



OPEN Serum metabolic characteristics associated with the deterioration of colorectal adenomas

Ze Dai^{1,2,3,5,7}, Tong Li^{4,7}, Kecong Lai⁶, Xiaomei Wang^{1,2,3,5}, Peng Zhou⁵, Kefeng Hu^{1,2,3} & Yuping Zhou^{1,2,3}✉

Colorectal cancer (CRC) can evolve from colorectal adenomas, which can be further classified into non-advanced adenomas (NAAs) and advanced adenomas (AAs) based on their clinical characteristics. Their prognoses are vastly different, with patients with NAAs having significantly lower recurrence and CRC-related mortality rates than those with AA or CRC. Although serum metabolomics has shown promise for the early diagnosis of CRC, the differences in serum metabolite composition between NAA and AA still need to be further elucidated. This study aimed to explore the mechanism of CRC occurrence and development based on the unique serum metabolic fingerprints of different stages of CRC and to discover a new non-invasive diagnostic method based on serum metabolomics. A clinical CRC progression cohort containing healthy control (NC; $n=40$), NAA ($n=40$), AA ($n=40$), and CRC ($n=22$) groups was constructed, and untargeted metabolomic analysis based on liquid chromatography/mass spectrometry was performed to analyze the serum metabolite characteristics of each group. A semi-quantitative analysis of intergroup metabolite differences was conducted, focusing on specific metabolites that differed in the NAA and AA groups. Finally, variable metabolites were selected based on least absolute shrinkage and selection operator (LASSO) regression analysis, and receiver operating characteristic curves were plotted to evaluate the efficacy of the serum metabolite-based diagnostic model in distinguishing NC/NAA populations from AA/CRC populations. Metabolomic analysis revealed significant differences in the composition of metabolites in the NC and NAA groups compared to the CRC group, whereas the serum metabolites of the AA group were similar to those of the CRC group. The levels of 33 metabolites were significantly different in the serum of AA/CRC patients compared to that of NAA patients, and their functions included glycerophospholipid, sphingolipid, and caffeine metabolism. LASSO regression analysis identified 57 differential metabolite variables between the NC/NAA and AA/CRC groups. The diagnostic model constructed using the random forest algorithm had the best discrimination effect, with areas under the curve of 1.000 (95% confidence interval [CI] 1.000–1.000) and 0.685 (95% CI 0.540–0.830) for the training and testing sets, respectively. The composition of serum metabolites is specific to the different stages of CRC development. The serum metabolite composition of patients with AAs was similar to that of patients with CRC. Auxiliary diagnostic measures based on serum metabolites have the potential to guide the follow-up and treatment of patients with adenoma.

Keywords Colorectal cancer, Colorectal adenoma, Serum metabolites, Machine learning, Diagnostic model

Abbreviations

AA	Advanced adenoma
AUC	Area under the curve
CEA	Carcinoembryonic antigen
CI	Confidence interval

¹Department of Gastroenterology, The First Affiliated Hospital of Ningbo University, Ningbo 315020, Zhejiang, China. ²Institute of Digestive Disease of Ningbo University, Ningbo University, Ningbo 315020, Zhejiang, China. ³Ningbo Key Laboratory of Translational Medicine Research on Gastroenterology and Hepatology, Ningbo Key Laboratory, Ningbo 315020, Zhejiang, China. ⁴Department of Colorectal-Anal Surgery, The First Affiliated Hospital of Ningbo University, Ningbo 315020, Zhejiang, China. ⁵Health Science Center, Ningbo University, Ningbo 315211, China. ⁶Digestive Department, The Second People's Hospital of Beilun District, Ningbo 315020, Zhejiang, China. ⁷Ze Dai and Tong Li contributed equally to this work. ✉email: fyzhouyuping@nbu.edu.cn

CRC	Colorectal cancer
GBM	Gradient boosting machine
LASSO	Least absolute shrinkage and selection operator
LR	Logistic regression
NAA	Non-advanced adenoma
NC	Normal control
NNET	Neural Network Classifier
OPLS-DA	Orthogonal-corrected partial least squares discriminant analysis
PCA	Principal component analysis
RF	Random forest
ROC	Receiver operating characteristic
SVM	Support vector machine
XGB	Extreme gradient boosting

Colorectal cancer (CRC) is the third most common cancer worldwide and the second most common cause of cancer-related death^{1,2}. In 2023, 153,020 new cases of CRC were diagnosed in the United States³. Most CRCs begin with abnormal crypts, which gradually develop into adenomas after the accumulation of genetic and epigenetic mutations^{4,5} and eventually progress to CRC in 10–15 years⁶.

Adenomas can be further divided into non-advanced/low-risk adenomas (NAAs) and advanced adenomas (AAs) based on the size of the lesion and its pathological manifestations⁷. Colorectal lesions that meet any of the following conditions can be defined as AAs: (1) ≥ 10 mm in size; (2) tubular villous/villous histology; and (3) high-grade dysplasia. No significant difference was observed in the incidence of CRC or CRC-related mortality between patients with NAA and individuals without adenomas, whereas the prognosis of patients with AA is similar to that of patients with CRC^{7,8}. Therefore, AAs may be a crucial intermediate in the development of CRC.

