

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Ley TJ, Ding L, Walter MJ, et al. *DNMT3A mutations in acute myeloid leukemia*. N Engl J Med 2010;363:2424-33. DOI: 10.1056/NEJMoa1005143.

(PDF last updated December 16, 2010.)

Supplementary Appendix

A. Supplementary Methods

Whole Genome Sequencing and mutation validation.

Whole genome sequencing (with paired end reads) and analysis was performed exactly as described.¹ Validation of putative somatic SNVs and indels with Sanger sequencing was performed as described.¹

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.¹

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.² 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 z-scored and Ward's unsupervised hierarchical clustering was performed (Spotfire DecisionSite 8.2 software, TIBCO).

Measuring 5-methylcytosine content by mass spectrometry.

Reagents and calibration solutions. Deoxyribonucleoside compounds were purchased from USB Corporation (Cleveland, OH) (5-methyl-2'-deoxycytidine monophosphate disodium salt (mdCMP), 2'-deoxyguanosine-5'-monophosphate disodium salt (dGMP), 2'-deoxyadenosine-5'-monophosphate free acid (dAMP)) or from Sigma Aldrich (2'-deoxycytidine monophosphate sodium salt (dCMP), thymidine (T), 2'-deoxyadenosine monohydrate (dA), 2'-deoxycytidine (dC), thymidine 5'-monophosphate disodium salt hydrate (TMP)). Water (LC-MS Chromasolv) and formic acid (1.00 mL ampoules, Puriss) were purchased from Fluka. Methanol (HPLC grade) was purchased from Fisher Scientific. The reference nucleosides and nucleoside monophosphates were weighed on an analytical balance and then dissolved in 1.00 mL of 5% methanol/0.1% formic acid. The concentrations of stock solutions accounted for the organic purity and water content as supplied by the manufacturer. The resulting stock solutions were used to prepare a 50 pmol/μL working stock solution, from which was prepared analytical solutions of 0.0050, 0.010, 0.050, 0.10, 0.50, 1.0, and 5.0 fmol/uL.

LC-ESI-MS/MS. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses were performed using a Shimadzu SCL-10A VP HPLC system (Columbia, MD) which was interfaced with a 4000 Q-Trap (MS/MS) mass

spectrometer (Applied Biosystems, Foster City, CA). Instrument operation and data acquisition were controlled with the Analyst software package (Version 1.4.2, Applied Biosystems). The HPLC system consisted of an autosampler, a binary pump, and a degasser. Chromatography and optimization of mass spectrometer source and MRM parameters was performed as previously described.³ Samples were analyzed at a flow rate of 0.22 ml/min on an Atlantis dC18 column (2.1 x 150 mm, 5 μ , Waters Corporation). Mobile phases were 0.1% formic acid (A) and 0.1% formic acid/100% methanol (B). Compounds of interest were eluted using the following gradient: linear gradient from 0% B to 12.5% B (0 to 15 min), linear gradient from 12.5% B to 80% B (15 to 17 min), 80% B (17 to 18 min), linear gradient from 100% B to 0% B (18 to 18.5 min), and 0% B (16 min. and after). Total run time was 25 min. Injection volume was 20 μ L. All MS/MS analyses were performed in positive ion mode by electrospray ionization (ESI) using a Turbo IonSpray source. Curtain, nebulizer, turbo, and collisionally-activated dissociation gases were 30 psig, 25 psig, 30 psig, 4 psig, respectively. Turbo heater temperature was 500 °C, and ion spray voltage was 5500 V. Dwell time for each MRM transition was 100 msec; total duty cycle time was 0.945 sec.

Preparation of DNA samples. Total genomic DNA was hydrolyzed exactly as described.⁴ From a test sample with a DNA concentration of ~ 35 ng/ μ L, tenfold dilutions gave signal responses within the working quantification range (0.1-100 fmols on column). After DNA hydrolysis, each of the test samples had a theoretical concentration of approximately 10 ng/ μ L DNA. Based on these initial estimates, the samples were analyzed after a three-fold dilution of the DNA hydrolysates. Several of the samples required repeat analysis with a five-fold dilution of the hydrolysate. Standard curves were prepared for each set of analyses on the same day. An injection volume of 20 μ L was used, resulting in on-column amounts of 0.10, 0.20, 1.0, 2.0, 10.0, 20.0, and 100.0 fmols for each nucleoside or nucleoside monophosphate compound. Peak areas and the corresponding molar amount of analyte were determined using Analyst software, and the resulting values were exported to an Excel spreadsheet for calculation of methylation ratios using two methods. The percentage methylation was calculated by dividing the measured level of mdCMP by the measured level of dGMP³ or by using the sum of mdCMP and dCMP.⁵

MeDIP-chip assays and analysis.

Genomic DNA from the bone marrow of five AML samples without the R882H mutation and five AML samples with the mutation were prepared for methylated DNA immunoprecipitation (MeDIP) exactly as described (http://app.roche-biochem.jp/pdf/products/microarray/trust_srv/DNA_Methylation_Sample_Prep_v4p1.pdf).

Comparative hybridization was performed on NimbleGen Human DNA Methylation 2.1M Deluxe Pro Arrays, which contain targets for all annotated CpG islands and most gene and miRNA promoters, using established protocols (http://www.nimblegen.com/products/lit/methylation_userguide_v7p1.pdf). The

log ratios of the probe intensities were scaled around zero by subtracting the biweight mean of all values from each log-ratio. The resulting 2.1M log ratios for ten samples were made available in GFF file format for data analysis.

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*.

1. The following criteria were applied to select probes among the 2.1M for the statistical test.
 - The mean \log_2 -ratio (signal intensity of immunoprecipitated sample/signal intensity of non-immunoprecipitated sample) for 5 *DNMT3A* mutant samples vs. the mean \log_2 -ratio of 5 *DNMT3A* wild type samples were different by a value of 1 or more
 - At least 4 out of 5 *DNMT3A* mutant \log_2 -ratios ≥ 1 , or at least 4 out of 5 *DNMT3A* wild type \log_2 -ratios ≥ 1 . This allows for the two outliers detected during unsupervised clustering.
2. For all probes selected, we calculated the mean of \log_2 -ratios for both groups using 5 consecutive probes, which include the current probe and two on either side.
3. We 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 , we considered this probe as differentially methylated.
4. We 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 rank-sum 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 analyses⁶ were performed with SAS PROC PHREG. Variables evaluated as potential confounders in the proportional hazards model were *FLT3* mutation, *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).

B. Supplementary Results and Discussion

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 high-dose 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.

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 risk-adapted 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.

Variant allele frequency of the *DNMT3A* mutation in sample 933124

Using PCR, we 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).⁷

5-Methylcytosine measurements in genomic DNA from AML samples

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.³ 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.⁴ 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 as described above in the Supplementary Methods. 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.³ 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. The percentages of mdCMP in the control and patient samples are summarized in **Supplementary Table 5**. Both calculation methods gave approximately the same values of 5-methyl deoxycytidine monophosphate in the patient samples.

Regions of differential methylation in genomes with R882H mutations detected by meDIP-chip

182 genomic regions had statistically different methylation levels (at specific genomic locations) that were related to *DNMT3A* mutation status (**Supplementary Figure 8B** and **Supplementary Table 6**). The methylation status of only two loci correlated precisely with the R882 mutation. An example of one such region near the *ESRP2* gene on chromosome 16 is shown in **Supplementary Figure 9**. In one region (shaded in gray), all non-mutated genomes (top 5 samples) have a methylation peak, and all *DNMT3A* mutant genomes (bottom 5 colored samples) have no detectable peak at the same location. Of note, there was no change in the level of expression of the nearest neighbor genes in the samples with the mutation.

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*.^{8,9} 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; **Supplementary Table 2**). 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.¹⁰ 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.¹⁰ 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.¹¹ 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.¹²,¹³ Since *DNMT3A* has been shown to interact with many proteins, including *DNMT3L*, SUMO-1, transcriptional repressors, histone modifying proteins, and TP53,¹⁴ 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.¹⁵

The diversity of mutations in the *DNMT3A* gene is reminiscent of the large number of mutations that inactivate classical tumor suppressor genes like *TP53* or *BRCA1*.^{16, 17} 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;¹⁸ several are associated with nonsense-mediated decay, clearly demonstrating loss of function (**Supplementary Figure 4**). The distribution of frameshift, nonsense, and missense mutations in *DNMT3A* are remarkably similar to the pattern of mutations seen in the *DNMT3B* gene in the ICF syndrome.^{19, 20} 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; see **Supplementary Figure 16**).^{18, 21} 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 (**Supplementary Table 2**). Of note, *DNMT3A* haploinsufficiency is not deleterious to mice,²² and complete loss of *DNMT3A* in the bone marrow cells of mice does not overtly alter hematopoiesis in the short term.²³ Although

many *DNMT3A* mutations clearly cause loss-of-function (nonsense mediated decay, truncation, deletion, some missense mutations), many more studies will be required to understand how these mutations contribute to AML pathogenesis.

Overall survival of patients with *DNMT3A* mutations and other common AML mutations

Several large studies have documented that adult AML patients harboring poor risk genotypes (*FLT3* mutations, *FLT3* mutations with *NPM* wild-type, and *IDH* mutants) have an inferior overall survival. We, therefore, examined the impact of *DNMT3A* mutations in patients with poor risk genotypes. *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 ($p=0.04$) (**Supplementary Figure 13**). 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* wild-type, *NPMc*, *IDH1* or *IDH2* wild-type ($p\leq 0.02$), but not for patients who were both *FLT3* wild-type and *NPMc* mutant (**Supplementary Figure 14**). 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 ($n=16$) (**Supplementary Figure 15**, $p=0.02$).

Multivariate Analysis of Event Free and Overall Survival

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 *DNMT3A* mutations, 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, *FLT3* mutations, 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)

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 (**Supplementary Table 7**). 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.

C. Supplementary Literature Cited.

1. Mardis ER, Ding L, Dooling DJ, et al. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 2009;361:1058-66.
2. Payton JE, Grieselhuber NR, Chang LW, et al. High throughput digital quantification of mRNA abundance in primary human acute myeloid leukemia samples. *J Clin Invest* 2009;119:1714-26.
3. Song L, James SR, Kazim L, Karpf AR. Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal Chem* 2005;77:504-10.
4. Burke DG, Griffiths K, Kassir Z, Emslie K. Accurate measurement of DNA methylation that is traceable to the international system of units. *Anal Chem* 2009;81:7294-301.
5. Friso S, Choi SW, Dolnikowski GG, Selhub J. A method to assess genomic DNA methylation using high-performance liquid chromatography/electrospray ionization mass spectrometry. *Anal Chem* 2002;74:4526-31.
6. Hosmer DW, Lemeshow S. Regression Modeling of Time to Event Data. In: Applied Survival Analysis: John Wiley & Sons, Inc.; 1999.
7. Ley TJ, Mardis ER, Ding L, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008;456:66-72.
8. Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 2009;462:739-44.
9. Ward PS, Patel J, Wise DR, et al. The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell*;17:225-34.
10. Yamashita Y, Yuan J, Suetake I, et al. Array-based genomic resequencing of human leukemia. *Oncogene*;29:3723-31.
11. Gowher H, Loutchanwoot P, Vorobjeva O, et al. Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. *J Mol Biol* 2006;357:928-41.
12. Xu GL, Bestor TH, Bourc'his D, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 1999;402:187-91.
13. Hansen RS, Wijmenga C, Luo P, et al. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci U S A* 1999;96:14412-7.
14. Wang YA, Kamarova Y, Shen KC, et al. DNA methyltransferase-3a interacts with p53 and represses p53-mediated gene expression. *Cancer Biol Ther* 2005;4:1138-43.

15. Jones PA, Liang G. Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 2009;10:805-11.
16. Walsh T, Casadei S, Coats KH, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA* 2006;295:1379-88.
17. Varley JM. Germline TP53 mutations and Li-Fraumeni syndrome. *Hum Mutat* 2003;21:313-20.
18. Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J Biol Chem* 2005;280:13341-8.
19. Ehrlich M, Sanchez C, Shao C, et al. ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. *Autoimmunity* 2008;41:253-71.
20. Jiang YL, Rigolet M, Bourc'his D, et al. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Hum Mutat* 2005;25:56-63.
21. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* 2007;449:248-51.
22. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;99:247-57.
23. Tadokoro Y, Ema H, Okano M, Li E, Nakauchi H. De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. *J Exp Med* 2007;204:715-22.
24. Klimasauskas S, Kumar S, Roberts RJ, Cheng X. Hhal methyltransferase flips its target base out of the DNA helix. *Cell* 1994;76:357-69.

D. Supplementary Figure Legends.

Supplementary Figure 1. Summary of *DNMT3A* SNVs found in CALGB AML samples. The locations of validated SNVs (single nucleotide variants) in the *DNMT3A* gene are shown for 94 AML samples obtained from CALGB are shown. These cases are described in **Supplementary Table 1** and Mardis, et al¹. 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 (proline-tryptophan-tryptophan-proline), ZNF (zinc finger), and MTase (methyltransferase) domains are shown. Each patient with a *DNMT3A* mutation is designated with a circle.

Supplementary Figure 2. Map of a 1.5 Mb deletion encompassing *DNMT3A*. 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*.

Supplementary Figure 3. Expression data for *DNMT* genes in AML cases.

Expression arrays were performed for 180 AML cases, and 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).²

Supplementary Figure 4. cDNA readcounts for *DNMT3A* mutations. Variant allele frequencies from cDNAs obtained from the bone marrow RNA of 21 different AML patients are shown. Regions containing the mutations were amplified by RT-PCR, and then sequenced using the 454-FLX platform to obtain deep readcounts for variant alleles. The data used to generate this figure are presented in **Supplementary Table 3**. Below each bar, the UPN for the patient sample is shown, followed by the mutation annotation.

Supplementary Figure 5. Total numbers of putative somatic single nucleotide variants (SNVs) in sequenced AML genomes with and without *DNMT3A* mutations. The data for this figure were extracted from whole genome sequencing data from 38 AML samples with normal cytogenetics, and a variety of FAB subtypes (M0,1,2,4, and 5). 11 of these samples had validated somatic mutations in *DNMT3A*, and 27 did not. The number of high confidence somatic SNVs for each “Tier” of each genome (described in Mardis, et al¹) are plotted as a function of *DNMT3A* mutation status. Each genome is represented by a single point. The mean number of SNVs called is shown as a bar. The mean number of Tier 1 (genic) SNVs is significantly reduced in samples with *DNMT3A* mutations ($p=0.0216$), but it is not different for any other Tier, nor for total mutations per genome (“All Calls”).

Supplementary Figure 6. Summary of somatic SNV types and frequencies in AML genomes. From the 38 sequenced AML genomes described in **Supplementary Figure 5**, we classified SNVs based on the nature of the base change identified, and the *DNMT3A* mutation status of that genome. All mutation types occur at nearly identical frequencies in genomes with *DNMT3A* mutations ($n=11$) and without ($n=27$). None of the differences is statistically different.

Supplementary Figure 7. Summary of 5-methylcytosine content data from 51 AML genomes. The percentage of 5-methylcytosine as a fraction of total cytosine is plotted from data presented in **Supplementary Table 5**. The 5-methylcytosine content of samples with and without *DNMT3A* mutations was not significantly different for all samples, nor for samples with R882 mutations.

