Am J Transl Res 2022;14(1):197-210
www.ajtr.org /ISSN:1943-8141/AJTR0139700

Original Article

SPINK2 is a prognostic biomarker related to immune infiltration in acute myeloid leukemia

Xiaohe Chen[1], Lifen Zhao[2], Tian Yu[3], Jue Zeng[1], Ming Chen[1]

[1]Department of Blood Transfusion, [2]Clinical Laboratory, The Third Affiliated Hospital of Wenzhou Medical University, Ruian 325200, Zhejiang, China; [3]Graduate School of Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing 100730, China

Received October 9, 2021; Accepted November 30, 2021; Epub January 15, 2022; Published January 30, 2022

Abstract: Background: Serine peptidase inhibitor Kazal type 2 (SPINK2) has been reported to be involved in certain cancers. We conducted an in-depth investigation on the role and mechanism of SPINK2 in acute myeloid leukemia (AML). Methods: The relationship between SPINK2 expression and AML clinicopathologic characteristics was determined using The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) database. Concomitantly, we used Kaplan-Meier survival analysis, as well as univariate and multivariate regression analyses to evaluate SPINK2 as a prognostic marker of AML. Additionally, we annotated the enrichment and function of SPINK2 using Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Sets Enrichment Analysis (GSEA). The CIBERSORT algorithm was used to analyze the relationship between SPINK2 expression and immune infiltration. Results: SPINK2 expression was significantly higher in AML patients compared to healthy individuals ($P<0.001$). The area under receiver operating characteristic curve in the GSE9476 dataset was 0.660, whereas that in the Genotype-Tissue Expression (GTEx) and TCGA datasets was 0.935. In addition, GSEA also showed that several pathways were enriched in the group with high SPINK2 expression, such as PI3K-AKT signaling, PD-L1 expression, and checkpoint pathways. Analysis of immune infiltration showed that SPINK2 expression was correlated with certain immune infiltrating cells. Cox multivariate analysis revealed that the level of SPINK2 was an independent risk factor for the progression of AML ($P<0.001$). Moreover, age, M1, M5, M6, and CytoRisk-Poor also affected the progression of AML ($P<0.05$). The C-index of the nomogram in our internal validation was 0.702. Conclusion: The high expression of SPINK2 in AML suggests that SPINK2 may play an important role in the immune microenvironment and thus could be a biomarker for diagnosis and prognosis of AML.

Keywords: Serine peptidase inhibitor Kazal-type 2, acute myeloid leukemia, immune cell infiltrate, prognosis, TCGA, GEO

Introduction

Acute myeloid leukemia (AML) is a myeloid precursor-derived hematopoietic malignancy and the most common acute leukemia in adults [1]. AML is known to usually impair normal hematopoietic function, leading to severe infections, anemia, and hemorrhage. Some patients also present with extramedullary disease with involvement of the central nervous system [2]. In recent years, immunotherapy has been considered a promising treatment for hematologic malignancies and solid tumors [3]. However, the prognosis for AML patients remains poor, with over half of patients dying from the disease due to the lack of high specificity of the target antigen and the heterogeneity of AML [4]. The pathogenesis of AML is complicated. Current research has shown that the most important causative factors are environmental influences and genetic factors [5, 6]. However, the molecular mechanisms involved in the development of AML remain largely unclear, further obstructing early diagnosis. Therefore, we need to explore the underlying molecular mechanisms and identify new biomarkers to halt or slow disease progression.

Serine peptidase inhibitor Kazal type 2 (SPINK2) is a member of the family of Kazal-type serine protease inhibitors. These inhibitors include at least one Kazal domain and six cyste-

ine residues for the formation of three interlinked disulfide bonds connected together in a 1-4, 2-5, and 3-6 pattern [7]. In particular, SPINK2 is a typical trypsin inhibitor that was originally identified and purified from human sperm. It is mainly synthesized in the testis, epididymis, and seminal vesicle, and its antibacterial activity has been suggested to be involved in fertility [8].

There have only been a few studies on the role of SPINK2 in tumors so far. For instance, Hoefnagel et al. [9] found that SPINK2 was highly expressed in primary cutaneous follicle center cell lymphomas. Interestingly, the expression of SPINK2 was absent or very low in other primary cutaneous large B-cell lymphomas. Nonetheless, some other studies have shown large differences in the expression of SPINK2 in CD34+ and CD133+ primitive hematopoietic stem/progenitor cells, indicating that SPINK2 might play important roles in stem cell function [10]. In a recent study of AML, researchers used bioinformatics methods to analyze the differentially-expressed genes (DEGs) data retrieved from the Gene Expression Omnibus (GEO) and the Gene Expression Profile Interactive Analysis database and confirmed that the high expression of SPINK2 in AML might be related to the poor prognosis [11]. However, the molecular mechanism(s) by which SPINK2 influences AML remains unknown.

In this study, we analyzed the gene expression profiles of AML patients in the Genotype-Tissue Expression (GTEx), The Cancer Genome Atlas (TCGA), and GEO datasets, respectively. We screened out DEGs and then performed functional analyses. In addition, we also constructed a protein-protein interaction (PPI) and ceRNA network, assessed immune cell infiltration, and performed correlation analyses using diagnostic markers. Finally, we also addressed the relevance of SPINK2 expression with respect to patient prognosis. Our results might provide novel insights into the mechanisms of leukemogenesis and reveal a diagnostic and prognostic value of SPINK2.

Materials and methods

Data download and data preprocessing

Both the entire RNA-sequencing profile data and corresponding clinical information of 151 AML patients were downloaded from TCGA database (https://portal.gdc.cancer.gov/). The 755 blood sample data of the normal group was downloaded from the GTEx database. The AML expression profiling dataset (GSE9476) was downloaded from the GEO (https://www.ncbi.nlm.nih.gov/geo/) database using the R/Bioconductor package GEOquery [12] (version 4.1.0, http://r-project.org/). All samples in the dataset were derived from Homo sapiens, based on the GPL96 Affymetrix Human Genome U133A Array platform (Affymetrix; Thermo Fisher Scientific, Inc.). Blood samples from 26 patients in the AML group and 38 healthy donors were selected for further analyses. The HTSeq_counts in TCGA_LB and GTEx datasets were used to remove batch effects and standardization disposal and were then converted to TPM standardized data. The gene expression matrix in the GSE9476 dataset was also standardized.

Differentially expressed single gene and organ distribution

We compared the expression of SPINK2 between normal and AML in the GTEx, TCGA, and GSE9476 datasets. To construct a receiver operating characteristic (ROC) prediction model for the normal and AML groups using the pROC package [13], we employed the GTEx_TCGA and GSE9476 dataset to distinguish the different status between the normal and the AML groups. We used the GTEx data to extract the level of SPINGK2 expression in each organ.

Screening of DEGs and functional analyses

Patients were divided into two groups according to the level of SPINK2 expression. The groups were filtered for the identification of DEGs in the GTEx_TCGA and GSE9476 datasets using the limma package [14]. We also generated volcano plots of DEGs using the ggplot2 package [15]. Filtration conditions were as follows: |log2FC| >0.3 and *P*-value <0.05. We accordingly identified common DEGs by taking the intersection of the two groups of DEGs. We then used the cluster profiler package for Gene Ontology (GO) and Kyoto Encyclopedia of genes and genomes (KEGG) enrichment analysis [16], in which differences with $P \leq 0.05$ were considered statistically significant. The gene expression matrix was used to conduct gene sets enrichment analysis (GSEA) using the clusterProfiler package. We

selected msigdb.v7.0.entrez.gmt (Download it from http://www.gsea-msigdb.org/gsea/msigdb/collections.jsp) as the reference gene set, and values with a false discovery rate (FDR) <0.25 and P<0.05 were considered as significantly enriched.

Construction of a PPI network

The PPI network was constructed using the STRING database [17] (version 11.0; http://string-db.org). We selected an interaction with a confidence score greater than 0.4 to construct the PPI network and analyze the relationship between each DEG. Cytoscape software [18] was used for visualizing the molecular interaction networks. The top 10 hub genes of the gene-gene interaction network were extracted by analyzing the networks using the Cytoscape plugin cytoHubba [19].

Construction of ceRNA network

We constructed the miRNA-mRNA interaction network by selecting the 79 identified DEGs from the luciferase reporter assay and western blot analyses, based on the miRTarBase database (http://mirtarbase.mbc.nctu.edu.tw/php/index.php) using the multiMiR package [20]. The LncRNA-miRNA interaction network was based on the starBase V3 database (http://starbase.sysu.edu.cn/), using the following thresholds: pancancerNun >5, clipExpNum >12. The lncRNA-miRNA-mRNA regulatory network was visualized using the ggalluvial package.

