Identification of hub genes associated with the pathogenesis of diffuse large B-cell lymphoma subtype one characterized by host response via integrated bioinformatic analyses

Lingna Zhou[1,4], Liya Ding[2,3], Yuqi Gong[2,3], Jing Zhao[2], Gong Xin[5], Ren Zhou[2,3] and Wei Zhang[1,4]

[1] Department of Pathology and Physiology, Cancer Institute, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[2] Department of Pathology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[3] Department of Pathology and Pathophysiology, Institute of Pathology and Forensic Medicine, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[4] Key Laboratory of Disease Proteomics of Zhejiang Province, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[5] School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China, Hangzhou, Zhejiang, China

Corresponding authors
Ren Zhou, zhouren@zju.edu.cn
Wei Zhang, zwei72@163.com

ABSTRACT

Background. Host response diffuse large B-cell lymphoma (HR DLBCL) shares features of histologically defined T-cell/histiocyte-rich B-cell lymphoma, including fewer genetic abnormalities, frequent splenic and bone marrow involvement, and younger age at presentation. HR DLBCL is inherently less responsive to the standard treatment for DLBCL. Moreover, the mechanism of infiltration of HR DLBCL with preexisting abundant T-cells and dendritic cells is unknown, and their associated underlying immune responses incompletely defined. Here, hub genes and pathogenesis associated with HR DLBCL were explored to reveal molecular mechanisms and treatment targets.

Methods. Differentially expressed genes were identified in three datasets (GSE25638, GSE44337, GSE56315). The expression profile of the genes in the GSE53786 dataset was used to constructed a co-expression network. Protein-protein interactions analysis in the modules of interest identified candidate hub genes. Then screening of real hub genes was carried out by survival analysis within the GSE53786 and GSE10846 datasets. Expression of hub genes was validated in the Gene expression profiling interactive analysis, Oncomine databases and human tissue specimens. Functional enrichment analysis and Gene set enrichment analysis were utilized to investigate the potential mechanisms. Tumor Immune Estimation Resource and The Cancer Genome Atlas were used to mine the association of the hub gene with tumor immunity, potential upstream regulators were predicted using bioinformatics tools.

Results. A total of 274 common differentially expressed genes were identified. Within the key module, we identified CXCL10 as a real hub gene. The validation of upregulated expression level of CXCL10 was consistent with our study. CXCL10 might have a

regulatory effect on tumor immunity. The predicted miRNA (hsa-mir-6849-3p) and transcription factor (IRF9) might regulate gene expression in the hub module.

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is the dominant subtype of non-Hodgkin lymphoma (*Malumbres et al., 2008*) and is characterized by clinical heterogeneity. This heterogeneity presents a challenge to the treatment of DLBCL. Although approximately 60% of the DLBCL patients achieved durable remission following the current standard chemotherapy R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone), 40% of patients suffered a relapse or became refractory, with limited treatment options (*Cristino et al., 2019*; *Jiang et al., 2017*).

Basing on gene expression profiling, a previous study classified patients into three subgroups: germinal center B-cell (GCB) group, activated B-cell group, and unclassified group (*Alizadeh et al., 2000*). The unclassified group indicated the molecular heterogeneity, which poses a barrier for better treatment and understanding of DLBCL. Thus, novel classification is urgently needed to better distinguish subtypes for DLBCL.

In 2005, *Monti et al. (2005)* identified three discrete DLBCL subtypes according to consensus clusters—"oxidative phosphorylation" (Oxphos) , "B-cell receptor/proliferation" (BCR), and "host response" (HR). HR DLBCL is T-cell/histiocyte-rich B-cell lymphoma, which suggests that tumor microenvironment may play an essential role in this subgroup of DLBCL. Tumor microenvironment has a closed relationship with the tumor initiation, progression, and invasion. Targeting the host immune response to HR DLBCL could be an effective treatment modality. However, the mechanism of infiltration of HR DLBCL with preexisting abundant T-cells and dendritic cells is unknown, and their associated underlying immune responses incompletely defined.

To further explore the mechanism of tumorigenesis and progression of HR DLBCL, we collected five published datasets of DLBCL from the Gene Expression Omnibus database (GEO), data from Gene Expression Profiling Interactive Analysis (GEPIA), Oncomine and The Cancer Genome Atlas (TCGA). Co-expression network analysis was applied to identify modules of common differentially expressed genes (DEGs) and determine how they correlated with HR DLBCL. To broadened our knowledge about the hub module and hub genes, comprehensive bioinformatics analyses were conducted, including functional enrichment analysis, gene set enrichment analysis, exploration of correlation of hub genes with tumor-infiltrating cells and immune signatures, upstream regulator prediction. By associating clinical data with molecular mechanisms, network-centric genes, new biomarkers for diagnosis, prognosis and treatment might be identified.

Table 1 Characteristics of the five studies inluded in the bioinformatic analysis.

Clinical features of included patients	$n = 629$
Platform	
GPL570[HG-U133_Plus_2]	629
Patient source	
USA	548
Spain	26
Denmark	55
Age	61.21764 ± 15.25502
Sex	
Male	292
Female	220
NA	117
Stage	
I	82
II	157
III	126
IV	153
NA	111

MATERIALS & METHODS

Data collection and preprocessing

The workflow of public microarray repositories search is shown in Fig. S1. After a systematic search, five datasets (GSE25638, GSE44337, GSE56315, GSE53786 and GSE10846) depended on the GPL570 platform met the inclusion criteria and were included in the integrated bioinformatics analyses (*Vicente-Duenas et al., 2012*; *Tompkins et al., 2013*; *Dybkaer et al., 2015*; *Scott et al., 2014*; *Cardesa-Salzmann et al., 2011*). Normalized datasets were obtained from the GEO database using the GEOquery package (*Davis & Meltzer, 2007*). The maximum mean expression values of probes were used for genes with multiple probes in the microarray data and were annotated using the hgu133plus2.db. The clinical characteristics of included DLBCL patients were summarized in Table 1. The GSE25638, GSE44337 and GSE56315 datasets were analyzed for DEGs respectively. The intersections of the DEGs of the three datasets were selected for further analysis. GSE53786 dataset with typing data, survival, and prognosis information was used as the training set to construct co-expression networks. Then screening of real hub genes was carried out by survival analysis within the GSE53786 and GSE10846 datasets. A schematic representation of the study methods is shown in Fig. 1.

Screening of DEGs

DEGs between DLBCL samples and normal tissues were identified through R package limma. DEGs with $P < 0.05$ and fold change (log FC) > Mean (log FC) + 2*SD (log FC) were regarded as significant DEGs.
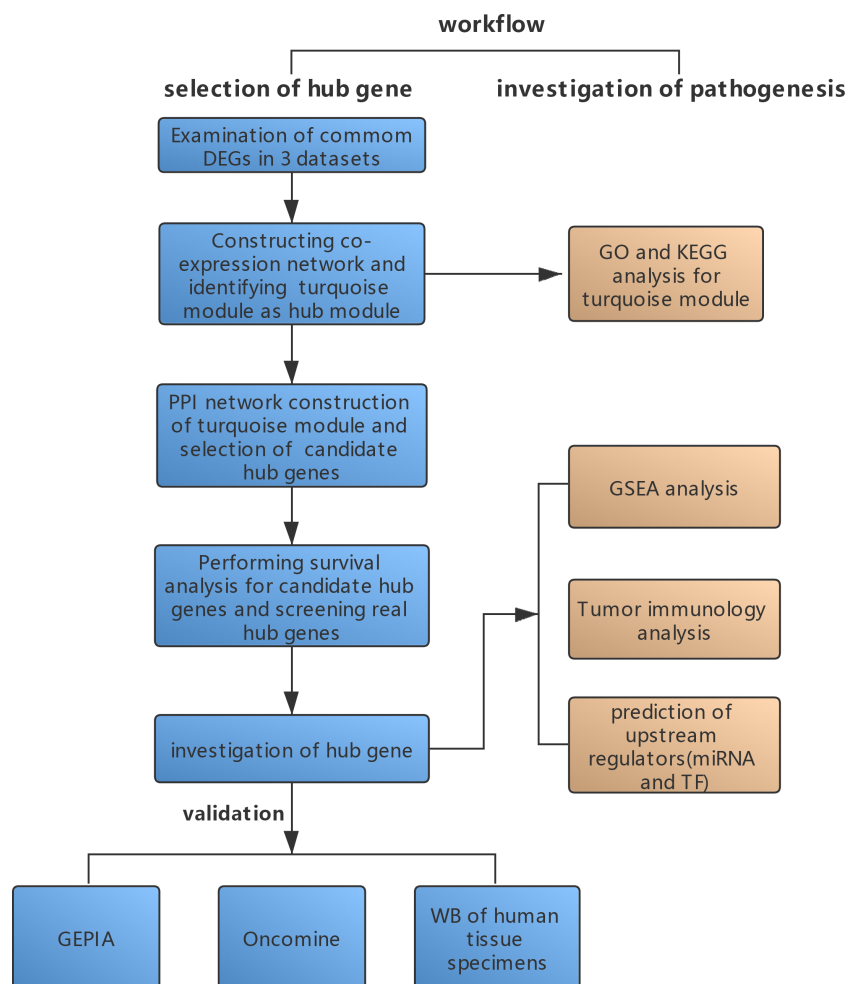
Figure 1 A schematic representation of the study methods. Methods regarding real hub gene selection and verification were shown in blue boxes, methods regarding molecular pathogenesis investigation were shown in orange boxes.

