World Journal of
Clinical Oncology

Contents

World Journal of Clinical Oncology

Monthly Volume 14 Number 10 October 24, 2023

ABOUT COVER

Peer Reviewer of *World Journal of Clinical Oncology*, Amr Ahmed El-Arabey, PhD, Assistant Professor, Department of Pharmacology and Toxicology, Faculty of Pharmacy, Al Azhar University, Nasr City 11751, Cairo, Egypt. ph.amrcapa@gmail.com

AIMS AND SCOPE

The primary aim of *World Journal of Clinical Oncology* (*WJCO*, *World J Clin Oncol*) is to provide scholars and readers from various fields of oncology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online.

   WJCO mainly publishes articles reporting research results and findings obtained in the field of oncology and covering a wide range of topics including art of oncology, biology of neoplasia, breast cancer, cancer prevention and control, cancer-related complications, diagnosis in oncology, gastrointestinal cancer, genetic testing for cancer, gynecologic cancer, head and neck cancer, hematologic malignancy, lung cancer, melanoma, molecular oncology, neurooncology, palliative and supportive care, pediatric oncology, surgical oncology, translational oncology, and urologic oncology.

INDEXING/ABSTRACTING

The *WJCO* is now abstracted and indexed in PubMed, PubMed Central, Emerging Sources Citation Index (Web of Science), Reference Citation Analysis, China National Knowledge Infrastructure, China Science and Technology Journal Database, and Superstar Journals Database. The 2023 Edition of Journal Citation Reports® cites the 2022 impact factor (IF) for *WJCO* as 2.8; IF without journal self cites: 2.8; 5-year IF: 3.0; Journal Citation Indicator: 0.36.

RESPONSIBLE EDITORS FOR THIS ISSUE

Production Editor: *Xiang-Di Zhang*; **Production Department Director:** *Xu Guo*; **Editorial Office Director:** *Xu Guo*.

NAME OF JOURNAL
World Journal of Clinical Oncology

ISSN
ISSN 2218-4333 (online)

LAUNCH DATE
November 10, 2010

FREQUENCY
Monthly

EDITORS-IN-CHIEF
Hiten RH Patel, Stephen Safe, Jian-Hua Mao, Ken H Young

EDITORIAL BOARD MEMBERS
https://www.wjgnet.com/2218-4333/editorialboard.htm

PUBLICATION DATE
October 24, 2023

COPYRIGHT
© 2023 Baishideng Publishing Group Inc

INSTRUCTIONS TO AUTHORS
https://www.wjgnet.com/bpg/gerinfo/204

GUIDELINES FOR ETHICS DOCUMENTS
https://www.wjgnet.com/bpg/GerInfo/287

GUIDELINES FOR NON-NATIVE SPEAKERS OF ENGLISH
https://www.wjgnet.com/bpg/gerinfo/240

PUBLICATION ETHICS
https://www.wjgnet.com/bpg/GerInfo/288

PUBLICATION MISCONDUCT
https://www.wjgnet.com/bpg/gerinfo/208

ARTICLE PROCESSING CHARGE
https://www.wjgnet.com/bpg/gerinfo/242

STEPS FOR SUBMITTING MANUSCRIPTS
https://www.wjgnet.com/bpg/GerInfo/239

ONLINE SUBMISSION
https://www.f6publishing.com

© 2023 Baishideng Publishing Group Inc. All rights reserved. 7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
E-mail: bpgoffice@wjgnet.com https://www.wjgnet.com

ORIGINAL ARTICLE

Basic Study

Hub genes and their key effects on prognosis of Burkitt lymphoma

Yan-Feng Xu, Guan-Yun Wang, Ming-Yu Zhang, Ji-Gang Yang

Yan-Feng Xu, Guan-Yun Wang, Ming-Yu Zhang, Ji-Gang Yang, Department of Nuclear Medicine, Beijing Friendship Hospital, Capital Medical University, Beijing 100050, China

Corresponding author: Ji-Gang Yang, MD, PhD, Chief Doctor, Professor, Researcher, Department of Nuclear Medicine, Beijing Friendship Hospital, Capital Medical University, No. 95 Yong An Road, Xicheng District, Beijing 100050, China. yangjigang@ccmu.edu.cn

Abstract

BACKGROUND
Burkitt lymphoma (BL) is an exceptionally aggressive malignant neoplasm that arises from either the germinal center or post-germinal center B cells. Patients with BL often present with rapid tumor growth and require high-intensity multi-drug therapy combined with adequate intrathecal chemotherapy prophylaxis, however, a standard treatment program for BL has not yet been established. It is important to identify biomarkers for predicting the prognosis of BLs and discriminating patients who might benefit from the therapy. Microarray data and sequencing information from public databases could offer opportunities for the discovery of new diagnostic or therapeutic targets.

AIM
To identify hub genes and perform gene ontology (GO) and survival analysis in BL.

METHODS
Gene expression profiles and clinical traits of BL patients were collected from the Gene Expression Omnibus database. Weighted gene co-expression network analysis (WGCNA) was applied to construct gene co-expression modules, and the cytoHubba tool was used to find the hub genes. Then, the hub genes were analyzed using GO and Kyoto Encyclopedia of Genes and Genomes analysis. Additionally, a Protein-Protein Interaction network and a Genetic Interaction network were constructed. Prognostic candidate genes were identified through overall survival analysis. Finally, a nomogram was established to assess the predictive value of hub genes, and drug-gene interactions were also constructed.

RESULTS
In this study, we obtained 8 modules through WGCNA analysis, and there was a significant correlation between the yellow module and age. Then we identified 10 hub genes (*SRC*, *TLR4*, *CD40*, *STAT3*, *SELL*, *CXCL10*, *IL2RA*, *IL10RA*, *CCR7* and *FCGR2B*) by cytoHubba tool. Within these hubs, two genes were found to be associated with OS (*CXCL10*, *P* = 0.029 and *IL2RA*, *P* = 0.0066) by survival ana-

lysis. Additionally, we combined these two hub genes and age to build a nomogram. Moreover, the drugs related to *IL2RA* and *CXCL10* might have a potential therapeutic role in relapsed and refractory BL.

CONCLUSION
From WGCNA and survival analysis, we identified *CXCL10* and *IL2RA* that might be prognostic markers for BL.

Key Words: Burkitt lymphoma; Weighted gene co-expression network analysis; Microarray data; Functional enrichment analysis; Prognosis; Therapeutic target

Core Tip: This study represents the pioneering investigation of gene expression in Burkitt lymphoma (BL) using weighted gene co-expression network analysis, coupled with functional enrichment analysis. In this study, we have successfully identified and validated 10 hub genes. Survival analysis has demonstrated that the overexpression of *CXCL10* and *IL2RA* in BL may serve as robust prognostic indicators. Furthermore, an integrated mRNA signature and age nomogram potentially provide valuable prognostic insights for patients with BLs.

INTRODUCTION

Burkitt lymphoma (BL) is a highly aggressive B-cell non-Hodgkin's lymphoma characterized by the *t* (8; 14) chromosomal translocation involving the *MYC* oncogene and the immunoglobulin heavy chain gene (IGH)[1]. Three distinct clinical subtypes of BL have been identified: Namely endemic (African), sporadic (non-endemic), and immunodeficiency-associated. Notably, chronic Epstein-Barr virus infection plays a pivotal role in the pathogenesis of BL, particularly in the endemic subtype[2]. Endemic BL is primarily found in countries located near the equator in Africa. The estimated annual incidence of endemic BL is 3-6 cases per 100000 children in African countries[3], which is approximately 50 times higher than that in the United States[4]. Sporadic BL predominantly occurs in the United States and Western Europe. The annual incidence of BL in the United States is approximately 3 cases per 1 million individuals, while in Europe it stands at around 2.2 cases per 1 million people[5]. Immunodeficiency-associated BL primarily affects individuals with HIV infection, typically those with relatively high CD4 counts and no opportunistic infections[6].

Patients with BL frequently exhibit rapid tumor growth, spontaneous tumor lysis, and elevated levels of serum lactate dehydrogenase. Currently, patients with BL necessitate a high-intensity multi-drug regimen in conjunction with adequate intrathecal central nervous system prophylaxis. However, the absence of an established standard treatment protocol for BL persists[7]. BL is an aggressive lymphoma, which can potentially be cured; however, patients with refractory and relapsed disease have an extremely poor prognosis[8]. Therefore, it is important to identify robust biomarkers for predicting the prognosis of BLs and discriminating patients who might benefit from therapy. The development of BL depends on the constitutive expression of the MYC gene located on chromosome 8q24, which encodes the transcription factor protein MYC[9]. MYC orchestrates the expression of target genes, regulating a variety of cellular processes, including cell growth, division, apoptosis, metabolism, adhesion, and motility[10]. MYC gene rearrangements are seen in the vast majority of BLs, and factors other than MYC translocation need to be present in the process of BL. However, it is not clear why and how B cells develop genetic alterations that result in increased MYC expression and ultimately lead to BL.

The Gene Expression Omnibus (GEO) is an international public repository constructed and maintained by the National Center for Biotechnology Information[11]. At the time of writing, the GEO database hosts more than 194000 public series. Weighted gene co-expression network analysis (WGCNA) is a widely used systematic biological method for generating gene co-expression networks[12,13]. In this study, WGCNA was first used to analyze genes of BL samples mined from the GEO database. Subsequently, we identified these hub genes and conducted a functional enrichment analysis. Additionally, a survival analysis was conducted to identify an mRNA signature that exhibits a significant association with prognosis. Finally, a prognostic nomogram was established based on the combination of gene signature and clinical characteristics.

MATERIALS AND METHODS

Data collection and preprocessing

The raw gene expressions and the corresponding clinical follow-up data of GSE4475 and GSE69051 were downloaded from the GEO database (http://www.ncbi.nlm.nih.gov/geo/. Accessed Jan 20, 2023)[14], and the two datasets were built based on the GPL96 platform (HG-U133A) and GPL14951 platform (Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip) respectively. Analysis was performed on the raw gene expression data of the BL datasets and the corresponding clinical follow-up obtained from GSE4475, which included a total of 36 BL samples. The survival data of the hub genes was verified by downloading another dataset, GSE69051, which included 77 BL samples. The mRNA sequencing data annotation information was used to match the probe with the corresponding gene and transform the gene name into a gene symbol. Probes that corresponded to more than one gene were excluded from the dataset.

