Prognostic and biologic significance of long non-coding RNA profiling in younger adults with cytogenetically normal acute myeloid leukemia

Dimitrios Papaioannou,[1] Deedra Nicolet,[1,2] Stefano Volinia,[3] Krzysztof Mrózek,[1] Pearlly Yan,[1] Ralf Bundschuh,[4] Andrew J. Carroll,[5] Jessica Kohlschmidt,[1,2] William Blum,[1] Bayard L. Powell,[6] Geoffrey L. Uy,[7] Jonathan E. Kolitz,[8] Eunice S. Wang,[9] Ann-Kathrin Eisfeld,[1] Shelley J. Orwick,[1] David M. Lucas,[1] Michael A. Caligiuri,[1] Richard M. Stone,[10] John C. Byrd,[1] Ramiro Garzon[1*] and Clara D. Bloomfield[1*]

**RG and CDB contributed equally to this work.*

[1]The Ohio State University Comprehensive Cancer Center, Columbus, OH, USA; [2]The Alliance for Clinical Trials in Oncology Statistics and Data Center, Mayo Clinic, Rochester, MN, USA; [3]Department of Morphology, Surgery and Experimental Medicine, University of Ferrara, Italy; [4]Department of Physics, Department of Chemistry & Biochemistry, Division of Hematology, Department of Internal Medicine, Center for RNA Biology, The Ohio State University, Columbus, OH, USA; [5]Department of Genetics, University of Alabama at Birmingham, AL, USA; [6]The Comprehensive Cancer Center of Wake Forest University, Winston-Salem, NC, USA; [7]Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA; [8]Hofstra North Shore-Long Island Jewish School of Medicine, Lake Success, NY, USA; [9]Roswell Park Cancer Institute, Buffalo, NY, USA and [10]Dana-Farber Cancer Institute, Harvard University, Boston, MA, USA

 doi:10.3324/haematol.2017.166215

Received: February 7, 2017.

Accepted: May 2, 2017.

Pre-published: May 4, 2017.

Correspondence: ramiro.garzon@osumc.edu or clara.bloomfield@osumc.edu

Supplementary Appendix

Prognostic and biologic significance of long non-coding RNA profiling in younger adults with cytogenetically normal acute myeloid leukemia

Dimitrios Papaioannou,[1] Deedra Nicolet,[1,2] Stefano Volinia,[3] Krzysztof Mrózek,[1] Pearlly Yan,[1] Ralf Bundschuh,[4] Andrew J. Carroll,[5] Jessica Kohlschmidt,[1,2] William Blum,[1] Bayard L. Powell,[6] Geoffrey L. Uy,[7] Jonathan E. Kolitz,[8] Eunice S. Wang,[9] Ann-Kathrin Eisfeld,[1] Shelley J. Orwick,[1] David M. Lucas,[1] Michael A. Caligiuri,[1] Richard M. Stone,[10] John C. Byrd,[1] Ramiro Garzon[1]* and Clara D. Bloomfield[1]*

[1]The Ohio State University Comprehensive Cancer Center, Columbus, OH; [2]Alliance for Clinical Trials in Oncology Statistics and Data Center, Mayo Clinic, Rochester, MN; [3]Department of Morphology, Surgery and Experimental Medicine, University of Ferrara, Ferrara, Italy; [4]Department of Physics, Department of Chemistry & Biochemistry, Division of Hematology, Department of Internal Medicine, Center for RNA Biology, The Ohio State University, Columbus, OH; [5]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; [6]The Comprehensive Cancer Center of Wake Forest University, Winston-Salem, NC; [7]Siteman Cancer Center, Washington University School of Medicine in St. Louis, St. Louis, MO; [8]North Shore Cancer Institute, Lake Success, NY; [9]Roswell Park Cancer Institute, Buffalo, NY; [10]Dana-Farber Cancer Institute, Harvard University, Boston, MA

*R.G and C.D.B. contributed equally to this study.

Participating institutions

The following Cancer and Leukemia Group B (CALGB)/Alliance for Clinical Trials in Oncology (Alliance) institutions participated in this study and contributed at least two patients. For each of these institutions, the current or last principal investigators are listed as follows:

The Ohio State University Medical Center, Columbus, OH: Clara D. Bloomfield; Wake Forest University School of Medicine, Winston-Salem, NC: Heidi Klepin; Washington University School of Medicine, St. Louis, MO: Nancy L. Bartlett; Dana Farber Cancer Institute, Boston, MA: Harold J. Burstein; North Shore University Hospital, Manhasset, NY: Daniel R. Budman; Roswell Park Cancer Institute, Buffalo, NY: Ellis G. Levine; University of Chicago Medical Center, Chicago, IL: Hedy L. Kindler; University of Iowa Hospitals, Iowa City, IA: Daniel A. Vaena; University of North Carolina, Chapel Hill, NC: Thomas C. Shea; Ft. Wayne Medical Oncology/Hematology, Ft. Wayne, IN: Sreenivasa Nattam; University of Maryland Cancer Center, Baltimore, MD: Martin J. Edelman; Christiana Care Health Services, Inc., Newark, DE: Gregory Masters; Dartmouth Medical School, Lebanon, NH: Konstantin Dragnev; Duke University Medical Center, Durham, NC: Jeffrey Crawford; University of Vermont Cancer Center, Burlington, VT: Claire Verschraegen; Eastern Maine Medical Center, Bangor, ME: Thomas H. Openshaw; Mount Sinai School of Medicine, New York, NY: Lewis R. Silverman; Weill Medical College of Cornell University, New York, NY: Scott Tagawa; University of Massachusetts Medical Center, Worcester, MA: William V. Walsh; University of California at San Francisco: Charalambos Andreadis; Western Pennsylvania Hospital, Pittsburgh, PA: John Lister; University of Puerto Rico School of Medicine, San Juan, PR: Eileen I. Pacheco; SUNY Upstate Medical University, Syracuse, NY:

