Leukemia (2013) **27**, 2023–2031

www.nature.com/leu

ORIGINAL ARTICLE

A stem cell-like gene expression signature associates with inferior outcomes and a distinct microRNA expression profile in adults with primary cytogenetically normal acute myeloid leukemia

KH Metzeler[1], K Maharry[1,2], J Kohlschmidt[1,2], S Volinia[1], K Mrózek[1], H Becker[1], D Nicolet[1,2], SP Whitman[1], JH Mendler[1], S Schwind[1], A-K Eisfeld[1], Y-Z Wu[1], BL Powell[3], TH Carter[4], M Wetzler[5], JE Kolitz[6], MR Baer[7], AJ Carroll[8], RM Stone[9], MA Caligiuri[1], G Marcucci[1,10] and CD Bloomfield[1,10]

Acute myeloid leukemia (AML) is hypothesized to be sustained by self-renewing leukemia stem cells (LSCs). Recently, gene expression signatures (GES) from functionally defined AML LSC populations were reported, and expression of a 'core enriched' (CE) GES, representing 44 genes activated in LCSs, conferred shorter survival in cytogenetically normal (CN) AML. The prognostic impact of the CE GES in the context of other molecular markers, including gene mutations and microRNA (miR) expression alterations, is unknown and its clinical utility is unclear. We studied associations of the CE GES with known molecular prognosticators, miR expression profiles, and outcomes in 364 well-characterized CN-AML patients. A high CE score (CE^{high}) associated with *FLT3*-internal tandem duplication, *WT1* and *RUNX1* mutations, wild-type *CEBPA* and *TET2*, and high *ERG*, *BAALC* and *miR-155* expression. CE^{high} patients had a lower complete remission (CR) rate ($P = 0.003$) and shorter disease-free (DFS, $P < 0.001$) and overall survival (OS, $P < 0.001$) than CE^{low} patients. These associations persisted in multivariable analyses adjusting for other prognosticators (CR, $P = 0.02$; DFS, $P < 0.001$; and OS, $P < 0.001$). CE^{high} status was accompanied by a characteristic miR expression signature. Fifteen miRs were upregulated in both younger and older CE^{high} patients, including miRs relevant for stem cell function. Our results support the clinical relevance of LSCs and improve risk stratification in AML.

Leukemia (2013) **27,** 2023–2031; doi:10.1038/leu.2013.181

Keywords: acute myeloid leukemia; leukemic stem cells; gene expression profiling; prognostication; gene mutations

INTRODUCTION

According to the cancer stem cell hypothesis, acute myeloid leukemia (AML) is organized hierarchically with the bulk of AML blasts originating from a distinct population of leukemia-initiating cells or leukemic stem cells (LSCs).[1] LSCs are defined by their unique capacity for unlimited self-renewal, and can be identified through their ability to cause long-term engraftment in immunodeficient mice.[2] Such xenotransplantation assays suggest that only a small fraction ($\sim$1 in 10^4–10^6 cells) of the leukemic cell population have LSC properties.[3] Studies on the clinical relevance of LSCs in human AML are hindered by the fact that there are no surface markers that reliably discriminate LSCs from non-LSC leukemic blasts. Instead, LSCs seem to be phenotypically heterogeneous and are enriched in, but not restricted to, certain defined cell populations such as the $CD34^+CD38^-$ subset.[4]

Recently, Eppert *et al.*[5] sorted primary human AML specimens into several fractions based on CD34 and CD38 surface antigen expression, and defined the frequency of LSCs in each cell fraction using a sensitive xenograft assay. By comparing gene expression profiles between cell populations containing such functionally defined LSCs and populations lacking detectable stem cell activity, they then derived an LSC-related gene expression signature (GES) comprising 42 genes. A comparison of this LSC GES with a signature derived from normal hematopoietic stem cells then led to a 'core enriched' (CE) hematopoietic stem cell-LSC signature, consisting of 44 stem cell-associated genes highly expressed in LSCs.[5] Patients with cytogenetically normal (CN) AML and CE signature gene expression above the median had worse survival than patients with low expression of these genes. However, it remained unknown whether expression of the CE signature associates with, and is potentially driven by, other molecular prognosticators including gene mutations and deregulated expression of microRNAs (miRs). The aim of our study was to clarify whether this LSC-like GES mainly is a surrogate of other, already known molecular alterations, or whether it provides additional prognostic information even when these other risk markers are taken into consideration. Therefore, we studied associations of the CE GES with known clinical and molecular prognosticators, miR-expression signatures, and outcomes in a comprehensively characterized cohort of CN-AML patients, and evaluated the prognostic relevance of the CE signature in the context of other recently described molecular markers.

[1]The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA; [2]Alliance for Clinical Trials in Oncology Statistics and Data Center, Mayo Clinic, Rochester, MN, USA; [3]Comprehensive Cancer Center of Wake Forest University, Winston-Salem, NC, USA; [4]University of Iowa, Iowa City, IA, USA; [5]Roswell Park Cancer Institute, Buffalo, NY, USA; [6]Monter Cancer Center, Hofstra North Shore-Long Island Jewish School of Medicine, Lake Success, NY, USA; [7]Greenebaum Cancer Center, University of Maryland, Baltimore, MD, USA; [8]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, USA and [9]Dana-Farber Cancer Institute, Boston, MA, USA. Correspondence: Prof. G Marcucci, The Ohio State University, Comprehensive Cancer Center, 410 Biomedical Research Tower, 460 West 12th Ave, Columbus, OH 43210, USA. E-mail: guido.marcucci@osumc.edu or Prof. CD Bloomfield, The Ohio State University, Comprehensive Cancer Center, 1216 James Cancer Hospital, 300 West 10th Ave, Columbus, OH 43210, USA. E-mail: clara.bloomfield@osumc.edu
[10]These authors contributed equally to this work.
Received 29 March 2013; revised 29 May 2013; accepted 7 June 2013; accepted article preview online 14 June 2013; advance online publication, 9 July 2013

MATERIALS AND METHODS

Patients

We studied 364 patients with primary CN-AML, including 164 younger patients aged 18–59 years and 200 older patients aged 60–83 years, who were enrolled on Cancer and Leukemia Group B (CALGB)/Alliance for Clinical Trials in Oncology (Alliance) protocols 20202, 8461 and 9665. The patients received cytarabine/daunorubicin-based first-line therapy on CALGB/Alliance trials (see Supplementary Information for details on treatment protocols). Per protocol, no patient received allogeneic stem cell transplantation in first complete remission (CR). Study protocols were in accordance with the Declaration of Helsinki and approved by the Institutional Review Boards at each center, and all patients provided written informed consent.

