A clinical measure of DNA methylation predicts outcome in de novo acute myeloid leukemia

Marlise R. Luskin,[1] Phyllis A. Gimotty,[2] Catherine Smith,[3] Alison W. Loren,[1] Maria E. Figueroa,[4] Jenna Harrison,[3] Zhuoxin Sun,[5] Martin S. Tallman,[6] Elisabeth M. Paietta,[7] Mark R. Litzow,[8] Ari M. Melnick,[9] Ross L. Levine,[6] Hugo F. Fernandez,[10] Selina M. Luger,[1] Martin Carroll,[1,11] Stephen R. Master,[12] and Gerald B.W. Wertheim[3,13]

[1]Division of Hematology and Oncology, Abramson Cancer Center, and [2]Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania (UPENN), Philadelphia, Pennsylvania, USA. [3]Department of Pathology, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [4]University of Michigan Medical School, Ann Arbor, Michigan, USA. [5]Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [6]Memorial Sloan Kettering Cancer Center, New York, New York, USA. [7]Montefiore Medical Center, New York, New York, USA. [8]Mayo Clinic, Rochester, Minnesota, USA. [9]Hematology and Oncology Division, Weill Cornell Medical College, New York, New York, USA. [10]Moffit Cancer Center, Tampa, Florida, USA. [11]Philadelphia Veterans Administration Medical Center, Philadelphia, Pennsylvania, USA. [12]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York, USA. [13]Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA.

BACKGROUND. Variable response to chemotherapy in acute myeloid leukemia (AML) represents a major treatment challenge. Clinical and genetic features incompletely predict outcome. The value of clinical epigenetic assays for risk classification has not been extensively explored. We assess the prognostic implications of a clinical assay for multilocus DNA methylation on adult patients with de novo AML.

METHODS. We performed multilocus DNA methylation assessment using xMELP on samples and calculated a methylation statistic (M-score) for 166 patients from UPENN with de novo AML who received induction chemotherapy. The association of M-score with complete remission (CR) and overall survival (OS) was evaluated. The optimal M-score cut-point for identifying groups with differing survival was used to define a binary M-score classifier. This classifier was validated in an independent cohort of 383 patients from the Eastern Cooperative Oncology Group Trial 1900 (E1900; NCT00049517).

RESULTS. A higher mean M-score was associated with death and failure to achieve CR. Multivariable analysis confirmed that a higher M-score was associated with death (P = 0.011) and failure to achieve CR (P = 0.034). Median survival was 26.6 months versus 10.6 months for low and high M-score groups. The ability of the M-score to perform as a classifier was confirmed in patients ≤ 60 years with intermediate cytogenetics and patients who achieved CR, as well as in the E1900 validation cohort.

CONCLUSION. The M-score represents a valid binary prognostic classifier for patients with de novo AML. The xMELP assay and associated M-score can be used for prognosis and should be further investigated for clinical decision making in AML patients.

Authorship note: S.R. Master and G.B.W. Wertheim contributed equally to this work.

Conflict of interest: M.R. Luskin, M. Carroll, S.R. Master, and G. Wertheim have the provisional patent application 62/040,821.

Submitted: February 26, 2016
Accepted: May 5, 2016
Published: June 16, 2016

Reference information:
JCI Insight. 2016;1(9):e87323. doi:10.1172/jci.insight.87323.

Introduction

The ability to predict therapeutic response is essential for improving care of patients with acute myeloid leukemia (AML). Recent efforts to understand AML variability have focused on the relationship between epigenetic abnormalities — including changes in DNA cytosine methylation — and AML phenotype (1–3).

While the mechanism by which aberrant methylation contributes to neoplasia remains incompletely understood, epigenetic alterations show significant correlation with patient outcome in several hematologic

Table 1. UPENN cohort: M-score by patient and AML characteristics

	Total Cohort (*n* = 166)					Age ≤ 60, Intermediate Cytogenetics (*n* = 82)				
	n	%	M-score (Mean)	95% CI	*P*	*n*	%	M-score (Mean)	95% CI	*P*
All Subjects	166	100	92.3	87.4–97.2	–	82	100	94.2	87.0–101.5	–
Age (years), diagnosis										
≤ 60	114	68.7	90.6	84.6–96.5	.297	–	–	–	–	–
> 60	52	31.3	96.2	87.4–105.0		–	–	–	–	–
Sex										
Male	98	59.0	90.8	84.2–97.5	.476	46	56.1	92.1	81.9–102.3	.516
Female	68	41.0	94.4	87.1–101.8		36	43.9	96.9	86.3–107.5	
WBC (× 10^9/l), diagnosis										
<100	131	78.9	93.5	87.7–99.4	.262	59	72.0	95.4	86.3–104.6	.604
≥100	35	21.1	87.8	79.3–96.2		23	28.0	91.2	79.5–102.9	
Cytogenetic risk group[A]										
Favorable	21	12.7	66.1	57.0–75.2	**.0002**	–	–	–	–	–
Intermediate	118	71.1	95.4	89.6–101.3		–	–	–	–	–
Unfavorable	27	16.3	99.1	86.9–111.3		–	–	–	–	–
FLT3-ITD										
Mutant	56	33.7	93.4	86.5–100.2	.742	37	45.1	92.9	83.5–102.3	.744
WT	110	66.3	91.8	85.2–98.4		45	54.9	95.3	84.3–106.3	
NPM1										
Mutant	58	34.9	93.3	84.0–96.9	.549	38	46.3	89.5	81.1–97.9	.212
WT	108	65.1	90.5	86.5–100.1		44	53.7	98.4	86.8–109.9	
NPM1$^+$ *FLT3*-ITD$^-$										
Yes	25	15.1	84.3	74.9–93.7	.084	17	20.7	83.1	74.1–92.1	**.025**
No	141	84.9	93.7	88.2–99.3		65	79.3	97.1	86.4–105.9	