Stephen L. Graziano; University of Alabama at Birmingham: Robert Diasio; Massachusetts General Hospital, Boston, MA: David Ryan, Justin Gainor; Rhode Island Hospital, Providence, RI: Howard Safran; University of Illinois, Chicago, IL: Arkadiusz Z. Dudek; Moores University of California San Diego Cancer Center, San Diego, CA: Barbara A. Parker; Walter Reed National Military Medical Center, Bethesda, MD: Mary Kwok; Virginia Commonwealth University, Richmond, VA: Steven Grossman; University of Minnesota, Minneapolis, MN: Bruce A. Peterson; Long Island Jewish Medical Center, Lake Success, NY: Daniel R. Budman; University of Missouri/Ellis Fischel Cancer Center, Columbia, MO: Clint Kingsley; Nevada Cancer Research Foundation NCORP, Las Vegas, NV: John Ellerton; University of Tennessee Cancer Center, Memphis, TN: Harvey B. Niell; University of Nebraska Medical Center, Omaha, NE: Apar Ganti.

Patients and Methods

Treatment protocols

All patients included in our study were treated on CALGB/Alliance first-line protocols for patients with acute myeloid leukemia (AML), and received cytarabine/daunorubicin-based induction therapy.[1] Per protocol, all patients were to receive at least one induction cycle. For patients with residual leukemia present in a bone marrow (BM) biopsy after one induction cycle, a second cycle of induction was administered. None of the protocols included allogeneic stem cell transplantation (SCT) in first complete remission (CR). Patients enrolled on the treatment protocols also provided written informed consent to participate in the companion protocols CALGB 20202 (molecular studies in AML), CALGB 8461 (prospective cytogenetic companion), and CALGB 9665 (leukemia tissue bank), which involved collection of pretreatment BM aspirates and blood samples.

Patients were enrolled on the following treatment protocols: CALGB 19808, CALGB 10503, CALGB 9621, CALGB 10603, CALGB 9222, CALGB 8525, CALGB 9022 and CALGB 8721, CALGB 8821 and CALGB 9120. Patients enrolled onto CALGB 19808 (n=114) were randomly assigned to receive induction chemotherapy with cytarabine/daunorubicin, and etoposide with or without PSC-833 (valspodar), a multidrug resistance protein inhibitor.[1] Upon attainment of CR, patients were assigned to intensification with high-dose cytarabine and etoposide for stem-cell mobilization followed by myeloablative treatment with busulfan and etoposide supported by autologous peripheral blood SCT. Patients on CALGB 10503 (n=113) received cytarabine/daunorubicin-based induction chemotherapy and those who achieved CR further received a two-step consolidation with chemo-mobilization and

autologous SCT if eligible, or high-dose cytarabine-based consolidation if not. Maintenance with decitabine began as soon as possible after recovery from consolidation.[2] Patients enrolled onto CALGB 9621 (n=61) were treated similarly to those on CALGB 19808, as previously reported.[3] Patients on CALGB 10603 (n=40) were stratified by *FLT3* mutation subtype [*FLT3*-tyrosine kinase domain mutations (*FLT3*-TKD) *versus FLT3*-internal tandem duplication (*FLT3*-ITD)-high allelic ratio (>0.7) *versus FLT3*-ITD low allelic ratio (0.05-0.7)], and were randomized to receive cytarabine/daunorubicin-based induction chemotherapy and high-dose cytarabine consolidation in combination with either the multi-kinase inhibitor midostaurin or placebo. One-year midostaurin or placebo maintenance was administered after the last cycle of consolidation therapy.[4] Patients enrolled on CALGB 9222 (n=27) received cytarabine/daunorubicin-based induction chemotherapy, and those who achieved CR received either three cycles of high-dose cytarabine or three cycles of a so-called non cross-resistant regimen (the first cycle of this regimen was high-dose cytarabine, the second was cyclophosphamide plus etoposide, and the third was mitoxantrone plus diaziquone).[5] Patients enrolled onto CALGB 8525 (n=17) who achieved CR after cytarabine/daunorubicin-based induction chemotherapy were randomly assigned to consolidation with different doses of cytarabine followed by maintenance treatment.[6] Patients who participated in CALGB 9022 (n=2) and achieved CR after cytarabine/daunorubicin-based induction chemotherapy received one course of high-dose cytarabine consolidation, followed by one course of cyclophosphamide and etoposide, followed by one course of mitoxantrone and diaziquone (AZQ).[7]

The patients analyzed in this work had similar disease-free survival (DFS, *P*=0.56) as the other CALGB/Alliance patients treated on the same protocols but not included in the study.

Transcriptome analysis: library generation, sequencing and data analysis

Extracted total RNA was assessed for quality on an Agilent 2100 Bioanalyzer (BioA) using the RNA 6000 Nanochip and for quantity on a Qubit 2.0 Fluorometer (Agilent Technologies, Santa Clara, CA) using the RNA HS Assay Kit. Samples with a RNA Integrity Number (RIN) greater than four, with no visible sign of genomic DNA (gDNA) contamination and a concentration of >40 ng/μL were used for total RNA library generation. RNA-seq libraries were prepared using the Illumina TruSeq Stranded Total RNA Sample Prep Kit with RiboZero Gold (#RS1222201) according to the manufacturer's instructions. Sequencing was performed with the Illumina HiSeq 2500 system using the HiSeq version 3 sequencing reagents to an approximate cluster density of 800,000/mm^2. Image analysis, base calling, error estimation, and quality thresholds were performed using the HiSeq Controller Software (version 2.2.38) and the Real Time Analyzer (RTA) software (version 1.18.64).

