EBioMedicine 40 (2019) 240–250

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Incorporation of long non-coding RNA expression profile in the 2017 ELN risk classification can improve prognostic prediction of acute myeloid leukemia patients

Cheng-Hong Tsai [a,b,c], Chi-Yuan Yao [a], Feng-Min Tien [a,b,d], Jih-Luh Tang [a,b], Yuan-Yeh Kuo [b], Yu-Chiao Chiu [e], Chien-Chin Lin [a,d,f], Mei-Hsuan Tseng [a], Yen-Ling Peng [a], Ming-Chih Liu [g], Chia-Wen Liu [g], Ming Yao [a], Liang-In Lin [h], Wen-Chien Chou [a,f], Chien-Yu Chen [c,i], Hsin-An Hou [a,*], Hwei-Fang Tien [a,*]

[a] Division of Hematology, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan
[b] Tai-Cheng Stem Cell Therapy Center, National Taiwan University, Taipei, Taiwan
[c] Genome and Systems Biology Degree Program, National Taiwan University, Taipei, Taiwan
[d] Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan
[e] Greehey Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA
[f] Departments of Laboratory Medicine, National Taiwan University Hospital, Taipei, Taiwan
[g] Pathology, National Taiwan University Hospital, Taipei, Taiwan
[h] Department of Clinical Laboratory Sciences and Medical Biotechnology, College of Medicine, National Taiwan University, Taipei, Taiwan
[i] Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 2 November 2018
Received in revised form 5 January 2019
Accepted 8 January 2019
Available online 17 January 2019

Keywords:
lncRNA
AML
Risk stratification
Prognostication

ABSTRACT

Background: Expression of long non-coding RNAs (lncRNAs) has recently been recognized as a potential prognostic marker in acute myeloid leukemia (AML). However, it remains unclear whether incorporation of the lncRNAs expression in the 2017 European LeukemiaNet (ELN) risk classification can further improve the prognostic prediction.

Methods: We enrolled 275 newly diagnosed non-M3 AML patients and randomly assigned them to the training ($n = 183$) and validation cohorts ($n = 92$). In the training cohort, we formulated a prognostic lncRNA scoring system composed of five lncRNAs with significant prognostic impact from the lncRNA expression profiling.

Findings: Higher lncRNA scores were significantly associated with older age and adverse gene mutations. Further, the higher-score patients had shorter overall and disease-free survival than lower-score patients, which were also confirmed in both internal and external validation cohorts (TCGA database). The multivariate analyses revealed the lncRNA score was an independent prognosticator in AML, irrespective of the risk based on the 2017 ELN classification. Moreover, in the 2017 ELN intermediate-risk subgroup, lncRNA scoring system could well dichotomize the patients into two groups with distinct prognosis. Within the ELN intermediate-risk subgroup, we found that allogeneic hematopoietic stem cell transplantation could provide better outcome on patients with higher lncRNA scores. Through bioinformatics approach, we identified high lncRNA scores were correlated with leukemia/hematopoietic stem cell signatures.

Interpretation: Incorporation of lncRNA scoring system in 2017 ELN classification can improve risk-stratification of AML patients and help clinical decision-making.

Fund: This work was supported Ministry of Science and Technology, and Ministry of Health and Welfare of Taiwan.

1. Introduction

Acute myeloid leukemia (AML) is an aggressive hematologic malignancy, characterized by uncontrolled proliferation and impaired differentiation of hematopoietic precursors. AML patients have substantial heterogeneity in the pathogenesis of the disease, clinical features, and treatment outcomes [1]. Therefore, risk-stratification is one of the most critical factors in the era of precision medicine.

Risk-stratification of AML patients by the United Kingdom Medical Research Council (MRC) cytogenetic classification and European LeukemiaNet (ELN) risk classification have been widely used [2][3].

* Correspondence authors.
E-mail addresses: hsinanhou@ntu.edu.tw (H.-A. Hou), hftien@ntu.edu.tw (H.-F. Tien).

https://doi.org/10.1016/j.ebiom.2019.01.022

Research in context

Evidence before this study

The 2017 European LeukemiaNet (ELN) expert panel is probably the most widely used risk-stratification model in current clinical practice. The model is composed of cytogenetic changes and recurrent gene mutations. However, there are still substantial patients classified into the intermediate-risk group, in which the treatment strategy is not well defined. Recently, besides the coding events, non-coding events have been discovered to have significant impact on patients' outcome. Although researchers have found that certain long non-coding RNAs (lncRNAs) have prognostic impact, whether incorporating lncRNA into the 2017 ELN classification can improve the prognostication is still unknown.

Added value of this study

We formulated a simple but precise lncRNA scoring system composed of only five lncRNAs and validated in both the internal and external cohorts, which were composed of different patient races and gene expression profile based on different platforms. The scoring system is closely associated with stem cell signature and can well risk-stratify intermediate-risk patients defined by the 2017 ELN classification.

Implications of all the available evidence

This lncRNA scoring system can help clinicians to provide proper treatment for the intermediate-risk patients defined by the 2017 ELN classification.

However, a large proportion of patients are in the intermediate-risk subgroup in whom there is substantial prognostic heterogeneity [4]. More sophisticated systems are needed to improve the risk-stratification of AML patients. Besides aberrations of the coding genes, a number of studies showed that non-coding RNAs, which lack protein-coding potential, also play important roles in the pathogenesis of AML and influence its prognosis [5–9]. Long non-coding RNAs (lncRNAs) are transcripts >200 nucleotides in length that act as gene expression regulators through epigenetic, transcriptional, or posttranscriptional regulation and are involved in various biological functions including cell cycle, apoptosis, differentiation, *etc.* [10–12] Recently, mounting evidences showed prognostic relevance of lncRNA expression in AML patients [13–15]. However, it remains unclear whether incorporation of the lncRNA expression in the newly developed 2017 ELN risk classification [16] can further improve the prognostic prediction. More investigations are needed in this field to give a clearer view of the clinical implications of lncRNA expression in AML. Furthermore, since the demographics and disease natures in AML patients are somewhat different between Asian and Western countries [17–19], it is unclear whether this new risk model combining both clinical parameters and molecular data is similarly useful for risk stratification of AML patients in Asia.

In this study, we profiled the lncRNA expression to investigate the clinical, biological, and prognostic implications of lncRNA expression in *de novo* AML patients. We constructed a succinct risk scoring system, which could independently predict the prognosis, irrespective of the risk based on the refined MRC cytogenetic or 2017 ELN classification. Further, the scoring system could well stratify the intermediate-risk group, defined by either classification, to two subgroups with distinct survival. To the best of our knowledge, this is the first study to show that incorporation of lncRNA expression in the 2017 ELN risk classification can improve stratification of intermediate-risk patients. Furthermore, we explored the differential mRNA expression between the two groups with high and low lncRNA scores to investigate the underlying mechanisms.

2. Materials and methods

2.1. Subjects

A total of 275 adult patients with newly diagnosed *de novo* non-M3 AML at the National Taiwan University Hospital (NTUH) were recruited. AML was diagnosed according to French-American-British (FAB) study group and 2016 World Health Organization (WHO) criteria [1]. Patients with antecedent hematological diseases, history of cytopenia, family history of myeloid neoplasms, or therapy-related AML were excluded. This study was approved by the Research Ethics Committee of the NTUH in accordance with the Declaration of Helsinki. The treatment protocol was described previously [20] and the choice of allogeneic hematopoietic stem cell transplantation (HSCT) was based on chromosomal findings, age, availability of donors and the response to induction treatment. We randomly divided the patients into the training ($n = 183$) and the validation cohorts ($n = 92$). The basic characteristics of the two subgroups were not significantly different regarding clinical features and treatment outcome (Supplementary Table S1). In addition, 151 AML patients with RNA-Seq data from The Cancer Genome Atlas (TCGA; https://tcga-data.nci.nih.gov/tcga/) were used as the external validation cohort [21].

2.2. Analyses of cytogenetics and genetic alterations

Cytogenetic analyses were performed and interpreted according to the International System for Human Cytogenetic Nomenclature [22]. Analyses of mutations in genes involving in activated signaling pathways, such as *FLT* [20], *NRAS* [23], *KRAS* [23], *JAK2* [23], *KIT* [20], and *PTPN11* [20]; transcription factors, such as *CEBPA* [24], *RUNX1* [17], and *GATA2* [25]; splicing factors, including *SRSF2*, *U2AF1*, and *SF3B1* [26]; epigenetic modifications, including *MLL*/PTD [20], *ASXL1* [27], *IDH1* [28], *IDH2* [28], *TET2* [29], and *DNMT3A* [30]; cohesin gene complex, including *STAG1*, *STAG2*, *SMC1A*, *SMC3*, and *RAD21* [31]; as well as *NPM1* [20], *WT1* [32], and *TP53* [33], were performed as previously described.

