Case Control Study
Comprehensive Proteomic Signature and Identification of CDKN2A as a Promising Prognostic Biomarker and Therapeutic Target of Colorectal Cancer

Qian-Qian Wang, Yuan-Chen Zhou, Ge-yu-jia Zhou, Geng Qin, Teng-Fei Yin, Dong-Yan Zhao, Chang Tan, Shu-Kun Yao

Abstract
BACKGROUND
The carcinogenesis of colorectal cancer (CRC) involves many different molecules and multiple pathways, and the specific mechanism has not been elucidated until now. Existing studies on the proteomic signature profiles of CRC are relatively limited. Therefore, we herein aimed to provide a more comprehensive proteomic signature profile and discover new prognostic markers and therapeutic targets by performing proteomic analysis of CRC and paired normal tissues.

AIM
To investigate the proteomic signature and identify novel protein prognostic biomarkers of CRC.

METHODS
Cancer tissues and paired normal tissues were collected from 48 patients who underwent surgical removal at the China-Japan Friendship Hospital from January 2020 to June 2021. Data independent acquisition (DIA) quantitative proteomic analysis was performed using high-performance liquid chromatography–mass spectrometry/mass spectrometry (nano-UHPLC-MS/MS) to identify differentially expressed proteins,
among which those with a P adj value (t test, BH correction) <0.05 and an absolute fold change (\(|\log_{2}FC|\) >2 were identified as potential markers. Differentially expressed proteins were selected by bioinformatics analysis and validated by immunohistochemical tissue microarrays, and their association with prognosis was further analyzed with the Gene Expression Profiling Interactive Analysis (GEPIA) database to identify prognostic protein biomarkers of CRC.

RESULTS
Significantly differential protein expression was observed between cancer tissues and normal tissues. Compared with normal tissues, 1115 proteins were upregulated and 705 proteins were downregulated in CRC based on P adj <0.05 and \(|\log_{2}FC|\) >2, and bioinformatics analysis revealed that the differentially expressed proteins were involved in multiple biological processes associated with tumorigenesis, including ribosome biogenesis in eukaryotes, focal adhesion, ECM-receptor interactions and other tumor metabolism processes. Moreover, cyclin-dependent kinase inhibitor 2A (CDKN2A) expression was markedly upregulated in CRC, as validated by immunohistochemistry (0.228 vs. 0.364, P = 0.0044), and was significantly enriched in tumor proliferation and signal transduction pathways such as the cell cycle and p53 signaling pathways. High CDKN2A expression was significantly correlated with poor prognosis (P = 0.021). These results demonstrated that CDKN2A functions as a driver of CRC.

CONCLUSION
Our study provides a comprehensive proteomic signature of CRC and highlights CDKN2A as a potential powerful prognostic marker and precision therapeutic target.

Key Words: Colorectal cancer; proteomic analysis; cyclin-dependent kinase inhibitor 2A; prognostic biomarker; therapeutic target; precision treatment

**Core Tip:** In this study, quantitative proteomic analysis of CRC was comprehensively performed, revealing many differentially expressed proteins that may be useful for mining novel targets. The results revealed the overexpression of thousands of tumor proteins, among which cyclin-dependent kinase inhibitor 2A (CDKN2A) was the highlight of this study, and high CDKN2A expression in CRC was significantly associated with poor prognosis and could serve as a powerful prognostic marker and precision therapeutic target.
INTRODUCTION

According to the newest global cancer statistics, colorectal cancer (CRC) has become the second leading cause of cancer-related morbidity and mortality worldwide, and these rates continue to increase every year[1]. In the past few decades, the primary CRC treatment methods, including surgical resection, chemo- or chemoradiotherapy, immunotherapy and targeted therapy, have significantly improved[2]. However, the prognosis of CRC patients, especially those diagnosed at an advanced stage, remains poor. Colorectal carcinogenesis encompasses mechanisms of abnormal proliferation, differentiation, resistance to apoptosis and surrounding invasion of colonic epithelial cells[3]. A variety of genes and the interplay of multiple signaling pathways have been proposed to underlie the tumorigenesis of CRC, but the complex mechanism remains incompletely understood.

