



OPEN

# Bioinformatics analysis of the pathogenic link between Epstein-Barr virus infection, systemic lupus erythematosus and diffuse large B cell lymphoma

Qian-Ying Zhu

Epstein-Barr virus (EBV) is a risk factor for diffuse large B-cell lymphoma (DLBCL) and systemic lupus erythematosus (SLE). While prior research has suggested a potential correlation between SLE and DLBCL, the molecular mechanisms remain unclear. The present study aimed to explore the contribution of EBV infection to the pathogenesis of DLBCL in the individuals with SLE using bioinformatics approaches. The Gene Expression Omnibus database was used to compile the gene expression profiles of EBV-infected B cells (GSE49628), SLE (GSE61635), and DLBCL (GSE32018). Altogether, 72 shared common differentially expressed genes (DEGs) were extracted and enrichment analysis of the shared genes showed that p53 signaling pathway was a common feature of the pathophysiology. Six hub genes were selected using protein–protein interaction (PPI) network analysis, including CDK1, KIF23, NEK2, TOP2A, NEIL3 and DEPDC1, which showed preferable diagnostic values for SLE and DLBCL and involved in immune cell infiltration and immune responses regulation. Finally, TF-gene and miRNA-gene regulatory networks and 10 potential drugs molecule were predicted. Our study revealed the potential molecular mechanisms by which EBV infection contribute to the susceptibility of DLBCL in SLE patients for the first time and identified future biomarkers and therapeutic targets for SLE and DLBCL.

Epstein-Barr virus (EBV) is considered to be one of eight human herpesviruses that contain protein capsids around the double-stranded linear DNA genome<sup>1</sup>. EBV infects over 90% of people worldwide. The oral cavity is assumed to be the site of primary EBV infection<sup>2</sup>. Lymphocytes and epithelial cells make up the majority of host cells for EBV<sup>3</sup>. In adolescents, primary EBV infection often causes infectious mononucleosis (IM). Several autoimmune disorders, including systemic lupus erythematosus (SLE), have been linked to EBV<sup>4</sup>. In addition, as the first tumorigenic virus to be identified, EBV causes approximately 200,000 new instances of cancer each year, comprising cancers originating from B cells like Burkitt lymphoma and diffuse large B-cell lymphoma (DLBCL)<sup>5</sup>.

SLE is an autoimmune disorder that usually occurs among women of childbearing age with multi-systemic involvement<sup>6</sup>. The phenotypes of SLE can vary from minor mucocutaneous signs to serious central nervous or kidney damage, which results in a high risk of death for patients<sup>7</sup>. The characteristic feature of SLE is the occurrence of autoantibodies against nuclear antigens (ANA), which can be detected up to 10 years before disease onset<sup>8</sup>. SLE is considered to be brought on by a confluence of hereditary and environmental factors. Among environmental factors, viral infection, especially EBV infection, is closely related to SLE<sup>9</sup>. Previous studies have shown that, compared to healthy controls, higher viral loads and elevated titers of antibodies against EBV are detected among SLE patients<sup>10</sup>. Moreover, a growing body of research suggests that EBV-infected B cells may become resistant to apoptosis, leading to the proliferation, activation and antibody production of autoreactive B cells, which may result in tissue damage for SLE<sup>11</sup>. However, the underlying genetic molecular mechanism of EBV infection on the development of SLE is still not fully elucidated.

Department of Laboratory Medicine, The Eighth Affiliated Hospital, Sun Yat-sen University, Shenzhen 518003, People's Republic of China. email: zhuqy37@mail.sysu.edu.cn

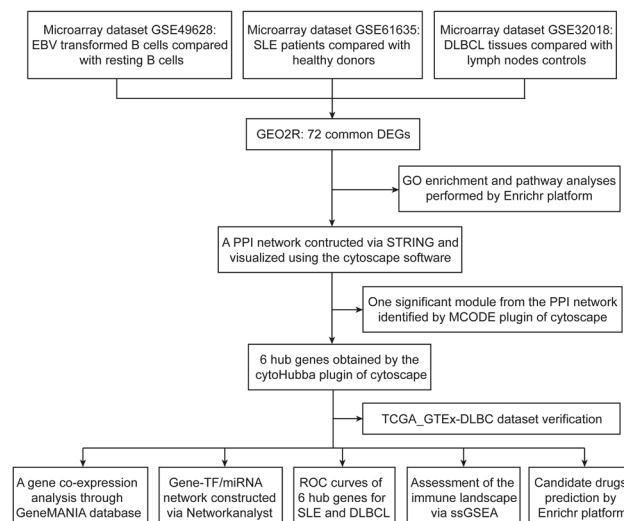
DLBCL refers to diffuse growth tumors with nuclei greater than 2 normal lymphocytes. It is the most prevalent kind of non-Hodgkin's lymphoma, making up roughly 40% of all B-cell lymphomas<sup>12</sup>. The R-CHOP therapy can cure approximately 60% of DLBCL. However, many patients still develop resistant to this treatment or experience a recurrence and eventually die<sup>13</sup>. At present, the etiology of DLBCL is not particularly clear, which is usually associated with gene abnormalities, EBV infection and other reasons<sup>14</sup>. In 2016, the World Health Organization classified EBV-associated DLBCL as a new subtype of DLBCL (EBV + DLBCL, NOS)<sup>15</sup>. Although a number of researches have demonstrated that EBV play extremely crucial roles in the process of inducing malignant transformation of lymphocytes<sup>16,17</sup>, the specific molecular mechanism of EBV regulating B cell signaling pathway still needs to be further studied.

Some studies have found an elevated incidence of malignancies, notably lymphoma, in people with SLE<sup>18</sup>. DLBCL composes 37–62% of all lymphomas identified in SLE<sup>19</sup>, but there is currently a lack of knowledge on the pathophysiology between DLBCL and SLE. Several determinants could be behind this altered risk, such as genetic factors, immunologic derangements and viruses, among which EBV infection is suggested to be the link between SLE and DLBCL<sup>20</sup>. Recent studies have shown that persistent EBV infection in SLE patients can promote malignant transformation of B cells due to the use of immunosuppressive drugs and other reasons<sup>21</sup>. However, molecular mechanisms underlying the contribution of EBV infection to the development of DLBCL in SLE patients are yet unknown.

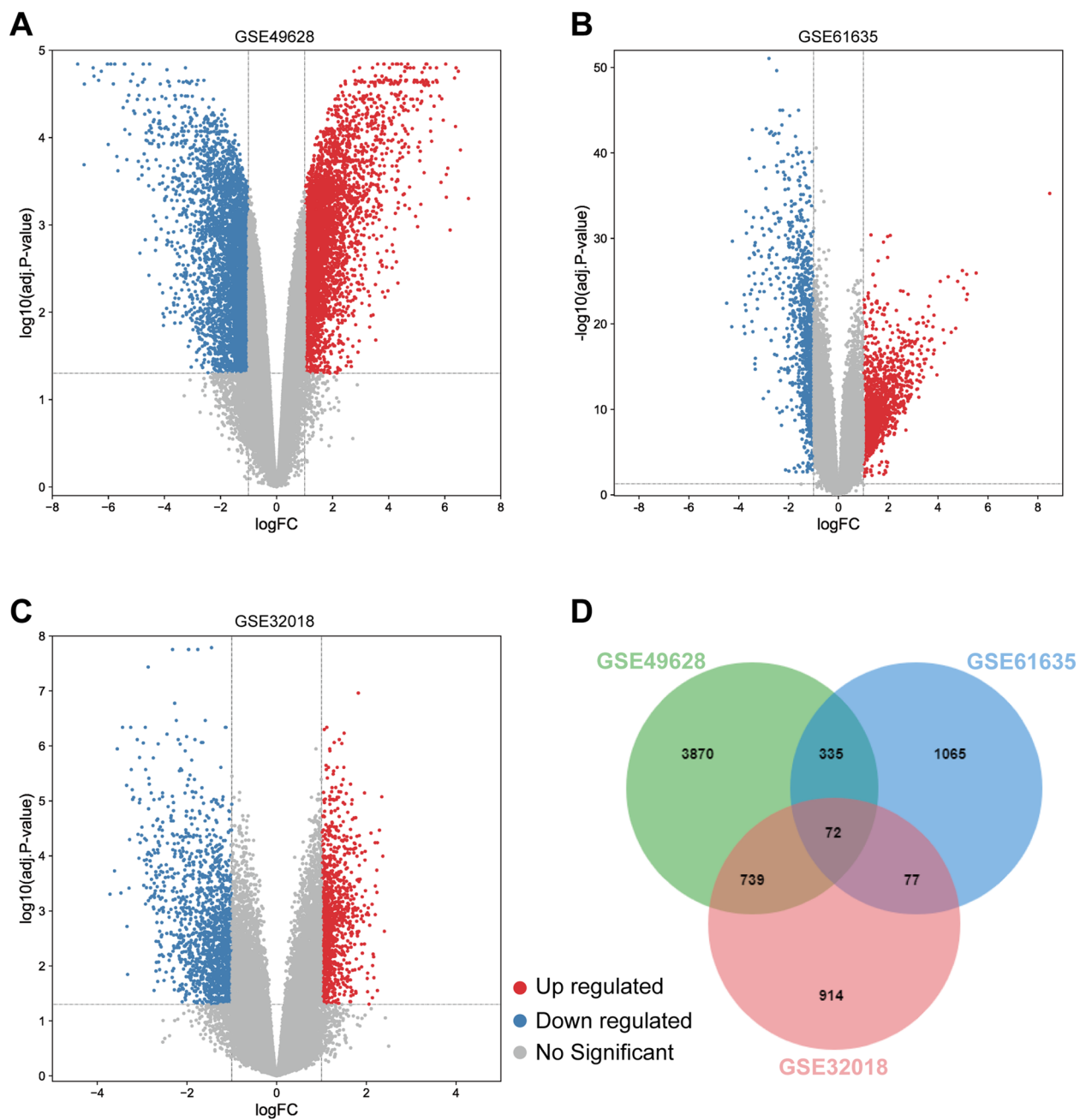
In this study, we attempted to identify the shared gene signatures between EBV infection, SLE and DLBCL and explore the possible biological effect of EBV infection to the pathogenesis of DLBCL in the context of SLE. Firstly, we employed three datasets in this investigation to explore the biological link between EBV infection, SLE, and DLBCL. GSE49628, GSE61635, and GSE32018 were chosen from the Gene Expression Omnibus (GEO) database for EBV infection, SLE, and DLBCL, respectively. We first identified the differentially expressed genes (DEGs) in each dataset before identifying the common DEGs in these three datasets. To comprehend genome-based biological processes, shared DEGs were employed as the key experimental genes throughout the study, including gene ontology (GO) analyses, pathway enrichment analyses, and protein–protein interaction (PPI) network construction. In addition, 6 hub genes were extracted utilizing Cytoscape software for gene regulatory investigation, such as transcription factors (TFs)-gene network and miRNAs-gene network generation and immune landscape assessment. Finally, receiver operating characteristic (ROC) curves and candidate drugs prediction for SLE and DLBCL were performed based on the hub genes. Taken together, our study explored the molecular mechanisms by which EBV infection contribute to the susceptibility of DLBCL in SLE patients for the first time and identified potential biomarkers and therapeutic targets for patients with SLE and DLBCL.

