



Rectal cancer

Predicting the pathological response to neoadjuvant chemoradiation using untargeted metabolomics in locally advanced rectal cancer

Huixun Jia^{a,b,1}, Xiaotao Shen^{c,f,1}, Yun Guan^{b,d}, Meimei Xu^{c,f}, Jia Tu^{c,f}, Miao Mo^{a,b}, Li Xie^g, Jing Yuan^{a,b}, Zhen Zhang^{b,d}, Sanjun Cai^{b,e}, Ji Zhu^{b,d,*}, ZhengJiang Zhu^{c,f,*}

^a Clinical Statistics Center, Fudan University Shanghai Cancer Center; ^b Department of Oncology, Shanghai Medical College, Fudan University; ^c Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences; ^d Department of Radiation Oncology, Fudan University Shanghai Cancer Center; ^e Department of Colorectal Surgery, Fudan University Shanghai Cancer Center; ^f University of Chinese Academy of Sciences, Beijing; and ^g Clinical Research Center, Shanghai Jiao Tong University School of Medicine, PR China

ARTICLE INFO

Article history:

Received 9 January 2018

Received in revised form 19 April 2018

Accepted 14 June 2018

Available online 2 July 2018

Keywords:

Rectal cancer

Neoadjuvant chemoradiotherapy

Metabolomics

ABSTRACT

Purpose: The present study aimed to identify a panel of potential metabolite biomarkers to predict tumor response to neoadjuvant chemo-radiation therapy (NCRT) in locally advanced rectal cancer (LARC).

Experimental design: Liquid chromatography–mass spectrometry (LC–MS)-based untargeted metabolomics was used to profile human serum samples ($n = 106$) from LARC patients treated with NCRT. The samples were collected from Fudan University Shanghai Cancer Center (FUSCC) from July 2014 to January 2016. Statistical methods, such as partial least squares (PLS) and Wilcoxon rank-sum test, were used to identify discriminative metabolites between NCRT-sensitive and NCRT-resistant patients according to their tumor regression grade (TRG). This trial is registered with ClinicalTrials.gov, number NCT03149978.

Results: A panel of metabolites was selected as potential predictive biomarkers of pathological response to NCRT. A total of 4810 metabolic peaks were detected, and 57 significantly dysregulated peaks were identified. These 57 metabolic peaks were used to differentiate patients using PLS in a dataset containing NCRT-sensitive ($n = 56$) and NCRT-resistant ($n = 49$) patients. The combination of 57 metabolic peaks had AUC values of 0.88, 0.81 and 0.84 in the prediction models using PLS, random forest, and support vector machine, respectively, suggesting that metabolomics has the potential ability to predict responses to NCRT. Furthermore, 15 metabolite biomarkers were identified and used to construct a logistic regression model and explore dysregulated metabolic pathways using untargeted metabolic profiling and data mining approaches.

Conclusions: A panel of metabolites has been identified to facilitate the prediction of tumor response to NCRT in LARC, which is promising for the generation of personalized treatment strategies for LARC patients.

© 2018 The Author(s). Published by Elsevier B.V. Radiotherapy and Oncology 128 (2018) 548–556 This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

For locally advanced rectal cancer (LARC) patients, preoperative neoadjuvant chemo-radiation therapy (NCRT) followed by total mesorectal excision (TME) is the standard treatment [1–3]. Previous studies have shown that NCRT has better local control and lower toxicity than adjuvant chemoradiotherapy [4–6]. In clinical practice, however, pathological responses to NCRT demonstrated obvious heterogeneity. Approximately 10–30% of patients show pathologic complete response, 40–45% show variant tumor regres-

sion, and the remaining 20–30% have no response to NCRT [7–9]. Therefore, more cost-effective, predictive biomarkers of NCRT for LARC patients are needed to maximize patient benefits and minimize adverse effects.

To date, the pathological tumor response to NCRT has been evaluated using imaging of tumor morphology. Monguzzi L et al. reported that the mean value of the apparent diffusion coefficient from MRI could be used to predict the pathological response to NCRT [10], which was confirmed by Genovesi D et al. in a single-site prospective study [11]. Similarly, Zhang et al. reported the use of the standardized uptake value from positron emission tomography (PET) scans to assess the pathological response of LARC patients to NCRT [12]. However, the changes in tumor morphology typically occur later than the changes at the biological

* Corresponding authors at: Department of Radiation Oncology, Fudan University Shanghai Cancer Center, Shanghai 200032, China (J. Zhu). Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, PR China (Z. Zhu).

E-mail addresses: leo.zhu@126.com (J. Zhu), jiangzhu@sioc.ac.cn (Z. Zhu).

¹ These authors made equally contributions.

and molecular levels. Therefore, image diagnosis is not an ideal method to evaluate the NCRT sensitivity at early stages and cannot accurately predict therapeutic responses prior to NCRT.

With recent advancements in molecular and systems biology, molecular biomarkers have been discovered to predict the response of rectal cancer patients to NCRT. Huh et al. reported that high levels of CD44 mRNA in pretreatment biopsies are associated with a poor tumor regression [13]. Yan et al. reported that patients with high Smac levels, low Ki-67 expression and negative vascular endothelial growth factor (VEGF) expression are more sensitive to NCRT [14]. Kim et al. reported that the candidate markers CORO2A rs1985859 and the putative marker FAM101A rs7955740 are valuable to predict sensitivity to NCRT [15]. These three studies focused on investigating the association between the specific molecular markers or single-nucleotide polymorphisms (SNPs) of genes and the pathological response to NCRT. However, the generalization of these studies is limited by a wide confidence interval of the prediction, a retrospective study design or a small sample size. In addition, these results explored only the association rather than the predictive capability of these biomarkers toward the therapeutic response of NCRT in LARC patients.

Metabolomics provides the global and quantitative measurement of endogenous small molecular metabolites within a biological system [16–18] and is well acknowledged in cancer research [19–21]. Metabolomics simultaneously measures thousands of metabolites, and describes a holistic and dynamic profile of disease progression [22]. Currently, metabolomics has been successfully utilized in biomarker discovery for the early diagnosis of cancer, targeted therapy and response prediction [23–27]. Studies confirmed the feasibility of metabolic markers in early diagnosis of colorectal cancer. Wei J et al. analyzed the serum samples of colorectal cancer patients and non-cancer subjects. Oleamine, pyruvic acid, three carboxylic acid and ornithine cycle related metabolites, which are closely associated with the occurrence of colorectal cancer, were identified as differentially expressed markers between colorectal cancer patients and non-intestine cancer patients [28]. In their subsequent study with a larger sample size, urine samples were analyzed to make early diagnosis of colorectal cancer and differentially expressed metabolic markers, including cresol and aminobutyric acid, etc. were screened out. The area under the receiver operating characteristic (ROC) curve was 0.998 in the test sample, while the diagnostic sensitivity and specificity of serum CEA in colorectal cancer is only about 50–70% [29]. Uchiyama et al. successfully distinguished patients between colorectal cancer and intestinal adenoma by the screened serum metabolic markers [30]. However, to the best of our knowledge, no studies have reported the use of metabolite biomarkers for predicting response to NCRT at a personalized level.

