



Research article

Identification of a novel snoRNA expression signature associated with overall survival in patients with lung adenocarcinoma: A comprehensive analysis based on RNA sequencing dataset

Linbo Zhang, Mei Xin and Peng Wang*

Department of Health Management and Division of Physical Examination, The First Affiliated Hospital of Guangxi Medical University, Shuang Yong Road 6, Nanning 530021, China

* **Correspondence:** Email: wangpeng@gxmu.edu.cn; Tel: +867715314316; Fax: +867715314316.

Abstract: Since multiple studies have reported that small nucleolar RNAs (snoRNAs) can be serve as prognostic biomarkers for cancers, however, the prognostic values of snoRNAs in lung adenocarcinoma (LUAD) remain unclear. Therefore, the main work of this study is to identify the prognostic snoRNAs of LUAD and conduct a comprehensive analysis. The Cancer Genome Atlas LUAD cohort whole-genome RNA-sequencing dataset is included in this study, prognostic analysis and multiple bioinformatics approaches are used for comprehensive analysis and identification of prognostic snoRNAs. There were seven LUAD prognostic snoRNAs were screened in current study. We also constructed a novel expression signature containing five LUAD prognostic snoRNAs (snoU109, SNORA5A, SNORA70, SNORD104 and U3). Survival analysis of this expression signature reveals that LUAD patients with high risk score was significantly related to an unfavourable overall survival (adjusted $P = 0.01$, adjusted hazard ratio = 1.476, 95% confidence interval = 1.096–1.987). Functional analysis indicated that LUAD patients with different risk score phenotypes had significant differences in cell cycle, apoptosis, integrin, transforming growth factor beta, ErbB, nuclear factor kappa B, mitogen-activated protein kinase, phosphatidylinositol-3-kinase and toll like receptor signaling pathway. Immune microenvironment analysis also indicated that there were significant differences in immune microenvironment scores among LUAD patients with different risk score. In conclusion, this study identified an novel expression signature containing five LUAD prognostic snoRNAs, which may be serve as an independent prognostic indicator for LUAD patients.

Keywords: RNA sequencing; snoRNA; lung adenocarcinoma; overall survival; the cancer genome atlas

1. Introduction

Lung adenocarcinoma (LUAD) is the most common subtype of lung cancer worldwide and is a non-small cell lung cancer. With the rapid development of medical technology and drugs, the diagnosis and treatment of LUAD has been greatly improved, but the survival rate of LUAD is still not satisfactory. Small nucleolar RNA (snoRNA) is a class of small non-coding RNAs widely found in the nucleoli of eukaryotic cells. Among them, Box C/D and Box H/ACA are the main known snoRNA types, which mainly guide the dioxymethylation and pseuduracil modification of ribosomal RNA by base pairings, respectively [1]. In addition, some studies have reported that snoRNA is involved in the post-transcriptional modification of snRNA, tRNA and mRNA [1]. There are also a considerable number of snoRNAs whose functions are unknown and are called orphan snoRNAs. In recent years, With the advancement of high-throughput sequencing, numerous dataset suggest that snoRNA is dysregulation and plays a role in multiple cancers [2,3]. Previous studies have reported that snoRNA plays a role in tumorigenesis and progression of cancers, and may become a new prognostic biomarker for cancers [4,5]. Since multiple studies have reported that snoRNAs can be serve as prognostic biomarkers for cancers, however, the prognostic values of snoRNAs in LUAD remain unclear [6,7]. In current study, the prognostic value of snoRNAs were explored based on the RNA sequencing dataset of The Cancer Genome Atlas (TCGA) LUAD cohort, and the bioinformatics approaches were further used for comprehensive analysis.

2. Materials and methods

2.1. Acquisition of RNA sequencing dataset

The whole genome RNA sequencing dataset of lung adenocarcinoma were downloaded from TCGA website (<https://portal.gdc.cancer.gov>) [8], and the raw dataset were normalized by edgeR. SnoRNAs with an average value greater than 0.5 was included in the follow-up analysis, and the rest were excluded. Inclusion criteria for patients included in the prognostic analysis: 1) LUAD patients have both survival information and RNA sequencing dataset; 2) RNA sequencing sample is the primary tumor tissue. Exclusion criteria: 1) Recurrent tumor tissue or duplicate samples of the same patient; 2) The patient lacks survival parameters or RNA sequencing dataset. We finally won the 500 cases LUAD patients included in the subsequent survival analysis [9]. Since all the dataset for the present study comes from TCGA database, the authors did not have any experiments involving humans or animals in this study. Therefore, the approval of the ethics committee is not required.

2.2. Identification of prognostic snoRNAs and signature construction

The identification of prognostic snoRNAs are calculated by the *survival* package in the R platform through the multivariate Cox proportional hazard regression model. The high- and low-expression groups of snoRNAs or genes were grouped according to the median value of expression. The clinical adjustment variables in the multivariate Cox model are these variables related to LUAD overall survival (OS). We use the step function of the survival package to screen the optimal expression

signature for the prognostic-related snoRNAs of LUAD in the R platform. The calculation formula of snoRNA expression signature is as follows: Risk score = expression of snoRNA1 \times β 1 + expression of snoRNA2 \times β 2 + ... expression of snoRNA n \times β n [9]. The *surcivalROC* package is used to evaluate the prediction accuracy of this snoRNA expression signature. The nomogram was executed by the *rms* package. Joint effect survival analysis was used to evaluate the combination of risk score and traditional clinical parameters to classify LUAD patients with different prognosis.

2.3. Functional enrichment analysis

In order to further understand the potential functional mechanism of this snoRNA expression signature. We conduct a functional enrichment analysis through three ways of snoRNA co-expression genes, differentially expressed genes (DEGs) and gene set enrichment analysis (GSEA). The acquisition of whole-genome expression profile dataset was derived from the same dataset as snoRNA, and the processing method of the original data was also the same. Co-expression genes were defined as the absolute value if Pearson correlation coefficient (r) between gene and snoRNA expression in LUAD tumor tissues greater than 0.2. DEGs were also identified by edgeR, and DEGs were defined as $\log_2|\text{fold change (FC)}| > 1$, $P < 0.05$ and false discovery rate (FDR) > 0.05 . The GSEA analysis was performed by the GSEA desktop application and the reference gene sets were C2 (c2.all.v7.4.symbols.gmt) and C5 (c5.all.v7.4.symbols.gmt). The criteria for GSEA significance results are as follows: $|\text{Normalized Enrichment Score (NES)}| > 1$, $\text{FDR} < 0.5$ and $P < 0.05$ [10,11]. Functional enrichment analysis of DEGs and co-expression genes were performed by Database for Annotation, Visualization and Integrated Discovery v6.8 (DAVID v6.8, <https://david.ncifcrf.gov/home.jsp>), and gene ontology (GO) term and kyoto encyclopedia of genes and genome (KEGG) pathways were annotated for these genes [12,13]. Subsequently, we also investigated the tumor immune microenvironment of LUAD. The ESTIMATE package is used to calculate the parameters of the tumor immune microenvironment [14].

2.4. Statistical analysis

The co-expression interaction was evaluated using Pearson's correlation coefficient r . The calculation of FDR is performed according to the Benjamini-Hochberg program [15]. All statistical analysis adopts SPSS version 22.0 and R version 3.6.2. $P < 0.05$ considered significant difference.

3. Results

3.1. Identification of prognostic snoRNAs

The demographics of TCGA LUAD cohort are shown in Table S1. In the patients' baseline dataset, we found that tumor stage is closely related to OS in TCGA LUAD cohort. Therefore, we included tumor stage as a correction factor in the multivariate Cox proportional hazards regression model. We obtained 940 snoRNAs from the raw RNA sequencing dataset of TCGA LUAD cohort, and finally obtained 288 snoRNAs that met the requirements for subsequent survival analysis. Seven snoRNAs that are significantly related to LUAD OS were screened out through multivariate Cox proportional hazards regression model analysis in the R platform (Table 1 and Figure 1). The seven prognostic snoRNAs are snoU109 (ENSG00000238832), U8 (ENSG00000239148), SNORA70 (ENSG00000206886), U3 (ENSG00000207119), SNORA5A (ENSG00000206838), SNORD7

(ENSG00000207297), SNORD104 (ENSG00000199753). The Kaplan-Meier survival curve of prognostic snoRNAs are shown in Figure 2A–G.

Table 1. Multivariate analysis results of snoRNAs related to the prognosis of LUAD patients.

ID	Adjusted P_{λ}	HR	Low 95% CI	High 95% CI
snoU109 ENSG00000238832	0.009846	1.482519	1.099398	1.999151
U8 ENSG00000239148	0.010419	0.677437	0.502876	0.91259
SNORA70 ENSG00000206886	0.017174	0.697903	0.51918	0.93815
U3 ENSG00000207119	0.030339	0.720103	0.534974	0.969297
SNORA5A ENSG00000206838	0.036391	1.369859	1.020138	1.83947
SNORD7 ENSG00000207297	0.037603	0.73046	0.543267	0.982155
SNORD104 ENSG00000199753	0.044011	0.736418	0.546781	0.991827

Notes: λ adjusted for tumor stage. Abbreviation: HR, hazard ratio; CI, confidence interval.

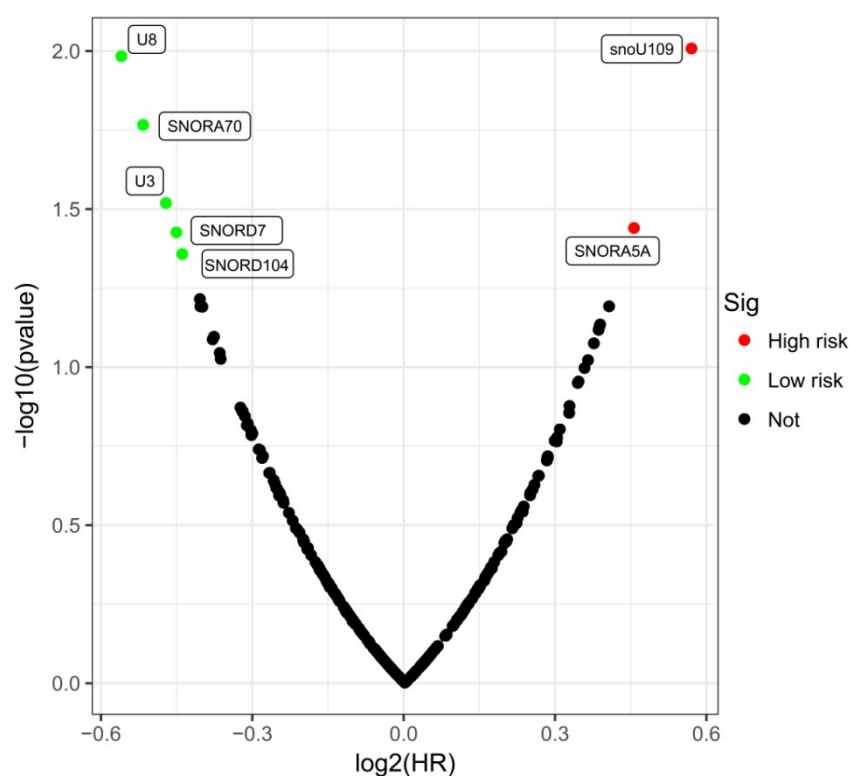


Figure 1. Volcano plot of prognostic-related snoRNAs in the TCGA LUAD cohort.

