Supporting Information

Base Editor Scanning Reveals Activating Mutations of DNMT3A

Emma M. Garcia[1,2], Nicholas Z. Lue[1,2], Jessica K. Liang[1], Whitney K. Lieberman[1], Derek D. Hwang[1], James Woods[1,2], Brian B. Liau*[1,2]

Affiliations:
[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA
[2]Broad Institute of Harvard and MIT, Cambridge, MA, USA

Table of Contents

Supplementary Figures:

Supplementary Tables:

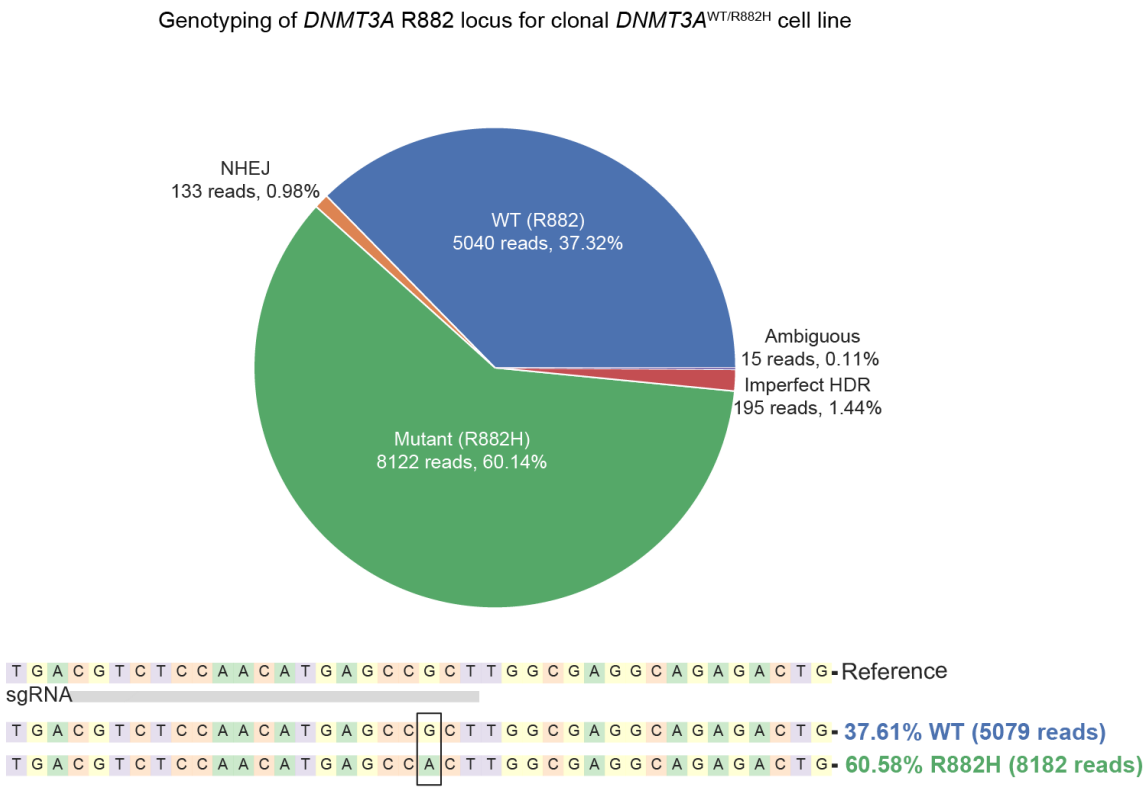


Figure S1: Pie chart and allele table showing targeted deep sequencing of R882 codon of *DNMT3A* in *DNMT3A*$^{WT/R882H}$ cells. K562 cells are triploid, and these genotyping results are consistent with our cell line containing 2 copies of *DNMT3A*^{R882H} and one copy of *DNMT3A*^{WT}. sequences of WT and R882H reads are shown below. Plots generated with CRISPResso2.

