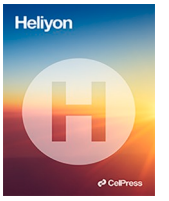
Heliyon 10 (2024) e30831

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Analyzing the involvement of diverse cell death-related genes in diffuse large B-cell lymphoma using bioinformatics techniques

Heyuan Feng [a,*], Xiyuan Zhang [b], Jian Kang [a]

[a] Flow Cytometry Room, Beijing Gaobo Boren Hospital, Beijing, China
[b] Department of Blood Transfusion, No.970 Hospital of PLA Joint Logistics Support Force, Shandong, China

ARTICLE INFO

Keywords:
Diffuse large B-Cell lymphoma
Programmed cell death
Prognosis
Risk model
Biomarkers

ABSTRACT

Diffuse large B-cell lymphoma (DLBCL) stands as the most prevalent subtype of non-Hodgkin's lymphoma and exhibits significant heterogeneity. Various forms of programmed cell death (PCD) have been established to have close associations with tumor onset and progression. To this end, this study has compiled 16 PCD-related genes. The investigation delved into genes linked with prognosis, constructing risk models through consecutive application of univariate Cox regression analysis and Lasso-Cox regression analysis. Furthermore, we employed RT-qPCR to validate the mRNA expression levels of certain diagnosis-related genes. Subsequently, the models underwent validation through KM survival curves and ROC curves, respectively. Additionally, nomogram models were formulated employing prognosis-related genes and risk scores. Lastly, disparities in immune cell infiltration abundance and the expression of immune checkpoint-associated genes between high- and low-risk groups, as classified by risk models, were explored. These findings contribute to a more comprehensive understanding of the role played by the 16 PCD-associated genes in DLBCL, shedding light on potential novel therapeutic strategies for the condition.

1. Introduction

Diffuse large B-cell lymphoma (DLBCL) is a type of non-Hodgkin's lymphoma [1], accounting for approximately 30 % of malignant lymphomas. DLBCL patients treated with R–CHOP have the potential for cure [2]. However, the treatment outcomes may vary considerably due to the high heterogeneity of DLBCL. Despite the presence of traditional clinical features (such as age, Eastern Cooperative Oncology Group (ECOG) performance status, Ann Arbor stage, and lactate dehydrogenase (LDH) level), patients with identical clinical characteristics may experience different treatment outcomes [3]. The advent of high-throughput sequencing technology enables the molecular exploration of DLBCL pathogenesis, leading to the identification of potential therapeutic targets.

According to the triggering mechanisms, cell death primarily occurs in two ways. Accidental cell death is an uncontrolled biological process, while programmed cell death (PCD) involves complex regulations and various mechanisms. Previous research has confirmed the roles of these PCD-related genes in tumor prognosis [4] and non-tumor diagnostics [5]. Various patterns of PCD play a crucial role in tumor progression. This paper categorizes 16 common modes of cell death, including apoptosis, necroptosis, ferroptosis, pyroptosis, netotic cell death, entotic cell death, lysosome-dependent cell death, parthanatos, autophagy-dependent cell death, oxeiptosis, and alkaliptosis.

Apoptosis is a process where the body eliminates injured and unwanted cells, morphologically characterized by cell membrane

* Corresponding author.
E-mail address: Fengheyuan123@outlook.com (H. Feng).

https://doi.org/10.1016/j.heliyon.2024.e30831
Received 28 December 2023; Received in revised form 26 April 2024; Accepted 6 May 2024
Available online 7 May 2024
2405-8440/© 2024 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

blistering, cell shrinkage, and the formation of apoptotic vesicles. Importantly, it does not trigger an inflammatory response [6]. Necrotizing apoptosis involves irreversible cell damage due to various pathological processes, ultimately leading to cell death [7]. Pyroptosis relies on caspase-1 to form plasma membrane pores, releasing pro-inflammatory cytokines and causing cell lysis. It is a vital process in the body's defense against inflammation and pathogenic microorganisms [8]. Iron death results from iron overload and ROS-dependent accumulation of lipid peroxides [9]. Cuproposis, distinct from oxidative stress-related cell death, is induced by intracellular copper [10]. Reticulocyte death is a form of PCD caused by extracellular traps released from neutrophils [11]. Parthanatos is a PARP1-dependent, cystitis-dependent cell death pathway [12]. Lysosome-dependent cell death occurs when altered lysosomal membrane permeability leads to the release of seed hydrolases into the cytoplasm [13]. Autophagy is characterized by cytoplasmic vacuolization, autophagic vesicle formation, and material removal via lysosomes [14]. Alkaliptosis, a recently discovered type of PCD, involves cell regulation through alkalinization [15]. Oxeiptosis is a caspase-independent apoptosis-like cell death pathway induced by ROS [16]. The majority of PCDs have been shown to participate in biological pathways influencing the development and progression of a wide range of cancers.

Previous research has confirmed the roles of certain PCD-related genes in DLBCL. For instance, Xiong et al. investigated the predictive significance of autophagy-related genes for the survival probability and drug resistance in DLBCL [17]. Wang et al. developed a novel prognostic model for DLBCL based on iron death-related genes [18]. However, there has been no study to date that utilizes other PCDs or all PCDs to construct a prognostic model for DLBCL and identify crucial prognostic genes.

To address this gap, we obtained transcriptome data for DLBCL and its control group from the GEO database (https://www.ncbi.nlm.nih.gov/geo/). Subsequently, we extracted the expression levels of 16 PCD-related genes from previous studies. Subsequently, single-factor Cox regression analysis was performed on PCD-related genes. In this analysis, the expression levels of PCD-related genes were used as independent variables, while survival time was considered as the dependent variable for regression analysis. Hazard ratios (HR) and corresponding p-values were calculated for each gene. After extracting PCD-related genes significantly associated with prognosis, a risk model was constructed based on Lasso-Cox regression analysis. In Lasso-Cox regression, a Lasso penalty was applied to identify 16 important PCD-related genes associated with survival risk and to construct a risk model capable of predicting patient survival.

Additionally, RT-qPCR was performed to validate the mRNA expression levels of the diagnosis-related genes. The validity of the risk model was assessed by generating KM (Kaplan-Meier) survival and ROC curves on an external test set. Nomogram models were also developed using prognosis-related genes and risk scores from the risk models. We evaluated differences in immune cell infiltration abundance and the expression of immune checkpoint-associated genes between high- and low-risk groups. Finally, interaction networks of prognosis-related genes with miRNAs were constructed separately.

2. Method

2.1. Data acquisition

In this study, "diffuse large B-cell lymphoma" and "GEO" were used as keywords to search the relevant literature of diffuse large B-cell lymphoma in NCBI, and the relevant data of diffuse large B-cell lymphoma were collected from the literature. In multiple datasets, GSE10846 and GSE31312 contained more samples, clinical survival information, and clinical staging, and ultimately, GSE10846 and GSE31312 were used for subsequent analyses. Transcriptome data for DLBCL (GSE10846 and GSE31312) were retrieved from the GEO database. Specifically, the GSE10846 dataset comprises transcriptomic data for 414 DLBCL patients, while the GSE31312 dataset encompasses transcriptomic data for 498 DLBCL patients. The specific preprocessing steps were as follows: we utilized the *averesps(.)* function from the *limma* package to average the expression levels of duplicated genes [19], followed by standardizing the data using the *normalizeBetweenArrays(.)* function. We retained the genes contained in both GSE10846 and GSE31312 datasets. Additionally, samples with survival time and survival status were retained for analysis. GSE10846 retained 385 patients, and GSE31312 retained 470 patients. Additionally, the survival data of patients from GSE10846 and GSE31312 are provided in the Supplementary Material files "GSE10846.clinical.xlsx" and "GSE31312.clinical.pdf", respectively. Among them, GSE31312 (300 alive, 170 dead, median survival time was 35.023 months) includes information such as follow-up time, survival status, survival time, age, gender, etc., while GSE10846 (237 alive, 151 dead, median survival time was 2.405 year) includes survival status, survival time, age, gender, Stage.

In the subsequent analysis, we utilized the DLBCL transcriptome data from GSE10846 and GSE31312 as the training and testing sets, respectively. The training set was employed for screening prognostic genes and constructing the risk model, while the testing set was used for validating the predictive performance of the risk model. Finally, a total of 1254 genes related to programmed cell death (PCD) were collected. The list of PCD-related genes is provided in the supplementary material file "PCDgene.txt."

2.2. Selection of prognostic genes and enrichment analysis

In this study, based on single-factor Cox regression [20], 28 genes significantly associated with prognosis (p-value <0.00001) were selected from 1254 PCD-related genes in the training set transcriptome data. The False Discovery Rate (FDR) is a statistical method used to control the number of false positives in experiments. In this step, we introduced the False Discovery Rate (FDR) [21] to assess the reliability of single-factor Cox regression analysis. Subsequently, the R package clusterProfiler was used to perform Gene Ontology (GO) enrichment analysis, and significant pathways involved in prognosis-related genes were screened using the hypergeometric distribution. In the screening process, the Benjamini-Hochberg correction (also known as FDR correction) was introduced to adjust the original p-values, ensuring the accuracy and reliability of statistical significance when conducting multiple hypothesis tests. Finally,

significant pathways with corrected p-values less than 0.05 were retained.

Subsequently, we employed the R package *ggplot2* to visually represent the results of the GO enrichment analysis [22]. Additionally, we conducted a Metascape enrichment analysis utilizing the Metascape database (https://metascape.org/gp/index.html#/main/step1) [23].

2.3. Approaches to screening for prognosis-related genes

This study employed univariate Cox regression analysis to conduct a screening for genes associated with DLBCL prognosis. FDR (False Discovery Rate) is a statistical method used to control the number of false positives occurring in experiments. In this step, we introduce False Discovery Rate (FDR) to assess the reliability of single-factor Cox regression analysis [21].

Additionally, KM survival analysis was performed on prognosis-related genes using the R package *survival* [24]. A significant difference in survival was considered when the p-value for survival between the high-risk and low-risk groups was less than 0.05.

