



OPEN

## Urinary charged metabolite profiling of colorectal cancer using capillary electrophoresis-mass spectrometry

Ryutaro Udo<sup>1</sup>, Kenji Katsumata<sup>1</sup>, Hiroshi Kuwabara<sup>1</sup>, Masanobu Enomoto<sup>1</sup>, Tetsuo Ishizaki<sup>1</sup>, Makoto Sunamura<sup>4</sup>, Yuichi Nagakawa<sup>1</sup>, Ryoko Soya<sup>1</sup>, Masahiro Sugimoto<sup>2,3</sup>✉ & Akihiko Tsuchida<sup>1</sup>

Colorectal cancer (CRC) has increasing global prevalence and poor prognostic outcomes, and the development of low- or less invasive screening tests is urgently required. Urine is an ideal biofluid that can be collected non-invasively and contains various metabolite biomarkers. To understand the metabolomic profiles of different stages of CRC, we conducted metabolomic profiling of urinary samples. Capillary electrophoresis-time-of-flight mass spectrometry was used to quantify hydrophilic metabolites in 247 subjects with stage 0 to IV CRC or polyps, and healthy controls. The 154 identified and quantified metabolites included metabolites of glycolysis, TCA cycle, amino acids, urea cycle, and polyamine pathways. The concentrations of these metabolites gradually increased with the stage, and samples of CRC stage IV especially showed a large difference compared to other stages. Polyps and CRC also showed different concentration patterns. We also assessed the differentiation ability of these metabolites. A multiple logistic regression model using three metabolites was developed with a randomly designated training dataset and validated using the remaining data to differentiate CRC and polyps from healthy controls based on a panel of urinary metabolites. These data highlight the changes in metabolites from early to late stage of CRC and also the differences between CRC and polyps.

More than 1.8 million new cases of colorectal cancer (CRC) and 881,000 deaths are estimated to have occurred in 2018. CRC ranks third in terms of incidence and second in terms of mortality among cancers worldwide<sup>1</sup>. The incidence and mortality of CRC show very similar patterns in both men and women, despite showing large geographical variation. The detection and removal of cancer precursors can reduce CRC incidence and mortality. Therefore, the early detection of CRC is needed to enable less invasive treatments and results in lower morbidity and mortality.

Serum carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) have been used as biomarkers for CRC. However, serum CEA is elevated in only approximately 50% of patients with CRC who have lymph node metastasis, and in 75% of patients with distant metastasis. Serum CA19-9 is positive in more than 70% of patients with distant metastasis and is useful for the diagnosis of recurrence and metastasis of CRC<sup>2</sup>. The sensitivity of serum CEA and CA19-9 is extremely low for the detection of stage II CRCs, at 33% and 11%, respectively<sup>3-6</sup>. Therefore, more accurate screening methods and novel biomarkers for the detection of CRC are required.

A less invasive screening test has been widely used in Japan for the detection of CRC, which detects microscopic amounts of blood by targeting either heme or human globin in the stool<sup>7</sup>. However, the rate of early detection of CRC remains only 37%<sup>8-11</sup>. This is because many patients refuse to undergo colonoscopy because of stigma, difficulty in preparing for the test, and the physical burden of the test. Some patients may not undergo an endoscopic examination because the rate of a positive diagnosis even after a positive faecal occult blood test

<sup>1</sup>Department of Gastrointestinal and Pediatric Surgery, Tokyo Medical University, 6-7-1, Nishijinjuku, Shinjuku, Tokyo 160-0023, Japan. <sup>2</sup>Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0811, Japan. <sup>3</sup>Research and Development Center for Minimally Invasive Therapies, Medical Research Institute, Tokyo Medical University, 6-1-1, Shinjuku, Tokyo 160-0022, Japan. <sup>4</sup>Department of Gastroenterological Surgery and Transplantation Surgery, Tokyo Medical University Hachioji Medical Center, 1163 Tatemachi, Hachioji City, Tokyo 193-0998, Japan. ✉email: mshrgmt@tokyo-med.ac.jp

is low. Furthermore, while people in Japan have access to high-quality medical care through the universal health insurance system, this may reduce patients' awareness of the importance of preventive medicine. Therefore, the development of less invasive screening tools for early detection of CRC, even in non-symptomatic patients, should be a priority.

There are many reports describing dynamic reprogramming of metabolic pathways in CRC<sup>12,13</sup>. The proto-oncogene *MYC* regulates various metabolic pathways in CRC, including the glycolysis<sup>14</sup> and tryptophan pathways<sup>15</sup>. *MYC* also activates polyamine synthesis in CRC by activating the first rate-limiting enzyme, ornithine decarboxylase<sup>16,17</sup>. Spermine/spermidine *N*<sup>1</sup>-acetyltransferase, *N*<sup>1</sup>-acetyl polyamine oxidase, and spermine oxidase are concurrently activated to metabolize spermine to *N*<sup>1</sup>-acetylspermine, spermidine to *N*<sup>1</sup>-acetylspermidine, and spermine to spermidine, respectively, and these metabolites are secreted into the surrounding tissues<sup>18–20</sup>. In particular, an increase in urinary *N*<sup>1</sup>,*N*<sup>12</sup>-diacetylspermine has been reported in CRC patients<sup>21–23</sup>. However, it is difficult to specifically diagnose CRC using this single marker on its own, because an increase in this metabolite in the blood of lung cancer patients and in the urine of breast cancer patients has also been reported. We previously performed a targeted analysis of urinary polyamines using liquid chromatography-triple quadrupole mass spectrometry (MS) and confirmed the elevation of *N*<sup>1</sup>,*N*<sup>12</sup>-diacetylspermine in urine samples from patients with CRC<sup>9</sup>.

Metabolomics has been used for the detection of metabolic abnormalities caused by tumours in biofluid samples that can be collected by minimally invasive methods, such as blood<sup>24,25</sup> and urine<sup>26,27</sup>. Potential biomarkers for the detection of CRC include metabolites that are dysregulated during the progression of CRC<sup>28</sup>. The combination of multiple markers into a biomarker panel has the potential to differentiate CRC from polyps<sup>27,29</sup>. Therefore, comprehensive metabolomic analysis using a sophisticated classification method is expected to contribute to the identification of patients with CRC.

In the present study, we utilized capillary electrophoresis-mass spectrometry (CE-MS) to perform non-targeted analysis of urine collected from patients with CRC or polyps (P), as well as from healthy controls, to profile hydrophilic urinary metabolites. To understand the change in metabolomic patterns depending on the progression, we conducted pathway analyses of the profiled metabolites. The discriminating ability of these profiles was assessed using a multiple logistic regression model (MLR).

## Materials and methods

**Study subjects.** This study was conducted according to the guidelines of the Declaration of Helsinki. The study protocol was approved by the Ethics Committee of Tokyo Medical University (study approval no. 2346). Written informed consent was obtained from each subject before participation in the study. Patients with CRC included subjects who underwent chemotherapy. Patients with chronic metabolic diseases, such as diabetes, were included.

