The Identification and Analysis of a Novel Model Based on Ferroptosis-Related Genes for Predicting the Prognosis of Diffuse Large B-Cell Lymphomas

Jiayi Wang
Central South University
Hongling Peng
Central South University
Guangsen Zhang
Central South University
Yunxiao Xu
Central South University
Wenzhe Yan (ywzwjy8890@csu.edu.cn)
Central South University

Research Article

Keywords: DLBCL, prognosis, ferroptosis, signature

Posted Date: December 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1119694/v1

Abstract

Background

Diffuse large B-cell lymphomas (DLBCLs) are the most common B-cell lymphoma featured as phenotypically and genetically heterogeneous. Ferroptosis is a new found programmed cell death and have a crucial role in the chemoresistance of tumor. We aim to build a ferroptosis-related genes (FRGs) prognostic signature to predict the outcome of DLBCLs.

Methods

Our study retrospectively investigated the mRNA expression level and clinical data of 604 DLBCL patients from 3 GEO public datasets. A series of bioinformatic approaches including Cox regression analysis, function enrich analysis, immune infiltration analysis, ROC curve analysis, Kaplan–Meier survival curve and the least absolute shrinkage and selection operator (LASSO) method by the corresponding R packages in R statistical software were combined to explored the heterogenicity of FRG based clusters and to build prognostic model. Immunohistochemistry was used to exam the protein expression of six FRGs in different molecular type of DLBCL.

Results

We first identified 19 FRGs with potential prognostic values and classfied the patients into cluster 1 and cluster 2, Results indicated that cluster 1 tend to have a shorter overall survival (OS) time, while patients in the two clusters have different patterns of infiltrating immune cells among. Furthermore, the LASSO was used to generated a six-genes (*GCLC*, *LPCAT3*, *NFE2L2*, *ABCC1*, *SLC1A5*, and *GOT1*) risk signature which constructed a risk score formula and prognostic model for the OS of DLBCL patients. Kaplan–Meier survival analysis proved that poorer OS was exhibited in higher risk patients stratified by the prognostic model in both the training cohort and test cohort. In addition, we constructed nomograms to predict the OS of DLBCL patients. Both the decision curve(DCA) and the calibration plots showed that the nomogram had good agreement between predicted results and actual observation. Finally, the validation by immunohistochemistry indicated the *GCLC*, *LPCAT3, ABCC1*, *SLC1A5*, and *GOT1* were high expressed in DLBCL with various prognostic adverse molecular factor.

Conclusion

In sum, we built a new FRG-based prognostic model which will help improve diagnosis and treatment for DLBCL patients.

Background

DLBCLs are heterogeneous disease comprising intermediate- and high-grade B cell lymphomas with various clinical, pathological, and molecular markers. Quickly determining which patients have poorer prognoses allows patients to be considered for novel targeted-treatment strategies faster. Clinical features are used to determine the International Prognostic Index (IPI) and the outcome in DLBCL, but in the tine of rituximab targeted therapy, the IPI could only interprets less than 25% of variation in OS and limited separates the highest-risk subgroups[1]. When it comes to molecular factors, three subtypes of DLBCLs, classified as activated B-cell–like [ABC], germinal-center B-cell–like [GCB], and unclassified, has been identified according to cell of origin and are widely used in clinical application[2]. Patients with ABC style are featured by a more aggressive profile and activation of NF-κB and B cell antigen receptor signaling pathways[3, 4], while GCB tumors are often related with aberrant chromatin-modification, PI3K signaling, and the overexpression of Bcl-2 and MYC through copy number gains or chromosomal translocations[5, 6]. Currently, researchers also identified four crucial genotypes in DLBCL with different mutant pattern, epigenetic, and clinical features: MCD defined as the co-occurrence of MYD88L265P and CD79B mutations, EZB defined as EZH2 mutations and Bcl-2 translocations, N1 defined as NOTCH1 mutations, and BN2 defined as Bcl-6 fusions and NOTCH2 mutations[7]. These subtypes are instructional for precision-medicine strategies, however, due to the characteristics of the gene-expression profiling subgroup, the prognosis prediction of DLBCL patients is still cursory and unsatisfactory and necessitates new biomarkers for accurate prognosis prediction in DLBCL patients.

Ferroptosis is a novel form of iron-based cell death characterized by lethal excessive accumulation of iron and lipid peroxidation[8, 9]. It is different from other well-known cell death including autophagy, necrosis, and apoptosis. Since it was firstly defined in 2012, many genes had been considered as the regulator or markers of ferroptosis. Increasing researches also proved that the new form of cell death is related to the tumor suppression of hematological tumor[10, 11]. A study of 114 cancer cell lines proved that DLBCLs were highly vulnerable to ferroptosis inducers-erastin[12]. Furthermore, ferroptosis may be induced by dimethyl fumarate, which is a promising option in the treatment of GCB DLBCLs[13]. Hence, we aimed to dissect the expression patterns of FRGs in the DLBCL patients to facilitate risk stratification.

In this study, we made use of trancriptome profiles and corresponding clinical parameters of DLBCL patients from GEO public databases to analyze the differential expression of FRGs to identify the function enriched pathways and their biological functions to investigate the potential biological mechanisms in DLBCL samples. Furthermore, we also constructed and validated a multigene prognostic model with six ferroptosis-related genes in the training and testing cohort, which might provide support to improving the treatment strategies of DLBCL.

Methods

Data sources

All datasets utilized in this study are publicly available. The mRNA expression data and corresponding clinical information of the three independent cohorts (GSE10846, GSE11318, GSE4475) from the Gene

Expression Omnibus (GEO) Database were systematical screening. The clinical data of the DLBCL patients are shown in Table 1. All patients with unclear clinicopathological characteristics, clinical stage, OS and survival status were excluded. Thereafter, GSE10846 and GSE11318 RNA-seq data were merged into one new 447 sample cohort. 155 samples from GSE4475 were used as an external validation cohort. All data were collected on 14 April 2021. The baseline characteristics of DLBCL patients were collected in the Table 1.

Table 1
The clinical data of DLBCL patients in the GEO dataset.

Clinical Features	GSE11318 & GSE10846	GSE4475
Gender: Female/male/NA	189/242/18	67/88/0
Age: >65/≤65/NA	181/231/37	69/86/0
Stage: I/II/III/IV/NA	66/122/96/121/44	25/39/46/31/14
OS: 1/0	179/270	73/82

Comprehensive analyses of FRG based clusters in DLBCL patients

We reviewed published high level literature on the subject of ferroptosis over the past 3 years and identified sixty FGRs[9, 14, 15]. The two cohorts were screened for the expression of sixty FRGs. Univariate Cox regression analysis was executed to explore the correlation between FRGs and survival data using the R package(Additional file 1). $P < 0.05$ was set as the filtering threshold, and 19 genes were determined to related with the prognosis of DLBCL. Next, we use the “ConsensusClusterPlus” software package to perform consensus clustering on the 19 genes. The kilometer and Euclidean distance were defined as cluster analysis and similarity measure. We further analyzed the expression of the 19 FGRs and survival rate by calculating the DEGs within the two clusters using the limma package. The filtering threshold was set at FDR < 0.05 and |log2FC| > 1.2. Further function analysis of biological processes (BP), molecular functions (MF), and cellular components (CC) regulated by the differently expressed genes in the two clusters were analyzed based on Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) data using R software, Clusterprofiler package. We used ESTIMATE package in R to evaluate the Stromal Score, Immune Score, and ESTIMATE Score. We also used MCPcounter to determine 10 immune cells scores and ssgsea method in GSVA package to assess immune cells scores.

Establishment and validation of FRG signature

449 samples from GSE10846 and GSE11318 were separated into the training set (70% (317)) and the testing set (30% (132)). Firstly, we utilized the R software package “survival” to execute univariate Cox regression analysis on the FRGs with significant prognostic values, and then performed the LASSO cox regression analysis to eliminate overfitting. Furthermore, we executed multivariate Cox regression analysis to screen the genes selected as independent prognostic factors of OS, and then calculated the

regression coefficients of the genes to select an optimal prognostic model. The risk score of every DLBCL patient was also evaluated according to the expression level and coefficient of the six genes, and the formula was presented as: Risk score = $coeffificient_{gene1} \times expression_{gene\ 1} + coeffificient_{gene\ 2} \times expression_{gene\ 2} + \ldots \ldots coeffificient_{gene\ n} \times expression_{gene\ n}$.

