

Published in final edited form as:

N Engl J Med. 2009 September 10; 361(11): 1058–1066. doi:10.1056/NEJMoa0903840.

Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome

Elaine R. Mardis, Ph.D., Li Ding, Ph.D., David J. Dooling, Ph.D., David E. Larson, Ph.D., Michael D. McLellan, B.S., Ken Chen, Ph.D., Daniel C. Koboldt, M.S., Robert S. Fulton, M.S., Kim D. Delehaunty, B.A., Sean D. McGrath, M.S., Lucinda A. Fulton, M.S., Devin P. Locke, Ph.D., Vincent J. Magrini, Ph.D., Rachel M. Abbott, B.S., Tammi L. Vickery, B.S., Jerry S. Reed, M.S., Jody S. Robinson, M.S., Todd Wylie, B.S., Scott M. Smith, Lynn Carmichael, B.S., James M. Eldred, Christopher C. Harris, B.S., Jason Walker, B.A., B.S., Joshua B. Peck, M.B.A., Feiyu Du, M.S., Adam F. Dukes, B.A., Gabriel E. Sanderson, B.S., Anthony M. Brummett, Eric Clark, Joshua F. McMichael, B.S., Rick J. Meyer, M.S., Jonathan K. Schindler, B.S., B.A., Craig S. Pohl, M.S., John W. Wallis, Ph.D., Xiaoqi Shi, M.S., Ling Lin, M.S., Heather Schmidt, B.S., Yuzhu Tang, M.D., Carrie Haipek, M.S., Madeline E. Wiechert, M.S., Jolynda V. Ivy, M.B.A., Joelle Kalicki, B.S., Glendoria Elliott, Rhonda E. Ries, M.A., Jacqueline E. Payton, M.D., Ph.D., Peter Westervelt, M.D., Ph.D., Michael H. Tomasson, M.D., Mark A. Watson, M.D., Ph.D., Jack Baty, B.A., Sharon Heath, William D. Shannon, Ph.D., Rakesh Nagarajan, M.D., Ph.D., Daniel C. Link, M.D., Matthew J. Walter, M.D., Timothy A. Graubert, M.D., John F. DiPersio, M.D., Ph.D., Richard K. Wilson, Ph.D., and Timothy J. Ley, M.D.

Departments of Genetics (E.R.M., L.D., V.J.M., R.K.W., T.J.L.), Medicine (R.E.R., P.W., M.H.T., S.H., W.D.S., D.C.L., M.J.W., T.A.G., J.F.D., T.J.L.), and Pathology and Immunology (J.E.P., M.A.W., R.N.); the Genome Center (E.R.M., L.D., D.J.D., D.E.L., M.D.M., K.C., D.C.K., R.S.F., K.D.D., S.D.M., L.A.F., D.P.L., V.J.M., R.M.A., T.L.V., J.S. Reed, J.S. Robinson, T.W., S.M.S., L.C., J.M.E., C.C.H., J.W., J.B.P., F.D., A.F.D., G.E.S., A.M.B., E.C., J.F.M., R.J.M., J.K.S., C.S.P., J.W.W., X.S., L.L., H.S., Y.T., C.H., M.E.W., J.V.I., J.K., G.E., M.A.W., R.K.W., T.J.L.); Siteman Cancer Center (P.W., M.H.T., M.A.W., S.H., W.D.S., R.N., D.C.L., M.J.W., T.A.G., J.F.D., R.K.W., T.J.L.); and the Division of Biostatistics (J.B.) — all at Washington University, St. Louis

Abstract

BACKGROUND—The full complement of DNA mutations that are responsible for the pathogenesis of acute myeloid leukemia (AML) is not yet known.

METHODS—We used massively parallel DNA sequencing to obtain a very high level of coverage (approximately 98%) of a primary, cytogenetically normal, de novo genome for AML with minimal maturation (AML-M1) and a matched normal skin genome.

RESULTS—We identified 12 acquired (somatic) mutations within the coding sequences of genes and 52 somatic point mutations in conserved or regulatory portions of the genome. All mutations appeared to be heterozygous and present in nearly all cells in the tumor sample. Four of the 64 mutations occurred in at least 1 additional AML sample in 188 samples that were tested.

Mutations in *NRAS* and *NPM1* had been identified previously in patients with AML, but two other

Copyright © 2009 Massachusetts Medical Society.

Address reprint requests to Dr. Ley at Washington University, 660 S. Euclid Ave., Campus Box 8007, St. Louis, MO 63110, or at timley@wustl.edu.

Dr. Westervelt reports receiving lecture fees from Celgene and Novartis; and Dr. DiPersio, receiving consulting and lecture fees from Genzyme. No other potential conflict of interest relevant to this article was reported.

mutations had not been identified. One of these mutations, in the *IDH1* gene, was present in 15 of 187 additional AML genomes tested and was strongly associated with normal cytogenetic status; it was present in 13 of 80 cytogenetically normal samples (16%). The other was a nongenic mutation in a genomic region with regulatory potential and conservation in higher mammals; we detected it in one additional AML tumor. The AML genome that we sequenced contains approximately 750 point mutations, of which only a small fraction are likely to be relevant to pathogenesis.

CONCLUSIONS—By comparing the sequences of tumor and skin genomes of a patient with AML-M1, we have identified recurring mutations that may be relevant for pathogenesis.

