



OPEN ACCESS

EDITED BY

Erxi Wu,
Baylor Scott and White Health,
United States

REVIEWED BY

Laisheng Li,
The First Affiliated Hospital of Sun
Yat-sen University, China
Jianming Lu,
Guangzhou First People's Hospital,
China
Yongzhen Liu,
Princeton University, United States

*CORRESPONDENCE

Mingwei Chen
✉ chenmw36@163.com

SPECIALTY SECTION

This article was submitted to
Thoracic Oncology,
a section of the journal
Frontiers in Oncology

RECEIVED 22 October 2022

ACCEPTED 12 December 2022

PUBLISHED 22 December 2022

CITATION

Zhang K, Zhang C, Wang K, Teng X
and Chen M (2022) Identifying
diagnostic markers and constructing
a prognostic model for small-cell lung
cancer based on blood
exosome-related genes and
machine-learning methods.
Front. Oncol. 12:1077118.
doi: 10.3389/fonc.2022.1077118

COPYRIGHT

© 2022 Zhang, Zhang, Wang, Teng and
Chen. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which
does not comply with these terms.

Identifying diagnostic markers and constructing a prognostic model for small-cell lung cancer based on blood exosome-related genes and machine-learning methods

Kun Zhang, Chaoguo Zhang, Ke Wang, Xiuli Teng
and Mingwei Chen*

Department of Respiratory and Critical Care Medicine, First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, Shaanxi, China

Background: Small-cell lung cancer (SCLC) usually presents as an extensive disease with a poor prognosis at the time of diagnosis. Exosomes are rich in biological information and have a powerful impact on tumor progression and metastasis. Therefore, this study aimed to screen for diagnostic markers of blood exosomes in SCLC patients and to build a prognostic model.

Methods: We identified blood exosome differentially expressed (DE) RNAs in the exoRBase cohort and identified feature RNAs by the LASSO, Random Forest, and SVM-REF three algorithms. Then, we identified DE genes (DEGs) between SCLC tissues and normal lung tissues in the GEO cohort and obtained exosome-associated DEGs (EDEGs) by intersection with exosomal DEmRNAs. Finally, we performed univariate Cox, LASSO, and multivariate Cox regression analyses on EDEGs to construct the model. We then compared the patients' overall survival (OS) between the two risk groups and assessed the independent prognostic value of the model using receiver operating characteristic (ROC) curve analysis.

Results: We identified 952 DEmRNAs, 210 DElncRNAs, and 190 DEcircRNAs in exosomes and identified 13 feature RNAs with good diagnostic value. Then, we obtained 274 EDEGs and constructed a risk model containing 7 genes (TBX21, ZFH2, HIST2H2BE, LTBP1, SIAE, HIST1H2AL, and TSPAN9). Low-risk patients had a longer OS time than high-risk patients. The risk model can independently predict the prognosis of SCLC patients with the areas under the ROC curve (AUCs) of 0.820 at 1 year, 0.952 at 3 years, and 0.989 at 5 years.

Conclusions: We identified 13 valuable diagnostic markers in the exosomes of SCLC patients and constructed a new promising prognostic model for SCLC.

KEYWORDS

small-cell lung cancer, exosome, machine learning, diagnostic markers, prognostic model

1 Introduction

Small-cell lung cancer (SCLC) constitutes approximately 15% of lung cancers and is characterized by a very high proliferation rate, susceptibility to early metastasis, and poor prognosis (1). Unlike non-small cell lung cancer (NSCLC), the survival rate of which has gradually increased, SCLC remains stable at a low survival rate of 14% to 15%, with a median survival of < 2 years for early stage SCLC patients and 1 year for metastatic patients (2). The prognosis of SCLC is worse because SCLC usually presents as an extensive disease at the time of diagnosis and lacks effective long-term treatment. Therefore, searching for convenient and sensitive diagnostic markers and new prognostic markers for SCLC are essential avenues to improve the outcome of SCLC patients.

Exosomes are broadly defined as secretory vesicles that “may have a physiological function” (3). Some researchers describe exosomes as extracellularly secreted organelles with a diameter of 30 to 200 nm (4). Exosomes have abundant proteins, lipids, and nucleic acids (5). Almost all mammalian cells can secrete exosomes (6), such as adipocytes (7) and immune cells (8, 9). The role of exosomes during tumor development has been extensively studied, and exosomal RNAs, proteins, and metabolites can influence cellular outcomes through signal transduction (10). In some recent studies on NSCLC, exosomes have shown good diagnostic and therapeutic value. For example, tumor-derived exosomal proteins can be used as diagnostic biomarkers for NSCLC (11), and membrane-bound proteins are also promising prognostic biomarkers (12). In addition, exosomes from distinct cells in NSCLC patients exhibit different functions. NSCLC cell-secreted exosomal circUSP7 was found to promote immunosuppression in NSCLC (13), whereas circulating NK cell-derived exosomes

showed antitumor activity (14). However, only a few studies have found that exosomes secreted by SCLC contribute to cancer growth, metastasis, and angiogenesis (15), and some circulating exosomal miRNAs also enhance angiogenesis in SCLC tumors (16). The diagnostic and prognostic role of circulating exosome-related genes in SCLC remains poorly explained.

Thus, this study aimed to screen valuable diagnostic markers from differentially expressed (DE) RNAs in blood exosomes of SCLC by machine-learning methods and to develop a prognostic model associated with exosome-associated differentially expressed genes (EDEGs). Figure 1 shows the study's flow.

2 Materials and methods

2.1 Data collection

We downloaded RNA sequencing (RNAseq) data (mRNA, lncRNA, and circRNA) of extracellular vesicles (mainly exosomes) in blood from 118 healthy individuals and 36 SCLC patients from the exoRBase 2.0 database (<http://www.exorbase.org/>) (17). The RNAseq and clinical data of tumor tissues from SCLC patients were extracted from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/theGEO/>, GSE60052). The GEO cohort included 7 control and 79 SCLC samples, while only 48 SCLC patients had complete follow-up data (Table S1). The baseline information of these two datasets is shown in Table S2.

2.2 Identification of exosomal DERNAs

In the RNAseq samples from SCLC and healthy individuals, duplicate gene data were averaged using the R package “limma”

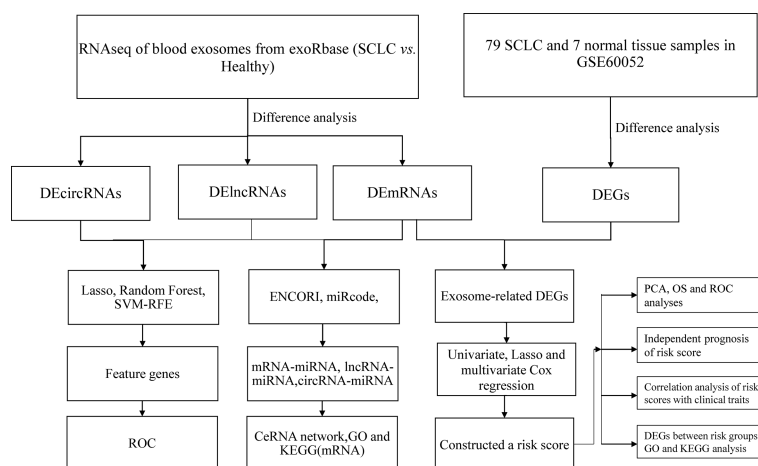


FIGURE 1
The research flow chart.

(18), and the difference analysis was performed using the Wilcoxon test (19). The DEG filtering criteria were $|\log_2FC| > 1$ and false discovery rate (FDR) < 0.05 . All analyses were performed with R software (version 4.2.1).

2.3 Machine learning to identify exosomal feature DERNAs

The least absolute shrinkage and selection operator (LASSO) (20), Random Forest (21), and support vector machine-recursive feature elimination (SVM-RFE) (22) were used to identify feature DEmRNAs, DElncRNAs, and DEcircRNAs of exosomes in blood, respectively. In LASSO, we calculate λ_{\min} and obtain the feature genes corresponding to the minimum point of cross-validation error. In Random Forest, the genes with MeanDecreaseGini values greater than 1 are the feature genes. In SVM-RFE, the genes corresponding to the minimum point of cross-validation error are the feature genes. The intersection of these three algorithms' results is considered the feature DERNAs. We used the R package "pROC" to evaluate the diagnostic values of the feature DERNAs.