Co-expression network construction

Weighted gene co-expression network analysis (WGCNA) (*Langfelder & Horvath, 2008*) was applied to construct the co-expression network of overlap significant DEGs. Hierarchical clustering was carried out using the average method with exclusion by static tree cutting with a cut height set to 100. By calculating the scale-free fit index with a soft-threshold power (β) capability of 1 to 20, the one with the highest average connectivity degree was regarded as the most appropriate one to construct a scale-free network. Co-expressed genes were assigned to modules via dynamic minimum tree cutting, and similar modules were merged into one (similarity $\geq$ 0.75). We applied module eigengenes (MEs) and module gene significance (MS) to estimate correlations between clustered modules and the DLBCL subtypes.

Candidate hub genes selection

All genes in the hub module were uploaded to Search Tool for the Retrieval of Interacting Genes (STRING) to predict protein-protein interactions(PPIs), then the PPIs were imported into Cytoscape for visualization (*Szklarczyk et al., 2017*; *Zhou et al., 2018*). To predict the important nodes, the plug-in CytoHubba was used to calculate the degree of genes in the hub module. We defined genes with node connectivity > ratio of total edges to total nodes as the candidate hub genes.

Real hub gene identification

Expressions of candidate hub genes were divided into high and low levels using median expression level for further Kaplan–Meier survival analyses in GSE53786 and GSE10846.

Genes significantly correlated with prognosis ($P < 0.05$) were selected. Then, the prognosis values of the selected genes were assessed in different subgroups of DLBCL separately. The gene with significant prognosis value both in overall DLBCL patients and HR subtype was identified as real hub gene.

Validation of expression level of the hub gene in DLBCL

The expression of the hub gene at the transcriptional level was displayed by Expression Profiling Interactive Analysis (GEPIA; http://gepia.cancer-pku.cn) (*Tang et al., 2017*) and Oncomine gene expression array database (http://www.oncomine.org) (*Rhodes et al., 2004*).

The expression of the real hub gene at the protein level was verified by western blot of human tissue specimens. All the human tissue specimens were obtained from the Department of Pathology, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. Our study was approved by the Ethics Committee of Zhejiang University (Hangzhou, China). Under the guidance of the World Health Organization Classification of Tumors of Hematopoietic and Lymphoid Tissues, specimens, collected with informed consent, were diagnosed as DLBCL. The protein samples were eluted by boiling in $1\times$ SDS loading buffer at 100 °C and then fractionated by 12% SDS/PAGE, followed by standard western blotting. Antibodies were specific for anti-CXCL10 (Abcam, ab254374), anti-GAPDH (Abcam, ab181602).

Functional enrichment analysis

Metascape (http://metascape.org/) integrates multiple databases, such as Gene Ontology (GO), Kyoyo Encyclopedia of Genes and Genomes (KEGG), UniProt, and DrugBank (*Zhou et al., 2019*). It provides the typical gene enrichment analysis and visualizes the results from lists of biological functions, pathways, and more (*Ye et al., 2019*). Functional enrichment analysis of the screened genes in the hub module was carried out by Metascape. $P < 0.01$ was set as the cutoff criterion, and significance was ranked by enrichment score ($-\log10(P\text{-value})$).

Gene set enrichment analysis (GSEA)

Samples of HR DLBCL in GSE10846 were divided into two groups according to the expression level of the real hub gene. GSEA (version 4.1.0) was utilized to detect significantly

different signaling in the gene rank between the two groups. False discovery rate (FDR) $\leq$ 0.25 and p-value $\leq$ 0.05 were recommended.

Tumor immunology analysis

Tumor Immune Estimation Resource (TIMER, https://cistrome.shinyapps.io/timer/), containing 10,897 samples across 32 cancer types from TCGA (*Li et al., 2017*), provides a comprehensive analysis of immune infiltrates with tumors. We analyzed the correlation of the hub gene with various immune infiltrates in DLBCL, including B cells, CD8+ T cells, CD4+ T cells, macrophages, neutrophils, dendritic cells, and the tumor purity.

To explore the correlation of the hub gene with different kinds of immune signatures in DLBCL, we obtain gene signatures of various types of tumor-infiltrating lymphocytes as well as markers of chemokine, Major histocompatibility complex (MHC), Immunoinhibitor, Immunostimulatory, Cytokine and cytokine receptor, Immune checkpoint, Immune cell infiltrate genes, Cancer testis antigen genes, Human leukocyte antigen (HLA) genes, and Pro-inflammatory genes from the previous study (*Liu et al., 2020*). Besides, we downloaded data of patients with DLBCL mined from TCGA and RNA sequencing (RNA-seq) expression results using the RTCGA Toolbox package in R.

Correlations were calculated by the Pearson correlation and the threshold used for significant expression correlation was a Pearson correlation coefficient | r | $\geq$ 0.3 and all corresponding P values < 0.05.

Construction of potential miRNA- and transcriptional factor (TF)-target regulatory networks

We used the miRNet to predict potential miRNAs in regulating genes in the hub module. The predicted miRNA with the highest was selected to construct miRNA-target networks. Transcription factors of the highly related module were analyzed by the iRegulon plugin in Cytoscape (*Janky et al., 2014*). TF-target network consisted of the predicted TF with the highest normalized enrichment score and targeted genes in the hub module.

RESULTS

DEGs screening

After data preprocessing and quality assessment, normalized expression data of each dataset was shown by box plots (Figs. 2A–2C). Then DEGs were identified from each of the three datasets. Volcano plots were used to visualize DEGs between the tumor and non-tumor groups (Figs. 2D–2F). A total of 834 DEGs were screened from the GSE25638 dataset. They comprised 778 upregulated genes and 56 downregulated genes. Nine hundred DEGs were screened from the GSE44337 dataset. They comprised 727 upregulated genes and 173 downregulated genes. Nine hundred thirty-three DEGs were screened from the GSE56351 dataset, comprising 472 upregulated genes and 461 downregulated genes (Fig. 2G). A total of 274 common DEGs from the three datasets were selected for further analysis. These 274 DEGs comprised 264 upregulated genes (Fig. 2H) and ten downregulated genes (Fig. 2I).
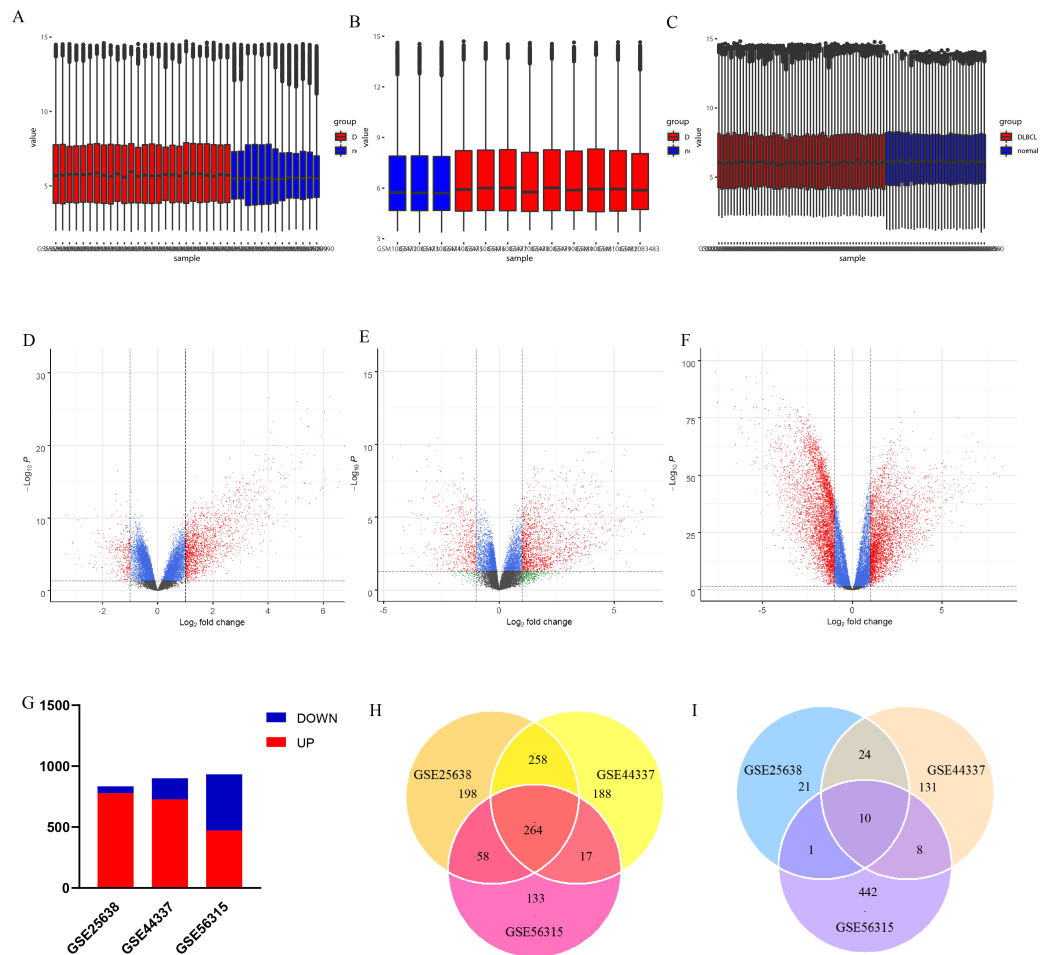
Figure 2 Statistics of differently expressed genes. (A–C) Box plots for the expression data in GSE25638, GSE44337 and GSE56315; (D–F) The volcano plots of DEGs in GSE25638, GSE44337, and GSE56315. The red dots represent upregulated DEGs while the green dots indicate downregulated DEGs; (G) DEGs in the three gene datasets; (H) Venn diagram analysis: overlapping upregulated differentially expressed genes from three datasets; (I) Venn diagram analysis: overlapping downregulated differentially expressed genes from three datasets.