Acute myeloid leukemia (AML) is a clonal hematopoietic disease caused by both inherited and acquired genetic alterations.¹⁻³ Current AML classification and prognostic systems incorporate genetic information but are limited to known abnormalities that have previously been identified with the use of cytogenetics, array comparative genomic hybridization (CGH), gene-expression profiling, and the resequencing of candidate genes (see the Glossary).

The karyotyping of AML cells remains the most powerful predictor of the outcome in patients with AML and is routinely used by clinicians.^{4,5} As an adjunct to cytogenetic studies, small subcytogenetic amplifications and deletions can be identified with the use of genomic methods, such as single-nucleotide-polymorphism (SNP) array and array CGH platforms (see the Glossary). However, these techniques remain investigational, and studies⁶⁻⁹ suggest that there are few recurrent acquired copy-number alterations in each AML genome. Gene-expression profiling has identified patients with known chromosomal lesions and genetic mutations and subgroups of patients with normal cytogenetic profiles who have variable clinical outcomes.^{10,11} Expression profiling has yielded single-gene predictors of outcome that are currently being evaluated for clinical use.¹²⁻¹⁶ Candidate-gene resequencing studies have also identified recurrent mutations in several genes — for example, genes encoding FMS-related tyrosine kinase 3 (*FLT3*) and nucleophosmin 1 (*NPM1*) — that can help to stratify patients with normal cytogenetic profiles according to risk and to identify patients for targeted therapy (e.g., those with mutated *FLT3*).^{3,12,17} However, the revised classification systems are imperfect, suggesting that important genetic factors for the pathogenesis of AML remain to be discovered.

We have previously described the sequence of an entire AML genome from a patient who had AML with minimal maturation (AML-M1) and a normal cytogenetic profile.¹⁸ Here we describe the genome sequence of another such tumor and recurring mutations in additional AML tumors.

METHODS

Details regarding the methods for library production, DNA sequencing with the Illumina Genome Analyzer II,¹⁹ evaluation of sequence coverage, identification of sequence variants, validation of variants and determination of the prevalence of variants in the index AML tumor, and screening of additional AML samples are provided in the Supplementary Appendix, available with the full text of this article at NEJM.org. All the high-quality single-nucleotide variants (SNVs) that were found in tumor and skin samples from this patient are available in the database of genotypes and phenotypes (dbGaP) of the National Center for Biotechnology Information (accession number, phs000159.v1.p1).

RESULTS

CASE REPORT

A previously healthy 38-year-old man of European ancestry presented with fatigue and a cough. The white-cell count was 39,800 cells per cubic millimeter, with 97% blasts; the hemoglobin level was 8.9 g per deciliter, and the platelet count was 35,000 per cubic millimeter. A bone marrow examination revealed 90% cellularity and 86% myeloperoxidase-positive blasts (Fig. 1 in the Supplementary Appendix). Routine cytogenetic analysis of bone marrow samples revealed a normal 46,XY karyotype. There was no family history of leukemia. The patient's mother had received the diagnosis of breast cancer at the age of 60 years and of non-Hodgkin's lymphoma at the age of 63 years; her half-sister had received the diagnosis of breast cancer at the age of 50 years.

Samples of the patient's bone marrow and skin were banked for whole-genome sequencing under a protocol approved by the institutional review board at Washington University. The patient provided written informed consent.

The patient was treated initially with a 7-day course of infusional cytarabine and with a 3-day course of daunorubicin. Within 5 weeks, he had complete morphologic remission and recovery of white-cell and platelet counts. The patient subsequently received consolidation therapy with four cycles of high-dose cytarabine without any further antileukemic therapy. He remained in complete remission 3 years later.

CHARACTERIZATION OF THE TUMOR GENOME

DNA samples from the patient's bone marrow sample at the time of initial presentation and a normal skin-biopsy specimen obtained after the patient's disease was in remission were labeled and genotyped with the use of the Affymetrix Genome-Wide Human SNP Array 6.0. The tumor genome had no detectable somatic copy-number alterations and no regions of partial uniparental disomy (Glossary, and Fig. 2 in the Supplementary Appendix). RNA that was derived from the same bone marrow sample was analyzed with the use of the Affymetrix GeneChip Human Genome U133 Plus 2.0 array, which revealed an expression signature similar to that of many other cytogenetically normal marrow samples from patients with AML-M1 (Fig. 2 in the Supplementary Appendix).

SEQUENCE COVERAGE AND POTENTIAL MUTATIONS

We sequenced 69.9 billion base pairs (23.3× haploid coverage) from DNA libraries that we generated from the tumor sample and 63.9 billion base pairs from libraries that we generated from the normal skin sample (21.3× haploid coverage) (Glossary and Table 1). Using Affymetrix 6.0 SNP arrays, we confirmed the detection of both alleles of 98.5% of the approximately 45,000 high-quality heterozygous SNPs in the tumor sample and 97.4% of the approximately 45,000 high-quality heterozygous SNPs in the skin sample.

A summary of the sequence differences between the patient's tumor genome and National Center for Biotechnology Information build 36 of the human reference genome is shown in Figure 1 (see the Glossary).²⁰ We identified 3,872,936 SNVs in the tumor genome, of which 3,464,449 passed a stringent calling filter. Of these SNVs, 3,377,680 (97.5%) were detected in the skin genome, indicating that they were inherited variants. Of the 86,769 potentially novel somatic SNVs, 66,513 had been described previously.

