ARPP19 promotes MYC expression and associates with patient relapse in acute myeloid leukemia

Eleonora Mäkelä[1,2,3], Eliisa Löyttyniemi[4], Urpu Salmenniemi[5], Otto Kauko[1], Taru Varila[1], Veli Kairisto[6], Maija Itälä-Remes[5] and Jukka Westermarck[1,2#]

[1]Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland; [2]Institute of Biomedicine, University of Turku, Turku, Finland; [3]Turku Doctoral Programme of Molecular Medicine, Turku, Finland; [4]Department of Biostatistics, University of Turku, Turku, Finland; [5]Department of Hematology, Turku University Hospital (TYKS), Turku, Finland; [6]Central Laboratory, Turku University Hospital (TYKS), Turku, Finland.

Running title: ARPP19 is a predictive AML relapse marker

Keywords: Cancer, ARPP-19, PME-1, SET, WT1

[#]Corresponding author: Jukka Westermarck, tel. +358-29-450 2880, jukwes@utu.fi

Abstract

Despite of extensive genetic analysis of acute myeloid leukemia (AML), we still do not understand comprehensively mechanism that promote disease relapse from standard chemotherapy. Based on recent indications for non-genomic inhibition of tumor suppressor protein phosphatase 2A (PP2A) in AML, we examined mRNA expression of PP2A inhibitor proteins in AML patient samples. Notably, out of examined PP2A inhibitor proteins, overexpression of ARPP19 mRNA was found independent of current AML risk classification. Functionally, ARPP19 promoted AML cell viability and expression of oncoproteins MYC, CDK1, and another PP2A inhibitor CIP2A. Clinically, ARPP19 mRNA expression was significantly lower at diagnosis (p=0.035) in patients whose disease did not relapse after standard chemotherapy. ARPP19 was an independent predictor for relapse both in univariable (p=0.007) and in multivariable analyses (p=0.0001); and gave additive information to EVI1 expression and risk group status (additive effect, p=0.005). Low ARPP19 expression also associated with better patient outcome in TCGA LAML cohort (p=0.019). In addition, in matched patient samples from diagnosis, remission and relapse phases, ARPP19 expression associated with disease activity (p=0.034).

Together, these data identify ARPP19 as a novel oncogenic PP2A inhibitor protein in AML, and demonstrate its risk group independent role in predicting AML patient relapse tendency.

Background

Acute myeloid leukemia (AML) is one of the most aggressive cancer types[1] and although up to 85% of the patients under the age of 60 achieve complete remission (CR) after standard induction therapy, only 35 to 40% can be fully cured [1,2]. In adult AML, the actual risk profile of a significant percentage of patients is not optimally reflected in current genetic classification schemes [3-5]. According to the European Leukemia Net risk group classification, most adult AML patients belong to intermediate risk group [6] which practically means that they have high relapse risk after conventional chemotherapy. These patients are therefore often directed to hematopoietic stem cell transplantation (HSCT). However, all intermediate risk patients would not need HSCT but could be cured with intensive chemotherapy. Together with mortality rate up to 25% among HSCT patients, and life-long need for immunosuppression with surviving patients, it would thus be of high clinical relevance to better understand the mechanisms that promote relapse tendency. Further, although patients in favourable risk group mostly can be cured with chemotherapy, some patients yet relapse. Thereby, understanding of the mechanisms behind high relapse risk would be useful to develop approaches to recognize the favorable risk group patients that would benefit from being directed to immediate HSCT.

Protein Phosphatase 2A (PP2A) is a tumor suppressor, which plays a critical role in plethora of cancer relevant cellular processes including regulation of cell cycle and apoptosis [7,8]. In cancer, the non-genomic inhibition of PP2A activity by elevated expression of endogenous PP2A inhibitor proteins (PAIPs), such as

CIP2A, SET, PME1, ARPP19 and TIPRL, greatly exceeds the frequency of genetic mutations on PP2A genes [9]. However, while in many solid cancers the non-genomic inhibition of PP2A has already been extensively studied, in haematological malignancies this understanding is still relatively poor.

Due to recent discovery of PP2A inhibition as a putative AML driver mechanism [10], PAIPs are also emerging as potentially interesting AML marker genes. However, none of the published studies have compared systematically the expression profiles of different PAIPs in AML. One of the PAIPs, ARPP19 (cAMP-regulated phosphoprotein 19), a member of the alpha-endosulfine (ENSA) family, has been shown to promote G2/M transition and the mitotic state in solid cancer cells [11]. ARPP19 overexpression has been linked to tumor progression in solid cancers such as glioma[12] and hepatocellular carcinoma[13] but its role in AML has not been studied as yet.

In this first study addressing landscape of PAIPs in AML, we discovered low ARPP19 mRNA expression as a novel predictive marker for estimation of low relapse risk in patients with AML. We also identified ARPP19 as an AML oncoprotein that increases cell viability and enhances expression of oncoproteins MYC and CDK1 and also of another oncogenic PP2A inhibitor protein CIP2A. Most importantly, we found that ARPP19 mRNA expression and its role as a predictive relapse marker was independent of current genetic risk classification schemes suggesting that ARPP19 mediates its functions in AML by mechanisms that are independent of the known genetic mechanisms. Together these novel results identify ARPP19 as a potential AML oncoprotein with clinical relevance.

Materials and methods

Patient cohorts

Patient cohort1: Consecutive bone marrow samples were collected between January 2000 and July 2010, a total of 80 patients aged 18-65 diagnosed with *de novo* or secondary AML at Turku University Hospital (TYKS). Patients with acute promyelocytic leukemia (t(15;17)(q22;q12)) were excluded from this cohort. Patient characteristics are presented in supplemental Table 1. Median age for the patients was 50 years (Q_1=38.8, Q_3=58.0), median overall survival was 5.4 years (95% CI, 2.8 to 7.9) and median follow-up time was 5.4 years (range 6 days - 16 years). The ELN risk classification, based on cytogenetic and molecular findings, was used as risk stratification (supplemental Table 2). Most patients (76) were enrolled in the Finnish Leukemia Study Group prospective protocols (supplemental Table 3). 32 patients were treated according to AML92 and 44 according to AML2003 protocol. Treatment of four patients was significantly modified due to patient related reasons. Although patients were treated with different schedules, all received regimens based on anthracycline and high-dose cytarabine as induction therapy. High-dose cytarabine, and allogenic stem cell transplantation when possible, were used as consolidation therapy. No significant differencies were found in the relapse or overall survival rates of patients on AML92 or AML2003 treatment. Informed consent was obtained from all patients and the local Ethical Review Board of TYKS approved the study protocol. No missing data imputation was performed.

Patient cohort2: Bone marrow samples from 48 AML patients, including nine AML patients with supplemental follow-up samples at first remission and/or at relapse,

were analyzed from The Finnish Hematology Registry and Clinical Biobank (FHRB) collection. Patient characteristics for the nine patients are presented in supplemental Table 4. All 48 patients had received intensive chemotherapy as an induction therapy and achieved CR. Additional follow-up samples at remission were available from four patients and at relapse from eight patients. Samples were collected from Finnish university hospitals and other hematological units between December 2011 and January 2017. Median age for the nine patients was 59.8 years (Q_1=50.7, Q_3=68.8), median overall survival was 1.7 years (95% CI, 1.3 to 3.9) and median follow-up time was 1.7 years (range 1 – 4.5 years). FHRB is authorized by the Finnish National Supervisory Authority for Welfare and Health (Valvira) and has been approved by the Finnish National Medical Ethics Committee. All patients signed an informed consent prior to biobanking.

Statistical analysis

Continuous variables were summarized by descriptive statistics (median, interquartile range and range) while frequencies and percentages were calculated for categorical data. Patients were stratified according to gene expression at diagnosis into high (>median expression of the studied gene in AML patients) and low (<median expression of the studied gene in AML patients). Additional analysis was performed by using overexpression (>mean expression of the studied gene in normal sample), underexpression (<mean expression of the studied gene in normal sample) or subpopulation analysis based on the distribution profile of the studied gene expression (also including quartiles). For continuous variables, when possible, transformations (ln, sqrt) were performed to achieve a normal distribution assumption. Wilcoxon rank sum test, Kruskal-Wallis

147 test, Student's t-test and paired t-test were used for analyzing continuous
148 variables.
149 Frequency tables were analyzed using Fisher's exact test for categorical
150 variables. Pairwise Pearson correlation analysis was performed on gene-to-gene
151 manner and further hierarchical clustering (average linkage) was performed.
152 Separate logistic regression model was fit for ARPP19 and EVI1 alone and
153 ARPP19+EVI1 together. Discriminative power of the three models was evaluated
154 using Receiver Operating Characteristic (ROC) curves. Chi-squared test was
155 used for comparison of AUC-values.
156 Univariable survival analysis for overall survival (OS) and time to relapse was
157 based on the Kaplan–Meier method where stratum-specific outcomes were
158 compared using log-rank statistics. To adjust for the explanatory variables
159 (diagnosis age, risk group stratification, FLT3-ITD status, NPM1 mutation status,
160 expression levels of CIP2A, SET, EVI1, WT1, ARPP19, TIPRL and PME1), a Cox
161 proportional hazards regression model was used for univariable and multivariable
162 analysis. We used type 1 approach where we report additive effect of the marker.
163 In multivariable analysis, covariates were entered in a stepwise backward
164 manner.
165 OS was defined for all patients measured from the date of diagnosis to the date
166 of death from any cause. Patients not known to have died at last follow-up were
167 censored on the date they were last known to be alive. Time to relapse was
168 defined for patients from the date of diagnosis until the date of relapse. Patients
169 not known to have relapsed were censored on the date they were last examined.
170

171 **RNA isolation and cDNA synthesis**

Total RNA was isolated from extracted mononuclear cells (patient bone marrow samples). Total RNA was extracted using the E.Z.N.A.® Total RNA Kit I (Omega Bio-Tek Inc, Norcross, GA, USA) according to the manufacturer's instructions. After isolation, RNA concentration was measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). cDNA was synthesized (with 1 µg of total RNA as a starting material) using SuperScript III Reverse Transcriptase (18080093, Invitrogen, Carlsbad, CA, USA), random primers (C1181, Promega), RiboLock(tm) Ribonuclease Inhibitor (#EO0381, Thermo scientific) and dNTP-mix (BIO-39028, Bioline, London, UK). RT-reactions were performed according to enzyme's manufacturer's instructions.

