



# Discovery of plasma messenger RNA as novel biomarker for gastric cancer identified through bioinformatics analysis and clinical validation

Wei Cao<sup>1,2,3,\*</sup>, Dan Zhou<sup>1,2,3,\*</sup>, Weiwei Tang<sup>4</sup>, Hanxiang An<sup>1</sup> and Yun Zhang<sup>1,2,3</sup>

<sup>1</sup>Department of Medical Oncology, Xiang'an Hospital of Xiamen University, School of Medicine, Xiamen University, Xiamen, Fujian, China

<sup>2</sup>Key Laboratory of Design and Assembly of Functional Nanostructures, Fujian Provincial Key Laboratory of Nanomaterials, Fujian Institute of Research on the Structure of Matter, Chinese Academy of Sciences, Fuzhou, China

<sup>3</sup>Department of Translational Medicine, Xiamen Institute of Rare Earth Materials, Chinese Academy of Sciences, Xiamen, China

<sup>4</sup>Department of Medical Oncology, Cancer Hospital, The First Affiliated Hospital of Xiamen University, School of Medicine, Xiamen University, Teaching Hospital of Fujian Medical University, Xiamen, Fujian, China

\*These authors contributed equally to this work.

## ABSTRACT

**Background.** Gastric cancer (GC) is the third leading cause of cancer-related death worldwide, partially due to the lack of effective screening strategies. Thus, there is a stringent need for non-invasive biomarkers to improve patient diagnostic efficiency in GC.

**Methods.** This study initially filtered messenger RNAs (mRNAs) as prospective biomarkers through bioinformatics analysis. Clinical validation was conducted using circulating mRNA in plasma from patients with GC. Relationships between expression levels of target genes and clinicopathological characteristics were calculated. Then, associations of these selected biomarkers with overall survival (OS) were analyzed using the Kaplan-Meier plotter online tool.

**Results.** Based on a comprehensive analysis of transcriptional expression profiles across 5 microarrays, top 3 over- and underexpressed mRNAs in GC were generated. Compared with normal controls, expression levels of *collagen type VI alpha 3 chain* (*COL6A3*), *serpin family H member 1* (*SERPINH1*) and *pleckstrin homology and RhoGEF domain containing G1* (*PLEKHG1*) were significantly upregulated in GC plasmas. Receiver-operating characteristic (ROC) curves on the diagnostic efficacy of plasma *COL6A3*, *SERPINH1* and *PLEKHG1* mRNAs in GC showed that the area under the ROC (AUC) was 0.720, 0.698 and 0.833, respectively. Combined, these three biomarkers showed an elevated AUC of 0.907. Interestingly, the higher *COL6A3* level was significantly correlated with lymph node metastasis and poor prognosis in GC patients. High level of *SERPINH1* mRNA expression was correlated with advanced age, poor differentiation, lower OS, and *PLEKHG1* was also associated with poor OS in GC patients.

**Conclusion.** Our results suggested that circulating *COL6A3*, *SERPINH1* and *PLEKHG1* mRNAs could be putative noninvasive biomarkers for GC diagnosis and prognosis.

Submitted 5 December 2018

Accepted 24 April 2019

Published 18 June 2019

Corresponding authors

Hanxiang An, an\_hanxiang@126.com

Yun Zhang, zhangyunfjirm@163.com

Academic editor

Shi-Cong Tao

Additional Information and  
Declarations can be found on  
page 12

DOI 10.7717/peerj.7025

© Copyright  
2019 Cao et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

**Subjects** Gastroenterology and Hepatology, Oncology, Translational Medicine

**Keywords** Biomarker, Bioinformatics, Gastric cancer, Circulating mRNA

## INTRODUCTION

Gastric cancer (GC) is one of the most common diagnosed cancers worldwide and the third leading cause of cancer-related mortality after lung and liver cancer (Ferlay *et al.*, 2015). The carcinogenesis and progression of GC is complex, involving the alternations of multi-step and multi-genes (Zhao *et al.*, 2017). The 5-year survival rate of GC diagnosed at a later stage is less than 20%, but rises to 90% for patients diagnosed at an early stage (Stahl *et al.*, 2017). Although traditional biomarkers including carbohydrate antigen 19-9 (CA 19-9) and carcino-embryonic antigen (CEA) have improved clinical outcomes for GC, their sensitivity and specificity are still limited (Ding *et al.*, 2017). Therefore, the identification of novel screening biomarkers may help better diagnose and improve the prognosis of GC (Stahl *et al.*, 2017).

Recently, there is a growing interest in circulating messenger RNAs (mRNAs) isolated from body fluids as potential minimally/non-invasive biomarkers for cancer detection (Kishikawa *et al.*, 2015; Sole *et al.*, 2019). Numerous studies have reported that circulating mRNAs can be detected in GC (Funaki *et al.*, 1996), melanoma (Kopreski *et al.*, 1999) and nasopharyngeal carcinoma (Lo *et al.*, 1999). These discoveries provide evidence for circulating mRNAs to be served as appealing non-invasive biomarker candidates in various cancers. However, few circulating mRNAs have been investigated in GC, and limited studies have compared the expression level of circulating mRNAs in plasma of GC patients with the clinicopathological characteristics.

The establishment of high-throughput molecular database such as microarray databases brings a new approach for biomarker identification. We have previously published an effective bioinformatics scheme to identify noninvasive biomarkers for lung cancer (Zhou *et al.*, 2017). In the present study, we applied a similar approach to explore mRNAs circulated in plasma from patients with GC as potential biomarkers. Subsequently, the associations between these biomarkers and clinicopathological characteristics were analyzed. Finally, the relationships between these selected noninvasive biomarkers and overall survival (OS) were investigated.

## MATERIALS & METHODS

### Genome-wide expression analysis by OncoPrint

Expression profiling, including 304 GC cancer samples and 174 normal controls, was obtained from OncoPrint microarray database (<http://www.oncoprint.org>) (Rhodes *et al.*, 2004). In order to analyze the expression pattern of cancer vs. normal tissue mRNA, we focused on primary tumors and the following cut-offs were employed  $p$ -value  $\leq 10^{-4}$ , fold change  $\geq 2$  and gene rank  $\leq 10\%$ . Heat maps of overexpressed and underexpressed mRNAs in GC were available for each study.

**Table 1** Clinicopathologic characteristics of gastric cancer patients and healthy subjects.

| Clinicopathologic factors         | Gastric cancer cases | Healthy subjects  |
|-----------------------------------|----------------------|-------------------|
| Total                             | 56                   | 14                |
| Mean Age $\pm$ SD                 | 62.93 $\pm$ 6.17     | 43.64 $\pm$ 14.91 |
| Sex                               |                      |                   |
| Male                              | 43                   | 9                 |
| Female                            | 13                   | 5                 |
| Stage <sup>a</sup>                |                      |                   |
| I + II + III                      | 37                   |                   |
| IV                                | 18                   |                   |
| N/A <sup>b</sup>                  | 1                    |                   |
| Lymph node metastasis             |                      |                   |
| <15                               | 31                   |                   |
| $\geq$ 15                         | 7                    |                   |
| N/A                               | 18                   |                   |
| Differentiation                   |                      |                   |
| Poor                              | 19                   |                   |
| Moderately or well-differentiated | 11                   |                   |
| N/A                               | 26                   |                   |

**Notes.**

<sup>a</sup>Tumor stages were determined according to Union Internationale Contre le Cancer (UICC) criteria.

