Clinical and molecular relevance of genetic variants in the non-coding transcriptome of patients with cytogenetically normal acute myeloid leukemia

Dimitrios Papaioannou,[1,2]* Hatice G. Ozer,[3]* Deedra Nicolet,[1,4,5] Amog P. Urs,[1] Tobias Herold,[6-9] Krzysztof Mrózek,[1,4] Aarif M.N. Batcha,[10,11] Klaus H. Metzeler,[12] Ayse S. Yilmaz,[3] Stefano Volinia,[13] Marius Bill,[1] Jessica Kohlschmidt,[1,4,5] Maciej Pietrzak,[3] Christopher J. Walker,[1,4] Andrew J. Carroll,[14] Jan Braess,[15] Bayard L. Powell,[16] Ann-Kathrin Eisfeld,[1,4] Geoffrey L. Uy,[17] Eunice S. Wang,[18] Jonathan E. Kolitz,[19] Richard M. Stone,[20] Wolfgang Hiddemann,[6-8] John C. Byrd,[1,4] Clara D. Bloomfield[1#] and Ramiro Garzon[1#]

[1]The Ohio State University, Comprehensive Cancer Center, Columbus, OH, USA; [2]Laura and Isaac Perlmutter Cancer Center, New York University School of Medicine, NYU Langone Health, New York, NY, USA; [3]The Ohio State University, Department of Biomedical Informatics, Columbus, OH, USA; [4]The Ohio State University Comprehensive Cancer Center, Clara D. Bloomfield Center for Leukemia Outcomes Research, Columbus, OH, USA; [5]Alliance Statistics and Data Center, The Ohio State University, Comprehensive Cancer Center, Columbus, OH, USA; [6]Laboratory for Leukemia Diagnostics, Department of Medicine III, University Hospital, LMU Munich, Munich, Germany; [7]German Cancer Consortium (DKTK), Heidelberg, Germany; [8]German Cancer Research Center (DKFZ), Heidelberg, Germany; [9]Research Unit Apoptosis in Hematopoietic Stem Cells, Helmholtz Zentrum München, German Center for Environmental Health (HMGU), Munich, Germany; [10]Institute for Medical Information Processing, Biometry and Epidemiology, LMU Munich, Munich, Germany; [11]Medical Data Integration Center (MeDIC), University Hospital, LMU Munich, Germany; [12]Department of Hematology, Cell Therapy & Hemostaseology, University Hospital Leipzig, Leipzig, Germany; [13]Department of Morphology, Surgery and Experimental Medicine, University of Ferrara, Ferrara, Italy; [14]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, USA; [15]Department of Oncology and Hematology, Hospital Barmherzige Brüder, Regensburg, Germany; [16]The Comprehensive Cancer Center of Wake Forest University, Winston-Salem, NC, USA; [17]Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO, USA; [18]Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA; [19]Monter Cancer Center, Hofstra Northwell School of Medicine, Lake Success, NY, USA and [20]Dana-Farber Cancer Institute, Harvard University, Boston, MA, USA

**DP and HGO contributed equally as co-first authors.*

#CDB and RG contributed equally as co-senior authors.

©2022 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2020.266643

Received: July 21, 2020.

Accepted: July 2, 2021.

Pre-published: July 15, 2021.

Correspondence: *RAMIRO GARZON* - ramiro.garzon@osumc.edu

SUPPLEMENTARY DATA

Clinical and molecular relevance of genetic variants in the non-coding transcriptome of patients with cytogenetically normal acute myeloid leukemia

Papaioannou et al.

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: Claire F. Verschraegen; 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: Jonathan E. Kolitz; 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: Umar Farooq; University of North Carolina, Chapel Hill, NC: Matthew I. Milowsky; Ft. Wayne Medical Oncology/Hematology, Ft. Wayne, IN: Sreenivasa Nattam; University of Maryland Greenebaum Cancer Center, Baltimore, MD: Heather D. Mannuel; Christiana Care Health Services, Inc., Newark, DE: Gregory Masters; Dartmouth Medical School, Lebanon, NH: Konstantin H. Dragnev; Duke University Medical Center, Durham, NC: Jeffrey Crawford; University of Vermont Cancer Center, Burlington, VT: Peter A. Kaufman; Eastern Maine Medical Center, Bangor, ME: Sarah J. Sinclair; Mount Sinai School of Medicine, New York, NY: Michael A. Schwartz; Weill Medical College of Cornell University, New York, NY: Scott T. Tagawa; University of Massachusetts Medical Center, Worcester, MA: William V. Walsh; University of California at San Francisco: Charalambos Andreadis; Western Pennsylvania Hospital, Pittsburgh, PA: Gene G. Finley; 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; Rhode Island Hospital, Providence, RI: Howard Safran; University of Illinois, Chicago, IL: John G. Quigley; University of California San Diego Moores Cancer Center, San Diego, CA: Lyudmila A. Bazhenova; Walter Reed National Military Medical Center, Bethesda, MD: Karen G. Zeman; Virginia Commonwealth University, Richmond, VA: Zhijian Chen; University of Minnesota, Minneapolis, MN: Robert A. Kratzke; Long Island Jewish Medical Center, Lake Success, NY: Jonathan E. Kolitz; University of Missouri/Ellis Fischel Cancer Center, Columbia, MO: Puja Nistala; 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.