2.4 Constructing a competitive endogenous RNA (ceRNA) network

The miRNAs interacting with exosomal DERNAs were predicted by the ENCORI (23) and miRcode databases (<http://www.mircode.org/>). Then, we built the lncRNA-miRNA-mRNA-circRNA network (ceRNA network) (24). The Cytoscape software (v3.7.2) was utilized to visualize the ceRNA network (25). Then, we conducted functional enrichment analysis on DEmRNAs in the ceRNA network.

2.5 Construction of the SCLC prognostic model based on EDEGs

DEGs between 79 SCLC samples and 7 normal tissue samples were intersected with exosomal DEmRNAs to obtain the EDEGs. We initially used univariate Cox (uniCox) regression to assess the correlation of each EDEG with the SCLC patient survival. The p -value < 0.15 was the prognosis-related EDEG, which was then included in the LASSO Cox regression analysis. Finally, we conducted a multivariate Cox (multiCox) regression of the EDEGs obtained from LASSO Cox and built a risk model (Risk Score = $\sum G_i \times C_i$ (G: gene expression, C: coefficients)). We divided the SCLC patients into two risk groups based on the median risk score. Survival curves were generated using the Kaplan-Meier method for the low-risk and high-risk groups. The "prcomp" function was applied to

conduct the principal component analysis (PCA). In addition, we utilized the "survival", "timeROC", and "survminer" R packages to conduct the ROC curve analysis.

2.6 Independent prognostic analysis

We combined clinical features and risk score data of SCLC patients and used uniCox and multiCox regression to analyze the independent prognosis of the risk model. In addition, we assessed the differences in risk scores between patients with different clinical characteristics, including age, sex, smoking status, Stage, T stage, and N stage.

2.7 Functional enrichment analysis

We performed a differential expression analysis of RNAseq between the two risk groups. The DEG screening criteria were $|\log_2FC| > 1$ and p value < 0.05 . Then, GO and KEGG analyses and visualization of the DEGs were performed using the R package "ClusterProfiler".

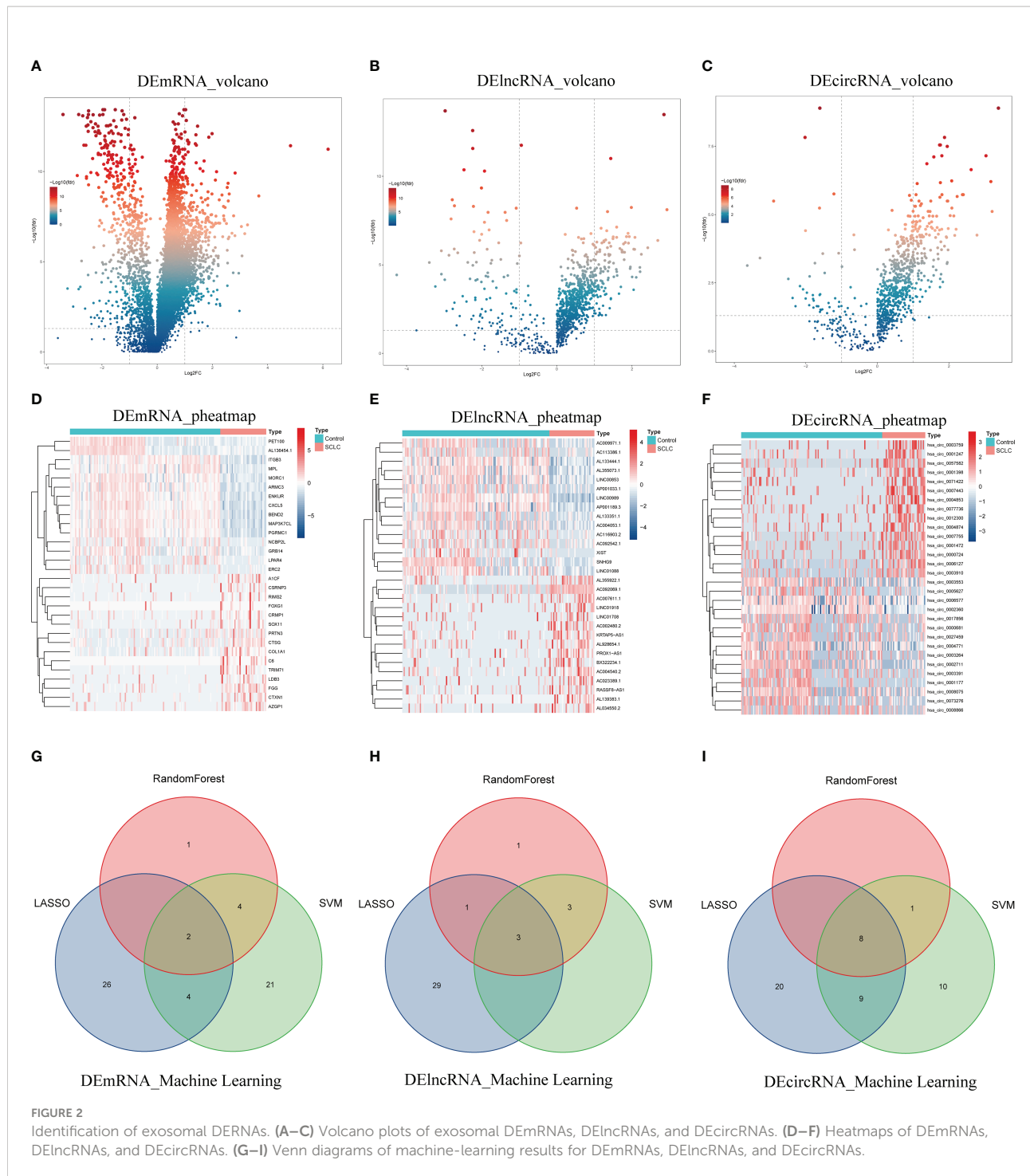
3 Results

3.1 Identification of exosomal DERNAs and feature DERNAs

We first identified 952 DEmRNAs (Figure 2A), 210 DElncRNAs (Figure 2B), and 190 DEcircRNAs (Figure 2C) of blood exosome RNAseq between 36 SCLC patients and 118 healthy individuals from the exoRBase database. The top 30 DEmRNAs, DElncRNAs, and DEcircRNAs are shown as heatmaps (Figures 2D–F). Three machine learning algorithms, including LASSO, Random Forest, and SVM-REF, were further applied to identify the most valuable 2 DEmRNAs (HIST1H1E, ID2, Figure 2G), 3 DElncRNAs (AP000547.3, AC092069.1, AC022150.4, Figure 2H) and 8 DEcircRNAs (hsa_circ_0001953, hsa_circ_0002360, hsa_circ_0007443, hsa_circ_0007637, hsa_circ_0005615, hsa_circ_0005455, hsa_circ_0001258, hsa_circ_0000437, Figure 2I) from the DERNAs obtained above (Table 1, Figure S1, Table S3).

3.2 Diagnostic value of the feature DERNAs

To assess the diagnostic value of exosomal feature DEmRNAs, DElncRNAs, and DEcircRNAs, we conducted the ROC curve analysis. The areas under the ROC curve (AUCs) of ID2 and HIST1H1E in DEmRNA were 0.909 (Figure 3A) and



0.970 (Figure 3B), respectively. The AUCs of AC022150.4, AC092069.1, and AP000547.3 in DElncRNA were 0.915 (Figure 3C), 0.928 (Figure 3D), and 0.925 (Figure 3E), respectively. The AUC of hsa_circ_0000437 in DEcircRNA was the maximum (0.885, Figure 3F), and the AUCs of the remaining feature DEcircRNAs are shown in Figure S2.

3.3 Constructing a ceRNA Network

Meanwhile, we predicted miRNAs that may bind to DEmRNAs, DElncRNAs, and DEcircRNAs and constructed a ceRNA network (Figure 4A). The ceRNA network included 198 mRNAs, 21 lncRNAs, 134 miRNAs, and 32 circRNAs (Table S4).

TABLE 1 Feature DERNAs screened by machine learning.

