SUPPLEMENTARY INFORMATION

The following Cancer and Leukemia Group B (CALGB)/Alliance institutions, principal investigators, and cytogeneticists participated in this study:

The Ohio State University Medical Center, Columbus, OH: Clara D. Bloomfield, Karl S. Theil, Diane Minka and Nyla A. Heerema (grant no. U10CA077658); Wake Forest University School of Medicine, Winston-Salem, NC: David D. Hurd, Wendy L. Flejter and Mark J. Pettenati; University of Iowa Hospitals, Iowa City, IA: Daniel A. Vaena and Shivanand R. Patil; North Shore–Long Island Jewish Health System, Manhasset, NY: Daniel R. Budman and Prasad R. K. Koduru (grant no. U10CA035279); Roswell Park Cancer Institute, Buffalo, NY: Ellis G. Levine and AnneMarie W. Block (grant no. U10CA059518); Duke University Medical Center, Durham, NC: Jeffrey Crawford, Mazin B. Qumsiyeh, John Eyre and Barbara K. Goodman (grant no. U10CA047577); University of Chicago Medical Center, Chicago, IL: Hedy L. Kindler, Diane Roulston, Yanming Zhang and Michelle M. Le Beau (grant no. U10CA041287); University of North Carolina, Chapel Hill, NC: Thomas C. Shea and Kathleen W. Rao (grant no. U10CA047559); Ft. Wayne Medical Oncology/Hematology, Ft. Wayne, IN: Sreenivasa Nattam and Patricia I. Bader; Washington University School of Medicine, St. Louis, MO: Nancy L. Bartlett, Michael S. Watson, Peining Li, and Jaime Garcia-Heras (grant no. U10CA077440); Dana Farber Cancer Institute, Boston, MA: Harold J. Burstein, Leonard L. Atkins, Paola Dal Cin and Cynthia C. Morton (grant no. U10CA032291); Eastern Maine Medical Center, Bangor, ME: Thomas H. Openshaw and Laurent J. Beauregard; Vermont Cancer Center, Burlington, VT: Steven M. Grunberg, Elizabeth F. Allen and Mary Tang; University of Massachusetts Medical Center, Worcester, MA: William W.

Walsh and Vikram Jaswaney; Mount Sinai School of Medicine, New York, NY: Lewis R. Silverman and Vesna Najfeld; University of Puerto Rico School of Medicine, San Juan, PR: Eileen I. Pacheco, Leonard L. Atkins and Cynthia C. Morton; SUNY Upstate Medical University, Syracuse, NY: Stephen L. Graziano and Constance K. Stein; University of Tennessee Cancer Center, Memphis,TN: Harvey B. Niell and Sugandhi A. Tharapel; University of Illinois at Chicago: David J. Peace, Maureen M. McCorquodale and Kathleen E. Richkind; University of Nebraska Medical Center, Omaha, NE: Apar Ganti and Warren G. Sanger; University of Maryland Cancer Center, Baltimore, MD: Maria R. Baer and Yi Ning; Western Pennsylvania Hospital, Pittsburgh, PA: John Lister and Gerard R. Diggans; Dartmouth Medical School, Lebanon, NH: Konstantin Dragnev and Thuluvancheri K. Mohandas; Georgetown University Medical Center, Washington, DC: Minnetta C. Liu and Jeanne M. Meck; Long Island Jewish Medical Center CCOP, Lake Success, NY: Kanti R. Rai and Prasad R. K. Koduru; Massachusetts General Hospital, Boston, MA: Jeffrey W. Clark, Leonard L. Atkins and Cynthia C. Morton; University of Minnesota, Minneapolis, MN: Bruce A. Peterson and Betsy A. Hirsch; University of Missouri/Ellis Fischel Cancer Center, Columbia, MO: Carl E. Freter and Tim H. Huang; Medical University of South Carolina, Charleston, SC: Mark R. Green and Daynna J. Wolff; Rhode Island Hospital, Providence, RI: William Sikov and Shelly L. Kerman; Christiana Care Health Services, Inc., Newark, DE: Stephen S. Grubbs and Digamber S. Borgaonkar.

Patients and treatment

Eighty patients with untreated *de novo* acute myeloid leukemia (AML) and sole +8, who were enrolled on CALGB/Alliance study protocols, were investigated. Fifty-nine patients who received treatment in accordance with the study protocols and who could be evaluated for treatment response were included in the outcome analyses. Patients received cytarabine/daunorubicin-based induction and consolidation regimens, as detailed below. Per study protocols, no patient received allogeneic hematopoietic stem-cell transplantation (HSCT) in first complete remission (CR).

A cohort of 483 adults with molecularly characterized *de novo* cytogenetically normal (CN)-AML was used for comparisons of pretreatment characteristics, outcome and expression profiles.[1] In addition, a cohort of 329 *de novo* AML patients assigned to the ELN Intermediate-II Genetic Group [i.e., patients with t(9;11)(p22;q23) and patients with cytogenetic aberrations not classified in the ELN Favorable or Adverse Genetic Groups other than sole +8] were considered for outcome analyses. The CN-AML and ELN Intermediate-II patients were enrolled on the same treatment protocols as the sole +8 AML patients (detailed below).[1]

All patients provided written informed consent for participation in the studies, and all study protocols were in accordance with the Declaration of Helsinki and approved by Institutional Review Boards at each treatment center.

Of the patients included in the outcome analyses, those younger than 60 years were treated on CALGB/Alliance protocols 19808 (n=211), 9621 (n=158), 8525 (n=30), 10503

(n=12), 9022 (n=4), 9120 (n=1), 9222 (n=1) and 10603 (n=1). Patients 60 years or older were treated on CALGB 9720 (n=192), 10201 (n=132), 8923 (n=70), 8525 (n=58), 9420 (n=21) and 10502 (n=1). Patients enrolled on CALGB 19808 were assigned to receive induction chemotherapy consisting of cytarabine, daunorubicin, and etoposide with or without the multidrug resistance protein modulator PSC-833 (valspodar).[2] Upon achievement of complete remission (CR), patients received high-dose cytarabine and etoposide for stem-cell mobilization followed by myeloablative treatment with busulfan and etoposide supported by autologous peripheral HSCT. Patients not eligible for HSCT received high-dose cytarabine. After intensification, patients received low-dose recombinant interleukin-2 maintenance or no further treatment. Patients enrolled on CALGB 9621 were treated similarly to those on CALGB 19808.[3] Patients enrolled on CALGB 10503 were treated similarly to those on CALGB 19808, but received no PSC-833 and were administered the DNA methyltransferase inhibitor decitabine for maintenance.[4] Patients on CALGB 8525 were treated with induction chemotherapy consisting of cytarabine and daunorubicin, and were randomly assigned to consolidation with different doses of cytarabine followed by maintenance treatment.[5] Patients enrolled on CALGB 9022 received induction treatment with cytarabine and daunorubicin that was followed by three sequential courses of intensive postremission therapy.[6] The patient enrolled on CALGB 9120 received induction treatment consisting of idarubicin and cytarabine and post-remission therapy off study.[7] CALGB 9222 tested a multiagent regimen including high-dose cytarabine as consolidation after cytarabine and daunorubicin.[8] In CALGB 10603, cytarabine and daunorubicin followed by consolidation with high-dose cytarabine were applied with or without PKC-412.[9] Patients on CALGB 9720 and 9420 received induction chemotherapy consisting of cytarabine in

combination with daunorubicin and etoposide, with/without (CALGB 9720) or with (CALGB 9420) PSC-833.[10,11] Patients on CALGB 9420 received postremission therapy with cytarabine (2 g/m^2/d) alone, and patients on CALGB 9720 received a single cytarabine/daunorubicin consolidation course and were randomly assigned to low-dose recombinant interleukin-2 maintenance therapy or none. Patients on CALGB 10201 received induction chemotherapy consisting of cytarabine and daunorubicin, with or without the *BCL2* antisense oblimersen sodium. The consolidation included two cycles of cytarabine (2 g/m^2/d) with or without oblimersen.[12] In CALGB 10502, bortezomib was added to induction consisting of daunorubicin and cytarabine and consolidation with high-dose cytarabine.[13] Patients on CALGB 8923 were treated with induction therapy consisting of cytarabine and daunorubicin and were randomly assigned to receive postremission therapy with cytarabine alone or in combination with mitoxantrone.[14]

Sample collection

Patients enrolled on the treatment protocols also provided written informed consent to participate in the companion protocols CALGB 8461 (prospective cytogenetic companion), CALGB 9665 (leukemia tissue bank) and CALGB 20202 (molecular studies), which involved collection of pretreatment bone marrow (BM) aspirates and blood samples.

