Current Medical Science
DOI https://doi.org/10.1007/s11596-022-2612-6
42(4):754-768,2022

PHOSPHO1 Serves as a Key Metabolism-Related Biomarker in the Tumorigenesis of Diffuse Large B-cell Lymphoma*

Tian-rui CHEN[1, 2], Huang-ming CAO[2], Yin WU[3], Jiang-tao XIE[1], Hai-feng LAN[4#], Li-na JIN[5#]
[1]*Department of Medicine, Shanghai Di An Medical Laboratory Co., Ltd , Shanghai 200433, China*
[2]*Department of Orthopedic Oncology, Changzheng Hospital, Naval Medical University, Shanghai 200003, China*
[3]*Department of Oncology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China*
[4]*Department of Hematology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China*
[5]*Department of Hematology, Myeloma & Lymphoma Center, Changzheng Hospital, Naval Medical University, Shanghai 200003, China*

[Abstract] Objective: Diffuse large B-cell lymphoma (DLBCL) is an aggressive type of non-Hodgkin lymphoma. Due to its genetic heterogeneity and abnormal metabolism, many DLBCL patients have a poor prognosis. This study investigated the key metabolism-related genes and potential mechanisms. **Methods:** Differentially expressed genes, differentially expressed transcription factors (TFs), and differentially expressed metabolism-related genes (DEMRGs) of glucose and lipid metabolic processes were identified using the edgeR package. Key DEMRGs were screened by Lasso regression, and a prediction model was constructed. The cell type identification by estimating relative subsets of RNA transcripts algorithm was utilized to assess the fraction of immune cells, and Gene Set Enrichment Analysis was used to determine immune-related pathways. A regulatory network was constructed with significant co-expression interactions among TFs, DEMRGs, immune cells/pathways, and hallmark pathways. **Results:** A total of 1551 DEMRGs were identified. A prognostic model with a high applicability (area under the curve=0.921) was constructed with 13 DEMRGs. Tumorigenesis of DLBCL was highly related to the neutrophil count. Four DEMRGs (PRXL2AB, CCN1, DECR2 and PHOSPHO1) with 32 TF–DEMRG, 36 DEMRG–pathway, 14 DEMRG–immune-cell, 9 DEMRG–immune-gene-set, and 67 DEMRG–protein-chip interactions were used to construct the regulatory network. **Conclusion:** We provided a prognostic prediction model based on 13 DEMRGs for DLBCL. We found that phosphatase, orphan 1 (PHOSPHO1) is positively regulated by regulatory factor X5 (RFX5) and mediates MYC proto-oncogene (MYC) targeting the V2 pathway and neutrophils.
Key words: diffuse large B-cell lymphoma; metabolism-related gene; immune microenvironment; regulatory network; PHOSPHO1

Lymphoma refers to the most common blood malignancy, including Hodgkin lymphoma and non-Hodgkin lymphoma (NHL)[1, 2]. Being the most aggressive among various kinds of NHL, diffuse large B-cell lymphoma (DLBCL) accounts for approximately 31% of patients newly diagnosed worldwide[2, 3]. In the USA, more than 30 000 people develop DLBCL every year. The primary therapeutic method for DLBCL is the combination of monoclonal antibody treatment and chemotherapy; for instance, the most frequently used regimen is rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP)[4]. Under most circumstances, despite being an aggressive malignancy, the initial treatment is effective. Rare tumor recurrence has been observed in patients after treatment, with a 3-year cancer-free survival rate of about 60%[4]. Nevertheless, owing to the genetic heterogeneity of this malignant lymphoma to a certain extent, most patients with DLBCL become refractory or relapse. Currently, the therapeutic regimens to improve clinical outcomes have mainly involved increased doses of standard agents in the context of autologous stem cell transplantation[5]. Hence, the identification of genetic heterogeneity and abnormalities is of urgent medical need to discover novel targets for DLBCL therapy.

Metabolic reprogramming is a significant characteristic of DLBCL, which may be caused by gene mutations and genetic abnormalities[6]. The alteration of metabolism influences the energy stores, stimulates the tumor extracellular environment, modulates cell

Tian-rui CHEN, E-mail: chentr@smmu.edu.cn
#Corresponding authors, Li-na JIN, Email: jinln2008@163.com; Hai-feng LAN, Email: lanhf2013@163.com
*This study was supported in part by the National Natural Science Foundation of China (No. 81702849).

proliferation and survival, and induces resistance to chemoradiotherapy[7]. Therefore, studies on metabolism-based mechanisms may provide critical insight into understanding tumor initiation and progression of DLBCL, benefiting the diagnosis and treatment of this disease. However, the associations between metabolism-related biomarkers and tumor progression as well as the corresponding target-directed treatment have rarely been explored in DLBCL. This study is innovative because it not only aimed to analyze the metabolism-related DLBCL biomarkers but also uncovered individualized therapeutic methods.

In this study, based on the RNA profiles of normal peripheral blood mononuclear cells (PBMCs) and DLBCL samples, a transcriptome bioinformatics analysis was conducted to explore key transcription factors (TFs), differentially expressed genes (DEGs), downstream DLBCL-related hallmark pathways, and essential ligand-receptor interactions. Additionally, potential TFs and pathways of key metabolism-related genes were integrated to construct a regulatory network, thus providing novel therapeutic targets and candidate predictors for DLBCL.

1 MATERIALS AND METHODS

1.1 Data Collection

The RNA sequencing profiles of 47 DLBCL patients were obtained from The Cancer Genome Atlas (TCGA) database as the experimental group (https://tcga-data.nci.nih.gov)[8]. The control group consisted of 337 normal PBMC samples obtained from the Genotype-Tissue Expression Portal (GTEx, https://commonfund.nih.gov/GTEx/)[9]. The demographics (age, sex, ethnicity, etc.), survival endpoints (vital status, death, and last follow-up), clinical stages, and histological types of these samples were also extracted. In addition, samples with incomplete clinical information were excluded. The batch effects introduced by comprehensive analysis of two datasets were diminished using the "normalize between array" function in the limma package. DLBCL-related hallmark signaling pathways were collected from the Molecular Signatures Database (MSigDB) v7.1 (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp)[10], and 318 TFs were retrieved from Cistrome Database (http://cistrome.org/)[11]. A total of 1551 metabolism-related genes were retrieved from two metabolism-related Gene Ontology (GO) pathways (http://www.geneontology.org/), including GO_GLUCOSE_METABOLIC_PROCESS and GO_LIPID_METABOLIC_PROCESS[12]. Reverse-phase protein array (RPPA) protein chip data were collected from TCGA database.

1.2 Identification of Differentially Expressed Genes and Functional Annotation

The "edgeR" R package was used to identify DEGs between normal PBMC samples and DLBCL samples. |Log2 Fold Change (FC)| >1.0 and False Discovery Rate (FDR) <0.05 served as the screening criteria for DEG selection[13]. The heatmap and volcano plot of DEGs were constructed, while differentially expressed RPPA protein chips were also identified.

GO enrichment analysis was conducted to explore the potential function associated with DEGs including biological processes (BPs), cellular components (CCs), and molecular functions (MFs). In addition, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis was utilized to explore the relevant functions. Both results were displayed by bar charts using the "cluster Profiler" R package with thresholds of P<0.01 and FDR <0.05[14].

1.3 Gene Set Over-representation Analysis

Gene Set Over-representation Analysis (GSORA) was used to evaluate the fraction of genes of interest (e.g., DEGs) that belongs to the test clusters (e.g., hallmark pathways). In this study, 54 DLBCL-related hallmark signaling pathways downloaded from the MSigDB were categorized into 9 clusters based on similar functional characteristics, including cluster C1 to 8 and H[10]. Furthermore, GSORA was performed to identify the functional enrichment of DEGs in the MSigDB gene sets (https://github.com/tomastokar/gsoap).

1.4 Identification of Differentially Expressed Metabolism-Related Genes

To identify the tumor metabolism in DLBCL, differential expression analysis was applied to explore metabolism-related genes. Differentially expressed metabolism-related genes (DEMRGs) were defined with thresholds of P<0.05 and |log2 FC| >1. The heatmap and volcano plot were constructed to illustrate the expression levels of DEMRGs.

1.5 Construction and Evaluation of the Prognostic Prediction Model

We followed the methods of Guo *et al*[15]. Least Absolute Shrinkage and Selection Operator (LASSO) regression[16] was utilized to reduce the over-fitting phenomenon and to identify the top 20 DEMRGs with great significance, which were incorporated into a proportional hazards model. In the proportional hazards model, the cross-validation for tuning parameter selection was performed. The β value represented the regression coefficient of the corresponding integrated DEMRG in this model. The risk score was calculated for the gene expression level with the following formula[17]:

$$\text{Risk score}=\beta_1\times\text{gene}_1+\beta_2\times\text{gene}_2+\beta_3\times\text{gene}_3\ldots\ldots+\beta_n\times\text{gene}_n$$

Specifically, the order number of relevant genes in the model was defined as "n"; the regression coefficient of a gene was defined as "β"; and gene indicated the

expression level of the nth key gene for each sample, respectively.