Quantitative real-time PCR (RQ-PCR)

Primers for each gene specific assays were designed to be located to different exonic sequences to avoid amplification of genomic DNA. Primer concentration in each reaction was 300 nM and probe concentration 200 nM. Specificity of RQ-PCR reactions was verified by agarose gel electrophoresis and melting curve analysis. Single band of the expected size and a single peak, respectively, were required. The amplification efficiency for each target was also assessed. shRNA control and standard curve analysis for amplification efficiency and melting curve analysis for ARPP19 RQ-PCR are shown in Supplemental Figure 1a-d. Amplification of target cDNAs was performed using KAPA PROBE FAST RQ-PCR Kit (Kapa Biosystems, Wilmington, MA, USA) and 7900 HT Fast Real-Time PCR System (Thermo Fisher) according to the manufacturers' instructions. Quantitative real-time PCR was executed under the following conditions: 95°C for 10 min followed by 45 cycles of 95°C for 15 s and 60°C for 1min. Relative

gene expression data was normalized to the expression level of endogenous house-keeping genes Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and beta-actin (ACTB) using 2^-ΔΔC(t) method with SDS software (version 2.4.1, Applied Biosystems, Foster City, CA, USA) or with Thermo Fisher Cloud Real-time qPCR Relative Quantification application (Patient cohort2). To estimate the degree of overexpression in AML, the expression of each gene was normalized to the expression level in a commercial normal pooled (from 56 males and females) bone marrow control sample (636591, lot 1002008, Clontech Laboratories, Fremont, CA, USA). Results were derived from the average of at least two independent experiments and two technical replicates. Primer and probe sequences used in this study for RQ-PCR analysis are listed in supplemental Table 5.

Cell culture

KG-1 (ACC-14), HL-60 (ACC-3), MOLM-14 (ACC-777) and KASUMI-1 (ACC-220) cell lines were obtained from Leibniz-Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany). All cell lines were maintained in RPMI-1640 medium (R5886, Sigma-Aldrich, Saint Louis, MO, USA) supplemented with 10% (KG-1, HL-60) or 20 % (MOLM-14, KASUMI-1) heat inactivated fetal bovine serum (FBS, Gibco, Thermo Fisher Scientific), 2 mM L-glutamine (Sigma-Aldrich), 50 units/ml penicillin (Sigma-Aldrich) and 50mg/ml streptomycin (Sigma-Aldrich). All cell lines were routinely tested for mycoplasma contamination.

Antibodies

The following antibodies were used: rabbit polyclonal anti-ARPP19 (11678-1-AP, Proteintech Group, Rosemont, Illinois, USA), mouse monoclonal anti-cMYC (sc-40, Santa Cruz Biotechnology, Dallas, Texas, USA), mouse monoclonal anti-b-actin (A1978, Sigma), mouse monoclonal anti-GAPDH (5G4 clone, HyTest, Turku, Finland), mouse monoclonal anti-cyclin D1 (sc-450, Santa Cruz),), mouse monoclonal anti-CDK1 (sc-51578), ECL HRP-linked secondary antibodies (Agilent Dako, Santa Clara, California, USA).

Western blot assay

Protein extracts were separated using SDS-PAGE under denaturing conditions (4–20% Mini-PROTEAN TGX Gels) and were transferred to the PVDF membrane (Bio-Rad Laboratories). Membranes were blocked with 5% milk-TBST (Tris-buffered saline and 0.1% Tween 20), incubated with the indicated primary antibodies overnight at 4°C, and then incubated ECL HRP-linked secondary antibodies at RT for 1 h. ECL Plus Western blotting reagent (GE Healthcare) was added to the membrane and film was developed. Band intensity was determined using ImageJ software (National Institutes of Health).

siRNAs, shRNAs and cell viability assay

The following siRNAs were used: siGENOME Human ARPP19 (10776) siRNA (D-015338-03, Dharmacon, Lafayette, CO, USA) and siRNA Scramble (CGUACGCGGAAUACUUCGA). Transfections were performed with Nucleofector II Device (Lonza, Basel, Switzerland) and optimized programs for each cell line.

shRNA constructs were ordered as lentiviral particles from Functional genomics unit's (FuGu, University of Helsinki, Finland) TRC1 library. ARPP19 shRNAs were TRCN0000158847 and TRCN0000160408. Control shSCR was SHC002 (Sigma). To establish the stable cell line, the ARPP19-RNAi lentivirus was transfected into HL-60 and KG1 cells with several different amounts of infectious virus. 24h after transduction spinoculation was performed and selection was done with puromycin at 72h time point. ARPP19 expression was determined through western blot analysis and qPCR. Differences in cell viability of shARPP19 transduced cell lines compared to control shRNA cell lines was measured with CellTiter-Glo® (Promega) luminescent assay (Promega) at 24h, 48h, 72h, 96h or 120h after plating the cells. Results were derived from the average of three independent experiments.

Results

PP2A inhibitor protein mRNA expression in AML patient cohort

We analyzed mRNA expression levels of PAIPs: CIP2A, PME1, TIPRL, SET and ARPP19 by real-time quantitative PCR (RQ-PCR) from 80 diagnosis phase AML patients' bone marrow (BM) samples. Patient characteristics and distribution of the patients to three clinically used risk groups (favorable n=21, intermediate n=37, adverse n=22) based on their genetic profiles were representative of an average AML patient population (supplemental Table 1). The representative nature of the study material was also confirmed by significant association between risk groups and overall survival (OS) of patients in this cohort (Figure 1a, p=0.003 by log-rank test). Five-year survival rate was 81% for the patients in favorable (Figure 1a, blue), 51% for the patients in intermediate (red) and 27% for the patients in adverse risk group (green). The median OS in the whole cohort was 5.4 years (95% CI, 2.8 to 7.9) and the probability of OS at five years was 52.5%.

To estimate the degree of overexpression in AML, expression of each gene was normalized to the expression level in a pooled normal BM control sample from 56 males and females (Clontech). The degree of overexpression as well as median expression of each target gene, are shown in supplemental Table 6. Waterfall blots of the expression patterns of the measured genes related to normal BM control set as 0, are shown in Figure 1.

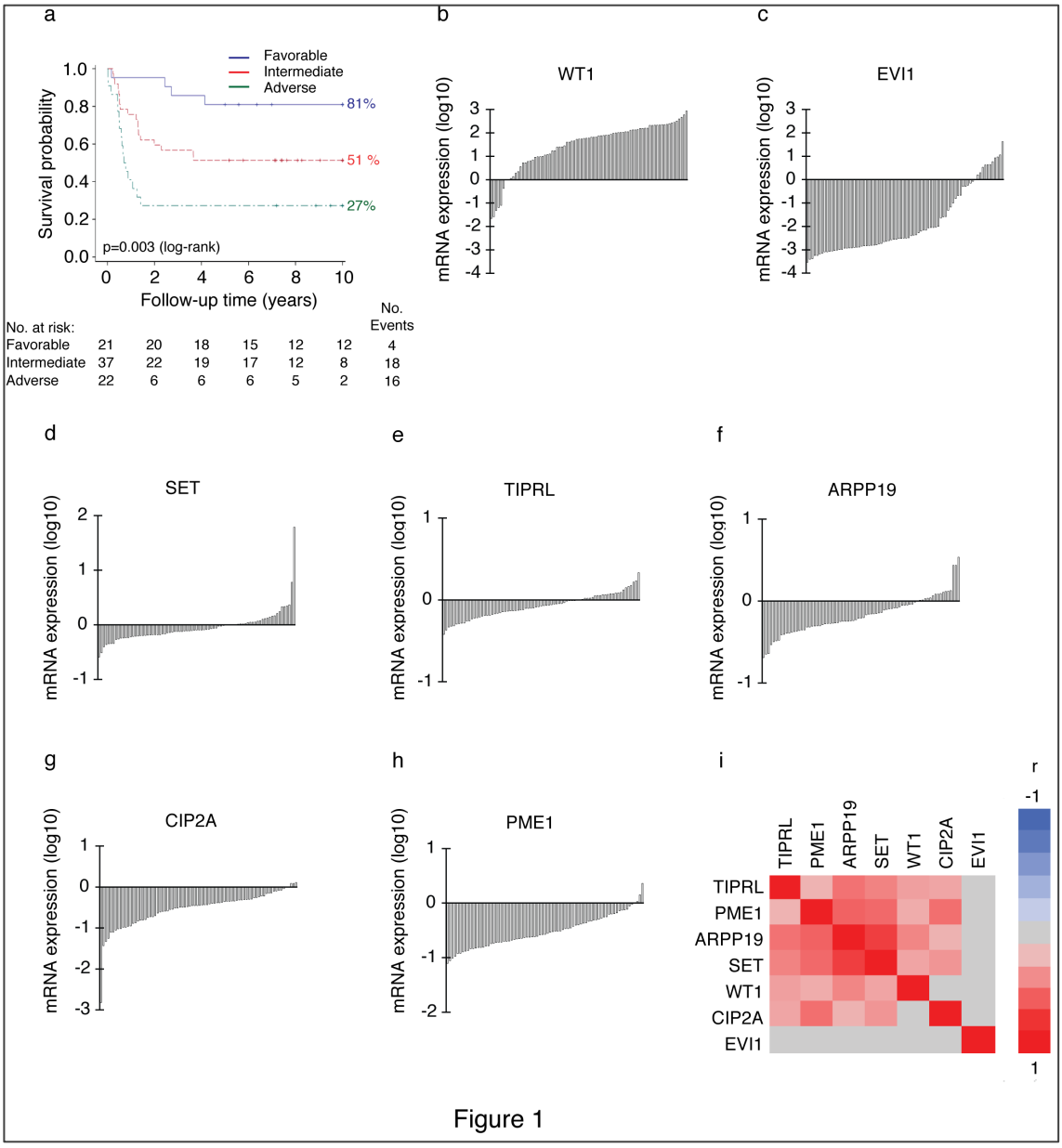


Figure 1

Figure 1. Expression profiles of PP2A inhibitors in AML patient samples. a) Higher risk group is significantly associated with poor survival of AML patients in patient cohort1. p=0.003 by log-rank test. Favorable n=21, intermediate n=37, adverse n=22. b) to h) Waterfall blots of analysed genes from the sample panel normalized to GAPDH & b-actin expression and a pooled (n=56) normal bone marrow sample. On y-axis log10 transformed RQ mRNA expression values derived from two technical replicates in two independent experiments. One bar represents one patient. b) WT1 mRNA expression was highly overexpressed (91%) in diagnosis phase AML patients bone marrow compared to normal bone marrow. c) EVI1 overexpression was 13%, d) SET overexpression was 30%, e) TIPRL overexpression was 30%, f) ARPP19 overexpression was 21%, g) CIP2A overexpression was 4% and h) PME1 overexpression was 4% in the sample

