Identification and validation of hub genes and molecular classifications associated with chronic myeloid leukemia

Fangmin Zhong, Fangyi Yao, Shuai Xu, Jing Zhang, Jing Liu* and Xiaozhong Wang*

Jiangxi Province Key Laboratory of Laboratory Medicine, Jiangxi Provincial Clinical Research Center for Laboratory Medicine, Department of Clinical Laboratory, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China

Background: Chronic myeloid leukemia (CML) is a kind of malignant blood tumor, which is prone to drug resistance and relapse. This study aimed to identify novel diagnostic and therapeutic targets for CML.

Methods: Differentially expressed genes (DEGs) were obtained by differential analysis of the CML cohort in the GEO database. Weighted gene co-expression network analysis (WGCNA) was used to identify CML-related co-expressed genes. Least absolute shrinkage and selection operator (LASSO) regression analysis was used to screen hub genes and construct a risk score model based on hub genes. Consensus clustering algorithm was used for the identification of molecular subtypes. Clinical samples and *in vitro* experiments were used to verify the expression and biological function of hub genes.

Results: A total of 378 DEGs were identified by differential analysis. 369 CML-related genes were identified by WGCNA analysis, which were mainly enriched in metabolism-related signaling pathways. In addition, CML-related genes are mainly involved in immune regulation and anti-tumor immunity, suggesting that CML has some immunodeficiency. Immune infiltration analysis confirmed the reduced infiltration of immune killer cells such as CD8+ T cells in CML samples. 6 hub genes (LINC01268, NME8, DMXL2, CXXC5, SCD and FBN1) were identified by LASSO regression analysis. The receiver operating characteristic (ROC) curve confirmed the high diagnostic value of the hub genes in the analysis and validation cohorts, and the risk score model further improved the diagnostic accuracy. hub genes were also associated with cell proliferation, cycle, and metabolic pathway activity. Two molecular subtypes, Cluster A and Cluster B, were identified based on hub gene expression. Cluster B has a lower risk score, higher levels of CD8+ T cell and activated dendritic cell infiltration, and immune checkpoint expression, and is more sensitive to commonly used tyrosine kinase inhibitors. Finally, our clinical samples validated the expression and diagnostic efficacy of hub genes, and the knockdown of LINC01268 inhibited the proliferation of CML cells, and promoted apoptosis.

Conclusion: Through WGCNA analysis and LASSO regression analysis, our study provides a new target for CML diagnosis and treatment, and provides a basis for further CML research.

Introduction

Chronic myeloid leukemia is a malignant tumor that affects the blood and bone marrow (1). It is mainly induced by the BCR-ABL1 fusion gene, which encodes a protein with strong tyrosine kinase activity and activates various signaling pathways (2). At present, the main therapeutic drugs for CML are tyrosine kinase inhibitors (TKIs) targeting BCR-ABL1 (2). The development of the first-generation TKI imatinib (IM) has changed the treatment status of CML, and the prognosis of patients has been significantly improved (3). It is widely used and has a good therapeutic effect. However, due to the existence of escape mechanisms, tumor cells often develop resistance to kinase drugs, leading to the malignant progression of the disease, which seriously affects the health of patients (4). In addition, the long-term use of TKI will also produce many complications, affecting the quality of life of patients (5). Therefore, there is an urgent need to identify novel molecular targets for the diagnosis and treatment of CML.

With the progress and development of sequencing technology, bioinformatics has been widely used to explore the genetic changes of tumors, and to find new targets for early diagnosis and therapeutic intervention of tumors. The Gene Expression Omnibus (GEO) database contains gene expression profiles of various diseases and tumor samples and corresponding clinical information, which can be used for in-depth analysis (6). Weighted gene co-expression network analysis (WGCNA) is a bioinformatics tool to screen genes with similar expression patterns related to disease phenotypes by constructing free-scale gene co-expression networks (7). The reliability of this method has been widely verified (8–10), and to a large extent, it overcomes the limitations caused by only focusing on differentially expressed genes (DEGs). Therefore, hub genes that are highly correlated with clinical phenotypes can be defined as potential biomarkers and therapeutic targets.

In this study, we systematically analyzed the CML dataset GSE13159 in the GEO database, combined with differential expressed expression analysis and WGCNA analysis, identified a set of co-expressed genes significantly associated with CML, and determined the biological functions of these genes by enrichment analysis. Subsequently, the least absolute shrinkage and selection operator (LASSO) analysis was used to screen out signature genes that had high diagnostic value for CML and could predict treatment response in CML patients. We also identified two molecular subtypes with distinct immune landscapes based on hub gene expression. Finally, the diagnostic performance of the risk score model constructed by hub genes was further improved. These signatures

were validated using an additional public cohort and our clinical real-world cohort. Therefore, these findings will help reveal more underlying mechanisms of CML, as well as the potential value of these targets in CML treatment.

Materials and methods

Data acquisition and processing

We downloaded the CML data sets (GSE13159, GSE144119) from the GEO database. GSE13159 contains 76 CML samples and 74 normal samples, and we normalized the original "cel" files. GSE144119 contained 48 newly diagnosed CML samples and 32 remission CML samples, as well as 17 normal samples, and the data were converted to transcripts per kilobase million (TPM) values for subsequent analyses. GSE13159 was used as the analysis cohort, and GSE144119 was used for subsequent validation. The normalized RNA-seq data (TPM values) of 173 TCGA-LAML (The Cancer Genome Atlas-Acute Myeloid Leukemia) samples containing clinical information were downloaded from the UCSC XENA database (https://xenabrowser.net/datapages/).

Pathway activity assessment and function enrichment analysis

The gene set variation analysis (GSVA) algorithm was used to calculate the enrichment score of the gene set to quantify the activity of the corresponding biological process or signaling pathway. The GSVA score was calculated based on the overall position of the gene set genes in the expression ranking of all genes, and the higher the overall expression level of these genes, the higher the GSVA score. KEGG enrichment analysis was used to analyze the function of phenotypic-related genes identified by WGCNA. We perform these analyses in the "clusterProfiler" package (11).

Analysis of immune cell infiltration

CIBERSORT algorithm based on support vector regression analysis was used to analyze the infiltration proportion of 22 kinds of immune cells in CML samples (12).

Weighted correlation network analysis

WGCNA is a tool for assessing gene expression correlations and visualizing co-expression networks. The "WGCNA" software package was used to identify CML-related genes in the GSE13159 cohort. Pearson correlation analysis was used to form an adjacency matrix for all matched genes, and the scale-free topology of the adjacency matrix was realized based on the optimal soft threshold power. Then, we further transform the adjacency matrix into a topological overlap matrix (TOM). Based on the TOM difference measure, the minimum module size was set to 30, the cutting height was set to 0.2, and the genes with similar expression patterns were divided into the same modules through average linkage hierarchical clustering. Then, the correlation between module characteristic genes (MEs) and CML was assessed, and the modules that met the purpose of the study were determined according to the degree of correlation.

Identification of DEGs between normal and CML samples

The empirical Bayesian approach via the "limma" package was used to determine DEGs between normal and CML samples (13). Genes with adjusted P-values < 0.05 and |logFC| > 1 were considered significantly different.

Construction of risk score model

Overlapping genes of CML-related genes and DEGs identified by WGCNA were used for the identification of CML hub genes. Then, the LASSO regression algorithm was used for dimensionality reduction analysis to screen out the most related genes with CML (14). In addition, based on the correlation of hub genes, LASSO regression analysis assigned a coefficient to each gene, and the expression of each gene was multiplied by its coefficient and added to obtain a risk score, which was used to analyze the diagnostic value of the combination of hub genes in CML. Risk score = NME8 × 1.160 + DMXL2 × 0.853 + CXXC5 × -0.126 + SCD × 0.610 + FBN1 × 0.405, where gene ID refers to the expression value of each gene.

Identification of molecular subtypes based on hub genes

Consensus cluster analysis was performed to identify CML molecular subtypes based on hub gene expression using the "consensusclusterplus" package. Clustering was performed for 1000 iterations to ensure reliable and stable results. t-distributedstochastic neighbor embedding (t-SNE) was used to validate the classification.

Construction of competing endogenous RNA network

Target miRNAs of hub genes were found in the miRTarBase, miRDB, and TargetScan databases. Perl programming language was used to perform the prediction analysis of the target lncRNAs of these miRNAs in the miRcode database.

Prediction of treatment response for different molecular subtypes

The half-maximal inhibitory concentrations (IC$_{50}$) of different CML samples to therapeutic drugs were predicted based on drug response data of blood cell lines from the Cancer Genome Project (CGP) database (https://cancer.sanger.ac.uk/cosmic) via the "pRRophetic" package. Tumor Immune Dysfunction and Exclusion (TIDE, http://tide.dfci.harvard.edu/) was considered a good predictor of immunotherapeutic response for molecular subtypes.

