



A Novel Defined Super-Enhancer Associated Gene Signature to Predict Prognosis in Patients With Diffuse Large B-Cell Lymphoma

Hong Xu^{1†}, Yuhang Li^{1†}, Yanan Jiang^{1†}, Jinhuan Wang², Huimeng Sun¹, Wenqi Wu¹, Yangyang LV¹, Su Liu¹, Yixin Zhai¹, LinYan Tian¹, Lanfang Li^{3*} and Zhigang Zhao^{1*}

¹Department of Hematology, Key Laboratory of Cancer Prevention and Therapy, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, ²Department of Oncology, Institute of Urology, Second Hospital of Tianjin Medical University, Tianjin, China, ³Departments 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, Tianjin, China

OPEN ACCESS

Edited by:

Sun Liang,
Shandong First Medical University,
China

Reviewed by:

Koen H. M. Prange,
University of Amsterdam, Netherlands
Yusuke Shiozawa,
Nippon Medical School, Japan

*Correspondence:

Lanfang Li
lilanfangmeng@163.com
Zhigang Zhao
zzhao01@tmu.edu.cn

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 02 December 2021

Accepted: 18 May 2022

Published: 14 June 2022

Citation:

Xu H, Li Y, Jiang Y, Wang J, Sun H,
Wu W, LV Y, Liu S, Zhai Y, Tian L, Li L
and Zhao Z (2022) A Novel Defined
Super-Enhancer Associated Gene
Signature to Predict Prognosis in
Patients With Diffuse Large B-
Cell Lymphoma.
Front. Genet. 13:827840.
doi: 10.3389/fgene.2022.827840

Background: Diffuse large B-cell lymphoma (DLBCL) is a genetically heterogeneous disease that can have profound differences in survival outcomes. A variety of powerful prognostic factors and models have been constructed; however, the development of more accurate prognosis prediction and targeted treatment for DLBCL still faces challenges. An explosion of research on super-enhancer (SE)-associated genes provide the possibility to use in prognostication for cancer patients. Here, we aimed to establish a novel effective prognostic model using SE-associated genes from DLBCL.

Methods: A total of 1,105 DLBCL patients from the Gene Expression Omnibus database were included in this study and were divided into a training set and a validation set. A total of 11 SE-associated genes (BCL2, SPAG16, PDK, BTG1, LRR37A2, EXT1, TGFB2, ANKRD12, MYCBP2, PAX5, and MYC) were initially screened and identified by the least absolute shrinkage and selection operator (Lasso) penalized Cox regression, univariate and multivariate Cox regression analysis. Finally, a risk score model based on these 11 genes was constructed.

Results: Kaplan–Meier (K–M) curves showed that the low-risk group appeared to have better clinical survival outcomes. The excellent performance of the model was determined via time-dependent receiver operating characteristic (ROC) curves. A nomogram based on the polygenic risk score was further established to promote reliable prognostic prediction. This study proposed that the SE-associated-gene risk signature can effectively predict the response to chemotherapy in DLBCL patients.

Conclusion: A novel and reliable SE-associated-gene signature that can effectively classify DLBCL patients into high-risk and low-risk groups in terms of overall survival was developed, which may assist clinicians in the treatment of DLBCL.

Keywords: super-enhancer, LASSO, diffuse large B-cell lymphoma, prognostic model, overall survival

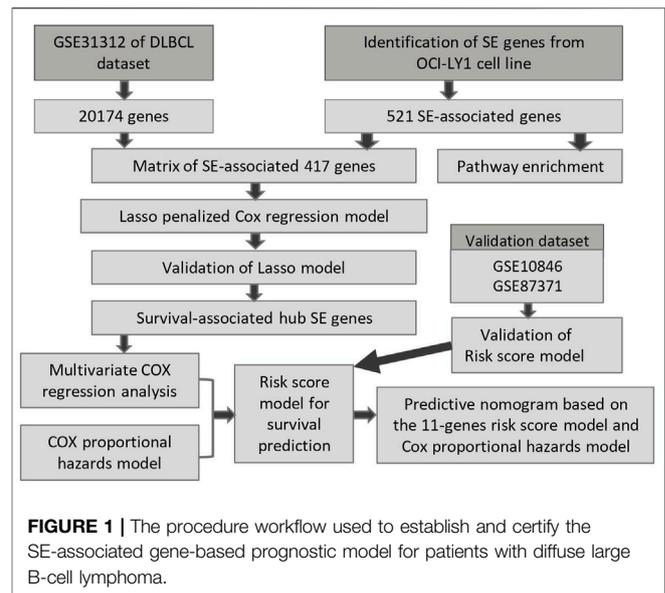
INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is the most common type of non-Hodgkin's lymphoma (NHL), accounting for 30%–40% of all newly diagnosed NHL cases (Armitage et al., 2017; Siegel et al., 2017). DLBCL is an aggressive, severe, and complex disease with broad genetic, phenotypic, and clinical heterogeneities (Abramson and Shipp, 2005). The heterogeneity of the disease results in different survival outcomes in DLBCL patients receiving standard therapy (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP)) (Younes, 2015). About 30–40% of patients do not respond well to standard treatment, with the highest mortality rate in the first 2 years after diagnosis (Yin et al., 2019).

In the era of rituximab, the International Prognostic Index (IPI) is one of the most important tools for prognostic risk stratification. The subsequent revisions have appeared to improve the prognostic evaluation system in DLBCL patients. Disappointingly, these prognostic indicators do not address the underlying biological heterogeneity of DLBCL. Therefore, it is urgent to explore novel and effective molecular markers for a more accurate prediction of the prognosis of patients with DLBCL.

Super-enhancers (SEs) have been described as a class of regulatory domains with unusually strong transcription-assisted activator binding capacity (Parker et al., 2013; Whyte et al., 2013). SE is a cluster of enhancers that has a stronger ability to promote transcription compared to the typical enhancers (TEs). Compared with normal cells, tumor cells construct SEs on oncogenes during tumorigenesis and recruit enhancer-binding proteins to drive gene expression (Lovén et al., 2013). SEs are generally occupied with abundant signals of H3K4me1, H3K27ac, p300, Mediator, RNA polymerase II, BRD4, CDK7, and other master transcription factors (Wang et al., 2019); among them, H3K27ac is the preferred marker for the identification of super-enhancers (Hnisz et al., 2013). The loss or gain of SEs has been reported in various tumors (He et al., 2019); similarly, SEs play a key role in the progression of DLBCL by activating the expression of downstream oncogenes (Chapuy et al., 2013). In addition, SE inhibitors (JQ1) used to treat DLBCL suppress the expression of these genes (Li et al., 2021). Therefore, the exploitation and identification of SEs-driven hub oncogenes will provide novel insights into the diagnosis, prognosis, and treatment of DLBCL.

The least absolute shrinkage and selection operator (Lasso) penalized Cox regression is a variable selection and contraction method in Cox's proportional risk model proposed by Tibshirani (1997). Lasso can reduce the number of variables compared to traditional stepwise regression because less influential variables will be regularized by shrinking their coefficients to zero (Zhang et al., 2018). Currently, Lasso is widely used to build survival prediction models based on complex, high-throughput genomic data. Wu et al. (2021) identified ten important immune-related genes most associated with the overall survival of DLBCL patients among the 26 immune-related genes by using Lasso regression analysis. Similarly, using group Lasso, an 11-SE-related-gene



signature effectively predicted overall survival in DLBCL. Thus, we applied the Lasso regression method to construct a prognostic model of DLBCL.

In this study, Lasso penalized Cox regression analysis was performed using 521 SE-associated genes. A gene cluster containing 11 SE-related genes (BCL2, SPAG16, PXX, BTG1, LRRK37A2, EXT1, TGFBR2, ANKRD12, MYCBP2, PAX5, and MYC) was screened. Subsequently, a risk score model based on these 11 genes was constructed, which was helpful for risk stratification and prognosis. Finally, based on the model, an interactive nomogram containing 11 gene risk groups and clinical characteristics was established, which provides a tool to predict the overall survival (OS) of DLBCL patients clinically. The workflow of our study is shown in **Figure 1**.

MATERIALS AND METHODS

Data Source

The microarray data and corresponding clinical information from GSE31312 as the training data and the two other independent datasets, GSE10846 and GSE80371, as the external validation datasets were obtained from Gene Expression Omnibus (GEO) database. 470 DLBCL samples were enrolled in GSE31312, 414 in GSE10856, and 221 in GSE80371.

Identification of Super-Enhancer-Associated Genes

The 521 SE-associated genes identified from the DLBCL cell line OCI-LY1 were obtained from the website <http://dbccrc.cam-su.org>. H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) signal was used to screen SE-associated genes in the OCI-LY1 cell line. The biological function of these genes was revealed by Gene Ontology (GO) enrichment and Kyoto

Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. To get the final expression matrix, we retained the genes that overlapped between GSE31312 datasets and the SE-associated genes in OCY-LY1.

Lasso Penalized Cox Regression Analysis

To screen the important and potential prognostic genes, Lasso penalized Cox regression analysis was performed to establish a predicting model using the R package “glmnet”. We identified the optimal lambda (λ) value based on ten-fold cross-validation. Two best-fit values (λ_{\min} and λ_{1se}) were chosen by minimizing the mean cross-validated error to construct the Lasso models. Subsequently, we performed the Wilcoxon test and ROC curve analysis to compare the two parameters.