2.4. Methods for constructing prognostic models

The Lasso algorithm was employed, utilizing the R package *glmnet* to conduct Lasso-Cox regression analysis on the training set and construct a risk model [25]. In Lasso-Cox regression, the lasso algorithm is employed for feature selection on survival data to build a model predicting survival risk. The core idea of the Lasso algorithm is to introduce an L1 regularization term on top of ordinary least squares regression, which is the sum of the absolute values of coefficients. By adjusting the regularization parameter, the lasso algorithm can shrink some coefficients to zero, achieving the goal of feature selection. We utilized the Lasso-Cox regression method to construct a risk model for calculating the risk scores of each patient. Subsequently, we divided patients into high-risk and low-risk groups based on the median values of these risk scores. Thus, the entire sample was divided into two groups, and KM survival curves were computed for each group to assess the survival differences between different risk groups. The method for calculating risk scores provided by the risk model is as follows.

$$\text{Risk score} = \sum_{1}^{n} \text{Coefficient}\,(\text{RNA}_i) \times \text{Expression}\,(\text{RNA}_i) \tag{2-1}$$

In our study, the log-rank test was employed to assess the survival differences between different risk groups. KM survival analysis and ROC analysis were performed using the *survivalROC* package in R. The risk model demonstrated its ability to provide and validate the model effectively. We put the main code in the "Code" folder of the supplementary material.

2.5. Construction method of the nomogram model

We first sequentially utilize univariate Cox regression and multivariate Cox regression analyses to screen independent prognostic models from clinical factors and risk scores. The introduction of False Discovery Rate (FDR) measures the reliability of the results. Subsequently, based on the independently selected prognostic factors, a column line chart model is constructed. Specifically, we developed a nomogram model utilizing the R package *rms* that relies on prognosis-related genes and risk scores. The accuracy of the nomogram model was evaluated using a calibration curve. Furthermore, we appraised the clinical utility of the nomogram model through decision curve analysis (DCA).

2.6. The analysis of immune infiltration and immunotherapy

In this paper, we employed the ssGSEA algorithm [26] to evaluate the abundance of 22 immune cell infiltrates in DLBCL samples. Additionally, we conducted a comprehensive analysis, employing Spearman correlation analysis, to investigate the relationships among risk scores, prognosis-related genes, and immune cells. The threshold p-value for significance for the association between the risk score and immune cells was 0.001. The threshold for the remaining correlation analyses was p-value <0.05. Utilizing the R package *estimate* [27], we calculated scores for the tumor microenvironment matrix, immune response, and tumor purity. Lastly, we explored differences in the expression of genes associated with immune checkpoints and HLA genes between high- and low-risk groups as classified by the risk model. When calculating the Spearman correlation coefficient between immune infiltration and the expression of prognostic genes, and evaluating the differential expression of immune checkpoint sites and HLA genes among high and low-risk group samples, False Discovery Rate (FDR) was introduced to control the false positive rate resulting from multiple comparisons.

2.7. Drug sensitivity analysis

The R package *oncoPredict* was used to obtain drug sensitivity (IC50 values) for 138 compounds in high- and low-risk groups [28]. The IC50 values between the two groups were considered to be significantly different when the p-value was less than 1e-13.

2.8. Interaction network construction of prognosis-related genes and miRNAs

In this study, miRNAs and transcription factors interacting with diagnosis-related genes were queried using the miRTarBase

database (http://mirtarbase.cuhk.edu.cn/php/index.php) [29]and the TargetScan database (http://www.targetscan.org/vert_72/), respectively. Subsequently, an interaction network was constructed.

2.9. Statistic analysis

In this paper, bioinformatics analysis and statistical analysis were completed based on R software (v4.0.2). The Wilcoxon-test was used for statistical analysis of differences. Specifically, we utilized the Wilcoxon test to compare the expression levels of immune checkpoint markers, HLA genes, ESTIMATE Score, Immune Score, and Stromal Score between samples from high and low-risk groups for significant differences. The Kaplan-Meier log-rank test was employed to analyze survival differences between high and low-risk groups. The Spearman correlation coefficient was calculated to assess the correlation. Specifically, we evaluated the correlation between immune cell infiltration abundance/immune function scores and prognostic genes in the training and testing datasets.

2.10. Experimental validation of prognosis-related genes

In order to detect the mRNA expression of prognosis-related genes in blood samples of DLBCL patients, RT-qPCR was performed to detect the expression of diagnosis-related genes. Specifically, we divided the samples into DLBCL group and control group to detect the expression of AEN, DNAJC10, DNM1L, ELL3 and HIF1A genes. GAPDH was used as an internal reference. The mRNA expression level was calculated with the 2-ΔΔCt. All data were expressed by means ± SD, and the statistical differences between groups were tested by T-Test, and p-value <0.05 indicated a significant difference. In RT-qPCR analysis, primer sequences play a crucial role. Primers are short DNA fragments used to amplify the target DNA segments, they anneal to the ends of the target DNA sequence and guide DNA replication during PCR. The selection of primer sequences can directly affect the specificity, efficiency, and accuracy of PCR. Therefore, primer sequence information is provided in Table 1.

3. Results

3.1. Acquisition of prognosis-related genes

In this section, we will provide a detailed overview of the process involved in acquiring both PCD-related genes and prognostic genes. The technology roadmap for this study is illustrated in Fig. 1. Specifically, we initially extracted the expression data of PCD genes from the training set.

To screen for genes significantly associated with the prognosis of DLBCL, we conducted univariate Cox regression analysis on the expression data of all PCD-related genes in the clinical samples of DLBCL from the GSE10846 dataset, resulting in 28 significant genes (p-value <0.00001). The forest plot resulting from this screening is presented in Fig. 2A. A total of 28 genes, including AEN, BAK1, BNIP3L, DNAJC10, DNM1L, ELL3, ERP29, HIF1A, ITGA6, ITGAV, MAEL, MMP9, NLE1, PDK1, PIK3R1, PPIF, PTPRC, YWHAB, YWHAE, SLC1A5, ATP13A2, MLST8, POLDIP2, TBC1D14, UBQLN2, USP30, BIRC2, and ATP7A, were found to be significantly correlated with prognosis (p-value <0.001). FDR values corrected by Benjamin Hochberg have been placed in the Supplementary material.

MetaScape analysis and GO enrichment analysis were performed on these genes. Fig. 2B represents a histogram of the significant pathways identified through MetaScape analysis. Fig. 2C–D depict circle and bar graphs, respectively, illustrating the significant pathways identified through GO enrichment analysis. These enriched pathways have been confirmed to be closely associated with the development of DLBCL, such as the regulation of proteolysis, malignant pleural mesothelioma, and regulation of autophagy, among others. The relationship between these pathways and DLBCL will be extensively discussed in the following sections.

Table 1
The primer sequences.

Primers	Sequence (5'→3')	
AEN	Forward	GAGATCCTTAAGCTCCTGAAGG
	Reverse	TCAATTCCTTCTGTCCTGTGC
DNAJC10	Forward	AGGATCTTATGAATCCTTCAGTGG
	Reverse	TTGAACGTTTTCCTGGGCAC
DNM1L	Forward	GTGTGCCAAAGGCAGTAAAAC
	Reverse	TTGACTGGCTCCTTGTAATGC
ELL3	Forward	AGGCTATTCTGAAGGAGATGC
	Reverse	TCTTGCTCCCAATCTTCTCC
HIF1A	Forward	TTGGCAGCAACGACACAGAA
	Reverse	TCGAAGTGGCTTTGGCGTTT
GAPDH	Forward	CGCTAACATCAAATGGGGTG
	Reverse	TGCCAGCCCCAGCGTCAAAG

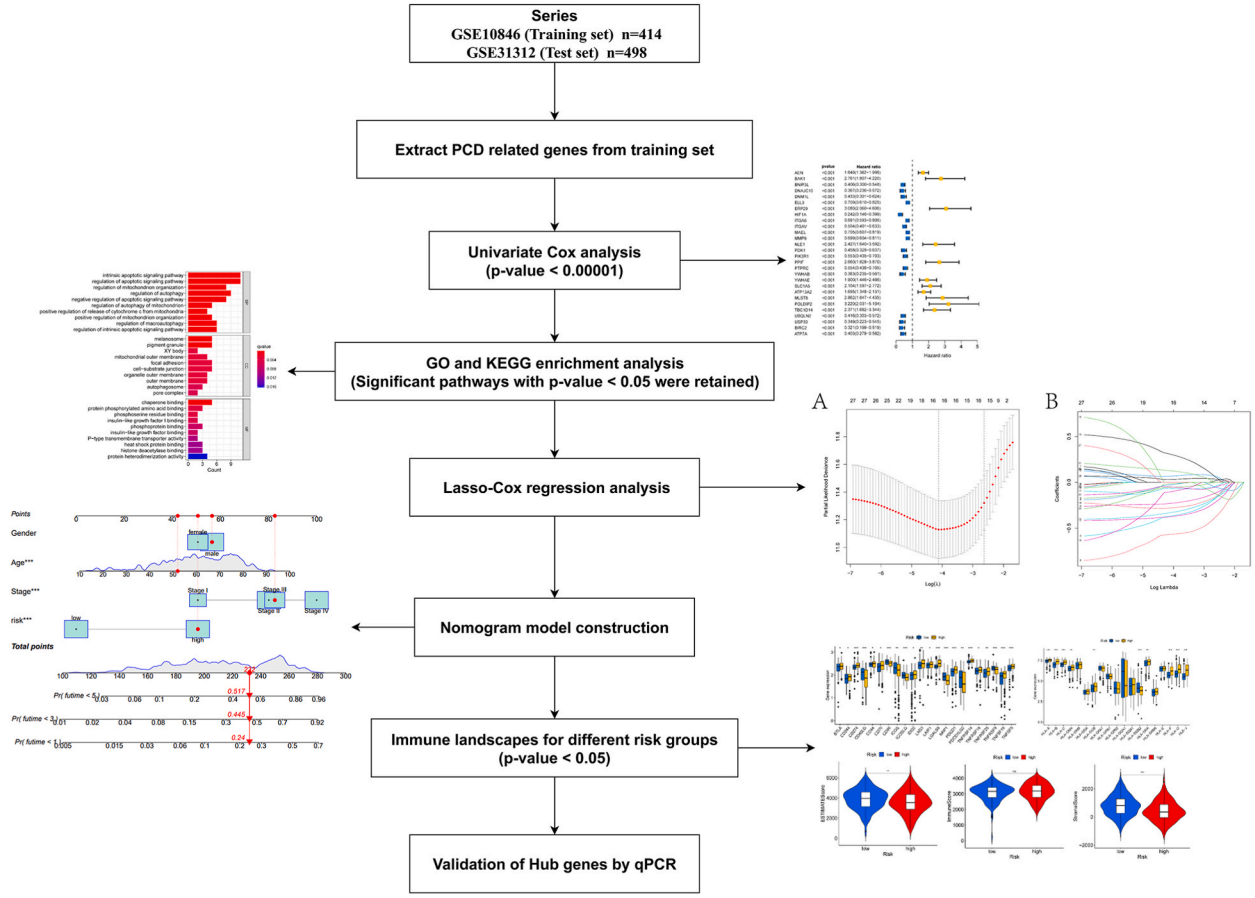


Fig. 1. The technological process diagram.