Clinical sample collection

CML samples and normal samples were collected for sequencing analysis in accordance with the Declaration of Helsinki and institutional guidelines, and informed consent was obtained from each patient and healthy volunteer who had undergone the appropriate workup. Our study was approved by the Ethics Committee of the Second Affiliated Hospital of Nanchang University, and sample processing was performed according to the norms. We collected samples from 5 untreated patients with newly diagnosed CML and 5 normal samples from healthy volunteers. The methods and details of sample collection, next-generation sequencing, and processing procedures were described in our previous report (15). Moreover, peripheral blood samples from 15 CML patients and 15 normal controls were collected for quantitative real-time polymerase chain reaction (RT-qPCR) assay to detect hub gene expression. RT-qPCR was performed using a Japanese TAKARA kit on an ABI7500 instrument. The primers are shown in Supplementary Table S1.

Cell culture and detection of cell proliferation and apoptosis

The CML cell line K562 was cultured in RPMI1640 medium supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin in a humidified atmosphere incubator at 37°C with 5% CO$_2$. Two different siRNAs targeting LINC01268 (si-LINC01268) and control siRNA (si-NC) were procured from Ribobio (China) and transfected into K562 cells using Lipofectamine 3000 (Thermofisher Scientific) (Supplementary Table S1). RT-qPCR was employed to assess the transfection efficiency. Cell proliferation was evaluated using the Cell Counting Kit-8 (CCK-8). For the CCK8 assay, a total of 2×10^4 cells from various treatment groups were seeded in individual wells of a 96-well plate, with each group being repeated five times. Subsequently, at time points of 0, 24, 48, and 72 hours, respectively, 10 µl of CCK8 solution was added to each well. After incubation at 37°C for two hours, the optical density (OD) value at a wavelength of 450 nm was measured using a microplate reader. Apoptosis assays were performed by staining the cells with Annexin V-PE/7-AAD Apoptosis Assay Kit followed by analysis on a flow cytometer.

Statistical analysis

We performed the Wilcoxon rank sum test and the Kruskal-Wallis test to determine differences between two or more groups, respectively. The "survminer" package divides patients into high- and low-gene expression groups based on the cut-off point at the minimum p-value of the log-rank test, and the Kaplan-Meier survival curve analysis was used to analyze survival differences between the two groups. The receiver operating characteristic (ROC) curve was used to analyze the diagnostic efficacy of genes. A two-sided P value < 0.05 was considered statistically significant.
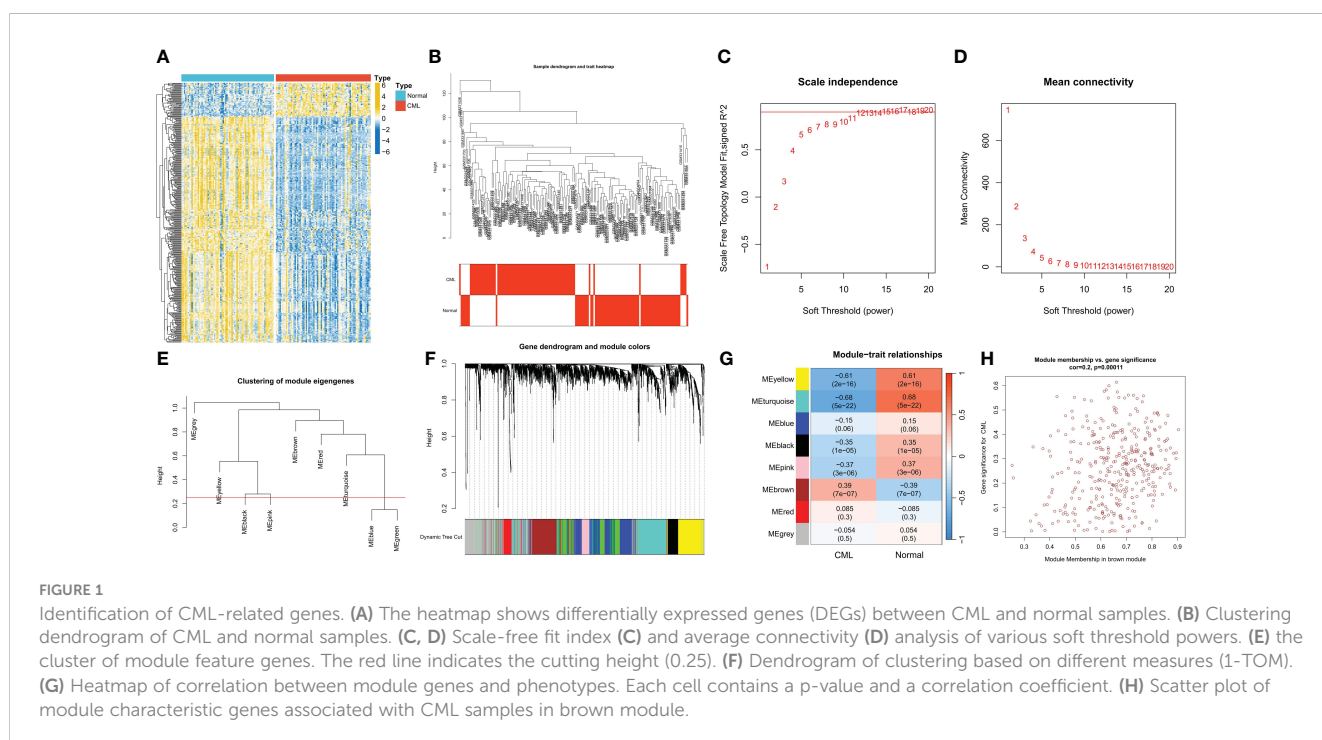
Results

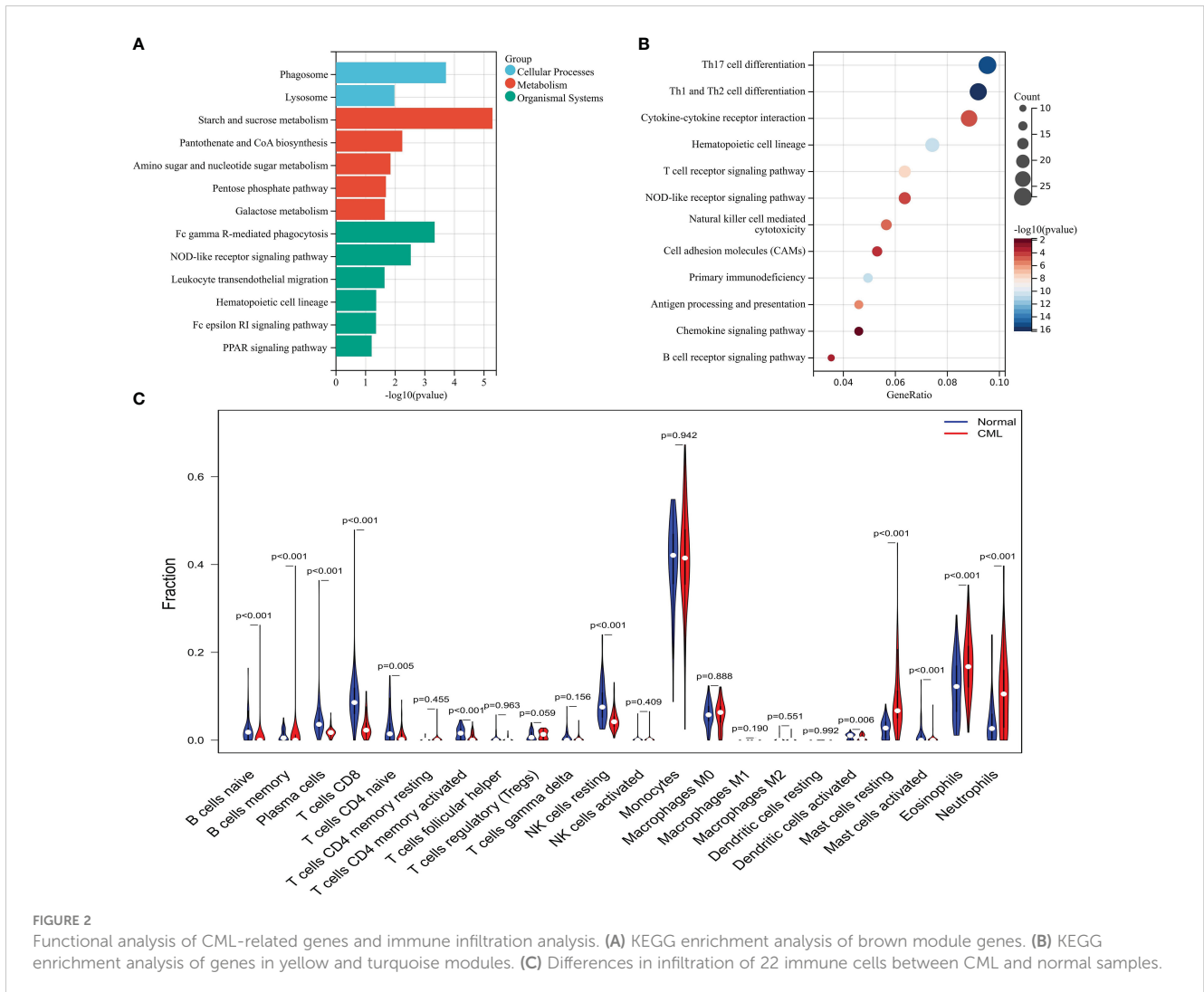### CML-related genes were identified by WGCNA analysis

We first performed differential expression analysis between CML and normal samples and obtained a total of 378 DEGs. Heatmap analysis showed that more DEGs were down-regulated in CML (Figure 1A). We further performed WGCNA analysis to identify more CML-related genes. The cluster tree diagram showed the clustering characteristics of the samples, and the CML samples had a high degree of discrimination from the normal samples (Figure 1B). Figures 1C, D show the scale-free fit exponent and average connectivity analysis for various soft threshold powers. We set cut height = 0.25 to merge the blue and green module feature genes (Figure 1E). According to the optimal soft threshold power β = 12 (unscaled R2 = 0.9), the 5000 genes with the highest standard deviation were divided into eight independent co-expression modules (Figure 1F). The correlogram of module-trait relationships showed that the brown module, which contains 369 genes, had the highest correlation with CML (Figures 1G, H) (Supplementary Table S1). We also found that the blue, green, yellow, black, and pink modules were negatively correlated with CML, and these results were associated with the downregulated expression of most genes in CML.

