REGULAR ARTICLE

blood advances

A 4-gene leukemic stem cell score can independently predict the prognosis of myelodysplastic syndrome patients

Yu-Hung Wang,[1,2,*] Chien-Chin Lin,[1-3,*] Chi-Yuan Yao,[2,3] Chia-Lang Hsu,[4] Hsin-An Hou,[2] Cheng-Hong Tsai,[2] Wen-Chien Chou,[2,3] and Hwei-Fang Tien[2]

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

Key Points

- A 4-gene LSC score based on LSC gene expression and its prognostic significance was constructed to improve risk stratification of MDS.
- A higher 4-gene LSC score is an independent adverse prognostic factor for both overall and leukemia-free survivals in MDS patients.

Myelodysplastic syndrome (MDS) comprised a heterogeneous group of diseases. The prognosis of patients varies even in the same risk groups. Searching for novel prognostic markers is warranted. Leukemic stem cells (LSCs) are responsible for chemoresistance and relapse in leukemia. Recently, expressions of 17 genes related to stemness of LSCs were found to be associated with prognosis in acute myeloid leukemia patients. However, the clinical impact of LSC genes expressions in MDS, a disorder arising from hematopoietic stem cells, remains unclear. We analyzed expression profile of the 17 stemness-related genes in primary MDS patients and identified expression of 4 genes (*LAPTM4B*, *NGFRAP1*, *EMP1*, and *CPXM1*) were significantly correlated with overall survival (OS). We constructed an LSC4 scoring system based on the weighted sums of the expression of 4 genes and explored its clinical implications in MDS patients. Higher LSC4 scores were associated with higher revised International Prognostic Scoring System (IPSS-R) scores, complex cytogenetics, and mutations in *RUNX1*, *ASXL1*, and *TP53*. High-score patients had significantly shorter OS and leukemia-free survival (LFS), which was also confirmed in 2 independent validation cohorts. Subgroup analysis revealed the prognostic significance of LSC4 scores for OS remained valid across IPSS-R lower- and higher-risk groups. Furthermore, higher LSC4 score was an independent adverse risk factor for OS and LFS in multivariate analysis. In summary, LSC4 score can independently predict prognosis in MDS patients irrespective of IPSS-R risks and may be used to guide the treatment of MDS patients, especially lower-risk group in whom usually only supportive treatment is given.

Introduction

Myelodysplastic syndromes are clonal myeloid malignancies arising from hematopoietic stem cells (HSCs), which are characterized by bone marrow (BM) ineffective hematopoiesis, peripheral blood cytopenias,[1-5] and a propensity of transformation to acute myeloid leukemia (AML).[1-3] The clinical features and outcomes of MDS patients vary considerably, underscoring the importance of individualized management. The International Prognostic Scoring System (IPSS)[6] and the revised IPSS (IPSS-R)[7] have been widely used to risk-stratify MDS patients and guide the choice of treatment. However, the prognosis of patients may be different even in the same risk groups.[8,9] The search for new prognostic markers is needed for better risk classification of MDS patients.

Recently, Ng et al proposed a 17-gene leukemic stem cell (LSC) scoring system (LSC17 score) that could accurately predict the outcomes of AML patients.[10] As the clonal origin of MDS and AML has

Submitted 4 November 2019; accepted 24 January 2020; published online 20 February 2020. DOI 10.1182/bloodadvances.2019001185.

*Y.-H.W. and C.-C.L. contributed equally to this study.

The raw and normalized microarray data reported in this article have been deposited in the Gene Expression Omnibus database (accession number GSE97064).

The full-text version of this article contains a data supplement.

been both demonstrated to lie within the stem cell compartment and tumor stemness is an established property pertaining to poor prognosis,[11-14] we surmise that LSC gene expression may harbor clinical significance in MDS patients. By using Z transformation and secondary multivariate Cox regression analysis, we identified expressions of 4 LSC genes were significantly correlated with overall survival (OS). We constructed a concise, integrated LSC signature–based 4-gene scoring system (LSC4) and found the LSC4 score was closely associated with clinical and biological features and could predict OS and leukemia-free survival (LFS) in MDS patients. The prognostic implication of the LSC4 score on OS remained significant in both IPSS-R lower-risk (very low, low, and intermediate risk) and higher-risk (high- and very-high-risk) groups. We also validated the prognostic significance of the LSC4 score in an independent internal cohort and 1 external cohort from GSE58831 in which microarray data were available.[15] Furthermore, it was an independent risk factor for OS and LFS, irrespective of age, IPSS-R, and mutation status. We believe that after prospective validation with independent cohorts, this scoring system might provide directives to therapeutic decision in MDS patients, especially lower-risk-group patients in whom usually only supportive treatment is given. More aggressive therapy may be indicated for IPSS-R lower-risk patients with higher LSC4 scores.

Materials and methods

Patients

We recruited 176 primary MDS patients diagnosed at the National Taiwan University Hospital (NTUH) from January 1992 to December 2010 who had cryopreserved BM samples for microarray analysis as a training cohort. The diagnosis was based on the 2016 World Health Organization (WHO) classification.[16] Patients with antecedent chemotherapy or hematologic malignancies were excluded. Another independent set of 30 patients diagnosed with the same criteria from January 2011 to May 2012 were recruited as an internal validation cohort. For external validation, we collected publically available data from GSE58831[15] in which 176 patients were annotated, but survival and microarray data for gene expression levels were both available in only 113 MDS patients. This study was approved by the institutional review board of the NTUH.

The median age of the 176 MDS patients in the training cohort was 68.7 years. Among the 164 patients who had karyotype data at diagnosis, 17.1% had IPSS-R very-high-risk MDS, 21.3% were high risk, 25% were intermediate risk, 32.9% were low risk, and 3.7% were very low risk (Table 1). Most patients (121 patients [68.8%]) received supportive care (only because a hypomethylation agent [HMA] was not reimbursed by National Health Insurance in Taiwan until 2013), and 55 received active treatment, including HMA, low-dose cytarabine, high-intensity chemotherapy, and/or HSC transplantation (HSCT). Nineteen (10.8%) patients, including 15 with MDS with excess blasts (EBs) and 4 with progressive disease during follow-up, underwent allogeneic HSCT. Approximately 10% patients in each group received supportive care due to the patient's choice and/or comorbidity. During a median follow-up duration of 37.3 months (range, 0.1-130.9 months), 83 patients died of the disease, and 38 progressed to AML.

