



## OPEN ACCESS

## EDITED BY

Fu Wang,  
Xi'an Jiaotong University, China

## REVIEWED BY

Xueyuan Cao,  
University of Tennessee Health Science  
Center (UTHSC), United States  
Xiaowei Ma,  
Chinese Academy of Sciences (CAS),  
China  
Xingru Zhao,  
Beijing QuantiHealth Technology Co.,  
Ltd., China

## \*CORRESPONDENCE

Zhimin Zhai,  
zzzm889@163.com

## SPECIALTY SECTION

This article was submitted to Cancer  
Genetics and Oncogenomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 23 July 2022

ACCEPTED 18 August 2022

PUBLISHED 15 September 2022

## CITATION

Liang X, Meng Y, Li C, Liu L, Wang Y, Pu L,  
Hu L, Li Q and Zhai Z (2022), Super-  
Enhancer–Associated nine-gene  
prognostic score model for prediction  
of survival in chronic lymphocytic  
leukemia patients.  
*Front. Genet.* 13:1001364.  
doi: 10.3389/fgene.2022.1001364

## COPYRIGHT

© 2022 Liang, Meng, Li, Liu, Wang, Pu,  
Hu, Li and Zhai. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Super-Enhancer–Associated nine-gene prognostic score model for prediction of survival in chronic lymphocytic leukemia patients

Xue Liang, Ye Meng, Cong Li, Linlin Liu, Yangyang Wang,  
Lianfang Pu, Linhui Hu, Qian Li and Zhimin Zhai\*

Department of Hematology/Hematological Lab, The Second Hospital of Anhui Medical University,  
Hefei, Anhui, China

Chronic lymphocytic leukemia (CLL) is a type of highly heterogeneous mature B-cell malignancy with various disease courses. Although a multitude of prognostic markers in CLL have been reported, insights into the role of super-enhancer (SE)–related risk indicators in the occurrence and development of CLL are still lacking. A super-enhancer (SE) is a cluster of enhancers involved in cell differentiation and tumorigenesis, and is one of the promising therapeutic targets for cancer therapy in recent years. In our study, the CLL-related super-enhancers in the training database were processed by LASSO-penalized Cox regression analysis to screen a nine-gene prognostic model including TCF7, VEGFA, MNT, GMIP, SLAMF1, TNFRSF25, GRWD1, SLC6AC, and LAG3. The SE-related risk score was further constructed and it was found that the predictive performance with overall survival and time-to-treatment (TTT) was satisfactory. Moreover, a high correlation was found between the risk score and already known prognostic markers of CLL. In the meantime, we noticed that the expressions of TCF7, GMIP, SLAMF1, TNFRSF25, and LAG3 in CLL were different from those of healthy donors ( $p < 0.01$ ). Moreover, the risk score and LAG3 level of matched pairs before and after treatment samples varied significantly. Finally, an interactive nomogram consisting of the nine-gene risk group and four clinical traits was established. The inhibitors of mTOR and cyclin-dependent kinases (CDKs) were considered effective in patients in the high-risk group according to the pRRophetic algorithm. Collectively, the SE-associated nine-gene prognostic model developed here may be used to predict the prognosis and assist in the risk stratification and treatment of CLL patients in the future.

## KEYWORDS

chronic lymphocytic leukemia, super-enhancer, prognostic model, overall survival, CLL

## Introduction

Chronic lymphocytic leukemia (CLL), a mature and monoclonal CD5+ CD23+ B cell malignancy, proliferates and accumulates in the bone marrow, blood, and lymphoid nodes (Hallek et al., 2018). It is often asymptomatic in the early stage. It is often found that painless lymphadenopathy or the absolute value of lymphocytes is increased for unknown reasons. Patients have mild fatigue, fatigue, and other non-specific manifestations. Once they enter the advanced stage, they can present with weight loss, repeated infection, bleeding, and anemia in addition to systemic lymph nodes and splenomegaly. CLL cases are fewer in Asia than those in the Western world, and it is reasonable to assume that genetic and environmental factors play roles in pathogenesis (Burger, 2020). During 2014–2018, the rate of new cases of CLL was 4.9 per 100,000 per year and the median age at diagnosis is 72 years, the death rate was 1.1 according to the aforementioned survey [The Surveillance Epidemiology and End Results (SEER) Program of the National Cancer Institute. Cancer fact sheets: chronic lymphocytic leukemia (CLL). <https://seer.cancer.gov/statfacts/html/clyl.html> (accessed 22 September 2021)].

CLL is widely known as a heterogeneous disease that exhibits variable clinical symptoms, time-to-treatment (TTT), easily progression and difficult prognosis. CLL patients are often diagnosed with incidental findings, and the clinical course ranges from an asymptomatic, indolent disease that requires no treatment to a rapidly progressive and chemotherapy-resistant disease until death within a short period (Burger, 2020). The indications for treatment mainly include the clinical stage and symptoms of patients, and the standard therapy is chemoimmunotherapy. Unfortunately, the majority of CLL patients are too old to tolerate intensive standard chemotherapy; therefore, an effective prognostic model is needed to predict the individual clinical courses and to improve the outcome. Over the past few decades, great advances have been made in figuring out the molecular and genetic biology of CLL to identify the indicators of progression and survival. These indicators include cytogenetics, age, IGHV gene mutation status,  $\beta$ 2-microglobulin ( $\beta$ 2-MG), clinical-stage (RAI/BINET stage), and so forth (Bosch and Dalla-Favera, 2019). In CLL, 13q14, 11q22-23, trisomy of 12q, and 17p deletions are found in 80% of the cases. 11q22-23 and 17p deletions are associated with poor survival, whereas 13q14 deletions and trisomy of 12q have a longer TTT and survival time (Dohner et al., 2000). TP53 aberrations (Zenz et al., 2010) indicate a more aggressive disease progression and extensive drug-resistant and worse outcome, and the same role applies to IGHV genes (Damle et al., 1999) and ZAP-70 (Crespo et al., 2003). Unmutated IGHV and high-expression of ZAP-70 have a comparatively aggressive disease course too, and other relevant risk markers include expression of CD38 (Rassenti et al., 2008), CD49d (Bulian et al., 2014), lipoprotein lipase (LPL) (Prieto and Oppezzo, 2017), serum concentrations of thymidine kinase (Hallek et al., 1999), and  $\beta$ 2-microglobulin (Hallek et al., 1996).

In this article, a SE-associated gene list was used to carry out LASSO-penalized Cox regression analysis, and construct a nine SE-associated gene prognostic model, namely, TCF7, VEGFA,

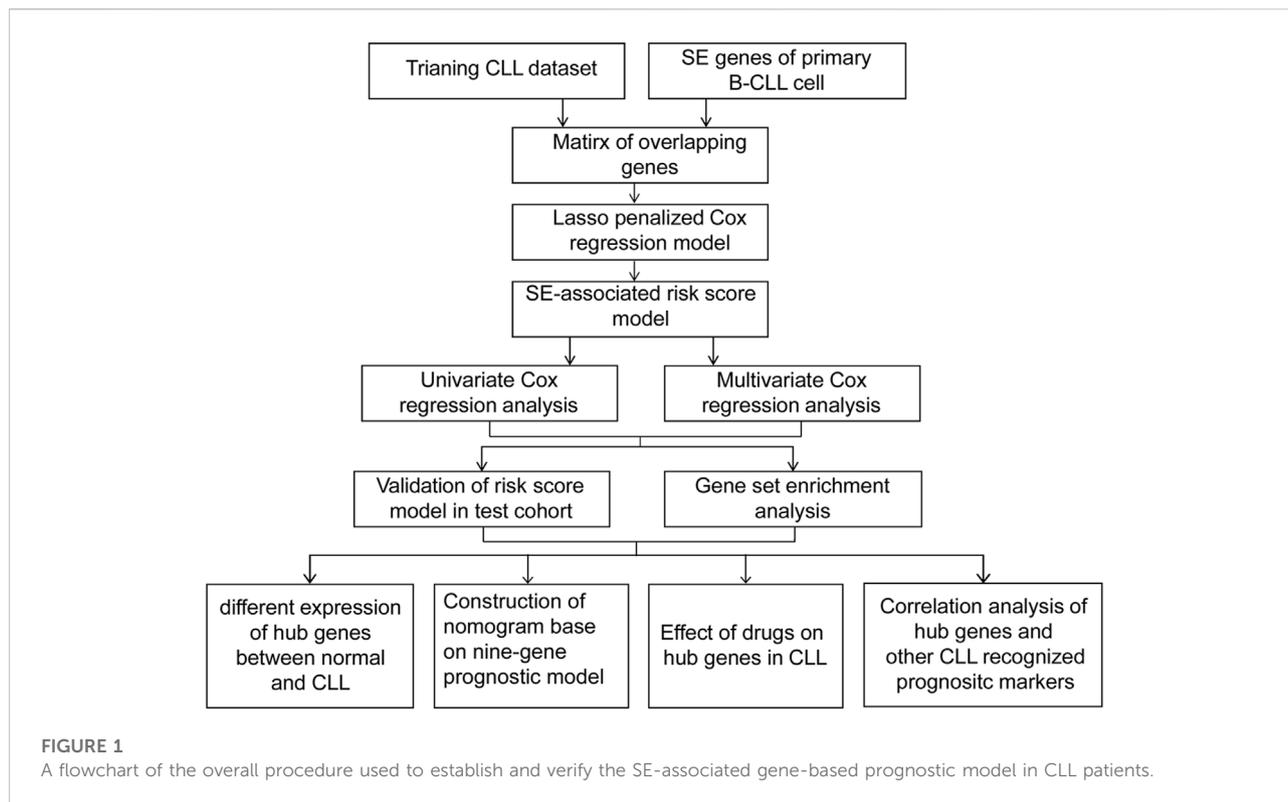
MNT, GMIP, SLAMF1, TNFRSF25, GRWD1, SLC6AC, and LAG3. Meanwhile, this model was verified by testing GEO datasets and the ICGC-CLL dataset, respectively. Univariate and multivariate Cox regression analyses, and the ROC curve were analyzed to evaluate the prognostic accuracy of this nine-gene model. Moreover, the aforementioned validated steps, the role of the nine-gene prognostic model, and the nine hub genes were further explored in CLL genesis and the relationship between this prognostic model and other known risk markers, such as IGHV status, FISH abnormality, and ZAP70 expression level. It was indicated that the model demonstrated predictive power and had an expected relationship with known risk markers. In addition, an interactive nomogram based on the nine-gene risk score and clinical traits was constructed. Finally, paired pre- and post-treatment datasets were used to examine the effects of treatments on the risk score or each of the nine hub genes' expression, and we predicted 25 clinical drugs that may be more sensitive to high-risk patients. The improved nine-gene prognostic model of this work provided a bright future for the diagnosis, disease stratification, and therapy of patients with CLL.