Construction of Weighted co-expression network and identification of the hub module

After data preprocessing, the expression matrix of common DEGs was obtained in the GSE53786 training set. No sample was removed from subsequent analysis in the dataset after the first quality check (Fig. S2). To ensure a scale-free network, the power of $\beta = 7$ (scale-free $R^2 = 0.91$) was selected (Fig. 3). The DEGs with similar expression patterns were clustered into modules. Three modules resulted in different colors (Fig. 4A). The grey module showed the genes that cannot be merged. According to the ME and MS, the turquoise module showed the highest correlation with HR DLBCL among all modules (Figs. 4B–4C). Moreover, Scatter plot of module eigengenes in the turquoise module was
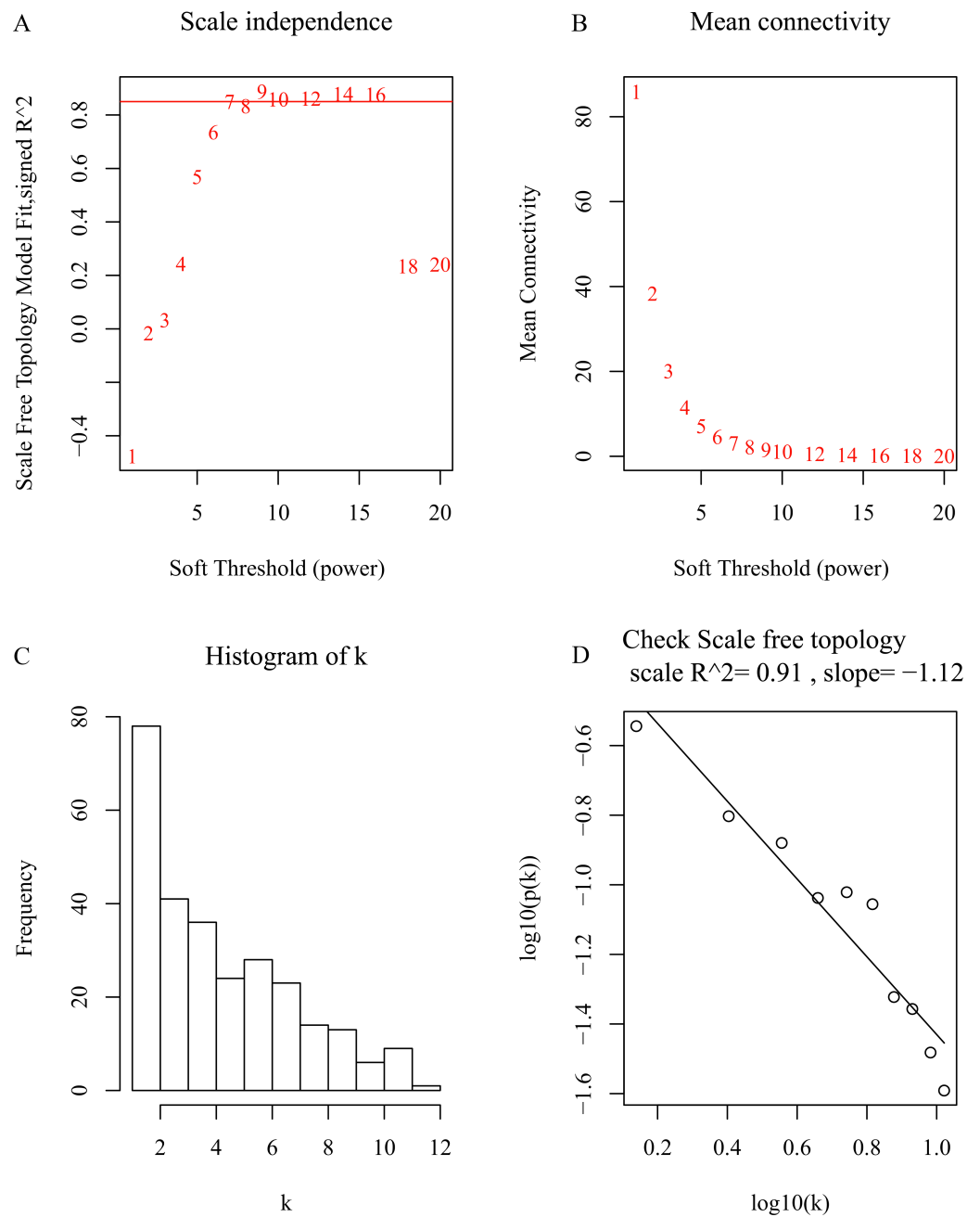
Figure 3 Soft-thresholding power selection in the weighted gene co-expression network analysis.
(A) Scale-free fit index of various soft-thresholding powers (β); (B) mean connectivity of various soft-thresholding powers(β); (C) Histogram of the connectivity distribution when $\beta = 7$; (D) scale-free topology checking when $\beta = 7$.

shown in Fig. 4D. Therefore turquoise module was identified as the hub module associated with HR DLBCL.
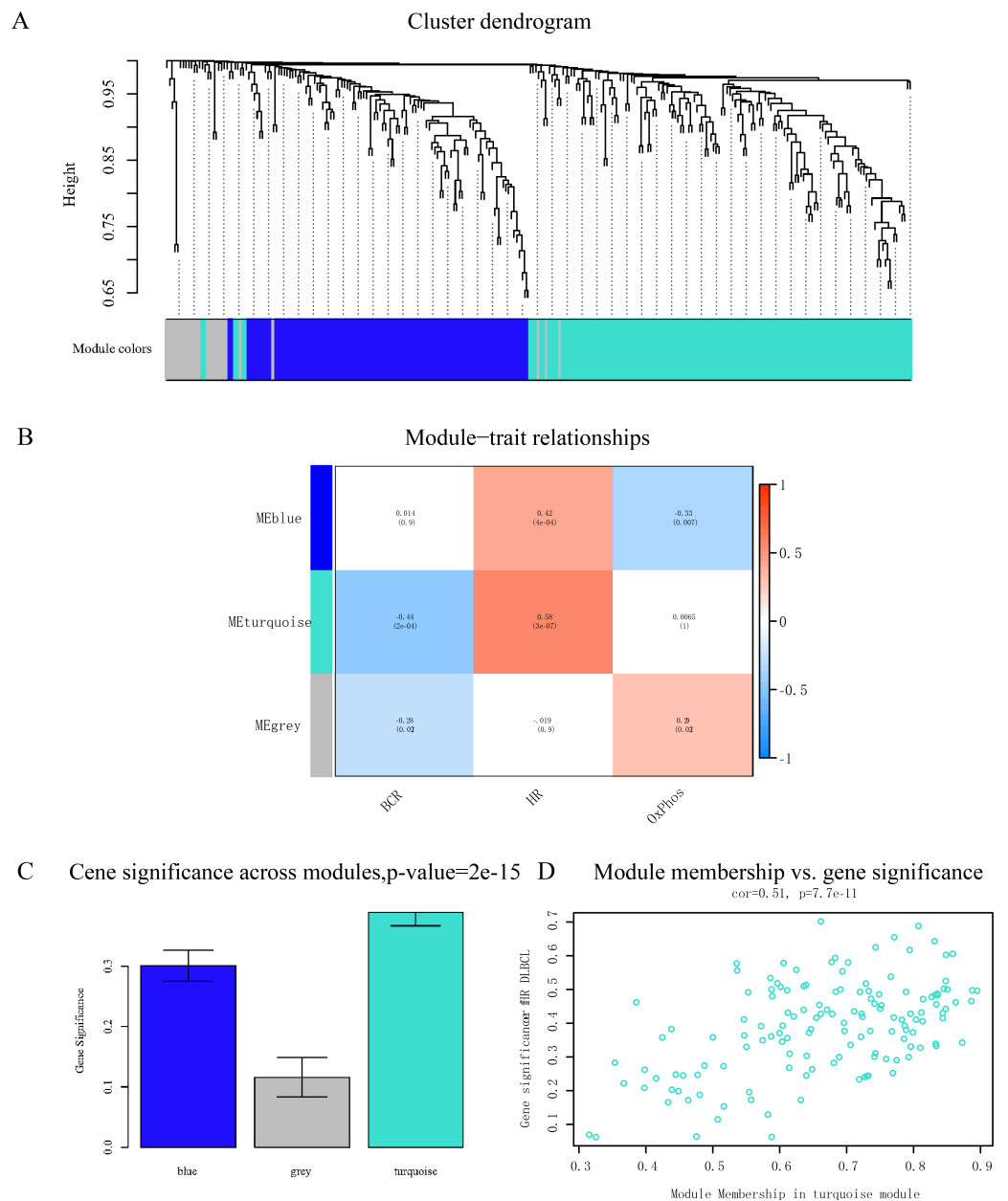
Figure 4 Analysis of modules correlated with the subtypes of DLBCL. (A) Dendrogram of all differentially expressed genes clustered according to a dissimilarity measure; (B) distribution of average gene significance and errors in the modules associated with DLBCL subtypes; (C) heatmap of the relevance between module eigengenes and DLBCL subtypes; (D) scatter plot of module eigengenes in the turquoise module.