Patients with limited early CRC lesions (American Joint Committee on Cancer stages 0, I, or II) have a good prognosis, with a 5-year survival rate of $>80\%$ ⁹; therefore, early diagnosis and intervention are crucial for prevention and treatment. Colonoscopy is the most effective diagnostic method for CRC¹⁰, as it can screen and treat patients in the early stages of disease. Unfortunately, the missed diagnosis rate of colonoscopy varies between 12 and 26%¹¹ because of the sporadic nature^{12,13} of adenomas and their small size. In addition, some patients with contraindications cannot undergo colonoscopy. Therefore, auxiliary screening tests are necessary. Several noninvasive methods have been used to detect CRC, such as the fecal occult blood and carcinoembryonic antigen (CEA) tests. However, the accuracy of these methods is relatively poor, and a novel auxiliary screening test with high accuracy needs to be explored.^{14,15}

Metabolomics is a discipline that studies small molecules (metabolites) present in biological systems at a given time point and can be used to identify changes in metabolite composition under the influence of exogenous (such as environment, lifestyle, and gut microbiota) and endogenous factors (such as genetic variations)^{16,17}. This molecular characterization reflects physiological and pathological phenotypes¹⁸. Metabolomics provides a unique understanding of the mechanisms of disease occurrence and development¹⁹. It is therefore an ideal method for examining early characteristic changes in CRC progression. There are multiple prospective and retrospective serum metabolomic studies using different metabolite identification methods and different research cohorts, which provide insights into the mechanisms of CRC occurrence, development, risk prediction, and diagnosis^{20–24}. However, current research cohorts only involve healthy controls and patients with adenomas or CRC and do not reflect the metabolic characteristics of key evolutionary stages from NAAs to AAs.

In this study, we performed a non-targeted metabolomic analysis using serum samples collected from healthy control individuals and those with NAAs, AAs, and CRC. This study aimed to explore the mechanism of CRC occurrence and development based on the unique serum metabolic fingerprints of different stages of CRC and to discover a new non-invasive diagnostic method based on serum metabolomics to differentiate high-risk populations (AA and CRC) and low-risk populations (NAA).

Results

Patient characteristics

After screening, 40 healthy individuals (NC group; age 55.94 ± 10.84 years), 40 patients with NAAs (NAA group; age 60.05 ± 9.13 years), 40 patients with AAs (AA group; age 56.45 ± 12.55 years), and 22 patients with early-stage cancer (CRC group; age 59.95 ± 9.94 years) were ultimately included. There were no statistically significant differences in age, sex, or tumor markers (CEA, CA19-9, CA125, and AFP) among the four groups (Table 1).

Serum metabolomic characteristics of the CRC progression cohort

The metabolites in the serum samples were identified and analyzed using liquid chromatography/mass spectrometry. The results of principal component analysis (PCA) and permutation multivariate analysis of variation (PERMANOVA) showed that differences in the metabolite composition among the four groups exist ($R^2 = 0.0783$; $P = 0.002$) (Fig. 1A). Further PERMANOVA between pairs showed that differences mainly existed between the NC group and NAA group ($R^2 = 0.1136$; $P = 0.002$), the NC group and AA group ($R^2 = 0.0585$; $P = 0.003$), the NC group and CRC group ($R^2 = 0.1022$; $P = 0.001$), the NAA group and CRC group ($R^2 = 0.0808$; $P = 0.006$). Unfortunately, although the inter group differences were slightly greater than the intra group differences when comparing the NAA and AA groups, this was not statistically significant ($R^2 = 0.0257$; $P = 0.117$). Unlike the NC and NAA groups, the serum metabolic profiles of patients in the AA group and early CRC group have higher homogeneity ($R^2 = 0.0123$; $P = 0.503$) (Table 2). Moreover, these changes in serum metabolite composition during the occurrence and development of cancer are synchronous and continuous (Fig. 1B).

Group	NC	NAA	AA	CRC	Total
Sample size	40	40	40	22	142
Gender (male/female)	21/19	30/10	27/13	14/8	95/50
Age(years)	55.94 ± 10.84	60.05 ± 9.13	56.45 ± 12.55	59.95 ± 9.94	
< 45	2	2	5	2	11
≥ 45 and ≤ 60	27	19	20	8	74
> 60	11	19	15	12	57
Tumor marker					
CEA	2.09 ± 1.76	2.23 ± 1.17	1.91 ± 1.31	2.35 ± 1.66	
CA 19–9	15.22 ± 33.42	12.32 ± 13.11	14.79 ± 36.96	11.87 ± 10.39	
CA 125	14.10 ± 20.61	8.63 ± 6.36	10.23 ± 5.35	7.20 ± 3.85	
AFP	3.42 ± 2.69	4.27 ± 2.11	3.36 ± 1.50	3.06 ± 1.17	

Table 1. General information of the subjects. NC: normal control; NAA: non-advanced adenoma; advanced adenoma; colorectal cancer; CEA: carcinoembryonic antigen; CA 19–9: carbohydrate antigen 19–9; CA 125: carbohydrate antigen 125; AFP: alpha-fetoprotein.