## Results

**Identification of DEGs among EBV infection, SLE and DLBCL.** Figure 1 depicts the whole work flow of this project. Firstly, 5016 genes, comprising 2883 up-regulated and 2133 down-regulated genes from the GSE49628 dataset, were differentially expressed for EBV infection (Fig. 2A). A total of 1549 DEGs were discovered using the SLE dataset (GSE61635), of which 992 genes showed up-regulation and 557 genes showed down-regulation (Fig. 2B). We discovered 1802 DEGs for the DLBCL dataset (GSE32018), comprising 791 up-regulated genes and 1011 down-regulated genes (Fig. 2C). With the use of the cutoff criteria ( $P$ -value < 0.05 and  $|\log_{2}FC| > 1$ ), all significant DEGs were retrieved. The summarized information of these datasets was listed in Table 1. Then by taking the intersection of DEGs of EBV, SLE and DLBCL datasets, 72 common DEGs were identified and visualized by Venn diagrams (Fig. 2D).



**Figure 1.** Flowchart for this investigation.



**Figure 2.** Volcano diagrams and Venn diagram. The volcano maps of GSE49628 (A), GSE61635 (B), and GSE32018 (C). Up-regulated genes are shown in red, while down-regulated genes are shown in blue. (D) There were 72 DEGs shared by these three datasets.

Disease name	GEO accession	GEO platform	Total DEGs count	Up regulated DEGs count	Down regulated DEGs count
EBV	GSE49628	GPL570	5016	2883	2133
SLE	GSE61635	GPL570	1549	992	557
DLBLC	GSE32018	GPL6480	1802	791	1011

**Table 1.** Summary of the datasets used in this investigation, together with their geo-features and quantitative metrics.

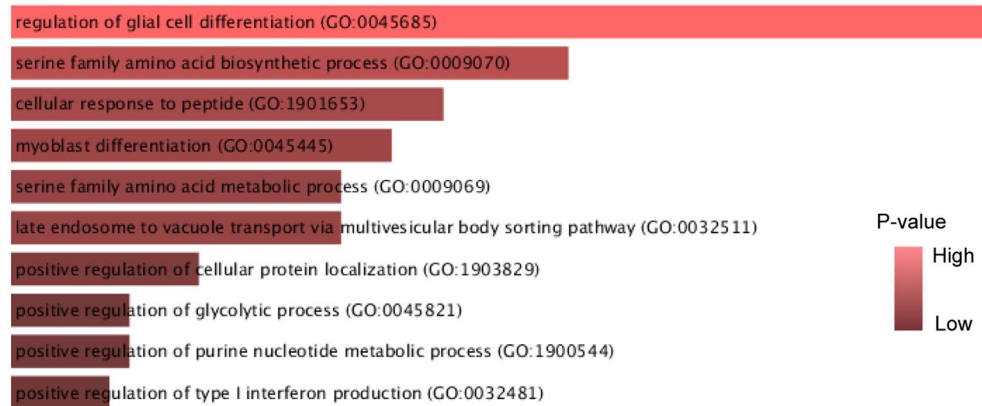
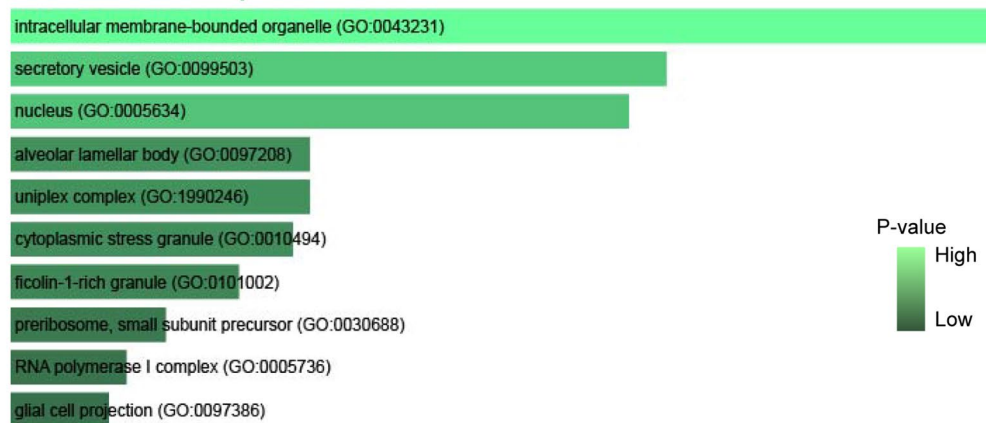
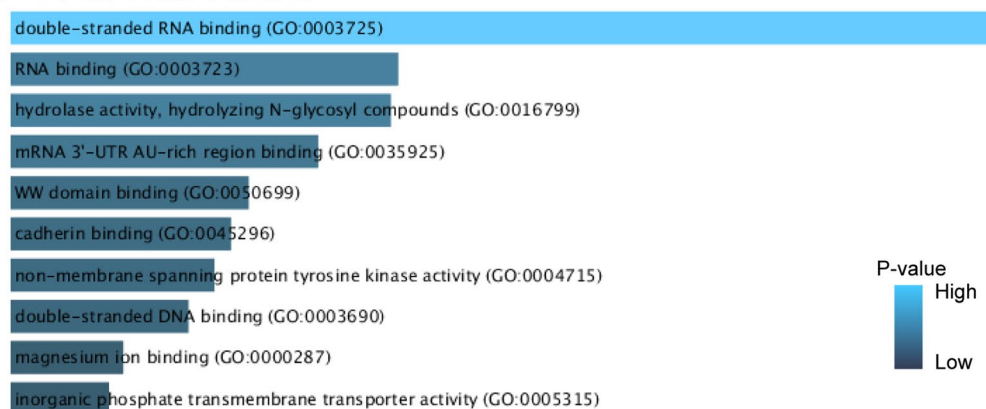
**GO and pathway enrichment analysis.** We used the Enrichr online tool to analyze GO and pathway enrichment of the common DEGs. The rank of significant terms was determined by the *P*-value. GO analysis consists of three categories, including biological process, cellular component and molecular function. The top 10 significant terms of different categories were summarized in Table 2 and listed as bar graphs in Fig. 3. The most impacted pathways of the common DEGs among EBV infection, SLE and DLBCL were identified from four databases (WikiPathways, Reactome, KEGG and BioCarta). The top 10 pathways gathered from the above datasets were listed in Table 3 and also shown precisely in bar graphs in Fig. 4.

**PPI network and submodule analysis.** The PPI network of 72 common DEGs was analyzed by STRING platform, and the result was further visualized with Cytoscape software. The 42 nodes and 91 edges of the PPI network were shown in Fig. 5A. The stronger the connection of the gene with other genes, the redder the node was in the PPI network. Besides, we applied the MCODE plug-in of Cytoscape to construct a key gene module, which contained 8 common DEGs (Fig. 5B).

**Hub gene identification and functional analysis.** Using 7 algorithms of cytoHubba plug-in of Cytoscape, the top 10 hub genes were screened. By applying the intersection of Venn diagrams, we finally identified 6 common hub genes, including CDK1, KIF23, NEK2, TOP2A, NEIL3 and DEPDC1 (Fig. 6A). The expression of the six hub genes were further verified in the EBV infection dataset (GSE49628), SLE dataset (GSE61635) and DLBCL datasets (GSE32018 and TCGA\_GTE<sub>x</sub>-DLBC) (Fig. 6B–E). All of the six genes expressions were higher in EBV infection, SLE and DLBCL group compared to the control group. Through the GeneMANIA database, a complex gene interaction network was constructed for understanding the biological roles of the common hub genes, with the co-expression of 92.3%, co-localization of 3.98%, physical interactions of 3.52%, pathway of 0.12% and shared protein domains of 0.08% (Fig. 6F). Based on the 6 hub genes, 20 related genes were identified, showing that they were mainly associated with mitotic nuclear division, chromosome segregation and cell cycle checkpoint. Besides, GO and pathway enrichment of the hub genes was analyzed, showing the similar results from DEGs (Supplementary Figs. 2 and 3).

Category	GO ID	Term	P-Values	Genes
GO Biological process	GO:0045685	Regulation of glial cell differentiation	2.65E-04	CDK1;TNFRSF21
	GO:0009070	Serine family amino acid biosynthetic process	8.24E-04	MTHFD1;PSPH
	GO:1901653	Cellular response to peptide	0.001138615	FYN;IGF1;CHMP5
	GO:0045445	Myoblast differentiation	0.001301854	MBNL1;IGF1
	GO:0009069	Serine family amino acid metabolic process	0.001484374	MTHFD1;PSPH
	GO:0032511	Late endosome to vacuole transport via multivesicular body sorting pathway	0.001484374	LEPROT;CHMP5
	GO:1903829	Positive regulation of cellular protein localization	0.002145527	CDK1;FYN;BICD1
	GO:0045821	Positive regulation of glycolytic process	0.002567618	ZBTB20;IGF1
	GO:1900544	Positive regulation of purine nucleotide metabolic process	0.002567618	ZBTB20;IGF1
	GO:0032481	Positive regulation of type I interferon production	0.002704008	ZBP1;RIOK3;LRRFIP1
GO Cellular Component	GO:0043231	Intracellular membrane-bounded organelle	0.001192748	TOP2A;NEK2;TMPO;KIF23;NEIL3;DEPDC1;CDK1
	GO:0099503	Secretory vesicle	0.004872687	IGF1;BICD1
	GO:0005634	Nucleus	0.005698888	TOP2A;NEK2;TMPO;KIF23;NEIL3;DEPDC1;CDK1
	GO:0097208	Alveolar lamellar body	0.021409026	LAMP3
	GO:1990246	Uniplex complex	0.021409026	MICU3
	GO:0010494	Cytoplasmic stress granule	0.022972237	MBNL1;ELAVL1
	GO:0101002	Ficolin-1-rich granule	0.028757131	CLEC4C;CAND1;TNFAIP6
	GO:0030688	Preribosome, small subunit precursor	0.038904184	RIOK3
	GO:0005736	RNA polymerase I complex	0.045815566	POLR1E
	GO:0097386	Glial cell projection	0.049252856	FYN
GO Molecular function	GO:0003725	Double-stranded RNA binding	6.52064E-06	ZBP1;MBNL1;STRBP;LRRFIP1;ELAVL1
	GO:0003723	RNA binding	0.001391275	TOP2A;ZBP1;MBNL1;STRBP;MRPS23;MSI2;ELAVL1
	GO:0016799	Hydrolase activity, hydrolyzing <i>N</i> -glycosyl compounds	0.001484374	NEIL3;MACROD2
	GO:0035925	mRNA 3'-UTR AU-rich region binding	0.002817823	ELAVL1;RBMS3
	GO:0050699	WW domain binding	0.005208639	PRRG4;PMEPA1
	GO:0045296	Cadherin binding	0.006086267	PPFIBP1;LRRFIP1;RPL15;TMPO;CHMP5
	GO:0004715	Non-membrane spanning protein tyrosine kinase activity	0.007042367	FYN;FRK
	GO:0003690	Double-stranded DNA binding	0.008866374	ZBP1;KLF12;NEIL3;ZBTB20;BACH2;KLF2;NR3C2
	GO:0000287	Magnesium ion binding	0.015749168	TOP2A;EPHX2;PSPH
	GO:0005315	Inorganic phosphate transmembrane transporter activity	0.017872498	ANKH

**Table 2.** Ontological analysis of shared DEGs between EBV, SLE, and DLBCL.

**A GO Biological Processes****B GO Cellular Component****C GO Molecular Function**

**Figure 3.** GO terms of common genes between EBV infection, SLE and DLBCL. (A) Biological Processes, (B) cellular component, (C) molecular function.