The patients' clinical data were obtained as described in a previous study<sup>9</sup>. Briefly, urinary samples were collected from patients with CRC (n = 209) and P (n = 16). Urinary samples were also collected from healthy controls (n = 22). Among them, five subjects provided multiple samples at various time (7:00–8:00, 11:00–12:00, and 17:00–18:00) on three consecutive days. Six to nine samples were collected from identical persons because of several missing samples. The first collected samples were defined as healthy control 1 (HC1, n = 22) (these samples were collected at 9:00 on the first day) and the other samples were defined as healthy control 2 (HC2, n = 50). The samples from patients with P and CRC were collected once between 9:00–16:00.

The resected specimens were pathologically classified according to the seventh edition of the Union for International Cancer Control TNM classification of malignant tumours. Serum CEA and CA19-9 levels were measured using a radioimmunoassay (Abbott, Chiba, Japan). The limit of detection of CEA was 0.5 ng/mL and that of CA19-9 was 2 U/mL. A high CEA level was defined as a level exceeding 5 ng/mL, and a high CA19-9 level was defined as a level exceeding 37 U/mL, according to the guidelines defined by the manufacturer of the test kit.

**Metabolomics of urine samples.** Urine was collected as previously described<sup>9</sup>. The collected samples were divided into two tubes and stored at –80 °C. One of the samples was used for liquid chromatography-mass spectrometry (LC-MS)-based polyamine analysis in our previous study<sup>9</sup> and the other was used in the present study for CE-MS-based metabolomic profiling. Frozen urine samples were thawed at 4 °C for approximately 1.5 h and subsequently homogenised using a vortex mixer at room temperature, and centrifuged through a 5-kDa cutoff filter (Millipore, Bedford, MA) at 9100 × g for at least 2.5 h at 4 °C. Twenty microlitre of each sample was placed in a 1.5 mL Eppendorf tube with 80 µL of Milli-Q water including 2 mM each of methionine sulfone, 2-[N-morpholino]-ethanesulfonic acid, D-camphor-10-sulfonic acid (sodium salt), 3-aminopyrrolidine, and trimesate.

The mixture was centrifugally filtered at 9100 × g for 30 min at 4 °C through a 5-kDa cutoff filter (Millipore) to remove large molecules. The filtrate was transferred to a vial and assayed by CE-time-of-flight (TOF)-MS. The parameters of CE-TOF-MS have been described elsewhere and were applied with slight modifications, as follows<sup>30</sup>. Each sample was measured in both positive and negative mode. The parameters in the negative mode were the same as those described previously<sup>30</sup>. For the measurement of cations, two conditions were modified as follows: (1) fused silica capillaries (50 µm i.d. × 95 cm total length) were used, and (2) before each injection, the capillary was equilibrated for 4 min by flushing with 1 M formate. To quantify each metabolite, standard mixtures, including 20 µL of each metabolite and 200 µL of each standard compound, were measured before analysis of the samples.

Raw CE-TOF-MS data were analysed with our proprietary software, MasterHands (ver. 2.17.2.15, Keio University, Yamagata, Japan)<sup>31,32</sup>. The raw data were adjusted by 0.02 m/z each to generate electropherograms. For each electropherogram, the background noise was removed and possible peaks with a signal-to-noise ratio

Group	Stage	n	Age		Gender	
			Mean	SD	Male	Female
HC		22	46.8	13.1	17	5
P		16	63.2	11.1	12	4
CRC	0	2	76.5	0.710	2	0
	I	52	68.9	10.6	26	26
	II	67	69.7	11.4	42	25
	III (N1)	46	67.9	14.7	24	22
	III (N2)	25	67.1	11.6	12	13
	IV	17	69.5	9.92	8	9

**Table 1.** Subject characteristics. CRC colorectal cancer, P polyps, HC healthy controls.

of greater than 2.0 were retained. Peaks that were detected in multiple samples were aligned, missing peaks were filled, and redundant features, for example, isotopic peaks, fragments, and noise peaks were eliminated. Metabolite identities were assigned to each feature by matching the corresponding  $m/z$  values and normalized migration times of standard compounds. The absolute concentration of each metabolite peak was calculated by dividing the peak area of by that of the internal standards (methionine sulfone and D-camphor-10-sulfonic acid (sodium salt) for cations and anions, respectively), and eliminating the drift in MS sensitivity. These absolute concentrations were normalised to the creatinine concentration of each sample and used for subsequent analysis.

**Data analysis.** To compare the metabolomic profiles of HC1, P, and CRC, we performed clustering analyses, principal component analysis (PCA), and partial least squares-discrimination analyses (PLS-DA). PLS-DA was also conducted by adding HC2 data, for evaluating the variation in data for healthy controls. To evaluate the discrimination ability of the combination of metabolites, two commonly used feature selection methods were employed. First, support vector machine-feature selection (SVM-FS) was initially performed to rank the identified metabolites by differential ability based on the margin of the data between HC1 and CRC<sup>33</sup>. Second, among the highly ranked metabolites, correlation-based feature selection (CFS) was performed to select a minimum set of features<sup>34</sup>. Stepwise feature selection (backward method) was also performed to eliminate unnecessary features. Features with a  $P$ -value greater than 0.05 were eliminated from the model. The MLR model was developed using highly ranked features. To assess the discriminatory ability of these potential biomarkers, we randomly split the data into a training and a validation dataset, while maintaining the ratio of HC1, P, and CRC subjects, with half of the data used as the training data and half used as the validation data. The MLR model was used to identify metabolites that could differentiate between two groups. We developed the MLR model to differentiate CRC + P from HC1, for evaluating the potential for detecting these diseases by non-invasive screening. As an internal validation, 10-fold cross-validation (CV) was performed using the training dataset. As external validation, the developed MLR model was evaluated using the validation data.

To eliminate potential bias caused by the random split, multiple times two-fold CV was performed using all the datasets. The selection of features was not performed in these experiments, and features selected in the previous experiment were used. Because of the limited sample size, resampling tests were also performed, that is, datasets were generated by random selection from the original datasets, which enabled redundant selection, and the MLR model was built and evaluated using the generated datasets.

Mann–Whitney and Kruskal–Wallis tests followed by the Dunn's post hoc test were used to analyse the differences in metabolite levels among 2 and 3 of the groups, respectively. The  $\chi^2$  test was used to assess gender bias in the dataset. For multiple testing correction,  $P$ -values were adjusted using a false discovery rate (FDR). The areas under the receiver operating characteristic curves were used to analyse the differentiation ability of the MLR models.