Development of the Prognostic Signature

To construct an optimal prognostic prediction model, we integrated the candidate genes' expression levels weighted by their regression coefficients and calculated the risk score for each patient, according to the formula $\text{RiskScore} = \sum \beta_i * X_i$. Here, X_i is the gene expression level, and β_i is the regression coefficient. Regarding the value obtained from the maximally standardized long-rank statistics as a cutoff point, DLBCL patients were separated into high- and low-risk groups.

Cox Proportional Hazard Regression Analyses

The univariate and multivariate Cox proportional hazard regression models were utilized to identify the correlation between the gene expression level of the candidate genes and OS, which was accomplished by R packages “survival” and “survminer”. The results were shown on the forest plot. The analyses were also applied to verify the independence of the constructed prognostic model with other clinical features. The parameters included the prognostic risk score and some important clinicopathological factors, such as age, gender, clinical stage, the situation of extranodal invasion, Eastern Cooperative Oncology Group (ECOG) score, lactate dehydrogenase (LDH), and IPI score. The p -value, hazard ratio (HR), and 95% confidence interval (CI) of each factor were calculated.

Kaplan–Meier Analysis and Time-dependent Receiver Operator Characteristic Curve Analysis

The Kaplan–Meier analysis method was used to compare the differences in OS and progression-free survival (PFS) between low- and high-risk groups, and the log-rank tests were performed to measure the statistical significance (p -value of less than 0.05). The R packages “survival” and “survminer” were used to execute the analysis. Moreover, we depicted the time-dependent ROC curve to assess the predictive capability for different factors by figuring out the area under the ROC (AUC) ($p < 0.05$).

Predictive Nomogram

In total, seven prognostic predictors (six clinical features and the 11-genes risk score) were enrolled to build the predictive nomogram, which was used to forecast the 1-year, 3-year, and 5-year OS of the patients *via* R package “rms”. We calculated the concordance index (C-index) by package “Hmisc” to evaluate the discrimination of the nomogram. Furthermore, calibration curves were plotted for intuitionistic comparison of the predicted against the actual survival probabilities. Data of one randomly selected patient from GSE31312 were used to validate the probability of 1–5-year OS, based on the predictors in the nomogram. Total points were calculated using the R package “nomogramEx”. Finally, the interactive nomogram was developed and visually displayed by the R package “regplot”.

Chemotherapy Response With Super-Enhancer-Associated Genes Signature

In order to predict the chemotherapy response in the low- and high-risk groups, the R package “pRRophetic” was applied for profiling. We straightforwardly compared the estimated half-maximal inhibitory concentration (IC_{50}) between low- and high-risk groups among the different chemotherapeutics, which exactly proved the hypothesis that the low-risk group was likely more sensitive to the chemotherapy.

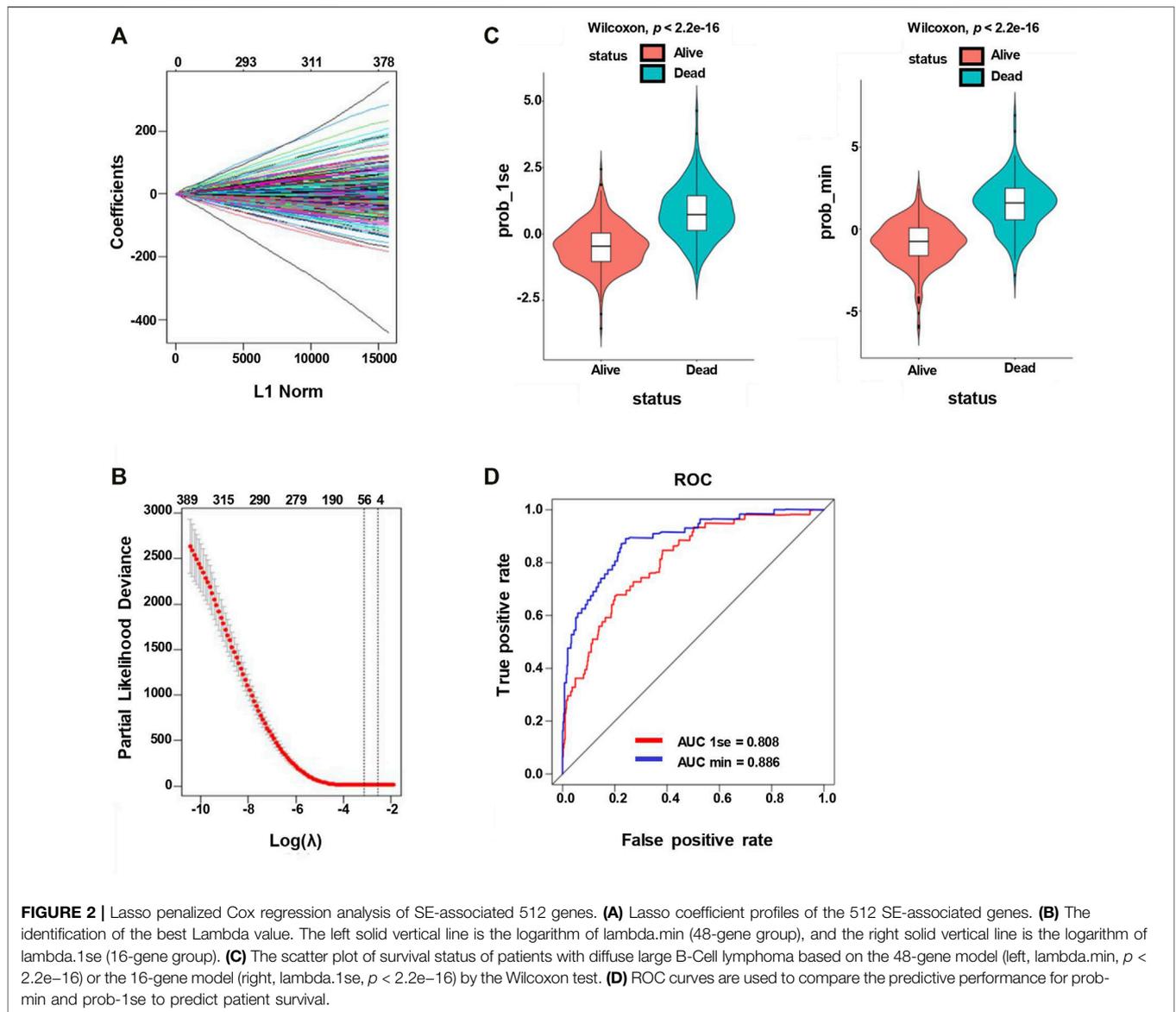
Protein–Chemical Interactions Analysis and Chromatin Immunoprecipitation Sequencing Profile for H3K27ac Signal Tracks

We established an interactive network of the hub genes and chemicals to probe into the chemicals correlated to these genes by “NetworkAnalyst 3.0”, based on the data from the Comparative Toxicogenomics Database (CTD). In the end, we used H3K27ac as SE biomarkers based on the ChIP-seq profiles data from Cistrome to visualize the location of the SEs regions and their target genes.

RESULT

Establishment of the Lasso Penalized Cox Regression Model

A 20,174-gene expression matrix of GSE31312 and the corresponding clinical information of 470 DLBCL patients were downloaded from the GEO database under the accession number GSE31312, as described in **Supplementary Table S1**. In total, 521 SE-associated genes identified from the DLBCL cell line OCY-LY1 were obtained from the website <http://dbcrc.cam-su.org>. Pathway enrichment analysis indicated that these SE-associated genes were closely related to lymphocyte activation and small GTPase mediated signal transduction (**Supplementary Figures S2A and C**). We extracted 417 genes that overlapped between GSE31312 datasets and the SE-associated



genes in OCY-LY1 to construct the expression matrix. The lasso penalized Cox regression analysis was applied to screen some potential and vital prognostic genes. We calculated the coefficient values at different levels of penalty (Figure 2A). First, we identified the optimal lambda (λ) value based on ten-fold cross-validation. Two best-fit values (lambda.min and lambda.1se) were chosen by minimizing the mean-square error to construct the Lasso models, and we selected two groups of genes (48-gene group of λ_{\min} and 16-gene group of λ_{1se} ; Figure 2B). As shown in Figure 2C, the lasso models were reconstructed according to the λ_{\min} and λ_{1se} , and both models performed well to separate the survival and death events (Wilcoxon test, $p < 2.2e-16$). The result of the ROC curves analysis for the two predictive models showed the AUCs were 0.808 (λ_{1se}) and 0.886 (λ_{\min}), suggesting that both models had a promising performance in predicting the probability of overall

survival (Figure 2D). Considering that there was no significant difference in the predictive performance of the two models according to AUC and Wilcoxon tests, we further studied the 16-gene model.