Supplementary Figure 8. Heatmaps of methylated genomic regions defined by Methyl-Chip analysis. (A) Global heatmap of MeDIP-chip analysis using the NimbleGen 2.1M deluxe promoter array (containing all annotated CpG islands and promoters) using DNA samples from 5 AML samples without *DNMT3A* mutations (WT) or patients with R882H mutations. Each mutant sample was matched with a WT sample for FAB type and bone marrow blast %. Regions with reduced methylation are shown in blue, and with increased methylation, in green. The average log2 ratios from 56,392 clusters with at least 20 probes in the cluster region were displayed by chromosomal position. The global pattern of methylation for CpG islands was unchanged in genomes with the R882H mutation. (B) Locations of 182 loci that were significantly hypomethylated (for at least three consecutive probes) in the five samples with *DNMT3A* mutations. Locations of each region and the methylation values are available in **Supplementary Table 6**. The data for the region near *ESRP2*, detailed in **Supplementary Figure 9**, is indicated.

Supplementary Figure 9. Differential methylation of the *ESRP2* locus in samples with *DNMT3A* mutations. Log2-normalized data are shown for a genomic region on chromosome 16 near the *ESRP2* gene. The data for five samples with wild-type *DNMT3A* genes (WT) are shown in black and gray at the top. The data for the five samples with R882H mutations are shown at the bottom. The shaded region is methylated in the WT samples, and unmethylated in the R882H samples. The center of this region is located 430 bp upstream from the 5' end of the *ESRP2* gene. The locations of known CpG islands are shown at the bottom.

Supplementary Figure 10. Lack of a gene expression signature correlated with *DNMT3A* mutations. Heatmap depicting unsupervised clustering of expression array data (Affymetrix U133+2) from 180 *de novo* AML cases. Relatively increased expression is indicated by red; relatively decreased expression by green; intermediate expression by black. The color legend below the heatmap indicates clinicopathological data; each row corresponds to the position of the row above the heatmap. The top row indicates the common AML karyotypes and cytogenetic risk groups: green = t(15;17), yellow = inv(16) or t(8;21), dark blue = normal karyotype, light blue = intermediate risk, gray = adverse risk including complex (>3 changes), white = not done. The middle row indicates the FAB subtype: orange = M0/M1, yellow = M2, dark green = M3, teal = M4, purple = M5, dark pink = M6, light pink = M7. The bottom row shows *DNMT3A* mutation status: white is wild type for *DNMT3A*, red indicates any of the R882 mutations, and yellow indicates non-R882 *DNMT3A* mutations.

Supplementary Figure 11. Heatmap of unsupervised clustering analysis of 76 AML cases with normal karyotypes. Heatmap depicting unsupervised clustering of expression array data (Affymetrix U133+2) from 76 *de novo* AML cases (Spotfire software). Relatively increased expression is indicated by red color; relatively decreased expression by green color; intermediate expression by

black color. The color legend below the heatmap indicates clinicopathological data. The top row shows *DNMT3A* mutation status: white is wild type for *DNMT3A*, red indicates one of the R882 mutations, and yellow indicates one of the non-R882 *DNMT3A* mutations. The bottom row indicates the FAB subtype: orange = M0/M1, yellow = M2, dark green = M3, teal = M4, purple = M5, dark pink = M6, light pink = M7.

Supplementary Figure 12. Overall survival of AML patients with *DNMT3A* and intermediate risk cytogenetics. Overall survival of AML patients without a *DNMT3A* mutation (dashed red line) compared to those with any *DNMT3A* mutation (blue line), with an R882 mutation (dashed green line), or with a non-R882 mutation (dashed brown line).

Supplementary Figure 13. Overall survival of AML patients with *DNMT3A* mutations and poor risk genotypes. (A) Overall survival of AML patients with (blue line) or without (dashed red line) *DNMT3A* mutations and a concurrent *FLT3* mutation (either ITD or any other mutation). Note that this dataset represents all *FLT3* mutations (ITD or any other mutation) in contrast to **Figure 3C**, where only *FLT3* ITD mutations are considered. (B) *NPM1* wild type patients, (C) patients with *FLT3* mutations (either ITD or any other mutation) and *NPM1* wild type, (D) patients with *IDH1* mutations, (E) patients with *IDH2* mutations, (F) patients with *IDH1* or *IDH2* mutations.

Supplementary Figure 14. Overall survival of AML patients with *DNMT3A* mutations and favorable risk genotypes. (A) Overall survival of AML patients with (blue line) or without (dashed red line) *DNMT3A* mutations that are *FLT3* wild type, (B) patients with *NPMc* mutations, (C) patients with wild type *FLT3* and *NPMc* mutations, (D) patients with wild type *IDH1*, (E) patients with wild type *IDH2*, or (F) patients with wild type *IDH1* and *IDH2*.

Supplementary Figure 15. Overall survival of AML patients with *DNMT3A* and *FLT3* ITD mutations. Overall survival of AML patients with *DNMT3A* mutations with (blue line) and without (dashed red line) *FLT3* ITD mutations.

Supplementary Figure 16. Locations of AML mutations within the three dimensional structure of *DNMT3A*. Side view of tetrameric *DNMT3A* (murine AA 623 – 908 (teal and dark blue) which corresponds to conserved human AA 627 - 912) and *DNMT3L* (light blue), interacting with S-adenosyl-L-methionine (AdoMet - white spacefill).^{21,24} Jia et al suggested that DNA binds this tetrameric complex along a structural groove (indicated by arrow).²¹ Amino acids mutated along the proposed DNA binding groove are highlighted in red; amino acids mutated along the *DNMT3L* interacting helix are highlighted in green. 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)²¹. The R882 side-chain protrudes into the DNA binding groove, and mutations at this position decrease both DNA binding and catalytic activity.¹¹

Amino acids P718, R729, K829, R803, and F909 all occur along the DNA binding groove, and their side-chains are positioned into the solute; mutations at these positions may therefore alter DNA affinity or specificity. R729 and R736 occur along the DNMT3L interacting helix, and mutations at these positions may alter heterodimerization. A741V does not obviously alter the DNMT3L interacting helix, but this mutation occurs in conjunction with E477* and may not be functionally relevant, since this mRNA is subject to nonsense mediated decay.

Supplementary Figure 17. 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*). Patients with favorable cytogenetic risk profiles had either no mutations in these genes, or in *FLT3* only. The presence of a *FLT3* mutation did not alter OS in this group (data not shown). For all other cases, the impact of the numbers of mutations in the common AML genes on overall survival is shown. *IDH1* and *IDH2* mutations were pooled (these mutations are mutually exclusive). Only 4 cases had 4 mutations, so they were pooled with cases with 3 mutations (n=23) for clarity.

Supplementary Figure 18. Overall survival of 187 AML patients collected at Washington University stratified by *DNMT3A* mutation status and allogeneic transplantation status. **(A)** 48 patients without *DNMT3A* mutations were treated with an allogeneic transplantation at any time during their treatment course. 41 patients had an allogeneic transplant only and 7 had an allogeneic transplant following an autologous transplant. 94 patients without *DNMT3A* mutations were treated with an autologous transplant (n=7) or no transplant (n=87) during their treatment course. **(B)** Patients with *DNMT3A* mutations were treated with (n=16) or without (n=29) an allogeneic transplant during their treatment course, as defined in Panel A. 13 patients had an allogeneic transplant only and 3 had an allogeneic transplant following an autologous transplant. 29 patients with *DNMT3A* mutations were treated with an autologous transplant (n=7) or no transplant (n=22) during their treatment course. **(C)** 64 patients were treated with an allogeneic transplantation at any time during their treatment course. 16 patients had a *DNMT3A* mutation and 48 patients did not have a *DNMT3A* mutation. Overall survival of patients treated with an allogeneic transplant was independent of *DNMT3A* mutation status (p=0.30).

Supplementary Table 1. Patient Characteristics.

Supplementary Table 2. Summary of *DNMT3A* Mutations.

Supplementary Table 3. *DNMT3A* cDNA variant allele readcounts.

Supplementary Table 4. Mutation types in *DNMT3A* mutant vs. wild type genomes.

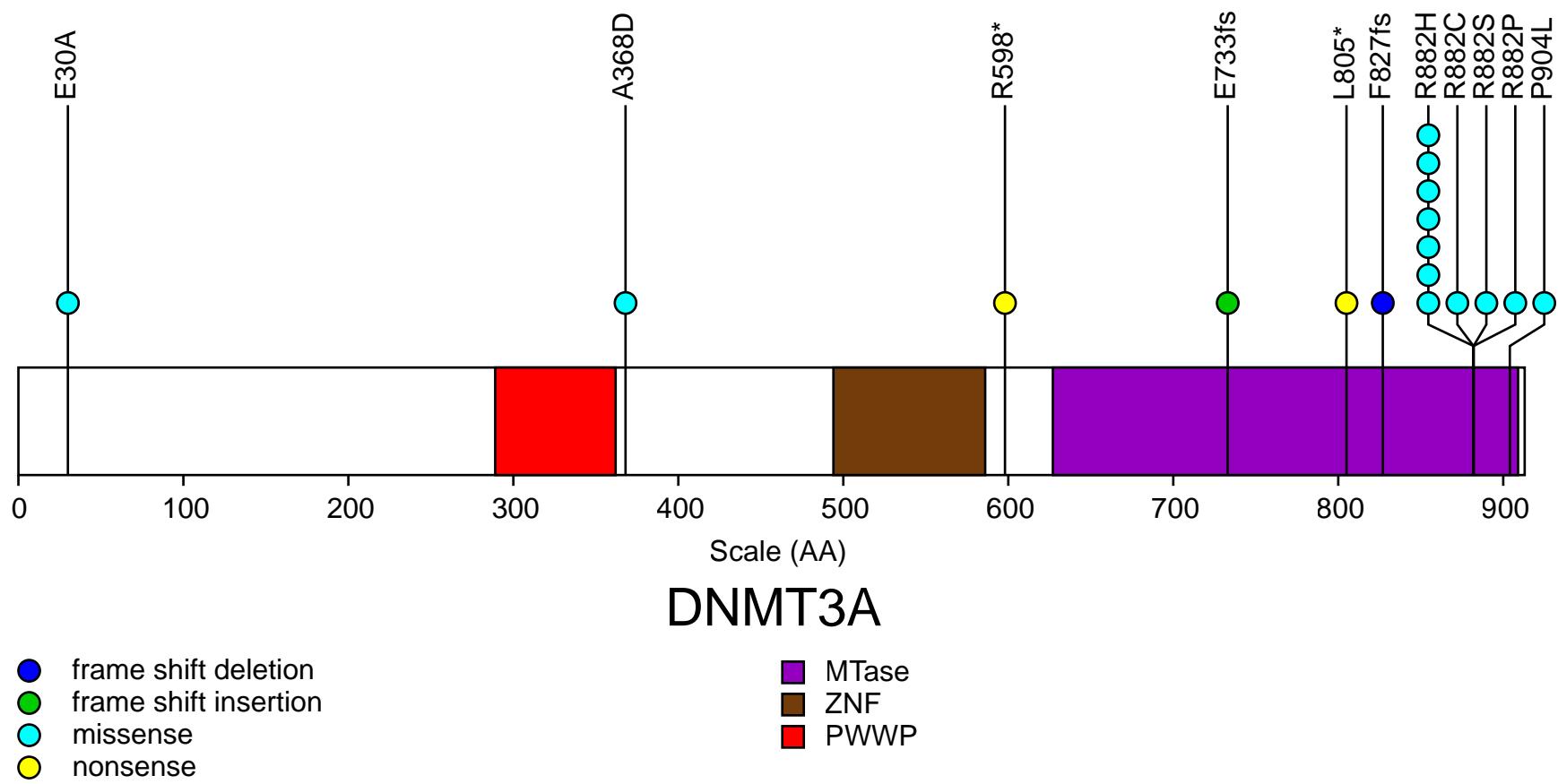
Supplementary Table 5. 5-methylcytosine content in *DNMT3A* mutant vs. wild type genomes.

Supplementary Table 6. Differentially methylated loci in *DNMT3A* genomes defined by meDIP-chip analysis.

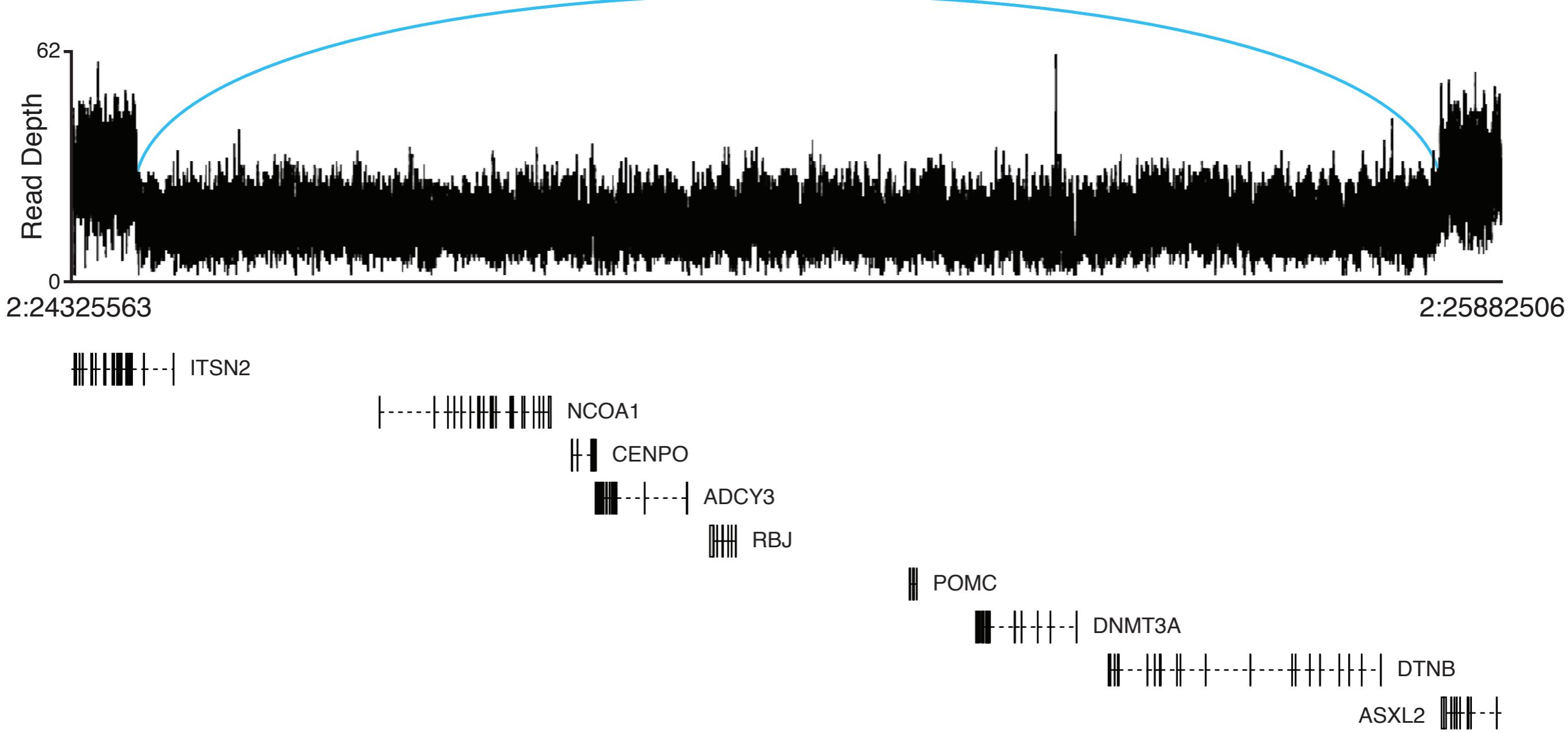
Supplementary Table 7. *DNMT3A* SNP frequencies.

Supplementary Table 8. *DNMT3A* primers used to sequence all coding exons.

Supplementary Table 9. *DNMT3L* primers used to sequence all coding exons.