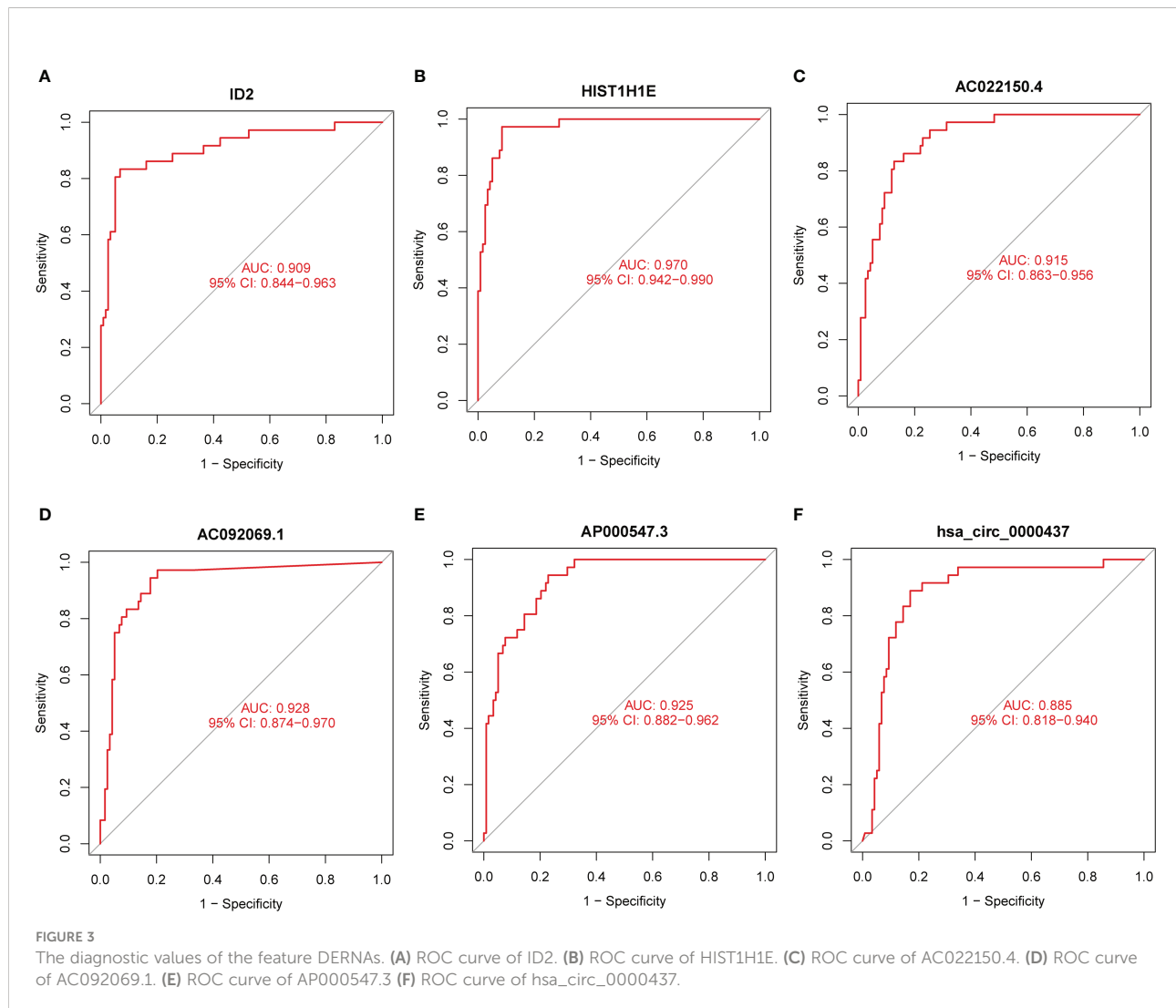
RNA	LASSO, Random Forest and SVM-RFE
mRNA	HIST1H1E, ID2
lncRNA	AP000547.3, AC092069.1, AC022150.4
circRNA	hsa_circ_0001953, hsa_circ_0002360, hsa_circ_0007443, hsa_circ_0007637, hsa_circ_0005615, hsa_circ_0005455, hsa_circ_0001258, hsa_circ_0000437
DE, differentially expressed.	

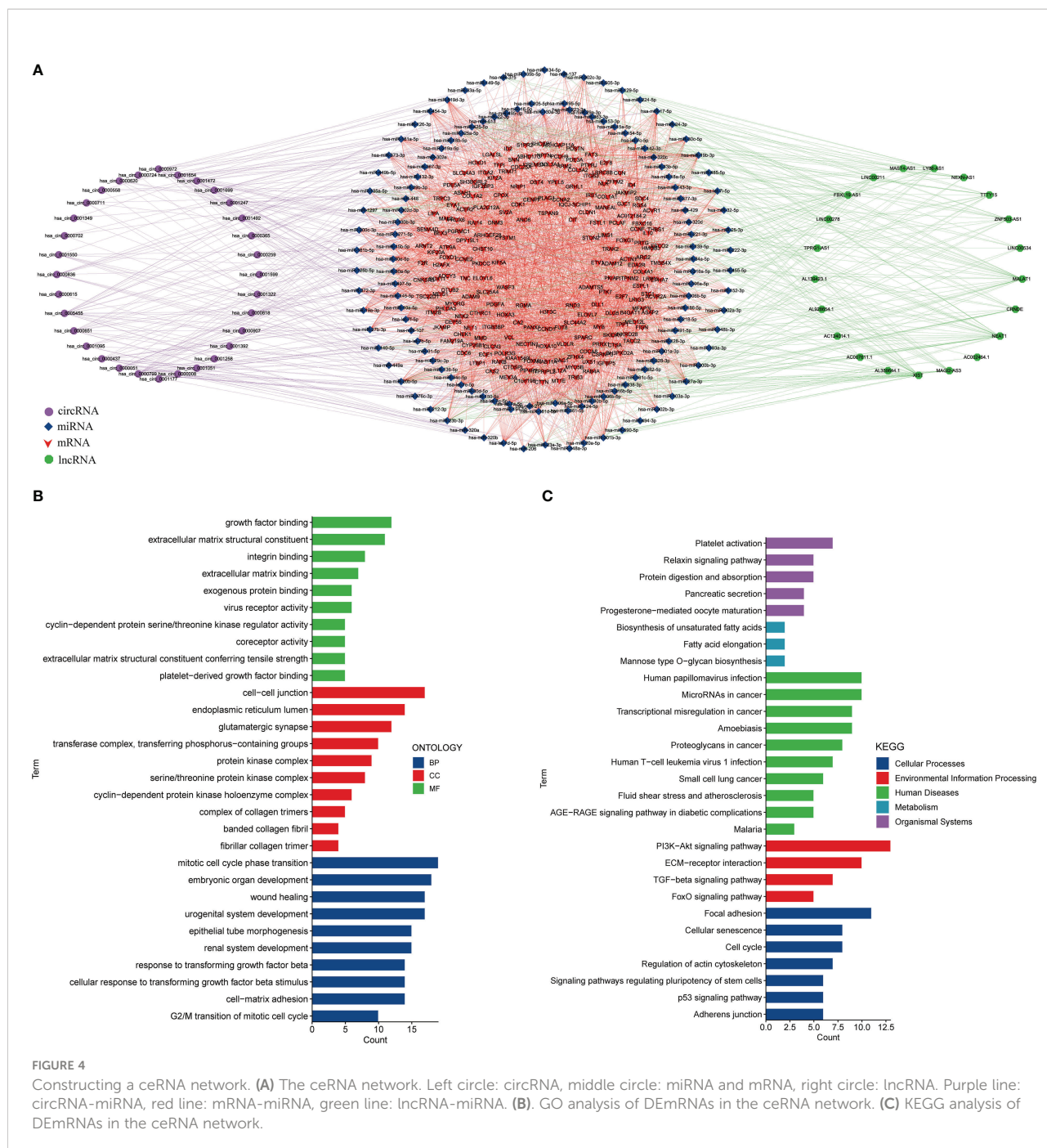
There were 2 miRNAs interacting with ID2, 19 miRNAs interacting with hsa_circ_0005455, 2 miRNAs interacting with hsa_circ_0001258, and 14 miRNAs interacting with hsa_circ_0000437 (Table S5). Then, we performed the functional enrichment analysis of the DEmRNAs in the ceRNA network. The top terms in the biological process (BP) category were growth factor binding and extracellular matrix structural constituent. The top

terms in the cell component (CC) category were endoplasmic reticulum lumen and cell-cell junction. In molecular function (MF), these genes were primarily abundant in mitotic cell cycle phase transition and embryonic organ development (Figure 4B). The KEGG analysis showed that these DEmRNAs were associated with the PI3K-Akt signaling pathway, ECM-receptor interaction, and TGF-beta signaling pathway (Figure 4C).