3.2. Results of prognostic model construction for 16 PCD-related genes

In this section, we provide a detailed overview of the construction process of the prognosis model and the validation results for differences in survival and prognosis between high and low-risk groups. Specifically, based on Lasso-Cox regression analysis, 16 genes were further selected from 28 significant genes to construct the risk model. Fig. 3A–B illustrate the distribution of Lasso coefficients and the partial likelihood deviation plots during the construction process. The dashed line on the left side of Fig. 3A corresponds to a lambda.min of 16, leading to the selection of 16 genes for building the prognosis model. Fig. 3B depicts the coefficient distribution, where each line represents a gene, and their endpoints indicate the corresponding coefficients. The final risk scoring formula is shown below.

$$\begin{aligned}\text{Risk score} = {} & 0.03286 \times \text{Expression(AEN)} - 0.02839 \times \text{Expression(DNAJC10)} - 0.3691 \times \text{Expression(DNM1L)} - 0.2305 \\ & \times \text{Expression(ELL3)} - 0.5697 \times \text{Expression(HIF1A)} - 0.0539 \times \text{Expression(ITGA6)} - 0.3320 \times \text{Expression(MAEL)} - 0.3754 \\ & \times \text{Expression(MMP9)} + 0.3081 \times \text{Expression(NLE1)} - 0.2801 \times \text{Expression(PDK1)} + 0.0663 \times \text{Expression(PPIF)} - 0.0661 \\ & \times \text{Expression(PTPRC)} - 0.1040 \times \text{Expression(YWHAB)} + 0.1268 \times \text{Expression(ATP13A2)} - 0.0497 \times \text{Expression(UBQLN2)} \\ & - 0.0182 \times \text{Expression (ATP7A)}\end{aligned} \tag{3-1}$$

All DLBCL samples were stratified into high- and low-risk groups based on the median values of the risk scores derived from the risk model. The KM survival curves for the risk models on the training and test sets are presented in Fig. 3C–D, respectively. As evident from the graphs, there is a significant difference in survival rates between the high- and low-risk groups. Figs. S1A–B showcase the ROC curves predicting the survival of DLBCL samples at 1, 3, and 5 years in the training and test sets. The ROC values for the training set were 0.782, 0.812, and 0.815 for 1, 3, and 5 years, respectively. Similarly, the ROC values for the test set were 0.636, 0.640, and 0.651 for 1, 3, and 5 years, respectively. We also calculated the trend of the Concordance index over time in both the training and testing datasets (Fig. 3E–F).

We conducted a comprehensive evaluation of the risk model's performance. Initially, we generated expression heat maps for the

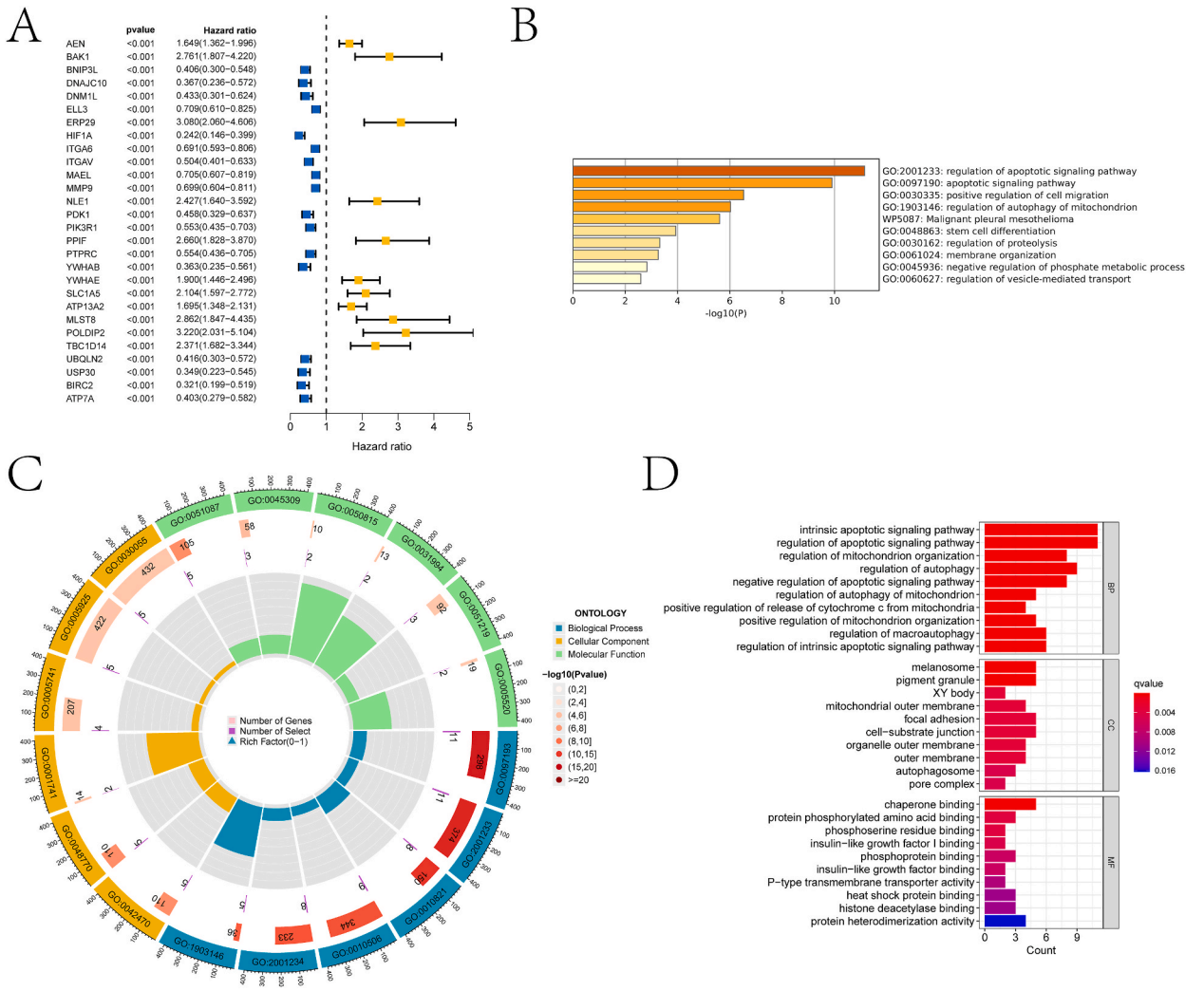


Fig. 2. Results of univariate Cox analysis and enrichment analysis of PCD genes. A is a forest plot obtained from univariate Cox regression analysis. B is the bar graph of metascape analysis of prognosis-related PCD genes. C and D are the circle chart and column chart of GO enrichment analysis.

identified risk genes in both high- and low-risk groups, the populations for which the risk model was developed. Fig. 4A–B depict the expression heatmaps in the training and test sets, respectively, revealing substantial differences in gene expression between the two groups. Our study results demonstrate a significant association between the high-risk group and adverse clinical outcomes in both the training and testing datasets (Fig. 4C–F). Notably, an increase in the risk score corresponds to a higher number of sample points in the high-risk group. Furthermore, we applied Principal Component Analysis (PCA) downsizing to the expression data of 28 genes in the two risk groups, resulting in a two-dimensional scatter plot represented in Fig. 4G–H. The graphs demonstrate that samples within the same group exhibit close proximity, whereas those from different groups are distinctly separated. To further evaluate the accuracy of the risk model in predicting progression-free survival (PFS), KM and ROC analyses were performed using the GSE31312 expression data and PFS survival data and are presented in Fig. S1 in the Supplementary Material.

3.3. Nomogram model construction and GSEA analysis

In this section, we conducted an independent prognosis analysis on DLBCL samples and constructed a nomogram model based on risk scores and clinical factors, aiming to further analyze the independent impact of risk scores and clinical factors on prognosis. Initially, through univariate Cox regression analysis, gender, age, stage, and risk score were identified as factors significantly associated with the prognosis of DLBCL patients (Fig. 5A). Subsequently, based on multivariate Cox regression analysis, age, stage, and risk score were identified as independent prognostic factors (Fig. 5B) and used to construct the nomogram model (Fig. 5C). FDR values corrected by Benjamin Hochberg have been placed in the Supplementary material. The decision curve analysis (DCA) curve in Fig. 5D and the calibration curve in Fig. 5E indicate that the nomogram model has excellent predictive performance. Fig. 5F shows that the