Moreover, according to the median risk score, samples were divided into low- and high-risk groups. The accuracy of this model was evaluated using receiver operator characteristic (ROC) curves. In addition, a Kaplan–Meier curve was made to assess the prognostic value of the multivariate model. Then the risk curve and scatter plot were constructed to reorder these samples. In order to test the prognostic value of the risk score in relation to demographics, clinical stages, and TNM stages, multivariate Cox analysis was performed.

1.6 Identification of Potential Immune Cells, Immune Gene Sets, and Hallmark Gene Sets

Due to the crucial roles of the immune microenvironment in the tumorigenesis of DLBCL, we further evaluated the fraction of immune cells and immune gene sets between normal PBMC and DLBCL blood, utilizing cell type identification by estimating the relative subsets of RNA transcripts (CIBERSORT). Samples with a CIBERSORT output of $P<0.05$ were extracted. Further, the Wilcoxon rank-sum test was performed to explore the relationship between immune cells and DLBCL. Single-sample Gene Set Enrichment Analysis (ssGSEA) was also performed to quantify 22 immune-related gene sets in each sample. Then we performed Pearson correlation analysis to identify the correlation characteristics of key DEMRGs with immune cells and immune gene sets.

Next, Gene Set Variation Analysis (GSVA) and Gene Set Enrichment Analysis (GSEA) were used to explore potential downstream hallmark pathways of DEMRGs. The absolute quantification of 50 hallmark signaling pathways was evaluated by GSVA to select differentially expressed pathways between normal PBMC and DLBCL samples. GSEA was conducted to reveal the noteworthy enrichment of upregulated and downregulated hallmark gene sets in normal PBMC and DLBCL samples. Furthermore, correlations of hallmark gene sets of GSVA and GSEA were extracted, and the interactional pathways were considered as key pathways in DLBCL.

1.7 Identification of Upstream TFs

To further explore the potential mechanism of DLBCL, a total of 318 cancer-related TFs were obtained from the Cistrome Database (http://cistrome.org/)[11]. Co-expression analysis was conducted to identify the upstream TFs that were significantly associated with the key DEMRGs. TFs with correlation coefficients greater than 0.50 were extracted for subsequent analysis.

1.8 Co-expression Analysis of TFs, DEMRGs, Protein Chips, Immune Genes, Immune Cells, and Signaling Pathways

Among DEMRGs, TFs, protein chips, immune genes, immune gene sets, and hallmark gene sets, Pearson correlation analysis was conducted to reconstruct the regulatory network using Cytoscape (3.7.1). Further, the protein–protein interaction (PPI) network was then established utilizing the String database[18]. The interaction relationships between DEMRGs and the remaining components were controlled with $P<0.05$ and |correlation coefficient| >0.40.

1.9 ChIP-seq Data Analysis

The ChIP-seq data of histone H3 lysine 27 acetylation (H3K27ac) including 4 DLBLC samples were downloaded from the Sequence Read Archive (SRA) database (Accession ID: SRP309761), while two GM12878 cell line samples were also downloaded from the SRA database (Accession ID: SRP007993) as the control samples. After removing the duplicate and low-quality reads by FastaQC and Trimmomatic software, the clean data were aligned to the reference genome file (version: Ensemble GRCh38) by HIAST2 software and sorted by SAMTools software. Only the unique mapped read was used for peak calling by MACS2 with the parameter of $q<0.05$. ChIPseeker was used for peak assignment annotation. Differentially bound region identification was performed by the DiffBind package, with FDR<0.05 and log 2-fold-change >1 or ≤1. Besides, the bamCoverage function from deepTools software was used to generate the bigwig files, which were visualized by the R package of Givz.

1.10 Statistics Analysis

In this study, it was statistically significant only when a two-sided *P*-value was less than 0.05. All statistical analyses were performed with R software (Institute for Statistics and Mathematics, Vienna, Austria; www.r-project.org, version 3.6.1).

2 RESULTS

2.1 DEGs and Functional Annotation

The analytical procedure is shown in fig. 1. In total, 15 303 DEGs were identified between 47 DLBCL samples and 337 normal PBMC samples, including 8722 upregulated genes and 6581 downregulated genes. The results are displayed in a heatmap plot (fig. 2A) and a volcano plot (fig. 2B).

GO and KEGG analyses were performed to annotate the functions of DEGs. The results showed that the most significant GO terms for BPs, CCs, and MFs were ncRNA metabolic process, mitochondrial inner membrane, and sulfur compound binding, respectively (fig. 2C). The results of KEGG analysis revealed that the functional similarities were mainly enriched in cytokine–cytokine receptor interactions (fig. 2D).

2.2 Gene Set Overrepresentation Analysis

In the cluster of C1 to 8 and H, the hallmarks

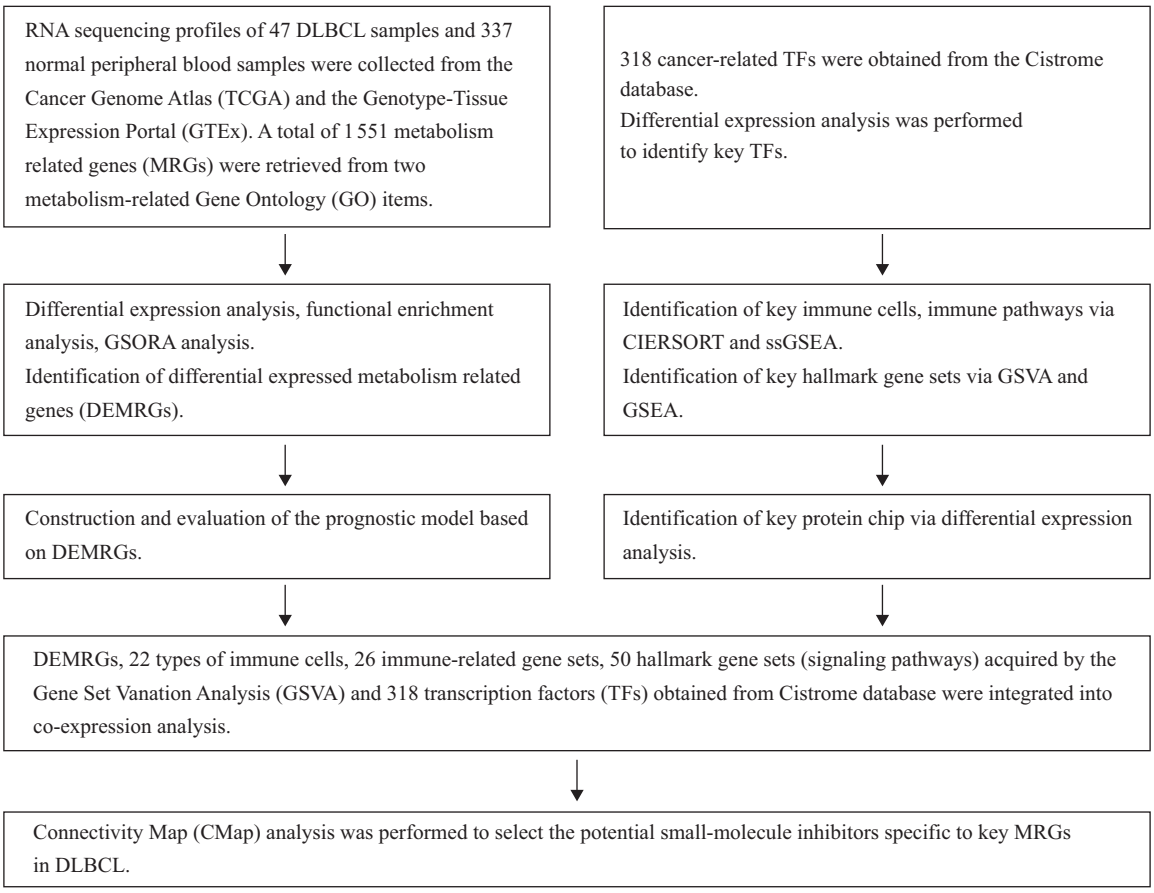


Fig. 1 The flow chart of the analytical processes

in which DEGs were mostly enriched were chr5q31, WONG embryonic stem cell core, ALKBH3 target genes, MODULE 54, GO ncRNA processing, RPS14 DN V1 DN, GSE22886 naive B cell *vs.* neutrophil DN, HAY bone marrow neutrophil, and HALLMARK E2F targets, respectively (fig. 3A). Further, the DEG enrichment degree of HAY bone marrow neutrophils ranked first among all hallmarks, indicating that the bone marrow immune microenvironment played a critical role in DLBCL. Subsequently, a total of 36 prominent gene sets/pathways were selected (fig. 3B). Importantly, apart from gene set chr7p15, almost all gene sets/pathways were significantly activated in the pathogenesis of DLBCL and were related to cell metabolism (REACTOME rRNA processing[19], GO ribosome biogenesis[20], GO ribosomal subunit[20], and hallmark HEME metabolism[21]), immunity (GO secretory granule membrane[22], GO specific granule[22], GSE22886 naive B cell *vs.* neutrophil DN, GSE29618 monocyte *vs.* PDC UP, GSE18893 TCONV *vs.* Treg 24h TNF STIM UP, HAY bone marrow neutrophil, and hallmark TNFA signaling via NFKB), DNA repair (ALKBH3 target genes)[23], and cell cycle progression (hallmark E2F targets)[24].

