

Article

WGCNA-based identification of anoikis-related subtypes, prognostic significance, and characterisation of the immune microenvironment in Philadelphia-negative acute lymphoblastic leukaemia

Na Li^{1,2}, Yang Hong^{1,2}, Ling Zhang^{1,2}, Aining Sun^{1,2,*}

¹ Department of Hematology, The First Affiliated Hospital of Soochow University, Jiangsu Institute of Hematology, National Clinical Research Center for Hematologic Diseases, Suzhou 215000, China

² Institute of Blood and Marrow Transplantation, Collaborative Innovation Center of Hematology, Soochow University, Suzhou 215000, China *** Corresponding author:** Aining Sun, aining sun@outlook.com

CITATION

Li N, Hong Y, Zhang L, Sun A. WGCNA-based identification of anoikis-related subtypes, prognostic significance, and characterisation of the immune microenvironment in Philadelphia-negative acute lymphoblastic leukaemia. Molecular & Cellular Biomechanics. 2024; 21: 90.

https://doi.org/10.62617/mcb.v21.90

ARTICLE INFO

Received: 22 March 2024 Accepted: 25 April 2024 Available online: 20 May 2024

COPYRIGHT



Copyright © 2024 by author(s). *Molecular & Cellular Biomechanics* is published by Sin-Chn Scientific Press Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. https://creativecommons.org/licenses/ by/4.0/ Abstract: The clinical outcomes and incidence of Philadelphia chromosome-negative B cell acute lymphoblastic leukaemia (ph-neg B-ALL) vary significantly across different age groups, influencing the prognosis. Despite recent advancements in diagnostic and therapeutic techniques, the detailed prognosis for ph-negative B-ALL across age demographics remains to be elucidated. In this study, clinical data were obtained from 80 patients with ph-neg B-ALL who were diagnosed at our centre. Ribonucleic acid sequencing was performed using their initial bone marrow aspirate samples. By employing weighted gene co-expression network analysis (WGCNA) on 408 anoikis-related genes (ARGs), four different modules were identified and subsequently analysed through bioinformatics. The WGCNA revealed distinct co-expression modules among ARGs. Specifically, the ARGs in the turquoise module might assess the risk associated with newly diagnosed ph-neg B-ALL. Additionally, the study revealed significant heterogeneity in the immune microenvironment and genome variance, highlighting the notable heterogeneity within the disease. 408 ARGs were screened out and four different co-expression modules were constructed by WGCNA algorithms from the RNAsequencing data of 80 ph-neg B-ALL patients; The ARGs in the turquoise module were the most, and it can be used to divide the de novo ph-neg B-ALL patients to different risk groups(high-risk and low-risk); The ph-neg B-ALL patients can be divided into PS-1 and PS-2, there is heterogeneity of genomes between PS-1 and PS-2; Immune infiltration difference exists in between PS-1 and PS-2. In conclusion, our study holds significant value in exploring the molecular pathways and mechanisms associated with anoikis implicated in ph-neg B-ALL, and in facilitating the development of treatments and prognostic tools for this disease

Keywords: ph-neg B-ALL; anoikis; WGCNA; immunmicroenvironment

1. Introduction

In recent years, significant progress has been made in the diagnosis and treatment of B cell acute lymphoblastic leukaemia (B-ALL). The advent of tyrosine kinase inhibitors has significantly improved the survival of patients with Philadelphia chromosome-positive B-ALL [1]. However, most B-ALL cases are negative in Philadelphia chromosome screening [2], and the prognosis of Philadelphia chromosome-negative B-ALL (ph-neg B-ALL) remains heterogeneous [3,4]. Chimeric antigen receptor T cells (CAR-T) that target B cell-lineage antigens such as cluster of differentiation (CD) 19, CD20, CD22, CD123, and B cell maturation antigens have shown promise in treating refractory or relapsed B-ALL. Nonetheless, the efficacy and safety of CAR-T therapy post-treatment remain significant concerns that researchers need to address immediately [5]. Therefore, it is crucial to explore new mechanisms that contribute to ph-neg B-ALL and to identify potential therapeutic targets for this disease.

Anoikis is a programmed cell death process that is activated by the loss of interaction between cells and the extracellular matrix (ECM) [6,7]. Cell-cell adhesion and interaction with the ECM are crucial for many essential cellular processes, such as migration and proliferation [8,9]. When cell-cell or cell-ECM attachments are lost, anoikis is triggered, resulting in a form of programmed apoptosis that eliminates misplaced or dislodged cells and helps maintain tissue homeostasis [10]. The induction of anoikis is primarily through two apoptotic pathways, i.e., interference with mitochondria or activation of cell surface death receptors [11,12]. Initially described in epithelial and endothelial cells, anoikis has been found to be an important mechanism in cancer invasion and metastasis [13]. Developing resistance to anoikis allows detached cells to evade cell death signalling pathways and survive under unfavourable conditions [14]. Numerous studies have reported that pyruvate dehydrogenase kinase 4 up-regulation is directly implicated in chemoresistance acquisition in lung cancer and promotes tumour cell proliferation in vivo and in vitro [15]. Additionally, the Nm23-integrin subunit alpha 5 pathway is essential for breast cancer cell invasion, and regulation of this pathway might prevent the establishment of breast cancer cell metastasis [16]. Despite the significant role of anoikis-related genes (ARGs) in tumourigenesis, tumour invasion, and tumour infiltration, few studies have systematically investigated the implications of anoikis in ph-neg B-ALL.

Although the complete understanding of anoikis remains unclear, several genes that are closely associated with anoikis progression have been identified by researchers. However, the specific role of anoikis in ph-neg B-ALL is not well established. Herein, we aimed to investigate the potential involvement of anoikis in ph-neg B-ALL by analysing ARGs previously reported in the literature [17]. Firstly, we aimed to evaluate the prognostic significance of ARGs in patients with ph-neg B-ALL using unsupervised clustering and perform bioinformatics-based analysis to reveal the mechanism of genetic and biological heterogeneity involved in anoikis. Then, we aimed to investigate whether the varying degrees of anoikis involvement is correlated with the immune microenvironment of leukaemia, given the emerging role of immune therapies in the field of cancer treatment.