2.3. Microarray analysis of lncRNAs and the dataset processing

The global gene expressions were profiled with the Affymetrix GeneChip Human Transcriptome Array (HTA) 2.0 (Affymetrix, Santa Clara, CA, USA). Please refer to the online supplement for the technical detail.

The probes were then mapped with LNCipedia [34] (V4.1) and GENCODE [35] (Release 27) databases to identify lncRNA probes. There were 70,523 probes corresponding to at least 18,638 known coding genes and many other non-coding regions. Among them, 19,614 probes were mapped to lncRNAs.

The external validation dataset was downloaded from TCGA in May 2018. The RNA-Seq of the 151 AML patients from TCGA were determined on the Illumina platform, in whom 142 patients had overall survival (OS) data. The lncRNAs were extracted based on the ENSG identifiers. The expression levels of each lncRNA were normalized in the form of fragments per kilobase of transcript per million mapped reads (FPKM) and were transformed into Log2 scale.

2.4. The lncRNA risk score

We first conducted probe-level *Z*-transformation on the 19,614 lncRNAs in the training cohort, which made zero mean and unit

standard deviation of each lncRNA across the patients. We then used the univariate Cox proportional hazards regression model to analyze the association between the expression levels of the individual lncRNAs and OS and disease-free survival (DFS). The lncRNAs with prominent significance on both OS and DFS ($P < .001$) were selected for further multivariate Cox model to identify the lncRNAs whose expression levels could independently predict survival. The lncRNAs with significant association with OS in the multivariable test (multivariate Cox $P < .05$) were then selected to establish the lncRNA risk scoring system. The lncRNA risk score was calculated as $\text{Risk}(j) = \sum_{lncRNA_i \in componentlncRNAs} lncRNA_i(j) \cdot \beta_i$, where j denotes the patient accession number; lncRNA_i represents the normalized expression level of the lncRNA probe i after Z-transformation, and β_i represents the weight of the particular lncRNA probe i.

To validate this scoring system, a k-fold cross-validation process was applied. Furthermore, in order to test the superiority of this scoring system, we performed a 10,000-time random permutation test. For each iteration, five lncRNAs were randomly selected to construct a random scoring system and multivariate Cox regression model was used to test the significance of each randomly assigned lncRNA score. The empirical P value was calculated as the fraction of random scoring systems with better P value than the proposed one. The smaller the empirical P value, the greater the probability that the proposed lncRNA scoring system could outperform the other randomly assembled lncRNA combinations.

2.5. Bioinformatics approaches for biological inference of the five lncRNAs included in the scoring system

Please refer to the online supplement.

2.6. Statistical analysis

Please refer to the online supplement.

3. Results

3.1. The lncRNA risk score

We identified 92 out of the 19,614 lncRNA probes whose expression levels had significant association with OS ($P < .001$, Supplementary Table S2), and 28 probes associated with DFS ($P < .001$, Supplementary Table S3). We then included expression levels of the overlapped 16 lncRNAs in a multivariate Cox model to select the lncRNAs whose expression levels had independent predictive power on survival. We discovered that high expressions of TC03000901.hg.1 (Chr3; 168,619,733-168,639,784; ENST00000484765), TC05001739.hg.1 (Chr5; 124,731,072-124,731,673; ENST00000509010), and TC21000723.hg.1 (Chr21; 40,218,171-40,220,568; ENST00000416842) independently predicted poor OS ($P = .001$, 0.002, and 0.026, respectively), while that of TC04000372.hg.1 (Chr4; 70,047,818-70082484;

Table 1
Comparison of clinical characteristics between AML patients with higher and lower lncRNA risk scores in the training cohort.

Variables	Total ($n = 183$)	Lower score ($n = 91$)	Higher score ($n = 92$)	P value
Sex[a]				0.372
Male	105	49 (53.8%)	56 (60.9%)	
Female	78	42 (46.2%)	36 (39.1%)	
Age (year)[b]	57 (18.4–91.3)	52.7 (19.1–91.3)	61.5 (18.4–89.8)	0.020
Lab data[b]				
WBC (/μL)	29,170 (890–423,720)	25,110(1020–423,720)	34,155 (890–417,500)	0.582
Hb (g/dL)	8.1 (2.5–14.0)	8.1 (3.7–14.0)	8.1 (2.5–13.2)	0.621
Platelet (×1000/μL)	44 (5–396)	38 (6–396)	56.5 (5–361)	0.028
Blast (/μL)	13,439 (0–398,297)	10,938 (0–398,296)	15,016 (0–369,070)	0.863
LDH (U/L)	1062 (242–8116)	1120 (279–7400)	982 (242–8116)	0.419
FAB[a]				
M0	2 (1.1%)	0 (0%)	2 (2.2%)	0.497
M1	37 (20.2%)	24 (26.4%)	13 (14.1%)	0.044
M2	59 (32.2%)	38 (41.8%)	21 (22.8%)	0.007
M4	68 (37.2%)	25 (27.5%)	43 (46.7%)	0.009
M5	14 (7.7%)	2 (2.2%)	12 (13.0%)	0.010
M6	3 (1.5%)	2 (2.2%)	1 (1.1%)	0.621
2016 WHO classification[a]				
AML with t(8;21)(q22;q22.1)	16 (8.7%)	13 (14.3%)	3 (3.3%)	0.009
AML with inv.(16)(p13.1q22) or t(16;16)(p13.1;q22)	6 (3.3%)	4 (4.4%)	2 (2.2%)	0.444
AML with t(9;11)(p21.3;q23.3)	3 (1.6%)	0 (0.0%)	3 (3.3%)	0.246
AML with t(6;9)(p23;q34.1)	0 (0.0%)	0 (0.0%)	0 (0.0%)	>0.999
AML with (q21.3q26.2) or t(3;3)(q21.3;q26.2)	2 (1.1%)	1 (1.1%)	1 (1.1%)	>0.999
AML with mutated *NPM1*	59 (32.2%)	29 (31.9%)	30 (32.6%)	>0.999
AML with biallelic mutations of *CEBPA*	18 (9.8%)	17 (18.7%)	1 (1.1%)	<0.001
AML with myelodysplasia-related changes	19 (10.4%)	7 (7.7%)	12 (13.0%)	0.333
AML, NOS				
AML with minimal differentiation	1 (0%)	0 (0%)	1 (0.5%)	>0.999
AML without maturation	14 (7.7%)	5 (5.5%)	9 (9.8%)	0.405
AML with maturation	30 (16.4%)	11 (12.1%)	19 (20.7%)	0.162
Acute myelomonocytic leukemia	12 (6.6%)	3 (3.3%)	9 (9.8%)	0.133
Acute monoblastic/monocytic leukemia	2 (1.1%)	0 (0%)	2 (2.2%)	0.497
Pure erythroid leukemia	1 (0.5%)	1 (1.1%)	0 (0%)	0.497
Induction response[a,c]				
CR	82 (69.5%)	51 (78.5%)	31 (58.5%)	0.027
PR/Refractory	28 (23.7%)	11 (16.9%)	17 (32.1%)	0.081
Induction death	8 (6.8%)	3 (4.6%)	5 (9.4%)	0.466
Relapse	54 (65.9%)	30 (58.8%)	24 (77.4%)	0.098

Abbreviations: FAB, French-American-British classification; CR, complete remission; WHO, World Health Organization; NOS, not otherwise specified; PR, partial remission.
[a] number of patients (%).
[b] median (range).
[c] Totally 118 patients received standard chemotherapy, 65 in the low-score group and 53 in the high-score group.

ENST00000505646) and TC10002453.hg.1 (Chr10; 6,973,425-7,002,440; lnc-SFMBT2-4:1), predicted better OS ($P = .023$ and 0.007, respectively). The details of the five lncRNAs were listed in Supplementary Table S4. By incorporating the beta values as statistical weights, we established the lncRNA risk score = 0.313 * [TC03000901.hg.1] - 0.353 * [TC04000372.hg.1] + 0.26 * [TC05001739.hg.1] - 0.371 * [TC10002453.hg.1] + 0.295 * [TC21000723.hg.1]. We used the median of the lncRNA risk score as a cut-off to define the groups of patients with higher and lower scores. To validate this scoring system, we applied five-fold cross-validation and found that the *P* values derived from log-rank were statistically significant.

