



# A Prognostic Model for Acute Myeloid Leukemia Based on IL-2/STAT5 Pathway-Related Genes

Yigang Tang<sup>†</sup>, Shujun Xiao<sup>†</sup>, Zhengyuan Wang, Ying Liang, Yangfei Xing, Jiale Wu and Min Lu<sup>\*</sup>

Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

## OPEN ACCESS

### Edited by:

João Pessoa,  
University of Coimbra, Portugal

### Reviewed by:

Priyanka Gupta,  
University of Alabama at Birmingham,  
United States  
Sarah E. Church,  
NanoString Technologies,  
United States

### \*Correspondence:

Min Lu  
min.lu@shsmu.edu.cn

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Molecular and Cellular Oncology,  
a section of the journal  
Frontiers in Oncology

Received: 29 September 2021

Accepted: 03 January 2022

Published: 02 February 2022

### Citation:

Tang Y, Xiao S, Wang Z, Liang Y,  
Xing Y, Wu J and Lu M (2022)  
A Prognostic Model for Acute Myeloid  
Leukemia Based on IL-2/STAT5  
Pathway-Related Genes.  
Front. Oncol. 12:785899.  
doi: 10.3389/fonc.2022.785899

Accurate prognostic stratification of patients can provide guidance for personalized therapy. Many prognostic models for acute myeloid leukemia (AML) have been reported, but most have considerable inaccuracies due to contained variables with insufficient capacity of predicting survival and lack of adequate verification. Here, 235 genes strongly related to survival in AML were systematically identified through univariate Cox regression analysis of eight independent AML datasets. Pathway enrichment analysis of these 235 genes revealed that the IL-2/STAT5 signaling pathway was the most highly enriched. Through Cox proportional-hazards regression model and stepwise algorithm, we constructed a six-gene STAT5-associated signature based on the most robustly survival-related genes related to the IL-2/STAT5 signaling pathway. Good prognostic performance was observed in the training cohort (GSE37642-GPL96), and the signature was validated in seven other validation cohorts. As an independent prognostic factor, the STAT5-associated signature was positively correlated with patient age and ELN2017 risk levels. An integrated score based on these three prognostic factors had higher prognostic accuracy than the ELN2017 risk category. Characterization of immune cell infiltration indicated that impaired B-cell adaptive immunity, immunosuppressive effects, serious infection, and weakened anti-inflammatory function tended to accompany high-risk patients. Analysis of in-house clinical samples revealed that the STAT5-associated signature risk scores of AML patients were significantly higher than those of healthy people. Five chemotherapeutic drugs that were effective in these high-risk patients were screened *in silico*. Among the five drugs, MS.275, a known HDAC inhibitor, selectively suppressed the proliferation of cancer cells with high STAT5 phosphorylation levels *in vitro*. Taken together, the data indicate that the STAT5-associated signature is a reliable prognostic model that can be used to optimize prognostic stratification and guide personalized AML treatments.

**Keywords:** acute myeloid leukemia, IL-2/STAT5 pathway, prognostic model, immune infiltration, drug screening, personalized therapy

## INTRODUCTION

Acute myeloid leukemia (AML) is the most common form of acute leukemia in adults, characterized by abnormal growth and differentiation of hematopoietic stem cells (HSCs) (1). The clinical outcomes of AML patients remain unsatisfactory, with 5-year overall survival rates of less than 50%, dropping to less than 20% for patients older than 60 years (2). The poor prognosis of AML may be attributed to the heterogeneity of therapeutic responses among patients (3) and conventional clinical therapies that have changed little over the past three decades (4). As a consequence, there is an urgent need to better stratify patients facilitating the development of personalized treatments for different patients with AML.

Signal transducer and activator of transcription 5 (STAT5), with its two isoforms STAT5A and STAT5B (5), is a key component of the janus tyrosine kinase (JAK)-signal transducer and activator of transcription (STAT) pathway (6). As a transcription factor, STAT5 can be phosphorylated upon interleukin-2 (IL-2) binding to its cognate receptor, followed by the activation of its downstream targets (7). Abnormal activation of STAT5 *via* phosphorylation frequently occurs in blast cells of patients with AML (8), where it is important for the proliferation of leukemic cells (9). High STAT5 levels are relevant to drug resistance and can desensitize BCR-ABL1<sup>+</sup> leukemia cells to tyrosine kinase inhibitors (10). Additionally, phosphorylated STAT5 can suppress antitumor immunity (11) and is also engaged in the pathogenesis of chronic osteomyelitis *via* immune dysregulation (12).

Transcriptomic variables have higher predictive accuracy than clinical or genetic variables in myelodysplastic syndrome (13), and similar trends were recently observed in AML (14). However, the current widely used risk-stratification system (European Leukemia Net (ELN)2017 risk category) recommended by the National Comprehensive Cancer Network (NCCN) AML guideline (15) was constructed based on genetic variables, without considering transcriptomic changes (14). Recently, increasing numbers of prognostic models for AML based on transcriptomic data were reported, encompassing distinct biological processes such as immunity (16–18), autophagy (19), etc. However, there was no comprehensive analysis of strongly survival-related genes in AML prior to this study, which hampered the development of more accurate prognostic models based on transcriptomic data.

## MATERIALS AND METHODS

### Retrieval of AML Datasets

We systematically retrieved AML datasets in the Gene Expression Omnibus (GEO) database (**Supplementary Table 1**). All datasets with more than 100 samples and available survival information were collected. The dataset with the most complete data and the largest sample size was selected when datasets overlapped. Eventually, five GEO datasets including GSE106291, GSE12417-

GPL96, GSE37642-GPL96, GSE37642-GPL570, and GSE71014 were screened out for the present study (20–23). In addition, an AML cohort from The Cancer Genome Atlas (TCGA) (24), an AML cohort from Therapeutically Applicable Research to Generate Effective Treatments (TARGET), and an AML cohort from a clinical study at Oregon Health & Science University (OHSU) (3) also met the inclusion criteria and were included in this study.

Total RNA samples isolated from bone marrow mononuclear cells were used for probing gene expression levels in cohorts GSE37642-GPL96, GSE37642-GPL570, GSE71014, and TCGA. Total RNA-isolated samples from bone marrow (BM) mononuclear cells and peripheral blood (PB) mononuclear cells were used for detecting gene expression levels in cohorts GSE106291 (details unavailable), GSE12417-GPL96 (161 BM and 2 PB), OHSU (251 BM and 160 PB), and TARGET (details unavailable).

Detection of gene expression levels in different cohorts was performed on different platforms: ~20,000 encoding genes detected using Illumina HiSeq 1500 in GSE106291; ~12,000 encoding genes detected using Affymetrix Human Genome U133A Array in GSE12417-GPL96; ~12,000 encoding genes detected using Affymetrix Human Genome U133A Array in GSE37642-GPL96; ~18,000 encoding genes detected using Affymetrix Human Genome U133 Plus 2.0 Array in GSE37642-GPL570; ~20,000 encoding genes detected using HumanHT-12 V4.0 expression beadchip in GSE71014; ~20,000 encoding genes detected using Illumina HiSeq 2500 in OHSU; ~20,000 encoding genes detected using Illumina HiSeq 2000 in TARGET; and ~20,000 encoding genes detected using Illumina HiSeq 2000 in TCGA.

Processed gene expression data with respective normalization method were downloaded for bioinformatical analysis in this study. All gene expression variables were scaled to a mean value of 0 and variance equal to 1 (Z-score) in GSE10621. Normalization was performed using the variance stabilizing normalization (VSN) algorithm, and probe set expression values were calculated by the median polish method in GSE12417-GPL96. Normalization was performed using the Robust Multichip Average (RMA) method in GSE37642-GPL96 and GSE37642-GPL570. Expression values were processed with log<sub>2</sub> transformation and quantile normalization in GSE71014. Normalization was performed using the conditional quantile normalization procedure in OHSU. Fragments per kilobase of exon model per million mapped fragments (FPKM) values of genes were log<sub>2</sub>(FPKM+1) transformed in TARGET. RNA-Seq by Expectation-Maximization (RSEM) normalized counts (norm\_count) of genes were log<sub>2</sub>(norm\_count +1) transformed in TCGA.

Normalized transcriptome data and clinical information were acquired from three different databases: GEO datasets from GEO database (<http://www.ncbi.nlm.nih.gov/geo/>), TCGA and TARGET datasets from the UCSC Xena database (<http://xena.ucsc.edu/>), as well as the OHSU dataset from cBioPortal (<https://www.cbioportal.org/>) (25, 26). Clinical variables of the eight cohorts were summarized in each dataset (**Supplementary Table 2**).

Clinical data of GSE71014 only contained survival information. Samples without survival information or transcriptome data were excluded in each dataset. After exclusion, sample size in each cohort was as follows: GSE106291 ( $n = 250$ ), GSE12417-GPL96 ( $n = 163$ ), GSE37642-GPL96 ( $n = 417$ ), GSE37642-GPL570 ( $n = 136$ ), GSE71014 ( $n = 104$ ), OHSU ( $n = 411$ ), TARGET ( $n = 156$ ), and TCGA ( $n = 151$ ).

## Screening of Robustly Survival-Related Genes

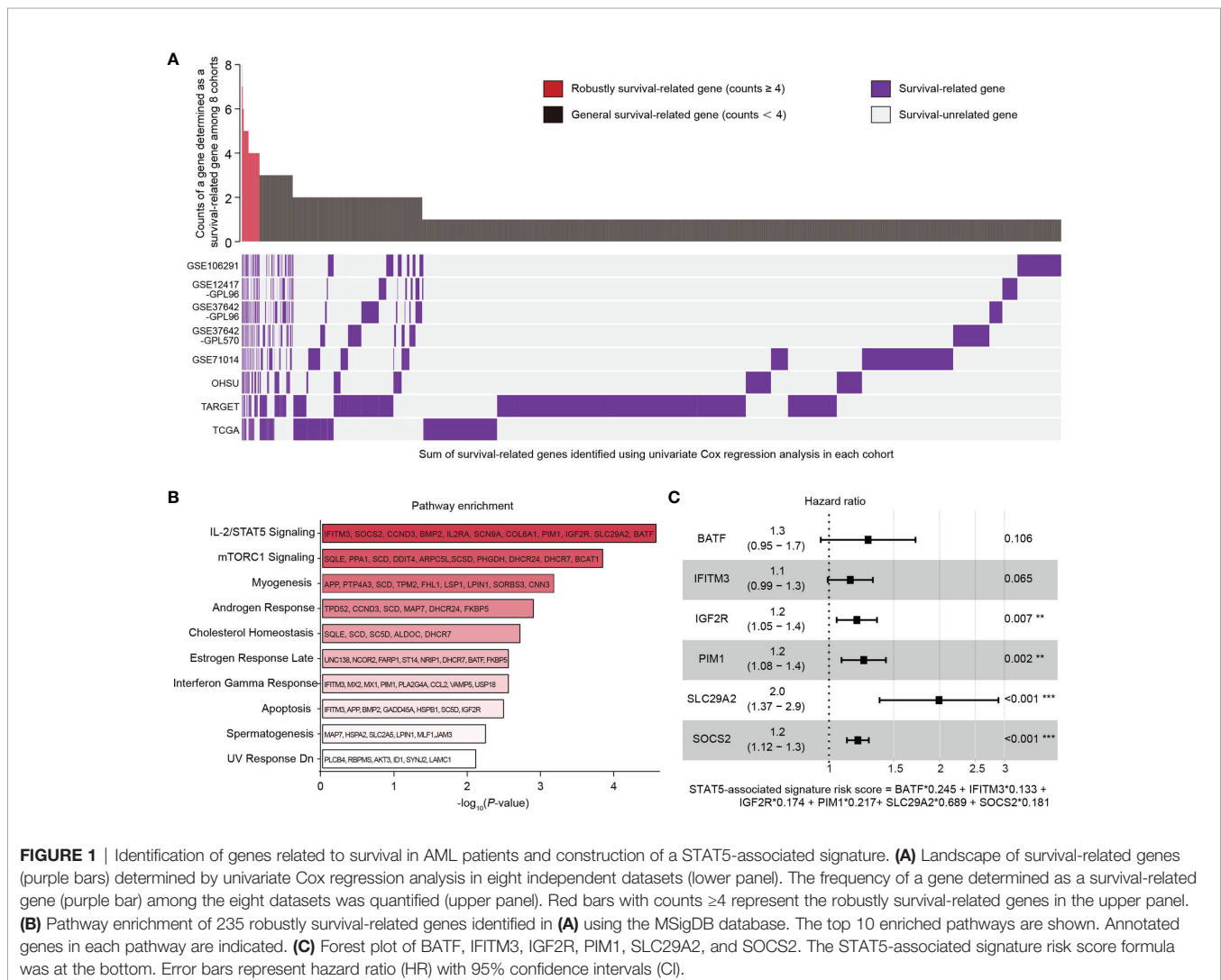
The univariate Cox regression analysis was performed individually in eight independent AML datasets (GSE106291, GSE12417-GPL96, GSE37642-GPL96, GSE37642-GPL570, GSE71014, OHSU, TARGET, TCGA). Survival-related genes ( $HR > 1$ ,  $p < 0.05$ ) in each dataset were screened out (Figure 1A, purple bars). A gene which was identified as a survival-related gene in at least four datasets was defined as a robustly survival-related gene in this study. Eventually, a total of 235 robustly survival-related genes were identified (red bars in upper panel, Figure 1A; Supplementary Table 3).