Small RNA sequencing (smRNA-seq) libraries were generated using the NEBNext Multiplex Small RNA Library Prep Set (Cat #: E7300L; New England Biolabs, Inc., Ipswich, MA). Library generation steps were performed as described by the manufacturer. The input RNA criteria for smRNA-seq were: a Qubit RNA concentration of >50 ng/μL and a BioA RNA RIN value >7. Generation of barcodes and enrichment of fragments with 3- and 5- adaptors for smRNA libraries were accomplished by 12 cycles of PCR amplification. Prior to pooling smRNA-seq libraries for enriching smRNA species, libraries generated from each sample were assessed for relative amount of smRNA fragments migrating between 140 to 160 bp using the Agilent Bioanalyzer HS DNA assay. Size selection/enrichment for smRNAs was accomplished using the Sage Science Pippin Prep (Beverly, MA) with 3% pre-

cast agarose gel. The profile of the resultant size-selected libraries was ascertained using the Agilent Bioanalyzer HS DNA assay. Each pool of the smRNA-seq libraries was sequenced with other samples with compatible barcodes on an Illumina HiSeq 2500 V3 single-read 50bp lane to achieve 5-8 million passed filter reads/sample.

Cutadapt and FastQC were used to apply quality control and adapter trimming to FastQ files. The Spliced Transcripts Alignment to a Reference (STAR) software[8] was used to align the short reads to the human genome (GENECODE ver22)[9] and the HTSeq program[10] to quantify and annotate long non-coding RNAs (lncRNAs). Raw data were transformed into reads per million (RPM) prior to statistical analysis. To minimize noise, mRNAs were evaluated in each sample only when at least nine reads were present in a total of 40 million reads.

Definition of clinical endpoints

Clinical endpoints were defined according to generally accepted criteria.[11] CR required a BM aspirate with cellularity >20% with maturation of all cell lines, <5% blasts and undetectable Auer rods; in peripheral blood, an absolute neutrophil count of ≥1.5 x 10^9/L, platelet count of >100 x 10^9/L, and leukemic blasts absent; and no evidence of extramedullary leukemia, all of which had to persist for ≥4 weeks.[9] Relapse was defined by the presence of ≥5% BM blasts, or circulating leukemic blasts, or the development of extramedullary leukemia. Disease-free survival (DFS) was measured from the date of CR until the date of relapse or death (from any cause); patients alive and in continuous first CR were censored at last follow-up. Overall survival (OS) was measured from the date of study entry until the date of death (from any cause); patients alive at last follow-up were censored. Event-free survival (EFS) was measured from the date of study entry until the date of failure to

achieve CR, relapse or death. Patients alive and in CR at last follow-up were censored.

Statistical analyses

Multivariable proportional hazards models were constructed for DFS, OS and EFS, using a limited backwards elimination procedure. Variables considered for model inclusion were: lncRNA score status (favorable *versus* unfavorable), age (as a continuous variable, in 10-year increments), sex (male *versus* female), race (white *versus* non-white), white blood cell count [(WBC) as a continuous variable, in 50-unit increments], hemoglobin (as a continuous variable, in 1-unit increments), platelet count (as a continuous variable, in 50-unit increments), extramedullary involvement (present *versus* absent), *ASXL1* mutations (mutated *versus* wild-type), *CEBPA* mutations (double-mutated *versus* single-mutated or wild-type), *DNMT3A* mutations (mutated *versus* wild-type), *FLT3*-ITD (present *versus* absent), *FLT3*-TKD (present *versus* absent), *IDH1* mutations (mutated *versus* wild-type), *IDH2* mutations (mutated *versus* wild-type), *NPM1* mutations (mutated *versus* wild-type), *TET2* mutations (mutated *versus* wild-type), *RUNX1* mutations (mutated *versus* wild-type), *WT1* mutations (mutated *versus* wild-type), *ERG* expression levels (high *versus* low), *BAALC* expression levels (high *versus* low), *MN1* expression levels (high *versus* low), miR-181a expression levels (high *versus* low), miR-3151 (expressed *versus* not expressed), and miR-155 expression levels (high *versus* low). For *ERG*, *BAALC*, *MN1*, miR-181a and miR-155 the median expression value was used as the cut point to divide patients into high and low expressers. Variables significant at α=0.20 from the univariable analyses were considered for multivariable analyses. Interactions between the variables and the lncRNA score were checked throughout the model

building process. For the time-to-event endpoints, the proportional hazards assumption was checked for each variable individually.

Derivation of gene mutation-related lncRNA signatures

First, to ensure that only lncRNAs, which are differentially expressed among patient samples would be analyzed, we removed transcripts with ≤1.5 fold change in expression levels in either direction from the median value in ≥20% of the analyzed samples. We also excluded lncRNAs with no detectable expression in ≥50% of the tested samples. We analyzed the training cohort (n=263) so as to identify the lncRNAs that were differentially expressed between patients harboring different recurrent gene mutations and those with wild-type alleles. Only gene mutations with at least nine cases in either the training or the validation set (n=114) were examined (i.e., *NPM1*, *CEBPA*, *DNMT3A* R882 and non-R882, *IDH2* R140, *TET2*, *WT1* and *FLT3*-ITD mutations). For this we used the univariable t-test within the BRB tools (linus.nci.nih.gov/BRB-ArrayTools.html) and included only lncRNA transcripts that had a $P<0.001$ and a fold change ≥2. The capacity of each of these mutation-related lncRNA signatures in distinguishing between samples that were positive or negative for their corresponding molecular alteration was tested on the validation set by repeated K-fold (K=10) cross-validations. The cross-validated estimate of misclassification error is an estimate of the prediction error for model fit applying the specified algorithm to the full dataset. Based on the 100 random permutations test, the *P*-values for the nearest centroid classifier and the support vector machines classifier of each signature were computed and only signatures significant ($P<0.01$) in both tests are reported. Heat maps of the lncRNA-expression signatures associated with *CEBPA* double-mutations, *NPM1* mutations and *FLT3*-ITD in younger patients with *de novo* CN-AML were derived using the lncRNAs included in each signature (supplemental Tables S7–S9). Patients were grouped by mutation status, and genes were ordered by hierarchical cluster analysis.

