ITRAQ and PRM-based quantitative saliva proteomics in gastric cancer: biomarker discovery

Zhanyan Liu[1,3†], Jieren Liu[2†], Jimao Mo[2], Haoran Chi[2], Lei Huang[2], Kaixin Luo[2], Runtao Wen[2], Guanlin Li[1], Yuanzhe Cai[2*], Feijuan Huang[1*], Zhengzhi Wu[1*]

[1] Shenzhen Institute of Geriatrics, Shenzhen Second People's Hospital, Shenzhen, China.
[2] Shenzhen Technique University, Shenzhen, China.
[3] Changsha Hospital of Traditional Chinese Medicine, Changsha, China.

*Corresponding author(s). E-mail(s): caiyuanzhe@sztu.edu.cn; huangfeijuan@163.com; szwzz001@163.com;
Contributing authors: lzyhappiness@163.com; liujierenxy@126.com; 202300803086@stumail.sztu.edu.cn; 202200501148@stumail.sztu.edu.cn; 202200501113@stumail.sztu.edu.cn; lantianyuuuuu@iCloud.com; 202200501085@stumail.sztu.edu.cn; ashton@lglws.com;
†These authors contributed equally to this work.

Abstract

Gastric cancer (GC) remains one of the leading causes of cancer-related mortality worldwide, and early detection is critical for improving patient prognosis. Traditional diagnostic methods, such as serum biopsy, endoscopy and CT scans, are invasive, costly, and often uncomfortable for patients. In addition, these traditional markers show limited sensitivity and specificity for early-stage GC, especially in asymptomatic patients. Therefore, this study was conducted to investigate salivary biomarkers for early GC detection using iTRAQ-based and PRM-based proteomics techniques to identify differentially expressed proteins in GC patients. We utilized iTRAQ technology to quantitatively analyze salivary proteins from two gastric cancer groups (GC group 1 and GC group 2) and healthy controls. Differential expression of proteins between these groups was assessed, and functional annotation was performed using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. Protein-protein interaction (PPI) networks were constructed to investigate potential molecular interactions. Validation of saliva-specific biomarkers for GC was performed using PRM. Kaplan–Meier (KM) survival analysis were employed to evaluate the clinical relevance of the identified proteins.In the screening cohort, 671 proteins with unique peptide segments were identified through iTRAQ analysis. Among them, 124 proteins were significantly differentially expressed in GC group 1, and 102 proteins were differentially expressed in GC group 2 compared to healthy controls. A total of 56 overlapping DEPs were identified between the two GC groups, with 24 proteins upregulated and 32 proteins downregulated. GO analysis revealed that these DEPs were involved in several biological processes including nucleosome, DNA packaging complex. KEGG analysis indicated these proteins involved in several pathways including transcriptional misregulation in cancer, alcoholism, shigellosis, IL-17 signaling pathway. In the validation cohort, we found consistent expression patterns for four proteins: S100A8, S100A9, CST4, and CST5. Results show that the levels of S100A8, S100A9 were upregulated and CST4, and CST5 downregulated in the saliva. The Kaplan-Meier (KM) survival analysis indicated that elevated CST5 and CST4 expression levels were associated with faster disease progression while high expression of S100A8 and S100A9 correlated with poor progress. Overall, S100A8, S100A9, CST4, and CST5 emerged as cancer-specific saliva biomarkers for early detection and diagnosis of GC, providing a scientific basis for clinical application.

Keywords: ITRAQ, PRM, Proteomic, Gastric Cancer, Saliva Biomarkers

1 Introduction

Gastric cancer (GC) is the major global health issue, with more than 1 million new cases that diagnosed worldwide in 2020 representing 5.6% of all cancer cases diagnosed[1][1]. The epidemiologic trends of GC vary significantly amongst the regions. Eastern Asia (Japan, Korea, and Mongolia) and Eastern Europe had the highest incidence rates[1][1].Currently, GC diagnosis primarily depends on endoscopic and imaging techniques, such as upper gastrointestinal endoscopy and CT scans. Serum biomarkers, including carcinoembryonic antigen (CEA), carbohydrate antigen 19-9 (CA19-9), and carbohydrate antigen 72-4 (CA72-4), are widely used for auxiliary GC diagnosis, particularly in monitoring advanced cases. However, these serum markers show limited sensitivity and specificity for early-stage GC, especially in asymptomatic patients.[2][3][4] Recent advancements in gastric cancer (GC) research have significantly improved our understanding of potential biomarkers for early diagnosis, particularly those identified in bodily fluids like blood. Circulating proteins, metabolites, and microRNAs (miRNAs) have emerged as promising non-invasive biomarkers. For example, circulating proteins such as sHLA-G, TrxR, and SHBG have shown potential in distinguishing early-stage GC patients from healthy individuals[5][2]. These findings highlight the potential of liquid biomarkers for early GC detection. However, these studies have predominantly focused on blood-based samples, which, although minimally invasive, may still deter some individuals due to the discomfort associated with venipuncture. Research on salivary biomarkers provides a novel approach to early GC detection, with the potential to become a safe and efficient screening method, thereby improving early diagnosis and patient prognosis in GC. Currently, ITRAQ technology is widely utilized in diverse fields such as cancer research and clinical biomarker. As this technology matures further stills its application extends into multi-omics joint analyses that promote advancements in precision medicine initiatives.[1, 6][1][2] Studies have shown that PRM can effectively differentiate actual analytes from biochemical matrix signals, leading to significantly improved data quality. [6][2]Furthermore, previous research has successfully utilized PRM to confirm the identities of several DEPs identified in iTRAQ analysis.[7–9] [3–5]Therefore, we employed PRM to analyze 10 DEPs to validate the reliability of the iTRAQ results. This study aims to investigate salivary biomarkers for early GC detection using iTRAQ-based and PRM-based proteomics techniques to identify differentially expressed proteins in GC patients. By focusing on saliva, we hope to leverage its unique advantages to contribute to non-invasive cancer diagnostics and improve early detection strategies for GC.

2 Methods

2.1 Sample source

Saliva samples were collected from GC patients and a normal group (non-GC) at Shenzhen Guangdong China from 2019/09-2020/09 with full ethical consent. Table 1 lists the clinical sample information in this experiment. Unstimulated saliva was collected into the sterile plastic tubes, then saliva tubes were centrifuged at 10000 rpm for 10 min at -4Co and separate aliquots of supernatants were stored frozen at -80Co until analysis.

Ethical approval declarations All experimental procedures in the present study were approved (no. KS20190418003) by the ethics committee of the first affiliated hospital of shenzhen university (Shenzhen, China). Written informed consent was obtained from all human saliva donors.

2.2 Identification and quantification of the saliva proteomics by iTRAQ experiment

These samples were processed routinely for the iTRAQ experiment. According to Noto[10][28], peptides from each group were labeled with the following tags: 117 and 118 tags for GC 1, 119 and 121 tags for GC 2, 113 and 114 tags for NC, respectively. The peptides were dissolved in 0.5 M TEAB and tagged in accordance with the iTRAQ kit's instructions as follows: The peptides and the labeled reagent were combined, and

Table 1 Demographic information on subjects in this study (n =68)

		Biomarker discovery phase iTRAQ				Biomarker validation phase PI
Demographic		GC 1	GC 2	Non-GC	GC	Non-GC
variable		12	13	11	16	16
Age, years		56.35 ± 3.04	52.45 ± 4.96	55.33 ± 6.78	56.2 ± 9.10	54.8 ± 10.4
Gender	Male	8	7	5	10	8
	Female	4	5	6	6	8

Note: This is an example of table footnote. This is an example of table footnote this is an example of table footnote this is an example of table footnote this is an example of table footnote.