Immune cell infiltration and biomarker correlation analyses

CIBERSORT is a deconvolution tool of the transcriptome expression matrix based on the principle of linear support vector regression, which is used to estimate the composition and abundance of immune cells in mixed cell populations [21]. To evaluate the immune cell infiltrate matrix, we uploaded the gene expression matrix data to CIBERPORT and set *P* values <0.05 as the screening condition. The distribution of the infiltration of the 22 immune cell types in each sample was plotted into bar charts using a graphics package. We next drew the relevant heatmap using the corrplot package, which showed the correlation of the 22 types of immune cell infiltration [22]. The violin diagram was drawn using the ggplot2 package, revealing the differences in the expression of the 22 types of immune cells.

Clinical correlation analysis

To investigative the prognostic value of SPINK2 and its correlation to clinical factors, we performed survival analysis and timeROC analysis based on TCGA_AML data using the survival and survminer packages. Clinical factors were selected using univariate and multivariate Cox regression analyses by forest plot. To predict the survival of AML patients, a nomogram was constructed based on the multivariable logistic model. The accuracy and clarity of the nomogram were evaluated using a correction curve.

Results

SPINK2 is highly expressed in AML

We extracted the SPINK2 expression values of the AML and normal groups in the GSE9476 and GTEx_TCGA datasets separately and performed the Wilcoxon-test to detect differences between the two groups. We accordingly found that the expression of SPINK2 was higher in the AML group compared with that in the normal group (**Figure 1A**, **1B**). In the GSE9476 dataset, we observed that the corresponding area under the ROC curve (AUC) for distinguishing between the AML and normal groups was 0.660 (**Figure 1C**). However, in the GTEx_TCGA dataset, the AUC was demonstrated to be 0.935 (**Figure 1D**). Therefore, we concluded that SPINK2 expression could be used to significantly distinguish the AML from normal groups. We further noted that the distribution of SPINK2 also differed across the whole body; the highest expression was found in the testis, whereas the lowest was in the bladder (**Figure 1E**).

Analysis of DEGs grouped according to SPINK2 expression

Differential gene analysis was performed by fold change method. We respectively analyzed the data from each dataset to screen for DEGs (P<0.05, and |log2FC| >3). We accordingly identified a total of 114 upregulated and 288 downregulated genes in GSE9476 and 1139 upregulated and 1102 downregulated genes in TCGA (**Figure 2**).

Functional enrichment analysis of SPINK2-related genes

To examine the functional roles of the SPINK2-related differentially expressed genes, we first

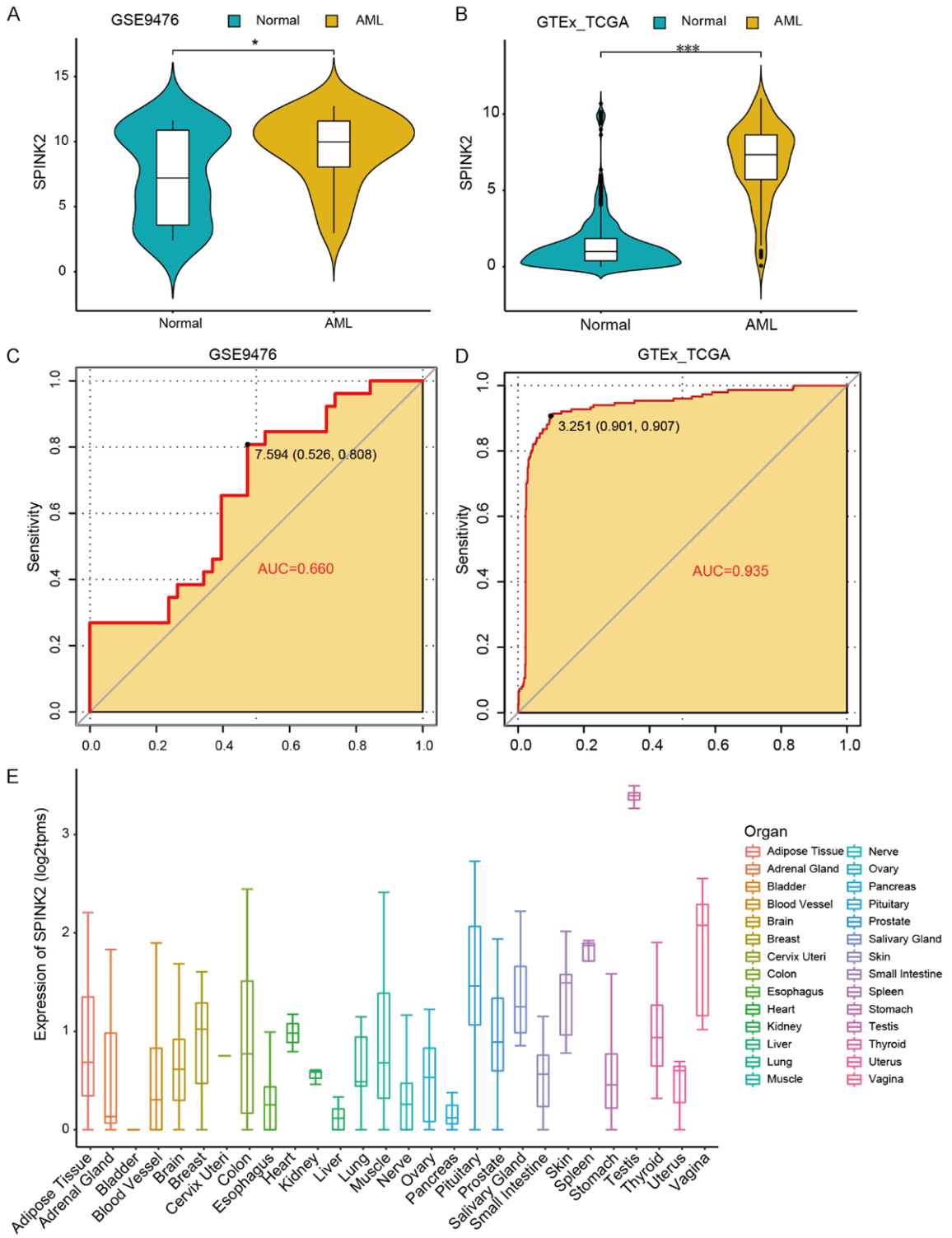

Figure 1. Expression of SPINK2. A and B: The level of SPINK2 expression was higher in AML compared with healthy individuals in different datasets (*P<0.05, ***P<0.001). C: Diagnostic value of SPINK2 expression in AML in GSE30219, AUC=0.660. D: Diagnostic value of SPINK2 expression in AML in GTEx_TCGA, AUC=0.9350. E: SPINK2 expression in the whole body.

took the intersection of 135 DEGs within the AML subset in GSE9476 and 2162 DEGs in TCGA, obtaining 79 intersection genes (**Figure 3A**). We then performed GO and KEGG analysis

