



OPEN ACCESS

EDITED BY
Andrea Ladányi,
National Institute of Oncology (NIO),
Hungary

*CORRESPONDENCE
Yang Liu,
liuyanglyon@uestc.edu.cn,
liuyang@scszlyy.org.cn

[†]These authors share first authorship

RECEIVED 08 September 2022

ACCEPTED 16 January 2023

PUBLISHED 02 February 2023

CITATION

Hou J, Guo P, Lu Y, Jin X, Liang K, Zhao N, Xue S, Zhou C, Wang G, Zhu X, Hong H, Chen Y, Lu H, Wang W, Xu C, Han Y, Cai S and Liu Y (2023), A prognostic 15-gene model based on differentially expressed genes among metabolic subtypes in diffuse large B-cell lymphoma. *Pathol. Oncol. Res.* 29:1610819. doi: 10.3389/pore.2023.1610819

COPYRIGHT

© 2023 Hou, Guo, Lu, Jin, Liang, Zhao, Xue, Zhou, Wang, Zhu, Hong, Chen, Lu, Wang, Xu, Han, Cai and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A prognostic 15-gene model based on differentially expressed genes among metabolic subtypes in diffuse large B-cell lymphoma

Jun Hou^{1†}, Peng Guo^{1†}, Yujiao Lu^{2†}, Xiaokang Jin², Ke Liang², Na Zhao¹, Shunxu Xue¹, Chengmin Zhou¹, Guoqiang Wang², Xin Zhu², Huangming Hong³, Yungchang Chen³, Huafei Lu², Wenxian Wang⁴, Chunwei Xu⁵, Yusheng Han², Shangli Cai² and Yang Liu^{1*}

¹Department of Pathology, Sichuan Cancer Hospital and Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China, ²Burning Rock Biotech, Guangzhou, China, ³Medical Oncology, Sichuan Cancer Hospital and Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China, ⁴Department of Clinical Trial, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China, ⁵Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, China

The outcomes of patients with diffuse large B-cell lymphoma (DLBCL) vary widely, and about 40% of them could not be cured by the standard first-line treatment, R-CHOP, which could be due to the high heterogeneity of DLBCL. Here, we aim to construct a prognostic model based on the genetic signature of metabolic heterogeneity of DLBCL to explore therapeutic strategies for DLBCL patients. Clinical and transcriptomic data of one training and four validation cohorts of DLBCL were obtained from the GEO database. Metabolic subtypes were identified by PAM clustering of 1,916 metabolic genes in the 7 major metabolic pathways in the training cohort. DEGs among the metabolic clusters were then analyzed. In total, 108 prognosis-related DEGs were identified. Through univariable Cox and LASSO regression analyses, 15 DEGs were used to construct a risk score model. The overall survival (OS) and progression-free survival (PFS) of patients with high risk were significantly worse than those with low risk (OS: HR 2.86, 95%CI 2.04–4.01, $p < 0.001$; PFS: HR 2.42, 95% CI 1.77–3.31, $p < 0.001$). This model was also associated with OS in the four independent validation datasets (GSE10846: HR 1.65, $p = 0.002$; GSE53786: HR 2.05, $p = 0.02$; GSE87371: HR 1.85, $p = 0.027$; GSE23051: HR 6.16, $p = 0.007$) and PFS in the two validation datasets (GSE87371: HR 1.67, $p = 0.033$; GSE23051: HR 2.74, $p = 0.049$). Multivariable Cox analysis showed that in all datasets, the risk model could predict OS independent of clinical prognosis factors ($p < 0.05$). Compared with the high-risk group, patients in the low-risk group predictively respond to R-CHOP ($p = 0.0042$), PI3K inhibitor ($p < 0.05$), and proteasome inhibitor ($p < 0.05$). Therefore, in this study, we developed a signature model of 15 DEGs among 3 metabolic subtypes, which could predict survival and drug sensitivity in DLBCL patients.

KEYWORDS

prognosis, risk score, drug sensitivity, DEGs, diffuse large B-cell lymphoma (DLBCL), metabolic subtypes

Introduction

Diffuse large B-cell lymphoma (DLBCL), a type of highly heterogeneous cancer, accounts for 30%–40% of non-Hodgkin lymphoma (1). The prognosis of DLBCL varies due to its distinct characteristics such as clinical factors, recurring mutations, cell of origin (COO) (2), etc. Currently, while approximately 60% of DLBCL patients could be cured by R-CHOP (a combined therapy of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone), the rest still suffer from a poor prognosis with fatal recurrent or progressive disease (3). Hence, effective prognostic stratification systems for DLBCL could benefit these patients in clinical decision-making and treatment.

At present, there are several commonly used prognosis-related classification systems for DLBCL in clinical practice, especially International Prognostic Index (IPI) (4) and COO (5, 6). Although IPI is convenient for application, its limitations are also obvious, in which only clinical factors are used and heterogeneous features of DLBCL, such as intrinsic genes and other biomarkers, are not considered. In the COO classification system, DLBCL is divided into germinal center B-cell-like (GCB) and non-GCB including activated B-cell-like (ABC) and type-III DLBCL based on immunohistochemistry algorithms (5) or gene expression profiling (GEP) analysis (6). Patients with GCB-DLBCL have better outcomes than those with non-GCB (6). However, the COO classification is mainly based on the B cell receptor (BCR) signaling pathway (7), and the isotype of BCR alone can predict the prognosis similar to COO (8). In addition, neither IPI nor COO classification can predict the efficacy of most drugs, although BTK inhibitor (BTKi) ibrutinib, was reported more effective in ABC than GCB DLBCL in a phase 1/2 clinical trial (9). Therefore, prognostic models of novel biomarkers are needed to assist the therapeutic drug selection in DLBCL.

Cancer cells autonomously alter their metabolic flux to meet the demands for rapid growth and survival, including increased bioenergetic and biosynthetic, mitigating oxidative stress, and immune evasion, etc. (10). For instance, the Warburg effect is a classic alteration in carbohydrate metabolism in tumors (11). Conversely, aberrantly accumulated metabolites also promote tumorigenesis (12). Targeting metabolic alterations has been considered a promising therapeutic strategy in some cancer types (13, 14). Meanwhile, there is a great interest in exploiting the relationship between metabolic gene expression and cancer prognosis stratification both in pan-cancer (15, 16, 17) and single solid tumors (18, 19, 20). In DLBCL, metabolism heterogeneity was reported a long time ago (21). However, just limited attention has been paid to correlating risk signatures to metabolic alterations. Therefore, the heterogeneity of metabolic expression profiles could be a novel perspective on prognosis stratification for DLBCL patients.

In this study, we identified three subtypes of DLBCL based on expression levels of genes in metabolism pathways. With the selected DEGs among these subtypes, we developed a risk score model to help stratify the survival of DLBCL patients and analyzed the relationships between the risk score and clinicopathological characteristics, drug sensitivity, and immune cell infiltrations. Our research provided a new prognosis model for DLBCL patients, rendering novel insights into the individual management of DLBCL.

Methods

Data obtaining

Microarray gene expression profiles and corresponding clinical-related information of the training dataset (GSE31312) and four validation datasets (GSE10846, GSE53786, GSE87371, and GSE23501) were downloaded from the NCBI Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). Excluding 0-month survival samples, 470, 412, 119, 221, and 69 DLBCL patients were respectively included in GSE31312, GSE10846, GSE53786, GSE87371, and GSE23501. For genes with multiple probes, the median value of gene expression was used in the following analyses. The clinicopathological data of the patients in each dataset were summarized in Supplementary Table S1.

Metabolic subtype classification

Overall, 1,916 metabolic genes of 7 major metabolic processes from Peng et al. (16) were initially studied and listed in Supplementary Table S2. The expression profiles of these 1,916 metabolic genes in GSE31312 were employed to perform Partitioning Around Medoids (PAM) clustering by the “ConsensusClusterPlus (v1.50.0)” R package with Euclidean distance. The distribution of clinical characteristics of patients was analyzed among subtypes. The “survival” and “survminer” R packages were used to analyze survival differences among subtypes, and the result was shown by the Kaplan-Meier survival curve.

ssGSEA

Single sample gene set enrichment analysis (ssGSEA) was applied by the “GSVA” R package using the 50 hallmark gene sets from the Molecular Signatures Database (MSigDB, <http://www.gsea-msigdb.org/gsea/index.jsp>). The result was shown as a heatmap by the “ComplexHeatmap” R package.

