# 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**
Second Xiangya Hospital
**Hongling Peng**
Second Xiangya Hospital
**Guangsen Zhang**
Second Xiangya Hospital
**Yunxiao Xu**
Second Xiangya Hospital
**Wenzhe Yan** ( yanwz1988@sina.com )
Second Xiangya Hospital https://orcid.org/0000-0002-1208-8711

**Research**

**Keywords:** DLBCL, prognosis, ferroptosis, signature

**Posted Date:** June 28th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-603170/v1

# Abstract

Backgroud: Diffuse large B-cell lymphomas (DLBCLs) are featured as phenotypically and genetically heterogeneous. Ferroptosis is a newly discovered regulated cell death pathway that plays a crucial role in the occurrence and progression of tumors. We aim to identify a ferroptosis-related gene (FRG) prognostic signature for DLBCLs by systematic analysis of transcriptional profiles.

Methods: This study retrospectively analysed the transcriptome profiles and clinical parameters of 604 DLBCL patients from 3 public datasets. A series of bioinformatic approaches including univariate and multivariate Cox regression analysis, function analysis, immune infiltration analysis, differential expression 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 software were combined to explored the heterogenicity of FRG based clusters and to built prognostic model. Immunohistochemistry was used to exam the protein expression of six FRGs in different type of DLBCL.

Results: We first identified 19 FRGs with potential prognostic values and classfied the patients into two subgroups (named cluster 1 and cluster 2), Results showed that there were different patterns of immune cell infiltration among patients in the two clusters. 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 overall survival (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 addtion, 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 predictive performance. Finally, the validation by immunohistochemistry indicated the *GCLC*, *LPCAT3*, *NFE2L2*, *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

DLBCL is a heterogeneous desease 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 prognosis in DLBCL, but in the era of rituximab based immunochemotherapy, the IPI explains less than 25% of variation in OS and poorly separates the highest-risk subgroups[1] (24264230). When it comes to molecular factors, gene expression profiling has identified three subtypes of DLBCLs (activated B-cell–like [ABC], germinal-center B-cell–like [GCB], and unclassified) according to cell of origin that are associated with a differential response to chemotherapy and targeted agents [2]. Patients with ABC cases are featured by a more

aggressive profile and active NF-κB and BCR signaling pathways [3, 4], while GCB tumors are related with alterations that drive aberrant chromatin-modification, PI3K signaling, and the overexpression of MYC and BCL2 through translocations or copy number gains [5, 6]. Currently, researcher also identified a new four prominent genetic subtypes in DLBCL: (1) MCD, based on the co-occurrence of MYD88L265P and CD79B mutations, (2) BN2, based on BCL6 fusions and NOTCH2 mutations, (3) N1, based on NOTCH1 mutations, and (4) EZB, based on EZH2 mutations and BCL2 translocations [7]. These subtypes provide a potential nosology 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 newly discovered form of iron-dependent oxidative cell death characterized by lethal accumulation of lipid-based reactive oxygen species (ROS) [8, 9]. It is different from other forms of cell death including apoptosis, necrosis, and autophagy in terms of morphology, biochemistry, and genetics. An increasing amount of evidence proves that the imbalanced iron metabolism is associated with the occurrence and development of cancer[10, 11]. Apart from ferroptosis-inducing agents, numerous genes have also been identified as modulators or markers of ferroptosis. An investigation of 114 cancer cell lines showed that DLBCLs were highly vulnerable to ferroptosis inducers-erastin[12]. Furthermore, ferroptosis may be induced by dimethyl fumarate, which is a promising novel therapeutic option in the treatment of GCB DLBCLs[13]. Hence, we aimed to analyze the expression patterns of ferroptosis-related genes in the DLBCL patients, as well as their prognostic values.

In the present study, we used mRNA expression profiles and corresponding clinical data of DLBCL patients from public databases to analyze the differential expression of ferroptosis-related genes to identify the enriched pathways and their biological functions to explore the underlying mechanisms in DLBCL samples. Furthermore, we also constructed and validated a prognostic multigene signature model with six ferroptosis-related differentially expressed genes (DEGs) in the training and testing cohort.

# Methods

# Data sources

All datasets used in this study are available to the public. The gene expression profile and corresponding clinical information of the three independent cohorts (GSE10846,GSE11318GSE4475) were retrieved from the Gene Expression Omnibus (GEO) Database (https://www.ncbi.nlm.nih.gov/geo/) after systematical screening. The clinical characteristics of the patients are shown in Table 1. All patients with unclear survival time, survival status, and clinicopathological characteristics were excluded. Thereafter, GSE10846 and GSE11318 RNA-seq data were merged into one new 447 sample cohort. 115 samples from GSE4475 were used as an external validation cohort. All data were collected on 14 April 2021.

Table 1
The clinical characteristics of 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

The two cohorts were screened for the expression of sixty FRGs that were collected from the published literature (Supplement Table 1). FRGs and survival data were analyzed by univariate Cox regression analysis using the R package. The filtering threshold was set at $p < 0.05$, and determined that 19 genes were 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 use 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 divided into the training set (70 % (317)) and the validation set (30 % (132)). Using the R software package "survival", we first performed univariate Cox regression analysis and then the LASSO regression analysis to further remove overfitting. Finally, the genes that could be used as independent prognostic factors of OS were screened by multivariate Cox regression analysis, and their regression coefficients were calculated. The risk score of each sample was calculated, and the formula was as follows: Risk score = $\text{coeffificient}_{\text{gene1}} \times \text{expression}_{\text{gene 1}} + \text{coeffificient}_{\text{gene 2}} \times \text{expression}_{\text{gene 2}} + \ldots \ldots \text{coeffificient}_{\text{gene n}} \times \text{expression}_{\text{gene n}}$.

Subsequently, all patients were divided into high-risk and low-risk subgroups based on the median risk score. The Kaplan–Meier method was performed to compare the survival difference between the two risk subgroups. The prediction accuracy and effectiveness in one year, three years, and five years of the multi-gene 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 divided into high-risk and low-risk groups.

In addition, to validate the robustness of the prognostic signatures, we calculated the risk score of patients in the validation cohorts. The Kaplan–Meier survival curve and survival ROC curve were developed to show the predictive ability of the prognostic signatures in the validation cohorts. The external dataset GSE4475 was also used to verify the risk score by the same signature and coefficients.

# Univariate and multivariate regression analyses of six gene signature

First, we performed univariate Cox regression analysis to evaluate the prognostic value of polygenic signatures and clinicopathological features. Multivariate Cox regression analysis was used to further determine the independent prognostic factors. Next, a nomogram was constructed using the survival rate and “RMS” R package for predicting overall survival, and a correction curve was drawn to evaluate the consistency between the actual and predicted survival rates. Calibration curves of nomograms were calculated to estimate the accuracy and consistency of the prognostic models.

