Molecular characterization of advanced colorectal cancer using serum proteomics and metabolomics

Jun Rao[1,a] · Xianghui Wan[1,a] · Fangfang Tou[1] · Qinsi He[1] · Aihua Xiong[1] · Xinyi Chen[2] · Li Hou[2] · Zhi Zheng[1,*]

[1]Jiangxi Cancer Hospital of Nanchang University, Jiangxi Provincial Key Laboratory of Translational Medicine and Oncology, Jiangxi Provincial Key Laboratory of digestive system malignant tumor, Jiangxi Cancer Center, Nanchang, 330029, People's Republic of China

[2]Department of Hematology and Oncology, Beijing University of Chinese Medicine, Beijing, 100029, People's Republic of China

[a]These authors contributed equally to this work.

*Corresponding Author:

Zhi Zheng: zhengxia_2007@163.com

Abstract

Introduction Colorectal cancer (CRC) is a growing public health concern with high mortality rate. However, there are no valid diagnostic biomarkers and few therapeutic strategies available for CRC, especially advanced CRC, since the pathogenic mechanisms remain poorly understood.

Objective To comprehensively reveal molecular characterization of advanced CRC, we applied integrated proteomic and metabolomic analyses on serum samples from 20 patients with CRC at stage III or IV.

Methods In the present study, we took advantage of nanoscale liquid chromatography and quadrupole time-of-flight tandem mass spectrometry (nanoLC/Q-TOF-MS/MS) and ultraperformance LC/Q-TOF-MS/MS technologies.

Results Overall, 551 proteins and 719 metabolites were identified in those serum samples, respectively. Hierarchical clustering analysis indicated much more remarkable diversity in proteomic profiles than metabolomic profiles. Further functional analysis suggested that ten key pathways associated with cancer cell metabolism were dissected including glycolysis/gluconeogenesis, biosynthesis of amino acids, glutathione metabolism, and arachidonic acid metabolism, based on which protein-protein interaction network analysis was thus constructed with 80 proteins and 21 metabolites. Moreover, the regulatory network in advanced CRC was established according to correlation analysis, indicating conserved roles of metabolome and lipids & lipid like molecules in human serum. Nevertheless, three metabolites and two proteins including hydroquinone, leucenol and sphingomyelin were supposed to be potential biomarkers, which were determined to be positively and significantly correlated with CEA and/or CA 19-9.

Conclusions Altogether, our work not only extended our understanding on the

physiopathology of advanced CRC, but provided potential biomarkers to improve the accuracy of the diagnosis and monitoring of the syndrome.

1 Introduction

Colorectal cancer (CRC) is the third most common malignancy and remains the second leading cause of cancer-related death worldwide (Bray et al., 2018; Cantor et al., 2020). It was estimated that by the year 2030 CRC probably account for 1 in 10 cancer cases and deaths, and its global health burden would unfortunately increase by 60% (Bray et al., 2018). The high mortality rate for CRC patients is mainly due to the delayed diagnosis in its advanced stage, while the metastasis has already occurred. Only 9% patients are practically diagnosed at stages I, with most (91%) diagnosed at stage II, III or IV (Hammond et al., 2016). As well documented, risk factors for the development of CRC are mainly cigarette smoking, physical inactivity, obesity, and high consumption of alcohol or red meat (Hissong et al., 2019). Besides, family history and certain medical conditions including inflammatory bowel disease are also associated with CRC (Xue et al., 2017).

The pathogenic mechanisms of CRC are complex and heterogeneous, while the molecular changes present in the tumor determine both the time to malignant transformation and the histologic type of premalignant lesion (Hissong et al., 2019). Chromosomal instability, usually secondary to inactivation of the adenomatous polyposis coli (APC) tumor suppressor gene, is the first characterized and most common molecular pathway in CRC, which subsequently results in the hyper-activation of the WNT signaling pathway (Fearon et al., 1990). The second molecular pathway, microsatellite instability, occurs in 15% of CRC (Hissong et al., 2019). There is an abnormality in MLH-1, PMS-2, MSH-2, MSH-6, or POL-E gene, which are all necessary for DNA mismatch repair. Furthermore in the past few years, the genomic, and transcriptomic landscapes for CRC have also identified many genomic alterations and extensive molecular heterogeneity of the disease (Vasaikar et

al., 2019). For example, genome-scale analysis of 276 samples was collected to characterize somatic alterations in CRC (Vasaikar et al., 2019). The results showed that twenty-four genes including APC, TP53, ARID1A, and SOX9, are significantly mutated, which suggests a number of therapeutic strategies to CRC.

