


RESEARCH

Open Access



# Machine learning-derived diagnostic model of epithelial ovarian cancer based on gut microbiome signatures

Cheng Chen<sup>1†</sup>, Chengyuan Deng<sup>3†</sup>, Yanwen Li<sup>4</sup>, Shuguang He<sup>1</sup>, Yunhong Liu<sup>2</sup>, Shuwen Pan<sup>2</sup>, Wenqian Xu<sup>5</sup>, Lu Fang<sup>6</sup>, Yixi Zhu<sup>1</sup>, Yingying Wang<sup>1</sup> and Xiaoxin Jiang<sup>2\*</sup> 

## Abstract

**Background** Prior studies have elucidated that alterations in gut microbiota are associated with a spectrum of tumors and metabolic disorders. However, the diagnostic value of gut microbiota in epithelial ovarian cancer remains insufficiently investigated.

**Methods** A total of 34 patients with a diagnosis of epithelial ovarian cancer (EOC), 15 patients with benign ovarian tumors (TB), and 30 healthy volunteers (NOR) were enrolled in this study. Fecal samples were collected, followed by sequencing of the V3–V4 region of the 16S rRNA gene. The clinical data and pathological characteristics were comprehensively recorded for further analysis, PICRUSt2 was utilized to conduct an analysis of microbial functional predictions, WGCNA networks were constructed by integrating microbiome and clinical data. LEfSe analysis was employed to identify microbial diagnostic markers, LASSO and SVM analyses were used to screen microbial diagnostic markers in conjunction with the Cally index, to establish a Microbial-Cally diagnostic model. Bootstrap resampling was utilized for the internal validation of the model, whereas the Hosmer–Lemeshow test and decision curve analysis (DCA) were employed to evaluate the diagnostic performance of the model. Plasma samples were subjected to untargeted metabolomics profiling, followed by differential analysis to identify key metabolites that are significantly altered in epithelial ovarian cancer. At the same time, Spearman correlation analysis was used to study the association between key microbiota and differential metabolites. The supernatants from *Escherichia coli* and *Bifidobacterium* cultures were co-cultured with SKOV3 cells. Cell proliferation, migration, and invasion were evaluated using Cell Counting Kit-8 (CCK-8) assay, Transwell migration and invasion assays. Apoptosis was assessed by flow cytometry analysis of fluorescence signals from Annexin V and propidium iodide (PI) staining.

**Results** Compared to Nor and TB populations, individuals diagnosed with EOC demonstrated a significantly diminished gut microbiota diversity when contrasted with both normal controls and those presenting benign conditions. Specifically, the relative abundance of *Bilophila*, *Bifidobacterium*, and other probiotics was significantly reduced in patients diagnosed with epithelial ovarian cancer (EOC), while *Escherichia* and *Shigella* demonstrated a marked enrichment within this cohort. Differential microorganisms were identified through the application of machine learning techniques to delineate the characteristic microbial profiles associated with the EOC patients. A significant correlation was identified between the Cally index and microorganisms. In conclusion, we utilized microbial biomarkers

<sup>†</sup>Cheng Chen and Chengyuan Deng have contributed equally.

\*Correspondence:

Xiaoxin Jiang

jiangxiaoxin168@126.com

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



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

alongside the Cally to establish a diagnostic model for epithelial ovarian cancer, receiver operating characteristic (ROC) curve Area Under Curve (AUC) of 0.976 (95%CI 0.943–1.00), The AUC obtained from the Bootstrap internal validation was 0.974. The Hosmer–Lemeshow test revealed a robust concordance between the observed probabilities and the predicted probabilities generated by the model. The decision curve analysis revealed that the model provided a significant net clinical benefit. A total of 233 differential metabolites were identified between the EOC group and the NT (NOR and TB) groups. Among these, eight specific metabolites (HMDB0243492, C09265, HMDB0242046, HMDB0240606, C04171, HMDB0060557, HMDB0252797, and C21412) were exclusively derived from the microbiome. Notably, metabolite HMDB0240606 exhibited a significant positive correlation with *Escherichia coli* and *Shigella*, while it showed a significant negative correlation with *Ruminococcus*. In vitro studies demonstrated that *Bifidobacterium* possessed anti-tumor activity, whereas *Escherichia coli* exhibited pro-tumor activity.

**Conclusion** This study provides the inaugural comprehensive analysis of gut microbiota composition and its differential profiles among patients with epithelial ovarian cancer, those with benign ovarian tumors, and healthy controls in Hunan province, China.

**Keywords** Epithelial ovarian cancer, Gut microbiota, Enterotype diagnostic model

## Introduction

Ovarian cancer is a common malignant tumor disease in women, epidemiological surveys show that ovarian cancer is the eighth-leading cause of cancer-related death among women globally [1]. Approximately 90% of ovarian tumors are of epithelial origin, thereby categorized as epithelial ovarian cancer (EOC). The unclear pathogenesis, insidious onset of EOC, and lack of effective early detection strategies result in ~ 75% of patients being diagnosed at an advanced stage, which correlates with a 5-year survival rate below 47% [2]. The late detection of the disease can largely explain the high mortality rate of ovarian cancer. Therefore, it is essential to find new earlier diagnostic biomarkers of EOC.

Researchers have been focusing on studying the gut microbiome in recent years. Previous evidence supports that changes in the gut microbiota are significantly associated with extraintestinal diseases, such as melanoma [3], rheumatoid arthritis [4], and polycystic ovary syndrome [5]. Recent evidence reported that gut microbiota shapes the liver inflammatory microenvironment through the liver–gut axis, thereby promoting the advancement of hepatocellular carcinoma (HCC) [6]. Moreover, Intestinal microbial metabolites, including bacteriocins, short-chain fatty acids, phenylpropane-derived metabolites, and secondary bile acids, were identified in multiple carcinomas can fine-tune the tumor microenvironment through different molecular mechanisms, and drive oncogenic transformation. These studies suggest that intestinal microbes play an important role in the occurrence and development of tumors. Moreover, David et al. observed that *Lactobacillus*-dominated vaginal communities were less common in EOC, as compared to existing datasets of similarly aged healthy controls [7]. In animal models, intestinal dysbiosis upregulates serum levels of

the proinflammatory cytokine IL-6 and stimulates the activation of tumor-associated macrophages in ovarian cancer [8], noting that gut microbial may interfere with the metabolism modulate the development of EOC.

Currently, the screening methods for EOC mainly rely on serum CA125, HE4 detection, and imaging examination, but the above methods are single and have certain limitations. The combined screening of multiple indicators can improve early cancer diagnostic sensitivity and reduce the risk of missed diagnosis. With the development of detection technology, the intestinal flora can be used as a special biomarker and is expected to become one of the combined indicators for ovarian cancer screening. A recent study showed that the microbiota in the fallopian tube of EOC patients is different from that of healthy people [9]. However, no study has established an EOC diagnostic model based on the gut microbiome.

In this study, we recruited EOC patients (EOC), benign ovarian tumor group (TB), and normal controls (Nor) and performed the 16S rRNA sequencing of their intestinal floras and examined multiple clinical and serum indicators. Noting that gut microbial differences may affect ovarian cancer risk through multiple pathways, we aimed to characterize the composition and biological functions of the intestinal floras among EOC, TB, and Nor through bioinformatic methods. In doing so, we hope to explore the differences in gut microbial diversity and elaborate on the effectiveness of gut microbiota as a tool for early diagnosis of EOC. We predicted the functions of each flora of EOC patients, laying the groundwork for the application of the intestinal microbiota and related clinical biomarkers to the early diagnosis of EOC. In this study, untargeted metabolomics analysis was conducted on plasma samples from the EOC and NT groups to identify significantly differing metabolites between the two groups and

investigate potential associations with microbiota. Furthermore, *E. coli* and *Bifidobacterium*, which exhibited significant differences in abundance between the EOC and NT groups, were selected for co-culture experiments with ovarian cancer cells. This approach aimed to systematically explore the specific effects of these bacterial metabolites on the proliferation, migration, and apoptosis of ovarian cancer cells.

Materials and methods

Study design and specimen collection

This research was a single-center cross-sectional study and the design is shown in the following flowchart (Fig. 1). Ethical approval was obtained from the Ethics Committee of The First Affiliated Hospital of University of South China. All methods were conducted in accordance with the Declaration of Helsinki and relevant guidelines and regulations. Participants were recruited from April to June 2020 in outpatient and inpatient gynecology departments. Inclusion criteria: (1) All recruited participants were pathologically confirmed for the presence or absence of ovarian cancer. (2) No concomitant or

previous malignancy, or gastrointestinal disease (including peptic ulcer, inflammatory bowel disease, acute gastroenteritis, etc.), pelvic abscess (3) The patient had not received any anti-tumor treatment, such as surgery, chemotherapy, radiotherapy, immunotherapy, Chinese herbal medicine, etc. (4) No unconscious or other cognitive impairment. Normal populations without metabolic, cardiovascular, or cerebrovascular diseases or cancer, matched for age and menstrual status, were selected as controls. All participants did not receive any antibiotic or probiotic therapy for at least one month prior to biosample collection. The participants' feces are stored in sterile containers and preserved in a − 80°C freezer. The venous blood sample was collected in the morning before eating, for routine blood tests and biochemical testing.

16S rRNA gene-sequencing

Total genomic DNA was extracted from the specimens using the Fecal genomic DNA extraction kit (TIANGEN, DP328, China). The microbial genomic DNA extracted from feces will be specifically amplified for the 16S V3–V4 region using the forward primer 341F (5′-ACTCCT

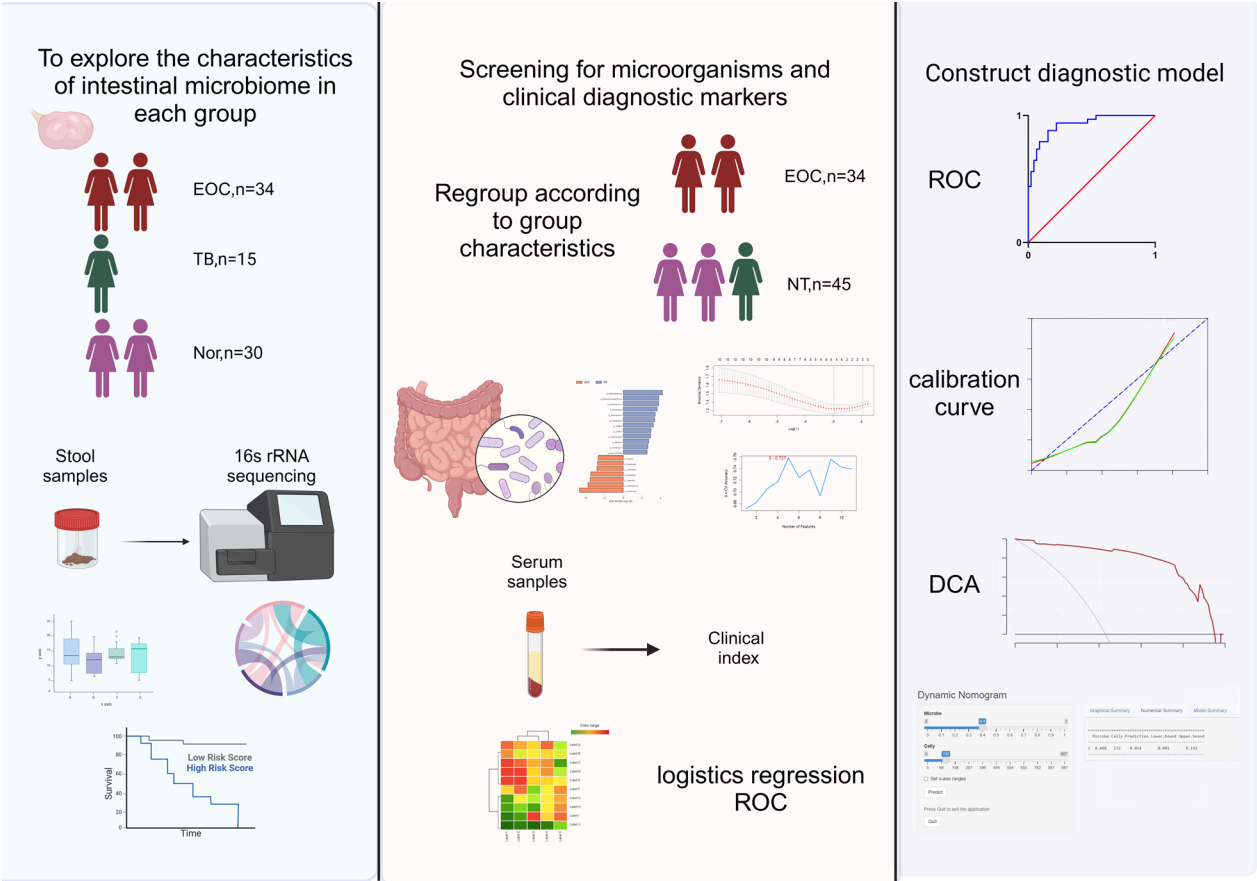


Fig. 1 Schematic representation of the study protocol, created with BioRender.com

ACGGGAGGCAGCAG-3') and reverse primer 806R (5'-GGACTACHVGGGTWTCTAAT-3'). Then the PCR amplification method uses primers containing adapters to enrich the target region, obtain a library of gut microbiota, and subsequently perform sequencing using Miseq PE300 (Illumina). Subsequently, QIIME (V1.9.0) was used to analyze the raw data of 16S rRNA sequencing.

