nature biotechnology

Engineered Cpf1 variants with altered PAM specificities

Linyi Gao[1,2], David B T Cox[1,3,4], Winston X Yan[1,4,5], John C Manteiga[3], Martin W Schneider[1], Takashi Yamano[6], Hiroshi Nishimasu[6,7], Osamu Nureki[6], Nicola Crosetto[8] & Feng Zhang[1,2,9,10]

The RNA-guided endonuclease Cpf1 is a promising tool for genome editing in eukaryotic cells[1–7]. However, the utility of the commonly used *Acidaminococcus sp. BV3L6* Cpf1 (AsCpf1) and *Lachnospiraceae bacterium ND2006* Cpf1 (LbCpf1) is limited by their requirement of a TTTV protospacer adjacent motif (PAM) in the DNA substrate. To address this limitation, we performed a structure-guided mutagenesis screen to increase the targeting range of Cpf1. We engineered two AsCpf1 variants carrying the mutations S542R/K607R and S542R/K548V/N552R, which recognize TYCV and TATV PAMs, respectively, with enhanced activities *in vitro* and in human cells. Genome-wide assessment of off-target activity using BLISS[7] indicated that these variants retain high DNA-targeting specificity, which we further improved by introducing an additional non-PAM-interacting mutation. Introducing the identified PAM-interacting mutations at their corresponding positions in LbCpf1 similarly altered its PAM specificity. Together, these variants increase the targeting range of Cpf1 by approximately threefold in human coding sequences to one cleavage site per ~11 bp.

Programmable endonucleases from class 2 microbial CRISPR–Cas systems have enabled a wide range of applications in eukaryotic genome editing[1–7]. Recent work has demonstrated that in addition to the widely used type II-A Cas9, the type V-A system Cpf1 can mediate efficient genome editing. Cpf1 has several advantages compared to Cas9; for instance, it has low mismatch tolerance[4–7], does not require a *trans*-activating crRNA, and can process its own CRISPR RNA (crRNA) array into mature crRNAs to facilitate targeting of multiple genes concurrently[2,3].

We previously identified two orthologs of Cpf1 with robust activity in mammalian cells, *Acidaminococcus sp. BV3L6* Cpf1 (AsCpf1) and *Lachnospiraceae bacterium ND2006* Cpf1 (LbCpf1)[1], both of which require a TTTV protospacer-adjacent motif (PAM), where V can be A, C, or G. For applications for which the location of the target site is critical, such as homology-directed repair or generation of loss-of-function mutations at specific exonic positions, the requirement of a TTTV PAM may limit the availability of suitable target sites, reducing the practical utility of Cpf1. To address this limitation, we aimed to engineer variants of Cpf1 that can recognize alternative PAM sequences in order to increase its targeting range.

Previous work has shown that the PAM preference of Cas9 can be altered by mutations to residues in close proximity to the PAM DNA duplex[8–11]. We sought to investigate whether the PAM preference of Cpf1, despite its strong evolutionary conservation across different orthologs[1], can also be modified. Based on the crystal structure of AsCpf1 in complex with crRNA and target DNA[12], we selected 60 residues in AsCpf1 in proximity to the PAM duplex for targeted mutagenesis (**Fig. 1a** and **Supplementary Table 1a**). By randomizing the codons at each position using cassette mutagenesis, we constructed a plasmid library of AsCpf1 variants encoding most single amino acid substitutions at these residues. The use of codon randomization allowed us to attain greater mutational coverage than would have been expected with error-prone PCR, since it prevents representational bias caused by the template sequence.

To identify variants within this library with cleavage activity at non-canonical PAMs, we adapted a plasmid interference-based depletion screen in *Escherichia coli*[1,8,13,14] (**Fig. 1b**). In our modified assay, a pool of *E. coli*, with each bacterium expressing crRNA and a variant of Cpf1 from a plasmid maintained with chloramphenicol, was transformed with a second plasmid carrying an ampicillin-resistance gene and a target site bearing a mutated PAM. Successful cleavage of the second plasmid resulted in the loss of ampicillin resistance and subsequent cell death when grown on ampicillin-selective media. By comparing the sequences of the original library to the sequences of Cpf1-carrying plasmid DNA in surviving bacteria, we determined the variants that were depleted as a result of their novel cleavage activity of the mutated PAM.

To effectively use this approach to distinguish variants with non-canonical PAM activity from wild-type (WT) AsCpf1, we first determined PAM sequences at which WT AsCpf1 had minimal activity. We evaluated the tolerance of WT AsCpf1 to substitution mutations in the PAM, as determined by *E. coli* death due to successful plasmid interference. We focused on PAMs with single-nucleotide substitutions (i.e., NTTV, TNTV, and TTNV, where V was arbitrarily chosen to be C).

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [2]Department of Biological Engineering, Massachusetts Institute of Technology Cambridge, Massachusetts, USA. [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [4]Harvard-MIT Division of Health Sciences and Technology, Harvard Medical School, Boston, Massachusetts, USA. [5]Graduate Program in Biophysics, Harvard Medical School, Boston, Massachusetts, USA. [6]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan. [7]JST, PRESTO, Tokyo, Japan. [8]Science for Life Laboratory, Division of Translational Medicine and Chemical Biology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [9]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [10]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. Correspondence should be addressed to F.Z. (zhang@broadinstitute.org).

Received 27 November 2016; accepted 16 May 2017; published online 5 June 2017; doi:10.1038/nbt.3900

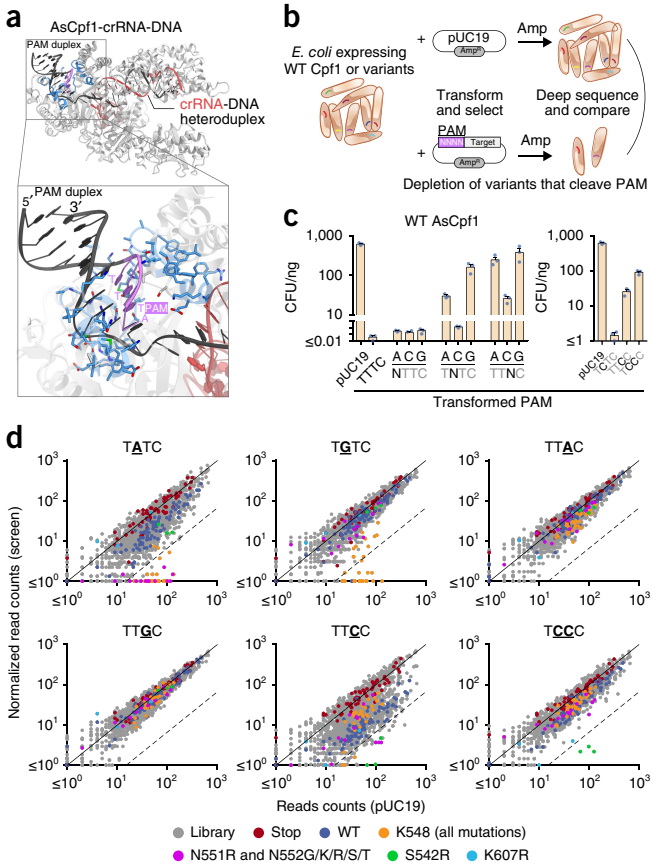


Figure 1 A bacterial interference-based negative selection screen identifies amino acid substitutions of AsCpf1 conferring activity at non-canonical PAMs. (**a**) Crystal structure of AsCpf1 (PDB ID: 5B43) in complex with crRNA and target DNA, highlighting the PAM nucleotides (magenta), and PAM-proximal residues selected for mutagenesis (blue). (**b**) Schematic of bacterial interference assay used to identify variants with altered PAM specificity. (**c**) Sensitivity of wild-type AsCpf1 to substitution mutations in the PAM as measured by bacterial interference. Error bars: mean ± s.e.m. of $n = 3$ plated transformations. (**d**) Scatter plots of screen readout, highlighting depleted variants. Each dot represents a distinct wild-type or mutant codon. The dashed lines indicate 15-fold depletion.

When transformed with NTTC and TCTC PAMs, *E. coli* expressing WT AsCpf1 had negligible survival on ampicillin media (**Fig. 1c**), indicating that these PAM sequences supported AsCpf1-mediated DNA plasmid cleavage and were not usable for screening the variant library. By contrast, the other five PAMs with a single mutation (TATC, TGTC, TTAC, TTCC, and TTGC) had notable survival rates. We subsequently screened the variant library for activity at these five PAMs, as well as an additional PAM with a double mutation (TCCC) (**Fig. 1d**).

Following deep sequencing readout, ~86% of the possible variants at the targeted residue positions were represented with at least 15 reads in the pUC19-transformed negative control to allow assessment of their depletion. For TATC, TGTC, TTCC, and TCCC PAMs, at least one AsCpf1 variant in the library was highly depleted (≥15-fold; **Fig. 1d** and **Supplementary Table 1b**). For TATC and TGTC, many of the depleted variants were at Lys548, a conserved residue that forms hydrogen bonds with the PAM duplex[12,15]. A number of hits were also observed for TTCC and TCCC, most notably an arginine substitution at Ser542, a non-conserved residue.

We evaluated whether variants identified in the screen had activity in HEK293T cells by targeting them to endogenous sites in two genes (*DNMT1* and *VEGFA*; **Fig. 2a** and **Supplementary Fig. 1a**). Most of the variants we tested generated indels at target sites with their corresponding PAMs; in particular, K548V was most active at a TATC target site, whereas S542R markedly increased activity for two TTCC target sites as well as a TCCC site. Combining the top single amino acid mutations into double and triple mutants further improved activity (**Fig. 2a** and **Supplementary Fig. 1b**). We selected the variants with the highest activity, S542R/K607R (hereafter referred to as RR) and S542R/K548V/N552R (hereafter referred to as RVR), for further investigation.

To assess the global PAM preference of the RR and RVR variants and compare them with WT AsCpf1, we adapted an *in vitro* PAM identification assay described previously (**Fig. 2b**)[1,16]. We incubated cell lysate from HEK293T cells expressing AsCpf1 (or an engineered variant) with *in vitro*-transcribed crRNA and a library of plasmid DNA containing a constant target preceded by a degenerate sequence (5′-NNNNNNNN-target). By amplifying and deep-sequencing the intact substrates and comparing them with the negative control, we determined which sequences were successfully cleaved. For each Cpf1 variant, nine reactions were carried out in parallel, each incubated for a different amount of time, in order to assess cleavage kinetics (**Supplementary Figs. 2** and **3**).

As expected, WT AsCpf1 was most active at TTTV PAMs (**Fig. 2c,d**) and had lower activity at TTTT, supporting the previously reported definition of the WT PAM as TTTV[1,6]. WT also cleaved other sequences including NTTV, TCTV, and TTCV at low rates, consistent with our observations in HEK293T cells (**Supplementary Fig. 4**) and in *E. coli*. By contrast, the RR and RVR variants had the highest activity at TYCV (where Y can be C or T) and TATV PAMs, respectively, compared to little or no activity for WT Cpf1 at those PAMs (**Fig. 2c,d**). The variant PAMs were also not as strictly defined as that of WT. The RR variant also cleaved ACCC and CCCC PAMs (and, to a lesser extent, VYCV), and the RVR variant also cleaved RATR PAMs (where R can be A or G).

To assess the robustness of the engineered PAM activity, we investigated the activity of the RR and RVR variants at their preferred PAMs (i.e., TYCV and TATV, respectively) across a diverse panel of endogenous target sites in HEK293T cells (**Fig. 2e** and **Supplementary Fig. 5**). The RR and RVR variants generated >50% indel for 14 out of 20 TYCV sites (70%) and 18 out of 23 TATV sites (78%), respectively, compared to little or no activity for WT AsCpf1 at most of these sites ($P < 0.0001$ for both variants; Wilcoxon signed-rank). By comparison, WT AsCpf1 achieved >50% indel for 8 out of 23 TTTV sites (35%). These data suggest that, at their respective preferred PAMs, the variants have comparable or slightly higher activity than the WT nuclease (**Fig. 2e**). The RR variant also exhibited substantial rates of editing in mouse Neuro2a cells (>20% indel for 6 out of 9 TYCV sites) (**Supplementary Fig. 6**).

