Hindawi

Research Article

Machine Learning Assistants Construct Oxidative Stress-Related Gene Signature and Discover Potential Therapy Targets for Acute Myeloid Leukemia

Jinhua Zhang,[1] **Zhenfan Chen,**[1] **Fang Wang,**[2,3] **Yangbo Xi,**[1] **Yihan Hu,**[1] **and Jun Guo**[1]

[1]*The First Clinical Medical College of Jinan University, First Affiliated Hospital of Jinan University, Guangzhou 510630, China*
[2]*School of Life Science, Northwestern Polytechnical University, Xian, 710072, China*
[3]*Department of Hematology and Immunology, Myeloma Center Brussels, Vrije Universiteit Brussel, Brussels B-1090, Belgium*

Correspondence should be addressed to Jun Guo; guojun2009@jnu.edu.cn

Background. Oxidative stress (OS) is associated with the development of acute myeloid leukemia (AML). However, there is lack of relevant research to confirm that OS-related genes can guide patients in risk stratification and predict their survival probability. *Method.* First, we Data from three public databases, respectively. Then, we use batch univariate Cox regression and machine learning to select important characteristic genes; next, we build the model and use receiver operating characteristic curve (ROC) to evaluate the accuracy. Moreover, GSEAs were performed to discover the molecular mechanism and conduct nomogram visualization. In addition, the relative importance value was used to identify the hub gene, and GSE9476 was to validate hub gene difference expression. Finally, we use symptom mapping to predict the candidate herbs, targeting the hub gene, and put these candidate herbs into Traditional Chinese Medicine Systems Pharmacology (TCMSP) to identify the main small molecular ingredients and then docking hub proteins with this small molecular. *Results.* A total of 313 candidate oxidative stress-related genes could affect patients' outcomes and machine learning to select six potential genes to construct a gene signature model to predict the overall survival (OS) of AML patients. Patients in a high group will obtain a short survival time when compared with the low-risk group (HR = 3.97, 95% CI: 2.48-6.36; $p < 0.001$). ROC results demonstrate the model has better prediction efficiency with AUC 0.873. GSEA suggests that this gene is enriched in several important signaling pathways. Nomogram is constructed and is robust. PLA2G4A is a hub gene of signature and associated with prognosis, and Nobiletin could target PLA2G4A for therapy AML. *Conclusion.* We use two different machine learning methods to build six oxidative stress-related gene signatures that could assist clinical decisions and identify PLA2G4A as a potential biomarker for AML. Nobiletin, targeting PLA2G4, may provide a third pathway for therapy AML.

1. Introduction

Acute myeloid leukemia (AML) was derived from abnormal stem cell precursors of the myeloid lineage [1]. These proliferative clonal hematopoietic precursor cells damage the normal hematopoiesis and cause a series of clinical symptoms. Although AML is a rare disease compared with other newly diagnosed cancers, it accounts for more than 15% of acute leukemia with more than 30% mortality [2, 3]. Currently, molecular and cytogenetic features are recognised as key prognostic factors for the disease diagnosis. Hematopoietic

stem cell transplantation (HSCT) remains the only way to cure this disease, but the overall survival (OS) is still stagnant due to severe infection and acute graft-versus-host disease (aGVHD) after transplantation [3, 4]. More than 70% of patients receiving conventional chemotherapy will eventually relapse or become refractory leukemia [5]. Thus, it is of great importance to gain an understanding of the genetic variations in this disease and explore new targets for predicting prognosis and direct treatment.

Oxidative stress (OS) refers to a state of imbalance between reactive oxygen species (ROS) and antioxidant
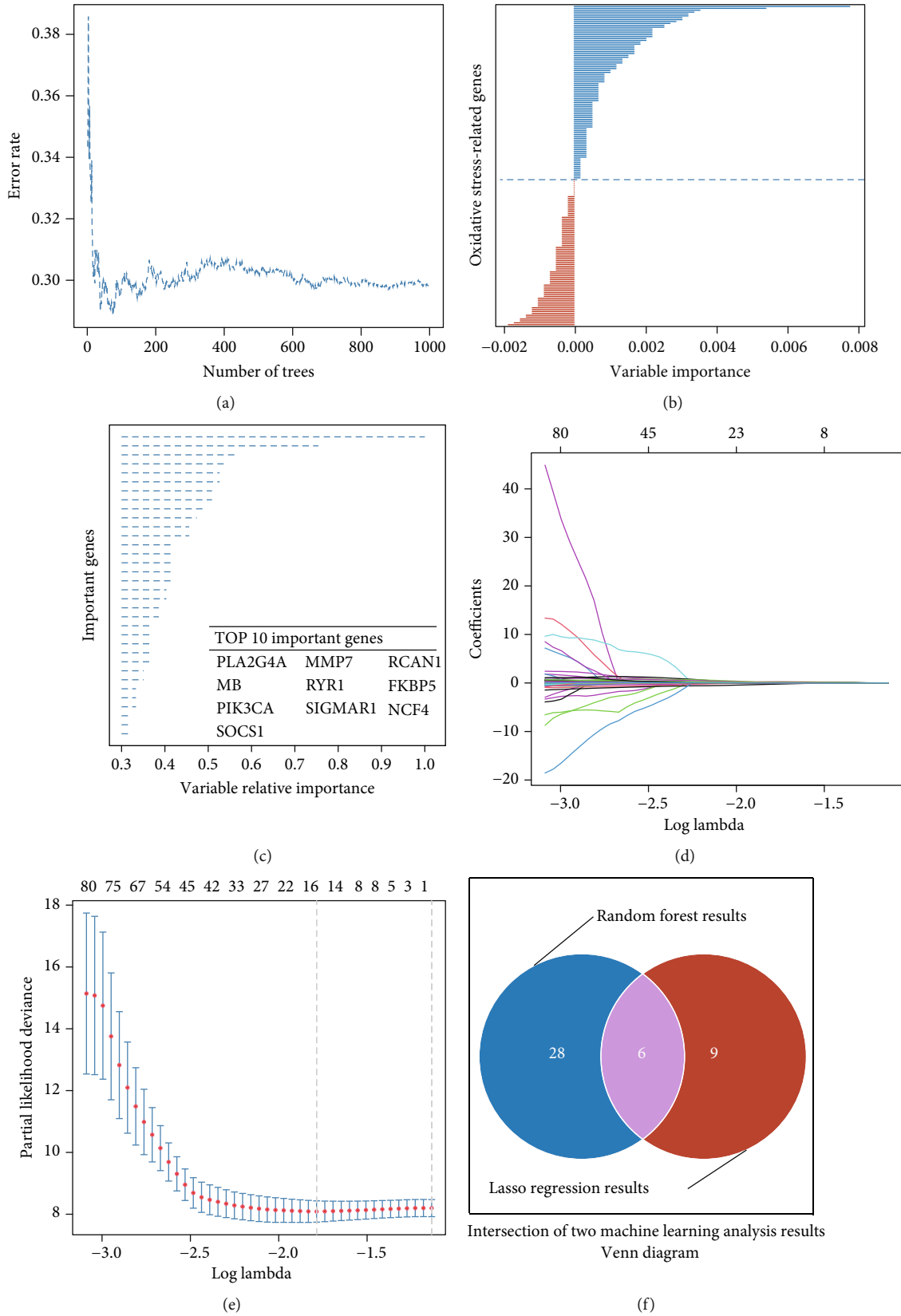
(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 1: Machine learning to identify important oxidative stress-related genes for the prognosis model. (a, b) The error rate of the random forest model is 29.83%, and the better and poor prognosis genes are ordered by their importance. (c) Thirty-four prognosis-related genes are important more than 0.3. (d, e) Fifteen candidate genes are extracted from the expression profile by the lasso regression model. (f) Six potential genes were identified after the merge of the results of the above two different machine learning algorithms.