Genetic analyses

Cytogenetic analyses were performed in CALGB/Alliance-approved institutional laboratories and confirmed by central karyotype review, and the diagnosis of normal cytogenetics was based on $\geqslant$20 analyzed metaphase cells in bone marrow specimens.[6] Patients were characterized for *FLT3*-internal tandem duplications (*FLT3*-ITD);[7] mutations in *NPM1*,[8] *CEBPA*,[9] *WT1*,[10] *RUNX1*,[11] *TET2*,[12] *DNMT3A*,[13] *ASXL1*[14] and *IDH1/IDH2*;[15] *FLT3*-tyrosine kinase domain (*FLT3*-TKD) mutations;[16] *MLL*-partial tandem duplications (*MLL*-PTD);[17,18] and expression of *BAALC*,[19] *ERG*,[19] *MN1*[20] and *miR-155*,[21] as previously reported. All molecular analyses were centrally performed at The Ohio State University Comprehensive Cancer Center (OSU-CCC).

Gene and miR expression profiling and calculation of the CE gene expression score

Gene and miR expression profiling was performed on pretreatment marrow or blood samples using Affymetrix HG-U133 plus 2.0 and OSU-CCC custom microarrays, respectively (see Supplementary Information for details on microarray analyses). The CE stem cell GES was derived as described by Eppert *et al.*[5] Briefly, summary measures of gene expression were computed for each probe-set using the robust multichip average method, which incorporates quantile normalization of arrays. The CE score was then calculated as the sum of the normalized expression values of the 44 probe sets included in the CE signature. As in the study by Eppert *et al.*,[5] patients were divided into groups with high (CE^{high}) or low (CE^{low}) CE score at the median, an approach that was also supported by our analyses of survival according to quartiles of CE score values (Supplementary Information and Supplementary Figure 1). Details on miR microarray data analysis are provided in the Supplementary Information.

Statistical analyses

Baseline characteristics were compared between CE^{high} and CE^{low} patients using Fisher's exact test for categorical variables and the Wilcoxon rank-sum test for continuous variables. Definitions of clinical endpoints (that is, CR, disease-free survival (DFS) and overall survival (OS)) are provided in the Supplement. For time-to-event analyses, we calculated survival estimates using the Kaplan–Meier method, and compared groups by the log-rank test. We constructed multivariable logistic regression models to analyze factors associated with the achievement of CR, and multivariable Cox proportional hazards models for factors associated with survival endpoints (see Supplementary Information). All analyses were performed by the Alliance for Clinical Trials in Oncology Statistics and Data Center, and the date of data lock was 11 October 2011.

RESULTS

Clinical and molecular characteristics associated with the CE stem cell gene expression score in CN-AML

We studied the associations between the CE score and clinical and molecular patient characteristics in our cohort of 364 primary CN-AML patients (Table 1). The proportion of patients with a high CE score (indicating a stem cell-like gene expression profile) was similar among younger and older patients ($P = 0.92$). Compared with CE^{low} patients, CE^{high} patients had higher peripheral blood (66 vs 54%; $P = 0.01$) and bone marrow blast percentages (70 vs 63%; $P = 0.008$), were more likely to carry *FLT3*-ITD (53 vs 19%; $P < 0.001$), mutated *WT1* (11 vs 4%; $P = 0.009$) and *RUNX1*

Table 1. Comparison of clinical and molecular characteristics according to the expression of the 'core enriched' stem cell gene expression score

Variable	*High CE score (n = 182)*	*Low CE score (n = 182)*	P-*value*
Age, years			0.53
Median	61	62	
Range	18–83	19–79	
Age group, n (%)			0.92
<60 years	83 (46)	81 (45)	
⩾60 years	99 (54)	101 (55)	
Male sex, *n* (%)	99 (54)	88 (48)	0.29
Race, n (%)			0.73
White	165 (91)	161 (89)	
Non-white	17 (9)	19 (11)	
White blood cell count, $\times 10^9$/l			0.36
Median	25.0	27.9	
Range	1.0–273	1.0–450	
Blood blasts (%)[a]			0.01
Median	66	54	
Range	1–99	1–97	
Bone marrow blasts (%)[a]			0.008
Median	70	63	
Range	15–97	4–97	
Hemoglobin, g/dl			0.42
Median	9.4	9.5	
Range	6.0–15.0	4.8–13.4	
Platelet count, $\times 10^9$/l			0.58
Median	67	59	
Range	4–850	5–510	
Extramedullary involvement, *n* (%)	38 (21)	60 (34)	0.01
FLT3-ITD, *n* (%)			<0.001
Positive	96 (53)	34 (19)	
Negative	86 (47)	148 (81)	
CEBPA, n (%)			<0.001
Mutated	15 (8)	40 (22)	
Single mutated	15	10	
Double mutated	0	30	
Wild-type	164 (92)	141 (78)	
NPM1, n (%)			0.67
Mutated	115 (64)	111 (61)	
Wild-type	66 (36)	71 (39)	
ELN Genetic Group, n (%)[b]			<0.001
Favorable	48 (27)	126 (70)	
Intermediate-I	130 (73)	55 (30)	
WT1, n (%)			0.009
Mutated	20 (11)	7 (4)	
Wild-type	158 (89)	174 (96)	
RUNX1, n (%)			0.01
Mutated	28 (18)	13 (8)	
Wild-type	132 (82)	150 (92)	
TET2, n (%)			0.01
Mutated	31 (18)	53 (30)	
Wild-type	139 (82)	123 (70)	

Table 1. (Continued)

Variable	*High CE score (n = 182)*	*Low CE score (n = 182)*	P-*value*
FLT3-TKD, *n* (%)			0.09
Present	14 (8)	25 (14)	
Absent	167 (92)	157 (86)	
DNMT3A, n (%)			0.25
Mutated	62 (38)	53 (31)	
R882	39	35	
Non-R882	23	18	
Wild-type	103 (62)	117 (68)	
ASXL1, n (%)			0.59
Mutated	19 (11)	16 (9)	
Wild-type	150 (89)	159 (91)	
IDH1, n (%)			0.63
Mutated	23 (13)	20 (11)	
Wild-type	149 (87)	156 (89)	
IDH2, n (%)			1.00
IDH2 mutated	27 (16)	28 (16)	
Codon R140 mutation	17	27	
Codon R172 mutation	10	1	
Wild-type	145 (84)	148 (84)	
MLL-PTD, *n* (%)			0.83
Present	11 (6)	13 (7)	
Absent	168 (94)	166 (93)	
ERG expression group, n (%)[c]			<0.001
High	131 (72)	54 (30)	
Low	51 (28)	128 (70)	
BAALC expression group, n (%)[c]			<0.001
High	118 (65)	70 (38)	
Low	64 (35)	112 (62)	
MN1 expression group, n (%)[c]			0.24
High	65 (56)	57 (48)	
Low	52 (44)	62 (52)	
miR-155 expression group, n (%)[c]			<0.001
High	111 (61)	63 (35)	
Low	71 (39)	119 (65)	

Abbreviations: CN-AML, cytogenetically normal acute myeloid leukemia; CE, core enriched; ELN, European LeukemiaNet; ITD, internal tandem duplication; TKD, tyrosine kinase domain; PTD, partial tandem duplication. [a]Peripheral blood and bone marrow blast percentages were centrally reviewed. [b]Within CN-AML patients, the ELN Favorable Genetic Group is defined as patients with mutated *CEBPA* and/or mutated *NPM1* without *FLT3*-ITD. All remaining CN-AML patients (that is, those with wild-type *CEBPA* and wild-type *NPM1* with or without *FLT3*-ITD or mutated *NPM1* with *FLT3*-ITD) belong to the ELN Intermediate-I Genetic Group.[22] [c]The median expression value was used as a cut point.