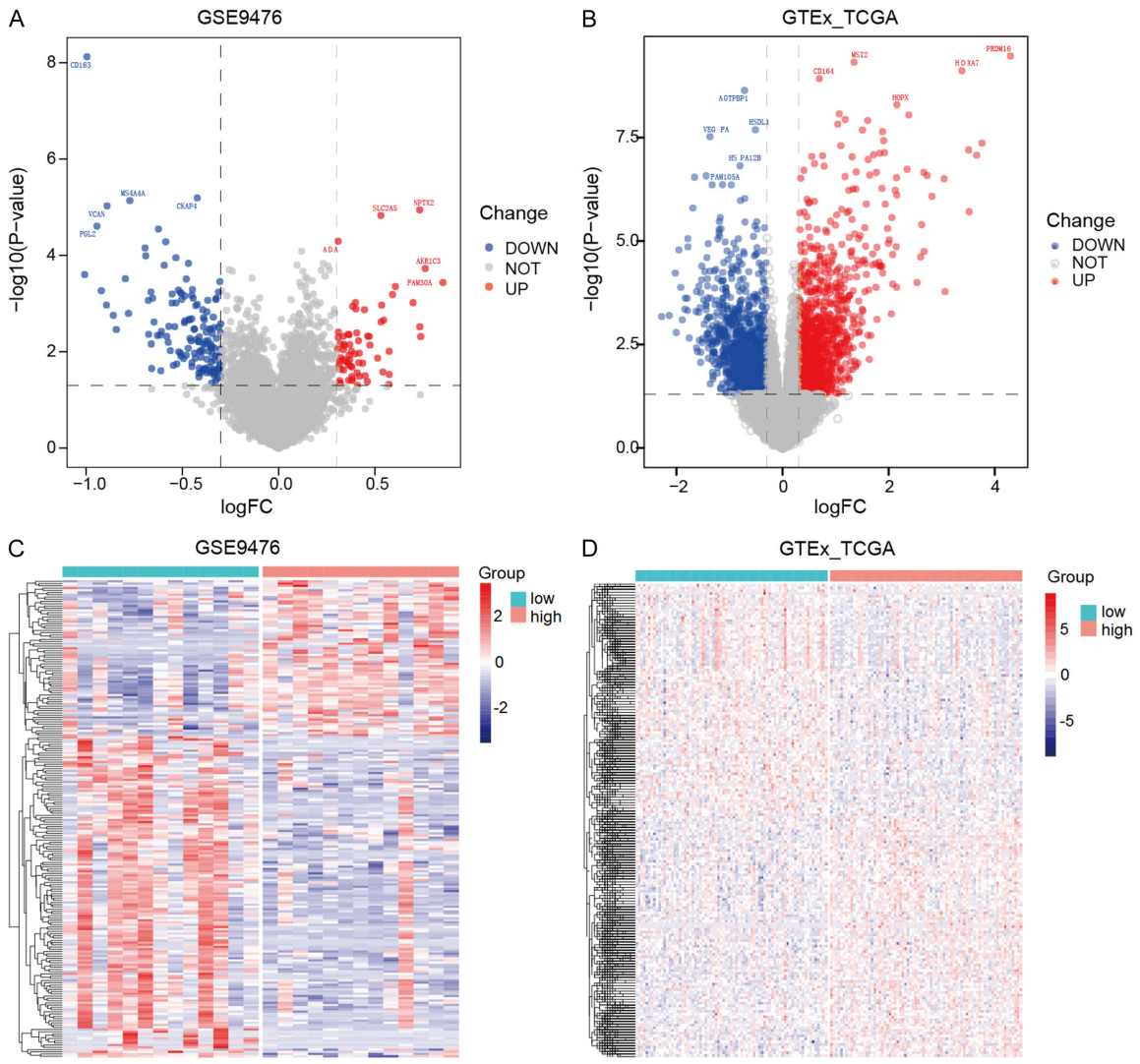

Figure 2. Differentially-expressed genes (DEGs) between high and low SPINK2 expression groups. A: Volcanic map of the DEGs between the high and low SPINK2 expression groups of the AML subset in GSE9476; red dots indicate high expression, whereas blue dots indicate low expression. B: Volcanic map of DEGs between the high and low SPINK2 expression groups in TCGA; red dots indicate high expression, whereas blue dots indicate low expression. C: Heatmap of the DEGs in different SPINK2 groups within the AML subset in GSE9476. D: Heatmap of DEGs in different SPINK2 groups in TCGA.

on the 79 intersection genes. KEGG analysis showed that these genes were related to complement and coagulation cascades, rheumatoid arthritis, phagosome, leishmaniasis, and the Staphylococcus aureus infection signaling pathway. GO analysis showed that they were associated with neutrophil degranulation that is related to neutrophil activation, which participates in the immune response, positively regulating the macrophage colony stimulating factor response, detection of external biotic stimulus, and the regulation of macrophage colony stimulating factor response (**Figure 3C**). Pathway enrichment analysis also showed a significant correlation of the gene expression with the PI3K-AKT pathway (Figure S1), expression of PD-L1, and checkpoint pathway (Figure S2).

To avoid the bias of simply attaining an intersection gene enrichment, we performed GSEA using the msigdb.v7.0.entrez.gmt reference gene set to analyze all DEGs related to SPINK2 in TCGA. Our results revealed an association with the following terms: “alcalay AML by npm1 localization up”, “boquest stem cell up”, “GO collagen containing extracellular matrix”, “GO extracellular matrix”, “GO extracellular structure organization”, “GSE10325 lupus B cell vs lupus myeloid dn”, and “GSE29618 B

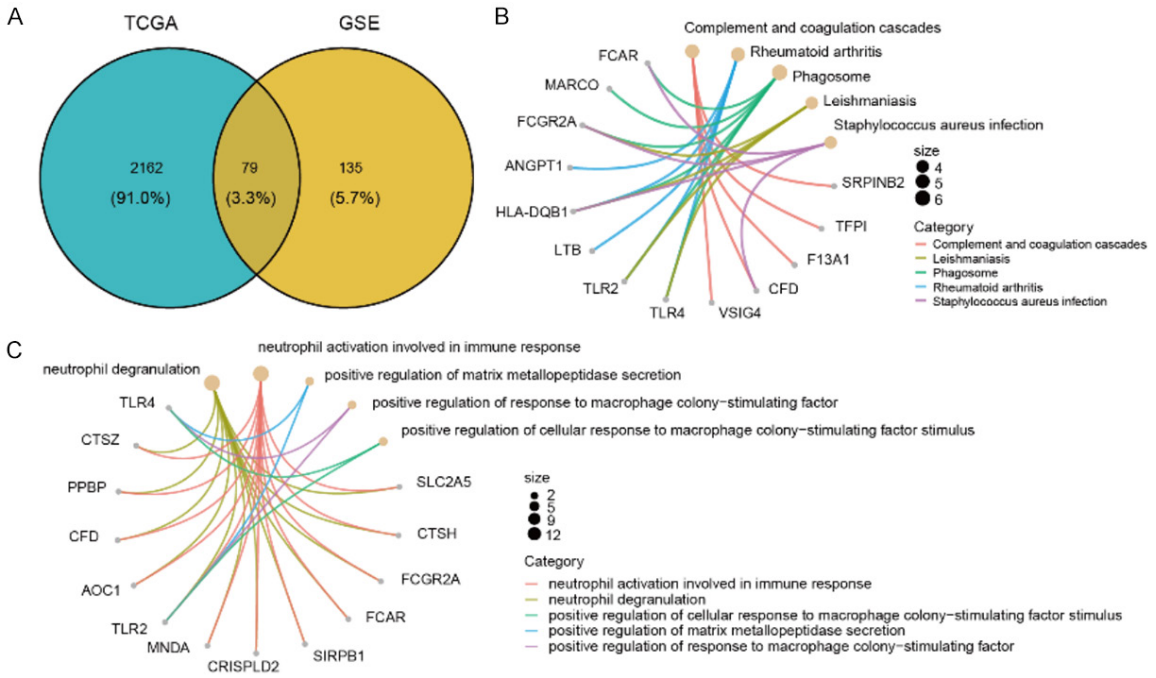

Figure 3. Significantly enriched GO annotations and KEGG pathways of SPINK2-related genes in AML. A: Venn diagram with the intersection of differentially-expressed genes in the AML subset within GSE9476 and TCGA. B: Intersection genes in KEGG analysis. C: Intersection genes in GO analysis.

cell vs monocyte day7 flu vaccine dn" dataset (**Figure 4A-H**).

Constructed PPI and ceRNA networks

To further explore the underlying mechanisms, we constructed a PPI network based on the STRING database using the Cytoscape software. We used a total of 79 intersection genes (**Figure 5A**), and then screened the top 10 hub genes and extensions using the cytoHubba plug-in. The top 10 genes were TLR4, TLR2, CD69, SERPINB2, SERPINB8, FCGR2A, CCR2, HLA-DQB1, F13A1, and PPBP. To further explore the upstream regulatory relationship, we constructed the mRNA-miRNA interaction network using the same 79 intersection genes in the multiMiR database, selected for luciferase reporter assay and western blot evidence, and identified interactions between 8 mRNAs and 11 miRNAs. For these 11 mRNAs, we predicted the interaction with 10 lncRNAs using the pancancerNun >5 and clipExpNum >12 settings on the starBase V3 database. The identified lncRNAs, miRNAs, and mRNAs constituted the ceRNA network (**Figure 5D**).

Immune cell infiltration analysis

To further confirm the correlation between SPINK2 expression and the immune microenvironment and to analyze the ratio of tumor immune infiltrating subgroups, we constructed a map of 22 immune cell types in the AML group (**Figure 6A**) and evaluated the correlation among them (**Figure 6B**). Pearson correlation analysis revealed an association between the three Tics and SPINK2 (**Figure 6C-E**). In addition, we found that eosinophils and plasma cells were positively correlated with the three Tics and SPINK2 expression, whereas resting mast cells were shown to be negatively correlated with the three Tics and SPINK2 expression. According to the correlation analysis between the level of immune cell infiltration in the SPINK2 high expression profile and SPINK2 low expression profile, a significant correlation was also observed between eosinophils, plasma cells, as well as resting mast cells and immune cells with the three Tics and SPINK2 (**Figure 6F**). These results further supported the existence of a correlation between the three Tics and the levels of SPINK2 expression that affect the activity of immune cells.