3.4 Construction of SCLC prognostic model based on EDEGs

We performed the differential expression analysis on samples from the GSE60052 cohort, yielding 4350 DEGs (Figures 5A, B). These DEGs intersected with 952 exosomal DEmRNAs to obtain 274 EDEGs (Figure 5C). We then merged these EDEGs expression data with the clinical follow-up data of 48 SCLC patients. Table 2 presents the clinical data of 48 SCLC patients. Based on the uniCox





regression, we obtained 28 prognosis-associated EDEGs (Figure 5D). The 28 prognosis-associated EDEGs were then included in the LASSO Cox regression analysis, and 13 representative genes were further screened (TBX21, ZFH2, MT1E, HIST1H2AB, HIST2H2BE, FAM83F, MMRN1, LTBP1, CCND1, SIAE, HIST1H2AL, DPY19L2, and TSPAN9) (Figures 5E, F). Finally, we conducted the multiCox regression analysis on the above 13 genes to establish a prognostic model consisting of 7 genes (TBX21, ZFH2, HIST2H2BE, LTBP1, SIAE, HIST1H2AL, and

TSPAN9) (Figure 5G, Table 3). Here is the formula: risk score = (exp. TBX21* -0.56) + (exp. ZFH2* 1.05) + (exp. E2F7*0.340957721) + (exp. HIST2H2BE* -0.58) + (exp. LTBP1* -0.54) + (exp. SIAE* 0.69) + (exp. HIST1H2AL* -0.61) + (exp. TSPAN9* 0.86). The median risk score assigned these 48 SCLC patients well into two different risk groups, as confirmed by the PCA analysis (Figure 5H). Patients with high-risk scores had higher mortality and shorter survival times than those with low-risk scores (Figures 5I–K). Additionally, we utilized the ROC curve analysis to

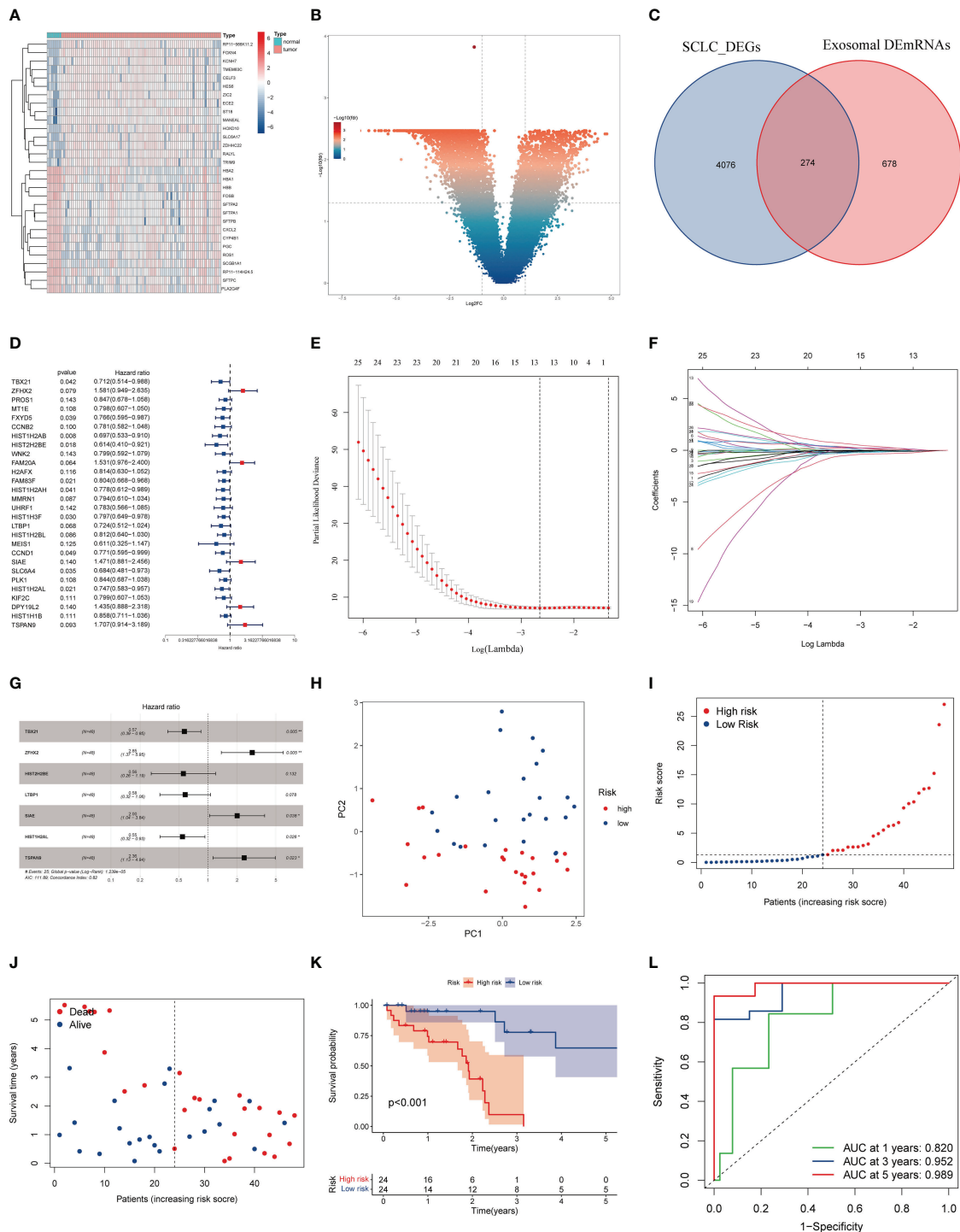


FIGURE 5 Construction of the SCLC prognostic model based on EDEGs. **(A)** Heatmap of the top 30 DEGs in the GEO cohort. **(B)** Volcano plot of DEGs in the GEO cohort. **(C)** The intersection of exosomal DE mRNAs with DEGs of the GEO cohort. **(D)** Univariate Cox regression analysis. **(E)** LASSO regression of survival-associated EDEGs **(F)** Cross-validation for LASSO regression. **(G)** Multivariate Cox regression analysis. **(H)** PCA plot. **(I, J)** The median risk score and the distribution of survival time. **(K)** Kaplan-Meier analysis. The numbers in the table represent the numbers of surviving patients. **(L)** ROC curve of the model.

TABLE 2 Baseline characteristics of SCLC patients in GSE60052.

	ALL (N=48)	Alive (N=23)	Dead (N=25)
Age, years			
>65	11 (22.9%)	3 (13.0%)	8 (32.0%)
<=65	37 (77.1%)	20 (87.0%)	17 (68.0%)
Sex			
Female	5 (10.4%)	3 (13.0%)	2 (8.00%)
Male	43 (89.6%)	20 (87.0%)	23 (92.0%)
Smoke			
Smoker	33 (68.8%)	17 (73.9%)	16 (64.0%)
No-smoker	15 (31.2%)	6 (26.1%)	9 (36.0%)
Stage			
I-II	16 (33.3%)	4 (17.4%)	12 (48.0%)
III-IV	32 (66.7%)	19 (82.6%)	13 (52.0%)
T			
T1-T2	36 (75.0%)	14 (60.9%)	22 (88.0%)
T3-T4	11 (22.9%)	8 (34.8%)	3 (12.0%)
Unknow	1 (2.08%)	1 (4.35%)	0 (0.00%)
N			
N0	11 (22.9%)	3 (13.0%)	8 (32.0%)
N1-N3	36 (75.0%)	19 (82.6%)	17 (68.0%)
Unknow	1 (2.08%)	1 (4.35%)	0 (0.00%)
M			
M0	46 (95.8%)	22 (95.7%)	24 (96.0%)
M1	1 (2.08%)	0 (0.00%)	1 (4.00%)
Unknow	1 (2.08%)	1 (4.35%)	0 (0.00%)
SCLC, small cell lung cancer.			

