Metabonomics as a tool to detect and monitor colorectal cancer

Yunping Qiu[*], Mingming Su[*], Guoxiang Cai[†], Sanjun Cai[†], Ye Xu[†], Yan ni[*], Yumin Liu[*], Mo Dan[*], Aihua Zhao[*], Shengli Yang[*], Lisa X. Xu[*], Wei Jia[§¶]

* School of Pharmacy, and Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai, P. R. China, 200240,

† Shanghai Cancer Hospital, Medical Center of Fudan University, Shanghai, P. R. China, 200032,

§ Department of Nutrition, University of North Carolina at Greensboro, North Carolina Research Campus, Kannapolis NC 28081, USA

Abbreviations

CRC, colorectal cancer; GC-MS, gas chromatography-mass spectrometry; GC-TOFMS, gas chromatography-time of flight mass spectrometry; TIC, total iron current; PCA, Principal component analysis; OPLS-DA, orthogonal partial least squares-discriminate analysis; FC, fold change; 5-HIAA, 5-hydroxyindoleacetic acid; 5-HTP, 5-hydroxytryptophan; VIP, variable importance; OSC, Orthogonal Signal Correction; FOBT, fecal occult blood test; BMI, body mass index; CEA, carcinoembryonic atigen; ECF, ethyl chloroformate.

Colorectal carcinogenesis involves overexpression of many immediate–early response genes associated with growth and inflammation, significantly altering downstream protein synthesis and small–molecule metabolite production. We have performed a comprehensive metabolic analysis to test

the hypothesis that the distinct metabolite profiles of malignant tumors are reflected in body fluids and can be quantified by a metabolic profiling approach. In this study, we have analyzed the serum and urine metabolites from 64 colorectal cancer (CRC) patients and 65 healthy volunteers using gas chromatography coupled with mass spectrometry. After automated mass spectral deconvolution, 225 metabolites were consistently detected in serum samples. Of these, 26 were selected as significantly different between CRC and healthy controls, using variable importance (VIP) values and the Wilcoxon-Mann-Whitney test. From human urine samples, 187 metabolites were detected, of which 31 were selected as differentially produced metabolites. Using the 3 serum metabolites that differed most between the groups—oleamide, pyruvate, and histidine—the orthogonal projection to latent structures-discriminant analysis (OPLS-DA) model predicted the classification of the biological samples with 97.78% sensitivity and 97.83% specificity, based on a 99% confidence interval for probability of class membership. The orthogonal signal corrected (OSC)/PLS-DA model successfully distinguished between two CRC phenotypes, fecal occult blood test (FOBT)–positive and FOBT–negative, and two CRC locations, colon cancer and rectal cancer. We conclude that the metabolic profiling approach is sufficiently robust and sensitive to develop further for early detection of CRC, and for CRC patient stratification into subgroups of pathological stages or clinical phenotypes.

Colorectal cancer; metabonomics; gas chromatography/mass spectrometry; oleamide; glycolysis

Colorectal cancer (CRC) is the third most common type of cancer and the fourth most frequent cause of cancer mortality in the world (1). The American Cancer Society estimates that a total of 1,437,180 new cancer cases and 565,650

deaths from cancer, including 148, 810 new CRC cases and 49,960 deaths from CRC, will occur in the United States in 2008 (2). Recent decades have also witnessed a rapid increase in CRC morbidity in fast developing countries like China, especially in major cities where significant lifestyle alterations have occurred (3). Early and accurate diagnosis of CRC is of central importance for five-year survival and for less complicated surgery (2). Although CRC is a highly treatable and often curable disease when localized to the bowel, the prognosis for late stage CRC (eg, recurrent metastatic disease) remains poor and is most often the ultimate cause of death(2).

A fundamental reason for the relative lack of progress worldwide in treating CRC is that the biology of this malignant disease is not sufficiently understood, not only in its early stages of development but also in the differences between CRC phenotypes and their therapeutic outcomes. Experimental studies have focused largely on understanding the transcriptional regulation of cancer-associated gene expression (4, 5), whereas less research has been devoted to determining how this perturbed post-transcriptional regulation leads to abnormal expression of downstream proteins and metabolites in this complex disease. Therefore, more effort is needed to improve our understanding of CRC biology to identify new molecular targets and to improve current cancer treatment and prevention strategies.

Overexpression of many immediate–early response genes associated with growth and inflammation (eg, proto-oncogenes, inflammatory mediators, and angiogenic growth factors) is commonly observed in CRC cells. These genetic modifications associated with colorectal carcinogenesis allow the transformed cell to escape apoptosis while promoting proliferation, angiogenesis, and metastasis (6). These changes lead to significant alterations in downstream biochemical substances such as proteins and small–molecule metabolites. Small–molecule metabolites are the products of systemic biochemical

regulations, and their expression levels can be regarded as the response of biological systems to genetic and environmental changes (7).

