Discover Oncology

Research

Research on the functions and potential mechanisms of STAT3 in chronic myelogenous leukemia

Xiaoyun Feng[1] · Yufeng Qin[2] · Yulong Feng[1] · Yingquan Zhuo[3]

Received: 21 December 2024 / Accepted: 24 April 2025
Published online: 12 May 2025

Abstract

Objective To explore the bioinformatics characteristics and potential mechanisms of signal transducer and activator of transcription (*STAT3*) in chronic myelogenous leukemia (CML).

Methods Through the cancerSEA and CCLE databases, the expression of *STAT3* in CML was verified and analyzed. Subsequently, K562 cells were treated with the *STAT3* inhibitor Stattic. Western blotting, cell counting, and flow cytometry were utilized to observe its impact on the functions of K562 cells. Then, Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) were applied to deeply explore the regulatory mechanism of *STAT3*. The "LIMMA" software package was used to calculate *STAT3*-related differentially expressed genes (DEGs). Machine-earning methods were utilized to screen the *STAT3*-related hub genes. The "pROC" software package was employed to perform Receiver Operating Characteristic (ROC) curve analysis on the hub genes. The "corrplot" software package was used to conduct a correlation analysis of the hub genes. The "RMS" software package was applied to construct a nomogram of the hub genes. Based on the DisGENET database, a disease network of the hub genes was constructed, and the DGIdb database was used to construct a drug network of the hub genes.

Results In CML, the expression of *STAT3* is upregulated compared to housekeeping genes. Among the 14 cell lines related to CML, *STAT3* has the highest expression level in K562 cells. Stattic at a concentration of 5 μM can inhibit the proliferation of K562 cells, promote their apoptosis, and block the cell cycle at the S phase ($P < 0.05$). GSEA and GSVA indicates that amino acid metabolism, NOD-like receptor of *STAT3*. LASSO and SVM-RFE show that *NCF4, PLAS1, IL7R*, and *TAGLN2* are hub differentially expressed genes (DEGs) related to *STAT3*. ROC and Nomogram indicate that the hub DEGs have high clinical diagnostic value. Correlation analysis shows that *PLAS1* and *NCF4* are negatively correlated, while *PLAS1* and *TAGLN2* are positively correlated. The construction of gene-disease networks reveals that these genes not only participate in the occurrence and development of CML but also jointly participate in multiple disease processes. The gene-drug network obtained 38 drugs targeting genes.

Conclusion STAT3 might serve as a potential target for the treatment of CML. In CML, *NCF4, PLAS1, IL7R,* and *TAGLN2* are hub genes associated with *STAT3*. These findings offer a fundamental theory for comprehending the pathogenesis of CML.

Xiaoyun Feng and Yufeng Qin have contributed equally as co-first authors.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02492-5.

✉ Xiaoyun Feng, 2946683319@qq.com; ✉ Yingquan Zhuo, Zyqgyfy1220@gmc.edu.cn; Yufeng Qin, 3175515989@qq.com; Yulong Feng, 2259468737@gmail.com | [1]Shizhen College of Guizhou University of Traditional Chinese Medicine, Guiyang 550200, Guizhou, China. [2]Department of Prosthodontics, Affiliated Stomatological Hospital of Guizhou Medical University, Guiyang 550004, Guizhou, China. [3]Department of Pediatric Surgery, The Afliated Hospital of Guizhou Medical University, Guiyang 550004, China.

Keywords Chronic myeloid leukemia · *STAT3* · Cell detection · Bioinformatics · Machine learning method

1 Introduction

Chronic myeloid leukemia (CML) is a highly life-threatening clonal malignant disease. The fundamental cause of its pathogenesis mainly lies in the reciprocal translocation between the long arms of chromosomes 9 and 22 [1]. This specific genetic alteration leads to the formation of the fusion oncoprotein BCR-ABL, which is a constitutively active tyrosine kinase with a molecular weight of 210 kDa [2]. The natural course of CML can be clearly divided into three distinct stages. In the chronic phase, most patients present with symptoms such as elevated white blood cell counts and splenomegaly [3]. Although CML in this phase is relatively stable to some extent, continuous monitoring and appropriate treatment are essential. As the disease progresses, it enters the accelerated phase, where the condition gradually deteriorates, accompanied by the emergence of various abnormal cellular characteristics and clinical manifestations [4]. Eventually, it enters the blast crisis phase, which is closely related to bone marrow failure. Unfortunately, the prognosis is extremely poor [5].

The advent of tyrosine kinase inhibitors (TKIs) has greatly revolutionized the treatment of CML patients. It has effectively transformed what was once a fatal leukemia into a manageable chronic condition treatable with oral medications [6]. Specifically, in the chronic phase, TKIs demonstrate remarkable therapeutic effects in CML patients. Patients who respond optimally to treatment can even expect a lifespan comparable to that of healthy individuals [7]. However, it is worth noting that over time, a considerable number of patients undergoing TKI treatment develop drug resistance, which poses a significant challenge to the long-term treatment of CML [7]. Therefore, addressing the issue of TKI resistance or actively exploring alternative drug targets holds great promise for opening up new therapeutic avenues and formulating innovative treatment strategies for CML [8].

During the occurrence and development of CML, multiple signal transduction pathways are abnormally activated [9]. Among them, *STAT3* is a key signal transduction molecule. A large amount of previous research evidence indicates that the abnormal activation of *STAT3* is closely associated with the occurrence and development of various tumors, including CML [10]. Structurally, the *STAT3* protein consists of 770 amino acids and contains six functionally conserved domains. These domains endow *STAT3* with the ability to regulate a wide range of genes involved in key aspects of cancer cell behavior, such as angiogenesis, survival, proliferation, drug resistance, invasion, metastasis, and immune evasion [11]. A large number of studies have shown that *STAT3*-mediated metabolic alterations can effectively counteract drug resistance in CML [12]. Under the influence of TKI treatment, CML cells exhibit BCR-ABL-independent activation of *STAT3*. In addition, genetic knockout of *STAT3* has been shown to significantly suppress the development of drug persistence in CML cells. Moreover, the application of small-molecule inhibitors targeting *STAT3* can enhance the sensitivity of CML cells to TKI treatment [13].

Among them, the *STAT3* inhibitor Stattic has shown unique effects in related studies. When treating the K562 cell line of CML with Stattic, it can significantly affect the *STAT3* pathway [14]. Detection by Western blotting technology reveals that it can effectively reduce the expression level of *STAT3,* thereby blocking the continuous activation of the *STAT3* pathway [15]. The abnormal activation of this pathway is one of the key factors in the development of CML. Meanwhile, in the study of apoptosis-related genes, it is found that after treatment with Stattic, the expression of the anti-apoptotic gene *BCL-2* is significantly downregulated, while the expression of the pro-apoptotic gene *BAX* is significantly upregulated [16]. This change indicates that Stattic can induce apoptosis in K562 cells by regulating the expression of apoptosis-related genes. In addition, evaluation using cell counting and flow cytometry shows that Stattic can effectively inhibit the proliferation of K562 cells, change the cell-cycle distribution, and cause more cells to be arrested in phases more sensitive to apoptosis, further promoting apoptosis [17].

Our research team has been intensively engaged in in-depth research on the epigenetic regulation of the malignant proliferation of leukemia cells. Relevant literature reports that the activity of JAK/STAT3 plays a crucial role in promoting epigenetic pluripotency. Specifically, it can promote de novo DNA methylation and contribute to the formation of chromatin in the late stage of reprogramming [18]. In addition, numerous independent studies have consistently demonstrated that *STAT3* has a significant impact on the survival and disease progression of CML cells through multiple mechanisms. For example, it can drive tumor cell proliferation, suppress apoptosis, and actively participate in the tumor immune-escape process [19]. Given these important findings, delving deeper into the potential functions and regulatory mechanisms of *STAT3* is of utmost theoretical and clinical significance for the treatment of CML.

In the era of rapid development of bioinformatics technology, the systematic analysis of gene expression data has emerged as a powerful tool for deciphering the biomolecular characteristics and pathological mechanisms underlying various diseases. In the present study, we aim to ingeniously combine functional experiments with bioinformatics approaches to conduct an in-depth exploration of the mechanism through which *STAT3* functions during the progression of CML. At the same time, we will vigorously search for differentially expressed genes that are significantly correlated with *STAT3*. We sincerely hope that these efforts will bring new possibilities for the clinical treatment of CML, thereby offering more hope and improved treatment outcomes for patients suffering from this debilitating disease.

2 Materials and methods

2.1 Materials

The K562 cell line was purchased from Wuhan Procell Life Science & Technology Co., Ltd. The *STAT3* inhibitor Stattic was purchased from MedChemExpress. The cell cycle and apoptosis detection kit was provided by Biosharp, and the phosphate-buffered saline was purchased from Solarbio. The Stat3, Bcl-2, and Bax antibodies were all purchased from Wuhan Sanying Biotechnology Co., Ltd. in China.

2.2 Data collection

In this study, we obtained the mRNA transcriptome datasets GSE5550 and GSE24739 related to chronic myeloid leukemia (CML) from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/). Specifically, the GSE5550 dataset encompasses blood samples from 9 CML patients and 8 healthy controls, while the GSE24739 dataset contains blood samples from 16 CML patients and 8 normal controls. The data were retrieved on August 11, 2024. To conduct in—depth analysis, we extracted the relevant gene sets c5.go.v2023.1.Hs.symbols.gmt and c2.cp.kegg.v2023.1.Hs.symbols.gmt from the Molecular Signatures Database v7.0 (MSigDB v7.0) for Gene Set Enrichment Analysis (GSEA) and Gene Set Variation Analysis (GSVA) to analyze the gene sets.