Co-expression network construction

WGCNA converts gene expression data into co-expression modules, establishing relationships between genes and focusing on gene modules rather than individual genes[15]. Besides, WGCNA can identify the gene modules related to clinical traits and has been widely used in cancer research. In this study, the top 5000 most variable genes were used to construct a co-expression network by using the package of WGCNA in R[13]. The power value was filtered out during the module construction process using the WGCNA algorithm. The mean connectivity and scale independence of network modules were analyzed using the gradient test under different power values, which ranged from 1 to 20. When the degree of independence was 0.85, the appropriate power value was determined. Then, the soft threshold test was performed. In this study, the soft threshold β was 12, and the network type was "signed". The WGCNA algorithm further identified co-expression modules under these conditions. The minimum size of the gene group was set at 100 to ensure the reliability of the results for this module. Then, the correlation between the characteristics of the module-trait association module and clinical traits was visually expressed. The relationship between the expression profile and traits was analyzed to make a scatter plot between gene significance and module membership.

Hub genes identification and functional analysis

The Protein-Protein Interaction (PPI) network of the interested module was constructed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (available from https://cn.string-db.org/. Accessed 25 Jan 2023)[16] and presented by Cytoscape software. The cytoHubba tool was used to screen the hub genes. Then, the selected hub genes were analyzed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis using Database for Annotation, Visualization, and Integrated Discovery (DAVID v.6.8: available from https://david.ncifcrf.gov/. Accessed Jan 29, 2023)[17]. The possible functions were analyzed by biological processes, cellular components, and molecular functions, while the potential signal pathways were analyzed using KEGG.

Construction of hub genes PPI and genetic interaction network

The PPI network was used to analyze the hub genes at the protein level, and the STRING database (available from https://string-db.org/. Accessed Jan 31, 2023) was used to check and predict the interaction between proteins[16]. The genetic interaction (GI) network, constructed using gene function prediction, aims to understand the complex interactions between genes. We used the Gene Multiple Association Network Integration Algorithm (GeneMANIA, available from https://genemania.org/. Accessed Jan 31, 2023) to analyze the hub genes[18]. The threshold of a collective score of 0.15 was implemented.

Statistical analysis

Based on the 50th percentile cut-off value of each hub gene mRNA, patients were divided into high-expression and low-expression groups. Log-rank test and Kaplan-Meier estimation were performed to obtain log-rank P value and evaluate hub genes in overall survival (OS). Cox regression analysis was conducted to examine the association between the risk score and clinical information, and a nomogram was generated. The survival curve and nomogram were carried out by R version 4.2.1. Additionally, $P < 0.05$ was statistically significant.

Drug-gene interaction

The DGIdb database (available from https://dgidb.org/. Accessed Feb 20, 2023) was used to investigate drug-gene interaction to identify drugs associated with hub genes[19]. The interaction network was visualized *via* Cytoscape.

RESULTS

Construction and screening of BL co-expression module

In this study, we obtained the BL dataset from GSE4475, resulting in a total of 13514 gene expression values. The clinical features of the BL samples are listed in Table 1. Then, we selected a total of 5000 genes with the highest average expression values for cluster analysis. Firstly, the clustering tree of 36 samples of BL was extracted from GSE4475 (Figure 1). Secondly, we calculated the soft threshold (power value), and when the weight was equal to 12, the independence exceeded 0.85, indicating higher average connectivity (Supplementary Figure 1). By utilizing this power value for hierarchical clustering analysis and combining similar analysis results, a total of 8 modules were identified,

Table 1 Clinical features of Burkitt lymphoma patients

| Clinical features	Total (*n* = 36)
Age, mean (range)	31.0 (2-90)
Gender	
Male	24
Female	11
Unknown	1
Stage	
I	4
II	10
III	5
IV	6
Unknown	11
Survival status	
Alive	20
Dead	7
Unknown	9
Ki 67	
≤ 75%	4
75%-90%	9
> 90%	22
Unknown	1
CCS	
< 10	29
≥ 10	5
Unknown	4

CCS: Chromosomal Complexity Score.

including black (1073 genes), blue (967 genes), brown (853 genes), green yellow (140 genes), grey (1019 genes), magenta (219 genes), pink (267 genes) and yellow (462 genes) (Figure 2A). Genes in grey were not included in any module, thus we analyzed the interactive relationships underlying the 7 co-expression modules (Figure 2B). Given the well-established association between age and prognosis in BL patients, we opted to investigate the module that exhibited the strongest correlation with age for subsequent analysis[20,21]. A significant correlation between the yellow module and age was discovered (Figure 3). The correlation between modules and samples is shown in Supplementary Figure 2. Finally, we conducted a scatter diagram of the correlation between the yellow module and age (Figure 4).

Hub genes identification

All of the genes from the yellow module were uploaded to the STRING database, and a PPI network was constructed using Cytoscape software (Supplementary Figure 3). And the top 10 hub genes (*SRC, TLR4, CD40, STAT3, SELL, CXCL10, IL2RA, IL10RA, CCR7* and *FCGR2B*) were screened out by cytoHubba tool. GeneMANIA showed the GI network of hub genes interaction at the mRNA expression level (Figure 5A). The STRING database generated the PPI co-expression network by analyzing the hub genes at the protein level (Figure 5B).

Functional and pathway enrichment analysis

Enrichment analyses of GO and KEGG were conducted to explore potential pathways of the hub genes. Forty-five GO-enriched terms were shown in Supplementary Table 1. The top 10 GO terms (Figure 6A) included inflammatory response, external side of plasma membrane, plasma membrane, cellular response to lipopolysaccharide, positive regulation of interleukin-12 production, receptor binding, positive regulation of MAP kinase activity, positive regulation of JNK cascade, immune response and positive regulation of humoral immune response. In the KEGG analysis, 14 pathways enriched by genes in the yellow module were shown in Supplementary Table 2, and the top 10 KEGG terms (Figure 6B)
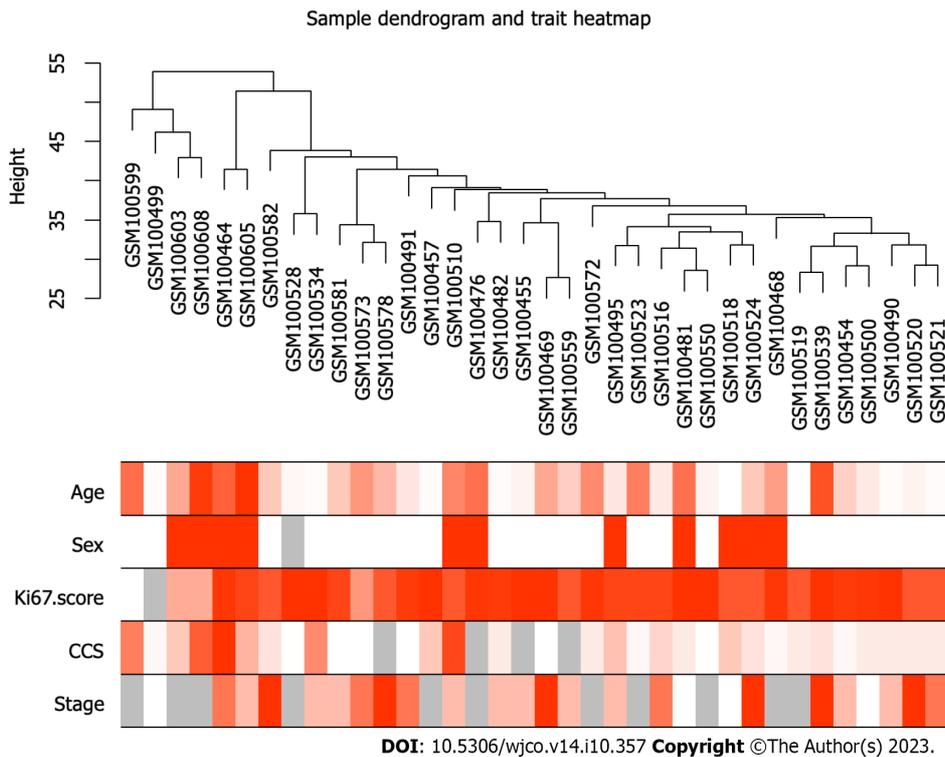
Figure 1 Clustering tree of 36 samples of Burkitt lymphoma extracted from GSE4475. Red indicated more gene expression, white less, and grey indicated deletion. CCS: Chromosomal Complexity Score.

included viral protein interaction with cytokine and cytokine receptor, cytokine-cytokine receptor interaction, toxoplasmosis, measles, tuberculosis, chemokine signaling pathway, lipid and atherosclerosis, Toll-like receptor signaling pathway, Hepatitis B and JAK-STAT signaling pathway.

Survival analysis

Additional survival analysis was conducted on the hub genes to evaluate their impact on BL patients' survival. Due to the small sample size of GSE4475, we opted for GSE69051 for survival analysis (Figure 7). Two of the 10 hub genes were significantly associated with OS: *CXCL10* ($P = 0.029$, Figure 7F) and *IL2RA* ($P = 0.0066$, Figure 7G).

Establishment of the nomogram and assessment of predictive value

Based on the hub genes and clinical data of the patients, a nomogram was developed to predict the 1- and 3-year OS of BL patients (Figure 8A). The model had a c-index of 0.84, and the calibration curve demonstrated strong agreement between predicted and observed survival times for both 1- and 3-year OS probabilities in the BL cohort (Figure 8B).

Identification of associated drugs

The drugs related to *IL2RA* and *CXCL10* were identified by the DGIdb database, as these were the only significant results from survival analysis (Figure 9). These results may provide new ideas for the treatment of BL with poor prognosis.

