ARTICLE

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

EUROPEAN HEMATOLOGY ASSOCIATION

Ferrata Storti Foundation

Haematologica 2017
Volume 102(8):1391-1400

ABSTRACT

Long non-coding ribonucleic acids (RNAs) are a novel class of RNA molecules, which are increasingly recognized as important molecular players in solid and hematologic malignancies. Herein we investigated whether long non-coding RNA expression is associated with clinical and molecular features, as well as outcome of younger adults (aged <60 years) with *de novo* cytogenetically normal acute myeloid leukemia. Whole transcriptome profiling was performed in a training (n=263) and a validation set (n=114). Using the training set, we identified 24 long non-coding RNAs associated with event-free survival. Linear combination of the weighted expression values of these transcripts yielded a prognostic score. In the validation set, patients with high scores had shorter disease-free (P<0.001), overall (P=0.002) and event-free survival (P<0.001) than patients with low scores. In multivariable analyses, long non-coding RNA score status was an independent prognostic marker for disease-free (P=0.01) and event-free survival (P=0.002), and showed a trend for overall survival (P=0.06). Among multiple molecular alterations tested, which are prognostic in cytogenetically normal acute myeloid leukemia, only double *CEBPA* mutations, *NPM1* mutations and *FLT3*-ITD associated with distinct long non-coding RNA signatures. Correlation of the long non-coding RNA scores with messenger RNA and microRNA expression identified enrichment of genes involved in lymphocyte/leukocyte activation, inflammation and apoptosis in patients with high scores. We conclude that long non-coding RNA profiling provides meaningful prognostic information in younger adults with cytogenetically normal acute myeloid leukemia. In addition, expression of prognostic long non-coding RNAs associates with oncogenic molecular pathways in this disease. *clinicaltrials.gov Identifier: 00048958 (CALGB-8461), 00899223 (CALGB-9665), and 00900224 (CALGB-20202).*

Introduction

Acute myeloid leukemia (AML) is a highly heterogeneous disease with regard to genetic abnormalities and clinical course.[1] The prognosis of adult AML is generally

Correspondence:

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

Received: February 7, 2017.

Accepted: May 2, 2017.

Pre-published: May 4, 2017.

doi:10.3324/haematol.2017.166215

Check the online version for the most updated information on this article, online supplements, and information on authorship & disclosures: www.haematologica.org/content/102/8/1391

poor. Only 40% of younger adult (aged <60 years) and 10% of older (aged ≥60 years) AML patients achieve long-term survival.[1] Currently, chromosomal aberrations[2-4] and recurrent gene mutations[5-8] are considered the most reliable and reproducible prognostic markers in AML, and are used in the clinic to identify patients at high risk of death and to guide treatment. Aberrant levels of messenger RNA (mRNA)[9-11] and microRNA (miR) transcripts[12,13] also have prognostic significance, and efforts have been made to incorporate gene-expression profiling into prognostic algorithms.[14-16]

Long non-coding RNAs (lncRNAs) are a novel class of RNA molecules that are longer than 200 nucleotides, have no protein coding potential and are either located within the intergenic stretches of the genome or overlap (in sense or antisense direction) protein coding genes.[17,18] These transcripts regulate key cellular functions, such as chromosome dosage compensation,[19] imprinting,[20] cell cycle progression,[21] and differentiation.[22] In cancer, individual lncRNAs have been shown to play an important role in malignant transformation.[23-25] Despite the growing understanding of the biologic significance of deregulated lncRNA expression in malignant diseases, the value of these molecules as potential biomarkers in the clinical setting has not been extensively studied.[26,27] With regard to cytogenetically normal AML (CN-AML), the prognostic and biologic significance of lncRNAs in younger adult patients remains unknown. Therefore, we analyzed, using whole transcriptome sequencing (RNA-seq), the lncRNA profiles of younger adults with *de novo* CN-AML, who were comprehensively characterized with regard to molecular abnormalities and outcome. Herein, we show that lncRNA profiling provides independent prognostic information in these patients. We also show that expression levels of prognostic lncRNAs correlate with distinct mRNA and miR signatures, and provide insights into the leukemogenic pathways that these lncRNAs potentially regulate.

Methods

Patients and treatment

Pretreatment bone marrow (BM) or blood samples were obtained from a training (n=263) and a validation set (n=114) of younger adult patients (aged 17-59 years) with *de novo* CN-AML, who received intensive cytarabine/anthracycline-based first-line therapy on Cancer and Leukemia Group B (CALGB)/Alliance for Clinical Trials in Oncology (Alliance) trials and were alive 30 days after initiation of treatment. Per protocol, no patient received allogeneic stem cell transplantation in first complete remission (CR). Details regarding treatment protocols are provided in the *Online Supplementary Appendix*. All patients provided written informed consent, and all study protocols were in accordance with the Declaration of Helsinki and approved by institutional review boards at each center.

Cytogenetic and molecular analyses

Cytogenetic analyses were performed in CALGB/Alliance-approved institutional laboratories and results confirmed by central karyotype review.[28] The diagnosis of normal karyotype was based on at least 20 metaphase cells analyzed in BM specimens subjected to short-term (24- or 48-hour) unstimulated cultures.

Targeted amplicon sequencing using the MiSeq platform (Illumina) was used to analyze DNA samples for presence of gene mutations that have been reported to associate with clinical outcome of CN-AML patients (i.e., mutations in the *ASXL1*, *DNMT3A* [R882 and non-R882], *IDH1, IDH2* [R140 and R172], *NPM1, RUNX1, TET2* or *WT1* genes, and *FLT3*-tyrosine kinase domain [*FLT3*-TKD] mutations), as described previously.[26,29] A variant allele frequency of ≥10% was used as the cutoff to distinguish between mutated *versus* wild-type alleles of these genes. The presence of mutations in the *CEBPA* gene and *FLT3*-internal tandem duplications (*FLT3*-ITD) were evaluated using Sanger sequencing[30] and fragment analysis,[31] as described previously. Since only double *CEBPA* mutations are favorable prognostic markers in CN-AML,[32] we considered only this genotype as mutated.