Cytogenetic analyses

Cytogenetic analyses at diagnosis were performed by CALGB-approved institutional cytogenetics laboratories as part of CALGB 8461, and the results confirmed by central

karyotype review.[15] The diagnosis of sole +8 was made from bone marrow (BM) in 76 (95%) and blood in 4 (5%) patients. At least 20 metaphase cells were analyzed in 77 (96%) sole +8 patients. Two patients had tetrasomy 8, and three had clones with trisomy 8 and tetrasomy 8. For each patient with +8 detected in BM, the percentage of metaphase cells containing +8 was calculated. For CN-AML patients, ≥20 metaphase cells from diagnostic BM had to be analyzed and the karyotype found to be normal.[15]

Analysis of molecular genetic markers

Mononuclear cells were enriched through Ficoll-Hypaque gradient centrifugation and cryopreserved until use. Genomic DNA and total RNA were extracted and quality control of the nucleic acids was performed as described previously.[16] Mutations were assessed centrally at The Ohio State University on genomic DNA from BM or blood with ≥20% blasts. Mutations in *NPM1* (ref. 17), *FLT3* [i.e., *FLT3*-internal tandem duplication (*FLT3*-ITD) and *FLT3*-tyrosine kinase domain mutation (*FLT3*-TKD)],[18,19] *CEBPA*,[16] *WT1* (ref. 20), *ASXL1* (ref. 21), *TET2* (ref. 22), *IDH1* (ref. 23), *IDH2* (ref. 23), *RUNX1* (ref. 24), and *DNMT3A*[25] were analyzed as previously reported. *NRAS* exons 2 and 3 (NM_002524) and *KRAS* exon 2 (NM_004985) were amplified and directly sequenced. *NRAS* exon 3 was analyzed by denaturing high-performance liquid chromatography before sequencing. The expression of *miR-155* was determined using Affymetrix arrays (Affymetrix, Santa Clara, CA) and of *miR-3151* using real-time reverse transcription-polymerase chain reaction (RT-PCR).[26,27] *BAALC* expression was assessed with RT-PCR (n=41)[28-30] and Affymetrix arrays (n=25). Because *BAALC* expression values obtained using the RT-PCR were strongly correlated with the values obtained from

Affymetrix platforms (r=0.82, *P*<0.001), we combined these measurements for statistical analyses, via regression imputation.

Definition of clinical end points

CR required an absolute neutrophil count ≥1,500/μL, platelet count ≥100,000/μL, no leukemic blasts in the blood, BM cellularity greater than 20% with maturation of all cell lines, no Auer rods, less than 5% BM blast cells, and no evidence of extramedullary leukemia, all of which had persisted for at least 1 month.[31] Relapse was defined by ≥5% BM blasts, 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; patients alive and relapse-free at last follow-up were censored. Overall survival (OS) was measured from the date on study until the date of death, and patients alive at last follow-up were censored.

Statistical analyses

All analyses were performed by the Alliance for Clinical Trials in Oncology Statistics and Data Center. The median follow-up for patients alive included in the outcome analyses was 4.7 years (range, 1.9-9.9 years). Pretreatment demographic, clinical, and molecular features were compared using the Fisher's exact and Wilcoxon rank-sum tests for categorical and continuous variables, respectively. Survival estimates were calculated using the Kaplan-Meier method, and the log-rank test evaluated differences between survival distributions. Logistic regression and Cox proportional hazards models were

generated to compare, respectively, CR achievement and survival (DFS, OS) end points. For these models, odds ratios were calculated for CR achievement and hazard ratios for survival, along with *P*-values from the Wald test.

Multivariable models in sole +8 AML patients

Multivariable logistic regression models and Cox proportional hazards models were constructed to analyze factors related to the probability of achieving CR and for survival end points (DFS and OS) respectively, using a limited backwards selection procedure. Variables considered for model inclusion were sex, race, hemoglobin, platelet count, white blood count (WBC), percentage of blood blasts, percentage of BM blasts, percentage of metaphases with +8 in BM (<80% vs ≥80%), *NPM1* (mutated vs wild-type), *RUNX1* (mutated vs wild-type), *ASXL1* (mutated vs wild-type), *FLT3*-ITD (present vs absent), *FLT3*-TKD (present vs absent), *IDH1* (mutated vs wild-type), *IDH2* (mutated vs wild-type), *TET2* (mutated vs wild-type), *DNMT3A* (mutated vs wild-type), *RAS* (*NRAS* or *KRAS* mutated vs *NRAS* and *KRAS* wild-type), *CEBPA* (mutated vs wild-type), *WT1* (mutated vs wild-type), *BAALC* expression (high vs low), *miR-155* expression (high vs low), and *miR-3151* expression (high vs low). Variables considered for inclusion were those significant at α=0.20 from the univariable models. For DFS and OS, the proportional hazards assumption was checked for each variable individually. If the proportional hazards assumption was not met for a particular variable, then an artificial time-dependent covariate was included in all models that contained that variable.[32]

Genome-wide gene- and microRNA-expression analyses

Gene-expression profiles were derived using U133 plus 2.0 arrays (Affymetrix) from pretreatment samples of 67 sole +8 (BM, n=53; blood, n=14) and 367 cytogenetically normal (CN) AML patients (BM, n=283; blood, n=84). Summary measures of gene expression were computed for each probe-set using the robust multichip average (RMA) method, which incorporates quantile normalization of arrays. Expression values were logged (base 2) before analysis. A filtering step was performed to remove probe-sets that did not display significant variation in expression across arrays; a gene was excluded when less than 20% of its expression data had at least a 1.5-fold change in either direction from the gene's median value, or percent of data missing exceeded 50%. Multiple probesets were reduced to one per gene symbol by using the maximally expressed probeset measured by average intensity across samples. A total of 7,899 genes passed the filtering criterion. A comparison of gene expression was made between sole +8 and CN-AML patients, using a univariable significance level of 0.001.

MicroRNA-expression profiles were assessed using custom microRNA arrays (OSU_CCC version 4.0) on pretreatment BM of sole +8 (n=53) and CN-AML patients (n=54). To account for batch effects, the microRNA profiles of sole +8 and CN-AML patients were analyzed in the same array batch. Due to limitations on material availability, the CN-AML cohort analyzed for microRNAs was restricted to patients aged ≥60 years. For microRNA-expression microarrays, signal intensities were calculated for each spot making an adjustment for local background. Intensities were log-transformed and log-intensities from replicate spots were averaged. Quantile normalization was

performed on arrays using all human microRNA-probes represented on the array. For each microRNA probe, an adjustment was made for batch effects (i.e., differences in expression related to the batch in which arrays were hybridized). Further analysis was limited to the 354 mature human microRNAs that passed filtering criteria. For microRNAs with more than one probe, the expression was determined as the average expression of the probes. A comparison of microRNA-expression was made between sole +8 and CN-AML patients using a univariable significance level of 0.005.

Analyses were performed using BRB-ArrayTools version 4.3.1 Stable Release developed by Dr. Richard Simon and Amy Peng Lam.