DISCUSSION

BL, a highly aggressive lymphoma identified and described by Denis Burkitt in the last century, continues to be the most common childhood malignancy in Africa nowadays[22]. A defining feature of BL is the translocation between the *c-MYC* gene and the *IgH* gene, which is found in 80% of cases [*t* (8; 14)], or between *c-MYC* and either the kappa or lambda light chain gene, which is found in the remaining 20% [*t* (2; 8) or *t* (8; 22)][23]. The proliferation rate and apoptosis rate of BL tumor cells are extremely high, indicating that nearly 100% of the cells are positive for Ki-67. Intensive, short-course combination chemotherapy is recommended for most BL patients. DA-EPOCH (dose-adjusted etoposide, doxorubicin, cyclophosphamide, vincristine, prednisone) + rituximab may be an option for patients who cannot tolerate more aggressive regimens[7]. As the standard of treatment for BL has not yet been established, strictly controlled clinical trials are also recommended. The prognosis of BL patients is associated with both clinical and laboratory characteristics[8,24]. The BL international prognostic index can be used to assess the prognosis of adult patients with disseminated or immunodeficiency-related BL, but it is not currently used for stratifying BL treatment[20]. Previous studies have demonstrated associations between *MYC* rearrangements, *TCF3* mutations or *ID3* alterations (its negative regulator), *TP53* modifications, *CCND3* and *CDKN2A* changes, as well as non-antigen-dependent B cell receptor signaling (tonic B

A

Gene dend rogram and module colors

B

Network heatmap plot
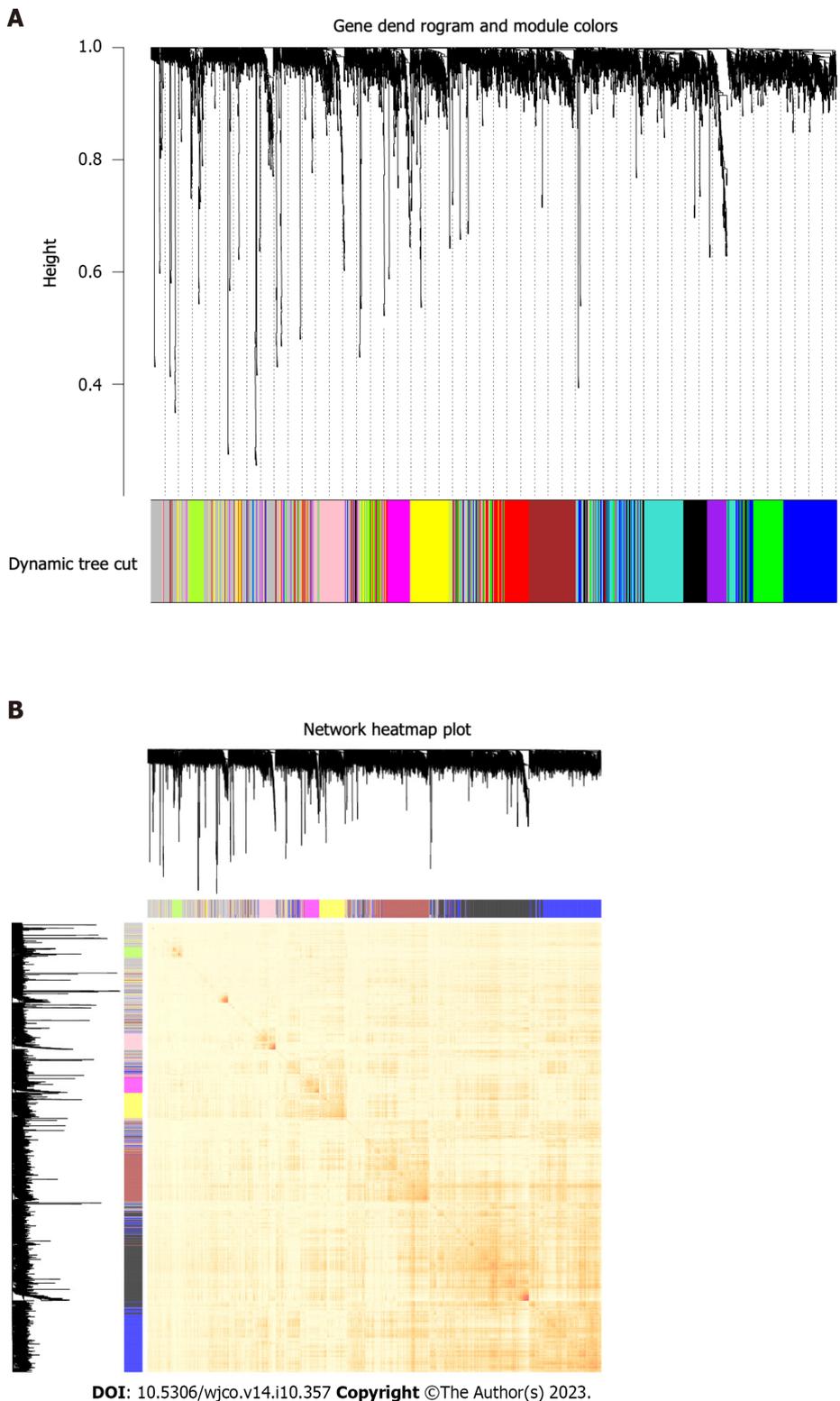
Figure 2 Sample clustering to detect outliers and construction of co-expression modules. A: The constructed co-expression modules of Burkitt lymphoma genes by weighted gene co-expression network analysis; B: Interaction analysis between gene co-expression modules. The heatmap showed the Topological Overlap Matrix among genes in the analysis. Different colors on the x-axis and y-axis represented different modules. The intensity of inter-module connections was visually represented by the yellow brightness in the central region, gradually transitioning into deeper shades of orange.

cell YF receptor signaling) with the development and prognosis of BL; however, a comprehensive investigation into the prognostic significance of molecular events associated with BL is lacking[25].

As a bioinformatics algorithm, WGCNA offers numerous advantages over conventional methods for differential expression analysis. It primarily focuses on elucidating co-expression patterns, facilitating the identification of biologically relevant modules comprising interconnected genes, and enabling the detection of pivotal hub genes[26-28]. So far, gene modules related to several cancers have been analyzed and verified using WGCNA[29,30].
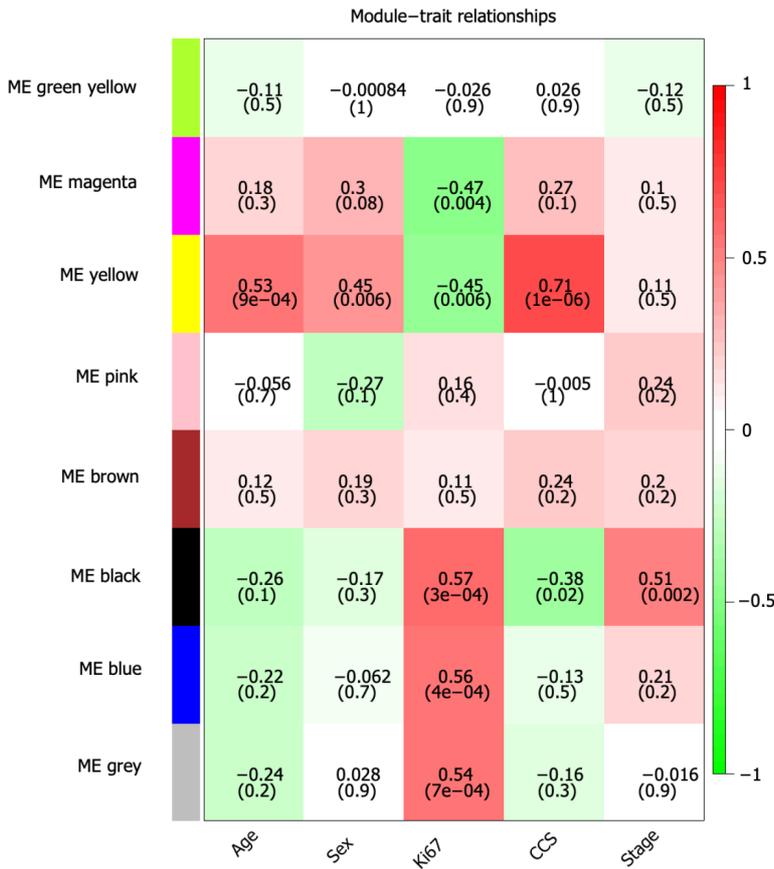
Figure 3 Module-trait association. Correlation thermography between modular feature genes and clinical features of Burkitt lymphoma. Each row corresponded to a module feature, and the column corresponded to a clinical feature. Each cell contained the correlation and the corresponding *P* value. CCS: Chromosomal Complexity Score; ME: Module membership.
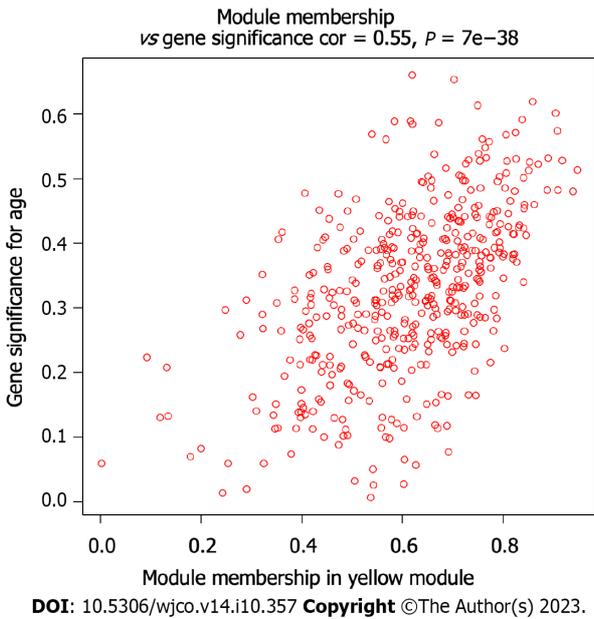
**Figure 4 The scatter plot of the correlation for an age-related gene between module membership and gene significance in the yellow module.**
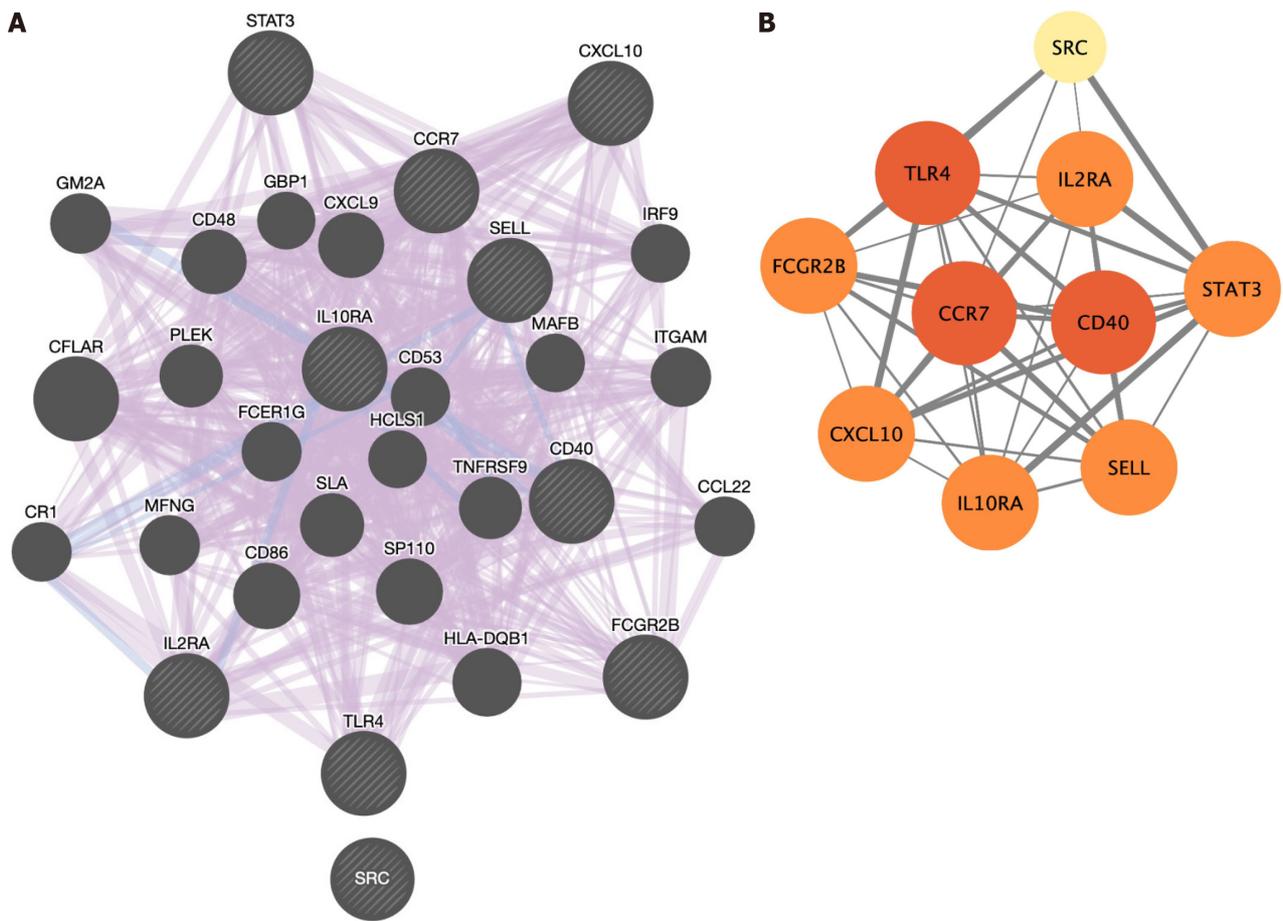
Figure 5 Genetic and Protein-Protein interaction network of hub genes. A: GeneMANIA was used to construct a genetic interaction network. The black nodes with a slash represent the query gene, while the other nodes represent the predicted genes. The purple edges indicate co-expression, whereas the blue edges signify co-localization; B: A physically and functionally connected Protein-Protein Interaction network implemented common goals through Search Tool for the Retrieval of Interacting Genes/Proteins, where nodes represented proteins and edges represented pairs of interactions between proteins. Node size and color indicated richness, while edge size and color reflected combined scores.