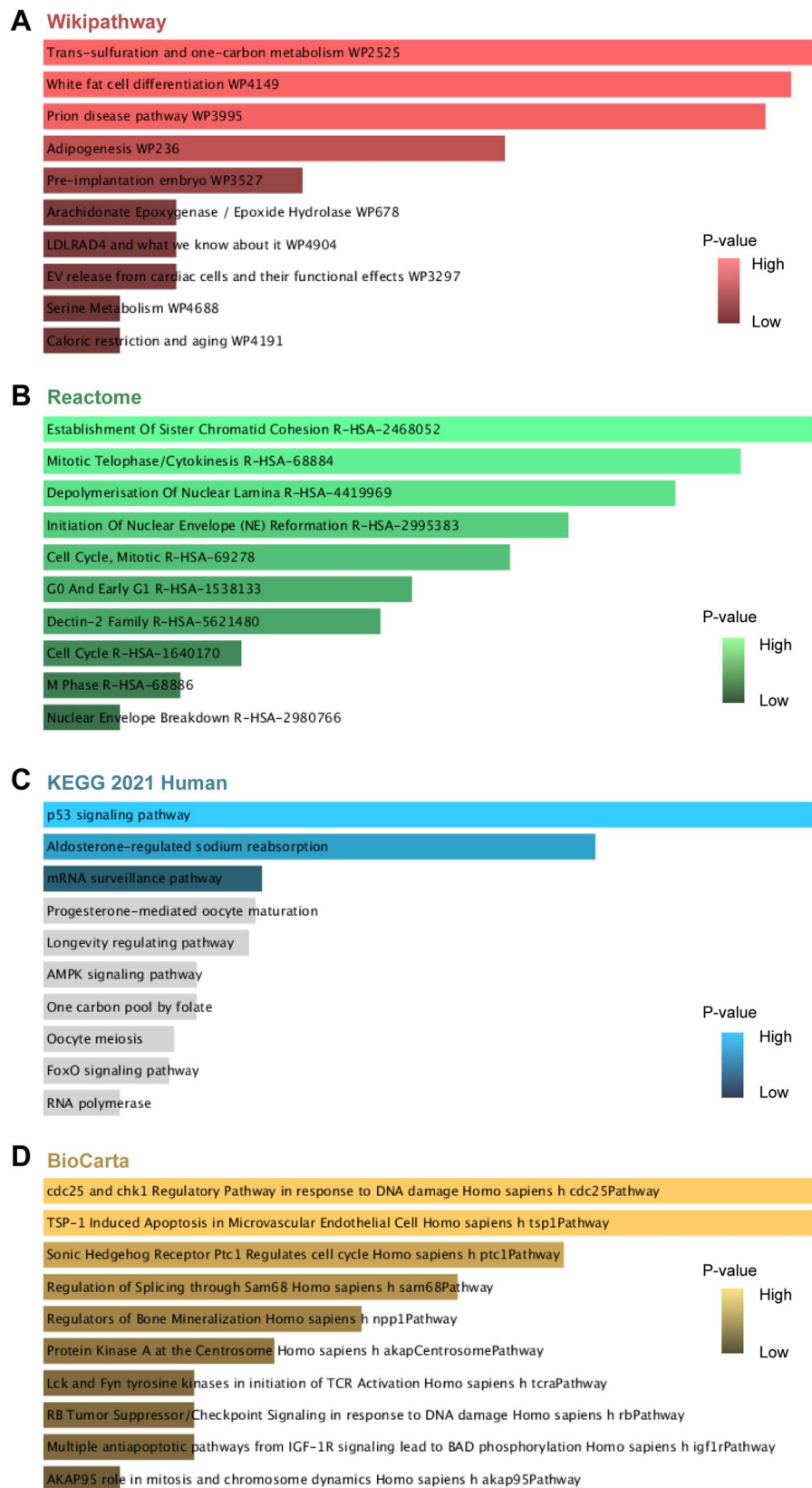
**Determination of regulatory signatures.** To identify the regulatory molecules of hub genes at the transcriptional level, we constructed TFs-gene and miRNAs-gene networks using NetworkAnalyst platform, which were visualized by Cytoscape. As shown in Fig. 7, the TFs-gene interaction network contains 50 nodes and 52 edges. CDK1 was modulated by 18 TF genes, and NEK2 was modulated by 17 TF genes. The TFs such as SIN3A, GABPA, ZNF18 and ZNF24 regulated many hub genes in the network. The miRNA-gene regulatory network was also predicted by NetworkAnalyst and created by Cytoscape, including 159 nodes and 185 edges (Fig. 8). It has been ascertained that 30 miRNAs regulated with more than one hub gene, which demonstrated the high interaction between them.

Category	Pathways	P-Values	Genes
WikiPathways	Trans-sulfuration and one-carbon metabolism WP2525	0.005554958	MTHFD1;PSPH
	White fat cell differentiation WP4149	0.005911567	EBF1;KLF2
	Prion disease pathway WP3995	0.006278388	EBF1;FYN
	Adipogenesis WP236	0.011549909	MBNL1;EBF1;IGF1
	Pre-implantation embryo WP3527	0.018552642	ELAVL1;NR3C2
	Arachidonate Epoxygenase / Epoxide Hydrolase WP678	0.024932997	EPHX2
	LDLRAD4 and what we know about it WP4904	0.024932997	PMEPA1
	EV release from cardiac cells and their functional effects WP3297	0.024932997	KLF2
	Serine Metabolism WP4688	0.028444455	PSPH
	Caloric restriction and aging WP4191	0.028444455	IGF1
Reactome	Establishment Of Sister Chromatid Cohesion R-HSA-2468052	6.88E-04	PMEPA1;ESCO2
	Mitotic Telophase/Cytokinesis R-HSA-68884	9.72E-04	PMEPA1;KIF23
	Depolymerisation Of Nuclear Lamina R-HSA-4419969	0.001301854	CDK1;TMPO
	Initiation Of Nuclear Envelope (NE) Reformation R-HSA-2995383	0.002100521	CDK1;TMPO
	Cell Cycle, Mitotic R-HSA-69278	0.002729699	TOP2A;CDK1;PMEPA1;NEK2;KIF23;ESCO2
	G0 And Early G1 R-HSA-1538133	0.004232191	TOP2A;CDK1
	Dectin-2 Family R-HSA-5621480	0.004872687	CLEC4C;FYN
	Cell Cycle R-HSA-1640170	0.009080403	TOP2A;CDK1;PMEPA1;NEK2;KIF23;ESCO2
	M Phase R-HSA-68886	0.011934571	CDK1;PMEPA1;NEK2;KIF23;TMPO
	Nuclear Envelope Breakdown R-HSA-2980766	0.015645345	CDK1;TMPO
KEGG 2021 Human	p53 signaling pathway	0.002322805	SESN3;CDK1;IGF1
	Aldosterone-regulated sodium reabsorption	0.007846291	IGF1;NR3C2
	mRNA surveillance pathway	0.04867657	MSI2;GSPT1
	Progesterone-mediated oocyte maturation	0.050463783	CDK1;IGF1
	Longevity regulating pathway	0.052274479	SESN3;IGF1
	AMPK signaling pathway	0.069559815	IGF1;ELAVL1
	One carbon pool by folate	0.069621662	MTHFD1
	Oocyte meiosis	0.078812459	CDK1;IGF1
	FoxO signaling pathway	0.080918832	IGF1;KLF2
	RNA polymerase	0.105852996	POLR1E
BioCarta	cdc25 and chk1 Regulatory Pathway in response to DNA damage Homo sapiens h cdc-25Pathway	0.024932997	CDK1
	TSP-1 Induced Apoptosis in Microvascular Endothelial Cell Homo sapiens h tsp1Pathway	0.024932997	FYN
	Sonic Hedgehog Receptor Ptc1 Regulates cell cycle Homo sapiens h ptc1Pathway	0.031943443	CDK1
	Regulation of Splicing through Sam68 Homo sapiens h sam68Pathway	0.035430005	CDK1
	Regulators of Bone Mineralization Homo sapiens h npp1Pathway	0.038904184	ANKH
	Protein Kinase A at the Centrosome Homo sapiens h akapCentrosomePathway	0.042366023	CDK1
	Lck and Fyn tyrosine kinases in initiation of TCR Activation Homo sapiens h tcraPathway	0.045815566	FYN
	RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage Homo sapiens h rbPathway	0.045815566	CDK1
	Multiple antiapoptotic pathways from IGF-1R signaling lead to BAD phosphorylation Homo sapiens h igf1rPathway	0.045815566	IGF1
AKAP95 role in mitosis and chromosome dynamics Homo sapiens h akap95Pathway	0.049252856	CDK1	

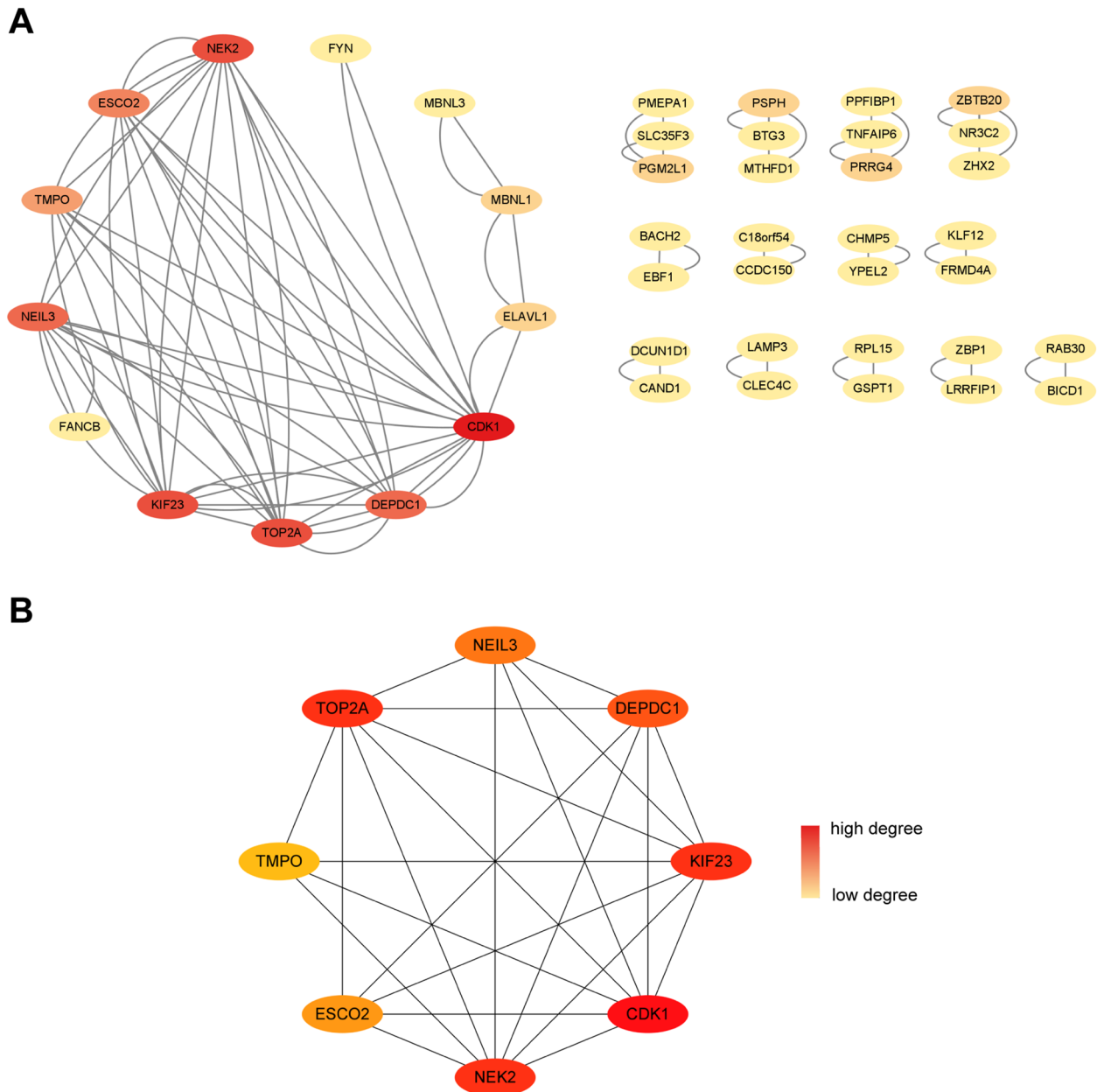
**Table 3.** Pathway enrichment analysis of shared DEGs between EBV, SLE, and DLBCL.

