

Development of a Correlative Strategy To Discover Colorectal Tumor Tissue Derived Metabolite Biomarkers in Plasma Using Untargeted Metabolomics

Zhuozhong Wang,^{†,‡,△} Binbin Cui,^{§,△} Fan Zhang,^{||} Yue Yang,[⊥] Xiaotao Shen,^{‡,#} Zhong Li,[§] Weiwei Zhao,[†] Yuanyuan Zhang,[†] Kui Deng,[†] Zhiwei Rong,[†] Kai Yang,[†] Xiwen Yu,^{||} Kang Li,^{*,†} Peng Han,^{*,§} and Zheng-Jiang Zhu^{*,‡,⊕}

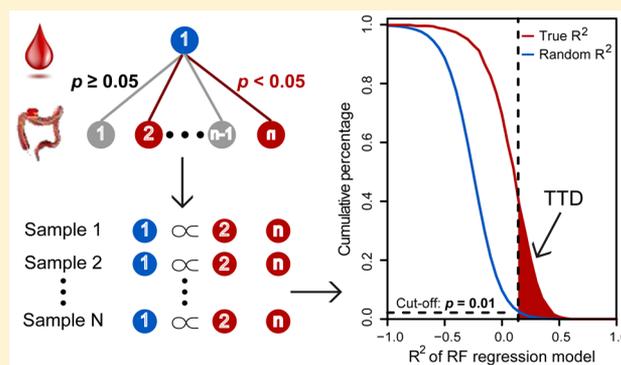
[†]Department of Epidemiology and Biostatistics, School of Public Health, [§]Department of Colorectal Surgery, Harbin Medical University Cancer Hospital, ^{||}Laboratory of Hematology Center, The First Affiliated Hospital of Harbin Medical University, [⊥]Department of Pathology, Harbin Medical University Cancer Hospital, and ^{||}Heilongjiang Academy of Medical Sciences, Harbin Medical University, Harbin 150086, P.R. China

[‡]Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, P.R. China

[#]University of Chinese Academy of Sciences, Beijing 100049, P.R. China

Supporting Information

ABSTRACT: The metabolic profiling of biofluids using untargeted metabolomics provides a promising choice to discover metabolite biomarkers for clinical cancer diagnosis. However, metabolite biomarkers discovered in biofluids may not necessarily reflect the pathological status of tumor tissue, which makes these biomarkers difficult to reproduce. In this study, we developed a new analysis strategy by integrating the univariate and multivariate correlation analysis approach to discover tumor tissue derived (TTD) metabolites in plasma samples. Specifically, untargeted metabolomics was first used to profile a set of paired tissue and plasma samples from 34 colorectal cancer (CRC) patients. Next, univariate correlation analysis was used to select correlative metabolite pairs between tissue and plasma, and a random forest regression model was utilized to define 243 TTD metabolites in plasma samples. The TTD metabolites in CRC plasma were demonstrated to accurately reflect the pathological status of tumor tissue and have great potential for metabolite biomarker discovery. Accordingly, we conducted a clinical study using a set of 146 plasma samples from CRC patients and gender-matched polyp controls to discover metabolite biomarkers from TTD metabolites. As a result, eight metabolites were selected as potential biomarkers for CRC diagnosis with high sensitivity and specificity. For CRC patients after surgery, the survival risk score defined by metabolite biomarkers also performed well in predicting overall survival time ($p = 0.022$) and progression-free survival time ($p = 0.002$). In conclusion, we developed a new analysis strategy which effectively discovers tumor tissue related metabolite biomarkers in plasma for cancer diagnosis and prognosis.



Untargeted metabolomics measures the alterations of metabolic profiles from tissue and biofluid samples in relevance to the disease phenotype and aims to discover metabolite biomarkers for clinical applications, such as cancer diagnosis and prognosis.^{1,2} However, in untargeted metabolomics studies of biofluid samples, unwanted confounding factors irrelevant to diseases often lead to the discovery of false positive biomarkers.³ For example, metabolic profiles of human blood and urine are significantly affected by gender, lifestyle, diet, etc.^{4–6} Some physiological confounding factors, such as age and gender, could be controlled or partially removed through a proper study design. Nevertheless, many other confounding factors, such as lifestyle, diet, and medication, are difficult to quantify and evaluate.³ Therefore, metabolite

biomarkers discovered in biofluids may not necessarily reflect the pathological status of tumor tissue, which makes these biomarkers difficult to reproduce.⁷ For example, Cheng et al.⁸ and Leichtle et al.⁹ found that alanine was decreased in urine and serum from colorectal cancer (CRC) patients, but Qiu et al.¹⁰ reported that alanine was significantly increased in colorectal tumor tissue. Sreekumar et al. revealed the potential role of sarcosine in tumor tissues for the pathological progression of prostate cancer, but they failed to discover

Received: November 9, 2018

Accepted: December 23, 2018

Published: December 23, 2018

any significant changes of sarcosine in plasma and urine samples from prostate cancer patients.¹¹ Accordingly, very few metabolite biomarkers in biofluids have been translated into clinical cancer diagnosis, due to their poor capability to indicate the metabolic dysregulation induced by tumor tissue.

To address this challenge, many efforts were made to improve the reliability and reproducibility through analyzing multiple types of samples. For instance, Hori et al. performed a metabolomics study on tissue and serum samples from lung cancer patients and healthy controls and found that metabolites simultaneously dysregulated in tumor tissue and serum exhibited similar changes in both sample types, including lactic acid, fumaric acid, malic acid, proline, and threonine.¹² However, no studies have yet systematically assessed the consistency of metabolic dysregulation between tumor tissue and biofluid samples. Particularly, no one has developed an effective method to integrate metabolic information from both tumor tissue and biofluid samples toward the reliable discovery of metabolite biomarkers. Here, we aim to analyze the global metabolic dysregulation of CRC in both tumor tissue and plasma samples and simultaneously develop a correlative analysis strategy to discover tumor tissue derived (TTD) metabolites in biofluids (e.g., plasma).