TABLE 1: Multivariate Cox regression for model genes.

Gene symbol	Coef	HR	p value	95% CI	
				Lower	Upper
AGRN	-0.391	0.676	0.507	0.902	0.008
ETFB	0.827	2.289	1.051	4.985	0.037
PLA2G4A	0.236	1.266	0.965	1.661	0.088
RYR1	0.650	1.917	1.239	2.966	0.003
SIGMAR1	0.404	1.498	1.026	2.187	0.036
SOCS1	0.473	1.605	1.177	2.189	0.003

effects in the body, which is characterized as inflammatory infiltration of neutrophils, increased secretion of proteases, and the production of a large number of oxidative compounds [6]. ROS have been confirmed to be associated with cellular signaling and gene expression in the normal cellular process [7, 8]. However, once the endogenous ROS is not adequately eliminated by the antioxidant system, the prooxidant/antioxidant balance is lost and leads to the occurrence of OS, which will damage the biological processes and the DNA repair mechanisms, leading to kinds of diseases and carcinogenesis, such as neurological disease, cardiovascular disease, breast cancer, prostate cancer, and hematologic malignancies [9–14]. On the other hand, ROS can undertake an opposite role in tumor therapy by inducing cell apoptosis [14]. Therefore, clarifying the dual role of OS in the pathogenesis and treatment of AML and selecting the potential beneficial population would lay the foundation for individualized precision treatment.

In the present work, we use two different machine learning methods to select candidate prognosis genes in oxidative stress-related gene sets and build a six-gene signature model to predict the AML patients' outcomes. In addition, we also identify PLA2G4A as a hub gene of signature and associated with prognosis and found that Nobiletin, a type of traditional Chinese medicine, targeting PLA2G4 may provide a third pathway for therapy AML.

2. Methods and Materials

2.1. Data Obtaining and Prepared.
Public datasets were applied to this study: The Cancer Genome Atlas Program (TCGA), Genotype-Tissue Expression (GTEx), and Gene Expression Omnibus (GEO), respectively. We obtain AML data from TCGA database, which includes RNAseq records and clinical information of patients. Original count data are transferred to TPM style and extract 1399 oxidative stress-related genes, recorded by gene card database, from the expression profile to construct a new oxidative stress-related gene matrix for model building. Overall survival (OS) was defined as the endpoint. RNAseq data of donor bone marrow are from the GTEx database. GSE9476, including 38 donors and 26 AML samples, was used to validate hub gene differential expression between healthy hematopoietic cells and leukemic blasts. The 3D structure of hub protein and small-molecule structures are sourced from PDB and Pub Chem, respectively.

2.2. Batch Univariate Cox Regression to Identify Prognosis-Related Genes.
Too many genes will affect patients' outcomes; here, we use batch univariate Cox regression to identify prognosis-related genes of oxidative stress. After analysis, genes with a p value less than 0.05 were identified as significant factors, and this gene will be input into the next model to perform dimensionality reduction.

2.3. Machine Learning to Select Important Characteristic Genes.
Random forest and lasso regression are performed to identify important characteristic genes of oxidative stress. The importance of the random forest algorithm was defined as 0.3, and then, all of these requirement genes were put into the lasso regression model, which model set significant criterion was lambda is minimal.

2.4. Predictive Model Construct and Validation.
For building the final prediction model, we use multivariate Cox regression to analyze the significant oxidative stress-related genes; as above standard, the p value is also set as less than 0.05. After selecting all of these requirement genes, we build the final model for the prediction of patient outcome, according to the regression coefficients. Each patient will obtain one risk score, patients will be divided into high- and low-risk groups, according to the median value. Survival analyses were conducted by log-rank test. Forty percent of the total data were set as a test dataset to validate robust of the above model. Area under ROC was used to evaluate the predictive accuracy.

2.5. Difference Expression Genes between High-Risk and Low-Risk Groups.
To identify the differential expression genes between the high-risk and low-risk groups, the limma package was used to conduct this procedure, and the criteria of significant genes were set as absolute of log fold change more than 2 and p value less than 0.05.

2.6. GSEA of Differential Expression Genes.
The molecular mechanism of differential expression genes between the high-risk and low-risk groups is unclear; here, we use the GSEA function, which provides by the cluster profile package, to do GO and KEGG pathway enrichment analysis. p value less than 0.05 was identified as a significant enrichment result.

2.7. Nomogram Construction and Evaluation.
Nomogram is more than eyes and provides a convenient tool for clinical physicians to assist clinical decisions. Significant genes from the above multivariate Cox regression results will be considered and put into the VRPM package to conduct visualization. The area under the curve (AUC) of receiver operating characteristic (ROC) and calibration curve were used to evaluate the model's robustness.

2.8. Hub Gene Selection and Validation.
In the gene panel, not all the genes play an important role in the model, so we select the importance value of the model gene and order it from high to low, and the biggest value of the important gene was identified as the hub gene. Donor patients' bone marrow, which obtains from the GTEx database, will be
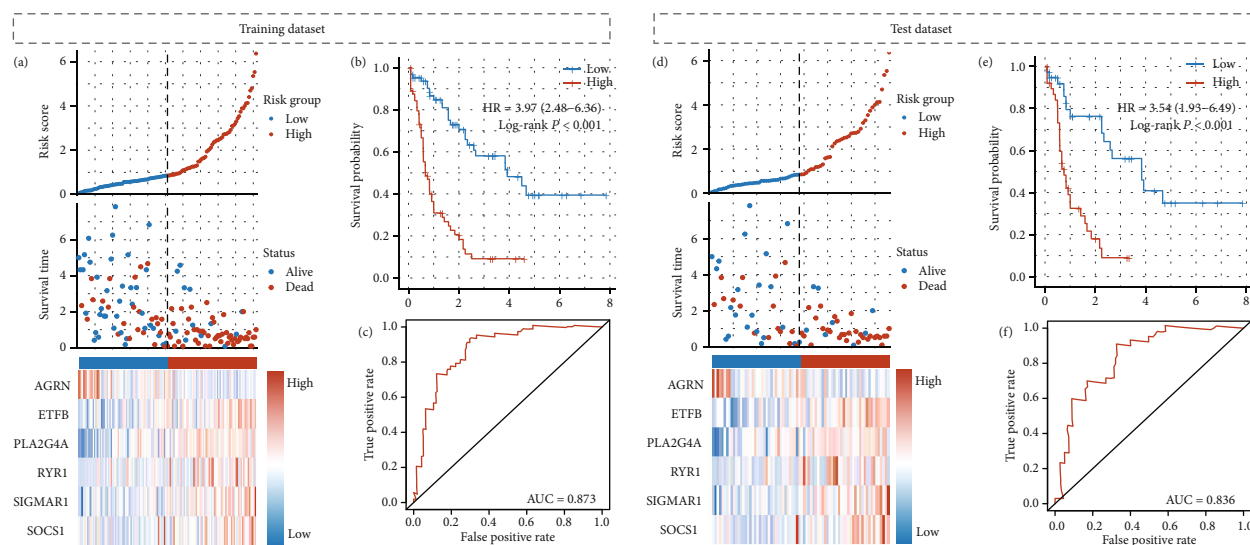
FIGURE 2: Six oxidative stress-related gene signature construction and internal validation. (a) Each patient will obtain a risk score, which is calculated by the model formula. (b) Patients in a high group will obtain a short survival time when compared with the low-risk group, and this difference is significant (HR = 3.97, 95% CI: 2.48-6.36; $p < 0.001$). (c) ROC demonstrates the model has better prediction efficiency with AUC is 0.873. (d–f) Internal validation results also support the above conclusions, and the AUC is 0.836.