2.3 Cell culture

For the resuscitation operation of K562 cells, RPMI 1640 cell culture medium, combined with 10% fetal bovine serum produced by American Gibco company and 1% double antibody, is cultured in a cell incubator containing 5% carbon dioxide and set at a temperature of 37 ℃. Ensure that the cell density is maintained at a level of 8×10^5, and replace the medium after culturing for 24 h.

2.4 Cell proliferation detection

Inoculate 8×10^5 cells in a six-well cell culture plate, and then add Stattic inhibitors with final concentrations of 0 μM, 2.5 μM, 5 μM, and 10 μM respectively. After 48 h and 96 h, measure the number of proliferating cells and conduct statistical analysis using GraphPad Prism software.

2.5 Western blot experiment

In the collected K562 cell samples, the protease inhibitor PMSF and lysis buffer RIPA were added to perform the cell lysis procedure. Subsequently, the samples were placed in a frozen high-speed centrifuge and centrifuged for 30 min to collect the supernatant. Then, the supernatant was placed in a metal bath and boiled at 100 ℃ for 10 min. Electrophoresis and transfer membrane operations were performed using a 15% concentration of SDS-PAGE premixed gel. After the transfer membrane was completed, skim milk powder was used to block for 1 h at room temperature. After that, the primary antibody was added to the sample and incubated overnight on a shaker at 4 ℃. The next day, the membrane was washed three times with PBST, then the secondary antibody was incubated at 37 ℃ for 1 h, and the membrane was washed three times again with PBST. Finally, the protein was detected using a biological imaging system.

2.6 Flow cytometry for detecting cell apoptosis

After conducting the cell proliferation experiment, select an appropriate concentration of Stattic based on the results to treat the K562 cell line in the logarithmic growth phase, and then collect cell samples. The samples need to be washed twice with pre-cooled PBS solution to remove residual medium components. Next, set up a Blank tube, a PE single-stained tube, and a 7-AAD single-stained tube to facilitate the subsequent adjustment of the voltage and compensation parameters of the flow cytometer. Add 5 μL of PE and 5 μL of 7-AAD dye to the samples respectively, gently mix well, and incubate at room temperature in the dark for 15 min to ensure that the dyes are fully bound to the cells. After incubation, add 400 μL of Binding Buffer to each tube, mix well again, and then filter the cell suspension through nylon cloth into a flow tube to ensure the preparation of a single-cell suspension. Finally, use the BD FACSCalibur flow cytometer to detect the stained cell samples and analyze the data through Flowjo software to distinguish the proportions of live cells, early apoptotic cells, late apoptotic cells, and dead cells. The statistical analysis of the obtained data is completed by GraphPad Prism software.

2.7 Flow cytometry for detecting cell cycle

After treating K562 cells in the logarithmic growth phase with Stattic, we collected various cell samples. Subsequently, we washed the samples twice with phosphate-buffered saline (PBS) and removed the supernatant. Then, we fixed the cells with 1 mL of pre-cooled 70% ethanol. For each sample, we added 25 μL of propidium iodide staining solution (20-fold dilution) and 10 μl of RNase A (50-fold dilution) to 0.5 mL of staining buffer and thoroughly mixed them. Then, we added 0.5 mL of propidium iodide staining solution to each cell sample, gently mixed and resuspended the cell pellet, and then incubated it at 37 °C in the dark for 30 min for flow cytometry detection. The stained samples were detected by BD FACSCalibur flow cytometer, detecting red fluorescence at an excitation wavelength of 488 nm and simultaneously recording light scattering. Finally, we analyzed the data using Modfit-1 software and accurately counted the percentage of cells in each stage of the cell cycle.

2.8 GEO chip data arrangement

Using R software, background correction, normalization processing, and conversion to base 2 logarithm were carried out on the two transcriptome datasets, GSE5550 and GSE24739. When multiple probes correspond to the same gene, the average value method was adopted to determine the expression level of that gene. Through the "SVA" software package, batch effects and other unwanted variations were successfully eliminated. Finally, these two datasets were integrated into a unified dataset for subsequent analysis.

2.9 Gene set enrichment analysis (GSEA) analysis

In order to further clarify the molecular mechanism through which *STAT3* acts on CML, the Spearman correlation between the *STAT3* gene and all the other genes within the CML dataset was computed. Based on the ranking derived from this correlation, the gene sets c5.go.v2023.1.Hs.symbols.gmt and c2.cp.kegg.v2023.1.Hs.symbols.gmt were adopted as the background sets for the enrichment analysis, and GSEA was then conducted on the *STAT3* gene. The minimum and maximum sizes of the gene sets were set at 10 and 500 respectively. A nominal p value less than 0.05 along with a false discovery rate (FDR), represented by q value, less than 0.25 were considered to be statistically significant.

2.10 Gene set variation analysis (GSVA) analysis

To gain a deeper understanding of the differences in pathways between the high and low expression groups of *STAT3*, the GSVA scores of each pathway in the high and low expression groups of *STAT3* within CML tumor samples were calculated.

2.11 Identification of *STAT3*-related hub genes

Under the conditions where the correlation coefficient was greater than zero point three and less than one, as well as less than minus zero point three and greater than minus one, the STAT3 correlated gene expression matrix of CML transcriptome data was computed through the "limma" package. Using the "limma" package, differential genes(DEGs) analysis was carried out on the high-expression and low-expression groups of *STAT3* within the *STAT3* correlated gene expression matrix to obtain the corresponding P-value and logFC (log2 fold change) value. The difference threshold was defined as: $P < 0.05$ and log2-fold change FC > 0.5. Visualization was conducted using the R package "Heatmap". The heatmap depicted the expression of DEGs in the high-expression and low-expression group samples. To comprehensively understand the distribution of DEGs, a volcano plot was utilized to display this distribution. In order to screen hub genes more precisely and deeply explore the key regulatory genes of *STAT3* in CML, this study adopted two advanced machine-learning methods. One of them is the Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithm. Through its unique regularization approach, it effectively improves the prediction accuracy and can efficiently and accurately identify a set of candidate genes of great research value from massive gene data. In the actual operation process, we utilized the powerful "glmnet" package in the R programming language to successfully execute the LASSO regression algorithm, ensuring the efficiency and accuracy of the analysis process. The other method is the Support Vector Machine (SVM). As a classic supervised machine-learning technique, it has demonstrated excellent performance in both classification and regression analysis and is extremely widely applied. However, to avoid overfitting in the model, which may affect the reliability of the analysis results, we introduced the Recursive Feature Elimination (RFE) algorithm. This algorithm can, based on specific evaluation criteria, screen layer by layer from the complex metadata queue and finally select the optimal genes that are most crucial for improving the model performance. Based on the complementary advantages of the above-mentioned methods, this study further adopted the Support Vector Machine-Recursive Feature Elimination (SVM-RFE) algorithm for feature selection. This algorithm combines the powerful classification ability of SVM with the feature-screening advantage of RFE, aiming to precisely screen out the hub genes that are most closely associated with *STAT3* in CML and most critical to the disease progression from numerous genes. In the specific implementation process of the SVM-RFE algorithm, we relied on the "e1071" package in the R language.

2.12 Differential expression and roc analysis of hub genes

The expression of hub genes was visualized by means of the "ggplot2" and "ggpubr" packages within R software. The area under the receiver operating characteristic curve (ROC), also known as the Area Under Curve (AUC), was calculated through the "pROC" package to gain a deeper understanding of the clinical diagnostic significance of hub genes for CML. When the AUC is greater than zero point seven, it indicates a relatively high level of clinical diagnostic accuracy.

2.13 Correlation analysis of hub genes

After clarifying the clinical diagnostic effect among hub genes, on the basis of the correlation between *STAT3* and hub genes, the correlation of hub genes was further calculated using the "corrplot" R package to uncover their functional similarity.

2.14 Construction and verification of diagnostic model

The "rms" R package was employed to predict the Nomogram model for CML. "Score" refers to the score of the subsequent related items, and "Total score" is the sum of all the elements above. Subsequently, the calibration curve was utilized to assess the prediction ability of the line graph model. Finally, decision curve analysis and clinical impact curve were employed to evaluate the clinical application value of the model.

2.15 Construction of STAT3 and its hub genes-disease network

To explore the diseases jointly involved with STAT3 and hub genes, diseases related to *STAT3* and hub genes were retrieved from the DisGeNET (https://www.disgenet.org/) database. Subsequently, the gene-disease relationship network was constructed using Cytoscape.

2.16 Gene-drug network construction

In an effort to explore potential therapeutic drugs for *STAT3* and its associated characteristic genes, drug prediction was implemented in the drug-gene interaction database DGIdb (https://dgidb.org).

2.17 Statistical analysis

In this study, the independent sample t-test was employed for comparing the means between two groups in terms of cell proliferation, apoptosis, and cell cycle statistics. A p value less than 0.05 was regarded as statistically significant. Bioinformatics analysis was carried out using R language version 4.2.0. Relevant language packages were directly obtained from Bioconductor (http://www.bioconductor.org/), including packages such as "rms", "SVA", "LIMMA", "clusterProfiler", "DOSE", "glmnet", "ggplot2", "ggpubr", "pROC", "pheatmap", "vioplot", and "corrplot". When $^{*}P < 0.05$, it is statistically considered that there is a notable difference.