Functional analysis of CML-related genes

The brown module genes were mainly related to metabolic-related signaling pathways such as Starch and sucrose metabolism, Pantothenate and CoA biosynthesis, Amino sugar and nucleotide sugar metabolism, Pentose phosphate pathway, and Galactose metabolism (Figure 2A). While yellow and turquoise module genes were negatively associated with CML, these genes were mainly enriched in immune-related signaling pathways such as Th17 cell differentiation, Th17 cell differentiation, Cytokine-cytokine receptor interaction, and Hematopoietic cell lineage, T cell receptor signaling pathway, NOD-like receptor signaling pathway, Natural killer cell mediated cytotoxicity (Figure 2B). These results indicate that CML has stronger metabolic activity and some immunodeficiency. Immune infiltration analysis showed that CML samples had fewer naive and memory B cells, plasma cells, CD8+ T cells, naive CD4+ T cells, activated memory CD4+ T cells, resting NK cells, and activated dendritic cells, and contained more regulatory T cells (Tregs) than normal samples (Figure 2C), which confirm the immunosuppressive features evident in CML samples.

FIGURE 1
Identification of CML-related genes. **(A)** The heatmap shows differentially expressed genes (DEGs) between CML and normal samples. **(B)** Clustering dendrogram of CML and normal samples. **(C, D)** Scale-free fit index **(C)** and average connectivity **(D)** analysis of various soft threshold powers. **(E)** the cluster of module feature genes. The red line indicates the cutting height (0.25). **(F)** Dendrogram of clustering based on different measures (1-TOM). **(G)** Heatmap of correlation between module genes and phenotypes. Each cell contains a p-value and a correlation coefficient. **(H)** Scatter plot of module characteristic genes associated with CML samples in brown module.

FIGURE 2
Functional analysis of CML-related genes and immune infiltration analysis. **(A)** KEGG enrichment analysis of brown module genes. **(B)** KEGG enrichment analysis of genes in yellow and turquoise modules. **(C)** Differences in infiltration of 22 immune cells between CML and normal samples.

Identification of CML hub genes

We intersected DEGs and WGCNA brown module genes and obtained 17 overlapping genes (Figure 3A), and the correlation coefficients of these genes with the brown module in WGCNA and with CML samples were greater than 0.4 (Supplementary Table S2), indicating that they were significantly positively correlated with both CML and brown module. LASSO regression analysis further reduced the dimension and screened out 6 hub genes most related to CML, which were LINC01268, NME8, DMXL2, CXXC5, SCD, and FBN1 (Figures 3B, C). Boxplots showed that LINC01268, NME8, DMXL2, SCD, and FBN1 were significantly up-regulated and CXXC5 was significantly down-regulated in CML samples compared with normal samples (Figure 3D).
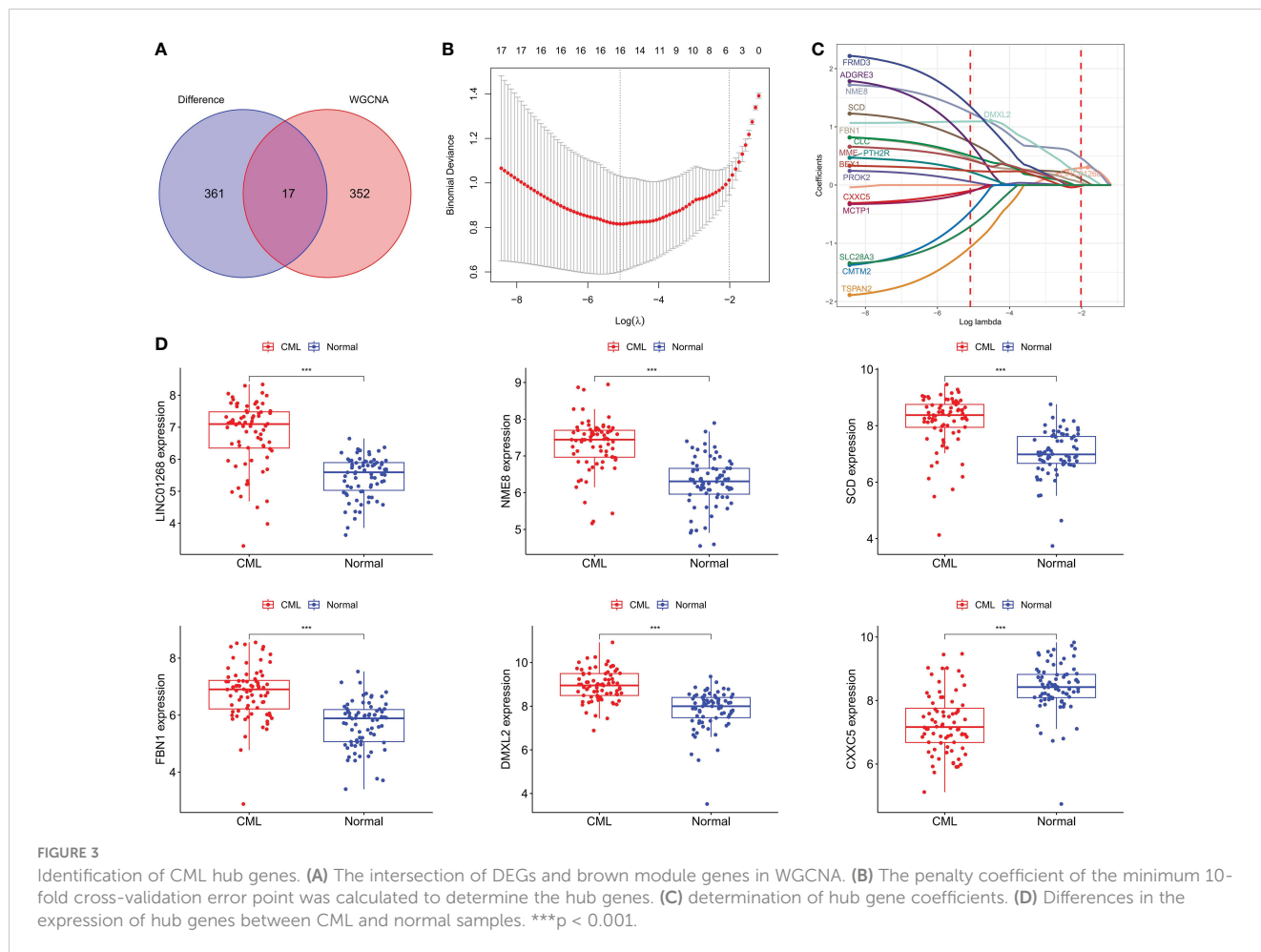
Diagnostic value and prognostic correlation of CML hub genes

We further analyzed the predictive value of CML hub genes for CML. ROC curve analysis showed that all 6 hub gens had high AUC

values for the diagnosis of CML, among which LINC01268 was 0.864 (95%CI: 0.796-0.924), NME8 was 0.869 (95%CI: 0.808-0.924), DMXL2 was 0.866 (95%CI: 0.805-0.91), CXXC5 was 0.831 (95%CI: 0.761-0.895), SCD was 0.856 (95%CI: 0.790-0.919), and FBN1 was 0.836 (95%CI: 0.767-0.900) (Figure 4A). In addition, considering that approximately 70% of CML cases in blast crisis progress to AML, we analyzed the prognostic predictive value of 6 hub genes in the TCGA-AML cohort. High expression groups of LINC01268, SCD, FBN1, and CXXC5 had significantly shorter overall survival than their low expression groups, respectively, while high expression groups of NME8 and DMXL2 showed better prognosis, but there was no statistical difference (Figure 4B).