## Results

### Construction of a nine-gene LASSO-penalized cox regression model and validation of independent prognostic factors

The flowchart featured the construction and validation of the SE-associated gene-based prognostic model of CLL and the correlation with other known risk markers (Figure 1). An 831 primary B-CLL cell-related SE list was downloaded from the website and the 18,887 gene matrix in CLL patients was provided in the GSE22762 column, and a 587 SE-associated gene matrix for CLL was gained via overlapping the aforementioned two gene sets. Immediately after, the gene matrix was done by LASSO-penalized Cox regression to screen the prognosis-related genes with potential. Figure 2A shows the coefficient values for each at various penalty levels as long as genes with non-zero coefficients had prognostic value in the LASSO-penalized regression model. Ten-fold cross-validation obtained the maximum lambda value, and we selected one model which produced a group of nine genes (Figure 2B). Principal component analysis (PCA) showed high-risk patients separate from low-risk ones evidently (Figure 2C). Also, the obvious distinction between survival and death was calculated by using the nine gene-based prognostic model, implying that the prognostic model functioned smoothly in the prediction on the OS of patients with CLL (Figure 2D).

To validate the LASSO-penalized Cox regression model, univariate Cox proportional hazard regression analysis determined that these genes affected the OS of patients with CLL independently, and all log-rank *p*-values of the nine genes were <



0.01 (Supplementary Figure S1A). Multivariate Cox proportional hazard regression analysis was also performed, and the global  $p$ -value of our model was only  $2.64 \times 10^{-16}$  (Supplementary Figure S1B), with an AIC of 124.96 and a C-index of 0.95. These indexes suggest that the nine genes are possibly prognostic markers for OS in CLL patients. Meanwhile, the results of K-M survival analysis showed that GRWD1, SLC6A3, and MNT had no significant association with survival (Supplementary Figure S2). Furthermore, we concluded that SLAMF1, TCF7, TNFRSF25, MNT, and VEGFA were protective factors, whereas GRWD1, SLC6A3, GMIP, and LAG3 appeared to be harmful factors in CLL, based on the aforementioned hazard ratios of univariate and multivariate regression analyses. Thus, the nine-gene SE-associated model by LASSO-penalized Cox regression possibly predicted the OS of CLL patients.

## Establishment and validation of the nine gene-based risk score model

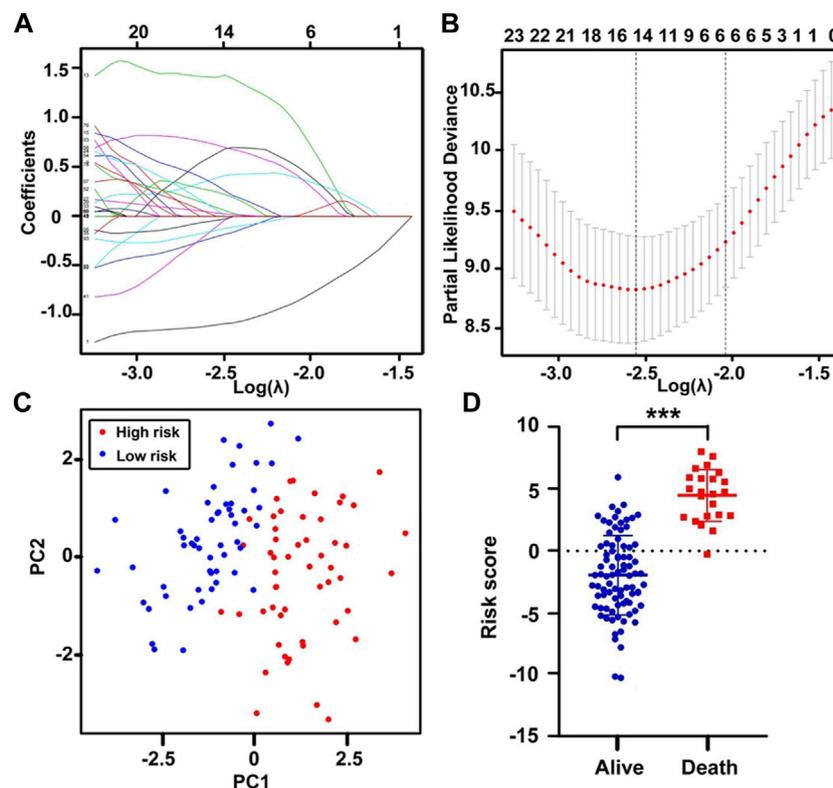
A total of 107 patients in the training dataset of GSE22762 (HGU-133plus2) were divided into high-risk (risk score  $> 0.7$ ) and low-risk groups (risk score  $< 0.7$ ) (Figure 3A). Figure 3B shows that death was more frequently observed in the high-risk group than in the low-risk group. The K-M survival analysis presented a much worse outcome in the high-risk group than that of the low-risk group (log-rank test,  $p = 3.561 \times 10^{-9}$ ) (Figure 3C). Also, the AUCs of a time-

dependent ROC curve of 1, 3, and 5 years calculated by the nine gene-based risk score model were 0.997, 0.958, and 0.996, respectively (Figure 3D), suggesting that the prediction was highly sensitive and specific. The testing column (GSE22762,  $N = 44$ , HGU-133A) verified the predictive values of the nine gene-based risk scores. The K-M curves of the high- and low-risk groups were noticeably different (log-rank test,  $p < 0.05$ ) and the AUCs of 1-, 3-, and 5-year ROC curves were 0.738, 0.679, and 0.628, respectively; these results showed that this prognostic model might be a potential predictor to judge the OS of patients with CLL (Supplementary Figure S3).

GSEA was carried out in two datasets on exploring enriched KEGG pathways in which the analysis suggested that vital enrichment was concentrated in the high-risk cohort, including base and nucleotide excision repair, DNA replication, and valine-leucine and isoleucine degradation (Supplementary Figures S4A,B). Other pathways including homologous recombination, oxidative phosphorylation, mismatch repair, RNA degradation, RNA polymerase, and one carbon pool by folate and lysine degradation were enriched in the high-risk group of the two cohorts.

## The prediction of the nine-gene model on time-to-treatment

In addition to survival, we also investigated the nine-gene prognostic model on TTT, and the results demonstrated that the



**FIGURE 2**

Construction of the SE-related prognostic model. (A) The LASSO coefficient values at various levels of penalty; each curve represents an SE-gene. (B) The confirmation of the best lambda value by LASSO Cox regression analysis. (C) Principal component analysis (PCA), red dots correspond to high-risk patients, and blue dots correspond to low-risk patients. (D) The scatter plot of the survival status of CLL patients based on the 9-gene model using the *t*-test. \*\*\* $p < 0.001$ .

nine-gene risk model performed well on predicting TTT in the training dataset (GSE22762). Low-risk patients showed a longer TTT than high-risk patients, and the  $p$ -value  $< 0.001$  (Figure 4A). Additionally, the time-dependent ROC curve analysis prompted that the AUCs of 1-, 3-, and 5-year TTT were 0.818, 0.840, and 1.000, respectively (Figure 4B). These results were in accordance with testing datasets (GSE39671) (Figures 4C,D), and it indicated that the prognostic model was equally effective in predicting TTT.

