haematologica
Journal of The Ferrata Storti Foundation

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

by Dimitrios Papaioannou, Hatice G. Ozer, Deedra Nicolet, Amog P. Urs, Tobias Herold, Krzysztof Mrózek, Aarif M.N. Batcha, Klaus H. Metzeler, Ayse S. Yilmaz, Stefano Volinia, Marius Bill, Jessica Kohlschmidt, Maciej Pietrzak, Christopher J. Walker, Andrew J. Carroll, Jan Braess, Bayard L. Powell, Ann-Kathrin Eisfeld, Geoffrey L. Uy, Eunice S. Wang, Jonathan E. Kolitz, Richard M. Stone, Wolfgang Hiddemann, John Byrd, Clara D. Bloomfield, and Ramiro Garzon

Haematologica 2021 [Epub ahead of print]

Citation: Dimitrios Papaioannou, Hatice G. Ozer, Deedra Nicolet, Amog P. Urs, Tobias Herold, Krzysztof Mrózek, Aarif M.N. Batcha, Klaus H. Metzeler, Ayse S. Yilmaz, Stefano Volinia, Marius Bill, Jessica Kohlschmidt, Maciej Pietrzak, Christopher J. Walker, Andrew J. Carroll, Jan Braess, Bayard L. Powell, Ann-Kathrin Eisfeld, Geoffrey L. Uy, Eunice S. Wang, Jonathan E. Kolitz, Richard M. Stone, Wolfgang Hiddemann, John Byrd, Clara D. Bloomfield, and Ramiro Garzon. Clinical and molecular relevance of genetic variants in the non-coding transcriptome of patients with cytogenetically normal acute myeloid leukemia. Haematologica. 2021; 106:xxx doi:10.3324/haematol.2021.266643

Article

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;
[2]Laura and Isaac Perlmutter Cancer Center, New York University School of Medicine, NYU Langone Health, New York, NY;
[3]The Ohio State University, Department of Biomedical Informatics, Columbus, OH;
[4]The Ohio State University Comprehensive Cancer Center, Clara D. Bloomfield Center for Leukemia Outcomes Research, Columbus, OH;
[5]Alliance Statistics and Data Center, The Ohio State University, Comprehensive Cancer Center, Columbus, OH;
[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;
[15]Department of Oncology and Hematology, Hospital Barmherzige Brüder, Regensburg, Germany;
[16]The Comprehensive Cancer Center of Wake Forest University, Winston-Salem, NC;

[17]Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO;
[18]Roswell Park Comprehensive Cancer Center, Buffalo, NY;
[19]Monter Cancer Center, Hofstra Northwell School of Medicine, Lake Success, NY;
[20]Dana-Farber Cancer Institute, Harvard University, Boston, MA, USA

* DP and HGO contributed equally to this work.
† CDB and RG contributed equally to this work as senior authors.

Running Head: lncRNA variants in CN-AML

Keywords: Long non-coding RNA, genetic variant, acute myeloid leukemia, clinical outcome, molecular pathways.

Corresponding Author: Ramiro Garzon, MD, The Ohio State University Comprehensive Cancer Center, 460 West 12th Avenue, Columbus, OH 43210. Phone: 614-685-9180; fax: 614-292-4080; e-mail: ramiro.garzon@osumc.edu

Disclosure of Potential Conflicts of Interest

The authors declare that there are no competing financial interests in relation to the work described in this manuscript.

Contributions: DP, HGO, DN, CDB and RG designed the study. DP, HGO, DN, TH, KM, AMNB, SV, ASY, MP, CJW, MB, WH, JCB, CDB and RG contributed to the data interpretation. DP, DN, KM, CDB and RG wrote the manuscript. HGO, DN and JK performed statistical analysis. DP and APU designed and performed experiments. TH, KM, KHM, AJC, JB, BLP, A-KE, GLU, ESW, JEK, RMS, WH, JCB, CDB and RG were involved directly or indirectly in the care of patients and/or sample procurement. KM and AJC oversaw the cytogenetic analyses. CDB and RG supervised the study. All authors read and approved the final version of the manuscript.

Word count: abstract: 233 words; main text: 3,754 words; tables: 2; figures: 4; references: 50; supplementary files: 1.

Clinicaltrials.gov Identifiers: NCT00048958 (CALGB-8461); NCT00899223 (CALGB-9665); and NCT00900224 (CALGB-20202).
Clinical and molecular relevance of genetic variants in the non-coding transcriptome of patients with cytogenetically normal acute myeloid leukemia

The authors would like to thank: The Alliance Hematology Malignancy Biorepository supported by Washington University subcontract WU-15-398/WU-16-51 and The Alliance NCTN Biorepository and Biospecimen Resource (U24CA196171) for sample processing and storage services, and Lisa J. Sterling (The Ohio State University, Comprehensive Cancer Center, Columbus, OH 43221) for data management. Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10CA180821 and U10CA180882 (to the Alliance for Clinical Trials in Oncology), U10CA077658, U10CA180850, U10CA180861, CA140158, CA16058, UG1CA233338 and R35CA197734. This work was also supported in part by the Leukemia Clinical Research Foundation, D. Warren Brown Foundation and the Pelotonia Fellowship Program. TH was supported by a Physician Scientists Grant (G-509200-004) from

the Helmholtz Zentrum München. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Celebrating the life and accomplishments of Dr. Clara D. Bloomfield (1942-2020).

ABSTRACT

Expression levels of long non-coding RNAs (lncRNAs) have been shown to associate with clinical outcome of patients with cytogenetically normal acute myeloid leukemia (CN-AML). However, the frequency and clinical significance of genetic variants in the nucleotide sequences of lncRNAs in AML patients is unknown. Herein, we analyzed total RNA sequencing data of 377 younger adults (aged <60 years) with CN-AML, who were comprehensively characterized with regard to clinical outcome. We used available genomic databases and stringent filters to annotate genetic variants unequivocally located in the non-coding transcriptome of AML patients. We detected 981 variants, which are recurrently present in lncRNAs that are expressed in leukemic blasts. Among these variants, we identified a cytosine-to-thymidine variant in the lncRNA *RP5-1074L1.4* and a cytosine-to-thymidine variant in the lncRNA *SNHG15*, which independently associated with longer survival of CN-AML patients. The presence of the *SNHG15* cytosine-to-thymidine variant was also found to associate with better outcome in an independent dataset of CN-AML patients, despite differences in treatment protocols and RNA sequencing techniques. To gain biological insights, we cloned and overexpressed both wild-type and variant versions of the *SNHG15* lncRNA. In keeping with its negative prognostic impact, overexpression of the wild-type *SNHG15* associated with higher proliferation rate of leukemic blasts when compared with the cytosine-to-thymidine variant. We conclude that recurrent genetic variants of lncRNAs that are expressed in the leukemic blasts of CN-AML patients have prognostic and potential biologic significance.

Introduction

Acute myeloid leukemia (AML) is heterogeneous with regard to the patients' clinical course and the underlying molecular lesions that drive the disease.[1-2] Research efforts of the past four decades have identified a growing list of genetic alterations associated with clinical outcome that could be used as biomarkers for the risk stratification of the patients' treatment. These alterations include chromosomal abnormalities,[3-5] gene mutations,[6-11] and aberrant expression of RNA transcripts.[12-16] The advent of next-generation sequencing has revealed that AML displays notable heterogeneity at the level of isolated cases; leukemic blasts of individual AML patients represent, in many instances, the sum of distinct clonal subpopulations, within which mutations in different genes coexist and cooperate.[17,18] While such sequencing efforts continue to expand our understanding of AML pathogenesis, the majority of them are focused on the protein-coding fraction of the genome, which is, comparatively, its smallest part.[19]

The non-protein-coding part of the genome, a large fraction of which is actively transcribed into non-coding RNAs, is gaining gradual recognition for its important regulatory role.[20,21] Long non-coding RNAs (lncRNAs), which are transcripts longer than 200 nucleotides and, per definition, lack protein-coding potential, regulate many key cellular functions in health and disease.[22-24] Deregulated expression of individual lncRNAs has been demonstrated to significantly affect the cancer phenotype and patients' clinical outcome.[25-29] We and others have previously shown that aberrant expression of small subsets of lncRNAs independently associate with the clinical outcome of patients with cytogenetically normal AML (CN-AML).[30-33] With regard to

variations in the nucleotide sequences of lncRNAs, it has previously been reported that disease-associated single nucleotide polymorphisms (SNPs) are enriched in the genetic loci encoding these transcripts.[34-36] In addition, acquired mutations of lncRNAs that are recurrently detectable in the leukemic blasts have previously been identified.[37] However, to our knowledge, their prognostic and biologic significance in AML have not yet been studied.

