A novel strategy to identify candidate diagnostic and prognostic biomarkers for gastric cancer

Lei Liu[1*†] , Honglin Pang[2†], Qiao He[3†], Biran Pan[4], Xiaobin Sun[5], Jing Shan[5], Liping Wu[5], Kaiwen Wu[2], Xue Yao[2] and Yuanbiao Guo[1*]

Abstract

Background: Gastric cancer (GC) is one of the most common cancer worldwide. It is essential to identify non-invasive diagnostic and prognostic biomarkers of GC. The aim of the present study was to screen candidate biomarkers associated with the pathogenesis and prognosis of GC by a novel strategy.

Methods: The expression level of gene higher in cancer than in adjacent non-cancer tissue was defined as "positive", and the top 5% genes with "positive rate" were filtered out as candidate diagnostic biomarkers in three Gene Expression Omnibus (GEO) datasets. Further, a prognostic risk model was constructed by multivariate Cox regression analysis in GEO dataset and validated in The Cancer Genome Atlas (TCGA). The expression level of candidate biomarkers was determined in serum and serum-derived exosomes of GC patients. Moreover, the effect of biomarkers in exosomes on migration of GC cells was analyzed by transwell assay.

Results: Ten candidate biomarkers (*AGT*, *SERPINH1*, *WNT2*, *LIPG*, *PLAU*, *COL1A1*, *MMP7*, *MXRA5*, *CXCL1* and *COL11A1*) were identified with efficient diagnostic value in GC. A prognostic gene signature consisted of *AGT*, *SERPINH1* and *MMP7* was constructed and showed a good performance in predicting overall survivals in TCGA. Consistently, serum levels of the three biomarkers also showed high sensitivity and specificity in distinguishing GC patients from controls. In addition, the expression level of the three biomarkers were associated with malignant degree and decreased after surgery in GC patients. Moreover, the expression level of AGT and MMP7 in exosomes correlated positively with serum level. The exosomes derived from serum of GC patients can promote migration of SGC-7901 cells. After neutralized the expression level of three proteins in exosomes with antibodies, the migration of GC cells was obviously suppressed.

Conclusions: Our findings provided a novel strategy to identify diagnostic biomarkers based on public datasets, and suggested that the three-gene signature was a candidate diagnostic and prognostic biomarker for patients with GC.

Keywords: Gastric cancer, Diagnosis, Prognosis, Serum biomarkers, Exosomes

Background

Gastric cancer (GC) is one of the most common cancer worldwide and is responsible for over 1,000,000 new cases in 2018 and an estimated 783,000 deaths, making it the fifth most frequently diagnosed cancer and the third leading cause of cancer death. The incidence rates of GC are markedly elevated in Eastern Asia [1]. The 5-year

*Correspondence: liulei@swtju.edu.cn; Guo.ybiao@yahoo.com
†Lei Liu, Honglin Pang and Qiao He contributed equally to this work
[1] Medical Research Center, The Third People's Hospital of Chengdu, The Affiliated Hospital of Southwest Jiaotong University, 82 Qinglong Road, Chengdu 610031, Sichuan, China
Full list of author information is available at the end of the article

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 2 of 15

survival of GC is low because more than 80% of patients are diagnosed at an advanced stage and lose the opportunity for the most effective surgical treatment [2, 3]. Thus, screening and early diagnosis is the most effective way to improve the survival rate.

The diagnosis of gastric cancer relies on endoscopy and biopsy, but the invasiveness of these means leads to poor compliance and potential risks to patients. Blood markers play an important role in disease screening and diagnosis due to its economic, convenient and minimally invasive characteristics. In gastric cancer, several blood markers have been used for diagnosis, determination of the clinical stage, evaluation of treatment responses and screening for recurrence after successful therapy [4]. Although many biomarkers for GC including carbohydrate antigen 72–4 (CA72-4), alphafetoprotein, carbohydrate antigen 12–5 (CA12-5), SLE, BCA225, hCG and pepsinogenI/II have been reported, carcinoembryonic antigen (CEA) and carbohydrate antigen 19–9 (CA19-9) are still the most frequently used biomarkers in clinical practice for GC [5]. However, low rates of sensitivity and specificity prevent the use of any of these serum markers in diagnosis of GC [4]. As current non-invasive tests are insufficient for GC screening or diagnosis, the discovery of alternative biomarkers is necessary, especially blood biomarkers.

In recent years, microarray and high throughput sequencing technologies have been used for discovering diagnostic or prognostic biomarkers of cancers [6]. Usually, based on the difference of genes expressed in tumor tissue and normal controls, genes with the most difference fold change (FC) are considered for candidate biomarkers. However, the most critical feature of a diagnostic marker is that it presents universally in almost all the cancer patients but not at the healthy individuals, although the marker is probably not the most different molecule between the tumor and healthy. Therefore, we think the positive ratio of a tumor biomarker, which is expressed higher in cancer than in adjacent non-cancer tissue, presents in the cancer patients determines its sensitivity. Based on this thought, a novel strategy was established for mining candidate blood biomarkers from the GEO datasets.

In the present study, three GEO datasets with large GC cohorts, which contain paired cancer and adjacent non-cancer tissues, were employed for candidate biomarkers screening using the mentioned mining method above. Then, we performed gene ontology (GO) term and find 16 proteins were located in extracellular region. The diagnostic value of the genes was validated using receiver operating characteristic (ROC) curve in TCGA and GEO datasets, and 10 genes were picked out for further analyses. Based on the Cox regression analysis, three candidate genes (*AGT*, *SERPINH1* and *MMP7*) were identified which associated with the overall survival (OS) of GC patients. The expression level of three candidate biomarkers in serum were associated with tumor malignant degree and decreased after surgery in GC patients. In additional, the expression level of three candidate biomarkers in exosomes correlated positively with serum level. Moreover, the three proteins probably promote progress of GC through exosomes. This study identified novel blood biomarkers in diagnosis and prognosis of GC.

Methods

Datasets and preprocessing

The gene expression data used in this study were obtained from GEO and TCGA database. Three GEO datasets (GSE66229, GSE27342, GSE63089) were selected to screen the candidate genes for diagnosis of GC, which consists of paired tumor tissue and adjacent non-cancer tissue with large sample size. The datasets TCGA and GSE54129 were employed to validate the efficiency of the candidate genes in diagnosis of GC. GSE15459 was selected to construct the prognosis risk model of GC and the risk model was tested in TCGA database. The Raw CEL files of the 5 GEO datasets were downloaded and the Robust Multichip Average (RMA) algorithm was used for background adjustment, quantile normalization and log-transformation. In addition, the gene expression profiles of TCGA database was transformed to the base-2 logarithm for further analysis.

Receiver operating characteristic (ROC) curve analysis

The datasets TCGA and GSE54129 were used for validating the efficiency of the candidate genes in diagnosis of GC. ROC curve analysis was performed to evaluate the efficiency of the candidate genes in diagnosis of GC. The area under the curve (AUC) value was calculated and used for evaluating the diagnostic value of these genes for GC diagnosis. The candidate genes with AUC value greater than 0.6 in the both datasets were regarded as acceptable for diagnosis and selected for further study [7, 8].

Construction of prognostic risk model

GSE15459 was used as the training dataset to construct the prognosis risk model. The patients without overall survival (OS) data were excluded, and finally 190 GC patients were included for analysis. First, the univariate Cox regression analysis was used for determining the association between the expression level of the ten candidate genes and overall survival (OS) of GC patients, and the genes with *P*-value < 0.05 were selected for multivariate Cox regression analysis to identify prognostic genes.

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 3 of 15

The coefficients of each prognostic gene were performed to construct the prognostic risk model. The formula is as follows: risk score = coefficient of $gene_1 \times$ expression of $gene_1$ + coefficient of $gene_2 \times$ expression of $gene_2$ + ... coefficient of $gene_n \times$ expression of $gene_n$.

Survival analysis

The risk model was used for validating prognosis of GC patients in datasets GSE15459 and TCGA. The risk score was calculated in each patient of GC and the patients were classified into low-risk and high-risk groups based on the median risk score. The Kaplan-Meier survival analysis was performed for the patients in different groups. Furthermore, the time-dependent ROC curve was performed by R package "survival ROC" to assess the predictive accuracy of the prognostic risk model. The AUC was calculated to measure the predictive ability of the candidate genes for time-dependent cancer death.

Patients

To validate the diagnosis and prognosis value of three genes in GC, the blood samples from GC patients and healthy volunteers were collected at the Sichuan Cancer Hospital between October 2018 and July 2020. The GC patients were newly diagnosed without received any chemotherapy or radiation therapy and histologically confirmed by two different pathologists. The subjects in healthy controls group showed no cancer and other serious disease in the physical examination. The postoperative blood samples were collected from 25 patients around 1 month after surgery. The present study was approved by the Ethics Committee of Sichuan Cancer Hospital. Informed consent was obtained from all of the participants before the start of the study.