(18 vs 8%; $P = 0.01$) and have high expression of *ERG* (72 vs 30%; $P < 0.001$), *BAALC* (65 vs 38%; $P < 0.001$) and *miR-155* (61 vs 35%; $P < 0.001$). On the other hand, CE^{high} patients were less likely to have extramedullary involvement (21 vs 34%; $P = 0.01$), mutations in *TET2* (18 vs 30%; $P = 0.01$) or *CEBPA* (8 vs 22%; $P < 0.001$) than CE^{low} patients. Of note, no patient in the CE^{high} group had a double *CEBPA* mutation, whereas double *CEBPA* mutations occurred in 17% of CE^{low} patients ($P < 001$). Single *CEBPA* mutations were equally common in both groups. As *FLT3*-ITD mutations were more frequent and *CEBPA* mutations less frequent in CE^{high} patients, whereas there was no significant difference with

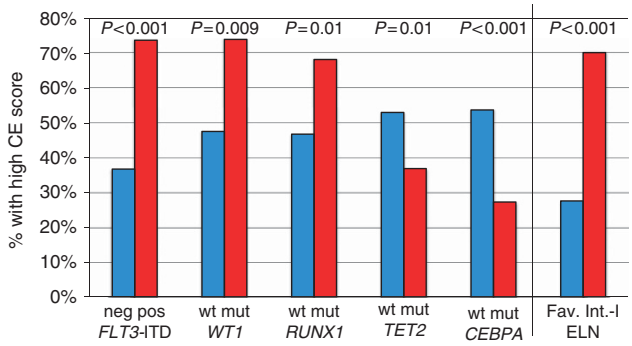

Figure 1. Association between CE stem cell gene expression scores and prognostic gene mutations in CN-AML. The CE score summarizes the expression of the 'CE' set of LSC-related genes as defined by Eppert *et al*.[5] The bar diagram shows the percentage of patients who have a high CE score, according to *FLT3*-ITD, *WT1*, *RUNX1*, *TET2* and *CEBPA* mutational status and ELN Genetic Group. Only mutations showing a significant association with CE scores were included.

respect to frequency of *NPM1* mutations, CE^{high} patients were less likely to belong to the European LeukemiaNet (ELN) Favorable Genetic Group (which comprises patients with mutated *CEBPA* and/or mutated *NPM1* without *FLT3*-ITD) than CE^{low} patients (27 vs 70%; $P < 0.001$).[22] Figure 1 illustrates the associations of the dichotomized CE gene expression score with individual gene mutations and the ELN Genetic Groups in CN-AML.

CE gene expression score and survival of CN-AML patients

In our entire cohort of 364 CN-AML patients, a high CE score associated with significantly lower odds of achieving a CR ($P = 0.003$; Table 2). Because of the baseline associations of a high CE score with established unfavorable molecular prognostic markers (that is, *FLT3*-ITD, mutated *WT1* and *RUNX1*; wild-type *CEBPA*; and high *ERG, BAALC* and *miR-155* expression), we constructed multivariable models evaluating these and other potentially confounding risk factors, including age group (Table 3; see Supplementary Information for a complete list of variables considered in the analyses). In a multivariable model for the achievement of CR, CE^{high} status remained associated with 47% lower odds of attaining a CR ($P = 0.02$). Other variables associated with lower odds of achieving CR were age $\geqslant$60 years, higher white blood count, absence of *NPM1* mutations and high *BAALC* expression. Among patients who reached CR, those with high CE expression had significantly shorter DFS than CE^{low} patients ($P < 0.001$; Figure 2a, Table 2). In a multivariable model for DFS, CE^{high} patients had a 2.2-fold higher risk of relapse or death than CE^{low} patients (Table 3; $P < 0.001$). Other variables associated with shorter DFS were age $\geqslant$60 years, higher white blood count, *WT1*, *ASXL1* and *DNMT3A* codon R882 mutations and high *miR-155* expression. Likewise, CE^{high} patients had shorter OS than CE^{low} patients ($P < 0.001$; Figure 2b). In a multivariable model for OS, CE^{high} patients had a 1.9-fold increased risk of death compared with CE^{low} patients (Table 3). Other factors associated with shorter OS were age $\geqslant$60 years, higher white blood count, *WT1, ASXL1* and *DNMT3A* codon R882 mutationss and high *BAALC* and *miR-155* expression.

As patients below the age of 60 received more intensive treatment than patients aged 60 years and above, we also studied the outcomes in these age groups separately (Table 2). In the younger age group, CE^{high} patients showed a trend toward a lower CR rate ($P = 0.09$), and had shorter DFS ($P < 0.001$; Supplementary Figure 2a) and OS ($P < 0.001$; Supplementary Figure 2b) compared with CE^{low} patients. Among older patients,

Table 2. Univariable analyses of outcomes according to expression of the 'core enriched' stem cell gene expression score

Group	*Endpoint*	*High CE score* (n = *182*)	*Low CE score* (n = *182*)	P-*value*
All patients (*n* = 364)	*Complete remission*, no. (%)	122 (67)	148 (81)	0.003
	Disease-free survival			<0.001
	Median (years)	0.7	1.7	
	% Disease-free at 3 years (95% CI)	17 (11–24)	41 (33–48)	
	% Disease-free at 5 years (95% CI)	16 (10–23)	36 (28–43)	
	Overall survival			<0.001
	Median (years)	1.0	2.5	
	% Alive at 3 years (95% CI)	20 (14–26)	45 (37–52)	
	% Alive at 5 years (95% CI)	16 (11–22)	39 (32–46)	
Younger patients (*n* = 164)	No. of patients	83	81	
	Complete remission, no. (%)	65 (78)	72 (89)	0.09
	Disease-free survival			<0.001
	Median, years	0.7	7.2	
	% Disease-free at 3 years (95% CI)	28 (17–39)	56 (43–66)	
	% Disease-free at 5 years (95% CI)	26 (16–37)	53 (41–63)	
	Overall survival			<0.001
	Median, years	1.2	n.r.	
	% Alive at 3 years (95% CI)	30 (21–40)	65 (54–74)	
	% Alive at 5 years (95% CI)	28 (19–38)	60 (49–70)	
Older patients (*n* = 200)	No. of patients	99	101	
	Complete remission, no. (%)	57 (57)	76 (75)	0.01
	Disease-free survival			<0.001
	Median (years)	0.6	1.1	
	% Disease-free at 3 years (95% CI)	5 (1–13)	26 (17–37)	
	% Disease-free at 5 years (95% CI)	4 (1–11)	20 (12–29)	
	Overall survival			<0.001
	Median (years)	0.8	1.5	
	% Alive at 3 years (95% CI)	11 (6–18)	28 (19–37)	
	% Alive at 5 years (95% CI)	6 (3–13)	22 (14–30)	

Abbreviations: n.r., not reached; CI, confidence interval. The median follow-up for those alive is 7.7 years, range: 2.3–13.1 years. The median follow-up for those who have not had an event is 7.9 years, range: 4.6–12.9 years.

