Immunophenotypic Landscape and Prognosis-Related mRNA Signature in Diffuse Large B Cell Lymphoma

Yanan Jiang
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Huimeng Sun
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Hong Xu
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Wenqi Wu
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Yangyang Lv
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Jinhuan Wang
Department of Oncology, Second Hospital of Tianjin Medical University, Tianjin, Institute of Urology Tianjin

Su Liu
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Yixin Zhai
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

LinYan Tian
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Yafei Wang
Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Zhigang Zhao (zzhao01@tmu.edu.cn)

Department of Hematology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin,

Research Article

Keywords: Diffuse large B cell lymphoma, immune phenotype, Immunity-H group, (LASSO)-Cox regression model

Posted Date: September 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-902940/v1

Abstract

Diffuse large B cell lymphoma (DLBCL) exhibits a tightly complexity immune landscape. In this study, we intended to identify different immune phenotype and to examine the immune related mRNA signature for clinical characteristic, therapeutic responsiveness as well as risk stratification and survival prediction in DLBCL. We separated 738 DLCBL patients into two robust subtypes with distinct immune features based on 28 immune cell types. GO analysis uncovered the concordant classification of two robust significant subtypes of DLBCL. Immunity-L group, associated with poorer outcomes, was characterized by significantly lower immune cell infiltration. Immunity-H group, associated with better outcomes, was characterized by significantly higher immune cell infiltration. Further, a prognostic gene signature related immune infiltration was established by a least absolute shrinkage and selection operator (LASSO)-Cox regression model. The comprehensive results showed that the high-risk group was correlated with lower immune infiltration, more aggressive phenotypes, lower overall survival and more sensitive to lenalidomide. In contrast, a low-risk group score was associated with higher immune infiltration, less aggressive phenotypes, better overall survival and more likely to benefit from PD-1/PD-L1 inhibitors. Finally, a nomogram comprised of the immune risk score and IPI score was verified to more accurately predict the overall survival of DLBCL than traditional clinical prediction models. Altogether, our data demonstrate the heterogeneity of immune patterns within DLBCL and deepen our molecular understanding of this tumor entity.

Introduction

Diffuse large B cell lymphoma (DLBCL) is the most common subtype of lymphoma in adults, accounting for approximately 30% of newly diagnosed non-Hodgkin's lymphomas (NHLs) annually[1]. Two distinct molecular subtypes, germinal center B cells (GCBs) and activated B cells (ABCs), were identified based on cell of origin (COO)[2]. The standard frontline therapy for DLBCL patients is cyclophosphamide in combination with doxorubicin, vincristine and prednisone (CHOP) with or without rituximab (R-CHOP), regardless of the subtype. However, one-third of patients will eventually fail R-CHOP treatment[3–4]. The International Prognostic Index (IPI) and Revised International Prognostic Index (R-IPI), which are based on five clinical characteristics (stage, age, lactic dehydrogenase (LDH) level, performance status, and extranodal sites involved), are common prognostic and predictive tools for DLBCL[5]. The limitations of IPI and R-IPI are their inability to predict individualized therapy for DLBCL patients. Therefore, great challenges exist regarding how to accurately predict survival and provide individualized treatment recommendations.

More recently, the critical role of the tumor microenvironment (TME) has been widely recognized in tumor initiation, proliferation and subsequent drug resistance, including in lymphoma[6–7]. Immune landscapes in lymphoma appear to be heterogeneous and can be categorized as "inflamed" and "noninflamed" or "immune excluded" lymphoma [8–10]. In recent years, the landscape for DLBCL immunotherapy strategies has become increasingly crowded. Novel therapies such as CAR-based cell therapies exhibit the most

promising results, particularly for patients who have demonstrated resistance to chemotherapy[11–13]. Similarly, immunomodulatory drugs such as lenalidomide also have a variety of effects on the immune system[14–16]. However, PD1/PD-L1 blockade seems to have unimpressive results[17–19]. The immune response to cancer is tightly correlated with the tumor microenvironment. A greater understanding of the types and roles of immune cells in the TME will enable us to identify candidate patients who will benefit from targeted immunotherapy and to discover biomarkers with important prognostic significance in DLBCL.

Here, we performed molecular subtyping of 738 DLBCL patients from the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases. We revealed that DLBCL patients could be classified into two robust subtypes with distinct immune features based on 28 immune cell types. Subsequently, we constructed an immune risk score based on the differentially expressed mRNA signature between these two immune subtypes. Our findings suggest that the immune risk score is a valuable biomarker for predicting the benefit of immunotherapy. In addition, it was an independent predictor of survival and can improve the prognostic value when combined with IPI.

Results

Molecular subtypes related to immune infiltration in DLBCL

This analysis included 1158 DLBCL patients with available survival data in 4 cohorts obtained from the TCGA and GEO databases. A total of 738 DLBCL patients, who were included in the TGCA database (n=47), GSE87371 (n=221) and GSE31312 (n=470), were considered the training cohort. GSE10846 (n=420) was considered the validation cohort. To further stratify the DLBCL patients into different immune infiltration groups, consensus cluster analysis was conducted. The optimal number of clusters was two, which was defined by CDF curves (Fig. 1A–C).

An unsupervised hierarchical cluster analysis of DLBCL patients based on immune infiltration-related genes separated DLBCL into high- and low-immune cell infiltration groups. The high immune cell infiltration group (Immunity-H, n=376) was enriched for almost all selected immune cell subtypes. Cases in the low immune cell infiltration group (Immunity-L, n=362) exhibited lower selected immune cell infiltration, except activated B cells (Fig. 1D-E). A comprehensive summary of the patient characteristics of the two groups is shown in Supplementary Table 1. Group membership within the two subtypes was associated with similar clinical characteristics, such as age, sex, stage, and COO. The Immunity-H group, with a longer overall survival, contained more patients with lower IPI scores. The Kaplan–Meier survival curves between the Immunity-H and Immunity-L groups are shown in Fig. 1F.

Immune Infiltration Defines a Biologically Distinct Subgroup Within DLBCL

To determine the accuracy of this grouping strategy, we determined the major histocompatibility complex (MHC), immune coinhibitor checkpoint (IAP), and immune costimulator checkpoint (ICP)-related gene expression in the two groups (Fig.2A-C). For most of the related genes, the bar chart analysis

showed that there was a significant positive correlation between the Immunity-H group and MHC, IAP and ICP. These results indicated that impaired immune surveillance of tumor cells is an escape mechanism in the Immunity-L group. Furthermore, we analyzed the functional context of these two subtypes by conducting Gene Ontology (GO). GO analysis based on the differentially expressed genes between the groups (Fig.2D, false discovery rate < 0.05) showed that the highly-expressed genes in Immunity-L group were mainly enriched in immune-related functions, such as activation of immune response, lymphocyte differentiation, neutrophil mediated immunity, positive regulation of MAPK cascade. We also analyzed somatic mutation data from the TCGA database to explore the difference in significantly mutated genes (SMGs) between these two subtypes. The SMG mutational profile of the DLBCL sample showed a distinct mutation ratio in IGLV3-1 (15/36 [41.7%]), BTG2 (13/36 [36.1%] and KMD2D (12/36 [33.3%]) (Supplementary Fig.1). The Immunity-L group showed more frequent mutations (101/169, [59.8%]). Because of the low sample size, we did not perform further detailed analysis between these two groups. In conclusion, these results identified extensive immune heterogeneity in DLBCL.