Recent advances in proteomics and metabolomics have extended our understanding of pathways that control cell proliferation, differentiation, and death (Chen et al., 2019). Identification of proteins or metabolites playing critical roles in the development of cancer can potentially discover biomarkers for the diagnosis of CRCs (Ritchie et al., 2010). Moreover, signaling proteins, metabolites and the corresponding pathways are usually attractive therapeutic targets for cancer treatment. The high-throughput approach, proteomics, allows for the simultaneous detection of thousands of peptides and/or proteins in specific cells, tissues, or body fluids, exhibiting great potentials for its application in large-scale clinical studies on different types of diseases including CRC. For example, Ward et al. (2016) employed surface-enhanced laser desorbtion/ionisation to characterize the serum proteome of 62 CRC patients and 31 normal subjects, which identified complement C3a des-arg, α1-antitrypsin and transferrin with diagnostic potential. Vasaikar et al. (2019) firstly conducted the proteogenomic study on prospectively collected tumor and their normal adjacent tissues. An association between increased glycolysis in microsatellite instability-high (MSI-H) tumors and decreased CD8 T cell infiltration was identified, which proposed glycolysis as a potential target to reverse the resistance of MSI-H tumors to immune check-point blockade. Apart from proteomics, the other unbiased and powerful analysis method is metabolomics, which enables comprehensively and semi-quantitatively to determine global metabolites in biological samples. It holds great potential on tumor diagnosis and therapy for various cancers such as CRC,

gastric cancer, liver cancer, and pancreatic cancer (Fan et al., 2018; Zheng et al., 2017; Zheng et al., 2018). For instance, Kim et al. (2019) performed urine-NMR metabolomics on 92 patients with colorectal neoplasia and 156 healthy individuals for screening of advanced adenoma and stage 0 CRC. The results revealed that 3-aminoisobutyrate, taurine, and alanine were good indicators for CRC patients according to receiver operating characteristics curve analysis.

Generally, different strategies such as surgery, radiation therapy, chemotherapy, targeted drug therapy, and immunotherapy have been adopted for CRC treatment (Khiavi et al., 2019). For advanced-stage CRC, chemotherapy is commonly recommended, while the targeted therapies including anti-epidermal growth factor receptor (anti-EGFR) agents are frequently used in combination with chemotherapy (Rawla et al., 2019; Wang et al., 2019). The best treatment of invasive CRC finally depends on its location, stage, and underlying molecular changes including genetic alterations. However, to the best of our knowledge, the comprehensive molecular characterization of advanced CRC, particularly at proteomic and/or metabolomic level, has not previously been elucidated. Hence, we here took advantage of the nontargeted technology to conduct integrated proteomic (nanoLC/Q-TOF-MS/MS) and metabolomic (UPLC/Q-TOF-MS/MS) analyses on serum samples obtained from 20 patients with advanced CRC. The present study aimed to identify the key regulatory elements (proteins or metabolites) and functional pathways in advanced CRC, which may provide potential biomarkers and the most suitable or novel therapeutic strategy for this syndrome.

2 Materials and methods

2.1 Subjects

In total, 20 patients were recruited, all diagnosed with advanced CRC at stage III or IV. The clinical characteristics of all the 20 patients were shown in Supplementary Table 1, who had been diagnosed without any disease of metabolic, liver kidney, and any other cancers. Meanwhile, the average age of the 20 patients with advanced CRC was about 51, ranging from 29 to 76. All participants had not received any drug treatment in the previous three months before sampling.

2.2 Proteomic analysis

As our previous study, we here employed the combination of a DDA-based ion library and DIA to perform high-throughput proteomics analysis (Chen et al., 2019). 2 μl serum samples were firstly diluted with lysis buffer containing 100 mM Tris-HCL (pH 8.5, Sigma, MO, USA), 8M Urea (Sigma, MO, USA), 1 mM EDTA, 1 mM PMSF, and then centrifuged to remove the sediment at 15000g for 15min at 4 °C. The quantification of protein was determined using a BCA protein assay kit (Bi Yuntian, Shanghai, China). Secondly, protein was reduced, alkylated and further digested by FASP (filter aided sample preparation) method with Trypsin (Promega, Madison, WI) (Wiśniewski et al., 2009). The NanoDrop 2000 instrument (Thermo Scientific, USA) was used for measurement of the peptide concentration at an absorbance of A280 nm.

Further, 3 ug peptides with iRT peptides (Biognosys, Schlieren, Switzerland) were loaded onto a C18 trap column in DDA (Data Dependent Acquisition) analysis ,which was performed on an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) equipped with an EASY-nLC 1000 system (Thermo Fisher Scientific) (Chen et al., 2019). Likewise, the data-independent acquisition (DIA)

analysis was performed the same mass spectrometer and LC system as DDA. The full scan in DIA analysis was set at a resolution of 60,000 over m/z range of 350 to 1500; followed by DIA scans with resolution 30,000; Collision Energy: 32%; AGC target: 5 e5 and maximal injection time: 74 ms. 44 variable DIA windows were set for DIA acquisition ranging from 350 to 1500 m/z. In addition, both the protein identification and quantification were finished by Spectronaut pulsar X 12.0 (Biognosys) with default setting. All results were filtered according to a Q value cutoff of 0.01, while P-value estimator was conducted by Kermel Density Estimator. Area was accordingly used for quantification. Each peptide contained at least 3 fragment-ions.