Currently, a lack of detailed information about the disease–associated metabonome limits the ability of cancer biologists to understand the roles of metabolic pathways associated with CRC and its treatment. To address this gap in knowledge, emerging metabonomics / metabolomics technology uses multivariate statistical techniques to analyze the highly complex data sets generated by high-throughput spectroscopy such as nuclear magnetic resonance (NMR) and mass spectrometry (MS) (8, 9). Identifying metabolites that account for the difference between the metabolic profiles of people with CRC and their healthy counterparts can reveal important underlying molecular mechanisms of the disease. To date, global metabolic profiling of clinical samples (eg, urine, sera, and tumor tissues) has been used to visualize the distinctive metabolic profiles of patients with coronary heart disease (10), inflammatory bowel disease (11), type 2 diabetes (12), ovarian cancer and ovarian borderline tumors –(13).

In this study, we conducted a comprehensive analysis of the urinary and serum metabolites from 129 participants (65 healthy individuals and 64 CRC patients diagnosed as stage I, II, III, or IV, TNM Classification) using gas chromatography-mass spectrometry (GC-MS) in conjunction with multivariate statistical techniques. The purpose of the comprehensive metabolic analysis was to determine whether variations in the CRC metabonome are reflected in body fluids, such as urine and serum, therefore, making alternative, noninvasive means for cancer detection possible. The study was also intended to gain knowledge of important metabolic variations associated with CRC morbidity, which can be utilized for improved CRC detection, diagnosis, and therapeutic strategies.

Results

Formatted: Keep with next

Serum metabolite profiles of CRC patients

Typical gas chromatography time-of-flight mass spectrometry (GC-TOF MS) total iron current (TIC) chromatograms of serum samples from a cancer patient and a healthy control are shown in Figure 1 A. Analysis with ChromaTOF software detected a total of 225 metabolites in more than 80% samples; most metabolites were organic acids, amines, amino acids, saccharine, and nucleic acids. Both principal component analysis (PCA) and orthogonal partial least squares projection to latent structures-discriminant analysis (OPLS-DA) were performed to analyze the differences between samples from CRC patients and those from healthy individuals. The OPLS-DA model demonstrated satisfactory modeling and predictive abilities using one predictive component and two orthogonal components (R2Ycum=0.793, Q2Ycum=0.618), achieving a distinct separation between the metabolite profiles of the two groups (Fig. 2A). Notably, the OPLS-DA model correctly discriminated all nine CRC patients diagnosed at stage I from the healthy controls. However, the model failed to distinguish CRC patients by different pathological stages (I to IV).

Twenty-eight significantly different metabolites were selected using VIP values, and of these metabolites, 26 were also determined to be significantly different by the Wilcoxon-Mann-Whitney test, with the critical *p*-value set at 0.05. Twenty-one of the 28 metabolites were identified using GC-MS spectral databases and thirteen were confirmed using reference standards (Table 1). Among the identified metabolites, oleamide was the serum metabolite found to be most depleted in the CRC patients, compared to controls, showing the greatest fold change (FC=-3). Pyruvate was the metabolite most increased (FC=2.1) in CRC patients. The most significantly altered serum metabolites included decreased serine, lysine, tryptophan, histidine, and valine, and elevated lactate, 2-hydroxybutrate, and 3-hydroxybutrate in the CRC patients.

Using the 5 most significantly different metabolites—oleamide, pyruvate, histidine, valine and serine—we classified patients with CRC and healthy individuals using SPSS software, and obtained high diagnostic accuracy with 100% sensitivity and 98.5% specificity (as shown in SI 1A). Furthermore, using the 3 most significantly different metabolites—oleamide, pyruvate and histidine—distinct separation between CRC patients and healthy controls was achieved (Fig. 3A).

Urinary metabolite profiles of CRC patients

Typical GC-MS TIC chromatograms of urine samples derived from a cancer patient and a healthy control are shown in Figure 1B. Using an automated peak deconvolution and library search procedure for GC-MS data (14, 15), 187 metabolites were consistently detected in at least 80% of the urine samples. As in the serum samples, most of the metabolites detected were organic acids, amines, and amino acids. PCA and OPLS-DA were used to discriminate urine samples from CRC patients from the healthy controls. The cross-validated OPLS-DA model had satisfactory predictive ability using one predictive component and three orthogonal components ($R^2Ycum=0.767$, $Q^2Ycum=0.594$) (Fig. 2B). All cancer patients were differentiated from the healthy controls in the first predictive component (PC1). Notably, 9 CRC patients at stage I were correctly classified as diseased individuals, suggesting that this method may hold a clinical potential for early diagnosis of CRC. However, like the serum metabolic profile, this method also failed to differentiate CRC patients among several pathological stages (stage I to IV).

Based on the VIP threshold from the OPLS-DA model, a total of 36 significant variables, expressed as retention time-mass to charge ratio (RT-M/Z) pairs, were obtained, but five of them were not verified as significant by univariate statistics. Through searching our compound library, 16 metabolites were identified, of which 10 metabolites were verified using

reference compounds available at our laboratory (Table 2). The urinary metabolites with the most significant differences between groups included 5-hydroxyindoleacetate (5-HIAA), 5-hydroxytryptophan (5-HTP), methyl-O-hydroxyhippurate, phenylacetylglutamine, at elevated levels, and isocitrate, succinate and citrate, at decreased levels.