Differentially expressed gene (DEG) analysis

Fold change (FC) was calculated by dividing the mean value of gene expression in one subtype group by that in the other subtype group. Wilcoxon signed-rank test was employed to calculate p value. The genes with $|\log_2FC| > 0.585$ and Benjamini-Hochberg-adjusted $p < 0.05$ were identified as DEGs. For DEGs in these subtype groups, KEGG and GO enrichment analyses were performed independently using “clusterProfiler” R package.

Construction and validation of the risk score model

First, the p value of <0.01 was used as the threshold to select DEGs with prognostic significance in univariable Cox regression analysis. Then, the selected DEGs were analyzed by least absolute shrinkage and selection operator (LASSO)-Cox regression analysis through the “glmnet” and “survival” packages. Ten-fold cross-validation was employed to determine the penalty parameter (λ) of the prognostic model and followed minimum criteria. The formula below was to calculate the risk score based on the expression level of each gene and its corresponding regression coefficient:

$$\text{Risk Score} = \sum_{i=0}^n \beta_i * \chi_i$$

β_i : weight coefficient of each gene; χ_i : expression quantity of each gene.

The high- and low-risk groups were divided according to the median value. Overall survival (OS) and progression-free survival (PFS) of patients between the two groups were compared by the Kaplan-Meier survival curves using the “survminer” package. Then, the signature was validated in four external datasets (GSE10846, GSE53786, GSE87371, and GSE23051).

Univariable and multivariable Cox analyses were conducted to determine whether the risk score was an independent prediction factor of OS and PFS in the training and four validation cohorts. Meanwhile, receiver operating characteristic (ROC) curves in the above datasets were constructed and determined the AUC values through the “timeROC” package.

Association analysis between clinicopathological characteristics and risk score

The risk scores were compared in different groups of age, sex, stage, extranodal sites, ECOG score, IPI, and GEP subtype, separately.

Drug sensitivity analysis

The drug sensitivity of different risk groups was predicted by the data from The Genomics of Drug Sensitivity in Cancer (GDSC) database (<https://www.cancerrxgene.org/>) (22). The half maximal inhibitory concentration (IC50) was analyzed using “pRRophic” R package (23).

Immune cell infiltration analysis

According to the LM22 gene signature of tumor-infiltrating immune cells (TIICs) pattern for distinguishing human immune cell phenotypes (24), the fraction of TIICs was analyzed by CIBERSORT (the Cell-type Identification by Estimating Relative Subsets of RNA Transcripts, <http://cibersort.stanford.edu/>).

Statistical analysis

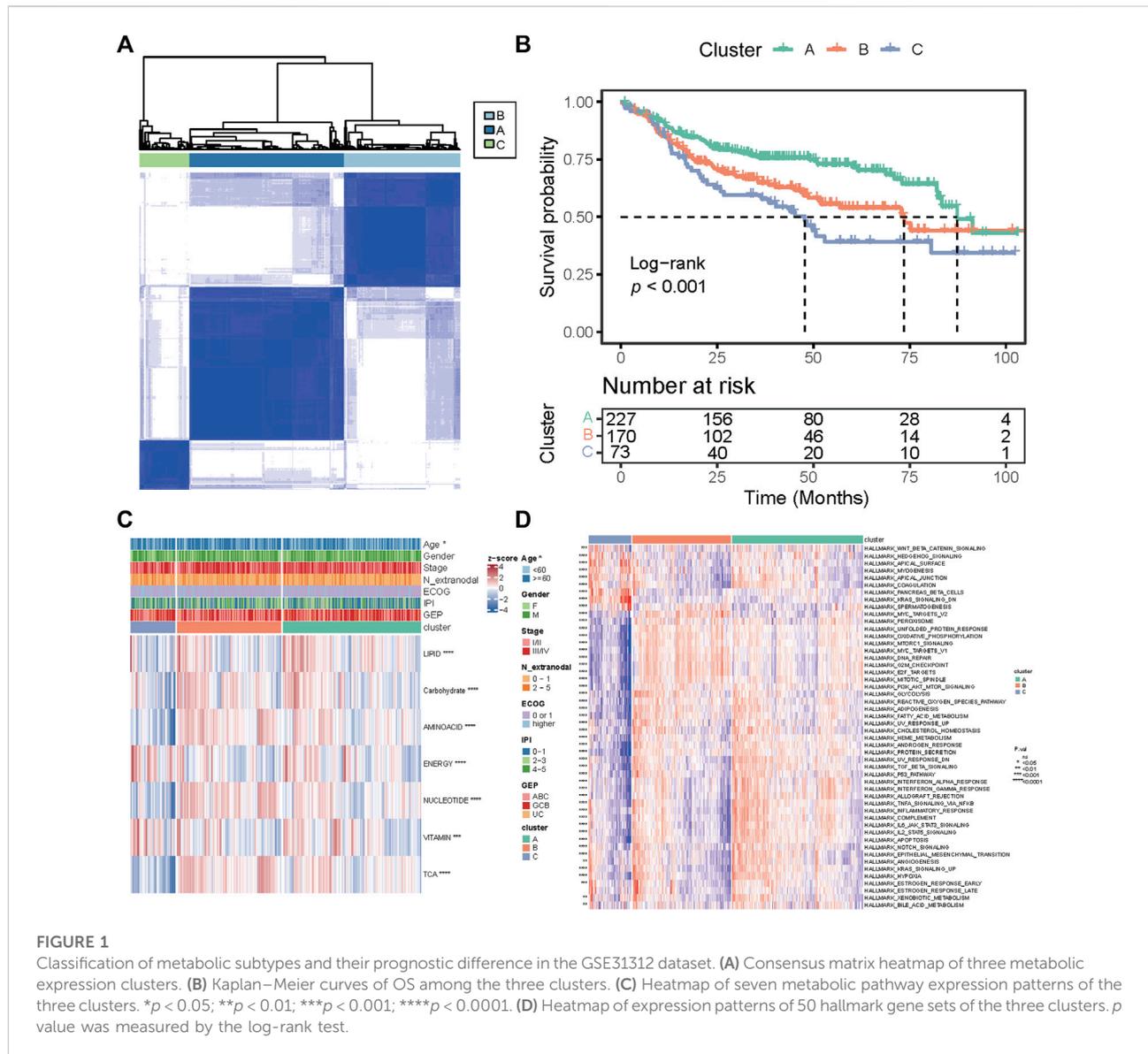
Statistical analyses and mapping of data were performed by the R software (version 4.1.2; <https://www.R-project.org>). Continuous variables were compared by Mann-Whitney U test or Kruskal-Wallis test. Categorical variables were compared by Fisher’s exact test. A two-tailed p value <0.05 indicated statistical significance.

Results

Classification of metabolic subtypes and their prognostic differences

The overall study plan was depicted in Supplementary Figure S1. To characterize the metabolic heterogeneity of DLBCL patients, we analyzed 1,916 genes in 7 metabolic pathways (16) in the GSE31312 cohort, including 766 genes in the lipid metabolism pathway, 286 genes in the carbohydrate metabolism pathway, 348 genes in the amino acid metabolism pathway, 110 genes in the integration of energy pathway, 90 genes in the nucleotide metabolism pathway, 168 genes in the vitamin cofactor metabolism pathway, and 148 genes in the tricarboxylic acid cycle (TCA cycle) pathway. Using an unsupervised consensus algorithm, 470 patients were divided into 3 metabolic subtypes (clusters A, B, and C with 227, 170, and 73 patients, respectively) in the training cohort (Figure 1A). The respective median OS of clusters A, B, and C was 87.3 months, 73.5 months, and 47.7 months (Log-rank $p < 0.001$; Figure 1B). Notably, the median OS of cluster C was significantly shorter than those of clusters A and B.