Computation of lncRNA score

To develop a prognostic lncRNA score, we first identified 24 lncRNAs that were associated with EFS ($P<10^{-6}$) by univariable Cox proportional hazard models in the 263 patients with *de novo* CN-AML (training set). The lncRNA score was derived as a linear combination of the expression of the 24 lncRNAs. The lncRNA score for patient i was ci = Σ wj xij, where xij was the expression value for gene j in patient i, and wj was the weight assigned to gene j. The univariable Cox regression coefficients for EFS for each of the 24 lncRNAs included in the score were used as the weights (wj) in the lncRNA score.

Correlation of lncRNA scores with mRNA and miRNA expression levels

To investigate interactions between the lncRNA score and mRNA and miR expression levels, 300 patients with available total transcriptome and smRNA-seq data were analyzed. The Spearman correlation coefficient was calculated to identify the miRNAs ($P<0.01$) and mRNAs ($P<0.001$) that had the strongest association with the lncRNA score. Independent analyses were performed in the patients in the training set (n=207) and those in the validation set (n=93), and the results were then intersected (i.e., only the transcripts that were found to significantly correlate with the lncRNA score in both sets of patients are reported).

Supplementary References

1. Kolitz JE, George SL, Marcucci G, et al. P-glycoprotein inhibition using valspodar (PSC-833) does not improve outcomes for patients younger than age 60 years with newly diagnosed acute myeloid leukemia: Cancer and Leukemia Group B study 19808. Blood. 2010;116(9):1413-1421.

2. Blum W, Sanford BL, Klisovic R, et al. Maintenance therapy with decitabine in younger adults with acute myeloid leukemia in first remission: a phase 2 Cancer and Leukemia Group B study (CALGB 10503). Leukemia. 2017;31(1):34-39.

3. Kolitz JE, George SL, Dodge RK, et al. Dose escalation studies of cytarabine, daunorubicin, and etoposide with and without multidrug resistance modulation with PSC-833 in untreated adults with acute myeloid leukemia younger than 60 years: final induction results of Cancer and Leukemia Group B study 9621. J Clin Oncol. 2004;22(21):4290-4301.

4. Stone R, Mandrekar S, Sanford B, et al. The multi-kinase inhibitor midostaurin (M) prolongs survival compared with placebo (P) in combination with daunorubicin (D)/cytarabine (C) induction (ind), high-dose C consolidation (consol), and as maintenance (maint) therapy in newly diagnosed acute myeloid leukemia (AML) patients (pts) age 18-60 with FLT3 mutations (muts): an international prospective randomized (pand) P-controlled double-blind trial (CALGB 10603/RATIFY [Alliance]). Blood. 2015;126(23). (abstract 6).

5. Moore JO, George SL, Dodge RK, et al. Sequential multiagent chemotherapy is not superior to high-dose cytarabine alone as postremission intensification therapy for acute myeloid leukemia in adults under 60 years of age: Cancer and Leukemia Group B study 9222. Blood. 2005;105(9):3420-3427.

6. Mayer RJ, Davis RB, Schiffer CA, et al. Intensive postremission chemotherapy in adults with acute myeloid leukemia. N Engl J Med. 1994;331(14):896-903.

7. Moore JO, Dodge RK, Amrein PC, et al. Granulocyte-colony stimulating factor (filgrastim) accelerates granulocyte recovery after intensive postremission chemotherapy for acute myeloid leukemia with aziridinyl benzoquinone and mitoxantrone: Cancer and Leukemia Group B study 9022. Blood. 1997;89(3):780-788.

8. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29(1):15-21.

9. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22(9):1760-1774.

10. Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 2015;31(2):166-169.

11. Cheson BD, Cassileth PA, Head DR, et al. Report of the National Cancer Institute-sponsored workshop on definitions of diagnosis and response in acute myeloid leukemia. J Clin Oncol. 1990;8(5):813-819.

12. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129(4):424-447.

Table S1. Comparison of clinical and molecular characteristics of training and validation sets of younger adult patients with cytogenetically normal acute myeloid leukemia.

Characteristic	Training set (n=263)	Validation set (n=114)	*P*
Age, years			0.45
Median	47	45	
Range	17-59	18-59	
Sex, n. (%)			0.82
Male	136 (52)	57 (50)	
Female	127 (48)	57 (50)	
Race, n. (%)			0.85
White	234 (91)	101 (90)	
Non-white	24 (9)	11 (10)	
Hemoglobin (g/dL)			0.16
Median	9.4	8.8	
Range	4.6-14.4	4.2-25.1	
Platelet count (x10^9/L)			0.75
Median	59	53	
Range	8-445	8-433	
WBC count (x10^9/L)			0.99
Median	27.2	30.3	
Range	0.6-308.8	0.9-475.0	
Blood blasts, %			0.03
Median	63	50	
Range	0-97	0-97	
Bone marrow blasts, %			0.41
Median	70	66	
Range	10-96	18-95	
Extramedullary involvement, n. (%)	77 (30)	33 (30)	1.00
NPM1, n. (%)			0.14
Mutated	149 (57)	74 (65)	
Wild-type	114 (43)	40 (35)	
FLT3-ITD, n. (%)			0.73
Present	99 (38)	45 (41)	
Absent	160 (62)	66 (59)	
CEBPA, n. (%)			0.63
Double Mutated	40 (16)	14 (13)	
Wild-type	214 (84)	91 (87)	
FLT3-TKD, n. (%)			0.02
Present	33 (13)	5 (5)	
Absent	225 (87)	105 (95)	

