Leukemia (2015) **29,** 1051–1059

www.nature.com/leu

ORIGINAL ARTICLE

A 3-microRNA scoring system for prognostication in *de novo* acute myeloid leukemia patients

M-K Chuang[1,5], Y-C Chiu[2,5], W-C Chou[1,3], H-A Hou[3], EY Chuang[2,4] and H-F Tien[3]

As a highly heterogeneous disease, acute myeloid leukemia (AML) needs fine risk stratification to get an optimal outcome of patients. MicroRNAs have florid biological functions and have critical roles in the pathogenesis and prognosis in AML. Expression levels of some single microRNAs are influential for prognosis, but a system integrating several together and considering the weight of each should be more powerful. We thus analyzed the clinical, genetic and microRNA profiling data of 138 *de novo* AML patients of our institute. By multivariate analysis, we identified that high expression of *hsa-miR-9-5p* and *hsa-miR-155-5p* were independent poor prognostic factors, whereas that of *hsa-miR-203* had a trend to be a favorable factor. We constructed a scoring system from expression of these three microRNAs by considering the weight of each. The scores correlated with distinct clinical and biological features and outperformed single microRNA expression in prognostication. In both ours and another validation cohort, higher scores were associated with shorter overall survival, independent of other well-known prognostic factors. By analyzing the mRNA expression profiles, we sorted out several cancer-related pathways highly correlated with the microRNA prognostic signature. We conclude that this 3-microRNA scoring system is simple and powerful for risk stratification of *de novo* AML patients.

Leukemia (2015) **29,** 1051–1059; doi:10.1038/leu.2014.333

INTRODUCTION

MicroRNA is a class of small, noncoding RNA, which is derived from precursor RNA through processing by protein complex including Dicer and Drosha. It regulates gene expression posttranscriptionally by either degradation of mRNA or inhibition of translation via binding at the 3′-untranslated region of their target genes.[1,2] The roles of microRNAs in carcinogenesis are very complex. In acute myeloid leukemia (AML), microRNAs are involved in hematopoietic cell differentiation, proliferation and survival, and have an impact on treatment response and outcome.[3,4]

Different microRNA expression profiles are seen in various cytogenetic groups of AML.[4–7] Moreover, AMLs with specific gene mutations also harbor distinct sets of microRNA signatures.[8–10] A particularly important feature of microRNA expression is its role in prognosis. More and more studies have demonstrated both positive and negative roles of microRNAs in gene regulations and their implications on the prognosis or leukemogenesis of AML.[4,5,11–15] Higher expression of a single microRNA, *miR-181a*, appears to be an independent good prognostic factor in cytogenetically normal AML.[11] On the other hand, high expression of individual *miR-191*, *miR-199a*[5] or *miR-155*,[16] and low expression of *miR-212*[12] or *miR-29*[13,14] were reported as poor prognostic factors in AML. We reason that multiple microRNAs are usually involved in specific physiological pathways and may in concert influence the response to chemotherapy in AML patients, so expression levels of multiple microRNAs may be more powerful to predict the prognosis in these patients. Incorporating expression of multiple relevant microRNAs together and taking into account each microRNA's weight in the survival analysis may provide more integrative information in prognostication. In this study, by comprehensive analysis of the microRNA expression profiles from our AML patients, we generated a simplified microRNA signature with prognostic significance. The power of this signature was then validated by another independent AML cohort from The Cancer Genome Atlas (TCGA; https://tcga-data.nci.nih.gov/tcga/). The possible molecular pathways underlying the signature were also explored through analyses of mRNA array data from our cohort.

MATERIALS AND METHODS

Patients

We recruited a total of 195 consecutive adult patients (⩾15 years of age) with newly diagnosed *de novo* AML from 1995 to 2007 at the National Taiwan University Hospital (NTUH), who had adequate cryopreserved bone marrow cells for microRNA analysis. Patients with antecedent hematological diseases or therapy-related AML were excluded. This study was performed in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of the NTUH. Among these primary AML patients, 138 (70.7%) received standard intensive chemotherapy as described previously.[17] The remaining 57 patients received palliative care or low-dose chemotherapy due to poor performance status or per patients' wish. All the 195 patients were included for correlation analysis between expression of specific microRNA and other parameters, but only the 138 patients who received standard intensive chemotherapy were included for survival analysis. We used AML cohort from TCGA

[1]Department of Laboratory Medicine, National Taiwan University Hospital, Taipei, Taiwan; [2]Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan; [3]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan and [4]Bioinformatics and Biostatistics Core, Center of Genomic Medicine, National Taiwan University, Taipei, Taiwan. Correspondence: Dr W-C Chou, Department of Laboratory Medicine, National Taiwan University Hospital, No. 7, Chung-Shan South Road, Taipei 10002, Taiwan or Dr H-F Tien, Department of Internal Medicine, National Taiwan University Hospital, No. 7, Chung-Shan South Road, Taipei 10002, Taiwan. E-mail: wchou@ntu.edu.tw or hftien@ntu.edu.tw
[5]These two authors contributed equally to this work.
Received 5 August 2014; revised 27 September 2014; accepted 7 November 2014; accepted article preview online 27 November 2014; advance online publication, 19 December 2014

(https://tcga-data.nci.nih.gov/tcga/), which contains publicly available data of microRNA and clinical information, as a validation cohort (Figure 1).[18]

Cytogenetic and mutation analysis

Chromosomal abnormalities and gene mutations were analyzed as described previously.[17,19–22]

Quantification of microRNA expression

Mononuclear cells were isolated from bone marrow samples obtained at diagnoses, followed by cryopreservation. We extracted RNA by TriZol method. One microgram RNA was subjected to TaqMan MicroRNA Reverse Transcription kit (Applied Biosystems, Grand Island, NY, USA) and the microRNA profiling was assayed by TaqMan Array Human microRNA A card (Applied Biosystems) on 7900HT real-time PCR machine. The amplification curves were converted into numeric tables using Applied Biosystems SDS2.3 Software. MicroRNAs that cannot be detected by 40 cycles ($C_t > 40$) were marked as undetermined.

mRNA microarray analysis

One hundred and eighty-one of the 195 patients also had mRNA microarray study using HumanHT-12 v4 Expression BeadChip (Illumina, San Diego, CA, USA) for global gene expression profiles. RNA concentration and integrity were verified with ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). For each sample, 1.5 μg cRNA was hybridized to Illumina HumanHT-12 v4 Expression BeadChip according to the manufacturer's instructions. Intensities of bead fluorescence were detected by Illumina BeadArray Reader, and the results were transformed to numeric values using GenomeStudio v2010.1 Software (Illumina).

Data sets processing

For the quantitative PCR with reverse transcription data, PCR cycle numbers (C_t) of microRNAs in each sample were normalized into ΔC_t value by subtracting average C_t of four probes of the endogenous control *MAMMU6*. $-\Delta C_t$ value represented expression abundance of microRNAs and were tabulated for further analysis. Global expression levels of mRNAs were Log2 transformed and quantile normalized across patients for the elimination of batch biases. The validation microRNA data sets were downloaded from The Cancer Genome Atlas (TCGA)[18] in May 2013. The microRNA expression levels of the 197 AML patients from TCGA were determined by Illumina HiSeq 2000 microRNA seq, in whom overall survival (OS) data were available in 174 patients. We transformed TCGA level 3 sequencing data (expression levels of each microRNA) into Log2 scale.

