Research Article

Ziyu Liang#, Dongxing Su#, Kang Liu, Haixing Jiang*

Comprehensive analysis of molecular mechanism and a novel prognostic signature based on small nuclear RNA biomarkers in gastric cancer patients

https://doi.org/10.1515/med-2022-0493
received February 17, 2022; accepted April 28, 2022

Abstract: Small nuclear RNAs (snRNAs) are rarely reported in cancer. This study is based on The Cancer Genome Atlas genome-wide data set to explore the prognostic value and molecular mechanism of snRNAs in gastric cancer (GC). Gene ontology, Kyoto Encyclopedia of Genes and Genomes, and gene set enrichment analysis were used to explore the molecular mechanism of snRNAs. A total of 351 patients were included in the survival analysis, and 14 prognostic snRNAs were identified using multivariate survival analysis. We constructed a prognostic signature containing nine snRNAs, which can signally classify patients into high- and low-risk phenotypes (adjusted $P < 0.0001$, hazard ratio = 2.671, 95% confidence interval = 1.850–3.858). Combining the molecular mechanisms obtained by the three functional enrichment approaches, we concluded that this prognostic signature snRNAs participated in classical tumor-related signaling pathways, including Notch, PI3K, toll-like receptor, etc.; cell adhesion; cell cycle; cell proliferation; and other biological processes that affect the biological phenotype of cancer cells. We also found significant downregulation of the abundance of immune cell infiltrates and immune microenvironment scores for high-risk phenotypes of GC patients. In conclusion, this study has identified 14 prognostic snRNAs signally associated with GC overall survival and also constructed a novel prognostic signature containing nine prognostic snRNAs.

Keywords: small nuclear RNA, molecular mechanism, gastric cancer, overall survival, The Cancer Genome Atlas

1 Introduction

Gastric cancer (GC) is a common type of cancer worldwide; 5-year relative survival rate is about 20% [1,2]. The main pathological type was gastric adenocarcinoma. The prognosis of GC is directly related to the stage at the time of diagnosis. To this day, surgery is still the major strategy for GC [3,4]. However, due to the low diagnosis rate of GC at the early stage, the survival rate is only about 10% [5,6]. Most GC patients are in an advanced stage at the time of diagnosis, and the 5-year survival rate is about 7–34% [1]. Hence, this is an impendency requirement to develop efficient biomarkers for GC diagnosis and prognostic assessment [7,8]. The study has reported that normally small nuclear RNA (snRNA) does not exist free, but combines with protein to form a complex and becomes small nuclear ribonucleoprotein particles (snRNPs) [9,10]. It has been confirmed that snRNAs do not participate in protein synthesis, and their main function is to play an indispensable role in RNA processing. The protein part of snRNA has nuclease and ligase activities, which can cut transcription at the intron–exon junction and connect the two free ends [9,11]. snRNA participates in the construction of nucleoprotein complexes and performs splicing functions by pairing bases with target site RNA or snRNA [12,13]. At the same time, it has also been reported that snRNA may be involved in

# These authors have contributed equally to this work.

*** Corresponding author: Haixing Jiang,** Department of Gastroenterology, The First Affiliated Hospital of Guangxi Medical University, Shuang Yong Road 6, Nanning, 530021, Guangxi Zhuang Autonomous Region, People's Republic of China,
e-mail: gxjianghx@163.com, tel: +86-771-5356501,
fax: +86-771-5350031
Ziyu Liang, Dongxing Su: Department of Gastroenterology, The Third Affiliated Hospital of Guangxi Medical University, Nanning, 530000, Guangxi Zhuang Autonomous Region, People's Republic of China
Kang Liu: Department of Radiation Oncology, The First Affiliated Hospital of Guangxi Medical University, Nanning, 530021, Guangxi Zhuang Autonomous Region, People's Republic of China.

gene activation and the processing of rRNA precursors. As a component of chromatin and nuclear structure, snRNA may play an important role in maintaining its special structure and may be involved in chromatin replication and transcription [11,14]. With the rapid development of high-throughput sequencing technology, more and more non-coding RNAs have been discovered, and their molecular mechanisms have been continuously verified and elucidated. By reviewing the literature, we note that snRNAs have been mentioned in a large number of non-coding RNA-related studies, but few studies have specifically focused on snRNAs, especially in cancer studies. To make up for the research gap of snRNAs in cancers. In this study, the clinical outcome and molecular mechanism of snRNAs in GC were comprehensively explored using The Cancer Genome Atlas (TCGA) genome-wide RNA sequencing data.

2 Material and methods

2.1 Data download and normalization

The RNA sequencing data of the GC cohort were downloaded from TCGA Data Portal (https://portal.gdc.cancer.gov) [15,16]. EdgeR was used to normalize raw RNA sequencing data [17]. We obtained RNA sequencing data from 407 samples in total, including 32 paracancerous tissue samples and 375 tumor tissue samples. We excluded 7 patients with missing survival information and 17 patients with overall survival (OS) time of zero. Finally, 351 GC patients were participated in the final survival analysis. We extracted 1,872 snRNAs from the RNA sequencing data set. After edgeR normalization, we filter out snRNAs with a mean value of less than 1 and finally got 554 snRNAs for inclusion in the final survival analysis. As the authors did not perform any animal or human experiments in this study, ethics committee approval was not required.

2.2 Survival analysis and construction of snRNA signature

We screened the prognostic snRNAs using a multivariate Cox proportional hazards regression model, and the step function was used to construct a prognostic signature model for these prognostic snRNAs. After screening by a step function, we will accumulate bonus points for multiple snRNAs and get a comprehensive score. A comprehensive risk assessment of the patient was conducted to separate high- and low-risk GC patients: Risk score = expression of snRNA1 $\times \beta 1$ + expression of snRNA2 $\times \beta 2$ + ... expression of snRNAn $\times \beta n$ [18–21]. Survival receiver operating characteristic (survivalROC) analysis can be applied to assess the prognostic accuracy of this signature. The nomogram is based on the prognostic signature and routine clinical information for scoring in individualized prognostic analysis.

2.3 Investigation of the molecular mechanism of snRNA prognostic signature

To investigate the molecular mechanisms of this snRNA signature, we employed three different functional enrichment analysis methods for molecular mechanism exploration.

First, we obtained a protein-coding gene data set from RNA-sequencing data, and the data normalization method was also carried out using edgeR. Screening of snRNA co-expressed genes by Pearson correlation coefficient (r), $|r| > 0.4$, and $P < 0.05$ were considered as snRNA co-expressed genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were performed using the DAVID V6.8 (https://david.ncifcrf.gov/home.jsp) online analysis tool [22–24]. Second, we also use clusterProfiler to perform gene set enrichment analysis (GSEA) in R platform [25,26]. Finally, we also screened differentially expressed genes (DEGs) for high- and low-risk group patients in the R platform by the edgeR package. Functional enrichment analysis of these DEGs was subsequently performed to screen out the molecular mechanisms underlying this prognostic signature. The criteria for identifying differentially expressed genes are as follows: |log2 fold change (FC)| >1, $P < 0.05$, and false discovery rate (FDR) <0.05. We also performed a co-expression network analysis of DEGs using weighted gene co-expression network analysis (WGCNA) to facilitate the identification of hub DEGs [27,28].

2.4 Immunomicroenvironment and ssGSEA analysis

Immune microenvironment scoring is performed in the R platform by the Estimation of STromal and Immune cells in MAlignant Tumours using Expression data' (ESTIMATE) package [29]. Single sample gene set enrichment analysis (ssGSEA) analysis was performed in the R platform with gene set variation analysis packages [30,31].

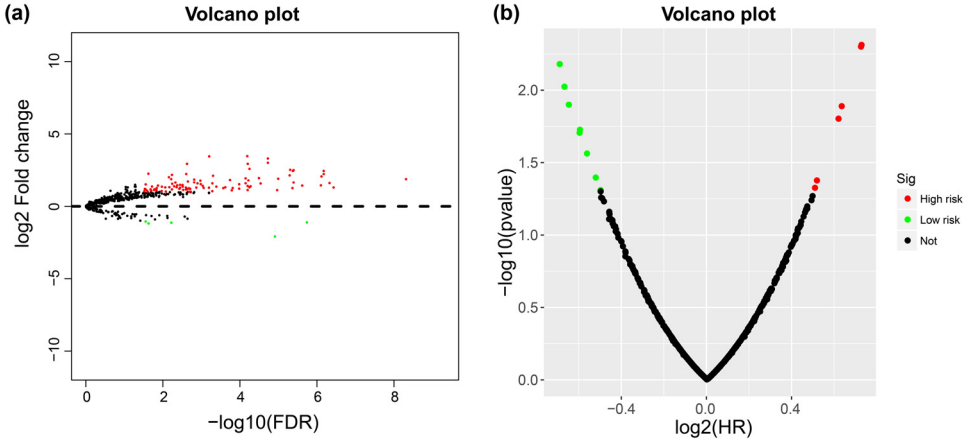

Figure 1: Volcano plot of fold change and survival analysis of snRNAs in GC: (a) volcano plot of the fold change of snRNAs in GC; (b) volcano plot of prognosis analysis of snRNAs in GC.

2.5 Statistical analysis

The co-expressed genes of snRNAs were determined using Pearson correlation coefficient r assessed using an independent sample t test. Kaplan–Meier analysis was applied to univariate survival analysis, and Cox proportional hazards regression model was used for multivariate survival analysis. Statistical analysis was performed using the SPSS version22 or R version 4.0.2. Statistical significance was determined at $P < 0.05$.