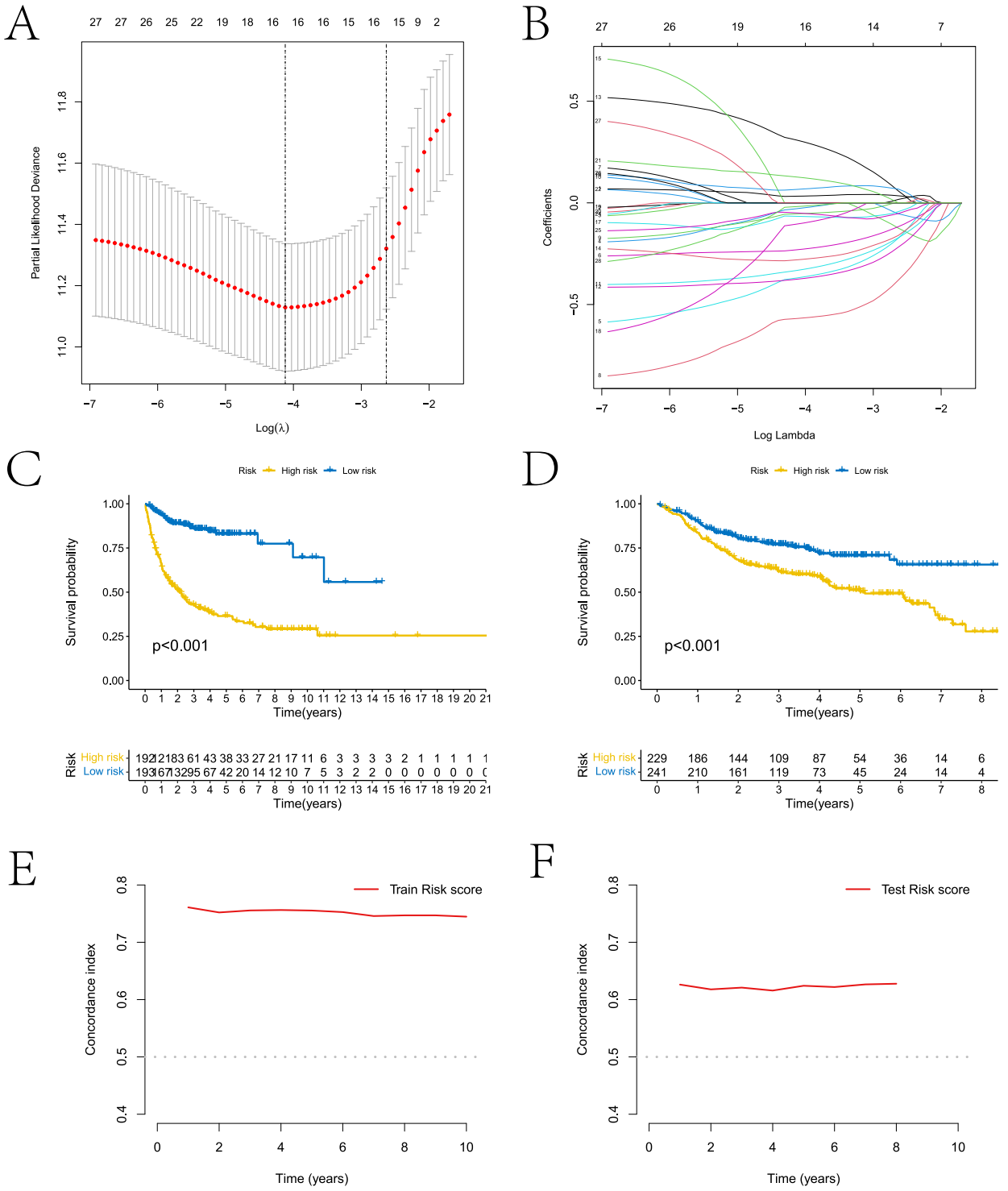


Fig. 3. Risk models constructed with prognostic correlation genes and model performance validation. A shows the partial likelihood deviation plots. B is the distribution of LASSO coefficients. C and D are the KM survival curves for the risk models on the training and test sets, respectively. E and F are line plots illustrating the relationship between survival time and Concordance index in the training and testing datasets, respectively.

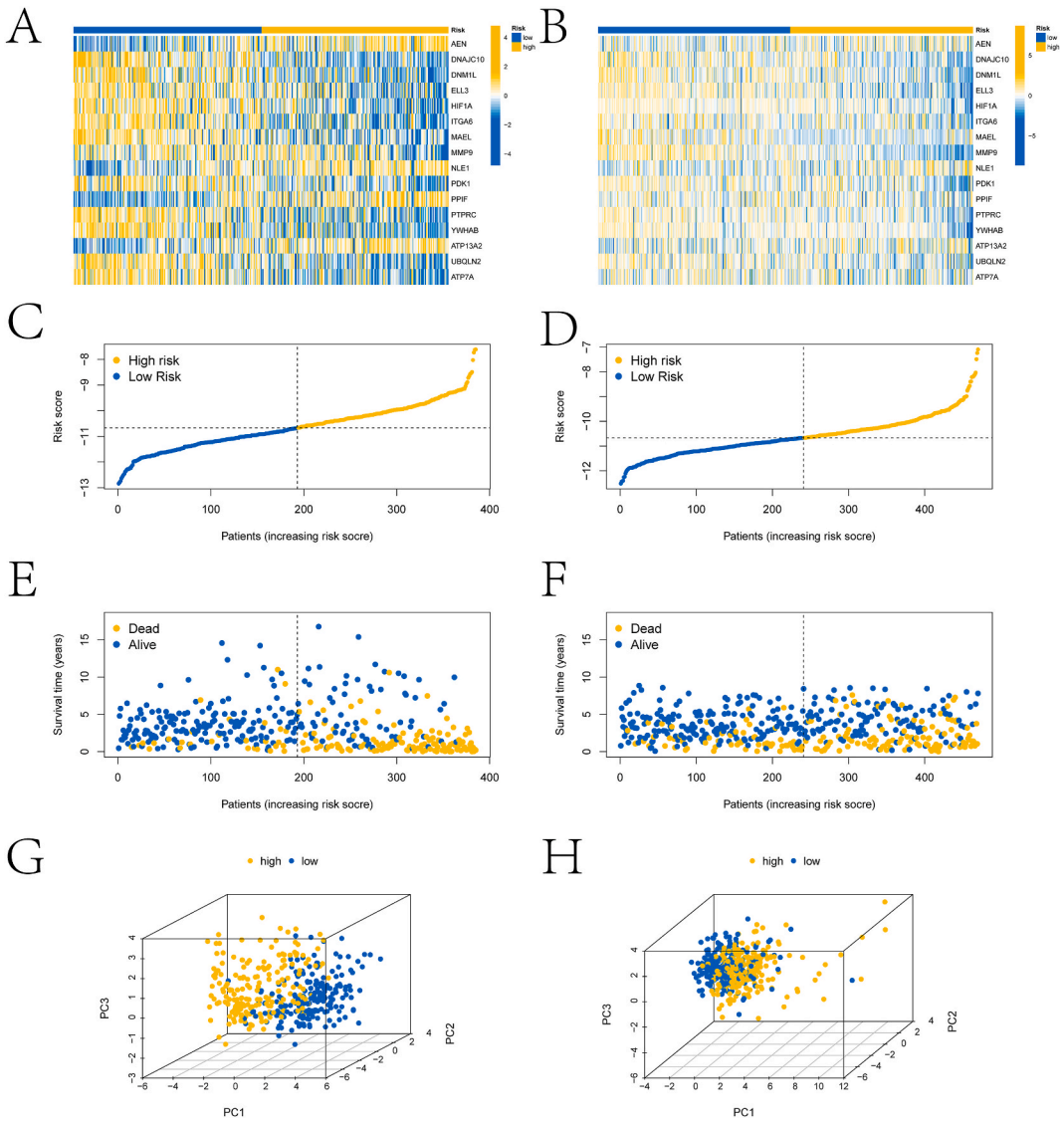


Fig. 4. Assessment of risk models. A and B are the expression heat maps of the risk genes constituting the risk model in the training and test sets, respectively. C and D are scatter plots reflecting the relationship between risk group assignment and risk scores in the training and testing datasets, respectively. E and F are scatter plots reflecting the relationship between risk scores and survival status in the training and testing datasets, respectively. G and H are visualizations after PCA reduction using the risk genes that make up the risk model on the training and test sets, respectively, with each point in the figure representing a sample.

AUC of 1-year, 3-year and 5-year survival predicted by this model reached 0.790, 0.816 and 0.825, respectively. To identify differences in pathway involvement between high and low-risk groups, GSEA analysis was conducted to identify the Top 5 pathways involved in samples from both high and low-risk groups (Fig. 6A–B). The relationship between these pathways and DLBCL will be elaborated in the discussion section. These findings underscore the potential of the prognostic nomogram model to provide valuable prognostic information for DLBCL.

3.4. Immune landscapes in different risk groups

In this section, we analyzed the differences between high and low-risk groups in terms of the abundance of immune cell infiltration and immune function scores. Utilizing the ssGSEA algorithm, this study investigated the correlation heat map depicting the immune infiltration abundance of various immune cells in the training and test sets of the samples, as well as the genes related to prognosis (Fig. 7A, Fig. 8A). Notably, a significant correlation was observed between the majority of immune cells and genes. Fig. 7B–I presents scatter plots illustrating the correlation analysis between risk scores and immune cells with significant correlations in the training set. FDR values corrected by Benjamin Hochberg have been placed in the Supplementary material. Similarly, Fig. 8B-L depict scatter plots

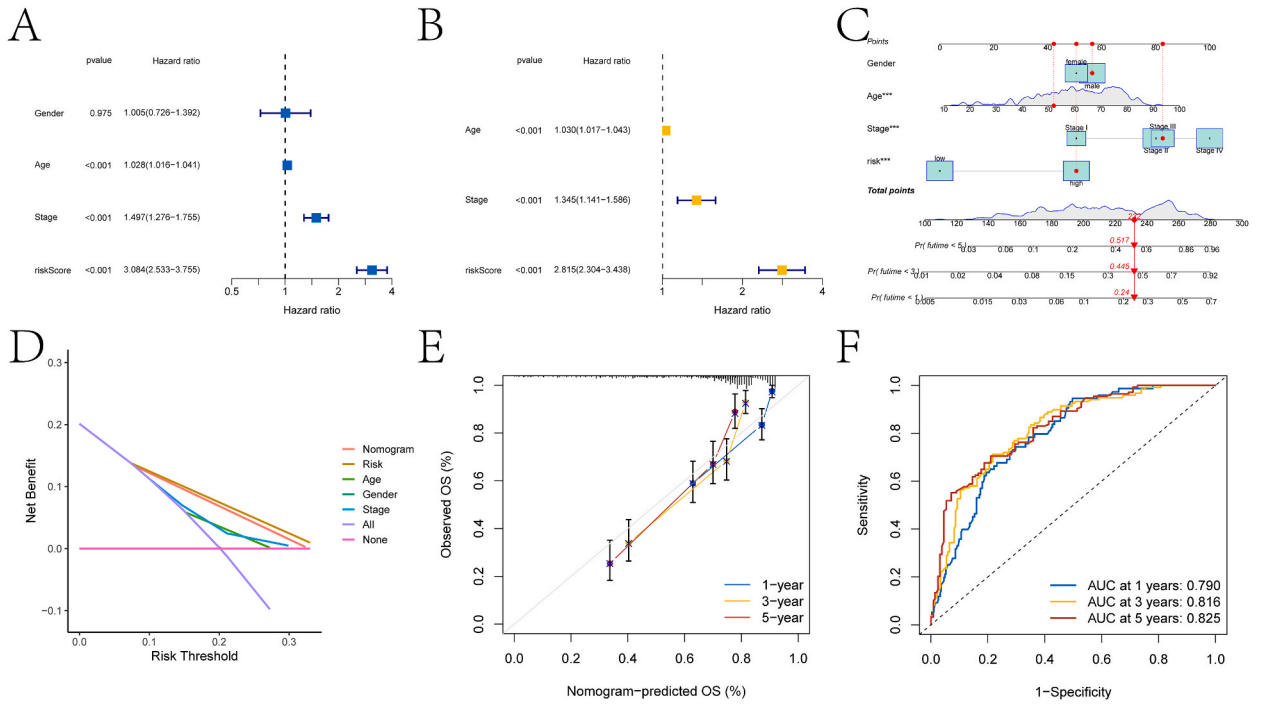


Fig. 5. Results of independent prognostic analysis. A and B are forest plots obtained from univariate Cox regression analysis and multivariate Cox regression analysis for risk scores and clinical factors, respectively. C is the nomogram model constructed from clinical factors and risk scores. D is the result of the DCA analysis. E is the calibration curve of the nomogram model. F is the result of the ROC analysis of the nomogram model.