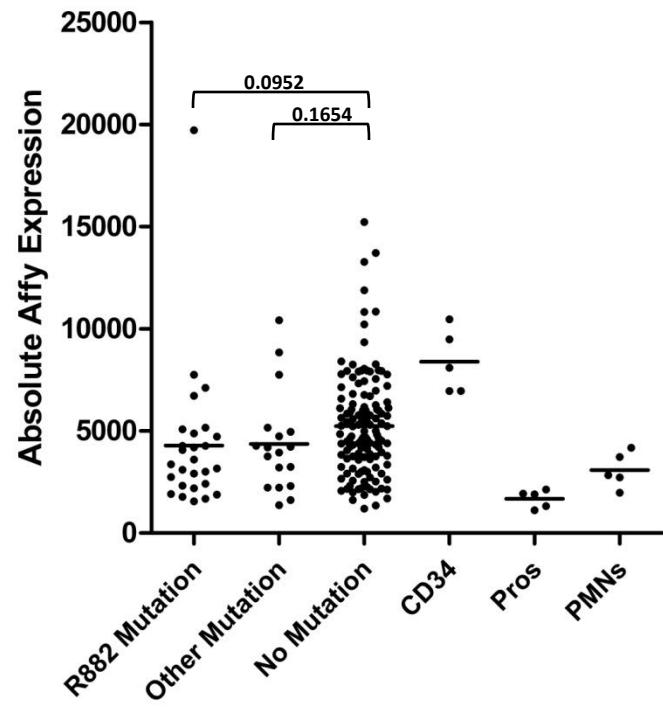


Supplementary Figure 1.

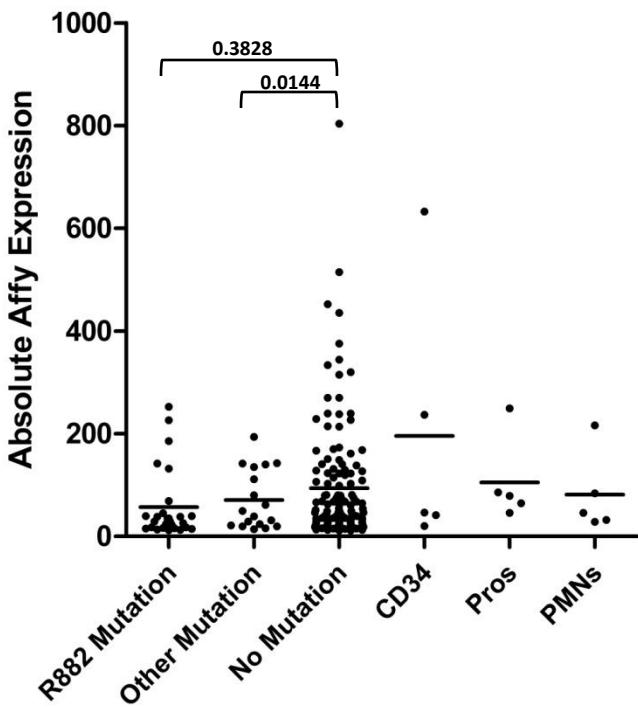


Supplementary Figure 2.

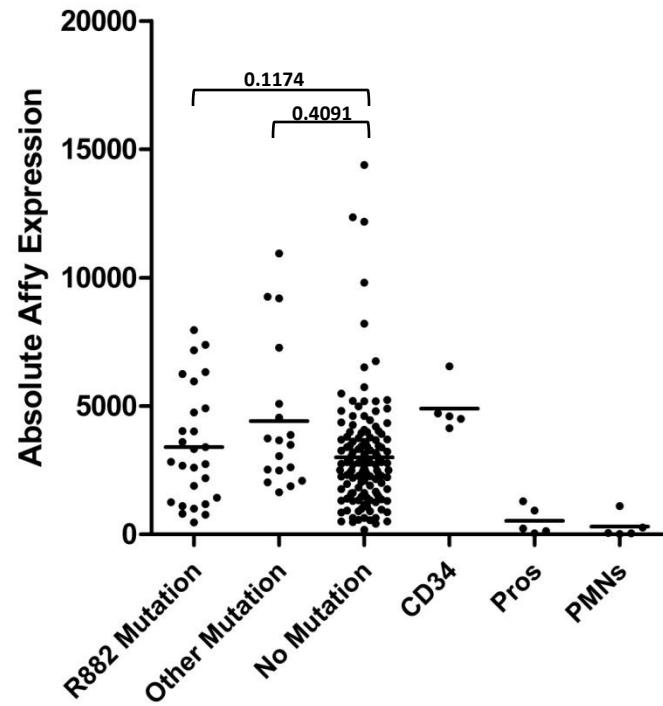
DNMT3A



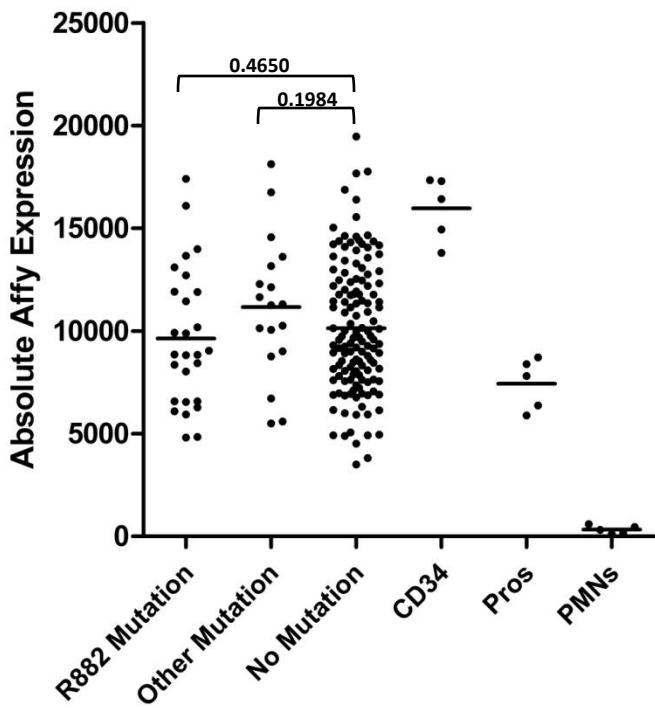
DNMT3L



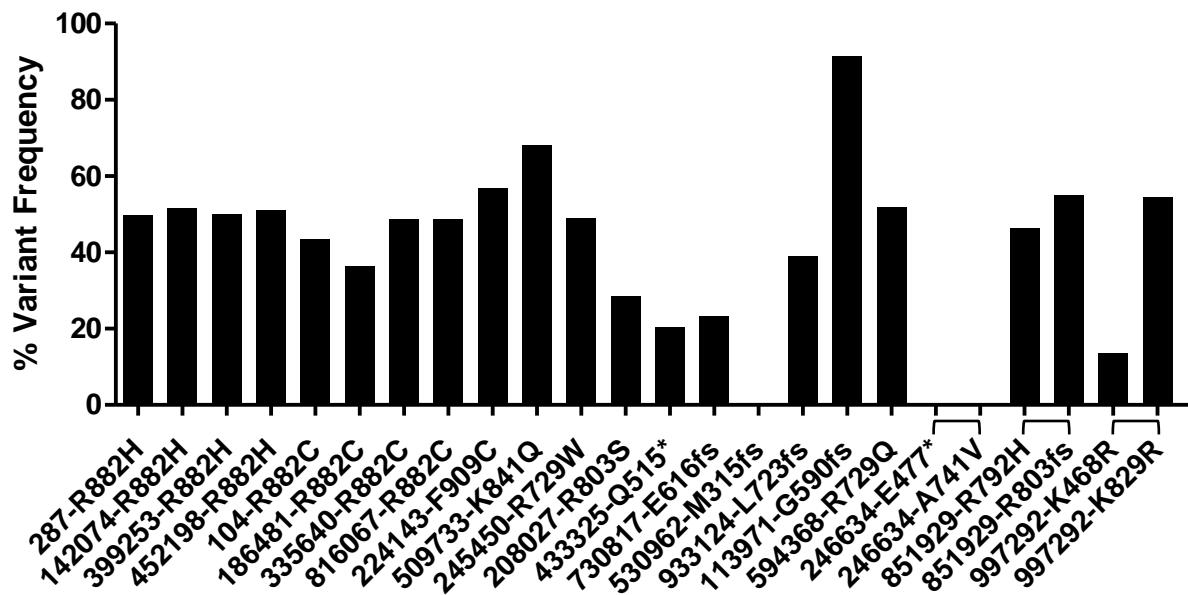
DNMT3B



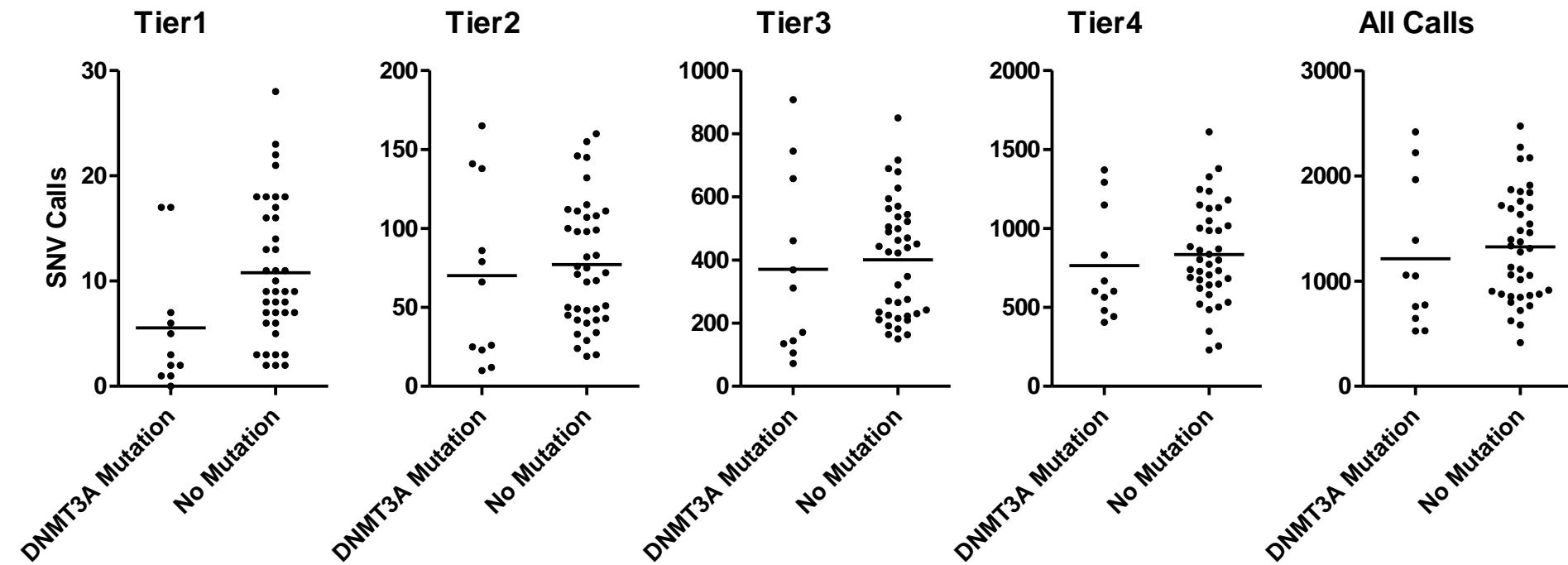
DNMT1



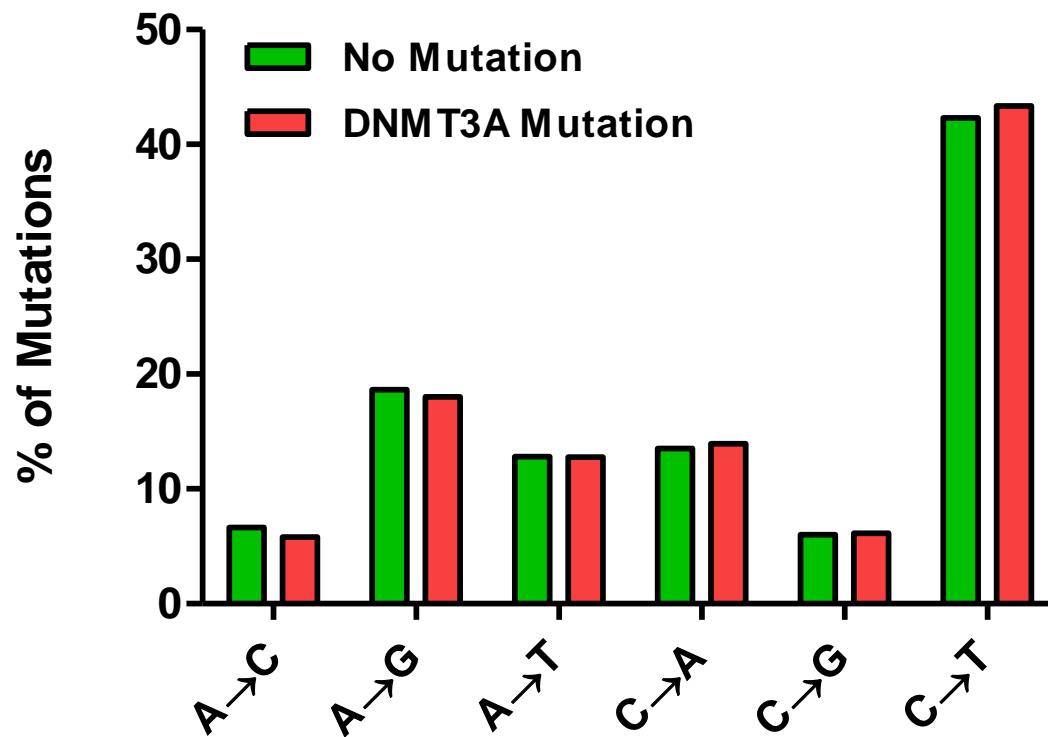
Supplementary Figure 3.



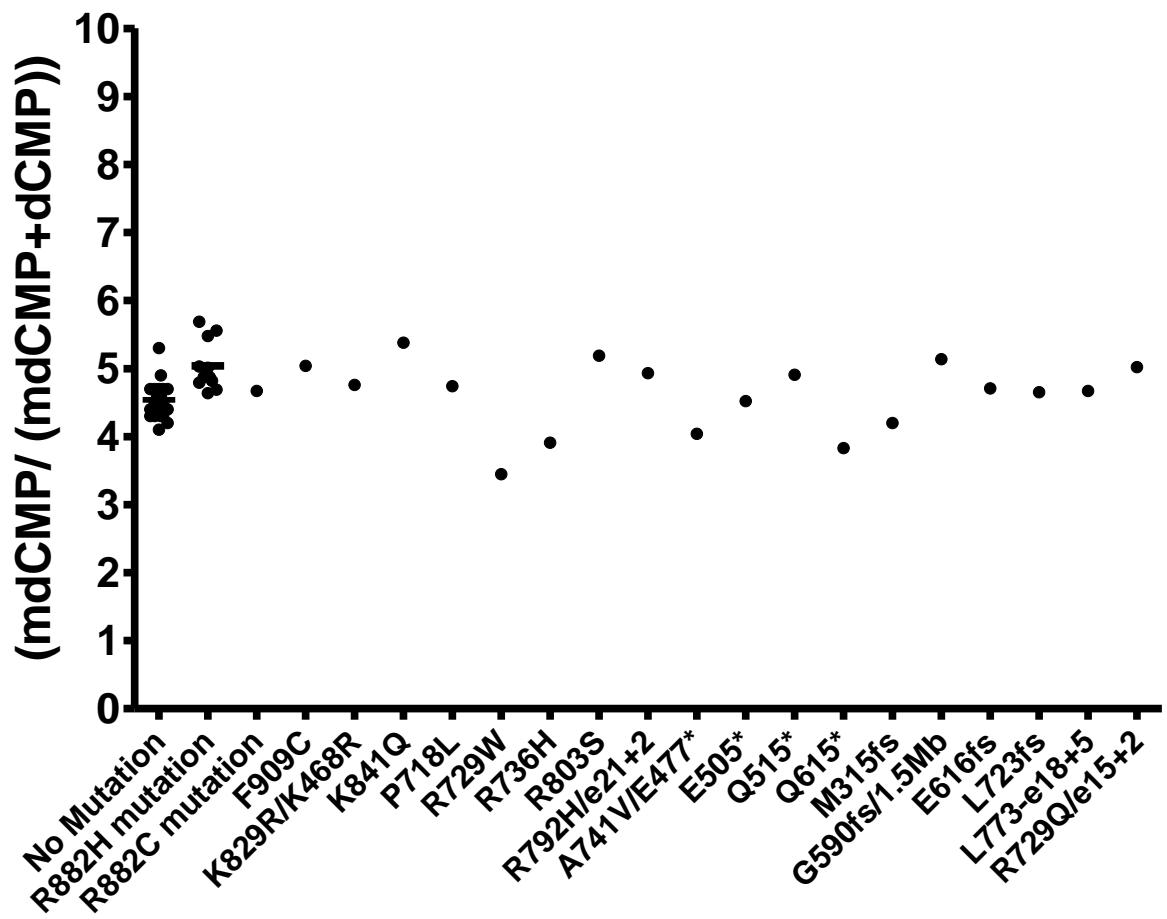
Supplementary Figure 4.