Colorectal cancer is one of the most common cancers worldwide with a high fatality rate. There were an estimated 1.4 million new cases and 700,000 deaths in 2012.¹³ Early screening can discover CRC in early stages, and subsequent surgical treatment provides a promising prognosis. For patients diagnosed at a localized stage, surgical removal of the tumor and nearby lymph nodes significantly improves the 5-year survival rate to as high as 90%. In contrast, the 5-year survival rate is less than 70% when patients are diagnosed at later stages (e.g., with regional or distant metastasis).¹⁴ Unfortunately, traditional methods for clinical CRC screening, including enteroscopy tests and fecal examinations, have significant limitations.^{15–17} In addition, molecular biomarkers for CRC including carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) are not sensitive enough to detect early stage CRC patients.¹⁸ Therefore, there is an urgent need to develop an alternative method with high accuracy, minimal invasion, and a low cost for CRC screening and early diagnosis. Blood metabolite biomarkers are considered to be an alternative choice.¹⁹

In this study, a set of paired colorectal cancer and adjacent normal tissues were excised from the CRC patients during surgery ($n = 68$), and paired preoperative and postoperative plasma samples ($n = 68$) from the same patients were also collected on the day before surgery and the seventh day after surgery, respectively. Next, untargeted metabolomics analyses were performed on both tissue and plasma samples. Particularly, a correlation analysis strategy was developed to discover the TTD metabolites in the plasma of CRC patients. To demonstrate the potential clinical application, plasma samples from 73 CRC patients and 73 gender-matched polyp controls were used as the training set and external validation set, respectively, to discover TTD metabolite biomarkers for CRC. Finally, eight potential metabolite biomarkers in plasma were discovered and demonstrated to have high sensitivity and specificity for CRC diagnosis. More importantly, metabolite biomarkers also performed well in predicting the prognosis of CRC patients after surgery, including the overall survival (OS) and progression-free survival (PFS). In summary, we developed a correlation analysis strategy to discover the

tumor tissue derived metabolite biomarkers in plasma which hold great clinical potential for CRC diagnosis and prognosis.

■ EXPERIMENTAL SECTION

Participant Eligibility Criteria. Between April 2015 and June 2016, all CRC patients and polyp controls were recruited from the Colorectal Surgery Department of Harbin Medical University Cancer Hospital (Harbin, China) with written informed consent. The eligibility criteria were set as follows: (1) participants were not receiving any medical treatment; (2) everyone was not diagnosed with any metabolic diseases, such as kidney diseases, liver diseases, or other cancers; (3) CRC patients had undergone colorectal surgical treatment, and their diagnosis had been confirmed and staged by histopathologic examinations. All of the patients were staged according to the Union for International Cancer Control (UICC) pathologic tumor-node-metastasis (TNM) classification system (eighth edition, 2016). This study was approved by the Ethics Committee of the Harbin Medical University Cancer Hospital.

Experimental details about sample collection, metabolite extraction, and liquid chromatography–mass spectrometry (LC-MS) data acquisition are provided in the [Supporting Information](#).

Data Processing. First, ProteoWizard (version 3.06150) was used to convert raw MS data (.d) files to the mzXML format, and R package “XCMS” (version 3.2) was used for data processing. The generated data matrix consisted of the mass-to-charge ratio (m/z) value, retention time (RT), and peak abundance. R package “CAMERA” was used for peak annotation after XCMS data processing. Metabolic peaks detected less than 50% in all the quality control (QC) samples were excluded. Subsequently, the R package “MetNormalizer” was used for the normalization of each metabolic peak in subject samples to remove unwanted system error that occurred among intra- and interbatches.²⁰ Minimum value (half of the least nonzero value) or a random forest (RF) regression model (R packages “missForest”) was used for missing data imputation before differential analysis and correlation analysis, respectively. The combination of accurate mass and experimental MS/MS match against our in-house tandem MS spectral library and other databases (NIST, METLIN, and MassBank) is used for metabolite identification. The final eight compounds (chenodeoxycholic acid, creatinine, dihydrothymine, histidine–glycine, L-gulonic γ -lactone, L-tryptophan, L-tyrosine, and xanthine) were confirmed using the purchased chemical standards.

Statistical Analyses. All statistical analyses were performed using R (version 3.3.2). An unsupervised principal component analysis (PCA) was applied to visualize the global metabolic profiles among groups (using R function “prcomp”). Next, a supervised model of orthogonal partial least-squares discriminant analysis (OPLS-DA) was applied to describe the global metabolic changes between cancer and control groups. To avoid the overfitting, a permutation test was performed 200 times to assess the validity of the discriminant models. The variable important in the projection (VIP) value was calculated for each variable in the OPLS-DA model. Additionally, a univariate analysis of paired t test or Wilcoxon Mann–Whitney U test with Benjamini–Hochberg-based false discovery rate (FDR) adjusted was performed using the R package “fdrtool” for differential analysis between self-paired and nonpaired samples, respectively. Dysregulated metabolites with a FDR adjusted p value less than 0.05 and VIP value larger than 1