Microarray and genetic alteration analysis

We profiled the global gene expression of BM mononuclear cells from the 206 patients by Affymetrix GeneChip Human Transcriptome Array 2.0 as described previously.[12] The raw and normalized microarray data reported in this article have been deposited in the Gene Expression Omnibus with the accession number GSE97064. The expression levels of the previously identified 17 LSC genes,[10] including *MMRN1*, *DPYSL3*, *CDK6*, *LAPTM4B*, *NGFRAP1*, *CD34*, *AKR1C3*, *EMP1*, *SOCS2*, *NYNRIN*, *KIAA0125*, *GPR56*, *SMIM24*, *DNMT3B*, *CPXM1*, *ZBTB46*, and *ARHGAP22*, were extracted for further analysis. Cytogenetic analyses were performed as described previously[17] and interpreted according to the International System for Human Cytogenetic Nomenclature.[18] We also analyzed the mutation statuses of 17 myeloid-relevant genes, including *ASXL1*, *IDH1*, *IDH2*, *EZH2*, *TET2*, and *DNMT3A*, genes related to the RNA-splicing machinery (including *SF3B1*, *U2AF1*, *SRSF2*, and *ZRSR2)*, as well as *FLT3*/internal tandem duplication, *NRAS*, *KRAS*, *RUNX1*, *MLL*/partial tandem duplication, *TP53*, and *SETBP1*, by Sanger sequencing as previously described.[17,19-24]

Establishment of the LSC prognostic score in MDS patients

We first conducted Z transformation for the expression of the 17 LSC genes at probe levels across the 176 MDS patients and set 0 as the mean and calculating unit standard deviation of each gene among the patients. We then used a multivariate approach to analyze the association between OS and the expression level of each LSC gene.

The LSC genes with significant association with OS were assigned for further multivariate Cox regression analysis to find genes independently associated with prognosis. These LSC genes were used to build the LSC prognostic scoring system. We performed the secondary multivariate Cox regression analysis to obtain the β values as the LSC genes' weights in the scoring system. The prognostic LSC score was calculated as the sum of the normalized expression level of each component multiplied by its weight as follows: $\text{risk(j)} = \Sigma\text{LSC component LSC}_i\text{ (j)} \times \beta_i$, where j denotes the patient accession number, LSCi represents the normalized expression level of the LSC probe i after Z transformation, and β_i is the weight of the particular LSC probe i.

Statistical analysis

We used the Mann-Whitney *U* test to compare medians and continuous variables of distribution. The Fisher exact test or the χ^2 test was performed to examine the difference among discrete variables, including gender, WHO classification, cytogenetic changes, IPSS-R, and genetic alterations between patients with lower and higher LSC scores. OS was the duration from the date of initial diagnosis to the time of last follow-up or death of any cause, whichever occurred first. The training set was used to build the LSC scoring system, which was then applied to the validation set to confirm its significance. The survival prediction power of this LSC score was evaluated by both the log-rank test and the univariate Cox proportional hazards model. After >100 000 iterations, the prediction rate of our proposed LSC score was calculated as the fraction of random scoring systems that achieved $P < .05$. We plotted the survival curves with Kaplan-Meier analysis and calculated the statistical significance with the log-rank test. The Cox proportional hazards model was used in multivariate regression analysis. $P < .05$ was considered statistically significant. All statistical analyses were performed with BRB-ArrayTools (version 4.5.1; Biometric Research Branch, National Cancer Institute, Rockville, MD) and IBM SPSS Statistics 23 for Windows. Time-dependent receiver operating

Table 1. Comparison of clinical and laboratory features between patients with lower and higher LSC4 scores

Clinical characteristics	Total (N = 176)	Low LSC4 score (n = 88)	High LSC4 score (n = 88)	*P*
Sex, n (%)				>.999
Male	121 (68.8)	60 (68.2)	61 (69.3)	
Female	55 (31.2)	28 (31.8)	27 (30.7)	
Age, median (range), y	68.65 (18.5-94.5)	67.9 (18.5-94.5)	69.4 (25.9-89.2)	.88
Laboratory data, median (range)				
WBC, $\times 10^9$/L	3.83 (0.49-20.44)	3.78 (1.71-9.99)	3.93 (0.49-20.44)	.88
ANC, $\times 10^9$/L	1.77 (0.1-12.73)	1.89 (0.1-7.0)	1.57 (0.1-12.73)	.45
Hb, g/dL	8.1 (3.5-14.6)	8.3 (3.5-14.6)	7.9 (3.7-14.4)	.88
Platelets, $\times 10^9$/L	86 (3-721)	96 (3-442)	75 (9-721)	.05
BM blasts, %	3 (0-18.8)	1.5 (0-14.6)	7.0 (0-18.8)	<.001
2016 WHO classification, n (%)				
MDS-SLD	42 (23.9)	30 (34.1)	12 (13.6)	.002
MDS-MLD	36 (20.5)	27 (30.7)	9 (10.2)	.001
MDS-RS-SLD	13 (7.4)	12 (13.6)	1 (1.1)	.002
MDS-RS-MLD	9 (5.1)	8 (9.1)	1 (1.1)	.034
MDS-EB1	32 (18.2)	3 (3.4)	29 (33.0)	<.001
MDS-EB2	44 (25.0)	8 (9.1)	36 (40.9)	<.001
IPSS-R,*† n (%)				
Very low	6 (3.7)	6 (6.8)	0 (0)	.029
Low	54 (32.9)	41 (46.6)	13 (14.8)	<.001
Intermediate	41 (25.0)	22 (25.0)	19 (21.6)	.722
High	35 (21.3)	6 (6.8)	29 (33.0)	<.001
Very high	28 (17.1)	4 (4.5)	24 (27.3)	<.001
Treatment, n (%)				
Palliative care	121 (68.8)	72 (81.8)	49 (55.7)	<.001
Active treatment‡				
HMA	17 (9.7)	5 (5.7)	12 (13.6)	.124
LDAraC§	19 (10.8)	4 (4.5)	15 (17)	.013
Chemotherapy	13 (7.4)	3 (3.4)	10 (11.4)	.08
HSCT	19 (10.8)	6 (6.8)	13 (14.8)	.143

$P < .05$ is considered statistically significant.

ANC, absolute neutrophil count; Hb, hemoglobin; LDAraC, low-dose cytarabine; MLD, multilineage dysplasia; RS, ring sideroblasts; SLD, single-lineage dysplasia; WBC, white blood cell count.

*One hundred sixty four patients, including 79 with low LSC4 scores and 85 with high LSC4 scores, had chromosome data at diagnosis.

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

‡Active treatment includes HMA, low-dose cytarabine, high-intensity chemotherapy, and HSCT. Some patients received >1 treatment modality: 2 received HMA and low-dose cytarabine; 4 received LDAraC and high-intensity chemotherapy; 1 received HMA, low-dose cytarabine, and high-intensity chemotherapy; 1 received high-intensity chemotherapy and HSCT; 2 received HMA and HSCT; 1 received HMA, high-intensity chemotherapy, and HSCT; and 15 received HSCT without bridging therapy.

§Low-dose cytarabine at 20 mg once or twice daily for 10 consecutive days every 4 to 6 weeks.

characteristic (ROC) curves analysis was performed using the R package time ROC. The Pearson's correlation coefficient (PCC) was calculated in R language.

Results

Applying LSC17 score in MDS patients

We first applied the LSC17 score constructed by Ng et al[10] to 176 MDS patients in our training cohort and divided patients into higher- and lower-score groups with the median value as a cutoff level. We noticed that the LSC17 score could truly stratify the total MDS cohort into 2 risk groups with different OS and LFS, but further subgroup analyses showed that the scoring system had no prognostic significance in either lower- or higher-risk MDS subgroup according to IPSS-R or WHO classification (supplemental Table 1). Therefore, the prognostic-predicting capability of the LSC17 score is not as good in MDS as in AML, suggesting the more heterogeneous nature of MDS.