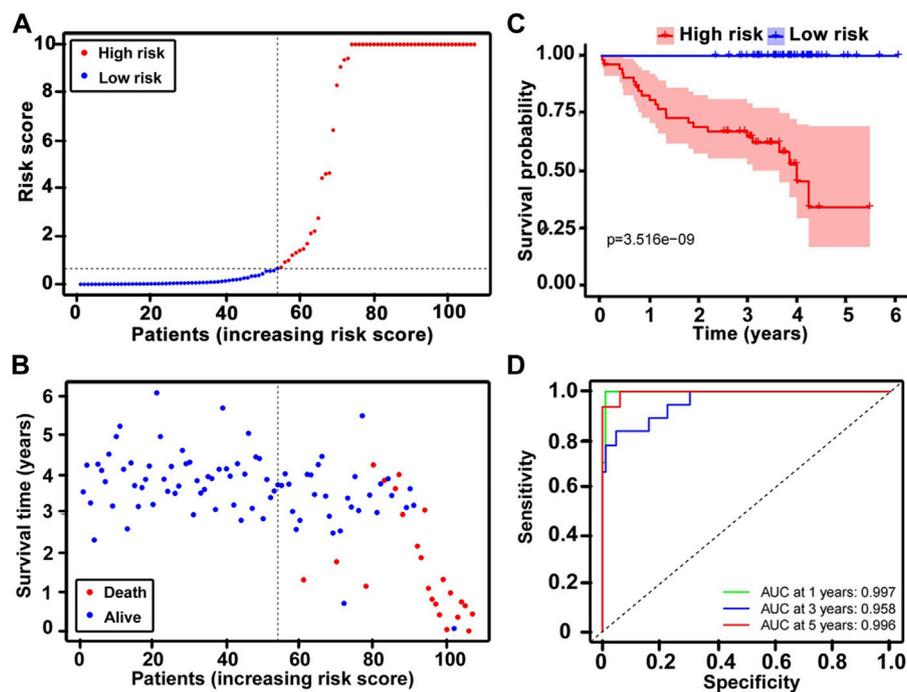
## Identification of SE-related hub genes in chronic lymphocytic leukemia using weighted gene co-expression network analysis

In addition to the prognostic value, we also expected a relationship between the nine-gene model and tumorigenesis. WGCNA was another statistical method for the analysis of finding the different genes between normal and CLL patients. As shown in Figure 5A, the best soft-thresholding value via prediction of the scale independence was  $\beta = 6$ . Then, genes were

divided into 9 different modules with 9 different colors, and a heatmap was developed according to Pearson's correlation coefficient (Figure 5B). An intersection between the SE matrix and the nine modules which presented a higher correlation with CLL showed that TCF7 and LAG3 appeared in the interaction genes between module purple, yellow, and SE-associated genes (Figure 5C). Simultaneously, TCF7, GMIP, SLAMF1, TNFRSF25, and LAG3 were found to express differently in normal and CLL patients when we compared the individual expression of nine SE-related hub genes in CLL (Figure 5D). The data indicated that the five genes may play a vital role in regulating the genesis of CLL.

## The validated nine-gene prognostic model and other risk factors

The performance of the nine-gene prognostic model was additionally evaluated in different subgroups defined by confirmed risk factors. Patients with mutated IGVH genes, 13q14 or single deletion or trisomy 12 on FISH analysis,



**FIGURE 3**

Nine-gene prognostic model for the GSE22762 dataset (N = 107, HG-U133\_Plus\_2). (A) Dot plots comparing the outcomes of subjects in the high- and low-risk cohorts. (B) The survival status and time in the high- and low-risk groups. (C) K–M survival curves showing the differences between the high- and low-risk groups. (D) Time-dependent ROC curve analysis for the prediction survival using the nine-gene model. K–M, Kaplan–Meier; ROC, receiver operating characteristic; AUC, area under the curve.

presented a favorable outcome, whereas patients with unmutated IGHV status, 17p13 or a 11q23 deletions, had an unfavorable prognosis. Unmutated IGHV patients had a higher risk score than mutated IGHV patients in three independent datasets (GSE9992, GSE16746, and GSE28654) (Figures 6A–C). Simultaneously, we analyzed the correlation between IGHV mutation status and each gene in the nine-gene prognostic model. The results reported that the expression of TCF7 and SLAMF1 had a strong positive correlation, and LAG3 showed a negative correlation with IGHV mutation (Figure 6D). Similarly, patients with del17p13 had a higher risk score compared to other chromosome types ( $p < 0.001$ , Figure 6E). The risk score of ZAP70-high patients was higher than that of ZAP70-low patients; the expression of MNT and SLAMF1 had a negative association, and LAG3 had a positive association with ZAP70, respectively (Figures 6F,G). Additionally, the variation of risk score and each gene expression before and after treatment was provided in Supplementary Figure S5A. The risk score was downregulated after processing with HDAC inhibitory *in vitro*, and VEGFA and MNT were upregulated accompanied by downregulated GMIP and TGFRSF25. In the other two *in vivo* treatment experiments, no significant change was found except LAG3, the LAG3 gene was

upregulated consistently after lenalidomide and thalidomide treatment respectively (Supplementary Figures S5B,C).

## Validation of nine-gene prognostic model in ICGC and construction of a nomogram to predict OS

International Cancer Genome Consortium (ICGC, <http://daco.icgc.org/>), which collected multiple genetic mutations, copy number variants, epigenetic modifications, and clinical data covering 50 tumor types, and we extracted 255 CLL patient data for following analysis. Again, high risk scores were significantly associated with shorter survival time,  $p < 0.001$  (Supplementary Figure S6A), and the AUCs of ROC curves of the 3-, 5-, and 10-year survival were 0.731, 0.718, and 0.800, respectively (Supplementary Figure S6B). CLL patients could be divided into two molecular subtypes according to the mutational status of the IGHV, with cases carrying unmutated IGHV (U-CLL) having more aggressive behavior than patients with mutated IGHV (M-CLL). Consistent with the most accepted view, the nine-gene risk score median value was obviously lower in the indolent CLL subtype (M-CLL) compared to the aggressive one (U-CLL) (Supplementary Figure S6D). The nine-gene risk

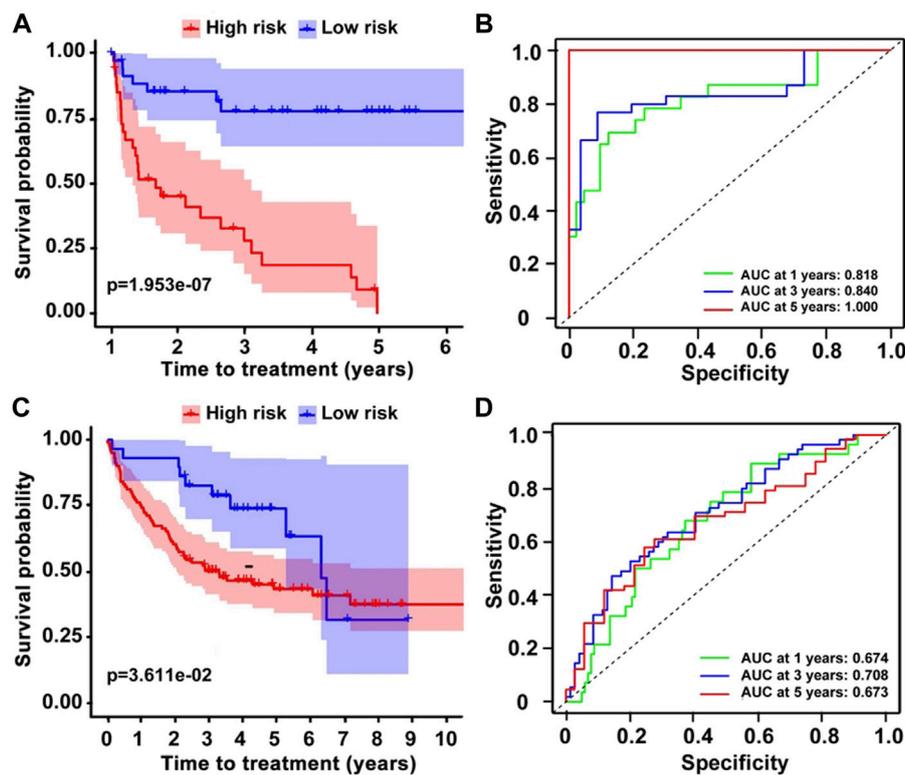


FIGURE 4

The prediction of TTT on CLL patients. (A,C) K–M survival curves showing the different TTT on two datasets and (B,D) ROC analysis for the prediction of TTT.

score was associated with the evolution of M-CLL with a median OS of 6.57 versus 8.87 years for patients with high and low risk scores, respectively ( $p = 0.005$ , Supplementary Figure S6C), while no differences were seen in U-CLL patients in relation to high- and low-risk scores (data not shown). Moreover, on the basis of the obtained sample clinical characteristics, we performed univariate as well as multivariate Cox survival analyses. Age, IGHV mutated status, and risk were identified to be independent prognostic factors for patients with CLL ( $p < 0.05$ ; Figures 7A,B). Based on the nine-gene risk score and clinical traits, a nomogram was constructed to accurately predict CLL patients' 1-, 3-, 5-, and 10-year survival rates by the using aforementioned clinical indicators and the nine-gene risk score. The C-index of this model was 0.82 (Figure 7C).