## Pathway Enrichment Analysis

A total of 235 identified robustly survival-related genes (Supplementary Table 3) were subjected to pathway enrichment analysis using the Molecular Signatures Database (MSigDB) on Enrichr (<https://maayanlab.cloud/Enrichr/>). Briefly, we entered the 235 gene symbols on each row in the text-box on the Enrichr (<https://maayanlab.cloud/Enrichr/>) and submitted these gene symbols. We then clicked “Pathways” module in the navigation (at the top). Detailed results including enriched pathways and  $p$ -values could be found after clicking the icon “MSigDB Hallmark 2020”.

## Expression Distribution of STAT5A and STAT5B Among 16 Different Organs

Online website The Human Protein Atlas (<https://www.proteinatlas.org/>) contains information on genome-wide RNA expression profiles of human protein-coding genes in 69 human cell lines. These 69 cell lines are derived from 16 different organs including: brain, liver and gallbladder, gastrointestinal tract, pancreas, male reproductive system, kidney and urinary



**FIGURE 1** | Identification of genes related to survival in AML patients and construction of a STAT5-associated signature. **(A)** Landscape of survival-related genes (purple bars) determined by univariate Cox regression analysis in eight independent datasets (lower panel). The frequency of a gene determined as a survival-related gene (purple bar) among the eight datasets was quantified (upper panel). Red bars with counts  $\geq 4$  represent the robustly survival-related genes in the upper panel. **(B)** Pathway enrichment of 235 robustly survival-related genes identified in **(A)** using the MSigDB database. The top 10 enriched pathways are shown. Annotated genes in each pathway are indicated. **(C)** Forest plot of BATF, IFITM3, IGF2R, PIM1, SLC29A2, and SOCS2. The STAT5-associated signature risk score formula was at the bottom. Error bars represent hazard ratio (HR) with 95% confidence intervals (CI).

bladder, skin, eye, proximal digestive tract, lung, female reproductive system, endothelial, muscle, mesenchymal, lymphoid, and myeloid. Downloaded gene expression levels of STAT5A (<https://www.proteinatlas.org/ENSG00000126561-STAT5A/cell+line>) and STAT5B (<https://www.proteinatlas.org/ENSG00000173757-STAT5B/cell+line>) in the 69 cell lines were used for investigating the expression distribution of STAT5A and STAT5B among the 16 different organs.

### Protein-Protein Interaction Network

The search tool for the retrieval of interacting genes (STRING) database (<http://string.embl.de/>) (27) was used to visualize the associations among robustly survival-related genes related to the IL-2/STAT5 pathway. Briefly, we selected function module “Multiple proteins” in the left navigation and enter protein names including STAT5A, STAT5B, and 11 robustly survival-related genes (IFITM3, SOCS2, CCND3, BMP2, IL2RA, SCN9A, COL6A1, PIM1, IGF2R, SLC29A2, and BATF) in the search box. We selected “*Homo sapiens*” in the pull-down list of “Organism” and clicked the icon “SEARCH” under the search box. We clicked the icon “CONTINUE” in the pop-up interface and waited for a moment. The results could be found in the next pop-up interface. We downloaded the scalable vector graphic from the “Exports” module.

### Construction and Validation of a STAT5-Associated Signature

Eleven robustly survival-related genes related to the IL-2/STAT5 pathway (IFITM3, SOCS2, CCND3, BMP2, IL2RA, SCN9A, COL6A1, PIM1, IGF2R, SLC29A2, and BATF; **Figure 1B**) were used for constructing a STAT5-associated signature. In the training cohort (GSE37642-GPL96), the Cox proportional-hazards model (28) was employed to estimate the optimal weighting coefficients of these 11 robustly survival-related genes with the function *coxph* in the package “survival” (29) on the basis of maximizing the partial likelihood techniques (30, 31). For building the best performing regression model, 6 genes (BATF, IFITM3, IGF2R, PIM1, SLC29A2, SOCS2; **Figure 1C**) out of the 11 robustly survival-related genes were selected for constructing the final Cox regression model with the function *step* in R language based on the stepwise algorithm (32). The STAT5-associated signature risk score was calculated according to the sum of the coefficients multiplied by the gene expression level of each selected gene. Patients were separated into low- and high-risk groups according to the median STAT5-associated signature risk score in each cohort. The prognostic performance of this model was assessed using Kaplan-Meier analysis. The specificity of this model was evaluated using curves with area under the receiver operating characteristic (ROC) curve (AUC) values. The prognostic independence of this model was confirmed by univariate and multivariate Cox analysis.

### Improvement of the European Leukemia Net 2017 Risk Stratification System

An integrated risk model was constructed based on STAT5-associated signature, age of patients, and ELN2017 risk category

using the Cox proportional-hazards model (28). This risk model was visualized by the nomogram produced by the R package “rms” (33). The predictive accuracy of this integrated risk model was assessed using calibration curves produced by the R package “rms” (33).

### Estimation of Immune Infiltration

Transcriptome data were used to estimate the composition of tumor-infiltrating immune cells based on the deconvolution algorithm of the Cell type Identification by Estimating Relative Subsets of RNA Transcripts (CIBERSORT) (34). The relative fractions of the 22 immune cell types in each sample were then determined using the function *CIBERSORT* in R language (34). An empirical *p*-value for the deconvolution was produced for each sample through Monte Carlo sampling (34). Only outputs with *p* < 0.05 were used for further analysis.

The correlations between STAT5-associated signature risk scores and fractions of tumor-infiltrating immune cells were further investigated using Spearman correlation analysis. Proportions of immune cells and stromal cells were estimated based on the immune score and stromal score, respectively. These two tumor microenvironment scores were calculated using the R package “estimate” (35).

### In-House Human Samples

Total RNA of peripheral blood mononuclear cells (from 6 healthy donors and 28 AML patients) and peripheral blood mononuclear cells (from 6 healthy donors and 6 AML patients) were collected at the Department of Hematology, the First Affiliated Hospital of Jinan University. These RNA samples were collected from June, 2020 to November, 2020 and were stored at  $-80^{\circ}\text{C}$ . Peripheral blood mononuclear cells were collected from June, 2020 to November, 2020 and were stored at liquid nitrogen. Written informed consent was obtained from all patients. The Ethics Committee of the First Affiliated Hospital of Jinan University approved the study (No. KY-2020-022 in 2020).

### Real-Time Quantitative PCR Assay

The real-time quantitative PCR (RT-qPCR) assay was performed exactly as reported previously (36). Briefly, total RNA was isolated using a Total RNA Purification Kit (B518651, Sangon Biotech, Shanghai, China) following the manufacturer’s protocol. Total RNA at 1  $\mu\text{g}$  was then reverse transcribed using the HiScript<sup>®</sup> II QRT SuperMix for qPCR (+ gDNA wiper) (R223-01, Vazyme Biotech, Nanjing, China) following the manufacturer’s protocol. PCR was performed in triplicate using ChamQ<sup>™</sup> SYBR<sup>®</sup> qPCR Master Mix (Low ROX Premixed) (Q331-02/03, Vazyme Biotech) and a ViiA<sup>™</sup> 7 Real-Time PCR System (Applied Biosystems, Waltham, MA, USA) under the following conditions: 10 min at  $95^{\circ}\text{C}$ , followed by 40 cycles of  $95^{\circ}\text{C}$  for 15 s and  $60^{\circ}\text{C}$  for 45 s.  $\beta$ -Actin was used as the housekeeping control. Primers used in this assay are as follows: BATF-forward: TATTGCCGCCGAGAAGAGC, BATF-reverse: GCTTGATCTCCTTGCGTAGAG; IFITM3-forward: AGGGACAGGAAGATGGTTGG, IFITM3-reverse: TGGGATGACGATGAGCAGAA; IGF2R-forward: CTGCCGCT



ATGAAATTGAGTGG, IGF2R-reverse: CGCCGCTCAGAG AACAAGTT; PIM1-forward: GAGAAGGACCGGATT TCCGAC, PIM1-reverse: CAGTCCAGGAGCCTAATGACG; SLC29A2-forward: TCAGTGCAGTCTACAGGG, SLC29A2-reverse: GGGCTGATAAAGTACCCAGG; SOCS2-forward: CAGATGTGCAAGGATAAGCGG, SOCS2-reverse: GCGGTT TGGTCAGATAAAGGTG;  $\beta$ -actin-forward: ACTTAGTTGCG TTACACCCCTTCT;  $\beta$ -actin-reverse: GACTGCTGTCAC CTTACACCGT. Relative gene expression was normalized to  $\beta$ -actin and calculated by the formula: relative target gene expression =  $2^{-\Delta CT}$  ( $\Delta CT = CT_{\text{target gene}} - CT_{\beta\text{-actin}}$ ). RT-qPCR cycles were uploaded as a supplementary material (**Supplementary Table 4**).

