Supplementary Information for

Base editor scanning charts the DNMT3A activity landscape

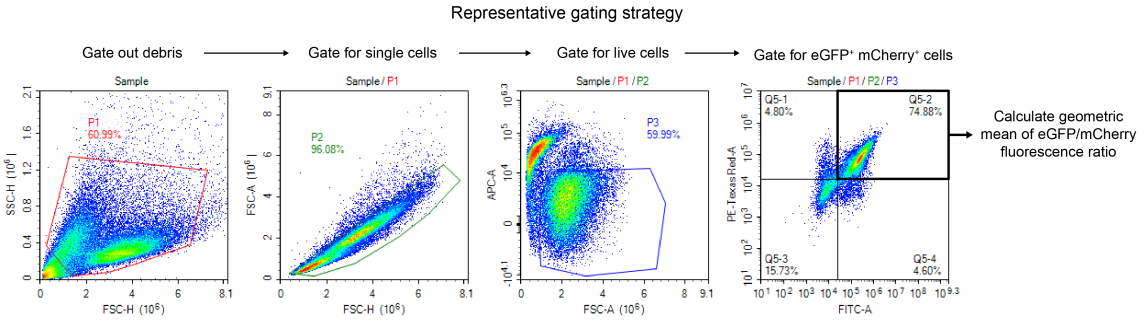
Nicholas Z. Lue, Emma M. Garcia, Kevin C. Ngan, Ceejay Lee, John G. Doench, Brian B. Liau*

Supplementary Figure

Supplementary Tables

Supplementary Data Legends

Supplementary Fig. 1 | Stability reporter gating strategy.
Representative gating strategy used for stability reporter assay measurements. Quadrant gate in mCherry vs. eGFP plot was set so that wild-type parent K562 cells were mCherry/eGFP double-negative.



Supplementary Table 1 | Plasmids constructed in this study.
DNMT3A2 refers to human DNMT3A isoform 2. Dnmt3a2 refers to mouse Dnmt3a isoform 2.