Table 3. Multivariable models evaluating the 'core enriched' stem cell gene expression score and other patient characteristics for outcome

Complete remission		
Variable	*OR (95% CI)*	P-*value*
CE score (high vs low)	0.53 (0.30–0.91)	0.02
Age group (⩾60 vs <60 years)	0.37 (0.22–0.65)	<0.001
WBC (per 50-unit increase)	0.60 (0.47–0.78)	<0.001
NPM1 (mutated vs wild-type)	1.94 (1.07–3.53)	0.03
BAALC expression (high vs low)	0.30 (0.16–0.56)	<0.001
Disease-free survival		
Variable	*HR (95% CI)*	P-*value*
CE score (high vs low)	2.17 (1.60–2.95)	<0.001
Age group (⩾60 vs <60 years)	2.30 (1.68–3.14)	<0.001
WBC (per 50-unit increase)	1.21 (1.05–1.40)	0.01
WT1 (mutated vs wild-type)	2.94 (1.66–5.19)	<0.001
ASXL1 (mutated vs wild-type)	2.07 (1.18–3.65)	0.01
DNMT3A (codon R882 mutation present vs absent)	1.52 (1.06–2.19)	0.02
miR-155 expression (high vs low)	1.48 (1.10–1.99)	0.01
Overall survival		
Variable	*HR (95% CI)*	P-*value*
CE score (high vs low)	1.92 (1.46–2.52)	<0.001
Age group (⩾60 years vs <60 years)	2.74 (2.08–3.62)	<0.001
WBC (per 50-unit increase)	1.16 (1.06–1.26)	<0.001
WT1 (mutated vs wild-type)	3.15 (2.00–4.97)	<0.001
ASXL1 (mutated vs wild-type)	1.68 (1.12–2.52)	0.01
DNMT3A (codon R882 mutation present vs absent)	1.48 (1.09–2.01)	0.01
BAALC expression (high vs low)	1.51 (1.14–1.98)	0.004
miR-155 expression (high vs low)	1.66 (1.27–2.17)	<0.001

Abbreviations: CI, confidence interval; CE, core enriched; CR, complete remission; HR, hazard ratio; OR, odds ratio; WBC, white blood count. An odds ratio greater than (less than) 1.0 means a higher (lower) CR rate for the higher values of the continuous variables and the first category listed for the categorical variables. A hazard ratio greater than 1 (less than 1) corresponds to a higher (lower) risk of an event for higher values of continuous variables and the first category listed of a dichotomous variable. Variables were considered for inclusion in the multivariable models if they had a univariable *P*-value of <0.2. See the Supplementary Information for a full list of variables evaluated in univariable analyses.

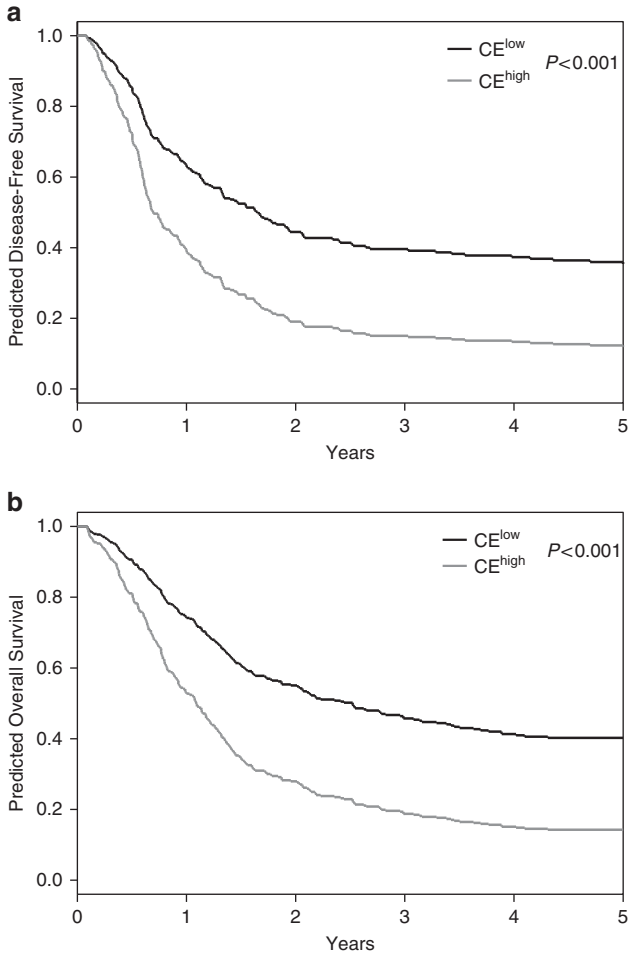

Figure 2. Survival of patients with CN-AML according to their CE stem cell gene expression score. (**a**) Disease-free survival, (**b**) overall survival. Kaplan–Meier curves are adjusted for age group (<60 vs ⩾60 years).