[1]Example for a first table footnote.

[2]Example for a second table footnote.

then the mixture was incubated for 2 hours at room temperature following desalting and vacuum-dried the labeled peptides. The raw data were processed using Proteome Discover 1.3 (Thermo Fisher Scientific, v.1.3) [11][29]. The P-value of the protein between GC and NC was calculated using the T-Test, the P¡0.05 and fold changes ¿ 1.5 or 0.67 were identified as DEPs.

2.3 Bioinformatics analysis of the DEPs

The DEPs were imported into the STRING database (https://cn.string-db.org/), and the protein-protein interaction (PPI) network map was constructed by Cytoscape (3.9.0). Functional classification of DEPs was evaluated by gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis using the ClusterProfiler package of the R language, the GO analysis includes three-term of biological processes, molecular function, and cellular components. GO terms with Bonferroni ¡0.05 and KEGG pathways with P ¡0.05 were considered statistically significant.

2.4 LC-PRM analysis

10 key DEPs were verified by PRM. Two pooled samples were prepared for PRM using peptides. The peptides were subjected to an NSI source, tandem mass spectrometry (MS/MS), in Q ExactiveTM Plus (Thermo), connected online to the UPLC, according to Xia(10). The fragments were then detected in the Orbitrap at a resolution of 17,500 after the peptides were chosen for MS/MS with the NCE setting as 27. 20 MS/MS scans were followed by one MS scan in a data-independent technique. AGC settings for complete MS and MS/MS were 3E6 and 1E5, respectively. The maximum IT for full MS and auto for MS/MS was set at 20 ms. MS/MS's isolation window was set to 2.0 m/z. Skyline (v.3.6) was used to process the generated MS data (11). Peptide settings: Trypsin Max missed cleavage was set at 2, the peptide length was set at 8–25, the maximum variable modifications were set to Carbamidomethyl on Cysine and oxidation on Methionine. 3. Precursor charges were set to 2, 3, ions to 1, 2, and ions to b, y, and p during the transition. Ion match tolerance was set to 0.02 Da, and the product ions were set from ion 3 to the final ion.

2.5 Establishment and estimation of prognostic signature

The RNA sequencing expression data were downloaded from the Cancer Genome Atlas (TCGA) dataset (https://portal.gdc.com), to compare these verified 4 DEPs between GC saliva with normal samples including differential expression analysis and patient survival analysis. Then KM curves with Log-rank test, was used to show the relationship between expression of candidate genes and disease-free survival (DFS) in GC patients. All the analysis methods and R packages were implemented in the R language. P ¡0.05 was considered statistically significant. Author Contributions: Conceptualization and methodology, Zhiyue Li; formal analysis and original draft preparation, Zhanyan Liu; samples collection, Guanbao Zhu and Yaqian Liu; review and editing, Feijuan Huang; funding acquisition, Zhengzhi Wu. All authors have read and agreed to the published version of the manuscript.

3 Results

3.1 Saliva protein identification and quantification of GC

671 proteins were obtained with one or more unique peptide segments and scores 20 in this study. 124 or 102 proteins were significantly differentially expressed in GC group 1 and 2 with P value 0.05, the fold changes 1.500 or 0.657. The proteins in GC group 1 and 2 were depicted by mapping the volcano map (Fig.1A and Fig.1B). 56 overlapping DEPs between GC group 1 and GC group 2 were identified (shown as Tab.1), 24 DEPs showed both over-expressed in GC group 1 and GC group 2 compared to non- GC group, while 32 DEPs were under-expressed in GC group 1 and GC group 2 (shown as Fig. 1C and Fig. 1D). A clustering heat map for the subset of DEPs depicts the differential expression of salivary proteins that were found changed between GC and non-GC group (Fig. 1E). These results demonstrate that the expression pattern of salivary proteins in the GC group is markedly distinct from that in the non-GC group, uncovering that GC might lead to a series of alterations in the salivary proteome. This perhaps bears a close relevance to the pathological mechanism of GC and offers significant biomarker cues for further investigations into the relationship between GC and the salivary proteome.

3.2 Functional enrichment analysis of the DEPs in GC

Bioinformatics analysis of DEPs can show potential interactions and discrepancies between proteins with reference to specific functionalities[8]. So, the DEPs were subjected to enrichment analysis of Gene Ontology (GO) classification, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. Biological process, cellular component, and molecular function are the three various observed, used by the GO annotation to explain the biological function of proteins. There were 65 terms of biological processes, 16 terms of cellular components, and 12 terms of molecular functions with p.adjust0.05 q value 0.05. the top 10 of the 3 parts were selected and shown in Fig.2A. According to the biological process classification, the majority of these proteins are utilized in defense response to bacterium, regulation of gene expression, epigenetic, protein-DNA complex assembly or organization, DNA replication-dependent chromatin assembly or organization, nucleosome assembly or organization, telomere or nucleosome organization, chromatin assembly, et al. The results of the cellular component classification revealed that the majority of these DEPs were components of nucleosome, DNA packaging complex, protein-DNA complex, secretory granule lumen, cytoplasmic vesicle lumen, et.al. These DEPs were categorized by their molecular functions into cadherin binding, protein heterodimerization activity, etc. KEGG is a collection of databases for comprehending biological processes, such as metabolic pathways., biomolecular complexes, and biochemical reactions. The biological pathways of DEPs were identified by KEGG biological pathway enrichment analysis. The 7 biological pathways with p.adjust0.05 q value 0.05 were shown in Fig.2B. We found that the genes of the DEPs mainly enriched in systemic lupus erythematosus, neutrophil extracellular trap formation, transcriptional misregulation in cancer, alcoholism, shigellosis, IL-17 signaling pathway, and salivary secretion.

Compared with normal samples, there are significantly differentially expressed proteins in the saliva of patients with gastric cancer. Through GO and KEGG analysis of the DEPs, we identified significant enrichment of these proteins in biological processes such as bacterial defense, gene expression regulation, and chromatin assembly, as well as in pathways related to immune and inflammatory responses, and transcriptional dysregulation in cancer. These findings suggest that the DEPs are potentially involved in immune responses, aberrant gene expression, chromatin remodeling, immune evasion, and inflammation in gastric cancer. This provides a molecular basis for understanding immune dysregulation and abnormal gene regulation in the gastric cancer microenvironment, supporting the potential of these proteins as candidate biomarkers for cancer diagnosis.

In order to gain a deeper insight into the interrelationships among the identified proteins, we conducted an in-depth analysis of the 56 DEPs using the STRING database. By selecting interaction pairs with a score greater than 0.4, a PPI network was constructed (Figure 3). The analysis yielded 39 key DEPs, comprising 18 significantly upregulated and 21 significantly downregulated proteins. These proteins are involved in fundamental cellular processes, including cell adhesion, growth, proliferation, and apoptosis.