3.2. Patient characteristics

Among the 183 *de novo* AML patients in the training cohort, 91 had lower lncRNA risk scores and 92, higher scores. The patients with higher lncRNA risk scores were significantly older (median 61.5 *vs.* 52.7, $P = .020$), had higher platelet count at diagnosis (median 56,500 *vs.* 38,000/μL, $P = .028$) and a higher portion of FAB M4 (46.7% *vs.* 27.5%, $P = .009$) or M5 subtype (13.0% *vs.* 2.2%, $P = .010$) but a lower proportion of M1 (14.1% *vs.* 26.4%, $P = .044$) or M2 subtype (22.8% *vs.* 41.8%, $P = .007$). Based on the 2016 WHO Classification, the high-score group had a lower incidence of t(8;21)(q22;q22.1) (3.3% *vs.* 14.2%, $P = .009$) and biallelic mutations of *CEBPA* (1.1% *vs.* 18.7%, $P < .001$) (Table 1). Other clinical features, including gender distribution, white blood cell (WBC) count at diagnosis, hemoglobin level, peripheral blood blast percentage, and lactate dehydrogenase level were similar between the two groups.

3.3. Comparison of cytogenetic abnormalities and gene mutations between patients with higher and lower lncRNA scores

Chromosome data were available in 174 patients at diagnosis, including 88 with lower lncRNA risk scores and 86 with higher scores (Supplementary Table S6). Compared with patients with lower scores, those with higher lncRNA scores had less frequently favorable-risk cytogenetic changes (5.8% *vs.* 19.3%, $P = .011$) based on the refined MRC classification [3]. The distribution of other chromosomal abnormalities was similar between the two groups.

To investigate the difference in gene mutations between high-score and low-score patients, we performed a complete mutational screening of 26 myeloid malignancies related genes (Table 2). The high-score patients harbored significantly more gene mutations at diagnosis compared with the low-score ones (median 3.0 *vs.* 2.0, $P = .005$). The high-score patients had significantly more *FLT3*/TKD (13.0% *vs.* 3.3%, $P = .028$), *KRAS* (6.5% *vs.* 0%, $P = .029$), *RUNX1* (19.6% *vs.* 7.8%, $P = .030$), *DNMT3A* (30.4% *vs.* 13.3%, $P = .007$), and *SF3B1* mutations (6.5% *vs.* 0%, $P = .029$), but less *KIT* mutations (1.1% *vs.* 9.9%, $P = .009$) and $CEBPA^{\text{double-mutations}}$ (1.1% *vs.* 18.7%, $P < .001$) than the low-score ones. Other genetic alterations were not significantly different between these two groups. The distributions of molecular gene mutations in these two groups seem different (Supplementary Fig. S1). Collectively, the high-score patients had a higher frequency to harbor one or more adverse gene mutations (including *FLT3*/ITD, *MLL*/PTD, *WT1*, *RUNX1*, *ASXL1*, *DNMT3A*, splicing factor genes, and *TP53* mutations) than the low-score ones (81.5% *vs.* 47.3%, $P < .001$), and the difference retained significant while two or more such gene mutations were counted (48.9% *vs.* 22.0%, $P < .001$).

3.4. Prognostic impact of lncRNA scores on OS and DFS

Among the 118 patients receiving standard chemotherapy in the training cohort, 82 (69.5%) patients achieved a complete remission (CR). The high-score patients had a lower CR rate (58.5% *vs.* 78.5%, $P = .027$) and a trend of higher relapse rate (77.4% *vs.* 58.5%, $P = .098$) than the low-score ones (Table 1). With a median follow-up of 91.1 months (ranges, 1 to 171.5 months), the high-score patients had significantly poorer OS and DFS than the low-score ones (median, 13.1 months *vs.* not reached (NR), $P < .001$, Fig. 1A, and median, 1.4 *vs.* 18.0 months, $P < .001$, Fig. 1B, respectively). The findings could be validated in an independent internal cohort (median OS, 25.2 months *vs.* NR, $P = .004$, Fig. 1C, and median DFS, 6.3 *vs.* 23.7 months, $P = .007$, Fig. 1D) and in the external TCGA cohort (median OS 274 *vs.* 822 days, $P = .006$, Fig. 1E). The comparison of clinical parameters between the lower and higher score groups in the TCGA cohort was shown in Supplementary Table S5. The differences between groups were similar to those in our cohort. In the subgroup of cytogenetically normal AML (CN-AML) patients, the lncRNA scoring system still served as a prognosticator for OS and DFS (both $P < .001$, Fig. 1F and G).

Though the refined MRC classification is widely used for risk-stratification in AML patients (Supplementary Fig. S2A and S2B), many patients are in the intermediate-risk group. Among the 118 patients receiving standard chemotherapy, 112 patients (63 lncRNA low-score and 49 high-score patients) could be categorized by MRC classification. Introducing the lncRNA scoring system, the intermediate-risk patients could be further categorized into two subgroups with distinct prognosis (median OS, NR *vs.* 12.8 months, $P < .001$, Fig. 2A and median DFS, 29.7 *vs.* 1.4 months, $P < .001$, Fig. 2B). In the patients with intermediate-risk cytogenetics, those who had lower lncRNA scores had OS and DFS similar to those with favorable-risk cytogenetics ($P = .121$ and 0.126, respectively), while those with higher lncRNA scores had OS and DFS similar to those with unfavorable-risk cytogenetics ($P = .091$ and 0.056, respectively) (Fig. 2A and B). In our cohort, the 2017 ELN risk classification could separate patients with favorable-risk from those with intermediate- or unfavorable-risk, but the last two were indistinguishable (Supplementary Fig. S3A and S3B). The lncRNA scoring system could well stratify the 2017 ELN intermediate-risk patients (Fig. 2C and D); those with lower lncRNA scores had OS and DFS similar to the favorable-risk group ($P = .238$ and 0.506, respectively), while those with higher lncRNA scores had OS and DFS similar to

Table 2
Association of lncRNA scores with genetic alterations in the training cohort.[a]

Variables	Number of tested	No. of patients with alteration (%)			*P* value
		Mutated patients	Lower score (n = 91)	Higher score (n = 92)	
FLT3/ITD	183	43 (23.5%)	18 (19.8%)	25 (27.2%)	0.296
FLT3/TKD	183	15 (8.2%)	3 (3.3%)	12 (13.0%)	0.028
NRAS	183	32 (17.5%)	20 (22.0%)	12 (13.0%)	0.123
KRAS	183	6 (3.3%)	0 (0%)	6 (6.5%)	0.029
PTPN11	183	12 (6.6%)	5 (5.5%)	7 (7.6%)	0.767
KIT	183	10 (5.5%)	9 (9.9%)	1 (1.1%)	0.009
WTI	182	13 (7.1%)	3 (3.3%)	10 (11.0%)	0.081
NPM1	183	61 (33.3%)	31 (34.1%)	30 (32.6%)	0.876
CEBPA	183	27 (14.8%)	22 (24.2%)	5 (5.4%)	<0.001
$CEBPA^{\text{double-mutations}}$	183	18 (9.8%)	17 (18.7%)	1 (1.1%)	<0.001
RUNX1	182	25 (13.7%)	7 (7.8%)	18 (19.6%)	0.030
MLL/PTD	183	9 (4.9%)	3 (3.3%)	6 (6.5%)	0.497
ASXL1	182	30 (16.5%)	13 (14.4%)	17 (18.5%)	0.550
IDH1	183	9 (4.9%)	2 (2.2%)	7 (7.6%)	0.169
IDH2	183	30 (16.4%)	14 (15.4%)	16 (17.4%)	0.842
TET2	181	36 (19.9%)	17 (18.9%)	19 (20.9%)	0.853
DNMT3A	182	40 (22.0%)	12 (13.3%)	28 (30.4%)	0.007
TP53	183	10 (5.5%)	5 (5.5%)	5 (5.4%)	0.999
GATA2	182	11 (6.0%)	9 (9.9%)	2 (2.2%)	0.058
Splicing gene	180	32 (17.8%)	7 (7.9%)	25 (27.5%)	0.001
SF3B1	180	6 (3.3%)	0 (0%)	6 (6.6%)	0.029
SRSF2	180	17 (9.4%)	5 (5.6%)	12 (13.2%)	0.125
U2AF1	180	9 (5.0%)	2 (2.2%)	7 (7.7%)	0.169
Cohesin	178	16 (9.0%)	8 (9.0%)	8 (9.0%)	>0.999
STAG1	181	2 (1.1%)	1 (1.1%)	1 (1.1%)	>0.999
STAG2	183	6 (3.3%)	1 (1.1%)	5 (5.4%)	0.211
SMC1A	182	4 (2.2%)	3 (3.3%)	1 (1.1%)	0.365
RAD21	183	4 (2.2%)	3 (3.3%)	1 (1.1%)	0.368

[a] No patients harbored *JAK2* or *SMC3* mutations.