Identification of a classification-related prognostic signature for DLBCL

Considering the convenient application of the immune infiltration subtypes for prognostic prediction, we developed an immune risk score based on the differentially expressed genes between the Immunity-H and Immunity-L groups. According to the cutoff thresholds of |Log2 Fold Change|>0.5 and FDR < 0.05, a total of 489 mRNAs that were differentially expressed were obtained, of which 18 were upregulated and 471 were downregulated (Fig.3A). We then performed univariate analysis and LASSO analysis to select the gene set with the best prognostic value (Fig.3B-C). A 16-gene signature (ANTXR1, CD3D, TIMP1, FPR3, NID2, CTLA4, LPAR6, GPR183, LYZ, PTGDS, ITK, FBN1, FRMD6, PLAU, MICAL2, C1S) was identified and the risk score of each case was computed with the gene expression level and regression coefficient (Fig.3D). The Kaplan–Meier analysis of the 16 immune-related genes is shown in Supplementary Fig.2. Next, patients in training cohorts were assigned to a low-risk or high-risk group based on the median value of risk scores, which was used as the cutoff value. We also calculated the risk scores of patients in the validation cohort with the same coefficients to validate this signature. According to the distribution of the risk score and survival status, we detected the association between this signature and the proportion of deaths (p□0.05). A higher risk score was associated with a higher proportion of deaths (Fig.3E-F). Subsequently, we analyzed the expression of the 16 genes included in the signature in the high- and low-risk groups. To test the value of the risk score for DLBCL, we performed survival analyses (Kaplan–Meier test). The results showed that the survival time of the high-risk group was significantly shorter than that of the low-risk group in the training and validation cohorts (P < 0.05) (Fig.4A). To address the question of whether the risk score based on the newly established molecular immune signature is a unique feature of DLBCL, we tested its role in different IPI scores. We found that the survival time of the high-risk group was also significantly shorter than that of the low-risk group in both the IPI ^{high} and IPI ^{low} groups (P < 0.05) (Fig.4B). Moreover, all 16 genes were expressed at low levels in the high-risk group in both the training and validation cohorts (Fig.4C, Supplementary Fig.3)

The Prognostic Signature Is Related to Immune Infiltration and Clinical Characteristics

Notably, the risk score was significantly correlated with immune infiltration. A high-risk score was associated with lower immune infiltration (Fig. 5A). To further confirm the immune heterogeneity between these two risk groups, we examined the distribution of stromal and immune content of each group. The risk score was significantly negatively correlated with the immune score and ESTIMATE score(Fig.5B) (P 0.05). Patients in the low-risk group had a lower tumor purity, suggesting that low-risk group patients contained a higher number of immune cells (Fig.5B). Pearson's analysis was performed to determine the correlations between the risk score and the specific type of each immune cell. Except for activated B cells, a negative correlation was observed between the risk score and immune infiltration cells, such as central memory CD4 T cell and Myeloid-derived suppressor cells(MDSC) (Fig.5C, Supplementary Fig. 4).By stratified analysis, we found that the risk score showed excellent prognostic value among subjects with different baseline characteristics. There was a tightly correlation between the risk score and IPI, age (Fig.5D). For high-risk group, there was a higher proportion of patients in the IPI^{high}, age 60 years old groups. Stratified analysis further revealed that the risk score was lower in CR patients than in PR patients, but there were no statistically significant differences between PR and PD/SD patients (Fig. 5D). In addition, the risk score was higher in ABC-DLBCL patients than in GCB-DLBCL patients (Fig. 5D).

The prognostic signature is related to the response to chemotherapy and immunotherapy

We next performed a prediction analysis of response to therapy in the two-risk group by applying the "pRRophetic" method. There were obvious differences in the risk score and response to different drugs (Fig.6A-H). Patients in the high-risk group had a lower estimated IC50 than those in the low-risk groups for the following chemotherapy drugs: etoposide, gemcitabine, lenalidomide, vorinostat and obRO.3306. Conversely, patients in the low-risk group had a lower estimated IC50 than those in the high-risk groups for the following chemotherapy drugs: bortezomib and AZD8055. ($P<0.05$).

The functional analysis of differentially expressed genes

Next, we compared gene expression between low- and high-risk patients to determine the functional differences. A total of 308 genes were differentially expressed between the two groups. Compared with the low-risk group, 189 genes were downregulated and 119 were upregulated in the high-risk group (Supplementary Fig.5). The KEGG enrichment analysis indicated that many of these pathways were linked to the immune response in DLBCL, including the T cell receptor signaling pathway, B cell receptor signaling pathway, NF−kappa B signaling pathway, PD−L1 expression and PD−1 checkpoint pathway, which were significantly enriched in the low-risk group (Fig.7A-D). These results indicated that the signature derived from the immune infiltration classification could represent similar biological differences in DLBCL.

Prognostic value of the established signature

Uni- and multivariable analyses were performed to evaluate the correlation between the risk score and survival. Univariate Cox regression analysis showed that the risk score, age and IPI were both important

risk factors for DLBCL (HR>1, P<0.001) (Fig.8A). Additionally, multivariate Cox analysis revealed that the risk score and IPI were independent prognostic factors for DLBCL survival (HR>1, P<0.001) (Fig.8B). Both univariate and multivariate Cox regression analyses indicated that the HR value of the risk score was greater than that of the IPI, which is currently considered to be the best prognostic criterion for DLBCL. We next constructed a survival prediction nomogram comprising the risk score and IPI to predict 3- and 5-year OS (Fig.8C-D). Calibration curves for the probability of 3- and 5-year survival showed good agreement between predictions and observations, which indicated that the established nomogram was reliable in predicting the prognosis of DLBCL (Fig.8E). Next, an ROC curve was plotted to observe the predicted value of the nomogram with or without the risk score. Interestingly, we noticed that the nomogram risk score AUC was 0.775, which was better than the IPI AUC (0.714) and risk score AUC (0.725) (Fig.8F). These data suggest that the nomogram was a better predictor for a poor prognosis than IPI in DLBCL.

Discussion

A better understanding of the relationship between the TME and lymphoma cells is urgently needed to improve the efficacy of DLBCL therapy and prognosis. Here, we describe a new stratification model for DLBCL patients with different immune cell infiltration. To improve its application in the clinic, a prognostic signature was constructed based on the different RNA sequencing data from these two subgroups. Then, a novel risk score was developed based on the signature. The risk score was a promising biomarker to predict the prognosis, molecular and immune characteristics, and the immune benefit from ICI therapy in DLBCL.

The tumor microenvironment (TME) is an integrative component of most tumors including lymphomas. Its main cellular components are reactive lymphocytes, macrophages, fibroblasts, endothelial cells and dendritic cells as well as various cytokines, growth factors, and chemokines[20]. There is a much closer relationship between TME cells and lymphoma than other cancers because lymphoma cells themselves stem from lymphocytes[21]. In contrast to Hodgkin lymphoma (HL) and indolent B-cell lymphomas, the role of the microenvironment in aggressive B-cell lymphomas is still a matter of debate. DLBCL was regarded as “effacement” of the TME or “non-inflamed” lymphomas. Li, L. et al. showed the complexity of the TME in DLBCL; these cases with high T-cell infiltrates showed an overall better prognosis, while the prognosis of T-cell-rich DLBCL cases was worsened by a high proportion of PD-1-positive T-cells [22]. Xu-Monette, Z. Y. et al. demonstrated the negative impact of PD-1- and CD8-positive T-cells in cases of PD-L1 positivity of lymphoma cells, and a high proportion of PD-L1-positive TME macrophages was linked to inferior outcomes[23]. In contrast to PD-1 and PD-L1, CTLA4 was associated with better overall survival, although it was only expressed in a minority of cases. A study by Leivonen, S. K. et al.showed that patients with high T-cell infiltration had a better response to rituximab-based immunochemotherapy[24]. Given its complexity in DLBCL, our study utilized artificial intelligence and computational diagnostics to provide new insights into the role of the TME in DLBCL and to identify a prognostic biomarker for immunotherapy.

We first separated DLBCL patients into the Immunity-L and Immunity-H groups. The reliability of this classification was demonstrated by the obviously different gene expression levels of MHC, IAP and ICP. Patients in the Immunity-L group exhibited decreased expression at these sites. In addition, we observed a higher proportion of activated B cells in Immunity-L group tumors, while tumors in the Immunity-H group displayed higher levels of activated CD4+/CD8 + T cells, T follicular helper cells and natural killer T cells, which contribute to the immune control of the malignant clone and impact the efficacy of chemoimmunotherapy. In addition, GO analyses found that Immunity-H group tumors showed higher expression of genes involved in the immune pathway.

Then, we developed an immune-related genetic prognostic index based on the different subgroups and analyzed its role in discriminating different molecular and immune characteristics, responses to different single drugs and outcomes of DLBCL. The comprehensive results showed that the high-risk group was correlated with lower immune infiltration, more aggressive phenotypes, lower overall survival and more sensitive to lenalidomide. In contrast, a low-risk group score was associated with higher immune infiltration, less aggressive phenotypes, better overall survival and more likely to benefit from PD-1/PD-L1 inhibitors.