Treatment protocols

Patients included in this 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. Patients with residual leukemia present in a bone marrow (BM) biopsy after one induction cycle received a second cycle of induction. CALGB/Alliance protocols did not include 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 achieving 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 [tyrosine kinase domain mutations of the *FLT3* gene (*FLT3*-TKD) *vs.* high allelic ratio of internal tandem duplications of the *FLT3* gene (*FLT3*-ITD) to *FLT3* wild-type (>0.7) *vs.* low allelic ratio of *FLT3*-ITD to *FLT3* wild-type (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.[7]

With regard to the German AML Cooperative Group (AMLCG) dataset, this consisted of patients that were recruited in the AMLCG 1999 and AMLCG 2008 clinical trials. The AMLCG 1999 trial (clinicaltrials.gov identifier NCT00266136) randomized patients <60 years to receive double induction with either one cycle of TAD-9 (thioguanine 100 mg/m^2 twice daily on days 3-9, cytarabine 100 mg/m^2/d continuous infusion on days 1 and 2 and 100 mg/m^2 twice daily on days 3-8, and daunorubicin 60 mg/m^2 on days 3-5) followed by one cycle of HAM (cytarabine 3 g/m^2 twice daily on days 1-3 and mitoxantrone 10 mg/m^2 on days 3-5) on day 21, or two cycles of HAM 21 days apart. Older patients (≥60 years) were randomized to receive induction therapy with one cycle of either TAD-9 or HAM. A second cycle of HAM was stipulated in the protocol if on day 21 ≥ 5% residual blasts were present in the BM at day 16. The trial recruited from 1999 to 2004.[8] The AMLCG 2008 trial (clinicaltrials.gov identifier NCT01382147) randomized patients <60 years and medically fit patients 60-70 years to receive either double induction chemotherapy with TAD-9 and HAM (21 days apart) as stipulated in the AMLCG 1999 trial, or dose-dense induction therapy [sHAM: cytarabine 3 g/m^2 (1 g/m^2 in patients ≥60 years) twice daily on days 1, 2, 8 and 9 and mitoxantrone 10 mg/m^2 on days 3, 4, 10 and 11]. Medically unfit and older patients were randomized

to receive either induction therapy according to the HAM regimen with reduced cytarabine dose (1g/m² per dose) and a second cycle of HAM if on day 16 ≥5% residual blasts were present in the BM or to receive dose-dense induction with sHAM (cytarabine, 1 g/m² per dose). The trial recruited from 2009 to 2012.[9] Younger patients in the AMLCG trial with available histocompatible family donors were treated with allogeneic SCT.[8,9]

Transcriptome analysis: library generation, sequencing and data analysis

For the CALGB/Alliance datasets, 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 software (version 1.18.64).

Cutadapt[10] and FastQC (Illumina, San Diego, CA) were used to apply quality control and adapter trimming to FastQ files. The STAR was used to align the short reads to the human genome (GENECODE ver22)[11] and the Htseq 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.

AMLCG RNAseq data were analyzed as described previously.[12]

With regard to detection of lncRNA variants, after quality control, adaptor-trimmed 50 base-pair-long paired-end reads were mapped to the human reference genome and variant calling was performed following the Genome Analysis Toolkit best practice recommendations for RNA Seq datasets. A two-pass variant calling approach was applied to ensure variant detection and depth of coverage. As a result, unique variant positions were identified on non-coding transcripts that do not overlap with coding exons and are located in low-complexity regions of the genome (i.e., excluding repeat masked regions and segmental duplications). These variants were further filtered based on their depth of coverage and distribution within the studied dataset and were then evaluated for associations with clinical outcome and the expression levels of other RNA transcripts.

For confirmatory analyses, we performed targeted analysis in the independent dataset of AMLCG patients. Specifically, we annotated and evaluated the prognostic value of the lncRNA variants that were detected in our initial cohort in the CALBG/Alliance patients.

To evaluate expression levels of lncRNAs in normal hematopoiesis, publicly available data deposited in the BloodSpot portal (www.bloodspot.eu) were used for analysis.

Mutational analyses

Targeted amplicon sequencing using the Miseq platform (Illumina, San Diego, CA) was used to analyze DNA samples for presence of recurrent prognostic gene mutations as described previously.[13,14] A variant allele frequency of ≥10% was used as the cut-off to distinguish between mutated versus wild-type alleles of these genes. *CEBPA* gene mutations and *FLT3*-ITD were evaluated using Sanger sequencing and fragment analysis, respectively, as described.[15,16] Per current guidelines, only patients with biallelic *CEBPA* mutations were considered to be mutated, whereas patients with single-allele mutations and those with wild-type *CEBPA* were grouped together for outcome analyses.

Molecular profiling of the AMLCG cohort were conducted as described previously.[8,9]

Definition of clinical endpoints

Clinical endpoints were defined according to generally accepted criteria.[17] 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.[17] 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

For each examined lncRNA variant, only patients with detectable expression of the lncRNA and adequate coverage of the variant position (i.e., depth of coverage >8) were analyzed. The expression distribution of each variant (i.e., continuous or bimodal) was used to divide the patients into groups. In the case of variants with a continuous distribution, a variant allele frequency cut-off of 10% was used to distinguish between the lncRNA variant expressers from the patients who expressed the wild-type lncRNA. In the case of bimodal distribution, the local minimum between the two distributions was used as the cut-off.