Association Between Candidate Genes and Prognosis

We utilized multivariate Cox regression analyses to explore whether each of the candidate genes is associated with the overall survival. As the outcome of the multivariate Cox regression analysis shown in Figure 3A, the global p -value of the predictive model was $1.8483e-30$, with the Akaike information criterion (AIC) of 1768.55 and C-index of 0.77. Multivariate Cox regression showed that BCL2, SPAG16, PXX, BTG1, LRRC37A2, EXT1, TGFB2, ANKRD12, MYCBP2,

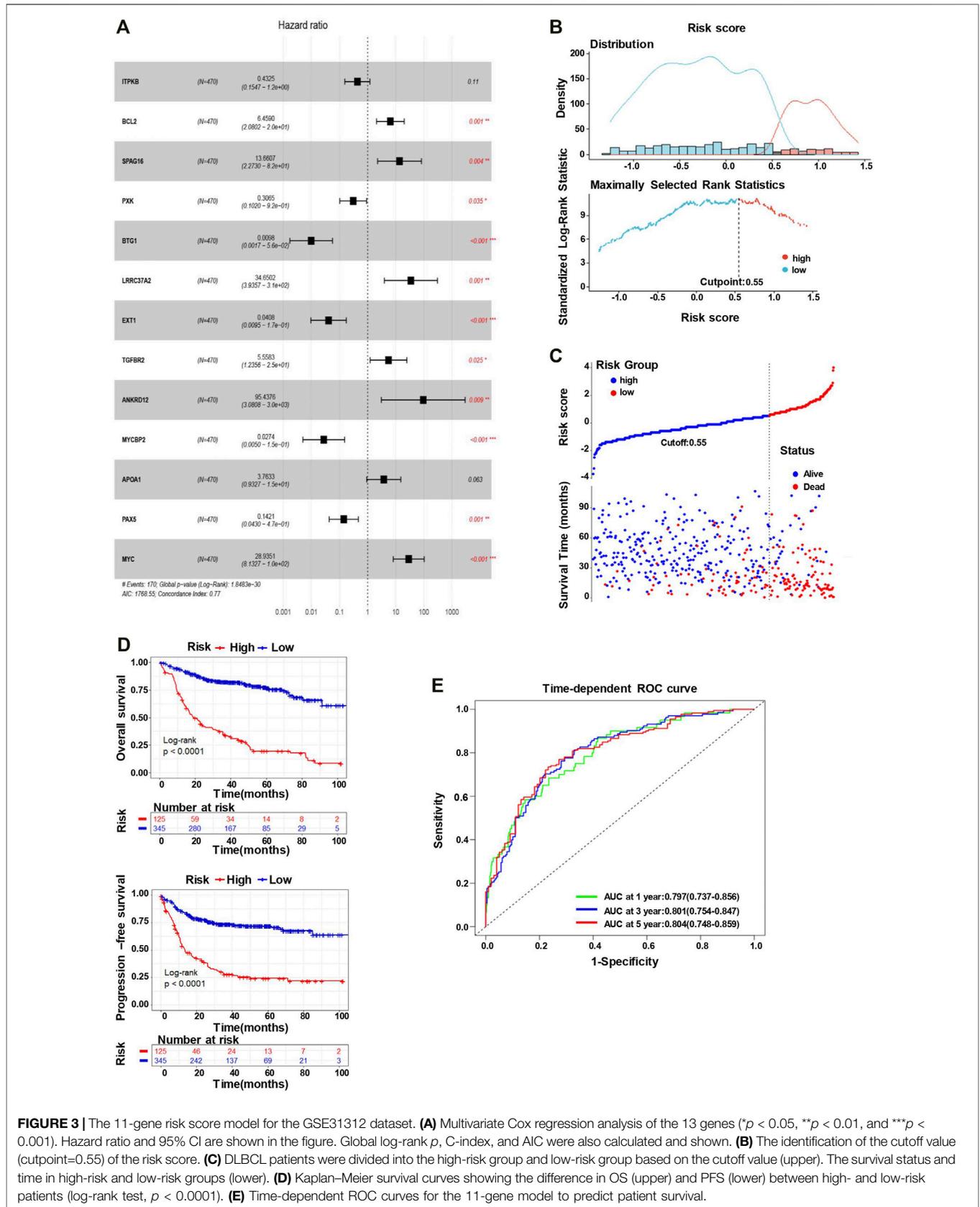
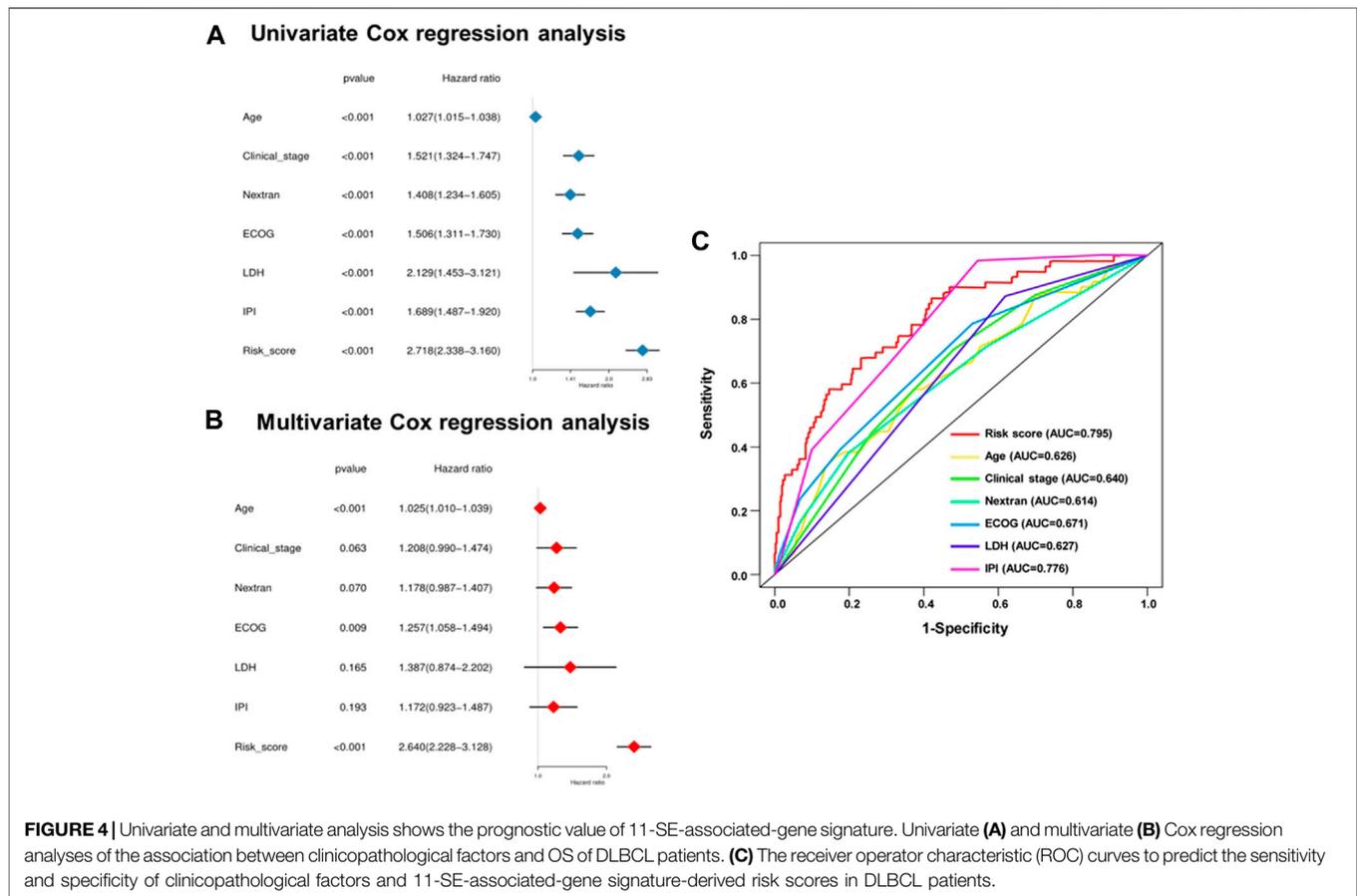


FIGURE 3 | The 11-gene risk score model for the GSE31312 dataset. **(A)** Multivariate Cox regression analysis of the 13 genes ($*p < 0.05$, $**p < 0.01$, and $***p < 0.001$). Hazard ratio and 95% CI are shown in the figure. Global log-rank p , C-index, and AIC were also calculated and shown. **(B)** The identification of the cutoff value (cutpoint=0.55) of the risk score. **(C)** DLBCL patients were divided into the high-risk group and low-risk group based on the cutoff value (upper). The survival status and time in high-risk and low-risk groups (lower). **(D)** Kaplan–Meier survival curves showing the difference in OS (upper) and PFS (lower) between high- and low-risk patients (log-rank test, $p < 0.0001$). **(E)** Time-dependent ROC curves for the 11-gene model to predict patient survival.



PAX5, and MYC were significantly associated with the overall survival of DLBCL patients. Among these genes, BCL2, SPAG16, LRRC37A2, TGFBR2, ANKRD12, and MYC may appear to be the risky factors (HR > 1), while PXX, BTG1, EXT1, MYCBP2, and PAX5 seemed to act as the protective factors (HR < 1). To optimize the predictive model, we selected these 11 SE-associated genes to forecast the OS of DLBCL patients.

Establishment and Validation of the 11-Gene Risk Score Model

The risk scores predicted by the coefficient of these 11 candidate genes from the multivariate Cox regression analysis (the equation for risk scores is shown in Materials and Method) stratify the patients into the low-risk ($n = 345$) and high-risk ($n = 125$) groups, with the cutoff point of 0.55 (Figure 3B). As the outcome shown, the number of alive events is significantly more in the low-risk group, while the death events are obviously more frequent in the high-risk group (Figure 3C). Subsequently, we conducted a K–M analysis to compare the differences in OS and PFS between low- and high-risk groups. The K–M survival curve of OS demonstrated an inferior outcome in the high-risk group (long-rank test, $p < 0.0001$), consistent with the analysis of PFS (Figure 3D). Furthermore, the time-dependent ROC

analysis also showed a favorable outcome, where the AUC was 0.797 at 1-year, 0.801 at 3-year, and 0.804 at 5-year (Figure 3E), indicating that the risk score model has a good performance to predict the prognostic outcomes.

