(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 June 2012 (14.06.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/078288 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2011/059843

(22) International Filing Date:
8 November 2011 (08.11.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/456,560 8 November 2010 (08.11.2010) US

(71) Applicant *(for all designated States except US)*: **WASHINGTON UNIVERSITY** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LEY, Timothy J.** [US/US]; 1879 Koenig Rd., Owensville, MO 85066 (US). **WILSON. Richard K.** [US/US]; 2234 Whitney Pointe Drive, Chesterfield, MO 63005 (US). **MARDIS, Elaine K.** [US/US]; 6 Mardis Ct., Troy, IL 62294 (US). **DING, Li** [US/US]; 11036 Ridgecorde Ln., Creve Couer, MO 63141 (US). **DIPERSiO. John K.** [US/US]; 6372 Forsyth, St. Louis, MO 63105 (US).

(74) Agent: **ZACKSON, Saul L.**; Zackson Law LLC, 1015 Locust Street, Ste. 750, St Louis, MO 63101 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

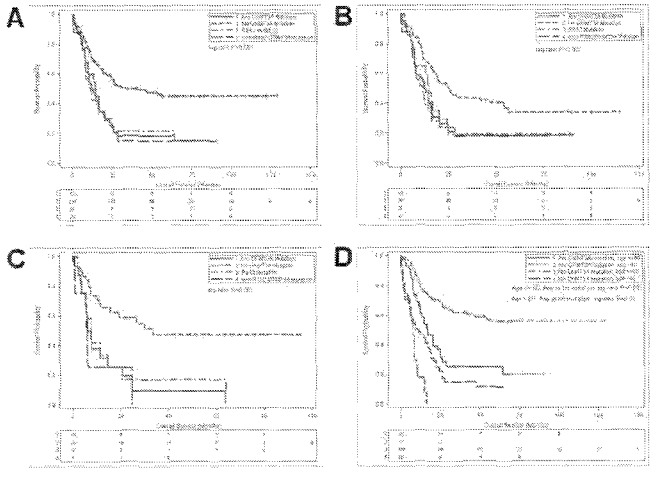
Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: METHODS OF DETERMINING RISK OF ADVERSE OUTCOMES IN ACUTE MYELOID LEUKEMIA

FIG. 10



(57) **Abstract**: Disclosed are methods of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia or myelodysplastic syndrome syndrome. The methods comprise providing a sample comprising bone marrow cells and/or peripheral blood from a subject diagnosed with acute myeloid leukemia or myelodysplastic syndrome syndrome, sequencing at least one exon of *DNMT3A* comprised by the sample, and identifying mutations, if any, in the *DNMT3A* gene. The presence of one or more mutations in *DNMT3A* can indicate that the subject is at high risk for an adverse outcome.

WO 2012/078288 A2

Methods of Determining Risk of Adverse Outcomes in Acute Myeloid Leukemia

Priority Statement

This application claims the benefit of and priority to U.S. Provisional Application No. 61/456,560, filed on November 8, 2010, which is incorporated herein by reference in its entirety.

Government Interest

This work was supported at least in part by grants from the NIH, including UL1 RR024992 from the National Center for Research Resources (NCRR), P41RR000954 , PO1 CA101937, and U54 HG003079. The government has certain rights in the invention.

Incorporation by Reference of Sequence Listing

The Sequence Listing, which is a part of the present disclosure, includes a computer readable form and a written sequence listing comprising nucleotide and/or amino acid sequences. The sequence listing information recorded in computer readable form is identical to the written sequence listing. The subject matter of the Sequence Listing is incorporated herein by reference in its entirety.

Introduction

The genetic alterations responsible for adverse outcomes in most persons with acute myeloid leukemia (AML) are unknown.

Whole genome sequencing is an unbiased approach for discovering somatic variations in cancer genomes. The inventors recently reported the DNA sequence and analysis of the genomes of two patients with normal karyotype acute myeloid leukemia (AML) (Ley, T.J., et al., Nature 456: 66-72, 2008; Mardis, E.R., et al., N. Engl. J. Med. 361: 1058-1066, 2009). The inventors did not find new recurring mutations in the first study, but did observe a recurrent mutation in *IDH1*, encoding isocitrate dehydrogenase 1, in the second (Mardis, E.R., et al., N Engl J Med. 361: 1058-66, 2009). Subsequent work has confirmed and extended this finding, showing that mutations in *IDH1* and the related gene *IDH2* are highly recurrent in cytogenetically normal AML (~20-30% frequency), and associated with poor prognosis in some subsets of patients (Marcucci G., et al., J Clin Oncol. 28: 2348-55, 2010; Paschka P., et al., J Clin Oncol., 28: 2348-55, 2010; Ward P.S., et al., Cancer Cell 17: 225-34, 2010).

The human genes *DNMT1*, *DNMT3A*, and *DNMT3B* encode DNA methyltransferases, enzymes that catalyze the addition of a methyl group to the cytosine residue of CpG dinucleotides. The sequence of *DNMT3A* gene is provided herein as SEQ ID NO: 1; The sequence of Dnmt3a polypeptide is provided herein as SEQ ID NO: 2. Clusters of CpG dinucleotides ("islands") are concentrated in regions upstream of genes; increased methylation of these CpG islands is often associated with reduced expression of the downstream gene. Aberrant DNA methylation has long been hypothesized to contribute to cancer pathogenesis (Esteller, M., N. Engl. J. Med. 358: 1148-59, 2008; Ting, A.H., et al., Genes Dev. 20: 3215-31, 2006; Sharma, S., et al., Carcinogenesis 31: 27-36, 2009; Ehrlich, M., Epigenomics 1: 239-59, 2009). Although cancer genomes tend to be globally hypomethylated relative to normal tissues, hypermethylation of CpG islands in the promoters of tumor suppressor genes is common in many tumors.

Genome-wide methylation profiling of AML samples using microarrays has demonstrated subgroups of samples that share similar methylation patterns (Figueroa, M.E., et al., Cancer Cell 17: 13-27, 2010). Samples obtained from patients with myelodysplastic syndromes (MDS) have a pattern of aberrant hypermethylation, compared with that of de novo AML (Figueroa, M.E., et al., Blood 114: 3448-3458, 2009). DNMT inhibitors are widely used to treat patients with AML and MDS, although the response rate is low, unpredictable in individual patients, and not clearly associated with methylation status (Cashen, A.F., et al., J. Clin. Oncol. 28: 556-561, 2009; Silverman, L.R., et al., J. Clin. Oncol. 20: 2429-2440, 2002; Fandy, T.E., et al., Blood 114: 2764-2773, 2009). In addition, a recent report demonstrated that higher expression of miR-29b (a microRNA that targets *DNMT3A*) in AML blasts was associated with improved clinical response to the DNA methyltransferase inhibitor decitabine (Blum, W., et al., Proc. Natl. Acad. Sci. USA. 107: 7473-7478, 2010).

Summary

Because persons with acute myeloid leukemia (AML) or myelodysplastic syndrome (MDS) can be at risk for adverse outcomes such as relapse or premature death, the present inventors have developed methods of determining risk of adverse outcome in a subject such as a human patient diagnosed with acute myeloid leukemia or myelodysplastic syndrome. These methods involve sequence analysis of bone marrow cells and/or peripheral blood of a subject such as a human previously diagnosed with AML or MDS. In various embodiments of these methods, a sample of bone marrow cells and/or peripheral blood can be obtained

from the subject. Sequencing of genomic DNA or mRNA encoding DNMT3A from cells of a bone marrow or peripheral blood sample from the subject can comprise sequencing at least one exon of *DMNT3A*. By sequencing at least one *DMNT3A* allele, mutations, if present, can be identified in at least one *DNMT3A* allele. The presence of one or more mutations in at least one *DNMT3A* sequence such as a *DNMT3A* exon sequence can indicate that the subject can be at high risk for an adverse outcome and therefore can be a candidate for intensive therapy, such as, for example but without limitation, a bone marrow transplant.

In various configurations, a mutant *DNMT3A* such as a *DNMT3A* allele that is predictive of high risk for an adverse outcome in AML or MDS can comprise one or more mutations of *DNMT3A* other than or in addition to a nucleic acid sequence encoding a substitution of arginine at position 882 of DNMT3A polypeptide. In some configurations, a *DNMT3A* mutant gene can be a *DNMT3A* allele that can comprise one or more mutations other than or in addition to nucleic acid sequence encoding arginine-to-histidine R882H or arginine-to-cysteine R882C. In some configurations, a mutation can be a missense mutation. In some configurations, a mutation can be a frameshift mutation. In some configurations, a mutation can be a frameshift in a nucleic acid sequence encoding a lysine codon at amino-acid position 723. In some configurations, a mutation can be a nonsense mutation. In some configurations, a mutation can be disruptive of splicing. In some configurations a mutation can include a *DNMT3A* sequence that encodes a splicing variant.

In various configurations, a mutant *DNMT3A* that is predictive of high risk for an adverse outcome in AML or MDS can comprise a missense mutation at the codon encoding R882. In some configurations, the mutation at the codon encoding R882 can encode an amino acid such as R882H, R882C, R882P or R882S.

In some aspects, a mutation can encode a missense mutation at a position other than or in addition to R882. In some configurations, a *DNMT3A* mutant allele can be a deletion.

In some configurations, sequencing at least one exon of at least one *DNMT3A* can comprise sequencing at least two exons of *DNMT3A* in a sample. In some configurations, sequencing at least one exon of a *DNMT3A* can comprise sequencing all exons of a *DNMT3A*. In some configurations, the sequencing of at least one exon of *DNMT3A* can comprise sequencing cDNA. In these configurations, cDNA can be generated from cellular DNA by standard methods such as reverse transcription. In some configurations, the sequencing of a *DNMT3A* gene comprised by a sample can comprise sequencing the *DNMT3A* gene in its entirety. In some configurations, the sequencing of at least one exon of at least one allele of *DNMT3A* can comprise sequencing the at least one *DNMT3A* comprised

by a sample in its entirety. In some configurations, the sequencing of at least one exon of at least one allele of *DNMT3A* can comprise sequencing at least a portion of a cDNA of *DNMT3A* RNA comprised by a sample. In some configurations, the sequencing of at least one exon of at least one allele of *DNMT3A* can comprise sequencing in its entirety a cDNA of *DNMT3A* RNA comprised by a sample such as a bone marrow sample or a peripheral blood sample.

In various aspects of the present teachings, a sample comprising bone marrow cells can be obtained by standard methods such as by performing a biopsy. In various aspects of the present teachings, a sample comprising peripheral blood can be obtained by standard methods such as standard blood draws.

In some aspects, intensive therapy can comprise a bone marrow transplant. In various aspects, a bone marrow transplant can be an allogeneic bone marrow transplant or a syngeneic bone marrow transplant.

Brief description of the drawings

FIG. 1 illustrates locations of AML mutations within the three dimensional structure of DNMT3A protein.

FIG. 2 illustrates overall survival of AML patients based on the number of mutations detected in the commonly mutated genes in AML (FLT3, DNMT3A, IDH1/2, and NPM1).

FIG. 3 illustrates overall survival of 187 AML patients collected at Washington University stratified by DNMT3A mutation status and allogeneic transplantation status.

FIG. 4 illustrates *DNMT3A* mutations in 188 AML cases analyzed.

FIG. 5 illustrates a map of a 1.5 Mb deletion encompassing DNMT3A from a tumor sample obtained from a subject.

FIG. 6 illustrates DNMT3A single nucleotide variants (SNVs) found in Cancer and Leukemia Group B (CALGB) AML samples.

FIG. 7 illustrates expression data for *DNMT* genes in AML cases.

FIG. 8. illustrates cDNA readcounts for DNMT3A mutations.

FIG. 9 illustrates total numbers of putative somatic single nucleotide variants (SNVs) in sequenced AML genomes with and without DNMT3A mutations.

FIG. 10 illustrates overall survival of adult AML patients with *DNMT3A* mutations.

FIG. 11 illustrates the overall survival of AML patients with DNMT3A and intermediate risk cytogenetics.

FIG. 12 illustrates overall survival of AML patients with DNMT3A mutations and poor risk (high risk) genotypes.

FIG. 13 illustrates overall survival of AML patients with DNMT3A mutations and favorable risk genotypes.

FIG. 14 illustrates overall survival of AML patients with DNMT3A and FLT3 ITD mutations.

Detailed Description

Using massively parallel DNA sequencing, the present inventors identified somatic mutations in *DNMT3A* (which encodes a de novo DNA methyltransferase) in the genome of an AML with a normal karyotype. The inventors sequenced the exons of *DNMT3A* in 280 additional de novo cases of AML to define recurring mutations.

The diversity of mutations in the DNMT3A gene is reminiscent of the large number of mutations that inactivate classical tumor suppressor genes like TP53or BRCA1 (Walsh, T., et al., Hum. Mutat. 21: 313-320, 2003; Varley, J.M., Hum. Mutat. 21: 313-220, 2003). The nonsense and frameshift mutations of DNMT3A are all predicted to result in truncated proteins that eliminate (8 of 11) or shorten (3 of 11) the methylase domain (Ehrlich, M., et al., Autoimmunity 41:253-271, 2008); several are associated with nonsense-mediated decay, clearly demonstrating loss of function. The distribution of frameshift, nonsense, and missense mutions in DNMT3A are remarkably similar to the pattern of mutations seen in the DNMT3B gene in the ICF syndrome (Ehrlich, M., et al., Autoimmunity 41:253-271, 2008; Jiang, Y.L., et al., Hum. Mutat 25: 56-63, 2005). Missense mutations are generally found within the methylase domain (37 of 38), near the homodimer interface (R882C, R882H, R882P), along the length of the DNMT3L interacting helix (R729Q, R729W, R736H, A741V), or near the DNA binding groove (P718L, R792H, R803S, K829R, R882C, R882H, R882P, F909C) (Ehrlich, M., et al., Autoimmunity 41:253-271, 2008; Jia, D., et al., Nature 449: 248-251, 2007). Mutations commonly occur at charged positions (10 of 12) and may alter or abolish homo/heterodimeric complexes and DNA binding properties. The SIFT/Polyphen algorithms suggest that many of the missense mutations are deleterious for protein function, but some are not. Of note, DNMT3A haploinsufficiency is not deleterious to mice (Okano, M., et al., Cell 99: 247-257, 1999), and complete loss of DNMT3A in the bone marrow cells of mice does not overtly alter hematopoiesis in the short term (Tadokoro, Y., et al., J. Exp. Med. 204:715-722, 2007).