In the present study, we designed a prospective cohort study of locally advanced rectal cancer to identify potential metabolite biomarkers for the prediction of tumor response to NCRT. The cohort recruitment began in July of 2014, with a target enrollment of 300 patients. For each patient, serum and urine samples before and during preoperative chemo-radiation therapy were obtained for metabolomics. Here, we report the results from the first 106 patients who were prospectively recruited in the study.

Materials and methods

Eligibility criteria

Between July 2014 and January 2016, a total of 106 patients with clinical T3–4 and/or N+ rectal cancer without distant metastasis were enrolled at Fudan University Shanghai Cancer Center (FUSCC). The following inclusion criteria were considered: (1) patients scheduled to receive CRT followed by TME surgery; (2)

without metabolic diseases, such as diabetes mellitus or hyperthyroidism; and (3) informed consent was signed and obtained before the treatment. The CRT included intensity-modulated radiation therapy (IMRT) of 50 Gy in 25 fractions concurrently with capecitabine-based chemotherapy. Two weeks after the completion of CRT, one additional cycle of chemotherapy was administered according to the guidelines of the center. Surgery was scheduled at 8 weeks after the completion of CRT. TME was mandatory, whereas the form of surgery (anterior resection or abdominal-perineal resection) and whether a temporary colostomy should be performed were decided by the surgeon. The study was approved by the institutional review board of FUSCC.

Pathological evaluation of tumor response

Pathological tumor response was evaluated according to the 2010 American Joint Committee on Cancer (AJCC) tumor regression grade (TRG) system, which recorded the degree and the volume of residual primary tumor cells. Details of AJCC TRG system are defined as follows: Grade 0, defined as no viable cancer cells; Grade 1, characterized by single or small groups of tumor cells; Grade 2, involves residual cancer outgrown by fibrosis, but fibrosis still predominates; and Grade 3, defined as the minimal or no tumor cells killed. The NCRT-sensitive patients were defined as those with TRG Grades 0–1, while the NCRT-resistant patients were defined as those with TRG Grades 2–3.

Collection of serum samples

According to a published protocol [31], biological samples of the enrolled patients were collected in five consecutive time-points: baseline (within two weeks before beginning of CRT); early-phase CRT (5 fractions after beginning); middle-phase CRT (15 fractions after beginning); late-phase CRT (25 fractions after beginning); and surgery (within 2 days before surgery). All participants were in an overnight fasting state, and 5 mL of peripheral venous blood was drawn in the morning. The blood was allowed to clot for 30 min, followed by centrifugation at 3000 rpm for 15 min. Then the serum supernatant was collected, separated into 5 aliquots (200 μ L of each aliquot) and immediately frozen in liquid nitrogen. The serum samples were then stored at -80°C until further analyses. However, only baseline serum samples were analyzed in the current study.

Reagents and materials

LC–MS-grade water (H_2O), acetonitrile (ACN), methanol (MeOH), 0.1% formic acid (FA) in water and 0.1% FA in ACN were purchased from Honeywell (Muskegon, MI, USA). Ammonium fluoride (NH_4F) was purchased from Sigma-Aldrich (St. Louis, MO, USA) and dissolved in LC–MS-grade water prior to use.

Serum sample preparation

The details of serum sample preparation, LC–MS analysis, data processing and preparation have been described in a previous publication [31] and are provided in the [Supporting Information](#).

LC–MS analysis

All serum samples were randomly injected during data acquisition. During data acquisition, blank samples (75% methanol in water) and QC samples (prepared by pooling aliquots of all subject samples) were injected every 8 samples, and the test mixture

(mixture of three internal standards) was injected every 20 samples to monitor the reproducibility of LC–MS system (Fig. S1).

The LC–MS analysis was performed using a UHPLC system (1290 series, Agilent Technologies, California, Santa Clara, USA) coupled to a quadrupole time-of-flight (QTOF) mass spectrometer (Agilent 6550 iFunnel QTOF, Agilent Technologies, USA) in positive mode. Tandem mass spectrometry (MS/MS) data acquisition was performed using another QTOF mass spectrometer (TripleTOF 6600, SCIEX, CANADA). The details of LC–MS analysis are provided in the [Supporting Information](#).

Data processing and data preparation

MS raw data (.d) files were converted to the mzXML format using ProteoWizard and processed using *R* package XCMS (version 3.2) [32,33]. The processing results generated a data matrix that consisted of the retention time (RT), mass-to-charge ratio (m/z) values, and peak abundance. *R* package CAMERA [34] was used for peak annotation after XCMS data processing. Metabolic peaks detected less than 80% [35] in all the QC samples were discarded. To remove the unwanted analytical variations occurring intra- and inter-batches, each metabolic peak in all subject samples was normalized using *R* package MetNormalizer [31]. Metabolic peaks with RSDs larger than 30% in QC samples were removed from the dataset.

Data quality assessment

The data quality was evaluated using the test mixture, which comprised three internal standards (Table S2) and was used to evaluate the status of the LC–MS system during data acquisition. The detailed information is provided in the [Supporting Information](#).

Metabolite identification

The acquired MS/MS spectra were matched against an in-house standard MS/MS spectral library and online METLIN MS/MS spectral database for metabolite identification [36]. The metabolic peaks that were not matched in the database were further analyzed manually using the characteristic fragment ions. The detailed information is provided in the [Supporting Information](#). Finally, the identifications were designated as levels 1, 2, 3, 4 and unknown according to the MSI (Metabolomics Standard Initiative) [37].

Statistical analyses

The categorical data are described as the frequency counts and percentages, and the values of all continuous variables are presented as the mean plus or minus the standard deviation. Comparisons between the patient characteristics in the NCRT-sensitive and -resistant patients were performed using Pearson's chi-square test (discontinuous variable) or Student's *t* test (continuous variables).

Partial least squares (PLS, *R* package "pls") analysis was used to understand global metabolic changes between NCRT-sensitive and -resistant patients, and corresponding variable importance in the projection (VIP values) were calculated in PLS model. Metabolic peaks with VIP larger than 1 were considered potential biomarkers for further study. A validation plot was used to assess the validity of PLS model by comparing the goodness of fit (R^2 and Q^2) of the PLS models with the goodness of fit of 200 Y-permuted models.

Unsupervised analysis, principle component analysis (PCA, *R* function "prcomp") and hierarchical cluster analysis (HCA, *R* package "pheatmap") were also utilized to understand global metabolic changes between NCRT-sensitive and -resistant patients.

Most metabolic peaks showed right-skewed distribution (Fig. S6); thus, the nonparametric Wilcoxon rank-sum test and the median fold-change between two groups were used to select the potential biomarkers. Metabolic peaks with $P < 0.05$ and extreme fold-change (fold-change values greater than the 99.75th percentile or less than the 0.25th percentile of all fold-change values) were identified.

