nature biotechnology

Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells

Daesik Kim[1–3], Jungeun Kim[1–3], Junho K Hur[1], Kyung Wook Been[1,2], Sun-heui Yoon[2] & Jin-Soo Kim[1,2]

Programmable clustered regularly interspaced short palindromic repeats (CRISPR) Cpf1 endonucleases are single-RNA-guided (crRNA) enzymes that recognize thymidine-rich protospacer-adjacent motif (PAM) sequences and produce cohesive double-stranded breaks (DSBs). Genome editing with CRISPR-Cpf1 endonucleases could provide an alternative to CRISPR-Cas9 endonucleases, but the determinants of targeting specificity are not well understood. Using mismatched crRNAs we found that Cpf1 could tolerate single or double mismatches in the 3′ PAM-distal region, but not in the 5′ PAM-proximal region. Genome-wide analysis of cleavage sites *in vitro* for eight Cpf1 nucleases using Digenome-seq revealed that there were 6 (LbCpf1) and 12 (AsCpf1) cleavage sites per crRNA in the human genome, fewer than are present for Cas9 nucleases (>90). Most Cpf1 off-target cleavage sites did not produce mutations in cells. We found mismatches in either the 3′ PAM-distal region or in the PAM sequence of 12 off-target sites that were validated *in vivo*. Off-target effects were completely abrogated by using preassembled, recombinant Cpf1 ribonucleoproteins.

© 2016 Nature America, Inc. All rights reserved.

CRISPR and CRISPR-associated (Cas) proteins, which function as adaptive immune systems in eubacteria and archaea, have been repurposed for targeted genome editing in human and other cells and organisms, and have broad applications in research, medicine and biotechnology[1–8]. Cas9, the single protein component in the type II CRISPR/Cas system, is complexed with two small RNAs named CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) to form a sequence-specific RNA-guided endonuclease (RGEN) whose target specificity is readily reprogrammed by either modifying the crRNA or using a single-chain guide RNA (sgRNA) composed of essential portions of crRNA and tracrRNA[9]. Cas9 RGENs cleave chromosomal DNA to produce site-specific DNA double-strand blunt-end breaks (DSBs) that are repaired by homologous recombination (HR) or non-homologous end-joining (NHEJ) to yield genetic modifications.

Cpf1 is a different RGEN that is derived from a type V (or class II) CRISPR system[10]. Cpf1 differs from Cas9 in several ways. First, Cpf1 is a single-RNA-guided nuclease and does not require a tracrRNA. Second, Cpf1 cleavage produces cohesive (not blunt) DSBs, which might facilitate NHEJ-mediated transgene knock-in at target sites. Third, Cpf1 recognizes thymidine-rich DNA sequences, for example, 5′-TTTN-3′, as the PAM at the 5′ end of target sequences. This should increase the selection of CRISPR-endonuclease-editable genomic sites beyond the guanosine-rich sequences recognized by *Streptococcus pyogenes* Cas9 enzymes (SpCas9). To harness the potential of Cpf1, we need to characterize the genome-wide target specificities of this RGEN.

First, we compared the efficiencies of Cpf1-mediated genome editing in HEK293 cells transfected with either plasmids or PCR amplicons encoding crRNAs (**Fig. 1a**). We found that the frequency of targeted mutagenesis was increased by 2–30-fold at three endogenous target sites using crRNA plasmids rather than amplicons. PCR amplicons can produce incorrect guide RNAs by transcription from synthesis-failed oligonucleotide templates, potentially causing off-target DNA cleavages at sites that seem to have an RNA bulge[11]. To avoid unwanted off-target effects at such sites and to optimize genome editing efficiency, we used plasmids encoding crRNAs throughout this study.