Validation of the diagnostic value of CML hub genes

The GSE144119 cohort contains samples from newly diagnosed and treatment-remission CML. Encourageously, the results of the differential analysis were consistent with the GSE13159 cohort, in which NME8, DMXL2, SCD, and FBN1 expression was

FIGURE 3
Identification of CML hub genes. **(A)** The intersection of DEGs and brown module genes in WGCNA. **(B)** The penalty coefficient of the minimum 10-fold cross-validation error point was calculated to determine the hub genes. **(C)** determination of hub gene coefficients. **(D)** Differences in the expression of hub genes between CML and normal samples. ***p < 0.001.

significantly increased and CXXC5 expression was significantly decreased in newly diagnosed (chronic phase) CML patients (The expression of LINC01268 was not detected). These hub genes also had predictive value for CML treatment remission. The expression levels of NME8, DMXL2, SCD, and FBN1 were significantly decreased in CML treatment-remission patients, while the CXXC5 expression level was significantly increased, and they all returned to normal control levels. ROC curve analysis confirmed the diagnostic value of these hub genes in CML (Figure 5A). The AUC values of NME8, SCD, FBN1, DMXL2, and CXXC5 were 0.906 (95% CI: 0.836-0.960), 0.958 (95% CI: 0.908-0.995), 0.933 (95% CI: 0.870-0.980), 0.795 (95% CI: 0.695-0.878), and 0.932 (95% CI: 0.868-0.982), respectively (Figure 5B). In our clinical cohort, we confirmed that SCD and FBN1 expression was significantly upregulated CXXC5 was significantly downregulated in CML, and NME8 and DMXL2 expression were not significantly different due to the small sample size (Figure 6A).

Potential biological mechanisms of CML hub genes

To better explore the biological functions of CML hub genes, we analyzed their correlation with tumor marker pathway activity and
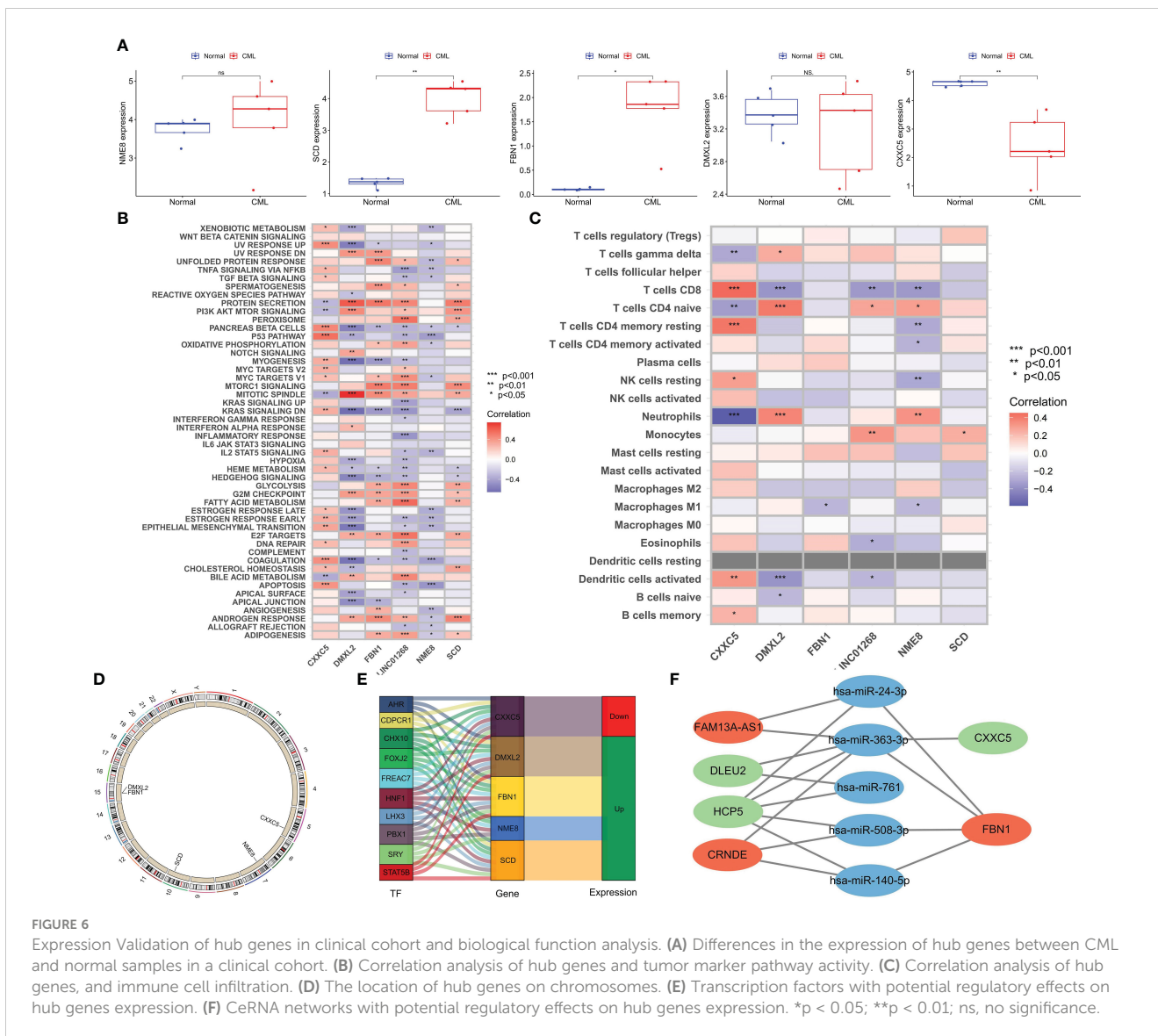
immune cell infiltration. CXXC5 expression was related to P53 PATHWAY, DNA REPAIR, MYC TARGETS, and APOPTOSIS, and may be involved in the regulation of CML cell proliferation. DMXL2 was positively correlated with cell cycle-related pathways such as MITOTIC SPINDLE, and G2M CHECKPOINT. FBN1, LINC01268, and SCD were related to the metabolic pathway activity of MTORC1 SIGNALING, GLYCOLYSIS, FATTY ACID METABOLISM, ADIPOGENESIS (Figure 6B). The expression of NME8 was negatively correlated with the activity of most tumor marker pathways. In addition, CXXC5 expression was positively correlated with infiltration of CD8+ T cells, resting memory CD4+ T cells, resting NK cells, activated dendritic cells, and memory B cells, suggesting that CXXC5 may be involved in CML anti-tumor immunity (Figure 6C). Figure 6D shows the location of five hub genes in chromosomes. In addition, we identified a group of transcription factors with potential regulatory effects on hub genes (Figure 6E). According to the construction of the CeRNA network (Figure 6F), lncRNA FAM13A-AS1 with upregulated expression may promote the expression of FBN1 by competitively binding hsa-miR-24-3p and hsa-miR-363-3p. lncRNA CRNDE may promote the expression of FBN1 by binding hsa-miR-363-3p, hsa-miR-508-3p and hsa-miR-140-5p. The downregulation of lncRNAs DLEU2 and HCP5 may reduce the binding of miR-363-3p, thereby inhibiting the expression of CXXC5.
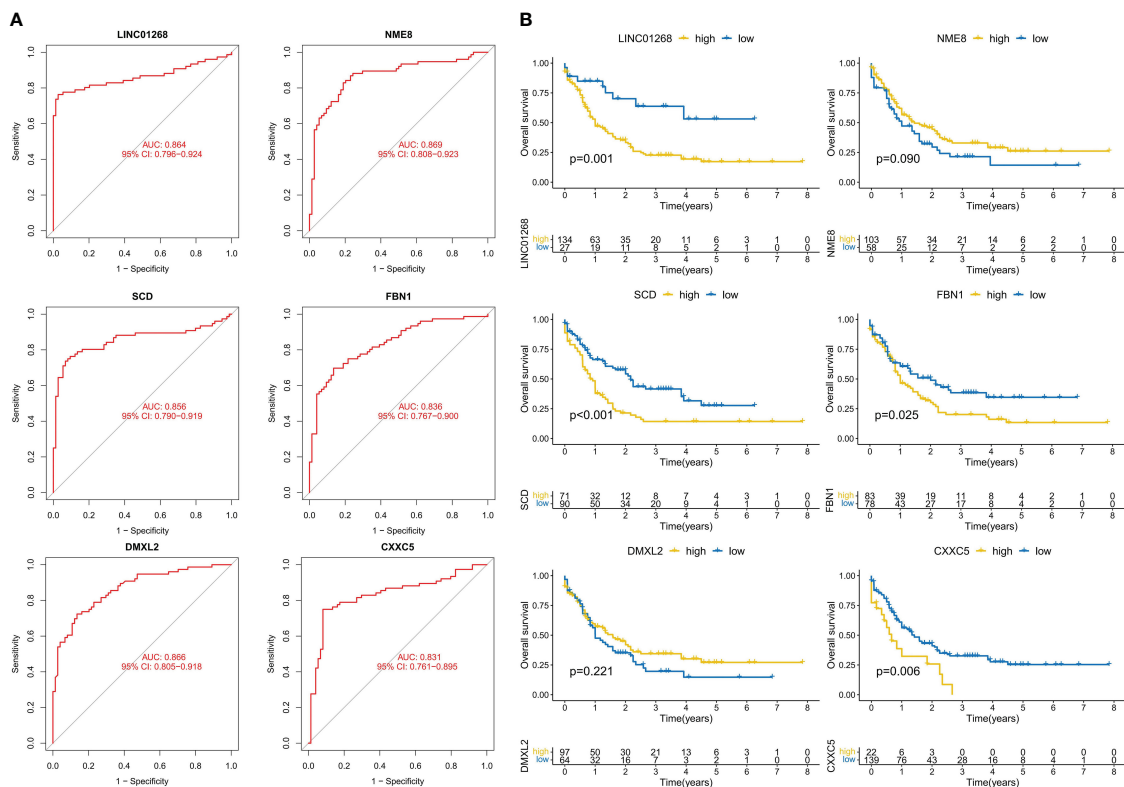
FIGURE 4