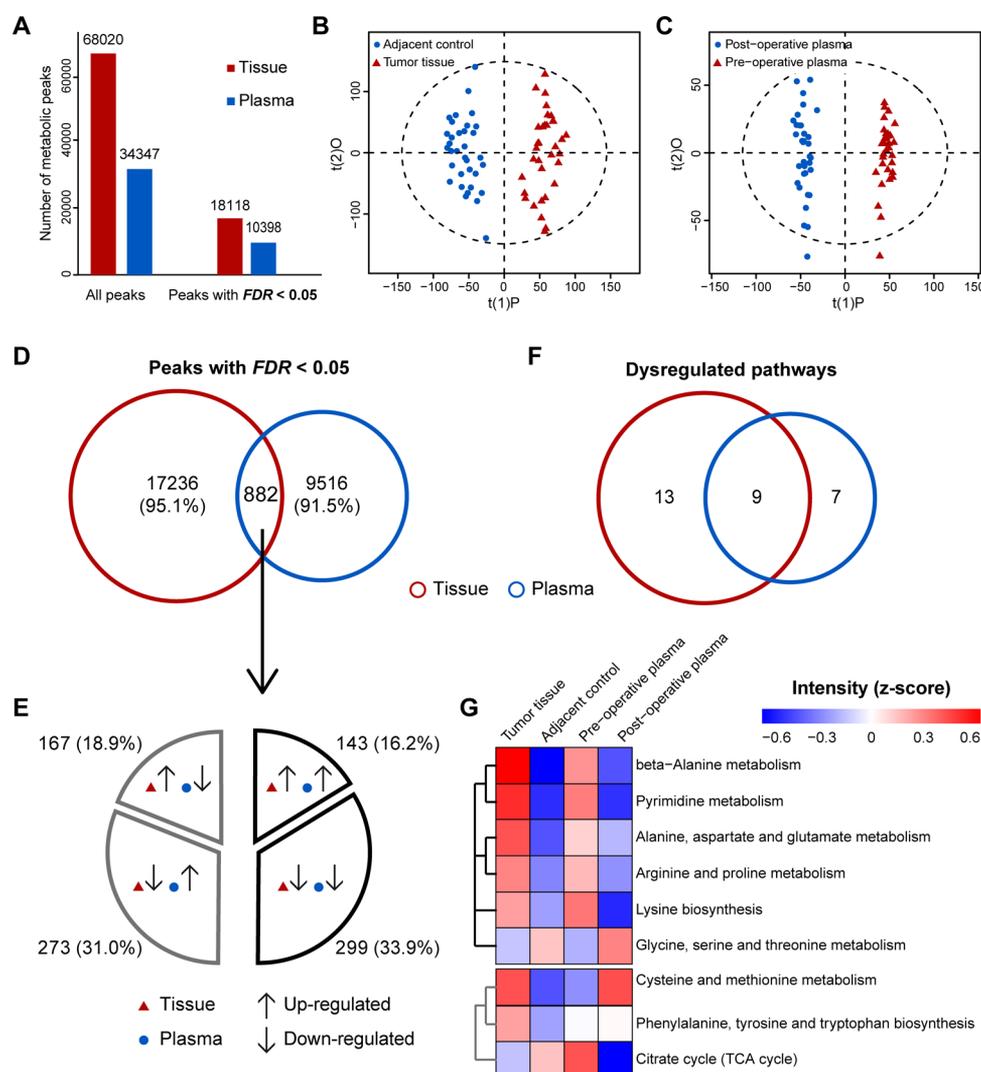


Figure 1. Metabolic dysregulation in CRC tissue and plasma samples. (A) Bar graphs demonstrating the global coverage of CRC tissue and plasma metabolic profiles. (B) Supervised OPLS-DA plot for metabolic profiles of CRC tissues and controls. (C) Supervised OPLS-DA plot for metabolic profiles of preoperative and postoperative plasma samples. (D) Venn diagram showing the shared and unique dysregulated metabolic peaks between CRC tissue and plasma samples. (E) Pie chart for shared dysregulated metabolic peaks is divided into four sections with different dysregulation trends. (F) Venn diagram showing shared and unique dysregulated pathways between CRC tissue and plasma samples. (G) Heatmap for shared dysregulated pathways between CRC tissue and plasma samples. Red represents up-regulated, and blue represents down-regulated.

were selected and mapped into the KEGG database (<http://www.genome.jp/kegg/>) for pathway enrichment analysis using the hypergeometric test. The median intensity of dysregulated metabolites from each pathway was used to represent the pathway expression level.

Spearman rank correlation analysis was used to select the correlative metabolite pairs between CRC plasma and tissue, and a p value less than 0.05 was set as the significance level. A multivariate correlation-based RF regression model with a permutation test (500 times) was then performed to evaluate the correlation of between dysregulated metabolites in plasma and tissue samples (using the R package “randomForest”). The least absolute shrinkage and selection operator (LASSO) regression with a 10-fold cross-validation was repeated 20 times for the discovery of metabolite biomarkers (using the R package “glmnet”). The network graph between tissue dysregulated metabolites and metabolite biomarkers in CRC plasma was drawn using Cytoscape (version 3.5.0). Finally, receiver operating characteristic (ROC) curves were plotted

using the R package “pROC” to evaluate the diagnostic performance of metabolite biomarkers. The area under the curve (AUC) value and 95% confidence interval (CI) were calculated to assess the accuracy of prediction. Survival curves were used to demonstrate the performance of predicting the prognosis using metabolite biomarkers, and a log-rank test was utilized to analyze the statistical differences between different groups of CRC patients (using the R package “survminer”).

RESULTS AND DISCUSSION

Metabolic Profiles of CRC Tissue and Plasma. LC-MS-based untargeted metabolomics was used to profile a set of paired tissue and plasma samples from 34 CRC patients (tumor tissue vs adjacent control; preoperative plasma vs postoperative plasma) and to investigate the global metabolic dysregulation in colorectal cancer. The clinical information on enrolled patients is listed in the Supporting Information, Table S1. To increase the coverage in untargeted metabolomics, both positive and negative modes of mass spectrometry measure-

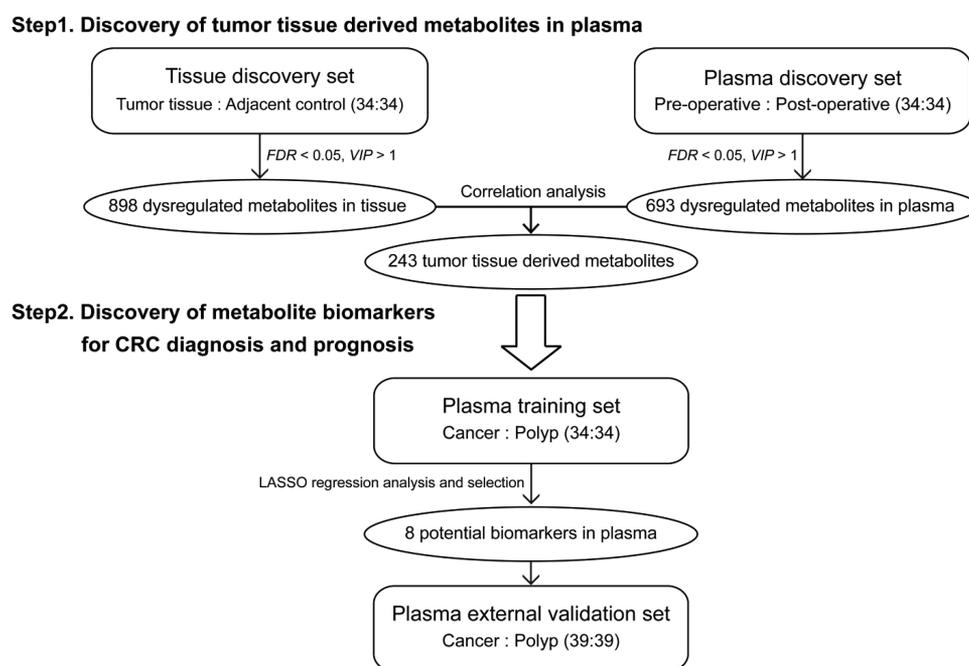


Figure 2. Analytical strategy to discover tumor tissue derived metabolites followed by the discovery of metabolite biomarkers in plasma for CRC diagnosis and prognosis.