Four statistical models, PLS (*R* package "pls"), random forest (*R* package "randomforest"), support vector machine (SVM) (*R* package "e1071") and logistic regression (*R* function "glm") were used to build prediction models. The bootstrap method was used to minimize the bias and improve the precision of prediction [38]. Briefly, 105 randomly sampled patients from the dataset sample with replacement (about 63% of the patients in average) were selected as discovery data to build the prediction model. The remaining about 37% of the patients in average were used as validation data. Notably, there is no overlap between discovery and validation data. The area under receiver operating characteristic (ROC) curve (AUC) (*R* package "pROC") was used to evaluate the performance of the prediction model. This procedure was repeated 1,000 times, and the median of AUCs was regarded as the final AUC. The 0.25th and 99.75th percentiles were used to calculate the 95% confidence interval (95% CI). All statistical analyses were performed using *R* software (version 3.1.3) [39].

Results

Characteristics of enrolled patients

In this prospective study, patient A092a was considered an outlier and was removed from the dataset (Fig. S3a and S3b); therefore, a total of 105 subject samples were used for subsequent statistical analysis. Among the patients who underwent NCRT followed by TME surgery, 56 were NCRT sensitive, and 49 were NCRT resistant (Table 1 and Fig. 1d). Detailed clinical characteristics for the participants are shown in Table 1 and Table S3. No significant differences were observed regarding sex, age, BMI, distance from anal verge, length, clinical T stage, clinical N stage, and TRG between the two groups ($P > 0.05$).

Metabolic profiles of serum samples

In the present study, all serum samples were obtained from patients prior to NCRT. A total of 4810 metabolic peaks in positive mode were detected using the untargeted metabolomics approach. The base peak chromatograms (BPC) of the NCRT-sensitive patients (Fig. 1a) and NCRT-resistant patients (Fig. 1b) showed no differences. A two-dimensional heat plot was used to visualize the distribution of 4810 metabolic peaks (Fig. 1c). The X-axis and Y-axis indicate the retention time (minutes) and mass-to-charge ratio (m/z), respectively, while the colors indicate the intensities with a log10 scale. The PLS score plot was demonstrated to visualize the classification performance of the metabolic profiles (Fig. 1e). From the PLS score plot, NCRT-sensitive and -resistant patients could be clearly discriminated, and 1661 metabolic peaks with VIP larger than 1 were considered potential biomarkers (Fig. 1f) and studied in the subsequent analyses.

Discovery of discriminative metabolites

For 1661 metabolic peaks with VIP larger than 1, Wilcoxon rank-sum test was employed to determine a subset of metabolic peaks with the highest and independent capability to predict the pathological response prior to NCRT. A total of 197 metabolic peaks with VIP > 1 and $P < 0.05$ were selected (Fig. 2a), which can also

Table 1

Clinical characteristics of NCRT-sensitive and NCRT-resistant patients.

	Neoadjuvant chemotherapy (NCRT)		P value
	Sensitive	Resistant	
Patient number	56	49	–
Sex, male, n (%)	41 (73.2%)	28 (57.1%)	0.1273 ^a
Age (median, min–max)			
BMI	23.0 ± 2.8	22.4 ± 2.8	0.2649 ^b
Distance from anal verge (median, cm),	5.2 ± 1.9	5.0 ± 1.7	0.5145 ^b
Length (median, cm)	5.5 ± 1.8	5.3 ± 1.9	0.9869 ^b
TRG			
0	26	–	–
1	30	–	–
2	–	43	–
3	–	6	–
Clinical T stage			0.9700 ^a
T3	48	42	
T4	8	7	
Clinical N stage			0.7396 ^a
N0	1	2	
N1	24	19	
N2	31	28	

^a Chi-squared test.^b Student *t* test.

effectively discriminate NCRT-sensitive and NCRT-resistant patients in the PLS analysis (Fig. 2b). Two unsupervised multivariate analysis approaches, PCA (Fig. 2c) and HCA (Fig. 2d), were also used to observe the performance of 197 metabolic peaks discriminating NCRT-resistant patients from most NCRT-sensitive patients.

However, the presence of 197 metabolic peaks would be challenging for clinical diagnosis. Then, the fold-change (NCRT-resistant patients divided by NCRT-sensitive patients) was further used to filter the significantly changed metabolic peaks. The cutoff of the fold-change values was set as larger than 1.2 or less than 0.83. As a result, 57 out of 197 metabolic peaks were selected (Fig. 3a) with 26 increasing and 31 decreasing in NCRT-resistant patients. The 57 metabolic peaks could also discriminate most NCRT-resistant patients from NCRT-sensitive patients in PLS score analysis (Fig. 3b). Next, we evaluated the power of predictive model using the bootstrap method (see Materials and Methods). The combination of these 57 metabolic peaks had AUC values of 0.87 (95% CI 0.75–0.96), 0.83 (95% CI 0.67–0.95) and 0.85 (0.73–0.94) in the PLS, random forest, and SVM prediction models, respectively. To avoid the over-fitting effect, we utilized two strategies. First, we randomly selected 57 metabolic peaks as reference peaks and built prediction models (Fig. S4d), and the AUC values were 0.55 (95% CI 0.45–0.68), 0.56 (95% CI 0.44–0.69) and 0.55 (0.45–0.67) in the PLS, random forest, and SVM prediction models, respectively. Second, we utilized the 57 metabolic peaks with the permuted Y labels of patients to build prediction models (Fig. S4e), and the AUC values were 0.58 (95% CI 0.44–0.73), 0.64 (95% CI 0.46–0.79) and 0.61 (0.45–0.77) in the PLS, random forest, and SVM prediction models, respectively. These results demonstrated that the prediction models with 57 metabolic peaks are not over-fitting.

Stepwise selection of potential metabolite biomarkers

We manually analyzed and evaluated the 57 discriminative metabolic peaks, and 19 metabolic peaks without good peak