Based on our observations that the RR variant also cleaves VYCV PAMs *in vitro*, albeit at a substantially lower rate than TYCV, we tested the activity of the RR variant at a separate panel of VYCV sites in HEK293T cells (**Supplementary Fig. 4**). Across the four genes assessed (*CFTR*, *DNMT1*, *EMX1*, and *VEGFA*), the RR variant achieved >20% indel for 24 out of 36 VYCV sites (67%), suggesting that, when necessary, target sites with VYCV PAMs can also be considered for editing with the RR variant.

To quantify how these Cpf1 PAM variants affect the targeting range of the CRISPR–Cpf1 system, we performed a computational analysis of the distribution of PAM sequences in the human genome (**Fig. 2f** and **Supplementary Fig. 7**), excluding Ns and masked repeats. When considering only the most active PAMs, the variants and WT collectively expand the targeting range of Cpf1 to one target site per ~11 bp in human coding sequences (corresponding to an approximately threefold increase relative to WT alone) and reduce the median distance to the

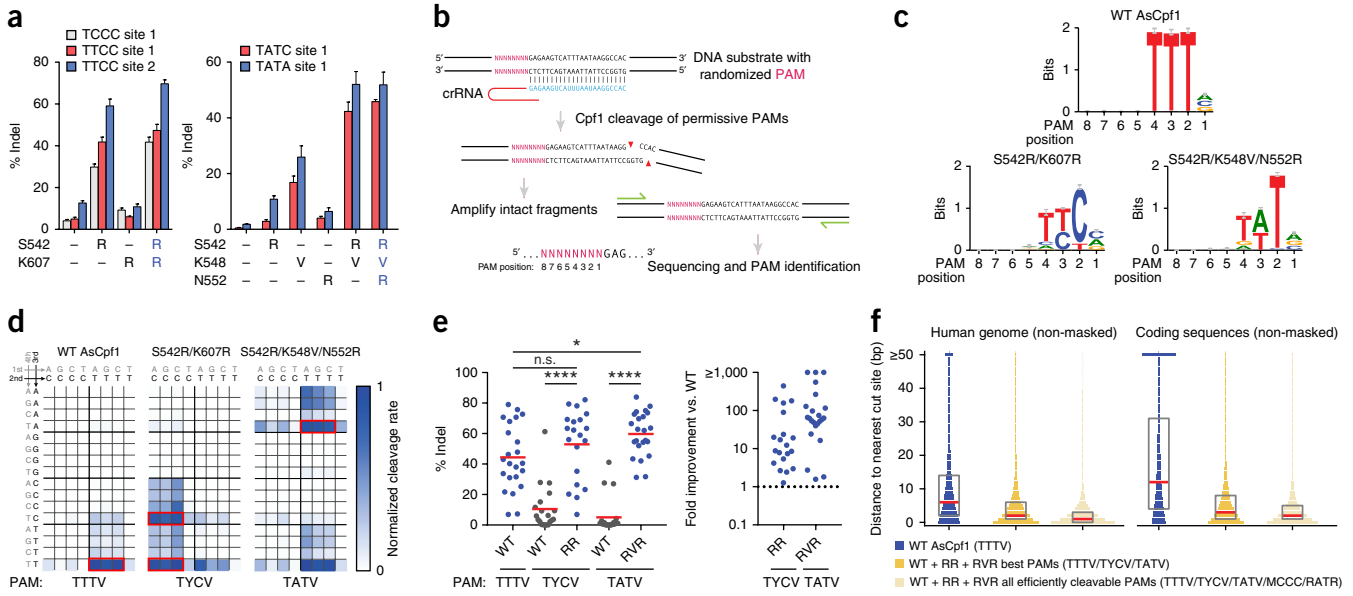


Figure 2 Construction and characterization of AsCpf1 variants with altered PAM specificities. (**a**) Combinatorial mutagenesis identifies AsCpf1 variants that cleave target sites with TYCV and TATV PAMs in HEK293T cells, where Y = C or T, and V = A, C, or G (see also **Supplementary Fig. 1**). Error bars: mean ± s.e.m. for $n = 4$ transfected cell cultures. (**b**) Schematic of *in vitro* cleavage assay used to determine global PAM specificity (see also **Supplementary Figs. 2** and **3**). (**c**) Web logos of the most rapidly cleaved PAMs for wild-type (WT), S542R/K607R (RR), and S542R/K548V/N552R (RVR) variants. (**d**) Normalized cleavage rates for all 4-base PAMs for WT and variants. NNRN PAMs are not shown due to negligible cleavage. The most active PAMs are boxed in red. (**e**) Comparison of the activity of WT, RR, and RVR at their preferred PAMs at a diverse panel of target sites in HEK293T cells (see also **Supplementary Fig. 5**). Each dot represents a target site. For indel percentages, each dot represents the mean of $n = 3$ transfected cell cultures, and the red lines indicate the overall means within each group. For fold change, each dot represents the ratio of the means of the corresponding indel replicates. n.s. $P > 0.05$ (Mann–Whitney); $^{*}P < 0.05$ (Mann–Whitney); $^{****}P < 0.0001$ (Wilcoxon signed-rank). (**f**) Targeting range of AsCpf1 variants in the human genome and in coding sequences (see also **Supplementary Fig. 7**). Plots show the probability mass function of the distance in base pairs to the nearest cleavage site. The boxplots indicate median and interquartile range. Genomic regions that contain Ns or masked repeats were ignored.

nearest cleavage site to 3 bp. Moreover, when considering a more broadly defined set of efficiently cleavable PAMs (in particular, the preferred PAMs plus MCCC and RATR, where M can be A or C), the targeting range is further expanded to one site per ~7 bp in human coding sequences, with a median distance to the nearest cleavage site of 2 bp.

We evaluated the genome-wide editing specificity of the RR and RVR variants using BLISS (double-strand breaks labeling *in situ* and sequencing), which quantifies DNA double-stranded breaks (DSBs) across the genome[7]. To fairly compare the variants to WT, we selected target sites bearing PAMs that can be reliably cleaved by all three nucleases; TTTV was the only PAM that met this criterion, although it has lower activity for the RR variant. For three of the four target sites evaluated (*VEGFA*, *GRIN2B*, and *DNMT1*), no off-target activity was detected from deep sequencing of the BLISS-identified loci (**Fig. 3a** and **Supplementary Table 2**) for any of the nucleases. For the fourth target site (*EMX1*), BLISS identified six off-target sites with detectable indels; all six sites had a TTCA PAM and no more than one mismatch in the first 19 nucleotides of the guide. As expected, both variants had increased activity at these off-target sites compared to WT, consistent with their increased ability to recognize TTCA PAMs. On the other hand, when targeting a site in the *RPL32P3* gene with known TTTV off-target sites[5], the variants exhibited similar or reduced off-target activity (**Fig. 3b**), which is also consistent with PAM preference. Collectively, these results indicate that the variants retain a high level of editing specificity that is comparable to WT AsCpf1. We note that a few of the off-target sites with low indel frequencies were not detected by BLISS at the time point we sampled, likely reflecting the dependence of BLISS on the timing of DSB formation[7].

We investigated whether the specificity of AsCpf1 can be improved by removing non-specific contacts between positively charged or polar residues and the target DNA, similar to strategies previously employed with *Streptococcus pyogenes* Cas9 (SpCas9)[17,18]. We identified K949A, which is located in the cleft of the protein that is hypothesized to interact with the non-target DNA strand, as a candidate (**Supplementary Fig. 8**). When combined with the RR and RVR variants, K949A reduced cleavage at all off-target sites assessed (**Fig. 3c**) while maintaining high levels of on-target activity (**Fig. 3d**).

Because Cpf1-family endonucleases have strong sequence and structural homology, the 542, 548, 552, and 607 positions in AsCpf1 have clear correspondences in other Cpf1 orthologs (**Supplementary Fig. 9** and **Supplementary Table 3**). Based on sequence alignment and the crystal structure[19], we hypothesized that LbCpf1 could also be engineered to recognize TYCV and TATV PAMs by introducing the mutations G532R/K595R and G532R/K538V/Y542R, respectively (**Supplementary Fig. 10a**). These mutations altered the PAM specificity of LbCpf1 in the predicted manner (**Supplementary Figs. 10b** and **11**), suggesting that this approach may be generally applicable across Cpf1 orthologs.

In summary, we have demonstrated that despite its evolutionary conservation, the PAM preference of Cpf1-family endonucleases can be altered by suitable mutations to residues close to the PAM duplex. Using a structure-guided mutagenesis screen, we engineered two variants, RR and RVR, which can robustly cleave target sites with TYCV and TATV PAMs, respectively, in mammalian cells. We extended this approach to similarly modify a second Cpf1 ortholog. Finally, we introduced an additional mutation that enhanced Cpf1 specificity. Collectively, these engineered variants increase the targeting range of Cpf1 to one cleavage site for every ~11 bp in human coding sequences and provide useful additions to the CRISPR–Cas genome engineering toolbox.

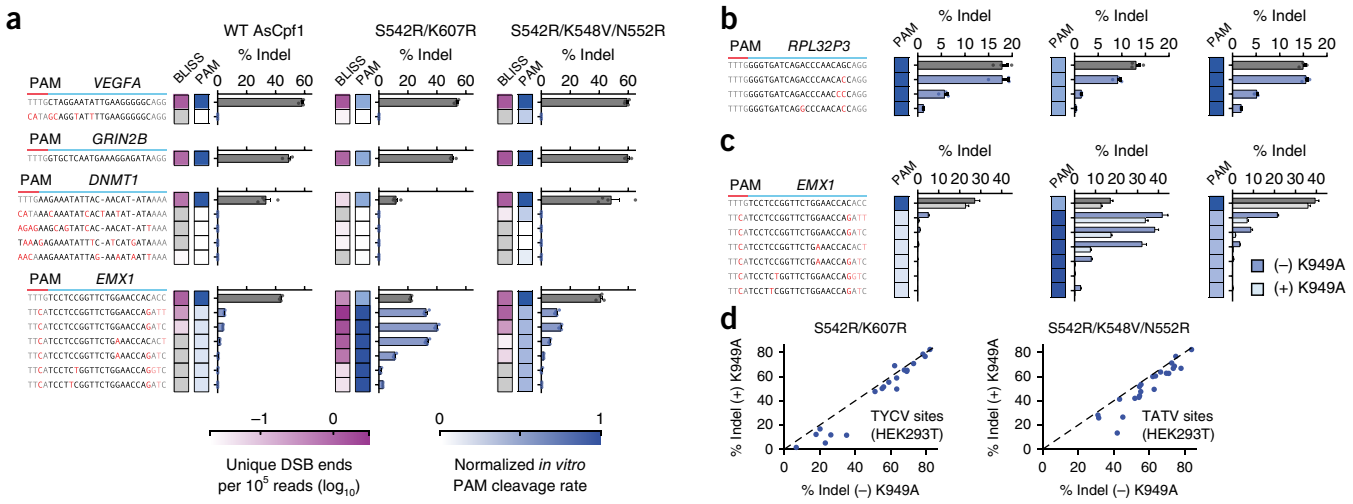


Figure 3 Specificity of AsCpf1 PAM variants. (**a**) DNA double-strand breaks labeling *in situ* and sequencing (BLISS) for four target sites (*VEGFA*, *GRIN2B*, *EMX1*, and *DNMT1*) in HEK293T cells. The $\log_{10}$ number of unique double-strand break (DSB) ends per 10^5 reads is indicated by the magenta heat map. The normalized PAM cleavage rates from the *in vitro* cleavage assay in **Figure 2d** are indicated by the blue heat map. Each BLISS-identified cleavage site was independently assessed for indel formation (bar graphs). Bars show mean ± s.e.m. for $n = 4$ transfected cell cultures. Mismatches in bases 21–23 of the target are grayed as they have minimal impact on cleavage efficiency[4,5]. (**b**) Evaluation of an additional target site in the *RPL32P3* gene with known TTTV off-target sites[5]. (**c**) Addition of a K949A mutation improves the specificity of WT AsCpf1 and variants (see also **Supplementary Fig. 8**). For **b** and **c**, bars show mean ± s.e.m. for $n = 3$ transfected cell cultures. (**d**) On-target efficiency of the RR and RVR variants ± K949A. Each dot represents a distinct target site (mean of $n = 3$ transfected cell cultures).

