Heliyon 9 (2023) e18242

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Analysis and application of RNA binding protein gene pairs to predict the prognosis of gastric cancer

Zhi-kun Ning [a,1], Hua-kai Tian [b,1,*], Jiang Liu [c,1], Ce-gui Hu [c], Zi-tao Liu [c], Hui Li [d,**], Zhen Zong [c,***]

[a] Department of Day Ward, The First Affiliated Hospital of Nanchang University, Nanchang, China
[b] Department of General Surgery, First Affiliated Hospital of Nanchang University, Nanchang, China
[c] Department of Gastrointestinal Surgery, The Second Affiliated Hospital of Nanchang University, Nanchang, China
[d] Department of Rheumatology, The First Affiliated Hospital of Nanchang University, Nanchang, China

ARTICLE INFO

Keywords:
RNA-Binding proteins
Gene pairs
Signature
Gastric cancer
The cancer genome atlas
Gene expression omnibus

ABSTRACT

Background: RNA-binding proteins (RBPs) are closely related to tumors, but little is known about the mechanism of RBPs in tumorigenesis and progression of gastric cancer (GC). As genes do not usually act alone in the pathway deregulation, gene pair combinations are more likely to become stable and accurate biomarkers. The purpose of our research is to establish a novel signature based on RBP gene pairs to predict the prognosis of gastric cancer patients.

Methods: We downloaded genetic and clinical information from the TCGA and GEO database. TCGA and GSE13911 were used for screening differentially expressed genes (DEGs). The RBP genes were gathered from previous studies and employed to screen out DE-RBP genes after intersecting with DEGs. Samples were classified according to the relative expression of each pair of DE-RBP genes. The univariate Cox regression analysis and random forest were used to identify hub gene pairs to construct signature for predicting the prognosis of gastric cancer. Time-dependent ROC curves and KM survival curves were performed to evaluate the signature. GSEA was performed in TCGA training cohort and GSE62254 testing cohort to analyze enrichment pathways. Finally, the influence of these gene pairs on the prognosis of GC patients was further elucidated respectively through the combination of high and low expression of the two genes in each hub gene pair.

Results: We screened out 6 hub RBP gene pairs (COL5A2/FEN1, POP1/GFRA1, EXO1/PLEKHS1, SLC39A10/CHI3L1, MMP7/PPP1R1 B and SLC5A6/BYSL) to predict the prognosis of patients with gastric cancer. Using the optimal cut-off value to divide patients into high-risk and low-risk groups in the training and testing cohort, we found that the overall survival (OS) of the low-risk group was higher than that of the high-risk group ($P < 0.05$). The area under the ROC curves for 1, 3, and 5 years were (0.659, 0.744, 0.758) and (0.624, 0.650, 0.653) in two cohorts. Univariate and multivariate Cox regression analysis showed that 6 RBP gene pairs signature were independent prognostic factors for gastric cancer ($P < 0.05$). In addition, the prognostic survival analysis showed that COL5A2-high/FEN1-low, POP1-low/GFRA1-high, EXO1-low/PLEKHS1-low,

* Corresponding author.
** Corresponding author.
*** Corresponding author. Department of Gastrointestinal Surgery, the Second Affiliated Hospital of Nanchang University, 1 MinDe Road, Nanchang 330006, China.
E-mail addresses: 972687734@qq.com (H.-k. Tian), lihui0791nc@126.com (H. Li), ndefy16133@ncu.edu.cn (Z. Zong).
[1] Zhi-kun Ning, Hua-kai Tian, and Jiang Liu contributed equally to this study.

https://doi.org/10.1016/j.heliyon.2023.e18242
Received 6 July 2022; Received in revised form 28 June 2023; Accepted 12 July 2023
Available online 17 July 2023

SLC39A10-high/CHI3L1-low, MMP7-high/PPP1R1 B-low, SLC5A6-low/BYSL-low had worse OS ($P < 0.05$). And the gene correlation analysis showed that there was no obvious correlation between the genes in each gene pairs except SLC5A6/BYSL and POP1/GFRA1. Finally, GSEA analysis showed that the high-risk group was enriched in tumor migration, invasion and growth-related pathways.

Conclusion: Our study identified a novel 6 RBP gene pairs signature to predict the prognosis of gastric cancer patients and provide potential targets for clinical gene therapy.

1. Introduction

Gastric cancer (GC) is the third cancer-related cause of death in the world and one of the most common malignant tumors [1]. In recent years, despite the continuous improvement and optimization of GC treatment methods and related technologies, the 5-year survival rate of GC patients still does not exceed 30% [2–4]. On the one hand, most of the patients were found to be advanced or metastasized due to the anatomical position of the stomach and the atypical clinical manifestations of GC; On the other hand, patients sometimes have different prognosis at the same stage. It is inevitable that accurately predicting the patient's prognosis can benefit patients in subsequent treatment [5,6]. At present, it has been reported that many clinicopathological factors, genes, etc. Can be used as prognostic factors, but most of them lack clinical practicality or other limitations [7]. Therefore, there is a clinical need for a marker predictor that more accurately predicts the prognosis of GC patients.

RNA binding protein (RBP) is a group of proteins that regulate gene transcription and mainly act on RNA processing, such as mRNA splicing, localization, polyadenylation, translocation, stability, translation, etc [8]. Recent studies have shown that RBP plays a vital role in the occurrence and development of tumors and is used to construct tumor prognostic models [9–11]. The occurrence and development of tumors may be under the joint action of multiple genes [12]. In recent years, people have proposed a gene expression data processing method based on the relative expression level of gene expression, which overcomes the shortcomings of standardization and large-scale in different data processing, and has achieved reliable results in a variety of studies [13,14]. The gene pairs obtained by CytoPred in view of top scoring pair (TSP)-based decision tree could well predict the survival and prognosis of acute myeloid leukemia (AML) patients [15]. However, there are few researches on RBP gene pairs, especially the impact on the prognosis of gastric cancer has not been reported.

Therefore, in this study, the expression levels of all RBP genes in each tumor sample were compared in pairs, and samples were classified with the relative expression of each gene pairs. We confirmed the role of the RBP gene pairs in predicting the prognosis of gastric cancer. At the same time, we further carried out correlation analysis and pathway enrichment analysis on the genes of each hub gene pair. And according to the high and low expression of the two genes in each gene pair, it is divided into different combinations of gene pairs to further verify its role in tumor progression.

2. Methods and materials

2.1. Data collection and pre-processing

Our study data were based on The Cancer Genome Atlas Program (TCGA) and Gene Expression Omnibus (GEO) database. RNA-sequencing dataset of GC patients was downloaded in TCGA (https://portal.gdc.cancer.gov/projects) and transformed into transcripts per kilobase million (TPM) value. Clinical information of these patients was also obtained together, of course, only with patient samples of complete follow-up information would be included in subsequent analysis. Next, GSE13911 and GSE62254 were acquired in GEO (https://www.ncbi.nlm.nih.gov/geo/). We removed some samples whose principal component analysis results were inconsistent with the grouping, including a tumor sample "GSM350469″ and 5 normal samples "GSM350415″, "GSM350423″, "GSM350427″, "GSM350431″ and "GSM350438". Finally, a total of 37 tumor samples and 26 normal samples were acquired in GSE13911. GSE62254 was considered as testing cohort for the later operation.

2.2. Screening of DEGs in TCGA and GSE13911, selection and bioinformatic analysis of DE-RBP genes

Differentially expressed genes (DEGs)were obtained using the "DESeq2" package in TCGA and the "limma" package in GSE13911. The screening criteria between tumor and normal samples were $|\log2 \text{ fold change (FC)}| > 1$ and adjusted p-value <0.05. We made an upset plot to show a list of the human RBPs from previous studies [16–20] (https://www.xrnax.com/) (http://geneontology.org/), (http://www.rbptd.com) (https://www.genscript.com/). The more detailed genetic information could be referred to in the Supplement Table 1. Next, we took the intersection of DEGs with RBP genes and then confirmed differentially expressed RBP genes (DE-RBP).

To further explore the function of DE-RBP genes, we carried out Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomics (KEGG) pathway enrichment analysis using the "clusterProfiler" R package. GO analysis included biological processes, cell components and molecular functions, and the generated diagram showed the top 10 items of each part. KEGG analysis displayed the 20 most enriched pathways.