All samples were departed into two subgroups as high-risk group and low-risk group according to the calculated median risk score. Subsequently, we executed the Kaplan–Meier survival analysis to contrast the survival probability between the two subgroups. The validity and effectiveness of prediction in one year, three years, and five years of the polygenic signature was estimated by receiver operating characteristic(ROC) analysis using “pROC” package. Furthermore, patients in different age, gender, or disease stage groups were tested by this formula and each group could be further departed into two subgroups as high-risk and low-risk.

Besides, we also calculated the risk score of DLBCL patients in the testing cohorts and the external dataset GSE4475 to certify the robustness of the predictive model by the same signature and coefficients. Then we constructed the Kaplan–Meier survival and survival ROC curve to demonstrate the predictive capability of the model in the internal and external validation cohorts.

Construction of a clinical FRGs nomogram for DLBCL Patients

First, the prognostic value of multi-gene signature and different variables like age, gender and clinical stage were assessed by performing univariate Cox regression analysis. We also conducted multivariate Cox regression analysis to confirm whether the risk score is the independent prognostic factor for OS of DLBCL. We further employed the “RMS”and “Survival” package to fit regression model and built nomogram. A correction curve was depicted to estimate the coherence between the nomogram predicted results and actual observation. Calibration curves of nomograms were calculated to evaluate the prediction accuracy of the nomogram.

Immunohistochemistry

15 tumor tissues were achieved from the hematology department of Second Xiangya Hospital. Tissue sections were incubated at 60℃ for 2 hours and then in xylene and in a series of descending ethanol concentrations for rehydration, The slides were boiled in sodium citrate in a microwave oven for 20 minutes, and then the 3% hydrogen peroxide solution was used to treat the slides for 5 minutes and 5% goat serum was used to block the non-specific antigen for half hour. Antibodies against *GCLC* (Affinity, USA), *LPCAT3* (Abcam, UK), *NFE2L2* (Affinity, USA), *ABCC1* (Affinity, USA), *SLC1A5* (Affinity, USA), and *GOT1* (Affinity, USA) incubate the sections for 1 hour. The slides was incubated by secondary antibody, then stained with 3,30 -diaminobenzidine and with hematoxylin, soaked in warm water for 10 minutes, dehydrated, incubated in dimethylbenzene, and mounted.

Statistical analysis

SPSS, R software and the GraphPad Prism 7 were employed for the data processing and statistical analyses. Univariate and multivariate Cox regression analyses, differential expression analysis, ROC curve analysis, and Kaplan–Meier survival analysis were constructed by the corresponding R packages in R software to determine the difference in the OS between the stratified groups. Wilcoxon test was performed to compare the two groups, and Kruskal–Wallis test was employed to compare prognosis between groups. For each analysis, $p < 0.05$ was considered as statistically significant.

Results

Molecular classification and verification based on FRGs

We determined 19 FGRs that were associated with the prognosis of DLBCL (Fig. 1A). The result of k = 2 from unsupervised consensus analysis of all samples seemed to be more accurate according to the CDF curve, which could separate all the samples into two subgroups with the least amount of correlation between groups (Fig. 1B, C). To find out the correlation between the clinical outcome and clustering result, we contrasted the OS between the two clusters of DLBCL patients and showed that cluster 1(C1) patients had poorer OS ($p < 0.0001$) than cluster 2(C2) (Fig. 1D), The heatmap shows that the 19 FRGs were expressed differently in the two clusters (Fig. 1E). Meanwhile there was no significant difference of age, gender, and stage between the two subtypes(Fig. 1F).

Identification and functional annotation analyses of DEGs

We detected the differential expression profile between C1 and C2, and uncovered that 248 genes were up-regulated, 1,242 genes were down-regulated (Additional file 2) in C2 than C1($p < 0.05$, $|\log2FC| > 1.2$). The representative 100 DEGs is displayed as volcano plot and heatmap in Figure 2A, 2B. These 1,490 DEGs were investigated by GO annotation and KEGG pathway enrichment to search their functions. As the chord plots show (Fig. 2C–2E), the main biological processes (BP) participating were the regulation of ion membrane transport and the modulation of chemical synaptic transmission, trans-synaptic signaling, and metal ion transport. The enriched cellular component (CC) terminology were the apical plasma membrane, collagen containing extracellular matrix, transmembrane transport complex, and ion channel complex. The molecule function (MF) mainly regulated amide binding, gated channel activity, and peptide binding (Additional file 3). Biological function related to iron ion molecular, such as ion channel and ion gated channel activities were significant enriched. As KEGG pathway analysis revealed (Fig. 2F, Additional file 4), the calcium signaling pathway, the PI3K-Akt signaling pathway, and the MAPK signaling pathway which are crucial for cellular proliferation and cell cycle regulation of lymphoma were significantly enriched[16].

FRG based cluster was significantly associated with immune function

We compare the immunity scores between the two ferroptosis subtypes demonstrated significant differences of tumor microenvironment (TME), including Stromal Score ($p < 0.001$) and Immune Score ($p < 0.001$). The C2 reflected higher Stromal Score and ESTIMATE Scores compare to C1(Fig. 3A). The Stromal Score present the frequency of stromal components in TME and ESTIMATE Score was the sum of Immune Score and Stromal Score reflecting the comprehensive proportion of both stromal components and immune cells in TME. MCPcounter analysis unveiled that immune scores of CD8 T cells, B lineage, NK cells, and neutrophils were significantly higher in C1 than C2, while the immune scores of cytotoxic lymphocytes and endothelial cells were significantly higher in C2 than C1(Fig. 3B). Implying that ferroptosis may have a profound impact on TME in DLBCL. Ssgsea analysis showed that activated B cell, activated dendritic cell, MDSC, NK T cell, effector memory CD8 T cell, and plasmacytoid dendritic cell scores had a higher abundance in C1 than C2, while activated CD4 T cell, activated CD8 T cell, effector memory CD4 T cell, type 2 T helper cell, type 17 T helper cell, regulatory T cell, gamma delta T cell, CD56 bright NK cell, macrophage, mast cell, NK cell, eosinophil, and neutrophil scores were higher in C2 than C1 (Fig. 3C). Furthermore, we plot immune score heatmaps between two molecular subgroups (Fig. 3D).

Construction of FRGs signature for OS in the training cohort

Seven FRGs were identified as associated with OS (Fig. 4A, Additional file 5), and the final multigene prognostic signature for OS used a six-gene signature composed of *GCLC*, *LPCAT3*, *NFE2L2*, *ABCC1*, *SLC1A5*, and *GOT1* based on the optimal value of lammda (0.008845861) (Fig. 4B). The prognostic Kaplan–Meier survival analysis of these six genes showed that the high expression of *GCLC*, *LPCAT3*, *SLC1A5*, *ABCC1* and *GOT1* were associated with poor prognosis, while *NFE2L2* was related to good outcome (Fig. 4C). The risk score was calculated as follows: (*GCLC* × 0.47) + (*LPCAT3* × 0.1) + (*NFE2L2* × -0.31) + (*ABCC1* × -0.1) + (*SLC1A5* × 0.2) + (*GOT1* × 0.35). The DLBCL patients were stratified into two subgroups, a high-risk group (n=158) or a low-risk group (n=159), according to the median cut-off value ($p = 0.00071$) (Fig. 4D). As shown in Figures 4E, A higher ratio of patient died in the high-risk group. The expression level of the six FRGs in two group were presented as heatmap(Fig. 4F). Consistently, the Kaplan-Meier survival curve calculated that the high-risk group had a poorer OS than the low-risk group (Fig. 4G). The time-dependent ROC curves estimated the predictive performance of the risk score for OS. The area under the curve (AUC) reached 0.659 at one year, 0.694 at three years, and 0.699 at five years (Fig. 4H).

Internal and external validation of the prognostic signature in the verification group

To evaluate the robustness of the six-genes signature constructed in the training cohort, the DLBCL patients from the verification cohort were also separated into two risk groups by the median value generated by the same formula performed in the training cohort. Similar to the results obtained in the training cohort. Patients with high-risk tend to correlated with worse OS status than those with low-risk($p<0.0001$), and the ROC curve of the prognostic signature identified that AUC was 0.718 at one year,

0.742 at three years, and 0.742 at five years which clarified a satisfactory accuracy of the model (Fig. 5A-5E). Furthermore, the patients from the independent cohort (GSE4475) were also categorized into two risk groups using the same formula (Fig. 5F), and the OS status based on risk score and heatmap of the expression level of the six genes was distributed in Figure 5G and 5H. The results were consistent with those obtained from both the testing cohort and the training cohort. The K-M plot presented that the DLBCL patients in high-risk group had a poorer OS compared with those in the low-risk group(p=0.0016) (Fig. 5J), and the ROC curve analysis showed that AUC of the six-gene signature was 0.612 at one year, 0.628 at three years, and 0.612 at five years (Fig. 5I).