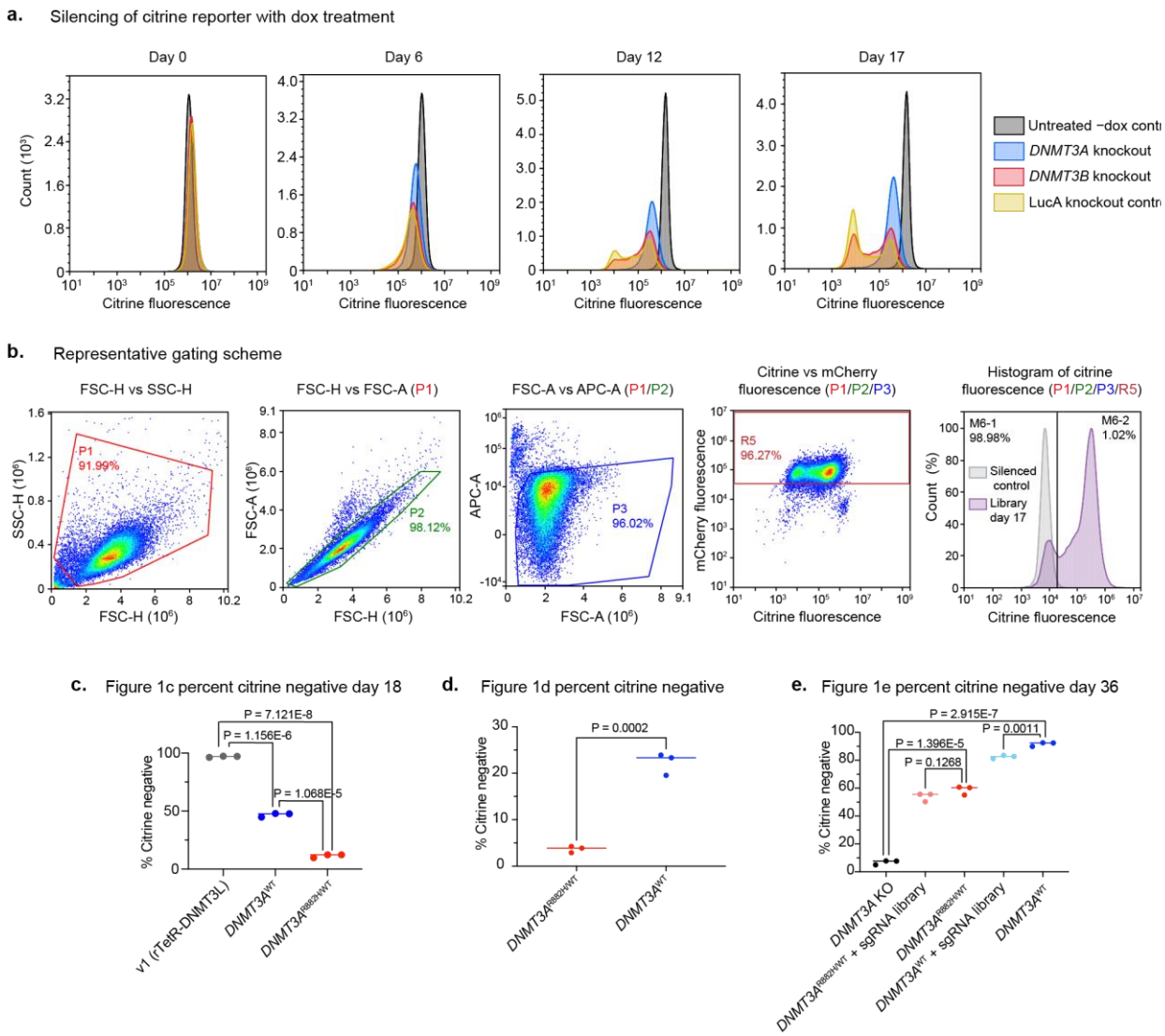


Figure S2: Silencing dynamics of dDNMT3A reporter system, representative gating scheme, and additional plots associated with Figure 1.

a.) Histograms of *DNMT3A*^{WT} reporter cells treated with *DNMT3A* knockout, *DNMT3B* knockout, or LucA knockout (control) sgRNAs and Cas9 after 0, 6, 12, and 17 days of 1 ug/mL doxycycline treatment.

b.) Representative gating scheme for sorting cells in our screen. Events were first gated for morphologically typical K562 cells, then for events corresponding to single cells. Then, helix-IR staining was used to gate out dead cells. Finally, the population corresponding to healthy, single cells was gated for mCherry+ cells and assessed for citrine fluorescence. Cells were sorted based on a citrine gate where ~1% of fully silenced control cells are citrine+.

c.) Plot of day 18 percent citrine negative cells from Figure 1c. P-values were calculated from two-sided student's t-tests.

d.) Plot of percent citrine negative cells from Figure 1d. P-value was calculated from a two-sided student's t-test.

e.) Plot of day 36 percent citrine negative cells from Figure 1e. P-values were calculated from two-sided student's t-tests.

