

Supplementary Information

Versatile and efficient genome editing with *Neisseria cinerea* Cas9

Zhiquan Liu^{1#}, Siyu Chen^{1#}, Wanhua xie², Hao Yu¹, Liangxue Lai^{1,3,4,5*}, Zhanjun Li^{1*}

```

Nme1Cas9 1 MAAFKPNSTINYILGLDIGIASVGVAMVEIDEEENPRLIDLGVRFERAEVPKTGDLSLAAARRLARSVRRLTRRAHRLRTRRLLKREGVLQAADFEN 100
NcCas9 1 MAAFKPNPMNYILGLDIGIASVGVAMVEIDEEENPRLIDLGVRFERAEVPKTGDLSLAAARRLARSVRRLTRRAHRLRTRRLLKREGVLQAADFEN 100
Nme2Cas9 1 MAAFKPNPIINYILGLDIGIASVGVAMVEIDEEENPRLIDLGVRFERAEVPKTGDLSLAAARRLARSVRRLTRRAHRLRTRRLLKREGVLQAADFEN 100

Nme1Cas9 101 GLIKSLPNTPWQLRAAALDRKLTPLWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYS 200
NcCas9 101 GLIKSLPNTPWQLRAAALDRKLTPLWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNTHALQTGDFRTPAELALNKFEKESGHIRNQRSDYS 200
Nme2Cas9 101 GLIKSLPNTPWQLRAAALDRKLTPLWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYS 200

Nme1Cas9 201 HTFSRKDLQAEILLLFEKQKEFGNPHVSGGLKEGETLLMTQRPALSGDAVQKMLGHCTFEPKAAKNTYTAERFVWLTCLNLRILEQGSERPLTDT 300
NcCas9 201 HTFNKDLQAEINLLFEKQKEFGNPHVSDGLKEGETLLMTQRPALSGDAVQKMLGHCTFEPKAAKNTYTAERFVWLTCLNLRILEQGSERPLTDT 300
Nme2Cas9 201 HTFSRKDLQAEILLLFEKQKEFGNPHVSGGLKEGETLLMTQRPALSGDAVQKMLGHCTFEPKAAKNTYTAERFVWLTCLNLRILEQGSERPLTDT 300

Nme1Cas9 301 ERATLMDEPYRKSCLTYAQARKLLGLLEDTAFFKGLRYGKDNAAEASTLMEKAYHAISRALKEKGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLK 400
NcCas9 301 ERATLMDEPYRKSCLTYAQARKLLGLLEDTAFFKGLRYGKDNAAEASTLMEKAYHAISRALKEKGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLK 400
Nme2Cas9 301 ERATLMDEPYRKSCLTYAQARKLLGLLEDTAFFKGLRYGKDNAAEASTLMEKAYHAISRALKEKGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLK 400

Nme1Cas9 401 DRIDQPEILEALLKHSFDFKVFQISL KALRRIVPLMEQGRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVLRALSQARKVINGVRRYVYSPAR 500
NcCas9 401 DRVQPEILEALLKHSFDFKVFQISL KALRRIVPLMEQGRYDEACTEYGDHYGKKNTEEKIYLPPIPADEIRNPVLRALSQARKVINGVRRYVYSPAR 500
Nme2Cas9 401 DRVQPEILEALLKHSFDFKVFQISL KALRRIVPLMEQGRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVLRALSQARKVINGVRRYVYSPAR 500

Nme1Cas9 501 HIHIEAREVGSFSDRKEIEKRQEENRKDEKAAAKFREYFPNFVGEPSKDILKLRLEYQQHGKCLYSYSGKEINLGRLEKGYVEIDHALPFSRTWDDSF 600
NcCas9 501 HIHIEAREVGSFSDRKEIEKRQEENRKDEKSAAKFREYFPNFVGEPSKDILKLRLEYQQHGKCLYSYSGKEINLGRLEKGYVEIDHALPFSRTWDDSF 600
Nme2Cas9 501 HIHIEAREVGSFSDRKEIEKRQEENRKDEKAAAKFREYFPNFVGEPSKDILKLRLEYQQHGKCLYSYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSF 600

Nme1Cas9 601 NNKVLVLGSENGKNGQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRI L LQKFDEDFGKERNLNDTRYVNRFLCOFVADHMLLTKGKRRVYFASNG 700
NcCas9 601 NNKVLVLGSENGKNGQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRI L LQKFDEDFGKERNLNDTRYVNRFLCOFVADHMLLTKGKRRVYFASNG 700
Nme2Cas9 601 NNKVLVLGSENGKNGQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRI L LQKFDEDFGKERNLNDTRYVNRFLCOFVADHMLLTKGKRRVYFASNG 700

Nme1Cas9 701 QITNLLRGFWGLRKYRAENDRHHALDAVVACSTVAMQQKITRFVRYKEMNAFDGKIDKETGEVLRHQTDFPQWVEFFAQEVMI R VFGKPDGKPEFEEA 800
NcCas9 701 QITNLLRGFWGLRKYRAENDRHHALDAVVACSTVAMQQKITRFVRYKEMNAFDGKIDKETGEVLRHQTDFPQWVEFFAQEVMI R VFGKPDGKPEFEEA 800
Nme2Cas9 701 QITNLLRGFWGLRKYRAENDRHHALDAVVACSTVAMQQKITRFVRYKEMNAFDGKIDKETGEVLRHQTDFPQWVEFFAQEVMI R VFGKPDGKPEFEEA 800

Nme1Cas9 801 DTLEKLRLLAEKLSRPEAVHEVYVPLFVSRAPNRKMSGQGHMETVKSARL---LDEGVSVLRVPLTQLKLDLEKMNRE---REPKLYEALKARLEAH 895
NcCas9 801 DTPEKLRLLAEKLSRPEAVHKYVYVPLFVSRAPNRKMSGQGHMETVKSARL---LDEGVSVLRVPLTQLKLDLEKMNRE---REPKLYEALKARLEAH 895
Nme2Cas9 801 DTPEKLRLLAEKLSRPEAVHEVYVPLFVSRAPNRKMSG-AHKDITLSAKRFVKHNEKISVSRVWLTETIKLADLENMNVNYKNGREIELYEALKARLEAY 899

Nme1Cas9 896 KDDPAKAF AEPFYKYDKAGNRTQQVKAVRVEQVQKTGWVWHNHG--IADNATMVRVDVFEKGDK-----YVLVPIYSWQVAKGILPDRAVVQKDEEDW 988
NcCas9 896 KDDPAKAF AEPFYKYDKAGNRTQQVKAVRVEQVQKTGWVWHNHG--IADNATMVRVDVFEKGDK-----YVLVPIYSWQVAKGILPDRAVVQKDEEDW 988
Nme2Cas9 900 GGNKAQAFDPKDNPFYKGG--QLVKAVRVEKTESEGVLLNKKNAYTIADNGDMVRVDVFEKVDKGGKNGYFIVPIYAWQVAENILPDIDCKGYR----- 992

Nme1Cas9 989 QLIDDSFNFKFSLHPNDLVEVITKKAR--MFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCLKRRPPVR 1082
NcCas9 989 TVMDSSEFKFVLYANDLIKLTAKKNE--FLGYEVLNRATGALDIRTHTDSTKGNKIFQSVGVTALSFQKYQIDELGKEIRPCLKRRPPVR 1082
Nme2Cas9 993 --IDDSYTCFSLHKYDLAFQKDEKSKVEFAYINQDSSNGRFYLAWHDKSKKEQ-----FRISTQNLVLIQKYQVNELGKEIRPCLKRRPPVR 1082