3 Results

3.1 Comprehensive survival analysis and construction of snRNA signature

By using edgeR to compare the expression abundance of snRNA between cancer and adjacent tissues in the TCGA GC queue, we obtained a total of 107 differentially expressed snRNAs. Among them, 5 snRNAs were significantly downregulated and 102 snRNAs were signally upregulated (Figure 1a, Figure S1 and Table S1). All snRNA differentially expressed fold changes are summarized in Table S1. The most signally downregulated snRNA was RNU1-70P (ENSG00000199488, log2FC = −2.09, $P < 0.0001$, and FDR < 0.0001), and the most signally upregulated snRNA was RNU6-438P (ENSG00000202431, log2FC = 3.460, $P < 0.0001$, and FDR <0.0001). Baseline parameters of patients in the TCGA GC cohort are summarized in Table S2. We observed that age (log-rank $P = 0.011$) and tumor stage (log-rank $P < 0.0001$) were signally associated with GC OS (Table S2), and these two clinical parameters were included in the subsequent multivariate Cox proportional hazards regression model. By performing survival analysis on all snRNAs, we observed that 14 snRNAs were signally related to GC OS in the TCGA cohort (Figure 1b). Among them, the most significant is RNU6–117P (ENSG00000202285, adjusted $P = 0.0049$, hazard ratio (HR) = 1.654, 95% confidence

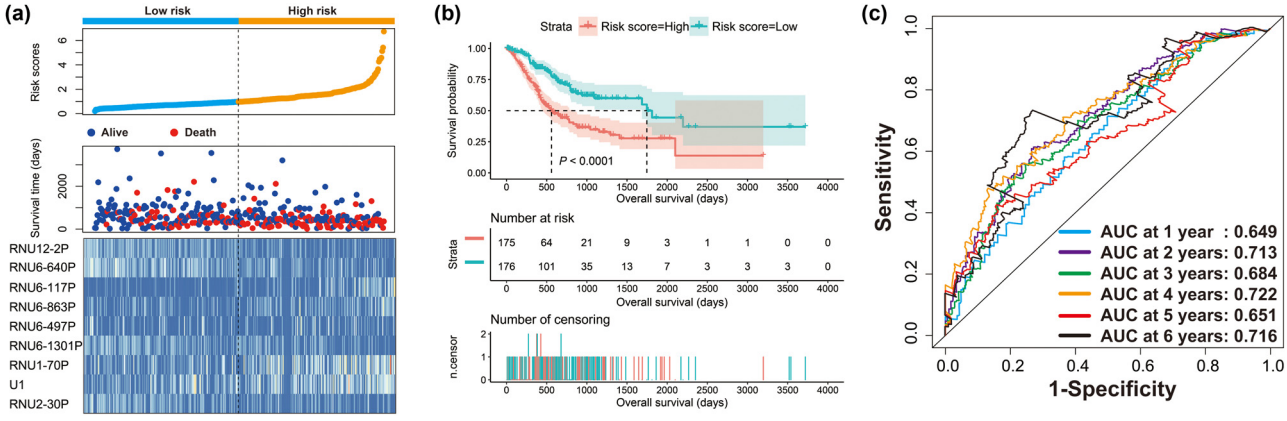

Figure 2: Prognostic value of this prognostic signature containing nine snRNAs: (a) risk score and survival time distribution plot; (b) Kaplan–Meier curve of the risk score in GC; (c) SurvivalROC curve of the risk score in GC.

interval [CI] = 1.165–2.348, Table S3). In subsequent prognostic signature construction, we screened out an optimal combination containing nine prognostic snRNAs, including RNU12–2P (ENSG00000201659), RNU6–640P (ENSG00000200563), RNU6–117P (ENSG00000202285), RNU6–863P (ENSG00000251798), RNU6–497P (ENSG00000202186), RNU6–1301P (ENSG00000199594), RNU1–70P (ENSG00000199488), U1 (ENSG00000274428), and RNU2–30P (ENSG00000252018). We observed that this prognostic signature could divide GC patients into two groups with significant prognostic differences. Patients in the high-risk phenotype had a signally higher risk of death than those in the low-risk phenotype (adjusted $P < 0.0001$, HR = 2.671, 95% CI = 1.850–3.858; Figure 2a and b). SurvivalROC evaluation showed that the prediction precision of the snRNAs signature was highest at the time point of 4 years, and the area under the curve was 0.722 (Figure 2c). Using Kaplan–Meier univariate survival analysis, we

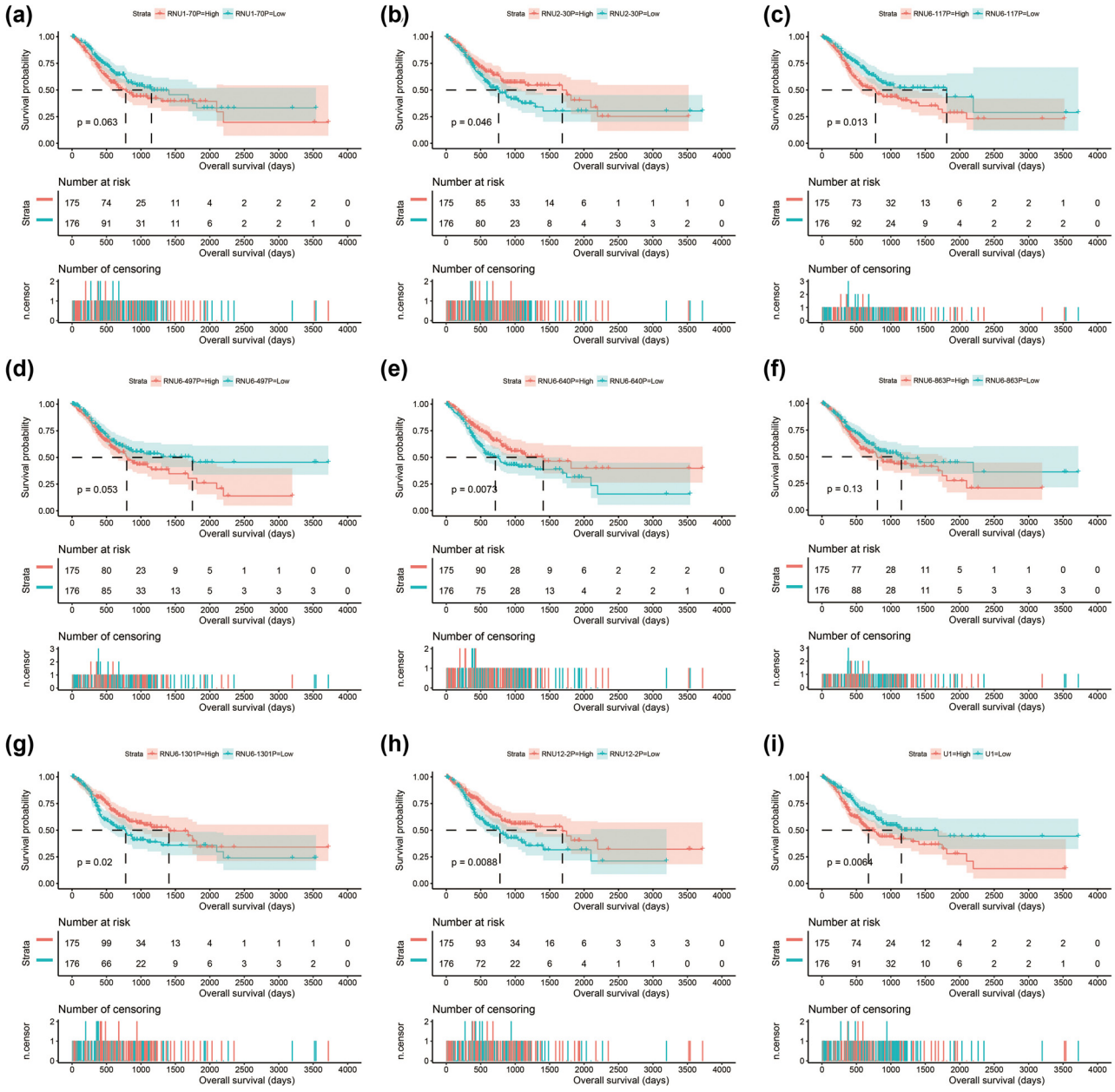

Figure 3: Kaplan–Meier curve of this prognostic signature containing nine snRNAs in GC: (a) RUN1–70P; (b) RUN2–30P; (c) RUN6–117P; (d) RUN6–497P; (e) RUN6–640P; (f) RUN6–863P; (g) RUN6–1301P; (h) RUN12–2P; and (i) U1.