### **In Silico Screening of Chemotherapy Drugs for the Treatment of High-Risk Patients**

Clinical drug responses could be predicted using baseline gene expression levels (37). In brief, a ridge regression model was fitted for baseline gene expression levels in the 700 cell lines against the *in vitro* 138 drug half-maximal inhibitory concentration ( $IC_{50}$ ) estimates, and this model was then applied to the baseline tumor expression data from the clinical trial, to yield drug sensitivity estimates (37). In the present study, the ridge regression model was used to estimate the  $IC_{50}$  of 138 chemotherapeutic agents for each AML patient based on transcriptomic data, followed by 10-fold cross-validation implemented using the R package “pRRophetic” (38). A chemotherapy drug with significantly lower  $IC_{50}$  in the high-risk group was determined as a targeted drug for high-risk patients in each cohort. The frequency with which a drug was identified as a targeted drug for high-risk patients in eight cohorts was quantified. Five drugs with the highest frequencies, including bexarotene, bortezomib, erlotinib, rapamycin, and MS.275 were screened as *in silico* hits.

### **Immunoblotting Assay**

A specific antibody against pSTAT5 (1:1,000 dilution; #9395, Cell Signaling Technology, Danvers, MA, USA) was used to determine the levels of phosphorylated STAT5 in cancer cell lines and peripheral blood mononuclear cells. The immunoblotting assay was performed exactly as reported previously (36). The protein bands were quantified densitometrically using ImageJ software. The full uncropped immunoblotting images were uploaded as a supplementary material (**Supplementary Figure 7**).

### **Cell Viability Assay**

Cell viability was assessed using the Cell Counting Kit-8 (CCK8; C0005, Targetmol) following the manufacturer’s instructions. Briefly, cells were seeded in triplicate into 96-well plates at a density of 1,500–3,000 cells/well in 100  $\mu$ l of medium. After treatment with the indicated chemotherapy drugs for 3 days, dye solution was added and the plates were incubated at 37°C for 3–4 h before the absorbance at 450 nm ( $A_{450}$ ) was measured. Cell viability was calculated by the formula: cell viability (%) = [(As-

Ab)/(Ac-Ab)]  $\times$  100 where As is the absorbance of the experimental well (absorbance of cells, medium, CCK8, and wells of the test compound), Ab is the blank well absorbance (absorbance of wells containing medium and CCK8), and Ac is the control well absorbance (absorbance of wells containing cells, medium and CCK8).

### **Statistical Analysis**

Statistical analyses were performed using R software (version 4.0.5; R foundation for statistical computing, Vienna, Austria) and SPSS version 23.0 (IBM Corp., Armonk, NY, USA). Uni- and multivariate Cox regression analyses were conducted using the “survival” R package (29). Selected 6 genes constituting the STAT5-associated signature were separated into low- and high-expression groups based on the optimal cutoff determined using the “survminer” package (39) with the minprop variable (the minimal proportion of the observations/group) set to 20% (40). Kaplan-Meier analysis was carried out using the packages “survminer” (39) and “survival” (29), and the significance of survival differences was determined using the log-rank test. Time-dependent and time-independent receiver operating characteristic (ROC) curves were generated using the packages “timeROC” (41) and “survivalROC” (42), respectively. Nomograms and calibration curves were generated using the “rms” package (33). The statistical significance of differences between mean values of two groups was assessed using unpaired two-tailed Student’s *t*-test. Chi-squared analysis was used to evaluate the relationship between risk categories and clinicopathological parameters. The *r*- and *p*-values were determined by Spearman correlation analyses. The “pRRophetic” R package (38) was used to predict the responses to chemotherapy. The  $IC_{50}$  values of different chemotherapeutics in six cancer cell lines were estimated using the online tool  $IC_{50}$  calculator (<https://www.aatbio.com/tools/ic50-calculator/>). Differences with *p* < 0.05 were considered statistically significant.

## **RESULTS**

### **Dataset Selection and Clinical Variables of Selective Eight Datasets**

Eight publicly available datasets with more than 100 samples and available survival information were selected (**Supplementary Table 1**). GSE37642-GPL96 (*n* = 417) with the largest sample size was used as a training cohort, and seven datasets including GSE106291 (*n* = 250), GSE12417-GPL96 (*n* = 163), GSE37642-GPL570 (*n* = 136), GSE71014 (*n* = 104), OHSU (*n* = 411), TARGET (*n* = 156), and TCGA (*n* = 151) were set aside as validation cohorts. Clinical variables of the eight cohorts were summarized in each dataset (**Supplementary Table 2**). Some clinically relevant features of the eight cohorts were observed. All of the patients in OHSU, for example, were under age 60 while about half of the patients in the seven other cohorts were under age 60. The ratio of men to women was close to 1:1 in the eight cohorts. Each cohort consisted mainly of M1, M2, M4, and M5 patients.

## Identification of Robustly Survival-Related Genes in AML

Univariate Cox regression analysis was performed individually in eight independent AML cohorts (GSE106291, GSE12417-GPL96, GSE37642-GPL96, GSE37642-GPL570, GSE71014, OHSU, TARGET, and TCGA; **Figure 1A**) and survival-related genes ( $HR > 1$ ,  $p < 0.05$ ) in each cohort were identified (**Figure 1A**; lower panel, purple bars). A gene which was determined to be survival related in at least four cohorts was defined as a robustly survival-related gene in the present study (**Figure 1A**; upper panel, red bars). A total of 235 identified robustly survival-related genes (**Figure 1A**; upper panel, red bars; **Supplementary Table 3**) were then subjected to pathway enrichment analysis using the Molecular Signatures Database (MSigDB), and the IL-2/STAT5 pathway was the most highly enriched item with 11 annotated genes including IFITM3, SOCS2, CCND3, BMP2, IL2RA, SCN9A, COL6A1, PIM1, IGF2R, SLC29A2, and BATF (**Figure 1B**). In addition, the gene expression levels of STAT5A and STAT5B were found to be organ specific and were significantly higher in cancer cells derived from lymphoid and myeloid organs (**Supplementary Figure 1A**). However, there was no difference in the expression levels of the two STAT5 genes between cancer cells derived from lymphoid and myeloid (**Supplementary Figure 1A**). The abnormal STAT5 expression pattern suggested that genes involved in STAT5-associated pathways might be alternative prognostic biomarkers for hematological malignancies.

## Construction of a STAT5-Associated Signature

Eleven identified robustly survival-related genes annotated in the IL-2/STAT5 pathway (**Supplementary Figure S1B**, text in white) were subjected to construct a STAT5-associated signature using Cox proportional-hazards regression model and stepwise algorithm in the training cohort (GSE37642-GPL96) (Details seen in the method section; **Figure 1C**). The STAT5-associated signature was described using the formula risk score =  $\text{Exp}_{\text{BATF}} * 0.245 + \text{Exp}_{\text{IFITM3}} * 0.133 + \text{Exp}_{\text{IGF2R}} * 0.174 + \text{Exp}_{\text{PIM1}} * 0.217 + \text{Exp}_{\text{SLC29A2}} * 0.689 + \text{Exp}_{\text{SOCS2}} * 0.181$  (**Figure 1C**). The STAT5-associated signature risk score of each

AML patient was then calculated and used to stratify patients into low- and high-risk groups according to the median risk score in each cohort.

The prognostic performance of the selected 6 genes that constitute this model was assessed using Kaplan-Meier analysis after classification into low- and high-expression groups in the training cohort (GSE37642-GPL96) (**Supplementary Figures 1C–H**). AML patients with high expression of any one of the six genes had significantly shorter overall survival (**Supplementary Figures 1C–H**). Among the six genes, SLC29A2 with the biggest weighting coefficient in the STAT5-associated signature might be the most significant prognostic marker to stratify AML patients (**Figure 1C**).

## Performance of the STAT5-Associated Signature

The prognostic performance of the signature was next assessed in the training cohort (GSE37642-GPL96), as well as the seven validation cohorts GSE106291, GSE12417-GPL96, GSE37642-GPL570, GSE71014, OHSU, TARGET, and TCGA. The relationship between STAT5-associated signature risk scores and survival status of patients in the cohorts is shown in **Supplementary Figures 2A–H**. Kaplan-Meier analysis indicated that patients in the high-risk group had significantly shorter overall survival in the training cohort (GSE37642-GPL96,  $p = 7.783e-10$ , **Table 1**; **Supplementary Figure 3A**). In line with the performance in the training cohort (GSE37642-GPL96), we found that the STAT5-associated signature also worked well in external validation cohorts, where patients in the high-risk group had shorter overall survival (GSE106291,  $p = 3.654e-04$ , **Table 1** and **Supplementary Figure 3B**; GSE12417-GPL96,  $p = 1.282e-02$ , **Table 1** and **Supplementary Figure 3C**; GSE37642-GPL570,  $p = 7.086e-04$ , **Table 1** and **Supplementary Figure 3D**; GSE71014,  $p = 2.618e-03$ , **Table 1** and **Supplementary Figure 3E**; OHSU,  $p = 1.478e-05$ , **Table 1** and **Supplementary Figure 3F**; TARGET,  $p = 8.45e-04$ , **Table 1** and **Supplementary Figure 3G** and TCGA,  $p = 8.631e-05$ , **Table 1** and **Supplementary Figure 3H**). The time-independent AUC value of this model reached 0.705 in the training cohort (GSE37642-GPL96), with time-dependent AUC values for 1-, 3- and 5-year survival of 0.705,

**TABLE 1** | Estimation of STAT5-associated signature risk scores: Kaplan-Meier analysis and AUC of time-independent and time-dependent ROC curves in the 8 cohorts.

	Prognostic Performance (Kaplan-Meier analysis)		Predictive Accuracy (AUC of ROC curves)			
	Higher Risk Score Indicated Poorer Prognosis?	<i>p</i>	Overall	1 Year	3 Years	5 Years
GSE37642-GPL96	Yes	<b>0.0000</b>	0.705	0.705	0.731	0.703
GSE10691	Yes	<b>0.0004</b>	0.675	0.677	0.627	0.691
GSE12417-GPL96	Yes	<b>0.0128</b>	0.723	0.726	0.668	NA
GSE37642-GPL570	Yes	<b>0.0007</b>	0.674	0.675	0.633	0.677
GSE71014	Yes	<b>0.0026</b>	0.720	0.734	0.746	0.729
OHSU	Yes	<b>0.0000</b>	0.667	0.662	0.616	0.538
TARGET	Yes	<b>0.0008</b>	0.608	0.608	0.662	0.651
TCGA	Yes	<b>0.0001</b>	0.708	0.698	0.678	0.660

Bold values indicate  $p < 0.05$ .

0.731, and 0.703, respectively (Table 1; Supplementary Figure 4A). Moreover, the STAT5-associated signature also showed high predictive accuracy in most of the validation cohorts (GSE106291, Table 1 and Supplementary Figure 4B; GSE12417-GPL96, Table 1 and Supplementary Figure 4C; GSE37642-GPL570, Table 1 and Supplementary Figure 4D; GSE71014, Table 1 and Supplementary Figure 4E; OHSU, Table 1 and Supplementary Figure 4F; TARGET, Table 1 and Supplementary Figure 4G; and TCGA, Table 1 and Supplementary Figure 4H).