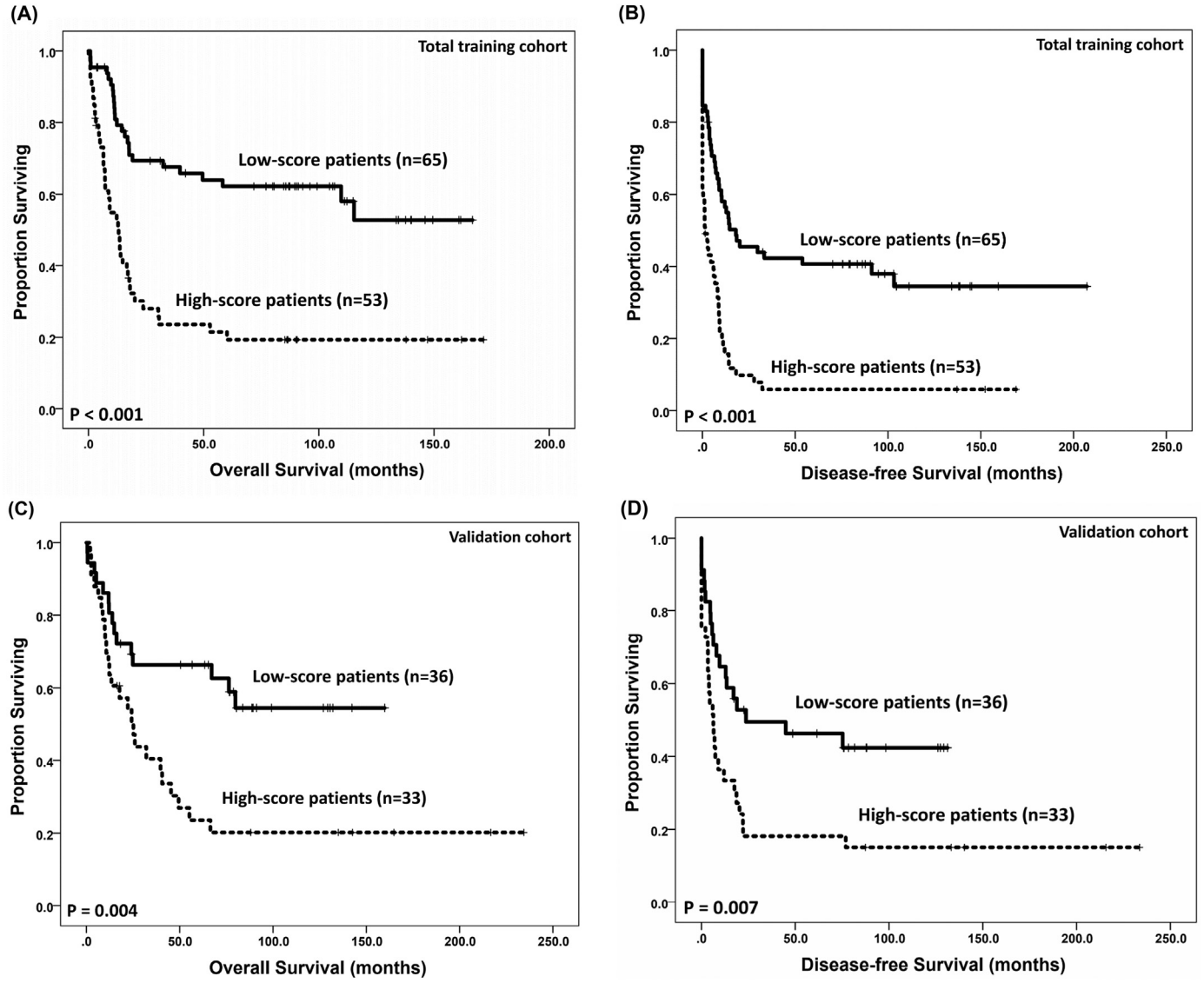


Fig. 1. The Kaplan-Meier survival curves for OS (A) and DFS (B) of AML patients stratified by lncRNA scores. The prognostic impact of the lncRNA scoring system could be validated in an independent internal cohort (C)(D), and the external cohort (E). Patients with normal cytogenetics could also be divided by lncRNA scores into two groups with distinct outcomes (F)(G).

the unfavorable-risk group ($P = .620$ and 0.377, respectively) (Fig. 2C and D).

In multivariate Cox proportional hazards regression analysis, we first included age [36], WBC counts at diagnosis [37], treatment modality and 2017 ELN risk classifications as the covariables (Table 3). Higher lncRNA score was an independent poor prognosticator, while ELN favorable-risk was a good prognostic factor for both OS and DFS. Receiving HSCT was an independent favorable factor for OS. Interestingly, among patients with higher lncRNA scores, HSCT did confer a better prognosis compared with post-remission chemotherapy alone (OS median 23.8 *vs.* 7.2 months, $P = .013$, Supplementary Fig. S4). Furthermore, allogeneic HSCT was also shown beneficial in the subgroup of ELN intermediate-risk patients with higher lncRNA scores (median OS, 30.6 *vs.* 4.6 months, $P < .001$) but not those with lower lncRNA scores (median OS, NR *vs.* 14.4 months, $P = .359$). We next applied cytogenetics and 11 gene mutations, including those that had significant survival impact in univariate analyses and those that were included in 2017 ELN risk classifications, as well as age and WBC counts as covariables into the multivariate analysis. Unfavorable-risk cytogenetics, higher lncRNA scores, and *DNMT3A* mutations were independent poor prognostic factors, while *NPM1*$^{+}$/*FLT3*-ITD^{-} was a favorable prognostic factor for both OS and DFS. HSCT and *CEBPA*$^{\text{double-mutations}}$ were independent favorable prognostic factors, but *TP53* mutations were independent poor prognostic factors for OS (Table 4). Importantly, the finding that the higher lncRNA score was an independent poor prognosticator for OS and DFS was confirmed in our validation cohort by multivariate analysis, which included variables with statistical significance in univariate analyses. (Supplementary Tables S7 and S8). In the 10,000-time permutation test, the same covariables as in Table 4 were chosen. The proposed scoring system surpassed almost all other randomly selected lncRNA combinations in multivariate OS prediction, with an empirical $P = 3 \times 10^{-4}$, suggesting the high performance and non-randomness of our scoring system.

3.5. Correlation of the lncRNA signature with gene expression to investigate potential underlying mechanisms

We profiled genome-wide RNA expression (the 70,523 probes corresponding to at least 18,638 known coding genes and many other non-coding regions) and compared the expression levels between patients with high and low lncRNA risk scores to get biological insight into the molecular mechanisms underlying the lncRNA signature. A