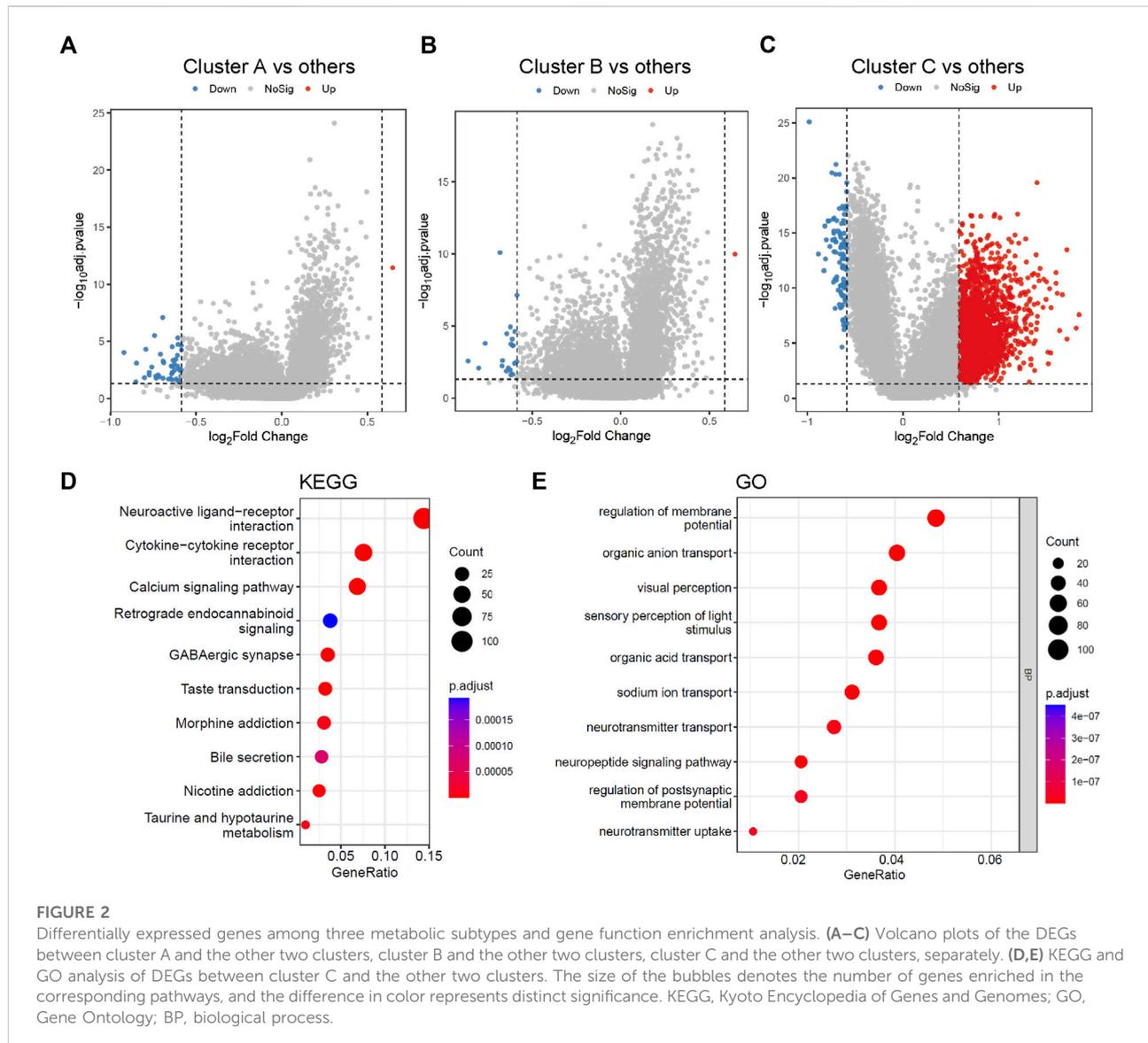
We further investigated the differences in pathway enrichment among the three metabolic subtypes. We found



that compared with clusters A and B, cluster C had lower expression of genes in the nucleotide metabolism, TCA cycle, and amino acid metabolism and higher expression of genes in the integration of energy and vitamin cofactor metabolism (Figure 1C). The clinical variable analyses showed that patients in cluster C were older ($p = 0.044$) than those in other clusters (Figure 1C). Then, we conducted ssGSEA to assess the differential expression levels of 50 biological hallmarks in the three groups. The expression pattern of those hallmarks in cluster C was different from that in other clusters (Figure 1D). These results indicate that the differences in the expression of genes in seven major metabolic pathways could stratify the prognosis of DLBCL.

Identification of DEGs and functional annotation

To select the metabolic genetic signature, DEG analysis was conducted and identified 1,854 DEGs. There were 43, 21, and 102 downregulated genes and 1, 1, 1,743 upregulated genes in clusters A, B, and C, respectively (Figures 2A–C, Supplementary Table S3). Functional analyses of those DEGs *via* KEGG pathway and GO analyses were not able to provide any results for clusters A and B, probably due to the small sizes of DEGs sets (44 vs. 22). DEGs in cluster C were more related to neuroactive ligand-receptor interaction, cytokine-cytokine receptor interaction, and calcium signaling pathway (Figure 2D). Biological processes associated with the regulation of membrane potential, organic



anion transport, organic acid transport, and sodium ion transport were also enriched in cluster C (Figure 2E).

Construction of the risk score model

To construct a risk score model, we analyzed 1,854 DEGs identified above using univariable Cox regression. The threshold of $p < 0.01$ were used to screen genes that were most related to the prognosis of DLBCL patients. A total of 108 prognosis-associated genes were selected in the training set (Supplementary Table S4). The top-20 prognosis-associated genes, according to the significance level (p value), are listed in Figure 3A. Then, we performed LASSO penalty regression to construct the prognostic model in the

training set (Figure 3B). A subset of 15 genes and their weighting coefficients were finally identified (Figure 3C; Supplementary Table S5). Furthermore, the risk score of individual patients was calculated, and all patients were dichotomized into high- or low-risk groups according to the median value of the risk score.

In the training dataset, patients in the high-risk group had both shorter OS (HR 2.86, 95% CI 2.04–4.01; $p < 0.001$; Figure 3D) and PFS (HR 2.42, 95% CI 1.77–3.31; $p < 0.001$; Figure 3E) than those in the low-risk group. The areas under the receiver operating characteristic curve (AUCs) for 1-, 3-, and 5- years OS were 0.701, 0.703, and 0.724, respectively (Supplementary Figure S2A). For 1-, 3-, and 5- years PFS, the AUCs were 0.667, 0.699, and 0.685, respectively (Supplementary Figure S2B). Moreover, the