ments with two LC columns (amide and T3) were utilized to profile polar and nonpolar metabolites. QC samples in an unsupervised model of PCA plots were clustered tightly, which indicated that the data quality was very good (Supporting Information, Figure S1). A total of 68,020 and 34,347 metabolic peaks were detected in CRC tissue and plasma, respectively. The numbers of dysregulated metabolic peaks (paired *t* test with FDR adjusted $p < 0.05$) were 18,118 and 10,398 in tissue and plasma, respectively (Figure 1A). In a supervised model of OPLS-DA, CRC tumor tissues were clearly separated with healthy controls, indicating different metabolic profiles between them (Figure 1B). Similarly, preoperative plasma samples from patients with CRC tumors were also significantly different from the paired postoperative plasma samples after tumor resection (Figure 1C). Permutation tests demonstrated that the OPLS-DA models were well fitted (Supporting Information, Figure S2). We also analyzed the numbers of metabolic peaks with fold change larger than three standard deviations and defined them as significantly dysregulated peaks toward CRC tumors. There were 1173 significantly dysregulated peaks discovered in tumor tissue samples, which is much larger than those discovered in plasma samples (299, Supporting Information, Figure S3). These results demonstrated that the CRC tumor induced significant shifts on metabolic profiles on both tumor tissue and plasma samples. In agreement, the metabolic changes in tumor tissue were significantly larger than those in plasma samples.

Metabolic Consistency between CRC Tissue and Plasma. We further compared the dysregulated metabolic peaks shared by CRC tissue and plasma and evaluated the consistency of metabolic profiles between two sample types. However, less than 10% of the dysregulated metabolic peaks were shared in both sample types (Figure 1D). More importantly, only half of these shared metabolic peaks displayed the same up- or down-regulation trends in both sample types (Figure 1E). Pathway enrichment analyses were also performed by mapping dysregulated metabolites into the

KEGG database. As a result, 22 and 16 enriched pathways were discovered from tumor tissue and plasma samples, respectively. Nine enriched pathways were shared between two sample types (Figure 1F; Supporting Information, Tables S2 and S3). Among them, five pathways were found to be up-regulated both in CRC tumor tissue and in preoperative plasma samples, whereas the glycine, serine, and threonine metabolism was down-regulated in both cancer sample types (Figure 1G). However, the other three pathways, including the citrate cycle, displayed completely different trends between tumor tissue and plasma samples. At the individual level, the trends for patient-specific changes of metabolic pathways were roughly consistent with the overall results. However, some exceptions existed and suggested that the patient-specific metabolic variation between tissue and plasma is more complex (Supporting Information, Figure S4). These results clearly demonstrated that the metabolic dysregulation from the same patients was sample-type-dependent, and metabolic dysregulation in CRC tumor tissue was significantly different from that in plasma samples. Therefore, it proved that the metabolite biomarkers discovered in tumor tissue may not be consistent with those discovered in plasma samples, and vice versa. This also confirmed that it is difficult to discover TTD metabolite biomarkers using biofluid samples.⁴

Additionally, we found that glycolysis is significantly increased in CRC tumor tissue ($p = 0.031$, Supporting Information, Table S2), due to the Warburg effect.²¹ Nevertheless, there is little change of glycolysis in CRC plasma ($p = 0.426$, Supporting Information, Table S3). A possible explanation is that the dysregulated metabolites in tissue are likely to be diluted or compensated during the circulation. Instead, a significant disturbance of the pyruvate metabolism (which is downstream of glycolysis) was unveiled in CRC plasma ($p = 0.002$, Supporting Information, Table S3). This suggested that TTD metabolites in plasma hold great potential to reflect the pathological status of CRC.