Analysis of the risk model and clinical characteristics

The risk score analysis in the training cohort found that the six-gene prognostic signature could also categorize the diverse age, gender, and stage groups into high- and low-risk. The DLBCL patients were stratified by age (>65y subtype, <65y subtype), gender (female subtype, male subtype) and stage (stage I/II subtype, stage III/IV subtype). All subtypes were tested by the FRG based model and showed significant shorter OS in the high-risk group (Fig. 6A-6F). These results further indicated that our model has good predictive function in different clinical signs. The risk score was also compared between different age/gender/stage/molecular subtypes. The results showed the risk score was significant higher in C1 than C2, as well as higher in >65y subtype than <65y subtype. (Fig. 6G-6J)

Construction of a nomogram of DLBCL patients

The Univariate Cox regression analyses indicated that risk score of the six-genes signature were significantly related to the OS status of DLBCL patients and that the prognostic signature could successfully predict OS in the merged cohort (Fig. 7A). Then the multivariate Cox regression analysis results also identified that risk score of the prognostic signature (HR = 3.1, p < 1e-5) were associated with the OS (Fig. 7B). Nomograms show risk model results intuitively and conveniently, with the length of the lines representing the impact of different variables on the outcome. We incorporated a multivariate analysis of the results which included clinical characteristics and the novel FRGs prognostic signature together, to generate a hybrid nomogram model for predicting the one-, three-, and five-year death rate (Fig. 7C). The nomogram showed that the risk score contributes more to the total score than other variables. The one-, three-, and five-year death rate of patients increased as the total score increased. The calibration plots approached 45° and showed satisfactory agreement between the predicted OS rates and actual observations at one, three and five years (Fig. 7D), indicating that the nomogram performed stable and accurate. The risk score and the nomogram both had good reliability based on the DCA. Rather than a single independent predictive factor, the hybrid nomogram could obtain the optimal net benefit at one, three and five years (Fig. 7E).

Expression levels of key genes in the DLBCL samples with different molecular feature

To investigate the clinical significance of this six genes, we used immunohistochemistry to exam the expression of the six key genes in DLBCL samples with different prognostic molecular feature, which include DLBCL with ABC type, DLBCL with GCB type, DLBCL with double hit(DLBCL-DH), DLBCL with double expressors(DLBCL-DE) and DLBCL with TP53 mutant(DLBCL-TP53). The result showed that the protein expression levels of *GCLC*, *LPCAT3*, *ABCC1*, *SLC1A5*, and *GOT1* are expressed relatively higher in DLBCL-ABC, DLBCL-DH, DLBCL-DE, DLBCL-TP53 than in DLBCL-GBC samples and the *NFE2L2* had a higher expression in DLBCL-GCB than that of DLBCL-ABC, DLBCL-DH, DLBCL-DE, DLBCL-TP53. (Fig. 8)

Discussion

DLBCL is known as a clinical and molecular heterogeneous malignant hematological tumor. Various classification methods, and molecular markers have been established to characterize this disease[17]. To develop practical molecular markers related to DLBCL prognosis, we have identified a signature of six FRGs by using high-throughput expression analysis. Ferroptosis is a novel explored programmed cell death modality. It could reverse resistance of tumor cells to chemotherapy drugs and as well promote removal of defective cells[18]. Consequently, ferroptosis is potentially a novel mechanism for tumor therapy. Several researches have concentrated on the role of ferroptosis in DLBCL treatment and development[12, 13].

In our study, based on the expression profile of FRGs, consensus clustering analysis divided the DLBCL patients in two clusters. The results showed obvious differences in OS between the two clusters. C1 patients had a much worse prognosis than C2. Further analyses uncovered DEGs between the two clusters. KEGG pathway analyses unveiled the genes participated in PI3K-Akt pathway and the calcium signaling pathway. Both of these pathways participate in the pathogenesis of lymphoma. The PI3K-Akt pathway is frequently activated in a variety of solid tumors and hematological malignancies, hence, PI3K was considered as an attractive therapeutic target in oncology. Currently, two PI3K inhibitors, copanlisib and idelalisib, have been approved for use in the leukaemias and B cell lymphoma[19–21]. Calcium signaling also paticipate the GA101-induced cell death in lymphoma cells[22]. Recent evidence indicates that Ca2+ ions play an important role in cell death mediated by oxidative glutamate toxicity or oxytosis, a form of programmed cell death similar to ferroptosis[23]. The GO enrichment mainly included several iron-related biological processes or molecular functions, such as ion gated channel activity and ion channel activity. We speculated that ferroptosis was related to ion transport in DLBCL pathogenesis. In this study, we also found the C1 group have a smaller ratio of the stromal components. MCPcounter analysis showed a larger percentage of CD8 T cells, B lineage, NK cells, and neutrophils in C1 than C2, and ssgsea analysis presented a larger ratio of activated B cell, effector memory CD8 T cell, activated dendritic cell, MDSC, NK T cell, and plasmacytoid dendritic cell scores in TME. Among these immune cell types, a higher proportion of NK cell was correlated with poorer DLBCL outcome, but dendritic cells might contribute to longer OS[24]. Since the main feature of MDSC is their potent immune suppressive activity, the stronger immunosuppressive effect of MDSC might contribute to the poor prognosis of the high-risk group[25]. Regulatory T cells and CD4+ T cells, which had a correlation with improved survival in DLBCL

has a lower ratio in the C1 subtype[24, 26]. These results confirm that ferroptosis has a regulatory effect on the TME.

Previous studies have proved that ferroptosis-inducer erastin is useful for inducing cell death in 114 DLBCL cell lines[12]. Ferroptosis might also be induced by dimethyl fumarate which is a promising novel therapeutic option in the treatment of GCB DLBCL, but the correlation between ferroptosis and DLBCL patients' OS remains largely unclear. In this study, we identified the novel risk scoring model constructed by six genes (*GCLC*, *LPCAT3*, *NFE2L2*, *ABCC1*, *SLC1A5*, and *GOT1*) to classify DLBCL patients into two classes and predicted the OS of DLBCL patients independently. Gene markers related to ferroptosis were established, and the expression level of the six genes are not influenced by the variances in the underlying diseases of DLBCL, indicating that the constructed prognosis signature could be well performed in different types of DLBCL patients. Moreover, the hybrid nomogram based on the six-gene signature also provide support to clinicians to develop optimal treatment strategies. By concentrating on the function of the six ferroptotic genes, former researches have showed that the six genes play key roles in the development of cancer cells.

GCLC is a rate-limited enzyme that primarily regulates *de novo* synthesis of glutathione. It has been revealed that activation of GCLC is related to drug resistance in various tumors, like breast, head, lung, and neck cancers[27–29]. LPCAT3 is an enzyme that primarily maintain systemic homeostasis by transfer lysophosphatidylcholine to phosphatidylcholine in the liver. It also participates in the intestinal stem cell growth and phospholipid remodeling and tumorigenesis[30, 31]. NFE2L2 is a main regulator of the antioxidant response and regulates the activity of several ferroptosis and lipid peroxidation-related proteins[32]. ABCC1 plays a pivotal role in protecting cells by removal a vast array of drugs to sub-lethal levels. In the last decade many studies clarified the mechanisms of function and molecular characteristics of ABCC1 [33]. SLC1A5 is a cell surface solute-carrying transporter that facilitates uptake of neutral amino acids[34]. Blocking SLC1A5 to prevent glutamine uptake successfully restrains tumor cell proliferation in melanoma[35], breast cancer[36], and acute myeloid leukemia[37]. GOT1 plays an active role in ROS balance and energy metabolism in chronic acidosis stress[38]. Few studies on the role of these genes played in DLBCL patients' prognosis have been reported and their influence on the development of ferroptosis needs to be clarified. We further evaluated the protein expression level of the six genes in the DLBCL clinical samples with different molecular types. Double-hit lymphoma, double-expressor lymphoma and DLBCL with TP53 mutant were all known for poor outcome, long term survivors are rare[39, 40], and GCB DLBCL have more favorable outcomes than those with ABC DLBCL when treated with standard immunochemotherapy[41]. In our results, the DLBCL with adverse molecular factor, which include DH, DE, TP53 and ABC type, manifested a more positive expression of *GCLC*, *LPCAT3*, *ABCC1*, *SLC1A5* and *GOT1* than the GCB DLBCL, while *NFE2L2* had the reverse tendency. These results supported that *GCLC*, *LPCAT3*, *ABCC1*, *SLC1A5* and *GOT1* are inclined to be the adverse prognostic biomarkers, while the *NFE2L2* is apt to be a protective factor in DLBCL.

