An IPI based immune prognostic model for diffuse large B-cell lymphoma

Shidai Mu[1,#,*], Deyao Shi[2,#], Lisha Ai[1], Fengjuan Fan[1], Fei Peng[1], Chunyan Sun[1*], Yu Hu[1*]

1 Institution of Hematology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430022, China

2 Department of Orthopaedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430022, China

[#] equally contributed

[*] Correspondence to: Shidai Mu, Chunyan Sun and Yu Hu

Shidai Mu: shidaimu@gmail.com

Deyao Shi: shideyao@hust.edu.cn

Lisha Ai: ailisha86@126.com

Fengjuan Fan: 4291125@qq.com

Fei Peng: pengfei0118@foxmail.com

Chunyan Sun: suncy0618@163.com

Yu Hu: dr_huyu@126.com

Abstract

Background: International Prognostic Index (IPI) was widely used to better discriminate prognosis of patients with diffuse large B-cell lymphoma (DLBCL) . However, there is a significant need to identify novel valuable biomarkers in the context of targeted therapies, such as immune checkpoint blockade (ICB) therapy.

Methods: Gene expression data and clinical information of DLBCL were obtained from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) datasets. 371 immune-related hub genes in DLBCL patients with different IPI levels were identified by weighted gene co-expression network analysis (WGCNA), and 8 genes were selected to construct an IPI-based immune prognostic model (IPI-IPM). Afterward, the genetic, somatic mutational and molecular profiles of IPI-IPM subgroups were analyzed, as well as the potential clinical response of ICB in different IPI-IPM subgroups.

Results: The IPI-IPM was constructed based on the expression of CMBL, TLCD3B, SYNDIG1, ESM1, EPHA3, HUNK, PTX3 and IL12A, where high-risk patients had shorter overall survival (OS) than low-risk patients, consistent with the results in the GEO cohorts. The comprehensive results showed that high IPI-IPM risk scores were correlated with immune-related signaling pathways, high KMT2D and CD79B mutation rates, as well as up-regulation of inhibitory immune checkpoints including PD-L1, BTLA and SIGLEC7, indicating more potential response to ICB therapy.

Conclusion: The IPI-IPM has independent prognostic significance for DLBCL patients, which provides an immunological perspective to elucidate the mechanisms on tumor progression, also sheds a light on developing immunotherapy for DLBCL.

Keywords: DLBCL, Immune prognostic model, Nomogram, Immunotherapy

Introduction

Diffuse large B-cell lymphoma (DLBCL) accounts for about 40% of non-Hodgkin B-cell lymphoma (NHL), with an annual incidence rate of over 100,000 cases worldwide[1, 2]. Although current frontline DLBCL therapy (the standard R-CHOP chemotherapy regimen) is associated with a high complete response rates of 70–80%, 10% to 15% of DLBCL patients are refractory, and almost 40% of cases experience relapse within 2–3 years after initial response[3, 4]. With the development of high-throughput technologies, germinal center B-cell-like (GCB) and activated B-cell-like (ABC) DLBCL subtypes were identified by gene expression profiling (GEP) based on cell-of-origin (COO) classification[5-7]. More recently, several key cytogenetic alterations including mutations, somatic copy number alterations (SCNA) and structural variants (SV), have been shown to classify distinct genetic subtypes within the COO subgroups, providing insights into disease heterogenous pathogenesis and candidate treatment targets[1, 3, 7-9]. Several prognostic factors including COO and the International Prognostic Index (IPI) have already been identified in the rituximab era, which still need to be further investigated in the context of targeted therapies[7, 9-12]. Therefore, there is an urgent need to explore potential molecular mechanism and identify more key biomarkers and therapeutic targets.

Accumulating evidence has shed light on the prognostic role of tumor microenvironments (TME) in immune checkpoint blockade therapy (ICB), which was mostly composed of a variety of immune cells (T-, NK-, and B-cells as well as macrophages) and stroma (blood vessels and extra-cellular matrix)[13-16]. Kotlov et al. characterized the DLBCL TME into 4 distinct microenvironment compositions including “germinal center-like” (GC), “mesenchymal” (MS), “inflammatory” (IN) and “depleted” (DP) form, which associate with distinct clinical behavior and provide novel potential targets for innovative therapeutic interventions[11].

In this study, we identified immune-related hub genes in DLBCL patients at different IPI levels by weighted gene co-expression network analysis (WGCNA), and constructed an IPI-based immune prognostic model (IPI-IPM). We then characterized the genetic, somatic mutational and molecular profile of IPI-IPM subgroups, investigated the expression of several inhibitory immune checkpoints between low- and high-risk subgroups, and applied. The results showed that IPI-IPM was a promising prognostic biomarker, which also had potential for use in patient management.

Materials and Methods

Data selection and acquisition

Data acquisition of the present study is fully under the TCGA publication guidelines (https://www.cancer.gov/tcga). Gene expression data (RNA-seq) of 570 samples, masked somatic mutation data of 37 samples and clinical follow up data with clinicopathological characteristics of 566 patients of DLBCL projects (TCGA-DLBC, CTSP-DLBCL1, NCICCR-DLBCL) were obtained from the Cancer Genome Atlas (TCGA) by using the TCGAbiolinks[17] R package in R software (version 4.0.2, https://www.r-project.org). Clinical information, gene expression subtype and genetic subtype of the DLBCL patients were supplemented by referring to open-access supplementary files of the GDC DLBCL

publication[4]. Matched gene expression data and survival follow-up data can be obtained in 563 samples. Matched gene expression data and survival follow-up data with available IPI data can be obtained in 445 samples (**Supplementary Table 1**). Gene expression data (Microarray) with matched clinical information of 414 DLBCL patients were obtained from GSE10846[12, 18] by accessing the GEO database (www.ncbi.nlm.nih.gov/geo). The immunologic gene lists used in the present study were downloaded from the ImmPort (www.immport.org/shared/home) and ImmuneSigDB[19] through MSigDB Collections (www.gsea-msigdb.org/gsea/msigdb). For the RNA-Seq data, HTSeq-count data was downloaded. Combat-Seq function of Sva[20] R package was used to remove batch-effect among different projects. The principal components analysis was performed and visualized to examine the batch effect[21]. For each gene, the effective gene length was extracted by using EDASeq[22] R package and a TPM (Transcripts Per Kilobase Million) gene expression data was calculated through the count2TPM function utilized in the IOBR[23] R package by using the batch-free count data and corresponding effective gene length. Homo sapiens GRCh38 annotation file downloaded from Ensembl[24] was used for gene symbol annotation. DESeq2[25] R package was applied for the normalization of RNA-seq counts data and variance stabilizing transformation (VST) data was used for downstream analysis. DLBCL microenvironmental (LME) signatures and LME subtype categorizing method were referred to Kotlov et al.'s publication on Cancer Discovery[11]. An anti-PD-L1 treatment cohort (IMvigor210 cohort) with the gene expression and overall survival information were downloaded from http://research-pub.gene.com/IMvigor210CoreBiologies[26].

Identification of differentially expressed genes

Samples with available IPI information were categorized to high, intermediate and low risk groups according to criterion of IPI, and the DESeq2[25] R package was applied to identify differentially expressed genes (DEGs) between high and low risk groups. The differential expression was defined with a fold-change of threshold at 1.5 and a false discovery rate (FDR) value < 0.05.

Gene functional enrichment analysis

The clusterProfiler[27] R package was used for both over representation analysis and pre-ranked gene set enrichment analysis (GSEA). Analysis of Gene Ontology (GO)[28], Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway[29], and Reactome pathway[30] was contained in the present study. Adjusted P value < 0.05 was considered statistically significance. The threshold for GSEA was set at the p-value < 0.05, FDR < 0.05 and | normalized enrichment score (NES) | > 1.0. The non-parametric gene set variation analysis was further performed with the GSVA[31] package of R.

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) is commonly used for analyzing high-throughput gene expression data with different characteristics, so as to mine gene co-expression networks

and intramodular hub genes based on pairwise correlations in genomic applications. In the present study, we applied the WGCNA[32, 33] R package to analyze key gene clusters that were most relevant to IPI scores in DLBCL samples.

Construction and validation of IPI-based immune prognostic model (IPI-IPM)

IPI-based immune related genes were selected to construct the prognostic risk model. The training cohort (563 patients with matched normalized RNA-seq data and survival data from TCGA) was used for the construction of IPI-IPM, and the testing cohort (414 patients with matched normalized microarray data and survival data from GSE10846) was used for validation of the prognostic risk model. The Survival R package was used to analyze the correlation between the expression of objective gene sets and DLBCL patients' overall survival (OS). Univariate Cox regression analysis was performed to screen genes, of which the expression was associated with OS with a P value < 0.05. Lasso (least absolute shrinkage and selection operator) regression analysis was applied for variable selection and regularization to enhance the prediction accuracy and interpretability by using the glmnet[34] R package. Multivariate Cox regression analysis was then carried out to select the optimal genes, based on the method of Akaike information criterion (AIC)[35]. For each sample, the risk score equals the sum of the normalized expression of each gene multiplying its corresponding regression coefficient. Time-dependent ROC curves were plotted by using the survivalROC[36] R package. 563 DLBCL patients in the training cohort were divided to low- and high- risk score groups according to the optimal cut-off value with largest AUC in the receiver operating characteristic (ROC) curve of the median survival time. Then, Kaplan-Meier survival analysis and time-dependent ROC curve analysis were performed to evaluate the prognostic significance and accuracy of IPI-IPM[37]. Besides, Harrell's concordance index (C-index) was calculated by using the survcomp[38] R package. Univariate and multivariate Cox regression analyses were performed on the risk score and all available clinicopathologic parameters, including age, gender, Ann Arbor clinical stage, LDH ratio, ECOG performance status, number of extranodal site. Then we utilized the rms[39] R packages set up a prognostic nomogram for OS probability assessment by enrolling all the independent prognostic factors. The discriminative efficacy, consistency and clinical judgment utility of the nomogram score was evaluated by time-dependent Calibration plots and decision curve analysis (DCA)[40] using the rmda[41] R package.

Comprehensive analysis of molecular and immune characteristics in different IPI-IPM subgroups

DEGs between high and low risk score groups were analyzed following the fold change of a threshold at 1.5 and FDR value < 0.05. The gene expression of samples between IPI-IPM subgroups were analyzed with the t-distributed stochastic neighbor embedding (t-SNE) method by using the Rtsne[42] package of R and then visualized on the 3D map with the scatterplot3d[43] package of R. Somatic mutations of IPI-IPM subgroups were analyzed by using the Maftools[44] R package. The intersection of DEGs and immune related gene set was used to construct a protein-protein interaction (PPI) network based on the STRING[45] database.

The Cytoscape[46] plugin MCODE[47] and cytoHubba[48] was utilized to identify top10 degree genes in the network. The TRRUST[49] database was browsed to explore the curated transcriptional regulatory networks. Gene expression data (TPM) of 570 DLBCL samples were imported into CIBERSORT[50], MCPCounter[51] and xCell[52] to calculate the score to estimate of the proportion of TME cells including immune and stromal cells.

Clustering analysis of expression pattern of LME signatures

Unsupervised clustering algorithm was applied to analysis the gene expression pattern of LME signatures in 563 DLBCL samples. By using ConsensusClusterplus[53] R package, we performed k-means clustering algorithm with 1000 repetitions to ensure the stability.

Data Analysis

All statistical data was analyzed in the R software (version 4.0.3). A Wilcoxon test was applied to compare continuous variables between two groups of sample data. A Kruskal-Wallis test was applied to compare continuous variables among three or more groups of sample data. A test is considered with statistical significance at $P < 0.05$. Pearson correlation was used to test the linear correlation between two sets of data and an absolute Pearson correlation coefficient larger than 0.3 was considered to be correlated. A Pearson correlation is considered with statistical significance at FDR < 0.05. We used ggplot2, ggstatsplot and ggpubr R packages[54, 55] for data analysis and visualization.

Results

Identification of immune related genes associated with IPI in DLBCL patients

A flowchart was diagramed to demonstrate the procedure and result of our study (**Figure 1**). RNA-seq data of 570 DLBCL patients were obtained from the TCGA (48 from TCGA-DLBC, 41 from CTSP-DLBCL1, 481 from NCICCR-DLBCL). Among the 566 DLBCL patients with survival (OS) data, 321 (56.71%) were male and 245 (43.29%) were female. Age of the patients at initial diagnosis ranged from 14 to 92 (median = 62). Other clinical characteristics including follow-up period, Ann Arbor stages, LDH ratio, ECOG performance status, and number of extranodal sites were all documented in **Table 1** and **Supplementary material 1**. To remove the batch effect among these 3 projects, we utilize the ComBat-seq function to transform the raw counts data with the sva R package. Then, the principal component analysis (PCA) was performed to show that there was no obvious batch effect among the samples (**Figure 2A**). After excluding samples with unrecorded IPI score or with IPI score crossing risk groups (such as 1-5 or 3-4), 458 DLBCL patients were divided into low-risk group (n = 118), intermediate-risk group (n = 221, 106 at low-intermediate risk and 92 at high-intermediate risk), and high-risk group (n = 109) (**Table 1**). Consistent with the previous publications, patients in the low-IPI risk group were shown with much longer overall survival (OS) (**Figure 2B**).

As shown in gene set variation analysis (GSVA), extracellular matrix (ECM) receptor interaction, CD28 dependent PI3K/AKT signaling and IL-6 type cytokine receptor ligand interaction and other immunologic signaling pathways were significantly enriched in low-risk group (**Figure 2C and Figure S1A-B**). A total 4651 genes (633 up-regulated and 4018 down-regulated) were detected significantly differentially expressed between the IPI high- and low-risk groups (**Figure S1C-D and Supplementary material 2**). By intersecting with the immunologic signature gene sets (combing 4686 genes from ImmuneSigDB and Immport, **Supplementary material 3**), 1927 immune DEGs (254 up-regulated and 1673 down-regulated) were identified for further analysis (**Figure 2D and S1E-F**).

We applied variance stabilizing transformed (VST) expression data via DESeq2 as the input data for WGCNA, including 13329 genes with the top 25% variance among all samples (**Figure 3A**). All clinical characteristics were enrolled as trait variables, and the best β value in the co-expression network was calculated to be 9 **(Figure S2B)**. Index for clustering of module eigengenes was modified to be 0.65, so as to construct reasonable number of merged modules **(Figure S2A-D)**. As shown in the Module-trait relationship, 8 modules were significantly correlated with IPI group (**Figure 3B**), and high correlation ($P < 0.0001$) between gene significance of immune score and gene module membership were found in the genes of 3 modules (Brown, Pink, Darkred) (**Figures 3C and S2E-G**). By intersecting 4106 genes from the top 3 IPI correlating modules with 1927 immune DEGs, a total of 371 genes were identified as immune related genes associated with IPI, which were used for prognostic risk model construction (**Figure S2H**).Several gene sets, such as ECM organization, ECM receptor interaction, PI3K/AKT signaling and integrin cell surface interaction, etc., were enriched in the top 3 IPI correlating modules (**Figure S2I and Supplementary material 4**).

