

A novel prognostic signature based on immune-related genes of diffuse large B-cell lymphoma

Zizheng Wu^{1,*}, Qingpei Guan^{1,*}, Xue Han^{1,*}, Xianming Liu¹, Lanfang Li¹, Lihua Qiu¹, Zhengzi Qian¹, Shiyong Zhou¹, Xianhuo Wang¹, Huilai Zhang¹

¹Departments of Lymphoma, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center of Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Sino-US Center for Lymphoma and Leukemia Research, Tianjin 300060, China

*Equal contribution

Correspondence to: Zhengzi Qian, Xianhuo Wang, Huilai Zhang; email: gzz@163.com, <https://orcid.org/0000-0002-2924-3212>; tizlyy_xianhuow@163.com, <https://orcid.org/0000-0002-7663-2923>; zhlwgg@126.com, <https://orcid.org/0000-0002-8558-3131>

Keywords: diffuse large B-cell lymphoma, immune-related genes, overall survival, gene expression omnibus database, prognostic signature

Received: May 20, 2021

Accepted: September 18, 2021

Published: October 5, 2021

Copyright: © 2021 Wu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Diffuse large B-cell lymphoma (DLBCL) presents a great clinical challenge and has a poor prognosis, with immune-related genes playing a crucial role. We aimed to develop an immune-related prognostic signature for improving prognosis prediction in DLBCL.

Samples from the GSE31312 dataset were randomly allocated to discovery and internal validation cohorts. Univariate Cox, random forest, LASSO regression and multivariate Cox analyses were utilized to develop a prognostic signature, which was verified in the internal validation cohort, entire validation cohort and external validation cohort (GSE10846). The tumor microenvironment was investigated using the CIBERSORT and ESTIMATE tools. Gene set enrichment analysis (GSEA) was further applied to analyze the entire GSE31312 cohort. We identified four immune-related genes (CD48, IL1RL, PSDM3, RXFP3) significantly associated with overall survival. Based on discovery and validation cohort analyses, this four-gene signature could classify patients into high- and low-risk groups, with significantly different prognoses. Activated memory CD4 T cells and activated dendritic cells were significantly decreased in the high-risk group, and these patients had lower immune scores. GSEA revealed enrichment of signaling pathways, such as T cell receptor, antigen receptor-mediated, antigen processing and presentation of peptide antigen via MHC class I, in the low-risk group. In conclusion, a robust signature based on four immune-related genes was successfully constructed for predicting prognosis in DLBCL patients.

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL), the most frequent pathologic subtype of malignant lymphoma, poses challenges for classification and treatment [1]. Moreover, evidence from clinical and biological studies has indicated that DLBCL is an aggressive severe and complicated disease with a broad spectrum of genetic, phenotypic and clinical heterogeneities [2]. Despite a vast

improvement in the survival rate (50% ~ 60%), the heterogeneous nature of this disease elicits different survival outcomes for DLBCL patients undergoing routine treatment (rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone (R-CHOP)) [3]. In general, finding novel anti-DLBCL pharmaceutical targets, either alone or in combination with R-CHOP therapy, is crucial for survival enhancement or alternative measures for ineligible, relapsed or refractory cases [4].

The international prognostic index (IPI), which includes age, tumor stage, Eastern Cooperative Oncology Group (ECOG) performance, number of extranodal sites and lactate dehydrogenase (LDH) level, is widely employed for the clinical evaluation of DLBCL patient prognosis [5]. However, IPI does not consider the molecular heterogeneity of DLBCL, and the marked differences in patient survival, even among patients with similar or the same IPI [6]. Current DLBCL studies focus on recognizing novel risk stratification and prognostic biomarkers to predict survival outcomes and treatment response or to identify patients eligible for more aggressive therapies. At the same time, prognostic biomarkers may shed light on current and future potential therapies.

With advances in human gene sequencing technology, increasing attention has been given to gene-based biomarkers [7]. Furthermore, there is growing evidence that immune-related genes and the tumor immune microenvironment (TME) are crucial for malignant tumor progression and response to therapy [8, 9]. Therefore, an immune-related gene signature that enables physicians to estimate DLBCL prognosis and characterize the TME in these patients is urgently needed.

Here, we developed a reliable prognostic signature (PS) for DLBCL using immune-related genes and validated the clinical feasibility of this signature in DLBCL patients. Immune cell infiltration and the TME of patients with different risk scores are comprehensively described.

RESULTS

Prognostic genes recognition and PS construction

Overall, 426 DLBCL patients from the GSE31312 dataset were arbitrarily placed into a discovery cohort (DC, $n = 213$) or an internal validation cohort (IVC, $n = 213$). We then conducted univariate Cox regression analysis on the immune-related genes expression profiles in the DC and 26 candidate genes were significantly correlated with overall survival (OS) ($p < 0.001$) (Figure 1A). To retrieve the most significant genes with predictive abilities, random forest and LASSO regression analyses were performed synchronously to reduce the volume and select the most relevant genes (Figure 1B, 1C). Eleven and eighteen genes were chosen, respectively; ten genes intersected and were deemed to have the greatest predictive power for OS (Figure 1D). Multivariable Cox regression was

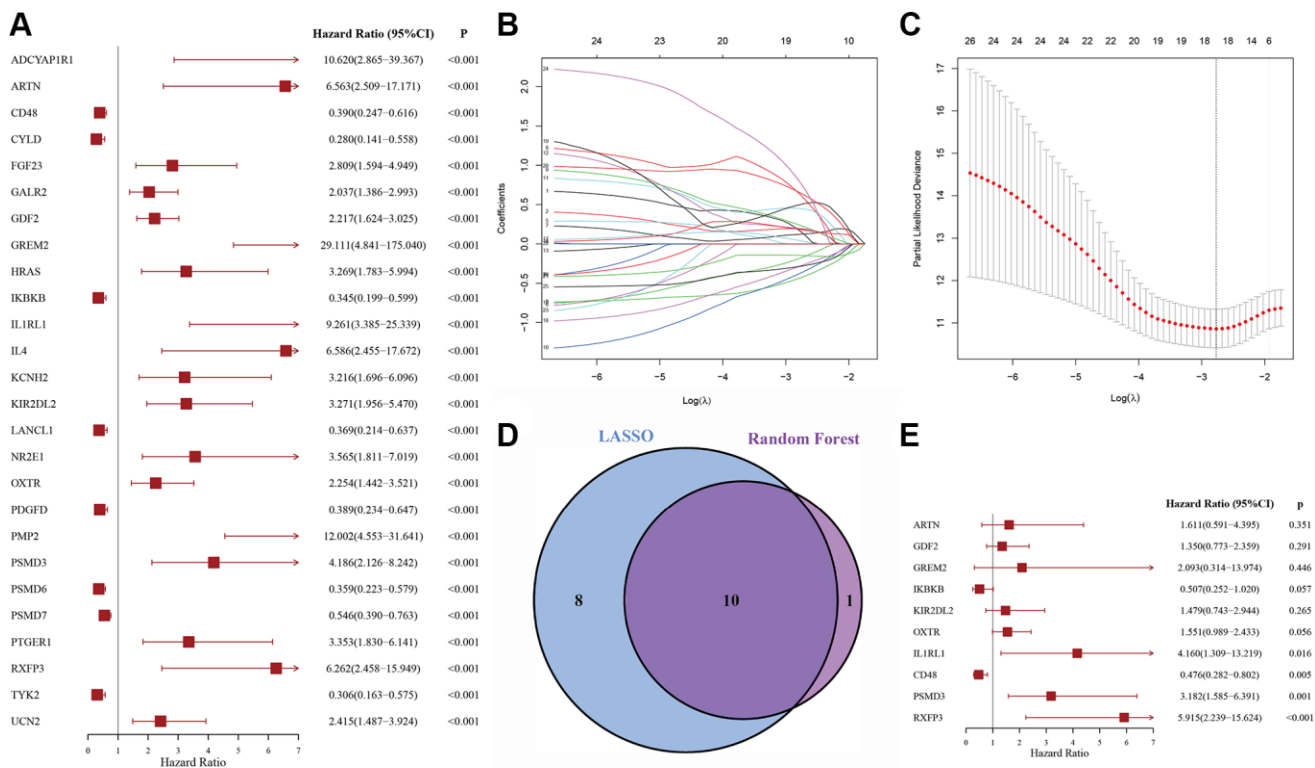


Figure 1. The process of variable selection. (A) Forest map of 26 candidate immune-related genes selected by univariate Cox regression analysis associated with DLBCL overall survival in the discovery cohort. (B, C) The performance of least absolute shrinkage and selection operator (LASSO) analysis. (D) LASSO and random forest analysis intersecting genes were selected. (E) Forest map of multivariate Cox regression analysis to establish a prognostic signature.

then carried out (Figure 1E), which revealed four risk genes: *CD48* (HR = 0.476, 95% CI: 0.282–0.802, $p = 0.005$), *IL1RL1* (HR = 4.160, 95% CI: 1.309–13.219, $p = 0.016$), *PSMD3* (HR = 3.182, 95% CI: 1.585–6.391, $p = 0.001$), and *RXFP3* (HR = 5.915, 95% CI: 2.239–15.624, $p < 0.001$).