Conclusion

In summary, our research used a large database to establish a novel six-gene signature model related to FRGs for predicting the prognosis of DLBCL patients. This prognostic model provides potential value for guiding the therapy and offer insights into the identification of potential biomarker of outcome in DLBCL. But we still need further intensive experimental verification and prospective cohort studies to prove the potential mechanism of FRGs in DLBCL.

Abbreviations

DLBCLs: Diffuse large B-cell lymphomas

FRG: ferroptosis-related gene

OS:overall survival

DCA:decision curve

IPI:International Prognostic Index

ABC:activated B-cell–like

GCB:germinal-center B-cell–like

GEO:Gene Expression Omnibus

BPbiological processes

MF:molecular functions

CC:cellular components

KEGG:Kyoto Encyclopedia of Genes and Genomes

GO:Gene Ontology

LASSO:least absolute shrinkage and selection operator

ROC:operating characteristic

TME:tumor microenvironment

AUC:area under the curve

DH:double hit

DE:double expressors

Declarations

Ethics approval and consent to participate

All methods were performed in accordance with the relevant guidelines and regulations

Consent for publication

Not applicable

Availability of data and materials

The gene expression profile and corresponding clinical information of the three independent cohorts (GSE10846 GSE11318 GSE4475) were retrieved from the Gene Expression Omnibus (GEO) Database (https://www.ncbi.nlm.nih.gov/geo/)

Competing interests

The authors declare that they have no competing interests

Funding

This study was supported by the National Natural Science Foundation of China (81900170) and Natural Science Foundation of Hunan(2020JJ5840)

Authors' contributions

W.Z.Y designed and performed research,analysed data and wrote the paper, J.Y.W performed research, collected and analyzed data and wrote the paper; H.L.P provided data collection; Y.X.X performed the experiment; G.S.Z contributed some study suggestions and reviewed the manuscript. All authors read and approved the final manuscript.

Acknowledgments

Not applicable

Data Availability Statement

All data included in this study are available upon request by contact with the corresponding author.

References

1. Zhou, Z., et al., *An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era.* Blood, 2014. **123**(6): p. 837-42.

2. Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.* Nature, 2000. **403**(6769): p. 503-11.
3. Compagno, M., et al., *Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma.* Nature, 2009. **459**(7247): p. 717-21.
4. Davis, R.E., et al., *Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma.* Nature, 2010. **463**(7277): p. 88-92.
5. Aukema, S.M., et al., *Double-hit B-cell lymphomas.* Blood, 2011. **117**(8): p. 2319-31.
6. Erdmann, T., et al., *Sensitivity to PI3K and AKT inhibitors is mediated by divergent molecular mechanisms in subtypes of DLBCL.* Blood, 2017. **130**(3): p. 310-322.
7. Schmitz, R., et al., *Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma.* N Engl J Med, 2018. **378**(15): p. 1396-1407.
8. Dixon, S.J., et al., *Ferroptosis: an iron-dependent form of nonapoptotic cell death.* Cell, 2012. **149**(5): p. 1060-72.
9. Stockwell, B.R., et al., *Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease.* Cell, 2017. **171**(2): p. 273-285.
10. Shen, Z., et al., *Emerging Strategies of Cancer Therapy Based on Ferroptosis.* Adv Mater, 2018. **30**(12): p. e1704007.
11. Tang, R., et al., *Ferroptosis, necroptosis, and pyroptosis in anticancer immunity.* J Hematol Oncol, 2020. **13**(1): p. 110.
12. Yang, W.S., et al., *Regulation of ferroptotic cancer cell death by GPX4.* Cell, 2014. **156**(1-2): p. 317-331.
13. Schmitt, A., et al., *Dimethyl fumarate induces ferroptosis and impairs NF-kappaB/STAT3 signaling in DLBCL.* Blood, 2021. **138**(10): p. 871-884.
14. Bersuker, K., et al., *The CoQ oxidoreductase FSP1 acts parallel to GPX4 to inhibit ferroptosis.* Nature, 2019. **575**(7784): p. 688-692.
15. Hassannia, B., P. Vandenabeele, and T. Vanden Berghe, *Targeting Ferroptosis to Iron Out Cancer.* Cancer Cell, 2019. **35**(6): p. 830-849.
16. Sapon-Cousineau, V., S. Sapon-Cousineau, and S. Assouline, *PI3K Inhibitors and Their Role as Novel Agents for Targeted Therapy in Lymphoma.* Curr Treat Options Oncol, 2020. **21**(6): p. 51.
17. Lacy, S.E., et al., *Targeted sequencing in DLBCL, molecular subtypes, and outcomes: a Haematological Malignancy Research Network report.* Blood, 2020. **135**(20): p. 1759-1771.
18. Mou, Y., et al., *Ferroptosis, a new form of cell death: opportunities and challenges in cancer.* J Hematol Oncol, 2019. **12**(1): p. 34.
19. Xu, Z.Z., et al., *Activation of the PI3K/AKT/mTOR pathway in diffuse large B cell lymphoma: clinical significance and inhibitory effect of rituximab.* Ann Hematol, 2013. **92**(10): p. 1351-8.
20. Yang, J., et al., *Sirt6 promotes tumorigenesis and drug resistance of diffuse large B-cell lymphoma by mediating PI3K/Akt signaling.* J Exp Clin Cancer Res, 2020. **39**(1): p. 142.

21. Curigliano, G. and R.R. Shah, *Safety and Tolerability of Phosphatidylinositol-3-Kinase (PI3K) Inhibitors in Oncology.* Drug Saf, 2019. **42**(2): p. 247-262.
22. Latour, S., et al., *Role of Calcium Signaling in GA101-Induced Cell Death in Malignant Human B Cells.* Cancers (Basel), 2019. **11**(3).
23. Maher, P., et al., *The role of Ca(2+) in cell death caused by oxidative glutamate toxicity and ferroptosis.* Cell Calcium, 2018. **70**: p. 47-55.
24. Ciavarella, S., et al., *Dissection of DLBCL microenvironment provides a gene expression-based predictor of survival applicable to formalin-fixed paraffin-embedded tissue.* Ann Oncol, 2019. **30**(12): p. 2015.
25. Kumar, V., et al., *The Nature of Myeloid-Derived Suppressor Cells in the Tumor Microenvironment.* Trends Immunol, 2016. **37**(3): p. 208-220.
26. Tzankov, A., et al., *Correlation of high numbers of intratumoral FOXP3+ regulatory T cells with improved survival in germinal center-like diffuse large B-cell lymphoma, follicular lymphoma and classical Hodgkin's lymphoma.* Haematologica, 2008. **93**(2): p. 193-200.
27. Fiorillo, M., et al., *Mitochondrial "power" drives tamoxifen resistance: NQO1 and GCLC are new therapeutic targets in breast cancer.* Oncotarget, 2017. **8**(12): p. 20309-20327.
28. Hiyama, N., et al., *Glutamate-cysteine ligase catalytic subunit is associated with cisplatin resistance in lung adenocarcinoma.* Jpn J Clin Oncol, 2018. **48**(4): p. 303-307.
29. Lin, L.C., et al., *gamma-Glutamylcysteine synthetase (gamma-GCS) as a target for overcoming chemo- and radio-resistance of human hepatocellular carcinoma cells.* Life Sci, 2018. **198**: p. 25-31.
30. Wang, B., et al., *Phospholipid Remodeling and Cholesterol Availability Regulate Intestinal Stemness and Tumorigenesis.* Cell Stem Cell, 2018. **22**(2): p. 206-220 e4.
31. Wang, B. and P. Tontonoz, *Phospholipid Remodeling in Physiology and Disease.* Annu Rev Physiol, 2019. **81**: p. 165-188.
32. Dodson, M., R. Castro-Portuguez, and D.D. Zhang, *NRF2 plays a critical role in mitigating lipid peroxidation and ferroptosis.* Redox Biol, 2019. **23**: p. 101107.
33. Lu, J.F., D. Pokharel, and M. Bebawy, *MRP1 and its role in anticancer drug resistance.* Drug Metab Rev, 2015. **47**(4): p. 406-19.
34. Kanai, Y. and M.A. Hediger, *The glutamate/neutral amino acid transporter family SLC1: molecular, physiological and pharmacological aspects.* Pflugers Arch, 2004. **447**(5): p. 469-79.
35. Wang, Q., et al., *Targeting glutamine transport to suppress melanoma cell growth.* Int J Cancer, 2014. **135**(5): p. 1060-71.
36. van Geldermalsen, M., et al., *ASCT2/SLC1A5 controls glutamine uptake and tumour growth in triple-negative basal-like breast cancer.* Oncogene, 2016. **35**(24): p. 3201-8.
37. Willems, L., et al., *Inhibiting glutamine uptake represents an attractive new strategy for treating acute myeloid leukemia.* Blood, 2013. **122**(20): p. 3521-32.