3.3 Verification of DEPs in GC via PRM

The application of Another proteomic analysis known as PRM aims to verify the reliability of the differentially expressed proteins (DEPs) identified based on iTRAQ results, supporting the discovery of potential biomarkers in the saliva of gastric cancer (GC) patients. The use of PRM provides higher sensitivity and specificity, serving as an effective strategy to confirm iTRAQ results and thereby enhancing our confidence in the data. A total of 10 key proteins identified as enriched in GO analysis, KEGG pathways, or PPI networks were included in the analysis. (Fig.4)

Through PRM validation of the iTRAQ results, we found consistent expression patterns for four proteins: S100A8, S100A9, CST4, and CST5. This finding not only validates the effectiveness of PRM technology in confirming iTRAQ data but also highlights the significance of these proteins in the exploration of biomarkers for gastric cancer. These significant differences support these proteins as valuable biomarkers for GC diagnosis, particularly noting the unique downregulation of CST4 and CST5 in saliva, which contrasts with their upregulation in GC tissue or serum, indicating a unique specificity in saliva diagnostics. (Fig.5)

3.4 Establishment and estimation of the four genes' prognostic signature

To assess the prognostic impact of CST5, CST4, S100A8, and S100A9 expression on progression-free survival (PFS) in gastric cancer patients, the Kaplan-Meier (KM) survival analysis was conducted using the log-rank test, with patients stratified into high-risk and low-risk groups based on the median risk score derived from a multigene prognostic model. Figure 6 presents the KM survival curves for these four genes (P ¡ 0.05). Patients with high CST5 and CST4 expression levels in cancer tissues had significantly shorter progression-free survival (PFS), identifying these two genes as risk factors. This indicates that elevated CST5 and CST4 expression levels are associated with faster disease progression. Similarly, high expression of S100A8 and S100A9 correlated with poor PFS, further suggesting their potential as risk factors in gastric cancer prognosis. In contrast, the low-risk group exhibited slower disease progression.

4 Discussion

GC is the main cause of cancer death in China, and 5-year survival rates of GC are low. It is estimated that about 288 500 Chinese people died from GC in 2016[8][4]. Since, many cases of cancer are diagnosed at an advanced stage due to low rate of early screening rates and the limited treatment options. In addition, the GC prognosis is mainly based on invasive procedures such as upper digestive endoscopy, the traditional biomarkers for GC prognosis show less sensitivity and specificity. It is urgent to find a less invasive but more accessible screening way for the diagnosis of GC.

In recent years, saliva has been recognized as an easy, fast, low-cost, and non-invasive approach for disease diagnosis. 9 Saliva has been found to have several markers for oral and systemic disorders that are proteomic, transcriptomic, and microbiological[12][7]. Proteins in saliva have good stability and can be easily developed into biomarkers to provide significant information on the underlying pathogenetic processes of oral and systemic diseases. 10,11 Consequently, an increasing number of proteomic analyses on saliva samples from various diseases (e.g., breast cancer, oral cancer, and lung cancer) have been conducted to identify salivary biomarkers associated with cancer. [1][1] Our previous study demonstrated the significant potential of SOD2 as a potential salivary biomarker for hepatocellular carcinoma detection. [] With the advancement of technology, saliva protein biomarkers offer considerable value for early GC diagnosis.

In the present study, we aimed to find saliva proteins that changed between the GC patients and normal healthy people based on the iTRAQ and PRM quantitative and figure out appropriate candidate proteins for new biomarkers related to GC. Our study identified 671 proteins with unique peptide segments and scores 20 in this study. The results revealed 124 and 102 different proteins in GC group 1 and 2 , compared to those in the normal group. Hierarchical clustering analysis and Venn diagrams detected that 58 saliva proteins were significantly changed in GC including 24 proteins upregulated and 23 proteins downregulated. These DEPs were implicated in several biological processes associated with gene expression regulation, epigenetics, and antimicrobial humoral response. The KEGG analysis showed that several screened upregulated DEPs involved in the in systemic lupus erythematosus, neutrophil extracellular trap formation, transcriptional misregulation in cancer, alcoholism, shigellosis, IL-17 signaling pathway, and salivary secretion. Furthermore,

the use of PRM-MS targeted proteomics validation method improves the accuracy of high-throughput validation. We screened 10 key DEPs verified by PRM through bioinformatic analysis, and the results found four overlapping DEPs including S100A8, S100A9, CST4, CST5.

In the most of cancer cases, S100 proteins dysregulation occurs and typically involves in upregulation. It has been said that different carcinoma showed different S100 protein signature[13][11]. So, these proteins are promising markers for the identification and prediction of staging of human tumors [13–15][11-13]. In iTRAQ analysis, we found that the S100A7, S100A7A, S100A8, and S100A9 were both upregulated in GC group 1 and 2. The PRM analysis verified that S100A8 and S100A9 were upregulated in GC saliva. It has been reported that S100A8 and S100A9 levels may be a potential prognosticator of DFS (disease-free survival) in tumor patients, high percentage of S100A8 and S100A9 means low DFS[16][14]. The normal range of calprotectin (S100A8 /S100A9 heterodimer) in human serum is less than 1 µg/ml, which increases in many types of cancers or inflammatory diseases[17][15]. Calprotectin has been documented overexpressed in neoplastic cancer cells and many other human tumor tissues or serum, such as nasopharyngeal carcinoma, GC, and bladder cancer[18–20][16-18]. Moreover, by Bax/Bcl-2 expression ratio and inhibiting ERK activation, highly concentrated calprotectin contains cytotoxicity and apoptosis-induction on AGS cell lines, the common type of gastric adenocarcinoma cell line[19][17]. Upregulation of S100A7 protein has been predicted in cancer from several tissues (oral, esophagus, breast) or serum [21–23] [19-21] with a strong correlation to poor prognosis.Upregulated S100A7 promotes cancer cell proliferation and migration through intracellular attachment to JAB1 as well as secretion and activation of RAGE receptors[22][20]. Therefore, the expression of S100 family proteins such as the S100A8, S100A9, or S100A7 was associated with GC progression which can be verified in GC saliva.

Recent studies linking cystatins (CSTs) to cancers have drawn more and more interest. There is a CST superfamily of endogenous and reversible proteins that has an impact on controlling cysteine peptidases' excessive activity in intracellular and extracellular environments.[24][22]. Moreover, it is imperative to precisely maintain the right balance between CSTs and cysteine proteases because it is thought that their breakdown can result in the development of malignancies (27). A possible method for early identification of gastrointestinal cancer in patients could be serum CST4 detection [25][23]. As a novel serum marker for gastrointestinal cancer, the positive detection rate of CST4 for gastrointestinal cancer is much higher than that of traditional markers like CEA, CA199, CA125, and CA724, showing great superiority in sensitivity [26][24]. CST4 is markedly upregulated in GC tissues, serum, or cells, which is connected with poor prognosis (OS) and progression-free survival (PFS) [26, 27][24, 25]. CST4 overexpression promotes invasion and migration abilities of the GC cell lines MKN-45 and SGC-7901 in vitro and pulmonary metastasis in vivo, whereas silencing endogenous CST4 causes an opposite outcome[27][25]. CST5 is a proposed tumor suppressor induced by the p53 or vitamin D3 pathway in colorectal cancer (CRC), which suppresses tumor progression and metastasis[28, 29][26, 27]. CST5 is identified as a significant mediator of tumor suppression by mediating mesenchymal-epithelial transition (MET) in CRC cells.[28][26] CST5 represses expression of the EMT inducers SNAI1, SNAI2, ZEB1, and ZEB2 respectively, induces expression of E-cadherin and other adhesion proteins[29][27]. Furthermore, CST5 restricted migration and anchorage-independent growth, antagonized the Wnt/-catenin signaling pathway, and suppressed c-MYC expression[29][27]. As a type of secreted protein, CST4 is secreted by the salivary gland, and lacrimal gland, but the expression in blood is lower. Gastric and intestinal tumor cells secrete CST4, which is transported to blood, so detection of serum CST4 has been well-defined to be conducive to diagnose some malignancies, mainly gastrointestinal tumors. Contrary to GC tissues or serum, we found the expression of CST4 was reduced in GC saliva via iTRAQ and PRM, which was a completely new discovery. The reason may be that CST4 is lower secreted in salivary gland or metabolized by microbes in the oral cavity of GC. Our research will focus on revealing the mechanism of lower expression of CST4 in GC saliva and continue to explore and verify the application of saliva CST4 in non-invasive diagnosis of GC.