[A]Medical Research Council criteria (26). Bold numbers indicate P values that are considered to be significant. AML, acute myeloid leukemia; *FLT3*-ITD, FMS-like kinase 3-internal tandem duplication; *NPM1*, nucleophosmin.

malignancies, including AML (2, 4–9). Despite the recognized relationship between DNA methylation and AML prognosis, clinical methylation assessment is not routine due to lack of a rapid, reliable assay that provides validated prognostic information. Recently, our group developed a microsphere-based assay for simultaneous assessment of DNA methylation status at multiple loci using commonplace clinical laboratory techniques (10, 11). This assay — xMELP — is an adaptation of the well-established HpaII tiny fragment enrichment by ligation mediated PCR (HELP) assay (12). We have shown that the technical parameters of xMELP — including precision, locus specificity, analytic sensitivity, and turn-around time — are appropriate for clinical use (10, 11).

In conjunction with a 17-locus xMELP assay, we developed a methylation-based risk score (M-score) for AML using random forest classification and demonstrated the association between M-score and overall survival (OS) on a limited cohort of AML patients (11). We hypothesized that M-score would independently predict clinical outcome in patients with de novo AML treated with intensive induction chemotherapy controlling for other prognostic markers.

Results

M-score is not associated with patient or sample characteristic. In total, 166 patients with de novo AML seen at UPENN were studied (Table 1). In response to 1 or 2 cycles of induction chemotherapy, 71% achieved complete remission (CR) and 38% were alive at 2 years (Supplemental Table 1; supplemental material available online with this article; doi:10.1172/jci.insight.87323DS1).

DNA methylation status at 17 previously identified prognostic loci was assessed by xMELP on a diagnostic sample from each patient (described in Supplemental Table 2), and the M-score statistic was calculated using our previously trained algorithm (11). The mean and median M-score for the UPENN cohort was 92.3 (95% CI, 87.4–97.2) and 91.4 (range, 30.8–197.3), respectively (Supplemental Figure 1). M-score was

