



OPEN ACCESS

EDITED BY

Qun Zhang,
First Affiliated Hospital,
Nanjing Medical University, China

REVIEWED BY

An Zhao,
Zhejiang Cancer Hospital, University of
(CAS), China
Bo Cao,
People's Liberation Army General
Hospital, China

*CORRESPONDENCE

Jinhua Jiang
scarli8880@126.com
Donglei Zhou
zhou_dl@sina.com
Lujian Liao
ljiao@bio.ecnu.edu.cn

[†]Department of Gastric Surgery, Fudan
University Shanghai Cancer Center,
Department of Oncology, Shanghai
Medical College, Fudan University,
Shanghai, China

[‡]These authors have contributed
equally to this work and share first
authorship

SPECIALTY SECTION

This article was submitted to
Gastrointestinal Cancers: Gastric and
Esophageal Cancers,
a section of the journal
Frontiers in Oncology

RECEIVED 22 September 2022

ACCEPTED 07 November 2022

PUBLISHED 24 November 2022

CITATION

Gu L, Chen J, Yang Y, Zhang Y, Tian Y,
Jiang J, Zhou D and Liao L (2022)
Data-independent acquisition mass
spectrometry identification of
extracellular vesicle biomarkers for
gastric adenocarcinoma.
Front. Oncol. 12:1051450.
doi: 10.3389/fonc.2022.1051450

Data-independent acquisition mass spectrometry identification of extracellular vesicle biomarkers for gastric adenocarcinoma

Lei Gu^{1‡}, Jin Chen^{2‡}, Yueying Yang^{2‡}, Yunpeng Zhang²,
Yuying Tian², Jinhua Jiang^{3*}, Donglei Zhou^{4,5*†}
and Lujian Liao^{2,6*}

¹Department of General Surgery, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China, ²Shanghai Key Laboratory of Regulatory Biology, School of Life Sciences, East China Normal University, Shanghai, China, ³Department of Interventional Oncology, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China, ⁴Department of Gastric Surgery, Fudan University Shanghai Cancer Center, Shanghai, China, ⁵Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China, ⁶Durbrain Medical Laboratory, Hangzhou, Zhejiang, China

Early diagnosis of gastric adenocarcinoma (GAC) can effectively prevent the progression of the disease and significantly improve patient survival. Currently, protein markers in clinical practice barely meet patient needs; it is therefore imperative to develop new diagnostic biomarkers with high sensitivity and specificity. In this study, we extracted extracellular vesicles (EV) from the sera of 33 patients with GAC and 19 healthy controls, then applied data-independent acquisition (DIA) mass spectrometry to measure protein expression profiles. Differential protein expression analysis identified 23 proteins showing expression patterns across different cancer stages, from which 15 proteins were selected as candidate biomarkers for GAC diagnosis. From this subset of 15 proteins, up to 6 proteins were iteratively selected as features and logistic regression was used to distinguish patients from healthy controls. Furthermore, serum-derived EV from a new cohort of 12 patients with gastric cancer and 18 healthy controls were quantified using the same method. A classification panel consisting of GSN, HP, ORM1, PIGR, and TFRC showed the best performance, with a sensitivity and negative predictive value (NPV) of 0.83 and 0.82. The area under curve (AUC) of the receiver operating characteristic (ROC) is 0.80. Finally, to facilitate the diagnosis of advanced stage GAC, we identified a 3-protein panel consisting of LYZ, SAA1, and F12 that showed reasonably good performance with an AUC of 0.83 in the validation dataset. In conclusion, we identified new protein biomarker panels from serum EVs for early diagnosis of gastric cancer that worth further validation.

KEYWORDS

proteomics, serum, extracellular vesicle, gastric adenocarcinoma, biomarker

Introduction

Gastric adenocarcinoma (GAC) is a highly invasive cancer with the third highest mortality rate world-wide (1–3). Although technology advancements have reduced the overall incidences and mortality of GAC (2), it remains one of the most common cancers (4). This situation is partly due to lack of sensitive methods to identify GAC at early stages, leading to unnoticed tumor progression and poor patient prognosis. Thus, early diagnosis of GAC has the potential to greatly improve the chance of patient survival (5).

Diagnostic methods used in clinical practice include upper gastrointestinal (UGI) radiography, endoscopy, histopathology and liquid biopsy. Endoscopy-guided biopsy and pathology is the “gold standard”, and is necessary for confirming the malignancy, stage, and tissue of origin (6). However, due to its invasive nature and poor patient compliance, extensive use of endoscopy in screening GAC is impractical. On the other hand, the emerging technology of liquid biopsy features non-invasiveness and low cost and has the potential to provide diagnostic information prior to the onset of symptoms, could provide a promising tool for screening gastric cancer at early stages. As potentially valuable diagnostic tools, individual protein markers provide relatively low sensitivity and specificity at present. For example, the sensitivity of CA72-4, CEA and CA125 in detecting GAC are all below 40%, but the sensitivity of combining the three proteins can rise to 66% (7). Still, sensitivity at this level remains too low to satisfy clinical demands. Discovering more efficient protein panels holds the promise to improve the sensitivity of detecting cancer at an early stage.

Previous studies have shown that EV is involved in many processes in the onset and development of gastric cancer (8–10). These vesicles carry RNAs, proteins as well as metabolites, which may reflect the pathological state of cancer cells. EV can transport specific proteins and nucleic acids into target cells in the tumor microenvironment, affecting tumor cell proliferation and metastasis, inhibiting immune surveillance and incurring drug resistance (11). In addition, the membrane structure of EV can preserve the molecular components. Due to its versatile functionalities, EV has caught tremendous attentions in cancer field.

In this study, we extracted EV from sera of GAC patients and healthy subjects, and applied LC-MS/MS technology to capture protein expression profiles. From this dataset we further screened reliable diagnostic protein biomarkers for GAC, aiming to explore the clinical usefulness of serum EV.

Materials and methods

Collection of serum samples

Peripheral blood samples were collected from healthy controls and GAC patients at Shanghai Tenth People's Hospital. Cohort 1 consisted of 19 controls and 33 patients, while Cohort 2 consisted of

18 controls and 12 GAC patients. The demographic data and staging information are listed in Table 1. To prepare serum samples, venal blood was drawn and placed at room temperature for 30 minutes, then centrifuged at $2000 \times g$ for 10 min. Serum was collected and stored at -80°C until use.

Isolation of EV from serum samples

EV was isolated from serum samples using ultracentrifugation (UC). Briefly, 500 μL of serum was centrifuged at $2500 \times g$ for 10 min (4°C) followed by another centrifugation at $10000 \times g$ for 30 min to pellet cell debris. The supernatant was then filtered through a 0.22- μm cellulose acetate centrifuge filter (Costar, USA), and the filtrate was diluted with PBS into a final volume of 5 mL. Crude EVs were pelleted by ultracentrifugation at $110,000 \times g$ (P70AT rotor, Hitachi, Japan) for 5 h. Afterwards, the pellets were resuspended with PBS and ultracentrifuge at $110,000 \times g$ for 70 min. The final EV pellets were resuspended with 50 μL of PBS and stored at -80°C for further analysis.