Blood sample collection and detection

Venous blood was collected in the morning before breakfast from all of the subjects. Blood samples were allowed to clot at room temperature for 30 min, and then centrifuged at 2000g for 10 min at 4 °C to separate serum. Then the serum was aliquoted and stored at − 80 °C for further study. The concentration of candidate markers AGT, SERPINH1 and MMP7 in serum were determined by a quantitative sandwich enzyme-linked immunosorbent assay (ELISA) kit (USCN, Wuhan, China) according to the manufacturer's instructions.

Exosomes isolation from serum

The exosomes were isolated from serum by precipitation or size exclusion chromatography (SEC). For precipitation, ExoQuick precipitation was carried out according to manufacturer's instructions (System Biosciences). Briefly, one milliliter of serum was thawed and centrifuged at 1500g for 10 min. The supernatant was collected and recentrifuged at 10,000g for 30 min. The supernatant was then incubated with ExoQuick for 30 min at 4 °C. The ExoQuick/serum sample was then centrifuged at 1500g for 30 and the pellet was resuspended in 200 μL of PBS. For SEC, 500 μL of serum was thawed and recentrifuged at 10,000g for 30 min. The clarified serum was overlaid on qEV size exclusion columns (Izon, New Zealand) followed by elution with PBS. The eluate was collected in 13 sequential fractions of 1 mL. Each fraction was aliquoted and stored at − 80 °C for subsequently study.

Identification of exosomes

The concentration and size distribution of particles in collected fractions was measured by Nanoparticle tracking analysis with ZetaView (Particle Metrix, Meerbusch, Germany). Each sample was diluted 1:1000 to 1:10,000 in PBS to get the ideal measurement concentration (50–200 particles/frame). The instrument set to a specific analysis parameter: Minimum brightness: 20, Maximum area: 1000, Minimum Area: 5, Laser Wavelength: 488 nm. Each measurement scan at 11 different positions, and after capture, the videos were analyzed by the software ZetaView (version 8.05.11).

Exosomes were visualized using transmission electron microscopy (TEM) (JEM-1400, JEOL, Japan). Briefly, freshly isolated exosomes were put on a copper grid and keep at room temperature for 5 min, then the grids were stained with 2% (v/v) uranyl acetate and the exosome samples were examined immediately.

Isolated exosomes were detected for the presence of exosomal markers TSG101 and CD63 by Western blotting. The Protein concentration was detected by bicinchoninic acid assay (BCA assay) (ThermoScientific, IL, USA). The equal amounts (30ug) of precipitated exosomes or equal volume (20 ul) of SEC fractions were loaded on 12% SDS-PAGE gels. Proteins were transferred to PVDF membrane (Millipore, MA, USA) and the membranes were incubated with primary antibodies TSG101 and CD63 (Proteintech, Wuhan, China) at 4 °C overnight. After incubated with horseradish peroxidase-conjugated species-specific secondary antibodies, the membranes were visualized with enhanced chemiluminescence reagent (Millipore, MA, USA).

Cell culture and uptake of exosomes

The GC cell line SGC-7901 were purchased from the cell bank of the Chinese Academy of Sciences (Shanghai, China) and cultured in RPMI-1640 medium supplemented with 10% FBS, 100 U/ml penicillin and 100 mg/ml streptomycin. Cells were cultured at 37 °C under a humidified atmosphere including 5% CO_2. For tracking exosomes, purified exosomes were incubated with 1uM

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 4 of 15

fluorescent dye DiI (AAT Bioquest, CA, USA) according to the manufacturer's guidelines. After washed with PBS, the DiI-labeled exosomes were co-cultured with SGC7901 for 24 h. The uptake of exosomes was observed by laser confocal microscope (SpinSR10, OLYMPUS, Japan).

Cell migration assay

Cell migration assay was performed using a 24-well migration chamber (Corning, NY, USA) with or without exosomes. In the exosomes group, exosomes derived from serum of GC patients were treated with neutralizing antibodies (anti-AGT, anti- SERPINH1, anti-MMP7 respectively or combined) or isotype control for 30 min, and then washed with PBS. The pretreated exosomes (30ug) and 5×10^4 SGC-7901 cells were co-cultured in the top chamber. The bottom chamber was filled with 600 µl medium containing 20% exosomes-free FBS. After incubation for 48 h, the cells remaining at the upper surface of the membrane were removed with a cotton swab, and those that adhered to the lower surface were fixed with paraformaldehyde and stained with crystal violet. Five fields were selected randomly to count the cells invaded through the membrane and imaged under a microscope (Leica Microsystems, Wetzlar, Germany) with magnification of 100 times. Each experiment was performed three times independently and the mean number of invaded cells were used for Transwell assay assessment.

Statistical analysis

The univariate and multivariate Cox regression analysis was used for determining predictive factors for GC prognosis. The Mann–Whitney U test was used for analyzing the difference of concentration of candidate markers between control and GC groups. The ROC curve was used for evaluating the diagnosis value of candidate markers. The Pearson's chi-square test was used for analyzing the associations between concentration of candidate markers and clinicopathological characteristics. The paired Student's t-tests was used for comparison of concentration of candidate markers between pre- and postoperative samples in GC. Pearson correlation assay was used for analyzing the expression level of three genes between serum and exosomes. All analysis was performed using SPSS 20.0 software (IBM Corporation, NY, USA) or GraphPad Prism 7.0 (GraphPad Software Inc., CA, USA). $P < 0.05$ was considered as significant difference.

Results

Screening the candidate genes from GEO datasets of gastric cancers

The most critical feature of a diagnostic blood marker is that it presents universally in almost all cancer patients but not at all or at lower levels in healthy individuals. Therefore, the expression level of gene higher in cancer than in adjacent non-cancer tissue were define as positive, and we believe that the positive rate of candidate biomarkers in cancer patients is as important as its difference in concentration between cancer patients and healthy people. A series of research were conducted following the schematic diagram in Fig. 1 in the hope of finding new blood candidate biomarkers.
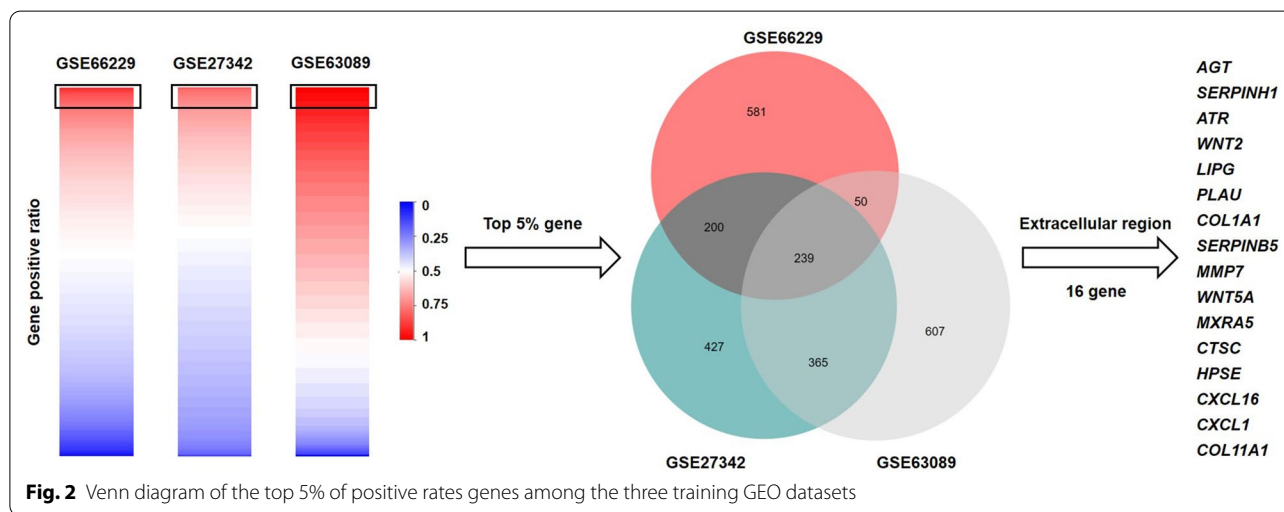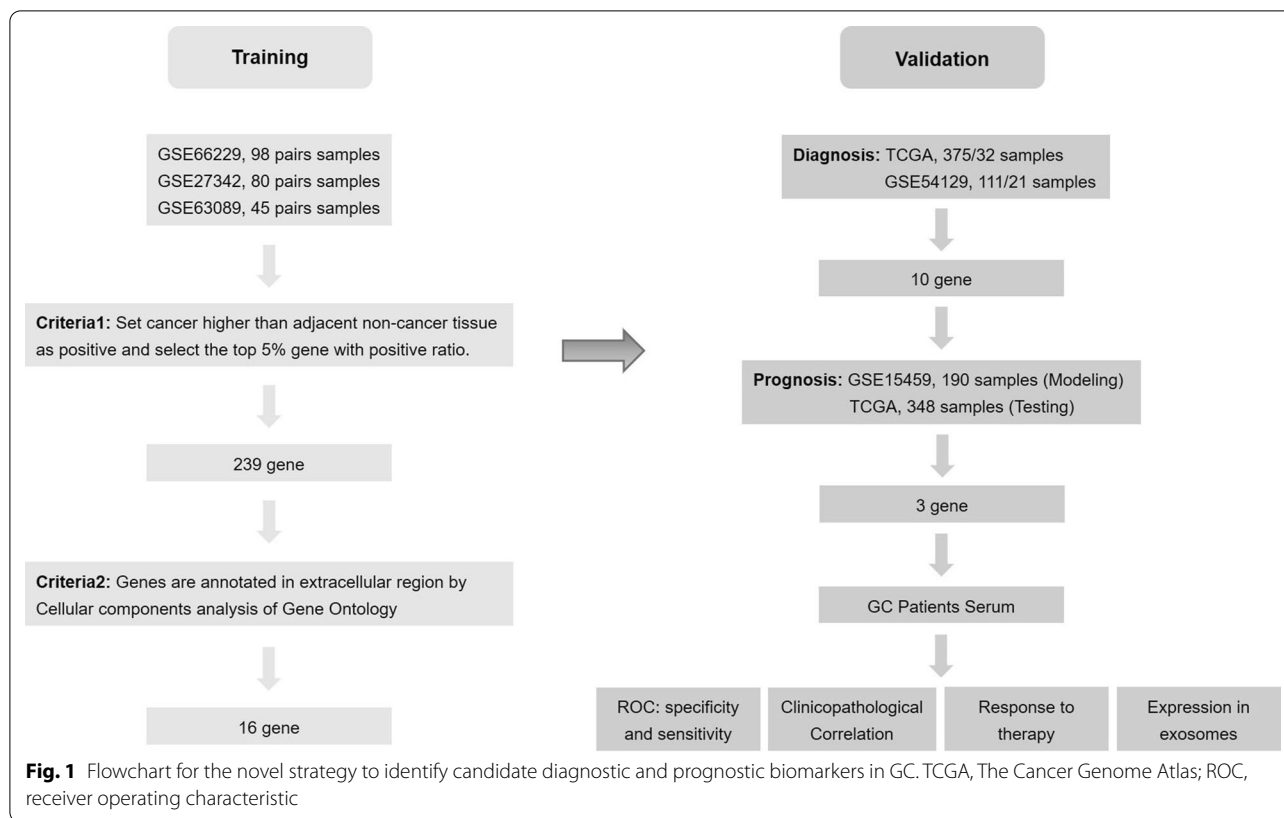
Three GEO datasets with large sample sizes, GSE66229 (n = 98), GSE27342 (n = 80) and GSE63089 (n = 80), were employed for candidate biomarkers screening. After ranking genes by their positive rates in cancer cases, the top 5% genes of each dataset were filtered out. GSE66229, GSE27342 and GSE63089, respectively, yielded 1070, 1231 and 1261 genes that were universally expressed in cancer tissue, and 239 overlapping genes were identified. To uncover the blood biomarkers, we focused on the extracellular proteins. The 239 genes were annotated by Cellular Components analysis of Gene Ontology (GO), and 16 proteins were located in extracellular region: *AGT, SERPINH1, ATR, WNT2, LIPG, PLAU, COL1A1, SERPINB5, MMP7, WNT5A, MXRA5, CTSC, HPSE, CXCL16, CXCL1* and *COL11A1* (Fig. 2).