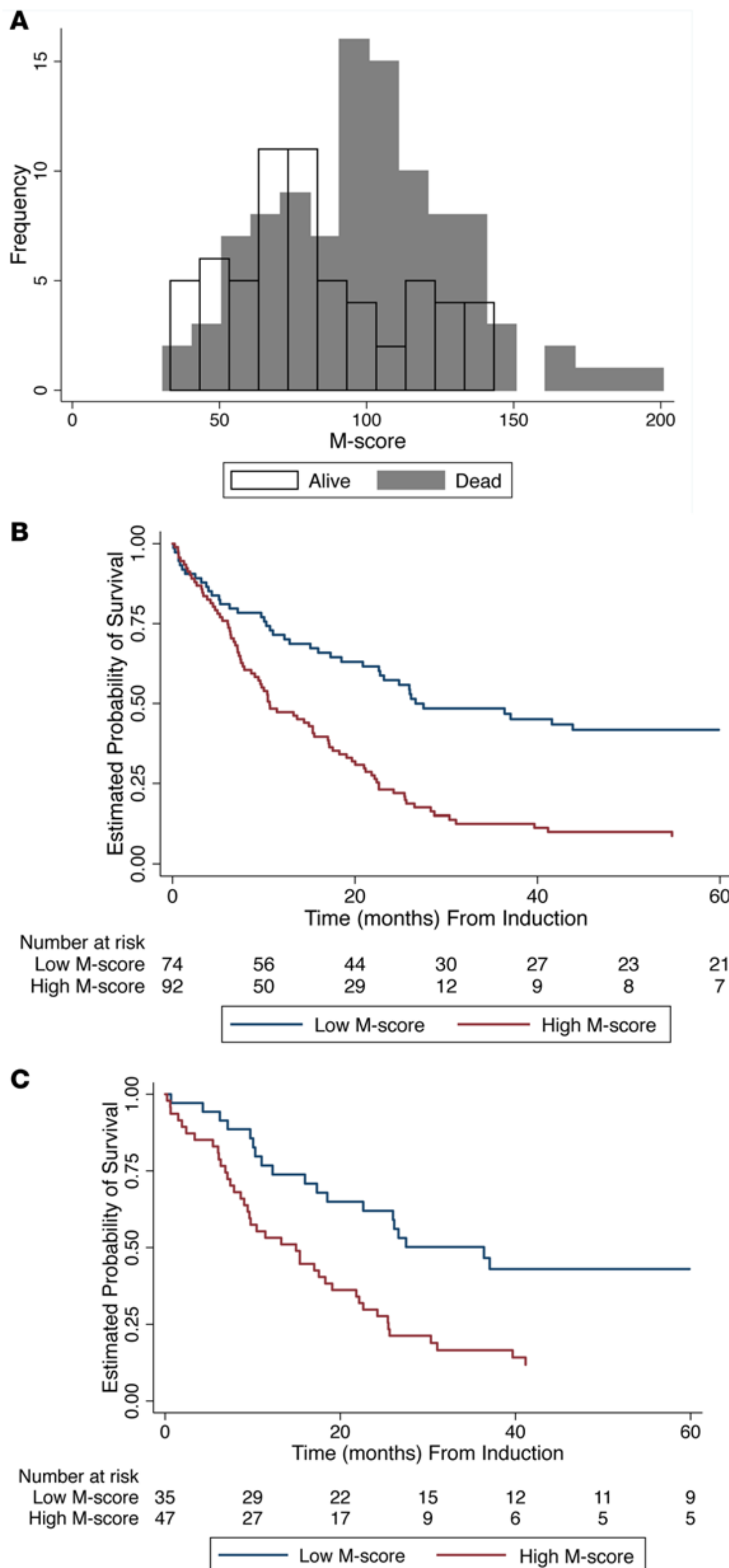

Figure 1. Overall survival by M-score. (**A**) Distribution of M-score by survival status at 2 years in the UPENN cohort (n = 163, n = 3 with unknown survival status at 2 years excluded). (**B** and **C**) Kaplan-Meier curves of overall survival in UPENN cohort. Subgroups are determined by the optimal M-score. Curves for the total cohort (**B**, n = 166) and patients ≤ 60 years with intermediate cytogenetics (**C**, n = 82; log-rank P = 0.001) are shown.

not significantly associated with patient age or sex (Table 1), specimen type (P = 0.809), or blast percentage (P = 0.415, Supplemental Figure 2).

M-score is significantly associated with AML clinical response. We next examined the relationship between M-score and both survival at 2 years and ability to achieve remission. The mean M-score for surviving patients was significantly lower than for deceased patients (81.8; 95% CI, 74.3–89.2, vs. 99.5; 95% CI 93.2–105.8, P = 0.0005, Figure 1A). Patients achieving CR also had a lower mean M-score compared with those who failed to achieve CR (86.8; 95% CI, 81.3–92.4, vs. 105.8; 95% CI, 96.5–115.0, P = 0.0005). Additionally, a univariate Cox survival analysis demonstrated that a 10-unit increase in the M-score was associated with a 10% increase in the hazard of death (P < 0.0001, Table 2) and a 20% increase in the odds of failing to achieve CR (Table 3).

M-score is associated with OS and failure to achieve CR in multivariable models. Given the relationship of genetic characteristics and outcome in AML, we assessed the association of M-score with AML genetic characteristics (Table 1). M-score was associated with cytogenetic risk; those with favorable cytogenetics had a lower mean M-score than both the intermediate and unfavorable cytogenetics (P < 0.0001 and P = 0.001, respectively) but there was no difference in mean M-score between intermediate and unfavorable groups (P = 1.0).

The M-score was not associated with established molecular markers of AML risk (*NPM1* and *FLT3*-ITD, Table 1) but was associated with mutations in *DNMT3A* and *IDH1*, genes involved in regulation of DNA methylation. M-score was not, however, associated with mutations in other methylation regulators, including *IDH2, TET2*, or *WT1* (refs. 13, 14, and Supplemental Table 3).

In multivariable Cox analysis, higher M-score and older age were associated with increased hazard of death, while *NPM1*$^+$ *FLT3*-ITD$^-$ status was associated with decreased hazard of death (Table 2). Interestingly, the reduced multivariable model for survival included only age and *NPM1*$^+$ *FLT3*-ITD$^-$ status in addition to M-score, and did not include cytogenetics (Table 2). Similarly, in a multivariable logistic analysis, M-score was associated with increased odds of failing to achieve CR. The reduced multivariable model for failure to achieve CR included M-score, age, and cytogenetics (Table 3). The association between M-score and hazard of death and odds of achieving CR was not significantly different between younger (≤60 years) and older (>60 years) patients.

Additional multivariable Cox and logistic regression analyses including *DNMT3A* and *IDH1* conducted on the subset of patients with extended molecular data (n = 136) confirmed that M-score remained significantly

Table 2. UPENN cohort: Cox model for overall survival (*n* = 166, events = 128)

	Univariate		Multivariable		Reduced		
Parameter	HR	*P*	Adj HR	*P*	Adj HR	95% CI	*P*
M-score[A]	1.1	**<.0001**	1.1	**.011**	1.1	1.0–1.2	**.002**
Age[B]	1.3	**<.0001**	1.3	**.001**	1.3	1.1–1.5	**<.0001**
Female	1.1	.461	1.1	.579	–	–	–
WBC, diagnosis[A]	1.0	.856	1.0	.210	–	–	–
Cytogenetic Risk[C] (reference unfavorable)							
Intermediate	0.7	.085	0.7	.226	–	–	–
Favorable	0.3	**.002**	0.5	.067	–	–	–
FLT3-ITD$^+$	1.4	.099	1.1	.733	–	–	–
NPM1$^+$ *FLT3*-ITD$^-$	0.5	**.017**	0.5	**.031**	0.5	0.3 – 0.8	**.011**