# Immunohistochemistry

15 tumor tissues were obtained from The Second Xiangya Hospital. Tissue sections were kept in an incubator at 60 C for 1 h and then in xylene and rehydration in a series of descending ethanol concentrations for deparaffinization, The slides were boiled in sodium citrate for 20 min in a microwave oven and then treated with 3% hydrogen peroxide solution for 10 min and blocked in 5% goat serum for 20 min. 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 1h. The secondary antibody was used to amplify and visualize the signal for each protein. Sections were stained with 3,30 -diaminobenzidine and with hematoxylin, back to blue by soaking in warm water, dehydrated, clarified with dimethylbenzene, and mounted.

# Statistical analysis

SPSS, GraphPad Prism 7, or the R software were used for the statistical analyses. Univariate and multivariate Cox regression analyses, differential expression analysis, ROC curve analysis, and Kaplan–Meier survival curve were performed by the corresponding R packages in R software to determine the difference in the OS between the stratified groups. Wilcoxon test was used for comparisons between the two groups, and Kruskal–Wallis test was used for comparisons of prognosis between groups. For each analysis, all $p$ values were two-sided, with $p < 0.05$ 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 divide all samples into two groups with the least amount of correlation between groups (Fig. 1B, C). To explore whether there was a correlation between the clustering result and clinical outcome, we compared the OS among the two clusters of patients and showed that patients in cluster 1(C1) had shorter 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). While there was also no difference of age, gender, and stage between the two subtypes, the patients in the C1 subtype were significantly associated with a higher rate of death(p = 3.6e-08)(Fig. 1F).

# Identification and functional enrichment analysis of DEGs

We detected the expression of FRGs by differential analysis and uncovered that 248 genes were up-regulated, and 1,242 genes were down-regulated in C2 than C1($p < 0.05$, |log2FC| > 1.2) (Supplementary Table 2). The distribution of the representative 100 DEGs is displayed in Fig. 2A, 2B. These 1,490 DEGs were further analyzed by GO and KEGG pathway to explore 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 most abundant cellular component (CC) terminology were the collagen containing extracellular matrix, apical plasma membrane, transmembrane transport complex, and ion channel complex. The molecule function (MF) mainly regulated amide binding, gated channel activity, and peptide binding (Supplementary Table 3). As expected, DEGs were enriched in several iron-related molecular functions, such as ion channel and ion gated channel activities. As KEGG pathway analysis revealed, the DEGs were involved in 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 [14](Fig. 2F, Supplementary Table 4).

# FRG based cluster was significantly associated with immune function

Comparison of immunity scores between the two clusters showed significant differences in the scores for tumor microenvironment (TME), including Stromal Score ($p < 0.001$) and ESTIMATE Score ($p < 0.001$) (Fig. 2A). The higher Stromal Scores reflected the higher frequency of stromal components in TME. ESTIMATE Score was the sum of Immune Score and Stromal Score denoting the comprehensive proportion of both components in TME. MCPcounter analysis showed that immune scores of CD8 T cells, B lineage, NK cells, and neutrophils were significantly higher in C1 than C2 (Fig. 3B). Implying that ferroptosis may have a profound impact on TME in DLBCL. Ssgsea analysis showed that activated B cell, effector memory CD8 T cell, activated dendritic cell, MDSC, NK T cell, and plasmacytoid dendritic cell scores were all higher in C1 than C2, while activated CD4 T cell, activated CD8 T cell, effector memory CD4 T cell, gamma delta T cell, regulatory T cell, type 17 T helper cell, type 2 T helper cell, CD56 bright NK cell, eosinophil, macrophage, mast cell, NK cell, and neutrophil scores were higher in C2 than C1 (Fig. 3C).

# Construction of FRGs signature for OS in the training cohort

Seven FEGs were identified as associated with OS (Fig. 4A, Supplementary Table 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 lamda (0.008845861) (Fig. 4B). The prognostic Kaplan–Meier survival analysis of these six genes showed that the high expression of *GCLC*, *LPCAT3*, *SLC1A5*, and *GOT1* were associated with poor prognosis, while *ABCC1* and *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 patients were stratified into a high-risk group (n = 158) or a low-risk group (n = 159) according to the median cut-off value ($p = 0.00071$). As shown in Figs. 4D-4F, high-risk patients had a higher probability of earlier death. Consistently, the Kaplan-Meier curve showed that the high-risk group demonstrated a poor OS than their low-risk counterparts (Fig. 4G). The predictive performance of the risk score for OS was evaluated by time-dependent ROC curves, and 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 six-gene signature for OS in the verification cohort

To test the robustness of the model constructed from the training cohort, the patients from the verification cohort were also categorized into high- or low-risk groups by the median value calculated with the same formula as that from the training cohort. Similar to the results obtained from the training cohort. High-risk scores correlated with worse OS than low-risk scores($p < 0.0001$), and the AUC of the six-gene signature was 0.718 at one year, 0.742 at three years, and 0.742 at five years (Fig. 5A-5E). Furthermore, the patients from the independent cohort (GSE4475) were also categorized into high- or low-risk groups using the same formula, and the results were consistent with those obtained from both the testing cohort and the training cohort. The high-risk group showed significantly reduced OS compared with patients in the low-risk group($p = 0.0016$), and the AUC of the six-gene signature was 0.612 at one year, 0.628 at three years, and 0.612 at five years (Fig. 5F-5J).

# Analysis of the risk model and clinical characteristics

The risk score analysis in the training cohort found that the six-gene signature could also categorize the diverse age, gender, and stage groups into high- or low-risk. The patients were stratified by gender (> 65y subtype, < 65y subtype), age (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 lower 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 compared between different age/gender/stage/molecular subtypes. The results showed the risk score was significant higher in C1.(Fig. 6G-6J)

# Construction of a nomogram of DLBCL patients based on OS

As the results showed, in the Univariate Cox regression analysis showed that age, stage, and risk score of the prognostic signature were significantly associated with the OS 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 identified that age, stage, and risk score of the prognostic signature (HR = 3.1,p < 1e-5) were independent predictive factors for 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 together to build a nomogram model. The results showed that hybrid nomogram, incorporating clinicopathological characteristics and the novel FRGs prognostic signature, was stable and accurate. All independent factors were combined to establish a nomogram for predicting the one-, three-, and five-year OS (Fig. 7D). As shown in Fig. 7D, the risk score contributes more to the total score than other variables. The one-, three-, and five-year OS scores of patients declined as the total score increased. The calibration plots approached 45° and showed great consistency between the predicted OS rates and actual observations at one, three and five years (Fig. 7C), indicating that the nomogram performed well. The risk score and the nomogram both had good reliability based on the DCA. Compared to a single independent predictive factor, the 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 explore the clinical significance of the signature, we used immunohistochemistry to validate the expression of the six key genes in DLBCL samples with different prognostic molecular feature, which include DLBCL with ABC type, DLBCL with ABC type, DLBCL with GCB type, DLBCL with double hit(DH), DLBCL with double expressors(DE) and DLBCL with TP53 mutant. The result showed that the protein expression levels of *GCLC*, *LPCAT3*, *NFE2L2*,, *SLC1A5*, and *GOT1* are expressed relatively higher in DLBCL-ABC, DLBCL-DH, DLBCL-DE, DLBCL-TP53 than in DLBCL-GBC samples and the *ABCC1* had the reverse tendency.