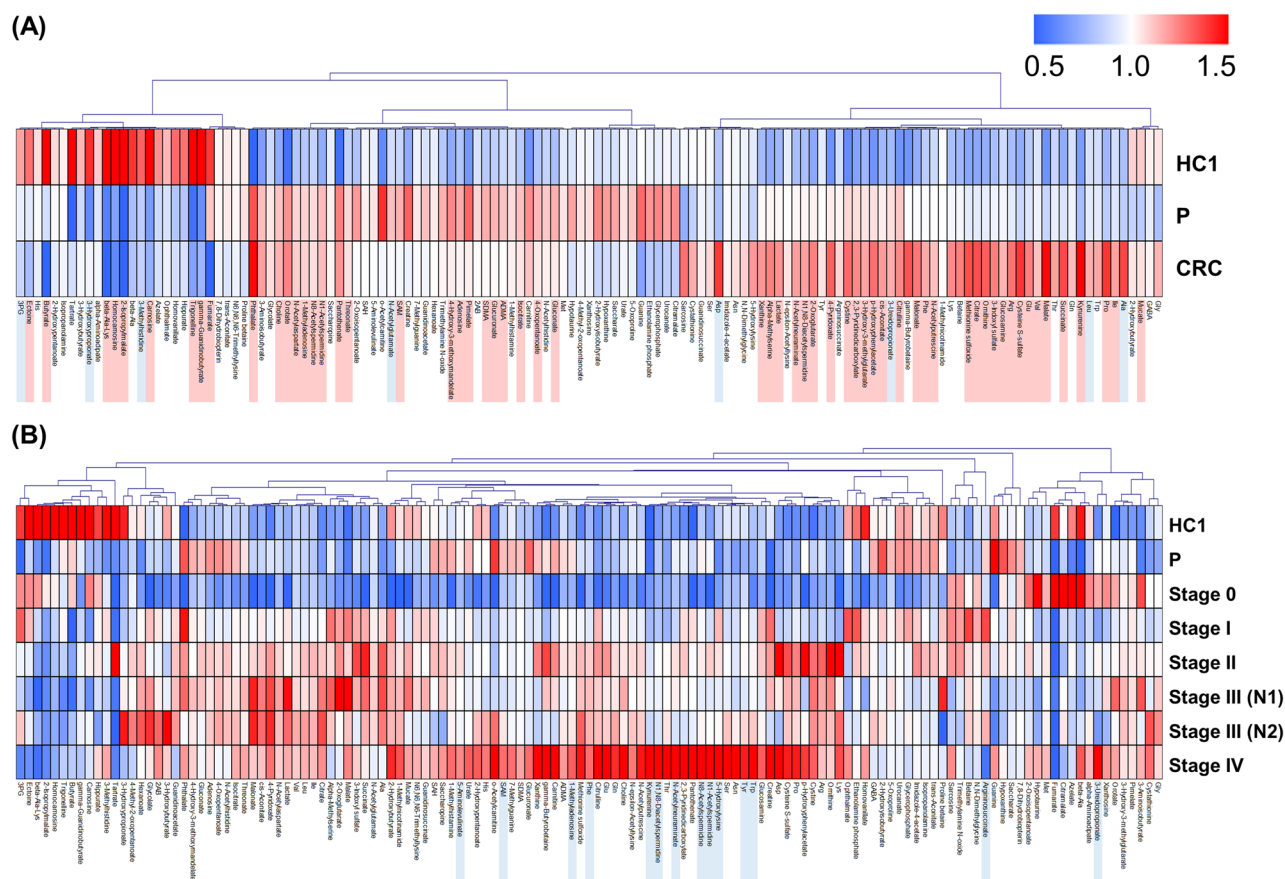
Pathway analyses, PCA, and PLS-DA were conducted using MetaAnalyst<sup>35</sup>. Log-transformation and mean centring options without sample normalization were used for PCA and PLS-DA. The SVM-FS and CFS were performed using the default parameters in Waikato Environment for Knowledge Analysis (Weka) (ver. 3.6.15, New Zealand)<sup>36</sup>. The metabolites ranked within 20% by SVM-FS were used for subsequent feature selection. Both the two-fold CV and resampling tests were performed 200 times with different random values. GraphPad Prism (ver. 8.4.2, GraphPad Software Inc., San Diego, CA, USA), JMP Pro (ver 14.1.0, SAS Institute Inc., Cary, NC, USA), R-language (ver 4.0.2, R foundation for statistical computing, Vienna, Austria), and MeV TM4 (ver. 4.7.4, <http://mev.tm4.org/>) were used for statistical analysis. The authors will provide raw data upon request.

## Results

**Urinary metabolomic profiling.** Characteristics of patients in the present cohort are summarized in Tables 1 and 2. Metabolomic analysis successfully identified and quantified 154 metabolites. The metabolites that were frequently identified among the samples were visualized in a heatmap (Fig. 1). HC1, P, and CRC groups clustered according to different metabolite profiles. The HC1 group included the lowest number of metabolites with higher concentrations than in the P and CRC groups. In the comparison between HC1 and CRC, 67 metabolites showed  $P < 0.05$  (Mann–Whitney test) and, among them, 59 metabolites showed  $P < 0.05$ , adjusted by FDR. Histidine, for example, had a higher concentration in the HC1 group than in the other two groups. The

Comparison		Age	Gender
HC	P	0.0001	0.8708
HC	CRC	<0.0001	0.0407
P	CRC	0.0641	0.1122

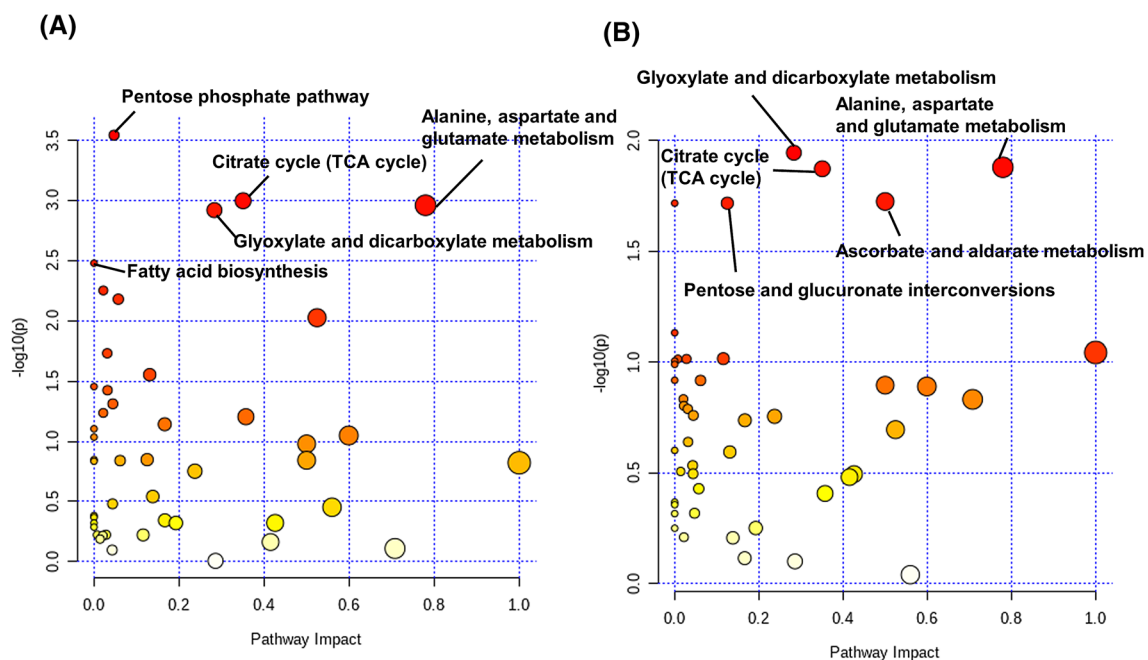
**Table 2.** Summarised characteristics of the three clinical groups. Mann–Whitney test and Chi-square test were used for age and gender, respectively. *CRC* colorectal cancer, *P* polyps, *HC* healthy controls.