Multivariable proportional hazards models were constructed for DFS, OS and EFS, using a limited backwards elimination procedure. Variables considered for model inclusion were: expression of a lncRNA variant versus expression of the wild-type lncRNA, age (as a continuous variable, in 10-year increments), sex (male *vs.* female), race (white *vs.* 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 *vs.* absent), *ASXL1* mutations (mutated *vs.* wild-type), *CEBPA* mutations (double-mutated *vs.* single-mutated or wild-type), *DNMT3A* mutations (mutated *vs.* wild-type), *FLT3*-ITD (present *vs.* absent), *FLT3*-TKD (present *vs.* absent), *IDH1* mutations (mutated *vs.* wild-type), *IDH2* mutations (mutated *vs.* wild-type), *NPM1* mutations (mutated *vs.* wild-type), *RUNX1* mutations (mutated *vs.* wild-type), *TET2* mutations (mutated *vs.* wild-type), *WT1* mutations (mutated *vs.* wild-type),

ERG expression levels (high *vs.* low), *BAALC* expression levels (high *vs.* low), *MN1* expression levels (high *vs.* low), miR-181a expression levels (high *vs.* low), miR-3151 (expressed *vs.* not expressed), and miR-155 expression levels (high *vs.* 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=0.2 from the univariable analyses were considered for multivariable analyses. For the time-to-event endpoints, the proportional hazards assumption was checked for each variable individually.

***In vitro* experiments**

To evaluate the functional relevance of lncRNA variants, we performed forced overexpression experiments in two AML cell lines (K-562 and THP-1). K-562 and THP-1 cells were cultured in Roswell Park Memorial Institute 1640 (RPMI-1640) medium (Sigma), supplemented with 10% fetal bovine serum (FBS, Sigma), 1% antibiotic-antimycotic and 1% L-glutamine. All cell lines were maintained at 37°C and 5% CO_2.

We isolated total RNA from AML cells using the Trizol reagent (Invitrogen) according to instructions of the manufacturer. RNA was transcribed into cDNA using the Superscript III first strand synthesis system (Invitrogen) and random hexamer primers. cDNA was used as template for the amplification of our target lncRNA transcripts. The Phusion high-fidelity DNA polymerase (NEB) was used for PCRs according to the instructions of the manufacturer. Amplicons were cloned into pcDNA3.1 vectors (obtained by Addgene), previously linearized via digestion with the BstBI restriction endonuclease. Cloning was performed with the Gibson technique, using the Gibson assembly kit (NEB). Primers for PCR amplification and Gibson cloning are provided in

the Supplementary Table S11. Correct direction and sequence of the amplicons was confirmed with Sanger sequencing.

Delivery of the *SNHG15*wt- or *SNHG15*varT-pcDNA3 vectors to K-562 and THP-1 cells was performed with electroporation with the Nucleofector device according to the instructions of the manufacturer. Two micrograms of vector were used per reaction. Expression levels were evaluated with real-time quantitative PCR. Custom primers were obtained by IDT (provided in supplementary Table S11).

Apoptosis was assessed with Annexin/PI staining (BD Pharmingen) and flow cytometry analysis on an LSRII instrument. Proliferation was assessed with the colorimetric MTT assay according to the instructions of the manufacturer.

SUPPLEMENTARY REFERENCES

1. Kolitz JE, George SL, Marcucci G, et al. P-glycoprotein inhibition using valspodar (PSC-833) does not improve outcomes for patients under 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 RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus chemotherapy for acute myeloid leukemia with a *FLT3* mutation. N Engl J Med. 2017;377(5):454–464.
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. Büchner T, Berdel WE, Schoch C, et al. Double induction containing either two courses or one course of high-dose cytarabine plus mitoxantrone and postremission therapy by either autologous stem-cell transplantation or by prolonged maintenance for acute myeloid leukemia. J Clin Oncol. 2006;24(16):2480–2489.

9. Braess J, Amler S, Kreuzer K-A, et al. Sequential high-dose cytarabine and mitoxantrone (S-HAM) versus standard double induction in acute myeloid leukemia—a phase 3 study. Leukemia. 2018;32(12):2558–2571.

10. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17(1):10–12.

11. 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.

12. Herold T, Jurinovic V, Batcha AMN, et al. A 29-gene and cytogenetic score for the prediction of resistance to induction treatment in acute myeloid leukemia. Haematologica. 2018;103(3):456–465.

13. Garzon R, Volinia S, Papaioannou D, et al. Expression and prognostic impact of lncRNAs in acute myeloid leukemia. Proc Natl Acad Sci U S A. 2014;111(52):18679–18684.

14. Eisfeld A-K, Mrózek K, Kohlschmidt J, et al. The mutational oncoprint of recurrent cytogenetic abnormalities in adult patients with *de novo* acute myeloid leukemia. Leukemia. 2017;31(10):2211–2218.

15. 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.

16. 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.

17. 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.

Supplementary Table S1. List of recurrent genetic variants located within expressed long non-coding RNAs in younger adult patients with cytogenetically normal acute myeloid leukemia (provided as an Excel sheet).

Supplementary Table S2. List of recurrent long non-coding RNA variants that were identified in the training set of younger adult patients with CN-AML (CALGB/Alliance dataset) and could also be detected in the TCGA dataset of AML patients (provided as an Excel sheet).

Supplementary Table S3. List of recurrent genetic variants that are located within expressed long non-coding RNAs and associate with more than one clinical outcome endpoint in younger adult patients with cytogenetically normal acute myeloid leukemia (provided as an Excel sheet).

Supplementary Table S4. Outcome of younger adult patients with cytogenetically normal acute myeloid leukemia by expression of the C-to-T variant of the *RP5-1074L1.4* long non-coding RNA (*RP5-1074L1.4*varT) versus the wild-type lncRNA (*RP5-1074L1.4*wt).

Endpoint	***RP5-1074L1.4* varT (n=156)**	***RP5-1074L1.4* wt (n=87)**	***P***	**OR/HR (95% CI)**
Complete remission, n (%)	134 (86)	70 (80)	0.28	1.48 (0.74-2.97)
Disease-free survival			<0.001	0.56 (0.40-0.79)
Median, years	2.4	1.0		
% Disease-free at 5 years (95% CI)	44 (36-52)	26 (16-36)		
Overall survival			0.09	0.76 (0.55-1.05)
Median, years	3.1	2.2		
% Alive at 5 years (95% CI)	43 (35-51)	34 (25-44)		
Event-free survival			<0.001	0.59 (0.44-0.80)
Median, years	1.7	0.8		
% Event-free at 5 years (95% CI)	38 (30-46)	21 (13-30)		

CI: confidence interval; HR: hazard ratio; n: number; OR: odds ratio; varT: C-to-T variant.