### Bioinformatics analysis

Sequences with  $\geq 97\%$  similarity were grouped into the same operational taxonomic units OTU, and representative sequences for each OTU were annotated using Usearch (<http://www.drive5.com/usearch/>), the sequence with the highest abundance in each OTU was selected as the representative sequence for that OTU. Using the Greengenes database to obtain the classification information corresponding to each OTU, and removing OTUs with abundance values below 0.001% (one ten-thousandth) of the total sequencing amount of all samples. Alpha and beta diversity analyses were performed using the QIIME software (V1.9.0) and R software (V4.4.1). Based on OUT data, use the Pareto data method, the oscoprespls algorithm, with a cross-validation multiplier of 7 and 200 permutations, to perform PLS-DA visualization. LEfSe analysis is used to identify key microbial taxa with significant differences between two groups, with an LDA threshold  $> 2.0$  and a P value  $< 0.05$ . Functional prediction and analysis are performed using PICRUSt2. Based on the abundance of microbial genera and clinical parameters, RDA/CCA analysis was conducted using the "vegan" R package. Enterotype analysis was performed using the OmicStudio tools [10], according to the Calinski-Harabasz (CH)-index value, the best cluster value was selected as 4, which was the number of sample classifications, and the visual analysis was performed by principal component analysis (PCA). The bacteria with the highest content of each enterotype were calculated as the representative bacteria of the enterotype. OTU data were used in the "WGCNA" R package was used for WGCNA analysis, Gephi software (V0.10.0) was used to visualize the interaction between the difference module and the clinical parameters.

### Prediction modeling and performance evaluation

The "glmnet" package and "e1071" package in R, two kinds of extensive machine learning methods, were used for the LASSO (Least Absolute Shrinkage and Selection Operator) model and the SVM (support vector machines) model building, to identify the microbiota biomarkers, and the "tidyverse" package was used for visualization. The accuracy of each model was assessed using the area under the curve (AUC) value of the area under the subject operating curve (ROC). The Bootstrap

resampling method was used for internal validation of the model, implemented using the "caret" R package. The Hosmer–Lemeshow test results were visualized by the "rms" package to obtain the calibration curve to evaluate the prediction probability accuracy of the model, and the Decision Curve Analysis (DCA), using the "rmda" R packages, was used to evaluate the clinical net benefit.

### LC-MS/MS analysis as well as annotation

A total of 15 blood samples were included in the metabolomics analysis: 8 from the EOC group, 3 from the TB group, and 4 from the Nor group. Methanol, acetonitrile, formic acid and isopropyl alcohol were purchased from ANPEL. All solvents were LC-MS grade. The sample were thawed slowly at 4 °C, added 4 times the volume of the extraction solution (methanol/acetonitrile, 1:1, v/v), homogenized 60 s, extracted 30 min by low-temperature ultrasonic, set still for 1 h to precipitate protein at  $-20^{\circ}\text{C}$ , centrifuged 10 min with 12,000 rpm at 4 °C, the supernatant solution were dried in vacuum, added 0.1 mL 30% acetonitrile solution, homogenized and centrifuged 10 min with 12,000 rpm at 4 °C, the supernatant were taken for computer detection. The extracted sample were analyzed using an UPLC-Orbitrap-MS system (UPLC, Vanquish; MS, HFX). HRMS data were recorded on a Q Exactive HFX Hybrid Quadrupole Orbitrap mass spectrometer equipped with a heated ESI source (Thermo Fisher Scientific) utilizing the Full-ms-ddMS2 acquisition methods. The raw data were first pre-processed using Progenesis QI (Waters Corporation, Milford, USA) software, including baseline filtering, peak identification, peak matching, retention time correction, peak alignment, etc., to obtain a data matrix containing retention time, mass-charge ratio, and peak intensity. The peaks containing secondary mass spectrometry data were identified by commercial databases and secondary mass spectrometry databases and corresponding cracking rules.

### In vitro experiments

The ovarian cancer cell line SKOV3 utilized in this study was obtained from Procell (Wuhan, China). The experimental bacterial strains included *Escherichia coli* (ATCC25922) and *Bifidobacterium longum subsp. longum* Reuter (ATCC15707). Both strains were procured from Bio science (Shanghai, China). SKOV3 cells were seeded into 6-well plates and incubated at 37 °C with 5% CO<sub>2</sub> until they reached 80% confluence. The culture medium was PMMI-1640, supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin solution. Cells were washed with sterile phosphate-buffered saline (PBS) to remove residual culture medium. *E. coli* was inoculated into LB liquid medium and cultured at 37 °C for 12–16 h to ensure the bacteria reached the



logarithmic growth phase. *Bifidobacterium* was inoculated into MRS liquid medium and cultured under anaerobic conditions at 37 °C for 12–16 h. *E. coli* and *Bifidobacterium* cultures were centrifuged and subsequently filtered to obtain supernatants at varying concentrations. SKOV3 cells were treated with varying concentrations of bacterial supernatant in 6-well plates and cultured for 24 or 72 h. The wound healing assay was employed to assess cell migration rate, the CCK-8 assay was used to evaluate cell proliferation activity, Transwell assays were conducted to assess cell migration and invasion capabilities, and flow cytometry was utilized to detect Annexin V and PI fluorescence signals to analyze apoptosis status.

### Statistical analysis

$$\text{CALLY index} : \text{Albumin} \times \text{Lymphocyte} \div (\text{CRP} \times 10).$$

Image J is used as a cell count. SPSS (V24.0) was used to analyze the data. The data with normal distribution and homogeneity of variance among the three groups were analyzed by One-Way ANOVA, and the data with non-normal distribution or heterogeneity of variance were analyzed by the Kruskal–Wallis test. The T-test was used for continuous variables with normal distribution between the two groups, and the Mann–Whitney U test was used for continuous variables without normal distribution. Spearman correlation analysis was used to analyze the correlation between microbiota, functional pathways, and clinical traits. The “ggplot2” package in R, GraphPad Prism (V5.0), Gephi (V0.10.0), and Adobe Illustrator 2024 were used for graph visualization. The difference is statistically significant when  $P < 0.05$ .

## Result

### Baseline characteristics of the participants

After strict screening and exclusion criteria, a total of 79 individuals were enrolled in this study, including 34 EOC patients, 15 TB patients, and 30 normal controls. As illustrated in Table 1, there were significant differences in white blood cells (WBC), neutrophils (NEU), monocytes (MON), basophils (BAS), blood platelets (PLT), albumin (ALB), aspartate aminotransferase (AST), serum ferritin (SF), tumor supplied group of factors (TSGF), high sensitivity c-reactive protein (hs-CRP), Cally, and CA125 among the three groups. It is worth noting that the inflammatory indicators such as WBC, NEU, and hs-CRP in the EOC group were significantly higher than those in the Nor group. No significant differences were observed in age, menopausal status, lymphocytes (LYM),

eosinophils (EOS), aminotransferase (ALT), total bile acid (TBA), globulin (GLO), total cholesterol (TCHO), triglyceride (TG), high-density lipoprotein (HDL), and low-density lipoprotein (LDL) among the three groups.

### Explore the characteristics of three groups of intestinal microbes

To ensure that our results were reliable, we first examined the sequence length and found that it was close to the length of the 16s rRNA amplified region, indicating good amplification characteristics (Fig. 2A). In addition, the results of the species accumulation curves and rarefaction curves indicated that our sample size was sufficient and that the current sequencing results were sufficient to reflect the diversity contained in the current sample (Additional file 1: S1A, B, C, D).

### Altered gut microbiota diversity in TB and EOC patients

Based on the OTU data from the three groups, we drew Venn diagram and heat maps (Fig. 2B, C) to show the basic situation of OTUs in each group, what's interesting to us is that the microbiota structure of TB patients was more similar to Nor in the heat map, while the EOC patients could be clearly distinguished from these two populations. Alpha diversity was assessed with Ace, Simpson, Shannon, and Chao1 indices among the three groups (Fig. 2D), the microbial diversity index of the EOC group significantly decreased ( $P$  value  $< 0.05$ ) compared to Nor, while TB was between the two groups with no significant difference, which indicates that the microbial diversity in EOC patients has changed and is significantly reduced. There were no significant differences statistically in the principal component analysis (PCA), principal coordinate analysis (PCoA), and non-metric multidimensional scaling analysis (NMDS) distance index among the three groups (Additional file 1: S1E–G), but Partial Least Squares Discriminant Analysis (PLS-DA) analysis suggested that EOC was significantly separated from the normal population while TB was between the two groups (Fig. 2E, F).

Composition and comparison of the gut microbiota in patients with EOC, TB groups, and health control.

Based on the species annotation results, we plotted the phylum, genus, and species level bar chart (Fig. 3A–C). Compared with Nor group, *p\_Firmicutes*, *g\_Escherichia*, and *s\_E.Coli* at the species level in the EOC patients were significantly increased. The chord diagram at the genus level also shows that the proportion of the EOC group *Escherichia* is much higher than that of Nor group (Fig. 3D). Next, we clustered gut microbiota into enterotypes by genus-level expression content, according to the Calinski-Harabasz (CH) index results (Additional file 2: S2A) there were four types

**Table 1** Baseline characteristics of the participants

	Nor, n = 30	TB, n = 15	EOC, n = 34	P value
Age, years (mean ± SD)	53.30 ± 5.96	54.60 ± 8.84	56.74 ± 7.65	0.173
BMI, kg/m <sup>2</sup> (mean ± SD)	22.07 ± 1.77	23.33 ± 2.66	23.64 ± 3.30	0.061
Menopause, n (%)				0.627 <sup>#</sup>
Menopause	20	9	25	
Premenopausal	10	6	9	
Differentiation, n (%)				
High			30(88.24%)	
Middle + low			4(11.76%)	
FIGO, n (%)				
I + II			9(26.47%)	
III + IV			25(73.53%)	
Histopathology, n (%)				
Serous ovarian cancer			27(79.41%)	
Mucinous ovarian cancer			7(20.59%)	
WBC, 10 <sup>9</sup> /L (mean ± SD)	5.41 ± 1.33	6.14 ± 1.63	7.36 ± 1.74	0.000
NEU, 10 <sup>9</sup> /L (mean ± SD)	3.22 ± 0.94	3.66 ± 1.27	5.11 ± 1.61	0.000 <sup>#</sup>
LYM, 10 <sup>9</sup> /L (mean ± SD)	1.78 ± 0.49	1.92 ± 0.69	1.55 ± 0.45	0.054
MON, 10 <sup>9</sup> /L (mean ± SD)	0.30 ± 0.08	0.41 ± 0.12	0.54 ± 0.15	0.000 <sup>#</sup>
EOS, 10 <sup>9</sup> /L M(IQR)	0.09(0.05 ~ 0.13)	0.09(0.06 ~ 0.16)	0.09(0.05 ~ 0.13)	0.798 <sup>#</sup>
BAS, 10 <sup>9</sup> /L (mean ± SD)	0.01 ± 0.01	0.04 ± 0.01	0.05 ± 0.09	0.000 <sup>#</sup>
PLT, 10 <sup>9</sup> /L (mean ± SD)	208.90 ± 73.34	263.33 ± 63.77	334.56 ± 117.38	0.000 <sup>#</sup>
ALB, g/L (mean ± SD)	45.53 ± 2.05	44.88 ± 3.99	40.92 ± 5.17	0.000 <sup>#</sup>
ALT, U/L (mean ± SD)	14.75 ± 6.99	23.33 ± 18.73	20.11 ± 15.83	0.227 <sup>#</sup>
AST, U/L (mean ± SD)	19.63 ± 4.45	24.71 ± 9.62	29.94 ± 14.67	0.000 <sup>#</sup>
TBA, μmol/L (mean ± SD)	2.91 ± 2.90	3.97 ± 3.09	3.24 ± 2.88	0.493 <sup>#</sup>
GLO, g/L (mean ± SD)	29.15 ± 2.89	27.47 ± 7.47	29.39 ± 3.91	0.066 <sup>#</sup>
TCHO, mmol/L (mean ± SD)	4.61 ± 0.75	4.98 ± 0.98	4.80 ± 1.14	0.391 <sup>#</sup>
TG, mmol/L (mean ± SD)	1.21 ± 0.60	1.30 ± 0.57	1.40 ± 0.41	0.112 <sup>#</sup>
HDL, mmol/L (mean ± SD)	1.44 ± 0.44	1.30 ± 0.27	1.21 ± 0.26	0.202 <sup>#</sup>
LDL, mmol/L (mean ± SD)	2.68 ± 0.63	2.97 ± 0.58	2.98 ± 0.77	0.198
SF, ng/mL (mean ± SD)	107.54 ± 83.19	124.18 ± 87.85	239.5 ± 145.09	0.000 <sup>#</sup>
TRSF, μmol/L (mean ± SD)	31.68 ± 5.46	34.71 ± 6.92	26.77 ± 9.24	0.000 <sup>#</sup>
hs-CRP, mg/dL M(IQR)	0.06(0.02 ~ 0.12)	0.13(0.07 ~ 0.25)	1.45(0.36 ~ 3.86)	0.000 <sup>#</sup>
TSGF, U/mL (mean ± SD)	39.21 ± 8.14	43.55 ± 10.41	53.28 ± 12.12	0.000
Cally M(IQR)	165.80(71.51 ~ 428.46)	52.83(21.5 ~ 141)	4.96(1.59 ~ 17.61)	0.000 <sup>#</sup>
CA125, U/mL M(IQR)	21.94(16.26 ~ 31.8)	30.89(9.62 ~ 55.76)	543.6(150.51 ~ 731.57)	0.000 <sup>#</sup>