Independence of 11-Gene Risk Score Model in Survival Prediction

Considering the effects of other important clinical indicators, such as age, gender, clinical stage, the situation of extranodal invasion, ECOG score, LDH level, and IPI score, we validated the independence of the polygenic prognostic predictive model via the univariate and multivariate Cox regression analyses. In the univariate Cox regression analysis, the risk score correlated with OS of the DLBCL patients (HR at 2.718, $p < 0.001$), similar to other important clinicopathological factors (Figure 4A). As for the multivariate Cox regression analysis, risk score appeared to be an independent and harmful factor for prediction (HR at 2.640, $p < 0.001$), while only Age and ECOG score among all clinical features showed statistical significance ($p < 0.001$ and $p = 0.009$, respectively) (Figure 4A; Table 1). The ROC curve analysis was a complement for verifying the predictive capacity of these indicators, which showed that the AUC of the risk score was 0.795, greater than other clinical indicators (Figure 4B). All these

TABLE 1 | Univariate and multivariate Cox regression analyses of the gene signature and overall survival of DLBCL patients in 3 independent datasets.

Variables	Patients(N)	Univariate analysis		Multivariate analysis		
		HR (95%CI)	p Value	HR (95%CI)	p Value	
GSE31312						
Age	>60/<=60	270/200	1.027 (1.015–1.038)	<0.001	1.025 (1.010–1.039)	<0.001
Clinical stage	III-IV/I-II	229/220	1.521 (1.324–1.747)	<0.001	1.208 (0.990–1.474)	0.063
Extranodal sites	>=2/<2	104/366	1.408 (1.234–1.605)	<0.001	1.178 (0.987–1.407)	0.07
ECOG	>=2/<2	96/374	1.506 (1.311–1.730)	<0.001	1.257 (1.058–1.494)	0.009
LDH	Evaluated/normal	278/148	2.129 (1.453–3.121)	<0.001	1.387 (0.874–2.202)	0.165
IPI score	>2/<=2	150/274	1.689 (1.487–1.920)	<0.001	1.172 (0.923–1.487)	0.193
Risk score	High/Low	125/345	2.718 (2.338–3.160)	<0.001	2.640 (2.228–3.128)	<0.001
GSE10846						
Age	>60/<=60	226/188	1.030 (1.018–1.041)	<0.001	1.030 (1.016–1.045)	<0.001
Clinical stage	III-IV/I-II	218/188	1.508 (1.293–1.758)	<0.001	1.313 (1.091–1.580)	0.004
Extranodal sites	>=2/<2	30/353	1.206 (1.001–1.452)	0.049	0.956 (0.755–1.210)	0.707
ECOG	>=2/<2	93/296	1.820 (1.551–2.136)	<0.001	1.534 (1.273–1.847)	<0.001
LDH	Evaluated/normal	178/173	1.137 (1.095–1.181)	<0.001	1.116 (1.059–1.177)	<0.001
Risk score	High/Low	126/288	2.718 (2.128–3.472)	<0.001	2.039 (1.542–2.697)	<0.001
GSE87371						
Age	>60/<=60	106/115	1.049 (1.025–1.072)	<0.001	1.010 (0.984–1.036)	0.454
Gender	Male/Female	116/105	1.499 (0.863–2.604)	0.151	1.336 (0.761–2.347)	0.313
Clinical stage	III-IV/I-II	150/71	1.802 (1.297–2.503)	<0.001	0.724 (0.441–1.186)	0.200
IPI score	>2/<=2	102/119	2.029 (1.622–2.538)	<0.001	1.666 (0.921–3.015)	0.092
Risk score	High/Low	79/142	2.718 (1.991–3.711)	<0.001	2.289 (1.690–3.099)	<0.001

results sufficiently confirmed that our 11-gene risk score model was an independent and robust predictor, which has promising application prospects in comparison with other well-established indicators.

Stratification Analysis

A stratification analysis was carried out to assess the predictive abilities of the risk score model within different clinical feature subgroups. Patients from the entire cohort were factitiously classified by age (>60 vs. <=60), gender (Male vs. Female), disease clinical stage (stage I–II vs. III–IV), the situation of extranodal invasion (extranodal sites >=2 vs. < 2), IPI score (>2 vs. <=2), and disease classification based on immunohistochemical (IHC) [activated B cell (ABC), germinal center B cell (GCB), and unclassified (UC)] as different subgroups. The risk score divided the patients in the same stratum into the low- and high-risk groups. We observed that the K–M curves could be distinguished by the risk score model irrespective of the subgroup, where all the high-risk groups had inferior survival outcomes (Supplementary Figure S1).

Development of Predictive Nomogram for Prognosis Prediction

There were seven prognostic predictors enrolled for building the predictive nomogram to forecast the 1-year, 3-year, and 5-year OS for the patients. The predictors of the nomogram involved the 11-genes risk score and the other six clinical indicators: age, clinical stage, ECOG, IPI, LDH, and extranodal sites (Figure 5A). Calibration curves were plotted for intuitionistic comparisons of the predicted against actual survival probabilities. The calibration curves of 1- to 5-year all appeared very close to the grey lines,

suggesting a powerful predictive ability of this nomogram (Figure 5B). In order to evaluate the predictive effect of the 11-genes risk score based on the nomogram, we randomly selected one specific patient from the entire cohort. We added up all the points from these clinical indicators and the 11-gene risk group; the total point was 551, compared with the total point of 382 when only considering the clinical variables. The probability of 1-, 3-, and 5-year OS were 0.335, 0.618, and 0.716, respectively, while taking both the clinical indicators and risk group into account. In reality, the patient died at 910 days, while the predictive probability of death at that day was 0.67. Meanwhile, when we only utilized the six clinical indicators, the probability of 1-, 3-, and 5-year OS were 0.194, 0.372, and 0.445, respectively. The predictive probability of death at 910 days was 0.408, obviously lower than the probability forecasted in consideration of the 11-genes risk score, as mentioned above (Figure 5C).

Validation of the 11-Genes Prognostic Signature in the External Datasets

To further validate the effect of the prognostic predictive model, we analyzed two independent external datasets, GSE10846 and GSE87371, with a similar working procedure as mentioned above. The detail of the clinical characteristics is also described in Supplementary Table S1. The risk scores of each cohort were calculated, which divided the patients into low- and high-risk groups. As the consistent result of the two datasets shown in Figure 6A, the overall survival was distinguished from different groups in K–M analysis (long-rank test, $p < 0.0001$). In addition, the time-dependent ROC curve analyses also performed favorable outcomes, in which the AUC of 1-year at 0.719, 3-year at 0.708, 5-

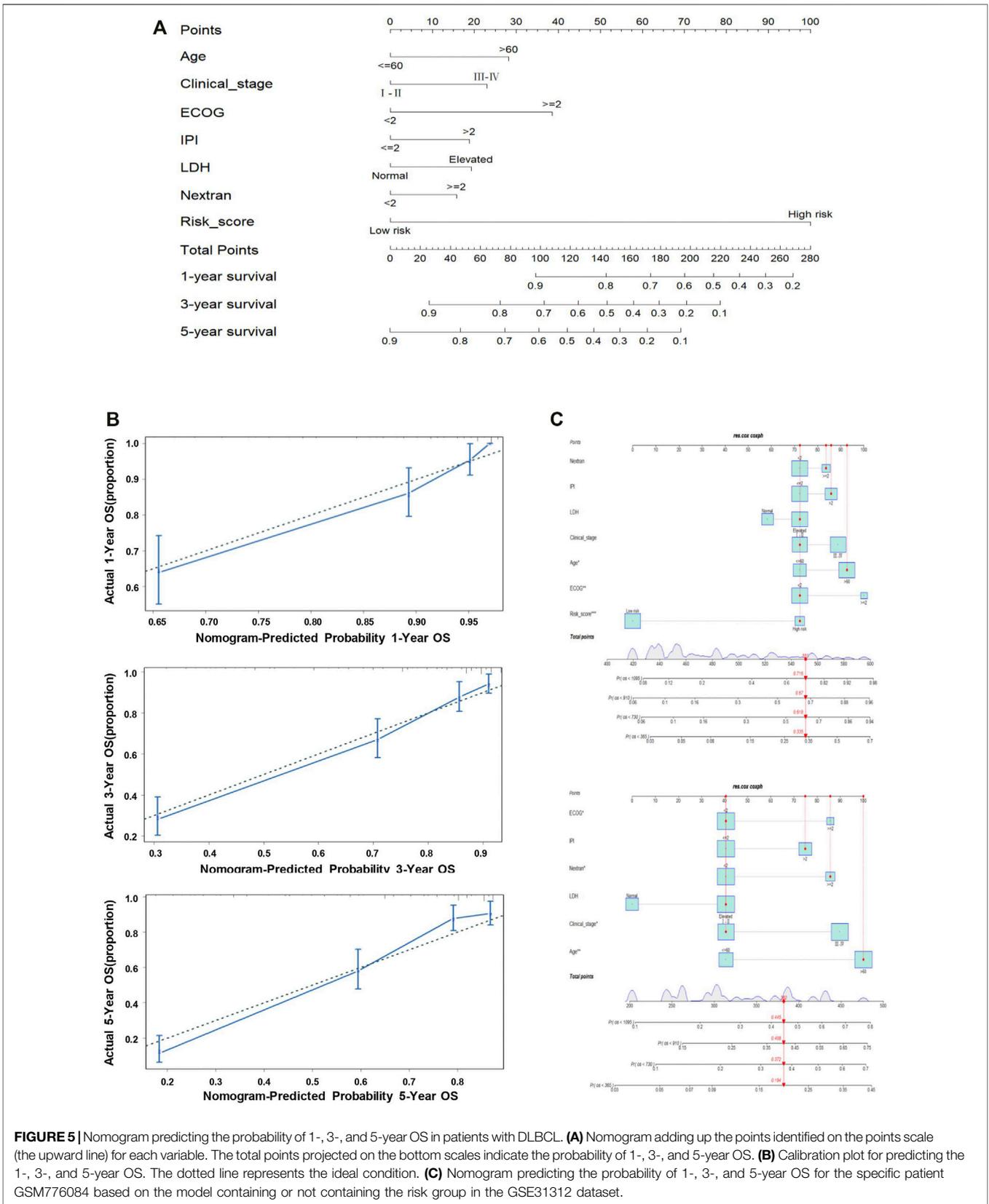


FIGURE 5 | Nomogram predicting the probability of 1-, 3-, and 5-year OS in patients with DLBCL. **(A)** Nomogram adding up the points identified on the points scale (the upward line) for each variable. The total points projected on the bottom scales indicate the probability of 1-, 3-, and 5-year OS. **(B)** Calibration plot for predicting the 1-, 3-, and 5-year OS. The dotted line represents the ideal condition. **(C)** Nomogram predicting the probability of 1-, 3-, and 5-year OS for the specific patient GSM776084 based on the model containing or not containing the risk group in the GSE31312 dataset.