Supplementary Table S5. Outcome of younger adult patients with cytogenetically normal acute myeloid leukemia by expression of the C-to-T variant of the *SNHG15* long non-coding RNA (*SNHG15*varT) versus the wild-type lncRNA (*SNHG15*wt).

Endpoint	***SNHG15*varT (n=239)**	***SNHG15*wt (n=67)**	***P***	**OR/HR (95% CI)**
Complete remission, n (%)	199 (83)	53 (79)	0.47	2.19 (0.90-5.33)
Disease-free survival			0.04	0.61 (0.41-0.90)
Median, years	1.7	1.0		
% Disease-free at 5 years (95% CI)	37 (30-43)	22 (12-34)		
Overall survival			0.07	0.64 (0.43-0.94)
Median, years	2.5	1.4		
% Alive at 5 years (95% CI)	41 (34-47)	32 (22-44)		
Event-free survival			0.04	0.58 (0.41-0.83)
Median, years	1.1	0.7		
% Event-free at 5 years (95% CI)	31 (25-36)	19 (11-30)		

CI: confidence interval; HR: hazard ratio; n: number; OR: odds ratio; varT: C-to-T variant.

Supplementary Table S6. Outcome of younger adult patients with cytogenetically normal acute myeloid leukemia by expression of the G-to-C variant of the *AL122127.25* long non-coding RNA (*AL122127.25*varC) versus the wild-type lncRNA (*AL122127.25*wt).

Endpoint	***AL122127.25 varC* (n=72)**	***AL122127.25 wt* (n=185)**	***P***	**OR/HR (95% CI)**
Complete remission, n (%)	53 (74)	155 (84)	0.08	0.54 (0.28-1.04)
Disease-free survival			0.01	1.57 (1.11-2.24)
Median, years	0.7	1.6		
% Disease-free at 5 years (95% CI)	17 (8-28)	35 (27-43)		
Overall survival			0.01	1.49 (1.08-2.04)
Median, years	1.1	2.2		
% Alive at 5 years (95% CI)	22 (17-25)	40 (32-47)		
Event-free survival			0.002	1.60 (1.19-2.16)
Median, years	0.7	1.1		
% Event-free at 5 years (95% CI)	12 (6-21))	30 (23-37)		

CI: confidence interval; HR: hazard ratio; n: number; OR: odds ratio; varC: G-to-C variant.

Supplementary Table S7. Comparison of clinical and molecular characteristics by expression of the C-to-T variant of the *RP5-1074L1.4* long non-coding RNA (*RP5-1074L1.4*varT) versus the wild-type lncRNA (*RP5-1074L1.4*wt) in younger adult patients with cytogenetically normal acute myeloid leukemia.

Characteristic	***RP5-1074L1.4* varT (n=156)**	***RP5-1074L1.4* wt (n=87)**	***P*[a]**
Age, years			0.37
Median	47	45	
Range	19-59	18-59	
Sex, n (%)			1.00
Male	85 (54)	48 (55)	
Female	71 (46)	39 (45)	
Race, n (%)			0.09
White	143 (94)	74 (87)	
Non-white	9 (6)	11 (13)	
Hemoglobin, g/dL			0.55
Median	9.1	9.6	
Range	4.9-25.1	4.2-13.7	
Platelet count, x10^9/L			0.36
Median	55	56	
Range	8-266	16-433	
WBC count, x10^9/L			0.39
Median	27.5	23.5	
Range	0.6-475.0	0.8-303.6	
Blood blasts, %			0.82
Median	63	60	
Range	0-97	0-97	
Bone marrow blasts, %			0.89
Median	69	68	
Range	21-95	18-92	
Extramedullary involvement, n (%)			0.65
Present	45 (29)	21 (26)	
Absent	109 (71)	60 (74)	
ASXL1, n (%)			0.10
Mutated	2 (1)	5 (6)	
Wild-type	147 (99)	79 (94)	
CEBPA, n (%)			1.00
Double Mutated	25 (17)	15 (17)	
Wild-type	120 (83)	71 (83)	
DNMT3A, n (%)			0.78
Mutated	57 (38)	34 (40)	
Wild-type	94 (62)	51 (60)	
FLT3-ITD, n (%)			0.50
Present	60 (40)	39 (45)	
Absent	91 (60)	48 (55)	
FLT3-TKD, n (%)			0.03
Present	18 (12)	3 (4)	
Absent	132 (88)	82 (96)	
IDH1, n (%)			0.21
Mutated	9 (6)	9 (11)	
Wild-type	142 (94)	76 (89)	
IDH2, n (%)			0.22
Mutated	10 (7)	10 (12)	
Wild-type	141 (93)	75 (88)	

