

## **Engineered Cpf1 Enzymes with Altered PAM Specificities**

Linyi Gao<sup>1,2</sup>, David B.T. Cox<sup>1,3,4</sup>, Winston X. Yan<sup>1,4,5</sup>, John Manteiga<sup>3</sup>, Martin Schneider<sup>1</sup>,  
Takashi Yamano<sup>6</sup>, Hiroshi Nishimasu<sup>6,7</sup>, Osamu Nureki<sup>6</sup>, and Feng Zhang<sup>1,2,8,9,\*</sup>

<sup>1</sup> Broad Institute of MIT and Harvard  
Cambridge, Massachusetts 02142, USA

<sup>2</sup> Department of Biological Engineering

<sup>3</sup> Department of Biology  
Massachusetts Institute of Technology  
Cambridge, MA 02139, USA

<sup>4</sup> Harvard-MIT Division of Health Sciences and Technology

<sup>5</sup> Graduate Program in Biophysics  
Harvard Medical School  
Boston, MA 02115, USA

<sup>6</sup> Department of Biological Sciences, Graduate School of Science

The University of Tokyo, Tokyo 113-0032, Japan

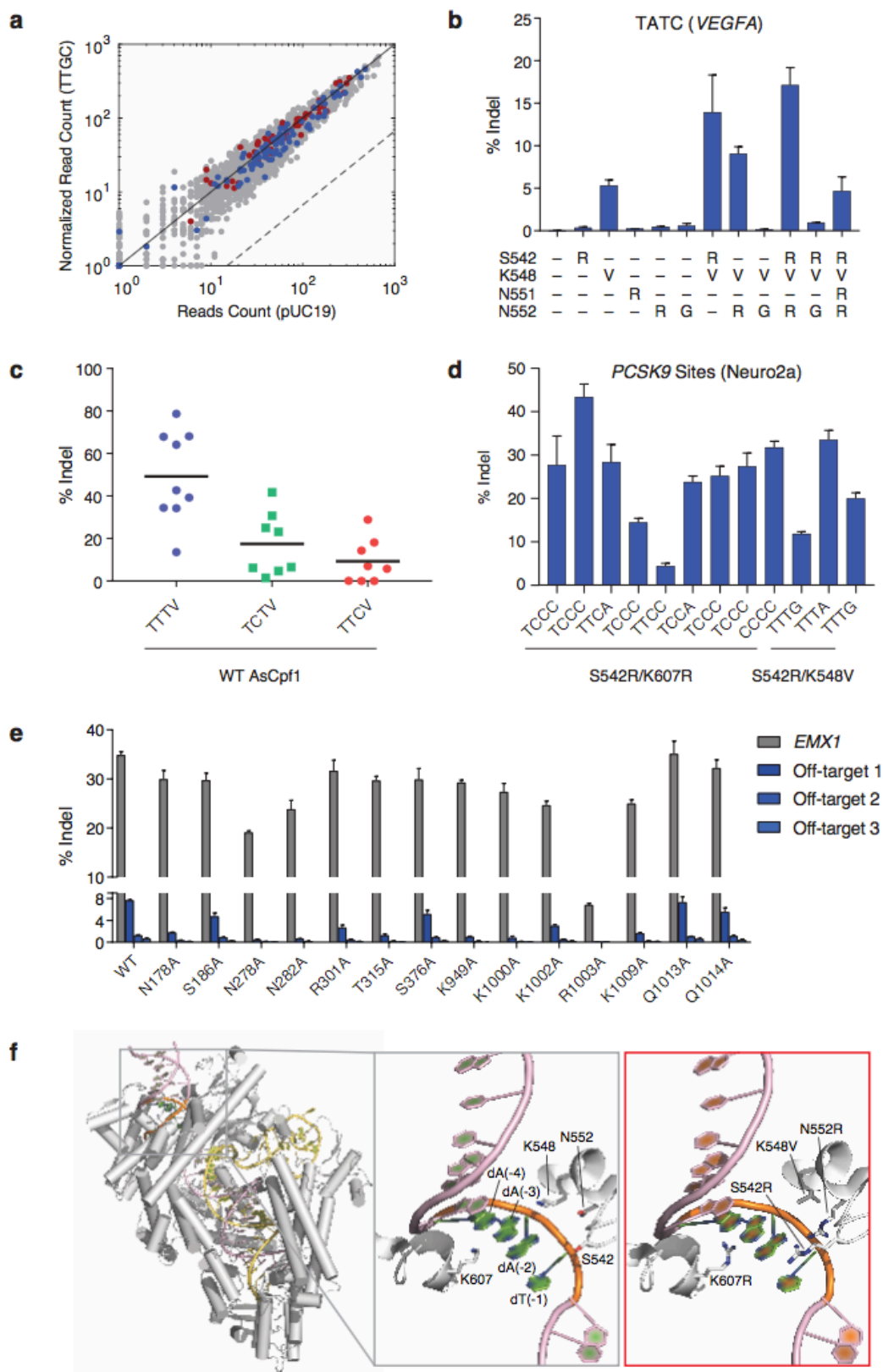
<sup>7</sup> JST, PRESTO, Tokyo 113-0032, Japan