# 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[15]. 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 recently recognized programmed cell death modality. It can overcome resistance of malignant cells to chemotherapy and as well facilitate removal of defective cells [16]. Consequently, it is potentially a novel approach for tumor treatment. Current studies mainly focus on the role of iron metabolism in DLBCL development and treatment [12, 13], as far as we know, this is the first attempt to identify the role of ferroptosis in DLBCL prognosis.

In this study, based the expression pattern of FRGs, two DLBCL subgroups were identified by consensus clustering analysis. The results revealed significant 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 analyses revealed the genes mainly 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 tumours and haematological 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[17–19]. Calcium signaling also plays a key role in GA101-induced cell death in lymphoma cells [20]. Recent evidence hints that Ca2 + ions play a fundamental role in cell death mediated by oxidative glutamate toxicity or oxytosis, a form of programmed cell death similar and possibly identical to ferroptosis[21]. The GO enrichment mainly included several iron-related biological processes or molecular functions, such as ion channel activity and ion gated 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[22]. 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[23]. Regulatory T cells and CD4 + T cells, which had a correlation with improved survival in DLBCL has a lower ratio in the C1 subtype[22, 24]. These results confirm that ferroptosis has a regulatory effect on the TME.

Previous studies have confirmed that ferroptosis-inducer erastin is effective for inducing 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 unknown. In this study, we identified, for the first time, the novel risk scoring model constructed by six genes (*GCLC*, *LPCAT3*, *NFE2L2*, *ABCC1*, *SLC1A5*, and *GOT1*) to classify DLBCL patients into two classes and independently predicted the OS of patients with superior prediction performance. Gene markers related to ferroptosis were established, and the expression characteristics of the six genes are not affected by the differences in the underlying diseases of DLBCL, suggesting that the constructed prognosis model can be applied to various types of DLBCL patients. Moreover, the corresponding nomogram based on the six-gene model also helps clinicians make better clinical decisions and develop treatment strategies. By focusing on the specific function of the six ferroptotic genes, previous studies have demonstrated that most of these genes play a pivotal role in cancer cells.

GCLC (glutamate-cysteine ligase catalytic subunit) is a rate-limited enzyme that primarily regulates *de novo* synthesis of glutathione. It has been revealed that GCLC activation is associated with anti-tumor drug resistance in breast, lung, liver, head, and neck cancers[25–27]. LPCAT3 is an enzyme that converts lysophosphatidylcholine to phosphatidylcholine in the liver in order to maintain systemic homeostasis. It also participates in the phospholipid remodeling and intestinal stem cell growth and tumorigenesis [28, 29]. NFE2L2 is a master regulator of the antioxidant response and regulates the activity of several

ferroptosis and lipid peroxidation-related proteins [30]. ABCC1 plays an active role in protecting cells by its ability to efflux a vast array of drugs to sub-lethal levels. There has been much effort in elucidating the mechanisms of action, structure, substrates, and substrate binding sites of ABCC1 in the last decade [31]. SLC1A5 is a cell surface solute-carrying transporter that mediates uptake of neutral amino acids, including glutamine[32]. Blocking SLC1A5 to prevent glutamine uptake successfully prevents tumor cell proliferation in melanoma[33], breast cancer[34], and acute myeloid leukemia[35]. GOT1 plays an important role in energy metabolism and ROS balance in chronic acidosis stress[36]. Few studies on the role these genes play in DLBCL patients' prognosis have been reported and their influence on the process of ferroptosis remains to be elucidated. We further evaluated the protein expression 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, and GCB DLBCL have more favorable outcomes than those with ABC DLBCL when treated with standard immunochemotherapy. 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*, *NFE2L2*, *SLC1A5* and *GOT1* than the GCB DLBCL, while *ABCC1* had the reverse tendency. These results confirmed that *GCLC*, *LPCAT3*, *NFE2L2*, *SLC1A5* and *GOT1* are inclinded to be the adverse prognostic biomarkers, while the *ABCC1* is apt to be a protective factor in DLBCL.

# Conclusion

In summary, our research used a large database to establish a six-gene signature related to FRGs for predicting the prognosis of DLBCL patients, which undoubtedly provides insights into the identification of therapeutic targets for DLBCL. The ferroptotic signature identified in our study is a potential biomarker of outcome in DLBCL patients. Furthermore, the immunohistochemistry identified the expression level of the six genes in DLBCL with different molecular type. But we still need further intensive experimental verification and prospective cohort studies in the future 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

Not applicable

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 were supported by the National Natural Science Foundation of China (81900170) and Nature 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 analysed 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

# 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. 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.
6. Aukema, S.M., et al., *Double-hit B-cell lymphomas.* Blood, 2011. **117**(8): p. 2319-31.
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. Stockwell, B.R., et al., *Ferroptosis: A Regulated Cell Death Nexus Linking Metabolism, Redox Biology, and Disease.* Cell, 2017. **171**(2): p. 273-285.
9. Dixon, S.J., et al., *Ferroptosis: an iron-dependent form of nonapoptotic cell death.* Cell, 2012. **149**(5): p. 1060-72.
10. Tang, R., et al., *Ferroptosis, necroptosis, and pyroptosis in anticancer immunity.* J Hematol Oncol, 2020. **13**(1): p. 110.
11. Shen, Z., et al., *Emerging Strategies of Cancer Therapy Based on Ferroptosis.* Adv Mater, 2018. **30**(12): p. e1704007.
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.
14. 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.
15. 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.
16. Mou, Y., et al., *Ferroptosis, a new form of cell death: opportunities and challenges in cancer.* J Hematol Oncol, 2019. **12**(1): p. 34.
17. 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.
18. 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.
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. Latour, S., et al., *Role of Calcium Signaling in GA101-Induced Cell Death in Malignant Human B Cells.* Cancers (Basel), 2019. **11**(3).
21. 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.
22. 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.
23. Kumar, V., et al., *The Nature of Myeloid-Derived Suppressor Cells in the Tumor Microenvironment.* Trends Immunol, 2016. **37**(3): p. 208-220.
24. 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.
25. 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.
26. 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.
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. Wang, B. and P. Tontonoz, *Phospholipid Remodeling in Physiology and Disease.* Annu Rev Physiol, 2019. **81**: p. 165-188.
29. Wang, B., et al., *Phospholipid Remodeling and Cholesterol Availability Regulate Intestinal Stemness and Tumorigenesis.* Cell Stem Cell, 2018. **22**(2): p. 206-220 e4.

