blood advances

A 4-lncRNA scoring system for prognostication of adult myelodysplastic syndromes

Chi-Yuan Yao,[1-3,*] Ching-Hsuan Chen,[4,*] Huai-Hsuan Huang,[1] Hsin-An Hou,[1] Chien-Chin Lin,[1-3] Mei-Hsuan Tseng,[1] Chein-Jun Kao,[1] Tzu-Pin Lu,[4] Wen-Chien Chou,[1,2] and Hwei-Fang Tien[1]

[1]Division of Hematology, Department of Internal Medicine, and [2]Department of Laboratory Medicine, National Taiwan University Hospital, Taipei, Taiwan; and [3]Graduate Institute of Clinical Medicine, College of Medicine, and [4]Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan

Key Points

- Through lncRNA profiling, we identified an MDS patient subset with distinct clinical and mutational patterns along with inferior outcomes.

- A concise yet powerful 4-lncRNA risk-scoring system was devised with the potential to improve current MDS risk stratification.

Long noncoding RNAs (lncRNAs) not only participate in normal hematopoiesis but also contribute to the pathogenesis of acute leukemia. However, their clinical and prognostic relevance in myelodysplastic syndromes (MDSs) remains unclear to date. In this study, we profiled lncRNA expressions in 176 adult patients with primary MDS, and identified 4 lncRNAs whose expression levels were significantly associated with overall survival (OS). We then constructed a risk-scoring system with the weighted sum of these 4 lncRNAs. Higher lncRNA scores were associated with higher marrow blast percentages, higher-risk subtypes of MDSs (based on both the Revised International Prognostic Scoring System [IPSS-R] and World Health Organization classification), complex cytogenetic changes, and mutations in *RUNX1*, *ASXL1*, *TP53*, *SRSF2*, and *ZRSR2*, whereas they were inversely correlated with *SF3B1* mutation. Patients with higher lncRNA scores had a significantly shorter OS and a higher 5-year leukemic transformation rate compared with those with lower scores. The prognostic significance of our 4-lncRNA risk score could be validated in an independent MDS cohort. In multivariate analysis, higher lncRNA scores remained an independent unfavorable risk factor for OS (relative risk, 4.783; $P < .001$) irrespective of age, cytogenetics, IPSS-R, and gene mutations. To our knowledge, this is the first report to provide a lncRNA platform for risk stratification of MDS patients. In conclusion, our integrated 4-lncRNA risk-scoring system is correlated with distinctive clinical and biological features in MDS patients, and serves as an independent prognostic factor for survival and leukemic transformation. This concise yet powerful lncRNA-based scoring system holds the potential to improve the current risk stratification of MDS patients.

Introduction

Myelodysplastic syndromes (MDSs) comprise a heterogeneous group of clonal hematopoietic disorders, and are the most common class of acquired bone marrow (BM) failure syndromes in adults. MDSs are clinically characterized by BM dysplasia with ineffective hematopoiesis, peripheral blood cytopenia, and an increased risk of transformation to acute myeloid leukemia (AML) in ~30% to 40% of patients.[1,2] The clinical presentations and treatment outcomes of MDS patients may vary considerably,[3,4] underscoring the necessity of individualized management.[5-8]

Although cytogenetic abnormalities are detected in about half of MDS patients,[9-11] recurrent somatic mutations, most notably those involving RNA-splicing machinery and epigenetic regulators, can be identified in over 80% of them.[12] Not only have these coding gene mutations been linked to disease progression and prognosis in MDSs,[13-17] but dysregulation of noncoding RNAs, such as the microRNAs, have also been implicated as important elements in the pathogenesis and disease evolution of MDSs.[18-21] Recent research works have revealed that the long noncoding RNAs (lncRNAs), a novel class of versatile noncoding RNAs operationally defined as transcripts longer than 200 nucleotides, play crucial roles in a constellation of biological processes including chromatin modification, regulation of gene expression in various signaling pathways, genomic imprinting, and epigenetic control.[22-25] Although a myriad of previous publications have established the association between deranged lncRNA expression and carcinogenesis, tumor progression, and metastasis in a wide variety of solid cancers,[26-33] relatively limited data exist so far on how these lncRNAs are involved in normal and malignant hematopoiesis.[29,34-37]

In this study, we aimed to perform lncRNA expression profiling in a large cohort of primary MDS patients to investigate the link between lncRNA expression and the clinical and genetic features of MDS patients. A concise, integrated risk-scoring system was constructed by incorporating the expression levels of 4 lncRNAs that were significantly associated with survival, taking into account the statistical weight of each component lncRNA. To our knowledge, this is the first report that provides a lncRNA platform for risk stratification of MDS patients. Furthermore, by analyzing the messenger RNA (mRNA) expression profiles, we identified several pathways highly correlated with the lncRNA prognostic signature.

Materials and methods

Patients

We recruited 176 consecutive adult patients with primary MDS diagnosed between November 1992 and December 2010 at the National Taiwan University Hospital (NTUH) as the training cohort; we confirmed their diagnoses according to the 2008 World Health Organization (WHO) classification.[38] Patients with therapy-related MDSs or antecedent hematological diseases were excluded.

Another independent set of 30 patients, who were diagnosed with MDSs with these criteria between January 2011 and May 2012 at NTUH, were collected as the validation cohort. This study was approved by the institutional review board of the NTUH. Informed consent was obtained from the patients in accordance with the Declaration of Helsinki. Cytogenetic analyses were performed as described previously,[39] and interpreted according to the International System for Human Cytogenetic Nomenclature.[40] Mutation analyses of 17 genes relevant to myeloid malignancies, such as genes related to epigenetic regulation (including ASXL1,[41] IDH1,[42] IDH2,[43] EZH2,[44] TET2,[45] and DNMT3A[46]), genes related to the RNA-splicing machinery (including SF3B1,[47] U2AF1,[48] SRSF2,[49] and ZRSR2[50]), as well as FLT3/ITD,[51] NRAS,[52] KRAS,[52] RUNX1,[53] MLL/PTD,[54] TP53,[55] and SETBP1,[52] were performed by Sanger sequencing as previously described. Abnormal sequencing results were confirmed by at least 2 repeated analyses.

Microarray analysis of lncRNA and data processing

We extracted RNA by the TRIzol method from the BM mononuclear cells obtained at diagnosis. One microgram of RNA from each sample was processed with the miRNeasy mini kit (Qiagen, Hilden, Germany); the global gene expressions of the 176 MDS patients in the training cohort, and the 30 MDS patients in the validation cohort, were profiled with Affymetrix GeneChip Human Transcriptome Array 2.0 (Affymetrix, Santa Clara, CA), following the manufacturer's instructions.

We used the robust multichip average algorithm for microarray preprocessing and normalization. We then mapped the probes with the most updated version of the LNCipedia[56] (V4.0) and GENCODE[57] (release 25) databases (as in September 2016); 19 614 lncRNA probes were identified in the Affymetrix GeneChip Human Transcriptome Array 2.0 microarray.

Establishment of the lncRNA risk score

We first conducted probe-level Z transformation on the 19 614 lncRNAs across 176 MDS patients, which made zero mean and unit standard deviation (SD) of each lncRNA across the patients. We then used the univariate Cox proportional hazards regression model to analyze the association between overall survival (OS) and the expression levels of the individual lncRNAs. The lncRNAs with the most prominent significance on OS (univariate Cox $P < 1 \times 10^{-6}$) 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 multivariate test (multivariate Cox $P < .05$) were then selected to build the lncRNA risk-scoring system, in which the expression of component lncRNAs was subjected to a second multivariate Cox regression test to obtain the β values as their weights in the risk model equation. The lncRNA risk score was calculated as the summation of the normalized expression level of each component lncRNA multiplied by its weight: $\text{Risk}(j) = \sum_{\text{lncRNA}_i \in \text{component lncRNAs}} \text{lncRNA}_i(j) \cdot \beta_i$, 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. We also performed a 100 000-time random permutation test to ensure the strength of our risk model. For each iteration of the random permutation, the same number of lncRNAs was randomly selected from the 19 614 lncRNAs to construct a random scoring system. Then, the univariate Cox regression model was used for testing the prognostic significance of each randomly assigned lncRNA score. After 100 000 iterations, the empirical P value of our proposed risk model was calculated as the fraction of random scoring systems that achieved a better P value than our proposed risk model. The smaller the empirical P value, the greater probability that our proposed lncRNA risk score outperformed the other randomly assembled lncRNA combinations.