## Evaluating Prognostic Independence of the STAT5-Associated Signature

The prognostic independence of the STAT5-associated signature was assessed through uni- and multivariate Cox regression analyses in the training cohort (GSE37642-GPL96) and all of the validation cohorts except for GSE71014, which lacked clinicopathological variables (Supplementary Table S5). In univariate Cox analysis, age, cytogenetic risk category, ELN2017 risk category, and STAT5-associated signature risk score were significantly correlated with overall survival of AML patients (Figure 2A). In multivariate Cox analysis, the STAT5-associated signature was proved to be an independent predictor of survival in the TCGA cohort, with HR of 1.49 (1.09–2.05,  $p = 0.0136$ , Figure 2A). The predictive independence was also confirmed in other validation cohorts and corresponding values were 1.67 (1.39–2.00;  $p < 0.001$ ) in GSE37642-GPL96, 1.05 (1.02–1.09;  $p = 0.003$ ) in GSE106291, 1.21 (1.13–1.31;  $p < 0.001$ ) in GSE12417-GPL96, 2.66 (1.66–4.25;  $p < 0.001$ ) in GSE37642-GPL570, and 1.01 (1.00–1.03;  $p = 0.035$ ) in OHSU (Supplementary Table S5). In low- and high-risk groups, subgroup survival analyses by ages, percentage bone marrow blasts, FLT3 status, gender, NPM1 status, and platelet counts were performed in the TCGA cohort (Figures 2B–G). The STAT5-associated signature was also a promising prognostic predictor of overall survival in subgroups of patients in the TCGA cohort (Figures 2B–G). These results indicated that the STAT5-associated signature was an independent prognostic biomarker for AML.

## Construction of an Integrated Risk Score

In multivariate analyses, the STAT5-associated signature risk score and age of patients were independent predictors of survival, respectively (Figure 2A and Supplementary Table S5). To improve the predictive accuracy of the ELN2017 risk categories, the STAT5-associated signature risk score, age of patients, and ELN2017 risk category were integrated into an integrated score to predict the 1-, 3-, and 5-year survival probabilities in the training cohort (GSE37642-GPL96), which was visualized by a nomogram (Supplementary Figure 5A). The calibration curves were used to assess the predictive accuracy of the integrated score and revealed that the predicted survival 1-, 3-, and 5-year probabilities by integrated score were in good accordance with the corresponding actual survival probabilities (The higher the overlap between the red lines and black dashed lines, the more accurate the integrated score; Supplementary Figure 5B). The integrated scores with higher AUC values had

higher predictive accuracy than the ELN2017 risk category alone (Supplementary Figures 5C, D). Furthermore, the advantage of the integrated score was confirmed in two additional validation cohorts (Supplementary Figures 5E, F).

## Distribution of STAT5-Associated Signature Risk Scores in Different Subgroups

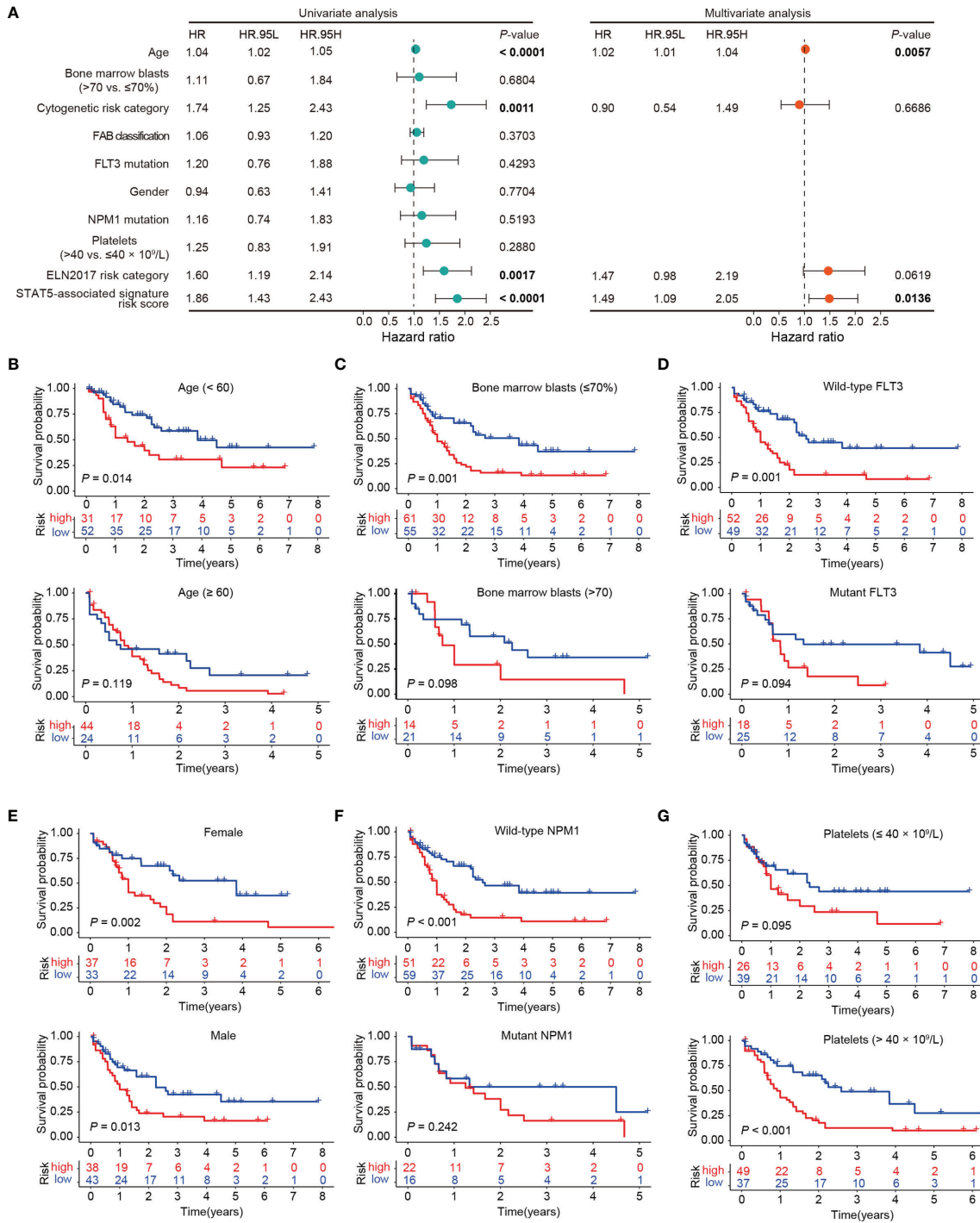
The distribution of STAT5-associated signature risk scores in diverse clinical and genetic risk subgroups was also investigated. Patients with an age of >60 years had significantly higher STAT5-associated signature risk scores compared with those with an age of ≤60 years in the training cohort (GSE37642-GPL96,  $p = 0.039$ ; Figure 3A). Similar associations were also observed in GSE106291 ( $p = 0.012$ ), OHSU ( $p = 0.010$ ), and TCGA ( $p = 0.013$ ) (Figure 3A). The STAT5-associated signature risk scores correlated well with the ELN2017 risk categories and increased along with the unfavorable ELN2017 risk levels in the training cohort (GSE37642-GPL96,  $r = 0.481$ ,  $p < 0.0001$ , Spearman correlation, Figure 3B). This correlation was confirmed in two other validation cohorts (OHSU,  $r = 0.446$ ,  $p < 0.0001$ , Spearman correlation; TCGA,  $r = 0.334$ ,  $p < 0.0001$ , Spearman correlation; Figure 3B). However, no correlation between STAT5-associated signature risk scores and percentage of bone marrow blasts was observed in the TARGET and TCGA cohorts, except for the OHSU cohort (Figure 3C).

## Characterization of Immune Cell Infiltration in Distinct STAT5-Associated Risk Groups

The characterization of immune-cell infiltration in distinct STAT5-associated risk groups was explored. The fractions of tumor-infiltrating immune cells were determined using CIBERSORT (34). The correlations between the fractions of tumor-infiltrating immune cells and STAT5-associated signature risk scores were assessed by Spearman correlation analysis in the training cohort (GSE37642-GPL96) and seven other cohorts (Figure 4A). The STAT5-associated signature risk scores were positively correlated with fractions of naïve B cells, naïve CD4<sup>+</sup> T cells, activated CD4<sup>+</sup> memory T cells, regulatory T cells (Tregs), activated NK cells, M0 macrophages, and neutrophils (Figure 4A). On the contrary, the STAT5-associated signature risk scores were negatively correlated with fractions of memory B cells, plasma cells, M2 macrophages, resting dendritic cells, resting mast cells (Figure 4A). In terms of the tumor microenvironment, patients in high-risk groups had significantly higher fractions of stromal cells and immune cells in some cohorts (Figures 4B, C).

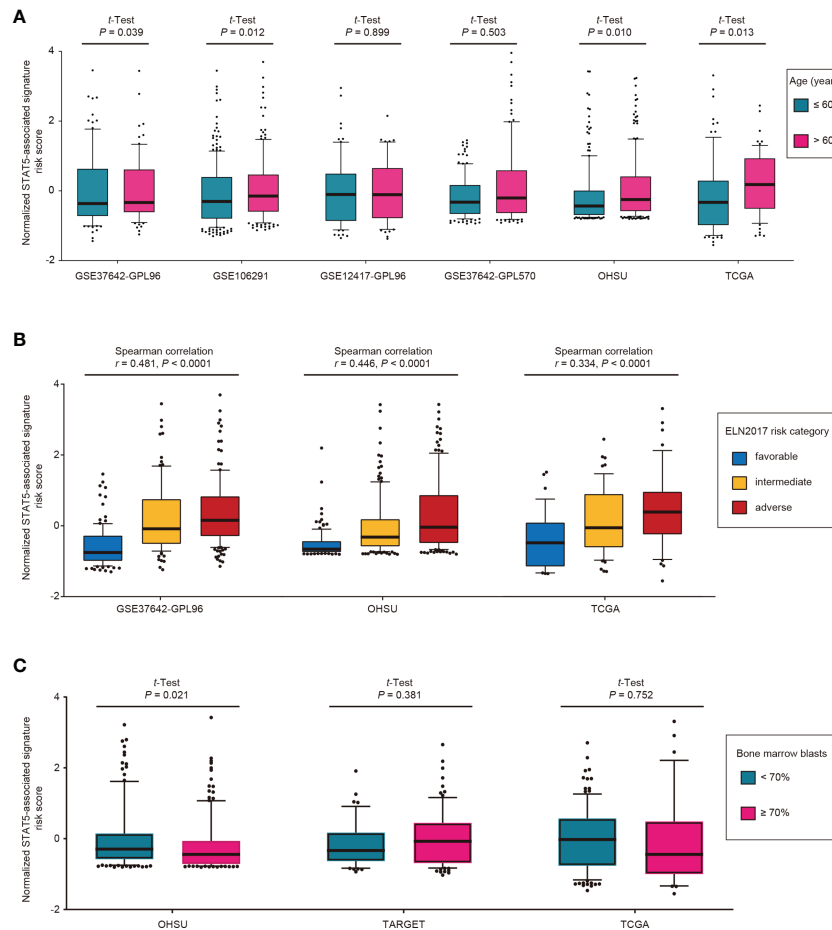
## Validation of the STAT5-Associated Signature by Analysis of In-House Clinical Samples

To validate the STAT5-associated signature, we detected the gene expression of the selected 6 genes for constructing the signature in peripheral blood mononuclear cells derived from 6 healthy donors and 28 AML patients using RT-qPCR



**FIGURE 2 |** Prognostic independence of the STAT5-associated signature. **(A)** Univariate and multivariate Cox regression analyses of the STAT5-associated signature and clinicopathological variables. Error bars represent hazard ratio (HR) with 95% confidence intervals (CI). **(B–G)** Kaplan-Meier survival analysis of subgroups stratified by age <60 and ≥60 **(B)**, bone marrow blasts ≤70% and >70% **(C)**, FLT3 wild-type and mutant subgroups **(D)**, female and male subgroups **(E)**, NPM1 wild-type and mutant subgroups **(F)**, platelets ≤40 and >40 × 10<sup>9</sup>/L **(G)**, respectively.