Construction and validation of an IPI-based immune prognostic risk model

In the training cohort (n = 563), 93 out of 371 IPI-related immune genes were significantly correlated with OS in the univariate Cox regression analysis (**Supplementary material 5**). Next, we applied the Lasso penalized Cox regression to identify the optimal number of genes (n = 10) for risk score model (**Figure 4A**). As a result of stepwise multivariate Cox regression and AIC analysis, 8 genes were selected to construct the most optimal IPI-based immune prognostic model (IPI-IPM) (**Figure S3A-B and 4B**). Risk score = (expression level of CMBL * 0.360 + expression level of TLCD3B * (-0.350) + expression level of SYNDIG1 * (-0.247) + expression level of ESM1 * (-0.238) + expression level of EPHA3 * (-0.163) + expression level of HUNK * (-0.156)+ expression level of PTX3 * 0.138 + expression level of IL12A * 0.111). As shown in the time-dependent ROC curves, AUCs were 0.703, 0.738, 0.733 for the 1, 3, 5-year, respectively (**Figure 4C**). The C-index of the risk score model was 0.732 (95% CI: 0.684-0.779, P = 1.38e-21). These results showed that IPI-IPM endowed a good capacity in OS prediction. According to the ROC curve of the median survival time (**Figure S3C**), we identified 0.982 as the cut-off value. Then we calculated the risk score of each patient and divided them into high- and low-risk groups. As shown in **Figure 4D**, Kaplan-Meier survival analysis showed shorter OS of patients in the high-risk score group (log-

rank P = 3.13e-14). The distribution of the risk score, survival status and the 8-gene expression between high- and low-risk score groups was shown in **Figure S3D and Figure 4E**.

Moreover, 335 patients with all available clinicopathologic parameters were enrolled for Multivariate Cox regression analysis, presenting the risk scores and age as independent prognostic factors of OS (**Figure 4F**). Risk scores along with age, Ann Arbor clinical stage, LDH ratio, ECOG performance status and number of extranodal sites were integrated to construct a nomogram model, showing IPI-IPM risk scores with the highest risk points (ranging from 0 to 100, **Figure 5A**). As shown in the DCA analysis, the nomogram along with risk score from IPI-IPM showed relatively high net benefit (**Figure 5B**). Moreover, the bias-corrected lines for the nomogram were shown close to the ideal line in 1,3,5-year and the median survival time periods (**Figure 5C**). The C-index for the nomogram was 0.790 (95% CI: 0.736-0.843, P = 2.38e-26). Altogether, these results suggested that the nomogram had excellent capacity and consistency for OS prediction in the training cohort.

414 patients from GEO (GSE10846) with survival data and microarray data were enrolled for further validation of IPI-IPM. The risk score for each patient was calculated and all patients were divided into high- and low-risk groups likewise. As shown in **Figure 5D**, AUCs were 0.619, 0.603, 0.601 for the 1, 3, 5-year, respectively. Kaplan-Meier survival analysis showed significantly shorter OS of patients in the high-risk group (log-rank P = 5.30e-05, **Figure 5E and S3E-F**). Moreover, patients in the high-risk group were shown to have shorter OS no matter which treatment regimens they received (**Figure 5F**). Taken the results of training and testing cohorts together, the IPI-IPM and the nomogram combining risk score with relevant clinical characteristics (age, Ann Arbor clinical stage, LDH ratio, ECOG performance and number of extranodal site) was an excellent model for predicting short-term or long-term OS in DLBCL patients, which might guide the therapeutic strategy decision and long-term prognosis follow-up.

Molecular characteristics of IPI-IPM subgroups

Compared to the low-risk group, A total 5980 genes (690 up-regulated and 5290 down-regulated), and 2731 immunologic genes (400 up-regulated and 2331 down-regulated) were detected significantly differentially expressed in high-risk group (**Figure S4A and 6A, Supplementary material 6**). The pre-ranked GSEA was performed to display that several gene sets including negative regulation of immune response, DNA repair and response to IL-12, etc., were enriched in high-risk group, and several gene sets including ECM receptor interaction, IL-10 synthesis and regulation of humoral immune response, etc., were enriched in low-risk group. Details were all documented in **Figure 6B and S4B-E** and **Supplementary material 6.** Additionally, t-SNE was applied to show an obvious genetic diversity between high- and low-risk groups (**Figure 6C**). Based on the Pearson correlation analysis, there existed either positive or negative correlation between these 8 genes from IPI-IPM (**Figure 6D)**. And **Figure 6E** presented the heatmap of differentially expressed genes correlating IPI-IPM risk scores between high- and low-risk groups.

To gain further molecular insight into the molecular characteristics of IPI-IPM, 176 genes correlating with risk scores and these 8 genes from IPI-IPM were identified as IPI-IPM associated immune genes (absolute

Pearson correlation coefficient ≥ 0.3, FDR < 0.05). Over representation analysis was applied to identify the enriched biological functions and pathways, such as ECM organization and T cell differentiation, activation and mediated immunity (**Figure 7A-B**). Detailed results were listed in **Supplementary material 6**. As shown in the PPI network, ECM organization, oncogenesis and tumor immunity associated regulatory genes (COL6A2, COL16A1, COL26A1, COL13A1, COL22A1, C3, ELN, MMP9, CLU, FOXP3,ADGRL1) were closely correlated, acting as hub genes (**Figure 7C**). In addition, TFTRUST database was applied to explore the transcription factors (TFs) regulating the 184 IPI-IPM associated immune genes, thus MMP9, FOXP3 and PLAU were identified in the TF network (**Figure 7D**).

Based on the somatic mutational data of 37 samples from TCGA, 18 out of 20 patients in high-risk group and 13 out of 17 patients in low-risk groups were found with altered gene expression (**Figure S5A-B and 8A**). Although most mutations were missense mutation, more nonsense and other mutations were identified in high-risk group (**Figure 8A**). Besides, the mutation frequency of top10 genes in high-risk group was much higher than that in low-risk group. Furthermore, we investigated specific mutation sites of key genes corresponding to their amino acids location, including KMT2D, MUC16, CARD11, LRP1B, BTG2 and PIM1, etc. (**Figure 8B and S5C**). As shown in the Oncodrive plot, MyD88, CD79B, KHL6 and MUC4 were identified as cancer driver genes in high-risk group while only PEG3 and ZNF337 were identified in low-risk group (**Figure 8C and Supplementary material 7**).

Immune landscape of IPI-IPM subgroups

CIBERSORT was applied to analyze the infiltrating abundances of various immune cell types in different IPI-IPM subgroups (**Figure 9A and S6A**). Activated memory $CD4^+$ T cells and resting NK cells were highly infiltrated in high-risk group, while memory B cells, $CD8^+$ T cells, follicular helper T cells, regulatory T cells (Tregs), and non-activated macrophages (M0) were more abundant in low-risk group (**Figure 9B**). In addition, MCPCounter and xCell algorithm were applied to show that myeloid dendritic cells (mDCs) and common lymphoid progenitors were highly infiltrated in high-risk group, while hematopoietic stem cells (HSCs), cancer associated fibroblasts (CAFs) and T cells especially $CD8^+$ T cells were more abundant in low-risk group (**Figure 9C and S6B**).

In addition, DLBCL subtypes varied between high- and low-risk groups. As shown in **Figure 10A**, more patients with higher IPI level were classified into IPI-IPM high-risk group, with similar results about age, Ann Arbor clinical stage, LDH ratio, ECOG performance and the number of extranodal sites (**Figure S6C**). Whereas high-risk group contained a higher proportion of ABC-DLBCLs, low-risk group was more enriched in GCB-DLBCLs (P = 1.34e-14). Schmitz et al.[4] identified four prominent genetic subtypes in DLBCL with different responses to immunochemotherapy, termed MCD (the co-occurrence of $MYD88^{L265P}$ and CD79B mutations), BN2 (BCL6 fusions and NOTCH2 mutations), N1 (NOTCH1 mutations), and EZB (EZH2 mutations and BCL2 translocations), and uneven distribution of these 4 subtypes were found between high- and low-risk groups. For example, the poor-prognostic MCD subtype represented 40% of high-risk group and 12% of low-risk group. Also, the correlation analysis showed IPI-

IPM risk scores positively correlated with the expression Bcl-2 and Myc, and negatively correlated with Bcl-6 expression (**Figure 10B**). Taken together these data indicates that IPI-IPM provide additional orthogonal information from previous lymphoma classifications.

To further explore the interaction between DLBCL cells and the microenvironment, Kotlov et al. [11]defined the lymphoma microenvironment (LME) into 4 major transcriptionally-defined categories with distinct biological properties and clinical behavior, including "germinal center-like" (GC), "mesenchymal" (MS), "inflammatory" (IN) and "depleted" (DP) form, respectively. Similarly, we utilized an unsupervised clustering method to assign the samples into 4 groups by using expression data of the 22 functional gene expression signatures (F^{GES}) sets (**Figure 10C**). GSVA enrichment scores were calculated to demonstrate the distinct TME characteristics among the 4 biological patterns (**Figure 10D**). Consistent with CIBERSORT results, the DP LME categories represented 35% of high-risk group whereas GC-like and MS LME categories were more enriched in low-risk group (**Figure 10E-F and Supplementary material 8**).

Potential therapeutic value of IPI-IPM

To further understand the effects of the risk score on drug response, 184 IPI-IPM associated immune genes mapped into the Connectivity Map database[56]. As shown in **Figure 11A and S7**, 14 genes (ADORA1, ADRA2A, CACNA1C, DBH, DNM1, ELN, ENPP1, EPHA1, IL12A, MMP9, PLAU, S1PR2, SLC1A2 and SV2A) were associated with 122 inhibitors, involved with 54 mechanisms of action (MOA). Details were documented in **Supplementary material 9**.

The clinical development of cancer immunotherapies, along with advances in genomic analysis, has validated the important role of TME in predicting cancer response to immune checkpoint blockade therapy (ICB)[13, 57, 58]. We investigate the expression of several inhibitory immune checkpoints between high- and low-risk groups using TMM normalized counts data. As shown in **Figure 11B**, the expression of PD-L1, BTLA and SIGLEC7 were significantly up-regulated in high-risk group. Details were documented in **Supplementary table 10**.

Since various biomarkers were reported to predict the response to immunotherapy, including tumor mutation burden and immune checkpoints expression such as PD-L1. We examined the value of the IPI-IPM to predict the response of patients to ICB therapy, based on a publicly accessible dataset - IMvigor210 Cohort[26]. Using the same cut-off point for group assigning, we observed that patients with high IPI-IPM risk had shorter OS after anti-PD-L1 treatment (log-rank P = 0.003, **Figure11C**).

Discussion

Recent groundbreaking insights into the pronounced genomic heterogeneity of DLBCL have identified potential biomarkers for patient diagnosis and prognosis, paving the way for a standardized application of precise medicine[3, 5, 13]. Multiple subtype classifications as well as IPI and enhanced NCCN-IPI were built to stratify prognostically relevant subgroups of DLBCL patients with R-CHOP therapy, the robustness of

which however needs to be further investigated in the context of targeted therapies[7, 11, 13, 59]. In the current study, we used WGCNA to profile IPI correlating immune gene sets, constructed and validated an 8-gene IPI-IPM (CMBL, TLCD3B, SYNDIG1, ESM1, EPHA3, HUNK, PTX3 and IL12A), with shorter OS in high-risk patients and longer OS in low-risk patients in both TCGA and GEO cohorts.

ESM1, also known as endocan, was initially proven to regulate endothelial cell function, also participate in the initiation and progression of human cancers including esophageal cancer, hepatocellular carcinoma, bladder cancer and breast cancer[60]. EphA3 receptor plays critical role in cell adhesion and migration during development and homeostasis of many tissues, as well as cancer growth, progression and angiogenesis[61]. Also, EphA3 high expression in tumors but not in normal tissues, together with antitumor properties of anti-EphA3mAb (chIIIA4), defined EphA3 as a potential target for antibody-based anticancer therapies[61]. PTX3 is secreted by dendritic cells, macrophages, and fibroblasts, and actively involved in regulation of inflammation, tissue remodeling and cancer[62]. PTX3 interacts with the PI3K/AKT/mTOR signaling pathway or the fibroblast growth factor-2 (FGF2) / FGF receptor (FGFR) system to regulate tumor cell proliferation, apoptosis and metastasis in lung cancer, breast cancer, melanoma, prostate cancer and multiple myeloma[62, 63]. IL12A is a potent immune suppressive cytokine produced by regulatory B (Breg) cells, Treg cells, macrophages, dendritic cells and tumor cells, which suppresses the effector functions of $CD4^+$ and $CD8^+$ T cells but strongly favors Tregs proliferation[64]. Furthermore, Larousserie et al. found that high levels of IL12A were associated with the poor survival of DLBCL patients[65].

High throughput gene expression databases and bioinformatic analysis have enabled systematic profiling of prognostic signatures in DLBCL. For example, Zhou et al. uncovered differentiated lncRNA expression pattern between GCB and ABC DLBCL and identified an immune-associated 17-lncRNA signature for subtype classification and prognosis prediction[66]. Chapuy et al. defined five distinct DLBCL subsets by integrating recurrent mutations, SCNAs and SVs[5]. And Hu et al. built a predictive model comprising drug resistance signature along with clinical factors including age at diagnosis, stage, the number of extra nodal sites and the ECOG performance score[59]. In the current study IPI-IPM risk score remained an independent prognostic factor after modification of clinical characteristics, thus we developed a nomogram model combining the risk score and other clinical features (age, Ann Arbor clinical stage, LDH ratio, ECOG performance and number of extranodal site) to predict OS probability of DLBCL patients in 1-, 3-, 5-year and the median survival time, respectively. Both the calibration curve and DCA analysis supported that our nomogram provides a complementary perspective on individualizing tumors and develops an individual scoring system for patients, thus arising to be a promising tool for clinicians in the future.

In the GSEA analysis between high- and low-risk groups, several immune-related gene sets, including negative regulation of immune response, B cell and T cell proliferation, response to IL-12 and γ-IFN, and NOD-like and TOLL-like receptor signaling, were enriched in high-risk group whereas ECM receptor interaction, IL-10 synthesis and regulation of humoral immune response, etc., were enriched in low-risk group. Therefore, we speculated that the local immune signature conferred a weaker immune phenotype in high-risk group but an intense immune phenotype in low-risk group. Moreover, over representation

analysis identified several immune-related pathways including ECM organization and T cell activation were enriched with IPI-IPM associated immune genes. ECM has important roles in supporting the cells and regulating intercellular interactions, thus contributing to progression of several malignancies[11, 14]. Lenz et al. built a survival model with 2 stromal gene signatures for DLBCL patients who received CHOP or R-CHOP, where the prognostically favorable stromal-1 signature reflected ECM deposition and histiocytic infiltration while the prognostically unfavorable stromal-2 signature reflected tumor blood-vessel density[12].