<sup>8</sup> McGovern Institute for Brain Research

<sup>9</sup> Department of Brain and Cognitive Sciences  
Massachusetts Institute of Technology  
Cambridge, MA 02139, USA

\* Correspondence should be addressed to [zhang@broadinstitute.org](mailto:zhang@broadinstitute.org) (F.Z.).

# Supplementary Figure 1



**Supplementary Figure 1** | (a) Bacterial selection screen readout for a TTGC PAM, highlighting depleted hits. Each dot represents a distinct Cpf1 wild-type (WT) or mutant codon. The dashed line indicates 15-fold depletion. Red = stop codon; blue = WT codon. (b) Combinatorial mutagenesis of top single amino acid mutations identified in the bacterial selection screen at a TATC target site in HEK293 cells. (c) Activity of WT AsCpf1 at TTTV, TCTV, and TTCV sites targeting *CFTR*, *EMX1*, *DNMT1*, and *VEGFA* in HEK293 cells. For TCTV and TTCV PAMs, the data shown is replotted from Fig. 2H. (d) Activity of the S542R/K607R variant at TYCV and CCCC target sites within *PCSK9* in Neuro2a cells. (e) Specificity mutagenesis of AsCpf1 using an alanine scan of residues hypothesized to contact one of the two strands of DNA that is unwound during Cpf1 binding and cleavage. (f) Crystal structure of AsCpf1, highlighting the S542, K548, N552 and K607 residues (gray inset) and *in silico* modeled mutations (red inset).

**Supplementary Table 1** | List of residues mutated in the bacterial selection screen.

#	Residue	Position	Domain	#	Res.	Pos.	Domain	#	Res.	Pos.	Domain
1	K	130	REC1	23	Q	536	WED-II	47	F	598	PI
2	G	131	REC1	24	M	537	WED-II	48	P	599	PI
3	L	132	REC1	25	P	538	WED-II	49	D	600	PI
4	F	133	REC1	26	T	539	WED-II	50	A	601	PI
5	K	134	REC1	27	L	540	WED-II	51	A	602	PI
6	A	135	REC1	28	A	541	WED-II	52	K	603	PI
7	F	162	REC1	29	S	542	WED-II	53	M	604	PI
8	D	163	REC1	30	G	543	WED-II	54	I	605	PI
9	K	164	REC1	31	W	544	WED-II	55	P	606	PI
10	F	165	REC1	32	D	545	WED-II	56	K	607	PI
11	T	166	REC1	33	V	546	WED-II	57	C	608	PI
12	T	167	REC1	34	N	547	WED-II	58	S	609	PI
13	Y	168	REC1	35	K	548	WED-II	59	T	610	PI
14	F	169	REC1	36	E	549	WED-II	60	Q	611	PI
15	S	170	REC1	37	K	550	WED-II	61	L	612	PI
16	G	171	REC1	38	N	551	WED-II	62	K	613	PI
17	F	172	REC1	39	N	552	WED-II	63	A	614	PI
18	Y	173	REC1	40	K	570	WED-II	64	V	615	PI
19	E	174	REC1	41	Q	571	WED-II				
20	N	175	REC1	42	K	572	WED-II				
21	R	176	REC1	43	G	573	WED-II				
22	K	177	REC1	44	Y	595	WED-II				
				45	D	596	WED-II				
				46	Y	597	WED-II				