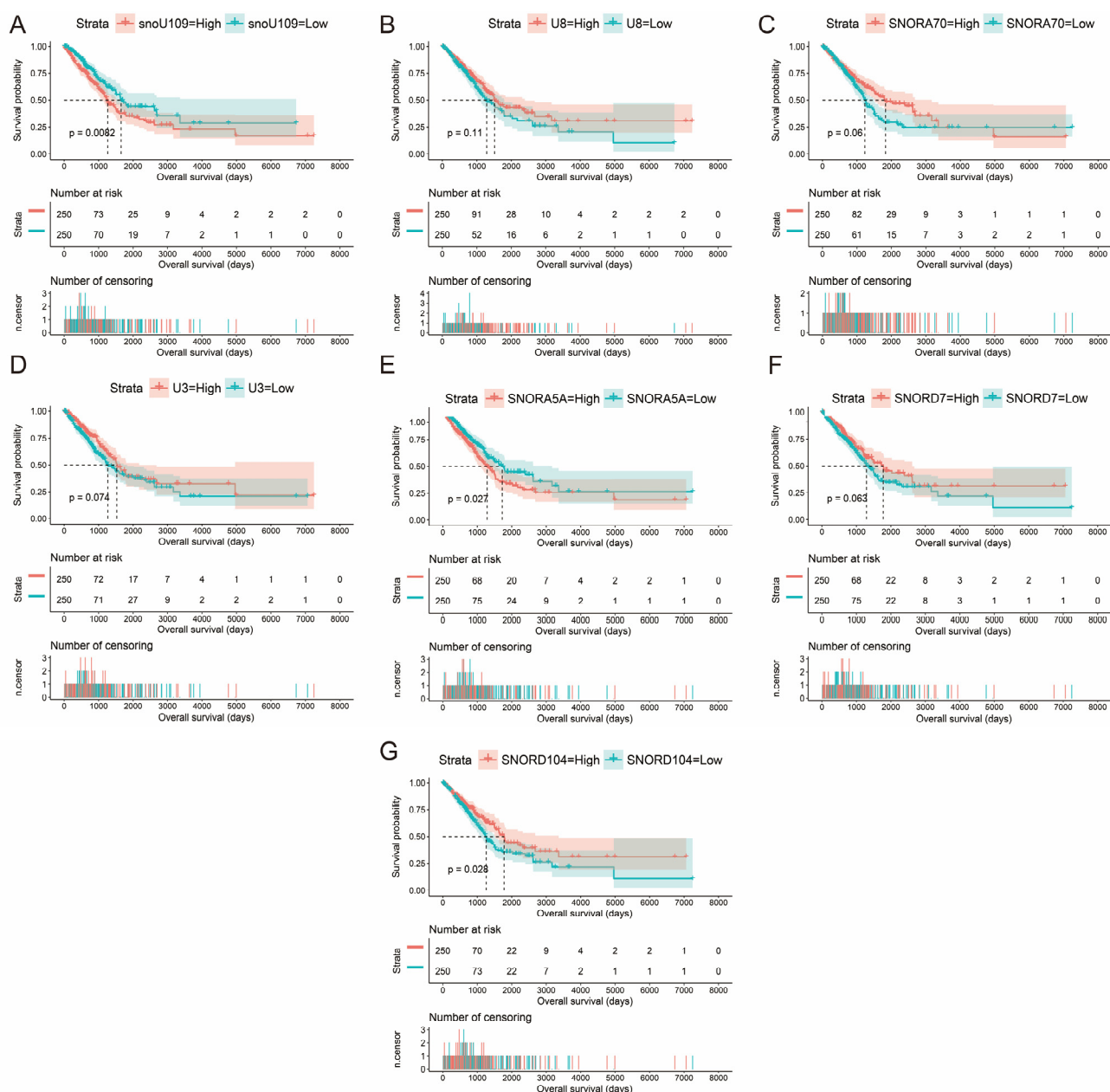


Figure 2. Kaplan Meier curve of prognostic snoRNAs. (A) Kaplan Meier curve of snoU109; (B) Kaplan Meier curve of U8; (C) Kaplan Meier curve of SNORA70; (D) Kaplan Meier curve of U3; (E) Kaplan Meier curve of SNORA5A; (F) Kaplan Meier curve of SNORD7; (G) Kaplan Meier curve of SNORD104.

3.2. Comprehensive survival analysis of snoRNA expression signature

Then we get the optimal combination signature through the step function in R platform. The formula of expression signature is as follows: Risk score = expression of snoU109 \times (0.1293) + expression of SNORA5A \times (0.1046) + expression of SNORA70 \times (-0.2012) + expression of SNORD104 \times (-0.1005) + expression of U3 \times (-0.1155). Through survival analysis, we found that in this snoRNA expression signature, the OS of LUAD patients with high risk was significantly reduced

(adjusted $P = 0.01$, adjusted hazard ratio = 1.476, 95% confidence interval = 1.096–1.987, Figure 3A,B). Subsequent time-dependent receiver operator characteristic curve (ROC) analysis found that this snoRNA expression signature has a certain application value in predicting LUAD OS. The area under curve (AUC) is 0.618 in 5-year OS, and the highest is 0.666 in 9-year OS (Figure 3C). Joint effect survival analysis suggests that the combination of risk score model and tumor stage can more accurately classify LUAD patients into subgroups with significant differences in prognosis (Table 2 and Figure 4A,B). The nomogram suggested that the contribution of risk score in the TCGA LUAD cohort was second only to the tumor stage (Figure 4C).

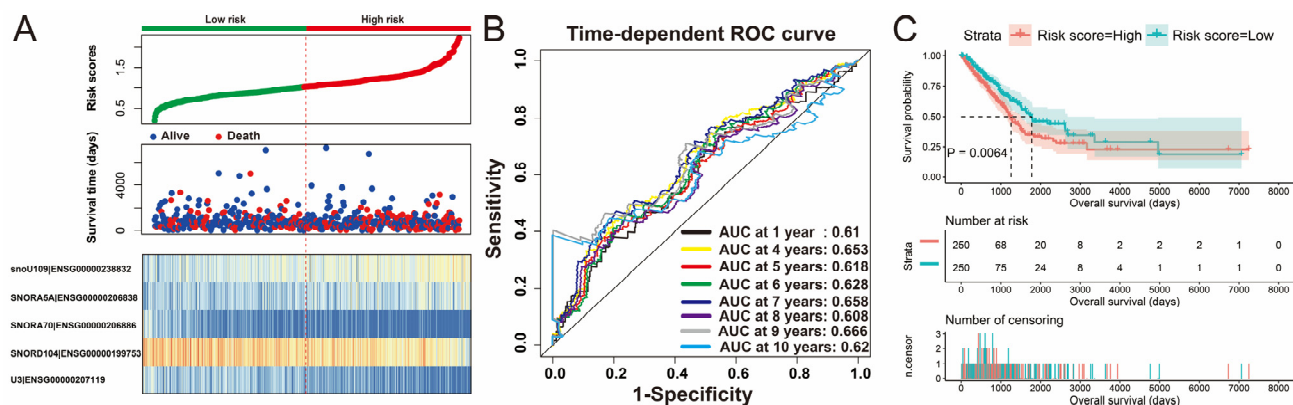


Figure 3. SnoRNA expression signature for LUAD OS. (A) Scatter plot of patients' survival time distribution and risk score; (B) Time dependent ROC curve of risk score; (C) Kaplan Meier curve of risk score.

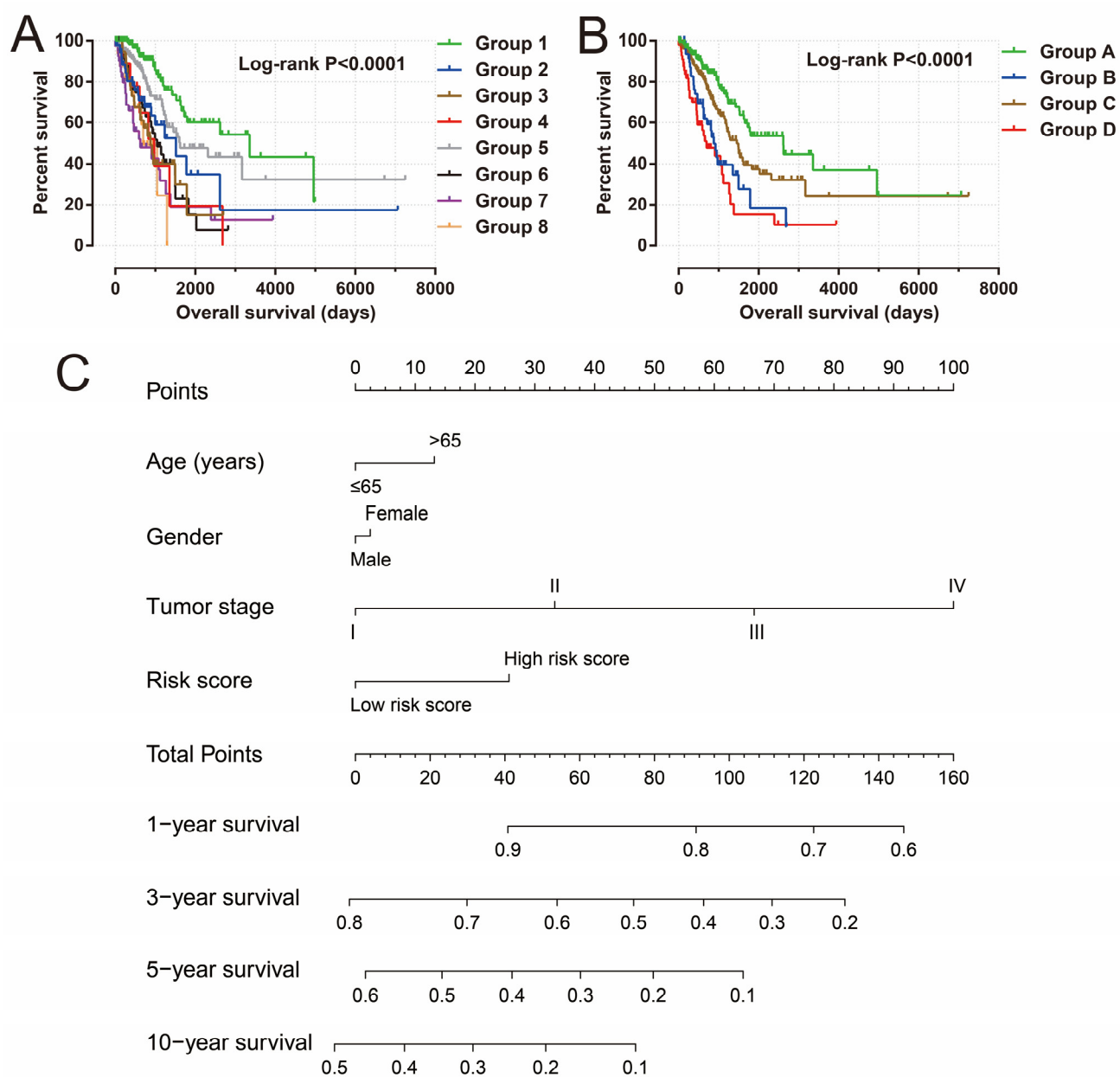


Figure 4. Joint effect survival analysis curve and nomogram of risk score and tumor stage. (A-B) Kaplan Meier curve of Joint effect analysis of risk score and tumor stage; (C) Nomogram of risk score and clinical parameters.