Gene set enrichment analysis (GSEA)

The GSEA analysis was performed to determine chromosomes whose genes were significantly up- or downregulated in sole +8 compared with CN-AML patients. Of the 7,899 filtered genes included in our study, 7,895 were studied in the GSEA. One thousand permutations were performed and a normalized enrichment score was used in conjunction with a familywise-error rate (FWER) p-value. FWER is a conservatively estimated probability that the normalized enrichment score represents a false positive finding.

Prediction of sole +8 status from expression profiles

We implemented the compound covariate predictor to predict sole +8 status of patients based on the expression patterns of 189 genes located on chromosome 8 and

significantly unregulated in sole +8 when compared to CN-AML.[33,34] A sample was classified to the sole +8 class if the compound covariate predictor for the 189 genes ($\sum_i w_i\ x_i$) was larger than the threshold of 7551.25. We estimated the prediction error of each model using leave-one-out cross-validation (LOOCV). For each LOOCV training set, the entire model building process was repeated, including the gene selection process. We also evaluated whether the cross-validated error rate estimate for a model was significantly less than one would expect from random prediction. The class labels were randomly permuted and the entire LOOCV process was repeated. The significance level is the proportion of the random permutations that gave a cross-validated error rate no greater than the cross-validated error rate obtained with the real data (1000 random permutations were used).

Supplementary References

1. Mrózek K, Marcucci G, Nicolet D, Maharry KS, Becker H, Whitman SP *et al*. Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. *J Clin Oncol* 2012; **30**: 4515-4523.
2. Kolitz JE, George SL, Marcucci G, Vij R, Powell BL, Allen SL *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**: 1413-1421.
3. Kolitz JE, George SL, Dodge RK, Hurd DD, Powell BL, Allen SL *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**: 4290-4301.
4. Blum W, Sanford B, Klisovic RB, DeAngelo DJ, Uy G, Powell BL *et al*. Maintenance therapy with decitabine in younger adults with acute myeloid leukemia (AML) in first remission: a phase II Cancer and Leukemia Group B study (CALGB 10503, Alliance). *Blood* 2012; **120**: (abstract 44).
5. Mayer RJ, Davis RB, Schiffer CA, Berg DT, Powell BL, Schulman P *et al*. Intensive postremission chemotherapy in adults with acute myeloid leukemia. *N Engl J Med* 1994; **331**: 896-903.
6. Moore JO, Dodge RK, Amrein PC, Kolitz J, Lee EJ, Powell B *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**: 780-788.

7. Cassileth PA, Harrington DP, Appelbaum FR, Lazarus HM, Rowe JM, Paietta E *et al*. Chemotherapy compared with autologous or allogeneic bone marrow transplantation in the management of acute myeloid leukemia in first remission. *New Engl J Med* 1998; **339**: 1649-1656.
8. Moore JO, George SL, Dodge RK, Amrein PC, Powell BL, Kolitz JE *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**: 3420-3427.
9. Stone RM, Dohner H, Ehninger G, Villeneuve M, Teasdale T, Virkus JD *et al*. CALGB 10603 (RATIFY): a randomized phase III study of induction (daunorubicin/cytarabine) and consolidation (high-dose cytarabine) chemotherapy combined with midostaurin or placebo in treatment-naive patients with FLT3 mutated AML. *J Clin Oncol* 2011; **29** (suppl): (abstract TPS199).
10. Lee EJ, George SL, Caligiuri M, Szatrowski TP, Powell BL, Lemke S *et al*. Parallel phase I studies of daunorubicin given with cytarabine and etoposide with or without the multidrug resistance modulator PSC-833 in previously untreated patients 60 years of age or older with acute myeloid leukemia: results of Cancer and Leukemia Group B study 9420. *J Clin Oncol* 1999; **17**: 2831-2839.
11. Baer MR, George SL, Sanford BL, Mrózek K, Kolitz JE, Moore JO *et al*. Escalation of daunorubicin and addition of etoposide in the ADE regimen in acute

myeloid leukemia patients aged 60 years and older: Cancer and Leukemia Group B study 9720. *Leukemia* 2011; **25**: 800-807.

12. Marcucci G, Moser B, Blum W, Stock W, Wetzler M, Kolitz JE *et al*. A phase III randomized trial of intensive induction and consolidation chemotherapy ± oblimersen, a pro-apopatientotic Bcl-2 antisense oligonucleotide in untreated acute myeloid leukemia patients >60 years old. *J Clin Oncol* 2007; **25**: 360s (abstract 7012).
13. Attar EC, Johnson JL, Amrein PC, Lozanski G, Wadleigh M, DeAngelo DJ *et al*. Bortezomib added to daunorubicin and cytarabine during induction therapy and to intermediate-dose cytarabine for consolidation in patients with previously untreated acute myeloid leukemia age 60 to 75 years: CALGB (Alliance) study 10502. *J Clin Oncol* 2012; **31**: 923-929.
14. Stone R, Berg D, George S, Dodge RK, Paciucci PA, Schulman PP *et al*. Postremission therapy in older patients with de novo acute myeloid leukemia: a randomized trial comparing mitoxantrone and intermediate-dose cytarabine with standard-dose cytarabine. *Blood* 2001; **98**: 548-553.
15. Mrózek K, Carroll AJ, Maharry K, Rao KW, Patil SR, Pettenati MJ *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**: 239-244.
16. Marcucci G, Maharry K, Radmacher MD, Mrózek K, Vukosavljevic T, Paschka P *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**: 5078-5087.

17. Becker H, Marcucci G, Maharry K, Radmacher MD, Mrózek K, Margeson D *et al*. Favorable prognostic impact of *NPM1* mutations in older patients with cytogenetically normal de novo acute myeloid leukemia and associated gene- and microRNA-expression signatures: a Cancer and Leukemia Group B study. *J Clin Oncol* 2010; **28**: 596-604.
18. Whitman SP, Maharry K, Radmacher MD, Becker H, Mrózek K, Margeson D *et al*. *FLT3* internal tandem duplication associates with adverse outcome and gene- and microRNA-expression signatures in patients 60 years of age or older with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* 2010; **116**: 3622-3626.
19. Whitman SP, Ruppert AS, Radmacher MD, Mrózek K, Paschka P, Langer C *et al*. *FLT3* D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with de novo cytogenetically normal acute myeloid leukemia lacking *FLT3* internal tandem duplications. *Blood* 2008; **111**: 1552-1559.
20. Becker H, Marcucci G, Maharry K, Radmacher MD, Mrózek K, Margeson D *et al*. Mutations of the Wilms tumor 1 gene (*WT1*) in older patients with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* 2010; **116**: 788-792.
21. Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrózek K *et al*. *ASXL1* mutations identify a high-risk subgroup of older patients with primary

cytogenetically normal AML within the ELN Favorable genetic category. *Blood* 2011; **118**: 6920-6929.

22. Metzeler KH, Maharry K, Radmacher MD, Mrózek K, Margeson D, Becker H *et al*. *TET2* mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol* 2011; **29**: 1373-1381.
23. Marcucci G, Maharry K, Wu Y-Z, Radmacher MD, Mrózek K, Margeson D *et al*. *IDH1* and *IDH2* gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol* 2010; **28**: 2348-2355.
24. Mendler JH, Maharry K, Radmacher MD, Mrózek K, Becker H, Metzeler KH *et al*. *RUNX1* mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and microRNA expression signatures. *J Clin Oncol* 2012; **30**: 3109-3118.
25. Marcucci G, Metzeler KH, Schwind S, Becker H, Maharry K, Mrózek K *et al*. Age-related prognostic impact of different types of *DNMT3A* mutations in adults with primary cytogenetically normal acute myeloid leukemia. *J Clin Oncol* 2012; **30**: 742-750.
26. Marcucci G, Maharry KS, Metzeler KH, Volinia S, Wu Y-Z, Mrózek K *et al*. Clinical role of microRNAs in cytogenetically normal acute myeloid leukemia: *miR-155* upregulation independently identifies high-risk patients. *J Clin Oncol* 2013; **31**: 2086-2093.
27. Eisfeld A-K, Marcucci G, Maharry K, Schwind S, Radmacher MD, Nicolet D *et a*l. *miR-3151* interplays with its host gene *BAALC* and independently affects

outcome of patients with cytogenetically normal acute myeloid leukemia. *Blood* 2012; **120**: 249-258.