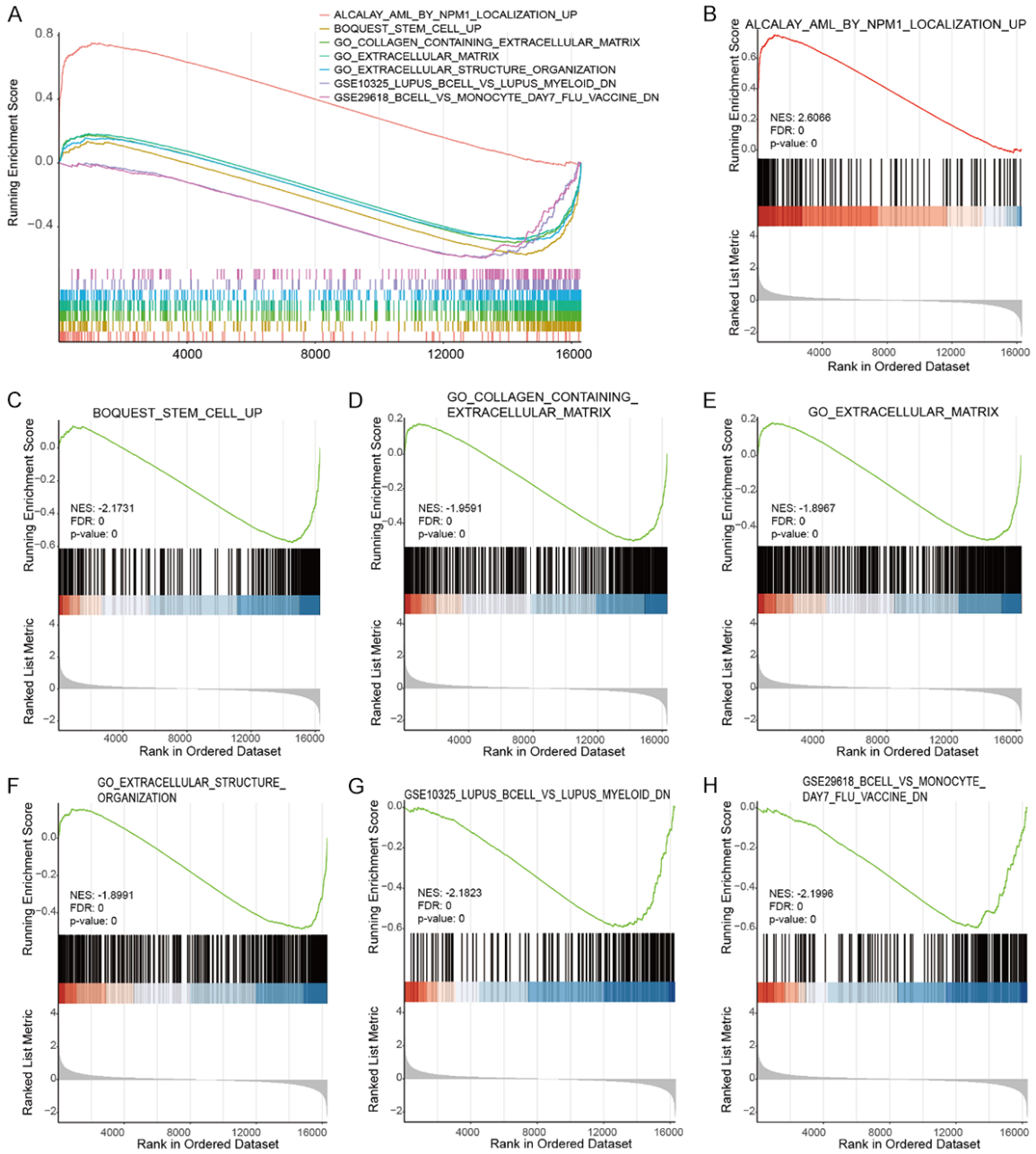

Figure 4. Gene set enrichment analysis (GSEA) plots. (A) TCGA differentially-expressed genes in GSEA; the seven most significantly enriched genes in the gene set. (B-H) Most significantly enriched gene sets: alcalay AML by npm1 localization up. (B) Bouquets stem cell up; (C) Go collagen containing extracellular matrix; (D) Go extracellular matrix; (E) Go extracellular structure organization; (F) GSE10325 lupus b cell vs lupus myeloid dn; (G) GSE29618 bcell vs monocyte day7 flu vaccine dn (H).

Clinical correlation analysis

To validate the clinical relevance of the correlation analysis between the three Tics and SPINK2, we constructed 1-year, 3-year, and 5-year survival ROC curves. We observed that the 1-year, 3-year, and 5-year AUCs were 0.617, 0.614, and 0.705, respectively. In particular, the 5-year survival AUC was the highest and was evidently associated with survival (**Figure 7A**). We used Kaplan-Meier survival analysis, as well as univariate and multivariate regression analyses to evaluate SPINK2 as a prognostic indicator of AML. We further noted that the overall survival curves were statistically significant (**Figure 7B**). To identify any correlation

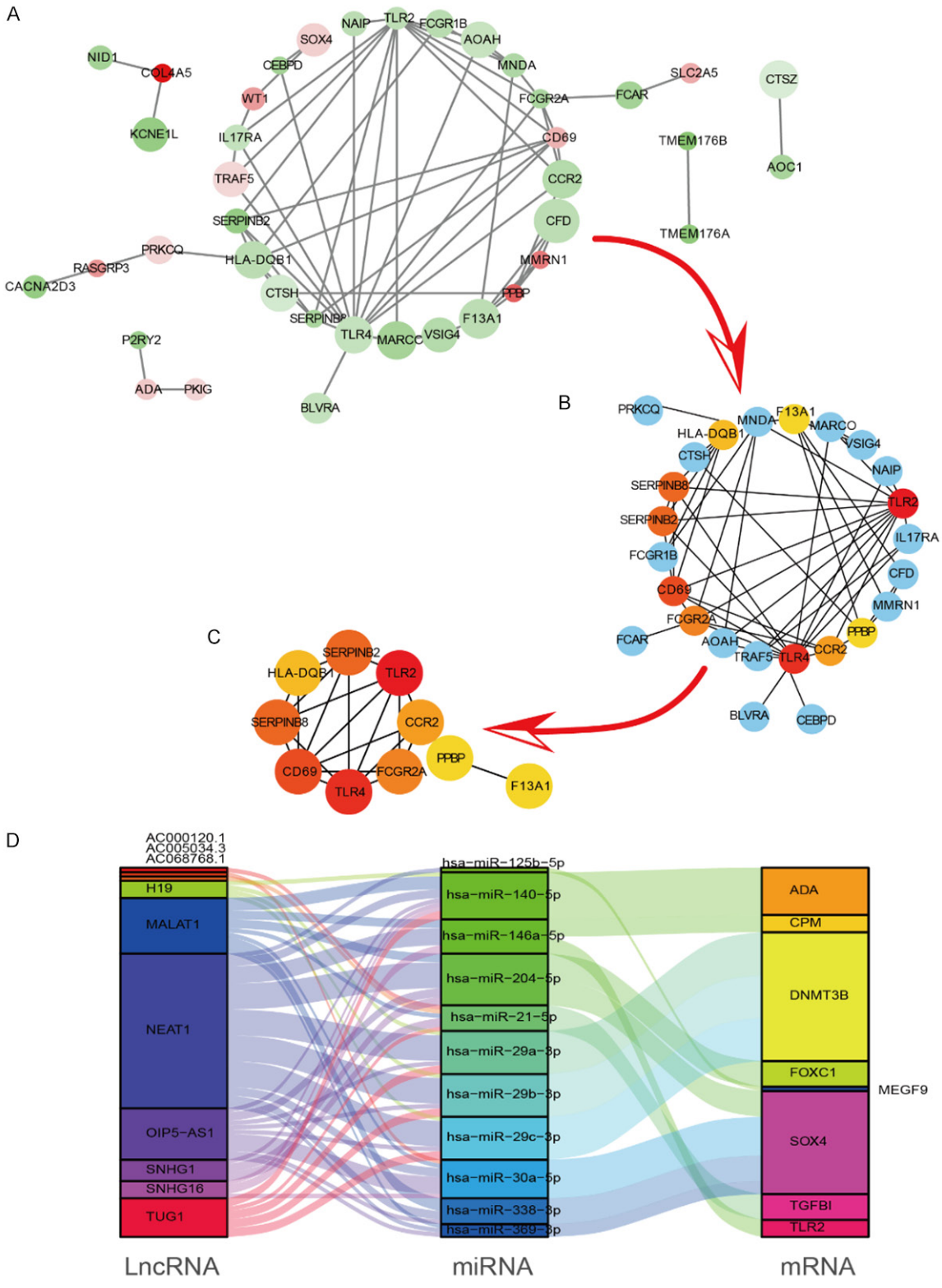

Figure 5. PPI and ceRNA network analyses. A: The PPI network of the SPINK2 interaction partners was generated using the STRING and Cytoscape; red represents upregulated genes, whereas green represents downregulated genes; the larger the adj. *P* value, the larger the diameter. B: Top 10 hub genes and extensions using the cytoHubba plug-in; the higher the enrichment scores, the darker the color; the color of expanding genes is light blue. C: Top 10 hub genes and extensions using the cytoHubba plug-in, the higher the enrichment scores the darker the color. D: The ceRNA network constructed based on the expression of the identified intersection genes.