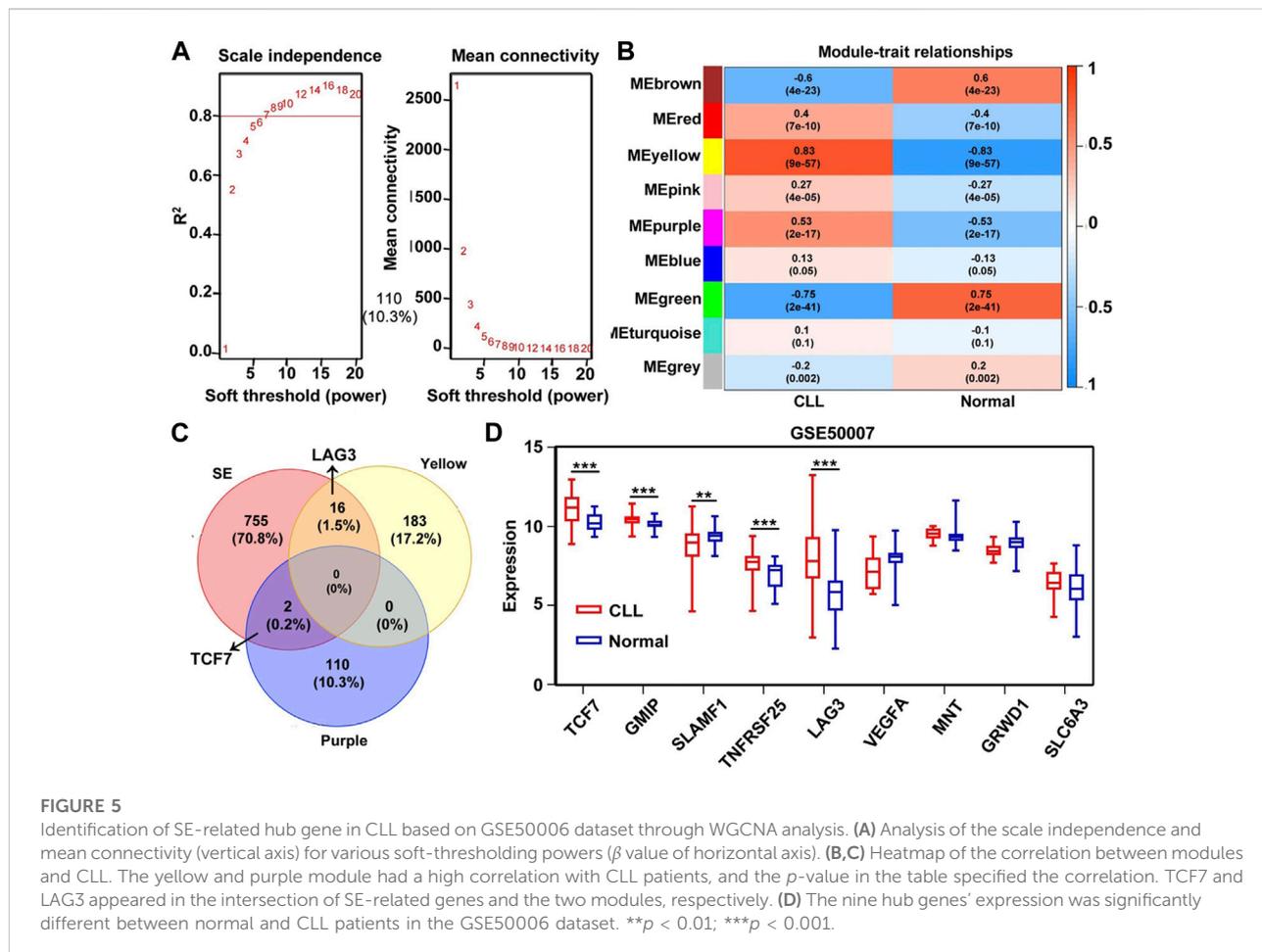
## Response of high- and low-risk patients to chemotherapeutic compounds

According to the pRRophetic algorithm, we predicted the IC50 of 130 chemotherapeutic agents and pathway inhibitors in both of high- and low-risk patients and found that 25 drugs had lower IC50 in high-risk patients ( $p < 0.05$ , additional file 1), which

indicated that the high-risk patients were more sensitive to these 25 drugs. Among these compounds, some have been reported to have pre-clinical anti-tumor activity in CLL, such as thapsigargin, which was found to be a potent cytotoxin that induced apoptosis by inhibiting the sarcoplasmic/endoplasmic reticulum Ca<sup>2+</sup> ATPase (SERCA) pump, which was necessary for cellular viability. Some have not been reported in CLL before, and therefore the therapeutic effect is still unknown. Interestingly, there were three kinds of compounds which could inhibit the mTOR pathway and CDKs in CLL, respectively, and these have been researched in CLL before and CDK inhibitors have entered clinical trials in patients with relapsed or refractory chronic lymphocytic leukemia. These results could be helpful for the precise treatment of CLL (Figure 8).

## Discussion

CLL is considered to have a highly heterogeneous clinical course, with time to first treatment varying from months to years and many patients eventually progressing and requiring chemotherapy, although initially, CLL is reported as an indolent malignancy. A review of the data so far, disease stratification, IGHV mutation status, 17p, and ZAP70 expression are the validated predictors of

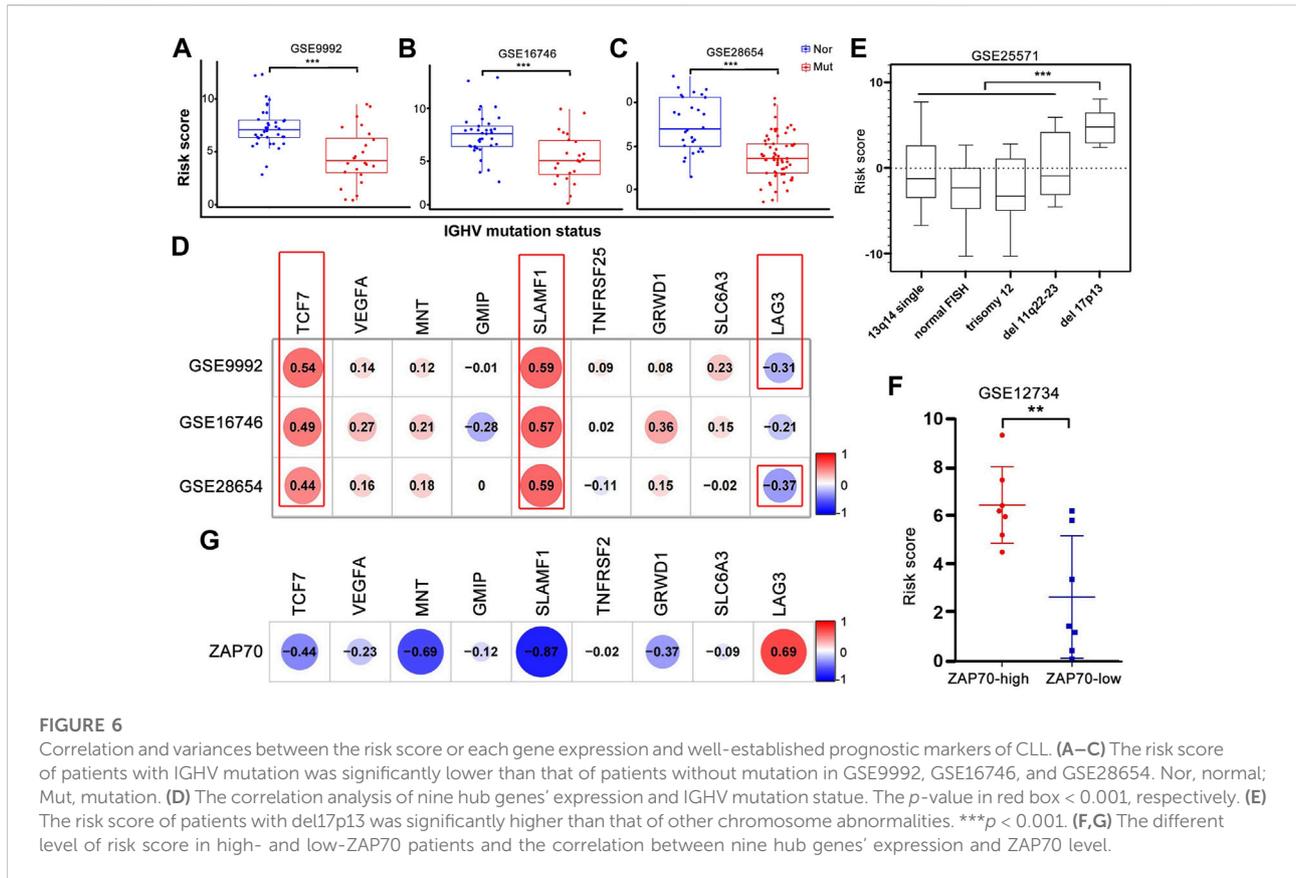


overall survival. Beyond that, gene expression analysis was carried out on various surrogate markers for genetic features and prognosis. A total of six surface antigens (CD62L, CD54, CD49c, CD49d, CD38, and CD79b) and prognostic risk models were put in place to diagnose and predict the OS for CLL (Zucchetto et al., 2006). Moreover, some large-scale gene expression profiling analyses generate different prognostic factors (Kienle et al., 2010; Herold et al., 2011a; Schweighofer et al., 2011). But the previous studies constructed no prognostic model according to SE-associated genes which regulate the expression of hub genes related to CLL tumorigenesis.

A super-enhancer is a new concept developed in recent years; a growing body of evidence indicates an explicit relationship between increasing tumorigenesis and malignancy of cancer and SEs. SEs drive not only the expression of genes but also non-coding RNA that regulates biological functions directly and indirectly. LASSO-penalized Cox regression has become popular in recent years because it could minimize overfitting (Ma et al., 2019). Hence, in our article, we use this novel bioinformatic strategy and the Cox proportional hazard regression models to screen and optimize hub genes related to survival.

In our research, the LASSO-penalized Cox regression analysis was carried out by filtering out the potential SE-associated genes and yielding a nine-gene prognostic model to foresee the OS of CLL patients. All of the individual markers in the nine-gene model associated with OS of CLL by Cox regression analysis were identical. K–M survival analysis also indicated that the majority of the nine genes correlated to OS. Beyond that, the nine-gene prognostic model was highly significant in the multivariate analysis of patients without treatment. The AUCs and C-index showed that our model performed well in the prediction of survival. The effectiveness of this prognostic model could be validated by an independent patient cohort. Moreover, this risk model was another indicator of TTT. We utilized the nine-gene risk score in the GSE22762 and GSE39671 datasets, and the results also indicated that the nine-gene model could be applied to predict TTT. The high-risk patients had less time-to-treatment than the low-risk patients. These data strongly indicated that the nine-gene prognostic model was a significant and valid risk forecaster.