Constructing the LSC4 score

To construct a more simplified and powerful prognostic scoring system based on relevant LSC signature, we put the 17 LSC-related genes in a multivariate Cox model to identify the genes whose expression could independently predict OS (supplemental Table 2).

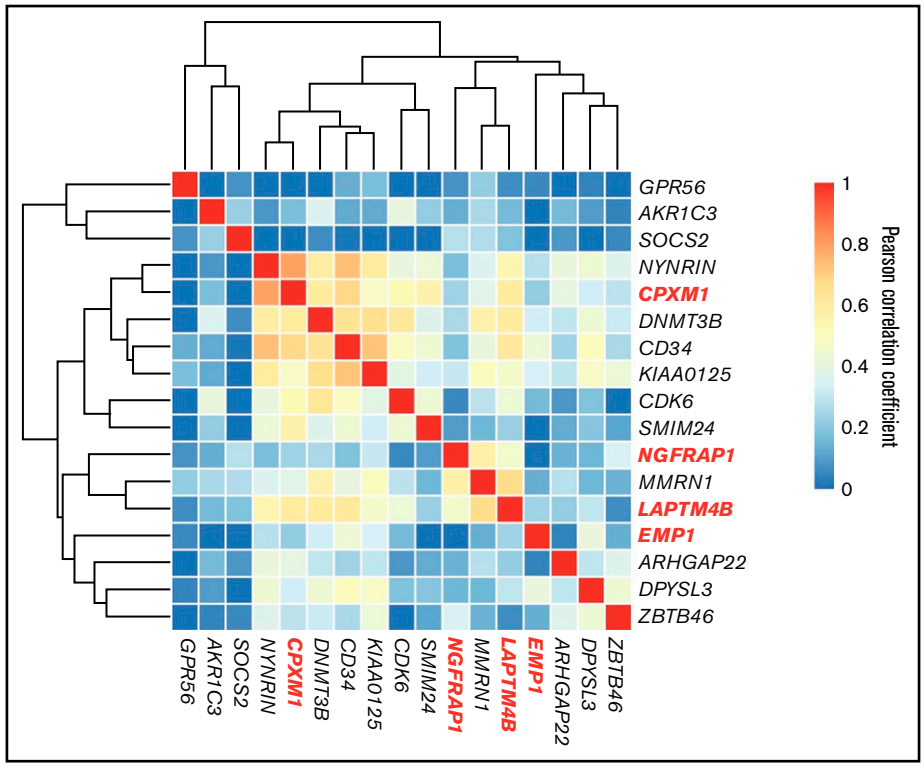

Figure 1. Expression correlation among stemness genes. The 4 genes colored in red are those included in our LSC4 equation.

We found the expression levels of *LAPTM4B*, *NGFRAP1*, *CPXM1*, *CDK6*, *NYNRIN*, and *EMP1* genes were correlated with survival (P = 0.017, 0.03, 0.042, 0.05, 0.072, and 0.078, respectively). We then performed another round of Cox regression analysis for these 6 genes. The expression levels of *LAPTM4B*, *NGFRAP1*, *EMP1*, and *CPXM1* remained significantly correlated with survival (P = .001, .027, .02, and .001, respectively). By integrating the β values as statistical weights, we constructed the LSC4 score, which was calculated with the following equation: [*LAPTM4B*] × 0.502 − [*NGFRAP1*] × 1.013 + [*EMP1*] X 0.181 + [*CPXM1*] × 0.381 (supplemental Table 3).

The LSC4 scores of the 176 MDS patients were calculated, and the relationship between LSC4 score and the expression of LSC-related genes were investigated (Figure 1; supplemental Figure 1). The LSC4 score had a strong correlation with *CPXM1* and *LAPTM4B* expression (PCC = 0.81 and 0.78, respectively), as well as a moderate correlation with *EMP1* (PCC = 0.55), but no correlation with *NGFRAP1* (PCC = 0.02), consistent with coefficients of the LSC4 equation. The large range of expression correlations among the 4 selected genes (PCC = −0.02 to 0.60) suggests the complement of these genes for the prognostic prediction. Although some genes are highly coexpressed with *CPXM1*, such as *NYNRIN* (PCC = 0.79), *CD34* (PCC = 0.68), *DNMT3B* (PCC = 0.60), and *KIAA0125* (PCC = 0.48), only *CPMX1* was selected as one of the predominant markers. It demonstrates that our procedure for marker selection can exclude redundant markers.

Comparison of clinical characteristics and genetic alterations between patients with high and low LSC4 score

The 176 MDS patients in the training set were divided into 2 groups by the median value of the LSC4 scores. A comparison of clinical and laboratory features between the 2 groups is shown in Table 1. The high-score group had higher BM blast percentages at diagnosis (P < .001) compared with the low-score group. Patients with higher scores more frequently had MDS-EB by the 2016 WHO classification, including EB-1 and EB-2, but less MDS-SLD, MDS-MLD, MDS-RS-SLD, and MDS-RS-MLD compared with patients with lower scores. High-score patients were more frequently categorized into IPSS-R high- and very high-risk subgroups but less frequently to the IPSS-R low- and very-low-risk subgroups (Table 1).

Moreover, high-score patients had significantly higher incidence of poor-risk cytogenetics (21.2% vs 5.1%, P = .003) and complex karyotypes (≥3 abnormalities, 17.6% vs 3.8%, P = .005) (supplemental Table 4). Overall, 108 patients (61.4%) had at ≥1 gene mutation, 36 (40.9%) in the low-score subgroup and 72 (81.8%) in the high-score subgroup (P < .001). As listed in supplemental Table 5, the most common mutation in the high-score patients was *ASXL1* mutation (36%), followed by *RUNX1* (26.7%), *SRSF2* (21.8%), and *DNMT3A* (18.4%) mutations. In contrast, *SF3B1* mutation was the most frequent mutation (23.9%) in the low-score patients. High LSC4 score was closely associated with *ASXL1*, *RUNX1*, *SRSF2*, *TP53*, *U2AF1*, and *ZRSR2* mutations but inversely associated with *SF3B1* mutation (supplemental Table 5).