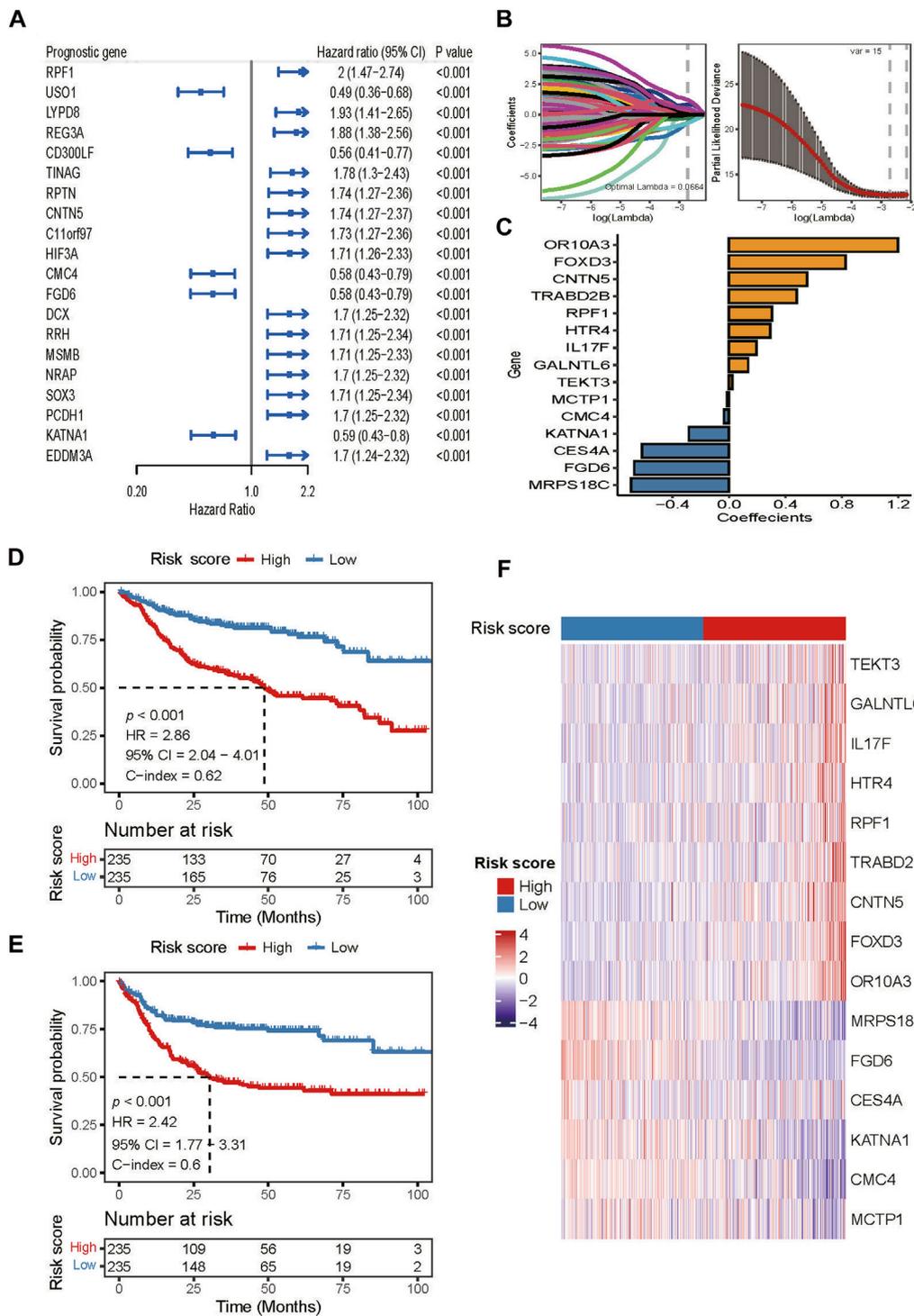


FIGURE 3

Construction of the risk score model in the training dataset. **(A)** Top 20 of the 108 prognosis-related DEGs identified by univariable Cox regression analysis. **(B)** LASSO regression analysis of the 108 prognosis-related DEGs. **(C)** Fifteen DEGs and their coefficients used for constructing the risk score. **(D,E)** Kaplan-Meier curves of OS and PFS of patients in the training dataset assigned to high and low-risk groups. **(F)** Heatmap of gene expression patterns of the 15 model-used genes in patients assigned to high- and low-risk groups.

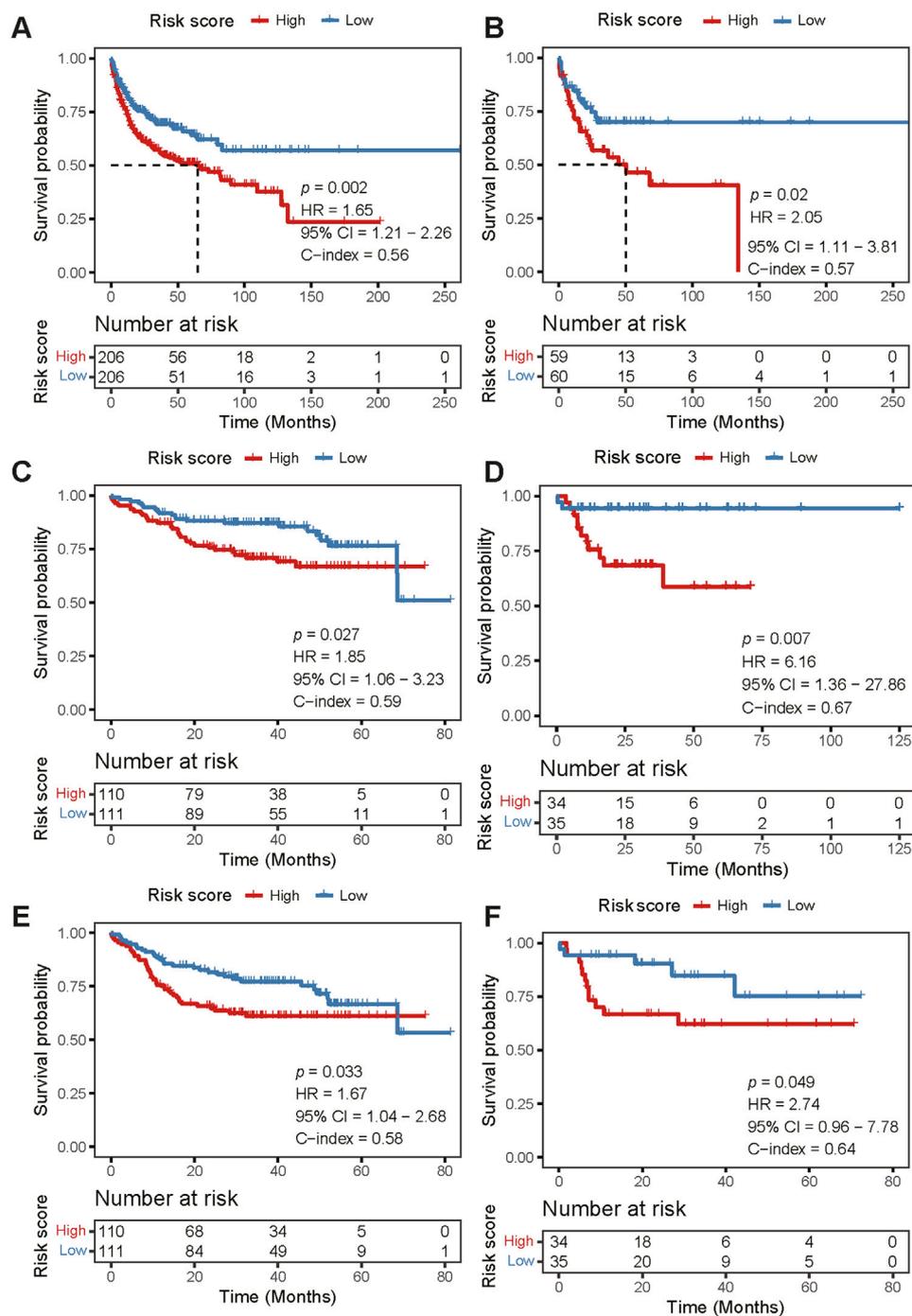


FIGURE 4

External validation of the risk score model. (A–D) Kaplan-Meier curve analysis of OS between the high- and low-risk groups in four independent validation cohorts [GSE10846 (A), GSE53786 (B), GSE87371 (C), and GSE23501 (D), respectively]. (E,F) Kaplan-Meier curve analysis of PFS between the high- and low-risk groups in two independent cohorts [GSE87371 (E) and GSE23501 (F), respectively]. There was no PFS information in GSE10846 and GSE53786. OS, overall survival; PFS, progression-free survival.

gene expression heatmap revealed that the high expressions of genes, *OR10A3*, *FOXD3*, *CNTN5*, *TRABD2B*, *RPF1*, *HTR4*, *IL17F*, *GALNTL6*, and *TEKT3* were observed in the

high-risk group, while in the low-risk group, *MCTP1*, *CMC4*, *KATNA1*, *CES4A*, *FGD6*, and *MRPS18C* were highly expressed (Figure 3F).

Validation of risk score model in the independent validation cohorts

We further validated the risk score model in four external cohorts, the robustness of the prognostic model was supported by significant prognostic values for OS in GSE10846 (HR 1.65, 95% CI 1.21–2.26; $p = 0.002$; Figure 4A), GSE53786 (HR 2.05, 95% CI 1.11–3.81; $p = 0.02$; Figure 4B), GSE87371 (HR 1.85, 95% CI 1.06–3.23; $p = 0.027$; Figure 4C), and GSE23051 (HR 6.16, 95% CI 1.36–27.86; $p = 0.007$; Figure 4D). For PFS validation, the results showed the same trend with statistical significance (GSE87371: HR 1.67, 95% CI 1.04–2.68; $p = 0.033$; Figure 4E; GSE23051: HR 2.74, 95% CI 0.96–7.78; $p = 0.049$; Figure 4F). In the validation datasets, the AUCs of the model were higher in GSE23501 for both OS (Supplementary Figures S2C–F) and PFS (Supplementary Figures S2G, H) for 1-, 3-, and 5- years. Therefore, compared with the high-risk group, patients in the low-risk group had better outcomes, indicating the predictive potential of our model.

Independent prognostic role of the risk score model

The independent prognostic value of the model was further studied, taking into consideration of age, sex, stage, ECOG, and IPI. The risk score was verified to be an independent prognostic factor of OS in all cohorts in both univariable and multivariable Cox regression analyses ($p < 0.05$; Supplementary Figures S3A–E) and an independent prognostic factor of PFS in GSE31312 ($p < 0.001$; Supplementary Figure S3F) and GSE87371 ($p = 0.015$; Supplementary Figure S3G). As expected, IPI was also an independent factor to predict PFS and OS (Supplementary Figure S3).