To further explore the immunological nature of the IPI-IPM subgroups, we analyzed 37 samples' somatic mutational profile and found higher mutation counts in high-risk group with more nonsense mutations, although missense mutations were the most common type. And the largest difference in mutations between high- and low-risk groups was in KMT2D mutation (40% in high-risk samples vs. 17.65% in low-risk samples)[67]. KMT2D is a tumor suppressor gene in DLBCL and genetic ablation of KMT2D in a BCL2-overexpression driven model promotes a higher penetrance of DLBCL[68]. Moreover, MyD88 and CD79B were identified as cancer driver genes in high-risk group, in good agreement with the previous findings that MYD88 and CD79B mutations have been associated with tumor response and survival in DLBCL patients[69-71]. Ngo et al. initially identified activating MyD88 mutations in DLBCL, where L265P was the most frequent and oncogenic form. MyD88 was shown to interact with IL-1 receptor-associated kinase (IRAK) 1 and IRAK4, activate the NF-κB and JAK-STAT3 pathways, promoting malignant cells proliferation and causing worse survival of DLBCL patients[72]. CD79B mutations frequently are detected in the first tyrosine (Y196) of the immunoreceptor tyrosine-based activation motif (ITAM). CD79B mutations were shown to cause chronic active BCR signaling and constitutive NF-κB activation, further promoting tumor cells growth within immunosuppressive TME.

Consistently, the composition of immune cells was different between two IPI-IPM subgroups, where memory B cells, $CD8^+$ T cells, Tregs, M0 and CAFs were more abundant in low-risk group. It's generally accepted that cytotoxic $CD8^+$ T cells, following successful priming, recognize tumor-specific (neoantigens) or tumor-associated antigens and exert anti-tumor function primarily via the release of cytotoxic molecules such as perforin and granzymes[73-75]. However, Tregs suppress $CD8^+$ T cells by direct cell contact and secretion of inhibitory cytokines including IL-10 and TGF-β[74, 76]. TAMs have been shown to mediate antibody-dependent cellular phagocytosis of rituximab engaged malignant B cells, but limit $CD8^+$ T cells activity through PD-L1 expression, as well as releasing IL-10 and TGF-β or inhibiting enzymes, thus regulating anti-tumor immunity and response to therapy[77]. CAFs, the resident fibroblasts activated in a chronic inflamed TME, have been shown to promotes recruitment and polarization of regulatory cells including Tregs, monocytes and M2 by secreting IL-6, CXCL12, Chi3L1, MCP-1 and SDF-1, thus actively shaping the immune infiltration in TME[78]. In addition, CAFs have been shown to impact the cytolytic activity of $CD8^+$ T cells through different mechanisms, such as producing prostaglandin E2 (PGE2) and NO to dampen $CD8^+$ T cells proliferation, expressing PD-L2 and FasL to promote $CD8^+$ T cells apoptosis, and inducing abnormal ECM deposition and remodeling to physically trap $CD8^+$ T cells and prevent effective tumor access[79].

Besides, DP LME represented more of high-risk group while immune rich GC LME was more enriched in low-risk group. Kotlov et al. characterized the microenvironment of DLBCL by analyzing gene expression profiles, developing TME-derived F^{GES}, analyzing ECM composition by proteomics and establishing patient-derived tumor xenografts (PDTX) models[11]. Four basic categories of DLBCL LME with distinct clinical and biological connotations were identified to uncover the bidirectional interaction between DLBCL cells and LME. Remarkably, immune rich GC LME confers a better prognosis than DP LME, suggesting the fundamental role of LME in preventing lymphomagenesis. In turn, DLBCL cells develop genetic and epigenetic traits that contribute to immune evasion from LME.

Finally, we applied Spearman correlation analysis to estimate the potential therapeutic effects of IPI-IPM; and explored the expression of several inhibitory immune checkpoints between IPI-IPM subgroups to predict the response to immunotherapy. The results shown that IPI-IPM associating genes were correlated with sensitivity to drugs targeting Aurora kinase, DNA methyltransferase (DNMT), Histone acetyltransferase (HAT), FLT3, EGFR and VEGFR signaling pathways, indicating that high-risk DLBCL patients may benefit from novel inhibitors targeting these signaling pathways. As for the inhibitory checkpoints expression, PD-L1, BTLA and SIGLEC7 were significantly up-regulated in high-risk group. Upregulation of checkpoints (PD-1, CTLA-4, and TIM-3) and their ligands (PD-L1 and PD-L2) in TME can mediate tumor cells to escape immune surveillance by modulating T-cell activity[13, 57, 80, 81]. Thus, ICB has exerted significant antitumor effect in both solid tumors and hematologic malignancies. Although nivolumab monotherapy showed a low efficacy in unselected DLBCL patients, pembrolizumab combined with R-CHOP was safe and associated with a high CR rate and improved 2-year PFS[58, 82]. BTLA has been reported to mark a high-checkpoint-expressing T-cell subset (PD-1, TIM-3, LIGHT, and LAG-3) with decreased cytolytic function and increased proliferation ability, thus correlating poor prognosis in DLBCL[13, 83].

Conclusion

Taken together, the IPI-IPM risk score was compatible with the ability of tumor-infiltrating immune cells to determine the expression of immune checkpoints, suggesting that the poor prognosis of high-risk group may be due to the stronger immunosuppressive TME, and high-risk patients will benefit more from immune checkpoint inhibitors than low-risk patients, thereby resulting in a better prognosis. Our research provides new insights into the TME and immune-related therapies for DLBCL. However, it is noteworthy that some limitations came out because the conclusion was drawn from data from retrospective studies, and prospective studies are warranted to further confirm our results. In addition, functional and mechanistic studies of the genes in our risk model should be conducted to support their clinical application.

Abbreviations

DLBCL - Diffuse large B-cell lymphoma	Lasso - least absolute shrinkage and selection operator
NHL - non-Hodgkin B-cell lymphoma	AIC - Akaike information criterion
GCB Subtype - germinal center B-cell-like subtype	ROC - receiver operating characteristic
ABC Subtype - activated B-cell-like subtype	C-index - concordance index
GEP - gene expression profiling	DCA - decision curve analysis
COO - cell-of-origin	t-SNE - t-distributed stochastic neighbor embedding
SCNA - somatic copy number alterations	PCA - principal component analysis
SV - structural variants	GSVA - gene set variation analysis
IPI - International Prognostic Index	ECM - extracellular matrix
TME - tumor microenvironment	TFs - transcription factors
ICB - immune checkpoint blockade therapy	Tregs - regulatory T cells
GC form - germinal center-like form	M0 - non-activated macrophages
MS form - mesenchymal form	mDCs - myeloid dendritic cells
IN form - inflammatory form	HSCs - hematopoietic stem cells
DP form – depleted form	CAFs - cancer associated fibroblasts
WGCNA - weighted gene co-expression network analysis	F^{GES} - functional gene expression signatures
IPI-IPM - IPI-based immune prognostic model	MOA - mechanisms of action
TCGA - the Cancer Genome Atlas	FGF - fibroblast growth factor
VST - variance stabilizing transformation	Breg - regulatory B cells
LME – DLBCL/Lymphoma microenvironment	IRAK - IL-1 receptor-associated kinase
DEGs - differentially expressed genes	ITAM - immunoreceptor tyrosine-based activation motif
GSEA - gene set enrichment analysis	PGE2 - prostaglandin E2
GO - Gene Ontology	PDTX - patient-derived tumor xenografts
KEGG - Kyoto Encyclopedia of Genes and Genomes	DNMT - DNA methyltransferase
NES - normalized enrichment score	HAT - Histone acetyltransferase
OS - overall survival	

Declaration

Ethical Approval and Consent to participate

Not applicable

Consent for publication

All the authors read and approved the final version of the manuscript, as well as consented for publication.

Availability of supporting data

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that there is no conflict of interest.

Funding

This work was supported by grants from the National Natural Science Foundation of China (No. 81974007 to Chunyan Sun) and the National Key R&D Program of China (Grant No 2019YFC1316204 to Yu Hu).

Authors' contributions

SM, DS, CS and YH conceived and designed the study. SM and DS did the literature research, performed study selection, data extraction, statistical analysis and wrote the draft. LA, FF and FP participated in the extraction and analysis of data. All the authors read and approved the final version of the manuscript.

Acknowledgements

The results here are based upon data generated by the TCGA Research Network. The study reported herein fully satisfies the TCGA publication requirements (https://www.cancer.gov/tcga). The authors would like to thank the TCGA database developed by National Institutes of Health.

Authors' information

Shidai Mu[1,#,*], Deyao Shi[2,#], Lisha Ai[1], Fengjuan Fan[1], Fei Peng[1], Chunyan Sun[1*], Yu Hu[1*]

1 Institution of Hematology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430022, China

2 Department of Orthopaedics, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430022, China

References

1. Wang, L, Li, LR, and Young, KH (2020). New agents and regimens for diffuse large B cell lymphoma. *J Hematol Oncol* **13**: 175.
2. Solimando, AG, Annese, T, Tamma, R, Ingravallo, G, Maiorano, E, Vacca, A, Specchia, G, and Ribatti, D (2020). New Insights into Diffuse Large B-Cell Lymphoma Pathobiology. *Cancers (Basel)* **12**.
3. El Hussein, S, Shaw, KRM, and Vega, F (2020). Evolving insights into the genomic complexity and immune landscape of diffuse large B-cell lymphoma: opportunities for novel biomarkers. *Mod Pathol* **33**: 2422-2436.
4. Schmitz, R, Wright, GW, Huang, DW, Johnson, CA, Phelan, JD, Wang, JQ, Roulland, S, Kasbekar, M, Young, RM, Shaffer, AL, *et al.* (2018). Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *The New England journal of medicine* **378**: 1396-1407.
5. Chapuy, B, Stewart, C, Dunford, AJ, Kim, J, Kamburov, A, Redd, RA, Lawrence, MS, Roemer, MGM, Li, AJ, Ziepert, M, *et al.* (2018). Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* **24**: 679-690.
6. Pasqualucci, L, and Dalla-Favera, R (2018). Genetics of diffuse large B-cell lymphoma. *Blood* **131**: 2307-2319.
7. Dubois, S, Tesson, B, Mareschal, S, Viailly, PJ, Bohers, E, Ruminy, P, Etancelin, P, Peyrouze, P, Copie-Bergman, C, Fabiani, B, *et al.* (2019). Refining diffuse large B-cell lymphoma subgroups using integrated analysis of molecular profiles. *EBioMedicine* **48**: 58-69.
8. Chen, YW, Hu, XT, Liang, AC, Au, WY, So, CC, Wong, ML, Shen, L, Tao, Q, Chu, KM, Kwong, YL, *et al.* (2006). High BCL6 expression predicts better prognosis, independent of BCL6 translocation status, translocation partner, or BCL6-deregulating mutations, in gastric lymphoma. *Blood* **108**: 2373-2383.
9. Huang, S, Nong, L, Wang, W, Liang, L, Zheng, Y, Liu, J, Li, D, Li, X, Zhang, B, and Li, T (2019). Prognostic impact of diffuse large B-cell lymphoma with extra copies of MYC, BCL2 and/or BCL6: comparison with double/triple hit lymphoma and double expressor lymphoma. *Diagn Pathol* **14**: 81.
10. Wight, JC, Chong, G, Grigg, AP, and Hawkes, EA (2018). Prognostication of diffuse large B-cell lymphoma in the molecular era: moving beyond the IPI. *Blood Rev* **32**: 400-415.
11. Kotlov, N, Bagaev, A, Revuelta, MV, Phillip, JM, Cacciapuoti, MT, Antysheva, Z, Svekolkin, V, Tikhonova, E, Miheecheva, N, Kuzkina, N, *et al.* (2021). Clinical and Biological Subtypes of B-cell Lymphoma Revealed by Microenvironmental Signatures. *Cancer discovery*.
12. Lenz, G, Wright, G, Dave, SS, Xiao, W, Powell, J, Zhao, H, Xu, W, Tan, B, Goldschmidt, N, Iqbal, J, *et al.* (2008). Stromal gene signatures in large-B-cell lymphomas. *The New England journal of medicine* **359**: 2313-2323.
13. Xu-Monette, ZY, Xiao, M, Au, Q, Padmanabhan, R, Xu, B, Hoe, N, Rodriguez-Perales, S, Torres-Ruiz, R, Manyam, GC, Visco, C, *et al.* (2019). 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.
14. Cioroianu, AI, Stinga, PI, Sticlaru, L, Cioplea, MD, Nichita, L, Popp, C, and Staniceanu, F (2019). Tumor Microenvironment in Diffuse Large B-Cell Lymphoma: Role and Prognosis. *Anal Cell Pathol (Amst)* **2019**: 8586354.
15. Ciavarella, S, Vegliante, MC, Fabbri, M, De Summa, S, Melle, F, Motta, G, De Iuliis, V, Opinto, G, Enjuanes, A, Rega, S, *et al.* (2018). Dissection of DLBCL microenvironment provides a gene expression-based predictor of survival applicable to formalin-fixed paraffin-embedded tissue. *Ann Oncol* **29**: 2363-2370.
16. Hopken, UE, and Rehm, A (2019). Targeting the Tumor Microenvironment of Leukemia and Lymphoma. *Trends Cancer* **5**: 351-364.
17. Colaprico, A, Silva, TC, Olsen, C, Garofano, L, Cava, C, Garolini, D, Sabedot, TS, Malta, TM, Pagnotta, SM, Castiglioni, I, *et al.* (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic acids research* **44**: e71.
18. Cardesa-Salzmann, TM, Colomo, L, Gutierrez, G, Chan, WC, Weisenburger, D, Climent, F, González-Barca, E, Mercadal, S, Arenillas, L, Serrano, S, *et al.* (2011). High microvessel density determines a poor outcome in patients with diffuse large B-cell lymphoma treated with rituximab

plus chemotherapy. *Haematologica* **96**: 996-1001.
19. Godec, J, Tan, Y, Liberzon, A, Tamayo, P, Bhattacharya, S, Butte, AJ, Mesirov, JP, and Haining, WN (2016). Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. *Immunity* **44**: 194-206.
20. Leek, JT, Johnson, W, Parker, H, Jaffe, A, and Storey, JJD (2014). sva: Surrogate Variable Analysis R package version 3.10. 0 **10**: B9.
21. Groth, D, Hartmann, S, Klie, S, and Selbig, J (2013). Principal components analysis. *Methods in molecular biology (Clifton, NJ)* **930**: 527-547.
22. Risso, D, Schwartz, K, Sherlock, G, and Dudoit, S (2011). GC-content normalization for RNA-Seq data. *BMC bioinformatics* **12**: 480.
23. Zeng, D, Ye, Z, Yu, G, Wu, J, Xiong, Y, Zhou, R, Qiu, W, Huang, N, Sun, L, Bin, J, *et al*. (2020). IOBR: Multi-omics Immuno-Oncology Biological Research to decode tumor microenvironment and signatures: 2020.2012.2014.422647.
24. Howe, KL, Achuthan, P, Allen, J, Allen, J, Alvarez-Jarreta, J, Amode, MR, Armean, IM, Azov, AG, Bennett, R, Bhai, J, *et al*. (2021). Ensembl 2021. *Nucleic acids research* **49**: D884-d891.
25. Love, MI, Huber, W, and Anders, S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**: 550.
26. Mariathasan, S, Turley, SJ, Nickles, D, Castiglioni, A, Yuen, K, Wang, Y, Kadel, EE, III, Koeppen, H, Astarita, JL, Cubas, R, *et al*. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* **554**: 544-548.
27. Yu, G, Wang, LG, Han, Y, and He, QY (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology* **16**: 284-287.
28. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic acids research* **47**: D330-d338.
29. Ogata, H, Goto, S, Sato, K, Fujibuchi, W, Bono, H, and Kanehisa, M (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research* **27**: 29-34.
30. Fabregat, A, Jupe, S, Matthews, L, Sidiropoulos, K, Gillespie, M, Garapati, P, Haw, R, Jassal, B, Korninger, F, May, B, *et al*. (2018). The Reactome Pathway Knowledgebase. *Nucleic acids research* **46**: D649-d655.
31. Hänzelmann, S, Castelo, R, and Guinney, J (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* **14**: 7.
32. Zhang, B, and Horvath, S (2005). A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* **4**: Article17.
33. Langfelder, P, and Horvath, S (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**: 559.
34. Friedman, J, Hastie, T, and Tibshirani, R (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software* **33**: 1-22.
35. Yamaoka, K, Nakagawa, T, and Uno, T (1978). Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations. *Journal of pharmacokinetics and biopharmaceutics* **6**: 165-175.
36. Heagerty, PJ, Lumley, T, and Pepe, MS (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56**: 337-344.
37. Steyerberg, EW, Vickers, AJ, Cook, NR, Gerds, T, Gonen, M, Obuchowski, N, Pencina, MJ, and Kattan, MW (2010). Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass)* **21**: 128-138.
38. Schröder, MS, Culhane, AC, Quackenbush, J, and Haibe-Kains, B (2011). survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics (Oxford, England)* **27**: 3206-3208.
39. Harrell Jr, FE (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*, Springer.
40. Tataranni, T, and Piccoli, C (2019). Dichloroacetate (DCA) and Cancer: An Overview towards Clinical Applications. *Oxidative medicine and cellular longevity* **2019**: 8201079.
41. Kerr, KF, Brown, MD, Zhu, K, and Janes, H (2016). Assessing the Clinical Impact of Risk Prediction Models With Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **34**: 2534-2540.