[A]Divided by 10; [B]10-year increase; [C]Medical Research Council criteria (2010). Hazard ratio, HR; *FLT3*-ITD, FMS-like kinase 3-internal tandem duplication; *NPM1*, nucleophosmin1. Bold numbers indicate P values that are considered to be significant.

associated with survival and achievement of CR (Supplemental Tables 4 and 5). Notably, *NPM1*$^+$ *FLT3*-ITD$^-$ was the only genetic marker included in both reduced Cox models, suggesting that M-score is more strongly associated with AML outcome than most individual genetic markers used in current clinical practice.

Risk classification using the M-score. After confirming the independent association of M-score with clinical outcome in AML, we designed a risk classifier for clinical application. Based on the maximization of the log-rank statistic, we determined the optimal binary M-score cut-point (Supplemental Figure 3). The Kaplan-Meier curves for the low and high M-score groups are shown in Figure 1B (characteristics of the 2 groups described in Supplemental Table 6). A high M-score was associated with an increased hazard of death alone (hazard ratio [HR] 2.5, $P < 0.0001$) and after adjustment for all other factors (HR 1.9, $P = 0.003$). Median survival for the low and high M-score groups was 26.6 and 10.6 months; 2-year OS was 56% (95% CI, 43.8–67.3) and 24% (95% CI, 15.2–33.1). The CR rate for low and high M-score group was 84% (95% CI, 75.2–92.4) and 61% (95% CI, 50.7–71.0; $P = 0.001$), respectively (Table 4). The M-score classifier identified groups with different outcome, regardless of whether patients did or did not receive allogeneic stem cell transplant (log-rank $P = 0.01$ and $P < 0.00001$, respectively).

AML patients aged ≤ 60 years with intermediate cytogenetics are in particular need of new tools for risk stratification; therefore, the binary M-score classifier was evaluated in this subgroup (described in Table 1). Standard prognostic factors were not different between patients with low and high M-scores within this subgroup (Supplemental Table 6). The M-score classifier defined groups with significantly different OS (log-rank $P = 0.001$; HR = 2.4, $P = 0.001$; Figure 1C and Table 4). Another major group in need of risk stratification is the group of patients with intermediate cytogenetics and *FLT3*-ITD mutation. Remarkably,

Table 3. UPENN cohort: Logistic model for failure to achieve complete remission (*n* = 166, events = 48)

	Univariate		Multivariable		Reduced		
Parameter	OR	*P*	Adj OR	*P*	Adj OR	95% CI	*P*
M-score[A]	1.2	**.001**	1.1	**.034**	1.2	1.0–1.3	**.012**
Age[B]	1.5	**.002**	1.5	**.007**	1.5	1.1–2.0	**.012**
Female	1.2	.642	1.3	.551	–	–	–
WBC, diagnosis[A]	1.0	.798	1.0	.329	–	–	–
Cytogenetic Risk[C] (reference unfavorable)							
Intermediate	0.4	.057	0.6	.236	0.5	0.1–1.1	.087
Favorable	0.1	**.008**	0.1	**.030**	0.1	0.0–1.0	**.047**
FLT3-ITD$^+$	0.9	.666	0.5	.168	–	–	–
NPM1$^+$ *FLT3*-ITD$^-$	0.4	.131	0.3	.081	–	–	–

[A]Divided by 10; [B]10-year increase; [C]Medical Research Council criteria (2010). OR, odds ratio; *FLT3*-ITD, FMS-like kinase 3-internal tandem duplication; *NPM1*, nucleophosmin1; Adj, adjusted. Bold numbers indicate P values that are considered to be significant.

Table 4. UPENN cohort: Clinical outcome by high versus low M-score

	Median OS (months)	2-Year OS (%)	CR Rate (%)
Total Cohort (n = 166)			
Low M-score	26.6	56%	84%
High M-score	10.6	24%	61%
Age ≤ 60 years, Intermediate Cytogenetics (n = 82)			
Low M-score	36.4	62%	91%
High M-score	14.9	30%	70%
Achieved CR (n = 118)			
Low M-score	43.9	67%	–
High M-score	17.2	36%	–

OS, overall survival; CR, complete remission.

Bold numbers indicate P values that are considered to be significant.

the M-score classifier defined sub-groups with significantly different OS within this traditionally high-risk subgroup (log-rank $P = 0.001$; HR = 3.1, $P = 0.002$; Supplemental Figure 4)

Finally, to investigate whether the ability of the M-score classifier to define groups with different OS was merely a reflection of achievement of CR, we restricted analysis to the 118 patients who had achieved CR. The M-score classifier continued to identify patients with significantly different OS (log-rank $P < 0.00001$; Supplemental Figure 5) with median survival 43.9 versus 17.2 months in low- and high-risk groups, respectively (Table 4). Additionally, we noted that patients with high M-score were more likely to need 2 cycles of induction chemotherapy than those with a low M-score in order to achieve CR (29% vs. 6%, $P = 0.001$).