2.3 Construction of Multivariate Prognostic Model

The expression levels of 1551 metabolism-related genes from two GO gene sets (GO GLUCOSE METABOLIC PROCESS and GO LIPID METABOLIC PROCESS) were analyzed. Based on them, a total of 853 DEMRGs were identified. The heatmap plot (fig. 4A) and volcano plot (fig. 4B) showed their distribution between normal PBMC samples and DLBCL samples.

To avoid over-fitting, Lasso regression was conducted, and the top 20 DEMRGs were identified (fig 4C and 4D). The results showed that GCKR, ACOT4, ADORA1, ANKRD1, CCN1, DECR2, EFR3B, GPC6, IDI2, PHOSPHO1, PLA2G4E, PRXL2B, and PTGDS were extracted for multivariate Cox regression analysis. These 13 DEMRGs were incorporated in a proportional hazards model to assess the prognosis of DLBCL, in which 6 DEMRGs (PLA2G4E, PHOSPHO1, EFR3B, IDI2, DECR2, and ACOT4) were considered as risk factors while 7 DEMRGs (ADORA1, PRXL2B, CCN1, GPC6, PTGDS, ANKRD1, and GCKR) were selected as protective factors (fig. 4E).

2.4 Evaluation of Prognostic Model and Identification of Independent Prognostic Factors

According to the median of the risk score, DLBCL samples were divided into low-risk and high-risk groups. The scatter dot plots indicated that high-risk DLBCL patients had worse survival outcomes than those with a low-risk score based on the prognostic-related DEMRG signature (fig. 5A). A significantly different risk score distribution was observed between

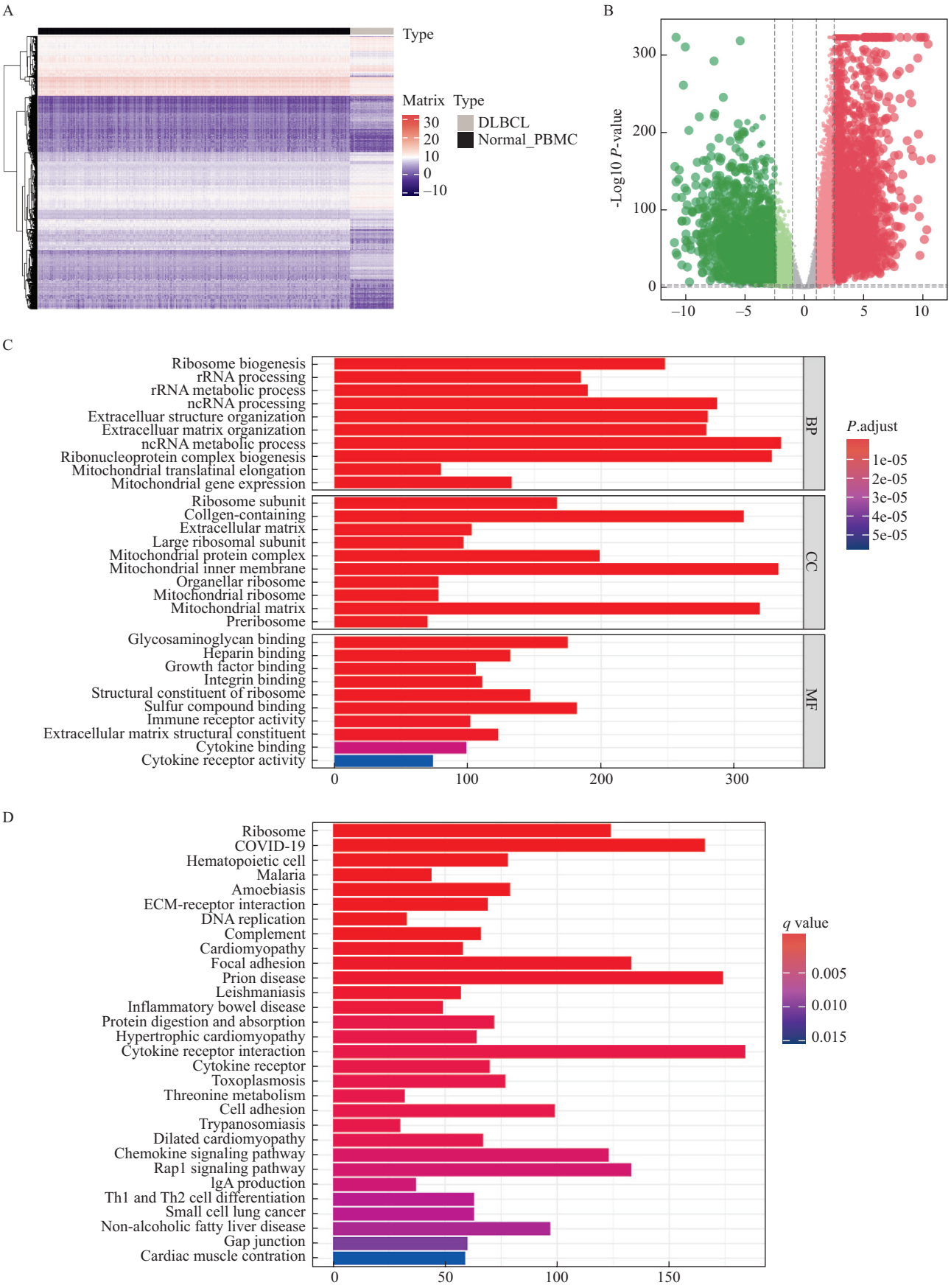


Fig. 2 The differentially expressed genes (DEGs) and functional enrichment analysis between diffuse large B-cell lymphoma (DLBCL) samples and normal peripheral blood mononuclear cells (PBMCs)

The heatmap (A) and volcano plot (B) for DEGs between DLBCL and normal control samples. The functional enrichment analysis for DEGs in GO terms (C) and KEGG pathways (D). GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes

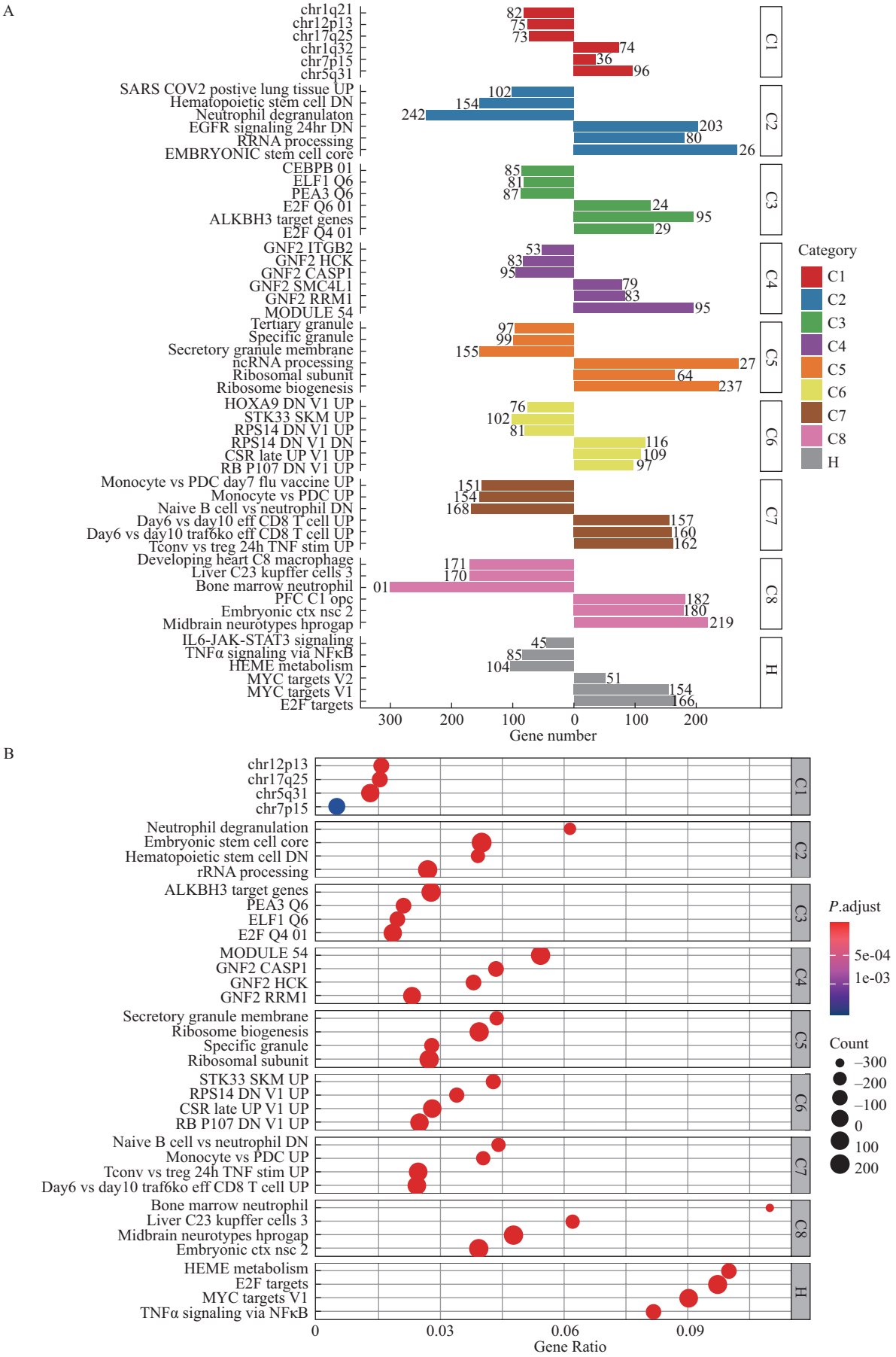


Fig. 3 The gene set over-representation analysis (GSORA) of differentially expressed genes (DEGs) and hallmark pathways
A: the bar plot for GSORA of DEGs and significant hallmark pathways; B: the dot plot for GSORA of DEGs and significant hallmark pathways

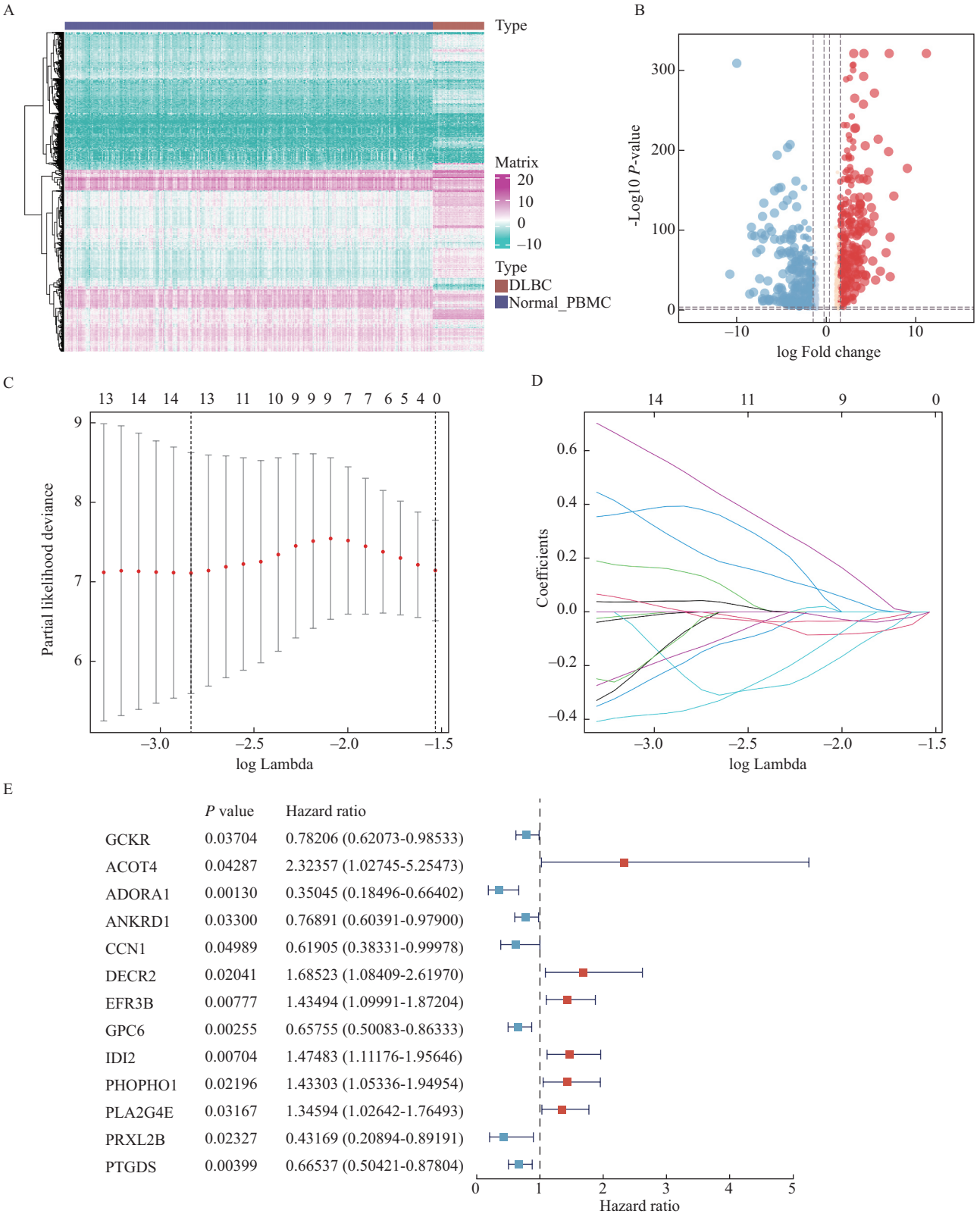
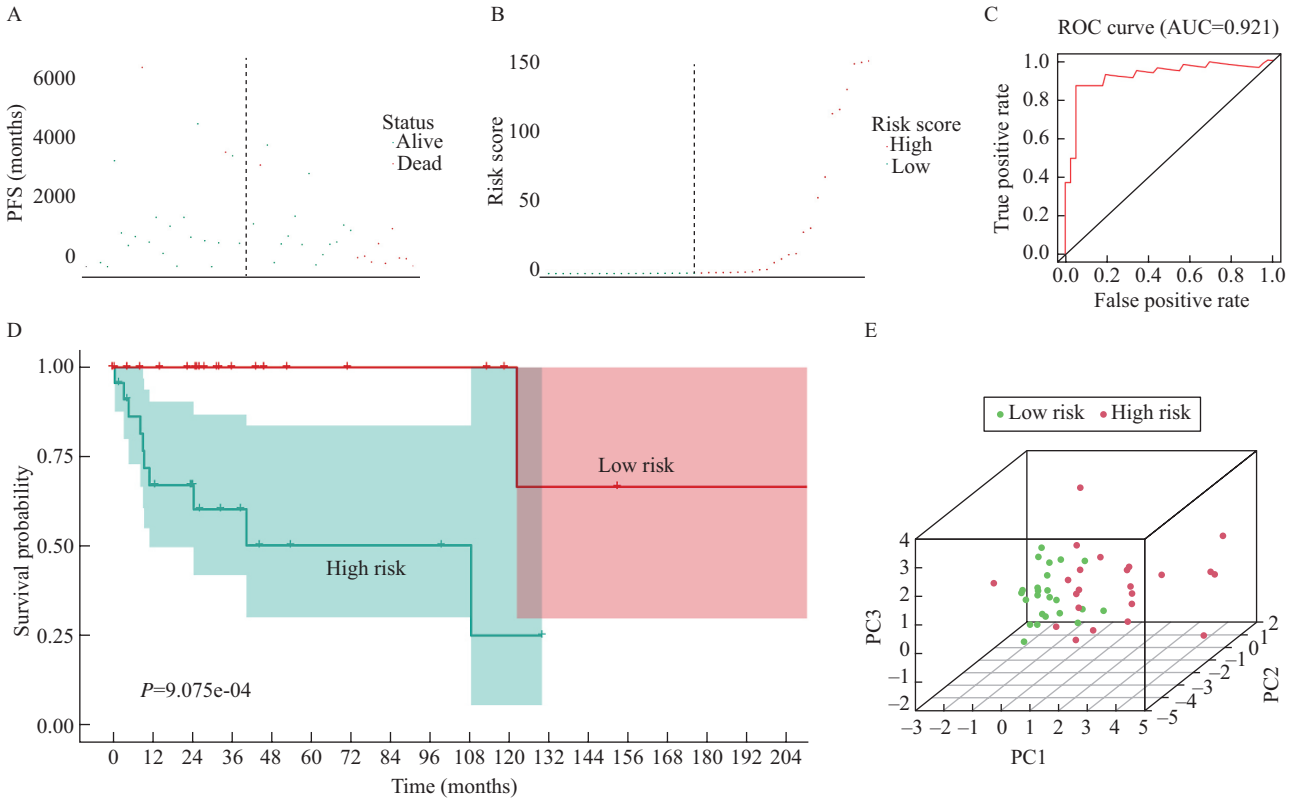


Fig. 4 The identification and prognostic model of differentially expressed metabolism-related genes (DEMRGs) between diffuse large B-cell lymphoma (DLBCL) and normal control samples

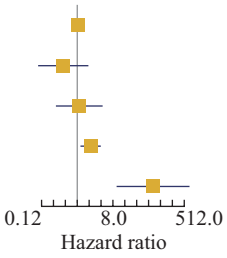
A and B: the heatmap plot (A) and volcano plot (B) of DEMRGs between DLBCL and normal control samples; C: cross-validation for tuning parameter selection in the proportional hazards model; D: the coefficients in the Lasso regression for identifying the top 20 DEMRGs; E: the proportional hazards model based on 13 key DEMRGs

low- and high-risk DLBCL patients (fig. 5B). ROC curve analysis was used to determine the specificity and sensitivity of the multivariate prognostic model (AUC=0.921, fig. 5C). Further, the Kaplan–Meier curve showed that the prediction model of the risk score had an excellent effectiveness (P<0.001, fig. 5D).