Next, we explored the relationship between the risk score and the response to different single drugs. Through a prediction analysis, we found that patients in the high-risk group were more sensitive to lenalidomide. It is an immunomodulatory agent that exerts tumor toxicities by altering the tumor microenvironment[14]. Moreover, lenalidomide has the ability to penetrate the blood–brain barrier. Combined with our results, patients in the high-risk group were also more sensitive to methotrexate and cytarabine, which are standard drugs for reducing the risk of CNS relapses. Thus, we speculate that the addition of lenalidomide treatment to chemotherapy may reduce the risk of CNS relapses in high-risk group patients, which needs to be validated in the future. Given the unimpressive efficacy of PD1/PD-L1 blockade in DLBCL, it is especially urgent to identify DLBCL patients who might benefit from anti-PD-1/PD-L1 immunotherapy. The KEGG enrichment analyses indicated that PD-L1 expression and the PD-1 checkpoint pathway were enriched in the low-risk group, which implying that patients in the low-risk group could benefit more from PD-1/PD-L1 inhibitors than high-risk patients. Together, our data provided predictive value for testing different immunotherapy in DLBCL patients.

Currently, IPIs are commonly used in clinical practice to predict the outcomes of DLBCL patients. However, the predictions are not as accurate as expected. Sixteen immune-related genes were used to establish our prognostic signature to improve the accuracy of 1-, 3-, and 5-year OS prediction. Meanwhile, we found that the risk score could separate the patients into two subgroups not only for patients in the IPI^{high} group but also in the IPI^{low} group. This suggested that DLBC patients in the same IPI group may have distinct pathogenic mechanisms and outcomes and may receive individualized therapy. Among these 16 genes, IL-2-inducible tyrosine kinase (ITK) belongs to the Tec family of kinases and is mainly expressed in T cells [25]. Some studies have focused on ITK as a key target for drug design. Recent studies have shown that CDC20 and PTGDS were able to predict overall survival (OS) in DLBCL, which was consistent with our study[26]. Wang, H.et al. reported that FBN1 promotes DLBCL cell migration by activating the Wnt/β-

catenin signaling pathway and regulating TIMP1[27]. Our studies found that these genes were related to overall survival. Further experiments are needed to verify the role of immune genes in DLBCL. Finally, ROC curves and calibration curves were used to construct a nomogram with precise accuracy for OS prediction. The predictive performance of the established prognostic signature was better for DLBCL patient outcomes than IPI. The nomogram, combined with the risk score and the IPI, would greatly improve the clinical prediction of DLBCL patient outcomes.

In conclusion, our study divided DLBCL patients into two subgroups based on immune infiltration. Then, we constructed an immune-related signature that was closely associated with prognosis, clinical characteristics, the immune response, and the tumor microenvironment. The prognostic signature revealed a significantly improved OS predictive ability compared to traditional prediction methods. However, this study was completed by using online datasets, and further clinical and basic experiments are needed to verify our results.

Materials And Methods

Expression and clinical datasets

This analysis included 1158 DLBCL patients with available survival data in 4 cohorts. Three microarray datasets, GSE87371, GSE31312, GSE10846, were derived from GEO database. TCGA-DLBC was extracted from the TCGA program. The “sva” package was used to remove batch effects. The R package “maftools” was used to identify the mutation status of TCGA-DLBC.738 DLBCL patients including TGCA database (n=47)GSE87371(n=221) and GSE31312 (n=470) was regarded as the training cohort. GSE10846 (n=420) was regarded as the validation cohort. Characteristics of the study patients for the DLBCL training cohorts are given in Supplementary Table 1.

Identification of gene expression-based subtypes

Subtype classification of DLBCL training cohort was based on Immune-Related Genes using the R package “ConsensusClusterPlus”. 80% item resampling, 100 resamplings, and a maximum evaluated K of 6 were selected for clustering. The cumulative distribution function (CDF) and consensus heatmap were used to assess the best K. Using the limma R package, we calculate differential gene expression based on RNA-seq counts data of two immune subgroups.

The Tumor Microenvironment Score of Immune subgroups

Gene expression levels of various genes including members of Immune complex of the major histocompatibility complex (MHC), immune costimulator checkpoint (ICP), and immune co-inhibitor checkpoint (IAP) was used to assess the differences between the two subtypes. Stromal Score, Immune Score, ESTIMATE Score were calculated using the ESTIMATE R package.

Gene Sets Enrichment Analysis

We used the R software package “clusterProfiler” for gene set enrichment analysis (GSEA) to study the biological process differences between immune infiltration subtypes. The results of Kyoto Encyclopedia of genes and genomes (KEGG) and Gene Ontology (go) are displayed through GSEA plot.

Mutated Genes Landscapes in AEG and Mutation Patterns in Two Immune subgroups

We extracted the Mutation characteristics of DLBCL data and compared them with the mutation database (COSMIC V2) by using the cosine similarity method (https://cancer.sanger.ac.uk/cosmic/).The R package “MutationalPatterns” and“Maftools” was used to compare the mutation signature of two immune subtypes.

Construction of Risk Assessment Model

We first get genes that were tightly correlated with prognosis by univariate Cox regression analysis($p<0.001$). Then the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm was used to select prognostic immune related signature and calculate variable coefficients with the “glmnet” package. Then, we calculated the risk score of each sample according to the following equation:

$$\text{Risk score} = Exp1 * Coe1 + Exp2 * Coe2 + Exp3 * Coe3 + \ldots Expi * Coei$$

where coei is equal to the gene coefficient, and Expi represents the gene expression level. ALL DLBCL patients were divided into high- and low-risk group based on the cut-off value.

Prognostic Value of the Prognostic Signature

Kaplan-Meier survival curves was used to compare the overall survival (OS) of DLBCL patients in the high- and low-risk groups. Receiver operating characteristic (ROC) analysis was used to compare the overall survival (OS) of DLBCL patients in the high- and low-risk groups. Receiver operating characteristic (ROC) curve analysis

was used to assessment the sensitivity and specificity of the Prognostic Signature in predicting OS.Univariate and Multivariate Analysis was used to assess whether clinical features and risk scores are prognostic risk factors.The R package “rms” was used to plot nomogram predict the of 3-, and 5-year OS DLBCL patients.Chemotherapeutic response in DLBCL patients was predicted by the R package “pRRophetic”.

Statistical Analysis

Kaplan-Meier analysis with log-rank tests was performed to compare differences in prognosis. $p < 0.05$ was considered statistically significant. Categorical variables were described as percentages. A chi-square test was performed to determine the difference of clinical and molecular parameters between two groups. Univariate and multivariate Cox regression analyses were conducted to determine factors with independent prognostic value. Pearson’s correlation analysis was carried out to compare correlations

between two groups. The Wilcoxon test was used to compare differences between two groups of nonnormally distributed data. R software (version 4.0.3), SPSS 22.0, and Prism 8 were used for statistical analysis and graphing.

Declarations

Data availability

The data involved in this study are freely available from the TCGA Hub at Xena datasets (https:// tcga. xenah ubs.net) and the GEO database with accession number of GSE87371GSE31312 and GSE10846 (http:// www. ncbi. nlm. nih. gov/geo/).

Author contributions

Y.N. J. and H.M. S. contributed to writing, reviewing, and editing the article. Conceptualization was by Z.G. Z. and Y.F. W. Y.N. J., W.Q. W. and Y.Y. L. contributed to Data Analysis and Interpretation, J.H. W., S.L.,Y.X. Z. and L.Y.T. contributed Collection and Assembly of Data, All authors contributed to the article and approved the submitted version.

Funding

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

Competing interests

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to ZG Z.