<sup>b</sup>N/A, not available.

**Clinical specimens**

Peripheral blood samples from 56 patients with gastric adenocarcinoma were addressed before therapeutic intervention by venipuncture and processed within 2 h at the First Affiliated Hospital of Xiamen University. We also collected blood samples from 14 healthy volunteers. The healthy volunteers presented neither a history of cancer nor other diseases. All patients were pathologically diagnosed as having gastric cancer using surgical specimens and biopsies. Plasma was isolated from 4 ml blood specimens after centrifugation at  $1,600 \times g$  for 10 min at 4 °C and  $10,000 \times g$  for 10 min, and then stored at  $-80$  °C until the next step. Demographic, clinical and histopathological parameters of all these cases were summarized in Table 1. All experimental protocols were approved by the Clinical Research Ethics Committee of the First Affiliated Hospital of Xiamen University. All methods were performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all human participants after complete description of the study (Fig. S1 and Table S2).

**RNA extraction**

RNA was extracted from 500  $\mu$ l plasma using TRIzol LS reagent (cat#10296018, Thermo Fisher Scientific Inc.) following manufacturer's instructions as previously described (Pucciarelli *et al.*, 2012; Zhang *et al.*, 2012). In brief, 500  $\mu$ l plasma was mixed with 500  $\mu$ l TRIzol reagent. After 5 min incubation at 4 °C, 500  $\mu$ l chloroform was added to the mixture, and violently shaken for 30 s. The mixture was immediately centrifuged at  $12,000 \times g$  for 5 min at 4 °C. The above aqueous layer was transferred into a fresh tube containing 800  $\mu$ l

isopropyl alcohol. Next, the mixture was centrifuged at  $12,000\times g$  for 5 min at  $4\text{ }^{\circ}\text{C}$  and washed with 1 ml 70% ethanol for twice. Lastly, the RNA pellet was dissolved in 15  $\mu\text{l}$  RNase-free water. Qubit RNA HS Assay Kit and Qubit 3.0 fluorometer (ThermoFisher Scientific Inc. cat#Q32851) were employed to quantify the concentration of RNA solution through spectrophotometry. The concentration of RNA isolated from plasma ranged was 35.24–168.18 ng/ $\mu\text{l}$ .

### Quantitative PCR (qPCR)

The extracted total RNA was reverse-transcription into cDNA using PrimeScript RT reagent Kit (TAKARA cat#RR047A) according to the manufacturer's protocol in triplicate. The resulting cDNA was stored at  $-80\text{ }^{\circ}\text{C}$  for next PCR amplification. Primer sequences were designed through web-based version 4.1.0 of Primer 3 and were shown in [Table S1](#). For clinical validation of the bioinformatics analysis results, qPCR was conducted by ABI ViiA 7 Real-Time PCR System (Applied Biosystems) with melting curve analysis. QPCR was carried out in triplicate at  $50\text{ }^{\circ}\text{C}$  for 2 min, denaturing at  $95\text{ }^{\circ}\text{C}$  for 5 min, followed by 40 cycles at  $95\text{ }^{\circ}\text{C}$  for 30 s,  $58\text{ }^{\circ}\text{C}$  for 1 min. *Glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) was selected as an internal control. Water negative controls contained all components for the qPCR reaction without target RNA. Positive controls of RNA were extracted from SGC7901 cells obtained from the Cancer Center of Xiamen University (Xiamen, China). The  $2^{-\Delta\text{CT}}$  algorithm ( $\Delta\text{CT} = \text{Ct. target} - \text{Ct. reference}$ ) was employed for data analysis ([Maru et al., 2009](#)).

### The Kaplan–Meier plotter analysis

The prognostic value of candidate circulating mRNAs was analyzed using the Kaplan–Meier plotter database, an online database containing 54,675 genes on survival based on 1065GC samples with a mean follow-up of 33 months ([Szasz et al., 2016](#)). Overall survival of patients with high and low expression levels of target genes was displayed using Kaplan–Meier survival curves. Hazard ratios (HR) with 95% confidence intervals (CI), and log-rank *p*-values were also calculated and summarized.

### Functional enrichment network

Gene functional network was performed using gene ontology enrichment (GO) analysis. Enrichment map were created using the Cytoscape (v3.6.0). FDR < 0.05 was considered to be significant.

### Statistical analysis

The Mann–Whitney *U* test was used to compare the expression status of circulating mRNAs in normal and GC groups and calculate the relationship between clinicopathologic characteristics and expression levels of relevant mRNA. Data was shown as median and range. Receiver operating characteristic (ROC) curves and the area under the curve (AUC) was used to identify the diagnosis value of selected mRNA ([Brumback, Pepe & Alonzo, 2006](#)). Cut-off values were assessed at different sensitivities and specificities and at the maximum Youden's index = (sensitivity + specificity – 1) ([Youden, 1950](#)). Then, the logistic regression model was performed to obtain a combined ROC curve. GraphPad

Prism 6.0 (GraphPad Software Inc., La Jolla, CA) and SPSS (version 22.0, IBM SPSS co., USA) was used for these statistical analyses. A two-sided  $p$  value of less than 0.05 was defined as statistically significant.

## RESULTS

### Identification of candidate mRNAs from Oncomine database

We compared the mRNA expression level in gastric cancer vs. normal samples obtained from Oncomine database. In total, 304 GC samples including 50 diffuse gastric adenocarcinoma, 21 gastric adenocarcinoma, 113 gastric intestinal type adenocarcinoma, 22 gastric mixed adenocarcinoma, 6 gastrointestinal stromal tumor and 92 other GC and 174 normal controls from 5 selected microarray datasets were analyzed (Chen *et al.*, 2003; Cho *et al.*, 2011; Cui *et al.*, 2011; D'Errico *et al.*, 2009; Wang *et al.*, 2012). As shown in Fig. 1, expression profiling analysis generated 40 altered mRNA expressions in GC. The top 10 overexpressed and underexpressed genes embodied important genes involved in carcinogenesis and progression as well as several uncharacterized candidates (*collagen type VI alpha 3 chain (COL6A3)*, *serpin family H member 1 (SERPINH1)*, *pleckstrin homology and RhoGEF domain containing G1 (PLEKHG1)*, *collagen type I alpha 2 chain (COL1A2)*, *claudin 1 (CLDN1)*, *metallopeptidase inhibitor 1 (TIMP1)*, *NOP56 ribonucleoprotein (NOP56)*, *collagen type IV alpha 1 chain (COL4A1)*, *immunoglobulin like domain containing receptor 1 (ILDR1)*; *cadherin 11 (CDH11)*, *pepsinogen 4, group I pepsinogen A (PGA4)*, *potassium voltage-gated channel subfamily E regulatory subunit 2 (KCNE2)*, *gastric intrinsic factor (GIF)*, *solute carrier family 9 member A4 (SLC9A4)*, *estrogen related receptor gamma (ESRRG)*, *ATPase H+/K+ transporting subunit alpha (ATP4A)*, *ATPase H+/K+ transporting subunit beta (ATP4B)*, *PRDM16 divergent transcript (FLJ42875)*, *MAS related GPR family member D (MRGPRD)* and *tripartite motif containing 50 (TRIM50)*, respectively).