**ROC curves of hub genes.** The diagnostic effectiveness of the 6 hub genes was evaluated using ROC curves. To identify SLE patients from healthy controls in the SLE dataset (GSE61635), CDK1 (AUC: 0.923) and TOP2A (AUC: 0.904) showed preferable diagnostic effectiveness (Fig. 9A). For separating DLBCL patients from healthy controls in the DLBCL dataset (GSE32018), TOP2A (AUC: 0.935), DEPDC1 (AUC: 0.935), NEIL3 (AUC: 0.922), and NEK2 (AUC: 0.916) showed good diagnostic performance (Fig. 9B). To distinguish DLBCL patients from healthy controls in the TCGA\_GETx dataset (Fig. 9C), TOP2A (AUC: 0.825), CDK1 (AUC: 0.799) and NEK2 (AUC: 0.782) ranked the top 3 genes.

**Immune infiltration assessment.** We investigated the relationship between the expression of hub genes and the infiltration of immune cells in DLBCL based on the TIMER database (Fig. 10A). The results showed that the expressions of CDK1, KIF23, NEK2, TOP2A, NEIL3 and DEPDC1 were all positively associated with T helper 2 (Th2) cells and T helper cells. On the contrary, the expressions of these six hub genes were negatively linked with plasmacytoid dendritic cells (pDCs) and NK CD56bright cells. In addition, immune checkpoint cor-



**Figure 4.** Pathway enrichment analysis of common genes between EBV infection, SLE and DLBCL. (A) Wikipathway, (B) Reactome Pathway, (C) KEGG Human Pathway, (D) BioCarta Pathway.



**Figure 5.** PPI network and gene module analysis. **(A)** PPI network diagram. **(B)** Significant gene module. The redder the color of the gene in the network, the higher the connectivity of the gene with other genes.

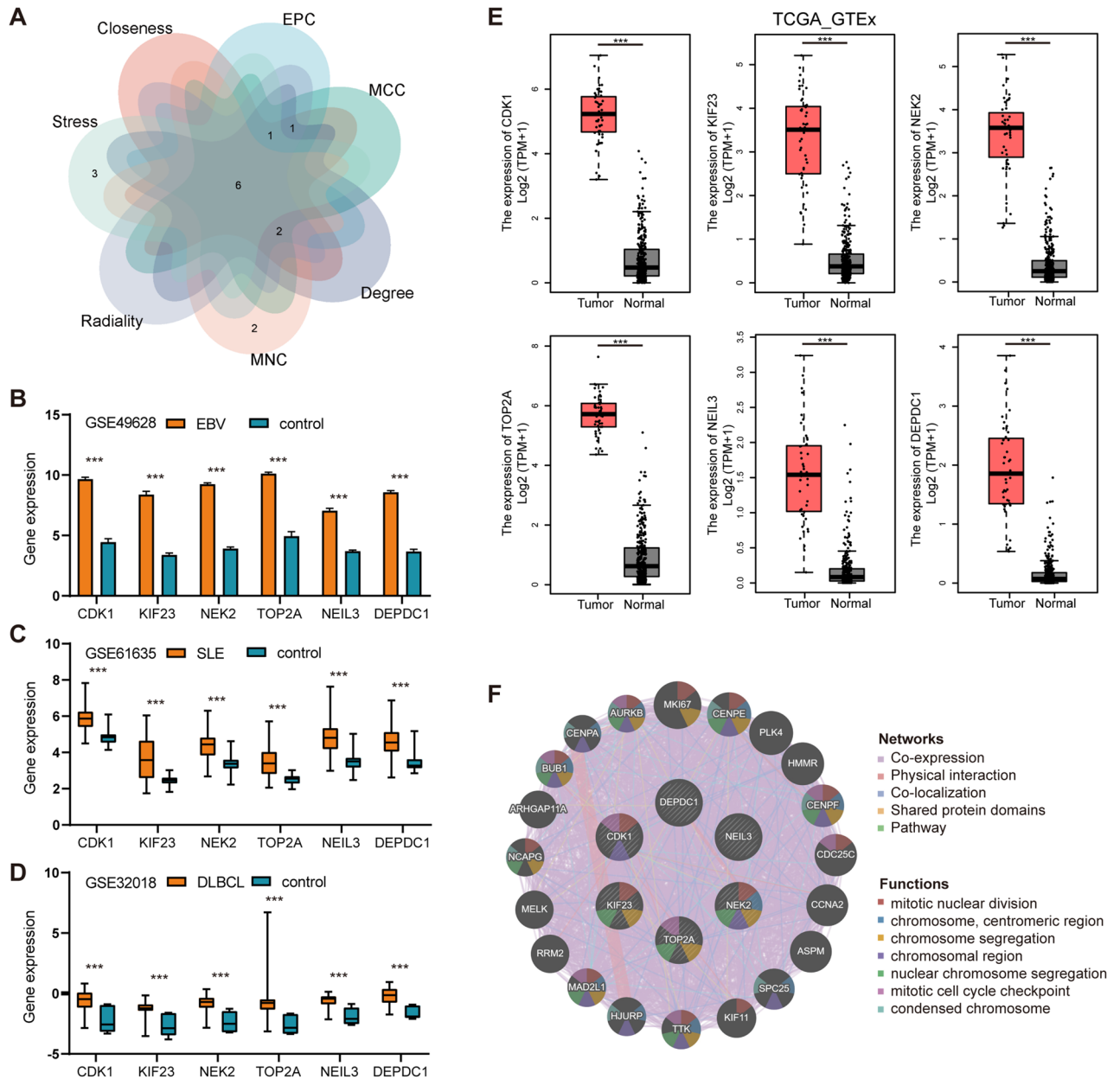
relation analysis showed that hub genes were closely associated with various immune checkpoint factors, such as NRP1, TNFSF18 and LGALS9 (Fig. 10B).

**Prediction of candidate drugs.** Using the Enrichr platform, which is based on the DSigDB database, the top 10 candidate therapeutic compounds were extracted and sorted by their *P*-value in the areas of hub genes as prospective pharmacological targets for EBV infection, SLE, and DLBCL (Table 4). It was discovered that the three pharmacological molecules that interacted with the majority of genes were testosterone (CTD 00006844), resveratrol (CTD 00002483), and calcitriol (CTD 00005558).

## Discussion

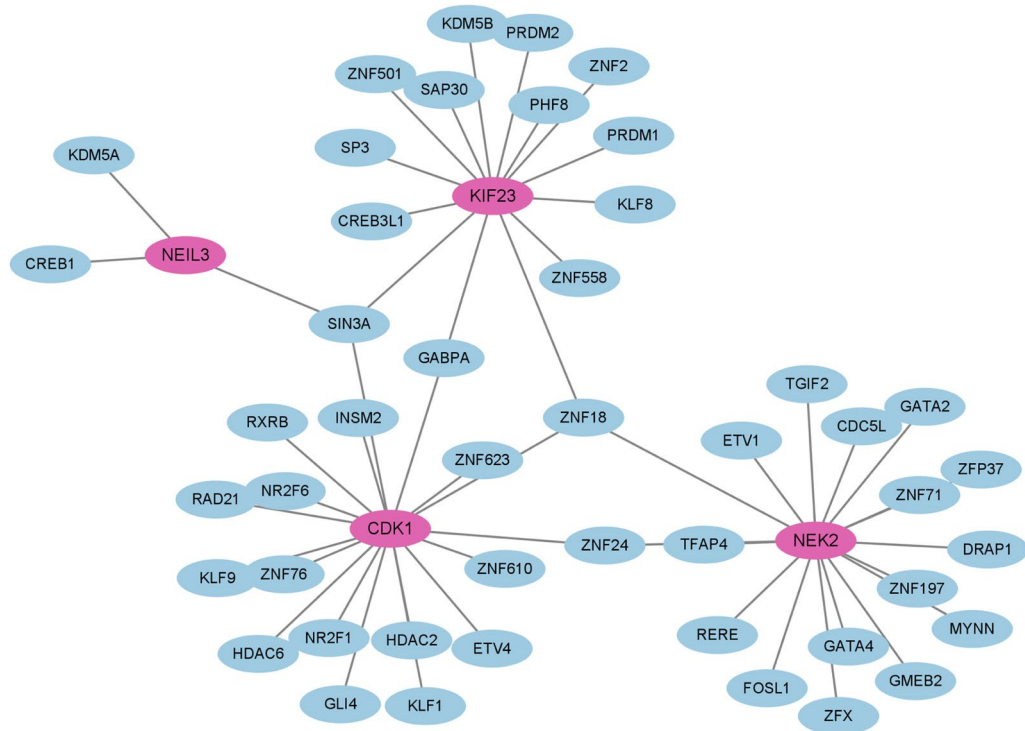
Several studies have indicated that EBV infections could trigger the occurrence of SLE and DLBCL. Additionally, SLE patients are more likely to develop DLBCL, suggesting that EBV infection may be the link between SLE and DLBCL. However, the underlying mechanism is still unclear. In this study, we performed a series of bioinformatics analysis and attempted to reveal the possible molecular mechanisms by which EBV infection contribute to the pathogenesis of DLBCL in SLE patients for the first time and discovered potential biomarkers and therapeutic targets for SLE and DLBCL.





**Figure 6.** Venn diagram and hub gene co-expression network. (A) The Venn diagram displayed six hub genes that were filtered by seven algorithms. (B) Differential expression of hub genes in EBV-transformed B cells and resting B cells in the GSE49628 dataset. (C) Differential expression of hub genes in SLE patients and healthy donors in the GSE61635 dataset. (D) Differential expression of hub genes between DLBCL samples and healthy controls in the GSE32018 dataset. (E) Differential expression of hub genes between DLBCL samples and healthy controls in the TCGA\_GTEX-DLBC dataset. (F) GeneMANIA was used to assess hub genes and the co-expressed genes. \* $P < 0.05$ , \*\* $P < 0.01$ , and \*\*\* $P < 0.001$ .