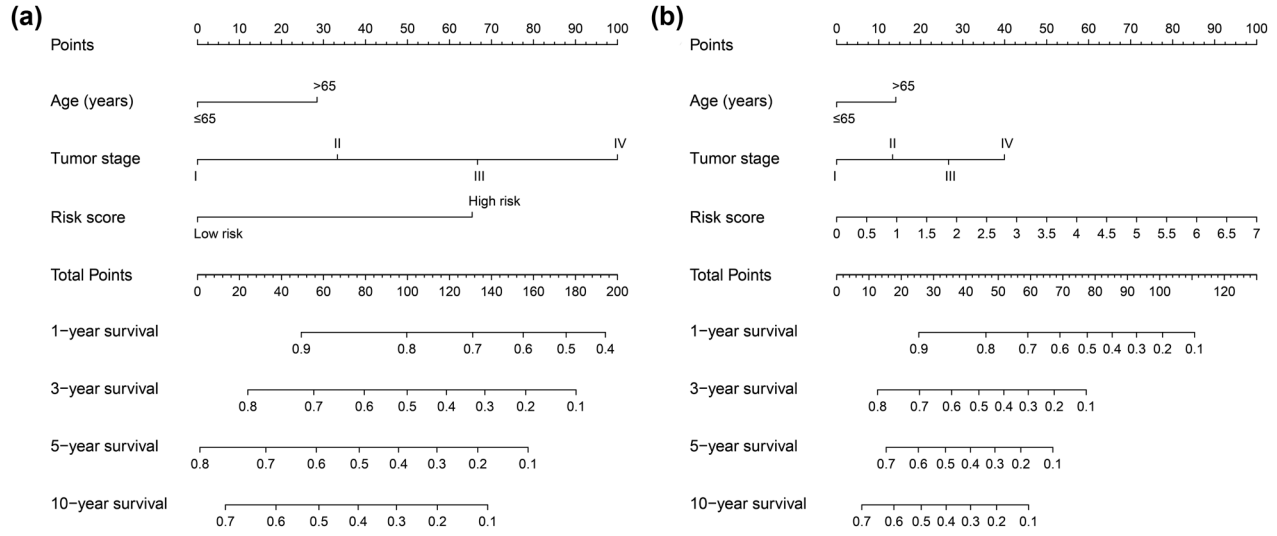

Figure 4: Nomogram for risk score in GC OS: (a) nomogram for low- and high-risk score phenotypes in GC OS; (b) nomogram for risk score in GC OS.

observed that three snRNAs (RNU1–70P, RNU6–497P, and RNU6–863P) were not signally related to OS, but after multivariate adjustment, they were found to be significantly associated with GC OS (Figure 3a–i). Nomogram analysis of this prognostic signature also showed that risk score contributed the most to GC death compared to traditional clinical parameters (Figure 4a–b).

3.2 Investigation of the molecular mechanism of snRNA prognostic signature

By screening the co-expression-related genes of nine snRNAs, we got 2,481 snRNA–mRNA co-expression relationship pairs (Figure 5, Table S4). Our survival analysis

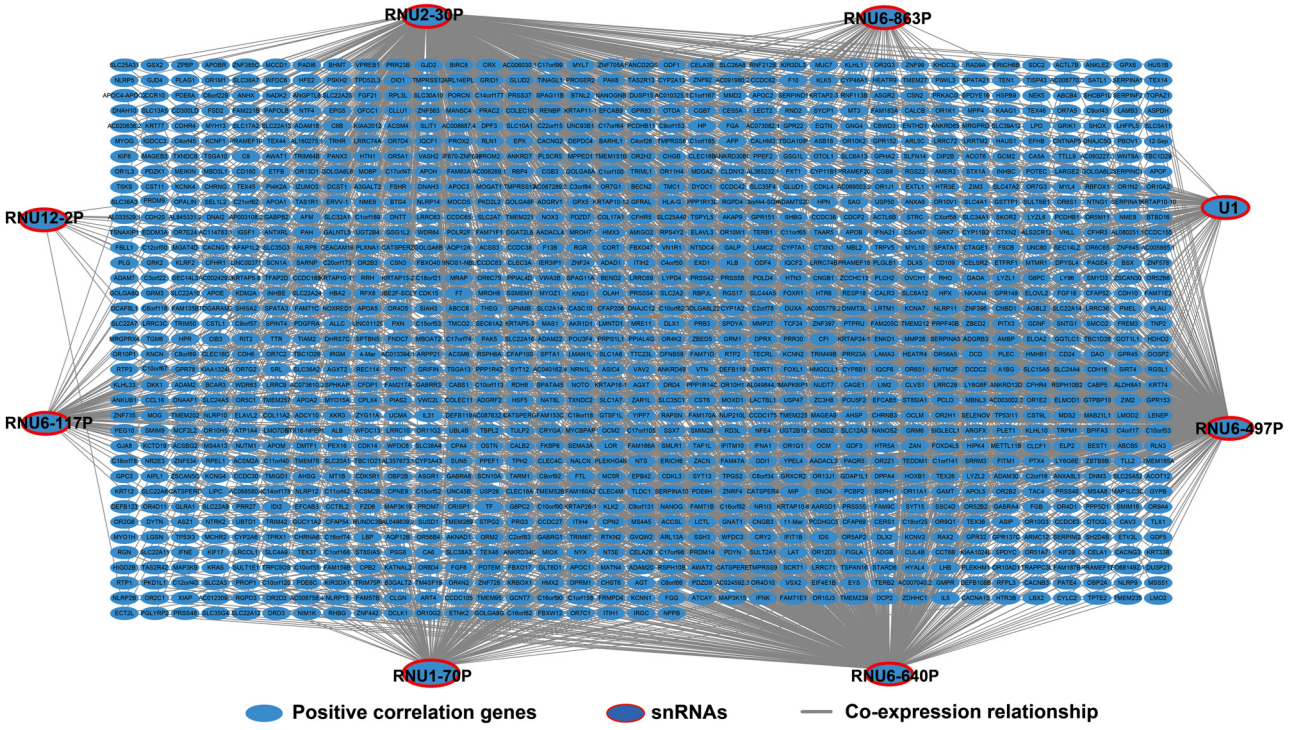

Figure 5: Interaction network of prognostic snRNAs and mRNA co-expression pairs.

of snRNA co-expressed genes observed that 122 protein-coding genes were significantly associated with GC OS (Figure 6a–d, Table S5). Functional enrichment investigation of these snRNA co-expressed genes revealed that they are involved in the following biological functions and pathway mechanisms: G-protein-coupled receptor activity, G-protein coupled receptor signaling pathway, retinoid metabolic process, regulation of Cdc42 protein signal transduction, cell differentiation, triglyceride catabolic process, and negative regulation of BMP signaling pathway (Table S6).

GSEA for pathway investigation revealed that GC patients in high-risk group are signally different from low-risk group in the following pathways: silenced by tumor microenvironment, P73 pathway, toll like receptor (TLR) TLR1/TLR2 cascade, signal transducer and activator of transcription 3 (STAT3) targets up, integrin3 pathway, metastasis epithelial–mesenchymal transition up, transforming growth factor beta 1 signaling, mitogen-activated protein kinase kinase (MAP2K) and mitogen-activated protein kinase (MAPK) activation, epidermal growth

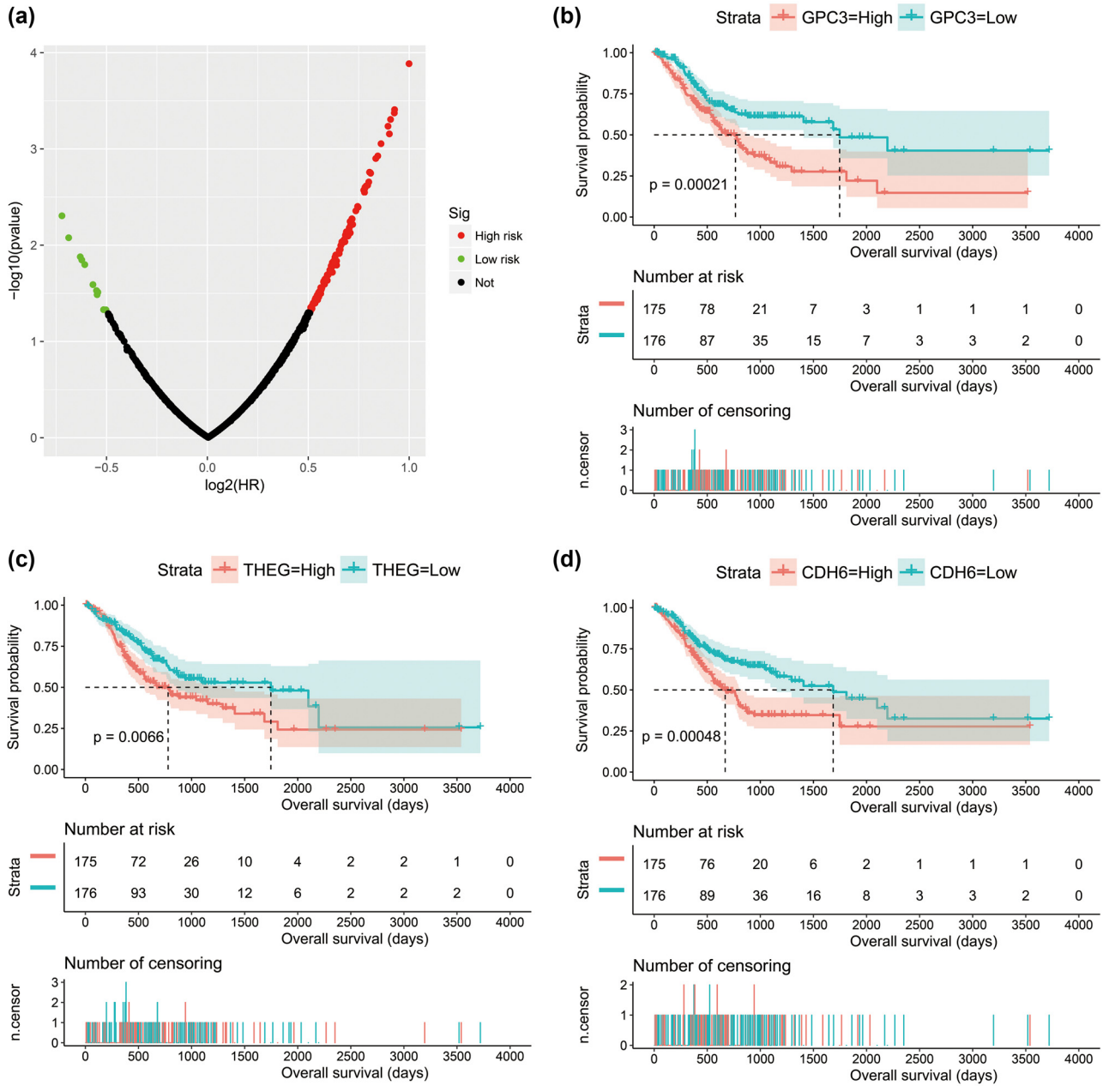

Figure 6: Survival analysis results of snRNA co-expressed genes: (a) volcano plot of survival analysis results; (b) Kaplan–Meier curve of glypican 3 (GPC3); (c) Kaplan–Meier curve of theg spermatid protein (THEG); and (d) Kaplan–Meier curve of cadherin 6 (CDH6).