Diagnostic evaluation of 16 candidate genes

To validate the efficiency of the novel strategy for biomarkers screening, the efficiency of 16 candidate genes in distinguish GC from NC tissues was analyzed in TCGA (375 GC tissues and 32 normal tissues). The results showed that all of the 16 candidate genes performed excellent. To analyze the diagnostic efficiency of these genes in GC, the ROC curve analysis was performed with dataset GSE54129 (111 GC tissues and 21 normal tissues). There are ten candidate genes with AUC value greater than 0.6 in GSE54129 dataset: *AGT, SERPINH1, WNT2, LIPG, PLAU, COL1A1, MMP7, MXRA5, CXCL1* and *COL11A1* (Table 1). These ten genes were used for further constructing of cox regression model.

Assess the prognostic risk of the biomarker candidates in the training dataset

To explore whether the 10 genes have the potential to predict prognosis, the correlation between gene expression level and survival time were analyzed by univariate and multivariate Cox proportional hazards regression in training dataset GSE15459 (n = 190). A total of eight genes significantly correlated with survival time ($P < 0.05$)

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 5 of 15

Fig. 1 Flowchart for the novel strategy to identify candidate diagnostic and prognostic biomarkers in GC. TCGA, The Cancer Genome Atlas; ROC, receiver operating characteristic

Fig. 2 Venn diagram of the top 5% of positive rates genes among the three training GEO datasets

were identified by the univariate Cox proportional hazards regression model. Among these genes, *AGT, SER-PINH1* and *MMP7* displayed significant prognostic values by the multivariate Cox proportional hazards regression model (Table 2). The three genes then were used as biomarker panel for the diagnostic and prognostic indication of gastric cancers.

Testing the prognostic risk models of the three-gene panel in the training and validating datasets

To investigate whether the three-gene panel could provide an accurate prediction of OS in GC patients, the prognostic risk scores of the panel was formulated in the training dataset (GSE15459) and validating dataset (TCGA) based on the coefficient of the three genes in multivariate analysis: Risk score $=$ (0.180 $\times$ expression

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 6 of 15

Table 1 Area under the curve of candidate biomarkers in distinguishing controls from GC patient**s**

Gene	TCGA					GSE54129				
	AUC	Std. error	P-value	95% CI		AUC	Std. error	P-value	95% CI	
				Lower	Upper				Lower	Upper
AGT	0.717	0.040	< 0.001	0.639	0.795	0.617	0.049	0.089	0.522	0.713
SERPINH1	0.913	0.020	< 0.001	0.873	0.952	0.957	0.016	< 0.001	0.925	0.989
ATR	0.908	0.026	< 0.001	0.856	0.959	0.328	0.049	0.013	0.232	0.425
WNT2	0.946	0.014	< 0.001	0.919	0.973	0.991	0.007	< 0.001	0.978	1.000
LIPG	0.829	0.032	< 0.001	0.766	0.892	0.890	0.039	< 0.001	0.813	0.966
PLAU	0.892	0.018	< 0.001	0.858	0.927	0.913	0.025	< 0.001	0.864	0.962
COL1A1	0.930	0.015	< 0.001	0.901	0.959	0.972	0.014	< 0.001	0.945	0.999
SERPINB5	0.661	0.059	0.003	0.545	0.776	0.403	0.057	0.159	0.292	0.514
MMP7	0.787	0.038	< 0.001	0.713	0.861	0.733	0.047	0.001	0.641	0.826
WNT5A	0.719	0.057	< 0.001	0.607	0.831	0.550	0.051	0.465	0.451	0.650
MXRA5	0.823	0.025	< 0.001	0.775	0.871	0.782	0.040	< 0.001	0.703	0.860
CTSC	0.706	0.050	< 0.001	0.609	0.804	0.545	0.058	0.512	0.432	0.658
HPSE	0.750	0.053	< 0.001	0.646	0.853	0.469	0.060	0.652	0.351	0.587
CXCL16	0.796	0.037	< 0.001	0.723	0.870	0.462	0.054	0.582	0.356	0.568
CXCL1	0.830	0.035	< 0.001	0.762	0.898	0.783	0.046	< 0.001	0.693	0.874
COL11A1	0.924	0.016	< 0.001	0.893	0.955	0.911	0.025	< 0.001	0.862	0.959

AUC area under curve, *Std.* standard, *CI* confidence interval

Table 2 Association between the ten gene signature and overall survival of GC patients in GSE15459

Gene	Univariate analysis		Multivariate analysis		
	HR (95% CI)	P-value	HR (95% CI)	P-value	Coefficient
AGT	1.269 (1.092–1.476)	0.002	1.197 (1.036–1.384)	0.015	0.180
SERPINH1	1.678 (1.289–2.184)	< 0.001	1.722 (1.296–2.289)	< 0.001	0.544
WNT2	1.185 (1.006–1.395)	0.042			
LIPG	1.176 (0.973–1.421)	0.094			
PLAU	1.406 (1.163–1.699)	< 0.001			
COL1A1	1.481 (1.187–1.849)	0.001			
MMP7	1.164 (1.061–1.278)	0.001	1.144 (1.044–1.254)	0.004	0.135
MXRA5	1.245 (1.026–1.511)	0.027			
CXCL1	0.958 (0.853–1.075)	0.464			
COL11A1	1.188 (1.068–1.322)	0.002			

HR hazard ratio, *CI* confidence interval

value of *AGT*) + (0.544 × expression value of *SER-PINH1*) + (0.135 × expression value of *MMP7*). In the training dataset, a total of 190 patients were divided into a high-risk group (*N*=95) and a low-risk group (*N*=95) according to the median risk score (Fig. 3A). The GC patients with high-risk scores had lower OS rates than those with low-risk scores (Fig. 3B). Moreover, tumor tissues obtained from patients with high-risk scores tended to express high level of prognostic genes (*AGT*, *SERPINH1* and *MMP7*) (Fig. 3C). As expected, GC patients with high-risk scores had lower OS than those with low-risk scores (*P*< 0.0001) (Fig. 3D). The prognostic gene signature presented a good performance in survival prediction, as the AUC was 0.727 for 5-year OS (Fig. 3E).

