Microbiome

RESEARCH

Open Access

The interplay between *Peptostreptococcus* and *Fusobacterium* as novel signatures in colorectal cancer recurrence

Yunhui Zhang[1†], Bowei Zhang[1†], Wenwen Pang[2†], Wentao Gu[1], Xiaotong Wang[3], Hao Yuan[3], Shiyu Fang[3], Jie Zhang[3], Xiang Li[1], Xiaolong Xing[1], Xuemeng Ji[1], Tianxu Liu[1], Fan Wei[1], Chunze Zhang[4,5*] and Shuo Wang[1*]

Abstract

Background After resection surgery, recurrence occurs in more than 30% of colorectal cancer (CRC) patients. Previous studies have highlighted the role of the gut microbiota in CRC initiation and progression; however, the microbial features associated with recurrence remain not completely understood. In particular, integrated multi-omics biomarkers, the interactions between microorganisms, as well as those between microbes and the host, and their relevance to recurrence risk require further investigation.

Results In this study, 120 tumor mucosal samples collected during surgery from CRC patients underwent 16S rRNA gene sequencing and LC–MS metabolomic profiling. Recurrence status was determined through postoperative follow-up. Compared to the relatively minor variations in mucosal microbial and metabolomic signatures across different tumor node metastasis (TNM) stages, recurrent patients exhibited a distinct landscape. Machine learning analysis identified an integrated signature comprising five bacterial genera (*Peptostreptococcus*, *Fusobacterium*, *Bacteroides*, *Porphyromonas*, and *Prevotella*) and five metabolites (alanylglutamic acid, putrescine, arginine, histidine, and sebacic acid), which demonstrated strong discriminatory performance between recurrent and non-recurrent patients (AUC = 0.89). By integrating 10 biomarkers, a comprehensive risk score for patient stratification was derived. After adjusting for TNM stage, patients classified as high-risk had a significantly shorter recurrence-free survival (adjusted HR = 1.59, 95% CI 1.35–1.88, $P < 0.0001$). The microbial biomarkers *Fusobacterium* and *Peptostreptococcus* displayed positive correlation and were observed to co-aggregate and form dense dual-species biofilms. Further cellular and murine experiments revealed that *P. anaerobius* significantly enhanced the adhesion of *F. nucleatum* to tumor cells and its colonization of colonic mucosa. KEGG pathway analysis of differential metabolites identified enrichment of arginine and proline metabolism pathways in the recurrence group. Concurrently, arginine was found to disrupt *F. nucleatum–P. anaerobius* co-aggregation, while its metabolite putrescine significantly promoted dual-species biofilm formation.

†Yunhui Zhang, Bowei Zhang, and Wenwen Pang contributed equally to this work.

*Correspondence:
Chunze Zhang
chunze.zhang@nankai.edu.cn
Shuo Wang
wangshuo@nankai.edu.cn
Full list of author information is available at the end of the article

Conclusion Our study identified integrated microbial and metabolic features associated with CRC recurrence and proposes an exploratory risk stratification framework. Furthermore, a previously unrecognized link between pathobionts and metabolites relevant to recurrence was reported, which requires further validation in a larger independent cohort.

Keywords Colorectal cancer, Microbiome, Metabolome, Machine learning, Recurrence prediction

Background

Colorectal cancer (CRC) represents the third most prevalent malignancy globally and is the second leading cause of cancer-related mortality [1, 2]. For patients with stage II or III CRC, surgical resection constitutes the standard of care; however, postoperative recurrence occurs in over 30% of these individuals [3, 4]. It is estimated that 80% of colon cancer recurrences manifest within the first 3 years following surgical resection [5]. Conventional histopathological criteria for tumors, including TNM stage, tumor grade, and the presence of carcinoma cells within tumor vasculature or lymphatic channels, have been associated with postoperative recurrence outcomes [6]. Given the considerable inter-observer variability in the assessment of these pathological entities [6, 7], the development of prognostic biomarkers targeting the molecular pathogenic pathways underlying CRC development and metastasis is required. This will allow for refined prognostic stratification of CRC patients following surgery.

Genetic and environmental factors are significant contributors to tumor development. Gastrointestinal malignancies, owing to their proximity, are profoundly influenced by the gut microbiota [8, 9]. Within CRC pathogenesis, the gut microbiota has been identified as a distinct environmental risk factor, primarily impacting CRC progression through genotoxicity, host immune modulation, and activation of carcinogenic signaling pathways [10]. For instance, colibactin secreted by *pks*+*E. coli* induces DNA interstrand crosslinks, leading to mutations in colonic epithelial cells and promoting CRC tumorigenesis [11, 12]. Numerous studies report that enrichment of *F. nucleatum* is associated with CRC initiation, progression, and poor prognosis [13–16]. Furthermore, *F. nucleatum* colonization established in primary CRC has been observed to persist within distant metastases coexisting with mixed bacterial communities [17]. Within these sites, *F. nucleatum* continues to facilitate tumor growth and promote chemoresistance by stimulating cancer cell autophagy [18]. The intestinal metabolome, serving as the chemical interface between the host and gut microbiota, reflects metabolomic alterations associated with tumor cell states and holds significant potential for application in CRC diagnosis and recurrence prediction [19–22]. Microbiota-derived metabolites influence CRC pathogenesis in a remarkably broad manner. For instance, bile acids tauro-β-muricholic acid (T-βMCA) and deoxycholic acid (DCA) antagonize intestinal farnesoid X receptor (FXR) signaling, thereby promoting the proliferation of intestinal cancer stem cells [23]. Conversely, the short-chain fatty acid butyrate acts as a histone deacetylase (HDAC) inhibitor, stimulating apoptosis to prevent CRC initiation [24]. However, the production of metabolites associated with CRC recurrence and their potential biological roles in both the host and microbiota remain poorly defined. Notably, mucosal sites represent a significant source of the intratumoral microbiota, while hematogenous dissemination may also contribute to pathogen invasion [25, 26]. The bacterial driver–passenger model further suggests that dominant pathogenic bacteria within tumors create ecological niches favorable for the colonization of other microbes [27, 28]. Therefore, integrated microbiome and metabolomic analysis of mucosal tissues is required to elucidate the interactions between the microbiota and metabolites within the tumor microenvironment and their role in CRC recurrence.

Untargeted metabolomics and microbiome profiling technologies enable comprehensive characterization of the diverse microbial taxa and metabolites present within tissue samples; however, the processing of these massive datasets and interpretation of their clinical relevance remain challenging. The machine learning-driven analysis has advantages in dealing with high-dimensional features and nonlinear relationships in omics data [29]. These approaches allow for the development of predictive models using screened biomarkers and support patient risk stratification. Recent studies have demonstrated the utility of machine learning algorithms in distinguishing CRC patients at different disease stages [21], and in predicting postoperative recurrence risk and survival outcomes using long non-coding RNA (lncRNA) signatures and histopathology images [30, 31]. However, studies specifically focused on identifying integrated multi-omics signatures to stratify postoperative recurrence risk among CRC patients remain limited. Machine learning algorithms could therefore be leveraged to uncover distinct microbiota compositions and metabolic pathways within the tumor microenvironment of

recurrent patients, thereby facilitating the development of novel prognostic biomarkers for postoperative recurrence risk management.

Here, we profiled the tumor mucosal microbiome of 125 CRC patients using 16S rRNA gene amplicon sequencing and characterized the corresponding metabolomic landscape in a subset of samples ($n = 120$) using LC–MS. Recurrent patients exhibited distinct microbial and metabolomic signatures compared to those with stage I-III CRC who remain recurrence-free. A machine learning analysis was conducted by integrating multi-omics data and clinical follow-up information to identify candidate features related to the risk of recurrence. The resulting model was able to distinguish patients with and without postoperative recurrence. These prognostic features were further integrated into a composite risk score, with higher-risk patients showing significantly shorter recurrence-free survival (RFS). Furthermore, correlation analysis between bacterial abundance and metabolite levels, combined with experimental validation, revealed that arginine and its metabolite putrescine modulate the adhesion and biofilm formation of CRC-associated pathobionts *F. nucleatum* and *P. anaerobius*. Collectively, these findings highlight the potential biological role of metabolic biomarkers in modulating bacterial interactions within the tumor microenvironment.

Materials and methods

Patient recruitment and sample collection

This study enrolled 126 pathologically confirmed CRC patients who underwent curative or palliative resection at Tianjin Union Medical Center between October 2021 and December 2022. Exclusion criteria included: (1) history of colorectal surgery or any type of malignancy (except for CRC); (2) previous chemotherapy, radiotherapy, immunotherapy, or neoadjuvant treatment; (3) intestinal obstruction or other severe systemic diseases; (4) use of antibiotics within the last month. All individuals in this study received standard preoperative mechanical bowel preparation. During surgery, mucosal tissues were aseptically collected from tumor sites, immediately placed into cryopreservation tubes, flash-frozen in liquid nitrogen, and stored at − 80 °C. According to the institution's clinical practice, all patients received prophylactic intravenous cefuroxime sodium injections after surgery. A total of 69% of patients received postoperative adjuvant chemotherapy. The TNM staging followed the American Joint Committee on Cancer (AJCC) staging handbook. Postoperative recurrence was determined through regular clinical examinations, telephone follow-ups, and radiological assessments. One tissue sample was excluded due to quality control (QC) failure prior to 16S rRNA sequencing. Ultimately, 16S rRNA sequencing was performed on 125 tissue samples, and untargeted metabolomic profiling was conducted on 120 samples with sufficient remaining tissue. The datasets used in each statistical and machine learning analysis are further illustrated in Supplementary information (Fig. S1). This study protocol adhered to the principles of the Declaration of Helsinki and received approval from the Ethics Committee of Tianjin Union Medical Center (2021-B37). Written informed consent for sample collection and subsequent analysis was obtained from all participants. Detailed patient-level clinical information, including TNM stage and recurrence status, is provided in Supplementary file S1.

DNA extraction and 16S rRNA amplicon sequencing

Genomic DNA was extracted from mucosal samples using the ZR Fungal/Bacterial DNA Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's protocol. DNA integrity and fragment size were verified by 1.0% agarose gel electrophoresis, and DNA concentration and purity were evaluated using NanoDrop spectrophotometry (Thermo Fisher Scientific, Wilmington, DE, USA). The V3–V4 hypervariable regions of the bacterial 16S rRNA genes were amplified using the following forward and reverse primers: 5′–TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC AG–3′and 5′–GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CHV GGG TAT CTA ATC C–3′. Amplicons were purified using VAHTS DNA Clean Beads (Vazyme), quantified, and sequenced on the Illumina MiSeq platform (San Diego, CA, USA) using 2 × 300 bp paired-end chemistry according to the manufacturer's instructions.

16S rRNA gene sequencing data analysis

Briefly, 16S rRNA gene sequencing data were processed using QIIME2 (v2019.4). Raw paired-end reads were quality filtered and denoised using the DADA2 plugin to generate amplicon sequence variants (ASVs). Taxonomic classification was performed using a Naïve Bayes classifier implemented in the q2-feature-classifier plug-in and trained on the SILVA database (v138.1). ASVs unclassified at the bacterial domain level were excluded from downstream analysis. Mitochondrial reads accounted for approximately 0.005% of the total sequences. Non-specific sequence features from chloroplast and mitochondria were further removed using the remove.lineage command in mothur. Alpha diversity indices (Ace, Chao, Sobs, and Shannon) and beta diversity metrics (Bray–Curtis and unweighted UniFrac distance) were calculated using mothur. The relative abundance of bacterial genera for all samples is shown in the Supplementary file S2.

Metabolite extraction and LC–MS analysis

Mucosal samples were extracted with 1 ml of ice-cold methanol aqueous solution (4:1, v/v), followed by ice-bath sonication for 30 min and incubation at − 20 °C for 1 h. After centrifugation (16,000×*g*, 20 min, 4 °C), supernatants were collected and evaporated to dryne1ss. For LC–MS analysis, residues were reconstituted in 100 µl methanol aqueous solution (1:1, v/v), centrifuged (20,000×*g*, 15 min, 4 °C), and supernatants injected into a SHIMADZU LC-30A UHPLC system equipped with an ACQUITY UPLC® HSS T3 column (2.1×100 mm, 1.8 µm; Waters, Milford, MA, USA). Chromatographic separation was performed at 40 °C with a flow rate of 0.3 ml/min using 0.1% formic acid in water (mobile phase A) and acetonitrile (mobile phase B) under the following gradient program: 0–2 min (0% B), 2–6 min (linear gradient 0–48% B), 6–10 min (48–100% B), 10–12 min (100% B held), 12–12.1 min (100–0% B), and 12.1–15 min (0% B re-equilibration). Mass spectrometric detection was performed on a Q Exactive™ Plus instrument (Thermo Scientific) operated in both positive and negative electrospray ionization (ESI) modes, with a scan range of *m/z* 70–1050. Raw LC–MS data were processed using MS-DIAL for peak detection, alignment, retention time (RT) correction, and peak area extraction, with metabolite identification achieved through interrogation of HMDB, MassBank, GNPS databases, and in-house standards, where only matched RT, MS, and MS/MS spectra were considered for annotation. System stability validation incorporated both QC sample base peak chromatogram consistency assessment and principal component analysis (PCA) clustering evaluation across all samples.

For the merged LC–MS data (positive and negative modes), peak areas were normalized to the total ion current and then multiplied by 10^6 for numerical scaling. Metabolic features with more than 50% missing values across samples or with a relative standard deviation (RSD) exceeding 30% in QC samples were removed. Furthermore, HMDB 5.0 was used to select endogenous metabolites [32], in order to remove those derived from food, drugs, or other exogenous sources [21]. Sparse partial least squares discriminant analysis (sPLS-DA) was applied for comparisons among three groups (stages I, II, and III), while orthogonal partial least squares discriminant analysis (OPLS-DA) was used for two-group comparisons, enabling dimensionality reduction and group discrimination for differential metabolite screening. Detailed metabolite annotations and intensity data are provided in Supplementary file S3.

Mfuzz soft clustering analysis

Following *z*-score normalization of bacterial and metabolite relative abundances, soft clustering analysis was performed using the Mfuzz algorithm based on fuzzy c-means clustering to characterize stage-associated abundance patterns across CRC progression. All analyses were conducted in R (v4.5.0) using the Mfuzz package (v2.68.0). To determine the optimal number of clusters, a continuous range of cluster numbers ($k=2$–9) was systematically evaluated using the Elbow method based on within-cluster variance. For each candidate k, the fuzzifier parameter (m) was estimated using the standard Mfuzz procedure, and the inflection point was identified based on quantitative changes in variance reduction [33]. Based on this assessment, four clusters were selected for microbial taxa and five clusters for metabolites, and these cluster numbers were used for downstream analysis. Cluster membership values were calculated to quantify the degree of association of each feature with individual clusters. For visualization, abundance patterns were displayed using raincloud plots generated with the ggplot2 (v3.5.2) and gghalves (v0.1.4) packages. Representative microbial genera were further summarized using a heatmap (Fig. S3), focusing on taxa with a relative abundance greater than 0.01% in at least 50% of samples. In addition, the top three metabolites with the highest membership values within each cluster were visualized using raincloud plots (Fig. S4C).