We binned the remaining 20,256 SNVs into four tiers, which are detailed in the Supplementary Appendix. Briefly, tier 1 contains all changes in the amino acid coding regions of annotated exons, consensus splice-site regions, and RNA genes (including

microRNA genes). Tier 2 contains changes in highly conserved regions of the genome or regions that have regulatory potential. Tier 3 contains mutations in the nonrepetitive part of the genome that does not meet tier 2 criteria, and tier 4 contains mutations in the remainder of the genome. We tentatively identified 113 potential tier 1 mutations, 749 potential tier 2 mutations, 3188 potential tier 3 mutations, and 16,206 potential tier 4 mutations. For each of the 113 putative tier 1 variants, we amplified the genomic region containing the mutation from both tumor and skin, using a polymerase-chain-reaction (PCR) assay, and performed Sanger sequencing. Of the 101 variants that were called with low confidence (the calling algorithm is summarized in the Supplementary Appendix), none were validated. Of the high-confidence variants, 10 of 12 were validated as somatic mutations. Similarly, we tested 178 low-confidence calls for tier 2, and only one was validated. In contrast, 51 of 104 high-confidence tier 2 calls were validated. We did not carry out validation studies of variants in tiers 3 and 4.

We also searched for somatic insertions and deletions (indels) using an algorithm described in the Supplementary Appendix. We identified 142 potential somatic indels (28 deletions and 114 insertions). Of these variants, 119 failed validation (i.e., they were falsely positive) in Sanger sequencing of the relevant PCR products, 21 were validated but were present in both tumor and skin, and 2 were validated as somatic mutations. One was a 4-bp insertion in exon 12 of the *NPM1* gene associated with aberrant cytoplasmic expression of nucleophosmin (*NPMc*). This insertion creates a frameshift mutation and a truncated protein that is known to have altered cellular localization, as described previously.²¹ The second mutation was a 3-bp insertion in the gene encoding centrosomal protein 170kDa (*CEP170*) at amino acid 177, predicted to result in the addition of a leucine residue at this position.

TIER 1 MUTATIONS

The genes with tier 1 mutations and the consequences of these mutations are summarized in Table 2, and in Table 1 in the Supplementary Appendix. Both the *NPMc* insertion and the *NRAS* mutation have been described previously in AML genomes, and both are known to be relevant for pathogenesis.³ Mutations in *IDH1* (encoding isocitrate dehydrogenase 1), which are predicted to affect the arginine residue at position 132, are found in malignant gliomas but have not been reported in patients with AML and are rare in other tumor types.²²⁻²⁴ Variants of the nine other tier 1 genes are discussed in the Supplementary Appendix.

Each of the 10 point mutations was amplified from tumor and skin samples by means of PCR, and the DNA species carrying the variant allele was assayed by sequencing the PCR products with the use of the Illumina platform. The entire experiment was replicated with amplified genomic DNA, with excellent concordance for all samples (Fig. 3 in the Supplementary Appendix). The variant allele frequencies of the two insertions were determined by sequencing PCR products containing these mutations. The representation of all but two of the mutations — in chromosome 19 open reading frame 62 (*C19orf62*), an unannotated gene of unknown function, and *CEP170* — was approximately 50%, suggesting that all the mutations were heterozygous and present in nearly all the cells in the tumor sample (Fig. 2A). Ten of the 12 genes in tier 1 had probe sets on the Affymetrix U133 Plus 2.0 array, and 9 of 10 were detectably expressed (Table 1). We also assayed expression of the 10 nonsynonymous mutant alleles by means of reverse-transcriptase PCR, using amplicons designed to span introns, followed by sequencing and counting of the sequenced PCR products. Eight of the mutant alleles were detected at frequencies of 35 to 85%. However, for two of the mutations (in *FREM2* and *IMPG2*) we did not detect complementary DNA carrying the variant allele (although we easily detected the wild-type allele), even though each variant was present in approximately 50% of the tumor DNA.

The individual bases that were mutated were highly conserved for 10 of the 12 variants, and all but 1 were found in highly conserved regions of the genome. The Sorting Intolerant from Tolerant (SIFT) algorithm (which gauges the likely effect of genic mutations on protein function) predicted that the mutations in *NRAS*, *IDH1*, *IMPG2*, and *ANKRD26* were deleterious.²⁵ The splice-site mutation at the 3' end of intron 4 of *C19orf62* caused exon 5 to be skipped (data not shown).

We then genotyped the tier 1 mutations in 187 additional samples from patients with AML whose clinical characteristics have been described previously²⁶ (Table 2 in the Supplementary Appendix). The *NPMc* mutation was previously shown to be present in 43 of 180 samples (23.9%), and activating *NRAS* mutations were present in 17 of 182 samples (9.3%).²⁶ We observed mutations in *IDH1*, which were predicted to cause substitution of the arginine residue at position 132, in 16 of 188 samples: R132C in 8 samples, R132H in 7 samples, and R132S in 1 sample (Table 2 in the Supplementary Appendix). The other nine mutations were not detected in the 187 additional samples. We detected no R172 mutations in *IDH2* in 188 samples (the sample from the index patient and the 187 additional samples), nor did we observe additional mutations in any of the exons of *IDH1* or *CDC42*.