38. Abrego, J., et al., *GOT1-mediated anaplerotic glutamine metabolism regulates chronic acidosis stress in pancreatic cancer cells.* Cancer Lett, 2017. **400**: p. 37-46.
39. Riedell, P.A. and S.M. Smith, *Double hit and double expressors in lymphoma: Definition and treatment.* Cancer, 2018. **124**(24): p. 4622-4632.
40. Zenz, T., et al., *TP53 mutation and survival in aggressive B cell lymphoma.* Int J Cancer, 2017. **141**(7): p. 1381-1388.
41. Rutherford, S.C. and J.P. Leonard, *DLBCL Cell of Origin: What Role Should It Play in Care Today?* Oncology (Williston Park), 2018. **32**(9): p. 445-9.

Figures

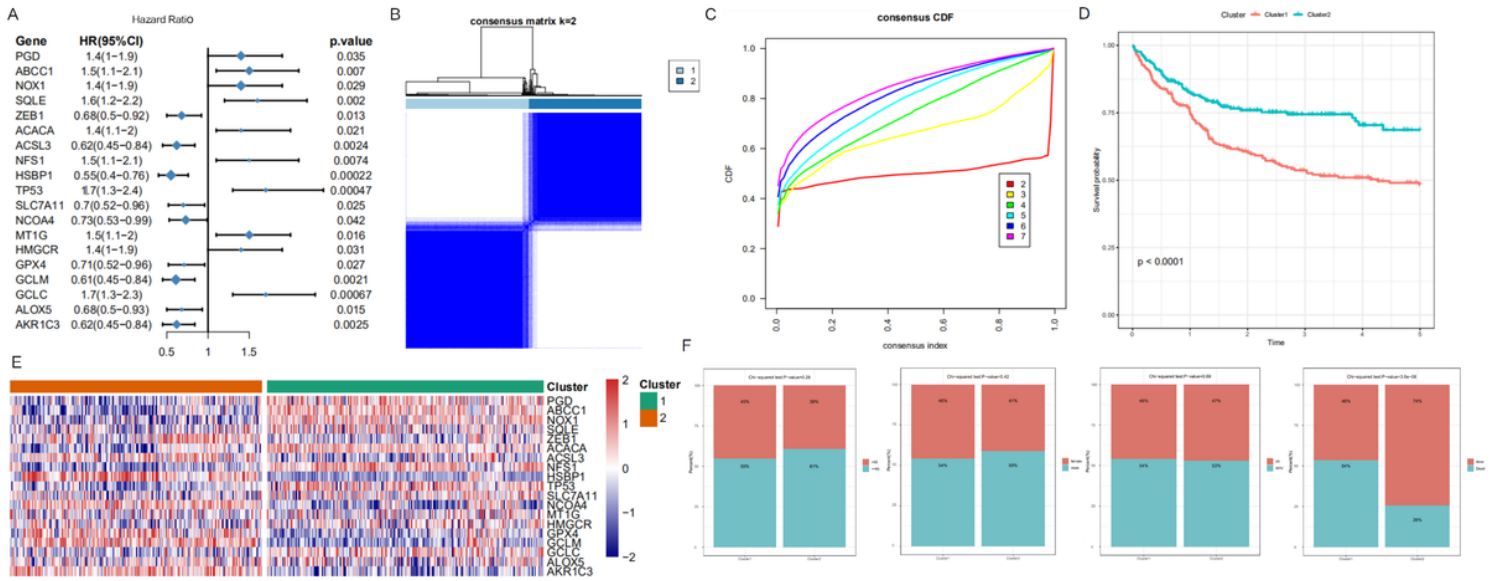

Figure 1

Molecular classification and verification based on FRGs. (A) Forest plots show the association between gene expression of FGRs and OS by univariate Cox regression analysis. (B) The consensus score matrix of all samples when k = 2. A higher consensus score between two samples indicates they are more likely to be grouped into the same cluster. (C) The cumulative distribution functions of the consensus matrix for each k (indicated by colors). (D) The OS of DLBCL patients in the two clusters demonstrated by Kaplan-Meier curves. (E) Heatmaps show the different expression levels of the 19 FRGs. (F) Composition ratio of the two subgroups in clinical characteristics such as stage, age, gender, and outcome.

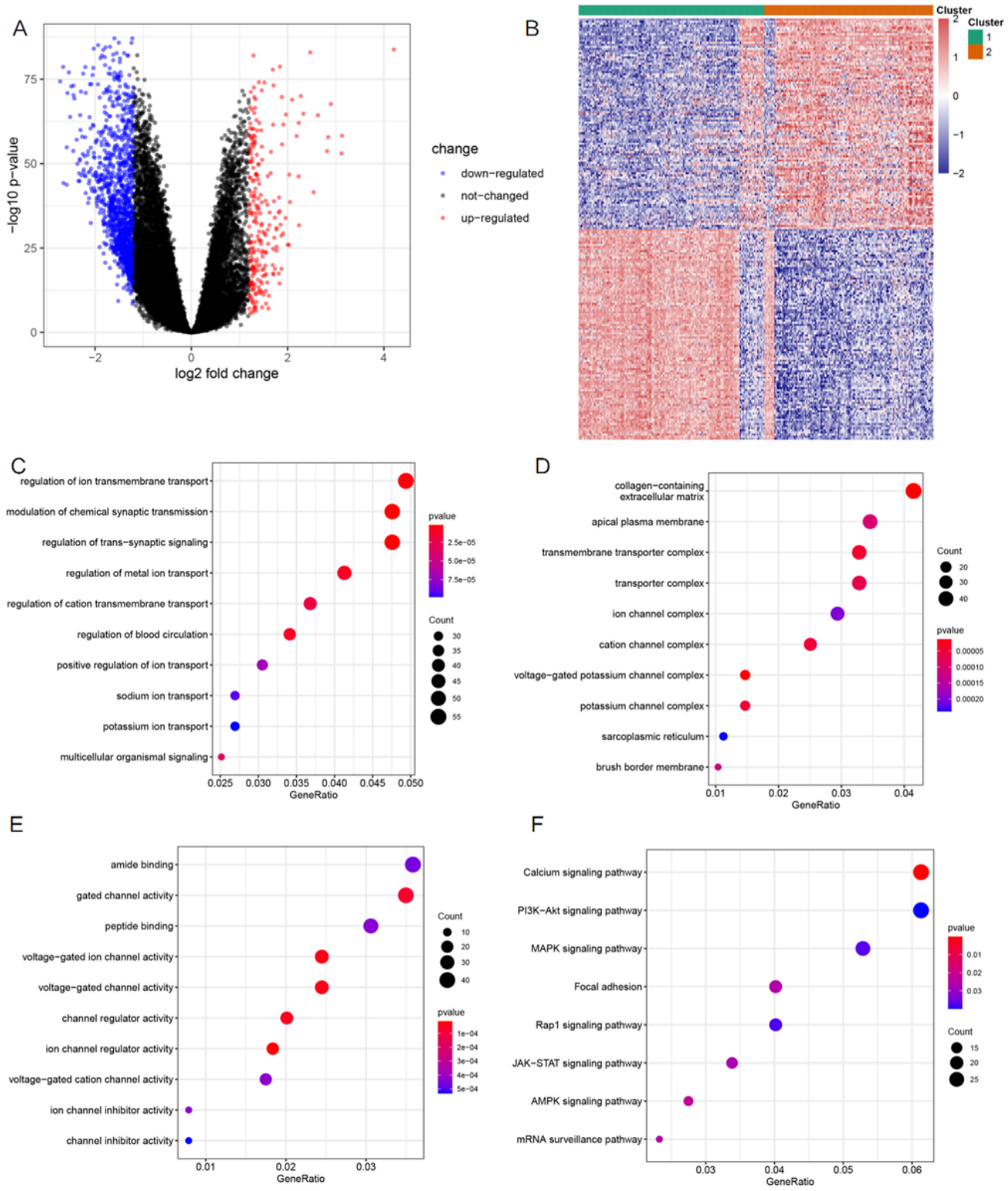

Figure 2

Identification and functional annotation analyses of the DEGs. (A) Volcano plot of the DEGs between C1 and C2 groups. Upregulated genes present as red nodes, downregulated genes present as blue nodes, and genes with no differences present as black nodes. (B) Heatmap of DEGs to visualize gene expression levels. (C-E) The top ten significant GO terms DEGs are displayed. (F)Top eight KEGG pathways of the DEGs are displayed.

