Journal of the National Cancer Center 3 (2023) 286–294

Contents lists available at ScienceDirect

Journal of the National Cancer Center

journal homepage: www.elsevier.com/locate/jncc

Full Length Article

Identification of plasma proteomic signatures associated with the progression of cardia gastric cancer and precancerous lesions

Jianhua Gu[1,2,†], Shuanghua Xie[1,3,†], Xinqing Li[1], Zeming Wu[4], Liyan Xue[5], Shaoming Wang[1,*], Wenqiang Wei[1,*]

[1] *Office of National Central Cancer Registry, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China*
[2] *Department of Emergency Medicine, Qilu Hospital of Shandong University, Jinan, China*
[3] *Department of Central Laboratory, Beijing Obstetrics and Gynecology Hospital, Capital Medical University. Beijing Maternal and Child Health Care Hospital, Beijing, China*
[4] *iPhenome Biotechnology (Dalian), Inc., Dalian, China*
[5] *Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China*

ARTICLE INFO

Keywords:
Cardia gastric cancer
Proteomics
Screening
Precancerous lesion

ABSTRACT

Objective: Considering that there are no effective biomarkers for the screening of cardia gastric cancer (CGC), we developed a noninvasive diagnostic approach, employing data-independent acquisition (DIA) proteomics to identify candidate protein markers.

Methods: Plasma samples were obtained from 40 subjects, 10 each for CGC, cardia high-grade dysplasia (CHGD), cardia low-grade dysplasia (CLGD), and healthy controls. Proteomic profiles were obtained through liquid chromatography-mass spectrometry (LC-MS/MS-based DIA proteomics. Candidate plasma proteins were identified by weighted gene co-expression network analysis (WGCNA) combined with machine learning and further validated by the Human Protein Atlas (HPA) database. The area under the receiver operating characteristic curve (AUC) was used to evaluate the performance of the biomarker panel.

Results: There was a clear distinction in proteomic features among CGC, CHGD, CLGD, and the healthy controls. According to the WGCNA, we found 42 positively associated and 164 inversely associated proteins related to CGC progression and demonstrated several canonical cancer-associated pathways. Combined with the results from random forests, LASSO regression, and immunohistochemical results from the HPA database, we identified three candidate proteins (GSTP1, CSRP1, and LY6G6F) that could together distinguish CLGD (AUC = 0.91), CHGD (AUC = 0.99) and CGC (AUC = 0.98) from healthy controls with excellent accuracy.

Conclusions: The panel of protein biomarkers showed promising diagnostic potential for CGC and precancerous lesions. Further validation and a larger-scale study are warranted to assess its potential clinical applications, suggesting a potential avenue for CGC prevention in the future.

1. Introduction

Cardia gastric cancer (CGC), arising from the upper part of the stomach adjoining the esophagus, is a subtype of gastric cancer (GC) with a high incidence in East Asia.[1] In China, the CGC was defined as a carcinoma with its center located within the region 1 cm above and 2 cm below the gastroesophageal junction (GEJ), corresponding to Siewert type II adenocarcinoma of the GEJ.[2] Due to a lack of specific symptoms, CGC is often diagnosed at late stages, with a subsequent poor prognosis for most patients.[3]

Endoscopy and biopsy have become the dominant modality for population-based CGC screening programs in China. Despite the great achievements in reducing mortality,[4] the high cost and invasive nature curtail their large-scale application. Blood-based biomarkers are of particular interest as cancer screening tests as a result of their convenience, low cost, and quantitative measure. In countries with a high incidence

* Corresponding authors.
E-mail addresses: wangshaoming@cicams.ac.cn (S. Wang), weiwq@cicams.ac.cn (W. Wei).
† These authors contributed equally to this work.

https://doi.org/10.1016/j.jncc.2023.10.003
Received 19 July 2023; Received in revised form 11 October 2023; Accepted 17 October 2023

of GC such as China,[5] Japan[6] and South Korea,[7] serum pepsinogen, gastrin-17 and *Helicobacter pylori* IgG antibodies have been used for risk stratification in screening programs of non-cardia gastric cancer (NCGC) at the population level. However, these strategies have been proven effective in the screening of NCGC, while having poor diagnostic effectiveness in CGC.[5] Therefore, robust detection biomarkers for CGC are desperately needed.

The development of high-throughput proteomic techniques based on liquid chromatography-mass spectrometry (LC-MS) provides a powerful platform for the exploration of biomarkers.[8] However, due to the bias toward high-abundance ions and limited dynamic range, the conventional data-dependent acquisition (DDA)-MS model has limitations in the quantification and identification of low-abundance proteins. Compared to this, the emerging strategy of data-independent acquisition (DIA) combines the protein-wide coverage of DDA with the reproducibility, sensitivity, and accuracy of targeted methods.[9,10] Motivated by the abundant unique advantages in the acquisition of MS data, DIA proteomics has been widely applied for the exploration of pathogenesis and potential biomarkers in sorts of cancers such as colorectal,[11] prostate[12] and lung cancer.[13]

In this study, we applied LC-MS/MS-based DIA proteomics to analyze plasma samples from CGC, cardia high-grade dysplasia (CHGD), cardia low-grade dysplasia (CLGD), and healthy controls, aiming to identify a set of potential plasma protein biomarkers for CGC early diagnosis.

2. Materials and methods

2.1. Study population and sample collection

CGC patients were recruited and treated at the Linzhou Cancer Hospital, Henan Province. Information related to pathologic diagnosis and tumor stage was obtained from the medical records. This study coded CGC as C16.0 and defined it as a carcinoma with its center located within 2 cm of the GEJ.[14] CHGD, CLGD, and healthy groups were recruited from residents who underwent endoscopic examination with subsequent pathological diagnoses in the same area. Details including inclusion and exclusion criteria have been reported previously.[14,15] Groups were matched 1:1 for age (±3 years), sex, and ethnicity (n = 10 per group). Plasma samples were obtained from peripheral blood by centrifugation at 3 000 g for 15 min at 4 °C and immediately stored at −80 °C until used.

2.2. Sample preparation

2.2.1. Sample preparation for spectral library construction

Total proteins from each plasma sample were extracted using the protein extraction kit (Beyotime Biotechnology, China) according to the manufacturer's instructions. In brief, the protein extraction buffer containing protease inhibitors (1:100) was added to the samples at a ratio of 1:2 and centrifuged at 15 000 g for 15 min twice at 4 °C to collect the supernatant. Protein concentrations were determined by BCA protein assay (Beyotime Biotechnology, China).