Herein, we analyzed whole transcriptome sequencing data of younger adults with the goal to evaluate the clinical and biologic relevance of lncRNA variants in CN-AML. We used a panel of currently available databases of human genomic polymorphisms to annotate recurrent genetic variants located within expressed lncRNAs. We show that a subset of these variants independently associates with clinical outcome of CN-AML patients.

Methods

Patients and treatment

Exploratory analysis was conducted in pretreatment bone marrow (BM) or blood samples from 377 younger adults (aged <60 years, range 18-59) with *de novo* CN-AML. Patients were treated with intensive, first-line chemotherapy on Cancer and Leukemia Group B (CALGB)/Alliance for Clinical Trials in Oncology (Alliance) trials. Confirmatory analyses were conducted in a set of 135 CN-AML patients (75 of whom were younger than 60 years and 60 were older) enrolled on clinical trials of the German AML Cooperative Group (AMLCG).[38,39]

All patients provided written informed consent regarding the research use of their specimens. 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 of CALGB/Alliance patients were performed in CALGB/Alliance-approved institutional laboratories and results were confirmed by central karyotype review.[40] The diagnosis of normal karyotype was based on analysis of ≥20 metaphases obtained from BM specimens subjected to short-term (24- or 48-hour) unstimulated cultures.[40]

Mutational analyses of patient samples were conducted with Sanger sequencing (for the *CEBPA* gene), fragment analysis [for detection of *FLT3*-internal tandem duplications (*FLT3*-ITD)] and targeted amplicon sequencing (for all other prognostic

gene mutations), as reported previously.[31,41-43] Molecular and cytogenetic profiling of the AMLCG cohort were obtained as described previously.[38,39]

Transcriptome analyses

RNA samples of the patients treated on CALGB/Alliance protocols were analyzed with total RNA Seq after depletion of ribosomal and mitochondrial RNA using the Illumina HiSeq 2500 platform. The results of the RNA Seq analysis have been deposited in the functional genomics data repository GEO and are publicly available under the accession number GSE137851. Patients in the AMLCG cohort were analyzed with RNA sequencing following selection for poly-adenylated transcripts (poly-A RNA Seq) as described previously.[16]

For exploratory analyses, 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.[44] A two-pass variant calling approach was applied to ensure variant detection and depth of coverage (Figure 1). 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 evaluated for associations with clinical outcome and the expression levels of other RNA transcripts.

Statistical analyses

Clinical endpoint definitions are provided in the Supplementary Data. 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 estimated probabilities of disease-free (DFS), overall (OS) and event-free (EFS) survival were calculated using the Kaplan–Meier method, and the log-rank test evaluated differences between survival distributions. Cox proportional hazard models were used to calculate hazard ratios for DFS, OS and EFS.[45] Multivariable proportional hazards models were constructed using a backward selection procedure.[45] All statistical analyses using CALGB/Alliance data were performed by the Alliance Statistics and Data Center.

In vitro experiments

LncRNA wild-type and variant transcripts were amplified with Phusion high fidelity polymerase by PCR. Amplicons were cloned into pcDNA using the Gibson technique, according to standard protocols. For primer sequences and further experimental details please see the Data Supplement. K-562 and THP-1 cells were transfected with vectors containing either a cytosine (C)-to-thymidine (T) variant in the lncRNA *SNHG15* (*SNHG15*varT) or wild-type lncRNA *SNHG15* (*SNHG15*wt); cells were also transfected with empty pcDNA3.1 and were used as controls. Cell viability and apoptosis were assessed with Annexin V staining. The colorimetric MTT assay was used to assess the proliferative capacity of the transfected blasts.

Results

Detection of genetic variants in the non-coding transcriptome of younger adults with CN-AML

To examine whether recurrent genetic variants are present in the non-coding transcriptomes of CN-AML patients, we first analyzed total RNA sequencing data of 377 younger adults with CN-AML. In order to identify unequivocally non-coding genetic variants and to avoid ambiguity in their genomic location, we excluded from further analyses all variants, which overlapped with exons of protein-coding genes and those that mapped to segmental duplications or other repeat regions of the genome.

To evaluate the clinical and functional relevance of the lncRNA variants, additional filters were applied. Specifically, we focused on the variants that displayed: 1) adequate expression and coverage in at least 100 samples (approximately 25% of a total number of samples), 2) detection of wild-type genotype in at least 5% of the samples, and 3) detection of variant allele frequency above 0.4 in at least 5% of the samples (Figure 1). Based on these criteria, 981 variants were selected for further analyses (Supplementary Table S1).

Detection of lncRNA variants in the Cancer Genome Atlas (TCGA) dataset

To examine the validity and reproducibility of our experimental pipeline and results, we queried the publicly available TCGA total RNA sequencing (RNA Seq) dataset.[8] It is noteworthy that the TCGA dataset was generated with a different RNA sequencing technique (i.e., poly-A RNA Seq), which is less suitable for the interrogation of the

non-coding fraction of the transcriptome.[46] In addition, a relatively small number of patients in the TCGA cohort represent CN-AML cases (i.e., 44 of the 196 available cases).[8] Despite these limitations, 277 out of the 981 variants that we tested were detectable in the transcriptomes of the CN-AML cases included in the TCGA study (Supplementary Table S2).

For a subset of TCGA cases, DNA sequencing data from both leukemic blasts and germline material are available in addition to the transcriptome sequencing data. We therefore sought: 1) to validate whether the presence of the detected variants in the transcriptome is also detectable at the DNA level and 2) to examine whether these variants are *bona fide* acquired genetic events or are present in the germline configuration of the AML patient genomes. As the sequencing technique that was used to analyze the TCGA dataset (i.e., exome sequencing) preferentially captures and interrogates the coding fraction of the genome, only 20 variant positions were available for analyses at the DNA level. Overall, there was a complete concordance between the detection of a variant in the transcriptome and its detection in the genome. Furthermore, 11 of these variants were detected in both leukemic blasts and non-leukemic tissues and thus could be considered as germline genetic variants, whereas nine variants were only detectable in leukemic samples and could therefore represent acquired mutations (Supplementary Table S2).

Prognostic significance of recurrent lncRNA variants in AML

We proceeded to examine the associations of the lncRNA variants with the clinical outcome of younger patients with CN-AML. Of the variants tested, 41 associated with more than one outcome endpoint (DFS, OS and/or EFS; Supplementary Table S3)

above a threshold level of significance (*P*<0.05). There was no association between the presence of lncRNA variants and complete remission rates in our cohort.

Among the variants that showed significant association with clinical outcome in our dataset were a C-to-T variant in the lncRNA *RP5-1074L1.4* (*RP5-1074L1.4*varT), and *SNHG15*varT, a C-to-T variant in the lncRNA *SNHG15*.

In patients with detectable *RP5-1074L1.4* expression (n=243), the presence of the *RP5-1074L1.4*varT was found in 156 (64%) of them. Patients with the *RP5-1074L1.4*varT had longer DFS (*P*<0.001) than patients with wild-type *RP5-1074L1.4* (*RP5-1074L1.4*wt); five years after diagnosis 44% of the *RP5-1074L1.4*varT patients were alive and leukemia-free in contrast to only 26% of those with *RP5-1074L1.4*wt. *RP5-1074L1.4*varT was also associated with longer EFS (*P*<0.001; 5-year rates: 38% *vs.* 21%) and showed a trend for longer OS (*P*=0.09; 5-year rates: 43% *vs.* 34%; Figures 2A-2C, Supplementary Table S4).

Among patients who expressed the *SNHG15* lncRNA (n=306), *SNHG15*varT was detected in 78% of them (n=239). Patients who expressed the *SNHG15*varT had longer DFS (*P*=0.04; 5-year rates: 37% *vs.* 22%) than patients who expressed the *SNHG15*wt. *SNHG15*varT expressers also had longer EFS (*P*=0.04; 5-year rates: 31% *vs.* 19%) and a trend for longer OS (*P*=0.07; 5-year rates: 41% *vs.* 32%) compared with *SNHG15*wt expressers (Figures 3A-3C, Supplementary Table S5).

Finally, we examined whether genetic variants were detectable in the set of 24 lncRNAs, whose expression levels were previously shown to associate with the clinical outcome of younger adults with CN-AML.[30] We found seven such variants in

three of the prognostic lncRNAs (annotated in bold lettering in Supplementary Table S1). A guanosine (G)-to-C variant in the lncRNA *AL122127.25* (*AL122127.25*varC) was the only one that associated with patient outcome. *AL122127.25*varC was detected in 72 of the 257 patients who expressed the *AL122127.25* lncRNA (i.e., 28% of the *AL122127.25* expressers). The presence of *AL122127.25*varC significantly associated with shorter DFS (*P*=0.01; 5-year rates: 17% *vs.* 35%), OS (*P*=0.01; 5-year rates: 22% *vs.* 40%) and EFS (*P*=0.002; 5-year rates: 12% *vs.* 30%; Supplementary Figures S1A-S1C, Supplementary Table S6) in younger adult CN-AML patients. Notably, the presence of *AL122127.25*varC had no impact on the expression levels of the *AL122127.25* transcript, when compared to the *AL122127.25*wt (Supplementary Figure S2).