Currently, it is important to identify new biomarkers of cancer and it has become major target for cancer studies. These biomarkers are useful in diagnosis, monitoring and therapeutic efficiency. In this study, we used iTRAQ and PRM-based quantitative proteomics to detect four saliva protein biomarkers of gastric including S100A8, S100A9, NUCB2, and CST4 which makes them potential novel biomarkers for the noninvasive diagnosis of GC. These identified biomarkers can be definitively validated for GC detection and it is a promising approach for screening of GC patients and reducing endoscopies. Our study enhances the aspect of salivary diagnostics in the finding of other systemic diseases. The novelity of our study is improvement in

detection techniques. Although, we faced some difficulties and limitations such as small sample size, however, more research is required to inight the diagnostic value of these biomarkers.

5 Conclusion

Gastric cancer (GC) is the major global health issue. Early diagnosis is crucial for effective treatment, and the identification of new biomarkers of GC is necessary to facilitate rapid diagnosis and proactive treatment. Our results show that the levels of S100A8, S100A9 were upregulated and CST4, and CST5 downregulated in the saliva of patients with GC compared with those of healthy controls. These findingsenhances the reliability and clinical applicability of salivary biomarkers, paving the way for future studies focused on non-invasive diagnostic strategies for gastric cancer. Based on these results, we hope to collect more patient samples and prove it through additional verification in order to apply it to clinical stages.

Supplementary information.

Acknowledgements.

Declarations

- Funding
  This work was financially supported by Guangdong TCM Strong Provincial science and technology special key project (20215002), and the National Natural Science Foundation of China (81804145), Shenzhen Science and Technology Innovation Commission (JSGG20220919091404008, JSGG20210901145539016, JCYJ20190812171005713, JCYJ20210324121611032, JCYJ20180227175929767, ZDSYS20190902093401689), Shenzhen Baoan Chinese Medicine Development Foundation (2020KJCX-KTYJ-215), GuangDong Basic and Applied Basic Research Foundation (2022A1515011777, 2021A1515012474). We are grateful for —for their assistance in the project. We would like to thank all the participants for contributing to the project.
- Conflict of interest/Competing interests (check journal-specific guidelines for which heading to use)
  There are no conflicts of interest to declare.
- Ethics approval and consent to participate
- Consent for publication
- Data availability
  All study data can be viewed in the manuscript.
- Materials availability
- Code availability
- Author contribution
  Conceptualization and methodology, Zhiyue Li; formal analysis and original draft preparation, Zhanyan Liu; samples collection, Guanbao Zhu and Yaqian Liu; review and editing, Feijuan Huang; funding acquisition, Zhengzhi Wu; introduction, result and discussion revision, Jieren Liu, Runtao Wen, Haoran Chi, Jimao Mo, Lei Huang, Guanlin Li, Kaixin Luo. All authors have read and agreed to the published version of the manuscript.

If any of the sections are not relevant to your manuscript, please include the heading and write 'Not applicable' for that section.

Appendix A Section title of first appendix

An appendix contains supplementary information that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem or it is information that is too cumbersome to be included in the body of the paper.

References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–249.

[2] Sun J, Li X, Wang Q, Chen P, Zhao L, Gao Y. Proteomic profiling and biomarker discovery for predicting the response to PD-1 inhibitor immunotherapy in gastric cancer patients. Frontiers in Pharmacology. 2024;15:1349459.

[3] Liu N. The Role of Inflammatory Biomarkers in Predicting the Prognosis of Gastric Cancer. Advances in Clinical Medicine. 2023;13:19890.

[4] Zúñiga-Pérez N, Muñoz-Navarro H, Rivera C. Salivary biomarkers for the diagnosis of gastric cancer: a systematic review. Revista Científica Odontológica. 2024;12(2):e199.

[5] Repetto O, Vettori R, Steffan A, Cannizzaro R, De Re V. Circulating proteins as diagnostic markers in gastric cancer. International Journal of Molecular Sciences. 2023;24(23):16931.

[6] Zong L, Abe M, Seto Y, Ji J. The challenge of screening for early gastric cancer in China. The Lancet. 2016;388(10060).

[7] Thrift AP, El-Serag HB. Burden of Gastric Cancer. Clinical Gastroenterology and Hepatology. 2020;18(3):534–542.

[8] Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in China, 2016. Journal of the National Cancer Center. 2022;2(1):1–9.

[9] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin. 2016;66(2):115–32.

[10] Noto JM, Rose KL, Hachey AaJ, Delgado AG, Romero-Gallo J, Wroblewski LE, et al. Carcinogenic Helicobacter pylori strains selectively dysregulate the in vivo gastric proteome, which may be associated with stomach cancer progression. Molecular & Cellular Proteomics. 2019;18(2):352–371.

[11] Chen W, Jiang Q, Yan G, Yang D. The oral microbiome and salivary proteins influence caries in children aged 6 to 8 years. BMC Oral Health. 2020;20:1–16.

[12] Kaczor-Urbanowicz KE, Wei F, Rao SL, Kim J, Shin H, Cheng J, et al. Clinical validity of saliva and novel technology for cancer detection. Biochim Biophys Acta Rev Cancer. 2019;1872(1):49–59.

[13] Bresnick AR, Weber DJ, Zimmer DB. S100 proteins in cancer. Nat Rev Cancer. 2015;15(2):96–109.

[14] Emberley ED, Murphy LC, Watson PH. S100 proteins and their influence on pro-survival pathways in cancer. Biochem Cell Biol. 2004;82(4):508–515.

[15] Allgower C, Kretz AL, von Karstedt S, Wittau M, Henne-Bruns D, Lemke J. Friend or Foe: S100 Proteins in Cancer. Cancers (Basel). 2020;12(8).

[16] Koh HM, Lee HJ, Kim DC. High expression of S100A8 and S100A9 is associated with poor disease-free survival in patients with cancer: a systematic review and meta-analysis. Transl Cancer Res. 2021;10(7):3225–3235.

[17] Shabani F, Farasat A, Mahdavi M, Gheibi N. Calprotectin (S100A8/S100A9): a key protein between inflammation and cancer. Inflamm Res. 2018;67(10):801–812.