3 Results

3.1 Expression of *STAT3* in CML-related cell lines

To explore the expression of *STAT3* in chronic myelogenous leukemia (CML), based on the CancerSEA (hrbmu.edu.cn) database, it was found that *STAT3* expression in CML is higher than that of housekeeping genes (Fig. 1A). To further investigate the expression of *STAT3* in CML-related cell lines, through the CCLE (aclbi.com) database, it was found that among 14 CML-related cell lines, *STAT3* expression is highest in the K562 cell line (Fig. 1B). Based on this, this study uses the K562 cell line as an in vitro model for functional research.

3.2 Effects of STAT3 inhibitor stattic on cell lines such as K562

This study demonstrated the changes in the number of K562 cells over time (0–4 days) after treatment with different concentrations of Stattic (0 μM, 2.5 μM, 5 μM, 10 μM). It revealed that Stattic at a concentration of 5 μM had a significant inhibitory effect on the growth of K562 cells, with the cell number increasing significantly less than that at other concentrations ($P < 0.001$) (Fig. 2A).Furthermore, this study presented the changes in the number of HL60 and THP-1 cells over time (0–4 days) after treatment with 5 μM Stattic. Similar to the K562 cell line, the effects of Stattic on the growth of these two cell lines were observed ($P < 0.001$) (Fig. 2B).Through WB experiments, this study detected the expression levels of Stat3, Bcl-2, and Bax proteins in K562, HL60, and THP—1 cell lines after treatment with 5 μM Stattic. The experimental results showed that Stattic significantly inhibited the expression of Stat3 and Bcl-2, while promoting the expression of Bax, and this trend was similar in different cell lines. The bar graphs below quantified the relative expression levels

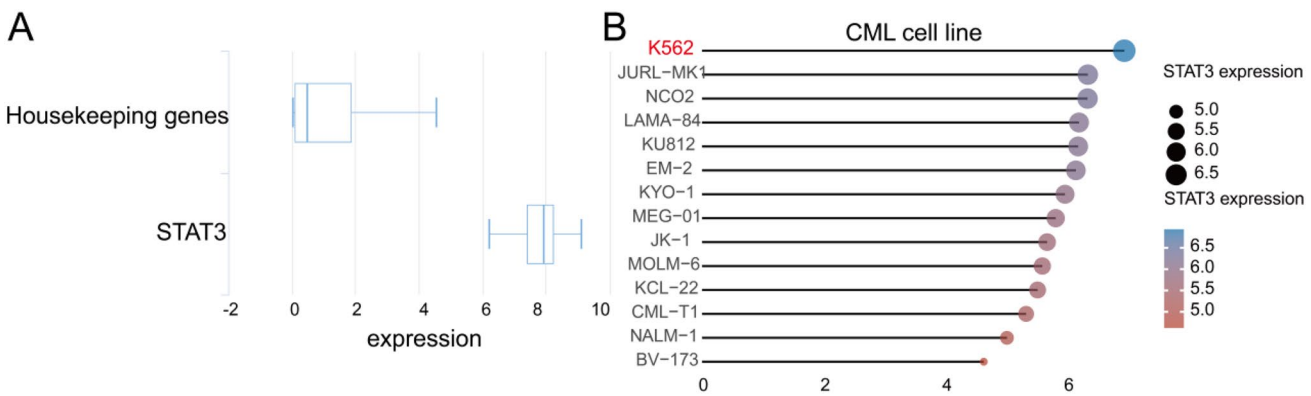


Fig. 1 Expression of *STAT3* in CML-related cell lines. **A** Analyze the expression of *STAT3* in the CancerSEA database. **B** Detect the expression of *STAT3* in 14 cell lines related to CML

of these proteins ($P < 0.001$) (Fig. 2C).Flow cytometry was used to analyze the apoptosis of K562 cells before and after treatment with 5 μM Stattic. Early and late apoptotic cells were distinguished by Annexin V-PE and 7-AAD staining. The experimental results indicated that the apoptosis rate of K562 cells increased significantly after Stattic treatment (Fig. 2D) ($P < 0.001$).Flow cytometry was also employed to detect the cell cycle distribution of K562 cells before and after treatment with 5 μM Stattic. The results showed that after Stattic treatment, the proportions of cells in the G_2/M and S phases decreased, and the proportion of cells in the G_0/G_1 phase changed. The bar graph on the right quantified and compared the proportions of cells in each phase (Fig. 2E) ($P < 0.001$).Overall, these data suggest that the *STAT3* inhibitor Stattic can inhibit the growth of cell lines such as K562, promote apoptosis by regulating the expression of related proteins, and affect the cell cycle distribution.

3.3 GSEA analysis

After clarifying the phenotypic function of *STAT3* on CML, GSEA was further used to deeply understand the molecular mechanism of *STAT3* acting on CML, so as to determine the impact of the synergistic changes of *STAT3* within this gene set on phenotypic changes. The pathways with significant functional enrichment of *STAT3* based on the c5.go.v2023.1.Hs.symbols.gmt gene set are amino acid metabolism, response to molecules of bacterial origin, mitochondrial matrix, organic acid catabolic process, and cellular amino acid metabolic process (Fig. 3A). The results with significant pathway enrichment of STAT3 based on the cp.kegg.v2023.1.Hs.symbols.gmt gene set are NOD-like receptor signaling pathway, cytokine receptor interaction, DNA replication, porphyrin and chlorophyll metabolism, and valine and isoleucine degradation (Fig. 3B).

3.4 GSVA analysis

To further understand the differences in pathways between the high and low expression groups of *STAT3*, the GSVA scores of each pathway in the high and low expression groups of *STAT3* in CML tumor samples were calculated. It was found that there were significant differences in the following pathways. Among them, the more significant pathways based on c5.go.v2023.1.Hs.symbols.gmt functional enrichment analysis are valine and isoleucine degradation, fatty acid metabolism, propionic acid metabolism, and cell cycle. The experimental results of this study also show that the expression of *STAT3* can regulate the cell cycle of chronic myeloid leukemia K562 cells (Fig. 4A). The more significant pathways based on c2.cp.kegg.v2023.1.Hs.symbols.gmt pathway enrichment analysis are pigment body apoptosis condensation, regulation of axon diameter, cellular magnesium ion balance, signaling pathways similar to nod-like receptors, and regulation of chondrocyte development (Fig. 4B). And some of the same pathways are enriched simultaneously with GSEA, further indicating the reliability of the study. These differentially expressed pathways are more biologically significant and more interpretable compared to genes.

3.5 Identification of hub genes of *STAT3*

A total of 1968 genes correlated with *STAT3* were obtained by calculating CML transcriptome data through the "limma" package, and the *STAT3*-correlated gene expression matrix was obtained. Differential gene analysis was performed on samples in the high and low expression groups of *STAT3* in the *STAT3*-correlated gene expression matrix, and a total of 412 significantly DEGs were obtained, including 254 up-regulated genes and 157 down-regulated genes (Fig. 5A, B). LASSO regression was used to identify differentially expressed genes related to STAT3, and 16 variables were determined as diagnostic markers for CML (Fig. 5C, D). The SVM-RFE algorithm determined six feature subsets (Fig. 5E). Venn analysis identified four overlapping feature genes between the two machine learning algorithms as *TAGLN2, NCF4, PIAS1*, and *IL7R* (Fig. 5F). The above four genes may be hub genes involved in the occurrence and development of CML.

3.6 DEGs and ROC analysis of *STAT3* significantly correlated genes

The research results show that the expressions of *TAGLN2* and *NCF4* are significantly up-regulated, and the expressions of PIAS1 and IL7R are significantly down-regulated (Fig. 6A). The results of ROC analysis show that *TAGLN2* (AUC = 0.895), NCF4 (AUC = 0.973), PIAS1 (AUC = 0.915), and IL7R (AUC = 0.907) have relatively high AUC values and strong diagnostic value (Fig. 6B).

Fig. 2 Effects of *STAT3* inhibitor Stattic on proliferation, apoptosis and cell cycle of K562 cells. **A** presents the changes in the number of K562 ▶ cells within 0–6 days after treatment with different concentrations of Stattic. **B** Shows the changes in the number of HL60 and THP-1 cells within 0–6 days after treatment with 5 μM Stattic. **C** Displays the WB results and quantitative analysis of the expression levels of Stat3, Bcl-2, and Bax proteins in K562, HL60, and THP-1 cell lines after treatment with 5 μM Stattic. **D** Uses flow cytometry to compare the apoptosis of K562 cells before and after treatment with 5 μM Stattic. **E** By means of flow cytometry, compares the cell cycle distribution of K562 cells before and after treatment with 5 μM Stattic

3.7 Correlation analysis of *TAGLN2, NCF4, PIAS1*, and *IL7R*

The study reveals a negative correlation between *PIAS1* and NCF4, suggesting their possible negative regulatory roles in the biological regulatory network. At the same time, the positive correlation between *PIAS1* and *TAGLN2* indicates significant functional similarity between the two, providing a new perspective for understanding their synergistic effects in cell signal transduction (Table 1) (Fig. 7).

3.8 Construction and validation of nomogram

With the assistance of the "rms" software package, this study constructed a Nomogram model for characteristic genes (*PLAS1* and *NCF4*) with a total score exceeding 2 (Fig. 8A). Its predictive ability was meticulously evaluated through calibration curves. In decision curve analysis, the Nomogram curve is significantly higher than the gray line. The *PLAS1* and *NCF4* curves reveal that under high-risk thresholds, patients can derive substantial benefits from the Nomogram model (Fig. 8B). The calibration curve further indicates that the disparity between the true risk and predicted risk of CML is minimal, fully demonstrating the high accuracy of the Nomogram model (Fig. 8C). To evaluate the clinical effect of the Nomogram model more intuitively, we drew a clinical impact curve based on the DCA curve. Within the high-risk threshold, the "high-risk number" curve is closely adjacent to the "high-risk number with events" curve, further indicating the excellent predictive ability of the Nomogram model (Fig. 8D).