Bioinformatics approaches for biological inference of the 4 lncRNAs

To obtain insights into the functionality of the 4 lncRNAs, we first investigated the differentially expressed mRNAs between patients with the highest ($>$ average + 1 SD) and the lowest ($<$ average − 1 SD) lncRNA risk scores, and then analyzed the correlated biological pathways with the Database for Annotation, Visualization and

Integrated Discovery (DAVID; v6.8)[58,59] and Ingenuity Pathway Analysis (IPA) software (Ingenuity Systems, Redwood City, CA).[60]

Statistical analysis

The Mann-Whitney U test was used to compare continuous variables and medians of distribution. The χ^2 test or the Fisher exact test were used to compare the difference of discrete variables such as sex, WHO classification, cytogenetic changes, distribution of the Revised International Prognostic Scoring System (IPSS-R), and genetic alterations between patients with lower and higher lncRNA risk scores. OS was measured from the date of initial diagnosis to the date of last follow-up or death from any cause. Twofold cross-validation methodology was performed to validate the scoring system. In this procedure, we randomly divided our 176 MDS patients into 2 subgroups, half as the training set and the other half as the validation set. The training set was used to build a lncRNA risk-scoring system with the procedure stated in "Establishment of the lncRNA risk score." This lncRNA risk-scoring system was then applied to the validation set to stratify patients into high-score and low-score subgroups. The survival prediction power of this particular lncRNA risk score was then evaluated by both the log-rank test and the univariate Cox proportional hazards model. After 100 000 iterations, the prediction rate of our proposed lncRNA risk score was calculated as the fraction of random scoring systems that achieved $P < .05$. Kaplan-Meier analysis was used to plot the survival curves, and the log-rank test was used to evaluate the statistical significance. The Cox proportional hazards model was used for multivariate regression analysis. P values $<.05$ were considered statistically significant. All statistical analyses were performed with BRB-ArrayTools (version 4.5.1; Biometric Research Branch, National Cancer Institute, Rockville, MD),[61] IBM SPSS Statistics for Windows, version 20 (IBM Corp, Armonk, NY), and R software (https://cran.r-project.org/).

Results

Patient population

The median age of the 176 MDS patients was 67.8 years. The 164 patients who had karyotype data at diagnosis belonged to the following subgroups according to the IPSS-R: very-low risk, 3.7%; low risk, 32.9%; intermediate risk, 25.0%; high risk, 21.3%; and very-high risk, 17.1%. Most patients (68.8%) received supportive care only. Seventeen patients (9.7%) received hypomethylating agents (HMAs), 19 (10.8%) received low-dose cytarabine, 13 (7.4%) received AML-directed intensive chemotherapy either as a bridge to transplantation or at the time of leukemic transformation, and 19 (10.8%) underwent allogeneic hematopoietic stem cell transplantation. With a median follow-up duration of 37.3 months (range, 0.1-130.9 months), 83 mortalities were recorded, whereas 38 patients progressed to AML.

Computing the lncRNA risk score

To construct an lncRNA scoring system that is predictive of survival in MDS patients, we first identified 30 of the 19 614 lncRNA probes whose expression levels had significant association with OS ($P < 1 \times 10^{-6}$), as listed in supplemental Table 2. We then included expression levels of the 30 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 4 lncRNA probes independently predicted poor

OS ($P < .05$), namely TC07000551.hg.1, TC08000489.hg.1, TC02004770.hg.1, and TC03000701.hg.1, with multivariate Cox P values being $<.001$, $<.001$, .013, and .018, respectively (the details of the 4 lncRNAs are listed in supplemental Table 3). By incorporating the β values as statistical weights, we constructed the following equation: 4-lncRNA risk score = [TC07000551.hg.1] × 0.578 + [TC08000489.hg.1] × 0.526 + [TC02004770.hg.1] × 0.289 + [TC03000701.hg.1] ± 0.183. In a 100 000-time random permutation test, the 4-lncRNA risk score outperformed almost all of the other 100 000 random combinations of 4 lncRNAs in survival prediction, with an empirical $P < 10^{-5}$, suggesting high performance and nonrandomness of our proposed scoring system.

Comparison of clinical characteristics and genetic alterations between patients with high vs low 4-lncRNA risk scores

The 176 MDS patients in the training cohort were divided into 2 groups by the median lncRNA risk score. The comparison of clinical and laboratory features between patients with higher and lower scores is listed in Table 1. The patients with higher lncRNA scores had higher BM blast percentages ($P < .001$) compared with low-score patients. There were no significant differences in distribution of age and sex, peripheral blood white blood cell count, absolute neutrophil count, hemoglobin, and platelet level between the 2 groups. High-score patients more frequently had refractory anemia with excess blasts-1 (RAEB-1) and RAEB-2 (27.3% vs 9.1%, $P = .003$ and 43.2% vs 6.8%, $P < .001$, respectively), but less commonly refractory cytopenia with unilineage dysplasia (RCUD), refractory anemia with ring sideroblasts (RARS), refractory cytopenia with multilineage dysplasia (RCMD), and RCMD and ring sideroblasts (RCMD-RS) compared with the low-score patients (14.8% vs 33.0%, $P = .008$; 2.3% vs 12.5%, $P = .018$; 12.5% vs 28.4%, $P = .014$; and 0% vs 10.2%, $P = .003$, respectively). High-score patients were more frequently assigned to IPSS-R high- and very-high-risk subgroups (36.0% vs 5.1%, $P < .001$ and 26.7% vs 6.4%, $P = .001$, respectively), but less frequently to the IPSS-R low-risk subgroup (17.4% vs 50.0%, $P < .001$).

The incidence of cytogenetic abnormalities tended to be higher in patients with higher lncRNA risk scores than those with lower scores (46.5% vs 30.8%, $P = .054$). High-score patients had statistically significantly higher incidence of poor-risk cytogenetics (20.9% vs 9.0%, $P = .049$) and complex karyotypes (16.3% vs 5.1%, $P = .026$), but lower incidence of good-risk cytogenetics (58.1% vs 76.9%, $P = .013$) than low-score patients (supplemental Table 4).