compared with AML patients' bone marrow to validate hub gene difference expression. In addition, sample from GSE9476 also repeats the above operation. Moreover, survival analysis was performed to conduct by survminer package, to compare survival probability difference between high and low expression patients.

2.9. Screening Candidate Herbs Targeting Hub Protein. Traditional Chinese medicine has been confirmed that has the potential ability to target tumor markers. We use a symptom mapping (symMap Version 2.0) database to predict the candidate herbs, which will target the hub gene. Herbs with FDR less than 0.05 will be selected. The top 10 requirement herbs will be selected and also checked by the previous literature, which has been reported to have the ability of anticancer function for preparing docking with hub proteins. Candidate herbs will be input into Traditional Chinese Medicine Systems Pharmacology (TCMSP) to identify the main small molecular ingredients, according to oral bioavailability (OB) more than 30 and drug-likeliness (DL) more than 0.18.

2.10. Docking Structures between Proteins and Small Molecular Drugs. Before docking both structures, we need prepared ligand and receptor structures. So, we download the hub protein's 3D structure from the protein data bank (PDB; https://www.rcsb.org/) and obtain small molecular drug structure form Pub Chem database (https://pubchem.ncbi.nlm.nih.gov/), respectively. Then, an online tool will conduct a docking program (https://cadd.labshare.cn/cb-dock2/php/index.php).

3. Results

3.1. Prognosis Genes in Oxidative Stress-Related Gene Set of AML. We use a batch univariate regression model to filter no significant genes, which could not affect patients' out-

comes, and some prognosis-related genes in the oxidative stress-related gene set were selected. The results of the batch univariate regression model show that a total of 313 candidates' oxidative stress-related genes could affect patients' outcomes (Supplement Table 1).

3.2. Machine Learning to Select Candidate Model Genes. Random forest and lasso regression models were used to select candidate model genes. On the one hand, we put the expression matrix to the random forest model. The results show that the error rate of the random forest model is 29.83%, and the better and poor prognosis genes are ordered by importance (Figures 1(a) and 1(b)). When we set variable relative importance to more than 0.3, 34 prognosis-related genes are selected, and the top 10 significant are shown in Figure 1(c) (Supplement Table 2). On the other hand, the above expression matrix is also put into the lasso regression model, when the model selects the minimal lambda value; 15 candidate genes are extracted from the expression profile (Figures 1(d) and 1(e)). Then, we merge the results of the above two different machine learning algorithms. Six potential genes were identified to construct a gene signature model (Figure 1(f)).

3.3. Six-Gene Signature Could Predict the OS of AML Patients. After multivariate Cox regression model analysis, we find six genes from machine learning methods are included in the model (Table 1). So, in the next step, we build a six-gene signature model to predict the OS of AML patients, according to the coefficient of multivariate Cox regression. After building the signature model, every patient will obtain a risk score, which calculated by the model formula, risk score $= (-0.391) \times$ AGRN $+ (0.827) \times$ ETFB $+ (0.236) \times$ PLA2G4A $+ (0.650) \times$ RYR1 $+ (0.404) \times$ SIGMAR $1 + (0.473) \times$ SOCS1. After the count, patients will be divided into low-risk and high-risk groups, based on median
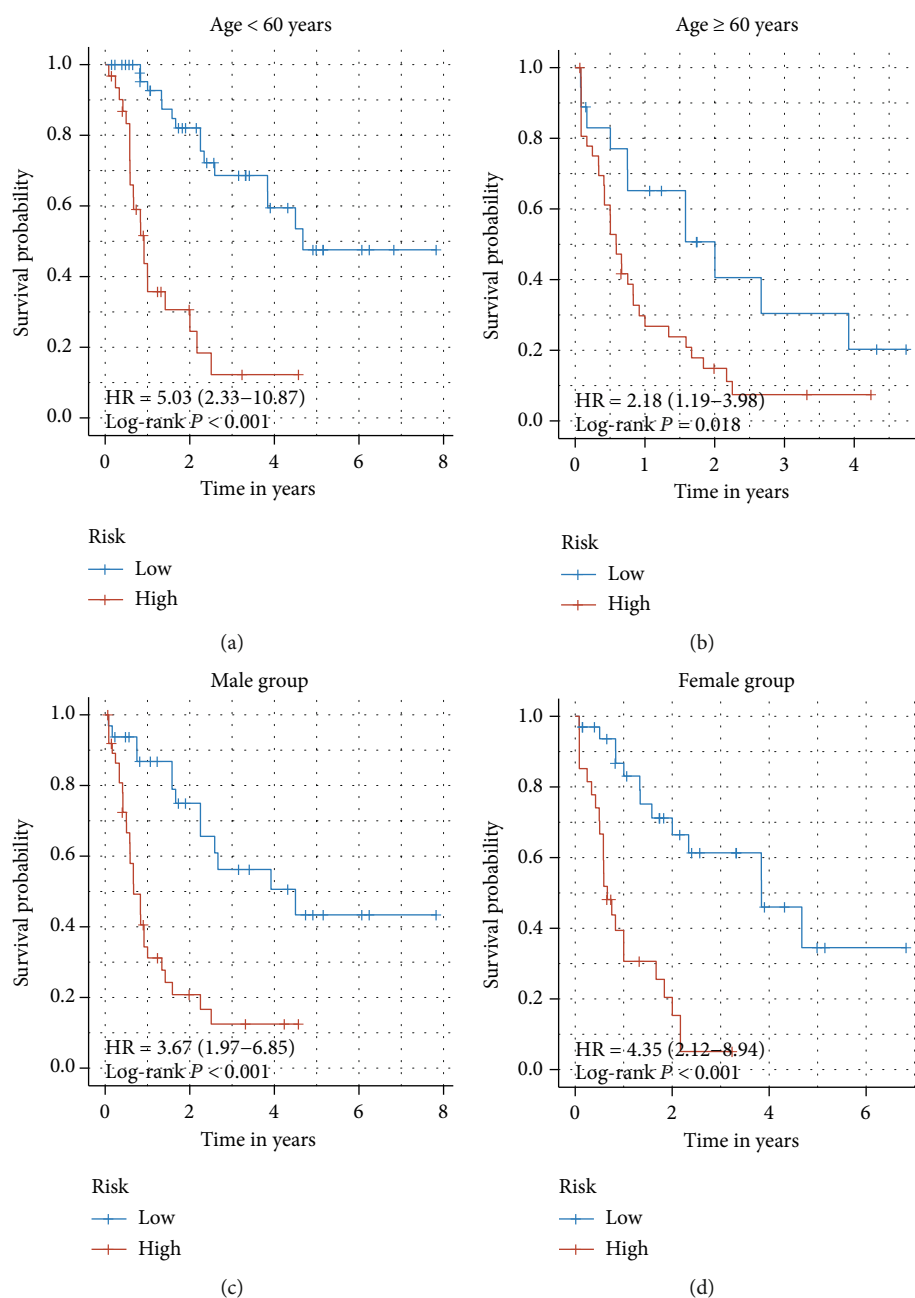
FIGURE 3: Signature with clinical variables. (a) In the age subgroup, patients under 60 years old, the high-risk score means shorter survival time, when compared with the low-risk group (HR = 5.03, 95% CI: 2.33-10.87; $p < 0.001$). (b) This conclusion also confirmed by patients more than 60 years old (HR = 2.18, 95% CI: 1.19-3.98; $p = 0.018$). (c, d) As for the sex subgroup, patients with low-risk scores always represent a better prognosis, when compared with a high-risk score, no matter which sex they are (HR = 3.67, 95% CI: 1.97-6.85; $p < 0.001$ vs. HR = 4.35, 95% CI: 2.12-8.94; $p < 0.001$, respectively).