3.9 Construction of gene-disease network.

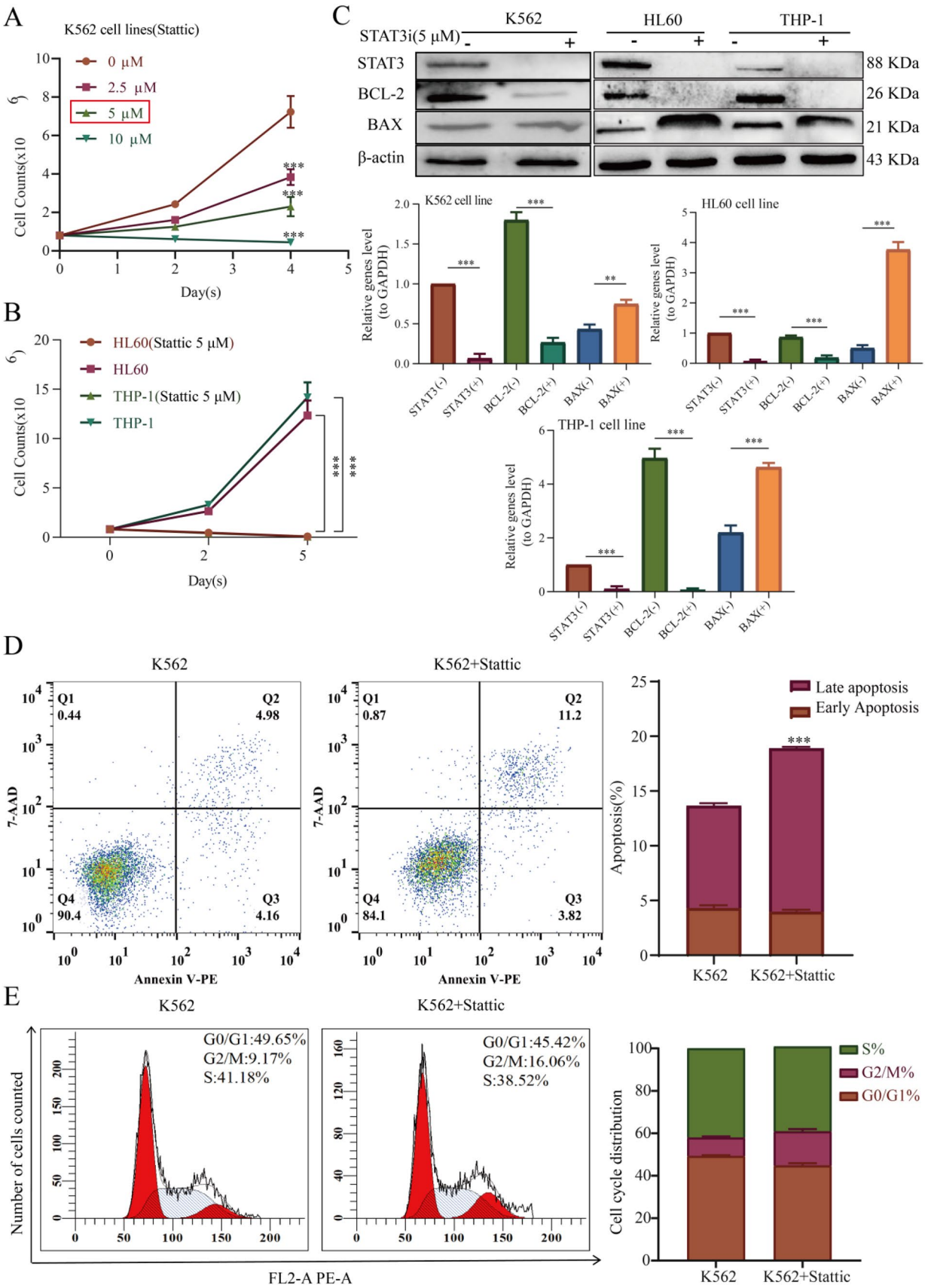
Research shows that genes such as *STAT3, NCF4, PLAS1, IL7R,* and *TAGLN2* not only play key roles in the occurrence and development of CML, but also jointly participate in the pathological processes of many other diseases, including systemic lupus erythematosus, asthma, adenocarcinoma of lung, breast carcinoma, neoplasms, primary malignant neoplasm, malignant neoplasms, and diabetes mellitus. This finding has been detailedly demonstrated in (Fig. 9).

3.10 Construction of gene-drug network

This study successfully identified 38 drugs targeting specific genes. Based on this, a detailed gene-drug interaction network was constructed. It is particularly noteworthy that the *IL7R* gene is dually regulated by recombinant human interleukin-7 and ruxolitinib. The *NCF4* gene is jointly affected by six drugs including cyclophosphamide, prednisone, rituximab, vincristine, idarubicin, and doxorubicin hydrochloride. In addition, the *STAT3* is regulated by as many as 30 drugs such as celecoxib and cucurbitacin E (Fig. 10).In the DGIdb database, we did not find any direct interactions between the two additional genes (*PLAS1* and *TAGLN2*) and drugs that we studied.

4 Discussion

CML accounts for approximately 15% of adult leukemias [20]. Although the introduction of TKIs has achieved remarkable results in treatment, some patients still cannot achieve the goal of cure due to drug resistance or intolerance issues [21]. As a member of the signal transducer and activator of transcription (STAT) protein family, *STAT3* undergoes abnormal activation within tumor cells. It regulates cell growth, differentiation, and gene expression, participates in numerous biological processes, and is closely associated with tumor drug resistance [22]. Therefore, the development of small-molecule inhibitors targeting *STAT3* has become an important goal in the field of tumor treatment research. In leukemia


A
K562 cell lines(Stattic)
0 μM
2.5 μM
5 μM
10 μM
Cell Counts(x10 6)
Day(s)
B
HL60(Stattic 5 μM)
HL60
THP-1(Stattic 5 μM)
THP-1
Cell Counts(x10 6)
Day(s)
C
STAT3i(5 μM)
K562
HL60
THP-1
STAT3
BCL-2
BAX
β-actin
88 KDa
26 KDa
21 KDa
43 KDa
K562 cell line
HL60 cell line
THP-1 cell line
Relative genes level (to GAPDH)
STAT3(-)
STAT3(+)
BCL-2(-)
BCL-2(+)
BAX(-)
BAX(+)
D
K562
K562+Stattic
Q1 0.44
Q2 4.98
Q4 90.4
Q3 4.16
Q1 0.87
Q2 11.2
Q4 84.1
Q3 3.82
7-AAD
Annexin V-PE
Late apoptosis
Early Apoptosis
Apoptosis(%)
K562
K562+Stattic
E
K562
K562+Stattic
G0/G1:49.65%
G2/M:9.17%
S:41.18%
G0/G1:45.42%
G2/M:16.06%
S:38.52%
Number of cells counted
FL2-A PE-A
S%
G2/M%
G0/G1%
Cell cycle distribution
K562
K562+Stattic