Equal amounts (5 ul per sample) of 40 plasma samples were pooled together for spectral library construction. The albumin, IgA, IgG, α1-antitrypsin, and other 14 interfering high-abundance proteins in pooled samples were removed through the Human 14 Multiple Affinity Removal LC Column (Agilent, USA).[16] Subsequently, the proteins were reduced to a 20 mM final concentration of 20 mM dithiothreitol (DTT) at 37 °C for 60 min. Alkylation was performed in the dark for 30 min by adding 1 M iodoacetamide (IAA) to a final concentration of 50 mM. Afterward, trypsin digestion was carried out in an enzyme-to-protein ratio of 1:50, overnight and in the dark at 37 °C. The digested peptides were desalted using a C18 Cartridge (Thermo, USA) and concentrated by vacuum centrifugation. An aliquot of 100 μg peptide samples was fractionated using high pH reversed-phase fraction chromatography (HPRP) followed by collecting 10 fractions. Finally, the peptides from each fraction were reconstituted in 0.1% formic acid and spiked with the indexed Retention Time standard (Biognosys, Switzerland) before LC-MS/MS.

2.2.2. Sample preparation for DIA analysis

Sample preparation for DIA analysis was identical to the one described above for the spectral library construction. Following completion of the protein extraction from 40 plasma, each protein sample was denatured, reduced with DTT, alkylated with IAA, and digested with trypsin sequentially, and then the DIA–MS was performed (Supplementary Fig. 1A).

2.3. LC-MS/MS analysis

LC-MS/MS analysis was performed on an EASY-nLC 1200 coupled to a Q-Exactive HF mass spectrometer (all from Thermo Scientific, USA). Detailed parameters for liquid chromatography and mass spectrometry procedures are provided in Supplementary materials.

2.4. Spectral library generation

DDA files were analyzed by MaxQuant software (version 1.5.3.17) and searched against the human-specific SwissProt-reviewed database (version 2019.12). After the completion of database retrieval, a spectral library was created from the MaxQuant output of the DDA runs using Spectronaut Pulsar X software (version 12.0.20491.4). More detailed parameter settings were displayed in Supplementary materials. Protein quantities were presented as relative quantification on the $\log_2$ scale of normalized protein expression (NPX) values.

2.5. Statistical analysis and bioinformatic analysis

The workflow of data processing and analysis is shown in Supplementary Fig. 1B.

2.5.1. Exploratory differential expression data analysis

The unsupervised dimensionality reduction of the multidimensional dataset was first performed employing principal component analysis (PCA) using SIMCA (version 14.1). As a complement, we also performed a supervised analysis using partial least squares discriminant analysis (PLS-DA). The up- or down-regulated proteins were defined as proteins with $\log_2$ fold changes ≥ 1 or ≤ -1 at FDR < 0.05.

2.5.2. Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) was conducted by the R (R version 4.0.2) package "WGCNA" to find a correlation between modules and disease courses. Pearson correlation coefficient was calculated for peptide abundance expression in the 40 samples, and an appropriate soft threshold was selected to ensure a scale-free co-expression network. The adjacency matrix was generated using the selected β and transformed into a Topological Overlap Matrix (TOM). Next, this TOM was hierarchically clustered, and the proteins with highly similar co-expression relationships were grouped and visualized by the dendrogram. Modules were identified using dynamic tree cutting and labeled in different colors. Whether a module was positively or negatively associated with disease courses was estimated by the sign of the Spearman correlation coefficient between principal components of individual modules and individuals sequenced by disease severity.

2.5.3. Gene ontology and pathway analysis

Gene ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Wikipathway cancer enrichment analysis of differential expression genes (DEGs) in the above key modules were performed by Metascape (http://metascape.org). The DEGs were assigned to functional groups based on Molecular Functions, Biological Processes, and Cellular Compartments.

2.5.4. Selection of biomarkers

We employed a random forest (RF) model to construct a classifier for discriminating among different groups (R package "randomForest") and the top 200 proteins were selected according to the mean decrease accuracy. The intersection between the WGCNA and RF was analyzed through Venn Diagram. The least absolute shrinkage and selection operator (LASSO) regression was implemented for further variable selection and shrinkage (R package "glmnet"). In addition, the Human Protein Atlas (HPA) (http://www.proteinatlas.org/) was used to confirm the expression of identified proteins. The receiver operating characteristics (ROC) curve from RF models with the area under the curve (AUC) was used to assess the performance of the final selected protein markers. To produce a smooth ROC curve, 100 cross-validations were performed and the results were averaged to generate the plot.

3. Results

3.1. Protein identification

The baseline characteristics of the groups were well-matched for sex and age (Supplementary Table 1). We performed DIA with targeted data extraction using sample-specific spectral libraries generated from the corresponding DDA run. Herein, an average of 757 proteins per sample were recovered and a total of 826 proteins were identified.

3.2. Proteins differently expressed between four groups

The results of PCA revealed that there was substantial variability in the distribution of protein abundance related to disease progression (Fig. 1A), and PLS-DA further supported that samples primarily cluster by the group (Fig. 1B). Volcano plots were used for the visualization of differentially expressed proteins between arbitrary two groups (Fig. 1C-E). Compared with the healthy controls, the CLGD group had 123 differential proteins (60 up-regulated and 63 down-regulated), the CHGD group had 287 differential proteins (152 up-regulated and 135 down-regulated) and the CGC group had 115 differential proteins (41 up-regulated and 74 down-regulated).

3.3. Identification of the most significant module by WGCNA

To identify the biologically significant protein modules associated with a course of CGC, we constructed gene-correlation networks using WGCNA. Outlier samples were examined using hierarchical clustering and no one was removed (Supplementary Fig. 2). The $\beta = 10$ (scale-free $R^2 = 0.935$) was selected as the soft-thresholding parameter to ensure a scale-free co-expression network (Supplementary Fig. 3). A total of seven modules were identified and marked by colors on the horizontal bar (Fig. 2A). According to the correlation analysis between modules and traits (Fig. 2B), the green module containing 42 proteins was identified as the positively correlated module ($r = 0.61$, $P < 0.05$), and the blue module containing 165 proteins was identified as the inversely correlated module ($r = -0.43$, $P < 0.05$). The plots of module membership and gene significance also illustrated significant correlations in the two modules (Fig. 2C and D). The layered clustering heat map showed the relative average expression for related proteins and the clustering results in different groups across the two modules (Supplementary Fig. 4 and 5).

3.4. GO functional enrichment and pathway enrichment analyses

To further elucidate the functional roles of the DEGs in the blue and green modules, GO functional enrichment (Fig. 3A and B) and pathway enrichment analyses were performed. According to the results of KEGG and Wikipathway cancer pathway analysis, the androgen receptor signaling pathway, Staphylococcus aureus infection, ubiquitin mediated proteolysis, glutathione metabolism, and selenium micronutrient network were most significantly overrepresented in the green module (Fig. 3C). For the blue module, we found that 20 pathways were significantly enriched, and the top 10 pathways are shown in Fig. 3D.