**Supplementary Table 2** | List of guide sequences used for BLISS and indel analysis.

<b>Figure</b>	<b>Gene or Description</b>	<b>PAM</b>	<b>Guide</b>
Fig. 1C	Plasmid interference	Varies	CCGATGGTCCATGTCTGTTACTCGCCTGTC
Fig. 2A	<i>VEGFA</i>	TATC	AAATTCCAGCACCGAGCGCCCTG
Fig. 2B			
Fig. S2			
Fig. 2B	<i>DNMT1</i>	TATA	AAGAAATATTACAACATATAAAA
Fig. 3D			
Fig. 2C	<i>DNMT1</i>	TCCC	GTCACCCCTGTTTCTGGCACCAG
Fig. 3D			
Fig. 2C	<i>DNMT1</i>	TTCC	TGGTGCCAGAAACAGGGGTGACG
Fig. 3D			
Fig. 2D	<i>In vitro</i> cleavage	NNNN	GAGAAGTCATTTAATAAGGCCACT
Fig. 2G	<i>CFTR</i>	TATA	GCAGTTGTCGCAGTTTTACAACC
Fig. 2G	<i>CFTR</i>	TATG	ACCCGGATAACAAGGAGGAACGC
Fig. 2G	<i>CFTR</i>	TATA	GAGTTGATTGGATTGAGAATAGA
Fig. 2G	<i>CFTR</i>	TATC	GCCTCTCCCTGCTCAGAATCTGG
Fig. 2G	<i>VEGFA</i>	TATG	TAGCTGTTTGGGAGGTCAGAAAT
Fig. 2G	<i>VEGFA</i>	TATA	GACATGTCCCATTTGTGGGAACT
Fig. 2G	<i>VEGFA</i>	TATG	TTCGGGTGCTGTGAACTTCCCTC
Fig. 2G	<i>EMX1</i>	TATG	ACCCACTGCGTGGGTTC CATGA
Fig. 2G	<i>EMX1</i>	TATC	CCAAGTCAA ACTTCTCTTCAGTC
Fig. 2G	<i>EMX1</i>	TATA	CCCTTTAGGACACATGCTGTCTA
Fig. 2G	<i>DNMT1</i>	TATC	AGTGCACCTTCGGCGTGCTGCAG
Fig. 2G	<i>DNMT1</i>	TATA	CCCACCATGACAGGAAGAACGGC
Fig. 2G	<i>DNMT1</i>	TATG	AGGCGCTTCCCCAGCACAACTG
Fig. 2H	<i>CFTR</i>	TTCG	GCGATGTTTTTTCTGGAGATTTA
Fig. 2H	<i>DNMT1</i>	TTCA	GCTAAAATAAAGGAGGAGGAAGC
Fig. 2H	<i>DNMT1</i>	TCCC	GTCACCCCTGTTTCTGGCACCAG
Fig. 2H	<i>DNMT1</i>	TTCC	TGGTGCCAGAAACAGGGGTGACG
Fig. 2H	<i>DNMT1</i>	TTCA	GTCTCCGTGAACGTTCCCTTAGC
Fig. 2H	<i>DNMT1</i>	TTCA	CGGAGACTGAACACTCCTCAAAC
Fig. 2H	<i>EMX1</i>	TTCG	TGGCAATGCGCCACCGGTTGATG
Fig. 2H	<i>VEGFA</i>	TTCC	CTGTGGTGGCCGAGCGCCCCCTA
Fig. 2H	<i>VEGFA</i>	TCCA	GTCCCAAATATGTAGCTGTTTGG
Fig. 2H	<i>VEGFA</i>	TCCG	CACGTAACCTCACTTTCCTGCTC
Fig. 2H	<i>VEGFA</i>	TCCC	TCTTTAGCCAGAGCCGGGGTGTG
Fig. 2H	<i>VEGFA</i>	TCCG	CCCCCGAAACTCTGTCCAGAGA
Fig. 2H	<i>VEGFA</i>	TCCG	GGGGCGGATGGGTAATTTTCAGG
Fig. 2H	<i>VEGFA</i>	TCCA	ATAGATCTGTGTGTCCCTCTCCC