2.3 Metabolomics analysis

Metabolites in the serum samples were extracted with 120 μL of 50% methanol buffer (Chen et al., 2019). For global metabolomics, an ultra-performance liquid chromatography (UPLC) system (SCIEX, Cheshire, UK) coupled to a high-resolution tandem mass spectrometer (Triple TOF 5600 plus; SCIEX) were used in the present study. An ACQUITY UPLC T3 column (100mm*2.1mm, 1.8μm, Waters, UK) was employed for the reversed phase separation. The mobile phase contained solvent A (water, 0.1% formic acid) and solvent B (Acetonitrile, 0.1% formic acid), while the flow rate was 0.4 mL/minute. Gradient elution conditions were set as follows: 0~0.5 minutes, 5% B; 0.5~7 minutes, 5% to 100% B; 7~8 minutes, 100% B; 8~8.1 minutes, 100% to 5% B; 8.1~10 minutes, 5% B. The injection volume for each sample was 4 μL. The Q-TOF was performed in both positive and negative ion modes (Chen et al., 2019). The Ionspray voltage floating in positive and negative ion mode were set at 5000 V and -4500V, respectively. The XCMS software was used here to carry out MS data pretreatments, including peak picking, peak grouping, retention time correction,

second peak grouping, and annotation of isotopes and adducts. Moreover, the online databases including Human Metabolome Database (HMDB) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were employed to annotate the metabolites. And More importantly, we here took advantage of an in-house fragment spectrum library of metabolites to identify the compounds by their MS^2 spectrums.

2.4 Data analysis

Proteomic and metabolomics data were normalized and analyzed, as our previous studies (Chen et al., 2019). Hierarchical cluster analysis was operated according to Mev (MultiExperiment Viewer) 4.8 software. For proteomic data, the categories of identified proteins were determined by using the online PANTHER (protein annotation through evolutionary relationship) classification system (www.pantherdb.org). Moreover, a multi-omics data analysis tool, OmicsBean (http://www.omicsbean.com), was used in the present study for bioinformatics analysis, including Gene Ontology (GO) analysis, KEGG pathway, and protein-protein interaction network analysis. Before correlation analysis, proteins/metabolites whose contents are the same values in more than ten samples were filtered due to their highly suspect correlations of significant. The correlation analysis was finished using Pearson's product-moment correlation by the R statistical software. And also, the corresponding *p*-values were calculated using the cor.test function. Additionally, *p*-Values were accordingly adjusted to controlling the false discovery rate (FDR) (Rao et al., 2014). The graphical presentation of correlation-ships was composed with Cytoscape version 3.4.0.

3 Results

3.1 Proteomic characterization in patients with advanced CRC

To comprehensively reveal the proteomic characterization, we took advantage of the data-independent mode (DIA)-based SWATH-MS for twenty serum samples from the patients with advanced CRC. A total of 551 proteins were identified, the majority of which functioned as defense/immunity protein, protein modifying enzyme, protein-binding activity modulator, and metabolite interconversion enzyme (Supplementary Table 2 and Fig. 1a). Other types of proteins were also detected in sera, such as extracellular matrix protein, signaling molecules, intercellular signal molecule, transmembrane signal receptor, transfer/carrier proteins, and cell adhesion molecule. To have an overview of the serum proteome among tested samples, hierarchical clustering analysis was performed, resulting in a plot of all the 551 proteins versus 20 serum samples (Fig. 1b). Except for 25 proteins at the top of the heat-map that were of low abundance in all the samples, the left 526 identified proteins displayed remarkable diversity in their abundances across the 20 tested samples. One one hand, the enrichment of certain proteins seemed to be sample-specific. For example, 34 proteins at the bottom of the heat-map including collectin-10, lithostathine-1-alpha, and osteopontin were abundant only in sample N5. On the other hand, several proteins were specifically enriched in one or more samples. For instance, the levels of Immunoglobulin kappa variable 1-39 and C-C motif chemokine 18 were higher in samples N15 and N16 than those in other samples.

The identified 551 proteins were further annotated according to GO databases (Supplementary Figs. 1, 2, and 3). Most of the biological process terms were cellular process, response to stimulus, and biological regulation. For cellular component, cellular anatomical entity, intracellular, and protein-containing complex accounted for

the main terms. The major of molecular function terms were binding, catalytic activity, and molecular function regular. To better understand the biological functions and/or interactions of the identified 551 proteins, we also carried out pathway annotation according to KEGG database. The results showed that 251 proteins were mapped to 189 pathways, the top fifteen of which included complement & coagulation cascades, PI3K-Akt signaling pathway, and pathways in cancer (Supplementary Table 2).

3.2 Metabolomic characterization in patients with advanced CRC

Taking advantage of a non-targeted metabolomic technology, UHPLC-Q-TOF-MS/MS, we then performed metabolite profiling for the same 20 serum samples as in proteomic profiling. The results showed that a total of 9193 positive-mode and 7571 negative-mode ion features were detected, respectively. Based on their MS/MS features, 567 positive-mode and 431 negative-mode metabolites were determined (Supplementary Tables 3 and 4). Totally, 719 metabolites were finally identified in sera, which could be classified into 15 categories according to HMDB database (Supplementary Fig. 4). Among these 719 metabolites, nearly 30% compounds were lipids and lipid-like molecules. Likewise, the unsupervised approach, hierarchical cluster analysis was employed to investigate the metabolic variations among patients with advanced CRC. Obviously when compared to proteins profiling, metabolites profiling exhibited less diversity in their abundances across the 20 tested samples for both positive and negative modes. As shown in Figs. 2a and 2b, samples N15, N16, N17, N18, and N19 were clearly clustered together, especially in positive mode. Pathway analysis was further conducted according to KEGG database, showing 135 mapped pathways (Supplementary Tables 5 and 6). The major of the mapped 135 pathways included

metabolic pathways, glycerophospholipid metabolism, and biosynthesis of amino acids.