In 62 out of the 281 (22%) de novo AMLs had mutations in *DNMT3A* that were predicted to affect translation. The inventors identified 18 different missense mutations, the most common of which was predicted to affect amino acid R882 (37 cases). The inventors also identified six frameshift, six nonsense, and three splice-site mutations, and a 1.5-Mb

deletion encompassing *DNMT3A*. These mutations were highly enriched in AMLs with intermediate risk cytogenetics (56/166–33.7%; $p<0.001$), and absent in AMLs with favorable cytogenetics (0/79; $p<0.001$). Genomic 5-methylcytosine content, the general pattern of CpG island methylation, and gene expression patterns were not consistently altered in genomes with *DNMT3A* mutations. The median overall survival of all AML patients with *DNMT3A* mutations was significantly less than those lacking such mutations (12.3 vs. 41.1 months, $p<0.001$); *DNMT3A* mutations correctly classified adverse outcomes for patients with intermediate risk cytogenetics or *FLT3* mutations, regardless of age, and were independently associated with poor outcomes in a Cox Proportional Hazards analysis.

DNMT3A mutations are highly recurrent in *de novo* AML cases with intermediate risk cytogenetics, and are independently associated with poor survival.

DNMT3A mutations are recurrent in patients with AML and, regardless of the type of mutation or location in the gene, are associated with poor event-free and overall survival, independent of age, *FLT3*, or *NPM1* mutation status. These findings indicate that *DNMT3A* mutations are relevant for AML pathogenesis. *DNMT3A* mutations do not cause genomic instability (since most genomes with mutations have normal cytogenetics, and unaltered numbers of total mutations), nor do they alter total 5-methylcytosine content or global patterns of methylation, or dramatically alter gene expression. Currently, the only clue regarding a pathogenetic mechanism is strong selection against *DNMT3A* mutations (and also *IDH1/2* and *NPM1* mutations) in cases with good-risk cytogenetics, indicating a biologic relationship that is not random.

There are two major "classes" of *DNMT3A* mutations in AML. The first is the highly recurrent set of mutations at R882 (also described recently by Yamashita, Y., et al., Oncogene 29: 3723-31, 2010, in 3/74 AML samples tested; the low prevalence in that study could reflect a different sample population, or a lower sensitivity of mutation detection), and the second is represented by all the other mutations in this gene. The locations of mutations in the *DNMT3A* gene are similar to that of *DNMT3B* mutations associated with the ICF syndrome (Immunodeficiency, centromere instability and facial anomalies syndrome) (Ehrlich, M., et al., Autoimmunity 41: 253-71, 2008; Jiang, Y.L., et al., Hum Mutat. 25: 56-63, 2005).

meDIP-chip experiments revealed a significant reduction in DNA methylation at 182 genomic locations. Both the R882H and R882C mutations are caused by a C-to-T transition at a CpG dinucleotide (R882H on the non-coding strand, and R882C on the coding strand), suggesting that these mutations may be caused by the deamination of methylcytosine on

either strand of this CpG dinucleotide (Rideout, W.M., et al., Science 249: 1288-90, 1990). Without be limited by theory many *DNMT3A* mutations are predicted to cause changes in the DNA binding groove of DNMT3A, and some are predicted to change its interaction with DNMT3L (FIG. 1).

In FIG. 1A, a side view is shown of tetrameric DNMT3A (murine AA 623 – 908 which corresponds to conserved human AA 627 - 912) and DNMT3L, interacting with S-adenosyl-L-methionine (AdoMet - white spacefill in FIG. 1B, C) (Jia, D., et al., Nature 449: 248-251, 2007; Tadokoro, Y., J. Exp. Med. 204: 715-722, 200). Jia, D., et al., Nature 449: 248-251, 2007 suggested that DNA binds this tetrameric complex along a structural groove (indicated by arrow in FIG. 1A. Amino acids mutated along the proposed DNA binding groove are indicated (R792, R803, K829, R882, F909). Amino acids mutated along the DNMT3L interacting helix are highlighted. R882 mutations occur near the homodimerization plane, but this residue has not been implicated in homodimerization (Jia et al calculated R885 to be a critical residue.) The R882 side-chain protrudes into the DNA binding groove, and mutations at this position decrease both DNA binding and catalytic activity (Gowher, H., et al., J. Mol. Biol. 357: 928-941, 2006). However, without being limited by theory, it is possible that R882 mutations may alter functions of DNMT3A that are not yet fully understood, including its ability to bind to other proteins involved in transcriptional regulation and localization to chromatin regions containing methylated DNA (Hervouet E, et al., Epigenetics 4: 487-99, 2009; Wang, Y.A., et al., Cancer Biol Ther. 4: 1138-43, 2005; Brenner, C., et al., EMBO J. 24: 336-46, 2005; Jones, P.A., et al., Nat Rev Genet. 10: 805-11, 2009). Regardless, all *DNMT3A* mutations are associated with poor overall survival, suggesting that they have an important common effect on the potential of AML cells to cause lethal disease.

The association of *DNMT3A* mutations with mutations in the four other most commonly mutated genes in AML (*FLT3, NPM1, IDH1* and *IDH2*) were studied. A large proportion of intermediate-risk patients have mutations in one or more of these genes; however, many intermediate and poor risk patients have no mutations in any of these genes, which is not a random event ($p<0.001$). Those with no mutations in any of these genes have outcomes similar to that of those with 1 or 2 mutations (FIG. 2) but may have a unique set of 'driver' mutations.

Remarkably, no *DNMT3A* mutations were found in 79 patients with a "good risk" cytogenetic profile, which includes patients with t(15;17), t(8;21), and inv(16). Mutations in *NPM1, IDH1* and *IDH2* were not detected in these cases either, consistent with data reported by others (Paschka P., et al., J Clin Oncol., 28: 2348-55, 2010). The virtual exclusion of

mutations in these four genes in good-risk AML cases is not random, and may reflect the leukemogenic properties of the fusion proteins created by the translocations. PML-RARA and AML-ETO fusion proteins (generated respectively by t(15;17) and t(8;21)) physically interact with *DNMT3A* (PML-RARA) or *DNMT1* (AML-ETO), and alter the methylation of specific promoters, respectively (Di Croce, L., et al., Science 295: 1079-1082, 2002; Liu, S., et al., Cancer Res 65: 1277-1284, 2005; Fazi, F., et al, Blood 109:4432-4440, 2007). Both *PML* and *DNMT3A* regulate telomere function, and ATRA (all trans retinoic acid, part of the therapy for t(15;17) patients) down-regulates *DNMT3A* expression (Gonzalo, S., et al., Nat Cell Biol. 8: 416-24, 2006; Fazi, F., et al., Oncogene 24: 1820-30, 2005). Together, these data indicate that *DNMT3A* mutations and good risk-associated translocations can be mutually exclusive because they both change DNMT3A functions, despite the fact that the mutations and translocations result in dramatically different clinical outcomes.

DNMT3A mutations do not change 5-methylcytosine content in AML genomes, and the R882H mutation minimally perturbs the methylation of CpG islands. The inventors did not detect changes in DNA methylation that are directly correlated with local changes in gene expression. These data might suggest that DNMT3A mutations do not directly affect the cytosine methyltransferase properties of DNMT3A. However, a recent study has suggested that DNMT3A may alter the methylation of non-promoter associated CpG regions, affecting gene expression indirectly. The discovery of highly recurrent mutations in *DNMT3A* provides a new tool for the classification of intermediate-risk AML. Furthermore, intensification of treatment, such as allogeneic bone marrow transplantation, can be considered for AML patients having somatic mutations in *DNMT3A*. In our small series, allogeneic transplantation provided statistically significant benefit to patients with DNMT3A mutations (FIG. 3).

Methods

Methods and compositions described herein utilize laboratory techniques well known to skilled artisans. Such techniques can be found in laboratory manuals such as Sambrook, J., et al., Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001; Spector, D. L. et al., Cells: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1998; Harlow, E., Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999.

The present teachings also include the following materials and methods:

Whole Genome Sequencing and mutation validation.

Whole genome sequencing (with paired end reads) and analysis was performed exactly as described (Mardis, E.R., et al., N. Engl. J. Med. 361:1058-1066, 2009). Validation of putative somatic SNVs and indels with Sanger sequencing was performed as described (Mardis, E.R., et al., N. Engl. J. Med. 361:1058-1066, 2009). Measuring expression levels of identified somatic mutations using RT-PCR followed by 454 readcounts. Deep readcounts of variant alleles in cDNA obtained from primary AML bone marrow samples was performed exactly as described(Mardis, E.R., et al., N. Engl. J. Med. 361:1058-1066, 2009).

RNA expression arrays and analysis.

Specimen RNA was extracted, assayed for quality, and subjected to Affymetrix Human Genome U133 Plus 2.0 Array GeneChip microarrays (Affymetrix) as described (Payton, J.E., et al., J. Clin. Invest. 119: 1714-1726, 2009). Profiling data for all specimens have been deposited on the Gene Expression Omnibus (GEO;http://www.ncbi.nlm.nih.gov/geo/; accession no. GSE12662). Cel files were processed using the MAS5 algorithm (Affymetrix) and only probesets called "present" in at least 75% of specimens in either DNMT3A mutated or wild type groups were retained. These filtered probesets were zscored and Ward's unsupervised hierarchical clustering was performed (Spotfire DecisionSite 8.2 software, TIBCO).

The top 2,000 differentially methylated probes among the 10 samples based on mean absolute deviation were selected, and both Pearson-Ward clustering and k-means clustering were performed (k=2). The samples that did not cluster in the appropriate groups were the M5 AML subtype UPN# 737451 with the R882H mutation, and the M4 AML subtype UPN# 418499 without the mutation. The following method was developed to identify regions with significant difference in methylation between AML samples with or without the R882H mutation in DNMT3A.

The following criteria were applied to select probes among the 2.1M for the statistical test:

1. The mean log2-ratio (signal intensity of immunoprecipitated sample/signal intensity of non-immunoprecipitated sample) for 5 DNMT3A mutant samples vs. the mean log2-ratio of 5 DNMT3A wild type samples were different by a value of 1 or more
2. At least 4 out of 5 DNMT3A mutant log2-ratios >= 1, or at least 4 out of 5 DNMT3A wild type log2-ratios >= 1. This allows for the two outliers detected during unsupervised clustering.

For all probes selected, the inventors calculated the mean of log2-ratios for both groups using 5 consecutive probes, which include the current probe and two on either side.

The inventors performed a paired t-test between the 5 DNMT3A mutant means and 5 DNMT3A wild type means, calculated above. If the resulting p-value is <= 0.05, the inventors considered this probe as differentially methylated.

The inventors clustered the differentially methylated probes identified above if they were 220bp or closer to each other. This resulted in 1,214 clusters. Among them, 423 clusters contained 2 or more differentially methylated probes and 182 clusters contained 3 or more differentially methylated probes.

Statistical analysis of survival by risk group.

Mean differences between patients with and without the DNMT3A mutation were compared with two-sided independent-sample t-tests. Differences in distribution of the highly skewed WBC measurements were compared with a Wilcoxon ranksum tests. Differences in the distribution of nominal variables between those with or without the DNMT3A mutation were compared with Fisher's Exact tests or Pearson chi-squared tests. Figures were generated with SAS PROC LIFETEST, which also calculated the log-rank and Wilcoxon test statistics comparing survival between the two DNMT3A-mutation groups. Multivariate Cox proportional hazards analyses (Hosmer, D.W. and Lemeshow, S., Regression Modeling of Time to Event Data. In: Applied Survival Analysis: John Wiley & Sons, Inc.; 1999) were performed with SAS PROC PHREG. Variables evaluated as potential confounders in the proportional hazards model were FLT3mutation, IDH1 mutation, NPMc mutation, male sex, white race, age, FAB category of M3, normal cytogenetics, and cytogenetic risk group. The analyses were generated using SAS/STAT software (Version 9.2 of the SAS System for Windows, SAS Institute Inc., Cary, NC).

Clinical AML treatment protocols. Among the 188 patients treated at Washington University, 72 were treated with a standard seven day induction regimen of infusional cytarabine plus three days of an anthracycline ("7+3"), and 61 received a similar regimen that also included 3 days of etoposide ("7+3+3"). Twenty-four patients with acute promyelocytic leukemia (AML M3), were treated with 7+3 plus concurrent ATRA. Thirteen patients were treated with either azacytidine or decitabine. Nine patients were treated with lenalidomide. Two patients received an induction regimen of cladribine, mitoxantrone, and highdose cytarabine ("CLAM"). Six patients received no chemotherapy (except hydroxyurea), and one patient was treated with infusional cytarabine alone. Seventy-nine patients subsequently

underwent stem cell transplantation. Of these, 24 underwent autologous stem transplantation, of which 10 also underwent allogeneic transplant after relapsing. Fifty-five patients underwent allogeneic transplant (without prior autologous transplant). One of the allogeneic transplant only patients failed DNMT3A genotyping.

The inventors have previously described the clinical features of the proband (Ley, T.J., et al., Nature 456: 66-72, 2008). The inventors have also previously described methods of sequencing, of determining whether mutations are recurrent, expression analysis, and outcomes analysis (Mardis, E.R., et al., N. Engl. J. Med. 361: 1058-1066, 2009).

Examples

The following non-limiting examples are provided to further illustrate the present teachings. Those of skill in the art can, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments that are disclosed and still obtain a like or similar result without departing from the spirit and scope of the present teachings.

Example 1

This example illustrates the identification of *DNMT3A* mutations.

The present inventors report that *DNMT3A* is recurrently mutated in AML samples, especially in those leukemias with a cytogenetic profile of intermediate risk. *DNMT3A* mutations were independently associated with poor survival.

In this example, the inventors obtained 116.4 billion base pairs with paired-end reads from the genome of the relapsed tumor, yielding 99.62% diploid coverage of the genome. The sequence coverage with paired-end reads and -mutation calling algorithms the inventors used allowed the inventors to identify several non-synonymous Tier 1 mutations that the inventors did not detect in our initial sequencing effort, including a one base-pair deletion in *DNMT3A*. This deletion causes a frameshift in the codon encoding lysine at amino-acid position 723, which predicts the synthesis of 56 "novel" amino acids (that is, amino acids that are not encoded by non-mutated *DNMT3A* at positions including and beyond amino-acid position 723), followed by an in-frame, premature stop codon.

Example 2

This example demonstrates *DNMT3A* Mutations in AML.

The exons of *DNMT3A* and *DNMT3L* were sequenced (DNMT3L forms heterodimers with both DNMT3A and its homologue, DNMT3B) by PCR amplification, and Sanger sequencing of 188 *de novo* AMLs banked at Washington University. Primer sequences are disclosed herein as SEQ ID NO: 3-58. All 188 tumors had matched normal DNA samples available for analysis. The inventors obtained 94 additional AML samples from the Cancer

and Leukemia Group B (CALGB); matched normal tissue was not available for these samples. Although these patients were not treated uniformly, the prevalance of common AML mutations in these patients is typical, and the clinical outcomes are similar to those of other recently described series of patients.