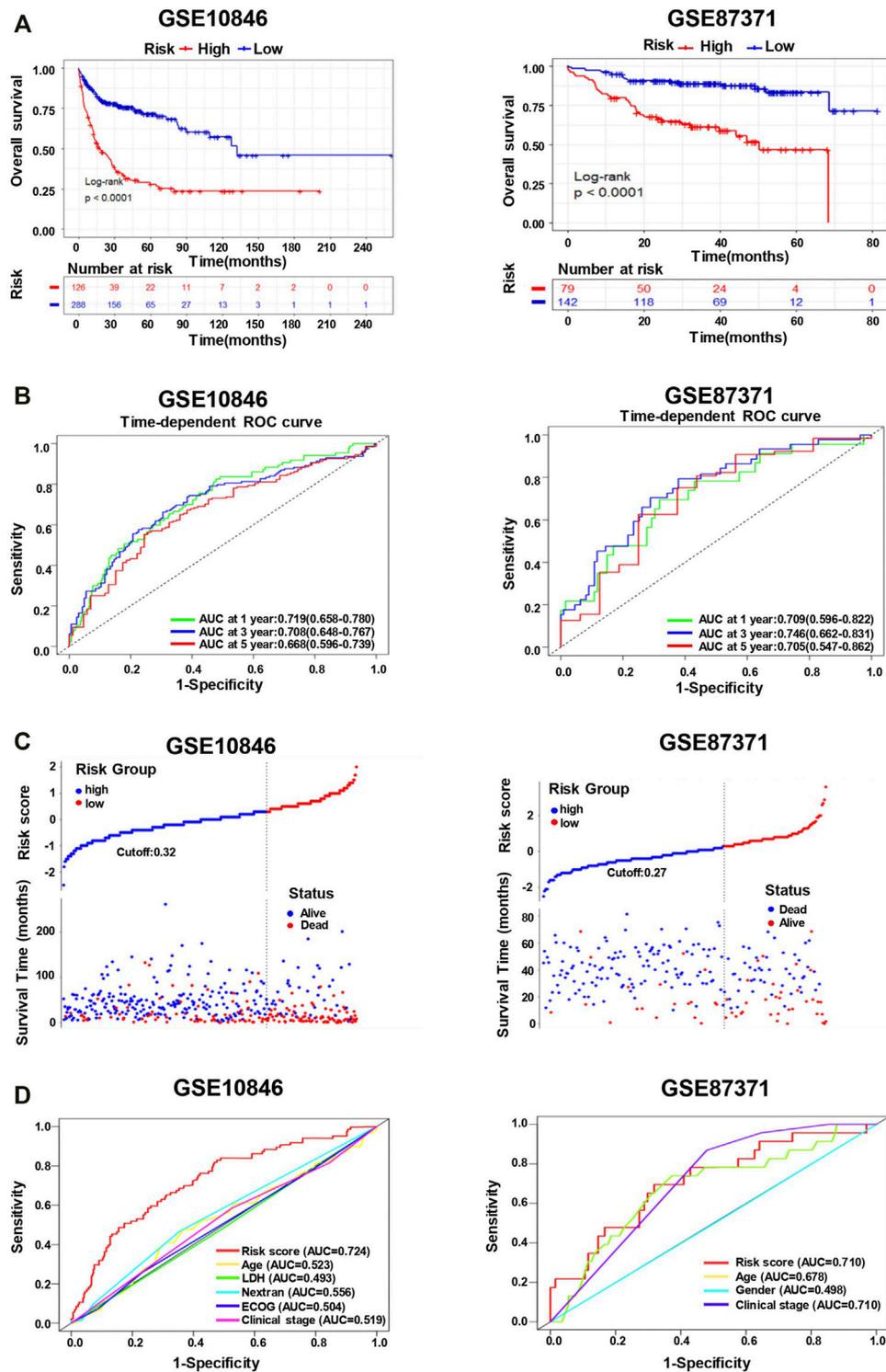
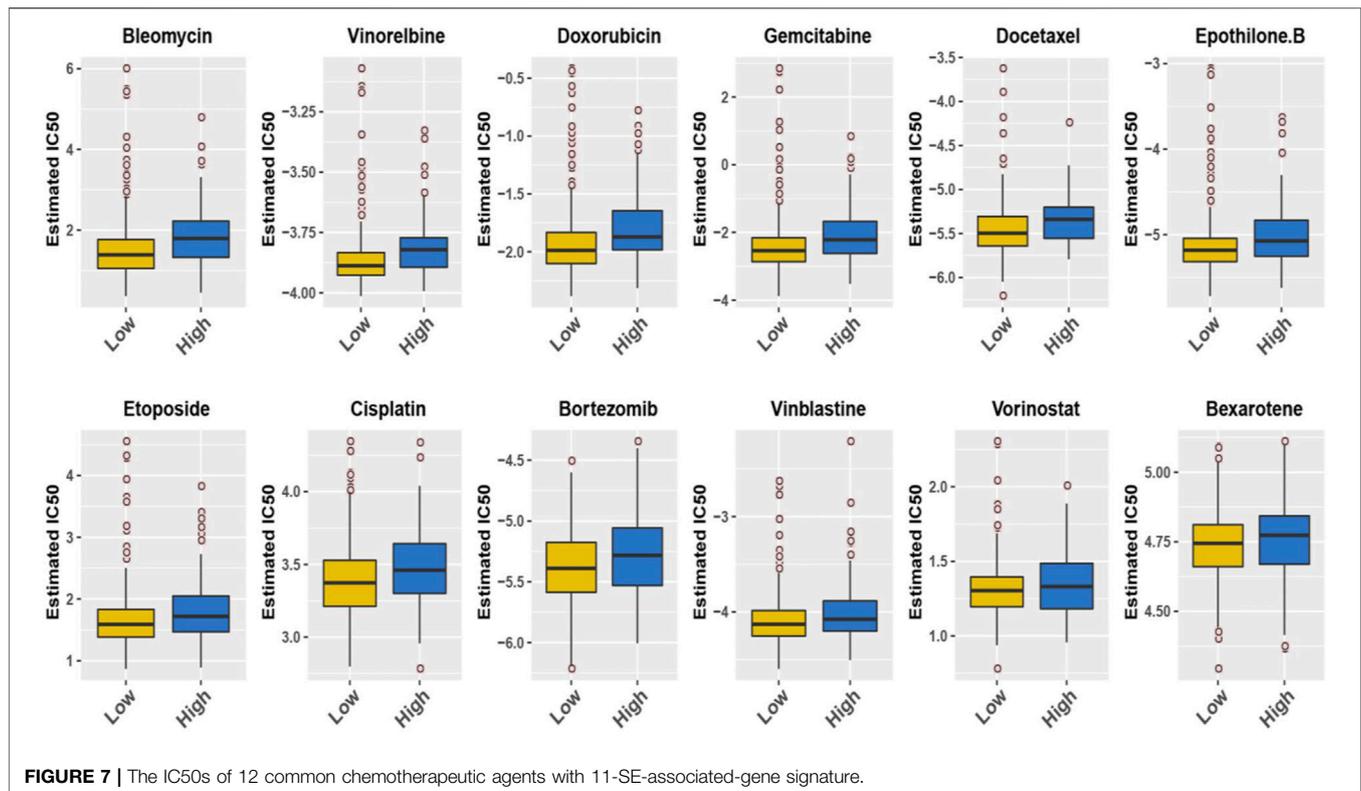


FIGURE 6 | The 11-gene risk score model for the validation datasets (GSE10846 and GSE87371). **(A)** Kaplan–Meier plots of overall survival in high-risk and low-risk subgroups in the validation datasets derived via Log-rank testing. **(B)** The time-dependent ROC curve and AUC in the validation datasets. **(C)** The survival status and time in high-risk and low-risk groups for the validation datasets. **(D)** The ROC curves to predict the sensitivity and specificity of clinicopathological factors and 11-gene signature-derived risk scores in DLBCL patients for the validation datasets.



year at 0.668 in GSE10846, and the AUC of 1-year at 0.709, 3-year at 0.746, 5-year at 0.705 in GSE87371 (**Figure 6B**). When the cutoff points were 0.32 and 0.27 in GSE10846 and GSE87371, respectively, the patients were separated into low- and high-risk groups subsequently. There were more death events in the high-risk group from both datasets (**Figure 6C**). Moreover, we also conducted the ROC curve analyses to evaluate the predictive performance of the 11-genes risk score model and some other clinical variables. The AUCs of the risk score were 0.724 in GSE10846 and 0.710 in GSE87371, significantly greater than that of any other clinical parameters (**Figure 6D**). The univariate and multivariate Cox regression analyses were also used for the two datasets, as shown in **Table 1**, and the outcome is consistent with the training dataset.