In this study, 8 modules were obtained through WGCNA. As many prior studies have shown a strong correlation between age and the prognosis of BL patients[20,21], we chose the yellow module that had the strongest correlation with age for further analysis. Ten hub genes (*SRC*, *TLR4*, *CD40*, *STAT3*, *SELL*, *CXCL10*, *IL2RA*, *IL10RA*, *CCR7* and *FCGR2B*) were identified using cytoHubba. GO and KEGG functional analyses were conducted on hub genes, and the PPI and GI analysis of these hub genes revealed their related biological functions. Based on survival analysis, CXCL10 and IL2RA have been identified as genes that affect survival. Afterward, we used a nomogram to develop a new risk assessment system for BL patients based on the aforementioned genes and age, aiming to aid in identifying high-risk groups for this disease.

CXCL10 is one of the three ligands for *CXCR3*, which is a chemokine receptor[31]. Various studies have demonstrated that in addition to attracting CD8+ and CD4+ effector T cells to tumor sites and sites of inflammation, *CXCL10* also governs the polarization and enhances the biological functionality of these cells. This makes *CXCL10* a key chemokine driver and a valid target for the therapy of autoimmune diseases such as Inflammatory Bowel Disease, Multiple Sclerosis, Rheumatoid arthritis, and others. Previous studies have also found that chemokines and their receptors are involved in supporting tumor development and metastatic spread[32-35]. In addition to inducing effector TH1 cells, *CXCL10* has recently been proven to be associated with the recruitment of CXCR3+ CD8+ T cells to the tumor site and the induction of Granzyme B production by these cells, thereby enhancing their anti-tumor activities[36]. Barreira da Silva *et al*[37] used Dipeptidyl peptidase 4 inhibitors to increase the endogenous level of *CXCL10*, thereby suppressing experimental melanoma. It has also been demonstrated that the combination of *CXCL10* gene therapy and radiotherapy improves therapeutic efficacy in cervical cancer using a HeLa cell murine xenograft tumor model[38]. Numerous studies have demonstrated a positive correlation between increased expression of *CXCL10* at the tumor site and improved prognosis in various human cancers[39-41]. However, the biological functions of *CXCL10* in BL have not been addressed so far. Our study initially discovered that the high expression of *CXCL10* appeared to be associated with a better prognosis. In our prognostic model, *CXCL10* outperforms age, which is an accepted prognostic factor for BL. Further studies are required to investigate and validate the mechanism of *CXCL10* in BL.