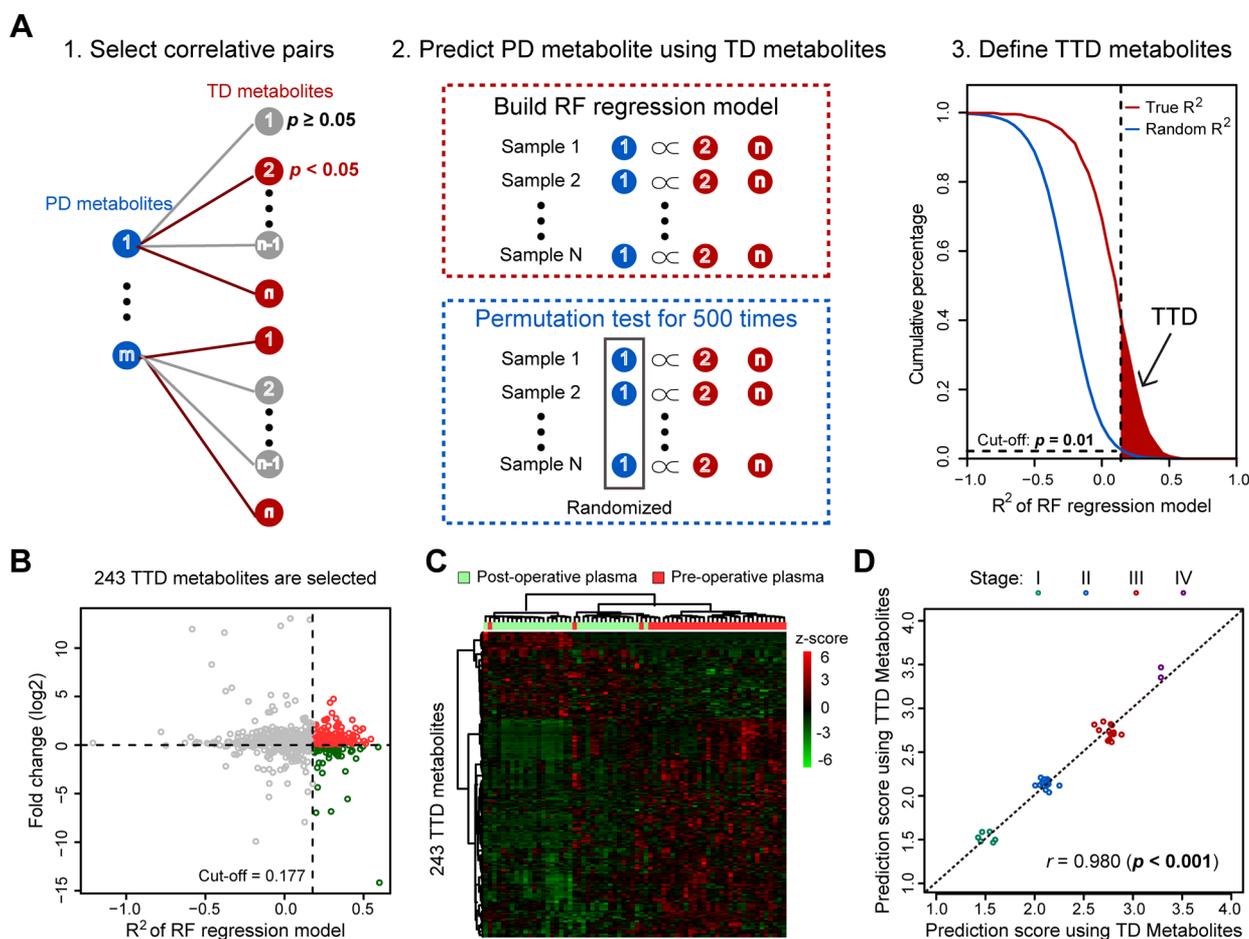


Figure 3. Correlation analysis strategy to discover tumor tissue derived metabolites in CRC plasma. (A) Analysis strategy to discover tumor tissue derived metabolites in plasma. PD, plasma dysregulated; TD, tissue dysregulated; TTD, tumor tissue derived. (B) Distributions of fold changes and R^2 values for TTD metabolites in CRC plasma. The cutoff of R^2 values is set at 0.177. Red, up-regulated; green, down-regulated; gray, unselected. (C) Heatmap for the discrimination between preoperative and postoperative plasma samples using 243 TTD metabolites. (D) Scatter plot displaying the prediction scores using the TD metabolites in tissue and TTD metabolites in plasma to predict the pathological stages of CRC patients.

Discovery of Tumor Tissue Derived Metabolites in Plasma. Here, we developed a new analysis strategy by integrating the univariate and multivariate correlation analysis to discover TTD metabolites in plasma and further to select a panel of metabolite biomarkers for clinical CRC diagnosis and prognosis (study workflow is illustrated in Figure 2, and analysis strategy is shown in Figure 3A). Paired tissue and plasma samples from 34 previous CRC patients were assigned to two data sets: tissue discovery set ($n = 68$) and plasma discovery set ($n = 68$). Dysregulated metabolites in both data sets with FDR adjusted $p < 0.05$ and $VIP > 1$ were selected (898 and 693 in tissue and plasma data sets, respectively). First, the univariate Spearman rank analysis was used to correlate each dysregulated metabolite in plasma with each of those in tissue samples, and correlative metabolite pairs with $p < 0.05$ were reserved. For each dysregulated metabolite in plasma, all of its correlated dysregulated metabolites in tissue were utilized to build a multivariate correlation-based RF regression model and to predict the corresponding dysregulated metabolite in plasma. The value of R^2 is calculated to characterize the fitting of the RF regression model. Importantly, metabolites in tumor tissue and preoperative plasma samples from the same CRC patient were paired to perform the RF regression. In contrast, the same regression

analysis was also performed using the same metabolites in the tissue and plasma samples but from randomly matched patients, which was used as a permutation test to evaluate the haphazard effect. The permutation test was repeated 500 times to generate the random distribution of the R^2 value. The R^2 value with a cumulative percentage less than 1% in the random distribution was determined as the threshold for the true correlation. Finally, 243 out of 693 dysregulated metabolites in CRC plasma were regarded as TTD metabolites ($R^2 > 0.177$). Among them, 173 metabolites were up-regulated and 70 metabolites were down-regulated (Figure 3B and Table S4 in the Supporting Information).

We further performed an unsupervised hierarchical cluster analysis (HCA) using 243 TTD metabolites in plasma to discriminate CRC patients before and after surgery. More than 90% preoperative plasma samples were clustered tightly and clearly separated with postoperative ones (Figure 3C). Furthermore, two independent RF regression models were constructed using either 243 TTD metabolites in plasma or 898 dysregulated metabolites in tumor tissue to predict the stages of the same CRC patients, respectively. Both of the two prediction models can accurately classify CRC patients according to their pathological stages. In particular, the prediction score using a plasma-metabolite-based prediction