factor receptor signaling 24 h up, apoptosis by cyclin dependent kinase inhibitor 1A via tumor protein P53 (TP53), GC early up, peroxisome proliferator-activated receptor signaling pathway, hypoxia inducible factor 1 subunit alpha pathway, oncogenic MAPK signaling, apoptosis via nuclear factor kappa-B (NFκB), Myc oncogenic signature, TP53 and tumor protein P63 (TP63) targets, cell cycle checkpoints, mechanistic target of rapamycin kinase (mTOR) 4pathway, cell cycle mitotic, and phosphatidylinositol 3-kinase (PI3K)/Akt signaling pathway (Figure 7a, Table S7). While the low-risk group was notably related to the following mechanisms: B-cell receptor signaling pathway, T-cell receptor pathway, PI3K cascade fibroblast growth factor receptor 2 (FGFR2), phospholipase-c-mediated cascade FGFR2, and interleukin 2 family signaling (Figure 7b, Table S7). GSEA for GO term investigation revealed that GC patients in the high-risk group are signally different from the low-risk group in the following biological processes: ncRNA processing, positive regulation of cell cycle, TLR signaling pathway, B-cell-mediated immunity regulation of receptor signaling pathway via STAT, regulation of Notch signaling pathway, cell–cell junction assembly, regulation of apoptotic signaling pathway, Notch signaling pathway, developmental cell growth, and regulation of cell–cell adhesion (Figure 7c, Table S8). While the low-risk group was significantly associated with the following mechanisms: B cell proliferation, mature B cell differentiation involved in immune response, 3′-UTR-mediated mRNA stabilization, T-cell receptor complex, regulation of B-cell receptor signaling pathway, and negative regulation of interleukin 8 production (Figure 7d, Table S8).

We screen for DEGs between high- and low-risk groups of GC patients, and 878 DEGs were generated, of which 623 were signally upregulated and 256 were

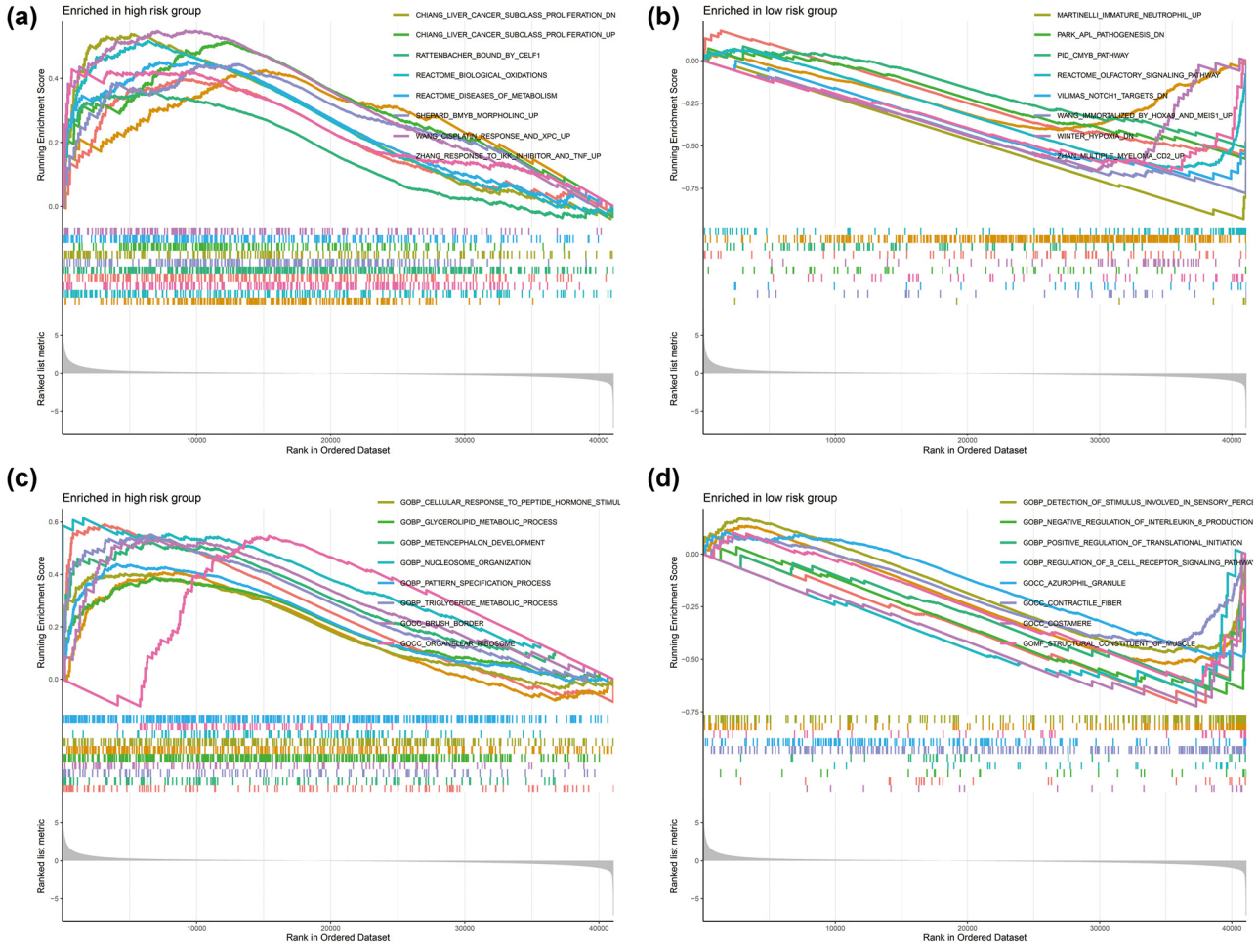

Figure 7: GSEA results between low- and high-risk score phenotypes in GC: (a) GSEA results in high-risk score group using the c2 reference gene set; (b) GSEA results in low-risk score group using the c2 reference gene set; (c) GSEA results in high-risk score group using the c5 reference gene set; and (d) GSEA results in low-risk score group using the c5 reference gene set.

signally downregulated (Figure 8, Figure S2 and Table S9). Survival analysis found that 83 DEGs were significantly associated with GC OS, among which the top three DEGs of significance were chorionic gonadotropin subunit beta 5 (CGB5), chorionic gonadotropin subunit beta 8 (CGB8), and secretin receptor (SCTR) (Figure 9a–d, Table S10). Then, we performed WGCNA and observed that these DEGs can be significantly divided into six modules: the top three modules are turquoise, blue, and brown modules, respectively (Figure 10a–f, Table S11). By analyzing the number of nodes of each DEG in the WGCNA network, we observed that the DEGs of the turquoise module have the highest degree, the highest degree value is 84, and there are three DEGs, namely cornulin, calmodulin-like 3, and cysteine-rich tail 1 (Figure 11, Table S12). These three genes may play the role of hub genes in this WGCNA network, especially in the turquoise module. GO term analysis revealed that these DEGs may be involved in G-protein-coupled receptor signaling pathway, cell differentiation, G-protein-coupled receptor activity, negative regulation of T-cell proliferation, cell–cell signaling, negative regulation of cytokine secretion involved in immune response, negative regulation of endothelial cell apoptotic process, opioid receptor signaling pathway, homophilic cell adhesion via plasma membrane adhesion molecules, negative regulation of nucleic acid-templated transcription, regulation of immune system process, retinoic acid metabolic process, negative regulation of endothelial cell apoptotic process, extracellular negative regulation of signal transduction, transcriptional activator activity, and RNA polymerase II transcription regulatory region sequence-specific binding (Table S13). KEGG suggested that these DEGs participated in the metabolism of xenobiotics by cytochrome P450, chemical carcinogenesis, and drug metabolism – cytochrome P450 (Table S13).

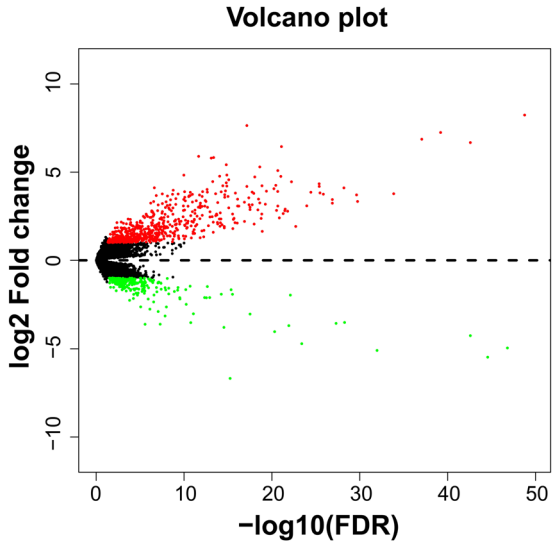

Figure 8: Volcano plot of DEGs between low- and high-risk score phenotypes in GC.

3.3 Immunomicroenvironment and ssGSEA analysis

By analyzing the immune microenvironment in GC tumor tissues, we observed that three immune microenvironment scores were signally downregulated in tumor tissues of GC patients in the high-risk phenotype (Figure 12a–c). Through ssGSEA analysis, we compared the infiltration abundance of 23 immune cells and found that the infiltration abundance of 9 immune cells was notably downregulated in the high-risk score group (Figure 13).