**FIGURE 3** | Distribution of STAT5-associated signature risk scores for different subgroups in the indicated cohorts. **(A)** Age, **(B)** ELN2017 risk category, and **(C)** bone marrow blasts. Among the total 8 cohorts, some cohorts were excluded due to inaccessibility of corresponding variables.

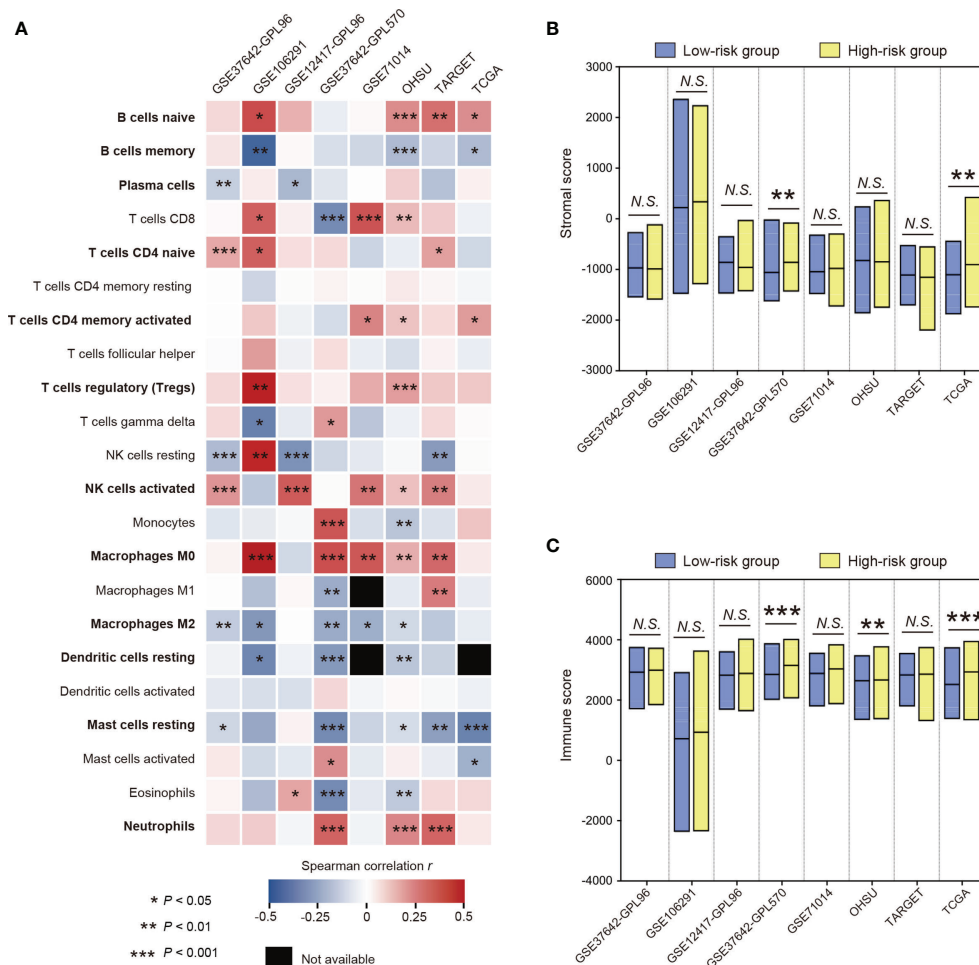
(Figure 5A). Gene expression levels of BATF, IFITM3, IGF2R, and SLCA29A2 were significantly higher in primary cells from AML patients than from healthy donors (Figure 5A). No difference in gene expression levels of PIM1 and SOCS2 was observed between primary cells from healthy donors and AML patients (Figure 5A). The STAT5-associated signature risk scores of all people were calculated based on the gene expression levels of the 6 genes in Figure 5A. AML patients had significantly higher STAT5-associated signature risk scores than healthy donors (Figure 5B). Phosphorylation of STAT5 is a prerequisite for activation of STAT5-associated pathways (43). As expected, phosphorylated STAT5 (pSTAT5) levels were higher in peripheral blood mononuclear cells from AML patients than healthy donors (Figure 5C).

### **In Silico Screening of Chemotherapy Drugs for Treatment of High-Risk AML Patients**

Half-maximal inhibitory concentrations ( $IC_{50}$ ) of 138 chemotherapeutic agents were estimated for each patient based on the transcriptomic data using the “PRRophetic” R package

(38) (details seen in the *Method* section). A drug with significantly lower  $IC_{50}$  in the high-risk group was determined as a targeted drug for high-risk patients in each cohort. The frequency with which a drug was determined as a targeted drug for high-risk patients among eight cohorts were quantified (Figure 6A), and five drugs with the highest frequencies were selected as screening hits, including bexarotene, bortezomib, erlotinib, rapamycin, and MS.275 (Figure 6A and Supplementary Figure 6A).

Further exploration of the underlying mechanism through which these five drugs targeted high-risk patients was also conducted. Phosphorylation of STAT5 is a prerequisite for activation of STAT5-associated pathways (43). Accordingly, the sensitivity of cell lines with different protein levels of phosphorylated STAT5 (Figure 6B) to these five drugs was determined in a cell proliferation assay (Figures 6C–G). The cell proliferation assay showed that the  $IC_{50}$  values of MS.275 were negatively correlated with the protein levels of phosphorylated STAT5 in six cancer cell lines ( $r = -0.812$ ,  $p = 0.0499$ ; Figure 6E). However, no correlation was observed for four other drugs



**FIGURE 4** | The characterization of immune cell infiltration based on the STAT5-associated signature. **(A)** Heatmap showing the relationship between fractions of tumor-infiltrating immune cells and STAT5-associated signature risk scores in each cohort. Twelve immune cell types that have strong correlations with STAT5-associated signature risk scores were highlighted in bold font. **(B, C)** Floating bars showing the differential composition of stromal cells **(B)** and immune cells **(C)** in the low- and high-risk groups. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , N.S., not significant.

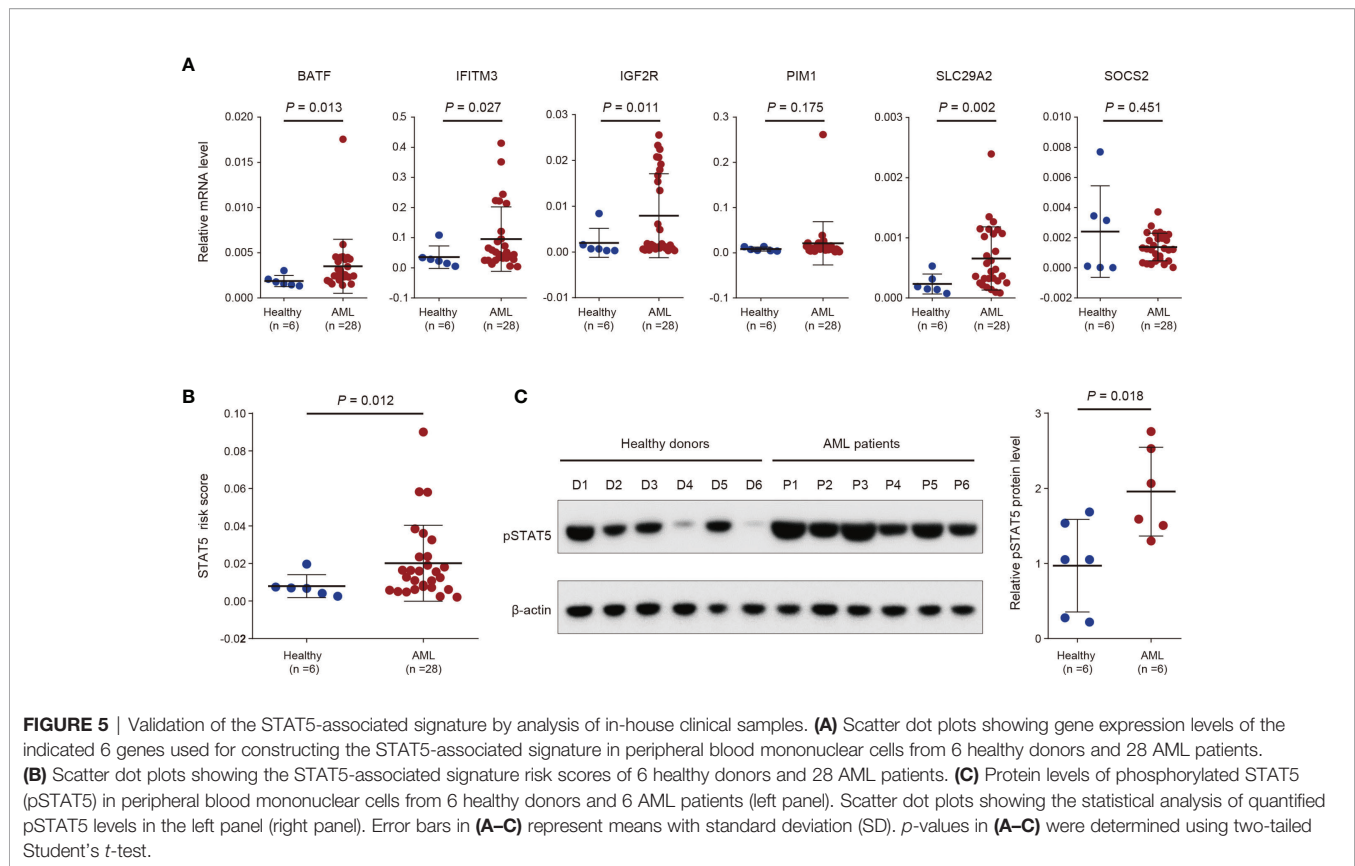
(Figures 6C, D, F, G). Overall, MS.275 might be a promising chemotherapy drug for the treatment of high-risk patients by targeting STAT5-associated pathways.

## DISCUSSION

In the present study, 235 robustly survival-related genes for AML were systematically identified through univariate Cox regression analysis of eight independent AML datasets. Pathway enrichment analysis with these 235 genes determined IL-2/STAT5 signaling pathway was the most highly enriched. In addition, it was reported that other enriched pathways including mechanistic target of rapamycin complex 1 (mTORC1) signaling pathway, androgen response, cholesterol homeostasis, estrogen response, and interferon gamma response were related to AML (44–48). Prognostic models based on these

gene pathways might be alternative candidates for predicting prognosis of AML patients.