```

Figure S1. Amino acid sequence alignment of Nme1Cas9, Nme2Cas9 and NcCas9 orthologues.

Alignment was performed with Clustal Omega¹ and rendered using Jalview². The PAM-interacting H1024 and N1029 residues are indicated in red.

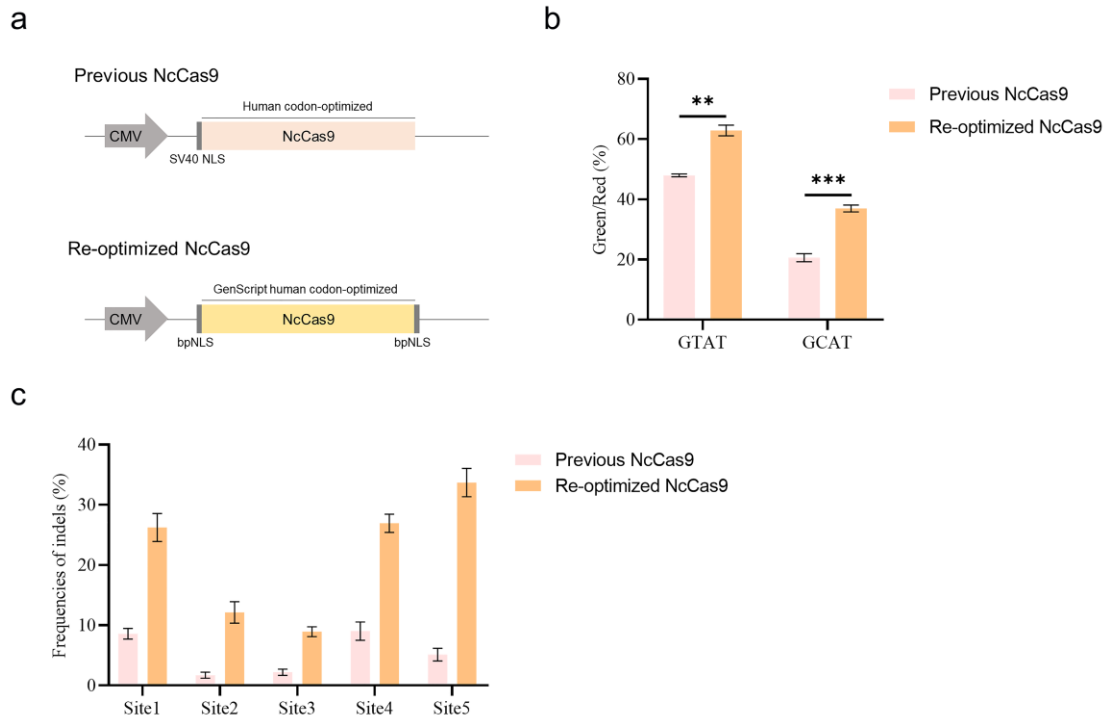


Figure S2. (a) Schematic architecture of previously reported NcCas9³ and our re-optimized NcCas9. **(b)** Comparison of the editing efficiency by pmTmG reporter assays. Red or green represents the tdTomato or EGFP signal, respectively. **(c)** Comparison of the editing efficiency by 5 endogenous loci. Error bars indicate the s.e.m. (n=3).