2. Methods and materials

2.1. Sample collection and ethics approval

Ethics approval and written informed consent were obtained for the use of bone marrow aspirate samples from 80 B-ALL patients. Diagnostic confirmation of B-ALL was achieved through comprehensive evaluations including complete blood count, bone marrow aspirates analysis, flow cytometry, morphological examination, chromosomal analysis, and molecular biology techniques, ensuring accurate classification of leukemia subtype. All samples were initially screened for the Philadelphia chromosome status to determine inclusion in the ph-neg or ph-pos B-ALL groups. The presence or absence of the chromosome was verified using both chromosome R-banding technique and fluorescence in situ hybridization (FISH),

ensuring accurate classification of each patient's condition, confirming the status of Ph-like ALL in the Ph-negative group, which is recognized as a distinct entity in the current ALL classifications. Prior to sample collection, the stage of disease and the clinical treatment protocols followed were documented for each patient to aid in subsequent analysis of survival data and treatment response.

To avoid selection bias, the study was designed meticulously with clear inclusion and exclusion criteria. Randomization methods such as simple randomization, block randomization, and stratified randomization were used to ensure balanced comparison groups aside from the study factor. Multiple control groups were established to represent different population types. Efforts were made to enhance response rates and minimize follow-up losses, including random sampling where necessary. Diagnostic criteria were strictly adhered to throughout the research and implementation phases, and data were collected using a blinded approach to minimize information bias. This retrospective study enrolled 80 patients diagnosed with acute B-lymphoblastic leukemia between October 2015 and January 2021. Inclusion criteria included newly diagnosed cases of B-ALL without prior treatment. Exclusion criteria comprised previous leukemia treatment, co-existing chronic diseases that could interfere with the study outcomes, and incomplete medical records. These criteria were established to ensure the collection of high-quality and relevant data for the study's objectives.

2.2. Whole transcriptome sequencing (ribonucleic acid sequencing [RNA-seq]) and data processing

For whole transcriptome sequencing, ribonucleic acid (RNA) from bone marrow aspirates was sequenced to generate comprehensive transcriptome data. This process, referred to as whole transcriptome sequencing or RNA-seq, was performed according to methodologies detailed in previously published references[18].

2.3. Targeted gene mutational analysis

The analysis of the method was in the previously published article[18].

2.4. Oncomine analysis

We also refer the reference previously published [18].

2.5. Construction of the weighted co-expression network based on ARGs

Following the establishment of gene co-expression modules, external validation was performed using an independent dataset of B-ALL patients to confirm the reproducibility and robustness of the identified modules. This step involved recalculating module preservation statistics, which verified that our findings were consistent and not due to random chance or specific to a single dataset. We used the weighted gene co-expression network analysis (WGCNA) method to construct the gene co-expression network and identify the functional modules [19]. For normalization, we applied quantile normalization to the RNA-seq data to ensure consistent distribution across samples, essential for accurate comparison and analysis. To safeguard against biases from technical artifacts or tissue contamination, comprehensive preprocessing was conducted to remove low-quality reads, followed

by batch effect correction using the ComBat method, ensuring our analysis reflects true biological signals. We conducted hierarchical clustering analysis using the R software package ggplots to identify sample outliers [20]. The WGCNA algorithm was subsequently employed to calculate the correlation between ARGs and the samples. Following this, we determined an appropriate soft threshold power and created a standard scale-free network. The correlation functional networks were then used to identify tissue-specific markers. Using a dynamic tree-cutting strategy, we identified the modules by clustering the genes hierarchically with a minimum module size of 30 and a deepSplit value of 2 to prevent over-splitting and preserve meaningful biological information. The sensitivity of the module detection was further fine-tuned by setting a cut height of 0.25, based on the dynamic tree cut method, ensuring precise module division. Using a dynamic tree-cutting strategy, we identified the modules by clustering the genes hierarchically. We used the WGCNA to establish gene co-expression modules and extract the gene information from each module. We created four different modules, namely, blue, brown, yellow, and turquoise, with the turquoise module comprising the most ARGs. Therefore, our subsequent analysis was primarily focused on the turquoise module. To select the ARGs, we initially performed an extensive literature review to identify genes implicated in anoikis resistance across various cancers, with a specific focus on hematological malignancies. This was complemented by bioinformatics analyses where we screened gene expression datasets for genes differentially expressed in B-ALL versus normal hematological samples. The robustness of our ARG model was ensured through the application of WGCNA, which not only facilitated the identification of co-expression modules correlated with B-ALL but also allowed for the validation of these modules in an independent patient dataset. This step was crucial in demonstrating the generalizability of our findings across different cohorts, thereby enhancing the confidence in our model's predictive power and stability.

2.6. Enrichment analysis of the turquoise modules and identification of hub ARGs

We conducted functional enrichment analysis on the functionally related modules to analyse, identify, and interpret various biological functions based on the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis and Gene Ontology (GO) annotation analysis. The ClusterProfiler and ggplot2 R packages were used for data visualisation and analysis (P-value cut-off, 0.05).

2.7. K-means clustering

The Scikit-learn package V0.24.2 in Python V3.8 was used to perform K-means clustering, which is a widely used algorithm in unsupervised machine learning. Before clustering, the FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values of 408 ARGs in 80 ph-neg B-ALL samples were standardised to [0,1] to eliminate the influence of dimension and variation range. Thereafter, the dimension was reduced from 24 to two using principal component analysis (PCA) in the turquoise module. The optimal K-value was determined using the "elbow" method, and 80 samples were grouped into two (PS-1 and PS-2) based on the K-means clustering

results. We used the adjusted Rand index, the adjusted mutual index, the V-measure score, the Fowlkes–Mallows index, the Silhouette coefficient, and the Calinski–Harabaz index to assess the clustering models.

2.8. Protein-protein interaction [21] network construction for selected modules and hub gene identification

We used the cytoHubba plugin based on Cytoscape to identify highly connected hub genes in the PPI network. To begin, we measured the module membership (MM) of each gene by Pearson's correlation with the module Eigengenes. In this study, we focused on selecting hub genes with an MM of >0.55 in the specific module.