Validation of the M-score classifier in the E1900 cohort. We sought to validate the M-score prognostic classifier for OS in the independent Eastern Cooperative Oncology Group Trial 1900 (E1900) cohort (Supplemental Tables 7–9). For these patients, the mean and median M-score were similar to the UPENN cohort: 98.2 (95% CI, 94.1–102.3) and 91.8 (range, 20.0–204.6), respectively (Supplemental Figure 1). Also similar to the UPENN cohort, the M-score was associated with survival on multivariable analysis ($P < 0.0001$), while the association with achievement of CR was marginally significant ($P = 0.076$).

The binary prognostic classifier derived in the UPENN cohort identified E1900 subgroups with different OS (log-rank $P < 0.00001$, Figure 2). The median OS in patients in the low M-score group was 29.5 months versus 12.6 months for those with high M-score. Among patients with intermediate cytogenetics ($n = 249$), OS was also significantly different (log-rank $P = 0.0003$), with median OS of 32.3 months versus 14.1 months in the low and high M-score groups, respectively. Additionally, we found that patients with low M-scores had better OS compared with those with high M-scores in the following subgroups: patients who had intermediate cytogenetics with a FLT3-ITD mutation (log-rank $P = 0.003$), younger and older patients (<50 years, log-rank $P = 0.0015$; ≥50 years, log-rank $P = 0.0015$), and in recipients of both standard and high dose anthracycline induction (standard-dose, log-rank $P < 0.00001$; high-dose, log-rank $P = 0.015$).

Since a primary objective of E1900 was to assess the impact of daunorubicin dose on AML outcome, the impact of treatment on patients with low and high M-scores was assessed. We found that high-dose daunorubicin was beneficial for patients with high M-scores (log-rank $P = 0.001$) but not for those with low M-scores (log-rank $P = 0.328$; Supplemental Figure 6).

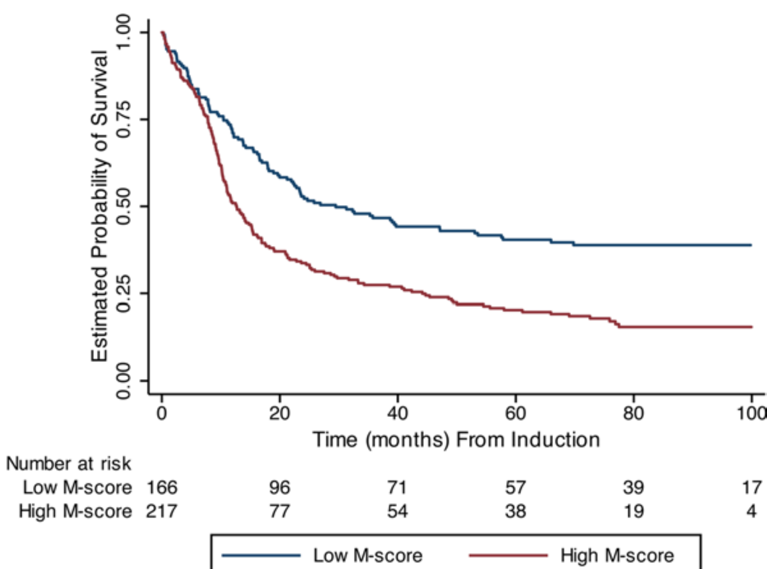

Figure 2. Kaplan-Meier curves of overall survival in E1900 cohort. Subgroups are determined by the optimal M-score (n = 383; log-rank $P < 0.00001$).

Discussion

Multivariable models incorporating standard AML prognostic characteristics have only a modest ability to predict clinical outcome, and the addition of molecular markers adds little to these models (15–17). To improve AML prognostication, we developed a clinically applicable assay for multilocus methylation assessment (xMELP), along with a corresponding M-score (10, 11). We now demonstrate that the M-score is robustly associated with CR and OS in both univariate and multivariable models in multiple independent AML cohorts. Remarkably, in the UPENN cohort, a reduced model for OS is based solely on the M-score, patient age, and *NPM1*$^{+}$ *FLT3*-ITD^{-} status; other information, including cytogenetics and presence of a *FLT3*-ITD mutation, provided little additional prognostic value. We also explored the association between M-score and recurrent intragenic mutations in the subgroup of UPENN patients with available extended mutation profiles obtained by next-generation sequencing. Mean M-score was higher in *DNMT3A* mutant patients and in the small group of *IDH1* mutant patients, but multivariable analyses that included *DNMT3A* and *IDH1* confirmed the independent association of M-score and clinical outcomes.

Determining the mutational profile in AML will remain important to clinical care, particularly in settings where mutations are able to predict response to targeted agents; however, clinical use of the xMELP assay and associated M-score may decrease the need for comprehensive genetic testing for risk stratification at diagnosis (18–20). Our reduced multivariable models indicate that the M-score has a stronger association with clinical outcome than many established prognostic factors, including cytogenetics and *FLT3*-ITD status, as well as other genetic lesions assessed by next-generation sequencing analysis.