Plasmid	Description
pRRL-5xTetO-pEF-H2B-Citrine-SV40	Lentiviral, mammalian expression of methylation reporter (Addgene #186966)
pRRL-pEF-H2B-mCherry-T2A-rTetR-DNMT3L-SV40	Lentiviral, mammalian expression of methylation reporter (Addgene #186967)
pRRL-pEF-H2B-mCherry-T2A-rTetR-Dest-SV40	Lentiviral destination vector for cloning protein of interest fused to rTetR for reporter (Addgene #186968)
pCilantro2-DNMT3A2	Lentiviral, mammalian expression of DNMT3A2 wild-type
pCilantro2-DNMT3A2_G293K/E294K	Lentiviral, mammalian expression of DNMT3A2 G293K/E294K
pCilantro2-DNMT3A2_R301W	Lentiviral, mammalian expression of DNMT3A2 R301W
pCilantro2-DNMT3A2_I310N	Lentiviral, mammalian expression of DNMT3A2 I310N
pCilantro2-DNMT3A2_S312F	Lentiviral, mammalian expression of DNMT3A2 S312F
pCilantro2-DNMT3A2_R326C	Lentiviral, mammalian expression of DNMT3A2 R326C
pCilantro2-DNMT3A2_W330R	Lentiviral, mammalian expression of DNMT3A2 W330R
pCilantro2-DNMT3A2_S337L	Lentiviral, mammalian expression of DNMT3A2 S337L
pCilantro2-DNMT3A2_E342K	Lentiviral, mammalian expression of DNMT3A2 E342K
pCilantro2-DNMT3A2_R366C	Lentiviral, mammalian expression of DNMT3A2 R366C
pcDNA3-DNMT3A2	Mammalian expression of Flag-DNMT3A2 wild-type
pcDNA3-DNMT3A2_K299N	Mammalian expression of Flag-DNMT3A2 K299N
pcDNA3-DNMT3A2_R301W	Mammalian expression of Flag-DNMT3A2 R301W
pcDNA3-DNMT3A2_W330R	Mammalian expression of Flag-DNMT3A2 W330R
pcDNA3-DNMT3A2_E342K	Mammalian expression of Flag-DNMT3A2 E342K
pcDNA3-DNMT3A2_K343E	Mammalian expression of Flag-DNMT3A2 K343E
pPB-Dnmt3a2-Ires2-mCherry	PiggyBac, mammalian expression of Flag-Dnmt3a2 wild-type (Addgene #186969)
pPB-Dnmt3a2_R297W-Ires2-mCherry	PiggyBac, mammalian expression of Flag-Dnmt3a2 R297W
pPB-Dnmt3a2_W326R-Ires2-mCherry	PiggyBac, mammalian expression of Flag-Dnmt3a2 W326R
pPB-Dnmt3a2_E338K-Ires2-mCherry	PiggyBac, mammalian expression of Flag-Dnmt3a2 E338K
pPB-Dnmt3a2_E752K-Ires2-mCherry	PiggyBac, mammalian expression of Flag-Dnmt3a2 E752K
pET28b-His_6-DNMT3A2_K299N	Bacterial expression of His_6-tagged DNMT3A2 K299N
pET28b-His_6-DNMT3A2_R301W	Bacterial expression of His_6-tagged DNMT3A2 R301W
pET28b-His_6-DNMT3A2_W330R	Bacterial expression of His_6-tagged DNMT3A2 W330R
pET28b-His_6-DNMT3A2_E342K	Bacterial expression of His_6-tagged DNMT3A2 E342K
pET28b-His_6-DNMT3A2_K343E	Bacterial expression of His_6-tagged DNMT3A2 K343E
pET28b-His_6-DNMT3A2_C520Y	Bacterial expression of His_6-tagged DNMT3A2 C520Y
pET28b-His_6-DNMT3A2_G532N	Bacterial expression of His_6-tagged DNMT3A2 G532N
pET28b-His_6-DNMT3A2_R556E	Bacterial expression of His_6-tagged DNMT3A2 R556E
pET28b-His_6-DNMT3A2_E756K	Bacterial expression of His_6-tagged DNMT3A2 E756K
pET28b-His_6-DNMT3A2_E907K	Bacterial expression of His_6-tagged DNMT3A2 E907K
pET28b-His_6-MBP-TEV-PWWP	Bacterial expression of His_6-MBP-tagged DNMT3A PWWP wild-type (Addgene #186970)
pET28b-His_6-MBP-TEV-PWWP_R301W	Bacterial expression of His_6-MBP-tagged DNMT3A PWWP R301W
pET28b-His_6-MBP-TEV-PWWP_R326C	Bacterial expression of His_6-MBP-tagged DNMT3A PWWP R326C
pET28b-His_6-MBP-TEV-PWWP_W330R	Bacterial expression of His_6-MBP-tagged DNMT3A PWWP W330R
pET28b-His_6-MBP-TEV-PWWP_S337L	Bacterial expression of His_6-MBP-tagged DNMT3A PWWP S337L
pET28b-His_6-MBP-TEV-PWWP_E342K	Bacterial expression of His_6-MBP-tagged DNMT3A PWWP E342K

pET28b-His_6-MBP-TEV-PWWP_R366C	Bacterial expression of His_6-MBP-tagged DNMT3A PWWP R366C

Supplementary Table 2 | Mutagenic primers used for plasmid construction.
Mutated codons are indicated in red. All primers were obtained from Sigma-Aldrich.