TABLE 3 Multivariate Cox proportional hazards regression analysis.

Gene	coef	HR	HR.95L	HR.95H	pValue
TBX21	-0.56022	0.571084	0.385641	0.8457	0.005165
ZFHX2	1.048958	2.854674	1.369668	5.949737	0.005118
HIST2H2BE	-0.58483	0.557201	0.260182	1.19329	0.132283
LTBP1	-0.54064	0.582376	0.319466	1.061652	0.077617
SIAE	0.691107	1.995924	1.038662	3.835425	0.038099
HIST1H2AL	-0.60605	0.545503	0.32045	0.928612	0.02556
TSPAN9	0.858402	2.359387	1.127123	4.938864	0.022759

assess the specificity and sensitivity of the risk model, which showed that the AUC was 0.820 at 1 year, 0.952 at 3 years, and 0.989 at 5 years (Figure 5L).

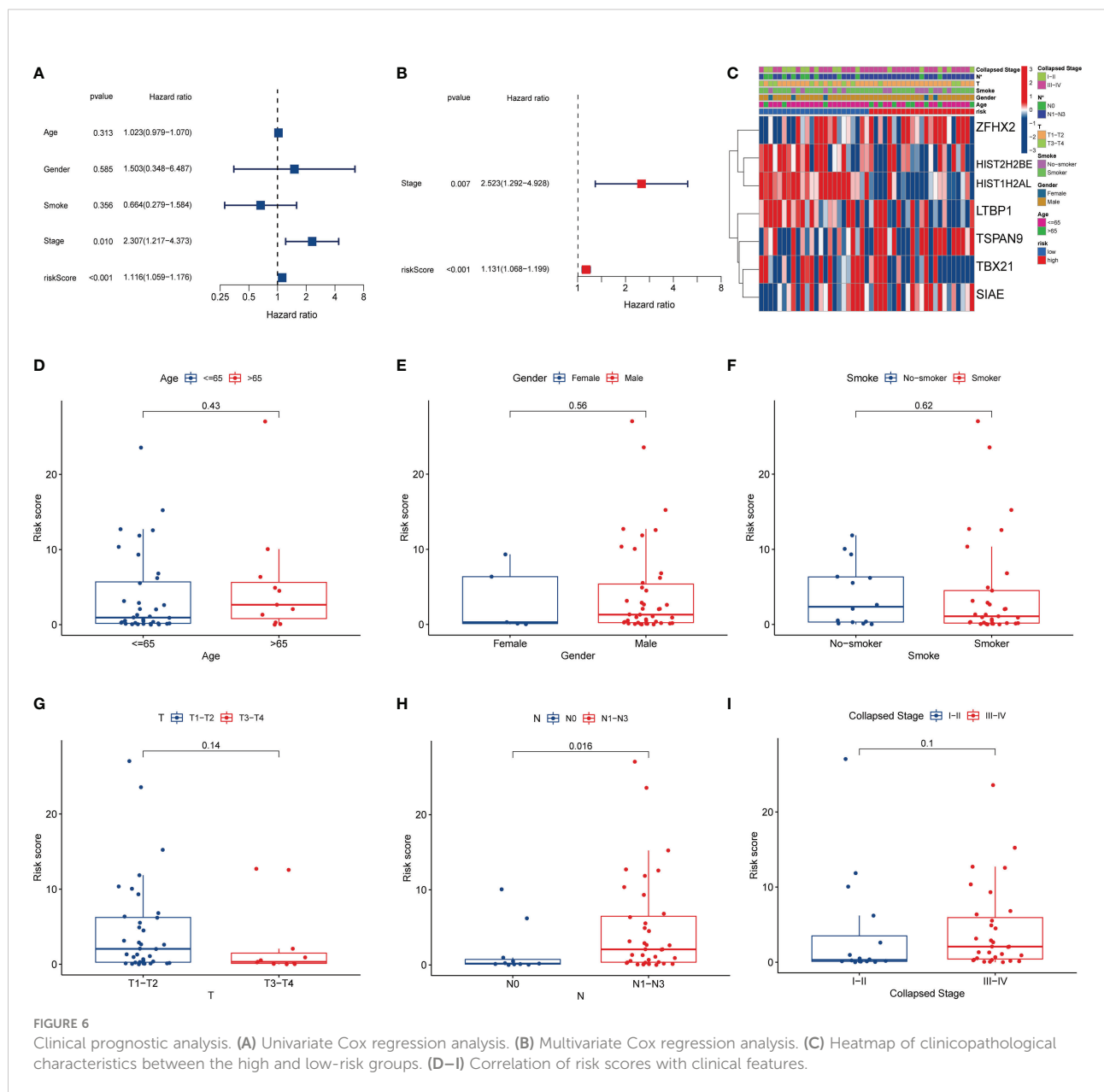
3.5 Clinical prognostic analysis

We evaluated the independent prognostic value of this risk model using uniCox and multiCox regression analyses. The uniCox regression indicated that a higher risk score was related to a lower survival rate (HR=1.116, 95% CI=1.059-1.176, Figure 6A). The multiCox regression revealed that this risk model was an independent prognostic indicator for SCLC patients after

adjusting for other confounding variables (HR=1.131, 95% CI=1.068-1.199, Figure 6B). Moreover, we evaluated the correlation between risk models and clinical features (Figures 6C–I). The results showed that N1-N3 stage patients scored higher risk than N0 stage patients, suggesting that this risk model was mainly associated with lymph node metastasis (Figure 6H).

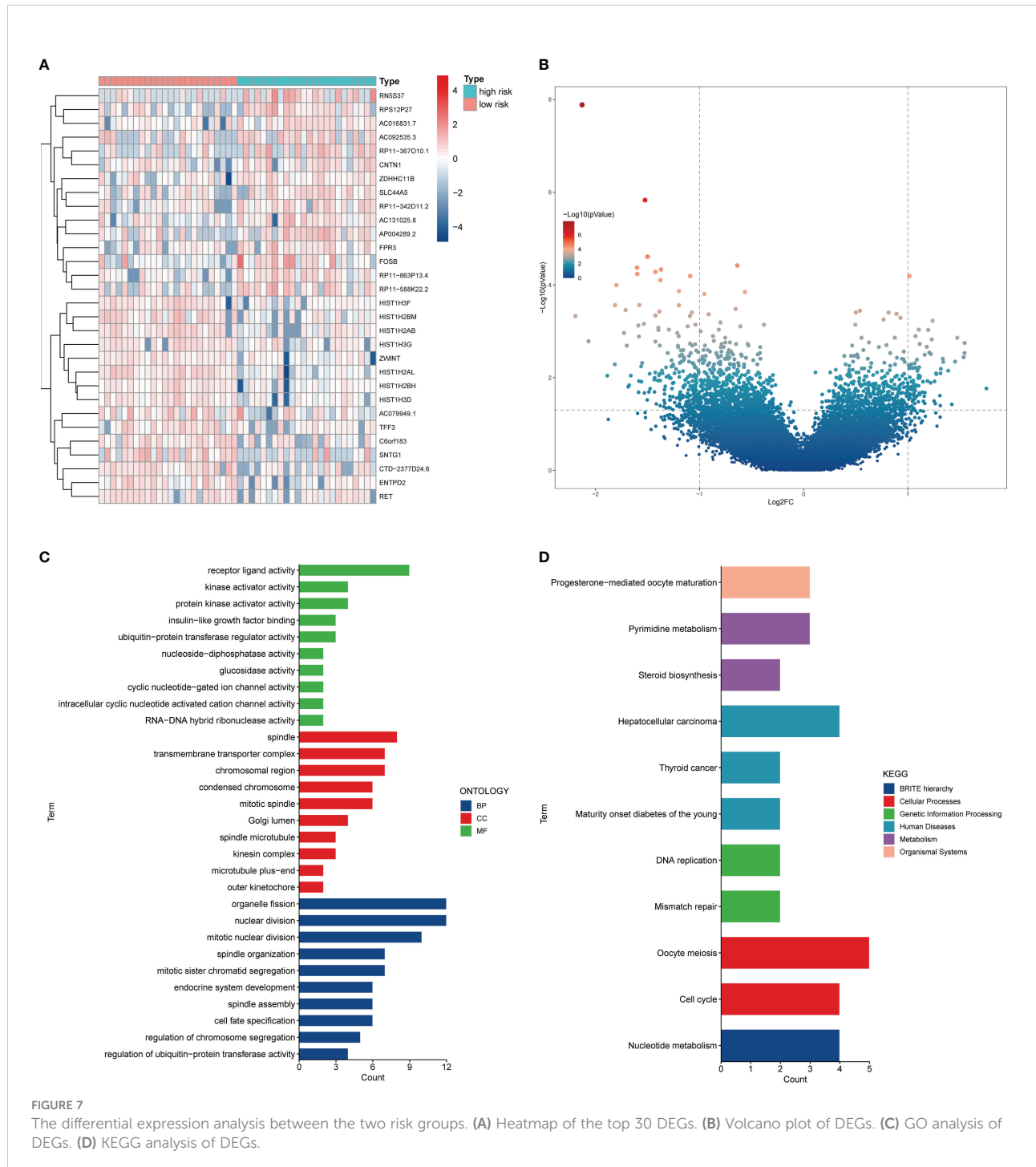
3.6 Functional enrichment analysis of DEGs between the two risk groups