28. Baldus CD, Tanner SM, Ruppert AS, Whitman SP, Archer KJ, Marcucci G *et al*. *BAALC* expression predicts clinical outcome of de novo acute myeloid leukemia patients with normal cytogenetics: a Cancer and Leukemia Group B Study. *Blood* 2003; **102**: 1613-1618.
29. Langer C, Radmacher MD, Ruppert AS, Whitman SP, Paschka P, Mrózek K *et al*. High *BAALC* expression associates with other molecular prognostic markers, poor outcome, and a distinct gene-expression signature in cytogenetically normal patients younger than 60 years with acute myeloid leukemia: a Cancer and Leukemia Group B (CALGB) study. *Blood* 2008; **111**: 5371-5379.
30. Schwind S, Marcucci G, Maharry K, Radmacher MD, Mrózek K, Holland KB *et al*. *BAALC* and *ERG* expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* 2010; **116**: 5660-5669.
31. Cheson BD, Cassileth PA, Head DR, Schiffer CA, Bennett JM, Bloomfield CD *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**: 813-819.
32. Klein JP, Moeschberger MP. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer-Verlag: New York, NY, 1997.
33. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. *J Comput Biol* 2002; **9**: 505-511.

34. Simon R. Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. *J Natl Cancer Inst* 2006; **98**: 1169-1171.

Supplementary Table S1. Comparison of pretreatment characteristics of all and among younger (<60 years) and older (≥60 years) patients with sole +8 and cytogenetically normal *de novo* acute myeloid leukemia within age groups

Characteristic	*Sole +8 AML (n = 80)*	*CN-AML (n = 483)*	*P*	*Sole +8 <60 years (n = 40)*	*CN-AML <60 years (n = 240)*	*P*	*Sole +8 ≥60 years (n = 40)*	*CN-AML ≥60 years (n = 243)*	*P*
Age, years			0.69			0.45			0.04
Median	59	60		43	46		71	68	
Range	18-84	18-83		18-59	18-59		60-84	60-83	
Male sex, no. (%)	50 (63)	237 (49)	0.03	26 (65)	113 (47)	0.04	24 (60)	124 (51)	0.31
Race, no. (%)			0.56			1.00			0.26
White	69 (87)	428 (90)		35 (90)	210 (88)		34 (85)	218 (91)	
Non-white	10 (13)	50 (10)		4 (10)	28 (12)		6 (15)	22 (9)	
Hemoglobin (g/dL)			0.36			0.91			0.15
Median	9.2	9.3		9.2	9.2		9.1	9.4	
Range	5.0-15.8	4.6-15.0		5.0-15.8	4.6-13.6		5.3-14.1	5.4-15.0	
Platelet count, x 10^9/L			<0.001			0.09			0.003
Median	46	64		49	61		41	69	
Range	5-233	4-850		11-148	5-466		5-233	4-850	
WBC, x 10^9/L			<0.001			0.68			<0.001
Median	8.9	23.7		20.7	22.4		4.2	25.6	
Range	0.6-302.3	0.7-450.0		0.6-302.3	0.7-295.0		0.8-187.0	0.9-450.0	
Blood blasts, %			0.18			0.82			0.07
Median	38	54		49	57		23	45	
Range	0-97	0-99		2-97	0-97		0-91	0-99	
Bone marrow blasts, %			0.98			0.20			0.32
Median	70	65		77	65		56	67	
Range	18-94	4-97		22-90	10-95		18-94	4-97	
FAB, no. (%)[a]			0.07			<0.001			0.12
M0	3 (5)	11 (3)	(M1/M2 vs	1 (3)	7 (4)	(M1/M2 vs	2 (7)	4 (2)	(M1/M2 vs
M1	9 (15)	93 (26)	M4/M5)	2 (6)	55 (28)	M4/M5)	7 (26)	38 (23)	M4/M5)
M2	17 (29)	117 (33)		6 (19)	67 (35)		11 (41)	50 (30)	
M4	9 (15)	81 (23)		5 (16)	43 (22)		4 (15)	38 (23)	
M5	19 (32)	49 (14)		17 (53)	19 (10)		2 (7)	30 (18)	
M6	2 (3)	7 (2)		1 (3)	3 (2)		1 (4)	4 (2)	
Extramedullary involvement, no. (%)	15 (20)	120 (25)	0.39	9 (24)	67(29)	0.70	6 (16)	53 (22)	0.52

Supplementary Table S1. Continued.

Characteristic	*Sole +8 AML (n = 80)*	*CN-AML (n = 483)*	*P*	*Sole +8 <60 years (n = 40)*	*CN-AML <60 years (n = 240)*	*P*	*Sole +8 ≥60 years (n = 40)*	*CN-AML ≥60 years (n = 243)*	*P*
RUNX1, no. (%)			<0.001			0.08			<0.001
Mutated	25 (32)	51 (13)		7 (18)	14 (8)		18 (46)	37 (17)	
Wild-type	54 (68)	344 (87)		33 (82)	160 (92)		21 (54)	184 (83)	
ASXL1, no. (%)			<0.001			0.04			<0.001
Mutated	22 (29)	46 (11)		5 (13)	7 (4)		17 (46)	39 (17)	
Wild-type	54 (71)	379 (89)		34 (87)	182 (96)		20 (54)	197 (83)	
FLT3-ITD, no. (%)			0.52			0.72			0.14
Present	23 (29)	159 (33)		15 (38)	81 (34)		8 (20)	78 (32)	
Absent	57 (71)	323 (67)		25 (62)	158 (66)		32 (80)	165 (68)	
IDH2, no. (%)			0.09			0.60			0.08
Mutated	21 (26)	79 (18)	(Mutated vs	6 (15)	23 (12)	(Mutated vs	15 (38)	56 (23)	(Mutated vs
R140 mutated	13	65	wild-type)	4	19	wild-type)	9	46	wild-type)
R172 mutated	8	14		2	4		6	10	
Wild-type	59 (74)	357 (82)		34 (85)	173 (88)		25 (62)	184 (77)	
DNMT3A, no. (%)			0.16			0.13			0.69
Mutated	17 (25)	142 (34)	(Mutated vs	8 (22)	64 (35)	(Mutated vs	9 (29)	78 (33)	(Mutated vs
R882	12	92	wild-type)	7	47	wild-type)	5	45	wild-type)
Non-R882	5	50		1	17		4	33	
Wild-type	51 (75)	273 (66)		29 (78)	117 (65)		22 (71)	156 (67)	
NPM1, no. (%)			<0.001			<0.001			<0.001
Mutated	18 (22.5)	280 (61)		12 (30)	140 (64)		6 (15)	140 (58)	
Wild-type	62 (77.5)	181 (39)		28 (70)	78 (36)		34 (85)	103 (42)	
FLT3-TKD, no. (%)			0.07			0.39			0.06
Present	13 (17)	43 (9)		5 (13)	19 (9)		8 (21)	24 (10)	
Absent	65 (83)	415 (91)		34 (87)	196 (91)		31 (79)	219 (90)	
IDH1, no. (%)			0.72			0.58			0.22
Mutated	11 (14)	54 (12)		3 (8)	23 (12)		8 (20)	31 (13)	
Wild-type	69 (86)	382 (88)		37 (92)	173 (88)		32 (80)	209 (87)	
TET2, no. (%)			0.01			0.002			0.68
Mutated	8 (11)	104 (24)		0 (0)	34 (18)		8 (24)	70 (30)	
Wild-type	64 (89)	323 (76)		39 (100)	156 (82)		25 (76)	167 (70)	

Supplementary Table S1. Continued.