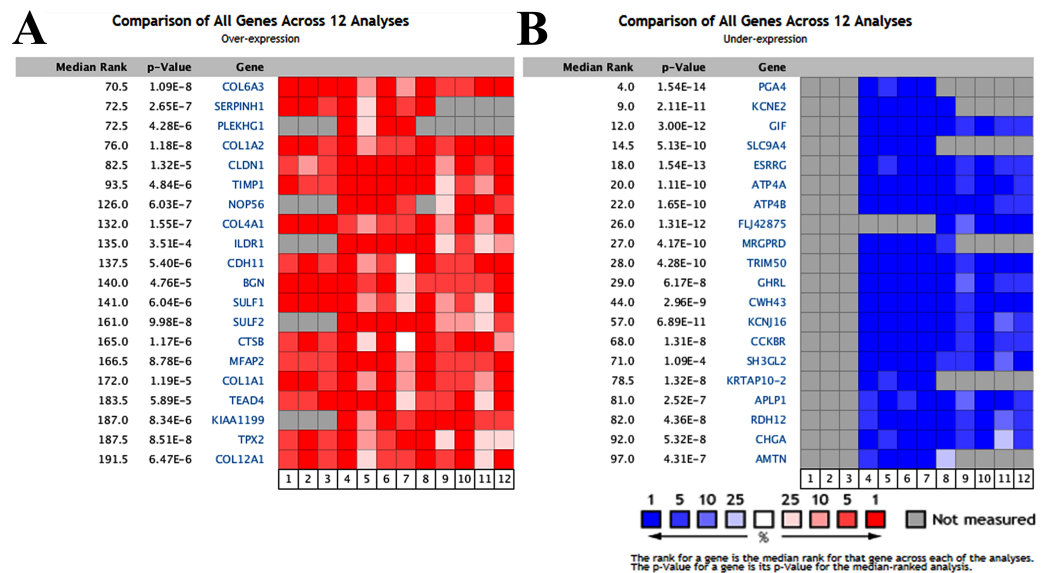
### Experimental validation of the potential noninvasive biomarkers

To validate our expression profiling analyses, six selected genes including top three overexpressed (*COL6A3*, *SERPINH1* and *PLEKHG1*) and underexpressed (*PGA4*, *KCNE2* and *GIF*) mRNAs were experimentally verified using circulating mRNA extracted from 56 GC patients and 14 healthy subjects.

The expression of *COL6A3*, *PLEKHG1*, *PGA4* and *KCNE2* was detected in 55/56, 41/56, 32/56 and 55/56 GC plasma samples respectively, whereas the expression of *SERPINH1* and *GIF* was detected in all GC samples. As shown in Fig. 2, the expression level of *COL6A3* ( $P = 0.0128$ ), *SERPINH1* ( $P = 0.0217$ ) and *PLEKHG1* ( $P = 0.0064$ ) was significantly higher in GC plasmas than those in normal controls. However, the expression levels of *PGA4*, *KCNE2* and *GIF* had no significant change.

### Diagnostic value of plasma *COL6A3*, *SERPINH1* and *PLEKHG1* for GC

We performed ROC curve analysis to assess the diagnostic value of these three circulating mRNAs (Fig. 3). The plasma level of *COL6A3* had a sensitivity of 100% and a specificity of 46.2% to differentiate the GC patients from the healthy subjects, with an AUC of 0.720 at



**Figure 1** Transcriptional heat map of the top 20 over- and underexpressed genes in gastric cancer samples compared with normal samples through OncoPrint analysis. The level plots depict the frequencies (%) of (A) over- and (B) underexpressed candidate messenger RNAs (mRNAs) in 12 analyses from five included studies (*Chen et al., 2003*; *Cho et al., 2011*; *Cui et al., 2011*; *D'Errico et al., 2009*; *Wang et al., 2012*). Red cells represent overexpression. Blue cells represent underexpression. Gray cells represent not measured.

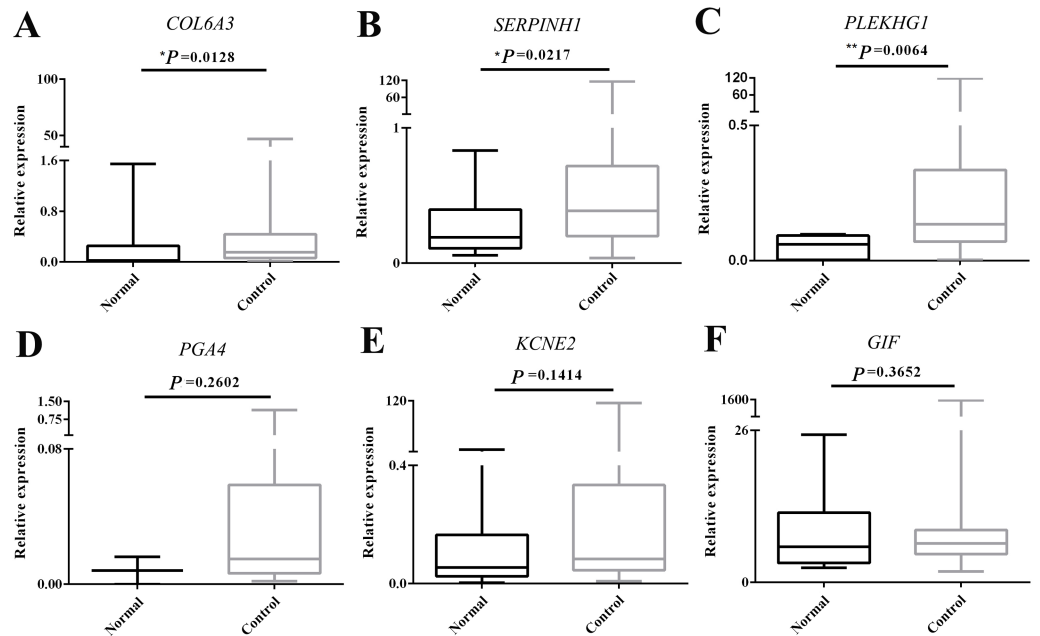
Full-size DOI: 10.7717/peerj.7025/fig-1

the optimal cut-off point (95% CI [0.522–0.919]). *SERPINH1* had a sensitivity of 58.9% and a specificity of 78.6%, with an AUC of 0.698 (95% CI [0.543–0.852]). *PLEKHG1* had a sensitivity of 68.3% and a specificity of 100%, with an AUC of 0.833 (95% CI [0.699–0.968]). Combining any two of the biomarkers had values for ROC area ranging from 0.676 to 0.870, sensitivities from 60% to 82.9%, specificities from 76.9% to 100% (Fig. 4). Furthermore, we used logistic regression analysis to combine these three circulating mRNAs and obtained an increased AUC value of 0.907 (95% CI [0.820–0.993]), with a sensitivity of 82.9% and a specificity of 100%.