Next, we examined crRNA orthogonality with four Cpf1 orthologs from *Lachnospiraceae bacterium* (LbCpf1), *Acidaminococcus sp.* (AsCpf1), *Francisella novicida* (FnCpf1) and *Moraxella bovoculi 237* (MbCpf1) (**Fig. 1b** and **Supplementary Fig. 1**). LbCpf1 and AsCpf1 recognize 5′-TTTN-3′ PAMs, and have been successfully used for genome editing in human cells, whereas FnCpf1 and MbCpf1 recognize 5′-TTN-3′ PAMs, but have been shown to be inefficient or inactive in human cells[10]. Cpf1 orthologs were co-transfected into HEK293 cells with plasmids encoding crRNA orthologs in various combinations. Each Cpf1 ortholog was most efficient with its cognate crRNA. However, all four Cpf1 orthologs in combination with non-orthogonal crRNAs cleaved chromosomal target sites. AsCpf1 and LbCpf1 were the most efficient endonucleases (**Fig. 1b**), and all subsequent experiments were carried out with these RGENs.

We compared the frequency of targeted mutations by LbCpf1, AsCpf1 and SpCas9 at ten chromosomal target sites. Each site contained two PAM sequences, one recognized by Cpf1 (5′-TTTN-3′) and the other recognized by SpCas9 (5′-NGG-3′) (**Fig. 1c** and **Supplementary Table 1**). Each RGEN had a wide range of mutation frequencies at these sited in HEK293cells. SpCas9 was more efficient than Cpf1, with an average mutation rate of 32 ± 4%. LbCpf1 and AsCpf1 had mutation frequencies of 19 ± 6% and 20 ± 5%, respectively, which is about the same frequency as *Staphylococcus aureus* Cas9 (ref. 12).

[1]Center for Genome Engineering, Institute for Basic Science (IBS), Seoul, Republic of Korea. [2]Department of Chemistry, Seoul National University, Seoul, Republic of Korea. [3]These authors contributed equally to this work. Correspondence should be addressed to J.-S.K. (jskim01@snu.ac.kr).

Received 25 February 2015; accepted 13 May 2016; published online 6 June 2016; doi:10.1038/nbt.3609

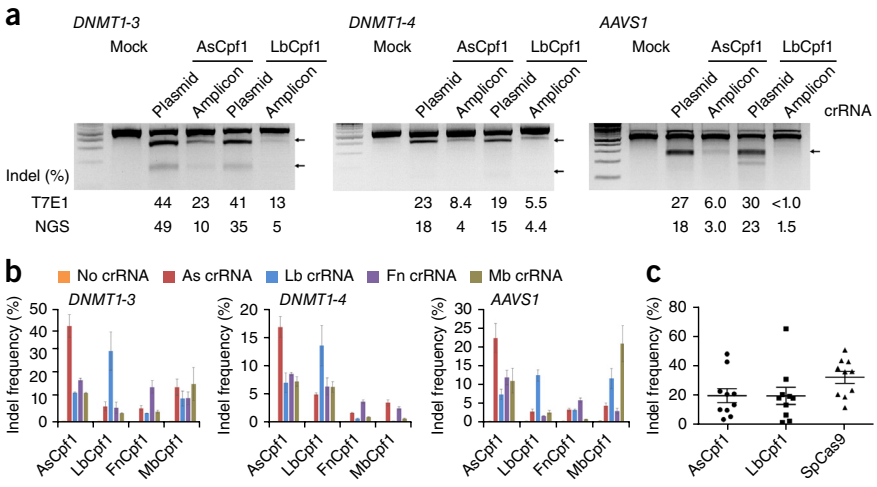


Figure 1 Genome editing in human cells using Cpf1. (**a**) Genome editing in HEK293T cells using plasmids or PCR amplicons encoding crRNAs is shown. Indel frequencies were measured with the T7E1 assay and next-generation sequencing (NGS). Arrows indicate cleaved DNA bands. Full-length gels are presented in **Supplementary Figure 11**. (**b**) crRNA orthogonality with four Cpf1 orthologs. Plasmids encoding Cpf1 orthologs were co-transfected with plasmids encoding crRNAs, in various combinations, into HEK293T cells. Indel frequencies were measured by targeted deep sequencing. Error bars indicate s.e.m. (**c**) Mutation frequencies obtained with LbCpf1, AsCpf1 and SpCas9 at ten endogenous target sites in HEK293T cells. Each target site contained two PAM sequences, one recognized by Cpf1 and the other recognized by SpCas9. Indel frequencies were measured by targeted deep sequencing. Mean indel frequencies ± s.e.m. are shown ($n = 10$ target sites).