The STAT5-associated prognostic signature for AML was constructed based on the genes *BATF*, *IFITM3*, *IGF2R*, *PIM1*, *SLC29A2*, and *SOCS2*. *BATF*, basic leucine zipper transcription factor ATF-like, is an important positive transcriptional regulator of the immune system that is particularly important in classical dendritic cell development, T follicular helper cell function and antibody production (49). *IFITM3*, interferon-induced transmembrane protein, plays a key role in cancer cell growth and maintenance, and is a marker of poor prognosis with high expression in many cancers, including AML (50). *IGF2R*, insulin-like growth factor 2 receptor, is currently considered a tumor suppressor gene, but it is upregulated and correlated with poor prognosis in cervical cancer (51) and glioblastomas (52). *PIM1*, proviral insertion site in murine leukemia virus (PIM) kinase 1, belongs to the PIM kinase family and has been implicated in the control of cancer cell proliferation, migration,

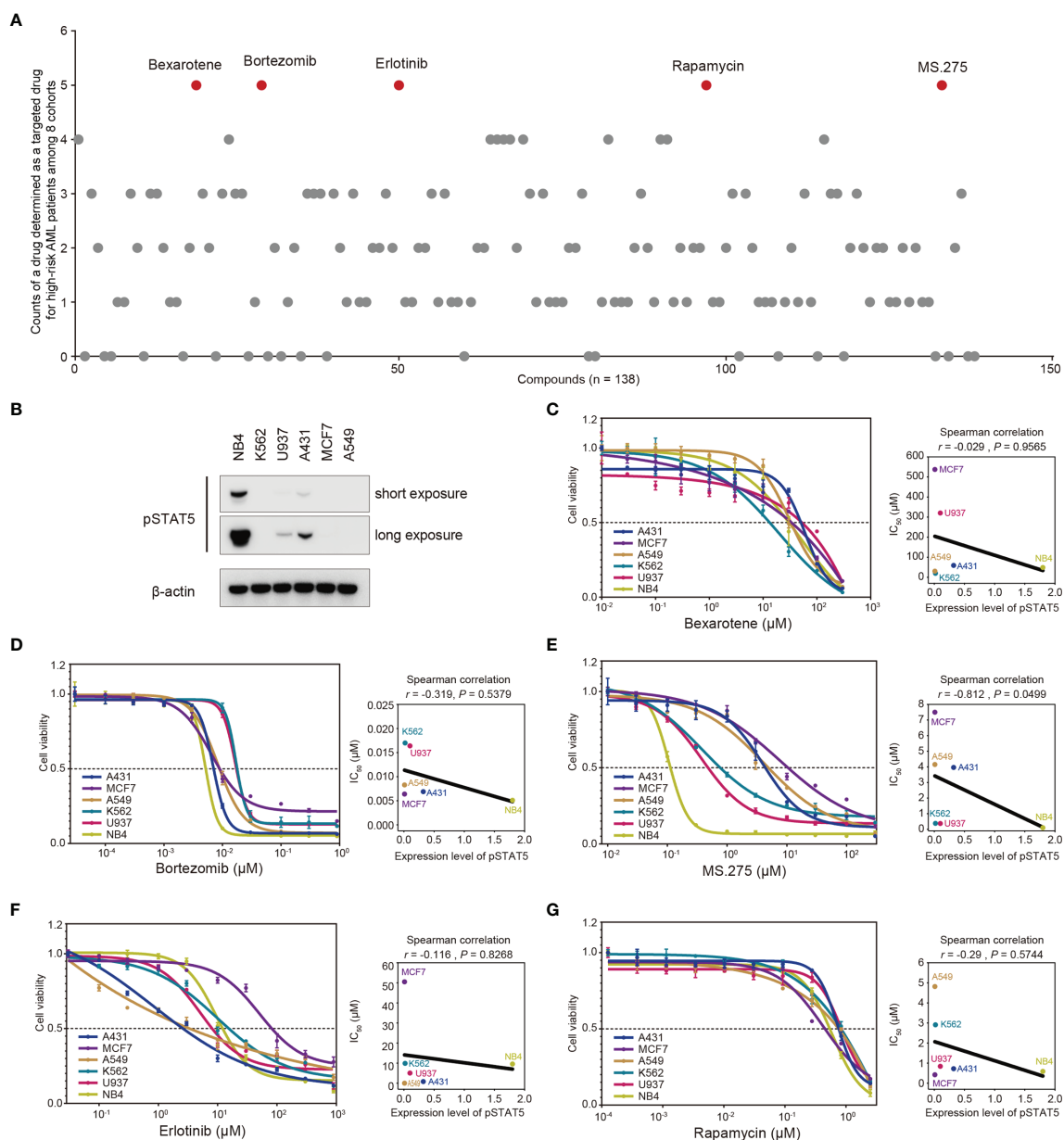


and apoptosis, particularly in prostate cancer and leukemia (53). SLC29A2, solute carrier family 29 member 2, is aberrantly upregulated and is a survival predictor in both hepatocellular carcinoma (54) and mantle cell lymphoma (55). SOCS2, suppressor of cytokine signaling-2, is highly upregulated and has tumor-promoting functions in the advanced stage of chronic myeloid leukemia (56) and in high-grade anal intraepithelial lesions (57). Furthermore, upregulation of SOCS2 is recognized as a potential prognostic marker for prostate cancer (58).

The good performance of the STAT5-associated signature was reproduced in most of the validation cohorts. Moreover, this signature was proven to be an independent prognostic factor upon multivariate Cox regression analysis and stratified survival analyses of several clinical characteristics. These results suggest that the STAT5-associated prognostic model may help predict the survival of AML patients.

It was reported that transcriptomic variables have higher predictive accuracy than genetic variables (13). However, the widely used clinical risk stratification system for AML, ELN2017, was constructed based on genetic and not transcriptomic variables (15). To complement this risk-assessment tool, an integrated score encompassing ELN2017 risk stratification, STAT5-associated signature risk scores and age of patients was constructed in the training cohort (GSE37642-GPL96). The STAT5-associated signature could improve the prognostic accuracy of ELN2017 risk categories in the training cohort (GSE37642-GPL96) as well as in two other independent cohorts.

Persistently phosphorylated STAT5 was found to suppress antitumor immunity (11). This suggests that immunological features also need to be investigated in myeloid neoplasms, since they will likely improve our knowledge of the underlying pathogenesis and inform novel therapies (18). Here, we characterized immune cell infiltration based on STAT5-associated risk stratification. The STAT5-associated signature risk scores were positively correlated with fractions of naïve B cells, and negatively correlated with fractions of memory B cells and plasma cells, which suggested impaired B-cell adaptive immunity in patients with high STAT5-associated signature risk scores (59, 60). Fractions of regulatory T cells (Tregs) and naïve CD4<sup>+</sup> T cells were found to be positively correlated with STAT5-associated signature risk scores, which implied immunosuppressive effects in the patients with high STAT5-associated signature risk scores (61, 62). Along with increasing STAT5-associated signature risk scores, we observed increasing neutrophils, increasing activated CD4<sup>+</sup> memory T cells, decreasing resting mast cells, and decreasing resting dendritic cells, which indicated severe infection in the patients with high STAT5-associated signature risk scores (63–66). At the same time, the anti-inflammatory function of high-risk patients might be weakened due to negative association of STAT5-associated signature risk scores with fractions of M2 macrophages and positive association of STAT5-associated signature risk scores with M0 macrophages (67). Unexpectedly, fractions of activated NK cells were positively correlated with STAT5-associated



**FIGURE 6 |** *In silico* screening of chemotherapy drugs for treatment of high-risk AML patients. **(A)** Half-maximal inhibitory concentration ( $\text{IC}_{50}$ ) of 138 chemotherapeutic agents for each patient was estimated based on the transcriptomic data. A drug with significantly lower  $\text{IC}_{50}$  in the high-risk group was determined as a targeted drug for high-risk patients in each cohort. The frequency with which a drug was determined as a targeted drug for high-risk patients among eight cohorts was quantified. Five drugs with the highest frequencies were screened out (red dots). **(B)** Baselines of phosphorylated STAT5 (pSTAT5) protein levels in six cancer cell lines. **(C–G)** Six cancer cell lines with different protein levels of phosphorylated STAT5 were treated with the indicated drugs for three days, followed by determination of cell viability. The correlations between  $\text{IC}_{50}$  values and protein levels of phosphorylated STAT5 were determined by Spearman correlation analysis.

signature risk scores. However, similar results were observed in another independent study (40), which might indicate tumor escape *via* defective expression of NK cell-triggering receptors by leukemic cells (68).

Chemotherapy remains the main treatment strategy for AML (69), and screening more effective chemotherapy drugs for high-risk patients might be a quick and economical strategy for

improving survival. To potentially improve the prognosis of high-risk patients, five chemotherapy drugs that were likely to be effective in high-risk patients were selected through *in silico* screening. The underlying mechanisms through which these five drugs target the high-risk patients were then investigated using cell viability assays. Among the five drugs, MS.275 selectively suppressed the cell lines with highly phosphorylated STAT5.



This result suggested that MS.275 might be a promising drug for the treatment of high-risk AML patients by targeting STAT5-associated pathways. MS.275 (Entinostat) is an oral class I histone deacetylase (HDAC) inhibitor that blocks cell proliferation and promotes apoptosis in breast cancer (70). The antitumor activity of MS.275 in AML was also reported, including the induction of robust differentiation of AML cell lines (71), inducing apoptosis in AML cell lines (72), and inhibited disease maintenance in a mouse model of AML (73). Clinical trials of MS.275 for the treatment of hematological cancers including AML were also performed by different groups (NCT00015925, NCT01159301, NCT01132573, NCT00313586, NCT01305499, NCT00462605, NCT00101179, NCT01383447). These concerted efforts will enrich the therapy regimen for AML in the clinic, and hopefully improve the prognosis of high-risk patients.

However, there are also some limitations to the current study. In the subgroup analysis used to validate the prognostic independence of this model, the difference was not statistically significant due to an insufficient number of patients in some subgroups, such as the mutant NPM1 subgroup. The underlying mechanisms through which the five chemotherapy drugs other than MS.275 target AML in high-risk patients are still unknown. Potential biases of this model exist due to heterogeneity of patients, therapy regimens, and disease stage. Additionally, this is a retrospective study with a few experiments, so the findings remain to be further validated in both the laboratory and the clinic.

In conclusion, we comprehensively analyzed the genes that are most strongly related to survival in AML. Pathway enrichment analysis of these robustly survival-related genes indicated that IL-2/STAT5 is the most highly enriched signaling pathway. A STAT5-associated signature was constructed on the basis of robustly survival-related genes related to the IL-2/STAT5 signaling pathway. The signature could independently predict survival of AML patients, and our prognostic model might complement and improve the current risk system based on genetic variables, such as the ELN2017 risk categories. The immune infiltration was also investigated based on the risk phenotype, which will contribute to immunotherapy of high-risk patients in the future. Analysis of in-house clinical samples revealed that the STAT5-associated signature risk scores of AML patients were significantly higher than those of healthy people. MS.275, a known HDAC inhibitor, was demonstrated as a targeted drug for high-risk patients by interfering with STAT5-associated pathways. This reliable model could be used for prognostic assessment and guidance for precision therapy for AML.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB of the First Affiliated Hospital of Jinan University. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

ML and YT designed this work. YT, ZW, YL, YX, JW, and SX retrieved the data and conducted the analyses. YT, SX, YL, YX, JW, and ZW performed the experiment. ML, YT, and SX wrote this manuscript. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## FUNDING

This work was funded by the National Natural Science Foundation of China (NSFC 81900157 & 82102756), the China Postdoctoral Science Foundation (2021M692111 & 2018M640399), and the SJTU Trans-med Awards Research and the Foundation of the National Facility for Translational Medicine (Shanghai) (TMSK-2020-003).