References

1. Bartlett, N. L. *et al.* Dose-Adjusted EPOCH-R Compared With R-CHOP as Frontline Therapy for Diffuse Large B-Cell Lymphoma: Clinical Outcomes of the Phase III Intergroup Trial Alliance/CALGB 50303. *J. Clin. Oncol.* **37**, 1790-1799, doi:10.1200/Jco.18.01994 (2019).
2. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511, doi:10.1038/35000501 (2000).
3. Oki, Y. *et al.* Double hit lymphoma: the MD Anderson Cancer Center clinical experience. *Br.J .Haematol.* **166**, 891-901, doi:10.1111/bjh.12982 (2014).
4. Pfreundschuh, M. *et al.* CHOP-like chemotherapy with or without rituximab in young patients with good-prognosis diffuse large-B-cell lymphoma: 6-year results of an open-label randomised study of the MabThera International Trial (MInT) Group. *Lancet Oncol.* **12**, 1013-1322. doi: 10.1016/S1470-2045(11)70235-2(2011).

5. Ziepert, M. *et al.* Standard International prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J. Clin. Oncol.* **28**, 2373-2380, doi:10.1200/JCO.2009.26.2493 (2010).
6. Khurana, A. & Ansell, S. M. Role of Microenvironment in Non-Hodgkin Lymphoma: Understanding the Composition and Biology. *Cancer J.* **26**, 206-216, doi:10.1097/PPO.0000000000000446 (2020).
7. Merryman, R. W., Houot, R., Armand, P. & Jacobson, C. Immune and Cell Therapy in Non-Hodgkin Lymphoma. *Cancer J.* **26**, 269-277, doi:10.1097/PPO.0000000000000445 (2020).
8. Menter, T., Tzankov, A. & Dirnhofer, S. The tumor microenvironment of lymphomas: Insights into the potential role and modes of actions of checkpoint inhibitors. *Hematol. Oncol.* **39**, 3-10, doi:10.1002/hon.2821 (2021).
9. El Hussein, S. A.-O., Shaw, K. R. M. & Vega, F. A.-O. X. Evolving insights into the genomic complexity and immune landscape of diffuse large B-cell lymphoma: opportunities for novel biomarkers. *Mod. Pathol.* **33**,2422-2436.doi: 10.1038/s41379-020-0616-y (2020).
10. Ciavarella, S. *et al.* Dissection of DLBCL microenvironment provides a gene expression-based predictor of survival applicable to formalin-fixed paraffin-embedded tissue. *Ann. Oncol. 29,*2363-2370. doi: 10.1093/annonc/mdy450(2018).
11. Neelapu, S. S. *et al.* Axicabtagene Ciloleucel CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. *N. Engl. J. Med.* **377**, 2531-2544, doi:10.1056/NEJMoa1707447 (2017).
12. Locke, F. L. *et al.* Long-term safety and activity of axicabtagene ciloleucel in refractory large B-cell lymphoma (ZUMA-1): a single-arm, multicentre, phase 1-2 trial. *Lancet Oncol.* **20**, 31-42, doi:10.1016/S1470-2045(18)30864-7 (2019).
13. Chavez JC, Locke FL. CAR T cell therapy for B-cell lymphomas. *Best. Pract. Res. Clin. Haematol.* **31**, 135-146, doi: 10.1016/j.beha.2018.04.001(2018)
14. Garciaz, S., Coso, D., Schiano de Colella, J. M. & Bouabdallah, R. Lenalidomide for the treatment of B-cell lymphoma. *Expert Opin. Investig. Drugs.* **25**, 1103-1116, doi:10.1080/13543784.2016.1208170 (2016).
15. Hernandez-Ilizaliturri, F. J. *et al.* Higher Response to Lenalidomide in Relapsed/Refractory Diffuse Large B-Cell Lymphoma in Nongerminal Center B-Cell-Like Than in Germinal Center B-Cell-Like Phenotype. *Cancer* **117**, 5058-5066, doi:10.1002/cncr.26135 (2011)
16. Feldman, T. *et al.* Addition of lenalidomide to rituximab, ifosfamide, carboplatin, etoposide (RICER) in first-relapse/primary refractory diffuse large B-cell lymphoma. *Br. J. Haematol.* **166**, 77-83, doi:10.1111/bjh.12846 (2014).
17. Sheikh S, Kuruvilla J. Pembrolizumab for the treatment of diffuse large B-cell lymphoma. *Expert Opin. Biol. Ther.* 19,1119-1126. doi: 10.1080/14712598.2019.1659777(2019)
18. Tomassetti, S., Chen, R. & Dandapani, S. The role of pembrolizumab in relapsed/refractory primary mediastinal large B-cell lymphoma. *Ther. Adv. Hematol.* **10**, 2040620719841591, doi:10.1177/2040620719841591 (2019).

19. Frigault, M. J. *et al.* PD-1 blockade for diffuse large B-cell lymphoma after autologous stem cell transplantation. *Blood Adv.* **4**, 122-126, doi:10.1182/bloodadvances.2019000784 (2020).
20. Scott, D. W. & Gascoyne, R. D. The tumour microenvironment in B cell lymphomas. *Nat Rev Cancer* **14**, 517-534, doi:10.1038/nrc3774 (2014).
21. Menter, T., Tzankov, A. & Dirnhofer, S. The tumor microenvironment of lymphomas: Insights into the potential role and modes of actions of checkpoint inhibitors. *Hematol Oncol.* **39**, 3-10, doi:10.1002/hon.2821 (2021).
22. Li, L. et al. PD-1/PD-L1 expression and interaction by automated quantitative immunofluorescent analysis show adverse prognostic impact in patients with diffuse large B-cell lymphoma having T-cell infiltration: a study from the International DLBCL Consortium Program. *Mod. Pathol.***32**, 741-754, doi:10.1038/s41379-018-0193-5 (2019).
23. Xu-Monette, Z. Y. et al. Immune Profiling and Quantitative Analysis Decipher the Clinical Role of Immune-Checkpoint Expression in the Tumor Immune Microenvironment of DLBCL. *Cancer Immunol. Res.* **7**, 644-657, doi:10.1158/2326-6066.CIR-18-0439 (2019).
24. Leivonen, S. K. et al. T-cell inflamed tumor microenvironment predicts favorable prognosis in primary testicular lymphoma. *Haematologica* **104**, 338-346, doi:10.3324/haematol.2018.200105 (2019).
25. Lechner, K. S., Neurath, M. F. & Weigmann, B. Role of the IL-2 inducible tyrosine kinase ITK and its inhibitors in disease pathogenesis. *J. Mol. Med (Berl).* **98**, 1385-1395, doi:10.1007/s00109-020-01958-z (2020).
26. Sun, C. *et al.* Gene expression profiles analysis identifies a novel two-gene signature to predict overall survival in diffuse large B-cell lymphoma. *Biosci. Rep.* **39**, doi:10.1042/BSR20181293 (2019).
27. Wang, H., Liu, Z. & Zhang, G. FBN1 promotes DLBCL cell migration by activating the Wnt/beta-catenin signaling pathway and regulating TIMP1. *Am. J. Transl. Res.* **12**, 7340-7353 (2020).