Establishment of a risk scoring system

To build a risk scoring system on the basis of microRNA expression levels, we first analyzed the association between OS and the expression levels of individual microRNAs. Here, the prognostic significance of each microRNA expression on survival was measured using univariate Cox proportional hazards regression model. The expression of microRNAs with top significance on survival (univariate Cox $P < 0.005$) was then applied to the multivariate Cox model to find those microRNAs whose expression could independently predict survival. The expression of microRNAs with significant association with OS from the multivariate tests (multivariate Cox $P < 0.1$) was selected to generate the risk scoring system, in which the expression of component microRNAs went through another round of multivariate Cox regression test to get beta values as weights. The expression of microRNAs with higher prognostic significance on survival was weighted more. The microRNA-based risk scoring system is defined as $\mathrm{Risk}(p) = \sum_{\mathrm{miRNA}_i \in \text{component miRNAs}} \mathrm{Beta}_i \cdot \mathrm{miRNA}_i(p)$, where p denotes the patient accession number. Beta_i, which is derived from multivariate Cox analysis, means the weight of the microRNA probe i, and miRNA_i means the expression levels of microRNA probe i after z-transformation (that is, subtracting the mean and then divided by the s.d.) across patients so that each microRNA has zero mean and unit s.d. The sum of Beta_i·$\mathrm{miRNA}_i(p)$ of all the component microRNAs is the estimated risk scores for Patient p.

We performed a 10 000-time random permutation test to ensure the performance of our scoring system. For each iteration of the random permutation, the same number of microRNAs were randomly selected from the microRNA data set into construction of a 'random scoring system', where appropriate weights were assigned according to the procedures discussed above. Each random scoring system was tested for prognostic significance using the univariate Cox model. After 10 000 iterations, the empirical P-value of our proposed risk system could be calculated as the fraction of random scoring systems that achieved better univariate Cox

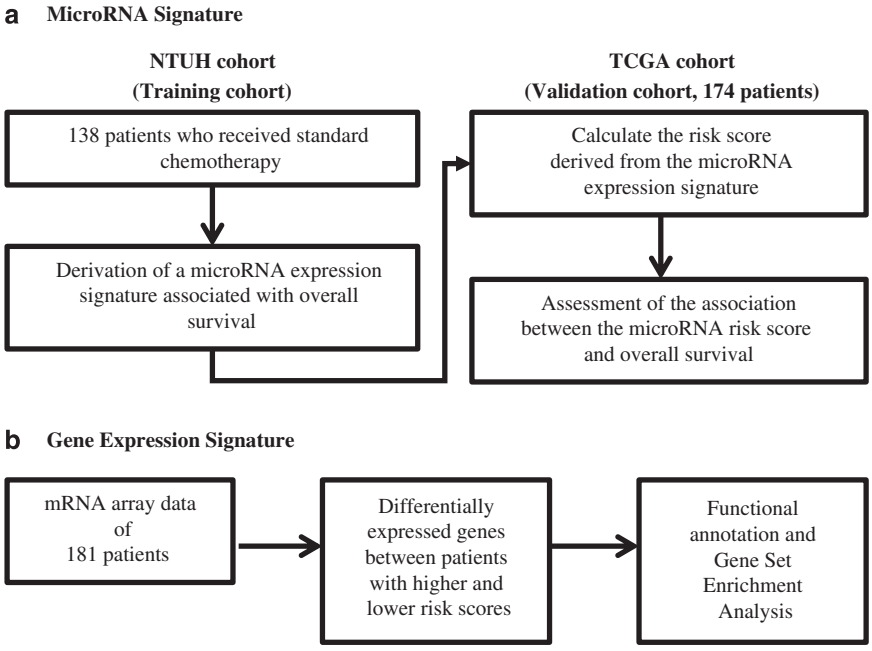

Figure 1. The flow chart of the study. (**a**) The patients from our institute, NTUH, serve as a training cohort. Totally 195 patients were recruited but only 138 patients who received standard intensive chemotherapy were included for construction of the microRNA risk scoring system. The result was then tested in 174 patients from TCGA database (as a validation cohort). (**b**) The NTUH cohort contains mRNA expression profiles so that the underlying genetic pathways associated with the highest quarter and the lowest quarter of microRNA scores can be sorted out by analyzing the mRNA expression signatures. In NTUH cohort, 181 of the 195 patients had both microRNA and mRNA expression data and were the subjects for integrated microRNA and mRNA expression analysis.

P-values than our proposed risk system. The smaller the empirical *P*-value is, the better our proposed risk scoring system outperforms random microRNA combinations.

Integrated gene expression analysis and pathway analysis

We analyzed the samples with paired mRNA and microRNA array data set ($n = 181$, Figure 1) to dissect the underlying molecular mechanisms that distinguish AML patients with different microRNA risk scores. Fold-change test and two-sided Student's *t*-test with unequal variance were used to identify differentially expressed genes between patients with high (> average +1 s.d.) and low (< average −1 s.d.) risk scores. We then analyzed the differentially expressed genes for associated biological functions using the knowledge-based software Ingenuity Pathway Analysis (IPA, Ingenuity Systems, Redwood City, CA, USA).[23] Biological functions of interest were further validated using Gene Set Enrichment Analysis (GSEA) software v2.1.0 (downloadable at www.broadinstitute.org/gsea/index.jsp). GSEA tests whether a gene set (list of genes involved in a common biological function) is enriched in the observed whole-genome gene expression profile. Unlike the pathway analysis which focused on analyzing differentially expressed genes, GSEA measures the global trend of changes in the expression profile and can thus reveal modest and subtle changes. Here the gene sets were manually downloaded from the database of IPA. Statistical significance of degree of enrichment was assessed using the 2000-time random permutation test among genes.

Statistical analysis

The discrete variables of patients with low and high microRNA-based risk scores were compared using the chi-square tests or Fisher's exact test. We used the Mann–Whitney *U*-test to compare continuous variables and medians of distributions. OS was measured from the date of first diagnosis to death from any cause or the last follow-up. To exclude any confounding influences from allogeneic hematopoietic stem cell transplantation, patients who received this procedure were censored on the day of cell infusion. We adopted Kaplan–Meier estimation to plot the survival curves and used log-rank tests to examine the difference between groups. The whole patient population was included for analyses of the correlation between microRNA expression-based risk scoring system and clinical characteristics; however, only those receiving standard intensive chemotherapy were included in the analyses of survivals. All statistical analyses were accomplished with XLSTAT statistical analysis software edition 2010.5.02 (Addinsoft, Deutschland, Germany).