295 panel. i) Hierarchical clustering of Pearson's pairwise correlations for the mRNA expression of
296 PP2A inhibitors in patient cohort1. Three potentially oncogenic PP2A inhibitors, PME1, ARPP19
297 and SET, form a cluster with correlated expression patterns. Red represents positive and blue
298 negative correlation. Grey indicates non-significant correlation (p-value >0.05).
299
300 As an additional indication for the representative nature of the sample material,
301 the expression patterns of established AML markers Wilms' tumor 1 (WT1)[14] and
302 ectopic viral integration site-1 (EVI1)[15] were in accordance to published literature.
303 WT1 mRNA was overexpressed in 91% of diagnosis phase AML patients' bone
304 marrow as compared to normal bone marrow (supplemental Table 6 and Figure
305 1b), whereas EVI1 overexpression was observed in 13% (Figure 1c) of the
306 patients. The overexpression pattern of PP2A inhibitor SET in 30% of patients
307 (Figure 1d) was also consistent with the published literature[16]. Of the other PP2A
308 inhibitors, TIPRL overexpression level was equal to SET (Figure 1e, 30%),
309 whereas ARPP19 overexpression was found in 21% of patients (Figure 1f). In
310 contrast, neither PME1 (Figure 1h, 4%), nor CIP2A (Figure 1g, 4%) were found
311 notably overexpressed in this AML patient cohort.
312
313 As some of the PAIPs have been previously associated to AML[10,17,18], but their
314 relatioships to each other are not clear, we used this first study addressing
315 landscape of PAIPs in AML to estimate their expression redundancies and mutual
316 dependencies by Pearson correlation analysis. We found that PME1 levels
317 correlated with CIP2A (Figure 1i, $r=0.52$, $p<0.001$), SET ($r=0.54$, $p<0.001$) and
318 ARPP19 ($r=0.58$, $p<0.001$) expression. Additionally, SET expression levels
319 correlated with TIPRL ($r=0.43$, $p<0.001$), and strongly with ARPP19 gene
320 expression ($r=0.75$, $p<0.001$). Furthermore, diagnosis phase ARPP19

expression levels also correlated with WT1 (r=0.42, p=0.001) and TIPRL (r=0.51, p<0.001) gene expression. Hierarchical clustering of the correlation matrix suggests that the expression of three PP2A inhibitors, ARPP19, PME1 and SET, form a cluster with similar expression pattern across AML patient samples (Figure 1i). EVI1 gene expression did not show any significant correlation with any other target gene in this patient cohort (for all correlations p>0.05).

Based on these analyses ARPP19 is overexpressed in AML, and it associates with SET that previously have been implicated in AML[17,18]. To validate the ARPP19 as a novel AML overexpressed gene in an independent patient cohort, we analysed 48 patients from The Finnish Hematology Registry and Clinical Biobank (FHRB)(cohort2) that had received intensive chemotherapy as an induction therapy. ARPP19 mRNA was overexpressed in 58% (n=28) of the cohort2 sample panel (Supplemental figure 2a), thus providing an independent validation for high rate of overexpression of ARPP19 in a subset of adult AML patients.

ARPP19 expression promotes AML cell survival

To explore functional role of ARPP19 in AML cells, we used four established cell lines that were chosen on the basis of their diverse genetic background (DSMZ Scientific data). Consistent with patient samples at mRNA level (Figure 1f), Western blot analysis demonstrated variable ARPP19 protein expression levels between AML cell lines (Figure 2a and 2b). Interestingly, eventhough ARPP19 and CIP2A did not strongly correlate at mRNA level (Figure 1i), ARPP19 protein

346 expression correlated CIP2A protein expression in these cell lines (Figure 2a). A
347 plausible explanation for this could be post-transcriptional stabilization of CIP2A
348 protein in ARPP19 positive AML cells.
349
350 To address role of ARPP19 in AML cell survival, ARPP19 was down-regulated
351 by lentiviral shRNAs in HL-60 and KG-1 cell lines expressing high endogenous
352 ARPP19 protein levels. Indicative of pivotal role for ARPP19 in AML cell survival
353 or proliferation, it was very challenging to maintain a long-term depletion of
354 ARPP19 by shRNA in cell clones. However, by using early cell clones with partial
355 ARPP19 protein knock-down, we were able to document significantly decreased
356 cell viability in ARPP19 shRNA transduced KG-1 cells (Figure 4c, 4d). Partial
357 ARPP19 inhibition also resulted in statistically significant inhibition of number of
358 KG-1 cells in M/G2 cell cycle state (Figure 2e). Similar to KG-1 cells, also HL-60
359 cells were sensitive to partial shRNA-mediated ARPP19 inhibition (Figure 2f,g).
360

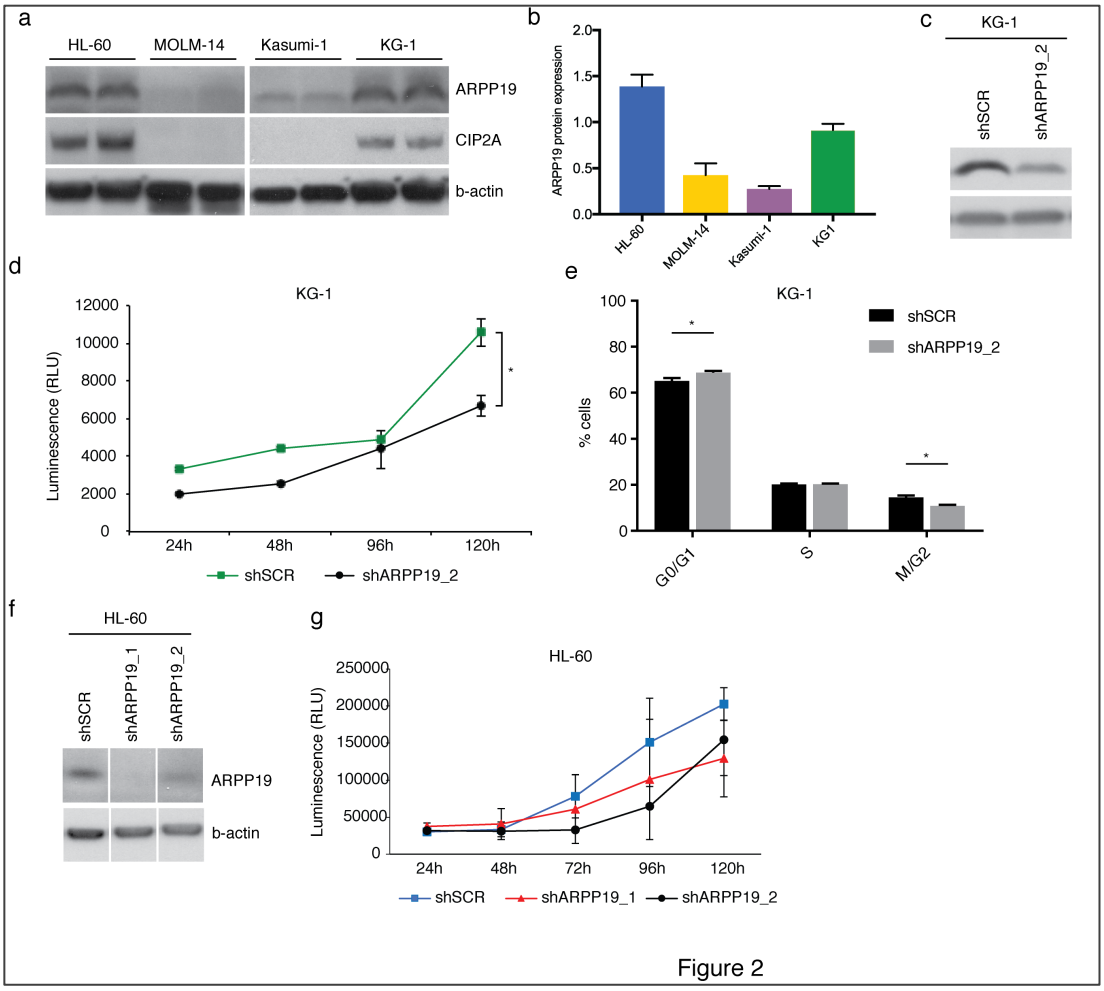


Figure 2

Figure 2. ARPP19 expression promotes AML cell survival

a) Representative western blot analysis of endogenous ARPP19 and CIP2A expression in HL-60, MOLM-14, Kasumi-1 and KG-1 cell lines. b) Quantitation of ARPP19 protein levels from a). B-actin was used as loading control and internal control was used to normalize the quantities on separate membranes. Internal control (T98G cell lysate) is set as 1. Shown is mean with SEM. c) Representative western blot analysis of ARPP19 expression in KG-1 cell line stably transduced with scrambled (shSCR) or shARPP19_2). d) KG-1 cell viability measured with CellTiter-Glo® (CTG) assay 24h, 48h, 96h or 120h after plating of stable shSCR or ARPP19 shRNA cells. Shown is mean ±SEM of three experiments. *(shSCR vs. shARPP19_2 at 120h) p=0.013 by Student's t-test. e) KG-1 cell cycle measured with propidium iodide staining and FACS analysis. Shown is mean ± SEM of two independent experiments, and significant values are indicated with asterisks (* $p < 0.05$, shSCR vs. shARPP19_2 by 2way ANOVA). f) Representative western blot analysis of ARPP19 expression in HL-60 cell line stably transduced with scrambled (shSCR) or two different ARPP19 shRNAs (shARPP19_1 and shARPP19_2). g) HL-60 cell viability measured

with CTG assay 24h, 48h, 72h, 96h or 120h after plating of stably with shSCR or two different ARPP19 shRNA transduced cells. Shown is mean ±SEM of three experiments.