Associations of lncRNA variants with pretreatment clinical and molecular characteristics

Next, we examined potential associations of *RP5-1074L1.4*varT, *SNHG15*varT and *AL122127.25*varC with pretreatment clinical characteristics and prognostic molecular features of younger adult CN-AML patients. Overall, there were only minor differences in these features between variant and wild-type lncRNA expressers. Patients expressing the *RP5-1074L1.4*varT were more likely to harbor tyrosine kinase domain mutations of the *FLT3* gene (*FLT3*-TKD) than patients who expressed the *RP5-1074L1.4*wt (*P*=0.03; 12% *vs.* 4%; Supplementary Table S7). Patients expressing the *SNHG15*varT were older (*P*=0.03), more likely to be Caucasian (*P*=0.02) and more likely to have low expression of the *MN1* gene (*P*=0.05; Supplementary Table S8) than *SNHG15*wt expressers. With regard to the lncRNA *AL122127.25*, there were no significant differences in the clinical features or

frequencies of prognostic gene mutations or gene expression between *AL122127.25*varC and *AL122127.25*wt expressers (Supplementary Table S9).

Multivariable analyses

To examine prognostic significance of the detected lncRNA variants in the context of other established clinical and molecular prognostic markers, we constructed multivariable models.

Expression of *RP5-1074L1.4*varT was a significant marker for longer DFS [hazard ratio (HR): 0.54; *P*<0.001] and longer EFS (HR: 0.54; *P*<0.001) after adjusting for other covariates (Table 1). Expression of *SNHG15*varT significantly associated with longer DFS (HR: 0.63; *P*=0.02), longer OS (HR: 0.63; *P*=0.008) and longer EFS (HR: 0.68; *P*=0.02) after adjusting for other variables for each outcome endpoint (Table 2). Finally, expression of the *AL122127.25*varC was an independent marker of shorter OS (HR: 1.57; *P*=0.009*)* and shorter EFS (HR: 1.59; *P*=0.004) after adjusting for other covariates (Supplementary Table S10).

RP5-1074L1.4 is an intronic lncRNA, which is embedded in intron 7 of the protein-coding *SLC16A4* transcript. To ensure that the prognostic effect of *RP5-1074L1.4*varT was not due to perturbation of the expression levels of *SLC16A4*, we compared the *SLC16A4* transcript abundance between 87 patients who expressed the *RP5-1074L1.4*wt and 156 patients who expressed the *RP5-1074L1.4*varT. We found no significant difference in expression levels of *SLC16A4* between these two patient groups (Supplementary Figure S3).

Evaluation of the prognostic significance of lncRNA variants in an independent cohort of CN-AML patients

To examine whether our findings were reproducible in an independent cohort of CN-AML patients, we examined patients treated on AMLCG protocols,[38,39] who had available clinical outcome data, and were analyzed with poly-A selected RNAseq (n=135). As was the case with the TCGA dataset,[8] the use of an alternative RNA sequencing technique limited the number of lncRNA variants that could be detected and analyzed. The lncRNA transcripts, whose variants showed the strongest association with prognosis in our initial cohort were not captured by the poly-A RNAseq and could not, therefore, be analyzed for associations with clinical outcome. Despite this limitation, there was concordance in the findings between the two cohorts. *SNHG15*varT was the one detectable lncRNA variant, which associated with multiple outcome endpoints in the AMLCG dataset. Specifically, the *SNHG15*varT was detected in 103 of the 120 AMLCG CN-AML patients who expressed the *SNHG15* lncRNA. In agreement with our findings in the CALGB/Alliance dataset, the presence of *SNHG15*varT associated with longer DFS (*P*=0.04; Figure 3D) and EFS (*P*=0.007, Figure 3E), but not OS (*P*=0.17), in the AMLCG cohort.

Expression levels of *SNHG15* in normal hematopoiesis

To further examine the functional significance of the lncRNAs, which harbor prognostic variants we sought to determine their expression patterns during normal hematopoiesis. To this end, we used publicly available datasets of normal hematopoietic cells analyzed with microarrays or RNA seq and deposited in the BloodSpot portal (www.bloodspot.eu). Of the lncRNAs with prognostic genetic variants, only *SNHG15* was annotated in the database and could be further analyzed. We found that *SNHG15* was most abundantly expressed in common

myeloid progenitors, granulocyte monocyte progenitors and megakaryocyte-erythroid progenitors. It was also highly expressed in hematopoietic stem cells and lymphoid cell populations. Among mature cell populations, *SNHG15* was overexpressed in monocytes, whereas its expression levels were the lowest in polymorphonuclear leucocytes of the bone marrow and peripheral blood (Supplementary Figure S4).

Functional relevance of prognostic lncRNA variants

We hypothesized that in addition to associations with outcome, the presence of genetic variants in lncRNAs could have a functional impact and affect AML blast viability and proliferation. We focused on the lncRNA *SNHG15*varT, whose prognostic significance was validated in an independent cohort of AML patients. We isolated total RNA from AML cell lines and amplified both the *SNHG15*wt and *SNHG15*varT transcripts. We cloned the amplicons into pcDNA3.1 expression vectors and transfected two AML cell lines (i.e. K-562 and THP-1 cells). Expression levels of the *SNHG15* lncRNA were similar in cells transfected with the *SNHG15*wt and those transfected with *SNHG15*varT-containing vectors, when compared with the empty vector controls (Figures 4A and 4B). Regarding cell viability, ectopic overexpression of the *SNHG15*wt and the *SNHG15*varT led to a discreet but consistent decrease in cell viability across cell lines, which did not reach statistical significance (Figures 4C and 4D).

To further evaluate the effect of *SNHG15* on the growth kinetics of the AML cells, we performed colorimetric MTT assays. Forced overexpression of the *SNHG15*varT had no significant effect on blast growth when compared with controls. In contrast, overexpression of the *SNHG15*wt led to increased proliferative capacity of the leukemic cells when compared with both controls and cells overexpressing the

*SNHG15*varT (Figures 4E and 4F). These findings are in line with our prognostic observations and the inferior outcome of CN-AML patients who express *SNHG15*wt, compared with those who express *SNHG15*varT.

Discussion

Long non-coding RNAs are gaining recognition as important molecular mediators and regulators of key cellular functions in health and disease.[22-29] In AML, lncRNAs have been shown to associate with the clinical outcome of both younger and older patients.[30-33] However, these previous studies have focused on the expression levels of the lncRNA transcripts; the effect of genetic variants within lncRNAs has not been extensively studied. Accumulating evidence suggests that such genetic variants could impact on the function of the lncRNAs and be relevant for disease pathogenesis.[47] In support of this view, disease-associated SNPs are more frequently found in regions of the genome that encode for non-coding RNA transcripts in several types of solid tumors[34,35] and in AML.[34,36] These findings suggest that the presence of a variant in the non-coding transcriptome could be the functional link that explains how genetic variants that do not alter the structure of protein molecules associate with malignant phenotypes. Recently, Klco et al.[37] have reported the acquisition of somatic mutations in lncRNA transcripts of AML patients, as demonstrated by analyses of the leukemic blasts in parallel with germline material. However, the prognostic value of these mutations could not be tested due to sample size limitations. Herein, we analyzed total RNAseq data of younger adult patients with CN-AML with the goal to detect recurrent lncRNA variants and evaluate their prognostic and biologic significance.

We used a stringent approach for detecting and filtering sequence variations of lncRNAs. We generated a list of 981 recurrent variants, which are located within lncRNAs in younger adult patients constituting the CALGB/Alliance cohort, and which had adequate sample sizes for meaningful survival analyses. To study potential associations of the detected variants with clinical outcome, we individualized analysis for each variant and limited comparisons to the group of patients that were expressers of the corresponding lncRNA transcript (i.e., we compared the expressers of a variant to the expressers of the wild-type lncRNA). Of the 981 candidate variants, a subset of 41 significantly associated with at least two clinical outcome endpoints of younger adults with CN-AML. LncRNA genetic variants *RP5-1074L1.4*varT and *SNHG15*varT were significantly associated with prognosis. *RP5-1074L1.4* has not been previously associated with cancer pathogenesis or clinical outcome of cancer patients. In contrast, *SNHG15*, a MYC-regulated lncRNA, which harbors a small nucleolar RNA in one of its introns, has been implicated in the pathogenesis of multiple solid malignancies. *SNHG15* has been shown to interact with the protein AIF and to associate with a clinically aggressive and prognostically unfavorable subset of colorectal carcinomas.[48] *SNHG15* has also been associated with aggressive phenotypes of hepatocellular and breast carcinomas via sponging and inhibiting the function of microRNAs 141-3p and 211-3p.[49,50]

Patients who expressed the *RP5-1074L1.4*varT had more favorable outcome (i.e., longer DFS, EFS and a trend for longer OS) than *RP5-1074L1.4*wt expressers. *SNHG15*varT also associated with better prognosis and *SNHG15*varT expressers had longer DFS and EFS compared with patients who expressed the *SNHG15*wt lncRNA.