Overall, 108 patients (61.4%) had at least 1 gene mutation, 42 (47.7%) in the low-score subgroup, compared with 66 (75.0%) in the high-score subgroup ($P < .001$). As listed in Table 2, the most common mutation in high-score patients was the *ASXL1* mutation (31.8%), followed by *RUNX1* (24.7%), *SRSF2* (20.7%), *DNMT3A* (19.5%), and *ZRSR2* mutations (16.3%). Compared with patients with lower lncRNA scores, those with higher scores more frequently harbored *RUNX1*, *ASXL1*, *TP53*, *SRSF2*, and *ZRSR2* mutations (24.7% vs 4.6%, $P < .001$; 31.8% vs 10.2%, $P = .001$; 12.8% vs 1.1%, $P = .002$; 20.7% vs 6.8%, $P = .009$; and 16.3% vs 1.2%, $P = .001$, respectively), but had lower incidence of *SF3B1* mutation (9.2% vs 23.9%, $P = .014$). We discovered distinctive patterns of concurrent gene mutations in MDS patients with either higher or

Table 1. Comparison of clinical and laboratory features between patients with high and low lncRNA scores

Clinical characters	Total, N = 176	Low lncRNA score, n = 88	High lncRNA score, n = 88	P
Sex (%)				.329
Male	121 (68.8)	57 (64.8)	64 (72.7)	
Female	55 (31.3)	31 (35.2)	24 (27.3)	
Age, median (range), y	67.8 (18.5-94.5)	68.7 (18.5-94.5)	68.7 (25.9-89.2)	.750
Laboratory data, median (range)				
WBC, ×10⁹/L	3.285 (0.490-20.440)	3.780 (1.490-10.880)	3.895 (0.490-2.040)	.970
ANC, ×10⁹/L	1.768 (0.103-12.728)	1.971 (0.258-6.953)	1.565 (0.103-12.728)	.395
Hb, g/dL	8.1 (3.5-14.6)	8.0 (3.5-13.1)	8.3 (4.0-14.6)	.409
Platelet, ×10⁹/L	86 (3-721)	104 (12-511)	77 (3-721)	.091
BM blast, %	3.0 (0-18.8)	1.5 (0-12.8)	8.2 (0-18.8)	<.001*
2008 WHO classification (%)				
RCUD	42 (23.9)	29 (33.0)	13 (14.8)	.008*
RARS	13 (7.4)	11 (12.5)	2 (2.3)	.018*
RCMD	36 (20.5)	25 (28.4)	11 (12.5)	.014*
RCMD-RS	9 (5.1)	9 (10.2)	0 (0)	.003*
RAEB1	32 (18.2)	8 (9.1)	24 (27.3)	.003*
RAEB2	44 (25)	6 (6.8)	38 (43.2)	<.001*
IPSS-R,†‡ (%)				
Very low	6 (3.7)	5 (6.4)	1 (1.2)	.103
Low	54 (32.9)	39 (50.0)	15 (17.4)	<.001*
Intermediate	41 (25.0)	25 (32.1)	16 (18.6)	.070
High	35 (21.3)	4 (5.1)	31 (36.0)	<.001*
Very high	28 (17.1)	5 (6.4)	23 (26.7)	.001*

ANC, absolute neutrophil count; Hb, hemoglobin; IPSS-R, Revised International Prognostic Scoring System; RAEB, refractory anemia with excess blasts; RARS, refractory anemia with ring sideroblasts; RCMD, refractory cytopenia with multilineage dysplasia; RCMD-RS, RCMD and ring sideroblasts; RCUD, refractory cytopenia with unilineage dysplasia; WBC, white blood cell.

*Statistically significant if $P < .05$.

†One hundred sixty-four patients, including 78 with low lncRNA scores and 86 with high lncRNA scores, had chromosome data at diagnosis.

‡IPSS-R: very low, ≤1.5; low, >1.5-3; intermediate, >3-4.5; high, >4.5-6; and very high, >6.

lower lncRNA risk scores (Figure 1). Patients with higher lncRNA scores had more complex interrelationships among the various mutations (Figure 1A), compared with those with lower lncRNA scores (Figure 1B). In addition, the mutation co-occurrences involving *SF3B1* and *TET2* were more frequent in patients with lower lncRNA scores (*P* = .019 and *P* = .015, respectively), whereas those involving *KRAS/NRAS*, *IDH1/2*, *SETBP1*, *TP53*, *MLL*/PTD, and *FLT3*/ITD were only found in patients with higher lncRNA scores.

Correlation of the 4-lncRNA risk score with survival and leukemic transformation

Patients with higher lncRNA scores had inferior OS than those with lower scores (median, 15.2 months vs not reached [NR], *P* < .001; Figure 2A). The cumulated incidence of transformation to AML at 5 years was 70.2% for higher-score patients and 12.5% for lower-score patients (*P* < .001; Figure 2B). The differences in outcomes remained significant in both IPSS-R lower-risk (very low, low, and intermediate risk) and IPSS-R higher-risk (high- and very-high-risk) subgroups (supplemental Figures 1 and 2). Likewise, when survival analysis was performed in the subgroup of patients with either a normal karyotype (N = 100), unfavorable cytogenetics (including complex karyotypes, monosomy 7, and del(7q); N = 24), WHO lower-

risk subtypes (RCUD, RARS, RCMD, and RCMD-RS; N = 100), or WHO higher-risk subtypes (RAEB-1 and RAEB-2; N = 76), patients with higher lncRNA scores consistently fared worse, and had a significantly higher incidence of acute leukemic transformation at 5 years than those with lower scores (supplemental Figures 3-6). By incorporating the lncRNA scoring system into the traditional IPSS-R classification, we could further refine the clinical outcome prediction in MDS patients (Figure 3), as we note that no matter which IPSS-R subgroup the patients had been originally assigned to, their clinical outcomes are more in line with the lncRNA score risk groups.

Furthermore, as the treatment strategies and options of MDS patients have evolved in the past 2 decades, with the advent of HMA being the most prominent advance, to test whether our lncRNA scoring system would continue to hold up in the populations managed with HMA, we also specifically looked into this selected subgroup in our cohort. Within the HMA-treated patient subset (N = 17), those patients with higher lncRNA scores did have significantly shorter median OS (supplemental Figure 7A; 17.3 months vs NR, *P* = .028), along with a marginally higher incidence of projected AML transformation at 5 years than those with lower scores (supplemental Figure 7B; 52.9% vs 16.7%, *P* = .129). A similar trend could also be observed in the patient subgroup that, beyond supportive care, had received active

Table 2. Comparison of genetic alterations between patients with low and high lncRNA scores

Gene	No. of specimens examined	Mutated (%)			
		Total	Low lncRNA score	High lncRNA score	P
SF3B1	175	29 (16.6)	21 (23.9)	8 (9.2)	.014*
FLT3/ITD	174	1 (0.6)	0 (0)	1 (1.2)	.494
NRAS/KRAS	175	5 (2.9)	1 (1.1)	4 (4.6)	.211
RUNX1	172	25 (14.5)	4 (4.6)	21 (24.7)	<.001*
MLL/PTD	173	1 (0.6)	0 (0)	1 (1.2)	.497
IDH1/2	175	4 (2.3)	0 (0)	4 (4.6)	.059
ASXL1	173	36 (20.8)	9 (10.2)	27 (31.8)	.001*
TET2	175	22 (12.6)	10 (11.4)	12 (13.8)	.655
DNMT3A	175	25 (14.3)	8 (9.1)	17 (19.5)	.054
TP53	174	12 (6.9)	1 (1.1)	11 (12.8)	.002*
SETBP1	175	4 (2.3)	0 (0)	4 (4.6)	.059
EZH2	176	10 (5.7)	4 (4.5)	6 (6.8)	.538
U2AF1	175	14 (8)	7 (8)	7 (8)	>.999
SRSF2	175	24 (13.7)	6 (6.8)	18 (20.7)	.009*
ZRSR2	171	15 (8.8)	1 (1.2)	14 (16.3)	.001*

*Statistically significant if $P < .05$.

treatment of their MDSs during disease course (supplemental Figure 8).

To validate our 4-lncRNA scoring system, we first applied the twofold cross-validation method and found that the P values derived from log-rank and univariate Cox regression models were both <.05. To further confirm the prognostic significance, we profiled the lncRNA expression of an independent validation cohort of 30 MDS patients with the same Affymetrix GeneChip HTA 2.0 platform. The clinical characteristics of patients in the validation cohort were comparable to those in the training cohort, as shown in supplemental Table 5, and the detailed description of the steps in the validation procedure is presented in "Methodology used to externally validate the 4-lncRNA score" in the supplemental Data. Similar to the training cohort, in the validation cohort, patients with higher lncRNA scores had a significantly shorter OS (P = .007), and a higher cumulated incidence of transformation to AML at 5 years (P = .001), compared with lower-score patients (Figure 4). In multivariate analysis of OS and leukemia-free survival in the 164 MDS patients who had diagnostic cytogenetic data (for calculating IPSS-R), we included age, karyotype, IPSS-R, and gene mutations that had P values <.10 in univariate Cox regression analysis (supplemental Table 6) as covariables. Higher lncRNA risk score was identified as an independent unfavorable prognostic factor for both OS (relative risk, 4.783; 95% confidence interval [CI], 2.491-9.182; P < .001) and leukemia-free survival (relative risk, 7.670; 95% CI, 2.702-21.772; P < .001) after adjusting for other clinical and molecular parameters (Table 3).