Identification of candidate hub genes in the turquoise module

According to the STRING database, the PPI network for all genes in the turquoise module consisted of 119 nodes and 660 edges. Moreover, under the cutoff of node connectivity > 6, 70 genes were identified as candidate hub genes (Fig. 5). The highly connected genes in the
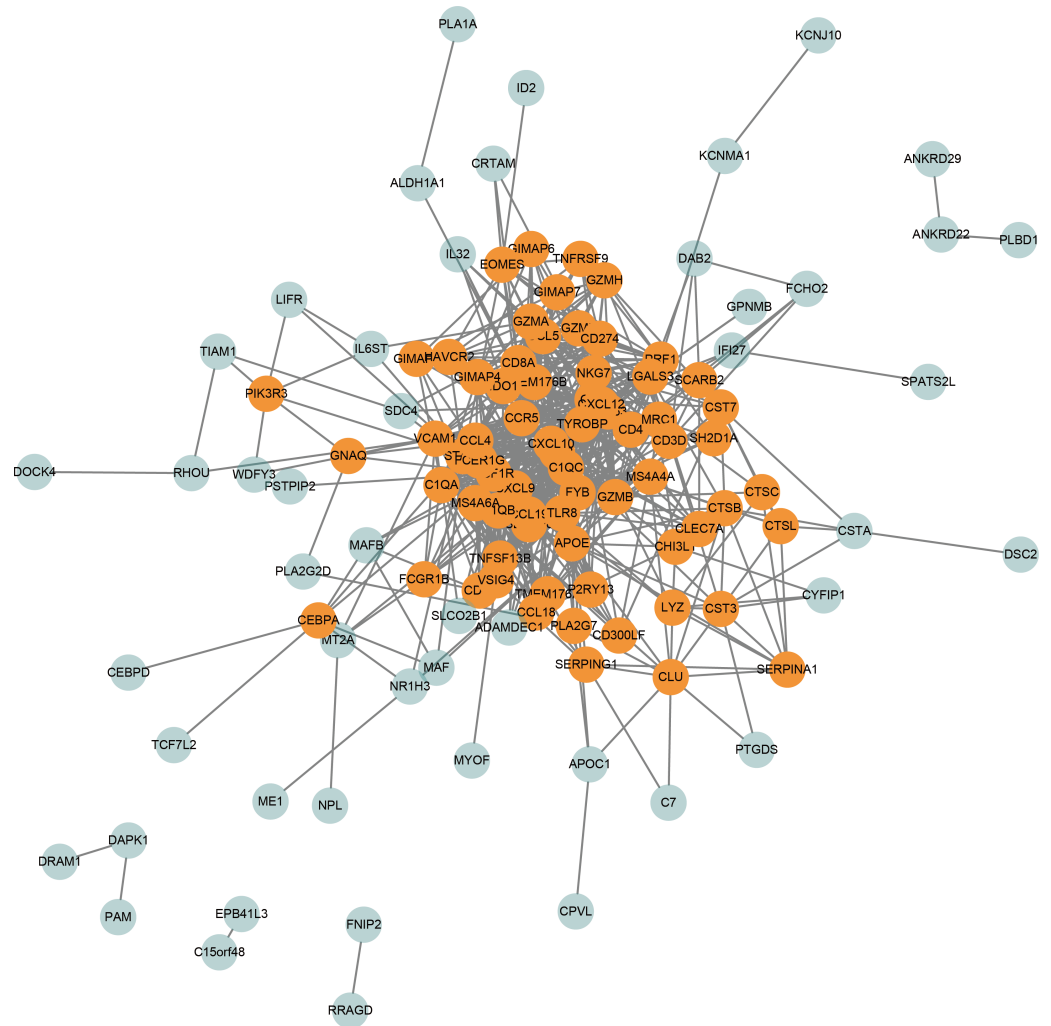
Figure 5 **PPI network of genes in the turquoise module.** The orange dots represent candidate hub genes within the turquoise module.

network are considered functionally important. Therefore the 70 genes were selected for further hub gene screening.

Hub gene screening

Kaplan–Meier survival analysis was used to determine survival with respect to candidate hub genes. Increased expression of *CCR5* and *CXCL10*, which were upregulated in the cancer group, predicted a significantly shorter overall survival (OS) (Figs. 6A–6B, Figs. 6F–6G). Next, the prognosis values of *CCR5* and *CXCL10* were assessed in different subtypes of DLBCL separately in GSE10846. Results show that high expression of CCR5 had no prognostic impact on survival in HR subtype and BCR subtype, but predicted a good prognosis in Oxphos subtype ($P = 0.042$) (Figs. 6C–6E). Besides, high expression of *CXCL10* had a significant adverse effect on the survival of HR subtype but had no significant prognostic impact on BCR subtype and Oxhos subtype (Figs. 6H–6J). Importantly, survival

analysis showed that high *CXCL10* expression was associated with poor outcome in DLBCL in general, and in the HR subtype in particular. Thus, *CXCL10* was identified as the real hub gene of HR DLBCL.

Validation of the expression level of the hub gene in DLBCL

Based on the GEPIA and Oncomine databases, we found that the expression of *CXCL10* was significantly elevated in DLBCL compared with normal tissues (Figs. 7A–7B). In addition, we examined the *CXCL10* expression profiles across all tumor samples and paired normal tissues. The log2-fold change in expression of *CXCL10* in DLBCL relative to that in paired normal tissues was significantly greater than that in any other type of tumor vs. normal (Fig. 7C).

The result of western blotting also revealed significantly higher expression of *CXCL10* in DLBCLs compared to lymphoid tissue (Figs. 7D–7E). Results suggested that the *CXCL10* was significantly upregulated in DLBCL and showed tumor specificity to some extent.

Functional enrichment analysis

To investigate the pathogenesis of HR DLBCL, we conducted functional enrichment analysis of genes of the turquoise module according to the Metascape database. Bar graphs of the top 20 enriched terms across input gene lists, colored according to *P*-values, were shown in Fig. 8. Based on the *P*-values of these biological processes, these genes in the turquoise module were particularly enriched in regulation of cell activation, cytokine-mediated signaling pathway, monocyte chemotaxis, and other activities. The results indicate that the immune response might play a pivotal role in HR DLBCL.

GSEA

To identify the potential function of the real hub gene in DLBCL, GSEA was conducted to search for KEGG pathways of the hub gene. Several immune-related pathways were enriched in the *CXCL10* highly expressed group, including "Graft versus host disease", "Toll-like receptor signaling pathway", "T-cell receptor signaling pathway", "NOD-like receptor signaling pathway" was enriched in the *CXCL10* highly expressed group (Figs. 9A–9D). Therefore, *CXCL10* might be the key effector gene and have an effect on tumor immunity in HR DLBCL.