RESULTS

Construction of the 3-microRNA risk scoring system

We sought to define a microRNA scoring system predictive of the risk in AML patients. We started by screening single microRNA expression that was individually associated with OS in the NTUH cohort (as a discovery data set, $n = 138$). The univariate Cox proportional regression model identified expression of 11 microRNAs with significant association with OS ($P < 0.005$), including *hsa-miR-9-5p*, *hsa-miR-146a*, *hsa-miR-222*, *hsa-miR-128*, *hsa-miR-181a*, *hsa-miR-125b*, *hsa-miR-196b*, *hsa-miR-155-5p*, *hsa-miR-224*, *hsa-miR-203* and *hsa-miR-339-3p* (ranked by increasing Cox *P*-values). To further pinpoint microRNAs expression with independent power of survival prediction, we introduced the expression of 11 microRNAs into a multivariate Cox model and identified high expression of *hsa-miR-9-5p* (accession number: MIMAT0000441) and *hsa-miR-155-5p* (MIMAT0000646) were independently associated with poor OS, whereas that of *hsa-miR-203* (MIMAT0000264) had a trend of association with favorable OS, with multivariate Cox $P = 0.005$, 0.033 and 0.080, respectively. By focusing on these three microRNAs, we constructed a risk scoring system according to the methods described above: Risk = 0.4908 [*hsa-miR-9-5p*]+0.2243[*hsa-miR-155-5p*] − 0.7187[*hsa-miR-203*]. The distribution of risk scores among the 138 patients in the NTUH discovery group was approximately normally shaped and ranged from −2.01 to 1.97, with median, mean and s.d. of 0.01, 2.87e−15 and 0.78, respectively (Supplementary Figure S1). The same situation could also apply to TCGA cohort (Supplementary Figure S1). The risk score served as a good survival predictor in our data set (univariate Cox $P = 1.41 \times 10^{-7}$, log-rank $P = 1.03 \times 10^{-5}$). The scoring system outperformed almost all the 10 000 random selections (empirical $P = 6 \times 10^{-4}$), suggestive of the high performance and nonrandomness of our proposed system.

Correlation of clinical and molecular characteristics with the scoring system

Higher scores were positively associated with older age, higher counts of white blood cells, platelets and blasts, but inversely associated with favorable cytogenetics (Table 1). Genetic mutation profiles were also different between the high and low score groups: patients with higher scores more often had *NPM1* mutation, *FLT3*-ITD and *MLL*-PTD, but less likely had *CEBPA* mutation (Table 2).

Survival analysis

AML patients with higher scores had significantly shorter OS compared with those with lower scores (median 13.5 months vs not reached, $P < 0.0001$, Figure 2a). The prognostic significance of the scoring system was validated in TCGA AML cohort (median 12.2 vs 26.4 months, $P = 0.008$, Figure 2b), which is the only publicly available AML cohort with survival and microRNA expression data. When we restricted the analysis in our patients with a normal karyotype, the OS of patients with higher scores still fared worse (median 17.0 months vs not reached, $P = 0.006$, Figure 2c). We sought to confirm this finding in TCGA cohort ($N = 86$) and found a trend of longer OS in patients with lower microRNA scores ($P = 0.131$). Although the difference in OS between the patients with lower and higher scores was not statistically significant, the survival curves looked separated with median survival duration of 19.3 vs 12.2 months for these two groups, respectively (Figure 2d). In univariate analysis, the poor risk factors for OS were older age, high white blood cell counts at diagnosis, unfavorable karyotype, *FLT3*-ITD, *RUNX1* mutation and *MLL*-PTD. CEBPA double mutations were favorable prognostic factors (Supplementary Table S1). We also investigated the association between the microRNA risk scores and expression of genes *BAALC*,[24] *ERG*[25,26] and *MN1*,[27] which were previously reported to be predictive of clinical outcome in AML patients. Notably, no significant difference was observed in the expression levels of these genes between the highest and lowest risk groups (Supplementary Table S2).

Multivariate analysis

Because high scores seemed to be associated with other poor prognostic variables (Tables 1 and 2), we sought to investigate whether the score in our scoring system functioned as an independent factor. We included several well-known prognostic factors as co-variables in multivariate analysis and high score appeared to be a highly independent risk factor. Notably, the independency of our scoring system still held true in TCGA cohort (Table 3). Further analysis by integrating the co-variables, we noted that the patients with more poor prognostic factors had shorter OS in both ours and TCGA patients (Figure 3). Even the mutation statuses of *RUNX1*, *WT1*, *DNMT3A* and *MLL*-PTD were included in the multivariate analysis, the 3-microRNA score was still a significantly independent prognostic factor in our patients (Supplementary Table S3; TCGA cohort was not calculated because these mutation statuses were not available). Furthermore, the microRNA score remained to be an independent prognostic factor by multivariate analysis in the subgroup of non-APL patients in ours and TCGA cohort ($P = 0.01$ and 0.002, respectively, Supplementary Table S4).

Table 1. Correlation between microRNA score and clinical data, FAB subtypes and chromosomal abnormalities in AML patients ($n = 195$)

Variant	*Total*	*MicroRNA score*		P
		Low (n = *97*)	*High* (n = *98*)	
Median age, years (range)	54 (15–89)	50 (18–89)	61.5 (15–88)	0.210
Age, in groups				
>60	80 (41.0%)	30 (30.9%)	50 (51.0%)	0.006
>50	108 (56.8%)	48 (49.5%)	60 (61.2%)	0.144
Gender				
Male	110 (56.4%)	57 (58.8%)	53 (54.1%)	0.564
Lab data				
WBC ($\times 10^3/\mu l$)	23.54 (0.38–423.0)	15.37 (0.38–189.1)	29.1 (0.65–423.0)	<0.001
Blasts ($\times 10^3/\mu l$)	12.03 (0–369.0)	5.95 (0–149.8)	16.45 (0–369.0)	<0.001
Hemoglobin, g/dl	7.8 (3.3–16.2)	7.6 (3.3–13.1)	8.1 (4.1–16.2)	0.212
Platelets ($\times 10^3/\mu l$)	41.0 (2–455)	33.0 (2–226)	47.0 (7–455)	0.018
LDH (U/l)	875.0 (271–8116)	867.0 (271–6885)	889.0 (274–8116)	0.587
FAB				0.007
M0	2 (1.0%)	1 (1.0)	1 (1.0%)	>0.999
M1	37 (19.0%)	18 (18.6%)	19 (19.4%)	>0.999
M2	59 (30.3%)	34 (35.1%)	25 (25.5%)	0.163
M3	17 (8.7%)	15 (15.5%)	2 (2.0%)	0.001
M4	63 (32.3%)	23 (23.7%)	40 (40.8%)	0.014
M5	11 (5.6%)	3 (3.1%)	8 (8.2%)	0.213
Karyotype SWOG[a]	*N* = 181	*N* = 94	*N* = 87	<0.001
Favorable	39 (21.5%)	32 (34.0%)	7 (8.0%)	<0.001
Intermediate	124 (68.5%)	52 (55.3%)	72 (82.8%)	0.005
Unfavorable	18 (9.9%)	10 (10.6%)	8 (9.2%)	0.630
Normal	92 (50.3%)	40 (42.6%)	52 (58.4%)	0.039
Isolated+8	8 (4.4%)	2 (2.1%)	6 (6.9%)	0.156

Abbreviations: Lab, laboratory; LDH, lactate dehydrogenase; WBC, white blood cell. [a]Southwest Oncology Group cytogenetic risk category: favorable: inv(16)/t(16;16)/del(16q), *t*(15;17) with/without secondary aberrations; *t*(8;21) lacking del(9q) or complex karyotypes; intermediate: normal, +8, +6, −*Y*, del(12p); unfavorable: del(5q)/−5, −7/del(7q), abn 3q, 9q, 11q, 20q, 21q, 17p, *t*(6;9), *t*(9;22) and complex karyotypes (⩾ 3 unrelated abn).

