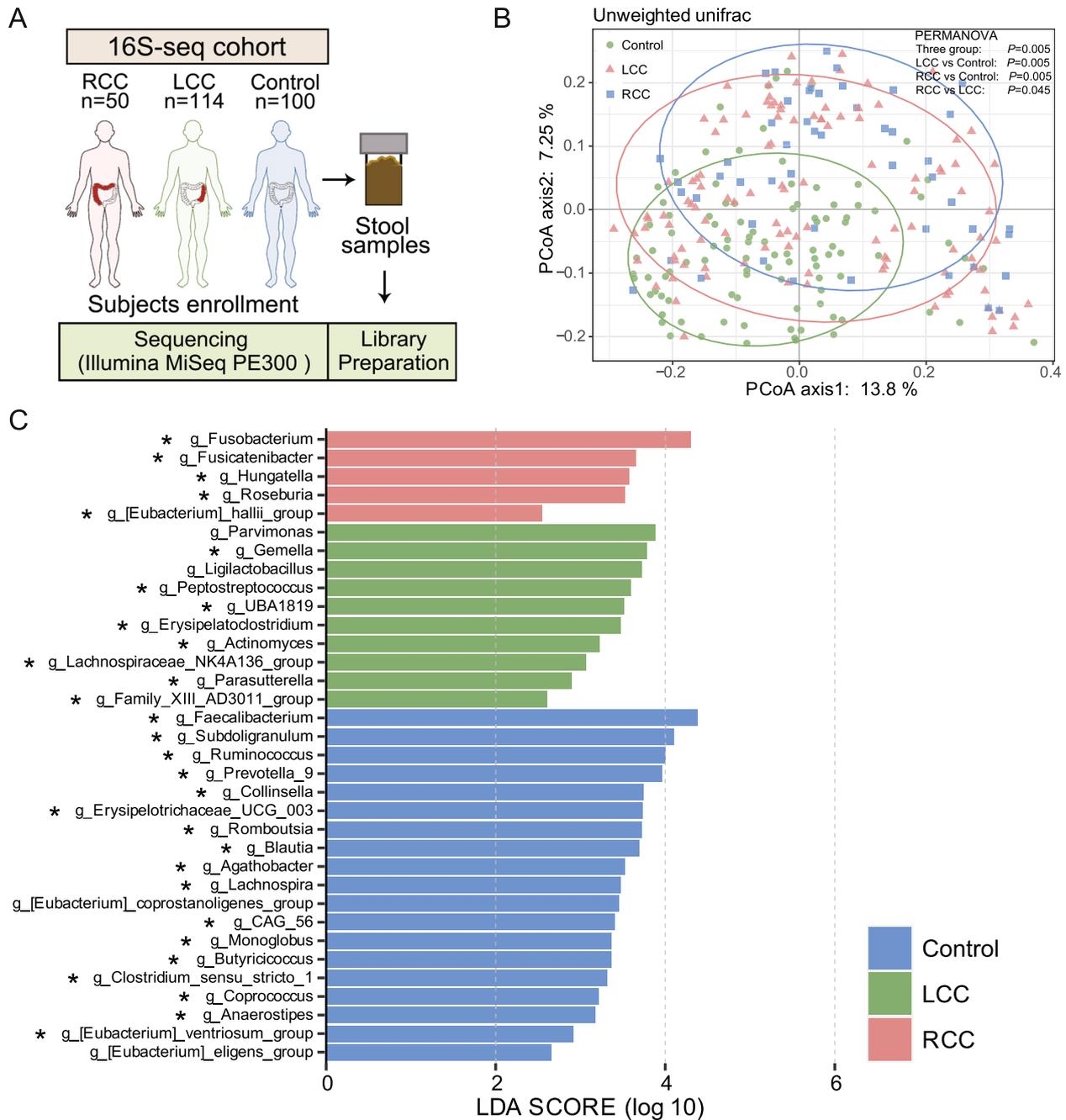


Fig. 2 Association of gut microbiota with tumor location in colon cancer. **A** Scheme of the clinical study on the 16S-seq cohort. To search for differences in the relative abundance of gut microbiota among fecal samples from healthy participants and patients with a primary tumor located in the right-sided colon (RCC) vs. left-sided colon (LCC), we globally analyzed the gut microbiota profiles of 264 individuals in the 16S-seq cohort, in which 16S rRNA sequencing gene data from 50 RCC, 114 LCC, and 100 control subjects were available. **B** Beta diversity was calculated using unweighted Unifrac measurements by principal coordinate analysis. **C** Bar plot of linear discriminant analysis coupled with effective size measurement based on the 16S rRNA gene sequencing among the RCC, LCC, and control groups. Asterisk "*" indicates the bacterial genera with distinct relative abundances between groups as detected by metagenomic sequencing after adjusting for the age and sex using MaAsLin2

the dominant taxa signature in the control group. Taken together, these results indicate significant associations between distinct gut microbiota and different tumor locations.

To further obtain deeper insights into the identification of gut microbiota associated with different tumor locations and to validate the results from the 16S-seq cohort, metagenomic sequencing with species-level taxonomic

resolution was performed with fecal samples from an additional 230 healthy participants and CRC patients in the Meta/LC-MS cohort (Fig. 3A). The clinical characteristics of age, sex, differentiation, TNM stage, serum CEA, serum CA19-9, lymphatic invasion, nerve invasion, and vascular invasion in the three groups were similar between the RCC and LCC groups (Table S4). Using the Chao1, Shannon, and Simpson indexes to assess alpha diversity, we found that there was no significant difference between RCC and LCC, which was consistent with the results observed in the 16S-seq cohort (Table S5). Furthermore, we employed the LEfSe algorithm to assess the differentially abundant species and genus signatures among the RCC, LCC, and control groups (Fig. 3B, LDA > 2 and p -value < 0.05, top 12 taxa in each group; Tables S6–S7). After correction for age and sex by MaAsLin2, the results indicated that the bacterial species *Bacteroides fragilis*, *Veillonella parvula*, *Flavonifractor plautii* (*Fp*), and *Fusobacterium nucleatum* were the key microbiota in RCC; whereas the species *Bacteroides* sp. A1C1 (*B.A1C1*), *Bacteroides ovatus*, *Bacteroides thetaiotaomicron*, *Porphyromonas asaccharolytica*, and *Parvimonas micra* were considered the dominant microbiota in LCC. Moreover, the species *Faecalibacterium prausnitzii*, *Eubacterium rectale*, *Anaerostipes hadrus*, and *Mesoplasma tabanidae* were identified as the key microbiota in the healthy controls. These findings indicate that unique profiles of the gut microbiome are present among the distinct tumor locations, which may potentially affect the clinicopathological characteristics of the tumor.