## ACKNOWLEDGMENTS

We thank Dr. Li Ran for kindly providing ELN2017 risk category data of TCGA. We thank Prof. Tobias Herold of the AMLCG study group for kindly providing ELN2017 risk category data of GSE37642-GPL96.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2022.785899/full#supplementary-material>

**Supplementary Figure 1** | Identification of genes related to survival in AML patients and construction of a STAT5-associated signature. **(A)** The gene expression levels of STAT5A and STAT5B in 69 cancer cell lines derived from 16 distinct organs. Each dot represents a cancer cell line. *P* values were determined using two-tailed Student's *t*-test. **(B)** The protein-protein interaction (PPI) network of robustly survival-related genes (text in white) mapping to the IL-2/STAT5 pathway. **(C–H)** Kaplan–Meier survival analysis of low- and high-expression groups of BATF, IFITM3, IGF2R, PIM1, SLC29A2, and SOCS2.

**Supplementary Figure 2** | The relationship between STAT5-associated signature risk scores and survival status. **(A–H)** STAT5-associated signature risk scores arrangement and survival status analyses in GSE37642-GPL96 **(A)**, GSE106291 **(B)**, GSE12417-GPL96 **(C)**, GSE37642-GPL570 **(D)**, GSE71014 **(E)**, OHSU **(F)**, TARGET **(G)**, and TCGA **(H)**.

**Supplementary Figure 3** | Kaplan–Meier survival analyses. **(A–H)** Kaplan–Meier curves for overall survival stratified by different risk levels in GSE37642-GPL96 **(A)**,

GSE106291 (B), GSE12417-GPL96 (C), GSE37642-GPL570 (D), GSE71014 (E), OHSU (F), TARGET (G), and TCGA (H).

**Supplementary Figure 4** | Time-independent and time-dependent ROC analyses. (A–H) AUC of time-independent (left) and time-dependent (right) ROC curves of STAT5-associated signature in GSE37642-GPL96 (A), GSE106291 (B), GSE12417-GPL96 (C), GSE37642-GPL570 (D), GSE71014 (E), OHSU (F), TARGET (G), and TCGA (H).

**Supplementary Figure 5** | Construction of an integrated risk score. (A) Nomogram visualizing the integrated risk model constructed based on the STAT5-associated signature risk score, patient age, and ELN2017 risk category in the

training cohort (GSE37642-GPL96). (B) Calibration curves of the nomogram in terms of agreement between predicted and observed 1-year, 3-year, 5-year survival in the training cohort (GSE37642-GPL96). Error bars represent actual overall survival probability with 95% confidence intervals (CI). (C) Comparison of the time-dependent ROC curves of the integrated risk score and its component single risk categories in the training cohort (GSE37642-GPL96). (D–F) Comparison of the time-independent ROC curves of the integrated risk score and its component single risk categories in GSE37642-GPL96 (D), OHSU (E), and TCGA (F).

**Supplementary Figure 6** | *In silico* screening of chemotherapy drugs for treatment of high-risk AML patients. (A) Estimated IC<sub>50</sub> for the five hits from in the indicated cohorts.

## REFERENCES

1. Khwaja A, Bjorkholm M, Gale RE, Levine RL, Jordan CT, Ehninger G, et al. Acute Myeloid Leukaemia. *Nat Rev Dis Primers* (2016) 2:16010. doi: 10.1038/nrdp.2016.10
2. Meyer SC, Levine RL. Translational Implications of Somatic Genomics in Acute Myeloid Leukaemia. *Lancet Oncol* (2014) 15(9):e382–94. doi: 10.1016/S1470-2045(14)70008-7
3. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional Genomic Landscape of Acute Myeloid Leukaemia. *Nature* (2018) 562(7728):526–31. doi: 10.1038/s41586-018-0623-z
4. Döhner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl J Med* (2015) 373(12):1136–52. doi: 10.1056/NEJMra1406184
5. Grimley PM, Dong F, Rui H. Stat5a and Stat5b: Fraternal Twins of Signal Transduction and Transcriptional Activation. *Cytokine Growth Factor Rev* (1999) 10(2):131–57. doi: 10.1016/S1359-6101(99)00011-8
6. Wingelhofer B, Maurer B, Heyes EC, Kumaraswamy AA, Berger-Becvar A, de Araujo ED, et al. Pharmacologic Inhibition of STAT5 in Acute Myeloid Leukemia. *Leukemia* (2018) 32(5):1135–46. doi: 10.1038/s41375-017-0005-9
7. Sockolovsky JT, Trotta E, Parisi G, Picton L, Su LL, Le AC, et al. Selective Targeting of Engineered T Cells Using Orthogonal IL-2 Cytokine-Receptor Complexes. *Sci (New York NY)* (2018) 359(6379):1037–42. doi: 10.1126/science.aar3246
8. Birkenkamp KU, Geugien M, Lemmink HH, Kruijer W, Vellenga E. Regulation of Constitutive STAT5 Phosphorylation in Acute Myeloid Leukemia Blasts. *Leukemia* (2001) 15(12):1923–31. doi: 10.1038/sj.leu.2402317
9. Hoelbl A, Schuster C, Kovacic B, Zhu B, Wickre M, Hoelzl MA, et al. Stat5 Is Indispensable for the Maintenance of Bcr/Abl-Positive Leukaemia. *EMBO Mol Med* (2010) 2(3):98–110. doi: 10.1002/emmm.201000062
10. Warsch W, Kollmann K, Eckelhart E, Fajmann S, Cerny-Reiterer S, Hölbl A, et al. High STAT5 Levels Mediate Imatinib Resistance and Indicate Disease Progression in Chronic Myeloid Leukemia. *Blood* (2011) 117(12):3409–20. doi: 10.1182/blood-2009-10-248211
11. Yu H, Pardoll D, Jove R. STATs in Cancer Inflammation and Immunity: A Leading Role for STAT3. *Nat Rev Cancer* (2009) 9(11):798–809. doi: 10.1038/nrc2734
12. Mao Q-F, Shang-Guan Z-F, Chen H-L, Huang K. Immunoregulatory Role of IL-2/STAT5/CD4+CD25+Foxp3 Treg Pathway in the Pathogenesis of Chronic Osteomyelitis. *Ann Trans Med* (2019) 7(16):384. doi: 10.21037/atm.2019.07.45
13. Gerstung M, Pellagatti A, Malcovati L, Giagounidis A, Porta MGD, Jädersten M, et al. Combining Gene Mutation With Gene Expression Data Improves Outcome Prediction in Myelodysplastic Syndromes. *Nat Commun* (2015) 6:5901. doi: 10.1038/ncomms6901
14. Oh C-K, Ha M, Han M-E, Heo HJ, Myung K, Lee Y, et al. FAM213A is Linked to Prognostic Significance in Acute Myeloid Leukemia Through Regulation of Oxidative Stress and Myelopoiesis. *Hematol Oncol* (2020) 38(3):381–9. doi: 10.1002/hon.2728
15. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, et al. Diagnosis and Management of AML in Adults: 2017 ELN Recommendations From an International Expert Panel. *Blood* (2017) 129(4):424–47. doi: 10.1182/blood-2016-08-733196
16. Yan H, Qu J, Cao W, Liu Y, Zheng G, Zhang E, et al. Identification of Prognostic Genes in the Acute Myeloid Leukemia Immune Microenvironment Based on TCGA Data Analysis. *Cancer Immunology Immunother CII* (2019) 68(12):1971–8. doi: 10.1007/s00262-019-02408-7
17. Chen C, Liang C, Wang S, Chio CL, Zhang Y, Zeng C, et al. Expression Patterns of Immune Checkpoints in Acute Myeloid Leukemia. *J Hematol Oncol* (2020) 13(1):28. doi: 10.1186/s13045-020-00853-x
18. Li R, Ding Z, Jin P, Wu S, Jiang G, Xiang R, et al. Development and Validation of a Novel Prognostic Model for Acute Myeloid Leukemia Based on Immune-Related Genes. *Front Immunol* (2021) 12:639634. doi: 10.3389/fimmu.2021.639634
19. Chen X-X, Li Z-P, Zhu J-H, Xia H-T, Zhou H. Systematic Analysis of Autophagy-Related Signature Uncovers Prognostic Predictor for Acute Myeloid Leukemia. *DNA Cell Biol* (2020) 39(9):1595–605. doi: 10.1089/dna.2020.5667
20. Metzler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland M-C, et al. An 86-Probe-Set Gene-Expression Signature Predicts Survival in Cytogenetically Normal Acute Myeloid Leukemia. *Blood* (2008) 112(10):4193–201. doi: 10.1182/blood-2008-02-134411
21. Li Z, Herold T, He C, Valk PJM, Chen P, Jurinovic V, et al. Identification of a 24-Gene Prognostic Signature That Improves the European LeukemiaNet Risk Classification of Acute Myeloid Leukemia: An International Collaborative Study. *J Clin Oncol Off J Am Soc Clin Oncol* (2013) 31(9):1172–81. doi: 10.1200/JCO.2012.44.3184
22. Chuang M-K, Chiu Y-C, Chou W-C, Hou H-A, Tseng M-H, Kuo Y-Y, et al. An mRNA Expression Signature for Prognostication in *De Novo* Acute Myeloid Leukemia Patients With Normal Karyotype. *Oncotarget* (2015) 6(36):39098–110. doi: 10.18632/oncotarget.5390
23. Herold T, Jurinovic V, Batcha AMN, Bamopoulos SA, Rothenberg-Thurley M, Ksienzyk B, et al. A 29-Gene and Cytogenetic Score for the Prediction of Resistance to Induction Treatment in Acute Myeloid Leukemia. *Haematologica* (2018) 103(3):456–65. doi: 10.3324/haematol.2017.178442
24. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, et al. Genomic and Epigenomic Landscapes of Adult *De Novo* Acute Myeloid Leukemia. *N Engl J Med* (2013) 368(22):2059–74. doi: 10.1056/NEJMoa1301689
25. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The Cbio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov* (2012) 2(5):401–4. doi: 10.1158/2159-8290.CD-12-0095
26. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the Cbioportal. *Sci Signal* (2013) 6(269):pl1. doi: 10.1126/scisignal.2004088
27. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING V10: Protein-Protein Interaction Networks, Integrated Over the Tree of Life. *Nucleic Acids Res* (2015) 43(Database issue):D447–52. doi: 10.1093/nar/gku1003
28. Cox DR. Regression Models and Life-Tables. *J R Stat Soc Ser B (Methodol)* (1972) 34(2):187–220. doi: 10.1111/j.2517-6161.1972.tb00899.x
29. Therneau T, Grambsch P. *Modeling Survival Data: Extending The Cox Model*. New York: Springer Press (2000).
30. McLain AC, Ghosh SK. Efficient Sieve Maximum Likelihood Estimation of Time-Transformation Models. *J Stat Theory Pract* (2013) 7(2):285–303. doi: 10.1080/15598608.2013.772835
31. Hothorn T, Möst L, Bühlmann P. Most Likely Transformations. *Scandinavian J Stat* (2018) 45(1):110–34. doi: 10.1111/sjos.12291