The impact of the LSC4 score on OS and leukemic transformation

Patients with higher LSC4 scores had an inferior OS and LFS than those with lower scores (median, 14.6 months vs 83.6 months, P < .001; and 10.3 months vs 83.6 months, P < .001, respectively; Figure 2A-B). Subgroups analysis showed that the prognostic

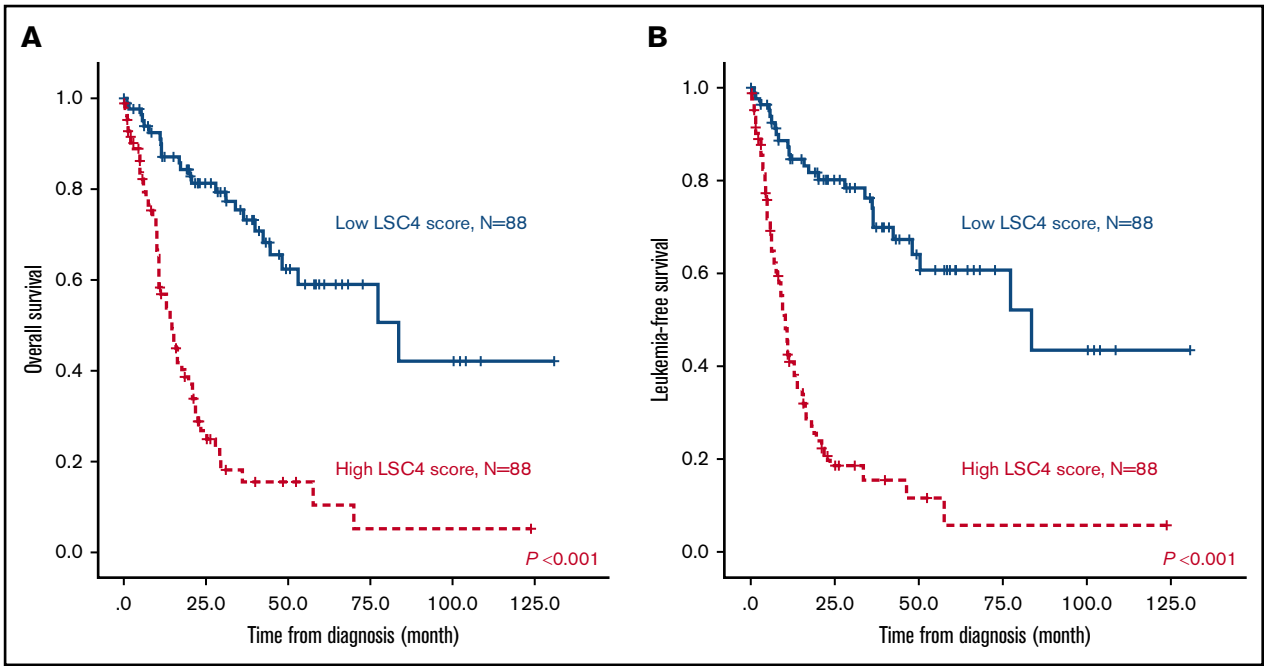

Figure 2. Kaplan-Meier plots stratified by LSC4 scores. OS (A) and LFS (B) of the 176 MDS patients in the training cohort. Patients with higher LSC4 scores had worse clinical outcomes than those with lower scores.

significance of LSC4 score for OS and LFS remained true in both IPSS-R lower-risk (very low, low, and intermediate risk) subgroup (Figure 3A-B) and IPSS-R higher-risk (high and very high risk) subgroup (Figure 3C-D). To further compare the power between LSC4 score and IPSS-R in prognostic prediction, the time-dependent ROC curves were estimated by the inverse probability of censoring weighting method. As illustrated in Figure 4, LSC4 score could be complementary to IPSS-R in prediction of both OS and LFS. Additionally, we performed bivariate analyses, including the LSC4 score and the IPSS-R, and LSC4 score and the presence of a complex karyotype, respectively. As a result, there was a moderate correlation between LSC4 score and IPSS-R (PCC = 0.536, $P < .001$) and a low correlation between LSC4 score and the presence of a complex karyotype (PCC = 0.221, $P = .004$).

The adverse implications of higher LSC4 scores on OS and LFS could also be demonstrated in the subgroups of patients with normal karyotype (n = 100; supplemental Figure 2A-B); patients without unfavorable cytogenetics such as complex karyotypes, monosomy 7, and del(7q) (n = 142, supplemental Figure 2C-D); and those with WHO lower-risk subtypes (MDS-SLD, MDS-MLD, MDS-RS-SLD, and MDS-RS-MLD; n = 100; supplemental Figure 3A-B) or WHO higher-risk subtypes (MDS-EB; n = 76) (supplemental Figure 3C-D).

We further analyzed separately the outcomes of MDS patients receiving different treatments. High-score patients consistently had a significantly inferior outcome in OS and LFS (supplemental Figure 4), regardless of whether they received palliative care (n = 121, 15.4 months vs 83.6 months, $P < .001$; and 13.1 months vs 83.6 months, $P < .001$, respectively) or active treatment (n = 55, 14.6 months vs 42.4 months, $P < .001$; 9 months vs 50.4 months, $P < .001$, respectively). For the patients treated with HMA (n = 17), those with a high LSC4 score also showed a significantly shorter OS and LFS (supplemental Figure 5; 14.6 months vs NR, $P = .047$, and 7 months vs not reached [NR], $P = .02$, respectively).

For multivariate analysis in the training cohort, we included parameters with $P < .05$ in univariate Cox regression analysis as covariates, including age, and mutations in *ASXL1*, *TP53*, *SRSF2*, and *ZRSR2* (supplemental Table 6). Higher LSC4 score, either divided by a median (Table 2) or regarded as continuous values (supplemental Table 7), appeared to be an independent adverse prognostic factor for OS ($P < .001$ and $P < .001$, respectively) and LFS ($P < .001$ and $P < .001$, respectively). To verify the prognostication power of the LSC4 scoring system, we analyzed the expression levels of 17 LSC genes in an independent internal validation cohort of 30 MDS patients. Characteristics of patients in the training cohort and internal validation cohort were generally comparable, as shown in supplemental Table 8. Consistent with the findings in the training cohort, patients with higher LSC4 scores had a shorter OS and LFS (Figure 5A-B; 6.9 months vs NR, $P = .005$, and 3.9 months vs NR, $P = .002$, respectively) than those with lower scores in the validation cohort. Moreover, we validated the prognostic significance of this LSC4 scoring system in one external validation cohort of 113 MDS patients from GSE58831.[15] A comparison of clinical and laboratory features between high- and low-score patients in this cohort is shown in supplemental Table 9. The high-score group had higher BM blast percentages at diagnosis ($P = .004$) and more frequently IPSS intermediate-2 subtype ($P = .003$) but lower absolute neutrophil counts ($P = .005$) and platelet counts ($P = .029$) compared with the low-score group. Low-score patients had a higher incidence of *SF3B1* mutation than high-score patients, but there was no difference in cytogenetic changes and other molecular gene mutations between the 2 groups (supplemental Tables 10 and 11). Patients with higher LSC4