Mutation	Gene	Expression host	Primer Sequences (5' → 3')	Forward Reverse
G293K/ E294K	Human *DNMT3A*	Mammalian	CTTTGGCATTAAAAAGCTGGTGTGGGGGAAAC CACACCAGCTTTTTAATGCCAAAGCCCCGG
K299N	Human *DNMT3A*	Mammalian	TGGGGGAATCTGCGGGGC CCCCGCAGATTCCCCCACA
R301W	Human *DNMT3A*	Mammalian	GGGGAAACTGTGGGGCTTCTC GAGAAGCCCCACAGTTTCCCC
I310N	Human *DNMT3A*	Mammalian	CCAGGCCGCAATGTGTCTTGG CCAAGACACATTGCGGCCTGG
S312F	Human *DNMT3A*	Mammalian	GCCGCATTGTGTTTTGGTGGATGAC GTCATCCACCAAAACACAATGCGGC
R326C	Human *DNMT3A*	Mammalian	TGAAGGCACCTGCTGGGTCAT ATGACCCAGCAGGTGCCTTCA
W330R	Human *DNMT3A*	Mammalian	CTGGGTCATGCGGTTCGGAGA TCTCCGAACCGCATGACCCAG
S337L	Human *DNMT3A*	Mammalian	ACGGCAAATTCTTAGTGGTGTGTGT ACACACACCACTAAGAATTTGCCGT
E342K	Human *DNMT3A*	Mammalian	GTGGTGTGTGTTAAAAAGCTGATGCCG CGGCATCAGCTTTTTAACACACACCAC
K343E	Human *DNMT3A*	Mammalian	GTGTGTTGAGGAGCTGATGCC GGCATCAGCTCCTCAACACAC
R366C	Human *DNMT3A*	Mammalian	GCAGCCCATGTATTGTAAAGCCATCTACG CGTAGATGGCTTTACAATACATGGGCTGC
R297W	Mouse *Dnmt3a*	Mammalian	GGGGGAAACTTTGGGGCTTCTCC GGAGAAGCCCCAAAGTTTCCCCC
W326R	Mouse *Dnmt3a*	Mammalian	CTGGGTCATGCGGTTCGGAGA TCTCCGAACCGCATGACCCAG
E338K	Mouse *Dnmt3a*	Mammalian	GTGGTGTGTGTGAAAAAGCTCATGCCG CGGCATGAGCTTTTTCACACACACCAC
E752K	Mouse *Dnmt3a*	Mammalian	TCTTCTGGCTCTTTAAAAATGTGGTGGCCA TGGCCACCACATTTTTAAAGAGCCAGAAGA
K299N	Human *DNMT3A*	Bacterial	TTTGGGGCAACCTCCGCGGTT AACCGCGGAGGTTGCCCCAAA
R301W	Human *DNMT3A*	Bacterial	GGGCAAACTCTGGGGTTTCAGTT AACTGAAACCCCAGAGTTTGCCC
R326C	Human *DNMT3A*	Bacterial	CGGAAGGTACGTGCTGGGTTATGTG CACATAACCCAGCACGTACCTTCCG
W330R	Human *DNMT3A*	Bacterial	CGCTGGGTTATGCGCTTCGGTGACGG CCGTCACCGAAGCGCATAACCCAGCG
S337L	Human *DNMT3A*	Bacterial	GACGGTAAGTTCCTGGTTGTTTGCGTG CACGCAAACAACCAGGAACTTACCGTC
E342K	Human *DNMT3A*	Bacterial	GTTGTTTGCGTGAAAAAACTCATGCC GGCATGAGTTTTTTCACGCAAACAAC
K343E	Human *DNMT3A*	Bacterial	TTTGCGTGGAAGAACTCATGCCG CGGCATGAGTTCTTCCACGCAAA
R366C	Human *DNMT3A*	Bacterial	CAGCCGATGTATTGCAAGGCGATCTAC GTAGATCGCCTTGCAATACATCGGCTG
C520Y	Human *DNMT3A*	Bacterial	GTGTCAGAATTGCAAGAACTATTTTCTGGAATGCG CGCATTCCAGAAAATAGTTCTTGCAATTCTGACAC
G532N	Human *DNMT3A*	Bacterial	AGTATGACGACGACAACTACCAGAGTTACT AGTAACTCTGGTAGTTGTCGTCGTCATACT

R556E	Human *DNMT3A*	Bacterial	ACAACTGCTGCGAATGCTTTTGCGT ACGCAAAAGCATTCGCAGCAGTTGT
E756K	Human *DNMT3A*	Bacterial	TGGCTGTTCAAAAACGTGGTGG CCACCACGTTTTTGAACAGCCA
E907K	Human *DNMT3A*	Bacterial	No forward primer CAAGCTTGTCGACGGAGCTCGAATTCCTACTAAA CGCACGCGAAATATTTTTTCAGCGG

Supplementary Table 3 | Oligonucleotides used for individual cloning of key sgRNAs. All oligonucleotides were obtained from Sigma-Aldrich. sgDNMT3A was previously reported in Liao et al., *Nat. Genet.* 47, 469–478, 2015.