**Figure 1.** Heatmap of urinary metabolomic profiles. **(A)** Comparison of metabolites identified in the urine of healthy controls 1 (HC1) and patients with polyps (P) and colorectal cancer (CRC). **(B)** Stage-specific profiles of **(A)**. Concentrations of each metabolite are represented as the average of all patients per group normalised to the total amount of that metabolite detected. Colours represent higher (red) or lower (blue) concentration compared to the average (white). Pearson's correlation was used for clustering. Metabolites detected in more than 50% of samples were included in this analysis. Mann–Whitney was conducted between HC1 and CRC for **(A)** and CRC at stage 0 to III and those with stage IV for **(B)**. The metabolites showing  $P < 0.05$  are coloured light blue and those showing  $P < 0.05$  (adjusted by FDR) are coloured light red. MeV TM4 (ver. 4.7.4, <http://mev.tm4.org/>) was used for generating heatmaps.

P and CRC groups showed substantial overlap, with several metabolites present at higher levels in the P and CRC groups than in the HC1 group; these included choline and single-acetylated polyamines.  $N^1$ -acetylspermidine and  $N^1$ -acetylspermine levels were higher in the P and CRC groups than in the HC1 group. The levels of methionine were highest in the P group. Most amino acids were present at higher levels in the CRC group (Fig. 1A). The metabolic profiles also showed stage-specific patterns in the CRC group. In particular, the profile of stage IV CRC included several metabolites that were present at a higher concentration compared with other stages (Fig. 1B). The comparison of CRC with stage 0 to III and those with stage IV revealed 14 metabolites with  $P < 0.05$  (Mann–Whitney test), whereas no metabolites had  $P < 0.05$ , adjusted by FDR. The kynurenine and polyamines, such as  $N^1$ -acetylspermidine and  $N^8$ -acetylspermidine, were included.

We conducted pathway analyses using the profiled metabolite concentrations. The comparison of HC1 and CRC resulted in the highest  $-\log_{10} P$ -value, indicating the lowest  $P$ -value for the pentose phosphate pathway. Glyoxylate and dicarboxylate metabolism, tricarboxylic acid (TCA) cycle and alanine, aspartate and glutamate metabolism also had high  $-\log_{10} P$ -values ( $> 2.5$ ) and higher pathway impact values (Fig. 2A). In the comparison



**Figure 2.** Pathway analyses using MetaboAnalyst (v.4.0, <https://www.metaboanalyst.ca>). Comparisons of healthy control 1 (HC1) and colorectal cancer (CRC) (A) and polyp (P) and CRC groups (B). The colour and size of each dots indicate  $P$ -value and pathway impact values based on the topological analysis, respectively. Yellow and red indicated higher and lower  $P$ -values. Larger dots indicate a higher impact score. The names of pathways with small  $P$ -values are described.

of P and CRC, glyoxylate and dicarboxylate metabolism, TCA cycle, and alanine, aspartate and glutamate metabolism, pentose and glucuronate interconversions, ascorbate and aldarate metabolism showed higher  $-\log_{10} P$ -values compared with the other pathways (Fig. 2B). Taken together, the TCA cycle is a CRC-specific difference.

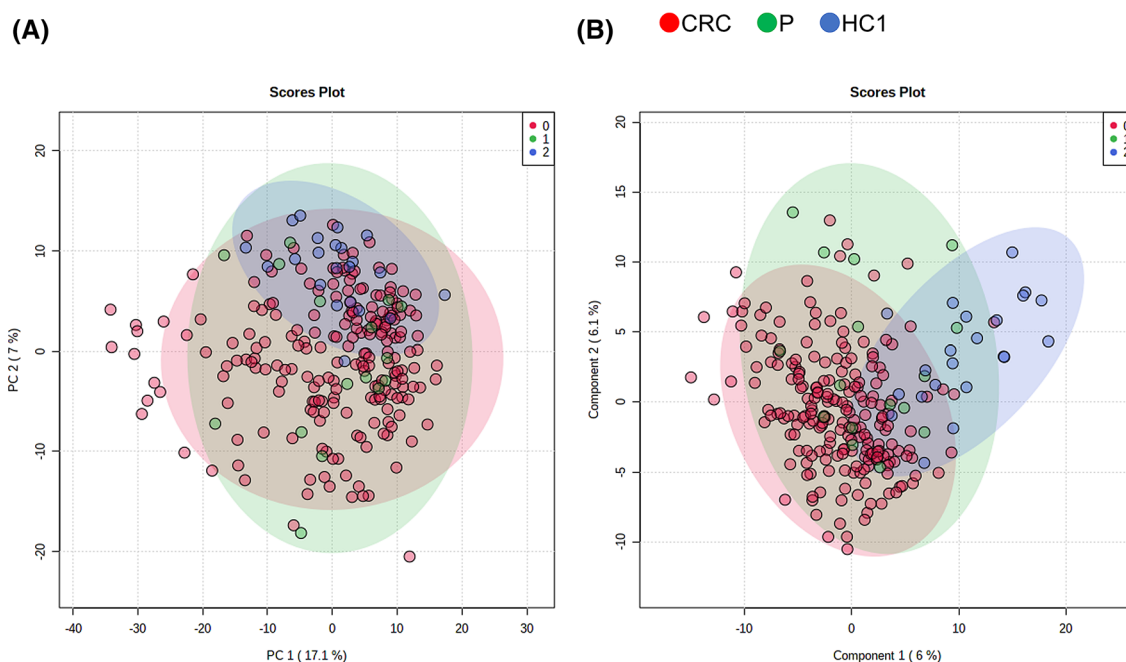
**Differentiation based on urinary metabolites.** We analysed the overall change in the concentrations of the profiled metabolites of the three groups by PCA and PLS-DA (Figs. 3 and S1). The PCA plots showed overlap between all the groups, with the HC1 samples grouped, and the P and CRC samples spread out. The HC1 samples grouped on the right of the PLS-DA, whereas the P and CRC samples were distributed in the area from the centre to the left. The metabolites showing top variable importance in projection (VIP) scores were depicted (Figure S2). Fumarate, 2-isopropylmalate, butyrate, and carnosine showed a significant difference in both HC1 vs P and HC1 vs CRC comparisons (Kruskal–Wallis and Dunn’s post-tests), and the other seven metabolites, such as guanine, showed a significant difference in the HC1 vs CRC comparison.

To assess the variations in the urinary samples collected at different times and days, PLS-DA was conducted using the HC2 data (Figure S3). Most of the plots of P and CRC overlapped at the left side and most of the plots of HC1 and HC2 overlapped at the right side of score plots (Figure S3A). The  $Q^2$  values of the analyses with HC2 (Figure S3C) were higher than those without HC2 (Figure S1C). The metabolites showing high VIP scores, such as fumarate, xanthine, and p-hydroxyphenylacetate, were common in both the PLS-DA analyses (Figure S1B and Figure S3B). These results indicate the high reproducibility between HC1 and HC2 data.