Further, we examined whether recurrent genetic variants could be detected in lncRNAs previously reported to be prognostic in younger adults with CN-AML.[30] A G-to-C variant in the prognostic lncRNA *AL122127.25* (*AL122127.25*varC) significantly associated with clinical outcome of CN-AML patients. Specifically, expression of the *AL122127.25*varC associated with shorter DFS, OS and EFS. Notably, the expression levels of the *AL122127.25* lncRNA were not significantly affected by the presence of the *AL122127.25*varC. Thus, both the abundance and the nucleotide sequence of the *AL122127.25* lncRNA associate with the clinical outcome of CN-AML patients via potentially independent mechanisms.

Currently, there is limited availability of AML datasets analyzed with RNA sequencing techniques (i.e., with total RNAseq) that are suitable for in-depth analyses of the non-coding transcriptome. To examine whether our observations could be reproduced in other datasets, we examined the presence of our curated variant list in an independent cohort of CN-AML patients who were treated on AMLCG protocols. As expected, the use of a different RNA sequencing technique significantly limited the number of lncRNA transcripts that we could interrogate. Nevertheless, *SNHG15*varT was detectable and associated with longer DFS and EFS of the AMLCG patients, similarly to its prognostic effect in the CALGB/Alliance dataset.

Given the constantly increasing number of prognostic molecular alterations in AML, it is important to examine the prognostic value of novel markers in the context of other established clinical and molecular prognosticators. It is thus noteworthy that the lncRNA variants that we examined (*RP5-1074L1.4*varT, *SNHG15*varT, and *AL122127.25*varC) did not show associations with gene mutations that are currently used for the risk stratification of the treatment of AML patients.[11] Mutations in the

CEBPA, *RUNX1*, *ASXL1* and *NPM1* genes as well as the presence of the *FLT3*-ITD were similarly distributed between expressers of the lncRNA variants and expressers of the wild-type transcripts. In addition, in formal multivariable analyses that included prognostic clinical and molecular parameters, *RP5-1074L1.4*varT, *SNHG15*varT, and *AL122127.25*varC retained their prognostic significance after adjusting for other covariates. Consequently, detection of these variants could provide additional prognostic information and further refine risk-stratification of CN-AML patients.

In addition to testing their prognostic significance, we sought to evaluate whether the presence of recurrent variants in lncRNA sequences has functional implications. We focused on *SNHG15* lncRNA and performed overexpression experiments with *SNHG15*wt- and *SNHG15*varT-containing vectors in the K-562 and THP-1 cell lines. Overexpression of the *SNHG15*varT had no significant effect on cell viability and no evident impact on cell proliferation, compared with controls. In contrast, overexpression of the *SNHG15*wt led to a significant increase in the proliferative capacity of the leukemic blasts in both cell lines that were tested. Despite the limitations of *in vitro* assays, these results indicate that expression of *SNHG15*wt associates with a more aggressive disease phenotype, when compared with *SNHG15*varT, and are in line with the inferior clinical outcome of *SNHG15*wt-expressers.

In summary, we have performed comprehensive characterization of genetic variants within lncRNAs in younger adults with CN-AML. We present analyses that support the prognostic and potential biologic significance of lncRNA variants in CN-AML. We believe that our work will serve as a useful starting point for further studies on the role of genetic variants in the non-coding transcriptome of cancer patients.

References

1. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. N Engl J Med. 2015;373(12):1136-1152.
2. 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.
3. 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.
4. 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.
5. Mrózek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev. 2004;18(2):115-136.
6. 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.
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. Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-2074.

9. 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.

10. Eisfeld A-K, Kohlschmidt J, Mrózek K, et al. Mutation patterns identify adult patients with de novo acute myeloid leukemia aged 60 years or older who respond favorably to standard chemotherapy: an analysis of Alliance studies. Leukemia. 2018;32(6):1338-1348.

11. Mrózek K, Marcucci G, Paschka P, Whitman SP, Bloomfield CD. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? Blood. 2007;109(2):431-448.

12. Valk PJM, Verhaak RGW, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med. 2004;350(16):1617-1628.

13. 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.

14. 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.

15. 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.

16. 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.

17. Welch JS, Ley TJ, Link DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012;150(2):264-278.

18. Klco JM, Spencer DH, Miller CA, et al. Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. Cancer Cell. 2014;25(3):379-392.

19. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

20. Kellis M, Wold B, Snyder MP, et al. Defining functional DNA elements in the human genome. Proc Natl Acad Sci U S A. 2014;111(17):6131-6138.

21. Taylor J. Clues to function in gene deserts. Trends Biotechnol. 2005;23(6):269-271.

22. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. Annu Rev Biochem. 2012;81:145-166.

23. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. Nature. 2012;482(7385):339-346.

24. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell. 2011;43(6):904-914.

25. 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.

26. Leucci E, Vendramin R, Spinazzi M, et al. Melanoma addiction to the long non-coding RNA *SAMMSON*. Nature. 2016;531(7595):518-522.

27. 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.

28. Papaioannou D, Petri A, Dovey OM, et al. The long non-coding RNA *HOXB-AS3* regulates ribosomal RNA transcription in *NPM1*-mutated acute myeloid leukemia. Nat Commun. 2019;10(1):5351.

29. Ji P, Diederichs S, Wang W, et al. MALAT-1, a novel noncoding RNA, and thymosin *β*4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene. 2003;22(39):8031-8041.

30. Papaioannou D, Nicolet D, Volinia S, et al. Prognostic and biologic significance of long non-coding RNA profiling in younger adults with cytogenetically normal acute myeloid leukemia. Haematologica. 2017;102(8):1391-1400.

31. 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.

32. Beck D, Thoms JAI, Palu C, et al. A four-gene LincRNA expression signature predicts risk in multiple cohorts of acute myeloid leukemia patients. Leukemia. 2018;32(2):263-272.

33. Mer AS, Lindberg J, Nilsson C, et al. Expression levels of long non-coding RNAs are prognostic for AML outcome. J Hematol Oncol. 2018;11(1):52.

34. Yan X, Hu Z, Feng Y, et al. Comprehensive genomic characterization of long non-coding RNAs across human cancers. Cancer Cell. 2015;28(4):529-540.

35. Iyer MK, Niknafs YS, Malik R, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015;47(3):199-208.

36. Schwarzer A, Emmrich S, Schmidt F, et al. The non-coding RNA landscape of human hematopoiesis and leukemia. Nat Commun. 2017;8(1):218.

37. Klco JM, Miller CA, Griffith M, et al. Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. JAMA. 2015;314(8):811-822.

38. 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.

39. Braes 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. Nekrutenko A, Taylor J. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. Nat Rev Genet. 2012;13(9):667-672.

45. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression methods in biostatistics: linear, logistic, survival and repeated measures models. New York, NY: Springer; 2005.

46. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics. 2014;15(1):419.

47. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. Nat Rev Genet. 2016;17(2):93-108.

48. Saeinasab M, Bahrami AR, González J, et al. *SNHG15* is a bifunctional MYC-regulated noncoding locus encoding a lncRNA that promotes cell proliferation, invasion and drug resistance in colorectal cancer by interacting with AIF. J Exp Clin Cancer Res. 2019;38(1):172.

49. Ye J, Tan L, Fu Y, et al. LncRNA SNHG15 promotes hepatocellular carcinoma progression by sponging miR-141-3p. J Cell Biochem. 2019;120(12):19775-19783.

50. Kong Q, Qiu M. Long noncoding RNA SNHG15 promotes human breast cancer proliferation, migration and invasion by sponging miR-211-3p. Biochem Biophys Res Commun. 2018;495(2):1594-1600.

Table 1. Multivariable analyses of outcome in 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).

Variables in final models	DFS		EFS	
	HR (95% CI)	***P***	**HR (95% CI)**	***P***
RP5-1074L1.4, varT *vs.* wild-type	0.54 (0.37-0.77)	<0.001	0.54 (0.40-0.75)	<0.001
FLT3-ITD, present *vs.* absent	2.29 (1.59-3.30)	<0.001	2.16 (1.54-3.02)	<0.001
WT1, mutated *vs.* wild-type	1.85 (1.04-3.27)	0.04	1.90 (1.20-3.02)	0.006
MN1 expression, high *vs.* low	1.55 (1.07-2.24)	0.02	-	-
CEBPA, double-mutated *vs.* wild-type	-	-	0.36 (0.21-0.61)	0.001
NPM1, mutated *vs.* wild-type	-	-	0.41 (0.28-0.58)	<0.001

DFS: disease-free survival; EFS: event-free survival; *FLT3*-ITD: internal tandem duplication of the *FLT3* gene; HR: hazard ratio; CI: confidence intervals; varT: C-to-T variant; *vs.*: versus.