Furthermore, a risk score (RS) system was established, according to the levels of these four genes and the corresponding coefficient obtained from multivariable Cox regression analysis. The formula was as follows: $RS = (-0.794 \times CD48 \text{ levels}) + (2.243 \times IL1RL1 \text{ levels}) + (1.440 \times PSMD3 \text{ levels}) + (1.348 \times RXFP3 \text{ levels})$. We then computed the RS for each patient and set the median as the cutoff to divide them into high-risk (HR) and low-risk (LR) groups.

As shown in Figure 2A, Kaplan-Meier survival analysis of the DC showed that HR patients exhibited significantly worse prognosis, compared to LR patients (log-rank $p < 0.001$). The OS rates at 3 and 5 years for

the HR patients were 69.33% and 54.54%, respectively, while the corresponding rates for the LR patients were 85.44% and 83.31%, respectively. The PS area under the curve (AUC) analysis demonstrated its excellent accuracy in estimating DLBCL patient OS (1-year AUC = 0.763, 3-year AUC = 0.767, 5-year AUC = 0.706) (Figure 2B).

We next plotted risk curves and survival status scatter plots to illustrate the RS and OS of each DLBCL patient in the DC, and a worse prognosis was significantly associated with a higher RS (Figure 2C, 2D). Moreover, using a heatmap, we demonstrated the HR and LR gene expression profiles of both groups (Figure 2E). We revealed that *CD48* was significantly upregulated in LR patients and that *IL1RL1*, *PSMD3*, and *RXFP3* were strongly elevated in HR patients.

Moreover, the PS correlated significantly with OS in our univariate analysis (HR = 4.113, 95% CI: 2.393–7.069, $p < 0.001$) (Figure 3A). After multivariable

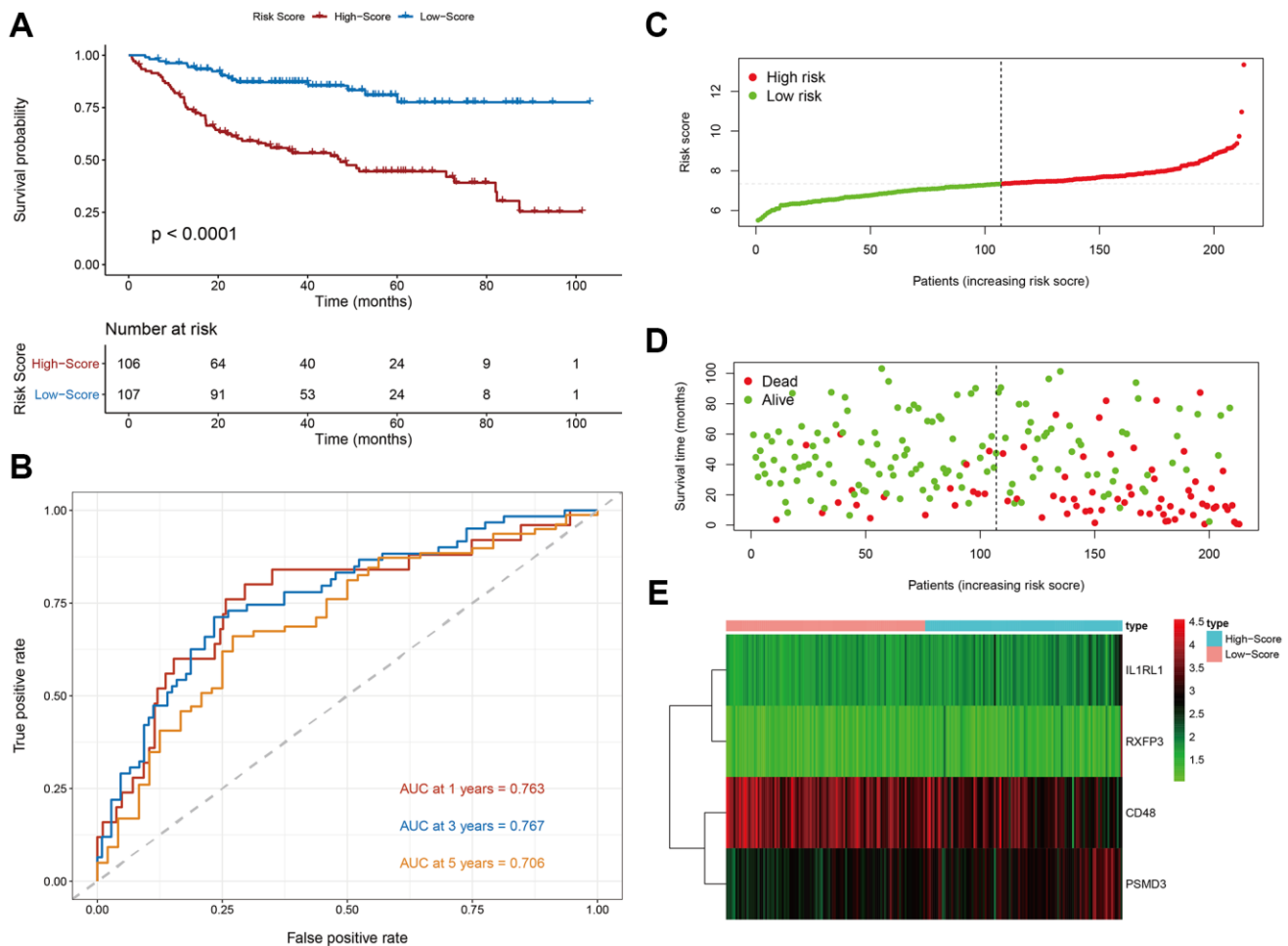


Figure 2. Evaluation of the prognostic signature in the discovery cohort. (A) Kaplan-Meier plots of overall survival between high- and low-risk patients. (B) Time-dependent receiver operating characteristic (ROC) curve analysis. (C) The risk score distribution. (D) The survival status distribution. (E) Expression heatmap of the four immune-related risk genes.

adjustment by other clinical factors, including stage, ECOG, LDH, number of sites of extranodal disease and subtype, the PS remained a significant and independent prognostic indicator in the DC (HR = 3.458, 95% CI: 1.992–6.005, $p < 0.001$) (Figure 3B).

Verification of the PS performance

To further assess the robustness of the PS, we performed similar analyses in the IVC, entire GSE31312 cohort (EGC) and external validation cohort

(EVC) from the GSE10846 dataset, and all cohorts yielded similar results. Kaplan-Meier survival analysis revealed markedly worse OS in HR patients (Figure 4A–4C). In univariable Cox regression analysis, an elevated RS was an OS risk factor in all validation cohorts (IVC: HR = 2.480, 95% CI: 1.532–4.014, $p < 0.001$ (Figure 4D); EGC: HR = 2.998, 95% CI: 2.104–4.272, $p < 0.001$ (Figure 4E); EVC: HR = 1.931, 95% CI: 1.310–2.846, $p = 0.001$ (Figure 4F)). A similar result was obtained in multivariate Cox regression analysis, in which the PS was analyzed in combination