### Associations between clinicopathological characteristics and plasma *COL6A3*, *SERPINH1* and *PLEKHG1*

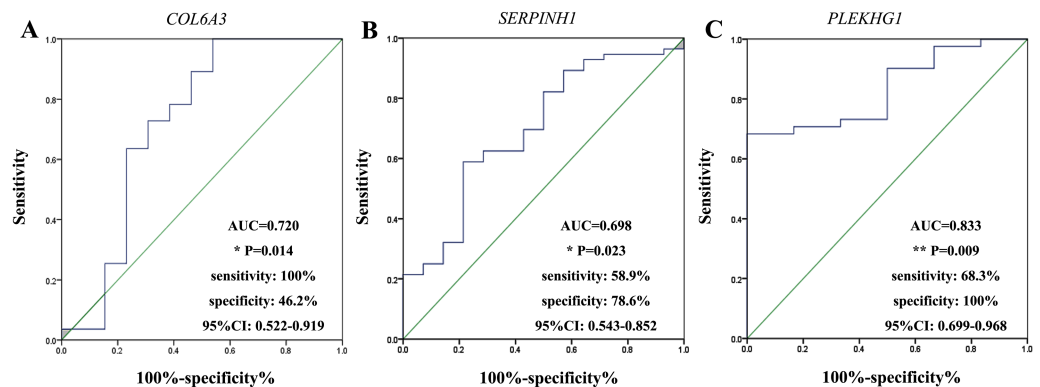
The associations between clinicopathological characteristics and these three circulating mRNAs were investigated. As shown in Table 2, the higher *COL6A3* level was significantly associated with increased lymph node metastasis ( $P = 0.0233$ ), whereas the elevated expression of *SERPINH1* was associated with advanced age ( $P = 0.0034$ ) and poor differentiation ( $P = 0.0231$ ) in GC patients. No significant association between *PLEKHG1* and any clinicopathological characteristics including age, sex, stage, lymph node metastasis or differentiation was found.





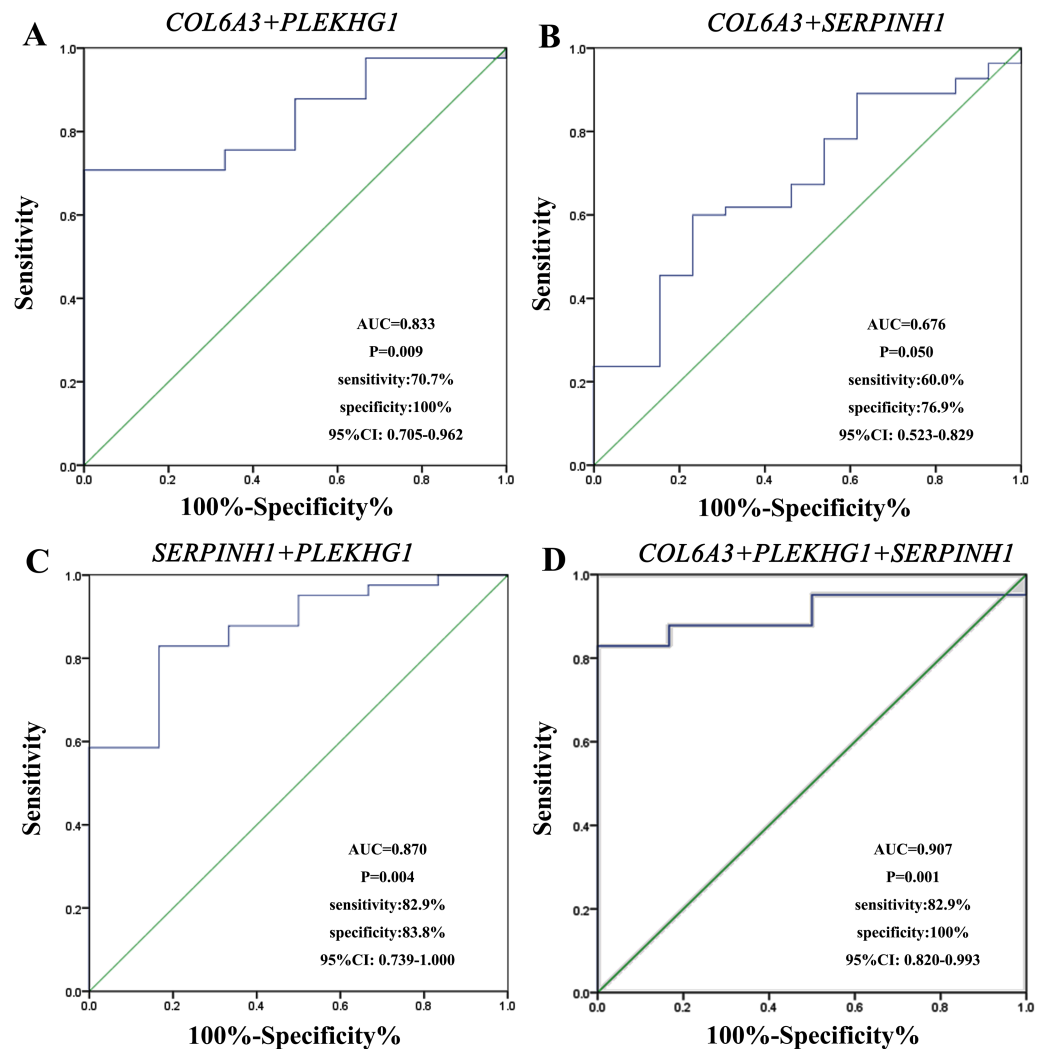
**Figure 2** Experimental validation of the top 6 over- and underexpressed messenger RNAs (mRNAs). The change in circulating mRNA levels of collagen type VI alpha 3 chain (*COL6A3*) (A), serpin family H member 1 (*SERPINH1*) (B), pleckstrin homology and RhoGEF domain containing G1 (*PLEKHG1*) (C), pepsinogen 4, group I (pepsinogen A) (*PGA4*) (D), potassium voltage-gated channel subfamily E regulatory subunit 2 (*KCNE2*) (E) and gastric intrinsic factor (*GIF*) (F) between gastric cancer patients and normal subjects detected by qPCR using Student's *t* test. The Mann–Whitney *U* test was used to compare the expression status of circulating mRNAs in normal and GC groups. Data was shown as box plots and the intersecting line represents the median value with the interquartile range. Results were shown with means  $\pm$  SEM.  $*p < 0.05$ ,  $**p < 0.01$ .