We not only evaluated the data by a rigorous training and validation design, but also concentrated on the connection

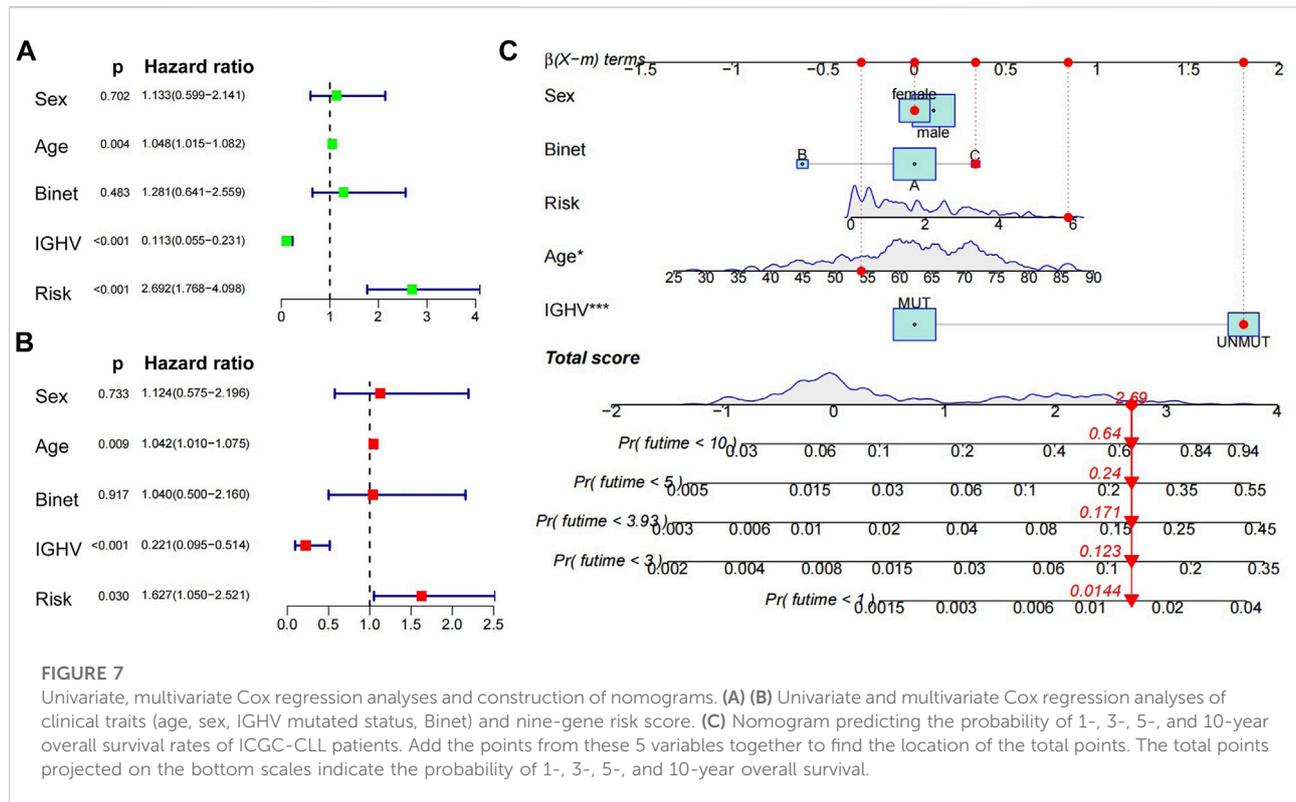


between individual genes and selected disease characteristics, such as IGHV mutation status, FISH abnormality, and ZAP70 expression level. The results of three of the markers (TCF7, SLAMF1, and LAG3) detected according to the association with IGHV status were expected. The lack of a public database that included both survival data and mutation information limited further research on a correlation between the nine-gene model and ZAP70, a FISH abnormality. But in the poor prognosis groups, like ZAP70-high and 17q-patients, the nine-gene risk score was significantly higher than that in the low-risk group, and we found that the low expression of SLAMF1 in CLL was associated with ZAP70-high expression. The quantitative relationship between TCF7, LAG3, and SLAMF1 expression and inferior overall survival was an accurate finding and indicated that these genes had a pathogenic role in CLL. Additionally, the nine-gene prognostic model also played an important role in CLL etiopathogenesis. The WGCNA of the GSE50006 dataset revealed that TCF7 and LAG3 belonged to two gene modules, respectively. In addition to this, the expression of GMIP, SLAMF1, and TNFRSF25 were also significantly different in normal and CLL patients. Therefore, the five genes contained in our model were possibly functionally vital in the pathogenesis of CLL. In the present study, SLAMF1, TCF7, TNFRSF25, MNT,

and VEGFA were protective factors, whereas GRWD1, SLC6A3, GMIP, and LAG3 appeared to be harmful factors in CLL; we subsequently discussed each gene in the prognostic model.

Transcription factor 7 (TCF7), the T-cell-specific transcription factor required for T-cell development in animal models, suggests that it probably functions as a tumor suppressor (Roose et al., 1999). TCF7 over-expression in mice led to a disease resembling CLL, indicating that it was probably involved in the CLL transformation in a direct way (Bichi et al., 2002). In CLL, TCF7 expression provided a high rate (74%) of correct assignment of patients at genetic risk (IGHV unmutated, V3-21 usage, 11q-, or 17p-) (Kienle et al., 2010). The aforementioned results are consistent with ours, and this indicates TCF7 plays an important role in CLL.

Signaling lymphocytic activation molecule family member 1 (SLAMF1), also known as CD150, regulates hematopoietic stem cell differentiation, leukocyte adhesion and activation, and humoral immune responses. SLAMF1 comparatively over-express in normal peripheral blood B cells according to the meta-analysis of three gene expression profiling studies. Recently, researchers found lower levels of SLAMF1 expression in cases with ZAP70-high (*p* < 0.001), IGHV-unmutated (*p* < 0.001), and 17q- (*p* = 0.003). In past studies, we believed that loss of SLAMF1 expression in CLL modulates genetic pathways regulating chemotaxis and



autophagy and that potentially affects drug responses, suggesting that the effects underlie unfavorable clinical outcomes experienced by SLAMF1-low patients (Bologna et al., 2016). Together, SLAMF receptors, the vital modulators of the BCR signaling axis, improve immune control in CLL by potentially interfering with NK cells (von Wenserski et al., 2021). In our research, the univariate and multivariate analyses presented that downregulated SLAMF1 levels had an independent negative prognostic impact on overall survival ( $p < 0.05$ ). We subsequently discovered that SLAMF1 is relatively overexpressed in IGHV-mutated and ZAP70-low CLL patients. The strict correlation among low levels of it and high-risk genetic features indicated that it probably represented a marker of surrogate genomic complexity; however, the mechanism of this correlation is still unknown.

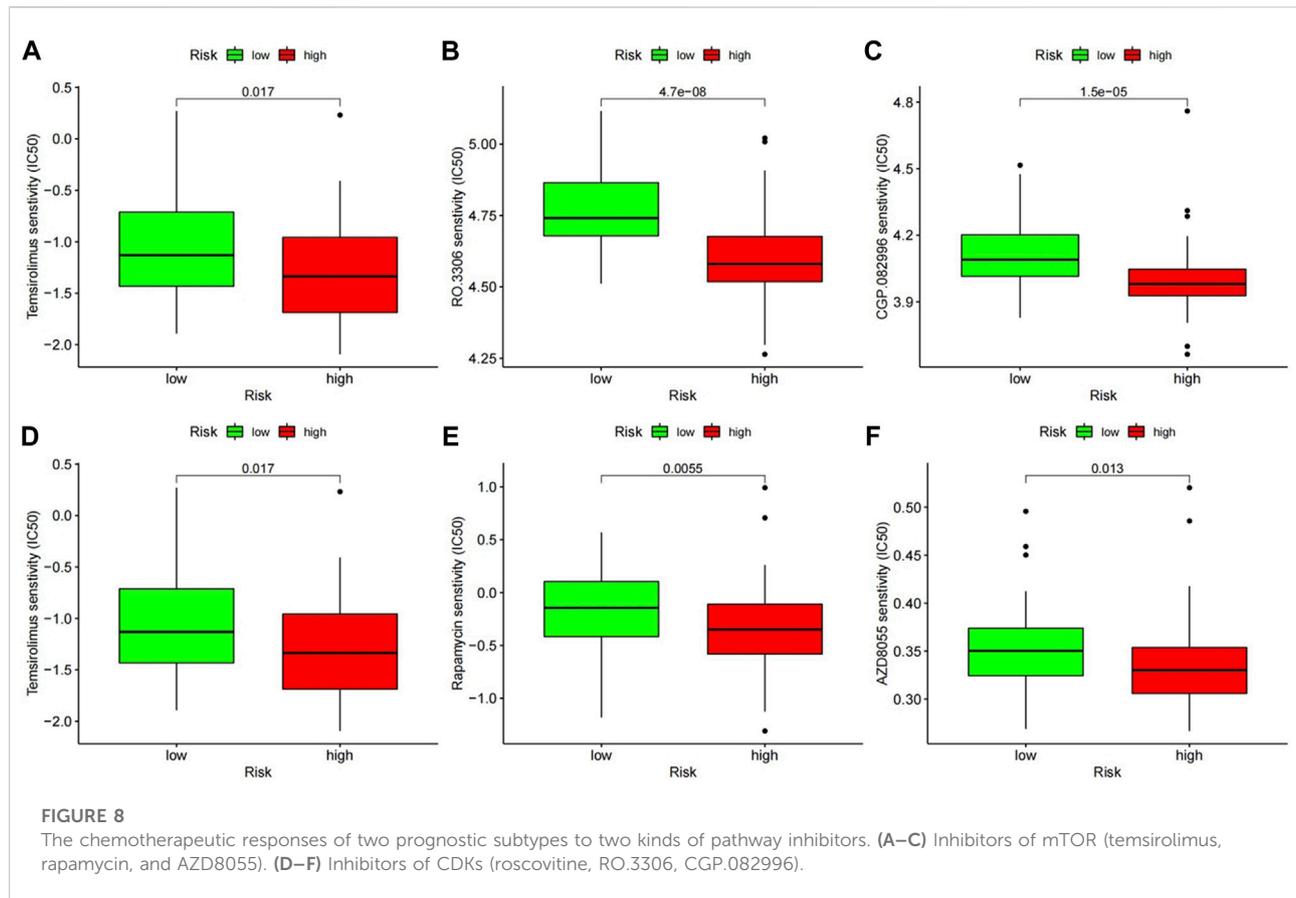
Lymphocyte activating 3 (LAG3), the immune inhibitory checkpoint receptor, is one of the immunoglobulin superfamily with about 20% amino acid homology with CD4. The expression of it activates and exhausts T, NK cells, B cells, dendritic cells, and regulatory T (Treg) cells. LAG3 high expression in CLL cells correlates with unmutated IGHV ( $p < 0.0001$ ) and decreased treatment-free survival ( $p = 0.0087$ ) (Kotaskova et al., 2010). Increased LAG-3 expression on leukemic cells correlates with shorter time-to-treatment and poor outcome in CLL; moreover, treatment with relatimab, a novel anti-LAG-3 blocking monoclonal antibody currently under clinical trial for different solid and hematological malignancies including CLL, restored, at least in part, NK and