2.9. Immune characteristics of the sample clusters

We used the CIBERSORTx algorithm to investigate the variance of the immune infiltration between the clusters and process the RNA-seq data [22]. This allowed us to calculate the infiltration levels of 65 variant immune cells, and we only included results with a significant p-value of <0.05 for further analysis. We also employed Kmeans clustering to divide the 80 qualified samples into five components based on their immune infiltration results to demonstrate the feature differences among the three groups with variant degrees of risk. We chose a K-value of two for clustering to exclude immune cells with extremely low infiltration levels and to more clearly demonstrate the differences in immune cell infiltration between the groups. We further used the ESTIMATE algorithm to estimate the purity of the tumour, and the R package "estimate" for storm and the immune cell ratio, following which the ESTIMATE score was used to evaluate the immune statement of leukaemia's micro-environment [23]. The cytolytic score generated from the average log10 value of five granzyme and perforin-1 gene expressions, as well as the inflammatory score calculated in previous reports [24,25], were assembled to assess the cytotoxic immune cell activity. The response to programmed cell death protein 1 (PD-1) blockage therapy was estimated based on a previously reported formula [26]. In this study, the median follow-up period was 23.5 months. During this time, follow-up visits were scheduled every three months to monitor disease progression and patient survival, allowing for a comprehensive assessment of long-term outcomes.

3. Statistical analysis

The statistical analyses were completed using R software (v4.0.2), GraphPad Prism8.0 (GraphPad Inc., San Diego, CA, USA), and Python software (v3.8). We then administered a two-way Student t-test to perform a numerical comparison and a chi-square test to analyse the categorical data to compare the clinical and molecular parameters between groups. Univariate Cox regression was performed to identify factors with independent prognostic value, while a multivariate Cox regression model was established as previously described. Statistical significance was set at P < 0.05 and was presented as *P < 0.05, **P < 0.01, or ***P < 0.001.

4. Results

4.1. Recognition of co-expression modules in ARGs

In this study, we included 408 ARGs to construct a weighted co-expression network (**Figure 1**). The WGCNA algorithm was used to cluster and construct gene modules. After observing that a soft threshold of three was acceptable in the scale-free network, (**Figure 2a, b**) we converted the representation matrix to adjacency and then to a topological matrix to build the co-expression network. Using the average-linkage hierarchy clustering method with a minimum of five genes for each network to restrict the hybrid dynamic shear tree with the cut height of 0.8, we obtained the ARGs clustered in each module and visualised the expression correlation as a TOM plot (**Figure 2c**). We identified four gene modules, while the genes that failed to cluster into any modules were excluded. Remarkably, the module marked in turquoise comprised the most ARGs (n = 68), significantly more than the blue (n = 14), brown (n = 9), and yellow (n = 5) modules (**Figure 2d**).



Figure 1. Flow diagram of the preparation, processing, analysis, and validation of data.



Figure 2. Determination of the soft threshold power in the weighted gene co-expression network analysis (WGCNA) and different identification modules with anoikis-related genes (ARGs). Hierarchical cluster tree showing co-expression modules based on the WGCNA. (a) Analysis of the scale-free index for various soft threshold powers; (b) Analysis of the mean connectivity for various soft threshold powers; (c) Co-expression heat map and correlations for ARGs in the modules; (d) Four different modules were identified based on the number of ARGs.

4.2. The turquoise module was involved in the apoptotic signalling pathway and focal adhesion

In this study, we identified a module of 68 ARGs, which was marked in turquoise. We hypothesised that these genes might work together in specific biological processes related to anoikis. We performed a gene enrichment analysis using the GO and KEGG databases to further explore this. Our analysis revealed that the biological processes (BPs) most significantly enriched in the turquoise module were regulation of the apoptotic signalling pathway and extrinsic apoptotic signalling pathway (**Figure 3a**). In terms of cellular components, focal adhesion and cell-substrate junction were among the enriched pathways, which are crucial in anoikis (**Figure 3b**). Tumour cells tended to generate functional molecules to resist the attack of the body's immune

system to realise the resistance to anoikis. Further, our analysis indicated that the genes in the turquoise module might help the leukaemia cells resist anoikis through enzyme inhibitor activity and protein serine/threonine kinase activity (**Figure 3c**). We also found that the KEGG pathway associated with the turquoise module included proteoglycans in cancer and apoptosis (**Figure 3d**).



Figure 3. Gene Ontology (GO) and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment of ARGs in the turquoise module. (a) The GO pathway enrichment-biological process; (b) The GO pathway enrichment-cellular components; (c) The GO pathway enrichment-molecular functions; (d) The KEGG pathway enrichment of ARGs in the turquoise module (|logFC| > 2, false discover rate < 0.05).

4.3. ARGs in the turquoise module helped distinguish the risk of de novo ph-neg B-ALL

The involvement of the turquoise module in anoikis raises the question of whether there is a correlation between patient survival and the gene expression in this module. We used PCA dimensionality reduction to represent module expression in each sample using two factors (PCA-1 and PCA-2) to consolidate the expression of the 68 genes in the turquoise model and represent the module expression in each sample. The genes significantly correlated with PCA factors were depicted in **Table 1**. Using the elbow method, we determined that K = 2 was an acceptable grouping number (**Figure 4a**) and applied the K-means clustering algorithm to differentiate patients into two clusters (PS-1 and PS-2, **Figure 4b**). Interestingly, we observed a significant difference in overall survival between the two clusters (P = 0.0399, **Figure 4c**).