value (Figure 2(a)). Figure 2(b) shows that patients in the high group will obtain a short survival time when compared with the low-risk group, and this difference is significant (HR = 3.97, 95% CI: 2.48-6.36; $p < 0.001$). We use ROC to evaluate the prediction accuracy of the model, and the results demonstrate the model has better prediction efficiency with AUC is 0.873 (Figure 2(c)). Internal validation results also support the above conclusions. Patients with low-risk scores mean longer living times when compared with high-risk score patients. In addition, the ROC of the

validation dataset also shows the model could predict patients' outcome accuracy with AUC equal to 0.836 (Figures 2(d)–2(f)).

3.4. Six-Gene Signature with Clinical Factors. Age and sex are both important clinical characteristics for AML patients. Here, we perform survival analysis to discover the difference between clinical subgroups. In the age subgroup, we find that in patients under 60 years old, the high-risk score means a shorter survival time, when compared with the low-risk
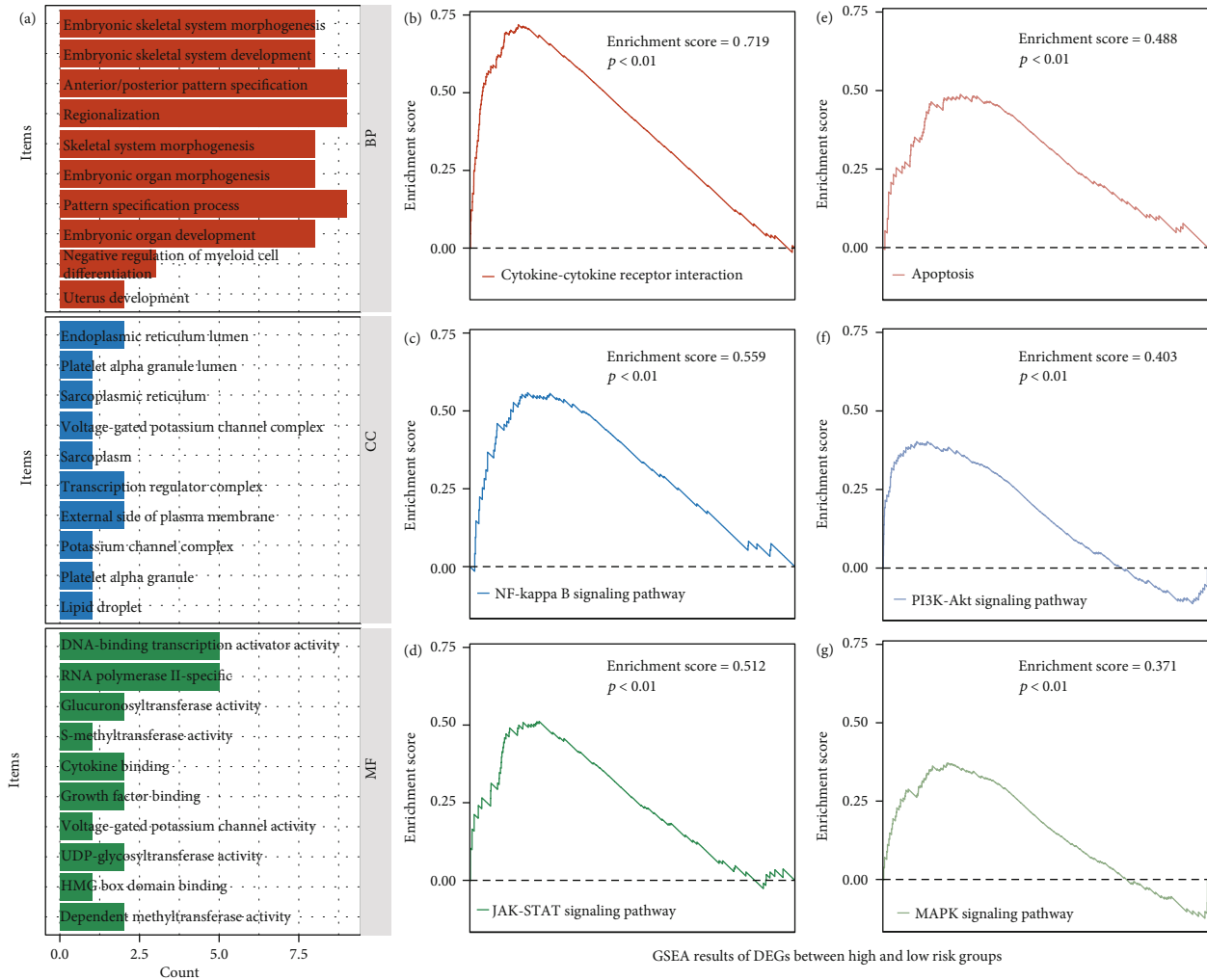
FIGURE 4: GSEA results of differential expression genes. (a) GSEA shows that this gene is enriched in embryonic skeletal system morphogenesis, endoplasmic reticulum lumen, and RNA polymerase II-specific, and the (b–g) involved pathways are cytokine-cytokine receptor interaction pathway, NF-kappa B signaling pathway, JAK-STAT signaling pathway, apoptosis, PI3K-AKT signaling pathway, and MAPK signaling pathway.

group (HR = 5.03, 95% CI: 2.33-10.87; $p < 0.001$) (Figure 3(a)). This conclusion also confirmed by patients more than 60 years old (HR = 2.18, 95% CI: 1.19-3.98; $p = 0.018$) (Figure 3(b)). As for the sex subgroup, patients with low-risk scores always represent have a better prognosis, no matter which sex they are (HR = 3.67, 95% CI: 1.97-6.85; $p < 0.001$ vs. HR = 4.35, 95% CI: 2.12-8.94; $p < 0.001$, respectively) (Figures 3(c) and 3(d)).