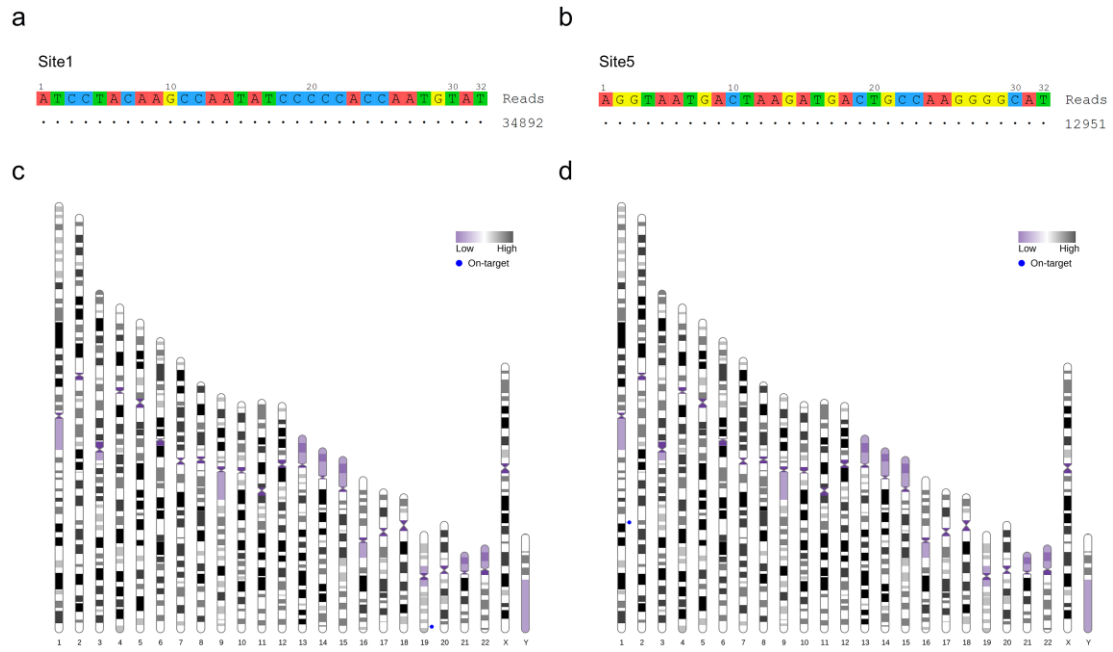


Figure S4. Guide-seq analysis. **(a-b)** On-target and off-target sites evaluated using GUIDE-seq with NcCas9 and sgRNA targeting Site1 **(a)** and Site5 **(b)**. **(c-d)** Visual image reveals on-target and off-target sites with dots on chromosomes in Site1 **(c)** and Site5 **(d)**.