ARPP19 promotes expression of oncogenic drivers MYC, CDK1, and CIP2A in AML cells

As a more direct support of the oncogenic role of ARPP19 in AML, transient ARPP19 knockdown by siRNA decreased expression of a well-validated PP2A target[20], and oncoprotein, MYC, and of cell cycle mediator CDK1 in both cell lines (Figure 3a,b). Very interestingly, acute depletion of ARPP19 also resulted in down-regulation of CIP2A in both cell lines, whereas ARPP19 protein levels were not affected upon CIP2A inhibition (Figure 3a,b). Down-regulation of CIP2A and of MYC upon ARPP19 silencing was validated in stably transduced HL-60 cells (Supplemental figure 2b). As shown previously in other cancer models[21], CIP2A promoted MYC protein levels in both of the studied AML cell lines (Figure 3a,b).

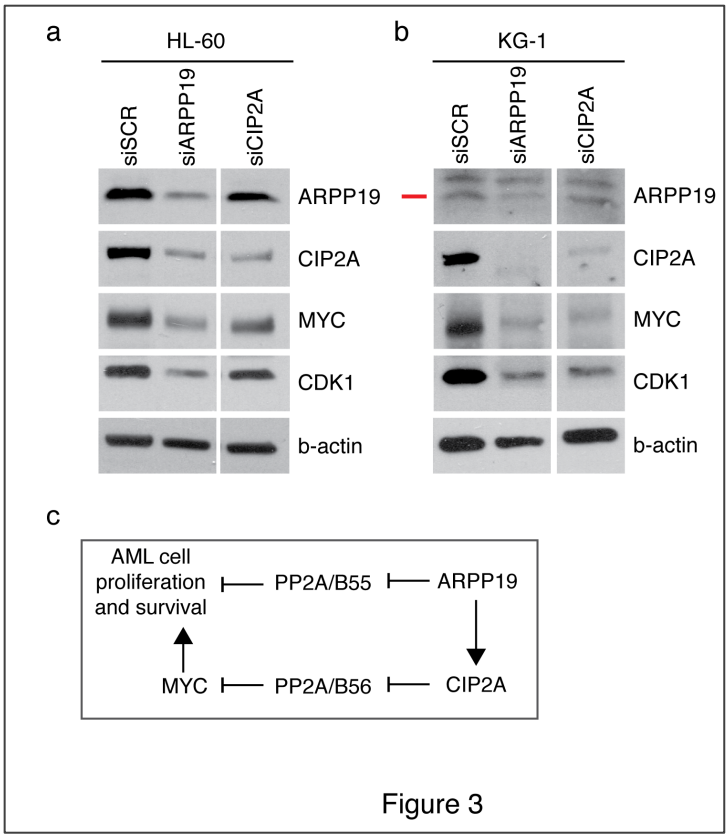


Figure 3

Figure 3. ARPP19 promotes expression of oncogenic drivers MYC, CDK1, and CIP2A in AML cells

a) Western blot analysis of ARPP19, CIP2A, MYC and CDK1 expression in HL-60 and b) KG1 cell line 48h after transfection with scrambled (SCR), ARPP19 (siARPP19) or CIP2A siRNA (siCIP2A). Red line marks the correct ARPP19 band. c) Schematic model of hierarchy of the two PP2A inhibitor proteins ARPP19, and CIP2A in regulation of AML cell proliferation and survival.

Together with cell survival analysis, these results support oncogenic role for ARPP19 in AML. The results also indicate a hierarchial co-regulation of two oncogenic PP2A inhibitors, ARPP19 and CIP2A at the protein level (Figure 3c). ARPP19 is known as an inhibitor of PP2A/B55 complex[11], whereas CIP2A regulates PP2A/B56[22]. The data indicate that by means of promoting CIP2A protein expression, ARPP19 could in principle control both these PP2A tumor suppressor complexes (Figure 3c).

Low ARPP19 mRNA expression is an independent predictive relapse marker

To assess potential clinical relevance of ARPP19 in human AML, we correlated expression of ARPP19, and of other mRNA markers, to clinical features of the patients from the cohort1. In this patent cohort, the median follow-up time of 68 AML patients who achieved complete remission (CR) was 7.2 years (range 0.2–15.9 years), and in line with a representative nature of this patient material, the patients that relapsed were more likely to be in adverse risk group than patients who did not relapse during the follow-up time (40% vs. 14%, p=0.038 by Fisher's exact test). Notably however, none of the PP2A inhibitor genes, including overexpressed ARPP19 ($p>0.05$ by Kruskal-Wallis test and supplemental Table 7), showed statistically significant association with the risk groups. On the other hand, and as expected, EVI1 mRNA expression at diagnosis was significantly different between the three risk groups, and its expression increased in relation to the risk group (p=0.005 by Kruskal-Wallis test). Together these results indicate that potential clinical correlations with ARPP19 are independent of genetic risk group classification of the patients.

Importantly, supportive of oncogenic role for ARPP19 in human AML, patients without relapse during the follow-up time had significantly lower ARPP19 expression than patients that relapsed during the follow-up time (Figure 4a, p=0.035 by Wilcoxon rank-sum test). However, there was no significant difference in the rate of CR (75% vs. 88%), resistance (25% vs. 8%) or death during induction therapy (0% vs. 3%) between patients with ARPP19

underexpression or overexpression. This indicates that low ARPP19 rather associates with low relapse tendency after remission, than with better induction therapy response. In regards to other evaluated mRNA markers, EVI1 was the only other marker which expression correlated with relapse (p=0.023 by Wilcoxon rank-sum test). There were no significant differences between non-relapsing and relapsing groups in any other clinical characteristics including patient's age, alloHSCT, secondary AML, extramedullary disease, normal karyotype, NPM1 mutation and FLT3-ITD gene fusion. Of a note, most of the patients with relapse (85%) did not have FLT3-ITD nor NPM1 mutation (FLT3-ITD-, NPM1-) at diagnosis.

Kaplan-Meier estimates were used to analyze association of markers with time to relapse. As expected, the risk group of the patients was a strong indicator of shorter time to relapse (p=0.008 by log-rank test). Notably, patients in the lowest quartile (Q_1) of ARPP19 expression were linked to longer relapse free time (Figure 4b, p=0.029; as compared to those over lowest quartile). Five-year relapse rate was only 7% for patients with the lowest quartile expression of ARPP19, while five-year relapse rate was 33% for patients that had ARPP19 expression higher than the lowest quartile. Importantly, directly underlining the risk group independent role for ARPP19 in relapse, patients in the lowest quartile ARPP19 expression (i.e. not relapsing patients) represented all risk groups and none of the intermediate risk group patients in this low ARPP19 cohort relapsed during > 10 years follow-up time (Figure 4c). On the other hand, 27 % of patients with high relapse tendency according to higher than Q_1 ARPP19 expression belonged to favorable risk group (Figure 4d). In addition to ARPP19, only EVI1,

458 and SET gene expressions had any role in predicting the prevalence of relapse
459 in this patient cohort. High EVI1 mRNA expression was a strong indicator of
460 shorter time to relapse (Supplementary figure S3a, p<0.0001 by log-rank test).

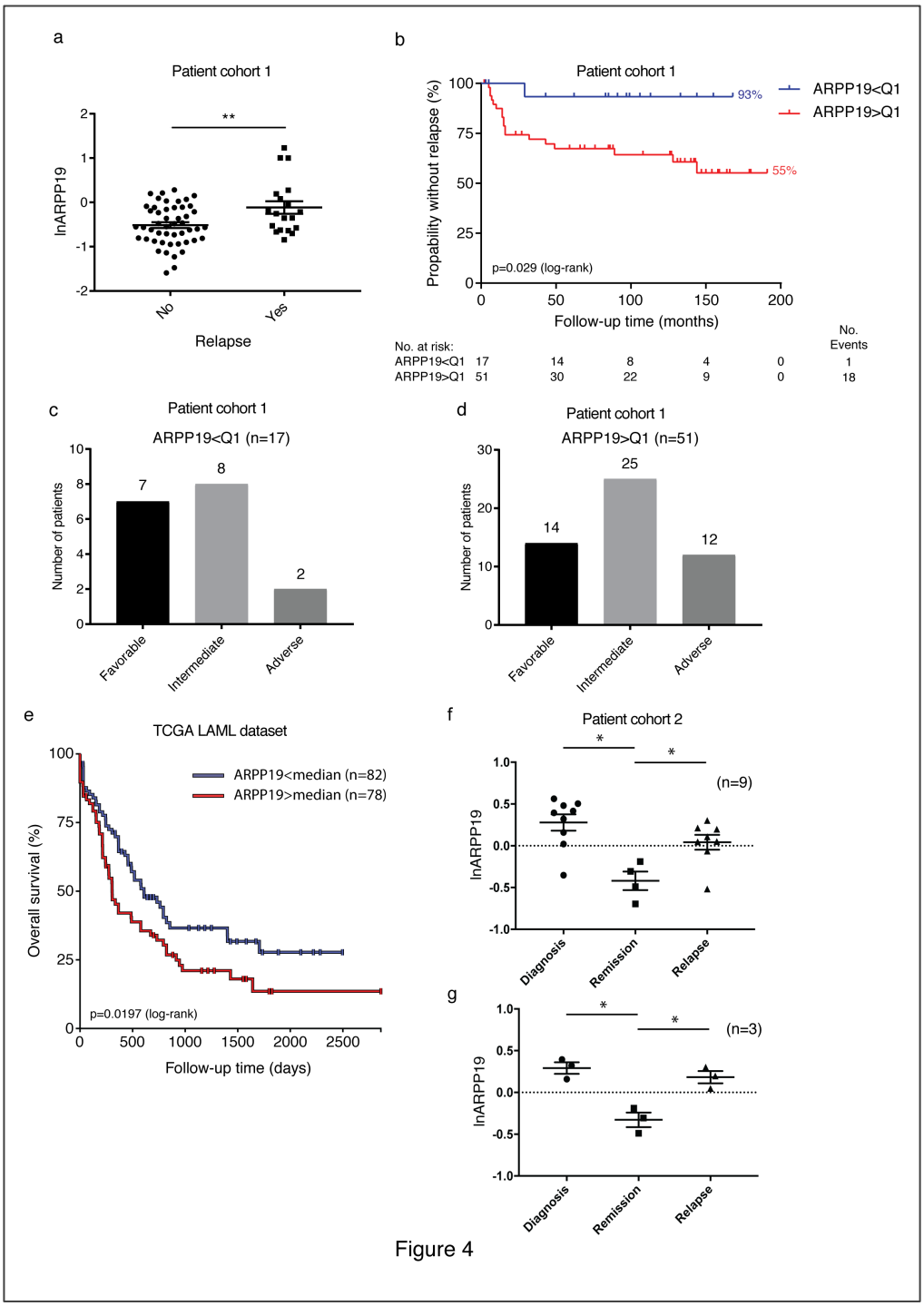


Figure 4

461

462 **Figure 4. High ARPP19 mRNA expression is a risk group independent predictive relapse**
463 **marker in AML.** a) Patients without relapse in cohort1 (n=49) during the follow-up time had lower
464 diagnostic ARPP19 mRNA expression than patients with relapse (n=19). Shown is logarithmic