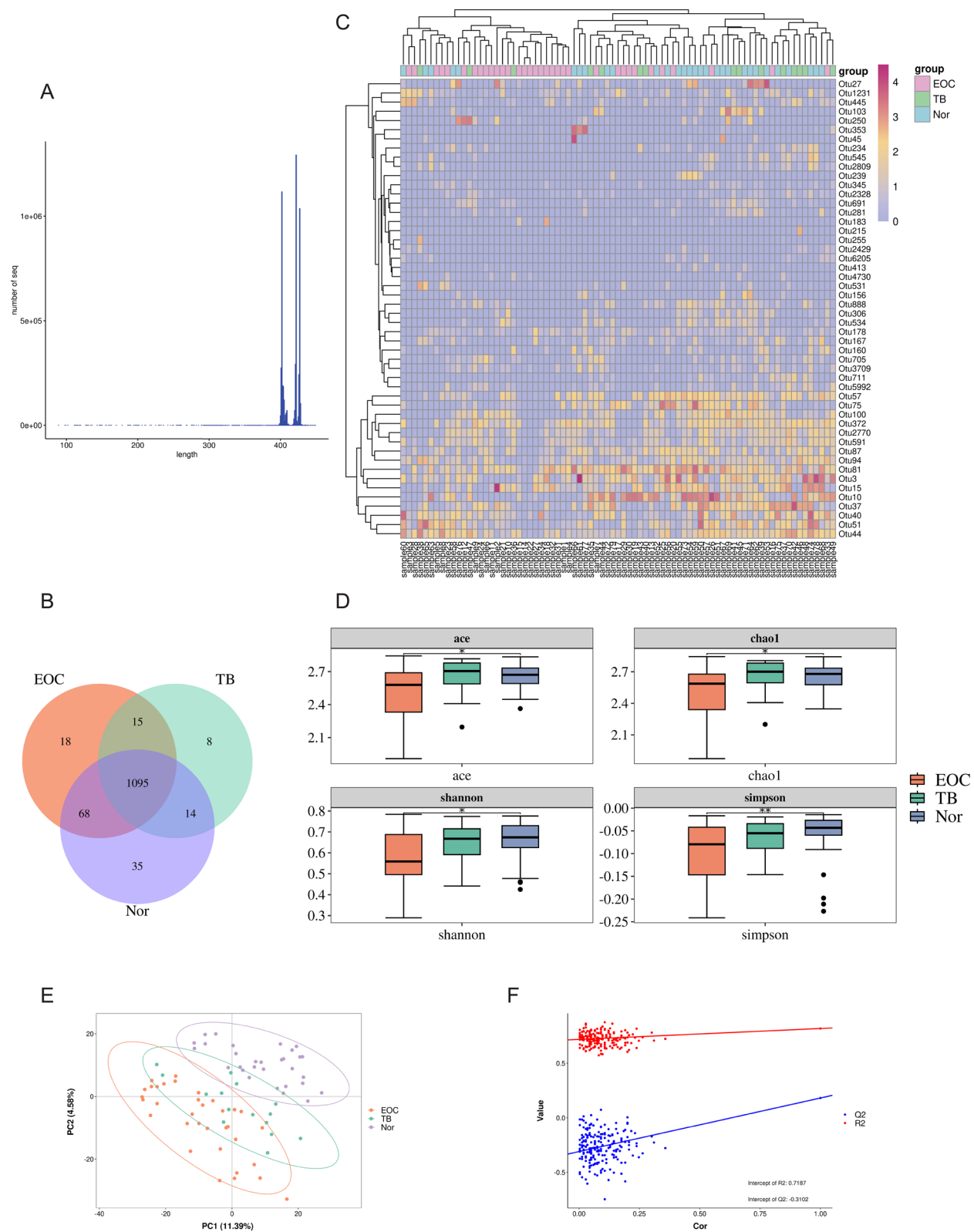
Results are presented as Mean ± SD for continuous variables and N (%) for categorical variables

N number of available values, SD standard deviation

<sup>#</sup> Non-parametric test

(See figure on next page.)

**Fig. 2** Data quality and  $\alpha$  and  $\beta$  diversity of microbial sequences. **A** The abscissa represents the length of the sequence, and the ordinate represents the number of sequences. The closer the length of the sequence obtained by sequencing to the length of the 16s rRNA amplified region, the better the amplification specificity. **B** Venn diagram shows the number of unique and common OUTs in EOC, TB, and Nor groups. **C** Heat map of OTUs distribution in samples from each group. **D** Ace, Simpson, Shannon, and Chao1 indexes of the three groups were compared by Wilcoxon test between the two groups and Kruskal–Wallis test between the three groups. “\*” represented  $p < 0.05$ , “\*\*\*” represented  $p < 0.01$ . **E** PLS-DA analysis of intestinal flora among the three groups. **F** PLS-DA model test plot



**Fig. 2** (See legend on previous page.)

(Fig. 3E). Representative bacteria for the four enterotypes are *Escherichia*, *Prevotella*, *Megamonas*, *Bacteroides*, respectively (Fig. 3F). There was the greatest number of type 1 cases in the EOC group (Additional file 2: S2B). The results of the ternary diagram likewise showed that EOC group *Escherichia* expressed the highest relative to TB and Nor groups (Fig. 3G). In addition, patients with EOC were followed up to obtain Disease Free Survival (PFS) data. Kaplan–Meier survival curves were drawn for the four enterotypes (median PFS: 17 vs 14.5 vs 19 vs 13 months,  $P=0.77$ ) and the high and low expression groups of *Escherichia* relative abundance (median PFS: 15 vs 19 months,  $P=0.2$ ), respectively (Additional file 2: S2C, D). Patients with enriched *Escherichia* had worse PFS, but the difference was not significant. In addition, *g\_Shigella* and *g\_escherichia* in the EOC group is significantly increased, and *g\_Ruminococcus*'s abundance in the Nor group is significantly higher than in EOC patients (Additional file 2: S2E). Additionally, we compared the *Bifidobacterium*/*Enterobacteriaceae* ratio and *Firmicutes*/*Bacteroidetes* ratio among the three groups (Fig. 3H), there was no significant difference in the F/B ratio between the two groups. However, the B/E ratio of the EOC group was significantly lower than that of Nor group, which indicated that EOC patients had impaired colonization ability of beneficial gut microbes.

#### Differences in gut microbial composition and microbial function prediction analysis

To explore the characteristic microorganisms in each group, LEfSe analysis was performed between pairs (Fig. 4D–F), meanwhile, the cladogram generated by the LEfSe method showed the phylogenetic distribution of the gut microbiota (Fig. 4A–C), it is obvious that the EOC group is significantly different from the Nor and TB groups, while the difference between Nor and TB patients is small. Compared with Nor group, the EOC group had significantly increased levels of opportunistic pathogens such as *Escherichia* and *Shigella*, and the expression of probiotics such as *Bifidobacterium* decreased. To investigate functional alterations in epithelial ovarian cancer microbial communities, we used PICRUST2 to predict gene functions and pathway potentials that might be affected by differential microbes. The results showed

61 pathways with different activity levels between the EOC and Nor groups (Fig. 4G). Among these pathways, 28 were activated in the EOC group ( $P<0.05$ ): Shigellosis, Pathogenic *Escherichia coli* infection, Amino acid metabolism, Nucleotide metabolism, etc. And 31 pathways were activated in the Nor group ( $P<0.05$ ): Arginine and proline metabolism, Transcription machinery, etc. However, there were few differential functional pathways between TB and EOC (Fig. 4H) and TB and Nor (Fig. 4I).

Overall, there was a significant difference in microbial diversity between EOC and Nor groups, mainly reflected in the increased expression of opportunistic pathogens, the decreased expression of probiotics, and the imbalance of gut microbiome in EOC patients, which leads to the differences in signal functional pathways. The above results revealed the dysregulated or activated functional pathways of gut microbiota associated with EOC occurrence.

#### Development and validation of a predictive model based on gut microbiome signature and clinical features

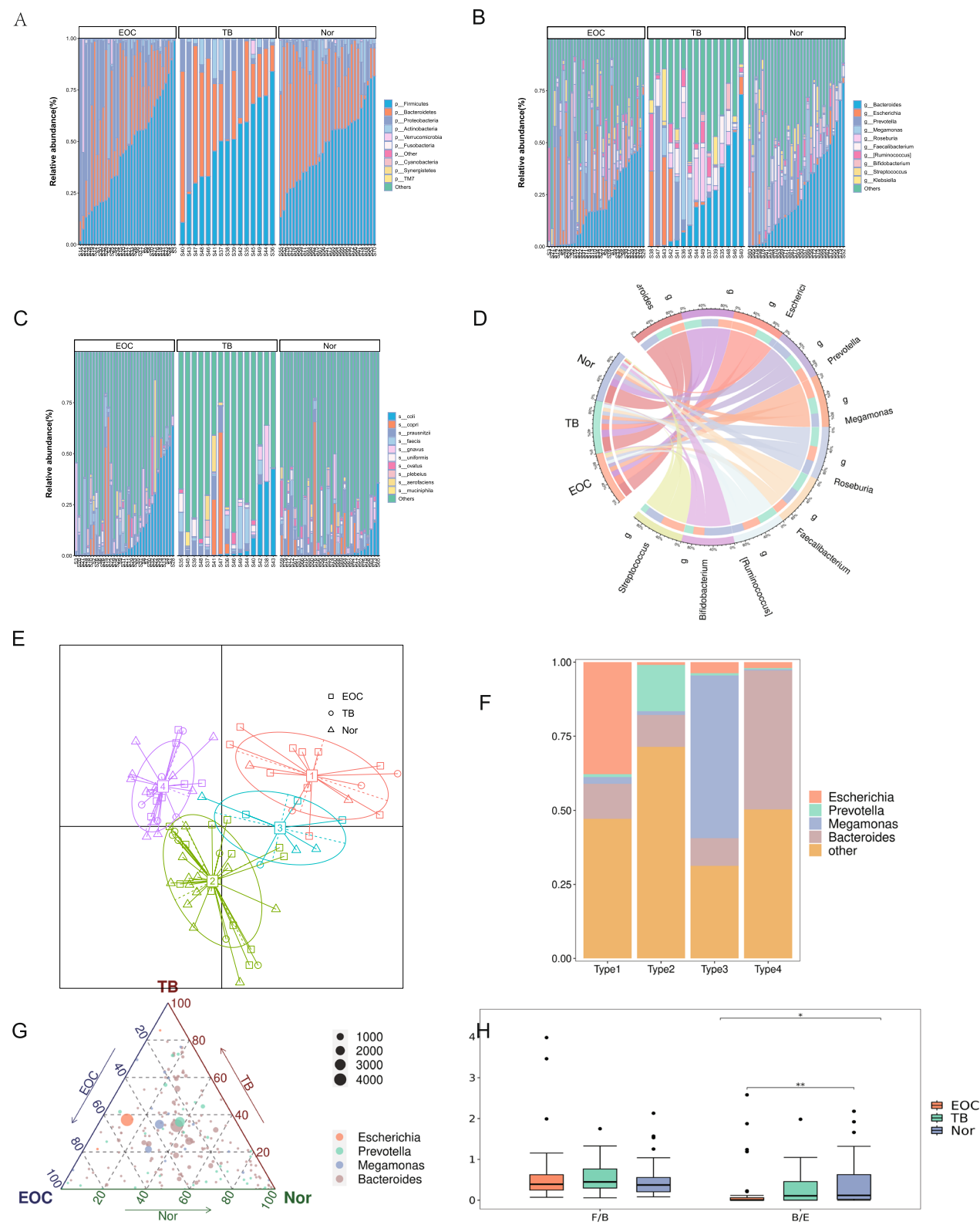
We explored microbiotas that showed significantly altered abundances between the different pairs of the population (EOC vs Nor, EOC vs TB, and Nor vs TB, adjusted  $p<0.05$ ), analyses based on the above revealed similar patterns in the normal population and benign ovarian tumor patients, from which the epithelial ovarian cancer population can be clearly distinguished. On further comparison of the significantly altered microbiotas in the three pairs, the Nor versus EOC pair showed the most obvious similarity with the TB versus EOC pair (Fig. 5A), LEfSe results also indicated that Nor and TB were highly similar. Therefore we grouped the TB and Nor groups into one group, the non-epithelial ovarian cancer (NT) group.