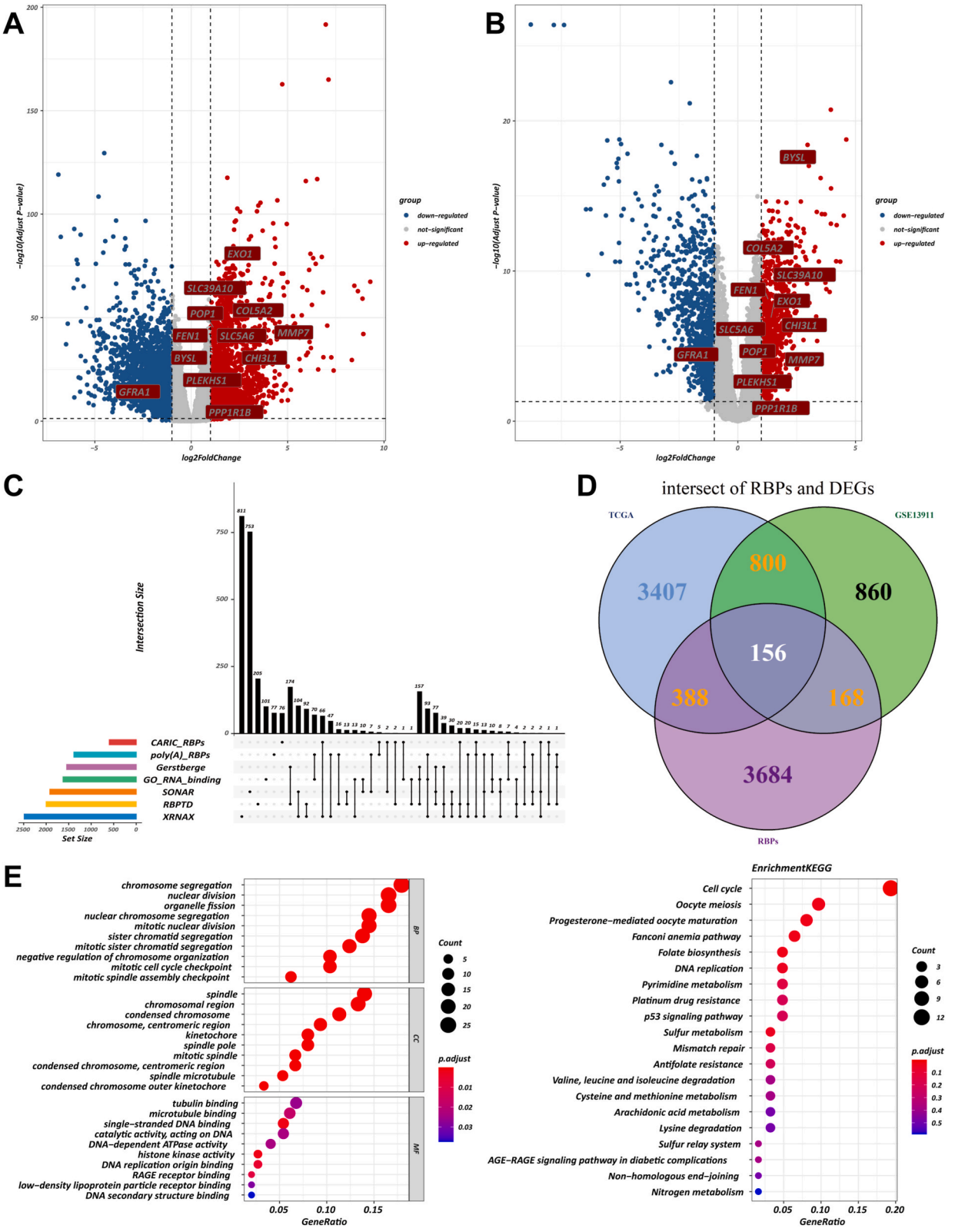

Fig. 1. Selection and functional enrichment analysis of differentially expressed RBP genes (DE-RBP genes). (A) Volcano plot of DEGs in TCGA. (B) Volcano plot of DEGs in GSE13911. (C) Upset diagram of integration from 7 sources. (D) Venn diagram: making intersection of 3 cohorts to screen DE-RBP genes. (E) Bubble plots of GO and KEGG enrichment analysis for DE-RBP genes.

2.3. Screening of gene pairs and construction of signature

After the above collection, aggregation and processing, we obtained a total of 156 DE-RBP genes. Each DE-RBP gene was paired with others separately to form a series of DE-RBP gene pairs (DRGPs), and the value of each gene pair was determined by the relative ranking of the expression levels of the two genes. If the expression value of the first gene was lower than that of the second gene, the value of this DRGP was output 1; otherwise, it was 0. Subsequently, we deleted some DRGPs that expressed a unique value with 0 or 1 in more than 80% samples, and the remaining DRGPs were as candidate for the following signature. Through calculation, 988 gene pairs were identified with common differences.

With a univariate Cox regression analysis performed in TCGA training cohort, 102 DRGPs with survival differences were finally determined through the limitation of p-value <0.05. Then, the "randomForestSRC" R package was applied to pick out the most important DRGPs which were subsequently brought into multivariate Cox analysis. Finally, 6 hub gene pairs were identified and used to build the prognostic signature which was performed by risk score $= \Sigma$ expression value of DRGPi $\times$ Cox coefficient of DRGPi.

2.4. Validation of signature

The optimal cut-off value was determined using the "survminer" R package in TCGA training cohort. Time-dependent ROC of TCGA and GSE62254 were respectively applied to evaluate the signature using the "survivalROC" R package. Based on the optimal cut-off value, we respectively divided the TCGA and GSE62254 samples into high- and low-risk groups. Heat maps were employed to display the expression tendency of the 6 gene pairs in different risk groups using the "pheatmap" R package. The Kaplan-Meier survival curves were plotted to reveal differences in overall survival of patients in high- and low-risk groups using the "survival" R package.

2.5. Analysis of the impact of high and low expression combinations of two genes in each hub gene pair on survival

We observed the expression differences of the 12 genes from the 6 hub gene pairs between high- and low-risk groups in the TCGA dataset. For further research, the expression of each gene was divided into high and low level according the optimal cut off. Subsequently, we combined the high and low expression of the two genes in each hub gene pair, so a cohort were sorted into 4 groups to show the effect of this gene pair on survival by comparison between groups. Meanwhile, the same operation was conducted in GSE62254 testing cohort.

In addition, we did a Pearson correlation analysis (Pearson' r) on the two genes in each hub gene pair to explore whether there was a linear relationship between them.

2.6. Independent prognostic factors

We performed univariate and multivariate Cox proportional-hazards analysis to identify independent prognostic factors using the "rms" R package in TCGA and GSE62254. Here, clinical information was brought into analysis including gender, age, stage, TNM stage, grade, tumor location and Lauren classification. And a value with $p < 0.05$ was considered significant.

2.7. GSEA

To observe the difference of pathway between high- and low-risk groups, we carried out gene set enrichment analysis (GSEA) using the "clusterProfiler" R package. The enrichment analysis of KEGG and HALLMARK was done on TCGA and GSE62254 cohort, respectively, and the enriched pathways were displayed in two directions of activation and suppression.

2.8. Statistical analysis

All the statistical analysis was performed on R (version 4.0.2) and each package we used had been explained in the above. Only $p < 0.05$ in all testes was considered statistically significant (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ and ****$p < 0.0001$).

3. Result

3.1. Selection and functional enrichment analysis of DE-RBP genes

In this study, we performed a difference analysis on the TCGA and the GSE13911 to screen DEGs. After removing some samples whose principal component analysis results were disagree with the grouping results, we got 37 tumor samples and 26 normal samples in GSE13911 dataset. According to the threshold of $|\log2 \text{ fold change (FC)}| > 1$ and adjusted p-value <0.05, we totally identified 4751 DEGs in TCGA and 1984 DEGs in GSE13911, and severally depicted volcano maps to distinguish up-regulated and down-regulated genes (Fig. 1A and B). Moreover, we integrated a list of human RBP genes from previous studies (Gerstberger, SONAR, CARIC, Gene Ontology project, Poly(A) binding protein, XRNAX, RBPTD) and a total of 4396 RBP genes were found and aggregated (Supplementary Table 1 and Fig. 1C). Venn diagram was drawn to show the total 156 DE-RBP genes of 3 cohorts (Fig. 1D).

To obtain a comprehensive function understanding of these DE-RBP genes, GO and KEGG were performed and visualized using bubble plots (Fig. 1E). GO analysis showed that primary enrichment pathways were chromosome segregation, nuclear division and

organelle fission in BP, spindle, chromosomal region and condensed chromosome in CC, tubulin binding, microtubule binding and single-stranded DNA binding in MF. The result of KEGG analysis showed that major enrichment pathways of IGs were cell cycle, oocyte meiosis and progesterone-mediated oocyte maturation.