Table 2. Correlation of microRNA score with other gene alterations

Mutation	*Total* (n = *189*)	*MicroRNA score*		P
		Low (n = *96*)	*High* (n = *93*)	
NPM1	46 (24.3%)	17 (17.7%)	29 (31.2%)	0.041
FLT3-ITD	48 (25.4%)	12 (12.5%)	36 (38.7%)	<0.001
NPM1$^+$/*FLT3*-ITD$^-$	25 (13.2%)	10 (10.4%)	15 (16.1%)	0.287
CEBPA^{double}	13 (6.9%)	10 (10.4%)	3 (3.2%)	0.082
CEBPA	19 (10.1%)	15 (15.6%)	4 (4.3%)	0.014
WT1	11 (5.8%)	3 (3.1%)	8 (8.6%)	0.129
RUNX1	26 (13.8%)	9 (9.4%)	17 (18.3%)	0.092
IDH1	11 (5.8%)	6 (6.3%)	5 (5.4%)	>0.999
IDH2	25 (13.2%)	10 (10.4%)	15 (16.1%)	0.287
FLT3-TKD	18 (9.5%)	9 (9.4%)	9 (9.7%)	>0.999
MLL-PTD	9 (4.8%)	1 (1.0%)	8 (8.6%)	0.017
KIT	8 (4.2%)	5 (5.2%)	3 (3.2%)	0.721
KRAS	7 (3.7%)	5 (5.2%)	2 (2.2%)	0.445
NRAS	28 (14.8%)	11 (11.5%)	17 (18.3%)	0.222
ASXL1	21 (11.1%)	14 (14.6%)	7 (7.5%)	0.165
TET2	30 (15.9%)	15 (15.6%)	15 (16.1%)	>0.999
DNMT3A	33 (17.5%)	12 (12.5%)	21 (22.6%)	0.086

Comparison of prognostic significance between the scoring system and single microRNA expression

To see if our scoring system was more powerful than single microRNA expression in predicting prognosis, we added the expression of individual component microRNAs, *hsa-miR-9-5p*, *hsa-miR-155-5p* and *hsa-miR-203*, and *hsa-miR-181a*, a microRNA whose upregulation is highly associated with favorable prognosis in AML,[4,11] into the multivariate analysis in addition to the co-variables shown in Table 3. The results showed that the 3-microRNA signature outperformed all the single microRNA expression (Table 4 and Supplementary Tables S5–S7).

Correlation of microRNA prognostic signature with gene expression and biological functions

To dissect possible underlying molecular mechanisms that distinguish patients with differential microRNA risk scores, we profiled both microRNA and mRNA microarrays in 181 patients. Comparing the highest-risk group (23 patients with the risk score higher than that of cohort mean +1 s.d.) and the lowest risk group (27 patients with the risk score lower than that of cohort mean −1 s.d.), we identified 225 unique genes with differential expression (> twofold change, *t*-test $P < 0.01$, heatmap in Supplementary Figure S2, and detailed list in Supplementary Table S8). Some homeobox genes (*HOXA3*, *HOXA5*, *HOXA6*, *HOXA9*, *HOXA10*, *HOXA11AS*, *HOXB2*, *HOXB3*, *HOXB5*, *HOXB6*, *HOXB7*, *HOXB8*, *MEIS1* and *PBX3*) were found significantly upregulated in the highest-risk cohort. The overexpression of the homeobox genes is known to be associated with important gene mutations and unfavorable prognosis in AML.[8,28] IPA revealed that the 225 differentially expressed genes were associated with abundant biological functions in a variety of categories (full list in Supplementary Table S9). Among them were genes that are associated with three hematopoietic/myeloid/blood cell-related functions including proliferation of hematopoietic cells ($P = 3.93 \times 10^{-7}$), differentiation of myeloid progenitor cells ($P = 5.52 \times 10^{-3}$) and apoptosis of blood cells ($P = 2.05 \times 10^{-3}$). As GSEA can test whether genes with a specific biological function show differential presentation

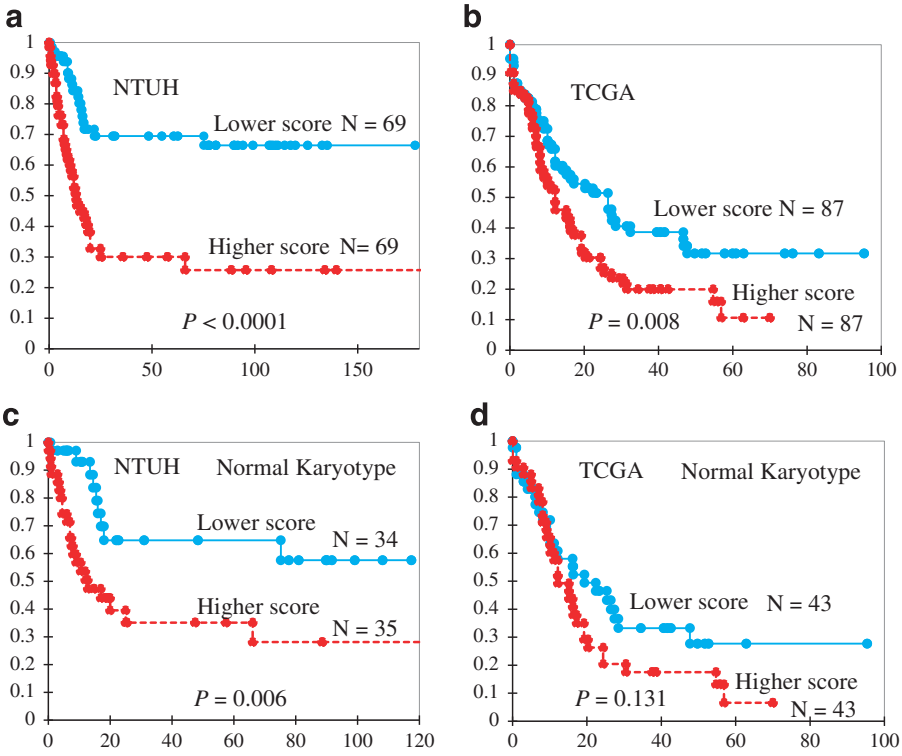

Figure 2. The Kaplan–Meier curves for overall survival (OS) according to the scores. (**a**) In NTUH discovery set, patients with lower scores have significantly longer OS than those with higher scores (median not reached vs 13.5 months, $P < 0.0001$). (**b**) In TCGA validation cohort, the scoring system still holds true (median 26.4 vs 12.2 months, $P = 0.008$). (**c**) For NTUH patients with a normal karyotype, lower scores also renders a longer OS (median not reached vs 17.0 months, $P = 0.006$). (**d**) For the TCGA patients with a normal karyotype, there is a trend ($P = 0.131$) for longer survival in those with lower microRNA scores with median survival durations of 19.3 vs 12.2 months for low and high scores, respectively.