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 7 of 15

To confirm our findings, the prognostic power of the three-gene panel was further evaluated in the validating dataset. According to the same risk formula and using the median risk score as the cutoff point, patients in TCGA dataset were divided into high-risk group (n = 174) and low-risk group (n = 174). Consistent with the findings in training dataset, patients in the high-risk group suffered significantly poorer OS than those in the low-risk group ($P < 0.05$). The predictive power of prognosis was consistent with those observed in the training dataset (Fig. 3F–J). These results indicated that the three-gene panel showed a good performance in prognosis assessment.

Prognostic risk model of the three-gene panel is independently associated with OS of GC patients in datasets

Next, we evaluated whether the three-gene panel was an independent predictor of GC patient's survival. The univariate and multivariate Cox regression analysis were performed in GSE15459 dataset, and the univariate Cox regression analysis showed that the prognostic risk score and the pathological stage significantly correlated with OS of GC patients ($P < 0.05$). Furthermore, multivariate Cox regression analysis was performed using the pathological stage and the risk scores. The results showed that the prognostic risk score and the pathological stage independently correlated with OS of GC patients ($P < 0.05$). In TCGA dataset, the univariate Cox regression analysis showed that the three-gene panel risk score, tumor invasion, lymph node invasion, metastasis and the pathological stage significantly correlated with OS of GC patients ($P < 0.05$). Furthermore, the multivariate Cox regression analysis showed that the tumor invasion and metastasis significantly correlated with OS ($P < 0.05$), and there is a similar tendency in the three-gene panel risk score (HR = 1.296, 95% CI: 0.918–1.829, P = 0.141) (Table 3). These results demonstrate that the three-gene panel prognostic risk model could be independently used for predicting OS in GC patients.

Performance of circulating levels of the three proteins in the diagnosis of GC patients

In order to assess the diagnostic capability of the three proteins, the circulating levels of them were quantified in 132 GC patients and 86 controls, whose age and sex were matched. According to the grade and WHO classification, there were 99 cases of well or moderately differentiated subtype; 24 cases of poorly differentiated subtype and 9 cases of signet ring cell subtype. The ELISA results showed that the concentration of AGT, SERPINH1 and MMP7 in GC patients (34.51 ± 22.35 ng/mL, 733.90 ± 204.65 pg/mL and 4.67 ± 1.30 ng/mL, respectively) are significantly higher than those in the control group (25.90 ± 16.08 ng/mL, 604.41 ± 185.04 pg/mL and 3.80 ± 2.27 ng/mL, respectively) (Fig. 4A–C). The diagnostic performance of the three proteins was evaluated by ROC curve. The AUCs of ROC of AGT, SERPINH1 and MMP7 are 0.6078, 0.7279 and 0.706 for distinguishing GC patients from controls, respectively. The optimal diagnostic cutoff value of AGT, SERPINH1 and MMP7 are 37.10 ng/mL, 677.00 pg/mL and 3.66 ng/mL respectively, at which with sensitivity (42.42%, 56.82% and 78.03%) and specificity (82.35%, 81.93% and 64.71%). Further, the three-protein panel showed more effective performance with sensitivity (72.73%) and specificity (71.60%) (Fig. 4D–G). These results indicate that the three-protein panel in serum displayed a high potential in the diagnosis of GC.

Correlations between clinicopathological characteristics and circulating levels of the three proteins in GC patients

Then, serum protein levels of the three genes were detected to study their clinicopathological values. GC patients were divided into 2 groups with high or low concentration of AGT, SERPINH1 and MMP7 by their median values, individually. High-AGT patients showed larger tumor size ($\chi^2 = 9.008$, $P = 0.003$), more depth of tumor invasion ($\chi^2 = 4.281$, $P = 0.039$) and more advanced TNM stage ($\chi^2 = 5.961$, $P = 0.0015$) than low-AGT patients. High-SERPINH1 patients also showed more depth of tumor invasion ($\chi^2 = 5.979$, $P = 0.014$) and more advanced TNM stage ($\chi^2 = 9.854$, $P = 0.002$) than the low-SERPINH1 ones. Consistently, compared to the low-MMP7 patients, patients with high MMP7 showed larger tumor size ($\chi^2 = 7.013$, $P = 0.008$), more depth of tumor invasion ($\chi^2 = 5.979$, $P = 0.014$) and more advanced TNM stage ($\chi^2 = 5.961$, $P = 0.015$). When evaluating the prognosis of GC patients by combination of the three proteins in serum, it still showed good performance (Table 4).

(See figure on next page.)

Fig. 3 Kaplan–Meier estimates of the survival for patients using the three-gene signature. **A**, **F** Three-gene signature risk score distribution in the training and validation datasets. The black dotted line represents the median three-gene signature risk score cutoff dividing patients into low-risk and high-risk groups. **B**, **G** GC patients' survival status in the training and validation datasets. **C**, **H** Heatmap of the three gene expression profiles in the training and validation datasets. Rows represent genes, and columns represent patients. Red: high expression; Blue: low expression. **D**, **I** Kaplan–Meier analysis for the overall survival of GC patients in the training and validation datasets. **E**, **J** The ROC curves for predicting OS in GC patients by the three-gene signature risk score
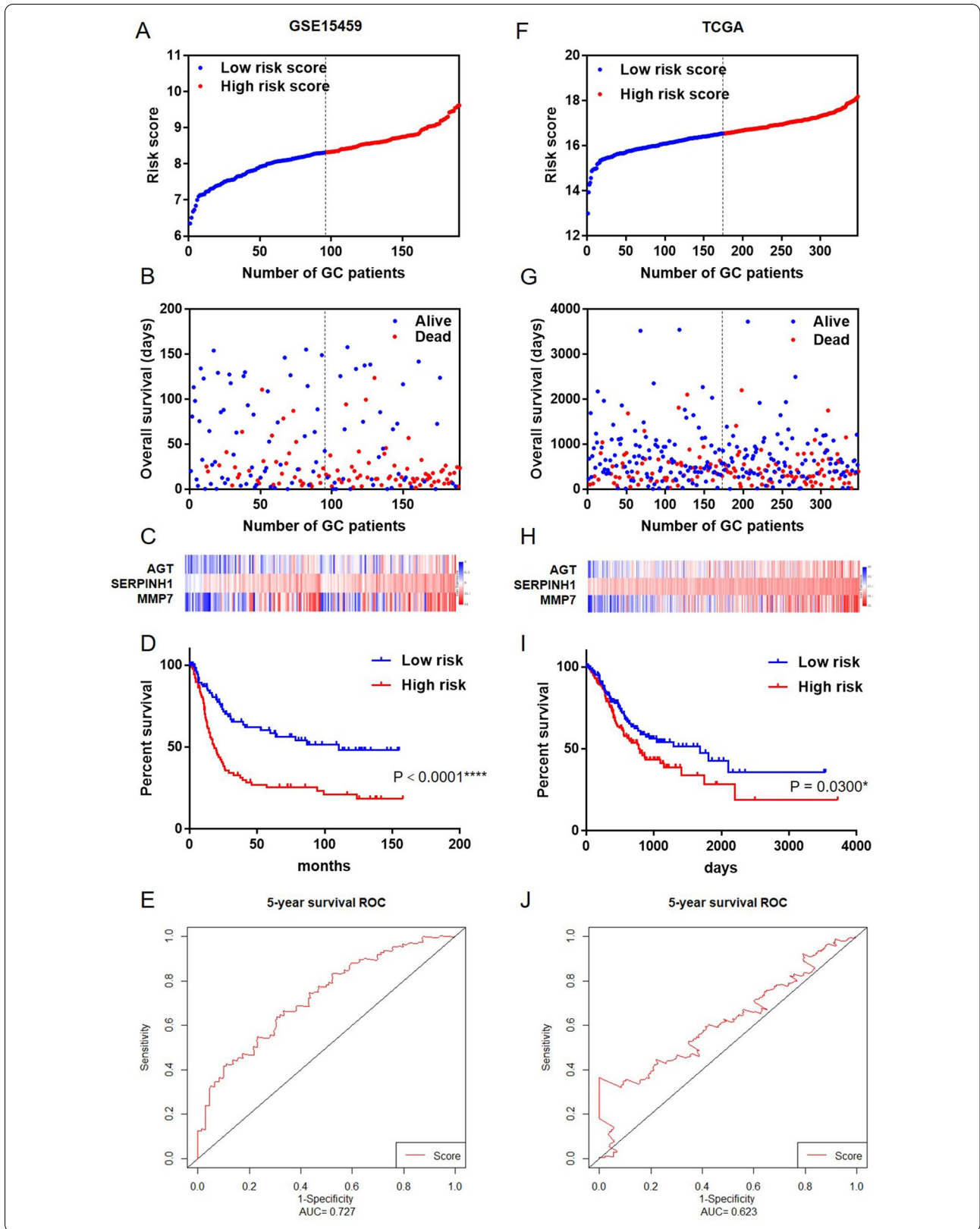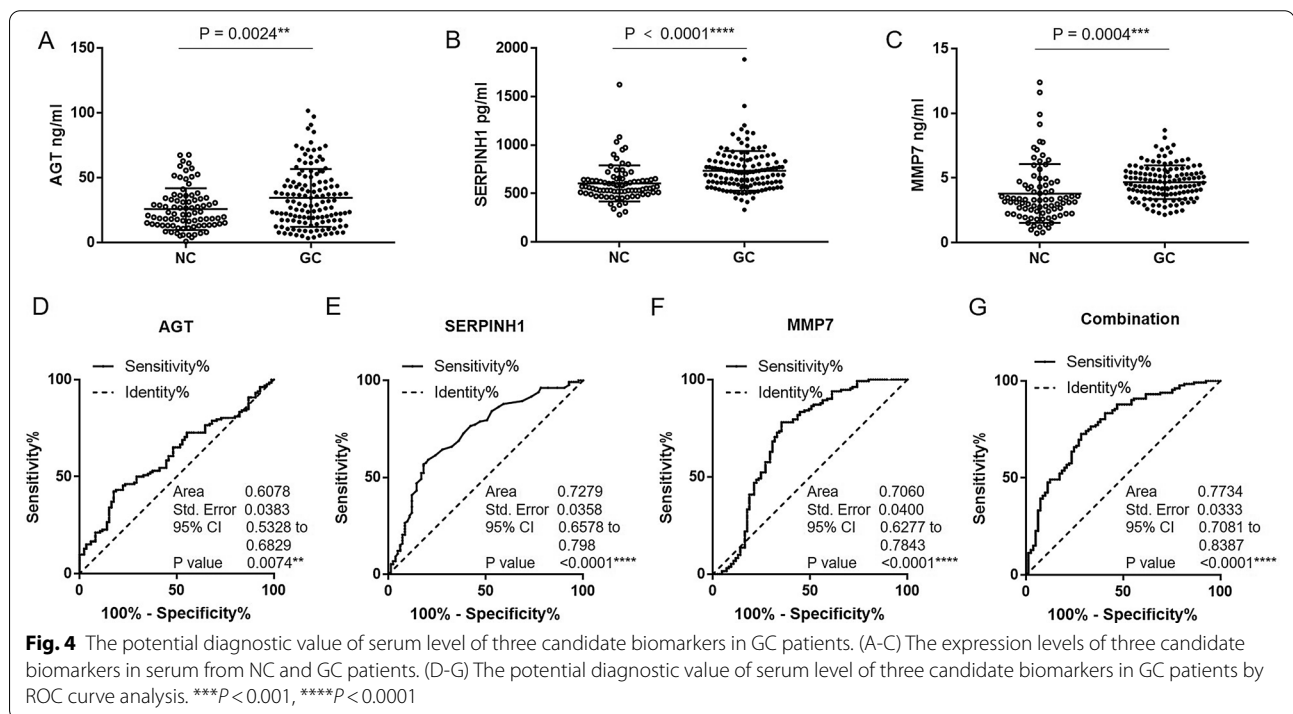
Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 8 of 15