T-cell-mediated anti-tumor responses (Sordo-Bahamonde et al., 2021). CART cell generation with the showing of ibrutinib created enhanced cell viability and expansion of CLL patient-derived CART cells. Also, ibrutinib enriched the mentioned cells with the less-differentiated naïve-like phenotype and declined expression of exhaustion markers (PD-1, TIM-3, and LAG-3) (Fan et al., 2021).

Vascular endothelial growth factor A (VEGFA) is a member of the PDGF/VEGF growth factor family. The angiogenesis process makes a significant contribution to the pathogenesis of B-cell chronic lymphocytic leukemia (B-CLL), the levels of VEGFA and bFGF being higher in patients than in healthy people (Ballester et al., 2020). Whereas, in our research, VEGFA has a protective role in CLL. The high expression of VEGFA indicated a good prognosis by the K–M survival analysis, and in normal samples, the level of VEGFA was higher even though it was not statistically significant.

The TNF receptor superfamily member 25 (TNFRSF25), the receptor expressed preferentially in the tissues of lymphocytes, possibly plays functions vital to the regulation of lymphocyte homeostasis. The receptor stimulates  $\kappa$ B activity and regulates cell apoptosis. TNFRSF25 was differentially expressed, activating CLL cells and predominantly detected in those with early clinical stage disease (Cavallini et al., 2015) and probably alters the balance between cell proliferation and death, influencing CLL physiopathology and results in the clinic.

A total of three genes (GRWD1, GMIP, and SLC6A3) have not been described in the context of CLL before, and all of them were upregulated in high-risk CLL patients. The results of the univariate



and K–M survival curves were not completely consistent with multivariate analysis. Glutamate-rich WD repeat containing 1 (GRWD1) was identified as one of the ribosomal/nucleolar proteins that promote tumorigenesis (Takafuji et al., 2017). Meanwhile, GRWD1 was also viewed as having histone-binding activity and regulating chromatin openness to specific chromatin locations (Sugimoto et al., 2015). Overexpression in colon carcinoma tissues was related to pathological grading, tumor size, N stage, TNM stage, and poor survival; knockdown of GRWD1 function as an inhibitor on cell proliferation and colony formation, and induced cell cycle arrest and more drug susceptibility, and suppressed the migration and invasion (Zhou et al., 2021). GEM interacting protein (GMIP), a RhoA-specific GAP, in a proteomics screen for proteins interacting with Girdin (Girders of actin), an actin-binding protein critical for neuronal migration to the olfactory bulbs, is identified as one of the major regulators of neuronal migration in the postnatal brain (Ota et al., 2014). Solute carrier family 6 member 3 (SLC6A3) involving in the metabolism of dopamine and catecholamine is the potential gene for Parkinson's disease and alcoholism. The significance of the aforementioned three genes in CLL remains to be further studied.

In GSE14973, the risk score was significantly downregulated after the valproic acid (VPA) treatment *in vitro*; meantime, protective factors (VEGFA and MNT) were highly expressed,

and pathogenic genes (GMIP) were less expressed than in the previous treatment, except TNFRSF25, and these results were almost consistent with our previous conclusion. VPA is a well-tolerated anti-epileptic drug with HDAC inhibitory activity. HDAC1 and HDAC3 inhibition or knockdown results could be figured out in HDAC7 downregulation, which was related to a decline in histone 3 lysine 27 acetylation (H3K27ac) at transcription start sites (TSS) and super-enhancers (SEs) prominently in stem-like BrCa cells. In GSE112953 and GSE15913, the only upregulated gene was LAG3, and it may suggest that combination drug treatment with an anti-LAG3 monoclonal antibody would have a better outcome.

In the present study, a nomogram based on the nine-gene risk score and other clinical traits was constructed, and to determine the predictive effect, we applied the nomogram to a specific patient in the ICGC project; moreover, the predictive model containing the nine-gene risk score was more accurate than the nomogram model containing only four clinical traits. Meanwhile, the risk score was strongly correlated with some known prognostic indicators, such as IGHV mutation state and chromosomal abnormalities. While, a further dissection of the nine-gene risk score on OS in the IGHV mutation state could identify that the nine-gene risk score value was apparent only in the less aggressive M-IGHV subtype, and this predicted trait corresponded to what

TABLE 1 Details of databases used in this research study.

GEO accession	Number	Subgroup of sample	Sample type	Application in article
GSE22762 (Herold et al., 2011a)	151	151 CLL	PBMC	Establishment of a survival model by LASSO and survival analysis of OS and TTT by a nine-gene model
GSE39671 (Chuang et al., 2012)	130	130 CLL	PBMC	Survival analysis of TTT by a nine-gene model
GSE50006	210	188 CLL 32 healthy donors	CD19+ B cells	Validation of expression difference of hub genes between CLL and healthy donors
GSE9992 (Fabris et al., 2008)	60	24 M-CLL 36 U-CLL	CD5+ CD19+ CD23+ B cells	Validation of the correlation of hub gene expression and risk score with IGHV status
GSE16746 (Mosca et al., 2010)	60	23 M-CLL 37 U-CLL	CD5+ CD19+ CD23+ B cells	Validation of the correlation of hub gene expression and risk score with IGHV status
GSE28654 (Trojani et al., 2011)	89	61 M-CLL 28 U-CLL	CD19+ cells	Validation of the correlation of hub gene expression and risk score with IGHV status
GSE25571 (Herold et al., 2011b)	109	FISH abnormality	PBMC	Validation of the correlation of hub gene expression and risk score with genotypic abnormality
GSE12734 (Stamatopoulos et al., 2009a)	14	7 high-ZAP70 7 low-ZAP70	CD19+ cells	Validation of the correlation of hub gene expression and risk score with ZAP70 expression level
GSE14973 (Stamatopoulos et al., 2009b)	28	14 CLL with and without VPA	B cells	Validation of the correlation of hub gene expression and risk score with before and after treatment
GSE112953	22	11 CLL before and after lenalidomide treatment	CD19+ cells	Validation of the correlation of hub gene expression and risk score with before and after treatment
GSE15913 (Giannopoulos et al., 2009)	40	20 CLL before and after thalidomide treatment	PBMC	Validation of the correlation of hub gene expression and risk score with before and after treatment

CLL, chronic lymphocytic leukemia; PBMC, peripheral blood mononuclear cells; M-CLL, IGHV mutated CLL; U-CLL, IGHV un-mutated CLL; OS, overall survival; TTT, time-to-treatment.

has been reported in an article which studied the relationship between the ENDOG expression and prognostic study of CLL. The reason why this situation occurred needed further exploration.

The introduction of fludarabine, fludarabine/cyclophosphamide, and either of these combined with rituximab has improved the outcome for younger patients with CLL. Treatment options available for patients in the setting of relapsed disease following receipt of chemoimmunotherapy are limited where most patients have high-risk genomic findings including IgVH un-mutated disease, del (17p13.1) and del (11q22.3) associated with poor treatment response (reviewed in Rassenti et al., 2008). Identifying therapies with novel mechanisms of action for this patient group is important (Johnson et al., 2012). In our research study, all patients were divided into two risk subtypes based on the nine-gene prognostic model, and we endeavored to estimate the drug response of each patient based on IC50 according to the activation of different pathways. ADZ8055 was a dual mTOR kinase inhibitor with inhibition of both mTORC1 and mTORC2 that preferentially decreased cell viability of poor prognostic CLL subsets like with del (11q) or del (17p). One class of drugs that has promise for the treatment of relapsed CLL is the cyclin-dependent kinase (CDK) inhibitors (Seftel et al., 2017). Interestingly, one research study has described that the pan-CDK inhibitor dinaciclib has potent pre-clinical *in vitro* activity against CLL cells independent of high-risk genomic features (Johnson et al., 2012). In our drug sensitivity prediction, there are three kinds of CDK inhibitors which seemed to be more effective for high-risk CLL patients. The reasons that could account for this difference may

include: 1) Different drugs have different mechanisms of action, although they are all one class of inhibitor. 2) The criteria of stratifying patients into “High-risk” and “Low-risk” were not consistent. 3) The most important point is the lack of experimental validation in our research.