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank A. Magnell for experimental assistance; R. Macrae for a critical review of the manuscript; and the entire Zhang laboratory for support and advice. D.B.T.C. is supported by T32GM007753 from the National Institute of General Medical Sciences. W.X.Y. is supported by T32GM007753 from the National Institute of General Medical Sciences and a Paul and Daisy Soros Fellowship. J.C.M. is supported by the NIH (training grant 2 T32 GM 7287-41). H.N. is supported by JST, PRESTO (JPMJPR13L8), JSPS KAKENHI (Grant Numbers 26291010 and 15H01463). O.N. is supported by the Basic Science and Platform Technology Program for Innovative Biological Medicine from the Japan Agency for Medical Research and Development, AMED, and the Council for Science, and Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology. N.C. is supported by the Karolinska Institutet, the Swedish Research Council (521-2014-2866), the Swedish Cancer Research Foundation (CAN 2015/585), and the Ragnar Söderberg Foundation. F.Z. is a New York Stem Cell Foundation–Robertson Investigator. F.Z. is supported by the NIH through NIMH (5DP1-MH100706 and 1R01-MH110049), NSF, Howard Hughes Medical Institute, the New York Stem Cell, Simons, Paul G. Allen Family, and Vallee Foundations; and James and Patricia Poitras, Robert Metcalfe, and David Cheng.

AUTHOR CONTRIBUTIONS

L.G., D.B.T.C., and F.Z. conceived this study. L.G. and D.B.T.C. performed experiments with help from all authors. J.C.M. contributed to the bacterial selection screen. M.W.S. processed BLISS samples, and W.X.Y. analyzed BLISS data. T.Y., H.N., and O.N. provided unpublished AsCpf1 crystal structure information. N.C. provided an unpublished BLISS protocol. F.Z. supervised research. L.G. and F.Z. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
2. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR–associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).
3. Zetsche, B. *et al.* Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
4. Kim, D. *et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
5. Kleinstiver, B.P. *et al.* Genome-wide specificities of CRISPR–Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34**, 869–874 (2016).
6. Kim, H.K. *et al. In vivo* high-throughput profiling of CRISPR–Cpf1 activity. *Nat. Methods* **14**, 153–159 (2017).
7. Yan, W.X. *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.* **8**, 15058 (2017).
8. Kleinstiver, B.P. *et al.* Engineered CRISPR–Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
9. Kleinstiver, B.P. *et al.* Broadening the targeting range of *Staphylococcus aureus* CRISPR–Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).
10. Hirano, S., Nishimasu, H., Ishitani, R. & Nureki, O. Structural basis for the altered PAM specificities of engineered CRISPR–Cas9. *Mol. Cell* **61**, 886–894 (2016).
11. Anders, C., Bargsten, K. & Jinek, M. Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. *Mol. Cell* **61**, 895–902 (2016).
12. Yamano, T. *et al.* Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
13. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR–Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
14. Esvelt, K.M. *et al.* Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
15. Gao, P., Yang, H., Rajashankar, K.R., Huang, Z. & Patel, D.J. Type V CRISPR–Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.* **26**, 901–913 (2016).
16. Ran, F.A. *et al. In vivo* genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
17. Slaymaker, I.M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
18. Kleinstiver, B.P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
19. Dong, D. *et al.* The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).

ONLINE METHODS

Library construction. Human codon-optimized AsCpf1 driven by a T7 promoter was cloned into a modified pACYC backbone, and unique restriction sites were introduced flanking the selected PAM-proximal AsCpf1 residues via suitable silent mutations. For each residue, a mutagenic insert was synthesized as short complementary oligonucleotides (Integrated DNA Technologies), with the mutated codon replaced by a degenerate NNK mixture of bases (where K can be G or T). Each degenerate codon position was also barcoded by creating a unique combination of silent mutations in neighboring codons in order to correct for sequencing errors during screen readout. The variant library was assembled by cassette mutagenesis, mini-prepped, pooled, and precipitated with isopropanol.

***E. coli* negative-selection screen.** NovaBlue(DE3) *E. coli* (Novagen) cells were transformed with the variant library and plated on LB agar (Affymetrix) containing 25 μg/mL chloramphenicol. Surviving colonies were scraped and cultured in ZymoBroth with 25 μg/mL chloramphenicol to an O.D. of 0.4–0.6 and made competent using a Mix & Go kit (Zymo). For each mutant PAM screened, the competent *E. coli* pool was transformed with 100-ng target plasmid containing the mutant PAM, incubated on ice for 15–30 min, heat shocked at 42 °C for 30s, and plated on LB agar containing 100 μg/mL ampicillin and 25 μg/mL chloramphenicol in the absence of IPTG. A negative control was obtained by transforming the *E. coli* with pUC19, which lacks the target site. Plasmid DNA from surviving colonies was isolated by midi-prep (Qiagen). The regions containing mutations were amplified with custom primers containing Illumina adaptors and paired-end sequenced with a 600-cycle MiSeq kit (Illumina). Reads were filtered by requiring perfect matches to silent codon barcodes; a Phred quality (Q score) of at least 30 for each of the three NNK bases; and consistency between forward and reverse reads, when applicable. The read count for each variant was normalized assuming that the mean abundance of TAG (stop) codons was equivalent to the negative control.

***In vitro* PAM identification assay.** Plasmids encoding the AsCpf1 variants were transfected into HEK293T cells as described below. Cell lysate was prepared with lysis buffer (20 mM HEPES, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 0.1% Triton X-100) supplemented with ETDA-free cOmplete Protease Inhibitor Cocktail (Roche). crRNA was transcribed *in vitro* using custom oligonucleotides and HiScribe T7 *in vitro* Transcription Kit (NEB) following the manufacturer's recommended protocol. For the PAM library, a degenerate 8-bp sequence preceding a 33-bp target site[1] was cloned into the multiple cloning site in pUC19, and the library was digested with AatII and LguI and gel purified before use. Each *in vitro* cleavage reaction consisted of 1 μL 10× CutSmart buffer (NEB), 25 ng PAM library, 250 ng *in vitro*-transcribed crRNA, 0.5 μL cell lysate, and water for a total volume of 10 μL. Reactions were incubated at 37 °C and quenched by adding 50 μL Buffer PB (Qiagen) followed by column purification. Purified DNA was amplified with two rounds of PCR over 29 total cycles using custom primers containing Illumina adaptors and sequenced with a 75-cycle NextSeq kit (Illumina). For each Cpf1 variant, separate *in vitro* cleavage reactions were carried out for 1.15 min, 4 min, 10 min, 15 min, 20 min, 30 min, 40 min, 90 min, and 175 min. The unmodified library of degenerate sequences was used as the 0 min time point. A negative control, using lysate from unmodified HEK293T cells, was taken at 10 min.

Computational analysis of PAM cleavage kinetics. See also **Supplementary Figures 2** and **3**. Sequencing reads were filtered by Phred quality (≥30 for all of the eight degenerate PAM bases). For each cleavage reaction, a depletion ratio for each of the 4^8 PAM sequences was calculated as (normalized read count in cleavage reaction)/(normalized read count in negative control). Each depletion ratio was then divided by the median depletion ratio of all NNNNVRRT sequences, which were not cleaved by WT AsCpf1 or either of the variants. The depletion ratios of each PAM sequence (4^8 total) across time points were fit using nonlinear least-squares to an exponential decay model $x(t) = c_0 + ce^{-kt}$, where $x(t)$ is the depletion ratio at time t, and the terms $c_0 \leq 0.2$, c, and k (the rate constant in min^{-1}) are parameters. For each variant, the estimated cleavage rate k of each 4-base PAM was computed as the median cleavage rate of the 256 8-base sequences corresponding to that PAM; for instance, the cleavage rate of TTTA was computed as the median cleavage rate of the 256 sequences of the form NNNNTTTA. Finally, all cleavage rates were adjusted so that the highest rate of any 4-base PAM was equal to 1 for each variant.

Cell culture and transfection. Human embryonic kidney 293 and Neuro2a cell lines were maintained in Dulbecco's modified Eagle's medium supplemented with 10% FBS (Gibco) at 37 °C with 5% CO_2 incubation. Cells were seeded one day before transfection in 24- or 96-well plates (Corning) at a density of approximately 1.2×10^5 cells per 24-well or 2.4×10^4 cells per 96-well and transfected at 50–80% confluency using Lipofectamine 2000 (Invitrogen), according to the manufacturer's recommended protocol. For cell lysates, 500 ng of Cpf1 plasmid was delivered per 24-well. For indel analysis in HEK293T cells, 400 ng of Cpf1 plasmid plus 100 ng crRNA plasmid was delivered per 24-well, or 100 ng Cpf1 plus 50 ng crRNA plasmid per 96-well. For BLISS and for indel analysis in Neuro2a cells, 500 ng of a plasmid encoding both Cpf1 and crRNA were delivered per 24-well. All indel and BLISS experiments used a guide length of 23 nucleotides.

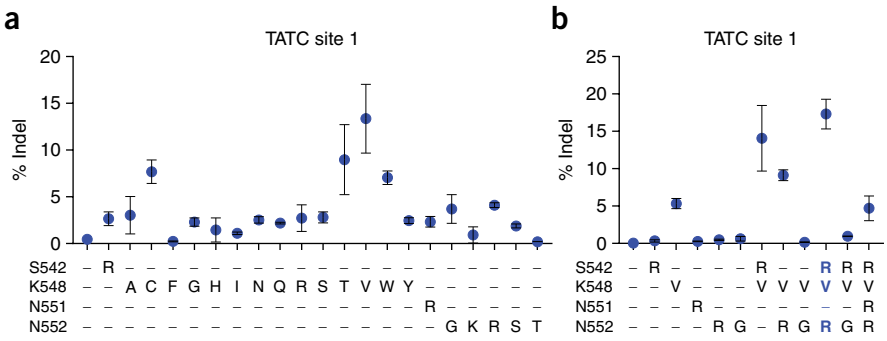
Indel quantification. All indel frequencies were quantified by targeted deep sequencing (Illumina). For indel library preparation, cells were harvested approximately 3 days after transfection, and genomic DNA was extracted using QuickExtract DNA extraction solution (Epicentre) by resuspending pelleted cells in QuickExtract (80 μL per 24-well, or 20 μL per 96-well), followed by incubation at 65 °C for 15 min, 68 °C for 15 min, and 98 °C for 10 min. Amplicons for deep sequencing were generated using two rounds of PCR to attach Illumina handles. Indels were counted computationally by searching each sequencing read for exact matches with strings delineating the ends of a 50- to 70-bp window around the cut site. The distance in bp between these strings was then compared to the corresponding distance in the reference genome, and the read was counted as an indel if the two distances differed. For each sample, the indel frequency was determined as (number of reads with an indel)/(number of total reads). Samples with fewer than 1,000 total reads were excluded. Where negative control data are not shown, indel percentages represent background-subtracted maximum likelihood estimates. In particular, for a sample with R total reads, of which $n \leq R$ are indels, and false-positive rate $0 \leq \alpha < 1$ (as determined by the negative control), the true indel rate was estimated as $\max\{0, [(n/R) - \alpha]/(1 - \alpha)\}$.

Computational analysis of Cpf1 targeting range. The complete GRCh38 human genome assembly and coding sequences, with repeats and low complexity regions masked, were downloaded from Ensembl and analyzed as described in **Supplementary Figure 7**.