Transcriptome analyses

RNA samples of all studied patients (n=377) were analyzed with total RNA sequencing (after depletion of ribosomal and mitochondrial RNA) using the Illumina HiSeq 2500 platform. Due to RNA quality restrictions, a subset of 300 patients could be additionally analyzed with small RNA sequencing, for profiling of miR expression. Further details are provided in the *Online Supplementary Appendix*. To determine the expression status of patients (i.e., high *versus* low expressers) with regard to prognostic expression markers (e.g., *BAALC*), the median values of normalized RNA sequencing reads were used as the cutoff.

Statistical analyses

Clinical endpoint definitions are given in the *Online Supplementary Appendix*. Baseline demographic, clinical, and molecular features were compared between patients with low and those with high lncRNA scores (later on referred to as favorable and unfavorable, see below), and between the training and validation sets using the Wilcoxon rank-sum and Fisher's exact tests for continuous and categorical variables, respectively.[33] The estimated probabilities of disease-free (DFS), overall (OS) and event-free survival (EFS) were calculated using the Kaplan–Meier method, and the log-rank test evaluated differences between survival distributions.[34] Cox proportional hazard models were used to calculate hazard ratios (HR) for DFS, OS and EFS.[33] Multivariable proportional hazards models were constructed using a backward selection procedure. All statistical analyses were performed by The Alliance Statistics and Data Center.

Results

Global expression of lncRNAs

To investigate the role of lncRNAs in AML, we first identified all known lncRNAs which were present in the transcriptomes of the younger CN-AML patients who were studied (n=377). After exclusion of contaminating ribosomal RNA molecules, we identified 22,166 non-coding RNA transcripts. According to the GENCODE v22 database,[35] 23% of these transcripts were categorized as processed pseudogenes, 21% as intergenic/intervening lncRNAs, 21% as antisense lncRNAs, 4% as sense intronic/overlapping lncRNAs and 31% were classified as other transcripts (e.g., as unitary pseudogenes, unprocessed pseudogenes etc.; Figure 1).

Generation of a prognostic lncRNA score in the training set

To assess the prognostic significance of lncRNA expression in younger adults with CN-AML, we performed exploratory analysis in a training set (n=263) of younger CN-AML patients and used a separate patient cohort to

validate our findings (validation set, n=114). Comparison of clinical and molecular characteristics at diagnosis between the training and validation sets showed that they were relatively similar, with the exceptions that patients in the training set had higher percentages of blood blasts (P=0.03), were more frequently *FLT3*-TKD-positive (P=0.02), and had higher *ERG* (P=0.01) and *BAALC* (P=0.002) expression levels (*Online Supplementary Table S1*).

We first identified all lncRNAs that were highly associated with EFS ($P<10^{-6}$) in the training set by univariable

Table 1. Comparison of clinical and molecular characteristics by favorable and unfavorable long non-coding RNA (lncRNA) score in the validation set of younger adults with cytogenetically normal acute myeloid leukemia.

Characteristic	Favorable lncRNA Score (n=57)	Unfavorable lncRNA Score (n=57)	P
Age, years			0.44
Median	44	47	
Range	18-59	18-59	
Sex, n. (%)			1.00
Male	28 (49)	29 (51)	
Female	29 (51)	28 (49)	
Race, n. (%)			1.00
White	51 (91)	50 (89)	
Non-white	5 (9)	6 (11)	
Hemoglobin (g/dL)			0.66
Median	9.1	8.8	
Range	4.2-25.1	4.8-13.4	
Platelet count (x10^9/L)			0.49
Median	52	55	
Range	10-271	8-433	
WBC count (x10^9/L)			0.009
Median	24.9	45.7	
Range	0.9-475.0	2.2-295.0	
Blood blasts, %			0.06
Median	45	63	
Range	0-90	0-97	
Bone marrow blasts, %			0.25
Median	63	68	
Range	21-91	18-95	
Extramedullary involvement, n. (%)	15 (28)	18 (32)	0.68
Autologous HCT in 1st CR, n. (%)	33 (65)	23 (48)	0.11
NPM1, n. (%)			1.00
Mutated	37 (65)	37 (65)	
Wild-type	20 (35)	20 (35)	
FLT3-ITD, n. (%)			0.007
Present	15 (27)	30 (54)	
Absent	40 (73)	26 (46)	
CEBPA, n. (%)			0.78
Double Mutated	8 (15)	6 (12)	
Wild-type	46 (85)	45 (88)	
FLT3-TKD, n. (%)			0.36
Present	4 (7)	1 (2)	
Absent	51 (93)	54 (98)	
WT1, n. (%)			0.09
Mutated	4 (7)	10 (19)	
Wild-type	52 (93)	44 (81)	
TET2, n. (%)			0.49
Mutated	6 (11)	3 (6)	
Wild-type	50 (89)	51 (94)	
IDH1, n. (%)			1.00
Mutated	4 (7)	3 (5)	
Wild-type	52 (93)	52 (95)	
IDH2, n. (%)			1.00
Mutated	7 (13)	6 (11)	
R140	4	6	
R172	3	0	
Wild-type	49 (88)	49 (89)	
ASXL1, n. (%)			1.00
Mutated	2 (4)	1 (2)	
Wild-type	54 (96)	51 (98)	

continued in the next page

continued in the previous page

Characteristic	Favorable lncRNA Score (n=57)	Unfavorable lncRNA Score (n=57)	P
DNMT3A, n. (%)			0.56
Mutated	20 (36)	23 (43)	
R882	14	20	
Non-R882	6	3	
Wild-type	36 (64)	31 (57)	
RUNX1, n. (%)			1.00
Mutated	3 (5)	2 (4)	
Wild-type	53 (95)	52 (96)	
ELN Risk Category,* n. (%)			0.02
Favorable	37 (71)	23 (43)	
Intermediate	11 (21)	20 (38)	
Adverse	4 (8)	10 (19)	
ERG expression group,† n. (%)			0.85
High	22 (39)	23 (41)	
Low	35 (61)	33 (59)	
BAALC expression group,† n. (%)			0.84
High	19 (36)	21 (39)	
Low	34 (64)	33 (61)	
MN1 expression group,† n. (%)			0.06
High	18 (33)	29 (52)	
Low	37 (67)	27 (48)	
miR-181a expression group,† n. (%)			0.41
High	24 (50)	18 (40)	
Low	24 (50)	27 (60)	
miR-3151, n. (%)			0.36
Expressed	8 (17)	4 (9)	
Not expressed	40 (83)	41 (91)	
miR-155 expression group,† n. (%)			<0.001
High	16 (33)	31 (69)	
Low	32 (67)	14 (31)	

*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 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 by a *FLT3*-ITD/*FLT3* wild-type allelic ratio of equal to or more than 0.5.[1] †The median expression value was used as the cut point. 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; lncRNA: long non-coding ribonucleic acid; miR: microRNA.