F

	P value	Hazard ratio
Age	0.268	1.030 (0.977-1.886)
Gender	0.266	0.449 (0.109-1.843)
Extranodal involvement	0.859	1.127 (0.300-4.238)
Therapy outcome	0.005	2.188 (1.264-3.788)
Risk score	<0.001	82.401 (10.370-654.760)



G

	P value	Hazard ratio
Age	0.729	1.018 (0.922-1.123)
Gender	0.101	0.069 (0.003-1.688)
Extranodal involvement	0.134	0.200 (0.024-1.642)
Therapy outcome	0.019	9.089 (1.432-57.708)
Risk score	0.002	1.036 (1.013-1.059)

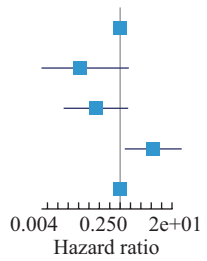


Fig. 5 The prognostic roles of differentially expressed metabolism-related genes (DEMRGs) in diffuse large B-cell lymphoma (DLBCL) patients

A: the scatter dot plot of each sample reordered by the status score; B: the scatter plot of the samples (green and red represent the low-risk and high-risk groups, respectively); C: the receiver operator characteristic (ROC) curve for prognostic DEMRGs [the area under the ROC curve (AUC)=0.921]; D: the Kaplan–Meier curve of overall survival for prognostic DEMRGs; E: the risk plot of samples in the low- and high-risk groups (green and red represent the low-risk and high-risk groups, respectively); F and G: the univariate (F) and multivariate (G) Cox regression models (yellow and blue represent the univariate and multivariate Cox regression model, respectively)

The scatter plot showed that the high-risk and low-risk groups had a high degree of separation. The patients in the high-risk group showed a higher mortality than in the low-risk group (fig. 5E).

2.5 Clinical Significance of Prognostic Model

To correct the gene expression levels by demographics, we integrated the risk score from the prognostic model and the clinicopathological

characteristics including age, sex, extra-nodal involvement, and therapeutic outcome. Next, univariate and multivariate Cox regression analyses were performed to explore the independent prognostic factors influencing overall survival, which indicated that the risk score was an independent prognostic factor in the univariate [HR=82.401, 95%CI (10.370–654.760), P<0.001, fig. 5F] and multivariate [HR=1.036, 95%CI (1.013–1.059), P=0.002, fig. 5G] Cox regression models.

Additionally, the therapeutic outcome was identified to have a significant impact on the overall survival in univariate [HR=2.188, 95%CI (1.264–3.788), P=0.005, fig. 5F] and multivariate [HR=9.089, 95%CI (1.432–57.708), P=0.019, fig. 5G] Cox regression models. Interestingly, no prognostic significance of age, sex, or extra-nodal involvement in DLBCL was observed, which was not consistent with our general cognition of malignancies. A possible reason for the inconsistency was the limited DLBCL sample size, which contributed to the lack of adequate representation of the DLBCL patient population.

2.6 Correlation Analysis of Immune Cells/Gene Sets and Hallmark Signaling Pathways

Due to the crucial roles of the bone marrow immune microenvironment in DLBCL, the immune cells and genes in DLBCL were further identified. In each sample, the fractions of 22 types of immune cells or gene sets are illustrated in fig. 6A. The results of the nonparametric test showed the fractions of naive B cells (P<0.001), memory B cells (P<0.001), plasma cells (P<0.001), $CD8^+$ T cells (P<0.001), memory activated T cells (P<0.001), follicular helper T cells (P<0.001), regulatory T cells (Tregs) (P<0.001), gamma delta T cells (P<0.001), activated natural killer (NK) cells (P<0.001), M0 macrophages (P<0.001), M1 macrophages (P<0.001), M2 macrophages (P<0.001), resting dendritic cells (P<0.001), and activated dendritic cells (P<0.05) were significantly greater in DLBCL samples than in normal PBMC samples, while infiltrates of $CD4^-$ T cells (P<0.001), resting NK cells (P<0.001), monocytes (P<0.001), activated mast cells (P<0.001), and neutrophils (P<0.001) were relatively fewer in DLBCL samples (fig. 6B). Moreover, ssGSEA was performed to identify immune cells and pathways that are significantly correlated with DEMRGs and are shown in a heatmap plot (fig. 6C).

2.7 Identification of Downstream Signaling Pathways and Upstream TFs

The differential expression levels of 50 hallmarks of cancer were illustrated between DLBCL and normal tissues in a heatmap (fig. 7A) and a volcano map (fig. 7B). Besides, the differential expression levels of 50 hallmarks of cancer and immune cells/gene sets were further assessed by GSVA and GSEA, respectively (fig. 7C and 7D). According to the results of the t-test, we illustrated the most significantly up- and down-regulated hallmark signaling pathways with red and green bars, respectively.

Meanwhile, edgeR was used to identify differentially expressed TFs. The heatmap and volcano plot illustrated 183 differentially expressed TFs extracted from 318 TFs (fig. 7E and 7F).

2.8 Co-expression Analysis

The co-expression interaction pairs of TFs, DEMRGs, protein chips, immune genes, immune cells, and hallmark signaling pathways were used to construct the regulation network by co-expression Pearson correlation analysis (fig. 8A). The co-expression heat map shows the regulatory patterns and expression levels of the six aforementioned factors (fig. 8B). In the co-expression heatmap, the TF regulatory factor X 5 (RFX5) exhibited a significant co-expression pattern with the metabolism-related gene phosphatase, orphan 1 (PHOSPHO1) (R=0.73, P<0.001). There was a significant relationship between PHOSPHO1 and the hallmark MYC proto-oncogene (MYC), targeting the V2 pathway (R= – 0.56, P<0.001) as well as neutrophil 1 (R=0.64, P<0.001). There was a significant co-expression trend between the protein chip INPP4B and PHOSPHO1 (R=0.56, P<0.001). Moreover, we identified a higher peak of H3K27ac in the region of the PHOSPHO1 gene in DLBCL than in human B cells (GM12878), indicating the enhanced transcription of PHOSPHO1 (fig. 8C).

3 DISCUSSION

DLBCL refers to the most common pathological type of NHL, accounting for about 35% of NHL in developed countries and up to 60% in developing countries[25]. In the tumorigenesis and recurrence of DLBCL, metabolic reprogramming is a significant feature. In this study, comprehensive bioinformatics analysis was performed on metabolism-related genes involved in glucose and lipid metabolic processes. Based on 13 identified DEMRGs, we constructed a prognostic model with a high applicability. In addition, our data indicated that the tumorigenesis of DLBCL was highly related to the bone marrow immune microenvironment; thus, the immune genes and immune cells were also identified. In the regulatory network, we found that the TF RFX5 might regulate the metabolism-related gene PHOSPHO1 and that PHOSPHO1 was also associated with MYC targeting the V2 pathway and neutrophils.