Characteristic	*Sole +8 AML (n = 80)*	*CN-AML (n = 483)*	*P*	*Sole +8 <60 years (n = 40)*	*CN-AML <60 years (n = 240)*	*P*	*Sole +8 ≥60 years (n = 40)*	*CN-AML ≥60 years (n = 243)*	*P*
CEBPA, no. (%)			0.08			0.005			1.00
Mutated	6 (8)	70 (15)	(Mutated vs	1 (3)	41 (19)	(Mutated vs	5 (13)	29 (12)	(Mutated vs
Single mutated	4	31	wild-type)	1	13	wild-type)	3	18	wild-type)
Double mutated	2	39		0	28		2	11	
Wild-type	74 (92)	385 (85)		39 (97)	171 (81)		35 (87)	214 (88)	
WT1, no. (%)			0.18			0.39			0.48
Mutated	3 (4)	40 (9)		2 (5)	24 (11)		1 (3)	16 (7)	
Wild-type	77 (96)	413 (91)		38 (95)	186 (89)		39 (97)	227 (93)	

Abbreviations: WBC, white blood count; FAB, French-American-British classification; *FLT3*-ITD, internal tandem duplication of the *FLT3* gene; *FLT3*-TKD, tyrosine kinase domain mutations of the *FLT3* gene.

[a] FAB morphology was centrally reviewed.

Supplementary Table S2. Pretreatment characteristics of patients with *de novo* acute myeloid leukemia and sole +8 included in outcome analyses

Characteristic	*Sole +8 AML n=59*	*Sole +8 AML <60 years (n=30)*	*Sole +8 AML ≥60 years (n=29)*	*P*[d]
Age, years				--
Median	59	43	71	
Range	23-84	23-59	60-84	
Male sex, no. (%)	35 (59)	19 (63)	16 (55)	0.60
Race, no. (%)				0.51
White	49 (83)	26 (87)	23 (79)	
Non-white	10 (17)	4 (13)	6 (21)	
Hemoglobin, g/dL				0.76
Median	9.3	9.3	9.3	
Range	5.9-14.1	6.3-13.1	5.9-14.1	
Platelet count, x10^9/L				0.60
Median	51	58	47	
Range	8-233	14-148	8-233	
WBC, x10^9/L				0.005
Median	11.2	24.9	5.9	
Range	0.6-268.0	0.6-268.0	0.8-54.4	
Blood blasts, %				0.04
Median	38	54	26	
Range	0-96	2-96	0-91	
Bone marrow blasts, %				0.16
Median	68	71	54	
Range	18-94	22-90	18-94	
FAB, no. (%)[a]				0.01
M0	2 (5)	1 (4)	1 (6)	(M1/M2 vs
M1	5 (12)	2 (8)	3 (17)	M4/M5)
M2	14 (33)	5 (21)	9 (50)	
M4	7 (17)	4 (17)	3 (17)	
M5	13 (31)	12 (50)	1 (6)	
M6	1 (2)	0 (0)	1 (6)	
Extramedullary Involvement, no. (%)	9 (16)	4 (14)	5 (19)	0.73
+8 metaphases in BM, no. (%)				0.79
≥ 80%	38 (64)	20 (67)	18 (62)	
RUNX1, no. (%)				0.03
Mutated	18 (31)	5 (17)	13 (45)	
Wild-type	41 (69)	25 (83)	16 (55)	
ASXL1, no. (%)				<0.001
Mutated	14 (25)	1 (3)	13 (48)	
Wild-type	42 (75)	28 (97)	14 (52)	
FLT3-ITD, no. (%)				0.01
Present	18 (31)	14 (47)	4 (14)	
Absent	41 (69)	16 (53)	25 (86)	
IDH2, no. (%)				0.05
Mutated	17 (29)	5 (17)	12 (41)	(Mutated vs
R140 mutated	12	4	8	wild-type)
R172 mutated	5	1	4	
Wild-type	42 (71)	25 (83)	17 (59)	

Supplementary Table S2. Continued.

Characteristic	*Sole +8 AML n=59*	*Sole +8 AML <60 years (n=30)*	*Sole +8 AML ≥60 years (n=29)*	*P* [d]
DNMT3A, no. (%)				1.00 (Mutated vs wild-type)
Mutated	13 (24)	7 (25)	6 (23)	
R882	10	6	4	
Non-R882	3	1	2	
Wild-type	41 (76)	21 (75)	20 (77)	
NPM1, no. (%)				0.36
Mutated	14 (24)	9 (30)	5 (17)	
Wild-type	45 (76)	21 (70)	24 (83)	
FLT3-TKD, no. (%)				0.33
Present	12 (21)	4 (14)	8 (28)	
Absent	46 (79)	25 (86)	21 (72)	
IDH1, no. (%)				0.42
Mutated	6 (10)	2 (7)	4 (14)	
Wild-type	53 (90)	28 (93)	25 (86)	
RAS, no. (%)				0.05
Mutated	7 (12)	1 (3)	6 (21)	
Wild-type	52 (88)	29 (97)	23 (79)	
TET2, no. (%)				.002
Mutated	8 (14)	0 (0)	8 (29)	
Wild-type	50 (86)	30 (100)	20 (71)	
CEBPA, no. (%)				0.05 (Mutated vs wild-type)
Mutated	4 (7)	0 (0)	4 (14)	
Single mutated	3	0	3	
Double mutated	1	0	1	
Wild-type	55 (93)	30 (100)	25 (86)	
WT1, no. (%)				0.49
Mutated	2 (3)	2 (7)	0 (0)	
Wild-type	57 (97)	28 (93)	29 (100)	
BAALC expression, no. (%) [b]				1.00
High	24 (48)	11 (50)	13 (46)	
Low	26 (52)	11 (50)	15 (54)	
miR-155 expression, no. (%) [c]				0.77
High	24 (50)	12 (55)	12 (46)	
Low	24 (50)	10 (45)	14 (54)	
miR-3151 expression, no. (%) [b]				1.00
High	21 (50)	8 (50)	13 (50)	
Low	21 (50)	8 (50)	13 (50)	

Abbreviations: WBC, white blood count; FAB, French-American-British classification; *FLT3*-ITD, internal tandem duplication of the *FLT3* gene; *FLT3*-TKD, tyrosine kinase domain mutations of the *FLT3* gene.

[a] FAB morphology was centrally reviewed.

[b] The median expression value was used as a cut point. It was calculated based on the expression levels assessed by RT-PCR.

[c] The median expression value was used as a cut point. It was calculated based on the expression levels on the Affymetrix array.

[d] *P*-values compare patients who are <60 years vs ≥60 years. For baseline continuous variables the Wilcoxon rank sum test was used, for baseline categorical variables the Fisher's exact test was used.