4 Discussion

We all know that non-coding RNAs act an increasingly indispensable role in tumors. As the main component of the post-transcriptional RNA spliceosome, snRNA is involved in the processing of mRNA precursors. Through literature review, we did not find any related reports on snRNA in GC. In this study, we first downloaded the whole-genome RNA sequencing data set from the TCGA website. By extracting snRNAs data set, we compared the expression levels of snRNAs in tumor and adjacent non-tumor tissues. We found that a large number of snRNAs were differential expressions between tumor and adjacent non-tumor tissues. These differentially expressed snRNAs may lead to abnormal processing of mRNA precursors, resulting in dysregulation of GC-related gene transcription, which induces tumorigenesis and progression. Then, based on snRNA prognosis analysis, we obtained 14 snRNAs associated with GC prognosis and constructed a prognostic signature including 9 snRNAs. After a literature search, we found that among these nine snRNAs, only U1 has been reported in previous studies, and the remaining eight are novel and unreported snRNAs. Yin et al. found that U1 can mediate the binding of interacting RNA to chromatin, thereby affecting the function of RNA on chromatin [13]. Spraggon and Cartegni reviewed the physiological role of U1 in inhibiting polyadenylation in cells and suggested that this physiological role might be used in the treatment of cancers [32]. Another study has confirmed that the inhibition of U1 in Hela cells can signally increase the invasion and migration of cancer cells, while the

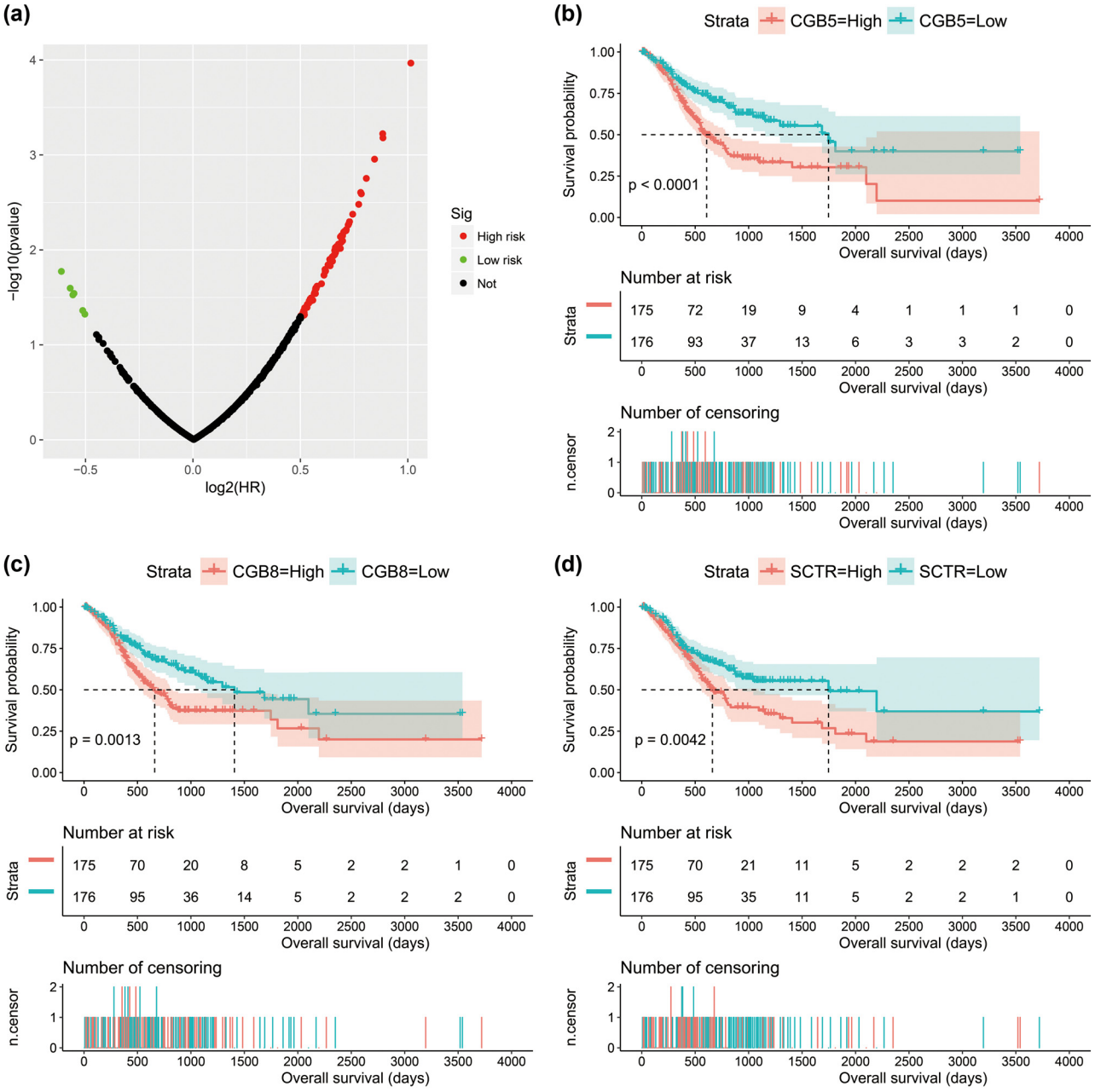

Figure 9: Survival analysis of DEGs between low- and high-risk score phenotypes in GC: (a) volcano plot of survival analysis results; (b) Kaplan–Meier curve of CGB5; (c) Kaplan–Meier curve of CBG8; and (d) Kaplan–Meier curve of SCTR.

opposite phenomenon was observed in Hela cell lines when U1 was overexpressed. At the same time, similar phenomena were observed in lung and breast cancer cell lines [33]. The study by Suzuki et al. found that U1 is frequently mutated in sonic hedgehogs, and the mutated U1 can significantly inhibit the tumor suppressor gene PTCH1 and activate the oncogenes GLI2 and CCND2. Their study concludes that U1 may be a potential therapeutic target for sonic hedgehogs [34].

Rahman et al. verified that U1 was significantly upregulated in Canine Melanoma using small RNA sequencing and quantitative reverse transcription-polymerase chain reaction [35]. Sadik et al. observed that U1 participates in immune regulation through TLR signaling pathway in the A549 lung cancer cell line, which may have anti-inflammatory effects [36]. Modification of U1 can significantly reduce the expression level of the human chorionic gonadotropin beta subunit and induce cervical

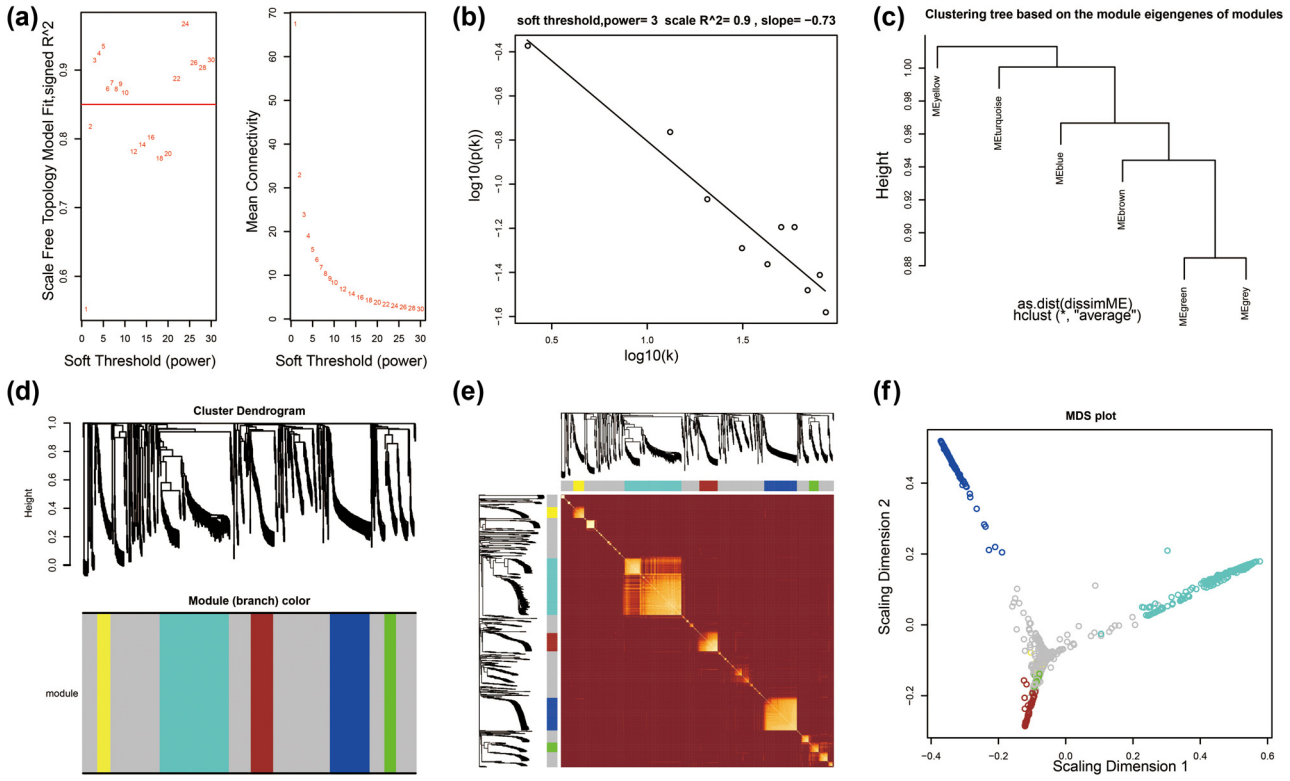

Figure 10: WGCNA analysis results of DEG: (a) soft thresholding; (b) scale-free plot; (c) clustering tree based on the module Eigen genes of modules; (d) cluster dendrogram; (e) topological overlap matrix (TOM) plot; and (f) multidimensional scaling plot.