Table 2. Joint effect survival analysis of the prognostic signature and tumor stage.

Group	Risk score	Tumor stage ^b	Patients (n = 488)	MST (days)	Crude HR (95% CI)	Crude <i>P</i>	Adjusted HR (95% CI)	Adjusted <i>P</i> λ
1	Low risk	I	142	3361	1		1	
2	Low risk	II	55	1516	2.834 (1.592-5.044)	0.0004	2.834 (1.592-5.044)	0.0004
3	Low risk	III	10	879	4.043 (2.267-7.212)	< 0.0001	4.043 (2.267-7.212)	< 0.0001
4	Low risk	IV	9	976	4.458 (1.943-10.351)	0.0004	4.458 (1.943-10.351)	0.0004
5	High risk	I	126	1622	1.749 (1.065-2.874)	0.027	1.749 (1.065-2.874)	0.027
6	High risk	II	64	1046	3.761 (2.239-6.316)	< 0.0001	3.761 (2.239-6.316)	< 0.0001
7	High risk	III	40	624	5.518 (3.162-9.628)	< 0.0001	5.518 (3.162-9.628)	< 0.0001
8	High risk	IV	16	697	5.839 (2.712-12.571)	< 0.0001	5.839 (2.712-12.571)	< 0.0001
A	Low risk	I + II	197	2617	1		1	
B	Low risk	III + IV	49	879	2.913 (1.820-4.662)	< 0.0001	4.099 (2.120-7.922)	< 0.0001
C	High risk	I + II	190	1454	1.639 (1.134-2.369)	0.009	1.548 (1.070-2.240)	0.020
D	High risk	III + IV	56	697	3.940 (2.517-6.167)	< 0.0001	5.524 (2.941-10.379)	< 0.0001

Notes: ^b Information of tumor stage was unavailable in 8 patients; λ adjusted for tumor stage. Abbreviation: MST, median survival time; HR, hazard ratio; CI, confidence interval; NA, not available.

3.3. Functional enrichment analysis

We also screened co-expression genes for the five snoRNAs in the risk score signature by Pearson correlation coefficient. A total of 917 SNORA5A co-expressed genes were screened (Figure 5 and Table S3). 158 co-expressed genes for SNORA70 (Figure 6 and Table S3). 1048 co-expressed genes for SNORD104 (Figure 7 and Table S3). 987 co-expressed genes for snoU109 (Figure 8 and Table S3). 303 co-expressed genes for U3 (Figure 9 and Table S3). Through merging, we found that a total of 3031 snoRNA co-expressed genes were obtained. Functional enrichment analysis revealed that these genes were significantly involved in the following biological functions: DNA replication-dependent nucleosome assembly, ubiquitin-protein transferase activity, cell cell adhesion, DNA repair, regulation of cell cycle, integrin binding, I-kappaB kinase/NF-kappaB signaling, negative regulation of TOR signaling, hippo signaling, and regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle (Table S4). KEGG analysis revealed that these snoRNA co-expressed genes may be involved in the following signaling pathways: focal adhesion, ubiquitin mediated proteolysis, proteoglycans in cancer, transcriptional misregulation in cancer, and Ras signaling pathway (Table S4). A total of 319 LUAD prognostic related genes were identified, including 169 low risk genes [hazard rate (HR) < 1] and 150 high risk genes (HR > 1) (Figure 10A and Table S5). The top three significant genes are transmembrane protein 125 (TMEM125, Figure 10B), phosphatidylglycerophosphate synthase 1 (PGS1, Figure 10C), and myozenin 1 (MYOZ1, Figure 10D).

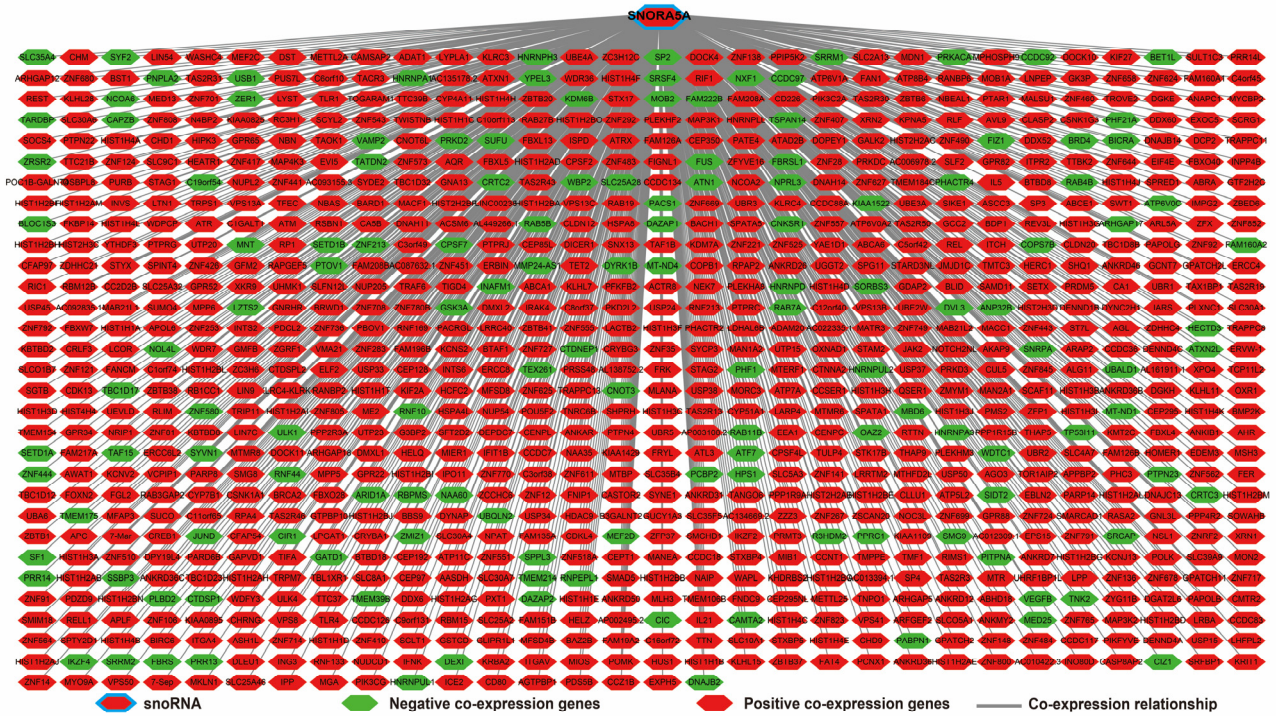


Figure 5. Gene-snoRNA co-expression interaction network of SNORA5A in LUAD tumor tissues.

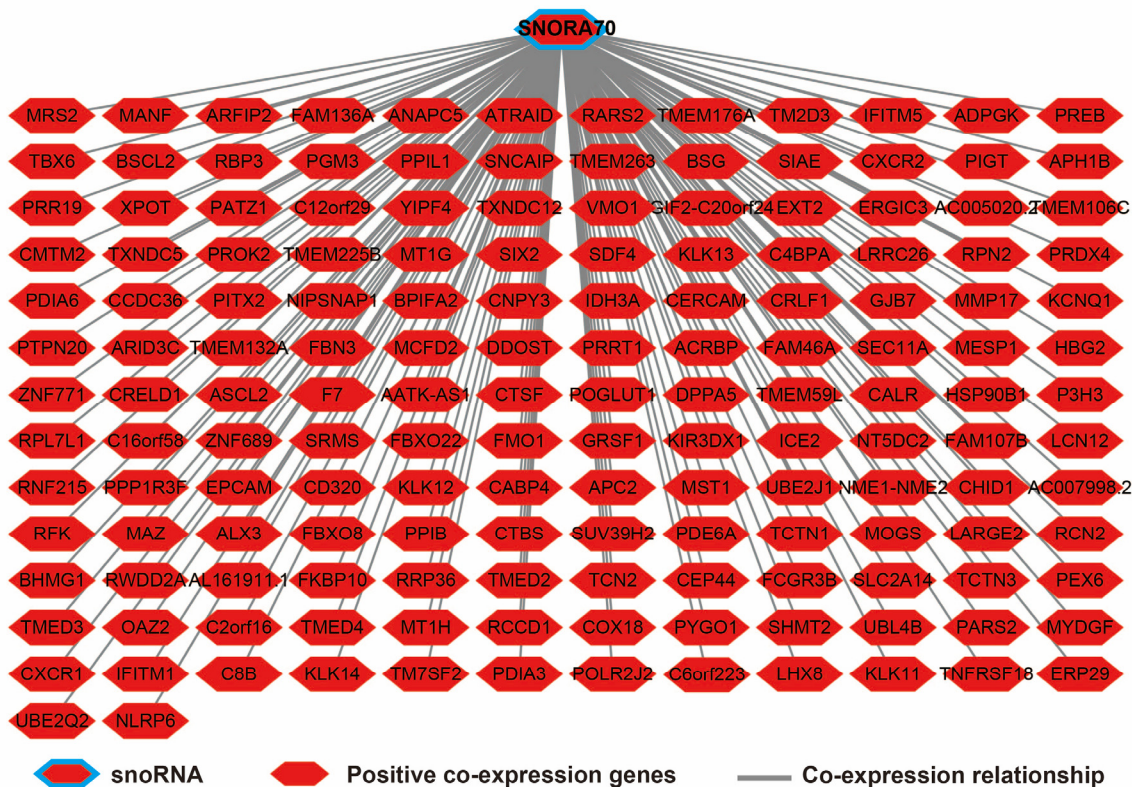


Figure 6. Gene-snoRNA co-expression interaction network of SNORA70 in LUAD tumor tissues.

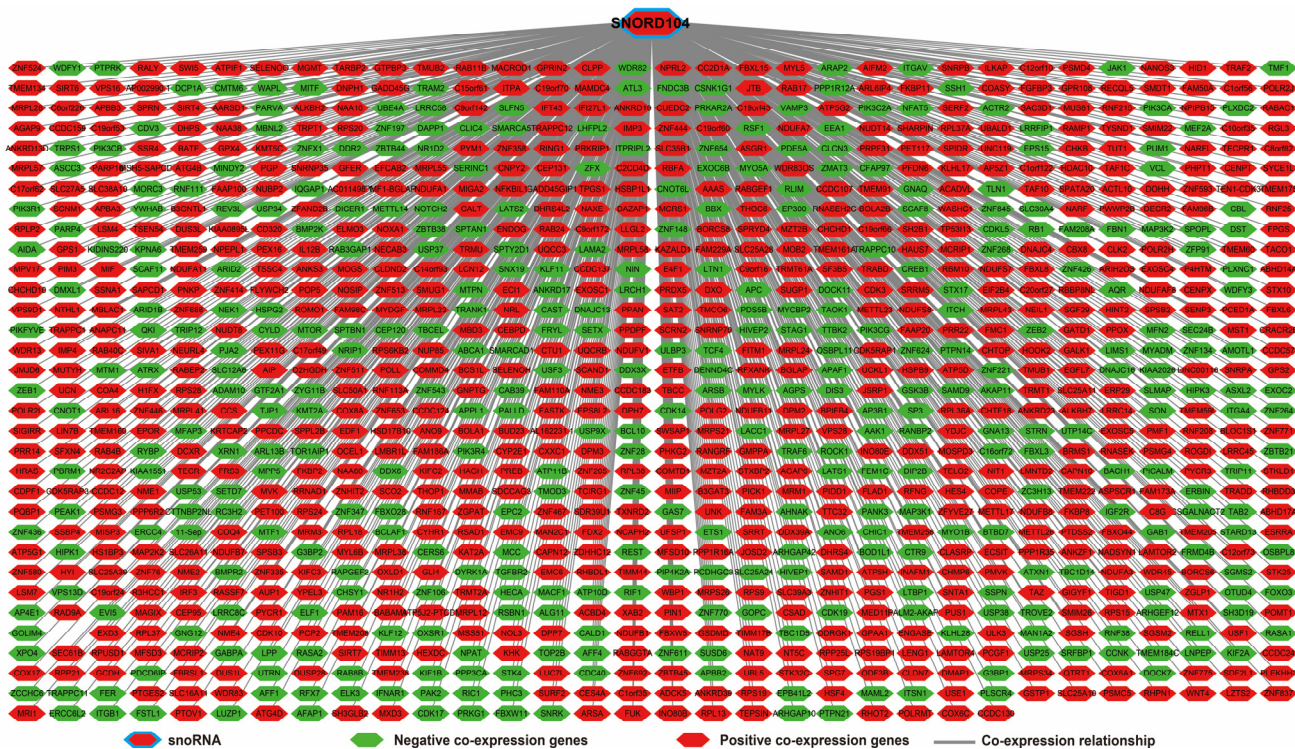


Figure 7. Gene-snoRNA co-expression interaction network of SNORD104 in LUAD tumor tissues.