Cox analysis (Figure 2). EFS was used because it comprehensively evaluates the lncRNAs that are associated with response to chemotherapy, probability of relapse and probability of survival. We detected 24 lncRNAs associated with EFS ($P<10^{-6}$; *Online Supplementary Table S2*). Next, we derived a prognostic lncRNA score by linear combination of the weighted expression values of these 24 lncRNAs. The median value of the lncRNA score was used to dichotomize the training set of patients. Patients with low lncRNA scores (n=132) had longer DFS ($P<0.001$), OS ($P<0.001$) and EFS ($P<0.001$) than patients with high lncRNA scores (n=131). We therefore classified low lncRNA scores as "favorable" and high as "unfavorable" (*Online Supplementary Table S3* and *Online Supplementary Figure S1*).

Association of lncRNA score with patient characteristics and clinical outcome in the training set

With regard to clinical and molecular characteristics, patients with favorable lncRNA scores in the training set were more likely to present with higher hemoglobin levels ($P=0.02$), lower white blood cell (WBC) counts ($P<0.001$), and lower percentages of BM blasts ($P=0.02$). They were also less likely to harbor *FLT3*-ITD ($P<0.001$), *DNMT3A* ($P=0.01$) and *RUNX1* ($P=0.009$) mutations and more likely to harbor double *CEBPA* mutations ($P<0.001$). Patients with favorable lncRNA scores in the training set differed with regard to their distribution in the Risk Categories of the European LeukemiaNet (ELN) classification of AML,[1] when compared with patients with unfavorable lncRNA scores ($P<0.001$); patients with favorable lncRNA scores were more frequently classified as favorable and less frequently as intermediate or adverse risk than those with unfavorable lncRNA scores (*Online Supplementary Table S4*). Favorable lncRNA score status also associated with high expression of miR-181a ($P<0.001$) and low expression of miR-155 ($P=0.03$, *Online Supplementary Table S4*). Association of a favorable lncRNA score with longer DFS, OS and EFS remained significant in multivariable analyses ($P<0.001$ for all 3 end points, *Online Supplementary Table S5*), after adjusting for other co-variates.

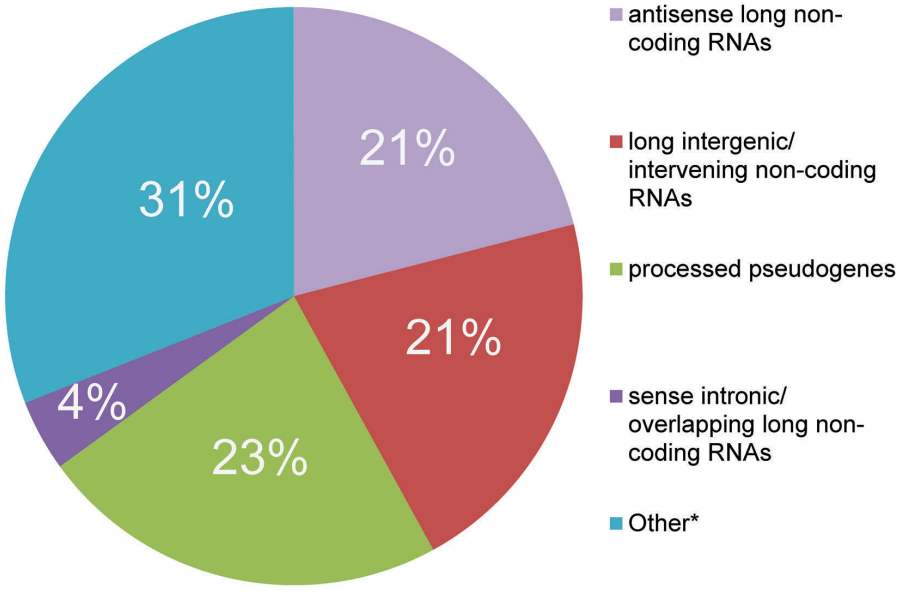


Figure 1. Distribution of the 22,166 detected non-coding RNA transcripts among different classes of non-coding RNA molecules. Annotation of transcripts was performed according to the GENCODE v22 database. lncRNA indicates long non-coding RNA and lincRNA denotes long intergenic/intervening non-coding RNA. *Other refers to: microRNAs, miscellaneous non-coding RNAs, unprocessed pseudogenes, small RNAs, translated unprocessed pseudogenes, processed transcripts, small nucleolar RNAs, transcribed processed pseudogenes, T-cell receptor pseudogenes, immunoglobulin genes, immunoglobulin pseudogenes, unitary pseudogenes, small cajal body specific RNAs, polymorphic pseudogenes, 3-prime overlapping non-coding RNAs, transcribed unitary pseudogenes and macro lncRNAs. lncRNA: long non-coding ribonucleic acid.

Association of lncRNA score with patient characteristics and clinical outcome in the validation set

We used the median value of the lncRNA score, as calculated in the training set to divide the validation set into favorable and unfavorable lncRNA score groups (Figure 2). Patients with favorable lncRNA scores (n=57) were less likely to present with higher WBC counts at the time of diagnosis (P=0.009) or to harbor *FLT3*-ITD (P=0.007). lncRNA score status also associated with significantly different distribution of the patients in the Risk Categories of the ELN guidelines (P=0.02).[1] Patients with favorable lncRNA scores were more likely to belong to the favorable and less likely to belong to the intermediate or adverse risk category. Patients with favorable lncRNA scores in the validation set were less likely to be miR-155 high-expressers (P<0.001) than patients with unfavorable lncRNA scores (n=57; Table 1).

Patients with favorable lncRNA scores had longer DFS than those with unfavorable lncRNA scores (P<0.001; Figure 3A). Five years after diagnosis, 51% of patients with favorable lncRNA scores remained alive and leukemia-free, in contrast to only 17% of those with unfavorable lncRNA scores. Favorable lncRNA score status also associated with longer OS (P=0.002, 5-year rates, 52% *versus* 26%; Figure 3B) and longer EFS (P<0.001, 5-year rates, 46% *versus* 16%; Figure 3C, *Online Supplementary Table S6*). The prognostic value of the lncRNA score in the validation set remained significant when it was analyzed as a continuous variable. Increasingly favorable lncRNA scores associated with longer DFS (P<0.001), OS (P=0.007) and EFS (P=0.002).