Analysis of the diagnostic and prognostic value of hub genes. **(A)** ROC curve analysis of hub genes. **(B)** K-M curve analysis of hub genes.
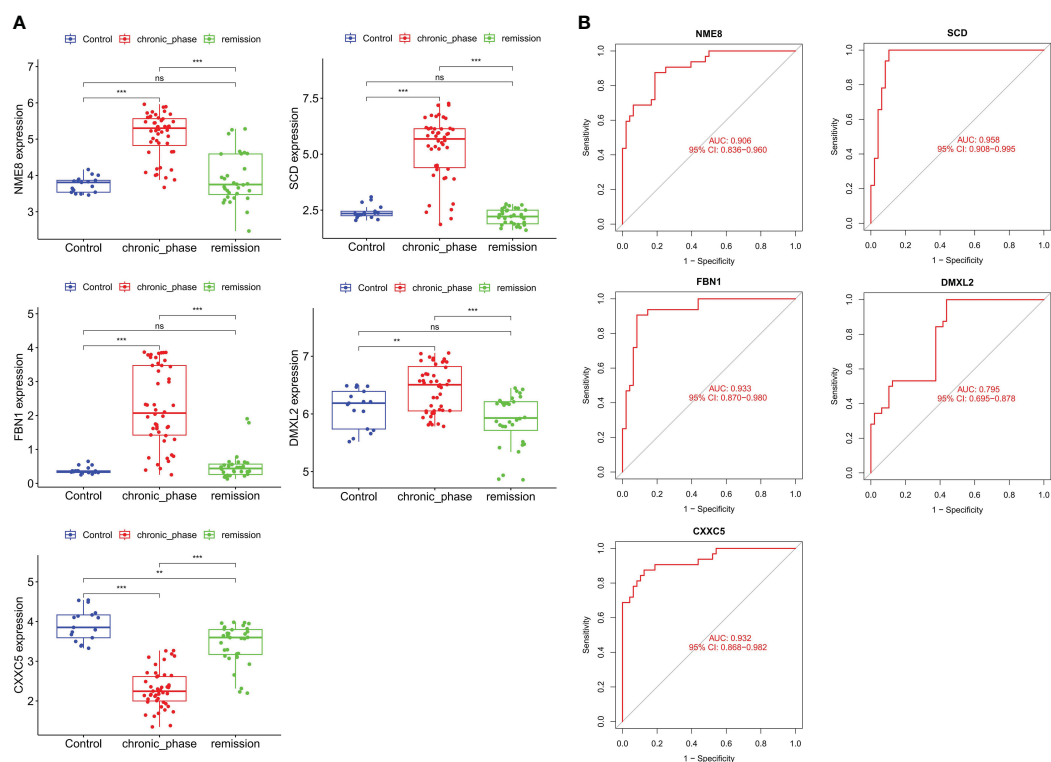
FIGURE 5

Validation of the expression and diagnostic value of hub genes in the validation cohort. **(A)** Differences in the expression of hub genes in normal samples, newly diagnosed CML samples, and treatment-remission samples. **(B)** ROC curve analysis of hub genes. **p < 0.01; ***p < 0.001; ns, no significance.

FIGURE 6
Expression Validation of hub genes in clinical cohort and biological function analysis. **(A)** Differences in the expression of hub genes between CML and normal samples in a clinical cohort. **(B)** Correlation analysis of hub genes and tumor marker pathway activity. **(C)** Correlation analysis of hub genes, and immune cell infiltration. **(D)** The location of hub genes on chromosomes. **(E)** Transcription factors with potential regulatory effects on hub genes expression. **(F)** CeRNA networks with potential regulatory effects on hub genes expression. *p < 0.05; **p < 0.01; ns, no significance.
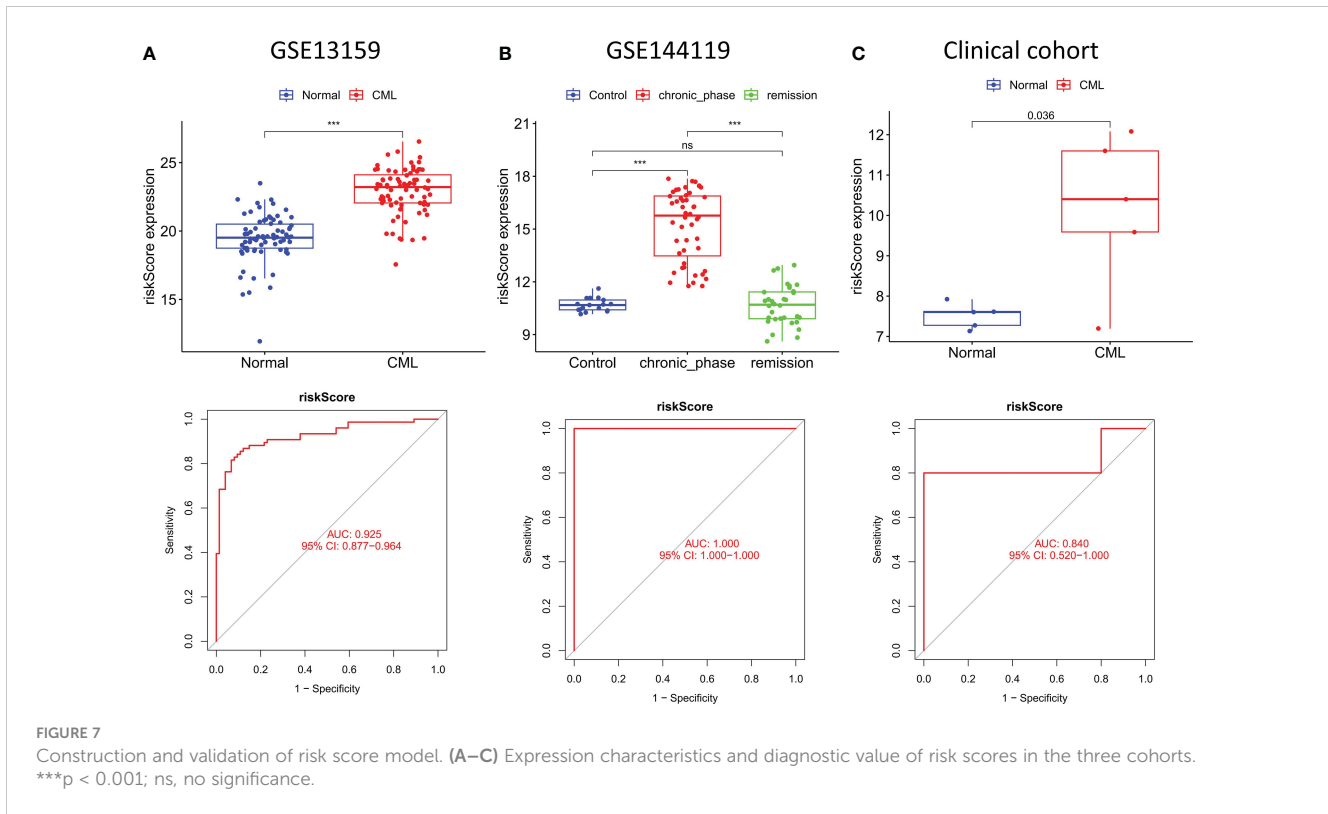
The construction of a risk score model can further improve the diagnostic value of hub genes

To better improve the diagnostic value of hub genes, we used LASSO regression analysis to construct a risk score model for 5 genes shared by the three cohorts. All three cohorts observed significantly higher risk scores in CML samples than in normal samples, and risk scores in patients in remission tended to be normal (Figures 7A–C). ROC curve analysis showed that the diagnostic AUC values in the GSE13159 cohort, GSE144119 cohort, and clinical cohort were 0.925 (95% CI: 0.877-0.964), 1.000 (95% CI: 1.000-1.000) and 0.840 (95% CI: 0.520-1.000), respectively, confirming that the diagnostic value of risk score of hub genes combination was further improved.