Figures

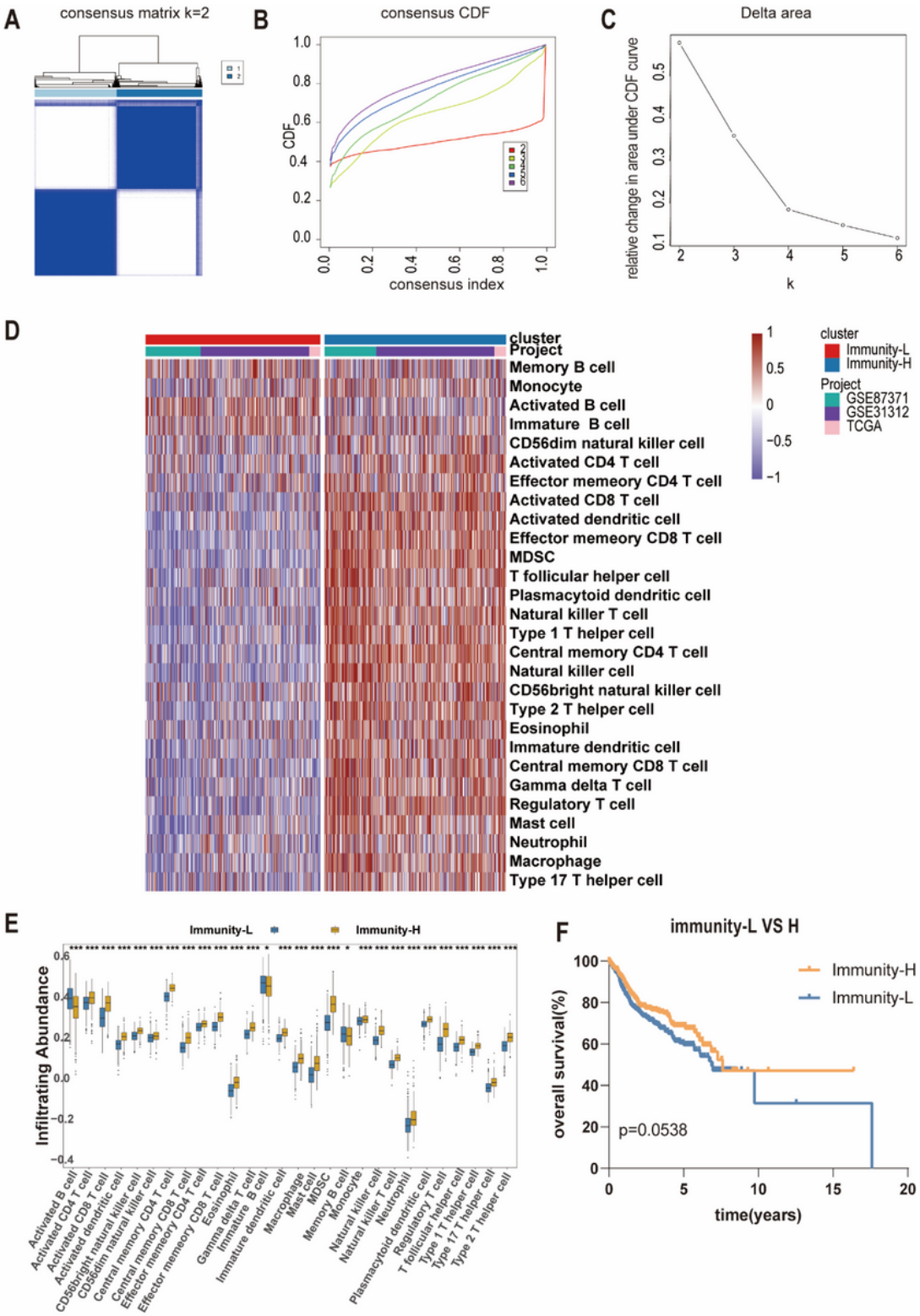

Figure 1

Establishment of a novel molecular immune subgroup (A–C) The optimal number of clusters (K=2) was determined from cumulative distribution function (CDF) curves, and the classification effect is the best. (D-E) Immune cell infiltration subtype is shown in the form of a heatmap, with the 738 patient samples shown as columns. (F) The Kaplan–Meier survival curves between Immunity-H and Immunity-L group.

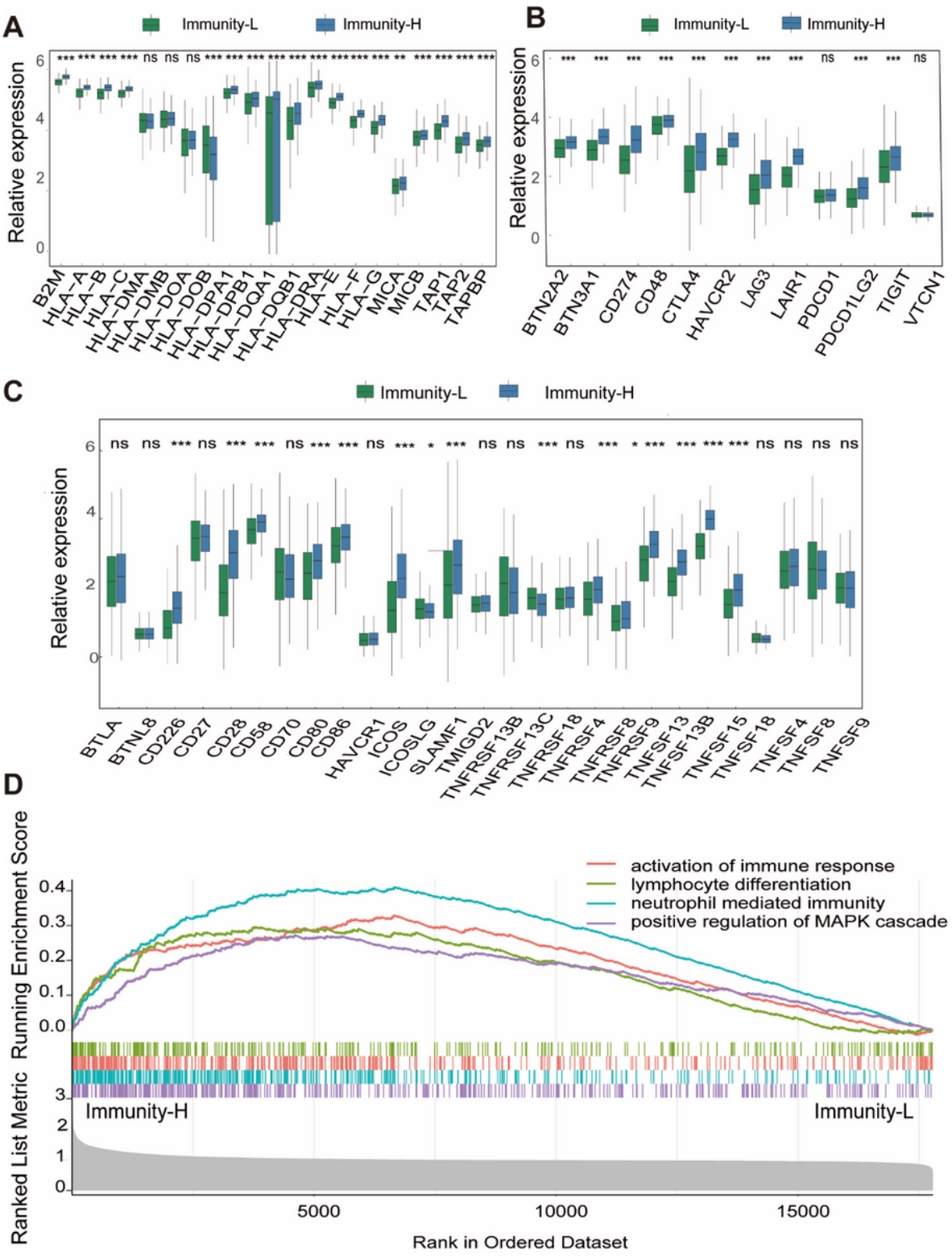

Figure 2

Immune Infiltration Defines a Biologically Distinct Subgroup Within DLBCL (A-C) The gene expression level of the gene set, including major histocompatibility complex (MHC), immune co-inhibitor checkpoints (IAP), and immune co-stimulator checkpoints (ICP) were all significantly different in the two immune infiltration subtypes ($P<0.05$). (D) GO enrichment analysis of the differentially expressed genes in the two immune infiltration subtypes.