Correlations of hub genes expression with immune infiltration level and immune makers in DLBCL

To investigate how the expression of *CXCL10* correlated with immune infiltration levels in DLBCL, we searched TIMER database. *CXCL10* expression is negatively associated with tumor purity ($r = -0.493$, $p = 9.16\mathrm{E}-04$) and positively associated with the infiltration levels of neutrophil, and dendritic cell (Figs. 10A–10G). We analyzed the association between *CXCL10* expression and various immune markers further to understand the crosstalk of the hub gene with immune genes. The top 10 strongest positive correlative signatures of *CXCL10* are shown in Fig. 10H. All the relative immune signatures are listed in Table S1 (|cor|>0.3 and $p < 0.05$ is recommended).
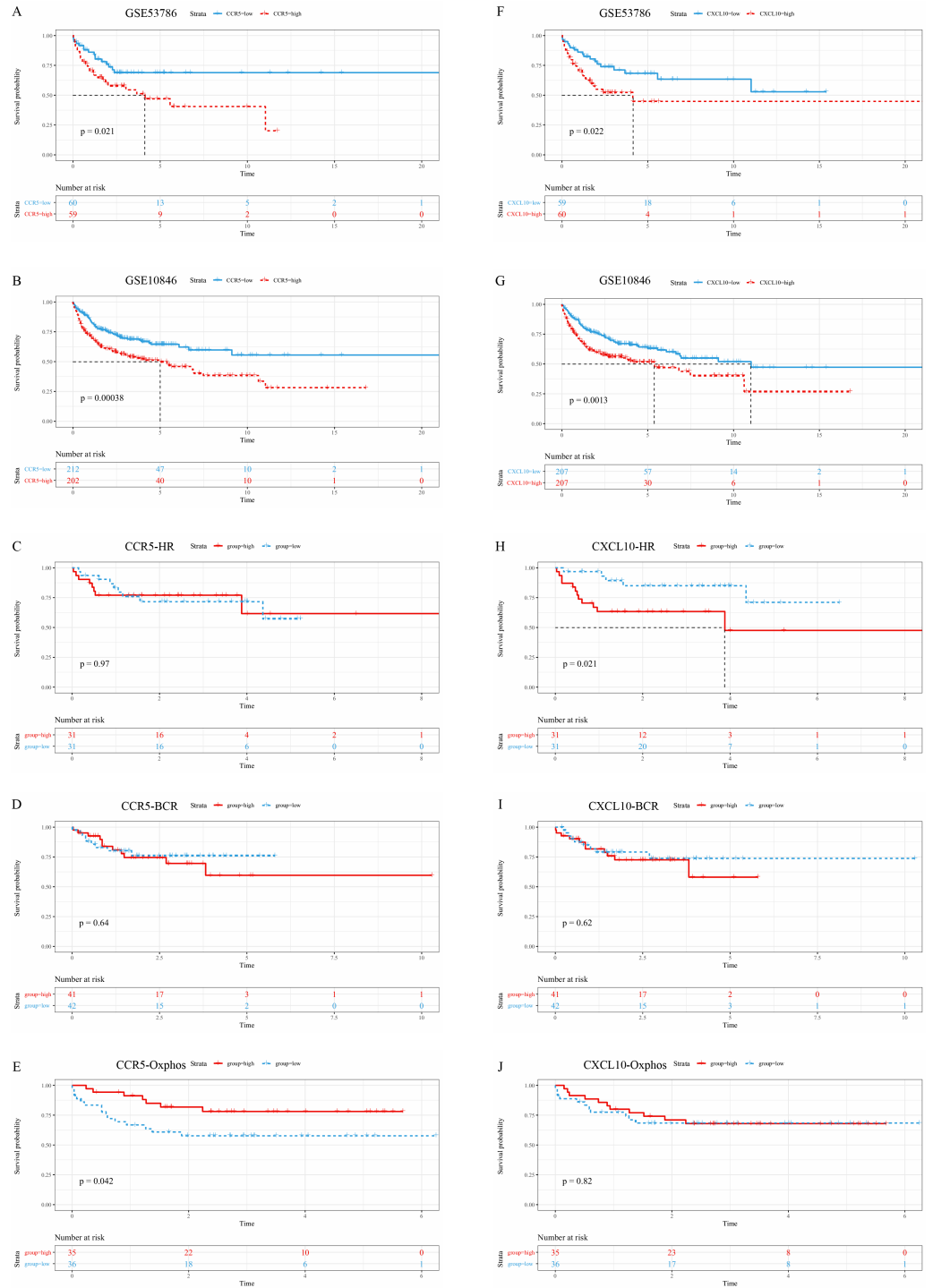
Figure 6 Survival analysis of hub genes. (A–B) Kaplan–Meier overall survival analysis of *CCR5* in GSE53786 and GSE10846; (C–E) Kaplan–Meier overall survival analysis of *CCR5* in the HR subtype, BCR subtype and Oxhos subtype; (F–G) Kaplan–Meier overall survival analysis of *CXCL10* in GSE53786 and GSE10846; (H–J) Kaplan–Meier overall survival analysis of *CXCL10* in HR subtype, BCR subtype and Oxhos subtype.

A

B

C

D

E
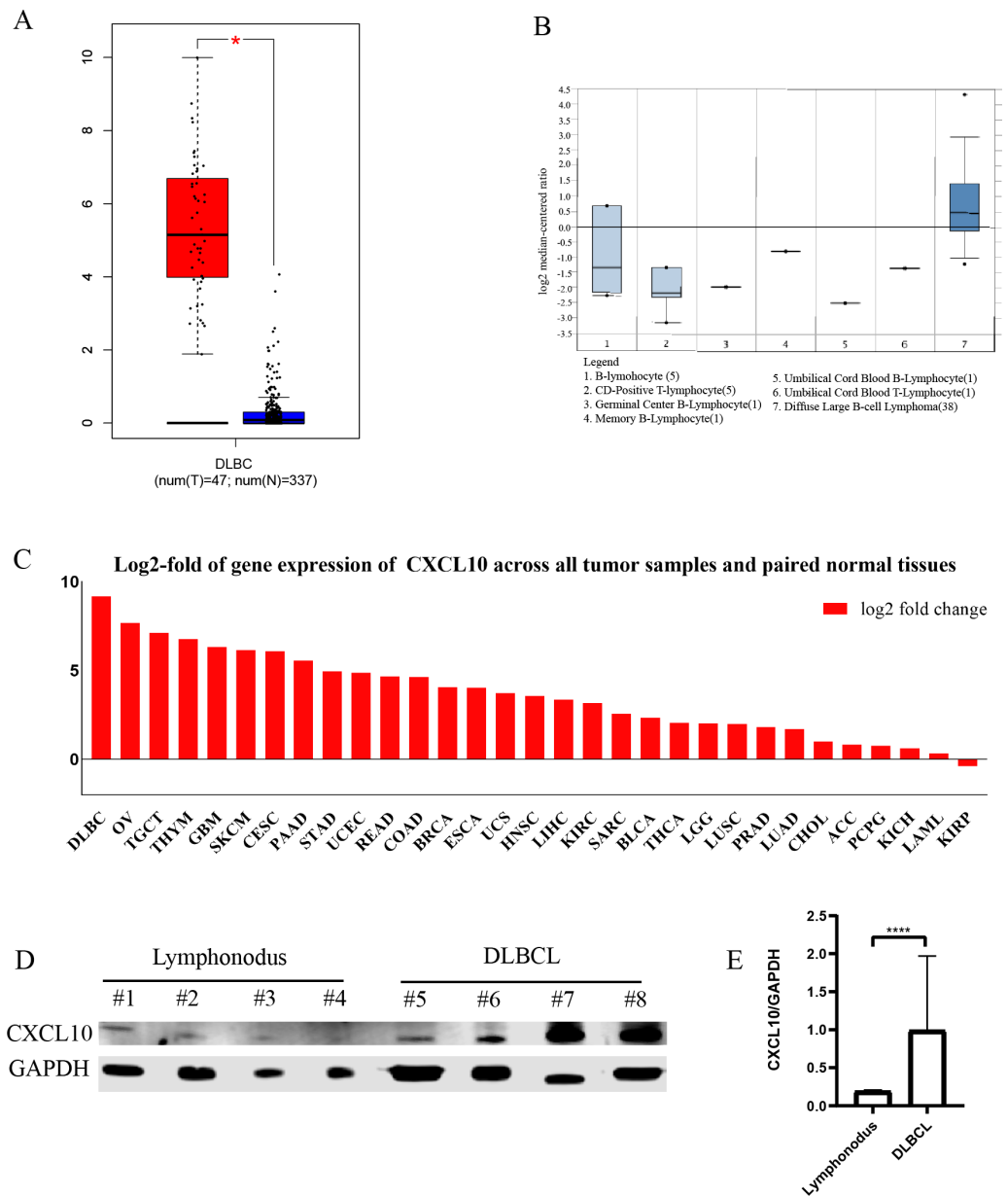
Figure 7 The expression of the hub gene in DLBCL. (A) The mRNA expression of *CXCL10* in DLBCL was obtained from GEPIA; (B) The mRNA expression of *CXCL10* in DLBCL was obtained from Oncomine; (C) Log2-fold of gene expression of *CXCL10* across all tumor samples vs. paired normal tissues. (D) Western blot result of *CXCL10* expression levels of lymph node samples ($n = 4$), patients with DLBCL ($n = 4$); (E) Bar chart of *CXCL10* relative protein expression (gray value of the target protein bands/gray value of the GAPDH protein bands). Mean expression ± SD is shown. (**** $P < 0.0001$, Student's *t*-test).