Supplementary Figure 5.



Supplementary Figure 6.



Supplementary Figure 7.

low methylation

high methylation

Supplementary Figure 8A

chromosome

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

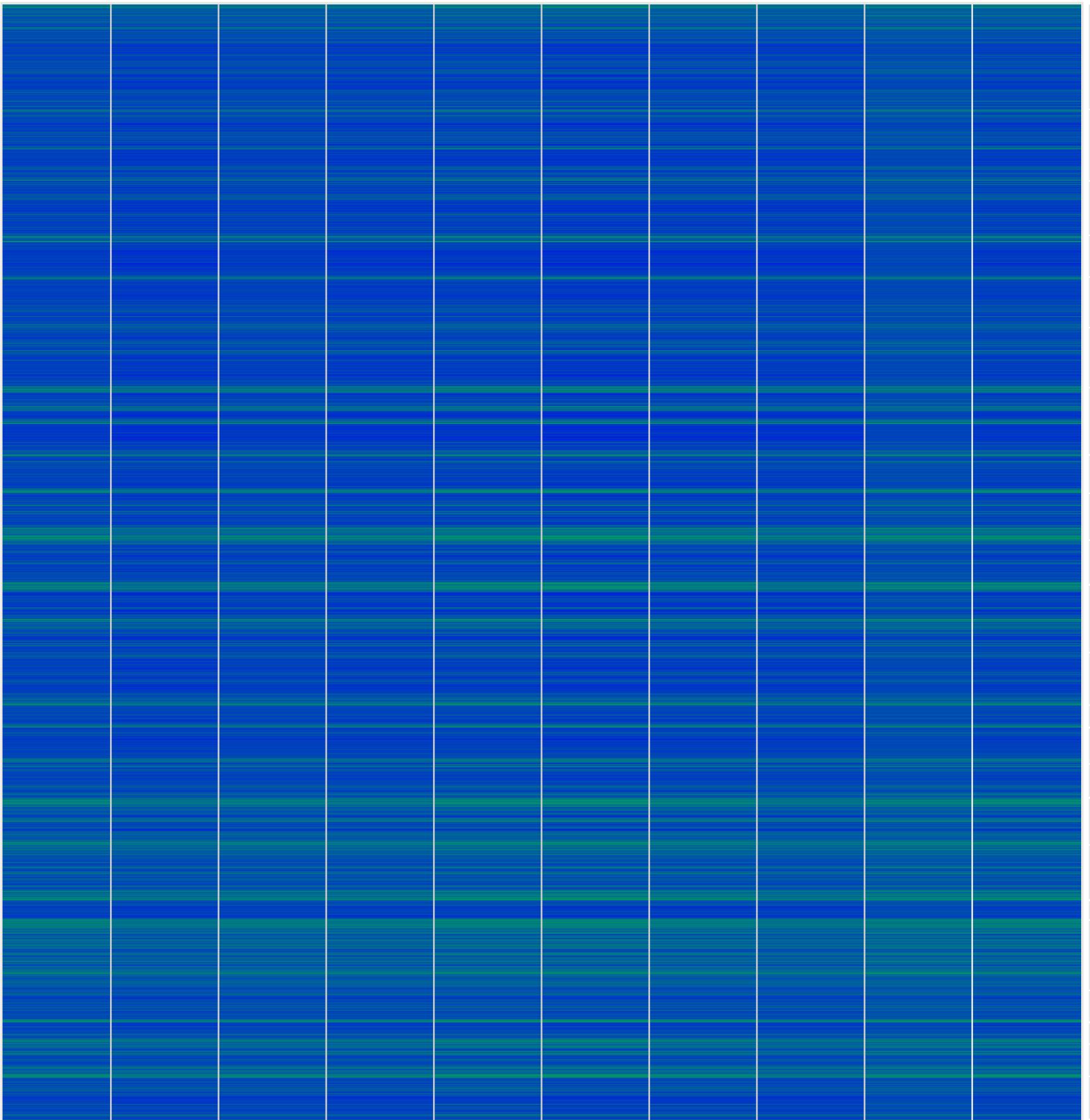
20

21

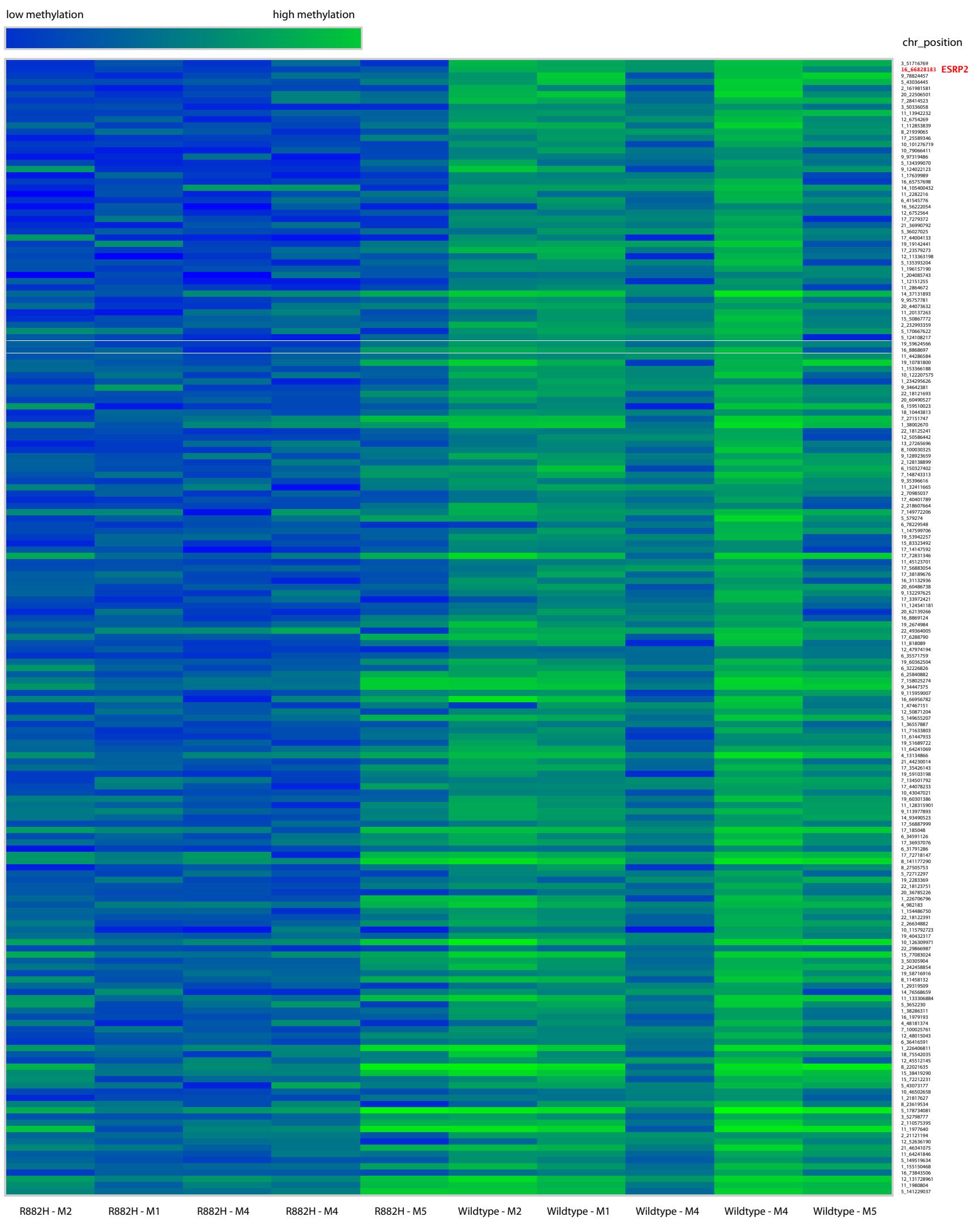
22

X

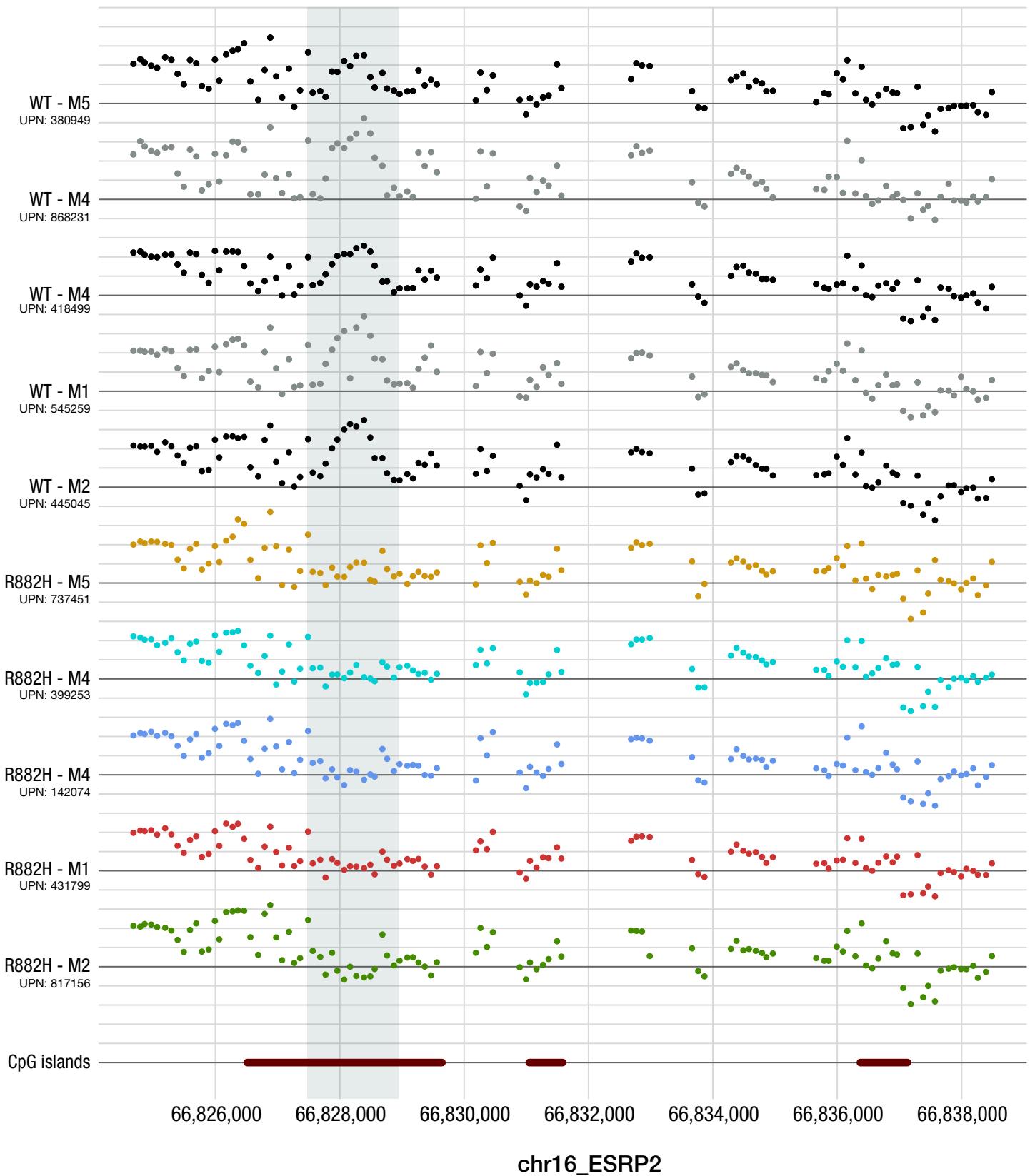
Y

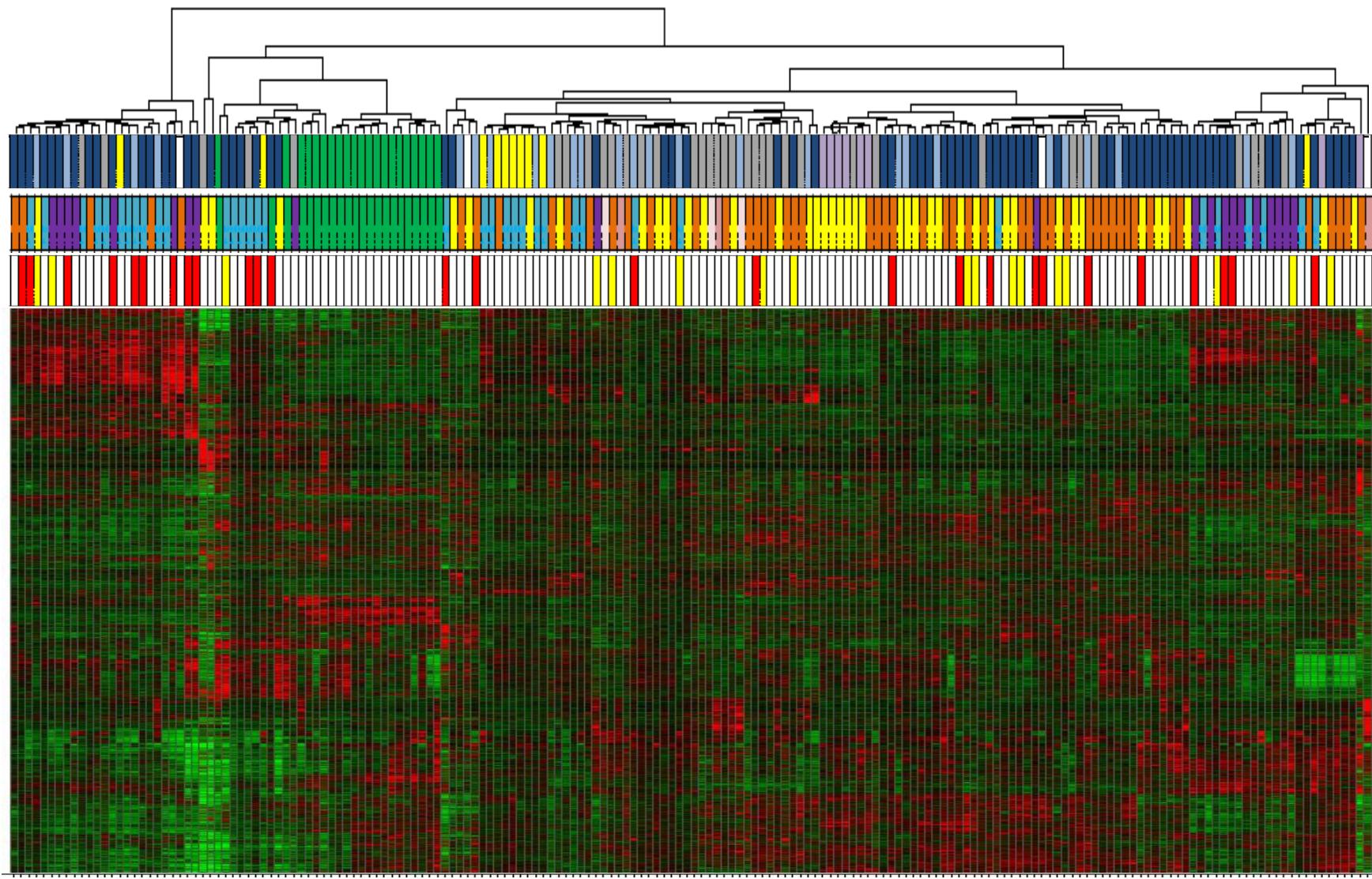
R882H - M2
UPN: 817156R882H - M1
UPN: 431799R882H - M4
UPN: 142074R882H - M4
UPN: 399253R882H - M5
UPN: 737451Wildtype - M2
UPN: 445045Wildtype - M1
UPN: 545259Wildtype - M4
UPN: 418499Wildtype - M4
UPN: 868231Wildtype - M5
UPN: 380949