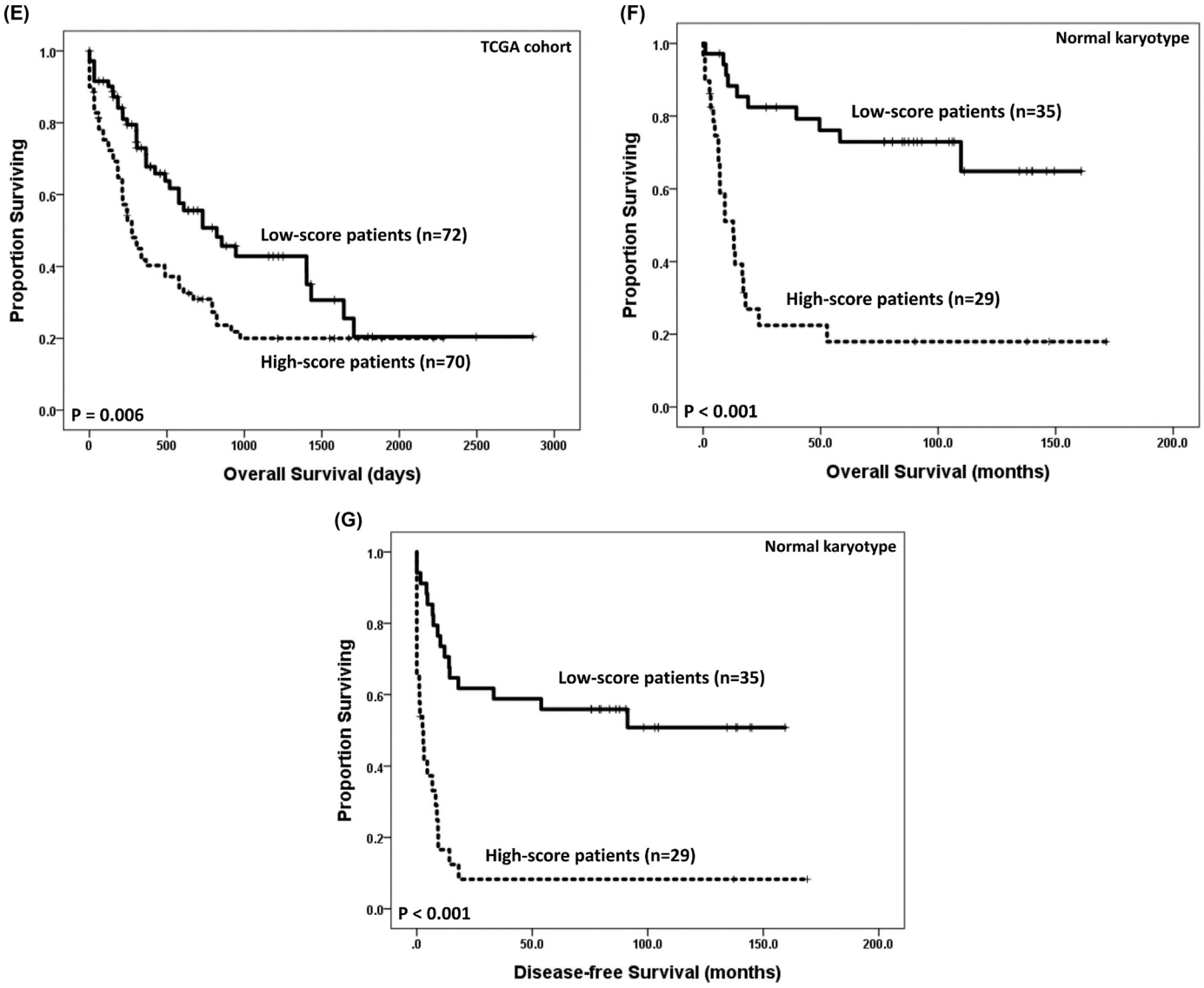


Fig. 1 (*continued*).

total of 595 differentially expressed genes were identified (>1.5-fold change and Student *t*-test $P < .0001$, Fig. 3A, B, and Supplementary Table S9). A network constructed by IPA showed *HOXA9*, *ERK1/2, NF-kB*, and *TRIM25* were the hub genes among the differentially expressed genes (Supplementary Fig. S5). Considering the roles of *HOX* gene family in determining stem cell fate, we further curated three published stem cell signatures [38–40]. The GSEA revealed significant up-regulation of leukemia stem cell and hematopoietic stem cell related genes in the high lncRNA score patients (Fig. 3C). We also made the differential gene expression analysis based on RNA-Seq data in the TCGA cohort, and the results were consistent (Supplementary Figs. 6 and 7).

4. Discussion

Besides the coding genes, aberrancy of non-coding genes are also involved in AML leukemogenesis and contribute to chemo-resistance, and may serve as prognosticators [5,6,9]. However, most studies regarding non-coding genes focused on miRNAs or certain subgroups of lncRNAs, and most importantly, the proposed scoring systems were mostly not further validated by the external cohort [13–15]. Garzon et al. from Cancer and Leukemia Group B (CALGB)/Alliance first established a lncRNA score composed of 48 lncRNAs as a prognosticator in elder (age ≧ 60 years) CN-AML patients [13]. Later on, the same group showed that a lncRNA score, made of 24 lncRNAs, had prognostic impacts on clinical outcome in younger (age < 60 years) CN-AML patients [15]. Because there was no overlap of lncRNAs between these two scoring systems, the mutual application may not be feasible in clinical practice. Beck et al. focused on long intervening/intergenic noncoding RNA (lincRNA), a subgroup of lncRNAs, instead of lncRNA we studied, and reported expression signatures composed of four out of 1664 lincRNAs could predict clinical outcomes in independent cohorts [41]. To our knowledge, this is the first study to elucidate the prognostic impacts of lncRNAs in *de novo* AML patients irrespective of various cytogenetic risks or gene mutation status. We developed a simple but concise scoring system composed of 5 lncRNAs, by analyzing expression levels of 19,614 lncRNA, and ensured the significant and independent prognostic implication through well-organized statistical modeling process. The reasons that the proposed lncRNAs in our scoring system were different from those in other reports could partly be explained by differences in ethnicity, age and treatment regimens.

Although there are several diverse risk-stratification models, most AML patients would be classified into an intermediate-risk group [4].

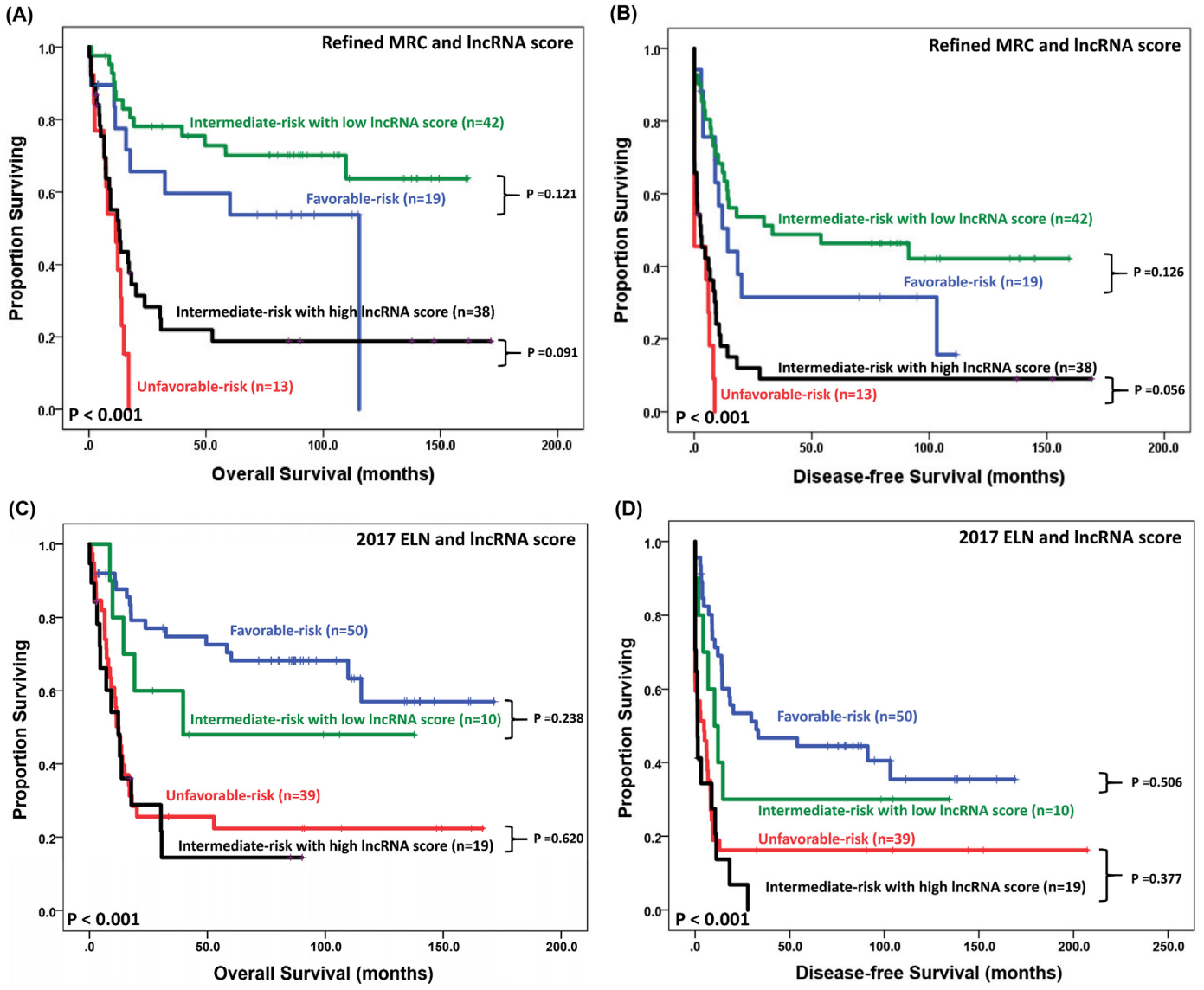


Fig. 2. The patients with MRC intermediate-risk cytogenetics could be well risk-stratified by lncRNA scoring system (A) (B). For the 2017 ELN intermediate-risk patients, the lncRNA score still served as a good prognosticator (C)(D).