Characteristic	*RP5-1074L1.4* varT (n=156)	*RP5-1074L1.4* wt (n=87)	*P*[a]
NPM1, n (%)			1.00
Mutated	91 (58)	51 (59)	
Wild-type	65 (42)	36 (41)	
RUNX1, n (%)			0.71
Mutated	6 (4)	2 (2)	
Wild-type	145 (96)	83 (98)	
TET2, n (%)			0.65
Mutated	13 (9)	9 (11)	
Wild-type	138 (91)	76 (89)	
WT1, n (%)			0.41
Mutated	20 (13)	8 (9)	
Wild-type	131 (87)	77 (91)	
ELN genetic group,[b] n (%)			0.32
Favorable	88 (61)	42 (51)	
Intermediate	37 (26)	27 (33)	
Adverse	19 (13)	14 (17)	
ERG expression,[c] n (%)			0.89
High	82 (53)	44 (51)	
Low	74 (47)	42 (49)	
BAALC expression,[c] n (%)			0.68
High	71 (51)	44 (54)	
Low	69 (49)	37 (46)	
MN1 expression,[c] n (%)			0.59
High	73 (49)	45 (54)	
Low	76 (51)	39 (46)	
miR-3151 expression, n (%)			1.00
Expressed	21 (17)	12 (17)	
Not expressed	102 (79)	57 (83)	
miR-155 expression,[c] n (%)			0.77
High	68 (54)	39 (56)	
Low	58 (46)	31 (44)	

n: number; ELN: European LeukemiaNet; *FLT3*-ITD: internal tandem duplications of the *FLT3* gene; *FLT3*-TKD: tyrosine kinase domain mutations of the *FLT3* gene; varT: C-to-T variant; WBC: white blood cell; wt: wild-type.

[a] *P*-values for categorical variables are from Fisher's exact test, *P*-values for continuous variables are from Wilcoxon rank sum test.

[b] Among patients with cytogenetically normal acute myeloid leukemia (CN-AML), the ELN Favorable-risk group comprises patients with biallelic mutations in *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN Intermediate-risk group includes patients with wild-type *CEBPA* and either wild-type *NPM1* without *FLT3*-ITD, wild-type *NPM1* and *FLT3*-ITD^{low} or mutated *NPM1* and *FLT3*-ITD^{high}. The ELN Adverse-risk group comprises patients with wild-type *CEBPA* and wild-type *NPM1* with *FLT3*-ITD^{high}, patients with mutated *TP53*, and patients with mutated *RUNX1* and/or mutated *ASXL1* (if these mutations do not co-occur with Favorable-risk AML subtype). *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.

[c] The median expression value was used as the cut point.

Supplementary Table S8. Comparison of clinical and molecular characteristics by expression of the C-to-T variant of the *SNHG15* long non-coding RNA (*SNHG15*varT) versus the wild-type lncRNA (*SNHG15*wt) in younger adult patients with cytogenetically normal acute myeloid leukemia.

Characteristic	*SNHG15*varT (n=297)	*SNHG15*wt (n=67)	*P*[a]
Age, years			0.03
Median	46	43	
Range	18-59	17-59	
Sex, n (%)			0.68
Male	120 (50)	36 (54)	
Female	119 (50)	31 (46)	
Race, n (%)			0.02
White	216 (93)	54 (82)	
Non-white	17 (7)	12 (18)	
Hemoglobin, g/dL			0.88
Median	9.3	9.1	
Range	5.2-15.1	4.6-13.2	
Platelet count, $x10^9/L$			0.56
Median	56	60	
Range	8-445	8-213	
WBC count, $x10^9/L$			0.49
Median	30.3	28.5	
Range	0.8-303.6	0.8-303.8	
Blood blasts, %			0.68
Median	63	62	
Range	0-97	0-97	
Bone marrow blasts, %			0.80
Median	70	66	
Range	18-94	21-96	
Extramedullary involvement, n (%)			0.65
Present	68 (29)	21 (32)	
Absent	167 (71)	45 (68)	
ASXL1, n (%)			0.73
Mutated	9 (4)	3 (5)	
Wild-type	220 (96)	63 (95)	
CEBPA, n (%)			0.13
Double mutated	33 (15)	15 (23)	
Wild-type	191 (85)	51 (77)	
DNMT3A, n (%)			0.57
Mutated	94 (41)	24 (36)	
Wild-type	135 (59)	42 (64)	
FLT3-ITD, n (%)			0.78
Present	92 (39)	27 (42)	
Absent	143 (61)	38 (58)	
FLT3-TKD, n (%)			0.15
Present	25 (11)	3 (5)	
Absent	204 (89)	62 (95)	
IDH1, n (%)			0.12
Mutated	22 (10)	2 (3)	
Wild-type	207 (90)	64 (97)	

Characteristic	*SNHG15*varT (n=297)	*SNHG15*wt (n=67)	*P*[a]
IDH2, n (%)			0.80
Mutated	19 (8)	6 (9)	
Wild-type	210 (92)	60 (91)	
NPM1, n (%)			0.21
Mutated	142 (59)	34 (51)	
Wild-type	97 (41)	33 (49)	
RUNX1, n (%)			0.76
Mutated	12 (5)	4 (6)	
Wild-type	217 (95)	62 (94)	
TET2, n (%)			0.52
Mutated	25 (11)	9 (14)	
Wild-type	204 (89)	57 (86)	
WT1, n (%)			0.40
Mutated	26 (11)	10 (15)	
Wild-type	203 (89)	56 (85)	
ELN genetic group,[b] n (%)			0.38
Favorable	126 (58)	36 (55)	
Intermediate	61 (28)	15 (23)	
Adverse	32 (15)	14 (22)	
ERG expression,[c] n (%)			0.68
High	124 (52)	37 (55)	
Low	113 (48)	30 (45)	
BAALC expression,[c] n (%)			0.16
High	113 (51)	40 (62)	
Low	108 (49)	25 (38)	
MN1 expression,[c] n (%)			0.05
High	115 (50)	42 (65)	
Low	115 (50)	23 (35)	
miR-3151 expression, n (%)			0.17
Expressed	33 (16)	14 (25)	
Not expressed	169 (84)	41 (75)	
miR-155 expression,[c] n (%)			1.00
High	104 (51)	28 (51)	
Low	98 (49)	27 (49)	

n: number; ELN: European LeukemiaNet; *FLT3*-ITD: internal tandem duplications of the *FLT3* gene; *FLT3*-TKD: tyrosine kinase domain mutations of the *FLT3* gene; varT: C-to-T variant; WBC: white blood cell; wt: wild-type.