The metabolites with statistically significant differences between the CRC and HC1 groups are depicted in Fig. 4. A total of 52 metabolites were present at different levels between these two groups ( $P < 0.05$ , Mann–Whitney test, adjusted by FDR). The top 10 metabolites with a large fold change (F.C.) difference between these two groups are shown in Fig. 4B and include intermediate metabolites of various pathways, such as malate and citrate (tricarboxylic acid [TCA] cycle), kynurenine (tryptophan pathway), and cystine (pentose phosphate pathway). Among P, CRC, and HC1, only citrate had a statistically significant difference between the HC1 and P groups and between the HC and CRC groups (Kruskal–Wallis test followed by Dunn’s post-hoc test).

We analysed the differentiation ability of a combination of a panel of metabolites. We randomly divided the data, and an MLR model was developed and analysed by tenfold CV using the training data. The model was evaluated using the validation data (Fig. 5A). The MLR differentiated CRC + P from HC1 and the distribution of their predicted probabilities using 10-fold CV and the validation data are shown in Fig. 5B,C. The parameters and selected metabolites of the MLR model are summarised in Table S1. The MLR included butyrate, 3-hydroxy-3-methylglutarate, and carnosine as features. The discriminatory abilities of these putative biomarker panels are summarised in Table S2 and ROC curves are depicted in Figure S4.

Using the selected features, we also developed an MLR model using all the data, that is only numerical coefficients were calculated using the same metabolites (Table S1). This model was also validated using two-fold CV and resampling using all the data (Figure S4C). Based on the ROC curves, the optimal thresholds were calculated for the MLR, and the positive rates based on tumour markers, CEA and CA19-9, predicted by the MLR model,



**Figure 3.** Score plots of principal component analysis (PCA) using MetaboAnalyst (v.4.0, <https://www.metaboolanalyst.ca>). (A) and partial least squares-discrimination analyses (PLS-DA) (B) of metabolites identified in the urine of healthy control 1 (HC1) and patients with polyps (P) or colorectal cancer (CRC), or healthy controls. The data was log-transformed and mean centred for both the analyses. Loading plots of PCA and the VIP score  $R^2$  and  $Q^2$  are summarized in Figure S1.

were compared (Table S3). Table S4 summarises the correlation among tumour markers, values predicted by the MLR model, and each metabolite used for the MLR model. Urinary polyamine concentrations in the data from the present study and those reported previously<sup>9</sup> were compared.

## Discussion

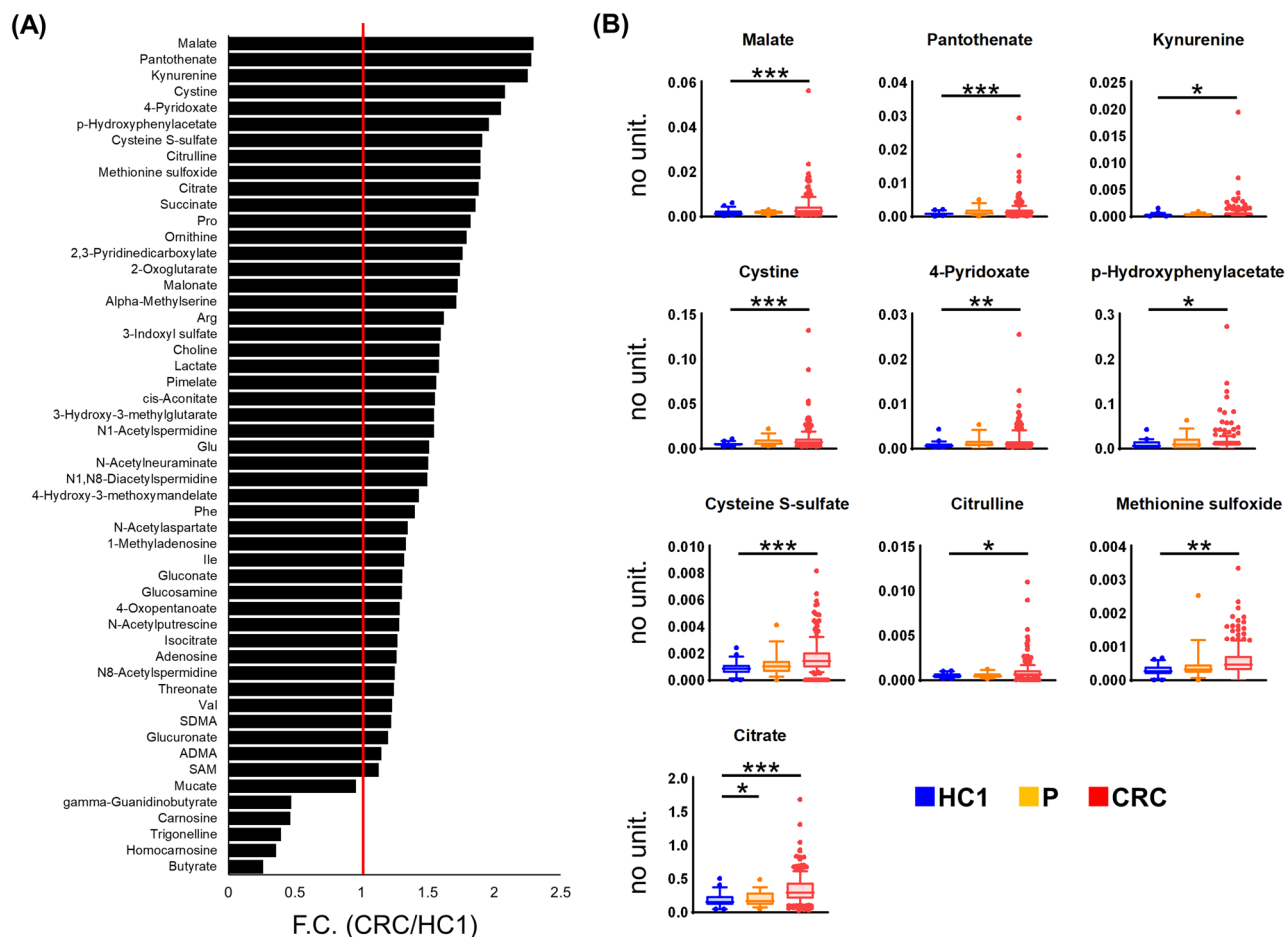
In the present study, non-targeted metabolomic analysis of urine samples collected from subjects with CRC, P, and HC1 was used to understand the changes in the metabolomic profiles. Based on this analysis, the metabolic profiles of each group were unique, and stage-specific differences were observed within the CRC group.

In the CRC group, many metabolites were present at high concentrations, and in particular, the metabolites in the samples of patients with stage IV CRC, which has metastatic potential, were higher than those in the other stages, as well as in the P and HC1 groups (Fig. 1). When comparing HC1 and CRC samples, 52 metabolites were significantly different (FDR-corrected  $P < 0.05$ ) (Fig. 4). However, PCA and PLS-DA analysis were not able to separate these groups based on their metabolic profiles (Figs. 3 and S1).