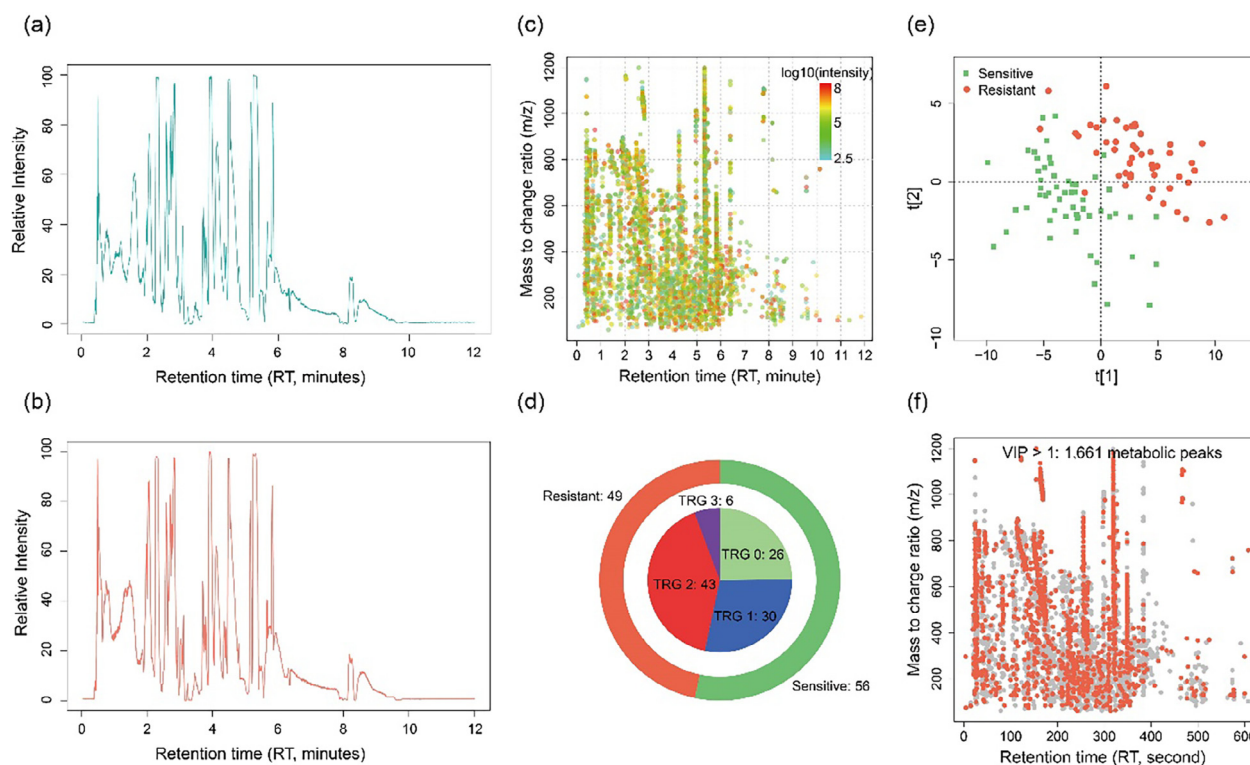


Fig. 1. Metabolic profiles of LARC patients. (a, b) The base peak chromatograms (BPC) of the serum samples from NCRT-sensitive (a) and NCRT-resistant patients (b). (c) Two-dimensional heat plot of metabolic peaks: x-axis is retention time (minute), y-axis is mass to charge ratio (*m/z*), and the colors indicate the scale of peak intensity with a log10 scale. (d) The distribution of TRG scores: TRG 0 and TRG 1 patients are grouped as the NCRT-sensitive group, while TRG 2 and 3 patients are grouped as the NCRT-resistant group. (e) Multivariate analysis PLS was used to observe the overall metabolic profile. (f) Two-dimensional heat plot of metabolic peaks with VIP values larger than 1 (labeled red).

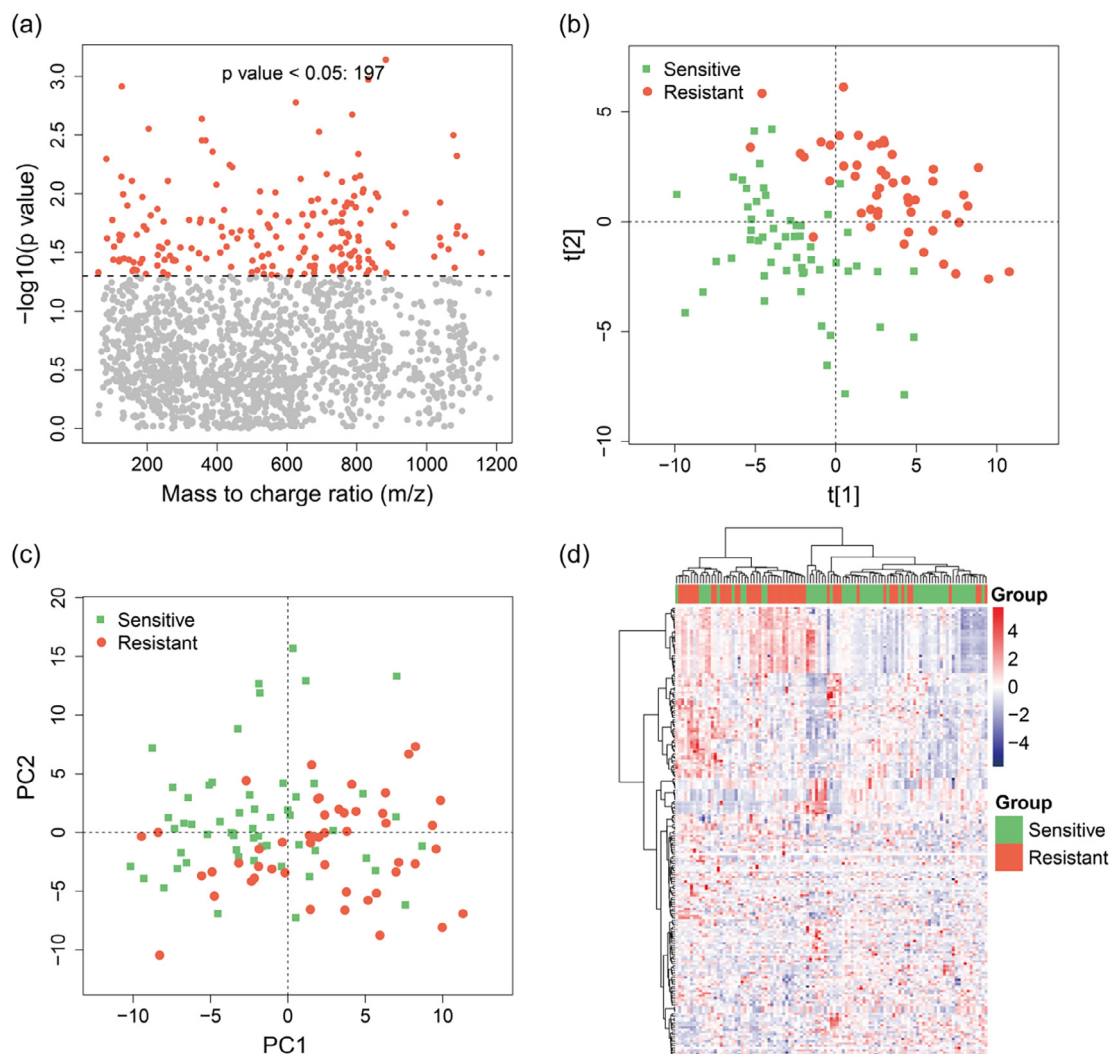


Fig. 2. Discovery of discriminative metabolites using univariate and multivariate analyses. (a) A total of 197 metabolic peaks with VIP > 1 and P values < 0.05 were selected for subsequent analysis. (b) PLS score plot of 197 metabolic peaks. (c) PCA score plot of 197 metabolic peaks. (d) Heat map of HCA of 197 metabolic peaks (Euclidian distance, Ward aggregation).