42. Van Der Maaten, LJTJoMLR (2014). Accelerating t-SNE using tree-based algorithms **15**: 3221-3245.
43. Ligges, U, and Mächler, M. Scatterplot3d-an r package for visualizing multivariate data, 2002.
44. Mayakonda, A, Lin, DC, Assenov, Y, Plass, C, and Koeffler, HP (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research* **28**: 1747-1756.
45. Szklarczyk, D, Gable, AL, Lyon, D, Junge, A, Wyder, S, Huerta-Cepas, J, Simonovic, M, Doncheva, NT, Morris, JH, Bork, P, *et al*. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research* **47**: D607-d613.
46. Shannon, P, Markiel, A, Ozier, O, Baliga, NS, Wang, JT, Ramage, D, Amin, N, Schwikowski, B, and Ideker, T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**: 2498-2504.
47. Bader, GD, and Hogue, CW (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics* **4**: 2.
48. Chin, CH, Chen, SH, Wu, HH, Ho, CW, Ko, MT, and Lin, CY (2014). cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC systems biology* **8 Suppl 4**: S11.
49. Han, H, Cho, JW, Lee, S, Yun, A, Kim, H, Bae, D, Yang, S, Kim, CY, Lee, M, Kim, E, *et al*. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic acids research* **46**: D380-d386.
50. Newman, AM, Liu, CL, Green, MR, Gentles, AJ, Feng, W, Xu, Y, Hoang, CD, Diehn, M, and Alizadeh, AA (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature methods* **12**: 453-457.
51. Becht, E, Giraldo, NA, Lacroix, L, Buttard, B, Elarouci, N, Petitprez, F, Selves, J, Laurent-Puig, P, Sautès-Fridman, C, Fridman, WH, *et al*. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome biology* **17**: 218.
52. Aran, D, Hu, Z, and Butte, AJ (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome biology* **18**: 220.
53. Wilkerson, MD, and Hayes, DNJB (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking **26**: 1572-1573.
54. Wickham, HJC, Switzerland (2016). ggplot2-Elegant Graphics for Data Analysis. Springer International Publishing.
55. Patil, I (2021). Visualizations with statistical details: The'ggstatsplot'approach.
56. Subramanian, A, Narayan, R, Corsello, SM, Peck, DD, Natoli, TE, Lu, X, Gould, J, Davis, JF, Tubelli, AA, Asiedu, JK, *et al*. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**: 1437-1452.e1417.
57. Kline, J, Godfrey, J, and Ansell, SM (2020). The immune landscape and response to immune checkpoint blockade therapy in lymphoma. *Blood* **135**: 523-533.
58. Xie, M, Huang, X, Ye, X, and Qian, W (2019). Prognostic and clinicopathological significance of PD-1/PD-L1 expression in the tumor microenvironment and neoplastic cells for lymphoma. *International immunopharmacology* **77**: 105999.
59. Hu, J, Xu, J, Yu, M, Gao, Y, Liu, R, Zhou, H, and Zhang, W (2020). An integrated prognosis model of pharmacogenomic gene signature and clinical information for diffuse large B-cell lymphoma patients following CHOP-like chemotherapy. *J Transl Med* **18**: 144.
60. Cui, Y, Guo, W, Li, Y, Shi, J, Ma, S, and Guan, F (2021). Pan-cancer analysis identifies ESM1 as a novel oncogene for esophageal cancer. *Esophagus* **18**: 326-338.
61. London, M, and Gallo, E (2020). Critical role of EphA3 in cancer and current state of EphA3 drug therapeutics. *Mol Biol Rep* **47**: 5523-5533.
62. Wang, Z, Wang, X, Zou, H, Dai, Z, Feng, S, Zhang, M, Xiao, G, Liu, Z, and Cheng, Q (2020). The Basic Characteristics of the Pentraxin Family and Their Functions in Tumor Progression. *Front Immunol* **11**: 1757.
63. Doni, A, Stravalaci, M, Inforzato, A, Magrini, E, Mantovani, A, Garlanda, C, and Bottazzi, B (2019). The Long Pentraxin PTX3 as a Link Between Innate Immunity, Tissue Remodeling, and Cancer. *Front Immunol* **10**: 712.
64. Mirlekar, B, and Pylayeva-Gupta, Y (2021). IL-12 Family Cytokines in Cancer and Immunotherapy. *Cancers (Basel)* **13**.
65. Larousserie, F, Kebe, D, Huynh, T, Audebourg, A, Tamburini, J, Terris, B, and Devergne, O (2019).

Evidence for IL-35 Expression in Diffuse Large B-Cell Lymphoma and Impact on the Patient's Prognosis. *Front Oncol* **9**: 563.
66. Zhou, M, Zhao, H, Xu, W, Bao, S, Cheng, L, and Sun, J (2017). Discovery and validation of immune-associated long non-coding RNA biomarkers associated with clinically molecular subtype and prognosis in diffuse large B cell lymphoma. *Mol Cancer* **16**: 16.
67. Bailey, MH, Tokheim, C, Porta-Pardo, E, Sengupta, S, Bertrand, D, Weerasinghe, A, Colaprico, A, Wendl, MC, Kim, J, Reardon, B, *et al.* (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**: 371-385 e318.
68. Saffie, R, Zhou, N, Rolland, D, Onder, O, Basrur, V, Campbell, S, Wellen, KE, Elenitoba-Johnson, KSJ, Capell, BC, and Busino, L (2020). FBXW7 Triggers Degradation of KMT2D to Favor Growth of Diffuse Large B-cell Lymphoma Cells. *Cancer Res* **80**: 2498-2511.
69. Takeuchi, T, Yamaguchi, M, Kobayashi, K, Miyazaki, K, Tawara, I, Imai, H, Ono, R, Nosaka, T, Tanaka, K, and Katayama, N (2017). MYD88, CD79B, and CARD11 gene mutations in CD5-positive diffuse large B-cell lymphoma. *Cancer* **123**: 1166-1173.
70. Visco, C, Tanasi, I, Quaglia, FM, Ferrarini, I, Fraenza, C, and Krampera, M (2020). Oncogenic Mutations of MYD88 and CD79B in Diffuse Large B-Cell Lymphoma and Implications for Clinical Practice. *Cancers (Basel)* **12**.
71. Singh, M, Jackson, KJL, Wang, JJ, Schofield, P, Field, MA, Koppstein, D, Peters, TJ, Burnett, DL, Rizzetto, S, Nevoltris, D, *et al.* (2020). Lymphoma Driver Mutations in the Pathogenic Evolution of an Iconic Human Autoantibody. *Cell* **180**: 878-894 e819.
72. Kraan, W, Horlings, HM, van Keimpema, M, Schilder-Tol, EJ, Oud, ME, Scheepstra, C, Kluin, PM, Kersten, MJ, Spaargaren, M, and Pals, ST (2013). High prevalence of oncogenic MYD88 and CD79B mutations in diffuse large B-cell lymphomas presenting at immune-privileged sites. *Blood Cancer J* **3**: e139.
73. Muris, JJ, Meijer, CJ, Cillessen, SA, Vos, W, Kummer, JA, Bladergroen, BA, Bogman, MJ, MacKenzie, MA, Jiwa, NM, Siegenbeek van Heukelom, LH, *et al.* (2004). Prognostic significance of activated cytotoxic T-lymphocytes in primary nodal diffuse large B-cell lymphomas. *Leukemia* **18**: 589-596.
74. Tamma, R, Ranieri, G, Ingravallo, G, Annese, T, Oranger, A, Gaudio, F, Musto, P, Specchia, G, and Ribatti, D (2020). Inflammatory Cells in Diffuse Large B Cell Lymphoma. *J Clin Med* **9**.
75. Kim, S, Nam, SJ, Park, C, Kwon, D, Yim, J, Song, SG, Ock, CY, Kim, YA, Park, SH, Kim, TM, *et al.* (2019). High tumoral PD-L1 expression and low PD-1(+) or CD8(+) tumor-infiltrating lymphocytes are predictive of a poor prognosis in primary diffuse large B-cell lymphoma of the central nervous system. *Oncoimmunology* **8**: e1626653.
76. Peng, F, Qin, Y, Mu, S, Li, J, Ai, L, and Hu, Y (2020). Prognostic role of regulatory T cells in lymphoma: a systematic review and meta-analysis. *Journal of cancer research and clinical oncology* **146**: 3123-3135.
77. McCord, R, Bolen, CR, Koeppen, H, Kadel, EE, 3rd, Oestergaard, MZ, Nielsen, T, Sehn, LH, and Venstrom, JM (2019). PD-L1 and tumor-associated macrophages in de novo DLBCL. *Blood Adv* **3**: 531-540.
78. Haro, M, and Orsulic, S (2018). A Paradoxical Correlation of Cancer-Associated Fibroblasts With Survival Outcomes in B-Cell Lymphomas and Carcinomas. *Front Cell Dev Biol* **6**: 98.
79. Autio, M, Leivonen, SK, Bruck, O, Mustjoki, S, Jorgensen, JM, Karjalainen-Lindsberg, ML, Beiske, K, Holte, H, Pellinen, T, and Leppa, S (2020). Immune cell constitution in the tumor microenvironment predicts the outcome in diffuse large B-cell lymphoma. *Haematologica*.
80. Chen, BJ, Dashnamoorthy, R, Galera, P, Makarenko, V, Chang, H, Ghosh, S, and Evens, AM (2019). The immune checkpoint molecules PD-1, PD-L1, TIM-3 and LAG-3 in diffuse large B-cell lymphoma. *Oncotarget* **10**: 2030-2040.
81. Fu, J, Li, K, Zhang, W, Wan, C, Zhang, J, Jiang, P, and Liu, XS (2020). Large-scale public data reuse to model immunotherapy response and resistance. *Genome Med* **12**: 21.
82. Xu-Monette, ZY, Zhou, J, and Young, KH (2018). PD-1 expression and clinical PD-1 blockade in B-cell lymphomas. *Blood* **131**: 68-83.
83. Quan, L, Lan, X, Meng, Y, Guo, X, Guo, Y, Zhao, L, Chen, X, and Liu, A (2018). BTLA marks a less cytotoxic T-cell subset in diffuse large B-cell lymphoma with high expression of checkpoints. *Exp Hematol* **60**: 47-56 e41.

Figure legends

Figure 1. A flow chart for the process of the present study.

Figure 2. Identification of immune-related genes in DLBCL patients with different IPI levels.

(A) PCA analysis. (B) Survival analysis of overall survival between high and low IPI groups. (C) Gene set variance analysis (GSVA) of enriched gene sets among different IPI groups. (D) Volcano plot of differentially expressed genes (DEGs) between the high and low IPI groups.

Figure 3. weighted gene co-expression network analysis (WGCNA) module identification and correlation analysis

(A) Network analysis of gene expression in DLBCL identifies distinct modules of co-expression genes. (B) Analysis and visualization of Module-trait relationship to identify IPI related modules. (C) Correlation between gene module membership and gene significance for IPI in 3 modules.

Figure 4. Construction of an IPI-based immune prognostic model.

(A) Plot of the Lasso penalized Cox regression. (B) A forest plot of the genes composing the IPI-based immune prognostic model (IPI-IPM). (C) time-dependent ROC curves for the IPI-IPM. (D) Survival analysis of overall survival between high and low IPI-IPM risk groups. (E) Heatmap of the gene expression in high and low IPI-IPM risk groups. (F) The multivariate analysis of IPI-IPM risk score and clinicopathologic parameters including Age, Ann Arbor clinical stage, LDH ratio, ECOG performance status and the number of extranodal sites.

Figure 5. Validation of the IPI-IPM and Construction of an IPI-based immune nomogram model.

(A) Nomogram for the prediction of the survival probability of 1-, 3-, and 5-year overall survival. (B) The DCA analysis of all parameters in the nomogram. (C) Calibration plots of nomogram-predicted probability of 1-, 3-, 5-year, and median survival. (D) time-dependent ROC curves for the Nomogram. (E-F) Survival analysis of overall survival between high and low risk groups in the testing (GSE10846) cohort, in patients with R-CHOP like regimens and those with CHOP like regimens (F).

Figure 6. Molecular characteristics of IPI-IPM associated immune genes.

(A) Volcano plot of immune related DEGs between high and low IPI-IPM risk group. (B) Preranked GSEA of enriched gene sets between high and low IPI-IPM risk group. (C) The t-SNE algorithm were applied to show the difference of DLBCL patients between the high and low IPI-IPM risk groups. (D) Pearson correlation analysis of the 8 genes composing the IPI-IPM. (E) Heatmap of immune related genes correlating the IPI-IPM risk groups.

Figure 7. Immune characteristics of IPI-IPM subgroups.

(A and B) Over representative analysis: A Chord map of the enriched Biological processes (A) and a Sankey plot of the enriched Reactome pathways (B). (C) Protein-protein interaction network (PPI) based on the STRING database. (D) Transcription factor network of immune related genes correlating the IPI-IPM risk groups.