3.5. Identification and validation of candidate protein biomarkers

The top 200 proteins were identified based on RF importance scores, indicating their contribution to classification performance grouped by disease status (Supplementary Fig. 6). Take the intersection of the TOP 200 proteins from the RF model and the proteins from key modules of the WGCNA analysis. The Venn diagrams showed that 67 proteins were shared between the two subsets (Fig. 4A).

To identify key protein biomarkers that could distinguish CGC and precancerous lesions from healthy controls, we merged CLGD, CHGD, and CGC in a single group. Subsequently, the LASSO analysis was performed to screen candidate protein biomarkers from the intersection, and five proteins were filtered with optimal lambda value (Fig. 4B and C). The expression of GSTP1, PRDX1, and RAB2B was higher in the case group than in the control group, whereas those of CSRP1 and LY6G6F were significantly lower than those in healthy controls (Fig. 4D).

We used immunohistochemical results from the HPA database to further validate the expression of the five candidate protein biomarkers between GC and normal tissues (Fig. 4E). The results in agreement with the relative protein expression levels were confirmed except for RAB2B and PRDX1, which exhibited negative or insignificant outcomes.

With the results obtained from immunohistochemistry, GSTP1, CSRP1, and LY6G6F were selected for further validation through the RF model to discriminate CGC and precancerous lesions from healthy controls. The ROC curve analysis demonstrated that the candidate biomarker panel has good discriminatory power, with an AUC of 0.91 (95% CI, 0.69–1.00) for control vs. CLGD (Fig. 5A), 0.99 (95% CI, 0.99–1.00) for control vs. CHGD (Fig. 5B), 0.98 (95% CI, 0.78–1.00) for control vs. CGC (Fig. 5C). The PLS-DA models obtained with the panel revealed a clear separation between CLGD (Fig. 5D), CHGD (Fig. 5E), and CGC (Fig. 5F) compared to the healthy control.

4. Discussion

Due to its nonspecific symptoms and lack of non-invasive detection methods, CGC is often diagnosed at advanced stages when the window for effective surgical treatment has passed. Therefore, a quest for potential biomarkers for CGC screening is urgently needed. In this study, we conducted LC-MS/MS-based DIA proteomics to explore dynamic changes in protein signatures in the evolution of CGC and precancerous lesions. We identified 42 positively associated and 164 inversely associated proteins related to CGC progression and demonstrated several canonical cancer-associated pathways. We also identified a panel of three candidate protein biomarkers that could distinguish CLGD, CHGD, and CGC from healthy controls with excellent accuracy.

DIA is a relatively novel technique in MS-based proteomics and has the advantage of both quantitative accuracy and increased data completeness when compared to DDA. Previous studies have investigated cancer-specific diagnostic biomarkers based on DIA mode, such as pancreatic cancer,[17] breast cancer[18] and colorectal cancer.[19] Noteworthy, the study of Su et al.[20] analyzed the differences in proteomic features in ten GC and adjacent non-tumor tissues using DIA-MS and identified three protein markers with excellent diagnostic capability. Nevertheless, these studies were mainly focusing on cancer and adjacent non-cancer tissues and did not include patients with precancerous lesions. The development of CGC is a lengthy, dynamic biological process that advances from normal tissue, precancerous lesions (CLGD and CHGD), and invasive carcinoma, in which the expression of biomarkers at various stages may change along with neoplastic progression.[15] For instance, the levels of *Helicobacter pylori* antibody in blood serum were highly expressed from atrophic gastritis and precancerous lesions stages, then dropped or