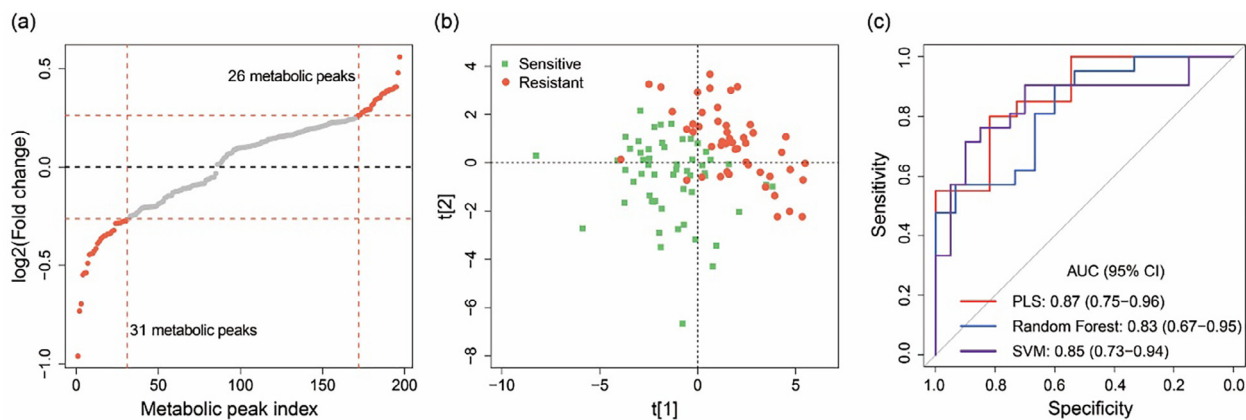


Fig. 3. The validation using selected 57 metabolitic peaks. (a) A total of 57 metabolitic peaks were selected by fold-change > 1.2 or < 0.83. (b) PLS score plot of 57 metabolitic peaks. (c) The ROCs of predictive models with 57 metabolitic peaks: red, PLS model; blue, random forest model; and purple, support vector machine (SVM) model.

shapes were removed (Fig. S5). The remaining 38 metabolic peaks were sorted according to their VIP values. First, we utilized all 38 metabolic peaks to build a PLS prediction model (see Materials and Methods). In brief, 105 randomly sampled patients from the dataset sample with replacement (about 63% of the patients in average) were selected as discovery data to build the PLS prediction model. The remaining about 37% of the patients in average were used as validation data. Notably, there is no overlap between discovery and validation data. This process was repeated 1000 times, and the AUCs with 38 metabolic peaks were recorded. Then, a stepwise procedure was performed, where the metabolic peak with the smallest VIP was removed in each step, and a PLS prediction model was built (about 63% of the patients in average) and validated with the remaining patients (remaining about 37% of the patients in average, Fig. 4a) as above described. The detailed code is provided in the [Supplementary](#). As shown in Fig. 4a, the top 15 metabolic peaks were appropriate and sufficient to build the predictive model with a good AUC value. Thus, we identified the metabolite structures of the 15 metabolic peaks (Fig. 4b and Fig. S7–9). Detailed structural information of the 15 metabolic peaks (metabolite biomarkers) is listed in Table 2. The AUC value of the 15 metabolite markers was 0.80, and the 95% CI was 0.67–0.91 in the NCRT response in the PLS prediction model (Fig. 4c).

Development of logistic regression models with metabolite biomarkers

We developed logistic regression models with the 15 metabolite biomarkers and risk factors to determine whether the risk factors can improve prediction performance. The first prediction model contained only risk factors (clinical T stage, clinical N stage, tumor distance to anal verge and tumor length), and the AUC value was 0.60 (95% CI 0.44–0.73). The second model contained only the 15 metabolite biomarkers, and the AUC value was 0.76 (95% CI 0.58–0.89). The third model integrated the risk factors and the metabolite biomarkers, with an AUC value of 0.75 (95% CI 0.57–0.88). These results demonstrated that the prediction model with metabolite biomarkers was better than the prediction models with risk factors (Fig. 4d). We also developed logistic regression models with the 5 metabolite biomarkers (4-Imidazoleacetic acid, δ -Valerolactam, N-Methylethanolamine phosphate, PC (16:0/18:1), PC (9:0)), and the AUC corresponding to this number of metabolites is 0.74 (95% CI: 0.60–0.86, Fig. S10). We also calculated the correlations of 15 metabolite biomarkers (Fig. S11), and there are not clear and significant correlations between those 15 metabolite biomarkers.

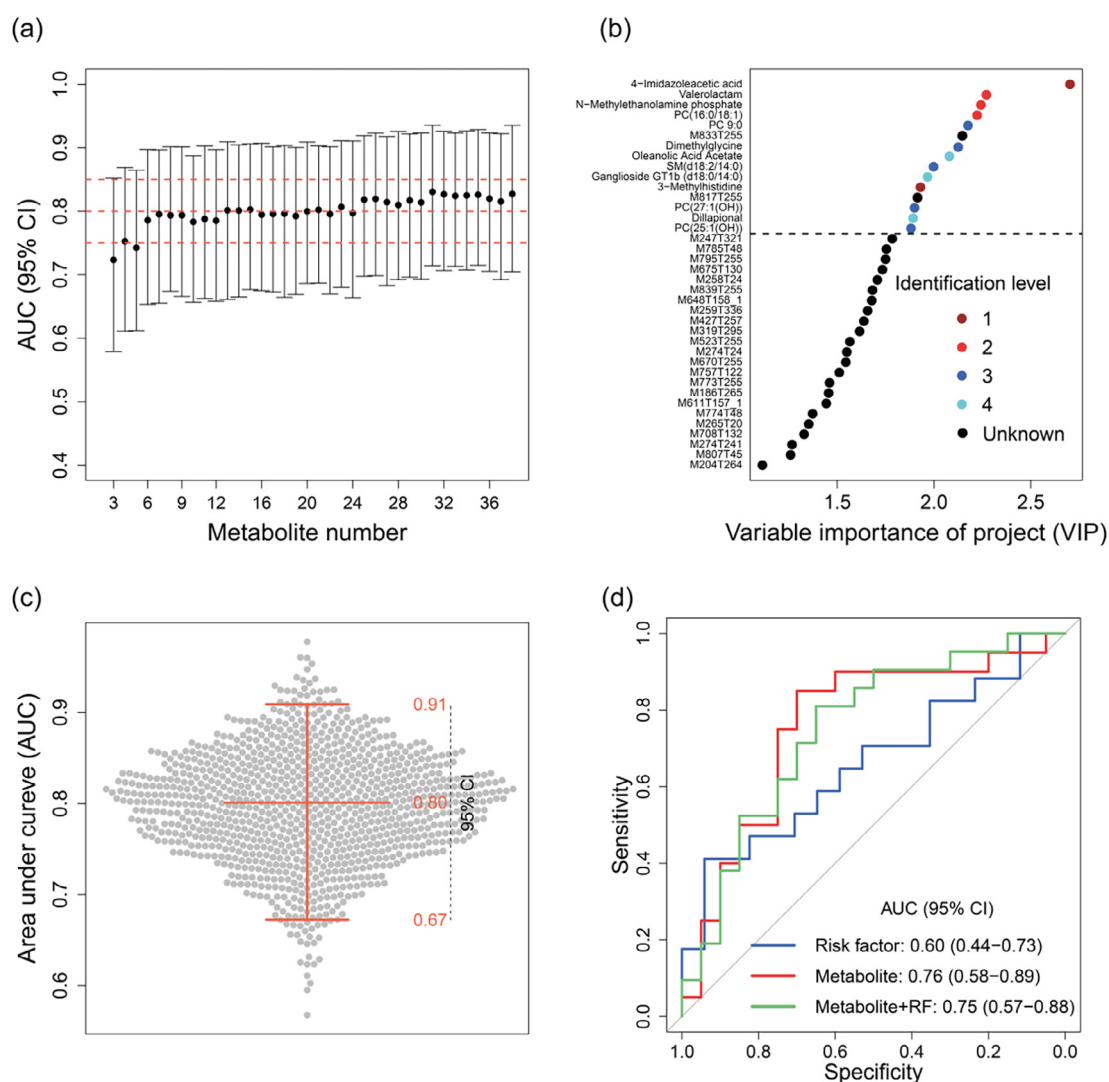


Fig. 4. Building of prediction model. (a) Step-wise selection of potential metabolite biomarkers. AUC distribution of PLS prediction models with different number of metabolite biomarkers. (b) Structure identification of top 15 metabolic peaks according to VIP values. (c) AUC distribution of PLS prediction model with the top 15 metabolic markers. (d) The ROCs of predictive models using logistic regression with risk factors, metabolite biomarkers and both combined.