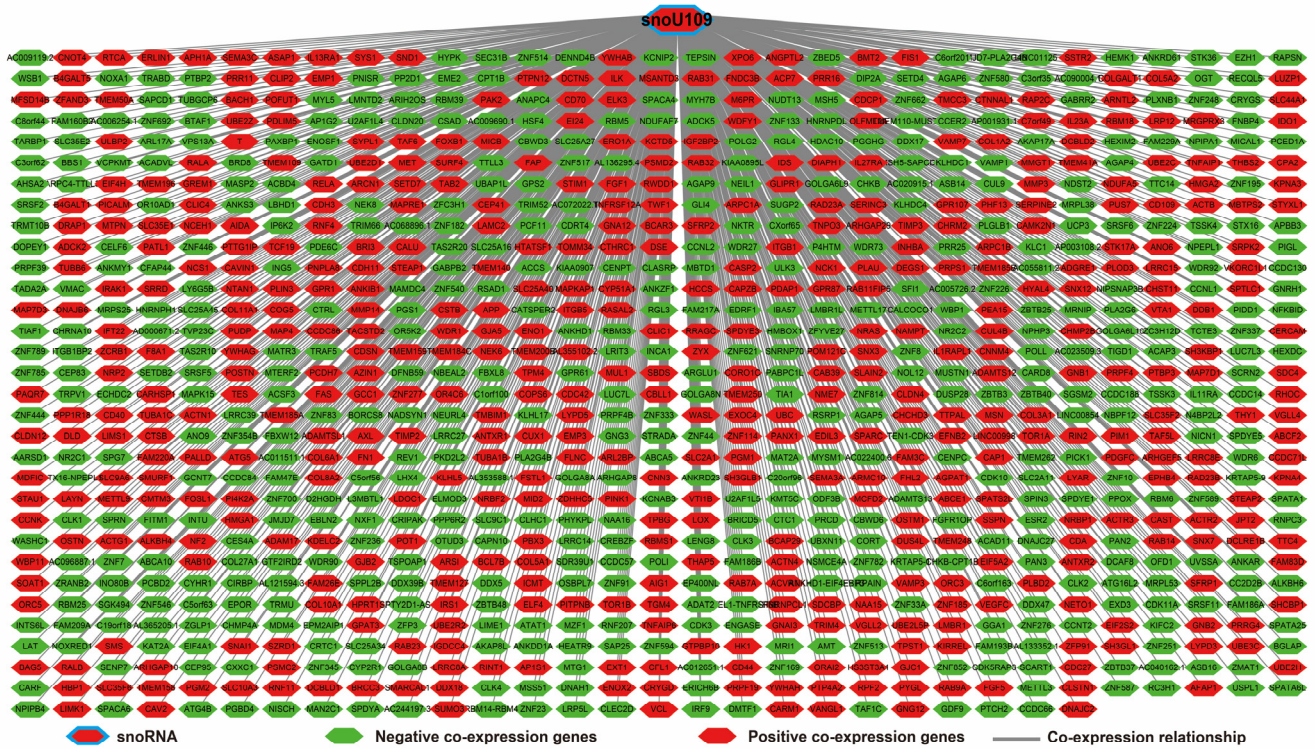


Figure 8. Gene-snoRNA co-expression interaction network of snoU109 in LUAD tumor tissues.

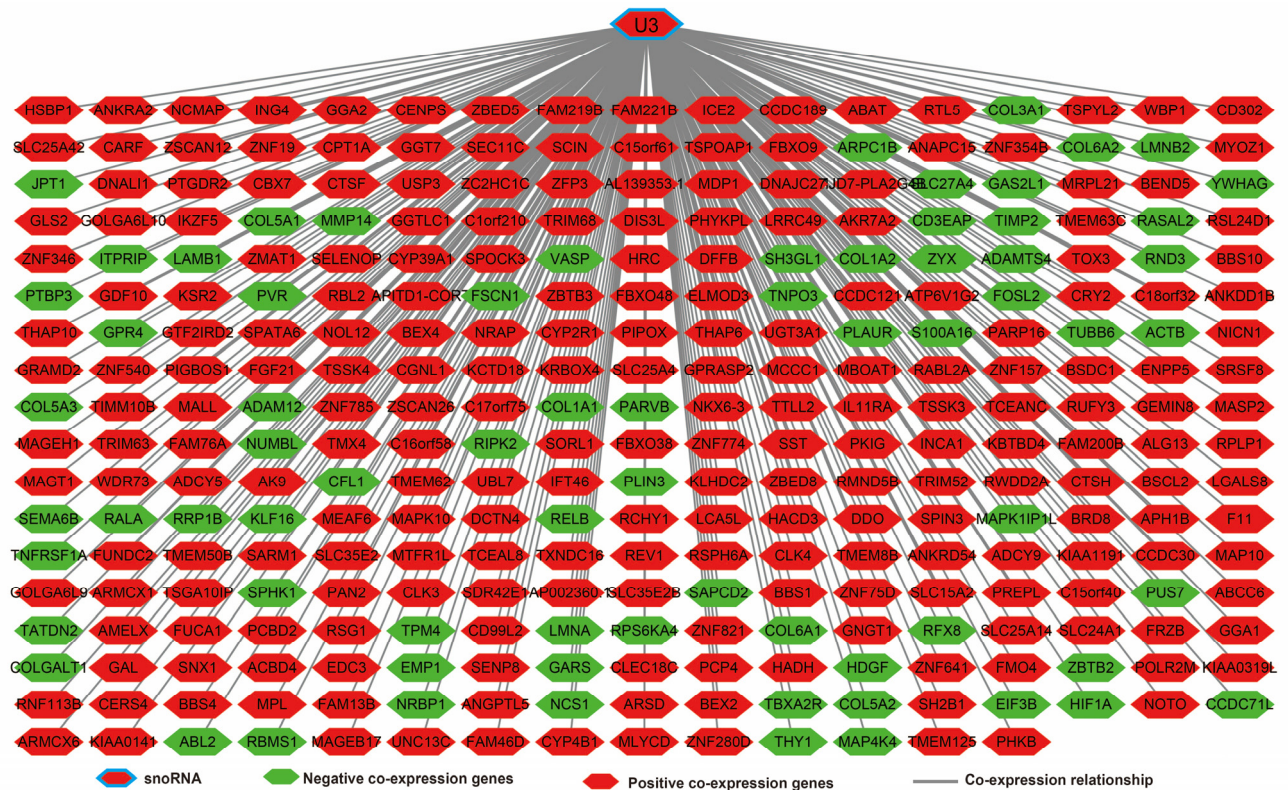


Figure 9. Gene-snoRNA co-expression interaction network of U3 in LUAD tumor tissues.

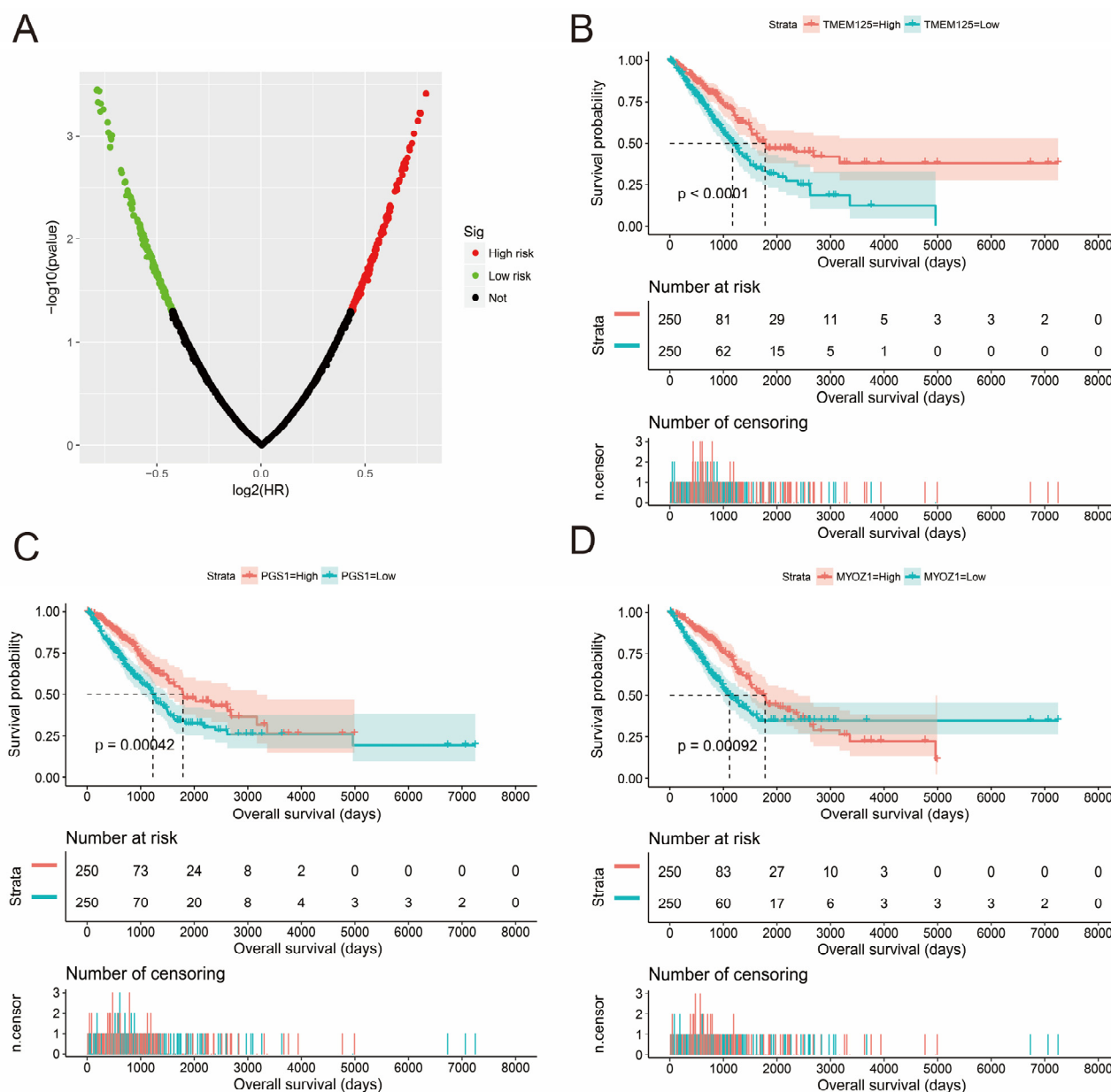


Figure 10. Survival analysis results of co-expression genes of snoRNA expression signature. (A) Volcano plot of prognostic analysis results of co-expressed genes; (B) Kaplan Meier curve of *TMEM125*; (C) Kaplan Meier curve of *PGS1*; (D) Kaplan Meier curve of *MYOZ1*.