#### Microbial diversity of EOC patients and NT populations

The alpha diversity in each group was evaluated using Ace, Chao, Simpson, and Shannon indices according to the genus level. The  $\alpha$ -diversity index of the EOC group was significantly lower than that of the NT group (Fig. 5B). Furthermore, the PCoA, PCA, and NMDS

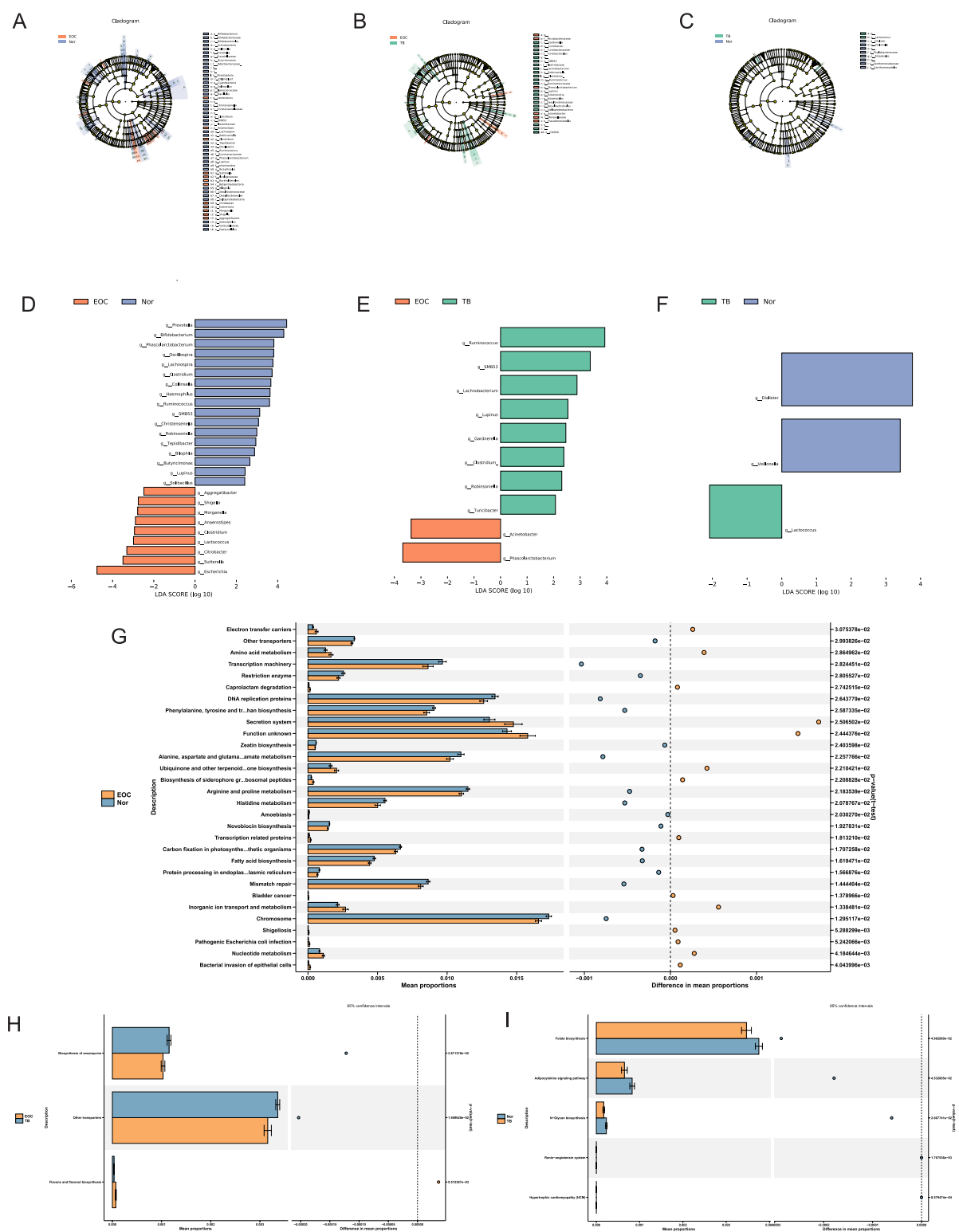
(See figure on next page.)

**Fig. 3** Microbiota Profile of Healthy Controls, TB patients, and EOC patients. **A** Bar chart of the top ten percent abundance of gut microbiota at the phylum level. **B** Bar chart of the top ten percent abundance of gut microbiota at the genus level. **C** Bar chart of the top ten percent abundance of gut microbiota at the species level. **D** Chord diagram of the top ten percent abundance of gut microbiota at the genus level. **E** The distribution of samples in the four cluster types is presented based on the unweighted Unifrac distance, different shapes represent different groups, and different colors represent different enterotypes. **F** Bar chart of representative bacteria percent abundance for each enterotype. **G** Trigram of the four enterotypes representing bacteria. **H** Boxplots of F/B ratio and B/E ratio. \*\*\* represented  $p<0.05$ , \*\*\*\* represented  $p<0.01$

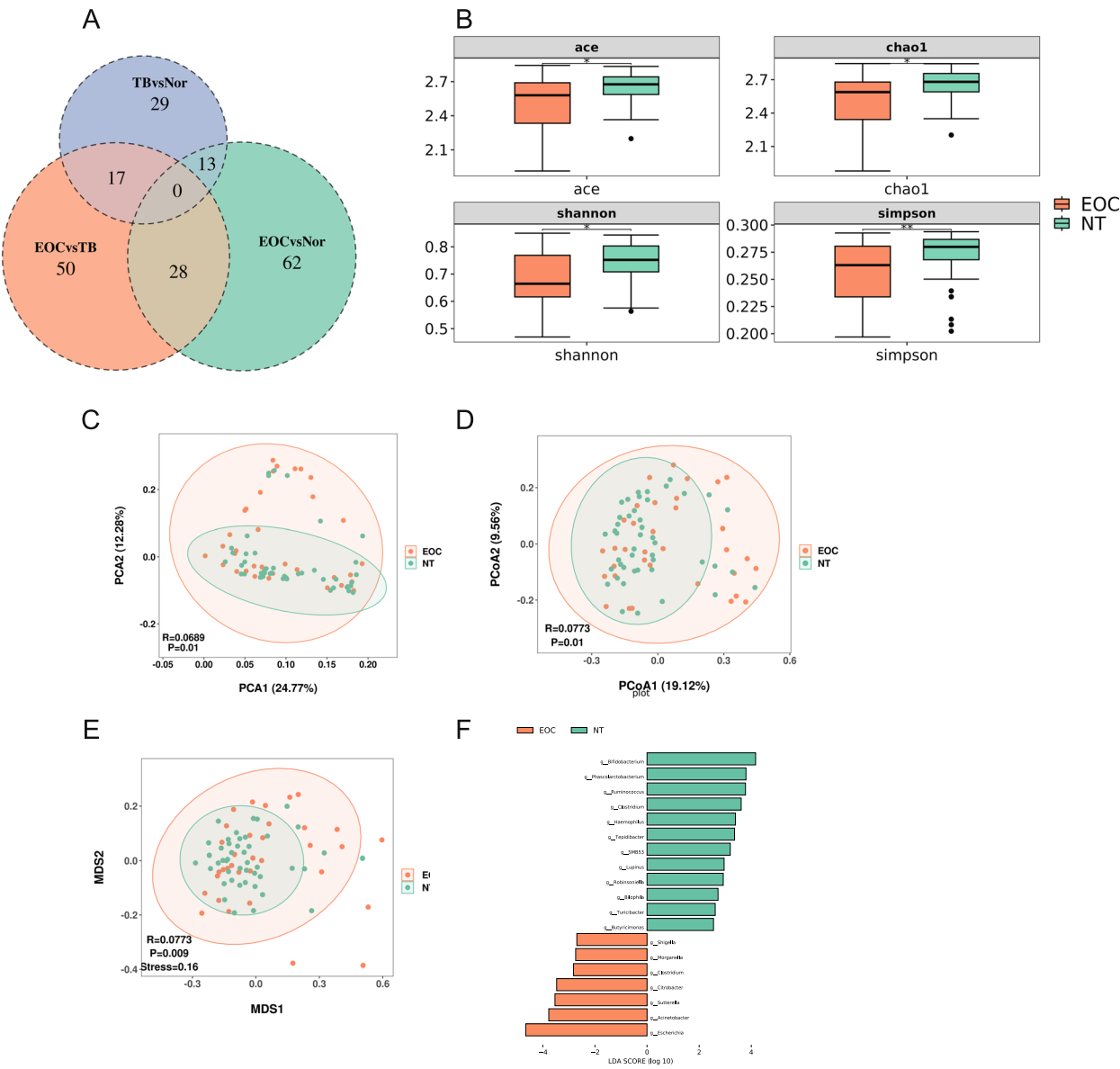


**Fig. 3** (See legend on previous page.)





**Fig. 4** The LDA effect size analysis (LEfSe) results highlight gut microbiota distinctions in the EOC, TB, and Nor groups. **A–C** The cladogram depicts the differential microbial branches and nodes, the node diameter size corresponds to the mean relative abundance of this microbiota. **D–F** The ordinate indicates microbiota with significant differences between groups, the abscissa shows the microbiota with LDA score > 2,  $P < 0.05$ , in the form of bar chart. **G–I** Results of the functional annotations in the three groups based on the PICRUSt2



**Fig.5** Investigation of gut microbiome diversity that are significantly altered in EOC patients. **A** Venn diagram showing the overlaps among the three altered microbiome pairs (EOC vs Nor, EOC vs TB, and Nor vs TB). **B** Ace, Chao, Simpson, and Shannon indices of the three groups. **C** PCA analysis. **D** PCoA analysis. **E** NMDS analysis. P values and R values for  $\beta$  analysis were calculated by the ANOSIM test. **F** The ordinate indicates microbiota with significant differences between groups, the abscissa shows the microbiota with LDA score > 2,  $P < 0.05$ , in the form of bar chart

utilizing the Bray–Curtis statistical method, as illustrated in Fig. 5C–E revealed significant differences in  $\beta$ -diversity between EOC patients and NT populations.

**LeFSe analysis was performed between EOC and NT groups, and the annotation of functional pathways in differential microbiota**

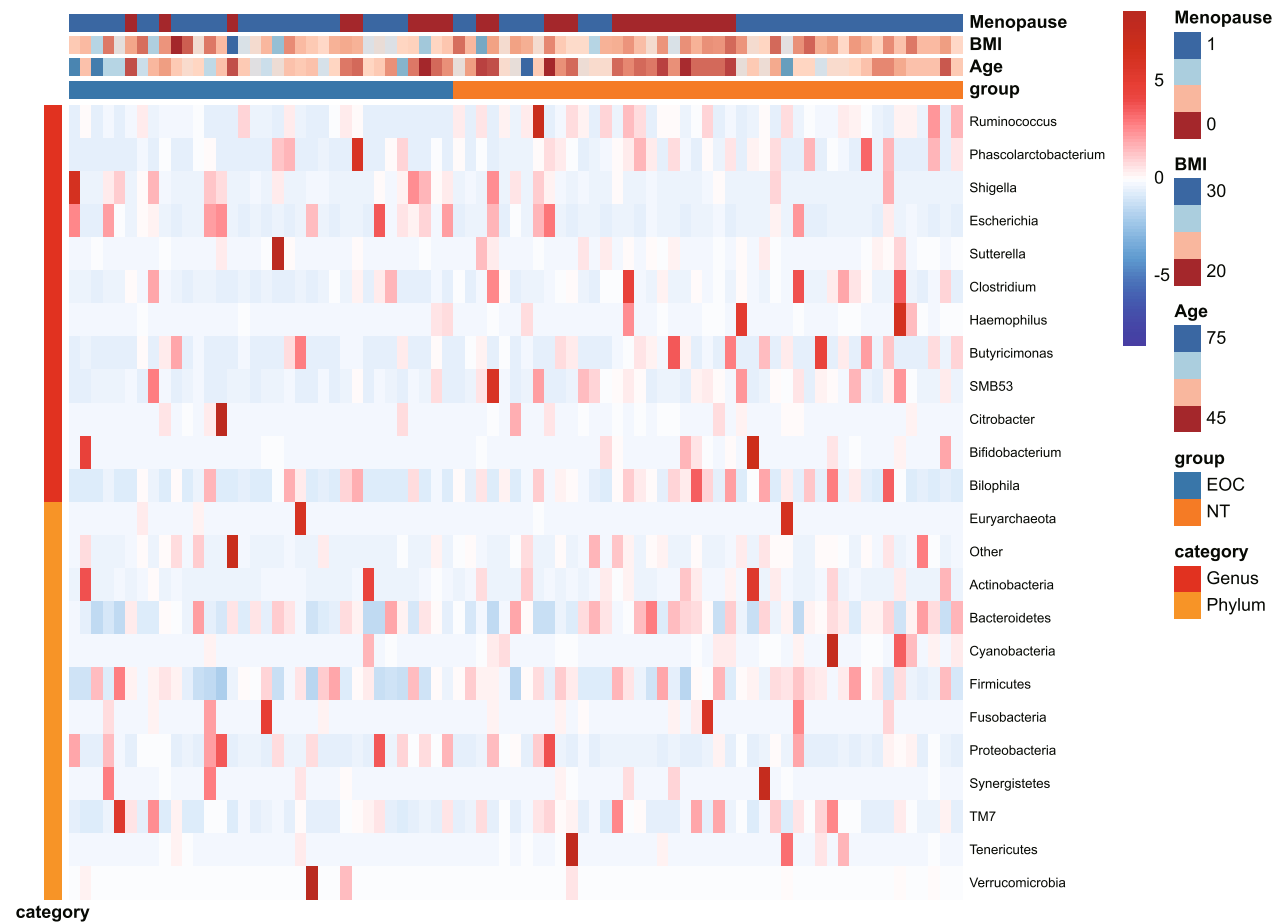
LeFSe analysis showed that EOC patients significantly increased the abundance of *g\_Escherichia* and *g\_Shigella*. However, *g\_Bifidobacterium*, *g\_Ruminococcus*, and other probiotics were significantly increased in the NT group (Fig. 5F, Additional file 3: S3A). Using functional prediction analysis, a total of 25 significantly

differed Encyclopedia of Genes and Genomes (KEGG) categories between the EOC and NT groups were obtained by PICRUSt2. 10 pathways including Pathogenic *Escherichia coli* infection and Shigellosis were enriched in the EOC, 15 pathways including Arginine and proline metabolism and Histidine metabolism were enriched in the NT (Additional file 3: S3B).