NOTE: 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 for the categorical variables. Variables considered for model inclusion were: *RP5-1074L1.4* (varT *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 expression (expressed *vs.* not expressed), and miR-155 expression levels (high *vs.* low).

Table 2. Multivariable analyses of outcome in 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).

Variables in final models	DFS		OS		EFS	
	HR (95% CI)	***P***	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
SNHG15, varT *vs.* wild-type	0.63 (0.34-0.81)	0.02	0.63 (0.45-0.89)	0.008	0.68 (0.22-0.57)	0.02
FLT3-ITD, present *vs.* absent	1.68 (1.08-2.60)	0.02	2.48 (1.83-3.35)	<0.001	2.08 (1.57-2.76)	<0.001
WT1, mutated *vs.* wild-type	1.91 (1.01-3.55)	0.04	1.78 (1.21-2.63)	0.004	-	-
MN1 expression, high *vs.* low	1.51 (1.04-2.20)	0.03	-	-	-	-
DNMT3A, mutated *vs.* wild-type	2.11 (1.43-3.12)	<0.001	-	-	-	-
miR-155 expression, high *vs.* low	1.85 (1.18-2.91)-	0.007	-	-	-	-
Extramedullary disease, Present *vs.* absent	0.61 (0.39-0.95)	0.03	-	-	-	-
Hemoglobin, continuous, 1-unit increments	0.87 (0.79-0.96)-	0.004	0.89 (0.83-0.96)	0.004	0.89 (0.83-0.95)	0.001
Age, continuous, 10 year increments	-	-	1.29 (1.12-1.48)	<0.001	-	-
NPM1, mutated *vs.* wild-type	-	-	0.51 (0.37-0.71)-	-	0.45 (0.34-0.62)	<0.001
CEBPA, double-mutated *vs.* wild-type	-	-	0.49 (0.30-0.80)	0.005	0.35 (0.22-0.57)	<0.01

DFS: disease-free survival; OS: overall survival; EFS: event-free survival; *FLT3*-ITD: internal tandem duplications of the *FLT3* gene; HR: hazard ratio; CI: confidence intervals; varT: C-to-T variant; *vs.*: versus.

NOTE: 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 for the

categorical variables. Variables considered for model inclusion were *SNHG15* (varT *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 [(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 expression (expressed *vs.* not expressed), and miR-155 expression levels (high *vs.* low).

FIGURE LEGENDS

Figure 1. Outline of the two-pass experimental approach for the identification of recurrent genetic variants located within long non-coding RNAs in younger adult patients with CN-AML. In the first pass, variant calling was performed on alignment results (i.e., BAM files) following the Genome Analysis Toolkit (GATK) best practice recommendations for RNA-sequencing datasets. Variants of non-coding transcripts that do not overlap with coding exons and are not located in low-complexity regions of the genome were selected. In the second pass, Samtools pile-up programs were used to identify sequencing depth, quality and alternative allele counts on selected unique variant positions. Resulting visual component framework (VCF) files were consolidated with annotation and the final variant call matrix was generated.

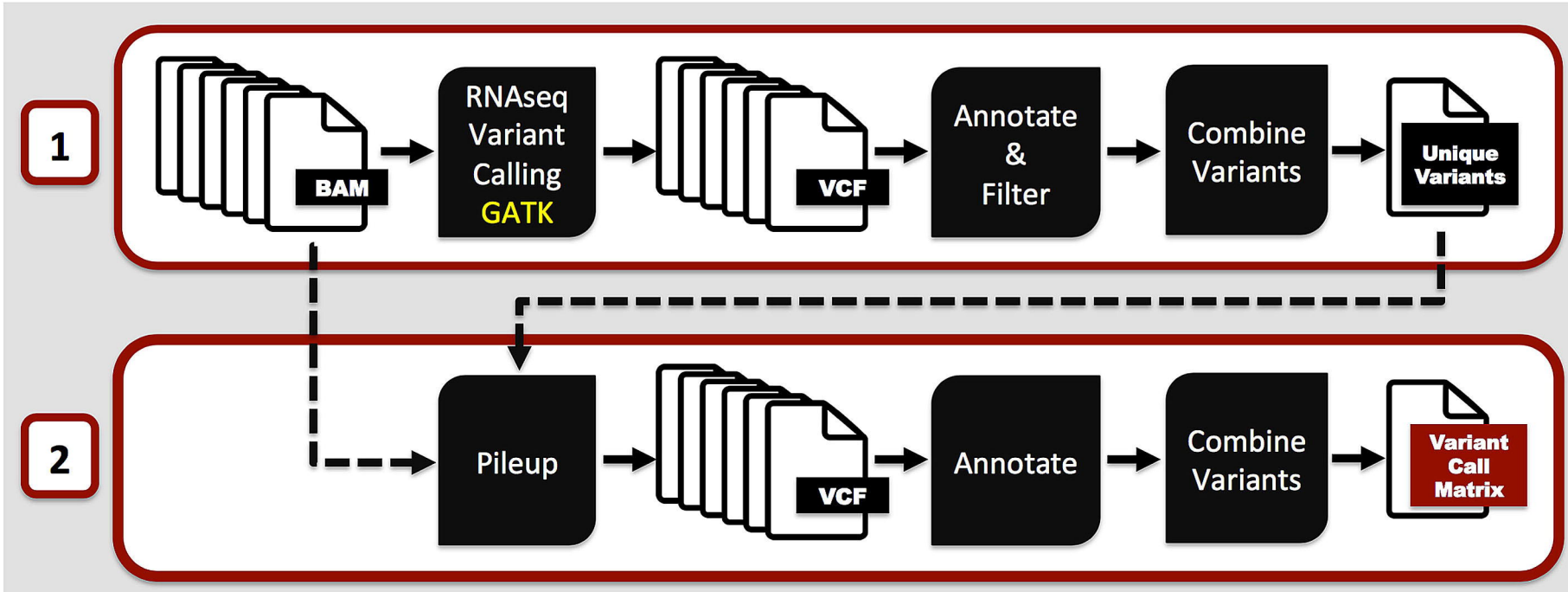
Figure 2. Outcome of younger adult patients with CN-AML who harbored the C-to-T variant of the *RP5-1074L1.4* long non-coding RNA (*RP5-1074L1.4*varT) and of those who had the wild-type lncRNA (*RP5-1074L1.4*wt). (A) Disease-free survival, (B) overall survival and (C) event-free survival.

Figure 3. Prognostic significance of long non-coding RNA variants across different CN-AML cohorts. (A) Disease-free survival, (B) overall survival and (C) event-free survival of younger adult CN-AML patients in the CALGB/Alliance cohort who had the C-to-T variant of the *SNHG15* long non-coding RNA (*SNHG15*varT) and of patients with the wild-type *SNHG15* lncRNA (*SNHG15*wt). (D) Disease-free survival and (E) event-free survival of CN-AML patients in the AMLCG cohort with the

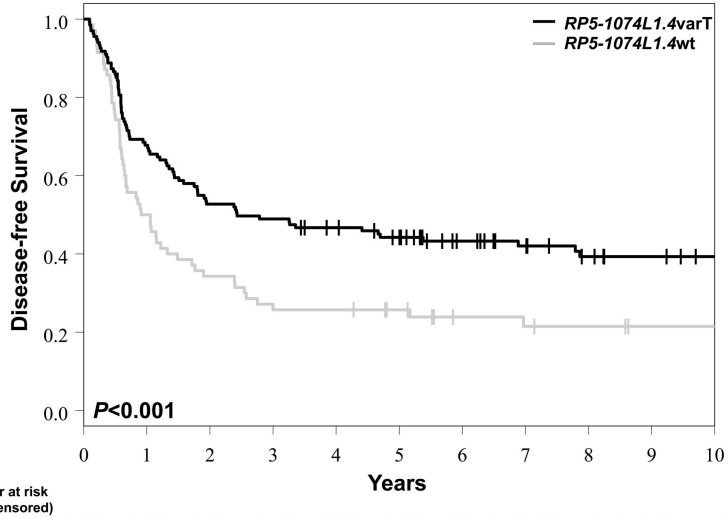
C-to-T variant of the *SNHG15* long non-coding RNA (*SNHG15*varT) and of those the wild-type *SNHG15* lncRNA (*SNHG15*wt).