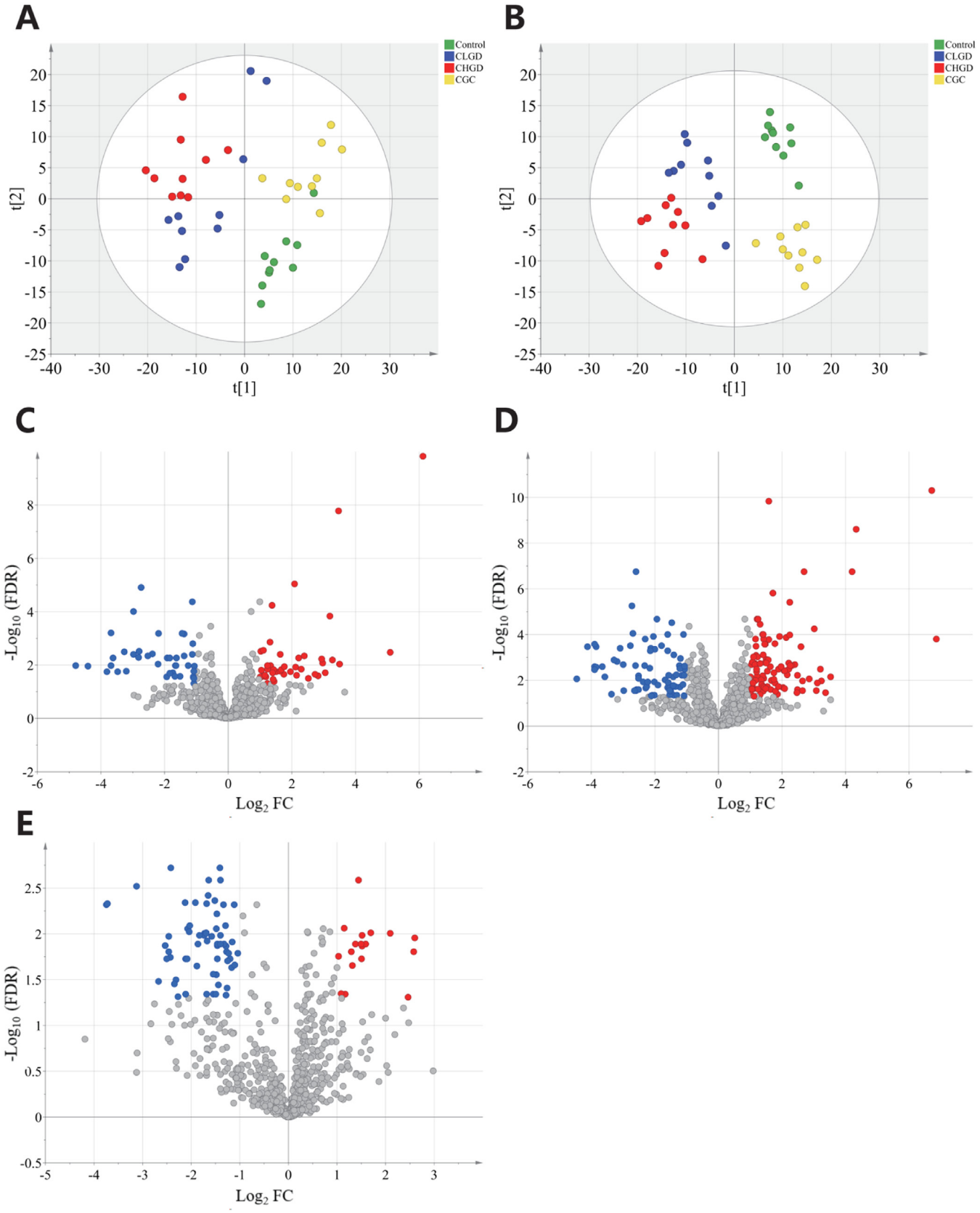

Fig. 1. The DIA-based quantitative proteomic landscape of the CGC and precancerous lesions. (A) Principal component analysis illustrating moderate clustering among four groups. (B) Partial Least Squares Discriminant Analysis illustrating obvious clustering among four groups. (C-E) Volcano plots showing the differentially expressed proteins in CGC and precancerous lesions versus healthy controls. The gray line represents the cutoff line with indicated significance criteria. Points having absolute log fold-change ≥ 1 and FDR adjusted *P*-value < 0.05 are shown in red, with absolute log fold-change ≤ 1 and *P*-value ≥ 0.05 are in blue, and the rest are in gray: (C) CLGD versus healthy controls, (D) CHGD versus healthy controls, and (E) CGC versus healthy controls. CGC, cardia gastric cancer; CHGD, cardia high-grade dysplasia; CLGD, cardia low-grade dysplasia; DIA, data-independent acquisition; FC, fold change; FDR, false discovery rate.

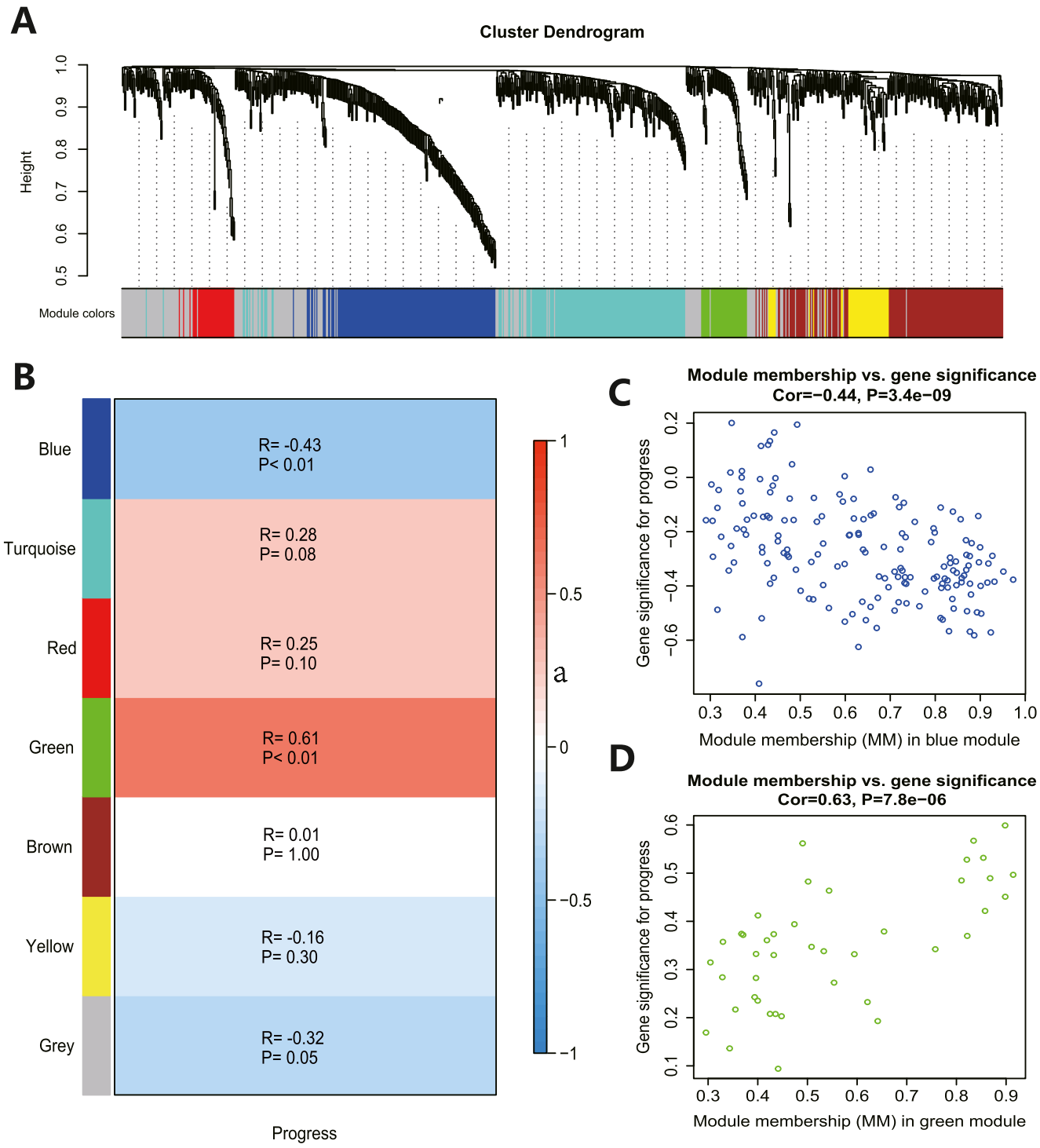

Fig. 2. The results of weighted gene co-expression network analysis. (A) WGCNA was performed to identify seven modules by unsupervised clustering. (B) Six modules (non-gray) were identified. The blue module was identified as the inversely correlated module ($r = -0.43$, $P < 0.01$) and the green module was identified as the positively correlated module ($r = 0.61$, $P < 0.01$). (C) The gene significance and module membership of the genes in the blue module exhibited an inverse correlation. (D) The gene significance and module membership of the genes in the blue module exhibited a positive correlation.

even disappeared during advanced gastric carcinoma.[21] Therefore, the identification of the more valuable biomarkers through the evolution of precancerous gastric cardia lesions and CGC development was required. In two studies conducted in Linqu, the researchers explored proteomic signatures associated with the progression of gastric lesions and risk of early GC using urine[22] and pathological tissues,[23] both achieving highly effective predictive results. The above results may have translational significance for defining high-risk populations of GC and its early detection; however, the patients were dominated by non-cardia gastric cancer and the DDA model was used. Given this, the present study explored dynamic changes in proteome and protein signatures in multiple stages of CGC through DIA-MS, which may help identify specific biomarkers and protein-associated pathway networks, thus providing a reference for population risk stratification.