A nonsynonymous acquired mutation (C328Y) was found in the mitochondrial gene *ND4*, which encodes NADH dehydrogenase subunit 4, a part of complex 1 of the electron transport chain. Two of 93 additional AML samples also had nonsynonymous mutations in this gene, but the importance of these mutations is not yet clear (Table 5 and the Results and Discussion section in the Supplementary Appendix).

TIER 2 MUTATIONS

We confirmed 52 mutations in tier 2. DNA segments, each containing 1 of the 52 mutations, were PCR-amplified from the tumor and skin samples and sequenced to determine the proportion of DNA molecules carrying the mutation (Fig. 2B, and Table 4 in the Supplementary Appendix). Three of these tier 2 mutations had variant frequencies of approximately 98%, and all were located on chromosome X or Y. Because only a single copy of these chromosomes was present in this male genome, the high representation of these three tier 2 mutations was consistent with the finding that an extremely high percentage of cells within the bone marrow sample were part of the malignant clone. One mutation (chromosome 4 at position 128,102,994) had a variant read frequency of approximately 78%, and we observed no somatic microamplification or deletion near this variant. Of the tier 2 mutations, 39 were present in approximately 50% of DNA species, and 9 were present in approximately 40%. We genotyped the 52 tier 2 mutations in 187 additional AML samples and detected the presence of just 1 of the mutations (on chromosome 10) in 1 other AML sample, from a patient with myelomonocytic leukemia (AML-M4), which bore a translocation and did not have a paired normal sample (Table 2 in the Supplementary Appendix). The proportion of DNA species in this sample that carried the mutation was 54%, suggesting that it was heterozygous.

PATIENTS WITH THE *IDH1* MUTATION

Of the 16 patients who had AML with an *IDH1* R132 mutation, 13 had tumors with normal cytogenetic profiles (of a total of 80 cytogenetically normal samples [16%]), 2 had trisomy 8, and 1 had trisomy 13. Ten of the 16 patients had AML-M1, three had AML with maturation (AML-M2), and three had AML-M4. The characteristics of patients with and those without the *IDH1* mutation are shown in Table 3, and in Tables 2 and 3 in the Supplementary Appendix. The mutation was detected only in patients with cytogenetic profiles associated with intermediate risk ($P < 0.001$).^{4,5} Although the patients who were analyzed in this study were not treated with a single uniform protocol, outcome data were

available for all 188 patients (Table 2 in the Supplementary Appendix). *IDH1* mutational status did not have independent prognostic value with respect to overall survival in multivariate analysis; subgroup analysis showed a possible adverse effect on overall survival among patients with normal-karyotype AML and wild-type *NPM1*, regardless of *FLT3* status (Fig. 4 in the Supplementary Appendix).

DISCUSSION

Our findings support the use of an unbiased sequencing approach to discover previously unsuspected, recurring mutations in a cancer genome. With improved sequencing techniques, we covered this genome more completely than the first one we sequenced (98% vs. 91% diploid coverage) and used fewer sequencing runs (16.5 vs. 98), resulting in a dramatically reduced cost of data generation. With better data quality and calling algorithms, we reduced the 96% false positive frequency of possible mutations for the first sequenced AML genome to a frequency of 47% of the high-confidence tier 1 and 2 mutations called in this genome. We predicted 1458 tumor-specific point mutations with high confidence; we tested 116 of these with validation sequencing and confirmed 61 of them (53%). Thus, this genome may contain approximately 750 somatic point mutations. We detected mutations in *NRAS*, *NPMc*, and *IDH1* and a tier 2 mutation on chromosome 10 in more than one AML genome, suggesting that these mutations are not random and are probably important for the pathogenesis of this tumor.

We suggest that the 12 nonsynonymous mutations are the most likely to be relevant for pathogenesis, since they could potentially alter the function of expressed genes. Consistent with this idea and with the results of our previous study¹⁸ is the finding that all these mutations were retained in the dominant clone. Surprisingly, we found that virtually all the 52 tier 2 mutations were also present in nearly every tumor cell in the sample, suggesting that they are also a part of the same dominant clone. However, one cannot conclude that these mutations (or any of the tier 3 or 4 mutations) are relevant for pathogenesis simply because they are found at a high frequency in the dominant clone. It is more likely that most of these mutations are random, benign sequence changes that existed in the hematopoietic cell that was transformed (i.e., they were preexisting and carried along as benign “passengers,” irrelevant for pathogenesis). The finding that the percentage of mutations found in each tier closely approximated the total amount of DNA assayed in that tier supports this hypothesis. Collectively, these data suggest that the vast majority of the mutations that we detected in this genome are random, background mutations in the hematopoietic stem cell that was transformed.²⁷ Functional validation will be required to prove which mutations are truly important.