Construction of predicted miRNA- and TF- target regulatory networks

To further explore the regulation mechanisms of the hub module and hub genes, upstream regulators were predicted. A total of 630 miRNAs were predicted in the turquoise module. The network of *hsa-mir-7110-3p* with the highest degree and targeted genes in the turquoise

GO:0050865: regulation of cell activation
GO:0019221: cytokine-mediated signaling pathway
GO:0002548: monocyte chemotaxis
M5885: NABA MATRISOME ASSOCIATED
GO:0002250: adaptive immune response
GO:0002274: myeloid leukocyte activation
GO:0071346: cellular response to interferon-gamma
M54: PID IL12 2PATHWAY
GO:0050727: regulation of inflammatory response
GO:0002521: leukocyte differentiation
GO:0042116: macrophage activation
GO:0071887: leukocyte apoptotic process
GO:0002253: activation of immune response
GO:0019835: cytolysis
GO:0030100: regulation of endocytosis
GO:0001819: positive regulation of cytokine production
hsa04620: Toll-like receptor signaling pathway
GO:0009617: response to bacterium
R-HSA-1280218: Adaptive Immune System
GO:0035821: modification of morphology or physiology of other organism

Figure 8 **Functional and pathway enrichment analysis of DEGs of turquoise module.** GO terms and KEGG pathway are presented, and each band represents one enriched term or pathway colored according to the −log 10 *P* value.

module was displayed (Fig. 11A). The predicted TF with the highest normalized enrichment score was displayed (Fig. 11B). *IRF9* (NES = 8.140) was predicted to regulate 9 targets in the turquoise module.

DISCUSSION

Patients with HR DLBCL display unique clinical features, with splenomegaly and bone marrow involvement being more common. In addition, HR DLBCL cases have less frequent genetic abnormalities and occur in younger patients (*Monti et al., 2005*). Despite its clinical significance, the pathogenetic mechanism of HR DLBCL is poorly understood. In this study, comprehensive bioinformatic analyses of multiple datasets were used to explore the hub genes and essential pathways associated with HR DLBCL.

Disease-specific differential expression genes reveal potential biological mechanism linked to disease development. Due to the high heterogeneity of DLBCL, we performed differential genes analysis in the 3 different datasets and get overlap DEGs, to identify highly and stably expressed genes as well as lowly and stably expressed genes for further exploration of biological mechanism. Two hundred and seventy-four DEGs were identified in the three datasets. Weighted correlation network analysis was used to establish a co-expression network of common DEGs. The analysis revealed the turquoise module comprising genes significantly associated with HR DLBCL in GSE53786.

Among the genes in the turquoise module, nodes with high degree were considered functionally important. Thus 70 candidates with the highest degree of PPI network were selected. After a series of survival analyses, *CXCL10,* having significant prognosis value both in overall DLBCL patients and HR subtype, was identified as the real hub gene. These findings might contribute to the guide differential diagnosis and prognosis prediction for patients with DLBCL.

All results from the GEPIA, Oncomine databases and human tissue specimens are consistent with the findings in our study. The expression level of *CXCL10* was higher in
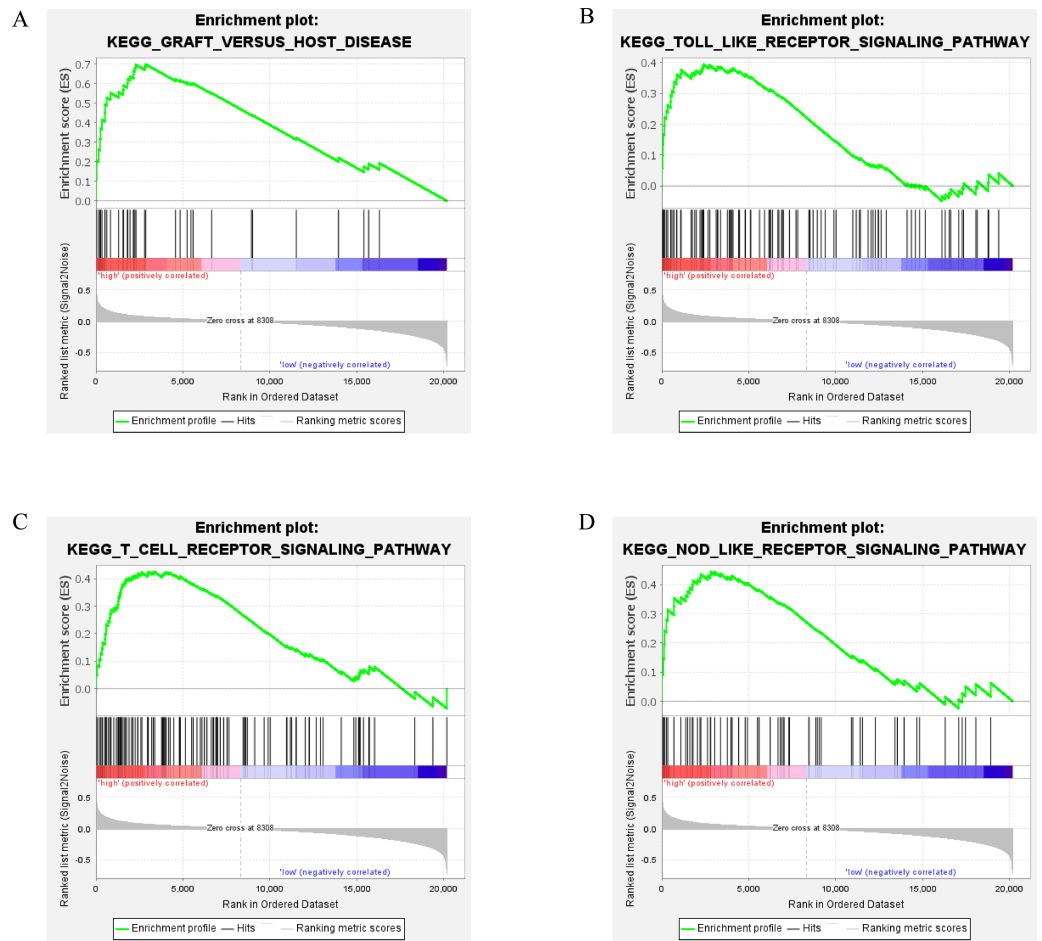
Figure 9 **Gene set enrichment analysis in HR DLBCL.** GSEA results indicating a significant correlation between the *CXCL10* expression level and the immune response, including (A) graft versus host disease, (B) Toll like receptor signaling pathway, (C) T cell receptor signaling pathway, (D) Nod like receptor signaling pathway.

DLBCL samples than the levels in normal tissue. According to expression profiles across all tumor samples and paired normal tissues in GEPIA, the log2-fold change in expression of *CXCL10* in DLBCL relative to that of paired normal tissues was significantly higher than that of any other type of tumor versus normal. The differential expression of *CXCL10* in DLBCL exhibited specificity to some extent. We identified that *CXCL10* is a tumor-specific gene for DLBCL. Importantly, *CXCL10* was associated with poor outcome in DLBCL, especially belonging to HR subtype. Therefore, we believe that targeting *CXCL10* may be a promising therapy with fewer side effects for DLBCL patients, especially for HR subtype.