32. Marhuenda Y, Morales D, Pardo M. Information Criteria for Fay-Herriot Model Selection. *Comput Stat Data Anal* (2014) 70:268–80. doi: 10.1016/j.csda.2013.09.016
33. Steyerberg EW, FRANK E, HARRELL, Jr., Regression Modeling Strategies: With Applications, to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, 2nd Ed. Heidelberg: Springer. *Biometrics* (2016) 72 (3):1006–7. doi: 10.1111/biom.12569
34. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells With CIBERSORT. *Methods Mol Biol (Clifton NJ)* (2018) 1711:243–59. doi: 10.1007/978-1-4939-7493-1\_12
35. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-García W, et al. Inferring Tumour Purity and Stromal and Immune Cell Admixture From Expression Data. *Nat Commun* (2013) 4(1):2612. doi: 10.1038/ncomms3612
36. Chen S, Wu J-L, Liang Y, Tang Y-G, Song H-X, Wu L-L, et al. Arsenic Trioxide Rescues Structural P53 Mutations Through a Cryptic Allosteric Site. *Cancer Cell* (2021) 39(2):225–39. doi: 10.1016/j.ccell.2020.11.013
37. Gleeleher P, Cox NJ, Huang RS. Clinical Drug Response can be Predicted Using Baseline Gene Expression Levels and *In Vitro* Drug Sensitivity in Cell Lines. *Genome Biol* (2014) 15(3):R47. doi: 10.1186/gb-2014-15-3-r47
38. Gleeleher P, Cox N, Huang RS. Prorhetic: An R Package for Prediction of Clinical Chemotherapeutic Response From Tumor Gene Expression Levels. *PLoS One* (2014) 9(9):e107468. doi: 10.1371/journal.pone.0107468
39. Alboukadel K, Marcin K, Przemyslaw B. *Survminer: Drawing Survival Curves Using 'ggplot2'*. (2021). Available at: <https://CRAN.R-project.org/package=survminer>.
40. Wang Y, Cai Y-Y, Herold T, Nie R-C, Zhang Y, Gale RP, et al. An Immune Risk Score Predicts Survival of Patients With Acute Myeloid Leukemia Receiving Chemotherapy. *Clin Cancer Res Off J Am Assoc Cancer Res* (2021) 27(1):255–66. doi: 10.1158/1078-0432.CCR-20-3417
41. Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and Comparing Time-Dependent Areas Under Receiver Operating Characteristic Curves for Censored Event Times With Competing Risks. *Stat Med* (2013) 32(30):5381–97. doi: 10.1002/sim.5958
42. Heagerty PJ, Lumley T, Pepe MS. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* (2000) 56(2):337–44. doi: 10.1111/j.0006-341x.2000.00337.x
43. Subramaniam D, Angulo P, Ponnurangam S, Dandawate P, Ramamoorthy P, Srinivasan P, et al. Suppressing STAT5 Signaling Affects Osteosarcoma Growth and Stemness. *Cell Death Dis* (2020) 11(2):149. doi: 10.1038/s41419-020-2335-1
44. Murohashi I, Hoang T. Interferon-Gamma Enhances Growth Factor-Dependent Proliferation of Clonogenic Cells in Acute Myeloblastic Leukemia. *Blood* (1991) 78(4):1085–95. doi: 10.1182/blood.V78.4.1085.1085
45. Park S, Chapuis N, Tamburini J, Bardet V, Cornillet-Lefebvre P, Willems L, et al. Role of the PI3K/AKT and mTOR Signaling Pathways in Acute Myeloid Leukemia. *Haematologica* (2010) 95(5):819–28. doi: 10.3324/haematol.2009.013797
46. Jurcic JG. Androgen Maintenance Therapy for Acute Myeloid Leukemia. *J Clin Oncol* (2016) 35(4):381–3. doi: 10.1200/JCO.2016.70.4999
47. Rota SG, Roma A, Dude I, Ma C, Stevens R, MacEachern J, et al. Estrogen Receptor  $\beta$  Is a Novel Target in Acute Myeloid Leukemia. *Mol Cancer Ther* (2017) 16(11):2618–26. doi: 10.1158/1535-7163.mct-17-0292
48. Brendolan A, Russo V. Targeting Cholesterol Homeostasis in Hematopoietic Malignancies. *Blood* (2022) 139(2):165–76. doi: 10.1182/blood.2021012788
49. Murphy TL, Tussiwand R, Murphy KM. Specificity Through Cooperation: BATF-IRF Interactions Control Immune-Regulatory Networks. *Nat Rev Immunol* (2013) 13(7):499–509. doi: 10.1038/nri3470
50. Rajapaksa US, Jin C, Dong T. Malignancy and IFITM3: Friend or Foe? *Front Oncol* (2020) 10:593245. doi: 10.3389/fonc.2020.593245
51. Takeda T, Komatsu M, Chiwaki F, Komatsuzaki R, Nakamura K, Tsuji K, et al. Upregulation of IGF2R Evades Lysosomal Dysfunction-Induced Apoptosis of Cervical Cancer Cells via Transport of Cathepsins. *Cell Death Dis* (2019) 10(12):876. doi: 10.1038/s41419-019-2117-9
52. Varghese RT, Liang Y, Guan T, Franck CT, Kelly DF, Sheng Z. Survival Kinase Genes Present Prognostic Significance in Glioblastoma. *Oncotarget* (2016) 7(15):20140–51. doi: 10.18632/oncotarget.7917
53. Brasó-Maristany F, Filosto S, Catchpole S, Marlow R, Quist J, Francesch-Domenech E, et al. PIM1 Kinase Regulates Cell Death, Tumor Growth and Chemotherapy Response in Triple-Negative Breast Cancer. *Nat Med* (2016) 22(11):1303–13. doi: 10.1038/nm.4198
54. Chen C-F, Hsu E-C, Lin K-T, Tu P-H, Chang H-W, Lin C-H, et al. Overlapping High-Resolution Copy Number Alterations in Cancer Genomes Identified Putative Cancer Genes in Hepatocellular Carcinoma. *Hepatology (Baltimore Md)* (2010) 52(5):1690–701. doi: 10.1002/hep.23847
55. Hartmann E, Fernández V, Moreno V, Valls J, Hernández L, Bosch F, et al. Five-Gene Model to Predict Survival in Mantle-Cell Lymphoma Using Frozen or Formalin-Fixed, Paraffin-Embedded Tissue. *J Clin Oncol Off J Am Soc Clin Oncol* (2008) 26(30):4966–72. doi: 10.1200/JCO.2007.12.0410
56. Schultheis B, Carapeti-Marootian M, Hochhaus A, Weisser A, Goldman JM, Melo JV. Overexpression of SOCS-2 in Advanced Stages of Chronic Myeloid Leukemia: Possible Inadequacy of a Negative Feedback Mechanism. *Blood* (2002) 99(5):1766–75. doi: 10.1182/blood.V99.5.1766
57. Arany I, Muldrow M, Tying SK. The Endogenous Interferon System in Anal Squamous Epithelial Lesions With Different Grades From HIV-Positive Individuals. *Int J STD AIDS* (2001) 12(4):229–33. doi: 10.1258/0956462011922977
58. Zhu J-G, Dai Q-S, Han Z-D, He H-C, Mo R-J, Chen G, et al. Expression of SOCSs in Human Prostate Cancer and Their Association in Prognosis. *Mol Cell Biochem* (2013) 381(1-2):51–9. doi: 10.1007/s11010-013-1687-6
59. Seifert M, Küppers R. Human Memory B Cells. *Leukemia* (2016) 30(12):2283–92. doi: 10.1038/leu.2016.226
60. Allen HC, Sharma P. Histology, Plasma Cells. In: *StatPearls*. Treasure Island, FL: StatPearls Publishing (2021).
61. Su S, Liao J, Liu J, Huang D, He C, Chen F, et al. Blocking the Recruitment of Naive CD4(+) T Cells Reverses Immunosuppression in Breast Cancer. *Cell Res* (2017) 27(4):461–82. doi: 10.1038/cr.2017.34
62. Xu Y, Mou J, Wang Y, Zhou W, Rao Q, Xing H, et al. Regulatory T Cells Promote the Stemness of Leukemia Stem Cells Through IL10 Cytokine-Related Signaling Pathway. *Leukemia* (2021). doi: 10.1038/s41375-021-01375-2
63. Banchereau J, Steinman RM. Dendritic Cells and the Control of Immunity. *Nature* (1998) 392(6673):245–52. doi: 10.1038/32588
64. MacLeod MK, Clambey ET, Kappler JW, Marrack P. CD4 Memory T Cells: What Are They and What Can They Do? *Semin Immunol* (2009) 21(2):53–61. doi: 10.1016/j.smim.2009.02.006
65. Krystal-Whittemore M, Dileepan KN, Wood JG. Mast Cell: A Multifunctional Master Cell. *Front Immunol* (2015) 6:620. doi: 10.3389/fimmu.2015.00620
66. Jenne CN, Liao S, Singh B. Neutrophils: Multitasking First Responders of Immunity and Tissue Homeostasis. *Cell Tissue Res* (2018) 371(3):395–7. doi: 10.1007/s00441-018-2802-5
67. Jablonski KA, Amici SA, Webb LM, Ruiz-Rosado JDD, Popovich PG, Partida-Sanchez S, et al. Novel Markers to Delineate Murine M1 and M2 Macrophages. *PLoS One* (2015) 10(12):e0145342. doi: 10.1371/journal.pone.0145342
68. Costello RT, Sivori S, Marcenaro E, Lafage-Pochitaloff M, Mozziconacci M-J, Reviron D, et al. Defective Expression and Function of Natural Killer Cell-Triggering Receptors in Patients With Acute Myeloid Leukemia. *Blood* (2002) 99(10):3661–7. doi: 10.1182/blood.V99.10.3661
69. Nair R, Salinas-Illarena A, Baldauf H-M. New Strategies to Treat AML: Novel Insights Into AML Survival Pathways and Combination Therapies. *Leukemia* (2021) 35(2):299–311. doi: 10.1038/s41375-020-01069-1
70. Trapani D, Esposito A, Criscitiello C, Mazzarella L, Locatelli M, Minchella I, et al. Entinostat for the Treatment of Breast Cancer. *Expert Opin Investig Drugs* (2017) 26(8):965–71. doi: 10.1080/13543784.2017.1353077
71. Blagitko-Dorfs N, Jiang Y, Duque-Afonso J, Hiller J, Yalcin A, Greve G, et al. Epigenetic Priming of AML Blasts for All-Trans Retinoic Acid-Induced Differentiation by the HDAC Class-I Selective Inhibitor Entinostat. *PLoS One* (2013) 8(10):e75258. doi: 10.1371/journal.pone.0075258
72. Zhou L, Ruvolo VR, McQueen T, Chen W, Samudio IJ, Conneely O, et al. HDAC Inhibition by SNDX-275 (Entinostat) Restores Expression of Silenced Leukemia-Associated Transcription Factors Nur77 and Nor1 and of Key Pro-Apoptotic Proteins in AML. *Leukemia* (2013) 27(6):1358–68. doi: 10.1038/leu.2012.366
73. Ramsey JM, Kettyle LMJ, Sharpe DJ, Mulgrew NM, Dickson GJ, Bijl JJ, et al. Entinostat Prevents Leukemia Maintenance in a Collaborating Oncogene-Dependent Model of Cytogenetically Normal Acute Myeloid

Leukemia. *Stem Cells (Dayton Ohio)* (2013) 31(7):1434–45. doi: 10.1002/stem.1398

**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.

*Copyright © 2022 Tang, Xiao, Wang, Liang, Xing, Wu and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.*