sgRNA	Protospacer	Sequences (5' → 3') Forward Reverse
sgLucA (Control)	GGATCTACTGGGTTACCTAA	CACCGGGATCTACTGGGTTACCTAA AAACTTAGGTAACCCAGTAGATCCC
sgDNMT3A (CRISPR KO)	GCATGATGCGCGGCCCAAGG	CACCGGCATGATGCGCGGCCCAAGG AAACCCTTGGGCCGCGCATCATGCC
sgDNMT3B (CRISPR KO)	CCCAACAACACGCAACCAGG	CACCGCCCAACAACACGCAACCAGG AAACCCTGGTTGCGTGTTGTTGGGC
sgW698	GCCCCACTCCTGGATCTGGG	CACCGGCCCCACTCCTGGATCTGGG AAACCCCAGATCCAGGAGTGGGGCC
sgE756	ATTCTCAAAGAGCCAGAAGA	CACCGATTCTCAAAGAGCCAGAAGA AAACTCTTCTGGCTCTTTGAGAATC
sgG532	GGTAGCCGTCGTCGTCGTAC	CACCGGGTAGCCGTCGTCGTCGTAC AAACGTACGACGACGACGGCTACCC
sgG293/E294	GCTCCCCAATGCCAAAGCCC	CACCGGCTCCCCAATGCCAAAGCCC AAACGGGCTTTGGCATTGGGGAGCC
sgR301	AACTGCGGGGCTTCTCCTGG	CACCGAACTGCGGGGCTTCTCCTGG AAACCCAGGAGAAGCCCCGCAGTTC
sgS312.1	CATTGTGTCTTGGTGGATGA	CACCGCATTGTGTCTTGGTGGATGA AAACTCATCCACCAAGACACAATGC
sgS312.2	TGTCTTGGTGGATGACGGGC	CACCGTGTCTTGGTGGATGACGGGC AAACGCCCGTCATCCACCAAGACAC
sgS337.1	AAATTCTCAGTGGTAAGTTG	CACCGAAATTCTCAGTGGTAAGTTG AAACCAACTTACCACTGAGAATTTC
sgS337.2	ATTCTCAGTGGTAAGTTGTG	CACCGATTCTCAGTGGTAAGTTGTG AAACCACAACTTACCACTGAGAATC
sgE342.1	CTTCTCAACACACACCTGGG	CACCGCTTCTCAACACACACCTGGG AAACCCCAGGTGTGTGTTGAGAAGC
sgE342.2	TCAGCTTCTCAACACACACC	CACCGTCAGCTTCTCAACACACACC AAACGGTGTGTGTTGAGAAGCTGAC
sgE342.3 (Not a hit)	GCTTCTCAACACACACCTGG	CACCGGCTTCTCAACACACACCTGG AAACCCAGGTGTGTGTTGAGAAGCC
sgR366	GTACCGCAAAGCCATCTACG	CACCGGTACCGCAAAGCCATCTACG AAACCGTAGATGGCTTTGCGGTACC
sgC520	AAGCAGTTCTAGACAGCAGC	CACCGAAGCAGTTCTAGACAGCAGC AAACGCTGCTGTCTAGAACTGCTTC
sgD641	CATCAAAGAGAGACAGCACC	CACCGCATCAAAGAGAGACAGCACC AAACGGTGCTGTCTCTCTTTGATGC
sgE664/V665	ACACCTCCGAGGCAATGTAG	CACCGACACCTCCGAGGCAATGTAG AAACCTACATTGCCTCGGAGGTGTC
sgD668	GGAGTCCTCACACACCTCCG	CACCGGGAGTCCTCACACACCTCCG AAACCGGAGGTGTGTGAGGACTCCC
sgL737	ACCGCCTCCTGCATGATGCG	CACCGACCGCCTCCTGCATGATGCG AAACCGCATCATGCAGGAGGCGGTC
sgP799	CCTTCCCGGTATGAACAGGT	CACCGCCTTCCCGGTATGAACAGGT AAACACCTGTTCATACCGGGAAGGC
sgT808	ATCCACTGTGAATGATAAGC	CACCGATCCACTGTGAATGATAAGC AAACGCTTATCATTCACAGTGGATC
sgE907	ACTCCTTCAGCGGAGCGAAG	CACCGACTCCTTCAGCGGAGCGAAG AAACCTTCGCTCCGCTGAAGGAGTC

Supplementary Table 4 | Primers used for amplicon deep sequencing.
All primers were obtained from Sigma-Aldrich. Bisulfite sequencing primers were designed using MethPrimer (Li & Dahiya, *Bioinformatics* 18, 1427–1431, 2002).