A

B
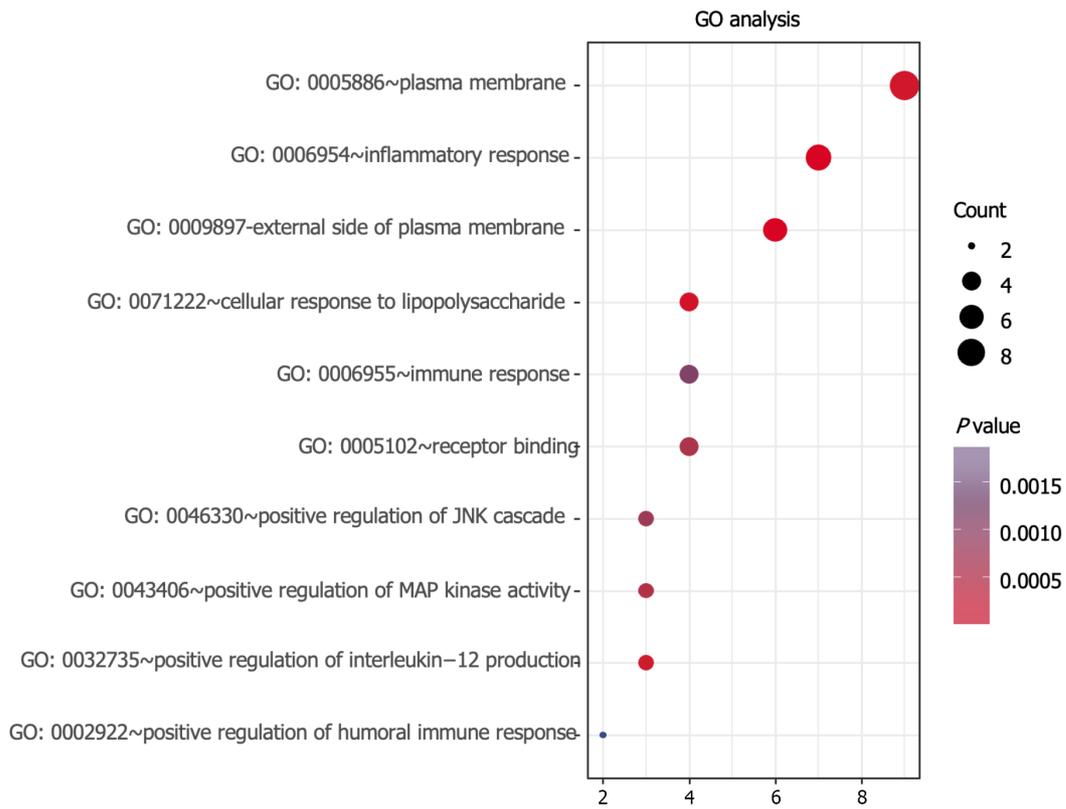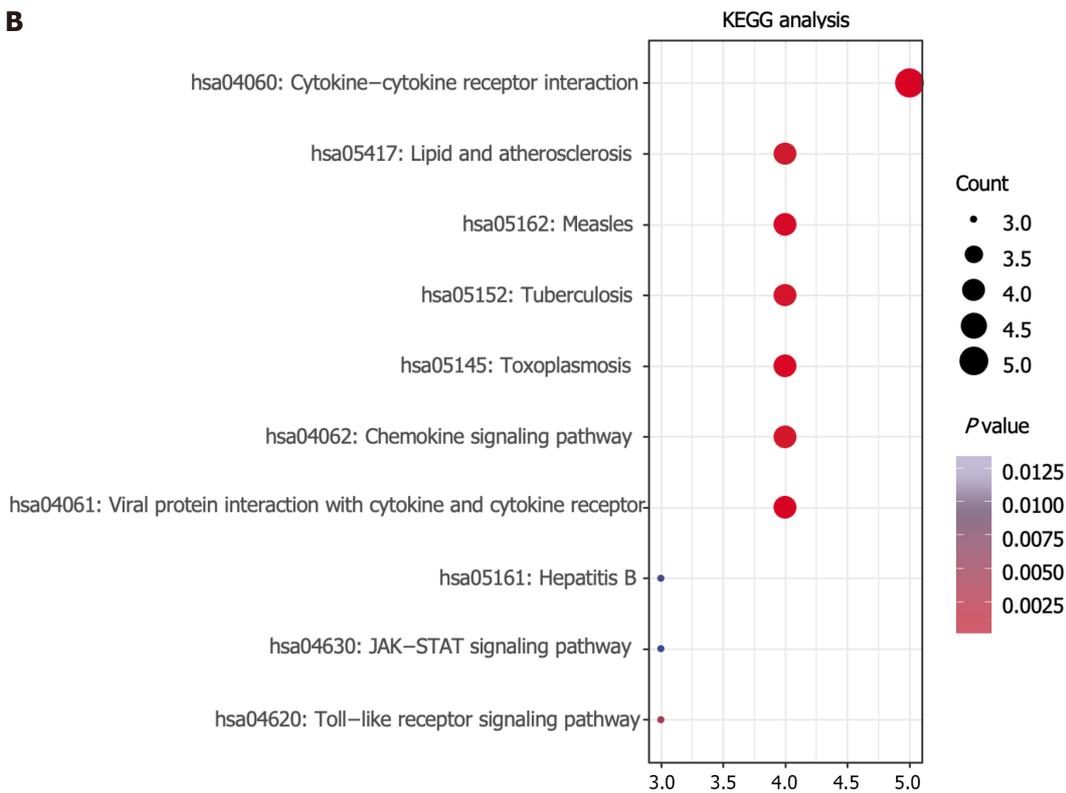
Figure 6 Functional enrichment analysis results of hub genes. A: The top 10 gene ontology terms of hub genes; B: The top 10 Kyoto Encyclopedia of Genes and Genomes pathways of hub genes. GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes.
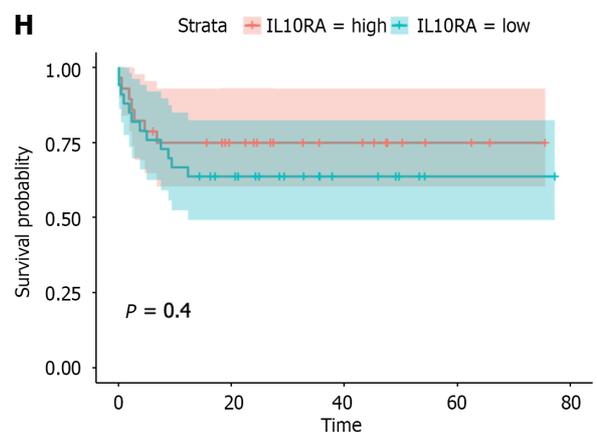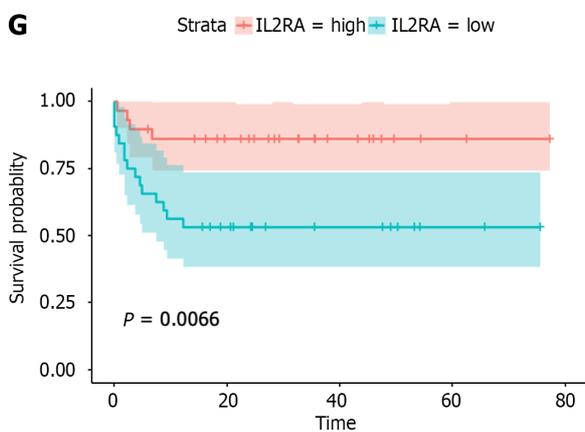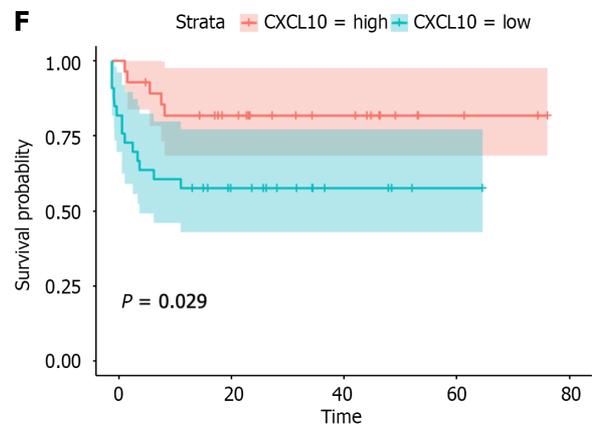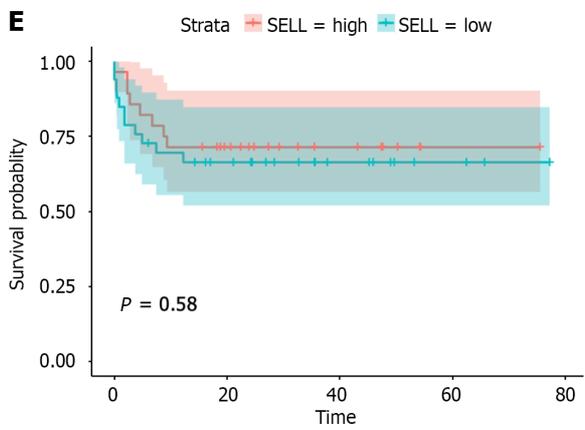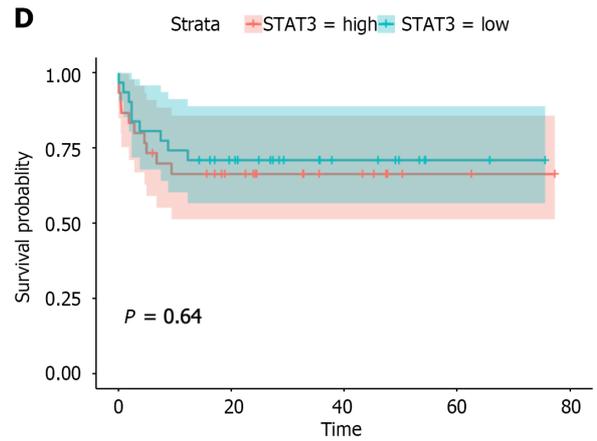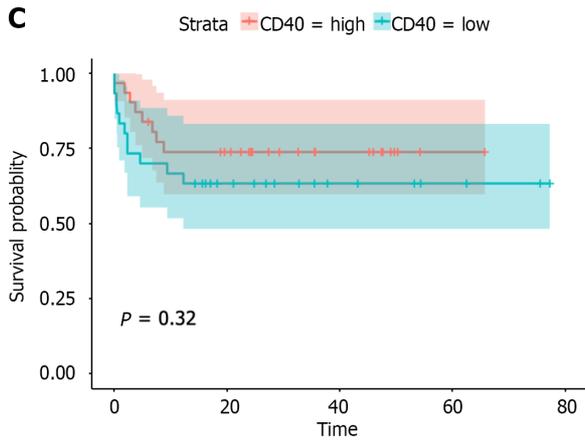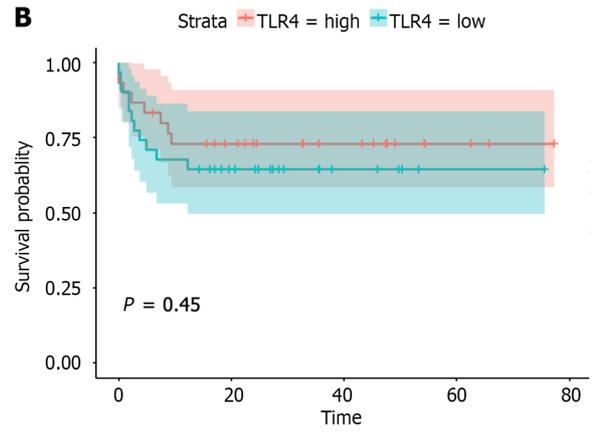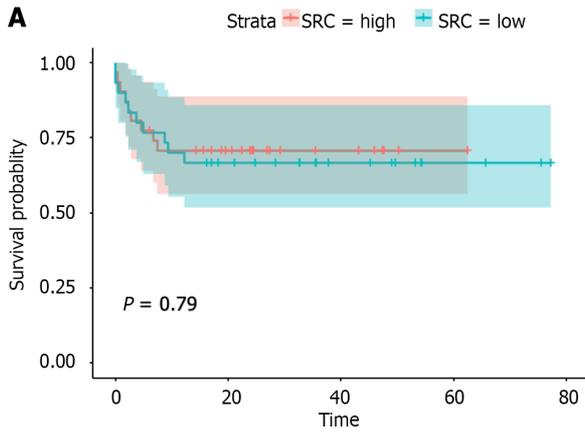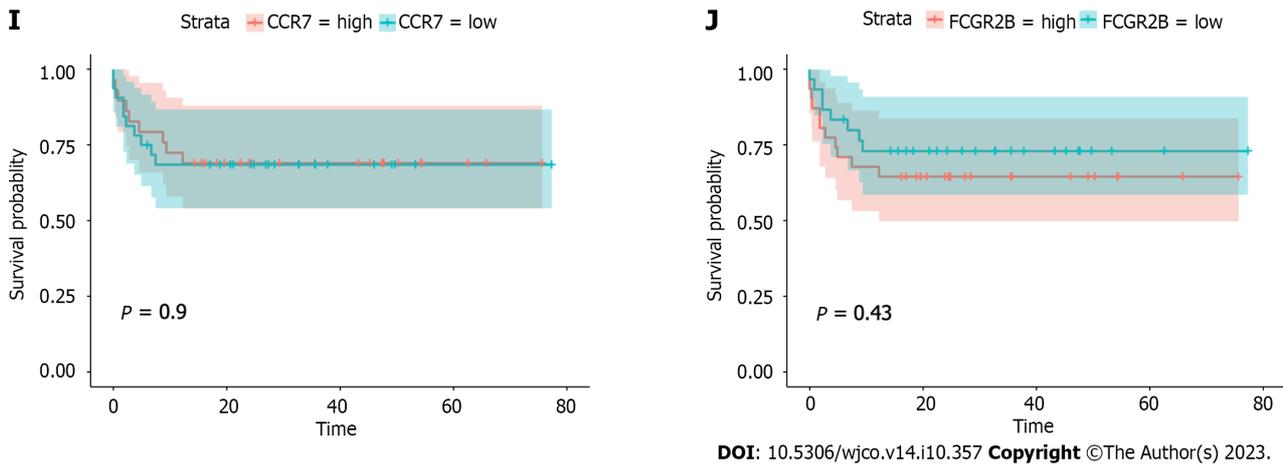
Figure 7 Kaplan–Meier survival curve. A to J: Kaplan–Meier survival curve of identified hub genes in GSE69051.