Tumor location-linked fecal metabolomic alterations and associations between the microbiota and metabolites

Considering the interplay between the gut microbiome and host metabolism, we first performed untargeted metabolomics on fecal samples from the Meta/LC-MS cohort (RCC, $n=63$; LCC, $n=79$; and controls, $n=88$). To explore the associations of each annotated metabolite with the tumor location, with comparisons among the RCC, LCC, and control groups, the LEfSe algorithm was used when the differentially accumulated metabolites were filtered by LDA score > 1.5, and p -value < 0.05

(Fig. 3C, Table S8). After correction for age and sex by MaAsLin2, 29 representative metabolites were found among the three groups, of which L-phenylalanine, N-acetyl-L-phenylalanine, and tryptophan were the key metabolites in RCC; whereas L-citrulline, D-ornithine, eicosadienoic acid, and L-tyrosine were considered the dominant metabolites in LCC. Specifically, lipid metabolites, including phosphoglyceride (PG) 15:0, PG (6:0/9:0), phosphatidic acid (PA) 21:0, and PA (2:0/19:0), as well as the short-chain fatty acids isonicotinic acid and isobutyric acid were significantly increased in the control group. Subsequently, we performed correlation analysis to examine the associations between the differentially abundant species and metabolites in the multi-omics cohort (simple Spearman correlation, $p < 0.05$; Fig. 3D). Also, the partial Spearman method was used to adjust for age, sex, tumor location, and whether the subject suffered from CRC (Table S9). In general, we observed strong positive associations between taxa and metabolites that were both elevated in the LCC and control groups as well as negative associations between the RCC- and control-enriched taxa and LCC-decreased metabolites. Notably, L-phenylalanine was positively associated with certain species, such as *Fp* and *B. fragilis*, which were enriched in RCC but decreased in LCC. In addition, the LCC-accumulated metabolites L-citrulline and D-ornithine were positively associated with *B.A1C1*, which was enriched in LCC. Altogether, our results suggest that patients with RCC vs. LCC have unique microbial metabolic processes.

Associations of host genomics with the tumor location in colon cancer

To further explore the potential mechanisms involved in RCC and LCC by gut microbiota and their metabolites, we analyzed the exome and transcriptome data from CRC tumors with different primary tumor locations in the WES/RNA-seq cohort ($n=59$) (Fig. 4A). The 59 tumor samples from the WES/RNA-seq cohort were all obtained from the colon cancer patients in the Meta/LC-MS cohort. Next, we analyzed the WES data to obtain the basic information of the mutated genes in LCC and RCC. The results showed that missense

(See figure on next page.)

Fig. 3 Tumor location-linked fecal metabolomic alterations and associations between the microbiota and metabolites. **A** Scheme of the clinical study on the Meta/LC-MS cohort. Metagenomic sequencing with species-level taxonomic resolution was performed with fecal samples from patients with right-sided colon cancer (RCC, $n=63$), left-sided colon cancer (LCC, $n=79$), and healthy controls (control, $n=88$) in the Meta/LC-MS cohort. **B** Bar plot of linear discriminant analysis coupled with effective size measurement based on the metagenomic sequencing among the RCC, LCC, and control groups (top 12 taxa in each group). **C** Bar plot of linear discriminant analysis coupled with effective size measurement based on the differential metabolites among the RCC, LCC, and control groups (filtered by linear discriminant analysis (LDA) > 1.5, and p -value < 0.05). Asterisk *** in panels **B** and **C** indicates bacterial species and metabolites with distinct relative abundances and levels between groups detected by metagenomic sequencing after adjusting for the age and sex using MaAsLin2. **D** The heatmap depicts the relationships between the taxa and metabolites that changed among groups (simple Spearman analysis, asterisk *** indicates p -value < 0.05)

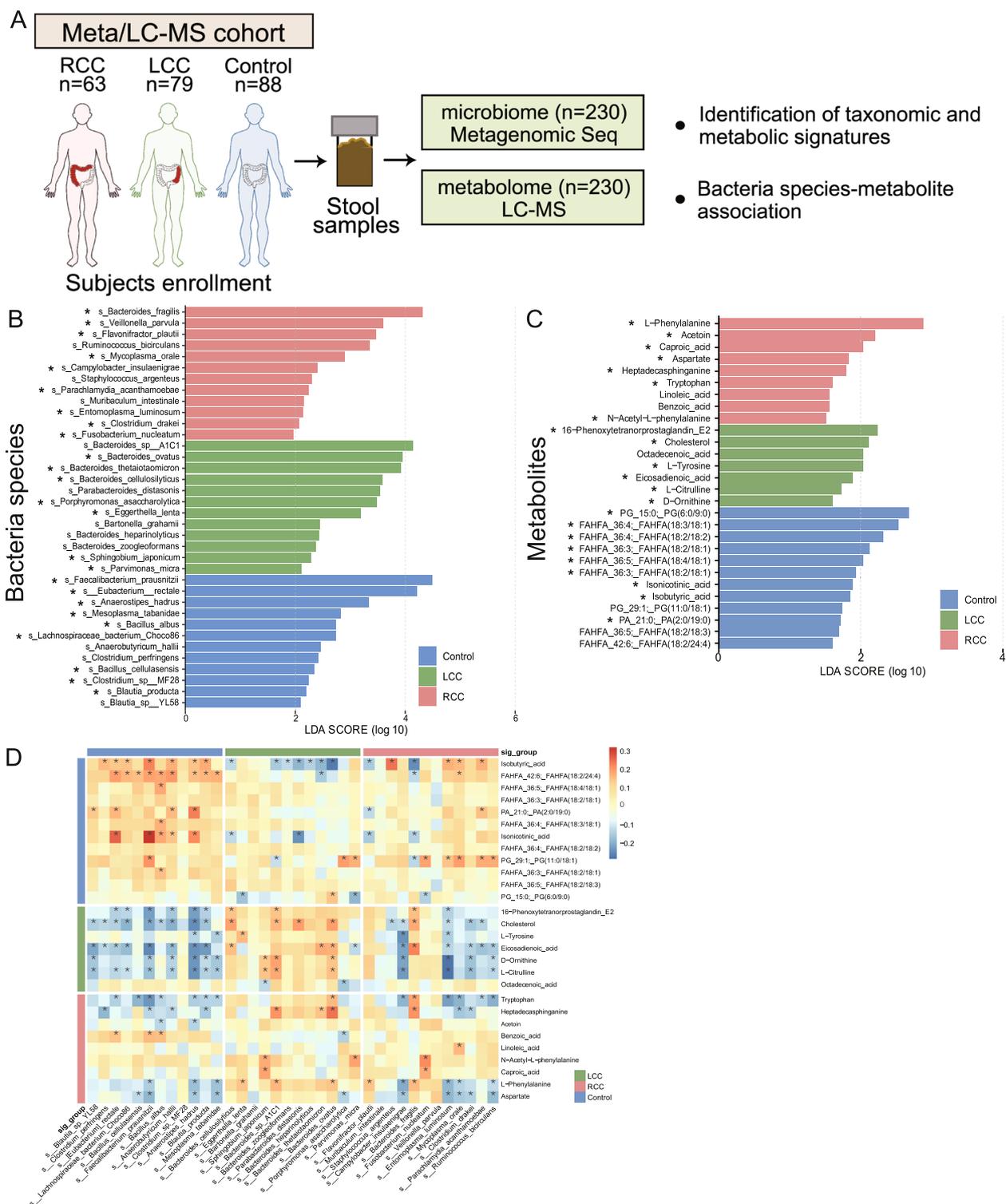


Fig. 3 (See legend on previous page.)