The 2017 ELN recommendations incorporate cytogenetic changes and gene mutation status, including *FLT3*-ITD allelic ratio, to propose a refined stratification model which largely enhances the stratification power compared with 2010 ELN recommendations and refined MRC classification [42]. Nonetheless, the survival of patients is still heterogeneous even in the same intermediate-risk group. The lncRNA scoring system shown in this study can further dissect this heterogeneous group and help guide the choice of treatment strategies; low-risk patients can be treated less intensively, while high-risk patients, more aggressively. Truly, we showed that HSCT did confer a benefit to the ELN intermediate-risk patients with higher lncRNA scores, but not those with lower scores. Analyses of the clinical significance of lncRNA in AML patients in previous studies were mostly based on the 2010 ELN recommendations [13,41,43], or did not incorporate 2017 ELN recommendations into the multivariate analysis [15]. To the best of our knowledge, this study was the first to show that a lncRNA scoring system was an independent prognostic factor for both OS and DFS, irrespective of the risk factors based on the 2017 ELN classification.

Spurred by the advancement of genomic and transcriptomic technologies, more and more molecular markers, including gene mutations, mRNA expression, non-coding RNA expression, *etc.*, were incorporated into risk-stratification models [19,41,43–47]. Previous studies focusing on the prognostic implications of lncRNA in AML exploited 12–48 lncRNA expression levels. However, this is not easy to put these comprehensive markers into clinical practice. The lncRNA scoring system in this study is composed of only 5 lncRNAs with significant prognostic impact, which makes it handy to use in clinical practice.

It is pertinent that the proposal of a new prognostic scoring system should merit further validation. In order to verify the prognostic impact of this lncRNA scoring system, we used both an internal and an external cohort. The total cohort in our study was randomly split to the independent training and validation cohorts, and the two cohorts were comparable in terms of clinical features and treatment outcome. The prognostic impact of the lncRNA scoring system was verified in this internal validation cohort, even in the multivariate analysis incorporating the 2017 ELN risk classifications or various kinds of cytogenetic changes and gene mutations. Most importantly, we chose the TCGA-AML database as the external validation cohort. The TCGA cohort has multi-omics information, facilitating the possibility of the validation process. Although we could only find the expression levels in three out of five lncRNAs included in this lncRNA scoring system in the TCGA cohort, the lncRNA scoring system could well stratify patients into subgroups

Table 3
Multivariate Cox proportional hazards regression analysis using the 2017 ELN risk classification as a covariable on the disease-free survival and overall survival in the training cohort.

Variables	Disease-free Survival				Overall Survival			
		95% CI				95% CI		
	RR	Lower	Upper	*P* value	RR	Lower	Upper	*P* value
Total cohort ($n = 118$)								
Age[a]	1.009	0.993	1.026	0.264	1.003	0.985	1.022	0.723
WBC counts (k/μL)[a]	1.000	0.997	1.002	0.923	1.000	0.997	1.003	0.819
HSCT[b]	1.200	0.734	1.962	0.468	0.534	0.303	0.941	0.030
ELN Favorable[c]	0.566	0.344	0.931	0.025	0.362	0.195	0.672	0.001
lncRNA score[d]	2.326	1.462	3.701	<0.001	2.296	1.338	3.939	0.003

Abbreviation: RR, relative risk; CI, confidence interval; HSCT, hematopoietic stem cell transplantation; WBC, white blood cell.
[a] Continuous variables.
[b] HSCT *vs.* without HSCT.
[c] 2017 ELN Favorable-risk *vs.* Others. Regarding patients without cytogenetic data, they were classified according to gene mutation status. The two patients within the favorable-risk group both had *NPM1* mutations; the four patients within the unfavorable-risk group had *RUNX1*, *RUNX1*, *ASXL1*, and *FLT3*-ITD^{high}, respectively; the three patients within the intermediate-risk group had *NPM1* mutations with *FLT3*-ITD^{high}, *FLT3*-ITD^{low}, and *NPM1* mutations with *FLT3*-ITD^{high}, respectively.
[d] High-score *vs* low-score.

with significantly distinct prognosis. Because the RNA-Seq data in TCGA is unstranded, it is noteworthy that the quantification of certain lncRNA may be confounded by its natural antisense transcript, which is complementary and overlapping with the interested target. Regarding the five components in the proposed lncRNA scoring system, only TC04000372.hg.1 (ENST00000505646) has a natural antisense transcript, ENST00000446444 (product of gene UGT2B11), which might introduce bias in the external validation analysis.

In order to put the lncRNAs scoring system into clinical practice, a standardized and accessible procedure is mandatory. Wang et al. validated six prognostic models in AML and described if the model's inputs are dichotomized by the median value rather than a fixed given value, it would be easier to implement in data from varied platforms irrespective of the original distribution of the data [4]. Our lncRNA scoring system using the median as the cut-off value may further pave the way for its clinical application. Furthermore, it has been reported that RNA-Seq can provide reproducible results compared with quantitative PCR [48–52]. The performance of the lncRNA scoring system in TCGA cohort suggests that our scoring system would work in another cohort even using quantitative PCR. It is feasible to use either microarray, RNA-Seq, or quantitative PCR data to implement the lncRNA scoring system.

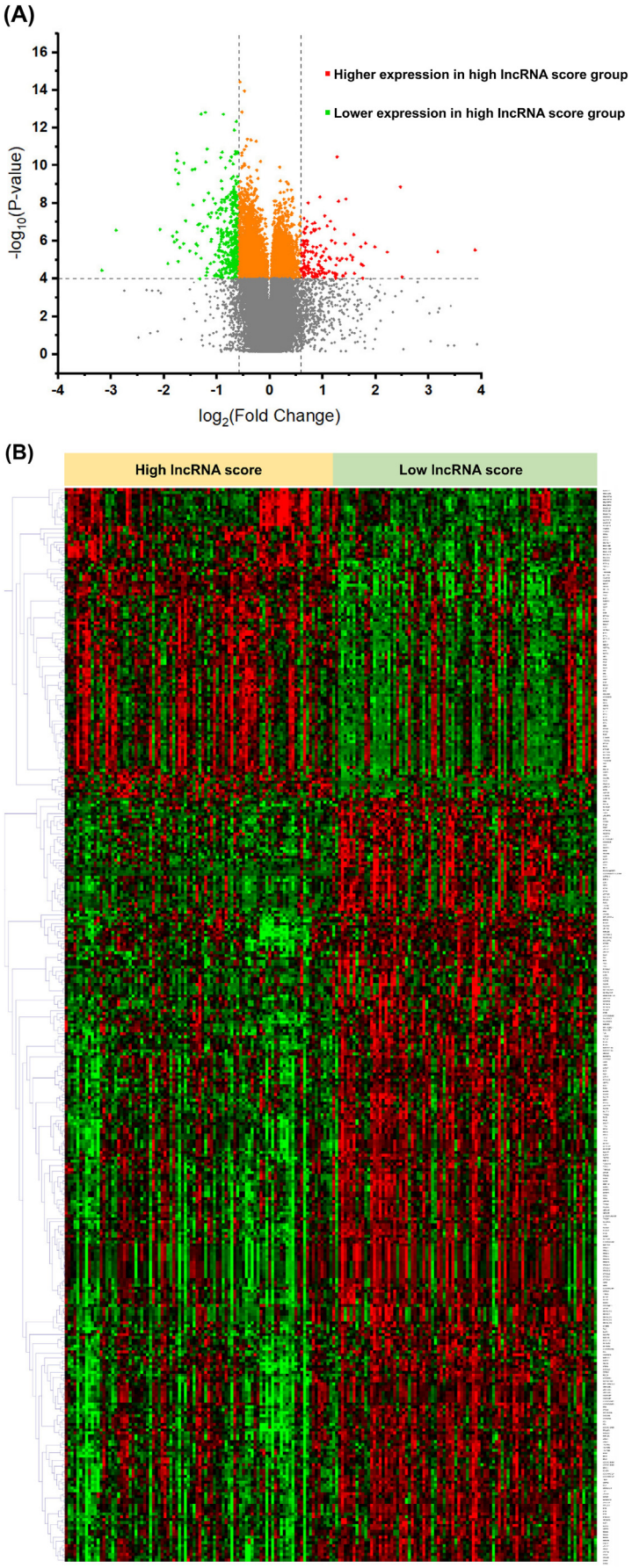
In this study, the higher lncRNA-score patients had significantly poorer OS and DFS, resulting from a lower CR rate and trend of higher relapse rate. Higher lncRNA scores were significantly associated with poor-risk molecular alterations, such as *RUNX1* [17], *DNMT3A* [30], and splicing gene mutations [26], but negatively associated with good-risk genetic alterations, such as favorable-risk cytogenetics and *CEBPA*$^{double\text{-}mutations}$. Further, the high-score patients harbored significantly more adverse gene mutations than the low-score patients did. Nevertheless, in multivariate analysis, higher lncRNA score was a poor prognostic factor independent from the clinical features, cytogenetic risks, gene mutations, and 2017 ELN risk classification. We also found allogeneic HSCT could significantly improve the survival of high-score patients. It implied that HSCT might ameliorate the poor survival impact of the adverse-risk lncRNA signatures. Further prospective studies with more patients recruited are needed to verify this point.