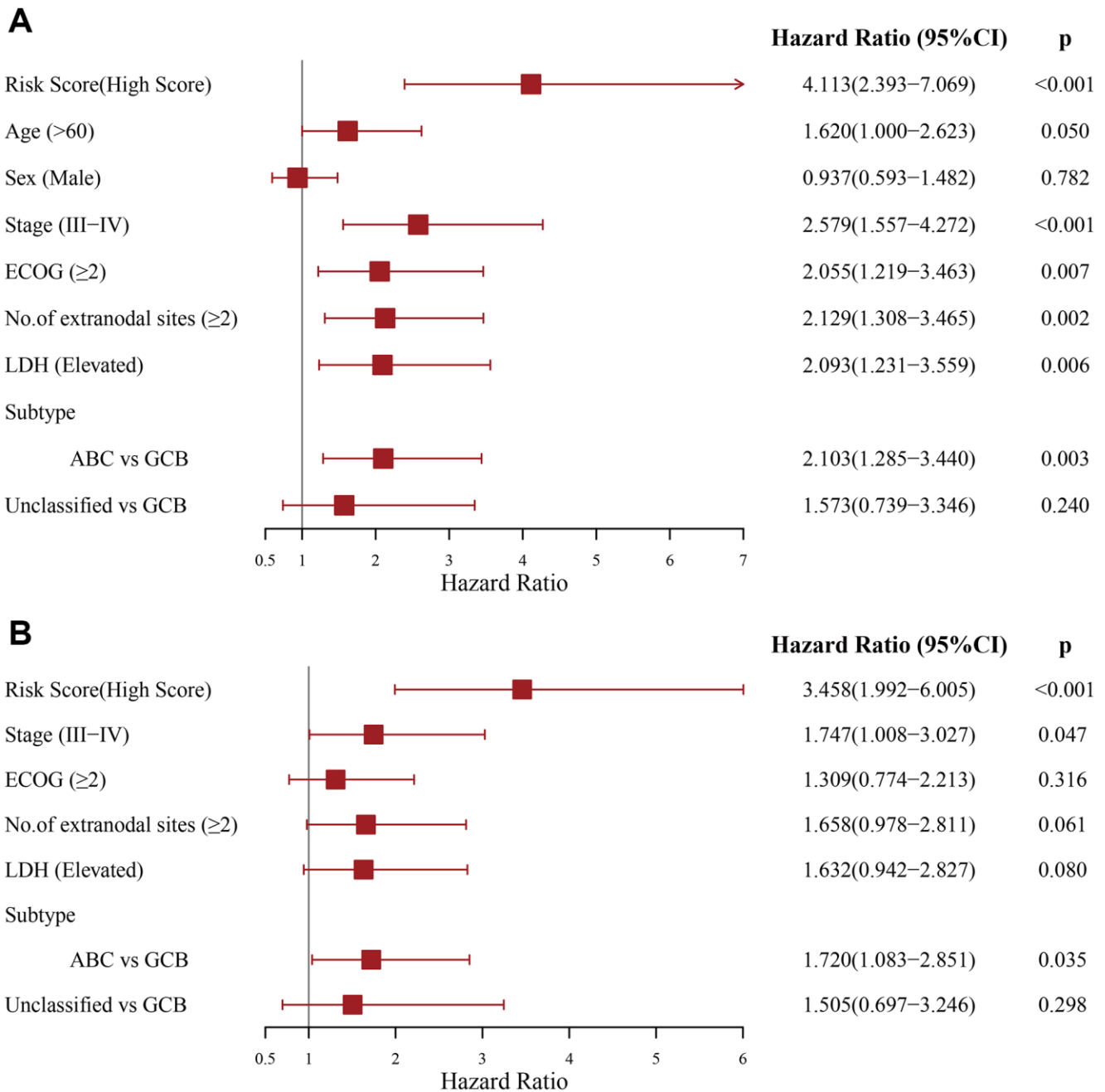


Figure 3. Evaluation of the independent prognostic value of the prognostic signature in the discovery cohort. (A) Univariate and (B) multivariate Cox regression analyses of the signature and clinical factors.

with other clinical factors (Figure 4G–4I). Furthermore, stratification analyses in the EGC indicated worse OS for HR patients in each stratum, including age, sex, stage, ECOG, LDH, number of sites of extranodal disease and subtype, than LR patients, except in the subgroup of those with the unclassified subtype (Supplementary Figure 1). Based on these data, our established PS is a robust and independent predictor of OS in different populations. In addition, HR patients tend to have shorter progression-free survival (PFS) (Supplementary Figure 2). HR patients achieved a remarkably low overall response rate (ORR) and complete remission (CR) rate (ORR, 86.38% vs. 94.84%, $p = 0.003$; CR, 66.20% vs. 84.04%, $p < 0.001$) (Table 1).

Evaluation of the relationship between the PS and clinical factors

The relationship between the PS and different clinical factors was assessed using the EGC. Compared with patients aged ≤ 60 , the RS was increased in patients aged > 60 ($p = 0.002$, Figure 5A). A similar phenomenon was observed for stage ($p = 0.014$, Figure 5B), number of sites of extranodal disease ($p = 0.001$, Figure 5C) and LDH ($p = 0.004$, Figure 5D). However, we did not observe differences in RS regarding

ECOG classification ($p = 0.12$, Figure 5E) or subtype classification ($p = 0.229$, Figure 5F).

Association between the PS and TME

The proportions of follicular helper T cells ($p = 0.044$), activated NK cells ($p = 0.001$), monocytes ($p = 0.002$), M2 macrophages ($p = 0.030$) and activated mast cells ($p = 0.001$) were markedly enhanced in HR patients. Conversely, the numbers of activated memory CD4 T cells ($p < 0.001$), gamma delta T cells ($p < 0.001$), stimulated dendritic cells ($p = 0.007$) and resting mast cells ($p = 0.031$) were significantly decreased in this group (Figure 6A). It was demonstrated that an elevated RS was strongly correlated with tumor purity ($p < 0.0001$) through the ESTIMATE algorithm. However, an elevated RS displayed an inverse correlation with the stromal ($p < 0.05$), immune ($p < 0.0001$) and ESTIMATE scores ($p < 0.0001$) (Figure 6B–6E). Hence, the PS may be reflective of the TME status in DLBCL patients.

Gene set enrichment analysis (GSEA) for functional annotation of the PS

According to our results, immune-related biological networks were enriched in LR patients, compared to HR

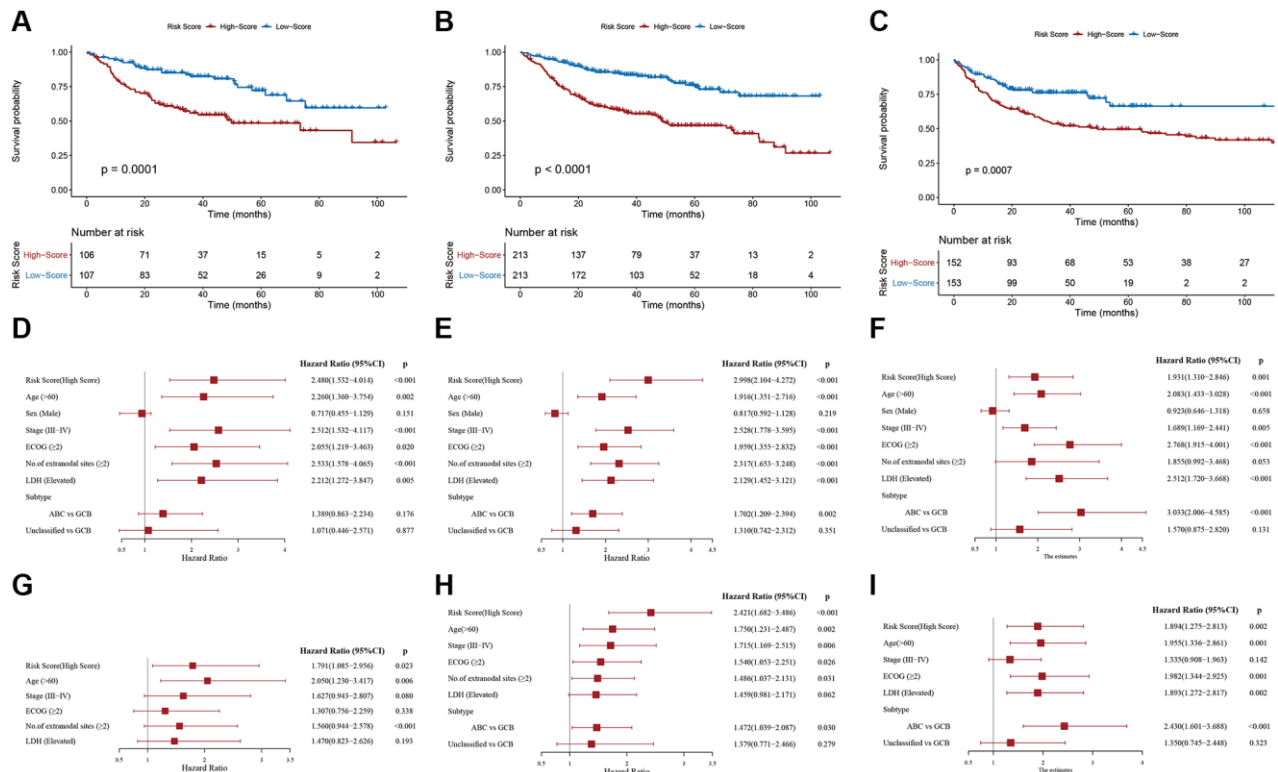


Figure 4. Validation of the prognostic signature. Kaplan-Meier plots of overall survival between high- and low-risk patients in the (A) internal validation cohort, (B) entire GSE31312 cohort, and (C) external validation cohort. Univariate Cox regression analyses in the (D) internal validation cohort, (E) entire GSE31312 cohort, and (F) external validation cohort. Multivariate Cox regression analysis in the (G) internal validation cohort, (H) entire GSE31312 cohort, and (I) external validation cohort.

Table 1. Treatment responses of patients in the entire GSE31312 cohort.

	Patients in the entire GSE31312 cohort (n = 426) nodal-DLBCL		p
	High-risk group n (%)	Low-risk group n (%)	
Total number	213 (50)	213 (50)	
ORR	184 (86.38)	202 (94.84)	0.003*
CR	141 (66.20)	179 (84.04)	<0.001**
PR	43 (20.18)	23 (10.80)	0.007*
PD/SD	15/14 (13.62)	6/5 (5.16)	0.003*

Abbreviations: ORR: overall response rate; CR: complete response; PR: partial response; SD: stable disease; PD: progressive disease. * $p < 0.05$; ** $p < 0.001$.

patients. We identified three immune-related GO terms in the GSEA results (Figure 7), including T cell receptor, antigen receptor mediated, antigen processing and presentation of peptide antigen via the MHC class I signaling pathway.