SPSS software analysis was used to classify participants by discriminant analysis using 5 of the 21 identified differentially produced metabolites in urine of CRC patients (ie, 5-HTP, succinate, glutamate, methyl-o-hydroxyhippurate, and phenylactate), achieving good discrimination with 76.6% sensitivity and 84.6% specificity (see SI 1B). The 3D scatter plot was generated with three metabolites, 5-HTP, succinate, and glutamate as the three axes, leading to good separation (Fig. 3B).

Prediction model and quantification of key differentially produced metabolites

SPSS software was used to identify potential "biomarkers" or characteristic metabolic patterns for clinical use. Approximately 70% of the samples (the 'training set') were randomly selected to build an OPLS-DA model using the 3 most significantly different metabolites, oleamide, pyruvate and histidine, used to predict the presence of CRC in the remaining 30% of samples (the 'test set'). The OPLS-DA model predicted the presence of CRC with 97.78% sensitivity and 97.83% specificity based on a 99% confidence interval for probability of class membership (Fig. 4A). The T-predicted scatter plot demonstrated a strong predictive ability distributing samples to class 1 (CRC patients) and class 2 (controls). The metabolites were quantified using internal reference standards, 2-chlorophenylalanine for pyruvate and histidine, and heptadecanoic acid for oleamide. Figure 4B-4D shows the difference in metabolite concentrations between the two groups. More details about the analysis are provided in SI 2.

With an orthogonal signal corrected (OSC) /PLS-DA discrimination technique, the serum and urinary metabolic profiles were also able to distinguish between two disease phenotypes (fecal occult blood test (FOBT)-positive and FOBT-negative) and two cancer locations (colon cancer and rectal cancer) among all CRC patients (Fig. 5).

Discussion

In this study, we detected significantly altered metabolic profiles of patients with CRC compared to their healthy counterparts, highlighting the diagnostic potential of this non-invasive analytical approach. Using the acquired serum and urinary metabolic profiles we were able to distinguish between CRC patients and the healthy subjects, as well as between two disease phenotypes (FOBT-positive and FOBT-negative) and between colon cancer and rectal cancer. However, our attempt to stratify CRC patients based on clinical classifications (stage Ito IV) was not successful. We believe that the inability to classify patients by their serum or urinary metabolite profile in this population was due to insufficient numbers of CRC patients, particularly at Stage I and Stage IV.

The comprehensive metabolic analysis was able to identify important metabolites and metabolic pathways significantly altered by CRC development. Oleamide (or *cis*-9, 10-octadecenoamide) was the most down-regulated metabolite observed in all of the CRC patients. It is a fatty acid primary amide believed to mediate conjugated linoleic acid inhibition of Caco-2 colon cancer cell growth (16). Oleamide has also been reported to enhance the activity of certain types of serotonin receptors (eg, 5-HT_{1A}, 5-HT_{2A}, and 5-HT_{2C}) (17, 18). In this study, disordered serotonin metabolism was reflected by increased expression of serotonin-related metabolites in the urine samples—the urinary levels of the intermediate precursor (5-HTP) and a metabolite of serotonin (5-HIAA) were both significantly elevated in CRC

patients compared to the healthy controls.

Pyruvate was the most significantly increased serum metabolite in CRC patients compared to healthy controls. As the first product of glycolysis, the abnormal accumulation of pyruvate in the serum is an indication of accelerated glycolysis in CRC patients. Recent proteomic analysis of colonic tissues from CRC patients also identified increased glycolysis as an important metabolic variation associated with CRC morbidity (*5*). The increased pyruvate converts to lactate in the tumor tissues, in the presence or absence of oxygen (19), as evidenced by the significantly elevated level of lactate observed CRC samples in this study. Therefore, we surmise that the abnormal accumulation of pyruvate and lactate in the CRC patients may be a result of a higher energy demand in the solid colorectal tumor tissues.

Significant changes in the levels of many phenyl-containing compounds were detected in the urine samples of diseased individuals compared to their healthy counterparts. These phenylic metabolites are produced mainly by gut microbiota through fermentation of dietary polyphenols and aromatic amino acids (eg, phenylalanine and tyrosine derived from dietary proteins) (20). In the present study, the significantly altered levels of p-cresol, phenylacetate, and p-hydroxyphenylacetate were detected in the urine of CRC patients, along with altered levels of phenylacetylglutamine, methyl-o-hydroxyhippurate, and 2-hydroxyhippurate. The significantly altered metabolic signatures reflected by those phenylic compounds strongly indicate a disrupted gut microbial metabolism associated with CRC.

We detected differently expressed metabolites in CRC patients with multivariate and univariate statistical significance, however, we were not able to use these acquired serum and urinary metabolic profiles for correct classification of CRC by TNM Classification (stages I to IV). We anticipated that our ability to classify patients by their serum and urinary metabolite profiles would be lower in this patient population due to insufficient numbers

of CRC patients, particularly at Stage I and Stage IV. However, using an OSC/PLS-DA discrimination technique, the serum and urinary metabolic profiles of FOBT-positive and FOBT-negative patients could be differentiated, as could metabolic profiles of colon cancer and rectal cancer among CRC patients (Fig. 5).