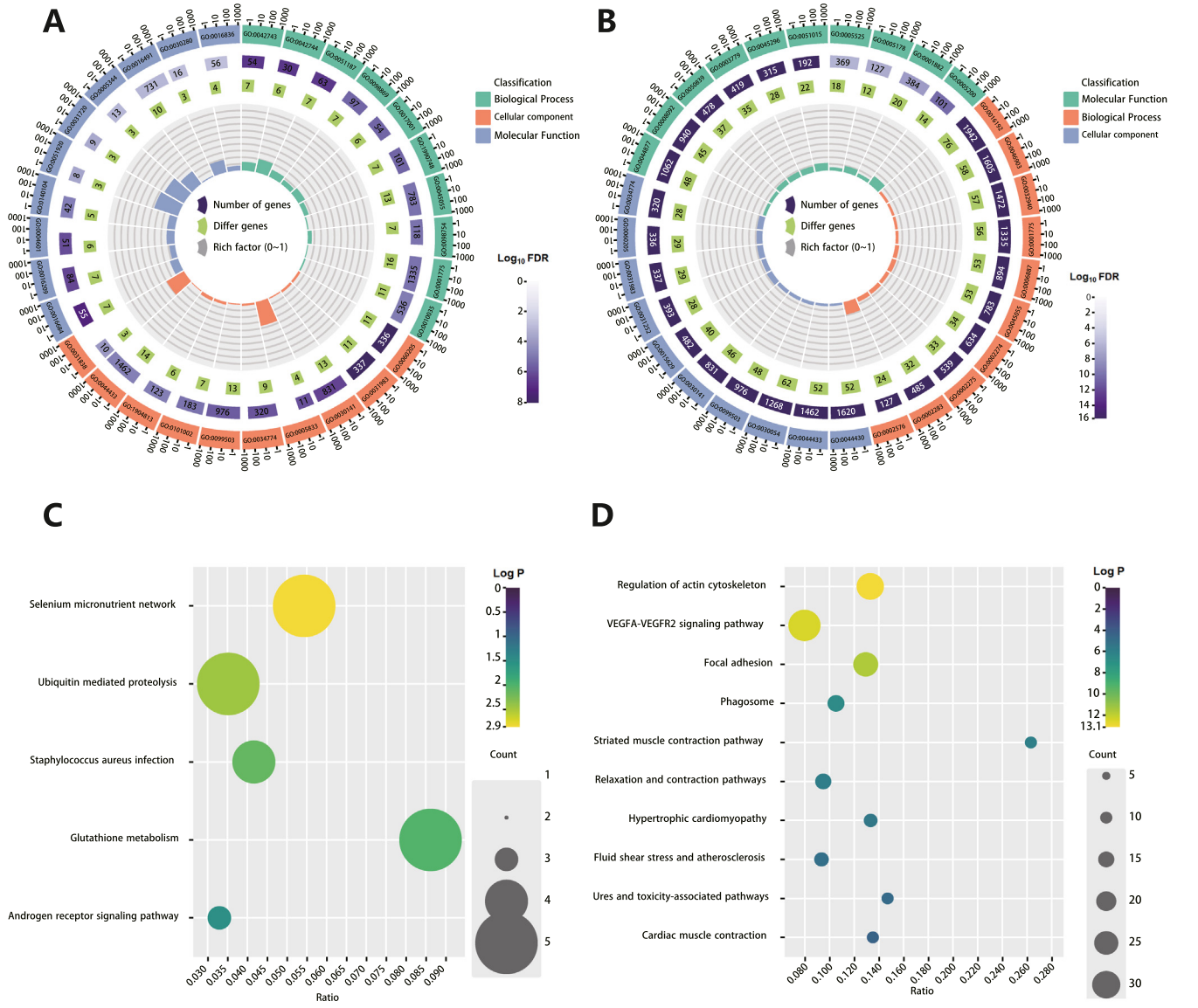

Fig. 3. GO functional enrichment and pathway enrichment analyses in blue and green modules. (A) Top 10 GO terms associated with the Molecular Functions, Biological Processes, and Cellular Compartments in the blue module. (B) Top 10 GO terms associated with the Molecular Functions, Biological Processes, and Cellular Compartments in the green module. (C) KEGG and Wikipathway enrichment analyses in green (green bar) module. Five pathways were enriched significantly. (D) KEGG and Wikipathway enrichment analyses in the blue module. Twenty pathways were enriched significantly and the top 10 were shown here.

Pathway enrichment analysis of the blue and green modules revealed that there are several signal pathways involved in cardia gastric carcinogenesis and development. The results of our pathway analysis can be summarized into three main processes: Selenium micronutrient network, muscle atrophy induced by cancer cachexia and sex hormone signaling pathways. Evidence from prospective studies suggests that individuals with lower blood Selenium levels have an increased risk of developing digestive tract cancers, while supplementation with Selenium can effectively antagonize the development of gastric cancer and esophageal cancer.[24-27] Muscle atrophy induced by cancer cachexia is a significant clinical characteristic of patients with gastric cancer, and there is preliminary evidence that it was associated with increased chemotherapy toxicity.[28] In addition, the ubiquitin-proteasome system (UPS) and the autophagic-lysosomal system (ALS) are the two major cellular protein degradation pathways.[29] A recent study noted that ALS and UPS were simultaneously activated in gastric cancer cachexia and might play complementary and sometimes synergistic roles in cachexia-induced muscle loss.[30] We also found the sex hormone signaling pathways, including estrogen and androgen signaling pathways, were associated with the occurrence and development of CGC. Previous studies suggested that a deficiency of estrogen may be associated with an increased risk of GC, while reduced androgen may also lead to a lower risk of GC.[31,32] The signaling pathways above are involved in the genesis, development and invasiveness of GC, providing compelling evidence for exploring biologically meaningful biomarkers.