mutations and single nucleotide polymorphisms (SNPs) were the most common variant types in both LCC and RCC, and the main base substitution mutation

frequencies were C > T and C > A (Fig. S1A–C for LCC and Fig. S2A–C for RCC). The variants per sample and boxplots of variant classification are shown in Fig.

S1D–E and Fig. S2D–E. Notably, the *TTN* and *MUC16* genes were prominent among the top 10 mutated genes in colon cancer (Fig. S1F for LCC and Fig. S2F for RCC), and the waterfall plot indicates that LCC and RCC have differential host genotype mutations (Fig. 4B). Moreover, somatic interaction analysis among the mutated genes showed a strong co-occurrence of mutational patterns in LCC and RCC, respectively (Fig. 4C for LCC and Fig. 4D for RCC, the pairwise Fisher's exact test was performed to detect mutually exclusive or co-occurring events, $p < 0.05$ and $p < 0.1$ are marked).

To obtain the maximum gene sequence variation, new transcript information, and the difference of gene expression profiles between RCC and LCC, we further analyzed the transcriptome sequencing results in the WES/RNA-seq cohort. Principal coordinate analysis calculated by Bray–Curtis measurements was performed to determine the sample distance between the gene expression profiles of RCC and LCC, and the results showed that there was a significantly separated distribution between the two groups (PERMANOVA, $p = 0.0218$, Fig. 5A). According to the gene expression analysis, using a cutoff of at least a two-fold difference and $p < 0.05$, 362 genes were found to be differentially expressed between RCC and LCC (Fig. 5B, Table S10, p -value was calculated by DESeq2, filtered by $\log_2(\text{foldchange}) > 1$ or < -1 ; FDR < 0.05). The genes *PHLDA1*, *VWF*, *IFITM3*, *Wbp1*, *GPR98*, *KLRC2*, *STAC*, and *SLIT2* were found to be enriched in the RCC group; while the genes *AC079354.1*, *KRTAP10-2*, *TCF25*, and *HLA-DRB5* were enriched in the LCC group. To identify the key genes involved in the pathogenesis of RCC and LCC, we screened for mutations in the WES data that showed significant expression differences in the transcriptomic data. With SNV changes in more than two samples and amino acid metabolic changes as screening conditions, we identified 3 and 10 key gene mutations in LCC and RCC, respectively (Fig. 5C–D). In RCC, the genes *GPR98*, *KLRC2*, *STAC*, and *SLIT2* may serve as potential key genes whose expression changed together. Meanwhile, the gene *AC079354.1* and its transcript expression changed together in LCC, which may play a key role in LCC pathogenesis and progression. Taken together, RCC and LCC have unique host gene and expression profiles, and key genes may be involved in the regulation of disease development.

Multi-omics interaction analysis between host genomics and metagenomics in RCC and LCC

The interaction between the gut microbiota and host genes has always been a frontier in the field of human microbiology. To clarify the association of host genes with taxa, metabolites, and transcripts, we first used PERMANOVA analysis to evaluate the driving effect of SNVs on species, metabolites, and transcriptome expression, respectively, based on the 59 samples that all underwent metagenomic sequencing, metabolomics analysis, WES, and transcriptome sequencing (PERMANOVA analysis, $p < 0.05$, Fig. 6A, Fig. S3A–B, Tables S11–S13). Among them, the RCC-enriched genes *IRS4* and *LARPI* and the LCC-enriched genes *TP53* and *PHLDA1* may have had a dominant driving effect on the species; while the LCC-enriched genes *ZNF267* and *IGHG3* may have had a lead driving effect on metabolites. Next, we calculated the correlation between SNVs and differential species and metabolites, respectively. As for the correlation between SNVs and species, simple Spearman's correlation analysis was firstly performed to analyze the correlations (p -value < 0.05), and the correlations were adjusted by microsatellite instability (MSI) status using partial Spearman analysis (p -value < 0.05 , Fig. 6B, Table S14). To better assess the correlations, the generalized linear model (GLM) with negative binomial distribution was used to adjust for age and sex was used secondly (p -value < 0.05 , Fig. 6B, Table S14). As for the correlation between SNVs and metabolites, simple Spearman's correlation analysis (p -value < 0.05) and GLM analysis (p -value < 0.05) were also performed respectively (Fig. S3C, Table S15). The results showed that RCC-enriched *Clostridium drakei* and *Fp* were significantly negatively correlated with the genes *PHLDA1* and *WBPI*, respectively. In addition, the *E. nucleatum* species enriched in RCC showed a positive correlation with the gene *VWF* and a negative correlation with the gene *OR2T2*. In LCC, the key species *B.A1C1* showed positive correlations with the genes *TCF25*, *HLA-DRB5*, and *AC079354.1* as well as a negative correlation with the gene *KRTAP10-2*. In determining the relationship between SNVs and metabolites, RCC-enriched L-phenylalanine and tryptophan were found to be significantly negatively correlated with the gene *HOXB7*. The analysis associated the presence/absence of an allele in a certain gene with microbial abundances.

(See figure on next page.)