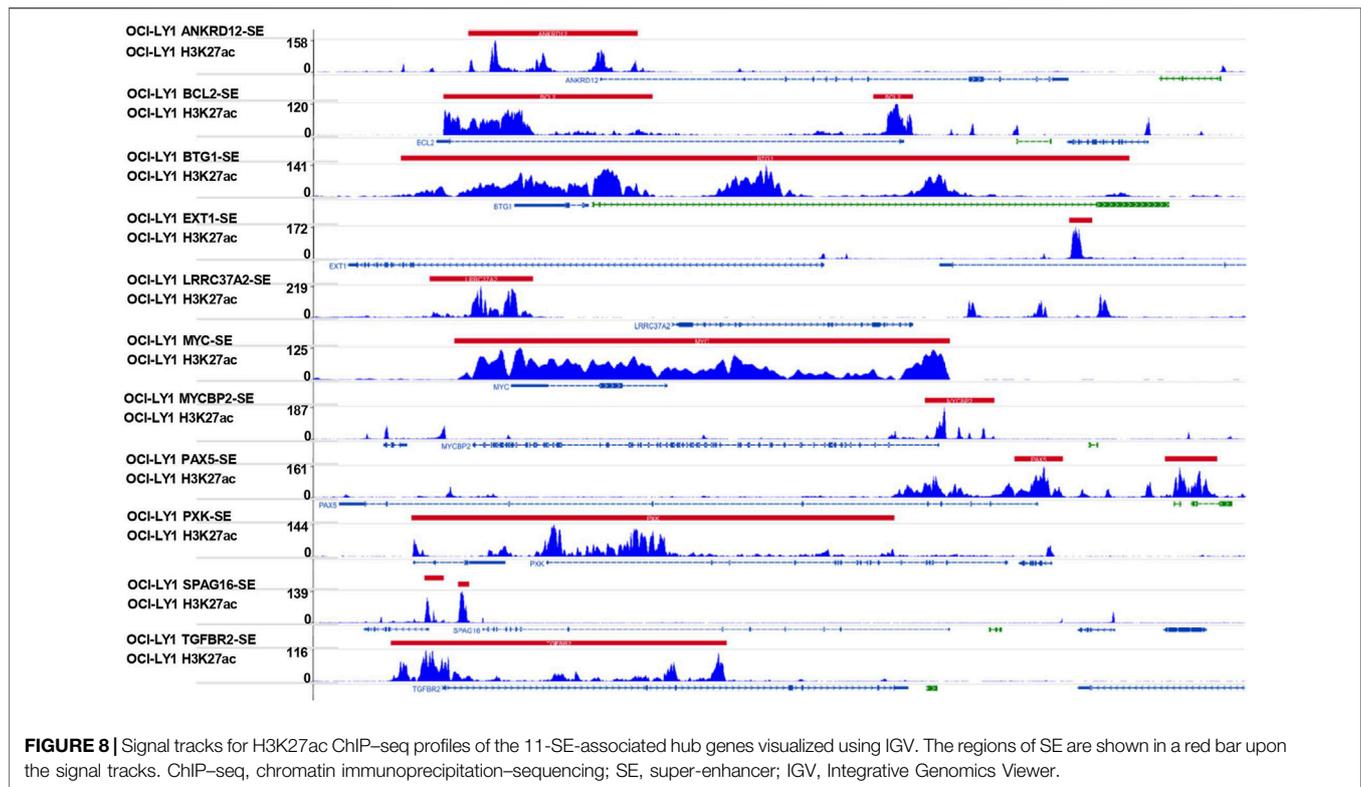
Chemotherapy Response With Super-Enhancer-Associated Genes Signature

In addition, we conducted a prediction analysis to evaluate the chemotherapy response in the low- and high-risk groups. Widely, all high-risk groups possessed higher estimated IC₅₀ for the different chemotherapeutics, which exactly proved the hypothesis that the high-risk group was not sensitive to the chemotherapy as the low-risk group (**Figure 7**). We took 12 chemotherapy drugs into account: bleomycin, vinorelbine, doxorubicin, gemcitabine, docetaxel, etoposide, cisplatin, bortezomib, vinblastine, vorinostat, and bexarotene.

In order to better improve the tricky problem, we additionally established an interactive network among these hub genes and chemicals to probe into the chemicals correlated to these genes by “NetworkAnalyst 3.0”. In total, six genes of these 11 hub genes interacted with JQ-1, a well-recognized SE inhibitor, which verified the regulating effect of SEs on these genes to some degree (**Supplementary Figure S2B**). In the end, we profiled the ChIP signal of H3K27ac-seq for these 11 genes (**Figure 8**). The predicted regions of SE were plotted as the red bar upon the signal tracks, and each of the predicted SEs located close to these 11 genes, suggesting that the SEs may play an influential role in the expression of the 11 genes. In addition, the SE inhibitor JQ1 may regulate the expression pattern in OCI-LY1 cells.

DISCUSSION

DLBCL is the most common lymphoma with high heterogeneity and invasiveness. It accounts for approximately one-third of the non-Hodgkin lymphoma, and plenty of patients suffer from insensitive to the typical treatment regimens (Lavacchi et al., 2021). Researchers aspired to identify optimal biomarkers and then establish various risk prediction models for predicting the survival rate, which can be used to improve the prognosis of DLBCL and contribute to personalized therapeutic decisions (Merdan et al., 2021). Enhancer is an important epigenetic regulatory element for DLBCL, which can determine the gene expression. Super-enhancers (SEs) are a large cluster of active enhancers critical for maintaining cell identity and driving the



expression of some oncogenes (Kai et al., 2021; Zhou et al., 2021). However, the previous studies had rarely constructed a risk prediction model based on SE-associated hub genes (Li, Duan and Hao, 2021). In this study, we succeeded in building a superior polygenic prognostic model by analyzing the data of the DLBCL patients from the GEO database, taking some clinical indicators into account as well, which was also rare in previous studies.

In the current study, Lasso penalized Cox regression was conducted to identify the candidate SE-associated genes, as the method has recently been prevalent in much research according to its ability to minimize overfitting (Zhu et al., 2019). In addition, we utilized univariate and multivariate Cox regression analyses to narrow the range of the selected genes. Then, we successfully constructed the gene risk score model for survival prediction. Moreover, we integrated the risk score and some other clinical indicators into developing the predictive nomogram and Cox proportion hazards model, which validated the predictive efficacy of the prognostic model. In our study, a total of 417 genes were filtered out by the Lasso penalized Cox regression. Subsequently, two best-fit values (λ_{\min} and λ_{1se}) were chosen, and then the 48-gene group of λ_{\min} and 16-gene group of λ_{1se} were initially screened out, respectively. Compared with the result of the AUC and Wilcoxon test, both models performed well. Furthermore, 11 genes were selected when statistically significant both in univariate and multivariate Cox regression analyses. To explore the influence of the 11 candidate genes on the OS and PFS of DLBCL patients, the patients were classified into two groups based on the 11-gene risk score model. The high-risk group had prominent inferior outcomes both in the K-M survival curve and

AUC. Combined with some clinical indicators, the univariate and multivariate Cox regression analyses and AUC were conducted to verify the independence of the risk score. Overall, the constructed 11-genes prognostic model demonstrated good predictive performance in the training dataset GSE31312 and the other two external validation sets, GSE10846 and GSE87371. In the training set, BCL2, SPAG16, LRRC37A2, TGFBR2, ANKRD12, and MYC appeared to be the risky factors, apparently upregulated, while PXX, BTG1, EXT1, MYCBP2, and PAX5 were downregulated in high-risk DLBCL patients.

BCL2 is considered an apoptosis suppressor gene. BCL2 is a cell survival protein that inhibits apoptosis by interacting with Bax, Bak, and other pro-apoptotic sensitizer proteins (Nabar et al., 2018) and also contributes to tumorigenesis by its promotion for survival, which already has a long and in-depth research history (Oltersdorf et al., 1998). Currently, many studies have shown a tight correlation between BCL2 expression levels in hematopoietic malignancies and drug resistance during therapy (Stewart et al., 2021). Previous studies have shown that DLBCL patients overexpressing the BCL2 protein may be strongly related to inferior survival and resistance to the standard therapy (de Jong et al., 2019). BCL2 is an important independent prognostic factor for DLBCL, consistent with our finding that the expression of BCL2 was significantly upregulated in the high-risk groups.

SPAG16 is a gene encoding sperm-associated antigen 16 that plays a role in sperm flagella function and motile ciliogenesis (Zhang et al., 2017; Alciaturi et al., 2019), correlated with the gene expression machinery of germ cells (Nagarkatti-Gude et al., 2011). Siliņa et al. (2011) have proposed that SPAG16 can be

a novel autoantibody target and serologic biomarker for cancers. Our study suggested that SPAG16 appears to be an independent predictor, but the specific mechanism to mediate tumorigenesis and its vulnerability to being an immunotherapeutic target remain unknown.

LRRC37A2 is a member of the LRRC37 gene family which is involved in the regulation of protein–ligand interactions and mapped to chromosome 17q21.31–q21.32 (Giannuzzi et al., 2013). Several studies suggested that LRRC37A2 is implicated in epilepsy, epileptic encephalopathy, and Parkinson's disease, while the effect on DLBCL has never been reported (Yao et al., 2021). In this study, high expression of LRRC37A2 corresponds with an inferior survival outcome that merits further exploitation.