Fig. 2H	<i>VEGFA</i>	TTCC	AAAGCCCATTCCCTCTTTAGCCA
Fig. 2H	<i>VEGFA</i>	TCCC	CCCACCCCCTTTCCAAAGCCCAT
Fig. 2H	<i>CFTR</i>	GTCG	AAAATTTTACACCACAAAATGTT
Fig. 2H	<i>CFTR</i>	ACCA	AAGATGATATTTTCTTTAATGGT
Fig. 2H	<i>CFTR</i>	ACCA	TTAAAGAAAATATCATCTTTGGT
Fig. 2H	<i>CFTR</i>	ATCC	TAAACTCATTAAATGCCCTTCGGC
Fig. 2H	<i>CFTR</i>	ATCC	AGGAAAACACTGAGAACAGAATGAA
Fig. 2H	<i>EMX1</i>	ATCA	CATCAACCGGTGGCGCATTGCCA
Fig. 2H	<i>EMX1</i>	GTCC	TCCCCATTGGCCTGCTTCGTGGC
Fig. 2H	<i>EMX1</i>	CCCG	GGCTTCAAGCCCTGTGGGGCCAT
Fig. 2H	<i>EMX1</i>	ATCG	ATGTCACCTCCAATGACTAGGGT
Fig. 2H	<i>EMX1</i>	ATCG	ATGTCCTCCCCATTGGCCTGCTT
Fig. 2H	<i>VEGFA</i>	CCCA	TTCCCTCTTTAGCCAGAGCCGGG
Fig. 2H	<i>VEGFA</i>	CTCG	GCCACCACAGGGAAGCTGGGTGA
Fig. 2H	<i>VEGFA</i>	GTCC	CAAATATGTAGCTGTTTGGGAGG
Fig. 2H	<i>VEGFA</i>	GCCG	AGCGCCCCCTAGTGACTGCCGTC
Fig. 2H	<i>VEGFA</i>	GCCC	ATTCCCTCTTTAGCCAGAGCCGG
Fig. 2H	<i>VEGFA</i>	CCCG	GCTCTGGCTAAAGAGGGAATGGG
Fig. 2H	<i>VEGFA</i>	GCCA	GAGCCGGGGTGTGCAGACGGCAG
Fig. 2H	<i>VEGFA</i>	CTCG	CTCCATTCACCCAGCTTCCCTGT
Fig. 2H	<i>VEGFA</i>	GTCA	GAAATAGGGGGTCCAGGAGCAAA
Fig. 2H	<i>VEGFA</i>	CTCC	AGTCCCAAATATGTAGCTGTTTG
Fig. 2H	<i>VEGFA</i>	GCCC	TGGGCTCTCTGTACATGAAGCAA
Fig. 2H	<i>VEGFA</i>	ACCA	CAGGGAAGCTGGGTGAATGGAGC
Fig. 2H	<i>VEGFA</i>	ACCC	CGGCTCTGGCTAAAGAGGGAATG
Fig. 2H	<i>VEGFA</i>	CCCA	GCTTCCCTGTGGTGGCCGAGCGC
Fig. 2H	<i>VEGFA</i>	GCCG	TCTGCACACCCCGGCTCTGGCTA
Fig. 2H	<i>VEGFA</i>	GCCC	CCTAGTGACTGCCGTCTGCACAC
Fig. 2H	<i>VEGFA</i>	ACCC	CCTATTTCTGACCTCCCAAACAG
Fig. 2H	<i>VEGFA</i>	GCCA	CCACAGGGAAGCTGGGTGAATGG
Fig. 2H	<i>VEGFA</i>	GTCC	TCACTCTCGAAGACGCTGCTCGC
Fig. 2H	<i>VEGFA</i>	GTCA	CTAGGGGGCGCTCGGCCACCACA
Fig. 2H	<i>VEGFA</i>	GCCG	GGGTGTGCAGACGGCAGTCACTA
Fig. 2H	<i>VEGFA</i>	CTCG	AAGACGCTGCTCGCTCCATTAC
Fig. 2H	<i>VEGFA</i>	CCCG	CTCCAACGCCCTCAACCCACAC
Fig. 2H	<i>VEGFA</i>	CTCC	TGGACCCCTATTTCTGACCTCC
Fig. 2H	<i>VEGFA</i>	ATCC	TGGAGTGACCCTGGCCTTCTCC
Fig. 2H	<i>VEGFA</i>	ACCC	CCTTTCCAAAGCCCATTCCCTCT
Fig. 2H	<i>VEGFA</i>	CCCC	CCACCCCTTTCCAAAGCCCATT
Fig. 3A	<i>VEGFA</i>	TTTG	CTAGGAATATTGAAGGGGGCAGG