Figure 4. Biologic significance of recurrent lncRNA variants in younger adult CN-AML patients. Fold changes of *SNHG15* lncRNA expression levels (A, B), percent of viable cells (C, D) and proliferation (E, F) of AML cell lines transfected with empty pcDNA3, *SNHG15*wt-containing or *SNHG15*varT-containing vectors. Results for K-562 (A, C, E) and THP-1 (B, D, F) cells are depicted. Proliferation is assessed with the MTT colorimetric assay, by light absorbance. * indicates *P*-value<0.05, ** indicates *P*-value<0.01, *** indicates *P*-value<0.001, N.S indicates not significant.

Figure 1

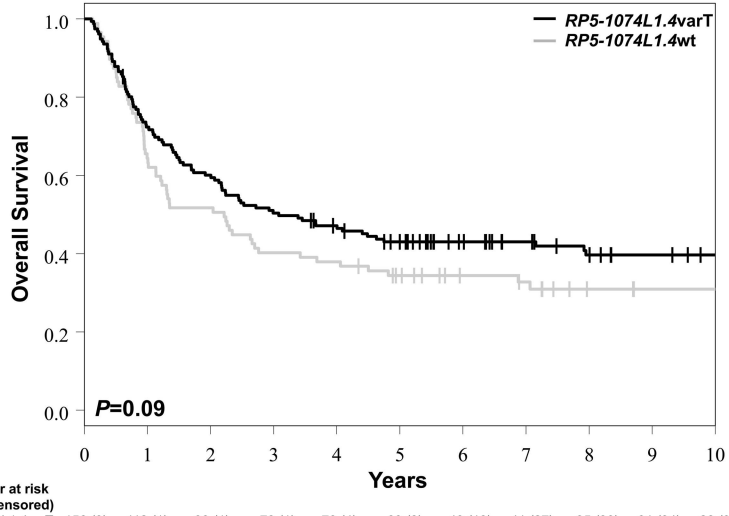

1
BAM
RNAseq
Variant
Calling
GATK
VCF
Annotate
&
Filter
Combine
Variants
Unique
Variants
2
Pileup
VCF
Annotate
Combine
Variants
Variant
Call
Matrix

2A



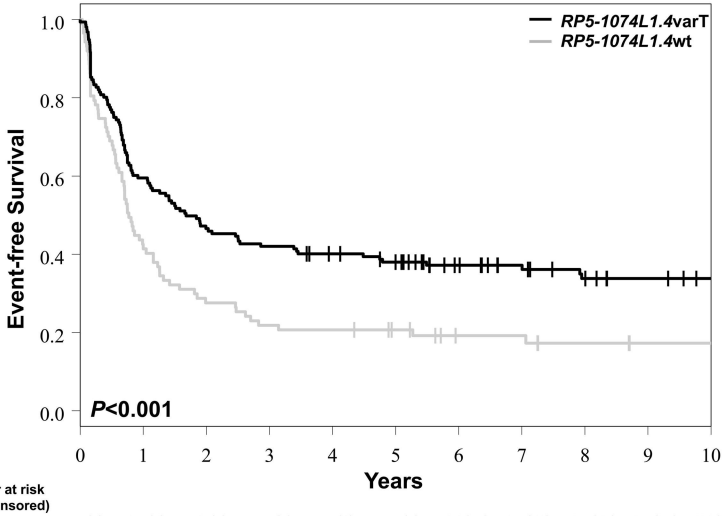
Number at risk (number censored)											
*RP5-1074L1.4*varT	134 (0)	90 (1)	70 (1)	65 (1)	59 (4)	53 (7)	41 (18)	34 (24)	28 (28)	25 (31)	22 (34)
*RP5-1074L1.4*wt	70 (0)	35 (0)	24 (0)	18 (0)	18 (0)	15 (3)	10 (7)	9 (7)	7 (9)	5 (11)	5 (11)

2B



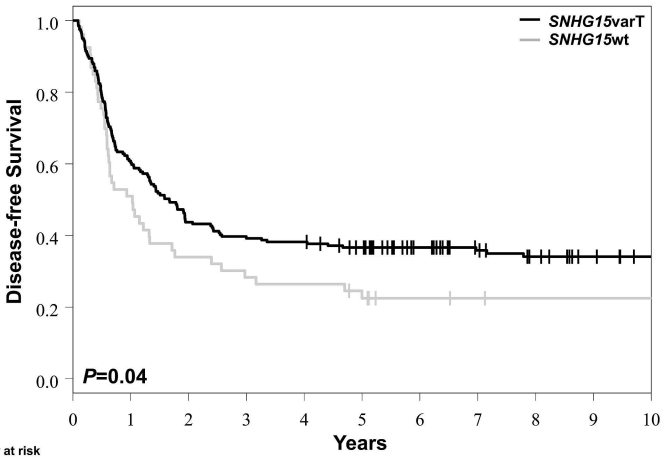
Number at risk (number censored)											
*RP5-1074L1.4*varT	156 (0)	112 (1)	93 (1)	78 (1)	70 (4)	60 (8)	49 (19)	41 (27)	35 (30)	31 (34)	28 (37)
*RP5-1074L1.4*wt	87 (0)	56 (0)	45 (0)	35 (0)	33 (0)	27 (3)	21 (9)	18 (11)	12 (16)	10 (18)	10 (18)

2C



Number at risk (number censored)											
*RP5-1074L1.4*varT	156 (0)	92 (1)	72 (1)	65 (1)	59 (4)	53 (7)	42 (17)	35 (24)	29 (27)	25 (31)	22 (34)
*RP5-1074L1.4*wt	87 (0)	36 (0)	24 (0)	19 (0)	18 (0)	15 (3)	10 (7)	10 (7)	7 (9)	5 (11)	5 (11)

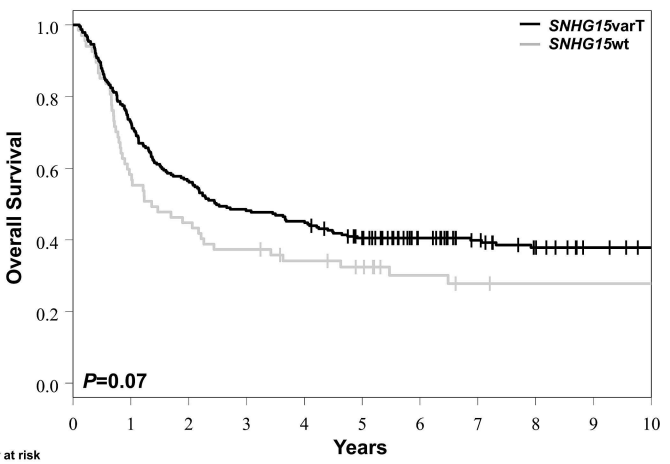
A



Number at risk (number censored)

*SNHG15*varT	199 (0)	120 (0)	87 (0)	78 (0)	76 (0)	68 (5)	54 (19)	44 (28)	37 (33)	31 (39)	27 (43)
*SNHG15*wt	53 (0)	27 (0)	18 (0)	15 (0)	14 (0)	11 (1)	8 (4)	7 (5)	6 (6)	6 (6)	6 (6)

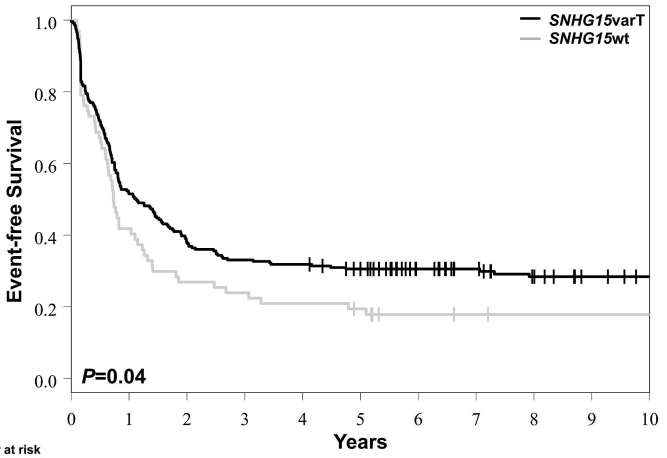
B



Number at risk (number censored)

*SNHG15*varT	239 (0)	174 (0)	135 (0)	115 (0)	108 (0)	91 (6)	74 (23)	62 (34)	52 (41)	44 (49)	40 (53)
*SNHG15*wt	67 (0)	39 (0)	30 (0)	25 (0)	21 (2)	18 (4)	13 (8)	11 (9)	10 (10)	10 (10)	10 (10)