3.3 Protein-protein interaction network analysis

When in combination of the 189 mapped pathways in proteins profiling and the 135 mapped pathways in metabolites profiling, 69 mapped pathways were determined to be in common, which included 238 proteins and 187 metabolites (Supplementary Table 7). Among those 69 pathways, the ten main pathways associated with cancer cell metabolism were pathways in cancer, glycolysis/gluconeogenesis, carbon metabolism, protein digestion & absorption, biosynthesis of amino acids, glutathione metabolism, vitamin digestion& absorption, central carbon metabolism in cancer, arachidonic acid metabolism, and tyrosine metabolism. Thus, protein-protein interaction network analysis was employed based on the ten pathways to provide further insights into the developmental and physiological processes underlying advanced CRC. All the 10 pathways except protein digestion & absorption were included in the network with 80 proteins and 21 metabolites. As shown in Fig. 3, proteins were mapped to pathway via KEGG database, while metabolites were connected to proteins via GO database. Obviously, biosynthesis of amino acids and carbon metabolism dominated the PPI network, associating with a large number of proteins including PKM, GAPDH, ALDOA, and ALDOB that plays important roles in cellular proliferation. Other crucial pathways such as glutathione metabolism, tyrosine metabolism, and glycolysis/gluconeogenesis were also shown in the network, which generates key products that promote cell survival and growth. Besides, arachidonic acid metabolism in the network, consisted of 9 proteins, was also suggested to play an important role in the development of various cancers (Borin et al.,

2017). The results here dissected key regulatory proteins/metabolites and pathways involved in the development of advanced CRC.

3.4 Correlation analysis between the detected proteins and metabolites

To deeply reveal the regulatory network in advanced CRC, network-based analysis (20) was employed to analyze the correlations among identified proteins and metabolites, as well as two tumor markers, CEA and CA 19-9. We calculated the values of Pearson pair-wise correlation across these 20 serum samples for the set of 1116 elements including 395 proteins and 719 metabolites. The results were visualized as a heat-map showing in Fig. 4, which displayed a total of 622170 correlations, ranging from -0.8978 for PE(18:2(9Z,12Z)/18:0) and PI 38:3 to 0.9991 for cholic acid and 1b,3a,7a-trihydroxy-5b-cholanoic acid. Further screening results revealed that there were 23201 significant correlations with $r^2 \geqslant 0.49$ and FDR $\leqslant$ 0.05. Among them, 22891 were positive correlations while only 310 were negative ones. Especially, there were much more significant correlations between metabolites than those between proteins or associated with tumor markers. Furthermore, lipids and lipid like molecules dominated the significant correlations, accounting for nearly 63% significant correlations. The significant correlations associated with two important metabolites, citric acid and glutamine were specially shown in Fig. 5, and interestingly, the correlation between them is positively and significantly high (Supplementary Table 8). 120 positive and significant correlations were determined to be associated with glutamine, while 63 positive and significant correlations were found to be related to citric acid. Totally, there were 182 positive and significant correlations between citric acid/glutamine and other 126 molecules, most of which were between citric acid/glutamine and lipids & lipid like molecules. Besides, other

classes of metabolites were also identified to be positively and significantly correlated with citric acid or glutamine, including 24 organic acids & derivatives, 14 benzenoids, 17 organoheterocyclic compounds, and 3 proteins. In addition, there were 18 positive and significant correlations between with CEA/CA 19-9 and 10 elements (5 metabolites and 5 proteins, Table 1). Notably, the correlation between CEA and CA 19-9 was very high, and most of those metabolites or proteins were positively and significantly correlated with both CEA and CA 19-9.

4 Discussion

CRC is one of the most common lethal cancers worldwide, highlighting the significance of understanding the pathogenesis given its poor diagnosis and few effective treatment options especially for patients with advanced CRC. We therefore employed the integrated proteomic and metabolomic strategy on serum samples from 20 patients with CRC at stage III or IV, to comprehensively reveal the molecular characterizations of advanced CRC. The ages of all 20 patients were widely distributed, ranging from 29 to 76. The types of advanced CRC also covered a wide range, including sigmoid colon cancer, right colon cancer, rectal cancer, ascending colon cancer, and adenocarcinoma of the junction of rectum & sigmoid. Moreover, both proteomic and metabolomic analyses were conducted on each serum sample from the same patient, which enabled us to explore the molecular alterations associated with advanced CRC in a more accurate and systematic manner.

Human has large amounts of body fluids, including blood (serum), breast milk, tears, urine and malignant pleural efusions, among which blood is supposed to be the most promising sample for discovering biomarkers to predict treatment effect and prognosis of diseases including cancers (Deng et al., 2020). Blood can be noninvasively collected in large quantity through a simple procedure. It is widely used for proteomic and metabolomic investigations since it often is used for reflecting physical or pathological status as a balanced and stable homeostatic system (Zhou et al., 2019). Accordingly in the present study, both nanoLC/Q-TOF-MS/MS and UPLC/Q-TOF-MS/MS technologies were used to investigate protomic and metabolomic characterization of human serum in relation to advanced CRC. On one hand, a total of 551 proteins were identified in the test 20 serum samples, most of which functioned as defense/immunity protein, protein modifying enzyme, and

protein-binding activity modulator. Other types of proteins including metabolite interconversion enzyme, extracellular matrix protein, and signaling molecules were also detected in sera. One the other hand, totally 719 named metabolites were determined in sera, among which 649 metabolites were classified into 14 known categories including lipids & lipid-like molecules, organic acids & derivatives, and organoheterocyclic compounds. These 649 identified metabolites covered most of the central metabolism pathways such as carbohydrate super pathway, amino acid super pathway, lipid super pathway, and nucleotide super pathway. When compared with previous studies, those 551 proteins and 719 metabolites uncovered so far much broader human serum proteome and metabolome, corroborating the power of this non-targeted proteins and metabolites profiling platforms in uncovering molecular characterization of human serum.