Molecular subtypes identified based on hub genes and prediction of treatment response

We performed cluster analysis of CML samples based on hub gene expression and identified two distinct molecular subtypes (Cluster A and Cluster B) (Figure 8A). The t-SNE algorithm verified the reliability of the clustering (Figure 8B). Compared with Cluster B, LINC01268, DMXL2, SCD, and FBN1 were up-regulated and CXXC5 was down-regulated in Cluster A (Figure 8C). Cluster A also had a significantly higher risk score than Cluster B (Figure 8D). Immune infiltration analysis showed that the infiltration levels of CD8+ T cells and activated NK cells were significantly higher in Cluster B than in Cluster A (Figure 8E). The expression of immune checkpoints PD-L1, CTLA4, HAVCR2, and PD-1 was also

FIGURE 7
Construction and validation of risk score model. **(A–C)** Expression characteristics and diagnostic value of risk scores in the three cohorts.
***p < 0.001; ns, no significance.

significantly up-regulated in Cluster B (Figure 8F). In addition, the TIDE score of Cluster B was significantly higher than that of Cluster A (Figure 8G), indicating significant immunosuppression in Cluster B. We also compared the activity of tumor-marker gene sets in the two molecular subtypes (Figure 8E). We found metabolic and cell proliferation-related pathways such as MYC targets V1, oxidative phosphorylation, G2M checkpoint, E2F targets, mTORC1 signaling and fatty acid metabolism were more active. In Cluster B, the enrichment scores of cancer-promoting pathways such as hedgehog, epithelial-mesenchymal transition, and TNFA signaling via NFKB were higher (Figure 8H). We then predicted the response of different molecular subtypes to TKIs commonly used for CML treatment, and the results showed that Cluster B patients had higher therapeutic sensitivity to imatinib, nilotinib, bosutinib, and dasatinib. Moreover, there was a significant positive correlation between the risk score and the IC$_{50}$ of the four drugs, that is, the higher the risk score, the less sensitive the treatment to the four drugs (Figure 8I).

Expand clinical sample size to validate the expression of hub genes and confirm the oncogenic role of LINC01268
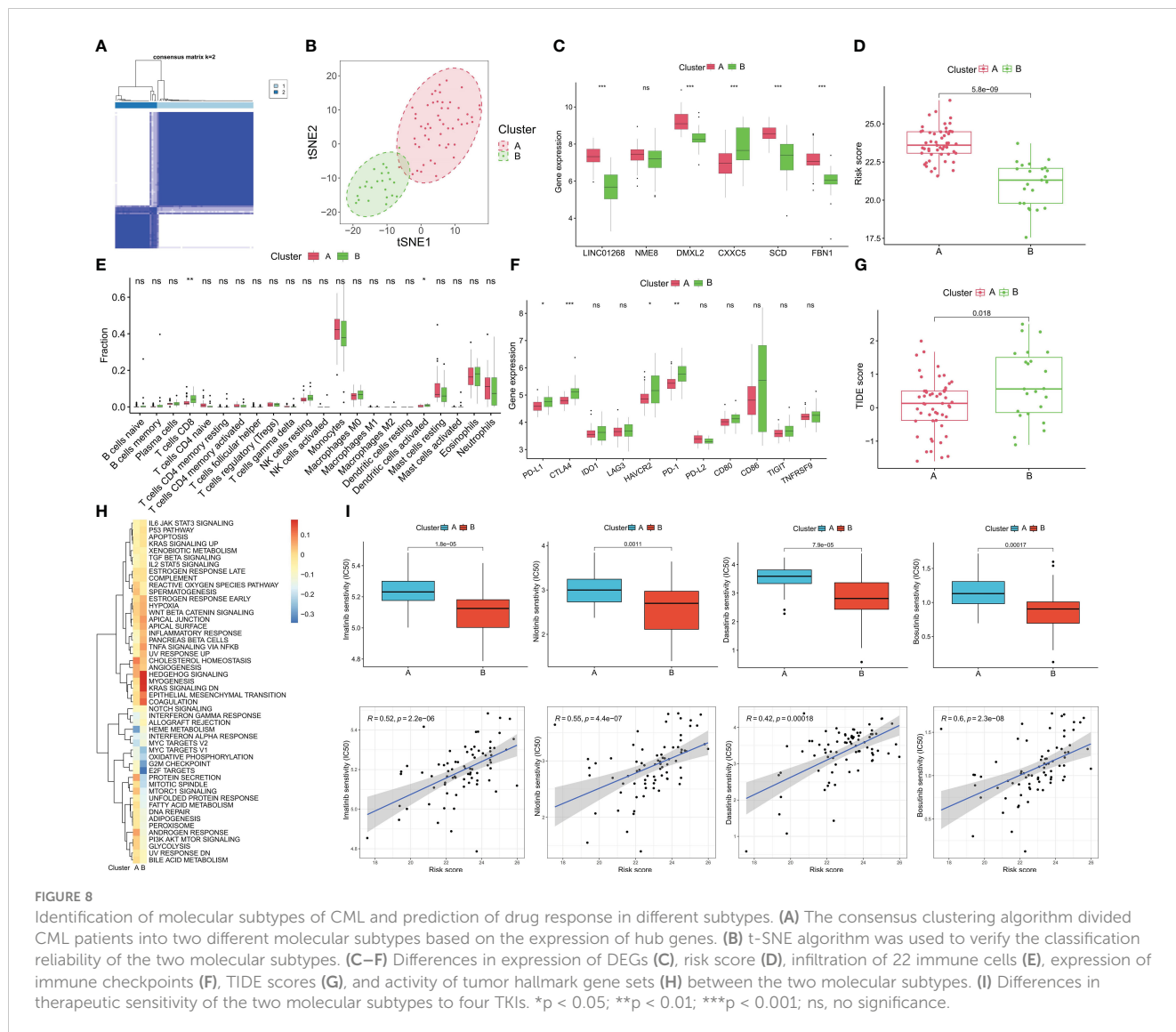
The expression of hub genes was validated by RT-qPCR in expanded clinical samples. Encouragingly, the results also confirmed that LINC01268, NME8, DMXL2, SCD, and FBN1 were up-regulated while CXXC5 was down-regulated in CML samples (Figure 9A). Previous studies have shown that DMXL2, NME8, and FBN1 primarily exert oncogenic roles through mutations and splice variants (16–18); moreover, the role of SCD

in CML has also been reported previously (19). Therefore, we chose to initially explore the biological function of LINC01268 in CML cells. The expression of LINC01268 was significantly inhibited by siRNA (Figure 9B). CCK8 assay showed that compared with the si-NC group, the proliferation ability of CML cells in the si-LINC01268 group was significantly reduced (Figure 9C). Moreover, the apoptosis rate of the si-LINC01268 group was higher than that of the si-NC group (Figures 9D–F). These results reveal the oncogenic role of LINC01268 and its potential as a therapeutic target for CML.

Discussion

The development and application of TKIs have significantly improved the prognosis of CML patients, but these drugs can only delay the progression of the disease, and cannot be used as a curative treatment (2). Due to the existence of resistance mechanisms, patients inevitably relapse (5). Therefore, it is particularly important to explore more potential therapeutic targets and markers for disease prediction and progression assessment in CML. In this study, we focused on the diagnostic markers of CML and their underlying biological mechanisms. Based on the DEGs between CML and normal samples and the CML-related genes identified by WGCNA analysis, we used LASSO regression analysis to screen out 6 hub genes (LINC01268, NME8, DMXL2, CXXC5, SCD, and FBN1).

We also focused on the co-expressed gene network identified by WGCNA analysis. The results showed that the brown module was significantly positively correlated with CML (Cor=0.39, P=7e-07). It

FIGURE 8

Identification of molecular subtypes of CML and prediction of drug response in different subtypes. **(A)** The consensus clustering algorithm divided CML patients into two different molecular subtypes based on the expression of hub genes. **(B)** t-SNE algorithm was used to verify the classification reliability of the two molecular subtypes. **(C–F)** Differences in expression of DEGs **(C)**, risk score **(D)**, infiltration of 22 immune cells **(E)**, expression of immune checkpoints **(F)**, TIDE scores **(G)**, and activity of tumor hallmark gene sets **(H)** between the two molecular subtypes. **(I)** Differences in therapeutic sensitivity of the two molecular subtypes to four TKIs. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$; ns, no significance.

reflects the correlation of the module as a whole with the CML phenotype. Although this correlation did not reach an exceptionally high level, a coefficient close to 0.4 suggests its reliability to some extent. Thus, it can be inferred that the brown module partially reflects gene co-expression patterns in CML transcriptome while uncovering underlying biological mechanisms. We found that the brown module genes positively correlated with CML were enriched in a variety of metabolic pathways, revealing the more active metabolic characteristics of CML cells. Several studies confirmed that targeting mitochondrial oxidative phosphorylation and glucose uptake is a potential therapeutic target for CML (20, 21). Most of the module genes negatively correlated with CML were involved in immune regulation and immune cell activation. Subsequent analysis showed that the infiltration of immune killer cells such as CD8+ T cells was significantly reduced in CML samples, confirming the immune deficiency characteristics. Cayssials et al. found that the sustained treat-free remission of CML was associated with an increased frequency of innate CD8+ T cells (22), and Harada et al. revealed that the inhibition of differentiation of dendritic

cells in the hematopoietic microenvironment, as well as the up-regulation of immune checkpoint expression such as PD-L1, were responsible for the impairment of CML immune function (23). Based on this, we believe that targeted inhibition of metabolism and enhancement of immune response are important strategies for CML treatment.