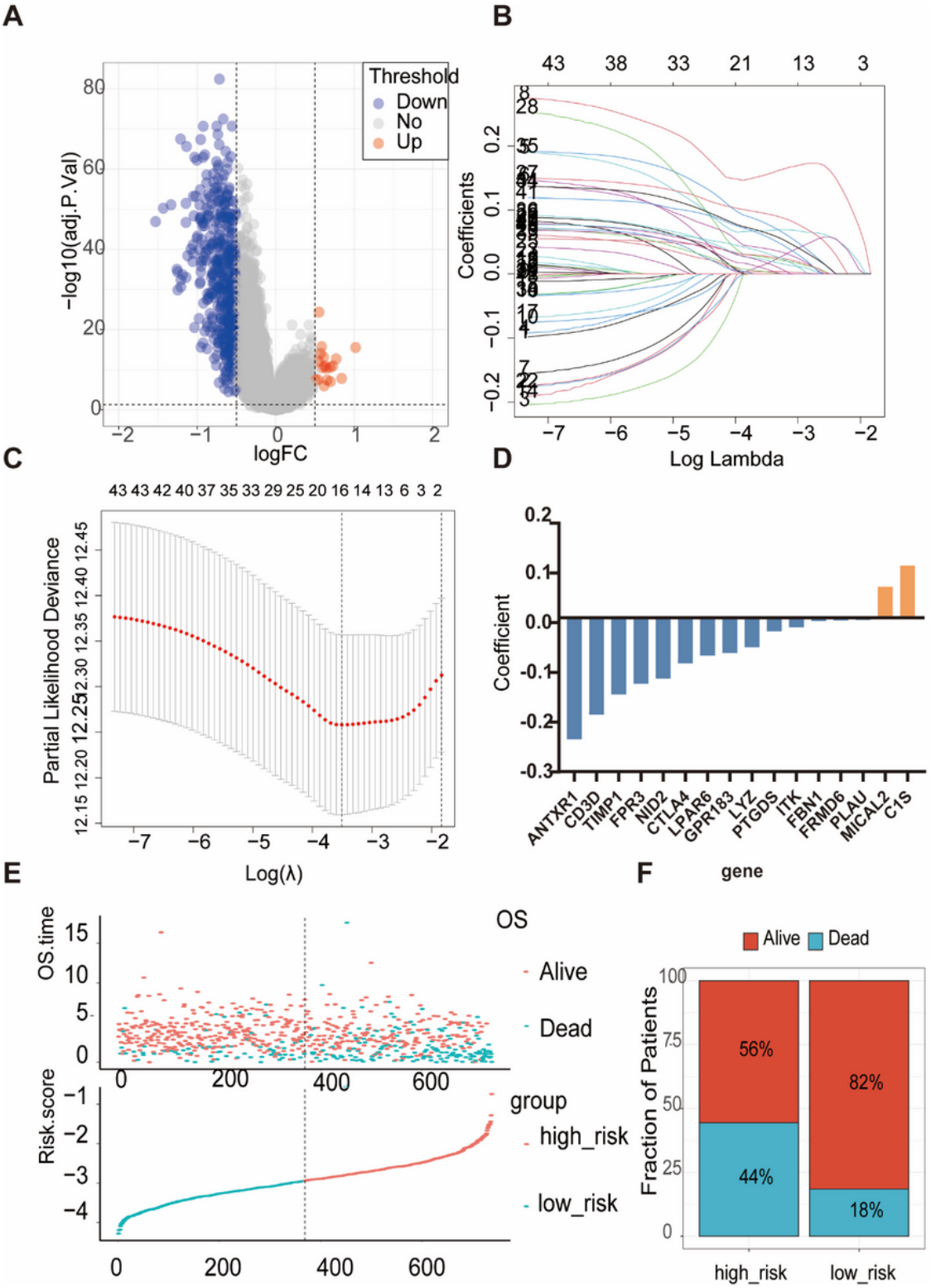

Figure 3

Construction of a prognostic signature for DLBCL (A) The volcano plot showed that mRNAs were up-regulated and down-regulated between the two immune infiltration subtypes. Each red dot showed an up-regulated mRNA, and each blue dot shows a downregulated lncRNA (|Log2 Fold Change| >0.5 and FDR < 0.05). (B-C) Sixteen genes were selected to construct the immune-related mRNA prognostic signature by

the LASSO regression algorithm. (D) The coefficients of the 16 genes. (E-F) The distribution of the risk Score and survival status of patients.

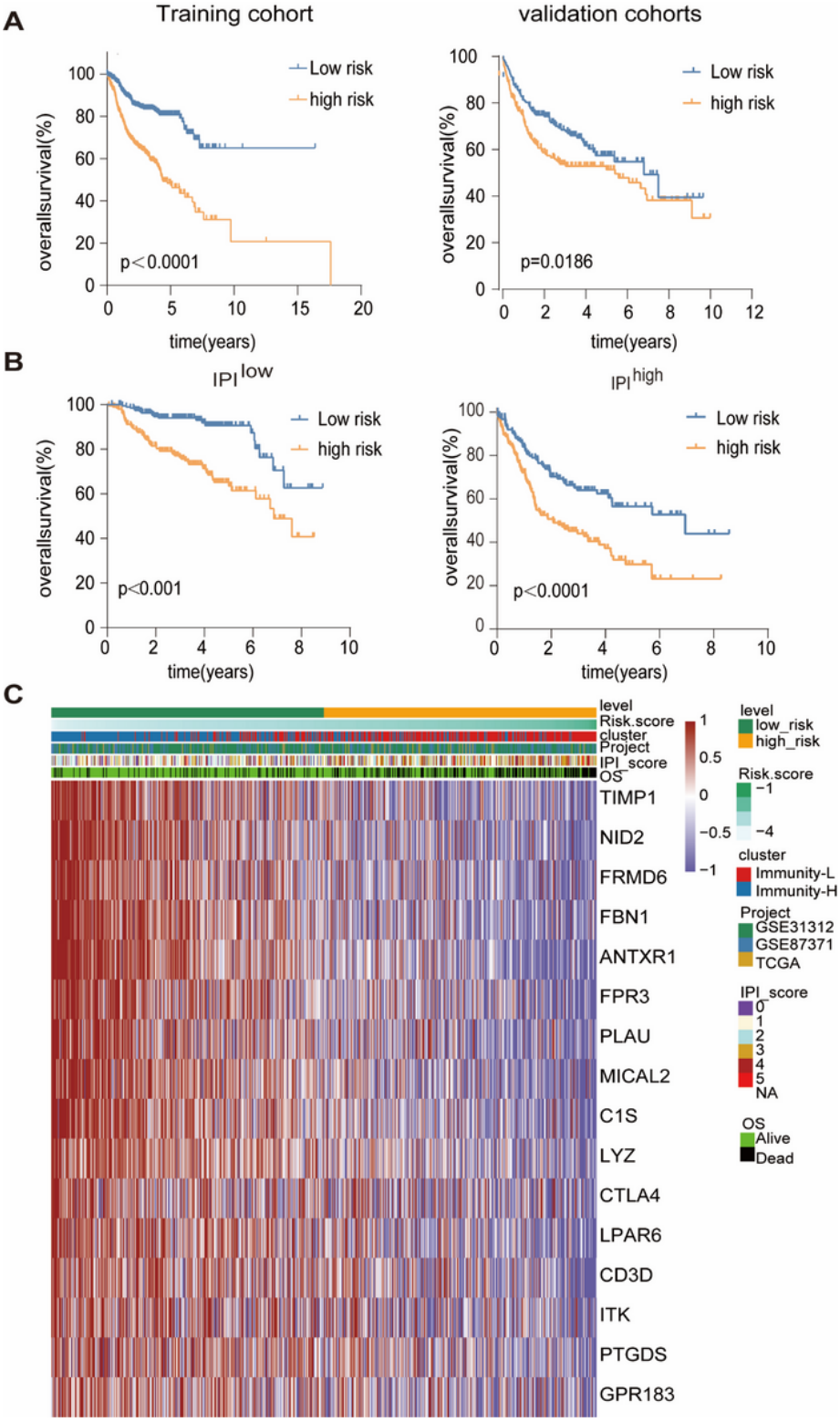

Figure 4

Identification of a classification-related prognostic signature for DLBCL (A) Kaplan–Meier analysis of the high versus low risk score in the training and validation cohort. (B) Kaplan–Meier analysis of the high

versus low risk score in the IPI low and IPI high group in the training cohort. (C) The differential expression of the 16 genes in the high- and low-risk groups in the training cohort.

