

76666\_Auto\_Edited.docx

**Name of Journal:** *World Journal of Gastrointestinal Oncology*

**Manuscript NO:** 76666

**Manuscript Type:** ORIGINAL ARTICLE

*Basic Study*

**Proteomic signatures of infiltrative gastric cancer by proteomic and bioinformatic analysis**

Proteomic signatures of IGC

**Abstract**

**BACKGROUND**

Proteomic signatures of Ming's infiltrative gastric cancer (IGC) remain unknown.

**AIM**

To elucidate the molecular characteristics of IGC at the proteomics level.

**METHODS**

Twelve pairs of IGC and adjacent normal tissues were collected and their proteomes were analyzed by high performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS). The identified peptides were sequenced de novo and matched against SwissProt database using the Maxquant software. The differentially expressed proteins (DEPs) were screened using  $|\log_2(\text{Fold change})| > 1$  &  $p\text{-adj} < 0.01$  as the thresholds. The expression levels of selected proteins were verified by Western blotting. The interaction network of the DEPs was constructed with the STRING database and visualized using Cytoscape with cytoHubba software. The DEPs were functionally annotated using clusterProfer, STRING and DAVID for gene ontology (GO) and KEGG pathways.  $P < 0.05$  was considered statistically significant.

## RESULTS

A total of 7361 DEPS were identified, of which 94 were significantly up-regulated and 223 were significantly down-regulated in the IGC relative to the normal gastric tissues. The top 10 up-regulated proteins were MRTO4, BOP1, PES1, WDR12, BRIX1, NOP2, POLR1C, NOC2L, MYBBP1A and TSR1, and the top 10 down-regulated proteins were NDUFS8, NDUFS6, NDUFA8, NDUFA5, NDUFC2, NDUF8, NDUF5, NDUF9, UQCRC2 and UQCRC1. The up-regulated proteins were enriched for 9 biological processes including DNA replication, ribosome biogenesis and initiation of DNA replication, and the cellular component MCM complex. Among the down-regulated proteins, 17 biological processes were enriched, including glucose metabolism, pyruvic acid metabolism and fatty acid  $\beta$ -oxidation. In addition, mitochondrial inner membrane, mitochondrial matrix and mitochondrial proton transport ATP synthase complex were among the 6 enriched cellular components, and 11 molecular functions including NADH dehydrogenase activity, acyl-CoA dehydrogenase activity and NAD binding were also enriched. The significant KEGG pathways for the up-regulated proteins were DNA replication, cell cycle and mismatch repair, whereas 18 pathways including oxidative phosphorylation, fatty acid degradation and phenylalanine metabolism were significantly enriched among the down-regulated proteins.

## CONCLUSION

The proteins involved in cell cycle regulation, DNA replication and mismatch repair, and metabolism were significantly altered in IGC, and the proteomic profile may enable the discovery of novel biomarkers.

**Key Words:** Infiltrative gastric cancer; IGC; Proteomics; Molecular biological characteristics; Ming's classification

Zhang LH, Zhuo HQ, Hou JJ, Yang Z, Cheng J, Cai JC. Proteomic signatures of infiltrative gastric cancer by proteomic and bioinformatic analysis. *World J Gastrointest Oncol* 2022; In press

**Core Tip:** A total of 7361 proteins were identified and 317 proteins were significantly abnormal expressed in IGC. 20 hub proteins were found and some of them were verified in IGC. The cell cycle regulation, DNA replication and mismatch repair, and metabolism were significantly altered in IGC.

## **INTRODUCTION**

Gastric cancer (GC) is the fourth most prevalent malignancy worldwide and ranks third in terms of mortality, especially in Asia<sup>[1, 2]</sup>. It is divided into infiltrative gastric cancer (IGC) and expanding gastric cancer (EGC) according to Ming's system of classification, which is related to Bormann classification (protrusion and ulcer type), Lauren classification (intestinal and diffuse type) and WHO classification (papillary adenocarcinoma, adenosquamous carcinoma, squamous cell carcinoma, carcinoid, *etc.*). The incidence of IGC is 61.5%, and its prognosis is worse compared to that of EGC<sup>[3-5]</sup>. Although the IGC classification is relevant in clinical diagnosis, treatment and prognosis assessment<sup>[6]</sup>, the molecular mechanism of IGC is not completely understood. In this study, we analyzed the proteomes of IGC and normal gastric tissues using high performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS), and identified the differentially expressed proteins (DEPs). The key proteins were screened and functionally annotated by gene ontology (GO) and KEGG pathway analyses. This study is the first to profile the IGC proteome, and may help unravel the molecular mechanisms and novel biomarkers of IGC.