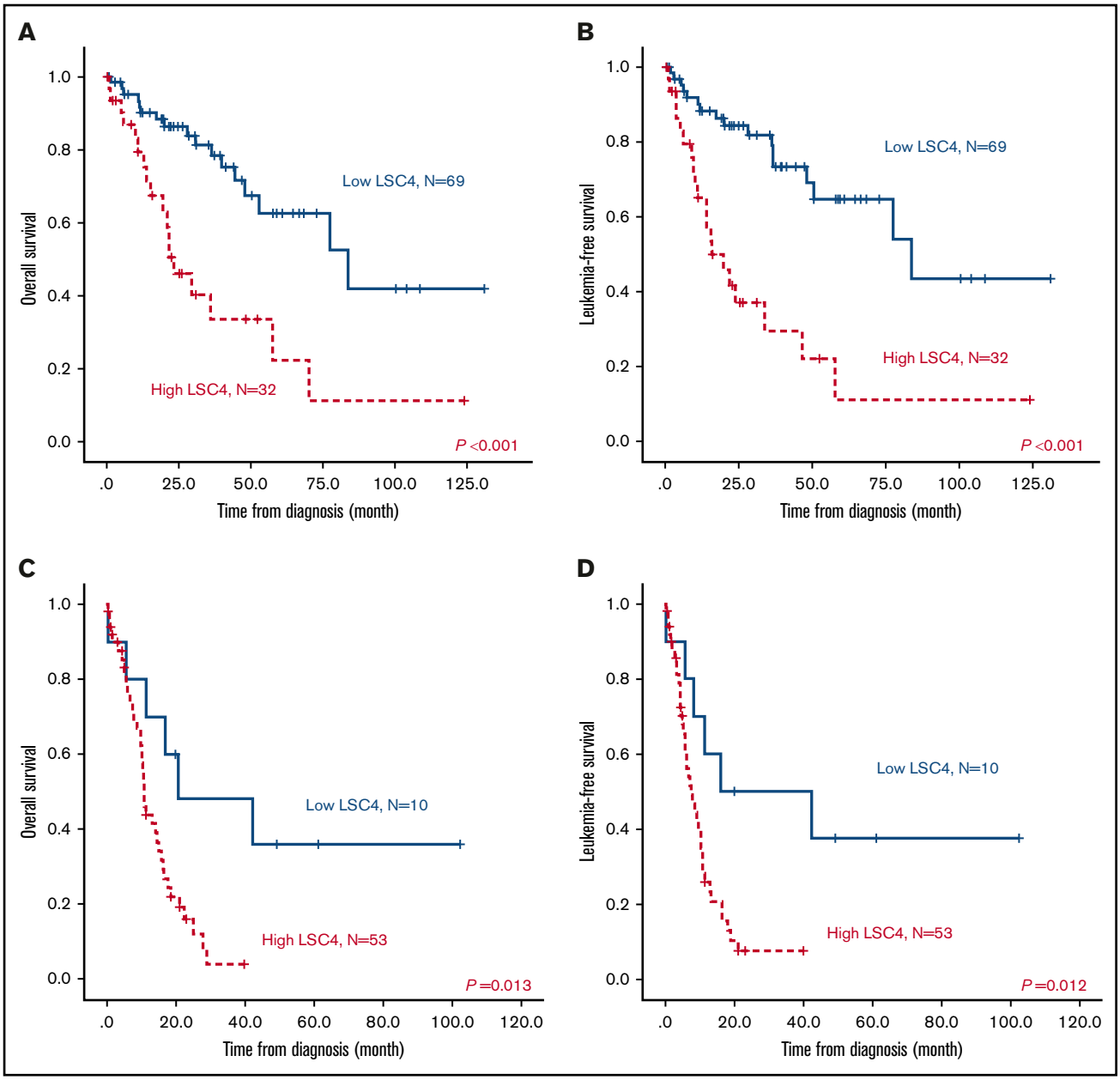

Figure 3. Kaplan-Meier plots stratified by LSC4 scores in IPSS-R lower- and higher-risk subgroups. Outcome of the 164 patients in the training cohort who had cytogenetic data at diagnosis (thus IRSS-R could be calculated). OS (A) and LFS (B) of patients in the IPSS-R lower-risk group (very low, low, and intermediate risk). OS (C) and LFS (D) of patients in IPSS-R higher-risk group (high and very high risk). Patients with higher LSC4 scores had worse clinical outcomes than those with lower scores in either IPSS-R subgroup.

scores had significantly shorter OS (Figure 5C; 37.3 months vs NR, $P = .01$) than those with lower scores.

Discussion

MDS is a heterogeneous disease with highly variable clinical presentations, survival, and rate of leukemia transformation among the patients. Traditional risk stratification systems like IPSS or IPSS-R have been widely adopted in the care of MDS patients for a long time. Compared with IPSS, the IPSS-R incorporated more parameters, including blast percentages, comprehensive cytogenetic classification, and the depth of cytopenias, with improved prediction power for prognosis assessment in MDS patients.[7] However, the clinical outcomes of MDS patients may vary even in the same IPSS-R risk groups. It is warranted to search for more prognostic markers for better risk stratification.

In this study, we tried to integrate expression of genes related to stemness into the prognostic model. In a recent report, Ng et al proposed a LSC17 scoring system that clearly predicted the outcomes of AML patients.[10] Since stemness of leukemia cells is a crucial factor of resistance to chemotherapy, it is reasonable that the expression profile of these LSC genes has significant impact on prognosis of AML patients.[5,10,13,14,25,26] MDS and

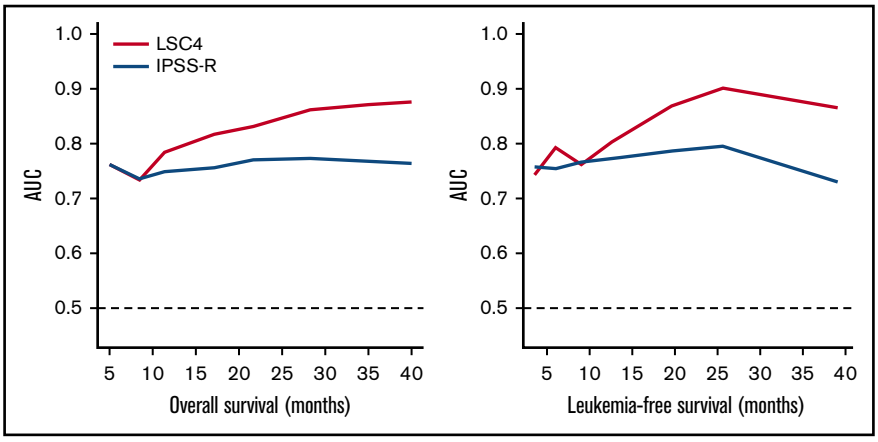

Figure 4. Time-dependent ROC curve analyses showing that the LSC4 score had better predictive power for OS and LFS than the IPSS-R. ROC curves were estimated by inverse probability of censoring weighting. AUC, area under the curve.

AML are both myeloid malignancies rising from stem/progenitor cells, and they share some clinical and biological features.[11,27-31] The aberrant stem and progenitor cell populations have some cellular features in common with normal HSCs, including sustained self-renewal and proliferation capacity.[28] They are responsible for disease initiation, transformation, and relapse and are also more resistant to chemotherapy.[28] Multiple reports have shown functionally defined subsets of AML stem cells using in vitro and xenograft model systems.[14,32-34] In contrast, while MDS has also long been considered as a stem cell disorder, the characterization of LSCs in MDS is still not clear yet. We hypothesize that expression levels of LSC genes have a significant impact on the prognosis of MDS patients, although the roles of these LSC gene signatures in prognostication in MDS patients have not been clarified yet. We first applied the LSC17 score proposed by Ng et al[10] to our MDS patient cohort but found that its correlation with clinical outcome was unsatisfactory. This might somewhat reflect the different nature of LSCs between MDS and AML.