Full-size [DOI: 10.7717/peerj.7025/fig-2](https://doi.org/10.7717/peerj.7025/fig-2)



**Figure 3** Receiver-operating characteristic curves (ROC) analysis of selected markers in gastric cancer. The results showed the performances of fold-change in collagen type VI alpha 3 chain (*COL6A3*) (A), serpin family H member 1 (*SERPINH1*) (B) and pleckstrin homology and RhoGEF domain containing G1 (*PLEKHG1*) (C) messenger RNAs (mRNAs) expression in predicting the gastric cancer.  $*p < 0.05$ ,  $**p < 0.01$ .

Full-size [DOI: 10.7717/peerj.7025/fig-3](https://doi.org/10.7717/peerj.7025/fig-3)



**Figure 4** Receiver-operating characteristic curves (ROC) for combining markers in gastric cancer. The results showed combination of collagen type VI alpha 3 chain (*COL6A3*) and pleckstrin homology and RhoGEF domain containing G1 (*PLEKHG1*) (A), *COL6A3* and serpin family H member 1 (*SERPINH1*) (B), *SERPINH1* and *PLEKHG1* (C), and combination of these three genes (D) to differentiate patients with gastric from normal subjects.

Full-size DOI: 10.7717/peerj.7025/fig-4

### Increased *COL6A3*, *SERPINH1* and *PLEKHG1* is associated with poor prognosis

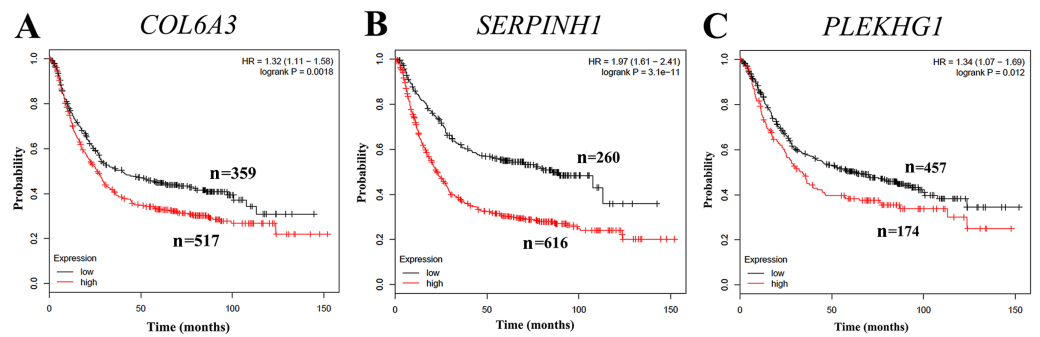
To further evaluate whether the expression levels of *COL6A3*, *SERPINH1* and *PLEKHG1* can predict prognosis, we performed a survival analysis based on publicly gene expression datasets from the Kaplan–Meier Plotter resource. As shown in Fig. 5, the higher expression of *COL6A3* (HR = 1.32, 95% CI [1.11–1.58],  $p = 0.0018$ ), *SERPINH1* (HR = 1.97, 95% CI [1.61–2.41],  $p = 3.1e^{-11}$ ) and *PLEKHG1* (HR = 1.34, 95% CI [1.07–1.69],  $p = 0.012$ ) were all significantly correlated with poor OS in GC. These results indicated that GC patients with high *COL6A3*, *SERPINH1* or *PLEKHG1* tend to have unfavorable outcome.



**Table 2** Association between the expression of circulating *COL6A3*, *SERPINH1* and *PLEKHG1* in gastric cancer and clinicopathologic characteristics.

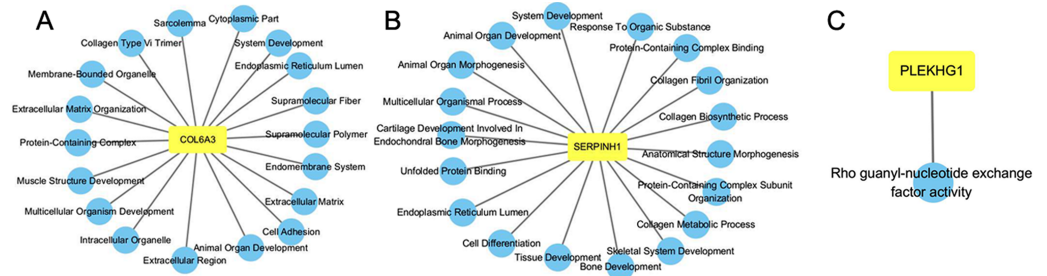
| Variable                          | <i>COL6A3</i> |                      |          | <i>SERPINH1</i> |                      |          | <i>PLEKHG1</i> |                      |          |
|-----------------------------------|---------------|----------------------|----------|-----------------|----------------------|----------|----------------|----------------------|----------|
|                                   | <i>n</i>      | Expression status    | <i>P</i> | <i>n</i>        | Expression status    | <i>P</i> | <i>n</i>       | Expression status    | <i>P</i> |
| Age                               |               |                      |          |                 |                      |          |                |                      |          |
| <50                               | 12            | 0.179 (0.028–1.250)  | 0.7091   | 12              | 0.190 (0.073–1.430)  | 0.0034** | 10             | 0.130 (0.003–1.052)  | 0.4517   |
| ≥50                               | 43            | 0.147 (0.011–47.014) |          | 44              | 0.47 (0.374–116.407) |          | 31             | 0.148 (0.036–119.46) |          |
| Sex                               |               |                      |          |                 |                      |          |                |                      |          |
| Male                              | 42            | 0.150 (0.011–10.386) | 0.5496   | 43              | 0.399 (0.374–8.283)  | 0.1856   | 30             | 0.126 (0.036–2.447)  | 0.9633   |
| Female                            | 13            | 0.162 (0.018–47.014) |          | 13              | 0.325 (0.038–116.41) |          | 11             | 0.161 (0.003–119.46) |          |
| Stage                             |               |                      |          |                 |                      |          |                |                      |          |
| I + II + III                      | 36            | 0.150 (0.011–47.014) | 0.1616   | 37              | 0.374 (0.038–116.41) | 0.4697   | 27             | 0.127 (0.003–119.46) | 0.5101   |
| IV                                | 18            | 0.190 (0.018–10.386) |          | 18              | 0.454 (0.118–3.089)  |          | 13             | 0.198 (0.065–1.214)  |          |
| Lymph node metastasis             |               |                      |          |                 |                      |          |                |                      |          |
| <15                               | 30            | 0.106 (0.229–47.014) | 0.0233*  | 31              | 0.374 (0.221–116.41) | 0.6848   | 20             | 0.118 (0.036–119.46) | 0.6207   |
| ≥15                               | 7             | 0.225 (0.155–0.698)  |          | 7               | 0.472 (0.039–3.623)  |          | 7              | 0.116 (0.037–0.336)  |          |
| Differentiation                   |               |                      |          |                 |                      |          |                |                      |          |
| Poor                              | 19            | 0.198 (0.011–47.01)  | 0.881    | 19              | 0.362 (0.127–116.41) | 0.0231*  | 12             | 0.110 (0.003–119.46) | 0.1059   |
| Moderately or well-differentiated | 11            | 0.033 (0.018–0.693)  |          | 11              | 0.678 (0.159–3.089)  |          | 10             | 0.106 (0.048–2.447)  |          |

**Notes.**\**P* value < 0.05.\*\**P* value < 0.01.



**Figure 5** Correlation of collagen type VI alpha 3 chain (*COL6A3*), serpin family H member 1 (*SERPINH1*) and pleckstrin homology and RhoGEF domain containing G1 (*PLEKHG1*) with survival outcomes in gastric cancer patients. Increased expression of (A) *COL6A3*, (B) *SERPINH1* and (C) *PLEKHG1* predicted worse overall survival (OS) in gastric cancer.