Figure 8. Somatic mutational profiles of IPI-IPM subgroups.

(A) Differentially mutated genes between high and low IPI-IPM risk groups. Top 10 Mutated genes (rows) are ordered by mutation rate. The color-coding legends indicates the mutation types and survival status of patients. (B) Lollipop plots for amino acid changes of KMT2D and MUC16. (C) Oncodrive plots of high and low risk IPI-IPM groups.

Figure 9. Tumor immune microenvironment (TME) characteristics of IPI-IPM subgroups.

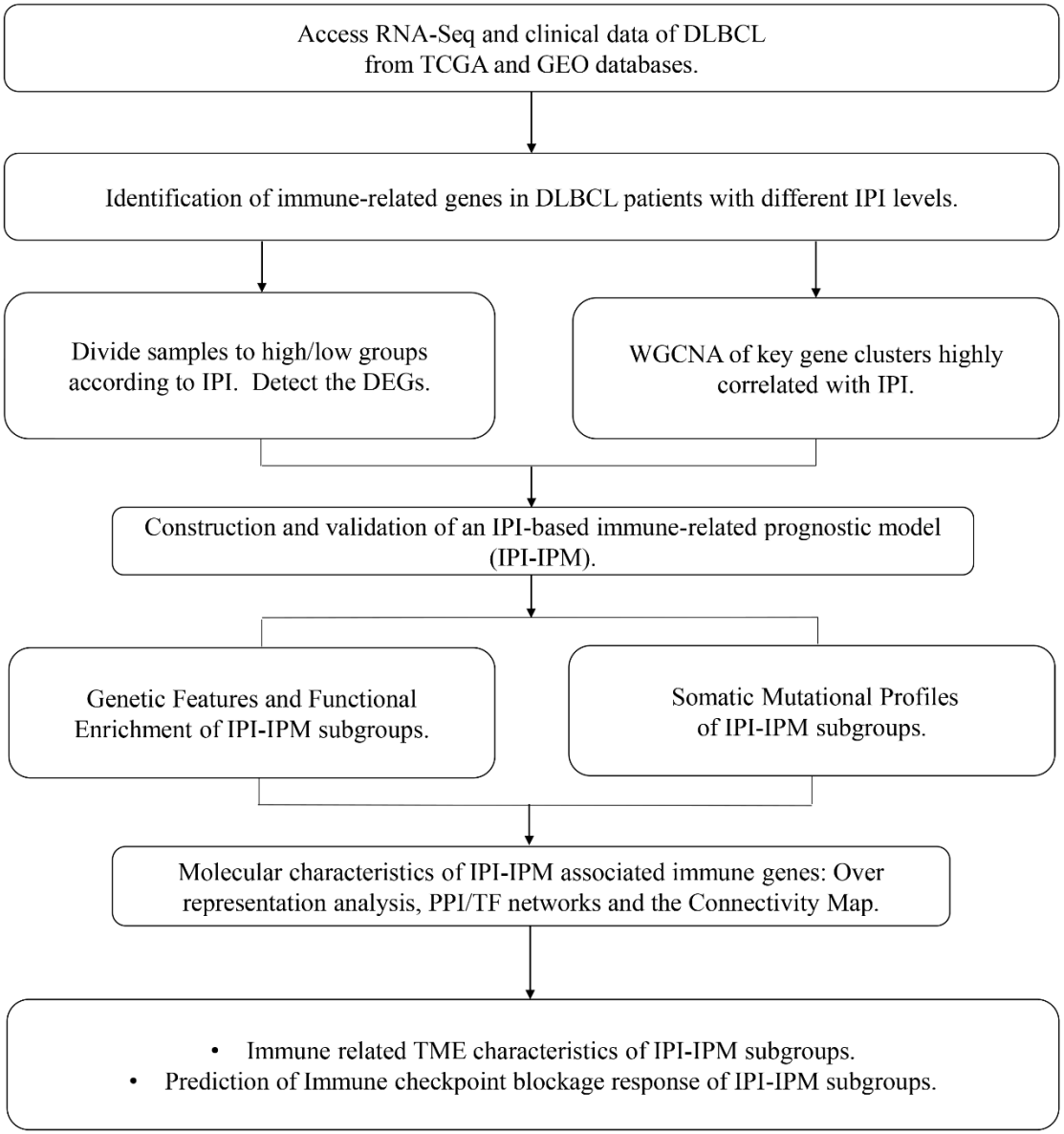
(A-B) Analysis of immune cells infiltration by using the CIBERSORT algorithm: Relative proportion of each type of cells infiltration in DLBCL patients (A) and a bar plot visualizing significantly differentially infiltrated cells between high and low IPI-IPM risk groups (B). (C) Analysis of immune cells infiltration by using the MCPcounter and xCell algorithm.

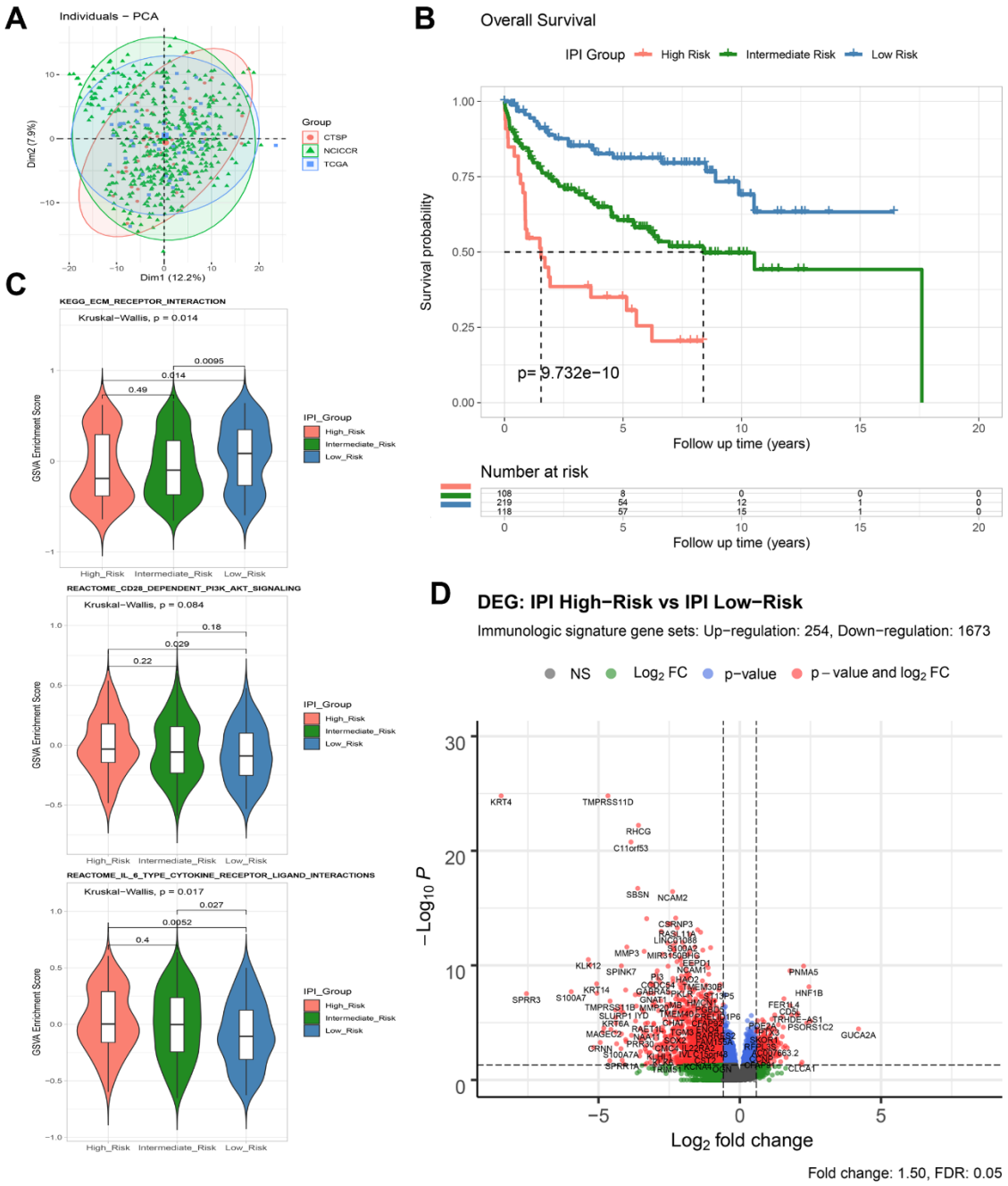
Figure 10. Molecular and microenvironmental subtypes of DLBCL.

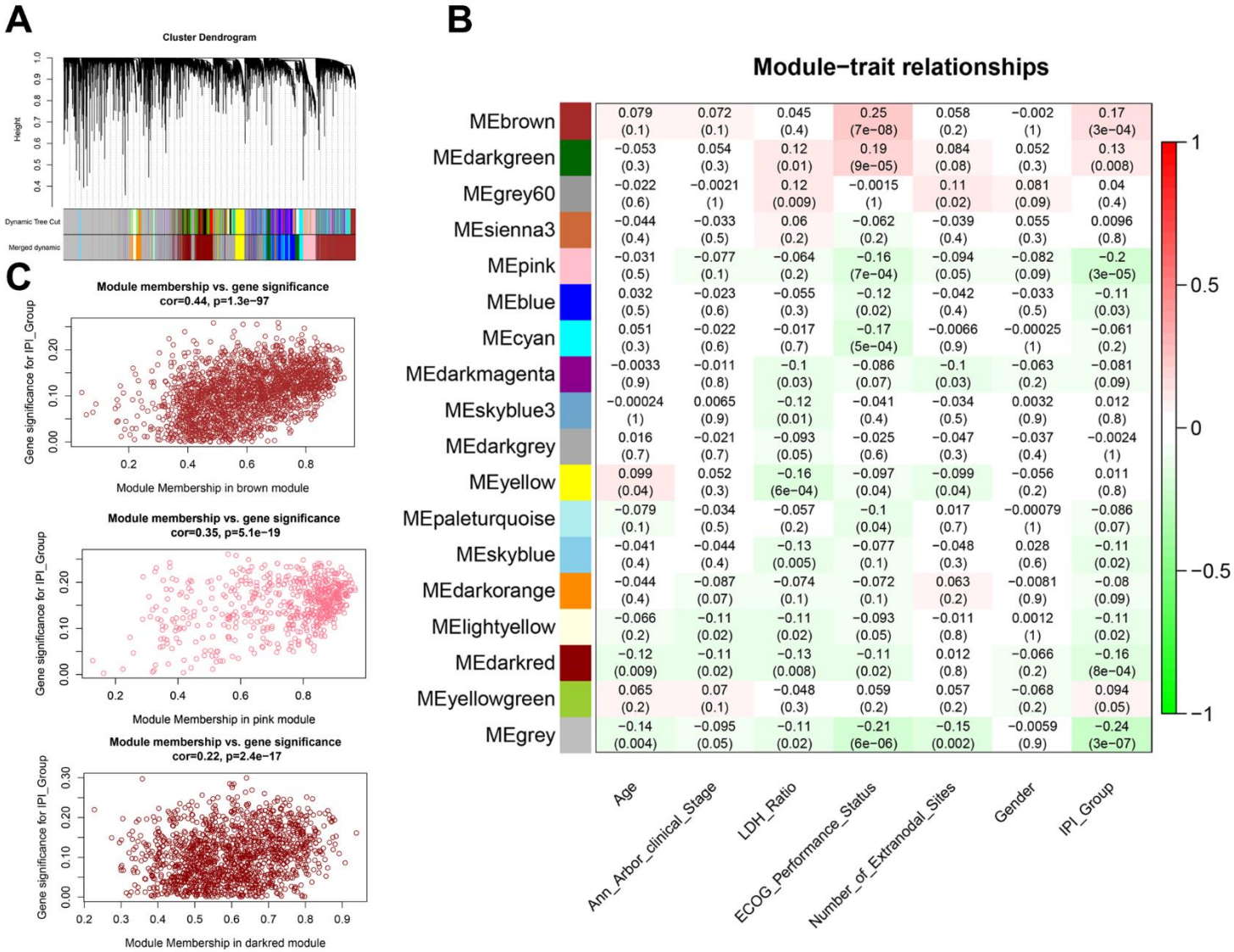
(A) Distribution of IPI groups, gene expression subtypes and genetic subtypes between high and low IPI-IPM risk groups. (B) Correlation analysis of IPI-IPM risk score and Bcl-2, Bcl-6 and c-Myc. (C) Consensus clustering and communities detection of lymphoma microenvironment (LME) clusters. (D-E) Distribution of LME patterns between high and low IPI-IPM risk groups. (F) GSVA enrichment score of LME functional signature between high and low IPI-IPM risk groups.

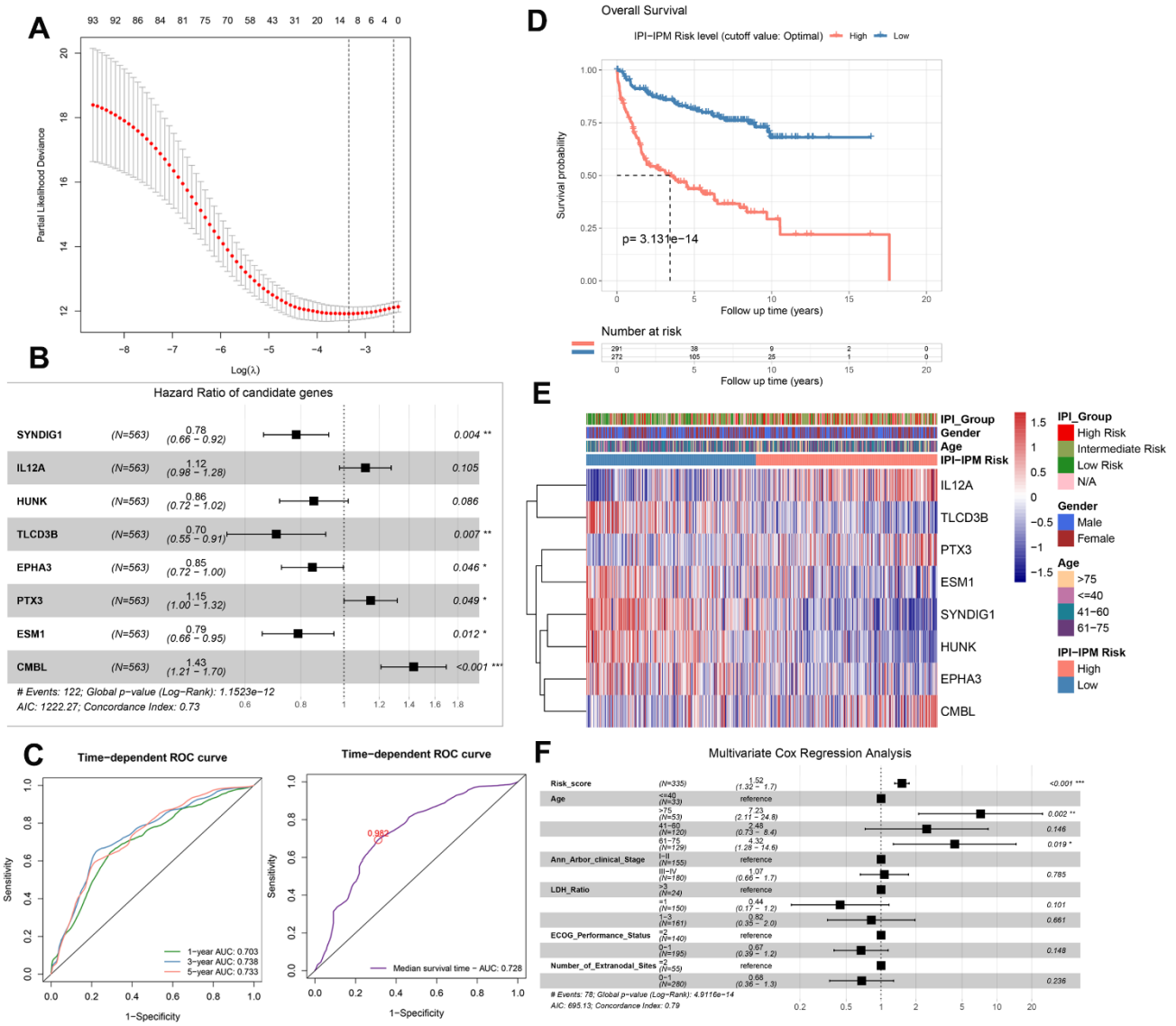
Figure 11. Potential therapeutic value of IPI-IPM.