Alongside, we identified three protein biomarkers with high diagnostic efficacy for CGC, including CSRP1, GSTP1 and LY6G6F. The practical applicability of these protein biomarkers is supported by their biological significance and their role in CGC development. The CSRP1 is a member of the cysteine protease family and is involved in cellular development and differentiation. The lower expression of CSRP1 may result in abnormal cell growth and differentiation, which in turn kindles tumorigenesis.[33] The GSTP1 belongs to the glutathione S-transferases family, which plays a critical role in detoxification and susceptibility to many

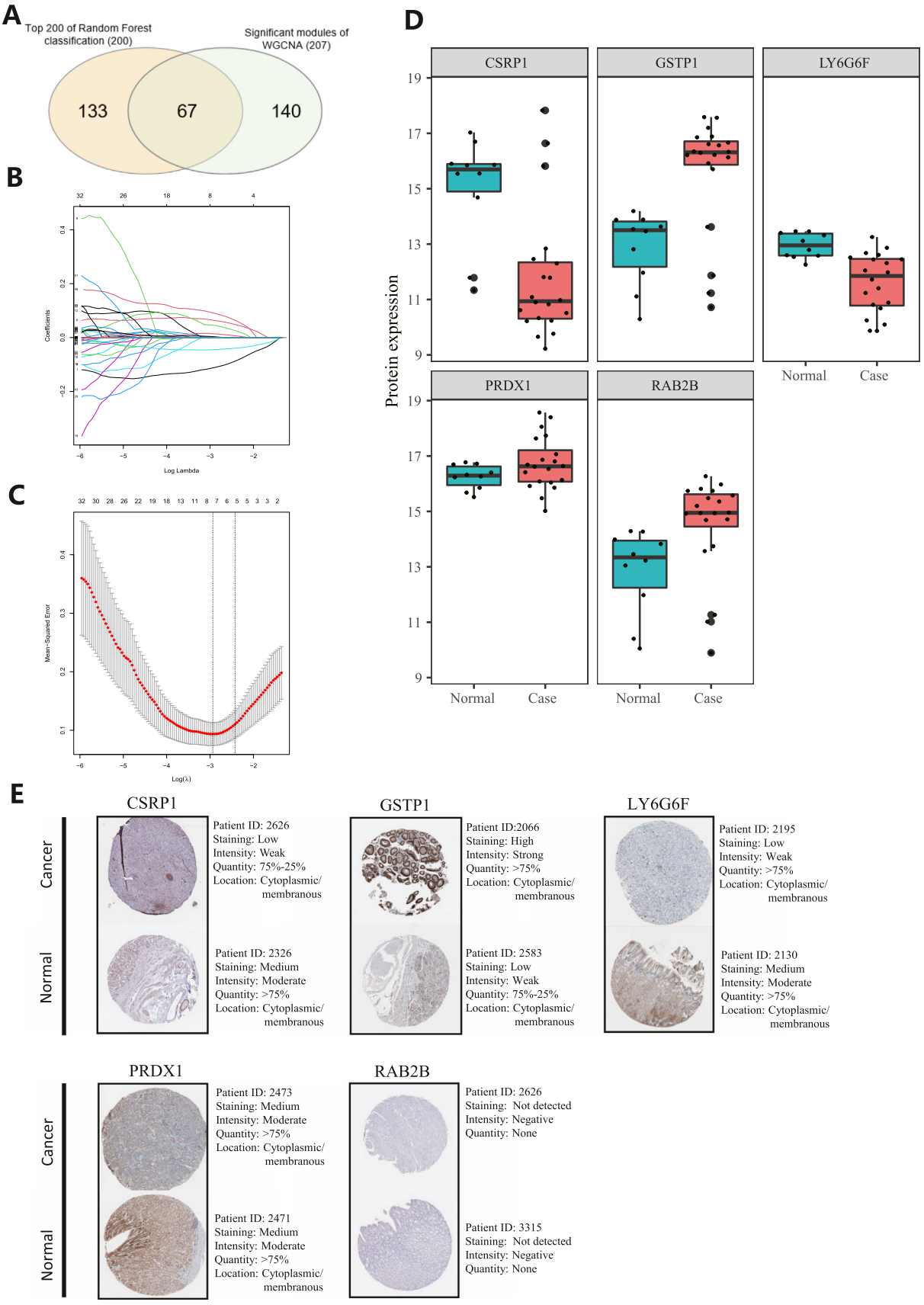

Fig. 4. Identification and validation of candidate protein biomarkers. (A) Venn diagrams showed that 67 proteins were shared between the results of WGCNA and random forest. (B) The extracted features were reduced via the LASSO regression. The left and right dotted vertical lines represent the optimal values of lambda when using the minimum criterion and the one-fold standard error of the minimum criterion, respectively. (C) LASSO coefficients of the five variables. (D) The expression level of five proteins in the case group (CLGD&CHGD&CGC) and health controls. (E) Immunohistochemistry of the five proteins based on the Human Protein Atlas.