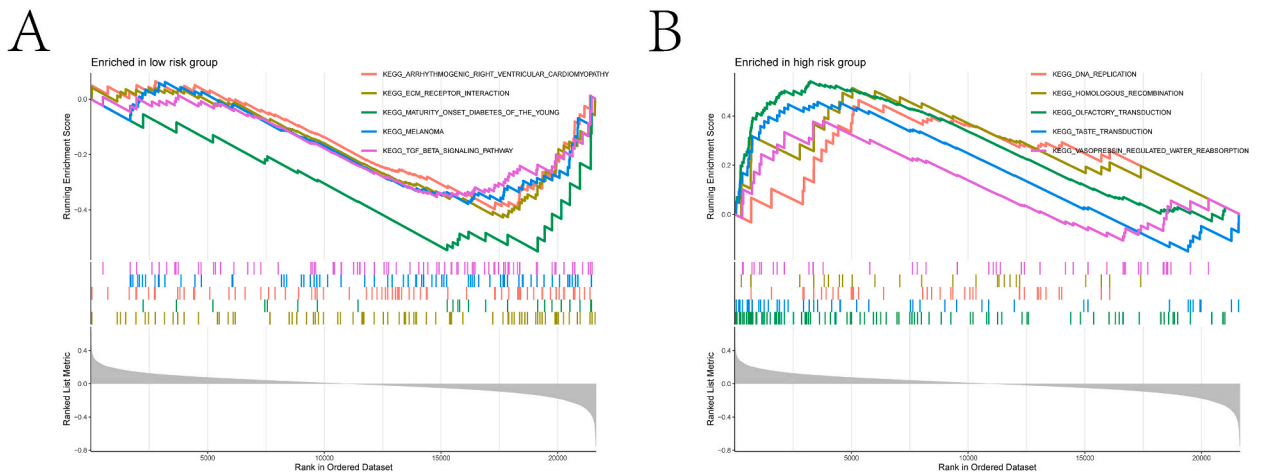


Fig. 6. GSEA analysis results. A and B are the upper five pathways obtained from GSEA analysis for the high- and low-risk groups, respectively.

of correlation analysis between risk scores and immune cells with significant correlations in the test set. Among them, multiple immune cell infiltration abundances/immune function scores showed a significant correlation with risk scores (p-value <0.001). In the training set, we confirmed a close association between risk scores and aDCs, APC_co_inhibition, DCs, iDCs, inflammation-promoting cells, pDCs, and Th2 cells. In the test set, we verified a strong correlation between risk scores and aDCs, APC_co_inhibition, HLA, iDCs, para-inflammation, Th1 cells, TIL, and Treg.

Additionally, we investigated variations in the expression of immune checkpoint-associated genes and HLA genes between samples from high- and low-risk groups (Fig. 9A–B). FDR values corrected by Benjamin Hochberg have been placed in the Supplementary material. All immune checkpoint-associated genes and the majority of HLA genes exhibited significant differences between the two groups. We will delve into the detailed roles of these genes in DLBCL in the Discussion section. Furthermore, we assessed differences in ESTIMATEScore, ImmuneScore, and StromalScore between the two groups through Estimate analysis (Fig. 9C–E). Notably,

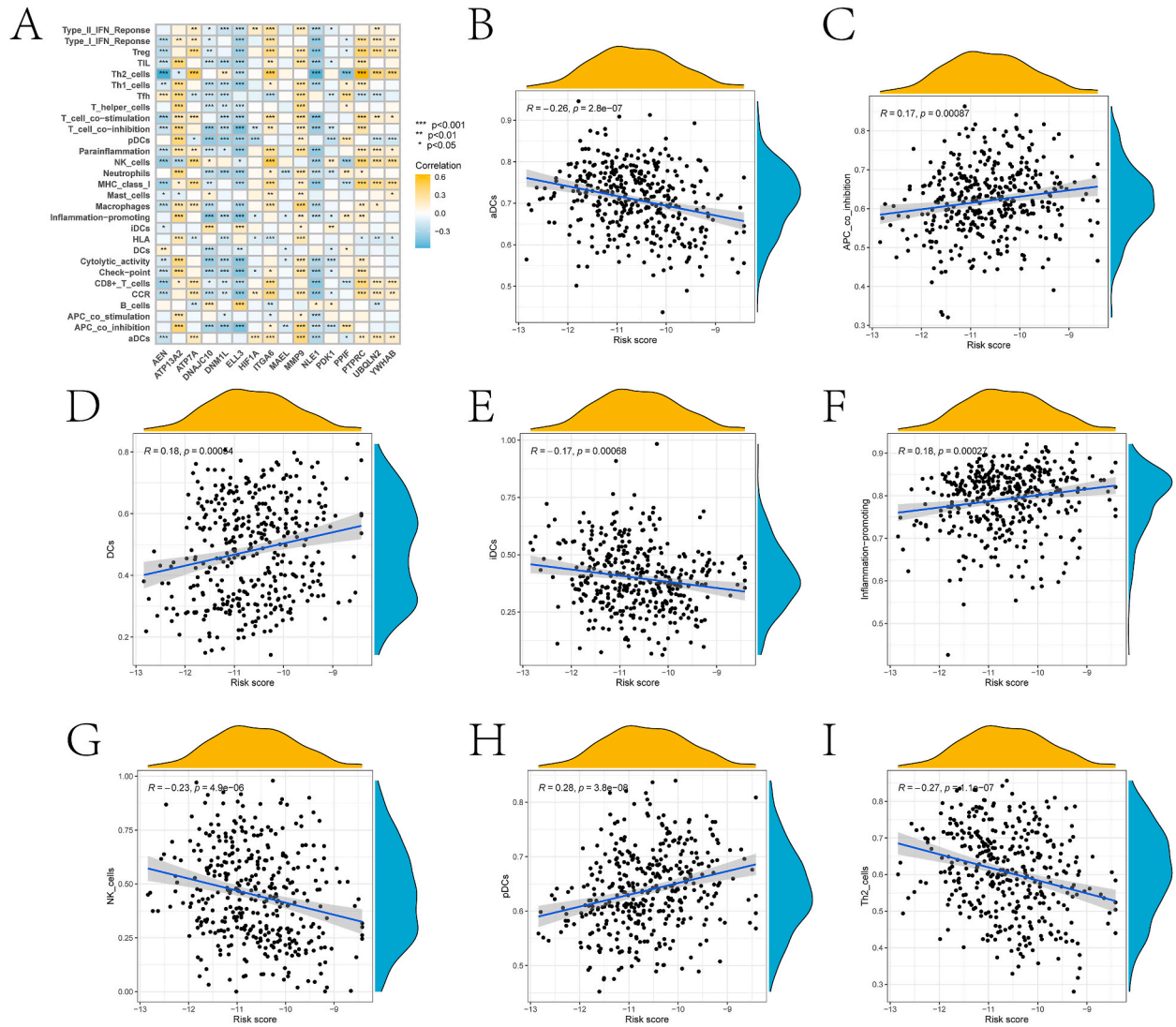


Fig. 7. Results of ssGSEA analysis on the training set. A is a heatmap of the correlation between prognosis-related genes and the abundance of immune cell infiltration obtained from ssGSEA analysis. B–I are correlation analyses between risk scores and immune cells with significant correlation.

ESTIMATEScore and StromalScore showed significant distinctions between the two groups.

3.5. Drug sensitivity and mRNA-miRNA interaction network construction

In this section, we aim to analyze the differences in drug sensitivity between high- and low-risk groups to guide stratified drug therapy for patients and identify miRNAs interacting with prognostic genes. The IC50 values of 135 compounds in the high- and low-risk groups were investigated using the R package "oncoPredict." We identified 16 compounds with significant differences between the two groups, applying a threshold of p-value <0.00001 (Fig. 10A-L). Notably, 5-Fluorouracil, AZD5153, cisplatin, and JAK1_8709 exhibited higher sensitivity in the low-risk group. Drugs such as AZD6482, BI-2536, KU-55933, doramapimod, NU7441, and RO-3306 may offer more effective treatment for the low-risk group. A detailed discussion of the roles of these compounds in DLBCL will be presented in the Discussion section. Additionally, we constructed mRNA-miRNA interaction networks (Supplementary Material Figs. S3–S6). Prognosis-related genes interact with various miRNAs, and the association of these miRNAs with DLBCL will be thoroughly examined in the Discussion section.