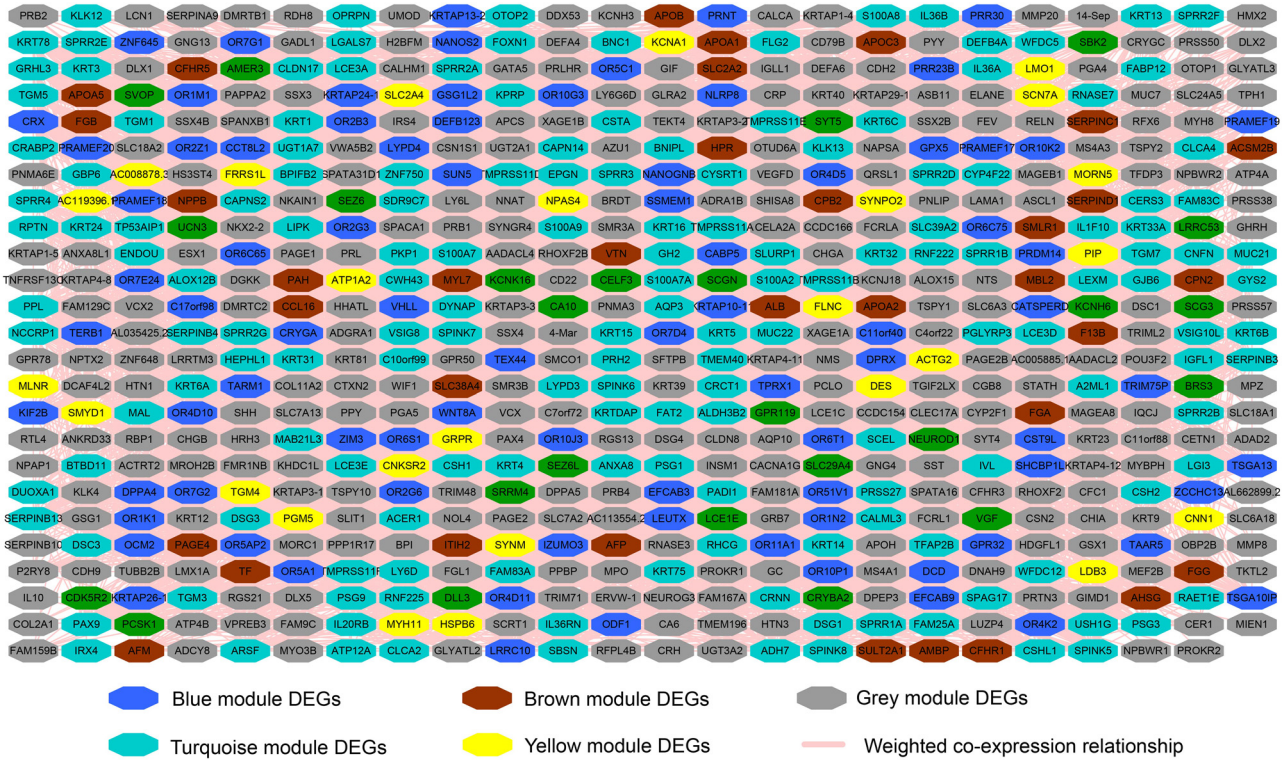

Figure 11: WGCNA interaction network of DEGs.

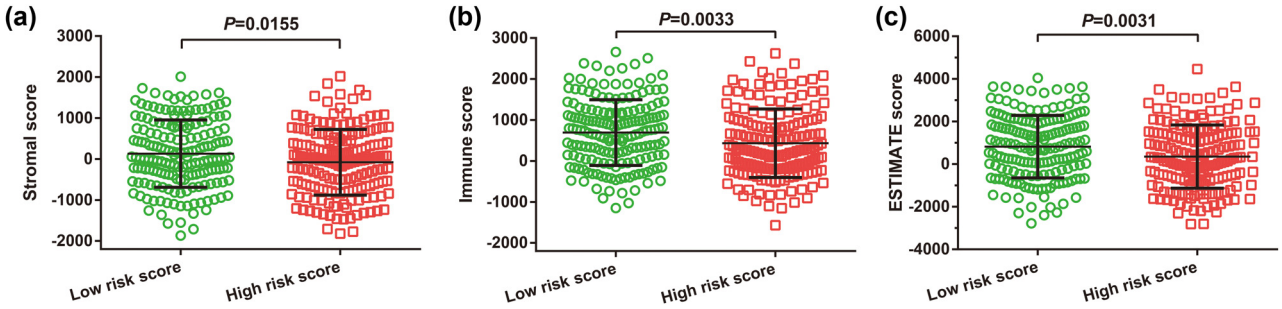

Figure 12: Results of tumor immune microenvironment analysis between low- and high-risk score phenotypes in GC: (a) Stromal score; (b) Immune score; and (c) ESTIMATE score.

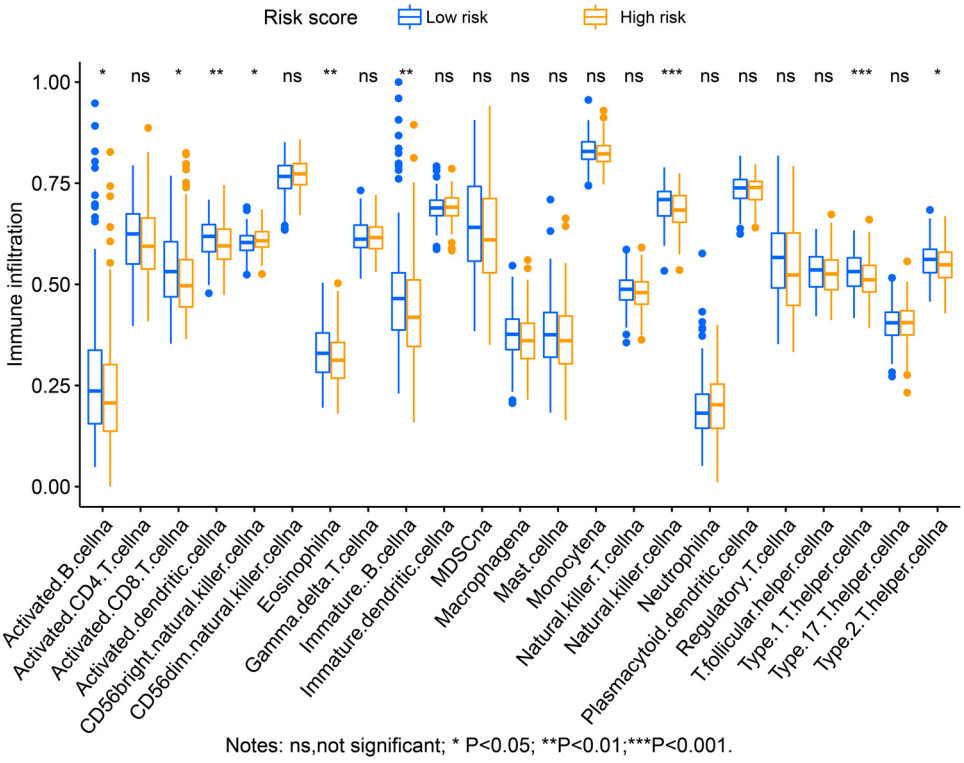

Figure 13: Immune infiltration analysis results of ssGSEA.

cancer cell apoptosis [37]. Dong et al. observed that U1 was signally downregulated in lung cancer patients by analyzing the serum of lung cancer patients, and the ROC curve suggested that U1 may be a potential diagnostic biomarker for lung cancer [38]. High-frequency mutations of the third base of U1 snRNA have been reported in multiple cancers. The mutated U1 changes the splicing connection of the 5′ splicing site, resulting in changes in the splicing pattern of multiple genes, including some known oncogenes, resulting in the occurrence, development, and poor prognosis of cancers [39]. Cheng et al. overexpressed U1 in PC-12 cell line, and functional enrichment of DEGs screened by genome-wide expression microarray can enrich a large number of cancer-related signaling pathways. It was further inferred that U1 may activate a number of cancer-related functional mechanisms in adrenal pheochromocytoma [40]. U1 can also act as an adaptor to mediate drug resistance and regulate downstream gene expression levels in cancers [41–43]. By reviewing the above literature, we found that among the known snRNAs, U1 has been reported to play a variety of functions in cancers and can be used as a biomarker for various cancers. However, it has not been reported in GC. This study is the first to report that U1 can be used as a diagnostic and prognostic biomarker in GC. At the same time, it also explores its molecular mechanism, which provides a theoretical basis for the clinical application and

functional mechanism exploration of U1 in GC. In the current study, we have found that U1 was signally downregulated in GC tumor tissues, which was consistent with previous reports. At the same time, it was previously found that U1 was related to the prognosis of CLL. We also observed that high expression of U1 was signally correlated with unfavorable prognosis in GC.

In terms of functional enrichment analysis, previous studies have reported that the physiological roles of snRNA are involved in alternative splicing, chromatin stabilization, and TLR signaling pathways. We also found that this prognostic signature snRNA could be significantly enriched in alternative splicing and TLR signaling pathways through functional enrichment analysis. In addition, we also enriched classical cancer-related signaling pathways such as Notch and PI3K. These results all suggest that snRNAs play an indispensable function in cancers. Since the previous literature has not conducted in-depth exploration and verification of snRNAs, our results still need to be further verified in the future.