Metabolic reprogramming is recognized as one of the ten new hallmarks of tumor cells[6]. Increased glycolysis under normoxic conditions, also known as the "Warburg effect," is considered as a key characteristic of human cancers as well as glutamine metabolism[26]. A recent study suggests that low pretreatment brain

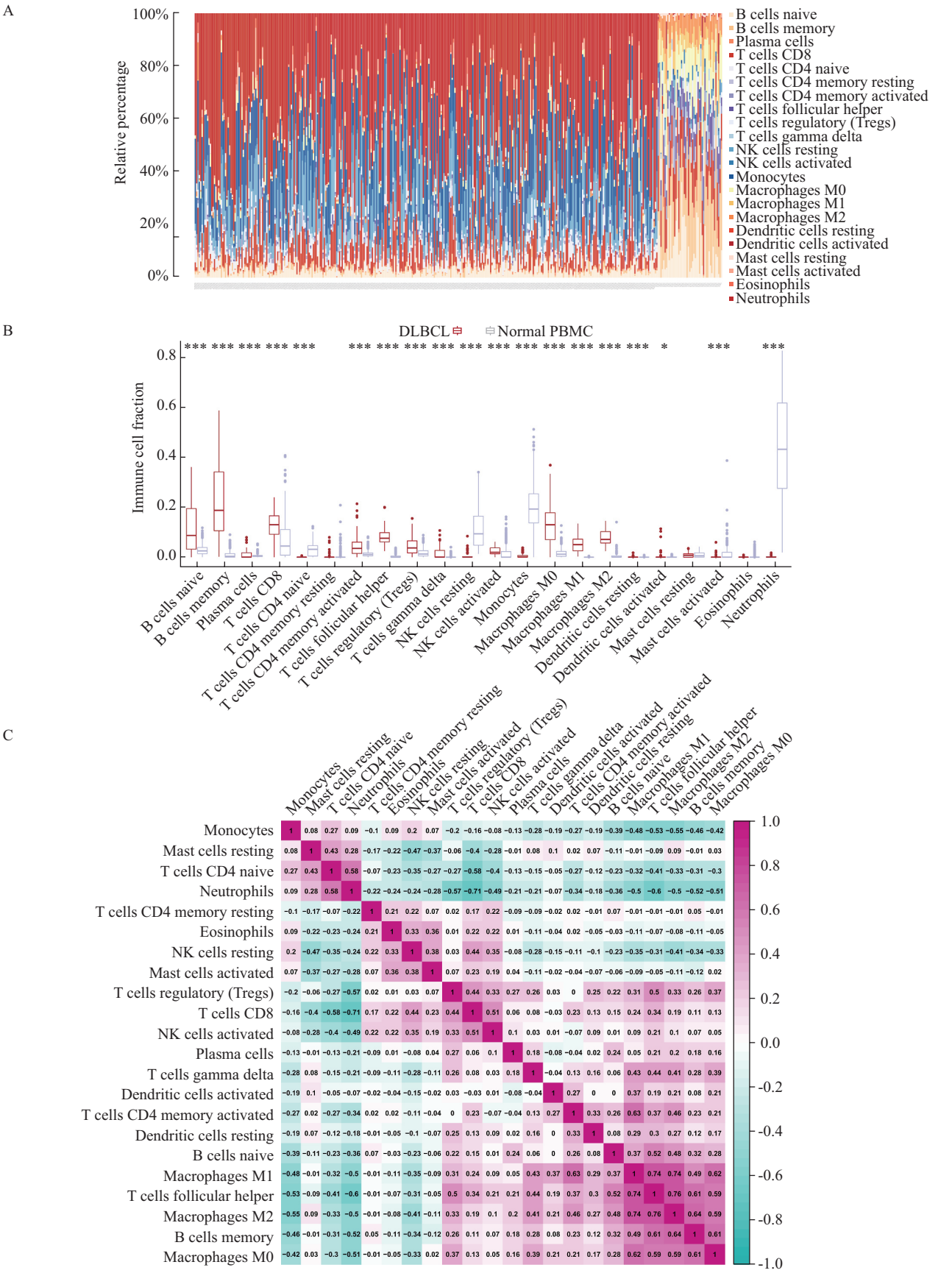


Fig. 6 The differentially expressed immune cell analysis between diffuse large B-cell lymphoma (DLBCL) samples and normal peripheral blood mononuclear cells (PBMCs)

A: the bar chart of differentially expressed immune cells between DLBCL and normal PBMC samples; B: the box plot of the fraction of immune cells between DLBCL and normal control samples via the Wilcoxon rank-sum test; C: the co-analysis of immune cells/pathways by single-sample Gene Set Enrichment Analysis

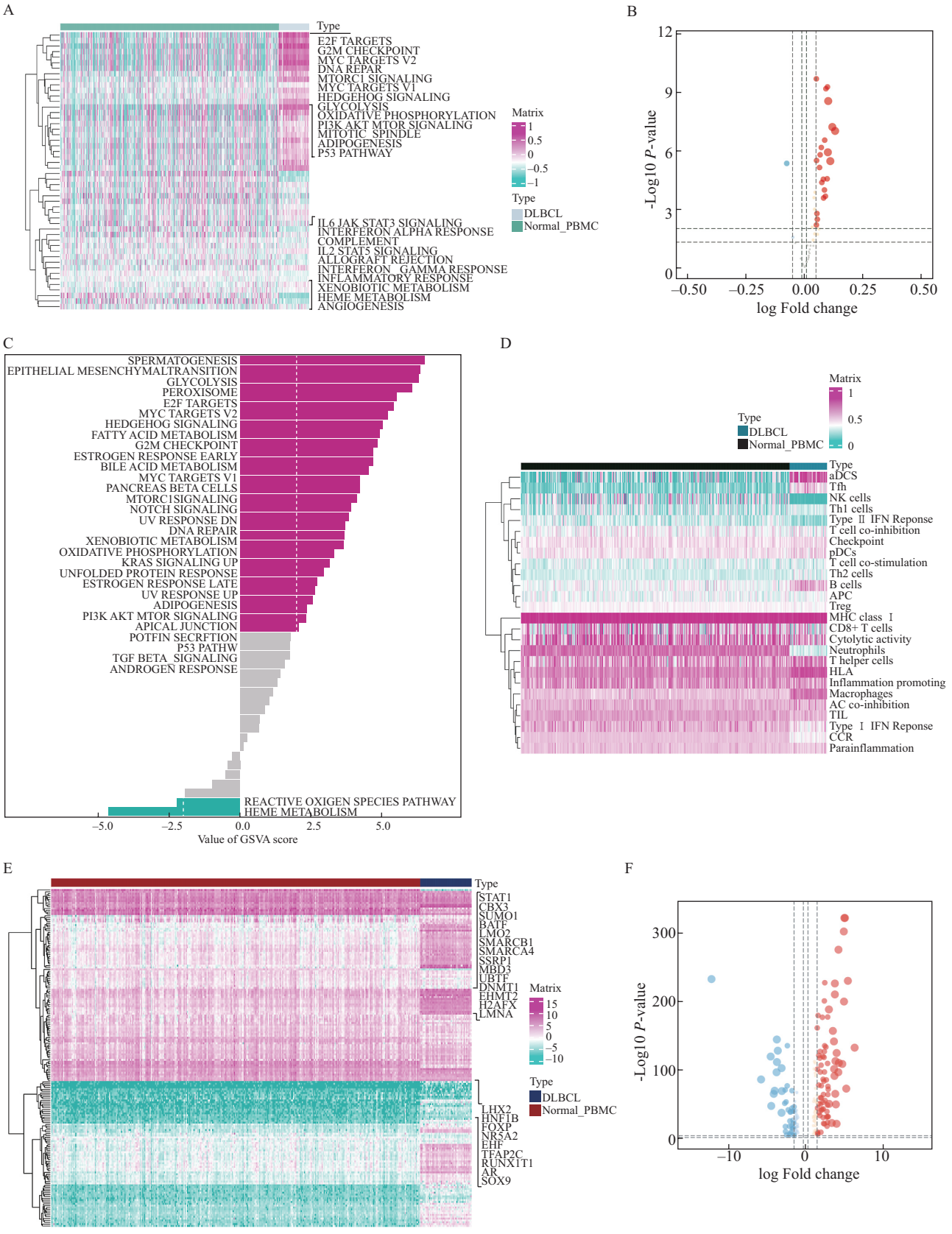


Fig. 7 Differential expression signaling pathway analysis between diffuse large B-cell lymphoma (DLBCL) and normal control samples A and B: the heatmap (A) and volcano plot (B) of signaling pathways between DLBCL and normal control samples by Gene Set Variation Analysis (GSVA); C: the bar plot of the *t* value of the GSVA score; D: the heatmap of immune cells/gene sets between DLBCL and normal control samples by GSEA; E and F: the heatmap (E) and volcano plot (F) of transcription factors between DLBCL and normal control samples based on the *P* value and log |Flod change| value