Subsequently, we performed the differential expression analysis of RNAseq in SCLC patients between the two risk



groups and the functional enrichment analysis of the DEGs. A total of 89 upregulated DEGs and 226 downregulated DEGs were obtained (Figures 7A, B). The GO analysis showed that the top BP terms were receptor ligand activity and kinase activator activity, and the top CC terms were spindle and transmembrane

transporter complex. In addition, the top MF terms were organelle fission and nuclear division (Figure 7C). Meanwhile, the KEGG analysis revealed that these DEGs were associated with the cell cycle, nucleotide metabolism, pyrimidine metabolism, and DNA replication (Figure 7D).



4 Discussion

As the most aggressive lung cancer type, SCLC has a poor prognosis, and most patients develop metastases at diagnosis (26). Exosomes are a class of secretory vesicles rich in proteins and nucleic acids. Exosomes have a crucial role in immunity, tissue homeostasis, and cancer (4, 27, 28). Several studies have found that exosomes are associated with growth, metastatic angiogenesis, and drug resistance in SCLC (15, 29, 30). However, more research on the role of exosomes in SCLC is needed.

In this study, we initially identified 952 DEmRNAs, 210 DElncRNAs, and 190 DEcircRNAs from blood exosomes and constructed a ceRNA network. Then, we identified the blood exosomal feature DERNAs from SCLC patients by machine learning, and these feature DERNAs showed good diagnostic value. Meanwhile, we extracted EDEGs of SCLC lung tissues by exosomal DEmRNAs. We used EDEGs to construct a prognostic model consisting of 7 genes. The model was mainly associated with lymph node metastasis, and the death risk of SCLC patients increased as the risk score increased. In addition, the model-based prognosis prediction of SCLC exhibited satisfactory efficiency.

Machine-learning algorithms have been increasingly used for automated diagnosis and prognosis prediction in precision medicine (31, 32). Based on 3 machine-learning methods, we identified blood exosomal feature DERNAs of diagnostic value in SCLC patients, including HIST1H1E, ID2, AP000547.3, AC092069.1, AC022150.4, hsa_circ_0001953, hsa_circ_0002360, hsa_circ_0007443, hsa_circ_0007637, hsa_circ_0005615, hsa_circ_0005455, hsa_circ_0001258, and hsa_circ_0000437. These RNAs can distinguish SCLC well from healthy individuals with AUC values from 0.781 to 0.970. HIST1H1E is a tumor suppressor whose overexpression inhibits lung cancer cell viability, migration, and invasion (33). ID2 is a transcription factor that is overexpressed in many cancers, such as prostate, breast, and gastric cancers (34, 35). We found that HIST1H1E and ID2 are overexpressed in the SCLC patient blood exosomes. However, the diagnostic value of HIST1H1E and ID2 in SCLC has not been reported. Of the three lncRNAs, only one study reported that AC022150.4 was associated with breast cancer prognosis, but their biological functions in cancer have not been investigated (36). The high abundance, stability, and conservation of circRNAs make them advantageous as diagnostic markers (37). It has been reported that hsa_circ_0002360 is highly expressed in NSCLC tissues and contributes to the malignant behavior of NSCLC (38–40). Circulating hsa_circ_0001953 can be used as a marker for proliferative diabetic retinopathy and active tuberculosis, which have not been investigated in cancer (41, 42). In addition, hsa_circ_0007637, hsa_circ_0005615, hsa_circ_0001258, and hsa_circ_0000437 are associated with the diagnosis and development of nasopharyngeal carcinoma, colorectal cancer, osteosarcoma, and hepatocellular carcinoma (43–46). However, hsa_circ_0005455 and hsa_circ_0007443 have not been reported. Our study provides blood exosomal RNAs with diagnostic value for

SCLC patients, which deserves further in-depth study. In addition, we constructed a ceRNA network of exosomal DERNAs, in which ID2 is the target gene of hsa-miR-19a-3p and hsa-miR-19b-3p. Reportedly, hsa-miR-19a-3p can predict the prognosis of lung adenocarcinoma (47), while hsa-miR-19b-3p can be used as a diagnostic and prognostic biomarker for prostate cancer (48). Unfortunately, there are few studies on the role of exosome-derived miRNAs in SCLC.

Our study constructed a model containing 7 prognosis-related genes (TBX21, ZFH2, HIST2H2BE, LTBP1, SIAE, HIST1H2AL, and TSPAN9) and found it to be a good predictor of OS in SCLC patients. TBX21 is a protein-coding gene specifically expressed in immune cells and highly expressed in tissues such as blood and lung (49). In lung adenocarcinoma, TBX21 expression enhances tumor cell recognition and clearance by the immune system (50, 51), while increased TBX21 expression in cutaneous melanoma is associated with a better prognosis (49). In addition, TBX21 expression in T cells was demonstrated in mouse experiments to control tumor progression and antimetastasis (52). In our study, TBX21 was a protective factor in SCLC, which is supported by previous studies. ZFH2 encodes a transcription factor with an unknown function that has not been studied in tumors, and we found that its overexpression predicts a poor prognosis for SCLC patients. One study found that high HIST2H2BE levels predicted low survival in NSCLC patients (53). However, we found the opposite result in SCLC patients, with the possible explanation that HIST2H2BE may be tumor-dependent in its molecular status and role. LTBP1 promotes the proliferation of lung adenocarcinoma cells (54), while its high expression in hepatocellular carcinoma inhibits cancer progression (55). The present study identified LTBP1 as a possible suppressor of SCLC. SIAE encodes an enzyme that removes the 9-O-acetylation modification from sialic acid. The SIAE gene regulates Siglec binding in lung cell lines, and NK-mediated cytotoxicity is increased in lung adenocarcinoma cells without SIAE. In other words, high expression of SIAE promotes lung adenocarcinoma progression. Our study found that SIAE is a risk factor for SCLC (56). HIST1H2AL encodes a core histone protein, which was identified in this study with a better prognosis of SCLC (57). However, its specific role in tumors remains unclear and needs further investigation. TSPANs are a family of four transmembrane segments of proteins that may promote tumor growth by affecting angiogenesis and immune function (58, 59). Among them, TSPAN9 has been shown to promote osteosarcoma metastasis (60) and affect the prognosis of breast cancer patients (61). Moreover, we found that high expression of TSPAN9 increases the risk of death in SCLC patients. Overall, the seven genes in our model have yet to be well studied in SCLC and are new prognostic markers for SCLC.

Our study has several advantages. First, multiple machine-learning approaches were used to identify valuable diagnostic markers in the blood exosomes of SCLC patients. Second, we utilized differential genes in blood exosomes to construct a new prognostic model for SCLC with satisfactory predictive effects. This

study also has some limitations. The sample size of our SCLC patients was small, and the model's accuracy needs to be validated in prospective studies with large samples. In addition, *in vitro* and *in vivo* experiments are needed to validate our results.

5 Conclusion

In summary, we identified valuable diagnostic markers in the exosomes of SCLC patients by machine-learning methods and constructed a novel promising prognostic model for SCLC using exosome-related genes.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

KZ and MC performed study concept and design; KZ, CZ, and MC performed development of methodology and writing; KZ, KW, and XT provided acquisition and analysis of data. All authors read and approved the final paper.