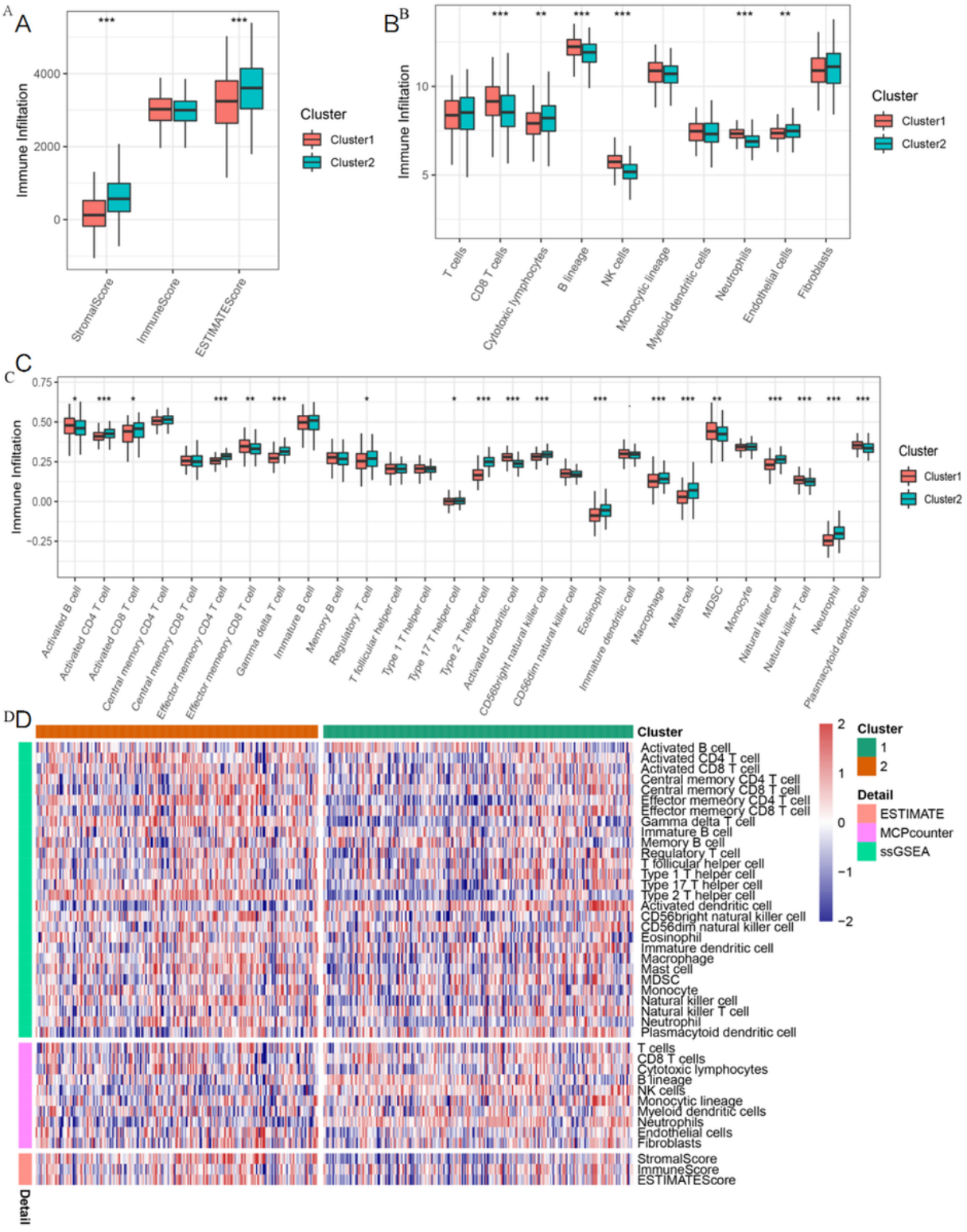

Figure 3

Relationship between TME and FRGs based subtypes. (A) Boxplot shows the Stromal Score, Immune Score, and ESTIMATE Score between the two clusters. (B) Boxplot shows the abundance of 10 immune cells in C1 and C2. (C) Boxplot shows the abundance of 28 immune cells between the two clusters. (D) Heatmap for immune responses based on ESTIMATE, MCPcounter, and ssGSEA algorithms in the two subtypes.

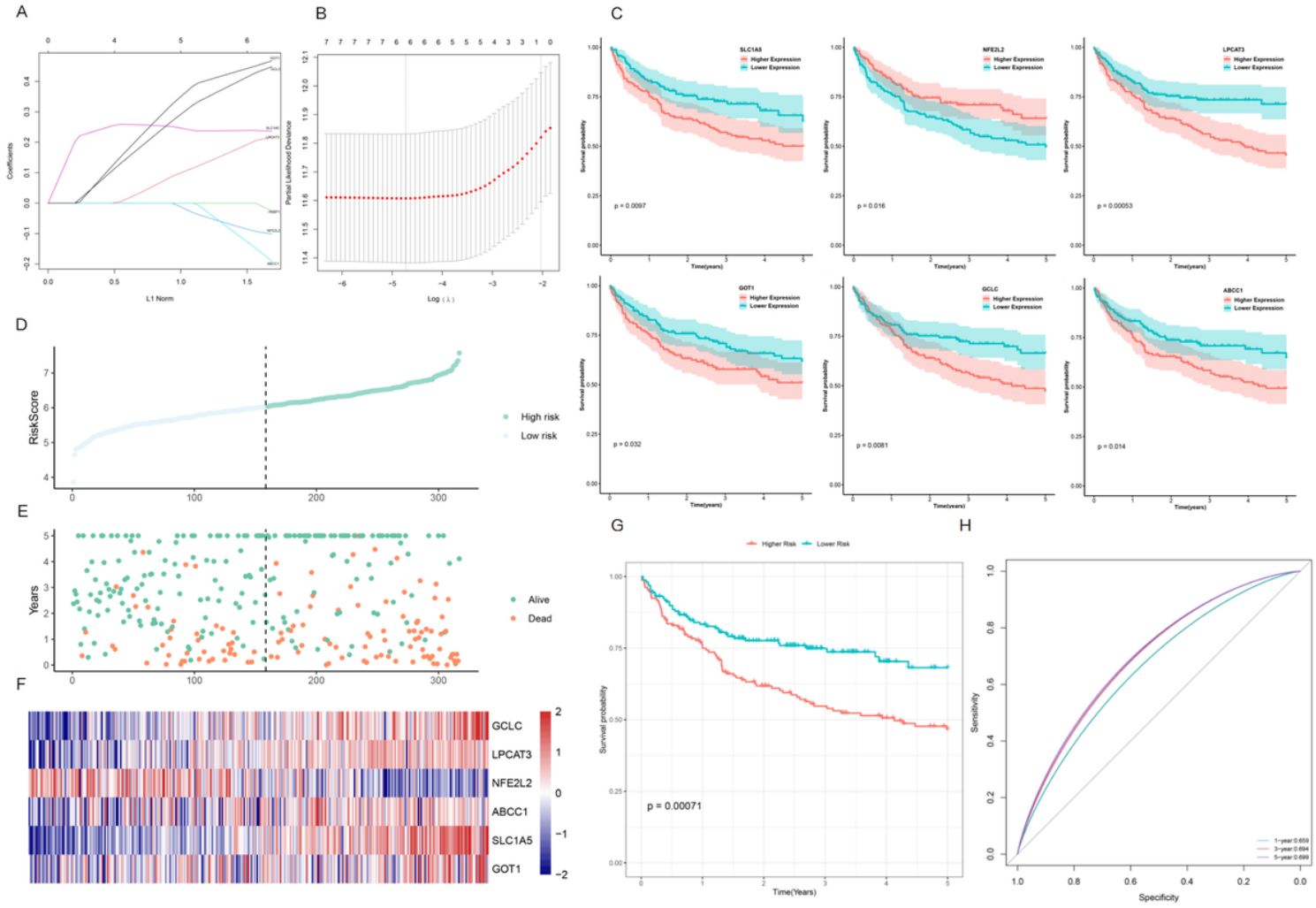

Figure 4

Identification of a six FRGs risk signature for OS of DLBCL in the training cohort. (A) LASSO coefficient spectrum of six genes in DLBCL patients. (B) Cross-validation for tuning parameter selection in the proportional hazards model. (C) Kaplan–Meier survival analysis of the DLBCL patients in low- and high-risk groups. (D) Risk score in low- and high-risk groups. (E)Survival time and survival state of the patients (F) Heatmap of the six genes in the two-risk group. (G) The KM survival curve of the two groups, the high-risk group showed worse OS status than that of low-risk group in the training cohort. (H) Time dependent ROC curve of the prognostic signature in the training cohort.