Supplementary Table S3. Outcome comparison of patients with *de novo* acute myeloid leukemia stratified by age group

Patients < 60 years	*Sole +8 (n=30)*	*CN-AML (n=240)*	*P* *OR/HR (95% CI)*	*t(9;11) (n=17)*	*Remaining ELN Int-II (n=121)*	*P* *OR/HR (95% CI)*	*P* *OR/HR (95% CI)*	*P* *OR/HR (95% CI)*
	I	*II*	*I vs II*	*III*	*IV*	*I vs III*	*I vs IV*	*III vs IV*
Complete remission, no. (%)	18 (60)	201 (84)	0.003 0.29 (0.13-0.65)	14 (82)	95 (79)	0.12 0.32 (0.08-1.36)	0.04 0.41 (0.18-0.96)	0.72 1.28 (0.34-4.78)
Disease-free survival			0.01 1.95 (1.16-3.30)			0.02 3.11 (1.21-8.00)	0.26 1.37 (0.79-2.36)	0.05 0.44 (0.19-1.01)
Median, years	1.1	2.4		NA	1.1			
% Disease-free at 3 years (95% CI)	21 (6-42)	47 (40-53)		57 (28-78)	32 (23-41)			
% Disease-free at 5 years (95% CI)	14 (3-35)	43 (36-49)		57 (28-78)	28 (20-38)			
Overall survival			0.003 1.94 (1.26-3.00)			0.02 2.53 (1.13-5.65)	0.05 1.58 (1.00-2.50)	0.20 0.62 (0.30-1.29)
Median, years	1.5	3.6		NA	2.0			
% Alive at 3 years (95% CI)	29 (14-46)	51 (45-57)		53 (28-73)	44 (35-52)			
% Alive at 5 years (95% CI)	19 (7-36)	46 (39-52)		53 (28-73)	38 (29-46)			
Patients ≥ 60 years	*Sole +8 (n=29)*	*CN-AML (n=243)*	*P* *OR/HR (95% CI)*	*t(9;11) (n=12)*	*Remaining ELN Int-II (n=179)*	*P* *OR/HR (95% CI)*	*P* *OR/HR (95% CI)*	*P* *OR/HR (95% CI)*
	I	*II*	*I vs II*	*III*	*IV*	*I vs III*	*I vs IV*	*III vs IV*
Complete remission, no. (%)	20 (69)	168 (69)	0.99 0.99 (0.43-2.28)	11 (92)	105 (59)	0.15 0.20 (0.02-1.81)	0.30 1.57 (0.68-3.63)	0.052 7.75 (0.98-61.35)
Disease-free survival			0.09 1.51 (0.95-2.41)			0.26 0.65 (0.31-1.36)	0.20 1.37 (0.85-2.23)	0.02 2.11 (1.12-3.96)
Median, years	0.6	0.9		0.5	0.5			
% Disease-free at 3 years (95% CI)	5 (0-21)	17 (12-23)		0	12 (7-19)			
% Disease-free at 5 years (95% CI)	5 (0-21)	13 (9-19)		0	7 (3-13)			
Overall survival			0.48 1.15 (0.78-1.71)			0.24 0.66 (0.34-1.31)	0.73 0.93 (0.63-1.39)	0.26 1.40 (0.78-2.53)
Median, years	1.2	1.2		0.8	0.9			
% Alive at 3 years (95% CI)	17 (6-33)	20 (15-25)		8 (1-31)	15 (10-21)			
% Alive at 5 years (95% CI)	10 (3-24)	14 (10-18)		8 (1-31)	7 (4-11)			

Abbreviations: OR, odds ratio; HR, hazard ratio; CI, confidence interval; Remaining ELN Int-II, European Leukemia Net Intermediate-II Genetic Group [i.e., t(9;11)(p22;q23) or cytogenetic abnormalities not classified as favorable or adverse without sole +8]; NA, not achieved.

Supplementary Table S4. Outcome comparison of younger patients with *de novo* AML and sole +8 according to *FLT3*-ITD mutation status, and of older patients with *de novo* AML and sole +8 according to *TET2* mutation status.

Sole +8 AML patients < 60 years	*FLT3-ITD positive (n=14)*	*FLT3-ITD negative (n=16)*	*P*[a] *OR/HR (95% CI)*
Complete remission, no. (%)	6 (43)	12 (75)	0.08 0.25 (0.05-1.18)
Disease-free survival			0.054 2.87 (0.98-8.40)
Median, years	0.5	1.4	
% Disease-free at 3 years (95% CI)	0	33 (10-59)	
% Disease-free at 5 years (95% CI)	0	22 (4-49)	
Overall survival			0.02 2.64 (1.14-6.10)
Median, years	0.7	2.4	
% Alive at 3 years (95% CI)	14 (2-37)	42 (17-65)	
% Alive at 5 years (95% CI)	7 (0-28)	31 (9-57)	
Sole +8 AML patients ≥ 60 years	*TET2 mutated (n=8)*	*TET2 wild-type (n=20)*	*P*[a] *OR/HR (95% CI)*
Complete remission, no. (%)	3 (38)	16 (80)	0.04 0.15 (0.03-0.91)
Disease-free survival			0.048 3.87 (1.01-14.78)
Median, years	0.2	0.7	
% Disease-free at 3 years (95% CI)	0	6 (0-25)	
% Disease-free at 5 years (95% CI)	0	6 (0-25)	
Overall survival			0.004 3.70 (1.51-9.11)
Median, years	0.4	1.4	
% Alive at 3 years (95% CI)	0	25 (9-45)	
% Alive at 5 years (95% CI)	0	15 (4-33)	

Abbreviations: OR, odds ratio; HR, hazard ratio; CI, confidence interval.
[a] *P*-values compare patients who are *FLT3*-ITD positive vs negative or *TET2* mutated vs wild-type. For CR, the Wald test was used from the logistic regression model. For overall and disease-free survival, the Wald test was used from the Cox regression models.

Supplementary Table S5. List of genes differentially expressed between sole +8 *versus* cytogenetically normal *de novo* acute myeloid leukemia patients

see separate Excel file

Supplementary Table S6. MicroRNAs in the expression signature comparing sole +8 and cytogenetically normal patients with *de novo* acute myeloid leukemia

Upregulated in sole +8 AML			*Downregulated in sole +8 AML*		
microRNA	*Chromosomal location*	*Fold-change*	*microRNA*	*Chromosomal location*	*Fold-change*
miR-34b	11q23.1	1.58	*miR-496*	14q32.3	0.55
miR-370	14q32.2	1.54	*miR-493*	14q32.2	0.80
miR-107	10q23.3	1.46	-	-	-
miR-342-3p	14q32.2	1.41	-	-	-
miR-96	7q32.2	1.35	-	-	-

Supplementary Table S7. microRNAs located exclusively on chromosome 8[a]

microRNA	*Fold-change: sole +8/CN*	*P-value*
miR-30b	1.68	0.05
miR-486-5p	1.63	0.07
miR-30d	1.44	0.07
miR-320a	1.24	0.12
miR-383	1.07	0.67
miR-151-3p	1.03	0.51
miR-599	1.01	0.44
miR-596	0.94	0.11
miR-661	0.93	0.73
*miR-30b**	0.85	0.07
miR-598	0.79	0.08

Abbreviation: CN, cytogenetically normal.
[a] microRNAs with gene copies located on different chromosomes are not included in the Table. The fold-changes and *P*-values pertain to the comparison of sole +8 *versus* cytogenetically normal *de novo* acute myeloid leukemia patients.

Supplementary Figure Legends

Supplementary Figure S1. Distribution of mutations among 80 patients with *de novo* AML and sole +8. Each row represents a mutation, and each column a patient. Red indicates a mutation, light gray, wild-type, and dark gray, undetermined. In the last row, the age group each patient belonged to is displayed, with light blue indicating age <60 years and dark blue, ≥60 years. For *CEBPA*, it is indicated whether a single (1) or double (2) mutation was present. At the bottom, asterisks indicate patients with a tetrasomy 8, and the number of mutations per each patient is shown in a separate graph underneath.

Supplementary Figure S2. Overall survival of (**a**) younger patients with *de novo* AML and sole +8 according to *FLT3*-ITD mutation status, and (**b**) older patients with *de novo* AML and sole +8 according to *TET2* mutation status. Outcome data are also provided in Supplementary Table S4.