Among the 94 CALGB patients, fifty-eight patients were treated according to CALGB 9621, in which they were randomized to receive PSC833 vs placebo with concurrent infusional cytarabine, daunorubicin, and etoposide, followed by riskadapted consolidation. Thirteen patients were treated with according to CALGB 9222, in which they received infusional cytarabine plus daunorubicin induction, followed by randomization to consolidation with high dose cytarabine, with or without cyclophosphamide and etoposide, or diaziquone and mitoxantrone. Thirteen patients with AML M3 were enrolled in CALGB 9191 and were randomized to induction therapy with ATRA vs cytarabine and daunorubicin. Fourteen AML M3 patients were enrolled in CALGB 9710 and were randomized to receive cytarabine, daunorubicin, and ATRA induction, with or without arsenic trioxide during consolidation. No patients in the CALGB group underwent stem cell transplantation.

The inventors obtained sequence coverage of all 24 exons of *DNMT3A* for 281 of the 282 tumor samples. Sixty-two tumors (22%) had mutations predicting translational consequences; 5 of these 62 tumors had two independent mutations (notably, none of these mutations is predicted to affect amino-acid position 882). The inventors determined that all mutations in the tumors obtained from Washington University were somatic events by virtue of their absence in DNA sequence obtained from matched skin samples; a summary of all confirmed somatic mutations is shown in FIG. 4. The locations of the PWWP (Stec, I., et al., FEBS Lett. 473: 1-5, 2000), PHD (zinc finger, Baker, L.A., et al., Mutat. Res. 647: 3-12, 2008), and methyltransferase (MTase) domains are shown in FIG. 4 and each patient with a *DNMT3A* mutation is designated with a circle. The deletion noted at position 1 is represents a 1.5 Mbp deletion that includes *DNMT3A* and part or all of 8 other genes (see FIG. 5).The inventors observed seven novel sequence changes in *DNMT3A* in the CALGB sample set, but could not determine whether they were somatic changes without matched DNA from an unaffected tissue (see FIG. 6 for putative somatic CALGB mutations). The locations of validated SNVs (single nucleotide variants) in the DNMT3A gene are shown for 94 AML samples obtained from CALGB (FIG. 6). Since no matched normal tissue was available for these tumor samples, these mutations are not proven to be somatic. However, none has been described in dbSNP, and none were inherited SNPs in our analysis of 188 additional AML

cases. The locations of the PWWP, PHD (zinc finger), and methyltransferase (MTase) domains are shown. Each patient with a DNMT3A mutation is designated with a circle.

None of the novel variants detected in the CALGB tumors have been previously identified as SNPs. The inventors identified 18 different missense mutations across both sets of tumors. The most common missense mutations are predicted to affect amino acid R882 (37 cases). 27 tumors harbored R882H, 7 harbored R882C, 2 harbored R882P, and 1 harbored R882S. The inventors also observed six frameshift, six nonsense, and three splice-site mutations, and a 1.5-Mb deletion that included *DNMT3A* and eight other genes (FIG. 5). A 1.5 Mb deletion was detected in UPN 113971 with whole genome sequencing, by defining low coverage of this region of chromosome 2 in the tumor sample only. The results were confirmed on Affymetrix 6.0 SNP arrays. The deleted region includes all or part of 9 genes, including DNMT3A. The inventors observed no somatic mutations in the coding exons of *DNMT3L* in 188 samples tested.

The relationship of *DNMT3A* mutations to the other common mutations in AML genomes was studied. Mutations in *FLT3*, *NPM1*, and *IDH1* are significantly enriched in samples with *DNMT3A* mutations. No *DNMT3A* mutations (and likewise no *NPM1* or *IDH1/2* mutations) were found in AML samples with translocations associated with favorable outcome (0/79, $p<0.001$). Likewise, none of the 11 patients with structural variations involving 11q23 (where the MLL gene resides) had *DNMT3A* mutations. *DNMT3A* mutations were significantly enriched in AMLs with a cytogenetic profile associated with intermediate risk (56/166=33.7%, $p<0.001$), in AMLs with normal cytogenetics (44/120=36.7%, $p<0.001$), and in stages M4 (20/61=32.8%, $p=0.04$) and M5 (12/21=57%, $p<0.001$) of the French-American-British (FAB) classification system (Bennett, J.M., et al., Br. J. Haematol. 33: 451-458, 1976). White blood cell counts at presentation were significantly higher in patients with R882 mutations. The single morphologic FAB M3 case with a *DNMT3A* mutation (UPN 287) was cytogenetically normal, did not have the characteristic expression signature of t(15;17) AML, and died 2.2 months after presentation (Payton, J.E., et al., J Clin Invest. 119: 1714-26, 2009).

The inventors obtained expression array data, using the Affymetrix 133 Plus 2 array, from 180 of the 188 samples obtained from Washington University. The inventors defined the expression of the DNMT genes in AML samples and in normal hematopoietic cells with this data (FIG. 7). In FIG. 7, data for the probesets measuring DNMT3A (upper left), DNMT3L (upper right), DNMT3B (lower left) and DNMT1 (lower right) are shown. Each dot represents one sample. Expression levels for R882 mutations, any other DNMT3A

mutation, or no DNMT3A mutation, are shown; expression levels are not significantly altered by DNMT3A mutations. All of the expression calls for the DNMT3L gene were called "absent" on the expression arrays. Expression levels for these genes are also shown for normal adult bone marrow derived CD34 cells, flow-purified promyelocytes (pros), and flow-purified polymorphonuclear leukocytes (PMNs). *DNMT3A* was expressed in all 180 AML samples, and in normal human CD34+ bone marrow cells. Its expression decreases with terminal myeloid differentiation. The inventors observed no difference in *DNMT3A* expression levels between AML samples with or without *DNMT3A* mutations. Likewise, *DNMT3B* and *DNMT1* are both highly expressed in AML samples and normal CD34 cells. The inventors observed no expression of *DNMT3L* in any AML sample, suggesting that either this gene is not expressed in AML, or that the Affymetrix 133 Plus 2 array is not capable of detecting *DNMT3L* mRNA.

To determine whether the mutant *DNMT3A* alleles are expressed, the inventors amplified the target regions from 21 primary bone marrow tumor samples using RT-PCR, and then sequenced the amplicons using the Roche 454-Titanium platform to obtain deep read countsFIG. 8). Variant allele frequencies from cDNAs obtained from the bone marrow RNA of 21 different AML patients are shown in FIG. 6. Regions containing the mutations were amplified by RT-PCR, and then sequenced using the 454-FLX platform to obtain deep readcounts for variant alleles.. Below each bar, the UPN for the patient sample is shown, followed by the mutation annotation. 30 to 60% of the cDNAs carried the mutant allele, suggesting that the variant and wild type alleles are expressed at near equal levels. One nonsense mutation (E477*) and one frameshift (M315fs) were not represented in cDNAs, suggesting the mRNAs carrying these variants are subject to nonsense-mediated decay. There was no evidence for homozygous loss of *DNMT3A* expression for any of the five samples carrying two mutations. The inventors did not detect variant cDNAs carrying the missense allele (A741V) or the nonsense allele (E477*) in sample 246634 ;although the inventors readily detected cDNA from the non-mutated allele in this sample, w e suspect that A741V and E477* were on the same allele, and thus were equally affected by nonsense-mediated decay.

Patterns of Mutations

The inventors have sequenced the genomes of 38 cytogenetically normal *de novo* AML cases (>25x haploid coverage of tumor and matched normal skin). 11 of these genomes have mutations in *DNMT3A*, but no mutations were detected in *DNMT3L*, *DNMT1*, or *DNMT3B* . The inventors identified with high confidence somatic single-nucleotide variants

(SNVs) in each genome, and determined that the total number of these variants is not significantly influenced by *DNMT3A* mutation status (FIG. 9). Further, the types of mutations detected in each genome (C to T, C to G, C to A, etc.) were not associated with *DNMT3A* mutation status.

Example 3

This example illustrates DNA methylation in AML genomes.

To determine whether 5-methylcytosine content was altered in AML genomes with *DNMT3A* mutations, the inventors hydrolyzed genomic DNA derived from the bone marrow of AML patients to nucleoside monophosphates, and assayed 5-methyl-2'-deoxycytidine monophosphate using liquid chromatography-tandem mass spectrometry. The LC-ESI-MS/MS conditions were optimized by infusing a 10 ng/μL solution of each standard compound in 5% methanol/0.1% formic acid(Song, L., et al., Anal Chem. 77: 504-10, 2005). The reported transition pairs 288.2/112.2, 308.0/112.0, 322.0/126.1, 332.1/136.2, 252.1/135.9, 348.8/152.1, 323.0/81.0, and 243.3/127.2 were used for dC, dCMP, mdCMP, dAMP, dA, dGMP, TMP, and T, respectively (Burke, D.G., et al., Anal. Chem. 81: 7294-7301, 2009). The source parameters were optimized by infusing a solution of mdCMP (10 ng/μL in 5% methanol/0.1% formic acid) at 0.22 mL/min in order to minimize in-source fragmentation. Quantification was performed by injecting standard solutions of the standard mixture. Under these conditions, all standard analytes gave linear curves with correlation coefficients greater than 0.99 in the range of 0.1 to 100 fmols injected. For five replicate standard curves, the coefficients of variation were 15-24% for mdCMP, 18-29% for dCMP, and 15-25% for dGMP. The LC-MS-MS analysis for patient samples was performed using 3.5-5 ng/μL of DNA hydrolysis products. The percentage of methylation was determined by either dividing the measured amount of mdCMP (in fmols) by the amount of dGMP in each sample or by dividing the sum of mdCMP and dCMP detected in each sample. The quantity of dGMP was used based on the assumption that dGMP = dCMP + mdCMP (Song, L., et al., Anal. Chem. 77: 504-510, 2005). For a control sample containing approximately 5% mdCMP, the average percentage of mdCMP was determined to be 4.2 +/- 0.1%, with a coefficient of variation of 3.5%, for four replicate measurements. Both calculation methods gave approximately the same values of 5-methyl deoxycytidine monophosphate in the patient samples. The 5-methylcytosine content of each genome with a *DNMT3A* mutation was virtually identical to that of AML genomes without *DNMT3A* mutations (WT= 4.89 +/- 0.57% 5-meC, vs. any *DNMT3A* mutation= 4.77 +/- 0.50% 5-meC, p=0.45).

The inventors determined the methylation patterns of five AML genomes with the R882H mutation, and five "matched" AML genomes (same FAB subtypes and myeloblast percentages) with non-mutated *DNMT3A*, using MeDIP-chip analysis. Nearly all the methylated regions were similar among the ten samples. 182 genomic regions had statistically different methylation levels (at specific genomic locations) associated with *DNMT3A* mutation status. All 182 regions had significantly reduced methylation (on average) in the mutant genomes. None of the differentially methylated regions was consistently correlated with altered expression of "nearest neighbor" genes.

Example 4

This example illustrates clinical outcomes of patients with DNMT3A mutations.

The inventors determined event-free and overall survival for all 281 AML cases for which *DNMT3A* mutation status was known. In FIG. 10, Panel A shows overall survival of 281 AML patients with (n=62) or without (n=219) a *DNMT3A* mutation; Panel B shows overall survival of 120 AML patients with normal karyotypes with (n=44) or without (n=76) a *DNMT3A* mutation; Panel C. shows overall survival of 54 AML patients with a *FLT3* ITD mutation with (n=16) or without (n=38) a *DNMT3A* mutation; Panel D shows overall survival of 281 AML patients stratified by age (>60, n=72; ≤60, n=209) and *DNMT3A* mutation status.

The two metrics were very similar, so only overall survival is shown (FIG. 10, Panel A). Patients with *DNMT3A* mutations had significantly worse survival. *DNMT3A* mutations are also associated with worse survival of patients with normal cytogenetics (FIG. 10, Panel B), and those with intermediate risk cytogenetics (FIG. 11, p=0.006). When patients with *FLT3 ITD (Internal Tandem Duplication)* mutations were classified by *DNMT3A* status, those carrying a *DNMT3A* mutation did significantly worse (FIG. 10, Panel C). Finally, *DNMT3A* mutations were associated with adverse outcomes for patients regardless of age (FIG. 10, Panel D). Patients of 60 years or greater with *DNMT3A* mutations have extremely poor outcomes. See FIG. 12-14 for data on *DNMT3A* mutations and overall survival of AML patients with all combinations of common AML mutations (*NPM1*, *IDH1/2*, and *FLT3*). DNMT3A mutations conferred a significantly worse overall survival for patients with FLT3 mutations (p=0.0006), for patients with WT NPM1 (p=0.01), and for patients with FLT3 mutations who are NPM1 wild-type (p=0.04) (FIG. 12). There was a trend towards worse overall survival for patients with DNMT3A mutations and IDH1 mutations (p=0.06), but not for the few patients with IDH2 mutations, or IDH1 and 2 mutations combined. DNMT3A mutations were also associated with a worse overall survival for patients with favorable risk

genotypes (FLT3 wildtype, NPMc, IDH1 or IDH2 wild-type ($p \leq 0.02$), but not for patients who were both FLT3 wild-type and NPMc mutant (FIG. 13). The presence of a FLT3 ITD mutation did confer a worse overall survival for patients with DNMT3A mutations, although the number of cases with both mutations is small (n=16) (FIG. 14, p=0.02).

The relationship between DNMT3A status and both overall and event-free survival was examined by multivariate Cox Proportional Hazards models. The variables that were independently associated with overall survival were DNMT3Amutations, age>60, and FLT3 mutations. FAB=M3 was used as a stratifying variable because it did not satisfy the proportionality assumption. After adjusting for age, FAB=M3, and FLT3 mutations, the hazard ratio for DNMT3A mutations was 1.90 (95% confidence interval: 1.34-2.71.) The variables that were independently associated with event-free survival included age>60, FLT3mutations, favorable cytogenetics, and DNMT3A mutations. As with overall survival, FAB=M3 was used as a stratifying variable. After adjusting for age>60, FAB=M3, FLT3 mutation, and favorable cytogenetics, the hazard ratio for DNMT3A mutation was 1.46 (95% confidence interval: 1.02 – 2.08) These Cox Proportional Hazards models show that the variables independently associated with overall survival were *DNMT3A* mutations, age, and *FLT3* mutations.

Example 5

This example illustrates variant allele frequency of the DNMT3A mutation in a sample.

Using PCR, the inventors amplified DNA containing the position of the frameshift mutation at L723, and performed "deep readcounts" to determine the variant allele frequency of this mutation in the de novo and relapse specimens. The variant allele frequency of this mutation was 38.86% (1819/4706 reads) in the de novo AML genome, and 29.75% (1600/5405 reads) at relapse. These data suggest that this mutation was present in most cells of the dominant clone at presentation (100% bone marrow blasts) and at relapse (78% blasts) (Ley, T.J.. et al., Nature 456: 66-72, 2008).