Although co-expressed genes are likely to have a cooperative effect on anoikis, it is known that the interactions among functional proteins are complex. Therefore, within a limited gene module, a functional gene's importance might depend on the number of connections it has with other genes. We used the STRING database to construct a PPI network with median confident evidence to explore this idea. We then used Cytoscape software and the HCC algorithm in cytoHubba to calculate the scores of each gene and identified the top 10 hub genes (**Figure 4d**). Given that the module gene-based PCA could distinguish patients with varying risks, it was important to investigate the relationship between hub gene expression and prognosis. Using a single-factor log-rank test, we discovered that high expression levels of all 10 genes were protective factors. Additionally, the five hub gene (CASP8, MCL1, NOTCH1,

STAT3, FOXO3) expressions were significantly associated with patient survival (Figure 4e). To compare the proposed gene module classification with established prognostic systems, we analyzed the survival data of the two clusters identified by the turquoise module with those classified by traditional prognostic markers. Our analysis revealed that while traditional prognostic markers focus primarily on genetic and molecular abnormalities, the turquoise module incorporates additional biological processes such as anoikis resistance, providing a more comprehensive insight into patient prognosis. The comparison data are presented in Table 1, illustrating that the turquoise module classification aligns well with established systems but offers enhanced predictive power for patient survival, especially in high-risk groups. Regarding the differences in survival data between previously published studies (as cited in reference [18]) and the current dataset, there are indeed variations observed. The differences stem primarily from the updated gene expression profiles used in this study, which include a broader range of anoikis-related genes. These variations are detailed in Figure 4, where survival analyses demonstrate that our current dataset shows a more distinct separation in survival curves between the two clusters, highlighting the robustness of the turquoise module in predicting clinical outcomes. This variance underscores the importance of continual updates to gene expression data and analytical methods in improving prognostic accuracy.

Id	<i>P</i> .value	PCA-1_corr	PCA-2_corr
CXCL8	0.015	0.272	
CEBPB	0.045	0.225	
PTGS2	0.013		0.278
CFLAR	0.005		-0.312
THBS1	0.037		-0.233
RHOG	0.002		-0.342
FAS	0.001		-0.376
S100A4	0.019		-0.262
SERPINA1	0.003		-0.332
CEBPB	0.000		-0.424
SDCBP	0.012		-0.280
SIRPA	0.024		-0.253
TNFRSF12A	0.020		-0.259
PXN	0.035		-0.236

Table 1. The genes significantly correlated with PCA factors.



Figure 4. Eighty patients with Philadelphia negative (ph-neg) B cell acute lymphoblastic leukaemia (B-ALL) were clustered into two groups by K-means clustering based on the anoikis-related genes (ARGs) and construction of the protein-protein interaction [21] network based on the turquoise module. (a) K=2 was an ideal K-value; (b) Eighty dots representing the enrolled ph-neg B-ALL patients after principal component analysis dimension reduction were located on a two-dimensioned plane and scattered into two groups based on the labels of the K-means clustering result; (c) The Kaplan–Meier analysis performed on the two clusters is demonstrated, with a significant difference between the two clusters (P=0.0399); (d) The PPI network was constructed based on the STRING databases; (e) Ten hub genes were identified and their relationship with prognosis was calculated. (f) Genetic characteristics of the 80 patients with ph-neg B-ALL; (g, h) The variance of the mutated genes of patients between the PS-1 and PS-2 groups; (i) Comparison of the top four mutated genes between the PS-1 and PS-2 groups.

4.4. The heterogeneity of genomes in PS-1 and PS-2

Nowadays, it is widely acknowledged that genomic variations play a crucial role in the development and progression of leukaemia. We initially gathered the nextgeneration sequencing (NGS) records of the included patients to investigate the genomic differences between the patient clusters. Of the 172 leukaemia recurrent mutated genes (**Table 2**), 65 were found to be positive in our samples (**Figure 4f**). Furthermore, we assessed the number of mutated genes in each sample and discovered that the patients in PS-1 had significantly more mutated genes than those in PS-2 (P = 0.0293, **Figure 4g**). In addition, PS-1 and PS-2 shared 13 identical mutated genes. In contrast to only one distinct mutated gene in PS-2, PS-1 had 51 distinct mutated genes, indicating a high degree of heterogeneity in PS-1 (**Figure 4h**). Due to the limited sample size, most genes barely recurred for further comparison. Therefore, we selected the top four mutated genes (neuroblastoma RAS viral oncogene homolog [NRAS], Kirsten rat sarcoma virus [KRAS], SET domain containing 2 [SETD2], and tumour protein p53 [TP53]) to demonstrate the difference between the clusters. As a result, the ratio of NRAS and KRAS mutation was higher in PS-1 than in PS-2, whereas the ratio of SETD2 and TP53 mutation did not significantly differ (**Figure 4i**).

Table 2. Gene mutation detection patients	anel
---	------

ABL1	BRAF	CD79B	CXCR4	EZH2	IKZF3	MED12	PDGFRA	RELN	STAG2	TP63
ANKRD2 6	BRCA1	CDKN1A	DDX3X	FAM46C	IL7R	MEF2B	PDGFRB	RHOA	STAT1	TP73
APC	BRCA2	CDKN2A	DDX41	FAT1	IRF4	MPL	PHF6	RPS15	STAT2	TRAF3
ARID1A	BRCA3	CDKN2B	DHX15	FBXW7	JAK1	MUM1	PIGA	RUNX1	STAT3	TYK2
ARID1B	BTG1	CDKN2C	DIS3	FLT3	JAK2	MYC	PIK3CA	SAMHD1	STAT4	U2AF1
ARID2	BTK	CEBPA	DNM2	FOXO1	JAK3	MYD88	PIM1	SETBP1	STAT5A	UNC13 D
ARID5B	CALR	CEBPE	DNMT3A	GATA1	KDM6A	NF1	PLCG1	SETD2	STAT5B	WT1
ASXL1	CARD11	CHD8	ECT2L	GATA2	KIT	NFKB1	PLCG2	SF3B1	STAT6	XPO1
ASXL2	CBL	CIITA	EED	GATA3	KLF2	NFKB2	PPM1D	SH2B3	SUZ12	ZAP70
ATM	CCND1	CRBN	EGFR	GNA13	KMT2A	NFKBIE	PRDM1	SMC1A	TBL1XR1	ZBTB7 A
B2M	CCND2	CREBBP	EGR2	HRAS	KMT2C	NOTCH1	PTEN	SMC2	TCF3	ZMYM3
BCL2	CCND3	CRLF2	EP300	ID3	KMT2D	NOTCH2	PTPN11	SMC3	TET1	ZRSR2
BCL6	CD274	CSF1R	EPOR	IDH1	KRAS	NPM1	PTPRD	SOS1	TET2	
BCOR	CD28	CSF3R	ERBB3	IDH2	MAP2K1	NRAS	RAD21	SPEN	TNFAIP3	
BCORL1	CD58	CSMD1	ETNK1	IGLL5	MAPK1	PAX5	RAF1	SRP72	TNFRSF14	
BIRC3	CD79A	CTCF	ETV6	IKZF1	MECOM	PDCD1LG2	RB1	SRSF2	TP53	