GO analysis was performed based on the 72 common DEGs identified by the three datasets of EBV infection, SLE and DLBCL. In terms of biological process terms, regulation of glial cell differentiation, serine family amino acid biosynthetic process, cellular response to peptide, myoblast differentiation and serine family amino acid metabolic process were the most significant. Serine is a significant supply of one-carbon units, which are necessary building blocks for the production of nucleotides. EBV upregulates import and synthesis of serine through its encoded protein EBNA2, which is essential for EBV-driven B cell survival and proliferation<sup>22</sup>. Moreover, reducing the amount of extracellular serine available or preventing the production of serine from glycolytic intermediates have also been proposed as innovative therapeutic approaches for the treatment of autoimmune diseases<sup>23</sup> and B-cell lymphomas<sup>24</sup>. Regarding the molecular function, double-stranded RNA binding, hydrolase activity, cadherin binding were the top terms. EBER-1 and EBER-2, two short RNAs that EBV encodes, are highly expressed in latently infected cells. EBERs can bind to the double-stranded RNA-activated protein kinase (PKR) and prevent it from being phosphorylated, leading to the resistance of IFN- $\alpha$ -induced apoptosis in



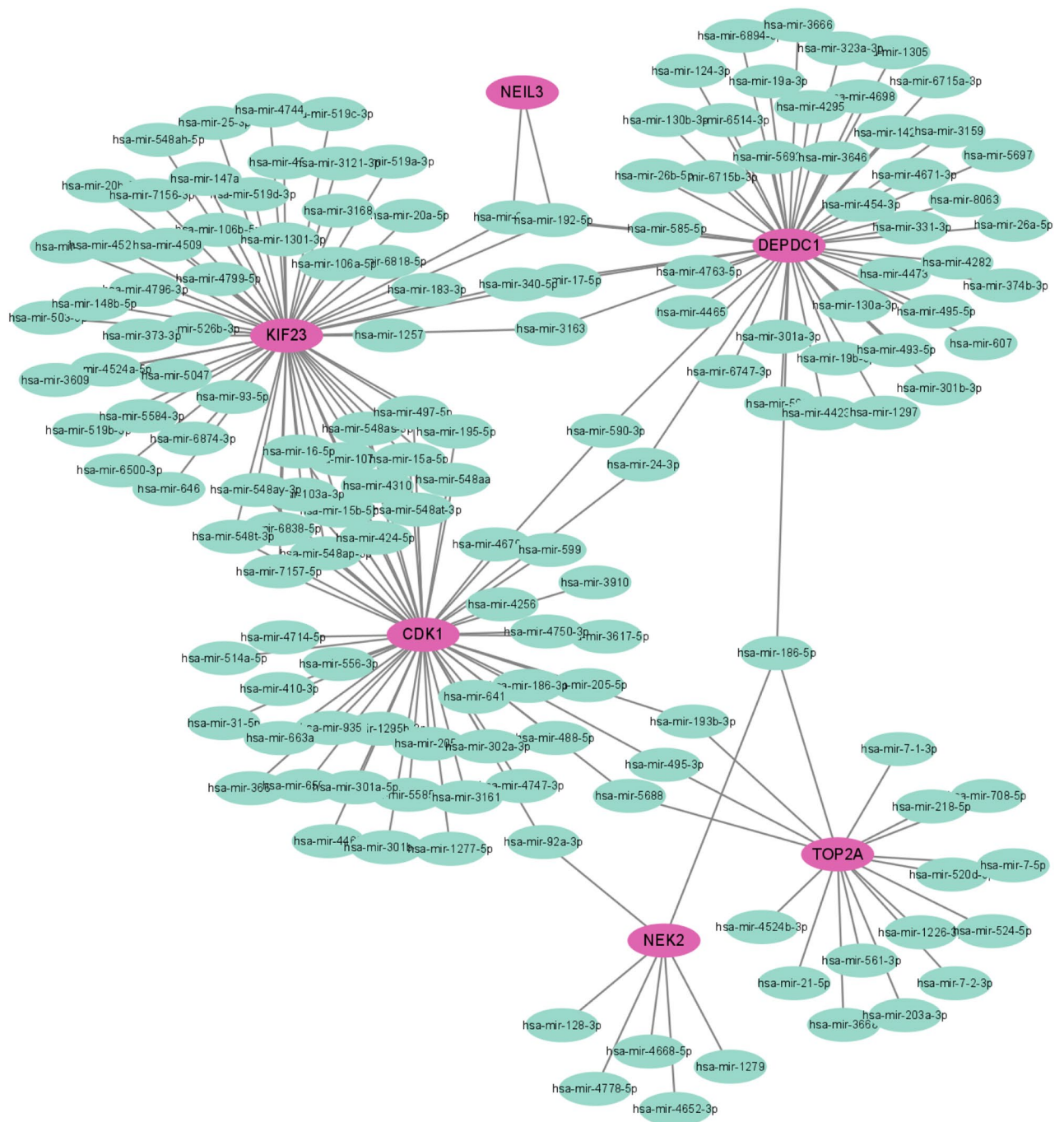
**Figure 7.** The TFs-gene regulatory network. Hub genes are represented by the pink nodes, and TF genes are represented by the other blue nodes.

EBV-positive lymphoma<sup>25</sup>. As for cellular components, intracellular membrane-bounded organelle, secretory vesicle and nucleus ranked the top 3. Secretory vesicles released from EBV-infected cells are capable of promoting inflammation and immune dysfunction, which may contribute to autoimmune diseases and numerous malignancies<sup>26</sup>.

The KEGG pathway enrichment analysis showed that p53 signaling pathway, aldosterone-regulated sodium reabsorption and mRNA surveillance pathway are the top 3 significant pathways. Mutations that inactivate p53 promote genomic instability and are hallmark of cancer<sup>27</sup>. EBV-encoded oncoprotein latent membrane protein 1 (LMP1) is reported to promote the growth of lymphoma cells by the degradation of p53<sup>28</sup>. Interestingly, mutations in p53 have been shown to increase autoimmune susceptibility in multiple strains of mice<sup>29,30</sup>. The autoantibodies against p53 have been detected in the serums of patients with SLE, which could functionally block p53 activation and affect apoptosis<sup>31</sup>. Results from WikiPathways indicated that trans-sulfuration and one-carbon metabolism, white fat cell differentiation and prion disease pathway are the most significant. The trans-sulfuration route is a biochemical process that connects methionine metabolism to the production of cellular redox-controlling molecules, which contributes to atherosclerosis and tumor development<sup>32</sup>. Reactome analysis showed that the most interacted gene pathways are establishment of sister chromatid cohesion, mitotic telophase/cytokinesis and depolymerisation of nuclear lamina. According to recent research, abnormal sister chromatid cohesion leads to chromosomal instability, which in turn promotes the growth of cancer<sup>33</sup>. As for BioCarta, the top pathways are as follows: cdc25 and chk1 regulatory pathway in response to DNA damage, TSP-1 induced apoptosis in microvascular endothelial cell and sonic hedgehog receptor Ptc1 regulates cell cycle.

Based on the PPI network, CDK1, KIF23, NEK2, TOP2A, NEIL3 and DEPDC1 were selected as hub genes. CDK1 is able to regulate cell cycle progression and transcription<sup>34</sup>. CDK1 expression was up-regulated in EBV-positive DLBCL and nasal natural killer/T-cell lymphoma (NKNL). Treatment with CDK1 inhibitors causes the death of EBV-transformed cells<sup>35–37</sup>. In addition, type I interferon (IFN) signaling is also thought to be a major pathogenic route in SLE. The excessively increased type I IFN signaling in SLE may be caused by overexpression of CDK1. KIF23, a kinesin 6 family member that is found at the interzone of mitotic spindles, is essential for cytokinesis<sup>38</sup>. KIF23 expression is increased in DLBCL and is a risk factor for this disease<sup>39</sup>. NEK2, a member of NIMA-related kinase family that regulates cell cycle, is up-regulated in a variety of malignancies, including DLBCL<sup>40,41</sup>. TOP2A is a multifunctional nuclear enzyme required during DNA replication, transcription and DNA damage repair<sup>42</sup>. It was identified that TOP2A-nucleolin interaction is essential for regulating Top2A targeting agent induced DLBCL cell death<sup>43</sup>. TOP2A is also proposed as potential biomarker for SLE diagnosis by comparative analysis<sup>44</sup>. For the remaining two hub genes, NEIL3 and DEPDC1, there are no publications describing their role in EBV, SLE or DLBCL, which highlights its significance for future research.

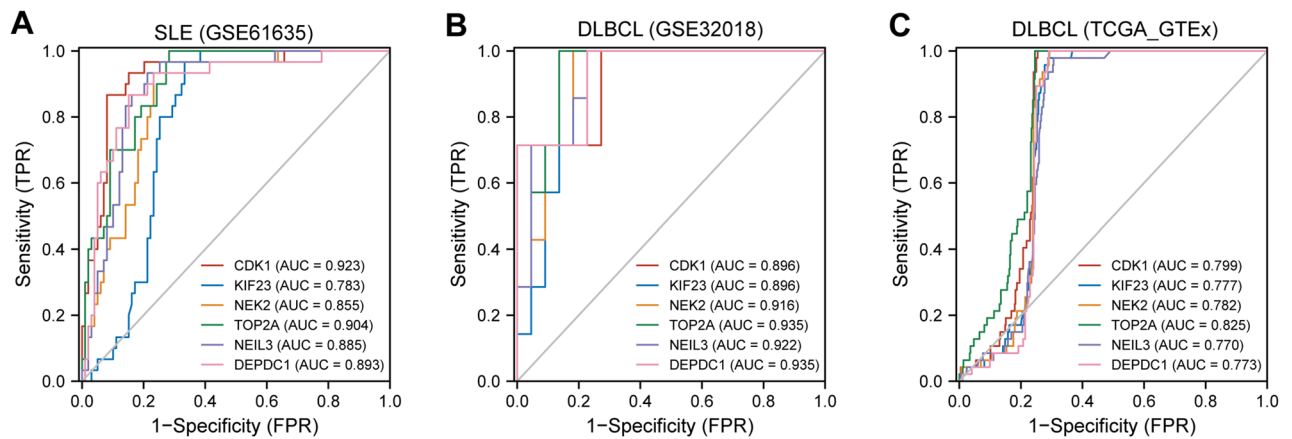
We also performed the TFs-gene and miRNAs-gene connection to discover the transcriptional and post-transcriptional regulators of the hub genes. The TFs such as SIN3A and ZNF18 regulated most hub genes in the network. SIN3A is hypothesized to control gene expression by acting as histone deacetylases, which is linked



**Figure 8.** The miRNAs-gene regulatory network. Hub genes are represented by the pink nodes, and miRNAs are represented by the other green nodes.

to tumor progression<sup>45</sup>. EBV nuclear antigen 3C is reported to recruit SIN3A to repress CDKN2A, which is important for immortal human B-lymphoblastoid cell line proliferation<sup>46</sup>. Furthermore, three miRNAs, including hsa-miRNA-186, hsa-miRNA-192 and hsa-miRNA-215 were identified to regulate most hub genes. These miRNAs play critical role in several types of cancers, such as breast cancer, gastric cancer and colorectal cancer<sup>47–49</sup>. However, their function in SLE or DLBCL has not been reported.

