Supporting Information

Base Editor Scanning Reveals Activating Mutations of DNMT3A

Emma M. Garcia^{1,2}, Nicholas Z. Lue^{1,2}, Jessica K. Liang¹, Whitney K. Lieberman¹, Derek D. Hwang¹, James Woods^{1,2}, Brian B. Liau*^{1,2}

Affiliations:

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

Table of Contents

Supplementary Figures:

- 1. Figure S1: Pie chart and allele table showing targeted deep sequencing of R882 codon of *DNMT3A* in *DNMT3A*^{WT/R882H} cells.
- 2. Figure S2: Silencing dynamics of dDNMT3A reporter system, representative gating scheme, and additional plots associated with Figure 1.
- 3. Figure S3: Western blot of reporter cells transduced with base editing sgRNAs.
- 4. Figure S4: Additional sgRNAs tested.
- 5. Figure S5: Genotyping of $DNMT3A^{WT/R882H}$ cells treated with sgRNAs.
- 6. Figure S6: D618_K621del mutation is found in colobinae monkeys.
- 7. Figure S7: Additional *in vitro* validation of DNMT3A variants.

Supplementary Tables:

- 1. Table S1: Plasmids constructed in this study.
- 2. Table S2: Primers and oligos used in this study.
- 3. Table S3: Melting temperatures of recombinant DNMT3A2 variants by DSF
- 4. Table S4: *DNMT3A*^{WT} cell line screen results. (separate xlsx)
- 5. Table S5: $DNMT3A^{WT/R882H}$ cell line screen results. (separate xlsx)
- 6. Table S6: Additional data on activating guides. (separate xlsx)

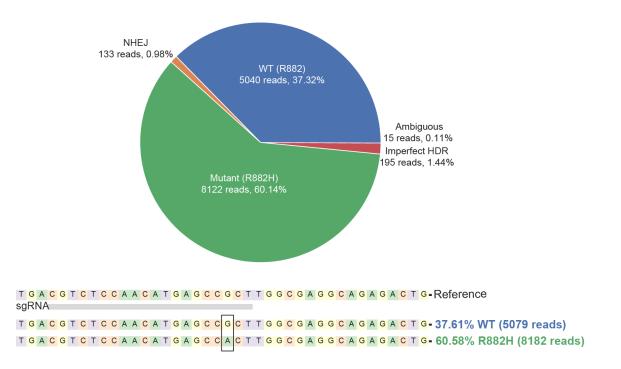


Figure S1: Pie chart and allele table showing targeted deep sequencing of R882 codon of *DNMT3A* **in** *DNMT3A*^{WT/R882H} **cells.** K562 cells are triploid, and these genotyping results are consistent with our cell line containing 2 copies of *DNMT3A*^{R882H} and one copy of *DNMT3A*^{WT}. sequences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

a. Silencing of citrine reporter with dox treatment

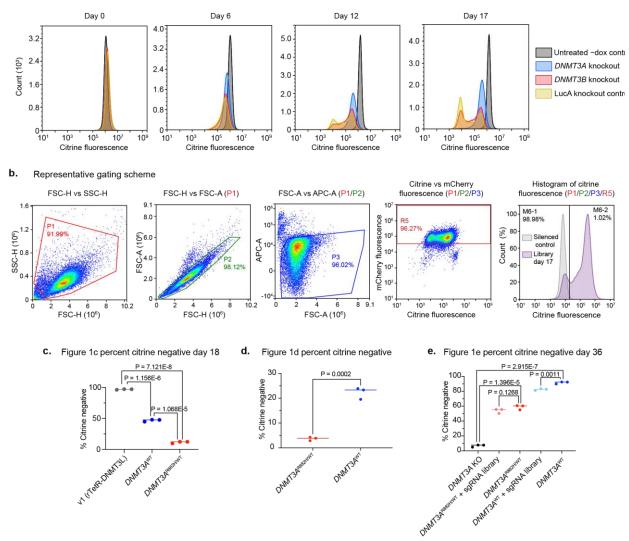
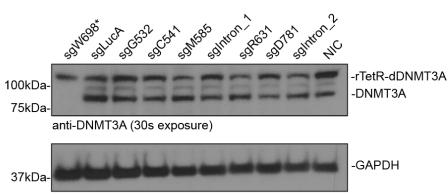


Figure S2: Silencing dynamics of dDNMT3A reporter system, representative gating scheme, and additional plots associated with Figure 1.

- 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 twosided 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.