30. 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.
31. 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.
32. 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.
33. Wang, Q., et al., *Targeting glutamine transport to suppress melanoma cell growth.* Int J Cancer, 2014. **135**(5): p. 1060-71.
34. 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.
35. Willems, L., et al., *Inhibiting glutamine uptake represents an attractive new strategy for treating acute myeloid leukemia.* Blood, 2013. **122**(20): p. 3521-32.
36. Abrego, J., et al., *GOT1-mediated anaplerotic glutamine metabolism regulates chronic acidosis stress in pancreatic cancer cells.* Cancer Lett, 2017. **400**: p. 37-46.

# Figures

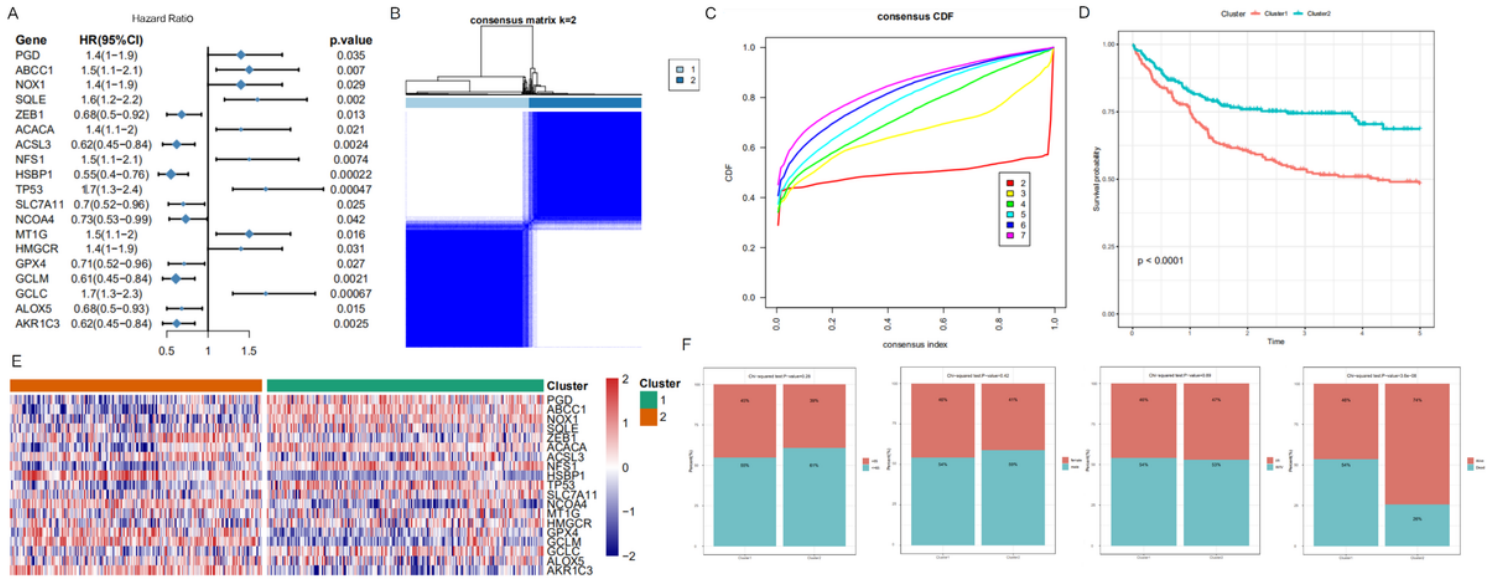
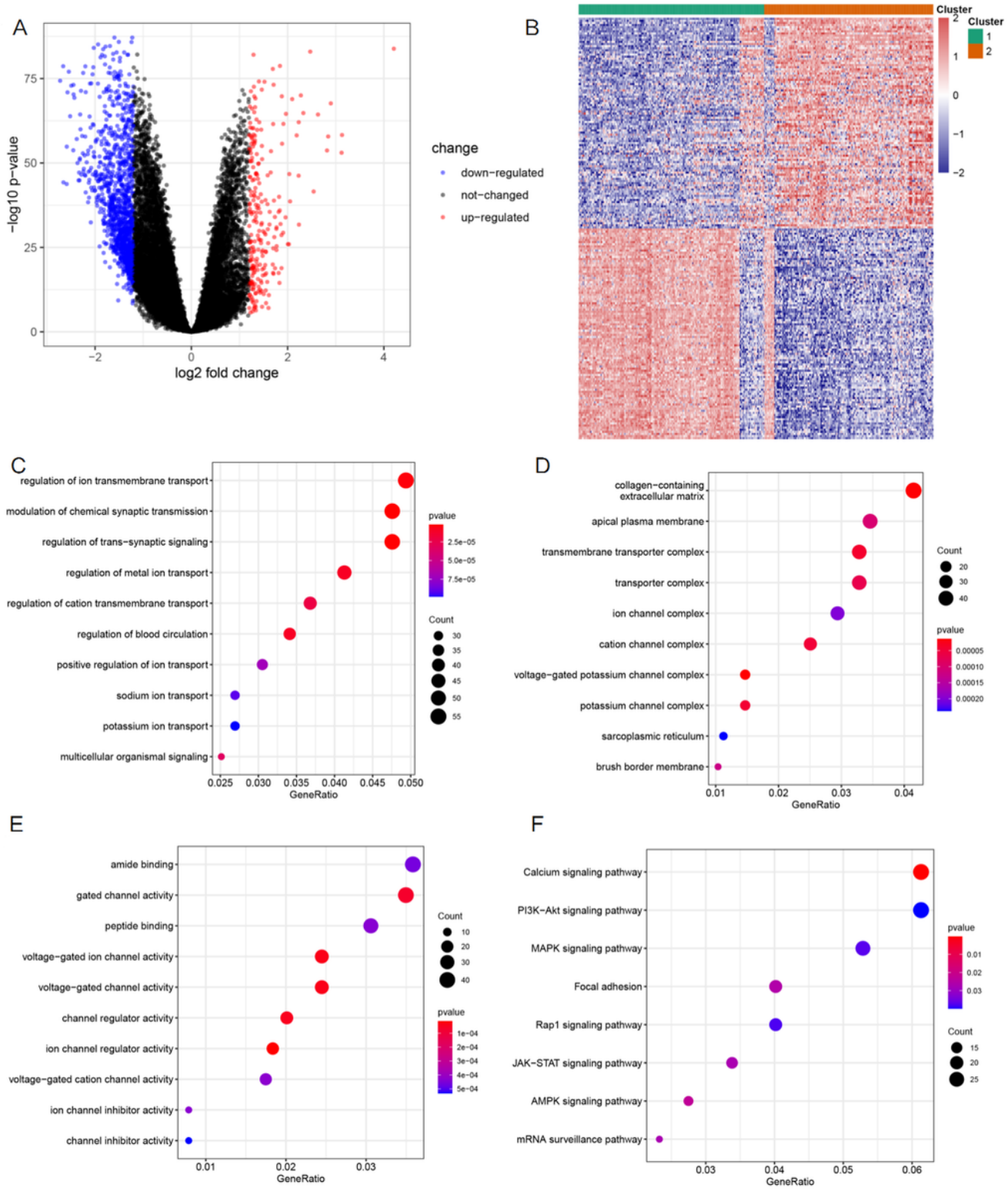