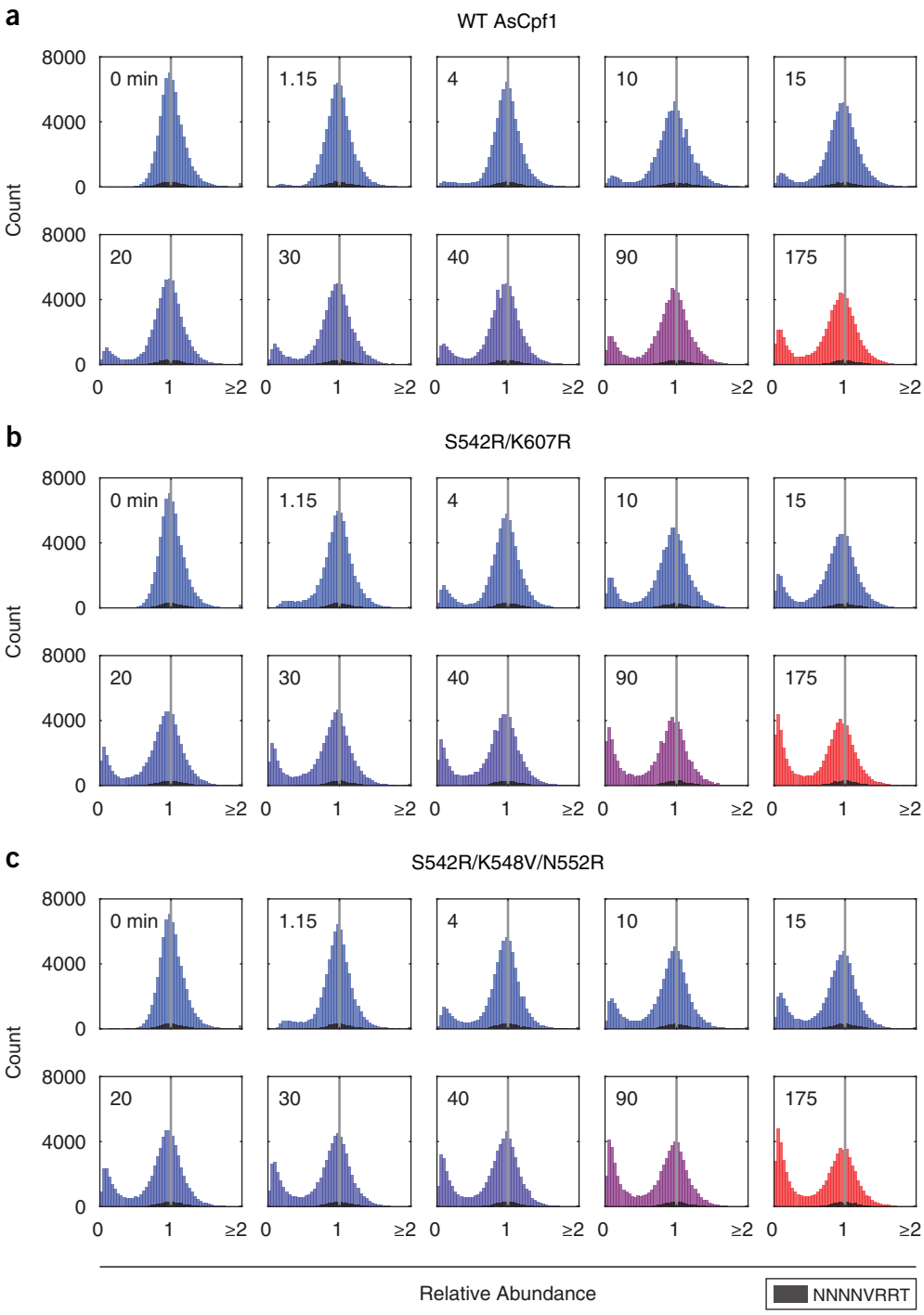
BLISS. All BLISS experiments and analyses were performed as previously described[7]. The data analysis for the staggered cut sites of Cpf1 was slightly modified from prior analysis[7,16] to increase sensitivity. Previously, to distinguish bona fide nuclease-induced events from the background DSBs in DSB hotspots, centromeres, and telomeres, we had used a cutoff based on the fraction of the pairwise reads that overlapped less than −6 bp. This cutoff was set at 0.95 based on empirical data from Cas9 off-target analysis, but to accommodate the variation produced by the staggered cut sites of Cpf1, we found that greater sensitivity to Cpf1 off-target sites could be obtained by relaxing this cutoff to 0.85. All other analyses, such as the guide homology score calculations, were as described[7].

Plasmids and guide sequences. A list of the plasmids and guide sequences used in this study can be found in **Supplementary Table 4**.

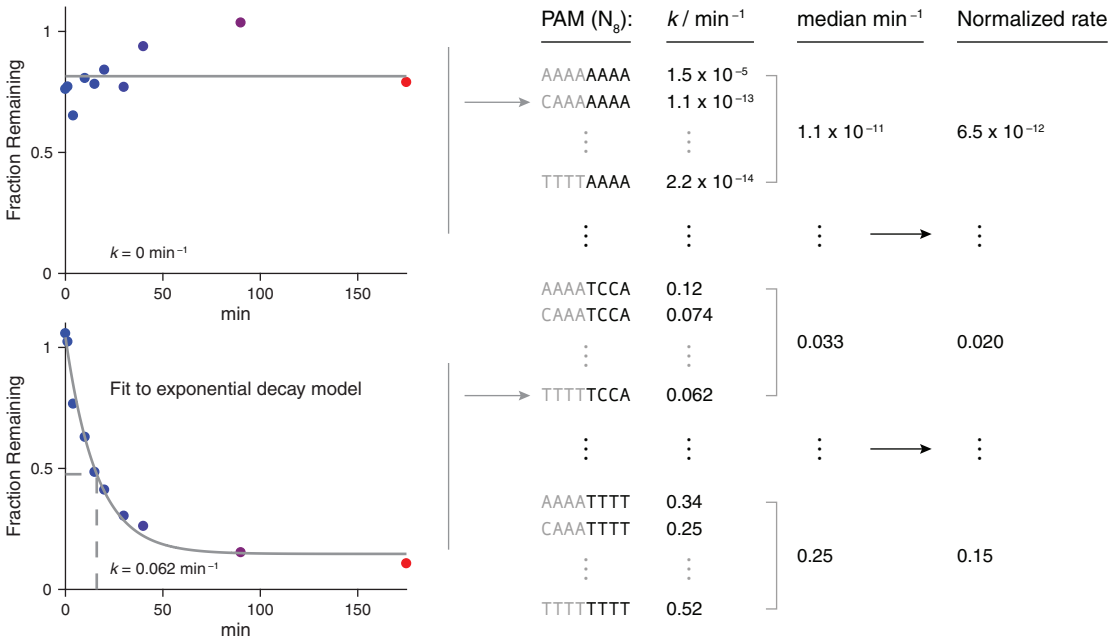
Data availability. Deep sequencing data are available on the NCBI Sequence Read Archive SRR5611789. Reagents and further information will be available to the academic community through Addgene and the Zhang laboratory website (http://www.genome-engineering.org/).



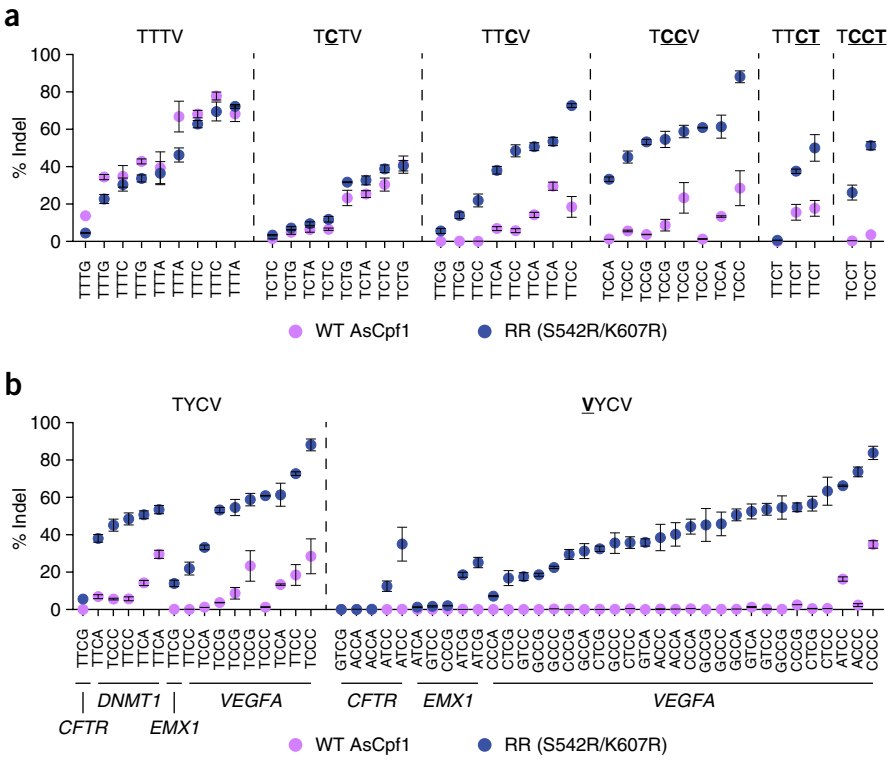
Supplementary Figure 1 Related to **Figure 2a**. Evaluation of (**a**) single amino acid mutations and (**b**) combination mutants to construct the AsCpf1 RVR variant, which is active at target sites with TATV PAMs. Dots show mean ± s.e.m. (n = 2).



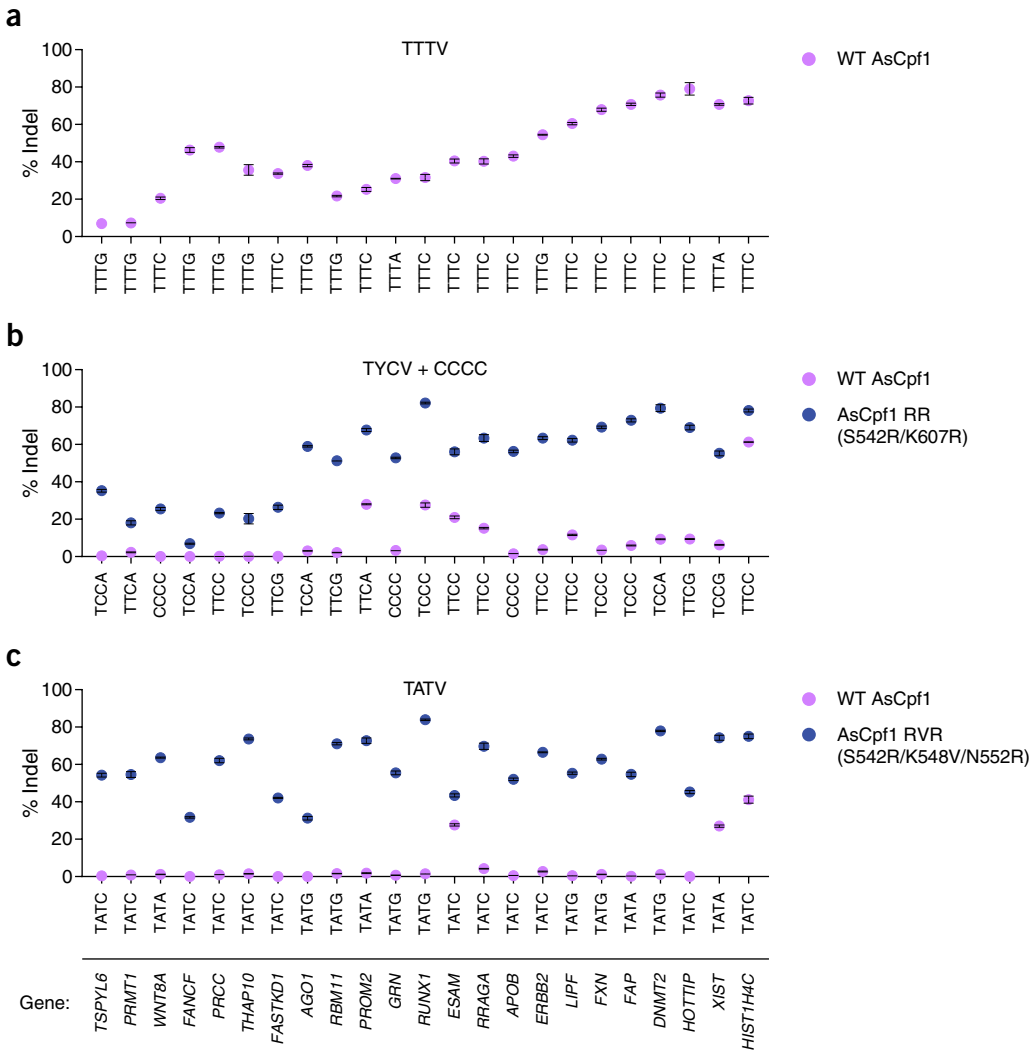
Supplementary Figure 2 Related to **Figure 2b-d**. Histograms of abundances of 4^8 PAMs (NNNNNNNN) at each *in vitro* cleavage time point for (**a**) WT AsCpf1, (**b**) S542R/K607R, and (**c**) S542R/K548V/N552R. The color of each histogram represents elapsed time. NNNNVRRT sequences, which were used to center the histograms, are shown in black.