With the ROC curves, we found that six hub genes exhibit good diagnostic performance in SLE and DLBCL. Furthermore, we used a cohort of DLBCL patients from TCGA dataset to investigate the association between the expression levels of the six hub genes and overall survival. By utilizing GEPIA2<sup>50</sup> tool, Kaplan–Meier survival analysis and Cox proportional hazards regression to assess the association between gene expression and patient survival were conducted (Supplementary Fig. 1). However, we observed that the expression levels of these hub genes were not associated with overall survival in DLBCL patients. These findings suggest that while these hub genes may be useful for diagnostic purposes, they may not be reliable prognostic markers for DLBCL.



**Figure 9.** Verification of common diagnostic biomarkers. **(A)** The ROC curve used to verify the diagnostic efficacy in GSE61635. **(B)** The ROC curve used to verify the diagnostic efficacy in GSE32018. **(C)** The ROC curve used to verify the diagnostic efficacy in TCGA\_GTEEx dataset.

With assessment of immune infiltration, we observed that hub genes expressions are positively correlated with Th2 cell infiltration and negatively linked with pDC in DLBCL. Th2 cells have been primarily associated with promoting tumor growth and suppressing anti-tumor immunity<sup>51</sup>, while pDCs can contribute to anti-tumor immunity by promoting the activation and expansion of effector immune cells, and by producing type I interferons<sup>52</sup>. In addition, we found that these 6 hub genes had the highest positive correlation with NRP1 expression. NRP1 can promote tumor angiogenesis, tumor cell migration and invasion, and the infiltration of immunosuppressive cells into the tumor microenvironment<sup>53,54</sup>. This finding suggested that hub genes may play a role in remodeling immune landscape of tumor microenvironment (TME) to promote DLBCL.

Six hub genes were applied to the DSignDB database for the prediction of potential medicines. The top 3 chemical molecules were listed as follow: lucanthone, testosterone and trifluridine. Recently, it was discovered that the anti-schistosomal drug lucanthone, which may pass the blood–brain barrier, inhibit autophagy and suppress the growth of breast cancer and glioblastoma<sup>55,56</sup>. In peripheral blood mononuclear cells from SLE patients, testosterone, one of the male hormones, is able to reduce the production of anti-DNA antibodies by inhibiting B cell hyperactivity, supporting the therapeutic effects for SLE<sup>57,58</sup>. Trifluridine, a thymidine-based nucleoside analog, is a novel oral cytotoxic chemotherapy licensed for the treatment of metastatic colorectal cancer refractory to standard therapies<sup>59</sup>. It is also an antiviral agent for topical use in the eye, which could effectively inhibit the replication of herpes simplex virus type 1<sup>60</sup>.

Certainly, our study has some limitations. On the one hand, there was a lack of information on patients with DLBCL secondary to SLE with EBV infection to further verify our finding. On the other hand, our findings were obtained by pure bioinformatics analysis, so the function of hub genes and prospective medicines is required to be further confirmed by scientific investigation *in vitro* and *in vivo*.

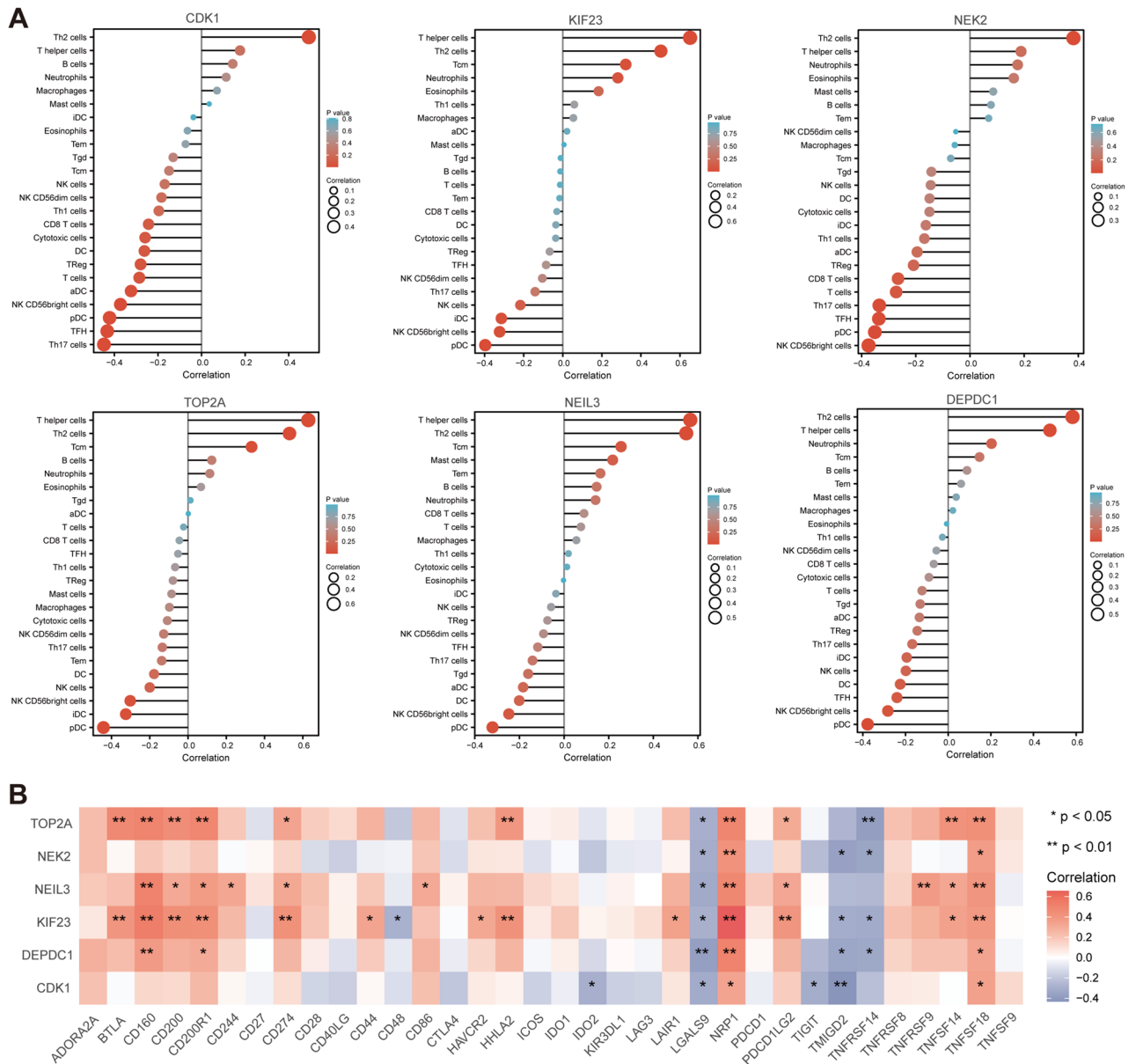
## Conclusion

To the best of our knowledge, our study is the first to reveal shared DEGs, GO and pathway enrichment and PPI network for EBV infection, SLE and DLBCL using bioinformatic analysis to explore the potential molecular mechanisms underlying the contribution of EBV infection to the development of DLBCL in SLE patients. Our research also identified immune-related biomarkers and future therapeutic targets for patients with SLE and DLBCL, which will help better manage SLE patients and provide early diagnosis and treatment for DLBCL.

## Methods

**Compilation of datasets.** GEO ([www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)) is a big database providing gene expression profiles for a variety of disorders. It is free of charge and publicly available<sup>61</sup>. GSE49628<sup>62</sup> dataset contains resting and EBV transformed B cells from 3 donors and GSE61635 dataset consists of 79 SLE patients and 30 healthy donors, which were both sequenced using the Affymetrix Human Genome U133 Plus 2.0 Array platform. GSE32018<sup>63</sup> is of 21 DLBCL samples and 7 healthy controls, which was sequenced by Agilent-014850 Whole Human Genome Microarray 4 × 44 K G4112F platform. Gene expression data and patient clinical information were downloaded from TCGA database (project ID: TCGA\_GTEEx-DLBC) (<https://portal.gdc.cancer.gov/>). In total, information for 41 DLBCL tissues and 447 adjacent normal tissues was obtained.

**Identification of shared DEGs between EBV infection, SLE and DLBCL.** An online program called GEO2R ([www.ncbi.nlm.nih.gov/geo/geo2r/](http://www.ncbi.nlm.nih.gov/geo/geo2r/)) can be used in the comparison and analysis between the groups with different gene expression<sup>64</sup>. GEO2R was used to identify DEGs for GSE49628, GSE61635 and GSE32018. DEGs were defined as genes with  $|\log_2 \text{fold change} (\log_2 \text{FC})| > 1.0$  and an adjusted *P*-value  $< 0.05$ . The volcano map of DEGs from each dataset and the Venn diagram of shared DEGs between these three datasets were both plotted by Bioinformatics (<https://www.bioinformatics.com.cn>).



**Figure 10.** Immune landscape assessment. (A) Relationship of hub genes expression and immune cell subtypes in DLBCL patients. (B) Heatmap of the correlation between hub genes expression and immune checkpoint. \*P < 0.05, \*\*P < 0.01, and \*\*\*P < 0.001.

Name of drugs	P-Values	Genes
LUCANTHONE CTD 00006227	7.96E-10	TOP2A;CDK1;DEPDC1;KIF23;NEK2
testosterone CTD 00006844	5.84E-08	TOP2A;NEIL3;CDK1;DEPDC1;KIF23;NEK2
trifluridine MCF7 DOWN	8.16E-08	TOP2A;DEPDC1;KIF23
monobenzene PC3 DOWN	1.55E-07	TOP2A;DEPDC1;KIF23;NEK2
0173570-0000 PC3 DOWN	1.84E-07	DEPDC1;KIF23;NEK2
troglitazone CTD 00002415	2.12E-07	TOP2A;NEIL3;CDK1;KIF23;NEK2
resveratrol CTD 00002483	2.62E-07	TOP2A;NEIL3;CDK1;DEPDC1;KIF23;NEK2
methotrexate MCF7 DOWN	3.30E-07	TOP2A;DEPDC1;KIF23
cidopirox MCF7 DOWN	7.13E-07	TOP2A;DEPDC1;KIF23
calcitriol CTD 00005558	8.77E-07	TOP2A;NEIL3;CDK1;DEPDC1;KIF23;NEK2

**Table 4.** Prediction of top 10 candidate drugs for EBV, SLE, and DLBCL.

**Investigation of GO and pathways enrichment.** In order to find the functions of the shared DEGs or hub genes in EBV infection, SLE and DLBCL, we utilized Enrichr<sup>65</sup>, a useful gene set enrichment online platform (<https://maayanlab.cloud/Enrichr/>), to conduct a series of enrichment analyses. Biological process, cellular component, and molecular function are the three elements of GO<sup>66</sup>. The shared pathways among DEGs or hub genes were identified based on four databases, including WikiPathways<sup>67</sup>, Reactome<sup>68</sup>, Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>69–72</sup>, and BioCarta<sup>73</sup>.

**PPI network construction and module analysis.** The assessment of PPI network is the cornerstone of cellular biology for understanding protein function and the mechanism of cellular machinery operations. Search Tool for the Retrieval of Interacting Genes (STRING, <http://string-db.org>) is a database for studying PPI network with physical and functional interactions<sup>74</sup>. We used STRING to create the PPI network of shared DEGs with an interaction score greater than 0.4 and displayed it by Cytoscape (Version 3.9.1)<sup>75</sup>. The core functional module was produced using a Cytoscape plug-in—Molecular Complex Detection (MCODE)<sup>76</sup>.