## Conclusion

To sum up, it was the initial study using the LASSO model to screen prognostic indicators from the profile of SE-associated genes in CLL. A useful prognostic score for OS in untreated CLL patients was presented, and the determination of the score can be achieved via the measurement of the expression levels of nine genes. It also could be done easily in a routine diagnosis. These nine SE-associated genes in this model were not only vital in the development and progression of CLL, but also could assist in guiding the development of alternative treatments.

## Materials and methods

### Data source and microarray analysis

The microarray data and clinical data of GSE22762 (Herold et al., 2011a) and GSE39671 (Chuang et al., 2012), which contain 107 and 130 CLL patients, respectively, were downloaded from the Gene

Expression Omnibus (GEO) database. These data were conducted by GPL570 and GPL96/GPL97. Here, 9 other datasets were also analyzed for different purposes, and the details were presented in Table 1 (Fabris et al., 2008; Stamatopoulos et al., 2009a; Stamatopoulos et al., 2009b; Giannopoulos et al., 2009; Mosca et al., 2010; Herold et al., 2011b; Trojani et al., 2011). In the meantime, the International Cancer Genome Consortium (ICGC) CLL sequencing data were extracted from the European Genome-Phenome Database (EGA).

## LASSO-penalized cox regression analysis

Super-enhancer-related genes list figured from the primary B-CLL cell was downloaded from SEA version 3.0, which was enriched with a post-translational modification histone mark, H3K27ac ChIP-seq signal. The gene matrix for subsequent analysis was obtained from the overlapping set of genes in the GSE22762 dataset and the SE-associated genes in the primary B-CLL cell. For narrowing and selecting the prognostic genes with potential, the overlapping gene matrix was weighted by the relative coefficients through the LASSO-penalized Cox regression. Ten-fold cross-validation derived the best-fit lambda value to decrease the mean cross-validated error as much as possible via the R package “glmnet”. We chose one median parameter to establish an ideal prognosis model. Then, we measured time-dependent ROC curves and calculated the area under the ROC (AUC).

## Risk score model establishment on predicting patient overall survival

After LASSO-penalized Cox regression analysis was carried out, a risk score model was built using the aforementioned nine genes and could calculate a risk score for each sample through this formula: Risk score = GRWD1 \*3.69 – TCF7\*2.09 – VEGFA\*0.90– MNT\* 2.14 + GMIP\*1.23 – SLAMF1\*0.91 – TNFRSF25\*2.29+SLC6A3\* 1.79 + LAG3\*0.44 + 8.67. Patients were separated into high- and low-risk cohorts (median risk score) using the R software “survival” and “survminer” packages, and a *t*-test was used to distinguish death and survival events according to the risk score.

## Cox proportional hazard regression model

Univariate Cox hazard regression analysis validated the correlation among the expression levels of nine genes and OS of each patient by the R package “survival” and “survminer”. At the same time, multivariate Cox hazard regression analyses were performed too. We foresaw the regression coefficient ( $\beta$ -value) and HR. The K–M survival curve and log-rank test of every single gene were also performed by the R package referred previously .

## Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) screened SE-associated hub genes differentially expressed between healthy donors and CLL patients. We counted out the optimal soft-threshold value under the scale independence and mean connectivity analyses. CLL-related genes were clustered into various modules and gained an intersection of significant models and SE-related gene lists via Venn diagrams.

## Gene set enrichment analysis

Under the standard of risk score, we separated the participants into high- and low-risk group sets. Kyoto Encyclopaedia of Genes and Genomes (KEGG) analysis revealed a potential signaling pathway underlying the two sets via gene set enrichment analysis (GSEA v4.1.0 software).  $p < 0.05$  and a false discovery rate  $q < 0.25$  were thought to be vital in the statistic.

## Predictive nomogram for prognostic prediction

A nomogram based on independent prognostic factors of clinical traits and the polygenic risk score was constructed to predict the probability of 1-, 3-, 5-, and 10-year OS of patients with CLL. Subsequently, the discrimination of the nomogram was verified using the C-index obtained through a bootstrap method with 1,000 resamples.

## Evaluation of the sensitivity of chemotherapeutic agents

To predict the half-maximal inhibitory concentration (IC50) of chemotherapy drugs in the high- and low-risk groups of CLL patients and to infer the sensitivity of the different patients, we used the “pRRophetic” package in R.

## Statistical analysis

SPSS software vision 25.0 (SPSS, Inc., Chicago, IL, United States) and R software vision 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria) were used to analyze the data in statistics. A two-sided  $p < 0.05$  was thought vital in a statistic.

## 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.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication. XL, CL, and ZZ have wrote the main manuscript text. YM, LL, LH, and YW have prepared figures 2, 3, 4, 5, 6 and figures s1, s2, s3, s4, s5. LP and QL have prepared figure 1, figures 7, 8, figure s6 and table1.

## Funding

This work was supported by the major subject of science and technology of Anhui province: (Grant number 201903a07020030).

## References

- Ballester, S., Pineda, B., Rodrigues, P., Tormo, E., Terol, M. J., and Eroles, P. (2020). Clinical relevance of +936 C>T VEGFA and c.233C>T bFGF polymorphisms in chronic lymphocytic leukemia. *Genes (Basel)* 11 (6), E686. doi:10.3390/genes11060686
- Bichi, R., Shinton, S. A., Martin, E. S., Koval, A., Calin, G. A., Cesari, R., et al. (2002). Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. *Proc. Natl. Acad. Sci. U. S. A.* 99 (10), 6955–6960. doi:10.1073/pnas.102181599
- Bologna, C., Buonincontri, R., Serra, S., Vaisitti, T., Audrito, V., Brusa, D., et al. (2016). SLAMF1 regulation of chemotaxis and autophagy determines CLL patient response. *J. Clin. Invest.* 126 (1), 181–194. doi:10.1172/JCI83013
- Bosch, F., and Dalla-Favera, R. (2019). Chronic lymphocytic leukaemia: From genetics to treatment. *Nat. Rev. Clin. Oncol.* 16 (11), 684–701. doi:10.1038/s41571-019-0239-8
- Bulian, P., Shanafelt, T. D., Fegan, C., Zucchetto, A., Cro, L., Nuckel, H., et al. (2014). CD49d is the strongest flow cytometry-based predictor of overall survival in chronic lymphocytic leukemia. *J. Clin. Oncol.* 32 (9), 897–904. doi:10.1200/JCO.2013.50.8515
- Burger, J. A. (2020). Treatment of chronic lymphocytic leukemia. *N. Engl. J. Med.* 383 (5), 460–473. doi:10.1056/NEJMra1908213
- Cavallini, C., Lovato, O., Bertolaso, A., Zoratti, E., Malpeli, G., Mimiola, E., et al. (2015). Expression and function of the TL1A/DR3 axis in chronic lymphocytic leukemia. *Oncotarget* 6 (31), 32061–32074. doi:10.18632/oncotarget.5201
- Chuang, H. Y., Rassenti, L., Salcedo, M., Licon, K., Kohlmann, A., Haferlach, T., et al. (2012). Subnetwork-based analysis of chronic lymphocytic leukemia identifies pathways that associate with disease progression. *Blood* 120 (13), 2639–2649. doi:10.1182/blood-2012-03-416461
- Crespo, M., Bosch, F., Villamor, N., Bellosillo, B., Colomer, D., Rozman, M., et al. (2003). ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N. Engl. J. Med.* 348 (18), 1764–1775. doi:10.1056/NEJMoa023143
- Damle, R. N., Wasil, T., Fais, F., Ghiotto, V., Aletto, A., Allen, S. L., et al. (1999). Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood* 94 (6), 1840–1847. doi:10.1182/blood.v94.6.1840
- Dohner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Krober, A., Bullinger, L., et al. (2000). Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* 343 (26), 1910–1916. doi:10.1056/NEJM200012283432602