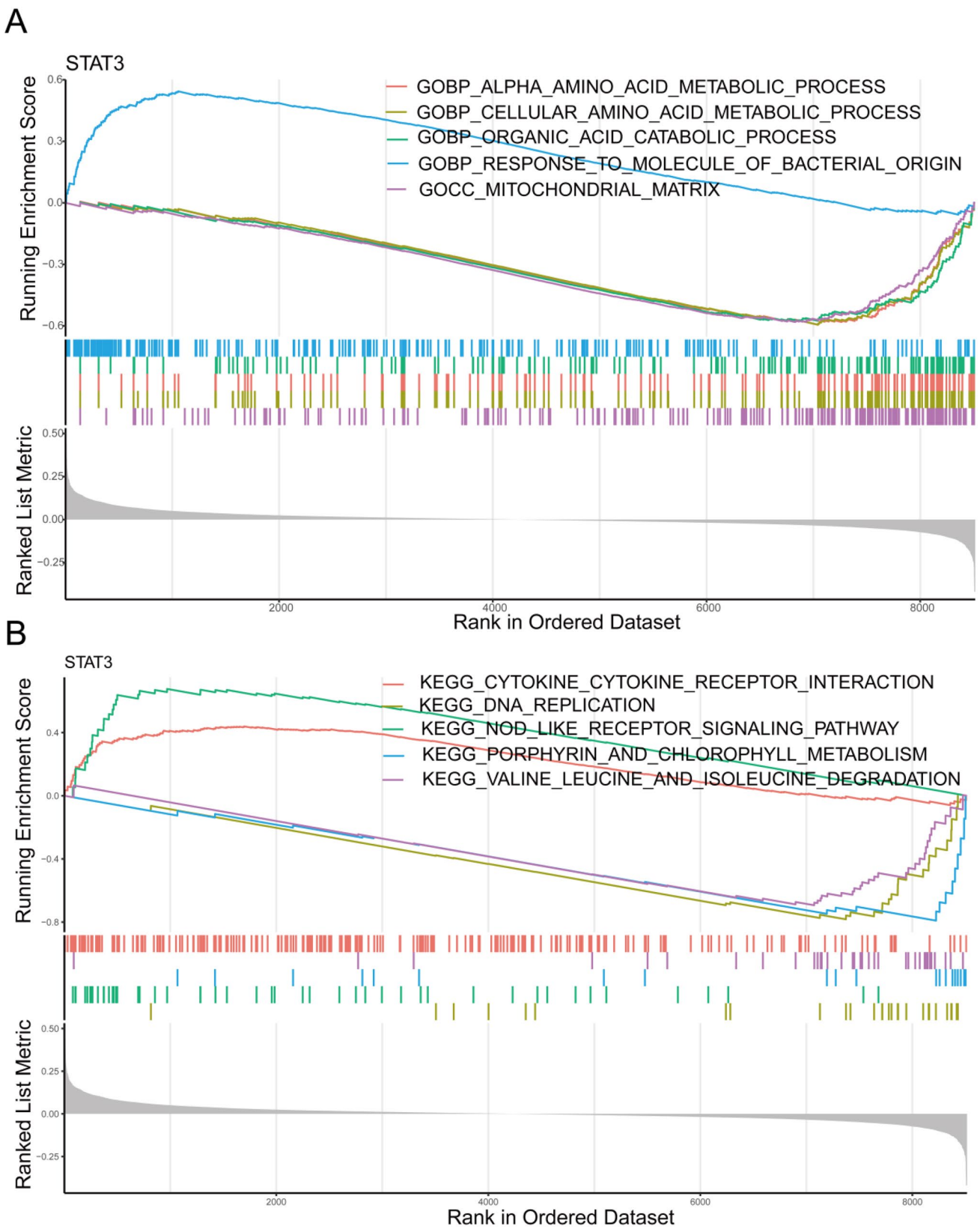


Fig. 3 GSEA analysis. **A** Conducted *STAT3* enrichment analysis based on the c5.go.v2023.1.Hs.symbols.gmt gene set. **B** Performed pathway enrichment analysis of *STAT3* based on the c2.cp.kegg.v2023.1.Hs.symbols.gmt gene set

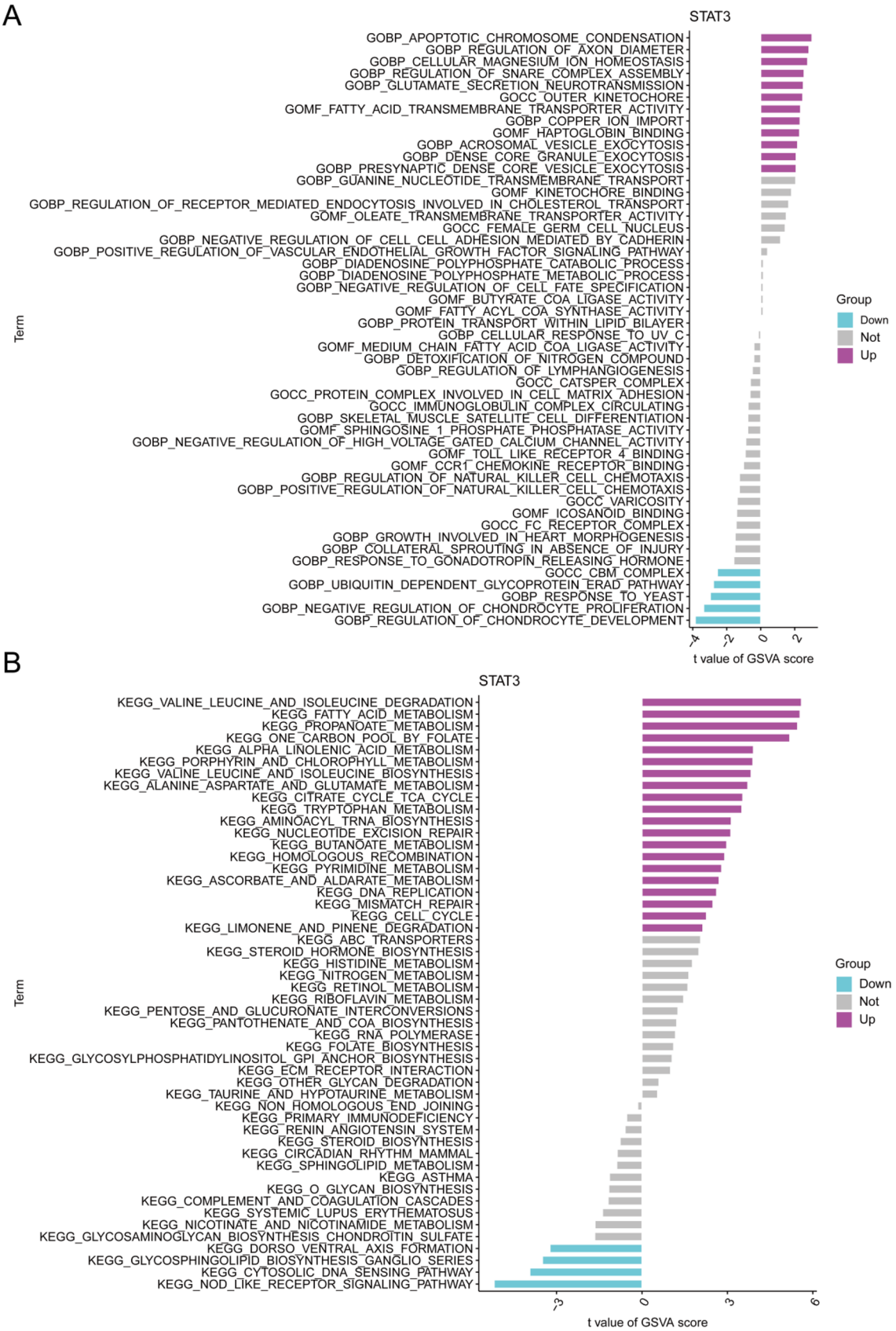


Fig. 4 Shows the results of GSVA enrichment analysis of high and low expression groups of *STAT3* in CML. **A** Perform gene set functional enrichment analysis of *STAT3* based on c5.go.v2023.1.Hs.symbols.gmt. **B** Conduct gene set pathway enrichment analysis of *STAT3* based on c2.cp.kegg.v2023.1.Hs.symbols.gmt

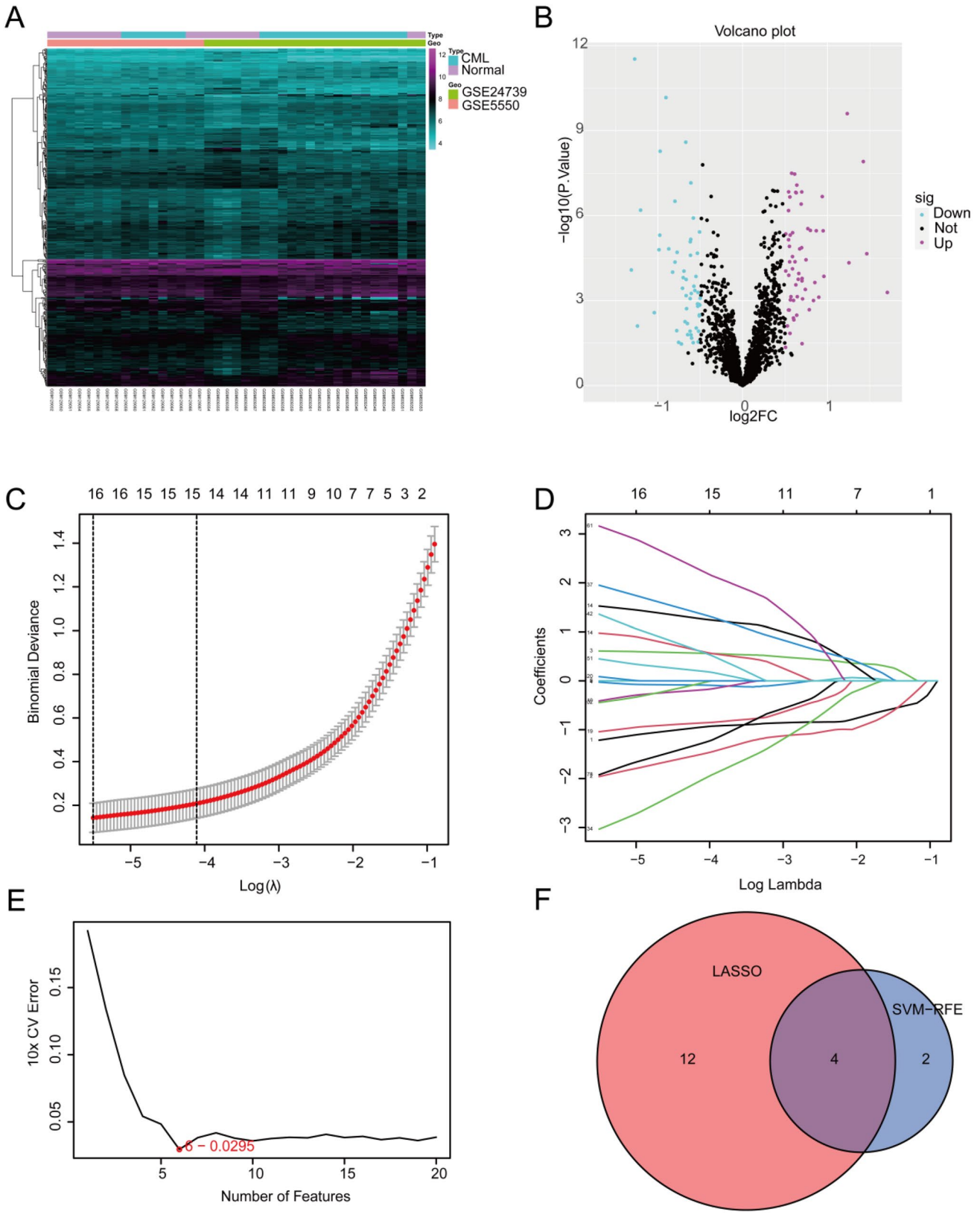


Fig. 5 Shows the screening results of hub DEGs related to *STAT3*. **A** Show the heatmap of *STAT3*-related DEGs. **B** Analyze the volcano plot of *STAT3*-related DEGs. **C** Present the pathway diagram of LASSO coefficients. **D** Show the pathway diagram of LASSO regularization. **E** Display the analysis results of SVM-RFE. **F** Venn diagram analysis