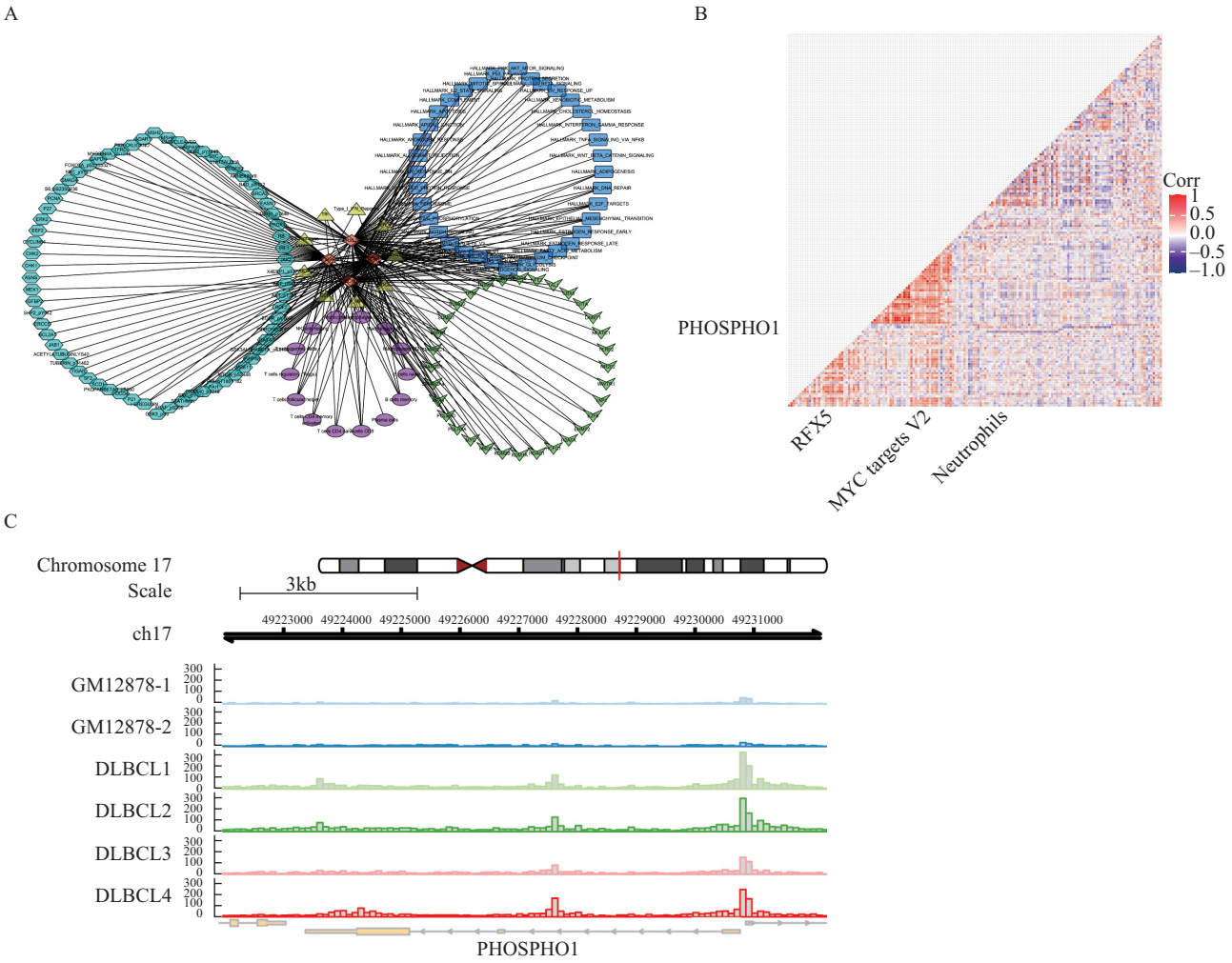


Fig. 8 The regulatory network of transcription factors (TFs), differentially expressed metabolism-related genes (DEMRGs), immune cells/gene sets, protein chips, and signaling pathways
A: the regulatory network of TFs, DEMRGs, immune cells/gene sets, protein chips, and signaling pathways (Arrows represent TFs. Diamonds represent DEMRGs. Ellipses and triangles represent immune cells/gene sets. Hexagons represent protein chips. Rectangles represent hallmark signaling pathways); B: the heatmap for the correlation analysis of DEMRGs, TFs, immune cells/gene sets, protein chips, and hallmark signaling pathways; C: the ChIP-seq of H3K27ac in the region of the PHOSPHO1 gene in diffuse large B-cell lymphoma and normal human B cells

glucose metabolism is correlated with a worse outcome of patients with DLBCL, and this phenomenon cannot be reversed by R-CHOP therapy[27]. In addition, overwhelming evidence supports the concept that dysregulation of lipid metabolic processes is also one of the principal metabolic markers of cancer cells and is tightly correlated with the tumorigenesis, metastasis, radiosensitivity, and chemosensitivity of multiple malignancies, including DLBCL[28]. In aggressive B cell lymphoma, substantial transcriptional reprogramming associated with increased lipid metabolism has been identified; in addition, relevant genetic heterogeneity and abnormalities have been observed[29]. Based on bioinformatics analysis and multidimensional validation, PHOSPHO1 was postulated to be positively regulated by RFX5 and to mediate MYC targeting the V2 pathway and neutrophils.

RFX5, a TF, belongs to the RFX family of DNA-binding proteins. RFX5 and CIITA form a complex that activates transcription from major histocompatibility class Ⅱ (MHCII) complex promoters[30]. Based on systematic identification of transcriptional target genes of RFX5, it may also be implicated in cancer-related pathways such as DNA damage repair, cell cycle, and proliferation pathways[31]. In a subset of DLBCL cases, loss of MHCII expression has been correlated with an extremely poor prognosis in a recent study[32]. These results confirmed that the loss of MHCII expression in DLBCL patients was probably due to altered transcriptional regulation induced by RFX5, which has profound biological, prognostic, and potential therapeutic implications.

PHOSPHO1, a soluble cytoplasmic phosphatase, consists of three peptide motifs belonging to the haloacid dehalogenase superfamily[33]. PHOSPHO1 exhibits high activities toward phosphocholine

and phosphoethanolamine (PE). Therefore, it is associated with glycerophospholipid biosynthesis[34, 35]. Generally, PHOSPHO1 is correlated with the production of inorganic phosphate (Pi) for matrix and skeletal mineralization[36]. Activation of cytosolic PE metabolism decreases the mitochondrial respiration activity and increases glycolysis[37]. Furthermore, PE values before cancer treatment are related to the response to medication at 6th month and the time to therapy failure in DLBCL patients[38]. Therefore, an increased level of PHOSPHO1 may be a novel biomarker to predict the treatment outcome of DLBCL patients and enhanced Warburg effects in DLBCL cellular energy metabolism via regulating PE levels.

MYC targets V2 pathway functions as a cancer-related pathway to activate cellular proliferation and tumorigenesis[39]. The translocation of the MYC gene locus (8q24) represents a characteristic biomarker of NHL[40]. In the current WHO lymphoma classification, high-grade B-cell lymphoma with MYC/BCL2 double-hit (DH) is considered to be a distinct entity with a dismal prognosis after classical immunochemotherapy[41]. DLBCL patients with MYC abnormalities demonstrate a high-grade morphology, poor clinical outcomes, and distinct gene-expression signatures[42]. For patients with follicular lymphoma (FL), histopathological transformation to high-grade FL and DLBCL is an unfavorable step in cancer progression[43]. Oncogene MYC activation and cancer microenvironment remodeling are implicated in FL progression. An increased MYC level might be an immediate outcome of genomic aberrations implicated in the MYC locus[44]. Various MYC-targeted gene inhibitors have been identified and tested, such as Alisertib and Romidepsin, which regulate the expression of MYC, and rituximab, which targets the translocation of MYC[40]. Thus, drugs targeting oncogenes involved in MYC targeting the V2 pathway may provide new strategies for DLBCL treatment.

CHOP chemotherapy is the first-line treatment for patients with DLBCL, and multiple variations have been reported with various clinical outcomes, such as immune therapies[45, 46]. Also, inflammation is considered as playing a fundamental role in lymphomagenesis and cancer invasion[47]. Immune cell subsets from peripheral blood such as neutrophils can offer an indication of inflammation[48]. The neutrophil-to-lymphocyte ratio (NLR) was identified as a significant prognostic marker in DLBCL. More recently, specific studies have found important associations between the NLR and the time-to-progression, lymphoma-specific survival, progression-free survival, and overall survival for DLBCL patients treated with CHOP-based chemotherapy[49]. Therefore, based on *in-silico* analyses and other studies, neutrophils were postulated to be significant therapeutic or prognostic biomarkers that might be beneficial for DLBCL prognosis prediction and improving therapy for patients with this disease.

There were still several inevitable limitations of this study. First, the quantity of related data acquired from online datasets was statistically incomplete. It is far too difficult to reduce the potential error and bias by obtaining the same number of samples with different sexes, ages, and races, leading to the lack of comprehensiveness. Second, despite the results being validated by external databases, the sample size was limited. Third, the scientific hypothesis was mainly based on bioinformatics, and it was not validated by exploring the underlying molecular mechanisms. Therefore, ChIP-seq and ATAC-seq can be further used to determine the direct transcriptional regulation pattern between TFs and DEMRGs.

To sum up, our data provide a well-applied model based on 13 DEMRGs for the prognosis of DLBCL. In addition, our data indicate that PHOSPHO1 is positively regulated by RFX5 and mediates MYC targeting the V2 pathway and neutrophils.