On the basis of the biological central dogma, gene mutation plays an important role in the occurrence and development of malignant tumors, including CRC. Previous studies have revealed possible gene mutations in patients with CRC; however, their effects are not completely consistent. For example, Pearlman et al[4] proposed that cyclin-dependent kinase inhibitor 2A (CDKN2A) mutation had nothing to do with CRC risk; in contrast, Lee et al[5] found that CDKN2A was a tumor suppressor, and Li et al[6] insisted that CDKN2A was an oncogenic gene that was significantly correlated with poor prognosis. Therefore, gene mutation alone could not fully explain the occurrence and metastasis of tumors: the perspective of proteins, the executors of function, were also needed.

Proteins play important physiological roles and have an enormous impact on cellular biology and human health, and proteomics has thus become an indispensable tool for mechanistic studies[7, 8]. Although rapidly accumulating omics studies, such as genomic, transcriptomic and epigenetic studies, have been conducted, genomic and epigenetic analyses cannot be used to fully elucidate the large variance in cancer mechanisms[9]. Proteomics analysis has been gradually applied to a variety of cancer types, such as pancreatic ductal adenocarcinoma and breast cancer, to characterize the tumor stage,
predict metastasis, identify candidate cancer biomarkers, and evaluate therapeutic effects\cite{10,11}. Nevertheless, our knowledge of the proteomics mechanisms of CRC is still limited.

Therefore, this study aimed to elucidate the proteomic profiles of CRC by performing proteomics analysis of CRC tissues and adjacent normal mucosal tissues. By performing tissue microarray validation, gene expression profiling interactive analysis (GEPIA) genomics and survival analyses in combination, we aimed to identify potential target proteins among numerous differentially expressed proteins and provide clinicians with more evidence for precision CRC treatment.

**MATERIALS AND METHODS**

**Study population**

A total of 48 CRC patients underwent surgical resection of colorectal tumors between January 2020 and June 2021 at the China-Japan Friendship Hospital. The inclusion criteria were as follows: The diagnosis of CRC was confirmed via identification by the pathology department. The exclusion criteria were as follows: 1) patients with familial adenomatous polyposis, hereditary nonpolyposis colorectal cancer, or synchronous multiple CRCs; 2) patients with inflammatory bowel disease; 3) patients who had received prior treatments, including chemoradiotherapy, targeted therapy, and immunosuppressive therapy, during the previous 6 mo before resection; and 4) patients were lacking informed consent or had missing data. Medical records were reviewed to obtain clinicopathological information, such as the patient’s age, sex, differentiation status, and TNM stage. The study was approved by the Ethics Committee of China-Japan Friendship Hospital (no. 2018-116-K85) and was conducted in accordance with the Declaration of Helsinki.

**Sample collection and preparation**

After resection of fresh colorectal cancer tissues along with the adjacent noncancerous tissues (>5 cm away from the tumor), the collected samples were cleaned with cold
normal saline, immediately frozen in liquid nitrogen and stored at -80 °C until use. Sample preparation includes multiple steps: washes, denaturation, reduction and alkylation, digestion by trypsin and extraction and cleanup of mixed peptides. Commercially available iST Sample Preparation kit (Pre Omics, Germany) was used following the manufacturer’s recommendation. Briefly, 50 μL of Lyse buffer was added and heated at 95 °C for 10 min at 1000 rpm with agitation. After cooling the sample to room temperature, trypsin digestion buffer was added, and the sample incubated at 37 °C for 2 h at 500 rpm with shaking. The digestion process was stopped with a stop buffer. Sample clean-up and desalting was carried out in the iST cartridge using the recommended wash buffers. Peptides were eluted with elution buffer (2×100 μL), and then lyophilized by SpeedVac.