Random forest modeling with LASSO-based feature selection

A Random Forest (RF) classifier, coupled with Least Absolute Shrinkage and Selection Operator (LASSO)-based feature selection, was developed to predict CRC recurrence [34]. The initial input feature matrix consisted of relative abundance profiles of 434 bacterial genera derived from tumor mucosal microbiome profiling. LASSO regression was employed for feature selection to address the high-dimensional nature of the data and to identify informative features for the binary classification task, as L1 regularization shrinks less important feature coefficients to zero, thereby reducing overfitting and improving model generalizability and interpretability [35]. Given that microbiome relative abundance data are compositional and thus exhibit feature dependence due to the sum-to-constant constraint, LASSO was applied in the context of predictive modeling, where the goal was biomarker selection for binary classification rather than inference of full community interaction structures. In this setting, LASSO can reduce redundancy by selecting representative features from correlated feature sets. By the end of follow-up, postoperative recurrence had developed in 27 CRC patients. To reduce potential confounding and class imbalance effects in machine learning, samples were balanced for TNM stage, age, gender, tumor location, and tumor differentiation. A matched cohort

of 27 recurrent and 27 non-recurrent CRC patients was included for comparative analysis and RF modeling. The cohort was randomly split once into a training set (70%) and an independent test set (30%) using stratified sampling to maintain class proportions. To ensure unbiased performance estimation, all feature selection, data transformation, and model training procedures were restricted to the training data. The test set was used solely for final evaluation and was not involved in any preprocessing or model optimization steps. Feature-wise mean-centering transformation was applied. Next, feature selection was performed using LASSO regression with an L1 regularization coefficient (α) of 0.01 and a maximum of 1000 iterations. Features with non-zero coefficients were selected. Crucially, to enhance the stability of the feature selection process, LASSO was fitted within a repeated stratified five-fold cross-validation (CV) framework, in which the composition of the training folds varied across iterations, and the frequency of feature selection was computed across all resampling iterations. Five microbial features with non-zero selection frequency were retained as the final candidate predictive feature set. An RF classifier (with 100 estimators and Gini impurity criterion) was trained using the selected features. The predictive performance was evaluated on a single independent test set using metrics including accuracy, precision, specificity, and sensitivity, along with confusion matrix visualization. The receiver operating characteristic (ROC) curves were generated, and the area under the curve (AUC) was calculated. To estimate the robustness of the model, a 95% confidence interval (CI) for AUC was estimated through stratified bootstrap resampling (1000 iterations) on the test set, preserving class proportions.

In parallel, a metabolome-based machine learning model was constructed following the aforementioned methodology. Initially, normalized intensity values of 46 differential metabolites derived from untargeted metabolomics analysis were used as candidate features for LASSO-based feature selection, and the analysis was applied to a cohort of 52 samples partitioned with a test size proportion of 0.3 (training set $n=36$, test set $n=16$). Notably, numerical features underwent z-score normalization. Similarly, the five top-selected metabolic predictors were identified using the LASSO-RF framework.

Furthermore, these metabolic features were integrated with the previously selected microbial features to construct a combined multi-omics RF classifier comprising ten predictors (five genera and five metabolites). All analyses were conducted using scikit-learn (v1.2.2) in Python (v3.11).

Model interpretability analysis

Model interpretability was assessed using Shapley Additive exPlanations (SHAP) [36]. Shapley values were computed using the TreeExplainer algorithm. Global model interpretation was performed using SHAP summary plots in a violin-style format, which visualize the distribution of Shapley values across samples for each feature within the model, thereby illustrating the model-based contribution of input features to the prediction of postoperative CRC recurrence.

Integrated microbiome-metabolite risk score construction and survival analysis

In the preceding analysis, LASSO-selected microbial and metabolic features were first evaluated using a conventional RF classifier to assess their combined discriminative performance for binary postoperative recurrence status, as an exploratory analysis without incorporating survival time or tumor stage. To extend this analysis, a random survival forest (RSF) framework was subsequently applied using the same feature set to derive an integrated risk score [37], enabling time-to-event modeling and downstream prognostic stratification analysis with adjustment for tumor stage.

The analysis was conducted on 115 patients out of a total of 120 patients in the complete cohort, with precise surgical dates, allowing accurate calculation of RFS. It included five microbial taxa and five metabolic features. Microbial abundances were zero-imputed using a pseudo-count of 1×10^{-6} and transformed by centered log-ratio (CLR), while metabolite intensities were standardized by z-score normalization. All modeling procedures were implemented in R using the randomForestSRC package (v3.4.1). To ensure unbiased risk estimation, the microbiome-metabolite risk score was generated within a stratified 5-fold cross-validation scheme based on recurrence event status. Within each training fold, an RSF model was fitted using RFS time (months) and recurrence status, with 1000 survival trees. The risk score was defined as the expected number of recurrence events for each patient. It was derived from the ensemble cumulative hazard estimated by the RSF model, using the predict function implemented in the randomForestSRC package. This metric aggregates the cumulative hazard across all trees in the forest, thereby capturing potential non-linear effects and high-order interactions between microbial and metabolic features. Specifically, for each patient in the held-out fold, the model trained on the remaining folds was used to predict their recurrence risk without access to their own outcome information.

Subsequently, patients were divided into low- and high-risk groups based on the median risk score (low-risk, $n=58$; high-risk, $n=57$). Survival analysis was subsequently performed to evaluate the prognostic value of the integrated microbiome-metabolite risk score, with TNM

stage included as an adjustment variable in Cox proportional hazards models. This approach was used to assess whether the risk score provided independent prognostic stratification beyond pathological stage.

Bacterial strains and culture conditions

Fusobacterium nucleatum subsp. *Nucleatum* (ATCC 25586) and *Peptostreptococcus anaerobius* strain VPI 4330 (catalogue number ATCC 27337) were obtained from the Guangdong Microbial Culture Collection Center (Guangdong, China). *F. nucleatum* and *P. anaerobius* were cultured under anaerobic conditions at 37 °C in brain heart infusion (BHI) broth (CM1135B, Thermo Scientific) supplemented with 5 μg/ml hemin (H8130, Solarbio), 1 μg/ml vitamin K1 (VK1, V8150, Solarbio), and 0.5 mg/ml L-cysteine hydrochloride monohydrate (C0011, Solarbio). *Escherichia coli* strain MG1655 (catalogue number 700926) was cultivated in Luria–Bertani (LB) medium at 37 °C under aerobic conditions.

Bacterial aggregation and inhibition assays

After reaching the logarithmic growth phase, bacterial cells were harvested by centrifugation at 5,000 rpm for 5 min, washed twice with sterile phosphate-buffered saline (PBS), and resuspended in PBS. Equal volumes (0.5 ml) of each bacterial suspension were mixed in a cuvette, gently vortexed for 10 s, and the optical density at 600 nm (OD_{600}) was measured immediately (designated as the 0 h time point). The mixture was then incubated under anaerobic conditions at 37 °C for 2 or 5 h, after which the OD_{600} was recorded (2 h or 5 h time point). The aggregation percentage was calculated using the following formula:

$$Aggregation(\%) = 100\% - \left(\frac{OD_{600}\text{ at 2h or 5h}}{OD_{600}\text{ at 0h}}\right) \times 100\%$$

For inhibition assays, *F. nucleatum* was pre-incubated with 50 mM L-arginine or D-galactose at 37 °C for 10 min. Subsequently, an equal volume of *P. anaerobius* suspension was added, and the co-culture was incubated anaerobically at 37 °C for 2 or 5 h prior to OD_{600} measurement.

Bacterial growth curves

Bacterial growth curves were determined by monitoring OD_{600}. *F. nucleatum* and *P. anaerobius* were initially cultured anaerobically in BHI broth for 24 h. For monoculture controls, each bacterium was inoculated at 1×10^6 CFU per well in 200 μL of fresh BHI broth. For co-culture experiments, the following inocula were prepared in 200 μL BHI: (1) a 1:1 mixture (5×10^5 CFU of each species); (2) a 1:10 mixture (5×10^5 CFU *F. nucleatum* and 5×10^6 CFU *P. anaerobius*); and (3) a high-density *P. anaerobius* control (1×10^7 CFU). Plates were incubated under anaerobic conditions at 37 °C, and OD_{600} was measured at 12, 24, 36, and 48 h post-inoculation.

To evaluate the effects of putrescine on bacterial growth dynamics, a 1:1 co-culture (total inoculum of 1×10^6 CFU) was supplemented with filter-sterilized putrescine at final concentrations of 50, 100, or 500 μM. Wells supplemented with an equal volume of PBS served as the vehicle control. For these assays, the OD_{600} was measured at 4, 8, 12, and 24 h post-inoculation.

Biofilm assays

Biofilm formation was assessed using a crystal violet staining assay. Overnight cultures of *F. nucleatum* and *P. anaerobius* were diluted to 5×10^8 cells/ml, and 1% (v/v) of each culture was inoculated into fresh BHI broth (200 μl per well) in a 96-well plate (3599, Corning). Plates were incubated at 37 °C for 24 h within an anaerobic workstation (Gene Science, Wilmington, DE, USA) to allow biofilm formation. After incubation, the culture medium was carefully aspirated, and the biofilms were gently washed 2–3 times with sterile PBS. The washed biofilms were air-dried at 37 °C for 1 h and then stained with 100 μl of 0.1% (w/v) crystal violet solution for 30 min. Excess stain was removed, and wells were washed 2–3 times with sterile PBS. After air-drying, the bound crystal violet was solubilized by adding 200 μl of 96% (v/v) ethanol aqueous solution and incubating for 30 min. Biofilm biomass was quantified by measuring absorbance at 595 nm (OD_{595}). To evaluate the effects of putrescine on biofilm formation, putrescine was added to the culture medium at final concentrations of 10 μM or 50 μM prior to the 24-h incubation period. Wells supplemented with an equal volume of PBS served as vehicle controls.

Fluorescence imaging of biofilms

F. nucleatum and *P. anaerobius* were labeled with 10 μM CellTracker™ Green CMFDA (C2925, Thermo Scientific) and CellTracker™ Red CMTPX (C34552, Thermo Scientific), respectively, for 1 h. Subsequently, bacterial cells were washed three times with PBS to remove excess dye. The labeled bacterial suspensions were then inoculated (1% inoculum) into fresh BHI medium (1 ml per well) in a 12-well plate (3513, Corning). For co-culture experiments, *F. nucleatum* and *P. anaerobius* were mixed at a 1:1 ratio, with equal volumes contributing to a total

inoculum of 1%. Plates were incubated anaerobically for 24 h to allow biofilm formation. Following incubation, the biofilms were gently washed three times with sterile PBS. Fluorescence imaging was performed using a Nikon Eclipse Ti2 inverted microscope.

Confocal microscopy

The human CRC cell line HCT116 was maintained in Dulbecco's modified Eagle's medium (DMEM; 8,123,298, Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS; 10,099-141C, Gibco) and 1% (v/v) penicillin–streptomycin (15,140,122, Gibco). Cells were cultured at 37 °C in a humidified incubator with 5% CO_2.

To visualize bacterial adhesion to CRC cells, HCT116 cells were seeded in 12-well plates and grown to approximately 80% confluence. Fluorescently pre-labeled *F. nucleatum* (CMFDA) and *P. anaerobius* (CMTPX) were added to the HCT116 monolayers either individually or in combination at a 1:1 ratio and co-incubated for 1 h at 37 °C at a multiplicity of infection (MOI) of 50. Following incubation, non-adherent bacteria were removed by washing twice with PBS. Cells were then fixed with 4% (w/v) paraformaldehyde and stained with 4′,6-diamidino-2-phenylindole (DAPI; D1306, Invitrogen) to visualize cell nuclei. Confocal imaging was performed using a Zeiss LSM 800 confocal microscope (Carl Zeiss) equipped with appropriate laser lines and filters.

Quantification of bacterial adhesion by fluorescence microplate reader

HCT116 cells were seeded in black 96-well plates (3916, Corning) and cultured to approximately 80% confluence. Two hundred microliters of pre-stained bacterial suspensions were added to cell monolayers at an MOI of 50 for individual strains (*F. nucleatum* alone or *P. anaerobius* alone) or at an MOI of 100 (1:1 co-culture ratio). Cells and bacteria were co-incubated for 1 h at 37 °C. After incubation, non-adherent bacteria were removed by gently washing the monolayers twice with pre-warmed PBS. Fluorescence intensity was then measured using a fluorescent microplate reader (Spark, Tecan, Switzerland).

Murine model and bacterial colonization

Eight-week-old male C57BL/6J mice of specific pathogen-free (SPF) grade were obtained from Vital River Laboratory Animal Technology Co., Ltd. (Beijing, China). All animal experiments were approved by the Animal Care and Use Committee of Nankai University (SYXK2019-0001) and conducted in accordance with institutional guidelines. Mice were housed under a 12-h light/dark cycle with ad libitum access to sterile water and standard chow.

After a one-week acclimatization period, mice were treated with an antibiotic cocktail consisting of streptomycin (1 mg/ml), gentamicin (1 mg/ml), ampicillin (0.5 mg/ml), and vancomycin (0.5 mg/ml) administered via drinking water for 3 consecutive days. Antibiotic-containing water was then replaced with sterile drinking water for an additional 2 days prior to bacterial inoculation.

Bacterial colonization was induced by oral gavage once daily for 5 consecutive days. Each mouse received 1×10^8 CFU of either live *F. nucleatum* or *P. anaerobius*, or co-culture bacteria suspended in 200 µl of PBS. Fresh fecal samples were collected at 1-, 3-, 6-, and 9-days post-inoculation cessation, immediately snap-frozen in liquid nitrogen, and stored at − 80 °C until further analysis.