Correlation of the lncRNA signature with gene expression and potential functionality

We first profiled the microarray data of 19 867 mRNA probes (corresponding to 18 638 unique coding genes) in the 176 patients and compared the expression levels between patients with the highest (> average + 1 SD; n = 22) and the lowest (< average − 1 SD; n = 22) lncRNA risk scores. Using the threshold of a fold-change of 2 and the Student t test P < .001, we identified
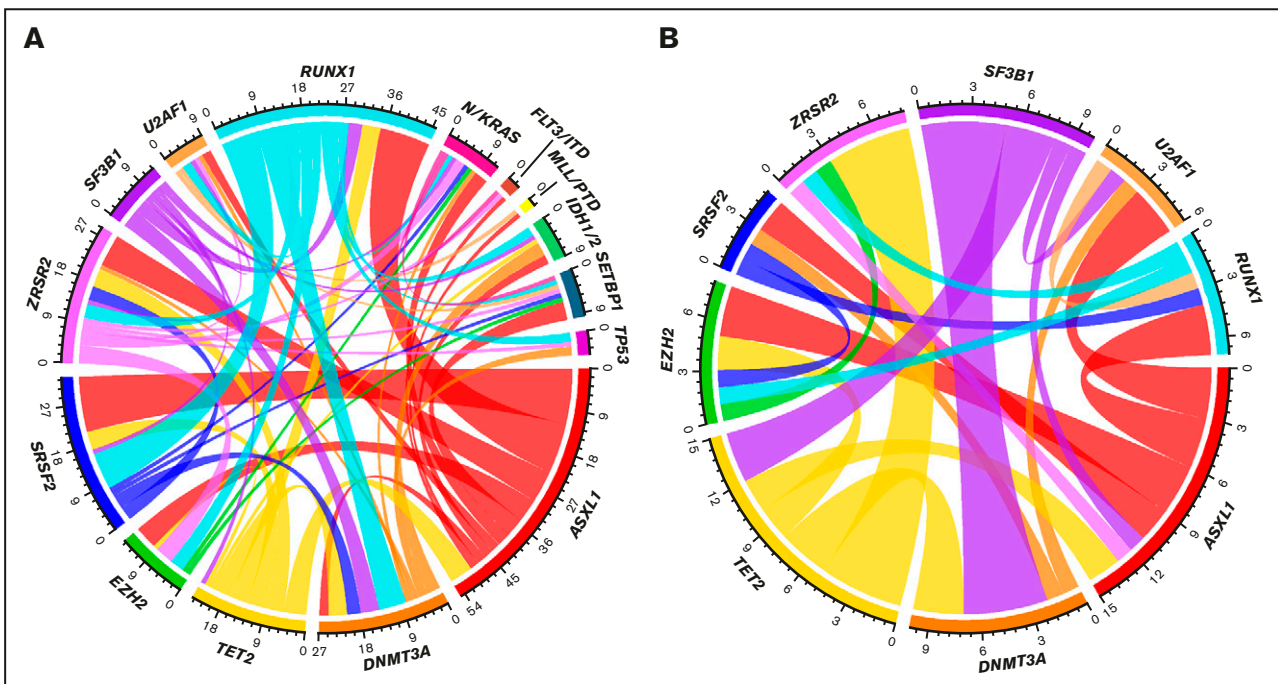
Figure 1. Circos diagrams illustrating the different comutation patterns. Panels A and B depict the co-occurrence of mutations in patients with higher vs lower lncRNA scores, respectively. Ribbon widths are proportional to the frequencies of mutation co-occurrence.
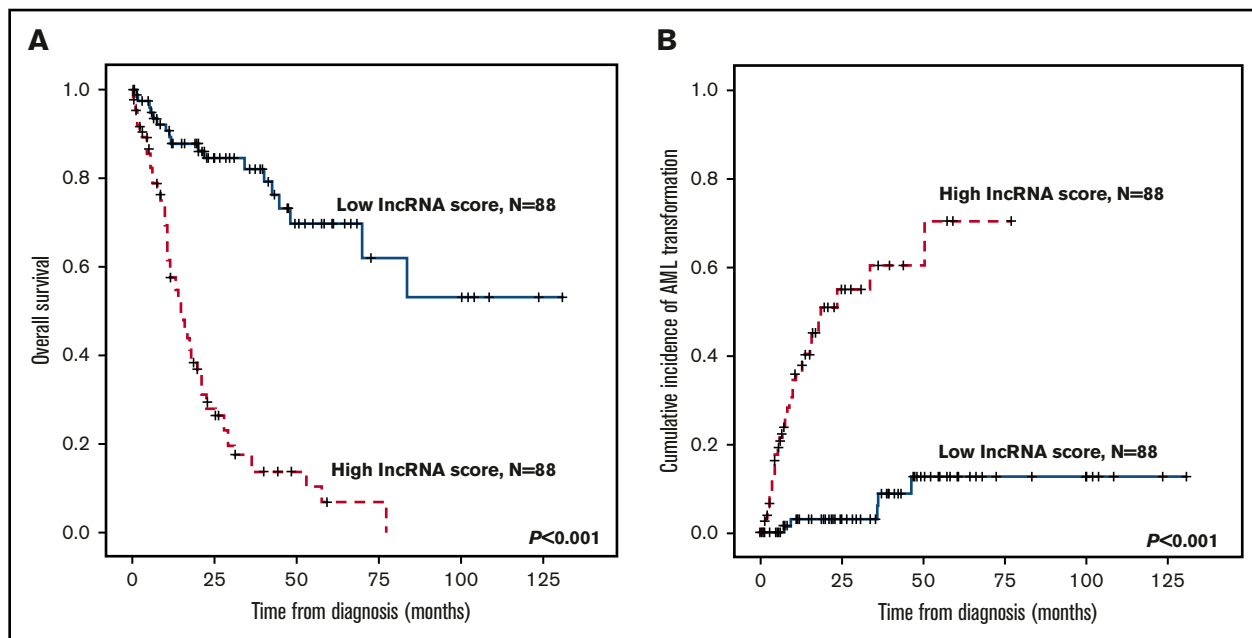
Figure 2. Kaplan-Meier plots, stratified by lncRNA scores. (A) OS and (B) cumulative incidence of AML transformation of the 176 MDS patients in the training cohort. Patients with higher lncRNA scores had worse clinical outcomes than those with lower scores.

255 genes that were differentially expressed between these 2 groups of patients (listed in supplemental Table 7; corresponding heatmap in Figure 5). Of note, several genes in the homeobox (*HOX*) family, such as *HOXA5*, *HOXA6*, *HOXA7*, and *MEIS1*, as well as stem cell markers such as *CD34* and *KIT*, were significantly upregulated in the patients with high lncRNA scores.