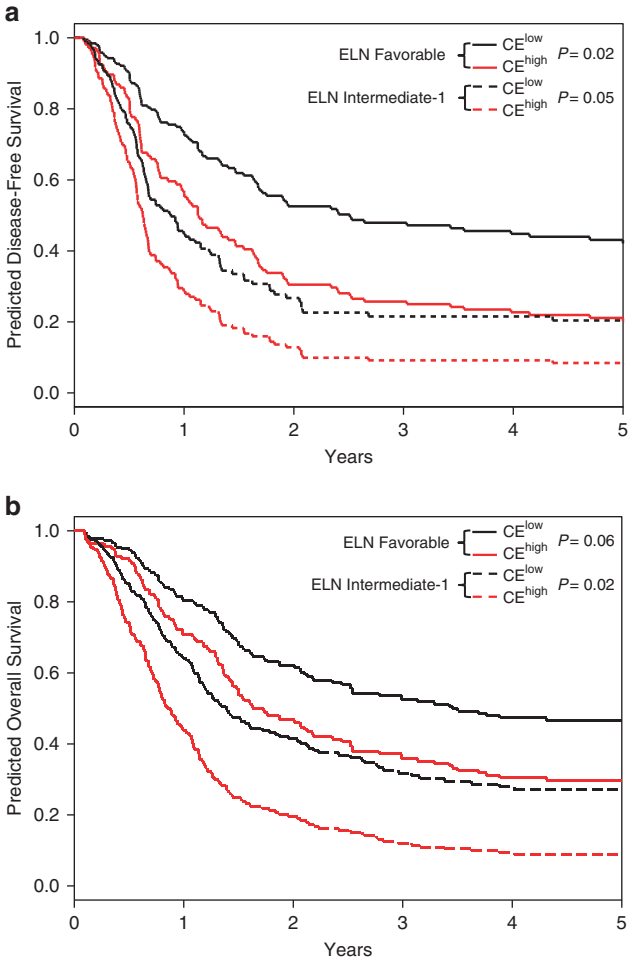

Figure 3. Survival of patients with CN-AML in the ELN Favorable and ELN Intermediate-I Genetic Group, according to CE stem cell gene expression score. (**a**) Disease-free survival, (**b**) overall survival. Kaplan–Meier curves are adjusted for age group (<60 vs ⩾60 years).

CE^{high} status was significantly associated a with a lower CR rate ($P=0.01$) and with shorter DFS ($P<0.001$; Supplementary Figure 2c) and OS ($P<0.001$; Supplementary Figure 2d). Older patients with a high CE score had particularly unfavorable outcomes, with a 3-year survival rate of only 11%, compared with 28% in CE^{low} older patients.

Prognostic value of the CE stem cell gene expression score in the context of the current ELN genetic classification of CN-AML

In 2010, an International expert panel working on behalf of the ELN proposed a standardized system for reporting cytogenetic and selected molecular abnormalities in AML.[22] Although the initial goal of the ELN classification was to facilitate comparisons between studies, the prognostic utility of the ELN Genetic Groups has been convincingly demonstrated.[23,24] Within the ELN classification, CN-AML patients are assigned to the ELN Favorable Genetic Group or the ELN Intermediate-I Genetic Group. Figure 3 shows the survival of patients in the ELN Favorable and Intermediate-I Groups according to their CE score. Within the ELN Favorable Genetic Group, CE^{high} patients, compared with CE^{low} patients, had comparable remission rates ($P=0.89$, CR rates, 88 vs 90% among younger and 83 vs 79% among older patients) but significantly shorter DFS ($P=0.02$), and showed a trend toward shorter OS ($P=0.06$). Within the ELN Intermediate-I Genetic Group, CE^{high} patients, compared with CE^{low} patients, had lower CR rates ($P=0.04$, CR rates, 74 vs 85% among younger and 50 vs 69% among older patients), and significantly inferior DFS ($P=0.05$) and OS ($P=0.002$). The survival of ELN Favorable/CE^{high} patients was very similar to those of ELN Intermediate-I/CE^{low} patients (Figure 3). When the ELN Genetic Groups, rather than individual molecular markers, were considered in multivariable analyses, a high CE score was not associated with CR rate but remained significantly associated with shorter DFS and OS (Supplementary Table 1). Thus, a single variable reflecting expression of a stem cell-like gene expression profile can refine the molecular risk stratification within both Genetic Groups of CN-AML patients defined by the ELN classification.

miR expression profiles associated with the CE gene expression score

miRs are important players involved in hematopoietic stem cell function, and deregulated expression of miRs has been shown to be clinically relevant in AML.[25,26] Therefore, we studied whether a more stem cell-like gene expression profile, indicated by a higher CE score, is accompanied by a characteristic miR expression signature. In these analyses, we identified a core set of 15 miRs that were consistently deregulated in patients with a high CE score in both age groups (<60 and ⩾60 years;

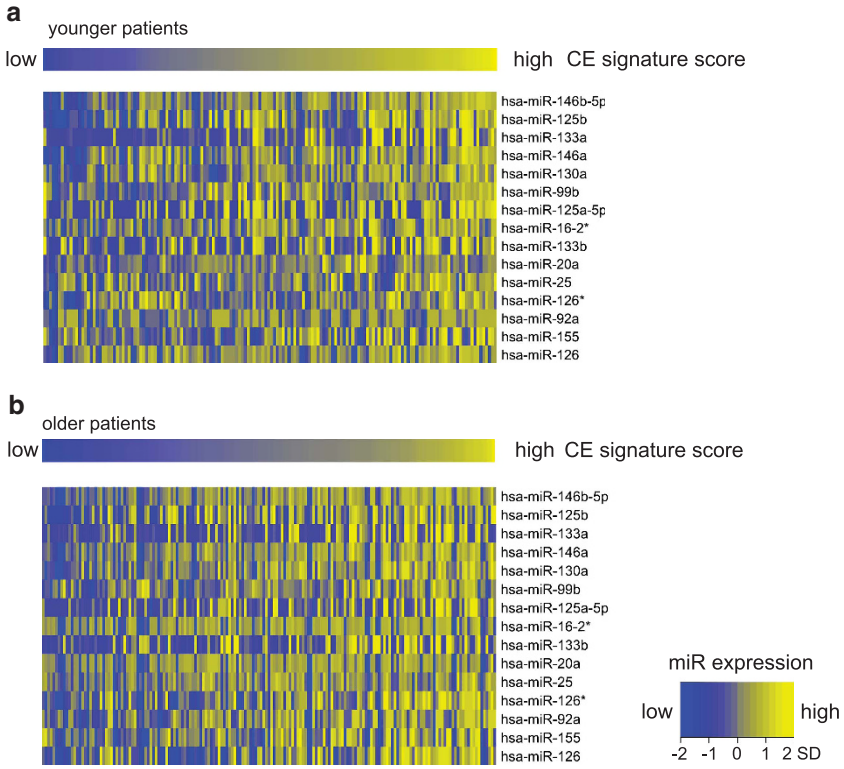

Figure 4. miR expression signatures associated with the CE stem cell gene expression score. (**a**) patients <60 years, (**b**) patients $\geqslant 60$ years.

see Supplementary Information for details). All 15 miRs showed higher expression in CE^{high} than in CE^{low} patients (Figure 4), indicating that they are overexpressed in CN-AML with a more stem cell-like gene expression profile. Table 4 summarizes the available data on the role of these miRs in normal hematopoiesis and AML.[27–41] Overexpressed miRs in patients with a high CE score include miRs known to be highly expressed and functionally relevant in embryonic (*miR-20a*)[37] or hematopoietic stem cells (*miR-99*,[27] *miR-125a/b*,[27,36] *miR-126*[27] and *miR-155*[33]). For some miRs in our signature, there are functional studies showing that their overexpression causes leukemia in model systems (*miR-92a*,[40] *miR-125b*[29]), or leads to increased survival and/or proliferation of normal or malignant myeloid cells (*miR-125a/b*,[28] *miR-126*[39]). Furthermore, several of the CE stem cell signature-associated miRs are known to be upregulated in CN-AML with prognostically unfavorable gene mutations (for example, *FLT3*-ITD or *IDH2* codon R172 mutations)[15,30,41] and/or downregulated in patients with favorable genetic changes (for example, mutated *NPM1* or translocation t(8;21)).[8,30,39] Our data illustrate that CN-AML blasts with a stem cell-like gene expression pattern also show other characteristics, such as expression of a characteristic set of miRs, known to be typical of stem cells. These findings suggest that in CE^{high} patients, the majority of leukemia cells have an overall cellular phenotype that more closely resembles LCSs compared with CE^{low} patients.