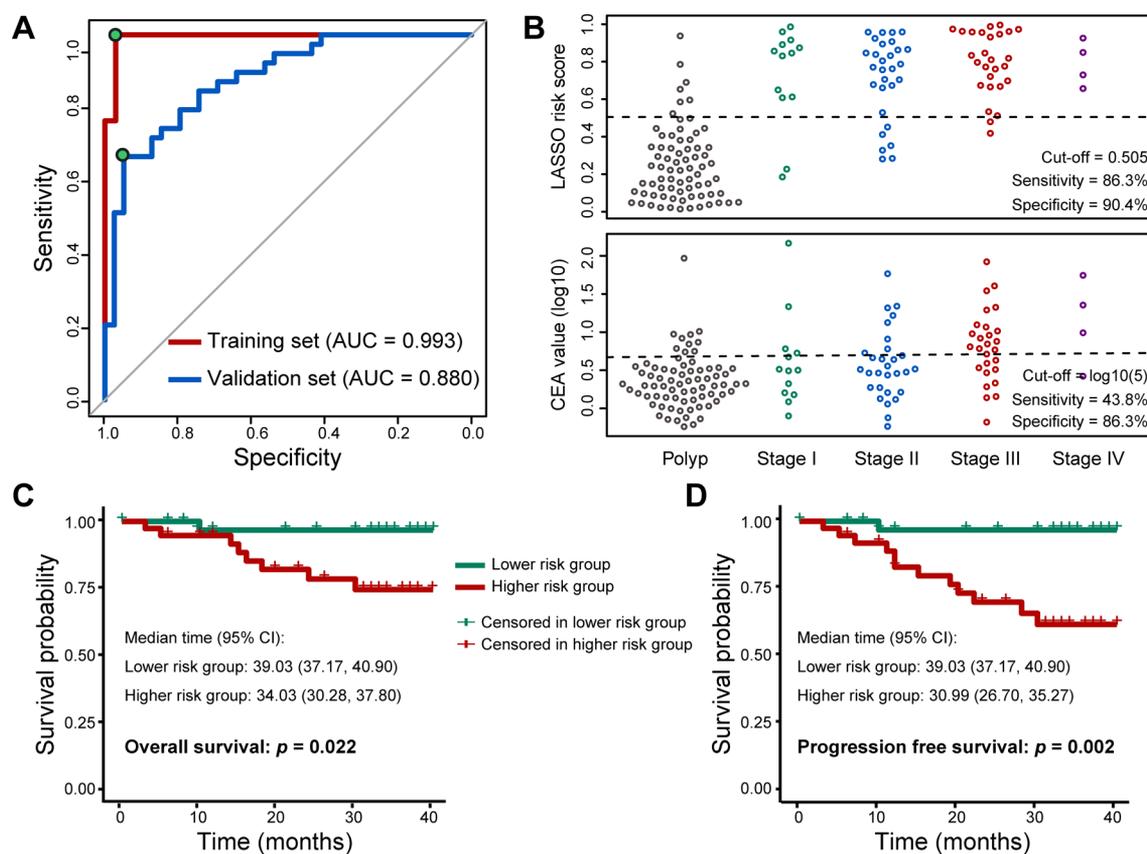


Figure 4. Clinical validation of potential metabolite biomarkers for CRC diagnosis and prognosis. (A) ROC curves of LASSO discriminant model performed in plasma training set and external validation set. (B) Scatter plot of the comparison result between the LASSO discriminant model and CEA in distinguishing CRC patients and polyp controls. (C,D) Kaplan–Meier survival curves for the overall survival and progression-free survival of CRC patients, respectively. Patients with a LASSO risk score greater than 0.798 were defined as the higher risk group, and other patients were defined as the lower risk group. Log-rank test was used to evaluate survival differences.

model displayed an excellent linear relationship with that using the tissue-metabolite-based model ($r = 0.980$, $p < 0.001$, Figure 3D). Collectively, these results confirmed that TTD metabolites in plasma accurately reflect the pathological status and tumor stages of CRC patients and have a high diagnostic potential for clinical applications.

Discovery of Metabolite Biomarkers for CRC Diagnosis. Most types of CRC are developed via the “adenoma–carcinoma sequence”.²² Therefore, detecting the transformation from polyps to carcinoma is critical for CRC screening and prevention. Accordingly, 146 recruited CRC patients and gender-matched polyp controls were enrolled to discover the potential metabolite biomarkers from the panel of TTD metabolites for CRC diagnosis (Figure 2). Plasma samples from 34 CRC patients and 34 gender-matched polyp controls were defined as the plasma training set. The rest of the plasma samples were assigned to plasma external validation set ($n = 78$). Clinical information for all participants is summarized in Table S1 in the Supporting Information.

In order to select a panel of metabolite biomarkers from the 243 TTD metabolites, we performed the LASSO regression analysis using the plasma training set. A 10-fold cross-validation approach was used to estimate the optimal parameter (i.e., lambda) of the model and to select the optimal combination of variables (Supporting Information, Figure S5A). In order to ensure the robustness and reproducibility, this analysis process was repeated 20 times. Consequently, eight metabolites (i.e., chenodeoxycholic acid

(CDCA), creatinine, dihydrothymine, histidine–glycine (His-Gly), L-gulonic γ -lactone, L-tryptophan, L-tyrosine, and xanthine) that recurred in all 20 models were selected as potential metabolite biomarkers for CRC diagnosis with strong statistical powers, and all of the identifications were confirmed by the commercial chemical standards (Supporting Information, Figure S6 and Table S5). These metabolites were generally involved in metabolism such as bile acids, amino acids, pyrimidine, and purine, which were reported to play important roles during carcinoma progression.^{23–28} More explanations are provided in the Supporting Information.

All of the potential metabolite biomarkers were significantly dysregulated in plasma samples from CRC patients (Supporting Information, Figure S5B and Table S6). Among them, dihydrothymine and L-gulonic γ -lactone were up-regulated in CRC patients, whereas CDCA, creatinine, His-Gly, L-tryptophan, L-tyrosine, and xanthine were down-regulated. Particularly, five out of them were previously reported being closely associated with CRC in biofluid samples. In the current study, we found that xanthine was down-regulated in CRC plasma, and Long et al. reported a similar result in CRC serum.²⁹ The primary bile acid CDCA was found to be down-regulated in CRC plasma compared with polyps, which is consistent with previous serum metabolomics studies.^{30,31} Down-regulated creatinine in CRC plasma was revealed in this study, which is broadly consistent with various metabolomics studies using CRC urine, serum, and tissue samples.^{8,32,33} Moreover, our finding supports that L-tryptophan and L-

tyrosine were down-regulated in CRC plasma samples, which are in agreement with the previous discovery using CRC urine and serum samples.^{8,9} These results confirmed that our discovery is excellent in reliability.

In the previous correlative analysis, 48 metabolites in tumor tissue were significantly correlated to the eight TTD metabolite biomarkers in plasma and enriched into 12 metabolic pathways. Among them, arginine and proline metabolism, cysteine and methionine metabolism, purine

metabolism, pyrimidine metabolism, and β -alanine metabolism were significantly dysregulated and enriched in CRC tissue with $p < 0.005$ and considered as potential therapeutic targets (Supporting Information, Figure S5C and Table S2).

Clinical Performance for CRC Diagnosis and Prognosis. In order to evaluate the clinical performance of metabolite biomarkers, a LASSO discriminant model (eq 1) was constructed in the plasma training set and applied to the plasma external validation set:

$$\begin{aligned} \text{prediction score} = & 3.040 - 7.74 \times 10^{-7} M_{\text{CDCA}} - 4.81 \times 10^{-5} M_{\text{creatinine}} + 3.91 \times 10^{-5} M_{\text{dihydrothymine}} \\ & + 7.41 \times 10^{-4} M_{\text{gulonic } \gamma\text{-lactone}} - 3.90 \times 10^{-5} M_{\text{His-Gly}} - 7.83 \times 10^{-6} M_{\text{tryptophan}} \\ & - 1.64 \times 10^{-7} M_{\text{tyrosine}} - 5.55 \times 10^{-6} M_{\text{xanthine}} \end{aligned} \quad (1)$$

ROC curves in Figure 4A demonstrated the excellent diagnostic performances for metabolite biomarkers in both data sets (plasma training set, AUC = 0.993, 95% CI: 0.979–1.000; plasma external validation set, AUC = 0.880, 95% CI: 0.807–0.978). The cutoff value for the prediction score (or called “LASSO risk score”) was set as 0.505. The diagnosis model using the metabolite biomarkers performed much better than using the conventional CEA (with a cutoff value of 5 ng/mL), especially in the detection of early stage CRC patients. The diagnostic sensitivities were 86.3 and 43.8% for metabolite biomarkers and CEA, respectively, and the specificities were 90.4 and 86.3% for metabolite biomarkers and CEA, respectively (Figure 4B). However, we also evaluated that the combination of the metabolite biomarkers and CEA did not improve the diagnostic performance (Supporting Information, Figure S7). These results indicate that the LASSO discriminant model is promising for the clinical diagnosis of CRC patients against polyp controls.