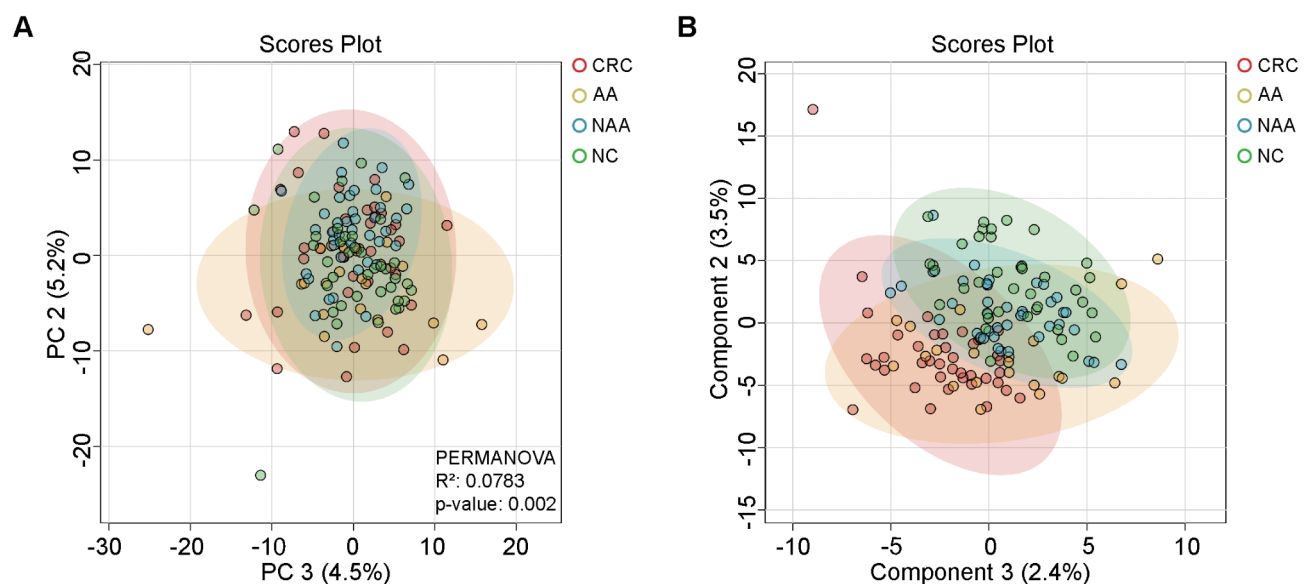


Fig. 1. The composition of serum metabolites changes during the occurrence and development of colorectal cancer. (A) Principal component analysis; (B) partial least squares discriminant analysis. NC: normal control; NAA: non-advanced adenoma; AA: advanced adenoma; CRC: colorectal cancer; PERMANOVA: permutational multivariate analysis of variance.

Groups	F-value	R-squared	p-value
NC vs. NAA	9.872	0.1136	0.002
NC vs. AA	4.726	0.0585	0.003
NC vs. CRC	6.721	0.1022	0.001
NAA vs. AA	2.032	0.0257	0.117
NAA vs. CRC	5.277	0.0808	0.006
AA vs. CRC	0.737	0.0123	0.503

Table 2. Permutational multivariate analysis of variance. NC: normal control; NAA: non-advanced adenoma; AA: advanced adenoma; CRC: colorectal cancer

Key metabolites associated with NAA deterioration

To determine the characteristic serum metabolites associated with the occurrence and development of colorectal adenomas, inter group comparisons were conducted for each metabolite. However, no metabolite content was found to gradually increase or decrease in the four groups with statistical significance.

Given that the serum metabolite compositions of patients with AA and CRC were relatively similar, and the prognoses of the two patient groups are also similar^{7,8}. We combined data from the two groups to identify key serum metabolites that changed during NAA progression. After comparing the semi-quantitative data of serum metabolites in the NC, NAA, and AA + CRC groups, 33 metabolites in the AA + CRC group changed compared to the NAA group; the abundances of 30 metabolites increased (Table 3). The functions of these metabolites mainly involved the glycerophospholipid metabolism, the sphingolipid signaling pathway, and caffeine metabolism (Fig. 2A–B).

Construction of the diagnostic model

To evaluate the potential of serum metabolites as diagnostic biomarkers for AA/CRC, we randomly divided the NC/NAA groups (n = 80) and the AA/CRC groups (n = 62) into training (55/41) and test sets (25/21) in a 2:1 ratio to construct and evaluate the diagnostic model. First, LASSO was used to identify serum metabolite variables in patients with AA/CRC compared to NAA/healthy controls. A total of 57 metabolites in the training set had the potential to serve as diagnostic biomarkers. Next, diagnostic models were constructed using the LR, SVM, GBM, NNET, RF, and XGB algorithms, and ROC curves were plotted. The RF and XGB algorithms performed best on the training set, with area under the curve (AUC) reaching 1.000 (95% confidence interval [CI], 1.000–1.000) for both. The diagnostic models using the LR (AUC = 0.708, 95% CI [0.570–0.847]), NNET (AUC = 0.729, 95% CI [0.595–0.863]), and RF (AUC = 0.685, 95% CI [0.540–0.830]) algorithms performed better on the test set (Fig. 3A–B). Overall, the model generated using the RF algorithm exhibited the best diagnostic performance. The importance of the feature in the RF algorithm diagnostic model was ranked, with PI (18:1 [9Z]/18:1 [9Z]) and SM (d18:1/22:0) making the most prominent contributions to the model (Fig. 3C–D).