Characteristic	Training set (n=263)	Validation set (n=114)	*P*
WT1, n. (%)			0.59
Mutated	27 (11)	14 (13)	
Wild-type	229 (89)	96 (87)	
TET2, n. (%)			0.28
Mutated	32 (13)	9 (8)	
Wild-type	224 (88)	101 (92)	
IDH1, n. (%)			0.53
Mutated	22 (9)	7 (6)	
Wild-type	235 (91)	104 (94)	
IDH2, n. (%)			0.34
Mutated	22 (9)	13 (12)	
R140	18	10	
R172	4	3	
Wild-type	235 (91)	98 (88)	
ASXL1, n. (%)			0.76
Mutated	10 (4)	3 (3)	
Wild-type	242 (96)	105 (97)	
DNMT3A, n. (%)			0.91
Mutated	102 (40)	43 (39)	
R882	72	34	
Non-R882	30	9	
Wild-type	154 (60)	67 (61)	
RUNX1, n. (%)			0.63
Mutated	16 (6)	5 (5)	
Wild-type	240 (94)	105 (95)	
ELN Genetic Group,* n. (%)			0.41
Favorable	135 (54)	60 (57)	
Intermediate-I	66 (27)	31 (30)	
Adverse	48 (19)	14 (13)	
ERG expression group,† n. (%)			0.01
High	142 (54)	45 (40)	
Low	119 (46)	68 (60)	
BAALC expression group,† n. (%)			0.002
High	133 (56)	40 (37)	
Low	106 (44)	67 (63)	
MN1 expression group,† n. (%)			0.07
High	133 (53)	47 (42)	
Low	117 (47)	64 (56)	
miR-181a expression group,† n. (%)			0.26
High	109 (53)	42 (45)	
Low	98 (47)	51 (55)	
miR-3151, n. (%)			0.25
Expressed	39 (19)	12 (13)	
Not expressed	168 (81)	81 (87)	
miR-155 expression group,† n. (%)			1.00
High	105 (51)	47 (51)	
Low	102 (49)	46 (49)	

WBC: white blood cell; ELN: European LeukemiaNet; *FLT3*-ITD: internal tandem duplication of the *FLT3* gene; *FLT3*-TKD: tyrosine kinase domain mutation in the *FLT3* gene.

* Among patients with cytogenetically normal acute myeloid leukemia (CN-AML), the ELN Favorable Risk Category comprises patients with double-mutated *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN Intermediate Risk Category includes patients with wild-type *NPM1* without *FLT3*-ITD or wild-type *NPM1* and *FLT3*-ITD^{low} or mutated *NPM1* and *FLT3*-ITD^{high}. The ELN Adverse Risk Category comprises patients with *NPM1* and *FLT3*-ITD^{high}, and/or mutated *RUNX1* (if it does not co-occur with a Favorable AML subtype) and/or mutated *ASXL1* (if it does not co-occur with a Favorable AML subtype) and/or mutated *TP53*. *FLT3*-ITD^{low} is defined by a *FLT3*-ITD/*FLT3* wild-type allelic ratio of less than 0.5 and *FLT3*-ITD^{high} is defined as by a *FLT3*-ITD/*FLT3* wild-type allelic ratio of equal to or more than 0.5.[12]

† The median expression value was used as a cut point.

Table S2. List of the 24 long non-coding RNAs which associate with event-free survival ($P<10^6$) in the training set of younger adult patients with cytogenetically normal acute myeloid leukemia.

Name	Class	Genomic location	Cox EFS regression co-efficients
AC006129.2	lincRNA	chr19:41,549,520-41,550,705	0.24
AF064858.11	lincRNA	chr21: 39,006,648-39,011,329	0.53
AL122127.25	lincRNA	chr14:105,894,646-105,896,577	0.22
CITF22-49E9.3	lincRNA	chr22:49,933,198-49,934,074	0.39
CTC-455F18.3	processed pseudogene	chr5:170,896,929-170,904,461	-0.55
DND1P1	processed pseudogene	chr17:45,585,871-45,586,929	0.34
GCNT1P3	processed pseudogene	chr3:190,624,020-190,625,744	0.37
KMT2E-AS1	antisense	chr7:105,013,425-105,014,321	0.39
MIR155HG	lincRNA	chr21:25,562,145-25,575,168	0.30
NPHP3-AS1	antisense	chr3:132,722,342-132,874,223	-0.26
PSMD6-AS2	antisense	chr3:64,004,022-64,012,148	-0.42
RP4-673M15.1	antisense	chr7:44,884,953-44,886,393	0.33
*RP4-728D4*2*	antisense	chr1: 35,569,813-35,577,729	0.37
*RP11-121A14*3*	sense intronic	chr9: 124,262,876-124,265,809	0.38
RP11-327P2.5	antisense	chr13: 51,803,838-51,813,832	0.41
RP11-333E13.4	transcribed processed pseudogene	chr4: 40,042,917-40,057,199	0.35
*RP11-440L14*1*	antisense	chr4: 762,387-781,849	0.53
RP11-815J21.1	sense intronic	chr15:85,701,109-85,702,771	0.47
*RP11-946L16*2*	antisense	chr12: 29,156,448-29,156,991	0.19
*RP11-1017G21*5*	lincRNA	chr14: 101,948,347-101,949,425	0.35
RP13-516M14.1	lincRNA	chr17: 82,293,716-82,294,910	0.47
SDHAP3	transcribed unprocessed pseudogene	chr5:1,572,222-1,593,289	0.33
SENCR	antisense	chr11:128,691,672-128,696,023	0.29
SMARCE1P1	processed pseudogene	chr11:107,403,404-107,404,589	-0.50

lincRNA: long intergenic non-coding RNA; EFSL event-free survival.