© 2016 Nature America, Inc. All rights reserved.

LbCpf1 and AsCpf1 recognize and cleave 27-nucleotide (nt) target DNA sequences composed of the 5′-TTTN-3′ PAM sequence followed by 23-nt protospacer sequences that match with crRNA sequences. We arbitrarily chose three endogenous target sites and transfected a series of plasmids encoding crRNAs with one or two mismatches into HEK293 cells to investigate whether Cpf1 can tolerate mismatches between on-target DNA sequences and crRNA sequences (**Fig. 2** and **Supplementary Fig. 2**). Notably, for the DNMT1–4 site, Cpf1 did not tolerate single mismatches at positions 1–17 (numbered 1–23 in the 5′ to 3′ direction; **Fig. 2**). Double mismatches at positions 1–18 resulted in an almost complete loss of Cpf1 activity. For the other two sites, single mismatches at most positions were tolerated by Cpf1, whereas double mismatches at positions 1–18 led to substantial loss of Cpf1 activity. These results indicate that Cpf1 is highly specific in human cells, although it can tolerate mismatches at the PAM-distal end.

Next, using Cas-OFFinder[13], we identified all of the potential off-target sites in the human genome that differed from each of the ten on-target sites that we initially examined by one to four or five nucleotides. A total of 130 sites were found using Cas-OFFinder. We measured whether these sites were cleaved using targeted deep sequencing in HEK293 cells (**Supplementary Table 2**). LbCpf1 and AsCpf1 induced off-target indels at 3 and 4 sites, respectively, of 87 sites that differed from 2 on-target sites (labeled as *DNMT1-3* and *EMX1-2* sites) by five nucleotides, but indel frequencies at these bona fide off-target sites (0.04–0.7%) were much lower than those at the two on-target sites (34% and 25% with LbCpf1; 47% and 13% with AsCpf1). Notably, Cpf1 distinguished the other two on-target sites (labeled *CCR5-1* and *HPRT-1*) from respective homologous sites with a single mismatch. Thus, LbCpf1 induced indels at the two on-target sites with a frequency of 19% and 10%, respectively, and at the single-base mismatched sites with a frequency of 0.4% and 0.04%, respectively, discriminating the off-target sites from the on-target sites by a factor of 48 (19%/0.4%) and 250, respectively. Overall, 9 of 130 sites that we examined were validated via targeted deep sequencing as bona fide off-target sites, but indel frequencies at most of these sites were below 1%. These results confirm that Cpf1 is highly specific in human cells.

To identify all genome-wide Cpf1 off-target sites *in vitro*, in an unbiased manner, we used a total of eight efficient Cpf1 nucleases and performed Digenome-seq[11,14]. Cell-free genomic DNA isolated from HeLa cells was digested with preassembled, recombinant AsCpf1 and LbCpf1 ribonucleoproteins (RNPs) at a high concentration (300 nM Cpf1 and 900 nM crRNA) and subjected to whole-genome sequencing (WGS). Uniform, rather than random, alignments of sequence reads, corresponding to on-target and off-target *in vitro* cleavage sites, were computationally identified. Cpf1 was highly specific, cleaving chromosomal DNA *in vitro* at just 1–46 sites (**Fig. 3a**,**b** and **Supplementary Table 3**). Notably, all eight on-target sites, and the eight off-target sites validated by deep sequencing, were captured by Digenome-seq (**Supplementary Table 2**). The number of *in vitro* cleavage sites identified in our monoplex Digenome-seq analyses was 6 ± 3 for LbCpf1 or 12 ± 5 for AsCpf1, which is far fewer than the 90 ± 30 that we found using SpCas9 in a previous multiplex Digenome-seq experiment (**Supplementary Fig. 3**)[11]. Note that the number of Cas9 cleavage sites was likely underestimated because 11 sgRNAs were mixed in that multiplex analysis, reducing the concentration of each Cas9 RNP by 11-fold (27 nM). At a higher concentration (300 nM), Cas9 RNPs cleaved ~50% more chromosomal sites.