[a] *P*-values for categorical variables are from Fisher's exact test, *P*-values for continuous variables are from Wilcoxon rank sum test.

[b] Among patients with cytogenetically normal acute myeloid leukemia (CN-AML), the ELN Favorable-risk group comprises patients with biallelic mutations in *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN Intermediate-risk group includes patients with wild-type *CEBPA* and either wild-type *NPM1* without *FLT3*-ITD, wild-type *NPM1* and *FLT3*-ITD^{low} or mutated *NPM1* and *FLT3*-ITD^{high}. The ELN Adverse-risk group comprises patients with wild-type *CEBPA* and wild-type *NPM1* with *FLT3*-ITD^{high}, patients with mutated *TP53*, and patients with mutated *RUNX1* and/or mutated *ASXL1* (if these mutations do not co-occur with Favorable-risk AML subtype). *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.
[c] The median expression value was used as the cut point.

Supplementary Table S9. Comparison of clinical and molecular characteristics by expression of the C-variant of the *AL122127.25* long non-coding RNA (*AL122127.25*varC) versus the wild-type lncRNA (*AL122127.25*wt) in younger adult patients with cytogenetically normal acute myeloid leukemia.

Characteristic	***AL122127.25*varC (n=72)**	***AL122127.25*wt (n=185)**	***P*[a]**
Age, years			0.22
Median	44	47	
Range	19-59	17-59	
Sex, n (%)			0.13
Male	29 (40)	95 (51)	
Female	43 (60)	90 (49)	
Race, n (%)			0.36
White	61 (86)	164 (91)	
Non-white	10 (14)	17 (9)	
Hemoglobin, g/dL			0.16
Median	9.4	9.0	
Range	4.8-14.4	4.2-15.1	
Platelet count, x10^9/L			0.08
Median	66	49	
Range	10-266	8-347	
WBC count, x10^9/L			0.85
Median	28.8	27.1	
Range	0.9-475.0	0.8-298.4	
Blood blasts, %			0.71
Median	58	65	
Range	0-97	0-97	
Bone marrow blasts, %			0.96
Median	66	68	
Range	19-96	20-95	
Extramedullary involvement, n (%)			0.21
Present	15 (21)	53 (30)	
Absent	55 (79)	126 (70)	
ASXL1, n (%)			0.75
Mutated	4 (6)	8 (5)	
Wild-type	66 (94)	168 (95)	
CEBPA, n (%)			0.20
Double mutated	5 (8)	25 (14)	
Wild-type	61 (92)	154 (86)	
DNMT3A, n (%)			0.15
Mutated	33 (47)	64 (36)	
Wild-type	37 (53)	112 (64)	
FLT3-ITD, n (%)			0.12
Present	38 (54)	79 (43)	
Absent	32 (46)	103 (57)	
FLT3-TKD, n (%)			0.26
Present	5 (7)	22 (13)	
Absent	65 (93)	153 (87)	
IDH1, n (%)			0.24
Mutated	10 (14)	15 (9)	
Wild-type	60 (86)	161 (91)	
IDH2, n (%)			0.17
Mutated	11 (16)	16 (9)	

Characteristic	*AL122127.25*varC (n=72)	*AL122127.25*wt (n=185)	P^a
Wild-type	59 (84)	160 (91)	
NPM1, n (%)			0.78
Mutated	44 (61)	109 (59)	
Wild-type	28 (39)	76 (41)	
RUNX1, n (%)			1.00
Mutated	5 (7)	13 (7)	
Wild-type	65 (93)	163 (93)	
TET2, n (%)			0.64
Mutated	8 (11)	16 (9)	
Wild-type	62 (89)	160 (91)	
WT1, n (%)			1.00
Mutated	11 (16)	27 (15)	
Wild-type	59 (84)	149 (85)	
ELN genetic group,[b] n (%)			0.47
Favorable	30 (45)	91 (54)	
Intermediate	23 (35)	48 (28)	
Adverse	13 (20)	30 (18)	
ERG expression,[c] n (%)			0.57
High	43 (60)	119 (64)	
Low	29 (40)	66 (36)	
BAALC expression,[c] n (%)			0.32
High	35 (51)	102 (58)	
Low	34 (49)	74 (42)	
MN1 expression,[c] n (%)			0.89
High	40 (57)	99 (55)	
Low	30 (43)	80 (45)	
miR-3151 expression, n (%)			1.00
Expressed	11 (20)	31 (20)	
Not expressed	44 (80)	122 (80)	
miR-155 expression,[c] n (%)			0.87
High	31 (56)	89 (58)	
Low	24 (44)	64 (42)	

n: number; ELN: European LeukemiaNet; *FLT3*-ITD: internal tandem duplications of the *FLT3* gene; *FLT3*-TKD: tyrosine kinase domain mutations of the *FLT3* gene; varC: G-to-C variant; WBC: white blood cell; wt: wild-type.

[a] *P*-values for categorical variables are from Fisher's exact test, *P*-values for continuous variables are from Wilcoxon rank sum test.

[b] Among patients with cytogenetically normal acute myeloid leukemia (CN-AML), the ELN Favorable-risk group comprises patients with biallelic mutations in *CEBPA* and patients with mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}. The ELN Intermediate-risk group includes patients with wild-type *CEBPA* and either wild-type *NPM1* without *FLT3*-ITD, wild-type *NPM1* and *FLT3*-ITD^{low} or mutated *NPM1* and *FLT3*-ITD^{high}. The ELN Adverse-risk group comprises patients with wild-type *CEBPA* and wild-type *NPM1* with *FLT3*-ITD^{high}, patients with mutated *TP53*, and patients with mutated *RUNX1* and/or mutated *ASXL1* (if these mutations do not co-occur with Favorable-risk AML subtype). *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.