**Nano-UHPLC-MS/MS analysis**

The peptides were redissolved in solvent A (A: 0.1% formic acid in water) and analyzed by an Orbitrap Fusion Lumos Tribrid coupled to an EASY-nanoLC 1200 system (Thermo Fisher Scientific, MA, USA). Next, 4 μL of the peptide sample was loaded onto a 25 cm analytical column (75 μm inner diameter, 1.9 μm resin (Dr Maisch)) and separated by gradient applied over 90 min as follows: 4% buffer B (80% ACN with 0.1% FA) for 3 min; a stepwise increase to 50% buffer B in 82 min; increase to 95% buffer B in 1 min; and a hold for 7 min. The column flow rate was maintained at 250 nL/min, and the column temperature was 55°C. The electrospray voltage was set to 2 kV. The mass spectrometer was run under data independent acquisition (DIA) mode with hybrid data strategy. A survey scan was acquired at 120,000 resolution, normalized AGC target of 250% and a maximum injection time of 100ms. In the DIA MS2 acquisition, variable Isolation window were performed. One full scan followed by 20 windows with resolution of 50,000, normalized AGC target of 200%, a maximum injection time of 86ms and normalized collision energy at 33.

**MS data analysis**
Raw DIA data were processed and analyzed by Spectronaut 15.0 (Biognosys AG, Switzerland) with default settings. Spectronaut was set up to search the UniProt-Homo sapiens database (version 201907, 20428 entries) assuming that trypsin was used as the digestion enzyme. Carbamidomethyl (C) was specified as the fixed modification. Oxidation (M) was specified as the variable modification. The retention time prediction type was set to dynamic iRT. Data extraction was performed by Spectronaut based on extensive mass calibration. Spectronaut was used to determine the ideal extraction window dynamically depending on the iRT calibration and gradient stability. The Q value (FDR) cutoff on the precursor level was 1%, and the protein level was 1%. Decoy generation was set to mutated, which was similar to scrambled but only applied a random number of AA position swamps (min=2, max=length/2). The normalization strategy was set to local normalization. The average top 3 filtered peptides that passed the 1% Q value cutoff were used to calculate the major group quantities. After application of the t test, differentially expressed proteins were identified by an P adj value <0.05 and an absolute fold change >2. The R package was used to visualize the differential expression for bioinformatics analysis. Gene set enrichment analysis (GSEA) was completed by using GSEA v4.2.3 (https://www.gsea-msigdb.org/gsea/index.jsp).

**Immunohistochemical (IHC) analysis**

The Human Protein Atlas (HPA, http://www.proteinatlas.org/) contains images of histological sections from normal and cancer tissues obtained by immunohistochemistry and is publicly available at v20.proteinatlas.org. Antibodies were labeled with DAB (3,3'-diaminobenzidine), and the resulting brown staining indicated where an antibody had bound to its corresponding antigen. The section was further counterstained with hematoxylin to enable visualization of microscopic features. Each sample was represented by 1 mm tissue cores. CRC pathology and normal colon tissue microarrays were downloaded to quantify protein expression. ImageJ software was used to assess the staining area and integrated optical density (IOD) to determine the average optical density (AOD) values. All images were manually evaluated by two
independent observers. The Mann-Whitney U test was used for statistical analyses, and the statistical significance level was set to 0.05.

**Gene expression and survival analysis**

GEPIA (http://geopia.cancer-pku.cn/index.html) provides customizable functions such as tumor/normal differential expression analysis and patient survival analysis[12]. A box plot was generated to compare gene expression in CRC. Genes with $|\log_2 FC|$ values >1 and q values <0.01 were considered to be differentially expressed. GEPIA was also used to perform overall survival (OS) analysis based on gene expression. We selected the median gene expression as a group cutoff for splitting the high- and low-expression cohorts. Kaplan-Meier plots and the log-rank test were used to analyze differences in survival times between patients with high and low expression.