3.6. Results of RT-qPCR experiment

In this section, we conducted qPCR validation of the gene expression involved in constructing the risk model in both the DLBCL

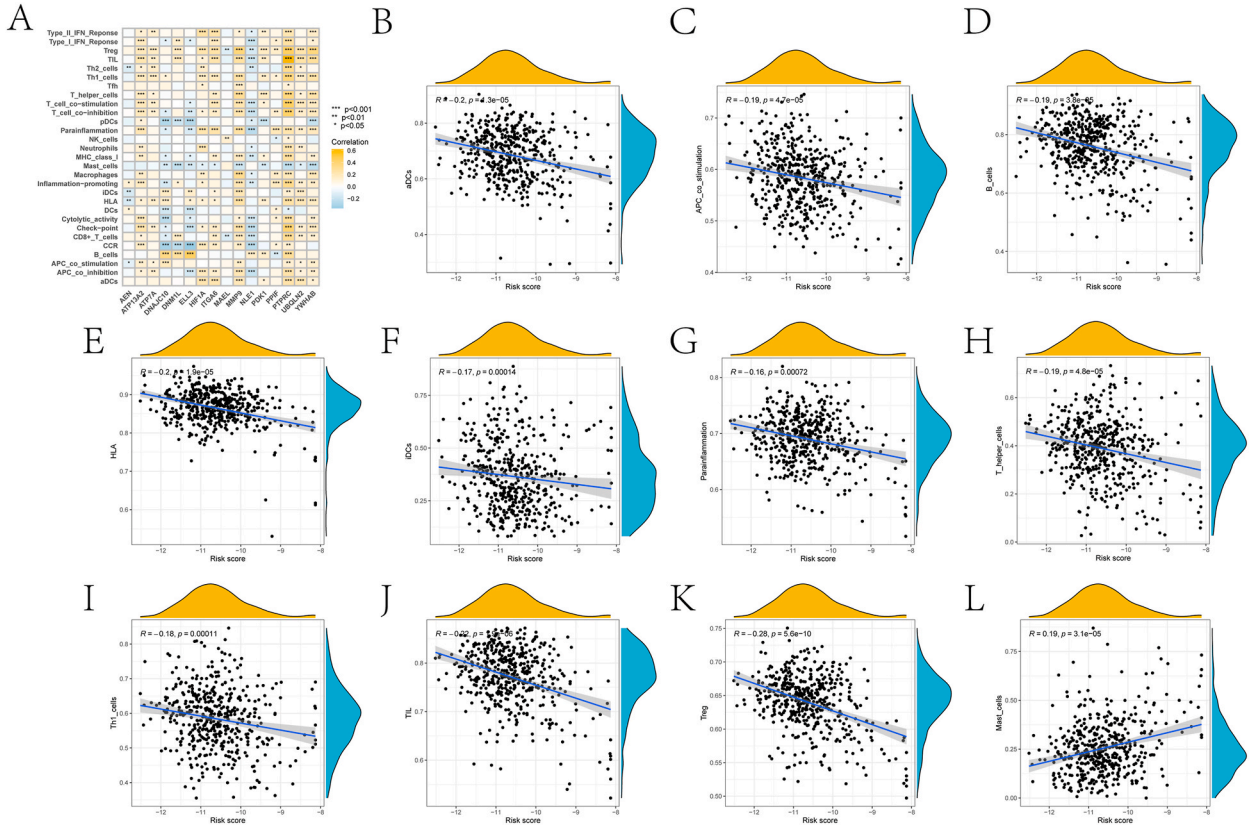


Fig. 8. Results of ssGSEA analysis on the test set. A is a heatmap of the correlation between prognosis-related genes and the abundance of immune cell infiltration obtained from ssGSEA analysis. B-L are the risk scores with significant correlation and the correlation analysis between immune cells.

group and the control group. As illustrated in Fig. 11A–E, when compared to the Normal Group, the serum expression levels of AEN, DNAJC10, DNM1L, ELL3, and HIF1A genes were markedly elevated in patients with DLBCL in the control group.

4. Discussion

DLBCL is a non-Hodgkin's lymphoma characterized by a high degree of heterogeneity. The majority of programmed cell deaths (PCDs) play a pivotal role in oncological diseases. Firstly, to identify genes significantly associated with the survival of DLBCL patients, the expression levels of PCD-related genes were extracted from the training set and subjected to univariate Cox regression analysis in this study, identifying 28 genes significantly associated with prognosis. Further analyses were conducted on these genes using GO and Metascape enrichment analyses, revealing multiple pathways closely associated with DLBCL. Apoptosis, a physiological death process, involves the moderate regulation of apoptosis-related signaling pathways, representing a promising strategy for cancer treatment. DCZ0858 has the potential to treat DLBCL patients by inactivating the JAK0/STAT1 pathway, inducing apoptosis, and causing cell cycle arrest in the G2/G3 phase [30]. Feng et al. demonstrated that DCZ0358 induces apoptosis and DNA damage in DLBCL cells [31]. Lipids and lipid metabolism play a crucial role in various non-apoptotic cell death pathways, and alterations in lipid metabolism may initiate mitochondrial autophagy, leading to necrotizing apoptosis [32]. Vera de Jonge et al. identified peripheral T cell and NK cell characteristics distinct from DLBCL patients in HGBL-MYC/BCL2 [33]. MORTALIN regulates the release of Ca2+ from the endoplasmic reticulum through the mitochondrial membrane and promotes the down-regulation of FAS in DLBCL cells, making it a potential drug for treating DLBCL [34].

Secondly, in order to stratify DLBCL patients, this study employed Lasso-Cox regression analysis to further refine gene selection and construct a risk model. The final risk model comprises 16 genes (AEN, DNAJC10, DNM1L, ELL3, HIF1A, ITGA6, MAEL, MMP9, NLE1, PDK1, PPIF, PTPRC, YWHAB, ATP13A2, UBQLN2, and ATP7A). Some of these genes have been confirmed to be closely associated with DLBCL. AEN is an apoptosis-enhancing nuclease. The protein encoded by DNM1L mediates mitochondrion and peroxisome division, related to cytochrome C release and the rate of caspase activation during apoptosis. Monti S et al. observed that the DLBCL cluster named "OxPhos" was significantly enriched in genes related to oxidative phosphorylation and mitochondrial function. There is a potential relationship between the imbalance of anti-apoptosis BCL2-2 and the disruption of mitochondrial membrane potential, the reduction of cytochrome C release, and caspase-mediated apoptosis in these tumors [35]. Broecker-Preuss M et al. investigated the

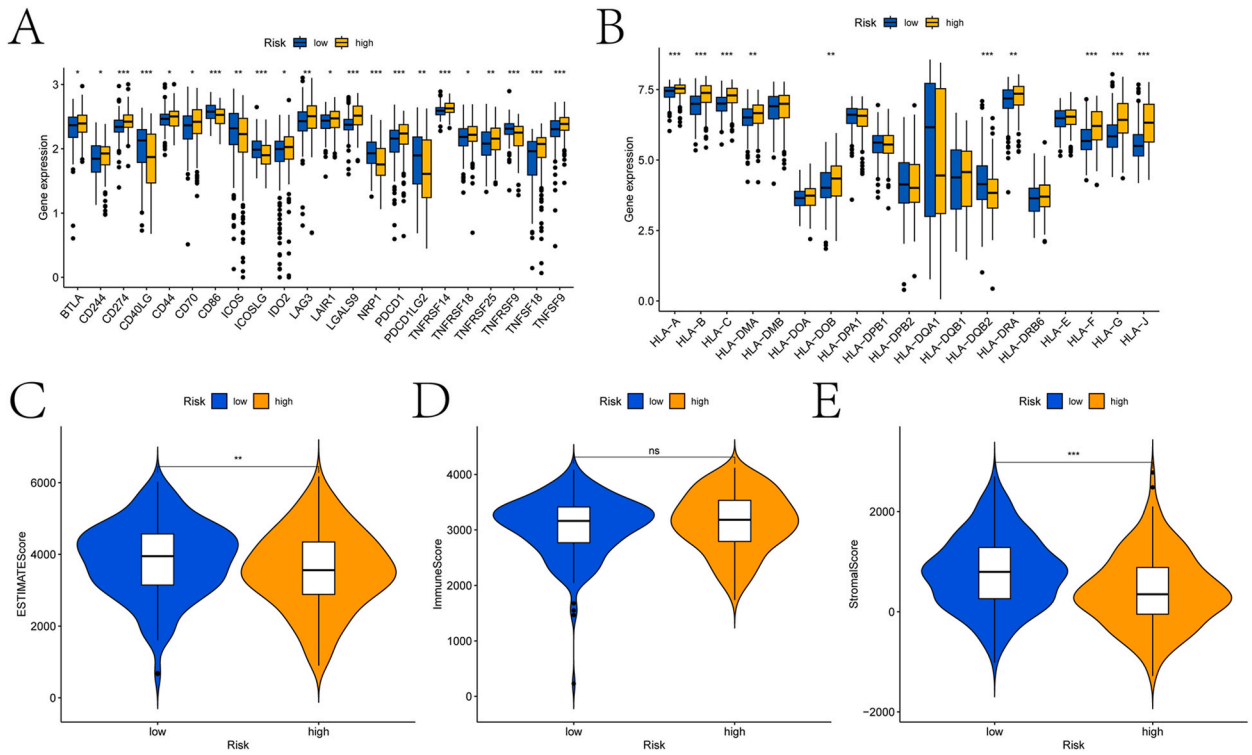


Fig. 9. Analysis of immunological differences between high- and low-risk groups. A and B are differential box plots of the expression of genes associated with immune check loci and HLA gene expression for the high-and low-risk groups, respectively. C, D, and E are the results of the Estimate analysis (ESTIMATEScore, ImmuneScore, and StromalScore).

impact of single nucleotide polymorphisms (SNPs) in the HIF1A gene on DLBCL and found that the HIF1A A588t SNP was associated with total metabolic tumor volume (TMTV) and progression time [36]. Based on a comprehensive gene expression database, Xie Y et al. identified MAEL as an RNA binding protein (RBP)-related gene influencing the prognosis of DLBCL [37]. The tumor microenvironment (TME) may play a role in DLBCL pathogenesis, and the overexpression of MMP9 in microenvironment cells may lead to favorable or unfavorable outcomes [38]. NLE1 is predicted to participate in the Notch signaling pathway and ribosome large subunit assembly, making it a potential target for treating DLBCL patients [39]. PDK1 induces the phosphorylation of PLK1, and then PLK1 directly induces the accumulation of c-Myc protein in a PDK1-dependent manner. Therefore, Feng Y et al. investigated the expression of PDK1, PLK1, and c-Myc in DLBCL and found that a high expression level of PDK1 and c-Myc predicts a poor prognosis for DLBCL patients. However, the role of the PDK1-PLK1-*c*-Myc pathway in DLBCL needs further exploration in the future [40]. ATP13A2 plays a role in lipid homeostasis [41].