We used DAVID for functional annotation of the 255 genes and identified a number of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways that were enriched in patients with higher lncRNA scores (supplemental Table 8), with "hematopoietic cell lineage" and "T-cell receptor signaling" pathways showing the top statistical significances.
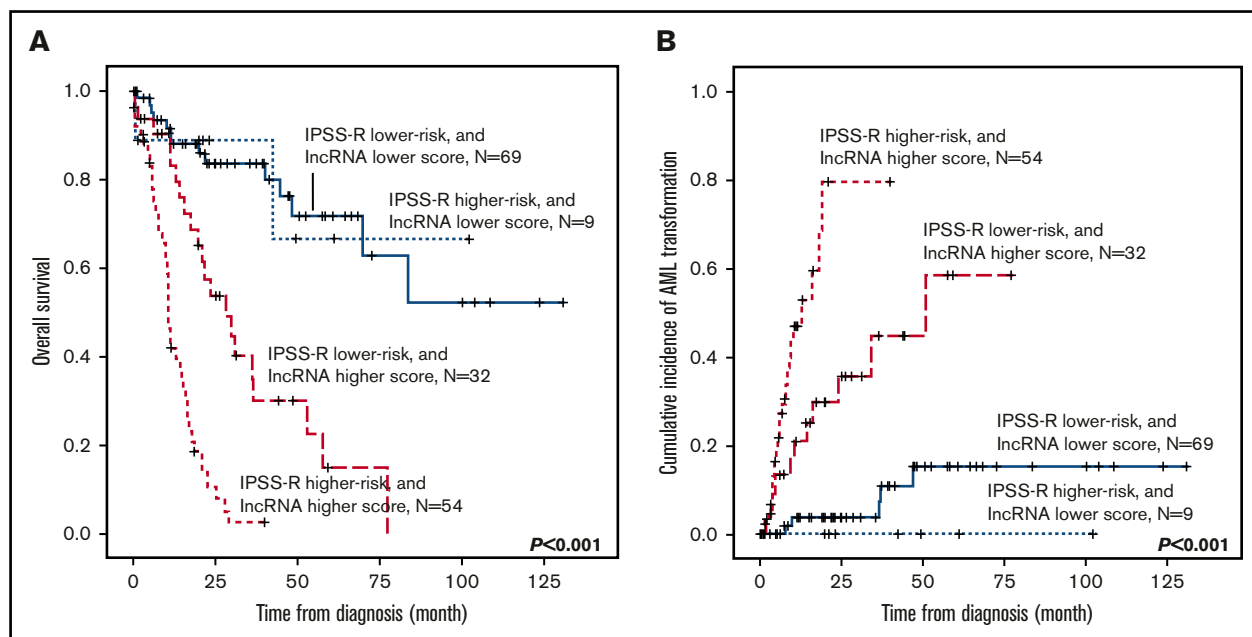

Figure 3. Kaplan-Meier plots, stratified by the integration of IPSS-R and lncRNA scores. (A) OS and (B) cumulative incidence of AML transformation of the 164 patients in the training cohort who had cytogenetic data at diagnosis (thus IRSS-R could be calculated). The incorporation of the lncRNA scoring system into the traditional IPSS-R classification could refine risk stratification of MDS patients.
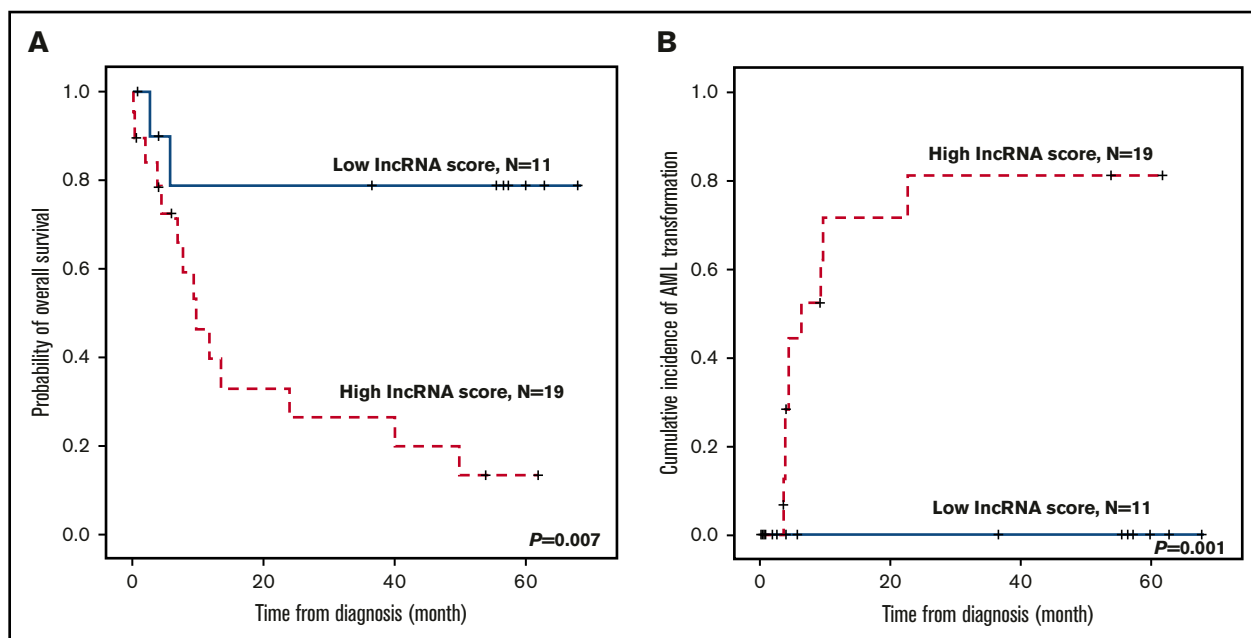
Figure 4. Kaplan-Meier plots of an independent validation cohort stratified by lncRNA scores. (A) OS and (B) cumulative incidence of AML transformation of the 30 MDS patients in the validation cohort.. Patients with higher lncRNA scores had shorter OS and higher AML transformation rates.

In IPA analysis, we also noted several implicated biological pathways related to the 255 differentially expressed genes (supplemental Figure 9). The pathways with top 5 significance levels were "T-cell receptor (TCR) signaling" ($P < .001$), "inducible T-cell costimulator (iCOS)-iCOS ligand (iCOSL) signaling in T helper cells" ($P < .001$), "natural killer (NK) cell signaling" ($P < .001$), "cytotoxic T-lymphocyte–associated antigen 4 (CTLA4) signaling in cytotoxic T lymphocytes" ($P < .001$), and "CD28 signaling in T helper cells" ($P < .001$).

Discussion

Establishing a powerful prognostic system to guide the appropriate treatment of MDSs is important because of the highly complex clinical presentations and the corresponding treatment strategies of this disease.

Gene expression profiling of mRNA levels had been found to outperform the traditional parameters, such as morphology and cytogenetics, in prognostication of MDSs whether CD34[+] selected cells[62-64] or unsorted bulk marrow cells[65] were used for analysis. As lncRNAs have emerged as important participants in the pathogenesis and disease progression of AML,[66-74] it is likely that profiling of their expression would have significant impact on MDS prognostication. Recently, it has been shown that lncRNA expression signatures were closely associated with specific mutations and a small subset of lncRNAs could predict clinical outcome in cytogenetically normal elderly AML patients,[75] but the clinical significance of lncRNA profiling in MDSs has not yet been reported. To our knowledge, this is the first report to provide a lncRNA platform for prognostication in MDS patients. A scoring system was constructed based on the expression levels and weights of only 4 lncRNAs through well-organized statistical modeling to ensure its significant and independent prognostic implication.