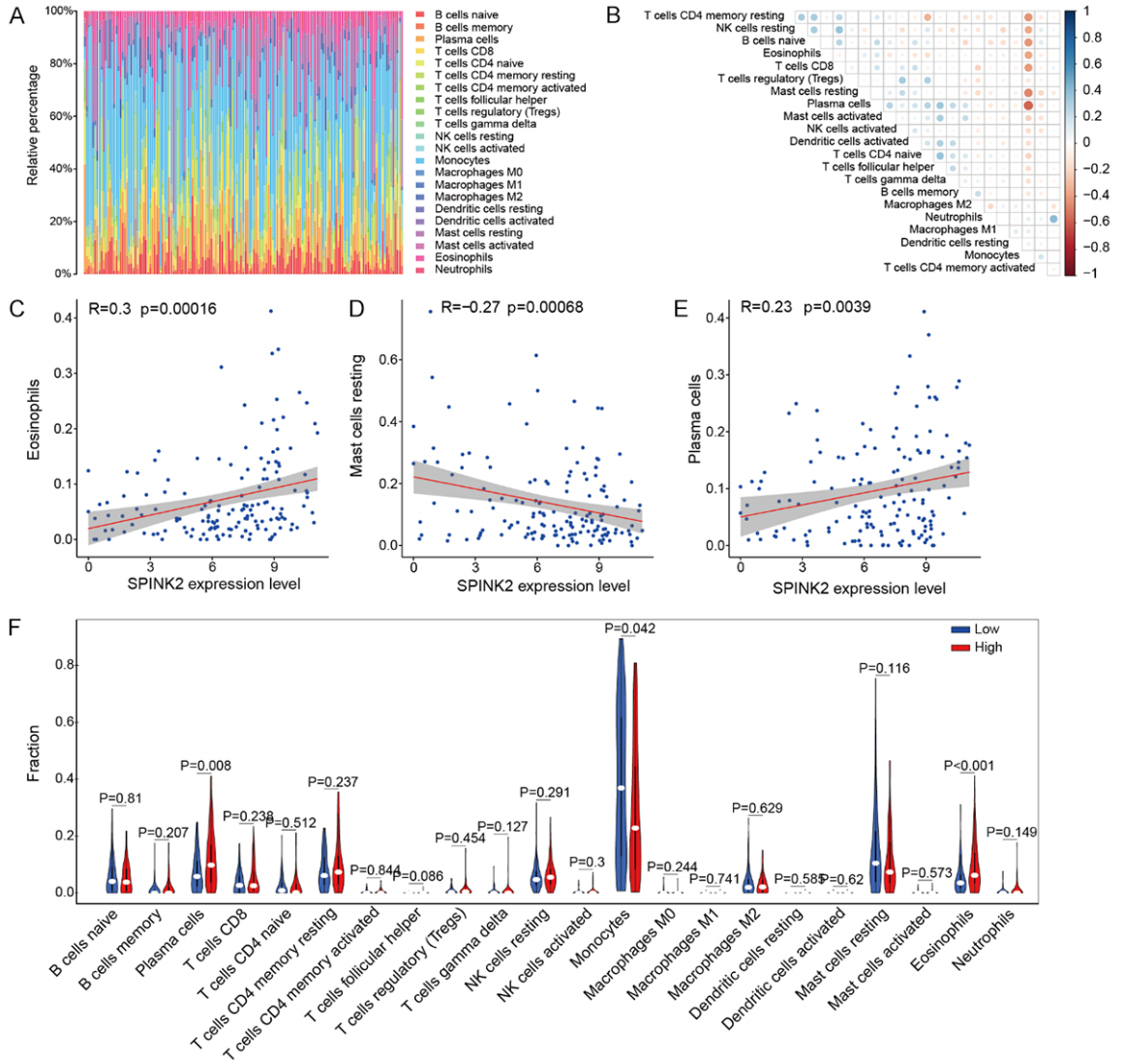

Figure 6. Evaluation of immune cell infiltration and correlation analysis. A: Bar plot showing the ratio of 22 immune cell types in TCGA_AML samples. Samples are shown in columns. B: Heatmap of the infiltration of the 22 immune cell types; blue indicates positive correlation, red indicates negative correlation; the darker the color, the stronger the correlation. C: Correlation between eosinophils and the three Tics and SPINK2 as dot plots, r=0.3, P=0.00016. D: Correlation between resting mast cells and the three Tics and SPINK2 as dot plots, r=0.27, P=0.00068. E: Correlation between plasma cells and the three Tics and SPINK2 as dot plots, r=0.23, P=0.0039. F: Comparison of the level of the 22 infiltrating immune cell types and the three Tics and SPINK2 between the high- and low-expression groups; red reflects high expression, whereas blue reflects low expression.

with other clinical factors, we performed univariate and multivariate Cox regression analyses for a number of factors, such as age, gender, leukemia, French American British morphology code (FAB) grade, AML cytogenetics risk category, previous administration of hydroxyurea, and SPINK2 expression (**Table 1**). Multivariate Cox regression analysis revealed that age, M1, M5, M6, CytoRisk Poor, and SPINK2 were significantly associated with survival (**Figure 7C**). To investigate the prediction of survival using clinical factors, we constructed a nomogram model based on the multivariate Cox regression analyses (**Figure 7D**). We further plotted the 1-year, 3-year, and 5-year survival calibration curve, and the C-index was 0.702; the calibration curve also confirmed the validity of the model.

Discussion

AML is a heterogeneous hematological malignancy [23]. The accumulation of myeloid precursor cells in the bone marrow, peripheral