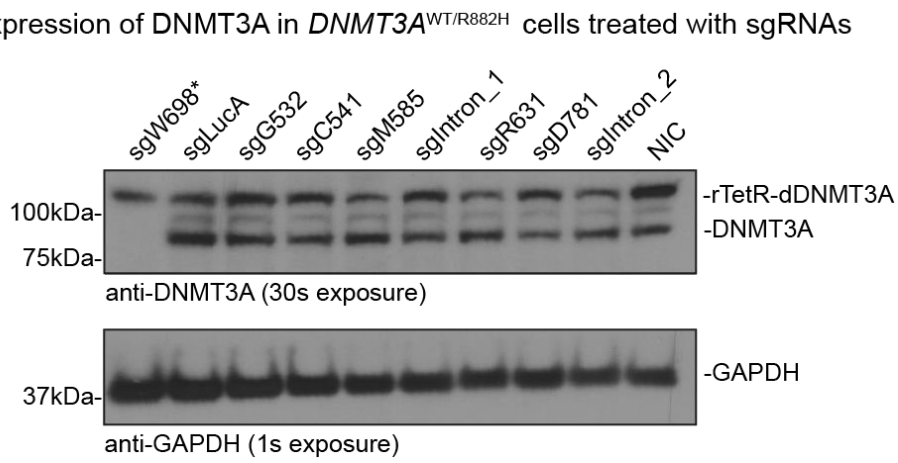


Figure S3: Western Blot of reporter cells transduced with base editing sgRNAs. sgW698* causes depletion of endogenous DNMT3A levels but not rTetR-dDNMT3A levels. Other sgRNAs tested (the same set as used in Figure 3b) do not dramatically affect protein expression levels.