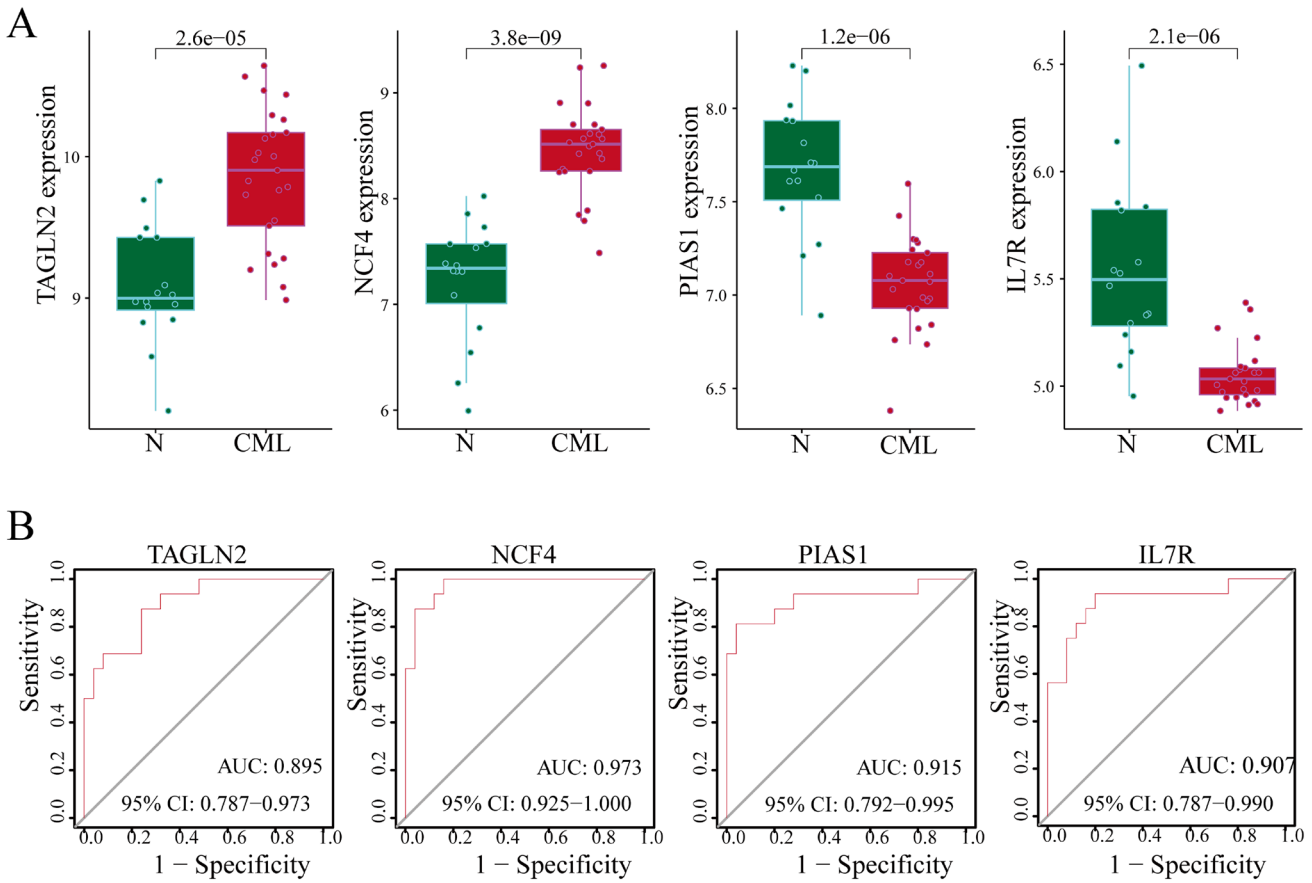


Fig. 6 DEGs and ROC analysis of *STAT3*. **A** Analyze the differential expression of *TAGLN2, NCF4, PIAS1* and *IL7R*. **B** Conduct ROC analysis on *TAGLN2, NCF4, PIAS1* and IL7R. $^{*}P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.001$

Table 1 Correlation analysis

Gene	NCF4	PIAS1	IL7R	TAGLN2
NCF4	R =1.0			
PIAS1	**R = − 0.58	R = 1		
IL7R	R = 0,09	R = 0.18	R = 1	
TAGLN2	R = − 0.17	*R = 0.4	R = − 0.19	R = 1

cells, the activation of *STAT3* promotes the transcription of angiogenesis—related genes, leading to the formation of leukemia phenotypes [19]. In addition, it maintains the undifferentiated state of leukemia cells by inducing the expression of myeloid differentiation genes [23].

In the regulatory mechanism of cell apoptosis, the *BCL-2* and *BAX* genes play crucial roles. *BCL-2* is an anti-apoptotic gene. Its high expression can inhibit cell apoptosis, enabling cancer cells to continue to survive and proliferate [24]. Conversely, *BAX* is a pro-apoptotic gene. When it is activated and highly expressed, it promotes the occurrence of cell apoptosis. Previous studies have shown that the abnormal activation of *STAT3* is closely related to abnormal cell apoptosis [25]. *STAT3* may affect the process of cell apoptosis by regulating the expression of apoptosis—related genes such as *BCL-2* and *BAX* [26]. In view of this, when studying the mechanism of action of STAT3 on chronic myeloid leukemia cells, choosing the *BCL-2* and *BAX* genes as research objects helps to deeply analyze the specific role of STAT3 in regulating cell apoptosis, providing key clues for understanding the pathogenesis of chronic myeloid leukemia and developing new treatment strategies [27].

The expression pattern of *STAT3* in leukemia has been widely studied. Research shows that the expression of *STAT3* is up-regulated in leukemia, indicating that *STAT3* plays a key role in this disease [28]. Targeting the *STAT3* signaling pathway has become a potential treatment strategy for leukemia. Currently, small—molecule inhibitors and antisense

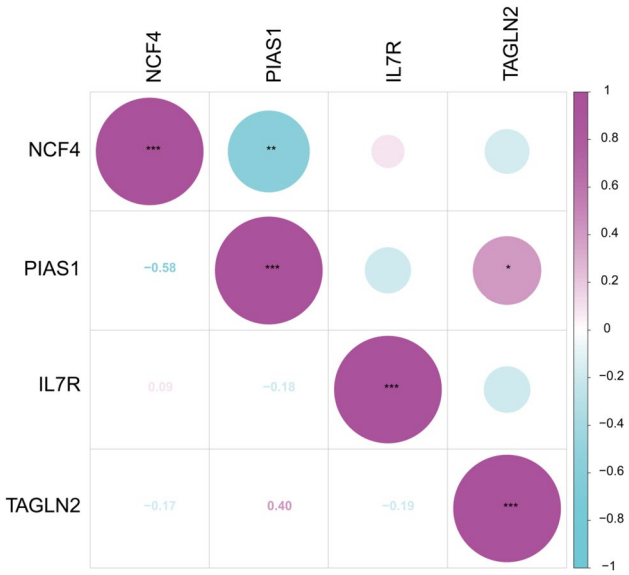


Fig. 7 Correlation between hub genes.*indicates that the correlation coefficient R ranges from 0.3 to 0.5 or from − 0.5 to − 0.3; **indicates that the correlation coefficient R ranges from 0.5 to 0.7 or from − 0.7 to − 0.5; ***indicates that the correlation coefficient R ranges from 0.7 to 1 or from − 1 to − 0.7