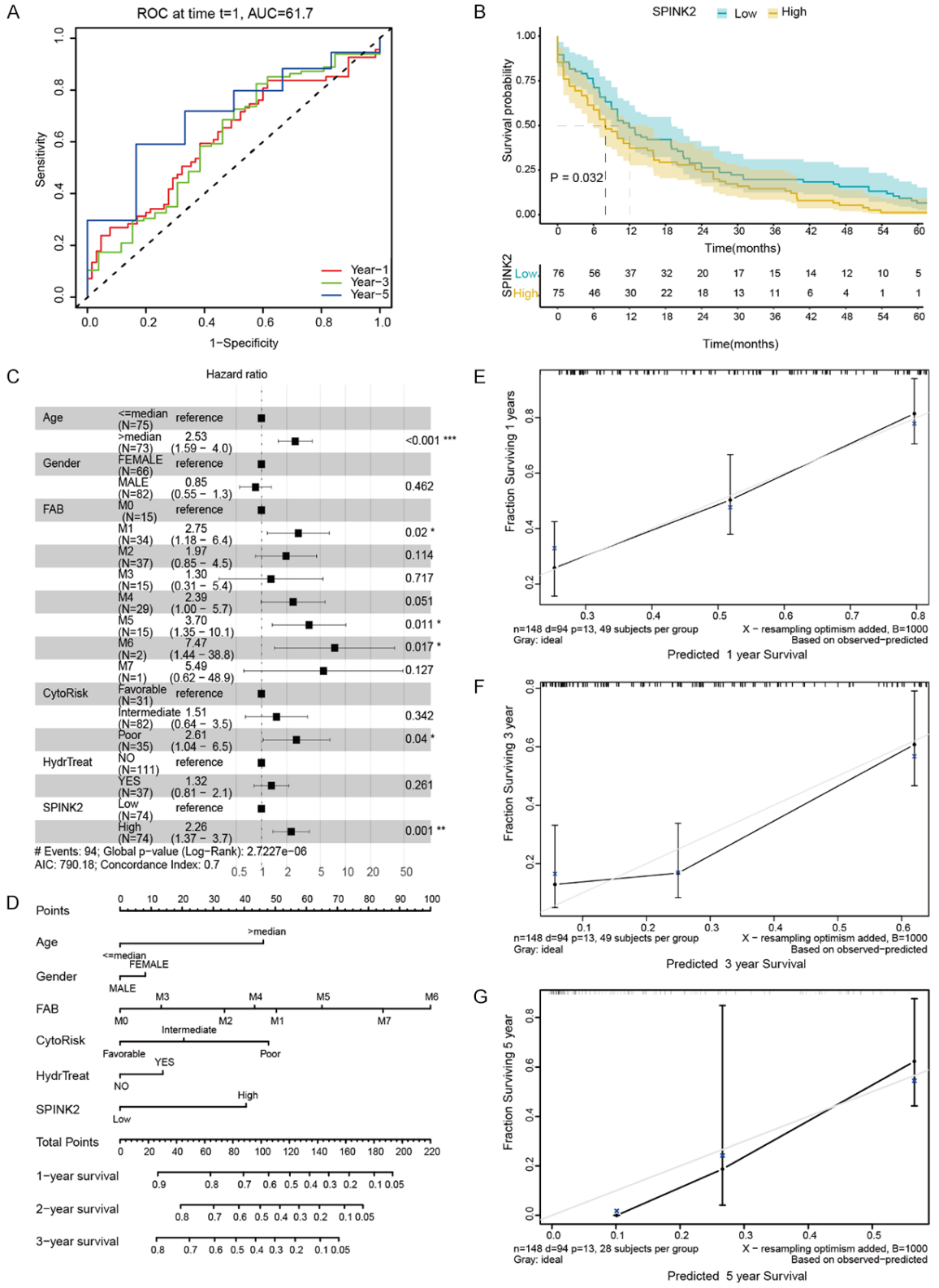

Figure 7. Clinical correlation analysis. A: Time-dependent ROC curves of SPINK2. B: Kaplan-Meier survival curves of different levels of SPINK2 expression. C: Forest plot of the results of the multivariate Cox regression analysis. D: Nomogram of the multivariate Cox regression analysis. E-G: Nomogram calibration plot used to predict the 1-year, 3-year, and 5-year overall survival.

Table 1. Univariate and multivariate Cox regression analyses for clinical factors

id	Univariate		Multivariate	
	HR (95% CL)	*P* value	HR (95% CL)	*P* value
Age	2.4542 (1.6125-3.7353)	<0.001	2.5296 (1.5912-4.0214)	<0.001
Gender	1.0151 (0.6755-1.5255)	0.942	0.8495 (0.5501-1.3120)	0.462
FAB				
M1	1.1513 (0.5474-2.4212)	0.710	2.7511 (1.1755-6.4383)	0.020
M2	1.0360 (0.4923-2.1805)	0.926	1.9656 (0.8510-4.5403)	0.114
M3	0.2718 (0.0847-0.8714)	0.028	1.3006 (0.3131-5.4022)	0.717
M4	1.1070 (0.5205-2.3544)	0.792	2.3889 (0.9966-5.7259)	0.051
M5	1.5970 (0.6593-3.8684)	0.300	3.6969 (1.3472-10.144)	0.011
M6	2.2385 (0.4844-10.345)	0.302	7.4704 (1.4394-38.771)	0.017
M7	2.0494 (0.2597-16.169)	0.496	5.4943 (0.6173-48.897)	0.127
CytoRisk				
Intermediate	3.1205 (1.5885-6.1298)	0.001	1.5123 (0.6443-3.5497)	0.342
Poor	4.6596 (2.2189-9.7848)	<0.001	2.6143 (1.0440-6.5465)	0.040
HydrTreat	1.5350 (0.9705-2.4278)	0.067	1.3186 (0.8140-2.1361)	0.261
SPINK2	1.8745 (1.2367-2.8412)	0.003	2.2610 (1.3726-3.7245)	0.001

blood, and organs can lead to systemic symptoms and death within weeks or months, if the disease is not treated [24]. Currently, the evaluation of molecular genetic damage as a predictor and the identification of prognostic markers is a relatively active research area [25]. In particular, increased focus has been given to the study of epigenetic mutations in DNMT3A, TET2, and ASXL1 [26]; however, the detailed immunological mechanism behind the development of AML remains unclear. Here, we reported that high SPINK2 expression could serve as a prognostic factor of AML. Furthermore, our analyses showed that the expression level of SPINK2 affected the activity of immune cells. Therefore, our study provided novel insights for understanding the role of SPINK2 in AML and its potential application as a biomarker.

The family of SPINK proteins is comprised of serine protease inhibitors that contain one or more Kazal domains and directly interact with the catalytic domain of proteases, blocking their activity. The Kazal domain contains three highly conserved disulfide bonds [27]. Different SPINK proteins are specifically expressed in different tissues and inhibit many serine proteases, such as pancreatic trypsin or sperm acrosin or kallikrein in the skin [28]. It is well known that the imbalance of a protein can led to serious diseases. For instance, SPINK5 knockout mice showed premature hydrolysis of the keratin connective tissue protein, resulting in Nertherton syndrome [29, 30]. Similarly, a human SPINK1 gene mutation has been associated with chronic pancreatitis [31]. Mice lacking SPINK3 (a direct homologue of human SPINK1) showed autophagic cell death and impaired pancreatic acinar cell regeneration due to increased trypsin activity [32]. Importantly, recent studies have shown that the upregulation of SPINK1 could lead to the resistance of cells to serine protease dependent apoptosis [33, 34].

Previous studies have shown that SPINK2 is highly expressed in testicular tissues. Deletion of the SPINK2 gene was reported to lead to nonobstructive azoospermia (NOA) [35]. SPINK2 is known to control intracellular proteolytic events in apoptosis. Another study revealed that SPINK2 interacted with tazarotene-induced gene 1 to inhibit the cellular invasion of testicular carcinoma cells [36]. In this study, we found that SPINK2 expression was increased in AML, and was clearly correlated with the AML FAB-type, CytoRisk, and survival time, consistent with the studies of Xue et al. [11].

To further investigate the role of SPINK2 in AML, we conducted GO and KEGG analyses and GSEA. We demonstrated that the highly expressed SPINK2 protein exhibited rich biological functions and interactions with signal

pathways and was related to PD-L1-associated tumorigenesis and the PI3K-Akt signal pathway. Programmed death-ligand 1 (PD-L1; encoded by the CD274 gene) is known to be involved in the maintenance of the complex regulation of the activity of T-cells under normal physiological conditions [37]. In many human cancers, the high expression of PD-L1 has been associated with poor prognosis [38]. In addition to cancer cells, other cells present in the tumor microenvironment and lymph nodes, such as dendritic cells, macrophages, fibroblasts, and T-cells, are also known to express PD-L1 [39]. These cells can cause the lack of antitumor immunity in the body. Based on these findings, therapeutic antibodies have been developed against PD-L1 and used in various AML cases [40]. The PI3K-AKT signaling pathway, which is involved in cell differentiation, proliferation, migration, and apoptosis is one of the most frequently dysregulated pathways in human cancers [41]. Therefore, many inhibitors targeting the pathway have been developed for the treatment of AML; however, the antileukemic effect was not as robust as expected [42].

In addition, the cellular immune microenvironment plays a very important role in the occurrence and development of tumors. To assess the correlation between SPINK2 expression and immune cell populations, we constructed a map of the 22 immune cell types using the CIBERSORT algorithm. We accordingly identified a correlation between SPINK2 and eosinophils, plasma cells, mast cells. These immune cells are known to be the key factors involved in tumorigenesis and development. As such, SPINK2 was suggested to affect the occurrence of AML through the immune infiltration mechanism.

Although we further revealed the mechanism of action of SPINK2 in AML, our study had still some limitations. To comprehensively analyze the specific roles of SPINK2 in the development of AML, we still need to consider other clinical factors, such as the treatment regimen of the patient. Second, the use of many datasets might lead to inter-batch differences that cannot be avoided or eliminated in the analysis process. Conclusively, although a multicenter study of public databases aims to compensate for the shortcomings of a single center study, a retrospective study also has its limitations, especially the inconsistent intervention measures and the lack of information. Therefore, a prospective study should be carried out in the future to balance out the bias caused by retrospective studies.

In summary, our study showed that SPINK2 was significantly upregulated in AML, and its high expression was related to the progression of AML and poor survival. We also reported the mechanism of action of SPINK2 in the tumor immune microenvironment. Our results indicated that SPINK2 might promote tumorigenesis through the occurrence of abnormal inflammation and immune response. This study provides new insights for further elucidating the pathogenesis and molecular targets of AML. However, the specific pathogenesis and molecular events during the progression of AML still require further exploration.

Disclosure of conflict of interest

None.