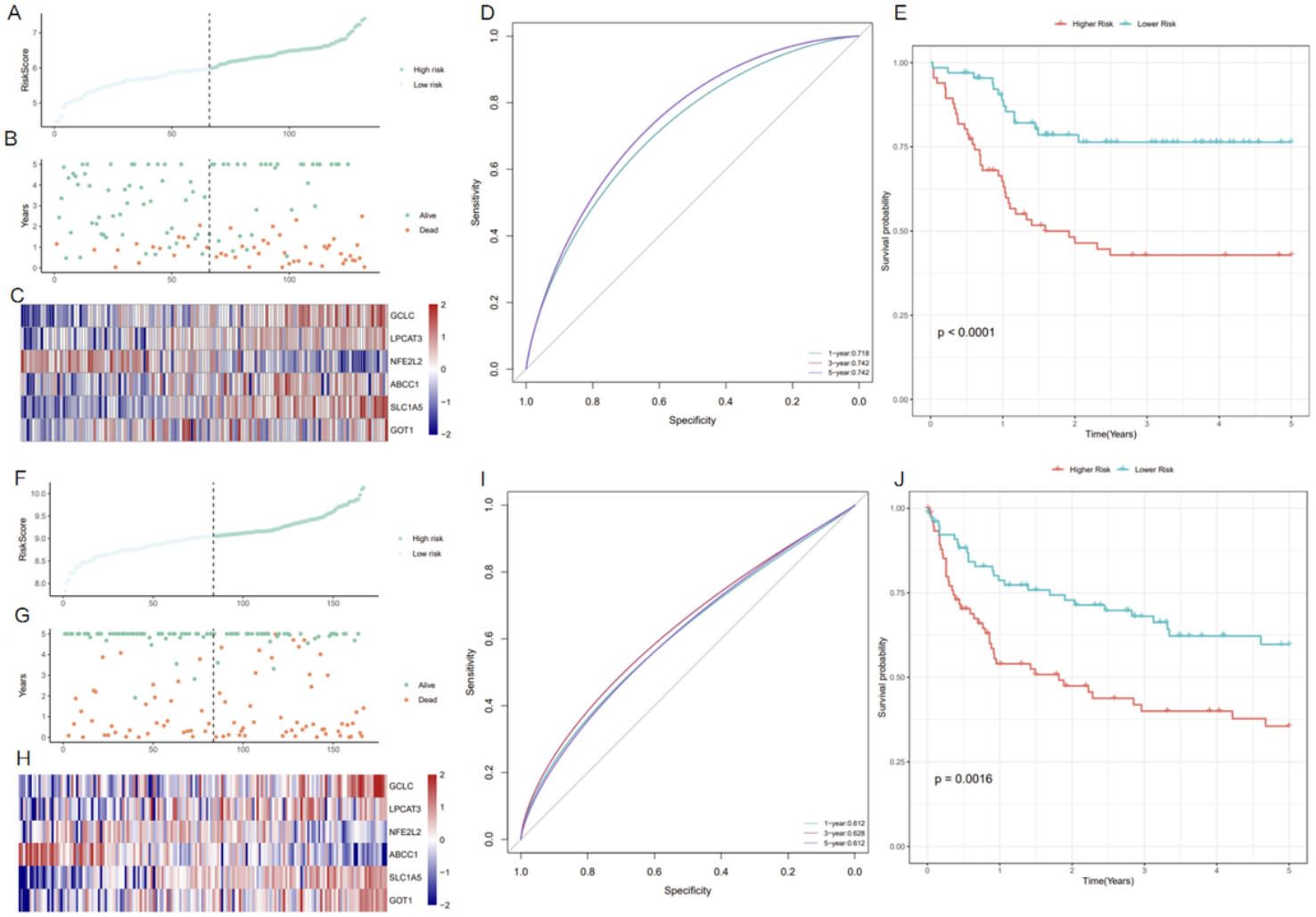

Figure 5

Internal and external validation of the six-gene signature. (A) Distribution of DLBCL patients and (B) OS status according to the median risk scores in the testing cohort. (C)Heatmap of the six genes in the two-risk group of testing group. (D) ROC curve and KM survival analysis(E) of the prognostic signature in testing cohort. (F)Distribution of DLBCL patients and (G) OS status according to the median risk scores in the independent cohort (GSE4475). (H)Heatmap of the six genes in the two-risk group of the independent cohort. (I) ROC curve and KM survival analysis(J) of the prognostic signature in the independent cohort (GSE4475).

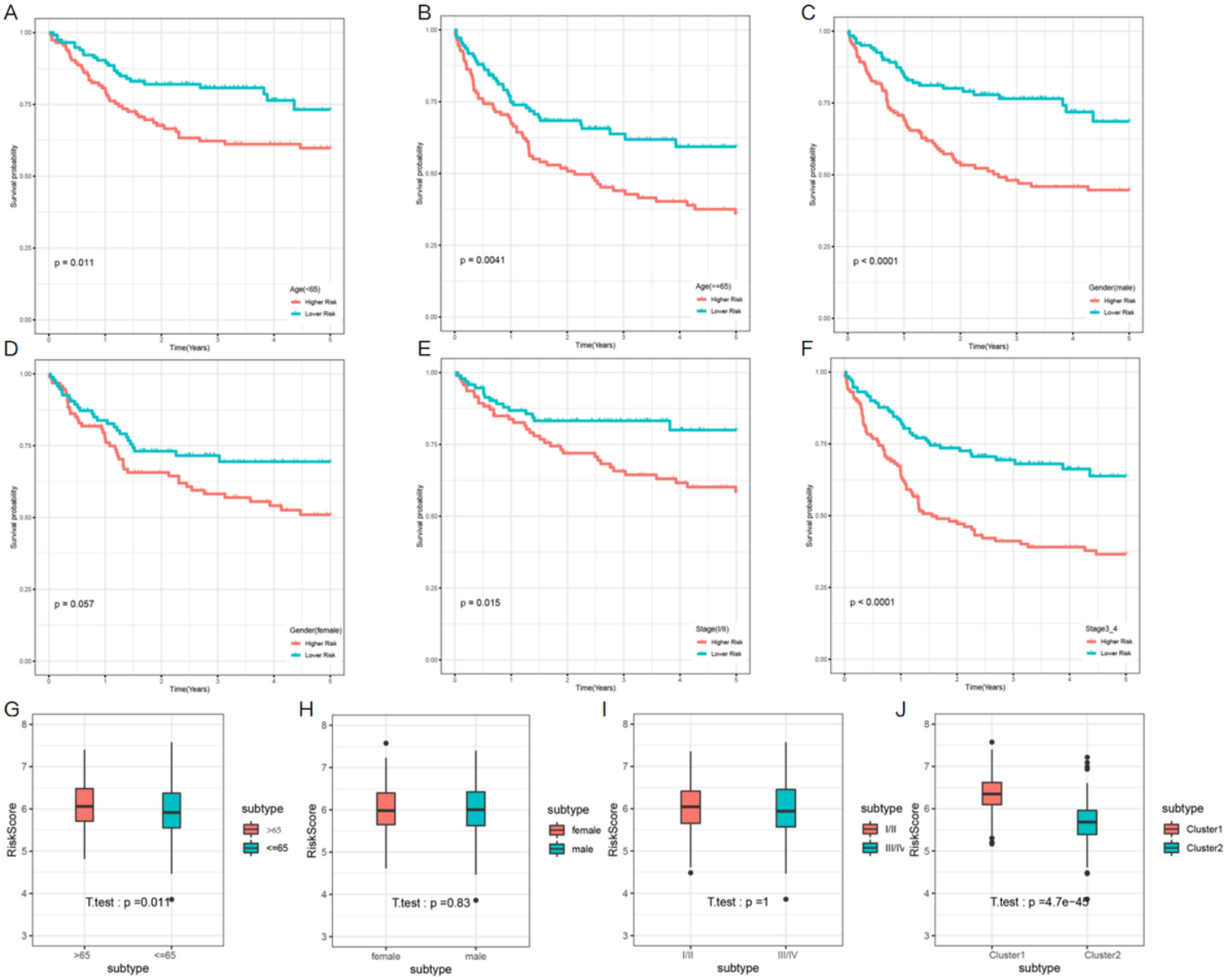

Figure 6

Connection between the FRGs signature and other clinical characteristics in the training cohort. The OS status for the two groups (high risk vs. low risk) assigned by age(A,B), gender(C,D) and clinical stage(E,F). The comparison of risk score in DLBCL patients stratified by different age, gender, stage and molecular subtypes(G-J).