DISCUSSION

Most traditional biomarkers used thus far have weak prognostic power and cannot reflect the status of tumor immune infiltration in DLBCL. Immune-related genes and cells participate throughout the process of malignant tumor initiation, proliferation, and progression [10]. Previous reports indicated that immune-associated long noncoding RNA [11], immune cell constitution [12] and B7-CD28 gene family expression [13] can estimate DLBCL patient prognosis.

Here, we developed a novel PS, based on immune-related genes to estimate DLBCL patient prognosis.

Our signature contains four immune-related genes with prognostic power. Among them, CD48 is a protective factor with $HR < 1$; the three other genes (IL1RL1, PSMD3, RXFP3) are risk factors with $HR > 1$. CD48 is a member of the signaling lymphocyte activation molecule family that contributes to the activation and proliferation of T cells, antigen presenting cells and granulocytes by binding to CD2 [14]. Wang et al. reported that high CD48 expression activates NK cell function and reverses acute myeloid leukemia immune escape [15]. IL1RL1, an IL-1-type receptor, is detected in a subcategory of T cells and mature myeloid cells [16]. The cytokine IL-33 is the only reported ligand for IL1RL1. The IL-33/IL1RL1 network was shown to

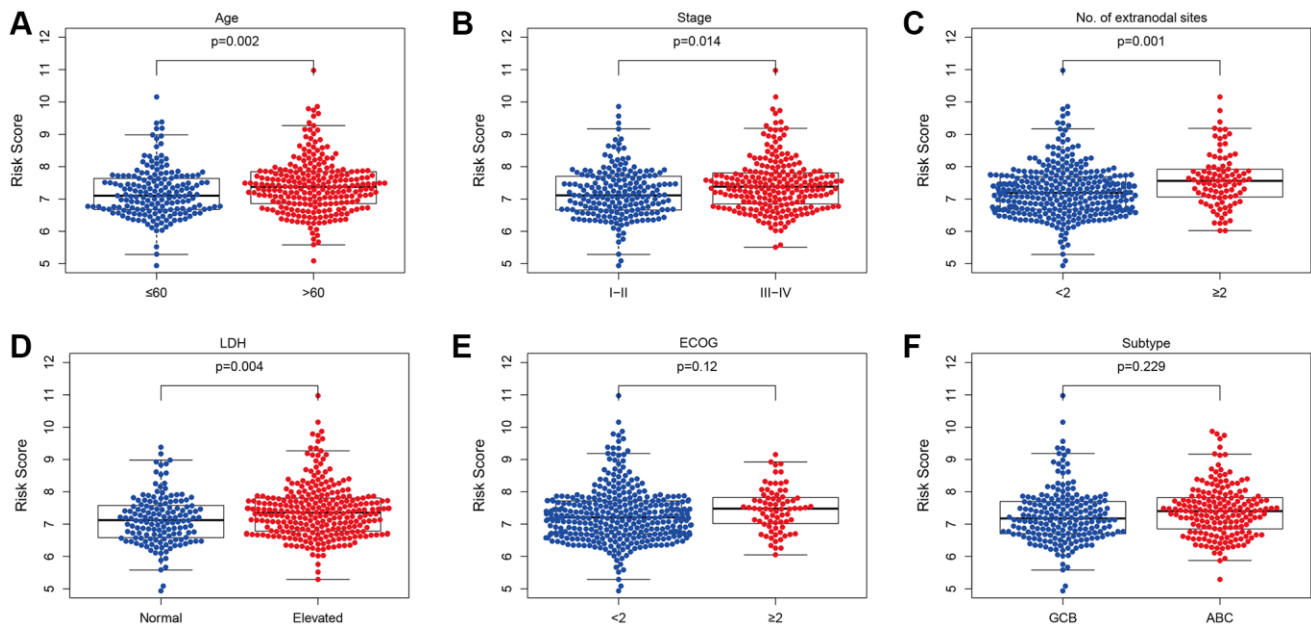


Figure 5. Relationship between the prognostic signature and clinical factors. (A) Age. (B) Stage. (C) Number of extranodal sites. (D) LDH. (E) Subtype. (F) ECOG (* $p < 0.05$, ** $p < 0.0001$).**

contribute to multiple types of blood malignancies [17]. The IL1RL1/IL-33 axis can remodel the tumor stroma or microenvironment to promote malignancy by recruiting a cohort of immune cells [18]. PSMD3 is a member of the 19 S regulatory complex in the 26 S proteasome, participating in cell cycle progression, apoptosis, and DNA damage repair [19]. In acute

myeloid leukemia, patients with high level of PSMD3 mRNA have a poor prognosis [20]. PSMD3 promoted NF- κ B protein expression and was upregulated in TKI-resistant chronic myeloid leukemia (CML) cells. The level of PSMD3 mRNA was higher in patients with blast phase than in patients with the chronic phase of the disease [21]. RXFP3 belongs to the insulin superfamily

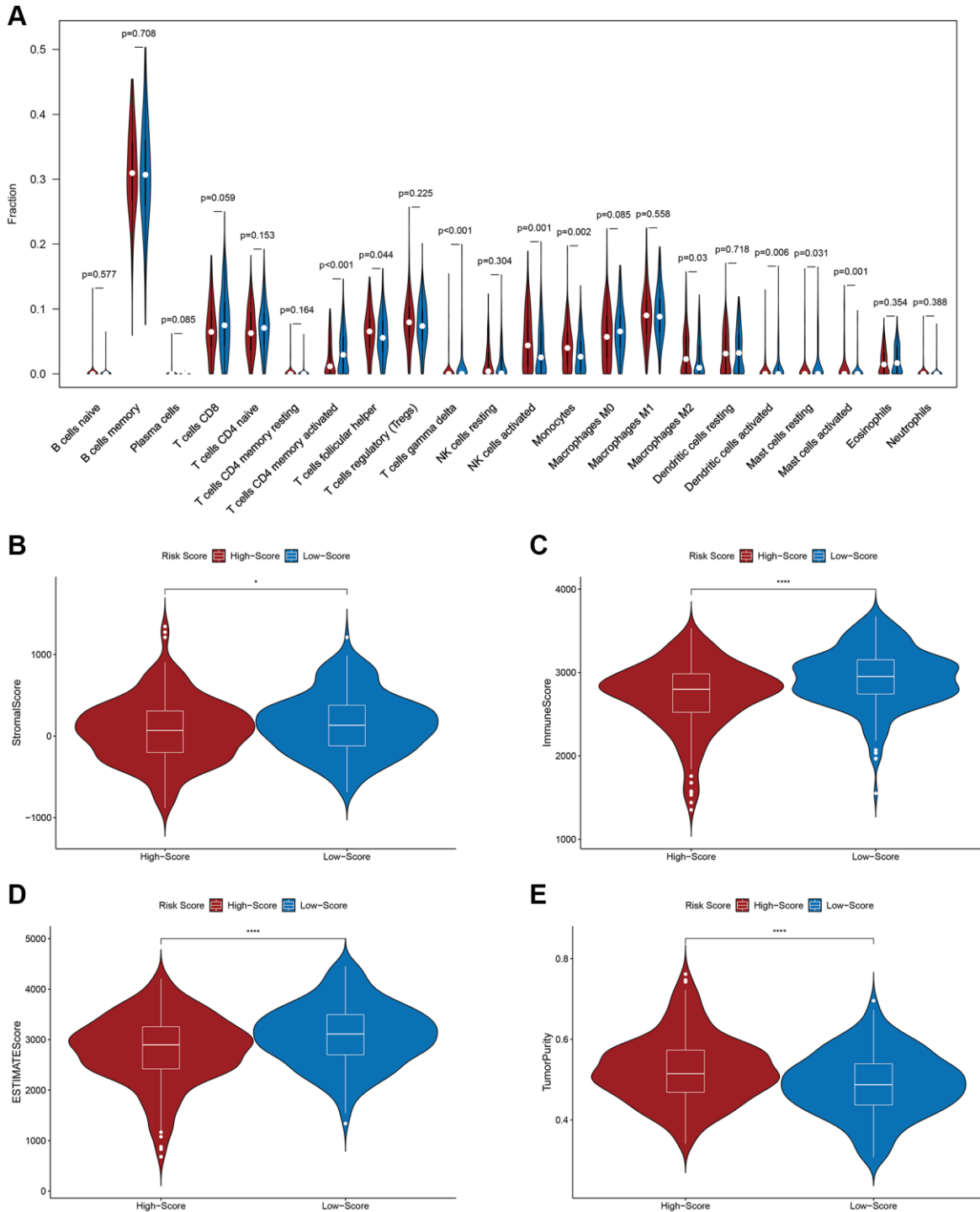


Figure 6. Relationship between the prognostic signature and the tumor immune microenvironment. (A) Relationship between the prognostic signature and immune cell infiltration. **(B)** Stromal score. **(C)** Immune score. **(D)** ESTIMATE score. **(E)** Tumor purity.