Table S3. Outcome of younger adult patients with cytogenetically normal acute myeloid leukemia in the training set according to long non-coding RNA (lncRNA) score status.

End point	Favorable lncRNA score (n=132)	Unfavorable lncRNA score (n=131)	*P*	OR/HR (95% CI)
Disease-free survival			<0.001	
Median, years	5.4	0.7		0.35
Disease-free at 5 years,% (95% CI)	51 (42-60)	16 (9-24)		(0.25-0.48)
Overall survival			<0.001	
Median, years	NR	1.1		0.30
Alive at 5 years, % (95% CI)	61 (52-69)	22 (15-29)		(0.22-0.42)
Event-free survival			<0.001	
Median, years	5.1	0.5		0.28
Event-free at 5 years,% (95% CI)	50 (41-58)	11 (6-17)		(0.21-0.38)

OR: odds ratio; HR: hazard ratio; CI: confidence interval; NR: not reached.

Table S4. Comparison of clinical and molecular characteristics by favorable and unfavorable long non-coding RNA (lncRNA) score in the training set of younger adult patients with cytogenetically normal acute myeloid leukemia.

Characteristic	Favorable lncRNA score (n=132)	Unfavorable lncRNA score (n=131)	*P*
Age, years			0.37
Median	45	48	
Range	17-59	18-59	
Sex, n. (%)			0.06
Male	76 (58)	60 (46)	
Female	56 (42)	71 (54)	
Race, n. (%)			0.67
White	120 (92)	114 (90)	
Non-white	11 (8)	13 (10)	
Hemoglobin (g/dL)			0.02
Median	9.5	9.1	
Range	4.9-13.5	4.6-14.4	
Platelet count. ($x10^9$/L)			0.69
Median	56	63	
Range	8-347	8-445	
WBC count ($x10^9$/L)			<0.001
Median	20.3	42.7	
Range	0.6-223.8	0.9-308.8	
Blood blasts, %			0.78
Median	63	63	
Range	0-97	0-97	
Bone marrow blasts, %			.02
Median	65	72	
Range	10-93	19-96	
Extramedullary involvement, n. (%)	33 (26)	44 (34)	0.17
Autologous HCT in 1^{st} CR, n. (%)	65 (51)	36 (41)	0.16
NPM1, n. (%)			.38
Mutated	71 (54)	78 (60)	
Wild-type	61 (46)	53 (40)	
FLT3-ITD, n. (%)			<0.001
Present	31 (24)	68 (53)	
Absent	99 (76)	61 (47)	
CEBPA, n. (%)			<0.001
Double Mutated	37 (29)	3 (2)	
Wild-type	92 (71)	122 (98)	
FLT3-TKD, n. (%)			1.00
Present	17 (13)	16 (13)	
Absent	114 (87)	111 (87)	
WT1, n. (%)			0.31
Mutated	11 (8)	16 (13)	
Wild-type	120 (92)	109 (87)	

Characteristic	Favorable lncRNA score (n=132)	Unfavorable lncRNA score (n=131)	*P*
TET2, n. (%)			0.71
Mutated	15 (11)	17 (14)	
Wild-type	116 (89)	108 (86)	
IDH1, n. (%)			0.51
Mutated	13 (10)	9 (7)	
Wild-type	118 (90)	117 (93)	
IDH2, n. (%)			1.00
Mutated	11 (8)	11 (9)	
R140	9	9	
R172	2	2	
Wild-type	120 (92)	115 (91)	
ASXL1, n. (%)			0.20
Mutated	3 (2)	7 (6)	
Wild-type	128 (98)	114 (94)	
DNMT3A, n. (%)			0.01
Mutated	42 (32)	60 (48)	
R882	32	40	
Non-R882	10	20	
Wild-type	89 (68)	65 (52)	
RUNX1, n. (%)			0.009
Mutated	3 (2)	13 (10)	
Wild-type	128 (98)	112 (90)	
ELN Genetic Group,* n. (%)			<0.001
Favorable	94 (73)	41 (34)	
Intermediate	21 (16)	45 (38)	
Adverse	14 (11)	34 (28)	
ERG expression group,† n. (%)			0.32
High	67 (51)	75 (58)	
Low	64 (49)	55 (42)	
BAALC expression group,† n. (%)			0.12
High	59 (50)	74 (61)	
Low	59 (50)	48 (39)	
MN1 expression group,† n. (%)			0.61
High	63 (51)	70 (55)	
Low	60 (49)	57 (45)	
miR-181a expression group,† n. (%)			<0.001
High	68 (65)	41 (40)	
Low	37 (35)	61 (60)	
miR-3151, n. (%)			0.86
Expressed	19 (18)	20 (20)	
Not Expressed	86 (82)	82 (80)	
miR-155 expression group,† n. (%)			0.03
High	45 (43)	60 (59)	
Low	60 (57)	42 (41)	

WBC: white blood cell; HCT: hematopoietic cell transplant; CR: complete remission; ELN: European LeukemiaNet; *FLT3*-ITD: internal tandem duplication of the *FLT3* gene; *FLT3*-TKD: tyrosine kinase domain mutation in the *FLT3* gene.