These distinctions have collectively constituted a metabolic window into the CRC morbidity, providing metabolic endpoints that complement the interpretation of genomic, proteomic, and epidemiological data. The results of this study indicate that this metabolic profiling approach is sufficiently robust and sensitive to develop further for early detection of CRC and for patient stratification into subgroups of different pathological stages or clinical phenotypes, which will increase survival among patients with such a deadly disease.

Materials and methods

Clinical samples. The protocol was approved by the Shanghai Cancer Hospital Institutional Review Board and all participants gave informed consent before they were involved in the study. The patients, age 42 to 74 years old and diagnosed with CRC (32 colon cancers and 32 rectal cancers), were categorized according to histopathological features and stages according to TNM classification of malignant tumors: stage I, 9 patients; stage II, 27 patients; stage III, 20 patients; stage IV, 8 patients. The clinical diagnosis and pathological reports of all the patients were obtained from hospitals. The healthy volunteers, aged 42 to 71 years old, were selected by a routine physical examination and patients with any gastrointestinal tract disorders were excluded. Body mass index (BMI) and carcinoembryonic antigen (CEA) level for each patient were also assessed. Clinical information of participants is provided in the Table 2. Urine and venous blood were collected in the morning before breakfast from a total of 64 CRC patients and 65 healthy

volunteers at Shanghai Cancer Hospital, Medical Center of Fudan University (Shanghai, China). Urine samples were immediately frozen on dry ice, and sera were separated by centrifugation at 2500rpm for 15 min. Until use in the assay, all samples were stored at -80°C

GC-TOFMS spectral acquisition of serum samples and data pretreatment. Metabolites in the serum were derivatized with trimethylsilyl and subsequently analyzed by gas chromatography-time of flight mass spectrometry (GC-TOFMS) with minor modifications to our previously published method (13). A 100-μl aliquot of serum sample was spiked with two internal standard solutions (10μl L-2-chlorophenylalanine in water, 0.3 mg/ml; 10μl heptadecanoic acid in methanol, 1 mg/ml) and vortexed for 10 seconds. The mixed solution was extracted with 300 μl of methanol:chloroform (3:1) and vortexed for 30 seconds. After storing for 10 minutes at -20°C, the samples were centrifuged at 10,000 rpm for 10 minutes. An aliquot of the 300-μl supernatant was transferred to a glass sampling vial to vacuum dry at room temperature. The residue was derivatized using a two-step procedure. First, 80 μl methoxyamine (15 mg/ml in pyridine,) to the vial at 30°C for 90 minutes followed by 80 μl BSTFA (1%TMCS) at 70°C for 60 minutes.

Each 1-μl aliquot of the derivatized solution was injected in spitless mode into an Agilent 6890N gas chromatograph coupled with a Pegasus HT time-of-flight mass spectrometer (Leco Corporation, St Joseph, USA). Separation was achieved on a DB-5 ms capillary column (30 m × 250 μm I.D., 0.25-μm film thickness; (5%-phenyl)-methylpolysiloxane bonded and crosslinked; Agilent J&W Scientific, Folsom, CA, USA) with helium as the carrier gas at a constant flow rate of 1.0 ml min. The temperature of injection, interface, and ion source was set to 270°C, 260°C, and 200°C, respectively. The GC temperature programming was set to 2 min isothermal heating at 80°C,

followed by 10°C/min oven temperature ramps to 180 °C, 5 °C/min to 240°C, and 25°C/min to 290 °C, and a final 9 min maintenance at 290°C. Electron impact ionization (70 eV) at full scan mode (m/z 30-600) was used, with an acquisition rate of 20 spectrum/second in the TOFMS setting. Chromatogram acquisition, data handing, automated peak deconvolution, unique mass area calculation, and library search were done by Leco ChromaTOF software (v3.30).

The data were mean-centered and pareto-scaled before multivariate statistical analysis in the SIMCA-P 11.0 Software package (Umetrics, Umeå, Sweden). Mean centering was performed columnwise to remove the offsets. Pareto scaling reduces the range of variance across each spectrum by dividing each variable by the square root of its standard deviation. Pareto-scaling is used in metabolic profiling studies as a compromise between no scaling, which may fail to detect small changes in concentration of metabolites, and unit variance scaling, which gives equal weight to baseline imperfections, noise, and defined signals in the mass spectrum (21).

GC-MS spectral acquisition of urine samples and data pretreatment. Urine samples from human were used for GC-MS analysis and spectral acquisition according to our previously developed method with minor modifications (15). Briefly, a 600-µl aliquot of urine sample was prepared for ethyl chloroformate (ECF) derivatization. After adding 400 µl anhydrous ethanol, 100 µl pyridine, and 50 µl ECF to the urine sample, the resultant mixture was sonicated at 40 kHz for 60 seconds. The extraction was carried out using 300 µl chloroform and adjusting the aqueous layer pH to 9–10 using 100 µl NaOH (7 M). The derivatization was repeated by adding 50µl ECF into the aforementioned products. After the samples were vortexed for 30 seconds and centrifuged at 3000 rpm for 10 minutes, the aqueous layer was aspirated off, and the remaining chloroform layer containing derivatives was dried with

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

anhydrous sodium sulfate for subsequent GC-MS analysis.