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 9 of 15

Table 3 Univariate and multivariate Cox regression analyses in GSE15459 and TCGA dataset

Variables	Univariate analyses		Multivariate analyses	
	HR (95% CI)	P-value	HR (95% CI)	P-value
GSE15459				
Three gene risk score	2.470 (1.620–3.767)	< 0.001	1.806 (1.178–2.771)	0.007
Gender	0.713 (0.462–1.101)	0.127		
Age	0.966 (0.645–1.445)	0.865		
Stage	6.520 (3.597–11.816)	< 0.001	5.684 (3.110–10.389)	< 0.001
TCGA				
Three gene risk score	1.446 (1.034–2.022)	0.031	1.296 (0.918–1.829)	0.141
Gender	0.917 (0.644–1.306)	0.630		
Age	0.933 (0.669–1.301)	0.682		
Tumor invasion	1.880 (1.218–2.902)	0.004	2.218 (1.405–3.500)	0.001
Lymph node invasion	1.727 (1.145–2.605)	0.009	1.482 (0.884–2.485)	0.135
Metastasis	1.931 (1.067–3.496)	0.030	1.893 (1.021–3.508)	0.043
Pathological stage	1.518 (1.082–2.128)	0.016	1.196 (0.768–1.861)	0.428
Histological grade	1.313 (0.924–1.866)	0.129		

HR hazard ratio, *CI* confidence interval

Fig. 4 The potential diagnostic value of serum level of three candidate biomarkers in GC patients. (A-C) The expression levels of three candidate biomarkers in serum from NC and GC patients. (D-G) The potential diagnostic value of serum level of three candidate biomarkers in GC patients by ROC curve analysis. ***$P < 0.001$, ****$P < 0.0001$

These data suggested that the circulating levels of three proteins could be a potential prognostic.

Circulating levels of three proteins were significantly decreased in postoperative GC patients

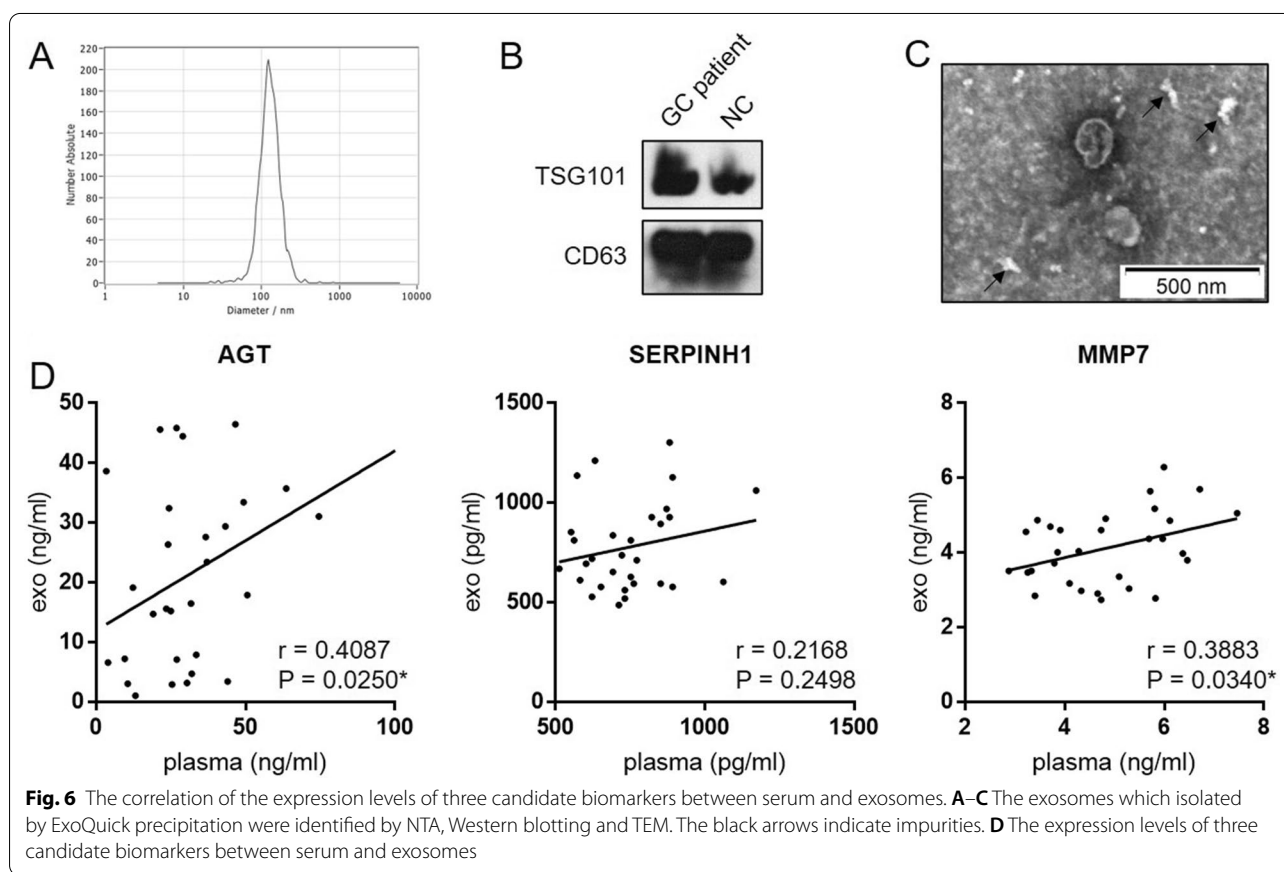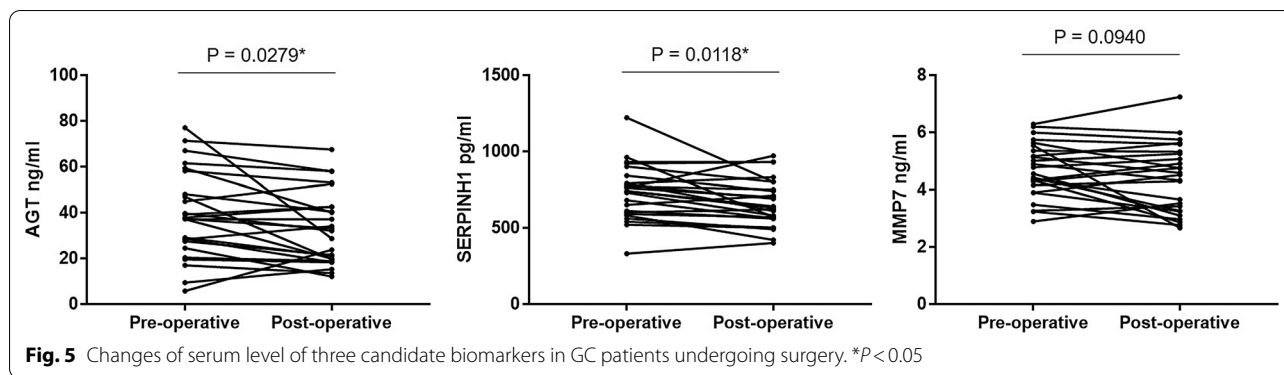For monitoring the efficiency of treatment in GC patients, the changes of circulating levels of three proteins between paired pre- and post-operative serum specimens were analyzed in 25 patients with GC. The serum level of AGT and SERPINH1 were significantly decreased one month after radical operation ($P < 0.05$). There is a similar tendency in the serum MMP7, however, which with no statistical significance (Fig. 5).