Name	Target	Sequence
3A-F1	*DNMT3A* exon 7, 5' end	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNACCACTGTGTAATGATTTCTGCTC
3A-R1	*DNMT3A* exon 7, 5' end	TGGAGTTCAGACGTGTGCTCTTCCGATCTGGCCCTGGGATCAAGAACCTT
3A-F2	*DNMT3A* exon 7, 3' end	TGGAGTTCAGACGTGTGCTCTTCCGATCTACCACTGTGTAATGATTTCTGCTC
3A-R2	*DNMT3A* exon 7, 3' end	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGGCCCTGGGATCAAGAACCTT
3A-F3	*DNMT3A* exon 8	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNGTAAGCCTCGGCAAACAAGG
3A-R3	*DNMT3A* exon 8	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGTGCTCTAGGCTCCTCCTC
3A-F4	*DNMT3A* exon 14	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCCTGGTGGTTTCTGACCCT
3A-R4	*DNMT3A* exon 14	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCAAGGTGTGCTACCTGGAA
3A-F5	*DNMT3A* exon 15	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTGGGAGCTTGGGACACC
3A-R5	*DNMT3A* exon 15	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCAGCACCTCTTGGGCCTG
3A-F6	*DNMT3A* exon 16	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGTAACGGTGCTGTCTGCTG
3A-R6	*DNMT3A* exon 16	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCCAGGTGCTGAGTGTG
3A-F7	*DNMT3A* exon 17	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTTCTTCCTGTCTGCCTCTGT
3A-R7	*DNMT3A* exon 17	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGGGTCATCGGGAATAGCTG
3A-F8	*DNMT3A* exon 18	TGGAGTTCAGACGTGTGCTCTTCCGATCTGATGGCTTTCTCTTCCGACCTC
3A-R8	*DNMT3A* exon 18	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCAAACCAAGGTTGCTGGCTAT
3A-F9	*DNMT3A* exon 19	TGGAGTTCAGACGTGTGCTCTTCCGATCTAGTGTGTGGCTCCTGAGAGA
3A-R9	*DNMT3A* exon 19	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCTATGGGTCATCCCACCTGC
3A-F10	*DNMT3A* exon 20	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNCCGCTGTTATCCAGGTTTCTGTT
3A-R10	*DNMT3A* exon 20	TGGAGTTCAGACGTGTGCTCTTCCGATCTGAGAAGCAGGCGGGACAAG
3A-F11	*DNMT3A* exon 22	TGGAGTTCAGACGTGTGCTCTTCCGATCTCAGCACTCACCCTGCCC
3A-R11	*DNMT3A* exon 22	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTCGCTACCTCAGTTTGCCC
Rep-BS-F1	Methylation reporter, 5' end, after bisulfite conversion	TGGAGTTCAGACGTGTGCTCTTCCGATCTGTTTATTTTTTATTAGTGATAGAGAA

Rep-BS-R1	Methylation reporter, 5’ end, after bisulfite conversion	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAA ACCAAACCTCAACTCAAACAC
Rep-BS-F2	Methylation reporter, 3’ end, after bisulfite conversion	TGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCAAACAA TAATTCAAAATTTTT
Rep-BS-R2	Methylation reporter, 3’ end, after bisulfite conversion	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNTT TGGTATGGTGGTAAGTTTGATAT

Primer binding sequence
Overhang for amplification with P5 site primer
Overhang for amplification with P7 site primer

Supplementary Table 5 | Oligonucleotide probe sequences.

Name	Description	Sequence (5' → 3')	Source
Probe-F	EMSA/FP probe, top	/5Cy3/CTGAATACTACTTGGCTCCTC TAACCTGAT	Integrated DNA Technologies
Probe-R	EMSA/FP probe, bottom	ATCAGGTTAGAGGAGCCAAGTAGTA TTCAG	Integrated DNA Technologies

Supplementary Table 6 | Coverage statistics for ChIP-seq.