Fecal colonization levels of *F. nucleatum* and *P. anaerobius* were quantified by quantitative PCR (qPCR). Bacterial genomic DNA (gDNA) was extracted using a Tiangen gDNA isolation kit (DP328, Tiangen) according to the manufacturer's instructions. Bacterial abundance in fecal and colonic mucosal samples was determined by qPCR using ChamQ Universal SYBR qPCR Master Mix (Q711, Vazyme) on a CFX Connect Real-Time PCR System (BIO-RAD, Redmond, WA, USA). The following primers were used: *F. nucleatum*, forward 5′-CTG TAT TGC GTT GGA AAC TGT GTA A-3′ and reverse 5′-TAC CAG GGT ATC TAA TCC TGT TTG C-3′; *P. anaerobius*, forward 5′-AGA CGA ATT CAA GTC AGT AAA TAC A-3′ and reverse 5′-CTC CTA TCC ACC AGG ATA TCA A-3′ [38]. Primer specificity was verified by negative controls and melting curve analysis. Using gDNA from untreated mouse fecal samples as negative controls, all reactions yielded cycle threshold (Ct) values greater than 37, indicating the absence of non-specific amplification. Single melting peaks were observed for all target amplicons, confirming amplification specificity and the absence of primer-dimer artifacts. Standard curves were generated using gDNA from reference strains of *F. nucleatum* and *P. anaerobius*. Serial four-fold dilutions of template DNA were subjected to qPCR alongside experimental samples. Ct values were plotted against the logarithm of known template copy numbers to generate linear regression models. Absolute bacterial loads were calculated using the following formula:

$$\mathrm{DNA}\frac{\mathrm{copies}}{\mu l} = 6.03\times10^{23}\times\mathrm{DNA\ concentration}\times10^{-9}/(\mathrm{DNA\ length}\times660)$$

Other statistical analyses and visualizations

For metabolomic analysis, differential metabolite abundance was assessed using the R package MetaboAnalystR (version 4.0). Two-sided Mann–Whitney *U* tests were applied for two-group comparisons of unpaired samples, while two-sided Kruskal–Wallis tests were used for three-group comparisons. Metabolites with a *P*-value corrected by FDR ≤ 0.05 were considered significantly different. Subsequently, based on the topological structure of the KEGG pathways, the metabolites with a *P* value ≤ 0.05 were used for metabolite set enrichment analysis (MSEA). Differential abundance analysis of microbiome data was performed using ANCOM-BC2 [39]. Multiple testing correction was conducted using the Benjamini–Hochberg method on the raw *P*-values to calculate *q*-values (FDR-adjusted *P* values). Permutation multivariate analysis of variance (PERMANOVA) was performed based on Bray–Curtis or unweighted UniFrac distance matrices with 999 permutations to statistically assess the differences in microbial community composition. For in vitro and in vivo experimental assays, including bacterial co-aggregation, biofilm formation, bacterial adhesion experiments, and intestinal colonization, one-way analysis of variance (ANOVA) followed by Dunnett's or Tukey's post hoc tests was applied for multi-group comparisons, as appropriate. Two-tailed Student's *t*-tests were used for two-group comparisons. All quantitative data were presented as mean ± standard deviation (SD). Clinical variables between groups were compared using the chi-squared test or Fisher's exact test, as appropriate. Spearman rank correlation coefficients were calculated using the rstatix package (v0.7.2) in R. Correlations meeting the dual criteria of $P < 0.05$ and an absolute correlation coefficient $|R| \geq 0.35$ were subsequently visualized as interaction networks using Cytoscape (v3.10.0). Statistical analysis and data visualization were performed using GraphPad Prism (v7.0) and R packages including ggplot2 (v3.5.2) and patchwork (v1.3.0).

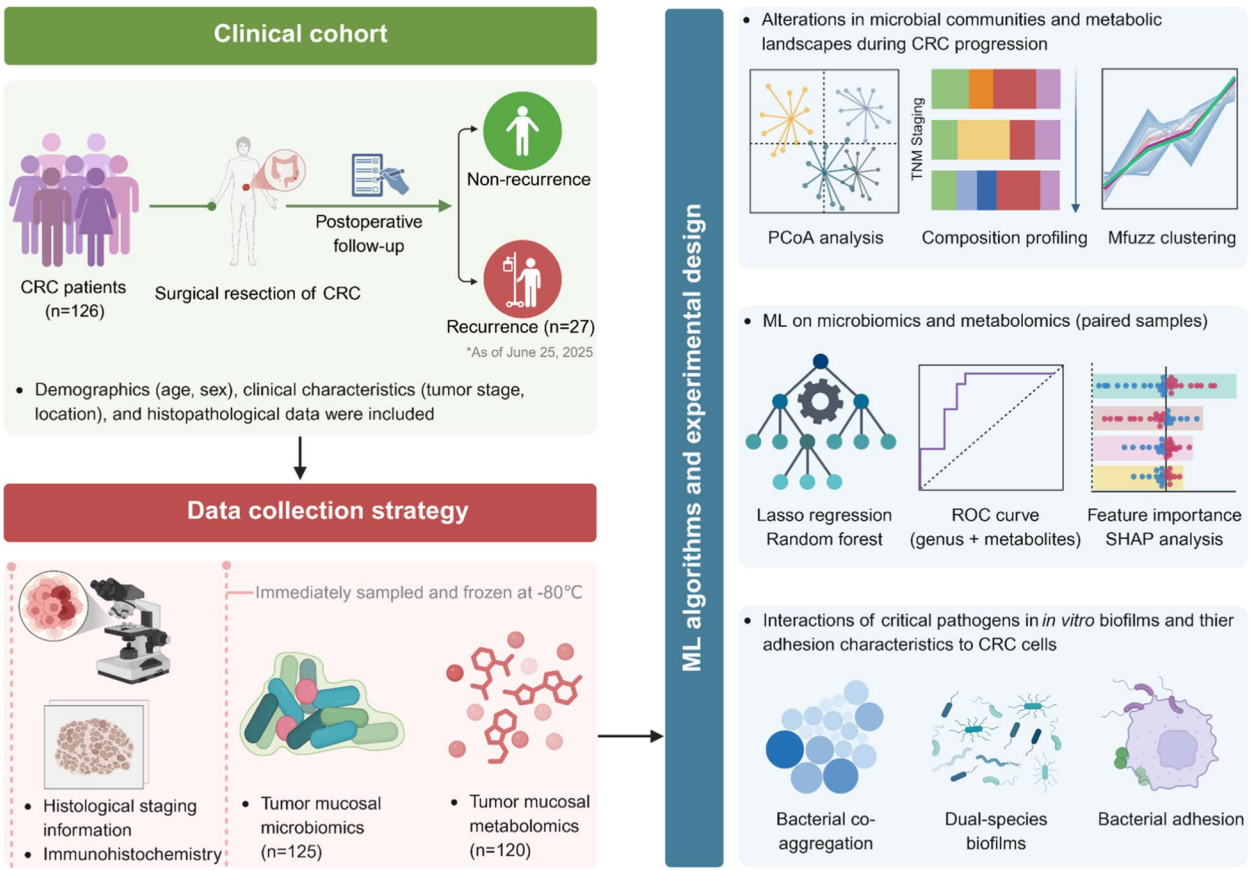

Fig. 1 Overview of the study design. A total of 126 CRC patients were enrolled in this study. Among these, 16S rRNA sequencing was performed on 125 tumor mucosal samples, while untargeted metabolomic assays were performed on 120 tumor mucosal samples. Integrated machine learning analysis of patient-derived microbiome, metabolomic, and clinical characteristics was utilized to develop a prognostic model for CRC recurrence. Furthermore, the interactions between biomarkers were mechanistically validated through in vitro cellular and in vivo animal experiments

Results

Sample characteristics and study design

The overall workflow of this study and exact sample numbers for each analysis are documented in the study design schematic (Fig. 1). A total of 126 CRC patients were included in the cohort analysis, with their demographic and clinical characteristics summarized in Table S1. At final follow-up (June 25, 2025), 27 of 126 surgically resected CRC patients (21.4%) developed disease recurrence. The median follow-up duration was 33.47 months (IQR 30.93–37.07 months) and the median time to recurrence was 12.80 months (IQR

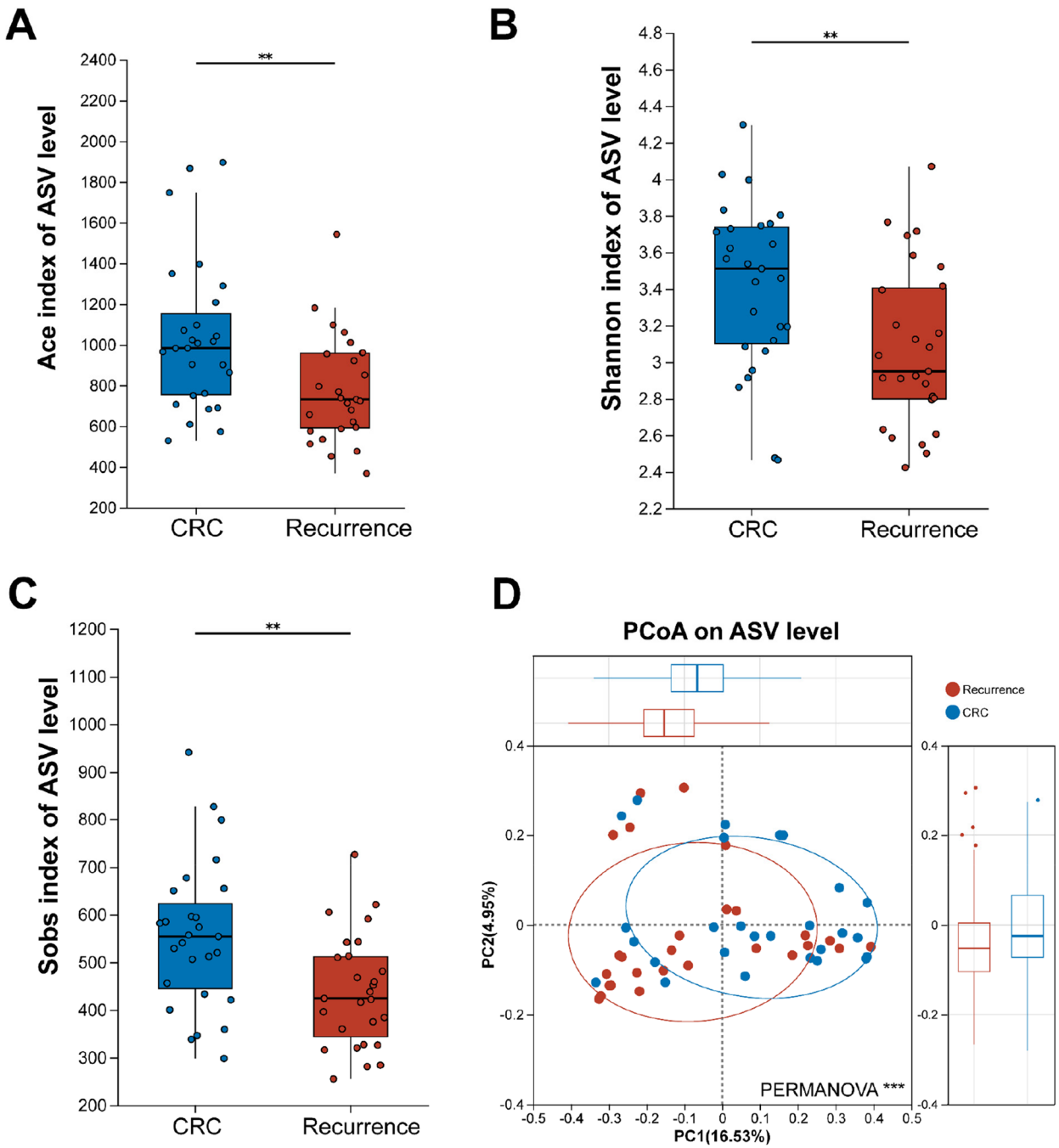

Fig. 2 Alterations in microbial structure between non-recurrent and postoperative recurrent CRC patients. **A**-**C** α-diversity indices. **D** PCoA analysis of microbial community structure based on unweighted UniFrac distances, with statistical significance assessed by PERMANOVA

7.12–21.25 months). Baseline clinical characteristics of matched recurrence and non-recurrence groups are shown in Table S2.

Differences in the microbiome features between CRC patients with and without postoperative recurrence

To characterize mucosal microbiome alterations associated with CRC progression and recurrence, we compared microbial diversity and community structure across disease stages (stage 0/I, II, III) and between patients with and without postoperative recurrence. While α-diversity (Ace, Chao, and Sobs indices) remained stable from stage 0/I to III patients (Fig. S2A–C), it was consistently and significantly reduced in recurrent patients compared with non-recurrent patients ($P < 0.01$ for Ace, Shannon, and Sobs indices; Fig. 2A–C). Principal coordinates analysis (PCoA) showed statistically significant differences in overall microbial community structure between recurrent and non-recurrent CRC patients (PERMANOVA, $P < 0.001$, Fig. 2D), although overlap between the two groups was observed, suggesting a modest yet significant compositional shift rather than complete community separation. No significant differences were identified among CRC patients at different stages (Fig. S2D). At the taxonomic level, the phylum *Proteobacteria* and the genus *Escherichia–Shigella* showed decreasing abundances with advancing disease stage (Fig. S2E, F). Mfuzz soft clustering analysis further identified multiple bacterial co-abundance clusters exhibiting synchronous dynamics across TNM stages (Fig. S3). Notably, Cluster