In this study, we constructed a simple 4-LSC gene signature–based score for prediction of clinical outcomes, considering the expression levels and weights of only 4 LSC genes. We demonstrated that higher LSC4 scores predicted poorer prognosis for both OS and LFS. Patients with higher LSC4 score had worse OS and LFS than others in the same IPSS-R or 2016 WHO subgroup. These findings indicate the clinical heterogeneity in the same risk groups. Patients in each IPSS-R risk group could be further stratified into different prognostic groups based on this LSC4 scoring system. In addition, patients with higher LSC4 score consistently had poorer OS and LFS across subgroups, including patients receiving supportive care ($P < .001$ for both OS and LFS; supplemental Figure 4) or active treatment ($P < .001$ for both OS and LFS). These analyses indicated the prognostic power and clinical relevancy of the score, since more proactive follow-up and treatment strategies can be considered if a patient harbors a high LSC4 score, regardless of which risk group he or she was initially categorized in.

Table 2. Multivariate analysis for OS and LFS in 164 MDS patients who had cytogenetic data at diagnosis

Variable	OS, 95% CI				LFS, 95% CI			
	HR	Lower	Upper	*P*	HR	Lower	Upper	*P*
Age*	1.024	1.007	1.040	.005	1.012	0.997	1.028	.114
IPSS-R†	1.418	1.116	1.803	.004	1.388	1.091	1.766	.008
ASXL1	1.277	0.684	2.386	.443	1.875	1.037	3.388	.037
SRSF2	0.766	0.372	1.579	.471	0.710	0.355	1.420	.333
TP53	3.108	1.254	7.702	.014	3.239	1.299	8.077	.012
ZRSR2	1.382	0.683	2.794	.368	1.023	0.508	2.060	.949
Higher LSC4 score‡	3.452	1.875	6.356	<.001	3.792	2.076	6.929	<.001

$P < .05$ is considered statistically significant. Only variables with $P \leq .05$ in univariate analysis were incorporated into the multivariate Cox proportional hazard regression analysis.
CI, confidence interval; HR, hazard ratio.
*Continuous variable.
†IPSS-R risk groups: very good, good, intermediate, poor, and very poor.
‡High vs low LSC4 risk scores (median as cutoff).

The prognostic significance of the LSC4 scoring system was confirmed in 2 independent validation cohorts. In bivariate analysis, there was a moderate correlation between LSC4 scores and IPSS-R, which is predictable, since both prognostic systems could well stratify patients' outcome, yet the underlying mechanism warrants further studies. Meanwhile, a low correlation between LSC4 score and the presence of a complex karyotype further underlines the independent prognostic significance of LSC4 score. Although higher LSC scores were closely associated with high-risk mutations such as *RUNX1*, *ASXL1*, *SRSF2*, and *TP53* mutations, multivariate analysis proved high LSC score to be an unfavorable prognostic factor independent of the IPSS-R and mutations.

To our knowledge, this is the first study integrating LSC gene signatures to predict the prognosis in MDS patients. This 4-LSC gene signature–based score is concise yet powerful and easier for clinical applications than using 17 genes. This LSC4 scoring system would be helpful for identifying those with poorer prognosis in the rather heterogeneous group of patients, especially those with IPSS-R lower-risk MDS; in patients with IPSS-R lower-risk but high LSC4 scores, more aggressive treatment, rather than traditionally palliative care, might be warranted.

From the 17 LSC genes, we identified that expression of *LAPTM4B*, *NGFRAP1*, *EMP1*, and *CPXM1* is most relevant to the prognosis of MDS patients. *LAPTM4B* (the lysosome-associated protein transmembrane-4β gene) is considered as an oncogene, and its overexpression has been proven to promote the proliferation of various tumor cells, boost invasion and metastasis, resist apoptosis, initiate autophagy, and assist drug resistance.[35-38] Overexpression