Association between clinicopathological characteristics and risk score

Further analysis of clinical characteristics showed that the risk score was higher in patients with the following characteristics: age >60 years ($p = 0.0018$), stage III/IV disease ($p = 0.029$), >1 extranodal sites ($p = 0.024$), higher IPI scores ($p = 0.00017$), or cluster C ($p < 0.01$) (Supplementary Figure S4).

Association between drug sensitivity and risk score

All patients in the training cohort were treated with the R-CHOP regimen, the first-line standard-of-care treatment for DLBCL (25). We studied the response rates to the

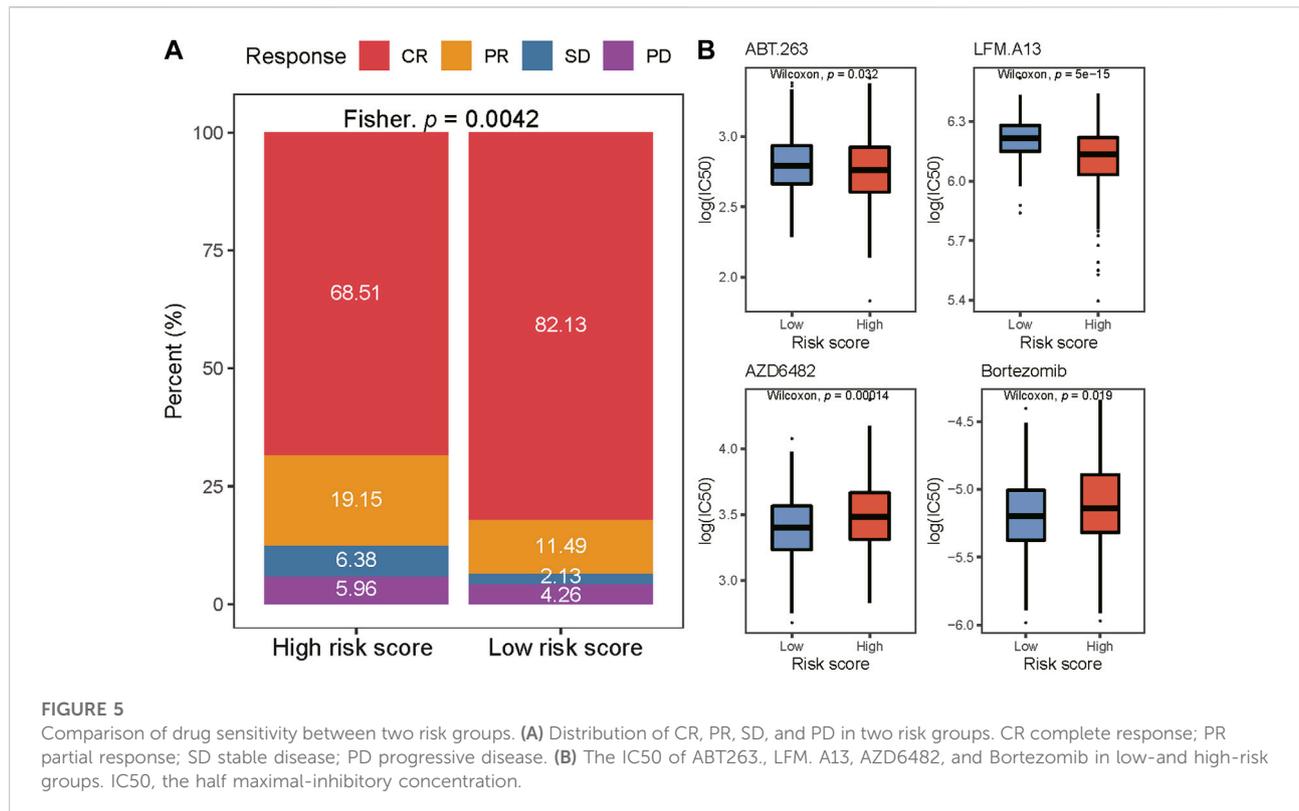
R-CHOP in the high- and low-risk groups. We found that the complete response (CR) rate was higher in the low-risk group (82.1% vs. 68.5%, $p = 0.0042$; Figure 5A). Then, we used the data from the GDSC database to predict the response to targeted agents in the two risk groups (Figure 5B). The estimated IC50s for BCL2i (ABT.263) and BTKi (LFM.A13) were lower in the high-risk group ($p < 0.05$), while estimated IC50s for PI3K inhibitor (PI3Ki and AZD6482) and proteasome inhibitors (Bortezomib) were lower in the low-risk group ($p < 0.05$).

Characteristics of immune cell infiltration of the two risk score groups

Tumor-infiltrating immune cells (TIICs), a component of the tumor microenvironment (TME), have been found to be associated with prognosis and treatment response. To explore the relationship between the risk score and TIICs, we analyzed the discrepancy of immune cell infiltration between two risk groups according to the LM22 gene signature[24] (Figure 6A). The proportion of naïve B cells, eosinophils, M1 Macrophages, activated CD4 memory T cells, resting CD4 memory T cells, follicular helper T cells, M0 Macrophages, and gamma delta ($\gamma\delta$) T cells were significantly higher in the low-risk group ($p < 0.05$). By contrast, neutrophils, resting NK cells, naïve CD4 T cells, and regulatory T cells (T-regs) were notably higher in the high-risk group ($p < 0.05$). In the training cohort, M1 macrophages, memory B cells, $\gamma\delta$ T cells, activated CD4 memory T cells, neutrophils, and CD8 T cells were found at the core of the correlation network (Figure 6B). The correlation heatmap showed that activated CD4 memory T cells correlated positively with $\gamma\delta$ T cells and negatively with T-regs (Figure 6C, Supplementary Figure S5). Altogether, these results suggest a remarkable discrepancy in immune cell infiltration between the high- and low-risk groups while the potential mechanism may be complex.

Discussion

DLBCL is a challenge in individualized treatments due to its significant heterogeneity. Although R-CHOP can cure over 60% of patients (2, 3) using traditional stratification systems like IPI (26) and COO (5, 6), a limitation of these systems is evident, either neglect biological factors (26) or only focus on the BCR signal (7, 8). Importantly, none of them can predict drug response. Metabolic signatures have been proposed for prognosis in many other neoplasms (18, 19, 27, 28) but not in DLBCL yet. Therefore, we distinguished three metabolic subtypes with different prognoses in DLBCL. Among them, 108 prognosis-associated DEGs were identified, of which