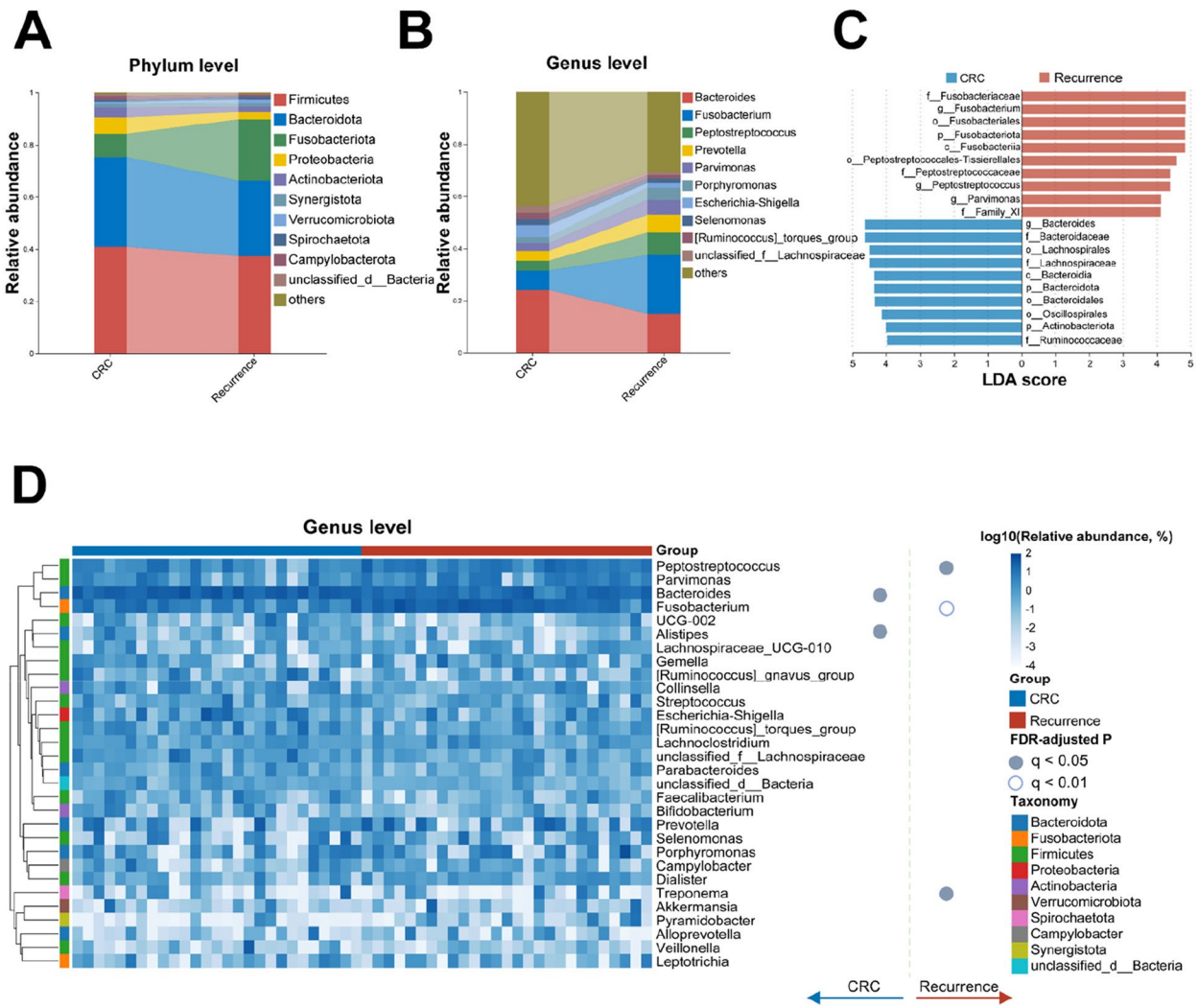

Fig. 3 Differential microbial features between recurrent and non-recurrent CRC patients. **A**, **B** Taxonomic profiles at the phylum level (**A**) and the genus level (**B**). **C** Bar plot of linear discriminant analysis. **D** Heatmap of the genera relative abundance. For differential microbial analysis, statistical significance was determined using FDR-adjusted *P*-values (*q* values)

4, dominated by *Parvimonas*, exhibited a pronounced stage-dependent increase, while *Bacteroides* in Cluster 1 peaked at Stage II before declining.

The phyla *Firmicutes*, *Bacteroidota*, *Fusobacteriota*, *Proteobacteria*, and *Actinobacteriota* were predominant in both the non-recurrence and recurrence groups (Fig. 3A), with *Bacteroides*, *Fusobacterium*, *Peptostreptococcus*, and *Prevotella* representing the most abundant genera (Fig. 3B). In addition, linear discriminant analysis of effect size (LEfSe) was performed with LDA score > 3.5 and q-value < 0.05 thresholds to identify the differentially abundant bacterial species between the recurrent and non-recurrent CRC patients. As shown in Fig. 3C, the CRC group was dominated by members of the *Ruminococcaceae* and *Lachnospiraceae* families and the genus *Bacteroides*. In contrast, the recurrence group showed enrichment of *Fusobacteriaceae* and *Peptostreptococcaceae* families and the genera *Fusobacterium*, *Peptostreptococcus*, and *Parvimonas*. Moreover, the relative abundances of genera *Alistipes* ($q = 0.025$), *Peptostreptococcus* ($q = 0.02$), *Bacteroides* ($q = 0.035$), *Fusobacterium* ($q = 0.006$), and *Treponema* ($q = 0.038$) were significantly altered between the CRC and recurrence groups (Fig. 3D).

Together, the tumor-associated mucosal microbiome exhibited greater compositional stability across disease stages (0/I–III) than between recurrence and non-recurrence groups, with recurrence-associated microbiomes characterized by reduced α-diversity and enrichment of specific bacterial taxa.

Distinct mucosal metabolomic landscape associated with postoperative CRC recurrence

Considering microbial-host metabolic interactions, we performed untargeted metabolomic profiling on tumor mucosal samples stratified by TNM stage and postoperative recurrence status. After removing food-derived and exogenous metabolites, a total of 721 endogenous metabolites were identified. Firstly, the sPLS-DA analysis exhibited that the mucosal metabolome of TNM stages I to III did not show significant separation (Fig. S4A). Subsequently, Mfuzz soft clustering identified progressive increases in anisic acid and N-acetylisoleucine throughout the progression of CRC disease, while the trajectories of creatine, L-asparagine, and uracil showed a downward trend, as shown in Fig. S4B and C.

In contrast, a pronounced and statistically significant metabolic shift was observed between postoperative recurrent and non-recurrent CRC patients. Metabolites with $|\log_2 FC| > 0.58$, $P < 0.05$, and variable importance in projection (VIP) ranking top 30 were highlighted in Fig. 4A. Metabolomic profiling identified outcome-specific metabolic alterations, with non-recurrent patients exhibiting elevated arginine ($q = 0.038$) and 3-methyluric acid ($q = 0.086$). In contrast, metabolites enriched in recurrent patients included putrescine ($q = 0.053$), cadaverine ($q = 0.068$), and hydrocinnamic acid ($q = 0.038$).

To gain insights into the functional significance of these recurrence-associated alterations, we performed the MSEA. Among tumor mucosal metabolites (Fig. 4B), the top four significantly perturbed pathways in recurrent patients compared with non-recurrent CRC patients were arginine biosynthesis (map00220, $q = 0.005$), pyrimidine metabolism (map00240, $q = 0.005$), alanine, aspartate, and glutamate metabolism (map00250, $q = 0.005$), and arginine and proline metabolism (map00330, $q = 0.096$). Subsequently, Spearman correlation analysis ($P < 0.05$) revealed significant associations between the five differentially abundant genera ($q < 0.05$) and metabolites (top 30 VIP) in the multi-omics cohort, depicting potential microbiota-metabolite crosstalk in CRC recurrence (Fig. S5). Metabolites enriched in recurrent cases were generally positively correlated with recurrence-enriched microbial taxa. In the context of recurrence-associated microbial dysbiosis, arginine exhibited a negative correlation with the enriched genus *Peptostreptococcus* ($r = -0.43$, $P = 0.001$) and a positive correlation with the depleted genus *Bacteroides* ($r = 0.54$, $P = 3 \times 10^{-5}$). Conversely, the biogenic amines cadaverine ($r = 0.41$, $P = 0.003$) and putrescine ($r = 0.43$, $P = 0.001$) exhibited strong positive associations with *Peptostreptococcus*. Furthermore, 3-methyluric acid was significantly negatively correlated with the recurrence-enriched microbe *Fusobacterium* ($r = -0.37$, $P = 0.008$). Collectively, postoperative recurrence was characterized by a distinct metabolic landscape and corresponding microbe-metabolite association patterns.

Identification of potential bacterial biomarkers for predicting CRC recurrence

Subsequently, machine learning algorithms were utilized to identify microbial biomarkers associated with CRC recurrence outcomes. Within this machine learning framework, LASSO regression consistently selected five discriminative bacterial genera as core predictors for postoperative outcome stratification. As depicted in Fig. 5A, B, an RF classifier was constructed using these five features and evaluated in an internal independent test cohort, achieving an AUC of 0.81 (95% CI 0.56–0.97). The confusion matrix demonstrated model performance metrics of 0.75 sensitivity, 0.71 accuracy, 0.67 specificity, and 0.67 precision.

In the variable importance (VIMP) plot (Fig. 5C), *Peptostreptococcus*, *Fusobacterium*, and *Bacteroides* contributed most prominently to model performance, each exhibiting a relative VIMP > 20%. To further enhance

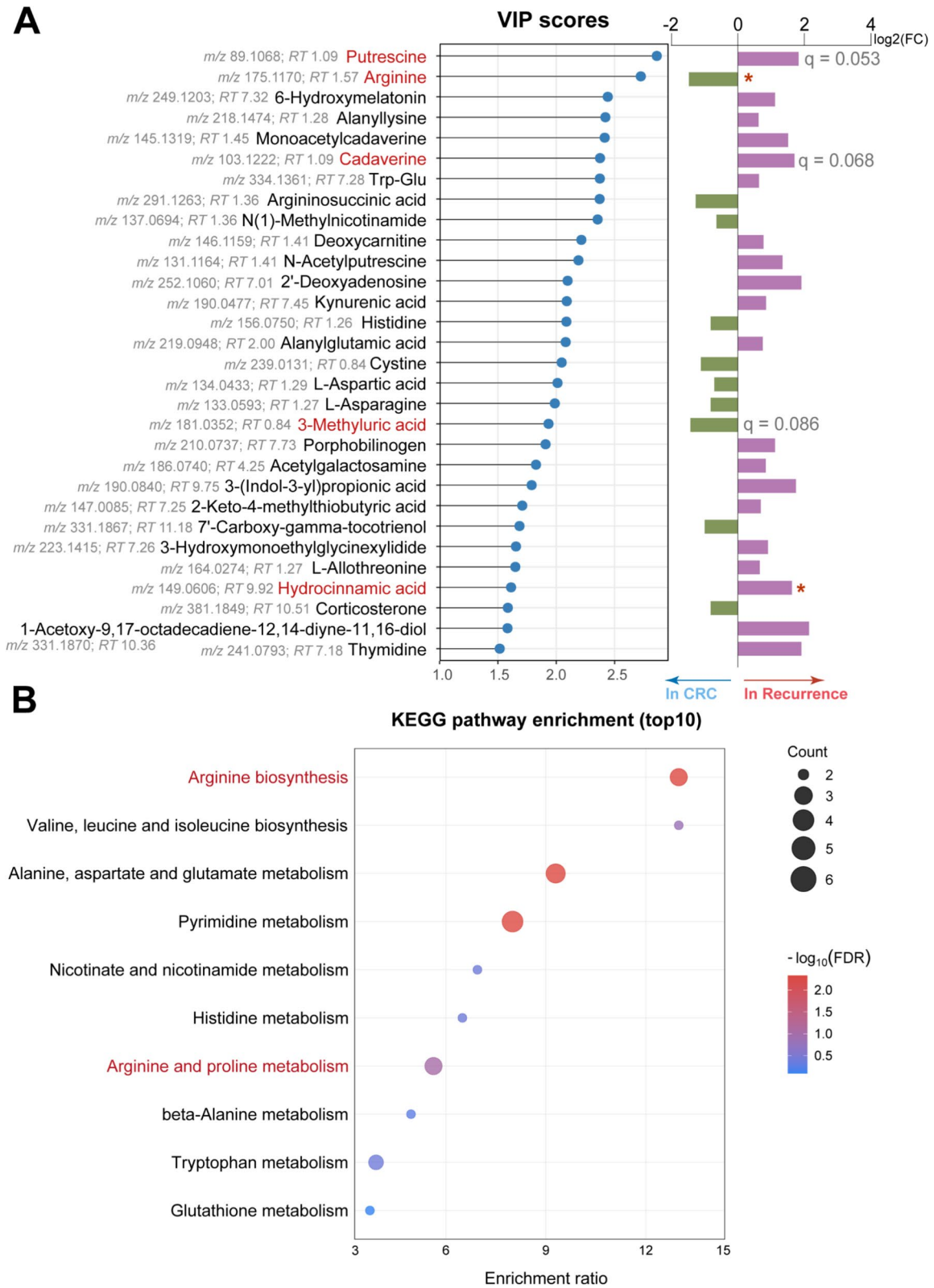

Fig. 4 Reprogrammed tumor mucosal metabolic landscape in postoperative recurrent CRC patients. **A** Differential metabolites with top 30 VIP scores between recurrent and non-recurrent CRC patients. **B** KEGG pathway enrichment analysis of differential metabolites. Differential metabolites and enriched pathways were identified based on FDR-adjusted *P*-values (*q* values)

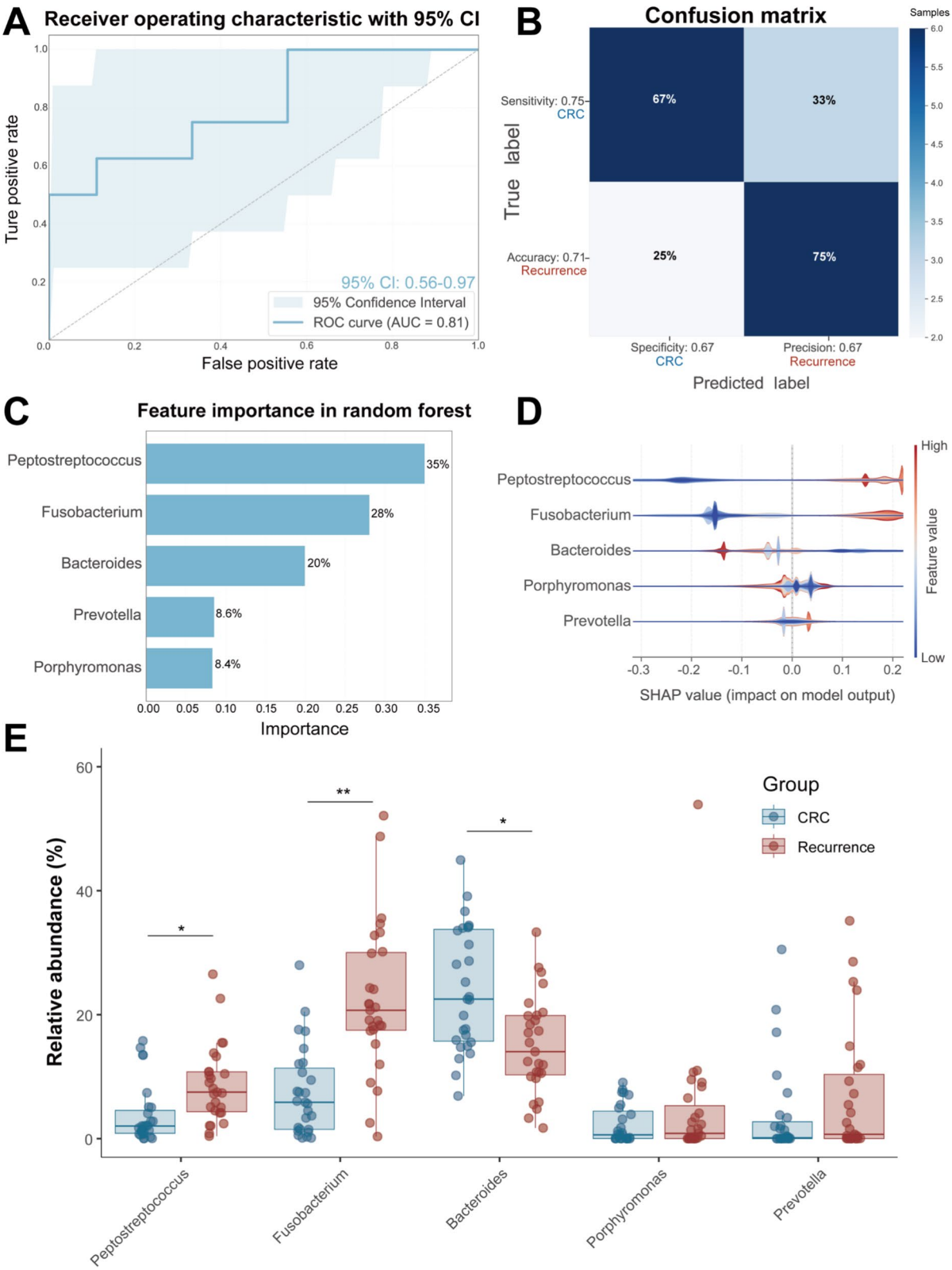

Fig. 5 Random forest model development and validation using microbiome data to discriminate postoperative recurrent and non-recurrent CRC patients. **A** Receiver operating characteristic (ROC) curve of the Random Forest classifier evaluated in the internal independent test cohort. **B** Confusion matrix. **C** Variable importance plot. **D** SHAP violin plot. **E** Differential abundance comparison of key microbial biomarkers between non-recurrent CRC and postoperative recurrence groups. FDR-adjusted *P*-values (*q* values) were applied for differential abundance analysis