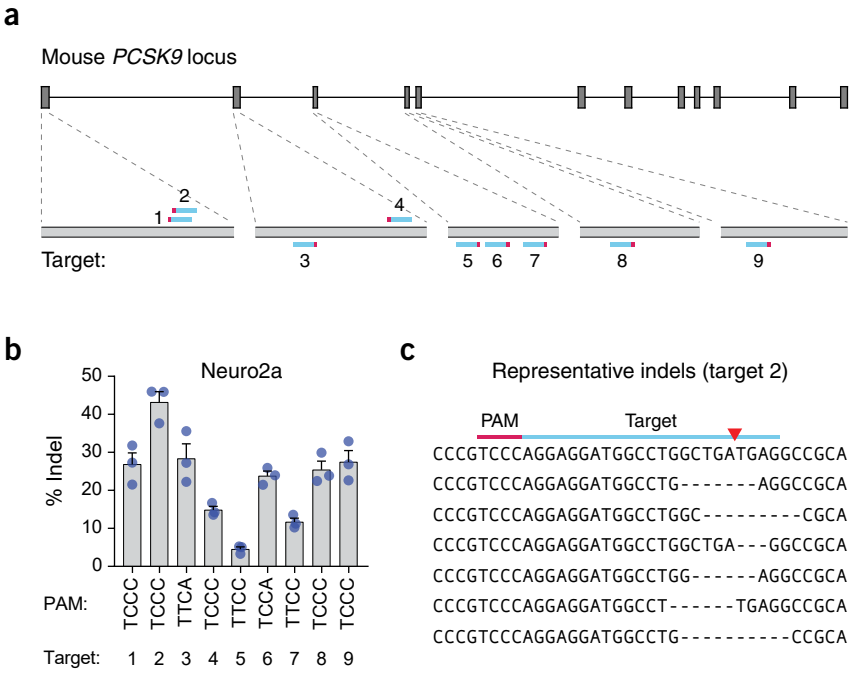
Supplementary Figure 3 Data processing pipeline for the *in vitro* cleavage assay used for **Figure 2d**.



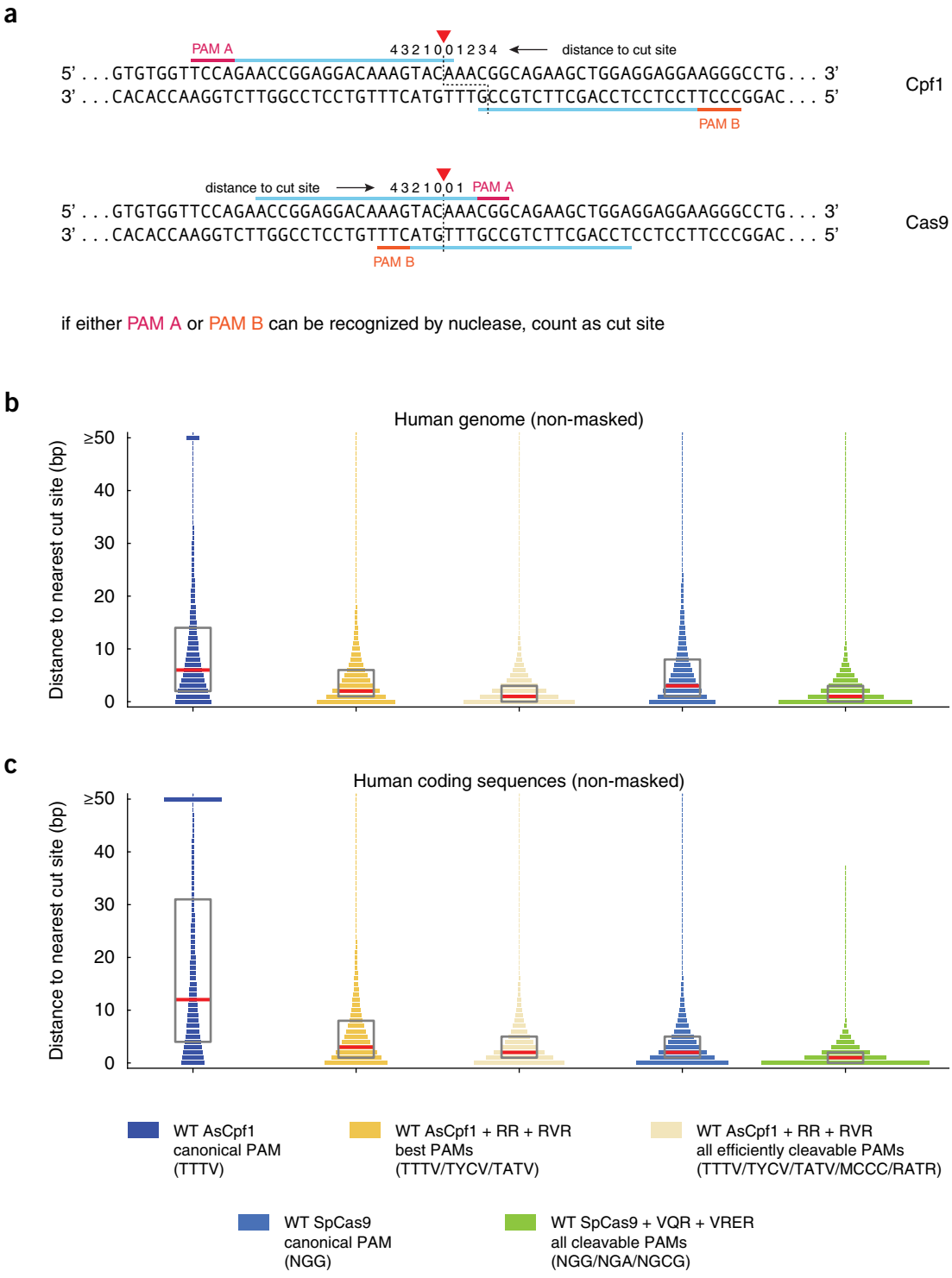
Supplementary Figure 4 (**a**) Comparison of the activity of WT AsCpf1 to the RR variant at target sites with cytosine-containing PAMs. (**b**) Activity of the RR variant at TYCV and VYCV sites (V = A, C, or G), demonstrating that the presence of a 5' T in the PAM sequence is not always required (*i.e.*, some NYCV PAMs can be recognized). The data for TYCV sites is the same as that shown in (**a**). All indel percentages were measured in HEK293T cells. Dots show mean ± s.e.m. (n = 2-3).



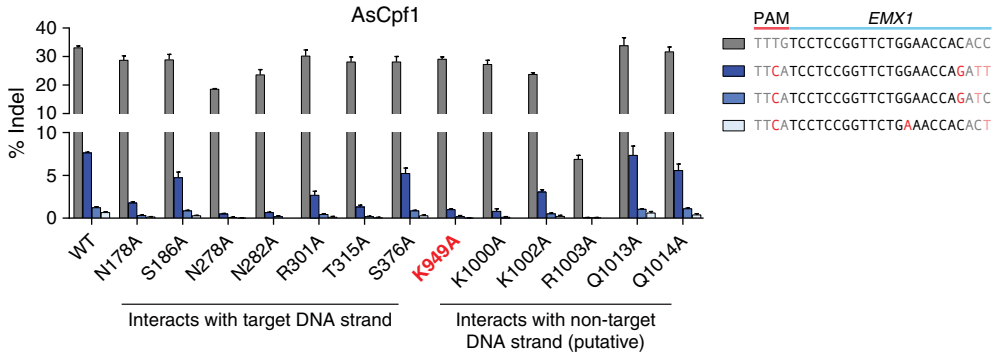
Supplementary Figure 5 Related to **Figure 2e**. Activity of (**a**) WT AsCpf1, (**b**) the RR variant, and (**c**) the RVR variant at target sites with highly active PAMs in HEK293T cells. Dots show mean ± s.e.m. ($n = 3$). **Figure 2e** shows these data in aggregate. For the AsCpf1 RR variant, the three CCCC sites are not included in **Figure 2e**.



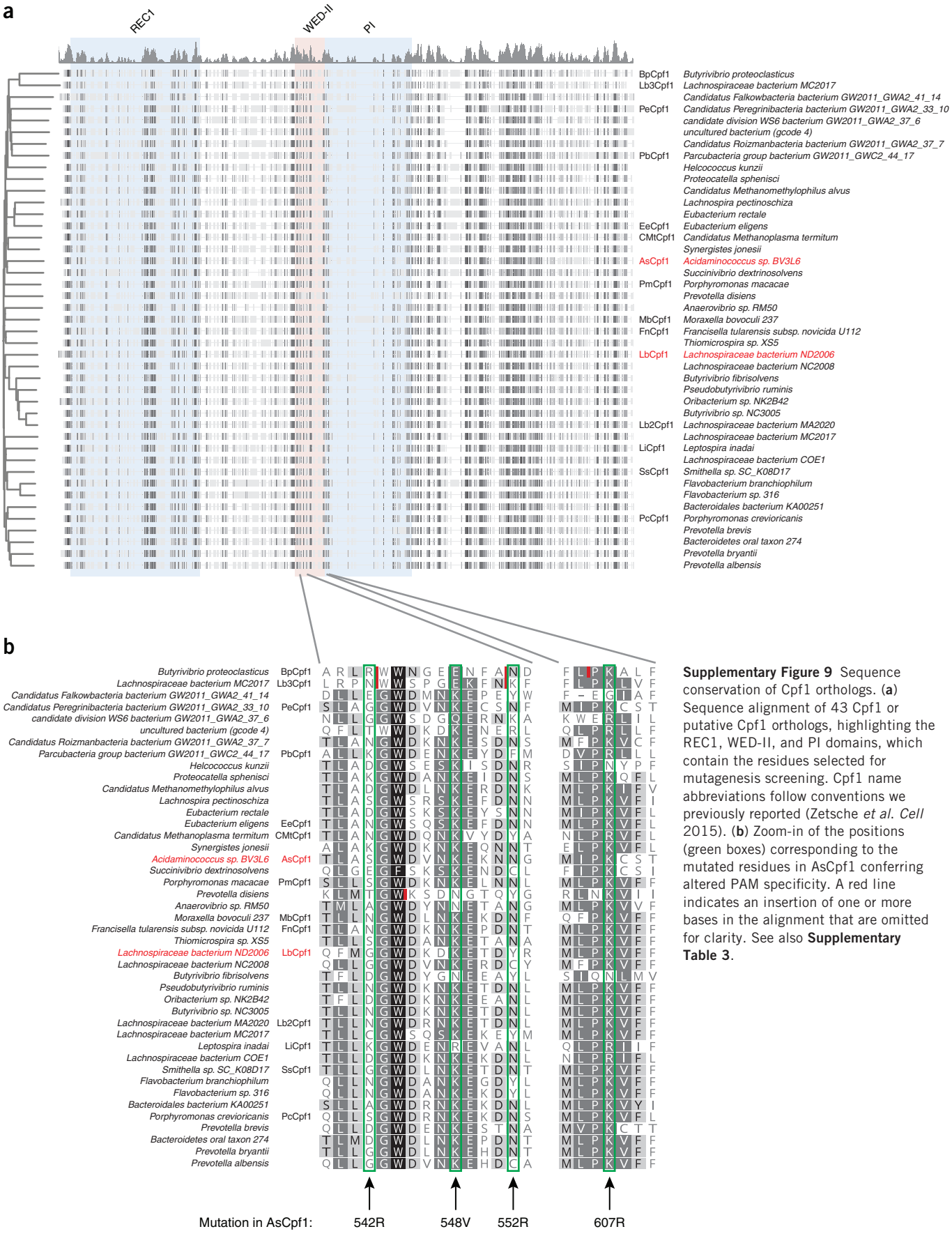
Supplementary Figure 6 Editing efficiency of the AsCpf1 RR variant at TYCV sites in mouse Neuro2a cells. (**a**) Diagram of the mouse *PCSK9* locus. Gray boxes represent coding sequences. (**b**) Indel percentages produced by the RR variant at *PCSK9* target sites with TYCV PAMs. Bars show mean ± s.e.m. ($n = 3$). (**c**) Representative indels at the target site (#2) with the highest editing efficiency. The red triangle represents the putative cleavage site on the top strand.