Address correspondence to: Ming Chen, Department of Blood Transfusion, The Third Affiliated Hospital of Wenzhou Medical University, No. 108 Wansong Road, Ruian 325200, Zhejiang, China. Tel: 0577-65866428; Fax: 0577-65866428; E-mail: chenming985017@wmu.edu.cn

References

[1] Tarumoto Y, Lin S, Wang J, Milazzo JP, Xu Y, Lu B, Yang Z, Wei Y, Polyanskaya S, Wunderlich M, Gray NS, Stegmaier K and Vakoc CR. Salt-inducible kinase inhibition suppresses acute myeloid leukemia progression in vivo. Blood 2020; 135: 56-70.

[2] Short NJ, Rytting ME and Cortes JE. Acute myeloid leukaemia. Lancet 2018; 392: 593-606.

[3] Im A and Pavletic SZ. Immunotherapy in hematologic malignancies: past, present, and future. J Hematol Oncol 2017; 10: 94-104.

[4] Dohner H, Weisdorf DJ and Bloomfield CD. Acute myeloid leukemia. N Engl J Med 2015; 373: 1136-1152.

[5] Schuz J and Erdmann F. Environmental exposure and risk of childhood leukemia: an overview. Arch Med Res 2016; 47: 607-614.

[6] Zhu B, Zhang J, Wang X, Chen J and Li C. Correlation between acute myeloid leukemia and IL-17A, IL-17F, and IL-23R gene polymorphism. Int J Clin Exp Pathol 2015; 8: 5739-5743.

[7] Kherraf ZE, Christou-Kent M, Karaouzene T, Amiri-Yekta A, Martinez G, Vargas AS, Lambert E, Borel C, Dorphin B, Aknin-Seifer I, Mitchell MJ, Metzler-Guillemain C, Escoffier J, Nef S, Grepillat M, Thierry-Mieg N, Satre V, Bailly M, Boitrelle F, Pernet-Gallay K, Hennebicq S, Faure J, Bottari SP, Coutton C, Ray PF and Arnoult C. SPINK2 deficiency causes infertility by inducing sperm defects in heterozygotes and azoospermia in homozygotes. EMBO Mol Med 2017; 9: 1132-1149.

[8] Moritz A, Lilja H and Fink E. Molecular cloning and sequence analysis of the cDNA encoding the human acrosin-trypsin inhibitor (HUSI-II). FEBS Lett 1991; 278: 127-130.

[9] Hoefnagel JJ, Dijkman R, Basso K, Jansen PM, Hallermann C, Willemze R, Tensen CP and Vermeer MH. Distinct types of primary cutaneous large B-cell lymphoma identified by gene expression profiling. Blood 2005; 105: 3671-3678.

[10] He X, Gonzalez V, Tsang A, Thompson J, Tsang TC and Harris DT. Differential gene expression profiling of CD34+ CD133+ umbilical cord blood hematopoietic stem progenitor cells. Stem Cells Dev 2005; 14: 188-198.

[11] Xue C, Zhang J, Zhang G, Xue Y, Zhang G and Wu X. Elevated SPINK2 gene expression is a predictor of poor prognosis in acute myeloid leukemia. Oncol Lett 2019; 18: 2877-2884.

[12] Davis S and Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007; 23: 1846-7.

[13] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC and Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011; 12: 77.

[14] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43: e47.

[15] Ito K and Murphy D. Application of ggplot2 to pharmacometric graphics. CPT Pharmacometrics Syst Pharmacol 2013; 2: e79.

[16] Chen L, Zhang YH, Wang S, Zhang Y, Huang T and Cai YD. Prediction and analysis of essential genes using the enrichments of gene ontology and KEGG pathways. PLoS One 2017; 12: e0184129.

[17] von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P and Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 2003; 31: 258-61.

[18] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003; 13: 2498-504.

[19] Chin CH, Chen SH, Wu HH, Ho CW, Ko MT and Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol 2014; 8 Suppl 4: S11.

[20] Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, Mahaffey S, Rossi S, Calin GA, Bemis L and Theodorescu D. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. Nucleic Acids Res 2014; 42: e133.

[21] Chen B, Khodadoust MS, Liu CL, Newman AM and Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol Biol 2018; 1711: 243-259.

[22] Salomé PA and Merchant SS. Co-expression networks in chlamydomonas reveal significant rhythmicity in batch cultures and empower gene function discovery. Plant Cell 2021; 33: 1058-1082.

[23] De Kouchkovsky I and Abdul-Hay M. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. Blood Cancer J 2016; 6: e441.

[24] Deschler B and Lübbert M. Acute myeloid leukemia: epidemiology and etiology. Cancer 2006; 107: 2099-2107.

[25] Xia T, Konno H, Ahn J and Barber GN. Deregulation of STING signaling in colorectal carcinoma constrains DNA damage responses and correlates with tumorigenesis. Cell Rep 2016; 14: 282-297.

[26] Sasaki K, Kanagal-Shamanna R, Montalban-Bravo G, Assi R, Jabbour E, Ravandi F, Kadia T, Pierce S, Takahashi K, Nogueras Gonzalez G, Patel K, Soltysiak KA, Cortes J, Kantarjian HM and Garcia-Manero G. Impact of the variant allele frequency of ASXL1, DNMT3A, JAK2, TET2, TP53, and NPM1 on the outcomes of patients with newly diagnosed acute myeloid leukemia. Cancer 2020; 126: 765-774.

[27] Wapenaar MC, Monsuur AJ, Poell J, van 't Slot R, Meijer JW, Meijer GA, Mulder CJ, Mearin ML and Wijmenga C. The SPINK gene family and celiac disease susceptibility. Immunogenetics 2007; 59: 349-357.

[28] Chen T, Lee TR, Liang WG, Chang WS and Lyu PC. Identification of trypsin-inhibitory site and structure determination of human SPINK2 serine proteinase inhibitor. Proteins 2009; 77: 209-219.

[29] Descargues P, Deraison C, Bonnart C, Kreft M, Kishibe M, Ishida-Yamamoto A, Elias P, Barrandon Y, Zambruno G, Sonnenberg A and Hovnanian A. Spink5-deficient mice mimic netherton syndrome through degradation of desmoglein

1 by epidermal protease hyperactivity. Nat Genet 2005; 37: 56-65.
[30] Yang T, Liang D, Koch PJ, Hohl D, Kheradmand F and Overbeek PA. Epidermal detachment, desmosomal dissociation, and destabilization of corneodesmosin in Spink5-/- mice. Genes Dev 2004; 18: 2354-2358.
[31] Witt H, Luck W, Hennies HC, Classen M, Kage A, Lass U, Landt O and Becker M. Mutations in the gene encoding the serine protease inhibitor, Kazal type 1 are associated with chronic pancreatitis. Nat Genet 2000; 25: 213-216.
[32] Ohmuraya M, Hirota M, Araki M, Mizushima N, Matsui M, Mizumoto T, Haruna K, Kume S, Takeya M, Ogawa M, Araki K and Yamamura K. Autophagic cell death of pancreatic acinar cells in serine protease inhibitor Kazal type 3-deficient mice. Gastroenterology 2005; 129: 696-705.
[33] Lamontagne J, Pinkerton M, Block TM and Lu X. Hepatitis B and hepatitis C virus replication upregulates serine protease inhibitor Kazal, resulting in cellular resistance to serine protease-dependent apoptosis. J Virol 2010; 84: 907-917.
[34] Lu X, Lamontagne J, Lu F and Block TM. Tumor-associated protein SPIK/TATI suppresses serine protease dependent cell apoptosis. Apoptosis 2008; 13: 483-494.
[35] Lee B, Park I, Jin S, Choi H, Kwon JT, Kim J, Jeong J, Cho BN, Eddy EM and Cho C. Impaired spermatogenesis and fertility in mice carrying a mutation in the Spink2 gene expressed predominantly in testes. J Biol Chem 2011; 286: 29108-29117.
[36] Shyu RY, Wang CH, Wu CC, Wang LK, Chen ML, Kuo CY, Lee MC, Lin YY and Tsai FM. Tazarotene-induced gene 1 (TIG1) interacts with serine protease inhibitor kazal-type 2 (SPINK2) to inhibit cellular invasion of testicular carcinoma cells. Biomed Res Int 2019; 2019: 6171065.
[37] Francisco LM, Sage PT and Sharpe AH. The PD-1 pathway in tolerance and autoimmunity. Immunol Rev 2010; 236: 219-242.
[38] Ohaegbulam KC, Assal A, Lazar-Molnar E, Yao Y and Zang X. Human cancer immunotherapy with antibodies to the PD-1 and PD-L1 pathway. Trends Mol Med 2015; 21: 24-33.
[39] Zou W, Wolchok JD and Chen L. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: mechanisms, response biomarkers, and combinations. Sci Transl Med 2016; 8: 328rv4.
[40] Isidori A, Cerchione C, Daver N, DiNardo C, Garcia-Manero G, Konopleva M, Jabbour E, Ravandi F, Kadia T, Burguera AF, Romano A, Loscocco F, Visani G, Martinelli G, Kantarjian H and Curti A. Immunotherapy in acute myeloid leukemia: where we stand. Front Oncol 2021; 11: 656218.
[41] Janku F, Yap TA and Meric-Bernstam F. Targeting the PI3K pathway in cancer: are we making headway? Nat Rev Clin Oncol 2018; 15: 273-291.
[42] Nepstad I, Hatfield KJ, Grønningsæter IS, Aasebø E, Hernandez-Valladares M, Hagen KM, Rye KP, Berven FS, Selheim F, Reikvam H and Bruserud Ø. Effects of insulin and pathway inhibitors on the PI3K-Akt-mTOR phosphorylation profile in acute myeloid leukemia cells. Signal Transduct Target Ther 2019; 4: 20-31.