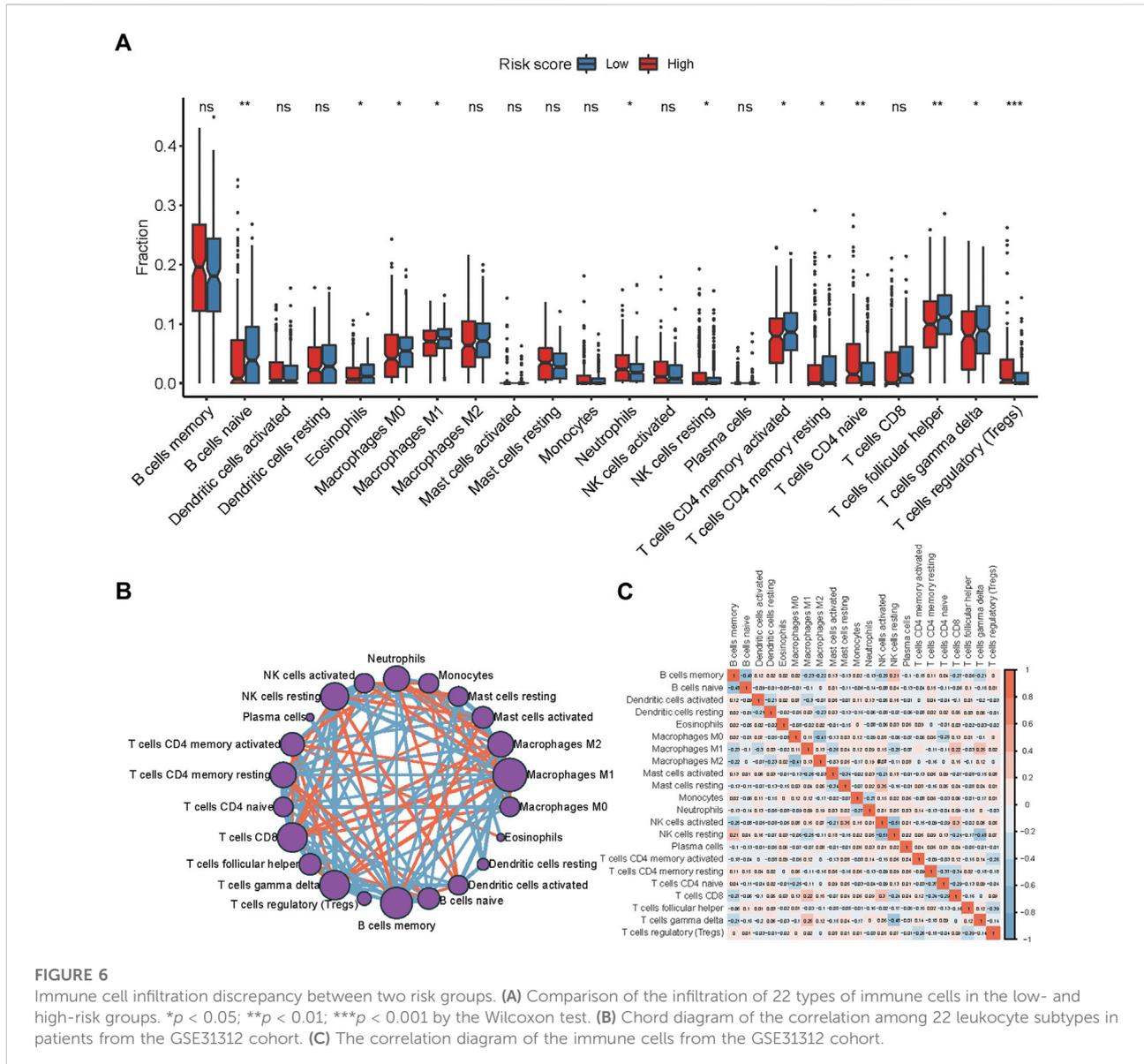


15 genes were used to construct a risk score model of DLBCL. The model showed a robust ability to predict the outcomes of DLBCL independently. As expected, there were distinctly different immune cell infiltration and clinicopathological characteristics between the high- and low-risk groups. The results also imply that patients in the high-risk group predictively respond to BCL2 and BTK inhibitors, while patients in the low-risk group might consider PI3Ki and proteasome inhibitors.

In this study, we explored the metabolic genetic features in DLBCL and identified three subtypes of patients with distinct survival. The most distinct differences were in cluster C, which had the shortest OS with lower expression of genes in nucleotide metabolism, TCA cycle, or amino acid metabolism and higher expression of genes in the integration of energy and vitamin cofactor metabolism. Interestingly, another metabolic expression subtype of 32 cancers also showed the clinical outcomes of subtypes with upregulated vitamin/cofactor metabolism were worse, while subtypes with upregulated nucleotide metabolism had a better prognosis (16). However, there were some inconsistent results. For instance, increased expressions of numerous nucleotide metabolism genes were associated with worse outcomes in breast cancer patients (29). The inconsistency may attribute to the following reasons: 1) previous studies merely focused on one or several metabolites and genes of an individual metabolic pathway (29,30,31), 2)

signaling pathways may interact with each other (20, 32), and 3) the same signaling could play distinct roles in different diseases. In summary, the results suggest metabolic heterogeneity in DLBCL can be used for prognostic stratification though it needs more validation.

Based on DEGs among the three metabolic subtypes, a risk score model of 15 genes was developed. Among the 15 genes, *OR10A3*, *FOXD3*, *CNTN5*, *TRABD2B*, *RPF1*, *HTR4*, *IL17F*, *GALNTL6*, and *TEKT3* were highly expressed in the high-risk group, which was associated with poor prognosis, while the high expression of *MRPS18C*, *FGD6*, *CES4A*, *KATNA1*, *CMC4*, and *MCTP1* were characterized in the low-risk group and related to favorable prognosis. Among the poor prognosis-related genes, *TRABD2B* (also known as *TIK12*), *IL17F*, and *GALNTL6* were reported to be related to the oncogenesis of renal cell carcinoma (33), cutaneous T-cell lymphoma (34), and thyroid carcinoma (35), respectively. Moreover, the mutations in *CNTN5* were reported to contribute to the metastatic process of pancreatic cancer (36), *HTR4* was found predominantly in only high-grade prostate cancer (37, 38), and the high expression of *TEKT3* could be influenced by HBV integration events in liver cancer (39). Therefore, these genes promoting pathogenesis or progression in several cancers may also lead to poor survival in DLBCL. However, *FOXD3*, a poor prognosis gene in our model, was reported as a suppressor factor of *H pylori* infection-induced gastric carcinoma (40) and melanoma (41). One possible



explanation for this difference is the dual role of forkhead Box D3 encoded by *FOXD3*. Forkhead Box D3, a member of the forkhead family of transcription factors, can function as both a transcriptional repressor and activator. Of the favorable prognosis-related genes, previous studies have reported that the downregulated *MCTP1* was related to drug-resistance of esophageal cancer (42), and *MRPS18C* was the least expressed *MRPS18* family member in malignantly transformed B-cells (43). Thus, the higher expression of these genes may be associated with favorable prognosis. The expression of *FGD* genes, a gene family comprising *FGD6*, were analyzed to predict the OS of head and neck squamous cell carcinoma (HNSC) (44), in which the OS was positively related to high expression of *FGD2* and *FGD3*

but not *FGD6*. The favorable effect of *FGD6* on prognosis in DLBCL patients needs further investigation. In addition, the over-expression of *CMC4* (also known as *MTCP1*), as a favorable prognosis gene in our model, was discordantly reported to produce clonal *CD5⁺/CD19⁺* leukemia in mice (45), which was thought to be a chronic lymphocytic leukemia driving gene. For the remaining genes, *OR10A3*, *RPF1*, *KATNA1*, and *CES4A*, in our risk score model, no specific relationship to cancer had been reported yet, further exploration should be carried out for their roles in the prognosis in DLBCL patients.

The relationship of risk score with clinical characteristics and treatment response was explored in our study. Patients with an age older than 60 years, advanced stage (stage III/IV), and high

IPI scores had higher risk scores. Older than 60 years, LDH (lactate dehydrogenase) greater than normal, PS score of 2–4 points, stage III/IV, and more than 1 extranodal sites are well-known high-risk factors in DLBCL (4), which is consistent with our results. Meanwhile, a greater proportion of patients in the cluster C subtype was observed in the high-risk group, which is also consistent with our result that cluster C was a poor prognosis factor in our research. In addition, our results showed that low-risk patients had a significantly higher CR rate than high-risk patients after R-CHOP treatment. Previous studies have confirmed that the higher CR/CRu rate of DLBCL patients after chemotherapy would improve the overall survival (46, 47), suggesting that the higher sensitivity to the R-CHOP regimen in the low-risk group may be another reason for its favorable prognosis.

DLBCL is a heterogeneous lymphoma (48). In this study, we observed immune cell infiltration discrepancy in two risk groups. A higher proportion of activated CD4 memory T cells, M1 macrophages, and $\gamma\delta$ T cells were in the low-risk group relevant to better prognosis. These results were consistent with previous studies. Chen et al. reported that when patients had a higher proportion of CD4 memory T cells and $\gamma\delta$ T cells, they were more sensitive to R-CHOP regimen so that more patients achieved CR/PR (49). Another study also showed that activated CD4 memory T cell was an independent factor of favorable prognosis in DLBCL patients (50). The reason may be that after chemotherapy, CD4⁺ T cells can produce multiple proinflammatory cytokines including IFN γ , TNF α , and IL2 (51). These factors may allow patients to achieve durable remissions through CD8⁺ effector cell-mediated antitumor immunity (51). For $\gamma\delta$ T cells, their cytotoxic effect and ability to secrete IFN can generate antitumor effects (52). Its subgroup, V γ 9V δ 2T cells, can increase antibody dependent cellular cytotoxicity (ADCC) of rituximab and further enhance the efficacy of the R-CHOP regimen (53). Meanwhile, Yan et al. (54) found that M1 macrophage infiltration was related to a lower risk of progression and improved overall survival, as FC γ R-dependent stimulation of M1 macrophage mediated ADCP (antibody-dependent cellular phagocytosis) maintained anti-lymphoma activity. In addition, T-regs were highly expressed in the high-risk group, suggesting that T-regs are relevant to poor prognosis. The prognostic role of T-regs is now a matter of debate. Autio et al. (55) found in the Nordic Lymphoma Group trial cohort (NCT01325194), a higher proportion of T-regs was associated with worse prognosis, but this could not be repeated in the Helsinki study cohort (NCT01502982). These controversial results may be caused by the heterogeneity of DLBCL. Interestingly, in this study, we found that CD4 memory-activated T cells were positively related to $\gamma\delta$ T cells but negatively correlated with T-regs. This mechanism of the correlation between those immune cells is needed in the future.