Table 3. Multivariate analysis (Cox regression) for the overall survival in NTUH and TCGA AML cohorts

Variables	*Hazard ratio*		*95% Confidence interval*		P-*value*	
	NTUH	*TCGA*	*NTUH*	*TCGA*	*NTUH*	*TCGA*
Age[a]	2.350	2.025	1.319–4.187	1.272–3.226	0.004	0.003
WBC[b]	2.189	0.864	1.118–4.284	0.550–1.355	0.022	0.524
Karyotype[c]	1.388	1.482	0.766–2.513	1.072–2.048	0.279	0.017
NPM1/*FLT3*-ITD[d]	0.118	0.823	0.016–0.885	0.460–1.472	0.038	0.512
CEBPA[double]	0.431	—	0.179–1.037	—	0.060	—
miRNA score	2.079	1.544	1.407–3.073	1.229–1.940	< 0.001	< 0.001

Abbreviations: NTUH, National Taiwan University Hospital; TCGA, The Cancer Genome Atlas; WBC, white blood cell. [a]Age older than 50 years relative to age 50 years or younger. [b]WBC > 50 000/μl vs ⩽50 000/μl. [c]Unfavorable cytogenetics vs others. [d]*NPM1*$^+$/*FLT3*-ITD$^-$ vs other subtypes. *CEBPA* double mutation vs others.

between AML with the high and low scores and can reveal more subtle trends than just comparing the expression levels, we used this technique to verify that genes associated with these three functions were enriched in AML with the high-risk scores. Truly, GSEA-confirmed genes related to these three functions showed significantly enriched in the high-risk cohort (empirical $P < 0.001$ for proliferation of hematopoietic cells and apoptosis of blood cells; empirical $P = 0.005$ for differentiation of myeloid progenitor cells; enrichment plots in Figure 4 and Supplementary Figure S3). Totally 12 genes (*CDKN1B*, *FLT3*, *HOXA9*, *HOXB4*, *IFNG*, *IL15*, *IL6*, *IRF8*, *KITLG*, *KLF4*, *MEIS1* and *SPI1*) appeared in the core-enriched genes of GSEA in all the three functions. All these 12 genes were upregulated in the high-risk cohort. Among them, *HOXA9*, *HOXB4* and *MEIS1* were homeobox genes. Co-overexpression of *HOXA9* and *MEIS1* was reported to induce rapid transformation of BM cells to AML[29] and *HOXB4* upregulation could dramatically enhance hematopoietic stem cell expansion both *in vivo* and *ex vivo*.[30,31] Others known to predispose to leukemogenesis included *FLT3*, which is expressed in immature hematopoietic cells.[32] *FLT3*-ITD is a common mutation in AML associated with poor prognosis;[33,34] IL6 and IL15 can promote leukemia cell growth, survival and proliferation;[35,36] *SPI1* (also known as PU.1) is deregulated in AML with *CEBPA* gene mutation and heterozygous mutation of *SPI1* is associated with AML.[37,38] The roles of the other five genes in AML remain relatively uncharacterized.

Clinically practical scoring system using real-time PCR microRNA assay

As the scoring system was designed by z-transformed microRNA expression levels as inputs, cohort mean and cohort s.d. of each of the three microRNAs are required for that formula. For a practical

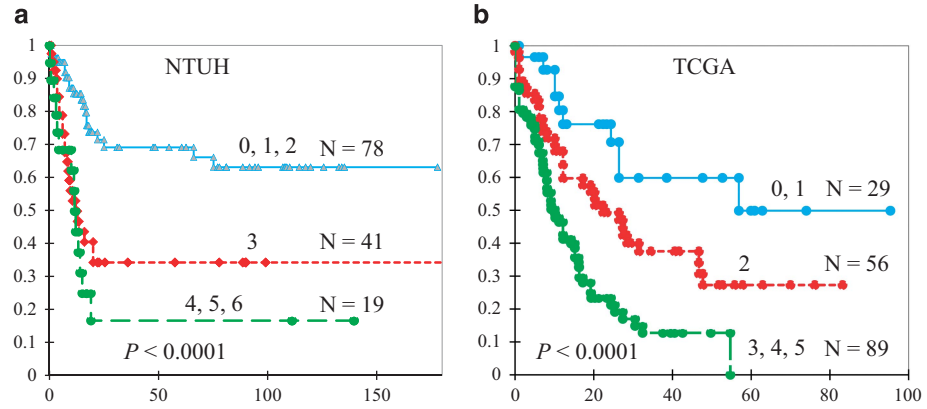

Figure 3. Kaplan–Meier curves for OS according to the numbers of unfavorable factors, as listed in Table 3. (**a**) The patients with 0, 1 or 2 unfavorable factors have longer OS than those with 3 and those with 4, 5 or 6 factors (median not reached, 12 and 11.5 months, respectively, $P < 0.0001$). (**b**) The same results also apply to TCGA cohort (median 56.9, 22.4 and 10.1 months for those with 0 or 1, 2 and ⩾3 unfavorable factors, $P < 0.0001$).

Table 4. Multivariate analysis (Cox regression) for the overall survival in NTUH and TCGA AML cohorts, including both microRNA score and *miR-181a* expression

Variables	*Hazard ratio*		*95% Confidence interval*		*P-value*	
	NTUH	*TCGA*	*NTUH*	*TCGA*	*NTUH*	*TCGA*
Age[a]	1.855	2.226	1.030–3.344	1.379–3.594	0.039	0.001
WBC[b]	1.658	1.064	0.937–2.934	0.657–1.722	0.082	0.801
Karyotype[c]	1.969	1.696	0.915–4.235	1.009–2.853	0.083	0.046
NPM1/FLT3-ITD[d]	0.572	0.933	0.240–1.361	0.523–1.663	0.206	0.813
CEBPA[double]	0.178	—	0.023–1.351	—	0.095	—
microRNA score	2.277	1.358	1.566–3.310	1.088–1.695	< 0.0001	0.007
miR-181a	0.777	0.824	0.607–0.995	0.660–1.029	0.046	0.088

Abbreviations: NTUH, National Taiwan University Hospital; TCGA, The Cancer Genome Atlas; WBC, white blood cell. [a]Age older than 50 years relative to age 50 years or younger. [b]WBC $>50\,000/\mu l$ vs $\leqslant 50\,000/\mu l$. [c]Unfavorable cytogenetics vs others. [d]*NPM1*$^{+}$/*FLT3*-ITD^{-} vs other subtypes. *CEBPA* double mutation vs others.