Expression of DNMT3A in DNMT3A^{WT/R882H} cells treated with sgRNAs



anti-GAPDH (1s exposure)

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.



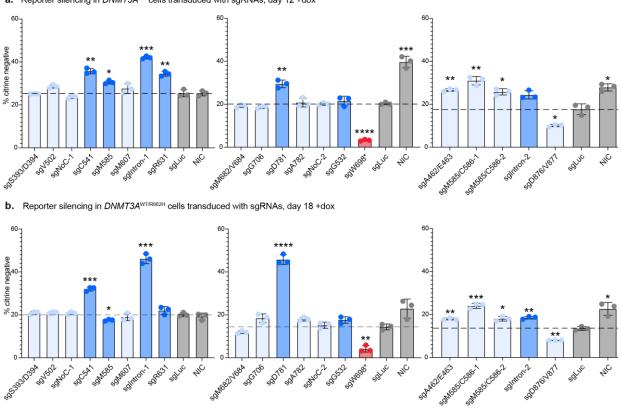


Figure S4: Additional sgRNAs tested.

- 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 \pm 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.

a.		A	lle	le	fre	qu	ier	ю	'a	na	lys	sis	of	ba	ise	e-edited cells	
sgC541 Wild-type:	<u>A</u>	T	С	T	×° G	С	Т	G	×Þ T	G	G	G	G	G	c	Product	Allele frequency
Top base editing outcomes	A A A	539 T T	с с с	T T	54 G G G	c c c	T T	A G A	T T T	G	G G G	2 G G G	A A	A A A	T T T	539 I C Y GG 543 539 I C C GG 543 539 I C Y R G 543	45.9% 34.8% WT 11.3%
sgM585 Wild-type:	A	A	с	т	G	с	т	A 7584	<mark>%</mark> ℃	A	т 158	G	× ^A T	G	C B	Product	Allele frequency
Top base editing outcomes	A	A A	C C	T T	G G	C C	T T	A A	C C	A A	T T	G A	T T	G G	C C	582NCYMC 586 582NCYIC 586	61.6% WT 34.1%
sgR631 Wild-type: pase ontcomes butcomes		A 62 A A A	6 9 0 0 0 9 0 0	A A A A	A (63) A A A	G G G G G G	A A A A	G G G G A	° G G G G G A A A		A (63) A A A	G G G A	С	C 633 C C C C		Product 639E K R K P 633 639E K K K P 633 639E K K K P 633 639E K K K P 633	Allele frequency 43.5% WT 18.4% 7.6% 6.8%
sgD781 Wild-type:	_	A T /77	G G 8	A	А Т И77	A G	A	A * T	A T	A G	A A 078	G T	G	C C	с с 2	• Product	6.5% Allele frequency
Top base editing outcomes	G G	T T	G G G	A A A	T T		A A A	T T T		A G A	A A A	Т	G G G	C C	C C	778VMINA782 778VMIDA782 778VIINA782	49.5% 30.9% WT 11.9%

b. Sequencing of sgIntron-1 locus DNA vs cDNA

DNA Wild-type:	° ¢	C	т	т	× A	G	G	A	с	с	с	т	с	с	A	A	A	G	G	т	т	Allele
g Jes		i	ntr	on			C	61	8	F	P61	9	F	P620	D	ĸ	62	1	١V	62	2	frequency
	G(С	т	Т	А	G	G	А	С	С	С	Т	С	С	А	А	А	G	G	Т	Т	64.2% WT
Top b: editir outcor	G	5	т	т	A	Α	Α	A	С	С	С	т	С	С	A	Α	A	G	G	т	т	28.5%
cDNA		(Dri	gin	al	sp	lice	si	te													
Wild-type:		-	-	-	-	-	G	А	С	С	С	т	С	С	А	А	А	G	G	т	т	Allele
g les		i	ntr	on			C	61	8	F	P61	9	F	P620	D	ĸ	62	1	٧	62	2	frequency
Top base editing outcomes		-	-	-	-	-	G	А	С	С	С	т	С	С	А	А	А	G	G	Т	т	51.7% WT
Top ba editin outcon			-	-	-	-	-	-	-	-	-	-	-	-	•	-	•	-	G	т	т	32.8%
															Ne	ws	spli	ce	sit	e		