model interpretability, SHAP analysis was applied to quantify feature-level contributions to individual predictions. These genera were ranked by mean absolute SHAP values and visualized in descending order in Fig. 5D. Positive SHAP values were associated with predictions toward postoperative recurrence, indicating an increased predicted risk. Specifically, higher relative abundances of *Peptostreptococcus* and *Fusobacterium* positively shifted model outputs toward the recurrence class, whereas reduced abundance of *Bacteroides* similarly increased the predicted recurrence risk. Collectively, these patterns identify a recurrence-associated microbial signature with consistent directionality across feature importance and model interpretation analysis. Finally, differential abundance analysis of these key predictive genera (Fig. 5E) demonstrated that *Fusobacterium* ($q = 0.006$) and *Peptostreptococcus* ($q = 0.02$) exhibited 3.0- and 2.28-fold higher relative abundances, respectively, in the recurrence group compared to the CRC group. Together, these two genera accounted for more than 30% of the total mucosal microbial abundance in recurrent patients, underscoring their potential utility as prognostic microbial features.

Optimization of CRC prognostic models via integrated metabolite signatures

Building upon the distinctive tumor mucosal metabolomic landscape associated with postoperative CRC recurrence, the 46 differential metabolites with VIP score > 1 were subjected to LASSO-based feature selection. This approach identified five key metabolite predictors (selection frequency > 0.3). Subsequent RF modeling using these selected metabolites achieved an AUC of 0.66 (Fig. 6A). As shown in Fig. S6, alanylglutamic acid, arginine, and putrescine exhibited the highest relative VIMP values, each exceeding 20%. These predictors were subsequently ranked by mean absolute SHAP values and visualized in descending order (Fig. 6B). Likewise, the relative abundance plots comparing the CRC and recurrence groups showed significant differences in arginine ($q = 0.038$) and putrescine ($q = 0.053$). Specifically, arginine, histidine, and sebacic acid were depleted in the recurrence group, whereas alanylglutamic acid and putrescine were enriched (Fig. 6C, D). These metabolites have previously been implicated in CRC development and progression [40–44]. Given the established interplay between the microbiome and host metabolism, we integrated the selected metabolite predictors with five key microbial genera to construct a combined multi-omics RF classifier (Fig. 6E). This integrative model demonstrated improved discriminative performance compared with microbiome-only models (AUC 0.89 vs 0.81). Performance metrics of the integrated model (Fig. 6F) revealed a 31.3% increase in specificity (from 0.67 to 0.88) accompanied by a 14.1% improvement in overall accuracy (from 0.71 to 0.81).

Moreover, an integrated microbiome-metabolite risk score was derived via the RSF framework, and 115 CRC patients were stratified into low- and high-risk groups based on the median predicted risk. Kaplan–Meier survival analysis demonstrated that RFS in the high-risk group was significantly shorter. Notably, this risk score remained an independent predictor of RFS in the multivariate Cox regression model (adjusted hazard ratio (HR) = 1.59, 95% CI: 1.35–1.88, $P < 0.0001$, Fig. 6G) after adjustment for TNM stage. Collectively, these results indicate that integrating microbial and metabolic signatures within a machine learning framework substantially enhances prognostic stratification and prediction of postoperative CRC recurrence.

Co-aggregation properties of F. nucleatum and P. anaerobius

Prior studies have revealed that *F. nucleatum* and *P. anaerobius* could exacerbate CRC initiation and progression, with *F. nucleatum* specifically implicated in promoting metastatic dissemination [38, 45, 46]. Notably, mucosal biofilms in CRC patients have been associated with procarcinogenic epithelial states, highlighting bacterial biofilm formation as a key epithelial-associated event in CRC pathogenesis [11, 47]. Moreover, host-microbiota polyamine metabolism has been reported to facilitate biofilm development and epithelial proliferation, thereby accelerating neoplastic transformation of colonic epithelial cells [48]. Given the limited understanding of how altered arginine metabolism within the tumor mucosa influences dual-species biofilm formation and tumor-adhesive properties in polymicrobial contexts, we first

(See figure on next page.)

Fig. 6 Optimization of prognostic prediction using integrated metabolomic and microbiome features. **A** ROC curve of the metabolomic-based RF predictive model evaluated in the internal dataset. **B** SHAP violin plot. **C**, **D** Differential abundance comparison of key metabolite biomarkers between non-recurrent CRC and postoperative recurrence groups. Differential abundance of metabolite biomarkers was assessed using FDR-adjusted *P*-values (*q* values). **E** ROC curve of the integrated microbiome–metabolomic RF predictive model evaluated in the internal dataset. **F** Confusion matrix. **G** Kaplan–Meier curves showing the recurrence-free survival of CRC patients ($n = 115$) stratified by prognostic risk scores. *P* values for Kaplan–Meier analysis were calculated using the log-rank test

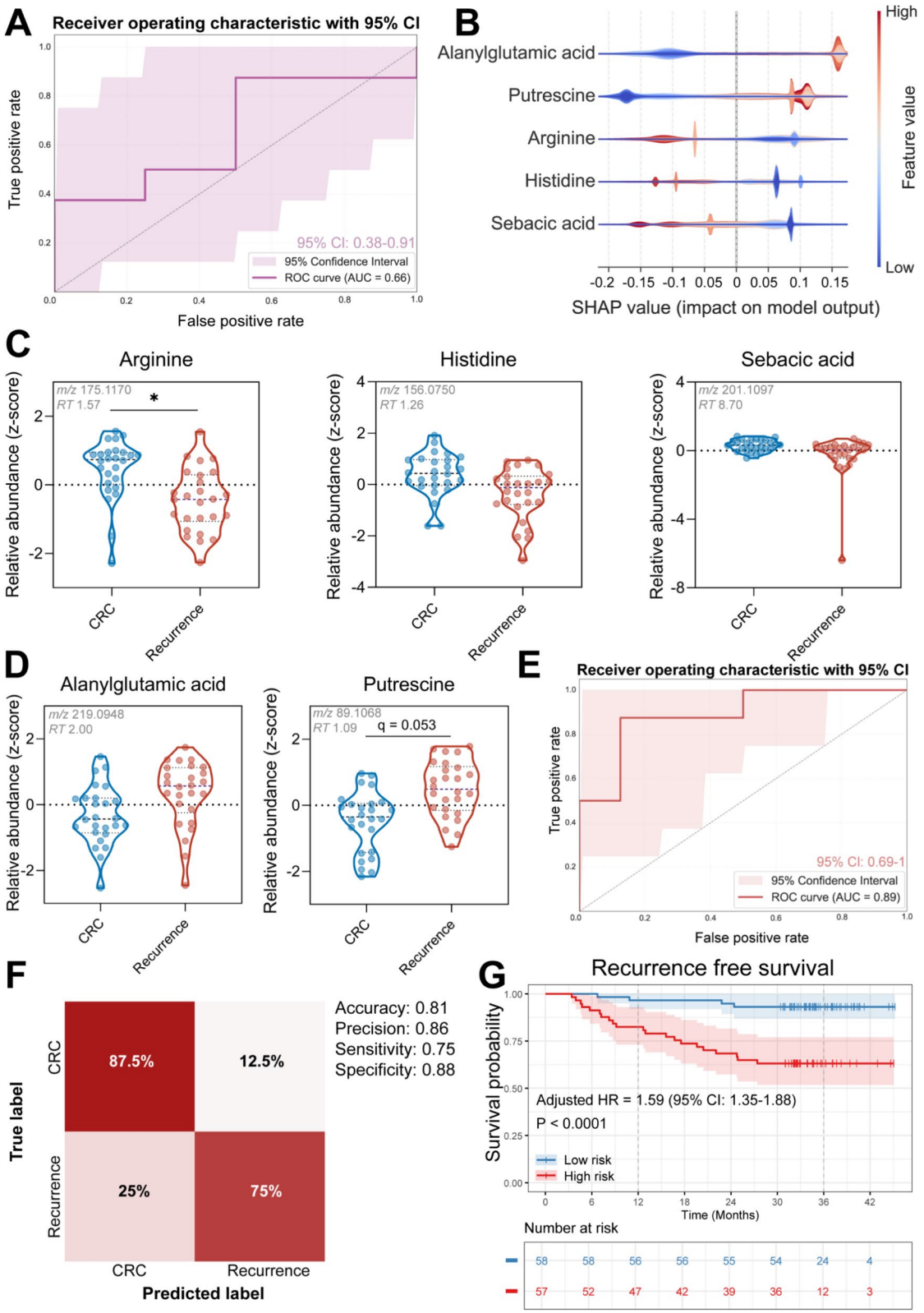

Fig. 6 (See legend on previous page.)

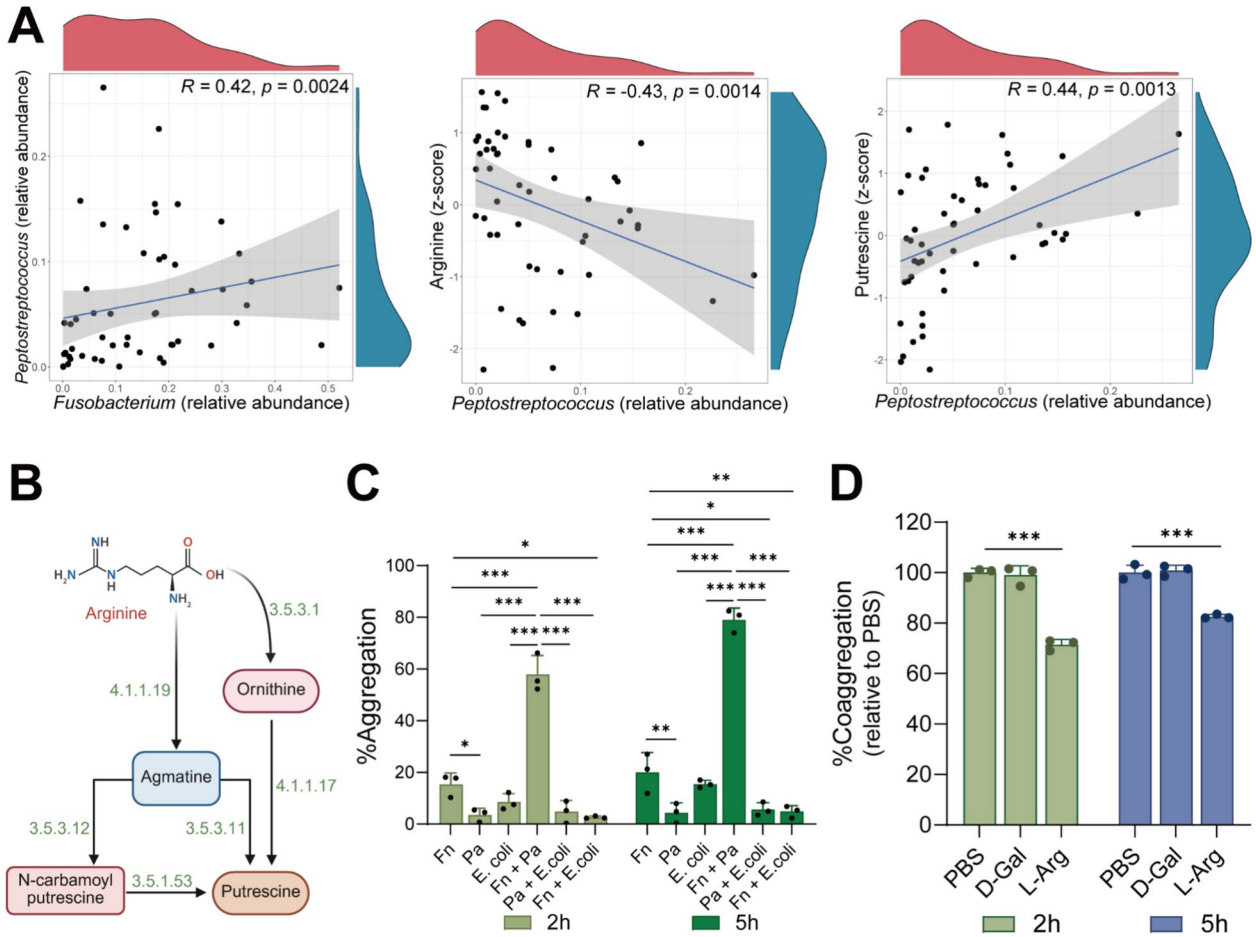

Fig. 7 Interbacterial co-aggregation characteristics. **A** Spearman correlation analysis. **B** Schematic representation of the arginine metabolism pathway yielding putrescine. **C** Quantification of interbacterial co-aggregation. **D** Effects of L-arginine or D-galactose on the co-aggregation at 2 or 5 h of F. nucleatum (Fn) and P. anaerobius (Pa) compared with PBS control

conducted exploratory correlation analysis to identify potential associations among microbial taxa and metabolites (Fig. 7A). A significant positive correlation was observed between the relative abundances of co-enriched *Fusobacterium* and *Peptostreptococcus* ($R=0.42$, $P=0.0024$), while *Peptostreptococcus* exhibited a negative correlation with arginine ($R=-0.43$, $P=0.0014$) and a positive correlation with putrescine ($R=0.44$, $P=0.0013$). These findings suggest a potential link between arginine-derived polyamine metabolism and the coordinated enrichment of these taxa. Subsequent interrogation of arginine metabolism in the KEGG database uncovered two convergent biosynthetic pathways for putrescine generation (Fig. 7B). The canonical route involves sequential conversion of arginine to ornithine by arginase (EC 3.5.3.1) followed by ornithine decarboxylase (ODC, EC 4.1.1.17)-mediated production of putrescine, while an alternative pathway proceeds through arginine decarboxylase (EC 4.1.1.19)-catalyzed formation of agmatine and its subsequent hydrolysis to putrescine by agmatinase (EC 3.5.3.11).