There are some deficiencies in this study that need to be stated. First, all the results of this study are resulting from the functional enrichment analysis of the whole genome, which needs to be verified by further cell and animal experiments. Second, the survival analysis in the present study was a single-cohort analysis and lacked other multicenter validation cohorts. Despite the above shortcomings of this study, our research still includes the prognostic analysis of snRNAs to the functional enrichment analysis of various approaches and preliminarily clarified the clinical significance and potential biological molecular mechanism of snRNAs in GC. These results can provide some theoretical support and a basis for future research.

5 Conclusion

In the present study, we obtained 14 prognostic snRNA markers significantly associated with GC OS and also constructed a prognostic signature containing nine prognostic snRNAs. Both survivalROC and nomogram model suggest that this prognostic signature can be a good indicator for predicting OS in GC patients. Molecular mechanism exploration found that this prognostic signature is involved in classical cancer-related signaling pathways such as Notch, PI3KAKT, TLR signaling pathways, and biological processes such as cell cycle, cell proliferation, and cell adhesion that affect the biological phenotype of cancer cells. Immune-related analyses also suggested significant differences in the tumor immune microenvironment and immune cell infiltration between high- and low-score patients classified according to this prognostic signature. As all the conclusions of this study were generated from the bioinformatics analysis and clinical analysis of whole-genome RNA sequencing data, all the results still need to be further verified in future studies.

Acknowledgments: The authors thank the contributors of The Cancer Genome Atlas for sharing their data set with open access. In addition, we would like to acknowledge the helpful comments on this article received from our reviewers.

Funding information: This work was supported in part by the National Natural Science Foundation of China (Grant Number: 81960119).

Author contributions: Ziyu Liang and Haixing Jiang designed this manuscript. Ziyu Liang, Dongxing Su, Kang Liu, and Haixing Jiang conducted the study, collected, and analyzed the data. Ziyu Liang wrote this manuscript and contributed to the data interpretation, and Haixing Jiang guided the writing.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The raw RNA-seq data set of GC can be obtained from the TCGA data portal (https://portal.gdc.cancer.gov/).

References

[1] Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. Lancet. 2020;396(10251):635–48. doi: 10.1016/S0140-6736(20)31288-5.

[2] Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA A Cancer J Clinicians. 2016;66(2):115–32. doi: 10.3322/caac.21338.

[3] Tan Z. Recent advances in the surgical treatment of advanced gastric cancer: A Review. Med Sci Monitor Int Med J Exp Clin Res. 2019;25:3537–41. doi: 10.12659/MSM.916475.

[4] Sexton RE, Al Hallak MN, Diab M, Azmi AS. Gastric cancer: a comprehensive review of current and future treatment strategies. Cancer Metastasis Rev. 2020;39(4):1179–203. doi: 10.1007/s10555-020-09925-3.

[5] Pasechnikov V, Chukov S, Fedorov E, Kikuste I, Leja M. Gastric cancer: prevention, screening and early diagnosis. World

J Gastroenterol. 2014;20(38):13842–62. doi: 10.3748/wjg.v20.i38.13842.
[6] Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F. Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention. Cancer epidemiology, biomarkers & prevention: a publication of the American association for cancer research, cosponsored by the American society of preventive. Oncology. 2014;23(5):700–13. doi: 10.1158/1055-9965.EPI-13-1057.
[7] Necula L, Matei L, Dragu D, Neagu AI, Mambet C, Nedeianu S, et al. Recent advances in gastric cancer early diagnosis. World J Gastroenterol. 2019;25(17):2029–44. doi: 10.3748/wjg.v25.i17.2029.
[8] Matsuoka T, Yashiro M. Biomarkers of gastric cancer: Current topics and future perspective. World J Gastroenterol. 2018;24(26):2818–32. doi: 10.3748/wjg.v24.i26.2818.
[9] Hombach S, Kretz M. Non-coding RNAs: Classification, biology and functioning. Adv Exp Med Biol. 2016;937:3–17. doi: 10.1007/978-3-319-42059-2_1.
[10] Jobert L, Pinzón N, Van Herreweghe E, Jády BE, Guialis A, Kiss T, et al. Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. EMBO Rep. 2009;10(5):494–500. doi: 10.1038/embor.2009.24.
[11] Wilkinson ME, Charenton C, Nagai K. RNA splicing by the spliceosome. Annu Rev Biochem. 2020;89:359–88. doi: 10.1146/annurev-biochem-091719-064225.
[12] Lee Y, Rio DC. Mechanisms and regulation of alternative Pre-mRNA splicing. Annu Rev Biochem. 2015;84:291–323. doi: 10.1146/annurev-biochem-060614-034316.
[13] Yin Y, Lu JY, Zhang X, Shao W, Xu Y, Li P, et al. U1 snRNP regulates chromatin retention of noncoding RNAs. Nature. 2020;580(7801):147–50. doi: 10.1038/s41586-020-2105-3.
[14] Will CL, Luhrmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011;3(7):a003707. doi: 10.1101/cshperspect.a003707.
[15] Cancer Genome Atlas Research, N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202–9. doi: 10.1038/nature13480.
[16] Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The cancer genome atlas pan-cancer analysis project. Nat Genet. 2013;45(10):1113–20. doi: 10.1038/ng.2764.
[17] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40. doi: 10.1093/bioinformatics/btp616.
[18] Deng T, Gong Y, Liao X, Wang X, Zhou X, Zhu G, et al. Integrative analysis of a novel eleven-small nucleolar RNA prognostic signature in patients with lower grade glioma. Front Oncol. 2021;11:650828. doi: 10.3389/fonc.2021.650828.
[19] Zhang L, Xin M, Wang P. Identification of a novel snoRNA expression signature associated with overall survival in patients with lung adenocarcinoma: A comprehensive analysis based on RNA sequencing dataset. Math Biosci Eng. 2021;18(6):7837–60. doi: 10.3934/mbe.2021389.
[20] Zhao Y, Yan Y, Ma R, Lv X, Zhang L, Wang J, 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. 2020;24(3):2215–28. doi: 10.1111/jcmm.14886.
[21] Xie J, Li H, Chen L, Cao Y, Hu Y, Zhu Z, et al. A novel pyroptosis-related lncRNA signature for predicting the prognosis of skin cutaneous melanoma. Int J Gen Med. 2021;14:6517–27. doi: 10.2147/IJGM.S335396.
[22] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57. doi: 10.1038/nprot.2008.211.
[23] Sherman BT, Huang da W, Tan Q, Guo Y, Bour S, Liu D, et al. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinforma. 2007;8:426. doi: 10.1186/1471-2105-8-426.
[24] Qiu Y, Li H, Xie J, Qiao X, Wu J. Identification of ABCC5 among ATP-binding cassette transporter family as a new biomarker for hepatocellular carcinoma based on bioinformatics analysis. Int J Gen Med. 2021;14:7235–46. doi: 10.2147/IJGM.S333904.
[25] Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov (N Y). 2021;2(3):100141. doi: 10.1016/j.xinn.2021.100141.
[26] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284–7. doi: 10.1089/omi.2011.0118.
[27] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinforma. 2008;9:559. doi: 10.1186/1471-2105-9-559.
[28] Xiao Y, Tu Y, Li Y. Expression level of long non-coding RNA colon adenocarcinoma hypermethylated serves as a novel prognostic biomarker in patients with thyroid carcinoma. Biosci Rep. 2021;41(4):BSR20210284. doi: 10.1042/BSR20210284.
[29] Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612. doi: 10.1038/ncomms3612.
[30] Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462(7269):108–12. doi: 10.1038/nature08460.
[31] Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinforma. 2013;14:7. doi: 10.1186/1471-2105-14-7.
[32] Spraggon L, Cartegni L. U1 snRNP-dependent suppression of polyadenylation: physiological role and therapeutic opportunities in cancer. Int J Cell Biol. 2013;2013:846510. doi: 10.1155/2013/846510.
[33] Oh JM, Venters CC, Di C, Pinto AM, Wan L, Younis I, et al. U1 snRNP regulates cancer cell migration and invasion in vitro. Nat Commun. 2020;11(1):1. doi: 10.1038/s41467-019-13993-7.
[34] Suzuki H, Kumar SA, Shuai S, Diaz-Navarro A, Gutierrez-Fernandez A, De Antonellis P, et al. Recurrent noncoding U1 snRNA mutations drive cryptic splicing in SHH medulloblastoma. Nature. 2019;574(7780):707–11. doi: 10.1038/s41586-019-1650-0.
[35] Rahman MM, Lai YC, Husna AA, Chen HW, Tanaka Y, Kawaguchi H, et al. Aberrantly expressed snoRNA, snRNA, piRNA and tRFs in canine melanoma. Vet Comp Oncol. 2020;18(3):353–61. doi: 10.1111/vco.12558.
[36] Sadik CD, Bachmann M, Pfeilschifter J, Muhl H. Activation of interferon regulatory factor-3 via toll-like receptor 3 and

immunomodulatory functions detected in A549 lung epithelial cells exposed to misplaced U1-snRNA. Nucleic Acids Res. 2009;37(15):5041–56. doi: 10.1093/nar/gkp525.

[37] Jankowska A, Gunderson SI, Andrusiewicz M, Burczynska B, Szczerba A, Jarmolowski A, et al. Reduction of human chorionic gonadotropin beta subunit expression by modified U1 snRNA caused apoptosis in cervical cancer cells. Mol Cancer. 2008;7:26. doi: 10.1186/1476-4598-7-26.