* Among patients with cytogenetically normal acute myeloid leukemia (CN-AML), the ELN Favorable Risk Category comprises patients with double-mutated *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN Intermediate Risk Category includes patients with wild-type *NPM1* without *FLT3*-ITD or wild-type *NPM1* and *FLT3*-ITD^{low} or mutated *NPM1* and *FLT3*-ITD^{high}. The ELN Adverse Risk Category comprises patients with wild-type *NPM1* and *FLT3*-ITD^{high}, and/or mutated *RUNX1* (if it does not co-occur with a Favorable AML subtype) and/or mutated *ASXL1* (if it does not co-occur with a Favorable AML subtype) and/or mutated *TP53*. *FLT3*-ITD^{low} is defined by a *FLT3*-ITD/*FLT3* wild-type allelic ratio of less than 0.5 and *FLT3*-ITD^{high} is defined as by a *FLT3*-ITD/*FLT3* wild-type allelic ratio of equal to or more than 0.5.[12]

† The median expression value was used as a cut point.

Table S5. Multivariable analyses for outcome in the training set of younger adults patients with cytogenetically normal acute myeloid leukemia.

Variables in final models	Disease-free survival		Overall survival		Event-free survival	
	HR (95% CI)	*P*	HR (95% CI)	*P*	HR (95% CI)	*P*
lncRNA score, favorable *versus* unfavorable	0.34 (0.24-0.48)	<0.001	0.34 (0.24-0.48)	<0.001	0.29 (0.21-0.4)	<0.001
FLT3-TKD, present *versus* absent	0.43 (0.23-0.8)	0.007	-	-	-	.-
FLT3-ITD, present *versus* absent	-	-	2.43 (1.74-3.93)	<0.001	1.94 (1.41-2.65)	<0.001
MN1, high *versus* low*	-	-	1.92 (1.37-2.69)	<0.001	-	-
Age, each 10-year increase	-	-	1.28 (1.10-1.49)	0.001	-	-
NPM1, mutated *versus* wild-type	-	-	-	-	0.53 (0.39-0.72)	<0.001

HR: hazard ratio; CI: confidence interval; lncRNA: long non-coding RNA; *FLT3*-TKD: tyrosine kinase domain mutations of the *FLT3* gene; *FLT3*-ITD: internal tandem duplication of the *FLT3* gene.
NOTE: Hazard ratios greater than (or less than) 1.0 indicate higher (or lower) risk for relapse or death (disease-free survival), for death (overall survival) or for failure to achieve complete remission, relapse or death (event-free survival) for the higher value of the continuous variables and the first category listed for the categorical variables. Variables considered for model inclusion were: lncRNA score status (favorable *versus* unfavorable), age (as a continuous variable, in 10-year increments), sex (male *versus* female), race (white *versus* non-white), white blood cell count [(WBC) as a continuous variable, in 50-unit increments], hemoglobin (as a continuous variable, in 1-unit increments), platelet count (as a continuous variable, in 50-unit increments), extramedullary involvement (present *versus* absent), *ASXL1* mutations (mutated *versus* wild-type), *CEBPA* mutations (double-mutated *versus* single-mutated or wild-type), *DNMT3A* mutations (mutated *versus* wild-type), *FLT3*-ITD (present *versus* absent), *FLT3*-TKD (present *versus* absent), *IDH1* mutations (mutated *versus* wild-type), *IDH2* mutations (mutated *versus* wild-type), *NPM1* mutations (mutated *versus* wild-type), *RUNX1* mutations (mutated *versus* wild-type), *TET2* mutations (mutated *versus* wild-type), *WT1* mutations (mutated *versus* wild-type), *BAALC* expression levels (high *versus* low), *ERG* expression levels (high *versus* low), *MN1*

expression levels (high *versus* low), miR-181a expression levels (high *versus* low), miR-3151 (expressed *versus* not expressed), and miR-155 expression levels (high *versus* low).
* The median expression value was used as the cut point.

Table S6. Outcome of younger adult patients with cytogenetically normal acute myeloid leukemia in the validation set of the studied cohort according to long non-coding RNA (lncRNA) score status.

End point	Favorable lncRNA score n=57	Unfavorable lncRNA score n=57	*P*	OR/HR (95% CI)
Disease-free survival			<0.001	
Median (years)	7.0	0.7		0.39
Disease-free at 5 years,% (95% CI)	51 (37-64)	17 (8-29)		(0.24-0.65)
Overall survival			0.002	
Median, (years)	7.1	1.5		0.48
Alive at 5 years, % (95% CI)	52 (38-64)	26 (16-38)		(0.30-0.78)
Event-free survival			<0.001	
Median (years)	2.6	0.8		0.45
Event-free at 5 years,% (95% CI)	46 (32-58)	16 (8-26)		(0.29-0.71)

OR: odds ratio; HR: hazard ratio; CI: confidence interval.

Table S7-S9. Signatures of long non-coding RNA (lncRNA) genes that significantly associated with double *CEBPA* mutations, *NPM1* mutations or presence of *FLT3*-ITD in the training set of younger adult patients with cytogenetically normal acute myeloid leukemia.

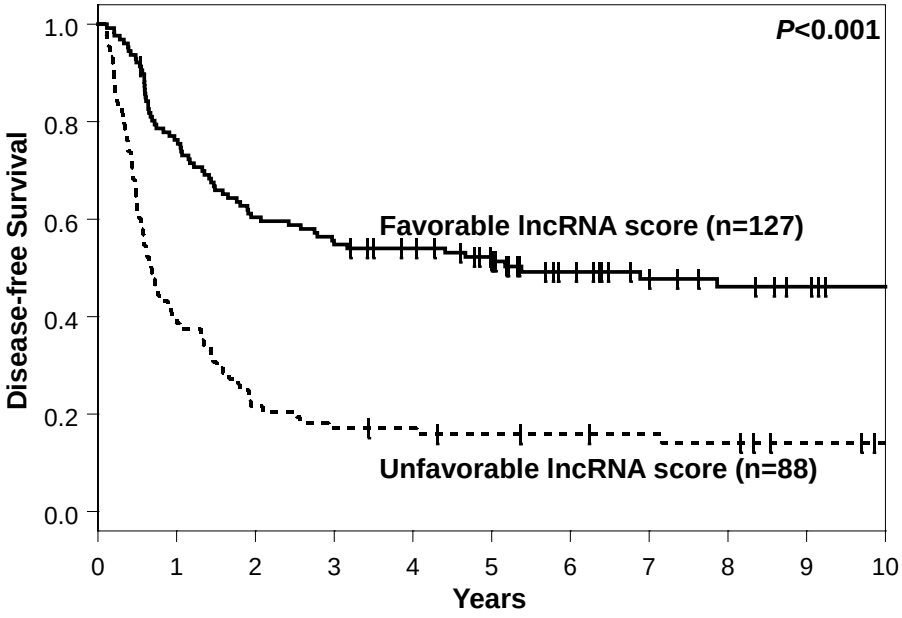
Table S10. Signature of 410 genes that significantly correlated with long non-coding RNA (lncRNA) scores in younger adult patients with cytogenetically normal acute myeloid leukemia.