3.5. Different Expression Genes and Enrichment Analysis.
Patients with different risk scores have different prognoses. Identifying different expression genes between two groups is good to discover the molecular mechanism in the future. The different expression analyses demonstrate a total of 49 different expression genes between low- and high-risk groups (Supplement Table 3). After doing GSEA, we found that this gene is enriched in embryonic skeletal system morphogenesis, endoplasmic reticulum lumen, and RNA polymerase II-specific (Figure 4(a)), and the involved

pathways are cytokine-cytokine receptor interaction pathway, NF-kappa B, PI3K-AKT, and MAPK signaling pathway (Figures 4(b)–4(g)).

3.6. Nomogram Is a Useful Tool for Assistant Clinical Decision.
We have demonstrated that six gene signatures could predict patients' outcome accuracy, so we build a nomogram, based on six gene expressions, to assist clinical decisions. This model is shown in Figure 5(a), and according to six gene expressions, patients will obtain six score values and accumulate six values to become one total score and projection onto the survival axis to obtain patients' 1-year, 3-year, and 5-year survival probability, respectively. The ROC of the model is 0.761, and the C index of this nomogram is 0.774. Calibration curve analysis results suggest that survival prediction results of 1-year, 3 year, and 5-year survival probability were close to the ideal line (Figure 5(b)).

3.7. PLA2G4A Is a Hub Gene of Signature and Associated with Prognosis.
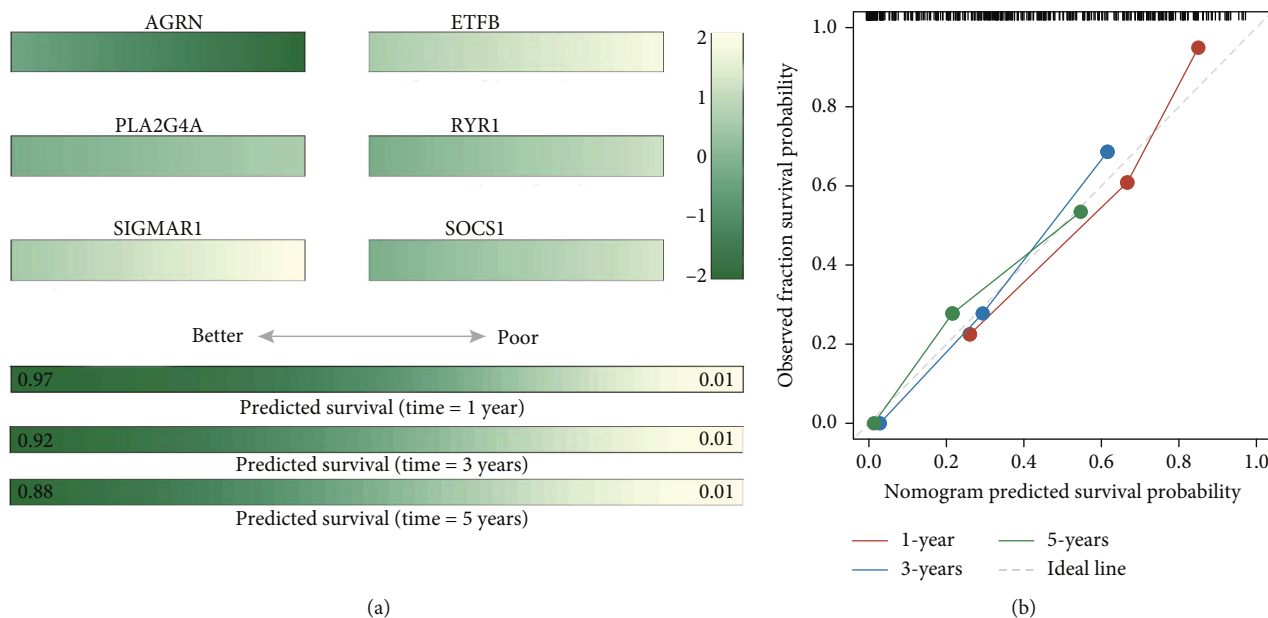We extract the relative importance of six

FIGURE 5: Nomogram for six model genes. (a) Nomogram, based on six-gene expression, to an assistant clinical decision, according to six-gene expression, patients will obtain six score values and accumulate six values to become one total score and projection onto the survival axis to obtain patients' 1-year, 3-year, and 5-year survival probability, respectively. (b) Calibration curve analysis results suggest that survival prediction probability was close to the ideal line.

model genes, and the PLA2G4A has the biggest value, so it was confirmed as the model hub gene (Figure 6(a)); to validate the potential value, we use external data set, GTEx, and GSE9476, to observe the expression difference between donor bone marrow and leukemic blasts from AML patients. The result demonstrates that PLA2G4 is a high expression in leukemic blasts and low expression in healthy hematopoietic cells ($p = 8.9e − 34$ vs. $p = 5.5e − 06$, respectively) (Figures 6(b) and 6(c)). Survival analysis also shows that high expression of this gene will lead to a poor outcome when compared with low expression patients (HR = 3.03, 95% CI: 1.97-4.67; $p < 0.001$) (Figure 6(d)).

3.8. Nobiletin Targeting PLA2G4A Provides a Third Pathway for Therapy AML. A total of 41 required herbs had been predicted, and the top 10 herbs are listed in Figure 6(e) and Supplement Table 4. Of these ten herbs, Zhiqiao has been reported to have a potential function as an anticancer. To identify which ingredients are important for these herbs, we input in into the TCMSP database. The results show that Hesperetin, Nobiletin, Naringenin, Marmin, and Beta-sitosterol are the main components of this drug (Table 2). During these components, Nobiletin has the best OB (61.67%) and highest DL (0.52); it was selected as a candidate small molecular drug to target PLA2G4A. The docking results also demonstrate the above conclusion (Figure 6(f)).