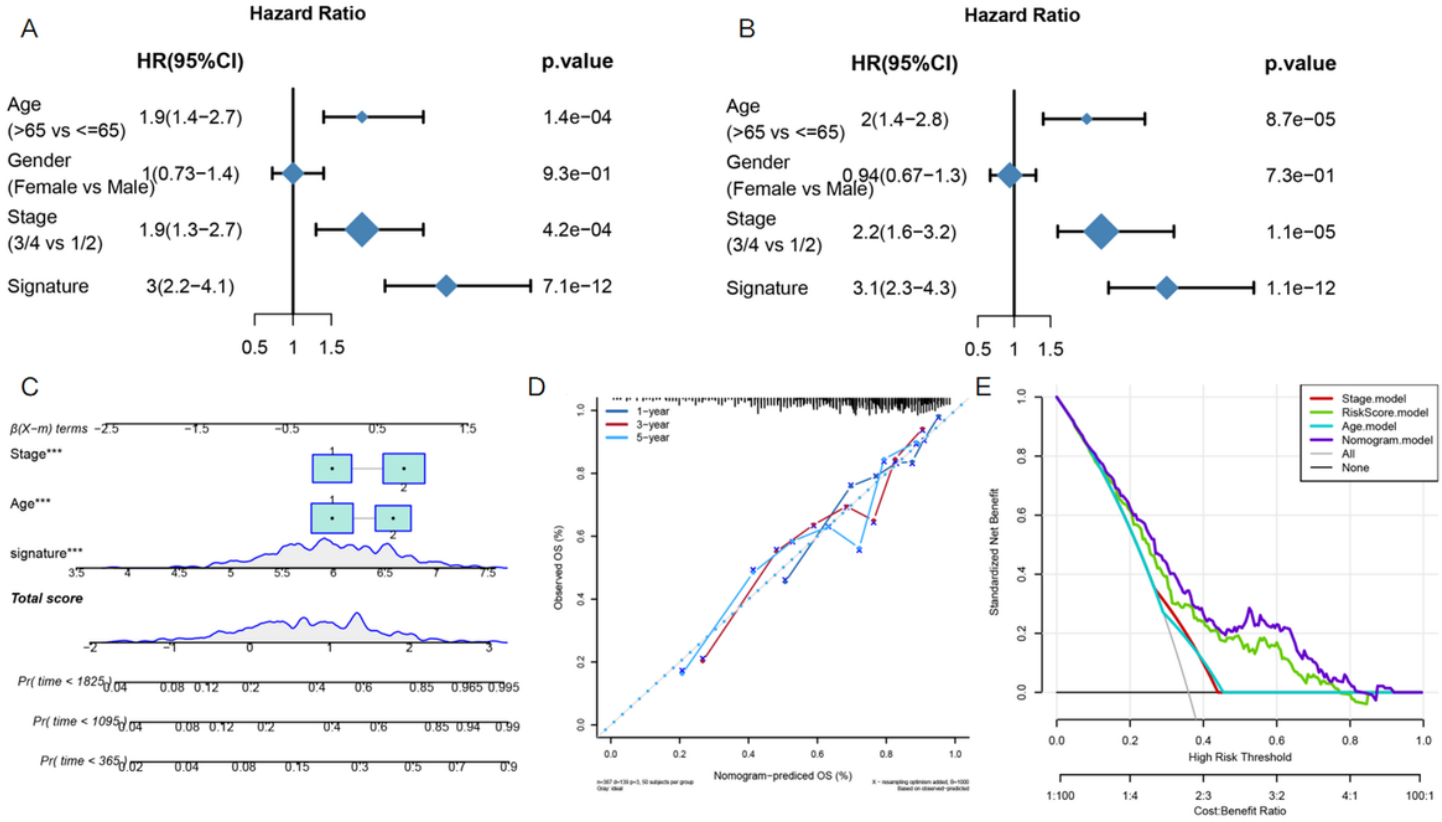

Figure 7

Construction of a nomogram of DLBCL patients. (A) Univariate Cox regression analysis presented as forest plot and showed the correlations between gender, age, clinical stage, risk factors and the OS survival of DLBCL patients. (B) Multivariate Cox regression analysis showed the correlation between risk score, clinical parameters and OS status. (C) The nomogram to predict the OS of patients. (D) The calibration curve revealed that the nomogram could predict OS with highly consistency with the actual OS. (E) DCA curves indicated that the nomogram had optimal clinical net benefits.

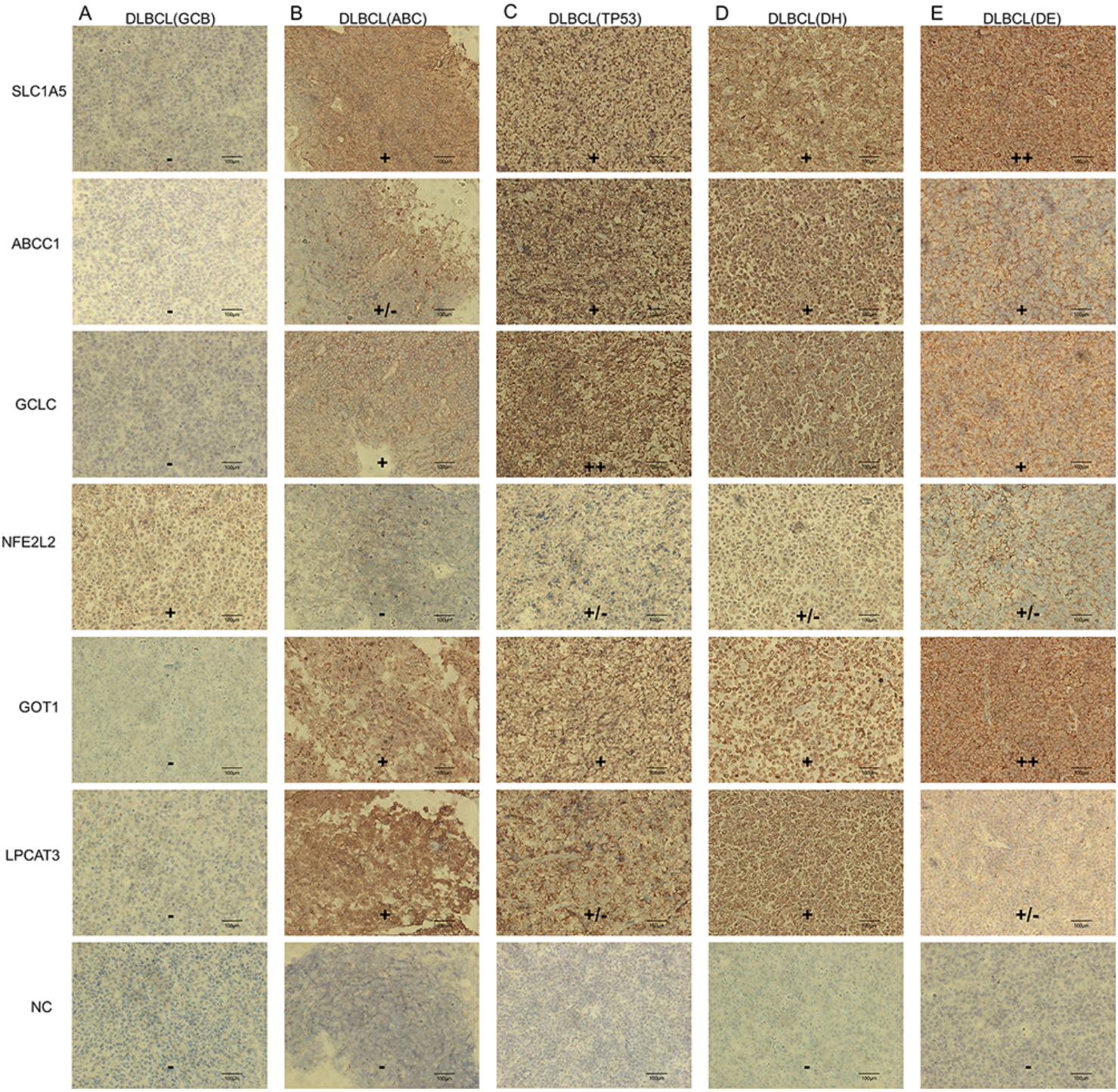

Figure 8

Differences in protein expression of the key genes in DLBCL tumor tissue by immunohistochemistry.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- S1.csv

- S2.csv
- S3.csv
- S4.csv
- S5.csv