Table S11. Gene Ontology (GO) functional groups that significantly correlated with unfavorable long non-coding RNA (lncRNA) scores in younger adult patients with cytogenetically normal acute myeloid leukemia.

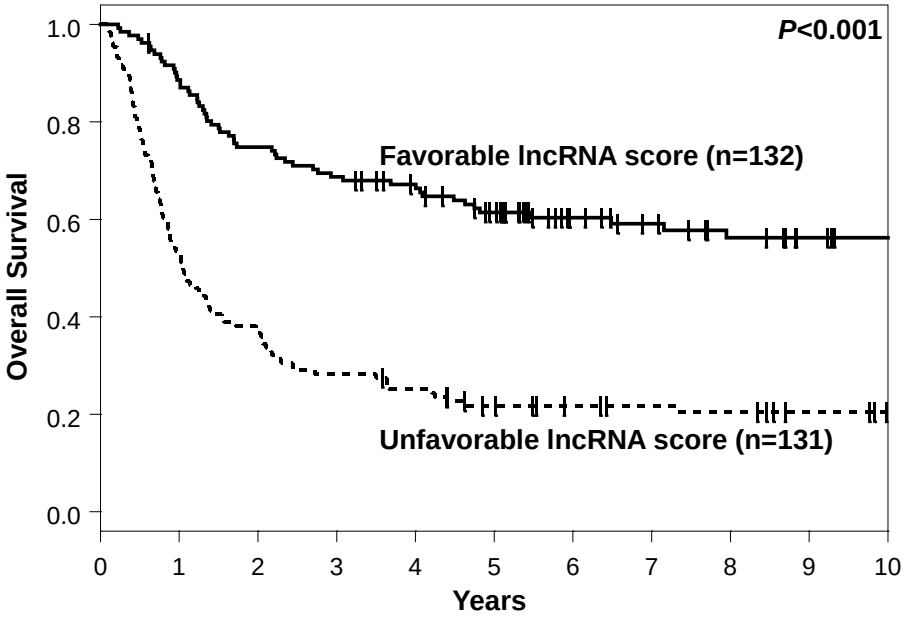
Table S12. Signature of 14 microRNAs (miRs) that significantly correlated with long non-coding RNA (lncRNA) scores in younger adult patients with cytogenetically normal acute myeloid leukemia.

Figure S1. Outcome of younger adult patients with cytogenetically normal acute myeloid leukemia with favorable and unfavorable long noncoding RNA (lncRNA) scores in the training set. (A) Disease-free survival, (B) overall survival and (C) event-free survival. The lncRNA score of each individual patient was computed as a weighted score of 24 prognostic lncRNAs.

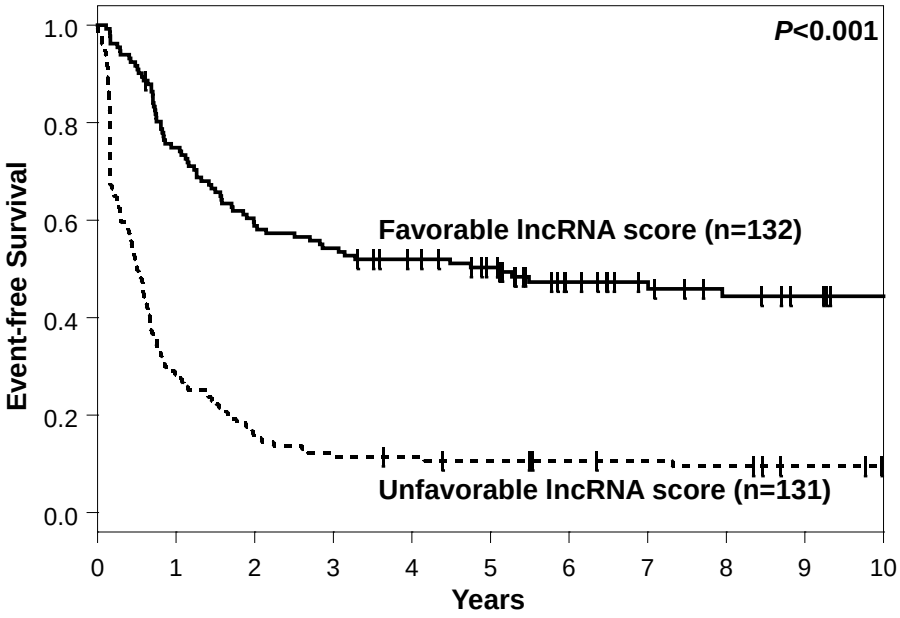
A



B


P<0.001
Favorable lncRNA score (n=132)
Unfavorable lncRNA score (n=131)
Overall Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8
9
10
Years

C


P<0.001
Favorable lncRNA score (n=132)
Unfavorable lncRNA score (n=131)
Event-free Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8
9
10
Years