Characterization of EVs

For transmission electron microscopy (TEM), 10 μL of PBS-diluted EV samples were added on top of the copper grids and incubated for 10 min at room temperature. The grids were washed with 6 μL of ultrapure water and negatively stained with 3% phosphotungstic acid for 10 min. Then, the grids were washed with ultrapure water and air-dried. Imaging was performed on a H-7700 transmission electron microscope (Hitachi, Japan).

To analyze particle size, EV samples were diluted 10 times with PBS and analyzed with Zetasizer Nano S instrument (Malvern, UK) according to manufacturer's instruction.

EV samples were lysed with RIPA buffer (150 mM NaCl, 0.5% SDC, 0.1% SDS, 50 mM Tris/HCl, 1% Triton-X 100, pH 7.6) and 20 μL of each lysed sample was separated on 12% SDS-PAGE. For immunoblotting, proteins were transferred onto a nitrocellulose membrane and incubated with anti-CD9 antibody (1:1000; Cat. #ab92726, Abcam, UK) and anti-Hsp70 antibody (1:1000; Cat. #ab181606, Abcam, UK) followed by anti-rabbit IgG secondary antibody (1:10000; Cat. #09-0034, Yeason, China). For silver staining, samples were washed with 50% methanol, 5% methanol and pure water successively, and then reduced by 0.0005% dithiothreitol (DTT; Cat. #43217, Sigma-Aldrich, USA) followed by incubation with 0.1% silver nitrate in the dark for 20 min. Finally, 3% sodium carbonate with 0.01% formaldehyde was applied for visualization.

Protein digestion

BCA kit (Cat. #23225, Thermo Science, USA) was used to determine the protein concentration in isolated EVs samples.

TABLE 1 Clinical information (X denotes information unavailable).

Characteristic	Cohort 1		Cohort 2		Total
	Healthy	Gastricadenocarcinoma	Healthy	Gastricadenocarcinoma	
Total cases	19	33	18	12	82
Age (years)					
Median	55	65	63	68	62
Range	27-74	35-82	46-75	56-85	27-85
Sex					
Male	9	26	9	6	50
Female	10	7	9	6	32
Stage					
I	–	8	–	1	9
II	–	9	–	3	12
III	–	13	–	7	20
IV	–	3	–	1	4
T					
1	–	6	–	1	7
2	–	2	–	0	2
3	–	13	–	5	18
4	–	11	–	5	16
X	–	1	–	1	2
N					
0	–	14	–	4	18
1	–	3	–	2	5
2	–	5	–	4	9
3	–	8	–	1	9
X	–	3	–	1	4
M					
0	–	30	–	11	41
1	–	3	–	1	4

From each sample, 20 µg protein was vacuum dried and resuspended in denaturing solution (7 M urea, 2 M thiourea, 10 mM DTT, 1 × protease inhibitor [Cat. # P8340, Sigma-Aldrich, USA]). The samples were then reduced for 30 min at 55°C, alkylated with 15 mM iodoacetamide (Cat. # I1149, Sigma-Aldrich, USA) in the dark for 20 min, diluted with 50 mM ammonium bicarbonate solution and digested with trypsin (1:50; Cat. # V5113, Promega, USA) overnight at 37°C. The resulting peptides were desalted with C18 column and vacuum dried for mass spectrometry analysis.

LC-MS/MS analysis

Protein digests were analyzed on an EASY-nLC 1000 LC (Thermo Science, USA) coupled with Q-Exactive mass spectrometer (Thermo Science, USA). The mobile phases consisted of buffer A (2% ACN, 0.1% formic acid) and buffer B (98% ACN, 0.1% formic acid). Tryptic peptides were resuspended in buffer A and spiked with iRT peptides (Omicsson, China).

Equivalent to 1 µg of protein digest from each sample was loaded onto a C18 column (Cat. #164534, Thermo Science, USA) linked with a pre-column (Cat. #164535, Thermo Science, USA) and separated at a flow rate of 250 nL/min. A 120 min gradient from 3% to 8% buffer B in 5 min, 8% to 28% in 95 min, 28% to 95% in 10 min, 95% for 5 min, 95% to 3% in 2 min and 3% for 3 min was used. The MS instrument was operated in the positive polarity and profile mode with a nano-electrospray through a heated ion transfer tube with a temperature setting of 275°C. For data dependent acquisition (DDA), one full scan MS from 400 to 1400 m/z followed by 12 MS2 scan were continuously acquired. MS spectra were acquired with resolution of 70000 for a maximum injection time (IT) of 100 ms with an automatic gain control (AGC) target value of 3e6. MS2 spectra were obtained in the higher-energy collisional dissociation (HCD) mode using a normalized collision energy of 27%, resolution at 17500, maximum IT of 60 ms, AGC target of 5e5 and isolation window at 2.0 m/z. For data independent acquisition (DIA), isolation window for MS2 was set to 20 Da with 1 Da overlap over a precursor mass window of 450–950 m/z, and other parameters were set to be the same as DDA method.

Analysis of proteomic data

Qualitative analysis of DDA raw files was performed by Proteome Discoverer (version 2.0) software searching against the UniProtKB database (2020 release, *Homo sapiens*) including the 11 synthetic iRT peptides. A maximum of 2 missed cleavages were allowed for trypsin digestion with fixed carbamidomethylation (+57.0251 Da) of cysteine and oxidation (+15.9949 Da) of methionine. The mass tolerance allowed was 15 ppm for precursor ions and 0.05 Da for fragmentation ions. A false discovery rate (FDR) of 1% was set at both peptide and protein levels.

Skyline (version 20.2.0.343) software was used for independent proteome spectral library construction. Briefly, fasta and pdresult files were imported with the following parameters: structural modifications: carbamidomethyl (C), oxidation (M), acetyl (N-term); minimum length: 6; maximum length: 30. The retention time of iRT was calibrated and the isolation scheme was set based on the isolation window of DIA MS parameter. Target list was added to obtain information of peptides and proteins. Finally, the peptides with reversed sequence were added as decoy peptides for the control of false discovery rate.

DIA raw files were imported into skyline and analyzed based on the aforementioned DDA spectral library, and filtered by a mProphet scoring model trained with decoy peptides. Q value and dot products were set to 0.01 and 0.65 respectively for selecting peptides with high confidence. Afterwards, the decoy peptides were removed. The exported files were used for statistical analysis.

Statistical and bioinformatic analysis

RStudio (version 1.3.1073) was used to perform all the statistical analysis, including evaluation of data quality, data preprocessing, differential expression analysis, principal component analysis (PCA), construction of classification models. For differential expression analysis, p value < 0.05 were considered statistically significant, and fold change > 1.5 or < 0.67 were considered as up- or down-regulated, respectively. Venn diagram and functional annotation were generated with FunRich (version 3.3.1) software.

Results

Study design and characterization of EV

The design for EV-based GAC biomarker discovery is shown in Figure 1A. DIA-based quantitative mass spectrometry (12) analysis of 52 samples from the first cohort was conducted and candidate biomarkers was screened through differential protein

expression analysis. Logistic regression classification was applied to identify a panel of candidate proteins whose expression were associated with GAC. These biomarkers were further validated by a second cohort of 30 samples.

EV was isolated from the serum using ultracentrifugation (UC). Western blot analysis detected classical EV markers Hsp70 and CD9 in UC fractions, indicating successful enrichment of EV (Figure 1B). Particle analysis of randomly selected EV samples showed that the majority of the isolated EV particles ranged between 10~100 nm (Figure 1C), which is consistent with the range distribution of exosomes. Furthermore, the morphology of the EV particles as visualized by TEM showed a typical cup-shaped structure with the size between 50~200 nm (Figure 1D).