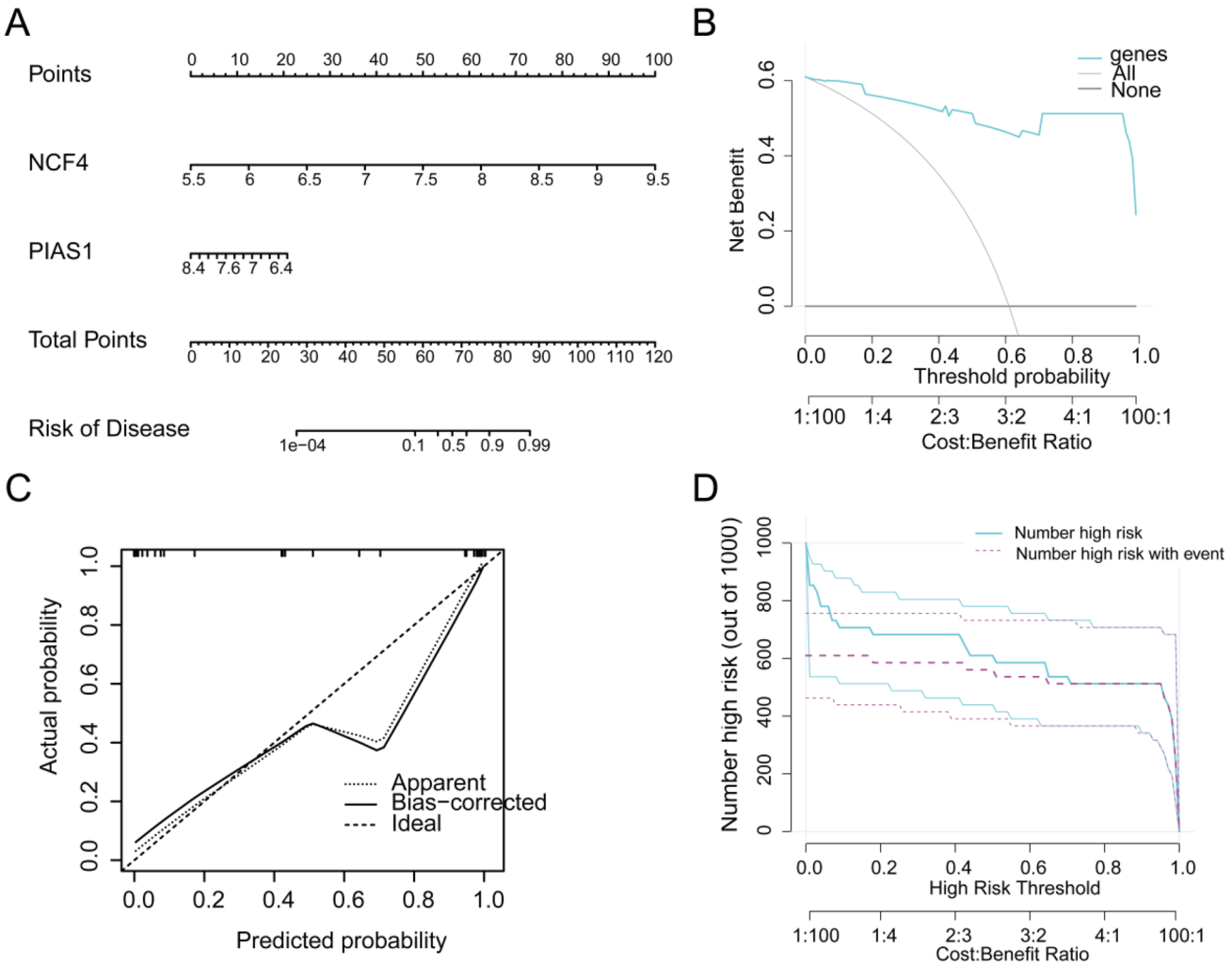


Fig. 8 Construction and validation of Nomogram. **A** Nomogram. **B** Decision curve. **C** Calibration curve. **D** Clinical impact curve

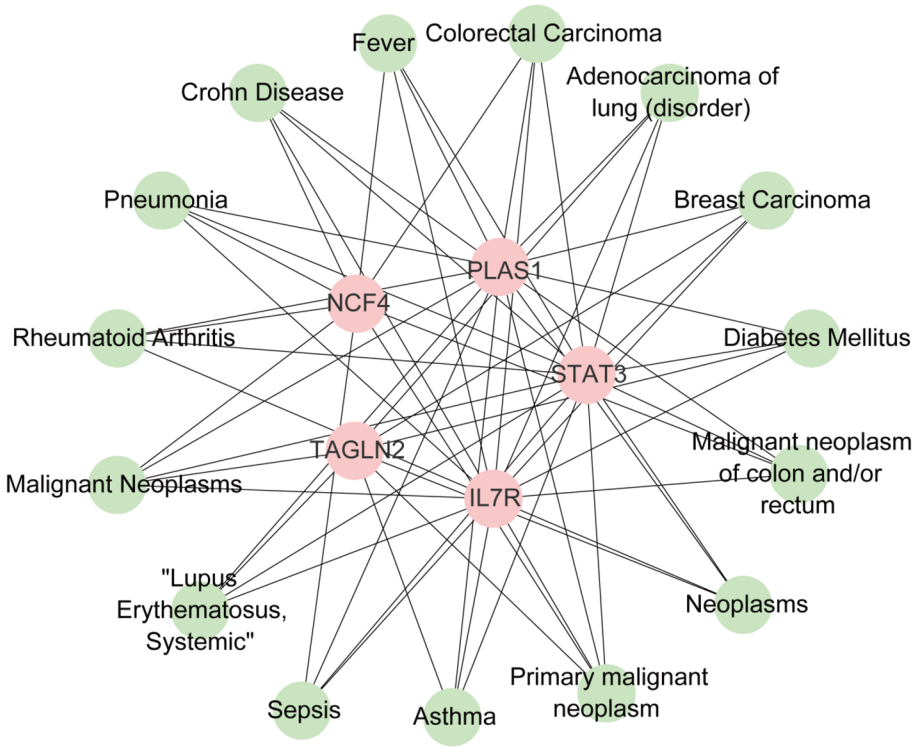


Fig. 9 Shows the construction of the network of STAT3 and its related hub genes and diseases

oligonucleotides that inhibit *STAT3* activity have been developed, and some of them have shown good effects in pre—clinical models. For example, the small—molecule inhibitor Stattic has been proven to selectively inhibit the phosphorylation of *STAT3*, thereby inducing the apoptosis of chronic myeloid leukemia cells without affecting normal cells [29]. Currently, clinical trials are investigating the efficacy of JAK inhibitors in treating chronic myeloid leukemia. This type of inhibitor can indirectly reduce the activation of *STAT3* [30]. In conclusion, the *STAT3* gene and its related signaling pathways play a vital role in the pathogenesis of chronic myeloid leukemia. Dysregulated activation of the *STAT3* signal is one of the characteristics of chronic myeloid leukemia, leading to uncontrolled proliferation and survival of leukemia cells [31]. In-depth exploration of the activation and functional role of *STAT3* in chronic myeloid leukemia is essential for the development of targeted therapies aimed at improving patient prognosis.

Bioinformatics methods play a key role in deeply understanding and analyzing the role of the *STAT3* signaling pathway in chronic myeloid leukemia. By analyzing the gene expression profiles of chronic myeloid leukemia patient samples in the GEOpublic database, it is found that the related genes of *STAT3* are generally up-regulated in chronic myeloid leukemia patients. To further reveal the specific role of *STAT3* in chronic myeloid leukemia, this study adopted algorithms such as machine learning and identified the hub genes related to *STAT3* in chronic myeloid leukemia as *NCF4, PLAS1, IL7R*, and *TAGLN2*. These genes may be closely related to the occurrence and development of chronic myeloid leukemia. The experimental results show that in K562 cells intervened by Stattic, the expression of *STAT3* is significantly down-regulated, cell proliferation is inhibited, the apoptosis rate increases, and the S—phase of the cell cycle is blocked. At the same time, the study also found that the expression of the *BCL-2* is significantly down—regulated, and the expression of the *BAX* is significantly up-regulated. This further verifies that the regulatory effect of *STAT3* on cell apoptosis is achieved by affecting the expression of *BCL-2* and *BAX* genes. This experimental result indicates that *STAT3* plays a crucial role in the regulation of proliferation and apoptosis of chronic myeloid leukemia cells.

In summary, this study explored the preliminary mechanism of action of *STAT3* on K562 cells and identified the hub genes *NCF4, PLAS1, IL7R*, and *TAGLN2* related to STAT3 in CML, providing more potential targets for the treatment of chronic myeloid leukemia. However, its limitation lies in the lack of verification with clinical samples. The significance of this study lies in identifying the hub genes *NCF4, PLAS1, IL7R,* and *TAGLN2* related to *STAT3* expression through bioinformatics analysis methods, and verifying the impact of *STAT3* on the proliferation, cycle, and apoptosis of K562 cells, providing a theoretical basis for gene therapy of leukemia and the search for drug targets.