Finally, we divided all of the 73 CRC patients into two groups using the median value of LASSO risk score (0.798). Accordingly, 36 CRC patients were defined as the lower risk group, and 37 others were assigned to the higher risk group. To date, seven patients in the lower risk group and five in the higher risk group were lost to follow-up, respectively. A Kaplan–Meier survival test was then used to predict the CRC progression using the OS time and PFS time (Figure 4C,D). The median OS time was 39.03 months (95% CI: 37.17–40.90) of the lower risk group, significantly longer than 34.03 months of the higher risk group (95% CI: 30.28–37.80, $p = 0.022$). Meanwhile, the median PFS time was 39.03 months (95% CI: 37.17–40.90) of the lower risk group, which was also longer than the 30.99 months of the higher risk group (95% CI: 26.70–35.27, $p = 0.002$). These results confirmed that the metabolite biomarkers are closely associated with the CRC prognosis.

CONCLUSIONS

In summary, we developed an effective analysis strategy for the discovery of tumor tissue derived metabolites in biofluids. Using this strategy, 243 tumor tissue derived metabolites were successfully revealed in plasma samples. This is the first time the consistency of metabolic dysregulation between CRC tumor tissue and biofluids samples was evaluated. As expected, metabolic changes in CRC tumor tissue were larger than those in plasma samples, and the shared parts between them were very limited. Most importantly, we have analyzed and

demonstrated the complex correlation between tumor tissue and plasma samples from patients and finally discovered eight potential metabolite biomarkers in plasma for CRC diagnosis and prognosis. These metabolite biomarkers were promising for further clinical applications. However, an MRM-based targeted metabolomics study focused on these potential metabolite biomarkers should be carried out with large-scale and multicenter cohorts for rigorous external validations and to determine the optimal cutoff value of diagnosis in clinical applications.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.8b05177.

Figures S1–S7, Tables S1–S6, supplemental experimental section and supplemental discussion section (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: likang@ems.hrbmu.edu.cn. Phone: 86-451-87502939.

*E-mail: leospiv@163.com. Phone: 86-451-86298375.

*E-mail: jiangzhu@sioc.ac.cn. Phone: 86-21-68582296.

ORCID

Zheng-Jiang Zhu: 0000-0002-3272-3567

Author Contributions

Δ Z.W. and B.C contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was funded by National Natural Science Foundation of China (Grant Nos. 21575151, 81773551, and 81702766) and Nn10 program of Harbin Medical University Cancer Hospital.

REFERENCES

- (1) Spratlin, J. L.; Serkova, N. J.; Eckhardt, S. G. *Clin. Cancer Res.* **2009**, *15*, 431–440.
- (2) Ni, Y.; Xie, G.; Jia, W. *J. Proteome Res.* **2014**, *13*, 3857–3870.
- (3) Li, Y.; Li, M.; Jia, W.; Ni, Y.; Chen, T. *Anal. Bioanal. Chem.* **2018**, *410*, 2689–2699.