IL2RA (CD25) is a low-affinity receptor for its ligand interleukin 2 (IL2), but when combined with *IL2RB* (CD122) and *IL2RG* (CD133), it forms the high-affinity IL2 receptor[42]. The binding of IL2 to IL2 receptor activates JAK1 and JAK3, which in turn activate several pathways that regulate cell survival and proliferation, including the PI3K/AKT, RAS/RAF/MEK/ERK, and STAT5 pathways[43]. *IL2RA* expression is elevated in a variety of cancers, especially hematologic tumors[44-46]. Fell *et al*[47] studied 69 patients with leukemia, lymphoma, or multiple myeloma and found that the expression of *IL2RA* in T cells was associated with frailty independent of age. This means that patients with high *IL2RA* expression showed better tolerance to chemotherapy and thus might have a superior prognosis. However, another study demonstrated that high *IL2RA* mRNA expression was an independent and adverse prognostic factor in acute myeloid leukemia, specifically stratifying patients into a worse prognosis[48], while reports on *IL2RA* in chronic myelogenous leukemia (CML) were controversial, they described it as either a promoter or an inhibitor of CML cell proliferation and disease aggressiveness[45,46]. This study demonstrated that BL patients with high expression of *IL2RA* exhibited a better prognosis. Due to the controversial reports on the function of *IL2RA* and the lack of research on BL, further studies are required to validate the prognostic value of *IL2RA* in BL.

As a predictive statistical tool, a nomogram visually displays the significant factors that impact outcomes in multifactor regression analyses and simplifies survival probability prediction through an easy-to-understand graphical representation[49]. The construction of the nomogram model in this study is based on age, *IL2RA*, and *CXCL10*. The nomogram effectively visualizes the impact of identified hub genes and facilitates survival prediction, with the multivariate regression analysis serving as the fundamental component of this model. However, the nomogram would benefit from a validation cohort to enhance its current model. Therefore, it is recommended that additional patients with long-term follow-up be included in future studies.

Based on *CXCL10* and *IL2RA*, we have also identified some drugs that may potentially play a therapeutic role in relapsed and refractory BL, which require further research on pharmacology and treatment protocols. There are also some limitations of the present study. Firstly, the sample size may not be sufficient and could result in selection bias. Secondly, the three different clinical types of BL have the same histological features and similar clinical behavior but differ in epidemiology, clinical presentation, and genetic characteristics, which might need to be classified and analyzed separately. What's more, additional genetic and experimental studies are required to explain the mechanism and the function of these hub genes in the carcinogenesis and progression of BLs. Due to the limited experimental conditions, our study exclusively utilized data sourced from publicly available databases. However, further validation is needed in larger samples or more external datasets.

CONCLUSION

In conclusion, this study is the first to investigate gene expression in BL using WGCNA. These findings provide a framework for identifying co-expression gene modules and discriminating key pathways and hub genes in BL. In the present study, we identified and verified 10 hub genes. Survival analysis showed that overexpression of *CXCL10* and *IL2RA* in BL may serve as superior prognostic indicators. Additionally, an integrated mRNA signature and age nomogram potentially offer prognostic value for patients with BLs.
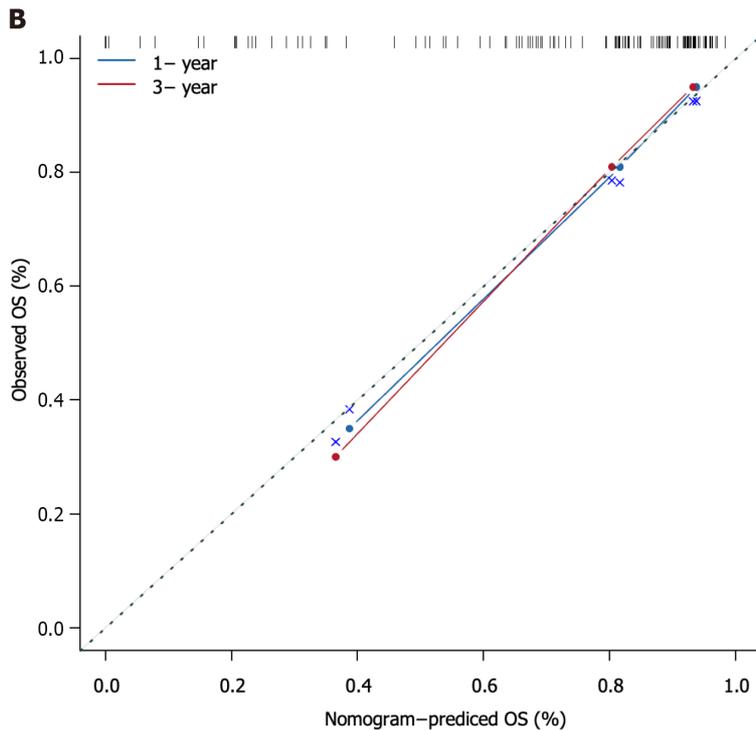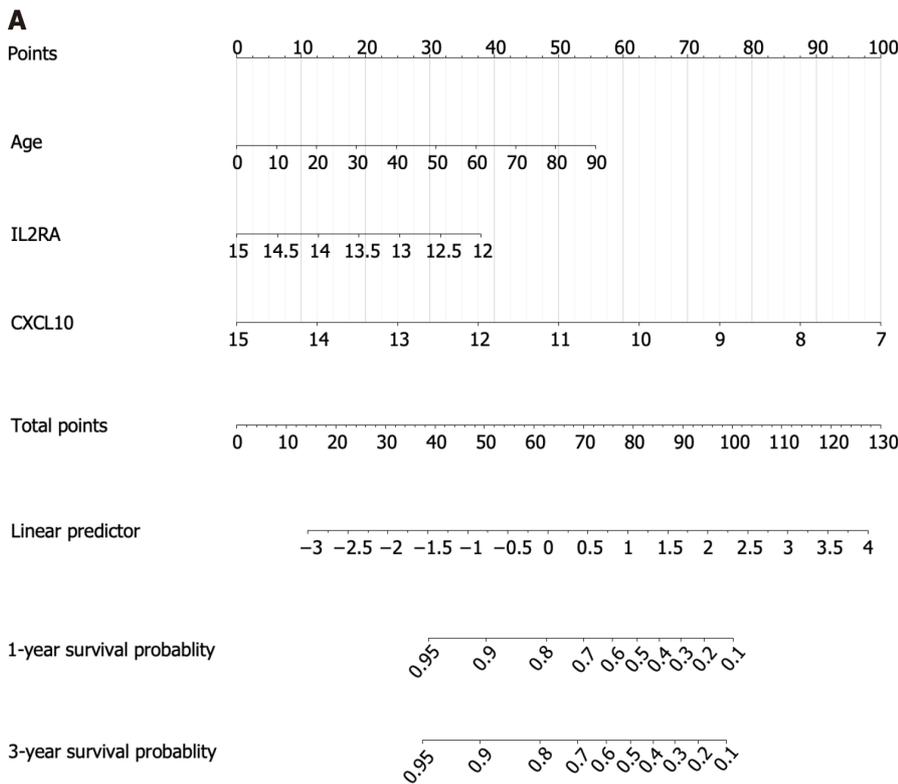
Figure 8 Nomogram and calibration plot for GSE69051 cohort. A: The nomogram was constructed to predicting1, 3-year survival rate of Burkitt lymphoma patients; B: The calibration curves for predicting patient survival at 1 and 3 years in the cohort. OS: Overall survival.
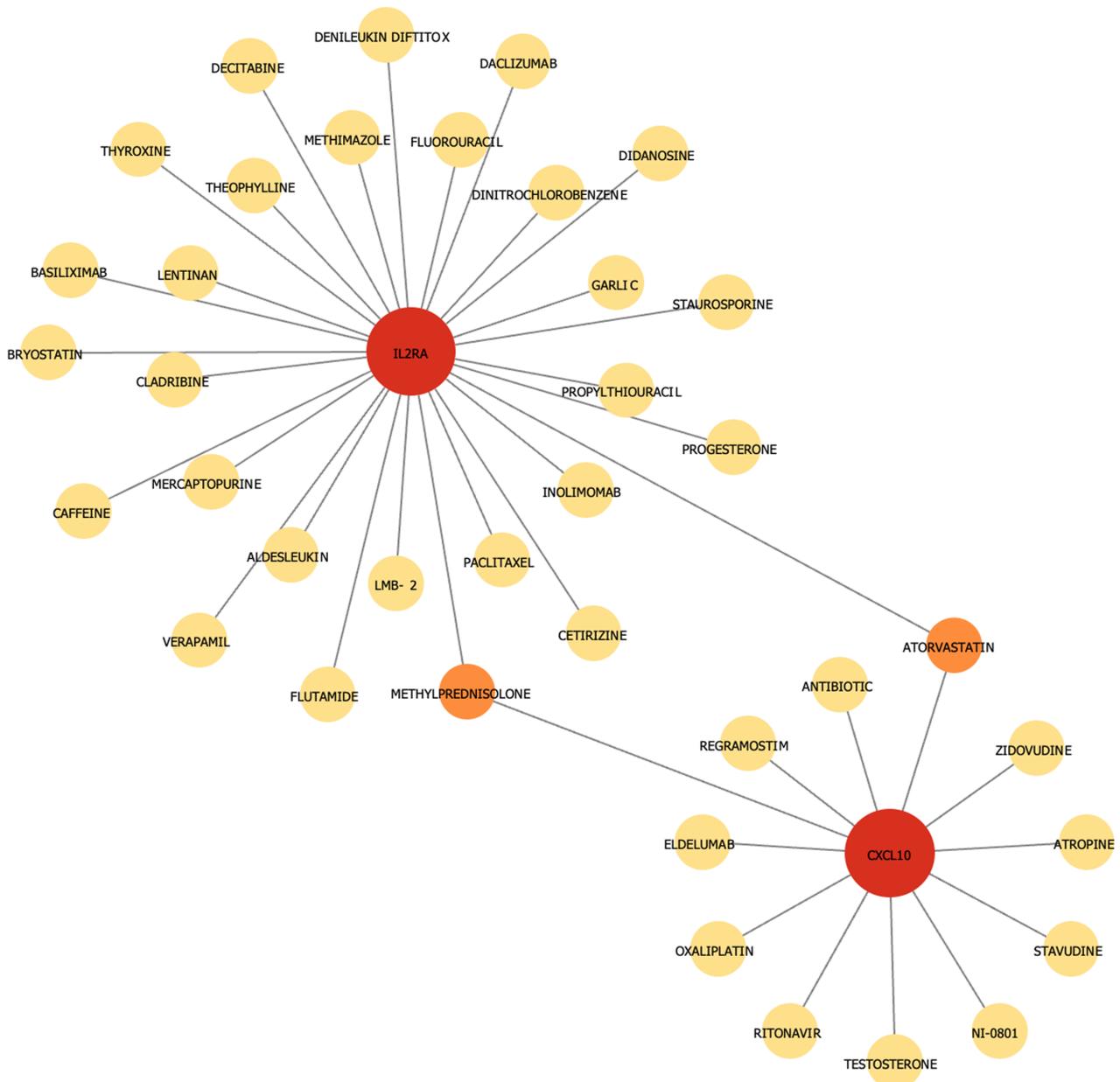
Figure 9 Drugs related to *IL2RA* and *CXCL10*.