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

Organism/sequence				Translation																
Human DNMT3A (Q9Y6K1):	A	Ν	Ν	н	D	Q	Е	F	D	Р	Ρ	κ	۷	Y	Ρ	Ρ	v	Ρ	<u>A</u>	528
																	V			
Rhinopithecus bieti (Black snub-nosed monkey):	A	Ν	Ν	Н	D	Q	Е	F	-	-	-	-	V	S	А	G	Ρ	G	Α	
Rhinopithecus roxellana (Golden snub-nosed monkey):	A	N	N	Н	D	Q	E	F	-	-			V	S	Α	G	Ρ	G	A	
Colobus angolensis palliatus (Peters' Angolan colobus):	A	Ν	N	Н	D	Q	E	F	-	-	-		V	S	А	G	Ρ	G	Α	

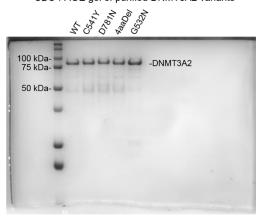
Overall sequence identity

	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

a. SDS-PAGE gel of purified DNMT3A2 variants

b. Replicate 2 DNMT3A mutant activation + H3K4me0



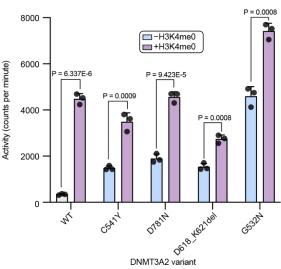


Figure S7: Additional in vitro validation of DNMT3A variants.

- 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 \pm 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.

Application	Plasmid	Description
Base editing screen	pEF-H2B-mCherry-T2A-rTetR-	Lentiviral, mammalian expression
and validation	dDNMT3A2	of updated reporter
	pET28b-His ₆ -DNMT3A2-C541Y	Expression of His ₆ -tagged
		DNMT3A2 C541Y
Bacterial Protein	pET28b-His ₆ -DNMT3A2-D781N	Expression of His ₆ -tagged
Expression		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
	Construction of pET28b-His ₆ -	CTTCCTTCGCATTGATCATAACCG
	DNMT3A2-D781N Forward	
	Construction of pET28b-His6-	CGGTTATGATCAATGCGAAGGAAG
	DNMT3A2-D781N Reverse	
	Construction of pET28b-His ₆ -	GGTGGATACACAAATTCTTGGTCATG
	DNMT3A2-4aaDel Forward	G
	Construction of pET28b-His ₆ -	GGTGGATACACAAATTCTTGGTCATG
	DNMT3A2-4aaDel Reverse	G
Plasmid	Construction of pET28b-His ₆ -	CCTCACGACCGCCATAGCAAATCGTA
Construction	DNMT3A2-C541Y Forward	C
Construction		
	Construction of pET28b-His6-	GTACGATTTGCTATGGCGGTCGTGAG
	DNMT3A2-C541Y Reverse	G
	Construction of pEF-H2B-	CCTTTTTTTGGCTCTTTGCGAACGTC
	mCherry-T2A-rTetR-	GTCGCTATGGGAG
	dDNMT3A2 Forward	
	Construction of pEF-H2B-	CTCCCATAGCGACGACGTTCGCAAA
	mCherry-T2A-rTetR-	GAGCCAAAAAAAGG
	dDNMT3A2 Reverse	
	DNMT3A2_552	CACCGGGCATCAATCATCACAGGGT
	$\mathbf{DNIMT}^{2} \mathbf{A} 2 \mathbf{A} 1 \mathbf{A}$	AAACACCCTGTGATGATGATGATGCCC
	DNMT3A2_414	CACCGTCCTAAGCAGTGAGCACAAC AAACGTTGTGCTCACTGCTTAGGAC
	DNMT3A2_207	CACCGTGTCACTCTCACTCGCTGTCG
	DINNI SA2_207	AAACCGACAGCGATGAGAGTGACAC
	DNMT3A2_341	CACCGAGCAGATGGTGCAGTAGGAC
	DINNI 3/12_3+1	AAACGTCCTACTGCACCATCTGCTC
	DNMT3A2 617	CACCGCATCTTATGGTGCACTGAAA
		AAACTTTCAGTGCACCATAAGATGC
Individual	DNMT3A2_310	CACCGGGTAACATTGAGGCTCCCAC
sgRNAs (top=		AAACGTGGGAGCCTCAATGTTACCC
Forward,	DNMT3A2_396	CACCGAAGAACATCTGGAGCCGGGA
bottom=	_	AAACTCCCGGCTCCAGATGTTCTTC
Reverse)	DNMT3A2_493	CACCGCCCCAATCACCAGATCGAAT
		AAACATTCGATCTGGTGATTGGGGC
	DNMT3A2_554	CACCGTCTTTGGCATCAATCATCAC
		AAACGTGATGATTGATGCCAAAGAC
	DNMT3A2_381	CACCGCGCACATGTAGCAGTTCCAG
		AAACCTGGAACTGCTACATGTGCGC
	DNMT3A2_422	CACCGTGGGCTTCCTCTTCTCAGCT
		AAACAGCTGAGAAGAGGAAGCCCAC
	DNMT3A2_343	CACCGCCCACAGCAGATGGTGCAGT
		AAACACTGCACCATCTGCTGTGGGC