It is worth noting that Figure 1G illustrates the association between module membership (MM) and gene significance (GS), it reflects the association of individual genes in the module with the module (x-axis, MM) and with the CML phenotype (y-axis, GS). If the correlation between MM and GS is high, the higher the correlation between the module gene and the module, the higher the correlation between the module gene and the CML phenotype, showing an overall distribution trend. We further calculated the correlation coefficient between these two types of coefficients; although Cor=0.2 with P=0.00011 indicates a weak positive relationship, it still signifies statistical significance. In this scatterplot analysis, we focused on points with strong correlations with both MM and GS. The correlation coefficients for both GS and
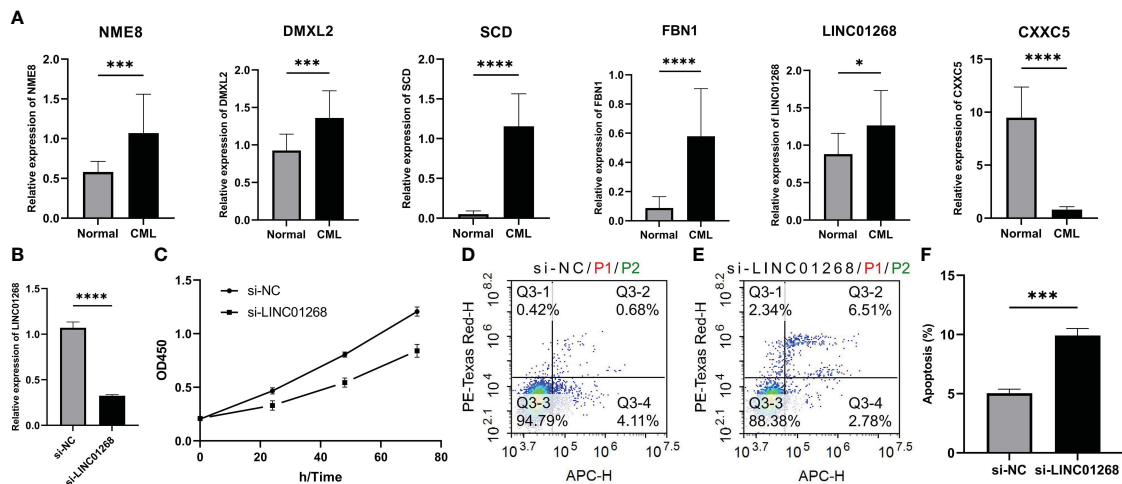
FIGURE 9
Expression characteristics of hub genes and its relationship with malignant phenotypes of CML cells. **(A)** Differences in mRNA expression of hub genes in peripheral blood samples from 15 CML patients and 15 normal controls. **(B)** mRNA expression level of LINC01268 in K562 cells in LINC01268 knockdown group (si- LINC01268) and control group (si-NC). **(C)** Absorbance at 450nm wavelength after CCK8 treatment in different LINC01268 treatment groups at different time nodes. **(D–F)** Apoptosis levels in different LINC01268 treatment groups. (*P < 0.05; *** P < 0.001; **** P < 0.0001).

MM of the 17 overlapping genes shared by differentially expressed genes and brown module genes were found to be greater than 0.4. Additionally, the correlation coefficients for both GS and MM of hub genes identified through LASSO regression analysis were greater than 0.5. This indicates that hub genes were significantly positively correlated with both the CML phenotype and the brown module. In this study, we utilized WGCNA analysis and LASSO regression analysis to identify hub genes of CML, and analyze their diagnostic value and potential biological functions. Therefore, WGCNA played a discriminating role to some extent. For the phenomenon that the correlation coefficients between module and phenotype and between MM and GS did not reach a high level, we believe that it may be due to the small size of CML samples included in the study. Since CML accounts for only about 15% of all leukemias, this disease is much less studied than other acute leukemias, and thus, the relevant sequencing data will be smaller. However, the two CML cohorts included in our study are currently the largest sample size cohorts with normal samples that can be found in public databases and are also representative.

The hub genes we identified are likely to be important molecules in CML metabolism and immune regulation. Stearoyl coenzyme A desaturase (SCD), a lipase that converts saturated fatty acids to monounsaturated fatty acids, is a key regulator of fatty acid metabolism pathways, its expression is also associated with poor prognosis in several cancer types (24), and elevated SCD levels also protect cancer cells from ferroptosis (25–27). Its upregulation in CML may also contribute to cancer cell growth and treatment resistance by affecting fatty acid metabolism. The high expression of LINC01268 promotes the progression of HCC by regulating MAP3K7 (28). Exosomal lncRNA LINC01268 is also a cancer-promoting factor for pancreatic cancer (29). NME/NM23 family member 8 (NME8) has been identified as a predisposition variant in breast cancer and a prognostic marker in diffuse large B-cell

lymphoma (30, 31). DMXL2 has also been proposed as a potential therapeutic target for breast cancer and oral mucosal melanoma (32, 33). CXXC5 is a member of the CXXC-type zinc finger protein family. It can regulate various signal transduction processes, including TGF-β, Wnt, and ATM-p53 pathways, thereby regulating cell proliferation, differentiation, and apoptosis, and has been implicated in cancer occurrence and progression in many studies (34). Fibrillin-1 (FBN1) promotes gastric cancer progression by activating TGF-β1 and PI3K/Akt pathways, and is targeted by miR-486-5p to inhibit the growth of thyroid cancer cells (35, 36). These studies have all revealed the promoting role of hub genes in a variety of cancers, however, their relationship to CML has not been elucidated, and more in-depth mechanistic exploration is expected to reveal their role and potential value as therapeutic targets.

Moreover, we confirmed the diagnostic value of hub genes in both the analysis and validation cohorts. The risk score model constructed by LASSO regression analysis further improved the diagnostic accuracy. The discovery of these markers provides new targets for the diagnosis and treatment of CML. Finally, we identified two distinct molecular subtypes based on hub gene expression, with Cluster B having a lower risk score and infiltrating a higher proportion of CD8+ T cells and activated dendritic cells. However, the expression of immune checkpoints such as PD-L1, CTLA4, HAVCR2, and PD-1 was significantly up-regulated in Cluster B, as well as the higher TIDE score, indicating that this molecular subtype has a certain degree of immunosuppression, which inhibits the tumor-killing function of immune cells. Therefore, immunotherapy of patients in this subtype may have a higher response. In addition, drug prediction analysis showed that Cluster B was more sensitive to commonly used TKIs. The identification of molecular subtypes provides a new strategy for precise treatment of CML. Finally, we verified the expression of hub genes in larger clinical sample sizes, and confirmed that inhibition

of LINC01268 expression significantly reduced CML cell viability and promoted apoptosis *in vitro*. These results reveal the oncogenic role of LINC01268 and its potential as a therapeutic target for CML. Another study showed that LINC01268, a lncRNA involved in the epigenetic regulation of AML, exerts deacetylation by directly activating HDAC2 and generating positive feedback with HDAC2. In addition, HDAC2 stimulates the transcription of LINC01268, and the expression of LINC01268 is also associated with poor prognosis and cell proliferation in AML (37). Therefore, combined with our findings, LINC01268 is most likely a malignant regulator of myeloid leukemia. However, our study also has some limitations, such as the still small size of clinical samples for the validation of diagnostic signatures and the lack of a more in-depth experimental analysis of hub genes function in CML cells. In addition, the correlation and biological mechanisms of hub genes with CML progression and drug resistance deserve further exploration, thus providing new targets for CML drug resistance treatment, which we will further refine in future studies.