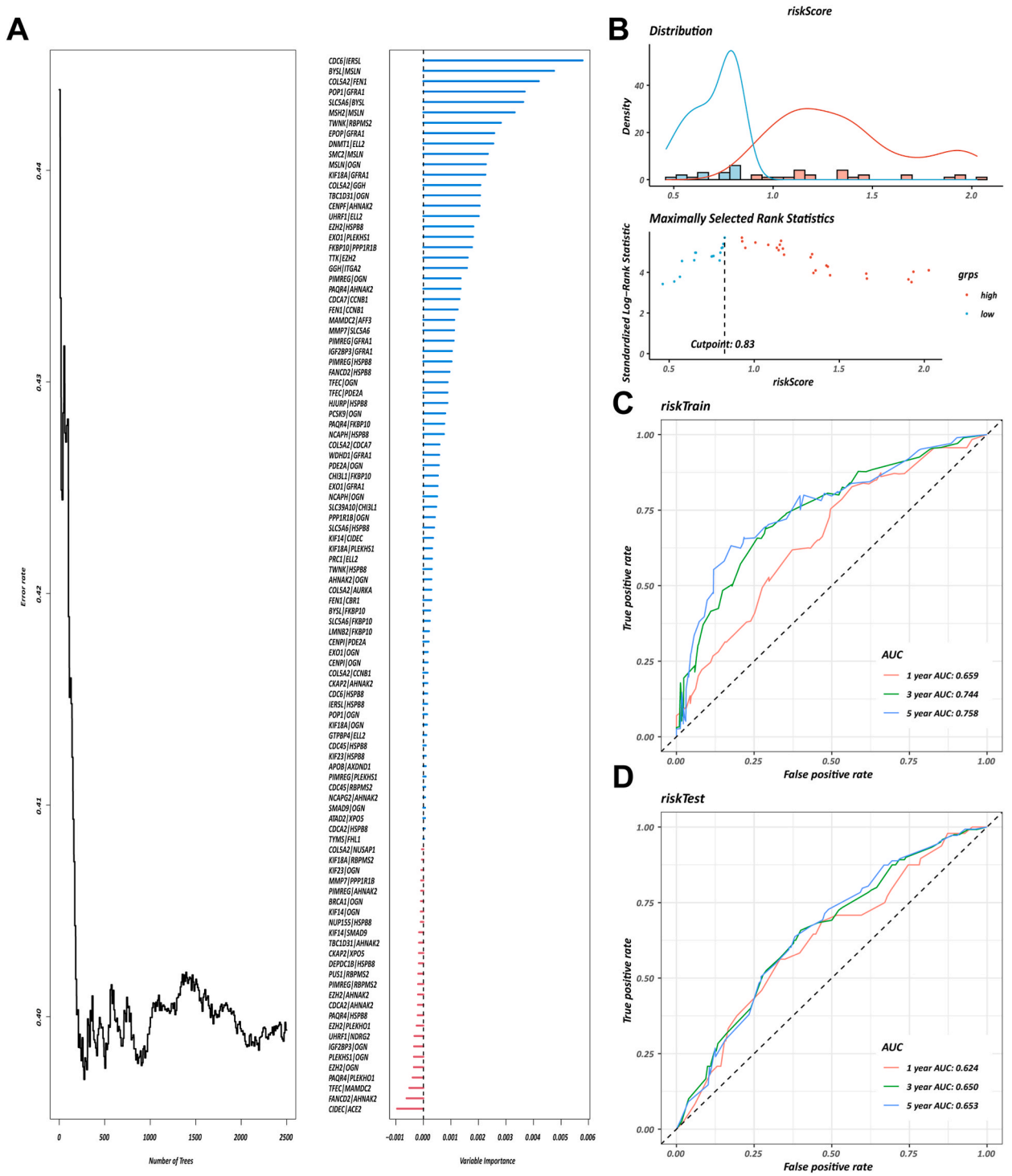

Fig. 2. Construction and validation of prognostic signature. (A) Visualization of Random Forest to screen hub gene pairs. (B) Selection of the optimal cut off value. (C) Time-dependent ROC curve of TCGA. (D) Time-dependent ROC curve of GSE62254.

3.2. Screening of RBP gene pairs, construction and validation of prognostic signature

After the process of pairing, calculation and screening of the above 156 DE-RBP genes, we obtained a total of 988 DRGPs with common differences. The gene pairs were calculated as follows.

1: DRGPi $<$ DRGPj; 0: DRGPi $\geq$ DRGPj.

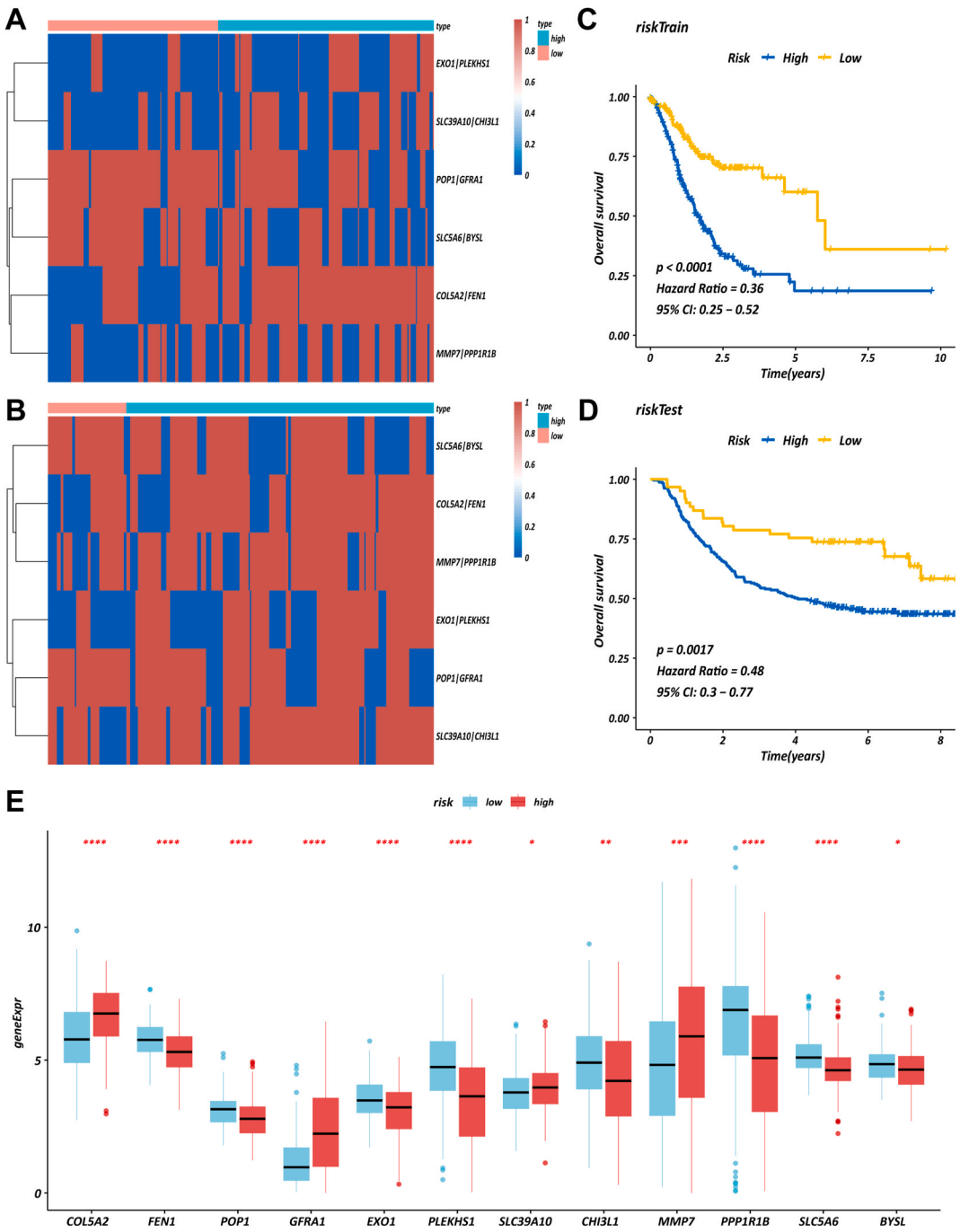

Fig. 3. Validation of prognostic signature and analysis of genes in hub gene pairs. (A) Heatmap of gene pairs in TCGA. (B) Heatmap of gene pairs in GSE62254. (C) Kaplan-Meier survival curve of overall survival (OS) between high- and low-risk groups in TCGA. (D) Kaplan-Meier survival curve of overall survival (OS) between high- and low-risk groups in GSE62254. (E) Differential expression analysis of each gene from 6 hub gene pairs between high- and low-risk groups in TCGA.

With a univariate Cox analysis in TCGA training cohort, 102 gene pairs were identified with survival differences according to p-value <0.05 (Supplementary Table 2). Random Forest algorithm was used to further select the most important DRGPs which were defined as hub gene pairs and put into multivariate Cox analysis to struct the prognostic signature (Fig. 2A and Supplementary Table 3). The final 6 hub gene pairs were COL5A2/FEN1, POP1/GFRA1, EXO1/PLEKHS1, SLC39A10/CHI3L1, MMP7/PPP1R1 B and SLC5A6/BYSL. The model was showed as risk score that was calculated by the gene pair expression value multiplied by the coefficient of multivariate Cox analysis, and the risk score of each patient was obtained. We regarded TCGA dataset as training cohort and GSE62254 as testing cohort, and divided each cohort into high- and low-risk groups according the optimal cut-off value calculated from the "survminer" R package (Fig. 2B). It was obvious that, compared with the low-risk group classification, the high-risk group in two cohorts both showed worse OS in KM curves (Fig. 3C–D). The areas under ROC curve for the risk score predicting OS at 1, 3 and 5 years were respectively 0.659, 0.744 and 0.758 in training cohort, and 0.624, 0.650 and 0.653 in testing cohort (Fig. 2C–D). The heat maps of the two cohorts both showed that POP1/GFRA1 and SLC5A6/BYSL were highly expressed in the low-risk group, while COL5A2/FEN1, EXO1/PLEKHS1, SLC39A10/CHI3L1 and MMP7/PPP1R1 B were highly expressed in the high-risk group (Fig. 3A–B). The risk score along with complete clinical information including gender, age, stage, TNM stage, grade, tumor location and Lauren classification were brought into univariate and multivariate Cox regression analysis to identify independent prognostic factors. The univariate Cox regression analysis in TCGA indicated that age (P = 0.038), stage (P < 0.001), T (P = 0.013) N (P < 0.001) M (P = 0.005) staging, risk score (P < 0.001) were prognostic factors for gastric cancer (Fig. 4A); the multivariate analysis showed that risk score (P < 0.001) was an independent risk factor for overall survival (Fig. 4B). The same operation was performed in the GSE62254 data set, and the results manifested Lauren (P < 0.001), Stage (P < 0.001), T (P < 0.001), N (P < 0.001), M (P < 0.001) staging, risk score (P < 0.001) in univariate analysis and Lauren (P = 0.036), M (P = 0.022), Stage (P = 0.035), risk score (P = 0.001) in multivariate analysis (Fig. 4C–D). In general, the analysis of the two datasets showed that risk score was an independent prognostic factor for gastric cancer, indicating that the established signature was reliable.