We further screened DEGs for LUAD patients between low- and high-risk phenotypes. A total of 772 DEGs were obtained, of which 394 DEGs were significantly down-regulated and 378 DEGs were significantly up-regulated (Figure 11 and Table S6). Functional enrichment analysis revealed that these genes were significantly involved in the following biological functions: DNA replication-dependent nucleosome assembly, DNA replication-independent nucleosome assembly, positive regulation of cell proliferation, positive regulation of mitotic nuclear division, positive regulation of cell division, cell surface receptor signaling pathway, and fibroblast growth factor

receptor binding (Table S7). KEGG analysis revealed that these DEGs may be involved in the following signaling pathways: drug metabolism-cytochrome P450, PPAR signaling pathway, transcriptional misregulation in cancer, and chemical carcinogenesis (Table S7). Through the survival analysis of these DEGs, we obtained a total of 88 LUAD prognostic-related DEGs (Figure 12A and Table S7). The top three significant DEGs are protein arginine methyltransferase 8 (PRMT8, Figure 12B), keratin 9 (KRT9, Figure 12C) and growth factor independent 1B transcriptional repressor (GFI1B, Figure 12D). In addition, we also constructed the co-expression interaction network of these DEGs through the weighted gene co-expression network analysis (WGCNA) approach. We have obtained three DEGs modules through WGCNA (Figure 13A–D). The WGCNA co-expression interaction network is displayed in Figure 14. In this WGCNA co-expression interaction network, there are two hub DEGs with the highest degree, which are histone cluster 1 H3b (HIST1H3B) and histone cluster 1 H1b (HIST1H1B) respectively, and both with the degree of 38 in turquoise module (Figure 14).

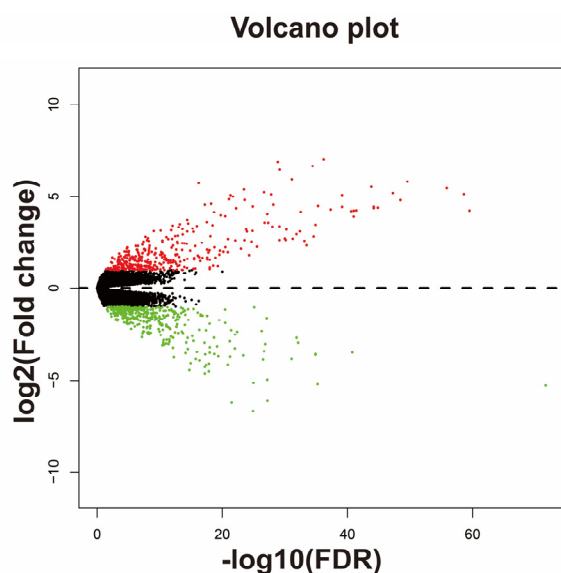


Figure 11. Volcano plot of fold change of DEGs between low- and high-risk score phenotypes.

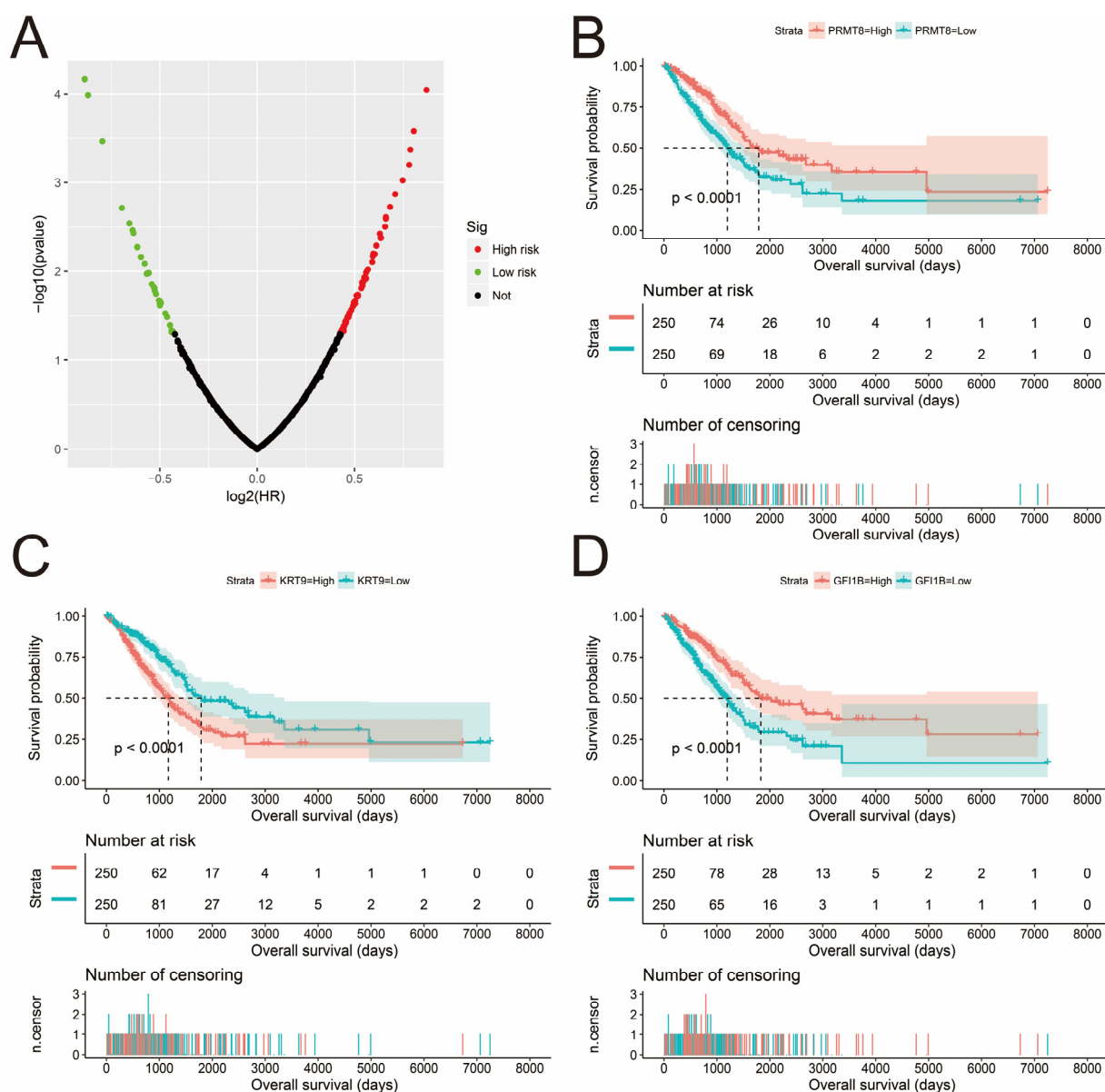


Figure 12. Survival analysis results of DEGs between low- and high-risk score phenotypes. (A) Volcano plot of prognostic analysis results of DEGs; (B) Kaplan Meier curve of *PRMT8*; (C) Kaplan Meier curve of *KRT9*; (D) Kaplan Meier curve of *GFI1B*.

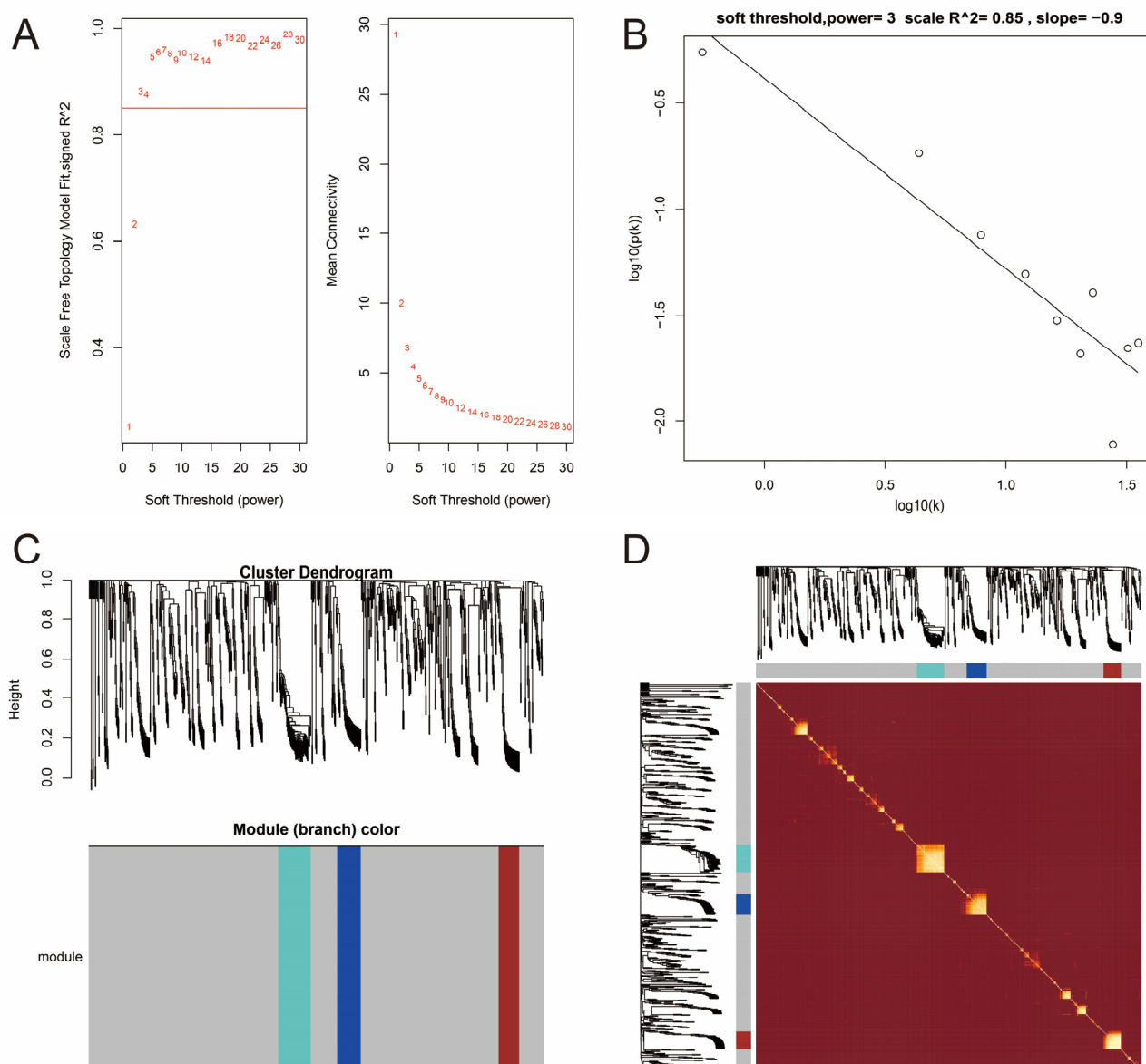


Figure 13. Results of WGCNA analysis. (A) Soft threshold (β value) distribution map; (B) Soft threshold correlation scatter plot; (C) Cluster dendrogram of WGCNA modules; (D) Tom plot of WGCNA modules.