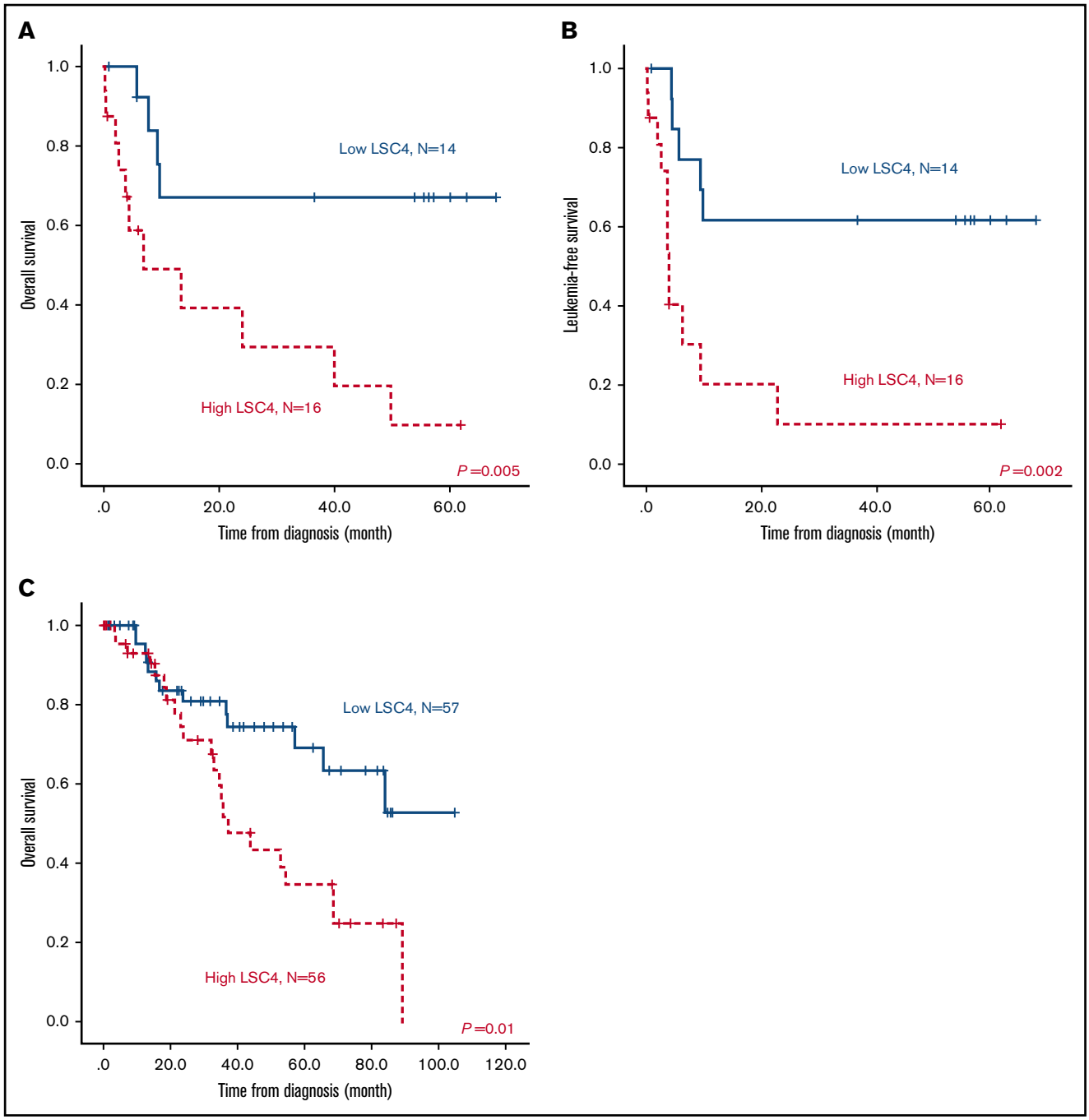

Figure 5. Kaplan-Meier plots of 2 independent validation cohorts stratified by LSC4 scores. OS (A) and LFS (B) of the 30 MDS patients in the internal validation cohort. Patients with higher LSC4 scores had shorter OS and LFS. (C) OS of the 113 MDS patients in an external validation cohort from GSE58831. Patients with higher LSC4 scores consistently had shorter OS.

of *LAPTM4B* in breast cancer cells results in resistance to anthracycline, and its knockdown can sensitize the drug response.[39] *NGFRAP1* (nerve growth factor receptor-associated protein 1) is an apoptosis-related gene, and its expression is downregulated in some solid organ malignancies and chronic lymphocytic leukemia.[40,41] Overexpression of *NGFRAP1* was reported to be correlated with an adverse outcome in AML patients.[42] *EMP1* (epithelial membrane protein 1) encodes proteins associated with membrane blebbing, cell proliferation, and squamous cell differentiation.[43-45] Its high expression was an independent predictor of poor outcome in pediatric leukemia,[46] while its roles in solid cancers are still not fully understood.[47-50] *CPXM1* (carboxypeptidase X, M14 family member 1) encodes a member of the carboxypeptidase family of proteins and is a positive regulator of adipogenesis, which may contribute to hyperplastic adipose tissue expansion via affecting extracellular matrix remodeling.[51] Recently, a gene expression

signature was developed to predict early molecular remission and long-term outcome in chronic myeloid leukemia patients and higher expression of *CPXM1* in this signature was positively correlated with early molecular remission failure rate in chronic-phase chronic myeloid leukemia patients on frontline imatinib.[52] However, the role of *CPXM1* in hematological and oncological malignancies is still largely unknown.[53] One interesting thing is that the directions of *NGFRAP1* and *CPXM1* in this LSC4 scoring system were different from those in the LSC17 equation for AML, implicating the different nature of LSCs between MDS and AML. Further studies of the underlying pathophysiology and mechanisms are worth exploring in the future. Overall, although these genes are regarded as LSC genes, their roles in MDS or other hematologic malignancies are still not fully understood, and further studies are needed to explore the functional pathways and pathogenesis of these genes in MDS.

There are limitations of this study. Firstly, we used array-based approaches rather than next-generation sequencing (NGS)–based methods to quantify gene expression. The advantages of NGS include higher sensitivity for genes with low expression and the broader dynamic range of expression levels in comparison with microarray. However, several reports have revealed that there is a strong concordance between microarray and RNA-sequencing data, and they have similar performance for the prediction of clinical end points.[54,55] Therefore, gene expression profiles generated by array-based methods are still valuable resources for biomarker discovery. Secondly, unlike mutations, which are clear-cut, standardizing the quantification of gene expression is a daunting task when it comes to applicability of this scoring system in clinical practice. To apply the scoring system in a clinical setting like *BCR-ABL1* levels in chronic myelogenous leukemia, there are still many problems to solve. Using novel technologies such as the NanoString nCounter system or targeted RNA sequencing, the LSC signature–based scoring system can be applied in clinical practice more easily. In conclusion, this is the first study integrating the LSC gene signature into risk stratification of MDS patients. This study provides a novel, powerful, and biologically significant scoring system, which serves as an independent prognostic factor for OS and LFS in MDS patients. This integrated prognostic system refines the prognostic prediction models and might guide the therapeutic decision and possible LSC-targeted therapy in the future. Lastly, although the reliability and reproducibility of this prognostic model had been validated by 2 independent cohorts, prospective studies with more standardization and more precise quantification methods, such as NGS or real-time polymerase chain reaction, are warranted.

Acknowledgments

The authors would like to acknowledge the service provided by Department of Laboratory Medicine, Department of Medical Research, and Division of Hematology, Department of Internal Medicine, NTUH.

This study was supported by grants from the Ministry of Health and Welfare, Taiwan (project number MOHW107-TDU-B-211-114009) and the Ministry of Science and Technology, Taiwan (project numbers MOST 107-2314-B-002-013 and MOST 108-2314-B-002-011).

Authorship

Contribution: Y.-H.W. and C.-C.L. were responsible for data collection and management, statistical analysis and interpretation, literature research, and manuscript writing; C.-Y.Y. was responsible for data management and statistical analysis; C.-L.H. assisted in statistical analysis; H.-A.H. and C.-H.T. were responsible for data collection and management; and 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.

ORCID profiles: Y.-H.W., 0000-0003-4483-5627; C.-C.L., 0000-0001-7160-2285; C.-Y.Y., 0000-0001-5052-9445; C.-L.H., 0000-0002-7447-8045.

Correspondence: Wen-Chien Chou, Department of Laboratory Medicine, National Taiwan University Hospital, No. 7, Chung-Shan S. Rd, Taipei City 10002, Taiwan; e-mail: wchou@ntu.edu.tw; and Hwei-Fang Tien, Department of Internal Medicine, National Taiwan University Hospital, No. 7, Chung-Shan S. Rd, Taipei City 10002, Taiwan; e-mail: hftien@ntu.edu.tw.

References

1. Heaney ML, Golde DW. Myelodysplasia. *N Engl J Med*. 1999;340(21):1649-1660.
2. Pellagatti A, Boultwood J. The molecular pathogenesis of the myelodysplastic syndromes. *Eur J Haematol*. 2015;95(1):3-15.
3. Tefferi A, Vardiman JW. Myelodysplastic syndromes. *N Engl J Med*. 2009;361(19):1872-1885.
4. Woll PS, Kjällquist U, Chowdhury O, et al. Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo [published corrections appear in *Cancer Cell*. 2014;25(6):861 and *Cancer Cell*. 2015;27(4):603-5]. *Cancer Cell*. 2014;25(6):794-808.
5. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer*. 2017;17(1):5-19.
6. Greenberg P, Cox C, LeBeau MM, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood*. 1997;89(6):2079-2088.
7. Greenberg PL, Tuechler H, Schanz J, et al. Revised international prognostic scoring system for myelodysplastic syndromes. *Blood*. 2012;120(12):2454-2465.
8. Fenaux P, Adès L. How we treat lower-risk myelodysplastic syndromes. *Blood*. 2013;121(21):4280-4286.
9. Gangat N, Patnaik MM, Tefferi A. Myelodysplastic syndromes: contemporary review and how we treat. *Am J Hematol*. 2016;91(1):76-89.
10. Ng SW, Mitchell A, Kennedy JA, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*. 2016;540(7633):433-437.