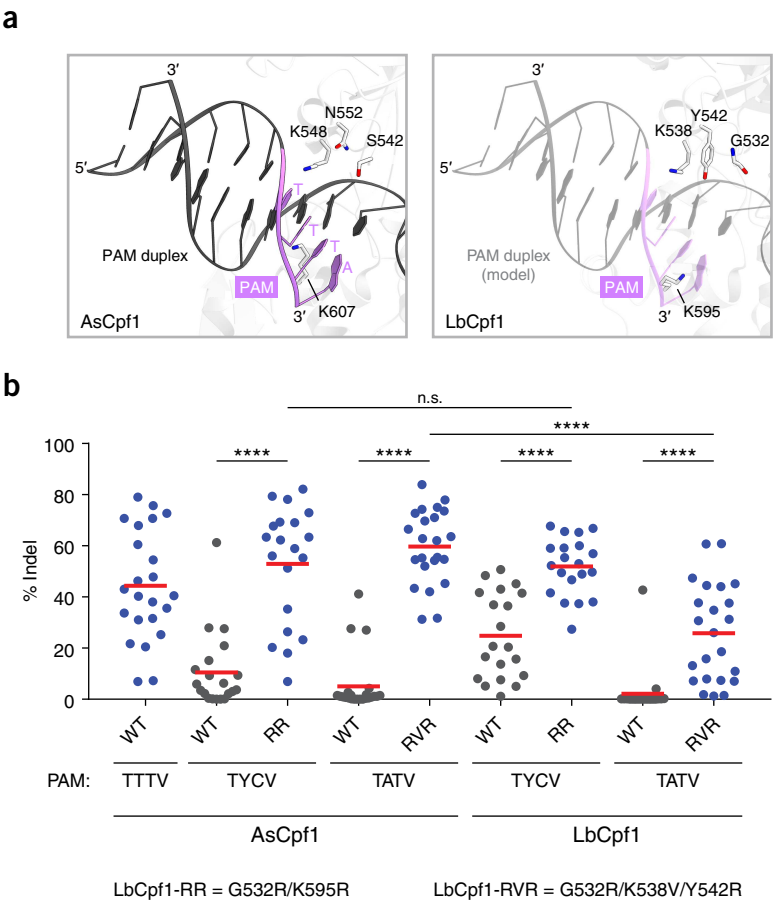
Supplementary Figure 7 Related to **Figure 2f**. (**a**) Definition of targeting range for Cpf1 and Cas9. Comparison of the targeting range of Cpf1 (+RR and RVR variants) to Cas9 (+VQR and VRER variants) in (**b**) the human genome and (**c**) coding sequences. Plots show the probability mass function of the distance (in base pairs) to the nearest cleavage site. The boxplots indicate median and interquartile range. Genomic regions that contain Ns or masked repeats were ignored in this analysis.



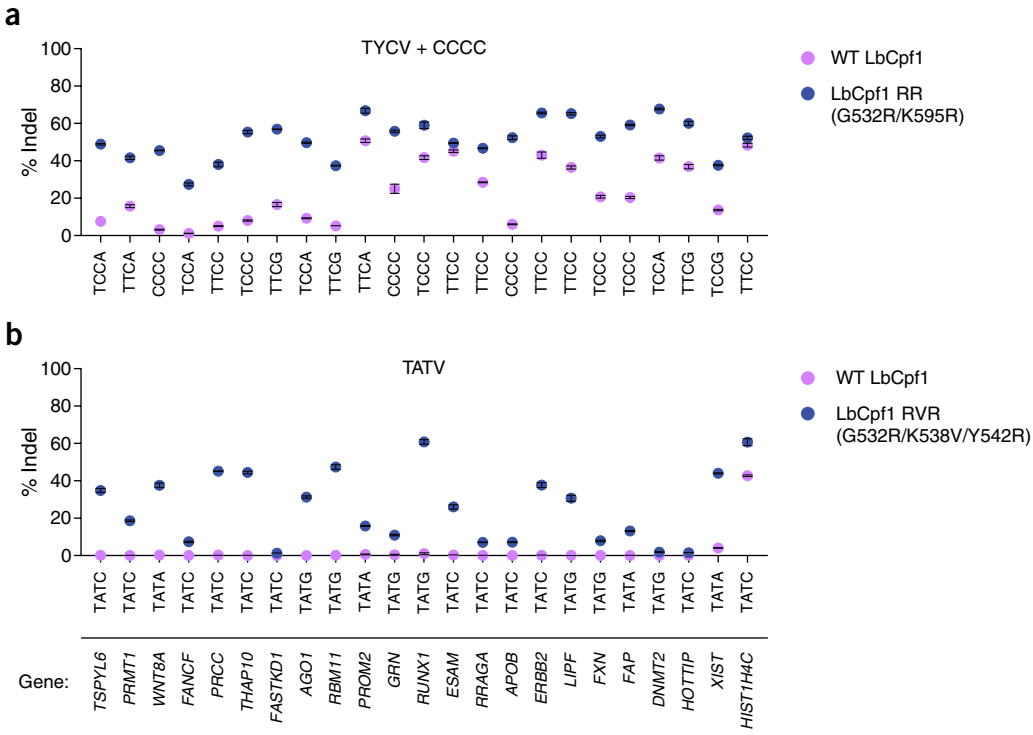
Supplementary Figure 8 Related to **Figure 3c**. Specificity mutagenesis of AsCpf1. An alanine scan of residues with interactions or putative interactions with the DNA strands. Bars show mean ± s.e.m. (n = 2-3). K949A was selected as a candidate for enhancing the specificity of AsCpf1. Lys949 is part of the bridge helix.



Supplementary Figure 9 Sequence conservation of Cpf1 orthologs. (**a**) Sequence alignment of 43 Cpf1 or putative Cpf1 orthologs, highlighting the REC1, WED-II, and PI domains, which contain the residues selected for mutagenesis screening. Cpf1 name abbreviations follow conventions we previously reported (Zetsche *et al. Cell* 2015). (**b**) Zoom-in of the positions (green boxes) corresponding to the mutated residues in AsCpf1 conferring altered PAM specificity. A red line indicates an insertion of one or more bases in the alignment that are omitted for clarity. See also **Supplementary Table 3**.



Supplementary Figure 10 Engineering the PAM recognition of LbCpf1. (**a**) Crystal structures of AsCpf1 (PBD ID: 5B43) and LbCpf1 (PDB ID: 5ID6), highlighting the corresponding residues mutated to alter PAM specificity. The PAM duplex shown for LbCpf1 is a model. (**b**) Activity of LbCpf1 G532R/K595R and G532R/K538V/Y542R at TYCV and TATV sites, respectively, in HEK293T cells. Each point represents the mean of three replicates, and the red lines indicate the overall means within each group. The data for AsCpf1 also appears in **Figure 2e**. n.s. $p > 0.05$ (Mann-Whitney); ****$p < 0.0001$ (Wilcoxon signed-rank).



Supplementary Figure 11 Related to **Supplementary Figure 10b**. Activity of the (**a**) LbCpf1 RR variant and (**b**) LbCpf1 RVR variant at target sites with preferred PAMs in HEK293T cells. Dots show mean ± s.e.m. (n = 3). **Supplementary Figure 10b** shows these data in aggregate. The target sites are the same as those shown in **Supplementary Figure 5b-c**. For the RR variant, the three CCCC sites are not included in **Supplementary Figure 10b**.

Supplementary Table 1a List of residue positions evaluated in mutagenesis screen

Position	Residue	Domain	Variants (/19)	Position	Residue	Domain	Variants (/19)
130	K	REC1	19	546	V	WED-II	19
131	G	REC1	18	547	N	WED-II	14
132	L	REC1	17	548	K	WED-II	19
133	F	REC1	19	550	K	WED-II	15
134	K	REC1	17	551	N	WED-II	15
135	A	REC1	18	552	N	WED-II	16
162	F	REC1	19	570	K	WED-II	19
163	D	REC1	13	571	Q	WED-II	19
164	K	REC1	14	572	K	WED-II	19
165	F	REC1	18	573	G	WED-II	19
166	T	REC1	19	595	Y	WED-II	16
168	Y	REC1	12	596	D	WED-II	19
169	F	REC1	17	597	Y	WED-II	15
171	G	REC1	10	599	P	PI	15
172	F	REC1	17	600	D	PI	12
173	Y	REC1	18	601	A	PI	17
174	E	REC1	15	602	A	PI	16
175	N	REC1	17	603	K	PI	16
176	R	REC1	19	604	M	PI	18
177	K	REC1	19	605	I	PI	16
536	Q	WED-II	15	606	P	PI	10
537	M	WED-II	14	607	K	PI	13
538	P	WED-II	19	608	C	PI	15
539	T	WED-II	18	609	S	PI	12
540	L	WED-II	18	610	T	PI	17
541	A	WED-II	18	611	Q	PI	12
542	S	WED-II	19	612	L	PI	11
543	G	WED-II	19	613	K	PI	18
544	W	WED-II	17	614	A	PI	18
545	D	WED-II	17	615	V	PI	13

Supplementary Table 1b List of variants depleted at least 15-fold relative to pUC19

TATC	
K164N	K548E
Y168V	K548A*
G171N	K548V*
G171M	K548G*
G171S	K548Y*
G171L	K548F*
E174L	K548C*
T539R	K548W*
L540F	N551R*
K548N*	N551Y
K548T*	N552G*
K548I*	N552K*
K548S*	N552R*
K548R*	N552S*
K548H*	N552T*
K548Q*	N552Q
K548P	A601C
K548L	A614R

TGTC
A135F
F169Y
F172A
F172I
K177L
K548T
K548M
K548S
K548R
K548H
K548Q
K548A
K548V
K548G
K548Y
K548C
K548W

TTCC	
K130Q	Q571S
K130L	Q571P
G131D	Q571A
G131F	Q571G
K134C	Q571C
F169W	K572P
F169Y	K572R
G171L	G573I
G171Y	G573L
K177R	G573E
Q536R	Y595L
M537R	D596N
S542L	D596S
S542R*	D596A
K548R	D596C
K550T	P599G
K550H	D600M
K550P	A602K
K550R	A602Y
N551I	K607H
N551S	S609G
N551R	A614Q
Q571T	A614L

TCCC
S542R*

* Evaluated in HEK293T cells in this study.

Supplementary Table 2 BLISS data

	WT AsCpf1			S542R/K607R			S542R/K548V/N552R				
Target	**BLISS**	**PAM**	**% Indel**	**BLISS**	**PAM**	**% Indel**	**BLISS**	**PAM**	**% Indel**	**PAM**	**Sequence**
VEGFA	0.833	0.88	58 (1.5)	0.988	0.42	54 (1.3)	0.895	0.89	59 (1.6)	TTTG	CTAGGAATATTGAAGGGGGCAGG
OT1	ND	0.00	<0.1	0.025	0.00	<0.1	ND	0.24	<0.1	CATA	GCAGGTATTTTGAAGGGGGCAGG
GRIN2B	0.590	0.88	49 (2.7)	0.664	0.42	51 (1.8)	0.969	0.89	60 (2.1)	TTTG	GTGCTCAATGAAAGGAGATAAGG
DNMT1	0.467	0.88	33 (5.9)	0.042	0.42	12 (2.4)	0.809	0.89	48 (12)	TTTG	AAGAAATATTACAACATATAAAA
OT1	ND	0.00	<0.1	ND	0.00	<0.1	0.030	0.24	<0.1	CATA	AACAAATATCACTAATATATAAAA
OT2	ND	0.00	<0.1	0.030	0.00	<0.1	ND	0.00	<0.1	AGAG	AAGCAGTATCACAACATATTAAA
OT3	ND	0.00	<0.1	0.025	0.00	<0.1	ND	0.00	<0.1	TAAA	GAGAAATATTTCATCATGATAAAA
OT4	0.022	0.00	<0.1	ND	0.02	<0.1	ND	0.10	<0.1	AACA	AAGAAATATTAGAAAATAATTAAA
EMX1	0.738	0.88	44 (0.8)	0.263	0.42	22 (0.6)	0.549	0.89	41 (2.4)	TTTG	TCCTCCGGTTCTGGAACCACACC
OT1	0.187	0.14	4.8 (0.2)	2.958	0.95	32 (1.5)	0.704	0.34	11 (1.3)	TTCA	TCCTCCGGTTCTGGAACCAGATT
OT2	0.048	0.14	3.7 (0.2)	1.173	0.95	40 (1.2)	0.189	0.34	13 (0.9)	TTCA	TCCTCCGGTTCTGGAACCAGATC
OT3	ND	0.14	1.5 (0.2)	0.794	0.95	33 (1.3)	0.026	0.34	6.0 (0.5)	TTCA	TCCTCCGGTTCTGAAACCACACT
OT4	ND	0.14	0.4 (0.1)	0.419	0.95	11 (0.9)	0.045	0.34	1.7 (0.1)	TTCA	TCCTCCGGTTCTGAAACCAGATC
OT5	ND	0.14	0.1 (0.2)	0.040	0.95	1.3 (0.4)	ND	0.34	0.3 (0.4)	TTCA	TCCTCTGGTTCTGGAACCAGGTC
OT6	ND	0.14	<0.1	0.036	0.95	2.8 (0.1)	ND	0.34	0.3 (0.0)	TTCA	TCCTTCGGTTCTGGAACCAGATC

BLISS data represent unique DSB ends per 10^5 reads (ND: not detected). PAM values are normalized cleavage rates from the *in vitro* PAM identification assay shown in **Figure 2d**. Indel values represent mean (std. dev.) of $n = 4$ replicates.