3.3. Analysis of the impact of high and low expression combinations of two genes in each hub gene pair on survival

Each gene from 6 hub gene pairs was performed differential expression analysis and found significant difference in the high- and

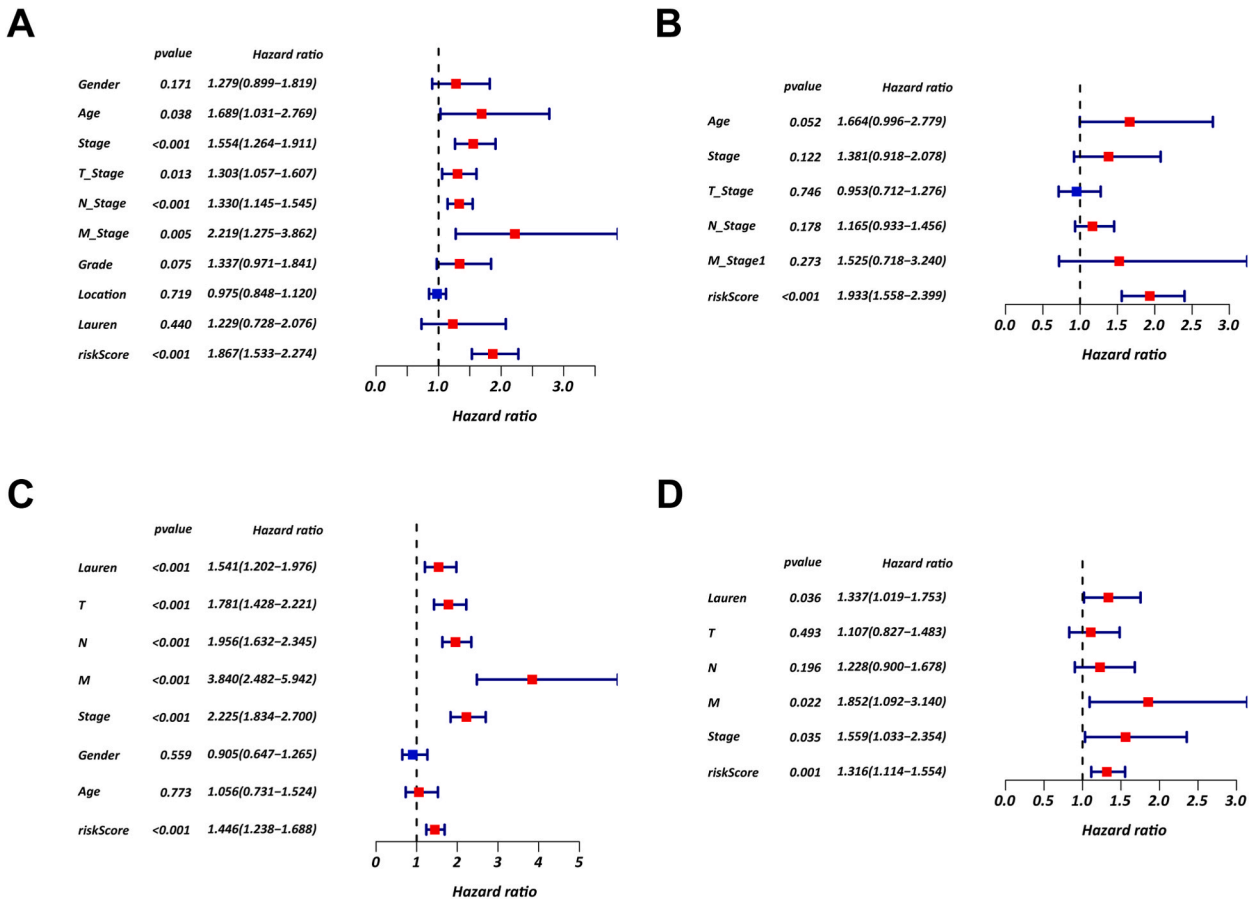

Fig. 4. The forest plots of univariate and multivariate Cox regression analysis to identify prognostic factors of gastric cancer. (A) Univariate Cox analysis in TCGA. (B) Multivariate Cox analysis in TCGA. (C) Univariate Cox analysis in GSE62254. (D) Multivariate Cox analysis in GSE62254.

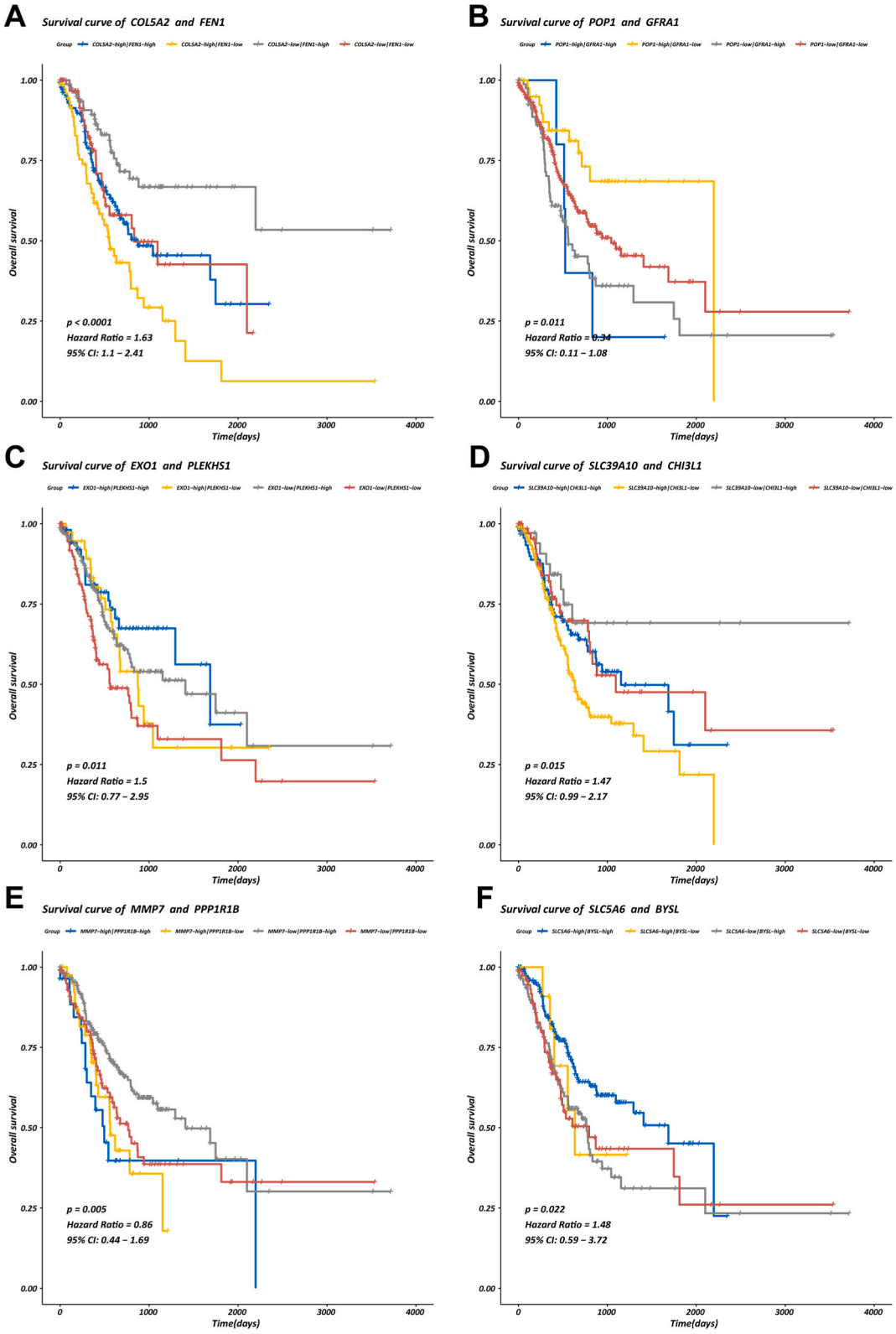

Fig. 5. Survival curves in TCGA cohort: the relationship between OS with the different combinations of high and low expression of the two genes in each hub gene pair. (A) COL5A2/FEN1. (B) POP1/GFRA1. (C) EXO1/PLEKHS1. (D) SLC39A10/CHI3L1. (E) MMP7/PPP1R1 B. (F) SLC5A6/BYSL.