Table 2

List of identified 15 metabolic markers between NCRT-sensitive and NCRT-resistant patients.

Metabolic peak name	<i>m/z</i> (Da)	RT (second)	Identification	Identification level	VIP ^a	Fold change ^b	<i>P</i> value ^c	CV
M127T112	127.0515	112.34	4-Imidazoleacetic acid	1	2.7027	0.82	0.0009	0.048
M100T44	100.0760	43.75	δ-Valerolactam	2	2.2707	0.60	0.0313	0.034
M156T112	156.0420	111.79	N-Methylethanolamine phosphate	2	2.2423	0.62	0.0233	0.119
M805T45	804.5522	45.25	PC(16:0/18:1)	2	2.223	1.32	0.0171	0.108
M412T212	412.2097	212.17	PC (9:0)	3	2.1755	0.82	0.0320	0.036
M833T255	833.0088	255.39	Unknown	–	2.1463	1.31	0.0091	0.131
M229T311_1	229.1189	310.93	Dimethylglycine	3	2.1212	1.31	0.0290	0.047
M501T111	501.3929	110.77	Oleanolic Acid Acetate	4	2.079	0.79	0.0464	0.180
M696T156	695.5097	156.40	SM(d18:2/14:0)	3	1.9978	1.27	0.0336	0.080
M692T255	692.0215	255.27	Ganglioside GT1b (d18:0/14:0)	4	1.9664	1.22	0.0149	0.030
M170T372	170.0929	372.26	3-Methylhistidine	1	1.9299	0.69	0.0479	0.080
M817T255	817.0395	255.25	Unknown	–	1.9148	1.27	0.0243	0.026
M678T134	678.4689	133.91	PC(27:1(OH))	3	1.8994	0.75	0.0479	0.022
M278T269	278.1022	269.09	Dillapional	4	1.8917	0.82	0.0464	0.121
M650T136	650.4397	136.11	PC(25:1(OH))	3	1.8809	0.79	0.0483	0.034

^a VIP: variable importance in the projection of PLS model.^b Fold-change: NCRT-resistant patients divided by NCRT-sensitive patients.^c Wilcoxon rank-sum test.

Metabolic dysregulation associated with the NCRT response

We mapped the top 15 metabolic biomarkers into the KEGG database and found that 4-imidazoleacetic acid and 3-methylhistidine were involved in the pathway of histidine metabolism, dimethylglycine was involved in the pathways of glycine, serine and threonine metabolism, and N-methylethanolamine phosphate was involved in the pathway of glycerophospholipid metabolism. These results suggested that metabolic dysregulation is closely associated with the response to NCRT.

Discussion

Currently, pathological TRG assessment is the most reliable standard for the evaluation of response to NCRT. However, TRG results are available only after the completion of both NCRT and surgery; thus, these results cannot be used as a reference to adjust the treatment strategy before or during therapy. In this cohort study, we proposed metabolomics as an approach for evaluating the sensitivity to NCRT using significantly changed metabolites from fasting serum samples. According to the discriminative analysis, the PLS model was constructed with 57 significant metabolic peaks that performed well with an AUC value of 0.88. Although there was no significant difference regarding the clinic pathological features between NCRT-sensitive and -resistant patients, the metabolomic profiles effectively categorized those patients into two distinct groups. Furthermore, four amino acids and eleven lipids were identified, and the PLS model constructed with these fifteen metabolites showed an ideal discriminant performance with an AUC value of 0.80. The results indicated that differentially expressed metabolite biomarkers facilitated the discrimination of the patients who are likely or unlikely to benefit from NCRT. The satisfactory prediction performance through the serum metabolites suggested that metabolomics might be a new approach in terms of selecting LARC patients for personalized treatment strategy most suitable in clinical practice. In addition, the metabolites detected by LC–MS provide better insights into the metabolism mechanism from molecular level.

The alternation of the amino acid metabolism is a consequence of the metabolic dysfunction in cancer. In the present study, 4 amino acids out of 15 identified metabolites showed significantly differential expression in the serum samples from the resistant group compared with the sensitive group. Previous studies also

demonstrated higher concentrations of various amino acids in colorectal cancer [40,41], but they have not to be explored as biomarkers for predicting tumor response to NCRT in LARC. The metabolites 4-imidazoleacetic acid and 3-methylhistidine were involved in the histidine metabolism pathway. Histidine is a half-essential amino acid, and histidine deficiency has been implicated in tumor growth and the promotion of inflammation-associated carcinogenesis [42]. We searched the Human Metabolome Database (HMDB), and histidine was found to be correlated with cancer [43–47], obesity [48] Alzheimer's disease [49], etc. Our study demonstrated that 4-imidazoleacetic acid and 3-methylhistidine were both down-regulated in the serum of NCRT-resistant patients. Thus, depleted 4-imidazoleacetic acid and 3-methylhistidine in the serum of resistant patients may indicate the higher absorption of 4-imidazoleacetic acid and 3-methylhistidine in tumor cells to sustain rapid cell proliferation.

We also found that dimethylglycine was involved in the glycine metabolism pathway. Sreekumar et al. confirmed that glycine was closely associated with tumor cell growth [50]. Lin et al. also reported the decreased expression of dimethylglycine in colorectal cancer patients compared with healthy controls [51]. Benito et al. suggested dimethylglycine as a potential biomarker for renal impairment [52]. Liu et al. also reported that the expression of dimethylglycine dehydrogenase is decreased in hepatocellular carcinoma [53]. In addition, according to the HMDB database, dimethylglycine was detected in Leukemia [54] and some solid cancers [53], especially in colorectal cancer [55,56]. The change and variation in the amino acid profiles might be attributable to the gluconeogenesis and protein synthesis [57]. Consequently, the fact that amino acids differed between sensitive and resistant patients might be a consequence of the destruction of the cellular proteins of the tumor.

Furthermore, the sample size in the present study was larger than those in other similar studies, which were usually between 40 and 80 patients [58–60]. However, compared with the high dimensional metabolomics data, the present sample size is still limited. Therefore, three statistical models were used to build prediction models and allow for the inclusion of a large number of variables. Since the training and testing were performed on the same set, a bootstrap method was used to minimize bias and improve prediction precision.