Cox regression analyses showed strong association between the M-score and clinical outcomes; however, it is difficult to apply continuous measures of association in clinical practice. The binary M-score classifier, which we validated in multiple clinically important subgroups and a large independent cohort, clearly enhances the usefulness of the M-score for practicing clinicians. Additionally, the different responses to daunorubicin seen in M-score–defined groups suggests that M-score may correlate with chemoresistance and identify patients that could benefit from high-dose chemotherapy.

It is important to recognize that the loci contributing to the M-score do not account for all sites subject to aberrant methylation in AML. These specific loci in combination represent a marker of prognosis rather than an explanatory model of AML biology. Previous studies have indicated that epigenetic dysregulation of multiple cellular activities, such as repression of tumor suppressors and DNA repair enzymes by DNA methylation, contribute to therapy response (21–23). Thus, M-score may reflect the overall epigenetic state of the leukemic cells, and tumors with high M-scores may be epigenetically predisposed to chemoresistance. Importantly, we showed that M-score is not correlated to blast percentage, suggesting that the assay reflects abnormal methylation in both blasts and more differentiated myeloid cells derived from leukemic precursors. This finding is consistent with our previous studies demonstrating that Ficoll centrifugation does not alter M-score (11), and it shows that M-score provides representative prognostic information regardless of the blast percentage in the specific sample submitted to the clinical lab.

The studied cohorts include only patients with de novo AML treated with induction chemotherapy, so our conclusions are currently limited to similarly treated patients. The prognostic value of the M-score for patients with AML arising in the setting of prior chemotherapy or myelodysplasia, or those treated with nonintensive regimens including hypomethylating agents, are areas of further research. Additionally, we have no information regarding the association between M-score and other prognostic markers, including minimal residual disease status (24). Of note, we do not have sufficient power in our cohort to assess whether allogeneic stem cell transplant alters M-score–based risk, which is clearly an essential area of future investigation.

In summary, the M-score provides valuable information in the clinical setting regarding the likelihood of long-term survival after AML induction. Those patients predicted to have poor outcomes based on M-score may be better served with more intensive postremission treatment or enrollment on a clinical trial.

Methods

Study population and patient samples

UPENN cohort. We performed xMELP (described below) on 166 AML samples collected at UPENN. Samples were selected from consecutive patients with de novo AML (25) who consented to donation

of a diagnostic sample to the Hematologic Malignancies Tissue Bank of the University of Pennsylvania between 2001 and 2012 and who consented to review of their medical records; whose banked sample yielded adequate quality DNA for xMELP analysis, for whom standard molecular (FLT3-ITD and NPM1) and cytogenetic studies were performed; and who had undergone induction with an anthracycline and cytarabine (see below). Of these patients, 66 had been included in a preliminary report describing the association of M-score with clinical outcome (11). Cytogenetic risk was classified according to the Medical Research Council criteria (26). *FLT3*-ITD and *NPM1* status was assessed in a CLIA-certified lab and classified as mutant or WT. For 136 patients, extensive molecular information was available from next-generation sequencing of 33 genes associated with hematologic malignancies (27, 28). Average read depth was 3,000×, minimal depth was 250×, and reporting frequency cutoff for variants was 5%. Mutations were classified as pathogenic, likely disease-associated, variant of uncertain significance (VUS), or likely benign based on review of publically available data. Only pathogenic or likely disease-associated mutations were considered abnormal. Patient and disease characteristics, treatment, and outcomes were obtained from medical records. Median follow-up was 68.1 months (range, 1.4–150.2) among 38 survivors and 10.5 months (range, 0.1–95.2) among those (n = 128) deceased.

E1900 cohort. The validation cohort was composed of 383 patients who enrolled on E1900 between 2002 and 2008 who had available DNA methylation, genetic, and clinical data. Methylation data is publically available (Gene Expression Omnibus [GEO] repository accession number GSE24505 [http://www.ncbi.nlm.nih.gov/geo]) (13). Patients with indeterminate cytogenetics were analyzed with intermediate-risk patients. Median follow-up was 83.2 months (range, 0.8–120.4) among 108 survivors and 11.0 months (range, 0.2–77.5) among those (n = 275) deceased.

Samples, xMELP, and the M-score

The xMELP assay was performed on UPENN samples as described (10, 11). Briefly, JHpaII 12XXXX primer (diluted to 3 optical density [260] units [OD] per ml, CGCCTGTTCAT) and JHpaII 24XXXX primer (6 OD/ml, CGACGTCGACTATCCATGAACAGG) were annealed by heating to 95°C for 3 minutes and slowly cooling to 25°C. Approximately 1–3 million cryopreserved cells from AML patients were thawed and washed once with PBS. Genomic DNA was extracted with Qiagen Gentra Puregene DNA isolation kit. Genomic DNA (500 ng) was then digested with either HpaII or MspI restriction enzymes and simultaneously ligated to the annealed primers by mixing genomic DNA with 7.5 μl of annealed primers, 2U T4 DNA ligase (Invitrogen), 0.5 μl ATP (100 mmol/L, pH 7.0, New England Biolabs), or either 4U of MspI (New England Biolabs) or 2U of HpaII (New England Biolabs) in a total reaction volume of 50 μl. After overnight incubation at 25°C, the reaction was diluted with 450 μl of ddH_2O, and 10 μl of the diluted DNA was subjected to PCR using 1 μl of JHpaII 24XXXX primer (6 OD/ml) and 1 μl Native Taq polymerase (ThermoFisher) in 50 μl total volume. Cycling conditions were 72°C for 10 minutes once, followed by 20 cycles of 95°C for 30 seconds and 72°C for 3 minutes, and a final 72°C extension for 10 minutes. PCR products (8 μl) were then hybridized to specific fluorescent microspheres following manufacturer's protocol (Quantigene Plex 2.0 Assay, Affymetrix). Following hybridization, microspheres were analyzed on a FlexMAP 3D instrument running XPONENT software (Luminex). As our studies have shown, xMELP is not significantly affected by freezing samples or Ficoll centrifugation; these variables were not standardized in the assay (11).