**Correlation between differential microorganisms and clinical features**

Next, we filtered out the low-abundance microbial (less than 20% of the samples contain the bacteria) communities in the LEfSe analysis, and the final included 18

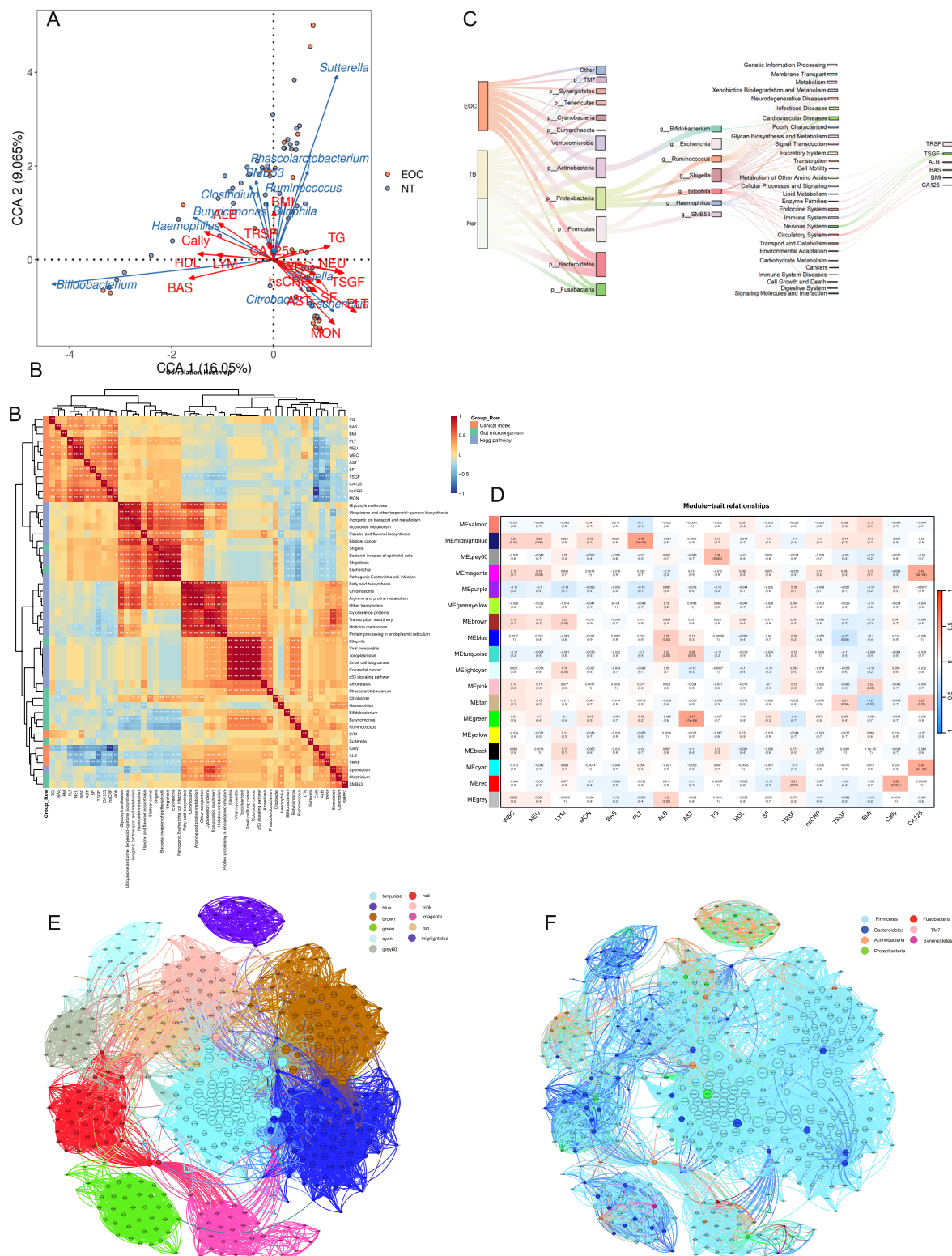
microorganisms. According to the 6 phyla level bacteria, 12 genera level bacteria and their relationship of clinical parameters draw heat maps, the basic information of each sample between the EOC and NT groups are shown in Fig. 6. To determine the impact of clinical factors on the distribution of our sample, we conducted RDA/CCA analysis and multi-parameter intercorrelation analysis among bacteria and clinical parameters and KEGG pathways, the results shown in Fig. 7A, B, Additional file 3: S3C. Microbes can be found to be closely related to clinical parameters, but there were significantly more intracorrelations than intercorrelations. *g\_Shigella* was positively correlated with CA125



**Fig. 6** Taxonomic features of the gut microbiota and clinical parameters of the EOC patients and the NT populations

(See figure on next page.)

**Fig. 7** Relationship between gut microbiota and clinical parameters. **A** CCA results of gut microbiota and clinical factors of patients with EOC and NT group. **B** Multi-parameter intercorrelation analysis among bacteria, clinical parameters, and KEGG pathways. **C** Sankey diagram containing groups, phylum level bacteria, genus level bacteria, functional pathways, and clinical parameters. **D** Module-trait-relationship heatmap. **E** Correlation network diagram, different colors represent different modules. **F** OTU co-occurrence network diagram was constructed based on the WGCNA module, and different colors represent different bacteria at the phyla level. \*\*\* represented  $p < 0.05$ , \*\*\*\* represented  $p < 0.01$



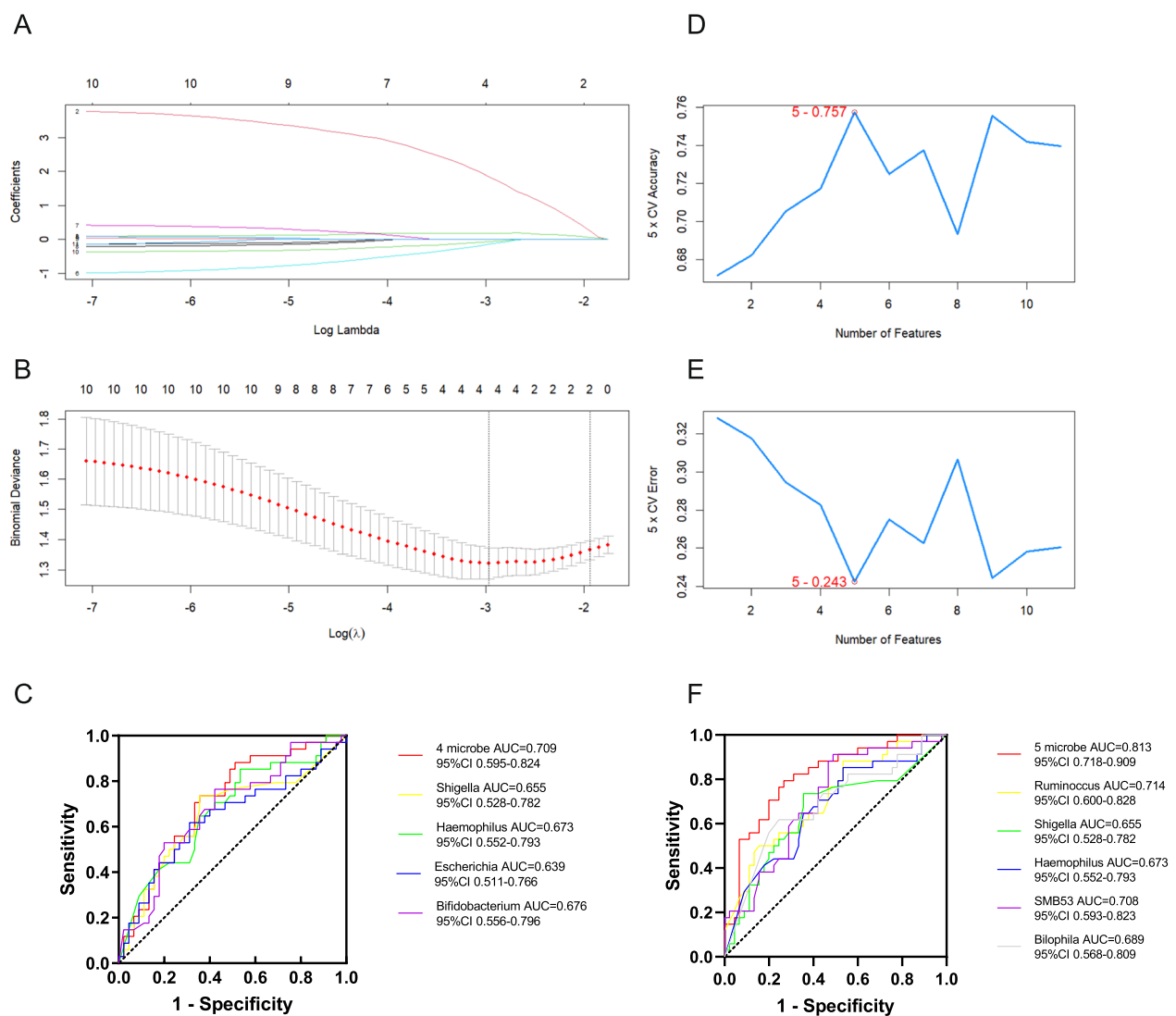
**Fig. 7** (See legend on previous page.)

and Bacterial invasion of epithelial cells pathway, negatively correlated with Cally; *g\_Ruminococcus* was negatively correlated with CA125; and *g\_Bifidobacterium* was positively correlated with Cally and Shigellosis pathway. Sankey diagram further demonstrated the interrelationship among gut microbiota, functional pathways, and clinical features (Fig. 7C).

In addition, we performed WGCNA analysis based on OUTs and constructed a heat map of the relationship between each OTU co-existing module and clinical parameters (Fig. 7D), the red module is closely related to Cally, and the cyan module is closely related

to CA125. Interaction network analysis was performed with modules related to clinical variation, and the complex correlations between modules are shown in Fig. 7E, *p\_Firmicutes* plays a dominant role in this (Fig. 7F).

In summary, we suggest that gut microbiota is closely related to clinical features, and this correlation is not caused by a single bacterium, but by the co-influence of multiple bacteria.



**Fig. 8** Identification of screening markers utilizing machine learning techniques. **A** LASSO regression coefficient plot. **B** LASSO regression cross-validation plot. **C** ROC curve of four bacterial species identified through LASSO regression and their collective detection. **D** Error rate of SVM analysis, the lowest error rate was 0.243 when 5 bacteria were included. **E** Accuracy Rate of SVM analysis, the highest accuracy was 0.757 when 5 bacteria were included. **F** ROC curve analysis of five bacteria identified through SVM and their integrated detection



### EOC diagnostic models based on gut microbiome and clinical parameters

According to the results of LEfSe analysis, 11 microorganisms were selected for Least Absolute Shrinkage and Selection Operator (LASSO) analysis after excluding the low-expression microorganisms (Fig. 8A, B). Finally, four genera were included in the diagnostic model, including *Shigella*, *Escherichia*, *Haemophilus*, and *Bifidobacterium*. According to the four selected gut microbial markers constructing the microbiome model, to assess the predictive value of the microbiome model, the AUC based on the ROC curves was plotted, reaching an AUC value of 0.709 (95% CI: 0.595–0.824) (Fig. 8C). These 11 bacteria were then included in the Support Vector Machines (SVM) analysis, and five genera were obtained after filtering, including *Ruminococcus*, *Shigella*, *Haemophilus*, *SMB53*, *Bilophila*, the accuracy rate was 0.757, and the error rate was 0.243 (Fig. 8D, E), the AUC area was 0.813 (95%CI: 0.718–0.909) (Fig. 8F). Considering that there are only two bacteria screened by LASSO and SVM together: *Shigella* and *Haemophilus*, we finally combined the 4 bacteria selected by LASSO with the 5 bacteria selected by SVM as the markers of the EOC diagnostic model, including *Shigella*, *Escherichia*, *Haemophilus*, *Bifidobacterium*, *Ruminococcus*, *Shigella*, *SMB53*, *Bilophila*. The above results suggest that the gut microbiome has a certain diagnostic efficacy, but the gut microbiome alone may not be enough. Therefore, we included clinical indicators in the diagnostic model to optimize its diagnostic performance.

To screen for clinical indicators, we used univariate logistic regression analysis to screen clinical indicators with  $P < 0.05$ , a total of 10 clinical indicators were included. Since age is an independent risk factor for many diseases, we performed multivariate logistic regression analysis with age and these 10 clinical indicators, the results showed that WBC, NEU, BAS, and Cally were independent risk factors for EOC, as illustrated in the Additional file 4: Stable1. Subsequently, we constructed ROC curves for these four indicators. Cally's AUC area exceeded the other clinical indicators, so we finally incorporated Cally into the diagnostic model (Additional file 4: S4A).