## **MATERIALS AND METHODS**

### ***Clinical samples***

Twelve pairs of IGC tissues and normal resection margin tissues were obtained from Zhongshan Hospital Affiliated to Xiamen University. The samples were fixed in formalin and snap frozen at -80°C. The frozen tissue sections were stained with hematoxylin and eosin (H&E) as per standard protocols, and examined by the chief surgeon and chief physician of the pathology department. All patients signed an informed consent form, and the study was approved by the ethics committee of Zhongshan Hospital affiliated to Xiamen University.

### ***Peptide preparation***

The frozen tissue samples were homogenized in liquid nitrogen, and ultrasonicated with lysis buffer (8 M urea, 1% protease inhibitor and 2 mmol/L EDTA). The protein samples were reduced with 5 mmol/L dithiothreitol at 56°C for 30 min, and then incubated with 11 mmol/L iodoacetamide for 15 min at room temperature in the dark. The urea concentration of the sample was diluted to less than 2 M.

### ***HPLC-MS/MS***

The peptides were fractionated by high pH reverse HPLC using Agilent 300 Extend C18 column (5 µm particle size, 4.6 mm inner diameter, 250 mm length), and a step gradient of 8% to 32% acetonitrile at pH 9. Sixty components were separated in 60 min. The peptides were pooled into 6 components and freeze-dried under vacuum. The lyophilized peptides were dissolved in 0.1% v/v aqueous formic acid and then separated using the EASY-nLC 1000 ultra-high performance liquid system. Mobile phase A consisted of 0.1% formic acid and 2% acetonitrile, and mobile phase B was an aqueous solution of 0.1% formic acid and 90% acetonitrile. The gradient setting was as follows: 0–62 min, 5%–22% B; 62–82 min, 22%–35% B; 82–86 min, 35%–80% B; 86–90 min, 80% B. The flow rate was maintained at 450 nL/min.

The separated peptides were then injected into the NSI ion source and analyzed by the Q Exactive™ Plus mass spectrometer. The ion source voltage was set to 2 kV, and the peptide precursor ions and their secondary fragments were detected and analyzed by high-resolution Orbitrap. The scanning range of the primary mass spectrometer was set to 350–1800 m/z, the scanning resolution to 70,000, and the secondary scanning

resolution to 17,500. Data were acquired using a data-dependent scanning (DDA) program. The entire process of HPLC-MS/MS is briefly summarized and shown in Figure 1.

#### ***Database search***

The secondary mass spectrum data were searched against the SwissProtHuman (20317 sequences) database using Maxquant (v1.5.2.8). The search parameters were as follows: restriction enzyme - Trypsin/P, number of missing cleavage sites - 2, mass error tolerance of the primary precursor ion of the first search and the main search - 20 ppm and 5 ppm respectively, mass error tolerance of the secondary fragment ion - 0.02 Da, fixed modification - cysteine alkylation, and variable modification - oxidation of methionine and the acetylation of the N-terminus. The FDR for protein identification and PSM identification was set to 1%.