[c] The median expression value was used as the cut point.

Supplementary Table S10. Multivariable analyses for outcome in younger adult patients with cytogenetically normal acute myeloid leukemia by *AL122127.25*varC versus *AL122127.25*wt expression.

Variables in final models	OS		EFS	
	HR (95% CI)	***P***	**HR (95% CI)**	***P***
AL122127.25, varC *vs.* wild-type	1.57 (1.12-2.21)	0.009	1.59 (1.16-2.18)	0.004
FLT3-ITD, present *vs.* absent	2.23 (1.62-3.07)	<0.001	1.97 (1.44-2.70)	<0.001
WT1, mutated *vs.* wild-type	1.66 (1.11-2.48)	0.01	1.90 (1.27-2.83)	0.002
MN1 expression, high *vs.* low	2.00 (1.44-2.79)	<0.001		
Age, continuous, 10 year increments	1.33 (1.15-1.55)	<0.001		
NPM1, mutated *vs.* wild-type	-	-	0.69 (0.48-0.97)	0.04
DNMT3A, mutated *vs.* wild-type	-	-	1.41 (1.02-1.95)	0.04
BAALC expression, high *vs.* low	-	-	1.47 (1.05-2.06)	0.03

n: number; CI: confidence interval; EFS: event-free survival; *FLT3*-ITD: internal tandem duplications of the *FLT3* gene; HR: hazard ratio; OS: overall survival; varC: G-to-C variant; *vs.*: versus; wt: wild-type.
NOTE: Hazard ratios greater than (less than) 1.0 indicate higher (lower) risk for death (overall survival) or for failure to achieve complete remission, relapse or death (event-free survival) for the first category listed for the categorical variables. Variables considered for model inclusion were: Variables considered for model inclusion were: *AL122127.25* (varC *vs.* wild-type), age (as a continuous variable, in 10-year increments), sex (male *vs.* female), race (white *vs.* 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 *vs.* absent), *ASXL1* mutations (mutated *vs.* wild-type), *CEBPA* mutations (double-mutated *vs.* single-mutated or wild-type), *DNMT3A* mutations (mutated *vs.* wild-type), *FLT3*-ITD (present *vs.* absent), *FLT3*-TKD (present *vs.* absent), *IDH1* mutations (mutated *vs.* wild-type), *IDH2* mutations (mutated *vs.* wild-type), *NPM1* mutations (mutated *vs.* wild-type), *RUNX1* mutations (mutated *vs.* wild-type), *TET2* mutations (mutated *vs.* wild-type), *WT1* mutations (mutated *vs.* wild-type), *ERG* expression levels (high *vs.* low), *BAALC* expression levels (high *vs.* low), *MN1* expression levels (high *vs.* low), miR-181a expression levels (high *vs.* low), miR-3151 (expressed *vs.* not expressed), and miR-155 expression levels (high *vs.* low).

Supplementary Table S11. List of oligonucleotides used in functional *in vitro* experiments.

	***SNHG15* RT-qPCR Primers**
Fw_Primer	5'- TTCCTTGGAACCTGTGCA G3'
Rev_Primer	5'- CTGGAAATTCCTGACTCCTTC C-3'
Probe	5'-56-FAM/ATTTGAATG/Zen/C AAG CCT TGG CAC CT/3IABkFQ-3'
	***SNHG15* Amplification Primers**
Fw_Primer	5'- GGACTACGCGGTGACGTCGAG-3"
Rev_Primer	5'-TCACATATTTAAATCCATATTTATTAAATATCCC TTGTTTTTCATTCAG- 3'
	***SNHG15* Gibson Cloning Primers**
Fw_Primer	5'- GCCTAGGCTTTTGCAAAAAGCTCCCGGACTACGC GGTGACGTCGAG-3'
Rev_Primer	5'-GATCAGATCCGAAAATGGATATACAAGCTCCCTCACAT ATTTAAATCCATATTTATTAAATATCCCTTGTTTTTCATTCAG -3'

Supplementary Figures

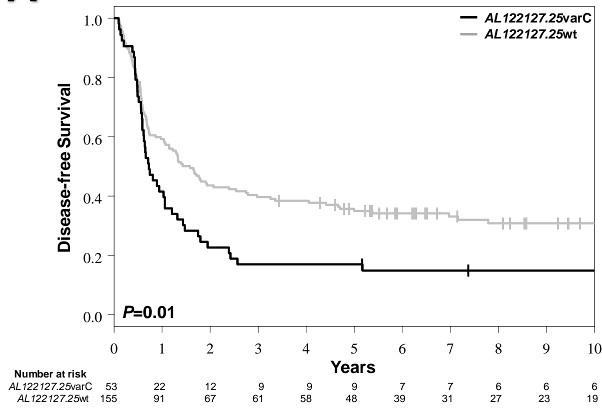
Supplementary Figure S1. Outcome of younger adult patients with cytogenetically normal acute myeloid leukemia by expression of the G-to-C variant of the long non-coding RNA *AL122127.25* (*AL122127.25*varC) versus the wild-type *AL122127.25* lncRNA (*AL122127.25*wt). (A) Disease-free survival, (B) overall survival and (C) event-free survival.