The AUC area of the Cally and gut microbiome combined diagnostic model was 0.976 (95%CI 0.943–1.00)

(Fig. 9A), and the diagnostic efficacy of the Cally and microbiome model was superior to that of the CA125, this indicated that this diagnostic model has a great application prospect in clinical practice. Bootstrap was used for internal validation of the model, and the AUC area of the ROC curve constructed by 1000 repeated sampling times also reached 0.974, sensitivity of 0.923, and specificity of 0.961 (Fig. 9B). The P-value derived from the calibration curve established by the Hosmer–Lemeshow test was 0.986 ( $p > 0.05$ ) (Fig. 9C), indicating that the diagnostic model exhibited good repeatability. The DCA curve indicated that the model demonstrated a substantial net clinical benefit in practical application (Fig. 9D). Ultimately, we developed a nomogram utilizing microbial data and the Cally index (Fig. 9E). Subsequently, we created an accompanying web version dynamic nomogram to facilitate its clinical application (Fig. 9F).

### Metabolic features in the EOC and NT groups

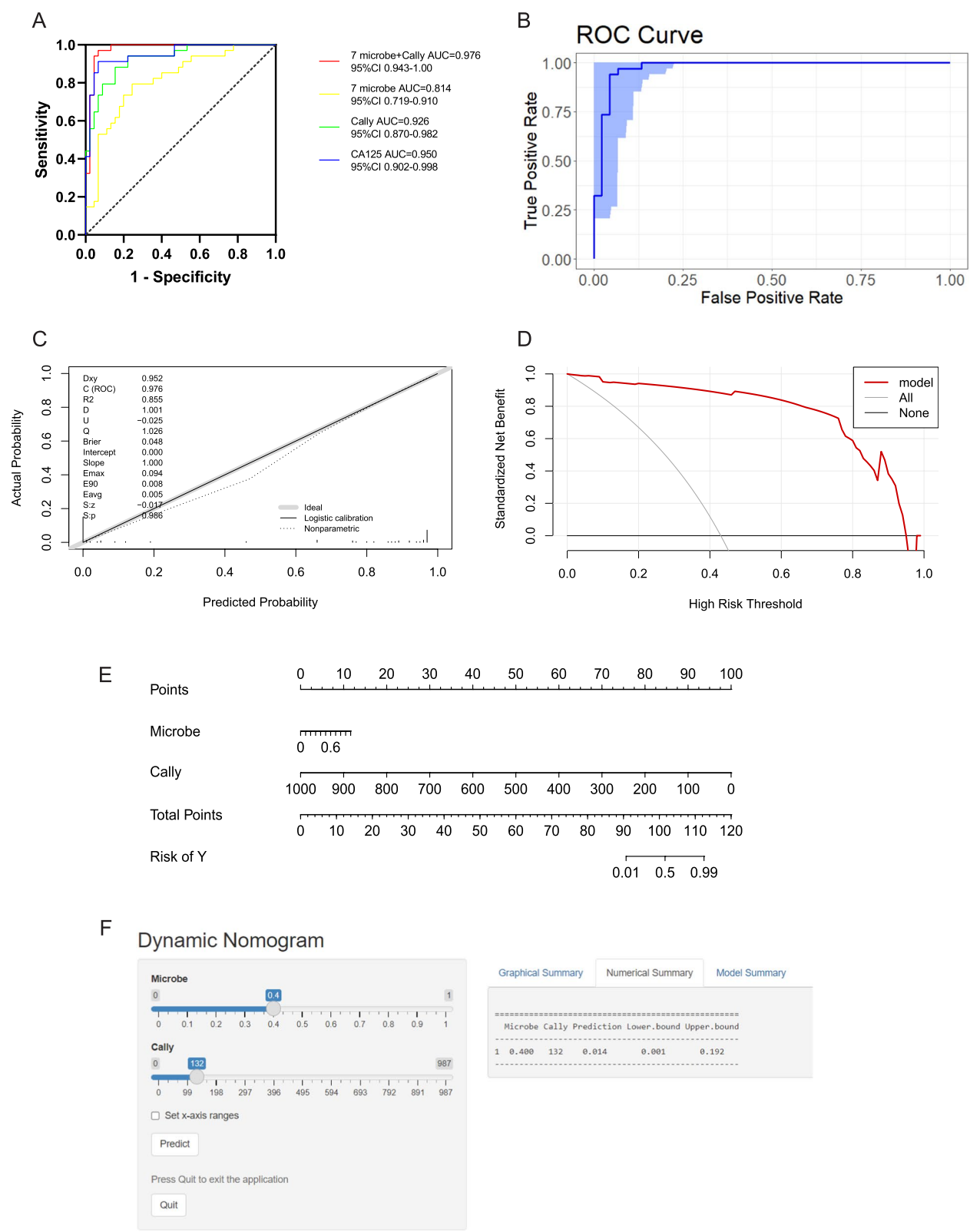
To elucidate the metabolite differences between the EOC group and the NT group, the following analyses were conducted. The Venn diagram analysis revealed that 141 metabolites were unique to the EOC group, whereas 68 metabolites were specific to the NT group (Fig. 10A). PLS-DA revealed significant differences in metabolite composition between the two groups (Fig. 10B, C). The volcano plot analysis revealed that, compared with the NT group, 82 metabolites were significantly upregulated and 151 metabolites were significantly downregulated in the EOC group (Fig. 10D). Using the MetOrigin platform [11] (<https://metorigin.met-bioinformatics.cn/home/>), we conducted a source analysis of the differentially expressed metabolites. The results indicated that these metabolites were primarily derived from food, microbiota, drugs, host pathways, and others (Fig. 10E). We further analyzed the 72 metabolites of microbial origin and found that 64 metabolites were associated with co-metabolism, whereas only 8 metabolites were exclusively of microbial origin. These included HMDB0243492, C09265, HMDB0242046, HMDB0240606, C04171, HMDB0060557, HMDB0252797, and C21412 (Fig. 10F). Finally, we conducted a heatmap analysis of the eight metabolites in conjunction with the previously identified microorganisms. The results showed that HMDB0240606 was positively correlated with *Escherichia* and *Shigella*,

(See figure on next page.)

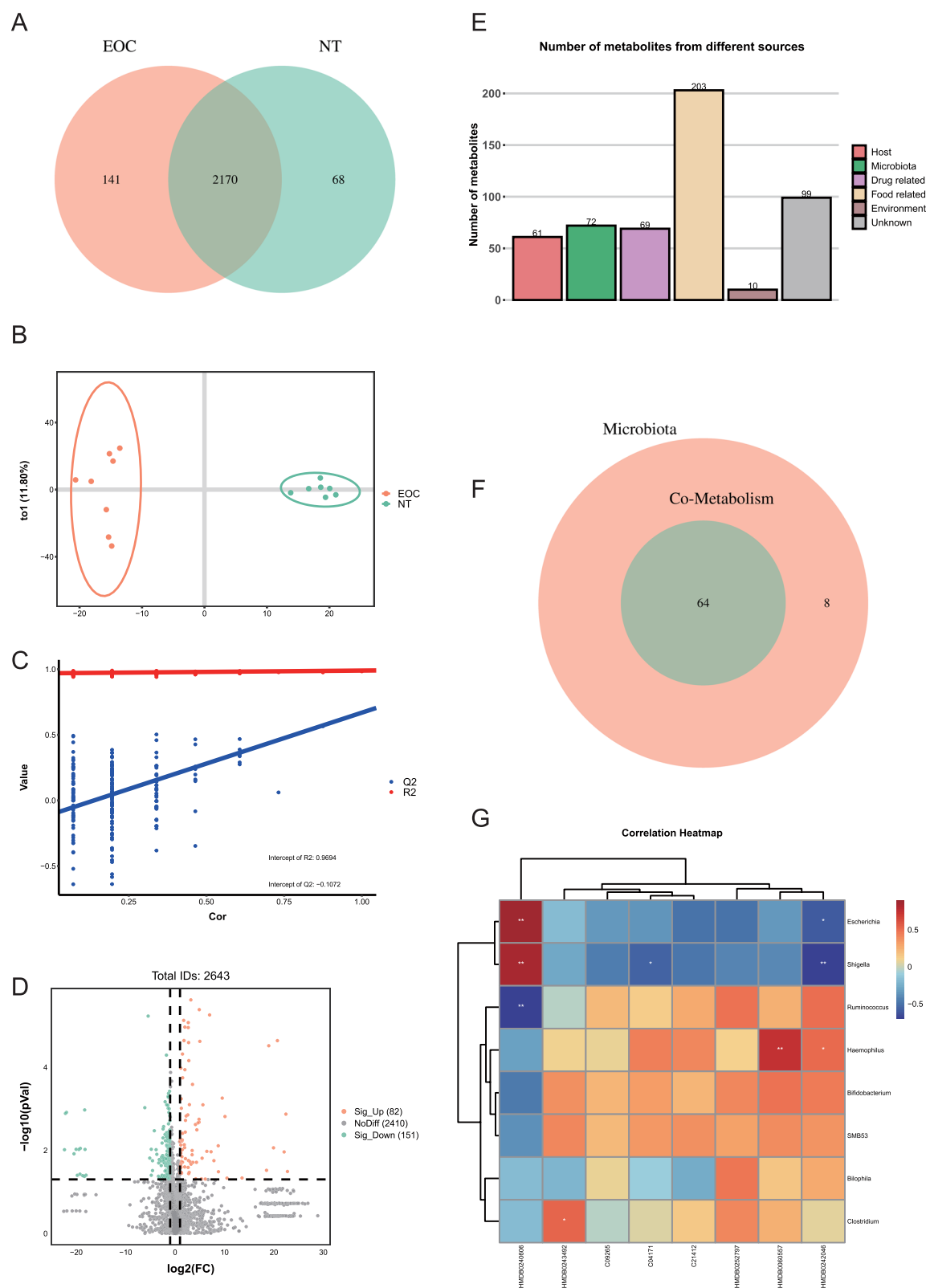
**Fig. 9** Construction and validation of the diagnostic model. **A** Receiver operating curves for the Microbial, Cally, Microbial-Cally joint indicator, and CA125. **B** The AUC for Microbial-Cally diagnostic model bootstrap internal validation. **C** Calibration curve of Hosmer–Lemeshow test.

**D** Decision curve analysis (DCA) of the Microbial-Cally diagnostic model showed the net benefit of using the model to diagnose EOC. **E**

A Microbial-Cally-Age-based nomogram for predicting the occurrence of EOC. **F** Web-based risk calculator (Dynamic Nomogram: <https://dcycc20200805.shinyapps.io/dynnomapp/>) to predict the incidence rate of EOC



**Fig. 9** (See legend on previous page.)



**Fig. 10** Classification of metabolites in EOC patients and NTs. **A** Venn diagram of the two groups of metabolites. **B** PLS-DA analysis of metabolites among the two groups. **C** OPLS-DA model test plot. **D** Volcano plot analysis. **E** Bar graph of metabolite traceability. **F** Venn diagram of metabolites associated with microorganisms. **G** Heat map analysis of metabolites and microorganisms

while it was negatively correlated with *Ruminococcus* (Fig. 10G).

#### Effect of bacteria on migration and proliferation of ovarian cancer cells in vitro

The human ovarian cancer cell lines SKOV3 were utilized in our study. The results of the wound healing assay demonstrated that the healing rate of SKOV3 cells was significantly increased after co-culture with 5% and 10% *E. coli* supernatants (Fig. 11A, B). CCK-8 results demonstrated that *E. coli* supernatant promoted the proliferation of SKOV3 cells, with the effect becoming more pronounced at higher concentrations (Fig. 11C). Additionally, the results of the Transwell migration and invasion assays demonstrated that a 10% concentration of *E. coli* supernatant promoted the migration and invasion of SKOV3 cells (Fig. 11S–G). Moreover, a 10% concentration of *E. coli* supernatant also inhibited the apoptosis rate of SKOV3 cells (Fig. 11H, I). However, the effects of *Bifidobacterium longum subsp. longum* Reuter supernatant on SKOV3 cell proliferation, migration, invasion, and apoptosis were opposite to those observed with *E. coli* supernatant (Additional file 5).