utilization of this scoring system in other clinical institutes or hospitals without cohort data set, we provide the calculating formula we used in the NTUH data set:

Risk = 0.4908 $(-\Delta C_{t\ hsa\text{-}miR\text{-}9\text{-}5p}+15.71)/3.60+0.2243(-\Delta C_{t\ hsa\text{-}miR\text{-}155\text{-}5p}+6.94)/1.45-0.7187(-\Delta C_{t\ hsa\text{-}miR\text{-}203}+17.16)/2.66$. Here, the ΔC_t values are C_t of the microRNA subtracting C_t of the endogenous control MAMMU6; 15.71 and 3.60 are the mean and s.d. of $\Delta C_{t\ hsa\text{-}miR\text{-}9\text{-}5p}$. The same annotation applies to *hsa-miR-155-5p* and *hsa-miR-203*. For each newly diagnosed patient, a four-well real-time PCR microRNA assay (probing *hsa-miR-9-5p*, *hsa-miR-155-5p*, *hsa-miR-203* and *MAMMU6*) is sufficient to get a prognostic score, which will then be compared with our cohort median score 0.0031 to stratify the risk group.

DISCUSSION

Although mounting evidences have shown the critical roles of microRNA in the pathogenesis and prognosis in AML, as far as we know, a global analysis of microRNA profiling to generate a scoring system for prediction of AML patients' treatment outcome is lacking.

In this report, we took advantage of an integration of comprehensive clinical, genetic and array data of our cohort to reach a simple 3-microRNA signature for prediction of clinical outcome, considering the expression levels and weights of only three microRNAs, which were sieved through repeated rounds of statistic calculation to ensure the strong and independent influence on prognosis. The power of this score was validated by TCGA cohort, an independent validation set of patients who were studied by a different microRNA quantification platform. Thus, our scoring system is independent of both patient populations and quantification methods. Although both ours and TCGA patients were not prospective cohorts, the very high significance of our 3-microRNA signature in prognosis prediction in these two independent populations suggests its high reliability for application in risk stratification. Notably, we proved that the integrated 3-microRNA scoring system outperformed the expression of individual microRNA, *hsa-miR-9-5p*, *hsa-miR-155-5p* and *hsa-miR-203*, all of which were component microRNAs in the scoring system, and *hsa-miR-181a*, whose upregulation was shown to be a highly favorable prognostic factor in AML.[4,11] Although low scores of our scoring system were associated with favorable prognostic factors such as good-risk cytogenetics and gene mutations, multivariate analyses in our cohort and that of TCGA confirmed the independence of this 3-microRNA signature to other important prognosis parameters.

There are several limitations in our study. First, although quantitative PCR-based microRNA quantification is an easy technique in the labs, the inter-laboratory standardization is a challenging task because the protocol, reference materials and quality control would pose significant problems. Second, the patient number, especially those with a normal karyotype, in ours and the TCGA cohort is relatively small. There is only a trend of OS difference between the patients with high and low microRNA scores among the patients with a normal karyotype in TCGA cohort, probably related to limited sample size and intertwining of

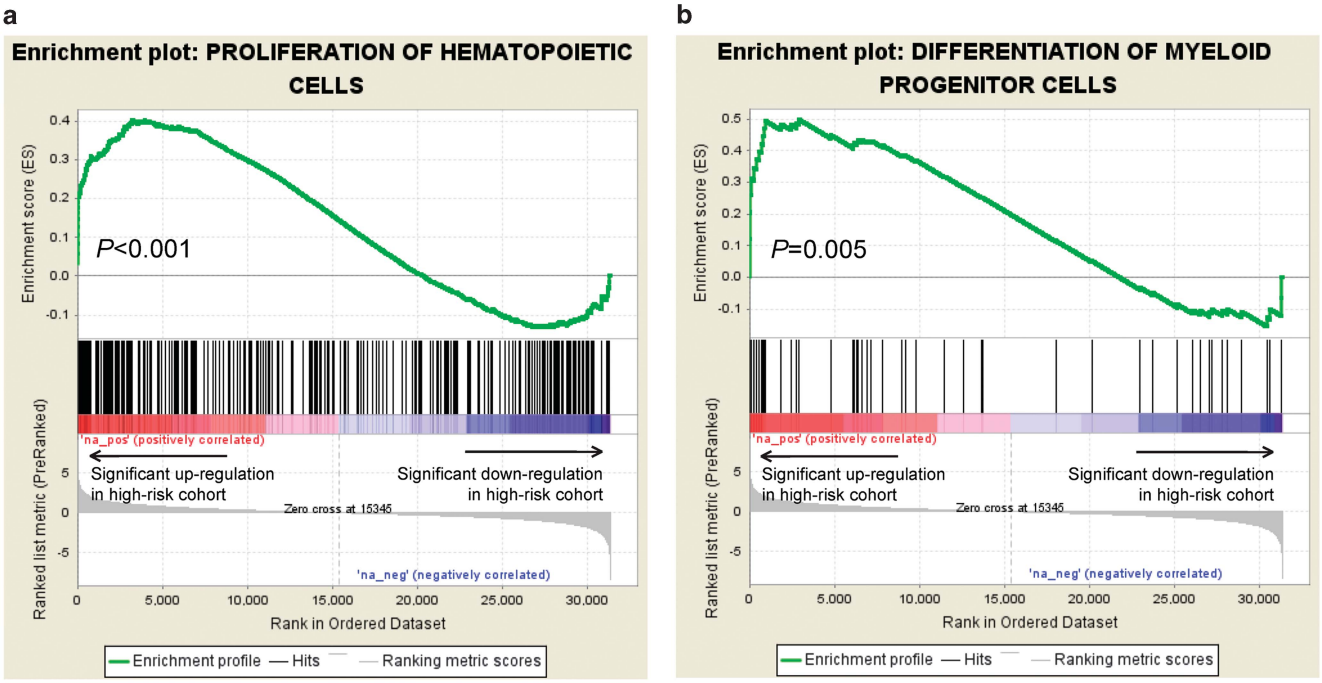

Figure 4. GSEA plots on genes associated with (**a**) proliferation of hematopoietic cells and (**b**) differentiation of myeloid progenitor cells. Genes relating to these two functions are highly upregulated in the patients with high microRNA scores, suggesting significant correlations between these two pathways and the scoring. GSEA, Gene Set Enrichment Analysis.

the survival curves in the first 10 months. Given the fact that there is no other public microRNA database with survival information for reference, we conclude that further validation by larger cohorts is necessary to confirm the prognostic significance of our microRNA scoring system in patients with a normal karyotype.

Taking advantage of available mRNA profiling data in our cohort, we had a chance to investigate the associating molecular pathways of the microRNA prognostic signature. We found expression of genes associated with cancer-related functions such as apoptosis, proliferation and stem cell properties were highly correlated with the scores. High expression of *HOX* clusters and their associated genes such as *MEIS1* and *PBX3* were associated with high scores in AML, suggesting de-regulation of stemness properties in these leukemia cells.