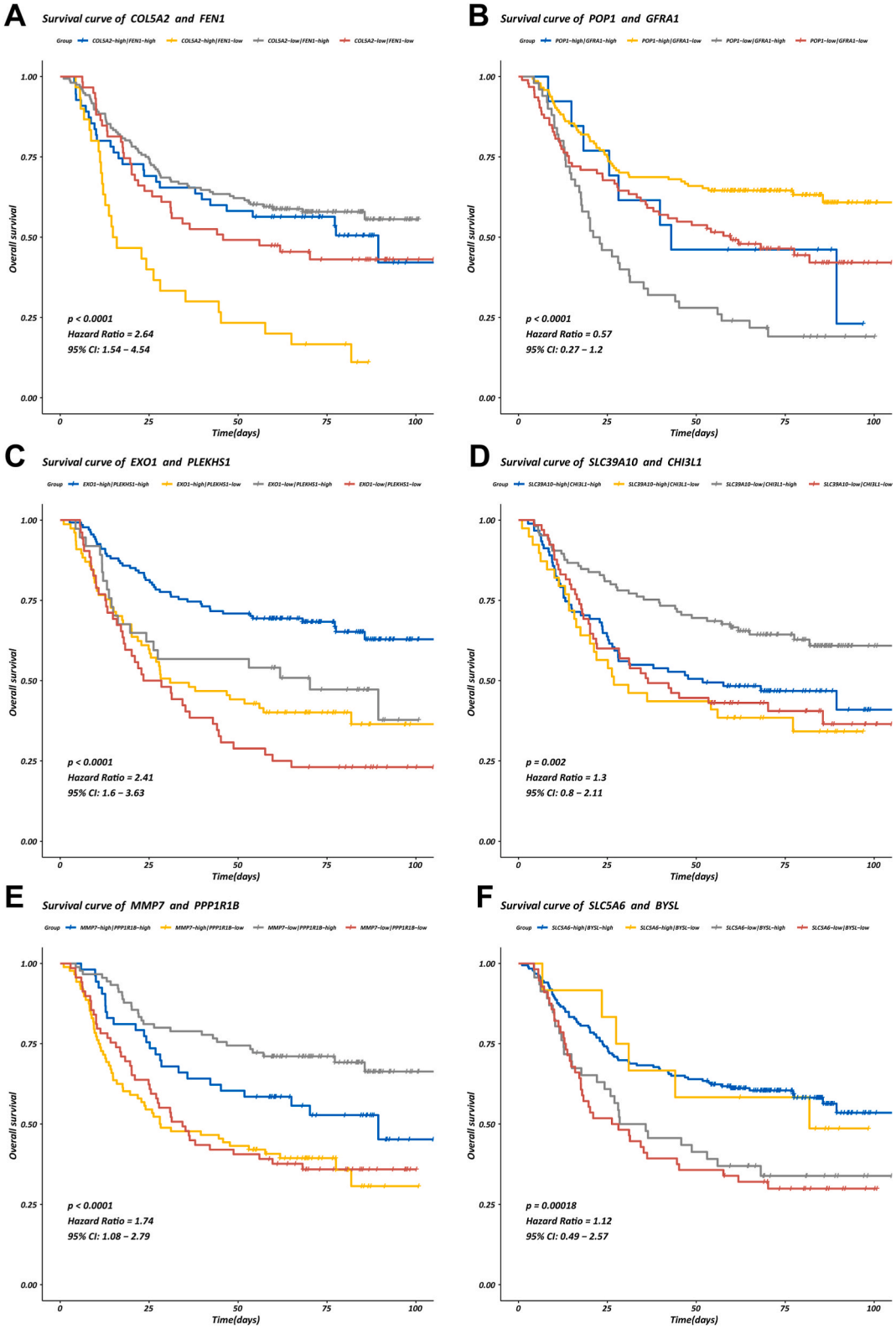

Fig. 6. Survival curves in GSE62254 cohort, the relationship between OS with the different combinations of high and low expression of the two genes in each hub gene pair. (A) COL5A2/FEN1. (B) POP1/GFRA1. (C) EXO1/PLEKHS1. (D) SLC39A10/CHI3L1. (E) MMP7/PPP1R1 B. (F) SLC5A6/BYSL.

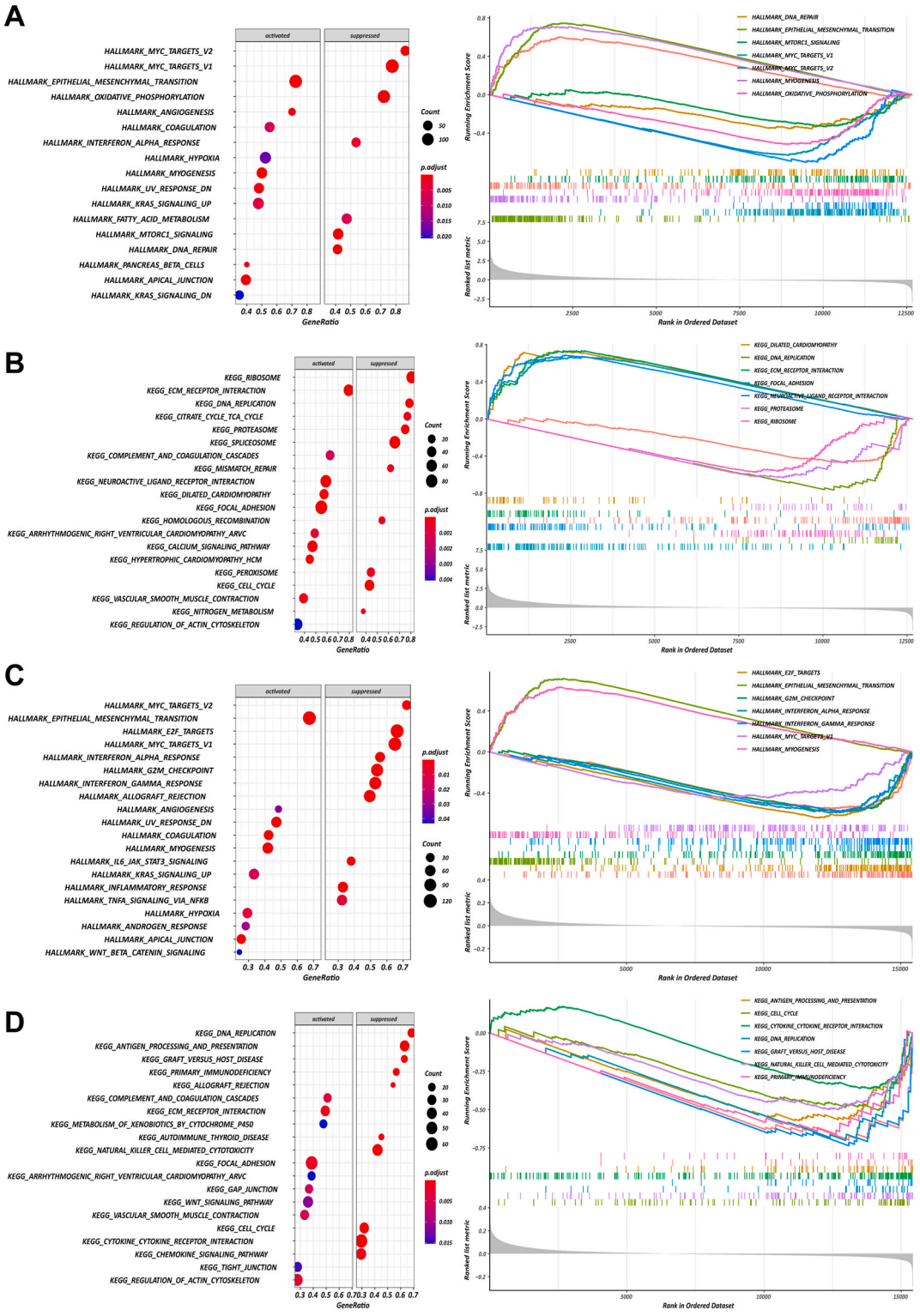

Fig. 7. Gene set enrichment analysis (GSEA) between high- and low-risk groups. (A) Visualization of HALLMARK pathway enrichment in TCGA. (B) Visualization of KEGG pathway enrichment in TCGA. (C) Visualization of HALLMARK pathway enrichment in GSE62254. (D) Visualization of KEGG pathway enrichment in GSE62254.

low-risk groups ($P < 0.05$). Meanwhile, the expression of each gene in high- and low-risk groups was analyzed comprehensively and visualized in the box plot (Fig. 3E). The COL5A2 gene in COL5A2/FEN1 gene pair had higher expression than FEN1 in the high-risk group that illustrated COL5A2 was an up-regulated gene and FEN1 was a down-regulated gene. Two genes in EXO1/PLEKHS1 were together highly expressed in the low-risk group, indicating that both genes were down-regulated. Similarly, we could understand that POP1 was a down-regulated gene and GFRA1 was an up-regulated gene in POP1/GFRA1, MMP7 was an up-regulated gene and PPP1R1 B was a down-regulated gene in MMP7/PPP1R1 B, SLC39A10 was an up-regulated gene and CHI3L1 was a down-regulated gene in SLC39A10/CHI3L1, SLC5A6 and BYSL were both down-regulated genes in SLC5A6/BYSL. Subsequently, we showed the effect of gene pairs on overall survival by combining the high and low expression levels of the two genes in a particular hub gene pair, which divided a cohort into 4 groups according to that gene pair. The analysis in TCGA indicated that COL5A2-high/FEN1-low resulted a worse OS (Fig. 5A); the similar consequence appeared while POP1-low/GFRA1-high (Fig. 5B), EXO1-low/PLEKHS1-low (Fig. 5C), SLC39A10-high/CHI3L1-low (Fig. 5D), MMP7-high/PPP1R1 B-low (Fig. 5E), SLC5A6-low/BYSL-low (Fig. 5F). And the same results occurred in GSE62254 (Fig. 6). This result further illustrated the role and relationship of the two genes in each gene pair in tumor progression. Interestingly, the gene correlation analysis in each gene pair showed that, except for SLC5A6/BYSL, POP1/GFRA1, there was no obvious correlation between the two genes of the other gene pairs. Even for the two gene pairs, the Pearson' r were only 0.61 and −0.32.