Supplementary Table 3 List of selected Cpf1 orthologs and their predicted PAM-altering mutations

#	Accession No.	Organism	Name	"542"	"548"	"552"	"607"
1	WP_013282991	*Butyrivibrio proteoclasticus*	BpCpf1	R527	E535	N540	K590
2	WP_044910712	*Lachnospiraceae bacterium MC2017*	Lb3Cpf1	N520	E528	K533	K582
3	KKR91555	*Candidatus Falkowbacteria bacterium GW2011_GWA2_41_14*		E633	K639	Y643	G705
4	KKP36646	*Candidatus Peregrinibacteria bacterium GW2011_GWA2_33_10*	PeCpf1	G623	K629	N633	K703
5	KKQ36153	*candidate division WS6 bacterium GW2011_GWA2_37_6*		G568	Q574	K578	R619
6	EKE28449	*uncultured bacterium (gcode 4)*		T552	K558	R562	R615
7	KKQ38174	*Candidatus Roizmanbacteria bacterium GW2011_GWA2_37_7*		N592	K598	N602	K660
8	KKT48220	*Parcubacteria group bacterium GW2011_GWC2_44_17*	PbCpf1	K624	K630	F634	R689
9	WP_005398606	*Helcococcus kunzii*		D554	K560	N564	N614
10	WP_028830240	*Proteocatella sphenisci*		K483	K489	N493	K551
11	WP_015504779	*Candidatus Methanomethylophilus alvus*		D515	K521	N525	K577
12	CUP14506	*Lachnospira pectinoschiza*		S548	K554	N558	K614
13	CUM80100	*Eubacterium rectale*		D529	K535	N539	K594
14	WP_012739647	*Eubacterium eligens*	EeCpf1	N535	K541	N545	K601
15	AIZ56868	*Candidatus Methanoplasma termitum*	CMtCpf1	N528	K534	Y538	R591
16	WP_037975888	*Synergistes jonesii*		K539	K545	N549	K602
17	WP_021736722	*Acidaminococcus sp. BV3L6*	**AsCpf1**	S542	K548	N552	K607
18	WP_031492824	*Succinivibrio dextrinosolvens*		E564	K570	C574	K629
19	WP_018359861	*Porphyromonas macacae*	PmCpf1	S559	K565	N569	K623
20	WP_050786240	*Prevotella disiens*		T588	N600	Y604	K674
21	WP_027407524	*Anaerovibrio sp. RM50*		A525	N531	N535	K594
22	KDN25524	*Moraxella bovoculi 237*	MbCpf1	N576	K582	N586	K637
23	AJI61006	*Francisella tularensis subsp. novicida U112*	FnCpf1	N607	K613	N617	K671
24	KUJ74576	*Thiomicrospira sp. XS5*		S575	K581	N585	K658
25	WP_051666128	*Lachnospiraceae bacterium ND2006* (*)	LbCpf1	G550	K556	Y560	K613
26	WP_027109509	*Lachnospiraceae bacterium NC2008*		G511	K517	C521	K574
27	WP_027216152	*Butyrivibrio fibrisolvens*		D510	N516	Y520	N573
28	WP_028248456	*Pseudobutyrivibrio ruminis*		N511	K517	N521	K574
29	WP_049895985	*Oribacterium sp. NK2B42*		D528	K534	N538	K591
30	WP_035798880	*Butyrivibrio sp. NC3005*		N512	K518	N522	K575
31	WP_044919442	*Lachnospiraceae bacterium MA2020*	Lb2Cpf1	N512	K518	N522	K575
32	WP_044910713	*Lachnospiraceae bacterium MC2017*		C537	K543	Y547	K599
33	WP_020988726	*Leptospira inadai*	LiCpf1	K580	R586	N590	R644
34	WP_016301126	*Lachnospiraceae bacterium COE1*		D545	K551	N555	R608
35	KIE18657	*Smithella sp. SC_K08D17*	SsCpf1	G561	K567	N571	K625
36	WP_014085038	*Flavobacterium branchiophilum*		N588	K594	Y598	K649
37	WP_045971446	*Flavobacterium sp. 316*		N586	K592	Y596	K647
38	KXB38146	*Bacteroidales bacterium KA00251*		A550	K556	N560	K613
39	WP_036890108	*Porphyromonas crevioricanis*	PcCpf1	S575	K581	N585	K641
40	WP_044110123	*Prevotella brevis*		D541	K547	N551	K603
41	WP_009217842	*Bacteroidetes oral taxon 274*		D564	K570	N574	K628
42	WP_006283774	*Prevotella bryantii*		G566	K572	N576	K629
43	WP_024988992	*Prevotella albensis*		G561	K567	C571	K624

(*) The version of LbCpf1 we used in in this study has the first 18 residues deleted relative to the NCBI record for accession number WP_051666128, consistent with previous work (Zetsche *et al*. 2015; Yamano *et al*. 2016). The numbering in figures and text is based on the truncated sequence (G532/K538/Y542/K595).

Supplementary Table 4a Selected list of plasmids used in study

Construct	Experiments Used	Figures	Addgene #
T7-AsCpf1(library)-T7-crRNA	Bacterial	1C-D	
CMV-AsCpf1(WT)-NLS-3xHA	Indel; lysate	2A-D, 3B-C, S1, S4, S8	69982
CMV-NLS-AsCpf1(WT)-NLS-3xHA	Indel	2E, S5, S10	
CMV-AsCpf1(RR)-NLS-3xHA	Indel; lysate	2A-D, 3B-C, S4	
CMV-NLS-AsCpf1(RR)-NLS-3xHA	Indel	2E, 3D, S5, S10	89351
CMV-AsCpf1(RVR)-NLS-3xHA	Indel; lysate	2A-D, 3B-C, S1	
CMV-NLS-AsCpf1(RVR)-NLS-3xHA	Indel	2E, 3D, S5, S10	89353
CMV-NLS-LbCpf1(WT)-NLS-3xHA	Indel	S10, S11	
CMV-NLS-LbCpf1(RR)-NLS-3xHA	Indel	S10, S11	89355
CMV-NLS-LbCpf1(RVR)-NLS-3xHA	Indel	S10, S11	
U6-crRNA-CBh-NLS-AsCpf1(WT)-NLS-3xHA	BLISS	3A	
U6-crRNA-CBh-NLS-AsCpf1(RR)-NLS-3xHA	BLISS; indel (Neuro2a)	3A, S6	89352
U6-crRNA-CBh-NLS-AsCpf1(RVR)-NLS-3xHA	BLISS	3A	89354
U6-crRNA(As)-CMV-mCherry	Indel	2A, 2E, 3B-D, S1, S4, S5, S8, S10	
U6-crRNA(Lb)-CMV-mCherry	Indel	S10, S11	

Supplementary Table 4b List of guide sequences used in study

Figure				Gene	PAM	Guide Sequence (5' to 3')
1C	1D			[Plasmid]	Varies	CCGATGGTCCATGTCTGTTACTCGCCTGTC
2A				*DNMT1*	TCCC site 1	GTCACCCCTGTTTCTGGCACCAG
2A				*DNMT1*	TTCC site 1	TGGTGCCAGAAACAGGGGTGACG
2A				*VEGFA*	TTCC site 2	AAAGCCCATTCCCTCTTTAGCCA
2A	S1			*VEGFA*	TATC site 1	AAATTCCAGCACCGAGCGCCCTG
2A				*DNMT1*	TATA site 1	GGAGAGATTTATTTGAAGAAATA
2B	2C	2D		[Plasmid]	NNNN	GAGAAGTCATTTAATAAGGCCACT
2E		S5A	S10B	*TSPYL6*	TTTG	ATGGCCGCTTGGAGCCAATCGTG
2E		S5A	S10B	*PRMT1*	TTTG	ACGATCTTCACCGCATAATCAGA
2E		S5A	S10B	*WNT8A*	TTTC	CCCAAATTCCACATTGTCGCTGC
2E		S5A	S10B	*FANCF*	TTTG	CACTATGACCTTCAGAAAGGCAT
2E		S5A	S10B	*PRCC*	TTTG	TCACCTGCAGGGCAGCACTCTTG
2E		S5A	S10B	*THAP10*	TTTG	ACGTCTCTTCGGTTATCCAGAAG
2E		S5A	S10B	*FASTKD1*	TTTC	TTCGAAATGTTAGATATCGTTAT
2E		S5A	S10B	*AGO1*	TTTG	ATGCAGGCATCACGAATGGCCAG
2E		S5A	S10B	*RBM11*	TTTG	CTGAATGGAATTCGTTTATATGG
2E		S5A	S10B	*PROM2*	TTTC	AGGCTCTGCAACTCCTGCCGTAG
2E		S5A	S10B	*GRN*	TTTA	CGTGTGACACGCAGAAGGGTACC
2E		S5A	S10B	*RUNX1*	TTTC	ACATTTGCTTCTCTTTACCATAG
2E		S5A	S10B	*ESAM*	TTTC	TCCTGGAGACCCTCCAGCCGCAG
2E		S5A	S10B	*RRAGA*	TTTC	CAGTTCGCGGCTCTCCACGTCAA
2E		S5A	S10B	*APOB*	TTTC	AGTGGATATTTCTGTTGCCACAT
2E		S5A	S10B	*ERBB2*	TTTG	TGGAAGGACATCTTCCACAAGAA
2E		S5A	S10B	*LIPF*	TTTC	CTCTGCTGTTGCCCAGCCACACA
2E		S5A	S10B	*FXN*	TTTC	CCAGTCCAGTCATAACGCTTAGG
2E		S5A	S10B	*FAP*	TTTC	GGTCCTGTCTATATGTGACTTCA
2E		S5A	S10B	*DNMT2*	TTTC	CAGAACACTGTATGCTGCCATCA
2E		S5A	S10B	*HOTTIP*	TTTC	CCTGAGAGCTGGGCCGAACGGGG
2E		S5A	S10B	*XIST*	TTTA	CCCTTGGCATTGCTGATCTTCAG
2E		S5A	S10B	*HIST1H4C*	TTTC	CGGTCTTATCTATGAGGAGACTC
2E	3D	S5B	S10B	*TSPYL6*	TCCA	AGCGGCCATCAAACATATCTGCC
2E	3D	S5B	S10B	*PRMT1*	TTCA	CCGCATAATCAGAGATACTGGAA
		S5B		*WNT8A*	CCCC	AAATTCCACATTGTCGCTGCAGC
2E	3D	S5B	S10B	*FANCF*	TCCA	ACCCAAATGCCTTTCTGAAGGTC
2E	3D	S5B	S10B	*PRCC*	TTCC	TGCGTGATCTGCTTTGTCACCTG
2E	3D	S5B	S10B	*THAP10*	TCCC	AGCGCCTGAGGCTGGTGGCAGGC
2E	3D	S5B	S10B	*FASTKD1*	TTCG	AAATGTTAGATATCGTTATCAAC
2E	3D	S5B	S10B	*AGO1*	TCCA	GTTTGATGCAGGCATCACGAATG
2E	3D	S5B	S10B	*RBM11*	TTCG	TTTATATGGAAGACCAATTAACG
2E	3D	S5B	S10B	*PROM2*	TTCA	GGCTCTGCAACTCCTGCCGTAGC
		S5B		*GRN*	CCCC	GCGGGACAGCAGTGTATGTGGTC
2E	3D	S5B	S10B	*RUNX1*	TCCC	TATGGTAAAGAGAAGCAAATGTG
2E	3D	S5B	S10B	*ESAM*	TTCC	GGGAGGGCATGGAGTAGACCAAG
2E	3D	S5B	S10B	*RRAGA*	TTCC	AGTTCGCGGCTCTCCACGTCAAA
		S5B		*APOB*	CCCC	AGGTCTCTTTCAGTGGATATTTC
2E	3D	S5B	S10B	*ERBB2*	TTCC	ACAAGAACAACCAGCTGGCTCTC
2E	3D	S5B	S10B	*LIPF*	TTCC	TCTGCTGTTGCCCAGCCACACAT