TGFBR2 encodes a protein named transforming growth factor-beta (TGF- β) receptor type 2. This receptor can transduce signals into the intracellular environment, triggering various responses such as cell proliferation, differentiation, motility, and apoptosis (Biswas et al., 2008). Previous studies have shown that acquisition of TGFBR2 somatic mutation may increase the risk of various tumorigenesis and different diseases (Li et al., 2020). This is in line with our result that high-risk patients have upregulated expression of TGFBR2 compared with the low-risk group.

ANKRD12 encodes a 224 kDa nuclear protein ankyrin repeat domain 12, also called ANCO-2. It has been reported that ANCO proteins can inhibit the transcriptional activity of nuclear receptors involved in carcinogenesis (Bai et al., 2013). As per our result, ANKRD12 can predict survival outcomes for DLBCL patients independently, but further investigation is needed to validate.

MYC, well-known as a key transcriptional effector that modulates cellular proliferative and metabolism in stem cells (MacDonald et al., 2010), is also involved in the diverse cellular processes such as adhesion, apoptosis, and DNA damage response, playing a role in the oncogenic effect (Finley et al., 2015). There has been an explosion of molecular, cellular, and animal experiments to illuminate the effect of MYC in the initial development of neoplasms. As for DLBCL patients, MYC rearrangement (MYC-R) may forebode poor prognostic. Rosenwald A et al. have evaluated a large cohort suggesting the adverse prognostic impact of MYC-R and the significant therapeutic potential in DLBCL (Rosenwald et al., 2019). This statement is corroborated again by our study.

As for the protective prognostic factors in our study, PXX encoding protein is involved in ligand-induced internalization, synaptic transmits, and degradation of epidermal growth factor receptors associated with some autoimmunity diseases (Takeuchi et al., 2010). B-cell translocation gene 1 (BTG1) belongs to an anti-proliferative gene family, which regulates autophagy and the cell cycle and is also implicated in DNA repair and mRNA stability (Xue et al., 2021). BTG1 is a well-characterized tumor suppressor for both solid tumors and hematopoiesis and recently has been reported to have a novel role in genotoxic and integrated stress responses. It is evident that the expression level of BTG1 is regarded as a prognostic biomarker for diverse cancers (Yuniati et al., 2019). EXT1 gene produces the protein exostosin-1, which is found in the Golgi apparatus. This protein can modify newly

produced enzymes and some proteins, which are critical for metastasis of cancer cells (Francannet et al., 2001). MYCBP2 encodes a ubiquitin (Ub) E3 ligase, which is essential for neurodevelopment (Mabbitt et al., 2020). The antitumor effect of this gene has been identified in various cancers. PAX5 is a member of the paired-box family of transcriptional factors, exclusively expressed in the B-cell lineage (Berek et al., 2008). This gene correlates with a heterogeneous subset of B cell non-Hodgkin lymphoma (B-NHL). The expression level and bio function of Pax5 play a role in normal B lymphopoiesis and prevent tumorigenesis (Medvedovic et al., 2011). The antitumor effect of the above genes is consistent with this study; every gene act as an independent protective prognostic factor, upregulated in the low-risk group. However, the concrete bio function and corresponding molecular machinery of each gene remain a ripe area for further investigation.

Since BCL2, SPAG16, LRRC37A2, TGFBR2, ANKRD12, MYC, PXX, BTG1, EXT1, MYCBP2, and PAX5 are SE-associated genes, the roles of the genes SPAG16, LRRC37A2, ANKRD12, PXX, and BTG1 have not been illuminated in DLBCL, which merits further in-depth analysis in the wet laboratory. In addition, to further assess the efficacy of the 11-gene risk model, large-scale prospective cohorts are still needed.

CONCLUSION

In summary, we succeeded in constructing a novel and reliable SE-associated-gene signature that can effectively classify DLBCL patients into high-risk and low-risk groups and perform well in predicting the overall survival. The prediction model can be used as a biomarker of prognosis for DLBCL, which may be a potential therapeutic target and can assist clinicians in the treatment of DLBCL.

DATA AVAILABILITY STATEMENT

The datasets provided in this study can be obtained from online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

HX and ZZ designed the research; YL and JW analyzed and wrote the first draft of the manuscript; YJ, LL, WW, and YL contributed to data analysis and interpretation; HS, SL, YZ, and LT contributed to the collection and assembly of data. All authors agreed to submit to the current journal, approved the final version to be published, and agreed to be responsible for all aspects of the work.

FUNDING

This study was funded by the National Natural Science Foundation of China (grant number 81870150).

ACKNOWLEDGMENTS

The authors express their sincere thanks for the data on gene expression and corresponding clinical information, which was provided by the GEO database.

REFERENCES

- Abramson, J. S., and Shipp, M. A. (2005). Advances in the Biology and Therapy of Diffuse Large B-Cell Lymphoma: Moving toward a Molecularly Targeted Approach. *Blood* 106, 1164–1174. doi:10.1182/blood-2005-02-0687
- Alcietari, J., Anesetti, G., Irigoien, F., Skowronek, F., and Sapiro, R. (2019). Distribution of Sperm Antigen 6 (SPAG6) and 16 (SPAG16) in Mouse Ciliated and Non-ciliated Tissues. *J. Mol. Hist.* 50, 189–202. doi:10.1007/s10735-019-09817-z
- Armitage, J. O., Gascoyne, R. D., Lunning, M. A., and Cavalli, F. (2017). Non-Hodgkin Lymphoma. *Lancet* 390, 298–310. doi:10.1016/s0140-6736(16)32407-2
- Bai, R., Li, D., Shi, Z., Fang, X., Ge, W., and Zheng, S. (2013). Clinical Significance of Ankyrin Repeat Domain 12 Expression in Colorectal Cancer. *J. Exp. Clin. Cancer Res.* 32, 35. doi:10.1186/1756-9966-32-35
- Berek, C., Radbruch, A., and Schroeder, H. W. (2008). Clinical Immunology (Third Edition). Editors R. R. Rich and Mosby, 113–125. doi:10.1016/B978-0-323-04404-2.10008-9
- Biswas, S., Trobridge, P., Romero-Gallo, J., Billheimer, D., Myeroff, L. L., Willson, J. K. V., et al. (2008). Mutational Inactivation of TGFBR2 in Microsatellite Unstable Colon Cancer Arises from the Cooperation of Genomic Instability and the Clonal Outgrowth of Transforming Growth Factor β Resistant Cells. *Genes Chromosom. Cancer* 47, 95–106. doi:10.1002/gcc.20511
- Chapuy, B., McKeown, M. R., Lin, C. Y., Monti, S., Roemer, M. G. M., Qi, J., et al. (2013). Discovery and Characterization of Super-enhancer-associated Dependencies in Diffuse Large B Cell Lymphoma. *Cancer Cell* 24, 777–790. doi:10.1016/j.ccr.2013.11.003
- de Jong, M. R. W., Langendonk, M., Reitsma, B., Nijland, M., van den Berg, A., Ammatuna, E., et al. (2019). Heterogeneous Pattern of Dependence on Anti-apoptotic BCL-2 Family Proteins upon CHOP Treatment in Diffuse Large B-Cell Lymphoma. *Int. J. Mol. Sci.* 20, 6036. doi:10.3390/ijms20236036
- Finley, L. W. S., and Thompson, C. B. (2015). The Molecular Basis of Cancer (Fourth Edition). Editors J. Mendelsohn and W. B. Saunders, 191–208. doi:10.1016/B978-1-4557-4066-6.00013-5
- Francannet, C., Cohen-Tanugi, A., Le Merrer, M., Munnich, A., Bonaventure, J., and Legeai-Mallet, L. (2001). Genotype-phenotype Correlation in Hereditary Multiple Exostoses. *J. Med. Genet.* 38, 430–434. doi:10.1136/jmg.38.7.430
- Giannuzzi, G., Siswara, P., Malig, M., Marques-Bonet, T., Mullikin, J. C., Ventura, M., et al. (2013). Evolutionary Dynamism of the Primate LRRC37 Gene Family. *Genome Res.* 23, 46–59. doi:10.1101/gr.138842.112
- He, Y., Long, W., and Liu, Q. (2019). Targeting Super-enhancers as a Therapeutic Strategy for Cancer Treatment. *Front. Pharmacol.* 10, 361. doi:10.3389/fphar.2019.00361
- Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., et al. (2013). Super-enhancers in the Control of Cell Identity and Disease. *Cell* 155, 934–947. doi:10.1016/j.cell.2013.09.053
- Kai, Y., Li, B. E., Zhu, M., Li, G. Y., Chen, F., Han, Y., et al. (2021). Mapping the Evolving Landscape of Super-enhancers during Cell Differentiation. *Genome Biol.* 22, 269. doi:10.1186/s13059-021-02485-x
- Lavacchi, D., Landini, I., Perrone, G., Roviello, G., Mini, E., and Nobili, S. (2021). Pharmacogenetics in Diffuse Large B-Cell Lymphoma Treated with R-CHOP: Still an Unmet Challenge. *Pharmacol. Ther.* 229, 107924. doi:10.1016/j.pharmthera.2021.107924
- Li, S., Liu, M., Do, M. H., Chou, C., Stamatiades, E. G., Nixon, B. G., et al. (2020). Cancer Immunotherapy via Targeted TGF- β Signalling Blockade in TH Cells. *Nature* 587, 121–125. doi:10.1038/s41586-020-2850-3
- Li, X., Duan, Y., and Hao, Y. (2021). Identification of Super Enhancer-Associated Key Genes for Prognosis of Germinal Center B-Cell Type Diffuse Large B-Cell