Acknowledgments

We are grateful for the use of the data provided by The Cancer Genome Atlas (TCGA), Genotype-Tissue Expression Portal (GTEx), and the Sequence Read Archive (SRA) database.

Conflict of Interest Statement

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

REFERENCES

1 Chapuy B, Stewart C, Dunford AJ, *et al*. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med, 2018,24(5):679-690
2 Naresh KN. Genetics of Diffuse Large B-Cell Lymphoma. New Engl J Med, 2018,379(5):493
3 Gribben JG, Fowler N, Morschhauser F. Mechanisms of Action of Lenalidomide in B-Cell Non-Hodgkin Lymphoma. J Clin Oncol, 2015,33(25):2803-2811
4 Coiffier B, Lepage E, Briere J, *et al*. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. New Engl J Med, 2002,346(4):235-242
5 Glass B, Ziepert M, Reiser M, *et al*. High-dose therapy followed by autologous stem-cell transplantation with and without rituximab for primary treatment of high-risk diffuse large B-cell lymphoma. Ann Oncol, 2010,21(11):2255-2261
6 Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell, 2011,144(5):646-674
7 Baenke F, Peck B, Miess H, *et al*. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. Dis Model Mech, 2013,6(6):1353-1363
8 Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer

Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn), 2015,19(1A):A68-A77
9 Ardlie KG, DeLuca DS, Segre AV, *et al*. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 2015,348(6235): 648-660
10 Liberzon A, Birger C, Thorvaldsdottir H, *et al*. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst, 2015,1(6):417-425
11 Zheng R, Wan C, Mei S, *et al*. Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res, 2019,47(D1):D729-D735
12 Harris MA, Clark J, Ireland A, *et al*. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res, 2004,32(Databasic issue):D258-D261
13 Meng T, Huang R, Jin J, *et al*. A comparative integrated multi-omics analysis identifies CA2 as a novel target for chordoma. Neuro-oncology, 2021,23(10):709-722
14 Yu GC, Wang LG, Han YY, *et al*. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. Omics, 2012,16(5):284-287
15 Guo H, Wang S, Ju M, *et al*. Identification of Stemness-Related Genes for Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma by Integrated Bioinformatics Analysis. Front Cell Dev Biol, 2021, 9:642724
16 Subramanian A, Narayan R, Corsello SM, *et al*. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell, 2017,171(6):1437-1452
17 Uhlen M, Fagerberg L, Hallstrom BM, *et al*. Proteomics. Tissue-based map of the human proteome. Science, 2015,347(6220):1260419
18 Snel B, Lehmann G, Bork P, *et al*. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res, 2000,28(18):3442-3444
19 Blazewicz SJ, Barnard RL, Daly RA, *et al*. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. Isme J, 2013,7(11):2061-2068
20 Brandman O, Hegde RS. Ribosome-associated protein quality control. Nat Struct Mol Biol, 2016,23(1):7-15
21 Poulos TL. Heme enzyme structure and function. Chem Rev, 2014,114(7):3919-3962
22 Acharya KR, Ackerman SJ. Eosinophil granule proteins: form and function. J Biol Chem, 2014,289(25):17406-17415
23 Duncan T, Trewick SC, Koivisto P, *et al*. Reversal of DNA alkylation damage by two human dioxygenases. Proc Natl Acad Sci USA, 2002,99(26):16660-16665
24 Di Stefano L, Jensen MR, Helin K. E2F7, a novel E2F featuring DP-independent repression of a subset of E2F-regulated genes. EMBO J, 2003,22(23):6289-6298
25 Garcia-Morales E, Lazaro-Martinez JL, Martinez-Hernandez D, *et al*. Impact of diabetic foot related complications on the Health Related Quality of Life (HRQol) of patients--a regional study in Spain. Int J Low Extrem Wounds, 2011,10(1):6-11
26 Brand K. Glutamine and Glucose-Metabolism during Thymocyte Proliferation - Pathways of Glutamine and Glutamate Metabolism. Biochem J, 1985,228(2):353-361
27 Adams HJ, de Klerk JM, Fijnheer R, *et al*. Brain glucose metabolism in diffuse large B-cell lymphoma patients as assessed with FDG-PET: impact on outcome and chemotherapy effects. Acta Radiol, 2016,57(6):733-741
28 Currie E, Schulze A, Zechner R, *et al*. Cellular Fatty Acid Metabolism and Cancer. Cell Metab, 2013,18(2):153-161
29 Kobayashi T, Lam PY, Jiang H, *et al*. Increased lipid metabolism impairs NK cell function and mediates adaptation to the lymphoma environment. Blood, 2020,136(26):3004-3017
30 Scholl T, Mahanta SK, Strominger JL. Specific complex formation between the type II bare lymphocyte syndrome-associated transactivators CIITA and RFX5. P Natl Acad Sci USA, 1997,94(12):6330-6334
31 Chen DB, Xie XW, Zhao YJ, *et al*. RFX5 promotes the progression of hepatocellular carcinoma through transcriptional activation of KDM4A. Sci Rep, 2020,10(1):14538
32 Rimsza L, Roberts R, Miller T, *et al*. Loss of MHC class II gene and protein expression in diffuse large B cell lymphoma is related to decreased tumor immumosurveillance and poor patient survival: A follow-up study to the NIH director's challenge leukemia and lymphoma molecular profiling project (LLMPP). Blood, 2004,103(11):4251-4258
33 Stewart AJ, Schmid R, Blindauer CA, *et al*. Comparative modelling of human PHOSPHO1 reveals a new group of phosphatases within the haloacid dehalogenase superfamily. Protein Eng, 2003,16(12):889-895
34 Roberts SJ, Stewart AJ, Sadler PJ, *et al*. Human PHOSPHO1 exhibits high specific phosphoethanolamine and phosphocholine phosphatase activities. Biochem J, 2004,382(Pt 1):59-65
35 Houston B, Seawright E, Jefferies D, *et al*. Identification and cloning of a novel phosphatase expressed at high levels in differentiating growth plate chondrocytes. Bba-Mol Cell Res, 1999,1448(3):500-506
36 Gohil VM, Zhu L, Baker CD, *et al*. Meclizine Inhibits Mitochondrial Respiration through Direct Targeting of Cytosolic Phosphoethanolamine Metabolism. J Biolog Chem, 2013,288(49):35387-35395
37 Arias-Mendoza F, Payne GS, Zakian K, *et al*. Noninvasive phosphorus magnetic resonance spectroscopic imaging predicts outcome to first-line chemotherapy in newly diagnosed patients with diffuse large B-cell lymphoma. Acad Radiol, 2013,20(9):1122-1129
38 Dang CV, O'Donnell KA, Zeller KI, *et al*. The c-Myc target gene network. Semin Cancer Biol, 2006,16(4):253-264
39 Zhang J, Meng L, Jiang W, *et al*. Identification of clinical molecular targets for childhood Burkitt lymphoma. Transl Oncol, 2020,13(12):100855
40 Swerdlow SH, Campo E, Pileri SA, *et al*. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood, 2016,127(20):2375-2390
41 Deng M, Xu-Monette ZY, Pham LV, *et al*. Aggressive B-cell Lymphoma with MYC/TP53 Dual Alterations Displays Distinct Clinicopathobiological Features and Response to Novel Targeted Agents. Mol Cancer Res,

2021,19(2):249-260
42 Lou X, Fu J, Zhao X, *et al*. MiR-7e-5p downregulation promotes transformation of low-grade follicular lymphoma to aggressive lymphoma by modulating an immunosuppressive stroma through the upregulation of FasL in M1 macrophages. J Exp Clin Cancer Res, 2020,39(1):237
43 Klapproth K, Wirth T. Advances in the understanding of MYC-induced lymphomagenesis. Brit J Haematol, 2010,149(4):484-497
44 Marconato L, Polton GA, Sabattini S, *et al*. Conformity and controversies in the diagnosis, staging and follow-up evaluation of canine nodal lymphoma: a systematic review of the last 15 years of published literature. Vet Comp Oncol, 2017,15(3):1029-1040
45 Marconato L, Frayssinet P, Rouquet N, *et al*. Randomized, placebo-controlled, double-blinded chemoimmunotherapy clinical trial in a pet dog model of diffuse large B-cell lymphoma. Clin Cancer Res, 2014,20(3):668-677
46 Carbone A, Tripodo C, Carlo-Stella C, *et al*. The role of inflammation in lymphoma. Adv Exp Med Biol, 2014,816:315-333
47 Howard R, Kanetsky PA, Egan KM. Exploring the prognostic value of the neutrophil-to-lymphocyte ratio in cancer. Sci Rep, 2019,9(1):19673
48 Marconato L, Martini V, Stefanello D, *et al*. Peripheral blood lymphocyte/monocyte ratio as a useful prognostic factor in dogs with diffuse large B-cell lymphoma receiving chemoimmunotherapy. Vet J, 2015,206(2): 226-230

(Received Jul. 22, 2021; accepted Dec. 29, 2021)