Fig. 4 Associations of the whole exome with the tumor location in left-sided colon cancer (LCC) and right-sided colon cancer (RCC). **A** Scheme of the clinical study on the WES/RNA-seq cohort. The exome and transcriptome data from CRC tumors with different primary tumor locations were analyzed to explore the potential mechanisms involved in LCC and RCC by gut microbiota and their metabolites. The 59 tumor samples from the WES/RNA-seq cohort were all obtained from the colon cancer patients in the Meta/LC–MS cohort. **B** The waterfall plot of host genotype mutations in LCC and RCC. **C–D** Somatic interaction analysis among mutated genes in LCC (**C**) and RCC (**D**). (The pairwise Fisher's exact test was performed to detect mutually exclusive or co-occurring events; the symbols * and · indicate $p < 0.05$ and $p < 0.1$, respectively)

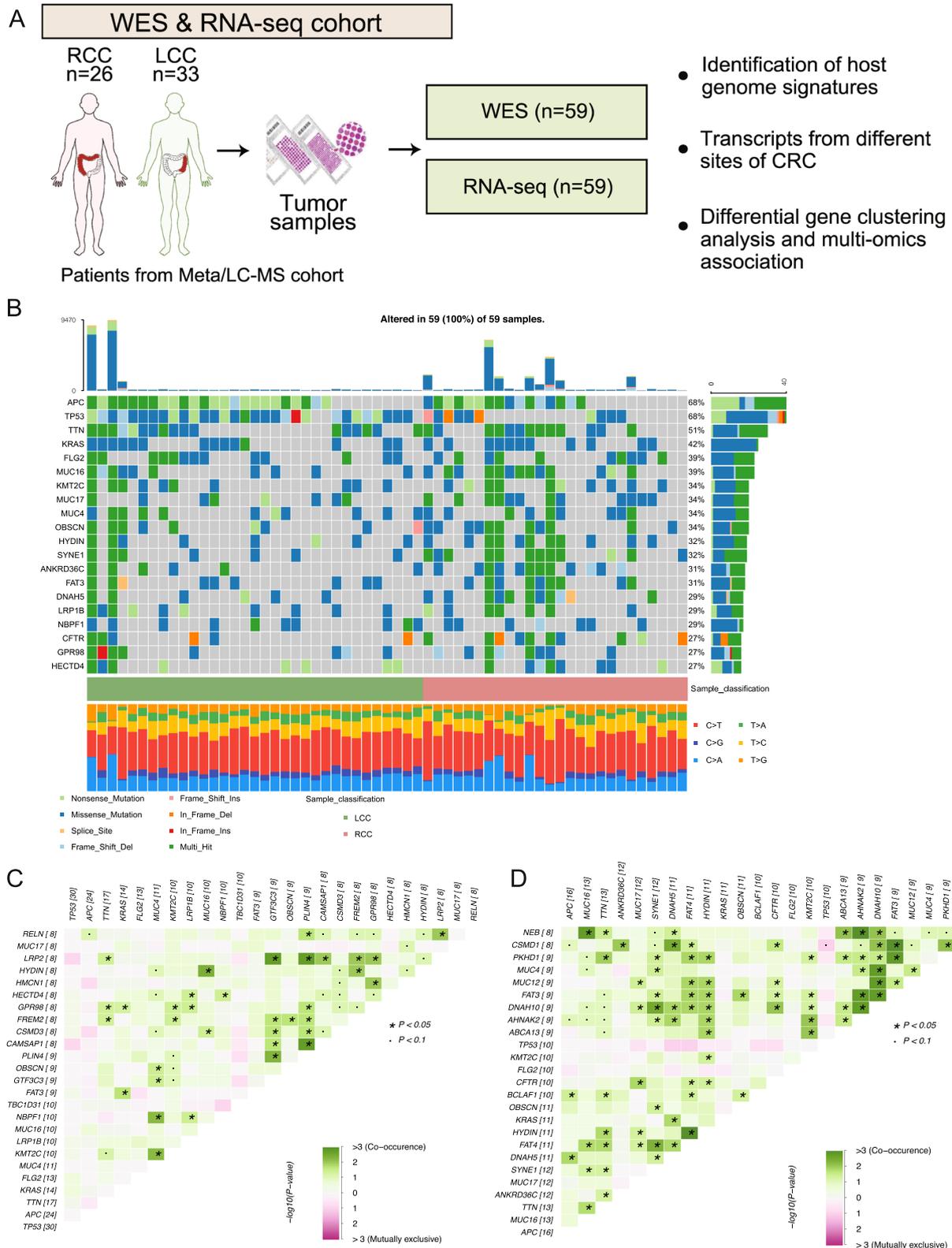


Fig. 4 (See legend on previous page.)