[18] Xu N, Zhang BB, Huang XN, Yi X, Yan XM, Cai Y, et al. S100A8/A9 Molecular Complexes Promote Cancer Migration and Invasion via the p38 MAPK Pathway in Nasopharyngeal Carcinoma. Bioinorg Chem Appl. 2021;2021:9913794.

[19] Shabani F, Mahdavi M, Imani M, Hosseinpour-Feizi MA, Gheibi N. Calprotectin (S100A8/S100A9)-induced cytotoxicity and apoptosis in human gastric cancer AGS cells: Alteration in expression levels of Bax, Bcl-2, and ERK2. Hum Exp Toxicol. 2020;39(8):1031–1045.

[20] Yasar O, Akcay T, Obek C, Turegun FA. Significance of S100A8, S100A9 and calprotectin levels in bladder cancer. Sc and J Clin Lab Invest. 2017;77(6):437–441.

[21] Dickinson A, Saraswat M, Mäkitie A, Silén R, Hagström J, Haglund C, et al. Label-free tissue proteomics can classify oral squamous cell carcinoma from healthy tissue in a stage-specific manner. Oral Oncol. 2018;86:206–215.

[22] Lu Z, Zheng S, Liu C, Wang X, Zhang G, Wang F, et al. S100A7 as a potential diagnostic and prognostic biomarker of esophageal squamous cell carcinoma promotes M2 macrophage infiltration and angiogenesis. Clin Transl Med. 2021;11(7):e459.

[23] Mayama A, Takagi K, Suzuki H, Sato A, Onodera Y, Miki Y, et al. OLFM4, LY6D and S100A7 as potent markers for distant metastasis in estrogen receptor-positive breast carcinoma. Cancer Sci. 2018;109(10):3350–3359.

[24] Breznik B, Mitrovic A, T TL, Kos J. Cystatins in cancer progression: More than just cathepsin inhibitors. Biochimie. 2019;166:233–250.

[25] Cai L, Tu M, Yin X, Zhang S, Zhuang W, Xia Y, et al. Combination of serum CST4 and DR-70 contributes to early diagnosis of colorectal cancer. Clin Chim Acta. 2022;531:318–324.

[26] Dou Y, Lv Y, Zhou X, He L, Liu L, Li P, et al. Antibody-s andwich ELISA analysis of a novel blood biomarker of CST4 in gastrointestinal cancers. Onco Targets Ther. 2018;11:1743–1756.

[27] Zhang YQ, Zhang JJ, Song HJ, Li DW. Overexpression of CST4 promotes gastric cancer aggressiveness by activating the ELFN2 signaling pathway. Am J Cancer Res. 2017;7(11):2290–2304.

[28] Hünten S, Hermeking H. p53 directly activates cystatin D/CST5 to mediate mesenchymal-epithelial transition: a possible link to tumor suppression by vitamin D3. Oncotarget. 2015;6(18):15842.

[29] Valle N, García JM, Peña C, Freije JM, Quesada V, Astudillo A, et al. Cystatin D is a c andidate tumor suppressor gene induced by vitamin D in human colon cancer cells. The Journal of clinical investigation. 2009;119(8):2343–2358.

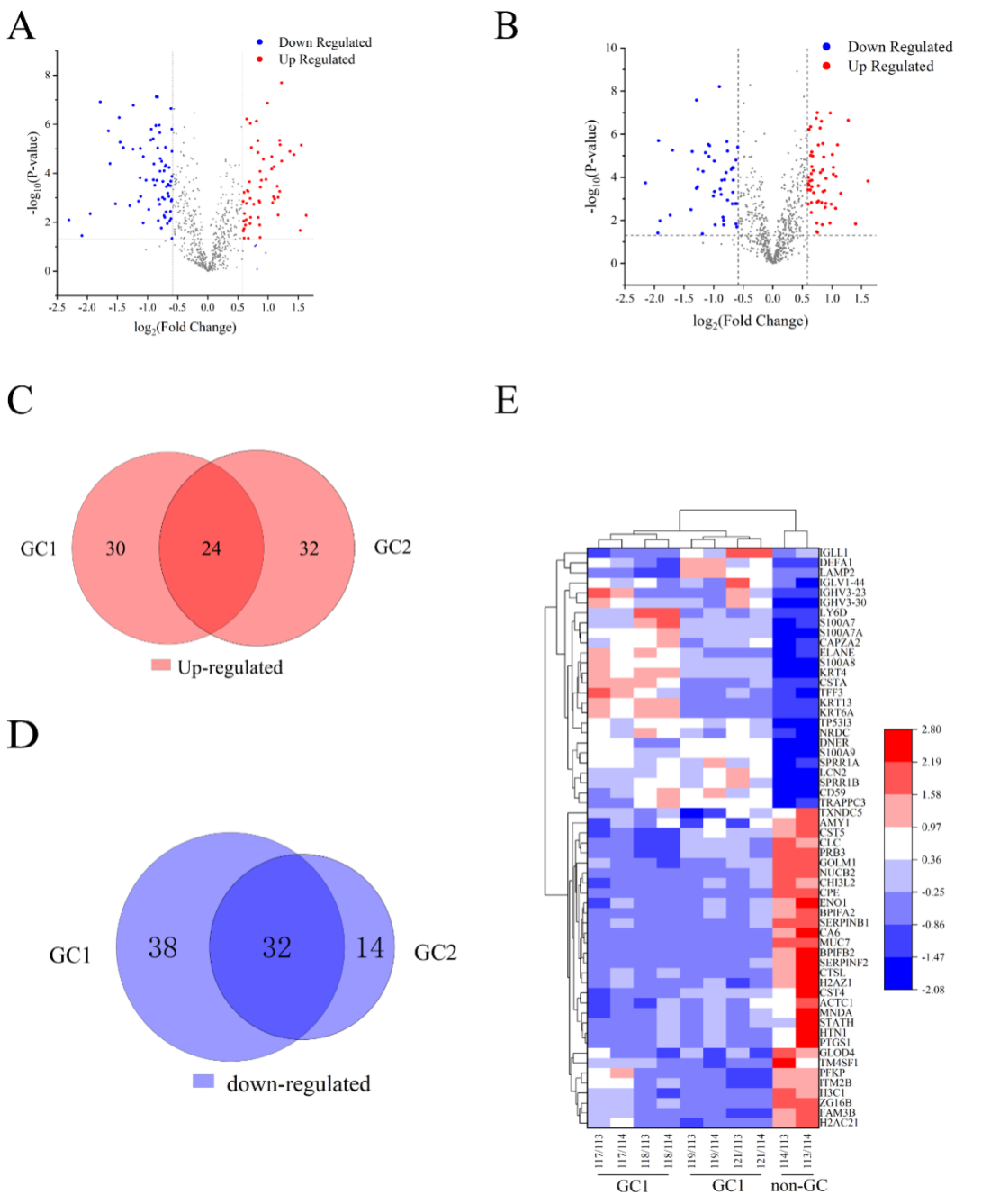

Fig. 1 Results from quantitative proteomic analysis screened from iTRAQ.(A) Volcano plot of DEPs in GC 1 group. (B) Volcano plot of DEPs in GC 2 group. (C) Venn diagrams of upregulated DEPs between GC 1 and GC 2. There were 24 overlapping DEPs between the 2 groups. (D) Venn diagrams of downregulated DEPs between GC 1 and GC 2. (E) Cluster analysis of DEPs in GC 1, GC 2, and non-GC groups. Red nodes represented the significantly up-regulated proteins with FC ¿1.5 and P ¡ 0.05. Blue nodes represented the significantly down-regulated proteins with FC ¡ 0.67 and P ¡ 0.05. The grey nodes represent non-differentiated proteins.