Example 6

This example illustrates predicted consequences of DNMT3A mutations.

The most common mutation in DNMT3A was found at amino acid position R882. Strong selection for mutations at this position suggest that the R882 mutations have a gain-of-function activity, similar in principle to that recently demonstrated for IDH1 and 2 (Dang, L., et al., Nature 462:739-844, 2009; Ward, P.S., et al., Cancer Cell 17: 225-234, 2010). In our

study, 37/62 (60%) of patients with DNMT3A mutations had mutations at this site (R to H most common, followed by R to C, and R to P or R to S). Yamashita, et al, recently reported R882H or R882C mutations in 3/74 AML samples tested; the low frequency could be related to technical issues, or the selection of patients used in their study (Yamashita, Y., et al., Oncogene Oncogene 29: 3723-3731, 2010). The pattern of mutations is remarkably similar to the pattern seen in the IDH1 gene, where recurring missense mutations (predominantly R to H or R to C) are found only at amino acid R132. Our data show that heterozygous R882H mutations are associated with reduced methylation at a small subset of genomic positions, suggesting that the mutation may in fact have a dominant negative effect on the methylase activity of this enzyme. Yamashita, et al, found that DNMT3A with an R882H mutation had reduced activity in a de novo methylation assay, and did not confer global changes in CpG methylation when transfected into Ba/F3 cells (Yamashita, Y., et al., Oncogene Oncogene 29: 3723-3731, 2010). In a domain-mapping experiment, Gowher, et al, also showed that a mutation at position R882 reduced the enzymatic activity of the methylase domain, and reduced DNA binding activity (Gowher, et al, J.Mol. Biol. 357:928-941, 2006). Importantly, an inherited, homologous mutation (R823G) in the DNMT3B gene has been identified in patients with the ICF syndrome (immunodeficiency, centromeric instability, and facial abnormalities), which is characterized in part by chromosomal rearrangements in lymphocytes (Xu, G.L., et al., Nature 402: 187-191, 1999; Hansen, R.S., et al. Proc.Natl. Acad. Sci. USA 96: 14412-14417, 1999). Since DNMT3A has been shown to interact with many proteins, including DNMT3L, SUMO-1, transcriptional repressors, histone modifying proteins, and TP53 (Wang, Y.A., et al., Cancer Biol. Ther 4:1138-1143, 2005), it is also possible that the R882 mutations alter one or more of these interactions, or create alternative activities that are currently unknown. In addition to the de novo methylation activity of DNMT3A, there is evidence that DNMT3A may also play a role in maintenance of methylation through its recruitment to nucleosomes and specific chromatin regions that contain methylated DNA (Jones, P.A., et al., Nat. Rev. Genet. 10: 805-811, 2009).

Example 7

This example illustrates inherited SNPs in the DNMT3A gene.

Six inherited single nucleotide polymorphisms (SNPs) were identified in the coding region of DNMT3A in normal DNA from AML cases. All six SNPs were synonymous, four were novel alleles, and two were previously known (rs2276598, rs41284843). The minor allele frequencies (MAFs) of the four novel alleles were all <0.01 and were not considered

further. For rs2276598 and rs41284843, there was no difference in allele or genotype frequencies between AML cases with or without DNMT3A mutations, or between AML cases (Caucasian only) and CEU controls from pilot 1 of the 1,000 Genomes project.

All references cited are incorporated by reference, each in its entirety.

What is claimed is:

1. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia, the method comprising:

providing a sample comprising bone marrow cells from a subject diagnosed with acute myeloid leukemia;

sequencing at least one exon of *DNMT3A* comprised by the sample; and

identifying mutations, if any, in the *DNMT3A*, whereby the presence of one or more mutations in the *DNMT3A* indicates that the subject is at high risk for an adverse outcome.

2. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the *DNMT3A* comprises one or more mutations other than a substitution of arginine at position 882.

3. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the *DNMT3A* comprises one or more mutations in addition to a substitution of arginine at position 882.

4. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the *DNMT3A* comprises one or more mutations other than R882H or R882C.

5. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the *DNMT3A* comprises one or more mutations in addition to R882H or R882C.

6. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the one or more mutations comprises a frameshift mutation.

7. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the one or more mutations comprises a frameshift in a codon encoding lysine at amino-acid position 723.

8. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the one or more mutations comprises a nonsense mutation.

9. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the one or more mutations comprises a missense mutation at the codon encoding R882 .

10. A method of determining risk of adverse outcome in a subject diagnosed with acute

myeloid leukemia in accordance with claim 9, wherein the mutation at the codon encoding R882 encodes an amino acid selected from the group consisting of R882H, R882C, R882P, and R882S.

11. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the one or more mutations encodes a missense mutation at a position other than R882.

12. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the *DNMT3A* comprises a deletion.

13. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the *DNMT3A* encodes a splicing variant.

14. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the sequencing at least one exon of a *DNMT3A* comprises sequencing of at least two exons of the *DNMT3A*.

15. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the sequencing at least one exon of the *DNMT3A* comprises sequencing of all exons of the *DNMT3A*.

16. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the sequencing at least one exon of the *DNMT3A* comprises sequencing the *DNMT3A* in its entirety.

17. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the providing a sample comprising bone marrow cells comprises performing a biopsy.

18. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the intensive therapy comprises a bone marrow transplant.

19. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 18, wherein the bone marrow transplant is an allogeneic bone marrow transplant.

20. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 1, wherein the adverse outcome is relapse or death.

21. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome, the method comprising:

providing a sample comprising bone marrow cells from a subject diagnosed with myelodysplastic syndrome;

sequencing at least one exon of *DNMT3A* comprised by the sample; and

identifying mutations, if any, in the *DNMT3A*, whereby the presence of one or more mutations in the *DNMT3A* indicates that the subject is at high risk for an adverse outcome.

22. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome in accordance with claim 21, wherein the *DNMT3A* comprises one or more mutations other than a substitution of arginine at position 882.

23. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the *DNMT3A* comprises one or more mutations in addition to a substitution of arginine at position 882.

24. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the *DNMT3A* comprises one or more mutations other than R882H or R882C.

25. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the *DNMT3A* comprises one or more mutations in addition to R882H or R882C.

26. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the one or more mutations comprises a frameshift mutation.

27. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the one or more mutations comprises a frameshift in a codon encoding lysine at amino-acid position 723.

28. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the one or more mutations comprises a nonsense mutation.

29. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the one or more mutations comprises a missense mutation at the codon encoding R882 .

30. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 29, wherein the mutation at the codon encoding R882 encodes an amino acid selected from the group consisting of

R882H, R882C, R882P, and R882S.

31. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the one or more mutations encodes a missense mutation at a position other than R882.

32. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the *DNMT3A* comprises a deletion.

33. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the *DNMT3A* encodes a splicing variant.

34. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the sequencing at least one exon of a *DNMT3A* comprises sequencing of at least two exons of the *DNMT3A*.

35. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the sequencing at least one exon of the *DNMT3A* comprises sequencing of all exons of the *DNMT3A*.

36. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the sequencing at least one exon of the *DNMT3A* comprises sequencing the *DNMT3A* in its entirety.

37. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the providing a sample comprising bone marrow cells comprises performing a biopsy.

38. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the intensive therapy comprises a bone marrow transplant.

39. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 38, wherein the bone marrow transplant is an allogeneic bone marrow transplant.

40. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 21, wherein the adverse outcome is relapse or death.

41. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia, the method comprising:

providing a sample comprising peripheral blood cells from a subject diagnosed with

acute myeloid leukemia;

sequencing at least one exon of *DNMT3A* comprised by the sample; and

identifying mutations, if any, in the *DNMT3A*, whereby the presence of one or more mutations in the *DNMT3A* indicates that the subject is at high risk for an adverse outcome.

42. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the *DNMT3A* comprises one or more mutations other than a substitution of arginine at position 882.

43. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the *DNMT3A* comprises one or more mutations in addition to a substitution of arginine at position 882.

44. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the *DNMT3A* comprises one or more mutations other than R882H or R882C.

45. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the *DNMT3A* comprises one or more mutations in addition to R882H or R882C.

46. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the one or more mutations comprises a frameshift mutation.

47. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the one or more mutations comprises a frameshift in a codon encoding lysine at amino-acid position 723.

48. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the one or more mutations comprises a nonsense mutation.

49. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the one or more mutations comprises a missense mutation at the codon encoding R882 .

50. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 49, wherein the mutation at the codon encoding R882 encodes an amino acid selected from the group consisting of R882H, R882C, R882P, and R882S.

51. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the one or more mutations encodes a

missense mutation at a position other than R882.

52. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the *DNMT3A* comprises a deletion.

53. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the *DNMT3A* encodes a splicing variant.

54. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the sequencing at least one exon of a *DNMT3A* comprises sequencing of at least two exons of the *DNMT3A*.

55. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the sequencing at least one exon of the *DNMT3A* comprises sequencing of all exons of the *DNMT3A*.

56. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the sequencing at least one exon of the *DNMT3A* comprises sequencing the *DNMT3A* in its entirety.

57. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the providing a sample comprising peripheral blood cells comprises drawing blood from the subject.

58. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the intensive therapy comprises a bone marrow transplant.

59. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 58, wherein the bone marrow transplant is an allogeneic bone marrow transplant.

60. A method of determining risk of adverse outcome in a subject diagnosed with acute myeloid leukemia in accordance with claim 41, wherein the adverse outcome is relapse or death.

61. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome, the method comprising:

providing a sample comprising peripheral blood cells from a subject diagnosed with myelodysplastic syndrome;

sequencing at least one exon of *DNMT3A* comprised by the sample; and

identifying mutations, if any, in the *DNMT3A*, whereby the presence of one or more mutations in the *DNMT3A* indicates that the subject is at high risk for an adverse outcome.

62. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome in accordance with claim 61, wherein the *DNMT3A* comprises one or more mutations other than a substitution of arginine at position 882.

63. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the *DNMT3A* comprises one or more mutations in addition to a substitution of arginine at position 882.

64. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the *DNMT3A* comprises one or more mutations other than R882H or R882C.

65. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the *DNMT3A* comprises one or more mutations in addition to R882H or R882C.

66. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the one or more mutations comprises a frameshift mutation.

67. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the one or more mutations comprises a frameshift in a codon encoding lysine at amino-acid position 723.

68. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the one or more mutations comprises a nonsense mutation.

69. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the one or more mutations comprises a missense mutation at the codon encoding R882 .

70. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 69, wherein the mutation at the codon encoding R882 encodes an amino acid selected from the group consisting of R882H, R882C, R882P, and R882S.

71. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the one or more mutations encodes a missense mutation at a position other than R882.

72. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the *DNMT3A* comprises a deletion.

73. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the *DNMT3A* encodes a splicing variant.

74. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the sequencing at least one exon of a *DNMT3A* comprises sequencing of at least two exons of the *DNMT3A*.

75. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the sequencing at least one exon of the *DNMT3A* comprises sequencing of all exons of the *DNMT3A*.

76. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the sequencing at least one exon of the *DNMT3A* comprises sequencing the *DNMT3A* in its entirety.

77. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the providing a sample comprising peripheral blood cells comprises drawing blood from the subject.

78. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the intensive therapy comprises a bone marrow transplant.

79. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 78, wherein the bone marrow transplant is an allogeneic bone marrow transplant.

80. A method of determining risk of adverse outcome in a subject diagnosed with myelodysplastic syndrome syndrome in accordance with claim 61, wherein the adverse outcome is relapse or death.

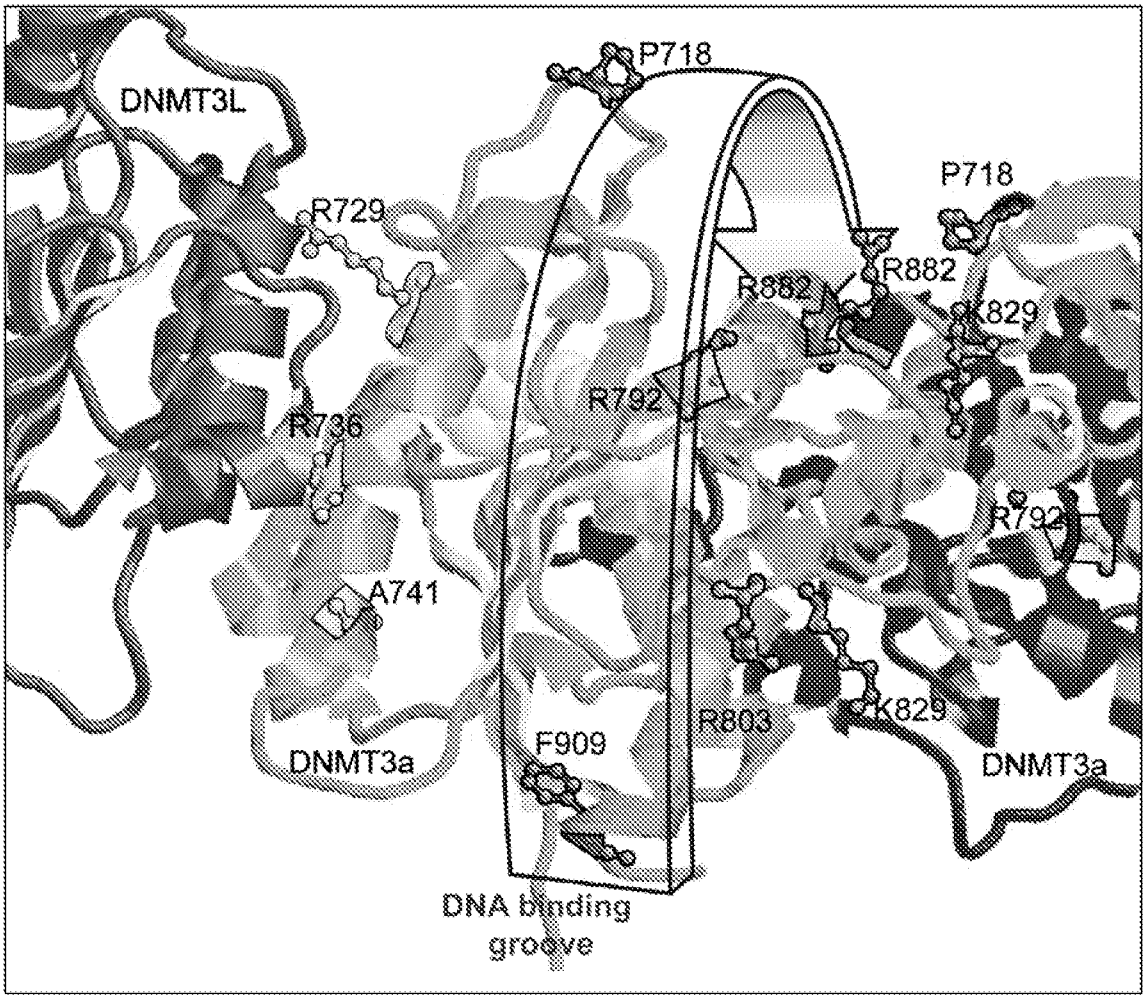

DNMT3L
P718
R729
P718
R882
R882
K829
R792
R736
R792
A741
R803
K829
F909
DNMT3a
DNMT3a
DNA binding
groove