In multivariable analyses, favorable lncRNA score status was an independent marker for longer DFS (P=0.01), after adjusting for miR-155 expression status, and EFS (P=0.002), after adjusting for the presence of *FLT3*-ITD (Table 2). With regard to OS, patients with a favorable lncRNA score had a trend for longer survival (P=0.06), after adjustment for *FLT3*-ITD and *MN1* expression status.

Associations of recurrent gene mutations with lncRNA expression

We evaluated if recurrent prognostic gene mutations in CN-AML associated with distinct expression patterns of lncRNAs in younger adults with CN-AML. For this purpose, mutation-related lncRNA signatures were derived in the training set using stringent criteria (for details see Methods and the *Online Supplementary Appendix*).

Double-mutated *CEBPA* showed the strongest association with lncRNA expression; 82 lncRNAs were upregulated and 186 lncRNAs were downregulated in patients who harbored double-mutated *CEBPA* (Figure 4A, *Online Supplementary Table S7*). Among the *CEBPA* mutation-related lncRNAs, *NEAT1* was significantly underexpressed in the group of patients with *CEBPA* mutations. This lncRNA has been involved in myeloid differentiation of acute promyelocytic leukemia cells after all-trans retinoic acid treatment.[36]

Mutations in the *NPM1* gene also strongly associated with a lncRNA signature, which comprised 35 transcripts upregulated and 37 transcripts downregulated in patients harboring *NPM1* mutations (Figure 4B, *Online Supplementary Table S8*). Thirty-three of the 35 lncRNAs overexpressed in patients with *NPM1* mutations, were downregulated in patients with *CEBPA* mutations. This finding is consistent with the observation that double *CEBPA* and *NPM1* mutations rarely co-occur in CN-AML. *NPM1* mutations were positively associated with lncRNAs embedded within the *HOX* gene loci (*HOXA-AS3*, *HOXB-AS3*) and other lncRNAs implicated in myelopoiesis (*EGOT*[37]) or carcinogenesis (e.g., *PCAT18*[38] and *LUCAT1*[39]).

The *FLT3*-ITD-related lncRNA signature consisted of 26 transcripts, 19 of which were upregulated and 7 downregulated in patients with this mutation (Figure 4C, *Online Supplementary Table S9*). The host gene of miR-155 (*MIR155HG*) was among the lncRNAs overexpressed in *FLT3*-ITD-positive patients. High *MIR155HG* expression independently associates with poor outcome in CN-

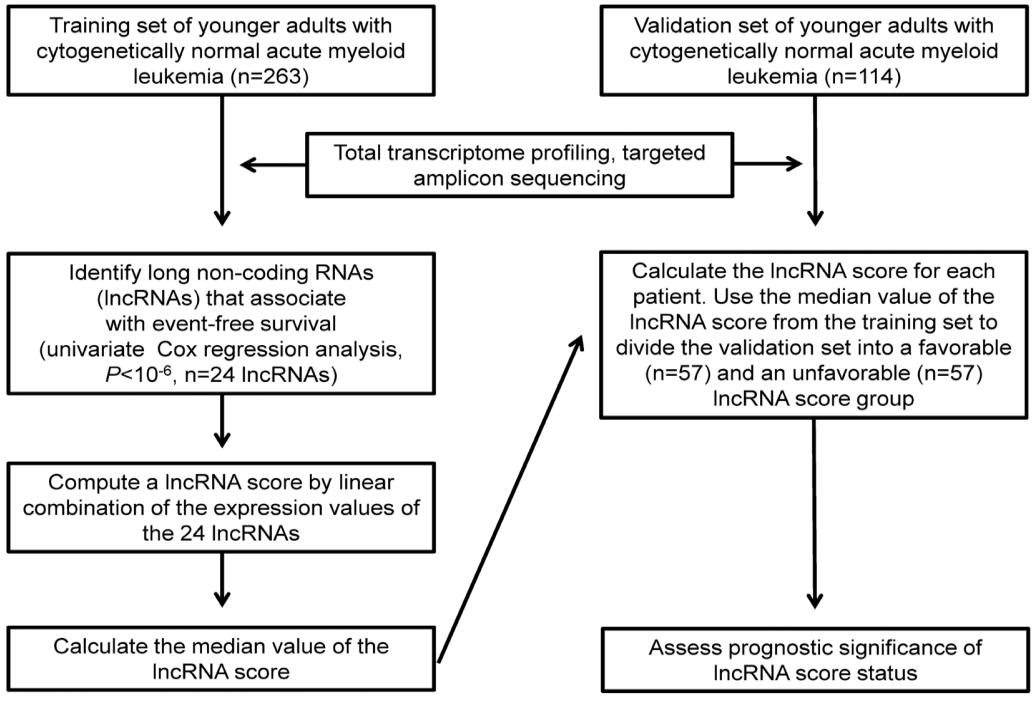


Figure 2. Overview of the study design.

AML.[40] The *WT1-AS* lncRNA was also highly expressed among *FLT3*-ITD-positive patients; it has been reported to post-translationally regulate the protein levels of WT1.[41]

To assess the capacity of gene mutation-related lncRNA signatures to detect their corresponding molecular alterations in CN-AML patients, we applied these signatures to the validation set. The mutated *CEBPA*-related signature showed the highest level of accuracy (specificity and sensitivity of mutated *CEBPA* detection: ≥93% and ≥98%, respectively), followed by the mutated *NPM1*-related (sensitivity: ≥80%, specificity ≥73%) and the *FLT3*-ITD-related signatures (sensitivity ≥70%, specificity: ≥76%). The remaining prognostic gene mutations that were tested either did not associate with differential expression of lncRNAs (i.e., *TET2* mutations) or generated signatures that failed to reliably detect the mutational status of patients in the validation set (e.g., *DNMT3A*, *WT1* mutations).

Biologic implications of the lncRNA score

To gain biologic insights into the molecular pathways that may be affected by differences in the lncRNA score, we examined the correlation between the lncRNA score and the mRNA/miRNA expression in 300 younger CN-AML patients who had available mRNA and miRNA profiling data.