Funding

This study was funded by the Key Research and Development Project of Shaanxi Province (No. 2021ZDLSF02-05).

References

- Rudin CM, Brambilla E, Faivre-Finn C, Sage J. Small-cell lung cancer. *Nat Rev Dis Primers* (2021) 7(1):3. doi: 10.1038/s41572-020-00235-0
- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
- Gould SJ, Raposo G. As we wait: Coping with an imperfect nomenclature for extracellular vesicles. *J Extracell Vesicles* (2013) 2:20389. doi: 10.3402/jev.v2i0.20389
- Pegtel DM, Gould SJ. Exosomes. *Annu Rev Biochem* (2019) 88:487–514. doi: 10.1146/annurev-biochem-013118-111902
- Mashouri L, Yousefi H, Aref AR, Ahadi AM, Molaei F, Alahari SK. Exosomes: Composition, biogenesis, and mechanisms in cancer metastasis and drug resistance. *Mol Cancer* (2019) 18(1):75. doi: 10.1186/s12943-019-0991-5
- Kowal J, Tkach M, Théry C. Biogenesis and secretion of exosomes. *Curr Opin Cell Biol* (2014) 29:116–25. doi: 10.1016/j.ccb.2014.05.004
- Thomou T, Mori MA, Dreyfuss JM, Konishi M, Sakaguchi M, Wolfrum C, et al. Adipose-derived circulating miRNAs regulate gene expression in other tissues. *Nature* (2017) 542(7642):450–5. doi: 10.1038/nature21365
- Raposo G, Nijman HW, Stoorvogel W, Liejendekker R, Harding CV, Melief CJ, et al. B lymphocytes secrete antigen-presenting vesicles. *J Exp Med* (1996) 183(3):1161–72. doi: 10.1084/jem.183.3.1161

Acknowledgments

The authors would like to acknowledge GEO and exoRBase for providing the primary data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.1077118/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

The results of LASSO, Random Forest, and SVM-REF.

SUPPLEMENTARY FIGURE 2

The AUCs of the feature DEcircRNAs.

- Zitvogel L, Regnault A, Lozier A, Wolfers J, Flament C, Tenza D, et al. Eradication of established murine tumors using a novel cell-free vaccine: Dendritic cell-derived exosomes. *Nat Med* (1998) 4(5):594–600. doi: 10.1038/nm0598-594
- Zhang L, Yu D. Exosomes in cancer development, metastasis, and immunity. *Biochim Biophys Acta Rev Cancer* (2019) 1871(2):455–68. doi: 10.1016/j.bbcan.2019.04.004
- Niu L, Song X, Wang N, Xue L, Song X, Xie L. Tumor-derived exosomal proteins as diagnostic biomarkers in non-small cell lung cancer. *Cancer Sci* (2019) 110(1):433–42. doi: 10.1111/cas.13862
- Sandfeld-Paulsen B, Aggerholm-Pedersen N, Bæk R, Jakobsen KR, Meldgaard P, Folkersen BH, et al. Exosomal proteins as prognostic biomarkers in non-small cell lung cancer. *Mol Oncol* (2016) 10(10):1595–602. doi: 10.1016/j.molonc.2016.10.003
- Chen SW, Zhu SQ, Pei X, Qiu BQ, Xiong D, Long X, et al. Cancer cell-derived exosomal Circusp7 induces Cd8(+) T cell dysfunction and anti-Pd1 resistance by regulating the mir-934/Shp2 axis in nscl. *Mol Cancer* (2021) 20(1):144. doi: 10.1186/s12943-021-01448-x
- Kang YT, Niu Z, Hadlock T, Purcell E, Lo TW, Zeinali M, et al. On-chip biogenesis of circulating nk cell-derived exosomes in non-small cell lung cancer exhibits antitumoral activity. *Adv Sci (Weinh)* (2021) 8(6):2003747. doi: 10.1002/advs.202003747