#### ***Western blotting***

Total protein was extracted using RIPA lysis buffer (Thermo Scientific, United Kingdom), and its concentration was determined with an enhanced bicinchoninic acid assay kit (CWBio, China). Around 40 mg protein per sample was separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and transferred to polyvinylidene difluoride membranes (CWBio, China). After blocking with 5% non-fat milk for 1h at room temperature (RT), the membranes were incubated overnight with primary antibodies targeting MRTO4 (1:2000, ab212044, Abcam, USA), BOP1 (1:3000, ab32053, Abcam, USA), PES1 (1:5000, ab56701, Abcam, USA), NDUFS8 (1:3000, ab226760, Abcam, USA), NDUFS6 (1:3000, ab226760, Abcam, USA), NDUFA8 (1:3000, ab226760, Abcam, USA) and  $\beta$ -actin (1:5000, ab8226, Abcam, USA) at 4°C. The membranes were washed thrice with TBST (tris-buffered saline with Tween 60), and probed with horseradish peroxidase-conjugated secondary antibody (1:5000) for 1.5 h at room temperature. The protein bands were visualized using an enhanced chemiluminescence system, and the membranes were exposed to X-ray films (Bio-Rad, USA). Densitometric analysis was performed using Image Pro-Plus software (Media Cybernetics, USA), and relative protein expression levels were normalized to  $\beta$ -actin.

### ***Protein-protein interaction network and hub protein screening***

Protein-protein interaction (PPI) network was analyzed using STRING (<https://string-db.org/>, version 11.0)<sup>[7]</sup> with *Homo sapiens* as the species. The DEPs were imported into the STRING website, and the Cytoscape plug-in cytoHubba was used to screen for hub proteins. The data were from experiments, databases, co-expression and co-occurrence, and the interaction score was 0.4. The TSV format file of PPI results was imported into Cytoscape (version 3.8.2) software for visual editing and network display.

### ***Gene Ontology (GO) and KEGG pathway enrichment analysis***

#### **(1) ClusterProfer enrichment analysis**

The bioconductor, org.Hs.eg.db and clusterProfiler packages<sup>[8]</sup> were simultaneously installed in the R software. The gene names of the DEPs were converted to the ENTERZID format using org.Hs.eg.db package. The enrichGO software was used for GO (biological processes, BP; cell components, CC; molecular functions, MF) enrichment analysis and enrichKEGG for KEGG pathway enrichment analysis. *P* value < 0.05 was considered as a significant enrichment.

#### **(2) String enrichment analysis**

The selected proteomic gene names were enriched and analyzed on the STRING website (<https://string-db.org/>, version 11.0) using the full network mode with *Homo sapiens* as the enriched species. The type of interactions between the protein was indicated by the network connection line. The data were retrieved from experiments, databases, and co-expression and co-occurrence analyses. The interaction score was set at 0.4, and *P* value < 0.05 was considered significant.

#### **(3) DAVID enrichment analysis**

The selected proteomic genes were enriched in DAVID version 6.8 (The Database for Annotation, Visualization and Integrated Discovery; <https://david.ncifcrf.gov/>)<sup>[9, 10]</sup> with *Homo sapiens* as the selected species. GO enrichment was analyzed by Gene\_Ontology, and KEGG pathways by KEGG\_PATHWAY. *P* value < 0.05 was considered statistically significant.

### ***Statistical methods and software***

The difference in protein expression levels between IGC and normal tissues was analyzed by the paired  $t$  test using R (version R4.0.3, <https://www.r-project.org/>). The volcano graph was plotted using the ggplot and ggrepel packages. The protein interaction network was visualized in Cytoscape (version 3.8.2, <https://cytoscape.org/>) software<sup>[11]</sup>. The pictures were edited using Adobe Photoshop CS6 software.

## **RESULTS**

### ***Proteomic signature of IGC relative to normal gastric tissues***

The proteomes of 12 pairs of histo-pathologically confirmed IGC and adjacent normal gastric tissues (Figure 2a/b) were profiled. The representative MS/MS fragments are shown in Figure 2c, and the expanded region of a single FTMS full scan with a mass range of 400-1800  $m/z$  is shown in Fig. 2d. There were a total of 7361 DEPs between IGC and normal tissues ( $P < 0.01$ ), of which 94 were up-regulated and 223 were downregulated in the IGC samples (Figure 2e).