4. Discussion

Although advances have been achieved in the therapeutic options which greatly improve the overall response rate, the long-term prognosis of this disease remains dismal, espe-cially among elder patients [15]. Besides, the complex molecular and cytogenetic abnormalities make AML a kind of heterogeneous disease with differential prognosis even in the same risk group by clinical practice guidelines [16]. All these revealed that insight into the genetic landscape of AML would benefit more patients. The development of various sequencing technologies has provided more information on the mechanism of pathogenesis, chemoresistance, and more refined prognostic stratification of AML in the past decades. Recently, Mer et al. [15] proposed a unique subtype of NPM1-mutated AML with different biological and therapeutical implications based on a stem cell signature. A set of mitochondrial metabolism proteins was also identified as potential targets associated with leukemia progression by multiomics [17]. Furthermore, some gene-based signatures have been constructed to predict the prognosis of AML as in other cancers [18, 19]. All these have brought new opportunities for the treatment of AML.

The maintenance of the quiescent state of hematopoietic stem cells (HSCs) depends on a condition of anaerobic gly-colysis with low ROS generation, while compelling evidence has indicated that leukemia stem cells (LSCs), which are considered the main part of drug resistance, are more dependent on oxidative respiration with high ROS levels compa-nied by an imbalance of oxidative and antioxidant, which promote the progression of leukemia by activating the path-ways involved in the cell proliferation, survival, and invasion [20, 21]. Previous studies have proved that the redundant ROS could be a risk factor for tumorigenesis and the drug resistance role of ROS in varied leukemia modes [22, 23]. Interestingly, LSCs are more susceptible to external antioxi-dants, and ROS and lipid peroxidation by-products can trig-ger cell apoptosis, which also brings new chemotherapy

(a)

(b)

(c)

(d)

	ID	Name	FDR
HERBs	SMHB00146	Gua di	<0.001
	SMHB00190	Huo ma ren	0.007
	SMHB00102	Di fuzi	0.011
	SMHB00485	Zhi qiao	0.017
	SMHB00258	Lu lu tong	0.018
	SMHB00181	Hua ju hong	0.021
	SMHB00399	Tian gua zi	0.021
	SMHB00323	Qie geng	0.025
	SMHB00391	Tian dong	0.026
	SMHB00008	Ba qia	0.028

Source from sym map database
(HerbsTarget PLA2G4A)

Source from nobiletin and PDB database
(docking of Nobiletin with PLA2G4A protein)

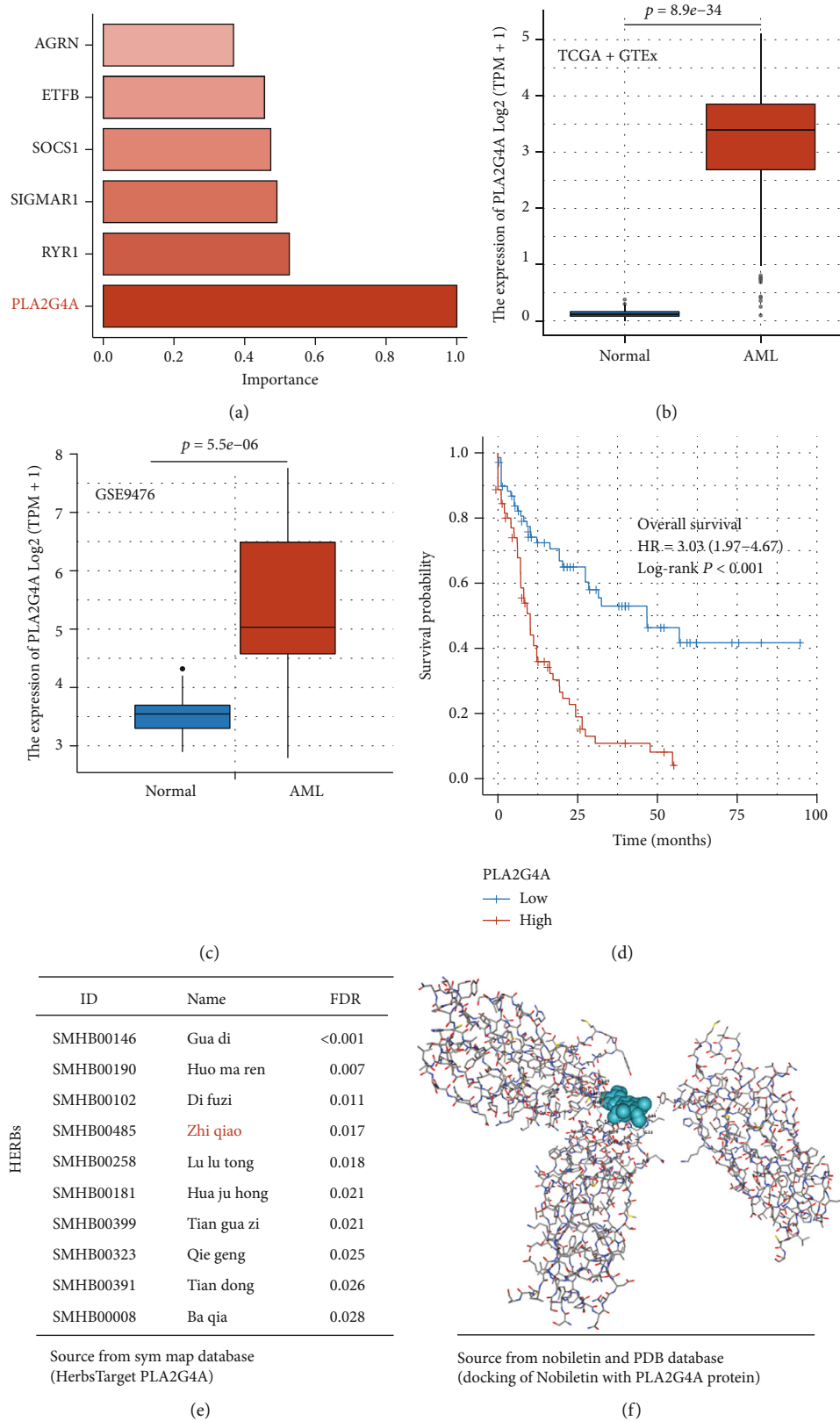(e)   (f)

FIGURE 6: PLA2G4A is a candidate therapeutic target for AML. (a) PLA2G4A has the biggest value of relative importance in six model genes, and (b, c) it is a high expression in leukemic blasts when compared with healthy hematopoietic cells ($p = 8.9e - 34$ vs. $p = 5.5e - 06$, respectively). (d) High expression of this gene will lead to a poor outcome when compared with low expression patients (HR = 3.03, 95% CI: 1.97-4.67; $p < 0.001$). (e) The top 10 herbs target PLA2G4A; (f) docking Nobiletin, the main components of this Zhiqiao, with PLA2G4A protein.

TABLE 2: Ingredients of Chinese traditional medicine Zhiqiao.