11. Chen J, Kao YR, Sun D, et al. Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level [published correction appears in *Nat Med*. 2019;25(3):529]. *Nat Med*. 2019;25(1):103-110.
12. Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. *Nat Rev Cancer*. 2005;5(4):275-284.
13. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA*. 2010;304(24):2706-2715.
14. Eppert K, Takenaka K, Lechman ER, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med*. 2011;17(9):1086-1093.
15. 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(1):5901.
16. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.
17. 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.
18. Simons A, Shaffer LG, Hastings RJ. Cytogenetic nomenclature: changes in the ISCN 2013 compared to the 2009 edition. *Cytogenet Genome Res*. 2013;141(1):1-6.
19. 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.
20. 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.
21. 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(7):e331.
22. 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.
23. 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.
24. 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.
25. Lin CC, Hsu YC, Li YH, et al. Higher HOPX expression is associated with distinct clinical and biological features and predicts poor prognosis in *de novo* acute myeloid leukemia. *Haematologica*. 2017;102(6):1044-1053.
26. Corces-Zimmerman MR, Majeti R. Pre-leukemic evolution of hematopoietic stem cells: the importance of early mutations in leukemogenesis. *Leukemia*. 2014;28(12):2276-2282.
27. Will B, Zhou L, Vogler TO, et al. Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations. *Blood*. 2012;120(10):2076-2086.
28. Shastri A, Will B, Steidl U, Verma A. Stem and progenitor cell alterations in myelodysplastic syndromes. *Blood*. 2017;129(12):1586-1594.
29. Pang WW, Pluvinage JV, Price EA, et al. Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes. *Proc Natl Acad Sci USA*. 2013;110(8):3011-3016.
30. Walter MJ, Shen D, Ding L, et al. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med*. 2012;366(12):1090-1098.
31. Tehranchi R, Woll PS, Anderson K, et al. Persistent malignant stem cells in del(5q) myelodysplasia in remission. *N Engl J Med*. 2010;363(11):1025-1037.
32. Hope KJ, Jin L, Dick JE. Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity. *Nat Immunol*. 2004;5(7):738-743.
33. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med*. 1997;3(7):730-737.
34. Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. 1994;367(6464):645-648.
35. Meng Y, Wang L, Chen D, et al. LAPTM4B: an oncogene in various solid tumors and its functions. *Oncogene*. 2016;35(50):6359-6365.
36. Xiao M, Yang S, Meng F, et al. LAPTM4B predicts axillary lymph node metastasis in breast cancer and promotes breast cancer cell aggressiveness in vitro. *Cell Physiol Biochem*. 2017;41(3):1072-1082.
37. Roy G, Roy P, Bhattacharjee A, et al. Expression signature of lysosomal-associated transmembrane protein 4B in hepatitis C virus-induced hepatocellular carcinoma. *Int J Biol Markers*. 2018;33(3):283-292.
38. Dong X, Tamura K, Kobayashi D, Ando N, Sumita K, Maehara T. LAPTM4B-35 is a novel prognostic factor for glioblastoma [published correction appears in *J Neurooncol*. 2017;132(2):305-306]. *J Neurooncol*. 2017;132(2):295-303.
39. Li Y, Zou L, Li Q, et al. Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat Med*. 2010;16(2):214-218.
40. Ruiz-Lafuente N, Alcaraz-García MJ, Sebastián-Ruiz S, et al. The gene expression response of chronic lymphocytic leukemia cells to IL-4 is specific, depends on ZAP-70 status and is differentially affected by an NFκB inhibitor. *PLoS One*. 2014;9(10):e109533.

41. Gao W, Li JZ, Chen SQ, Chu CY, Chan JY, Wong TS. BEX3 contributes to cisplatin chemoresistance in nasopharyngeal carcinoma. *Cancer Med*. 2017; 6(2):439-451.

42. Niavarani A, Herold T, Reyal Y, et al. A 4-gene expression score associated with high levels of Wilms Tumor-1 (WT1) expression is an adverse prognostic factor in acute myeloid leukaemia. *Br J Haematol*. 2016;172(3):401-411.

43. Wilson HL, Wilson SA, Surprenant A, North RA. Epithelial membrane proteins induce membrane blebbing and interact with the P2X7 receptor C terminus. *J Biol Chem*. 2002;277(37):34017-34023.

44. Ben-Porath I, Benvenisty N. Characterization of a tumor-associated gene, a member of a novel family of genes encoding membrane glycoproteins. *Gene*. 1996;183(1-2):69-75.

45. Marvin KW, Fujimoto W, Jetten AM. Identification and characterization of a novel squamous cell-associated gene related to PMP22. *J Biol Chem*. 1995; 270(48):28910-28916.

46. Ariës IM, Jerchel IS, van den Dungen RE, et al. EMP1, a novel poor prognostic factor in pediatric leukemia regulates prednisolone resistance, cell proliferation, migration and adhesion. *Leukemia*. 2014;28(9):1828-1837.

47. Sun GG, Wang YD, Cui DW, Cheng YJ, Hu WN. Epithelial membrane protein 1 negatively regulates cell growth and metastasis in colorectal carcinoma. *World J Gastroenterol*. 2014;20(14):4001-4010.

48. Ahmat Amin MKB, Shimizu A, Zankov DP, et al. Epithelial membrane protein 1 promotes tumor metastasis by enhancing cell migration via copine-III and Rac1. *Oncogene*. 2018;37(40):5416-5434.

49. Sun G, Zhao G, Lu Y, Wang Y, Yang C. Association of EMP1 with gastric carcinoma invasion, survival and prognosis. *Int J Oncol*. 2014;45(3): 1091-1098.

50. Liu C, Wei X, Li F, et al. The prognostic value of epithelial membrane protein 1 (EMP-1) in patients with laryngeal carcinoma. *Med Sci Monit*. 2017;23: 3795-3800.

51. Kim YH, Barclay JL, He J, et al. Identification of carboxypeptidase X (CPX)-1 as a positive regulator of adipogenesis. *FASEB J*. 2016;30(7):2528-2540.

52. Kok CH, Yeung DT, Lu L, et al. Gene expression signature that predicts early molecular response failure in chronic-phase CML patients on frontline imatinib. *Blood Adv*. 2019;3(10):1610-1621.

53. Kim YH, O'Neill HM, Whitehead JP. Carboxypeptidase X-1 (CPX-1) is a secreted collagen-binding glycoprotein. *Biochem Biophys Res Commun*. 2015; 468(4):894-899.

54. Raghavachari N, Barb J, Yang Y, et al. A systematic comparison and evaluation of high density exon arrays and RNA-seq technology used to unravel the peripheral blood transcriptome of sickle cell disease. *BMC Med Genomics*. 2012;5(1):28.

55. Zhang W, Yu Y, Hertwig F, et al. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol*. 2015;16(1):133.