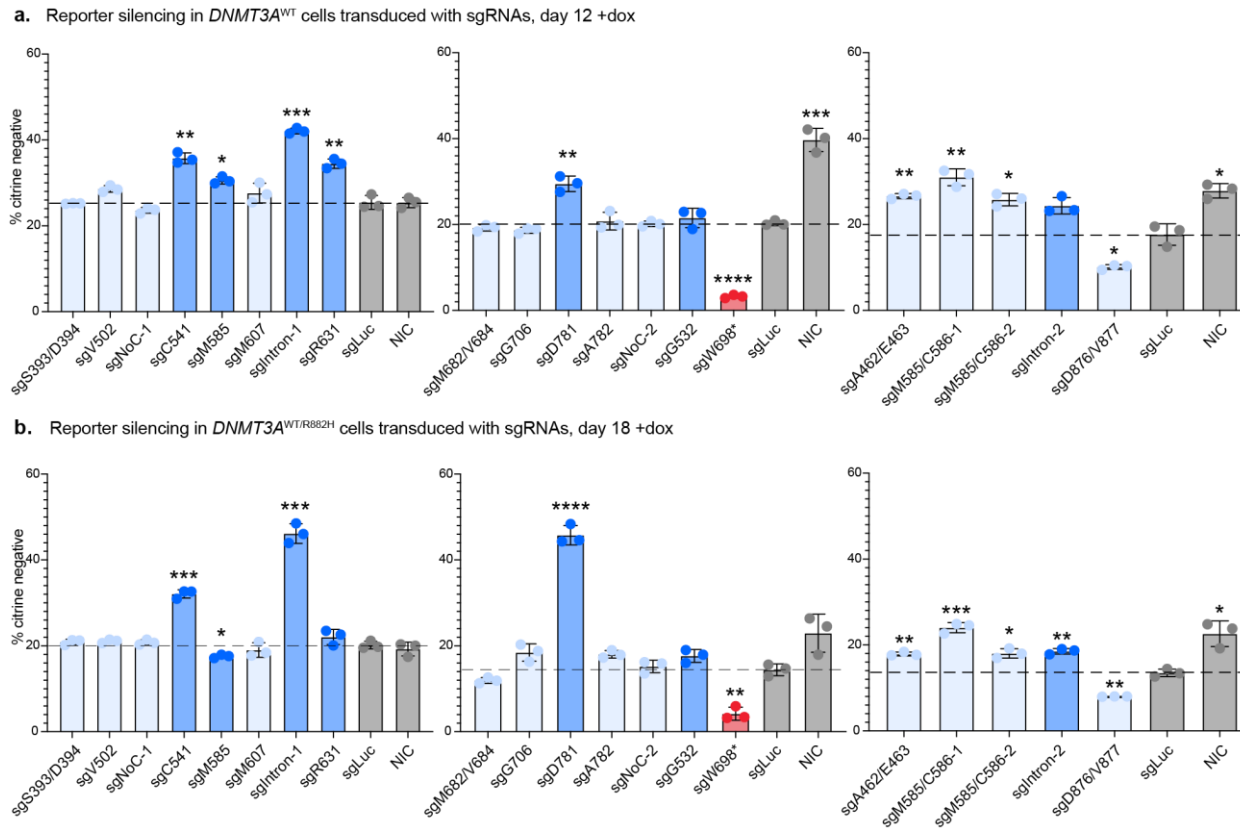


Figure S4: Additional sgRNAs tested.

a. Bar plots showing % silenced *DNMT3A*^{WT} cells treated with sgRNAs after 12 days of dox treatment. LucA = Luciferase targeting sgRNA, NIC = Non-infection control. Dark blue, red and gray bars indicate biological replicates for guides shown in Figure 3b, and light blue bars indicate one biological replicate in technical triplicate for additional sgRNAs. Data are mean ± SD of n = 3 replicates. Stars correspond to significance testing after two-sided students t-tests comparing each guide to sgLuc with Bonferroni multiple hypothesis correction. See Table S6 for a summary of all single guides tested.

b. Bar plots showing % silenced *DNMT3A*$^{WT/R882H}$ cells after 18 days of dox treatment. Data are mean ± SD of n = 3 replicates. Stars correspond to significance testing after two-sided Students t-tests comparing each guide to sgLuc with Bonferroni multiple hypothesis correction. See table S6 for a summary of all single guides tested.

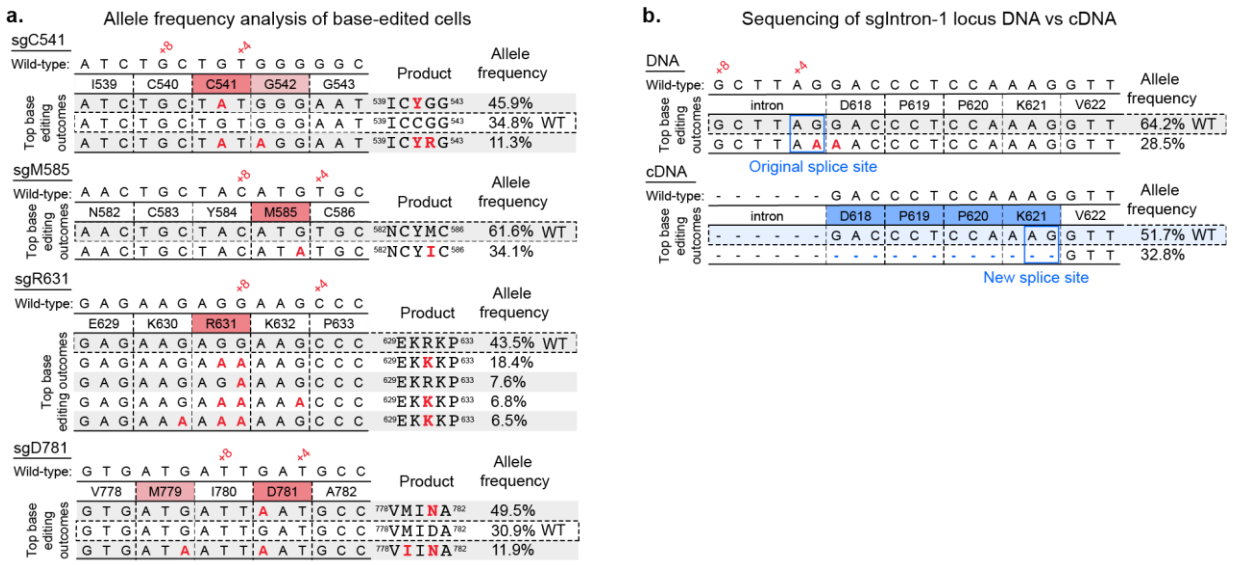
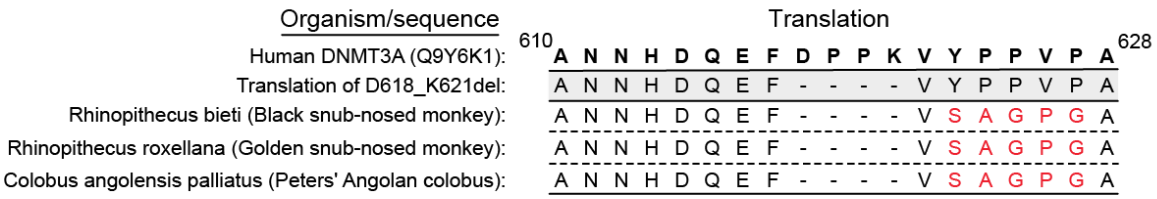