Based on these results, we developed an MLR model to identify a panel of metabolites that could discriminate between the two groups. The MLR model to discriminate CRC + P from HC showed significant differences in the predicted probabilities (Figs. 5 and S4, and Table S2). Because of the small sample size, especially in the HC1 and P groups, the results may have been caused by bias introduced during the random split to eliminate unexpected bias caused by the difference of the sample size. Therefore, we also developed an MLR model using the metabolites that were already selected in the protocol (Fig. 5A). We conducted 200 times two-fold CVs and confirmed the variance in the obtained AUC values (Figure S4AB). We performed resampling analysis and observed a larger variance in the AUC values compared to the two-fold CV (Figure S4C). In the two-fold CV analysis, all data were also randomly split into two datasets. While the ratio of HC, P, and CRC in the training and validation datasets were almost the same, these ratios in the resampling tests changed, which could have an impact on the accuracy of the resampling analysis. To address this, the model should be tested in a larger cohort with more participants in each of the groups.

The developed MLR model showed higher positive rates (87.1%) compared with the tumour markers CEA (69.9%) and CA19-9 (83.3%) (Table S3). The correlation coefficient ( $R$ ) between CEA and CA19-9 was 0.408 (Table S4), which was the highest correlation coefficient between each tumour marker and the MLR model. Thus, the MLR model can be used as an independent factor to detect CRC patients.

Among the identified metabolites, amino acids were particularly elevated in CRC patients (Fig. 1A), especially in the advanced stages of cancer (Fig. 1B). Profiling of CRC tumour tissue revealed higher concentrations of amino acids than in the adjacent noncancerous tissues<sup>37</sup>. The intermediate metabolites of glycolysis and the TCA cycle are actively used for amino acid synthesis. The Warburg effect, commonly observed in tumour tissue, induces rapid synthesis of adenosine triphosphate (ATP) as an energy source, as well as of nucleic acids, amino acids, and lipids. Activated glycolysis has a rapid reaction rate of oxidative phosphorylation despite the low production efficiency per glucose, which is advantageous for the synthesis of these metabolites without the use of

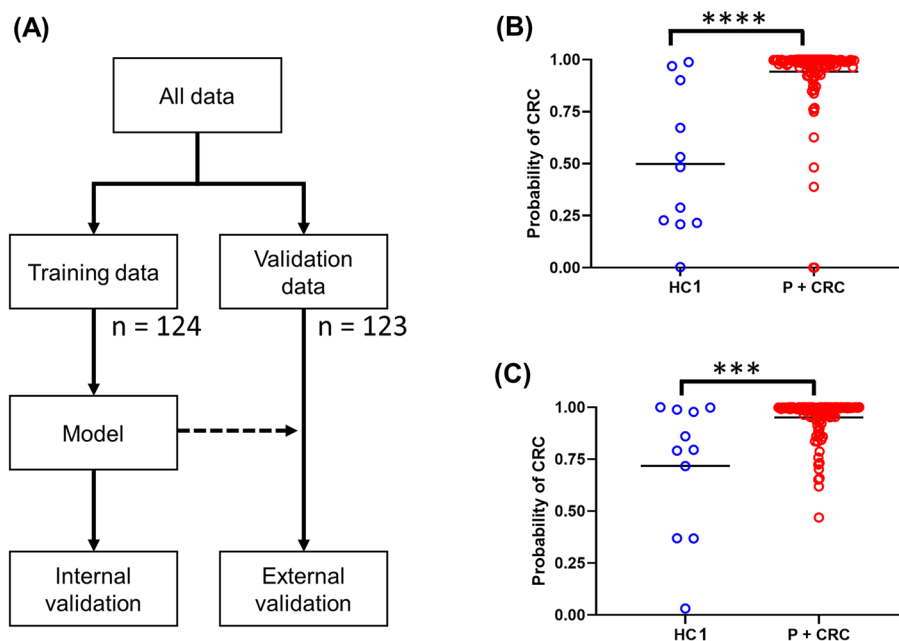


**Figure 4.** The average fold change of metabolites with significantly different levels between healthy control 1 (HC1) and patients with colorectal cancer (CRC) (Mann–Whitney test;  $P < 0.05$  adjusted by FDR) (A). The representative box plots of metabolite concentrations within the top tenfold change values (B). Metabolite concentrations are normalized to that of creatinine for each sample. The comparison among HC1, polyp (P), and CRC groups was performed using Kruskal–Wallis and Dunn’s post-tests. All metabolites shown here had  $P < 0.005$  in the Kruskal–Wallis test, and \*\*\*, \*\*, and \* indicate  $P < 0.0001$ ,  $P < 0.01$ , and  $P < 0.05$  in the post-tests, respectively. GraphPad Prism (ver. 8.4.2, <https://www.graphpad.com/>) was used.

oxidative phosphorylation. The pathway analysis revealed that the TCA cycle exhibited a high difference between the HC1 and CRC groups and also between the CRC and P groups, which indicates that it is highly activated in CRC (Fig. 2). Among the TCA cycle metabolites, malate showed CRC-specific elevation (Fig. 4B). Citric acid, which was elevated in CRC (Fig. 4B), contributes to fatty acid production in cancer cells as an intermediate metabolite in the citric acid cycle. Cancer cells use fatty acids as an energy source, and they have increased de novo lipid metabolism. Many cancer cells undergo apoptosis or growth inhibition upon suppression of their lipid-metabolizing enzymes. The change in citrate concentration of serum samples from CRC patients was also reported<sup>38</sup>. Polyamines, such as *N*<sup>1</sup>-acetylspermidine and *N*<sup>8</sup>-acetylspermidine, were elevated in CRC (Fig. 1A), consistent with our previous study<sup>9</sup>. The synthesis of these metabolites was activated by the activation of ornithine decarboxylase and spermidine/spermine *N*<sup>1</sup>-acetyltransferase<sup>18,39</sup> and also their elevation was frequently detected in urine samples collected from the patients with CRC<sup>40</sup>.

Amino acid synthesis by glutaminolysis, which is part of glutamine metabolism, plays an important role in cancer cell growth<sup>41</sup>. The high concentration of these metabolites in CRCs, which can metastasize to biofluids, especially in the advanced stages of CRC, is possibly caused by increased synthesis of metabolites. The concentration of kynurenine, a major degradation product of tryptophan, was significantly high in the urine collected from the patients with CRC (Fig. 4B). The synthesis of kynurenine is activated by MYC<sup>15</sup> and this metabolite functions to promote cell proliferation in colon cancer cells<sup>42</sup>. Pantothenate, a type of vitamin B, was also significantly elevated in CRC (Fig. 4B). The elevated concentration of pantothenate in colon tissue was reported previously<sup>43</sup>. The elevation of urinary polyamines and amino acid-related metabolites observed in this study is consistent with the elevation in CRC tissue samples reported in various studies.

The MLR model included three metabolites (Table S1). Butyrate, which was found in the developed MLR model, is used as an energy source by colon cells in vivo. It is metabolized into acetyl CoA by  $\beta$ -oxidation and acts in the citric acid cycle to produce energy. Butyrate induces differentiation of undifferentiated T cells into regulatory T cells through epigenetic regulation<sup>44</sup>. Regulatory T cells function in the suppression of the immune



**Figure 5.** Multiple logistic regression (MLR)-based discrimination tests used for both MLR (A) design of data split, (B) internal validation and (C) external validation results of MLR. The corresponding receiver operating characteristic (ROC) curves are depicted in Figure S2. \*\*\*\* $P < 0.0001$  and \*\*\* $P < 0.001$ . GraphPad Prism (ver. 8.4.2, <https://www.graphpad.com/>) was used.

response and play a role in autoimmune diseases, inflammatory diseases, and allergic diseases. Regulatory T cells may suppress the immune response to cancer cells and promote excessive metabolism in these cells.