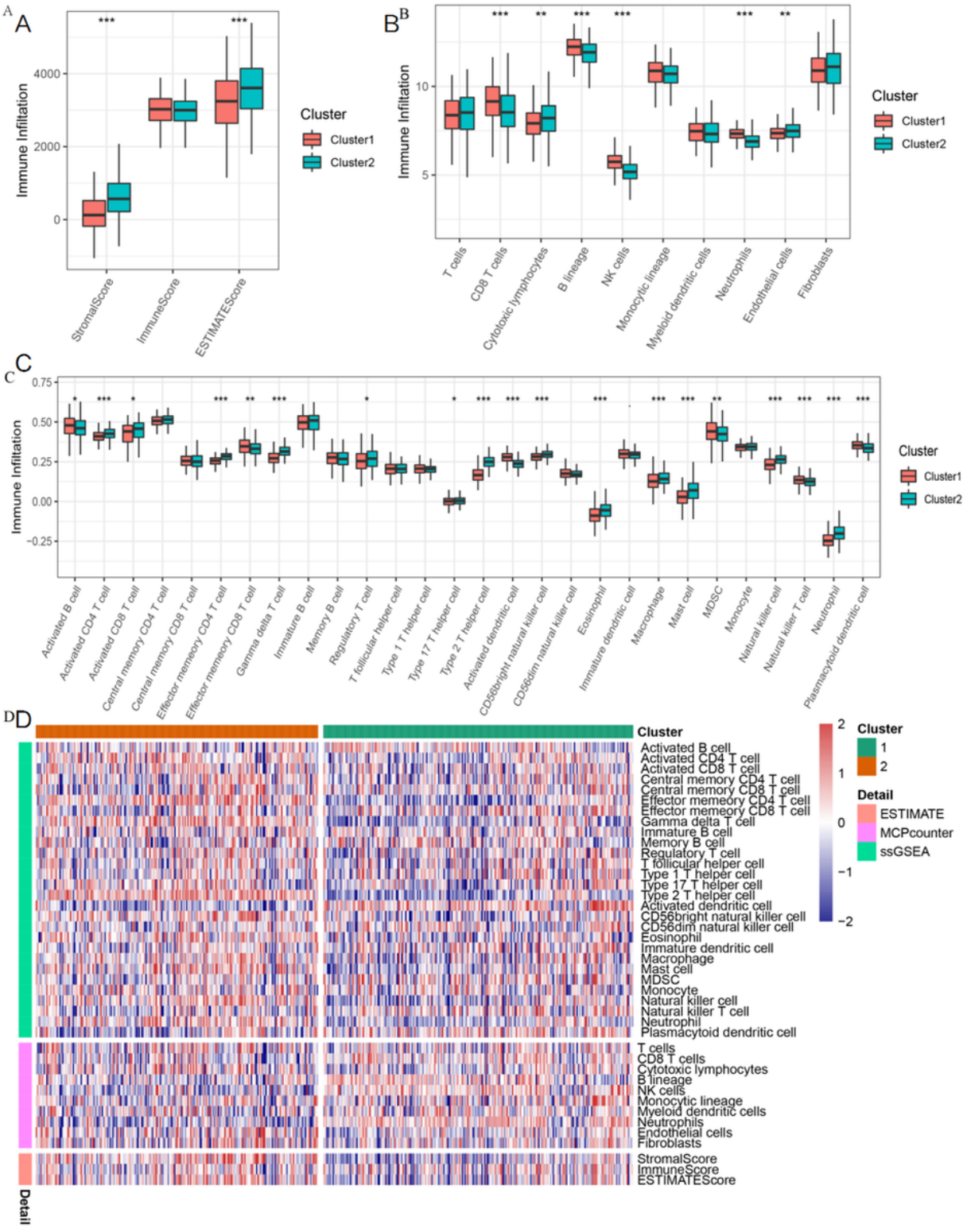

**Figure 1**

Molecular classification and verification based on FRGs. (A) Forest plots showing the results of the univariate Cox regression analysis between gene expression and OS. (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 in different iterations. (C) The cumulative distribution functions of the consensus matrix for each k (indicated by colors). (D) Kaplan-Meier curves for the OS of patients in the two clusters. (E) Heatmaps show the different expression levels of the 19 FRGs. (F) Composition percentage of the two subtypes in clinical characteristics such as stage, age, gender, and outcome.