[22]. When combined with the ligand, RXFP3 activates downstream kinase pathways via multiple networks, such as, protein kinase C. The *RXFP3* methylation status has a strong correlation with microsatellite instability in endometrial cancer [23]. Overall, the role of these four genes in DLBCL development and immune regulation deserves further investigation.

The signature developed in this study demonstrated excellent predictive performance and effectively classified DLBCL patients into HR and LR categories. In the DC, the HR patients exhibited markedly worse OS, than the LR patients. Furthermore, the immune-related signature is an independent stand-alone prognostic factor, based on multivariate analysis. The results from the validation cohorts agreed with the above results. In addition, the signature showed a strong correlation with clinicopathologic factors. Therefore, this PS may serve as a reliable tool in guiding clinical work.

Immune cells identify tumor cells and destroy them via immune surveillance [24, 25], and immune cell infiltration is a major determinant of DLBCL prognosis [26, 27]. In this study, the proportions of activated memory CD4 T and dendritic cells were lower, but those of M2 macrophages, monocytes and NK cells were significantly higher in HR patients. In general, CD4⁺ T, dendritic, and NK cells are crucial factors in antitumor immunity and have critical significance for cancer immunotherapy [28–30]. Previous studies have reported that high levels of CD4⁺ T cells are associated with improved survival outcomes in many malignancies [31], and mouse models of B-cell lymphoma suggest that CD4⁺ T cells are key to the establishment of an antitumor microenvironment [32]. Dendritic cells have a strong ability to present antigens, and improved antigen presentation has been shown to be a key

determinant of survival in patients with DLBCL [33, 34]. Indeed, Ciavarella et al. concluded that DLBCL patients with elevated amounts of dendritic and CD4⁺ T cells experienced prolonged OS. However, patients with a high number of activated NK cells experienced a worse prognosis [35], which is consistent with the results of this study. M2 macrophages are immunosuppressive cells and are thought to be involved in tumor immune evasion [36]. M2 macrophages are a crucial factor for poor survival outcomes and a stand-alone indicator of short OS and PFS [37]. Larger amounts of immune and stromal cells equated to lower quantities of tumor cells [38]. Here, we found that an elevated RS correlated positively with tumor purity but negatively with immune, stromal and ESTIMATE scores. Hence, patients with elevated RSs have more tumor cells and fewer stromal cells. Finally, GSEA showed enrichment of immune-related biological processes in LR patients. Based on this evidence, the TME of patients with low RSs tends to display active immune status and enhanced immune defense. In contrast, that of patients with high-RSs tends to be suppressed. This may explain why the prognosis of HR patients was quite poor.

The advantage of this study is that both internal and external validation were used to evaluate the PS. However, there were several limitations. First, as the data analyzed were downloaded from an online database, our study was retrospective. Second, there are no experimental data to confirm our findings. Thus, more functional studies on these four genes, alone and in combination, are needed to verify the predictive power of the PS and to explore possible immune-related pathways in depth. Our work was the first to report an immune-related gene PS that predicts OS in DLBCL patients.

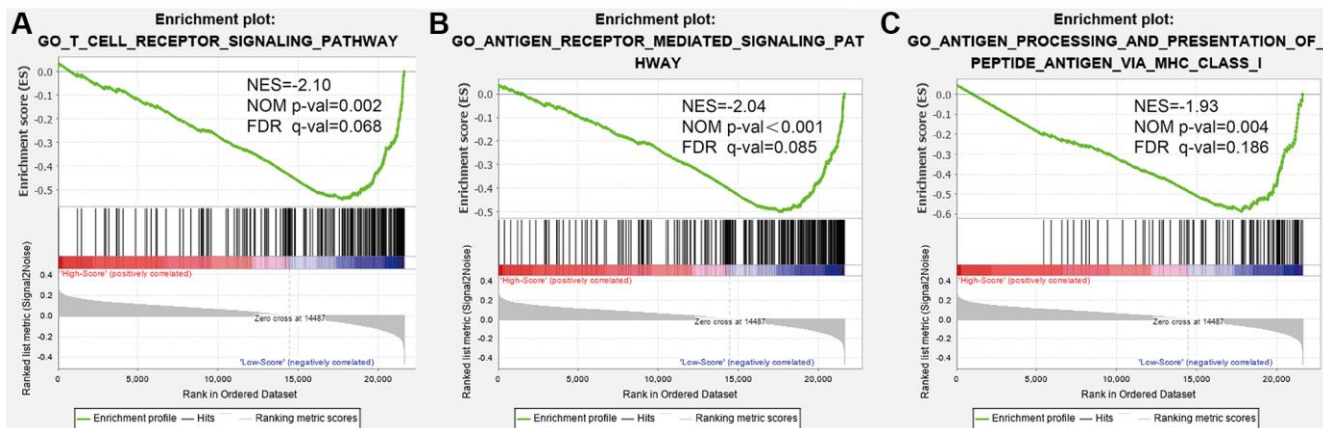


Figure 7. Gene set enrichment analysis (GSEA) for functional annotation of the prognostic signature. (A) The T cell receptor signaling pathway. **(B)** The antigen receptor-mediated signaling pathway. **(C)** Antigen processing and presentation of peptide antigen via MHC class I.

In conclusion, the signature developed in this study can both predict DLBCL patient survival outcomes and reveal the immunologic status of DLBCL. The PS may be clinically employed to improve patient OS and individualized therapy methods based on RSs. However, both experimental and clinical research efforts are warranted to confirm the findings of this research.

MATERIALS AND METHODS

Data source and preprocessing

Transcriptome data (.CEL files) for DLBCL patients were extracted from the GSE31312 and GSE10846 datasets in the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). The raw data were uniformly normalized with the robust multichip average (RMA) technique [39] using the “affy” and “affyPLM” packages, and gene expression profiles were performed on the GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array) platform. For genes with several probes, gene expression values were generated with the median of multiple probes.

After excluding patients without complete clinical information, 426 DLBCL patients (GSE31312 cohort) and 305 DLBCL patients (GSE10846 cohort) were examined in this study; the detailed clinical data are shown in Table 2. Immune-related genes were retrieved from the Immunology Database and Analysis Portal (ImmPort) database (<https://www.immport.org>) [40].

Generation of an immune-related gene PS

In total, 426 samples from the GSE31312 dataset were randomly allocated at a 1:1 ratio to discovery and IVCs using R software. A description and comparison of the baseline features of patients from the DC and IVC was conducted (Supplementary Table 1). Variables were analyzed via the chi-square test, and $p < 0.05$ was the significance threshold.

Univariate analysis was performed to identify immune-related genes strongly related to OS in DC ($p < 0.001$). Next, random forest analysis and least absolute shrinkage and selection operator (LASSO) analysis were conducted simultaneously. The overlapping genes were subjected to multivariate Cox analysis to fulfill variable selection, with $p < 0.05$ as the criterion. We then calculated the scorers, according to a linear combination of the gene levels and regression coefficient of the multivariate Cox analysis: $RS = \exp_{mRNA1} \times \beta_{mRNA1} + \exp_{mRNA2} \times \beta_{mRNA2} \dots + \exp_{mRNAi} \times \beta_{mRNAi}$, where \exp_{mRNAi} is the expression value of each gene, and β_{mRNAi} is the regression coefficient of the multivariate analysis for the candidate gene. The

patients were then separated into HR and LR groups, based on the median RS. The Kaplan-Meier log-rank test and time-dependent receiver operating characteristic (ROC) curve analysis were applied to validate PS performance. An area under the ROC curve (AUC) > 0.60 was considered an acceptable prediction value; AUC > 0.75 was regarded as excellent for prediction. RS distribution plots, survival status scatter plots, and expression heatmaps of the four immune-related risk genes between the HR and LR groups were generated to assess PS, and stand-alone prognostic analysis was carried out to evaluate whether this signature was indeed an independent stand-alone predictor of OS.

Application and verification of the immune-related gene PS

RSs for each patient in the IVC, the EGC and the EVC from the GSE10846 dataset were computed and assigned to two groups, according to the median. Kaplan-Meier log-rank tests and univariate and multivariate analyses were conducted to compare OS between the HR and LR groups. Kaplan-Meier log-rank tests were performed to compare PFS between the HR and LR groups in the DC, IVC and EGC. The treatment response in the EGC was analyzed by Pearson’s chi-square test.

Evaluation of the relationship between the PS and clinical factors

To assess the predictive value of PS in DLBCL, we examined the correlation between PS and clinical factors in the EGC. Intergroup differences were analyzed by independent Student’s *t*-tests. Two-tailed $p < 0.05$ was set as the significance threshold.

Association between the PS and TME

The proportion of 22 immune cell subtypes based on expression profile data in the EGC was assessed using the CIBERSORT package [41], with permutations set at 1000. Cases with $p < 0.05$ according to the CIBERSORT results were included in further analyses. The Wilcoxon test was employed to compare differences in immune cell subtypes between the HR and LR groups. To further evaluate the association between this signature and the TME, the stromal, immune, and tumor purity scores were computed using the ESTIMATE algorithm [42]. Significance was considered at $p < 0.05$.