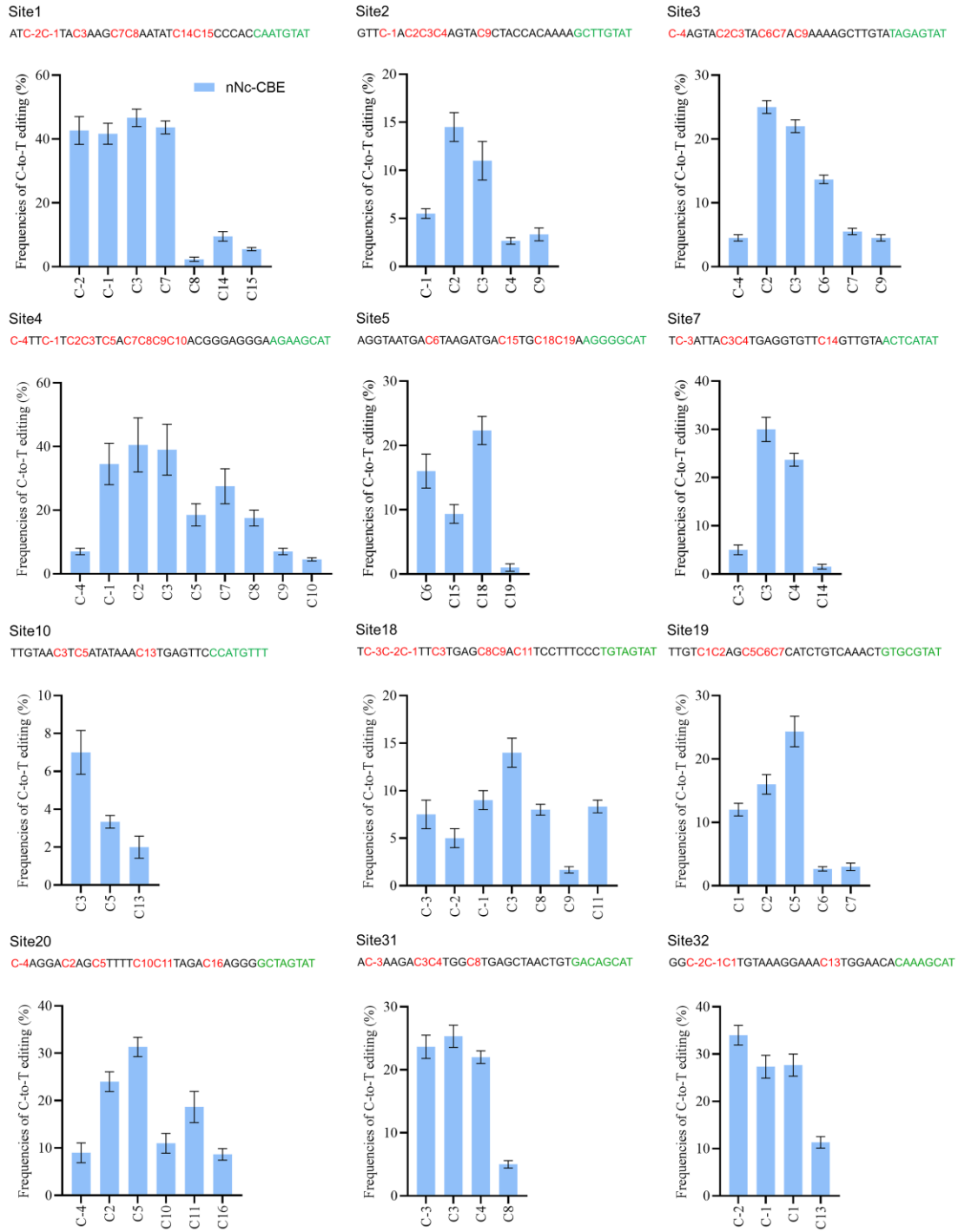


Figure S5. The C-to-T editing frequencies induced by nNc-CBE on each cytosine at 12 target sites.

The edited Cs and PAM are shown in red and green, respectively.