We identified 410 mRNA transcripts whose expression levels correlated with the lncRNA score, 172 of which correlated positively and 238 negatively with unfavorable lncRNA scores (Figure 5A, *Online Supplementary Table S10*). Among highly expressed genes in patients with unfavorable lncRNA scores, putative oncogenes and key mediators of the oncogenic AP-1 pathway such as *ATF3, FOS, FOSB, JUN*, and *MAFF* were identified. With regard to hematopoiesis, the AP-1 pathway has been shown to regulate proliferation of erythroleukemia cells,[42] to mediate monocyte/macrophage differentiation of myeloid cells[43] and to co-regulate miR-155 expression in stimulated macrophages.[44] Genes that regulate immune responses (e.g., *IL1B, IRF7, CD80*) and genes that mediate immune evasion (e.g., *IER3, LILRB4*) were also highly expressed in patients with unfavorable lncRNA profiles. Finally, oncogenes promoting proliferation of malignant cells (e.g., *RET, ETS2, PLK2, NEK6, PLK3* and *SRC)* were found to be overexpressed in patients with unfavorable lncRNA scores. Gene ontology analysis revealed that genes involved in lymphocyte/leukocyte activation, inflammation, response to wounding and regulation of apoptosis were enriched in the subset of patients with unfavorable lncRNA scores (Figure 5B, *Online Supplementary Table S11*).

Among mRNA molecules downregulated in patients with unfavorable lncRNA scores, we detected transcripts with reported tumor-suppressive function (*APC, JADE1, BRMS1L,* and *ING3*). Gene ontology analysis showed that genes that participate in the regulation of transcription, the regulation of RNA metabolic processes and DNA binding were underexpressed in the group of patients with unfavorable lncRNA scores (Figure 5C, *Online Supplementary Table S11*).

With regard to miR expression, 10 miRs were found to correlate positively (miR-660, miR-502, miR-532-5p, miR-155, miR-500a-3p, miR-500a-5p, miR-532-3p, miR-362, miR-339 and miR-23a) and 4 miRs to correlate negatively (miR-192, miR-625, miR-100 and miR-194) with unfavorable lncRNA scores (*Online Supplementary Table S12*). Among the 10 miRs that positively correlated with unfavorable lncRNA scores, 7 were located in close proximity on chromosome X; miR-660, miR-502, miR-532-5p, miR-500a-3p, miR-500a-5p, miR-532-3p and miR-362 are all imbedded in intron 3 of the *CLCN5* gene. This miR cluster mediates an anti-apoptotic effect in chronic lymphocytic leukemia cells.[45] miR-155, which also positively correlated

with unfavorable lncRNA scores, is an established adverse prognosticator in CN-AML[40] and has been implicated in leukemogenesis of *FLT3*-ITD-positive AML.[46]

Discussion

Over the past 5 years, lncRNAs have emerged as new players in cancer biology and biomarker discovery.[47] Our group has previously reported that distinctive lncRNA signatures are associated with prognostic gene mutations in older CN-AML patients, and that expression levels of a small group of lncRNAs have prognostic significance in these patients.[26] Since CN-AML in younger adults differs with regard to clinical features, associated molecular abnormalities and outcome from that in older patients, we investigated the prognostic value and biologic implications of lncRNA expression in a total of 377 CN-AML adult patients younger than 60 years.

First, we identified 24 lncRNAs highly correlated with EFS. Similarly to the previously reported older CN-AML patients,[26] only a small number of these prognostic lncRNAs associated with prognostic gene mutations: *MIR155HG* was upregulated in patients who harbor *FLT3*-ITD, *AC006129.2* was upregulated in patients with double *CEBPA* mutations, whereas *AL122127.25*, *RP11-946L16.2*,

Table 2. Multivariable analyses for outcome in the validation set of younger adults with cytogenetically normal acute myeloid leukemia.

Variables in final models	Disease-free survival HR (95% CI)	*P*	Overall survival HR (95% CI)	*P*	Event-free survival HR (95% CI)	*P*
lncRNA score, favorable *versus* unfavorable	0.46 (0.26-0.83)	0.01	0.6 (0.35-1.03)	0.06	0.48 (0.30-0.77)	0.002
miR-155, high *versus* low*	1.81 (1.01-3.24)	0.05	–	–	–	–
FLT3-ITD, present *versus* absent	–	–	1.96 (1.17-3.29)	0.01	2.15 (1.36-3.41)	0.001
MN1, high *versus* low*	–	–	1.92 (1.16-3.17)	0.01	–	–

Hazard ratios greater than (less than) 1.0 indicate higher (lower) risk for relapse or death (disease-free survival) or death (overall survival) or for failure to achieve complete remission, relapse or death (event-free survival) for the first category listed. 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 (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. HR: hazard ratio; CI: confidence interval; lncRNA: long non-coding RNA; *FLT3*-ITD: internal tandem duplication of the *FLT3* gene.