mean ± standard error of mean (SEM). ARPP19 expression in pooled (n=56) normal bone marrow sample is set as 0. **(no vs. yes relapse) p=0.0046 by Student's t-test. b) Lower ARPP19 expression is associated with longer time to relapse (Kaplan-Meier estimate in months) in AML patients, p=0.029 by log-rank test. Q_1= lowest quartile of ARPP19 mRNA expression (n=17). c) Patients in the lowest quartile ARPP19 mRNA expression are assigned to all risk groups. d) Patients over the lowest quartile ARPP19 mRNA expression are assigned to all risk groups. e) Kaplan–Meier survival curve for overall survival (OS) by ARPP19 gene expression (exon IlluminaHiSeq RNAseq) in TCGA acute myeloid leukemia (LAML) patients (n = 160). Median serves as cut-off value. Lower ARPP19 expression is associated with longer OS in AML patients, p=0.0197 by log-rank test. f) ARPP19 mRNA expression in diagnosis (n=9), remission (n=4) and relapse (n=8) samples from patients with AML. ARPP19 expression in pooled (n=56) normal bone marrow sample is set as 0 (dashed line). Shown is logarithmic mean ±SEM. *(diagnosis vs. remission) p=0.021, *(remission vs. relapse) p=0.034 by paired t-test. g) ARPP19 mRNA expression in matched diagnosis, remission and relapse samples from three patients with AML. *(diagnosis vs. remission) p=0.047, *(remission vs. relapse) p=0.034 by paired t-test.

Uni -and multivariant analyses

Supportive of independent role for ARPP19 in regulating AML relapse, Cox's univariable analysis also revealed that ARPP19 (Table 1, p=0.007, HR 2.87 (95% CI, 1.33 to 6.22)), EVI1 (p=0.0005, HR 1.26 (95% CI, 1.11 to 1.44)), and SET (p=0.035, HR 2.36 (95% CI, 1.06 to 5.25)) expressions at the diagnosis had significant role in predicting the time to relapse.

Table 1. Univariable analysis for time to relapse in entire patient cohort1.

Parameter	Time to relapse		
	Hazard Ratio	95% CI	p
Riskgroup			0,017
EVI1	1,26	1.11 to 1.44	0,0005
SET	2,36	1.06 to 5.25	0,035
ARPP19	2,87	1.33 to 6.22	0,007

Multivariable Cox's proportional hazard model for relapse included age, FLT3-ITD status, NPM1 mutation status and mRNA expression of ARPP19, CIP2A, SET, TIPRL, PME1, EVI1 and WT1 at diagnosis. In the initial model the significant markers for time to relapse were diagnosis age (p=0.024), NPM1 mutation positivity (p=0.035), EVI1 (p=0.0004), SET (p=0.021) and ARPP19 (p=0.0008) mRNA expression. After excluding the non-significant markers, age (p=0.023, HR: 1.07, 95% CI, 1.01 to 1.13), NPM1 mutation positivity (p=0.048, HR: 0.031 (95% CI, 0.001 to 0.97), EVI1 (p=0.0005, HR: 1.41 (95% CI, 1.16 to 1.71), SET (p=0.0097, HR: 0.12 (95% CI, 0.022 to 0.59) and ARPP19 (p=0.0001, HR: 58.8 (95% CI, 7.39 to 467.2) expressions were independent prognostic factors for the time to relapse (Table 2).

Table 2. Multivariable analysis for time to relapse and overall survival in the entire patient cohort1.

Parameter	Time to relapse		Overall survival	
	Hazard Ratio	p	Hazard Ratio	p
Diagnosis age	1,07	0,023	1,07	0,0004
NPM1 mutation	0,03	0,048	0,21	0,017
EVI1	1,41	0,0005	1,14	0,026
SET	0,12	0,010	-	-
ARPP19	58,77	0,0001	2,05	0,046

Very importantly, Cox's type1 analysis revealed that ARPP19 expression (p=0.005) gave additional information in AML patients relapse prognosis after risk group, and EVI1 mRNA expression, were depicted as significant factors in explaining the probability of relapse. Receiver operating characteristic (ROC) analysis also showed that ARPP19 together with EVI1 could be more accurate predictor of relapse than EVI1 alone (EVI1 AUC 0.69 (95% CI, 0.48 to 0.89), ARPP19 AUC 0.67 (95% CI, 0.48 to 0.83) and ARPP19+EVI1 AUC together 0.76 (95% CI, 0.57 to 0.91); EVI1 AUC vs. EVI1+ARPP19 AUC p=0.07 by Chi-Squared test, supplemental Figure 3b).

Together these results identify low ARPP19 expression as a novel risk group independent gene associated with low relapse risk in human AML. Importantly, the predictive role of ARPP19 was additive when the currently used clinicopathological markers, including risk group classification, were taken into account.

Survival analysis based on PP2A inhibitor protein mRNA expression in AML patients

Next, we analyzed whether the risk group independent predictive role of ARPP19 for relapse is reflected in the overall survival of all 80 cases treated with intensive chemotherapy in cohort1. For this purpose, we used Cox's proportional multivariable hazard model for OS, which included diagnosis age, FLT3-ITD status, NPM1 mutation status, and diagnosis phase mRNA expression levels of ARPP19, CIP2A, SET, TIPRL, PME1, EVI1 and WT1. In the initial model the

significant markers for OS were diagnosis age (p=0.024) and EVI1 (p=0.0127) mRNA expression. After excluding the non-significant markers, diagnosis age (Table 2, p=0.0004, HR: 1.07 (95% CI, 1.03 to 1.11), NPM1 mutation positivity (p=0.0165, HR: 0.21 (95% CI, 0.057 to 0.75)) and EVI1 expression (p=0.0263, HR: 1.14 (95% CI, 1.02 to 1.28) were found as independent prognostic factor for OS. Notably, out of the PAIPs, only ARPP19 mRNA expression was found as an independent prognostic factor for OS, and its HR was found to be even higher than either EVI1 or diagnosis age (p=0.0456, HR: 2.05 (95% CI, 1.01 to 4.15)).

To evaluate these results in an independent AML patient cohort, we used RNA sequencing dataset available from TCGA (TCGA LAML, survival data available for n=160, exon expression IlluminaHiSeq)[23] and analyzed the correlation between OS and ARPP19 gene expression using UCSC Xena Browser [24]. Based on median as a cut-off value, the data was categorized into two groups, low ARPP19 and high ARPP19. Consitent with other results, ARPP19 expression alone was able to act as an independent prognostic marker for OS (Figure 4e). Patient group with low ARPP19 gene expression (n = 82) showed better OS (p=0.019 by log-rank test) than the patients with high ARPP19 expression (n = 78).

In summary, better overall survival of low ARPP19 mRNA expressing AML patients' supports the observed lower relapse risk of these patients after standard therapy.

ARPP19 expression correlates with AML disease activity after remission

552

553 Finally, we wanted to study whether ARPP19 expression levels correlate with
554 emerge of relapse after patients have achieved clinical remission. For this
555 purpose we could identify bone marrow samples of nine patients from cohort2 for
556 which in addition to diagnostic samples, also follow-up samples at first remission
557 (n=4) or relapse (n=8) were available (Figure 4f). Three patients among these
558 nine had complete follow-up set of diagnosis, remission and relapse samples
559 (Figure 4g). Consistently with the overexpression of ARPP19 mRNA in cohort1
560 (Figure 1f), seven out of nine samples in follow-up series had higher ARPP19
561 mRNA expression than in the normal BM indicated by dashed line (Figure 4f).
562 Notably, ARPP19 expression dropped below the control level in remission
563 samples, whereas it was found overexpressed again in relapse samples (Figure
564 4f; diagnosis vs. remission p=0.021, remission vs. relapse p=0.034 by paired t-
565 test). These findings were confirmed in the complete matched set of samples
566 (diagnosis, remission and relapse) from three patients (Figure 4g; diagnosis vs.
567 remission p=0.047, remission vs. relapse p=0.034 by paired t-test).

568

569 These results provide an independent validation for association between high
570 ARPP19 expression, and emerge of relapse from standard AML therapy.

571

Discussion

Upon diagnosis of AML, multiple molecular markers are used to define the risk group for AML patients, but also for stratifying patients between chemotherapy and HSCT. Whereas risk grouping is sufficient to predict the relapse risk in large fraction of patients, some patients from favorable risk group yet relapse whereas not all intermediate-adverse risk group patients are suitable for HSCT. Therefore, a better understanding of risk grouping independent mechanisms that affect the AML relapse tendency would be of high medical relevance. In this study we identified ARPP19 as a novel oncogenic protein that is associated with AML relapse independently of risk groups, and ofother existing AML diagnostic markers. Further understanding of mechanisms by which ARPP19 promotes relapse tendency could lead to future patient stratification strategies to quide patients with low relapse risk to chemotherapy, whereas high relapse tendency patients (regardless of their genetic risk group) should be treated more intensively, such as with HSCT.