	DNMT3A2_613	CACCGTCATGAAGACAGGAAAATGC					
	DINMISA2_015	AAACGCATTTTCCTGTCTTCATGAC					
	DNMT3A2_468	CACCGCGACGTACATGATCTTCCCC					
	DINM13A2_408						
		AAACGGGGAAGATCATGTACGTCGC					
	DNMT3A_260	CACCGTTCTCCGCTGTGCTCTTCCG					
		AAACCGGAAGAGCACAGCGGAGAAC					
	DNMT3A_382	CACCGCCGCACATGTAGCAGTTCCA					
		AAACTGGAACTGCTACATGTGCGGC					
	DNMT3A_384	CACCGCCCGCACATGTAGCAGTTCC					
		AAACGGAACTGCTACATGTGCGGGC					
	DNMT3A_568	CACCGTTTCACCAACCTGTTCATAC					
		AAACGTATGAACAGGTTGGTGAAAC					
	sgRNA for CRISPR-Cas9	CCAACATGAGCCGCTTGGCG					
	R882H knock-in						
R882H Knock-	HDR template for R882H	CAGGGTATTTGGTTTCCCAGTCCACT					
in	knock-in	ATACTGACGTCTCCAACATGAGCCAC					
		TTGGCGAGGCAGAGACTGCTGGGCC					
		GGTCATGGAGCGTGCCAGTCATCC					
	pNL1.155 F3-sequence	ACACTCTTTCCCTACACGACGCTCTT					
	around R882	CCGATCTNNNNCAGCACTCACCCTG					
		CCC					
	pNL1.155 R3-sequence	TGGAGTTCAGACGTGTGCTCTTCCGA					
	around R882	TCTTCGCTACCTCAGTTTGCCC					
	pNL1.175 F1-sequence	ACACTCTTTCCCTACACGACGCTCTT					
	around G532	CCGATCTNNNNTCCTGGTGGTTTCT					
		GACCCT					
	pNL1.175 R1-sequence	TGGAGTTCAGACGTGTGCTCTTCCGA					
	around G532	TCTCCAAGGTGTGCTGCTGCGAA					
	pEG1.103 3F-sequence exon	TGGAGTTCAGACGTGTGCTCTTCCGA					
	10	TCTGCCCTCCAGCCAGGCTCCTA					
Constanting							
Genotyping	pEG1.103 3R-sequence exon	ACACTCTTTCCCTACACGACGCTCTT					
(bold=gDNA	10	CCGATCTNNNNTTGCGTGGAGTGTG					
binding region)		TGGACCT					
	pEG1.103 4F-sequence exon	TGGAGTTCAGACGTGTGCTCTTCCGA					
	11	TCTCCCATCCTGGGACAAGGCGG					
	pEG1.103 4R-sequence exon	ACACTCTTTCCCTACACGACGCTCTT					
	11	CCGATCTNNNNGGTTGTGGGGGCCT					
		GAGCTGT					
	pNL1.243 F4- sequence exon	TGGAGTTCAGACGTGTGCTCTTCCGA					
	11	TCTGTGGGAGCTTGGGACACC					
	pNL1.243 R4- sequence exon	ACACTCTTTCCCTACACGACGCTCTT					
	11	CCGATCTNNNNCAGCACCTCTTGGG					
		ССТБ					
	pNL1.243 F6- sequence exon	TGGAGTTCAGACGTGTGCTCTTCCGA					
1	15	TCTAGTGTGTGGGCTCCTGAGAGA					

pNL1.243 R6.2- sequence exon 15	ACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNNCTATGGGTCATCCC ACCTGC
pEG1.136 1F- sequence around sgIntron-1	ACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNNATCAAAGAGAGACA GCACCCG
pEG1.136 1R- sequence around sgIntron-1	TGGAGTTCAGACGTGTGCTCTTCCGA TCT 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