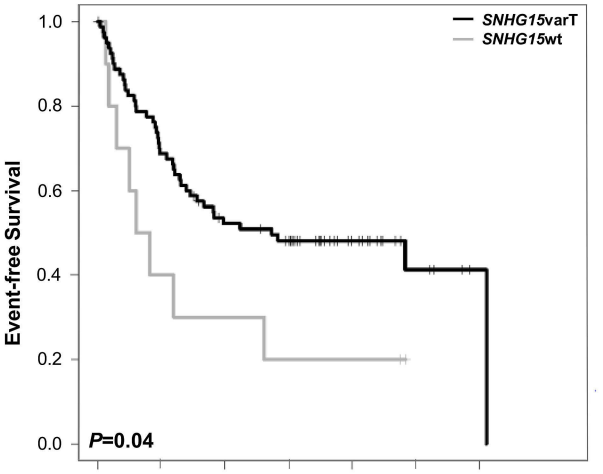
C



Number at risk (number censored)

*SNHG15*varT	239 (0)	123 (0)	91 (0)	79 (0)	76 (0)	68 (5)	54 (19)	46 (27)	38 (32)	31 (39)	27 (43)
*SNHG15*wt	67 (0)	28 (0)	18 (0)	16 (0)	14 (0)	12 (1)	8 (4)	7 (5)	6 (6)	6 (6)	6 (6)

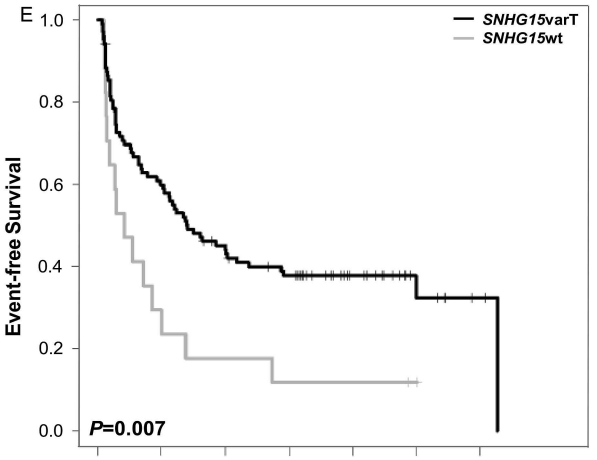
D



Number at risk

*SNHG15*varT	81	55	39	33	16	6	2
*SNHG15*wt	10	4	3	2	2	1	0

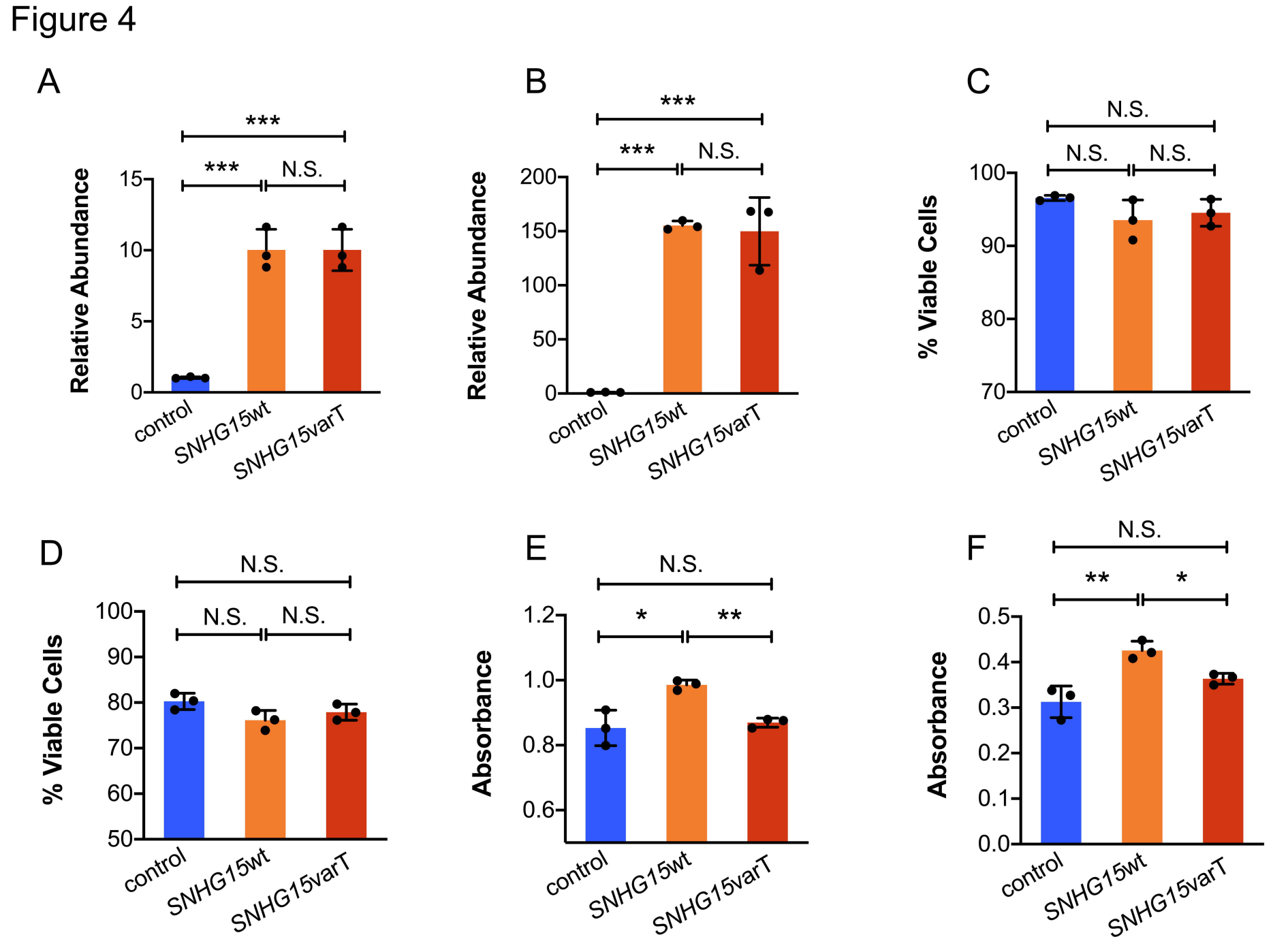
E



Number at risk

*SNHG15*varT	103	61	43	35	18	6	2
*SNHG15*wt	17	4	3	2	2	1	0

Figure 4


A
Relative Abundance
15
10
5
0
***
***
N.S.
control
SNHG15wt
SNHG15varT
B
Relative Abundance
200
150
100
50
0
***
***
N.S.
control
SNHG15wt
SNHG15varT
C
% Viable Cells
100
90
80
70
N.S.
N.S.
N.S.
control
SNHG15wt
SNHG15varT
D
% Viable Cells
100
90
80
70
60
50
N.S.
N.S.
N.S.
control
SNHG15wt
SNHG15varT
E
Absorbance
1.2
1.0
0.8
0.6
N.S.
*
**
control
SNHG15wt
SNHG15varT
F
Absorbance
0.5
0.4
0.3
0.2
0.1
0.0
N.S.
**
*
control
SNHG15wt
SNHG15varT

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/m² twice daily on days 3-9, cytarabine 100 mg/m²/d continuous infusion on days 1 and 2 and 100 mg/m² twice daily on days 3-8, and daunorubicin 60 mg/m² on days 3-5) followed by one cycle of HAM (cytarabine 3 g/m² twice daily on days 1-3 and mitoxantrone 10 mg/m² 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/m² (1 g/m² in patients ≥60 years) twice daily on days 1, 2, 8 and 9 and mitoxantrone 10 mg/m² 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% CO2.

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 Median, years % Disease-free at 5 years (95% CI)	 2.4 44 (36-52)	 1.0 26 (16-36)	<0.001	0.56 (0.40-0.79)
Overall survival Median, years % Alive at 5 years (95% CI)	 3.1 43 (35-51)	 2.2 34 (25-44)	0.09	0.76 (0.55-1.05)
Event-free survival Median, years % Event-free at 5 years (95% CI)	 1.7 38 (30-46)	 0.8 21 (13-30)	<0.001	0.59 (0.44-0.80)

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 Median, years % Disease-free at 5 years (95% CI)	 0.7 17 (8-28)	 1.6 35 (27-43)	0.01	1.57 (1.11-2.24)
Overall survival Median, years % Alive at 5 years (95% CI)	 1.1 22 (17-25)	 2.2 40 (32-47)	0.01	1.49 (1.08-2.04)
Event-free survival Median, years % Event-free at 5 years (95% CI)	 0.7 12 (6-21))	 1.1 30 (23-37)	0.002	1.60 (1.19-2.16)

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.