- (4) Cross, A. J.; Moore, S. C.; Boca, S.; Huang, W. Y.; Xiong, X.; Stolzenberg-Solomon, R.; Sinha, R.; Sampson, J. N. *Cancer* **2014**, *120*, 3049–3057.
- (5) Moore, S. C.; Matthews, C. E.; Sampson, J. N.; Stolzenberg-Solomon, R. Z.; Zheng, W.; Cai, Q.; Tan, Y. T.; Chow, W. H.; Ji, B. T.; Liu, D. K.; Xiao, Q.; Boca, S. M.; Leitzmann, M. F.; Yang, G.; Xiang, Y. B.; Sinha, R.; Shu, X. O.; Cross, A. J. *Metabolomics* **2014**, *10*, 259–269.
- (6) Slupsky, C. M.; Rankin, K. N.; Wagner, J.; Fu, H.; Chang, D.; Weljie, A. M.; Saude, E. J.; Lix, B.; Adamko, D. J.; Shah, S.; Greiner, R.; Sykes, B. D.; Marrie, T. J. *Anal. Chem.* **2007**, *79*, 6995–7004.
- (7) Zhang, F.; Zhang, Y.; Zhao, W.; Deng, K.; Wang, Z.; Yang, C.; Ma, L.; Openkova, M. S.; Hou, Y.; Li, K. *Oncotarget* **2017**, *8*, 35460–35472.
- (8) Cheng, Y.; Xie, G.; Chen, T.; Qiu, Y.; Zou, X.; Zheng, M.; Tan, B.; Feng, B.; Dong, T.; He, P.; Zhao, L.; Zhao, A.; Xu, L. X.; Zhang, Y.; Jia, W. *J. Proteome Res.* **2012**, *11*, 1354–1363.
- (9) Leichtle, A. B.; Nuoffer, J. M.; Ceglarek, U.; Kase, J.; Conrad, T.; Witzigmann, H.; Thiery, J.; Fiedler, G. M. *Metabolomics* **2012**, *8*, 643–653.
- (10) Qiu, Y.; Cai, G.; Zhou, B.; Li, D.; Zhao, A.; Xie, G.; Li, H.; Cai, S.; Xie, D.; Huang, C.; Ge, W.; Zhou, Z.; Xu, L. X.; Jia, W.; Zheng, S.; Yen, Y.; Jia, W. *Clin. Cancer Res.* **2014**, *20*, 2136–2146.
- (11) Sreekumar, A.; Poisson, L. M.; Rajendiran, T. M.; Khan, A. P.; Cao, Q.; Yu, J.; Laxman, B.; Mehra, R.; Lonigro, R. J.; Li, Y.; Nyati, M. K.; Ahsan, A.; Kalyana-Sundaram, S.; Han, B.; Cao, X.; Byun, J.; Omenn, G. S.; Ghosh, D.; Pennathur, S.; Alexander, D. C.; Berger, A.; Shuster, J. R.; Wei, J. T.; Varambally, S.; Beecher, C.; Chinnaiyan, A. M. *Nature* **2009**, *457*, 910–914.
- (12) Hori, S.; Nishiumi, S.; Kobayashi, K.; Shinohara, M.; Hatakeyama, Y.; Kotani, Y.; Hatano, N.; Maniwa, Y.; Nishio, W.; Bamba, T.; Fukusaki, E.; Azuma, T.; Takenawa, T.; Nishimura, Y.; Yoshida, M. *Lung Cancer* **2011**, *74*, 284–292.
- (13) Torre, L. A.; Bray, F.; Siegel, R. L.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A. *Ca-Cancer J. Clin.* **2015**, *65*, 87–108.
- (14) Miller, K. D.; Siegel, R. L.; Lin, C. C.; Mariotto, A. B.; Kramer, J. L.; Rowland, J. H.; Stein, K. D.; Alteri, R.; Jemal, A. *Ca-Cancer J. Clin.* **2016**, *66*, 271–289.
- (15) Vernon, S. W. *J. Natl. Cancer Inst.* **1997**, *89*, 1406–1422.
- (16) Walsh, J. M.; Terdiman, J. P. *JAMA, J. Am. Med. Assoc.* **2003**, *289*, 1288–1296.
- (17) Sali, L.; Mascalchi, M.; Falchini, M.; Ventura, L.; Carozzi, F.; Castiglione, G.; Delsanto, S.; Mallardi, B.; Mantellini, P.; Milani, S.; Zappa, M.; Grazzini, G. *J. Natl. Cancer Inst.* **2016**, *108*, djv319.
- (18) Locker, G. Y.; Hamilton, S.; Harris, J.; Jessup, J. M.; Kemeny, N.; Macdonald, J. S.; Somerfield, M. R.; Hayes, D. F.; Bast, R. C. *J. Clin. Oncol.* **2006**, *24*, 5313–5327.
- (19) Farshidfar, F.; Weljie, A.; Kopciuk, K.; Hilsden, R.; McGregor, S.; Buie, W.; MacLean, A.; Vogel, H.; Bathe, O. *Br. J. Cancer* **2016**, *115*, 848–857.
- (20) Shen, X.; Gong, X.; Cai, Y.; Guo, Y.; Tu, J.; Li, H.; Zhang, T.; Wang, J.; Xue, F.; Zhu, Z.-J. *Metabolomics* **2016**, *12*, 89.
- (21) Vander Heiden, M. G.; Cantley, L. C.; Thompson, C. B. *Science* **2009**, *324*, 1029–1033.
- (22) Erichsen, R.; Baron, J. A.; Hamilton-Dutoit, S. J.; Snover, D. C.; Torlakovic, E. E.; Pedersen, L.; Froslev, T.; Vyberg, M.; Hamilton, S. R.; Sørensen, H. T. *Gastroenterology* **2016**, *150*, 895–902e895.
- (23) Matsuzaki, J.; Suzuki, H.; Tsugawa, H.; Watanabe, M.; Hossain, S.; Arai, E.; Saito, Y.; Sekine, S.; Akaike, T.; Kanai, Y.; Mukai-sho, K.; Auwerx, J.; Hibi, T. *Gastroenterology* **2013**, *145*, 1300–1311.
- (24) Ma, C.; Han, M.; Heinrich, B.; Fu, Q.; Zhang, Q.; Sandhu, M.; Agdashian, D.; Terabe, M.; Berzofsky, J. A.; Fako, V.; Ritz, T.; Longrich, T.; Theriot, C. M.; McCulloch, J. A.; Roy, S.; Yuan, W.; Thovarai, V.; Sen, S. K.; Ruchirawat, M.; Korangy, F.; Wang, X. W.; Trinchieri, G.; Greten, T. F. *Science* **2018**, *360*, No. eaan5931.
- (25) Payne, C. M.; Bernstein, C.; Dvorak, K.; Bernstein, H. *Clin. Exp. Gastroenterol.* **2008**, *1*, 19–47.
- (26) Simińska, E.; Koba, M. *Amino Acids* **2016**, *48*, 1339–1345.
- (27) Heidelberger, C.; Chaudhuri, N. K.; Danneberg, P.; Mooren, D.; Griesbach, L.; Duschinsky, R.; Schnitzer, R. J.; Plevin, E.; Scheiner, J. *Nature* **1957**, *179*, 663–666.
- (28) Pedley, A. M.; Benkovic, S. J. *Trends Biochem. Sci.* **2017**, *42*, 141–154.
- (29) Long, Y.; Sanchez-Espirdion, B.; Lin, M.; White, L.; Mishra, L.; Raju, G. S.; Kopetz, S.; Eng, C.; Hildebrandt, M. A. T.; Chang, D. W.; Ye, Y.; Liang, D.; Wu, X. *Cancer* **2017**, *123*, 4066–4074.
- (30) Costarelli, V.; Key, T. J.; Appleby, P. N.; Allen, D. S.; Fentiman, I. S.; Sanders, T. A. *Br. J. Cancer* **2002**, *86*, 1741–1744.
- (31) Tan, B.; Qiu, Y.; Zou, X.; Chen, T.; Xie, G.; Cheng, Y.; Dong, T.; Zhao, L.; Feng, B.; Hu, X.; Xu, L. X.; Zhao, A.; Zhang, M.; Cai, G.; Cai, S.; Zhou, Z.; Zheng, M.; Zhang, Y.; Jia, W. *J. Proteome Res.* **2013**, *12*, 3000–3009.
- (32) Zhu, J.; Djukovic, D.; Deng, L.; Gu, H.; Himmati, F.; Chiorean, E. G.; Raftery, D. *J. Proteome Res.* **2014**, *13*, 4120–4130.
- (33) Phua, L. C.; Chue, X. P.; Koh, P. K.; Cheah, P. Y.; Ho, H. K.; Chan, E. C. *Cancer Biol. Ther.* **2014**, *15*, 389–397.