DISCUSSION

Recently, LSC-associated GESs have been reported in AML.[5,42] Eppert *et al.*[5] used functionally defined stem cell-enriched populations to define a LSC GES, and showed that high expression of these genes was associated with inferior OS and event-free survival in CN-AML. They also demonstrated that their approach was superior to using a purely phenotypic definition (that is, $CD34^+/CD38^-$ immunophenotype) to identify LSCs. These results, albeit intriguing, remained to be fully validated in independent patient cohorts. In order to assess the potential clinical utility of the stem cell-associated GES, it needs to be evaluated in the context of a comprehensive panel of other molecular prognosticators. Thus, our aim was to analyze the impact of the stem cell-associated 'CE' GES in a relatively large cohort of primary CN-AML patients that have been well characterized for molecular aberrations, including recently described gene mutations, deregulated expression of individual genes and miR expression profiles, that were not included in the original report by Eppert *et al.*[5]

In the largest patient cohort that has been studied for the prognostic relevance of stem cell-associated GESs, we demonstrate that high expression of the CE signature associates with inferior patient outcomes. Eppert *et al.*[5] showed that high expression of their stem cell GESs associated with inferior OS and event-free survival, but they did not report on other clinically relevant endpoints, including response to induction therapy (that is, CR) and DFS. Our study not only confirms their findings, but for the first time demonstrates that CN-AML patients with a high CE score (indicating a robust stem cell-like GES) have a lower chance of disease eradication than patients with a low CE score, as supported by a lower CR rate and shorter DFS. Our study also is the first to demonstrate the independent prognostic relevance of a stem cell GES in multivariable models considering a comprehensive set of known prognostic molecular markers, beyond the *FLT3*, *NPM1* and *CEBPA* mutations analyzed by Eppert *et al.*[5] We show that patients with a high CE score had a lower CR rate and shorter DFS and OS after adjusting for known clinical and molecular prognosticators. Furthermore, we show that the stem cell GES provides additional, clinically relevant prognostic information even in the context of the current ELN Genetic Classification of AML. A high CE score was associated with inferior DFS and OS both among low-risk (ELN Favorable) and high-risk (ELN Intermediate-I) CN-AML patients, thereby

Table 4. List of miRs associated with a high 'core enriched' stem cell gene expression score in younger and in older CN-AML patients

miR name	*Known functional role in hematopoiesis and leukemia*
hsa-miR-146b-5p	—
hsa-miR-125b	Highly expressed in murine hematopoietic stem cells[27] and in CN-AML with *IDH2* codon R172 mutation;[15] enhances proliferation and disturbs differentiation of myeloid progenitors;[28] and overexpression causes acute leukemia in mice[29]
hsa-miR-133a	Downregulated in AML with t(8;21)[30] and upregulated in CN-AML with *IDH2* codon R172 mutation.[15]
hsa-miR-146a	Lost in myelodysplastic syndrome with del(5)(q31)[31] overexpression in hematopoietic stem cells causes transient myeloid cell expansion;[32] and associated with downregulation of immune-response pathway genes[33,34]
hsa-miR-130a	Highly expressed in murine hematopoietic stem cells;[27] involved in cell cycle regulation in granulocytic progenitors;[35] associated with high expression of zinc finger transcription factors including WT1;[33] and downregulated in *NPM1*-mutated CN-AML[8]
hsa-miR-99b	Part of the miR-99b/let-7e/miR-125a cluster and highly expressed in hematopoietic stem cells[27]
hsa-miR-125a-5p	Highly expressed in hematopoietic stem cells,[27] and in CN-AML with *IDH2* codon R172 mutation;[15] increases hematopoietic stem cell numbers;[36] and enhances proliferation and disturbs differentiation of myeloid progenitors;[28]
*hsa-miR-16-2**	—
hsa-miR-133b	Downregulated in AML with t(8;21)[30]
hsa-miR-20a	Member of the miR-17-92 cluster; highly expressed in embryonic stem cells;[37] and associated with high expression of *HOX* genes including *HOXA5*[33]
hsa-miR-25	Promotes reprogramming of somatic cells into induced pluripotent stem cells[38]
*hsa-miR-126**	Increases survival/inhibits apoptosis of AML blasts;[39] and downregulated in *NPM1*-mutated CN-AML[8]
hsa-miR-92a	Member of the miR-17-92 cluster and overexpression causes erythroleukemia in mice through *p53* and *gata1* downregulation[40]
hsa-miR-155	Highly expressed in hematopoietic stem cells[27] and upregulated in *FLT3*-ITD-positive AML[30,41]
hsa-miR-126	Highly expressed in murine hematopoietic stem cells;[27] increases survival/inhibits apoptosis of AML blasts;[39] and downregulated in *NPM1*-mutated CN-AML[8]

Abbreviations: miR, microRNA; CN-AML, cytogenetically normal acute myeloid leukemia; ITD, internal tandem duplication. Two different microarray versions were used for younger (<60 years) and older ($\geqslant 60$ years) patients. A total of 535 miR probes were common to both platforms. Separate miR signatures were generated for each age group, and the 15 miRs listed above represent the overlap between those two signatures. The degree of overlap between the signatures in younger and older patients was statistically highly significant ($P = 1.1 \times 10^{-13}$ by Fisher's exact test).

suggesting that the stem cell GES may be useful to improve current prognostic cytogenetic- and molecular-based AML classifications.[23] Of note, our study was limited to previously untreated, primary CN-AML, and the prognostic importance of the signature remains to be validated in other cytogenetic subgroups and for patients with secondary or relapsed disease.

So far, it has also remained unknown whether expression of a stem cell-like gene expression profile is linked to other known molecular alterations in AML. Our data show that, although the CE score is an independent prognosticator, patients with a high CE score are more likely to be positive for multiple unfavorable prognostic markers in primary CN-AML (*FLT3*-ITD, *RUNX1* and *WT1* mutations and high *BAALC*, *ERG* and *miR-155* expression). Moreover, patients harboring favorable prognostic markers (that is, double *CEBPA* mutations) had low CE scores. These data may indicate a biologic interplay between these relatively frequent prognostic molecular markers and the 'stemness' features of AML blasts. Given that molecular markers are not only prognostic indicators, but also frequently represent suitable therapeutic targets, it is possible that the success of novel molecular therapeutic approaches may be determined and evaluated by their ability to modify the patients' CE score. Should this be the case, the CE score could represent a useful surrogate endpoint for early activity evaluation of novel therapies in CN-AML.