We found that 50 and 98 *in vitro* cleavage sites obtained using eight LbCpf1 and AsCpf1 nucleases, respectively, contained mismatches, mostly in the 13-nt PAM-distal region rather than in the 10-nt PAM-proximal region (**Supplementary Fig. 4** and **Supplementary Table 3**). In addition, 46 of the 50 sites (92%) cleaved by eight LbCpf1 nucleases were also cleaved by AsCpf1. Five sites had a missing nucleotide, as compared with their respective on-target sites, potentially creating an RNA bulge in the DNA-crRNA duplex region. Three LbCpf1 and AsCpf1 nucleases cleaved four and eight sites, respectively, that contained non-canonical PAM sequences, such as 5′-TCTN-3′ and 5′-TTCN-3′. Of 2,956 homologous sites with five or six mismatches, identified using Cas-OFFinder, only 24 sites (0.8%) were cleaved *in vitro* (**Fig. 3c**). Homologous sites with four or less mismatches were more likely to be cleaved, and therefore captured by Digenome-seq, but these sites were rare (6 ± 2 such sites per crRNA) in the human genome. Thus, 10 of 44 such sites (23%) were identified by Digenome-seq using eight different crRNAs (**Fig. 3c**). We found that there was no correlation ($R^2 = 0.0$) between the number of *in vitro* cleavage sites identified using Digenome-seq and the number of homologous sites with six or less mismatches in the human genome (**Supplementary Fig. 5**). This is in sharp contrast with our previous results with Cas9, which revealed a strong correlation ($R^2 = 0.93$)[11].

To compare the genome-wide specificity of SpCas9 with that of LbCpf1 and AsCpf1, we carried out Digenome-seq using two Cas9 sgRNAs that target the two Cpf1 sites in the *DNMT1* locus (**Fig. 3a**,**b**). Plots of the genome-wide distribution of *in vitro* cleavage sites revealed that Cas9 and Cpf1 mainly cut chromosomal DNA at different sites.

De novo motifs or sequence logos obtained by comparing DNA sequences at *in vitro* cleavage sites showed that Cpf1 was more specific than Cas9 (**Fig. 3a** and **Supplementary Fig. 6**). Notably, both LbCpf1 and AsCpf1 targeted to the *DNMT1-4* site cleaved only the single on-target site present in the entire human genome (**Fig. 3b**). Note that both Cpf1 and SpCas9 nucleases targeting the two *DNMT1* sites showed comparable indel frequencies in cells (**Supplementary Table 1**), suggesting that the high specificity of Cpf1 was not caused by reduced editing efficiencies in cells, at least at the sites studied. We also found that there was no substantial correlation between indel frequencies and genome-wide specificities, as defined by the number of Digenome-captured sites (**Supplementary Fig. 7**).