Further KEGG pathway analysis of proteomic and metabolomic data revealed 69 identical pathways, among which ten main pathways were associated with cancer cell metabolism including glycolysis/gluconeogenesis, biosynthesis of amino acids, glutathione metabolism, and arachidonic acid metabolism. Protein-protein interaction network was firstly constructed by OmicsBean for advanced CRC with nine cancer-associated pathways, 80 proteins and 21 metabolites. Metabolic reprogramming, as a hallmark of cancer, has become a topic research area in cancer biology over the past decades (Koppenol et al., 2011). The well understood Warburg effect, associated with glycolysis/gluconeogenesis, is characterized by an increase in glucose uptake and lactate production but a decrease in oxidative phosphorylation. As shown in Fig. 3, 13 proteins were related to glycolysis/gluconeogenesis, most of which play important roles in the development of cancers including CRC (Ahmad et al., 2013; Cui et al., 2014; Dai et al., 2018; Dayton et al., 2016; Duell et al., 2012;

Gimm et al., 2001; He et al., 2016; Leithner et al., 2015; Patel et al., 2008; Tsai et al., 2010; Yun et al., 2015). For example, phosphoglycerate kinase 1 (PGK1) is reported to be a promoter of metastasis in CRC while high expression of ALDOA is associated with poor prognosis in human CRC. Moreover, glutathione metabolism plays both beneficial and pathogenic roles in a series of malignancies (Bansal et al., 2018). Here, eight important proteins including G6PD, GPX3, and LAP3 were connected to glutathione metabolism, which function as colon cancer markers and involve in colon cancer cell growth (Pelosof et al., 2017; Yang et al., 2018; Zhang et al., 2017). Apart from glycolysis/gluconeogenesis and glutathione metabolism, arachidonic acid metabolism, consists of nine proteins in the constructed network may also play necessary roles in the advanced CRC. Many studies have demonstrated arachidonic acid metabolism involved in carcinogenesis (Hong et al., 2004). Habermann et al. (2013) revealed that SNPs of those proteins such as PTGS1, ALOX5, ALOX12, and ALOX1 were interacted with fatty acids in CRC (Habermann et al., 2013). In addition, 21 metabolites including citrate, oxaloacetate, arachidonate and nine standard amino acid were participated in those nine cancer-associated pathways via the related proteins and collectively play important roles in the development of CRC and its metastasis.

Likewise, we for the first time constructed the regulatory network to reveal key elements or pathways in advanced CRC according to correlation analysis, which is proved to be very helpful for dissecting putative key regulatory (Rao et al., 2014). The results here showed a large number of significant correlations in patients with advanced CRC, most of which were positive. The highly positive associations between every two metabolites indicated the conserved roles of metabolome in the human serum, which was in line with the observation in hierarchical clustering

analysis. Especially, lipids and lipid like molecules dominated the significant correlations, displaying the essential roles during cancer development (Beloribi-Djefaflia et al., 2016). Moreover, most of the significant correlations associated with two important cancer-related metabolites, citric acid and glutamine, were observed to be connected to lipids and lipid like molecules. Additionally, we found that there were 18 positive and significant correlations in relation to CEA or CA 19-9. CEA and CA19-9 are acknowledged markers for diagnosing early stages of CRC and predicting treatment effect, but not ideally (Hissong et al., 2019). In the present study the correlation between CEA and CA 19-9 was determined to be very high. And more importantly, 5 metabolites and 5 proteins were found to be positively and significantly associated with CEA or CA 19-9, some of which were demonstrated to be involved in modulating cancer cell growth and functioned as a biomarker (Byeon et al., 2018). For example, hydroquinone was determined to increase the risk of skin cancer by removing the top layer of skin, while the biosynthesis of sphingomyelin was reported to modulate cancer cell death and growth (Lewis et al., 2018). Meanwhile, Kulp et al. (1996) found that by chelating Iron mimosine block cell cycle progression in asynchronous human breast cancer cells. In addition, activation peptide of the coagulation factor XIII (AP-F13A1) and plasma kallikrein (fragment) were identified to be a novel biomarker for the screening of CRC and lung cancer, respectively (Chee et al., 2008; Peltier et al., 2018). These three metabolites and two proteins were supposed to be potential biomarkers, together with CEA and CA 19-9, to improve the accuracy of   the diagnosis and monitoring of CRC.

5 Conclusions

Taken together, the present study comprehensively revealed molecular characterization of advanced CRC by serum proteomics and metabolomics. Our results revealed key regulatory elements or pathways involved in the syndrome by protein-protein interaction network analysis and correlation analysis. The study here not only provided new insights into the mechanism of CRC development, but also present potential biomarkers for the diagnosis and monitoring of CRC. Since the number and types of samples were limited, extensive validation of our findings would be further performed in a larger scale study.