GC-MS raw data files were initially converted into NetCDF format using DataBridge (Perkin-Elmer Inc., USA), then directly processed by our custom scripts in MATLAB 7.0 (The MathWorks, Inc. USA), to carry out baseline correction, peak deconvolution and alignment, exclusion of the internal standard peak, and normalization to the total sum of the chromatogram. (22). The resultant three-dimensional matrix encompassing peak indices (retention time-m/z pairs), sample names (observations), and normalized peak areas (variables) were imported into the SIMCA-P 11.0 Software package (Umetrics, Umeå, Sweden). The data were also mean-centered and pareto-scaled prior to multivariate statistical analysis.

Statistical analysis. The spectral processing and multivariate statistics for metabolic profiling were performed with Matlab and SIMCA-P software as previously reported (23). The detailed process is described in Supporting Information. PCA, a widely-ascribed unsupervised method, was used in this study to reduce complex spectral data sets to a two- or three-dimensional scores map that indicates inherent relationship such as clusters and groupings among observations. The first principal component (PC1) explains the most variance in the data. The second principal component, (PC2), is orthogonal to PC1, and represents maximum amount of variance not explained by PC1. The remaining components are constructed in a similar manner (24). In this study, PCA was used to analyze the MS data derived from samples prior to the use of supervised methods.

OPLS-DA (21, 25), a more sophisticated supervised classifier, decomposes the **X** matrix into blocks of structured variation correlated to and orthogonal (unrelated) to **Y**. Therefore, the discriminating information is concentrated in the first predictive component by removing systematic variation in the quantified serum or urine profiles in **X** related to the response **Y**. $\mathbf{R^2X}$ and $\mathbf{R^2Y}$

Formatted: German (Germany)

represent the fraction of the variance while **Q^2Y** suggests the predictive accuracy of the model. The cumulative values of **R^2X**, **R^2Y,** and **Q^2Y** close to 1 indicate a satisfactory model with a reliable predictive ability.

The technique of orthogonal signal correction OSC(26) was used to remove between-sample variation in the data matrix that is not correlated with Y-vector. The resulting dataset was filtered to eliminate unwanted information and describe maximum separation based on class.

Metabolites identified in serum and urine samples were classified by discriminant analysis to identify the primary metabolites accounting for the difference between CRC and control samples, using SPSS for Windows software (14.0, Chicago, IL). The three most important metabolites were plotted in a 3D scatter plot using Matlab software. An OPLS-DA predication model was generated in SIMCA-p, using the most outstanding metabolites selected from SPSS software. The most important metabolites were quantitated using reference standards. The peak area ratio of key metabolites to internal standard was used to calculate metabolite concentrations.

Identification of metabolites differentially produced in CRC. Based on a VIP threshold from a typical 7-fold cross-validated OPLS-DA model, a number of metabolites responsible for the difference in the metabolic profiles of diseased individuals and healthy controls could be obtained. In this study, the default 7-round cross-validation in SIMCA-P software package was applied with $1/7^{th}$ of the samples being excluded from the mathematical model in each round, in order to guard against over-fitting. In parallel, the metabolites identified by the OPLS-DA model were validated at a univariate level using nonparametric Wilcoxon-Mann-Whitney test from the Matlab statistical toolbox with the critical *p*-value set to 0.05. The corresponding fold change shows how these selected differential metabolites varied in the CRC patients from those of healthy controls. Additionally, compound were

identified by comparing the mass fragments with those present in the commercially available mass spectral databases (eg, NIST, Wiley, NBS, in ChromaTOF or Turbomass (v 4.1.1) software) and finally verified by reference compounds available.

Acknowledgement

This work was funded by the National Basic Research Program of China (Project number 2007CB914700).