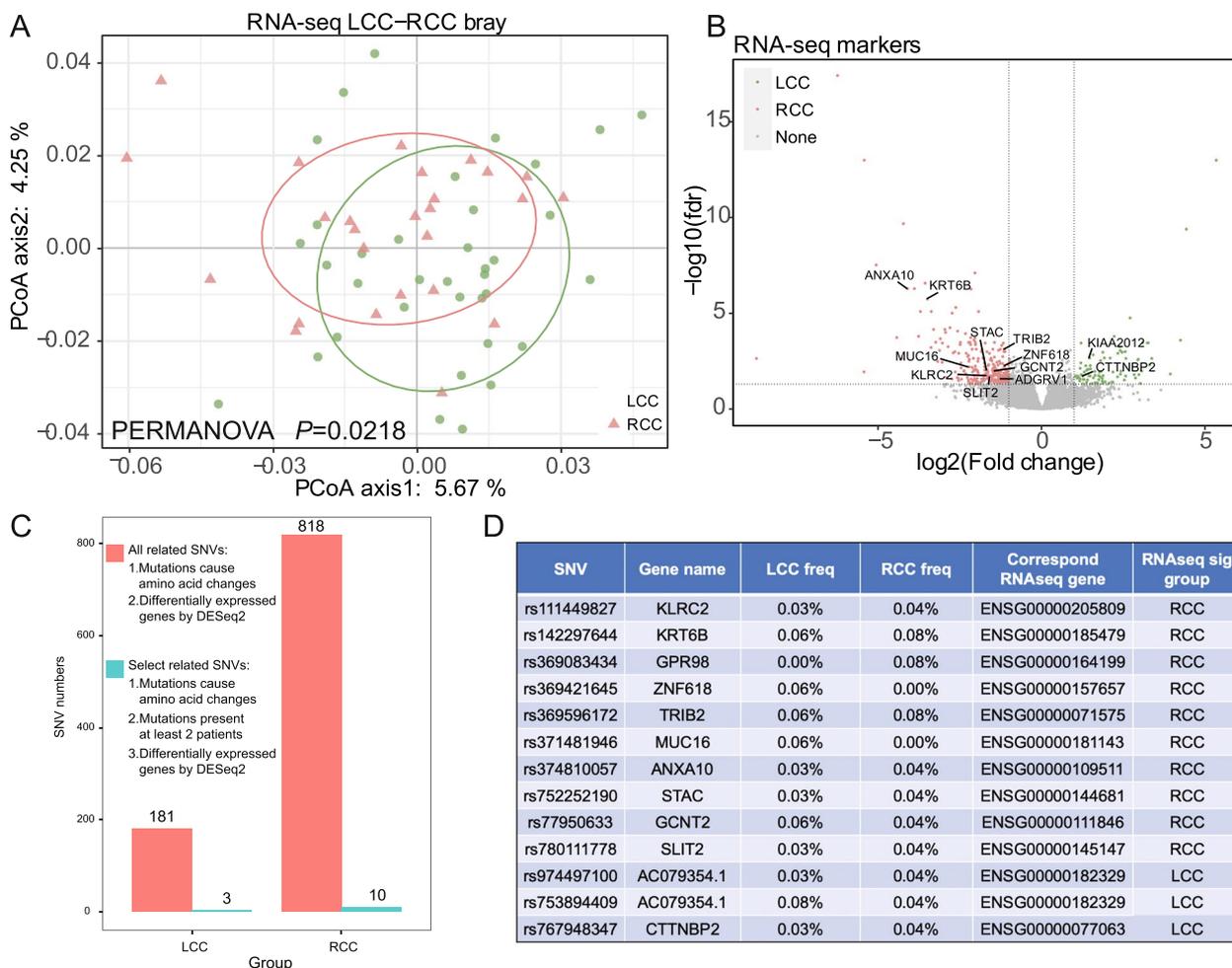


Fig. 5 Associations of transcriptome with tumor location in colon cancer. **A** Principal coordinate analysis calculated by Bray–Curtis measurements was performed to determine the sample distance between the gene expression profiles of left-sided colon cancer (LCC) and right-sided colon cancer (RCC) (Bray–Curtis measurements, PERMANOVA, $p=0.0218$). **B** The volcano plot showed the differentially expressed genes between LCC and RCC using a cutoff of at least a two-fold difference and $p < 0.05$ (FDR value was calculated by DESeq2). **C–D** Genes in whole-exome sequencing whose expression profiles in transcriptomic sequencing changed together were screened to identify the key genes involved in the pathogenesis of LCC and RCC. With single-nucleotide variant (SNV) changes in more than two samples and amino acid metabolic changes as screening conditions, 3 and 10 key changed genes in RCC and LCC were identified, respectively

Besides focusing on distinguishing RCC from LCC, we hypothesized that a combined analysis of RCC and LCC samples compared against controls across omics layers would be interesting to provide context. We separately included the analysis results for RCC compared to the control and LCC compared to the control. These results are now available in Tables S16–S21. (The criteria for these results were $LDA > 2$ and $p < 0.05$). Combining the associations between bacteria and metabolites as well as between bacteria and genes, we propose that two key signaling axes may represent the

multi-omics profiles of RCC and LCC, respectively, and may play a key role in guiding pathogenesis research and disease therapy. In RCC, the enrichment of *Fp* may be related to the accumulation of the metabolite L-phenylalanine, which can potentially affect the *WBPI* signal. Moreover, LCC-enriched *B.AIC1* may be associated with the accumulation of the metabolites D-ornithine and L-citrulline and activation of the genes *TCF25*, *HLA-DRB5*, and *AC079354.1*, which are involved in colorectal carcinogenesis (summarized in Fig. 1).

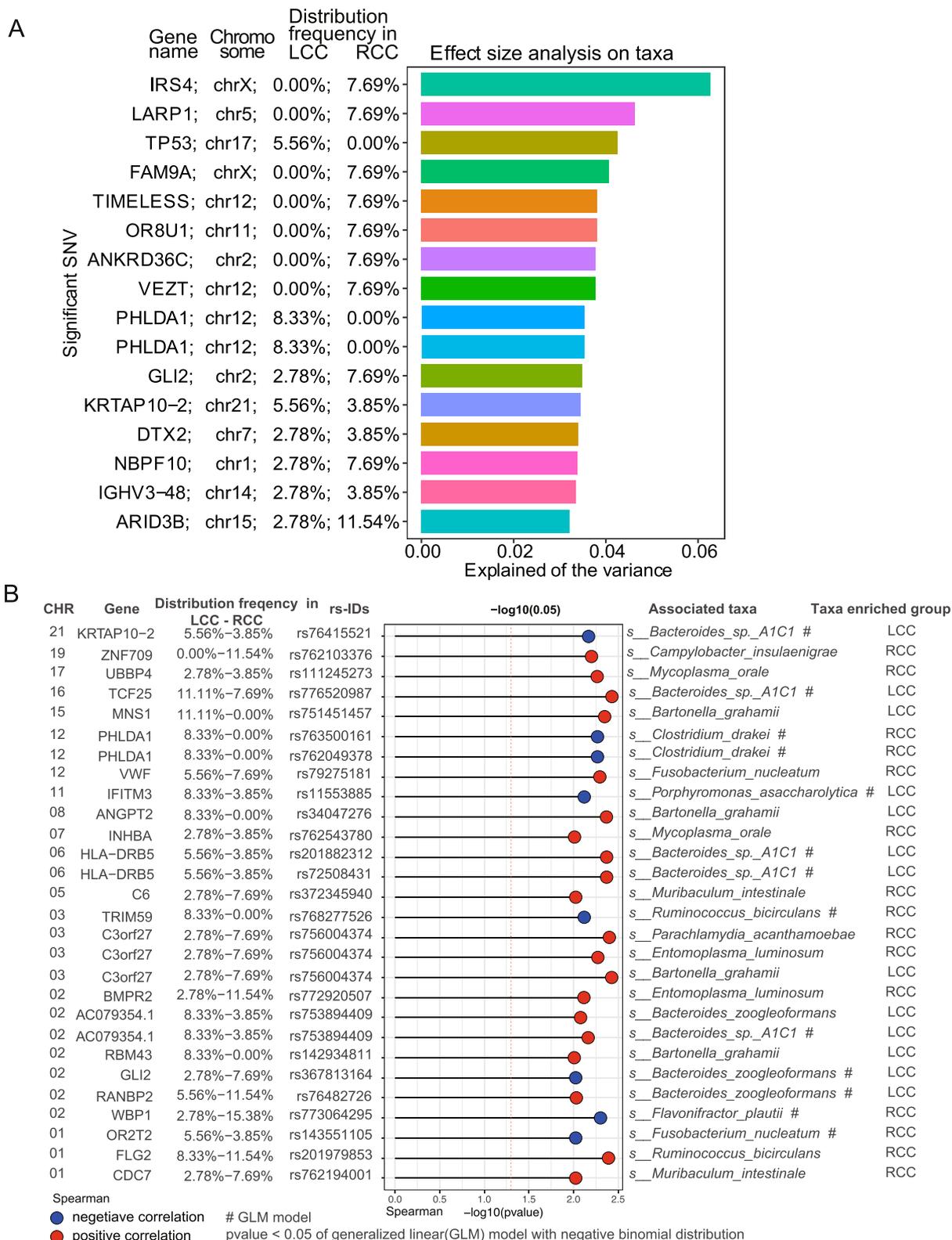


Fig. 6 Multi-omics interaction analysis between host genomics and metagenomics in left-sided colon cancer (LCC) and right-sided colon cancer (RCC). **A** PERMANOVA analysis was used to evaluate the driving effect of single-nucleotide variants (SNVs) on species based on the 59 samples that all underwent metagenomic sequencing, metabolomics analysis, whole-exome sequencing, and transcriptome sequencing ($p < 0.05$). **B** The correlation between filtered SNVs and differential species. (Spearman’s correlation analysis was performed to analyze the correlations; $p < 0.05$)