Compliance with ethical standards

Conflict of interest The authors declared no conflict of interest.

Ethics approval The study here was approved by the local ethics committee of Jiangxi Cancer Hospital and was performed in accordance with the Declaration of Helsinki, while all participants provided their written informed consent.

Supporting Information

Supplementary Fig. 1 Biological process (BP) terms of the identified 551 proteins in GO enrichment analysis

Supplementary Fig. 2 Cellular component (CC) terms of the identified 551 proteins in GO enrichment analysis

Supplementary Fig. 3 Molecular function (MF) terms of the identified 551 proteins in GO enrichment analysis

Supplementary Fig. 4 Category information of the 719 metabolites identified in serum samples according to the database from HMDB

Supplementary Table 1 The characteristics of 20 patients in the present study

Supplementary Table 2 KEGG pathway analysis of the 551 proteins detected in serum samples from 20 patients with advanced CRC

Supplementary Table 3 The list of 567 positive-mode metabolites identified in serum samples

Supplementary Table 4 The list of 431 negative-mode metabolites identified in serum samples

Supplementary Table 5 KEGG pathway analysis of 567 positive-mode metabolites identified in serum samples

Supplementary Table 6 KEGG pathway analysis of 431 negative-mode metabolites identified in serum samples

Supplementary Table 7 69 identical pathways in both proteomic and metabolomic data with 238 proteins and 187 metabolites

Supplementary Table 8 The list of significant correlations associated with glutamine or citric acid

References

Ahmad, S. S., Glatzle, J., Bajaeifer, K., Buehler, S., Lehmann, T., Koenigsrainer, I., et al. (2013). Phosphoglycerate kinase 1 as a promoter of metastasis in colon cancer. *International Journal of Oncology,* 43(2), 586-590.

Bansal, A., & Simon, M. C. (2018). Glutathione metabolism in cancer progression and treatment resistance. *Journal of Cell Biology,* 217(7), 2291-2298.

Beloribi-Djefaflia, S., Vasseur, S., & Guillaumond, F. (2016). Lipid metabolic reprogramming in cancer cells. *Oncogenesis,* 5(1), e189-e189.

Borin, T. F., Angara, K., Rashid, M. H., Achyut, B. R., & Arbab, A. S. (2017). Arachidonic acid metabolite as a novel therapeutic target in breast cancer metastasis. *International Journal of Molecular Sciences,* 18(12), 2661.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer Journal for Clinicians,* 68(8), 394-424.

Byeon, S. E., Yi, Y. S., Lee, J., Yang, W. S., Kim, J. H., Kim, J., et al. (2018). Hydroquinone exhibits in vitro and in vivo anti-cancer activity in cancer cells and mice. *International Journal of Molecular Sciences,* 19(3), 903.

Cantor, D. I., Cheruku, H. R., Westacott, J., Shin, J. S., & Ahn, S. B. (2020). Proteomic investigations into resistance in colorectal cancer. *Expert Review of Proteomics,* 17(1), 49-65.

Chee, J., Naran, A., Misso, N. L., Thompson, P. J., & Bhoola, K. D. (2008). Expression of tissue and plasma kallikreins and kinin B1 and B2 receptors in lung cancer. *Biological Chemistry,* 389(9), 1225-1233.

Chen, X., Rao, J., Zheng, Z., Yan, Y., Lou, S., Liu, L. P., et al. (2019). Integrated tear

proteome and metabolome reveal panels of inflammatory-related molecules via key regulatory pathways in dry eye syndrome. *Journal of Proteome Research,* 18(5), 2321-2330.

Cui, J., Shi, M., Xie, D., Wei, D., Jia, Z., Zheng, S., et al. (2014). FOXM1 promotes the warburg effect and pancreatic cancer progression via transactivation of LDHA expression. *Clinical Cancer Research,* 20(10), 2595-2606.

Dai, L., Pan, G., Liu, X., Huang, J., Jiang, Z., Zhu, X., et al. (2018). High expression of ALDOA and DDX5 are associated with poor prognosis in human colorectal cancer. *Cancer Management and Research,* 10, 1799.

Dayton, T. L., Jacks, T., & Vander Heiden, M. G. (2016). PKM 2, cancer metabolism, and the road ahead. *EMBO Reports,* 17(12), 1721-1730.

Deng, W., Rao, J., Chen, X., Li, D., Zhang, Z., Liu, D., et al. (2020). Metabolomics study of serum and urine samples reveals metabolic pathways and biomarkers associated with pelvic organ prolapse. *Journal of Chromatography B,* 1136, 121882.

Duell, E. J., Sala, N., Travier, N., Muñoz, X., Boutron-Ruault, M. C., Clavel-Chapelon, F., et al. (2012). Genetic variation in alcohol dehydrogenase (ADH1A, ADH1B, ADH1C, ADH7) and aldehyde dehydrogenase (ALDH2), alcohol consumption and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort. *Carcinogenesis,* 33(2), 361-367.

Fan, X., Rao, J., Zhang, Z., Li, D., Cui, W., Zhang, J., et al. (2018). Macranthoidin B modulates key metabolic pathways to enhance ROS generation and induce cytotoxicity and apoptosis in colorectal cancer. *Cellular Physiology and Biochemistry,* 46(4), 1317-1330.