Mol ID	Molecule name	OB (%)	DL	MW	A logp	FASA-	HL
MOL002341	Hesperetin	70.31	0.27	302.3	2.28	0.33	15.78
MOL005828	Nobiletin	61.67	0.52	402.43	3.04	0.13	16.2
MOL004328	Naringenin	59.29	0.21	272.27	2.3	0.4	16.98
MOL013381	Marmin	38.23	0.31	332.43	3.11	0.31	4.68
MOL000358	Beta-sitosterol	36.91	0.75	414.79	8.08	0.23	5.36

options. These all suggest a bidirectional role of ROS in leukemia [24]. Given the above complex mechanism, we aim to construct an accurate model which can provide precise predictivity and guide stratification therapy for clinical application.

As shown in our work, we used machine learning to select six hub OS-related genes, which demonstrated robust predictive ability in AML populations. Agrin (AGRN) has been described as a multifunctional heparan sulfate proteoglycan, which can regulate angiogenesis and has a board-ranging impact on the tumor microenvironment (TME) in HCC and papillary thyroid carcinoma (PTC) [25, 26]. However, nothing has been reported on the exact function of AGRN in AML. Electron transfer flavoprotein β subunit (ETFB) can transfer the electrons to the electron transport chain (ETC) and maintain the generation of ATP. Caplan et al. [17] recently found that increased mitochondrial stress and apoptosis in AML mouse models can be induced by silencing ETFB, which suggests that ETFB could be a potential therapeutic target for AML. Moreover, the placental phospholipase A24A (PLA2G4A), sigma 1 receptor (Sig1R), and suppressors of cytokine signaling 1 (SOCS1) have been reported in varied diseases which are involved in the stress response biological procedure [27–29]. However, to the best of our knowledge, abscisic acid receptor (PYR1) has been shown as a stomatal regulation response to drought stress in plants, and no studies regarding abscisic acid receptor (PYR1) have been reported in human diseases or cancers [30]. Among these six genes, PLA2G4A has the most important in our model, which is one of the cytosolic placental phospholipases A2 (cPLA2) family and can catalyze the hydrolysis of membrane phospholipids to release arachidonic acid (AA) and lysophospholipid. It has been identified as an index in response to oxidative stress in preeclampsia and might be due to the oxidation of AA [27]. Higher expression of PLA2G4A is positively correlated with the migration and invasion of lung cancer cells and unfavorable prognosis in breast cancers [31, 32]. Previous studies also revealed that PLA2G4A expression could be an independent diagnostic and prognostic marker in patients with non-M3/NPM1 WT AML patients [33], which was also confirmed in our study. Nevertheless, whether this differential prognosis is caused by PLA2G4A through oxidative stress remains to be further investigated.

To explore the specific mechanism of OS-related genes in AML, we carried out functional annotation of DEGs between the high and low gene expression groups. The OS-related genes were enriched in the cytokine-cytokine receptor interaction pathway, NF-kappa B signaling pathway, JAK-STAT signaling pathway, apoptosis, PI3K-AKT signaling pathway, and MAPK signaling pathway as revealed by the GSEA results, which have been identified by previous studies serving as an important role in the pathogenesis and progression of AML. The PI3K/AKT pathway was proved to play important roles in regulating cell proliferation, differentiation, apoptosis, and migration in kinds of human diseases and cancers, such as diabetes, colorectal cancer, and AML [34, 35]. Some scholars also reported the PI3K/AKT pathway is associated with oxidative stress-mediated survival of melanoma and when targeting the PI3K/AKT and MAPK/ERK signaling pathway exerts an anticancer effect in leukemia cells by induction of oxidative stress and the cellular antioxidant defense mechanisms, which suggest PI3K/AKT and MAPK/ERK signaling pathway might involve in the leukemia cell apoptosis caused by oxidative stress [36, 37].

However, there are some limitations to our study. First, we did not distinguish between OS-related genes that promote leukemia proliferation and invasion with genes that mediate leukemia cell apoptosis via chemotherapy-induced OS. More datasets with pre/posttreatment information need to be included to clarify this bidirectional effect of OS. Second, our results need to be validated in a clinical trial in the further.

5. Conclusion

Our six oxidative stress-related gene signatures could predict AML patients' outcome accuracy, and this model is robust. It may become a useful tool to assist clinical decisions. In addition, we identify PLA2G4A as a potential biomarker for AML. Nobiletin, targeting PLA2G4, may provide a third pathway for therapy AML.

Data Availability

All data used in the study are available in The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases.

Conflicts of Interest

The authors declare no potential conflict of interest.

Authors' Contributions

Acknowledgments

Supplementary Materials

Supplement Table 1: prognosis-related oxidative stress genes by filter batch univariate Cox regression. Supplement Table 2: genes with a relative importance of more than 0.3 in the random forest model. Supplement Table 3: different expression genes between low- and high-risk groups. Supplement Table 4: candidate herbs targeting PLA2G4A protein. (Supplementary Materials)

References

[1] J. E. Rubnitz, B. Gibson, and F. O. Smith, "Acute myeloid leukemia," *Pediatric Clinics of North America*, vol. 55, no. 1, pp. 21–51, 2008.

[2] Y. Xie, S. M. Davies, Y. Xiang, L. L. Robison, and J. A. Ross, "Trends in leukemia incidence and survival in the United States (1973–1998)," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 97, no. 9, pp. 2229–2235, 2003.

[3] A. Pelcovits and R. Niroula, "Acute myeloid leukemia: a review," *Rhode Island Medical Journal*, vol. 103, no. 3, pp. 38–40, 2020.

[4] A. Burnett, M. Wetzler, and B. Lowenberg, "Therapeutic advances in acute myeloid leukemia," *Journal of Clinical Oncology*, vol. 29, no. 5, pp. 487–494, 2011.

[5] X.-X. Chen, Z. P. Li, J. H. Zhu, H. T. Xia, and H. Zhou, "Systematic analysis of autophagy-related signature uncovers prognostic predictor for acute myeloid leukemia," *DNA and Cell Biology*, vol. 39, no. 9, pp. 1595–1605, 2020.

[6] H. Sies and D. P. Jones, "Reactive oxygen species (ROS) as pleiotropic physiological signalling agents," *Nature Reviews Molecular Cell Biology*, vol. 21, no. 7, pp. 363–383, 2020.

[7] R. Gerschman, D. L. Gilbert, S. W. Nye, P. Dwyer, and W. O. Fenn, "Oxygen poisoning and x-irradiation: a mechanism in common," *Science*, vol. 119, no. 3097, pp. 623–626, 1954.

[8] M. Valko, D. Leibfritz, J. Moncol, M. T. D. Cronin, M. Mazur, and J. Telser, "Free radicals and antioxidants in normal physiological functions and human disease," *The International Journal of Biochemistry & Cell Biology*, vol. 39, no. 1, pp. 44–84, 2007.

[9] J. Clerkin, R. Naughton, C. Quiney, and T. G. Cotter, "Mechanisms of ROS modulated cell survival during carcinogenesis," *Cancer Letters*, vol. 266, no. 1, pp. 30–36, 2008.

[10] S. C. Bondy, "The relation of oxidative stress and hyperexcitation to neurological disease," *Proceedings of the Society for Experimental Biology and Medicine*, vol. 208, no. 4, pp. 337–345, 1995.