References

1. Labianca, R., Beretta, G., Gatta, G., de Braud, F., & Wils, J. (2004) *Critical reviews in oncology/hematology* **51,** 145-170.
2. Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., & Thun, M. J. (2007) *Ca-A Cancer Journal for Clinicians* **57,** 43-66.
3. Lianzhi, D., Youlin, Q., & Liandi, L. (2002) *Chin Cancer* **11,** 250-260.
4. Cardoso, J., Boer, J., Morreau, H., & Fodde, R. (2007) *Biochimica et biophysica acta* **1775,** 103-137.
5. Bi, X., Lin, Q., Foo, T. W., Joshi, S., You, T., Shen, H. M., Ong, C. N., Cheah, P. Y., Eu, K. W., & Hew, C. L. (2006) *Mol Cell Proteomics* **5,** 1119-1130.
6. J. Mendelsohn, P. Howley, M. Israel, & Liotta, L. (2001) in *The Molecular Basis of Cancer*, eds. MENDELSOHN J. , HOWLEY P., & M., I. (W. B. Saunders, Philadelphia), pp. 289-312.
7. Denkert, C., Budczies, J., Kind, T., Weichert, W., Tablack, P., Sehouli, J., Niesporek, S., Konsgen, D., Dietel, M., & Fiehn, O. (2006) *Cancer research* **66,** 10795-10804.
8. Nicholson, J. K., Lindon, J. C., & Holmes, E. (1999) *Xenobiotica* **29,** 1181-1189.
9. Williams, R., Lenz, E. M., Wilson, A. J., Granger, J., Wilson, I. D., Major, H., Stumpf, C., & Plumb, R. (2006) *Molecular bioSystems* **2,** 174-183.
10. Brindle, J. T., Antti, H., Holmes, E., Tranter, G., Nicholson, J. K., Bethell, H. W., Clarke, S., Schofield, P. M., McKilligin, E., Mosedale, D. E., *et al.* (2002) *Nature medicine* **8,** 1439-1444.
11. Marchesi, J. R., Holmes, E., Khan, F., Kochhar, S., Scanlan, P., Shanahan, F., Wilson, I. D., & Wang, Y. (2007) *Journal of proteome research* **6,** 546-551.
12. van Doorn, M., Vogels, J., Tas, A., van Hoogdalem, E. J., Burggraaf, J., Cohen, A., & van der Greef, J. (2007) *British journal of clinical pharmacology* **63,** 562-574.
13. Odunsi, K., Wollman, R. M., Ambrosone, C. B., Hutson, A., McCann, S. E., Tammela, J., Geisler, J. P., Miller, G., Sellers, T., Cliby, W., *et al.* (2005) *International journal of cancer* **113,** 782-788.
14. Ni, Y., Su, M., Qiu, Y., Chen, M., Liu, Y., Zhao, A., & Jia, W. (2007) *FEBS Lett* **581,** 707-711.
15. Qiu, Y., Su, M., Liu, Y., Chen, M., Gu, J., Zhang, J., & Jia, W. (2007) *Anal Chim Acta* **583,** 277-283.
16. Kim, E. J., Jun, J. G., Park, H. S., Kim, S. M., Ha, Y. L., & Park, J. H. (2002) *Anticancer research* **22,** 2193-2197.
17. Boger, D. L., Henriksen, S. J., & Cravatt, B. F. (1998) *Current pharmaceutical design* **4,** 303-314.
18. Thomas, E. A., Carson, M. J., Neal, M. J., & Sutcliffe, J. G. (1997) *Proceedings of the National Academy of Sciences of the United States of America* **94,** 14115-14119.
19. Gatenby, R. A. & Gillies, R. J. (2004) *Nature reviews* **4,** 891-899.
20. Rechner, A. R., Smith, M. A., Kuhnle, G., Gibson, G. R., Debnam, E. S., Srai, S. K., Moore, K. P., & Rice-Evans, C. A. (2004) *Free Radic Biol Med* **36,** 212-225.
21. Cloarec, O., Dumas, M. E., Trygg, J., Craig, A., Barton, R. H., Lindon, J. C., Nicholson, J. K., & Holmes, E. (2005) *Anal Chem* **77,** 517-526.
22. Chen, M., Su, M., Zhao, L., Jiang, J., Liu, P., Cheng, J., Lai, Y., Liu, Y., & Jia, W. (2006) *Journal of proteome research* **5,** 995-1002.
23. Li, H., Ni, Y., Su, M., Qiu, Y., Zhou, M., Qiu, M., Zhao, A., Zhao, L., & Jia, W. (2007) *Journal of proteome research* **6,** 1364-1370.
24. Jackson, J. E. (1991) *A User's Guide to Principal Components*.
25. Trygg, J. & Wold, S. (2002) *Journal of Chemometrics* **16,** 119-128.
26. Gavaghan, C. L., Wilson, I. D., & Nicholson, J. K. (2002) *FEBS letters* **530,** 191-196.

Formatted: French (France)

Figure legends

Fig. 1. Typical total ion current (TIC) chromatograms of biological samples obtained from a CRC patient and a healthy control. (A) Human serum TICs. (B) Human urine TICs. The TIC chromatogram for the control groups are shown below with the CRC samples. The keys can be found in Table 1.