Although the marrow cells in MDSs are heterogeneous with regard to lineages and maturation stages, studies have demonstrated the clonal nature of these cells regardless of the blast counts.[76]

Although the amount of immature cells may influence the expression levels of lncRNAs, in our study, we clearly demonstrated that the prognostic significance of the lncRNA score held true not only in an unselected cohort of MDS patients, but also in various selected patient subsets, such as the lower- or higher-risk subgroups based on IPSS-R and WHO classifications. Furthermore, we found that this lncRNA score could maintain its performance on outcome prediction in the HMA-treated MDS patient subset. This observation is clinically relevant as HMAs are the mainstay treatment of the

Table 3. Multivariate analysis for OS and LFS in 164 MDS patients who had cytogenetic data at diagnosis

	OS				LFS			
		95% CI				95% CI		
Variable	RR	Lower	Upper	*P*	RR	Lower	Upper	*P*
Age*	1.022	1.006	1.039	.006†	0.984	0.964	1.004	.122
Karyotype‡	0.590	0.251	1.387	.226	0.454	0.118	1.740	.249
IPSS-R§	1.621	1.220	2.153	.001†	1.820	1.206	2.746	.004†
ASXL1	1.039	0.491	2.201	.920	4.065	1.576	10.483	.004†
TP53	3.735	1.327	10.514	.013†	4.795	0.790	29.081	.088
EZH2	1.847	0.643	5.300	.254	0.960	0.205	4.485	.958
SRSF2	0.985	0.448	2.168	.971	0.678	0.214	2.151	.510
ZRSR2	1.068	0.502	2.272	.863	0.442	0.121	1.614	.216
lncRNA score‖	4.783	2.491	9.182	<.001†	7.670	2.702	21.772	<.001†

Only variables with $P \leq .10$ in univariate analysis were incorporated into the multivariate Cox proportional hazard regression analysis.

CI, confidence interval; LFS, leukemia-free survival; RR, relative risk.

*Age as a continuous variable.

†Statistically significant if $P < .05$.

‡Unfavorable cytogenetics vs others. Patients without chromosome data at diagnosis were not included in the analysis.

§IPSS-R risk groups: very good, good, intermediate, poor, very poor.
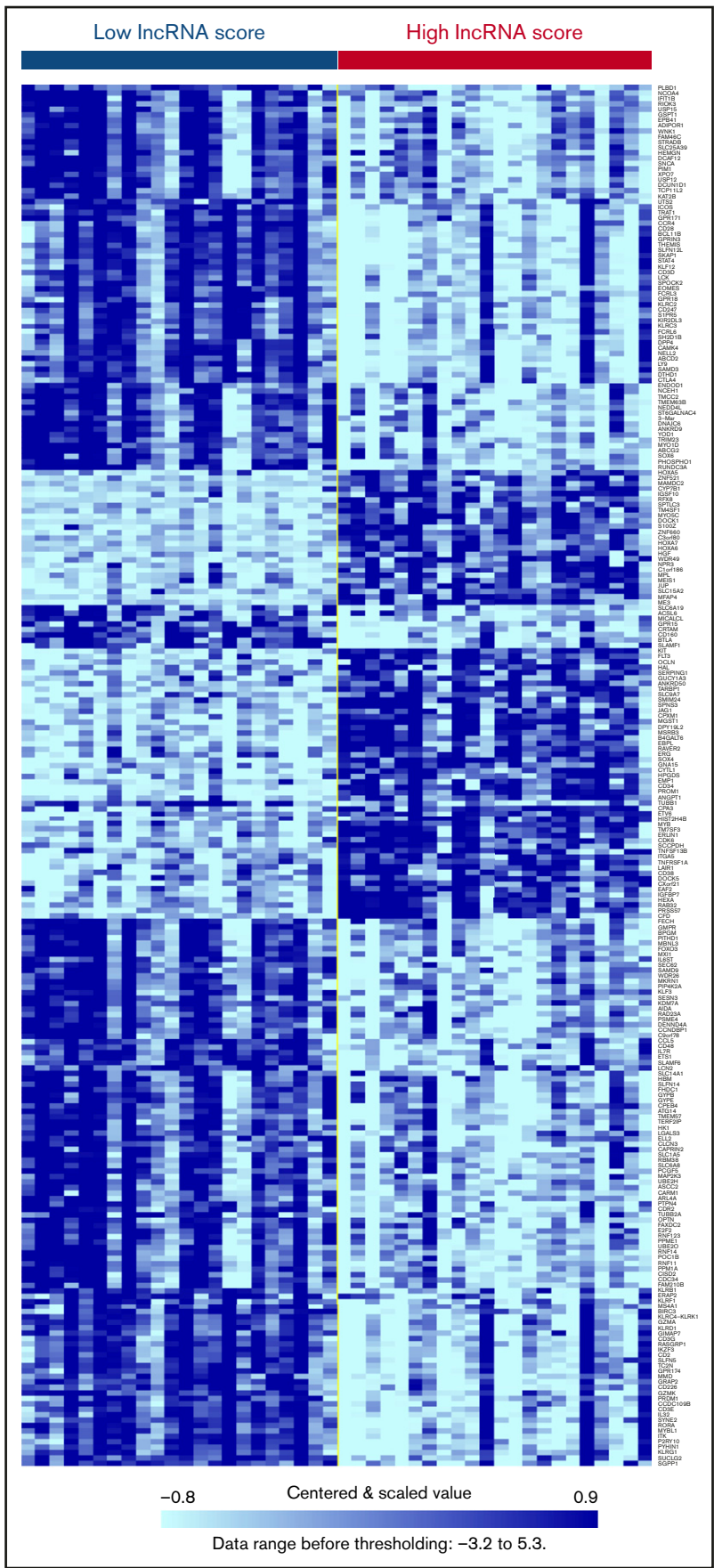
‖High vs low lncRNA risk scores.

Figure 5. Heatmap demonstrating differential gene expression. Depicted are the 255 genes that are differentially expressed between patients with the highest (> average +1 SD, n = 22) and the lowest (< average − 1 SD, n = 22) lncRNA risk scores.

high-risk MDS patients nowadays. In multivariate analysis, we demonstrated that the lncRNA score remained an independent prognostic factor with regard to both OS and leukemia-free survival, after adjusting for other clinical parameters (which included IPSS-R, and thus taking the influence of marrow blast percentage into consideration) and gene mutations. Therefore, the poor prognostic implication of higher lncRNA score is indeed independent of the maturation stages of BM cells, or a more advanced disease status at diagnosis.

The finding that several *HOX* family genes were upregulated in patients with high lncRNA scores was intriguing. The *HOX* genes have been known to be a conserved family of homeodomain-containing transcription factors acting as critical effectors in cell-fate determination and differentiation, and consequently implicated in hematopoiesis and leukemogenesis.[77] Overexpression of the *HOX* genes is associated with unfavorable prognosis in AML,[78] and their deregulation is also linked to higher risk of AML transformation in MDSs.[65] *HOXA5* expression can inhibit erythropoiesis but promote myeloid lineage commitment.[79] On the other hand, overexpression of *HOXA6*, *HOXA7*, *HOXA9*, and *MEIS1* have been reported to correlate with AML with *MLL* rearrangements.[80] The association of high lncRNA scores and upregulation of HOX genes could possibly contribute to the more advanced disease stages and worse clinical outcomes of MDS patients with high lncRNA scores. On the other hand, DAVID and IPA analysis showed that the 255 differentially expressed genes between highest and lowest lncRNA scores were highly associated with hematopoietic cell lineage–related pathways, and those related to T-cell and NK-cell functions. These findings might reflect recent discoveries that dysregulated innate and adaptive immunity lead to the pathogenesis of MDSs. For instance, clonal T-cell expansion and skewed TCR repertoire in CD4 and CD8 T cells have been implicated in the pathogenesis of MDSs.[81,82] There is also evidence that the compromised regulatory T-cell and NK-cell functions can contribute to ineffective immune surveillance, culminating in the pathogenesis of MDSs, and govern disease risk and progression.[83-87]

In summary, we provide evidence that higher expression levels of the 4 lncRNAs in our scoring system are associated with distinct clinical and genetic features and could predict inferior outcomes of MDS patients. To our knowledge, this is the first study to develop a risk prediction model incorporating lncRNA expressions and their weights for MDS patients. We believe our lncRNA risk score, in combination with IPSS-R, may further improve risk stratification of MDS patients and guide treatment options accordingly. In this era of high-throughput technologies, the incorporation of multidisciplinary parameters including molecular biomarkers would undoubtedly move us one step further toward the holy grail of precision medicine. We acknowledge that further basic research would be necessary to delve into the biological functions and networking of these lncRNAs, and prospective clinical studies are anticipated to validate the prognostic power of our lncRNA risk score.