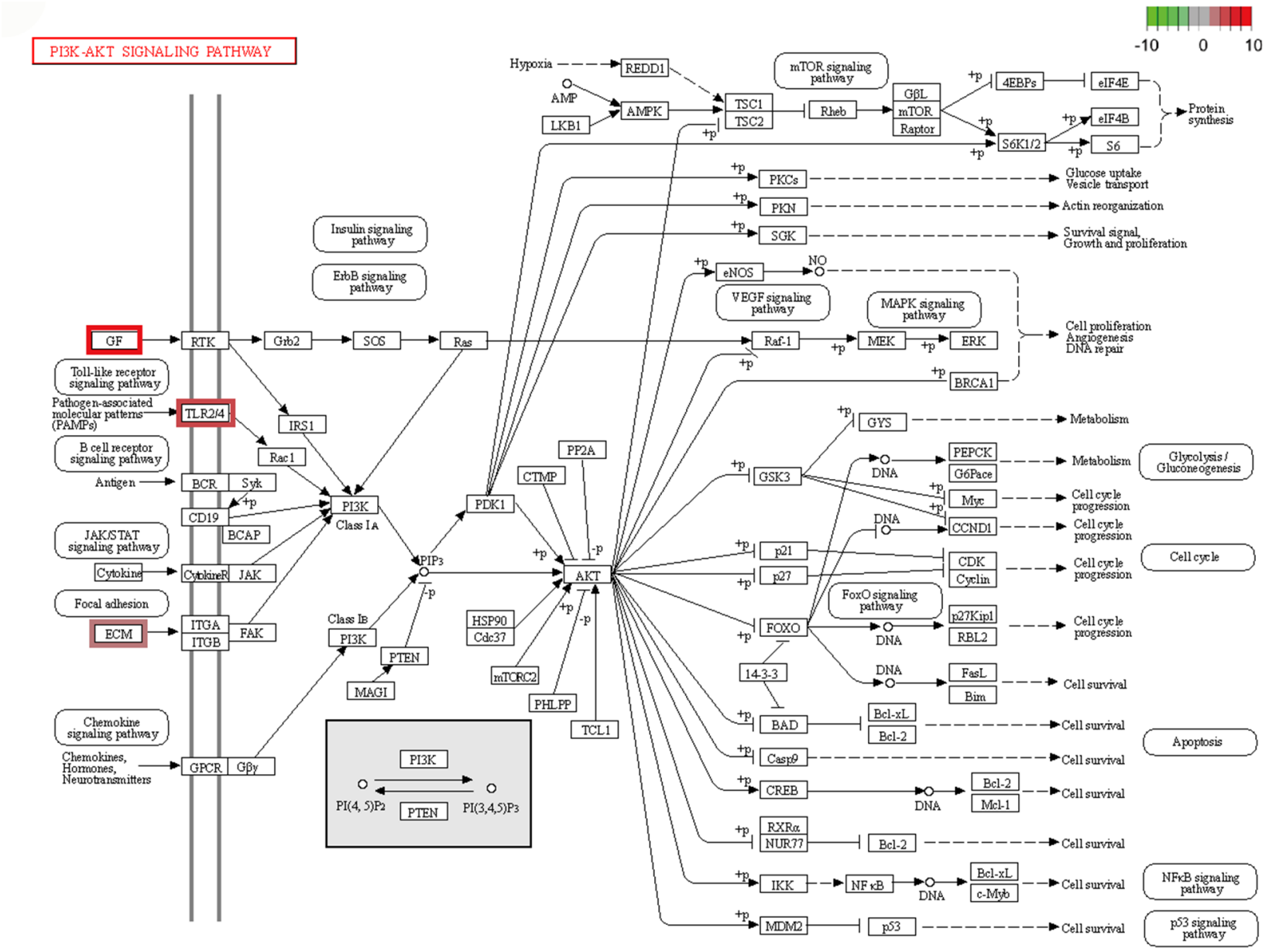

Figure S1. Intersection genes in enriched PI3K-AKT pathway.

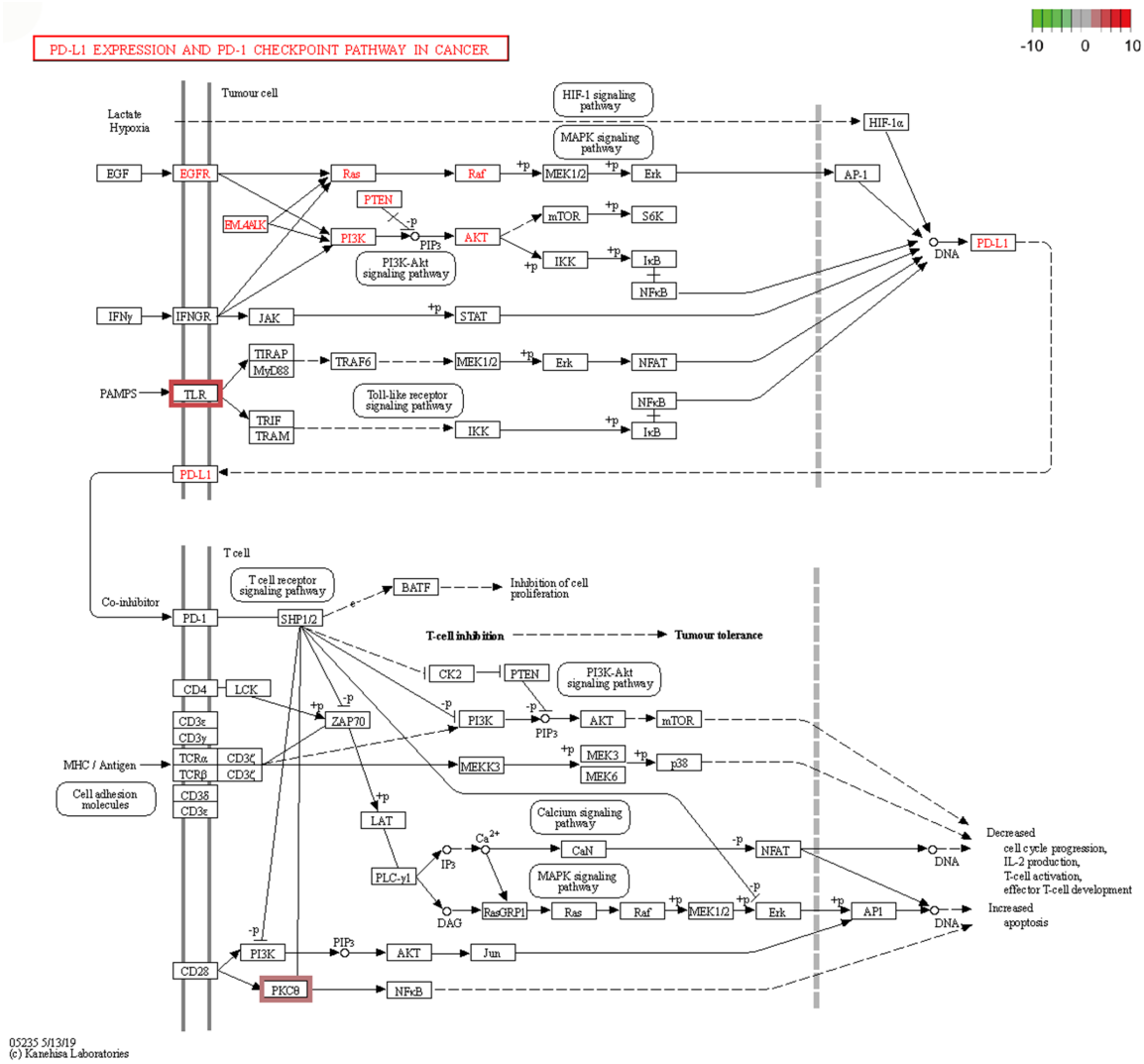

Figure S2. Intersection genes in enriched PD-L1 expression and checkpoint pathway.