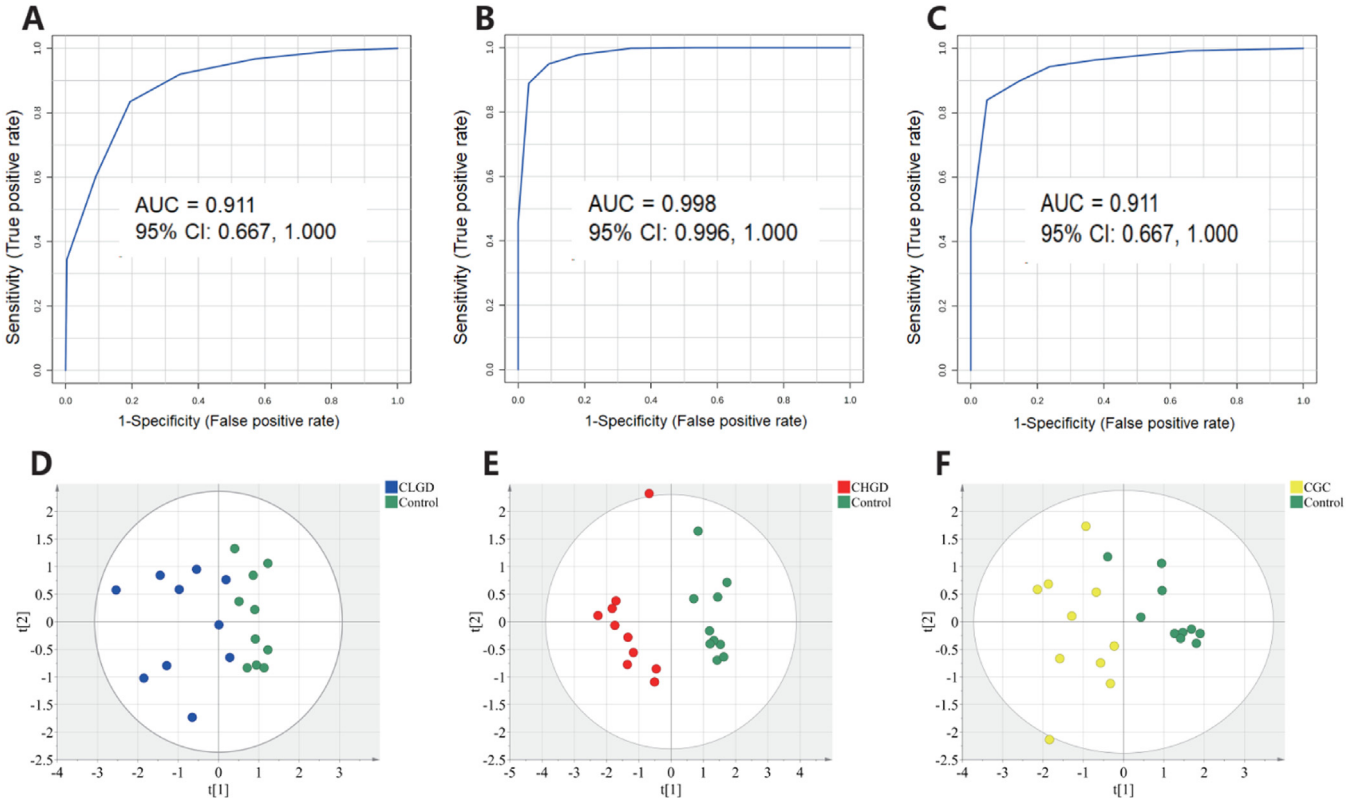

Fig. 5. Diagnostic performances of three protein combinations. (A-C) Receiver operating characteristic analysis of the three candidate biomarkers combined to discriminate CGC and precancerous lesions versus healthy controls: (A) CLGD versus healthy controls, (B) CHGD versus healthy controls, and (C) CGC versus healthy controls. (D-F) Partial Least Squares Discriminant Analysis of the three candidate biomarkers combined to discriminate CGC and precancerous lesions versus healthy controls: (D) CLGD versus healthy controls, (E) CHGD versus healthy controls, and (F) CGC versus healthy controls. AUC, area under curve; CGC, cardia gastric cancer; CHGD, ardia high-grade dysplasia; CI, confidence interval; CLGD, cardia low-grade dysplasia.

diseases including cancer. It has been suggested that the GSTP1 Val allele shows interaction with *Helicobacter pylori* infection to increase the risk of GC.[34] Although there is no direct evidence for the involvement of LY6G6F in cancer development, evidence from previous investigations on the molecular mechanisms revealed that LY6G6F may operate downstream of the Grb7-dependent signal transduction pathway and indirectly participate in the pathogenesis of digestive tract tumors.[35] Combined with the quantitative information of immunohistochemistry images from the HPA public database, we further verified their expression levels in normal and GC tissues, which are consistent with the results of DIA-MS.

Strengths of our study included the representative plasma samples from multiple stages of disease progression, well-matched results for sex and age, and applied DIA strategy with better stability and reproducibility. There were several limitations to our study as well. Firstly, although the candidate protein biomarkers were validated by the HPA database, external validation involving more representative populations is still warranted. Secondly, we cannot unravel the complex mechanisms underlying these results. Since the data mining and visualization for quantitative proteomics data were only exploratory in nature, the mechanisms of the protein biomarkers in CGC development need to be elucidated in future studies.

5. Conclusions

Overall, our study conducted a comprehensive proteomic analysis utilizing the DIA strategy, resulting in the identification of three potential candidate protein biomarkers. These biomarkers hold promise for identifying high-risk populations for CGC and precancerous lesions. However, it is important to emphasize that further validation in larger, diverse cohorts is essential to ascertain their true diagnostic potential. If validated in future studies, these biomarkers may have the potential to complement existing diagnostic approaches, such as endoscopic examinations, aiding in CGC prevention and control.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Author Zeming Wu serves as a technical advisor for “iPhenome Biotechnology (Dalian)” company. In the Methods section of the manuscript, he provided recommendations for bioinformatics analysis. There are no financial relationships or conflicts of interest between the author’s affiliation and “iPhenome Biotechnology (Dalian)” company. This study did not receive any form of funding or support from “iPhenome Biotechnology (Dalian)” company. The research results are objective and truthful, and Zeming Wu’s involvement did not influence the authenticity of the results. Other authors declare that they have no conflict of interests.

Ethics statement

This study was approved by the Institutional Review Board of the Cancer Hospital, Chinese Academy of Medical Sciences (approval number: 20/387-2583) and all participants provided written informed consent.

Data availability

The data generated in this study are available from the corresponding author on reasonable request.

Acknowledgments

This work was supported by grants from the Beijing Nova Program (grant number: Z201100006820069), the Hope Star Talent Incentive Program of the Cancer Hospital of Chinese Academy of Medical Sciences, the National Natural Science Fund (grant number: 81573224), the CAMS Innovation Fund for Medical Sciences (grant number: 2021-I2M-1-023), and the Shandong Province Natural Science Foundation Youth Branch (grant number: ZR202211080124).

Author contributions

J.G. and S.X. conducted the study design, data analysis and writing. X.L. performed the data collection. L.X. verified the pathological results. Z.W. provided guidance on bioinformatics methodologies. W.W. and S.W. acted as the principal investigator, and conducted the trial design and final approval.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jncc.2023.10.003.