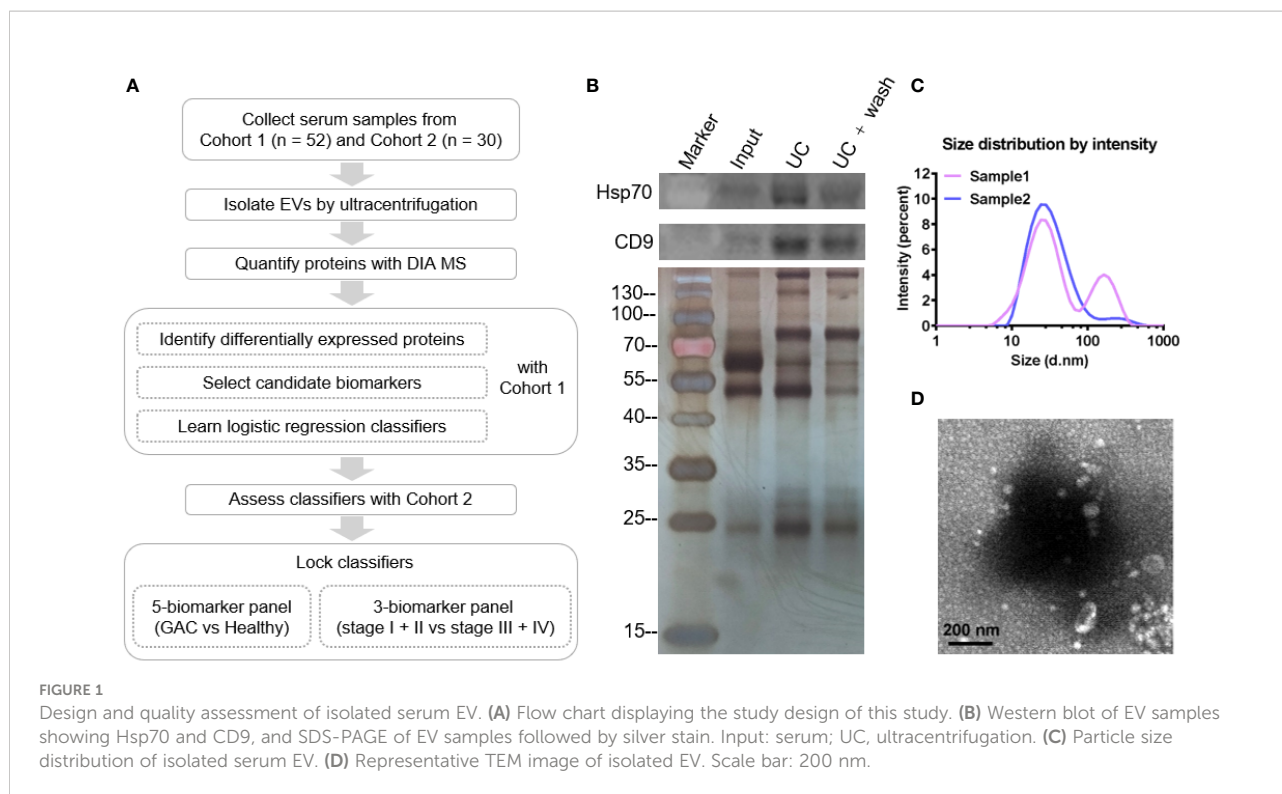
Proteomic analysis of the EV

For quantitative proteomic profiling, EV samples from 19 healthy subjects and 33 GAC patients were analyzed using DIA-based mass spectrometry. To assess the quality of our data as the result of a complex procedure of sample collection and handling, we interspersed quality control (QC) samples during mass spectrometry data acquisition. The distribution of the correlation coefficients of all the QC samples was between 0.71 and 0.92 (Figure 2A), indicating reasonable reproducibility. Signal intensity of mass spectrum spanned a dynamic range of six orders of magnitude, with the majority of precursor mass accuracy within ± 5 ppm (Figures 2B, D), indicating that our analysis achieved high accuracy and depth.

In total, we identified 448 proteins and quantified 352 proteins from 52 EV samples from cohort 1 (Figure 2C). We also performed EV enrichment on cohort 2 of 12 GAC patients and 18 healthy controls, followed by quantitative proteomic analysis, resulting in quantification of 321 proteins. A total of 249 proteins were quantified in both cohorts (Supplemental Figure 1A). Gene ontology analysis showed that more than 65% of these proteins localized in extracellular region and exosomes (Supplemental Figure 1C), further confirming the successful enrichment of EV.

Proteomic profiles of serum EV samples from cohort 1

Comparing the EV proteome profiles of GAC patients and healthy controls from cohort 1, we found a total of 26 significantly changed proteins, among which 13 were up-regulated and 13 down-regulated (Figure 3A). The fold change of up-regulated proteins had a wider range compared to that of down-regulated proteins (Figure 3B). The heatmap of differentially expressed proteins displayed distinct patterns, with gender, age and TNM stages of GAC were displayed together (Figure 3C). GO analysis



showed that the differentially expressed proteins mainly involved in protein activation cascade, hydrogen peroxide catabolic process, regulated exocytosis, hemostasis, acute-phase reaction, response to bacterium (Figure 3D). Regulated exocytosis is highly correlated with the secretion of EVs, and contains up-regulated proteins including apolipoprotein B (APOB), haptoglobin (HP), hemoglobin subunit alpha (HBA1) and hemoglobin subunit beta (HBB). On the other hand, the majority of the down-regulated proteins involved in complement and coagulation cascade including von Willebrand factor (VWF), coagulation factor XIII A chain (F13A1) and component 6 (C6).

Discovery of a serum EV biomarker panel for GAC diagnosis

To identify biomarkers with an increased accuracy to differentiate GAC from normal subjects, we performed multi-group differential protein expression analysis using the proteomic data from cohort 1. In addition to comparing GAC patients with healthy controls, patients with early-stage GAC (stage I + II, N=17) and late-stage GAC (stage III + IV, N=16) were also compared to healthy controls (Figures 4A, B), and the proteins with high consistency of expression trend were selected as candidate markers. Venn diagram showed the aforementioned three comparisons (Supplemental Figure 2). In

total, there are 23 intersecting proteins, of which 15 are EV proteins (Figure 4C). These 15 proteins were then used as candidate serum EV biomarkers for GAC diagnosis (Table 2).