Papaioannou et al. correlated biological function with lncRNA scoring system established in younger AML patients and disclosed enrichment of genes involved in lymphocyte/leukocyte activation, inflammation, and apoptosis in patients with higher lncRNA scores [15]. No biological function analysis of lncRNA or lincRNA involved in scoring systems was performed in other reports in literature. In this study, the mechanistic investigation suggested that the high lncRNA score was strongly associated with hematopoietic and leukemic stem cells signatures, including those involving *HOX* genes, which might not only contribute to leukemogenesis but also lead to chemoresistance [53,54], and confer to the poor outcome. Similarly, the relationships of lncRNAs with cancer stem cell and drug resistance were also reported in solid cancers [55].

Table 4
Multivariate Cox proportional hazards regression analysis using genetic alterations as covariables on the disease-free survival and overall survival in the training cohort.

Variables	Disease-free Survival				Overall Survival			
		95% CI				95% CI		
	RR	Lower	Upper	*P* value	RR	Lower	Upper	*P* value
Total cohort (n = 118)								
Age[a]	1.013	0.995	1.030	0.164	1.001	0.981	1.022	0.890
WBC counts (k/μL)[a]	1.001	0.998	1.004	0.419	1.003	0.999	1.006	0.121
Karyotype[b]	2.687	1.312	5.507	0.007	3.125	1.432	6.819	0.004
HSCT[c]	1.009	0.590	1.725	0.973	0.349	0.182	0.669	0.002
NPM1^{+}/*FLT3*-ITD^{-}[d]	0.281	0.114	0.694	0.006	0.253	0.083	0.768	0.015
CEBPA$^{double\text{-}mutations}$ [e]	0.539	0.238	1.218	0.137	0.150	0.034	0.654	0.012
WT1[e]	0.993	0.480	2.055	0.985	1.824	0.811	4.104	0.146
Splicing genes[e,f]	1.775	0.778	4.049	0.173	2.089	0.743	5.877	0.163
DNMT3A[e]	1.877	1.021	3.451	0.043	2.142	1.024	4.481	0.043
RUNX1[e]	0.865	0.337	2.224	0.764	0.932	0.299	2.909	0.904
ASXL1[e]	0.534	0.248	1.148	0.108	0.488	0.187	1.276	0.144
TP53[e]	1.363	0.384	4.840	0.632	5.143	1.345	19.664	0.017
lncRNA score[g]	2.015	1.199	3.387	0.008	2.174	1.191	3.969	0.011

Abbreviation: RR, relative risk; CI, confidence interval, HSCT, hematopoietic stem cell transplantation; WBC, white blood cell.
[a] Continuous variables.
[b] Unfavorable cytogenetics *vs.* others.
[c] HSCT *vs* without HSCT.
[d] *NPM1*$^{+}$/*FLT3*-ITD^{-} *vs.* other genotypes.
[e] Mutated *vs* wild type.
[f] Including *SF3B1*, *SRSF2*, and *U2AF1*.
[g] High-score *vs.* low-score.


(A)
Higher expression in high lncRNA score group
Lower expression in high lncRNA score group
-log10(P-value)
log2(Fold Change)
(B)
High lncRNA score
Low lncRNA score

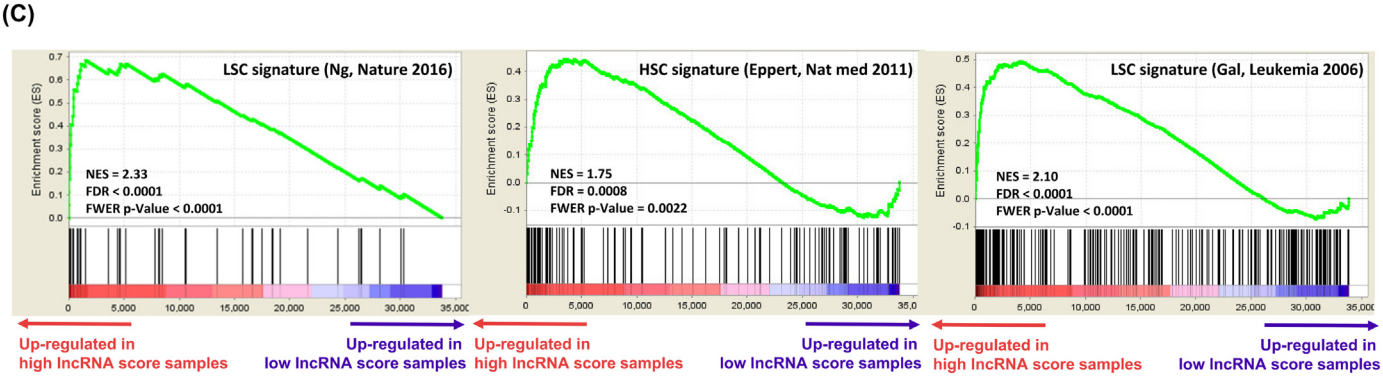


Fig. 3. Comparison of the genome-wide RNA expressions between the patients with higher and lower lncRNA scores. (A) The x-axis specifies the fold-changes (FC) and the y-axis specifies the negative logarithm to the base 10 of the *t*-test *P* values. The vertical and horizontal lines reflect the filtering criteria (FC = ±1.5 and P value = .0001). Red and green dots represented probe sets for transcripts expressed at significantly higher ($n = 174$) or lower ($n = 421$) levels in high-score patients, respectively. (B) Heatmap of the selected differential expressed genes. (C) GSEA confirmed that leukemia or hematopoietic stem cell signature containing genesets were enriched in higher lncRNA score patients.

5. Conclusion

We proposed an easy-to-use but concise lncRNA score composed of five lncRNAs. High lncRNA score was associated with distinct clinical features and gene mutation profiles, and was an independent poor prognostic factor. Incorporating lncRNA score into risk assessment could further refine the 2017 ELN or refined MRC risk classification by sub-dividing intermediate-risk patients. Further prospective large-scale studies are warranted to confirm our findings.

Acknowledgments

We would like to acknowledge the service provided by the DNA Sequencing Core of the First Core Laboratory, National Taiwan University College of Medicine.

Funding

MOST 103-2628-B-002-008-MY3, 104-2314-B-002-128-MY4, and 106-2314-B-002-226-MY3 from the Ministry of Science and Technology (Taiwan), MOHW 107-TDU-B-211-114,009 from the Ministry of Health and Welfare (Taiwan), NTUH 102P06, from the Department of Medical Research, National Taiwan University Hospital, and Taiwan Health Foundation.

Declaration of interests

The authors declare no competing interests.

Authors' contributions

C.-H.T. was responsible for data management and interpretation, statistical analysis, bioinformatic analysis, and manuscript writing; Y.-Y.K. and L.I.L were responsible for mutation analysis and interpretation; C.-Y.Y., F-M.T., J.-L.T., C.-C.L., M.-Y., W.-C.C. contributed patient samples and clinical data; M.-H.T., Y.-L.P. M.-C.L, and C.-W.L performed the gene mutation and cytogenetic studies. Y.-C.C. and C.-Y.C. contributed to the bioinformatic analysis. H.-A.H. and H.-F.T. designed, planned and coordinated the study over the entire period and wrote the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.01.022.