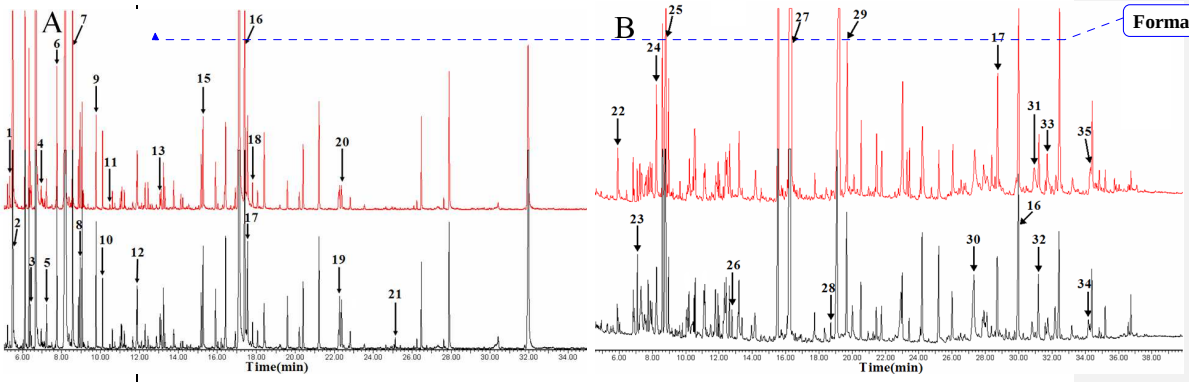
Fig. 2. Metabolic profiles of samples from CRC patients and their healthy counterparts, derived from OPLS-DA models. (A) Human serum samples. (B) Human urine samples.

Fig. 3. Scatter plots of (A) human serum samples from CRC patients and healthy subjects constructed using three most significantly altered serum metabolites—oleamide, pyruvate, and histidine—identified by the SPSS discriminant analysis; (B) human urine samples from CRC patients and healthy subjects constructed using three most significantly altered urinary metabolites—glutamate, succinate, and 5-HTP.

Fig. 4. Three most altered metabolites in CRC patients: oleamide, pyruvate, and histidine. (A) PLS-DA prediction model constructed using oleamide, pyruvate, and histidine from serum samples of 45 CRC patients (red circles) and 45 control individuals (black dots), as the 'training set'. This model was used to predict the samples that were not used in the construction of the model (the 'test set', 39 individuals (blue triangles)). Predictions are made using T-predicted scatter plot with the cut-off of 0 for class membership. The dashed line separates the CRC group from the control group (sample number). The OPLS-DA model predicted the presence of CRC with a sensitivity of 97.78% and a specificity of 97.83% in the test set, based on a 99% confidence interval for probability of class membership.
(B–D) Differences in serum concentration of oleamide, pyruvate, and histidine between CRC group and control group, showing the mean concentration of the metabolite in each group with standard deviation.

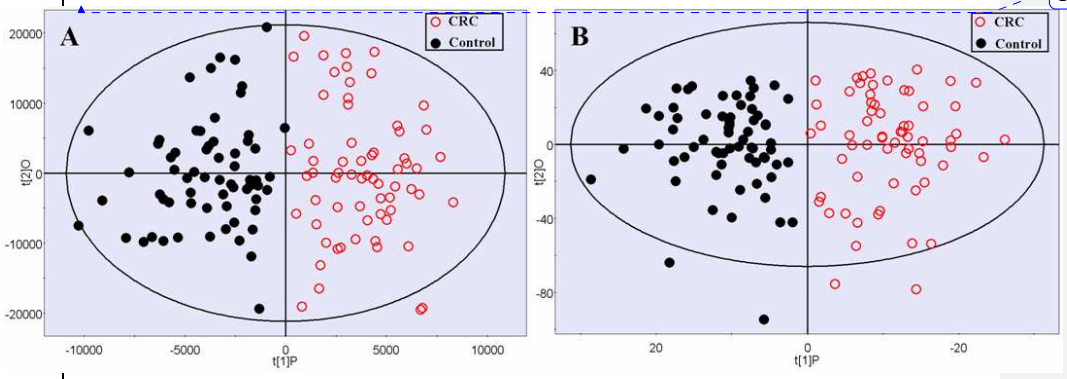
Fig. 5. OSC/PLS-DA discrimination between phenotypes and cancer location in CRC patients. (A) Serum samples to discriminate between FOBT-positive and FOBT-negative CRC. (B) Urine samples to discriminate

between FOBT-positive and FOBT-negative CRC. (C) Serum samples to discriminate between colon cancer and rectal cancer. (D) Urine samples to discriminate between colon and rectal cancers. (A: R2Y=0.468, Q2=0.264; B: R2Y=0.618, Q2=0.337; C: R2Y=0.375, Q2=0.271; and D: R2Y=0.774, Q2=0.587)



Formatted: French (France)

Fig 1



Formatted: French (France)

Fig 2

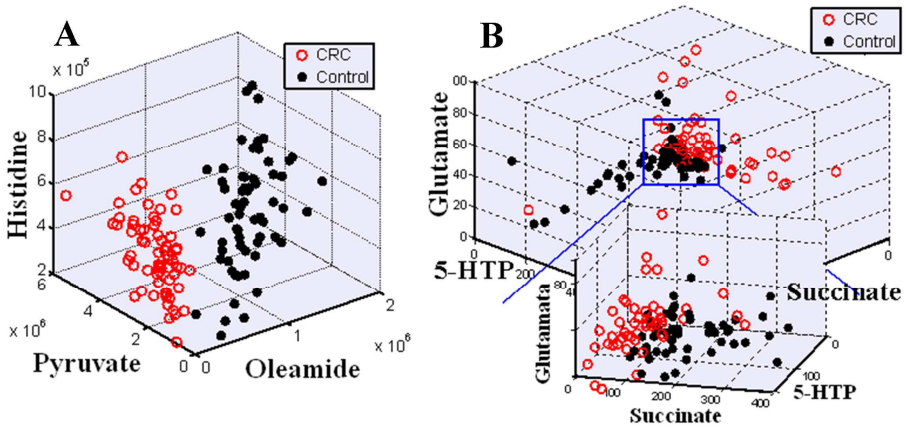

A
CRC
Control
Histidine
Pyruvate
Oleamide
B
CRC
Control
Glutamate
5-HTP
Glutamata
Succinate
5-HTP
Succinate