References

1. Arnold M, Ferlay J, van Berge Henegouwen MI, et al. Global burden of esophageal and gastric cancer by histology and subsite in 2018. *Gut*. 2020;69:1564–1571. doi:10.1136/gutjnl-2020-321600.
2. Buas MF, Vaughan TL. Epidemiology and risk factors for gastroesophageal junction tumors: understanding the rising incidence of this disease. *Semin Radiat Oncol*. 2013;23:3–9. doi:10.1016/j.semradonc.2012.09.008.
3. Zeng H, Chen W, Zheng R, et al. Changing cancer survival in China during 2003-15: a pooled analysis of 17 population-based cancer registries. *Lancet Glob Health*. 2018;6:e555–e567. doi:10.1016/S2214-109X(18)30127-X.
4. Chen R, Liu Y, Song G, et al. Effectiveness of one-time endoscopic screening program in prevention of upper gastrointestinal cancer in China: a multicentre population-based cohort study. *Gut*. 2021;70:251–260. doi:10.1136/gutjnl-2019-320200.
5. Cai Q, Zhu C, Yuan Y, et al. Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study. *Gut*. 2019;68:1576–1587. doi:10.1136/gutjnl-2018-317556.
6. Charvat H, Sasazuki S, Inoue M, et al. Prediction of the 10-year probability of gastric cancer occurrence in the Japanese population: the JPHC study cohort II. *Int J Cancer*. 2016;138:320–331. doi:10.1002/ijc.29705.
7. Suh YS, Lee J, Woo H, et al. National cancer screening program for gastric cancer in Korea: nationwide treatment benefit and cost. *Cancer*. 2020;126:1929–1939. doi:10.1002/cncr.32753.
8. Kang C, Lee Y, Lee JE. Recent advances in mass spectrometry-based proteomics of gastric cancer. *World J Gastroenterol*. 2016;22:8283–8293. doi:10.3748/wjg.v22.i37.8283.
9. Ludwig C, Gillet L, Rosenberger G, et al. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol*. 2018;14:e8126. doi:10.15252/msb.20178126.
10. Hansen FM, Tanzer MC, Brüning F, et al. Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology. *Nat Commun*. 2021;12:254. doi:10.1038/s41467-020-20509-1.
11. Zheng X, Xu K, Zhou B, et al. A circulating extracellular vesicles-based novel screening tool for colorectal cancer revealed by shotgun and data-independent acquisition mass spectrometry. *J Extracell Vesicles*. 2020;9:1750202. doi:10.1080/20013078.2020.1750202.
12. Keam SP, Gulati T, Gamell C, et al. Exploring the oncoproteomic response of human prostate cancer to therapeutic radiation using data-independent acquisition (DIA) mass spectrometry. *Prostate*. 2018;78:563–575. doi:10.1002/pros.23500.
13. Ortea I, Rodríguez-Ariza A, Chicano-Gálvez E, et al. Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH-MS data-independent acquisition and targeted data extraction. *J Proteomics*. 2016;138:106–114. doi:10.1016/j.jprot.2016.02.010.
14. Wei WQ, Hao CQ, Guan CT, et al. Esophageal histological precursor lesions and subsequent 8.5-year cancer risk in a population-based prospective study in China. *Am J Gastroenterol*. 2020;115:1036–1044. doi:10.14309/ajg.0000000000000640.
15. Gu J, Xie S, Wang S, et al. Surveillance of premalignant gastric cardia lesions: a population-based prospective cohort study in China. Int J Cancer. 2021;149(9):1639–1648. doi:10.1002/ijc.33720. Published correction appears in Int J Cancer. 2023. doi:10.1002/ijc.34735.
16. Pappireddi N, Martin L, Wühr M. A review on quantitative multiplexed proteomics. *ChemBioChem*. 2019;20:1210–1224. doi:10.1002/cbic.201800650.
17. Azizian NG, Sullivan DK, Nie L, et al. Selective labeling and identification of the tumor cell proteome of pancreatic cancer in vivo. *J Proteome Res*. 2021;20:858–866. doi:10.1021/acs.jproteome.0c00666.
18. Bouchal P, Schubert OT, Faktor J, et al. Breast cancer classification based on proteotypes obtained by SWATH Mass spectrometry. *Cell Rep*. 2019;28 832-843.e7. doi:10.1016/j.celrep.2019.06.046.
19. Rao J, Wan X, Tou F, et al. Molecular characterization of advanced colorectal cancer using serum proteomics and metabolomics. *Front Mol Biosci*. 2021;8:687229. doi:10.3389/fmolb.2021.687229.
20. Su F, Zhou FF, Zhang T, et al. Quantitative proteomics identified 3 oxidative phosphorylation genes with clinical prognostic significance in gastric cancer. *J Cell Mol Med*. 2020;24:10842–10854. doi:10.1111/jcmm.15712.
21. Yang L, Kartsonaki C, Yao P, et al. The relative and attributable risks of cardia and non-cardia gastric cancer associated with Helicobacter pylori infection in China: a case-cohort study. *Lancet Public Health*. 2021;6:e888–e896. doi:10.1016/S2468-2667(21)00164-X.
22. Fan H, Li X, Li ZW, et al. Urine proteomic signatures predicting the progression from premalignancy to malignant gastric cancer. *EBioMedicine*. 2022;86:104340. doi:10.1016/j.ebiom.2022.104340.
23. Li X, Zheng NR, Wang LH, et al. Proteomic profiling identifies signatures associated with progression of precancerous gastric lesions and the risk of early gastric cancer. *EBioMedicine*. 2021;74:103714. doi:10.1016/j.ebiom.2021.103714.
24. Hashemian M, Poustchi H, Abnet CC, et al. Dietary intake of minerals and risk of esophageal squamous cell carcinoma: results from the Golestan Cohort Study. *Am J Clin Nutr*. 2015;102:102–108. doi:10.3945/ajcn.115.107847.
25. Nasir A, Bullo MMH, Ahmed Z, et al. Nutrigenomics: epigenetics and cancer prevention: a comprehensive review. *Crit Rev Food Sci Nutr*. 2020;60:1375–1387. doi:10.1080/10408398.2019.1571480.
26. Wang SM, Taylor PR, Fan JH, et al. Effects of nutrition intervention on total and cancer mortality: 25-year post-trial follow-up of the 5.25-year Linxian nutrition intervention trial. *J Natl Cancer Inst*. 2018;110:1229–1238. doi:10.1093/jnci/djy043.
27. Li WQ, Zhang JY, Ma JL, et al. Effects of Helicobacter pylori treatment and vitamin and garlic supplementation on gastric cancer incidence and mortality: follow-up of a randomized intervention trial. *BMJ*. 2019;366:l5016. doi:10.1136/bmj.l5016.
28. Bossi P, Delrio P, Mascheroni A, et al. The spectrum of malnutrition/cachexia/sarcopenia in oncology according to different cancer types and settings: a narrative review. *Nutrients*. 2021;13:1980. doi:10.3390/nu13061980.
29. Sun-Wang JL, Ivanova S, Zorzano A. The dialogue between the ubiquitin-proteasome system and autophagy: implications in ageing. *Ageing Res Rev*. 2020;64:101203. doi:10.1016/j.arr.2020.101203.
30. Zhang Y, Wang J, Wang X, et al. The autophagic-lysosomal and ubiquitin proteasome systems are simultaneously activated in the skeletal muscle of gastric cancer patients with cachexia. *Am J Clin Nutr*. 2020;111:570–579. doi:10.1093/ajcn/nqz347.
31. Errors in figures 4, 5, and 6. Correction. *Nat Commun*. 2018;9:3599. doi:10.1038/s41467-018-06061-z.
32. McMenamin ÚC, Liu P, Kunzmann AT, et al. Circulating sex hormones are associated with gastric and colorectal cancers but not esophageal adenocarcinoma in the UK biobank. *Am J Gastroenterol*. 2021;116:522–529. doi:10.14309/ajg.0000000000001045.
33. Zhou CZ, Qiu GQ, Wang XL, et al. Screening of tumor suppressor genes on 1q31.1-32.1 in Chinese patients with sporadic colorectal cancer. *Chin Med J*. 2008;121:2479–2486.
34. Zhang Y, Sun LP, Xing CZ, et al. Interaction between GSTP1 Val allele and H. pylori infection, smoking and alcohol consumption and risk of gastric cancer among the Chinese population. *PLoS One*. 2012;7:e47178. doi:10.1371/journal.pone.0047178.
35. Kishi T, Sasaki H, Akiyama N, et al. Molecular cloning of human GRB-7 co-amplified with CAB1 and c-ERBB-2 in primary gastric cancer. *Biochem Biophys Res Commun*. 1997;232:5–9. doi:10.1006/bbrc.1997.6218.