FIG. 1A

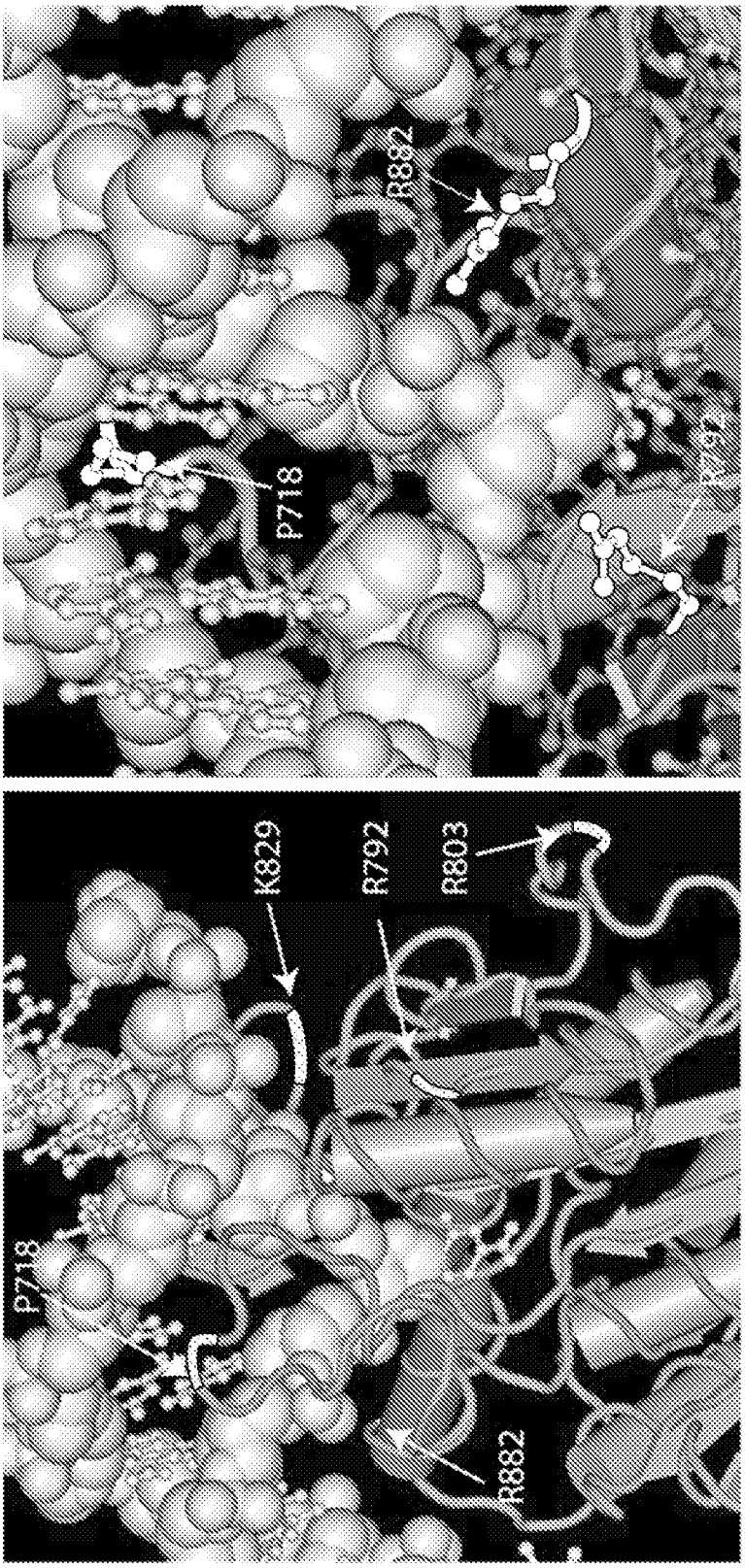


FIG. 1B

FIG. 1C

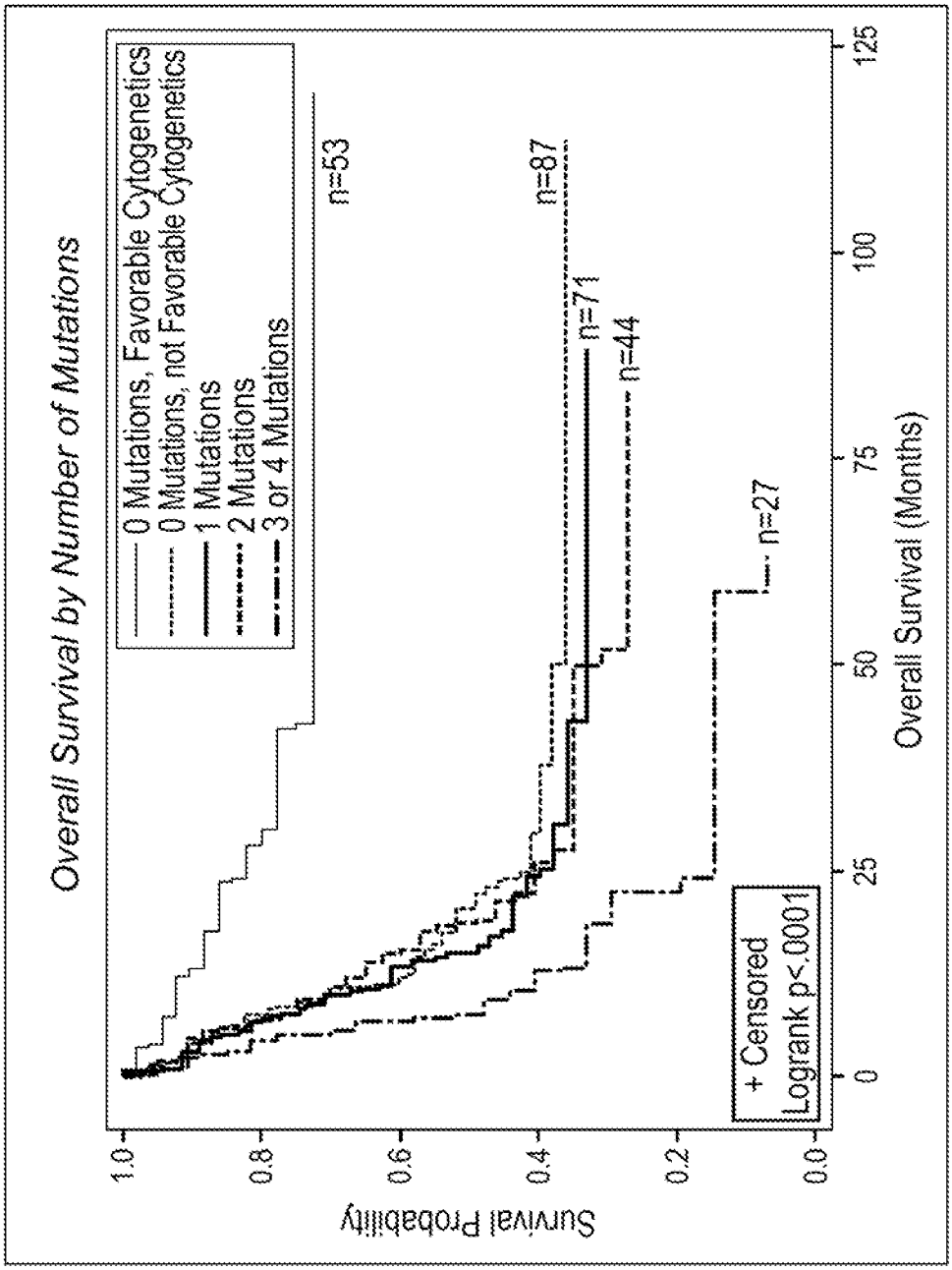

Overall Survival by Number of Mutations
0 Mutations, Favorable Cytogenetics
0 Mutations, not Favorable Cytogenetics
1 Mutations
2 Mutations
3 or 4 Mutations
n=53
n=87
n=71
n=44
n=27
+ Censored
Logrank p<.0001
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
25
50
75
100
125
Overall Survival (Months)

FIG. 2

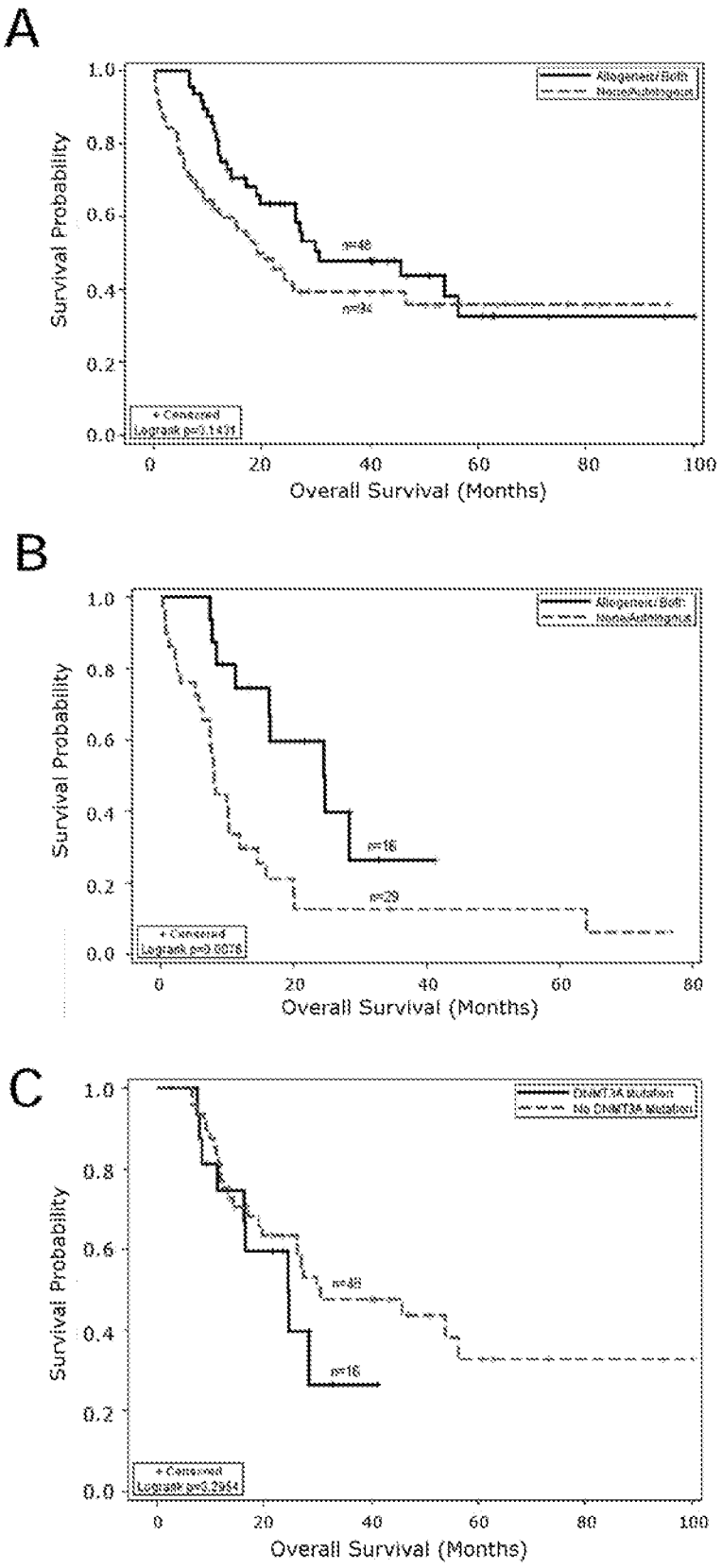

A
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
Overall Survival (Months)
n=48
n=34
B
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
Overall Survival (Months)
n=18
n=29
C
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
Overall Survival (Months)
n=48
n=18

FIG. 3

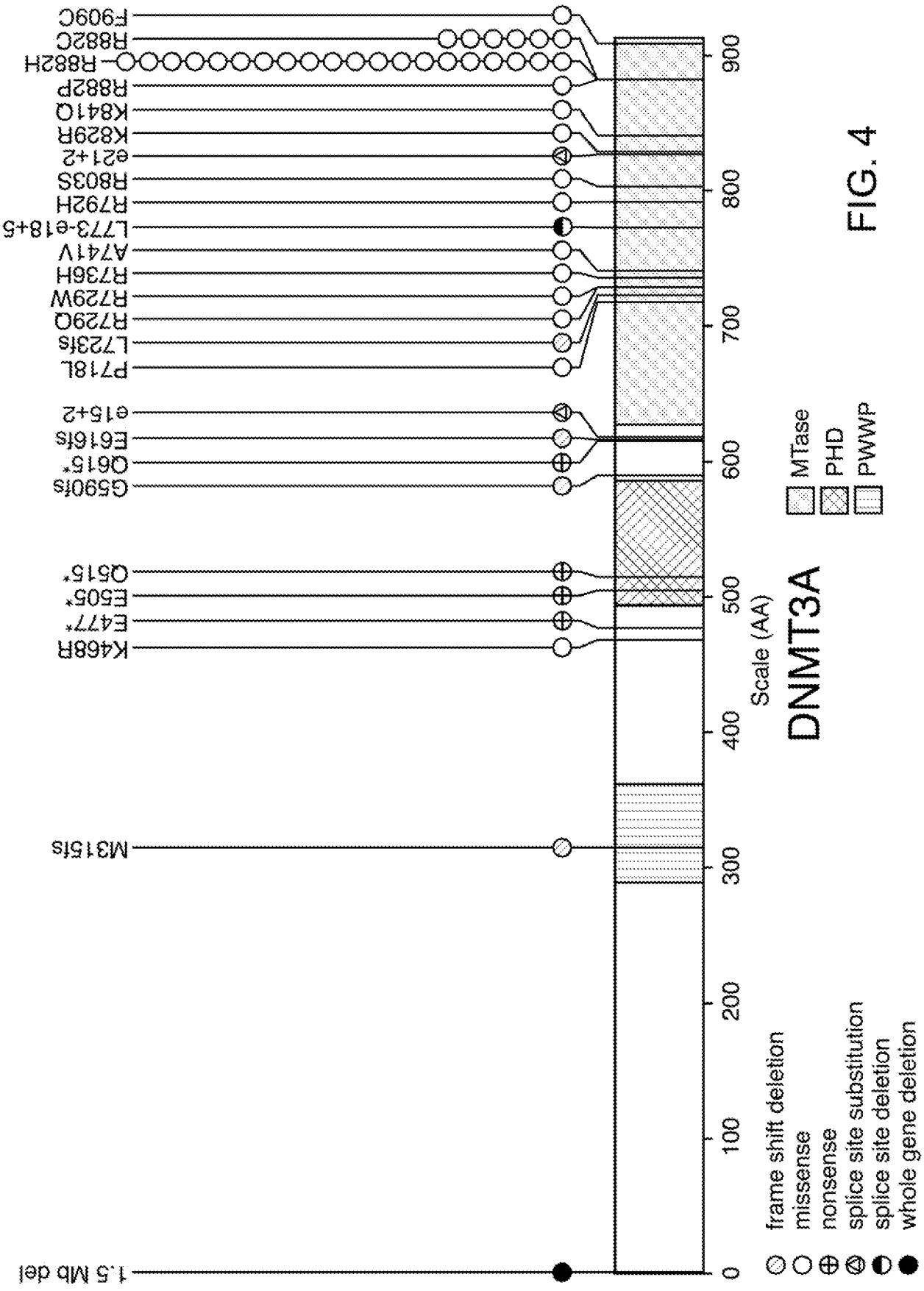

1.5 Mb del
M315fs
K468R
E477*
E505*
Q515*
G590fs
Q615*
E616fs
e15+2
P718L
L723fs
R729Q
R729W
R736H
A741V
L773-e18+5
R792H
R803S
e21+2
K829R
K841Q
R882p
R882H
R882C
F909C
frame shift deletion
missense
nonsense
splice site substitution
splice site deletion
whole gene deletion
0
100
200
300
400
500
600
700
800
900
Scale (AA)
DNMT3A
MTase
PHD
PWWP