Table 4 Correlation between serum level of the three biomarkers and clinicopathological characteristics in GC patients

Characteristics	AGT level			SERPINH1 level			MMP7 level			Score 1 level			Score 2 level		
	Low (n=66)	High (n=66)	P-value	Low (n=66)	High (n=66)	P-value	Low (n=66)	High (n=66)	P-value	Low (n=66)	High (n=66)	P-value	Low (n=66)	High (n=66)	P-value
Gender															
Male	45	45	1.00	49	41	0.135	47	43	0.455	49	41	0.135	51	38	0.196
Female	21	21		17	25		19	23		17	25		19	23	
Age (years)															
≤65	37	42	0.375	41	38	0.594	40	39	0.859	41	38	0.594	42	36	0.909
>65	29	24		25	28		26	27		25	28		28	25	
Tumor size (cm)															
≤4	47	30	0.003	42	35	0.217	46	31	0.008	42	35	0.217	50	27	0.002
>4	19	36		24	31		20	35		24	31		20	34	
Tumor invasion															
T1–T2	26	15	0.039	27	14	0.014	27	14	0.014	27	14	0.014	30	11	0.002
T3–T4	40	51		39	52		39	52		39	52		40	50	
Lymph node invasion															
Negative	22	22	0.368	27	22	0.368	27	22	0.368	27	22	0.368	30	19	0.167
Positive	44	44		39	44		39	44		39	44		40	42	
Metastasis															
Negative	57	54	0.475	58	53	0.234	59	52	0.096	58	53	0.234	63	47	0.044
Positive	9	12		8	13		7	14		8	13		7	14	
TNM stage															
I–II	38	24	0.015	40	22	0.002	38	24	0.015	40	22	0.002	42	20	0.002
III–IV	28	42		26	44		28	42		26	44		28	41	

Score 1 = (0.180 × concentration of AGT) + (0.544 × concentration of SERPINH1) + (0.135 × concentration of MMP7), patients were divided into a high-risk group and a low-risk group according to the median of score 1

Score 2, two or more of the three proteins are high-expressed, the patients were divided into a high-risk group; otherwise, the patients were divided into a low-risk group
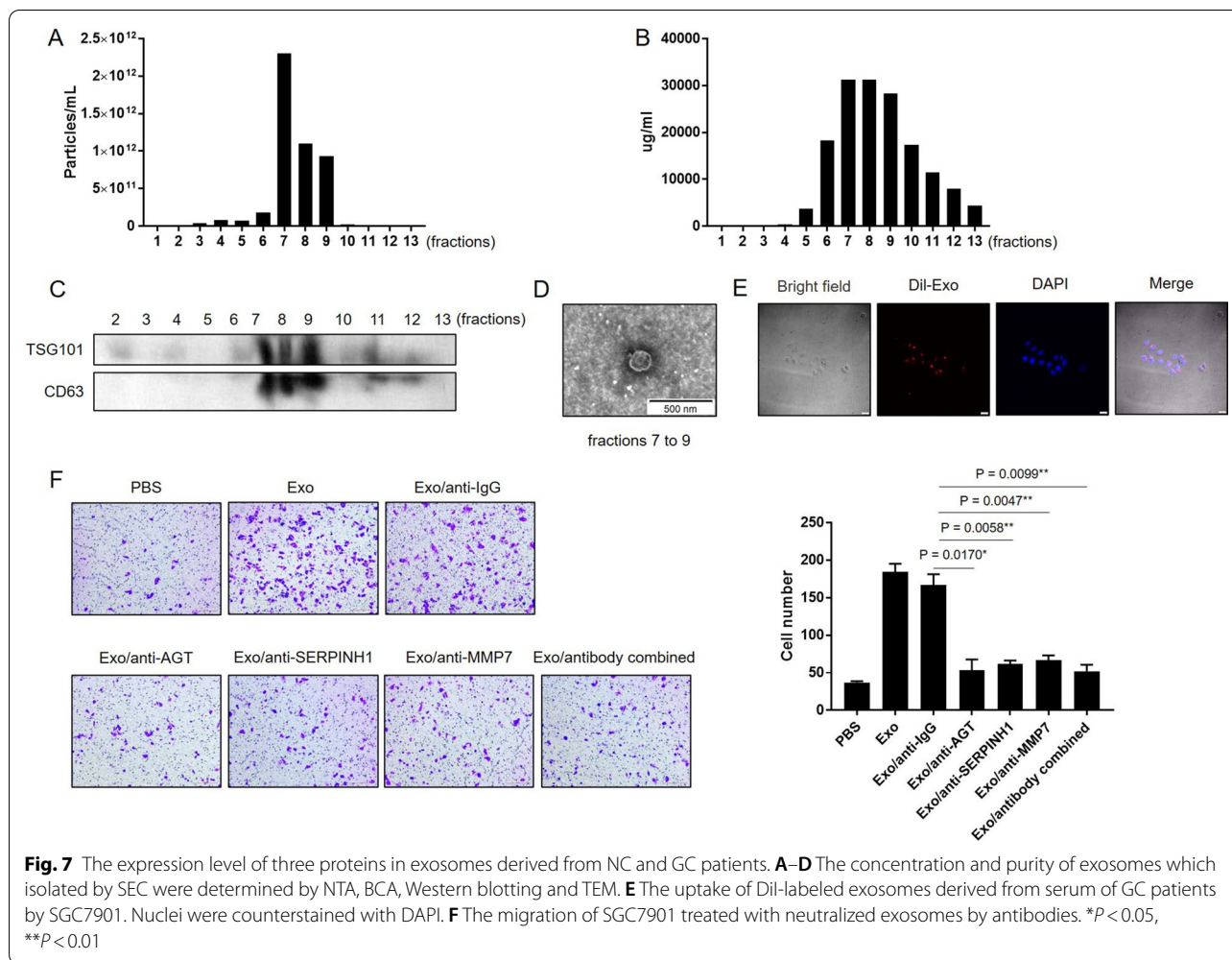
Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 11 of 15

Fig. 5 Changes of serum level of three candidate biomarkers in GC patients undergoing surgery. *$P < 0.05$

Fig. 6 The correlation of the expression levels of three candidate biomarkers between serum and exosomes. **A**–**C** The exosomes which isolated by ExoQuick precipitation were identified by NTA, Western blotting and TEM. The black arrows indicate impurities. **D** The expression levels of three candidate biomarkers between serum and exosomes

The expression level of three proteins in serum-derived exosomes from GC patients

Given the three genes located in extracellular region by GO annotation, we wanted to know their protein levels in serum exosomes. The exosomes derived from serum were isolated by ExoQuick precipitation and identified by NTA, Western blotting and TEM. As expected, the mean diameter of exosomes was about 120 nm (Fig. 6A). The exosomal markers (TSG101 and CD63) were also detected (Fig. 6B) and the exosomes were clearly visible

by TEM (Fig. 6C). The exosomal SERPINH1, AGT and MMP7 were thereafter measured by ELISA. The exosomes level of AGT and MMP7, except SERPINH1, correlated positively with the serum level (Fig. 6D).