Microbial surface adhesion and interspecies co-aggregation are fundamental biological processes governing polymicrobial biofilm formation and structural organization [49]. To functionally assess whether *F. nucleatum* and *P. anaerobius* are capable of direct physical association, bacterial strains were co-incubated in PBS for 2 or 5 h, and aggregation was monitored via OD_{600} measurements. Compared to control groups (*F. nucleatum*, *P. anaerobius*, or *E. coli* incubated individually), co-aggregation rates were calculated based on the reduction in OD_{600} after 2 or 5 h of anaerobic incubation at 37 °C. While *F. nucleatum* exhibited a self-aggregation rate of 15.4%, *P. anaerobius* (3.6%) and *E. coli* (8.5%) showed minimal self-aggregation. Notably, co-incubation of *F. nucleatum* and *P. anaerobius* for 2 h resulted in pronounced aggregation (Fig. S7), with 58.0% of cells forming aggregates (Fig. 7C). In contrast, combinations of *F. nucleatum*–*E. coli* and *P.*

anaerobius–E. coli displayed negligible co-aggregation (< 5.7%) even after 5 h, whereas *F. nucleatum–P. anaerobius* co-incubation reached 79.1% aggregation. *F. nucleatum* expresses multiple adhesion-associated proteins, including Fap2 (a Gal-GalNAc-binding lectin) and RadD (an arginine-dependent adhesin) [26, 46]. To explore the potential involvement of these adhesins in interspecies co-aggregation, PBS (control), D-galactose (Fap2 inhibitor), or L-arginine (RadD inhibitor) were added to the *F. nucleatum–P. anaerobius* co-culture system. While neither PBS nor D-galactose affected aggregation (Fig. 7D), L-arginine markedly reduced co-aggregation at both 2 h and 5 h ($P < 0.001$), supporting a central role for RadD-mediated adhesion in *F. nucleatum–P. anaerobius* intermicrobial interactions.

Effect of putrescine on F. nucleatum and P. anaerobius dual-species biofilms

Biofilm formation initiates through bacterial aggregation, which represents a critical determinant of microbial colonization. Building upon our observation of robust co-aggregation between *F. nucleatum* and *P. anaerobius*, we subsequently investigated whether *F. nucleatum* influences *P. anaerobius* biofilm development. Fluorescent tracer assays revealed that while monocultures of *P. anaerobius* failed to establish substantial biofilms after 24 h (Fig. 8A), co-culture with *F. nucleatum* promoted the formation of structured dual-species biofilms. Crystal violet quantification further demonstrated a synergistic enhancement of biofilm biomass in *F. nucleatum–P. anaerobius* co-cultures compared to *F. nucleatum* monocultures (Fig. 8B, $P < 0.001$). qPCR analysis of mature

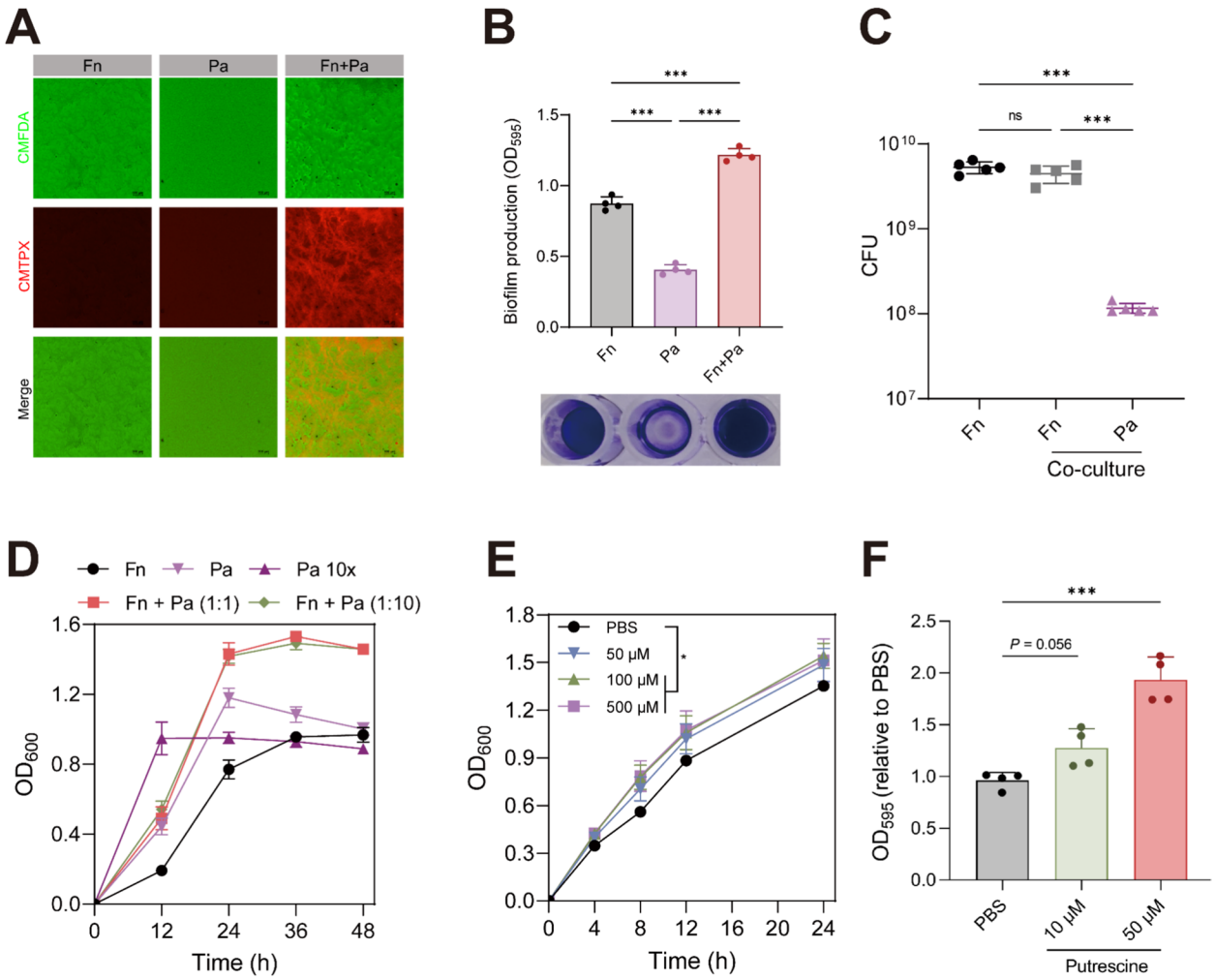

Fig. 8 Formation of F. nucleatum–P. anaerobius dual-species biofilms and the impact of putrescine. **A** Fluorescence micrograph depicting bacterial biofilm formation. **B** Crystal violet-stained image of bacterial biofilms and quantification of adherent biomass. **C** Quantification of Fn and Pa within dual-species biofilms by qPCR. **D** Growth curves of Fn and Pa inoculated alone and co-inoculated in different proportions. **E** Effect of putrescine on growth capacity in Fn:Pa = 1:1 inoculation. **F** The promotion of dual-species biofilm formation by different concentrations of putrescine

biofilms identified F. nucleatum as the dominant constituent (Fig. 8C). Growth kinetic analysis demonstrated higher culture densities in 1:1 *F. nucleatum–P. anaerobius* co-cultures compared with monocultures over 48 h of incubation. Notably, this increase in growth was not solely attributable to *P. anaerobius*, as a 10-fold increase in *P. anaerobius* inoculum did not result in a comparable elevation in culture density (Fig. 8D).

Prior studies have documented that polyamine-mediated metabolic cross-feeding among oral pathogens (*Streptococcus gordonii*, *F. nucleatum*, and *Porphyromonas gingivalis*) drives biofilm expansion [50]. To evaluate whether putrescine modulates growth and biofilm formation in the *F. nucleatum–P. anaerobius* dual-species system, we first assessed bacterial growth responses to putrescine supplementation at concentrations ranging from 50 to 500 μM. Supplementation with 100 and 500 μM putrescine significantly increased OD_{600} of *F. nucleatum–P. anaerobius* co-cultures at 24 h (Fig. 8E, $P<0.05$ compared with PBS-treated control). Subsequent biofilm assays using sub-growth-stimulatory concentrations of putrescine (10–50 μM) revealed a concentration-dependent enhancement of biofilm formation (Fig. 8F, at least $P\leq0.056$). These findings reveal that *F. nucleatum* serves as a keystone species in *F. nucleatum* and *P. anaerobius* dual-species biofilm formation, while arginine metabolite putrescine potentiates biofilm development in a concentration-dependent manner.

Promotive effect of P. anaerobius on F. nucleatum adhesion to CRC cells

Bacterial aggregation in biofilms enhances tissue adherence and facilitates tumor invasion by pathogenic bacteria, thereby potentiating carcinogenesis and metastatic progression [38, 45, 46, 51]. Confocal microscopy revealed that both *F. nucleatum* (green fluorescence) and *P. anaerobius* (red fluorescence) were individually capable of adhering to tumor cell surfaces. Strikingly, when co-incubated with HCT116 cells, *P. anaerobius* predominantly adhered to *F. nucleatum* rather than directly binding to host cells, forming bacterial consortia that collectively anchored to tumor surfaces (Fig. 9A). Quantitative fluorescence assays confirmed that *P. anaerobius* significantly augmented the adhesion capacity of *F. nucleatum* to tumor cells ($P<0.05$), while co-culture resulted in only a modest and statistically non-significant increase in *P. anaerobius* adhesion ($P>0.05$, Fig. 9B). Taken together, this hierarchical adhesion pattern suggests that *F. nucleatum* may serve as a bridging organism, facilitating the incorporation of *P. anaerobius* into the tumor-associated biofilm consortium.

P. anaerobius enhances intestinal colonization and mucosal adherence of F. nucleatum in vivo

Biofilm-mediated bacterial growth establishes favorable niches near the intestinal epithelial barrier [47]. To evaluate in vivo effects of *P. anaerobius* on *F. nucleatum* mucosal colonization, mice were administered equivalent CFUs of *F. nucleatum* alone, *P. anaerobius* alone, or *F. nucleatum–P. anaerobius* co-culture (experimental design: Fig. 10A). Initial fecal analysis revealed comparable *F. nucleatum* abundance between co-culture and *F. nucleatum*-only groups at day 1 post-gavage. However, longitudinal monitoring showed a progressive increase in fecal *F. nucleatum* abundance in the co-culture group, resulting in significantly higher levels compared with the *F. nucleatum*-only group at day 9 (two-tailed Student's *t*-test, $P<0.01$), whereas *F. nucleatum* colonization in the mono-colonized group remained stable throughout the 9-day observation period. Conversely, fecal *P. anaerobius* became undetectable (below detection limit) in both *P. anaerobius*-only and co-culture groups by day 3 (Fig. 10B). Critically, administration of the *F. nucleatum–P. anaerobius* co-culture resulted in enhanced colonization of *F. nucleatum* within the colonic mucosa compared with *F. nucleatum* monoculture ($P=0.051$), suggesting that *P. anaerobius* facilitates mucosal adherence and invasion of *F. nucleatum* despite its transient persistence in fecal samples (Fig. 10C). Collectively, these in vivo findings extend our in vitro observations, supporting a facilitative role of *P. anaerobius* in modulating *F. nucleatum* colonization dynamics within the intestinal mucosal niche.

Discussion

Despite the curative-intent surgery, postoperative recurrence remains a major clinical challenge in CRC, suggesting that biological heterogeneity beyond conventional clinicopathological factors may contribute to disease progression. Although TNM staging remains the standard framework for recurrence risk assessment, substantial heterogeneity in clinical outcomes is frequently observed among patients within the same stage, highlighting the need for additional biomarkers to improve risk stratification [52–54]. In this prospective cohort, integrated microbiome and metabolome profiling of tumor mucosal samples, combined with longitudinal follow-up, revealed microbial and metabolic features associated with postoperative recurrence in CRC. Primary tumors from recurrent patients exhibited distinct microbial and metabolic signatures, supporting the potential utility of tumor-associated microbiota and metabolites as prognostic indicators.