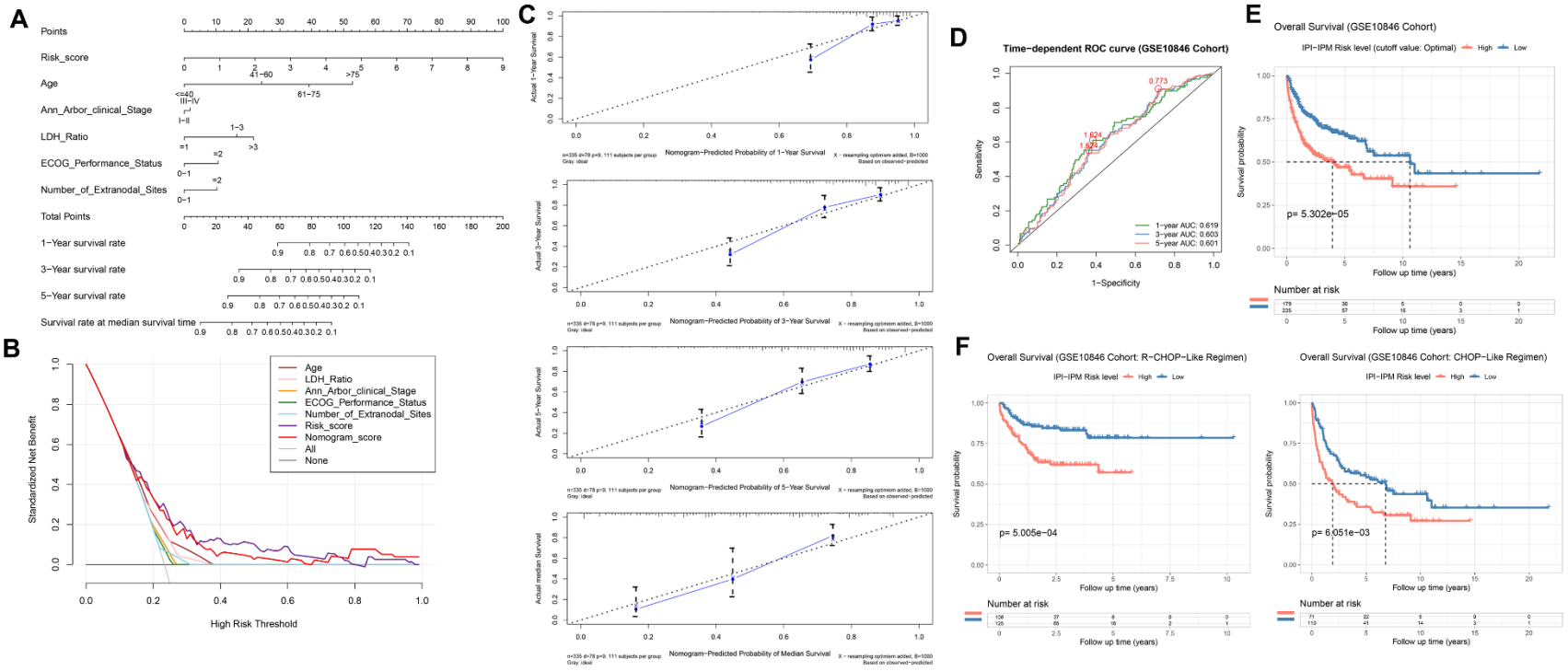
(A) Connectivity map of top enriched immune related genes correlating the IPI-IPM risk groups. (B) The expression of inhibitory immune checkpoints between high and low IPI-IPM risk groups. (C) Survival analysis of IMvigor210 Cohort.

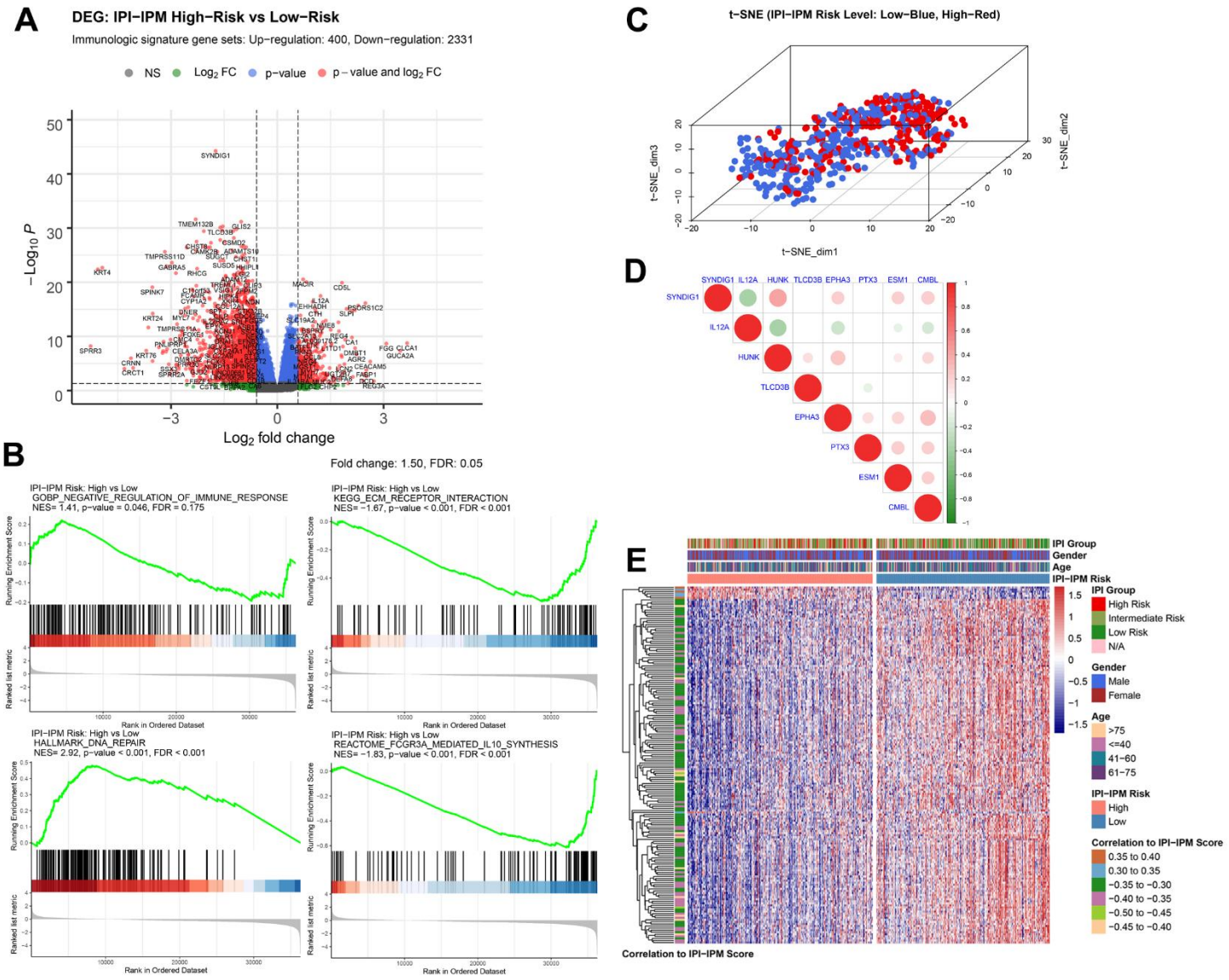

Access RNA-Seq and clinical data of DLBCL from TCGA and GEO databases.
Identification of immune-related genes in DLBCL patients with different IPI levels.
Divide samples to high/low groups according to IPI. Detect the DEGs.
WGCNA of key gene clusters highly correlated with IPI.
Construction and validation of an IPI-based immune-related prognostic model (IPI-IPM).
Genetic Features and Functional Enrichment of IPI-IPM subgroups.
Somatic Mutational Profiles of IPI-IPM subgroups.
Molecular characteristics of IPI-IPM associated immune genes: Over representation analysis, PPI/TF networks and the Connectivity Map.
• Immune related TME characteristics of IPI-IPM subgroups.
• Prediction of Immune checkpoint blockage response of IPI-IPM subgroups.

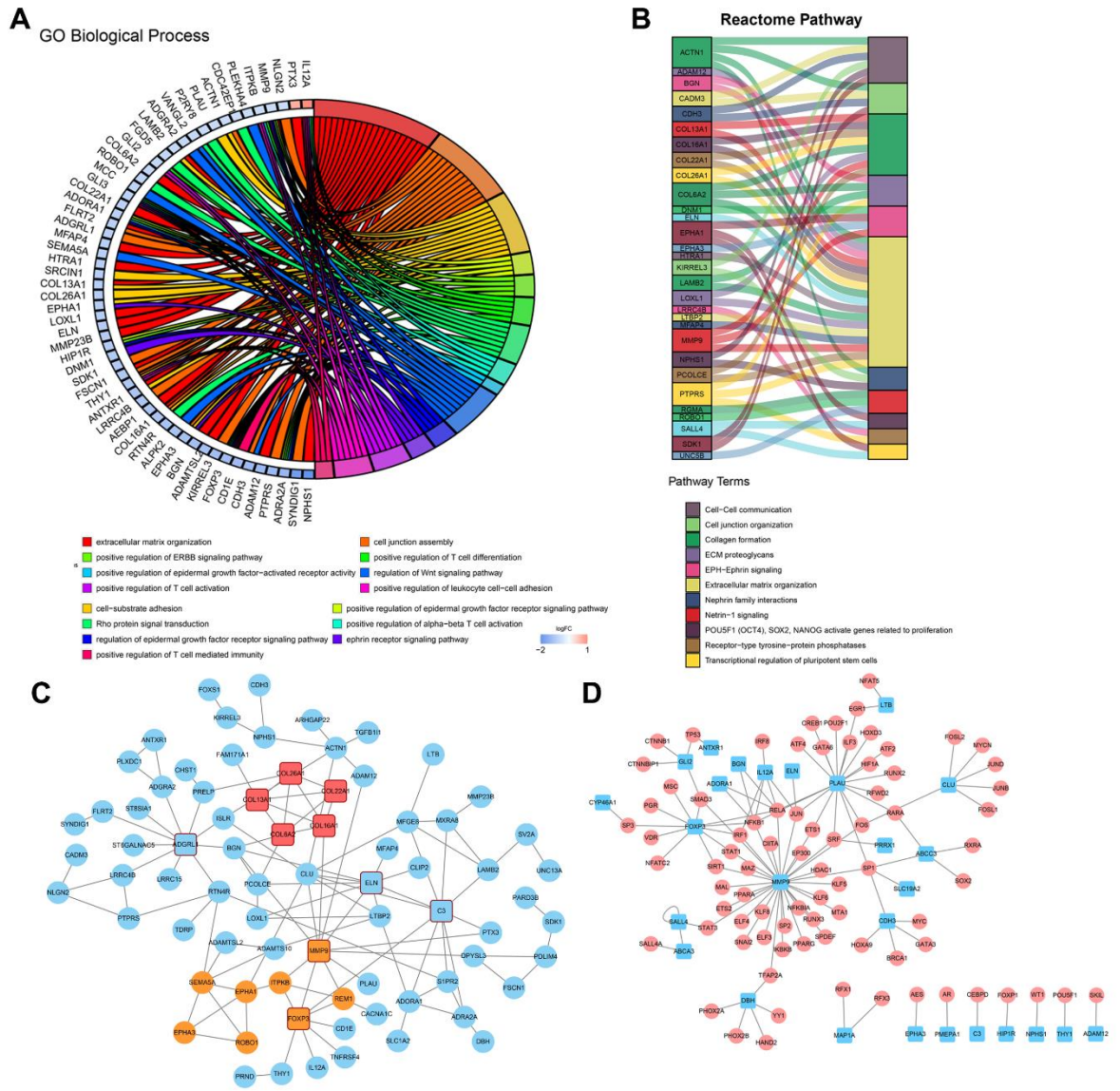

A
Individuals – PCA
Dim1 (12.2%)
Dim2 (7.9%)
Group
CTSP
NCICCR
TCGA
B
Overall Survival
IPI Group
High Risk
Intermediate Risk
Low Risk
Survival probability
p= 9.732e−10
Follow up time (years)
Number at risk
C
KEGG_ECM_RECEPTOR_INTERACTION
Kruskal−Wallis, p = 0.014
REACTOME_CD28_DEPENDENT_PI3K_AKT_SIGNALING
Kruskal−Wallis, p = 0.084
REACTOME_IL_6_TYPE_CYTOKINE_RECEPTOR_LIGAND_INTERACTIONS
Kruskal−Wallis, p = 0.017
GSVA Enrichment Score
IPI_Group
High_Risk
Intermediate_Risk
Low_Risk
D
DEG: IPI High−Risk vs IPI Low−Risk
Immunologic signature gene sets: Up−regulation: 254, Down−regulation: 1673
NS
Log2 FC
p−value
p − value and log2 FC
−Log10 P
Log2 fold change
Fold change: 1.50, FDR: 0.05