Decreased PP2A tumor suppressor activity due to an increased expression of PAIPs has been reported to promote malignant growth of several cell types [9,25], including leukemic cells[10]. In AML, SET promotes both malignant growth and drug resistance [17,26], and CIP2A inhibition in AML cells reduces proliferation and MYC expression [27]. Prevalent role for PP2A inhibition in AML [10], and in other cancer types [9,25], provides a strong scientific rationale for clinical association between low ARPP19 expression and a lower risk for AML relapse newly discovered in this study. In a direct support of oncogenic role of ARPP19 in AML, we demonstrate that ARPP19 knockdown decreased expression of a well-

598 validated oncogenic PP2A target MYC. Interestingly, our data also show that
599 ARPP19 positively regulates CIP2A protein expression even though we did not
600 observe any particular strong assocation between ARPP19 and CIP2A mRNA
601 expression in AML patient samples. This data suggests that similarly to CML [19,28],
602 CIP2A may be regulated at the protein level in AML. In fact, a recent study did
603 indicate that CIP2A protein levels function as a biomarker for AML [29]. Thereby,
604 further studies on regulation of CIP2A protein expression by ARPP19 in AML
605 cells are clearly warranted. The functional hierarchy between ARPP19 and
606 CIP2A proteins provides a plausible explanation why ARPP19 may have a
607 stronger clinical role than CIP2A in AML. This can be rationalized as ARPP19
608 can control both directly its own PP2A/B55-subunit targets[30], but also PP2A/B56-
609 subunit targets via CIP2A[22] (Figure 3c). Therapeutically, it is tempting to envision
610 that decreased PP2A activity due to ARPP19 overexpression could be restored
611 by blocking ARPP19 effects on PP2A. However, development of ARPP19
612 targeted therapies awaits for structural analysis of the ARPP19 protein.
613
614 In summary, our results identify ARPP19 as a potential novel AML oncoprotein.
615 Most importantly, ARPP19 gene expression, and its relapse predicting role were
616 found to be independent of the current genetic risk classification. This suggests
617 that better understanding of ARPP19 function in AML could provide clinically
618 relevant additional value to existing diagnostic and therapeutic approaches.
619

Ethics approval and consent to participate

Patient cohort1: Informed consent was obtained from all patients and the local Ethical Review Board of TYKS approved the study protocol.

Patient cohort2: The Finnish Hematology Registry and Clinical Biobank (FHRB) is authorized by the Finnish National Supervisory Authority for Welfare and Health (Valvira) and has been approved by the Finnish National Medical Ethics Committee. All patients signed an informed consent prior to biobanking.

Availability of data and material

For original data, please contact jukwes@utu.fi. Information about the datasets supporting the conclusions of this article are included within the article (and its additional files)."

Competing interests

J.W. and E.M have a patent pending for “ARPP19 as prognostic biomarker for haematological cancers” (FI20185436)

Funding

This work was supported by funding from The Sigrid Jusélius Foundation, Kalevi Soininen Foundation, Turku Doctoral Program of Molecular Medicine, University of Turku faculty of medicine, Turku University Hospital ERVA (13283) and Business Finland TUTL (2445/31/2017).

Authors' contributions

E.M., O.K. and J.W. conceived the study and experiments; E.M., O.K. and T.V. performed the experiments; E.M. and E.L. analyzed the data; U.S., M.I-R. and V.K. collected samples and data from leukemia patients; E.M. and J.W. wrote the

643 manuscript, with input from O.K., U.S., E.L., V.K. and M.I-R. All authors reviewed
644 and approved the final manuscript.

645 **Acknowledgements**

646 We want to thank Kimmo Porkka (Hematology Research Unit Helsinki, University
647 of Helsinki, Helsinki, Finland) for expert comments and for advice with FHRB
648 samples. Taina Kalevo-Mattila is acknowledged for superior technical assistance.
649 We thank Maria D. Odero (University of Navarra, Pamplona, Spain) for
650 constructive discussions and advice on the material and methodology related to
651 AML cell lines. We also thank Artur Padzik and The Viral Vector Facility at Turku
652 Centre for Biotechnology (Supported by Biocenter Finland) for technical
653 assistance with lentiviral work.

654

655 **References**

656

657 1. Döhner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl*
658 *J Med.* 2015;373(12):1136-1152.

659 2. Petersdorf SH, Kopecky KJ, Slovak M, et al. A phase 3 study of
660 gemtuzumab ozogamicin during induction and postconsolidation therapy in
661 younger patients with acute myeloid leukemia. *Blood.* 2013;121(24):4854-4860.

662 3. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification
663 and Prognosis in Acute Myeloid Leukemia. *The New England journal of medicine.*
664 2016;374(23):2209-2221.

665 4. Walter RB, Othus M, Burnett AK, et al. Resistance prediction in AML:
666 analysis of 4601 patients from MRC/NCRI, HOVON/SAKK, SWOG and MD
667 Anderson Cancer Center. *Leukemia.* 2015;29(2):312-320.

668 5. Parkin B, Ouillette P, Yildiz M, Saiya-Cork K, Shedden K, Malek SN.
669 Integrated genomic profiling, therapy response, and survival in adult acute
670 myelogenous leukemia. *Clin Cancer Res.* 2015;21(9):2045-2056.

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

674 7. Meeusen B, Janssens V. Tumor suppressive protein phosphatases in
675 human cancer: Emerging targets for therapeutic intervention and tumor
676 stratification. *Int J Biochem Cell Biol.* 2018;96:98-134.

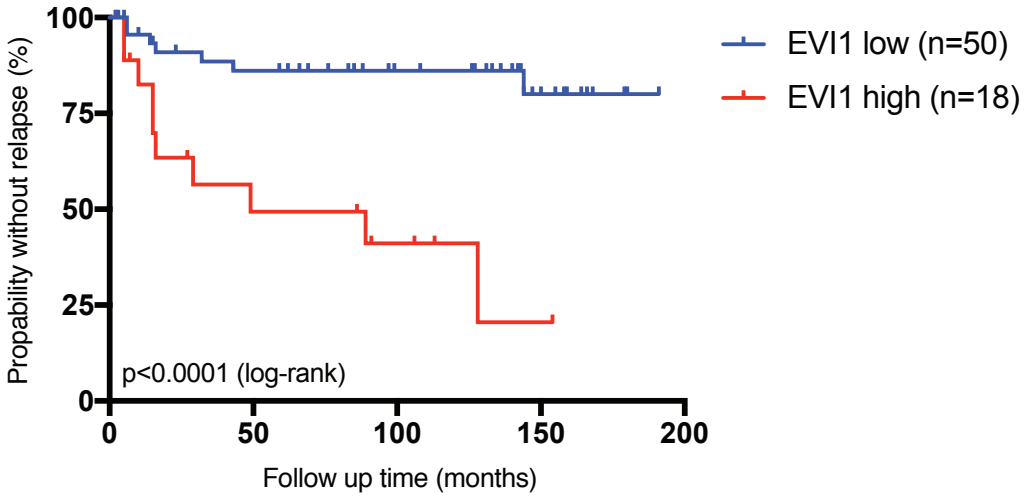
677 8. Perrotti D, Neviani P. Protein phosphatase 2A: a target for anticancer
678 therapy. *The lancet oncology.* 2013;14(6):e229-238.

679 9. Kauko O, Westermarck J. Non-genomic mechanisms of protein
680 phosphatase 2A (PP2A) regulation in cancer. *Int J Biochem Cell Biol.*
681 2018;96:157-164.
682 10. Arriazu E, Pippa R, Odero MD. Protein Phosphatase 2A as a Therapeutic
683 Target in Acute Myeloid Leukemia. *Frontiers in oncology.* 2016;6:78.
684 11. Gharbi-Ayachi A, Labbe JC, Burgess A, et al. The substrate of Greatwall
685 kinase, Arpp19, controls mitosis by inhibiting protein phosphatase 2A. *Science.*
686 2010;330(6011):1673-1677.
687 12. Jiang T, Zhao B, Li X, Wan J. ARPP-19 promotes proliferation and
688 metastasis of human glioma. *Neuroreport.* 2016;27(13):960-966.
689 13. Song H, Pan J, Liu Y, et al. Increased ARPP-19 expression is associated
690 with hepatocellular carcinoma. *Int J Mol Sci.* 2014;16(1):178-192.
691 14. Ujj Z, Buglyó G, Udvardy M, et al. WT1 Expression in Adult Acute Myeloid
692 Leukemia: Assessing its Presence, Magnitude and Temporal Changes as
693 Prognostic Factors. *Pathol Oncol Res.* 2016;22(1):217-221.
694 15. Hinai AA, Valk PJ. Review: Aberrant EVI1 expression in acute myeloid
695 leukaemia. *Br J Haematol.* 2016;172(6):870-878.
696 16. Cristóbal I, Garcia-Orti L, Cirauqui C, et al. Overexpression of SET is a
697 recurrent event associated with poor outcome and contributes to protein
698 phosphatase 2A inhibition in acute myeloid leukemia. *Haematologica.*
699 2012;97(4):543-550.
700 17. Agarwal A, MacKenzie RJ, Pippa R, et al. Antagonism of SET using
701 OP449 enhances the efficacy of tyrosine kinase inhibitors and overcomes drug
702 resistance in myeloid leukemia. *Clin Cancer Res.* 2014;20(8):2092-2103.