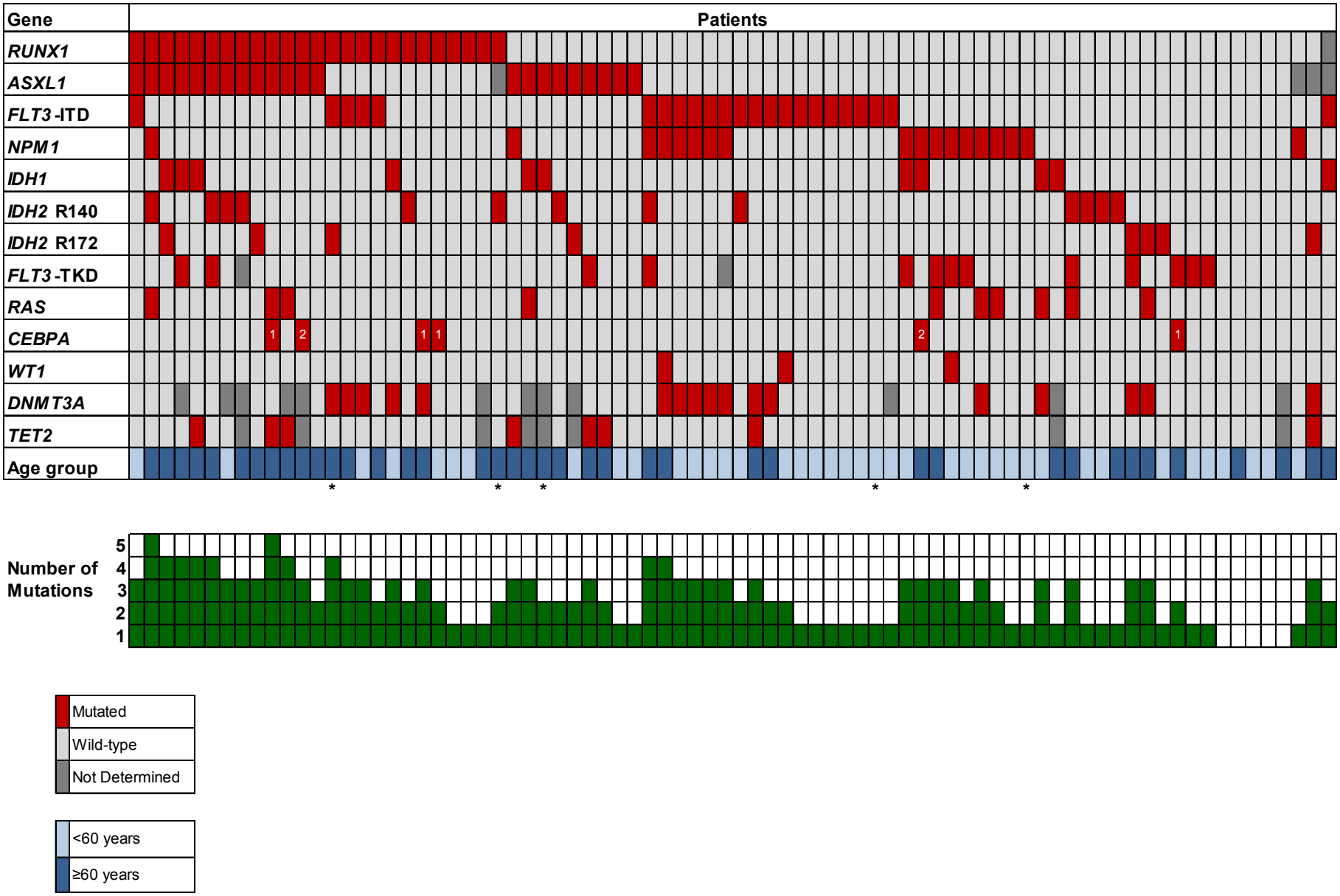
Supplementary Figure S3. Chromosomal localization of genes comprising the gene-expression signature, and expression patterns of genes located on chromosome 8. (**a**) Chromosome distribution of genes significantly differentially expressed between *de novo* AML patients with sole +8 and those with cytogenetically normal AML (CN-AML). The bars represent the proportion of genes mapped to the pertinent chromosome of all filtered genes analyzed by gene-expression profiling (blue bars), the genes significantly upregulated (red bars) and the genes significantly downregulated (green bars) in sole +8 AML compared with CN-AML. One gene was omitted from the graph due to missing

information on chromosome location. (**b**) Chromosome 8 expression map. The sole +8 versus CN-AML fold-changes are plotted on a log scale for each filtered gene (n=309) mapped to chromosome 8 according to its location on chromosome 8. All chromosome 8-genes (i.e., not only those included in the signature) are displayed. Each gene is represented by a dot. Dots may superimpose due to the density of genes.

Supplementary Figure S4. Gene Ontology (GO) analysis of the 81 genes whose expression was ≥1.5-fold higher in sole +8 versus CN-AML. Cytoscope and BinGO were used to evaluate the significance of the cellular processes and molecular functions associated with the genes over-expressed in +8 AML. BiNGO determines which GO categories are statistically overrepresented in a set of genes. It uses the hypergeometric test (without replacement) to provide an accurate answer in the form of a *P*-value, and it maps the predominant functional themes of a given gene set on the GO hierarchy as a Cytoscape graph. The False Discovery Rate (FDR), i.e., the expected proportion of false positives among the positively identified tests is controlled using the Benjamini & Hochberg correction, which provides strong control over the FDR under positive regression dependency of the null hypotheses. The genes overrepresented in +8 AML and associated with response to chemical stimulus (corrected *P*-value=4.75×10^{-3}) were *LPL*, *EIF2C2*, *LY96*, *GGH*, *FOXO1*, *GJA1*, *ABCB1*, *AK4*, *DPYSL2*, *RRM2B*, *SDC2*, *DDIT4*, *THBD*, *EPS8*, *ADM*, *ID1*, *ALDH2*, *CTSB*, and *ADAM9*; and genes associated with organ development were *EGR3*, *FHL1*, *FOXO1*, *GJA1*, *NID1*, *DPYSL2*, *RRM2B*, *TPD52*, *FZD6*, *HOOK3*, *TNS3*, *PTK2*, *APP*, *ADM*, *ID1*, *ALDH2*, *CTSB*, *LAMC1*, *EMP1*, and *ADAM9*. The graph visualizes the GO categories significantly overrepresented in the context of the GO hierarchy. The size (area) of the nodes is proportional to the

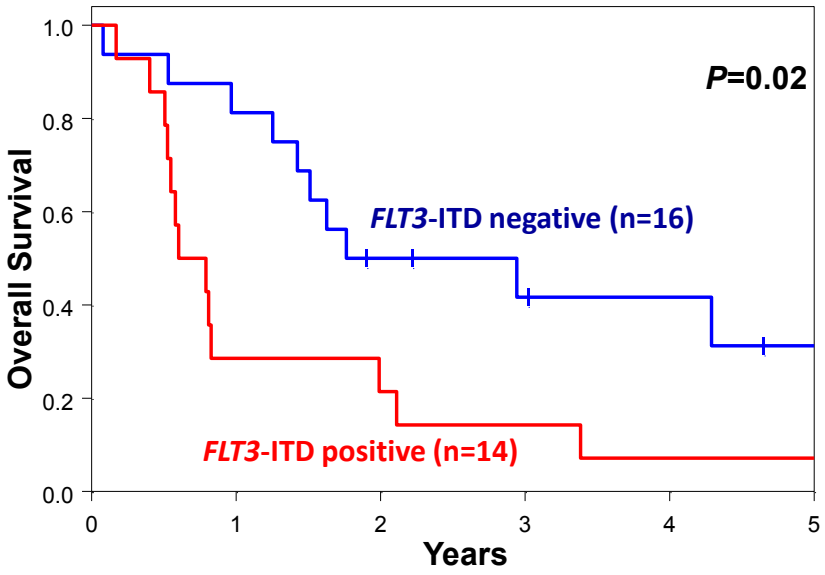
number of genes in the test set which are annotated to that node. The color of the node represents the corrected *P*-value. White nodes are not significantly overrepresented, the other ones are, with a color scale ranging from yellow (*P*-value <=0.05) to dark orange (*P*-value $<10^{-6}$). In addition to the biological processes displayed in the graph, genes associated with extracellular matrix organization (corrected *P*-value=4.75×10^{-3}) were *NID1*, *SPARC*, *PLEKHA2*, and *ADAM9*.