Ideally, response prediction should be performed prior to NCRT to offer the patient personalized treatment with a higher chance of success and better benefits. Although the present results showed

that metabolite biomarkers have the potential predictive capability for the response of NCRT, external validation of this prediction model is necessary for a systematic evaluation of this method. Another interesting issue is whether the changes of metabolite levels during NCRT can further enhance the performance of this model, which should be studied in the future.

In conclusion, we demonstrated that metabolomics studies on patient serum samples were able to discriminate NCRT-sensitive from NCRT-resistant patients, facilitating the development of a new predictive tool for LARC with a serum metabolite test. The results of the present prospective study showed promising results that a total of 15 differently expressed metabolites could be used to identify the potential NCRT-sensitive LARC patients; thus, beneficial functional and clinical outcomes can be achieved without complete resection. In contrast, the successful identification of NCRT-resistant patients would reduce their likelihood to receive ineffective NCRT treatment, and directing these patients to surgery would be a better course of action. As oncology research evolves toward personalized medicine, the method proposed in the present study has important implications in clinical practice. However, these findings require further large-scale and multi-institutional validation for confirmation of the applicability of metabolomics for predicting the response to NCRT.

Translational relevance

Reliable predictive markers of NCRT are needed to maximize patient benefits. Dynamic changes of metabolites in LARC patients make these molecules potential candidate biomarkers. In the present study, we identified a panel of metabolite markers with great potential for an early assessment of NCRT response and described the underlying pathways to support potential metabolic mechanisms. The results revealed the great value of metabolomics in tailoring treatment strategies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.radonc.2018.06.022>.

References

- [1] van de Velde CJ, Boelens PG, Borras JM, Coebergh JW, Cervantes A, Blomqvist L, et al. EURECCA colorectal: multidisciplinary management: European consensus conference colon & rectum. *Eur J Cancer* 2014;50:1.
- [2] van de Velde CJ, Aristei C, Boelens PG, Beets-Tan RG, Blomqvist L, Borras JM, et al. EURECCA colorectal: multidisciplinary mission statement on better care for patients with colon and rectal cancer in Europe. *Eur J Cancer* 2013;49:2784–90.
- [3] Valentini V, Aristei C, Glimelius B, Minsky BD, Beets-Tan R, Borras JM, et al. Multidisciplinary rectal cancer management: 2nd European rectal cancer consensus conference (EURECCA-CC2). *Radiother Oncol* 2009;92:148–63.
- [4] Sauer R, Becker H, Hohenberger W, Rodel C, Wittekind C, Fietkau R, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med* 2004;351:1731–40.
- [5] Rodel C, Liersch T, Becker H, Fietkau R, Hohenberger W, Hothorn T, et al. Preoperative chemoradiotherapy and postoperative chemotherapy with fluorouracil and oxaliplatin versus fluorouracil alone in locally advanced rectal cancer: initial results of the German CAO/ARO/AIO-04 randomised phase 3 trial. *Lancet Oncol* 2012;13:679–87.
- [6] Sauer R, Liersch T, Merkel S, Fietkau R, Hohenberger W, Hess C, et al. Preoperative versus postoperative chemoradiotherapy for locally advanced rectal cancer: results of the German CAO/ARO/AIO-94 randomized phase III trial after a median follow-up of 11 years. *J Clin Oncol* 2012;30:1926–33.
- [7] Onaitis MW, Noone RB, Hartwig M, Hurwitz H, Morse M, Jowell P, et al. Neoadjuvant chemoradiation for rectal cancer: analysis of clinical outcomes from a 13-year institutional experience. *Ann Surg* 2001;233:778–85.
- [8] Kaminsky-Forreth MC, Conroy T, Luporsi E, Peiffert D, Lapeyre M, Boissel P, et al. Prognostic implications of downstaging following preoperative radiation therapy for operable T3–T4 rectal cancer. *Int J Radiat Oncol Biol Phys* 1998;42:935–41.
- [9] Kuremsky JG, Tepper JE, McLeod HL. Biomarkers for response to neoadjuvant chemoradiation for rectal cancer. *Int J Radiat Oncol Biol Phys* 2009;74:673–88.
- [10] Monguzzi L, Ippolito D, Bernasconi DP, Trattenero C, Galimberti S, Sironi S. Locally advanced rectal cancer: value of ADC mapping in prediction of tumor response to radiochemotherapy. *Eur J Radiol* 2013;82:234–40.
- [11] Genovesi D, Filippone A, Ausili CG, Trignani M, Vinciguerra A, Augurio A, et al. Diffusion-weighted magnetic resonance for prediction of response after neoadjuvant chemoradiation therapy for locally advanced rectal cancer: preliminary results of a mono-institutional prospective study. *Eur J Surg Oncol* 2013;39:1071–8.
- [12] Zhang C, Tong J, Sun X, Liu J, Wang Y, Huang G. 18F-FDG-PET evaluation of treatment response to neo-adjuvant therapy in patients with locally advanced rectal cancer: a meta-analysis. *Int J Cancer* 2012;131:2604–11.
- [13] Huh JW, Lee JH, Kim HR. Pretreatment expression of 13 molecular markers as a predictor of tumor responses after neoadjuvant chemoradiation in rectal cancer. *Ann Surg* 2014;259:508–15.
- [14] Yan H, Wang R, Yu J, Jiang S, Zhu K, Mu D, et al. Predictive value of Smac, VEGF and Ki-67 in rectal cancer treated with neoadjuvant therapy. *Oncol Lett* 2010;1:641–7.
- [15] Kim JC, Ha YJ, Roh SA, Cho DH, Choi EY, Kim TW, et al. Novel single-nucleotide polymorphism markers predictive of pathologic response to preoperative chemoradiation therapy in rectal cancer patients. *Int J Radiat Oncol Biol Phys* 2013;86:350–7.
- [16] Nicholson JK, Lindon JC. Systems biology: Metabonomics. *Nature* 2008;455:1054–6.
- [17] De Preter V. Metabonomics and systems biology. *Methods Mol Biol* 2015;1277:245–55.
- [18] van der Greef J, van Wietmarschen H, van Ommen B, Verheij E. Looking back into the future: 30 years of metabolomics at TNO. *Mass Spectrom Rev* 2013;32:399–415.
- [19] Farshidfar F, Weljie AM, Kopciuk KA, Hilsden R, McGregor SE, Buie WD, et al. A validated metabolomic signature for colorectal cancer: exploration of the clinical value of metabolomics. *Br J Cancer* 2016;115:848–57.
- [20] Corona G, Polesel J, Fratino L, Miolo G, Rizzolio F, Crivellari D, et al. Metabonomics biomarkers of frailty in elderly breast cancer patients. *J Cell Physiol* 2014;229:898–902.
- [21] Snyder NW, Mesaros C, Blair IA. Translational metabolomics in cancer research. *Biomark Med* 2015;9:821–34.
- [22] Gowda GA, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabonomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn* 2008;8:617–33.
- [23] Spratlin JL, Serkova NJ, Eckhardt SG. Clinical applications of metabolomics in oncology: a review. *Clin Cancer Res* 2009;15:431–40.
- [24] Zhang T, Wu X, Ke C, Yin M, Li Z, Fan L, et al. Identification of potential biomarkers for ovarian cancer by urinary metabolomic profiling. *J Proteome Res* 2013;12:505–12.
- [25] Tenori L, Oakman C, Morris PG, Gralka E, Turner N, Cappadona S, et al. Serum metabolomic profiles evaluated after surgery may identify patients with oestrogen receptor negative early breast cancer at increased risk of disease recurrence. Results from a retrospective study. *Mol Oncol* 2015;9:128–39.
- [26] Ke C, Hou Y, Zhang H, Fan L, Ge T, Guo B, et al. Large-scale profiling of metabolic dysregulation in ovarian cancer. *Int J Cancer* 2015;136:516–26.
- [27] Hou Y, Yin M, Sun F, Zhang T, Zhou X, Li H, et al. A metabolomics approach for predicting the response to neoadjuvant chemotherapy in cervical cancer patients. *Mol Biosyst* 2014;10:2126–33.
- [28] Qiu Y, Cai G, Su M, Chen T, Zheng X, Xu Y, et al. Serum metabolite profiling of human colorectal cancer using GC-TOFMS and UPLC-QTOFMS. *J Proteome Res* 2009;8:4844–50.
- [29] Cheng Y, Xie G, Chen T, Qiu Y, Zou X, Zheng M, et al. Distinct urinary metabolic profile of human colorectal cancer. *J Proteome Res* 2012;11:1354–63.
- [30] Uchiyama K, Yagi N, Mizushima K, Higashimura Y, Hirai Y, Okayama T, et al. Serum metabolomics analysis for early detection of colorectal cancer. *J Gastroenterol* 2017;52:677–94.
- [31] Shen X, Gong X, Cai Y, Guo Y, Tu J. Normalization and integration of large-scale metabolomics data using support vector regression. *Metabolomics* 2016;12:89.
- [32] Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 2006;78:779–87.
- [33] Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinform* 2008;9:504.
- [34] Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem* 2012;84:283–9.
- [35] Dunn WB, Broadhurst D, Begley P, Zelena E, Francis-McIntyre S, Anderson N, et al. Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2011;6:1060–83.