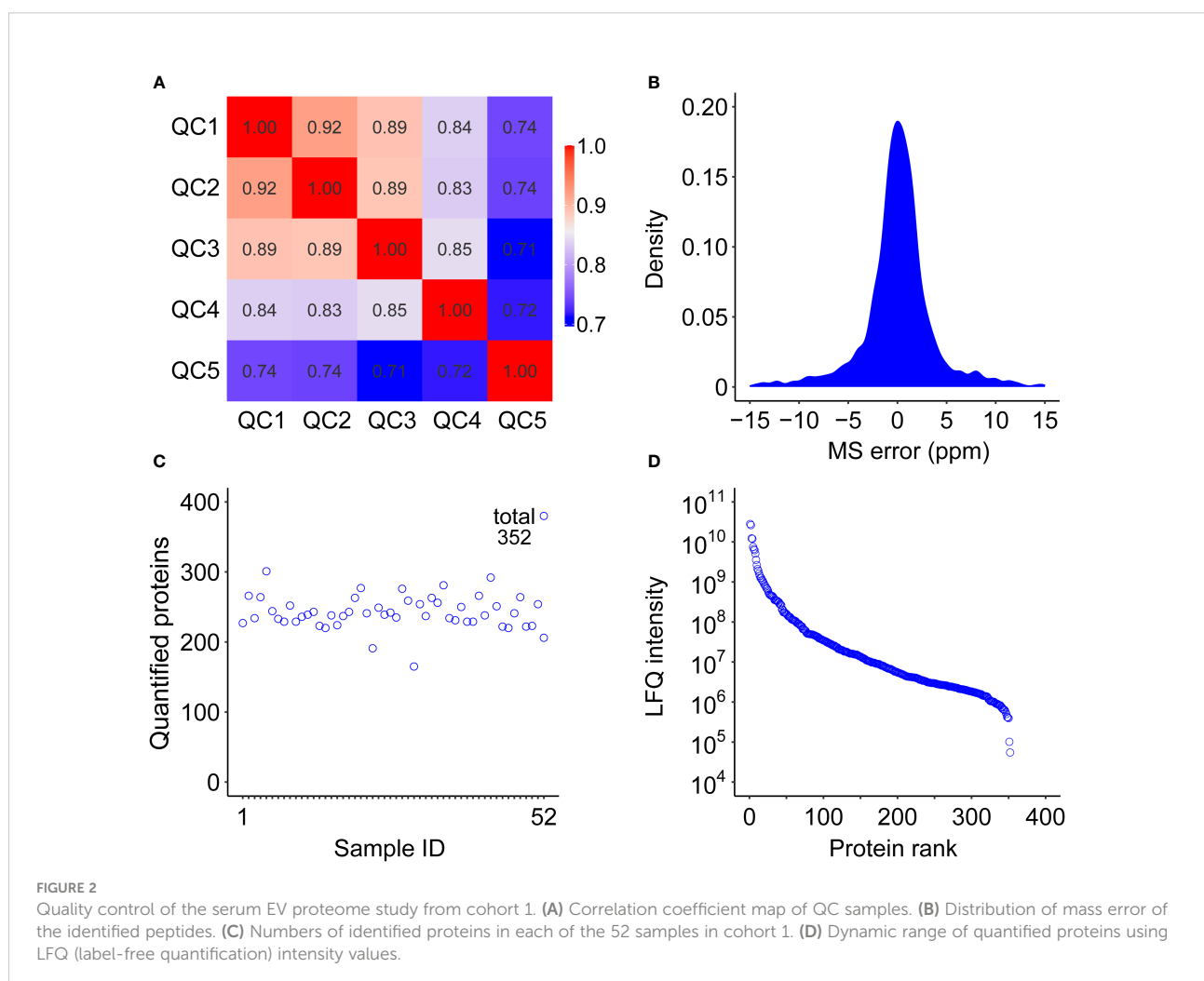
For construction of GAC diagnostic models, panels containing 2 to 6 proteins were randomly selected from the 15 candidate proteins through an exhaustive method, and resulting in a total of 9828 combinations. Using cohort 1 as the training set, we built a logistic regression classification model for each panel and calculate the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). To reduce the false negative predictions, we set the sensitivity greater than 0.9 and retained 2774 classifiers.

We then used Cohort 2 as the testing set to assess the classification accuracy of the models, and found a five-protein panel containing glycine N-acyltransferase (GSN), transferrin receptor protein 1 (TFRC), alpha-1-acid glycoprotein 1 (ORM1), haptoglobin (HP), and polymeric immunoglobulin receptor (PIGR) that showed the best classification performance. This panel of classifiers showed an accuracy, sensitivity, NPV and AUC of 0.97, 0.93 and 0.93 respectively in the training set (Figures 4E, F). In the validation set, the sensitivity, NPV and AUC were all above 0.8, indicating that the classifier maintained a good classification performance on new data set. The parameter of the logistic regression model is displayed in Table 3. Principal component analysis (Figure 4D) also confirmed the effectiveness of the five proteins in distinguishing GAC from healthy samples.

Discovery of a serum EV biomarker panel for diagnosis of advanced stage GAC

Since lymph node is a frequent tumor metastatic site, lymph node metastasis (LNM) is highly informative in selection of treatment strategies (13, 14). At present, the most commonly used blood-based diagnostic markers for GAC in clinic usage are the universal tumor markers carcinoembryonic antigen (CEA), carcinoembryonic antigen (CA19-9, CA72-4, CA24-2, CA50, CA125), and alpha-fetoprotein (AFP) (15–17). Therefore, we used CEA, CA19-9, and AFP as a panel to construct a classifier for diagnosis of advanced GAC, since these proteins are routinely measured in our patients. Based on the patient information in cohort 1, we combined patients in stage I and II as non-LNM group, and patients in stage III + IV as LNM group. In order to distinguish LNM from non-LNM groups, logistic regression was applied using cohort 1 as training set and cohort 2 as validation set, which resulted in a sensitivity, NPV and AUC of 0.66, 0.50 and 0.75 in validation set (Figures 5C, D, G and Table 4).

In contrast, we used our EV proteomic data to discover potential biomarkers for diagnosis of advanced GAC. We performed differential protein expression analysis comparing patients with stage III + IV (LNM) to that of stage I + II (non-LNM), as shown in Figure 5A. We identified 6 differentially expressed proteins including 3 up-regulated proteins and 3 down-regulated proteins (Figure 5B and Table 5), among which serum amyloid A (SAA1) and immunoglobulin heavy constant alpha 2 (IGHA2) play key roles in receptor-mediated endocytosis. These 6 proteins were used as candidate biomarkers, from which 2 to 5 proteins were randomly selected as panels to construct logistic regression classifiers. Fifty-six classifiers were constructed and trained with cohort 1 to evaluate the classification performance. Then 15 panels were retained with a cutoff of 0.9 for sensitivity and NPV. Applying these classifiers to cohort 2, we found an EV protein panel consisting of lysozyme (LYZ), SAA1, and coagulation factor XII (F12) that showed the best performance, with a sensitivity of 1, NPV of 1 and AUC of 0.83 (Figures 5E, F, H and Table 6).