Metabolite name	NC	NAA	AA + CRC	P-value	Trend
p-Cymene	0.255 ± 0.014	0.256 ± 0.014	0.271 ± 0.038	0.0138	Up
Prolylhydroxyproline	0.012 ± 0.004	0.010 ± 0.004	0.015 ± 0.010	0.0039	Up
L-Palmitoylcarnitine	0.097 ± 0.019	0.094 ± 0.015	0.108 ± 0.026	0.0031	Up
L-Serine	0.008 ± 0.003	0.008 ± 0.003	0.010 ± 0.004	0.0428	Up
Oxypurinol	0.029 ± 0.011	0.024 ± 0.010	0.031 ± 0.01	0.0121	Up
25-Hydroxycholesterol	0.032 ± 0.010	0.031 ± 0.010	0.036 ± 0.012	0.0415	Up
Serylalanine	0.007 ± 0.003	0.007 ± 0.003	0.009 ± 0.003	0.0032	Up
2-Methylhippuric acid	0.0030 ± 0.0009	0.0032 ± 0.0008	0.0038 ± 0.0018	0.0445	Up
Lactosylceramide (d18:1/16:0)	0.040 ± 0.016	0.045 ± 0.011	0.052 ± 0.010	0.0101	Up
SM(d17:1/24:1(15Z))	0.149 ± 0.048	0.170 ± 0.042	0.197 ± 0.035	0.0027	Up
SM(d18:1/24:1(15Z))	0.896 ± 0.298	1.041 ± 0.1943	1.178 ± 0.145	0.0035	Up
SM(d18:1/20:0)	0.177 ± 0.069	0.200 ± 0.055	0.244 ± 0.051	0.0005	Up
SM(d18:1/22:0)	0.170 ± 0.048	0.185 ± 0.039	0.217 ± 0.047	0.0018	Up
PC(P-18:0/18:0)	0.016 ± 0.007	0.017 ± 0.005	0.020 ± 0.006	0.0423	Up
SM(d17:1/24:0)	0.043 ± 0.010	0.046 ± 0.009	0.051 ± 0.007	0.0071	Up
PC(20:4(8Z,11Z,14Z,17Z)/15:0)	0.115 ± 0.041	0.102 ± 0.041	0.135 ± 0.059	0.0025	Up
SM(d18:0/20:0)	0.127 ± 0.007	0.013 ± 0.007	0.018 ± 0.007	0.0007	Up
4-Dodecylbenzenesulfonic Acid	0.808 ± 0.233	0.703 ± 0.247	0.860 ± 0.285	0.0068	Up
SM(d18:1/24:0)	0.133 ± 0.039	0.152 ± 0.027	0.177 ± 0.021	< 0.0001	Up
16-Hydroxy hexadecanoic acid	0.0052 ± 0.0010	0.0051 ± 0.0012	0.0059 ± 0.0020	0.0378	Up
Deoxyuridine	0.0055 ± 0.1380	0.0054 ± 0.0014	0.0062 ± 0.0020	0.0320	Up
PIP(16:2(9Z,12Z)/22:3(10Z,13Z,16Z))	0.045 ± 0.028	0.051 ± 0.027	0.074 ± 0.058	0.0229	Up
SM(d18:0/18:0)	0.035 ± 0.025	0.043 ± 0.029	0.056 ± 0.027	0.0309	Up
LysoPE(0:0/20:4(5Z,8Z,11Z,14Z))	0.132 ± 0.036	0.122 ± 0.034	0.143 ± 0.045	0.0217	Up
N-3-Methyluridine	0.041 ± 0.006	0.041 ± 0.006	0.045 ± 0.011	0.0326	Up
Nb-Hexacosanoyltryptamine	0.0016 ± 0.0013	0.0015 ± 0.0010	0.0022 ± 0.0015	0.0270	Up
PC(14:0/P-18:0)	0.085 ± 0.023	0.093 ± 0.018	0.113 ± 0.036	0.0016	Up
Elaidic carnitine	0.147 ± 0.032	0.138 ± 0.035	0.163 ± 0.047	0.0050	Up
Caffeine	0.0068 ± 0.0007	0.0069 ± 0.0007	0.0073 ± 0.0010	0.0485	Up
Myristic acid	0.0048 ± 0.0027	0.0045 ± 0.0028	0.0061 ± 0.0039	0.0480	Up
Oxoadipic acid	0.011 ± 0.007	0.016 ± 0.011	0.010 ± 0.007	0.0018	Down
D-1-Amino-2-pyrrolidinecarboxylic acid	0.036 ± 0.010	0.034 ± 0.008	0.030 ± 0.006	0.0441	Down
13 S-hydroxyoctadecadienoic acid	0.0046 ± 0.0019	0.0049 ± 0.0016	0.0041 ± 0.0017	0.0139	Down

Table 3. Permutational multivariate analysis of variance. NC: normal control; NAA: non-advanced adenoma; AA + CRC: group of patients with advanced adenoma or colorectal cancer.