GSEA

GSEA [43] was performed to examine different biological processes between the HR and LR patients.

Table 2. Clinical and pathological characteristics of patients with DLBCL in this study.

Variables	GSE31312 N (%)	GSE10846 N (%)
Age, years		
≤60	179 (42.02)	146 (47.87)
>60	247 (57.98)	159 (52.13)
Sex		
Female	183 (42.96)	134 (43.93)
Male	243 (57.04)	171 (56.07)
Stage		
I–II	200 (46.95)	144 (47.21)
III–IV	226 (53.05)	161 (52.79)
No. of extranodal sites		
<2	331 (77.70)	282 (92.46)
≥2	95 (22.30)	23 (7.54)
ECOG		
<2	350 (82.16)	230 (75.41)
≥2	76 (17.84)	75 (24.59)
LDH		
Normal	148 (34.74)	153 (50.16)
Elevated	278 (65.26)	152 (49.87)
Subtype		
GCB	203 (47.65)	133 (43.61)
ABC	183 (42.96)	125 (40.98)
Unclassified	40 (9.39)	47 (15.41)

Abbreviations: ECOG: Eastern Cooperative Oncology Group; LDH: lactate dehydrogenase; GCB: germinal center B-cell-like lymphoma; ABC: activated B-cell-like lymphoma.

The gene expression profiles of the EGC were evaluated with regard to Gene Ontology (GO) gene sets. The number of random sample permutations was set at 1000, and enriched gene sets with a nominal $p < 0.05$ and FDR < 0.25 were regarded as significant.

Statistical analysis

In this study, all statistical analyses were carried out with R version 3.6.3 (<https://www.r-project.org/>) and the corresponding packages.

Abbreviations

DLBCL: diffuse large B-cell lymphoma; ECOG: Eastern Cooperative Oncology Group; LDH: lactate dehydrogenase; GCB: germinal center B-cell-like lymphoma; ABC: activated B-cell-like; UC: unclassified type; OS: overall survival; CI: confidence interval; HR: hazard ratio; NES: normalized enrichment score; FDR: false discovery rate; DC: discovery cohort; IVC: internal validation cohort; EGC: entire GSE31312 cohort; PS: prognostic signature; RS: risk score.

AUTHOR CONTRIBUTIONS

HLZ and XHW conceived and designed the study and reviewed the manuscript. ZZW, QPG, and XH collected, arranged, and analyzed the data and wrote the manuscript. XM L, LFL, LHQ, and SYZ revised the statistical methodology. ZZQ designed and prepared the figures and tables. All authors reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

The authors thank GEO for sharing the DLBCL data.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

This work was supported by the National Natural Science Foundation of China (NO. 81770213), Natural

Science Foundation of Tianjin (NO.19JCYBJC26500 and 18JCZDJC45100), Clinical Oncology Research Fund of CSCO (NO. Y-XD2019-162, Y-Roche20192-0097), National Key New Drug Creation Special Programs (NO. 2018ZX09201015) and National Human Genetic Resources Sharing Service Platform/Cancer Biobank of Tianjin Medical University Cancer Institute and Hospital (NO. 2005DKA21300).

REFERENCES

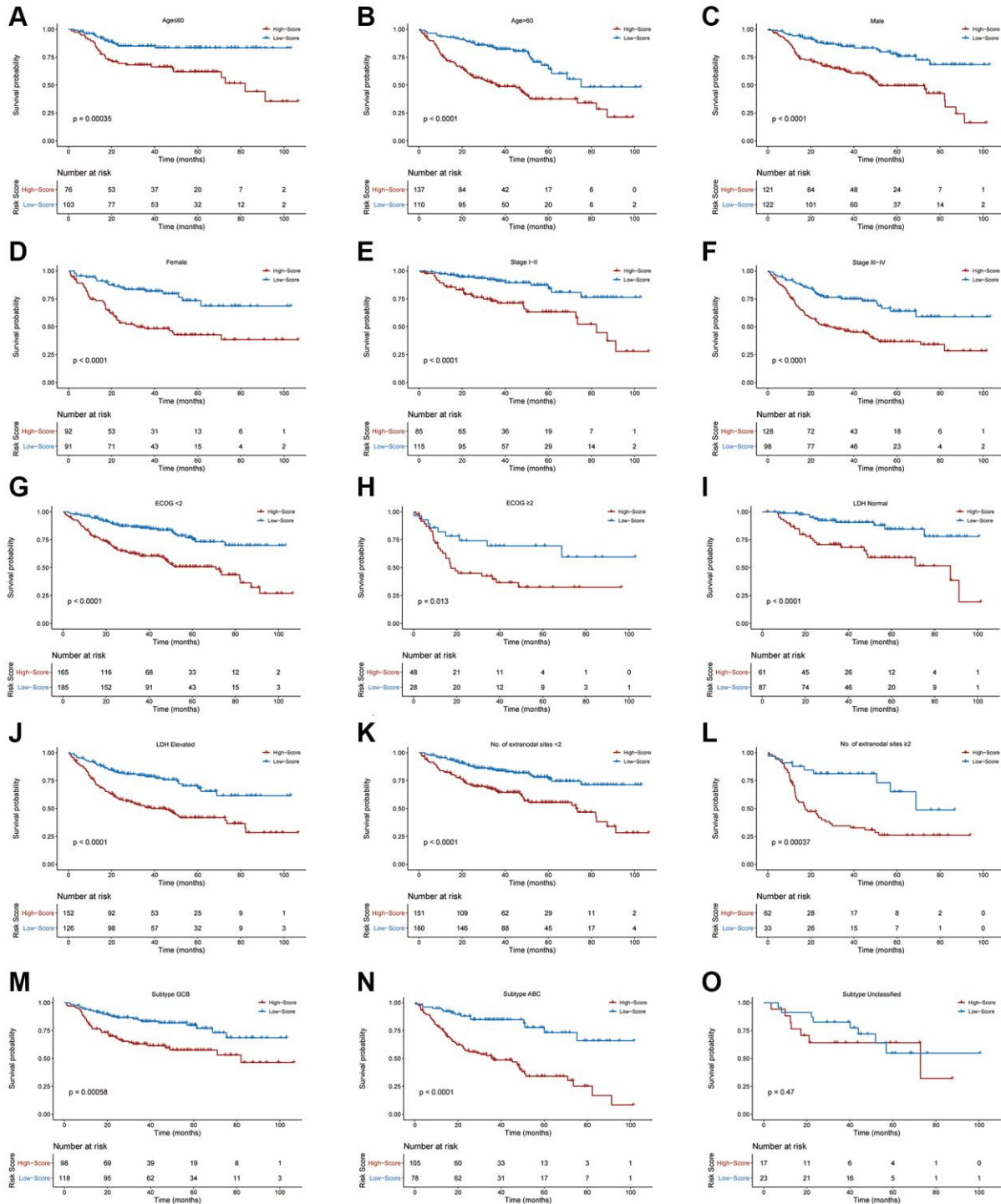
1. Pasqualucci L, Dalla-Favera R. Genetics of diffuse large B-cell lymphoma. *Blood*. 2018; 131:2307–19. <https://doi.org/10.1182/blood-2017-11-764332> PMID:29666115
2. Abramson JS, Shipp MA. Advances in the biology and therapy of diffuse large B-cell lymphoma: moving toward a molecularly targeted approach. *Blood*. 2005; 106:1164–74. <https://doi.org/10.1182/blood-2005-02-0687> PMID:15855278
3. Younes A. Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin: Seeing the Forest and the Trees. *J Clin Oncol*. 2015; 33:2835–36. <https://doi.org/10.1200/JCO.2015.61.9288> PMID:26261249
4. Sehn LH, Gascoyne RD. Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood*. 2015; 125:22–32. <https://doi.org/10.1182/blood-2014-05-577189> PMID:25499448
5. International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med*. 1993; 329:987–94. <https://doi.org/10.1056/NEJM199309303291402> PMID:8141877
6. Shipp MA, Abeloff MD, Antman KH, Carroll G, Hagenbeek A, Loeffler M, Montserrat E, Radford JA, Salles G, Schmitz N, Symann M, Armitage JO, Coiffier B, Philip T. International Consensus Conference on high-dose therapy with hematopoietic stem-cell transplantation in aggressive non-Hodgkin's lymphomas: report of the jury. *Ann Oncol*. 1999; 10:13–19. <https://doi.org/10.1023/a:1008397220178> PMID:10076716
7. Angell H, Galon J. From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. *Curr Opin Immunol*. 2013; 25:261–67. <https://doi.org/10.1016/j.coi.2013.03.004> PMID:23579076
8. Li Y, Jiang T, Zhou W, Li J, Li X, Wang Q, Jin X, Yin J, Chen L, Zhang Y, Xu J, Li X. Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun*. 2020; 11:1000. <https://doi.org/10.1038/s41467-020-14802-2> PMID:32081859
9. Zhu R, Tao H, Lin W, Tang L, Hu Y. Identification of an Immune-Related Gene Signature Based on Immunogenomic Landscape Analysis to Predict the Prognosis of Adult Acute Myeloid Leukemia Patients. *Front Oncol*. 2020; 10:574939. <https://doi.org/10.3389/fonc.2020.574939> PMID:33330048
10. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat Rev Cancer*. 2020; 20:662–80. <https://doi.org/10.1038/s41568-020-0285-7> PMID:32753728
11. Zhou M, Zhao H, Xu W, Bao S, Cheng L, Sun J. Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Mol Cancer*. 2017; 16:16. <https://doi.org/10.1186/s12943-017-0580-4> PMID:28103885
12. Autio M, Leivonen SK, Brück O, Mustjoki S, Mészáros Jørgensen J, Karjalainen-Lindsberg ML, Beiske K, Holte H, Pellinen T, Leppä S. Immune cell constitution in the tumor microenvironment predicts the outcome in diffuse large B-cell lymphoma. *Haematologica*. 2021; 106:718–29. <https://doi.org/10.3324/haematol.2019.243626> PMID:32079690
13. Wang G, Fu X, Chang Y, Li X, Wu X, Li L, Zhang L, Sun Z, Zhang X, Zhang M. B7-CD28 gene family expression is associated with prognostic and immunological characteristics of diffuse large B-cell lymphoma. *Aging (Albany NY)*. 2019; 11:3939–57. <https://doi.org/10.18632/aging.102025> PMID:31195368
14. McArdel SL, Terhorst C, Sharpe AH. Roles of CD48 in regulating immunity and tolerance. *Clin Immunol*. 2016; 164:10–20. <https://doi.org/10.1016/j.clim.2016.01.008> PMID:26794910
15. Wang Z, Xiao Y, Guan W, Wang M, Chen J, Zhang L, Li Y, Xiong Q, Wang H, Wang M, Li Y, Lv N, Li Y, et al. Acute myeloid leukemia immune escape by epigenetic CD48 silencing. *Clin Sci (Lond)*. 2020; 134:261–71. <https://doi.org/10.1042/CS20191170> PMID:31922199