Table 2 Overlapping differentially expressed proteins profile

Uniprot ID	Protein Names	Score	Coverage	Unique Peptides	GC1 FC	GC1 -Log10P	GC2 FC	GC2 -Log10P
P15515	HTN1	54.380	28.070	1.000	0.203	2.091	0.267	1.980
P02808	STATH	47.310	48.390	1.000	0.235	1.446	0.260	1.409
P23219	PTGS1	25.180	2.340	1.000	0.259	2.345	0.300	2.238
Q04118	PRB3	395.670	3.240	1.000	0.291	6.914	0.504	4.750
Q8TAX7	MUC7	40.930	9.810	5.000	0.319	5.728	0.309	5.257
Q05315	CLC	45.780	25.350	3.000	0.325	4.392	0.470	5.513
P68032	ACTC1	721.310	38.200	3.000	0.346	2.749	0.561	2.002
P28325	CST5	243.860	49.300	5.000	0.361	6.273	0.626	4.451
Q96DR5	BPIFA2	234.630	47.790	12.000	0.365	5.264	0.442	4.272
P23280	CA6	1667.150	64.940	13.000	0.379	5.046	0.389	5.190
P01036	CST4	7066.280	68.790	5.000	0.408	2.673	0.656	1.691
P80303	NUCB2	181.320	32.380	12.000	0.425	6.775	0.478	5.460
P0C0S5	H2AZ1	58.670	12.500	1.000	0.459	3.813	0.415	3.558
P07711	CTSL	52.770	16.520	4.000	0.463	3.050	0.409	3.490
P41218	MNDA	29.050	4.180	1.000	0.476	1.965	0.562	1.790
P08697	SERPINF2	47.710	4.280	2.000	0.492	3.712	0.540	3.450
P04745	AMY1	38226.830	75.930	6.000	0.508	2.519	0.558	2.146
Q8N4F0	BPIFB2	450.570	29.040	10.000	0.538	4.377	0.579	4.215
Q96DA0	ZG16B	1072.930	48.080	7.000	0.551	5.945	0.409	7.574
P16870	CPE	42.110	11.340	5.000	0.562	7.112	0.535	8.202
Q15782	CHI3L2	38.390	28.460	8.000	0.563	5.034	0.615	4.361
Q8IUE6	H2AC21	20.210	12.310	1.000	0.576	2.731	0.416	4.355
P30408	TM4SF1	69.450	4.460	1.000	0.586	1.763	0.384	2.500
P06733	ENO1	2199.190	63.590	17.000	0.605	4.096	0.630	3.865
Q9Y287	ITM2B	35.470	4.890	1.000	0.605	2.267	0.225	3.737
P58499	FAM3B	33.210	23.830	5.000	0.613	4.271	0.454	5.139
Q8NBJ4	GOLM1	90.060	22.690	9.000	0.616	4.328	0.630	4.427
Q8NBS9	TXNDC5	27.500	4.170	2.000	0.633	3.184	0.508	1.791
Q9HC38	GLOD4	28.200	20.770	7.000	0.652	2.437	0.622	2.763
P68431	H3C1	43.340	22.060	4.000	0.661	2.922	0.589	5.211
Q01813	PFKP	36.910	3.570	1.000	0.662	1.335	0.262	5.694
P30740	SERPINB1	878.060	51.720	20.000	0.663	4.894	0.663	5.400
Q8NFT8	DNER	58.450	8.410	5.000	1.504	2.849	1.665	6.735
P01699	IGLV1-44	32.070	14.410	2.000	1.519	1.347	1.673	1.470
O43617	TRAPPC3	20.550	4.440	1.000	1.542	1.879	1.685	6.997
P59665	DEFA1	186.450	20.210	4.000	1.594	1.343	2.215	3.253
P80188	LCN2	197.050	58.080	10.000	1.653	5.050	1.790	4.333
P35321	SPRR1A	70.530	87.640	2.000	1.784	4.842	1.802	3.334
P13987	CD59	47.200	25.000	3.000	1.818	2.814	2.032	4.141
P01764	IGHV3-23	175.780	41.030	2.000	1.827	1.380	1.674	1.871
P01773	IGHV3-30	28.830	15.970	2.000	1.832	3.427	1.774	2.831
P47755	CAPZA2	60.380	6.590	3.000	1.855	3.712	1.705	3.604
P22528	SPRR1B	42.940	70.790	1.000	1.957	3.746	2.098	4.044
P19013	KRT4	1157.030	55.990	29.000	1.988	6.867	1.558	6.346
O43847	NRDC	30.570	0.960	1.000	2.085	4.165	1.836	3.381
Q53FA7	TP53I3	33.670	6.020	2.000	2.090	4.682	2.032	4.459
P31151	S100A7	310.280	68.320	5.000	2.110	2.796	1.745	4.218
Q07654	TFF3	129.550	22.500	2.000	2.225	3.473	1.553	3.404
P06702	S100A9	3348.350	82.460	9.000	2.289	5.339	2.418	6.639
P08246	ELANE	204.640	36.330	6.000	2.299	5.164	1.613	3.283
P01040	CSTA	141.150	75.510	6.000	2.340	7.692	1.595	4.979
P13646	KRT13	1133.620	50.220	17.000	2.343	4.497	1.529	6.213
P05109	S100A8	3608.280	94.620	14.000	2.573	4.888	1.962	6.975
Q86SG5	S100A7A	207.240	47.520	1.000	2.698	4.758	2.135	5.500
P02538	KRT6A	1166.900	44.150	1.000	2.936	5.146	1.519	3.991
Q14210	LY6D	53.140	23.440	3.000	3.104	2.277	1.858	2.794

Source: There are 56 overlapping DEPs between GC group 1 and GC group 2.