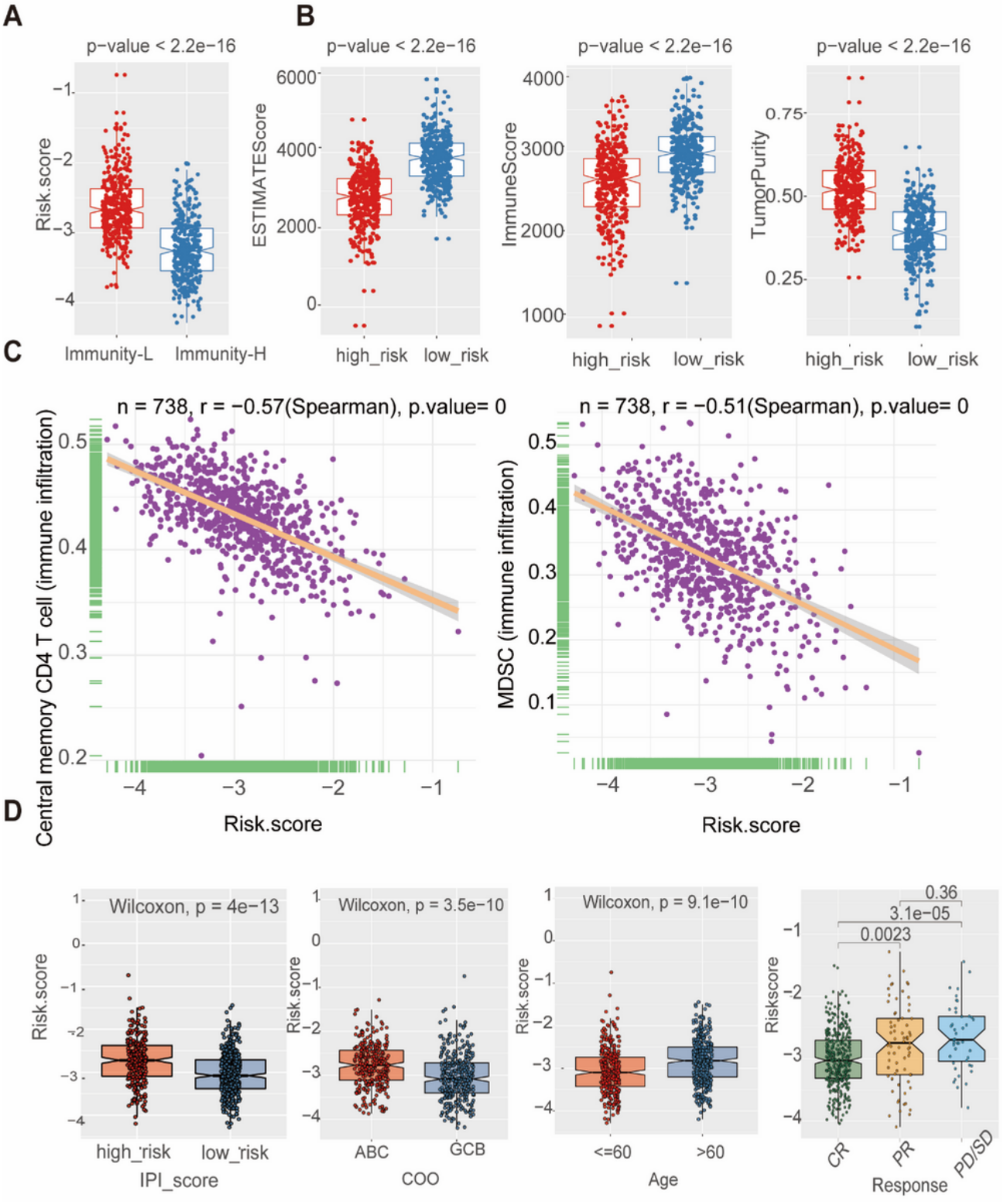

Figure 5

The prognostic signature was related to immune cells and clinical characteristics (A) The negative correlation between the risk Score and immune infiltration. (B) The boxplot showed that there was a statistical difference in Immune Score, ESTIMATE Score, and Tumor Purity between the two immune

infiltration subtypes (P<0.05). (C)Assessment of immune cell infiltration abundance by the risk Score. There were negative correlation between the risk Score and central memory CD4 T cell □MDSC. (D) The box plot shows the risk score for indicated subgroups, with a significantly low risk score for IPI low□age ≤60 years old , GCB , the CR group as compared with other groups.

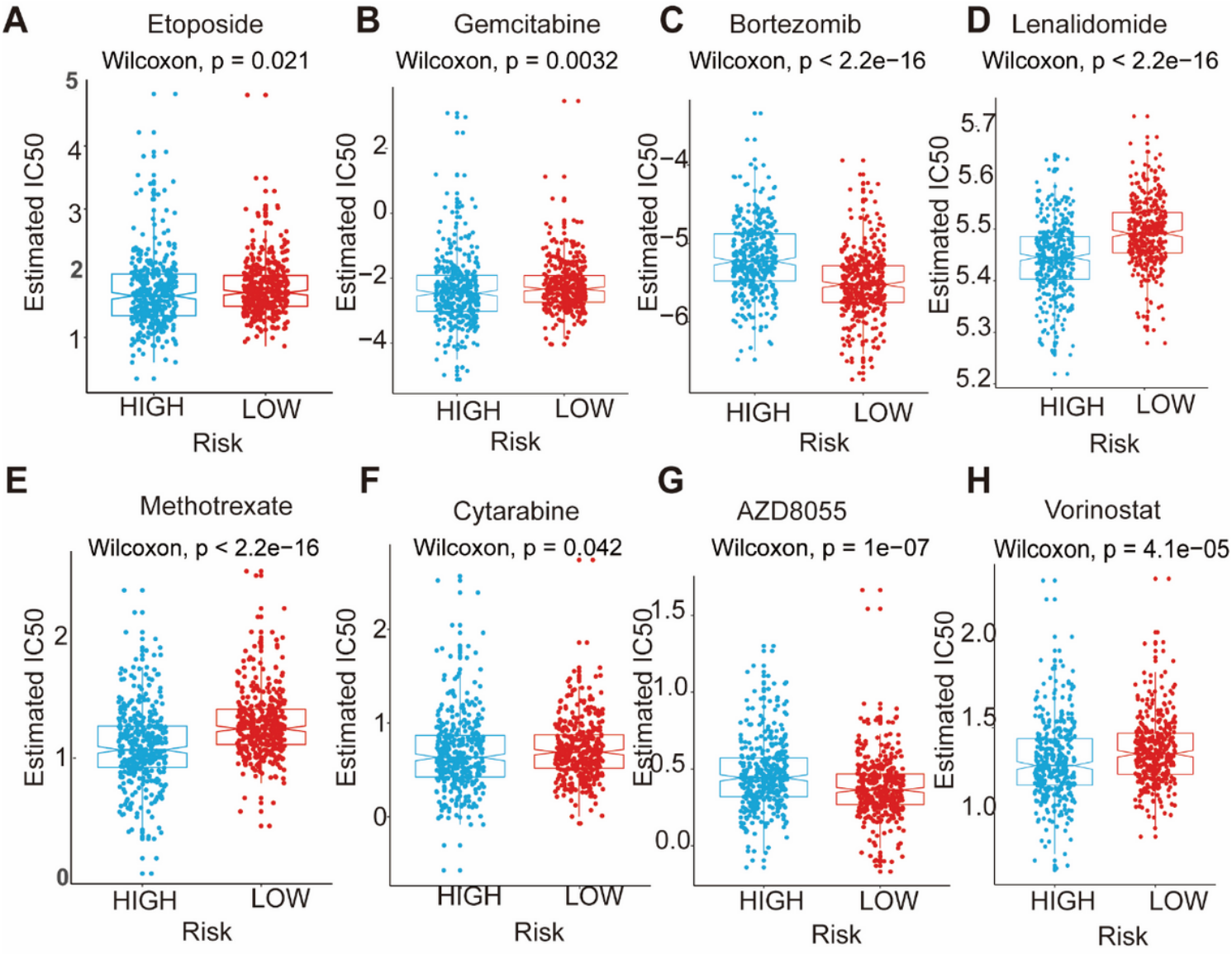

Figure 6

The IC50s of chemotherapeutic agents with mRNA signature. (A) Etoposide, (B) Gemcitabine, (C) Bortezomib, (D) Lenalidomide, (E) Methotrexate, (F) Cytarabine, (G) AZD8055, (H) Vorinostat.