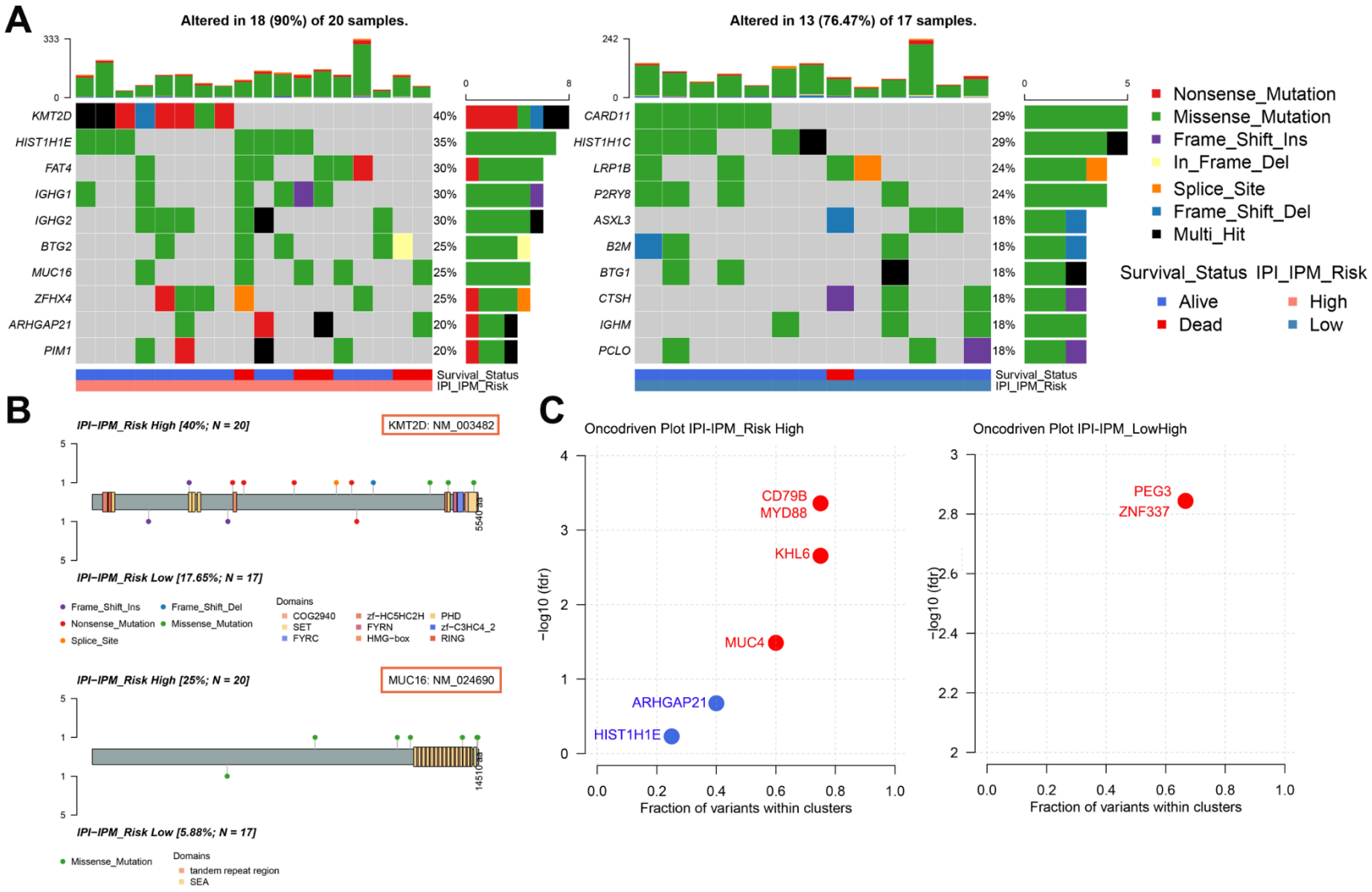

A
Cluster Dendrogram
Height
Dynamic Tree Cut
Merged dynamic
B
Module−trait relationships
MEbrown 0.079 (0.1) 0.072 (0.1) 0.045 (0.4) 0.25 (7e−08) 0.058 (0.2) −0.002 (1) 0.17 (3e−04)
MEdarkgreen −0.053 (0.3) 0.054 (0.3) 0.12 (0.01) 0.19 (9e−05) 0.084 (0.08) 0.052 (0.3) 0.13 (0.008)
MEgrey60 −0.022 (0.6) −0.0021 (1) 0.12 (0.009) −0.0015 (1) 0.11 (0.02) 0.081 (0.09) 0.04 (0.4)
MEsienna3 −0.044 (0.4) −0.033 (0.5) 0.06 (0.2) −0.062 (0.2) −0.039 (0.4) 0.055 (0.3) 0.0096 (0.8)
MEpink −0.031 (0.5) −0.077 (0.1) −0.064 (0.2) −0.16 (7e−04) −0.094 (0.05) −0.082 (0.09) −0.2 (3e−05)
MEblue 0.032 (0.5) −0.023 (0.6) −0.055 (0.3) −0.12 (0.02) −0.042 (0.4) −0.033 (0.5) −0.11 (0.03)
MEcyan 0.051 (0.3) −0.022 (0.6) −0.017 (0.7) −0.17 (5e−04) −0.0066 (0.9) −0.00025 (1) −0.061 (0.2)
MEdarkmagenta −0.0033 (0.9) −0.011 (0.8) −0.1 (0.03) −0.086 (0.07) −0.1 (0.03) −0.063 (0.2) −0.081 (0.09)
MEskyblue3 −0.00024 (1) 0.0065 (0.9) −0.12 (0.01) −0.041 (0.4) −0.034 (0.5) 0.0032 (0.9) 0.012 (0.8)
MEdarkgrey 0.016 (0.7) −0.021 (0.7) −0.093 (0.05) −0.025 (0.6) −0.047 (0.3) −0.037 (0.4) −0.0024 (1)
MEyellow 0.099 (0.04) 0.052 (0.3) −0.16 (6e−04) −0.097 (0.04) −0.099 (0.04) −0.056 (0.2) 0.011 (0.8)
MEpaleturquoise −0.079 (0.1) −0.034 (0.5) −0.057 (0.2) −0.1 (0.04) 0.017 (0.7) −0.00079 (1) −0.086 (0.07)
MEskyblue −0.041 (0.4) −0.044 (0.4) −0.13 (0.005) −0.077 (0.1) −0.048 (0.3) 0.028 (0.6) −0.11 (0.02)
MEdarkorange −0.044 (0.4) −0.087 (0.07) −0.074 (0.1) −0.072 (0.1) 0.063 (0.2) −0.0081 (0.9) −0.08 (0.09)
MElightyellow −0.066 (0.2) −0.11 (0.02) −0.11 (0.02) −0.093 (0.05) −0.011 (0.8) 0.0012 (1) −0.11 (0.02)
MEdarkred −0.12 (0.009) −0.11 (0.02) −0.13 (0.008) −0.11 (0.02) 0.012 (0.8) −0.066 (0.2) −0.16 (8e−04)
MEyellowgreen 0.065 (0.2) 0.07 (0.1) −0.048 (0.3) 0.059 (0.2) 0.057 (0.2) −0.068 (0.2) 0.094 (0.05)
MEgrey −0.14 (0.004) −0.095 (0.05) −0.11 (0.02) −0.21 (6e−06) −0.15 (0.002) −0.0059 (0.9) −0.24 (3e−07)
Age
Ann_Arbor_clinical_Stage
LDH_Ratio
ECOG_Performance_Status
Number_of_Extranodal_Sites
Gender
IPI_Group
1
0.5
0
−0.5
−1
C
Module membership vs. gene significance
cor=0.44, p=1.3e−97
Gene significance for IPI_Group
Module Membership in brown module
Module membership vs. gene significance
cor=0.35, p=5.1e−19
Gene significance for IPI_Group
Module Membership in pink module
Module membership vs. gene significance
cor=0.22, p=2.4e−17
Gene significance for IPI_Group
Module Membership in darkred module

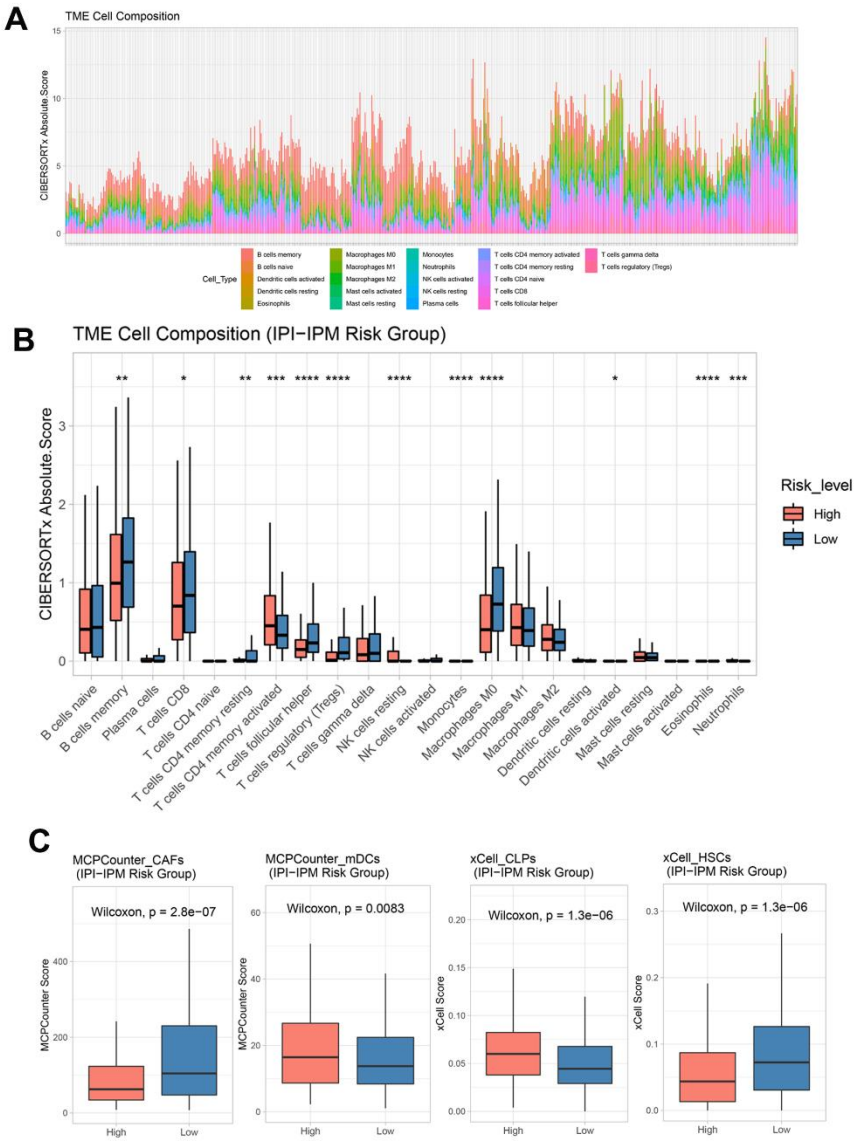

A
93 92 86 84 81 75 70 58 43 31 20 14 8 6 4 0
Partial Likelihood Deviance
Log(λ)
B
Hazard Ratio of candidate genes
SYNDIG1 (N=563) 0.78 (0.66 − 0.92) 0.004 **
IL12A (N=563) 1.12 (0.98 − 1.28) 0.105
HUNK (N=563) 0.86 (0.72 − 1.02) 0.086
TLCD3B (N=563) 0.70 (0.55 − 0.91) 0.007 **
EPHA3 (N=563) 0.85 (0.72 − 1.00) 0.046 *
PTX3 (N=563) 1.15 (1.00 − 1.32) 0.049 *
ESM1 (N=563) 0.79 (0.66 − 0.95) 0.012 *
CMBL (N=563) 1.43 (1.21 − 1.70) <0.001 ***
Events: 122; Global p−value (Log−Rank): 1.1523e−12
AIC: 1222.27; Concordance Index: 0.73
C
Time−dependent ROC curve
1−year AUC: 0.703
3−year AUC: 0.738
5−year AUC: 0.733
Sensitivity
1−Specificity
Time−dependent ROC curve
0.982
Median survival time − AUC: 0.728
D
Overall Survival
IPI−IPM Risk level (cutoff value: Optimal) High Low
Survival probability
p= 3.131e−14
Follow up time (years)
Number at risk
E
IPI_Group
Gender
Age
IPI−IPM Risk
IL12A
TLCD3B
PTX3
ESM1
SYNDIG1
HUNK
EPHA3
CMBL
High Risk
Intermediate Risk
Low Risk
N/A
Male
Female
>75
<=40
41−60
61−75
High
Low
F
Multivariate Cox Regression Analysis
Risk_score (N=335) 1.52 (1.32 − 1.7) <0.001 ***
Age <=40 (N=33) reference
>75 (N=53) 7.23 (2.11 − 24.8) 0.002 **
41−60 (N=120) 2.48 (0.73 − 8.4) 0.146
61−75 (N=129) 4.32 (1.28 − 14.6) 0.019 *
Ann_Arbor_clinical_Stage I−II (N=155) reference
III−IV (N=180) 1.07 (0.66 − 1.7) 0.785
LDH_Ratio >3 (N=24) reference
=1 (N=150) 0.44 (0.17 − 1.2) 0.101
1−3 (N=161) 0.82 (0.35 − 2.0) 0.661
ECOG_Performance_Status >2 (N=140) reference
0−1 (N=195) 0.67 (0.39 − 1.2) 0.148
Number_of_Extranodal_Sites >2 (N=55) reference
0−1 (N=280) 0.68 (0.36 − 1.3) 0.236
Events: 78; Global p−value (Log−Rank): 4.9116e−14
AIC: 695.13; Concordance Index: 0.79