#### Discussion

The gut microbiota represents the intricate genetic composition of all microorganisms inhabiting the human gastrointestinal tract, including bacteriophages, viruses, bacteria, protists, helminths, and fungi, which are integral to the maintenance of human health. Recent investigations have elucidated that intestinal microbiota can potentiate insulin sensitivity through the modulation of carbohydrate metabolism [12], and uphold intestinal barrier integrity by diminishing pathogen colonization [13, 14]. Moreover, there are significant differences in gut microbial profiles between patients and healthy individuals. For instance, the *Blautia* species, specifically *B. caecimuris* and *B. producta*, along with the genera *Lachnospirillum* and *Oscilibacter*, as well as *Dialister invisus*, were observed to be significantly enriched in individuals diagnosed with depression [15]; Two *Streptococcus* species, commonly detected in the oral cavity, exhibited a significant enrichment in individuals diagnosed with type 2 diabetes [16]. Accordingly, we commenced a comprehensive analysis of the variations in gut

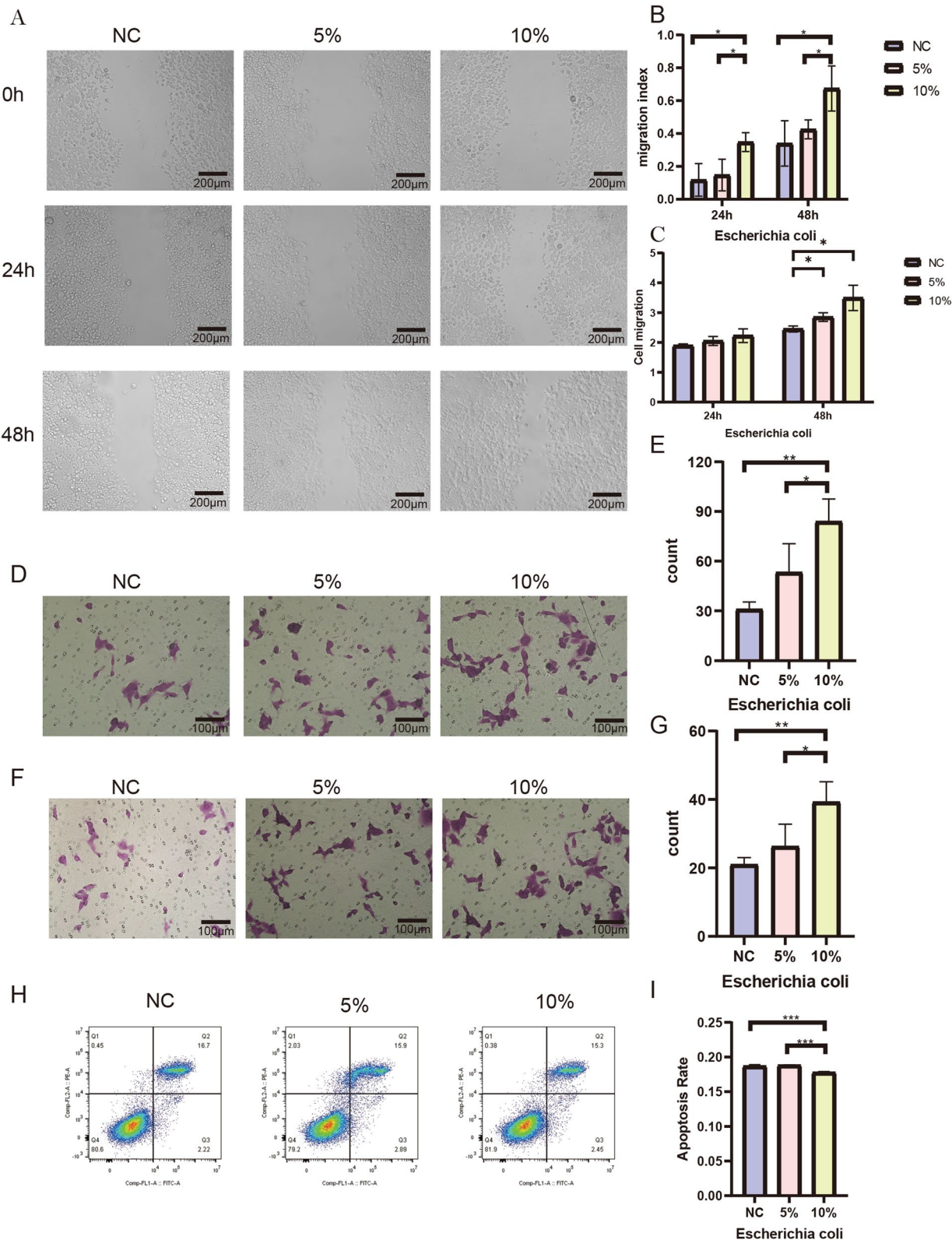
microbiota profiles between ovarian cancer patients and other demographic cohorts.

Our investigation demonstrated that relative to healthy controls and patients with benign ovarian tumors, individuals diagnosed with epithelial ovarian cancer displayed a significantly reduced alpha diversity of intestinal microbiota, an increased prevalence of opportunistic pathogens, a decreased abundance of probiotics, and a diminished B/E ratio. Transcriptome-wide sequencing data from The Cancer Genome Atlas (TCGA) revealed that, across 11 distinct cancer types, the alpha diversity index exhibited a statistically significant disparity between tumor tissue and adjacent normal tissue [17]. It is posited that notable disparities in diversity exist between cancer patients and healthy individuals, this is in accordance with the findings of our investigation. Nonetheless, our study did not demonstrate a decrease in the expression of *Akkermansia* in ovarian cancer, which contrasts with the findings presented by other scholars [18, 19]. In light of the regional disparities inherent in the sampled populations, it is essential to acknowledge that regional and environmental factors cannot be disregarded.

Based on the CH-index results, the enterotypes of our study population can be classified into four distinct categories. Enterotype is a classification concept based on the structural characteristics and compositional diversity of intestinal microbial communities initially proposed by Arumugam et al. in 2011. This analytical framework is invariant to variables including age, gender, cultural background, and geographical location [20]. According to findings published in Nature Microbiology, the enterotypes serves as a valuable diagnostic tool for evaluating an individual's disease status and may also act as a significant indicator of risk or susceptibility to various human conditions. Additionally, variations within the enterotypes can modulate the metabolism of diverse compounds [21]. Yu et al. reported that the proportion of enterotype in patients with thyroid cancer exhibited a significant deviation from that observed in a healthy control population, thereby implying alterations in the enterotype among cancer-affected individuals [22]. In our study, we found that the distribution of enterotypes in patients with epithelial ovarian cancer (EOC) significantly diverged from that observed in the healthy control

(See figure on next page.)

**Fig. 11** *E. coli* supernatants were co-cultured with SKOV3. **A** Scratch assay of *E. coli* and SKOV3 cells. **B** Comparison of healing rates among the three groups. **C** Comparison of the proliferative activity of the three groups. **D** Transwell assay of *E. coli* and SKOV3 cells. **E** Comparison of Transwell migration bar charts among the three groups. **F** Diagram of Transwell invasion assay. **G** Bar graph of Transwell invasion among the three groups. **H** Cell apoptosis was detected by Annexin V/PI assay. **I** Bar graph of apoptosis among the three groups. \*\*\* represented  $p < 0.05$ , \*\*\*\* represented  $p < 0.01$ , \*\*\*\*\* represented  $p < 0.001$





population, characterized by a marked predominance of enterotype 1 (predominantly expressed in *Escherichia*). This finding further highlights the alterations present within the intestinal microbiota of individuals diagnosed with epithelial ovarian cancer. Charles M. Rudin reported that in a cohort of non-small cell lung cancer patients undergoing treatment with immune checkpoint inhibitors, those with positive *Escherichia coli* demonstrated significantly longer median overall survival (OS) and progression-free survival (PFS) compared to those with negative *Escherichia coli* [23]. Moreover, Oncomicrobial Community Subtype 3 (OCS3), characterized by a predominance of Proteobacteria, has been recognized as a critical determinant influencing the prognosis of colorectal cancer (CRC) [24]. Based on the aforementioned studies, we conducted a survival analysis of EOC patients categorized by enterotypes and *Escherichia* expression levels. However, the PFS of patients with enterotype 1 and high *Escherichia* expression did not show significant differences when compared to other groups, which is insufficient to support the use of *Escherichia* as a prognostic marker for EOC, this finding may be ascribed to the constraints of our sample size, in conjunction with the number of tumor stages present among EOC patients.

The parameters of the TB group were positioned between those of the Nor group and the EOC group. Incorporating the results from LEfSe and Venn diagrams, we ultimately merged the TB group with the Nor group into an NT group, subsequently analyzing it alongside the EOC group. In comparison to the NT group, the EOC group exhibited a significant increase in the abundance of *Escherichia* and *Shigella*, while *Ruminococcus* showed a notable decrease. Furthermore, *Escherichia*, *Shigella*, and *Ruminococcus* were associated with inflammation to varying extents. The lipopolysaccharides (LPS) produced by *Escherichia* can surpass the regulatory capacity of the host immune system, resulting in an excessive inflammatory response that compromises the integrity of the intestinal barrier and further disrupts the balance of gut microbiota. Certain members of *Ruminococcus* are known to produce short-chain fatty acids, which exhibit notable anti-inflammatory properties. Recent studies have indicated that *Escherichia coli* modifies the tumor microenvironment, thereby facilitating colorectal cancer progression and conferring resistance to chemotherapy [25]. *Shigella* predominantly infects the human gastrointestinal tract and demonstrates a capacity to attenuate the bactericidal function of natural killer (NK) cells [26]. Clinical data revealed that the inflammation-associated markers: WBC, NEU, and levels of hs-CRP in patients diagnosed with EOC were significantly elevated compared to those observed in NT subjects.

Inflammation is a pivotal hallmark of EOC, recent studies have demonstrated that pelvic inflammation significantly enhances the risk of EOC development [27]. Inflammation-associated risk scores were significantly correlated with reduced OS in black patients with EOC [28]. The Cally index consists of CRP, ALB, and LYM, which indicate inflammation levels, nutritional status, and immune function. Cally has been recognized as a promising prognostic biomarker for EOC following surgical intervention [29]. Correlation analysis showed a positive association between *Shigella* and hs-CRP, and a negative association with Cally. Additionally, *Bifidobacterium* had an inverse correlation with hs-CRP. Thus, we suggest that changes in gut microbiota may influence inflammation onset and progression, which can facilitate EOC development.

Moreover, notable differences in clinical indicators were found between the EOC and NT groups, including CA125, TSGF, TG, and other markers, among them, CA125 is an effective method for the detection of EOC [30, 31], it is used in the follow-up management of ovarian cancer patients. The heat map and Sankey diagram illustrated correlations among flora, signaling pathways, CA125 levels, and other clinical indicators. WGCNA is a systems biology method for elucidating gene association patterns across various samples, used to be used to investigate therapeutic targets of HCC [32], gene modules related to bone mineral density (BMD) [33], and those associated with depression [32]. We used WGCNA analysis to clarify the relationships between modules and clinical indicators. Additionally, the correlation network and heatmap provided insights into the intricate interactions among microbiota. Prior research suggests that changes in intestinal microbial diversity or community composition in tumor and healthy populations are not due to a single bacterial genus [34–36]. Consequently, we propose that the pathological state results from the disruption of the gut microbiota's overall balance.

Recent research shows that gut microbiota can act as diagnostic biomarkers for cancer [37, 38], and may serve as a biomarker for the next generation of cancer immunotherapy [39]. As a sophisticated artificial intelligence technology, machine learning has found extensive application in the identification of diagnostic markers for various diseases [40–42]. In this investigation, we utilized two well-established machine learning methodologies, LASSO and SVM, to discern potential biomarkers. In conclusion, we have established the pioneering Microbial-Cally model for diagnosing epithelial ovarian cancer. Subsequent validation revealed that the model exhibited high robustness and consistency in performance. This model presents the advantages of non-invasive gut microbiota analysis and easy-to-obtain Cally index in

clinical practice, moreover, it exhibits superior predictive efficacy.

Metabolomics analyzes small molecules from cellular metabolism, providing insights into biochemical reactions in living systems. Recent studies show distinct metabolomic profiles in tumor patients [43, 44] compared to healthy individuals, indicating its potential as a biomarker for cancer diagnosis. Our results also found that there were metabolite differences between the EOC group and the NT group, and some metabolites were related to microorganisms. Nevertheless, metabolomics was not incorporated into the diagnostic model, as a comprehensive metabolomic analysis of all samples was not conducted.

Ultimately, our findings revealed that the supernatant of *E. coli* significantly enhanced the proliferation and migration of SKOV3 cells while concurrently inhibiting their apoptosis. However, the opposite result was observed for *Bifidobacterium*. Apoptosis is a form of programmed cell death that is genetically regulated. It plays a crucial role in organismal growth and development, as well as in maintaining tissue and organ homeostasis. Additionally, apoptosis is involved in numerous pathological processes. Numerous studies have demonstrated that the majority of anticancer therapies eradicate tumor cells through the induction of apoptosis [45]. Previous studies have demonstrated that sodium citrate exhibits anti-tumor activity through the induction of apoptosis in ovarian cancer cells [46]. In conjunction with our experimental findings, we propose that *E. coli* and *Bifidobacterium* may modulate pro- or anti-tumor activities by influencing apoptosis via their respective metabolites. Unfortunately, we were unable to conduct in vivo experiments to validate this hypothesis.

This study acknowledges several limitations, having been conducted as a single-center, cross-sectional analysis with a relatively limited cohort size, unable to investigate the longitudinal dynamics of microbiota or their modifications subsequent to the onset of cancer. Secondly, we exclusively performed 16S rRNA sequencing on the collected stool samples and did not conduct metabolite profiling or whole genome sequencing analysis, which constrained our capacity to offer a comprehensive insight into the pathogenesis. In addition, we only conducted cell experiments, but not in vivo experiments. In future research, it is imperative to undertake large-scale, multi-center studies and to integrate metagenomic and multi-omics analyses in order to accurately elucidate the distinctions in gene function and metabolic pathways. Furthermore, both in vitro and in vivo experiments should be conducted to verify, thereby enhancing our comprehensive understanding of disease mechanisms.

## Conclusion

In summary, our current study elucidates that epithelial ovarian cancer patients demonstrate diminished microbial diversity and pronounced alterations in community composition, predominantly marked by an enrichment of opportunistic pathogens alongside a reduction in beneficial probiotics. Moreover, a significant correlation is observed between microbiota and clinical traits, which is not exclusively determined by any individual genus but rather emerges from the complex interactions among diverse microbial communities. In conclusion, we established a non-invasive diagnostic model leveraging gut microbiota and Cally, which presents a novel methodology for the diagnosis of epithelial ovarian cancer.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06339-z>.