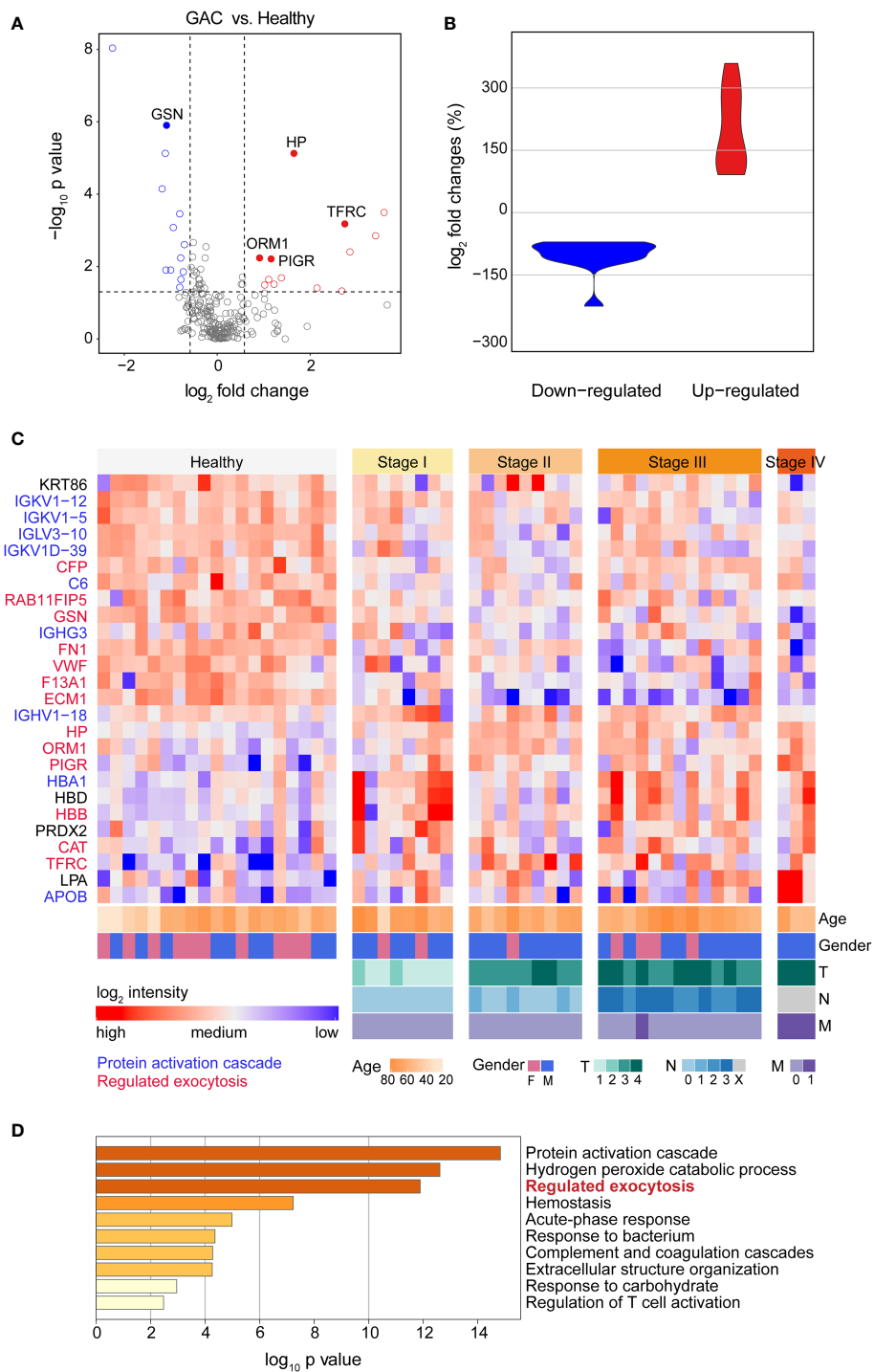


FIGURE 3 Proteomic profiles of serum EV proteins between healthy subjects and GAC patients. **(A)** Volcano plot of statistical significance value against \log_2 -fold change between GAC patients (N=33) and healthy controls (N=19) from cohort 1, showing differentially expressed proteins in blue (down) or red (up) circles. **(B)** Violin plot showing fold changes of up- and down-regulated proteins. **(C)** Heat map of 26 differentially expressed proteins between GAC patients and healthy subjects. Intensities of proteins were \log_2 -transformed. Different color in protein names indicates different biological processes derived from these proteins. **(D)** Gene Ontology (GO) analysis of differentially expressed proteins between GAC patients and healthy controls.

TABLE 2 Expression data of 15 candidate protein markers.

Gene name	Protein name	Fold change			p value		
		GC/HH	S1/HH	S2/HH	GC/HH	S1/HH	S2/HH
ECM1	Extracellular matrix protein 1	0.21	0.25	0.21	9.2E-09	7.8E-06	1.3E-06
GSN	Glycine N-acyltransferase	0.47	0.39	0.56	1.3E-06	5.1E-06	2.8E-04
FN1	Fibronectin	0.46	0.44	0.48	7.5E-06	2.2E-04	1.1E-04
VWF	von Willebrand factor	0.52	0.68	0.36	8.4E-04	1.3E-02	2.9E-03
IGKV1-5	Immunoglobulin kappa variable 1-5	0.58	0.62	0.54	5.8E-03	5.2E-02	5.8E-03
C6	Complement component C6	0.46	0.39	0.54	1.3E-02	2.2E-03	1.5E-01
IGHG3	Immunoglobulin heavy constant gamma 3	0.57	0.50	0.64	3.7E-02	3.3E-02	1.7E-01
HP	Haptoglobin	3.14	2.82	3.48	7.5E-06	1.4E-04	1.7E-04
HBB	Hemoglobin subunit beta	12.05	16.19	7.65	3.2E-04	4.8E-03	1.0E-03
TFRC	Transferrin receptor protein 1	6.71	8.42	4.87	6.6E-04	5.4E-03	4.2E-03
HBA1	Hemoglobin subunit alpha	7.24	7.80	6.66	4.0E-03	6.1E-02	2.3E-03
ORM1	Alpha-1-acid glycoprotein 1	1.88	1.82	1.95	5.8E-03	1.9E-02	2.0E-02
PIGR	Polymeric immunoglobulin receptor	2.23	1.96	2.52	6.2E-03	3.3E-02	1.4E-02
CAT	Catalase	2.60	3.00	3.00	2.0E-02	7.6E-02	5.2E-03
PRDX2	Peroxiredoxin-2	2.03	2.29	1.91	3.2E-02	3.1E-02	3.1E-02

GC, GAC patients; HH, healthy controls; S1, GAC patients in stage I + II; S2, GAC patients in stage III + IV.

Discussion

Based on analysis of protein expression in serum EV and exhaustive feature selection, this study identified a 5-protein panel consisting of GSN, PIGR, TFRC, ORM1, and HP that classifies GAC samples from healthy controls with high accuracy, warranted for further validation. These proteins have been reported in literature and have shown various connections to cancer. GSN is a tumor suppressor down-regulated in gastric cancer cells and gastric tumor tissues, and is a potential therapeutic target (18). TFRC is highly expressed in *H. pylori*-positive tissues and is a potential indicator for gastrointestinal metaplasia (19). The expression of PIGR is associated with the prognosis of gastric adenocarcinoma, esophageal carcinoma, endometrial carcinoma, hepatocellular carcinoma, and other tumors (20). ORM1 plays an important role in acute phase reaction and inflammatory response, and is highly expressed in plasma of multiple cancers, including gastric cancer (21). HP is the main glycoprotein in the acute phase response, and abnormal glycosylation is associated with several cancers and

inflammatory diseases (22). Our study showed that a logistic regression model utilizing these five proteins largely improved the accuracy of distinguishing GAC from healthy subjects.