4.5. Multiple essential signalling pathways were enriched with differentially expressed genes (DEGs) from the comparison of the two clusters

Using the edgeR algorithm, we identified 939 DEGs ($\log 2|FC| > 2$, P < 0.05, **Figure 5a**) from these two groups. These DEGs were indicative of biological function variance; thus, we conducted the gene set enrichment analysis (GSEA) with the KEGG and GO databases. The results revealed that neutrophil extracellular trap formation was enriched in KEGG pathways, indicating a potential influence of the immune microenvironment on patient prognosis (**Figure 5b**). Additionally, several GO terms were enriched, including cell activation in immune response, immune effector process, and myeloid activation involved in immune response, which supported the immune microenvironment difference related to anoikis (**Figure 5c**). Notably, secretory granule and secretory granule membrane of cell components (CCs) were also enriched in GO terms (**Figure 5d**). Furthermore, we integrated the different genes expressed in the DEGs, ARGs, and turquoise module using PCA and identified 13 co-expressed genes (**Figure 5e**). Besides, the Log-rank test of thirteen key DEGs was displayed in **Table 3**. The results indicated that genes (CFLAR, RHOG, S100A4 and TNFRSF12A) were of significance. These 13 genes were enriched in different GO terms (BPs, CC, and molecular functions), as presented in (**Figure 5f**). We then focused on the gene caspase 8 And Fas-associated via death domain like apoptosis regulator (CFLAR) and investigated its expression and relationship with patient survival. The results demonstrated that higher CFLAR expression was associated with better patient survival (**Figure 5g**). This finding suggests that the prognosis of patients with ph-neg B-ALL may correlate with the expression of ARGs expression.

Genes	Hazard Ratio	95% CI	<i>p</i> .value
CXCL8	0.9968	(0.9921,1.002)	0.186
PTGS2	0.9566	(0.9104,1.005)	0.0794
CFLAR*	0.8361	(0.7034,0.9937)	0.0422
THBS1	0.9728	(0.879,1.077)	0.594
RHOG*	0.9784	(0.9579,0.9994)	0.0438
FAS	0.8327	(0.6824,1.016)	0.0715
S100A4*	0.9941	(0.9886,0.9995)	0.0335
SERPINA1	0.9826	(0.9633,1.002)	0.0843
CEBPB	0.9928	(0.9852,1)	0.0665
SDCBP	0.9689	(0.9353,1.004)	0.0797
SIRPA	0.952	(0.9022,1.005)	0.0727
TNFRSF12A*	0.5287	(0.2806,0.9962)	0.0486
PXN	0.9469	(0.8962,1)	0.052

Table 3. Log-rank test to thirteen key DEGs.



Figure 5. Identification of the differentially expressed overlapping anoikis-related genes (ARGs) in the turquoise module based on the principal component analysis (PCA) method. The differentially expressed genes (DEGs) were sorted and the Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment with these DEGs between the two clusters was exhibited. (a) A total of 939 DEGs were sorted using the edgeR algorithm ($\log_2|FC| > 2$, P<0.05). Yellow dots indicate statistically significant up-regulated genes and blue dots represent statistically significant down-regulated genes. The x-axis represents the adjusted P-value based on the false discovery rate correction method, and the y-axis represents the logarithm of the fold change of DEGs. (b, c, d) KEGG

and GO analyses of the DEGs. (e) Venn diagram of overlapping genes in DEGs and ARGs of the turquoise module based on the PCA method. (f) The 13 overlapping genes are distributed to different GO terms and categories (biological processes, cellular components, and molecular functions) based on the PCA method. (g) The graph of the survival curve of the overlapped gene caspase 8 And Fas-associated via death domain like apoptosis regulator is presented. P=0.023, hazard ratio=0.350.

4.6. Infiltrated immune cell difference among subgroups

"In addition to differences in immune status between PS-1 and PS-2, variations were also observed in immune infiltration, including the infiltration of immune cells and stromal cells within the tumor microenvironment. Immune cells, primarily consisting of lymphocytes, macrophages, and dendritic cells, play crucial roles in tumor defense and immunoregulation. Stromal cells, including fibroblasts, endothelial cells, and mesenchymal stem cells, contribute to the structural and supportive framework of the tumor microenvironment. We further investigated the detailed infiltration patterns of these cells between the two subgroups using the CIBERSORTx method to analyze 64 common immune cells in 80 variant samples. This analysis helped us understand how the interactions between immune and stromal cells can influence tumor behavior and patient prognosis, there were also differences in immune infiltration. We further investigated the immune cell infiltration between the two subgroups using the CIBERSORTx method to analyse 64 common immune cells in 80 variant samples (Figure 6a). Specifically, after comparing the "PS-1" and "PS-2" groups, we found that granulocyte-macrophage progenitors (GMPs), multipotent progenitors (MPPs), common myeloid progenitors, osteoblasts, type 1 T helper cells, M2 macrophages, memory B cells, B cells, naive B cells, mesenchymal stem cells (MSCs), common lymphoid progenitors (CLP), and pro-B cells were enriched in the "PS-1" group. However, erythrocytes, macrophages, gamma delta T cells, megakaryocytes, platelets, melanocytes, sebocytes, eosinophils, monocytes, neutrophils, natural killer T cells, epithelial cells, and mesangial cells were enriched in the "PS-2" group. Besides, the enrichment score of different immune cells between two groups was indicated in (Figure 6b). We also calculated the relationship between the six ARG expressions and the infiltration degree of 10 immune cell types using the Pearson correlation coefficient to demonstrate the correlation between the anoikis and immune cell infiltration (Figure 6d). The results indicated that the expression of these genes was found positively correlated with the infiltration of the B cells, CLPs, GMPs, MPPs, MSCs, osteoblasts, and pro-B cells, while negatively correlated with the infiltration of eosinophils, monocytes, and neutrophils. As immunotherapy has become a focus of cancer treatment in recent years, investigating the tumour microenvironment (TME) is of great significance. The "ESTIMATE" R package is an ideal method for describing the immune status of cancers based on gene transcriptome data. After processing the gene expression data of 80 samples with the "ESTIMATE" algorithm, the stromal, immune, and estimate scores were calculated. Compared with the "PS-1" group, the "PS-2" group had significantly higher stromal, immune, and estimate scores (Figure 6c). These results suggest that the "PS-1" group had relatively fewer immune cells in the TME, which benefited the survival of leukaemia. Furthermore, we estimated the function of T cells attacking against leukaemia with