### ***Proteomic signature of IGC tissues***

The interaction networks of the upregulated and downregulated proteins in IGC are respectively shown in Figure 3a and Figure 3c. According to the MCC algorithm, the top 10 up-regulated proteins were MRTO4, BOP1, PES1, WDR12, BRIX1, NOP2, POLR1C, NOC2L, MYBBP1A and TSR1 (Figure 3b), whereas NDUFS8, NDUFS6, NDUFA8, NDUFA5, NDUFC2, NDUFB8, NDUFB5, NDUFB9, UQCRC2 and UQCRC1 were the key down-regulated proteins (Figure 3d). The expression levels of MRTO4, BOP1, PES1 (Figure 3e), NDUFS8, NDUFS6 and NDUFA8 (Figure 3f) were verified in IGC tissues by western blotting.

### ***Proteomic signatures of GO analysis in IGC***

The significantly enriched GO terms for the up-regulated and down-regulated proteins and their interactive networks are shown in Figure 4a-d. The biological processes enriched in the upregulated proteins included DNA replication, ribosome biogenesis, and DNA replication initiation (Figure 4e), and MCM complex was the key cellular component (Figure 4f). The molecular functions did not show co-enrichment (Figure



4g). Among the significantly down-regulated proteins, 17 biological processes including glucose metabolism, pyruvate metabolism and fatty acid beta-oxidation were significantly enriched (Figure 4h). In addition, 6 cell components including mitochondrial inner membrane, mitochondrial matrix and mitochondrial proton-transporting ATP synthase complex (Figure 4i), and 11 molecular functions including NADH dehydrogenase activity, acyl-CoA dehydrogenase activity and NAD binding (Figure 4j) were enriched.

#### ***Proteomic signatures of KEGG pathways in IGC***

Three KEGG pathways were significantly enriched in the up-regulated proteins (Figure 5a), including DNA replication, cell cycle and mismatch repair (Figure 5b). Among the down-regulated proteins, 18 KEGG pathways were enriched (Figure 5c), such as oxidative phosphorylation, fatty acid degradation, and phenylalanine metabolism (Figure 5d).

## **DISCUSSION**

The top 20 hub proteins identified in the IGC tissues were MRTO4, BOP1, PES1, WDR12, BRX1, NOP2, POLR1C, NOC2L, MYBBP1A, TSR1, NDUF8, NDUF6, NDUF8, NDUF5, NDUF2, NDUF8, NDUF5, NDUF9, UQCRC2 and UQCRC1. PES1 is highly expressed in GC tissues, and knocking down PES1 in GC cells inhibited their proliferation<sup>[12]</sup>. On the other hand, UQCRC2 is downregulated in GC and its overexpression inhibited the migration and invasion of the tumor cells<sup>[13]</sup>. Therefore, these proteins are promising prognostic biomarkers of IGC.

The proteins that were upregulated in IGC tissues showed significant enrichment of DNA replication, ribosome biogenesis and DNA replication initiation, MCM complex, and the cell cycle and mismatch repair signal pathways. This strongly suggests that these proteins exert a pro-proliferative and oncogenic function in IGC, most likely by promoting DNA replication, cell cycle progression and mismatch repair. Changes in the MCM complex have been previously reported in IGC cells, and are regulated by miRNAs<sup>[14]</sup>. In addition, DNA replication and cell cycle signaling pathways are key

factors involved in the proliferation of GC cells<sup>[15, 16]</sup>, whereas the mismatch repair pathway affects drug resistance and metastasis<sup>[17-21]</sup>.

The down-regulated proteins were enriched in glucose metabolism, pyruvate metabolism, phenylalanine metabolism, fatty acid  $\beta$ -oxidation, oxidative phosphorylation, mitochondrial inner membrane, mitochondrial matrix, mitochondrial proton transport ATP synthase complex, NADH dehydrogenase activity, acyl-CoA dehydrogenase activity and NAD binding. This is indicative of aberrant mitochondrial metabolism in the IGC cells. Consistent with our findings, a previous study reported dysregulated oxidative phosphorylation in GC<sup>[22]</sup>.

The significant DEPs identified in our study are potential diagnostic and prognostic markers, as well as therapeutic targets in IGC, and will have to be validated in a large cohort from multiple centers. Furthermore, the proteomic signatures of IGC provide insights into the possible mechanisms underlying IGC progression, which likely involve DNA replication, cell cycle, mismatch repair, and energy metabolism pathways, and will also contribute to precision medicine for more accurate diagnosis and better treatment effect.