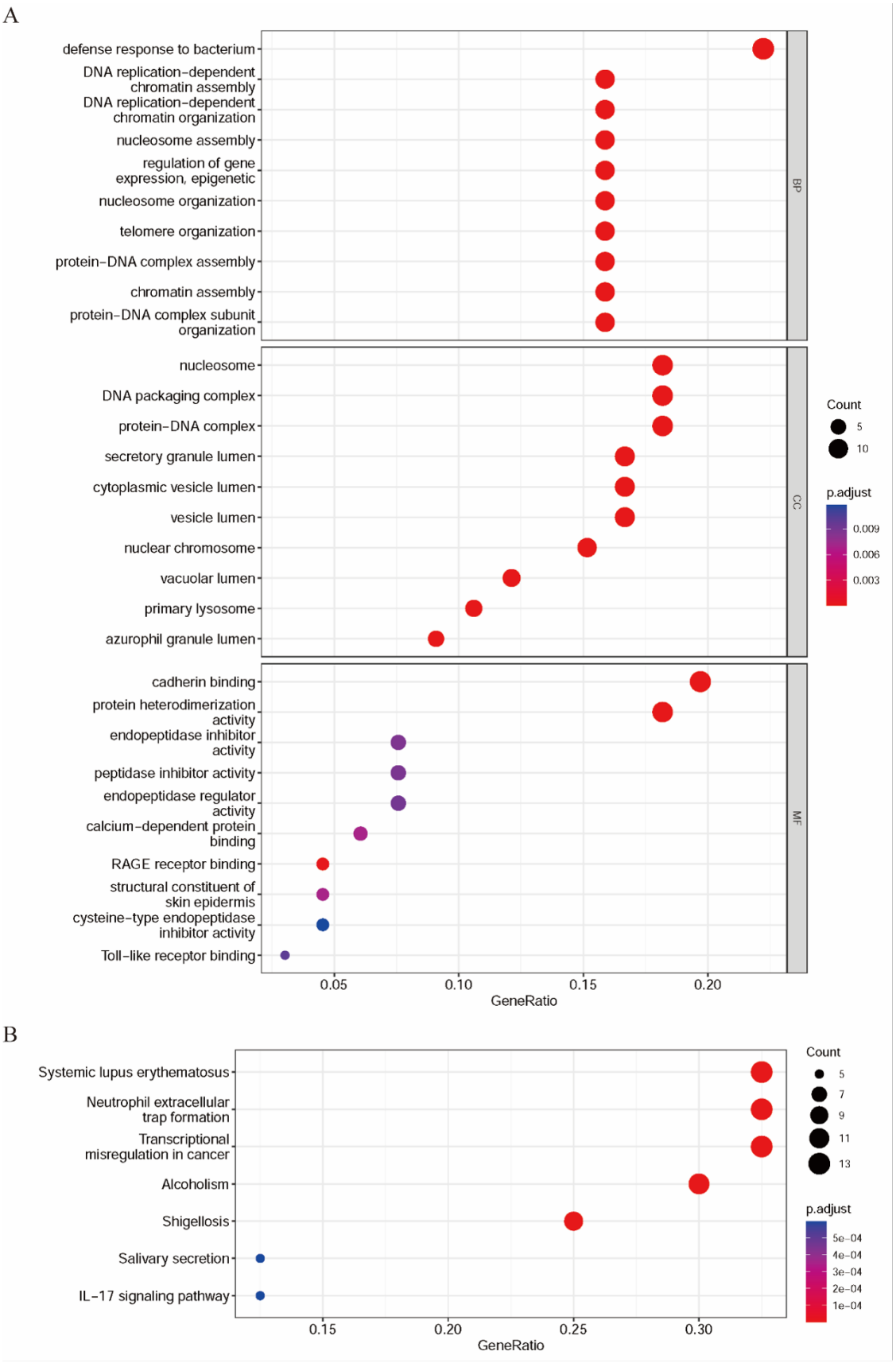

Fig. 2 Functional enrichment analysis and PPI of the DEPs in GC. (A) GO analysis of DEPs, (B) KEGG pathway of DEPs.

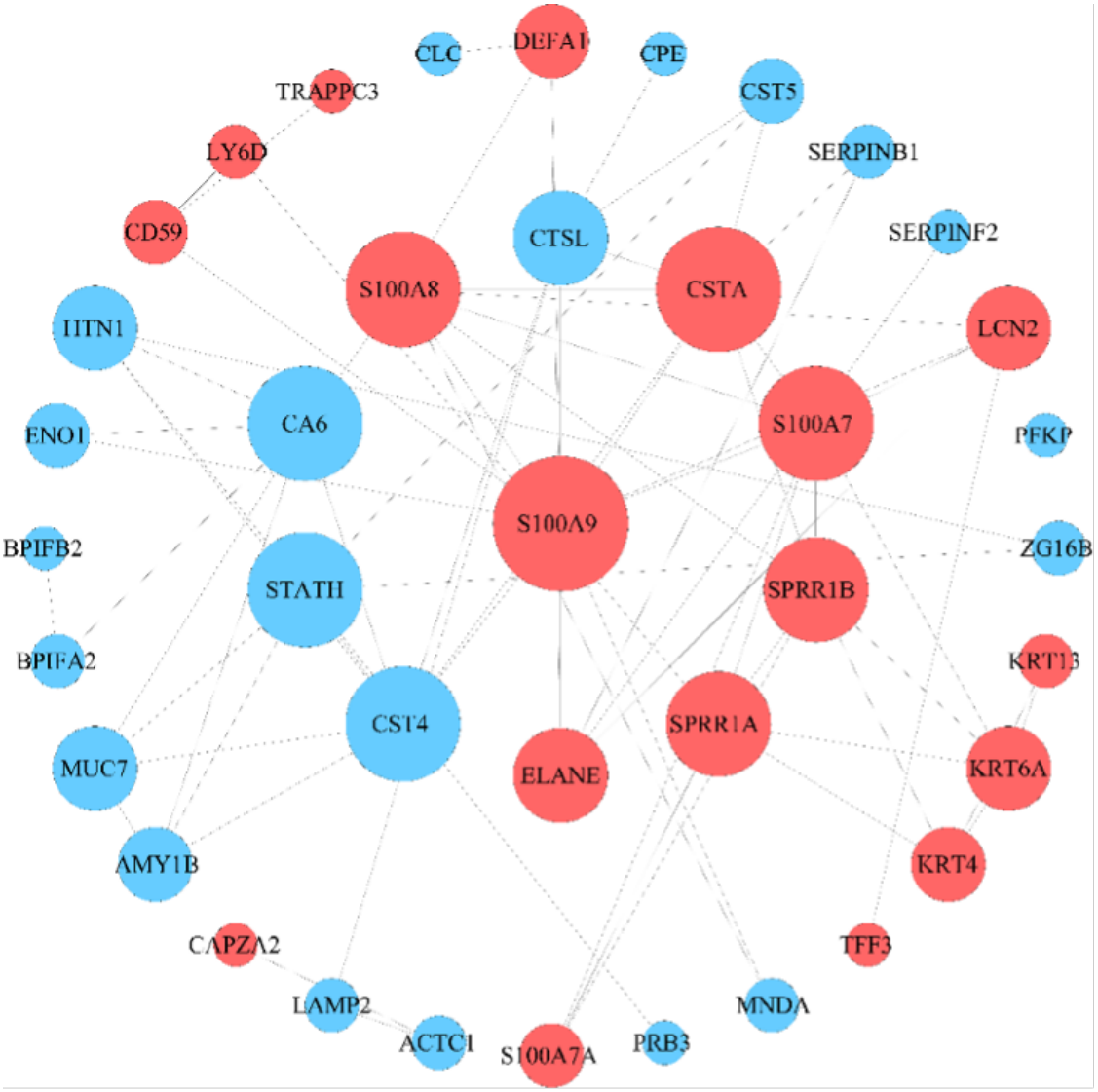

Fig. 3 PPI network of DEPs. The size of nodes was proportional to degree of centrality by topology analysis. Upregulated proteins were marked in red, downregulated proteins were marked in blue. The color and size of nodes are determined by the P value or the number of associated genes. The color, which ranges from blue to red, is indicative of the value of P, with blue indicating a lower value and red indicating a higher value. The size of the nodes, which ranges from small to large, is indicative of the number of associated genes. Smaller sizes are indicative of a lower number of genes, while larger sizes are indicative of a higher number of genes.