3.4. GSEA between high- and low-risk groups

In order to analyze enrichment pathways of gene pairs between high and low groups, we performed GSEA on the TCGA training cohort and GSE62254 testing cohort. From HALLMAKER in TCGA, we found that EPITHELIAL MESENCHYMAL TRANSITION, MYOGENESIS were enriched in the high-risk group, and DNA REPAIR, MTORC1 SIGNALING, MYC TARGETS V1, MYC TARGETS V2 and OXIDATIVE PHOSPHORYLATION were enriched in the low-risk group (Fig. 7A). While KEGG analysis in TCGA indicated that ECM RECEPTOR INTERACTION, NEUROACTTIVE LIGAND RECEPTOR INTERACTION, FOCAL ADHESION and DILATED CARDIOMYOPATHY were enriched in the high-risk group, and DNA REPLICATION, PROTEASOME, RIBOSOME were enriched in the low-risk group (Fig. 7B). In GSE62254 cohort, the enriched result of HALLMAKER was same as that in TCGA, while major pathways were enriched in the low-risk group, like ANTIGEN PROCESSING AND PRESENTATION, CELL CYCLE, CYTOKINE RECEPTOR INTERACTION, DNA REPLICATION, GRAFT VERSUS HOST DISEASE, NATURAL KILLER CELL MEDIATED CYTOTOXICITY, PRIMARY IMMUNODEFICIENCY. The results revealed that the selected gene pairs might affect the occurrence and development of gastric cancer through these pathways, such as epithelial-mesenchymal transition, myogenesis, and ECM receptor interaction in enriched activation pathways.

4. Discussion

As a highly malignant tumor, aggressiveness and high recurrence rate of GC make the prognosis of patients challenging [21]. A good prognosis cannot be guaranteed under a simple radical surgical excision, so comprehensive treatment is needed to improve the prognosis of patients [22]. However, this requires sensitive and reliable prognostic biomarkers to identify the prognosis of patients and determine which patients can significantly benefit from comprehensive treatment [23]. Although many gene-based prognostic markers have been found to predict the prognosis of GC patients [24–26], most of them lack clinical utility or other limitations. Therefore, there is an urgent need for highly relevant predictive markers that can be used to accurately predict and improve the prognosis of GC patients. Recent studies have shown that RBP appear to be closely related to the occurrence and development of tumors [9,27]. Some RBP genes participate in the progression of tumors through certain pathways, and effective predictive models have been established to predict the prognosis [28,29]. With the update of gene sequencing methods and data, however, it would be a challenge in clinical application when standardizing data from different sequencing platforms. Therefore, we introduce the concept of gene pairs to eliminate the influence of different data platforms and inter-individual standardization on the results. By forming a pair of specific RBP gene expression values through the specific expression relationship between each gene, we have obtained a new prediction signature, which is more suitable for individual research and clinical application. Moreover, we further explored the related pathways of RBP gene pairs and the correlation analysis between genes comprised gene pairs.

Our results suggested that the risk score composed by 6 RBP gene pairs could be used as a predictor of the prognosis of GC. Through univariate and multivariate Cox regression analysis of risk score with the clinicopathological parameters, it was proved that risk scores could be used as independent risk factors for the prognosis of GC, and we also verified this conclusion through testing dataset. Recent studies have shown that it is the pathway deregulation rather than that of a single gene, which may play a crucial role in triggering cancer, and inactivation of a pathway is usually caused by multiple genes [30,31]. At the same time, pairing of genes avoids false positives caused by fluctuations in the expression of a single gene and improves the stability and accuracy of the results [32].

It is worth noting that each gene in the 6 RBP gene pairs has different expressions in high and low groups, and shows different functions. Therefore, two genes in each hub gene pair were grouped by combining with different high and low expression levels. Finally, we saw interesting results that some gene pairs showed the same expression trend for the two genes, while others showed opposite. For example, the heat map manifested the gene pair COL5A2/FEN1 expressed highly in the high-risk group, but individual gene expression analysis showed that COL5A2 was a tumor-promoting gene and FEN1 was a tumor suppressor gene, and the COL5A2-high/FEN1-low combination had the worst prognosis. As far as we know, COL5A2 is involved in the occurrence of tumors and is related to the poor clinical prognosis and survival rate of tumor patients [33,34]. And FEN1 plays a role in tumor DNA replication and repair. FEN1 inhibitors have the potential to treat homologous recombination-deficient cancers [35,36]. This just proves that our conclusion is reasonable. Similarly, the remaining three gene pairs (POP1/GFRA1, SLC39A10/CHI3L1, MMP7/PPP1R1 B) showed the same trend.

SLC39A10/CHI3L1 and MMP7/PPP1R1 B were highly expressed in the high-risk group and POP1/GFRA1 in the low-risk group, while the two genes showed opposite expression trends in the single gene expression analysis. POP1, CHI3L1 and PPP1R1 B are tumor suppressor genes, GFRA1, SLC39A10 and MMP7 are tumor promoter genes, and POP1-low/GFRA1-high, SLC39A10-high/CHI3L1-low and MMP7-high/PPP1R1 B-low have the worst prognosis. Previous studies have shown that breast cancer patients with high POP1 expression benefit from immunotherapy and that patients are more likely to respond to immunotherapy [37]. CHI3L1 can promote tumor metastasis in patients with gastric cancer. Serum levels of CHI3L1 are significantly elevated in patients with gastric or breast cancer and can be used as a marker for patients with metastatic gastric cancer [38]. Overexpression of PPP1R1 B in pancreatic cancer significantly enhanced the invasive ability and metastatic activity of tumor cells [39]. These studies appear to have some discrepancies with our findings. It has been shown that GFRA1 is reactivated by DNA demethylation in patients with colorectal cancer and is associated with poor patient prognosis [40]. Increased expression of the SLC39A10 gene may be a detrimental treatment for survival in patients with hepatocellular carcinoma, with its ability to promote tumor aggressiveness [41]. MMP7 was elevated in patients with gastric cancer and the number of MMP7 positive cells was significantly lower in patients who showed significant improvement after treatment [42]. EXO1/PLEKHS1 gene pair had a high expression in the high-risk group according the heat map, while EXO1 and PLEKHS1 were both tumor suppressor genes in individual gene expression analysis and the combination of EXO1-low/PLEKHS1-low had the worse prognosis. Exonuclease-1 (Exo-1) is an important nuclease involved in the mismatch repair system, helping to maintain genome stability, regulate DNA recombination and mediate cell cycle arrest. Errors in DNA repair and replication can lead to the accumulation of mutations, which result in the development of cancer. EXO-1 mutation causes the premature termination of amino acid synthesis in EXO1 protein, so similar to a typical mutation of function loss, this mutation may inactivate DNA damage repair and apoptosis in affected cancer cells [43]. There is a lack of relevant research on PLEKHS1. Although it is highly expressed in gastric cancer, our results are similar to previous studies: PLEKHS1 is a protective factor for GC patients, and its expression is higher in low-risk groups [44,45]. Maybe the EXO1/PLEKHS1 gene pair triggers an unknown pathway which leads to different results. SLC5A6/BYSL was highly expressed in the low risk group and both SLC5A6 and BYSL were tumor suppressor genes in the individual gene expression analysis. SLC5A6-low/BYSL-low has the worst prognosis. The SLC5A6 gene functions in the body to maintain the body's uptake of biotin and ubiquitin. SLC5A6 promotes intestinal uptake of biotin through encoding Smvt, maintains intestinal flora balance and controls intestinal tumourigenesis [46]. BYSL is a key factor in embryo implantation and development and plays a role in a variety of cancers, and mutations in the BYSL gene will significantly increase tumourigenesis and progression [47,48]. These findings suggest that the pairing of two genes with the same or opposite effect can predict tumor prognosis, and that there may be an unknown pathway between the two genes that act together to predict the course of the tumor. Correlation analysis also confirmed this possibility, due to most gene pairs have no obvious correlation between the two composed genes but they could predict the prognosis of GC after pairing. The discovery of this result provided evidence for previous studies that the occurrence and development of tumors were determined by multiple genes, and two or more genes with no obvious correlation affected a certain pathway and led to tumor progression.