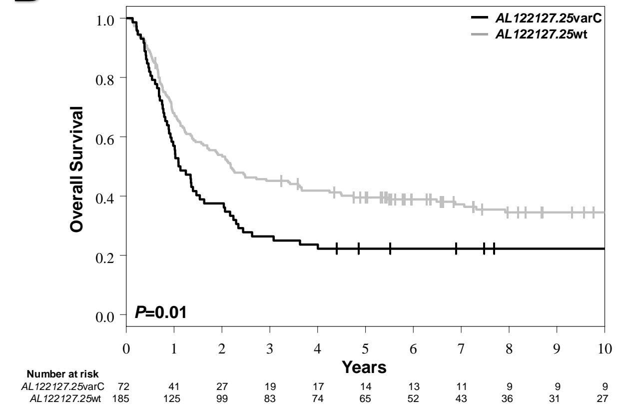
Supplementary Figure S2. Correlation between the frequency of the G-to-C variant and the expression levels of the long non-coding RNA *AL122127.25*. Distance of correlation (dcor) is used as a metric for evaluating the strength of correlation between the allele frequency of the variant and the abundance of the long non-coding RNA (dcor<0.5 indicates no correlation).

Supplementary Figure S3. Expression levels of the *SLC16A4* gene in CN-AML patients who express the wild-type *RP5-1074L1.4* lncRNA (*RP5-1074L1.4*wt) and in those who express the the C-to-T variant of the *RP5-1074L1.4* lncRNA (*RP5-1074L1.4*varT).

Supplementary Figure S4. Expression levels of *SNHG15* in subpopulations of stem cells, progenitor and mature blood cells during healthy hematopoiesis. The image was generated using publicly available data in the Bloodspot portal (www.bloodspot.eu). In the figure, HSC indicates hematopoietic stem cells; CMP, common myeloid progenitors; GMP, granulocyte-monocyte progenitors; MEP, megakaryocyte-erythroid progenitors; PM, promyelocytes; MY, myelocytes; PMN, polymorphonuclear cells; mDC, myeloid dendritic cells; pCD, plasmacytoid dendritic cells. BM denotes samples obtained from bone marrow, whereas PB denotes samples obtained from peripheral blood.

Figure S1





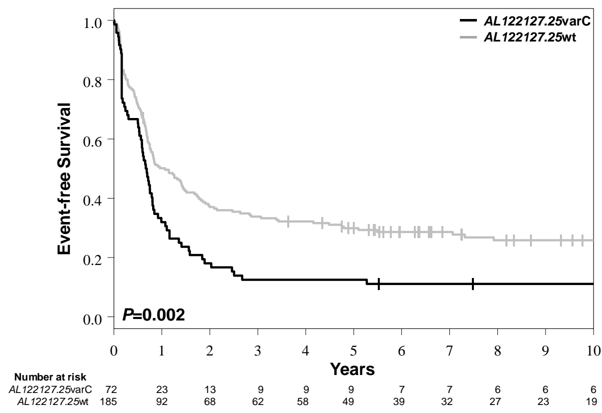


Figure S2

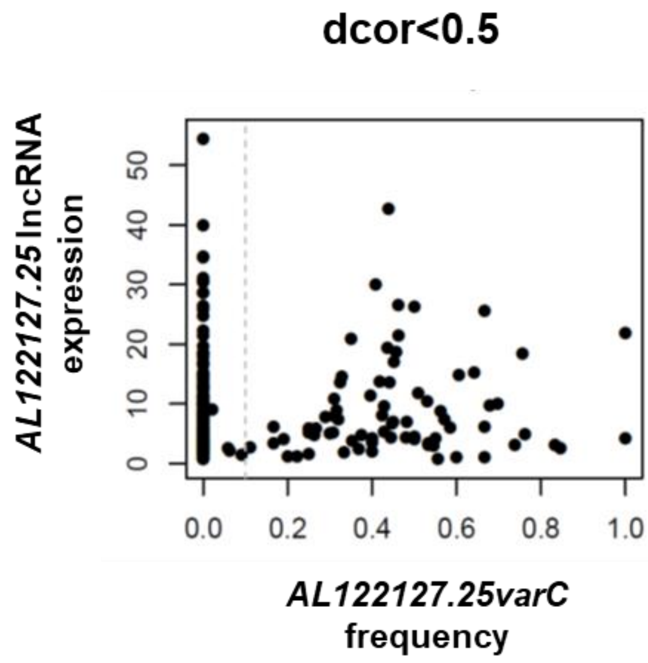


Figure S3

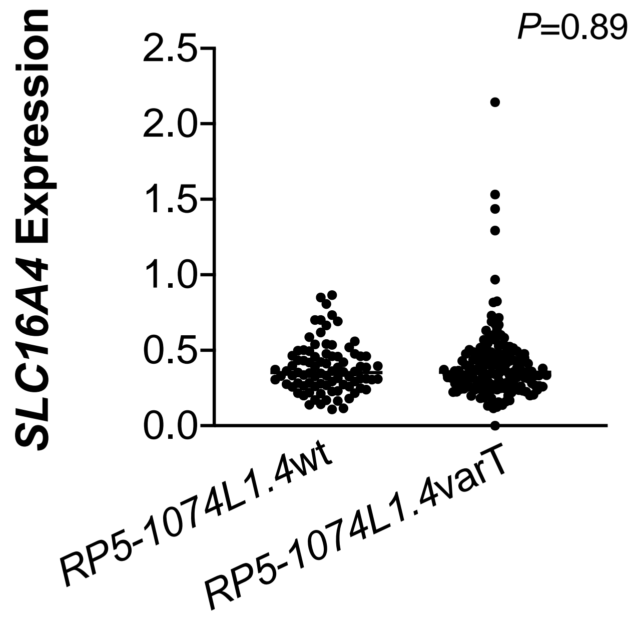


Figure S4