Supplementary Material 1.  
Supplementary Material 2.  
Supplementary Material 3.  
Supplementary Material 4.  
Supplementary Material 5.

## Acknowledgements

We wish to express our profound appreciation to all participants who contributed to this experiment. CC, CD, YL, WX, LF, XJ used to conduct research and academic endeavors at The First Affiliated Hospital of University of South China, expressing our appreciation to the institution for its provision of a robust scientific research platform and invaluable educational opportunities.

## Author contributions

CC, CD, WX, XJ conceived and designed the experiments; CC, CD, YL, YHL, XJ analyzed the data and visualization; CD, SH, YZ, YW, SP, XJ helped with reagents/materials/analysis tools; CC, YL, WX, LF collected the clinical samples, underwent the ethical evaluation process, and obtained informed consent from patients before conducting questionnaire data sheets; CC, YL, YHL, SH, SP, YZ, YW, XJ project administration, writing—review & editing, Funding acquisition. All authors have contributed to the manuscript's writing and approved the final manuscript.

## Funding

The study was supported by the Hunan Natural Science Foundation (2023JJ60350), Shenzhen Science and Technology Program (JCYJ20210324131414040), and Hunan Natural Science Foundation (2019JJ80064).

## Availability of data and materials

The datasets presented in this study can be found in online repositories. The names of the repository/repository and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/sra/PRJNA1152737>.

## Declarations

### Ethics approval and consent to participate

This study was approved by the Ethics Committee of The First Affiliated Hospital of University of South China.

### Consent for publication

Not applicable.

# Competing interests

The author reports no competing interest in this work.

# Author details

<sup>1</sup>Department of Clinical Laboratory, The First Affiliated Hospital of Hunan Traditional Chinese Medical College, Hunan Province Directly Affiliated TCM Hospital, Zhuzhou, China. <sup>2</sup>Department of Clinical Laboratory, People's Hospital of Longhua, Shenzhen, China. <sup>3</sup>Hunan Traditional Chinese Medical College, Zhuzhou, China. <sup>4</sup>Department of Clinical Laboratory Shenzhen Longhua Maternity and Child Healthcare Hospital, Shenzhen, China. <sup>5</sup>Changsha Center for Disease Prevention and Control, Changsha, China. <sup>6</sup>Xuzhou First People's Hospital, Xuzhou, China.

Received: 14 September 2024 Accepted: 3 March 2025

Published online: 13 March 2025

# References

- Webb P, Jordan S. Global epidemiology of epithelial ovarian cancer. *Nat Rev Clin Oncol*. 2024. <https://doi.org/10.1038/s41571-024-00881-3>.
- Torre L, Trabert B, DeSantis C, et al. Ovarian cancer statistics, 2018. *CA Cancer J Clin*. 2018. <https://doi.org/10.3322/caac.21456>.
- Karla AL, Andrew Maltez T, Laura AB, et al. Cross-cohort gut microbiome associations with immune checkpoint inhibitor response in advanced melanoma. *Nat Med*. 2022. <https://doi.org/10.1038/s41591-022-01695-5>.
- Vinod KG, Kevin YC, Benjamin H, et al. Gut microbial determinants of clinically important improvement in patients with rheumatoid arthritis. *Genome Med*. 2021. <https://doi.org/10.1186/s13073-021-00957-0>.
- Xinyu Q, Chuyu Y, Lulu S, et al. Gut microbiota-bile acid-interleukin-22 axis orchestrates polycystic ovary syndrome. *Nat Med*. 2019. <https://doi.org/10.1038/s41591-019-0509-0>.
- Robert FS, Tim FG. Gut microbiome in HCC—mechanisms, diagnosis and therapy. *J Hepatol*. 2020. <https://doi.org/10.1016/j.jhep.2019.08.016>.
- David J, Kathleen M, Camille G, et al. Shifts in gut and vaginal microbiomes are associated with cancer recurrence time in women with ovarian cancer. *PeerJ*. 2021. <https://doi.org/10.7717/peerj.11574>.
- Shuyun X, Zhenzhen L, Meihua L, et al. Intestinal dysbiosis promotes epithelial–mesenchymal transition by activating tumor-associated macrophages in ovarian cancer. *Pathog Dis*. 2019. <https://doi.org/10.1093/femspd/ftz019>.
- Bo Y, Congzhou L, Sean CP, et al. Identification of fallopian tube microbiota and its association with ovarian cancer. *Elife*. 2024. <https://doi.org/10.7554/eLife.89830>.
- Lyu F, Han F, Ge C, et al. OmicStudio: a composable bioinformatics cloud platform with real-time feedback that can generate high-quality graphs for publication. *iMeta*. 2023. <https://doi.org/10.1002/imt2.85>.
- Gang Y, Cuifang X, Xiaoyan W, et al. MetOrigin 2.0: advancing the discovery of microbial metabolites and their origins. *iMeta*. 2025. <https://doi.org/10.1002/imt2.246>.
- Takeuchi T, Kubota T, Nakanishi Y, et al. Gut microbial carbohydrate metabolism contributes to insulin resistance. *Nature*. 2023. <https://doi.org/10.1038/s41586-023-06466-x>.
- Spragge F, Bakkeren E, Jahn M, et al. Microbiome diversity protects against pathogens by nutrient blocking. *Science (New York NY)*. 2023. <https://doi.org/10.1126/science.adj3502>.
- Gavzy S, Kensiski A, Lee Z, et al. Bifidobacterium mechanisms of immune modulation and tolerance. *Gut Microbes*. 2023. <https://doi.org/10.1080/19490976.2023.2291164>.
- Bosch J, Nieuwdorp M, Zwinderman A, et al. The gut microbiota and depressive symptoms across ethnic groups. *Nature Commun*. 2022. <https://doi.org/10.1038/s41467-022-34504-1>.
- Mei Z, Wang F, Bhosle A, et al. Strain-specific gut microbial signatures in type 2 diabetes identified in a cross-cohort analysis of 8,117 metagenomes. *Nat Med*. 2024. <https://doi.org/10.1038/s41591-024-03067-7>.
- Sheng D, Jin C, Yue K, et al. Pan-cancer atlas of tumor-resident microbiome, immunity and prognosis. *Cancer Lett*. 2024. <https://doi.org/10.1016/j.canlet.2024.217077>.
- Wang Z, Qin X, Hu D, et al. Akkermansia supplementation reverses the tumor-promoting effect of the fecal microbiota transplantation in ovarian cancer. *Cell Rep*. 2022. <https://doi.org/10.1016/j.celrep.2022.111890>.
- Hu X, Xu X, Zeng X, et al. viaGut microbiota dysbiosis promotes the development of epithelial ovarian cancer regulating Hedgehog signaling pathway. *Gut Microbes*. 2023. <https://doi.org/10.1080/19490976.2023.2221093>.
- Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. *Nature*. 2011. <https://doi.org/10.1038/nature09944>.
- Costea P, Hildebrand F, Arumugam M, et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol*. 2018. <https://doi.org/10.1038/s41564-017-0072-8>.
- Yu X, Jiang W, Kosik R, et al. Gut microbiota changes and its potential relations with thyroid carcinoma. *J Adv Res*. 2022. <https://doi.org/10.1016/j.jare.2021.04.001>.
- Elkrief A, Montesio M, Sivakumar S, et al. Intratumoral Escherichia is associated with improved survival to single-agent immune checkpoint inhibition in patients with advanced non-small-cell lung cancer. *J Clin Oncol*. 2024. <https://doi.org/10.1200/jco.23.01488>.
- Mouradov D, Greenfield P, Li S, et al. Oncomicrobial community profiling identifies clinicomolecular and prognostic subtypes of colorectal cancer. *Gastroenterology*. 2023. <https://doi.org/10.1053/j.gastro.2023.03.205>.
- de Oliveira Alves N, Dalmasso G, Nikitina D, et al. Escherichia coli The colibactin-producing alters the tumor microenvironment to immunosuppressive lipid overload facilitating colorectal cancer progression and chemoresistance. *Gut Microbes*. 2024. <https://doi.org/10.1080/19490976.2024.2320291>.
- Hansen J, de Jong M, Wu Q, et al. Pathogenic ubiquitination of GSDMB inhibits NK cell bactericidal functions. *Cell*. 2021. <https://doi.org/10.1016/j.cell.2021.04.036>.
- McAlpine J, Lisonkova S, Joseph K, et al. Pelvic inflammation and the pathogenesis of ovarian cancer: a cohort study. *Int J Gynecol Cancer*. 2014. <https://doi.org/10.1097/igc.0000000000000235>.
- Johnson C, Albergo A, Bandera E, et al. Association of inflammation-related exposures and ovarian cancer survival in a multi-site cohort study of Black women. *Br J Cancer*. 2023. <https://doi.org/10.1038/s41416-023-02385-w>.
- Wang W, Gu J, Liu Y, et al. Pre-treatment CRP-albumin-lymphocyte index (CALLY Index) as a prognostic biomarker of survival in patients with epithelial ovarian cancer. *Cancer Manage Res*. 2022. <https://doi.org/10.2147/cmar.s359968>.
- Funston G, Hamilton W, Abel G, et al. The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study. *PLoS Med*. 2020. <https://doi.org/10.1371/journal.pmed.1003295>.
- Karam A, Karlan B. Ovarian cancer: the duplicity of CA125 measurement. *Nat Rev Clin Oncol*. 2010. <https://doi.org/10.1038/nrclinonc.2010.44>.
- Wang W, Li W, Wu Y, et al. Genome-wide DNA methylation and gene expression analyses in monozygotic twins identify potential biomarkers of depression. *Transl Psychiatry*. 2021. <https://doi.org/10.1038/s41398-021-01536-y>.
- Farber C. Identification of a gene module associated with BMD through the integration of network analysis and genome-wide association data. *J Bone Mineral Res*. 2010. <https://doi.org/10.1002/jbmr.138>.
- Wong S, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. *Nat Rev Gastroenterol Hepatol*. 2019. <https://doi.org/10.1038/s41575-019-0209-8>.
- Ni B, Kong X, Yan Y, et al. Combined analysis of gut microbiome and serum metabolomics reveals novel biomarkers in patients with early-stage non-small cell lung cancer. *Front Cell Infect Microbiol*. 2023. <https://doi.org/10.3389/fcimb.2023.1091825>.
- Ren Z, Li A, Jiang J, et al. Gut microbiome analysis as a tool towards targeted non-invasive biomarkers for early hepatocellular carcinoma. *Gut*. 2019. <https://doi.org/10.1136/gutjnl-2017-315084>.
- Yang J, Ma Y, Tan Q, et al. Streptococcus Gut is a microbial marker for the occurrence and liver metastasis of pancreatic cancer. *Front Microbiol*. 2023. <https://doi.org/10.3389/fmicb.2023.1184869>.
- Trivieri N, Pracella R, Cariglia M, et al. BRAF mutation impinges on gut microbial markers defining novel biomarkers for serrated colorectal cancer effective therapies. *J Exp Clin Cancer Res*. 2020. <https://doi.org/10.1186/s13046-020-01801-w>.
- Thomas A, Fidelle M, Routy B, et al. Gut OncoMicrobiome Signatures (GOMS) as next-generation biomarkers for cancer immunotherapy. *Nat Rev Clin Oncol*. 2023. <https://doi.org/10.1038/s41571-023-00785-8>.

40. Mitrović K, Savić A, Radojičić A, et al. Machine learning approach for Migraine Aura Complexity Score prediction based on magnetic resonance imaging data. *J Headache Pain*. 2023. <https://doi.org/10.1186/s10194-023-01704-z>.
41. Smith L, Oakden-Rayner L, Bird A, et al. Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Lancet Digital Health*. 2023. [https://doi.org/10.1016/s2589-7500\(23\)00177-2](https://doi.org/10.1016/s2589-7500(23)00177-2).
42. Jannusch K, Dietzel F, Bruckmann N, et al. Prediction of therapy response of breast cancer patients with machine learning based on clinical data and imaging data derived from breast [F]FDG-PET/MRI. *Eur J Nucl Med Mol Imag*. 2024. <https://doi.org/10.1007/s00259-023-06513-9>.
43. Sun Y, Zhang X, Hang D, et al. Integrative plasma and fecal metabolomics identify functional metabolites in adenoma-colorectal cancer progression and as early diagnostic biomarkers. *Cancer Cell*. 2024. <https://doi.org/10.1016/j.ccell.2024.07.005>.
44. Hamed M, Wasinger V, Wang Q, et al. Prostate cancer-derived extracellular vesicles metabolic biomarkers: emerging roles for diagnosis and prognosis. *J Contr Rel*. 2024. <https://doi.org/10.1016/j.jconrel.2024.05.029>.
45. Mohammad R, Muqbil I, Lowe L, et al. Broad targeting of resistance to apoptosis in cancer. *Semin Cancer Biol*. 2015. <https://doi.org/10.1016/j.semcancer.2015.03.001>.
46. Wu Y, Jia C, Liu W, et al. Sodium citrate targeting Ca/CAMKK2 pathway exhibits anti-tumor activity through inducing apoptosis and ferroptosis in ovarian cancer. *J Adv Res*. 2024. <https://doi.org/10.1016/j.jare.2024.04.033>.

## Publisher's Note

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