Full-size [DOI: 10.7717/peerj.7025/fig-5](https://doi.org/10.7717/peerj.7025/fig-5)



**Figure 6** GO functional enrichment analysis of *COL6A3* (A), *SERPINH1* (B) and *PLEKHG1* (C).

Full-size [DOI: 10.7717/peerj.7025/fig-6](https://doi.org/10.7717/peerj.7025/fig-6)

## GO functional enrichment analysis

Functional enrichment network of *COL6A3*, *SERPINH1* and *PLEKHG1* was constructed. As shown in Fig. 6A, *COL6A3* was predicted to have the main functions: extracellular matrix organization, cell adhesion, multicellular organism development, animal organ development and system development. *SERPINH1* played roles in extracellular matrix organization, collagen fibril polymerization, skeletal system development, collagen metabolic process, and animal organ morphogenesis (Fig. 6B). *PLEKHG1* was found associated with Rho guanyl-nucleotide exchange factor activity (Fig. 6C).

## DISCUSSION

The survival of GC affected patients depends mainly on early detection (Wang et al., 2006). The common screening approaches are gastroscopy and computed tomography, which are invasive and expensive (Ke et al., 2017). Therefore, easily accessible and noninvasive biomarkers derived from body fluid are prevalent (Shen et al., 2017). Exploration of circulating biomarkers for various cancer types can be conducted through different approaches. In the present study, we identified candidate biomarkers according to our previous strategy which combined comprehensive analysis of microarray data and

experimental validation using plasma samples (Zhou et al., 2017). We first identified three circulating mRNA markers (*COL6A3*, *SERPINH1* and *PLEKHG1*) that carry diagnostic potential for GC. We also found that the combination of these three markers exhibited better diagnostic performance for GC than each individual marker. Our strategy can also be flexibly applied to various diseases.

Circulating RNA including lncRNA, mRNA and microRNA can be isolated and detected in serum, plasma, urine and lymph. In bodily fluid, RNA molecules are directly exposed to RNase, resulting in the degradation of RNAs and the difficulty to identify RNA based biomarkers (Hasselmann et al., 2001; Sisco, 2001). However, some studies have suggested that circulating RNA is especially stable due to the protection from phospholipids (Elhefnawy et al., 2004; Halicka, Bedner & Darzynkiewicz, 2000; Ma, Tao & Kang, 2012). In addition, another study demonstrates that the concentration of circulating RNA in GC patients is higher than that of healthy controls, which is associated with tumor growth and metastasis metabolism (Elhefnawy et al., 2004; Rykova et al., 2006). In this study, we found that the levels of plasmatic *COL6A3*, *SERPINH1* and *PLEKHG1* were significantly increased in patients with GC than in healthy subjects.

*COL6A3*, located at chromosome 2q37, encodes the alpha-3 chain for type VI collagen (Dankel et al., 2014). Collagen VI has been initially defined as an extracellular matrix protein and it is expressed in various tissues such as muscle, skin and cartilage. *COL6A3* is a secreted protein and have received growing attention due to its abnormal expression in colon, pancreatic, bladder and prostate cancer (Kang et al., 2014; Thorsen et al., 2008). In a previous study, *COL6A3* has been shown to be a potential plasma marker of colorectal cancer and is associated with tumor metastasis (Qiao et al., 2015). However, the expression pattern and function of *COL6A3* in the tumorigenesis of GC remain unclear. Our present study indicated that *COL6A3* was overexpressed in plasma of GC patients and was associated with increased lymph node metastasis.

*SERPINH1*, also known as heat shock protein 47 (HSP47), belongs to the serpin superfamily involving serine proteinase inhibitors (Ito & Nagata, 2016). The location of *SERPINH1* is at chromosome 11q13.5, a domain frequently abnormal in various human cancers. Numerous studies have demonstrated that *SERPINH1* is overexpressed in various human cancers, including lung cancer, pancreatic cancer, cervical cancer and glioma (Wu et al., 2016; Yamamoto et al., 2013). In addition, serum *SERPINH1* has been reported to be used as a possible target for patients with scirrhous gastric cancer. Our results showed that the overexpressed *SERPINH1* was associated with advanced age and poor differentiation in plasma from GC patients. These data indicated that *SERPINH1* played a critical role in GC, although further studies are necessary to clarify the biological mechanism of *SERPINH1* in GC.

*PLEKHG1* contains a Rho guanine nucleotide exchange factor domain and a pleckstrin homology domain. *PLEKHG1* acts as a signaling platform in various cells, but the detail functions are not clear. In this study, we found that the plasma level of *PLEKHG1* mRNA was significantly increased in GC patients compared with normal subjects, with a markedly high AUC value of 0.8333. These results suggested that *PLEKHG1* mRNA have a high diagnosis capability.

We carried out the ROC curve to analyze the diagnostic value of *COL6A3*, *SERPINH1* and *PLEKHG1* in plasma from GC patients. The results demonstrated that *PLEKHG1* had higher diagnostic value for GC than that of *COL6A3* or *SERPINH1*. More powerful diagnostic values were observed when combining these three mRNAs, resulting in an AUC of 0.907. In addition, the prognostic roles of these three potential biomarkers in GC patients were rarely reported and all these biomarkers were correlated with worse OS for patients with GC.

## CONCLUSIONS

Accordingly, we have identified potential noninvasive biomarkers for gastric cancer using bioinformatics analysis through a public database and verified their value using GC clinical tumor and plasma specimens. *COL6A3*, *SERPINH1* and *PLEKHG1* are three prospective biomarkers for GC. The combination of plasma *COL6A3*, *SERPINH1* and *PLEKHG1* represent a promising diagnostic method. The clinical samples employed in this study were relatively limited. Hence, large-scale studies should be performed to investigate the clinical significance of *COL6A3*, *SERPINH1* and *PLEKHG1* in GC in the future. Moreover, further investigation of their biological function and their potential as therapeutic targets in GC is warranted.

## ACKNOWLEDGEMENTS

We would like to thank the patients and healthy subjects for consenting to provide material for this study.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This study was supported by the National Natural Science Foundation of China (No. 81802323 and No. 81702414), Natural Science Foundation of Fujian Province of China (No. 2017Y0084 and No. 2015J01557), Science and Technology Service Network Initiative Foundation of CAS (No. 2016T3009), the Fujian Provincial Health and Family Planning Commission Foundation of Youth Scientific Research Project (No. 2015-2-43), and Xiamen Science and Technology Bureau Foundation of Science and Technology Project for the Benefit of the People (No. 3502Z20164010). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

National Natural Science Foundation of China: 81802323, 81702414.