M-score of each sample was determined using the random forest classification algorithm previously trained on an independent cohort of 344 AML samples collected by the Dutch-Belgian Hemato-Oncology Cooperative Group (HOVON) (R-scripts for M-score derivation are described and publicly available) (11). For E1900 samples, HELP-derived methylation data from BM samples were transformed to MELP-associated values using previously described regression coefficients (10).

Treatment and definitions

UPENN cohort. The induction chemotherapy regimen in all cases included an anthracycline and cytarabine (Supplemental Table 1). Patients with residual leukemia at nadir BM assessment were frequently retreated with an anthracycline-based regimen or high-dose cytarabine at clinician discretion. Endpoints were OS and CR. OS was time from induction chemotherapy to death from any cause. For living patients, OS times were censored at last follow-up. CR was defined as morphologic leukemic-free state on BM examination

after 1 or 2 cycles of induction chemotherapy (assessment required within 90 days of induction) (25).

E1900 cohort. The treatment schema and endpoint definitions for the E1900 cohort have been previously described (29). E1900 was a randomized trial of high-dose versus standard-dose daunorubicin that accrued patients aged ≤ 60 from 2002–2008 (NCT00049517).

Statistics

Continuous variables were summarized by median and range, and categorical variables by count and relative frequency. Comparisons of M-score between groups of AML patients was done by the parametric unpaired 2-sample t test (adjusted using Satterthwaite's method when variances unequal) and ANOVA test (for comparing ≥ 2 groups). The association between M-score and blast percentage was assessed by Pearson's correlation coefficient.

Univariate and multivariable logistic models were used to assess the association of M-score with response to induction chemotherapy (failure to achieve CR) alone and controlling for age, sex, WBC count at diagnosis, cytogenetics, and molecular status. Survival distributions for OS were computed using the Kaplan-Meier method and were compared using the log-rank statistic. Univariate and multivariable Cox regression analyses were used to examine the association between M-score and OS controlling for the same covariates. Backward selection was used in multivariable logistic and Cox models to develop the most parsimonious model.

An optimal cut-point for M-score was determined by identifying the cut-point that maximized the log-rank statistic between high and low M-score groups. P values were considered significant when $P < 0.05$ (2-sided). Analyses were performed using Stata 12.1 (StataCorp LP).

Study approval

IRB approval was obtained from UPENN and the Eastern Cooperative Oncology Group.

Author contributions

MR Luskin, PAG, AWL, MC, SRM, and GBWW designed the study. CS, JH, GBWW, and MEF conducted experiments. MR Luskin, PAG, ZS, and SRM conducted data analysis. AWL, MST, MR Litzow, HFF, and SML contributed patients. MR Luskin and GBWW wrote the manuscript. All authors contributed to interpretation of data and review of the manuscript. EMP, AMM, and RLL conducted data analysis.

Acknowledgments

M.R. Luskin received funding from NIH (T32 CA 09679-22); G.B.W. Wertheim received funding from Perelman School of Medicine Hematologic Malignancies Translational Center of Excellence Pilot Project Grant, American Cancer Society (IRG-78-002-35), and UPENN Institute for Translational Medicine and Therapeutics (ITMAT) through a grant from the National Center for Research Resources, grant UL1RR02434 (now at the National Cancer for Advancing Translational Sciences, grant UL1TR000003); M. Carroll received funding from NIH (1R21CA185365-01 and 1R01CA149566-01A1) and Veterans Affairs Administration (1I01BX000918-01); M. Carroll and P.A. Gimotty received funding from NIH (5R21CA185365-02); and S.R. Master received funding from NIH (1R21CA185365-01). M.R. Luskin, M. Carroll, S.R. Master, G.B.W. Wertheim obtained provisional patent application 62/040,821.

Address correspondence to: Stephen R. Master, Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, 525 E. 68th St., Suite F-715, Box 79, New York, New York 10065, USA. Phone: 212.746.3265; E-mail: srm9012@med.cornell.edu. Or to: Gerald B.W. Wertheim, Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, 5199b Main Building, 324 S. 34th St., Philadelphia, Pennsylvania 19104, USA. Phone: 267.970.5918; E-mail: wertheimg@email.chop.edu.

M.R. Luskin's present address is Department of Medicine, Brigham and Women's Hospital, Boston, MA.