Immunohistochemistry (IHC) provides valuable information on protein expression profiles in tumor tissues. Although no IHC experiment is conducted in this study, incidentally, we found that in Human Protein Atlas (HPA) database, there are IHC data on all of the marker proteins discovered in our study, in normal and gastric cancer tissues. The data shows that the expression of GSN (<https://www.proteinatlas.org/ENSG00000148180-GSN/pathology/stomach+cancer#ihc>) is down-regulated in gastric cancer tissues, while TFRC (<https://www.proteinatlas.org/ENSG00000072274-TFRC/pathology/stomach+cancer#ihc>) and PIGR (<https://www.proteinatlas.org/ENSG00000162896-PIGR/pathology/stomach+cancer#ihc>) are upregulated in gastric cancer tissues.

The model based on the panel of LYZ, SAA1, and F12 has the potential to identify advanced GAC. Comparing to known protein biomarkers in clinical use, the sensitivity, NPV and AUC of our panel are clearly improved. These three proteins have also been documented in literature. Elevated concentration of SAA1 is

TABLE 3 The five-protein logistic regression classifier for GAC diagnosis.

Protein (Human)	Coefficient	95% CI	p value
HP	3.745	(0.751, 7.545)	0.02601048
PIGR	-0.956	(-3.713, 1.668)	0.47226928
ORM1	-0.206	(-4.143, 3.334)	0.91019726
TFRC	2.070	(0.341, 4.608)	0.05118695
GSN	-3.817	(-8.158, -0.813)	0.03281620
Constant	-12.609	(-56.222, 33.111)	0.56981117

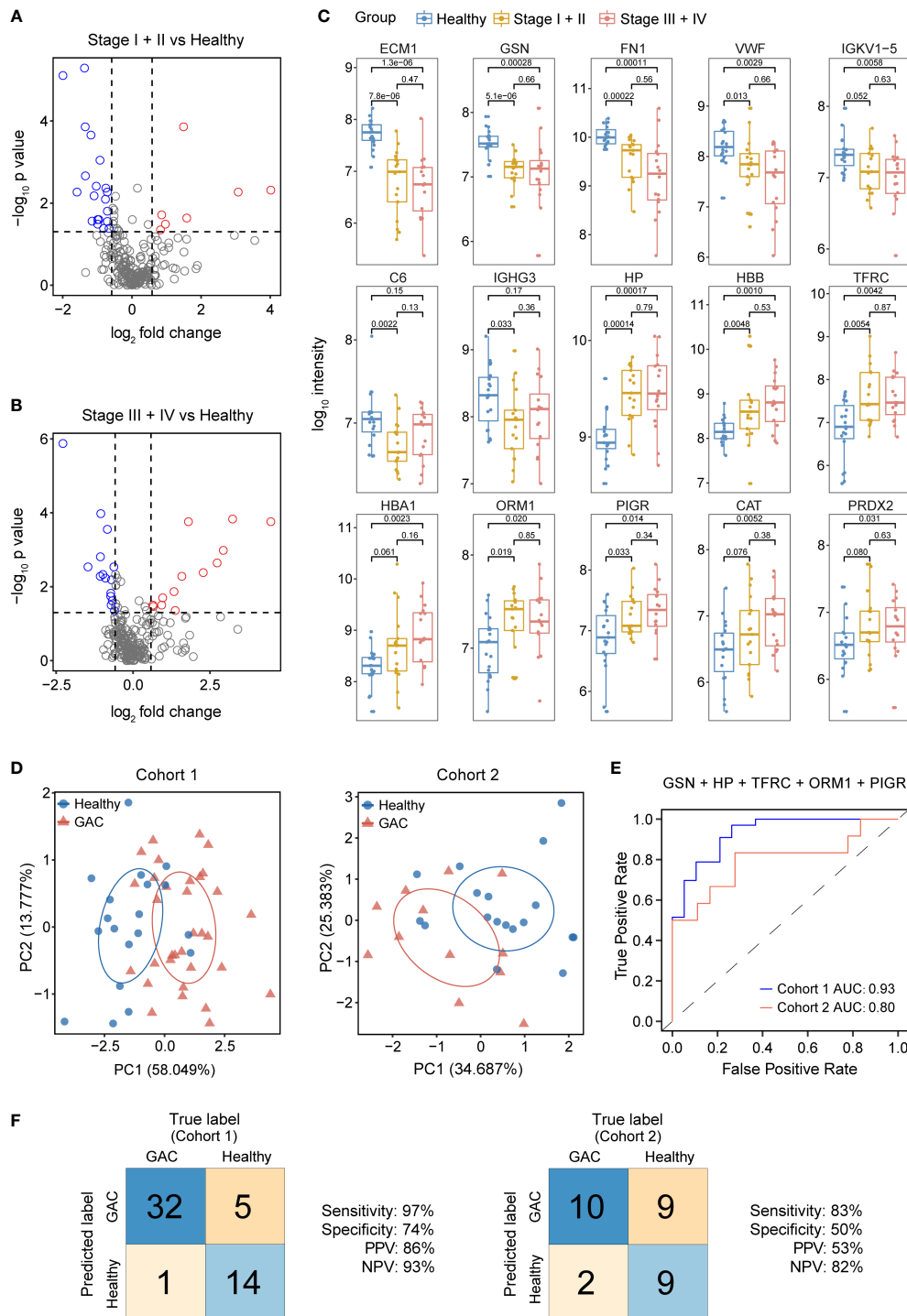


FIGURE 4

Discovery and validation of serum EV biomarkers for GAC diagnosis. (A) Volcano plot of significance value against log₂-fold change between stage I + II GAC patients (N=17) and healthy controls (N=19) from cohort 1, with significantly changed proteins shown in blue or red circles. (B) Volcano plot of significance value against log₂-fold change between stage III + IV GAC patients (N=16) and healthy controls (N=19) from cohort 1. (C) Box-whisker and dot plots showing distribution of intensity values of 15 candidate proteins across three groups from cohort 1: healthy controls (N=19), GAC stage I + II (N=17) and GAC stage III + IV (N=16). (D) Principal component (PC) analysis of healthy control and GAC samples from cohort 1 (left) and cohort 2 (right) using 5 candidate proteins (GSN, HP, TFRC, ORM1 and PIGR). (E) ROC curves of the 5-protein logistic regression classifier for GAC diagnosis in cohort 1 and cohort 2. AUC, area under the curve. (F) Classification error matrix of the 5-protein logistic regression classifier from E in cohort 1 and cohort 2. The number of samples is noted in each box.