Using RF modeling and SHAP analysis, we identified several bacterial genera associated with CRC recurrence,

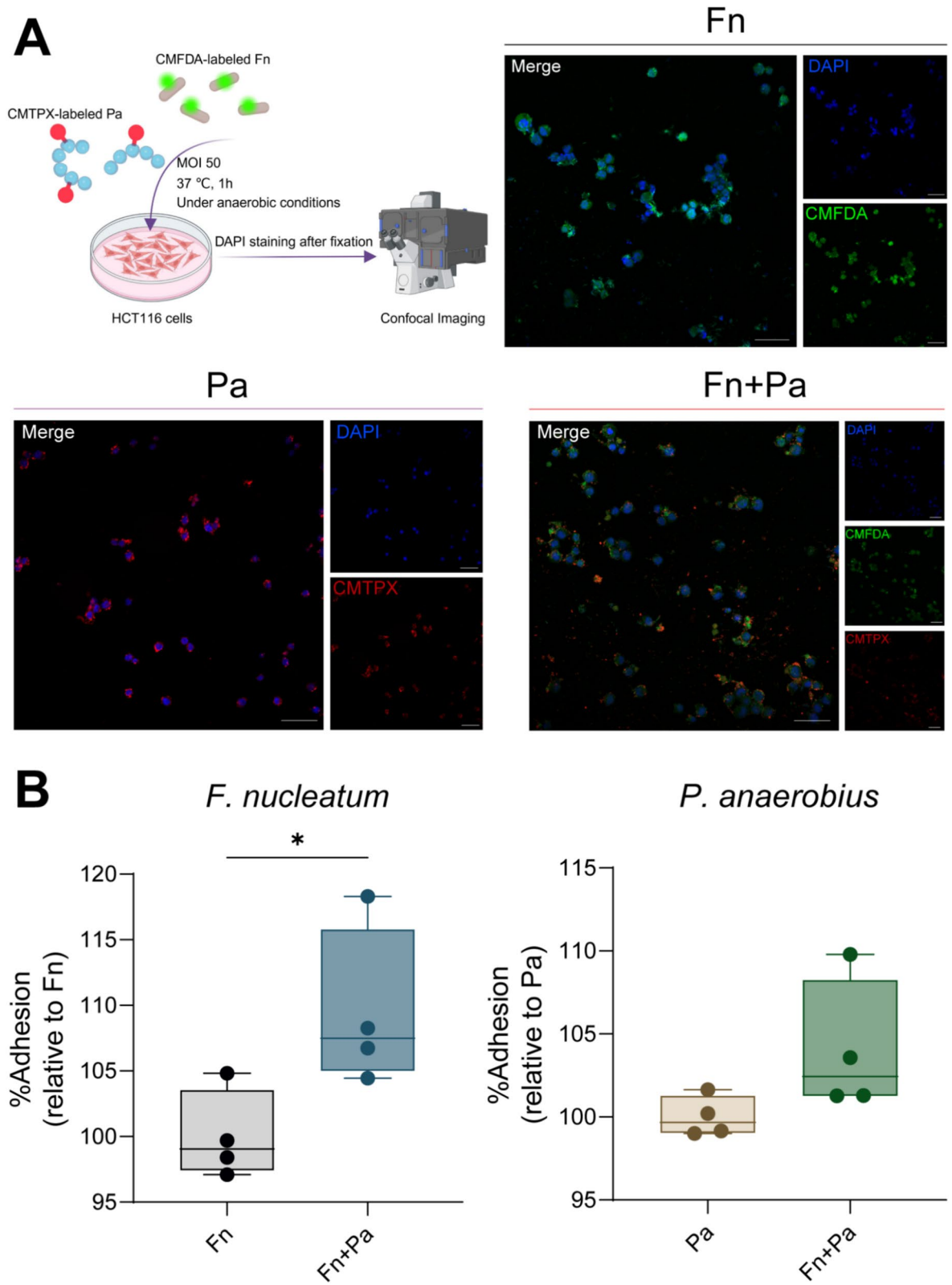

Fig. 9 Effect of P. anaerobius on the tumor cell adhesion capacity of F. nucleatum. **A** Representative confocal images showing CMFDA-labelled Fn (green) and CMTPX-labelled Pa (red) binding to DAPI-stained (blue) HCT116 cells. **B** Fluorescence microplate reader analysis of bacterial attachment

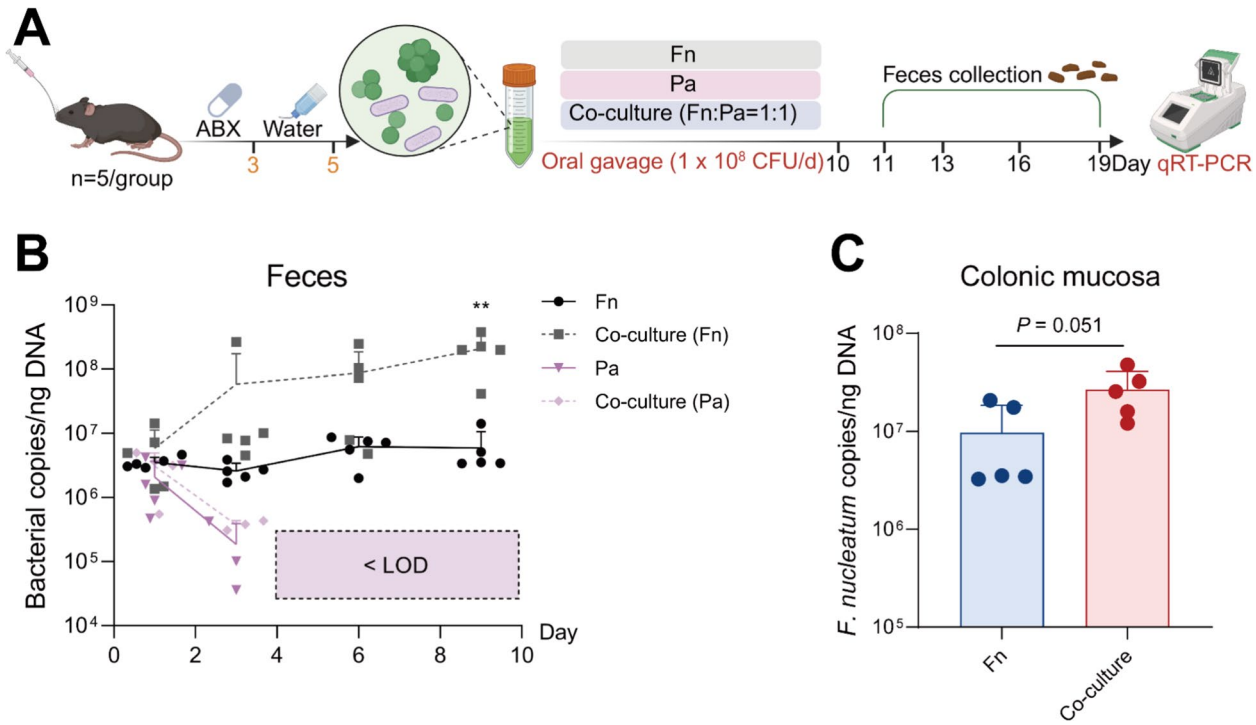

Fig. 10 Effect of P. anaerobius on the intestinal colonization capacity of F. nucleatum in mice. **A** Schematic diagram of the experimental design. **B**–**C** Quantification of bacterial genomic DNA copies from fecal samples collected during the experimental period (**B**) and from colonic mucosa at the endpoint (**C**). Data is presented as mean ± SD

including *Peptostreptococcus*, *Fusobacterium*, *Bacteroides*, *Porphyromonas*, and *Prevotella*. These taxa have been previously implicated in CRC initiation, progression, and adverse clinical outcomes, supporting their potential value as prognostic biomarkers [16, 18, 55–57]. Among them, *F. nucleatum* has been extensively linked to poor prognosis, recurrence, and chemoresistance in CRC [13, 15, 18, 58]. Notably, we further observed significant enrichment of *Peptostreptococcus* in the tumor mucosa of recurrent CRC patients, where it contributed positively to recurrence classification in our RF model. Previous studies have shown that *P. anaerobius*, a representative species of this genus, promotes CRC progression by adhering to tumor tissues, activating oncogenic and inflammatory signaling pathways, and modulating antitumor immunity [38, 59–61]. However, synergistic associations between *F. nucleatum* and co-enriched pathobionts within the tumor microenvironment remain insufficiently explored.

A key finding of this study is the significant positive correlation between *Fusobacterium* and *Peptostreptococcus*, the predominant genera enriched in the tumor mucosa of CRC patients with postoperative recurrence, suggesting a synergistic relationship within the tumor-associated microbial community. Polymicrobial biofilms represent a prominent higher-order spatial organization in the gut and have been implicated in promoting pro-inflammatory and carcinogenic effects in CRC [47, 48, 51]. Previous studies have shown that tumor-associated pathobionts, including *F. nucleatum*, have been implicated in facilitating distant metastasis through disruption of the gut vascular barrier and enhanced bacterial dissemination. In addition, *F. nucleatum* adheres to CRC cells via its adhesins Fap2 and RadD, and higher RadD expression has been linked to an increased risk of disease recurrence [25, 26, 46]. Expanding on these observations, we demonstrate that *P. anaerobius* co-aggregates with *F. nucleatum* to form compact dual-species biofilms, significantly enhancing *F. nucleatum* adhesion to tumor cells and intestinal colonization. This effect seems to be partially dependent on RadD-mediated interactions and suggests a potential role for *P. anaerobius* in modulating *F. nucleatum*-associated biofilm formation. These findings suggest that RadD-mediated polymicrobial interactions may represent a potential mechanistic target for disrupting recurrence-associated bacterial colonization.

Metabolic reprogramming is a hallmark of CRC progression, and metabolomics provides a powerful

approach for identifying prognostic biomarkers within the tumor microenvironment [62–64]. In this study, untargeted metabolomics revealed significant enrichment of polyamines and related metabolites (cadaverine, putrescine, and monoacetylcadaverine) in the tumor mucosa of recurrent CRC patients. The correlation analysis demonstrated a strong positive association between putrescine and *Peptostreptococcus*, and both features were retained as predictors in the RSF model, correlating with poorer postoperative outcomes. Notably, we found that putrescine supplementation significantly enhances dual-species biofilm formation by *P. anaerobius* and *F. nucleatum*, suggesting that tumor-associated metabolites may directly modulate the interaction of pathogenic bacteria. Previous studies have also documented elevated fecal putrescine levels in CRC patients [65], while ODC-dependent putrescine production has been implicated in chemotherapy resistance mechanisms [44]. Prior research has established that acetylated polyamine metabolites influence CRC-associated biofilm development [48], though their species-specific impacts on biofilm assembly remain uncharacterized. Putrescine genesis in the gut involves dual origins: host-cell biosynthesis through ODC-catalyzed ornithine decarboxylation, and microbial metabolism where arginine is converted via diverse bacterial enzymes [66, 67]. Although the relative contributions of host versus microbial arginine metabolism to putrescine production remain unclear, these findings reveal a novel link between metabolic alterations and recurrence-associated microbial synergy.

Previous studies have demonstrated the prognostic potential of the gut microbiome and metabolome for predicting recurrence and survival outcomes following CRC surgery [16, 57, 62, 63]. However, most investigations relied on single-omics profiling, limiting insight into the coordinated microbial–metabolic interactions within the tumor microenvironment. Given the complexity of microbial community dysregulation and metabolic reprogramming in CRC, integrated multi-omics analysis is necessary to better elucidate mechanisms underlying postoperative recurrence. Here, integrated analysis of the tumor mucosal microbiome and metabolome revealed recurrence-associated signatures that were partially independent of tumor stage, suggesting complementary value to conventional staging.

Machine learning has been widely applied in CRC research, spanning applications from diagnosis to prognosis [31]. Using LASSO-based feature selection and subsequent random forest modeling, we identified a panel of 10 integrated microbial and metabolic biomarkers associated with recurrence status in CRC patients. Notably, the model captured predictive taxa such as *Porphyromonas* and *Prevotella*, which exhibited only moderate abundance differences but contributed meaningfully to recurrence classification, suggesting that machine learning can uncover latent, non-linear microbial associations not detected by conventional differential analysis [34].

Currently, circulating tumor DNA (ctDNA) represents a clinically relevant biomarker for detecting molecular residual disease and predicting CRC recurrence [68–70]. In this study, the integrated RSF-derived risk score effectively stratified 115 CRC patients, with the high-risk group exhibiting significantly shorter RFS. Multivariate analysis confirmed the risk score as an independent prognostic indicator (adjusted HR = 1.59, 95% CI 1.35–1.88, $P < 0.0001$), supporting the prognostic relevance of tumor mucosal multi-omics signatures and highlighting their potential complementarity to existing molecular approaches. However, this study has several limitations. Although patients received relatively uniform adjuvant chemotherapy, the detailed patient-level treatment information was not available, which limited our ability to model chemotherapy exposure as a potential confounder. In addition, the predictive model was derived from a relatively small cohort and lacked validation in large independent cohorts with long-term follow-up. Future studies incorporating larger cohorts, comprehensive treatment annotation, and longitudinal validation are warranted to enhance robustness and generalizability.

Nevertheless, this work identifies discriminant microbial and metabolic features distinguishing primary from recurrent CRC, revealing synergistic interactions between *F. nucleatum* and the newly established prognostic genus *Peptostreptococcus*. These findings provide an analytical and mechanistic framework for refining postoperative risk stratification following CRC surgery.

Conclusions

Machine learning-driven integration of the tumor mucosal microbiome and metabolome in CRC patients was performed to predict postoperative recurrence risk. Furthermore, interaction studies of prognostic biomarkers have revealed that *P. anaerobius* could enhance the tumor adhesion capacity of *F. nucleatum*, alongside the promotion of dual-species biofilm formation by putrescine. These findings will facilitate the prediction of postoperative recurrence outcomes in CRC, while the distinct arginine metabolism-microbiota correlation may offer novel strategies for therapeutic intervention.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40168-026-02378-w.

Supplementary Material 1: Table S1. Demographic, clinical, and histological data of CRC cohort. Table S2. Clinical characteristics of matched CRC patients for recurrence prediction. Figure S1. Datasets used in statistical and machine learning analyses. Figure S2. Microbial community structure and composition of tumor mucosa across different TNM stages. Figure S3. Mfuzz clustering of genera across TNM stages. Figure S4. Metabolomic landscape across different TNM stages. Figure S5. Spearman correlation analysis. Figure S6. Feature importance ranking. Figure S7. Images of bacterial aggregation.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements
The authors express sincere gratitude to all participating donors and collaborating colleagues from Tianjin Union Medical Center for their invaluable contributions.

Authors' contributions
SW and CZZ conceived and designed the study. YHZ, BWZ, WWP, and CZZ collected clinical samples and organized the data. YHZ performed experiments. YHZ and BWZ interpreted and analyzed the data. YHZ, BWZ, and WWP drafted and revised the manuscript. WTG, XTW, HY, SYF, JZ, XL, XLX, XMJ, TXL, and FW critically reviewed the manuscript and provided substantive intellectual revisions. All authors have read and approved the final manuscript.

Funding
This study was supported by the Beijing-Tianjin-Hebei Basic Research Cooperation Project (23JCZXJC00390-J230038), National Science and Technology Major Project (2024ZD0521200-2024ZD0521203), Tianjin Health Technology Project (TJWJ2023QN040), Natural Science Foundation of Tianjin (No. 25JCYBJC01130), Tianjin Municipal Public Health Program (25ZXWQSY00010), and National Natural Science Foundation of China (82473806). The authors express sincere gratitude to all participating donors and collaborating colleagues from Tianjin Union Medical Center for their invaluable contributions.