- [36] Schultz AW, Wang J, Zhu ZJ, Johnson CH, Patti GJ. Liquid chromatography quadrupole time-of-flight characterization of metabolites guided by the METLIN database. *Nat Protoc* 2013;8:451–60.
- [37] Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. *Metabolomics* 2007;3:211–21.
- [38] Zhao J, Zhu Y, Hyun N, Zeng D, Uppal K, Tran VT, et al. Novel Metabolic Markers for the Risk of Diabetes Development in American Indians. *Diabetes Care* 2015;38:220–7.
- [39] R: A Language and Environment for Statistical Computing.
- [40] Denkert C, Budczies J, Weichert W, Wohlgemuth G, Scholz M, Kind T, et al. Metabolite profiling of human colon carcinoma—deregulation of TCA cycle and amino acid turnover. *Mol Cancer* 2008;7:72.
- [41] Ong ES, Zou L, Li S, Cheah PY, Eu KW, Ong CN. Metabolic profiling in colorectal cancer reveals signature metabolic shifts during tumorigenesis. *Mol Cell Proteomics* 2010.
- [42] Yang XD, Ai W, Asfaha S, Bhagat G, Friedman RA, Jin G, et al. Histamine deficiency promotes inflammation-associated carcinogenesis through reduced myeloid maturation and accumulation of CD11b+Ly6G+ immature myeloid cells. *Nat Med* 2011;17:87–95.
- [43] Ni Y, Xie G, Jia W. Metabonomics of human colorectal cancer: new approaches for early diagnosis and biomarker discovery. *J Proteome Res* 2014;13:3857–70.
- [44] Ritchie SA, Ahiaonu PW, Jayasinghe D, Heath D, Liu J, Lu Y, et al. Reduced levels of hydroxylated, polyunsaturated ultra long-chain fatty acids in the serum of colorectal cancer patients: implications for early screening and detection. *BMC Med* 2010;8:13.
- [45] Goedert JJ, Sampson JN, Moore SC, Xiao Q, Xiong X, Hayes RB, et al. Fecal metabolomics: assay performance and association with colorectal cancer. *Carcinogenesis* 2014;35:2089–96.
- [46] Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M. Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. *Metabolomics* 2010;6:78–95.
- [47] De Angelis M, Piccolo M, Vannini L, Siragusa S, De Giacomo A, Serrazanetti DI, et al. Fecal microbiota and metabolome of children with autism and pervasive developmental disorder not otherwise specified. *PLoS One* 2013;8:e76993.
- [48] Wahl Simone, Holzapfel Christina, Zhonghao Yu, Breier Michaela, Kondofersky Ivan, Fuchs Christiane, et al. Metabolomics reveals determinants of weight loss during lifestyle intervention in obese children. *Metabolomics* 2013.
- [49] Fonteh AN, Harrington RJ, Tsai A, Liao P, Harrington MG. Free amino acid and dipeptide changes in the body fluids from Alzheimer's disease subjects. *Amino Acids* 2007;32:213–24.
- [50] Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 2009;457:910–4.
- [51] Lin Y, Ma C, Liu C, Wang Z, Yang J, Liu X, et al. NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. *Oncotarget* 2016;7:29454–64.
- [52] Benito S, Sanchez A, Unceta N, Andrade F, Aldamiz-Echevarria L, Goicolea MA, et al. LC-QTOF-MS-based targeted metabolomics of arginine-creatine metabolic pathway-related compounds in plasma: application to identify potential biomarkers in pediatric chronic kidney disease. *Anal Bioanal Chem* 2016;408:747–60.
- [53] Liu G, Hou G, Li L, Li Y, Zhou W, Liu L. Potential diagnostic and prognostic marker dimethylglycine dehydrogenase (DMGDH) suppresses hepatocellular carcinoma metastasis in vitro and in vivo. *Oncotarget* 2016;7:32607–16.
- [54] Peng CT, Wu KH, Lan SJ, Tsai JJ, Tsai FJ, Tsai CH. Amino acid concentrations in cerebrospinal fluid in children with acute lymphoblastic leukemia undergoing chemotherapy. *Eur J Cancer* 2005;41(8):1158–63.
- [55] Wang X, Wang J, Rao B, Deng L. Gut flora profiling and fecal metabolite composition of colorectal cancer patients and healthy individuals. *Exp Ther Med* 2017;13:2848–54.
- [56] Lundholm K, Bylund AC, Holm J, Schersten T. Skeletal muscle metabolism in patients with malignant tumor. *Eur J Cancer* 1976;12:465–73.
- [57] Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal cancer: assessment of neoadjuvant chemoradiation outcome based on radiomics of multiparametric MRI. *Clin Cancer Res* 2016;22:5256–64.
- [58] Kim K, Yeo SG, Yoo BC. Identification of hypoxanthine and phosphoenolpyruvic Acid as serum markers of chemoradiotherapy response in locally advanced rectal cancer. *Cancer Res Treat* 2015;47:78–89.
- [59] Roh K, Yeo SG, Yoo BC, Kim KH, Kim SY, Kim MJ. Seven low-mass ions in pretreatment serum as potential predictive markers of the chemoradiotherapy response of rectal cancer. *Anticancer Drugs* 2016;27:787–93.