The best test of the relevance of individual mutations for pathogenesis (in the absence of functional validation) is recurrence in other AML samples or other cancers. Of the 12 tier 1 mutations, 3 (occurring in *NPM1*, *NRAS*, and *IDH1*) were recurrent in patients with AML and therefore were likely to be important in the pathogenesis of this tumor. R132 mutations in the *IDH1* gene had not previously been detected in the 45 patients with AML who were tested²³ and are detected only rarely in tumor types other than malignant gliomas.^{22,24} The *IDH1* R132H, C, and S mutations dramatically reduce the catalytic activity of the *IDH1* enzyme; it has been suggested that *IDH1* is a tumor suppressor that is inactivated by dominant mutations in R132.²⁸ There are significant differences, however, between the *IDH1* mutations found in gliomas and those in AML. We detected the R132C mutation in 8 of 16 patients with AML who carried an *IDH1* mutation (50%). In contrast, the mutation was reported in only 7 of 161 patients with gliomas (4%, $P < 0.001$ by Fisher’s exact test). The most common mutation in gliomas (R132H) was detected in 142 of 161 patients (88%) but in only 7 of 16 patients with AML (44%, $P = 0.13$). When the R132H mutation was

overexpressed in a glioblastoma cell line, induction of messenger RNAs for several target genes of hypoxia-inducible factor 1 α (HIF1 α) was detected (*GLUT1*, *VEGF*, and *PGKI*).²⁸ However, in 13 patients with AML — 5 with R132H and 8 with R132C — there were no significant alterations in the expression of any of these genes (Fig. 3 in the Supplementary Appendix).

Assuming that the number of point mutations in most AML genomes is similar to the number in the first 2 patients we studied (approximately 750), the likelihood that 2 of 188 patients will carry an identical mutation at the same position in the genome is extremely small (1.1×10^{-9}). This suggests that the tier 2 somatic mutation at position 108,115,590 of chromosome 10 is unlikely to be a random event. It falls in a conserved region with regulatory potential, and its detection in a second patient with AML suggests that this region may contribute to pathogenesis through a novel mechanism that remains to be defined.

Although the potential of next-generation sequencing platforms for uncovering the genetic rules of cancer is great, the sequencing of thousands of additional cancer genomes will be required to fully unravel this complex and heterogeneous disease.^{29,30}

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by grants from the National Institutes of Health (PO1-CA101937, to Dr. Ley; and U54-HG003079, to Dr. Wilson) and the Barnes–Jewish Hospital Foundation (00335-0505-01, to Dr. Ley).

We thank Jennifer Ivanovich for obtaining the detailed family histories of the patients; Nancy Reidelberger for administrative support; Dr. Rob Culverhouse for statistical support; Todd Hepler, William Schroeder, Justin Lolofie, Scott Abbott, Shawn Leonard, Ken Swanson, Indraniel Das, and Michael Kiwala for their contributions to the Laboratory Information Management System; Gary Stiehr, Richard Wohlstadter, Matt Weil, and Kelly Fallon for information-technology support; Drs. Clara Bloomfield, Michael Caligiuri, and James Vardiman for providing the AML samples from the Cancer and Leukemia Group B Leukemia Bank; the nursing staff of the Siteman Cancer Center and Barnes–Jewish Hospital; and all the patients who participated in the study.

References

1. Song WJ, Sullivan MG, Legare RD, et al. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet.* 1999; 23:166–75. [PubMed: 10508512]
2. Owen C, Barnett M, Fitzgibbon J. Familial myelodysplasia and acute myeloid leukaemia — a review. *Br J Haematol.* 2008; 140:123–32. [PubMed: 18173751]
3. Schlenk RF, Döhner K, Krauter J, et al. Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. *N Engl J Med.* 2008; 358:1909–18. [PubMed: 18450602]
4. Byrd JC, Mrózek K, Dodge RK, et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood.* 2002; 100:4325–36. [PubMed: 12393746]
5. Grimwade D, Walker H, Harrison G, et al. The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial. *Blood.* 2001; 98:1312–20. [PubMed: 11520776]
6. Rucker FG, Bullinger L, Schwaenen C, et al. Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. *J Clin Oncol.* 2006; 24:3887–94. [PubMed: 16864856]

7. Suela J, Alvarez S, Cifuentes F, et al. DNA profiling analysis of 100 consecutive de novo acute myeloid leukemia cases reveals patterns of genomic instability that affect all cytogenetic risk groups. *Leukemia*. 2007; 21:1224–31. [PubMed: 17377590]
8. Tyybäkinöja A, Elonen E, Piippo K, Porkka K, Knuutila S. Oligonucleotide array-CGH reveals cryptic gene copy number alterations in karyotypically normal acute myeloid leukemia. *Leukemia*. 2007; 21:571–4. [PubMed: 17268525]
9. Walter MJ, Payton JE, Ries RE, et al. Acquired copy number alterations in adult acute myeloid leukemia genomes. *Proc Natl Acad Sci U S A*. in press.
10. Bullinger L, Döhner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med*. 2004; 350:1605–16. [PubMed: 15084693]
11. Valk PJ, Verhaak RG, Beijnen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004; 350:1617–28. [PubMed: 15084694]
12. Baldus CD, Mrózek K, Marcucci G, Bloomfield CD. Clinical outcome of de novo acute myeloid leukaemia patients with normal cytogenetics is affected by molecular genetic alterations: a concise review. *Br J Haematol*. 2007; 137:387–400. [PubMed: 17488484]
13. Heuser M, Beutel G, Krauter J, et al. High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. *Blood*. 2006; 108:3898–905. [PubMed: 16912223]
14. Langer C, Radmacher MD, Ruppert AS, 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; 111:5371–9. [PubMed: 18378853]
15. Lughart S, van Drunen E, van Norden Y, et al. High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated. *Blood*. 2008; 111:4329–37. [PubMed: 18272813]
16. Marcucci G, Maharry K, Whitman SP, et al. High expression levels of the ETS-related gene, ERG, predict adverse outcome and improve molecular risk-based classification of cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol*. 2007; 25:3337–43. [PubMed: 17577018]
17. Pratz K, Levis M. Incorporating FLT3 inhibitors into acute myeloid leukemia treatment regimens. *Leuk Lymphoma*. 2008; 49:852–63. [PubMed: 18452067]
18. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*. 2008; 456:66–72. [PubMed: 18987736]
19. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456:53–9. [PubMed: 18987734]
20. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 2008; 18:1851–8. [PubMed: 18714091]
21. Falini B, Mecucci C, Tiacci E, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med*. 2005; 352:254–66. Erratum, *N Engl J Med* 2005;352:740. [PubMed: 15659725]
22. Bleeker FE, Lamba S, Leenstra S, et al. IDH1 mutations at residue p.R132 (IDH1(R132)) occur frequently in high-grade gliomas but not in other solid tumors. *Hum Mutat*. 2009; 30:7–11. [PubMed: 19117336]
23. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 mutations in gliomas. *N Engl J Med*. 2009; 360:765–73. [PubMed: 19228619]
24. Kang MR, Kim MS, Oh JE, et al. Mutational analysis of IDH1 codon 132 in glioblastomas and other common cancers. *Int J Cancer*. 2009; 125:353–5. [PubMed: 19378339]
25. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res*. 2002; 12:436–46. [PubMed: 11875032]
26. Link DC, Kunter G, Kasai Y, et al. Distinct patterns of mutations occurring in de novo AML versus AML arising in the setting of severe congenital neutropenia. *Blood*. 2007; 110:1648–55. [PubMed: 17494858]

27. Sjöblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006; 314:268–74. [PubMed: 16959974]
28. Zhao S, Lin Y, Xu W, et al. Gliomaderived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1 α . *Science*. 2009; 324:261–5. [PubMed: 19359588]
29. Dulbecco R. A turning point in cancer research: sequencing the human genome. *Science*. 1986; 231:1055–6. [PubMed: 3945817]
30. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458:719–24. [PubMed: 19360079]

Glossary

Build 36 of the human reference genome	The most current version of the assembled human genome reference sequence, available online at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/).
Comparative genomic hybridization (CGH)	A comparison of DNA abundance, throughout the genome, between two DNA samples to identify regions where DNA copies have been gained or lost.
dbSNP	A publicly available database of known DNA variants, housed at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/SNP).
Diploid coverage	A metric used in whole-genome sequencing studies to describe the likelihood of detecting both alleles at any given nucleotide position in a genome.
Genic	Regions of the genome that contain genes, including their exons and introns.
Haploid coverage	A metric used in whole-genome sequencing studies to describe detection of each nucleotide position in a genome for at least one allele “1 \times coverage” is equivalent to the size of the genome (e.g., approximately 3 billion base pairs for the human genome).
Next-generation sequencing	A variety of new techniques that have in common the generation of DNA sequence from single molecules of DNA, rather than pools of DNA templates hundreds of millions of DNA fragments can be sequenced at the same time on a single platform (massively parallel sequencing).
Paired-end reads	DNA sequences that are produced on next-generation sequencing platforms by sequencing both ends of DNA fragments, resulting in higher confidence in assigning the sequence a position in the reference genome and allowing the detection of structural variations.
Partial uniparental disomy	An acquired somatic recombination event that causes the duplication of a part of a chromosome from one parent, resulting in a “copy-number neutral” loss of heterozygosity for a chromosomal segment.
Resequencing	Obtaining the DNA sequence of additional members of a species for which a completed reference sequence is known and to which comparisons can be made.

Sequencing run	The sequence that is generated by a complete Illumina flow cell or a similar next-generation sequencing platform one sequencing run generates many billions of base pairs of sequence.
Single-nucleotide polymorphism (SNP)	A position in the genome where some individuals in a population inherit a change in a single nucleotide that differs from the reference genome.
Single-nucleotide variant (SNV)	A difference in a DNA sequence of a sample at a single position in the genome, as compared with the reference genome each variant may represent either an inherited or an acquired change.
SNP array	A microarray-based assay system that allows for simultaneous measurement of nucleotide sequence and abundance in a DNA sample at possibly hundreds of thousands of positions in the genome.