Figure S5: Genotyping of *DNMT3A*$^{WT/R882H}$ cells treated with sgRNAs.

a. Tables showing base editing outcomes with >5% allele frequency from treatment of *DNMT3A*$^{WT/R882H}$ reporter cells with single missense sgRNAs.

b. Scheme showing genomic DNA and transcript cDNA sequencing of cells treated with sgIntron-1.

Alignment of human and colobinae monkey DNMT3A



Overall sequence identity

	Human	**Black snub-nosed**	**Golden snub-nosed**	**Peters' Angolan**
Human	100%	96.8%	96.8%	95.3%
Black snub-nosed	96.8%	100%	100%	98.1%
Golden snub-nosed	96.8%	100%	100%	98.1%
Peters' Angolan	95.3%	98.1%	98.1%	100%

Figure S6: D618_K621del mutation is found in colobinae monkeys. Despite additional local variation in sequence between human and colobinae DNMT3A, the proteins are overall very similar. UniProt IDs: A0A2K6PNQ5_RHIRO, A0A2K6KW24_RHIBE, A0A2K5JG23_COLAP

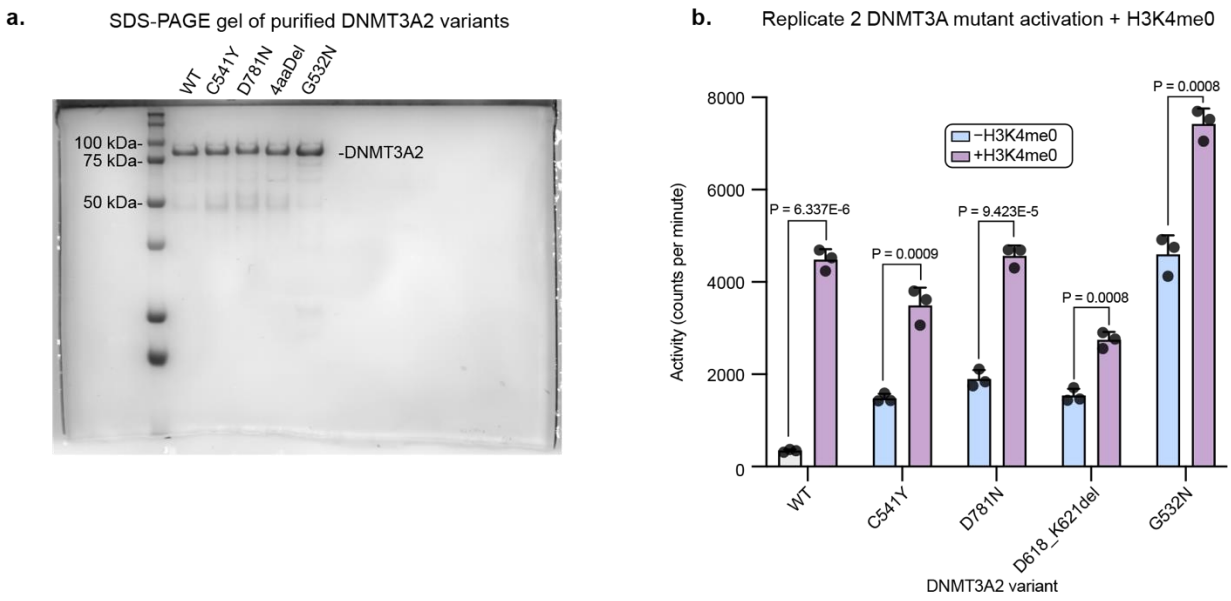


Figure S7: Additional *in vitro* validation of DNMT3A variants.