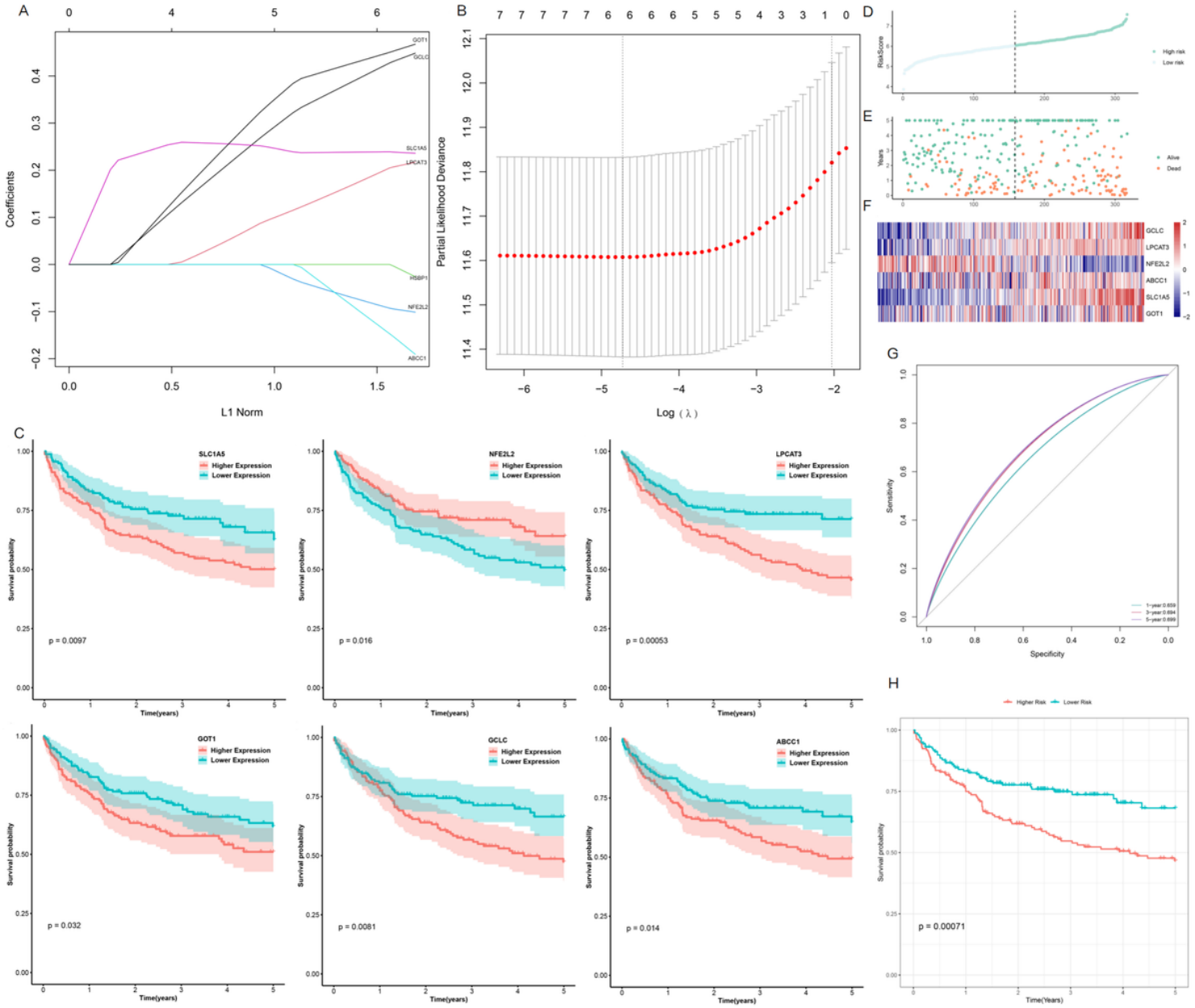

## Figure 2

Identification and functional enrichment analysis of the ferroptosis-related DEGs. (A) Volcano plot of DEGs. Red dots represent up-regulated genes, blue dots represent down-regulated genes, and black dots represent genes with no differences. (B) Heatmap of DEGs to visualize gene expression levels. (C-F) The most significant or shared GO enrichment and KEGG pathways of the DEGs are displayed



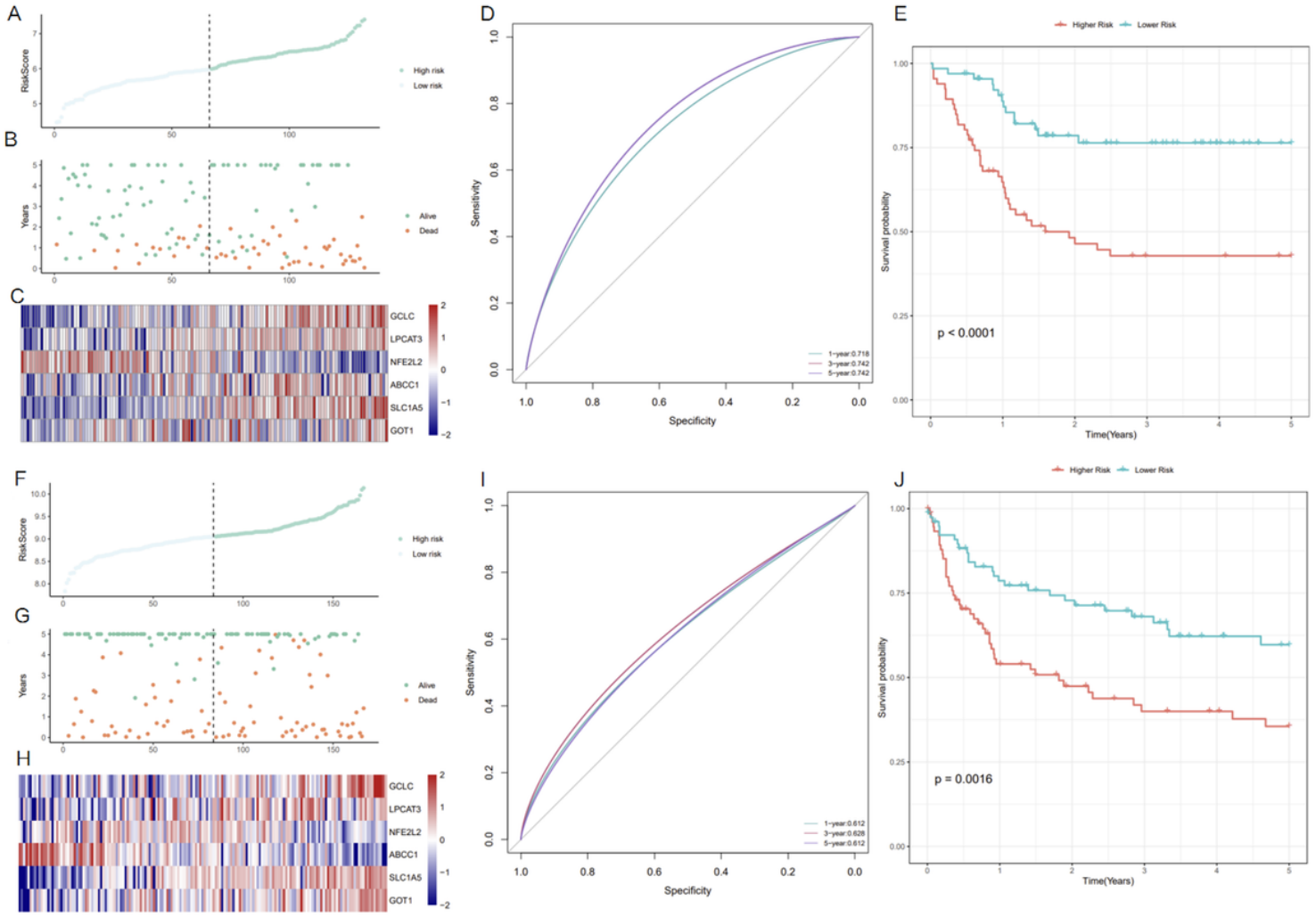

## Figure 3

Comparison of the immune score among different clusters. (A) The Stromal Score, Immune Score, and ESTIMATE Score between the two clusters. (B) The infiltration difference of 10 immune cells between the two subtypes. (C) Comparison of 28 immune cells between the two clusters. (D) Heatmap for immune responses based on ESTIMATE, MCPcounter, and ssGSEA algorithms in the two subtypes.