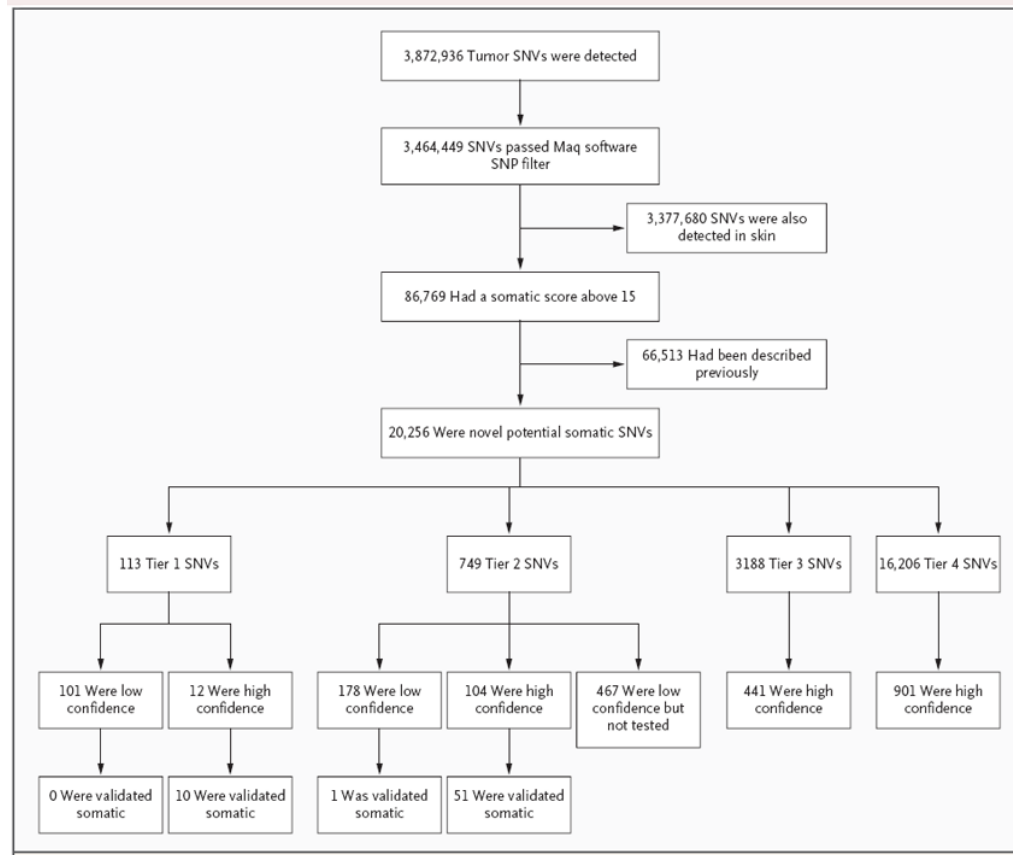


Figure 1. Flow Chart for Identification of Somatic Point Mutations in the Acute Myeloid Leukemia Genome

Maq denotes Mapping and Assembly with Quality, SNP single-nucleotide polymorphism, and SNV single-nucleotide variant.

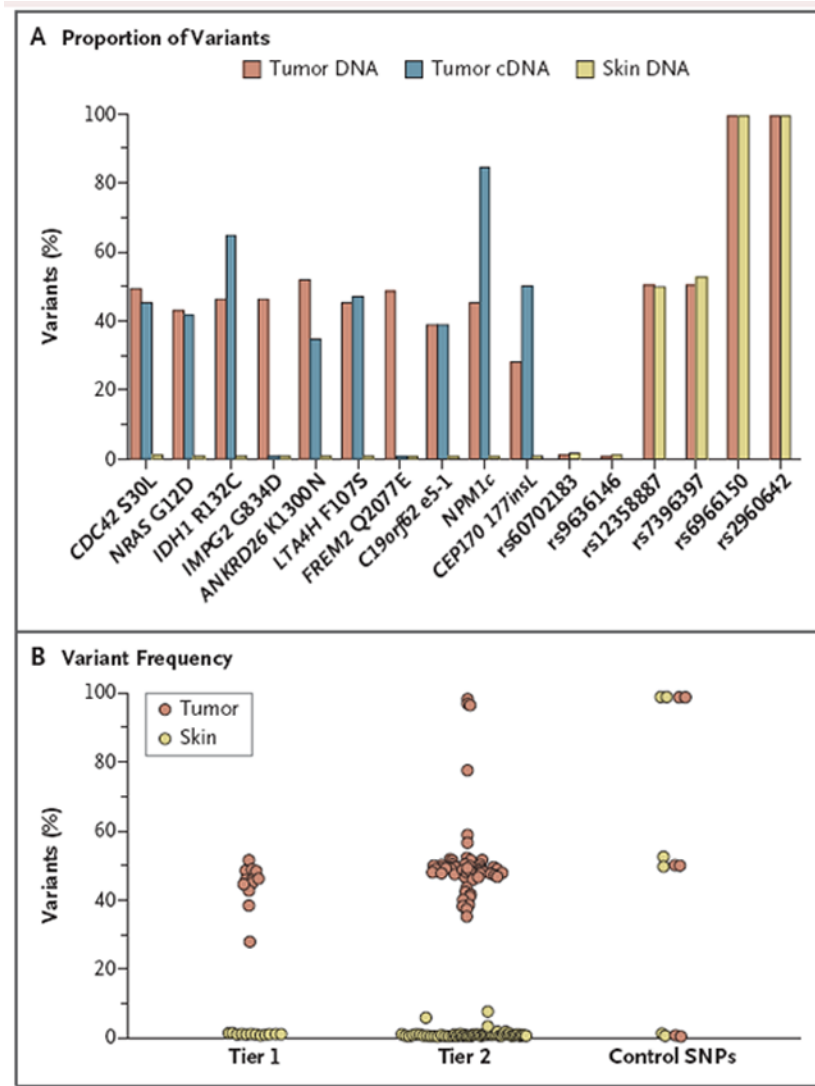


Figure 2. Allele Frequency in Tumor DNA, Tumor Complementary DNA, and Skin DNA
 Panel A shows the percentage of variant alleles that were detected in tumor DNA, tumor complementary DNA (cDNA), and skin DNA for the 10 validated nonsynonymous tier 1 somatic mutations in the index patient. For comparison, variant allele frequencies are shown for six known single-nucleotide polymorphisms (SNPs). The patient was homozygous for the reference sequence for the first two variants, heterozygous for the next two variants, and homozygous for a rare SNP for the last two variants. Panel B shows variant allele frequencies for all validated tier 1 and tier 2 mutations and the six control SNPs for tumor DNA and skin DNA.