Supplemental Figure 8B



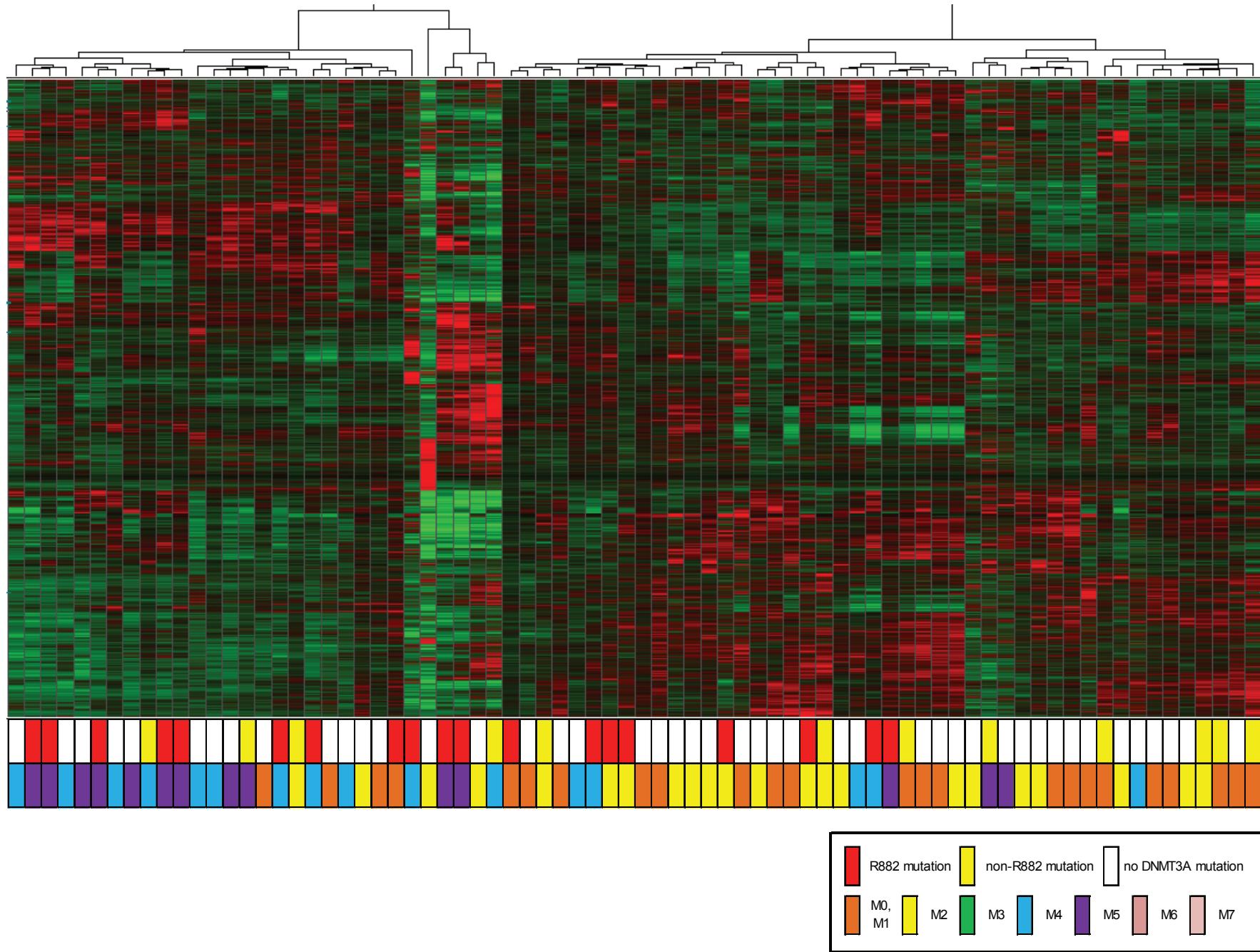
Supplementary Figure 9



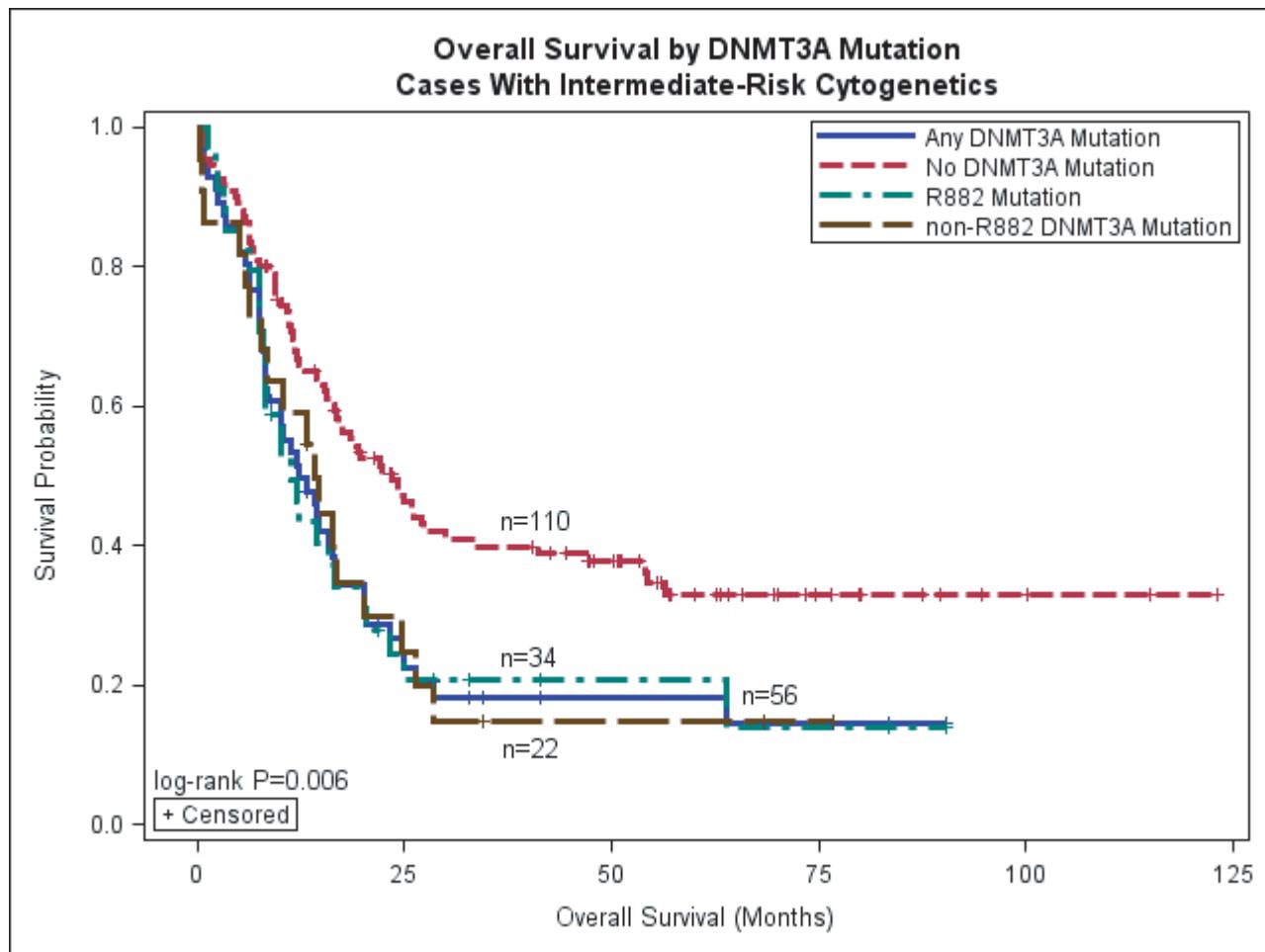


Supplementary Figure 10.

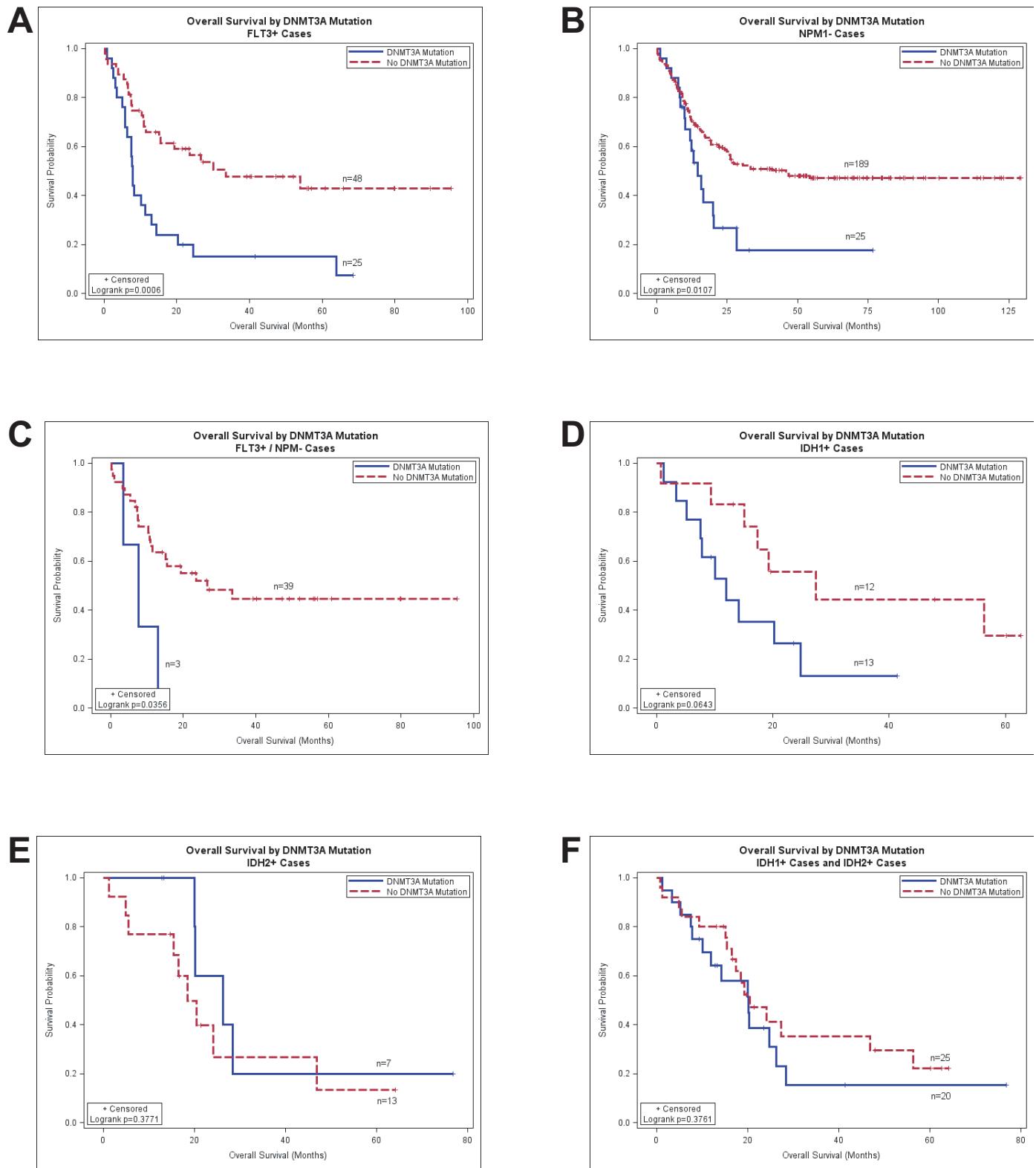
t(15;17)	t(8;21), inv(16)	Normal	Intermediate (not Normal)	Poor incl. complex	Not Done	
M0, M1	M2	M3	M4	M5	M6	M7
R882 mutation	non-R882 mutation	no DNMT3A mutation				



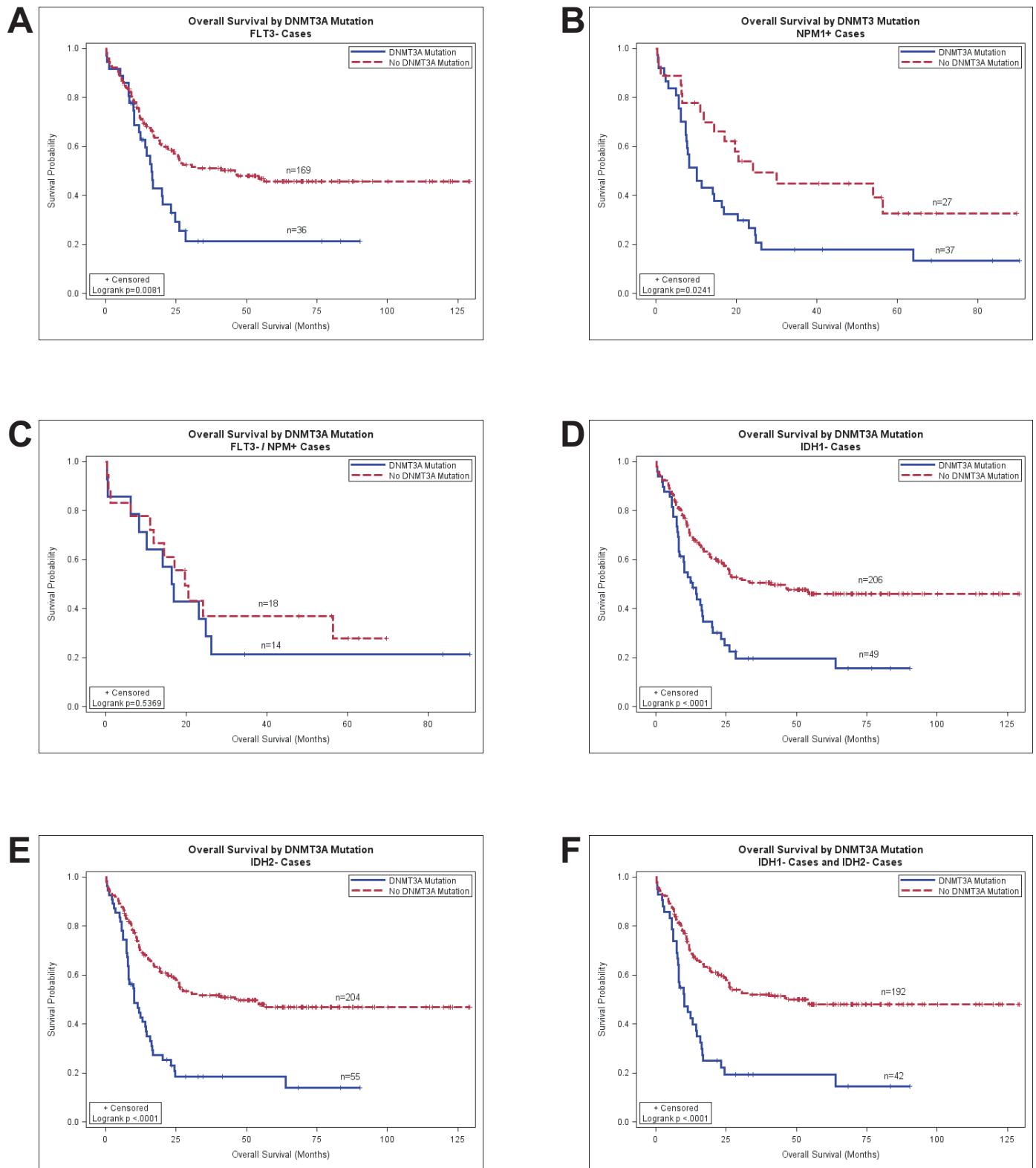
Supplementary Figure 11.