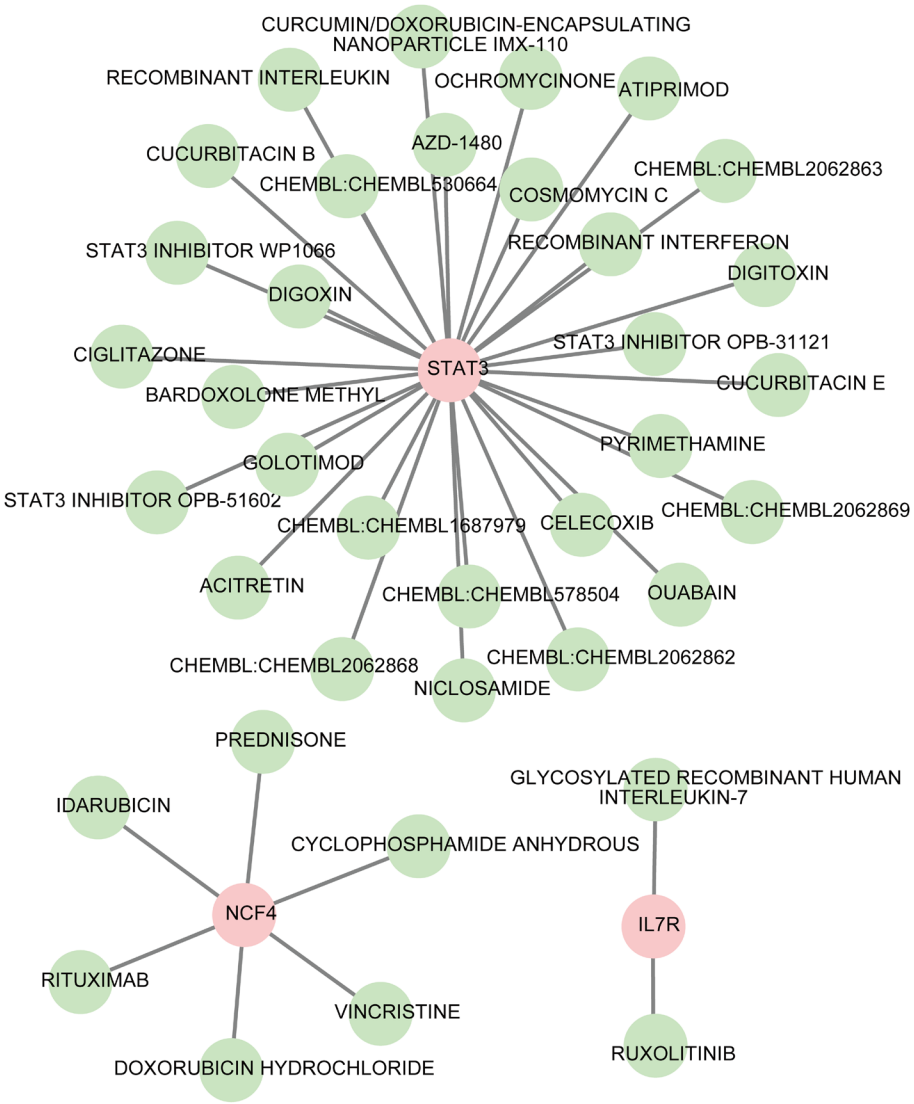


Fig. 10 *STAT3* and its associated hub genes drug prediction

5 Conclusion

This study has thoroughly investigated the mechanism of action of *STAT3* in K562 cells and successfully identified hub genes that are closely associated with *STAT3*. The aim of the study is to expand new targets for the treatment strategy of CML. Although the research findings have unveiled the crucial role of *STAT3* in the processes of K562 cell proliferation, cell cycle regulation, and apoptosis, and have provided a solid theoretical foundation for gene therapy and the discovery of drug targets for leukemia, this study still has certain limitations, mainly manifested in the lack of further verification using clinical samples. Additionally, the study found that the activity of *STAT3* is closely connected to certain specific signaling pathways, such as the JAK-STAT pathway, which offers a new direction for future research. The research team also noted that the role of *STAT3* may vary in different types of leukemia cells, suggesting the need for more personalized treatment strategies.

Author contributions Xiaoyun Feng designed the study and performed the analysis. Yufeng Qin and Yulong Feng performed the validation in the independent cohort. Yingquan Zhuo revised the manuscript. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the National Natural Science Foundation of China (Grant No. gzwkj2024-275).

Data availability In this study, the relevant codes for bioinformatics analysis are provided. For detailed information, please consult the (supplementary material file).

Declarations

Competing interests The authors declare no competing interests.

References

1. Minciacchi VR, Kumar R, Krause DS. Chronic myeloid leukemia: a model disease of the past, present and future. Cells. 2021;10(1):117.
2. Hai A, Kizilbash NA, Zaidi SH, et al. Differences in structural elements of Bcr-Abl oncoprotein isoforms in chronic myelogenous leukemia. Bioinformation. 2014;10(3):108–14.
3. Yassin MA, Ata F, Mohamed SF, et al. Ophthalmologic manifestations as the initial presentation of chronic myeloid leukemia: a review. Surv Ophthalmol. 2022;67(2):530–43.
4. Senapati J, Jabbour E, Kantarjian H, et al. Pathogenesis and management of accelerated and blast phases of chronic myeloid leukemia. Leukemia. 2023;37(1):5–17.
5. Shah NP, Bhatia R, Altman JK, et al. Chronic myeloid leukemia, version 2.2024, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw. 2024;22(1):43–69.
6. Saussele S, Richter J, Guilhot J, et al. Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): a prespecified interim analysis of a prospective, multicentre, non-randomised, trial. Lancet Oncol. 2018;19(6):747–57.
7. Huang L, Fu L. Mechanisms of resistance to EGFR tyrosine kinase inhibitors. Acta Pharm Sin B. 2015;5(5):390–401.
8. Yung Y, Lee E, Chu HT, et al. Targeting abnormal hematopoietic stem cells in chronic myeloid leukemia and philadelphia chromosome-negative classical myeloproliferative neoplasms. Int J Mol Sci. 2021;22(2):659.
9. Cao ZR, Chen XP, Hu J. Research advance on classic Wnt pathway in chronic myelogenous leukemia-review. Zhongguo Shi Yan Xue Ye Xue Za Zhi. 2020;28(1):350–3.
10. Bi L, Yu Z, Wu J, et al. Honokiol inhibits constitutive and inducible STAT3 signaling via PU.1-induced SHP1 expression in acute myeloid leukemia cells. Tohoku J Exp Med. 2015;237(3):163–72.
11. Fan Y, Mao R, Yang J. NF-κB and STAT3 signaling pathways collaboratively link inflammation to cancer. Protein Cell. 2013;4(3):176–85.
12. Patel SB, Nemkov T, Stefanoni D, et al. Metabolic alterations mediated by STAT3 promotes drug persistence in CML. Leukemia. 2021;35(12):3371–82.
13. Kim J, Lee HJ, Park JH, et al. Nilotinib modulates LPS-induced cognitive impairment and neuroinflammatory responses by regulating P38/STAT3 signaling. J Neuroinflam. 2022;19(1):187.
14. Xie L, Li Y, Tang W, Zhang Q, Luo C, Long X. Stattic alleviates pulmonary fibrosis in a mouse model of rheumatoid arthritis-relevant interstitial lung disease. Exp Biol Med. 2023;248(8):712–21.
15. Jafari S, Lavasanifar A, Hejazi MS, Maleki-Dizaji N, Mesgari M, Molavi O. STAT3 inhibitory stattic enhances immunogenic cell death induced by chemotherapy in cancer cells. Daru. 2020;28(1):159–69. https://doi.org/10.1007/s40199-020-00326-z.
16. Yuan C, Yuan M, Li W, et al. The STAT3 inhibitor stattic overcome bortezomib-resistance in multiple myeloma via decreasing PSMB6. Exp Cell Res. 2023;429(1):113634.
17. Guha P, Gardell J, Darpolor J, et al. STAT3 inhibition induces Bax-dependent apoptosis in liver tumor myeloid-derived suppressor cells. Oncogene. 2019;38(4):533–48.
18. Tang Y, Tian XC. JAK-STAT3 and somatic cell reprogramming. JAKSTAT. 2013;2(4):e24935.
19. Shin JE, Kim SH, Kong M, et al. Targeting FLT3-TAZ signaling to suppress drug resistance in blast phase chronic myeloid leukemia. Mol Cancer. 2023;22(1):177.
20. Zhou T, Medeiros LJ, Hu S. Chronic myeloid leukemia: beyond BCR-ABL1. Curr Hematol Malig Rep. 2018;13(6):435–45.
21. Cortes JE, Kim DW, Pinilla-Ibarz J, et al. Ponatinib efficacy and safety in Philadelphia chromosome-positive leukemia: final 5-year results of the phase 2 PACE trial. Blood. 2018;132(4):393–404.
22. Semenzato G, Calabretto G, Teramo A, et al. The constitutive activation of STAT3 gene and its mutations are at the crossroad between LGL leukemia and autoimmune disorders. Blood Cancer J. 2024;14(1):13.
23. Song J, Wang J, Tian S, et al. Discovery of STAT3 inhibitors: recent advances and future perspectives. Curr Med Chem. 2023;30(16):1824–47.
24. Al-Zubaidy HFS, Majeed SR, Al-Koofee DAF. Evaluation of Bax and BCL 2 genes polymorphisms in Iraqi women with breast cancer. Arch Razi Inst. 2022;77(2):799–808.
25. Yao K, Xing HC, Wu B, et al. Effect of TIEG1 on apoptosis and expression of Bcl-2/Bax and Pten in leukemic cell lines. Genet Mol Res. 2015;14(1):1968–74.

26. Chen YJ, Liu WH, Kao PH, et al. Involvement of p38 MAPK- and JNK-modulated expression of Bcl-2 and Bax in *Naja nigricollis* CMS-9-induced apoptosis of human leukemia K562 cells. Toxicon. 2010;55(7):1306–16.
27 Ghasemian M, Mahdavi M, Zare P, Ali HosseinpourFeizi M. Spiroquinazolinone-induced cytotoxicity and apoptosis in K562 human leukemia cells: alteration in expression levels of Bcl-2 and Bax. J Toxicol Sci. 2015;40(1):115–26.
28. Niehues T, von Hardenberg S, Velleuer E. Rapid identification of primary atopic disorders (PAD) by a clinical landmark-guided, upfront use of genomic sequencing. Allergol Select. 2024;8:304–23.
29. Bhansali RS, Rammohan M, Lee P, et al. DYRK1A regulates B cell acute lymphoblastic leukemia through phosphorylation of FOXO1 and STAT3. J Clin Invest. 2021;131(1):e135937.
30. Teramo A, Barilà G, Calabretto G, et al. STAT3 mutation impacts biological and clinical features of T-LGL leukemia. Oncotarget. 2017;8(37):61876–89.
31. Johnson DE, O'Keefe RA, Grandis JR. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. Nat Rev Clin Oncol. 2018;15(4):234–48.