[38] Dong X, Ding S, Yu M, Niu L, Xue L, Zhao Y, et al. Small nuclear RNAs (U1, U2, U5) in tumor-educated platelets are downregulated and act as promising biomarkers in lung cancer. Front Oncol. 2020;10:1627. doi: 10.3389/fonc.2020.01627.

[39] Shuai S, Suzuki H, Diaz-Navarro A, Nadeu F, Kumar SA, Gutierrez-Fernandez A, et al. The U1 spliceosomal RNA is recurrently mutated in multiple cancers. Nature. 2019;574(7780):712–6. doi: 10.1038/s41586-019-1651-z.

[40] Cheng Z, Sun Y, Niu X, Shang Y, Ruan J, Chen Z, et al. Gene expression profiling reveals U1 snRNA regulates cancer gene expression. Oncotarget. 2017;8(68):112867–74. doi: 10.18632/oncotarget.22842.

[41] Goraczniak R, Wall BA, Behlke MA, Lennox KA, Ho ES, Zaphiros NH, et al. U1 adaptor oligonucleotides targeting BCL2 and GRM1 suppress growth of human melanoma xenografts in vivo. Mol Ther Nucleic Acids. 2013;2:e92. doi: 10.1038/mtna.2013.24.

[42] Weirauch U, Grunweller A, Cuellar L, Hartmann RK, Aigner A. U1 adaptors for the therapeutic knockdown of the oncogene pim-1 kinase in glioblastoma. Nucleic Acid Ther. 2013;23(4):264–72. doi: 10.1089/nat.2012.0407.

[43] Tsang AT, Dudgeon C, Yi L, Yu X, Goraczniak R, Donohue K, et al. U1 adaptors suppress the KRAS-MYC oncogenic axis in human pancreatic cancer xenografts. Mol Cancer Therapeutics. 2017;16(8):1445–55. doi: 10.1158/1535-7163.MCT-16-0867.

Appendix

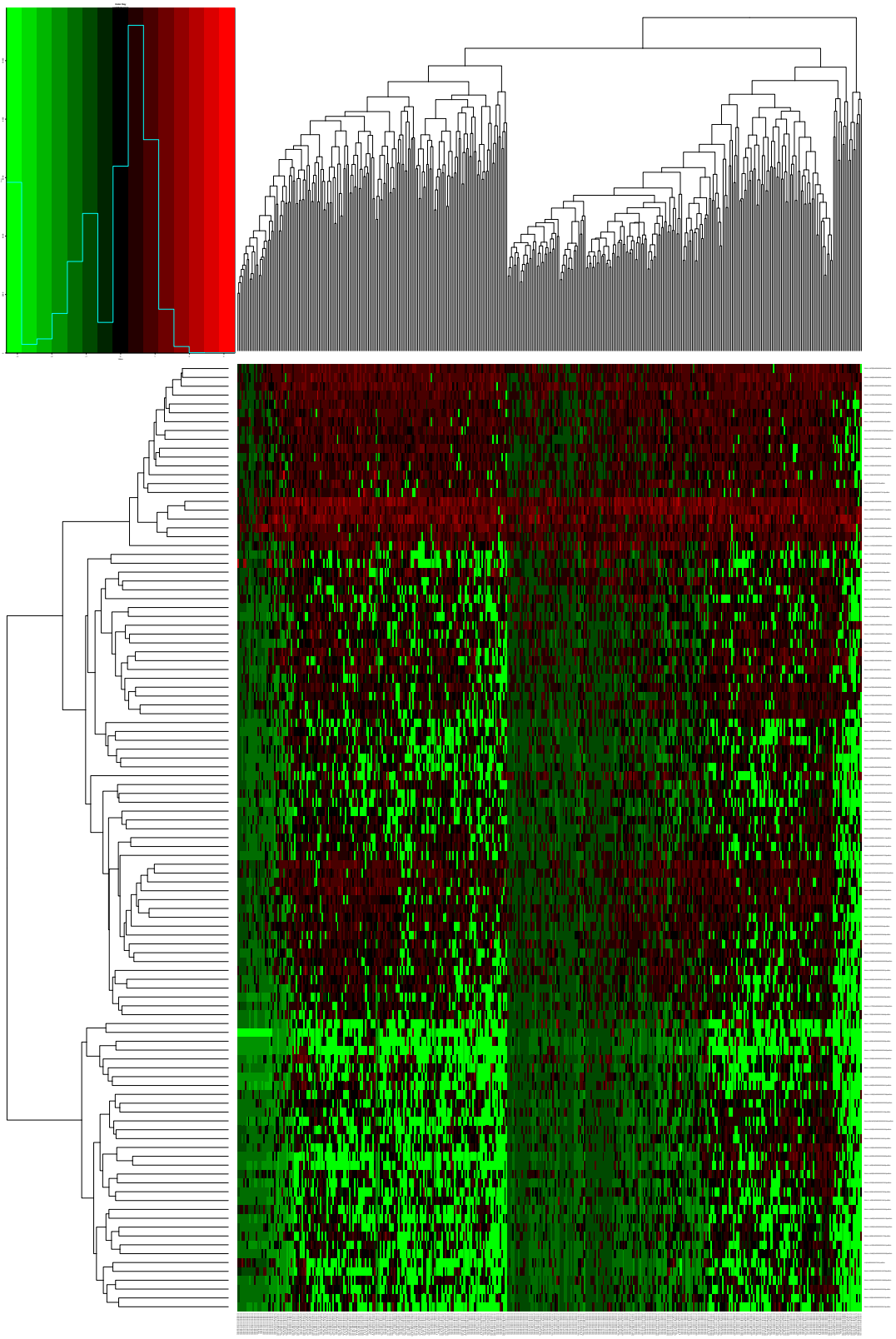

Figure S1: Heat map of differentially expressed snRNAs between tumor and paracancer tissues of gastric cancer.

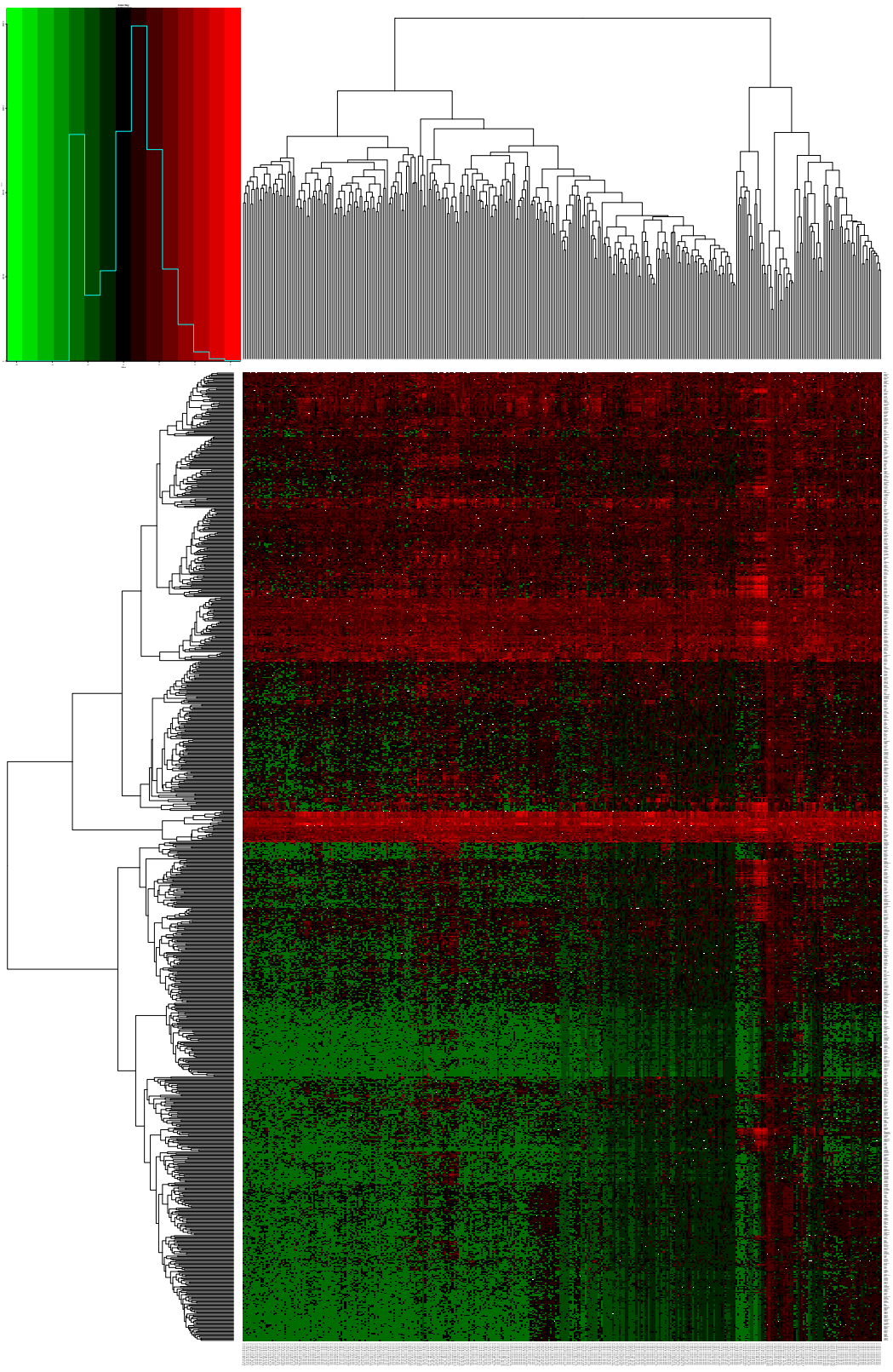

Figure S2: Heat map of DEGs between low- and high-risk score phenotypes of gastric cancer patients.