FIG. 4

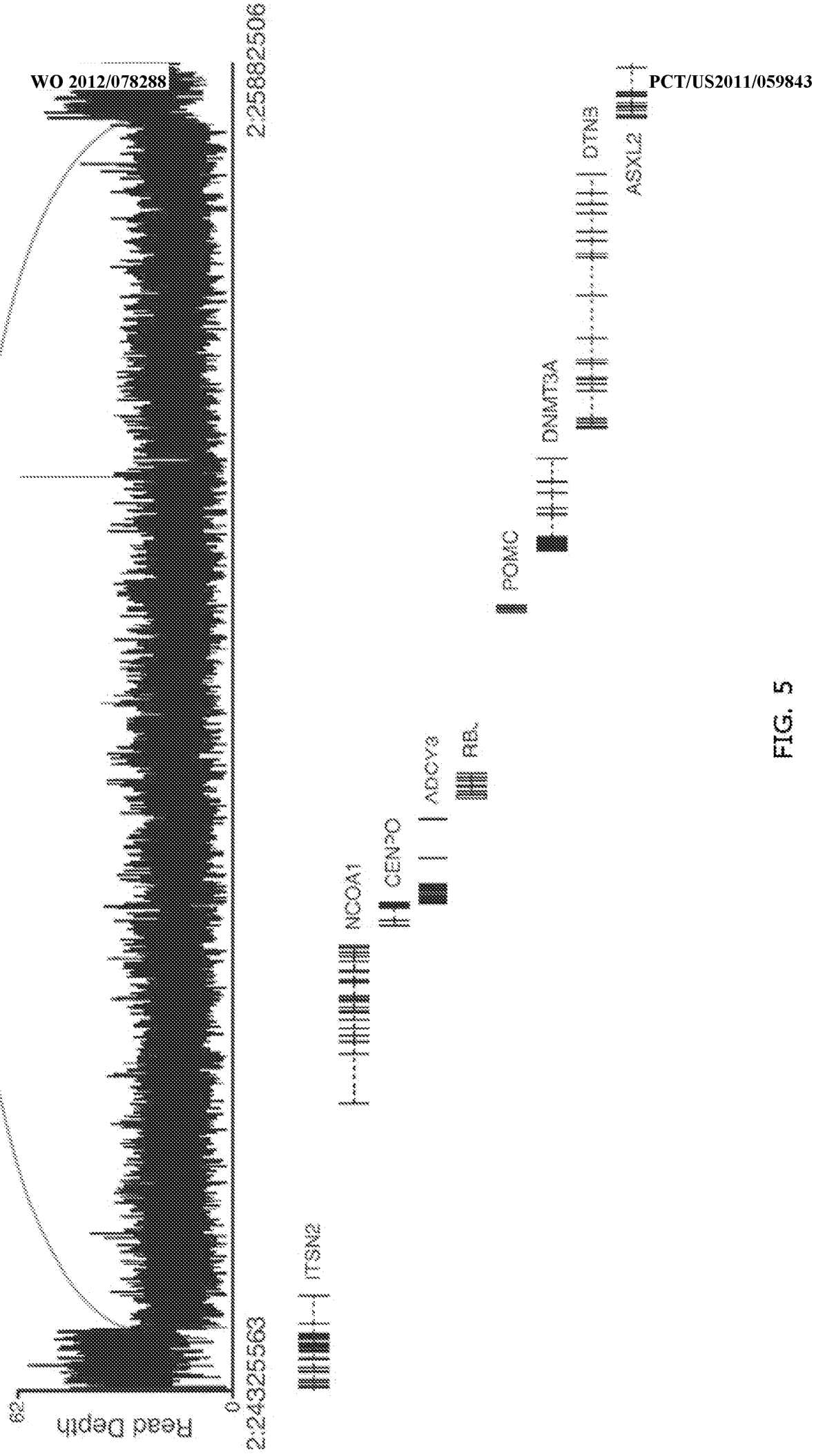

Read Depth
62
0
2:24325563
2:25882506
ITSN2
NCOA1
CEN2O
ADCY8
RBL
POMC
DNMT3A
DTN3
ASXL2

FIG. 5

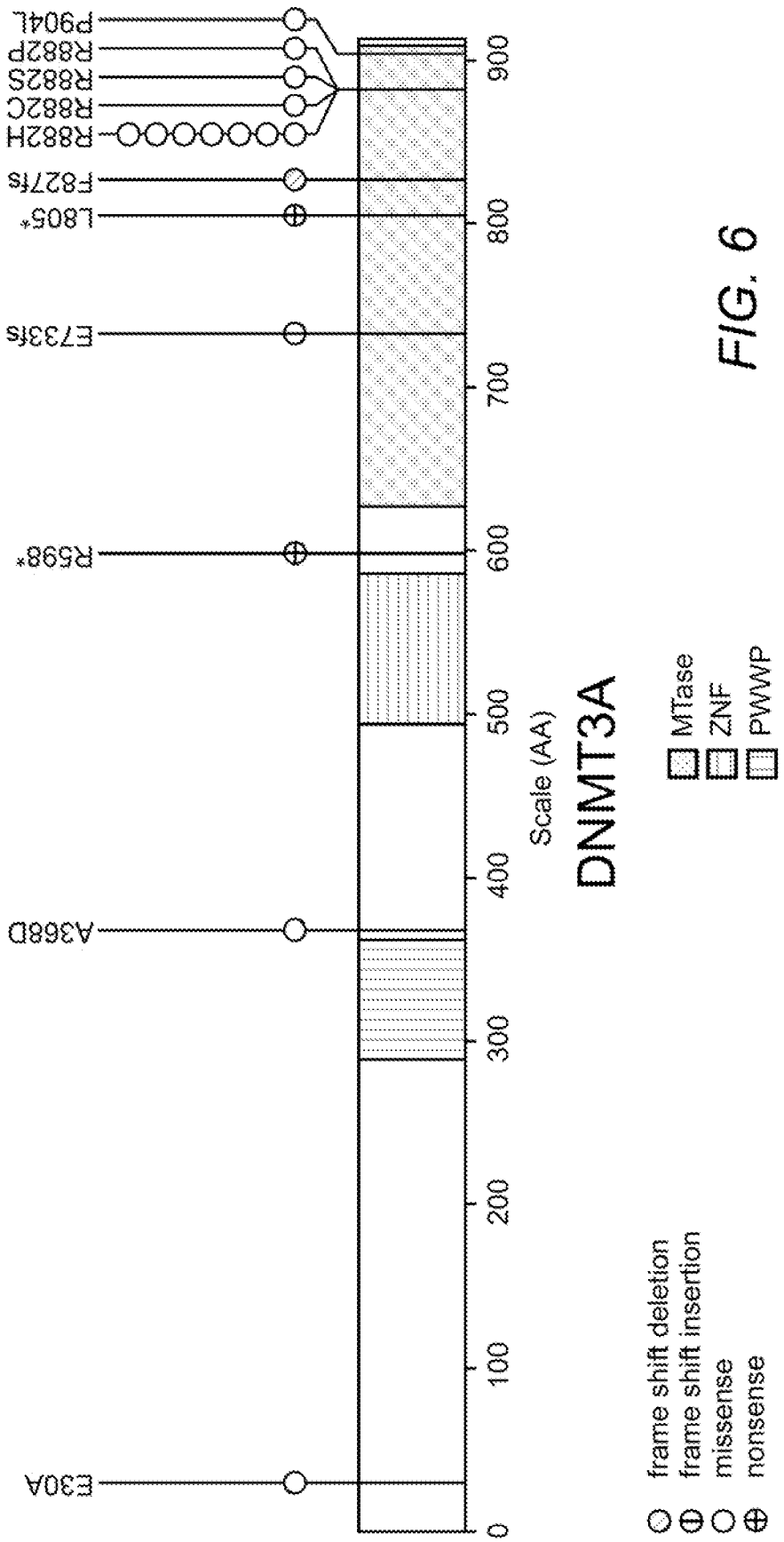

P904L
R882P
R882S
R882C
R882H
F827fs
L805*
E733fs
R598*
A368D
E30A
0
100
200
300
400
500
600
700
800
900
Scale (AA)
DNMT3A
MTase
ZNF
PWWP
frame shift deletion
frame shift insertion
missense
nonsense

FIG. 6

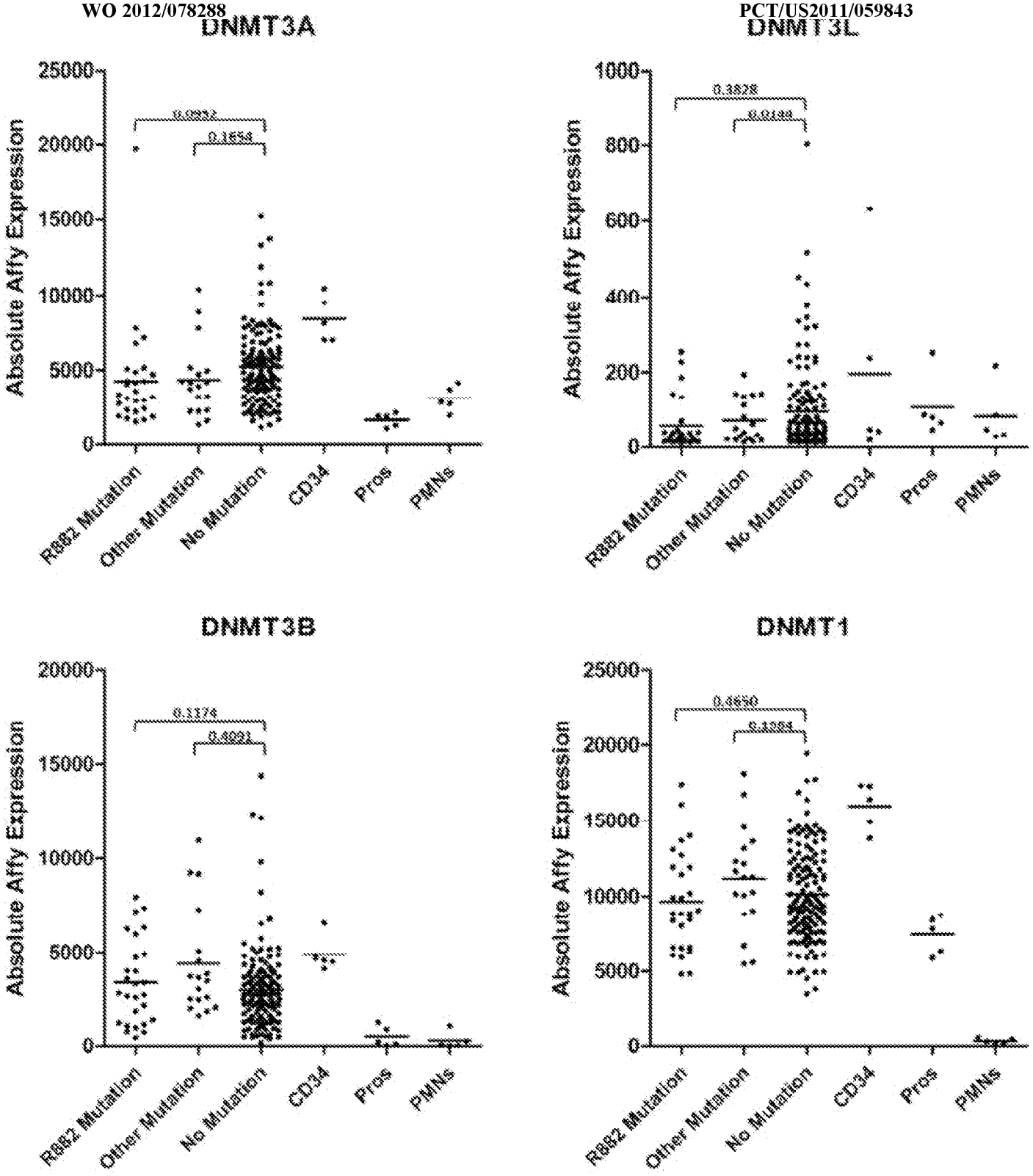

DNMT3A
Absolute Affy Expression
25000
20000
15000
10000
5000
0
0.0832
0.1858
R882 Mutation
Other Mutation
No Mutation
CD34
Pros
PMNs
DNMT3L
Absolute Affy Expression
1000
800
600
400
200
0
0.3828
0.0168
R882 Mutation
Other Mutation
No Mutation
CD34
Pros
PMNs
DNMT3B
Absolute Affy Expression
20000
15000
10000
5000
0
0.1174
0.4021
R882 Mutation
Other Mutation
No Mutation
CD34
Pros
PMNs
DNMT1
Absolute Affy Expression
25000
20000
15000
10000
5000
0
0.4690
0.1284
R882 Mutation
Other Mutation
No Mutation
CD34
Pros
PMNs