Discussion

In the present study, we combined multi-omics sequencing to profile the gut microbiome, metabolome, and host genome in RCC and LCC, revealing new metagenome–genome-associated interactions. Our approach identified that the RCC-associated species *Fp* with increased L-phenylalanine shaped the gut ecology and correlated with the mutation and dysregulation profiles of the host gene *WBP1*. We also identified LCC-associated features, including the species *B.A1C1* and *B. thetaiotaomicron*, the metabolites D-ornithine and L-citrulline, and the genes *TCF25*, *HLA*, and *KRTAP10-2*. Furthermore, we summarized the microbiota–host interaction networks as the *Fp*–phenylalanine–*WBP1* axis in RCC patients and the *B.A1C1*–D-ornithine/L-citrulline–*TCF25/HLA/KRTAP10-2* axis in LCC patients, potentially suggesting the distinct clinicopathological characteristics of RCC and LCC.

The 16S rRNA and metagenomic sequencing showed that the key microbiome of RCC consisted of the genus *Fusobacterium* and the species *B. fragilis*, *V. parvula*, *Fp*, and *F. nucleatum*, while that of LCC consisted of the genera *Parvimonas* and *Gemella* and the species *B. ovatus*, *B. thetaiotaomicron*, *P. asaccharolytica*, and *P. micra*; these findings are partially consistent with the published literature [14, 15, 28]. *Fusobacterium* is a common oral bacteria that form biofilms on teeth and epithelial tissues [29]. Fusobacteria-dominant polymicrobial biofilms have been detected in RCC to contain *Bacteroidetes*, *Lachnospiraceae*, and *Enterobacteriaceae* [14, 15]. Moreover, *F. nucleatum* is a type of oncogenic bacteria that is significantly elevated in tumor tissues, which increases the mutation frequency of CRC and affects CRC proliferation, metastasis, and chemoresistance [30–39]. The amount of *F. nucleatum* DNA in CRC tissue has been associated with a shorter survival time and an MSI-high status [16]. Furthermore, metagenomic and transcriptomic analyses have revealed that intestinal microbes, especially *F. nucleatum*, are involved in CRC development, consistent with our results [40]. We also identified that *F. nucleatum* was positively associated with the von Willebrand factor (vWF) polymorphism. vWF is reported as a plasma protein marker for the early detection of CRC, indicating the early alteration of vWF in CRC initiation [41]. The deposition of vWF and blood coagulation may drive tumor metastasis [42]. Our results suggest that the vWF mutation associated with *F. nucleatum* may have an impact on platelet adhesion, potentially contributing to chronic bleeding in the tumor and dysbiosis of the intestinal microenvironment. This finding may point to a possible mechanism of carcinogenesis. Nevertheless, further research is required to establish a causal relationship.

We are the first to report the RCC-associated axis of *Fp*–phenylalanine–*WBP1*. We have reported previously that *Fp* is a flavonoid-degrading species that is enriched in young-onset CRC [43, 44]. *WBP1*, which encodes a ligand of the WW domain of the Yes-associated protein (YAP), has been demonstrated to have an antitumor role via binding to YAP and reducing the transcriptional activity of YAP [45, 46]. *WBP1* also has been reported to be essential for oligosaccharide transfer and N-glycosylation in yeast. In addition, depleted *WBP1* may reduce the intake of oligosaccharides and the level of N-glycosylation [47]. Our data suggest that enrichment of *Fp* in RCC may be related to the mutation of *WBP1*. We also found *Fp* to be significantly negatively correlated with *PHLDA1*, a novel identified p53 target acting as a tumor-suppressor gene that is essential for repressing Akt involved in the PI3K signaling pathway [48]. These findings bridge the links among diet-associated flavonoids, microbiota, and the host genome.

The LCC-associated metagenome–genome interaction comprised the key species *B.A1C1* and various SNVs, including in *TCF25*, *HLA-DRB5*, and *KRTAP10-2*. *TCF25*, which acts as a transcriptional repressor in cell death control, has been proven to repress the transcription of the serum response factor (SRF). Overexpression of *TCF25* may aggravate intestinal ulceration by depleting re-epithelialization and regeneration of submucosal smooth muscular cells via suppressing SRF [49]. Additionally, the dysregulation of the circular RNA *TCF25* has been reported in bladder carcinoma [50]. Meanwhile, *HLA-DRB5* polymorphisms are associated with various diseases such as type 1 diabetes and multiple sclerosis [51]. However, the relationship between *HLA-DRB5* polymorphisms and gut microbiota and its role in carcinogenesis have not been elucidated. We observed a correlation between *B.A1C1* and the mutation of the host genome in LCC, which may be associated with promoting tumorigenesis. Our multi-omics research provides correlational evidence, serving as supportive findings for other research results, thus contributing to the consensus in the literature. Nevertheless, further basic experiments should be conducted to validate these correlation findings in the future and to enhance our understanding of the underlying mechanisms.

Conclusion

We performed a comprehensive multi-omics analysis of the microbiome, metabolome, and host genome in patients with RCC vs. LCC to explain the site-specific difference in CRC, revealing many new metagenome–genome interactions. In this regard, these data contribute to a growing body of literature to understand the differences between different locations of CRC from an

integrated multi-omics perspective, which is the key to exploring the differences between RCC and LCC in tumor initiation and progression. Since metagenome–genome association studies are an emerging field to elucidate the role of gut microbiota in human physiological processes and pathogenesis, the causal relationship between the gut microbiota and the host genome requires further investigation [52]. Our findings extend our insights into the relationship between the gut microbiota at the species level, metabolites, and host genomics in RCC and LCC, pointing to possible future modalities for CRC intervention targeting the gut microbiota–host interplay.

Abbreviations

RCC	Right-sided colon cancer
LCC	Left-sided colon cancer
WES	Whole-exome sequencing
RNA-seq	Transcriptome sequencing
<i>Fp</i>	<i>Flavonifractor plautii</i>
<i>B.A1C1</i>	<i>Bacteroides</i> sp. A1C1
LEFSe	Linear discriminant analysis effect size
MaAsLin2	Multivariate association with linear models
MSI	Microsatellite instability
SNP	Single nucleotide polymorphism
SNV	Single-nucleotide variant
YAP	Yes-associated protein

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-024-01987-7>.