In our study, we also explored the potential response of the high- and low-risk groups to targeted drugs, in which BCL2 inhibitor and PI3K inhibitor were suggested for the high- and low-risk groups, respectively. Interestingly, in the previous classification based on genetic heterogeneity (3), the BCL2 SVs were associated with poor outcomes of GCB-DLBCLs and PI3K with good-risk GCB-DLBCLs, while BCL2 was considered in the EZB subtype of DLBCL with favorable outcomes in another genetic classification (56). A therapeutic classification of DLBCL was constructed based on the responses to drugs targeted at genetic alteration (57) and both BCL2 inhibitor and PI3K inhibitor were suggested for the MCD subgroup with poor survival and EZB subgroup with good survival. Hopefully, combining the study of Chapuy et al. (3), our results would provide more clues to make the decision of treatment for DLBCL.

Some limitations are in our study. Although we included a total of 1,291 patients and 4 independent validation sets, which suggests that the results may be highly reliable, the risk score model should be further verified through a prospective study. Second, our study was based on bioinformatic analyses of public data, while validations by clinical specimens are needed to be studied. Last, the mechanism of how the 15 genes in the risk score model affect prognosis in DLBCL needs to be further explored.

Conclusion

Overall, we identified three metabolic subtypes in DLBCL patients with different clinical outcomes and further constructed a prognostic 15-gene model based on DEGs among the three subtypes, which indicates that the differentially expressed gene profile of metabolic heterogeneity may provide a new strategy for prognosis stratification in DLBCL patients. Additionally, the risk score model demonstrated a remarkable predictive value of survival and drug sensitivity, which may benefit individualized prognosis management and personalized therapeutic intervention in DLBCL.

Data availability statement

The datasets generated and analyzed in the present study are available in the public data repository, GEO: <https://www.ncbi.nlm.nih.gov/geo/>.

Author contributions

JH, PG, YuL, XJ, and KL: processing data and conducting bioinformatic analysis. NZ, SX, CZ, and GW: writing original draft of the manuscript. XZ, HH, YC, HL, WW, CX, and YH:

reviewing and editing the manuscript. SC: final review of the manuscript. YaL: conception, design, supervision, and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by 2018 Entrepreneurial Leading Talent of Guangzhou Huangpu District and Guangzhou Development District (No. 2022-L023).

Conflict of interest

Authors YuL, XJ, KL, GW, XZ, HL, YH, and SC were employed by the company Burning Rock Biotech.

References

- Susanibar-Adaniya S, Barta SK. 2021 update on diffuse large B cell lymphoma: A review of current data and potential applications on risk stratification and management. *Am J Hematol* (2021) 96(5):617–29. doi:10.1002/ajh.26151
- Liu Y, Barta SK. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *Am J Hematol* (2019) 94(5):604–16. doi:10.1002/ajh.25460
- Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* (2018) 24(5):679–90. doi:10.1038/s41591-018-0016-8
- International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *New Engl J Med* (1993) 329(14):987–94. doi:10.1056/NEJM199309303291402
- Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* (2004) 103(1):275–82. doi:10.1182/blood-2003-05-1545
- Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New Engl J Med* (2002) 346(25):1937–47. doi:10.1056/NEJMoa012914
- Miao Y, Medeiros LJ, Xu-Monette ZY, Li J, Young KH. Dysregulation of cell survival in diffuse large B cell lymphoma: Mechanisms and therapeutic targets. *Front Oncol* (2019) 9:107. doi:10.3389/fonc.2019.00107
- Ruminy P, Etancelin P, Couronné L, Parmentier F, Rainville V, Mareschal S, et al. The isotype of the BCR as a surrogate for the GCB and ABC molecular subtypes in diffuse large B-cell lymphoma. *Leukemia* (2011) 25(4):681–8. doi:10.1038/leu.2010.302
- Wilson WH, Young RM, Schmitz R, Yang Y, Pittaluga S, Wright G, et al. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. *Nat Med* (2015) 21(8):922–6. doi:10.1038/nm.3884
- Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell* (2011) 144(5):646–74. doi:10.1016/j.cell.2011.02.013
- Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: The metabolic requirements of cell proliferation. *Science (New York, NY)* (2009) 324(5930):1029–33. doi:10.1126/science.1160809
- Martinez-Reyes I, Chandel NS. Cancer metabolism: Looking forward. *Nat Rev Cancer* (2021) 21(10):669–80. doi:10.1038/s41568-021-00378-6
- Fendt SM, Frezza C, Erez A. Targeting metabolic plasticity and flexibility dynamics for cancer therapy. *Cancer Discov* (2020) 10(12):1797–807. doi:10.1158/2159-8290.CD-20-0844
- Ricci JE, Chiche J. Metabolic reprogramming of non-hodgkin's B-cell lymphomas and potential therapeutic strategies. *Front Oncol* (2018) 8:556. doi:10.3389/fonc.2018.00556

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

Acknowledgments

We thank Wenchuan Xie, Jing Zhao and Leo Li for their useful discussion and suggestions.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.por-journal.com/articles/10.3389/pore.2023.1610819/full#supplementary-material>

- Gaude E, Frezza C. Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival. *Nat Commun* (2016) 7:13041. doi:10.1038/ncomms13041
- Peng X, Chen Z, Farshidfar F, Xu X, Lorenzi PL, Wang Y, et al. Molecular characterization and clinical relevance of metabolic expression subtypes in human cancers. *Cell Rep* (2018) 23(1):255–69.e4. doi:10.1016/j.celrep.2018.03.077
- Sung JY, Cheong JH. Pan-cancer analysis reveals distinct metabolic reprogramming in different epithelial-mesenchymal transition activity states. *Cancers (Basel)* (2021) 13(8):1778. doi:10.3390/cancers13081778
- Cui L, Xue H, Wen Z, Lu Z, Liu Y, Zhang Y. Prognostic roles of metabolic reprogramming-associated genes in patients with hepatocellular carcinoma. *Aging (Albany NY)* (2020) 12(21):22199–219. doi:10.18632/aging.104122
- Tan Z, Lei Y, Xu J, Shi S, Hua J, Zhang B, et al. The value of a metabolic reprogramming-related gene signature for pancreatic adenocarcinoma prognosis prediction. *Aging (Albany NY)* (2020) 12(23):24228–41. doi:10.18632/aging.104134
- Zhang Z, Zhu H, Li Q, Gao W, Zang D, Su W, et al. Gene expression profiling of tricarboxylic acid cycle and one carbon metabolism related genes for prognostic risk signature of colon carcinoma. *Front Genet* (2021) 12:647152. doi:10.3389/fgene.2021.647152
- Caro P, Kishan AU, Norberg E, Stanley IA, Chapuy B, Ficarro SB, et al. Metabolic signatures uncover distinct targets in molecular subsets of diffuse large B cell lymphoma. *Cancer cell* (2012) 22(4):547–60. doi:10.1016/j.ccr.2012.08.014
- Cokelaer T, Chen E, Iorio F, Menden MP, Lightfoot H, Saez-Rodriguez J, et al. GDSCTools for mining pharmacogenomic interactions in cancer. *Bioinformatics* (2018) 34(7):1226–8. doi:10.1093/bioinformatics/btx744
- Geeleher P, Cox N, Huang RS. pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* (2014) 9(9):e107468. doi:10.1371/journal.pone.0107468
- Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12(5):453–7. doi:10.1038/nmeth.3337
- Visco C, Li Y, Xu-Monette ZY, Miranda RN, Green TM, Li Y, et al. Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: A report from the international DLBCL rituximab-CHOP consortium program study. *Leukemia* (2012) 26(9):2103–13. doi:10.1038/leu.2012.83
- Ruppert AS, Dixon JG, Salles G, Wall A, Cunningham D, Poeschel V, et al. International prognostic indices in diffuse large B-cell lymphoma: A comparison of IPI, R-IPI, and NCCN-IPI. *Blood* (2020) 135(23):2041–8. doi:10.1182/blood.2019002729
- Nie Y, Liu L, Liu Q, Zhu X. Identification of a metabolic-related gene signature predicting the overall survival for patients with stomach adenocarcinoma. *PeerJ* (2021) 9:e10908. doi:10.7717/peerj.10908