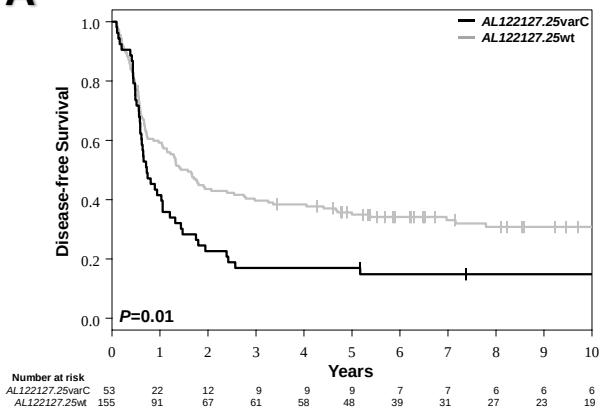
Supplementary Figure S2. Correlation between the frequency of the G-to-C variant and the expression levels of the long non-coding RNA *AL122127.25*. Distance of correlation (dcor) is used as a metric for evaluating the strength of correlation between the allele frequency of the variant and the abundance 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).

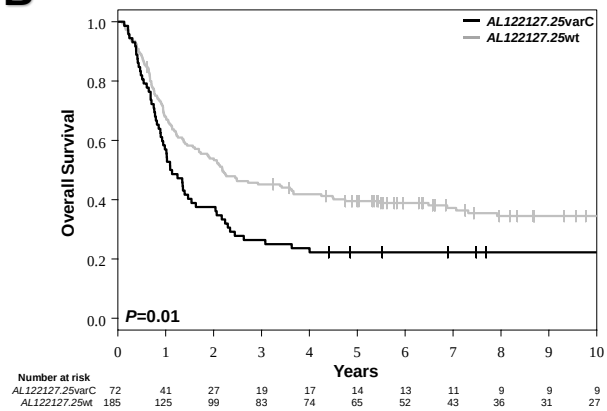
Supplementary Figure S4. Expression levels of *SNHG15* in subpopulations of stem cells, progenitor and mature blood cells during healthy hematopoiesis. The image was generated using publicly available data in the Bloodspot portal (www.bloodspot.eu). In the figure, HSC indicates hematopoietic stem cells; 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

A



B



C

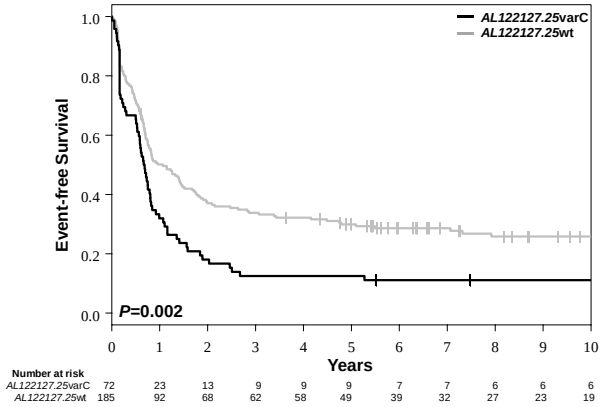


Figure S2

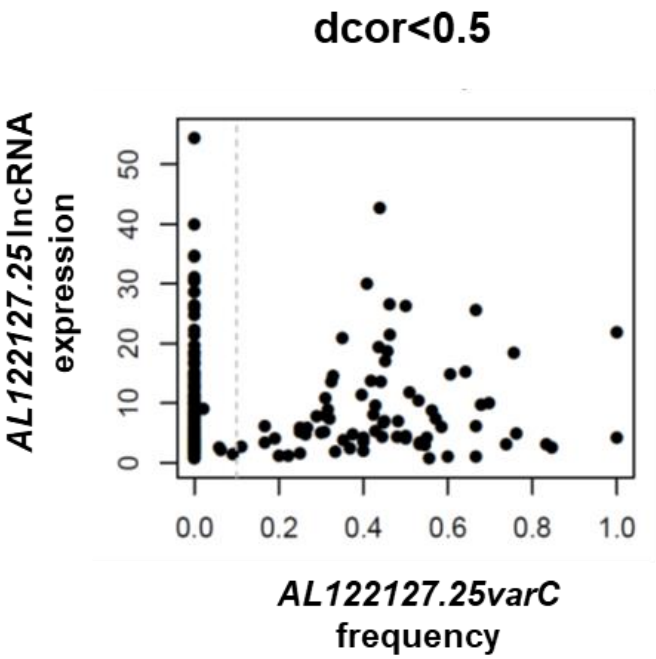

dcor<0.5
AL122127.25 lncRNA expression
0
10
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0
AL122127.25varC frequency

Figure S3

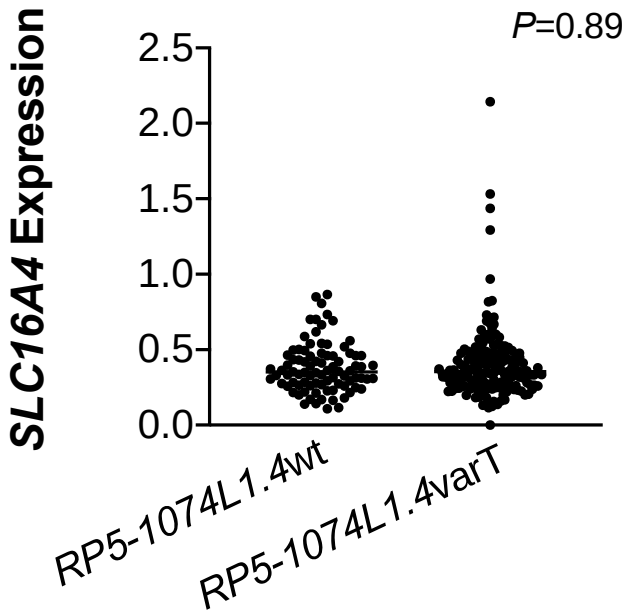
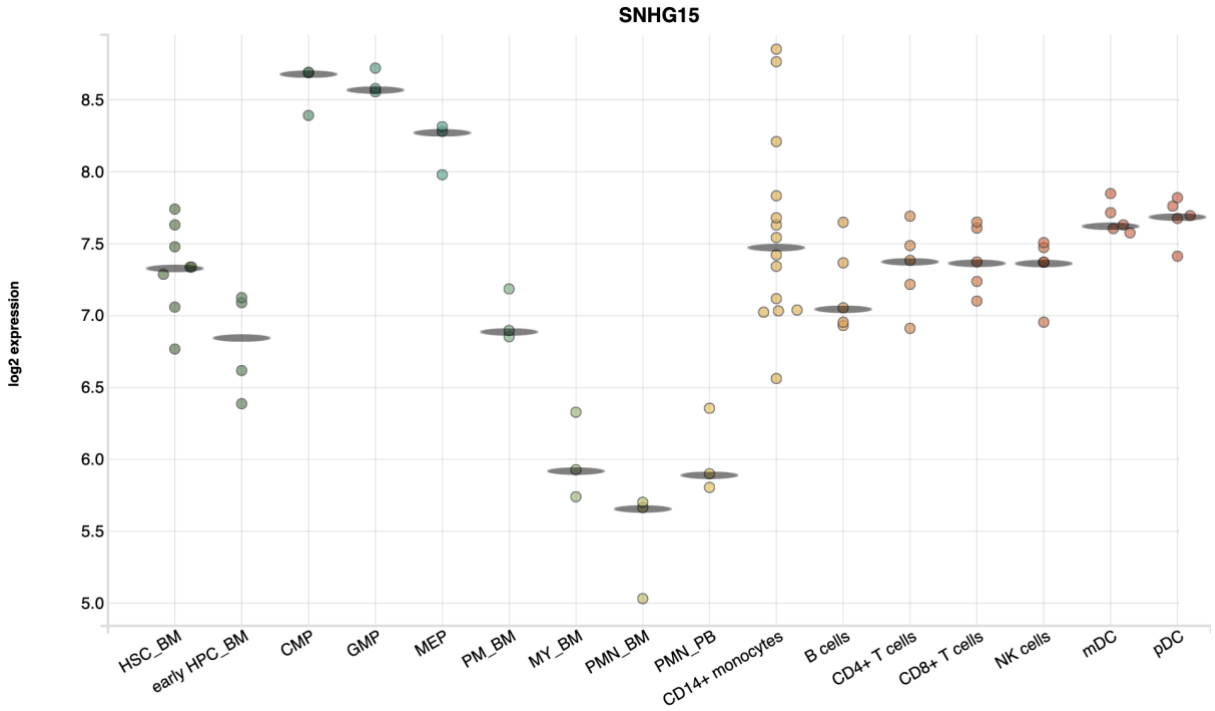

SLC16A4 Expression
P=0.89
2.5
2.0
1.5
1.0
0.5
0.0
RP5-1074L1.4wt
RP5-1074L1.4varT

Figure S4


SNHG15
log2 expression
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
HSC_BM
early HPC_BM
CMP
GMP
MEP
PM_BM
MY_BM
PMN_BM
PMN_PB
CD14+ monocytes
B cells
CD4+ T cells
CD8+ T cells
NK cells
mDC
pDC