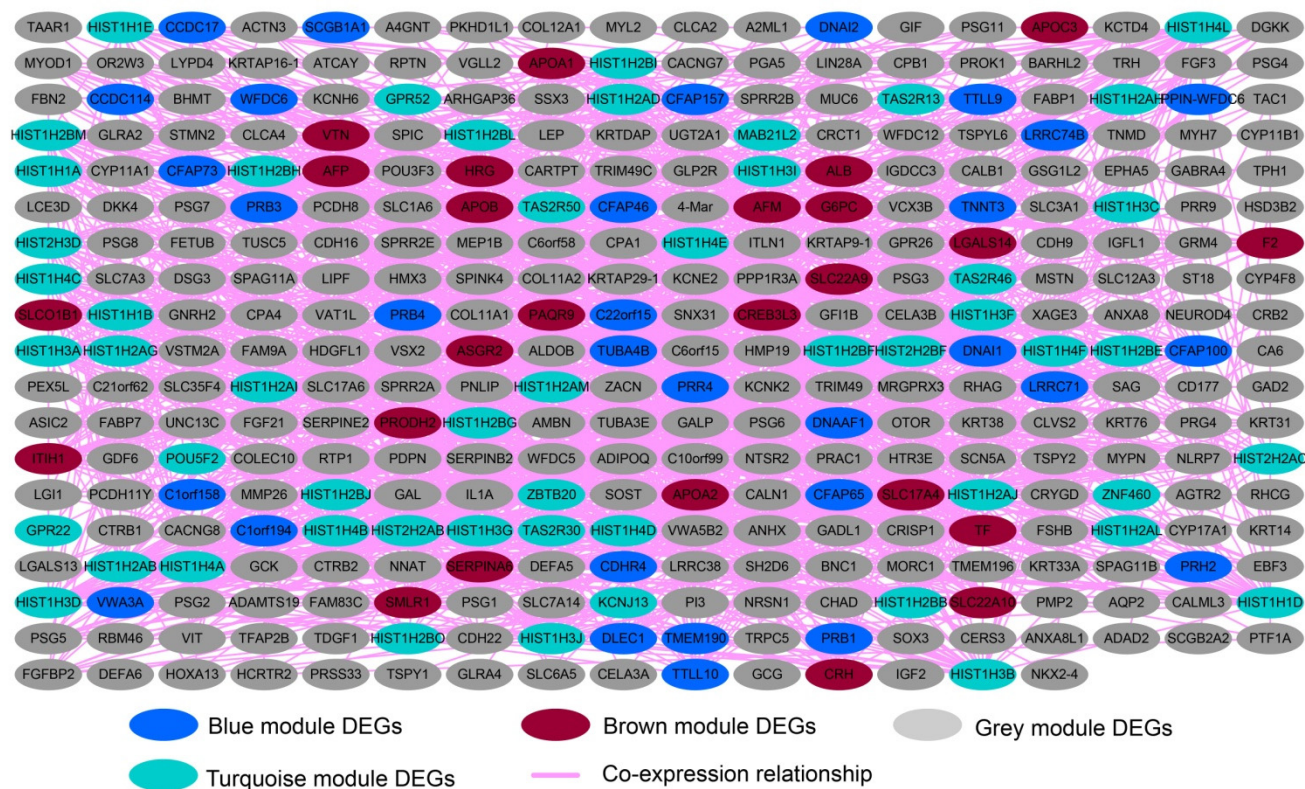


Figure 14. WGCNA co-expression interaction network.

We also use GSEA to explore the mechanism of this snoRNA expression signature. GSEA using the c5 reference gene set suggested that LUAD patients with high-risk score phenotypes are markedly different from low-risk score phenotypes in the following biological processes: ErbB signaling pathway, toll like receptor signaling pathway, regulation of cell cell adhesion, nuclear factor (NF)-kappaB-Inducing Kinase (NIK)/NF kappa B signaling, non canonical Wnt signaling pathway, cell cell G1/S phase transition, integrin mediated signaling pathway, extrinsic apoptotic signaling pathway, epidermal growth factor receptor signaling pathway, platelet derived growth factor receptor signaling pathway, transforming growth factor beta receptor signaling pathway, positive regulation of Wnt signaling pathway, receptor signaling pathway via signal transducers and activators of transcription (STAT), regulation of apoptotic signaling pathway, regulation of mitogen-activated protein kinase (MAPK) activity and phosphatidylinositol 3 kinase (PI3K) binding (Figure 15A–P). GSEA using the c2 reference gene set suggested that LUAD patients with high-risk score phenotypes are markedly different from low-risk score phenotypes in the following pathways: epidermal growth factor receptor (ErbB1) receptor proximal pathway, p38/MAPK pathway, toll like receptor signaling pathway, metastasis epithelial-to-mesenchymal transition (EMT) up, PI3K/AKT pathway, NFκB signaling, signaling by the B cell receptor (BCR), Notch pathway, apoptosis, cell cycle pathway, p53 signaling pathway, hypoxia-inducible factor (HIF) pathway, tumorigenesis up, tumor angiogenesis up, TGFB signaling pathway and integrin pathway (Figure 16A–P).

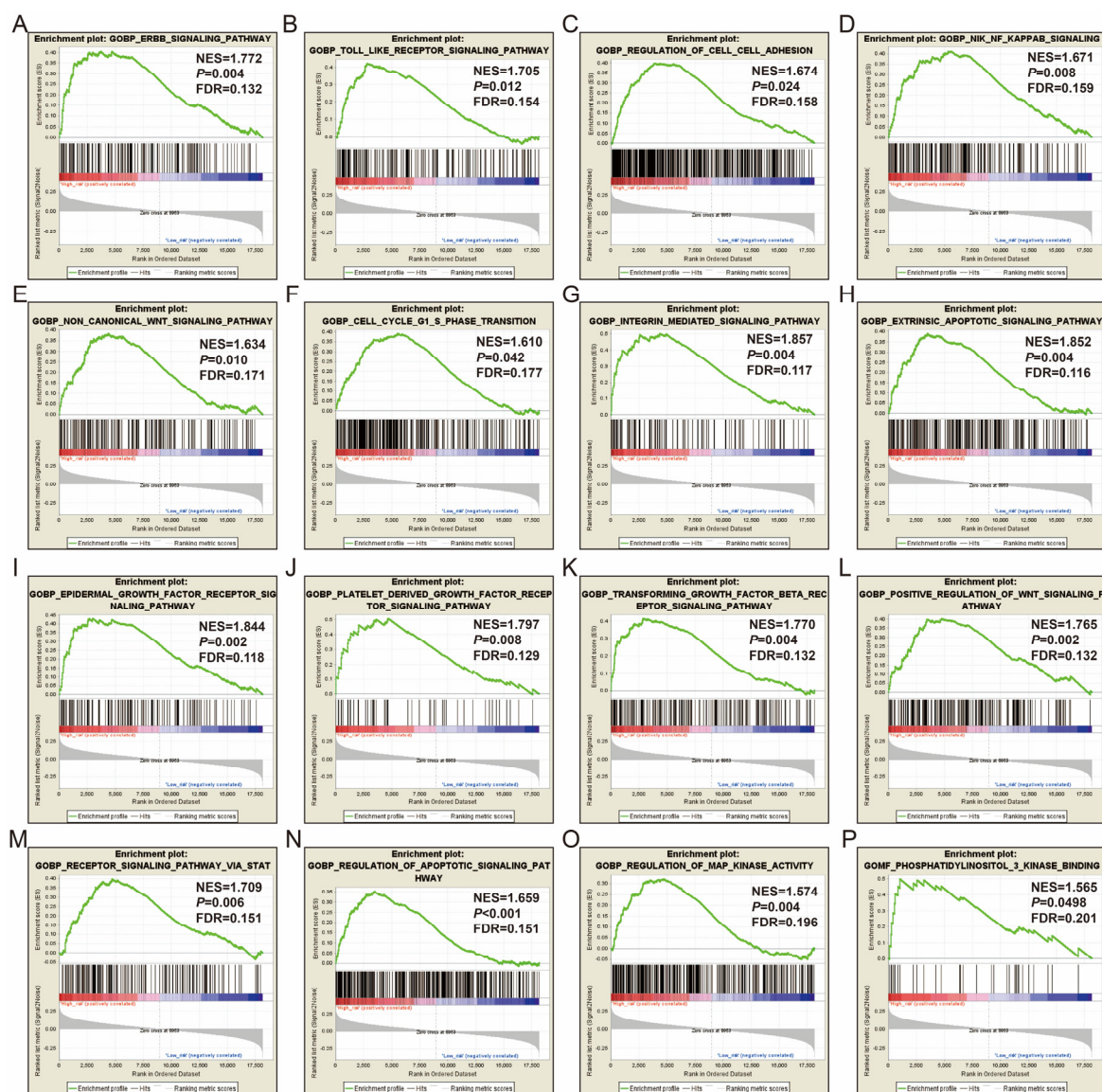


Figure 15. GSEA results between high-risk and low-risk phenotypes by using C5 reference gene set. (A) ErbB signaling pathway; (B) toll like receptor signaling pathway; (C) regulation of cell cell adhesion; (D) NIK/NF kappa B signaling; (E) non canonical Wnt signaling pathway; (F) cell cell G1/S phase transition; (G) integrin mediated signaling pathway; (H) extrinsic apoptotic signaling pathway; (I) epidermal growth factor receptor signaling pathway; (J) platelet derived growth factor receptor signaling pathway; (K) transforming growth factor beta receptor signaling pathway; (L) positive regulation of Wnt signaling pathway; (M)receptor signaling pathway via signal transducers and activators of transcription (STAT); (N)regulation of apoptotic signaling pathway; (O)regulation of mitogen-activated protein kinase (MAPK) activity; (P)phosphatidylinositol 3 kinase (PI3K) binding.