cytolytic and inflammatory scores and found differences in cytolytic scores between the "PS-1" and "PS-2" groups (**Figure 6e**), suggesting that the dysfunction of T cells in the TME might play an important role in the relatively poor prognosis of the "PS-1" group. Conversely, the human leukocyte antigen [27] expression was often reduced in cancers to escape immune surveillance. In our study, the HLA expression (including HLA-A, B, and C) was relatively lower in the "PS-1" group compared with the "PS-2" group (**Figure 6g**), which was associated with HLA-downregulation in promoting tumor immune escape and tumor development. Immune checkpoint inhibitors (ICIs) have been recognised as a promising therapy in solid tumours. However, their role in B-ALL remains unknown. Referring to the interferon- γ signature, there were significant differences in the clinical response to PD-1 blockade between these two groups (**Figure 6f**), indicating that there might be potential benefits from PD-1 blockage therapy in the "PS-1" group compared with the "PS-2" group.





5. Discussion

In this study, we retrospectively analysed clinical data from 80 patients with phneg B-ALL who were treated at our centre. Using the WGCNA algorithm, we constructed co-expression modules with 408 ARGs and investigated the relationship between infiltrating immune cells and prognostic genes and risk scores. Through the WGCNA, we identified four gene modules (the brown, blue, yellow, and turquoise modules), with the latter containing the highest number of ARGs (n = 68). We performed GO and KEGG pathway enrichment analyses to gain insight into the complex mechanisms of the turquoise module [28]. Our findings indicated that the turquoise module is involved in the apoptosis signalling pathway and focal adhesion. We also assessed whether there was a correlation between patient survival and the turquoise module's gene expression. The results revealed that the 68 ARG expressions in the turquoise module could be used to cluster these patients into two groups (PS-1 and PS-2) using the unsupervised machine learning algorithm K-means clustering [29]. Additionally, we constructed a PPI network to identify 10 hub genes with high scores and evaluated their association with patient prognosis. Remarkably, elevated expression levels of these ten genes were found to be protective factors. Consistent with previous studies, we also observed significant genomic heterogeneity between the PS-1 and PS-2 groups, which was recognised as a driving factor in the development and progression of leukaemia. Furthermore, we analysed RNA-seq data from 80 patients with ph-neg B-ALL and identified 939 DEGs. Using GSEA with KEGG and GO databases, we integrated the DEG, ARG, and the turquoise module gene sets [30] and identified 13 co-expressed genes enriched in various GO terms. Among these, CLAFR might play a crucial role in patient survival. Moreover, we found a significant correlation between the immune microenvironment and ARG-based clustering. Based on these findings, we postulate that anoikis might play a crucial role in the onset and progression of ph-neg B-ALL [31].

Anoikis is a type of programmed cell death that occurs when there is a loss of interaction between the cell and ECM [32]. Numerous studies have demonstrated that anoikis is involved in tumourigenesis and plays a crucial role in cancer therapy [33]. However, the role of anoikis in ph-neg B-ALL has not been extensively studied, and the outcomes of ph-neg B-ALL treatment vary significantly due to the lack of targeted therapies like dasatinib in Philadelphia positive B-ALL [34]. Herein, we employed the NGS technology to explore the heterogeneity of ph-neg B-ALL at the genetic and transcriptomeal levels using 68 ARGs [35]. Our results demonstrated that the expression of ARGs in B-ALL was significantly different from that in normal individuals, and the combination of ARG expression could classify ph-neg B-ALL into different risk groups using K-means clustering. These findings provided clinical evidence that the progression of leukaemia is closely related to the resistance of anoikis [36].

Our attention was drawn to investigating the leukaemia immune microenvironment in light of the potential of immunotherapy [37]. However, the role of the anoikis mechanism in the leukaemia immune microenvironment remains unclear and warrants further research. Our findings indicate that there are differences in immune microenvironment aberration and anoikis between the "PS-1" and "PS-2"

groups. Specifically, the "PS-1" group showed higher infiltration of naive B cells, lower CD4+ memory-activated T cells [38], and an outstanding polarisation of macrophages to an M2 phenotype. This phenomenon might contribute to anoikis and ultimately stimulate leukaemia cell growth. Furthermore, our research suggests that PD-1 blockade therapy might provide greater benefits in the "PS-1" group compared with the "PS-2" group, which is an important finding given the limited knowledge of the role of ICIs in B-ALL [39]. These results raise the possibility of implementing anoikis inducers in the "PS-1" group, which is characterised by low immune cell infiltration. This is because a leukaemia microenvironment with low immune cell infiltration facilitates leukaemia cells to evade immune surveillance [40], and the inflammatory environment induced by anoikis inducers has a chemotaxis effect on immune cells and transforms "cold" leukaemia into "hot" leukaemia, making it more susceptible to chemotherapy.

In line with other models examining the correlation between anoikis and various cancer types [41,42], we systematically investigated anoikis involvement in our samples and confirmed its important yet largely unexplored role in ph-neg B-ALL. Despite the success of imatinib in improving the prognosis of patients with Philadelphia positive B-ALL, a considerable number of patients with ph-neg B-ALL show low sensitivity to standard therapies, particularly in adult patients. Additionally, the lack of a prognostic evaluation system has hampered the development of Individualised treatment options. Using 68 ARGs in the turquoise module, we employed a series of bioinformatics analyses to predict the prognosis of patients with ph-neg B-ALL.