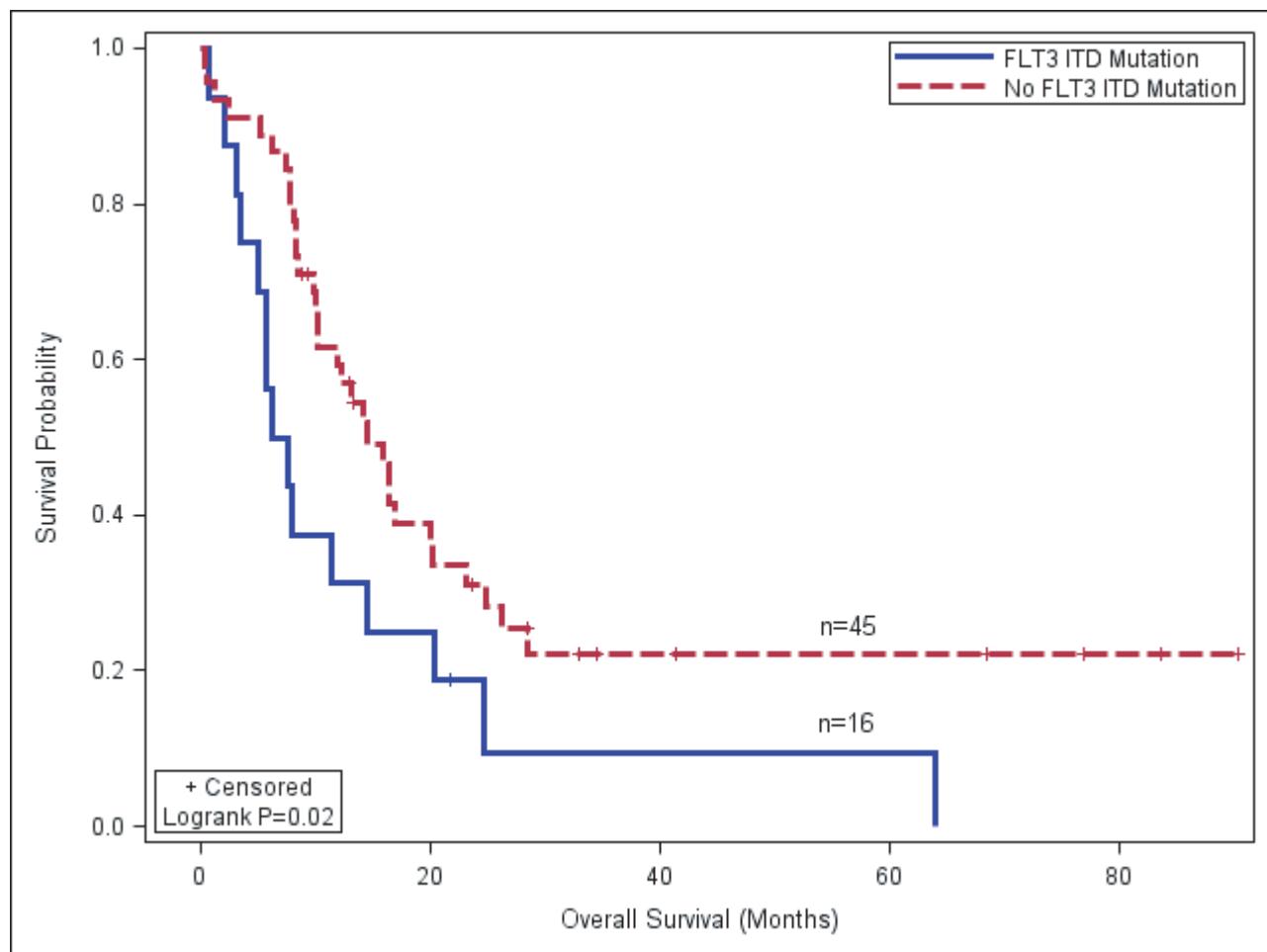
Supplementary Figure 12.



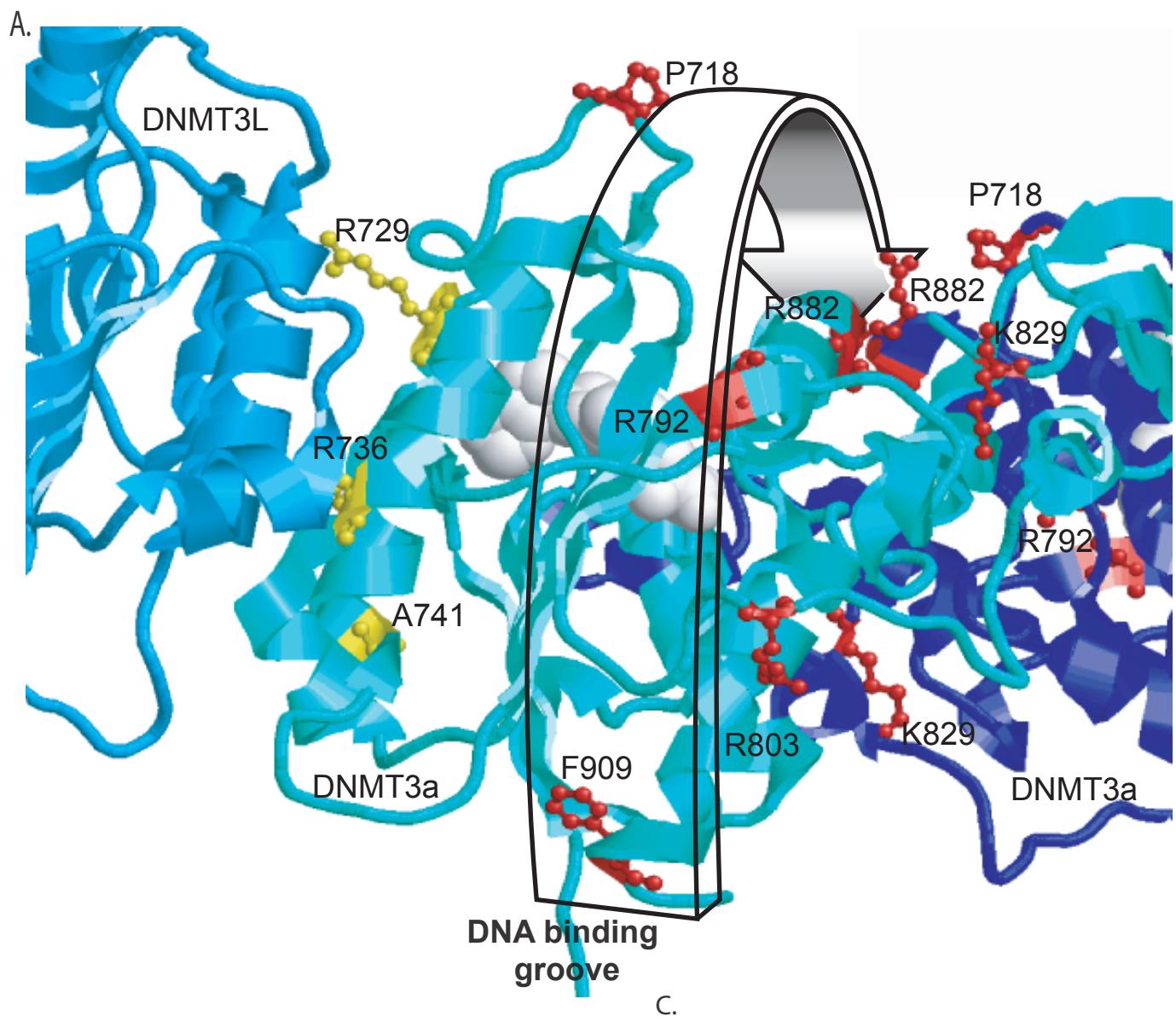
Supplementary Figure 13.



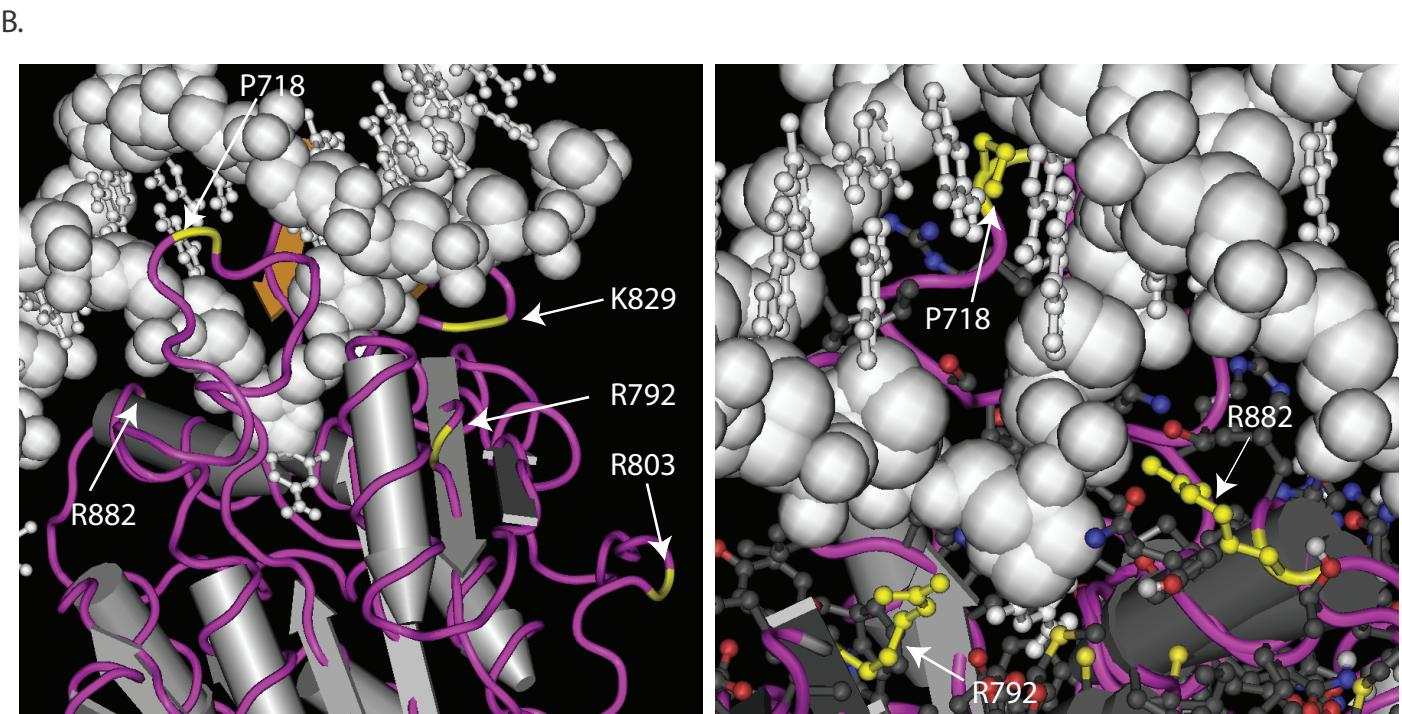
Supplementary Figure 14.



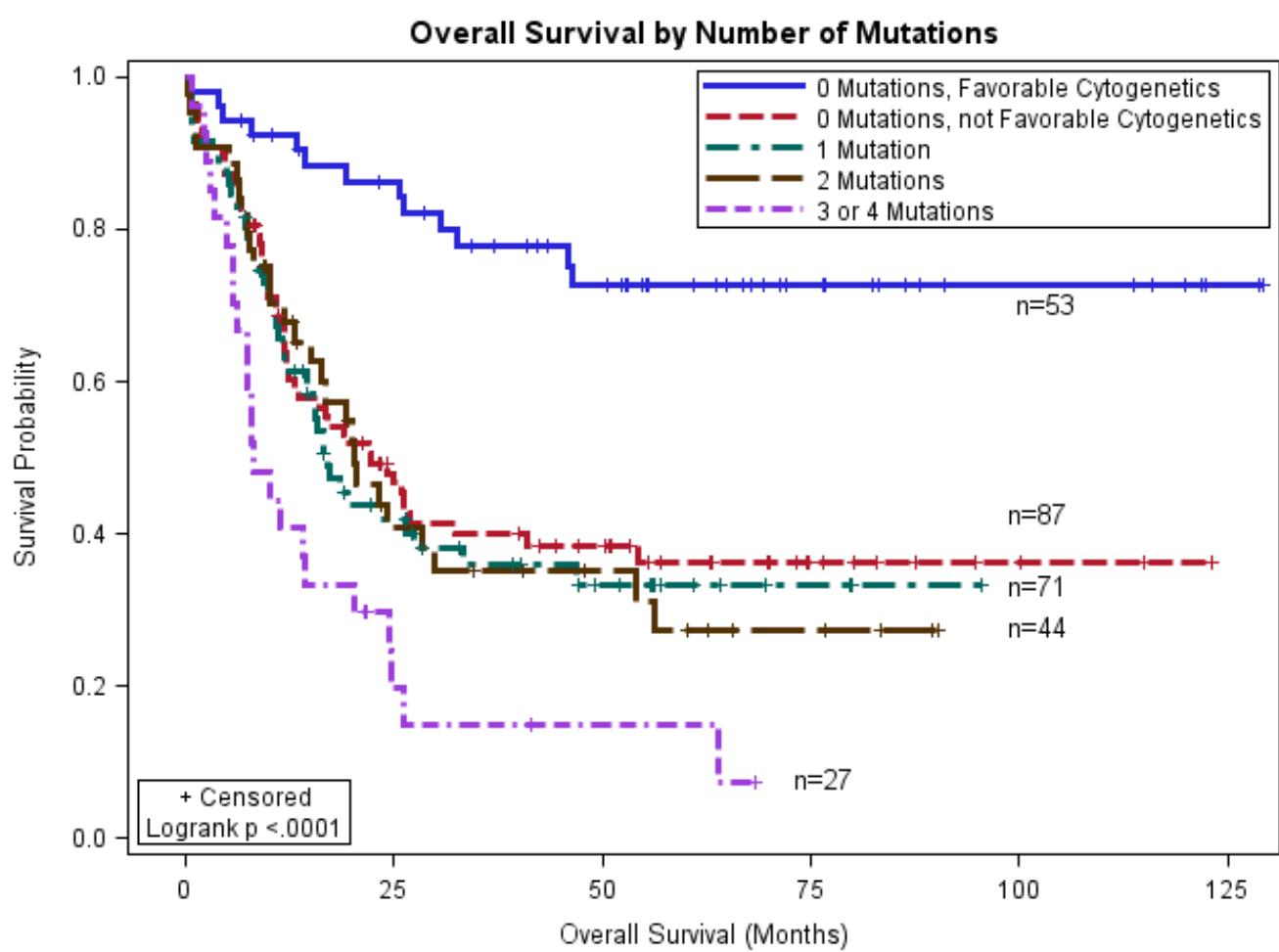
Supplementary Figure 15.