a. SDS-PAGE gel showing purified DNMT3A2 variants used in Figure 4.

b. Second replicate of activity assay shown in Figure 4b. Data are mean ± SD of n = 3 replicates. P-values were calculated by unpaired two-tailed Students' T-tests. P-values for no peptide mutant comparisons are as follows: WT vs C541Y = 3.877E-5, WT vs D781N = 0.0001, WT vs D618_K621del = 0.0002, WT vs G532N = 6.012E-5.

Table S1: Plasmids constructed in this study.

Application	Plasmid	Description
Base editing screen and validation	pEF-H2B-mCherry-T2A-rTetR-dDNMT3A2	Lentiviral, mammalian expression of updated reporter
Bacterial Protein Expression	pET28b-His_6-DNMT3A2-C541Y	Expression of His_6-tagged DNMT3A2 C541Y
	pET28b-His_6-DNMT3A2-D781N	Expression of His_6-tagged DNMT3A2 D781N
	pET28b-His_6-DNMT3A2-4aaDel	Expression of His_6-tagged DNMT3A2 4aaDel

Table S2: Primers and oligos used in this study.

Application	Description	Sequence
Plasmid Construction	Construction of pET28b-His$_6$-DNMT3A2-D781N Forward	CTTCCTTCGCATTGATCATAACCG
	Construction of pET28b-His$_6$-DNMT3A2-D781N Reverse	CGGTTATGATCAATGCGAAGGAAG
	Construction of pET28b-His$_6$-DNMT3A2-4aaDel Forward	GGTGGATACACAAATTCTTGGTCATGG
	Construction of pET28b-His$_6$-DNMT3A2-4aaDel Reverse	GGTGGATACACAAATTCTTGGTCATGG
	Construction of pET28b-His$_6$-DNMT3A2-C541Y Forward	CCTCACGACCGCCATAGCAAATCGTAC
	Construction of pET28b-His$_6$-DNMT3A2-C541Y Reverse	GTACGATTTGCTATGGCGGTCGTGAGG
	Construction of pEF-H2B-mCherry-T2A-rTetR-dDNMT3A2 Forward	CCTTTTTTTTGGCTCTTTGCGAACGTCGTCGCTATGGGAG
	Construction of pEF-H2B-mCherry-T2A-rTetR-dDNMT3A2 Reverse	CTCCCATAGCGACGACGTTCGCAAAGAGCCAAAAAAAAGG
Individual sgRNAs (top= Forward, bottom= Reverse)	DNMT3A2_552	CACCGGGCATCAATCATCACAGGGTAAACACCCTGTGATGATTGATGCCC
	DNMT3A2_414	CACCGTCCTAAGCAGTGAGCACAACAAACGTTGTGCTCACTGCTTAGGAC
	DNMT3A2_207	CACCGTGTCACTCTCATCGCTGTCGAAACCGACAGCGATGAGAGTGACAC
	DNMT3A2_341	CACCGAGCAGATGGTGCAGTAGGACAAACGTCCTACTGCACCATCTGCTC
	DNMT3A2_617	CACCGCATCTTATGGTGCACTGAAAAAACTTTCAGTGCACCATAAGATGC
	DNMT3A2_310	CACCGGGTAACATTGAGGCTCCCACAAACGTGGGAGCCTCAATGTTACCC
	DNMT3A2_396	CACCGAAGAACATCTGGAGCCGGGAAAACTCCCGGCTCCAGATGTTCTTC
	DNMT3A2_493	CACCGCCCCAATCACCAGATCGAATAAACATTCGATCTGGTGATTGGGGC
	DNMT3A2_554	CACCGTCTTTGGCATCAATCATCACAAACGTGATGATTGATGCCAAAGAC
	DNMT3A2_381	CACCGCGCACATGTAGCAGTTCCAGAAACCTGGAACTGCTACATGTGCGC
	DNMT3A2_422	CACCGTGGGCTTCCTCTTCTCAGCTAAACAGCTGAGAAGAGGAAGCCCAC
	DNMT3A2_343	CACCGCCCACAGCAGATGGTGCAGTAAACACTGCACCATCTGCTGTGGGC