ARTICLE HIGHLIGHTS

Research background

Burkitt lymphoma (BL) is an exceptionally aggressive malignant neoplasm originating from either the germinal center or post-germinal center B cells. However, a standardized treatment regimen for BL has yet to be established. The utilization of microarray data and sequencing information retrieved from public databases presents promising prospects for the identification of novel diagnostic or therapeutic targets.

Research motivation

It is crucial to identify biomarkers that can predict the prognosis of BLs and distinguish patients who would benefit from specific therapies.

Research objectives

The aim of our study was to identify hub genes and conduct gene ontology analysis specifically in BL, as well as perform functional enrichment analysis. Additionally, we performed survival analysis and developed a novel prognostic model incorporating candidate genes along with clinical features.

Research methods

The gene expression profiles and clinical traits of BL patients were obtained from the Gene Expression Omnibus database. Weighted gene co-expression network analysis (WGCNA) was employed to construct gene co-expression modules, while the cytoHubba tool was utilized to identify hub genes. Prognostic candidate genes were identified through overall survival (OS) analysis. A nomogram was developed to evaluate the predictive value of the hub genes.

Research results

In this study, we identified 8 modules through WGCNA analysis and found a significant correlation between the yellow module and age. By using the cytoHubba tool, we identified 10 hub genes (*SRC*, *TLR4*, *CD40*, *STAT3*, *SELL*, *CXCL10*, *IL2RA*, *IL10RA*, *CCR7*, and *FCGR2B*). Among these hubs, two genes (*CXCL10* with $P = 0.029$ and *IL2RA* with $P = 0.0066$) were associated with OS based on our survival analysis.

Research conclusions

This study is the first to investigate gene expression in BL using WGCNA. We have identified and validated 10 hub genes, demonstrating that the overexpression of *CXCL10* and *IL2RA* in BL can serve as robust prognostic indicators. Furthermore, the integration of an mRNA signature with age nomogram holds promising potential for predicting patient outcomes in BLs.

Research perspectives

Further genetic and experimental investigations are imperative to elucidate the underlying mechanism and functional significance of these hub genes in the carcinogenesis and progression of BLs.

FOOTNOTES

Institutional review board statement: Institutional review board statement is not applied to our manuscript.

Institutional animal care and use committee statement: Institutional animal care and use committee statement is not applied to our manuscript.

Data sharing statement: The datasets analyzed (GSE4475 and GSE69051) during this study are publicly available in the GEO database (https://www.ncbi.nlm.nih.gov/geo/), the original contributions presented in this study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author (yangjigang@ccmu.edu.cn).

ARRIVE guidelines statement: The authors have read the ARRIVE Guidelines, and the manuscript was prepared and revised according to the ARRIVE Guidelines.

Country/Territory of origin: China

ORCID number: Guan-Yun Wang 0000-0003-1880-7163; Ji-Gang Yang 0000-0002-6489-7780.

REFERENCES

1   **Bishop PC**, Rao VK, Wilson WH. Burkitt's lymphoma: molecular pathogenesis and treatment. *Cancer Invest* 2000; **18**: 574-583 [PMID: 10923106 DOI: 10.3109/07357900009012197]
2   **Sall FB**, Shmakova A, Karpukhina A, Tsfasman T, Lomov N, Canoy RJ, Boutboul D, Oksenhendler E, Toure AO, Lipinski M, Wiels J, Germini D, Vassetzky Y. Epstein-Barr virus reactivation induces MYC-IGH spatial proximity and t(8;14) in B cells. *J Med Virol* 2023; **95**: e28633 [PMID: 36866703 DOI: 10.1002/jmv.28633]

3   **Magrath I**. Epidemiology: clues to the pathogenesis of Burkitt lymphoma. *Br J Haematol* 2012; **156**: 744-756 [PMID: 22260300 DOI: 10.1111/j.1365-2141.2011.09013.x]

4   **Ogwang MD**, Bhatia K, Biggar RJ, Mbulaiteye SM. Incidence and geographic distribution of endemic Burkitt lymphoma in northern Uganda revisited. *Int J Cancer* 2008; **123**: 2658-2663 [PMID: 18767045 DOI: 10.1002/ijc.23800]

5   **Sant M**, Allemani C, Tereanu C, De Angelis R, Capocaccia R, Visser O, Marcos-Gragera R, Maynadié M, Simonetti A, Lutz JM, Berrino F; HAEMACARE Working Group. Incidence of hematologic malignancies in Europe by morphologic subtype: results of the HAEMACARE project. *Blood* 2010; **116**: 3724-3734 [PMID: 20664057 DOI: 10.1182/blood-2010-05-282632]

6   **Guech-Ongey M**, Simard EP, Anderson WF, Engels EA, Bhatia K, Devesa SS, Mbulaiteye SM. AIDS-related Burkitt lymphoma in the United States: what do age and CD4 Lymphocyte patterns tell us about etiology and/or biology? *Blood* 2010; **116**: 5600-5604 [PMID: 20813897 DOI: 10.1182/blood-2010-03-275917]

7   **Smeland S**, Blystad AK, Kvaløy SO, Ikonomou IM, Delabie J, Kvalheim G, Hammerstrøm J, Lauritzsen GF, Holte H. Treatment of Burkitt's/Burkitt-like lymphoma in adolescents and adults: a 20-year experience from the Norwegian Radium Hospital with the use of three successive regimens. *Ann Oncol* 2004; **15**: 1072-1078 [PMID: 15205201 DOI: 10.1093/annonc/mdh262]

8   **Castillo JJ**, Winer ES, Olszewski AJ. Population-based prognostic factors for survival in patients with Burkitt lymphoma: an analysis from the Surveillance, Epidemiology, and End Results database. *Cancer* 2013; **119**: 3672-3679 [PMID: 23913575 DOI: 10.1002/cncr.28264]

9   **Saleh K**, Michot JM, Camara-Clayette V, Vassetzky Y, Ribrag V. Burkitt and Burkitt-Like Lymphomas: a Systematic Review. *Curr Oncol Rep* 2020; **22**: 33 [PMID: 32144513 DOI: 10.1007/s11912-020-0898-8]

10   **Hecht JL**, Aster JC. Molecular biology of Burkitt's lymphoma. *J Clin Oncol* 2000; **18**: 3707-3721 [PMID: 11054444 DOI: 10.1200/JCO.2000.18.21.3707]

11   **Edgar R**, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; **30**: 207-210 [PMID: 11752295 DOI: 10.1093/nar/30.1.207]

12   **Botía JA**, Vandrovcova J, Forabosco P, Guelfi S, D'Sa K; United Kingdom Brain Expression Consortium, Hardy J, Lewis CM, Ryten M, Weale ME. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst Biol* 2017; **11**: 47 [PMID: 28403906 DOI: 10.1186/s12918-017-0420-6]

13   **Langfelder P**, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**: 559 [PMID: 19114008 DOI: 10.1186/1471-2105-9-559]

14   **Barrett T**, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 2013; **41**: D991-D995 [PMID: 23193258 DOI: 10.1093/nar/gks1193]

15   **Niemira M**, Collin F, Szalkowska A, Bielska A, Chwialkowska K, Reszec J, Niklinski J, Kwasniewski M, Kretowski A. Molecular Signature of Subtypes of Non-Small-Cell Lung Cancer by Large-Scale Transcriptional Profiling: Identification of Key Modules and Genes by Weighted Gene Co-Expression Network Analysis (WGCNA). *Cancers (Basel)* 2019; **12** [PMID: 31877723 DOI: 10.3390/cancers12010037]

16   **Szklarczyk D**, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res* 2021; **49**: D605-D612 [PMID: 33237311 DOI: 10.1093/nar/gkaa1074]

17   **Huang da W**, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; **37**: 1-13 [PMID: 19033363 DOI: 10.1093/nar/gkn923]

18   **Warde-Farley D**, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010; **38**: W214-W220 [PMID: 20576703 DOI: 10.1093/nar/gkq537]

19   **Wagner AH**, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, Krysiak K, Pan D, McMichael JF, Eldred JM, Walker JR, Wilson RK, Mardis ER, Griffith M, Griffith OL. DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res* 2016; **44**: D1036-D1044 [PMID: 26531824 DOI: 10.1093/nar/gkv1165]

20   **Olszewski AJ**, Jakobsen LH, Collins GP, Cwynarski K, Bachanova V, Blum KA, Boughan KM, Bower M, Dalla Pria A, Danilov A, David KA, Diefenbach C, Ellin F, Epperla N, Farooq U, Feldman TA, Gerrie AS, Jagadeesh D, Kamdar M, Karmali R, Kassam S, Kenkre VP, Khan N, Kim SH, Klein AK, Lossos IS, Lunning MA, Martin P, Martinez-Calle N, Montoto S, Naik S, Palmisiano N, Peace D, Phillips EH, Phillips TJ, Portell CA, Reddy N, Santarsieri A, Sarraf Yazdy M, Smeland KB, Smith SE, Smith SD, Sundaram S, Zayac AS, Zhang XY, Zhu C, Cheah CY, El-Galaly TC, Evens AM. Burkitt Lymphoma International Prognostic Index. *J Clin Oncol* 2021; **39**: 1129-1138 [PMID: 33502927 DOI: 10.1200/JCO.20.03288]