FIG. 7

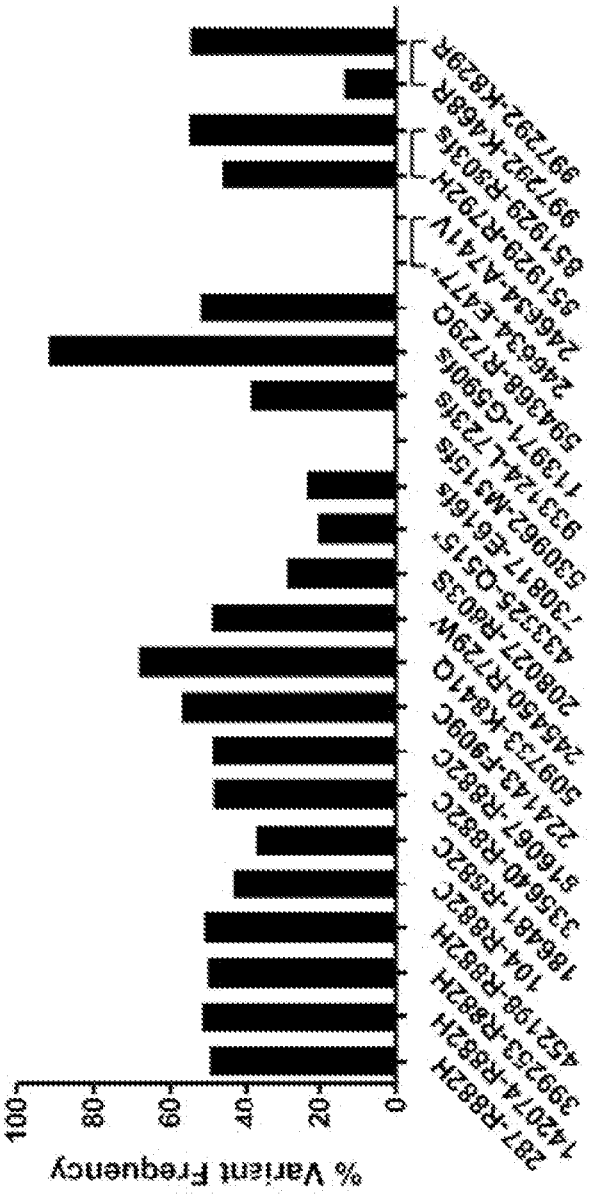

% Variant Frequency
0
20
40
60
80
100
287-R882H
1420T4-R882H
399253-R882H
452108-R882H
104-R882H
186481-R882C
335640-R882C
516067-R882C
224143-R882C
509723-P909C
245450-K841Q
208027-R729W
433225-D603S
730817-S515F
530962-E616S
933124-M315S
113971-L735S
59A368-G590S
246634-R729Q
246634-E477*
951929-A741V
851929-R729H
997292-R903S
997292-K468R
997292-K829R

FIG. 8

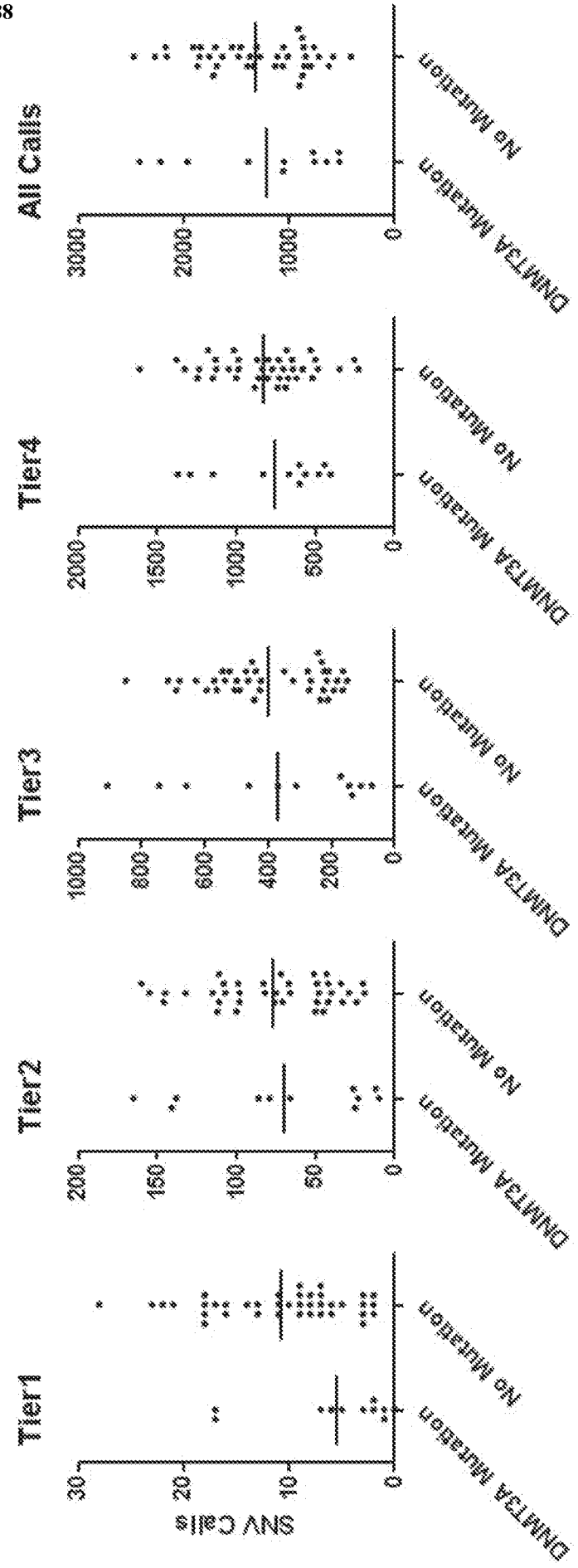

All Calls
Tier4
Tier3
Tier2
Tier1
SNV Calls
DNMT3A Mutation
No Mutation

FIG. 9

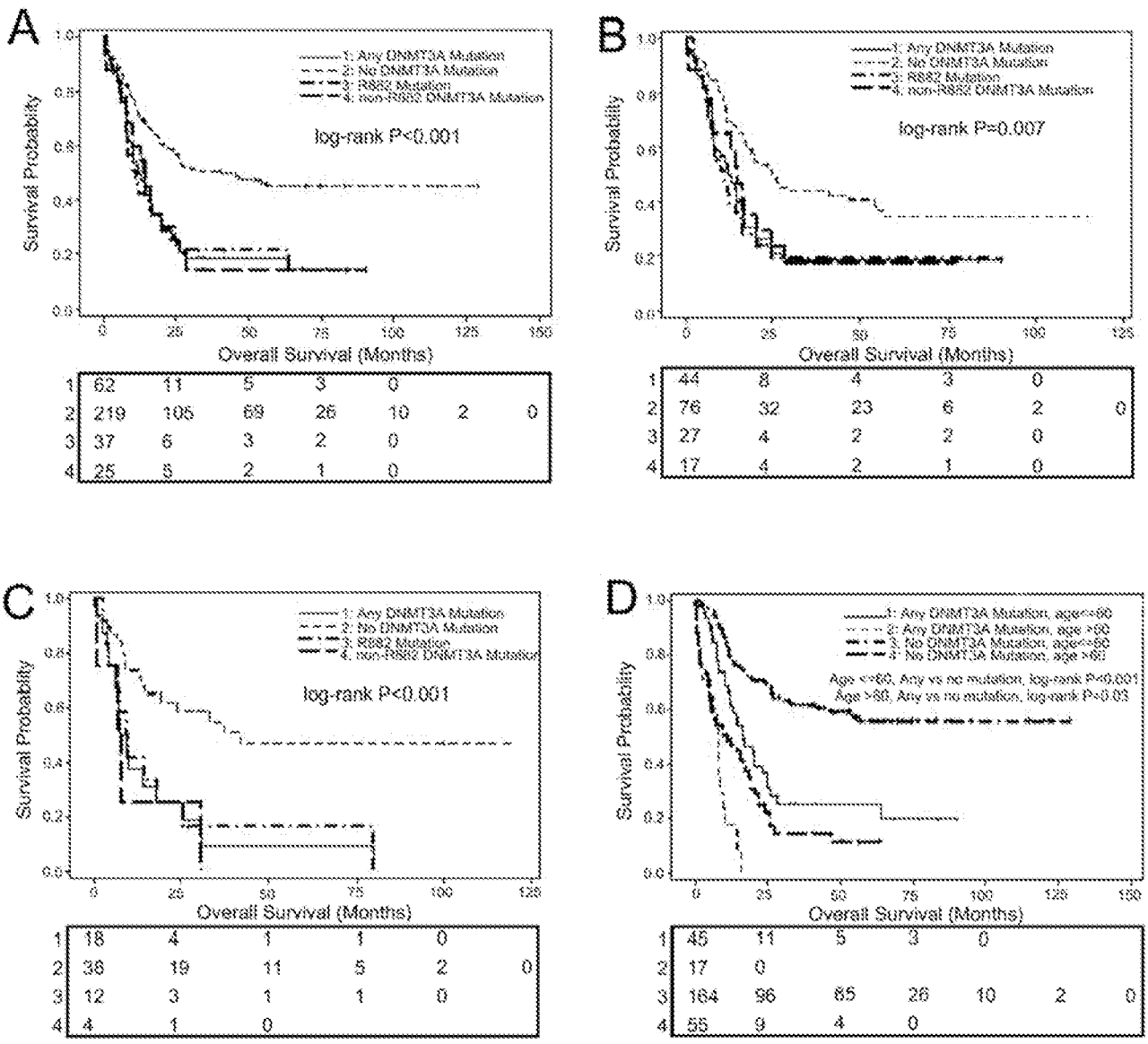

A
Survival Probability
1: Any DNMT3A Mutation
2: No DNMT3A Mutation
3: R882 Mutation
4: non-R882 DNMT3A Mutation
log-rank P<0.001
Overall Survival (Months)
1 62 11 5 3 0
2 219 105 69 26 10 2 0
3 37 6 3 2 0
4 25 6 2 1 0
B
Survival Probability
1: Any DNMT3A Mutation
2: No DNMT3A Mutation
3: R882 Mutation
4: non-R882 DNMT3A Mutation
log-rank P=0.007
Overall Survival (Months)
1 44 8 4 3 0
2 76 32 23 6 2 0
3 27 4 2 2 0
4 17 4 2 1 0
C
Survival Probability
1: Any DNMT3A Mutation
2: No DNMT3A Mutation
3: R882 Mutation
4: non-R882 DNMT3A Mutation
log-rank P<0.001
Overall Survival (Months)
1 18 4 1 1 0
2 38 19 11 5 2 0
3 12 3 1 1 0
4 4 1 0
D
Survival Probability
1: Any DNMT3A Mutation, age<=60
2: Any DNMT3A Mutation, age >60
3: No DNMT3A Mutation, age<=60
4: No DNMT3A Mutation, age >60
Age <=60, Any vs no mutation, log-rank P<0.001
Age >60, Any vs no mutation, log-rank P=0.03
Overall Survival (Months)
1 45 11 5 3 0
2 17 0
3 164 96 65 26 10 2 0
4 55 9 4 0

FIG.10

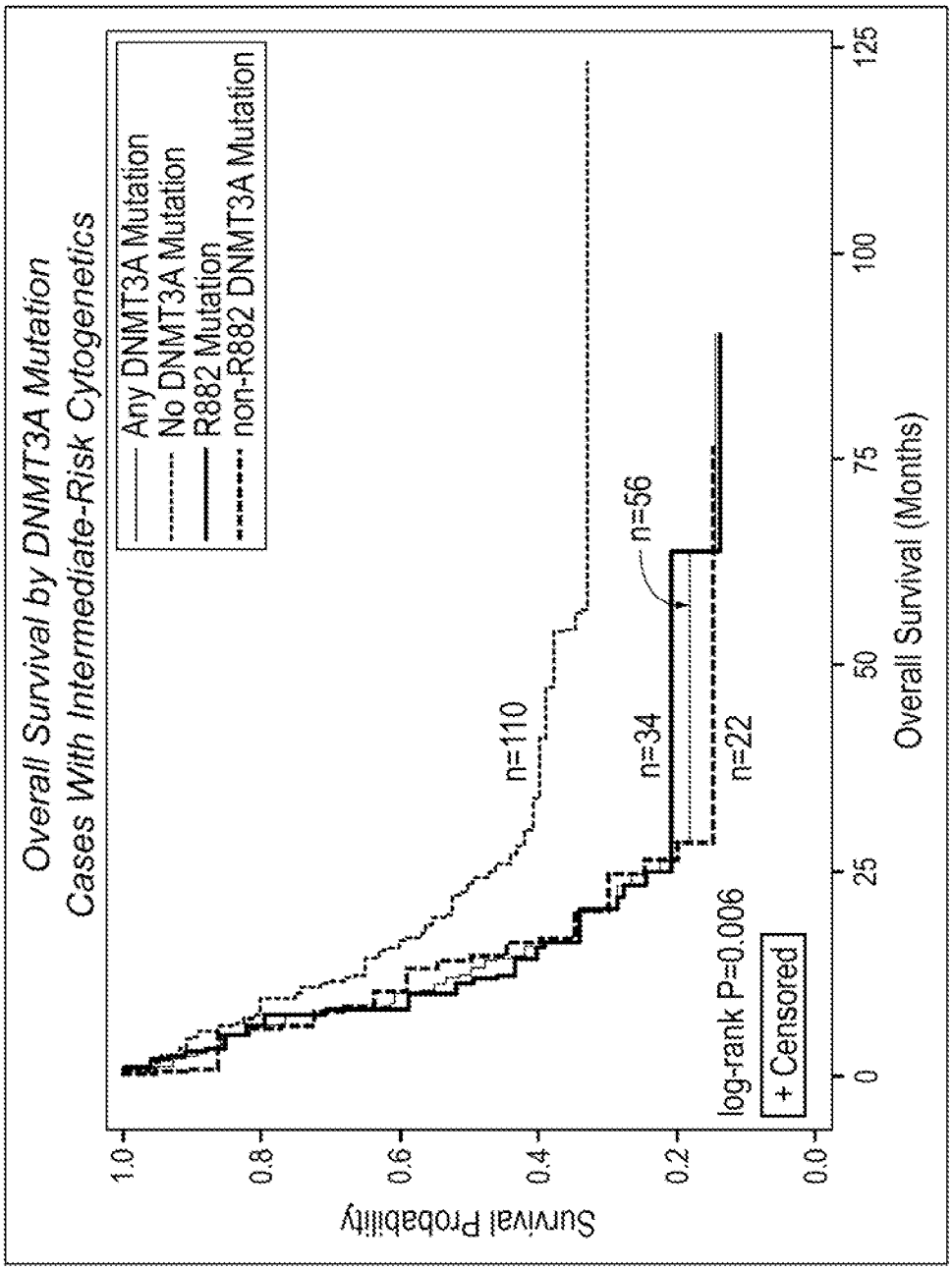

Overall Survival by DNMT3A Mutation
Cases With Intermediate-Risk Cytogenetics
Any DNMT3A Mutation
No DNMT3A Mutation
R882 Mutation
non-R882 DNMT3A Mutation
Survival Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
25
50
75
100
125
Overall Survival (Months)
n=110
n=56
n=34
n=22
log-rank P=0.006
+ Censored