[11] S. V. Lakshmi, *Oxidative Stress in Cardiovascular Disease*, 2009.

[12] N. S. Brown and R. Bicknell, "Hypoxia and oxidative stress in breast cancer oxidative stress-its effects on the growth, metastatic potential and response to therapy of breast cancer," *Breast Cancer Research*, vol. 3, no. 5, pp. 323–327, 2001.

[13] G. Gupta-Elera, A. R. Garrett, R. A. Robison, and K. L. O'Neill, "The role of oxidative stress in prostate cancer," *European Journal of Cancer Prevention*, vol. 21, no. 2, pp. 155–162, 2012.

[14] U. K. Udensi and P. B. Tchounwou, "Dual effect of oxidative stress on leukemia cancer induction and treatment," *Journal of Experimental & Clinical Cancer Research*, vol. 33, no. 1, pp. 106–115, 2014.

[15] A. S. Mer, E. M. Heath, S. A. Madani Tonekaboni et al., "Biological and therapeutic implications of a unique subtype of NPM1 mutated AML," *Nature Communications*, vol. 12, no. 1, pp. 1–13, 2021.

[16] M. Bienz, M. Ludwig, B. U. Mueller et al., "Risk assessment in patients with acute myeloid leukemia and a normal karyotype," *Clinical Cancer Research*, vol. 11, no. 4, pp. 1416–1424, 2005.

[17] M. Caplan, K. J. Wittorf, K. K. Weber et al., "Multi-omics reveals mitochondrial metabolism proteins susceptible for drug discovery in AML," *Leukemia*, vol. 36, no. 5, pp. 1296–1305, 2022.

[18] W.-L. Chen, J. H. Wang, A. H. Zhao et al., "A distinct glucose metabolism signature of acute myeloid leukemia with prognostic value," *Blood, The Journal of the American Society of Hematology*, vol. 124, no. 10, pp. 1645–1654, 2014.

[19] W. Shi, D. Hu, S. Lin, and R. Zhuo, "Five-mRNA signature for the prognosis of breast cancer based on the ceRNA network," *BioMed Research International*, vol. 2020, 17 pages, 2020.

[20] U. Testa, C. Labbaye, G. Castelli, and E. Pelosi, "Oxidative stress and hypoxia in normal and leukemic stem cells," *Experimental Hematology*, vol. 44, no. 7, pp. 540–560, 2016.

[21] C. L. Bigarella, R. Liang, and S. Ghaffari, "Stem cells and the impact of ROS signaling," *Development*, vol. 141, no. 22, pp. 4206–4218, 2014.

[22] Y.-F. Chen, H. Liu, X. J. Luo et al., "The roles of reactive oxygen species (ROS) and autophagy in the survival and death of leukemia cells," *Critical Reviews in Oncology/Hematology*, vol. 112, pp. 21–30, 2017.

[23] G. Bossis, J. E. Sarry, C. Kifagi et al., "The ROS/SUMO axis contributes to the response of acute myeloid leukemia cells to chemotherapeutic drugs," *Cell Reports*, vol. 7, no. 6, pp. 1815–1823, 2014.

[24] M. E. Irwin, N. Rivera-Del Valle, and J. Chandra, "Redox control of leukemia: from molecular mechanisms to therapeutic opportunities," *Antioxidants & Redox Signaling*, vol. 18, no. 11, pp. 1349–1383, 2013.

[25] C.-C. Wu, J. D. Lin, J. T. Chen et al., "Integrated analysis of fine-needle-aspiration cystic fluid proteome, cancer cell secretome, and public transcriptome datasets for papillary thyroid cancer biomarker discovery," *Oncotarget*, vol. 9, no. 15, pp. 12079–12100, 2018.

[26] S. Chakraborty, M. Lakshmanan, H. L. F. Swa et al., "An onco-genic role of Agrin in regulating focal adhesion integrity in hepatocellular carcinoma," *Nature Communications*, vol. 6, no. 1, pp. 1–16, 2015.

[27] M. Brien, J. Larose, K. Greffard, P. Julien, and J. F. Bilodeau, "Increased placental phospholipase A_2 gene expression and free F_2-isoprostane levels in response to oxidative stress in pre-eclampsia," *Placenta*, vol. 55, pp. 54–62, 2017.

[28] J. Wang, A. Saul, P. Roon, and S. B. Smith, "Activation of the molecular chaperone, sigma 1 receptor, preserves cone function in a murine model of inherited retinal degeneration," *Proceedings of the National Academy of Sciences*, vol. 113, no. 26, pp. E3764–E3772, 2016.

[29] X. Liu, M. G. Mameza, Y. S. Lee et al., "Suppressors of cytokine-signaling proteins induce insulin resistance in the retina and promote survival of retinal cells," *Diabetes*, vol. 57, no. 6, pp. 1651–1658, 2008.

[30] Z. Jin, S. Xue, Y. Luo et al., "Hydrogen sulfide interacting with abscisic acid in stomatal regulation responses to drought stress in Arabidopsis," *Plant Physiology and Biochemistry*, vol. 62, pp. 41–46, 2013.

[31] W. Zhang, X. Wang, L. Zhang et al., "Inhibition of PLA2G4A reduces the expression of lung cancer-related cytokines," *DNA and Cell Biology*, vol. 37, no. 12, pp. 1076–1081, 2018.

[32] S. Mishra, M. Charan, R. K. Shukla et al., "cPLA2 blockade attenuates S100A7-mediated breast tumorigenicity by inhibiting the immunosuppressive tumor microenvironment," *Journal of Experimental & Clinical Cancer Research*, vol. 41, no. 1, pp. 1–19, 2022.

[33] H. Bai, M. Zhou, M. Zeng, and L. Han, "PLA2G4A is a potential biomarker predicting shorter overall survival in patients with non-M3/NPM1 wildtype acute myeloid leukemia," *DNA and Cell Biology*, vol. 39, no. 4, pp. 700–708, 2020.

[34] M. Y. Follo, L. Manzoli, A. Poli, J. A. McCubrey, and L. Cocco, "PLC and PI3K/Akt/mTOR signalling in disease and cancer," *Advances in biological regulation*, vol. 57, pp. 10–16, 2015.

[35] S. A. Danielsen, P. W. Eide, A. Nesbakken, T. Guren, E. Leithe, and R. A. Lothe, "Portrait of the PI3K/AKT pathway in colorectal cancer," *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, vol. 1855, no. 1, pp. 104–121, 2015.

[36] H. G. Hambright, P. Meng, A. P. Kumar, and R. Ghosh, "Inhibition of PI3K/AKT/mTOR axis disrupts oxidative stress-mediated survival of melanoma cells," *Oncotarget*, vol. 6, no. 9, pp. 7195–7208, 2015.

[37] E. Jasek-Gajda, H. Jurkowska, M. Jasińska, and G. J. Lis, "Targeting the MAPK/ERK and PI3K/AKT signaling pathways affects NRF2, Trx and GSH antioxidant systems in leukemia cells," *Antioxidants*, vol. 9, no. 7, p. 633, 2020.