Based on our findings, further validation of our results on DNA variance and RNA level quantifications for ARGs in a larger and prospective cohort is necessary. Moreover, as anoikis is a newly discovered and complex BP, it requires further investigation. In addition, the limited number of cases in our study, especially with the introduction of CAR-T therapy, underscores the importance of exploring the role of anoikis in patients with ph-neg B-ALL patients in the new era. Thus, our current knowledge and data suggest the need for updates and further validation in future research. One of the primary limitations of this study is its reliance solely on bioinformatics and computational data analysis. While these approaches are instrumental in uncovering potential genetic interactions and pathways implicated in B-ALL, they lack the empirical validation that can only be provided through direct laboratory experiments and clinical trials. Future research will focus on validating the predictive markers identified in this study through in vitro and in vivo experiments, as well as clinical studies to confirm their functional significance and relevance to patient care. Such steps are crucial for bridging the gap between theoretical predictions and practical, clinical applications.

6. Conclusion

Altogether, this study defined four different co-expression modules and divided ph-neg B-ALL patients samples into two different subgroups according to ARGs by WGCNA algorithms. The correlation of ARGs with leukemic immune microenvironment and genomic mutations was also investigated. The results showed that the ARGs were negatively correlated with the occurrence of tumor and immune cell infiltration. Overall, our findings have great significance for investigating the molecular pathways and mechanisms associated with ARGs involved in ph-neg B-ALL and for developing treatments and prognoses of ph-neg B-ALL.

Author contributions: Contributed to the concept development, study design, wrote the manuscript, collected the clinical data, followed up the enrolled patients, performed the data check and analysis, NL and YH; participated in the figure preparation, revision of the paper, LZ and AS. All authors have read and agreed to the published version of the manuscript.

Data availability statement: The data that support the findings of this study are openly available in reference [18] at https://doi.org/10.1186/s12885-021-09076-w.

Funding: This work was supported by grants from the National Natural Science Foundation of China (grant no. 81873449). The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Ethical approval: This study was performed in accordance with the Declaration of Helsinki and approved by the Ethics Committee and Institutional Review Board of The First Affiliated Hospital of Soochow University, Suzhou, China, with a reference number of 2021–105. And written informed consent forms signed by the patients were achieved before this research. The Ethics Committee and Institutional Review Board of The First Affiliated Hospital of Soochow University can be reached at sdfyec@163.com.

Conflict of interest: The authors declare no conflict of interest.

References

- 1. Abou Dalle I, Jabbour E, Short NJ, et al. Treatment of Philadelphia Chromosome-Positive Acute Lymphoblastic Leukemia. Current Treatment Options in Oncology. 2019; 20(1). doi: 10.1007/s11864-019-0603-z
- Burmeister T, Schwartz S, Bartram CR, et al. Patients' age and BCR-ABL frequency in adult B-precursor ALL: a retrospective analysis from the GMALL study group. Blood. 2008; 112(3): 918-919. doi: 10.1182/blood-2008-04-149286
- 3. Jabbour E, Richard-Carpentier G, Sasaki Y, et al. Hyper-CVAD regimen in combination with ofatumumab as frontline therapy for adults with Philadelphia chromosome-negative B-cell acute lymphoblastic leukaemia: a single-arm, phase 2 trial. Lancet Haematol. 2020; 7(7): 523-533.doi:10.1016/s2352-3026(20)30144-7
- 4. Parikh SA, Litzow MR. Philadelphia chromosome-negative acute lymphoblastic leukemia: therapies under development. Future Oncology. 2014; 10(14): 2201-2212. doi: 10.2217/fon.14.81
- 5. Hong M, Clubb JD, Chen YY. Engineering CAR-T Cells for Next-Generation Cancer Therapy. Cancer Cell. 2020; 38(4): 473-488. doi: 10.1016/j.ccell.2020.07.005
- 6. Adeshakin FO, Adeshakin AO, Afolabi LO, et al. Mechanisms for Modulating Anoikis Resistance in Cancer and the Relevance of Metabolic Reprogramming. Frontiers in Oncology. 2021; 11. doi: 10.3389/fonc.2021.626577
- 7. Kim YN, Koo KH, Sung JY, et al. Anoikis Resistance: An Essential Prerequisite for Tumor Metastasis. International Journal of Cell Biology. 2012; 2012: 1-11. doi: 10.1155/2012/306879
- 8. Frisch S, Francis H. Disruption of epithelial cell-matrix interactions induces apoptosis. The Journal of cell biology. 1994; 124(4): 619-626. doi: 10.1083/jcb.124.4.619
- 9. Frisch SM, Ruoslahti K. Integrins and anoikis. Current Opinion in Cell Biology. 1997; 9(5): 701-706. doi: https://doi.org/10.1016/S0955-0674(97)80124-X
- 10. Han H jun, Sung JY, Kim SH, et al. Fibronectin regulates anoikis resistance via cell aggregate formation. Cancer Letters.