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.827840/full#supplementary-material>

- Lymphoma by Integrated Analysis. *BMC Med. Genomics* 14, 69. doi:10.1186/s12920-021-00916-z
- Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., et al. (2013). Selective Inhibition of Tumor Oncogenes by Disruption of Super-enhancers. *Cell* 153, 320–334. doi:10.1016/j.cell.2013.03.036
- Mabbitt, P. D., Loreto, A., Déry, M.-A., Fletcher, A. J., Stanley, M., Pao, K.-C., et al. (2020). Structural Basis for RING-Cys-Relay E3 Ligase Activity and its Role in Axon Integrity. *Nat. Chem. Biol.* 16, 1227–1236. doi:10.1038/s41589-020-0598-6
- MacDonald, R. J., Swift, G. H., and Real, F. X. (2010). Transcriptional Control of Acinar Development and Homeostasis. *Prog. Mol. Biol. Transl. Sci.* 97, 1–40. doi:10.1016/b978-0-12-385233-5.00001-5
- Medvedovic, J., Ebert, A., Tagoh, H., and Busslinger, M. (2011). Pax5: a Master Regulator of B Cell Development and Leukemogenesis. *Adv. Immunol.* 111, 179–206. doi:10.1016/b978-0-12-385991-4.00005-2
- Merdan, S., Subramanian, K., Ayer, T., Van Weyenbergh, J., Chang, A., Koff, J. L., et al. (2021). Gene Expression Profiling-Based Risk Prediction and Profiles of Immune Infiltration in Diffuse Large B-Cell Lymphoma. *Blood Cancer J.* 11, 2. doi:10.1038/s41408-020-00404-0
- Nagarkatti-Gude, D. R., Jaimez, R., Henderson, S. C., Teves, M. E., Zhang, Z., and Strauss, J. F. (2011). Spag16, an Axonemal Central Apparatus Gene, Encodes a Male Germ Cell Nuclear Speckle Protein that Regulates SPAG16 mRNA Expression. *PLoS one* 6, e20625. doi:10.1371/journal.pone.0020625
- Nabar, N. R., Shi, C.-S., and Kehrl, J. H. (2018). *Immunology*. Editors M. A. Hayat (Academic Press). doi:10.1016/B978-0-12-809819-6.00006-X
- Oltersdorf, T., and Fritz, L. C. (1998). *Annual Reports in Medicinal Chemistry*. Editors J. A. Bristol (Academic Press), 253–262. doi:10.1016/S0065-7743(08)61090-8
- Parker, S. C. J., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., et al. (2013). Chromatin Stretch Enhancer States Drive Cell-specific Gene Regulation and Harbor Human Disease Risk Variants. *Proc. Natl. Acad. Sci. U.S.A.* 110, 17921–17926. doi:10.1073/pnas.1317023110
- Rosenwald, A., Bens, S., Advani, R., Barrans, S., Copie-Bergman, C., Elsensohn, M.-H., et al. (2019). Prognostic Significance of MYC Rearrangement and Translocation Partner in Diffuse Large B-Cell Lymphoma: A Study by the Lunenburg Lymphoma Biomarker Consortium. *J. Clin. Oncol.* 37, 3359–3368. doi:10.1200/jco.19.00743
- Siegel, R. L., Miller, K. D., and Jemal, A. (2017). Cancer Statistics, 2017. *CA a cancer J. Clin.* 67, 7–30. doi:10.3322/caac.21387
- Siliņa, K., Zayakin, P., Kalniņa, Z., Ivanova, L., Meistere, I., Endzeliņš, E., et al. (2011). Sperm-associated Antigens as Targets for Cancer Immunotherapy. *J. Immunother. Hagerst. Md* 34, 28–44. doi:10.1097/CJI.0b013e3181fb64fa
- Stewart, C. M., Michaud, L., Whiting, K., Nakajima, R., Nichols, C., De Frank, S., et al. (2021). Phase I/Ib Study of the Efficacy and Safety of Buparlisib and Ibrutinib Therapy in MCL, FL, and DLBCL with Serial Cell-free DNA Monitoring. *Clin. Cancer Res.* 28, 45–56. doi:10.1158/1078-0432.Ccr-21-2183
- Takeuchi, H., Takeuchi, T., Gao, J., Cantley, L. C., and Hirata, M. (2010). Characterization of PDK as a Protein Involved in Epidermal Growth Factor Receptor Trafficking. *Mol. Cell Biol.* 30, 1689–1702. doi:10.1128/mcb.01105-09
- Tibshirani, R. (1997). The Lasso Method for Variable Selection in the Cox Model. *Stat. Med.* 16, 385–395. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3
- Wang, X., Cairns, M. J., and Yan, J. (2019). Super-enhancers in Transcriptional Regulation and Genome Organization. *Nucleic Acids Res.* 47, 11481–11496. doi:10.1093/nar/gkz1038
- Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., et al. (2013). Master Transcription Factors and Mediator Establish

- Super-enhancers at Key Cell Identity Genes. *Cell* 153, 307–319. doi:10.1016/j.cell.2013.03.035
- Wu, Z., Guan, Q., Han, X., Liu, X., Li, L., Qiu, L., et al. (2021). A Novel Prognostic Signature Based on Immune-Related Genes of Diffuse Large B-Cell Lymphoma. *Aging* 13, 22947–22962. doi:10.18632/aging.203587
- Xue, K., Wu, J.-C., Li, X.-Y., Li, R., Zhang, Q.-L., Chang, J.-J., et al. (2021). Chidamide Triggers BTG1-Mediated Autophagy and Reverses the Chemotherapy Resistance in the Relapsed/refractory B-Cell Lymphoma. *Cell Death Dis.* 12, 900. doi:10.1038/s41419-021-04187-5
- Yao, S., Zhang, X., Zou, S.-C., Zhu, Y., Li, B., Kuang, W.-P., et al. (2021). A Transcriptome-wide Association Study Identifies Susceptibility Genes for Parkinson's Disease. *NPJ Park. Dis.* 7, 79. doi:10.1038/s41531-021-00221-7
- Yin, X., Xu, A., Fan, F., Huang, Z., Cheng, Q., Zhang, L., et al. (2019). Incidence and Mortality Trends and Risk Prediction Nomogram for Extranodal Diffuse Large B-Cell Lymphoma: An Analysis of the Surveillance, Epidemiology, and End Results Database. *Front. Oncol.* 9, 1198. doi:10.3389/fonc.2019.01198
- Younes, A. (2015). Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin: Seeing the Forest and the Trees. *J. Clin. Oncol.* 33, 2835–2836. doi:10.1200/jco.2015.61.9288
- Yuniati, L., Scheijen, B., Meer, L. T., and Leeuwen, F. N. (2019). Tumor Suppressors BTG1 and BTG2: Beyond Growth Control. *J. Cell. Physiol.* 234, 5379–5389. doi:10.1002/jcp.27407
- Zhang, F., Tapera, T. M., and Gou, J. (2018). Application of a New Dietary Pattern Analysis Method in Nutritional Epidemiology. *BMC Med. Res. Methodol.* 18, 119. doi:10.1186/s12874-018-0585-8
- Zhang, L., Liu, Y., Li, W., Zhang, Q., Li, Y., Liu, J., et al. (2017). Transcriptional Regulation of Human Sperm-Associated Antigen 16 Gene by S-SOX5. *BMC Mol. Biol.* 18, 2. doi:10.1186/s12867-017-0082-3
- Zhou, J., Wang, S., Nie, D., Lai, P., Li, Y., Li, Y., et al. (2021). Super-enhancer Landscape Reveals Leukemia Stem Cell Reliance on X-Box Binding Protein 1 as a Therapeutic Vulnerability. *Sci. Transl. Med.* 13, eabh3462. doi:10.1126/scitranslmed.abh3462
- Zhu, F. X., Wang, X. T., Zeng, H. Q., Yin, Z. H., and Ye, Z. Z. (2019). A Predicted Risk Score Based on the Expression of 16 Autophagy-related Genes for Multiple Myeloma Survival. *Oncol. Lett.* 18, 5310–5324. doi:10.3892/ol.2019.10881
- Conflict of Interest:** The 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.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Xu, Li, Jiang, Wang, Sun, Wu, LV, Liu, Zhai, Tian, Li and Zhao. 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.

GLOSSARY

ABC	activated B cell	IHC	immunohistochemical
AIC	Akaike information criterion	IPI	International Prognostic Index
AUC	area under the ROC	KEGG	Kyoto Encyclopedia of Genes and Genomes
B-NHL	B cell non-Hodgkin lymphoma	K–M	Kaplan–Meier
ChIP-seq	chromatin immunoprecipitation sequencing	Lasso	least absolute shrinkage and selection operator
CI	confidence interval	LDH	lactate dehydrogenase
C-index	concordance index	MYC-R	MYC rearrangement
CTD	Comparative Toxicogenomics Database	NHL	non-Hodgkin's lymphoma
DLBCL	diffuse large B-cell lymphoma	OS	overall survival
ECOG	Eastern Cooperative Oncology Group	PFS	progression-free survival
GCB	germinal center B cell	R-CHOP	rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone
GEO	Gene Expression Omnibus	ROC	receiver operating characteristic
GO	Gene Ontology	SE	super-enhancer
HR	hazard ratio	TE	typical-enhancer
IC₅₀	half-maximal inhibitory concentration	Ub	ubiquitin
IGV	Integrative Genomics Viewer	UC	unclassified