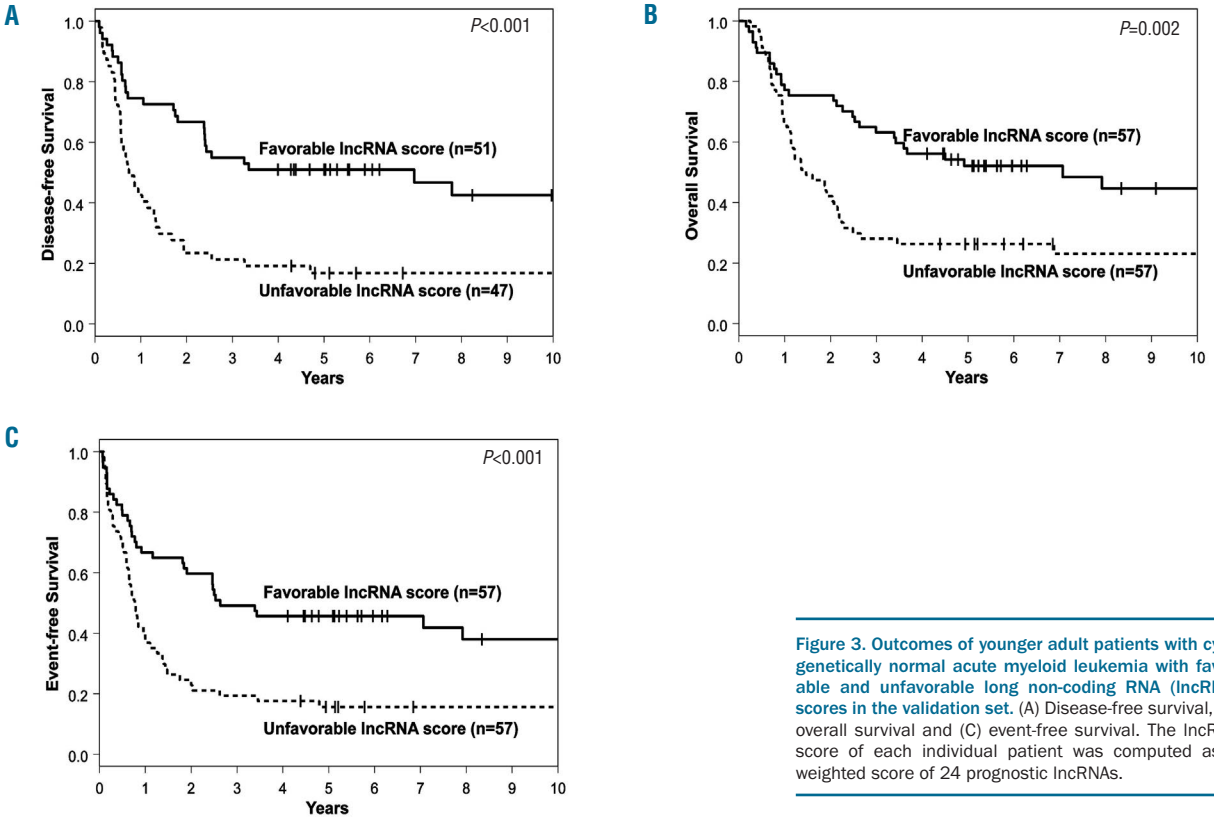


Figure 3. Outcomes of younger adult patients with cytogenetically normal acute myeloid leukemia with favorable and unfavorable long non-coding RNA (lncRNA) scores in the validation set. (A) Disease-free survival, (B) overall survival and (C) event-free survival. The lncRNA score of each individual patient was computed as a weighted score of 24 prognostic lncRNAs.

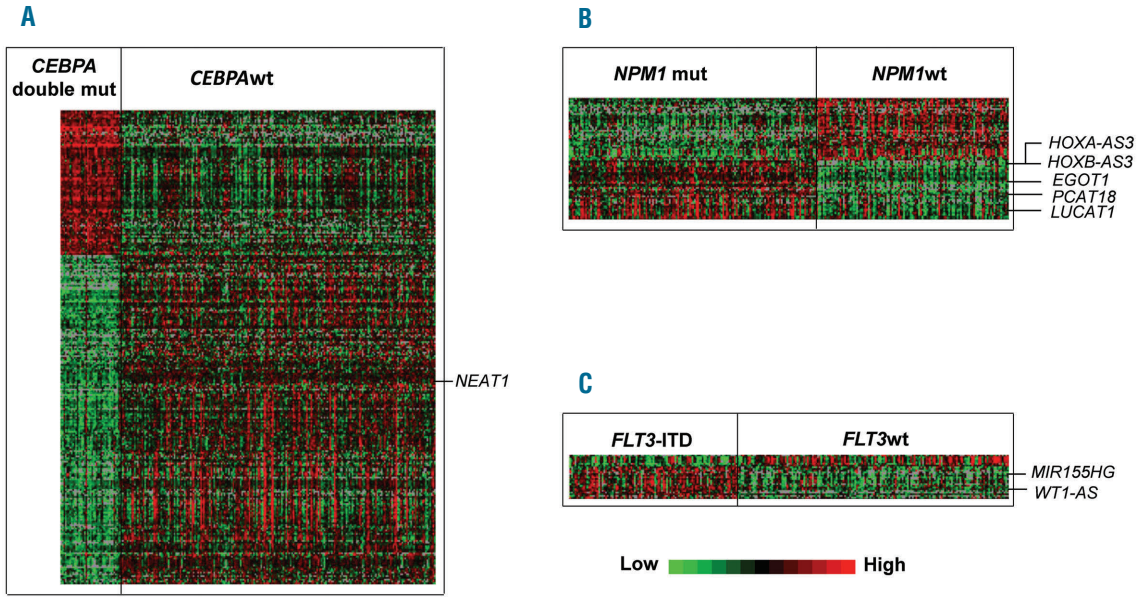


Figure 4. Long non-coding RNA (lncRNA) expression signatures associated with prognostic gene mutations in younger adult patients with cytogenetically normal acute myeloid leukemia. Heat maps for (A) double *CEBPA*, (B) *NPM1* and (C) *FLT3*-ITD mutation-related lncRNA signatures are presented. The lncRNA signatures were derived in the training set of the studied cohort. Expression values of the lncRNA transcripts are represented by color, with green indicating expression less than and red indicating expression greater than the median value for the lncRNA transcript. Gray color indicates lack of detectable expression. Rows represent lncRNA transcripts, and columns represent patients. Patients are grouped by the gene mutational status (i.e., mutated [mut] *versus* wild-type [wt]). For a full list of the lncRNAs that associated with prognostic gene mutations see the *Online Supplementary Appendix*.

SDHAP3 and *SENC3* were downregulated in patients with double *CEBPA* mutations. Of the 24 prognostic lncRNA genes, only *MIR155HG* has previously been associated with the clinical outcome of CN-AML patients.[40,48]

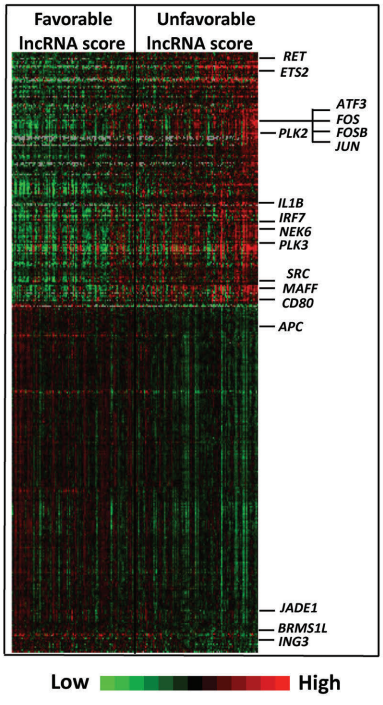
Linear combination of the weighted expression values of lncRNA transcripts yielded a prognostic score, which strongly associated with DFS, EFS and OS duration in younger adult CN-AML patients. Favorable lncRNA score status was an independent marker for longer DFS and EFS (and also showed a strong trend towards significance for longer OS). We were intrigued to find no overlap between the 48 prognostic lncRNAs that we previously identified in older CN-AML patients[26] and the 24 transcripts reported herein in younger patients. This finding could be interpreted as an additional biologic difference between CN-AML of younger and that of older patients, similar to the age-dependent difference in frequency of some recurrent prognostic gene mutations (e.g., mutations in the *ASXL1* and *RUNX1* genes).[1]