The intermediate metabolite of leucine, 3-hydroxy-3-methylglutarate, was identified in the developed MLR model. Higher leucine levels were observed in the stools of patients with intramucosal cancer<sup>45</sup>. Carnosine, identified in the MLR model, suppresses cancer cell growth. Wu et al. reported that carnosine inhibits the activation of nuclear factor-kappa B (NF- $\kappa$ B) signalling, and subsequently inhibits metastatic cell adhesion and extravasation in CRCs<sup>46</sup>. Carnosine, which is involved in the suppression of metastasis, might also be useful as a CRC screening marker. Although these metabolites were identified in the present study as potential biomarkers for CRC, the biological mechanisms underlying their high concentration in the urine of CRC patients are not clear.

There are several limitations to this study. The samples size, especially the number of healthy controls and P patients, was small. In the present study, we did not compare the levels of these metabolites in other cancer types to establish the specificity of these biomarkers for CRC. Bias may also be introduced because of the demographic makeup of the cohort, and a larger number of subjects with various backgrounds should be compared to ascertain the specificity of the putative biomarkers. The prediction of metastasis and prognosis using biofluids is also challenging<sup>47</sup>, but was beyond the scope of the present study. The comparison of polyamines measured in the present study and in our previous study (Figure S5) reveals clear linearity between these data.  $N^1$ -acetylspermidine,  $N^8$ -acetylspermidine, and  $N^1, N^8$ -diacetylspermidine were successfully separated and quantified, whereas  $N^1, N^{12}$ -diacetylspermine was not successfully separated by CE under the present conditions.

In the present study, metabolites were identified in the urine of patients with CRC, P, and in HCs. The MLR was used to identify panels of metabolites with the ability to differentiate between these clinical groups. A method to differentiate between the different stages within the HC, CRC, and P groups remains elusive, and further analysis of a larger number of patients is required, particularly regarding early diagnosis at the polyp stage. Further research is expected to lead to the development of an accurate diagnostic method by measuring novel biomarkers for tumour identification. With the development of this future technology, it may be possible to confirm the presence of P or CRC in advance, before performing a colonoscopy, which would enable minimally invasive examination and treatment to be performed. This is expected to contribute to the further spread of medical examinations for CRC in Japan, and finally, reduce the harm caused by this disease.

Received: 5 July 2020; Accepted: 19 November 2020

Published online: 03 December 2020

## References

1. Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).