2021; 508: 59-72. doi: 10.1016/j.canlet.2021.03.011

- 11. Amoedo ND, Rodrigues MF, Rumjanek FD. MITOCHONDRIA: Are mitochondria accessory to metastasis? The International Journal of Biochemistry & Cell Biology. 2014; 51: 53-57. doi: 10.1016/j.biocel.2014.03.009
- 12. Zhong X, Rescorla FJ. Cell surface adhesion molecules and adhesion-initiated signaling: Understanding of anoikis resistance mechanisms and therapeutic opportunities. Cellular Signalling. 2012; 24(2): 393-401. doi: 10.1016/j.cellsig.2011.10.005
- Kakavandi E, Shahbahrami R, Goudarzi H, et al. Anoikis resistance and oncoviruses. Journal of Cellular Biochemistry. 2017; 119(3): 2484-2491. doi: 10.1002/jcb.26363
- 14. Di Micco R, Krizhanovsky V, Baker D, et al. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. Nature Reviews Molecular Cell Biology. 2020; 22(2): 75-95. doi: 10.1038/s41580-020-00314-w
- 15. Yu S, Li Y, Ren H, et al. PDK4 promotes tumorigenesis and cisplatin resistance in lung adenocarcinoma via transcriptional regulation of EPAS1. Cancer Chemotherapy and Pharmacology. 2020; 87(2): 207-215. doi: 10.1007/s00280-020-04188-9
- 16. Wong AW, Paulson QX, Hong J, et al. Alcohol promotes breast cancer cell invasion by regulating the Nm23-ITGA5 pathway. Journal of Experimental & Clinical Cancer Research. 2011; 30(1). doi: 10.1186/1756-9966-30-75
- 17. Gilmore AP. Anoikis. Cell Death & Differentiation. 2005; 12(S2): 1473-1477. doi: 10.1038/sj.cdd.4401723
- 18. Hong Y, Zhang L, Tian X, et al. Identification of immune subtypes of Ph-neg B-ALL with ferroptosis related genes and the potential implementation of Sorafenib. BMC Cancer. 2021; 21(1). doi: 10.1186/s12885-021-09076-w
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9(1). doi: 10.1186/1471-2105-9-559
- 20. Karim MR, Beyan O, Zappa A, et al. Deep learning-based clustering approaches for bioinformatics. Briefings in Bioinformatics. 2020; 22(1): 393-415. doi: 10.1093/bib/bbz170
- 21. Dourthe ME, Rabian F, Yakouben K, et al. Determinants of CD19-positive vs CD19-negative relapse after tisagenlecleucel for B-cell acute lymphoblastic leukemia. Leukemia. 2021; 35(12): 3383-3393. doi: 10.1038/s41375-021-01281-7
- 22. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature Methods. 2015; 12(5): 453-457. doi: 10.1038/nmeth.3337
- 23. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature Communications. 2013; 4(1). doi: 10.1038/ncomms3612
- 24. Rooney MS, Shukla SA, Wu CJ, et al. Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. Cell. 2015; 160(1-2): 48-61. doi: 10.1016/j.cell.2014.12.033
- 25. Wu F, Li G, Liu H, et al. Molecular subtyping reveals immune alterations in IDH wild-type lower-grade diffuse glioma. The Journal of Pathology. 2020; 251(3): 272-283. doi: 10.1002/path.5468
- 26. Ayers M, Lunceford J, Nebozhyn M, et al. IFN-γ–related mRNA profile predicts clinical response to PD-1 blockade. Journal of Clinical Investigation. 2017; 127(8): 2930-2940. doi: 10.1172/jci91190
- 27. Lonial, Weiss, Usmani et al. Daratumumab monotherapy in patients with treatment-refractory multiple myeloma (SIRIUS): an open-label, randomised, phase 2 trial. The Lancet. 2016; 387(10027): 1551-1560. doi:10.1016/S0140-6736(15)01120-4
- 28. Chen L, Zhang YH, Wang S, et al. Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. Liu B, ed. PLOS ONE. 2017; 12(9): e0184129. doi: 10.1371/journal.pone.0184129
- 29. Eckhardt CM, Madjarova SJ, Williams RJ, et al. Unsupervised machine learning methods and emerging applications in healthcare. Knee Surgery, Sports Traumatology, Arthroscopy. 2022; 31(2): 376-381. doi: 10.1007/s00167-022-07233-7
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences. 2005; 102(43): 15545-15550. doi: 10.1073/pnas.0506580102
- Aldoss I, Yang D, Malki MMA, et al. Allogeneic Hematopoietic Cell Transplantation for Relapsed and Refractory Philadelphia Negative B Cell ALL in the Era of Novel Salvage Therapies. Transplantation and Cellular Therapy. 2021; 27(3): 255.e1-255.e9. doi: 10.1016/j.jtct.2020.12.020
- 32. Sattari Fard F, Jalilzadeh N, Mehdizadeh A, et al. Understanding and targeting anoikis in metastasis for cancer therapies. Cell Biology International. 2022; 47(4): 683-698. doi: 10.1002/cbin.11970
- 33. Qu J, Luo M, Zhang J, et al. A paradoxical role for sestrin 2 protein in tumor suppression and tumorigenesis. Cancer Cell International. 2021; 21(1). doi: 10.1186/s12935-021-02317-9
- Korashy HM, Rahman AFMM, Kassem MG. Dasatinib. Profiles of Drug Substances, Excipients and Related Methodology. Published online 2014: 205-237. doi: 10.1016/b978-0-12-800173-8.00004-0

- 35. van Dijk EL, Jaszczyszyn Y, Naquin D, et al. The Third Revolution in Sequencing Technology. Trends in Genetics. 2018; 34(9): 666-681. doi: 10.1016/j.tig.2018.05.008
- 36. Khan SU, Fatima K, Malik F. Understanding the cell survival mechanism of anoikis-resistant cancer cells during different steps of metastasis. Clinical & Experimental Metastasis. 2022; 39(5): 715-726. doi: 10.1007/s10585-022-10172-9
- Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumorinfiltrating immune cells and their therapeutic implications. Cellular & Molecular Immunology. 2020; 17(8): 807-821. doi: 10.1038/s41423-020-0488-6
- Serafim Junior V, Fernandes GM de M, Oliveira-Cucolo JG de, et al. Role of Tropomyosin-related kinase B receptor and brain-derived neurotrophic factor in cancer. Cytokine. 2020; 136: 155270. doi: 10.1016/j.cyto.2020.155270
- Bagchi S, Yuan R, Engleman EG. Immune Checkpoint Inhibitors for the Treatment of Cancer: Clinical Impact and Mechanisms of Response and Resistance. Annual Review of Pathology: Mechanisms of Disease. 2021; 16(1): 223-249. doi: 10.1146/annurev-pathol-042020-042741
- Chen LY, Kang LQ, Zhou HX, et al. Successful application of anti-CD19 CAR-T therapy with IL-6 knocking down to patients with central nervous system B-cell acute lymphocytic leukemia. Translational Oncology. 2020; 13(11): 100838. doi: 10.1016/j.tranon.2020.100838
- Sun Z, Zhao Y, Wei Y, et al. Identification and validation of an anoikis-associated gene signature to predict clinical character, stemness, IDH mutation, and immune filtration in glioblastoma. Frontiers in Immunology. 2022; 13. doi: 10.3389/fimmu.2022.939523
- 42. Ye G, Yang Q, Lei X, et al. Nuclear MYH9-induced CTNNB1 transcription, targeted by staurosporin, promotes gastric cancer cell anoikis resistance and metastasis. Theranostics. 2020; 10(17): 7545-7560. doi: 10.7150/thno.46001