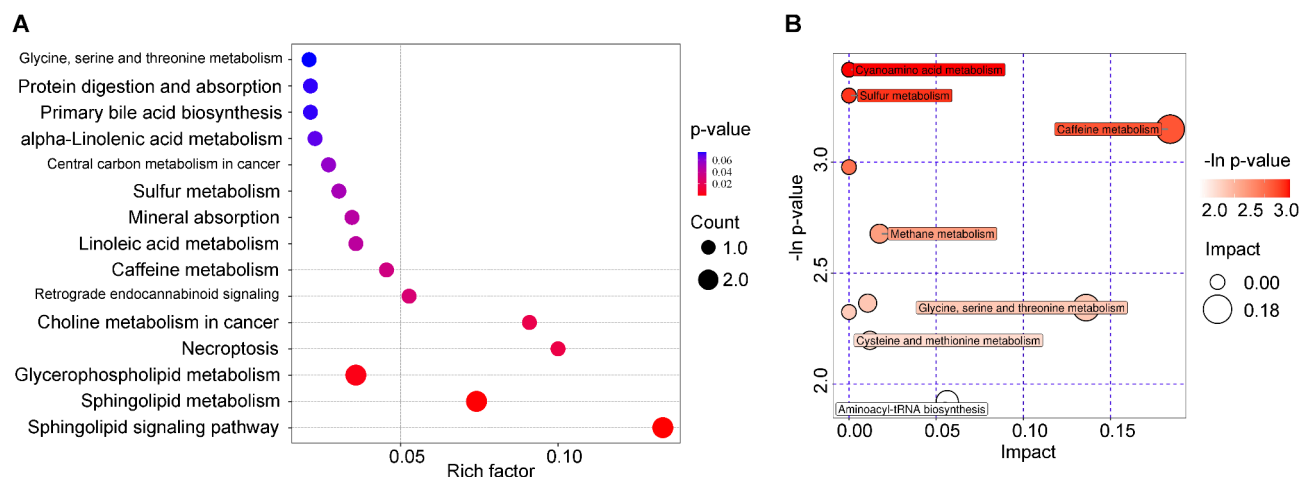


Fig. 2. Pathway and functional prediction of serum metabolites associated with non-advanced adenoma deterioration. **(A)** Functional annotation of differential metabolites based on KEGG database; **(B)** functional annotation of differential metabolites based on GO database.

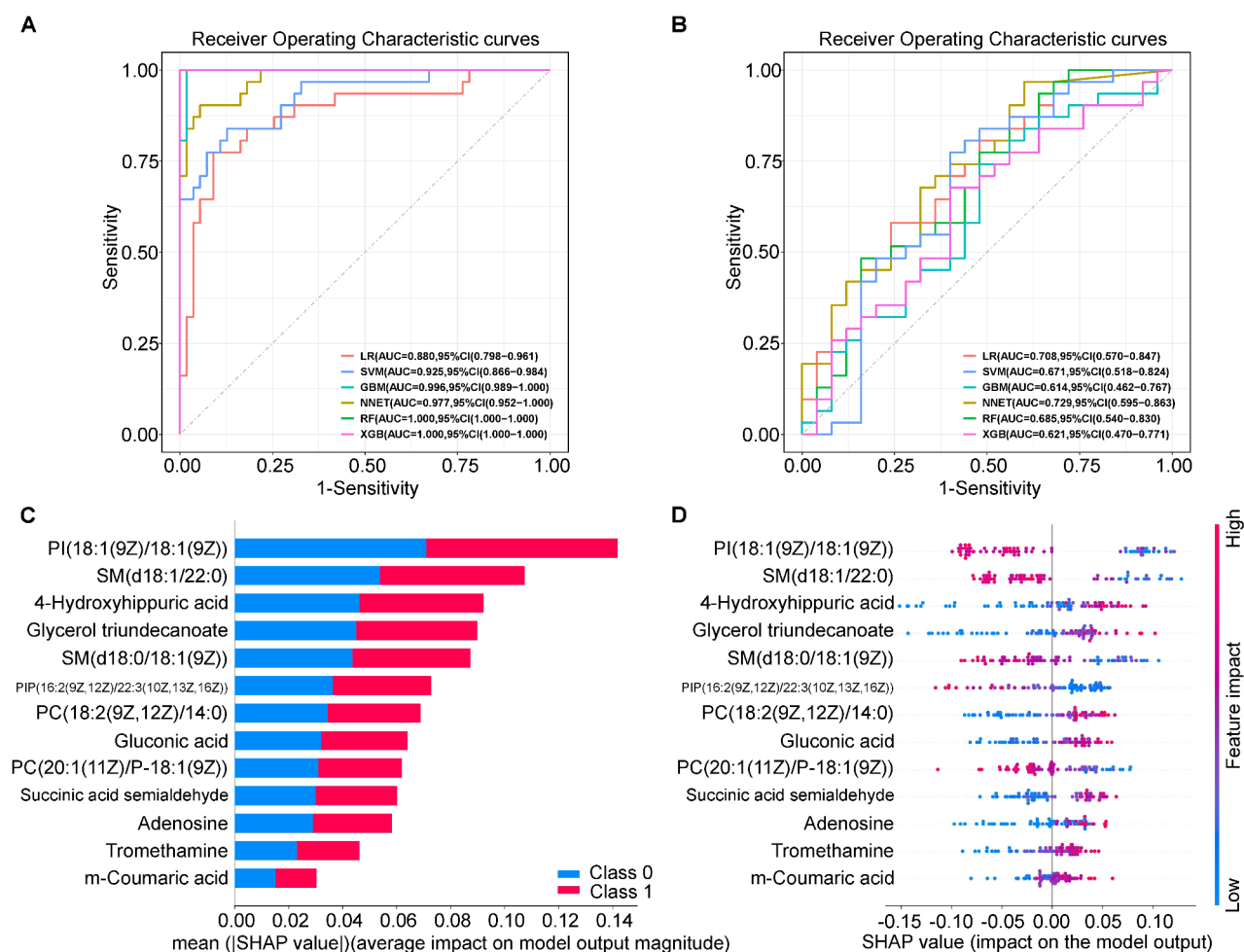


Fig. 3. Construction of a diagnostic model for patients with advanced adenomas/colorectal cancer. **(A)** Receiver operating characteristic curves of the training set; **(B)** receiver operating characteristic curves of the test set; **(C–D)** Permutation importance. LR: logistic regression; SVMs: support vector machine; GBM: gradient boosting machine; NNET: neural network classifier; RF: random forest; XGB extreme gradient boosting; AUC: area under curve; CI: confidence interval.