**RESULTS**

**Proteomic signatures**

We collected a total of 48 pairs of cancer tissues and matched adjacent normal tissues after surgical treatment of CRC for quantitative proteomic detection. We found differences in the protein signatures between the CRC group and the normal group, and PLS-DA could clearly distinguish the two groups (Fig 1A). In total, we identified 7643 proteins as well as 1820 differentially expressed proteins with an FDR < 1% in the CRC group vs the normal group, which were used for subsequent enrichment analysis and tumor target mining. There were 1115 proteins upregulated in the tumor group, including CEAM6/S, IFM1, LAT1, RAB3, VSL1, TACC3, HSZL1, IMA1, DIXF, and UPAR, while 705 proteins, including SYUG, NOS1, RT26, GEMI4, HAUS5, TET1, NB5R2, PERI, LIPS and PPLA, were downregulated in the tumor group compared with the normal group (Fig 1B, Fig S1).

GO analysis of the differentially expressed proteins was performed, and the differentially expressed proteins were significantly enriched in biological processes (BPs; RNA metabolic processing, RNA processing, biological adhesion, and other
processes), cellular components (CCs, binding, protein binding, catalytic activity, RNA binding, etc.), and molecular functions (MFs, extracellular exosome, intracellular nonmembrane-bounded organelle, nucleoplasm, etc.) (Fig 1C). The differentially expressed proteins were further subjected to subsequent KEGG pathway analysis, which identified the PI3K-Akt signaling, focal adhesion, ribosome biogenesis in eukaryotes, cell cycle, ECM-receptor interaction and DNA replication pathways as being significantly enriched (Fig 1D). To explore the mechanisms of the differentially expressed proteins in cancer development to the greatest extent possible, we performed further GSEA, which showed that transcription factors, the cell cycle, the p53 signaling pathway and other processes were significantly enriched in CRC tissues (Fig 1E, F).

Visualizing and displaying the bioinformatics analysis results led to the identification of a special protein, CDKN2A, which was significantly upregulated in tumors compared with the normal tissues. KEGG pathway enrichment analysis showed that CDKN2A was involved in the cell cycle, PI3K-Akt signaling pathway, and p53 signaling pathway. GSEA further confirmed that CDKN2A was involved in regulating the cell cycle and p53 signaling pathway and functioned as a core gene in the gene set. Therefore, we speculated that CDKN2A plays a pivotal role in colorectal carcinogenesis and has potential as a prognostic biomarker.

**IHC validation of CDKN2A**

We collected protein microarrays from a total of 3 normal colon tissues and 12 CRC tissues from the HPA database that had been subjected to antigen antibody labeling reactions using CAB0093 antibodies. The AOD value of each chip was calculated using ImageJ. The protein expression of CDKN2A was significantly different between CRC and normal samples (P<0.05) (Fig 2A, B). Low or no protein expression of CDKN2A was observed in the normal samples, while the staining intensity and area were significantly increased in CRC samples, implying that CDKN2A protein expression was significantly upregulated (P = 0.0044). These results were consistent with those of
proteomic analysis and further confirmed that the protein expression of CDKN2A was significantly increased in CRC. The results are presented as a bar graph (Fig 2B).

**CDKN2A gene expression and survival analysis**

To explore the expression profile of the CDKN2A gene in normal mucosa and CRC tissues as well as its association with patient prognosis, we performed differential gene expression analysis and survival analysis using the GEPIA database. We found that CDKN2A gene expression differed significantly between CRC and normal tissues, with significantly higher expression in colon cancer (P<0.05), and the same trend was observed in rectal cancer compared with normal tissue (P<0.05) (Fig 2C). Further overall survival analysis using GEPIA showed that a total of 362 colorectal cancer patients were matched to normal tissues from the TCGA and GTEx databases and that patients with high expression of the CDKN2A gene had a significantly decreased overall survival (P = 0.021), while patients with low expression of the CDKN2A gene had a relatively good prognosis (Fig 2D).