Supplementary Material 1: Figure S1. Basic information of mutated genes from the whole-exome sequencing data in left-sided colon cancer (LCC) (related to Figure 4). (A) Bar chart showing the frequency distribution of the different mutation types, categorized into types based on their effect on protein coding, including missense mutations. (B) Bar chart displaying the frequency distribution of the different mutation categories, divided into single nucleotide polymorphism (SNP), insertion (INS), and deletion (DEL) categories. (C) Bar chart illustrating the frequency distribution of the single nucleotide variation (SNV) mutation types, considering combinations of mutations and strand orientation, categorized into six types. (D) Stacked bar chart showing the distribution of the different mutation types in each sample, corresponding to the types mentioned in panel A. (E) Box plot depicting the distribution of the different mutation types in each sample, corresponding to the types in panel A. (F) Stacked bar chart displaying the different mutation types for the top 10 genes with the highest mutation frequencies, corresponding to the types shown in panel A.

Supplementary Material 2: Figure S2. Basic information of mutated genes from the whole-exome sequencing data in right-sided colon cancer (RCC) (related to Figure 4). (A–C) The variant classification (A), variant type (B), and single nucleotide variant (SNV) class of the main base substitution mutation frequency (C) in RCC based on the whole-exome sequencing data. (D–E) The variants per sample (D) and box and line diagram of variant classification (E) in RCC. (F) The top 10 mutated genes in RCC.

Supplementary Material 3: Figure S3. Multi-omics interaction analysis between host genomics and metagenomics in left-sided colon cancer (LCC) and right-sided colon cancer (RCC) (related to Figure 6). (A–B) PERMANOVA analysis was used to evaluate the driving effect of single nucleotide variants (SNVs) on metabolites (A) and transcriptome expression (B) based on the 59 samples that all underwent metagenomic sequencing, metabolomics analysis, whole-exome sequencing, and transcriptome sequencing ($p < 0.05$). (C) The correlation between filtered SNVs and differential metabolites. (Spearman's correlation analysis was performed to analyze the correlations. $p < 0.05$).

Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.
Supplementary Material 9.
Supplementary Material 10.
Supplementary Material 11.
Supplementary Material 12.
Supplementary Material 13.
Supplementary Material 14.
Supplementary Material 15.
Supplementary Material 16.
Supplementary Material 17.
Supplementary Material 18.
Supplementary Material 19.
Supplementary Material 20.
Supplementary Material 21.
Supplementary Material 22.
Supplementary Material 23.
Supplementary Material 24.

Acknowledgements

We thank all donors and colleagues of the Fudan University Shanghai Cancer Center for their participation and cooperation.

Authors' contributions

L.L., C.K., G.L., J.W., and Y.M. designed the experiments. L.L., C.K., J.W., G.W., X.L., D.S., and Y.Y. provided the clinical samples and performed the experiments. L.L., C.K., G.L., J.W., and J.L. analyzed the data. L.L., C.K., J.W., J.L., and Y.M. wrote the manuscript. All authors edited the manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (81920108026, 82372594, and 82303276), Shanghai Science and Technology Development Fund (No.23410710500), Shanghai Pujiang Program (No. 22PJJD013).

Data availability

The 16S rRNA gene sequences that support the findings of this study were provided and are publicly available at the NIH National Center for Biotechnology Information Sequence Read Archive with BioProject ID PRJNA763023. The raw data for multi-omics cohort are deposited in the National Genomics Data Center website (<https://ngdc.cncb.ac.cn/>) under bioproject PRJCA017288 (<https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA017288>). The metagenomic data are accessible with the identifier HRA005038. The WES data are accessible with the identifier HRA005847. The RNA-seq data are accessible with the identifier HRA005810. The metabolomics data are accessible with the identifier OMIX004606 and OMIX006518. The metabolomics data are also deposited to the EMBL-EBI MetaboLights database with the identifier MTBLS10232 (www.ebi.ac.uk/metabolights/MTBLS10232).

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the Institutional Review Board of Fudan University Shanghai Cancer Center (ID: 050432–4–1911D), and written informed consent was provided by all subjects before sampling.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Colorectal Surgery, Fudan University Shanghai Cancer Center, Shanghai, China. ²Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China. ³Guangdong Hongyuan Pukang Medical Technology Co., Ltd., Guangdong, China. ⁴GenomiCare Biotechnology Co. Ltd., Shanghai, China.