Sample	Raw reads	Aligned reads after deduplication
TKO ESC ChIP Input	30,946,078	22,369,501
TKO ESC H3K4me3 ChIP	21,919,238	16,852,907
TKO ESC H3K36me2 ChIP	89,059,296	52,330,555

Supplementary Table 7 | Coverage statistics for RRBS.

Sample	Raw reads	Aligned reads	CpGs with 5× coverage	% Bisulfite conversion
WT rep 1	19,313,422	10,675,643	1,361,494	99.75
WT rep 2	20,746,007	11,638,622	1,182,394	99.76
R297W rep 1	25,744,902	14,445,407	1,267,765	99.76
R297W rep 2	19,648,555	11,051,245	1,187,723	99.76
W326R rep 1	24,182,744	13,169,409	1,385,240	99.77
W326R rep 2	18,131,613	9,847,876	1,246,368	99.78
E338K rep 1	24,404,294	14,249,868	1,442,271	99.77
E338K rep 2	16,019,954	8,737,686	1,259,164	99.78
E752K rep 1	21,181,866	11,858,068	1,413,761	99.79
E752K rep 2	21,597,024	11,915,160	1,300,865	99.77

Supplementary Data Legends

Supplementary Data 1–7 are supplied as a separate Excel file.

Supplementary Data 1 | DNMT3A sgRNA library sequences and annotations.
sgRNA_pos and Edit_window annotations refer to genomic coordinates. sgRNA classifications are in the Mut_type column (No_C is equivalent to no predicted edits, Non-exon is equivalent to intron/UTR). Mut_list_3A1 and Edit_site_3A1 contain the expected base editing mutation(s), if any, and the target site assigned to each sgRNA, respectively, using the canonical isoform 1 numbering (out of 912 residues). Mut_list_3A2 and Edit_site_3A2 contain the same, except using isoform 2 numbering (out of 723 residues). Note that the library was designed against isoform 2, and residues 1–24 are not shared with isoform 1.

Supplementary Data 2 | Full DNMT3A base editor scanning data.
DNMT3A base editor scanning read counts after (Log_2 + 1)-transformation and normalization to plasmid library counts. Annotations as in Supplementary Table 1. Mutation_list displays the expected mutations using isoform 1 numbering.

Supplementary Data 3 | Summary of DNMT3A base editor scanning results.
Summary of processed DNMT3A base editor scanning data for citrine$^+$ cells at day 9, showing sgRNA scores and whether each sgRNA is enriched, depleted, or unchanged. sgRNAs tested individually in this paper are indicated along with the result of validation. Other columns are defined as in Supplementary Tables 1 and 2.

Supplementary Data 4 | Summary of PWES analysis and clustering results.
sgRNA scores (sgRNA_score_day9_citrine_pos), cluster assignments (Cluster), and summed ΔPWES scores (Summed_delta_PWES) for missense sgRNAs considered in the PWES analysis (n = 118).

Supplementary Data 5 | Top editing outcomes for key sgRNAs.
Editing outcomes with at least 1% allele frequency are shown for each sgRNA. Genotyping results shown here correspond to bulk base-edited cells (not sorted based on citrine expression) prior to treatment with dox. The Allele column shows the nucleotide sequence within a 40 bp window centered around the sgRNA protospacer (i.e., extending 10 bp past the protospacer in each direction), in the direction of the *DNMT3A* gene. Exonic sequences are shown in uppercase, and intronic sequences are shown in lowercase. The Product column shows the translated protein product within that region.

Supplementary Data 6 | Frequencies of top alleles in base-edited clones.
Alleles with at least 5% frequency are shown for each clone. The Allele column shows the nucleotide sequence within a 40 bp window centered around the sgRNA protospacer (i.e., extending 10 bp past the protospacer in each direction), in the direction of the *DNMT3A* gene (sgR301) or antisense to it (others). Annotation corresponds to legend in Extended Data Fig. 3a.

Supplementary Data 7 | Conservation of PWWP domain residues.
Conservation by residue within a diverse set of proteins containing PWWP domains (Conservation_all_PWWP) or within only proteins sharing the DNMT3A domain structure (Conservation_only_DNMT3A_like). See Methods for details.