## 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/fgene.2022.1001364/full#supplementary-material>

- Fabris, S., Mosca, L., Todoerti, K., Cutrona, G., Lionetti, M., Intini, D., et al. (2008). Molecular and transcriptional characterization of 17p loss in B-cell chronic lymphocytic leukemia. *Genes Chromosom. Cancer* 47 (9), 781–793. doi:10.1002/gcc.20579
- Fan, F., Yoo, H. J., Stock, S., Wang, L., Liu, Y., Schubert, M. L., et al. (2021). Ibrutinib for improved chimeric antigen receptor T-cell production for chronic lymphocytic leukemia patients. *Int. J. Cancer* 148 (2), 419–428. doi:10.1002/ijc.33212
- Giannopoulos, K., Dmoszynska, A., Kowal, M., Wasik, S., Szczepanek, E., Bojarska-Junak, A., Rolinski, J., et al. (2009). Thalidomide exerts distinct molecular antileukemic effects and combined thalidomide/fludarabine therapy is clinically effective in high-risk chronic lymphocytic leukemia. *Leukemia* 23 (10), 1771–1778. doi:10.1038/leu.2009.98
- Hallek, M., Langenmayer, I., Nerl, C., Knauf, W., Dietzfelbinger, H., Adorf, D., et al. (1999). Elevated serum thymidine kinase levels identify a subgroup at high risk of disease progression in early, non-smoldering chronic lymphocytic leukemia. *Blood* 93 (5), 1732–1737.
- Hallek, M., Shanafelt, T. D., and Eichhorst, B. (2018). Chronic lymphocytic leukaemia. *Lancet* 391 (10129), 1524–1537. doi:10.1016/S0140-6736(18)30422-7
- Hallek, M., Wanders, L., Ostwald, M., Busch, R., Senekowitsch, R., Stern, S., et al. (1996). Serum beta(2)-microglobulin and serum thymidine kinase are independent predictors of progression-free survival in chronic lymphocytic leukemia and immunocytoma. *Leuk. Lymphoma* 22 (5-6), 439–447. doi:10.3109/10428199609054782
- He, Y., Long, W., and Liu, Q. (2019). Targeting super-enhancers as a therapeutic strategy for cancer treatment. *Front. Pharmacol.* 10, 361. doi:10.3389/fphar.2019.00361
- Herold, T., Jurinovic, V., Metzler, K. H., Boulesteix, A. L., Bergmann, M., Seiler, T., et al. (2011a). An eight-gene expression signature for the prediction of survival and time-to-treatment in chronic lymphocytic leukemia. *Leukemia* 25 (10), 1639–1645. doi:10.1038/leu.2011.125
- Herold, T., Jurinovic, V., Mulaw, M., Seiler, T., Dufour, A., Schneider, S., et al. (2011b). Expression analysis of genes located in the minimally deleted regions of 13q14 and 11q22-23 in chronic lymphocytic leukemia—unexpected expression pattern of the RHO GTPase activator ARHGAP20. *Genes Chromosom. Cancer* 50 (7), 546–558. doi:10.1002/gcc.20879
- Johnson, A. J., Yeh, Y. Y., Smith, L. L., Wagner, A. J., Hessler, J., Gupta, S., et al. (2012). The novel cyclin-dependent kinase inhibitor dinaciclib (SCH727965)

promotes apoptosis and abrogates microenvironmental cytokine protection in chronic lymphocytic leukemia cells. *Leukemia* 26 (12), 2554–2557. doi:10.1038/leu.2012.144

Kienle, D., Benner, A., Laufle, C., Winkler, D., Schneider, C., Buhler, A., et al. (2010). Gene expression factors as predictors of genetic risk and survival in chronic lymphocytic leukemia. *Haematologica* 95 (1), 102–109. doi:10.3324/haematol.2009.010298

Kotaskova, J., Tichy, B., Trbusek, M., Francova, H. S., Kabathova, J., Malcikova, J., et al. (2010). High expression of lymphocyte-activation gene 3 (LAG3) in chronic lymphocytic leukemia cells is associated with unmutated immunoglobulin variable heavy chain region (IGHV) gene and reduced treatment-free survival. *J. Mol. Diagn.* 12 (3), 328–334. doi:10.2353/jmoldx.2010.090100

Ma, H., Qu, J., Luo, J., Qi, T., Tan, H., Jiang, Z., et al. (2019). Super-enhancer-associated hub genes in chronic myeloid leukemia identified using weighted gene Co-expression network analysis. *Cancer Manag. Res.* 11, 10705–10718. doi:10.2147/CMAR.S214614

Mosca, L., Fabris, S., Lionetti, M., Todoerti, K., Agnelli, L., Morabito, F., et al. (2010). Integrative genomics analyses reveal molecularly distinct subgroups of B-cell chronic lymphocytic leukemia patients with 13q14 deletion. *Clin. Cancer Res.* 16 (23), 5641–5653. doi:10.1158/1078-0432.CCR-10-0151

Ota, H., Hikita, T., Sawada, M., Nishioka, T., Matsumoto, M., Komura, M., et al. (2014). Speed control for neuronal migration in the postnatal brain by Gmip-mediated local inactivation of RhoA. *Nat. Commun.* 5, 4532. doi:10.1038/ncomms5532

Prieto, D., and Oppezco, P. (2017). Lipoprotein lipase expression in chronic lymphocytic leukemia: New insights into leukemic progression. *Molecules* 22 (12), E2083. doi:10.3390/molecules22122083

Rassenti, L. Z., Jain, S., Keating, M. J., Wierda, W. G., Grever, M. R., Byrd, J. C., et al. (2008). Relative value of ZAP-70, CD38, and immunoglobulin mutation status in predicting aggressive disease in chronic lymphocytic leukemia. *Blood* 112 (5), 1923–1930. doi:10.1182/blood-2007-05-092882

Roose, J., Huls, G., van Beest, M., Moerer, P., van der Horn, K., Goldschmeding, R., et al. (1999). Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target Tcf1. *Science* 285 (5435), 1923–1926. doi:10.1126/science.285.5435.1923

Schweighofer, C. D., Coombes, K. R., Barron, L. L., Diao, L., Newman, R. J., Ferrajoli, A., et al. (2011). A two-gene signature, SKI and SLAMF1, predicts time-to-treatment in previously untreated patients with chronic lymphocytic leukemia. *PLoS One* 6 (12), e28277. doi:10.1371/journal.pone.0028277

Seftel, M. D., Kuruvilla, J., Kouroukis, T., Banerji, V., Fraser, G., Crump, M., et al. (2017). The CDK inhibitor AT7519M in patients with relapsed or refractory chronic lymphocytic leukemia (CLL) and mantle cell lymphoma. A Phase II study of the Canadian Cancer Trials Group. *Leuk. Lymphoma* 58 (6), 1358–1365. doi:10.1080/10428194.2016.1239259

Sordo-Bahamonde, C., Lorenzo-Herrero, S., Gonzalez-Rodriguez, A. P., Payer, A. R., Gonzalez-Garcia, E., Lopez-Soto, A., et al. (2021). LAG-3 blockade with relatlimab (BMS-986016) restores anti-leukemic responses in chronic lymphocytic leukemia. *Cancers (Basel)* 13 (9), 2112. doi:10.3390/cancers13092112

Stamatopoulos, B., Haibe-Kains, B., Equeter, C., Meuleman, N., Soree, A., De Bruyn, C., et al. (2009). Gene expression profiling reveals differences in microenvironment interaction between patients with chronic lymphocytic leukemia expressing high versus low ZAP70 mRNA. *Haematologica* 94 (6), 790–799. doi:10.3324/haematol.2008.002626

Stamatopoulos, B., Meuleman, N., De Bruyn, C., Mineur, P., Martiat, P., Bron, D., et al. (2009). Antileukemic activity of valproic acid in chronic lymphocytic leukemia B cells defined by microarray analysis. *Leukemia* 23 (12), 2281–2289. doi:10.1038/leu.2009.176

Sugimoto, N., Maehara, K., Yoshida, K., Yasukouchi, S., Osano, S., Watanabe, S., et al. (2015). Cdt1-binding protein GRWD1 is a novel histone-binding protein that facilitates MCM loading through its influence on chromatin architecture. *Nucleic Acids Res.* 43 (12), 5898–5911. doi:10.1093/nar/gkv509

Takafuji, T., Kayama, K., Sugimoto, N., and Fujita, M. (2017). GRWD1, a new player among oncogenesis-related ribosomal/nucleolar proteins. *Cell Cycle* 16 (15), 1397–1403. doi:10.1080/15384101.2017.1338987

Trojani, A., Di Camillo, B., Tedeschi, A., Lodola, M., Montesano, S., Ricci, F., et al. (2011). Gene expression profiling identifies ARSD as a new marker of disease progression and the sphingolipid metabolism as a potential novel metabolism in chronic lymphocytic leukemia. *Cancer Biomark.* 11 (1), 15–28. doi:10.3233/CBM-2012-0259

von Wenserski, L., Schultheis, C., Bolz, S., Schliffke, S., Simnica, D., Willscher, E., et al. (2021). SLAMF receptors negatively regulate B cell receptor signaling in chronic lymphocytic leukemia via recruitment of prohibitin-2. *Leukemia* 35 (4), 1073–1086. doi:10.1038/s41375-020-01025-z

Wang, Y., Nie, H., He, X., Liao, Z., Zhou, Y., Zhou, J., et al. (2020). The emerging role of super enhancer-derived noncoding RNAs in human cancer. *Theranostics* 10 (24), 11049–11062. doi:10.7150/thno.49168

Zenz, T., Eichhorst, B., Busch, R., Denzel, T., Habe, S., Winkler, D., et al. (2010). TP53 mutation and survival in chronic lymphocytic leukemia. *J. Clin. Oncol.* 28 (29), 4473–4479. doi:10.1200/JCO.2009.27.8762

Zhou, X., Shang, J., Liu, X., Zhuang, J. F., Yang, Y. F., Zhang, Y. Y., et al. (2021). Clinical significance and oncogenic activity of GRWD1 overexpression in the development of colon carcinoma. *Oncotargets Ther.* 14, 1565–1580. doi:10.2147/OTT.S290475

Zucchetto, A., Bomben, R., Dal Bo, M., Sonogo, P., Nanni, P., Rupolo, M., et al. (2006). A scoring system based on the expression of six surface molecules allows the identification of three prognostic risk groups in B-cell chronic lymphocytic leukemia. *J. Cell. Physiol.* 207 (2), 354–363. doi:10.1002/jcp.20570