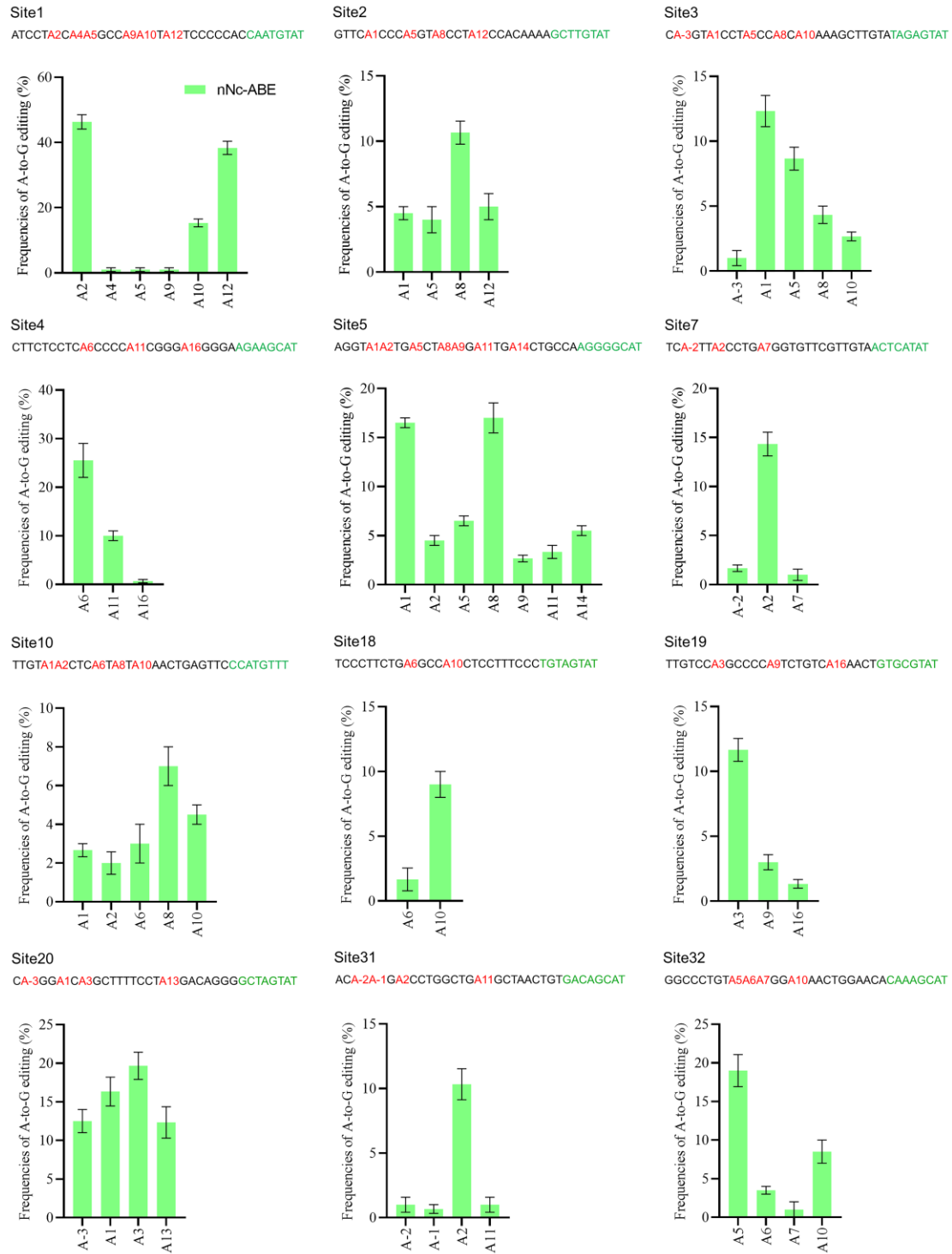


Figure S6. The A-to-G editing frequencies induced by nNc-ABE on each adenine at 12 target sites.

The edited As and PAM are shown in red and green, respectively.

Table S1. Primers used for site-directed mutation in this study.

Mutation	Primers (5'-3')
nNcCas9 (D16A)	F: ACATCCTGGGGCTGGCCATCGGAATCGCAA R: GCCAGCCCCAGGATGTAGTTCATAGGGTTT

Table S2. Target sites used in this study. Target sequence (black), PAM region (green).

PAM	Target site	Locus	Sequence (5'-3')
N4GTAT	Site1	<i>AAVS1</i>	ATCCTACAAGCCAATATCCCCCACAATGTAT
	Site2	<i>RNF2</i>	G TTCACCCAGTACCTACCACAAAAAGCTTGTAT
	Site3	<i>RNF2</i>	CAGTACCTACCACAAAAAGCTTGTATAGAGTAT
	Site13	<i>AAVS1</i>	CCAGGGGGTCTGTGAACTTTCTGACGTTGTAT
	Site14	<i>RNF2</i>	AATTTATATTCAAGCATTCTGACTTCTGTAT
	Site15	<i>CHM13 chromosome 4</i>	TTAATAGATACTTATGAACACTAGCCTTGTAT
	Site16	<i>CHM13 chromosome 4</i>	ACTTACCTTATCCTGGAGTCTATGCTTAGTAT
	Site17	<i>CHM13 chromosome 4</i>	GAGACACATACAAGGCTAGTGTTCAATAAGTAT
	Site18	<i>CHM13 chromosome 4</i>	TCCCTTCTGAGCCACTCCTTTCCCTGTAGTAT
	Site19	<i>VISTA hs267</i>	TTGTCCAGCCCCATCTGTCAAAGTGTGCGTAT
	Site20	<i>LINC01509</i>	CAGGACAGCTTTTCCTAGACAGGGGCTAGTAT
	Site21	<i>LINC01509</i>	CTCCAGCCCAGCCAAACTTGTCACCAGTAT
	Site22	<i>DYRK1A</i>	CCTCTAACTAACCAGGTAAGTTCATGGAGTAT
	Site23	<i>DYRK1A</i>	ACTCAACACAACAGTTATGTTTTAATGGTAT
	Site24	<i>DYRK1A</i>	TAAAGGAAATATTCATATCCTAATGTTTGTAT
Site25	<i>POR</i>	CTGGTGCACCCCAGGCTCAGTCTGCCGTGTAT	
N4GCAT	Site4	<i>AAVS1</i>	CTTCTCCTCACCCCACGGGAGGGAAGAAGCAT
	Site5	<i>RNF2</i>	AGGTAATGACTAAGATGACTGCCAAGGGGCAT
	Site6	<i>FANCF</i>	GGGAGATTGACATGCATTTTCGACCAATAGCAT
	Site26	<i>FANCF</i>	TGCGCCACATCCATCGGCGCTTTGGTCGGCAT
	Site27	<i>EMX1</i>	AAGGGCTCCCATCACATCAACCGGTGGCGCAT
	Site28	<i>CHM13 chromosome 4</i>	AATGAAGTTATGATCGAAAGAATACTAAGCAT
	Site29	<i>CHM13 chromosome 4</i>	AACACTAGCCTTGTATGTGTCTCAGGTGGCAT
	Site30	<i>CHM13 chromosome 4</i>	CAGCTTCCTCGGACTTTGGGGTAGGTTTGCAT
	Site31	<i>VISTA hs267</i>	ACAAGACCTGGCTGAGCTAACTGTGACAGCAT
	Site32	<i>VISTA hs267</i>	GGCCCTGTAAAGGAAACTGGAACAACAAGCAT
	Site33	<i>LINC01509</i>	AGGCCAGGCTGGAGAAGCAGAAAAAAGCAT
	Site34	<i>LINC01509</i>	ACCCAGAGCCCAGTGGCTTTCAGCACCTGCAT
	Site35	<i>DYRK1A</i>	TTATCTTCTGTAGGAGGAGAGACTTCAGCAT
	Site36	<i>DYRK1A</i>	ACGGTGCAAGCCGAACAGATGAAGTTTGCAT
Site37	<i>POR</i>	CGGCCTCACCCCTGGTCTCCCCTTCCAGCAT	
Site38	<i>POR</i>	GCCCATCTCCCCATGTACACCTGGCCGCAT	
N4ATAT	Site7	<i>RNF2</i>	TCATTACCTGAGGTGTTTCGTTGTAACATCATAT
	Site8	<i>RNF2</i>	CCAACATACAGAAGTCAGGAATGCTTGAATAT
	Site9	<i>RNF2</i>	TAAAAATGTATCCCAGTTTACACGTCTCATAT
	Site39	<i>CHM13 chromosome 4</i>	AAGCATAGACTCCAGGATAAGGTAAGTAATAT
	Site40	<i>CHM13 chromosome 4</i>	TTGTATGTGTCTCAGGTGGCATGAGGCAATAT
	Site41	<i>VISTA hs267</i>	AGTACCACCAGTTTTATGAAGAAAATGATAT
	Site42	<i>DYRK1A</i>	CTAAAGTTCTTATTTAAGGAACCATTAGATAT
	Site43	<i>DYRK1A</i>	AGATATGTCAAATGATACAAACATTAGGATAT