Conclusion

In summary, through WGCNA analysis and LASSO regression analysis, this study provides a better understanding of the role of biomarkers LINC01268, NME8, DMXL2, CXXC5, SCD, and FBN1, and provides a biological basis for further investigation of CML diagnosis and treatment.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The studies involving humans were approved by ethics committee of The Second Hospital of Nanchang University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FZ: Data curation, Formal analysis, Funding acquisition, Methodology, Software, Validation, Visualization, Writing – original draft. FY: Validation, Visualization, Writing – original draft. SX: Validation, Visualization, Writing – original draft. JZ: Validation, Visualization, Writing – original draft. JL: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. XW: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1297886/full#supplementary-material

References

1. Jabbour E, Kantarjian H. Chronic myeloid leukemia: 2022 update on diagnosis, therapy, and monitoring. *Am J Hematol* (2022) 97:1236–56. doi: 10.1002/ajh.26642

2. Osman AEG, Deininger MW. Chronic Myeloid Leukemia: Modern therapies, current challenges and future directions. *Blood Rev* (2021) 49:100825. doi: 10.1016/j.blre.2021.100825

3. Rosti G, Castagnetti F, Gugliotta G, Baccarani M. Tyrosine kinase inhibitors in chronic myeloid leukaemia: which, when, for whom? *Nat Rev Clin Oncol* (2017) 14:141–54. doi: 10.1038/nrclinonc.2016.139

4. Poudel G, Tolland MG, Hughes TP, Pagani IS. Mechanisms of resistance and implications for treatment strategies in chronic myeloid leukaemia. *Cancers (Basel)* (2022) 14(14). doi: 10.3390/cancers14143300

5. Alves R, Gonçalves AC, Rutella S, Almeida AM, De Las Rivas J, Trougakos IP, et al. Resistance to tyrosine kinase inhibitors in chronic myeloid leukemia-from molecular mechanisms to clinical relevance. *Cancers (Basel)* (2021) 13(19). doi: 10.3390/cancers13194820

6. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol (Clifton N.J.)* (2016) 1418:93–110. doi: 10.1007/978-1-4939-3578-9_5

7. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* (2008) 9:559. doi: 10.1186/1471-2105-9-559

8. Li W, Gao Y, Jin X, Wang H, Lan T, Wei M, et al. Comprehensive analysis of N6-methylandenosine regulators and m6A-related RNAs as prognosis factors in colorectal cancer. *Mol Ther Nucleic Acids* (2022) 27:598–610. doi: 10.1016/j.omtn.2021.12.007

9. Liu K, Chen S, Lu R. Identification of important genes related to ferroptosis and hypoxia in acute myocardial infarction based on WGCNA. *Bioengineered* (2021) 12:7950–63. doi: 10.1080/21655979.2021.1984004

10. Yu H, Du X, Zhao Q, Yin C, Song W. Weighted gene Co-expression network analysis (WGCNA) reveals a set of hub genes related to chlorophyll metabolism process in chlorella (Chlorella vulgaris) response androstenedione. *Environ pollut (Barking Essex 1987)* (2022) 306:119360. doi: 10.1016/j.envpol.2022.119360

11. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J Integr Biol* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

12. Newman A, Liu C, Green M, Gentles A, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* (2015) 12:453–7. doi: 10.1038/nmeth.3337

13. Ritchie M, Phipson B, Wu D, Hu Y, Law C, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* (2015) 43:e47. doi: 10.1093/nar/gkv007

14. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* (1997) 16:385–95. doi: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

15. Li SQ, Liu J, Zhang J, Wang XL, Chen D, Wang Y, et al. Transcriptome profiling reveals the high incidence of hnRNPA1 exon 8 inclusion in chronic myeloid leukemia. *J Adv Res* (2020) 24:301–10. doi: 10.1016/j.jare.2020.04.016

16. Esposito A, Falace A, Wagner M, Gal M, Mei D, Conti V, et al. Biallelic DMXL2 mutations impair autophagy and cause Ohtahara syndrome with progressive course. *Brain J Neurol* (2019) 142:3876–91. doi: 10.1093/brain/awz326

17. Liu Y, Yu JT, Wang HF, Hao XK, Yang YF, Jiang T, et al. Association between NME8 locus polymorphism and cognitive decline, cerebrospinal fluid and neuroimaging biomarkers in Alzheimer's disease. *PloS One* (2014) 9:e114777. doi: 10.1371/journal.pone.0114777

18. Milleron O, Arnoult F, Delorme G, Detaint D, Pellenc Q, Raffoul R, et al. Pathogenic FBN1 genetic variation and aortic dissection in patients with marfan syndrome. *J Am Coll Cardiol* (2020) 75:843–53. doi: 10.1016/j.jacc.2019.12.043

19. Zhang H, Li H, Ho N, Li D, Li S. Scd1 plays a tumor-suppressive role in survival of leukemia stem cells and the development of chronic myeloid leukemia. *Mol Cell Biol* (2012) 32:1776–87. doi: 10.1128/mcb.05672-11

20. Kominsky DJ, Klawitter J, Brown JL, Boros LG, Melo JV, Eckhardt SG, et al. Abnormalities in glucose uptake and metabolism in imatinib-resistant human BCR-ABL-positive cells. *Clin Cancer Res* (2009) 15:3442–50. doi: 10.1158/1078-0432.Ccr-08-3291

21. de Beauchamp L, Himonas E, Helgason GV. Mitochondrial metabolism as a potential therapeutic target in myeloid leukaemia. *Leukemia* (2022) 36:1–12. doi: 10.1038/s41375-021-01416-w

22. Cayssials E, Jacomet F, Piccirilli N, Lefèvre L, Roy L, Guilhot F, et al. Sustained treatment-free remission in chronic myeloid leukaemia is associated with an increased frequency of innate CD8(+) T-cells. *Br J Haematol* (2019) 186:54–9. doi: 10.1111/bjh.15858

23. Harada I, Sasaki H, Murakami K, Nishiyama A, Nakabayashi J, Ichino M, et al. Compromised anti-tumor-immune features of myeloid cell components in chronic myeloid leukemia patients. *Sci Rep* (2021) 11:18046. doi: 10.1038/s41598-021-97371-8

24. Tracz-Gaszewska Z, Dobrzyn P. Stearoyl-coA desaturase 1 as a therapeutic target for the treatment of cancer. *Cancers (Basel)* (2019) 11(7). doi: 10.3390/cancers11070948

25. Sen U, Coleman C, Sen T. Stearoyl coenzyme A desaturase-1: multitasker in cancer, metabolism, and ferroptosis. *Trends Cancer* (2023) 9:480–9. doi: 10.1016/j.trecan.2023.03.003

26. Zhao Y, Li M, Yao X, Fei Y, Lin Z, Li Z, et al. HCAR1/MCT1 regulates tumor ferroptosis through the lactate-mediated AMPK-SCD1 activity and its therapeutic implications. *Cell Rep* (2020) 33:108487. doi: 10.1016/j.celrep.2020.108487

27. Chen H, Qi Q, Wu N, Wang Y, Feng Q, Jin R, et al. Aspirin promotes RSL3-induced ferroptosis by suppressing mTOR/SREBP-1/SCD1-mediated lipogenesis in PIK3CA-mutatnt colorectal cancer. *Redox Biol* (2022) 55:102426. doi: 10.1016/j.redox.2022.102426

28. Jin X, Fu W, Li D, Wang N, Chen J, Zeng Z, et al. High expression of LINC01268 is positively associated with hepatocellular carcinoma progression via regulating MAP3K7. *Onco Targets Ther* (2021) 14:1753–69. doi: 10.2147/ott.S295814

29. Liu S, Di Y, Li Q, Chen L, Ma Y, He X, et al. Exosomal lncRNA LINC01268 promotes pancreatic cancer progression via the miR-217-KIF2A-PI3K/AKT axis. *Genes Dis* (2023) 10:1799–801. doi: 10.1016/j.gendis.2022.12.018

30. Shahi RB, De Brakeleer S, Caljon B, Pauwels I, Bonduelle M, Joris S, et al. Identification of candidate cancer predisposing variants by performing whole-exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families. *BMC Cancer* (2019) 19:313. doi: 10.1186/s12885-019-5494-7

31. Wu W, Liu S, Tian L, Li C, Jiang Y, Wang J, et al. Identification of microtubule-associated biomarkers in diffuse large B-cell lymphoma and prognosis prediction. *Front Genet* (2022) 13:1092678. doi: 10.3389/fgene.2022.1092678

32. Menon R, Panwar B, Eksi R, Kleer C, Guan Y, Omenn GS. Computational inferences of the functions of alternative/noncanonical splice isoforms specific to HER2+/ER-/PR- breast cancers, a chromosome 17 C-HPP study. *J Proteome Res* (2015) 14:3519–29. doi: 10.1021/acs.jproteome.5b00498

33. Lyu J, Song Q, Chen J, Shepard MJ, Song H, Ren G, et al. Whole-exome sequencing of oral mucosal melanoma reveals mutational profile and therapeutic targets. *J Pathol* (2018) 244:358–66. doi: 10.1002/path.5017

34. Xiong X, Tu S, Wang J, Luo S, Yan X. CXXC5: A novel regulator and coordinator of TGF-β, BMP and Wnt signaling. *J Cell Mol Med* (2019) 23:740–9. doi: 10.1111/jcmm.14046

35. Wang X, Shi X, Lu H, Zhang C, Li X, Zhang T, et al. Succinylation inhibits the enzymatic hydrolysis of the extracellular matrix protein fibrillin 1 and promotes gastric cancer progression. *Adv Sci (Weinheim Baden-Wurttemberg Germany)* (2022) 9: e2200546. doi: 10.1002/advs.202200546

36. Ma X, Wei J, Zhang L, Deng D, Liu L, Mei X, et al. miR-486-5p inhibits cell growth of papillary thyroid carcinoma by targeting fibrillin-1. *Biomedicine pharmacotherapy = Biomedecine pharmacotherapie* (2016) 80:220–6. doi: 10.1016/j.biopha.2016.03.020

37. Lei L, Xia S, Liu D, Li X, Feng J, Zhu Y, et al. Genome-wide characterization of lncRNAs in acute myeloid leukemia. *Brief Bioinform* (2018) 19:627–35. doi: 10.1093/bib/bbx007