Authorship

Contribution: C.-Y.Y. was responsible for data collection and management, statistical analysis and interpretation, literature research, and manuscript writing; C.-H.C. was responsible for data management and statistical analysis; H.-H.H. assisted in statistical analysis; H.-A.H. and C.-C.L. were responsible for data collection and management; M.-H.T. and C.-J.K. performed the gene mutation studies; and T.-P.L., W.-C.C., and H.-F.T. planned, designed, and coordinated the study over the entire period and wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Hwei-Fang Tien, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; e-mail: hftien@ntu.edu.tw; Wen-Chien Chou, Department of Laboratory Medicine, National Taiwan University Hospital, Taipei, Taiwan; e-mail: wchou@ntu.edu.tw; and Tzu-Pin Lu, Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan; e-mail: tplu@ntu.edu.tw.

References

1. Tefferi A, Vardiman JW. Myelodysplastic syndromes. *N Engl J Med*. 2009;361(19):1872-1885.

2. Adès L, Itzykson R, Fenaux P. Myelodysplastic syndromes. *Lancet*. 2014;383(9936):2239-2252.

3. Greenberg P, Cox C, LeBeau MM, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood*. 1997;89(6):2079-2088.

4. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood*. 2012;120(12):2454-2465.

5. Stone RM. How I treat patients with myelodysplastic syndromes. *Blood*. 2009;113(25):6296-6303.

6. Sekeres MA, Cutler C. How we treat higher-risk myelodysplastic syndromes. *Blood*. 2014;123(6):829-836.

7. Fenaux P, Adès L. How we treat lower-risk myelodysplastic syndromes. *Blood*. 2013;121(21):4280-4286.

8. Gangat N, Patnaik MM, Tefferi A. Myelodysplastic syndromes: contemporary review and how we treat. *Am J Hematol*. 2016;91(1):76-89.

9. Haase D, Germing U, Schanz J, et al. New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood*. 2007;110(13):4385-4395.

10. Morel P, Hebbar M, Lai JL, et al. Cytogenetic analysis has strong independent prognostic value in de novo myelodysplastic syndromes and can be incorporated in a new scoring system: a report on 408 cases. *Leukemia*. 1993;7(9):1315-1323.

11. Bernasconi P, Alessandrino EP, Boni M, et al. Karyotype in myelodysplastic syndromes: relations to morphology, clinical evolution, and survival. *Am J Hematol*. 1994;46(4):270-277.

12. Bejar R. Clinical and genetic predictors of prognosis in myelodysplastic syndromes. *Haematologica*. 2014;99(6):956-964.

13. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. *Blood*. 2013;122(25):4021-4034.

14. Bejar R, Steensma DP. Recent developments in myelodysplastic syndromes. *Blood*. 2014;124(18):2793-2803.

15. Papaemmanuil E, Gerstung M, Malcovati L, et al; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616-3627, quiz 3699.

16. Visconte V, Tiu RV, Rogers HJ. Pathogenesis of myelodysplastic syndromes: an overview of molecular and non-molecular aspects of the disease. *Blood Res*. 2014;49(4):216-227.

17. Epling-Burnette PK, List AF. Advancements in the molecular pathogenesis of myelodysplastic syndrome. *Curr Opin Hematol*. 2009;16(2):70-76.

18. Liao R, Xu Y, Chen M, Chen X, Zhan X, Sun J. Molecular mechanism of microRNA involvement in genesis of myelodysplastic syndrome and its transformation to acute myeloid leukemia. *Hematology*. 2013;18(4):191-197.

19. Dostalova Merkerova M, Krejcik Z, Votavova H, Belickova M, Vasikova A, Cermak J. Distinctive microRNA expression profiles in CD34+ bone marrow cells from patients with myelodysplastic syndrome. *Eur J Hum Genet*. 2011;19(3):313-319.

20. Fang J, Varney M, Starczynowski DT. Implication of microRNAs in the pathogenesis of MDS. *Curr Pharm Des*. 2012;18(22):3170-3179.

21. Rhyasen GW, Starczynowski DT. Deregulation of microRNAs in myelodysplastic syndrome. *Leukemia*. 2012;26(1):13-22.

22. Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*. 2014;15(1):7-21.

23. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*. 2009;23(13):1494-1504.

24. Kung JT, Colognori D, Lee JT. Long noncoding RNAs: past, present, and future. *Genetics*. 2013;193(3):651-669.

25. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009;10(3):155-159.

26. Zhang H, Chen Z, Wang X, Huang Z, He Z, Chen Y. Long non-coding RNA: a new player in cancer. *J Hematol Oncol*. 2013;6:37.

27. Qi P, Zhou XY, Du X. Circulating long non-coding RNAs in cancer: current status and future perspectives. *Mol Cancer*. 2016;15(1):39.

28. Schmitt AM, Chang HY. Long noncoding RNAs in cancer pathways. *Cancer Cell*. 2016;29(4):452-463.

29. Meseure D, Drak Alsibai K, Nicolas A, Bieche I, Morillon A. Long noncoding RNAs as new architects in cancer epigenetics, prognostic biomarkers, and potential therapeutic targets. *Biomed Res Int*. 2015;2015:320214.

30. Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015;21(11):1253-1261.

31. Zhang F, Zhang L, Zhang C. Long noncoding RNAs and tumorigenesis: genetic associations, molecular mechanisms, and therapeutic strategies. *Tumour Biol*. 2016;37(1):163-175.

32. Fang Y, Fullwood MJ. Roles, functions, and mechanisms of long non-coding RNAs in cancer. *Genomics Proteomics Bioinformatics*. 2016;14(1):42-54.

33. Bartonicek N, Maag JL, Dinger ME. Long noncoding RNAs in cancer: mechanisms of action and technological advancements. *Mol Cancer*. 2016; 15(1):43.

34. Luo M, Jeong M, Sun D, et al. Long non-coding RNAs control hematopoietic stem cell function. *Cell Stem Cell*. 2015;16(4):426-438.

35. Heuston EF, Lemon KT, Arceci RJ. The beginning of the road for non-coding RNAs in normal hematopoiesis and hematologic malignancies. *Front Genet*. 2011;2:94.

36. Morlando M, Ballarino M, Fatica A. Long non-coding RNAs: new players in hematopoiesis and leukemia. *Front Med (Lausanne)*. 2015;2:23.

37. Alvarez-Dominguez JR, Hu W, Gromatzky AA, Lodish HF. Long noncoding RNAs during normal and malignant hematopoiesis. *Int J Hematol*. 2014;99(5): 531-541.

38. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 2009;114(5):937-951.

39. Tien HF, Tang JH, Tsay W, et al. Methylation of the p15(INK4B) gene in myelodysplastic syndrome: it can be detected early at diagnosis or during disease progression and is highly associated with leukaemic transformation. *Br J Haematol*. 2001;112(1):148-154.

40. Shaffer LG, Slovak ML, Campbell LJ. ISCN 2009: An International System for Human Cytogenetic Nomenclature. Basel, Switzerland: Karger; 2009.

41. Chen TC, Hou HA, Chou WC, 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(1):e177.

42. Chou WC, Hou HA, Chen CY, et al. Distinct clinical and biologic characteristics in adult acute myeloid leukemia bearing the isocitrate dehydrogenase 1 mutation. *Blood*. 2010;115(14):2749-2754.