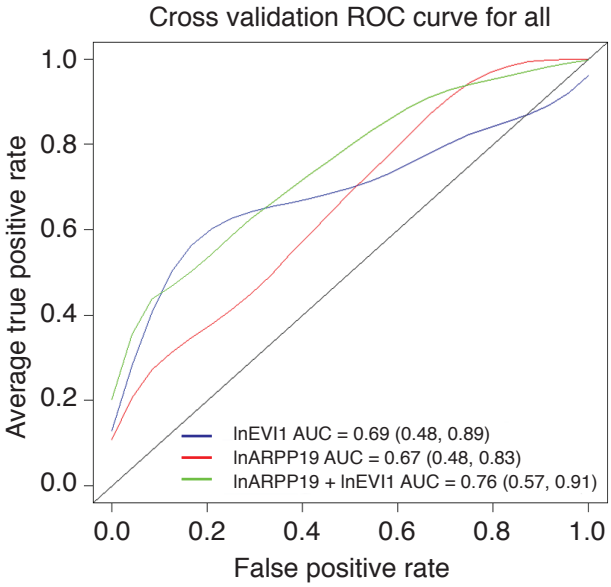
703 18. Li M, Makkinje A, Damuni Z. The myeloid leukemia-associated protein
704 SET is a potent inhibitor of protein phosphatase 2A. *J Biol Chem.*
705 1996;271(19):11059-11062.
706 19. Lucas CM, Harris RJ, Giannoudis A, Copland M, Slupsky JR, Clark RE.
707 Cancerous inhibitor of PP2A (CIP2A) at diagnosis of chronic myeloid leukemia is
708 a critical determinant of disease progression. *Blood.* 2011;117(24):6660-6668.
709 20. Junttila MR, Westermarck J. Mechanisms of MYC stabilization in human
710 malignancies. *Cell Cycle.* 2008;7(5):592-596.
711 21. Khanna A, Pimanda JE, Westermarck J. Cancerous inhibitor of protein
712 phosphatase 2A, an emerging human oncoprotein and a potential cancer therapy
713 target. *Cancer Res.* 2013;73(22):6548-6553.
714 22. Wang J, Okkeri J, Pavic K, et al. Oncoprotein CIP2A is stabilized via
715 interaction with tumor suppressor PP2A/B56. *EMBO reports.* 2017;18(3):437-
716 450.
717 23. Ley TJ, Miller C, Ding L, et al. Genomic and epigenomic landscapes of
718 adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22):2059-2074.
719 24. Goldman M, Craft B, Brooks AN, Zhu J, Haussler D. The UCSC Xena
720 Platform for cancer genomics data visualization and interpretation. In. Vol
721 https://doi.org/10.1101/326470. bioRxiv2018.
722 25. Khanna A, Pimanda JE. Clinical significance of Cancerous Inhibitor of
723 Protein Phosphatase 2A (CIP2A) in human cancers. *Int J Cancer.* 2015.
724 26. Yang Y, Huang Q, Lu Y, Li X, Huang S. Reactivating PP2A by FTY720 as
725 a novel therapy for AML with C-KIT tyrosine kinase domain mutation. *J Cell*
726 *Biochem.* 2012;113(4):1314-1322.

727 27. Barragán E, Chillón MC, Castelló-Cros R, et al. CIP2A high expression is
728 a poor prognostic factor in normal karyotype acute myeloid leukemia.
729 *Haematologica.* 2015;100(5):e183-185.
730 28. Lucas CM, Harris RJ, Holcroft AK, et al. Second generation tyrosine
731 kinase inhibitors prevent disease progression in high-risk (high CIP2A) chronic
732 myeloid leukaemia patients. *Leukemia.* 2015;29(7):1514-1523.
733 29. Lucas CM, Scott LJ, Carmell N, et al. CIP2A- and SETBP1-mediated
734 PP2A inhibition reveals AKT S473 phosphorylation to be a new biomarker in
735 AML. *Blood Adv.* 2018;2(9):964-968.
736 30. Hunt T. On the regulation of protein phosphatase 2A and its role in
737 controlling entry into and exit from mitosis. *Adv Biol Regul.* 2013;53(2):173-178.
738

a

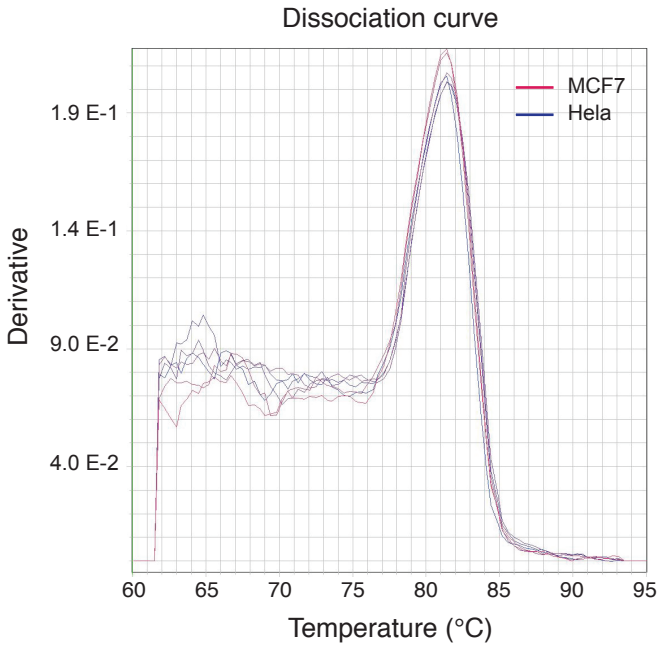


b

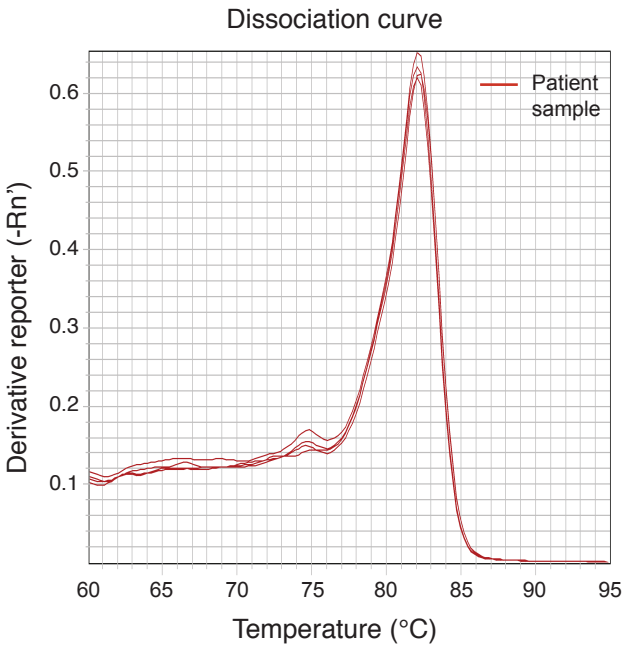


Supplemental Figure 3 Kaplan-Meier curve for time to relapse according to EVI1 mRNA expression in cohort1. a) High EVI1 expression is significantly associated with shorter relapse free time of AML patients, p<0.0001 by log-rank test. b) The ROC curve analysis on the abilities of ARPP19 and EVI1 gene expression at diagnosis in prediction of relapse in AML patients. AUC values and 95% confidence intervals are shown for each analysis.

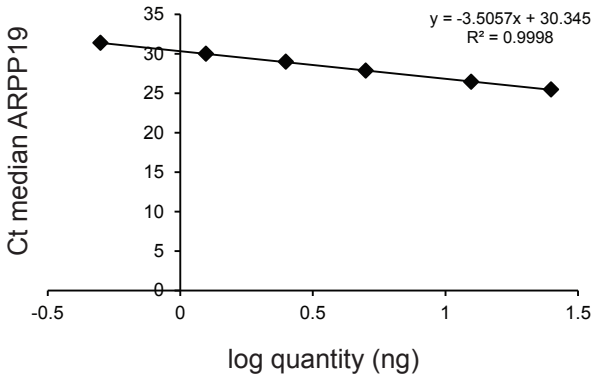
a



b

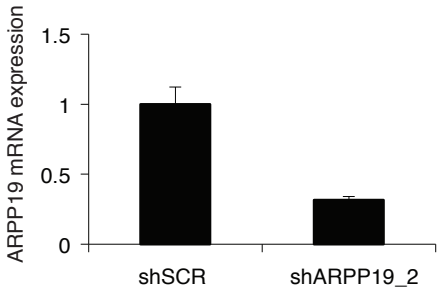


c



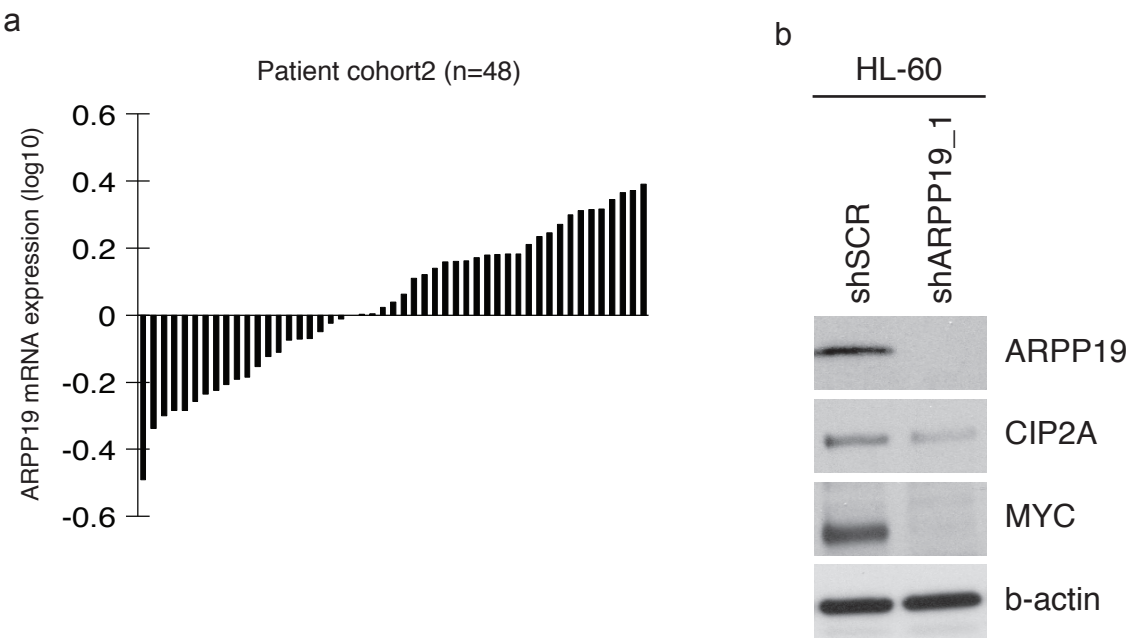
Means from two independent experiments:
ARPP19 amplification efficiency 93%
GAPDH amplification efficiency 93%
b-actin amplification efficiency 81%

d



ARPP19 mRNA levels from KG-1 cell line measured with ARPP19 qPCR normalized to shSCR

Supplemental Figure 1 Validation of ARPP19 specific RQ-PCR. a) and b) Melting curves from RQ-PCR of ARPP19 gene with primers indicated in Supplemental table 7. Amplicons from both MCF7 and Hela cell line samples, as well as diagnosis phase AML patient sample, reveal a single peak following melting curve analysis. c) Amplification efficiency of ARPP19 RQ-PCR used in this study was 93% when measured with standard curve analysis. Shown is representative standard curve, but similar results were extracted from two independent experiments. Control genes used in the study, GAPDH and b-actin, showed amplification efficiencies of 93% and 81%, respectively. d) RQ-PCR specificity validation of ARPP19 assay on ARPP19 shRNA transduced KG1 cells.