In order to study the pathways that RBP pairs may affect, we performed GSEA analysis on the high-risk and low-risk groups. The result of HALLMARK revealed that EPITHELIAL MESENCHYMAL TRANSITION and MYOGENESIS pathways were enriched in high-risk group. Studies have shown that EPITHELIAL MESENCHYMAL TRANSITION plays an important role in migration and invasion of tumor cell [49,50]. And MYOGENESIS plays an important part in tumor growth, indicating a poor prognosis [51,52]. In the results of KEGG enrichment, we have reached a similar conclusion. The enriched pathways in the high-risk group were ECM RECEPTOR INTERACTION, NEUROACTTIVE LIGAND RECEPTOR INTERACTION, FOCAL ADHESION, DILATED CARDIOMYOPATHY. The ECM RECEPTOR INTERACTION pathway is related to tumor metastasis and is an accomplice of the occurrence, development and poor survival of GC [53,54]. The FOCAL ADHESION pathway is also associated with tumor migration and invasion [55]. It is not difficult to find from our results that the RBP gene pair seems to be involved in tumor growth, invasion, metastasis and prognosis. Previous research mostly only focused on a single gene and the pathway it affects, but pathway disorders are mostly caused by multiple genes. Only for tumor-related single gene therapy, there may be little effect. The study of gene pairs can greatly compensate for this defect. The joint action of multiple genes may affect a certain process or multiple processes of the tumor, and effective targeted treatment will help improve the accuracy and effectiveness of the treatment.

Although the exact mechanism is still unclear, more and more studies point that RBPs have a vital role in the occurrence and development of tumors. Our research also proves this view that the risk score of RBP gene pairs can predict the prognosis of GC and is an independent risk factor for prognosis. Compared with other studies on individual genes, this study matched RBP genes to form gene pairs with specific expression value, which removed the limitations of different data processing and more accurately predicted the prognosis of GC. Meanwhile, the paired genes reflecting no obvious correlation between the two genes did influence the progression of some pathways, further suggesting that the pathway deregulation was caused by multiple genes. The combination of gene pairs can use these potential biological responses to provide tumors with better biomarkers and curative effects than a single gene.

Finally, several limitations of our study are worth noticing. The prognostic signature is mainly based on the TCGA database that means it is a retrospective study. Although verified by the GEO database, the model is necessary to use prospective data with complete clinical information and gene expression information to verify its clinical practicality. Additionally, our study only indicates that RBP gene pairs affect correlated pathways, so functional analysis of related genes from multiple angles are indispensable to explore the possible internal connections and common mechanisms between genes that may provide more accurate guidance for clinical treatment in the future.

5. Conclusion

Our study identified 6 RBP gene pairs to predict the prognosis of gastric cancer patients and provide potential targets for clinical gene therapy. At the same time, it was confirmed that tumor progression was caused by the out-of-control of multiple genes, providing a theoretical basis for future research on specific mechanisms.

Author contributions statement

Zhi-kun Ning: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.
Hua-kai Tian, Jiang Liu: Contributed reagents, materials, analysis tools or data; Wrote the paper.
Ce-gui Hu: Analyzed and interpreted the data; Wrote the paper.
Zi-tao Liu: Contributed reagents, materials, analysis tools or data; Wrote the paper.
Zhen Zong, Hui Li: Conceived and designed the experiments; Wrote the paper.

Data availability statement

The authors are unable or have chosen not to specify which data has been used.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e18242.

References

[1] E.C. Smyth, M. Nilsson, H.I. Grabsch, N.C. van Grieken, F. Lordick, Gastric cancer, Lancet 396 (10251) (2020) 635–648.
[2] L. Zong, M. Abe, Y. Seto, J. Ji, The challenge of screening for early gastric cancer in China, Lancet 388 (10060) (2016) 2606.
[3] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, Ca - Cancer J. Clin. 65 (1) (2015) 5–29.
[4] H.H. Kim, S.U. Han, M.C. Kim, W. Kim, H.J. Lee, S.W. Ryu, G.S. Cho, C.Y. Kim, H.K. Yang, D.J. Park, K.Y. Song, S.I. Lee, S.Y. Ryu, et al., Effect of laparoscopic distal gastrectomy vs open distal gastrectomy on long-term survival among patients with stage I gastric cancer: the KLASS-01 randomized clinical trial, JAMA Oncol. 5 (4) (2019) 506–513.
[5] T. Li, X. Mo, L. Fu, B. Xiao, J. Guo, Molecular mechanisms of long noncoding RNAs on gastric cancer, Oncotarget 7 (8) (2016) 8601–8612.
[6] P. Li, S. Chen, H. Chen, X. Mo, T. Li, Y. Shao, B. Xiao, J. Guo, Using circular RNA as a novel type of biomarker in the screening of gastric cancer, Clin. Chim. Acta (2015) 444132–444136.
[7] J.H. Lee, K.K. Chang, C. Yoon, L.H. Tang, V.E. Strong, S.S. Yoon, Lauren histologic type is the most important factor associated with pattern of recurrence following resection of gastric adenocarcinoma, Ann. Surg. 267 (1) (2018) 105–113.
[8] S. Mohibi, X. Chen, J. Zhang, Cancer the'RBP'eutics-RNA-binding proteins as therapeutic targets for cancer, Pharmacol. Ther. (2019), 203107390.
[9] F. Gebauer, T. Schwarzl, J. Valcárcel, M.W. Hentze, RNA-binding proteins in human genetic disease, Nat. Rev. Genet. 22 (3) (2021) 185–198.
[10] L. Wurth, Versatility of RNA-binding proteins in cancer, Comp. Funct. Genom. (2012), 2012178525.
[11] D.J. Hodson, M. Screen, M. Turner, RNA-binding proteins in hematopoiesis and hematological malignancy, Blood 133 (22) (2019) 2365–2373.
[12] P. Chopra, J. Lee, J. Kang, S. Lee, Improving cancer classification accuracy using gene pairs, PLoS One 5 (12) (2010), e14305.
[13] B. Li, Y. Cui, M. Diehn, R. Li, Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer, JAMA Oncol. 3 (11) (2017) 1529–1537.
[14] X.Y. Sun, S.Z. Yu, H.P. Zhang, J. Li, W.Z. Guo, S.J. Zhang, A signature of 33 immune-related gene pairs predicts clinical outcome in hepatocellular carcinoma, Cancer Med. 9 (8) (2020) 2868–2878.
[15] S. Salunkhe, N. Chandran, P. Chandrani, A. Dutt, S. Dutt, CytoPred: 7-gene pair metric for AML cytogenetic risk prediction, Briefings Bioinf. 21 (1) (2020) 348–354.
[16] S. Gerstberger, M. Hafner, T. Tuschl, A census of human RNA-binding proteins, Nat. Rev. Genet. 15 (12) (2014) 829–845.
[17] K.W. Brannan, W. Jin, S.C. Huelga, C.A. Banks, J.M. Gilmore, L. Florens, M.P. Washburn, E.L. Van Nostrand, G.A. Pratt, M.K. Schwinn, D.L. Daniels, G.W. Yeo, SONAR discovers RNA-binding proteins from analysis of large-scale protein-protein interactomes, Mol. Cell. 64 (2) (2016) 282–293.
[18] R. Huang, M. Han, L. Meng, X. Chen, Transcriptome-wide discovery of coding and noncoding RNA-binding proteins, Proc. Natl. Acad. Sci. U. S. A. 115 (17) (2018) E3879–e3887.
[19] J. Trendel, T. Schwarzl, R. Horos, A. Prakash, A. Bateman, M.W. Hentze, J. Krijgsveld, The human RNA-binding proteome and its dynamics during translational arrest, Cell 176 (1–2) (2019) 391–403.e319.
[20] K. Li, Z.W. Guo, X.M. Zhai, X.X. Yang, Y.S. Wu, T.C. Liu, RBPTD: a Database of Cancer-Related RNA-Binding Proteins in Humans, Database, Oxford), 2020, p. 2020.
[21] G. Spolverato, A. Ejaz, Y. Kim, M.H. Squires, G.A. Poultsides, R.C. Fields, C. Schmidt, S.M. Weber, K. Votanopoulos, S.K. Maithel, T.M. Pawlik, Rates and patterns of recurrence after curative intent resection for gastric cancer: a United States multi-institutional analysis, J. Am. Coll. Surg. 219 (4) (2014) 664–675.