28. Zhang ZY, Yao QZ, Liu HY, Guo QN, Qiu PJ, Chen JP, et al. Metabolic reprogramming-associated genes predict overall survival for rectal cancer. *J Cel Mol Med* (2020) 24(10):5842–9. doi:10.1111/jcmm.15254
29. Kollareddy M, Dimitrova E, Vallabhaneni KC, Chan A, Le T, Chauhan KM, et al. Regulation of nucleotide metabolism by mutant p53 contributes to its gain-of-function activities. *Nat Commun* (2015) 6:7389. doi:10.1038/ncomms8389
30. Cao Y, Chen P, Cai M, Shi Q, Xu P, Wang L, et al. Prognostic impact of B-vitamins involved in one-carbon metabolism in patients with diffuse large B-cell lymphoma. *Hematol Oncol* (2020) 38(4):456–66. doi:10.1002/hon.2752
31. Shen N, Korm S, Karantanos T, Li D, Zhang X, Ritou E, et al. DLST-dependence dictates metabolic heterogeneity in TCA-cycle usage among triple-negative breast cancer. *Commun Biol* (2021) 4(1):1289. doi:10.1038/s42003-021-02805-8
32. Vacanti NM, Divakaruni AS, Green CR, Parker SJ, Henry RR, Ciaraldi TP, et al. Regulation of substrate utilization by the mitochondrial pyruvate carrier. *Mol Cel* (2014) 56(3):425–35. doi:10.1016/j.molcel.2014.09.024
33. Yuan X, Dong B, Xu Y, Dong L, Huang J, Zhang J, et al. TIKI2 is upregulated and plays an oncogenic role in renal cell carcinoma. *Oncotarget* (2016) 7(13):17212–9. doi:10.18632/oncotarget.7873
34. Krejsgaard T, Litvinov IV, Wang Y, Xia L, Willerslev-Olsen A, Korolov SB, et al. Elucidating the role of interleukin-17F in cutaneous T-cell lymphoma. *Blood* (2013) 122(6):943–50. doi:10.1182/blood-2013-01-480889
35. Iqbal MA, Li M, Lin J, Zhang G, Chen M, Moazzam NF, et al. Preliminary study on the sequencing of whole genomic methylation and transcriptome-related genes in thyroid carcinoma. *Cancers (Basel)* (2022) 14(5):1163. doi:10.3390/cancers14051163
36. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* (2010) 467(7319):1114–7. doi:10.1038/nature09515
37. Dizeyi N, Bjartell A, Hedlund P, Taskén KA, Gadaleanu V, Abrahamsson PA. Expression of serotonin receptors 2B and 4 in human prostate cancer tissue and effects of their antagonists on prostate cancer cell lines. *Eur Urol* (2005) 47(6):895–900. doi:10.1016/j.eururo.2005.02.006
38. Nakamura Y, Ise K, Yamazaki Y, Fujishima F, McNamara KM, Sasano H. Serotonin receptor 4 (5-hydroxytryptamine receptor Type 4) regulates expression of estrogen receptor beta and cell migration in hormone-naive prostate cancer. *Indian J Pathol Microbiol* (2017) 60(1):33–7. doi:10.4103/0377-4929.200022
39. Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, et al. The landscape of viral associations in human cancers. *Nat Genet* (2020) 52(3):320–30. doi:10.1038/s41588-019-0558-9
40. Cheng AS, Li MS, Kang W, Cheng VY, Chou JL, Lau SS, et al. *Helicobacter pylori* causes epigenetic dysregulation of FOXD3 to promote gastric carcinogenesis. *Gastroenterology* (2013) 144(1):122–33. doi:10.1053/j.gastro.2012.10.002
41. Rosenbaum SR, Knecht M, Mollaei M, Zhong Z, Erkes DA, McCue PA, et al. FOXD3 regulates VISTA expression in melanoma. *Cel Rep* (2020) 30(2):510–24. doi:10.1016/j.celrep.2019.12.036
42. Kong L, Yang W, Chen L, Qian L. The DNA methylation-regulated MCTP1 activates the drug-resistance of esophageal cancer cells. *Aging (Albany NY)* (2021) 13(3):3342–52. doi:10.18632/aging.104173
43. Kovalevska L, Kashuba E. Expression pattern of MRPS18 family genes in malignantly transformed b-cells. *Exp Oncol* (2020) 42(4):295–9. doi:10.32471/exp-oncology.2312-8852.vol-42-no-4.15366
44. Ma C, Li H, Li X, Lu S, He J. The prognostic value of faciogenital dysplasias as biomarkers in head and neck squamous cell carcinoma. *Biomarkers Med* (2019) 13(16):1399–415. doi:10.2217/bmm-2019-0273
45. Walker JS, Hing ZA, Sher S, Cronin J, Williams K, Harrington B, et al. Rare t(X;14)(q28;q32) translocation reveals link between MTCPI1 and chronic lymphocytic leukemia. *Nat Commun* (2021) 12(1):6338. doi:10.1038/s41467-021-26400-x
46. Pfreundschuh M, Kuhnt E, Trumper L, Osterborg A, Trnny M, Shepherd L, et al. CHOP-Like chemotherapy with or without rituximab in young patients with good-prognosis diffuse large-B-cell lymphoma: 6-year results of an open-label randomised study of the MabThera international trial (MInT) group. *Lancet Oncol* (2011) 12(11):1013–22. doi:10.1016/S1470-2045(11)70235-2
47. Pfreundschuh M, Trumper L, Osterborg A, Pettengell R, Trnny M, Imrie K, et al. CHOP-Like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: A randomised controlled trial by the MabThera international trial (MInT) group. *Lancet Oncol* (2006) 7(5):379–91. doi:10.1016/S1470-2045(06)70664-7
48. Lang R, Gill MJ. Diffuse large B-cell lymphoma. *New Engl J Med* (2021) 384(23):2261–2. doi:10.1056/NEJMc2105452
49. Chen W, Liang W, He Y, Liu C, Chen H, Lv P, et al. Immune microenvironment-related gene mapping predicts immunochemotherapy response and prognosis in diffuse large B-cell lymphoma. *Med Oncol* (2022) 39(4):44. doi:10.1007/s12032-021-01642-3
50. Keane C, Gill D, Vari F, Cross D, Griffiths L, Gandhi M. CD4(+) tumor infiltrating lymphocytes are prognostic and independent of R-IP1 in patients with DLBCL receiving R-CHOP chemo-immunotherapy. *Am J Hematol* (2013) 88(4):273–6. doi:10.1002/ajh.23398
51. Emens LA. Chemoimmunotherapy. *Cancer J* (2010) 16(4):295–303. doi:10.1097/PPO.0b013e3181eb5066
52. Zhao Y, Niu C, Cui J. Gamma-delta ($\gamma\delta$) T cells: Friend or foe in cancer development? *J translational Med* (2018) 16(1):3. doi:10.1186/s12967-017-1378-2
53. Tokuyama H, Hagi T, Mattarollo SR, Morley J, Wang Q, So HF, et al. V gamma 9 V delta 2 T cell cytotoxicity against tumor cells is enhanced by monoclonal antibody drugs-rituximab and trastuzumab. *Int J Cancer* (2008) 122(11):2526–34. doi:10.1002/ijc.23365
54. Yan M, Chang YM, Raghavan V, Dong E, Klein C, Nielsen TG, et al. Lymphoma microenvironment deconvolution links M1 macrophage infiltration to clinical outcome in diffuse large B-cell lymphoma. *Blood* (2020) 136(1):29–30. doi:10.1182/blood-2020-134867
55. Autio M, Leivonen SK, Bruck O, Mustjoki S, Meszaros Jorgensen J, Karjalainen-Lindsberg ML, et al. Immune cell constitution in the tumor microenvironment predicts the outcome in diffuse large B-cell lymphoma. *Haematologica* (2021) 106(3):718–29. doi:10.3324/haematol.2019.243626
56. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *New Engl J Med* (2018) 378(15):1396–407. doi:10.1056/NEJMoa1801445
57. Wright GW, Huang DW, Phelan JD, Coulbaly ZA, Roulland S, Young RM, et al. A probabilistic classification tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications. *Cancer cell* (2020) 37(4):551–68. doi:10.1016/j.ccell.2020.03.015