**Recognition of hub genes.** The hub genes of this study were identified using CytoHubba<sup>77</sup>, a plug-in of Cytoscape. Subsequently, the final hub genes were confirmed by seven algorithms (MCC, MNC, EPC, Closeness, Degree, Radiality and Stress) and visualized by Venn diagram. Based on the final hub genes, co-expression networks were constructed by GeneMANIA (<http://genemania.org>), an online platform for gene interactions prediction<sup>78</sup>.

**Validation of the hub genes in EBV infection, SLE and DLBCL.** To confirm hub shared genes in EBV infection, SLE and DLBCL, we conducted the differentially expressed gene analysis on validation datasets (GSE49628, GSE61635 and GSE32018). GEPIA2<sup>50</sup> (<http://gepia2.cancer-pku.cn/#index>) is an analysis tool containing RNA sequence expression data of tumors and normal tissue samples. We used GEPIA2 to analyze TCGA and GTEx databases of gene expression profiles in DLBCL. Unpaired comparisons of hub gene expression between two groups were analyzed by Wilcoxon rank-sum test. We also performed prognostic analysis of hub genes in DLBCL with the “survival plots” module using a Kaplan–Meier curve with GEPIA2.

**Identification of TFs and miRNAs interactions with hub genes.** Proteins called TFs control how quickly genes are transcribed by binding to certain DNA regions. TFs-gene networks were created by utilizing ENCODE database<sup>79</sup> on NetworkAnalyst platform (<https://www.networkanalyst.ca/>)<sup>80</sup>. Furthermore, miRNAs are a group of short non-coding RNAs, which can impede translation or degrade the target mRNA. The network of miRNAs-gene was acquired from mirTarbase<sup>81</sup> via NetworkAnalyst. Cytoscape was used to display TFs-gene and miRNAs-gene regulatory networks.

**Receiver operating characteristic curves of hub genes.** Receiver operating characteristic (ROC) curves of the hub genes on SLE (GSE61635) and DLBCL (GSE32018 and TCGA\_GTEx-DLBC) were both plotted by Bioinformatics platform. The diagnostic ability of each hub gene was assessed using the calculation of area under ROC curve (AUC).

**Assessment of the immune landscape.** We used immune cell infiltration and gene expression data from the TIMER database<sup>82</sup> to identify relationships between the expression of hub genes and immune cell abundance in DLBCL, and plotted bubble plots to show these results. Spearman’s correlation analysis was used to analyze the relationship between hub genes and a range of immune-related genes, such as immune checkpoint-associated genes and immune cell subpopulation-associated genes.

**Prediction of candidate drugs.** It is important to evaluate protein–drug interactions in this research. On the basis of hub genes, drug molecules were obtained from Drug Signatures database (DsigDB)<sup>83</sup> on Enrichr platform. *P*-value was used to rank the candidate medications from small to large.

**Copyright permission of KEGG.** We have contacted Kanehisa Laboratories. We do not directly use these KEGG Pathway map “images” in the article, we need not obtain copyright permission of KEGG. However, they believe that we have written our article using their data, they kindly ask us to cite the following articles in it<sup>69–72</sup>.

## Data availability

The datasets used in this investigation are accessible through online repositories. The article contains information on the repository names and accession numbers.

Received: 25 February 2023; Accepted: 15 April 2023

Published online: 18 April 2023

## References

1. Young, L. S., Yap, L. F. & Murray, P. G. Epstein-Barr virus: More than 50 years old and still providing surprises. *Nat. Rev. Cancer* **16**, 789–802. <https://doi.org/10.1038/nrc.2016.92> (2016).
2. Speck, P., Haan, K. M. & Longnecker, R. Epstein-Barr virus entry into cells. *Virology* **277**, 1–5. <https://doi.org/10.1006/viro.2000.0624> (2000).

3. Odumade, O. A., Hogquist, K. A. & Balfour, H. H. Jr. Progress and problems in understanding and managing primary Epstein-Barr virus infections. *Clin. Microbiol. Rev.* **24**, 193–209. <https://doi.org/10.1128/CMR.00044-10> (2011).
4. Dasari, V., Bhatt, K. H., Smith, C. & Khanna, R. Designing an effective vaccine to prevent Epstein-Barr virus-associated diseases: Challenges and opportunities. *Expert. Rev. Vaccines* **16**, 377–390. <https://doi.org/10.1080/14760584.2017.1293529> (2017).
5. Cohen, J. I., Fauci, A. S., Varmus, H. & Nabel, G. J. Epstein-Barr virus: An important vaccine target for cancer prevention. *Sci. Transl. Med.* **3**, 107. <https://doi.org/10.1126/scitranslmed.3002878> (2011).
6. Justiz Vaillant, A. A., Goyal, A. & Varacallo, M. In *StatPearls* (2022).
7. Luo, S., Long, H. & Lu, Q. Recent advances in understanding pathogenesis and therapeutic strategies of Systemic Lupus Erythematosus. *Int. Immunopharmacol.* **89**, 107028. <https://doi.org/10.1016/j.intimp.2020.107028> (2020).
8. Choi, M. Y. & Fritzler, M. J. Autoantibodies in SLE: Prediction and the p value matrix. *Lupus* **28**, 1285–1293. <https://doi.org/10.1177/0961203319868531> (2019).
9. Jog, N. R. & James, J. A. Epstein barr virus and autoimmune responses in systemic lupus erythematosus. *Front. Immunol.* **11**, 623944. <https://doi.org/10.3389/fimmu.2020.623944> (2020).
10. Iwata, S. & Tanaka, Y. Association of viral infection with the development and pathogenesis of systemic lupus erythematosus. *Front. Med. (Lausanne)* **9**, 849120. <https://doi.org/10.3389/fmed.2022.849120> (2022).
11. Ghabeshi, S. *et al.* Evaluation of molecular apoptosis signaling pathways and its correlation with EBV viral load in SLE patients using systems biology approach. *Hum. Antibodies* **30**, 37–46. <https://doi.org/10.3233/HAB-211505> (2022).
12. Martelli, M. *et al.* Diffuse large B-cell lymphoma. *Crit. Rev. Oncol. Hematol.* **87**, 146–171. <https://doi.org/10.1016/j.critrevonc.2012.12.009> (2013).
13. Cheson, B. D., Nowakowski, G. & Salles, G. Diffuse large B-cell lymphoma: New targets and novel therapies. *Blood Cancer J.* **11**, 68. <https://doi.org/10.1038/s41408-021-00456-w> (2021).
14. Schmitz, R. *et al.* Genetics and pathogenesis of diffuse large B-cell lymphoma. *N. Engl. J. Med.* **378**, 1396–1407. <https://doi.org/10.1056/NEJMoa1801445> (2018).
15. Chabay, P. Advances in the pathogenesis of EBV-associated diffuse large B cell lymphoma. *Cancers (Basel)* **13**, 2021. <https://doi.org/10.3390/cancers13112717> (2021).
16. Li, C. *et al.* EBNA2-deleted Epstein-Barr virus (EBV) isolate, P3HR1, causes Hodgkin-like lymphomas and diffuse large B cell lymphomas with type II and Wp-restricted latency types in humanized mice. *PLoS Pathog.* **16**, e1008590. <https://doi.org/10.1371/journal.ppat.1008590> (2020).
17. Healy, J. A. & Dave, S. S. The role of EBV in the pathogenesis of diffuse large B cell lymphoma. *Curr. Top. Microbiol. Immunol.* **390**, 315–337. [https://doi.org/10.1007/978-3-319-22822-8\\_13](https://doi.org/10.1007/978-3-319-22822-8_13) (2015).
18. Klein, A., Polliack, A. & Gafter-Gvili, A. Systemic lupus erythematosus and lymphoma: Incidence, pathogenesis and biology. *Leuk. Res.* **75**, 45–49. <https://doi.org/10.1016/j.leukres.2018.11.004> (2018).
19. Tessier-Cloutier, B. *et al.* Cell of origin in diffuse large B-cell lymphoma in systemic lupus erythematosus: Molecular and clinical factors associated with survival. *Lupus Sci. Med.* **6**, e000324. <https://doi.org/10.1136/lupus-2019-000324> (2019).
20. Gumber, M. R. *et al.* Lupus nephritis associated with CD20+ B-cell NHL. *J. Assoc. Phys. India* **59**, 520–523 (2011).
21. Abenavoli, L. *et al.* Aggressive large B-cell lymphoma in a systemic lupus erythematosus patient with chronic active Epstein-Barr virus infection: A case report. *Int. J. Immunopathol. Pharmacol.* **24**, 1083–1086. <https://doi.org/10.1177/039463201102400427> (2011).
22. Wang, L. W. *et al.* Epstein-Barr-virus-induced one-carbon metabolism drives B cell transformation. *Cell Metab.* **30**, 539–555. <https://doi.org/10.1016/j.cmet.2019.06.003> (2019).
23. Wang, T., Jiao, Y. & Zhang, X. Immunometabolic pathways and its therapeutic implication in autoimmune diseases. *Clin. Rev. Allergy Immunol.* **60**, 55–67. <https://doi.org/10.1007/s12016-020-08821-6> (2021).
24. Eraslan, Z., Papatzikas, G., Cazier, J. B., Khanim, F. L. & Gunther, U. L. Targeting asparagine and serine metabolism in germinal centre-derived B cells non-hodgkin lymphomas (B-NHL). *Cells* **10**, 253. <https://doi.org/10.3390/cells10102589> (2021).
25. Nanbo, A., Inoue, K., Adachi-Takasawa, K. & Takada, K. Epstein-Barr virus RNA confers resistance to interferon-alpha-induced apoptosis in Burkitt's lymphoma. *EMBO J.* **21**, 954–965. <https://doi.org/10.1093/emboj/21.5.954> (2002).
26. Cone, A. S., York, S. B. & Meckes, D. G. Jr. Extracellular vesicles in Epstein-Barr virus pathogenesis. *Curr. Clin. Microbiol. Rep.* **6**, 121–131. <https://doi.org/10.1007/s40588-019-00123-6> (2019).
27. Bareke, H. *et al.* Autoimmune responses in oncology: Causes and significance. *Int. J. Mol. Sci.* **22**, 8030. <https://doi.org/10.3390/ijms22158030> (2021).
28. Zeng, M., Chen, Y., Jia, X. & Liu, Y. The anti-apoptotic role of EBV-LMP1 in lymphoma cells. *Cancer Manag. Res.* **12**, 8801–8811. <https://doi.org/10.2147/CMAR.S260583> (2020).
29. Leech, M. *et al.* The tumour suppressor gene p53 modulates the severity of antigen-induced arthritis and the systemic immune response. *Clin. Exp. Immunol.* **152**, 345–353. <https://doi.org/10.1111/j.1365-2249.2008.03629.x> (2008).
30. Okuda, Y., Okuda, M. & Bernard, C. C. Regulatory role of p53 in experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* **135**, 29–37. [https://doi.org/10.1016/s0165-5728\(02\)00428-9](https://doi.org/10.1016/s0165-5728(02)00428-9) (2003).
31. Herkel, J. *et al.* Autoimmunity to the p53 protein is a feature of systemic lupus erythematosus (SLE) related to anti-DNA antibodies. *J. Autoimmun.* **17**, 63–69. <https://doi.org/10.1006/jaut.2001.0518> (2001).
32. Rosado, J. O., Salvador, M. & Bonatto, D. Importance of the trans-sulfuration pathway in cancer prevention and promotion. *Cell Biochem.* **301**, 1–12. <https://doi.org/10.1007/s11010-006-9389-y> (2007).
33. Sajesh, B. V., Lichtensztein, Z. & McManus, K. J. Sister chromatid cohesion defects are associated with chromosome instability in Hodgkin lymphoma cells. *BMC Cancer* **13**, 391. <https://doi.org/10.1186/1471-2407-13-391> (2013).
34. Nurse, P. A long twentieth century of the cell cycle and beyond. *Cell* **100**, 71–78. [https://doi.org/10.1016/s0092-8674\(00\)81684-0](https://doi.org/10.1016/s0092-8674(00)81684-0) (2000).
35. Bernasconi, M. *et al.* Early gene expression changes by Epstein-Barr virus infection of B-cells indicate CDKs and survivin as therapeutic targets for post-transplant lymphoproliferative diseases. *Int. J. Cancer* **133**, 2341–2350. <https://doi.org/10.1002/ijc.28239> (2013).
36. Nagato, T. *et al.* Cyclin-dependent kinase 1 and survivin as potential therapeutic targets against nasal natural killer/T-cell lymphoma. *Lab. Invest.* **99**, 612–624. <https://doi.org/10.1038/s41374-018-0182-9> (2019).
37. Zhao, X. F. & Gartenhaus, R. B. Phospho-p70S6K and cdc2/cdk1 as therapeutic targets for diffuse large B-cell lymphoma. *Expert Opin. Ther. Targets* **13**, 1085–1093. <https://doi.org/10.1517/14728220903103833> (2009).
38. Liu, X., Gong, H. & Huang, K. Oncogenic role of kinesin proteins and targeting kinesin therapy. *Cancer Sci.* **104**, 651–656. <https://doi.org/10.1111/cas.12138> (2013).
39. Gong, Y. *et al.* KIF23 is a potential biomarker of diffuse large B cell lymphoma: Analysis based on bioinformatics and immunohistochemistry. *Med. (Baltim.)* **101**, e29312. <https://doi.org/10.1097/MD.00000000000029312> (2022).
40. Zhou, L. *et al.* NEK2 promotes cell proliferation and glycolysis by regulating PKM2 abundance via phosphorylation in diffuse large B-cell lymphoma. *Front. Oncol.* **11**, 677763. <https://doi.org/10.3389/fonc.2021.677763> (2021).
41. Andreasson, U. *et al.* Identification of molecular targets associated with transformed diffuse large B cell lymphoma using highly purified tumor cells. *Am. J. Hematol.* **84**, 803–808. <https://doi.org/10.1002/ajh.21549> (2009).
42. Nitiss, J. L. DNA topoisomerase II and its growing repertoire of biological functions. *Nat. Rev. Cancer* **9**, 327–337. <https://doi.org/10.1038/nrc2608> (2009).