References

[1] Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016;127(20):2391–405.
[2] Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood 2010;115(3):453–74.
[3] Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood 2010;116(3):354–65.
[4] Wang M, Lindberg J, Klevebring D, Nilsson C, Mer AS, Rantalainen M, et al. Validation of risk stratification models in acute myeloid leukemia using sequencing-based molecular profiling. Leukemia 2017;31(10):2029–36.
[5] Garzon R, Volinia S, Liu CG, Fernandez-Cymering C, Palumbo T, Pichiorri F, et al. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. Blood 2008;111(6):3183–9.
[6] Marcucci G, Radmacher MD, Maharry K, Mrozek K, Ruppert AS, Paschka P, et al. MicroRNA expression in cytogenetically normal acute myeloid leukemia. N Engl J Med 2008;358(18):1919–28.
[7] Schwind S, Maharry K, Radmacher MD, Mrozek K, Holland KB, Margeson D, et al. Prognostic significance of expression of a single microRNA, miR-181a, in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol 2010;28(36):5257–64.
[8] Marcucci G, Maharry KS, Metzeler KH, Volinia S, Wu YZ, Mrozek K, et al. Clinical role of microRNAs in cytogenetically normal acute myeloid leukemia: miR-155 upregulation independently identifies high-risk patients. J Clin Oncol 2013;31(17):2086–93.
[9] Chuang MK, Chiu YC, Chou WC, Hou HA, Chuang EY, Tien HF. A 3-microRNA scoring system for prognostication in de novo acute myeloid leukemia patients. Leukemia 2015;29(5):1051–9.
[10] Kitagawa M, Kitagawa K, Kotake Y, Niida H, Ohhata T. Cell cycle regulation by long non-coding RNAs. Cell Mol Life Sci 2013;70(24):4785–94.
[11] Hu W, Yuan B, Flygare J, Lodish HF. Long noncoding RNA-mediated anti-apoptotic activity in murine erythroid terminal differentiation. Genes Dev 2011;25(24):2573–8.
[12] Geisler S, Coller J. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. Nat Rev Mol Cell Biol 2013;14(11):699–712.
[13] Garzon R, Volinia S, Papaioannou D, Nicolet D, Kohlschmidt J, Yan PS, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci U S A 2014;111(52):18679–84.
[14] De Clara E, Gourvest M, Ma H, Vergez F, Tosolini M, Dejean S, et al. Long non-coding RNA expression profile in cytogenetically normal acute myeloid leukemia identifies a distinct signature and a new biomarker in NPM1-mutated patients. Haematologica 2017;102(10):1718–26.
[15] Papaioannou D, Nicolet D, Volinia S, Mrozek K, Yan P, Bundschuh R, et al. Prognostic and biologic significance of long non-coding RNA profiling in younger adults with cytogenetically normal acute myeloid leukemia. Haematologica 2017;102(8):1391–400.
[16] Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017;129(4):424–47.
[17] Tang JL, Hou HA, Chen CY, Liu CY, Chou WC, Tseng MH, et al. AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. Blood 2009;114(26):5352–61.

[18] Kihara R, Nagata Y, Kiyoi H, Kato T, Yamamoto E, Suzuki K, et al. Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. Leukemia 2014;28(8):1586–95.
[19] Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016;374(23):2209–21.
[20] Hou HA, Lin CC, Chou WC, Liu CY, Chen CY, Tang JL, et al. Integration of cytogenetic and molecular alterations in risk stratification of 318 patients with de novo non-M3 acute myeloid leukemia. Leukemia 2014;28(1):50–8.
[21] Cancer Genome Atlas Research Network, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368(22):2059–74.
[22] Hou HA, Tsai CH, Lin CC, Chou WC, Kuo YY, Liu CY, et al. Incorporation of mutations in five genes in the revised International Prognostic Scoring System can improve risk stratification in the patients with myelodysplastic syndrome. Blood Cancer J 2018;8(4):39.
[23] Hou HA, Kuo YY, Tang JL, Chou WC, Yao M, Lai YJ, et al. Clinical implications of the SETBP1 mutation in patients with primary myelodysplastic syndrome and its stability during disease progression. Am J Hematol 2014;89(2):181–6.
[24] Lin LI, Chen CY, Lin DT, Tsay W, Tang JL, Yeh YC, et al. Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin Cancer Res 2005;11(4):1372–9.
[25] Tien FM, Hou HA, Tsai CH, Tang JL, Chiu YC, Chen CY, et al. GATA2 zinc finger 1 mutations are associated with distinct clinico-biological features and outcomes different from GATA2 zinc finger 2 mutations in adult acute myeloid leukemia. Blood Cancer J 2018;8(9):87.
[26] Hou HA, Liu CY, Kuo YY, Chou WC, Tsai CH, Lin CC, et al. Splicing factor mutations predict poor prognosis in patients with de novo acute myeloid leukemia. Oncotarget 2016;7(8):9084–101.
[27] Chen TC, Hou HA, Chou WC, Tang JL, Kuo YY, Chen CY, et al. Dynamics of ASXL1 mutation and other associated genetic alterations during disease progression in patients with primary myelodysplastic syndrome. Blood Cancer J 2014;4:e177.
[28] Lin CC, Hou HA, Chou WC, Kuo YY, Liu CY, Chen CY, et al. IDH mutations are closely associated with mutations of DNMT3A, ASXL1 and SRSF2 in patients with myelodysplastic syndromes and are stable during disease evolution. Am J Hematol 2014;89(2):137–44.
[29] Chou WC, Chou SC, Liu CY, Chen CY, Hou HA, Kuo YY, et al. TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. Blood 2011;118(14):3803–10.
[30] Hou HA, Kuo YY, Liu CY, Chou WC, Lee MC, Chen CY, et al. DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. Blood 2012;119(2):559–68.
[31] Tsai CH, Hou HA, Tang JL, Kuo YY, Chiu YC, Lin CC, et al. Prognostic impacts and dynamic changes of cohesin complex gene mutations in de novo acute myeloid leukemia. Blood Cancer J 2017;7(12):663.
[32] Hou HA, Huang TC, Lin LI, Liu CY, Chen CY, Chou WC, et al. WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system. Blood 2010;115(25):5222–31.
[33] Hou HA, Chou WC, Kuo YY, Liu CY, Lin LI, Tseng MH, et al. TP53 mutations in de novo acute myeloid leukemia patients: longitudinal follow-ups show the mutation is stable during disease evolution. Blood Cancer J 2015;5:e331.
[34] Volders PJ, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res 2015;43(Database issue):D174–80.
[35] Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for the ENCODE project. Genome Res 2012;22(9):1760–74.
[36] Tsai CH, Hou HA, Tang JL, Liu CY, Lin CC, Chou WC, et al. Genetic alterations and their clinical implications in older patients with acute myeloid leukemia. Leukemia 2016;30(7):1485–92.
[37] Tien FM, Hou HA, Tsai CH, Tang JL, Chen CY, Kuo YY, et al. Hyperleukocytosis is associated with distinct genetic alterations and is an independent poor-risk factor in de novo acute myeloid leukemia patients. Eur J Haematol 2018;101(1):86–94.
[38] Ng SW, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature 2016;540(7633):433–7.
[39] Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. Nat Med 2011;17(9):1086–93.
[40] Gal H, Amariglio N, Trakhtenbrot L, Jacob-Hirsh J, Margalit O, Avigdor A, et al. Gene expression profiles of AML derived stem cells; similarity to hematopoietic stem cells. Leukemia 2006;20(12):2147–54.
[41] Beck D, Thoms JAI, Palu C, Herold T, Shah A, Olivier J, et al. A four-gene LincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. Leukemia 2018;32(2):263–72.
[42] Harada Y, Nagata Y, Kihara R, Ishikawa Y, Asou N, Ohtake S, et al. Prognostic analysis according to the 2017 ELN risk stratification by genetics in adult acute myeloid leukemia patients treated in the Japan Adult Leukemia Study Group (JALSG) AML201 study. Leuk Res 2018;66:20–7.
[43] Mer AS, Lindberg J, Nilsson C, Klevebring D, Wang M, Gronberg H, et al. Expression levels of long non-coding RNAs are prognostic for AML outcome. J Hematol Oncol 2018;11(1):52.
[44] Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med 2012;366(12):1079–89.
[45] Li Z, Herold T, He C, Valk PJ, 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 2013;31(9):1172–81.
[46] Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland MC, et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood 2008;112(10):4193–201.
[47] Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004;350(16):1605–16.
[48] Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science (New York, NY) 2008;320(5881):1344–9.
[49] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10(1):57–63.
[50] Camarena L, Bruno V, Euskirchen G, Poggio S, Snyder M. Molecular mechanisms of ethanol-induced pathogenesis revealed by RNA-sequencing. PLoS Pathog 2010;6(4):e1000834.
[51] Fang Z, Cui X. Design and validation issues in RNA-seq experiments. Brief Bioinform 2011;12(3):280–7.
[52] Feng L, Liu H, Liu Y, Lu Z, Guo G, Guo S, et al. Power of deep sequencing and agilent microarray for gene expression profiling study. Mol Biotechnol 2010;45(2):101–10.
[53] Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nat Rev Cancer 2005;5(4):275–84.
[54] Ho TC, LaMere M, Stevens BM, Ashton JM, Myers JR, O'Dwyer KM, et al. Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. Blood 2016;128(13):1671–8.
[55] Heery R, Finn SP, Cuffe S, Gray SG. Long non-coding RNAs: key regulators of epithelial-mesenchymal transition, tumour drug resistance and cancer stem cells. Cancer 2017;9(4).