Thirdly, in order to explore differences in the immune microenvironment among different risk groups, we applied ssGSEA analysis to obtain the abundance of immune cell infiltration and immune function scores in all pathological samples. Additionally, we investigated the expression differences of immune checkpoint-related genes and HLA genes in two risk subgroups. Various immune checkpoint genes and HLA genes with significantly differential expression in both groups have been confirmed to be associated with the progression of DLBCL. Studies have shown that DLBCL patients with overexpression of CD274 (high-risk group) are often resistant to drugs such as rituximab, cyclophosphamide, and prednisolone [42]. This confirms the accuracy of the division into high and low-risk groups in our study. Furthermore, the expression level of CD40 in the plasma of DLBCL was significantly reduced before and after clinical treatment [43]. Moreover, changes in HLA are linked to DLBCL susceptibility. Alcoceba M and colleagues analyzed the impact of HLA-A, B, C, DRB1, and DQB1 polymorphisms on the development of 250 DLBCL patients [44].

Fourth, in order to explore the potential differences in clinical drug use between the two risk groups, this study conducted a drug sensitivity analysis and identified significant differences in the sensitivity of various drugs between the two groups. Some drugs have been confirmed to be effective in the treatment of DLBCL. Kong et al.'s clinical research confirmed the combined benefit of irditinib with the RDHAP regimen in the treatment of relapsed/refractory DLBCL [45]. Derek B Oien et al. found that in heterologous transplantation models derived from DLBCL cell lines and patients, acatnib combined with AZD5153 demonstrated synergistic benefits [46]. Cisplatin and AZD5153 showed high expression in the low-risk group, suggesting that they may have better therapeutic effects in the treatment of the low-risk group. On the other hand, JQ1 might be more effective in treating patients in the high-risk group. This information could guide stratified drug therapy for DLBCL patients in clinical practice.

Fifth, to explore miRNAs potentially associated with the prognosis of DLBCL, an interactive network involving prognosis-related genes and miRNAs was constructed in this study. The prognosis of DLBCL patients might be linked to specific miRNAs. Researchers

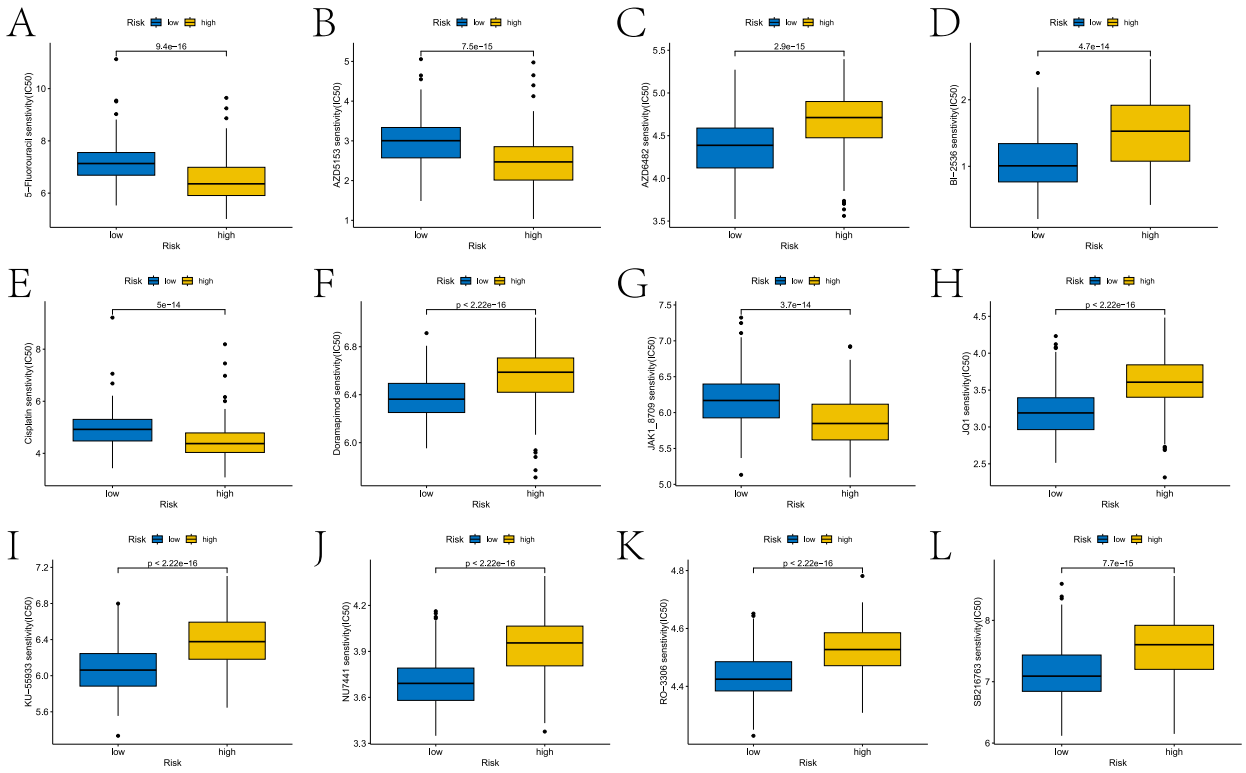


Fig. 10. Results of drug sensitivity analysis. A-L is a box plot of compounds with significantly different IC50 values between high- and low-risk subgroups.

discovered a significant increase in serum levels of miR-15a-5p in DLBCL patients [47,48]. Sromek M et al. analyzed paired cerebrospinal fluid (CSF) tumor samples from CNS DLBCL patients, revealing substantially reduced levels of miR-26a, miR-15a-5p, miR-15b-5p, miR-19a-3p, miR-106b-3p, miR-221-3p, and miR-423-5p in CSF, while these miRNAs exhibited higher levels in tumor samples [49].

Finally, the research process and results of this study are subject to certain limitations. During the construction of the prognostic model using Lasso-Cox regression, the sparse selection of variables may lead to different variable choices in different datasets or runs, challenging the stability of the model. This uncertainty could make it difficult to consistently interpret the research conclusions.

Additionally, the penalty term in Lasso-Cox regression may cause estimates to deviate from the true values. Particularly in cases of high collinearity, Lasso's coefficient estimates may tend to set the coefficients of some correlated variables to zero, introducing bias into the research results. When assessing immune cell infiltration abundance using the ssGSEA algorithm, it assumes that gene expression data are uniformly distributed across the entire gene set. However, in reality, gene expression data may not meet this assumption, leading to biased estimates of gene set enrichment levels. Furthermore, in situations with a small sample size, the results of ssGSEA may lack stability and reliability. Especially in small sample studies, the interpretation of results may be significantly affected by substantial uncertainty. It is worth noting that we have compiled the gene symbols and their full names related to this study and placed the organized table in the supplementary material file "symbol.xlsx".

5. Conclusion

DLBCL is a highly heterogeneous tumor. This study identified genes significantly associated with the prognosis of DLBCL and constructed risk models using various PCD-related genes. The two risk subgroups exhibited significant differences in immune landscape and drug sensitivity. In conclusion, the genes identified in relation to prognosis and the risk models developed in this paper can serve as valuable references for the advancement of targeted therapies and related drugs for DLBCL.

Data availability statement

The current dataset for the study can be found in the GEO database, which are publicly accessible, under the following Accession Numbers: GSE10846 and GSE31312.

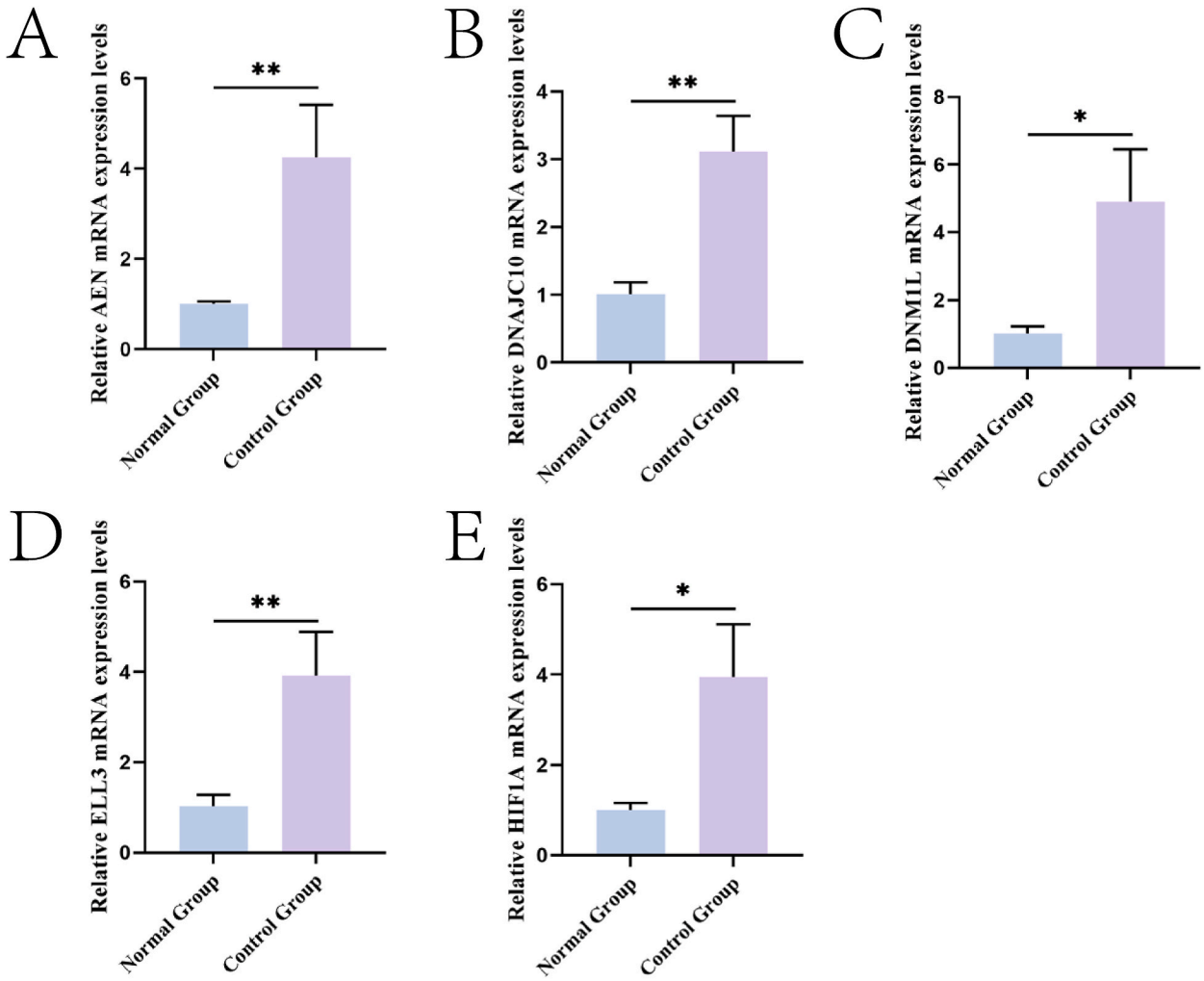


Fig. 11. The mRNA expression levesl of AEN、DNAJC10、DNM1L、ELL3 and HIF1A.