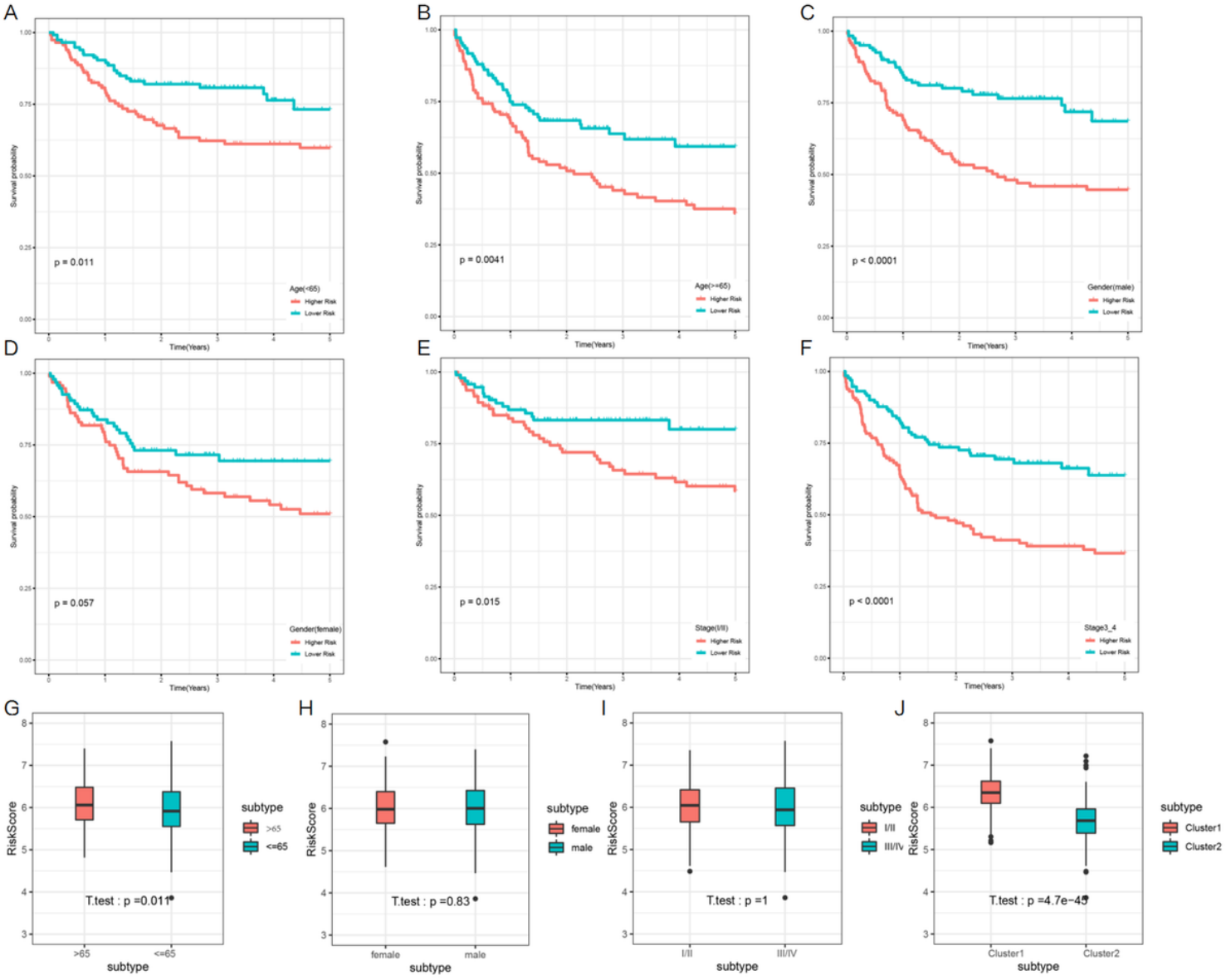

## Figure 4

Identification of a six-gene risk signature for OS by LASSO regression analysis in the training cohort. (A) LASSO coefficient spectrum of six genes in DLBCL. (B) Cross-validation for tuning parameter selection in the proportional hazards model. (C) Kaplan–Meier survival analysis for DLBCL patients with low and high expression of the six genes. (D-E) Risk survival status plot. (F) Heatmap of the six genes. (G) The survival curve shows that the OS status of high-risk group is significantly worse than that of low-risk group in the training cohort. (H) Receiver operating characteristic curve of the prognostic signature in the training cohort.