FIG. 11

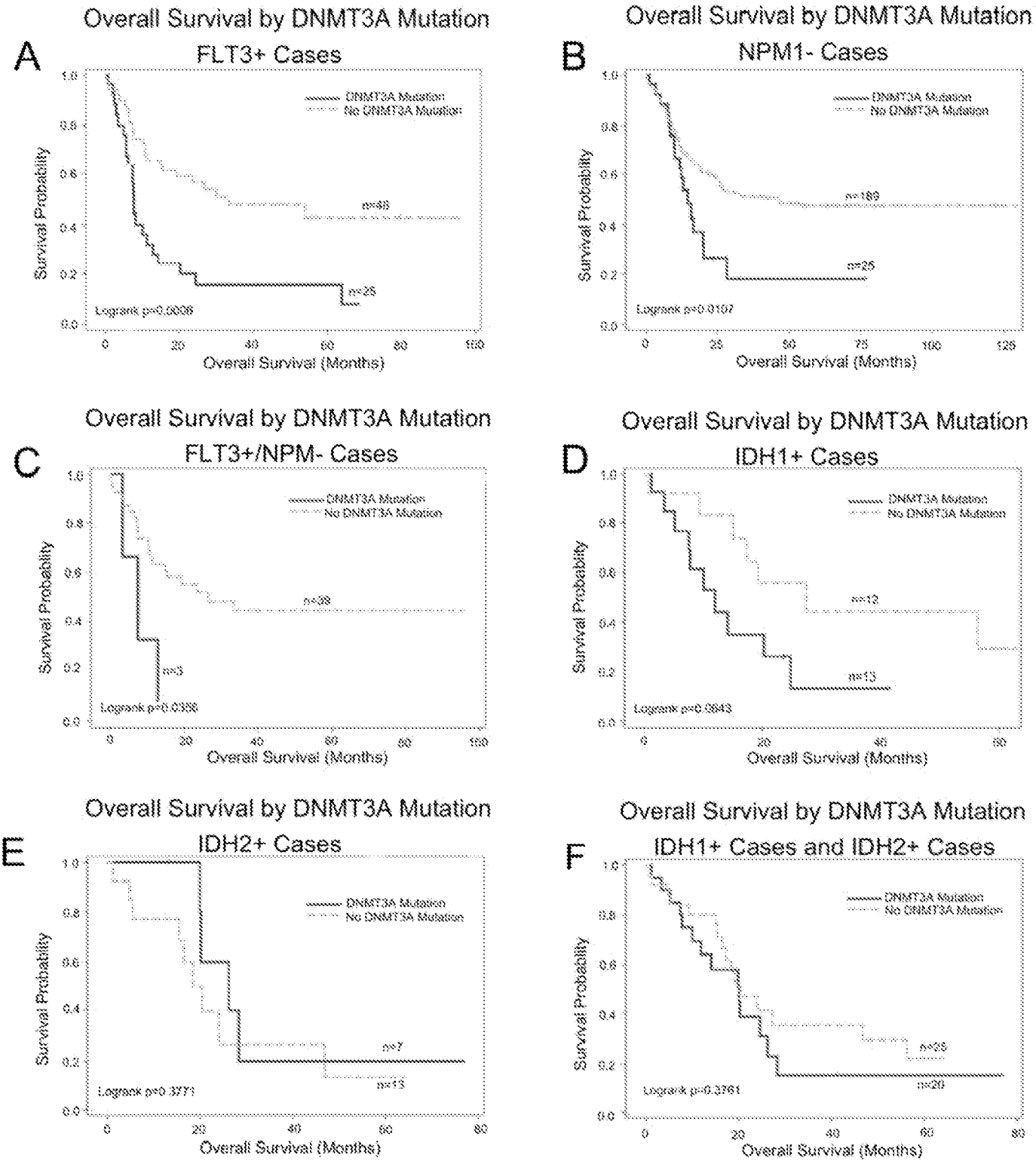

A
Overall Survival by DNMT3A Mutation
FLT3+ Cases
DNMT3A Mutation
No DNMT3A Mutation
n=46
n=25
Logrank p=0.0008
Survival Probability
Overall Survival (Months)
B
Overall Survival by DNMT3A Mutation
NPM1- Cases
DNMT3A Mutation
No DNMT3A Mutation
n=189
n=25
Logrank p=0.0107
Survival Probability
Overall Survival (Months)
C
Overall Survival by DNMT3A Mutation
FLT3+/NPM- Cases
DNMT3A Mutation
No DNMT3A Mutation
n=38
n=3
Logrank p=0.0358
Survival Probability
Overall Survival (Months)
D
Overall Survival by DNMT3A Mutation
IDH1+ Cases
DNMT3A Mutation
No DNMT3A Mutation
n=12
n=13
Logrank p=0.0843
Survival Probability
Overall Survival (Months)
E
Overall Survival by DNMT3A Mutation
IDH2+ Cases
DNMT3A Mutation
No DNMT3A Mutation
n=7
n=13
Logrank p=0.3771
Survival Probability
Overall Survival (Months)
F
Overall Survival by DNMT3A Mutation
IDH1+ Cases and IDH2+ Cases
DNMT3A Mutation
No DNMT3A Mutation
n=20
Logrank p=0.3761
Survival Probability
Overall Survival (Months)

FIG.12

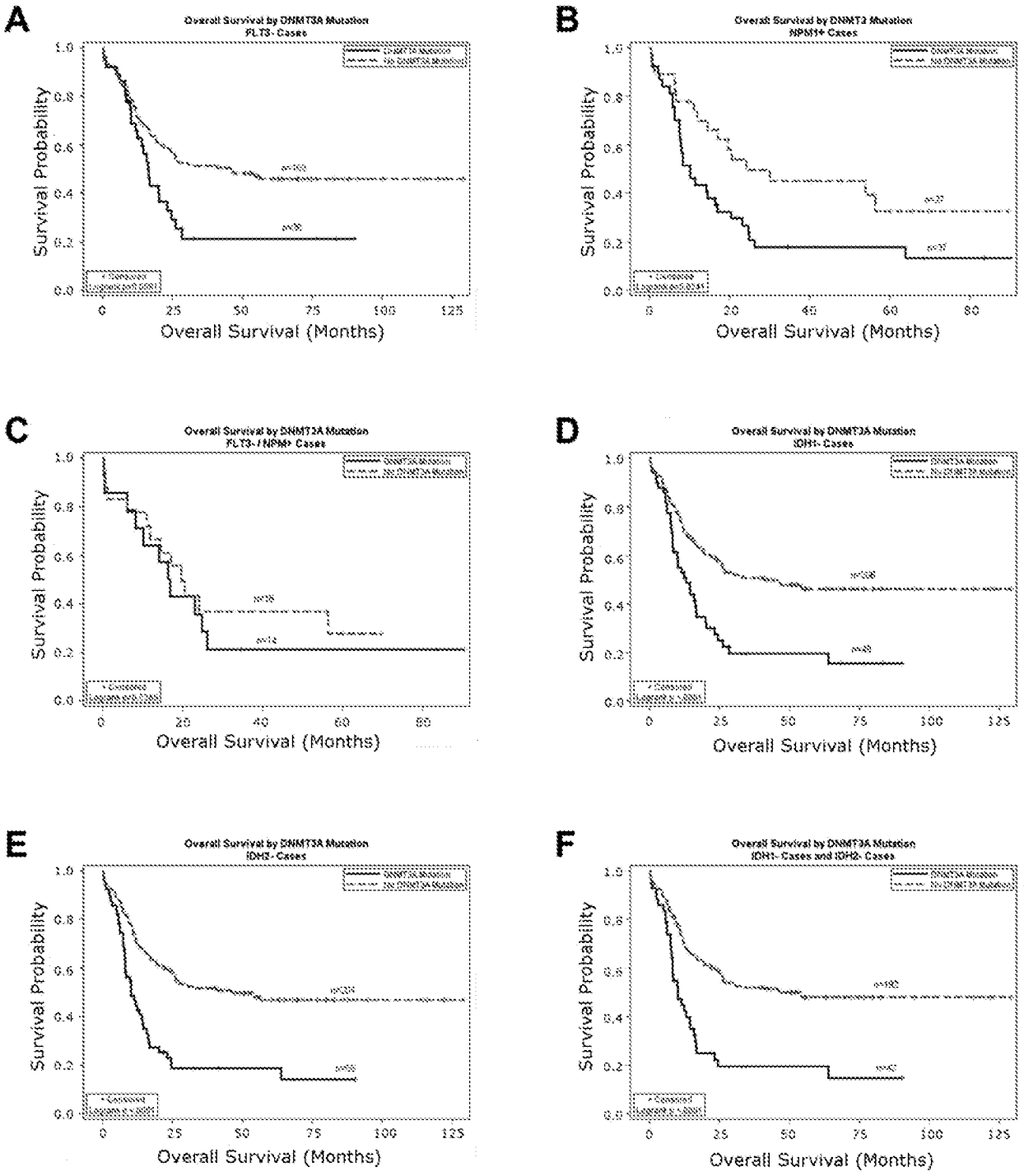

A
Overall Survival by DNMT3A Mutation
FLT3- Cases
B
Overall Survival by DNMT3 Mutation
NPM1+ Cases
C
Overall Survival by DNMT3A Mutation
FLT3- / NPM+ Cases
D
Overall Survival by DNMT3A Mutation
IDH1- Cases
E
Overall Survival by DNMT3A Mutation
IDH2- Cases
F
Overall Survival by DNMT3A Mutation
IDH1- Cases and IDH2- Cases
Survival Probability
Overall Survival (Months)

FIG. 13

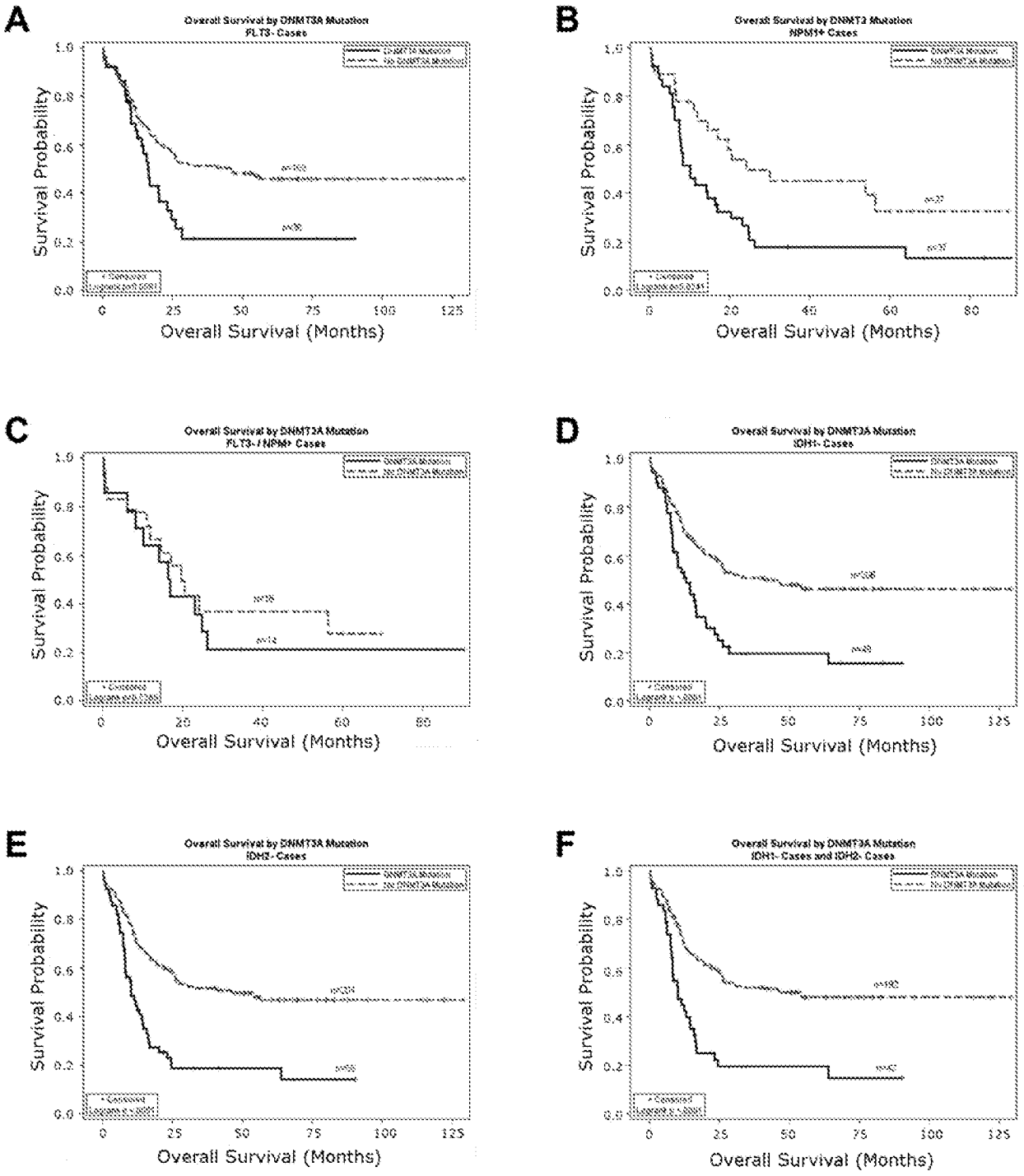

A
Overall Survival by DNMT3A Mutation
FLT3- Cases
Survival Probability
Overall Survival (Months)
B
Overall Survival by DNMT3 Mutation
NPM1+ Cases
Survival Probability
Overall Survival (Months)
C
Overall Survival by DNMT3A Mutation
FLT3- / NPM+ Cases
Survival Probability
Overall Survival (Months)
D
Overall Survival by DNMT3A Mutation
IDH1- Cases
Survival Probability
Overall Survival (Months)
E
Overall Survival by DNMT3A Mutation
IDH2- Cases
Survival Probability
Overall Survival (Months)
F
Overall Survival by DNMT3A Mutation
IDH1- Cases and IDH2- Cases
Survival Probability
Overall Survival (Months)

FIG. 14