We also examined the associations between recurrent prognostic gene mutations and lncRNA expression, and found double *CEBPA* and *NPM1* mutations and *FLT3*-ITD to associate with distinct lncRNA signatures. We identified several lncRNAs that were commonly associated with these gene mutations in both younger and older CN-AML patients[26] (e.g., the *HOX*-loci embedded lncRNAs in the *NPM1* mutation-related lncRNA signature, *WT1-AS* in the *FLT3*-ITD-related signature, etc.). On the other hand, such gene mutations as *RUNX1* and *ASXL1* that are more frequent in older CN-AML patients and were found to associate with differential expression of lncRNAs[26] could not be tested in younger CN-AML patients, because too few younger patients harbored these mutations. Of note, mutations in the *TET2* gene showed no correlation with differential expression of lncRNA molecules in either older[26] or younger CN-AML patients, despite their impact on the epigenome[49] and adequate numbers of patients in both studied cohorts.

To gain insights into biologic pathways affected by differences in the lncRNA score, we investigated correlations between mRNA and miR expression signatures and lncRNA scores. In concordance with the adverse outcome that unfavorable lncRNA scores bestow, a number of previously described oncogenes and oncomiRs were found overexpressed in patients with unfavorable lncRNA score status. Similarly, genes with reported tumor-suppressive activity were found downregulated in this patient group. Only a small fraction of these transcripts have been reported in gene mutation-related mRNA signatures or other prognostic gene-expression signatures.[14-16] These findings indicate that, in addition to being independent of prognostic mutations, the differential expression of prognostic lncRNAs may regulate distinct molecular pathways in CN-AML. Notably, 5 important mediators of the AP-1 pathway (*ATF3, FOS, FOSB, JUN*, and *MAFF*) were found upregulated in patients with unfavorable lncRNA scores. The high number of cell cycle regulators and proliferation-inducing kinases that were also upregulated in this patient group is consistent with aberrant activation of the AP-1 pathway.

In this work, we used whole transcriptome sequencing techniques to identify and measure the expression of prognostic lncRNA molecules in younger adults with CN-AML. While this technology is becoming rapidly cheaper and widely available, it will most likely continue to serve

A



B

Gene Ontology functional groups that positively correlated with unfavorable lncRNA scores				
Biologic process	*P*	Fold enrichment	n of genes	FDR
Positive regulation of lymphocyte/leukocyte activation	2.16×10^{-5}	9.38	8	3.6×10^{-2}
Regulation of cell activation	5.51×10^{-6}	9.22	9	9×10^{-3}
Inflammatory response	1.44×10^{-7}	5.60	16	2.37×10^{-4}
Response to wounding	2.90×10^{-5}	4.50	21	4.78×10^{-5}
Regulation of apoptosis	4.85×10^{-6}	4.23	16	8×10^{-3}

C

Gene Ontology functional groups that negatively correlated with unfavorable lncRNA scores				
Biologic process	*P*	Fold enrichment	n of genes	FDR
Nucleoplasm part	1.81×10^{-9}	4.73	23	2.307×10^{-6}
Nucleoplasm	1.18×10^{-9}	3.75	29	1.507×10^{-6}
Nuclear lumen	5.86×10^{-10}	2.99	38	7.486×10^{-7}
Transcription	1.71×10^{-15}	2.63	71	2.620×10^{-12}
Intracellular organelle lumen	1.35×10^{-8}	2.57	40	1.730×10^{-5}

Abbreviations: n, number; FDR, false discovery rate.

Figure 5. Messenger RNA (mRNA) transcripts which associate with the long non-coding RNA (lncRNA) score in younger adults with cytogenetically normal acute myeloid leukemia (CN-AML). (A) Heat map of the gene-expression signature associated with the lncRNA score. Rows represent protein-coding genes and columns represent patients. Patients are grouped by lncRNA score: favorable on the left and unfavorable on the right. The lncRNA score of each individual patient was computed as a weighted score of 24 prognostic lncRNAs. Expression values of the lncRNA transcripts are represented by color: green: expression less than median value; red: expression greater than median value; gray: lack of detectable expression. Top 5 gene ontology terms that positively (B) or negatively (C) correlate with unfavorable lncRNA scores in younger patients with CN-AML are displayed. Gene ontology terms in (B) and (C) are ranked according to fold enrichment.

as a research tool rather than a diagnostic test to guide patient treatment. Despite this, alternative techniques for measuring RNA transcripts in a clinically applicable manner are available and are used to risk stratify patients with certain solid malignancies.[50] Similar assays could be developed in order to obtain targeted measurements of prognostic lncRNAs in a fast and clinically meaningful manner. The potential of such assays to improve risk stratification of AML patients should be evaluated in future prospective clinical studies.

Acknowledgments

The authors would like to thank: Donna Bucci and Wacharaphon Vongchucherd of The Alliance NCTN Biorepository and Biospecimen Resource for sample processing and storage services, Karl Kroll for technical support, and Lisa J. Sterling and Christine Finks of The Ohio State University, Comprehensive Cancer Center, Columbus, OH for data management.

Funding

This work was supported by the National Cancer Institute of the National Institutes of Health under Award Numbers CA180821 and CA180882 (to the Alliance for Clinical Trials in Oncology), CA077658, CA180850, CA180861, CA140158, CA16058, and CA197734. This work was also supported in part by the Leukemia Clinical Research Foundation, D Warren Brown Foundation, and the Pelotonia Fellowship Program. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health.