16. Griesenauer B, Paczesny S. The ST2/IL-33 Axis in Immune Cells during Inflammatory Diseases. *Front Immunol.* 2017; 8:475.
<https://doi.org/10.3389/fimmu.2017.00475>
PMID:[28484466](https://pubmed.ncbi.nlm.nih.gov/28484466/)
17. Wang Y, Luo H, Wei M, Becker M, Hyde RK, Gong Q. IL-33/IL1RL1 axis regulates cell survival through the p38 MAPK pathway in acute myeloid leukemia. *Leuk Res.* 2020; 96:106409.
<https://doi.org/10.1016/j.leukres.2020.106409>
PMID:[32652328](https://pubmed.ncbi.nlm.nih.gov/32652328/)
18. Larsen KM, Minaya MK, Vaish V, Peña MMO. The Role of IL-33/ST2 Pathway in Tumorigenesis. *Int J Mol Sci.* 2018; 19:2676.
<https://doi.org/10.3390/ijms19092676>
PMID:[30205617](https://pubmed.ncbi.nlm.nih.gov/30205617/)
19. Fararjeh AS, Chen LC, Ho YS, Cheng TC, Liu YR, Chang HL, Chang HW, Wu CH, Tu SH. Proteasome 26S Subunit, non-ATPase 3 (PSMD3) Regulates Breast Cancer by Stabilizing HER2 from Degradation. *Cancers (Basel).* 2019; 11:527.
<https://doi.org/10.3390/cancers11040527>
PMID:[31013812](https://pubmed.ncbi.nlm.nih.gov/31013812/)
20. Dai YJ, Hu F, He SY, Wang YY. Epigenetic landscape analysis of lncRNAs in acute myeloid leukemia with DNMT3A mutations. *Ann Transl Med.* 2020; 8:318.
<https://doi.org/10.21037/atm.2020.02.143>
PMID:[32355762](https://pubmed.ncbi.nlm.nih.gov/32355762/)
21. Bencomo-Alvarez AE, Rubio AJ, Olivas IM, Gonzalez MA, Ellwood R, Fiol CR, Eide CA, Lara JJ, Barreto-Vargas C, Jave-Suarez LF, Nteliopoulos G, Reid AG, Milojkovic D, et al. Proteasome 26S subunit, non-ATPases 1 (PSMD1) and 3 (PSMD3), play an oncogenic role in chronic myeloid leukemia by stabilizing nuclear factor-kappa B. *Oncogene.* 2021; 40:2697–710.
<https://doi.org/10.1038/s41388-021-01732-6>
PMID:[33712704](https://pubmed.ncbi.nlm.nih.gov/33712704/)
22. van der Westhuizen ET, Werry TD, Sexton PM, Summers RJ. The relaxin family peptide receptor 3 activates extracellular signal-regulated kinase 1/2 through a protein kinase C-dependent mechanism. *Mol Pharmacol.* 2007; 71:1618–29.
<https://doi.org/10.1124/mol.106.032763>
PMID:[17351017](https://pubmed.ncbi.nlm.nih.gov/17351017/)
23. Huang YW, Luo J, Weng YI, Mutch DG, Goodfellow PJ, Miller DS, Huang TH. Promoter hypermethylation of CIDEA, HAAO and RFXP3 associated with microsatellite instability in endometrial carcinomas. *Gynecol Oncol.* 2010; 117:239–47.
<https://doi.org/10.1016/j.ygyno.2010.02.006>
PMID:[20211485](https://pubmed.ncbi.nlm.nih.gov/20211485/)
24. Han S, Huang K, Gu Z, Wu J. Tumor immune microenvironment modulation-based drug delivery strategies for cancer immunotherapy. *Nanoscale.* 2020; 12:413–36.
<https://doi.org/10.1039/c9nr08086d>
PMID:[31829394](https://pubmed.ncbi.nlm.nih.gov/31829394/)
25. Corthay A. Does the immune system naturally protect against cancer? *Front Immunol.* 2014; 5:197.
<https://doi.org/10.3389/fimmu.2014.00197>
PMID:[24860567](https://pubmed.ncbi.nlm.nih.gov/24860567/)
26. Merdan S, Subramanian K, Ayer T, Van Weyenbergh J, Chang A, Koff JL, Flowers C. Gene expression profiling-based risk prediction and profiles of immune infiltration in diffuse large B-cell lymphoma. *Blood Cancer J.* 2021; 11:2.
<https://doi.org/10.1038/s41408-020-00404-0>
PMID:[33414466](https://pubmed.ncbi.nlm.nih.gov/33414466/)
27. Risueño A, Hagner PR, Towfic F, Fontanillo C, Djebbari A, Parker JS, Drew CP, Nowakowski GS, Maurer MJ, Cerhan JR, Wei X, Ren Y, Lee CW, et al. Leveraging gene expression subgroups to classify DLBCL patients and select for clinical benefit from a novel agent. *Blood.* 2020; 135:1008–18.
<https://doi.org/10.1182/blood.2019002414>
PMID:[31977005](https://pubmed.ncbi.nlm.nih.gov/31977005/)
28. Borst J, Ahrends T, Bąbała N, Melief CJM, Kastenmüller W. CD4⁺ T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol.* 2018; 18:635–47.
<https://doi.org/10.1038/s41577-018-0044-0>
PMID:[30057419](https://pubmed.ncbi.nlm.nih.gov/30057419/)
29. Wculek SK, Cueto FJ, Mujal AM, Melero I, Krummel MF, Sancho D. Dendritic cells in cancer immunology and immunotherapy. *Nat Rev Immunol.* 2020; 20:7–24.
<https://doi.org/10.1038/s41577-019-0210-z>
PMID:[31467405](https://pubmed.ncbi.nlm.nih.gov/31467405/)
30. Shimasaki N, Jain A, Campana D. NK cells for cancer immunotherapy. *Nat Rev Drug Discov.* 2020; 19:200–18.
<https://doi.org/10.1038/s41573-019-0052-1>
PMID:[31907401](https://pubmed.ncbi.nlm.nih.gov/31907401/)
31. Fridman WH, Pagès F, Sautès-Fridman C, Galon J. The immune contexture in human tumours: impact on clinical outcome. *Nat Rev Cancer.* 2012; 12:298–306.
<https://doi.org/10.1038/nrc3245>
PMID:[22419253](https://pubmed.ncbi.nlm.nih.gov/22419253/)
32. Ding ZC, Huang L, Blazar BR, Yagita H, Mellor AL, Munn DH, Zhou G. Polyfunctional CD4⁺ T cells are essential for eradicating advanced B-cell lymphoma after chemotherapy. *Blood.* 2012; 120:2229–39.
<https://doi.org/10.1182/blood-2011-12-398321>
PMID:[22859605](https://pubmed.ncbi.nlm.nih.gov/22859605/)