Fearon, E. R., Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell,* 61(5), 759-767.

Gimm, O., Chi, H., Dahia, P. L., Perren, A., Hinze, R., Komminoth, P., et al. (2001). Somatic mutation and germline variants of MINPP1, a phosphatase gene located in proximity to PTEN on 10q23. 3, in follicular thyroid carcinomas. *The Journal of Clinical Endocrinology & Metabolism,* 86(4), 1801-1805.

Habermann, N., Ulrich, C. M., Lundgreen, A., Makar, K. W., Poole, E. M., Caan, B., et al. (2013). PTGS1, PTGS2, ALOX5, ALOX12, ALOX15, and FLAP SNPs: interaction with fatty acids in colon cancer and rectal cancer. *Genes & Nutrition,* 8(1), 115-126.

Hammond, W. A., Swaika, A., & Mody, K. (2016). Pharmacologic resistance in colorectal cancer: a review. *Therapeutic Advances in Medical Oncology,* 8(1), 57-84.

He, J., Jin, Y., Chen, Y., Yao, H. B., Xia, Y. J., Ma, Y. Y., et al. (2016). Downregulation of ALDOB is associated with poor prognosis of patients with gastric cancer. *OncoTargets and Therapy,* 9, 6099.

Hissong, E., & Pittman, M. E. (2019). Colorectal carcinoma screening: established methods and emerging technology. *Critical Reviews in Clinical Laboratory Sciences,* 57(1), 1-15.

Hong, J., Bose, M., Ju, J., Ryu, J. H., Chen, X., Sang, S., et al. (2004). Modulation of arachidonic acid metabolism by curcumin and related β-diketone derivatives: effects on cytosolic phospholipase A2, cyclooxygenases and 5-lipoxygenase. *Carcinogenesis,* 25(9), 1671-1679..

Khiavi, M. A., Safary, A., & Somi, M. H. (2019). Recent advances in targeted therapy of colorectal cancer: impacts of monoclonal antibodies nanoconjugates.

BioImpacts, 9(3), 123.

Kim, E. R., Kwon, H. N., Nam, H., Kim, J. J., Park, S., & Kim, Y. H. (2019). Urine-NMR metabolomics for screening of advanced colorectal adenoma and early stage colorectal cancer. *Scientific Reports,* 9(1), 1-10.

Koppenol, W. H., Bounds, P. L., & Dang, C. V. (2011). Otto Warburg's contributions to current concepts of cancer metabolism. *Nature Reviews Cancer,* 11(5), 325-337.

Kulp, K. S., & Vulliet, P. R. (1996). Mimosine blocks cell cycle progression by chelating iron in asynchronous human breast cancer cells. *Toxicology and Applied Pharmacology,* 139(2), 356-364.

Leithner, K., Hrzenjak, A., Trötzmüller, M., Moustafa, T., Köfeler, H. C., Wohlkoenig, C., et al. (2015). PCK2 activation mediates an adaptive response to glucose depletion in lung cancer. *Oncogene,* 34(8), 1044-1050..

Lewis, A. C., Wallington-Beddoe, C. T., Powell, J. A., & Pitson, S. M. (2018). Targeting sphingolipid metabolism as an approach for combination therapies in haematological malignancies. *Cell Death Discovery,* 4(1), 1-11.

Patel, M., Lu, L., Zander, D. S., Sreerama, L., Coco, D., & Moreb, J. S. (2008). ALDH1A1 and ALDH3A1 expression in lung cancers: correlation with histologic type and potential precursors. *Lung Cancer,* 59(3), 340-349.

Pelosof, L., Yerram, S., Armstrong, T., Chu, N., Danilova, L., Yanagisawa, B., et al. (2017). GPX3 promoter methylation predicts platinum sensitivity in colorectal cancer. *Epigenetics,* 12(7), 540-550.

Peltier, J., Roperch, J. P., Audebert, S., Borg, J. P., & Camoin, L. (2018). Activation peptide of the coagulation factor XIII (AP-F13A1) as a new biomarker for the screening of colorectal cancer. *Clinical Proteomics,* 15(1), 15.

Rao, J., Cheng, F., Hu, C., Quan, S., Lin, H., Wang, J., et al. (2014). Metabolic map of mature maize kernels. *Metabolomics, 10*(5), 775-787.

Rawla, P., Barsouk, A., Hadjinicolaou, A. V., & Barsouk, A. (2019). Immunotherapies and targeted therapies in the treatment of metastatic colorectal cancer. *Medical Sciences, 7*(8), 83.

Ritchie, S. A., Ahiahonu, P. W., Jayasinghe, D., Heath, D., Liu, J., Lu, Y., et al. (2010). Reduced levels of hydroxylated, polyunsaturated ultra long-chain fatty acids in the serum of colorectal cancer patients: implications for early screening and detection. *BMC Medicine, 8*(1), 13.

Tsai, S. T., Chien, I. H., Shen, W. H., Kuo, Y. Z., Jin, Y. T., Wong, T. Y., et al. (2010). ENO1, a potential prognostic head and neck cancer marker, promotes transformation partly via chemokine CCL20 induction. *European Journal of Cancer, 46*(9), 1712-1723.

Vasaikar, S., Huang, C., Wang, X., Petyuk, V. A., Savage, S. R., Wen, B., et al. (2019). Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell, 177*(4), 1035-1049.