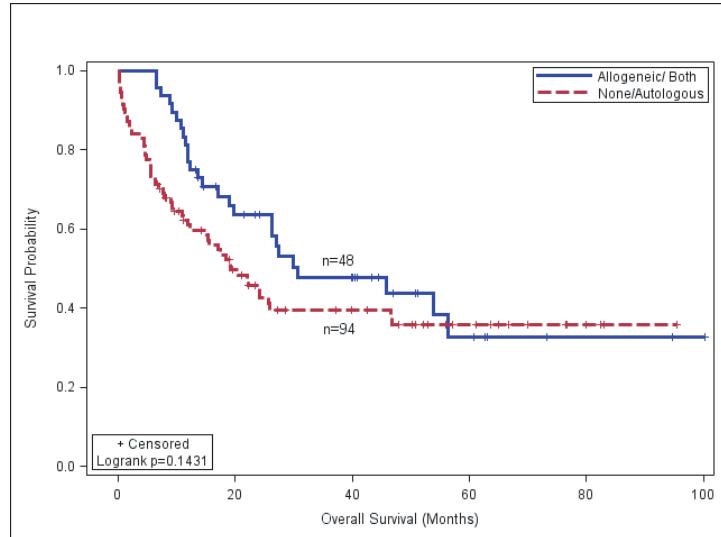
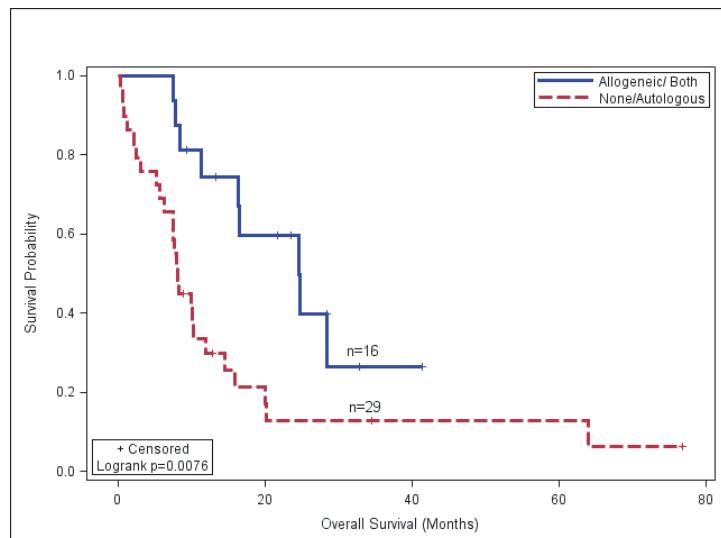
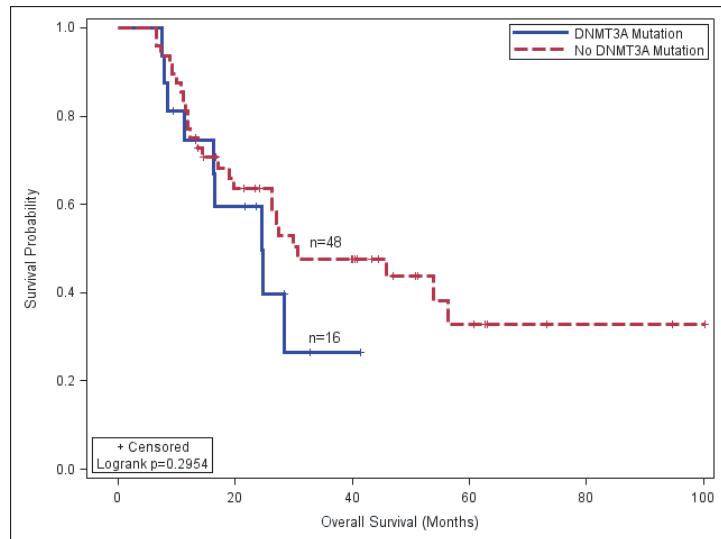
C.



Supplementary Figure 16.



Supplementary Figure 17.

A**B****C**

Supplementary Figure 18.

C-99-474	25	M	W	M2	normal		48	NA	NA	Dead	9.7	13.1	E30A ^a	missense		D324N		
C-99-483	28	F	W	M2	normal		53	NA	NA	Alive	68.4	68.4	A368D ^a	missense		D324N		W288fs
C-99-508	55	M	H	M2	(18;10;21)		60	NA	NA	Alive	69.5	69.5						
C-99-550	38	M	W	M4	inv(16)		45	NA	NA	Alive	71.2	71.2						
C-99-627	38	M	W	M2	(18;21)		60	NA	NA	Dead	5.1	13.5						
C-99-67	35	F	W	M1	normal		76	NA	NA	Dead	47.7	54.3						
C-99-709	59	F	W	M1	normal		75	NA	NA	Alive	56.9	56.9						
C-99-860	54	M	W	M4	normal		45	NA	NA	Dead	2	7.3		ITD				
C-99-901	38	F	B	M4	(116;16)		42	NA	NA	Alive	47.2	47.2				D835Y,V571I		
C-99-934	37	F	W	M1	normal		80	NA	NA	Alive	65.7	65.7				D835Y		W288fs

^anovel mutation, not
confirmed somatic
0=no data obtained

Supplementary Table 2. Summary of *DNMT3A* Mutations.

Mutation	Predicted Consequence	Evidence	n	PolyPhen [#]	SIFT ^{&}
R882H	missense	Prediction based on NM_022552.2	27	benign	deleterious
R882C	missense	Prediction based on NM_022552.2	7	probably damaging	deleterious
R882P	missense	Prediction based on NM_022552.2	2	probably damaging	deleterious
R882S	missense	Prediction based on NM_022552.2	1	probably damaging	deleterious
R803S	missense	Prediction based on NM_022552.2	1	probably damaging	no prediction
R792H	missense	Prediction based on NM_022552.2	1	benign	tolerated
R736H	missense	Prediction based on NM_022552.2	1	benign	no prediction
R729W	missense	Prediction based on NM_022552.2	1	probably damaging	tolerated
R729Q	missense	Prediction based on NM_022552.2	1	possibly damaging	tolerated
P904L [^]	missense	Prediction based on NM_022552.2	1	possibly damaging	deleterious
P718L	missense	Prediction based on NM_022552.2	1	probably damaging	deleterious
K841Q	missense	Prediction based on NM_022552.2	1	benign	no prediction
K829R	missense	Prediction based on NM_022552.2	1	possibly damaging	deleterious
K468R	missense	Prediction based on NM_022552.2	1	benign	tolerated
F909C	missense	Prediction based on NM_022552.2	1	probably damaging	deleterious
E30A [^]	missense	Prediction based on NM_022552.2	1	benign	deleterious
A741V	missense	Prediction based on NM_022552.2	1	benign	deleterious
A368D [^]	missense	Prediction based on NM_022552.2	1	probably damaging	deleterious
M315fs*	p.M315fs*	Prediction based on NM_022552.2	1	no prediction	no prediction
E616fs	p.E616fsNLTLQRFTTHSQLRRGSPSGCCLSLMESLQGSWC*	Prediction based on NM_022552.2	1	no prediction	no prediction
F827fs [^]	p.F827fsLAK*	Prediction based on NM_022552.2	1	no prediction	no prediction
L723fs	p.L723fsPTRALAGSSLSTASCMMRGPRREMIAPSSGSLRMWW/PVALVTRGTSRDFSSPTL*	Prediction based on NM_022552.2	1	no prediction	no prediction
E733fs [^]	p.E733fs*	Prediction based on NM_022552.2	1	no prediction	no prediction
G590fs	p.G590fsVPTGCCGGERTGPPGSRCSLLTTTRNLTQRFTHLSQLRRGSPSGCCLSLMESLQGSWC*	Prediction based on NM_022552.2	1	no prediction	no prediction
E477*	nonsense	Prediction based on NM_022552.2	1	no prediction	no prediction
R598 [^] +	nonsense	Prediction based on NM_022552.2	1	no prediction	no prediction
Q615*	nonsense	Prediction based on NM_022552.2	1	no prediction	no prediction
Q515*	nonsense	Prediction based on NM_022552.2	1	no prediction	no prediction
E505*	nonsense	Prediction based on NM_022552.2	1	no prediction	no prediction
L805 [^] +	nonsense	Prediction based on NM_022552.2	1	no prediction	no prediction
e21+2	mutation at donor splice site at 3' end of exon 21; causes p.R803fsSSAK*	RT-PCR indicates skipping of exon21 (fusion of exon 20 to 22)	1	no prediction	no prediction
e15+2	mutation at donor splice site at 3' end of exon 15; causes p.D618fsRVLGLGRGLELPWLSSLGCGA*	RT-PCR confirms intron 15 retention	1	no prediction	no prediction
L773-e18+5	deletion at L773 also deletes splice donor site; causes p.L773fsRQQPWFGQLTNGFYLGLLLHPVSSALWSWGLVTSALQGACLGSPRLPG*	Predicted based on intron retention	1	no prediction	no prediction
Deletion	del chr2:24,460,563-25,807,506, removes all or part of 9 genes	Prediction based on NM_022552.2	1	no prediction	no prediction

[^]=CALGB sample, not verified as somatic[&]=deleterious, SIFT score <= 0.05; tolerated, SIFT score > 0.05 (http://sift.jcvi.org/www/SIFT_help.html#SIFT_OUTPUT)[#]version 1.13

Supplementary Table 3. DNMT3A cDNA variant allele readcounts.

UPN	DNMT3A Mutation	WT Reads	Variant Reads	Total Reads	% Variant Reads
287	R882H	5668	5582	11250	49.62%
142074	R882H	1833	1951	3784	51.56%
399253	R882H	756	759	1515	50.10%
452198	R882H	827	861	1688	51.01%
104	R882C	1368	1048	2416	43.38%
186481	R882C	1040	597	1637	36.47%
335640	R882C	1247	1179	2426	48.60%
816067	R882C	762	725	1487	48.76%
224143	F909C	95	125	220	56.82%
509733	K841Q	977	2071	3048	67.95%
245450	R729W	2227	2141	4368	49.02%
208027	R803S	2231	885	3116	28.40%
433325	Q515*	140	36	176	20.45%
730817	E616fs	66	20	86	23.26%
530962	M315fs	106	0	106	0.00%
933124	L723fs	2862	1819	4681	38.86%
113971	G590fs^	603	6421	7024	91.42%
246634	E477*	8282	0	8282	0.00%
246634	A741V	86940	69	87009	0.08%
594368	D618fs/e15+2^^	445	12182	12627	96.48%
594368	R729Q	13610	14665	28275	51.87%
851929	R792H	32291	27959	60250	46.40%
851929	R803fs/e21+2^^^	30030	36605	66635	54.93%
997292	K468R	4585	714	5299	13.47%
997292	K829R	14458	17368	31826	54.57%

UPN, unique patient number; WT, wildtype

^Mutant allele frequency is 91% because one copy of *DNMT3A* was deleted in this sample

^^Primer pairs designed to detect aberrantly spliced transcript only (retained intron 15)

^^^^54.93% of transcripts reveal skipping of exon 21 (i.e. fusion of exon 20 to 22)

Supplementary Table 4. DNMT3A Mutation Types.

Reference Base	Variant Base	Total in 11 DNMT3A Mutant* genomes (%)	Total in 27 DNMT3A WT genomes (%)
A	C	773 (5.8)	1661 (6.9)
A	G	2403 (18.0)	4580 (19.1)
A	T	1703 (12.8)	3004 (12.5)
C	A	1856 (13.9)	3195 (13.3)
C	G	820 (6.2)	1488 (6.2)
C	T	5778 (43.3)	10074 (42.0)
	Total:	13,333	24,002

Mutations in each genome included: A741V/E477, E505*, F909C, L723fs, G590fs/1.5Mb deletion, R882H (n=4), and R882C (n=2).

Supplementary Table 5. LC-MS quantification of deoxyribonucleoside phosphates.¹

Sample Name	DNMT3A Mutation	Peak Area (counts)			Calculated amount of analyte (fmol)			Percentage of mdCMP (mdCMP/(mdCMP/mGMP))	
		mdCMP	dCMP	dGMP	mdCMP	dCMP	dGMP	(mdCMP/mGMP)	(mdCMP/(mdCMP+dCMP))
246634	A741V/E477*	4.92E+05	8.48E+06	1.04E+07	0.922	21.9	22.8	4.00%	4.04%
456892	E505*	2.13E+05	3.78E+06	4.65E+06	0.463	9.78	10.2	4.50%	4.52%
730817	E616fs	6.26E+05	1.00E+07	1.26E+07	1.28	25.9	27.4	4.70%	4.71%
224143	F909C	1.00E+06	1.47E+07	1.87E+07	2.01	37.9	40.7	4.90%	5.04%
113971	G590fs/deletion	5.37E+05	7.86E+06	1.02E+07	1.1	20.3	22.2	5.00%	5.14%
997292	K829R/K468R	6.71E+05	1.06E+07	1.29E+07	1.37	27.4	28.1	4.90%	4.76%
509733	K841Q	6.46E+05	8.98E+06	1.17E+07	1.32	23.2	25.5	5.20%	5.38%
933124	L723fs	3.42E+05	5.69E+06	7.23E+06	0.717	14.7	15.8	4.50%	4.65%
369065	L773-e18+5	6.92E+05	1.11E+07	1.43E+07	1.41	28.8	31.1	4.50%	4.67%
530962	M315fs	5.16E+05	9.36E+06	1.16E+07	1.06	24.2	25.4	4.20%	4.20%
498463	P718L	4.22E+05	6.81E+06	8.58E+06	0.875	17.6	18.8	4.70%	4.74%
433325	Q515*	6.81E+05	1.04E+07	1.33E+07	1.39	26.9	29	4.80%	4.91%
957664	Q615*	3.76E+05	7.63E+06	9.32E+06	0.784	19.7	20.4	3.80%	3.83%
594368	R729Q/e15+2	4.45E+05	6.74E+06	8.49E+06	0.92	17.4	18.6	4.90%	5.02%
245450	R729W	4.46E+05	7.16E+06	9.12E+06	0.662	18.5	19.9	3.30%	3.45%
294154	R736H	2.34E+05	4.32E+06	5.18E+06	0.81	19.9	18.8	4.30%	3.91%
851929	R792H/e21+2	1.30E+06	1.94E+07	2.57E+07	2.6	50.1	55.9	4.70%	4.93%
208027	R803S	2.14E+06	3.02E+07	3.97E+07	4.26	77.9	86.2	4.90%	5.19%
186481	R882C	5.68E+05	9.15E+06	1.15E+07	1.16	23.7	25.2	4.60%	4.67%
142074	R882H	1.70E+06	2.72E+07	2.59E+07	3.95	74.9	81.3	4.90%	5.01%
375182	R882H	5.01E+06	7.16E+07	8.81E+07	11.6	197	277	4.20%	5.56%
431799	R882H	9.37E+05	1.57E+07	1.58E+07	2.19	43.2	49.8	4.40%	4.82%
452198	R882H	1.00E+06	1.73E+07	1.72E+07	2.34	47.6	54	4.30%	4.69%
721214	R882H	1.12E+06	1.88E+07	1.67E+07	2.61	51.9	52.5	5.00%	4.79%
737451	R882H	1.33E+06	2.12E+07	2.15E+07	3.09	58.4	67.5	4.60%	5.03%
740266	R882H	4.66E+06	6.51E+07	7.38E+07	10.8	179	232	4.70%	5.69%
817156	R882H	4.04E+06	5.83E+07	6.94E+07	9.34	161	218	4.30%	5.48%
869922	R882H	3.50E+05	5.98E+06	5.55E+06	0.839	16.4	17.5	4.80%	4.87%
399253	R882H	5.30E+05	8.22E+06	1.06E+07	1.09	21.3	23.2	4.70%	4.87%
807615	R882H	6.40E+05	1.03E+07	1.35E+07	1.3	26.7	29.4	4.40%	4.64%
103342	Wild Type	1.67E+06	2.63E+07	2.81E+07	3.89	72.5	88.3	4.40%	5.09%
179223	Wild Type	3.11E+06	4.46E+07	5.38E+07	7.21	123	169	4.30%	5.54%
329614	Wild Type	7.34E+05	1.20E+07	1.03E+07	1.72	33	32.4	5.30%	4.95%
804168	Wild Type	7.13E+06	8.93E+07	1.08E+08	16.5	246	339	4.90%	6.29%
804168	Wild Type	1.76E+06	2.69E+07	2.78E+07	4.1	74	87.3	4.70%	5.25%
906708	Wild Type	6.93E+06	8.66E+07	1.07E+08	16	238	337	4.70%	6.30%
907786	Wild Type	6.84E+05	1.18E+07	1.14E+07	1.61	32.6	35.9	4.50%	4.71%
982009	Wild Type	3.00E+06	4.61E+07	5.22E+07	6.95	127	164	4.20%	5.19%
380949	Wild Type	5.61E+05	9.60E+06	1.24E+07	1.15	24.8	27	4.30%	4.43%
418499	Wild Type	6.44E+05	1.00E+07	1.36E+07	1.31	25.9	29.7	4.40%	4.81%