This work was partially presented in abstract form at ASH Annual Meeting, San Francisco, California, USA, on December 6–9, 2014; at the USCAP Annual Meeting, Boston, Massachusetts, USA, on March 21–27, 2015; and at the ASH Annual Meeting, Orlando, Florida, USA, on December 4–7, 2015.

1. Akalin A, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS Genet*. 2012;8(6):e1002781.
2. Figueroa ME, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell*. 2010;17(1):13–27.
3. Schoofs T, Berdel WE, Müller-Tidow C. Origins of aberrant DNA methylation in acute myeloid leukemia. *Leukemia*. 2014;28(1):1–14.
4. Bullinger L, et al. Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood*. 2010;115(3):636–642.
5. Nordlund J, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol*. 2013;14(9):r105.
6. Borssén M, et al. Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. *PLoS One*. 2013;8(6):e65373.
7. Jiang D, et al. The diagnostic value of DNA methylation in leukemia: a systematic review and meta-analysis. *PLoS One*. 2014;9(5):e96822.
8. Marcucci G, et al. Epigenetics meets genetics in acute myeloid leukemia: clinical impact of a novel seven-gene score. *J Clin Oncol*. 2014;32(6):548–556.
9. Deneberg S, et al. Gene-specific and global methylation patterns predict outcome in patients with acute myeloid leukemia. *Leukemia*. 2010;24(5):932–941.
10. Wertheim GB, et al. Microsphere-based multiplex analysis of DNA methylation in acute myeloid leukemia. *J Mol Diagn*. 2014;16(2):207–215.
11. Wertheim GB, et al. Validation of DNA methylation to predict outcome in acute myeloid leukemia by use of xMELP. *Clin Chem*. 2015;61(1):249–258.
12. Figueroa ME, Melnick A, Greally JM. Genome-wide determination of DNA methylation by Hpa II tiny fragment enrichment by ligation-mediated PCR (HELP) for the study of acute leukemias. *Methods Mol Biol*. 2009;538:395–407.
13. Figueroa ME, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*. 2010;18(6):553–567.
14. Rampal R, et al. DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell Rep*. 2014;9(5):1841–1855.
15. Walter RB, et al. Resistance prediction in AML: analysis of 4601 patients from MRC/NCRI, HOVON/SAKK, SWOG and MD Anderson Cancer Center. *Leukemia*. 2015;29(2):312–320.
16. Krug U, et al. Complete remission and early death after intensive chemotherapy in patients aged 60 years or older with acute myeloid leukaemia: a web-based application for prediction of outcomes. *Lancet*. 2010;376(9757):2000–2008.
17. Walter RB, et al. Effect of genetic profiling on prediction of therapeutic resistance and survival in adult acute myeloid leukemia. *Leukemia*. 2015;29(10):2104–2107.
18. Man CH, et al. Sorafenib treatment of FLT3-ITD(+) acute myeloid leukemia: favorable initial outcome and mechanisms of subsequent nonresponsiveness associated with the emergence of a D835 mutation. *Blood*. 2012;119(22):5133–5143.
19. Pemmaraju N, Kantarjian H, Andreeff M, Cortes J, Ravandi F. Investigational FMS-like tyrosine kinase 3 inhibitors in treatment of acute myeloid leukemia. *Expert Opin Investig Drugs*. 2014;23(7):943–954.
20. Stein EM AJ, et al. AG-221, an oral, selective, first-in-class, potent inhibitor of the IDH2 mutant metabolic enzyme, induces durable remissions in a phase I study in patients with IDH2 mutation positive advanced hematologic malignancies. Presented at: 56th ASH Annual Meeting and Exposition; December 7, 2014; San Francisco, California, USA. Abstract 115.
21. Grövdal M, et al. Negative effect of DNA hypermethylation on the outcome of intensive chemotherapy in older patients with high-risk myelodysplastic syndromes and acute myeloid leukemia following myelodysplastic syndrome. *Clin Cancer Res*. 2007;13(23):7107–7112.
22. Glasspool RM, Teodoridis JM, Brown R. Epigenetics as a mechanism driving polygenic clinical drug resistance. *Br J Cancer*. 2006;94(8):1087–1092.
23. Figueroa ME, et al. MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood*. 2009;114(16):3448–3458.
24. Chen X, et al. Relation of clinical response and minimal residual disease and their prognostic impact on outcome in acute myeloid leukemia. *J Clin Oncol*. 2015;33(11):1258–1264.
25. Cheson BD, et al. Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. *J Clin Oncol*. 2003;21(24):4642–4649.
26. Grimwade D, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*. 2010;116(3):354–365.
27. Luskin MR, et al. NPM1 mutation is associated with leukemia cutis in acute myeloid leukemia with monocytic features. *Haematologica*. 2015;100(10):e412–e414.
28. Sehgal AR, et al. DNMT3A Mutational Status Affects the Results of Dose-Escalated Induction Therapy in Acute Myelogenous Leukemia. *Clin Cancer Res*. 2015;21(7):1614–1620.
29. Fernandez HF, Sun Z, Yao X, et al. Anthracycline dose intensification in acute myeloid leukemia. *N Engl J Med*. 2009;361(13):1249–1259.