N4GTTT	Site10	<i>RNF2</i>	TTGTAAGTCATATAAACTGAGTTCCCATGTTT
	Site11	<i>RNF2</i>	ATTTATTACTCTATGTTCTATTTAAGTTT
	Site12	<i>RNF2</i>	AGTTTTTCATGTTCTAAAAATGTATCCCAGTTT
	Site44	<i>EMX1</i>	TCAGGCCCTTCCTCCTCCAGCTTCTGCCGTTT
	Site45	<i>EMX1</i>	AGCAGCAAGCAGCACTCTGCCCTCGTGGGTTT
	Site46	<i>DYRK1A</i>	TAAGTGTGTGTTGAGTAACATATACCTGTTT
	Site47	<i>DYRK1A</i>	TAACATATACCTGTTTGTAGTTAGAAAAGTTT
	Site48	<i>DYRK1A</i>	TCGCCAGCCAAACATAAGTGACCAACAGGTTT
N4GYAT	Tyr-1	<i>Tyr</i>	TAATGATATCAACATCTACGACCTCTTTGTAT
	Tyr-2	<i>Tyr</i>	TCAACATCTACGACCTCTTTGTATGGATGCAT
	Pcsk9-1	<i>Pcsk9</i>	GGAAGATGGAAGCAGCCAGGTGGAGGTGTAT
	Pcsk9-2	<i>Pcsk9</i>	GAGGAAGACTCCTTTGTCTTCGCCAGAGCAT

Table S3. Primers used for genotyping in this study.

Primers	Sequence (5'-3')	Product size (bp)	Target sites
AAVS1-F: AAVS1-R:	GCAGTCGCCTTGTGATTTG GCGTTTCTCCTTGACATTCTTT	391	Site1, 4, 13
RNF2-F: RNF2-R:	GTCTTCCTTGGTGCCTTATCA TGGTGGGACACATACATTCAAG	524	Site2, 3, 5, 7-12, 14
FANCF-F: FANCF-R:	ATCTGCTCTCCCTCCACTAA GGCCTGGAAGTTTCGCTAAT	483	Site6
CHM13-F: CHM13-R:	CATCTTCTCCGCAGACAGAAATA TCCTGCCCTTAGGTCAAATAAC	465	Site15-18, 28-30, 39, 40
VISTA-F: VISTA-R:	GGACGTCTGCCAATATGTAA GCATTAATCCACAGCAACACC	455	Site19, 31, 32, 41
LINC-F: LINC-R:	GAAACGCCCATGCAATTAGTC ATCACTTTCTGGCCTCAGTTT	443	Site20, 21, 33, 34
DYRK1A-F: DYRK1A-R:	GGAGCTGGTCTGTTGGAGAA TCCAATCCATAATCCCACGTT	692	Site22-24, 35, 36, 42, 43, 46-48
POR-F: POR-R:	CTCCCCTGCTTCTTGTCTGAT ACAGGTCGTGGACTCACA	380	Site 25, 37, 38
EMX1-F: EMX1-R:	CCTTCTGTGAATGTTAGACCCA TGCTTGTCCCTCTGTCAATG	520	Site27, 44, 45
Tyr-F: Tyr-R:	CAGTGCTCAGGCAACTTCAT CGACCTCCCAAGTACTCATCT	497	Tyr-1, Tyr-2
Pcsk9-F1: Pcsk9-R1:	CCCCTCCTTGCTGTAAGATT CACACTTGCTCGCCTGAT	445	Pcsk9-1
Pcsk9-F2: Pcsk9-R2:	CCCAGCCAGAGAGAAATGAG GACTAAGACAGGAAGGGATGAAG	489	Pcsk9-2

Supplementary Note 1

Within the sequences below, NcCas9 sequence is in gray and bpNLS sequences are in green.