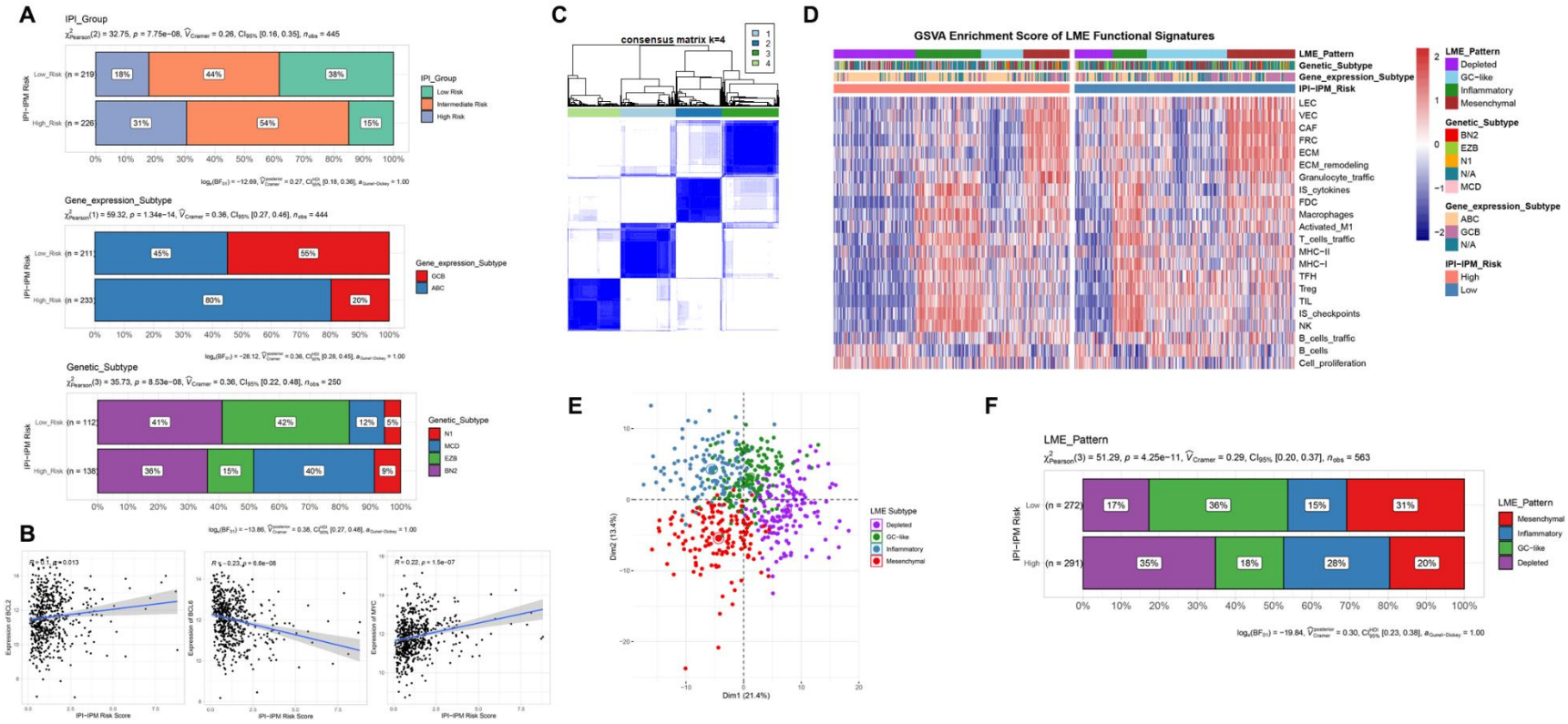

A
Points
Risk_score
Age
Ann_Arbor_clinical_Stage
LDH_Ratio
ECOG_Performance_Status
Number_of_Extranodal_Sites
Total Points
1−Year survival rate
3−Year survival rate
5−Year survival rate
Survival rate at median survival time
B
Standardized Net Benefit
High Risk Threshold
Age
LDH_Ratio
Ann_Arbor_clinical_Stage
ECOG_Performance_Status
Number_of_Extranodal_Sites
Risk_score
Nomogram_score
All
None
C
Actual 1−Year Survival
Nomogram−Predicted Probability of 1−Year Survival
Actual 3−Year Survival
Nomogram−Predicted Probability of 3−Year Survival
Actual 5−Year Survival
Nomogram−Predicted Probability of 5−Year Survival
Actual median Survival
Nomogram−Predicted Probability of Median Survival
D
Time−dependent ROC curve (GSE10846 Cohort)
Sensitivity
1−Specificity
1−year AUC: 0.619
3−year AUC: 0.603
5−year AUC: 0.601
E
Overall Survival (GSE10846 Cohort)
IPI−IPM Risk level (cutoff value: Optimal)
High
Low
Survival probability
p= 5.302e−05
Follow up time (years)
Number at risk
F
Overall Survival (GSE10846 Cohort: R−CHOP−Like Regimen)
IPI−IPM Risk level
p= 5.005e−04
Overall Survival (GSE10846 Cohort: CHOP−Like Regimen)
p= 6.051e−03

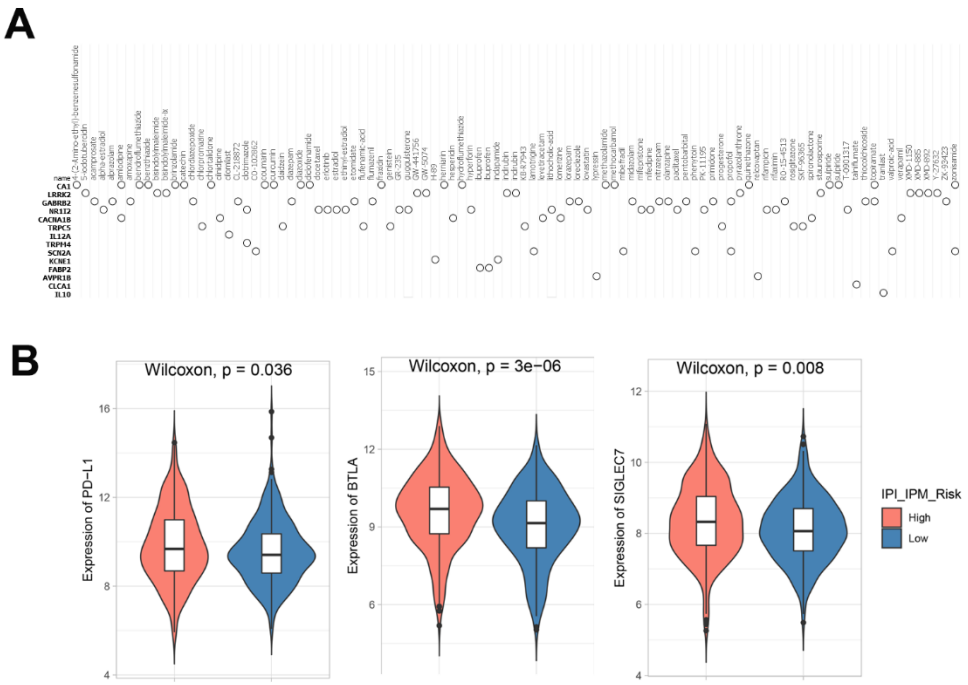

A
DEG: IPI−IPM High−Risk vs Low−Risk
Immunologic signature gene sets: Up−regulation: 400, Down−regulation: 2331
NS
Log2 FC
p−value
p − value and log2 FC
−Log10 P
Log2 fold change
B
Fold change: 1.50, FDR: 0.05
IPI−IPM Risk: High vs Low
GOBP_NEGATIVE_REGULATION_OF_IMMUNE_RESPONSE
NES= 1.41, p−value = 0.046, FDR = 0.175
IPI−IPM Risk: High vs Low
KEGG_ECM_RECEPTOR_INTERACTION
NES= −1.67, p−value < 0.001, FDR < 0.001
IPI−IPM Risk: High vs Low
HALLMARK_DNA_REPAIR
NES= 2.92, p−value < 0.001, FDR < 0.001
IPI−IPM Risk: High vs Low
REACTOME_FCGR3A_MEDIATED_IL10_SYNTHESIS
NES= −1.83, p−value < 0.001, FDR < 0.001
Running Enrichment Score
Ranked list metric
Rank in Ordered Dataset
C
t−SNE (IPI−IPM Risk Level: Low−Blue, High−Red)
t−SNE_dim1
t−SNE_dim2
t−SNE_dim3
D
SYNDIG1
IL12A
HUNK
TLCD3B
EPHA3
PTX3
ESM1
CMBL
E
IPI Group
Gender
Age
IPI−IPM Risk
High Risk
Intermediate Risk
Low Risk
N/A
Male
Female
>75
<=40
41−60
61−75
High
Low
Correlation to IPI−IPM Score
0.35 to 0.40
0.30 to 0.35
−0.35 to −0.30
−0.40 to −0.35
−0.50 to −0.45
−0.45 to −0.40

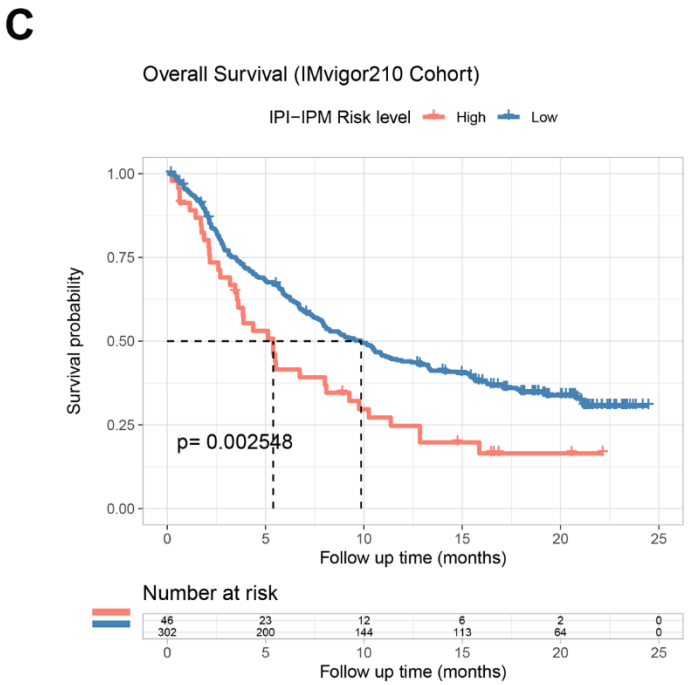

A
GO Biological Process
extracellular matrix organization
cell junction assembly
positive regulation of ERBB signaling pathway
positive regulation of T cell differentiation
positive regulation of epidermal growth factor−activated receptor activity
regulation of Wnt signaling pathway
positive regulation of T cell activation
positive regulation of leukocyte cell−cell adhesion
cell−substrate adhesion
positive regulation of epidermal growth factor receptor signaling pathway
Rho protein signal transduction
positive regulation of alpha−beta T cell activation
regulation of epidermal growth factor receptor signaling pathway
ephrin receptor signaling pathway
positive regulation of T cell mediated immunity
logFC
−2
1
B
Reactome Pathway
Pathway Terms
Cell−Cell communication
Cell junction organization
Collagen formation
ECM proteoglycans
EPH−Ephrin signaling
Extracellular matrix organization
Nephrin family interactions
Netrin−1 signaling
POU5F1 (OCT4), SOX2, NANOG activate genes related to proliferation
Receptor−type tyrosine−protein phosphatases
Transcriptional regulation of pluripotent stem cells
C
D


A
Altered in 18 (90%) of 20 samples.
KMT2D 40%
HIST1H1E 35%
FAT4 30%
IGHG1 30%
IGHG2 30%
BTG2 25%
MUC16 25%
ZFHX4 25%
ARHGAP21 20%
PIM1 20%
Survival_Status
IPI_IPM_Risk
Altered in 13 (76.47%) of 17 samples.
CARD11 29%
HIST1H1C 29%
LRP1B 24%
P2RY8 24%
ASXL3 18%
B2M 18%
BTG1 18%
CTSH 18%
IGHM 18%
PCLO 18%
Nonsense_Mutation
Missense_Mutation
Frame_Shift_Ins
In_Frame_Del
Splice_Site
Frame_Shift_Del
Multi_Hit
Survival_Status
Alive
Dead
IPI_IPM_Risk
High
Low
B
IPI−IPM_Risk High [40%; N = 20]
KMT2D: NM_003482
5540 aa
IPI−IPM_Risk Low [17.65%; N = 17]
Frame_Shift_Ins
Frame_Shift_Del
Nonsense_Mutation
Missense_Mutation
Splice_Site
Domains
COG2940
SET
FYRC
zf−HC5HC2H
FYRN
HMG−box
PHD
zf−C3HC4_2
RING
IPI−IPM_Risk High [25%; N = 20]
MUC16: NM_024690
14510 aa
IPI−IPM_Risk Low [5.88%; N = 17]
Missense_Mutation
Domains
tandem repeat region
SEA
C
Oncodriven Plot IPI-IPM_Risk High
CD79B
MYD88
KHL6
MUC4
ARHGAP21
HIST1H1E
−log10 (fdr)
Fraction of variants within clusters
Oncodriven Plot IPI-IPM_LowHigh
PEG3
ZNF337
−log10 (fdr)
Fraction of variants within clusters


A
TME Cell Composition
CIBERSORTx Absolute.Score
Cell_Type
B cells memory
B cells naive
Dendritic cells activated
Dendritic cells resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast cells activated
Mast cells resting
Monocytes
Neutrophils
NK cells activated
NK cells resting
Plasma cells
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD4 naive
T cells CD8
T cells follicular helper
T cells gamma delta
T cells regulatory (Tregs)
B
TME Cell Composition (IPI−IPM Risk Group)
CIBERSORTx Absolute.Score
Risk_level
High
Low
B cells naive
B cells memory
Plasma cells
T cells CD8
T cells CD4 naive
T cells CD4 memory resting
T cells CD4 memory activated
T cells follicular helper
T cells regulatory (Tregs)
T cells gamma delta
NK cells resting
NK cells activated
Monocytes
Macrophages M0
Macrophages M1
Macrophages M2
Dendritic cells resting
Dendritic cells activated
Mast cells resting
Mast cells activated
Eosinophils
Neutrophils
C
MCPCounter_CAFs (IPI−IPM Risk Group)
Wilcoxon, p = 2.8e−07
MCPCounter_mDCs (IPI−IPM Risk Group)
Wilcoxon, p = 0.0083
xCell_CLPs (IPI−IPM Risk Group)
Wilcoxon, p = 1.3e−06
xCell_HSCs (IPI−IPM Risk Group)
Wilcoxon, p = 1.3e−06
MCPCounter Score
xCell Score
High
Low


A
IPI_Group
IPI−IPM Risk
Low_Risk (n = 219)
High_Risk (n = 226)
18%
44%
38%
31%
54%
15%
Low Risk
Intermediate Risk
High Risk
Gene_expression_Subtype
Low_Risk (n = 211)
High_Risk (n = 233)
45%
55%
80%
20%
GCB
ABC
Genetic_Subtype
Low_Risk (n = 112)
High_Risk (n = 138)
41%
42%
12%
5%
36%
15%
40%
9%
N1
MCD
EZB
BN2
B
IPI−IPM Risk Score
C
consensus matrix k=4
1
2
3
4
D
GSVA Enrichment Score of LME Functional Signatures
LME_Pattern
Genetic_Subtype
Gene_expression_Subtype
IPI−IPM_Risk
LEC
VEC
CAF
FRC
ECM
ECM_remodeling
Granulocyte_traffic
IS_cytokines
FDC
Macrophages
Activated_M1
T_cells_traffic
MHC−II
MHC−I
TFH
Treg
TIL
IS_checkpoints
NK
B_cells_traffic
B_cells
Cell_proliferation
Depleted
GC−like
Inflammatory
Mesenchymal
BN2
EZB
N1
N/A
MCD
ABC
GCB
High
Low
E
LME Subtype
Depleted
GC−like
Inflammatory
Mesenchymal
Dim1 (21.4%)
Dim2 (13.4%)
F
LME_Pattern
IPI−IPM Risk
Low (n = 272)
High (n = 291)
17%
36%
15%
31%
35%
18%
28%
20%
Mesenchymal
Inflammatory
GC−like
Depleted


A
B
Wilcoxon, p = 0.036
Wilcoxon, p = 3e−06
Wilcoxon, p = 0.008
Expression of PD−L1
Expression of BTLA
Expression of SIGLEC7
IPI_IPM_Risk
High
Low


C
Overall Survival (IMvigor210 Cohort)
IPI−IPM Risk level
High
Low
Survival probability
p= 0.002548
Follow up time (months)
Number at risk
46 23 12 6 2 0
302 200 144 113 64 0
0 5 10 15 20 25