C-X-C motif ligand 10 (*CXCL10*) is one of the chemokines commonly released by inflammatory cells (*Gupta et al., 2012*). Several studies (*Ansell et al., 2012*; *Hong et al., 2017*) identified the serum level of *CXCL10* as an inflammatory prognostic biomarker in patients with DLBCL. Presently, the elevated expression of *CXCL10* in DLBCL tumor tissue samples predicted disease prognosis in DLBCL. However, the exact mechanisms of
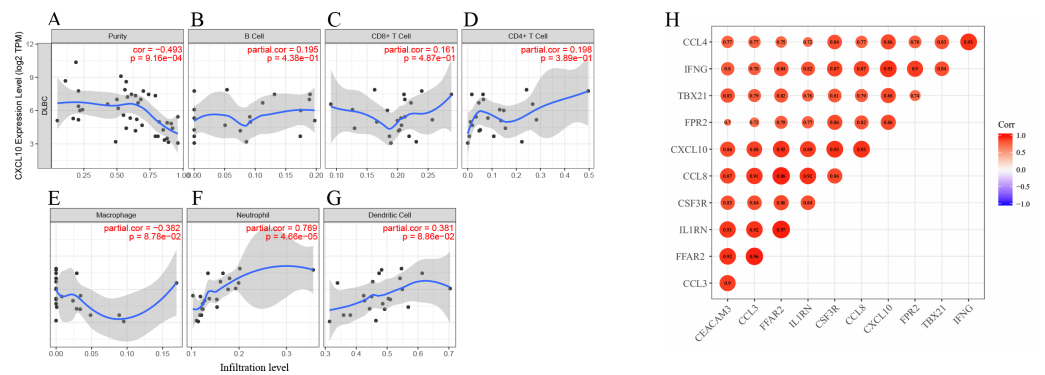
Figure 10 Correlations of the hub gene expression with immune microenvironment. (A–G) *CXCL10* expression level has significant correlations with tumor purity, macrophage, neutrophil, dendritic cell; (H) The correlation plot of top ten *CXCL10* correlated immune markers.
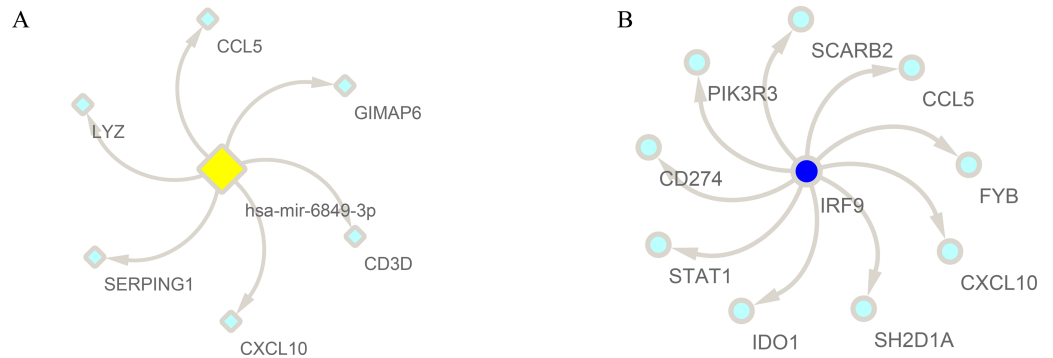
Figure 11 Prediction of potential regulators in the module of interest. (A) The turquoise nodes in the circle represent DEGs within the modules that regulated by hsa-mir-6849-3p; (B) the turquoise nodes in the circle represent DEGs within the modules that regulated by IRF9.

CXCL10 are yet to be reported in DLBCL. *Lee et al. (2012)* reported an in vivo study that monocytes promote migration and invasion of tumor cells via *CXCL10* expression in B-cell lymphoma cell lines. Natural killer (NK) cells are innate lymphocytes that are crucial in the immune response against tumor. *Bernardini et al. (2017)* demonstrated that NK cell compartment in MM could be modulated by the expression levels of *IP10/CXCL10*. *Wendel et al. (2008)* identified *CXCL10-CXCR3* signaling in NK cells as prerequisites for NK cell infiltration into tumors. The authors suggested that high levels of *CXCL10* in the tumor microenvironment represent a valuable target for therapeutic intervention by affecting myeloma NK cell surveillance.

We preliminarily explore the biological characteristics and function of the hub module and the hub gene by functional enrichment analysis and GSEA analysis. The functional enrichment analysis revealed that the DEGs in the turquoise module were markedly enriched in the regulation of cell activation, cytokine-mediated signaling pathway, monocyte chemotaxis, and other activities. GSEA analysis suggested that *CXCL10* get

involved several immune-associated pathways. Of these, Toll-like receptors (TLRs) participate the regulation of immune responses to infection as innate immune receptors, and also get involved in noninfectious inflammatory diseases, like tumor invasion, survival, and tumorigenicity (*Hua et al., 2009*). Nucleotide-binding oligomerization domain (NOD)-like receptor signaling pathway is involved in the formation of inflammasomes, and numerous types of cancer are associated with inflamed tissue (*Castano-Rodriguez et al., 2015*). Both Toll-like receptors and Nod-like receptors utilize the NF-κB pathway (*Van Driel et al., 2015*). It has been reported that Constitutive NF-κB activation existed in DLBCL and provide advantages for proliferation and survival of these tumor cells (*Compagno et al., 2009*). The results have shown that the pathogenesis of HR DLBCL is closely associated with tumor immunity, and these potential mechanisms require further exploration.

Herein, the relationship between the hub gene and tumor immunity was assessed by TIMER database, which applies a deconvolution method to estimate tumor purity and the abundance of tumor-infiltrating immune cells from gene expression profiles. It was inferred that *CXCL10* is negatively associated with tumor purity and positively related to dendritic cells and neutrophil. With the increased intratumoral T cells and dendritic cells, there was a relative decrease of neoplastic B cells in HR DLBCL. Furthermore, we investigated the association of *CXCL10* with immune markers in DLBCL patients. Among the top ten immune makers correlated with *CXCL10*, chemokines *CCL3* and *CCL4* are identified as biomarkers for B cell receptor pathway activation and prognostic serum markers in DLBCL (*Takahashi et al., 2015*). Based on these results of the analyses, it is inferred that *CXCL10* might have a regulatory effect on tumor immunity.

Our upstream regulator analysis showed that predicted miRNA (hsa-mir-6849-3p) might regulate genes expression in the hub module. Interferon regulatory factor 9 (IRF9) , the upstream transcriptional factor of hub module, belongs to IRF family and has an established role in type I interferon responses (*Sobhkhez et al., 2014*).

CONCLUSIONS

All in all, by using comprehensive bioinformatics analyses, we identified *CXCL10* as the real hub gene associated with HR DLBCL, which specific highly expressed in DLBCL and may also serve as a prognostic biomarker. And our further analysis showed that CXCL10 might have a regulatory effect on tumor immunity, which is in accordance with characteristics of immune infiltration in DLBCL. Although our results are preliminary, they provide novel insights into the molecular mechanisms of HR DLBCL. These findings will inform the development of clinically useful treatments. The tumor microenvironment is composed of non-cancerous cells and cytokines present in and around a tumor, having a major impact on the genomic analysis of tumor samples (*Aran, Sirota & Butte, 2015*). Tumor treatment strategies targeting both tumor cells and tumor microenvironment may contribute to HR DLBCL treatment.

Funding
This work was supported by the grant from the Nature Science Foundation of Zhejiang Province (Nos. Y17H160070). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures
The following grant information was disclosed by the authors:
Nature Science Foundation of Zhejiang Province: Y17H160070.

Competing Interests
The authors declare there are no competing interests.

Author Contributions
- Lingna Zhou, Ren Zhou and Wei Zhang conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Liya Ding, Yuqi Gong, Jing Zhao and Gong Xin performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Microarray Data Deposition
The following information was supplied regarding the deposition of microarray data:
Data are available at NCBI GEO (GSE25638, GSE44337, GSE56315, GSE53786, GSE10846) and TCGA (TCGA-DLBCL).

Data Availability
The datasets used in our study are available in the Supplementary Files.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.10269#supplemental-information.

REFERENCES

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**:503–511 DOI 10.1038/35000501.

Ansell SM, Maurer MJ, Ziesmer SC, Slager SL, Habermann TM, Link BK, Witzig TE, Macon WR, Dogan A, Cerhan JR, Novak AJ. 2012. Elevated pretreatment serum levels of interferon-inducible protein-10 (CXCL10) predict disease relapse and prognosis in diffuse large B-cell lymphoma patients. *American Journal of Hematology* **87**:865–869 DOI 10.1002/ajh.23259.

Aran D, Sirota M, Butte AJ. 2015. Systematic pan-cancer analysis of tumour purity. *Nature Communications* **6**:8971 DOI 10.1038/ncomms9971.

Bernardini G, Vulpis E, Bonanni V, Stabile H, Ricciardi MR, Petrucci MT, Gismondi A, Santoni A, Zingoni A. 2017. High expression levels of IP10/CXCL10 are associated with modulation of the natural killer cell compartment in multiple myeloma. *Leukemia and Lymphoma* **58**:2493–2496 DOI 10.1080/10428194.2017.1295144.

Cardesa-Salzmann TM, Colomo L, Gutierrez G, Chan WC, Weisenburger D, Climent F, Gonzalez-Barca E, Mercadal S, Arenillas L, Serrano S, Tubbs R, Delabie J, Gascoyne RD, Connors JM, Mate JL, Rimsza L, Braziel R, Rosenwald A, Lenz G, Wright G, Jaffe ES, Staudt L, Jares P, Lopez-Guillermo A, Campo E. 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 DOI 10.3324/haematol.2010.037408.

Castano-Rodriguez N, Kaakoush NO, Goh KL, Fock KM, Mitchell HM. 2015. The NOD-like receptor signalling pathway in Helicobacter pylori infection and related gastric cancer: a case-control study and gene expression analyses. *PLOS ONE* **10**:e0117870 DOI 10.1371/journal.pone.0117870.

Compagno M, Lim WK, Grunn A, Nandula SV, Brahmachary M, Shen Q, Bertoni F, Ponzoni M, Scandurra M, Califano A, Bhagat G, Chadburn A, Dalla-Favera R, Pasqualucci L. 2009. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature* **459**:717–721 DOI 10.1038/nature07968.