Supplementary Figure S1

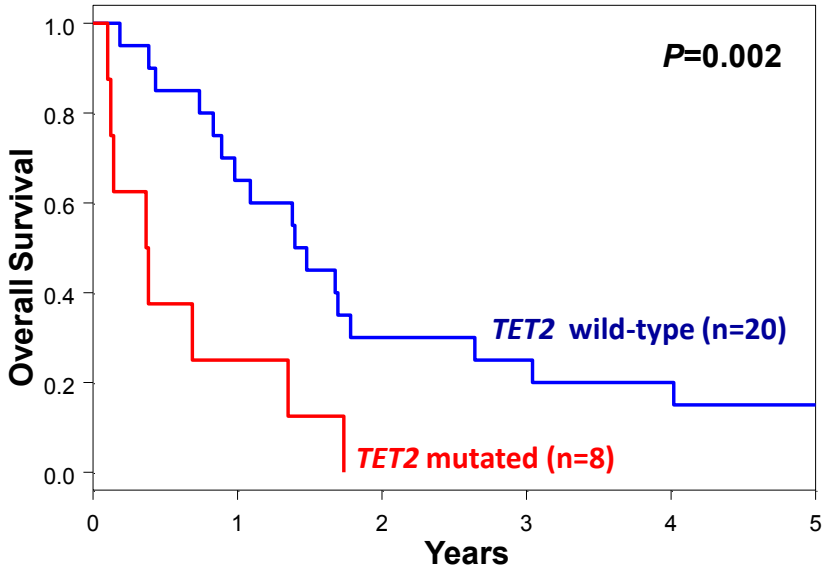


Supplementary Figure S2

a

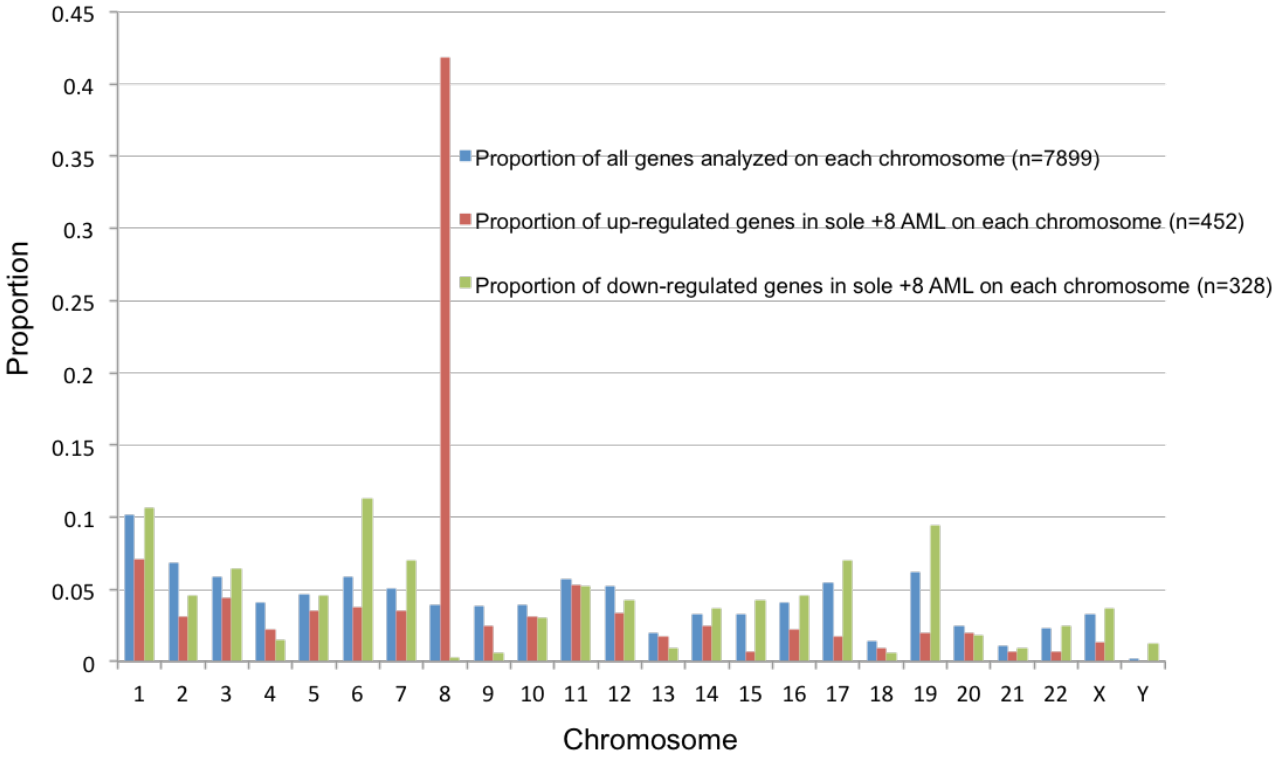


b

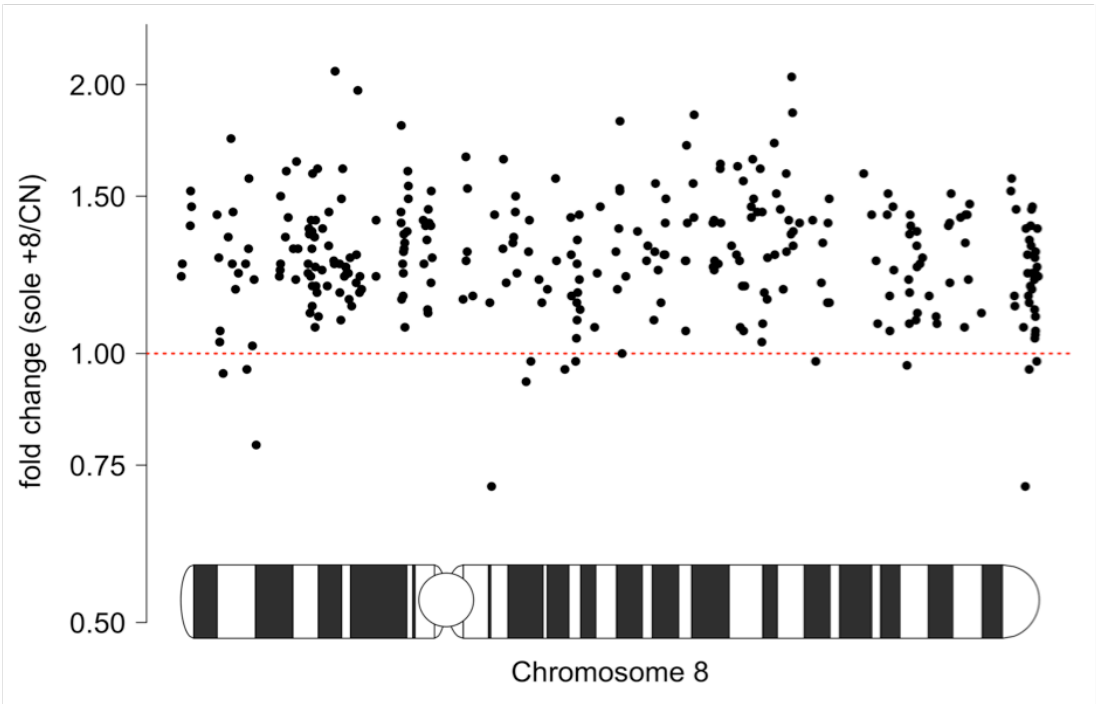


Supplementary Figure S3

a



b



Supplementary Figure S4