The three proteins promote migration of GC cells through exosomes

To further study the role of the three proteins in exosomes, more purer exosomes were isolated by SEC method. The concentration of particles in the SEC

Fig. 7 The expression level of three proteins in exosomes derived from NC and GC patients. **A–D** The concentration and purity of exosomes which isolated by SEC were determined by NTA, BCA, Western blotting and TEM. **E** The uptake of Dil-labeled exosomes derived from serum of GC patients by SGC7901. Nuclei were counterstained with DAPI. **F** The migration of SGC7901 treated with neutralized exosomes by antibodies. *$P < 0.05$, **$P < 0.01$

fractions were detected by NTA. It showed that the highest concentration of particles emerged in fractions 7–9 (Fig. 7A). In addition, the data of bicinchoninic acid assay also found the highest protein content appeared in fractions 7–9 (Fig. 7B), where the presence of exosomes were further confirmed by the exosomal markers TSG101 and CD63 (Fig. 7C). Therefore, the fractions 7–9 were collected for TEM detection and for subsequent study. The TEM scan showed that the purified exosomes were enriched in fractions 7–9 (Fig. 7D). Since the three proteins are associated with tumor invasion and TNM stage, the effect of three proteins on the migration of GC cells were studied. The exosomes were pretreated with or without neutralizing antibodies to the three proteins, and co-cultured with SGC-7901 cells. Then the migration of GC cells was observed. Immunofluorescence staining assay indicated that exosomes can transfer into SGC-7901 cells (Fig. 7E). The exosomes derived from serum of GC patients can promote the migration of SGC-7901 cells. After neutralized the expression level of three proteins

in exosomes with antibodies, the migration of SGC-7901 cells was obviously suppressed ($P < 0.05$) (Fig. 7F). These results indicated that the three proteins probably promote progress of GC through exosomes.

Discussion
Although the gold standard of diagnosis of GC is endoscopy and biopsy, the invasiveness of this method leads to poor compliance and cause a great economic burden. Blood markers play an important role in screening and aiding in diagnosis of GC due to its economic, convenient and minimally invasive characteristics. The traditional molecular biomarkers for GC include CEA [9] and CA 19–9 [10]. CEA was identified in 1965 and applied in the clinical diagnosis of GC in 1980 [11]. However, the circulating level of CEA is not a GC specific marker and is generally increased in a lot of cancers [12], and also be artificially affected by other factors, such as smoking [13]. CA 19–9 is also a commonly used in pancreatic cancer [14], colorectal cancer [15] and gastric cancer [16].

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 13 of 15

However, the low rates of sensitivity and specificity of the markers made it is insufficient for GC screening or diagnosis. Therefore, the discovery of alternative biomarkers is necessary, especially blood biomarkers.

The current usual strategy for screening biomarkers from omics data is based on the different expression folds between cancer patients and normal controls. However, this method easily overlooks many low abundant genes that are widely expressed in cancers but not in the healthy. In fact, the sensitivity of a biomarker is primarily decided by its universal expression in cancer. Therefore, in this study we established a biomarker-mining strategy based on the positive rate of a gene in all individuals with cancer, which balanced the coverage and difference of gene expression. To finding the candidate blood biomarkers, we focused on the extracellular molecules annotated by Cellular Components analysis of Gene Ontology (GO). Sixteen genes were screened out and confirmed in TCGA dataset. Further, ten of them showed high diagnostic efficiency in GSE54129 dataset. These results indicated that the novel strategy for screening biomarkers is reliable. Ideal biomarker has possible excellent properties for diagnosis, therapeutic and prognostic evaluation. We construct a prognostic risk models, which contains three genes *AGT*, *SERPINH1* and *MMP7*, based on the OS of GC patients by multivariate Cox proportional hazards regression. The three candidate biomarkers in the prognostic risk model presented a good performance in survival prediction in the training and validating datasets, and the prognostic risk model is independently associated with OS of GC patients.

To compare the efficiency of the novel method and traditional method in screening of biomarkers, the differential gene expression analysis was performed. Based on the comparability with the 'positive ratio'-based approach and the traditional differential expression analysis, the genes that met the cutoff criteria of a fold change > 2 and an adjusted P-value < 0.05 were considered differentially expressed genes (DEGs). GSE66229, GSE27342 and GSE63089, respectively, yielded 1070, 211 and 1261 genes that were met the criteria, and 101 overlapping genes were identified. The 101 genes were annotated by Cellular Components analysis of Gene Ontology (GO), and 17 proteins were located in extracellular region: *COL12A1, COL5A2, CXCL5, AGT, LAMC2, GPRC5A, PLBD1, BGN, WNT5A, WISP1, PLAU, CXCL1, CXCL16, CHI3L1, ADAMTS12, CCL18* and *SERPINB5*. The diagnostic efficiency of these genes was assessed by ROC curve in the GSE54129 dataset. Then, there are 11 candidate genes with AUC value greater than 0.6: *COL12A1, COL5A2, CXCL5, AGT, BGN, WISP1, PLAU, CXCL1, CHI3L1, ADAMTS12* and *CCL18*. The 11 genes were used for constructing of cox regression model in training

dataset GSE15459. Among the 11 genes, *AGT* and *BGN* displayed significant prognostic values. Risk score = $(0.184 \times$ expression value of *AGT*) + $(0.487 \times$ expression value of *BGN*). Further, when evaluating the prognostic power of this model in validating dataset TCGA, patients in the high-risk group suffered the tendency of poorer OS than those in the low-risk group, however, there was on statistical difference ($P = 0.063$). These results indicated that the novel strategy may screen out more biomarkers with high efficiency.

AGT was rarely studied in GC. Previous studies focus on the association between AGT gene polymorphism and Helicobacter pylori infection-related GC or high-salt diet-related GC [17, 18]. Recent study reports that AGT was aberrantly methylated and associated with prognosis in GC [19]. Our study revealed that AGT was high expressed in serum and exosomes derived from GC patients. It with sensitivity 43.15% and specificity 82.35% for distinguishing GC patients from controls. High-AGT patients showed larger tumor size, more depth of tumor invasion and more advanced TNM stage. Further, the serum level of AGT was significantly decreased one month after radical operation. These results suggested that AGT has the potential to be a non-invasive biomarker in GC. Some studies, based on bioinformatics analysis of public database, reported that SERPINH1 was up-regulated in GC tissues [20–24]. Tian et al. revealed that SERPINH1 expression was significantly higher in the GC cell lines than in the normal gastric mucosal cell line. SERPINH1 regulates EMT and GC progression via the Wnt/β-catenin pathway [25]. SERPINH1 is considered to be the target gene of antitumor miR-148a-5p in GC cells. Knockdown of SERPINH1 resulted in inhibition of the aggressive phenotype of GC cells [26]. We focus on the concentration of SERPINH1 in serum and the results revealed that the level of serum SERPINH1 had high diagnostic efficacy in GC, with the AUC of 0.7355 (sensitivity 60.27% and specificity 79.52%). High-SERPINH1 patients showed more depth of tumor invasion and more advanced TNM stage than the low-SERPINH1 ones. The serum level of SERPINH1 was also significantly decreased one month after radical operation. These data indicated that SERPINH1 could be considered as a serum marker in diagnosis and prognosis assessment of GC. Previous studies confirmed that EGFR/MMP7 signaling pathway was activated in GC and played a role in GC metastasis [27–29]. MMP7 was directly or indirectly regulated by some microRNA and participated in GC metastasis [27, 30]. However, the content of MMP7 in serum, especially in exosomes, were not studied sufficiently. In the present study, we found that MMP7 showed high sensitivity and specificity in distinguishing GC patients from controls, and high expression of MMP7 was related

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 14 of 15

to the malignant characteristics of GC. When evaluating the prognosis of GC patients by combination of the three proteins in serum, it still showed good performance. These results indicated that the three candidate biomarkers showed excellent potential as the GC blood markers.

In recent years, extracellular exosomes have emerged as a novel biomarker pool for discovering the diagnostic and prognostic signatures of many diseases, especially in regard to cancer diagnosis [31, 32]. Proteins which carried by exosomes are ideal blood biomarker candidates, since they bear the traits of the original tissue and the signatures of the homing organs or cells [33, 34]. Unsurprisingly, several exosomal proteins have been demonstrated as potential diagnostic markers in GC. Yen et al. analyzed TGF-beta1 expressions in the exosomes isolated from the gastroepiploic veins of 61 GC patients by ELISA and revealed that increased exosomal TGF-β1 expression level was correlated with lymph node metastasis [35]. In the present study, the exosomes level of AGT and MMP7 were correlated positively with the serum level. Although there is a trend of positive correlation between exosomes level and serum level of SERPINH1, however, there was no statistical significance. This may be due to the small sample size. In addition, antibodies of the three proteins inhibited migration of GC cells, which mediated by exosomes derived from GC patients. Exosomal contents are relative stability and the proteins in blood derived exosomes are suitable as diagnostic and prognostic biomarkers. The results revealed that exosomal AGT, SERPINH1 and MMP7 may serve as biomarkers for gastric cancer diagnosis and prognosis and involved in GC progression.