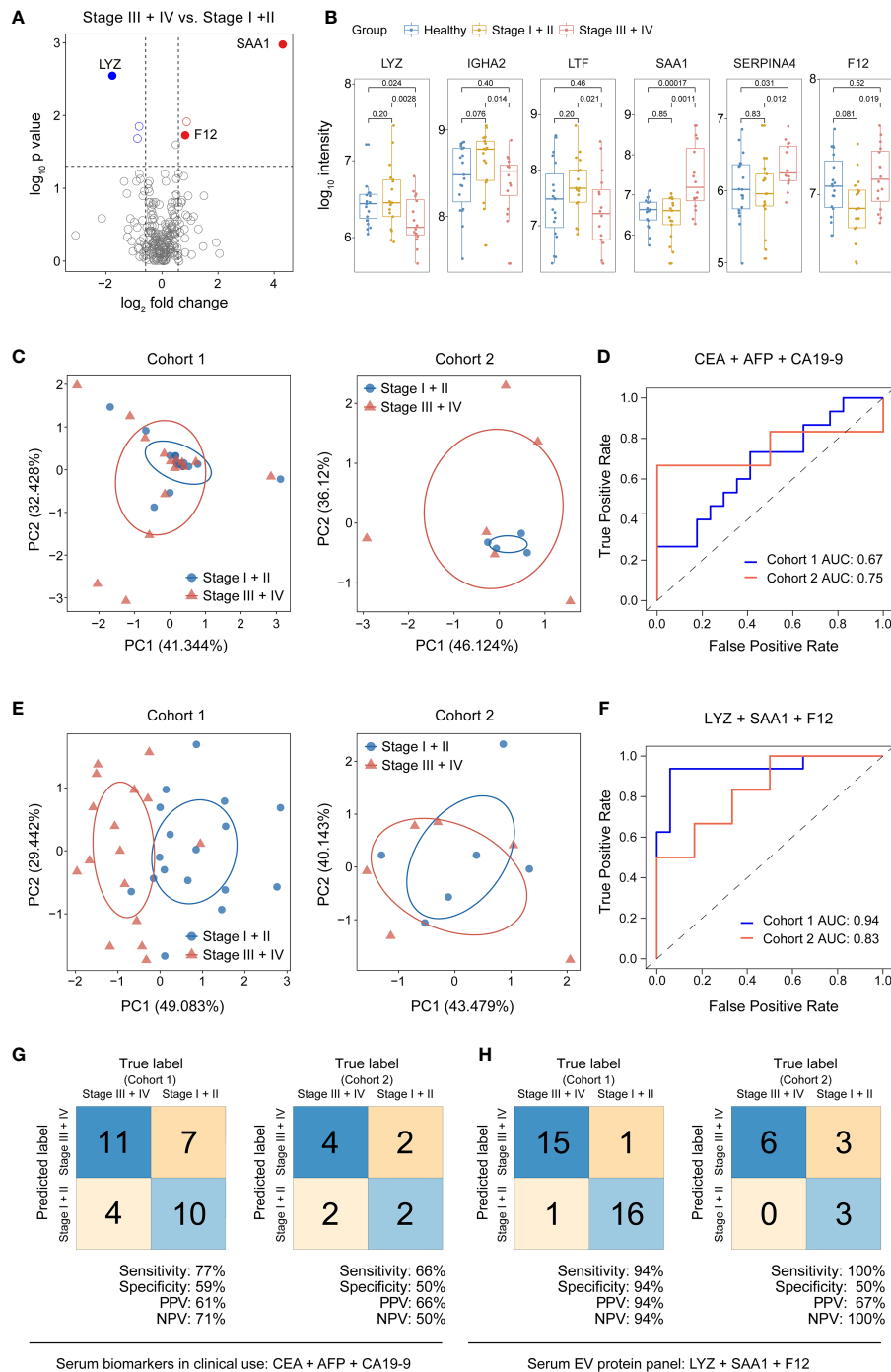


FIGURE 5

Discovery and validation of serum EV biomarkers for diagnosis of advanced stage GAC. (A) Volcano plot of significance values against log₂-fold change between stage III + IV (N=16) and stage I + II (N=17) GAC patients from cohort 1, showing significantly changed proteins in blue or red circles. (B) Box-whisker and dot plots showing distribution of intensity values of 6 candidate proteins across three groups from cohort 1: healthy controls (N=19), stage I + II (N=17) and stage III + IV GAC patients (N=16). (C) Principal component (PC) analysis of healthy and GAC samples from cohort 1 (left) and cohort 2 (right) using clinically used serum proteins (CEA, AFP and CA19-9) for diagnosis of advanced GAC. (D) ROC curves of the 3-protein (CEA, AFP and CA19-9) logistic regression classifier for diagnosis of advanced GAC. (E) Principal component (PC) analysis of healthy controls and GAC samples from cohort 1 (left) and cohort 2 (right) using 3 serum EV proteins (LYZ, SAA1 and F12) for diagnosis of advanced GAC. (F) ROC curves of the 3-protein (LYZ, SAA1 and F12) logistic regression classifier for diagnosis of advanced GAC. (G) Classification error matrix of the 3-protein logistic regression classifier from D in cohort 1 and cohort 2. (H) Classification error matrix of the 3-protein logistic regression classifier from F in cohort 1 and cohort 2. In both (G, H), the number of samples is noted in each box.

TABLE 4 The three-protein logistic regression classifier for diagnosis of advanced stage in GAC.

Protein (Human)	Coefficient	95% CI	p value
CA199	0.008	(-0.001, 0.025)	0.1674706
AFP	0.009	(-0.137, 0.135)	0.8883499
CEA	0.034	(-0.034, 0.12)	0.3503110
Constant	-0.796	(-2.093, 0.396)	0.2022071

TABLE 5 Expression data of 6 candidate protein markers for diagnosis of advanced GAC.

Gene name	Protein name	Fold change	p value
LYZ	Lysozyme	0.29	2.8E-03
IGHA2	Immunoglobulin heavy constant alpha 2	0.57	1.4E-02
LTF	Lactotransferrin	0.54	2.1E-02
SAA1	Serum amyloid A	19.79	1.1E-03
SERPINA4	Serpin family A member 4	1.83	1.2E-02
F12	Coagulation factor XII	1.77	1.9E-02

TABLE 6 The three-protein logistic regression classifier for diagnosis of advanced GAC.

Protein (Human)	Coefficient	95% CI	p value
SAA1	2.577	(0.814, 5.842)	0.03063643
F12	3.719	(0.301, 8.446)	0.06144730
LYZ	-3.818	(-9.187, -0.471)	0.07836439
Constant	-19.814	(-60.292, 12.35)	0.25727669

associated with occurrence, recurrence and survival of gastric cancer (21). LYZ is associated with incidence of colorectal cancer and lymph node metastasis (23). F12 is a plasma protease which promotes the production of inflammatory bradykinin by activating the kallikrein-kinin system (24). Nevertheless, the specificity and PPV of this model were dramatically decreased in testing data set. We could not rule out the possibility of overfitting due to limited sample size, and further studies with much increased sample size could be the key to address this issue.

In addition to the relatively small sample size, limitation of this study includes the apparent age discrepancy between patients and healthy controls in cohort 1. To rule out the possibility of protein expression changes due to aging, we removed some patients with extremely high ages in cohort 1 to make the median age matching that of the control group and performed differential protein expression analysis. The result shows that the protein markers in our model remains differentially expressed (STable 1). In addition, there were essentially no age difference between case and control groups in the validation cohort. Thus, we have strong reason to believe that the age difference between the two groups in cohort 1 was not the major contributing factor for the differential protein

expression, which is the basis for our selection of marker proteins. In conclusion, the abnormal expression of these marker proteins appears to have strong association with the growth and progression of GAC tumors, and has the predictive value for identifying GAC at early stage. Further validation of these proteins with increased sample size is warranted.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors. The mass spectrometry raw data are available *via* ProteomeXchange with identifier PXD027535.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Reviewing Board of The Ten's People's Hospital of Tongji University. The patients/participants

provided their written informed consent to participate in this study.