15. Cao Q, Liu Y, Wu Y, Hu C, Sun L, Wang J, et al. Profilin 2 promotes growth, metastasis, and angiogenesis of small cell lung cancer through cancer-derived exosomes. *Aging (Albany NY)* (2020) 12(24):25981–99. doi: 10.18632/aging.202213
16. Mao S, Lu Z, Zheng S, Zhang H, Zhang G, Wang F, et al. Exosomal mir-141 promotes tumor angiogenesis *Via* Klf12 in small cell lung cancer. *J Exp Clin Cancer Res* (2020) 39(1):193. doi: 10.1186/s13046-020-01680-1
17. Lai H, Li Y, Zhang H, Hu J, Liao J, Su Y, et al. Exorbase 2.0: An atlas of Mrna, Lncrna and Circrna in extracellular vesicles from human biofluids. *Nucleic Acids Res* (2022) 50(D1):D118–28. doi: 10.1093/nar/gkab1085
18. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43(7):e47. doi: 10.1093/nar/gkv007
19. Li Y, Ge X, Peng F, Li W, Li JJ. Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biol* (2022) 23(1):79. doi: 10.1186/s13059-022-02648-4
20. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models *Via* coordinate descent. *J Stat Softw* (2010) 33(1):1–22. doi: 10.18637/jss.v033.i01
21. Petralia F, Wang P, Yang J, Tu Z. Integrative random forest for gene regulatory network inference. *Bioinformatics* (2015) 31(12):i197–205. doi: 10.1093/bioinformatics/btv268
22. Guyon I, Weston J, Barnhill SD, Vapnik VN. Gene selection for cancer classification using support vector machines. *Mach Learn* (2004) 46:389–422. doi: 10.1023/A:1012487302797
23. Li JH, Liu S, Zhou H, Qu LH, Yang JH. Starbase V2.0: Decoding mirna-cerna, mirna-ncrna and protein-rna interaction networks from Large-scale clip-seq data. *Nucleic Acids Res* (2014) 42(Database issue):D92–7. doi: 10.1093/nar/gkt1248
24. Salmerna L, Poliseño L, Tay Y, Kats L, Pandolfi PP. A cerna hypothesis: The Rosetta stone of a hidden rna language? *Cell* (2011) 146(3):353–8. doi: 10.1016/j.cell.2011.07.014
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* (2003) 13(11):2498–504. doi: 10.1101/gr.1239303
26. Chan JM, Quintanal-Villalonga Á, Gao VR, Xie Y, Allaj V, Chaudhary O, et al. Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. *Cancer Cell* (2021) 39(11):1479–96.e18. doi: 10.1016/j.ccell.2021.09.008
27. Bebelman MP, Smit MJ, Pegtel DM, Baglio SR. Biogenesis and function of extracellular vesicles in cancer. *Pharmacol Ther* (2018) 188:1–11. doi: 10.1016/j.pharmthera.2018.02.013
28. Meng W, Hao Y, He C, Li L, Zhu G. Exosome-orchestrated hypoxic tumor microenvironment. *Mol Cancer* (2019) 18(1):57. doi: 10.1186/s12943-019-0982-6
29. Xu ZH, Miao ZW, Jiang QZ, Gan DX, Wei XG, Xue XZ, et al. Brain microvascular endothelial cell exosome-mediated S100a16 up-regulation confers small-cell lung cancer cell survival in brain. *FASEB J* (2019) 33(2):1742–57. doi: 10.1096/fj.201800428R
30. Li M, Shan W, Hua Y, Chao F, Cui Y, Lv L, et al. Exosomal mir-92b-3p promotes chemoresistance of small cell lung cancer through the Pten/Akt pathway. *Front Cell Dev Biol* (2021) 9:661602. doi: 10.3389/fcell.2021.661602
31. Chen S, Jiang L, Zheng X, Shao J, Wang T, Zhang E, et al. Clinical use of machine learning-based pathomics signature for diagnosis and survival prediction of bladder cancer. *Cancer Sci* (2021) 112(7):2905–14. doi: 10.1111/cas.14927
32. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. Edoctor: Machine learning and the future of medicine. *J Intern Med* (2018) 284(6):603–19. doi: 10.1111/joim.12822
33. Chang S, Yim S, Park H. The cancer driver genes *Idh1/2*, *Jarid1c*, *Kdm5c*, and *utx/ Kdm6a*: Crosstalk between histone demethylation and hypoxic reprogramming in cancer metabolism. *Exp Mol Med* (2019) 51(6):1–17. doi: 10.1038/s12276-019-0230-6
34. Fong S, Debs RJ, Desprez PY. Id genes and proteins as promising targets in cancer therapy. *Trends Mol Med* (2004) 10(8):387–92. doi: 10.1016/j.molmed.2004.06.008
35. Chan JYW, Tsui JCC, Law PTW, So WKW, Leung DYP, Sham MMK, et al. Profiling of the silica-induced molecular events in lung epithelial cells using the rna-seq approach. *J Appl Toxicol* (2017) 37(10):1162–73. doi: 10.1002/jat.3471
36. Shi GJ, Zhou Q, Zhu Q, Wang L, Jiang GQ. A novel prognostic model associated with the overall survival in patients with breast cancer based on lipid metabolism-related long noncoding rnas. *J Clin Lab Anal* (2022) 36(6):e24384. doi: 10.1002/jcla.24384
37. Salzman J. Circular rna expression: Its potential regulation and function. *Trends Genet* (2016) 32(5):309–16. doi: 10.1016/j.tig.2016.03.002
38. Ding L, Li L, Tang Z. Cisplatin resistance and malignant behaviors of lung cancer cells are promoted by Circ_0002360 *via* targeting mir-6751-3p to regulate the expression of *Znf300*. *Thorac Cancer* (2022) 13(7):986–96. doi: 10.1111/1759-7714.14342
39. Cui X, Zhang B, Li B, Li X. Circular rna Circ_0002360 regulates the taxol resistance and malignant behaviors of taxol-resistant non-small cell lung cancer cells by microrna-585-3p-Dependent modulation of G protein regulated inducer of neurite outgrowth 1. *Bioengineered* (2022) 13(4):9070–85. doi: 10.1080/21655979.2022.2053803
40. Zhang Y, Zeng S, Wang T. Circular rna Hsa_Circ_0002360 promotes non-small cell lung cancer progression through upregulating matrix metalloproteinase 16 and sponging multiple micrornas. *Bioengineered* (2021) 12(2):12767–77. doi: 10.1080/21655979.2021.1999370
41. Wu Z, Liu B, Ma Y, Chen H, Wu J, Wang J. Discovery and validation of Hsa_Circ_0001953 as a potential biomarker for proliferative diabetic retinopathy in human blood. *Acta Ophthalmol* (2021) 99(3):306–13. doi: 10.1111/aos.14585
42. Huang Z, Su R, Qing C, Peng Y, Luo Q, Li J. Plasma circular rnas Hsa_Circ_0001953 and Hsa_Circ_0009024 as diagnostic biomarkers for active tuberculosis. *Front Microbiol* (2018) 9:2010. doi: 10.3389/fmicb.2018.02010
43. Wu H, Liu Y, Duan H, Fan X, Wang Y, Song J, et al. Identification of differentially expressed circular rnas in human nasopharyngeal carcinoma. *Cancer Biomark* (2020) 29(4):483–92. doi: 10.3233/cbm-201731
44. Ma Z, Han C, Xia W, Wang S, Li X, Fang P, et al. Circ5615 functions as a cerna to promote colorectal cancer progression by upregulating tnks. *Cell Death Dis* (2020) 11(5):356. doi: 10.1038/s41419-020-2514-0
45. Zhu KP, Zhang CL, Ma XL, Hu JP, Cai T, Zhang L. Analyzing the interactions of mrnas and ncRNAs to predict competing endogenous rna networks in osteosarcoma chemo-resistance. *Mol Ther* (2019) 27(3):518–30. doi: 10.1016/j.ymt.2019.01.001
46. Chen G, Xie D, Zhang P, Zhou H. Circular rna Hsa_Circ_0000437 may be used as a new indicator for the diagnosis and prognosis of hepatocellular carcinoma. *Bioengineered* (2022) 13(6):14118–24. doi: 10.1080/21655979.2022.2081458
47. Li T, Liu H, Dong C, Lyu J. Application of mirna biomarkers in predicting overall survival outcomes for lung adenocarcinoma. *BioMed Res Int* (2022) 2022:5249576. doi: 10.1155/2022/5249576
48. Duca RB, Massillo C, Dalton GN, Farré PL, Graña KD, Gardner K, et al. Mir-19b-3p and mir-101-3p as potential biomarkers for prostate cancer diagnosis and prognosis. *Am J Cancer Res* (2021) 11(6):2802–20.
49. Zhang X, Wen X, Feng N, Chen A, Yao S, Ding X, et al. Increased expression of T-box transcription factor protein 21 (Tbx21) in skin cutaneous melanoma predicts better prognosis: A study based on the cancer genome atlas (Tcga) and genotype-tissue expression (Gtex) databases. *Med Sci Monit* (2020) 26:e923087. doi: 10.12659/msm.923087
50. Andreev K, Trufa DI, Siegmund R, Rieker R, Hartmann A, Schmidt J, et al. Impaired T-Bet-Pstat1 α and perforin-mediated immune responses in the tumoral region of lung adenocarcinoma. *Br J Cancer* (2015) 113(6):902–13. doi: 10.1038/bjc.2015.255
51. Kachler K, Holzinger C, Trufa DI, Sirbu H, Finotto S. The role of Foxp3 and tbet Co-expressing treg cells in lung carcinoma. *Oncoimmunology* (2018) 7(8):e1456612. doi: 10.1080/2162402x.2018.1456612
52. Lee K, Min HJ, Jang EJ, Hong JH, Hwang ES. *In vivo* tumor suppression activity by T cell-specific T-bet restoration. *Int J Cancer* (2010) 127(9):2129–37. doi: 10.1002/ijc.25238
53. Zeng Z, Lu J, Wu D, Zuo R, Li Y, Huang H, et al. Poly(Adp-ribose) glycohydrolase silencing-mediated H2b expression inhibits Benzo(a)Pyrene-induced carcinogenesis. *Environ Toxicol* (2021) 36(3):291–7. doi: 10.1002/tox.23034
54. Taki T, Shiraki Y, Enomoto A, Weng L, Chen C, Asai N, et al. Cd109 regulates *in vivo* tumor invasion in lung adenocarcinoma through tgf-B signaling. *Cancer Sci* (2020) 111(12):4616–28. doi: 10.1111/cas.14673
55. Chen H, Cai W, Chu ESH, Tang J, Wong CC, Wong SH, et al. Hepatic cyclooxygenase-2 overexpression induced spontaneous hepatocellular carcinoma formation in mice. *Oncogene* (2017) 36(31):4415–26. doi: 10.1038/onc.2017.73
56. Grabenstein S, Barnard KN, Anim M, Armoos A, Weichert WS, Bertozzi CR, et al. Deacetylated sialic acids modulates immune mediated cytotoxicity *Via* the sialic acid-siglec pathway. *Glycobiology* (2021) 31(10):1279–94. doi: 10.1093/glycob/cwab068
57. Monteiro FL, Baptista T, Amado F, Vitorino R, Jerónimo C, Helguero LA. Expression and functionality of histone H2a variants in cancer. *Oncotarget* (2014) 5(11):3428–43. doi: 10.18632/oncotarget.2007
58. Hemler ME. Targeting of tetraspanin proteins—potential benefits and strategies. *Nat Rev Drug Discovery* (2008) 7(9):747–58. doi: 10.1038/nrd2659
59. Qi Y, Qi W, Liu S, Sun L, Ding A, Yu G, et al. Tspan9 suppresses the chemosensitivity of gastric cancer to 5-fluorouracil by promoting autophagy. *Cancer Cell Int* (2020) 20:4. doi: 10.1186/s12935-019-1089-2
60. Shao S, Piao L, Wang J, Guo L, Wang J, Wang L, et al. Tspan9 induces emt and promotes osteosarcoma metastasis *Via* activating fak-Ras-Erk1/2 pathway. *Front Oncol* (2022) 12:774988. doi: 10.3389/fonc.2022.774988
61. Mendaza S, Ulazia-Garmendia A, Monreal-Santesteban I, Córdoba A, Azúa YR, Aguiar B, et al. Adam12 is a potential therapeutic target regulated by hypomethylation in triple-negative breast cancer. *Int J Mol Sci* (2020) 21(3):903. doi: 10.3390/ijms21030903