Supplemental Figure 2 Validation of ARPP19 overexpression in AML cohort2. a) Waterfall blot of ARPP19 mRNA expression in patient cohort2 (n=48) normalized to GAPDH & b-actin expression and a pooled normal bone marrow sample (n=56). ARPP19 was 58% (n=28) overexpressed in the sample panel. b) Representative western blot analysis of ARPP19, CIP2A and MYC expression in HL-60 cell line stably transduced with scrambled (shSCR) or ARPP19 shRNAs (shARPP19_1).

Table S1. Clinical and molecular characteristics of 80 patients with AML (cohort1)

Characteristic	No. (%)
Sex	
Male	39 (49)
Female	41 (51)
Age	
≤ 60 years	67 (84)
> 60 years	13 (16)
Complete remission	
No	12 (15)
Yes	68 (85)
Diagnosis	
AML-M0	5 (6)
AML-M1	21 (26)
AML-M2	12 (15)
AML-M3	0 (0)
AML-M4	7 (9)
AML-M5	15 (19)
AML-M6	1 (1)
AML-M7	1 (1)
Not specified	18 (23)
Cytogenetic group (AML 2012)	
Favorable	21 (26)
Intermediate	37 (46)
Adverse	22 (28)
Normal karyotype	
No	50 (63)
Yes	30 (37)
Secondary AML	
No	61 (76)
Yes	19 (24)
Extramedullary disease	
No	71 (89)
Yes	9 (11)
FLT3-ITD	
No	67 (84)
Yes	10 (12)
Not analyzed	3 (4)
NPM1 mutated	
No	58 (72)
Yes	19 (24)
Not analyzed	3 (4)
FLT3-ITD and NPM1 co-association	
FLT3-ITD+, NPM1mutation+	6 (8)
FLT3-ITD+, NPM1mutation-	4 (5)
FLT3-ITD-, NPM1mutation+	13 (16)
FLT3-ITD-, NPM1mutation-	54 (67)
Not analyzed	3 (4)

Table S2. Risk stratification that was applied in retrospective analyses (modified from European LeukemiaNet; Döhner et al. 2010)

Favorable
t(8;21)
Inv(16)/t(16;16)
Normal karyotype and NPM1 mutation without FLT3-ITD
Intermediate
Normal karyotype without NPM1 mutation and FLT3-ITD
Normal karyotype with NPM1 mutation and FLT3-ITD
Cytogenetic abnormaltied not classified as favourable or adverse
Adverse
Inv(3)/t(3;3)
t(6;9)
t(v;11)
-5, del(5q), -7, del(7q), abn(17p)
t(9;22)
Normal karyotype with FLT3-ITD
Complex karyotype (≥3 chromosome abnormalities)
Monosomal karyotype (≥2 monosomies or single monosomy with structural abnormality)

Table S3. Chemotherapy according to AML92 and AML2003 protocols

AML92	Days	**AML2003**	Days
1. IAT		**1. IdAraC-Ida vesrus IAT**	
Idarubicin 12 mg/m2	1, 3, 5	Cytarabine 1 g/m2	1-4
Cytarabine 100 mg/m2	1-9	Idarubicin 12 mg/m2	5-7
Thioguanine 75 mg/m2	1-9	(2nd induction MEA)	
2. HDAraC-Ida		**2. HDAraC-Ida**	
Cytarabine 1,5 mg/m2 every 12h	1-5	Cytarabine 1,5 mg/m2 every 12h	1-4
Idarubicin 8 mg/m2	6-8	Idarubicin 8 mg/m2	5-7
3. MEA		**3. Mito-IdAraC**	
Mitoxantrone 12 mg/m2	2-5	Mitoxantrone 12 mg/m2	2-5
Etoposide 100 mg/m2	1-4	Cytarabine 1-1.5 g/m2 every 12h	1-4
Cytarabine 1 g/m2 every 12h	1-4		
		4. MACE	
4. Amsa-HDAraC		Amsacrine 100 mg/m2	1-(4)5
Amsacrine 115 mg/m2	1-5	Cytarabine 100-200 mg/m2	1-5
Cytarabine 3 g/m2 every 12h	1-2	Etoposide 100 mg/m2	1-5
5. HDAraC-Ida		**5. IAE**	
If 2 induction cycles		If 2 induction cycles	

Table S4. Clinical and molecular characteristics of nine patients with AML (cohort2)

Characteristic	No. (%)
Sex	
Male	4 (44)
Female	5 (56)
Age	
≤ 60 years	5 (56)
> 60 years	4 (44)
Complete remission	
No	0 (0)
Yes	9 (100)
Cytogenetic group (AML 2012)	
Favorable	3 (33)
Intermediate	4 (45)
Adverse	2 (22)
Allogeneic HSCT	
No	5 (56)
Yes	4 (44)
Relapse	
No	1 (11)
Yes	8 (89)
FLT3-ITD	
No	5 (56)
Yes	2 (22)
Unknown	2 (22)
NPM1 mutated	
No	2 (22)
Yes	4 (45)
Unknown	3 (33)
FLT3-ITD and NPM1 co-association	
FLT3-ITD+, NPM1mutation+	1 (11)
FLT3-ITD+, NPM1mutation-	0 (0)
FLT3-ITD-, NPM1mutation+	3 (33)
FLT3-ITD-, NPM1mutation-	2 (23)
Not analyzed	3 (33)

Table S5. Primer and probe sequences used in this study for qPCR analysis

Target	Forward primer	Probe	Reverse primer
CIP2A	cagtctggactgagaatattattgga	tccactgc*	ggcattgtttgctgctatacttt
PPME1	acaggtttgcagaacccatc	tccagtgt*	ggacagcaggtcactaacagc
ARPP19	cagagggagcactatgtctgc	aggagcag*	gcttttaattttgcttcttctgct
TIPRL	catgatgatccacggcttc	ggccctgg*	tcagggagagatggcatatgta
EVI1 (MECOM)	agtgccctggagatgagttg	ccccagtgaggtataaagagga	tttgaggctatctgtgaagtgc
WT1	gggcgtgtgaccgtagct	agcacggtcaccttcgacggg	cgctattcgcaatcagggtta
b-actin	tcacccacacrgtgcccatctacgc	atgccctcccccatgccatcctgcgt	cagcggaaccgctcattgccaatgg
GAPDH	acccactcctccacctttga	acgaccactttgtcaagctcatttcctggt	ttgctgtagccaaattcgttgt

For **SET** detection, custom TaqMan® gene expression assay AIY9AXG (Thermo Fisher Scientific) was used.

* Roche Universal ProbeLibrary (UPL) probe

Table S6. Overexpression of WT1, EVI1, SET, TIPRL, ARPP19, CIP2A and PME1 in 80 patients with AML (cohort1)

Target	Median	Quartiles	Overexpression in patient cohort No. (%)	FC > 2 No. (%)
WT1	63.3	9.68, 151.3	73 (91)	70 (88)
EVI1	0.003	0.001, 0.067	10 (13)	9 (11)
SET	0.8	0.64, 1.08	24 (30)	6 (8)
TIPRL	0.84	0.65, 1.07	23 (30)	1 (1)
ARPP19	0.62	0.48, 0.91	16 (21)	3 (4)
CIP2A	0.35	0.18, 0.5	3 (4)	0 (0)
PME1	0.27	0.18, 0.48	3 (4)	1 (1)

Table S7. Clinical and genetic characteristics of AML patients according to ARPP19 expression (cohort1)

ARPP19 categorical expression					
	ARPP19low		ARPP19high		
Characteristic	**Median (range)**	**No. (%)**	**Median (range)**	**No. (%)**	**p**
Overall		60 (79)		16 (21)	
Sex					
Male		31 (52)		5 (31)	ns
Female		29 (48)		11 (69)	
Age	50 (18-64)		49 (24-65)		ns
Complete remission					
No		7 (12)		4 (25)	ns
Yes		53 (88)		12 (75)	
Diagnosis					
AML-M0		3 (5)		2 (12.5)	**0.02**
AML-M1		11 (18)		10 (62.5)	
AML-M2		10 (17)		2 (12.5)	
AML-M3		0 (0)		0 (0)	
AML-M4		7 (12)		0 (0)	
AML-M5		15 (25)		0 (0)	
AML-M6		1 (1.5)		0 (0)	
AML-M7		1 (1.5)		0 (0)	
Not specified		12 (20)		2 (12.5)	
Cytogenetic group (AML 2012)					
Favorable		15 (25)		6 (38)	ns
Intermediate		30 (50)		5 (31)	
Adverse		15 (25)		5 (31)	
Normal karyotype					
No		42 (70)		5 (31)	**0.005**
Yes		18 (30)		11 (69)	
Secondary AML					
No		46 (77)		13 (81)	ns
Yes		14 (23)		3 (19)	
Extramedullary disease					
No		52 (87)		15 (94)	ns
Yes		8 (13)		1 (6)	
FLT3-ITD					
No		52 (87)		13 (81)	ns
Yes		6 (10)		3 (19)	
Not analyzed		2 (3)		0 (0)	
NPM1 mutated					
No		48 (80)		7 (44)	**0.002**
Yes		10 (17)		9 (56)	
Not analyzed		2 (3)		0 (0)	
FLT3-ITD and NPM1 co-association					
FLT3-ITD+, NPM1mutation+		3 (5)		3 (19)	**0.03**
FLT3-ITD+, NPM1mutation-		3 (5)		0 (0)	
FLT3-ITD-, NPM1mutation+		7 (12)		6 (37)	
FLT3-ITD-, NPM1mutation-		45 (75)		7 (44)	
Not analyzed		2 (3)		0 (0)	