Author contributions

LL conceived the concept and directed the research. JC performed the experiments and YY analyzed the data. YT, and YZ contributed to some experimental data. LG and DL contributed to clinical samples. JJ contributed to clinical consultation. LL and JC wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Shanghai Natural Science Foundation grant to LL (19ZR1416400 and 19JC1411900); the East China Normal University National Natural Science Foundation of China grant No. 82273390 to LL (11300-120215-10321).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet* (2020) 396(10251):635–48. doi: 10.1016/S0140-6736(20)31288-5
- Nie Y, Wu K, Yu J, Liang Q, Cai X, Shang Y, et al. A global burden of gastric cancer: the major impact of China. *Expert Rev Gastroenterol Hepatol* (2017) 11(7):651–61. doi: 10.1080/17474124.2017.1312342
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492
- 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. doi: 10.3322/caac.21338
- Hamashima C, Shabana M, Okada K, Okamoto M, Osaki Y. Mortality reduction from gastric cancer by endoscopic and radiographic screening. *Cancer Sci* (2015) 106(12):1744–9. doi: 10.1111/cas.12829
- Sumiyama K. Past and current trends in endoscopic diagnosis for early stage gastric cancer in Japan. *Gastric Cancer* (2017) 20(Suppl 1):20–7. doi: 10.1007/s10120-016-0659-4
- Yang AP, Liu J, Lei HY, Zhang QW, Zhao L, Yang GH. CA72-4 combined with CEA, CA125 and CA19-9 improves the sensitivity for the early diagnosis of gastric cancer. *Clin Chim Acta* (2014) 437:183–6. doi: 10.1016/j.cca.2014.07.034
- Huang T, Song C, Zheng L, Xia L, Li Y, Zhou Y. The roles of extracellular vesicles in gastric cancer development, microenvironment, anti-cancer drug resistance, and therapy. *Mol Cancer* (2019) 18(1):62. doi: 10.1186/s12943-019-0967-5
- Fu M, Gu J, Jiang P, Qian H, Xu W, Zhang X. Exosomes in gastric cancer: roles, mechanisms, and applications. *Mol Cancer* (2019) 18(1):41. doi: 10.1186/s12943-019-1001-7
- Kahroba H, Hejazi MS, Samadi N. Exosomes: from carcinogenesis and metastasis to diagnosis and treatment of gastric cancer. *Cell Mol Life Sci* (2019) 76(9):1747–58. doi: 10.1007/s00018-019-03035-2

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.1051450/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Proteome data assessment of the two cohort. (A) Venn diagram of identified proteins in training cohort (cohort 1) and testing cohort (cohort 2). (B) Venn diagram of the identified proteins with the Vesiclepedia database. (C) Gene ontology analysis of shared EV proteins between training cohort and testing cohort.

SUPPLEMENTARY FIGURE 2

Venn diagrams of differentially expressed proteins obtained in multi-group differential protein expression analysis for cohort 1. (A) Venn diagram of down-regulated proteins. (B) Venn diagram of up-regulated proteins. GC, GAC patients; HH, healthy individuals; S1, GAC patients of stage I + II; S2, GAC patients of stage III + IV.

- Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. *Science* (2020) 367(6478). doi: 10.1126/science.aau6977
- Egerton JD, MacLean B, Johnson R, Xuan Y, MacCoss MJ. Multiplexed peptide analysis using data-independent acquisition and skyline. *Nat Protoc* (2015) 10(6):887–903. doi: 10.1038/nprot.2015.055
- Hu B, El Hajj N, Sittler S, Lammert N, Barnes R, Meloni-Ehrig A. Gastric cancer: Classification, histology and application of molecular pathology. *J Gastrointest Oncol* (2012) 3(3):251–61. doi: 10.3978/j.issn.2078-6891.2012.021
- Zhang X, Li M, Chen S, Hu J, Guo Q, Liu R, et al. Endoscopic screening in Asian countries is associated with reduced gastric cancer mortality: A meta-analysis and systematic review. *Gastroenterology* (2018) 155(2):347–54.e9. doi: 10.1053/j.gastro.2018.04.026
- Sturgeon CM, Duffy MJ, Hofmann BR, Lamerz R, Fritsche HA, Gaarenstroom K, et al. National academy of clinical biochemistry laboratory medicine practice guidelines for use of tumor markers in liver, bladder, cervical, and gastric cancers. *Clin Chem* (2010) 56(6):e1–48. doi: 10.1373/clinchem.2009.133124
- Acharya A, Markar SR, Matar M, Ni M, Hanna GB. Use of tumor markers in gastrointestinal cancers: Surgeon perceptions and cost-benefit trade-off analysis. *Ann Surg Oncol* (2017) 24(5):1165–73. doi: 10.1245/s10434-016-5717-y
- Shimada H, Noie T, Ohashi M, Oba K, Takahashi Y. Clinical significance of serum tumor markers for gastric cancer: A systematic review of literature by the task force of the Japanese gastric cancer association. *Gastric Cancer* (2014) 17(1):26–33. doi: 10.1007/s10120-013-0259-5
- Wang HC, Chen CW, Yang CL, Tsai IM, Hou YC, Chen CJ, et al. Tumor-associated macrophages promote epigenetic silencing of gelsolin through DNA methyltransferase 1 in gastric cancer cells. *Cancer Immunol Res* (2017) 5(10):885–97. doi: 10.1158/2326-6066.CIR-16-0295
- Hamedl Asl D, Naserpour Farivar T, Rahmani B, Hajmanoochehri F, Emami Razavi AN, Jahanbin B, et al. The role of transferrin receptor in the helicobacter pylori

pathogenesis; l-ferritin as a novel marker for intestinal metaplasia. *Microb Pathog* (2019) 126:157–64. doi: 10.1016/j.micpath.2018.10.039

20. Fristedt R, Gaber A, Hedner C, Nodin B, Uhlen M, Eberhard J, et al. Expression and prognostic significance of the polymeric immunoglobulin receptor in esophageal and gastric adenocarcinoma. *J Transl Med* (2014) 12:83. doi: 10.1186/1479-5876-12-83

21. Subbannayya Y, Mir SA, Renuse S, Manda SS, Pinto SM, Puttamallesh VN, et al. Identification of differentially expressed serum proteins in gastric adenocarcinoma. *J Proteomics* (2015) 127(Pt A):80–8. doi: 10.1016/j.jprot.2015.04.021

22. Lee J, Hua S, Lee SH, Oh MJ, Yun J, Kim JY, et al. Designation of fingerprint glycopeptides for targeted glycoproteomic analysis of serum haptoglobin: Insights into gastric cancer biomarker discovery. *Anal Bioanal Chem* (2018) 410(6):1617–29. doi: 10.1007/s00216-017-0811-y

23. Yin HR, Zhang L, Xie LQ, Huang LY, Xu Y, Cai SJ, et al. Hyperplex-MRM: A hybrid multiple reaction monitoring method using mTRAQ/iTRAQ labeling for

multiplex absolute quantification of human colorectal cancer biomarker. *J Proteome Res* (2013) 12(9):3912–9. doi: 10.1021/pr4005025

24. Nickel KF, Long AT, Fuchs TA, Butler LM, Renne T. Factor XII as a therapeutic target in thromboembolic and inflammatory diseases. *Arterioscler Thromb Vasc Biol* (2017) 37(1):13–20. doi: 10.1161/ATVBAHA.116.308595

COPYRIGHT

© 2022 Gu, Chen, Yang, Zhang, Tian, Jiang, Zhou and Liao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.