445045	Wild Type	4.00E+05	6.53E+06	8.33E+06	0.832	16.9	18.2	4.60%	4.69%
545259	Wild Type	1.05E+06	1.68E+07	2.24E+07	2.12	43.4	48.7	4.40%	4.66%
868231	Wild Type	6.50E+05	9.71E+06	1.29E+07	1.32	25.1	28.1	4.70%	5.00%
311636	Wild Type	1.16E+06	1.99E+07	2.49E+07	3.92	91.5	90.2	4.30%	4.11%
103342	Wild Type	7.03E+05	1.13E+07	1.42E+07	1.43	29.2	31	4.60%	4.67%
179223	Wild Type	8.90E+05	1.48E+07	1.91E+07	1.8	38.2	41.5	4.30%	4.50%
329614	Wild Type	6.40E+05	9.64E+06	1.28E+07	1.3	24.9	27.8	4.70%	4.96%
573988	Wild Type	9.81E+05	1.54E+07	1.99E+07	1.97	39.8	43.3	4.50%	4.72%
804168	Wild Type	6.17E+05	1.02E+07	1.25E+07	1.26	26.3	27.3	4.60%	4.57%
804168	Wild Type	7.97E+05	1.28E+07	1.60E+07	1.61	33.1	34.9	4.60%	4.64%
906708	Wild Type	7.64E+05	1.19E+07	1.52E+07	1.55	30.8	33	4.70%	4.79%
907786	Wild Type	1.02E+06	1.83E+07	2.32E+07	3.46	84.2	84	4.10%	3.95%
982009	Wild Type	5.45E+05	9.11E+06	1.10E+07	1.12	23.5	24.1	4.60%	4.55%
CONTROL 1 ²	Control	1.08E+06	1.79E+07	2.45E+07	2.17	46.1	53.3	4.10%	4.50%
CONTROL 2 ²	Control	1.16E+06	1.92E+07	2.44E+07	2.33	49.7	53.1	4.40%	4.48%
CONTROL 3 ²	Control	1.54E+06	2.47E+07	3.46E+07	3.07	63.9	75.2	4.10%	4.58%
CONTROL 4 ²	Control	8.30E+05	1.47E+07	1.84E+07	2.82	67.6	66.7	4.20%	4.00%

¹Determined from 3.5-5 ug DNA with 23% of total hydrolysate analyzed.

²The control samples were derived from a single pool of genomic DNA derived from the spleens of 6 young C57Bl/6 male mice. The same control sample was run in the assay 4 independent times.

Nearby Gene: ENSG00000216862														
chr1	226706706	226706755	0.99	0.69	0.36	0.52	3	3.32	2.4	0.27	3.39	2.51	0.0014966832	0.0489592741
chr1	226706776	226706825	1.04	0.54	0.4	0.45	3.06	3.17	2.38	0.38	3.29	2.5	0.0001404431	0.0287990422
chr1	226706886	226706935	1.2	0.38	0.42	0.45	3.57	3.38	2.51	-0.07	3.06	2.26	0.0024254359	0.0564372399
Nearby Gene: GJC2														
chr1	226406704	226406753	2.25	1.34	1.92	1.89	4.11	4.51	3.67	1.51	4.4	4.11	0.0059194867	0.0760236858
chr1	226406829	226406878	2.12	1.25	1.72	1.41	3.36	3.57	3.06	1.32	3.58	3.61	0.0066112750	0.0791939228
chr1	226406919	226406968	2.65	1.55	2.21	2.17	4.44	4.59	3.96	1.28	4.08	4.13	0.0114571396	0.0984634730
Nearby Gene: PM20D1														
chr1	204085603	204085652	-0.88	0.38	-0.7	1.52	0.88	1.32	1.35	1.19	1.37	1.94	0.0308053580	0.1508980136
chr1	204085683	204085732	-2.06	0.28	-2.12	1.38	0.41	1.31	1.37	1.04	1.87	1.78	0.0052841880	0.0730874803
chr1	204085883	204085932	-1.49	0.49	-1.48	1.49	0.71	1.03	1.38	1.09	1.79	1.95	0.0332018060	0.1562079455
Nearby Gene: PEAR1														
chr1	155150366	155150415	1.27	1.07	1.07	1.47	2.23	3.07	2.28	1.4	3.21	2.77	0.0062223244	0.0775930908
chr1	155150471	155150520	1.56	0.88	1.2	1.27	2.85	3.28	2.35	1.22	3.38	2.72	0.000903933	0.0264076198
chr1	155150571	155150620	1.28	0.86	1.12	1.05	2.51	2.92	2.23	1.12	3.1	2.45	0.0038358647	0.0653266507
Nearby Gene: SMG5														
chr1	154486638	154486687	1.25	1.18	1.13	-0.28	1.44	2.01	1.87	1.48	2.59	1.94	0.0117956978	0.0996572570
chr1	154486763	154486812	1.31	1.18	1.21	-0.12	2.26	2.31	2.24	1.62	3.01	2.31	0.0035746149	0.0638631274
chr1	154486863	154486912	0.72	0.7	0.7	-0.55	1.79	2.38	1.86	1.34	2.45	2.14	0.0005105192	0.0367458364
Nearby Gene: LOC100131974														
chr1	147599626	147599675	0.3	1.06	1.3	1.48	0.84	1.34	2.48	1.41	2.83	2.33	0.0078575878	0.0845543707
chr1	147599691	147599740	0.12	0.63	0.92	1.2	0.99	1.69	2.52	0.99	3.56	2.73	0.0013409318	0.0474718099
chr1	147599786	147599835	0.44	-0.17	0.14	0.09	0.85	1.33	2.33	-0.23	2.57	1.87	0.0023009935	0.0555334433
Nearby Gene: WNT2B														
chr1	112853677	112853726	1.42	0.03	0.62	0.12	2.36	3.8	3.65	1.09	4.08	2.41	0.0376206105	0.1655391443
chr1	112853782	112853831	1.28	0.07	0.61	0.28	1.38	2.83	2.7	1.35	4.04	2.06	0.0059510566	0.0761878250
chr1	112854002	112854051	0.85	0.13	0.32	0.25	0.72	2.26	2.24	1.49	3.15	1.69	0.0287674793	0.1464027898

Supplementary Table 7. Inherited DNMT3A coding variants in AML cases vs. controls.

Genome coordinate	rsID	Variant	Codon	AML patients*			CEU [#]			P-value [^]			
				n	AA	AB	BB	n	AA	AB	BB	Allele	Genotype
chr2:25,323,006	rs2276598	C>T	L422	166	120	40	6	60	43	15	2	0.884	0.987
chr2:25,390,331	rs41284843	G>A	P9	168	122	43	3	60	46	12	2	0.879	0.561
chr2:25,324,464	novel	G>A	S267	167	165	2	0	--	--	--	--	NA	NA
chr2:25,324,473	novel	G>A	P264	168	166	2	0	--	--	--	--	NA	NA
chr2:25,324,506 [@]	novel	G>A	P253	168	168	0	0	--	--	--	--	NA	NA
chr2:25,390,336	novel	C>T	G8S	168	167	1	0	--	--	--	--	NA	NA

*Caucasian only, [#]from 1,000 Genomes (build 130), [@]polymorphic in African American cases only, [^]by Fisher's exact test (allele) or chi-square test (genotype).

Supplementary Table 9. Primers used for DNMT3L sequencing.

Amplicon Name	L Primer Coord	R Primer Coord	L Amp Coord	R Amp Coord	Primer 1 Sequence	Primer 2 Sequence	Target Name	+	Target Start	Target Stop	Build	Hugo Name	TranscriptID	Exon#	ExonID	Exon Start	Exon Stop	Entrez ID	Chrom
0229947_04B	44505296	44505643	44505316	44505621	TGTAAAACGACGGCCAGTGCATCAGGATCTGAGCAGG	CAGGGAAACAGCTATGACCTGTGTTAACGCCCATACC	29947-10		44505465	44505570	36	DNMT3L	ENST000002070172	2	ENSE00000952757	44505465	44505577	29947	21
0229947_04D	44505294	44505276	44505298	44505275	TGTAAAACGACGGCCAGTGCATCAGGATCTGAGCAGG	CAGGGAAACAGCTATGACCTGTGTTAACGCCCATACC	29947-9		44505195	44505150	36	DNMT3L	ENST000002070172	3	ENSE00000952758	44505106	44505150	29947	21
0229947_04E	44505074	44505096	44505075	44505095	TGTAAAACGACGGCCAGTGCATCAGGATCTGAGCAGG	CAGGGAAACAGCTATGACCTGTGTTAACGCCCATACC	29947-9		44505106	44505150	36	DNMT3L	ENST000002070172	3	ENSE00000952758	44505106	44505150	29947	21
0229947_04C	44505074	44505296	44505096	44505276	TGTAAAACGACGGCCAGTGCATCAGGATCTGAGCAGG	CAGGGAAACAGCTATGACCTGTGTTAACGCCCATACC	29947-9		44505106	44505150	36	DNMT3L	ENST000002070172	3	ENSE00000952758	44505106	44505150	29947	21
0229947_059	44503802	44504095	44503829	44504073	TGTAAAACGACGGCCAGTGCATCAGGATAGGCGAGGGAGC	CAAGGGAAACAGCTATGACCTCTGTGCGGAGATGATGC	29947-8		44503943	44504022	36	DNMT3L	ENST000002070172	4	ENSE00000952759	44503943	44504022	29947	21
0229947_057	44503740	44503947	44503762	44503923	TGTAAAACGACGGCCAGTGCATCAGGATAGGCGAGGGAGC	CAAGGGAAACAGCTATGACCTCTGTGCGGAGATGATGC	29947-7		44503748	44503860	36	DNMT3L	ENST000002070172	5	ENSE00000952760	44503748	44503860	29947	21
0229947_059	44503802	44504095	44503829	44504073	TGTAAAACGACGGCCAGTGCATCAGGATAGGCGAGGGAGC	CAAGGGAAACAGCTATGACCCAGGGAAATCATCCCGTG	29947-7		44503748	44503860	36	DNMT3L	ENST000002070172	5	ENSE00000952760	44503748	44503860	29947	21
0229947_038	44503637	44503884	44503657	44503860	TGTAAAACGACGGCCAGTGCATCAGGATAGGCGAGGGAGC	CAAGGGAAACAGCTATGACCCAGGGAAATCATCCCGTG	29947-7		44503748	44503860	36	DNMT3L	ENST000002070172	5	ENSE00000952760	44503748	44503860	29947	21
0229947_037	44503733	44503884	44503903	44503914	TGTAAAACGACGGCCAGTGCATCAGGATAGGCGAGGGAGC	CAAGGGAAACAGCTATGACCCAGGGAAATCATCCCGTG	29947-6		44503748	44503860	36	DNMT3L	ENST000002070172	5	ENSE00000952760	44503748	44503860	29947	21
0229947_030	44502771	44503011	44502793	44502991	TGTAAAACGACGGCCAGTGCATCAGGATAGGCGAGTCAGG	CAAGGGAAACAGCTATGACCCAGGGAAATCATCCCGTG	29947-6		44502834	44503005	36	DNMT3L	ENST000002070172	6	ENSE00001236963	44502834	44503005	29947	21
0229947_04M	44500254	44500524	44500277	44500505	TGTAAAACGACGGCCAGTAACTCAAGGGCTCTGTAGCCAAG	CAAGGGAAACAGCTATGACCCGACATCAGCAGTTGGCGC	29947-5		44500378	44500465	36	DNMT3L	ENST000002070172	7	ENSE000012397260	44500378	44500465	29947	21
0229947_04S	44498752	44498916	44498774	44499094	TGTAAAACGACGGCCAGTGTAGAGGGTCTCTGTAGTCAGC	CAAGGGAAACAGCTATGACCCGACTCTGTAGCAGCCACAGG	29947-4		44498928	44499016	36	DNMT3L	ENST000002070172	8	ENSE00001311957	44498928	44499016	29947	21
0229947_05d	H_21_000000059	44498935	44498152	44498857	TGTAAAACGACGGCCAGTGCCTGCCTCACTCTCAAGC	CAAGGGAAACAGCTATGACCTGACAATTTCAGTTCTGCC	29947-3		44499534	44499609	36	DNMT3L	ENST000002070172	9	ENSE000012391091	44499534	44499609	29947	21
H_21_000000059	44498885	44498978	44498907	44498987	TGTAAAACGACGGCCAGTCGGCTGTGAGCCCTCTGAAGC	CAAGGGAAACAGCTATGACCCCTCAGCCCTTCAGTTCAGCC	29947-2		44498912	44498920	36	DNMT3L	ENST000002070172	10	ENSE00000952765	44498912	44498920	29947	21
0229947_03P	44493269	44493474	44493298	44493495	TGTAAAACGACGGCCAGTGGCGCTTAACCGTCAGGT	CAAGGGAAACAGCTATGACCCCTCTGTGGCTCATGTT	29947-1		44493324	44493422	36	DNMT3L	ENST000002070172	11	ENSE00000952766	44493324	44493422	29947	21
0229947_03T	4449093	44490953	44490950	44490971	TGTAAAACGACGGCCAGTAACTGAAGGCCAGGCTGTC	CAAGGGAAACAGCTATGACCCCTCTGTGGCTCATGTT	29947-0		44490708	44490876	36	DNMT3L	ENST000002070172	12	ENSE00000952767	44490651	44490956	29947	21

Annotation based on Ensembl release 54_36p.