Wang, B., Li, F., Zhou, X., Ma, Y., & Fu, W. (2019). Is microsatellite instability-high really a favorable prognostic factor for advanced colorectal cancer? A meta-analysis. *World Journal of Surgical Oncology, 17*(1), 169.

Ward, D. G., Suggett, N., Cheng, Y., Wei, W., Johnson, H., Billingham, L. J., et al. (2006). Identification of serum biomarkers for colon cancer by proteomic analysis. *British Journal of Cancer, 94*(12), 1898-1905.

Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods, 6*(5), 359-362.

Xue, X., Bredell, B. X., Anderson, E. R., Martin, A., Mays, C., & Nagao-Kitamoto, H.,

et al. (2017). Quantitative proteomics identifies steap4 as a critical regulator of mitochondrial dysfunction linking inflammation and colon cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 114(45), 9608-9617.

Yang, Q., Roehrl, M. H., & Wang, J. Y. (2018). Proteomic profiling of antibody-inducing immunogens in tumor tissue identifies PSMA1, LAP3, ANXA3, and maspin as colon cancer markers. *Oncotarget,* 9(3), 3996.

Yun, J., Mullarky, E., Lu, C., Bosch, K. N., Kavalier, A., Rivera, K., et al. (2015). Vitamin C selectively kills KRAS and BRAF mutant colorectal cancer cells by targeting GAPDH. *Science,* 350(6266), 1391-1396.

Zhang, X., Zhang, X., Li, Y., Shao, Y., Xiao, J., Zhu, G., & Li, F. (2017). PAK4 regulates G6PD activity by p53 degradation involving colon cancer cell growth. *Cell Death & Disease,* 8(5), e2820-e2820.

Zheng, Z., Liu, P., Xu, L., Peng, Z., Zhang, Y., Chen, X., et al. (2018). Metabolomics analysis of salvage chemotherapy on refractory acute myeloid leukemia patients. *RSC Advances,* 8(26), 14445-14453.

Zheng, Z., Xu, L., Zhang, S., Li, W., Tou, F., He, Q., et al. (2017). Peiminine inhibits colorectal cancer cell proliferation by inducing apoptosis and autophagy and modulating key metabolic pathways. *Oncotarget,* 8(29), 47619.

Zhou, W., Yang, F., Peng, J., Wang, F., Lin, Y., Jiang, W., et al. (2019). High pretreatment serum CA19-9 level predicts a poor prognosis for patients with stage III colon cancer after curative resection and adjuvant chemotherapy. *Journal of Cancer,* 10(16), 3810.

Figure Legends

Fig. 1 a Functional classification of the identified 551 proteins in serum samples using the PANTHER classification system (www.pantherdb.org). **b** Hierarchical clustering analysis of the identified 551 proteins in serum samples

Fig. 2 a Hierarchical clustering analysis of 567 positive-mode metabolites identified in serum samples. **b** Hierarchical clustering analysis of 431 negative-mode metabolites identified in serum samples

Fig. 3 Protein-protein interaction network analysis based on ten key pathways associated with cancer cell metabolism

Fig. 4 The heat-map generated from correlation analysis. X and Y-axes were categorized into proteins/metabolites/CA-19/CEA. Both r and p values of the correlations were displayed in distinct colors

Fig. 5 Regulatory network associated with citric acid or glutamine based on significant correlations ($r^2 \geqslant 0.49$ & FDR $\leqslant 0.05$). Metabolites and proteins were respectively represented as circular and rectangle, and their relations as edges. Metabolites in different pathway were displayed in different node colors. The positive correlations were shown in red

Tables

Table 1 The list of significant correlations associated with CEA or CA 19-9

Protein/metabolite 1	Protein/metabolite 2	Correlation coefficient	*p* value
CEA	Hydroquinone	0.74	1.99E-04
CEA	CA19-9	0.96	1.46E-11
CEA	Immunoglobulin heavy variable 1-69D	0.74	2.18E-04
CA19-9	Immunoglobulin lambda variable 4-60	0.78	4.22E-05
CEA	Immunoglobulin lambda variable 4-60	0.82	1.00E-05
CEA	Immunoglobulin kappa variable 2-40	0.71	4.84E-04
CA19-9	16beta-16-Hydroxy-3-oxo-1,12-oleanadien-28-oic acid	0.71	4.82E-04
CEA	16beta-16-Hydroxy-3-oxo-1,12-oleanadien-28-oic acid	0.73	2.50E-04
CA19-9	SM 28:3; SM(d14:2/14:1)	0.76	9.02E-05
CEA	SM 28:3; SM(d14:2/14:1)	0.79	3.42E-05
CA19-9	Coagulation factor XIII A chain	0.75	1.37E-04
CEA	Coagulation factor XIII A chain	0.82	1.16E-05
CA19.9	Leucenol	0.72	3.40E-04
CEA	Leucenol	0.76	9.88E-05
CEA	Plasma kallikrein (Fragment)	0.70	5.39E-04
CA19.9	Plasma kallikrein (Fragment)	0.74	1.73E-04
CA19.9	1-(1',3'-Benzodioxol-5'-yl)-2-butanamine	0.81	1.34E-05
CEA	1-(1',3'-Benzodioxol-5'-yl)-2-butanamine	0.87	8.56E-07