43. Jain, N. *et al.* Targeting nucleolin for better survival in diffuse large B-cell lymphoma. *Leukemia* **32**, 663–674. <https://doi.org/10.1038/leu.2017.215> (2018).
44. Fang, Q. *et al.* Comparative Analysis on abnormal methylation of differentially expressed genes and disease pathways in the immune cells of RA and SLE. *Front. Immunol.* **12**, 668007. <https://doi.org/10.3389/fimmu.2021.668007> (2021).
45. Das, T. K., Sangodkar, J., Negre, N., Narla, G. & Cagan, R. L. Sin3a acts through a multi-gene module to regulate invasion in *Drosophila* and human tumors. *Oncogene* **32**, 3184–3197. <https://doi.org/10.1038/onc.2012.326> (2013).
46. Jiang, S. *et al.* Epstein-Barr virus nuclear antigen 3C binds to BATF/IRF4 or SPI1/IRF4 composite sites and recruits Sin3A to repress CDKN2A. *Proc. Natl. Acad. Sci. USA* **111**, 421–426. <https://doi.org/10.1073/pnas.1321704111> (2014).
47. Deng, S. *et al.* miRNA-192 and -215 activate Wnt/beta-catenin signaling pathway in gastric cancer via APC. *J. Cell Physiol.* **235**, 6218–6229. <https://doi.org/10.1002/jcp.29550> (2020).
48. Gao, J. B., Zhu, M. N. & Zhu, X. L. miRNA-215-5p suppresses the aggressiveness of breast cancer cells by targeting Sox9. *FEBS Open Bio* **9**, 1957–1967. <https://doi.org/10.1002/2211-5463.12733> (2019).
49. Radwan, A. F., Shaker, O. G., El-Boghdady, N. A. & Senousy, M. A. Association of MALAT1 and PVT1 variants, expression profiles and target miRNA-101 and miRNA-186 with colorectal cancer: Correlation with epithelial-mesenchymal transition. *Int. J. Mol. Sci.* **22**, 6147. <https://doi.org/10.3390/ijms22116147> (2021).
50. Tang, Z., Kang, B., Li, C., Chen, T. & Zhang, Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **47**, W556–W560. <https://doi.org/10.1093/nar/gkz430> (2019).
51. Chen, Y. *et al.* Pharmaceutical targeting Th2-mediated immunity enhances immunotherapy response in breast cancer. *J. Transl. Med.* **20**, 615. <https://doi.org/10.1186/s12967-022-03807-8> (2022).
52. Schiavoni, G., Mattei, F. & Gabriele, L. Type I Interferons as stimulators of DC-mediated cross-priming: Impact on anti-tumor response. *Front. Immunol.* **4**, 483. <https://doi.org/10.3389/fimmu.2013.00483> (2013).
53. Hong, T. M. *et al.* Targeting neuropilin 1 as an antitumor strategy in lung cancer. *Clin. Cancer Res.* **13**, 4759–4768. <https://doi.org/10.1158/1078-0432.CCR-07-0001> (2007).
54. Chuckran, C. A., Liu, C., Bruno, T. C., Workman, C. J. & Vignali, D. A. Neuropilin-1: A checkpoint target with unique implications for cancer immunology and immunotherapy. *J. Immunother. Cancer* **8**, 2. <https://doi.org/10.1136/jitc-2020-000967> (2020).
55. Radin, D. P. *et al.* Lucanthone targets lysosomes to perturb glioma proliferation, chemoresistance and stemness, and slows tumor growth in vivo. *Front. Oncol.* **12**, 852940. <https://doi.org/10.3389/fonc.2022.852940> (2022).
56. Carew, J. S. *et al.* Lucanthone is a novel inhibitor of autophagy that induces cathepsin D-mediated apoptosis. *J. Biol. Chem.* **286**, 6602–6613. <https://doi.org/10.1074/jbc.M110.151324> (2011).
57. Kanda, N., Tsuchida, T. & Tamaki, K. Testosterone suppresses anti-DNA antibody production in peripheral blood mononuclear cells from patients with systemic lupus erythematosus. *Arthritis Rheum.* **40**, 1703–1711. <https://doi.org/10.1002/art.1780400921> (1997).
58. Bizzarro, A. *et al.* Influence of testosterone therapy on clinical and immunological features of autoimmune diseases associated with Klinefelter's syndrome. *J. Clin. Endocrinol. Metab.* **64**, 32–36. <https://doi.org/10.1210/jcem-64-1-32> (1987).
59. Kawazoe, A. & Shitara, K. Trifluridine/tipiracil for the treatment of metastatic gastric cancer. *Expert Rev. Gastroenterol. Hepatol.* **14**, 65–70. <https://doi.org/10.1080/17474124.2020.1715209> (2020).
60. Carmine, A. A., Brogden, R. N., Heel, R. C., Speight, T. M. & Avery, G. S. Trifluridine: A review of its antiviral activity and therapeutic use in the topical treatment of viral eye infections. *Drugs* **23**, 329–353. <https://doi.org/10.2165/00003495-198223050-00001> (1982).
61. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets—10 years on. *Nucleic Acids Res.* **39**, D1005–1010. <https://doi.org/10.1093/nar/gkq1184> (2011).
62. Hansen, K. D. *et al.* Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. *Genome Res.* **24**, 177–184. <https://doi.org/10.1101/gr.157743.113> (2014).
63. Gomez-Abad, C. *et al.* PIM2 inhibition as a rational therapeutic approach in B-cell lymphoma. *Blood* **118**, 5517–5527. <https://doi.org/10.1182/blood-2011-03-344374> (2011).
64. Barrett, T. *et al.* NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–995. <https://doi.org/10.1093/nar/gks1193> (2013).
65. Kuleshov, M. V. *et al.* Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–97. <https://doi.org/10.1093/nar/gkw377> (2016).
66. Gene Ontology, C. The Gene Ontology resource: Enriching a GOLD mine. *Nucleic Acids Res.* **49**, D325–D334. <https://doi.org/10.1093/nar/gkaa1113> (2021).
67. Martens, M. *et al.* WikiPathways: Connecting communities. *Nucleic Acids Res.* **49**, D613–D621. <https://doi.org/10.1093/nar/gkaa1024> (2021).
68. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503. <https://doi.org/10.1093/nar/gkz1031> (2020).
69. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M. & Tanabe, M. KEGG: Integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551. <https://doi.org/10.1093/nar/gkaa970> (2021).
70. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
71. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
72. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592. <https://doi.org/10.1093/nar/gkac963> (2023).
73. Darryl, N. BioCarta. *Biotech. Softw. Internet Rep.* **2**, 117–120. <https://doi.org/10.1089/152791601750294344> (2001).
74. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613. <https://doi.org/10.1093/nar/gky1131> (2019).
75. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
76. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **4**, 2. <https://doi.org/10.1186/1471-2105-4-2> (2003).
77. Chin, C. H. *et al.* cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**(4), S11. <https://doi.org/10.1186/1752-0509-8-S4-S11> (2014).
78. Franz, M. *et al.* GeneMANIA update 2018. *Nucleic Acids Res.* **46**, W60–W64. <https://doi.org/10.1093/nar/gky311> (2018).
79. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* **46**, D794–D801. <https://doi.org/10.1093/nar/gkx1081> (2018).
80. Xia, J., Benner, M. J. & Hancock, R. E. NetworkAnalyst—integrative approaches for protein-protein interaction network analysis and visual exploration. *Nucleic Acids Res.* **42**, W167–174. <https://doi.org/10.1093/nar/gku443> (2014).
81. Hsu, S. D. *et al.* miRTarBase: A database curates experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **39**, D163–169. <https://doi.org/10.1093/nar/gkq1107> (2011).
82. Li, T. *et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.* **48**, W509–W514. <https://doi.org/10.1093/nar/gkaa407> (2020).



83. Yoo, M. *et al.* DSigDB: Drug signatures database for gene set analysis. *Bioinformatics* **31**, 3069–3071. <https://doi.org/10.1093/bioinformatics/btv313> (2015).

## Acknowledgements

I appreciate the contributions made by the authors of the GSE49628, GSE61635, and GSE32018 datasets.

## Author contributions

Data analysis and manuscript writing were both done by Q.Z.

## Funding

This research was supported financially by the National Natural Science Foundation of China (82202509), Shenzhen Science and Technology Program (RCBS20221008093104015) and Innovative Talents Program of the Eighth Affiliated Hospital, Sun Yat-sen University (YXYXCXRC202305).

## Competing interests

The author declares no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-33585-2>.

**Correspondence** and requests for materials should be addressed to Q.-Y.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023