Discussion

A large number of studies have reported classification models for colorectal adenomas or cancers based on gut metabolites^{25,26}, circulating metabolites^{20–24}, and complex microbial metabolite characteristics, demonstrating their potential in diagnosing CRC and exploring the pathogenesis of different types of CRC²⁷. However, the differences in serum metabolomics between NAA and AA with clear significance for clinical classification require further investigation. In this study, a cohort study of CRC progression was conducted, and differences in the composition of metabolites were identified among the four serum groups.

Corresponding to the clinical outcome and prognosis⁸, the serum metabolite composition of the AA and CRC groups was similar, while the NC and NAA groups showed significant differences compared to the CRC group. A study in Austria analyzed the serum metabolome of adenomas (excluding healthy controls) and found no significant difference in the composition of metabolites between patients with AAs and NAAs²⁸. However, this study found that NAAs and AAs can be partially distinguished based on serum metabolites (but without statistical differences). A larger sample size or more precise targeted metabolomics may better explain these differences.

The occurrence and development of CRC involve multifactorial accumulation and slow progression⁶. AAs are important intermediate nodes in terms of serum metabolites. Patients with AA and CRC had significant changes in 33 metabolites compared to patients with NAA; these metabolites may indicate or affect the deterioration of NAAs. Altered metabolic functions mainly involved glycerophospholipid metabolism, sphingolipid signaling pathway, and caffeine metabolism. The metabolism of glycerophospholipids may be a hallmark of NAA deterioration. By analyzing the metabolomic differences among stage II CRC tissues, adjacent tissues, and normal tissues, Lin et al. found a significant relationship between glycerophospholipids and CRC development²⁹. In a prospective cohort of 250 patients with CRC, dysregulation of glycerophospholipids in plasma was identified as a factor that may increase the risk of CRC³⁰. Our results suggest that metabolic abnormalities in glycerophospholipids may be observed in the serum of patients with AAs at an earlier stage of the disease. Cancer cells must continuously generate glycerophospholipids for the production of cell membranes to match their high proliferative capacity, which may explain the increased glycerophospholipid content in the serum of patients with AAs³¹. At the same time, abnormal activity of the sphingomyelin metabolic pathway is one of the important serum metabolite characteristics during the deterioration of non-advanced adenomas (Fig. 2, Table 2). Elevated sphingomyelin in the cell membrane reduces membrane fluidity and permeability while increasing rigidity and strength³². This, to some extent, leads to immune suppression and abnormal proliferation. Moreover, abnormal accumulation of sphingomyelin can lead to a decrease in ceramide production. And ceramides are important signaling molecules in cancer biology, which can affect the apoptosis, proliferation, migration, aging, and inflammation of tumor cells^{33,34}.

Unlike glycerophospholipids and sphingolipid, the role of caffeine in CRC occurrence and development remains controversial. Caffeine can interact with signaling pathways such as transforming growth factor beta, phosphoinositol kinase/AKT/mammalian target of rapamycin, and mitogen-activated protein kinase, playing an important role in the onset, metastasis, and prognosis of CRC. In addition, caffeine can act as an antioxidant to protect cells from oxidative stress and as a cell cycle regulator that regulates the DNA repair system³⁵. The controversy over the role of caffeine in the progression of CRC may stem from the spatiotemporal heterogeneity of the disease progression cycle, where it may play a role in specific processes of adenoma deterioration during advanced stages.

Several studies have constructed early diagnostic models for CRC based on serum metabolites^{20–24}. However, the classification in these models has certain shortcomings, and no further division of the developmental stages of colorectal adenomas are present. As the risk of colorectal cancer in patients with AAs and NAAs are significantly different⁸, developing a convenient diagnostic model that can distinguish AAs and NAAs is necessary. Therefore, this study used LASSO to screen for variables in serum metabolites and constructed models to distinguish the NC/NAA groups and AA/CRC groups based on multiple machine learning algorithms. The RF algorithm model performed the best in both the training (AUC=1.000, 95% CI [1.000–1.000]) and test sets (AUC=0.685, 95% CI [0.540–0.830]), with PI (18:1 [9Z]/18:1 [9Z]) and SM (d18:1/22:0) being the highest contributing metabolite variables. These results indicate that the serum metabolite diagnostic model based on PI (18:1 [9Z]/18:1 [9Z]) and SM (d18:1/22:0) has the potential to be used as a supplementary diagnostic tool to provide additional information about the disease situation of patients with colorectal adenoma and suggest corresponding prevention and treatment measures.

Serum metabolites show specificity at different stages of the development of colorectal adenoma into CRC. Significant differences in the composition of serum metabolites were observed between the NC/NAA groups and CRC group, whereas the serum metabolite compositions of the AA and CRC groups were similar. Auxiliary diagnostic measures based on serum metabolites have the potential to guide the follow-up and treatment of patients with colorectal adenomas.

Limitation

However, our study may be considered preliminary, as research based on a single group may be affected by group selection and analytical biases. Independent validation in a larger, multicenter, prospective trial is necessary to explore the diagnostic performance under real-world conditions. Such large studies reduce the effect of confounding factors (e.g., age and comorbidities) that may limit the model's applicability or reduce its overall performance. Subsequently, the sources and biological functions of 33 serum metabolites related to the development of colorectal adenomas require further *in vitro* and *in vivo* experiments for interpretation.