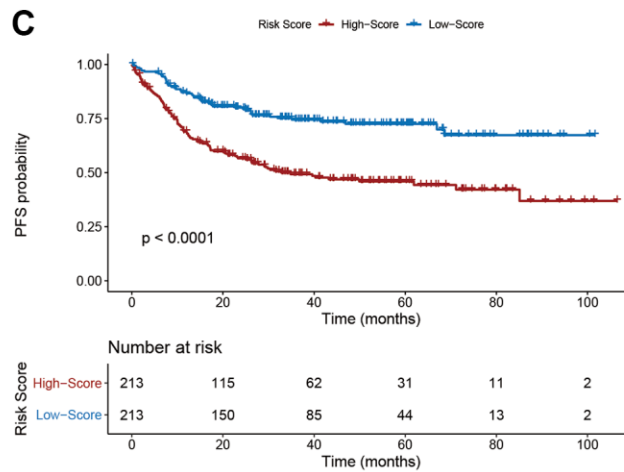
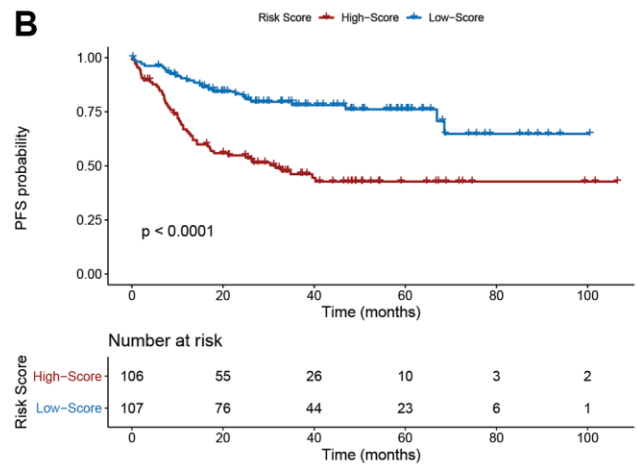
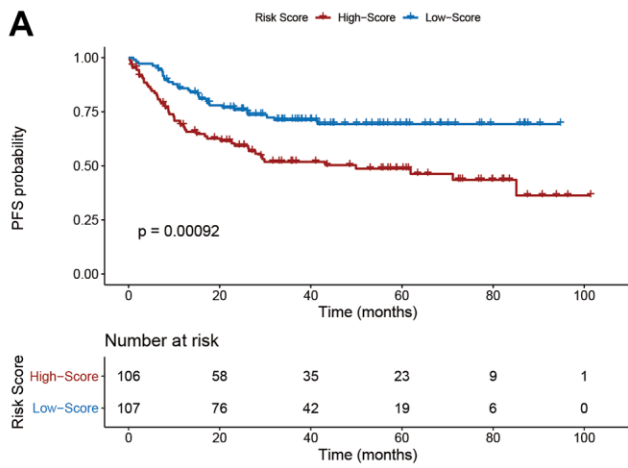
33. Ott G, Ziepert M, Klapper W, Horn H, Szczepanowski M, Bernd HW, Thorns C, Feller AC, Lenze D, Hummel M, Stein H, Müller-Hermelink HK, Frank M, et al. Immunoblastic morphology but not the immunohistochemical GCB/nonGCB classifier predicts outcome in diffuse large B-cell lymphoma in the RICOVER-60 trial of the DSHNHL. *Blood*. 2010; 116:4916–25.
<https://doi.org/10.1182/blood-2010-03-276766>
PMID:[20736456](https://pubmed.ncbi.nlm.nih.gov/20736456/)
34. Rimsza LM, Roberts RA, Miller TP, Unger JM, LeBlanc M, Braziel RM, Weisenberger DD, Chan WC, Muller-Hermelink HK, Jaffe ES, Gascoyne RD, Campo E, Fuchs DA, et al. Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors: a follow-up study from the Leukemia and Lymphoma Molecular Profiling Project. *Blood*. 2004; 103:4251–58.
<https://doi.org/10.1182/blood-2003-07-2365>
PMID:[14976040](https://pubmed.ncbi.nlm.nih.gov/14976040/)
35. Ciavarella S, Vegliante MC, Fabbri M, De Summa S, Melle F, Motta G, De Iulii V, Opinto G, Enjuanes A, Rega S, Gulino A, Agostinelli C, Scattone A, et al. Dissection of DLBCL microenvironment provides a gene expression-based predictor of survival applicable to formalin-fixed paraffin-embedded tissue. *Ann Oncol*. 2018; 29:2363–70.
<https://doi.org/10.1093/annonc/mdy450>
PMID:[30307529](https://pubmed.ncbi.nlm.nih.gov/30307529/)
36. DeNardo DG, Ruffell B. Macrophages as regulators of tumour immunity and immunotherapy. *Nat Rev Immunol*. 2019; 19:369–82.
<https://doi.org/10.1038/s41577-019-0127-6>
PMID:[30718830](https://pubmed.ncbi.nlm.nih.gov/30718830/)
37. Wada N, Zaki MA, Hori Y, Hashimoto K, Tsukaguchi M, Tatsumi Y, Ishikawa J, Tominaga N, Sakoda H, Take H, Tsudo M, Kuwayama M, Morii E, Aozasa K, and Osaka Lymphoma Study Group. Tumour-associated macrophages in diffuse large B-cell lymphoma: a study of the Osaka Lymphoma Study Group. *Histopathology*. 2012; 60:313–19.
<https://doi.org/10.1111/j.1365-2559.2011.04096.x>
PMID:[22211289](https://pubmed.ncbi.nlm.nih.gov/22211289/)
38. Scott DW, Gascoyne RD. The tumour microenvironment in B cell lymphomas. *Nat Rev Cancer*. 2014; 14:517–34.
<https://doi.org/10.1038/nrc3774>
PMID:[25008267](https://pubmed.ncbi.nlm.nih.gov/25008267/)
39. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4:249–64.
<https://doi.org/10.1093/biostatistics/4.2.249>
PMID:[12925520](https://pubmed.ncbi.nlm.nih.gov/12925520/)
40. Bhattacharya S, Dunn P, Thomas CG, Smith B, Schaefer H, Chen J, Hu Z, Zalocusky KA, Shankar RD, Shen-Orr SS, Thomson E, Wiser J, Butte AJ. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. *Sci Data*. 2018; 5:180015.
<https://doi.org/10.1038/sdata.2018.15>
PMID:[29485622](https://pubmed.ncbi.nlm.nih.gov/29485622/)
41. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015; 12:453–57.
<https://doi.org/10.1038/nmeth.3337>
PMID:[25822800](https://pubmed.ncbi.nlm.nih.gov/25822800/)
42. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013; 4:2612.
<https://doi.org/10.1038/ncomms3612>
PMID:[24113773](https://pubmed.ncbi.nlm.nih.gov/24113773/)
43. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50.
<https://doi.org/10.1073/pnas.0506580102>
PMID:[16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)

SUPPLEMENTARY MATERIALS

Supplementary Figures



Supplementary Figure 1. Stratification analyses of overall survival between high- and low-risk patients in different subgroups. (A) Age ≤ 60. (B) Age > 60. (C) Male. (D) Female. (E) Stage I-II. (F) Stage III-IV. (G) ECOG < 2. (H) ECOG ≥ 2. (I) LDH Normal. (J) LDH Elevated. (K) Number of extranodal sites < 2. (L) Number of extranodal sites ≥ 2. (M) Subtype GCB. (N) Subtype ABC. (O) Subtype Unclassified.



Supplementary Figure 2. Kaplan-Meier plots of progression-free survival between high- and low-risk patients. (A) Discovery cohort, (B) internal validation cohort, (C) entire GSE31312 cohort.

Supplementary Table

Supplementary Table 1. Clinical and pathological characteristics of patients with DLBCL in discovery and internal validation cohort.

Variables	GSE31312 N (%)	GSE10846 N (%)
Age, years		
≤60	179 (42.02)	146 (47.87)
>60	247 (57.98)	159 (52.13)
Sex		
Female	183 (42.96)	134 (43.93)
Male	243 (57.04)	171 (56.07)
Stage		
I–II	200 (46.95)	144 (47.21)
III–IV	226 (53.05)	161 (52.79)
No. of extranodal sites		
<2	331 (77.70)	282 (92.46)
≥2	95 (22.30)	23 (7.54)
ECOG		
<2	350 (82.16)	230 (75.41)
≥2	76 (17.84)	75 (24.59)
LDH		
Normal	148 (34.74)	153 (50.16)
Elevated	278 (65.26)	152 (49.87)
Subtype		
GCB	203 (47.65)	133 (43.61)
ABC	183 (42.96)	125 (40.98)
Unclassified	40 (9.39)	47 (15.41)

Abbreviations: ECOG: Eastern Cooperative Oncology Group; LDH: lactate dehydrogenase; GCB: germinal center B-cell-like lymphoma; ABC: activated B-cell-like lymphoma. Variables were checked via χ^2 -test.