Supplementary Table 4b (continued)

2E	3D	S5B	S10B	*FXN*	TCCC	AGTCCAGTCATAACGCTTAGGTC
2E	3D	S5B	S10B	*FAP*	TCCC	TGAAGTCACATATAGACAGGACC
2E	3D	S5B	S10B	*DNMT2*	TCCA	GAACACTGTATGCTGCCATCAAA
2E	3D	S5B	S10B	*HOTTIP*	TTCG	GCCCAGCTCTCAGGGAAACGAAG
2E	3D	S5B	S10B	*XIST*	TCCG	TTTACCCTTGGCATTGCTGATCT
2E	3D	S5B	S10B	*HIST1H4C*	TTCC	GGTCTTATCTATGAGGAGACTCG
2E	3D	S5C	S10B	*TSPYL6*	TATC	TGCCATTACCTCTGTCGCCTTGC
2E	3D	S5C	S10B	*PRMT1*	TATC	TCTGATTATGCGGTGAAGATCGT
2E	3D	S5C	S10B	*WNT8A*	TATA	TAGGAGGCCATGGCTGGATCTGG
2E	3D	S5C	S10B	*FANCF*	TATC	TGGGTCTGCTAACAGACTGGGGT
2E	3D	S5C	S10B	*PRCC*	TATC	AAGGCTGCTGCCAAGAGTGCTGC
2E	3D	S5C	S10B	*THAP10*	TATC	CAGAAGAACCTGCGCTTCTCCCA
2E	3D	S5C	S10B	*FASTKD1*	TATC	TAACATTTCGAAGAAACTTTGCT
2E	3D	S5C	S10B	*AGO1*	TATG	AGCTACTGGCCATTCGTGATGCC
2E	3D	S5C	S10B	*RBM11*	TATG	CCATAGCTTTGCTGAATGGAATT
2E	3D	S5C	S10B	*PROM2*	TATA	CCAACAAGCTACGGCAGGAGTTG
2E	3D	S5C	S10B	*GRN*	TATG	TGGTCCTCACAGCACACAGCCTA
2E	3D	S5C	S10B	*RUNX1*	TATG	GTAAAGAGAAGCAAATGTGAAAC
2E	3D	S5C	S10B	*ESAM*	TATC	CTTGGTCTACTCCATGCCCTCCC
2E	3D	S5C	S10B	*RRAGA*	TATC	TTCCGTAACGTGGAAGTTTTGAT
2E	3D	S5C	S10B	*APOB*	TATC	CACTGAAAGAGACCTGGGGCAGT
2E	3D	S5C	S10B	*ERBB2*	TATC	AGTGTGAGAGCCAGCTGGTTGTT
2E	3D	S5C	S10B	*LIPF*	TATG	ATGTGTGGCTGGGCAACAGCAGA
2E	3D	S5C	S10B	*FXN*	TATG	ACTGGACTGGGAAAAACTGGGTG
2E	3D	S5C	S10B	*FAP*	TATA	TGTGACTTCAGGGAAGACTGGCA
2E	3D	S5C	S10B	*DNMT2*	TATG	CTGCCATCAAAGCTAATATTTGG
2E	3D	S5C	S10B	*HOTTIP*	TATC	AAAAGTTCTTACTGAGCGCTTCG
2E	3D	S5C	S10B	*XIST*	TATA	CTGGGATATTCCGTTTACCCTTG
2E	3D	S5C	S10B	*HIST1H4C*	TATC	TATGAGGAGACTCGAGGTGTGCT
3A				*VEGFA*	TTTG	CTAGGAATATTGAAGGGGGCAGG
3A				*GRIN2B*	TTTG	GTGCTCAATGAAAGGAGATAAGG
3A				*DNMT1*	TTTG	AAGAAATATTACAACATATAAAA
3A	3C	S8		*EMX1*	TTTG	TCCTCCGGTTCTGGAACCACACC
3B				*RPL32P3*	TTTG	GGGTGATCAGACCCAACAGCAGG
S4A				*EMX1*	TTTG	GGGAGGCCTGGAGTCATGGCCCC
S4A				*EMX1*	TTTG	TGGTTGCCCACCCTAGTCATTGG
S4A				*DNMT1*	TTTC	CCTTCAGCTAAAATAAAGGAGGA
S4A				*DNMT1*	TTTG	AGGAGTGTTCAGTCTCCGTGAAC
S4A				*CFTR*	TTTA	ATGGTGCCAGGCATAATCCAGGA
S4A				*VEGFA*	TTTA	GCCAGAGCCGGGGTGTGCAGACG
S4A				*DNMT1*	TTTC	CTGATGGTCCATGTCTGTTACTC
S4A				*VEGFA*	TTTC	CAAAGCCCATTCCCTCTTTAGCC
S4A				*DNMT1*	TTTA	GCTGAAGGGAAATAAAAGGAAAA
S4A				*CFTR*	TCTC	AGTTTTCCTGGATTATGCCTGGC
S4A				*EMX1*	TCTG	GCCACTCCCTGGCCAGGCTTTGG
S4A				*CFTR*	TCTA	TATTCATCATAGGAAACACCAAA
S4A				*VEGFA*	TCTC	TGTACATGAAGCAACTCCAGTCC
S4A				*DNMT1*	TCTG	CCCTCCCGTCACCCCTGTTTCTG
S4A				*VEGFA*	TCTA	TTGGAATCCTGGAGTGACCCCTG

Supplementary Table 4b (continued)

S4A				*DNMT1*	TCTC	CGTGAACGTTCCCTTAGCACTCT
S4A				*VEGFA*	TCTG	GCTAAAGAGGGAATGGGCTTTGG
S4A				*EMX1*	TTCT	TCTTCTGCTCGGACTCAGGCCCT
S4A				*DNMT1*	TTCT	GCCCTCCCGTCACCCCTGTTTCT
S4A				*VEGFA*	TTCT	GACCTCCCAAACAGCTACATATT
S4A				*DNMT1*	TCCT	GGTGCCAGAAACAGGGGTGACGG
S4A				*DNMT1*	TCCT	GATGGTCCATGTCTGTTACTCGC
S4B	S4A			*CFTR*	TTCG	GCGATGTTTTTTCTGGAGATTTA
S4B	S4A			*DNMT1*	TTCA	GCTAAAATAAAGGAGGAGGAAGC
S4B	S4A			*DNMT1*	TCCC	GTCACCCCTGTTTCTGGCACCAG
S4B	S4A			*DNMT1*	TTCC	TGGTGCCAGAAACAGGGGTGACG
S4B	S4A			*DNMT1*	TTCA	GTCTCCGTGAACGTTCCCTTAGC
S4B	S4A			*DNMT1*	TTCA	CGGAGACTGAACACTCCTCAAAC
S4B	S4A			*EMX1*	TTCG	TGGCAATGCGCCACCGGTTGATG
S4B	S4A			*VEGFA*	TTCC	CTGTGGTGGCCGAGCGCCCCCTA
S4B	S4A			*VEGFA*	TCCA	GTCCCAAATATGTAGCTGTTTGG
S4B	S4A			*VEGFA*	TCCG	CACGTAACCTCACTTTCCTGCTC
S4B	S4A			*VEGFA*	TCCG	CCCCCGGAAACTCTGTCCAGAGA
S4B	S4A			*VEGFA*	TCCG	GGGGCGGATGGGTAATTTTCAGG
S4B	S4A			*VEGFA*	TCCC	TCTTTAGCCAGAGCCGGGGTGTG
S4B	S4A			*VEGFA*	TCCA	ATAGATCTGTGTGTCCCTCTCCC
S4B	S4A			*VEGFA*	TTCC	AAAGCCCATTCCCTCTTTAGCCA
S4B	S4A			*VEGFA*	TCCC	CCCACCCCCTTTCCAAAGCCCAT
S4B				*CFTR*	GTCG	AAAATTTTACACCACAAAATGTT
S4B				*CFTR*	ACCA	AAGATGATATTTTCTTTAATGGT
S4B				*CFTR*	ACCA	TTAAAGAAAATATCATCTTTGGT
S4B				*CFTR*	ATCC	TAAACTCATTAATGCCCTTCGGC
S4B				*CFTR*	ATCC	AGGAAAACTGAGAACAGAATGAA
S4B				*EMX1*	ATCA	CATCAACCGGTGGCGCATTGCCA
S4B				*EMX1*	GTCC	TCCCCATTGGCCTGCTTCGTGGC
S4B				*EMX1*	CCCG	GGCTTCAAGCCCTGTGGGGCCAT
S4B				*EMX1*	ATCG	ATGTCACCTCCAATGACTAGGGT
S4B				*EMX1*	ATCG	ATGTCCTCCCCATTGGCCTGCTT
S4B				*VEGFA*	CCCA	TTCCCTCTTTAGCCAGAGCCGGG
S4B				*VEGFA*	CTCG	GCCACCACAGGGAAGCTGGGTGA
S4B				*VEGFA*	GTCC	CAAATATGTAGCTGTTTGGGAGG
S4B				*VEGFA*	GCCG	AGCGCCCCCTAGTGACTGCCGTC
S4B				*VEGFA*	GCCC	ATTCCCTCTTTAGCCAGAGCCGG
S4B				*VEGFA*	CCCG	GCTCTGGCTAAAGAGGGAATGGG
S4B				*VEGFA*	GCCA	GAGCCGGGGTGTGCAGACGGCAG
S4B				*VEGFA*	CTCG	CTCCATTCACCCAGCTTCCCTGT
S4B				*VEGFA*	GCCC	TGGGCTCTCTGTACATGAAGCAA
S4B				*VEGFA*	CTCC	AGTCCCAAATATGTAGCTGTTTG
S4B				*VEGFA*	GTCA	GAAATAGGGGGTCCAGGAGCAAA
S4B				*VEGFA*	ACCC	CGGCTCTGGCTAAAGAGGGAATG
S4B				*VEGFA*	ACCA	CAGGGAAGCTGGGTGAATGGAGC
S4B				*VEGFA*	CCCA	GCTTCCCTGTGGTGGCCGAGCGC
S4B				*VEGFA*	GCCG	TCTGCACACCCCGGCTCTGGCTA
S4B				*VEGFA*	GCCC	CCTAGTGACTGCCGTCTGCACAC

Supplementary Table 4b (continued)

S4B				*VEGFA*	GCCA	CCACAGGGAAGCTGGGTGAATGG
S4B				*VEGFA*	GTCA	CTAGGGGGCGCTCGGCCACCACA
S4B				*VEGFA*	GTCC	TCACTCTCGAAGACGCTGCTCGC
S4B				*VEGFA*	GCCG	GGGTGTGCAGACGGCAGTCACTA
S4B				*VEGFA*	CCCG	CTCCAACGCCCTCAACCCCACAC
S4B				*VEGFA*	CTCG	AAGACGCTGCTCGCTCCATTCAC
S4B				*VEGFA*	CTCC	TGGACCCCCTATTTCTGACCTCC
S4B				*VEGFA*	ATCC	TGGAGTGACCCCTGGCCTTCTCC
S4B				*VEGFA*	ACCC	CCTTTCCAAAGCCCATTCCCTCT
S4B				*VEGFA*	CCCC	CCACCCCCTTTCCAAAGCCCATT
S6				*PCSK9* (mouse)	TCCC	GTCCCAGGAGGATGGCCTGGCTG
S6				*PCSK9* (mouse)	TCCC	AGGAGGATGGCCTGGCTGATGAG
S6				*PCSK9* (mouse)	TTCA	ATCTGTAGCCTCTGGGTCTCCTC
S6				*PCSK9* (mouse)	TCCC	TGGCTTCTTGGTGAAGATGAGCA
S6				*PCSK9* (mouse)	TTCC	TCAATGTACTCCACATGGGGCAA
S6				*PCSK9* (mouse)	TCCA	TGGGATGCTCTGGGCGAAGACAA
S6				*PCSK9* (mouse)	TTCC	TCTGTCTGGTGCCATGCTGGGAT
S6				*PCSK9* (mouse)	TCCC	GATGGGCACCCTGGATGCTGGTA
S6				*PCSK9* (mouse)	TCCC	GGCCGCTGACCACACCTGCCAGG