Received: 20 January 2023 Accepted: 22 November 2024

Published online: 28 December 2024

References

- Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet* (London, England). 2019;394:1467–80.
- Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66:683–91.
- Iacopetta B. Are there two sides to colorectal cancer? *Int J Cancer*. 2002;101:403–8.
- Wexner SD, Gervaz P, Bouzourene H, Cerottini JP, Chaubert P, Benhattar J, et al. Dukes B colorectal cancer - distinct genetic categories and clinical outcome based on proximal or distal tumor location - Reply. *Dis Colon Rectum*. 2001;44:373–373.
- Gonzalez EC, Roetzheim RG, Ferrante JM, Campbell R. Predictors of proximal vs. distal colorectal cancers. *Dis Colon Rectum*. 2001;44:251–8.
- Distler P, Holt PR. Are right- and left-sided colon neoplasms distinct tumors? *Dig Dis*. 1997;15:302–11.
- Bufill JA. Colorectal-cancer - evidence for distinct genetic categories based on proximal or distal tumor location. *Ann Intern Med*. 1990;113:779–88.
- Delattre O, Law DJ, Remvikos Y, Sastre X, Feinberg AP, Olschwang S, et al. Multiple genetic alterations in distal and proximal colorectal-cancer. *Lancet*. 1989;2:353–6.
- Gao XH, Yu GY, Gong HF, Liu LJ, Xu Y, Hao LQ, et al. Differences of protein expression profiles, KRAS and BRAF mutation, and prognosis in right-sided colon, left-sided colon and rectal cancer. *Sci Rep*. 2017;7(1):7882.
- Huang C-W, Tsai H-L, Huang M-Y, Huang C-M, Yeh Y-S, Ma C-J, et al. Different clinicopathologic features and favorable outcomes of patients with stage III left-sided colon cancer. *World J Surg Oncol*. 2015;13:257.
- Benedix F, Kube R, Meyer F, Schmidt U, Gasting I, Lippert H, et al. Comparison of 17,641 patients with right- and left-sided colon cancer: differences in epidemiology, perioperative course, histology, and survival. *Dis Colon Rectum*. 2010;53:57–64.
- Bustamante-Lopez LA, Nahas SC, Nahas CSR, Pinto RA, Marques CFS, Cecconello I. Is there a difference between right- versus left- sided colon cancers? Does side make any difference in long-term follow-up? *Abcd-Arquivos Brasileiros De Cirurgia Digestiva-Brazilian Archives of Digestive Surgery*. 2019;32(4):e1479.
- Imperial R, Ahmed Z, Toor OM, Erdoğlan C, Khaliq A, Case P, et al. Comparative proteogenomic analysis of right-sided colon cancer, left-sided colon cancer and rectal cancer reveals distinct mutational profiles. *Mol Cancer*. 2018;17(1):177.
- Dejea CM, Wick EC, Hechenbleikner EM, White JR, Mark Welch JL, Rossetti BJ, et al. Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc Natl Acad Sci*. 2014;111:18321–6.
- Zhong MY, Xiong YB, Ye ZJ, Zhao JB, Zhong LF, Liu Y, et al. Microbial community profiling distinguishes left-sided and right-sided colon cancer. *Front Cell Infect Microbiol*. 2020;10:498502.
- 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;65:1973–80.
- Lee MS, Menter DG, Kopetz S. Right versus left colon cancer biology: integrating the consensus molecular subtypes. *J Natl Compr Canc Netw*. 2017;15:411–9.
- Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37:852–7.
- Liu G, Li T, Zhu X, Zhang X, Wang J. An independent evaluation in a CRC patient cohort of microbiome 16S rRNA sequence analysis methods: OTU clustering, DADA2, and Deblur. *Front Microbiol*. 2023;14:1178744.
- Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. *Nat Protoc*. 2022;17:2815–39.
- Thompson RS, Vargas F, Dorrestein PC, Chichlowski M, Berg BM, Fleshner M. Dietary prebiotics alter novel microbial dependent fecal metabolites that improve sleep. *Sci Rep*. 2020;10:3848.
- Xiang S, Ye K, Li M, Ying J, Wang H, Han J, et al. Xylitol enhances synthesis of propionate in the colon via cross-feeding of gut microbiota. *Microbiome*. 2021;9:62.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol*. 2016;17:122.
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28:1747–56.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–50.
- Rühlemann MC, Hermes BM, Bang C, Doms S, Moitinho-Silva L, Thingholm LB, et al. Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat Genet*. 2021;53:147–55.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Phipps O, Quraishi MN, Dickson EA, Steed H, Kumar A, Acheson AG, et al. Differences in the on- and off-tumor microbiota between right- and left-sided colorectal cancer. *Microorganisms*. 2021;9(5):1108.
- Kolenbrander PE, Palmer RJ Jr, Periasamy S, Jakubovics NS. Oral multispecies biofilm development and the key role of cell-cell distance. *Nat Rev Microbiol*. 2010;8:471–80.
- Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med*. 2019;25:968–76.
- Brennan CA, Garrett WS. Fusobacterium nucleatum — symbiont, opportunist and oncobacterium. *Nat Rev Microbiol*. 2019;17:156–66.
- 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;170:548–563.e16.
- Yang Y, Weng W, Peng J, Hong L, Yang L, Toiyama 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- κ B, and up-regulating expression of MicroRNA-21. *Gastroenterology*. 2017;152:851–866.e24.
- Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of fusobacterium persistence and antibiotic response in colorectal cancer. *Science*. 2017;358:1443–8.
- Liu G, Su L, Kong C, Huang L, Zhu X, Zhang X, Ma Y, Wang J. Improved diagnostic efficiency of CRC subgroups revealed using machine learning based on intestinal microbes. *BMC Gastroenterol*. 2024;24(1):315.
- Serna G, Ruiz-Pace F, Hernandez J, Alonso L, Fasani R, Landolfi S, et al. Fusobacterium nucleatum persistence and risk of recurrence after preoperative treatment in locally advanced rectal cancer. *Ann Oncol*. 2020;31:1366–75.
- Hong J, Guo F, Lu SY, Shen C, Ma D, Zhang X, et al. F. nucleatum targets lncRNA ENO1-IT1 to promote glycolysis and oncogenesis in colorectal cancer. *Gut*. 2021;70(11):2123–37.
- Chen Y, Chen Y, Zhang J, Cao P, Su W, Deng Y, et al. Fusobacterium nucleatum promotes metastasis in colorectal cancer by activating autophagy signaling via the upregulation of CARD3 expression. *Theranostics*. 2020;10:323–39.
- Chen S, Su T, Zhang Y, Lee A, He J, Ge Q, et al. Fusobacterium nucleatum promotes colorectal cancer metastasis by modulating KRT7-AS/KRT7. *Gut Microbes*. 2020;11:511–25.

40. 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.
41. Rho J-H, Ladd JJ, Li CI, Potter JD, Zhang Y, Shelley D, et al. Protein and glycomic plasma markers for early detection of adenoma and colon cancer. *Gut.* 2018;67:473–84.
42. Feinauer MJ, Schneider SW, Berghoff AS, Robador JR, Tehranian C, Karremann MA, et al. Local blood coagulation drives cancer cell arrest and brain metastasis in a mouse model. *Blood.* 2021;137:1219–32.
43. Yang Y, Du L, Shi D, Kong C, Liu J, Liu G, et al. Dysbiosis of human gut microbiome in young-onset colorectal cancer. *Nat Commun.* 2021;12:6757.
44. Kong C, Liang L, Liu G, Du L, Yang Y, Liu J, et al. Integrated metagenomic and metabolomic analysis reveals distinct gut-microbiome-derived phenotypes in early-onset colorectal cancer. *Gut.* 2023;72(6):1129–42.
45. Totaro A, Panciera T, Piccolo S. YAP/TAZ upstream signals and downstream responses. *Nat Cell Biol.* 2018;20:888–99.
46. Chen HI, Einbond A, Kwak S-J, Linn H, Koepf E, Peterson S, et al. Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands. *J Biol Chem.* 1997;272:17070–7.
47. Te Heesen S, Janetzky B, Lehle L, Aebi M. The yeast WBP1 is essential for oligosaccharyl transferase activity in vivo and in vitro. *EMBO J.* 1992;11:2071–5.
48. Chen Y, Takikawa M, Tsutsumi S, Yamaguchi Y, Okabe A, Shimada M, et al. PHLDA1, another PHLDA family protein that inhibits Akt. *Cancer Sci.* 2018;109:3532–42.
49. Ro S. Multi-phenotypic role of serum response factor in the gastrointestinal system. *J Neurogastroenterol Motil.* 2016;22:193–200.
50. Zhong Z, Lv M, Chen J. Screening differential circular RNA expression profiles reveals the regulatory role of circTCF25-miR-103a-3p/miR-107-CDK6 pathway in bladder carcinoma. *Sci Rep.* 2016;6:30919.
51. Caillier SJ, Briggs F, Cree BA, Baranzini SE, Fernandez-Viña M, Ramsay PP, et al. Uncoupling the roles of HLA-DRB1 and HLA-DRB5 genes in multiple sclerosis. *J Immunol.* 2008;181:5473–80.
52. Liu X, Tong X, Zou Y, Lin X, Zhao H, Tian L, et al. Mendelian randomization analyses support causal relationships between blood metabolites and the gut microbiome. *Nat Genet.* 2022;54:52–61.

Publisher's Note

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