Fig. 3A	<i>GRIN2B</i>	TTTG	GTGCTCAATGAAAGGAGATAAAGG
Fig. 3A	<i>DNMT1</i>	TTTG	AAGAAATATTACAACATATAAAA
Fig. 3A	<i>EMX1</i>	TTTG	TCCTCCGGTTCTGGAACCACACC
Fig. 3C			
Fig. S5			
Fig. 3B	<i>RPL32P3</i>	TTTG	GGGTGATCAGACCCAACAGCAGG
Fig. S3	<i>CFTR</i>	TTTA	ATGGTGCCAGGCATAATCCAGGA
Fig. S3	<i>DNMT1</i>	TTTC	CCTTCAGCTAAAATAAAGGAGGA
Fig. S3	<i>DNMT1</i>	TTTG	AGGAGTGTTTCAGTCTCCGTGAAC
Fig. S3	<i>DNMT1</i>	TTTC	CTGATGGTCCATGTCTGTTACTC
Fig. S3	<i>DNMT1</i>	TTTA	GCTGAAGGGAAATAAAAGGAAAA
Fig. S3	<i>EMX1</i>	TTTG	GGGAGGCCTGGAGTCATGGCCCC
Fig. S3	<i>EMX1</i>	TTTG	TGGTTGCCACCCTAGTCATTGG
Fig. S3	<i>VEGFA</i>	TTTA	GCCAGAGCCGGGGTGTGCAGACG
Fig. S3	<i>VEGFA</i>	TTTC	CAAAGCCCATTCCTCTTTAGCC
Fig. S4	<i>PCSK9</i> (Neuro2a)	TCCC	GTCCCAGGAGGATGGCCTGGCTG
Fig. S4	<i>PCSK9</i> (Neuro2a)	TCCC	AGGAGGATGGCCTGGCTGATGAG
Fig. S4	<i>PCSK9</i> (Neuro2a)	TTCA	ATCTGTAGCCTCTGGGTCTCCTC
Fig. S4	<i>PCSK9</i> (Neuro2a)	TCCC	TGGCTTCTTGGTGAAGATGAGCA
Fig. S4	<i>PCSK9</i> (Neuro2a)	TTCC	TCAATGTACTCCACATGGGGCAA
Fig. S4	<i>PCSK9</i> (Neuro2a)	TCCA	TGGGATGCTCTGGGCGAAGACAA
Fig. S4	<i>PCSK9</i> (Neuro2a)	TCCC	GATGGGCACCCTGGATGCTGGTA
Fig. S4	<i>PCSK9</i> (Neuro2a)	TCCC	GGCCGCTGACCACACCTGCCAGG
Fig. S4	<i>PCSK9</i> (Neuro2a)	CCCC	GATGGGCACCCACTGCTCTGCGT
Fig. S4	<i>PCSK9</i> (Neuro2a)	TTTG	TTCAATCTGTAGCCTCTGGGTCT
Fig. S4	<i>PCSK9</i> (Neuro2a)	TTTA	TGACCTCTTCCCTGGCTTCTTGG
Fig. S4	<i>PCSK9</i> (Neuro2a)	TTTG	TCTTCGCCCAGAGCATCCCATGG

**Supplementary Table 3** | List of plasmids used in the study.

<b>Name</b>	<b>Backbone</b>	<b>Description</b>	<b>Experiments Used</b>	<b>Addgene #</b>	<b>Ref.</b>
pY010	pcDNA3.1	CMV promoter-driven AsCpf1	Cell lysate and indel (HEK293)	69982	Zetsche et al. 2015
	pcDNA3.1	CMV promoter-driven AsCpf1 (S542R/K607R)	Cell lysate and indel (HEK293)		
	pcDNA3.1	CMV promoter-driven AsCpf1 (S542R/K548V/N552R)	Cell lysate and indel (HEK293)		
pY036		U6-crRNA(BbsI)-CBh-AsCpf1	BLISS (HEK293) and indel (Neuro2a)		
		U6-crRNA(BbsI)-CBh-AsCpf1 (S542R/K607R)	BLISS (HEK293) and indel (Neuro2a)		
		U6-crRNA(BbsI)-CBh-AsCpf1 (S542R/K548V/N552R)	BLISS (HEK293) and indel (Neuro2a)		
pY020		U6-crRNA(BsmBI)-CMV-mCherry	Indel (HEK293)		
	pACYC (modified)	T7-AsCpf1-T7-crRNA	Bacterial screen		