**DISCUSSION**

Signaling proteins are very attractive for the precise treatment of cancer. However, research on the proteomic spectrum of CRC is still limited. In this study, 48 pairs of cancer and adjacent normal tissues were collected for proteomics analysis, and 1115 upregulated proteins and 705 downregulated proteins were observed in the CRC group. First, we made an enormous effort to provide more comprehensive proteomic signatures for CRC. Second, we identified many proteins with abnormal expression abundance in CRC and validated their potential as tumor targets. Interestingly, among all the differentially expressed proteins, we highlighted one upregulated protein, cyclin-dependent kinase inhibitor 2A (CDKN2A), which is involved in both the cell cycle and the p53 signaling pathway, and a tissue microarray confirmed that the CDKN2A protein was highly expressed in CRC and was associated with low OS survival and
poor prognosis. This suggests that CDKN2A plays a cancer-driving role in CRC and is predictive of poor survival.

CDKN2A, a well-known tumor suppressor located at chromosome 9p21\cite{[13, 14]}. Two proteins are encoded, p16INK4a and p14ARF, which exert different regulatory effects on the cell cycle: p14ARF interacts with and degrades MDM2, preventing p53 inactivation by ubiquitin-mediated proteolysis or transcriptional silencing, and P16INK4a binds CDK4 and CDK6 to prevent phosphorylation of the Rb protein\cite{[13, 15]}. Loss-of-function mutations or homozygous deletions of CDKN2A resulted in the loss of both proteins, releasing the G1-S and G2-M cell cycle checkpoints and resulting in uninhibited cell proliferation and tumor formation\cite{[16]}. A literature database search revealed that CDKN2A has been extensively researched in the context of melanoma, pancreatic cancer and other tumor types, but the findings are controversial\cite{[15, 17]}. CDKN2A plays a role in CRC initiation and progression via multiple mechanisms. Ferroptosis, a newly defined form of cell death that differs from apoptosis and autophagy and is characterized by iron overload, lipid reactive oxygen species and lipid peroxidation, is involved in the carcinogenesis, progression, and treatment of CRC\cite{[18]}. A study showed that CDKN2A is a ferroptosis-associated gene that is involved in the iron metabolism and tumorigenesis of CRC by enhancing p53-dependent transactivation and ferroptosis. The upregulation of this gene in CRC correlated with poor prognosis and could be considered part of a predictive model\cite{[19]}. However, our differential protein enrichment analysis did not reveal the involvement of CDKN2A in this process. In addition, CDKN2A is a prominent hallmark of cellular senescence, whereas reactive oxygen species, DNA damage, and chronic inflammation all induce cellular senescence\cite{[20-22]}. In mice, p53 genes in senescent cells might transform these cells into highly aggressive, cancer-initiating cells with long-term accumulation progressing to a cancerous state\cite{[23]}. Another recent study on the inflammatory microenvironment in CRC found significant differences in CDKN2A expression between the epithelium and stroma, with a 10-fold decrease in the epithelium and a 17-fold increase in the stroma\cite{[24, 25]}. This study speculated that the activation of CDKN2A expression in the stroma was
influenced by cellular senescence and oncogene activation, and senescent fibroblasts prematurely accumulated in the stroma as a result of chronic inflammation and oxidative stress\textsuperscript{[24]}. Moreover, CDKN2A participates in the development of CRC through the Wnt/\beta-catenin signaling pathway\textsuperscript{[13, 26]}.