43. Lin CC, Hou HA, Chou WC, 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):137-144.

44. Ernst T, Chase AJ, Score J, et al. Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*. 2010;42(8): 722-726.

45. Chou WC, Chou SC, Liu CY, et al. TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood*. 2011;118(14):3803-3810.

46. Hou HA, Kuo YY, Liu CY, et al. DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. *Blood*. 2012; 119(2):559-568.

47. Lin CC, Hou HA, Chou WC, et al. SF3B1 mutations in patients with myelodysplastic syndromes: the mutation is stable during disease evolution. *Am J Hematol*. 2014;89(8):E109-E115.

48. Hou HA, Liu CY, Kuo YY, et al. Splicing factor mutations predict poor prognosis in patients with de novo acute myeloid leukemia. *Oncotarget*. 2016;7(8): 9084-9101.

49. Wu SJ, Kuo YY, Hou HA, et al. The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood*. 2012;120(15):3106-3111.

50. Thol F, Kade S, Schlarmann C, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. 2012;119(15):3578-3584.

51. Chou WC, Hou HA, Liu CY, et al. Sensitive measurement of quantity dynamics of FLT3 internal tandem duplication at early time points provides prognostic information. *Ann Oncol*. 2011;22(3):696-704.

52. Hou HA, Kuo YY, Tang JL, 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):181-186.

53. Tang JL, Hou HA, Chen CY, 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):5352-5361.

54. Shiah HS, Kuo YY, Tang JL, et al. Clinical and biological implications of partial tandem duplication of the MLL gene in acute myeloid leukemia without chromosomal abnormalities at 11q23. *Leukemia*. 2002;16(2):196-202.

55. Hou HA, Chou WC, Kuo YY, 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.

56. Volders PJ, Helsens K, Wang X, et al. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res*. 2013;41(Database issue):D246-D251.

57. Harrow J, Denoeud F, Frankish A, et al. GENCODE: producing a reference annotation for ENCODE. *Genome Biol*. 2006;(7 suppl 1):S4.1-S4.9.

58. Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4(1):44-57.

59. Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1-13.

60. Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30(4):523-530.

61. Simon R, Lam A, Li MC, Ngan M, Menenzes S, Zhao Y. Analysis of gene expression data using BRB-ArrayTools. *Cancer Inform*. 2007;3:11-17.

62. Pellagatti A, Benner A, Mills KI, et al. Identification of gene expression-based prognostic markers in the hematopoietic stem cells of patients with myelodysplastic syndromes. *J Clin Oncol*. 2013;31(28):3557-3564.

63. Sridhar K, Ross DT, Tibshirani R, Butte AJ, Greenberg PL. Relationship of differential gene expression profiles in CD34+ myelodysplastic syndrome marrow cells to disease subtype and progression. *Blood*. 2009;114(23):4847-4858.

64. Gerstung M, Pellagatti A, Malcovati L, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun*. 2015;6:5901.

65. Mills KI, Kohlmann A, Williams PM, et al. Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of AML transformation of myelodysplastic syndrome. *Blood*. 2009;114(5):1063-1072.

66. Hirano T, Yoshikawa R, Harada H, Harada Y, Ishida A, Yamazaki T. Long noncoding RNA, CCDC26, controls myeloid leukemia cell growth through regulation of KIT expression. *Mol Cancer*. 2015;14:90.

67. Zeng C, Xu Y, Xu L, et al. Inhibition of long non-coding RNA NEAT1 impairs myeloid differentiation in acute promyelocytic leukemia cells. *BMC Cancer*. 2014;14:693.

68. Zhang X, Weissman SM, Newburger PE. Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells. *RNA Biol*. 2014;11(6):777-787.

69. Hughes JM, Legnini I, Salvatori B, et al. C/EBPα-p30 protein induces expression of the oncogenic long non-coding RNA UCA1 in acute myeloid leukemia. *Oncotarget*. 2015;6(21):18534-18544.

70. Wu S, Zheng C, Chen S, et al. Overexpression of long non-coding RNA HOTAIR predicts a poor prognosis in patients with acute myeloid leukemia. *Oncol Lett*. 2015;10(4):2410-2414.

71. Xing CY, Hu XQ, Xie FY, et al. Long non-coding RNA HOTAIR modulates c-KIT expression through sponging miR-193a in acute myeloid leukemia. *FEBS Lett*. 2015;589(15):1981-1987.

72. Zeng C, Yu X, Lai J, Yang L, Chen S, Li Y. Overexpression of the long non-coding RNA PVT1 is correlated with leukemic cell proliferation in acute promyelocytic leukemia. *J Hematol Oncol*. 2015;8:126.

73. Chen L, Wang W, Cao L, Li Z, Wang X. Long non-coding RNA CCAT1 acts as a competing endogenous RNA to regulate cell growth and differentiation in acute myeloid leukemia. *Mol Cells*. 2016;39(4):330-336.

74. McCarty G, Loeb DM. Hypoxia-sensitive epigenetic regulation of an antisense-oriented lncRNA controls WT1 expression in myeloid leukemia cells. *PLoS One*. 2015;10(3):e0119837.

75. Garzon R, Volinia S, Papaioannou D, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. *Proc Natl Acad Sci USA*. 2014; 111(52):18679-18684.

76. Walter MJ, Shen D, Ding L, et al. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med*. 2012;366(12):1090-1098.

77. Alharbi RA, Pettengell R, Pandha HS, Morgan R. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia*. 2013;27(5): 1000-1008.

78. van Oostveen J, Bijl J, Raaphorst F, Walboomers J, Meijer C. The role of homeobox genes in normal hematopoiesis and hematological malignancies. *Leukemia*. 1999;13(11):1675-1690.

79. Fuller JF, McAdara J, Yaron Y, Sakaguchi M, Fraser JK, Gasson JC. Characterization of HOX gene expression during myelopoiesis: role of HOX A5 in lineage commitment and maturation. *Blood*. 1999;93(10):3391-3400.

80. De Braekeleer E, Douet-Guilbert N, Basinko A, Le Bris MJ, Morel F, De Braekeleer M. Hox gene dysregulation in acute myeloid leukemia. *Future Oncol*. 2014;10(3):475-495.

81. Melenhorst JJ, Eniafe R, Follmann D, Nakamura R, Kirby M, Barrett AJ. Molecular and flow cytometric characterization of the CD4 and CD8 T-cell repertoire in patients with myelodysplastic syndrome. *Br J Haematol*. 2002;119(1):97-105.

82. Epling-Burnette PK, Painter JS, Rollison DE, et al. Prevalence and clinical association of clonal T-cell expansions in myelodysplastic syndrome. *Leukemia*. 2007;21(4):659-667.

83. Kotsianidis I, Bouchliou I, Nakou E, et al. Kinetics, function and bone marrow trafficking of CD4+CD25+FOXP3+ regulatory T cells in myelodysplastic syndromes (MDS). *Leukemia*. 2009;23(3):510-518.

84. Kordasti SY, Ingram W, Hayden J, et al. CD4+CD25high Foxp3+ regulatory T cells in myelodysplastic syndrome (MDS). *Blood*. 2007;110(3):847-850.

85. Kiladjian JJ, Bourgeois E, Lobe I, et al. Cytolytic function and survival of natural killer cells are severely altered in myelodysplastic syndromes. *Leukemia*. 2006;20(3):463-470.

86. Epling-Burnette PK, Bai F, Painter JS, et al. Reduced natural killer (NK) function associated with high-risk myelodysplastic syndrome (MDS) and reduced expression of activating NK receptors. *Blood*. 2007;109(11):4816-4824.

87. Hejazi M, Manser AR, Fröbel J, et al. Impaired cytotoxicity associated with defective natural killer cell differentiation in myelodysplastic syndromes. *Haematologica*. 2015;100(5):643-652.