Methods

Patients and specimen collection

This study was approved by the Medical Ethics Committee of First Affiliated Hospital of Ningbo University (KS202111002) and conducted in accordance with relevant guidelines. All participants and/or their legal guardians provide informed consent. Patients who visited the Gastroenterology Department of the First Affiliated Hospital of Ningbo University between December 2021 and January 2023 were enrolled in the study. The inclusion criteria were as follows: (1) age > 18 years and complete medical record information; (2) satisfactory bowel preparation and examination process and successful removal of tumor-like tissue; (3) clear staging of tissue pathological biopsy; and (4) provided informed consent for this study. The exclusion criteria were as follows: (1) familial polyposis of the colon, P-J syndrome, and other malignant tumors of the digestive system; (2) receiving antibiotics, immunosuppressants, or probiotics within 2 months prior to enrollment; (3) chronic diseases such as hypertension and diabetes; and (4) inability to actively cooperate. Patients with lesions < 1 cm in diameter and pathological results of tubular adenoma with or without low-grade intraepithelial neoplasia were included in the NAA group; patients with lesions ≥ 1 cm in diameter and pathological results of tubular adenoma with or without low-grade intraepithelial neoplasia or tubular villous adenoma or villous adenoma were included in the AA group; and patients with pathological results of high-grade intraepithelial neoplasia/submucosal carcinoma/submucosal superficial carcinoma were included in the CRC group.

Three milliliters of blood from each participant were centrifuged at 8000 rpm for 8 min, and the upper serum samples were collected. The serum samples were stored at -80°C until metabolomic testing was performed. Clinical data such as routine blood tests, biochemistry, coagulation function, tumor markers, pathological results, and colonoscopy reports were collected from electronic medical records.

Metabolite extraction

One hundred microliter of serum sample was transferred to an EP tube, and 400 µL of extraction solution (methanol: acetonitrile = 1:1 (V/V), with isotope-labeled internal standard mixture) was added. Subsequently, the solution was mixed by vortexing for 30 s, followed by sonication for 10 min in an ice-water bath. The mixture was left to stand at -40°C for 1 h and then centrifuged for 15 min at 4°C and 12,000 rpm (centrifugal force 13,800 × g, radius 8.6 cm). The supernatant was then added to a sample bottle and tested. An equal amount of supernatant was collected from all samples and mixed to form a QC sample for machine testing.

Metabolite analysis

The Vanquish (Thermo Fisher Scientific) ultra-high performance liquid chromatography system and Waters ACQUITY UPLC BEH Amide (2.1 mm × 100 mm, 1.7 µm) liquid chromatography column were used for chromatographic separation of the target compounds. The A phase of the liquid chromatograph was aqueous and contained 25 mmol/L ammonium acetate and 25 mmol/L ammonium hydroxide; the B phase was acetonitrile. The temperature of the sample tray was 4°C, and the injection volume was 2 µL. The Orbitrap Exploris 120 mass spectrometer is capable of collecting primary and secondary mass spectrometry data using the control software (Xcalibur, version 4.4, Thermo). The detailed parameters are as follows: Sheath gas flow rate: 50 Arb; Aux gas flow rate: 15 Arb; Capillary temperature: 320°C; Full ms resolution: 60,000; MS/MS resolution: 15,000; Collision energy: 10/30/60 in NCE mode; and Spray Voltage: 3.8 kV (positive) or -3.4 kV (negative).

Data processing

The raw data were converted into mzXML format using ProteoWizard software and processed using a self-developed R program package (with an XCMS kernel) for peak recognition, extraction, alignment, and integration. The processed data were matched with the BiotreeDB (V2.1) self-built secondary mass spectrometry database for substance annotation, and the cutoff value for algorithm scoring was set to 0.3. OmicsBean was used for pathway enrichment analysis, and MetaboAnalyst was used for principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). Permutational multivariate analysis of variance (PERMANOVA) was used to evaluate the differences in metabolomics between groups. Paired non-parametric Wilcoxon test was used to determine metabolites with significant changes, and $P < 0.05$ was considered to have statistical significance.

Construction of the diagnostic model

All clinical samples were randomly divided into training and testing sets in a 2:1 ratio, and least absolute shrinkage and selection operator (LASSO) regression analysis was used to screen variables for the samples in the training set. Logistic Regression (LR), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), Neural Network Classifier (NNET), Random Forest (RF), and Extreme Gradient Boosting (XGB) algorithms were used to construct diagnostic models and draw receiver operating characteristic (ROC) curves. Diagnostic performance was tested in the test set. These steps were performed using R (4.3.2).

Data availability

Requests to access the datasets should be directed to the corresponding author.