Natural Science Foundation of Fujian Province of China: 2017Y0084, 2015J01557.

Science and Technology Service Network Initiative Foundation of CAS: 2016T3009.

Fujian Provincial Health and Family Planning Commission Foundation of Youth Scientific Research Project: 2015-2-43.

Xiamen Science and Technology Bureau Foundation of Science and Technology Project for the Benefit of the People: 3502ZZ20164010.

### Competing Interests

The authors declare there are no competing interests.

### Author Contributions

- Wei Cao performed the experiments, analyzed the data, prepared figures and/or tables.
- Dan Zhou conceived and designed the experiments.
- Weiwei Tang contributed reagents/materials/analysis tools.
- Hanxiang An and Yun Zhang authored or reviewed drafts of the paper, approved the final draft.

### Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

All experimental protocols were approved by the Clinical Research Ethics Committee of the First Affiliated Hospital of Xiamen University. All methods were performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all human participants after complete description of the study.

### Data Availability

The following information was supplied regarding data availability:

The raw data is available as [Dataset S1](#).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.7025#supplemental-information>.

## REFERENCES

- Brumback LC, Pepe MS, Alonzo TA. 2006.** Using the ROC curve for gauging treatment effect in clinical trials. *Statistics in Medicine* **25**(4):575–590 DOI [10.1002/sim.2345](https://doi.org/10.1002/sim.2345).
- Chen X, Leung SY, Yuen ST, Chu KM, Ji J, Li R, Chan AS, Law S, Troyanskaya OG, Wong J, So S, Botstein D, Brown PO. 2003.** Variation in gene expression patterns in human gastric cancers. *Molecular Biology of the Cell* **14**(8):3208–3215 DOI [10.1091/mbc.E02-12-0833](https://doi.org/10.1091/mbc.E02-12-0833).
- Cho JY, Lim JY, Cheong JH, Park YY, Yoon SL, Kim SM, Kim SB, Kim H, Hong SW, Park YN. 2011.** Gene expression signature-based prognostic risk score in gastric cancer. *Clinical Cancer Research* **17**(7):1850–1857 DOI [10.1158/1078-0432.CCR-10-2180](https://doi.org/10.1158/1078-0432.CCR-10-2180).
- Cui J, Chen Y, Chou WC, Sun L, Chen L, Suo J, Ni Z, Zhang M, Kong X, Hoffman LL, Kang J, Su Y, Olman V, Johnson D, Tench DW, Amster IJ, Orlando R, Puett D, Li F, Xu Y. 2011.** An integrated transcriptomic and computational analysis for biomarker identification in gastric cancer. *Nucleic Acids Research* **39**(4):1197–1207 DOI [10.1093/nar/gkq960](https://doi.org/10.1093/nar/gkq960).

- Dankel SN, Svard J, Mattha S, Claussnitzer M, Kloting N, Glunk V, Fandalyuk Z, Grytten E, Solsvik MH, Nielsen HJ, Busch C, Hauner H, Bluher M, Skurk T, Sagen JV, Mellgren G. 2014. *COL6A3* expression in adipocytes associates with insulin resistance and depends on PPARgamma and adipocyte size. *Obesity* 22(8):1807–1813 DOI 10.1002/oby.20758.
- D’Errico M, De Rinaldis E, Blasi MF, Viti V, Falchetti M, Calcagnile A, Sera F, Saieva C, Ottini L, Palli D, Palombo F, Giuliani A, Dogliotti E. 2009. Genome-wide expression profile of sporadic gastric cancers with microsatellite instability. *European Journal of Cancer* 45(3):461–469 DOI 10.1016/j.ejca.2008.10.032.
- Ding D, Song Y, Yao Y, Zhang S. 2017. Preoperative serum macrophage activated biomarkers soluble mannose receptor (sMR) and soluble haemoglobin scavenger receptor (sCD163), as novel markers for the diagnosis and prognosis of gastric cancer. *Oncology Letters* 14(3):2982–2990 DOI 10.3892/ol.2017.6547.
- Elhefnawy T, Raja S, Kelly L, Bigbee WL, Kirkwood JM, Luketich JD, Godfrey TE. 2004. Characterization of amplifiable, circulating RNA in plasma and its potential as a tool for cancer diagnostics. *Clinical Chemistry* 50(3):564–573 DOI 10.1373/clinchem.2003.028506.
- Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* 136(5):E359–E386 DOI 10.1002/ijc.29210.
- Funaki NO, Tanaka J, Kasamatsu T, Ohshio G, Hosotani R, Okino T, Imamura M. 1996. Identification of carcinoembryonic antigen mRNA in circulating peripheral blood of pancreatic carcinoma and gastric carcinoma patients. *Life Sciences* 59(25–26):2187–2199 DOI 10.1016/S0024-3205(96)00576-0.
- Halicka HD, Bedner E, Darzynkiewicz Z. 2000. Segregation of RNA and separate packaging of DNA and RNA in apoptotic bodies during apoptosis. *Experimental Cell Research* 260(2):248–256 DOI 10.1006/excr.2000.5027.
- Hasselmann DO, Rapp G, Tilgen W, Reinhold U. 2001. Extracellular tyrosinase mRNA within apoptotic bodies is protected from degradation in human serum. *Clinical Chemistry* 47(8):1488–1489.
- Ito S, Nagata K. 2016. Biology of Hsp47 (Serpine H1), a collagen-specific molecular chaperone. *Seminars in Cell & Developmental Biology* 62:142–151 DOI 10.1016/j.semcdb.2016.11.005.
- Kang CY, Wang J, Axell-House D, Soni P, Chu ML, Chipitsyna G, Sarosiek K, Sendekci J, Hyslop T, Al-Zoubi M, Yeo CJ, Arafat HA. 2014. Clinical significance of serum *COL6A3* in pancreatic ductal adenocarcinoma. *Journal of Gastrointestinal Surgery* 18(1):7–15 DOI 10.1007/s11605-013-2326-y.
- Ke D, Li H, Zhang Y, An Y, Fu H, Fang X, Zheng X. 2017. The combination of circulating long noncoding RNAs AK001058, INHBA-AS1, MIR4435-2HG, and CEBPA-AS1 fragments in plasma serve as diagnostic markers for gastric cancer. *Oncotarget* 8:21516–21525 DOI 10.18632/oncotarget.15628.