For hematopoietic cancers, *miR-155* upregulation was a poor prognostic factor in some studies,[16,39] but contrary result was shown in another.[40] The pathways mediating the functions of *miR-155* are very complex, but the biological consequences are largely promotion of cell proliferation, cell cycle progression and invasion/metastasis.[41,42] In our cohort, *miR-155* is an independent unfavorable prognostic factor, compatible with the previous report.[16] *MiR-9* has been shown to inhibit tumorigenicity of cancers,[43–45] but this molecule can promote metastasis of solid cancers too.[46] To make things more complicated, its increased expression was shown to be a favorable prognostic factor in medulloblastoma in one report,[47] but a poor prognostic factor in glioma in another study.[48] For AML, *miR-9* is the most specifically upregulated microRNA in *MLL*-rearranged AML compared with other types of AML, as it is a target of MLL fusion proteins, and its expression directly correlates with disease progression.[49] In our study, *miR-9* was an unfavorable prognostic factor. There are less studies about *miR-203*. This molecule functions as a suppressor of skin stemness by regulating the transition between proliferative basal progenitors and terminally differentiating suprabasal cells in the skin.[50] Its prognostic significance in AML has not been clarified, but it can target *ABL1* and suppress *BCR-ABL1* expression in chronic myeloid leukemia or some acute lymphoblastic leukemia.[51] In other cancers, it usually acts as a tumor suppressor,[52–54] but upregulation of *miR-203* in ovarian cancers is correlated with tumor progression and poor prognosis.[55] In our study, we found that *miR-203* was a favorable prognostic factor.

In conclusion, we present a simple and user-friendly 3-microRNA signature as a powerful prognostic factor for AML through multiple rounds of statistical analyses on our cohort and further validation by another independent patient group. This scoring system seems to outperform the expression of single microRNA in multivariate analysis. Paired microRNA–mRNA analyses suggest association between this signature and the common cancer-related molecular pathways.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by grants MOST103-2923-B-002-001, MOST102-2325-B-002-028 and MOST103-2314-B-002-131-MY3 from Ministry of Science and Technology (Taiwan), UN103-051 from National Taiwan University Hospital and National Taiwan University College of Medicine and MOHW102-TD-C-111-001 and MOHW103-TD-B-111-04 from Ministry of Health and Welfare (Taiwan).

AUTHOR CONTRIBUTIONS

M-KC, Y-CC and W-CC analyzed the data and wrote the paper. W-CC and H-FT designed the study and wrote the paper. H-AH and EYC provided important materials and help in the study.

REFERENCES

1 Ambros V. The functions of animal microRNAs. *Nature* 2004; **431**: 350–355.
2 Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009; **136**: 215–233.
3 Marcucci G, Mrozek K, Radmacher MD, Garzon R, Bloomfield CD. The prognostic and functional role of microRNAs in acute myeloid leukemia. *Blood* 2011; **117**: 1121–1129.