Conclusions

In summary, using the novel biomarker-mining strategy, we identified three genes (*AGT*, *SERPINH1* and *MMP7*) expression profiles with good performance in diagnosis and prognosis of GC patients. We confirmed the diagnostic and prognostic value of the three-gene signature not only in public datasets but also in clinical serum samples. In addition, we detected the expression level of the three candidate biomarkers in exosomes derived from serum, and found that these proteins promote migration of GC cells through exosomes. These findings suggested that AGT, SERPINH1 and MMP7 have great potential as diagnostic and prognostic blood biomarkers and involved in GC progression.

Author details
[1]Medical Research Center, The Third People's Hospital of Chengdu, The Affiliated Hospital of Southwest Jiaotong University, 82 Qinglong Road, Chengdu 610031, Sichuan, China. [2]College of Medicine, Southwest Jiaotong University, Chengdu 610036, Sichuan, China. [3]Department of Clinical Laboratory, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu 610031, Sichuan, China. [4]Assisted Reproductive Center, The Maternal and Child Health Hospital of Qinzhou, Qinzhou 535000, Sichuan, China. [5]Department of Gastroenterology, The Third People's Hospital of Chengdu, The Affiliated Hospital of Southwest Jiaotong University, Chengdu 610031, Sichuan, China.

References
1. 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.
2. Zong L, Abe M, Seto Y, Ji J. The challenge of screening for early gastric cancer in China. Lancet. 2016;388(10060):2606.
3. Coburn N, Cosby R, Klein L, Knight G, Malthaner R, Mamazza J, Mercer CD, Ringash J. Staging and surgical approaches in gastric cancer: a systematic review. Cancer Treat Rev. 2018;63:104–15.
4. 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.
5. Matsuoka T, Yashiro M. Biomarkers of gastric cancer: Current topics and future perspective. World J Gastroenterol. 2018;24(26):2818–32.
6. Kulasingam V, Diamandis EP. Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat Clin Pract Oncol. 2008;5(10):588–99.
7. Wan B, Liu B, Yu G, Huang Y, Lv C. Differentially expressed autophagy-related genes are potential prognostic and diagnostic biomarkers in clear-cell renal cell carcinoma. Aging (Albany NY). 2019;11(20):9025–42.

Liu *et al. Cancer Cell Int*   (2021) 21:335

Page 15 of 15

8.   Yang J, Guo X, Zhang H, Zhang W, Song J, Xu H, Ma X. Differential diagnosis of pancreatic serous cystadenoma and mucinous cystadenoma: utility of textural features in combination with morphological characteristics. BMC Cancer. 2019;19(1):1223.

9.   Gold P, Freedman SO. Specific carcinoembryonic antigens of the human digestive system. J Exp Med. 1965;122(3):467–81.

10. Szymendera JJ. Clinical usefulness of three monoclonal antibody-defined tumor markers: CA 19–9, CA 50, and CA 125. Tumour Biol. 1986;7(5–6):333–42.

11. Tatsuta M, Itoh T, Okuda S, Yamamura H, Baba M, Tamura H. Carcinoembryonic antigen in gastric juice as an aid in diagnosis of early gastric cancer. Cancer. 1980;46(12):2686–92.

12. Hao C, Zhang G, Zhang L. Serum CEA levels in 49 different types of cancer and noncancer diseases. Prog Mol Biol Transl Sci. 2019;162:213–27.

13. Fukuda I, Yamakado M, Kiyose H. Influence of smoking on serum carcinoembryonic antigen levels in subjects who underwent multiphasic health testing and services. J Med Syst. 1998;22(2):89–93.

14. Haab BB, Huang Y, Balasenthil S, Partyka K, Tang H, Anderson M, Allen P, Sasson A, Zeh H, Kaul K, et al. Definitive characterization of CA 19–9 in resectable pancreatic cancer using a reference set of serum and plasma specimens. PLoS ONE. 2015;10(10):e0139049.

15. Wang YR, Yan JX, Wang LN. The diagnostic value of serum carcinoembryonic antigen, alpha fetoprotein and carbohydrate antigen 19–9 for colorectal cancer. J Cancer Res Ther. 2014;10(Suppl):307–9.

16. Feng F, Tian Y, Xu G, Liu Z, Liu S, Zheng G, Guo M, Lian X, Fan D, Zhang H. Diagnostic and prognostic value of CEA, CA19-9, AFP and CA125 for early gastric cancer. BMC Cancer. 2017;17(1):737.

17. Sugimoto M, Furuta T, Shirai N, Kodaira C, Nishino M, Ikuma M, Sugimura H, Hishida A. Role of angiotensinogen gene polymorphism on Helicobacter pylori infection-related gastric cancer risk in Japanese. Carcinogenesis. 2007;28(9):2036–40.

18. Shibata T, Tahara T, Arisawa T, Hirata I. Polymorphism of the salt sensitivity gene angiotensinogen and gastric cancer risk. Mol Med Rep. 2011;4(4):723–6.

19. Zhang C, Liang Y, Ma MH, Wu KZ, Dai DQ. KRT15, INHBA, MATN3, and AGT are aberrantly methylated and differentially expressed in gastric cancer and associated with prognosis. Pathol Res Pract. 2019;215(5):893–9.

20. Zhang X, Yang JJ, Kim YS, Kim KY, Ahn WS, Yang S. An 8-gene signature, including methylated and down-regulated glutathione peroxidase 3, of gastric cancer. Int J Oncol. 2010;36(2):405–14.

21. Li T, Gao X, Han L, Yu J, Li H. Identification of hub genes with prognostic values in gastric cancer by bioinformatics analysis. World J Surg Oncol. 2018;16(1):114.

22. Cao W, Zhou D, Tang W, An H, Zhang Y. Discovery of plasma messenger RNA as novel biomarker for gastric cancer identified through bioinformatics analysis and clinical validation. Peer J. 2019;7:e7025.

23. Zheng S, Yang L, Dai Y, Jiang L, Wei Y, Wen H, Xu Y. Screening and Survival Analysis of Hub Genes in Gastric Cancer Based on Bioinformatics. J Comput Biol. 2019;26(11):1316–25.

24. Wang F, Xue Q, Xu D, Jiang Y, Tang C, Liu X. Identifying the hub gene in gastric cancer by bioinformatics analysis and in vitro experiments. Cell Cycle. 2020;19(11):1326–37.

25. Tian S, Peng P, Li J, Deng H, Zhan N, Zeng Z, Dong W. SERPINH1 regulates EMT and gastric cancer metastasis via the Wnt/beta-catenin signaling pathway. Aging (Albany NY). 2020;12(4):3574–93.

26. Kawagoe K, Wada M, Idichi T, Okada R, Yamada Y, Moriya S, Okubo K, Matsushita D, Arigami T, Kurahara H, et al. Regulation of aberrantly expressed SERPINH1 by antitumor miR-148a-5p inhibits cancer cell aggressiveness in gastric cancer. J Hum Genet. 2020;65(8):647–56.

27. Liu G, Jiang C, Li D, Wang R, Wang W. MiRNA-34a inhibits EGFR-signaling-dependent MMP7 activation in gastric cancer. Tumour Biol. 2014;35(10):9801–6.

28. Ye Y, Zhou X, Li X, Tang Y, Sun Y, Fang J. Inhibition of epidermal growth factor receptor signaling prohibits metastasis of gastric cancer via down-regulation of MMP7 and MMP13. Tumour Biol. 2014;35(11):10891–6.

29. Xu L, Hou Y, Tu G, Chen Y, Du YE, Zhang H, Wen S, Tang X, Yin J, Lang L, et al. Nuclear Drosha enhances cell invasion via an EGFR-ERK1/2-MMP7 signaling pathway induced by dysregulated miRNA-622/197 and their targets LAMC2 and CD82 in gastric cancer. Cell Death Dis. 2017;8(3):e2642.

30. Sakamoto N, Naito Y, Oue N, Sentani K, Uraoka N, Oo HZ, Yanagihara K, Aoyagi K, Sasaki H, Yasui W. MicroRNA-148a is downregulated in gastric cancer, targets MMP7, and indicates tumor invasiveness and poor prognosis. Cancer Sci. 2014;105(2):236–43.

31. Thery C. Cancer: diagnosis by extracellular vesicles. Nature. 2015;523(7559):161–2.

32. Rahbarghazi R, Jabbari N, Sani NA, Asghari R, Salimi L, Kalashani SA, Feghhi M, Etemadi T, Akbariazar E, Mahmoudi M, Rezaie J. Tumor-derived extracellular vesicles: reliable tools for Cancer diagnosis and clinical applications. Cell Commun Signal. 2019;17(1):73.

33. Belov L, Matic KJ, Hallal S, Best OG, Mulligan SP, Christopherson RI. Extensive surface protein profiles of extracellular vesicles from cancer cells may provide diagnostic signatures from blood samples. J Extracell Vesicles. 2016;5:25355.

34. Buzas EI, Toth EA, Sodar BW, Szabo-Taylor KE. Molecular interactions at the surface of extracellular vesicles. Semin Immunopathol. 2018;40(5):453–64.

35. Yen EY, Miaw SC, Yu JS, Lai IR. Exosomal TGF-beta1 is correlated with lymphatic metastasis of gastric cancers. Am J Cancer Res. 2017;7(11):2199–208.

Publisher's Note