It should be also recognized that, although our results suggest that the prognostic significance of the CE score may be partially related to already known molecular alterations, it is very likely that some additional, as yet unknown, genetic and/or epigenetic alterations functionally contribute to the negative clinical impact of a stem cell-like GES. To this end, miRs are emerging as important contributors to myeloid leukemogenesis.[43,44] Deregulated expression can cause miRs to act as tumor suppressors or oncogenes, and the expression levels of several miRs have been shown to carry prognostic information in CN-AML.[25,26,43] The relationship between the stem cell-like GES and miR expression has not yet been reported. We found that the expression of a stem cell-like GES is accompanied by a characteristic miR signature. In younger and older CN-AML patients, we identified a core set of 15 miRs that were consistently overexpressed in CE^{high} patients. Several of these miRs, including *miR-92a*, *miR-125a/b*, *miR-126* and *miR-146a,* have been functionally implicated in normal stem cell or LSC biology before, thus supporting a potential role of noncoding RNAs in maintaining the LSC compartment-like phenotype in a subset of CN-AML patients. Others (for example, *miR-155*) have been shown to independently impact on the prognosis of CN-AML patients.[21] These findings suggest that a complex network of aberrantly expressed genes and miRs and gene mutations collectively define a stem cell-like phenotype associated with clinically aggressive disease. It is also possible that targeting miRs[45] may directly have an impact on the self-renewal ability of AML blast subpopulations enriched for LSCs.

In summary, we validated that high expression of a stem cell-associated GES has negative prognostic impact in primary CN-AML. Although we showed that a high CE signature associates with known unfavorable molecular alterations in CN-AML, it provides additional prognostic information not reflected by these markers. Our results suggest that the discovery of additional genetic and epigenetic mechanisms may be necessary to fully explain the functional role of the LSC signature genes during leukemogenesis. In support of this hypothesis, we have shown that the stem cell-associated GES associates with a characteristic miR expression profile comprising miRs known to be involved in conferring 'stemness' to normal and malignant blasts. Future studies of newly discovered gene mutations and aberrantly expressed genes and miRs occurring in LSCs may not only improve patients' molecular risk stratification but also potentially reveal novel therapeutic targets.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported in part by the National Cancer Institute (CA101140, CA114725, CA31946, CA33601, CA16058, CA77658, CA129657 and CA140158), The Coleman Leukemia Research Foundation, the Deutsche Krebshilfe-Dr Mildred Scheel Cancer Foundation (HB), the Pelotonia Fellowship Program (A-KE) and the Conquer Cancer Foundation (JHM). The CALGB/Alliance institutions, principal investigators and cytogeneticists participating in this study are listed in the Supplementary Information. We thank Donna Bucci and the CALGB/Alliance Leukemia Tissue Bank at The Ohio State University Comprehensive Cancer Center, Columbus, OH, for sample processing and storage services, and Lisa J Sterling, Christine Finks and Colin G Edwards, PhD, for data management.