4 Marcucci G, Radmacher MD, Maharry K, Mrozek K, Ruppert AS, Paschka P *et al.* MicroRNA expression in cytogenetically normal acute myeloid leukemia. *N Engl J Med* 2008; **358**: 1919–1928.
5 Garzon R, Volinia S, Liu CG, Fernandez-Cymering C, Palumbo T, Pichiorri F *et al.* MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood* 2008; **111**: 3183–3189.
6 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.
7 Popovic R, Riesbeck LE, Velu CS, Chaubey A, Zhang J, Achille NJ *et al.* Regulation of mir-196b by MLL and its overexpression by MLL fusions contributes to immortalization. *Blood* 2009; **113**: 3314–3322.
8 Becker H, Marcucci G, Maharry K, Radmacher MD, Mrozek 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 Garzon R, Garofalo M, Martelli MP, Briesewitz R, Wang L, Fernandez-Cymering C *et al.* Distinctive microRNA signature of acute myeloid leukemia bearing cytoplasmic mutated nucleophosmin. *Proc Natl Acad Sci USA* 2008; **105**: 3945–3950.
10 Jongen-Lavrencic M, Sun SM, Dijkstra MK, Valk PJ, Lowenberg B. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. *Blood* 2008; **111**: 5078–5085.
11 Schwind S, Maharry K, Radmacher MD, Mrozek K, Holland KB, Margeson D *et al.* Prognostic significance of expression of a single microRNA, miR-181a, in cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol* 2010; **28**: 5257–5264.
12 Sun SM, Rockova V, Bullinger L, Dijkstra MK, Dohner H, Lowenberg B *et al.* The prognostic relevance of miR-212 expression with survival in cytogenetically and molecularly heterogeneous AML. *Leukemia* 2013; **27**: 100–106.
13 Xiong Y, Li Z, Ji M, Tan AC, Bemis J, Tse JV *et al.* MIR29B regulates expression of MLLT11 (AF1Q), an MLL fusion partner, and low MIR29B expression associates with adverse cytogenetics and poor overall survival in AML. *Br J Haematol* 2011; **153**: 753–757.
14 Zhu C, Wang Y, Kuai W, Sun X, Chen H, Hong Z. Prognostic value of miR-29a expression in pediatric acute myeloid leukemia. *Clin Biochem* 2013; **46**: 49–53.
15 Zhu YD, Wang L, Sun C, Fan L, Zhu DX, Fang C *et al.* Distinctive microRNA signature is associated with the diagnosis and prognosis of acute leukemia. *Med Oncol* 2012; **29**: 2323–2331.
16 Marcucci G, Maharry KS, Metzeler KH, Volinia S, Wu YZ, Mrozek K *et al.* Clinical role of microRNAs in cytogenetically normal acute myeloid leukemia: miR-155 upregulation independently identifies high-risk patients. *J Clin Oncol* 2013; **31**: 2086–2093.
17 Chou WC, Chou SC, Liu CY, Chen CY, Hou HA, Kuo YY *et al.* TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood* 2011; **118**: 3803–3810.
18 The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N Engl J Med* 2013; **368**: 2059–2074.
19 Falini B, Mecucci C, Tiacci E, Alcalay M, Rosati R, Pasqualucci L *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med* 2005; **352**: 254–266.
20 Chou WC, Hou HA, Liu CY, Chen CY, Lin LI, Huang YN *et al.* Sensitive measurement of quantity dynamics of FLT3 internal tandem duplication at early time points provides prognostic information. *Ann Oncol* 2011; **22**: 696–704.
21 Bacher U, Haferlach C, Kern W, Haferlach T, Schnittger S. Prognostic relevance of FLT3-TKD mutations in AML: the combination matters--an analysis of 3082 patients. *Blood* 2008; **111**: 2527–2537.
22 Hou HA, Kuo YY, Liu CY, Chou WC, Lee MC, Chen CY *et al.* DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. *Blood* 2012; **119**: 559–568.
23 Kramer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2014; **30**: 523–530.
24 Baldus CD, Tanner SM, Ruppert AS, Whitman SP, Archer KJ, Marcucci G *et al.* BAALC expression predicts clinical outcome of *de novo* acute myeloid leukemia patients with normal cytogenetics: a Cancer and Leukemia Group B Study. *Blood* 2003; **102**: 1613–1618.
25 Marcucci G, Baldus CD, Ruppert AS, Radmacher MD, Mrozek K, Whitman SP *et al.* Overexpression of the ETS-related gene, ERG, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J Clin Oncol* 2005; **23**: 9234–9242.
26 Marcucci G, Maharry K, Whitman SP, Vukosavljevic T, Paschka P, Langer C *et al.* High expression levels of the ETS-related gene, ERG, predict adverse outcome and improve molecular risk-based classification of cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B Study. *J Clin Oncol* 2007; **25**: 3337–3343.
27 Heuser M, Beutel G, Krauter J, Dohner K, von Neuhoff N, Schlegelberger B *et al.* High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. *Blood* 2006; **108**: 3898–3905.
28 Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R *et al.* Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* 2004; **350**: 1605–1616.
29 Kroon E, Krosl J, Thorsteinsdottir U, Baban S, Buchberg AM, Sauvageau G. Hoxa9 transforms primary bone marrow cells through specific collaboration with Meis1a but not Pbx1b. *EMBO J* 1998; **17**: 3714–3725.
30 Sauvageau G, Thorsteinsdottir U, Eaves CJ, Lawrence HJ, Largman C, Lansdorp PM *et al.* Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations *in vitro* and *in vivo*. *Genes Dev* 1995; **9**: 1753–1765.
31 Antonchuk J, Sauvageau G, Humphries RK. HOXB4-induced expansion of adult hematopoietic stem cells . *Cell* 2002; **109**: 39–45.
32 Gilliland DG, Griffin JD. The roles of FLT3 in hematopoiesis and leukemia. *Blood* 2002; **100**: 1532–1542.
33 Yanada M, Matsuo K, Suzuki T, Kiyoi H, Naoe T. Prognostic significance of FLT3 internal tandem duplication and tyrosine kinase domain mutations for acute myeloid leukemia: a meta-analysis. *Leukemia* 2005; **19**: 1345–1349.
34 Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J *et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* 2012; **366**: 1079–1089.
35 Nakajima K, Yamanaka Y, Nakae K, Kojima H, Ichiba M, Kiuchi N *et al.* A central role for Stat3 in IL-6-induced regulation of growth and differentiation in M1 leukemia cells. *EMBO J* 1996; **15**: 3651–3658.
36 Meazza R, Basso S, Gaggero A, Detotero D, Trentin L, Pereno R *et al.* Interleukin (IL)-15 induces survival and proliferation of the growth factor-dependent acute myeloid leukemia M-07e through the IL-2 receptor beta/gamma. *Int J Cancer* 1998; **78**: 189–195.
37 Mueller BU, Pabst T, Osato M, Asou N, Johansen LM, Minden MD *et al.* Heterozygous PU.1 mutations are associated with acute myeloid leukemia. *Blood* 2002; **100**: 998–1007.
38 Marcucci G, Maharry K, Radmacher MD, Mrozek 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.
39 Cui B, Chen L, Zhang S, Mraz M, Fecteau JF, Yu J *et al.* MicroRNA-155 influences B-cell receptor signaling and associates with aggressive disease in chronic lymphocytic leukemia. *Blood* 2014; **124**: 546–554.
40 Palma CA, Al Sheikha D, Lim TK, Bryant A, Vu TT, Jayaswal V *et al.* MicroRNA-155 as an inducer of apoptosis and cell differentiation in acute myeloid leukaemia. *Mol Cancer* 2014; **13**: 79.
41 Chen Z, Ma T, Huang C, Hu T, Li J. The pivotal role of microRNA-155 in the control of cancer. *J Cell Physiol* 2014; **229**: 545–550.
42 Neumeister P, Sill H. Novel face of microRNA-155. *Blood* 2014; **123**: 5–7.
43 Zhang H, Qi M, Li S, Qi T, Mei H, Huang K *et al.* microRNA-9 targets matrix metalloproteinase 14 to inhibit invasion, metastasis, and angiogenesis of neuroblastoma cells. *Mol Cancer Ther* 2012; **11**: 1454–1466.
44 Selcuklu SD, Donoghue MT, Rehmet K, de Souza Gomes M, Fort A, Kovvuru P *et al.* MicroRNA-9 inhibition of cell proliferation and identification of novel miR-9 targets by transcriptome profiling in breast cancer cells. *J Biol Chem* 2012; **287**: 29516–29528.
45 Gomez GG, Volinia S, Croce CM, Zanca C, Li M, Emnett R *et al.* Suppression of microRNA-9 by mutant EGFR signaling upregulates FOXP1 to enhance glioblastoma tumorigenicity. *Cancer Res* 2014; **74**: 1429–1439.
46 White RA, Neiman JM, Reddi A, Han G, Birlea S, Mitra D *et al.* Epithelial stem cell mutations that promote squamous cell carcinoma metastasis. *J Clin Invest* 2013; **123**: 4390–4404.
47 Fiaschetti G, Abela L, Nonoguchi N, Dubuc AM, Remke M, Boro A *et al.* Epigenetic silencing of microRNA-9 is associated with HES1 oncogenic activity and poor prognosis of medulloblastoma. *Br J Cancer* 2014; **110**: 636–647.
48 Wu Z, Wang L, Li G, Liu H, Fan F, Li Z *et al.* Increased expression of microRNA-9 predicts an unfavorable prognosis in human glioma. *Mol Cell Biochem* 2013; **384**: 263–268.
49 Chen P, Price C, Li Z, Li Y, Cao D, Wiley A *et al.* miR-9 is an essential oncogenic microRNA specifically overexpressed in mixed lineage leukemia-rearranged leukemia. *Proc Natl Acad Sci USA* 2013; **110**: 11511–11516.
50 Yi R, Poy MN, Stoffel M, Fuchs E. A skin microRNA promotes differentiation by repressing 'stemness'. *Nature* 2008; **452**: 225–229.
51 Bueno MJ, Perez de Castro I, Gomez de Cedron M, Santos J, Calin GA, Cigudosa JC *et al.* Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression. *Cancer Cell* 2008; **13**: 496–506.

52 Okumura T, Shimada Y, Moriyama M, Takei Y, Omura T, Sekine S *et al.* MicroRNA-203 inhibits the progression of esophageal squamous cell carcinoma with restored epithelial tissue architecture *in vivo*. *Int J Oncol* 2014; **44**: 1923–1932.
53 Taube JH, Malouf GG, Lu E, Sphyris N, Vijay V, Ramachandran PP *et al.* Epigenetic silencing of microRNA-203 is required for EMT and cancer stem cell properties. *Sci Rep* 2013; **3**: 2687.
54 Takeshita N, Mori M, Kano M, Hoshino I, Akutsu Y, Hanari N *et al.* miR-203 inhibits the migration and invasion of esophageal squamous cell carcinoma by regulating LASP1. *Int J Oncol* 2012; **41**: 1653–1661.
55 Wang S, Zhao X, Wang J, Wen Y, Zhang L, Wang D *et al.* Upregulation of microRNA-203 is associated with advanced tumor progression and poor prognosis in epithelial ovarian cancer. *Med Oncol* 2013; **30**: 681.

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