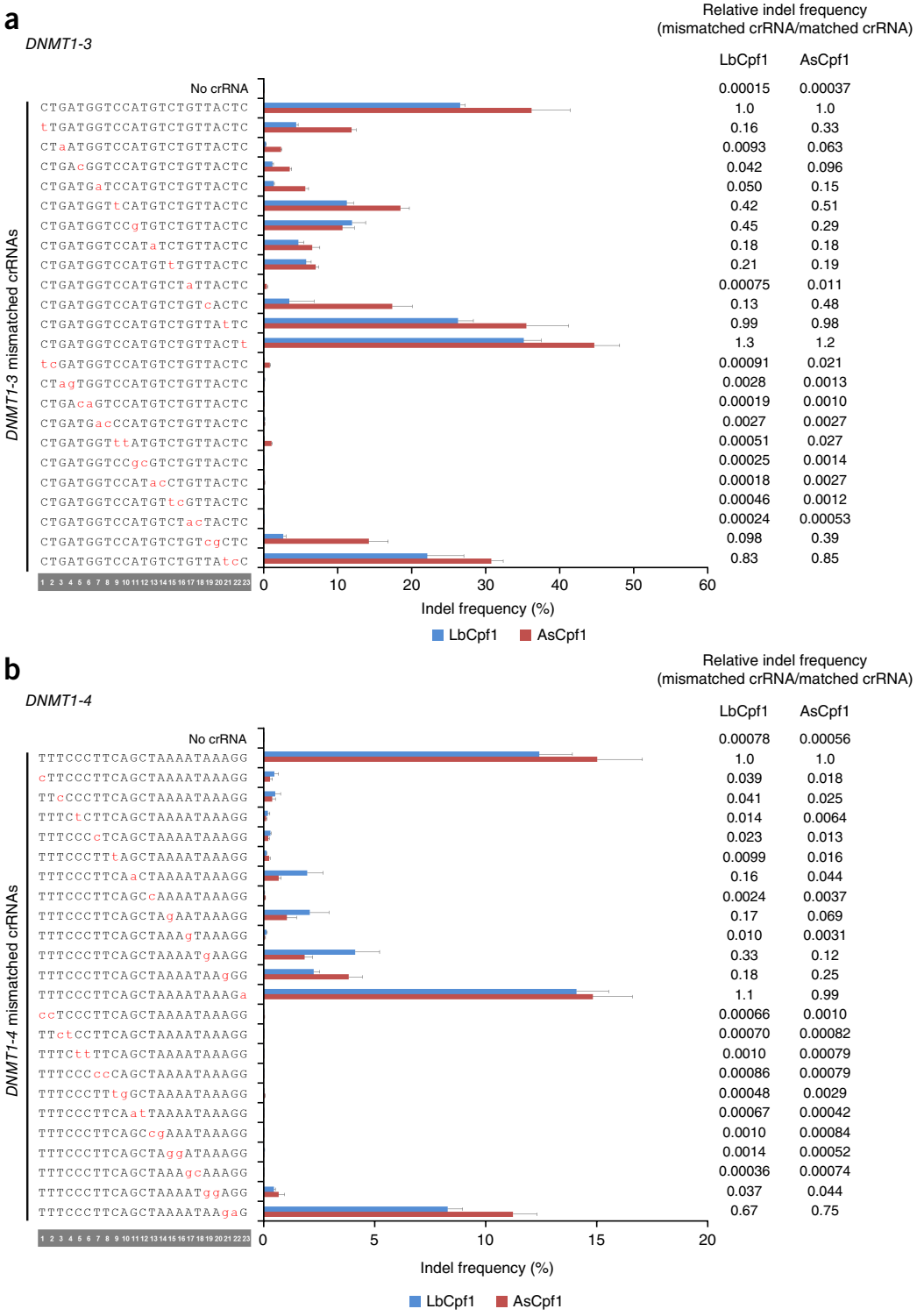


Figure 2 Specificity of Cpf1. (**a**,**b**) Mismatched crRNAs that differed from the two *DNMT1* target sites by 1 or 2 nt were tested in HEK293T cells. Indel frequencies were measured by targeted deep sequencing. Error bars indicate s.e.m. ($n = 3$ independent experiments).

© 2016 Nature America, Inc. All rights reserved.