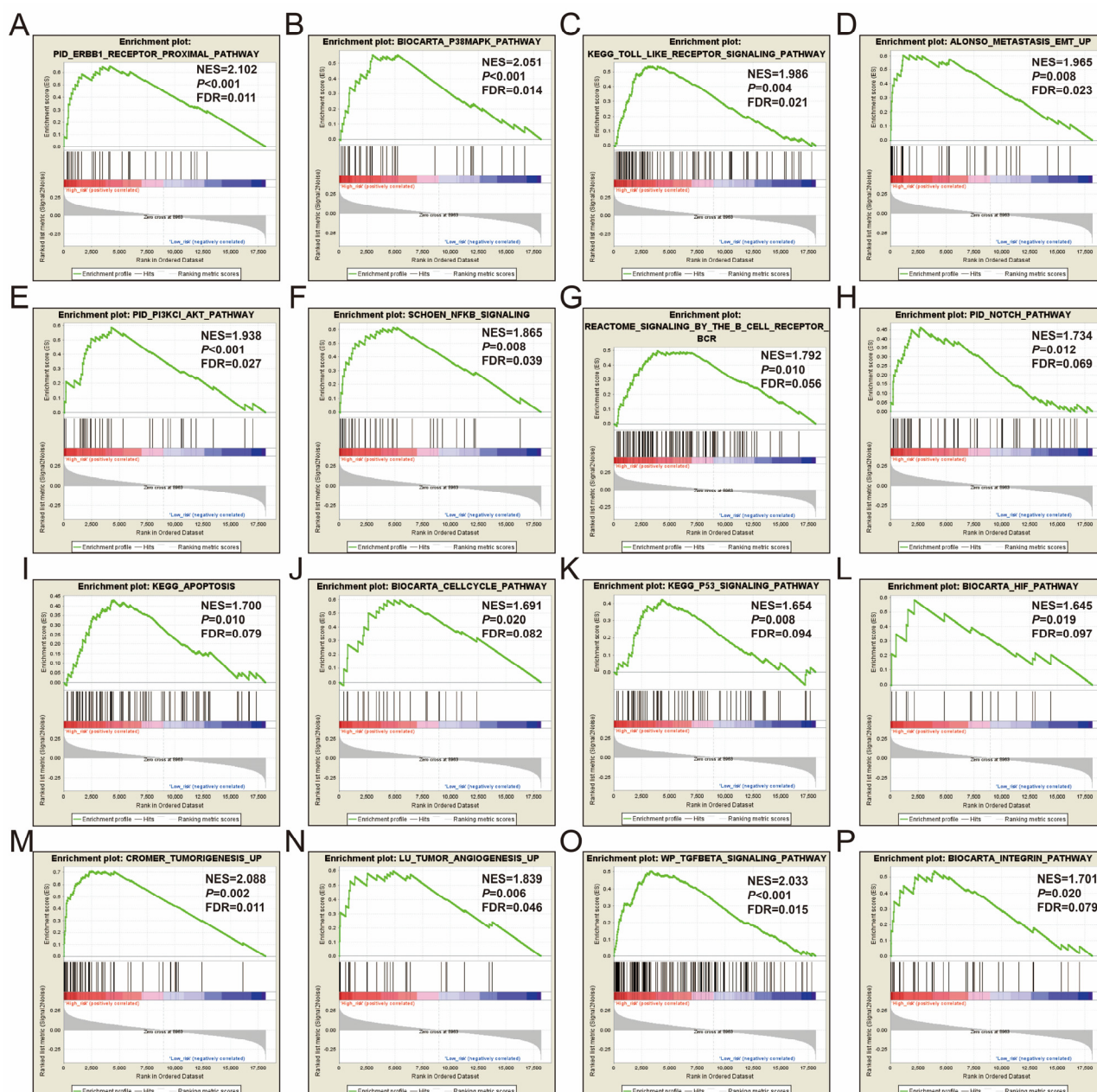


Figure 16. GSEA results between high-risk and low-risk phenotypes by using C2 reference gene set. (A) epidermal growth factor receptor (ErbB1) receptor proximal pathway; (B) p38/MAPK pathway; (C) toll like receptor signaling pathway; (D) metastasis epithelial-to-mesenchymal transition (EMT) up; (E) PI3KCI/AKT pathway; (F) NF κ B signaling; (G) signaling by the B cell receptor (BCR); (H) Notch pathway; (I) apoptosis; (J) cell cycle pathway; (K) p53 signaling pathway; (L) hypoxia-inducible factor (HIF) pathway; (M) tumorigenesis up; (N) tumor angiogenesis up; (O) TGF β signaling pathway; (P) integrin pathway.

3.4. Investigation of tumor immune microenvironment

We also used whole-genome RNA sequencing data to score the immune microenvironment in TCGA Luad patients. Stromal score was significantly increased in LUAD patients with high risk and high snoU109 (Figure 17A). Stromal score was markedly reduced in LUAD patients with high expression of SNORD104 (Figure 17A). Immune score was markedly increased in LUAD patients with high risk (Figure 17B). ESTIMATE score was markedly increased in LUAD patients with high risk and high snoU109, and was markedly reduced in LUAD patients with high expression of SNORD104 (Figure 17C).

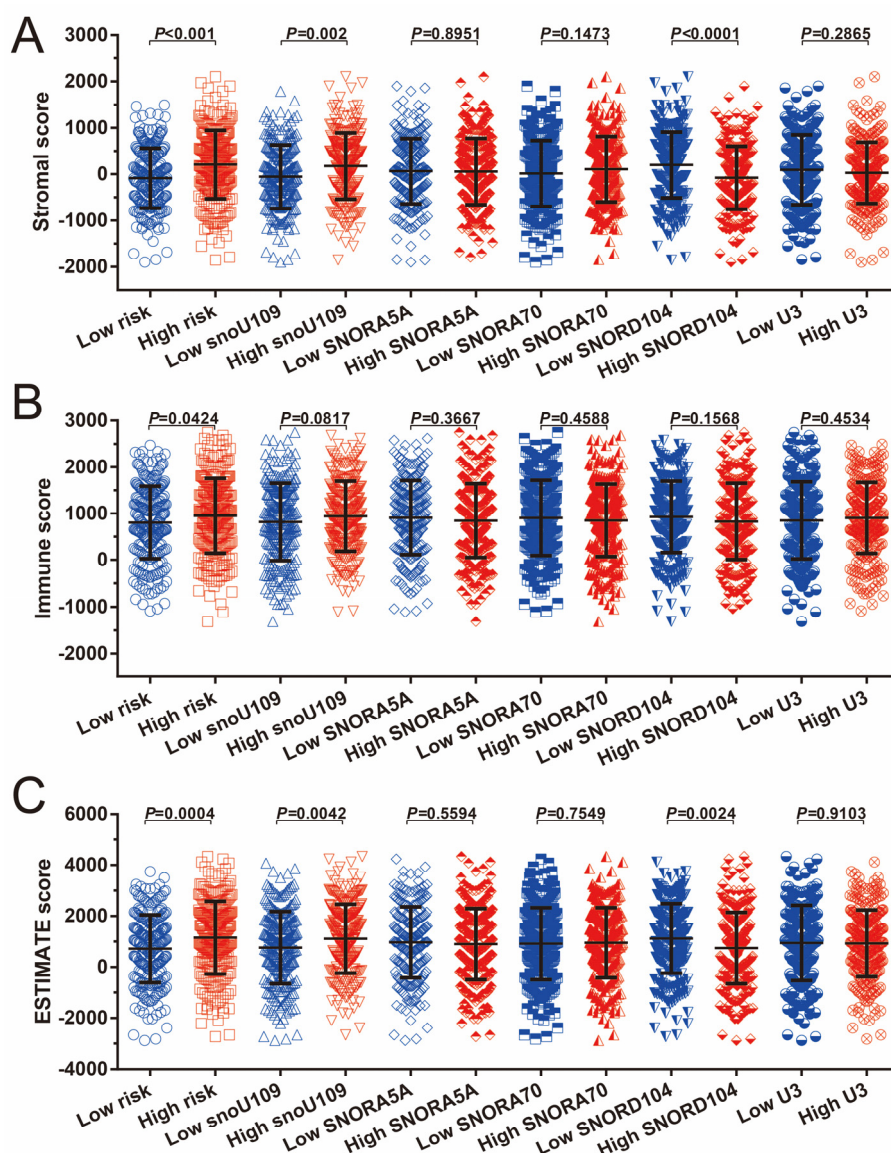


Figure 17. SnoRNA expression signature immune microenvironment score results. (A) Stromal score of risk score and its snoRNAs in LUAD tumor tissues; (B) Immune score of risk score and its snoRNAs in LUAD tumor tissues; (C) ESTIMATE score of risk score and its snoRNAs in LUAD tumor tissues.

4. Discussion

With the wide application of next-generation sequencing technology in tumors, more and more non-coding RNAs have been found to be abnormally expressed in tumor tissues, and their biological functions play a certain role in tumorigenesis, development and prognosis [16,17]. SNORNA belongs to a class of non-coding RNAs. Recently, increasing evidences have shown that SNORNA plays an important role in the tumorigenesis, development, diagnosis and prognosis of cancers [18]. Previous studies have mined the high-throughput sequencing data of TCGA and constructed a snoRNA-related online analysis tool to explore the prognosis and drug response for snoRNA in TCGA pan-cancer cohort [2,3,19]. Shuwen et al. reviewed snoRNAs related to hepatocellular carcinoma (HCC). Multiple snoRNAs have been reported to be closely related to EMT and Wnt/ β -catenin signaling pathways [4]. Dsouza et al. also reviewed the breast-cancer-related snoRNA and gave a comprehensive overview of the role of snoRNA in breast cancer. In addition, the regulation, biological function, signaling pathway and clinical efficacy of the abnormal expression of snoRNAs in breast cancer were summarized [5]. Previous studies have also identified cancer prognostic related snoRNAs and constructed prognostic snoRNAs expression signature based on the TCGA RNA sequencing dataset. Deng et al. used the TCGA lower grade glioma (LGG) patient RNA sequencing data set to screen for prognostic-related snoRNA, and initially constructed an prognostic expression signature based on eleven prognostic snoRNAs, and initially identified the potential biological function mechanisms and targeted drugs of this expression signature in LGG [7]. Cao et al. used the TCGA bladder cancer patient RNA sequencing data set to screen prognostic-related snoRNA, and then used the least absolute shrinkage and selection operator cox regression model to construct an expression signature based on five prognostic snoRNAs. And constructed a bladder cancer prognostic nomogram based on the risk score model. SuivivalROC is used to evaluate the accuracy of this risk score model to predict the 5-year survival of bladder cancer patients, and the AUC is greater than 0.7 [20]. Liu et al. also identified 15 prognostic-related snoRNAs by analyzing TCGA sarcoma cohort dataset, and constructed a expression signature based on four prognostic snoRNAs [6]. Pan et al. built an incremental feature selection based on the TCGA pan-cancer database and using the monte carlo feature selection approach to predict the prognosis of cancer patients [21]. Zhao et al. confirmed that snoRNA is a diagnostic indicator of clear cell renal cell carcinoma through TCGA data and the serum samples of the validation cohort, and also constructed a six-snoRNA signature that can be used as an independent prognostic indicator of clear cell renal cell carcinoma [22]. In this study, a prognostic expression signature of five-snoRNA was constructed based on the multivariate Cox proportional hazard regression model by LUAD prognostic snoRNAs. In a review of previous studies, no one in LUAD has ever comprehensively analyzed prognostic related snoRNA. This study is the first to report that snoRNA can be used as a prognostic marker for LUAD.