Table 1

Sequence Coverage for Tumor and Skin Genomes.*

Variable	Tumor	Skin
Sequencing runs — no. [†]	16.5	13.125
Haploid coverage	23.3×	21.3×
SNVs — no.	3,464,465	3,448,797
Concordance with dbSNP build 129 — no. (%)	3,053,215 (88.1)	2,992,069 (86.8)
High-quality heterozygous SNPs		
By array — no.	45,111	44,778
By sequence — no. (% of array)	44,442 (98.5)	43,629 (97.4)
High-quality homozygous rare SNPs		
By array — no.	28,295	27,735
By sequence — no. (% of array)	28,252 (99.8)	27,685 (99.8)

* The term dbSNP denotes a National Center for Biotechnology Information database of known DNA variants, SNP single-nucleotide polymorphism, and SNV single-nucleotide variant.

[†] A single sequencing run uses all eight lanes of an Illumina flow cell (see the Glossary).

Table 2

Tier 1 Mutations.*

Annotated Gene	Mutation Type	Annotation	SIFT Prediction	Conservation Score	Base Conservation	Skin Tumor %	Variant Frequency	Best Probe [†]
						cDNA		
<i>CDC42</i>	Missense	S30L	Tolerated	597	1	1.03	49.27	46.3 27,990
<i>NRAS</i>	Missense	G12D	Deleterious	616	1	0.66	43.00	42.0 7,468
<i>IDH1</i>	Missense	R132C	Deleterious	445	1	0.81	46.06	63.9 11,400
<i>IMPG2</i>	Missense	G834D	Deleterious	472	0.018	0.67	46.22	0.4 NA
<i>ANKRD26</i>	Missense	K1300N	Deleterious	444	1	0.70	51.73	33.1 514
<i>LTA4H</i>	Missense	F107S	Tolerated	539	0.946	0.68	45.28	47.9 12,138
<i>FREM2</i>	Missense	Q2077E	Tolerated	464	1	0.37	48.92	0 [‡] NA
<i>C19orf62</i>	Splice-site	Exon 5-1	NA	444	1	0.27	38.71	38.8 5,021
<i>SRRM1</i>	Silent	P69I	NA	553	0.988	0.97	46.61	ND 12,858
<i>PCDHA6</i>	Silent	A73I	NA	NS	0.423	0.66	49.75	ND Absent
<i>CEP170</i>	In-frame insertion	Codon 177 in-frame ins L	NA	513	1	0.28	28.57	52.0 15,298
<i>NPM1</i>	Frame-shift insertion	W288fs	NA	689	1	0	45.46	85.4 27,150

*The term cDNA denotes complementary DNA, ins L insertion of Leu, NA not available, ND not done, NS no score, and SIFT Sorting Intolerant from Tolerant.

[†]The best probe refers to the signal value for the most highly expressed probe on the Affymetrix GeneChip Human Genome U133 Plus 2.0 array, transformed by statistical algorithms (MAS 5.0).[‡]The variant frequency was calculated from cDNA subclones.

Table 3Characteristics of the Patients, According to *IDH1* Genotype.*

Variable	Without <i>IDH1</i> Mutation (N = 172)	With <i>IDH1</i> Mutation (N = 16)	P Value
Age at study entry — yr	46.3±15.8	48.9±15.4	0.52 [†]
Race — no. (%) [‡]			0.88 [§]
White	140 (81)	13 (81)	
Black	14 (8)	1 (6)	
Other	18 (10)	2 (12)	
Male sex — no. (%)	101 (59)	9 (56)	1.00 [§]
Bone marrow blasts at diagnosis — %	69.3±18.1	76.7±16.4	0.12 [†]
Cytogenetic profile — no. (%)			0.001 [§]
Normal	67 (39)	13 (81)	
Other	105 (61)	3 (19)	
Cytogenetic risk group — no./total no. (%) [¶]			0.001 [§]
Favorable	58/169 (34)	0/16	
Intermediate or normal	97/169 (57)	16/16 (100)	
Poor	14/169 (8)	0/16	
AML-M3 subtype — no. (%)	40 (23)	0/16	0.03 [§]
Underwent transplantation — no. (%)	27 (16)	3 (19)	0.72 [§]
Mutation — no. (%)			
<i>NPM1</i>	36 (21)	7 (44)	0.06 [§]
<i>FLT3</i>			
Internal tandem duplication	36 (21)	4 (25)	0.75 [§]
D835	10 (6)	1 (6)	1.00 [§]
<i>RAS</i>	19 (11)	1 (6)	1.00 [§]

* Plus-minus values are means ±SD. Percentages may not total 100 because of rounding. AML-M3 denotes acute promyelocytic leukemia, *FLT3* FMS-related tyrosine kinase 3, and *IDH1* isocitrate dehydrogenase 1.

[†]The P value was calculated with the use of the two-sided t-test.

[‡]Race was self-reported.

[§]The P value was calculated with the use of Fisher's exact test.

[¶]The definitions for cytogenetic risk groups are from the Medical Research Council and the Cancer and Leukemia Group B.^{4,5} Adequate cytogenetic data were not available for 3 of the 172 patients who did not have an *IDH1* mutation.