Received: 11 November 2024; Accepted: 20 February 2025

Published online: 26 February 2025

References

1. Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J. Clin.* **74**, 12–49 (2024).
2. Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263 (2024).
3. Siegel, R. L., Wagle, N. S., Cercek, A., Smith, R. A. & Jemal, A. Colorectal cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 233–254 (2023).
4. Feng, G. et al. Recompensation in cirrhosis: Unravelling the evolving natural history of nonalcoholic fatty liver disease. *Nat. Rev. Gastroenterol. Hepatol.* **21**, 46–56 (2024).
5. Medema, J. P. Cancer stem cells: the challenges ahead. *Nat. Cell Biol.* **15**, 338–344 (2013).
6. Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M. & Wallace, M. B. colorectal cancer. *Lancet*. **394**, 1467–1480 (2019).
7. Gupta, S. et al. Recommendations for follow-up after colonoscopy and polypectomy: A consensus update by the US multi-society task force on colorectal cancer. *Am. J. Gastroenterol.* **115**, 415–434 (2020).
8. Click, B., Pinsky, P. F., Hickey, T., Doroudi, M. & Schoen, R. E. Association of colonoscopy adenoma findings with long-term colorectal cancer incidence. *JAMA* **319**, 2021–2031 (2018).
9. O'Connell, J. B., Maggard, M. A. & Ko, C. Y. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. *J. Natl. Cancer Inst.* **96**, 1420–1425 (2004).
10. Levin, B. et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J. Clin.* **58**, 130–160 (2008).
11. Church, J. Colonoscopy: What are we missing? *Surg. Oncol Clin N Am.* **23**, 1–9 (2014).
12. Lichtenstein, P. et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.* **343**, 78–85 (2000).
13. Zauber, A. G. et al. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *N. Engl. J. Med.* **366**, 687–696 (2012).
14. Fecal occult blood test for colorectal cancer screening. an evidence-based analysis. *Ont. Health Technol. Assess. Ser.* **9**, 1–40 (2009).
15. Forones, N. M. & Tanaka, M. CEA and CA 19–9 as prognostic indexes in colorectal cancer. *Hepatogastroenterology* **46**, 905–908 (1999).
16. Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* **12**, 661–672 (2014).
17. Wishart, D. S. Metabolomics for investigating physiological and pathophysiological processes. *Physiol. Rev.* **99**, 1819–1875 (2019).
18. Bathe, O. F. & Farshidfar, F. From genotype to functional phenotype: Unraveling the metabolomic features of colorectal cancer. *Genes (Basel)*. **5**, 536–560 (2014).
19. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459 (2016).
20. Farshidfar, F. et al. A validated metabolomic signature for colorectal cancer: exploration of the clinical value of metabolomics. *Br. J. Cancer*. **115**, 848–857 (2016).
21. Geijsen, A. et al. Plasma metabolites associated with colorectal cancer: A discovery-replication strategy. *Int. J. Cancer* **145**, 1221–1231 (2019).
22. Qiu, Y. et al. A distinct metabolic signature of human colorectal cancer with prognostic potential. *Clin. Cancer Res.* **20**, 2136–2146 (2014).
23. Chen, F. 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* **71**, 1315–1325 (2022).
24. Gao, R. et al. Integrated analysis of colorectal cancer reveals cross-cohort gut microbial signatures and associated serum metabolites. *Gastroenterology* **163**, 1024–1037.e1029 (2022).
25. Yang, Y. et al. Integrated microbiome and metabolome analysis reveals a novel interplay between commensal bacteria and metabolites in colorectal cancer. *Theranostics* **9**, 4101–4114 (2019).
26. Coker, O. O. et al. Altered gut metabolites and microbiota interactions are implicated in colorectal carcinogenesis and can be non-invasive diagnostic biomarkers. *Microbiome* **10**, 35 (2022).
27. Kong, C. et al. Integrated metagenomic and metabolomic analysis reveals distinct gut-microbiome-derived phenotypes in early-onset colorectal cancer. *Gut* **72**, 1129–1142 (2023).
28. Gumpenberger, T. et al. Untargeted metabolomics reveals major differences in the plasma metabolome between colorectal cancer and colorectal adenomas. *Metabolites* **11** (2021).
29. Lin, L. et al. Biomarkers of coordinate metabolic reprogramming and the construction of a co-expression network in colorectal cancer. *Ann. Transl. Med.* **10**, 1115 (2022).
30. Shu, X. et al. Prospective study of blood metabolites associated with colorectal cancer risk. *Int. J. Cancer*. **143**, 527–534 (2018).
31. Dolce, V., Cappello, A. R., Lappano, R. & Maggiolini, M. Glycerophospholipid synthesis as a novel drug target against cancer. *Curr. Mol. Pharmacol.* **4**, 167–175 (2011).
32. Tallima, H., Azzazy, H. M. E. & El Ridi, R. Cell surface sphingomyelin: Key role in cancer initiation, progression, and immune evasion. *Lipids Health Dis.* **20**, 150 (2021).
33. D'Angelo, G., Moorthi, S. & Luberto, C. Role and function of sphingomyelin biosynthesis in the development of cancer. *Adv. Cancer Res.* **140**, 61–96 (2018).
34. Ogretmen, B. Sphingolipid metabolism in cancer signalling and therapy. *Nat. Rev. Cancer*. **18**, 33–50 (2018).
35. Cui, W. Q., Wang, S. T., Pan, D., Chang, B. & Sang, L. X. Caffeine and its main targets of colorectal cancer. *World J. Gastrointest. Oncol.* **12**, 149–172 (2020).

Acknowledgements

Supported by Zhejiang Provincial Natural Science Foundation of China (No. LTGD23C040008, LBY23H200006, LQ22H030004), Ningbo Top Medical and Health Research Program (No. 2023020612), the project of Ningbo leading Medical & healthy Discipline (2022-S04) and Ningbo Public Welfare Project(2023-S013).

Author contributions

Y.Z. conceived and designed the project. Z.D., T.L., and P.Z. recruited patients and collected tissue samples. Z.D., T.L., K.L., X.W., and K.H. analysed the data. Z.D. and Y.Z. prepared the manuscript. Z.D. and Y.Z. wrote and

reviewed the final version of the text. All authors contributed to finalizing the manuscript, read, and approved the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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

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/>.

© The Author(s) 2025