- Kishikawa T, Otsuka M, Ohno M, Yoshikawa T, Takata A, Koike K. 2015. Circulating RNAs as new biomarkers for detecting pancreatic cancer. *World Journal of Gastroenterology* 21(28):8527–8540 DOI 10.3748/wjg.v21.i28.8527.
- Kopreski MS, Benko FA, Kwak LW, Gocke CD. 1999. Detection of tumor messenger RNA in the serum of patients with malignant melanoma. *Clinical Cancer Research* 5(8):1961–1965.
- Lo KW, Lo YM, Leung SF, Tsang YS, Chan LY, Johnson PJ, Hjelm NM, Lee JC, Huang DP. 1999. Analysis of cell-free epstein-barr virus associated RNA in the plasma of patients with nasopharyngeal carcinoma. *Clinical Chemistry* 45(8 Pt1):1292–1294.
- Ma R, Tao J, Kang X. 2012. Circulating microRNAs in cancer: origin, function and application. *Journal of Experimental & Clinical Cancer Research* 31(1):38–46 DOI 10.1186/1756-9966-31-38.
- Maru DM, Singh RR, Hannah C, Albarracin CT, Li YX, Abraham R, Romans AM, Yao H, Luthra MG, Anandasabapathy S. 2009. MicroRNA-196a is a potential marker of progression during Barrett's metaplasia-dysplasia-invasive adenocarcinoma sequence in esophagus. *American Journal of Pathology* 174(5):1940–1948 DOI 10.2353/ajpath.2009.080718.
- Pucciarelli S, Rampazzo E, Briarava M, Maretto I, Agostini M, Digito M, Keppel S, Friso ML, Lonardi S, Paoli AD. 2012. Telomere-Specific Reverse Transcriptase (hTERT) and cell-free RNA in plasma as predictors of pathologic tumor response in rectal cancer patients receiving neoadjuvant chemoradiotherapy. *Annals of Surgical Oncology* 19(9):3089–3096 DOI 10.1245/s10434-012-2272-z.
- Qiao J, Fang C-Y, Chen S-X, Wang X-Q, Cui S-J, Liu X-H, Jiang Y-H, Wang J, Zhang Y, Yang P-Y, Liu F. 2015. Stroma derived COL6A3 is a potential prognosis marker of colorectal carcinoma revealed by quantitative proteomics. *Oncotarget* 6:29929–29946 DOI 10.18632/oncotarget.4966.
- Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. 2004. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 6(1):1–6 DOI 10.1016/S1476-5586(04)80047-2.
- Rykova EY, Wunsche W, Brizgunova OE, Skvortsova TE, Tamkovich SN, Senin IS, Laktionov PP, Sczakiel G, Vlassov VV. 2006. Concentrations of circulating RNA from healthy donors and cancer patients estimated by different methods. *Annals of the New York Academy of Sciences* 1075(1):328–333 DOI 10.1196/annals.1368.044.
- Shen J, Kong W, Wu Y, Ren H, Wei J, Yang Y, Yang Y, Yu L, Guan W, Liu B. 2017. Plasma mRNA as liquid biopsy predicts chemo-sensitivity in advanced gastric cancer patients. *Journal of Cancer* 8(3):434–442 DOI 10.7150/jca.17369.
- Sisco KL. 2001. Is RNA in serum bound to nucleoprotein complexes? *Clinical Chemistry* 47(9):1744–1745.
- Sole C, Arnaiz E, Manterola L, Otaegui D, Lawrie CH. 2019. The circulating transcriptome as a source of cancer liquid biopsy biomarkers. *Seminars in Cancer Biology* In Press DOI 10.1016/j.semcan.2019.01.003.



- Stahl D, Braun M, Gentles AJ, Lingohr P, Walter A, Kristiansen G, Gütgemann I.** 2017. Low BUB1 expression is an adverse prognostic marker in gastric adenocarcinoma. *Oncotarget* 8:76329–76339 DOI [10.18632/oncotarget.19357](https://doi.org/10.18632/oncotarget.19357).
- Szasz AM, Lanczky A, Nagy A, Forster S, Hark K, Green JE, Boussioutas A, Busuttil R, Szabo A, Gyorffy B.** 2016. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget* 7:49322–49333 DOI [10.18632/oncotarget.10337](https://doi.org/10.18632/oncotarget.10337).
- Thorsen K, Sorensen KD, Brems-Eskildsen AS, Modin C, Gaustadnes M, Hein AM, Kruhoffer M, Laurberg S, Borre M, Wang K, Brunak S, Krainer AR, Torring N, Dyrskjot L, Andersen CL, Orntoft TF.** 2008. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Molecular & Cellular Proteomics* 7:1214–1224 DOI [10.1074/mcp.M700590-MCP200](https://doi.org/10.1074/mcp.M700590-MCP200).
- Wang Q, Wen YG, Li DP, Xia J, Zhou CZ, Yan DW, Tang HM, Peng ZH.** 2012. Upregulated INHBA expression is associated with poor survival in gastric cancer. *Medical Oncology* 29(1):77–83 DOI [10.1007/s12032-010-9766-y](https://doi.org/10.1007/s12032-010-9766-y).
- Wang L, Zhu J-S, Song M-Q, Chen G-Q, Chen J-L.** 2006. Comparison of gene expression profiles between primary tumor and metastatic lesions in gastric cancer patients using laser microdissection and cDNA microarray. *World Journal of Gastroenterology* 12(43):6949–6954 DOI [10.3748/wjg.v12.i43.6949](https://doi.org/10.3748/wjg.v12.i43.6949).
- Wu ZB, Cai L, Lin SJ, Leng ZG, Guo YH, Yang WL, Chu YW, Yang SH, Zhao WG.** 2016. Heat shock protein 47 promotes glioma angiogenesis. *Brain Pathology* 26(1):31–42 DOI [10.1111/bpa.12256](https://doi.org/10.1111/bpa.12256).
- Yamamoto N, Kinoshita T, Nohata N, Yoshino H, Itesako T, Fujimura L, Mitsuhashi A, Usui H, Enokida H, Nakagawa M, Shozu M, Seki N.** 2013. Tumor-suppressive microRNA-29a inhibits cancer cell migration and invasion via targeting HSP47 in cervical squamous cell carcinoma. *International Journal of Oncology* 43(6):1855–1863 DOI [10.3892/ijco.2013.2145](https://doi.org/10.3892/ijco.2013.2145).
- Youden WJ.** 1950. Index for rating diagnostic tests. *Cancer* 3(1):32–35.
- Zhang X, Wang C, Wang L, Du L, Wang S, Zheng G, Li W, Zhuang X, Zhang X, Dong Z.** 2012. Detection of circulating Bmi-1 mRNA in plasma and its potential diagnostic and prognostic value for uterine cervical cancer. *International Journal of Cancer* 131(1):165–172 DOI [10.1002/ijc.26360](https://doi.org/10.1002/ijc.26360).
- Zhao H, Wen J, Dong X, He R, Gao C, Zhang W, Zhang Z, Shen L.** 2017. Identification of AQP3 and CD24 as biomarkers for carcinogenesis of gastric intestinal metaplasia. *Oncotarget* 8:63382–63391 DOI [10.18632/oncotarget.18817](https://doi.org/10.18632/oncotarget.18817).
- Zhou D, Tang W, Liu X, An H-X, Zhang Y.** 2017. Clinical verification of plasma messenger RNA as novel noninvasive biomarker identified through bioinformatics analysis for lung cancer. *Oncotarget* 8:43978–43989 DOI [10.18632/oncotarget.16701](https://doi.org/10.18632/oncotarget.16701).