Among the five prognostic snoRNAs in the signature of this study, only U3 is closely related to cancers in previous studies, and the other four snoRNAs have not been reported in cancers. expression had a poor prognosis [6]. In this study, we also observed that LUAD patients with low U3 expression have a poor prognosis for patients with higher expression. Some DEGs or co-expressed genes of snoRNAs screened in this study that are related to the prognosis of LUAD have also been reported to be closely related to cancers in previous studies. TMEM125 has been reported to be significantly associated with prognosis in non-small cell lung cancer [23]. Previous studies have confirmed that PRMT8 plays a certain role in the proliferation, invasion, migration and drug resistance of colon cancer,

and is closely related to the formation of stem cells of colon cancer cells [24]. PRMT8 is highly expressed in breast cancer, ovarian cancer, and gastric cancer. High PRMT8 expression is associated with increased survival in breast and ovarian cancer patients, while high PRMT8 expression is associated with reduced survival in gastric cancer patients [25]. KRT9 and HSP70 interact to influence bladder cancer cells invasion and metastasis in vitro [26]. GFI1B is necessary for the development of erythrocyte and megakaryocyte cell lines, and its abnormal function often leads to cancer or hematologic malignancies [27]. Meanwhile, study have shown that GFI1B can regulate the immobility of hematopoietic stem cells and the differentiation of red blood cells and platelets, and the loss of GFI1B heterozygosity can accelerate the progression of acute myeloid leukemia (AML) [28]. Low expression of GFI1B is associated with poor prognosis in patients with myelodysplastic syndrome (MDS) and AML [29]. Spi-1 proto-oncogene (SPI1) is a regulatory target of GFI1B, and there is a negative correlation between the two genes. GFI1B can affect the development of erythroid and bone marrow mononuclear cells by regulating SPI1 [30]. GFI1B p32 mutation can reduce the number of platelets, p32 is elevated in acute and chronic leukemia cells, and cancer-related genes in humans are significantly dysregulated, confirming that GFI1B may play a role in carcinogenic regulation [31]. Significant overexpression of GFI1B in hematological malignancies can increase the proliferation of leukemia cells and make the disease progress. Targeted inhibition of GFI1B can significantly inhibit the proliferation of hematological malignancies and can be used as a potential therapeutic target for hematological malignancies [32,33]. In the functional enrichment analysis, we enriched a large number of classic tumor-related signaling pathways, which may be the potential molecular mechanism of snoRNA's function in LUAD.

The objective evaluation of this research is not perfect. First, this study is a single-center cohort study and lacks additional validation cohorts. Second, this study is based on RNA sequencing data for functional enrichment analysis, and the biological functional mechanisms selected still need to be verified by in vivo and in vitro experiments. Third, due to the limited clinical parameters and sample size provided by the TCGA official website, our results still need further verification. Despite the above shortcomings, our research still innovative. First of all, this study is the first comprehensive screening study of LUAD prognostic-related snoRNAs based on RNA sequencing dataset, and no one has ever reported on LUAD prognostic snoRNAs before. Secondly, we also use prognostic-related snoRNA to construct a prognostic signature model. Third, we also explored the molecular mechanism of this snoRNA feature based on RNA sequencing data. In summary, this study innovatively identified the prognostic related snoRNAs of LUAD for the first time, and comprehensively explored them from the perspective of prognosis and molecular mechanisms. It has certain clinical application value and provides theoretical basis for future study.

5. Conclusions

In summary, seven LUAD prognostic related snoRNAs were screened in this study. We also constructed an expression signature containing five LUAD prognostic related snoRNAs (snoU109, SNORA5A, SNORA70, SNORD104 and U3). Functional analysis indicated that LUAD patients with different risk score phenotypes had significant differences in cell cycle, apoptosis, integrin, TGFB, ErbB, NFκB, MAPK, PI3K and toll like receptor signaling pathway. These biological functional mechanisms may be the potential factors for the significant differences in prognosis between different risk score phenotype LUAD patients. Immune microenvironment analysis also indicated that there

were significant differences in immune microenvironment scores among LUAD patients with different risk score. However, our results still need to be validated in future studies.

Acknowledgement

This study was supported in part by the self-raised Scientific Research Fund of the Ministry of Health of Guangxi Province (Z20190998 and Z20210384), The Basic Ability Improvement Project for Middle-aged and Young Teachers in Colleges and Universities in Guangxi (2020KY03050).

Conflict of interest

The authors declare that they have no competing interests.

References

1. T. Bratkovic, J. Bozic, B. Rogelj, Functional diversity of small nucleolar RNAs, *Nucleic Acids Res.*, **48** (2020), 1627–1651.
2. J. Gong, Y. Li, C. J. Liu, Y. Xiang, C. Li, Y. Ye, et al., A pan-cancer analysis of the expression and clinical relevance of small nucleolar RNAs in human cancer, *Cell Rep.*, **21** (2017), 1968–1981.
3. J. Liang, J. Wen, Z. Huang, X. P. Chen, B. X. Zhang, L. Chu, Small nucleolar RNAs: Insight into their function in cancer, *Front. Oncol.*, **9** (2019), 587.
4. H. Shuwen, Y. Xi, Q. Quan, J. Yin, D. Miao, Can small nucleolar RNA be a novel molecular target for hepatocellular carcinoma?, *Gene*, **733** (2020), 144384.
5. V. L. Dsouza, D. Adiga, S. Sriharikrishnaa, P. S. Suresh, A. Chatterjee, S. P. Kabekkodu, Small nucleolar RNA and its potential role in breast cancer-A comprehensive review, *Biochem. Biophys. Acta Rev. Cancer*, **1875** (2021), 188501.
6. J. Liu, X. Liao, X. Zhu, P. Lv, R. Li, Identification of potential prognostic small nucleolar RNA biomarkers for predicting overall survival in patients with sarcoma, *Cancer Med.*, **9** (2020), 7018–7033.
7. T. Deng, Y. Gong, X. Liao, X. Wang, X. Zhou, G. Zhu, et al., Integrative analysis of a novel eleven-small nucleolar RNA prognostic signature in patients with lower grade glioma, *Front. Oncol.*, **11** (2021), 650828.
8. Cancer Genome Atlas Research Network, Comprehensive molecular profiling of lung adenocarcinoma, *Nature*, **511** (2014), 543–550.
9. L. Zhang, G. Zhu, X. Wang, X. Liao, R. Huang, C. Huang, et al., Genomewide investigation of the clinical significance and prospective molecular mechanisms of kinesin family member genes in patients with lung adenocarcinoma, *Oncol. Rep.*, **42** (2019), 1017–1034.
10. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Nat. Acad. Sci. U. S. A.*, **102** (2005), 15545–15550.
11. A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, J. P. Mesirov, P. Tamayo, The molecular signatures database (MSigDB) hallmark gene set collection, *Cell Syst.*, **1** (2015), 417–425.
12. G. Dennis, Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, et al., DAVID: Database for annotation, visualization and integrated discovery, *Genome Biol.*, **4** (2003), P3.

13. X. Jiao, B. T. Sherman, W. H. Da, R. Stephens, M. W. Baseler, H. C. Lane, et al., DAVID-WS: a stateful web service to facilitate gene/protein list analysis, *Bioinformatics*, **28** (2012), 1805–1806.
14. K. Yoshihara, M. Shahmoradgoli, E. Martinez, R. Vegesna, H. Kim, W. Torres-Garcia, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, *Nat. Commun.*, **4** (2013), 2612.
15. Y. Benjamini, D. Drai, G. Elmer, N. Kafkafi, I. Golani, Controlling the false discovery rate in behavior genetics research, *Behav. Brain Res.*, **125** (2001), 279–284.
16. M. Esteller, Non-coding RNAs in human disease, *Nat. Rev. Genet.*, **12** (2011), 861–874.
17. G. Romano, D. Veneziano, M. Acunzo, C. M. Croce, Small non-coding RNA and cancer, *Carcinogenesis*, **38** (2017), 485–491.
18. G. T. Williams, F. Farzaneh, Are snoRNAs and snoRNA host genes new players in cancer?, *Nat. Rev. Cancer*, **12** (2012), 84–88.
19. Y. Liu, H. Ruan, S. Li, Y. Ye, W. Hong, J. Gong, et al., The genetic and pharmacogenomic landscape of snoRNAs in human cancer, *Mol. Cancer*, **19** (2020), 108.
20. R. Cao, B. Ma, L. Yuan, G. Wang, Y. Tian, Small nucleolar RNAs signature (SNORS) identified clinical outcome and prognosis of bladder cancer (BLCA), *Cancer Cell Int.*, **20** (2020), 299.
21. X. Pan, L. Chen, K. Y. Feng, X. H. Hu, Y. H. Zhang, X. Y. Kong, et al., Analysis of expression pattern of snoRNAs in different cancer types with machine learning algorithms, *Int. J. Mol. Sci.*, **20** (2019).
22. Y. Zhao, Y. Yan, R. Ma, X. Lv, L. Zhang, J. Wang, et al., Expression signature of six-snoRNA serves as novel non-invasive biomarker for diagnosis and prognosis prediction of renal clear cell carcinoma, *J. Cell Mol. Med.*, **24** (2020), 2215–2228.
23. H. Yu, Z. Pang, G. Li, T. Gu, Bioinformatics analysis of differentially expressed miRNAs in non-small cell lung cancer, *J. Clin. Lab. Anal.*, **35** (2021), e23588.
24. H. Lin, B. Wang, J. Yu, J. Wang, Q. Li, B. Cao, Protein arginine methyltransferase 8 gene enhances the colon cancer stem cell (CSC) function by upregulating the pluripotency transcription factor, *J. Cancer*, **9** (2018), 1394–1402.
25. S. J. Hernandez, D. M. Dolivo, T. Dominko, PRMT8 demonstrates variant-specific expression in cancer cells and correlates with patient survival in breast, ovarian and gastric cancer, *Oncol. Lett.*, **13** (2017), 1983–1989.
26. C. Andolino, C. Hess, T. Prince, H. Williams, M. Chernin, Drug-induced keratin 9 interaction with Hsp70 in bladder cancer cells, *Cell Stress Chaperones*, **23** (2018), 1137–1142.
27. E. Anguita, F. J. Candel, A. Chaparro, J. J. Roldan-Etcheverry, Transcription factor GFI1B in health and disease, *Front. Oncol.*, **7** (2017), 54.
28. A. N. Cheng, E. L. Bao, C. Fiorini, V. G. Sankaran, Macrothrombocytopenia associated with a rare GFI1B missense variant confounding the presentation of immune thrombocytopenia, *Pediatr. Blood Cancer*, **66** (2019), e27874.
29. A. Thivakaran, L. Botezatu, J. M. Hones, J. Schutte, L. Vassen, Y. S. Al-Matary, et al., Gfi1b: a key player in the genesis and maintenance of acute myeloid leukemia and myelodysplastic syndrome, *Haematologica*, **103** (2018), 614–625.
30. E. Anguita, R. Gupta, V. Olariu, P. J. Valk, C. Peterson, R. Delwel, et al., A somatic mutation of GFI1B identified in leukemia alters cell fate via a SPI1 (PU.1) centered genetic regulatory network, *Dev. Biol.*, **411** (2016), 277–286.

31. M. Faleschini, N. Papa, M. C. Morel-Kopp, C. Marconi, T. Giangregorio, F. Melazzini, et al., Dysregulation of oncogenic factors by GFI1B p32: investigation of a novel GFI1B germline mutation, *Haematologica*, 2021.
32. A. H. Elmaagacli, M. Koldehoff, J. L. Zakrzewski, N. K. Steckel, H. Ottinger, D. W. Beelen, Growth factor-independent 1B gene (GFI1B) is overexpressed in erythropoietic and megakaryocytic malignancies and increases their proliferation rate, *Br. J. Haematol.*, **136** (2007), 212–219.
33. L. Vassen, C. Khandanpour, P. Ebeling, B. A. van der Reijden, J. H. Jansen, S. Mahlmann, et al., Growth factor independent 1b (Gfi1b) and a new splice variant of Gfi1b are highly expressed in patients with acute and chronic leukemia, *Int. J. Hematol.*, **89** (2009), 422–430.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)