CRediT authorship contribution statement

Heyuan Feng: Writing – original draft, Visualization, Validation. **Xiyuan Zhang:** Writing – review & editing, Supervision, Data curation, Conceptualization. **Jian Kang:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e30831.

References

[1] P. Dawiec, et al., Automatic subtyping of Diffuse Large B-cell Lymphomas (DLBCL): Raman-based genetic and metabolic classification, Spectrochim. Acta Mol. Biomol. Spectrosc. 309 (2024) 123795.
[2] S. Poletto, et al., Treatment strategies for patients with diffuse large B-cell lymphoma, Cancer Treat Rev. 110 (2022) 102443.
[3] D. Ennishi, et al., Toward a new molecular taxonomy of diffuse large B-cell lymphoma, Cancer Discov. 10 (9) (2020) 1267–1281.

[4] Y. Zou, et al., Leveraging diverse cell-death patterns to predict the prognosis and drug sensitivity of triple-negative breast cancer patients after surgery, Int. J. Surg. 107 (2022) 106936.
[5] X. Wang, et al., Exploring the role of different cell-death-related genes in sepsis diagnosis using a machine learning algorithm, Int. J. Mol. Sci. 24 (19) (2023).
[6] X. Lin, et al., Focus on ferroptosis, pyroptosis, apoptosis and autophagy of vascular endothelial cells to the strategic targets for the treatment of atherosclerosis, Arch. Biochem. Biophys. 715 (2022) 109098.
[7] M.S. D'Arcy, Cell death: a review of the major forms of apoptosis, necrosis and autophagy, Cell Biol. Int. 43 (6) (2019) 582–592.
[8] Y. Fang, et al., Pyroptosis: a new frontier in cancer, Biomed. Pharmacother. 121 (2020) 109595.
[9] S. Xu, et al., Selenium deficiency causes iron death and inflammatory injury through oxidative stress in the mice gastric mucosa, Biol. Trace Elem. Res. 202 (3) (2024) 1150–1163.
[10] P. Tsvetkov, et al., Copper induces cell death by targeting lipoylated TCA cycle proteins, Science 375 (6586) (2022) 1254–1261.
[11] S. Wen, et al., Androgen receptor (AR) positive vs negative roles in prostate cancer cell deaths including apoptosis, anoikis, entosis, necrosis and autophagic cell death, Cancer Treat Rev. 40 (1) (2014) 31–40.
[12] Y. Zhou, et al., Parthanatos and its associated components: promising therapeutic targets for cancer, Pharmacol. Res. 163 (2021) 105299.
[13] A.L. Berg, et al., Engaging the lysosome and lysosome-dependent cell death in cancer, in: Breast Cancer, H.N. Mayrovitz (Eds.), Exon Publications Copyright: the Authors.; the Authors Confirm that the Materials Included in This Chapter Do Not Violate Copyright Laws. Where Relevant, Appropriate Permissions Have Been Obtained from the Original Copyright Holders, and All Original Sources Have Been Appropriately Acknowledged or Referenced, 2022. Brisbane (AU).
[14] Q. Jia, et al., Neuroprotective effects of chaperone-mediated autophagy in neurodegenerative diseases, Neural Regen Res 19 (6) (2024) 1291–1298.
[15] D. Que, et al., ACSS2-mediated NF-κB activation promotes alkaliptosis in human pancreatic cancer cells, Sci. Rep. 13 (1) (2023) 1483.
[16] J. Shen, et al., Different types of cell death in diabetic endothelial dysfunction, Biomed. Pharmacother. 168 (2023) 115802.
[17] D. Xiong, et al., Prediction significance of autophagy-related genes in survival probability and drug resistance in diffuse large B-cell lymphoma, Aging (Albany NY) 15 (2024).
[18] J. Wang, et al., Identification of a novel model based on ferroptosis-related genes for predicting the prognosis of diffuse large B-cell lymphomas, Hematology 28 (1) (2023) 2198862.
[19] W.E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods, Biostatistics 8 (1) (2007) 118–127.
[20] T. Emura, S. Matsui, H.Y. Chen, compound.Cox: univariate feature selection and compound covariate for predicting survival, Comput. Methods Progr. Biomed. 168 (2019) 21–37.
[21] A. Farcomeni, A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion, Stat. Methods Med. Res. 17 (4) (2008) 347–388.
[22] M. Ashburner, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (1) (2000) 25–29.
[23] Y. Zhou, et al., Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat. Commun. 10 (1) (2019) 1523.
[24] G.E. Dinse, S.W. Lagakos, Nonparametric estimation of lifetime and disease onset distributions from incomplete observations, Biometrics 38 (4) (1982) 921–932.
[25] R. Tibshirani, The lasso method for variable selection in the Cox model, Stat. Med. 16 (4) (1997) 385–395.
[26] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (43) (2005) 15545–15550.
[27] E. Becht, et al., Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, Genome Biol. 17 (1) (2016) 218.
[28] Z. Liu, X.S. Zhang, S. Zhang, Breast tumor subgroups reveal diverse clinical prognostic power, Sci. Rep. 4 (2014) 4002.
[29] H.Y. Huang, et al., miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions, Nucleic Acids Res. 50 (D1) (2022) D222–d230.
[30] K. Lu, et al., A novel silicone derivative of natural osalmid (DCZ0858) induces apoptosis and cell cycle arrest in diffuse large B-cell lymphoma via the JAK2/STAT3 pathway, Signal Transduct. Targeted Ther. 5 (1) (2020) 31.
[31] Q. Feng, et al., Berberine derivative DCZ0358 induce oxidative damage by ROS-mediated JNK signaling in DLBCL cells, Int. Immunopharm. 125 (Pt A) (2023) 111139.
[32] Z. Zhang, et al., A novel lipid metabolism-based risk model associated with immunosuppressive mechanisms in diffuse large B-cell lymphoma, Lipids Health Dis. 23 (1) (2024) 20.
[33] A.V. de Jonge, et al., Distinct peripheral T-cell and NK-cell profiles in HGBL-MYC/BCL2 versus DLBCL NOS patients, Blood Adv 8 (5) (2024) 1094–1104. https://doi.org/10.1182/bloodadvances.2023011687.
[34] Q. Sun, et al., MORTALIN-Ca(2+) axis drives innate rituximab resistance in diffuse large B-cell lymphoma, Cancer Lett. 537 (2022) 215678.
[35] S. Monti, et al., Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response, Blood 105 (5) (2005) 1851–1861.
[36] M. Broecker-Preuss, et al., Impact of germline polymorphisms in genes regulating glucose uptake on positron emission tomography findings and outcome in diffuse large B-cell lymphoma: results from the PETAL trial, J. Cancer Res. Clin. Oncol. 148 (10) (2022) 2611–2621.
[37] Y. Xie, et al., Identification of an individualized RNA binding protein-based prognostic signature for diffuse large B-cell lymphoma, Cancer Med. 10 (8) (2021) 2703–2713.
[38] A.I. Cioroianu, et al., Tumor microenvironment in diffuse large B-cell lymphoma: role and prognosis, Anal. Cell Pathol. 2019 (2019) 8586354.
[39] W. Wu, et al., Prognostic significance of ribosome-related genes signature in diffuse large B cell lymphoma, J. Cancer 14 (3) (2023) 403–416.
[40] Y. Feng, et al., Investigation of expressions of PDK1, PLK1 and c-Myc in diffuse large B-cell lymphoma, Int. J. Exp. Pathol. 100 (1) (2019) 32–40.
[41] A.L. Marcos, et al., The Parkinson-associated human P5B-ATPase ATP13A2 modifies lipid homeostasis, Biochim. Biophys. Acta Biomembr. 1861 (10) (2019) 182993.
[42] X. Wu, et al., IGH::CD274 (PD-L1) rearrangement in diffuse large B cell lymphoma and its therapeutic implication, EJHaem 4 (2) (2023) 442–445.
[43] L.E. Martínez, et al., Plasma extracellular vesicles bearing PD-L1, CD40, CD40L or TNF-RII are significantly reduced after treatment of AIDS-NHL, Sci. Rep. 12 (1) (2022) 9185.
[44] M. Alcoceba, et al., HLA specificities are related to development and prognosis of diffuse large B-cell lymphoma, Blood 122 (8) (2013) 1448–1454.
[45] X. Kong, et al., Decitabine combined with RDHAP regimen in relapsed/refractory diffuse large B cell lymphoma, Cancer Med. 12 (7) (2023) 8134–8143.
[46] D.B. Oien, et al., BET inhibition targets ABC-DLBCL constitutive B-cell receptor signaling through PAX5, Blood Adv 7 (17) (2023) 5108–5121.
[47] C. Fang, et al., Serum microRNAs are promising novel biomarkers for diffuse large B cell lymphoma, Ann. Hematol. 91 (4) (2012) 553–559.
[48] A. Beheshti, et al., Identification of circulating serum multi-MicroRNA signatures in human DLBCL models, Sci. Rep. 9 (1) (2019) 17161.
[49] M. Sromek, et al., A set of 17 microRNAs common for brain and cerebrospinal fluid differentiates primary central nervous system lymphoma from non-malignant brain tumors, Biomolecules 11 (9) (2021).