21   **Lu J**, Tan H, Li B, Chen S, Xu L, Zou Y. Status and prognostic nomogram of patients with Burkitt lymphoma. *Oncol Lett* 2020; **19**: 972-984 [PMID: 31897210 DOI: 10.3892/ol.2019.11155]

22   **Ferry JA**. Burkitt's lymphoma: clinicopathologic features and differential diagnosis. *Oncologist* 2006; **11**: 375-383 [PMID: 16614233 DOI: 10.1634/theoncologist.11-4-375]

23   **Burmeister T**, Schwartz S, Horst HA, Rieder H, Gökbuget N, Hoelzer D, Thiel E. Molecular heterogeneity of sporadic adult Burkitt-type leukemia/Lymphoma as revealed by PCR and cytogenetics: correlation with morphology, immunology and clinical features. *Leukemia* 2005; **19**: 1391-1398 [PMID: 15973450 DOI: 10.1038/sj.leu.2403847]

24   **Wästerlid T**, Jonsson B, Hagberg H, Jerkeman M. Population based study of prognostic factors and treatment in adult Burkitt lymphoma: a Swedish Lymphoma Registry study. *Leuk Lymphoma* 2011; **52**: 2090-2096 [PMID: 21718134 DOI: 10.3109/10428194.2011.593274]

25   **Schmitz R**, Ceribelli M, Pittaluga S, Wright G, Staudt LM. Oncogenic mechanisms in Burkitt lymphoma. *Cold Spring Harb Perspect Med* 2014; **4** [PMID: 24492847 DOI: 10.1101/cshperspect.a014282]

26   **Fuller TF**, Ghazalpour A, Aten JE, Drake TA, Lusis AJ, Horvath S. Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm Genome* 2007; **18**: 463-472 [PMID: 17668265 DOI: 10.1007/s00335-007-9043-3]

27   **Mason MJ**, Fan G, Plath K, Zhou Q, Horvath S. Signed weighted gene co-expression network analysis of transcriptional regulation in murine embryonic stem cells. *BMC Genomics* 2009; **10**: 327 [PMID: 19619308 DOI: 10.1186/1471-2164-10-327]

28   **Saris CG**, Horvath S, van Vught PW, van Es MA, Blauw HM, Fuller TF, Langfelder P, DeYoung J, Wokke JH, Veldink JH, van den Berg LH, Ophoff RA. Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics* 2009; **10**: 405 [PMID: 19712483 DOI: 10.1186/1471-2164-10-405]

29   **Chou WC**, Cheng AL, Brotto M, Chuang CY. Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer. *BMC Genomics* 2014; **15**: 300 [PMID: 24758163 DOI: 10.1186/1471-2164-15-300]

30   **Liu R**, Cheng Y, Yu J, Lv QL, Zhou HH. Identification and validation of gene module associated with lung cancer through coexpression network analysis. *Gene* 2015; **563**: 56-62 [PMID: 25752287 DOI: 10.1016/j.gene.2015.03.008]

31   **Karin N**, Razon H. Chemokines beyond chemo-attraction: CXCL10 and its significant role in cancer and autoimmunity. *Cytokine* 2018; **109**: 24-28 [PMID: 29449068 DOI: 10.1016/j.cyto.2018.02.012]

32   **Nagarsheth N**, Wicha MS, Zou W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat Rev Immunol* 2017; **17**: 559-572 [PMID: 28555670 DOI: 10.1038/nri.2017.49]

33   **Reymond N**, d'Água BB, Ridley AJ. Crossing the endothelial barrier during metastasis. *Nat Rev Cancer* 2013; **13**: 858-870 [PMID: 24263189 DOI: 10.1038/nrc3628]

34   **Nibbs RJ**, Graham GJ. Immune regulation by atypical chemokine receptors. *Nat Rev Immunol* 2013; **13**: 815-829 [PMID: 24319779 DOI: 10.1038/nri3544]

35   **Murdoch C**, Muthana M, Coffelt SB, Lewis CE. The role of myeloid cells in the promotion of tumour angiogenesis. *Nat Rev Cancer* 2008; **8**: 618-631 [PMID: 18633355 DOI: 10.1038/nrc2444]

36   **Zumwalt TJ**, Arnold M, Goel A, Boland CR. Active secretion of CXCL10 and CCL5 from colorectal cancer microenvironments associates with GranzymeB+ CD8+ T-cell infiltration. *Oncotarget* 2015; **6**: 2981-2991 [PMID: 25671296 DOI: 10.18632/oncotarget.3205]

37   **Barreira da Silva R**, Laird ME, Yatim N, Fiette L, Ingersoll MA, Albert ML. Dipeptidylpeptidase 4 inhibition enhances lymphocyte trafficking, improving both naturally occurring tumor immunity and immunotherapy. *Nat Immunol* 2015; **16**: 850-858 [PMID: 26075911 DOI: 10.1038/ni.3201]

38   **Zhao M**, Ma Q, Xu J, Fu S, Chen L, Wang B, Wu J, Yang L. Combining CXCL10 gene therapy and radiotherapy improved therapeutic efficacy in cervical cancer HeLa cell xenograft tumor models. *Oncol Lett* 2015; **10**: 768-772 [PMID: 26622567 DOI: 10.3892/ol.2015.3281]

39   **Bronger H**, Singer J, Windmüller C, Reuning U, Zech D, Delbridge C, Dorn J, Kiechle M, Schmalfeldt B, Schmitt M, Avril S. CXCL9 and CXCL10 predict survival and are regulated by cyclooxygenase inhibition in advanced serous ovarian cancer. *Br J Cancer* 2016; **115**: 553-563 [PMID: 27490802 DOI: 10.1038/bjc.2016.172]

40   **Toiyama Y**, Fujikawa H, Kawamura M, Matsushita K, Saigusa S, Tanaka K, Inoue Y, Uchida K, Mohri Y, Kusunoki M. Evaluation of CXCL10 as a novel serum marker for predicting liver metastasis and prognosis in colorectal cancer. *Int J Oncol* 2012; **40**: 560-566 [PMID: 22038159 DOI: 10.3892/ijo.2011.1247]

41   **Liu M**, Guo S, Stiles JK. The emerging role of CXCL10 in cancer (Review). *Oncol Lett* 2011; **2**: 583-589 [PMID: 22848232 DOI: 10.3892/ol.2011.300]

42   **Flynn MJ**, Hartley JA. The emerging role of anti-CD25 directed therapies as both immune modulators and targeted agents in cancer. *Br J Haematol* 2017; **179**: 20-35 [PMID: 28556984 DOI: 10.1111/bjh.14770]

43   **Olejniczak K**, Kasprzak A. Biological properties of interleukin 2 and its role in pathogenesis of selected diseases--a review. *Med Sci Monit* 2008; **14**: RA179-RA189 [PMID: 18830208]

44   **Saito Y**, Kitamura H, Hijikata A, Tomizawa-Murasawa M, Tanaka S, Takagi S, Uchida N, Suzuki N, Sone A, Najima Y, Ozawa H, Wake A, Taniguchi S, Shultz LD, Ohara O, Ishikawa F. Identification of therapeutic targets for quiescent, chemotherapy-resistant human leukemia stem cells. *Sci Transl Med* 2010; **2**: 17ra9 [PMID: 20371479 DOI: 10.1126/scitranslmed.3000349]

45   **Sadovnik I**, Hoelbl-Kovacic A, Herrmann H, Eisenwort G, Cerny-Reiterer S, Warsch W, Hoermann G, Greiner G, Blatt K, Peter B, Stefanzl G, Berger D, Bilban M, Herndlhofer S, Sill H, Sperr WR, Streubel B, Mannhalter C, Holyoake TL, Sexl V, Valent P. Identification of CD25 as STAT5-Dependent Growth Regulator of Leukemic Stem Cells in Ph+ CML. *Clin Cancer Res* 2016; **22**: 2051-2061 [PMID: 26607600 DOI: 10.1158/1078-0432.CCR-15-0767]

46   **Kobayashi CI**, Takubo K, Kobayashi H, Nakamura-Ishizu A, Honda H, Kataoka K, Kumano K, Akiyama H, Sudo T, Kurokawa M, Suda T. The IL-2/CD25 axis maintains distinct subsets of chronic myeloid leukemia-initiating cells. *Blood* 2014; **123**: 2540-2549 [PMID: 24574458 DOI: 10.1182/blood-2013-07-517847]

47   **Fell G**, Rosko AE, Abel GA, Dumontier C, Higby KJ, Murillo A, Neuberg DS, Burd CE, Lane AA. Peripheral blood CD3(+) T-cell gene expression biomarkers correlate with clinical frailty in patients with haematological malignancies. *Br J Haematol* 2022; **199**: 100-105 [PMID: 35766906 DOI: 10.1111/bjh.18336]

48   **Du W**, He J, Zhou W, Shu S, Li J, Liu W, Deng Y, Lu C, Lin S, Ma Y, He Y, Zheng J, Zhu J, Bai L, Li X, Yao J, Hu D, Gu S, Li H, Guo A, Huang S, Feng X. High IL2RA mRNA expression is an independent adverse prognostic biomarker in core binding factor and intermediate-risk acute myeloid leukemia. *J Transl Med* 2019; **17**: 191 [PMID: 31171000 DOI: 10.1186/s12967-019-1926-z]

49   **Raghav K**, Hwang H, Jácome AA, Bhang E, Willett A, Huey RW, Dhillon NP, Modha J, Smaglo B, Matamoros A Jr, Estrella JS, Jao J, Overman MJ, Wang X, Greco FA, Loree JM, Varadhachary GR. Development and Validation of a Novel Nomogram for Individualized Prediction of Survival in Cancer of Unknown Primary. *Clin Cancer Res* 2021; **27**: 3414-3421 [PMID: 33858857 DOI: 10.1158/1078-0432.CCR-20-4117]