References

1. 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.
2. Byrd JC, Mrózek K, Dodge RK, et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). Blood. 2002;100(13):4325-4336.
3. Grimwade D, Hills RK, Moorman AV, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood. 2010;116(3):354-365.
4. Mrózek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev. 2004;18(2):115-136.
5. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med. 2016;374(23):2209-2221
6. Metzeler KH, Herold T, Rothenberg-Thurley M, et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. Blood. 2016;128(5):686-698.

7. Patel JP, Gönen M, Figueroa ME, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med. 2012;366(12):1079-1089.
8. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. N Engl J Med. 2015;373(12):1136-1152.
9. Baldus CD, Tanner SM, Ruppert AS, 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(5):1613-1618.
10. Marcucci G, Baldus CD, Ruppert AS, 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(36):9234-9242.
11. Schwind S, Marcucci G, Kohlschmidt J, 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(15): 4188-4198.
12. Garzon R, Volinia S, Liu C-G, et al. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. Blood. 2008;111(6):3183-3189.
13. Marcucci G, Radmacher MD, Maharry K, et al. MicroRNA expression in cytogenetically normal acute myeloid leukemia. N Engl J Med. 2008;358(18):1919-1928.
14. Bullinger L, Döhner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med. 2004; 350(16):1605-1616.
15. Metzeler KH, Hummel M, Bloomfield CD, et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood. 2008;112(10):4193-4201.
16. Li Z, Herold T, He C, et al. Identification of a 24-gene prognostic signature that improves the European LeukemiaNet risk classification of acute myeloid leukemia: an international collaborative study. J Clin Oncol. 2013;31(9):1172-1181.
17. Gibb EA, Brown CJ, Lam WL. The functional role of long non-coding RNA in human carcinomas. Mol Cancer. 2011;10:38.
18. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. Cell. 2009;136(4):629-641.
19. Zhao J, Sun BK, Erwin JA, Song J-J, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science. 2008;322(5902):750-756.
20. Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. Nature. 1991;351(6322):153-155.
21. Hung T, Wang Y, Lin MF, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet. 2011;43(7):621-629.
22. Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell. 2011;147(7):1537-1550.
23. Yildirim E, Kirby JE, Brown DE, et al. Xist RNA is a potent suppressor of hematologic cancer in mice. Cell. 2013;152(4):727-742.
24. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature. 2010;464(7291):1071-1076.
25. Trimarchi T, Bilal E, Ntziachristos P, et al. Genome-wide mapping and characterization of Notch-regulated long noncoding RNAs in acute leukemia. Cell. 2014;158(3): 593-606.
26. Garzon R, Volinia S, Papaioannou D, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci USA. 2014;111(52):18679-18684.
27. Zhou M, Guo M, He DA, et al. A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. J Transl Med. 2015;17 (13)231.
28. Mrózek K, Carroll AJ, Maharry K, 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(2):239-244.
29. Kroll KW, Eisfeld AK, Lozanski G, Bloomfield CD, Byrd JC, Blachly JS. MuCor: mutation aggregation and correlation. Bioinformatics. 2016; 32(10):1557-1558.
30. Marcucci G, Maharry K, Radmacher MD, 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(31):5078-5087.
31. Whitman SP, Archer KJ, Feng L, et al. Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of *FLT3*: a Cancer and Leukemia Group B study. Cancer Res. 2001;61(19):7233-7239.
32. Wouters BJ, Löwenberg B, Erpelinck-Verschueren CAJ, van Putten WLJ, Valk PJM, Delwel R. Double *CEBPA* mutations, but not single *CEBPA* mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood. 2009;113(13):3088-3091.
33. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression methods in biostatistics: linear, logistic, survival and repeated measures models. New York, NY: Springer, 2005.
34. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457-481.
35. 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.
36. Zeng C, Xu Y, Xu L, et al. Inhibition of long non-coding RNA NEAT1 impairs myeloid differentiation in acute promyelocytic leukemia cells. BMC Cancer. 2014;14:693.
37. Wagner LA, Christensen CJ, Dunn DM, et al. EGO, a novel, noncoding RNA gene, regulates eosinophil granule protein transcript expression. Blood. 2007;109(12): 5191-5198.
38. Crea F, Watahiki A, Quagliata L, et al. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. Oncotarget. 2014;5(3):764-774.
39. Thai P, Statt S, Chen CH, Liang E, Campbell C, Wu R. Characterization of a novel long noncoding RNA, SCAL1, induced by cigarette smoke and elevated in lung cancer cell lines. Am J Respir Cell Mol Biol. 2013;49(2):204-211.
40. Marcucci G, Maharry KS, Metzeler KH, 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(17):2086-2093.
41. Moorwood K, Charles AK, Salpekar A, Wallace JI, Brown KW, Malik K. Antisense WT1 transcription parallels sense mRNA and protein expression in fetal kidney and can elevate protein levels in vitro. J Pathol. 1998;185(4):352-359.
42. Smith MJ, Prochownik EV. Inhibition of c-*jun* causes reversible proliferative arrest and withdrawal from the cell cycle. Blood. 1992;79(8):2107-2115.
43. Gaynor R, Simon K, Koeffler P. Expression of c-*jun* during macrophage differentiation of HL-60 cells. Blood. 1991;77(12):2618-2623.
44. O'Connell RM, Taganov KD, Boldin MP, Cheng G, Baltimore D. MicroRNA-155 is induced during the macrophage inflammatory response. Proc Natl Acad Sci USA. 2007;104(5):1604-1609.
45. Ruiz-Lafuente N, Alcaraz-García MJ, Sebastián-Ruiz S, et al. IL-4 up-regulates miR-21 and the miRNAs hosted in the CLCN5 gene in chronic lymphocytic leukemia. PLoS One. 2015;10(4):e0124936.
46. Gerloff D, Grundler R, Wurm AA, et al. NF-κB/STAT5/miR-155 network targets PU.1 in FLT3-ITD-driven acute myeloid leukemia. Leukemia. 2015;29(3):535-547.
47. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015;21(11):1253-1261.
48. Marcucci G, Yan P, Maharry K, et al. Epigenetics meets genetics in acute myeloid leukemia: clinical impact of a novel seven-gene score. J Clin Oncol. 2014;32(6):548-556.
49. Ko M, Huang Y, Jankowska AM, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant *TET2*. Nature. 2010;468(7325):839-843.
50. Liu MC, Pitcher BN, Mardis ER, et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). NPJ Breast Cancer. 2016;(2):15023.