REFERENCES

1 Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* 1997; **3**: 730–737.
2 Lapidot T, Fajerman Y, Kollet O. Immune-deficient SCID and NOD/SCID mice models as functional assays for studying normal and malignant human hematopoiesis. *J Mol Med (Berl)* 1997; **75**: 664–673.
3 Sarry J-E, Murphy K, Perry R, Sanchez PV, Secreto A, Keefer C *et al.* Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rγc-deficient mice. *J Clin Invest* 2011; **121**: 384–395.
4 Vargaftig J, Taussig DC, Griessinger E, Anjos-Afonso F, Lister TA, Cavenagh J *et al.* Frequency of leukemic initiating cells does not depend on the xenotransplantation model used. *Leukemia* 2012; **26**: 858–860.
5 Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med* 2011; **17**: 1086–1093.
6 Mrózek K, Carroll AJ, Maharry K, Rao KW, Patil SR, Pettenati MJ *et al.* Central review of cytogenetics is necessary for cooperative group correlative and clinical studies of adult acute leukemia: the Cancer and Leukemia Group B experience. *Int J Oncol* 2008; **33**: 239–244.
7 Whitman SP, Maharry K, Radmacher MD, Becker H, Mrózek K, Margeson D *et al. FLT3* internal tandem duplication associates with adverse outcome and gene- and microRNA-expression signatures in patients 60 years of age or older with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* 2010; **116**: 3622–3626.
8 Becker H, Marcucci G, Maharry K, Radmacher MD, Mrózek K, Margeson D *et al.* Favorable prognostic impact of *NPM1* mutations in older patients with cytogenetically normal de novo acute myeloid leukemia and associated gene- and microRNA-expression signatures: a Cancer and Leukemia Group B study. *J Clin Oncol* 2010; **28**: 596–604.
9 Marcucci G, Maharry K, Radmacher MD, Mrózek K, Vukosavljevic T, Paschka P *et al.* Prognostic significance of, and gene and microRNA expression signatures associated with, *CEBPA* mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a Cancer and Leukemia Group B study. *J Clin Oncol* 2008; **26**: 5078–5087.
10 Becker H, Marcucci G, Maharry K, Radmacher MD, Mrózek K, Margeson D *et al.* Mutations of the Wilms tumor 1 gene (*WT1*) in older patients with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* 2010; **116**: 788–792.
11 Mendler JH, Maharry K, Radmacher MD, Mrózek K, Becker H, Metzeler KH *et al. RUNX1* mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and microRNA expression signatures. *J Clin Oncol* 2012; **30**: 3109–3118.
12 Metzeler KH, Maharry K, Radmacher MD, Mrózek K, Margeson D, Becker H *et al. TET2* mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol* 2011; **29**: 1373–1381.
13 Marcucci G, Metzeler KH, Schwind S, Becker H, Maharry K, Mrózek K *et al.* Age-related prognostic impact of different types of *DNMT3A* mutations in adults with primary cytogenetically normal acute myeloid leukemia. *J Clin Oncol* 2012; **30**: 742–750.
14 Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrózek K *et al. ASXL1* mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. *Blood* 2011; **118**: 6920–6929.
15 Marcucci G, Maharry K, Wu Y-Z, Radmacher MD, Mrózek K, Margeson D *et al. IDH1* and *IDH2* gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol* 2010; **28**: 2348–2355.
16 Whitman SP, Ruppert AS, Radmacher MD, Mrózek K, Paschka P, Langer C *et al. FLT3* D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with *de novo* cytogenetically normal acute myeloid leukemia lacking *FLT3* internal tandem duplications. *Blood* 2008; **111**: 1552–1559.
17 Whitman SP, Caligiuri MA, Maharry K, Radmacher MD, Kohlschmidt J, Becker H *et al.* The *MLL* partial tandem duplication in adults aged 60 years and older with *de novo* cytogenetically normal acute myeloid leukemia. *Leukemia* 2012; **26**: 1713–1717.
18 Caligiuri MA, Strout MP, Schichman SA, Mrózek K, Arthur DC, Herzig GP *et al.* Partial tandem duplication of *ALL1* as a recurrent molecular defect in acute myeloid leukemia with trisomy 11. *Cancer Res* 1996; **56**: 1418–1425.
19 Schwind S, Marcucci G, Maharry K, Radmacher MD, Mrózek K, Holland KB *et al. BAALC* and *ERG* expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with *de novo* cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* 2010; **116**: 5660–5669.
20 Schwind S, Marcucci G, Kohlschmidt J, Radmacher MD, Mrózek K, Maharry K *et al.* Low expression of *MN1* associates with better treatment response in older patients with de novo cytogenetically normal acute myeloid leukemia. *Blood* 2011; **118**: 4188–4198.
21 Marcucci G, Maharry K, Metzeler KH, Volinia S, Wu Y-Z, Mrózek K *et al.* Clinical role of microRNAs in cytogenetically normal acute myeloid leukemia: *miR-155* upregulation independently identifies high-risk patients. *J Clin Oncol* 2013; **31**: 2086–2093.
22 Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK *et al.* Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 2010; **115**: 453–474.
23 Mrózek K, Marcucci G, Nicolet D, Maharry KS, Becker H, Whitman SP *et al.* Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. *J Clin Oncol* 2012; **30**: 4515–4523.
24 Röllig C, Bornhäuser M, Thiede C, Taube F, Kramer M, Mohr B *et al.* Long-term prognosis of acute myeloid leukemia according to the new genetic risk classification of the European LeukemiaNet recommendations: evaluation of the proposed reporting system. *J Clin Oncol* 2011; **29**: 2758–2765.
25 Schwind S, Maharry K, Radmacher MD, Mrózek K, Holland KB, Margeson D *et al.* Prognostic significance of expression of a single microRNA, *miR-181a*, in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol* 2010; **28**: 5257–5264.
26 Eisfeld A-K, Marcucci G, Maharry K, Schwind S, Radmacher MD, Nicolet D *et al. miR-3151* interplays with its host gene *BAALC* and independently affects outcome of patients with cytogenetically normal acute myeloid leukemia. *Blood* 2012; **120**: 249–258.
27 Gerrits A, Walasek MA, Olthof S, Weersing E, Ritsema M, Zwart E *et al.* Genetic screen identifies microRNA cluster 99b/let-7e/125a as a regulator of primitive hematopoietic cells. *Blood* 2012; **119**: 377–387.
28 Shaham L, Binder V, Gefen N, Borkhardt A, Izraeli S. MiR-125 in normal and malignant hematopoiesis. *Leukemia* 2012; **26**: 2011–2018.
29 Bousquet M, Harris MH, Zhou B, Lodish HF. MicroRNA miR-125b causes leukemia. *Proc Natl Acad Sci USA* 2010; **107**: 21558–21563.
30 Jongen-Lavrencic M, Sun SM, Dijkstra MK, Valk P, Lowenberg B. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. *Blood* 2008; **111**: 5078–5085.
31 Starczynowski DT, Kuchenbauer F, Argiropoulos B, Sung S, Morin R, Muranyi A *et al.* Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype. *Nat Med* 2010; **16**: 49–58.
32 Starczynowski DT, Kuchenbauer F, Wegrzyn J, Rouhi A, Petriv O, Hansen CL *et al.* MicroRNA-146a disrupts hematopoietic differentiation and survival. *Exp Hematol* 2011; **39**: 167–178.
33 Havelange N, Stauffer N, Heaphy CCE, Volinia S, Andreeff M, Marcucci G *et al.* Functional implications of microRNAs in acute myeloid leukemia by integrating microRNA and messenger RNA expression profiling. *Cancer* 2011; **117**: 4696–4706.
34 Quinn EM, Wang JH, Redmond HP. The emerging role of microRNA in regulation of endotoxin tolerance. *J Leukoc Biol* 2012; **91**: 721–727.
35 Häger M, Pedersen CC, Larsen MT, Andersen MK, Hother C, Grønbæk K *et al.* MicroRNA-130a–mediated down-regulation of Smad4 contributes to reduced sensitivity to TGF-β1 stimulation in granulocytic precursors. *Blood* 2011; **118**: 6649–6659.
36 Guo S, Lu J, Schlager R, Zhang H, Wang JY, Fox MC *et al.* MicroRNA miR-125a controls hematopoietic stem cell number. *Proc Natl Acad Sci USA* 2010; **107**: 14229–14234.
37 Rizzo M, Mariani L, Pitto L, Rainaldi G, Simili M. miR-20a and miR-290, multifaceted players with a role in tumourigenesis and senescence. *J Cell Mol Med* 2010; **14**: 2633–2640.

38 Lu D, Davis MPA, Abreu-Goodger C, Wang W, Campos LS, Siede J *et al.* MiR-25 regulates Wwp2 and Fbxw7 and promotes reprogramming of mouse fibroblast cells to iPSCs. *PLoS One* 2012; **7**: e40938.
39 Li Z, Lu J, Sun M, Mi S, Zhang H, Luo RT *et al.* Distinct microRNA expression profiles in acute myeloid leukemia with common translocations. *Proc Natl Acad Sci USA* 2008; **105**: 15535–15540.
40 Li Y, Vecchiarelli-Federico LM, Li Y-J, Egan SE, Spaner D, Hough MR *et al.* The miR-17-92 cluster expands multipotent hematopoietic progenitors whereas imbalanced expression of its individual oncogenic miRNAs promotes leukemia in mice. *Blood* 2012; **119**: 4486–4498.
41 Garzon R, Volinia S, Liu C-G, Fernandez-Cymering C, Palumbo T, Pichiorri F *et al.* MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood* 2008; **111**: 3183–3189.
42 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**: 2706–2715.
43 Marcucci G, Mrózek K, Radmacher MD, Garzon R, Bloomfield CD. The prognostic and functional role of microRNAs in acute myeloid leukemia. *Blood* 2011; **117**: 1121–1129.
44 Hickey CJ, Schwind S, Radomska HS, Dorrance AM, Santhanam R, Mishra A *et al.* Lenalidomide-mediated enhanced translation of C/EBPα-p30 protein upregulates expression of the antileukemic *microRNA-181a* in acute myeloid leukemia. *Blood* 2013; **121**: 159–169.
45 Janssen HL, Reesink HW, Lawitz EJ, Zeuzem S, Rodriguez-Torres M, Patel K *et al.* Treatment of HCV infection by targeting microRNA. *N Engl J Med* 2013; **368**: 1685–1694.

Supplementary Information accompanies this paper on the Leukemia website (http://www.nature.com/leu)