[22] Z.J. Ward, A.M. Scott, H. Hricak, M. Abdel-Wahab, D. Paez, M.M. Lette, H.A. Vargas, T.P. Kingham, R. Atun, Estimating the impact of treatment and imaging modalities on 5-year net survival of 11 cancers in 200 countries: a simulation-based analysis, Lancet Oncol. 21 (8) (2020) 1077–1088.
[23] A. Digklia, A.D. Wagner, Advanced gastric cancer: current treatment landscape and future perspectives, World J. Gastroenterol. 22 (8) (2016) 2403–2414.
[24] H. Ren, J. Zhu, H. Yu, A.V. Bazhin, C.B. Westphalen, B.W. Renz, S.N. Jacob, C. Lampert, J. Werner, M.K. Angele, F. Bösch, Angiogenesis-related gene expression signatures predicting prognosis in gastric cancer patients, Cancers 12 (12) (2020).
[25] T. Luo, Y. Li, R. Nie, C. Liang, Z. Liu, Z. Xue, G. Chen, K. Jiang, Z.X. Liu, H. Lin, C. Li, Y. Chen, Development and validation of metabolism-related gene signature in prognostic prediction of gastric cancer, Comput. Struct. Biotechnol. J. (2020) 183217–183229.
[26] R. Mao, Z. Wang, Y. Zhang, Y. Chen, Q. Liu, T. Zhang, Y. Liu, Development and validation of a novel prognostic signature in gastric adenocarcinoma, Aging (Albany NY) 12 (21) (2020) 22233–22252.
[27] T.A. Cooper, L. Wan, G. Dreyfuss, RNA and disease, Cell 136 (4) (2009) 777–793.
[28] L. Zhou, Y. Wu, L. Xin, Q. Zhou, S. Li, Y. Yuan, J. Wang, D. Wu, Development of RNA binding proteins expression signature for prognosis prediction in gastric cancer patients, Am. J. Transl. Res. 12 (10) (2020) 6775–6792.
[29] K.E. Lukong, K.W. Chang, E.W. Khandjian, S. Richard, RNA-binding proteins in human genetic disease, Trends Genet. 24 (8) (2008) 416–425.
[30] L. Kachuri, R.E. Graff, K. Smith-Byrne, T.J. Meyers, S.R. Rashkin, E. Ziv, J.S. Witte, M. Johansson, Pan-cancer analysis demonstrates that integrating polygenic risk scores with modifiable risk factors improves risk prediction, Nat. Commun. 11 (1) (2020) 6084.
[31] A.C. Fahed, M. Wang, J.R. Homburger, A.P. Patel, A.G. Bick, C.L. Neben, C. Lai, D. Brockman, A. Philippakis, P.T. Ellinor, C.A. Cassa, M. Lebo, K. Ng, et al., Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions, Nat. Commun. 11 (1) (2020) 3635.
[32] Z. Zhao, X. Fan, L. Yang, J. Song, S. Fang, J. Tu, M. Chen, J. Li, L. Zheng, F. Wu, D. Zhang, X. Ying, J. Ji, Recognition of lung adenocarcinoma-specific gene pairs based on genetic algorithm and establishment of a deep learning prediction model, Comb. Chem. High Throughput Screen. 22 (4) (2019) 256–265.
[33] X.T. Zeng, X.P. Liu, T.Z. Liu, X.H. Wang, The clinical significance of COL5A2 in patients with bladder cancer: a retrospective analysis of bladder cancer gene expression data, Medicine (Baltim.) 97 (10) (2018), e0091.
[34] S. Wei, J. Lu, J. Lou, C. Shi, S. Mo, Y. Shao, J. Ni, W. Zhang, X. Cheng, Gastric cancer tumor microenvironment characterization reveals stromal-related gene signatures associated with macrophage infiltration, Front. Genet. (2020), 11663.
[35] E. Guo, Y. Ishii, J. Mueller, A. Srivatsan, T. Gahman, C.D. Putnam, J.Y.J. Wang, R.D. Kolodner, FEN1 endonuclease as a therapeutic target for human cancers with defects in homologous recombination, Proc. Natl. Acad. Sci. U. S. A. 117 (32) (2020) 19415–19424.
[36] M.A. Sparks, P.M. Burgers, R. Galletto, Pif1, RPA, and FEN1 modulate the ability of DNA polymerase δ to overcome protein barriers during DNA synthesis, J. Biol. Chem. 295 (47) (2020) 15883–15891.
[37] X. He, J. Wang, H. Yu, W. Lv, Y. Wang, Q. Zhang, Z. Liu, Y. Wu, Clinical significance for diagnosis and prognosis of POP1 and its potential role in breast cancer: a comprehensive analysis based on multiple databases, Aging (Albany NY) 14 (17) (2022) 6936–6956.
[38] Y. Chen, S. Zhang, Q. Wang, X. Zhang, Tumor-recruited M2 macrophages promote gastric and breast cancer metastasis via M2 macrophage-secreted CHI3L1 protein, J. Hematol. Oncol. 10 (1) (2017) 36.
[39] A. Tiwari, K. Tashiro, A. Dixit, A. Soni, K. Vogel, B. Hall, I. Shafqat, J. Slaughter, N. Param, A. Le, E. Saunders, U. Paithane, G. Garcia, et al., Loss of HIF1A from pancreatic cancer cells increases expression of PPP1R1B and degradation of p53 to promote invasion and metastasis, Gastroenterology 159 (5) (2020) 1882–1897.e1885.
[40] W.R. Ma, P. Xu, Z.J. Liu, J. Zhou, L.K. Gu, J. Zhang, D.J. Deng, Impact of GFRA1 gene reactivation by DNA demethylation on prognosis of patients with metastatic colon cancer, World J. Gastroenterol. 26 (2) (2020) 184–198.
[41] Z. Ma, Z. Li, S. Wang, Q. Zhou, Z. Ma, C. Liu, B. Huang, Z. Zheng, L. Yang, Y. Zou, C. Zhang, S. Huang, B. Hou, SLC39A10 upregulation predicts poor prognosis, promotes proliferation and migration, and correlates with immune infiltration in hepatocellular carcinoma, J. Hepatocell. Carcinoma (2021) 8899–8912.
[42] L. Lu, G.Q. Ma, X.D. Liu, R.R. Sun, Q. Wang, M. Liu, P.Y. Zhang, Correlation between GDF15, MMP7 and gastric cancer and its prognosis, Eur. Rev. Med. Pharmacol. Sci. 21 (3) (2017) 535–541.
[43] Y.R. Kim, N.J. Yoo, S.H. Lee, Somatic mutation of EXO1 gene in gastric and colorectal cancers with microsatellite instability, Acta Oncol. 49 (6) (2010) 859–860.
[44] J. Chen, A. Wang, J. Ji, K. Zhou, Z. Bu, G. Lyu, J. Ji, An innovative prognostic model based on four genes in asian patient with gastric cancer, Cancer Res. Treat. 53 (1) (2021) 148–161.
[45] X. Liu, J. Wu, D. Zhang, Z. Bing, J. Tian, M. Ni, X. Zhang, Z. Meng, S. Liu, Identification of potential key genes associated with the pathogenesis and prognosis of gastric cancer based on integrated bioinformatics analysis, Front. Genet. (2018) 9265.
[46] C. Neophytou, C. Pitsouli, Biotin controls intestinal stem cell mitosis and host-microbiome interactions, Cell Rep. 38 (10) (2022), 110505.
[47] H. Wang, W. Xiao, Q. Zhou, Y. Chen, S. Yang, J. Sheng, Y. Yin, J. Fan, J. Zhou, Bystin-like protein is upregulated in hepatocellular carcinoma and required for nucleologenesis in cancer cell proliferation, Cell Res. 19 (10) (2009) 1150–1164.
[48] K. Adachi, C. Soeta-Saneyoshi, H. Sagara, Y. Iwakura, Crucial role of Bysl in mammalian preimplantation development as an integral factor for 40S ribosome biogenesis, Mol. Cell Biol. 27 (6) (2007) 2202–2214.
[49] S. Zhang, Z. Shi, K. Pang, H. He, J. Chen, Z. Zhang, Q. Zhang, L. Hao, C. Han, MAP30 inhibits bladder cancer cell migration and invasion in vitro through suppressing akt pathway and the epithelial/mesenchymal transition process, DNA Cell Biol. 39 (11) (2020) 1948–1960.
[50] H.H. Zhang, R. Li, Y.J. Li, X.X. Yu, Q.N. Sun, A.Y. Li, Y. Kong, eIF4E-related miR-320a and miR-340-5p inhibit endometrial carcinoma cell metastatic capability by preventing TGF-β1-induced epithelial-mesenchymal transition, Oncol. Rep. 43 (2) (2020) 447–460.
[51] M.E. O'Brien, J. Londino, M. McGinnis, N. Weathington, J. Adair, T. Suber, V. Kagan, K. Chen, C. Zou, B. Chen, J. Bon, R.K. Mallampalli, Tumor necrosis factor alpha regulates skeletal myogenesis by inhibiting SP1 interaction with cis-acting regulatory elements within the Fbxl2 gene promoter, Mol. Cell Biol. 40 (12) (2020).
[52] K.A. Hogan, D.S. Cho, P.C. Arneson, A. Samani, P. Palines, Y. Yang, J.D. Doles, Tumor-derived cytokines impair myogenesis and alter the skeletal muscle immune microenvironment, Cytokine (2018) 1079.
[53] T. Machackova, P. Vychytilova-Faltejskova, K. Souckova, K. Trachtova, D. Brchnelova, M. Svoboda, I. Kiss, V. Prochazka, Z. Kala, O. Slaby, MiR-215-5p reduces liver metastasis in an experimental model of colorectal cancer through regulation of ECM-receptor interactions and focal adhesion, Cancers 12 (12) (2020).
[54] X. Yang, L. Chen, Y. Mao, Z. Hu, M. He, Progressive and prognostic performance of an extracellular matrix-receptor interaction signature in gastric cancer, Dis. Markers (2020), 20208816070.
[55] H.H. Gari, G.D. DeGala, R. Ray, M.S. Lucia, J.R. Lambert, PRL-3 engages the focal adhesion pathway in triple-negative breast cancer cells to alter actin structure and substrate adhesion properties critical for cell migration and invasion, Cancer Lett. 380 (2) (2016) 505–512.