Amino acid sequence of NcCas9.

MKRTADGSEFESPKKKRKVAAFKPNPMNYILGLDIGIASVGWAIVEIDEEENPIRLIDLGVR
VFERA EVPKTGDSLAAARRLARSVRLTRRRRAHLLRARRLLKREGVLQAADFENGLI
KSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGV
ADNTHALQTGDFRTPAELALNKFEEKSGHIRNQRGDYSHTFNRKDLQAEINLLFEKQKEF
GNPHVSDGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPTPKAAKNTYTAERFVWLT
KLNNLRILEQGSERPLTDTERATLMDEPYRKSCLTYAQARKLLDLDDETAFFKGLRYGKDN
AEASTLMEMKAYHAISRALKKEGLKDKKSPNLSPQLQDEIGTAFSLFKTDEDITGRLKDR
VQPEILEALLKHISFDKVFQISLKALRRIVPLMEQGNRYDEACTEYGDHYGKKNTEEKIY
LPPIPADEIRNPVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEE
NRKDREKSAKFREYFPNFVGEPKSKDILKRLRYEQQHGKCLYSGKEINLGRNLEKGYVE
IDHALPFSRTWDDSFNNKVLALGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRF
PRSKKQRILLQKFDEEDGFKERNLNDTRYINRFLCQFVADHMLLTGKGKRRVFASNGQITNL
LRGFWGLRKVRAENDRRHALDAVVVACSTIAMQQKITRFVRYKEMNAFDGKTIDKETGE
VLHQKAHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRLLAEKLSSRPEAVHKY
VTPLFISRPNRKMMSGQGHMETVKSARKLDEGISVLRVPLTQLKLDLEKMVNREREPKL
YEALKARLEAHKDDPAKFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVVHNHNGI
ADNATIVRVDVFEKGGKYYLVPIYSWQVAKGILPDRVVQKDEEDWTVMDDSFEFKVF
LYANDLIKLTAKKNEFLGYFVSLNRATGAIDIRTHDSTDSTKGKNGIFQSVGVTALSFOKY
QIDELGKEIRPCRLKKRPPVRSGGSKRTADGSEFEPKKRKV

DNA sequence of NcCas9.

atgaaacggacagccgacggaagcagttcgagtcaccaaagaagaagcggaaagtcGCCGCCTTCAAACCAAACCC
TATGAACTACATCCTGGGGCTGGACATCGGAATCGCAAGCGTGGGCTGGGCTATCGTG
GAAATCGACGAGGAGGAGAACCAATCCGGCTGATCGACCTGGGCGTGAGAGTGTTT
GAGAGGGCAGAGGTGCCAAAGACCGGCGATTCCCTGGCAGCAGCACGGAGACTGGC
ACGGTCTGTGAGGCGCCTGACACGGAGAAGGGCACACAGACTGCTGAGGGCACGCC
GGCTGCTGAAGAGAGAGGGCGTGCTGCAGGCAGCAGACTTCGATGAGAACGGCCTGA
TCAAGAGCCTGCCTAATACCCCATGGCAGCTGAGAGCAGCCGCCCTGGACAGGAAGC
TGACACCACTGGAGTGGTCCGCCGTGCTGCTGCACCTGATCAAGCACCGCGGCTACCT
GTCTCAGCGGAAGAATGAGGGCGAGACAGCCGATAAGGAGCTGGGCGCCCTGCTGAA
GGGAGTGGCCGACAACACACACGCACTGCAGACCGGCGACTTCGGACACCTGCAGA
GCTGGCCCTGAATAAGTTCGAGAAGGAGAGCGGACACATCAGGAACCAGAGGGGGCGA
CTATTCCCACACCTTCAACCGCAAGGATCTGCAGGCCGAGCTGAATCTGCTGTTTGAG
AAGCAGAAGGAGTTCGGCAATCCTCACGTGTCCGACGGCCTGAAGGAGGGCATCGAG
ACACTGCTGATGACCCAGAGGCCAGCCCTGTCTGGCGATGCAGTGCAGAAGATGCTG
GGCCACTGCACATTTGAGCCCACCGAGCCTAAGGCCGCAAGAACACCTACACAGCC
GAGAGGTTCTGTGGCTGACAAAGCTGAACAATCTGAGGATCCTGGAGCAGGGAAGC
GAGAGGCCTCTGACCGACACAGAGAGGGCCACCCTGATGGATGAGCCATACAGGAAG