## **CONCLUSION**

This study has several limitations that ought to be considered. First, only 12 paired IGC and adjacent normal tissues were analyzed, and the sample size will have to be increased by involving multiple centers in the follow-up study. Second, few proteins could be verified, and the number will have to be increased in future studies by mass spectrometry.

## **ARTICLE HIGHLIGHTS**

### ***Research background***

The prognosis of infiltrative gastric cancer (IGC) patients remains relatively poor. Therefore, it is necessary to explore the molecular mechanisms underlying the occurrence and development of IGC.

### ***Research motivation***

The proteomic signatures of IGC remain unknown.

### ***Research objectives***

To profile the proteome of IGC.

### ***Research methods***

The proteins from IGC and normal <sup>3</sup> tissue samples were analyzed by high performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS) and searched against the database *via* Maxquant software. The differentially expressed proteins (DEPs) were screened using  $|\log_2(\text{Fold Change})| > 1$  &  $p\text{-adj} < 0.01$  as the thresholds. The expression levels of some proteins were verified by Western blotting. The interaction network of DEPs was constructed using the STRING database, and the key proteins were visualized using Cytoscape cytoHubba software. Finally, clusterProfer, String and DAVID were used for gene ontology (GO) and KEGG pathway enrichment analysis of DEPs, with  $P < 0.05$  as the threshold.

### ***Research results***

<sup>1</sup> A total of 7361 DEPS were identified, of which 94 were significantly up-regulated and 223 were significantly down-regulated in the IGC relative to the normal gastric tissues. The top 10 up-regulated proteins were MRTO4, BOP1, PES1, WDR12, BRIX1, NOP2, POLR1C, NOC2L, MYBBP1A and TSR1, and the top 10 down-regulated proteins were NDUFS8, NDUFS6, NDUFA8, NDUFA5, NDUFC2, NDUFB8, NDUFB5, NDUFB9, UQCRC2 and UQCRC1. The up-regulated proteins were enriched for 9 biological processes including DNA replication, ribosome biogenesis and initiation of DNA replication, and the cellular component MCM complex. Among the down-regulated proteins, 17 biological processes were enriched, including glucose metabolism, pyruvic acid metabolism and fatty acid  $\beta$ -oxidation. In addition, mitochondrial inner

membrane, mitochondrial matrix and mitochondrial proton transport ATP synthase complex were among the 6 enriched cellular components among the down-regulated proteins, and 11 molecular functions including NADH dehydrogenase activity, acyl-CoA dehydrogenase activity and NAD binding were also enriched. The significant KEGG pathways for the up-regulated proteins were DNA replication, cell cycle and mismatch repair, whereas 18 pathways including oxidative phosphorylation, fatty acid degradation and phenylalanine metabolism were significantly enriched among the down-regulated proteins.

### ***Research conclusions***

The proteins involved in cell cycle regulation, DNA replication and mismatch repair, and metabolism were significantly altered in IGC, which provides a basis for the future identification of novel biomarkers.

### ***Research perspectives***

This study reveals the proteomic signature of IGC.

3%

SIMILARITY INDEX

PRIMARY SOURCES

1	<a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a> Internet	29 words — 1%
2	Abdullah Saleh Al-wajeeh, Salizawati Muhamad Salhimi, Majed Ahmed Al-Mansoub, Imran Abdul Khalid et al. "Comparative proteomic analysis of different stages of breast cancer tissues using ultra high performance liquid chromatography tandem mass spectrometer", PLOS ONE, 2020 Crossref	16 words — 1%
3	<a href="http://cancerres.unboundmedicine.com">cancerres.unboundmedicine.com</a> Internet	16 words — 1%
4	<a href="http://www.spkx.net.cn">www.spkx.net.cn</a> Internet	12 words — < 1%

EXCLUDE QUOTES ON

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE SOURCES < 12 WORDS

EXCLUDE MATCHES < 12 WORDS