Data availability
The 16S rRNA gene sequencing data generated in this study have been deposited in the NCBI Sequence Read Archive and are publicly available at (https:/www.ncbi.nlm.nih.gov/sra/PRJNA1309319) (accession number: PRJNA1309319). Metabolomics data have been deposited in the National Genomics Data Center (NGDC) and are publicly available at (https:/ngdc.cncb.ac.cn/omix/release/OMIX013927) (accession number: OMIX013927). Code in this study is available at https://github.com/hui-bio/ML_Analysis.

Declarations

Ethics approval and consent to participate
Ethical approval was obtained from the Ethics Committee of Tianjin Union Medical Center (2021-B37), 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
[1]Tianjin Key Laboratory of Food Science and Health, School of Medicine, Nankai University, Tianjin 300071, China. [2]Department of Clinical Laboratory, Tianjin Union Medical Center, Nankai University, Tianjin 300121, China. [3]School of Medicine, Nankai University, Tianjin 300071, China. [4]Department of Colorectal Surgery, Tianjin Union Medical Center, The First Affiliated Hospital of Nankai University, Tianjin 300121, China. [5]Tianjin Institute of Coloproctology, Tianjin 300121, China.

Received: 27 August 2025 Accepted: 9 February 2026
Published online: 18 March 2026

References

1. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin. 2020;70:145–64.
2. Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. CA Cancer J Clin. 2025;75:10–45.
3. Kotani D, Oki E, Nakamura Y, Yukami H, Mishima S, Bando H, et al. Molecular residual disease and efficacy of adjuvant chemotherapy in patients with colorectal cancer. Nat Med. 2023;29:127–34.
4. Andre T, Boni C, Navarro M, Tabernero J, Hickish T, Topham C, et al. Improved overall survival with oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment in stage II or III colon cancer in the MOSAIC trial. J Clin Oncol. 2009;27:3109–16.
5. Midgley R, Kerr D. Colorectal cancer. Lancet. 1999;353:391–9.
6. Yang L, Yang J, Kleppe A, Danielsen HE, Kerr DJ. Personalizing adjuvant therapy for patients with colorectal cancer. Nat Rev Clin Oncol. 2024;21:67–79.
7. Klaver CEL, Bulkmans N, Drillenburg P, Grabsch HI, van Grieken NCT, Karrenbeld A, et al. Interobserver, intraobserver, and interlaboratory variability in reporting pT4a colon cancer. Virchows Arch. 2020;476:219–30.
8. Gagnaire A, Nadel B, Raoult D, Neefjes J, Gorvel JP. Collateral damage: insights into bacterial mechanisms that predispose host cells to cancer. Nat Rev Microbiol. 2017;15:109–28.
9. Chen F, Dai X, Zhou CC, Li KX, Zhang YJ, Lou XY, et al. Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma. Gut. 2022;71:1315–25.
10. White MT, Sears CL. The microbial landscape of colorectal cancer. Nat Rev Microbiol. 2024;22:240–54.
11. Dejea CM, Fathi P, Craig JM, Boleij A, Taddese R, Geis AL, Wu XQ, Shields CED, Hechenbleikner EM, Huso DL, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. Science. 2018;359:592-+.
12. Arthur JC, Perez-Chanona E, Muhlbauer M, Tomkovich S, Uronis JM, Fan TJ, et al. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science. 2012;338:120–3.
13. Serna G, Ruiz-Pace F, Hernando 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.
14. 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.
15. 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.
16. Huh JW, Kim MJ, Kim J, Lee HG, Ryoo SB, Ku JL, Jeong SY, Park KJ, Kim D, Kim JF, et al. Enterotypical Prevotella and three novel bacterial biomarkers in preoperative stool predict the clinical outcome of colorectal cancer. Microbiome. 2022;10:203.
17. 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.
18. Yu TC, Guo FF, Yu YN, Sun TT, Ma D, Han JX, et al. *Fusobacterium nucleatum* promotes chemoresistance to colorectal cancer by modulating autophagy. Cell. 2017;170:548.
19. Kim M, Vogtmann E, Ahlquist DA, Devens ME, Kisiel JB, Taylor WR, White BA, Hale VL, Sung J, Chia N, et al. Fecal metabolomic signatures in colorectal adenoma patients are associated with gut microbiota and early events of colorectal cancer pathogenesis. mBio. 2020;11:e03186–19.
20. 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;72:1129–42.

21. Sun Y, Zhang X, Hang D, Lau HC, Du J, Liu C, et al. Integrative plasma and fecal metabolomics identify functional metabolites in adenoma-colorectal cancer progression and as early diagnostic biomarkers. Cancer Cell. 2024;42:1386–1400 e8.
22. Farshidfar F, Weljie AM, Kopciuk KA, Hilsden R, McGregor SE, Buie WD, et al. A validated metabolomic signature for colorectal cancer: exploration of the clinical value of metabolomics. Br J Cancer. 2016;115:848–57.
23. Fu T, Coulter S, Yoshihara E, Oh TG, Fang S, Cayabyab F, et al. FXR regulates intestinal cancer stem cell proliferation. Cell. 2019;176:1098-1112 e18.
24. Donohoe DR, Holley D, Collins LB, Montgomery SA, Whitmore AC, Hillhouse A, et al. A gnotobiotic mouse model demonstrates that dietary fiber protects against colorectal tumorigenesis in a microbiota- and butyrate-dependent manner. Cancer Discov. 2014;4:1387–97.
25. Xie Y, Xie F, Zhou X, Zhang L, Yang B, Huang J, et al. Microbiota in tumors: from understanding to application. Adv Sci. 2022;9:e2200470.
26. 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.
27. Tjalsma H, Boleij A, Marchesi JR, Dutilh BE. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol. 2012;10:575–82.
28. Li S, Konstantinov SR, Smits R, Peppelenbosch MP. Bacterial biofilms in colorectal cancer initiation and progression. Trends Mol Med. 2017;23:18–30.
29. Cammarota G, Ianiro G, Ahern A, Carbone C, Temko A, Claesson MJ, et al. Gut microbiome, big data and machine learning to promote precision medicine for cancer. Nat Rev Gastroenterol Hepatol. 2020;17:635–48.
30. Liu Z, Liu L, Weng S, Guo C, Dang Q, Xu H, et al. Machine learning-based integration develops an immune-derived lncRNA signature for improving outcomes in colorectal cancer. Nat Commun. 2022;13:816.
31. Tsai PC, Lee TH, Kuo KC, Su FY, Lee TM, Marostica E, et al. Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients. Nat Commun. 2023;14:2102.
32. Wishart DS, Guo AC, Oler E, Wang F, Anjum A, Peters H, et al. HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Res. 2022;50:D622–31.
33. Schwammle V, Jensen ON. A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. Bioinformatics. 2010;26:2841–8.
34. Chen Y, Wang B, Zhao Y, Shao X, Wang M, Ma F, et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. Nat Commun. 2024;15:1657.
35. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. 1996;58:267–88.
36. Lundberg SM and Lee SI. A unified approach to interpreting model predictions. in 31st Annual Conference on Neural Information Processing Systems (NIPS). 2017. Long Beach, CA.
37. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat. 2008;2:841–60.
38. Long XH, Wong CC, Tong L, Chu ESH, Szeto CH, Go MYY, et al. Peptostreptococcus anaerobius promotes colorectal carcinogenesis and modulates tumour immunity. Nat Microbiol. 2019;4:2319–30.
39. Lin H and Das Peddada S. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. Nat Met. 2024;21:83–91.
40. Du T, Han J. Arginine metabolism and its potential in treatment of colorectal cancer. Front Cell Dev Biol. 2021;9:658861.
41. Tan J, Wang HL, Yang J, Liu QQ, Li CM, Wang YQ, et al. JMJD2B-induced amino acid alterations enhance the survival of colorectal cancer cells under glucose-deprivation via autophagy. Theranostics. 2020;10:5763–77.
42. Brown DG, Rao S, Weir TL, O'Malia J, Bazan M, Brown RJ, et al. Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. Cancer Metab. 2016;4:11.
43. Vidman L, Zheng R, Boden S, Ribbenstedt A, Gunter MJ, Palmqvist R, et al. Untargeted plasma metabolomics and risk of colorectal cancer-an analysis nested within a large-scale prospective cohort. Cancer Metab. 2023;11:17.
44. Zhang X, Chen YJ, Hao LJ, Hou A, Chen XZ, Li YF, et al. Macrophages induce resistance to 5-fluorouracil chemotherapy in colorectal cancer through the release of putrescine. Cancer Lett. 2016;381:305–13.
45. Lu XX, Xu QL, Tong YA, Zhang ZJ, Dun GD, Feng YY, Tang J, Han D, Mao YL, Deng L, et al. Long non-coding RNA EVADR induced by Fusobacterium nucleatum infection promotes colorectal cancer metastasis. Cell Rep. 2022;40:111127.
46. Zhang L, Leng XX, Qi JX, Wang N, Han JX, Tao ZH, Zhuang ZY, Ren YM, Xie YL, Jiang SS, et al. The adhesin RadD enhances Fusobacterium nucleatum tumour colonization and colorectal carcinogenesis. Nat Microbiol. 2024;9:2292–307.
47. Dejea CM, Wick EC, Hechenbleikner EM, White JR, Welch JLM, Rossetti BJ, et al. Microbiota organization is a distinct feature of proximal colorectal cancers. Proc Natl Acad Sci USA. 2014;111:18321–6.
48. Johnson CH, Dejea CM, Edler D, Hoang LT, Santidrian AF, Felding BH, et al. Metabolism Links Bacterial Biofilms and Colon Carcinogenesis. Cell Metab. 2015;21:891–7.
49. He XS, Hu W, Kaplan CW, Guo LH, Shi WY, Lux R. Adherence to streptococci facilitates *Fusobacterium nucleatum* integration into an oral microbial community. Microb Ecol. 2012;63:532–42.
50. Sakanaka A, Kuboniwa M, Shimma S, Alghamdi SA, Mayumi S, Lamont RJ, Fukusaki E, and Amano A. Fusobacterium nucleatum metabolically integrates commensals and pathogens in oral biofilms. Msystems. 2022;7:e0017022.
51. Pedrogo DAM, Sears CL, Melia JMP. Colorectal cancer in inflammatory bowel disease: a review of the role of gut microbiota and bacterial biofilms in disease pathogenesis. J Crohns Colitis. 2024;18:1713–25.
52. Pages F, Mlecnik B, Marliot F, Bindea G, Ou FS, Bifulco C, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. Lancet. 2018;391:2128–39.
53. Kong J, Kim J, Kim D, Lee K, Lee J, Han SK, et al. Information about immune cell proportions and tumor stage improves the prediction of recurrence in patients with colorectal cancer. Patterns (N Y). 2023;4:100736.
54. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, et al. Towards the introduction of the "Immunoscore" in the classification of malignant tumours. J Pathol. 2014;232:199–209.
55. Wang TT, Cai GX, Qiu YP, Fei N, Zhang MH, Pang XY, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J. 2012;6:320–9.
56. Flemer B, Lynch DB, Brown JMR, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut. 2017;66:633–43.
57. Byrd DA, Damerell V, Morales MFG, Hogue SR, Lin TD, Ose J, et al. The gut microbiome is associated with disease-free survival in stage I-III colorectal cancer patients. Int J Cancer. 2025;157:64–73.
58. Yan XB, Liu LG, Li H, Qin HL, Sun ZL. Clinical significance of *Fusobacterium nucleatum*, epithelial-mesenchymal transition, and cancer stem cell markers in stage III/IV colorectal cancer patients. Oncotargets Ther. 2017;10:5031–46.
59. Liu YL, Wong CC, Ding YQ, Gao MX, Wen J, Lau HCH, Cheung AHK, Huang D, Huang H, and Yu J. Peptostreptococcus anaerobius mediates anti-PD1 therapy resistance and exacerbates colorectal cancer via myeloid-derived suppressor cells in mice. Nat Microbiol. 2024;9:1467–82.
60. Tito RY, Verbandt S, Vazquez MA, Lahti L, Verspecht C, Lloréns-Rico V, et al. Microbiome confounders and quantitative profiling challenge predicted microbial targets in colorectal cancer development. Nat Med. 2024;30:1339–48.
61. Cui WW, Guo M, Liu D, Xiao P, Yang CC, Huang HD, Liang CH, Yang YH, Fu XL, Zhang YD, et al. Gut microbial metabolite facilitates colorectal cancer development via ferroptosis inhibition. Nat Cell Biol. 2024;26:124-+.
62. Ose J, Gigic B, Brezina S, Lin T, Baierl A, Geijsen A, van Roekel E, Robinot N, Gicquiau A, Achaintre D, et al. Targeted plasma metabolic profiles and risk of recurrence in stage II and III colorectal cancer patients: results from an international cohort consortium. Metabolites. 2021;11:129.
63. Zhang F, Zhang Y, Zhao W, Deng K, Wang Z, Yang C, et al. Metabolomics for biomarker discovery in the diagnosis, prognosis, survival and recurrence of colorectal cancer: a systematic review. Oncotarget. 2017;8:35460–72.
64. Sedlak JC, Yilmaz OH, Roper J. Metabolism and colorectal cancer. Annu Rev Pathol. 2023;18:467–92.
65. Yang Y, Misra BB, Liang L, Bi D, Weng W, Wu W, et al. Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. Theranostics. 2019;9:4101–14.
66. Nakamura A, Ooga T, Matsumoto M. Intestinal luminal putrescine is produced by collective biosynthetic pathways of the commensal microbiome. Gut Microbes. 2019;10:159–71.
67. Nakamura A, Kurihara S, Takahashi D, Ohashi W, Nakamura Y, Kimura S, Onuki M, Kume A, Sasazawa Y, Furusawa Y, et al. Symbiotic polyamine metabolism

regulates epithelial proliferation and macrophage differentiation in the colon. Nat Commun. 2021;12:2105.
68. Malla M, Loree JM, Kasi PM, Parikh AR. Using circulating tumor DNA in colorectal cancer: current and evolving practices. J Clin Oncol. 2022;40:2846–57.
69. Jin S, Zhu D, Shao F, Chen S, Guo Y, Li K, Wang Y, Ding R, Gao L, Ma W, et al. Efficient detection and post-surgical monitoring of colon cancer with a multi-marker DNA methylation liquid biopsy. Proc Natl Acad Sci U S A. 2021;118:e2017421118.
70. Suzuki T, Suzuki T, Yoshimura Y, Yahata M, Yew PY, Nakamura T, et al. Detection of circulating tumor DNA in patients of operative colorectal and gastric cancers. Oncotarget. 2020;11:3198–207.

Publisher's Note