TCCAAGCTGACATATGCACAGGCAAGGAAGCTGCTGGACCTGGACGATACCGCCTTCT
TTAAGGGCCTGAGATACGGCAAGGATAATGCCGAGGCCTCTACACTGATGGAGATGAA
GGCCTATCACGCCATCAGCCGCGCCCTGGAGAAGGAGGGCCTGAAGGACAAGAAGTC
TCCTCTGAACCTGAGCCCAGAGCTGCAGGATGAGATCGGCACCGCCTTTTCCCTGTTC
AAGACCGACGAGGACATCACAGGCCGCCTGAAGGACAGGGTGCAGCCTGAGATCCTG
GAGGCCCTGCTGAAGCACATCTCTTTTGATAAGTTCGTGCAGATCAGCCTGAAGGCC
TGAGAAGGATCGTGCCACTGATGGAGCAGGGCAATCGCTACGACGAGGCCTGCACCG
AGATCTACGGCGATCACTATGGCAAGAAGAACACAGAGGAGAAGATCTATCTGCCCC
TATCCAGCCGACGAGATCAGAAATCCAGTGGTGCTGAGGGCCCTGAGCCAGGCACG
CAAAGTGATCAACGGAGTGGTGCGCCGGTACGGCTCCCCGCAAGGATCCACATCGA
GACAGCCAGAGAAGTGGGCAAGTCTTTTAAGGACCGGAAGGAGATCGAGAAGAGAC
AGGAGGAGAATAGAAAGGATAGGGAGAAGAGCGCCGCAAGTTCAGGGAGTACTTTC
CCAACTTCGTGGGCGAGCCTAAGTCCAAGGACATCCTGAAGCTGCGCCTGTACGAGCA
GCAGCACGGCAAGTGTCTGTATAGCGGCAAGGAGATCAACCTGGGCCGGCTGAATGA
GAAGGGCTATGTGGAGATCGATCACGCCCTGCCCTTTAGCAGAACCTGGGACGATTCC
TTCAACAATAAGGTGCTGGCCCTGGGCTCTGAGAACCAGAATAAGGGCAATCAGACAC
CATACGAGTATTTCAACGGCAAGGACAATAGCAGGGAGTGGCAGGAGTTTAAGGCAA
GGGTGGAGACAAGCAGGTTCCCACGGTCTAAGAAGCAGCGGATCCTGCTGCAGAAGT
TTGACGAGGATGGCTTCAAGGAGAGAAACCTGAATGACACCAGATACATCAACAGGTT
TCTGTGCCAGTTCGTGGCCGATCACATGCTGCTGACCGGCAAGGGCAAGAGAAGGGT
GTTTGCCTCCAACGGCCAGATCACAAATCTGCTGAGGGGCTTCTGGGGCCTGAGAAAG
GTGAGGGCAGAGAATGACAGGCACCACGCACTGGATGCAGTGGTGGTGGCATGTAGC
ACCATCGCCATGCAGCAGAAGATCACACGCTTTGTGCGGTATAAGGAGATGAACGCCT
TCGACGGCAAGACCATCGATAAGGAGACAGGGCAGGTGCTGCACCAGAAGGCCCACT
TTCCACAGCCCTGGGAGTCTTTGCCAGGAAGTGATGATCCGGGTGTTTGGCAAGCC
AGACGGCAAGCCGAGTTCGAGGAGGCCGATACCCAGAGAAGCTGCGGACACTGCT
GGCAGAGAAGCTGAGCTCCAGACCTGAGGCCGTGCACAAGTACGTGACCCCTCTGTT
CATCTCCCGCGCCCCAAATCGGAAGATGTCTGGCCAGGGCCACATGGAGACAGTGAA
GTCTGCCAAGCGGCTGGACGAGGGCATCAGCGTGCTGAGAGTGCCCCTGACACAGCT
GAAGCTGAAGGATCTGGAGAAGATGGTGAACCGCGAGCGGGAGCCTAAGCTGTATGA
GGCCCTGAAGGCCAGACTGGAGGCCACAAGGACGATCCCGCAAGGCCTTTGCCGA
GCCTTTCTACAAGTATGACAAGGCCGGCAATCGGACCCAGCAGGTGAAGGCCGTGAG
AGTGGAGCAGGTGCAGAAGACAGGCGTGTGGGTGCACAACCACAATGGCATCGCCGA
CAACGCCACAATCGTGCGGGTGGACGTGTTTCGAGAAGGGCGGCAAGTACTATCTGGT
GCCAATCTACTCCTGGCAGGTGGCAAAGGGAATCCTGCCAGACAGAGCCGTGGTGCA
GGGCAAGGACGAGGAGGATTGGACCGTGATGGACGATTCTTTCGAGTTTAAGTTCGTG
CTGTACGCCAACGACCTGATCAAGCTGACCGCCAAGAAGAATGAGTTTCTGGGCTATT
TCGTGTCTCTGAACAGGGCAACAGGAGCAATCGACATCAGGACCCACGACACAGATA
GCACCAAGGGCAAGAACGGCATCTTTCAGAGCGTGGGCGTGAAGACCGCCCTGTCCT
TCCAGAAGTATCAGATTGACGAACTGGGAAAAGAGATTAGACCCTGTAGGCTGAAAA
AGAGACCACCTGTGAGAtctggcggctcaaaaagaaccgccagcggcagcgaattcgagccaagaagaagaggaaagt

Supplementary references

1. Sievers F, *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7**, 539 (2011).
2. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)* **25**, 1189-1191 (2009).
3. Ran FA, *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191 (2015).
4. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* **31**, 3406-3415 (2003).