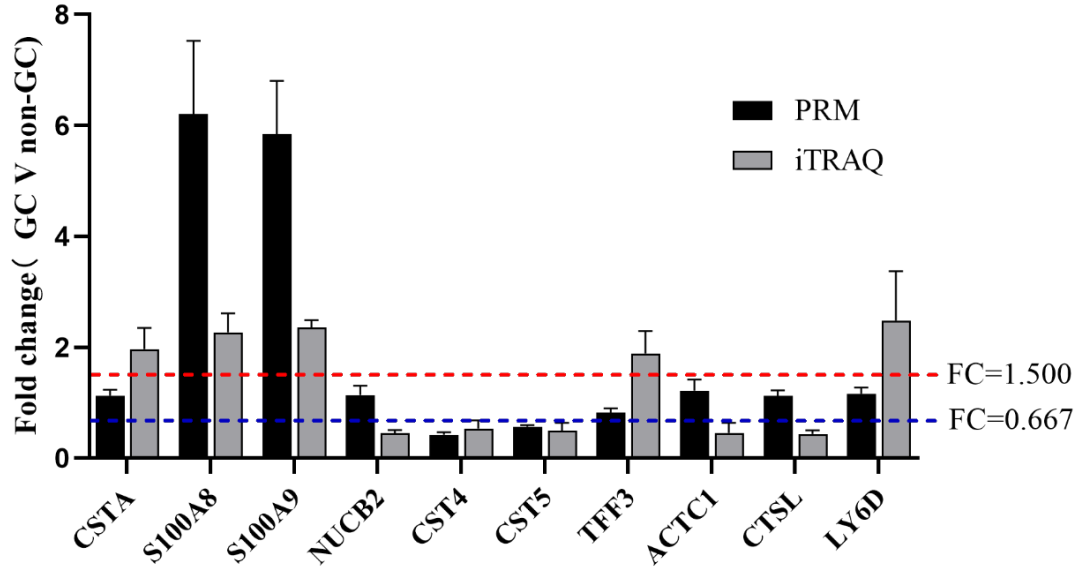

Fig. 4 Comparison between iTRAQ-based results and PRM-based results.

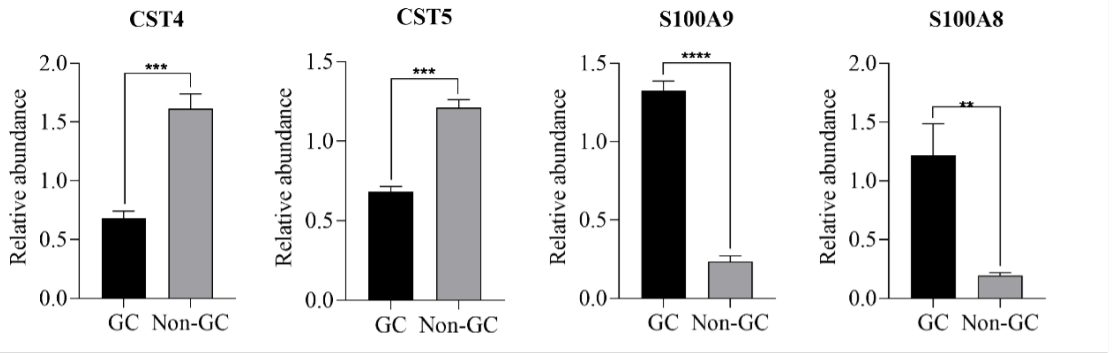

Fig. 5 The comparison of protein expression in the GC group and non-GC group by PRM. (*P0.05, **P0.01,***P0.001, **** P0.0001)

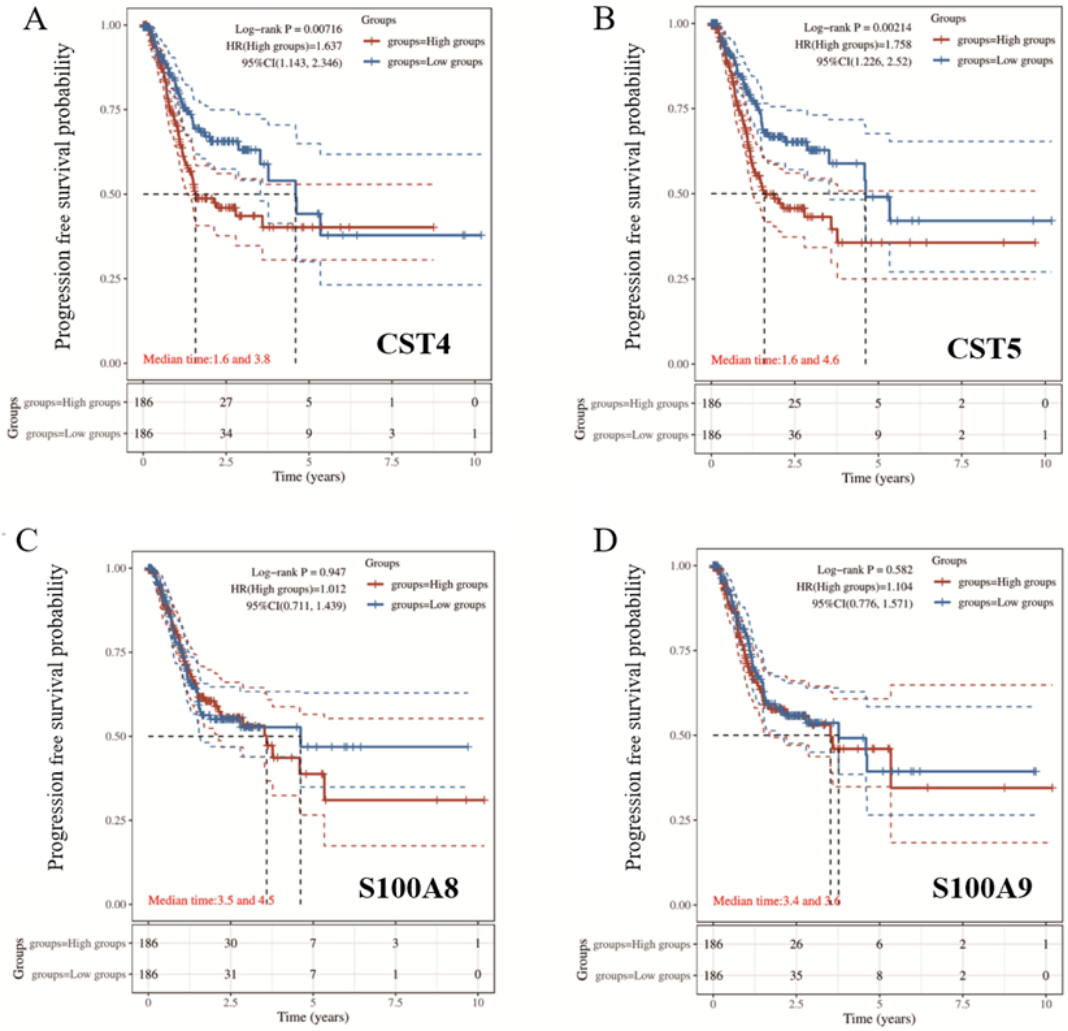

Fig. 6 Kaplan–Meier survival analysis of CST4, CST5, S100A8, and S100A9. A hazard ratio (HR) greater than 1 indicates that a gene serves as a risk factor associated with reduced PFS, whereas an HR less than 1 suggests a protective factor, linked with better PFS outcomes.