Cristino AS, Nourse J, West RA, Sabdia MB, Law SC, Gunawardana J, Vari F, Mujaj S, Thillaiyampalam G, Snell C, Gough M, Keane C, Gandhi MK. 2019. EBV microRNA-BHRF1-2-5p targets the 3′ UTR of immune checkpoint ligands PD-L1 and PD-L2. *Blood* **134**:2261–2270 DOI 10.1182/blood.2019000889.

Davis S, Meltzer PS. 2007. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**:1846–1847 DOI 10.1093/bioinformatics/btm254.

Dybkaer K, Bogsted M, Falgreen S, Bodker JS, Kjeldsen MK, Schmitz A, Bilgrau AE, Xu-Monette ZY, Li L, Bergkvist KS, Laursen MB, Rodrigo-Domingo M, Marques SC, Rasmussen SB, Nyegaard M, Gaihede M, Moller MB, Samworth RJ, Shah RD, Johansen P, El-Galaly TC, Young KH, Johnsen HE. 2015. Diffuse large B-cell lymphoma classification system that associates normal B-cell subset phenotypes with prognosis. *Journal of Clinical Oncology* **33**:1379–1388 DOI 10.1200/JCO.2014.57.7080.

Gupta M, Han JJ, Stenson M, Maurer M, Wellik L, Hu G, Ziesmer S, Dogan A, Witzig TE. 2012. Elevated serum IL-10 levels in diffuse large B-cell lymphoma: a mechanism of aberrant JAK2 activation. *Blood* **119**:2844–2853 DOI 10.1182/blood-2011-10-388538.

Hong JY, Ryu KJ, Lee JY, Park C, Ko YH, Kim WS, Kim SJ. 2017. Serum level of CXCL10 is associated with inflammatory prognostic biomarkers in patients with diffuse large B-cell lymphoma. *Hematological Oncology* **35**:480–486 DOI 10.1002/hon.2374.

Hua D, Liu MY, Cheng ZD, Qin XJ, Zhang HM, Chen Y, Qin GJ, Liang G, Li JN, Han XF, Liu DX. 2009. Small interfering RNA-directed targeting of Toll-like receptor 4 inhibits human prostate cancer cell invasion, survival, and tumorigenicity. *Molecular Immunology* **46**:2876–2884 DOI 10.1016/j.molimm.2009.06.016.

Janky R, Verfaillie A, Imrichova H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D, Svetlichnyy D, Kalender Atak Z, Fiers M, Marine JC, Aerts S. 2014. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLOS Computational Biology* **10**:e1003731 DOI 10.1371/journal.pcbi.1003731.

Jiang M, Chen P, Ruan X, Ye X, Pan Y, Zhang J, Huang Q, Zhou W, Wu H, Wang Q. 2017. Interim (18)F-FDG PET/CT improves the prognostic value of S-IPI, R-IPI and NCCN-IPI in patients with diffuse large B-cell lymphoma. *Oncology Letters* **14**:6715–6723 DOI 10.3892/ol.2017.7093.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**:559 DOI 10.1186/1471-2105-9-559.

Lee Y, Chittezhath M, Andre V, Zhao H, Poidinger M, Biondi A, D'Amico G, Biswas SK. 2012. Protumoral role of monocytes in human B-cell precursor acute lymphoblastic leukemia: involvement of the chemokine CXCL10. *Blood* **119**:227–237 DOI 10.1182/blood-2011-06-357442.

Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, Li B, Liu XS. 2017. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Research* **77**:e108–e110 DOI 10.1158/0008-5472.CAN-17-0307.

Liu Y, Yang Y, Luo Y, Wang J, Lu X, Yang Z, Yang J. 2020. Prognostic potential of PRPF3 in hepatocellular carcinoma. *Aging* **12**:912–930 DOI 10.18632/aging.102665.

Malumbres R, Chen J, Tibshirani R, Johnson NA, Sehn LH, Natkunam Y, Briones J, Advani R, Connors JM, Byrne GE, Levy R, Gascoyne RD, Lossos IS. 2008. Paraffin-based 6-gene model predicts outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. *Blood* **111**:5509–5514 DOI 10.1182/blood-2008-02-136374.

Monti S, Savage KJ, Kutok JL, Feuerhake F, Kurtin P, Mihm M, Wu B, Pasqualucci L, Neuberg D, Aguiar RC, Dal Cin P, Ladd C, Pinkus GS, Salles G, Harris NL, Dalla-Favera R, Habermann TM, Aster JC, Golub TR, Shipp MA. 2005. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* **105**:1851–1861 DOI 10.1182/blood-2004-07-2947.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. 2004. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**:1–6 DOI 10.1016/s1476-5586(04)80047-2.

Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES, Rosenwald A, Campo E, Chan WC, Connors JM, Smel EB, Mottok A, Braziel RM, Ott G, Delabie J, Tubbs RR, Cook JR, Weisenburger DD, Greiner TC, Glinsmann-Gibson BJ, Fu K, Staudt LM, Gascoyne RD, Rimsza LM. 2014. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* **123**:1214–1217 DOI 10.1182/blood-2013-11-536433.

Sobhkhez M, Skjesol A, Thomassen E, Tollersrud LG, Iliev DB, Sun B, Robertsen B, Jorgensen JB. 2014. Structural and functional characterization of salmon STAT1, STAT2 and IRF9 homologs sheds light on interferon signaling in teleosts. *FEBS Open Bio* **4**:858–871 DOI 10.1016/j.fob.2014.09.007.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* **45**:D362–D368 DOI 10.1093/nar/gkw937.

Takahashi K, Sivina M, Hoellenriegel J, Oki Y, Hagemeister FB, Fayad L, Romaguera JE, Fowler N, Fanale MA, Kwak LW, Samaniego F, Neelapu S, Xiao L, Huang X, Kantarjian H, Keating MJ, Wierda W, Fu K, Chan WC, Vose JM, O'Brien S, Davis RE, Burger JA. 2015. CCL3 and CCL4 are biomarkers for B cell receptor pathway activation and prognostic serum markers in diffuse large B cell lymphoma. *British Journal of Haematology* **171**:726–735 DOI 10.1111/bjh.13659.

Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. 2017. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research* **45**:W98–W102 DOI 10.1093/nar/gkx247.

Tompkins VS, Han SS, Olivier A, Syrbu S, Bair T, Button A, Jacobus L, Wang Z, Lifton S, Raychaudhuri P, Morse 3rd HC, Weiner G, Link B, Smith BJ, Janz S. 2013. Identification of candidate B-lymphoma genes by cross-species gene expression profiling. *PLOS ONE* **8**:e76889 DOI 10.1371/journal.pone.0076889.

Van Driel B, Wang G, Liao G, Halibozek PJ, Keszei M, O'Keeffe MS, Bhan AK, Wang N, Terhorst C. 2015. The cell surface receptor Slamf6 modulates innate immune responses during Citrobacter rodentium-induced colitis. *International Immunology* **27**:447–457 DOI 10.1093/intimm/dxv029.

Vicente-Duenas C, Fontan L, Gonzalez-Herrero I, Romero-Camarero I, Segura V, Aznar MA, Alonso-Escudero E, Campos-Sanchez E, Ruiz-Roca L, Barajas-Diego M, Sagardoy A, Martinez-Ferrandis JI, Abollo-Jimenez F, Bertolo C, Penuelas I, Garcia-Criado FJ, Garcia-Cenador MB, Tousseyn T, Agirre X, Prosper F, Garcia-Bragado F, McPhail ED, Lossos IS, Du MQ, Flores T, Hernandez-Rivas JM, Gonzalez M, Salar A, Bellosillo B, Conde E, Siebert R, Sagaert X, Cobaleda C, Sanchez-Garcia I, Martinez-Climent JA. 2012. Expression of MALT1 oncogene in hematopoietic stem/progenitor cells recapitulates the pathogenesis of human

lymphoma in mice. *Proceedings of the National Academy of Sciences of the United States of America* **109**:10534–10539 DOI 10.1073/pnas.1204127109.

Wendel M, Galani IE, Suri-Payer E, Cerwenka A. 2008. Natural killer cell accumulation in tumors is dependent on IFN-gamma and CXCR3 ligands. *Cancer Research* **68**:8437–8445 DOI 10.1158/0008-5472.CAN-08-1440.

Ye Z, Wang F, Yan F, Wang L, Li B, Liu T, Hu F, Jiang M, Li W, Fu Z. 2019. Bioinformatic identification of candidate biomarkers and related transcription factors in nasopharyngeal carcinoma. *World Journal of Surgical Oncology* **17**:60 DOI 10.1186/s12957-019-1605-9.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications* **10**:1523 DOI 10.1038/s41467-019-09234-6.

Zhou Z, Cheng Y, Jiang Y, Liu S, Zhang M, Liu J, Zhao Q. 2018. Ten hub genes associated with progression and prognosis of pancreatic carcinoma identified by co-expression analysis. *International Journal of Biological Sciences* **14**:124–136 DOI 10.7150/ijbs.22619.