The role of CDKN2A in CRC is also controversial. In 2017, a comprehensive cancer center conducted exploratory research on CRC cancer susceptibility genes and found that CDKN2A had high- or moderate-level gene mutations. This study pointed out that CDKN2A gene mutation was not traditionally associated with CRC risk\textsuperscript{[4, 27]}. Some studies have shown that CDKN2A usually acts as a tumor suppressor. On the one hand, CRC patients have a poor prognosis due to the occurrence of promoter region hypermethylation, CDKN2A gene silencing or dysfunction, which causes uncontrolled cell proliferation and carcinogenesis\textsuperscript{[5, 28]}. On the other hand, CDKN2A participates in the ILF3-AS1/EZH2/CDKN2A/H3K27me3 axis, and downregulation of CDKN2A accelerates CRC proliferation and metastasis\textsuperscript{[29]}. However, the results of some other studies were opposite but consistent with our findings. A recent convincing study demonstrated that CDKN2A was a cancer-driver gene that could effectively predict the poor prognosis and clinical status of CRC patients\textsuperscript{[6]}. Concordant with this, several studies found that CDKN2A was a risk gene for overexpression in CRC and highlighted that increased CDKN2A protein expression, rather than loss of protein expression, was associated with poor prognosis in CRC\textsuperscript{[3, 19, 25, 30]}.

Possible explanations for these contradictory results are as follows: 1) It is well known that CDKN2A protein expression is affected by genomic deletions, point mutations, and promoter region hypermethylation\textsuperscript{[28, 31]}. The promoter region of the CDKN2A gene was hypermethylated, combined with methylase, and thus could not bind to the RNA binding enzyme, resulting in decreased or silent protein expression. Therefore, we speculated that this result was due to hypermethylation of gene body regions rather than promoter regions, and CDKN2A protein expression was thus positively altered in the tumor with the help of histone modification\textsuperscript{[32]}.

2) An additional speculation was that CDKN2A protein expression was influenced by the stage of tumor development.
CDKN2A is known to be involved in the epithelial-stromal microenvironment during CRC development, with different stages of tumor progression and diverse states of the tumor microenvironment, and CDKN2A protein expression thus varies substantially in multiple studies\textsuperscript{[24].} It has also been proposed that the upregulation of CDKN2A protein expression is a consequence of tumor metastasis\textsuperscript{[30].} 3) The degree of cellular senescence varies, leading to different trends in the expression of CDKN2A and therefore to conflicting inferences about its role in tumors\textsuperscript{[22].} Additionally, since the mRNA expression of CDKN2A could be promoted by nontumor cells present in the tumor microenvironment, interpreting of bulk transcriptional data remains challenging, requiring single-cell RNA-seq analysis to analyze population-specific transcription of this gene\textsuperscript{[33].} Because of the limitations of the study design, small sample size, and publication bias, the mechanism of CDKN2A in CRC has not been fully elucidated despite multiple investigations being conducted. This problem needs to be resolved in the future to fully construct the genomic, transcriptomic, epigenetic and proteomic landscape of CDKN2A in CRC.

Our study more comprehensively revealed the proteomic signature of CRC, although not for the first time, and provided important biological and clinical insights. Of course, our study has some limitations. First, similar to the shortcomings of previous studies, this was a single-center study, and the findings may therefore not have general applicability. Second, the lack of genomic and epigenetic data from our own tissues prevented the complete elucidation of the gene-mRNA-protein chain. To some extent, it is possible to complement the understanding of the CRC mechanism at the protein level. Finally, to fully substantiate the tumor suppressor or cancer-promoting role of the CDKN2A gene, multicenter studies are needed.

**CONCLUSION**

In conclusion, this study more comprehensively revealed the proteomic signature of CRC. CDKN2A was identified among all the differentially expressed proteins and shown to be involved in the cell cycle, p53 signaling pathway and other mechanisms.
Moreover, CDKN2A is associated with poor overall survival and may serve as a prognostic biomarker and treatment target for CRC.
<table>
<thead>
<tr>
<th>Source</th>
<th>Title</th>
<th>Words</th>
<th>Originality (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td><a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a></td>
<td>20</td>
<td>1</td>
</tr>
</tbody>
</table>

Exclude Quotes: ON
Exclude Bibliography: ON
Exclude Sources: < 1%
Exclude Matches: < 10 Words