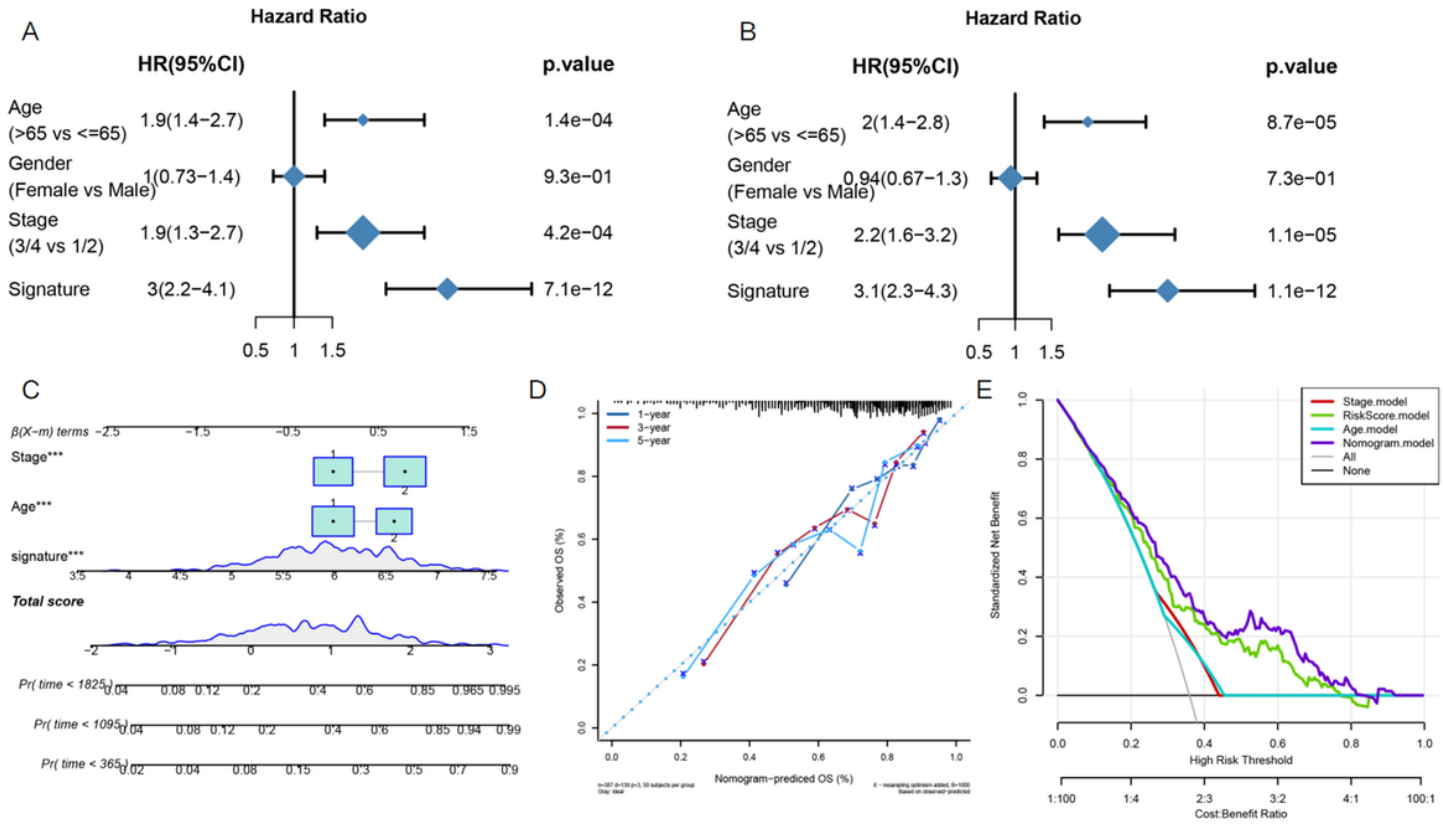

## Figure 5

Validation of the six-gene signature. (A-E) Validation of the six-gene signature for overall survival in the verification cohort. (F-J) Validation of the six-gene signature for overall survival in the independent cohort (GSE4475).




## Figure 6

Association between the ferroptosis-related signature and other clinical features in the training cohort. Survival of the FRG signature in patients stratified by gender, age, and stage. (A, B) The difference in OS between high- and low-risk groups stratified by age, A: <65y, B: >65y. (C, D) The difference in OS between high- and low-risk groups stratified by gender, C: male, D: female. (E, F) The difference in OS between high- and low-risk groups stratified by stage, E: stage I/II, F: stage III/IV. (G-J) The comparison of risk score in different age/gender/stage/molecular subtypes.




## Figure 7

Univariate and multivariate Cox analysis of prognostic risk scores for DLBCL. (A) Univariate Cox regression analysis. Forest plot of associations between age, gender, stage, risk factors and the survival of DLBCL patients. (B) Multivariate Cox regression analysis. The FRG signature is an independent predictor of DLBCL. (C) The calibration curve shows that using the nomogram to predict OS is highly consistent with the actual OS. (D) Establish a nomogram to predict the OS of patients. (E) DCA curves determined that the nomogram could provide optimal clinical decision-making benefits.

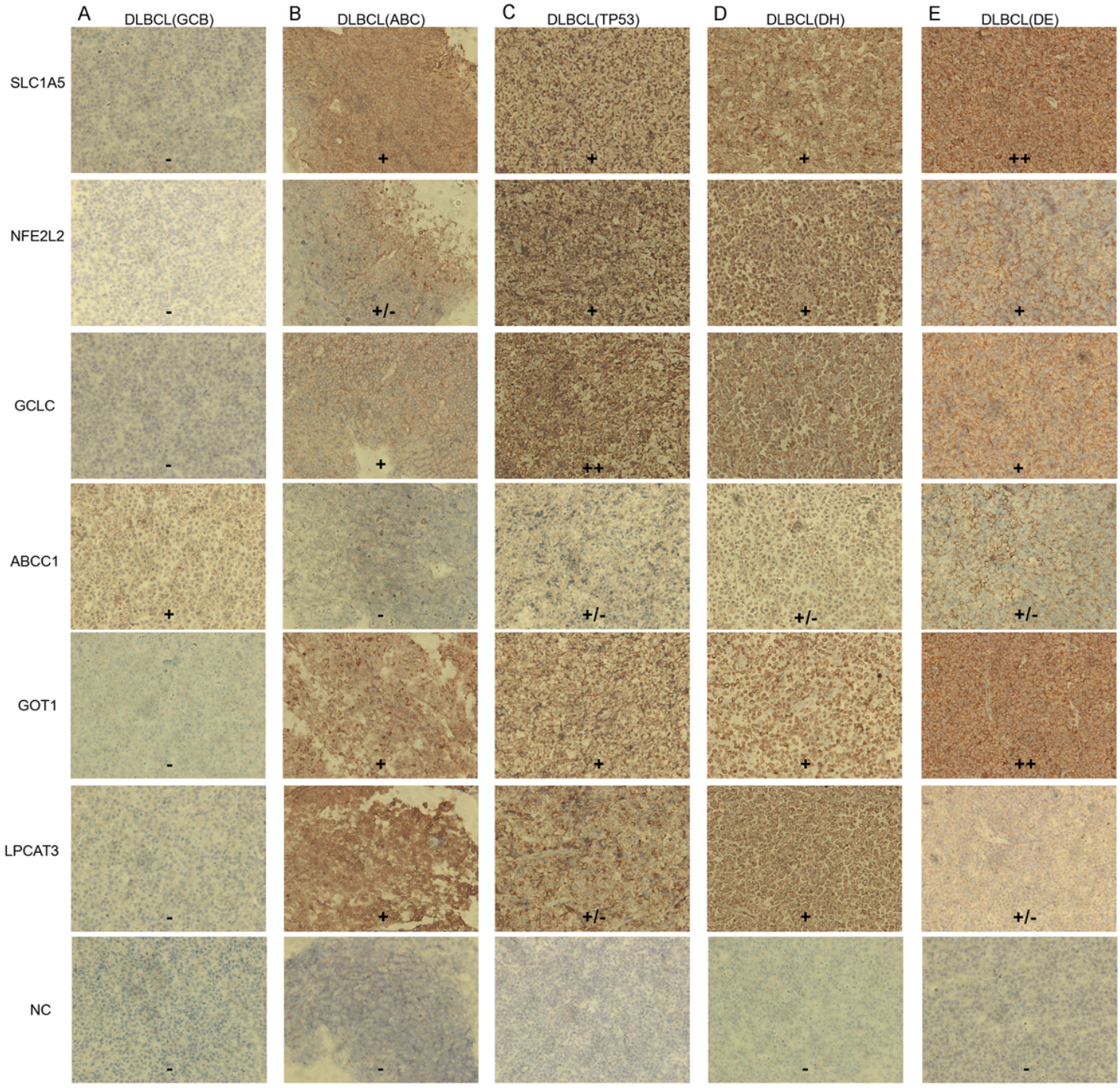



**Figure 8**

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