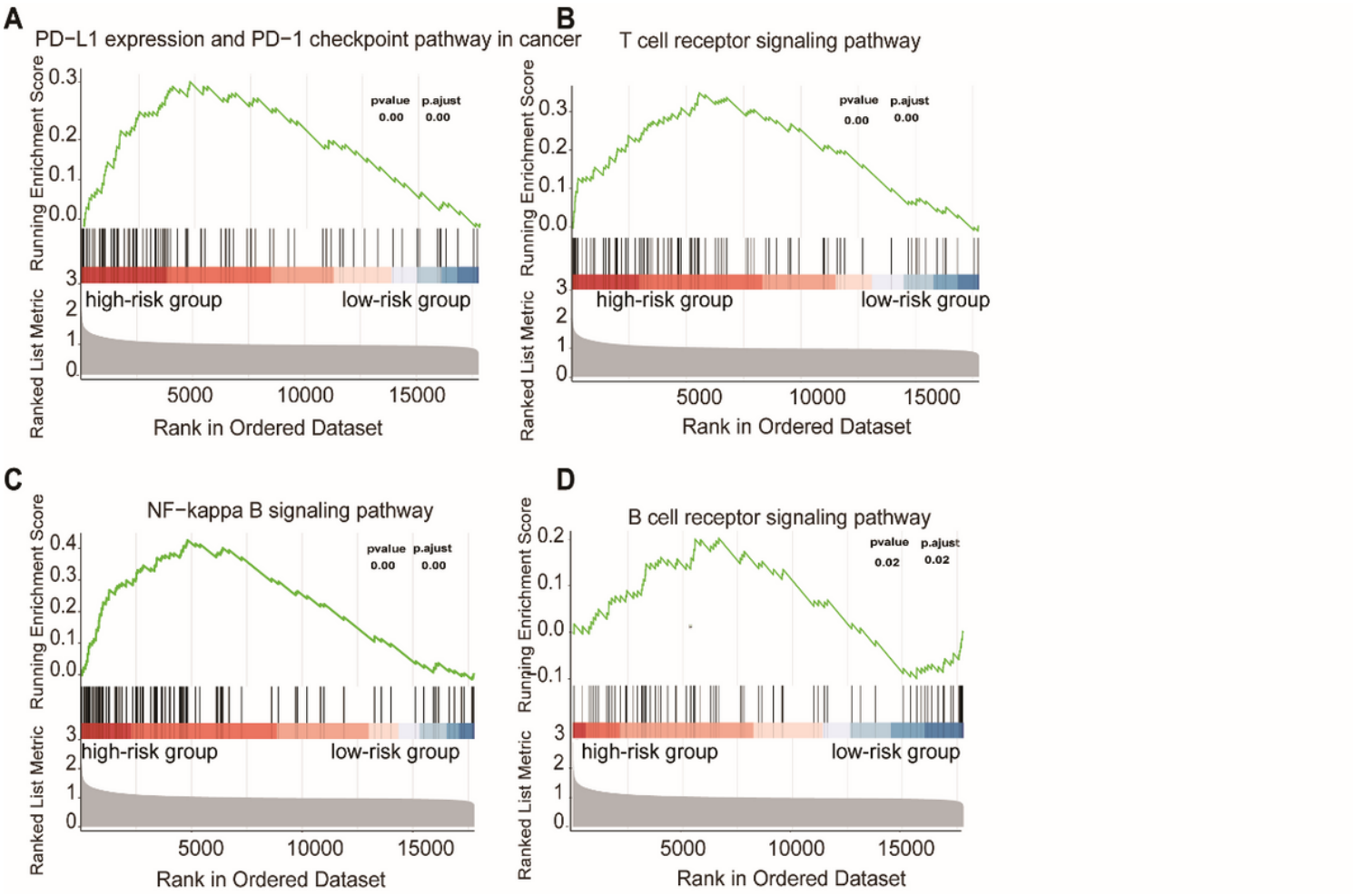

Figure 7

Functional annotation of the two risk subtypes. (A-D) Enriched gene pathways/functions in distinct risk groups from the DLBCL cohort were assessed by using the KEGG algorithm.

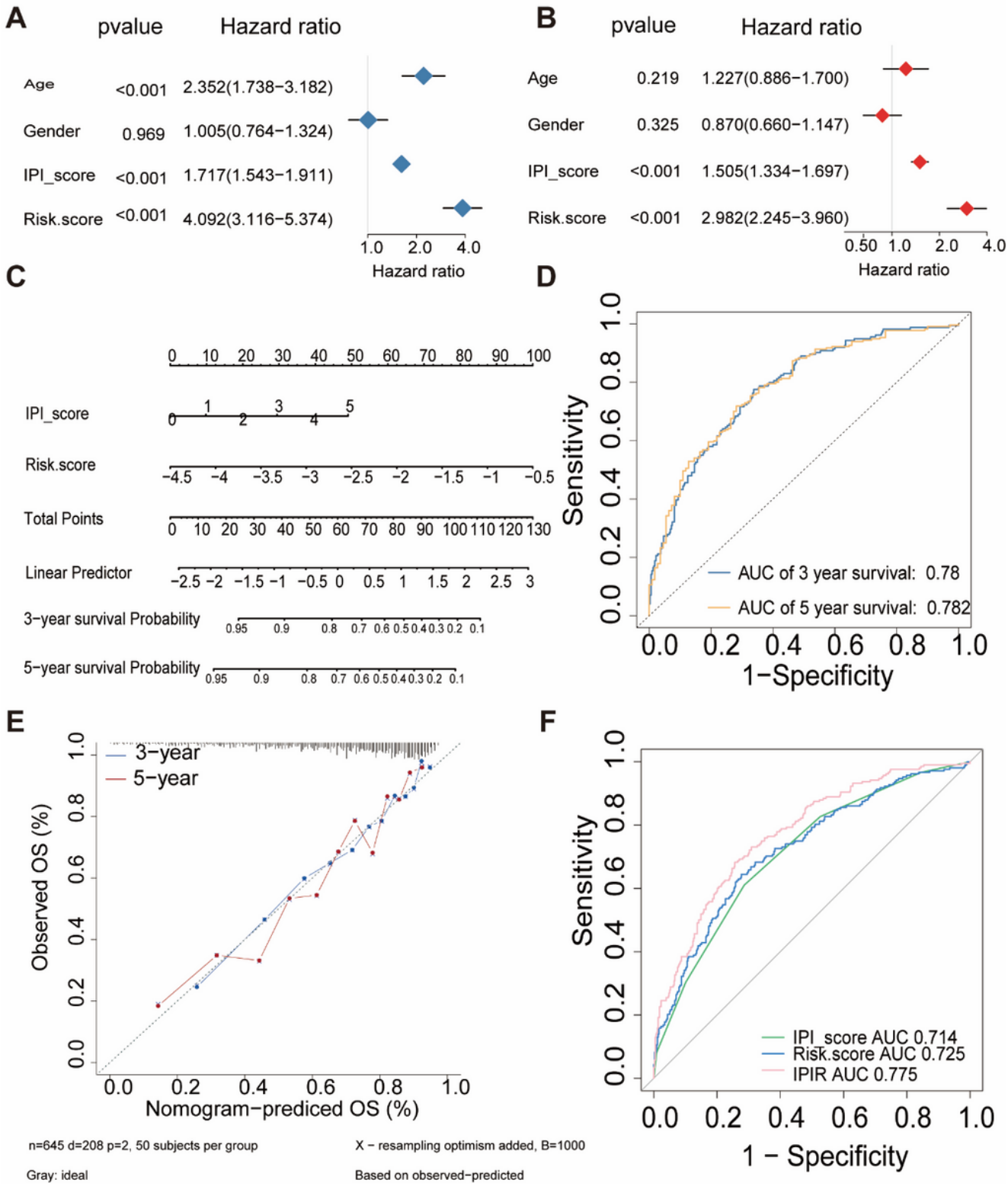

Figure 8

Prognostic value of the established signature (A-B) Univariate and multivariate Cox regression analyses of the association between clinicopathological factors and OS of DLBCL patients. (C) The nomogram of IPI score and the risk score. (D) ROC curve analysis for OS prediction by the nomogram. (E) Calibration curve of the nomogram for predicting the OS rates of DLBCL patients (F) ROC curves and AUCs for evaluating the prediction accuracy of the nomogram(IPIR), immune risk score and IPI score.

Supplementary Files

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

- SupplementaryFigureS1.pdf
- SupplementaryFigureS2.pdf
- SupplementaryFigureS3.pdf
- SupplementaryFigureS4.pdf
- SupplementaryFigureS5.pdf
- Supplementaryfiguresandtablelegends.docx
- SupplementaryTable1.docx