Fig 3

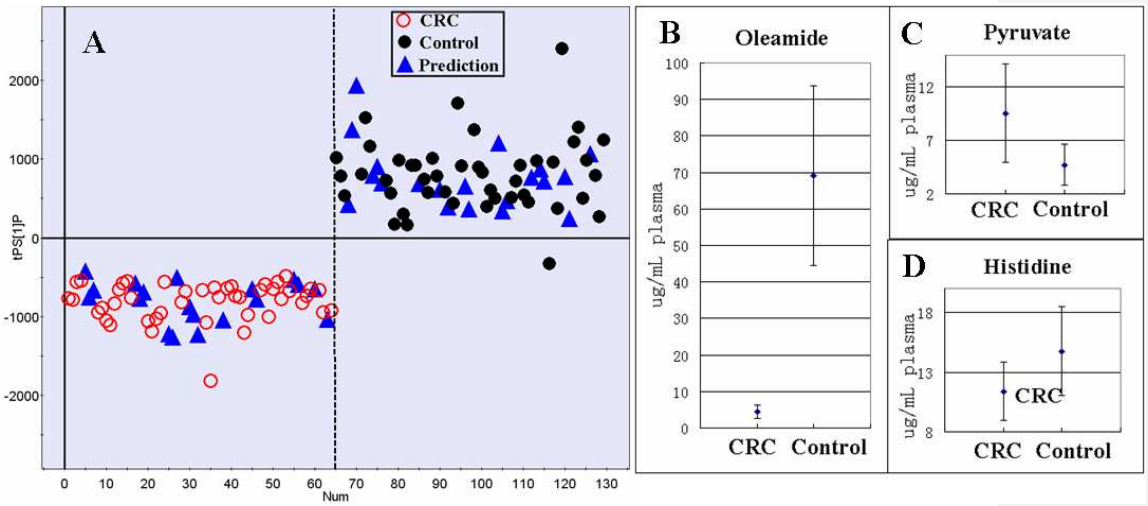

A
CRC
Control
Prediction
tPS[1]P
Num
B
Oleamide
ug/mL plasma
CRC Control
C
Pyruvate
ug/mL plasma
CRC Control
D
Histidine
ug/mL plasma
CRC
CRC Control

Fig. 4

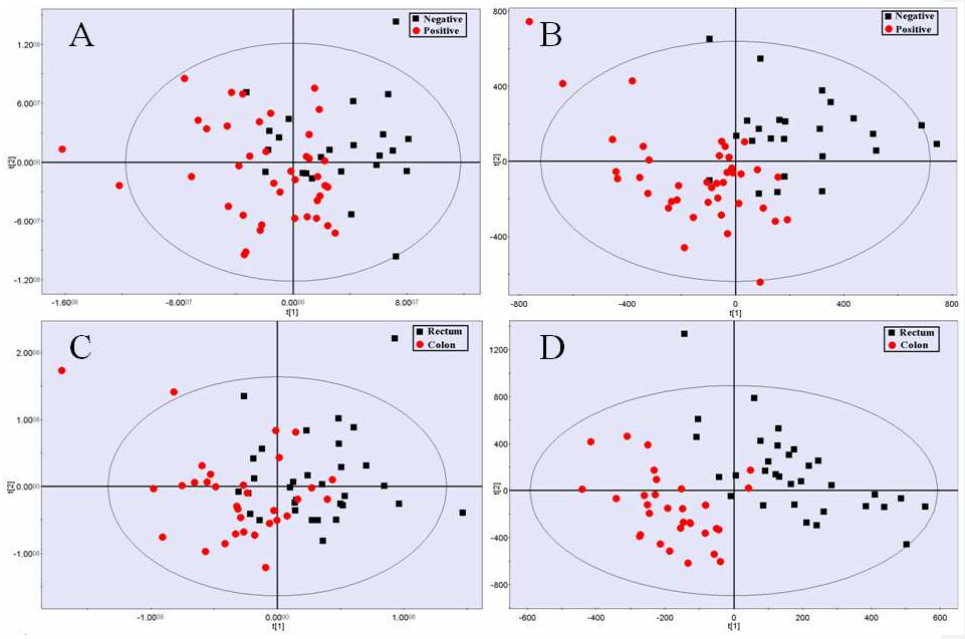

A
Negative
Positive
t[1]
t[2]
B
Negative
Positive
t[1]
t[2]
C
Rectum
Colon
t[1]
t[2]
D
Rectum
Colon
t[1]
t[2]

Fig. 5