	DNMT3A2_613	CACCGTCATGAAGACAGGAAAATGC AAACGCATTTTCCTGTCTTCATGAC
	DNMT3A2_468	CACCGCGACGTACATGATCTTCCCC AAACGGGGAAGATCATGTACGTCGC
	DNMT3A_260	CACCGTTCTCCGCTGTGCTCTTCCG AAACCGGAAGAGCACAGCGGAGAAC
	DNMT3A_382	CACCGCCGCACATGTAGCAGTTCCA AAACTGGAACTGCTACATGTGCGGC
	DNMT3A_384	CACCGCCCGCACATGTAGCAGTTCC AAACGGAACTGCTACATGTGCGGGC
	DNMT3A_568	CACCGTTTCACCAACCTGTTCATAC AAACGTATGAACAGGTTGGTGAAAC
R882H Knock-in	sgRNA for CRISPR-Cas9 R882H knock-in	CCAACATGAGCCGCTTGGCG
	HDR template for R882H knock-in	CAGGGTATTTGGTTTCCCAGTCCACT ATACTGACGTCTCCAACATGAGCCAC TTGGCGAGGCAGAGACTGCTGGGCC GGTCATGGAGCGTGCCAGTCATCC
Genotyping (bold=gDNA binding region)	pNL1.155 F3-sequence around R882	ACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNN**CAGCACTCACCCTG CCC**
	pNL1.155 R3-sequence around R882	TGGAGTTCAGACGTGTGCTCTTCCGA TCT**TCGCTACCTCAGTTTGCCC**
	pNL1.175 F1-sequence around G532	ACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNN**TCCTGGTGGTTTCT GACCCT**
	pNL1.175 R1-sequence around G532	TGGAGTTCAGACGTGTGCTCTTCCGA TCT**CCAAGGTGTGCTACCTGGAA**
	pEG1.103 3F-sequence exon 10	TGGAGTTCAGACGTGTGCTCTTCCGA TCT**GCCCTCCAGCCAGGCTCCTA**
	pEG1.103 3R-sequence exon 10	ACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNN**TTGCGTGGAGTGTG TGGACCT**
	pEG1.103 4F-sequence exon 11	TGGAGTTCAGACGTGTGCTCTTCCGA TCT**CCCATCCTGGGACAAGGCGG**
	pEG1.103 4R-sequence exon 11	ACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNN**GGTTGTGGGGCCT GAGCTGT**
	pNL1.243 F4- sequence exon 11	TGGAGTTCAGACGTGTGCTCTTCCGA TCT**GTGGGAGCTTGGGACACC**
	pNL1.243 R4- sequence exon 11	ACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNN**CAGCACCTCTTGGG CCTG**
	pNL1.243 F6- sequence exon 15	TGGAGTTCAGACGTGTGCTCTTCCGA TCT**AGTGTGTGGCTCCTGAGAGA**

	pNL1.243 R6.2- sequence exon 15	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN**CTATGGGTCATCCCACCTGC**
	pEG1.136 1F- sequence around sgIntron-1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN**ATCAAAGAGAGACAGCACCCG**
	pEG1.136 1R- sequence around sgIntron-1	TGGAGTTCAGACGTGTGCTCTTCCGATCT**GGGCACAAGGGTACCTACG**

Table S3: Melting temperatures of recombinant DNMT3A2 variants by DSF

Variant	**Replicate 1 (°C)**	**Replicate 2 (°C)**
DNMT3A2_WT	43.4 ± 0.1	43.2 ± 0.1
DNMT3A2_C541Y	44.5 ± 0.2	44.1 ± 0.0
DNMT3A2_D781N	43.9 ± 0.1	43.9 ± 0.0
DNMT3A2_4aaDel	45.8 ± 1.0	43.2 ± 0.1
DNMT3A2_G532N	45.1 ± 0.2	44.7 ± 0.2