2. American Society of Clinical Oncology. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. Adopted on May 17, 1996 by the American Society of Clinical Oncology. *J. Clin. Oncol.* **14**, 2843–2877. <https://doi.org/10.1200/jco.1996.14.10.2843> (1996).
3. Locker, G. Y. *et al.* ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clin. Oncol.* **24**, 5313–5327. <https://doi.org/10.1200/jco.2006.08.2644> (2006).
4. Bagaria, B., Sood, S., Sharma, R. & Lalwani, S. Comparative study of CEA and CA19-9 in esophageal, gastric and colon cancers individually and in combination (ROC curve analysis). *Cancer Biol. Med.* **10**, 148–157. <https://doi.org/10.7497/j.issn.2095-3941.2013.03.005> (2013).
5. Flamini, E. *et al.* Free DNA and carcinoembryonic antigen serum levels: an important combination for diagnosis of colorectal cancer. *Clin. Cancer Res.* **12**, 6985–6988. <https://doi.org/10.1158/1078-0432.Ccr-06-1931> (2006).
6. Tan, E. *et al.* Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. *Surg. Oncol.* **18**, 15–24. <https://doi.org/10.1016/j.suronc.2008.05.008> (2009).
7. Schreuders, E. H. *et al.* Colorectal cancer screening: a global overview of existing programmes. *Gut* **64**, 1637–1649. <https://doi.org/10.1136/gutjnl-2014-309086> (2015).
8. Morikawa, T. *et al.* A comparison of the immunochemical fecal occult blood test and total colonoscopy in the asymptomatic population. *Gastroenterology* **129**, 422–428. <https://doi.org/10.1016/j.gastro.2005.05.056> (2005).
9. Nakajima, T. *et al.* Urinary polyamine biomarker panels with machine-learning differentiated colorectal cancers, benign disease, and healthy controls. *Int. J. Mol. Sci.* **19**, 756. <https://doi.org/10.3390/ijms19030756> (2018).
10. Saito, H. Screening for colorectal cancer by immunochemical fecal occult blood testing. *Jpn. J. Cancer Res.* **87**, 1011–1024. <https://doi.org/10.1111/j.1349-7006.1996.tb03103.x> (1996).
11. van Rossum, L. G. *et al.* Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. *Gastroenterology* **135**, 82–90. <https://doi.org/10.1053/j.gastro.2008.03.040> (2008).
12. Milovic, V. & Turchanowa, L. Polyamines and colon cancer. *Biochem. Soc. Trans.* **31**, 381–383. <https://doi.org/10.1042/bst0310381> (2003).
13. Ning, W. *et al.* Identification of differential metabolic characteristics between tumor and normal tissue from colorectal cancer patients by gas chromatography-mass spectrometry. *Biomed. Chromatogr. BMC* <https://doi.org/10.1002/bmc.3999> (2017).
14. Satoh, K. *et al.* Global metabolic reprogramming of colorectal cancer occurs at adenoma stage and is induced by MYC. *Proc. Natl. Acad. Sci. USA* **114**, E7697–E7706. <https://doi.org/10.1073/pnas.1710366114> (2017).
15. Venkateswaran, N. *et al.* MYC promotes tryptophan uptake and metabolism by the kynurenine pathway in colon cancer. *Genes Dev.* **33**, 1236–1251. <https://doi.org/10.1101/gad.327056.119> (2019).
16. Gerner, E. W. & Meyskens, F. L. Jr. Polyamines and cancer: old molecules, new understanding. *Nat. Rev. Cancer* **4**, 781–792. <https://doi.org/10.1038/nrc1454> (2004).
17. Casero, R. A. Jr., Murray Stewart, T. & Pegg, A. E. Polyamine metabolism and cancer: treatments, challenges and opportunities. *Nat. Rev. Cancer* **18**, 681–695. <https://doi.org/10.1038/s41568-018-0050-3> (2018).
18. Soda, K. The mechanisms by which polyamines accelerate tumor spread. *J. Exp. Clin. Cancer Research CR* **30**, 95. <https://doi.org/10.1186/1756-9966-30-95> (2011).
19. Loser, C., Folsch, U. R., Paprotny, C. & Creutzfeldt, W. Polyamines in colorectal cancer. Evaluation of polyamine concentrations in the colon tissue, serum, and urine of 50 patients with colorectal cancer. *Cancer* **65**, 958–966 (1990).
20. Venalainen, M. K. *et al.* Altered polyamine profiles in colorectal cancer. *Anticancer Res.* **38**, 3601–3607. <https://doi.org/10.21873/anticancerres.12634> (2018).
21. Hiramatsu, K. *et al.* N(1), N(12)-Diacylspermine as a sensitive and specific novel marker for early- and late-stage colorectal and breast cancers. *Clin. Cancer Res.* **11**, 2986–2990. <https://doi.org/10.1158/1078-0432.CCR-04-2275> (2005).
22. Kawakita, M., Hiramatsu, K., Yanagiya, M., Doi, Y. & Kosaka, M. Determination of N(1), N(1)(2)-diacylspermine in urine: a novel tumor marker. *Methods Mol. Biol.* **720**, 367–378. [https://doi.org/10.1007/978-1-61779-034-8\\_23](https://doi.org/10.1007/978-1-61779-034-8_23) (2011).
23. Umemori, Y. *et al.* Evaluating the utility of N1,N12-diacylspermine and N1,N8-diacylspermidine in urine as tumor markers for breast and colorectal cancers. *Clinica Chimica Acta Int J. Clin. Chem.* **411**, 1894–1899. <https://doi.org/10.1016/j.cca.2010.07.018> (2010).
24. Farshidfar, F. *et al.* A validated metabolomic signature for colorectal cancer: exploration of the clinical value of metabolomics. *Br. J. Cancer* **115**, 848–857. <https://doi.org/10.1038/bjc.2016.243> (2016).
25. Zhang, Y. *et al.* Profiling of serum metabolites in advanced colon cancer using liquid chromatography-mass spectrometry. *Oncol. Lett.* **19**, 4002–4010. <https://doi.org/10.3892/ol.2020.11510> (2020).
26. Deng, L. *et al.* Urinary metabolomics to identify a unique biomarker panel for detecting colorectal cancer: a multicenter study. *Cancer Epidemiol. Biomarkers Prev.* **28**, 1283–1291. <https://doi.org/10.1158/1055-9965.EPI-18-1291> (2019).
27. Eisner, R., Greiner, R., Tso, V., Wang, H. & Fedorak, R. N. A machine-learned predictor of colonic polyps based on urinary metabolomics. *Biomed. Res. Int.* **2013**, 303982. <https://doi.org/10.1155/2013/303982> (2013).
28. Erben, V., Bhardwaj, M., Schrotz-King, P. & Brenner, H. Metabolomics biomarkers for detection of colorectal neoplasms: a systematic review. *Cancers (Basel)* **10**, 246. <https://doi.org/10.3390/cancers10080246> (2018).
29. Qiu, Y. *et al.* Urinary metabolomic study on colorectal cancer. *J. Proteome Res.* **9**, 1627–1634. <https://doi.org/10.1021/pr901081y> (2010).
30. Ishikawa, S. *et al.* Identification of salivary metabolomic biomarkers for oral cancer screening. *Sci. Rep.* **6**, 31520. <https://doi.org/10.1038/srep31520> (2016).
31. Sugimoto, M., Kawakami, M., Robert, M., Soga, T. & Tomita, M. Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. *Curr. Bioinform.* **7**, 96–108. <https://doi.org/10.2174/157489312799304431> (2012).
32. Sugimoto, M., Wong, D. T., Hirayama, A., Soga, T. & Tomita, M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics* **6**, 78–95. <https://doi.org/10.1007/s11306-009-0178-y> (2010).
33. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **46**, 389–422 (2002).
34. Hall, M. A. & Smith, L. A. Practical feature subset selection for machine learning (1998).
35. Chong, J., Wishart, D. S. & Xia, J. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr. Protoc. Bioinform.* **68**, e86. <https://doi.org/10.1002/cpbi.86> (2019).
36. Frank, E., Hall, M., Trigg, L., Holmes, G. & Witten, I. H. Data mining in bioinformatics using Weka. *Bioinformatics* **20**, 2479–2481 (2004).
37. Hirayama, A. *et al.* Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Can. Res.* **69**, 4918–4925. <https://doi.org/10.1158/0008-5472.Can-08-4806> (2009).
38. Gu, J. *et al.* Metabolomics analysis in serum from patients with colorectal polyp and colorectal cancer by (1)H-NMR spectrometry. *Dis. Markers* **2019**, 3491852. <https://doi.org/10.1155/2019/3491852> (2019).
39. Weiss, T. S. *et al.* Polyamine levels of human colorectal adenocarcinomas are correlated with tumor stage and grade. *Int. J. Colorectal Dis.* **17**, 381–387. <https://doi.org/10.1007/s00384-002-0394-7> (2002).
40. O'Brien, B. L. *et al.* Urinary N1-acetylspermidine and N8-acetylspermidine excretion in normal humans and in patients with colorectal cancer. *Dig. Dis. Sci.* **40**, 1269–1274 (1995).

41. DeBerardinis, R. J. *et al.* Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl. Acad. Sci. USA* **104**, 19345–19350. <https://doi.org/10.1073/pnas.0709747104> (2007).
42. Venkateswaran, N. & Conacci-Sorrell, M. Kynurenine: an oncometabolite in colon cancer. *Cell Stress* **4**, 24–26. <https://doi.org/10.15698/cst2020.01.210> (2020).
43. Baker, H. *et al.* Elevated vitamin levels in colon adenocarcinoma as compared with metastatic liver adenocarcinoma from colon primary and normal adjacent tissue. *Cancer* **47**, 2883–2886. [https://doi.org/10.1002/1097-0142\(19810615\)47:12%3c2883::aid-cnrcr2820471222%3e3.0.co;2-i](https://doi.org/10.1002/1097-0142(19810615)47:12%3c2883::aid-cnrcr2820471222%3e3.0.co;2-i) (1981).
44. Hori, S., Nomura, T. & Sakaguchi, S. Control of regulatory T cell development by the transcription factor Foxp3. *Science (New York, N.Y.)* **299**, 1057–1061. <https://doi.org/10.1126/science.1079490> (2003).
45. Yachida, S. *et al.* Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* **25**, 968–976. <https://doi.org/10.1038/s41591-019-0458-7> (2019).
46. Wu, C. C., Lai, P. Y., Hsieh, S., Cheng, C. C. & Hsieh, S. L. Suppression of carnosine on adhesion and extravasation of human colorectal cancer cells. *Anticancer Res.* **39**, 6135–6144. <https://doi.org/10.21873/anticancer.13821> (2019).
47. Martín-Blázquez, A. *et al.* Untargeted LC-HRMS-based metabolomics to identify novel biomarkers of metastatic colorectal cancer. *Sci. Rep.* **9**, 20198. <https://doi.org/10.1038/s41598-019-55952-8> (2019).

## Acknowledgements

We thank all the sample providers. This work was supported by grants from Yamagata Prefecture and Tsuruoka city.

## Author contributions

R.U, K.K, H.K, M.E, T.I. R.S., and Y.N. collected samples. M.S. and A.T. designed this study. M.S. analysed the results. R.U, K.K., and M.S. wrote the manuscript. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-78038-2>.

**Correspondence** and requests for materials should be addressed to M.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2020