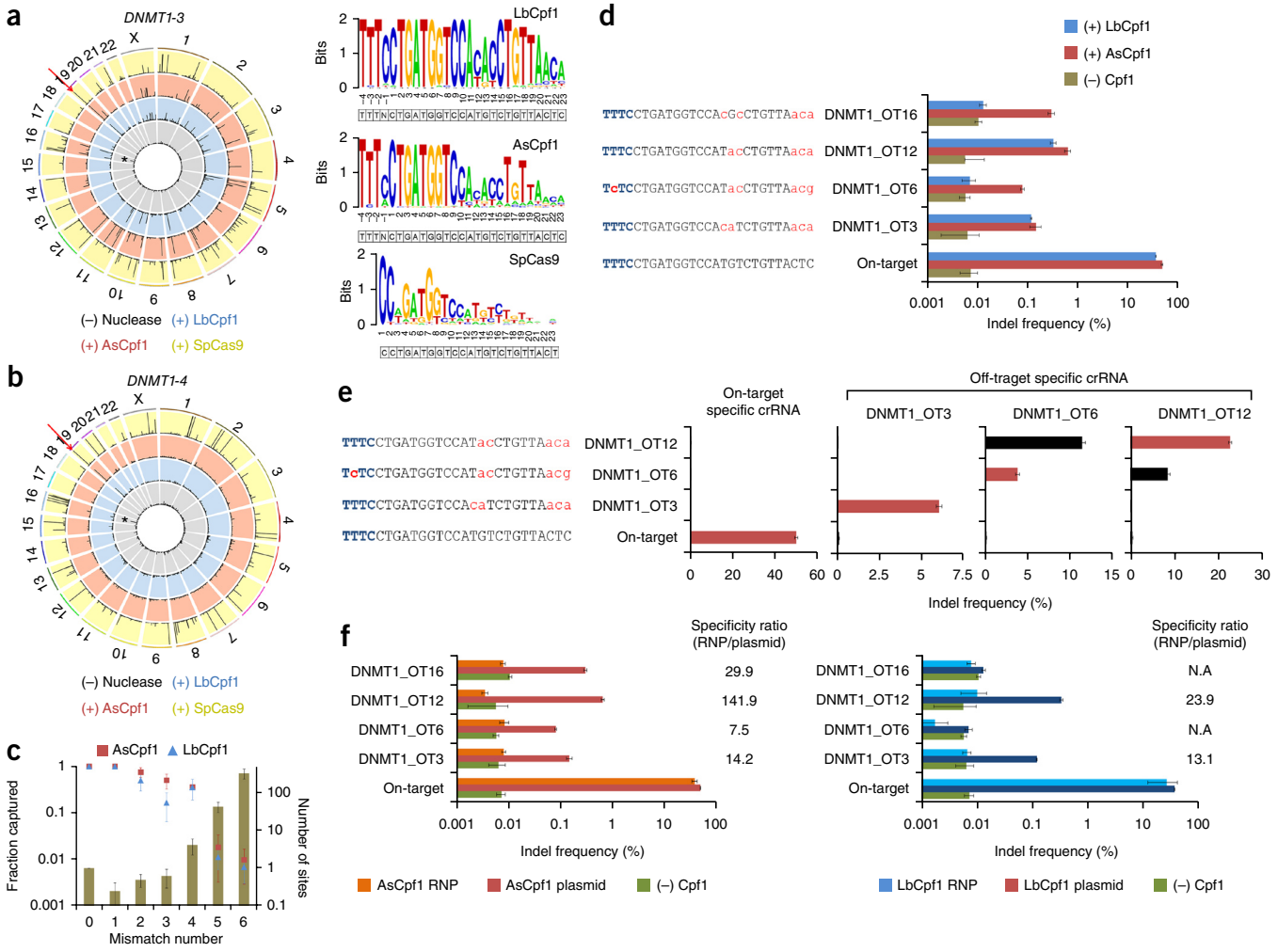


Figure 3 Using Digenome-seq to analyze genome-wide specificity of Cpf1 and SpCas9 RGENs. (**a**,**b**) Genome-wide Circos plots showing DNA cleavage scores. Intact genomic DNA (gray) and genomic DNA digested by LbCpf1 (blue), AsCpf1 (red) or SpCas9 (yellow) were subjected to whole-genome sequencing and Digenome-seq analysis. The asterisk indicates one false-positive site found in the intact genomic DNA. The arrow indicates the on-target site. Sequence logos were obtained via WebLogo using DNA sequences at *in vitro* cleavage sites identified by Digenome-seq. (**c**) Fractions (left, red dots for AsCpf1 and blue dots for LbCpf1) of homologous sites captured by Digenome-seq and numbers (right, bars) of homologous sites that differed from eight Cpf1 on-target sites by up to 6 nt binned by the mismatch number. Error bars indicate s.e.m. (**d**) Off-target sites validated in human cells by targeted deep sequencing. DNA sequences at on-target and off-target sites are shown. PAM sequences are shown in blue. Mismatched bases are shown in red. (**e**) Targeted mutagenesis at AsCpf1 off-target sites using redesigned crRNAs that match with off-target sites. (**f**) Cpf1 off-target effects were avoided using Cpf1 RNPs rather than plasmids encoding Cpf1 and crRNA. The specificity ratio indicates the fold difference between the ratio of on-target indel frequency to off-target indel frequency obtained using Cpf1 RNPs with that obtained using plasmids.

We validated 49 *in vitro* cleavage sites, identified by Digenome-seq, in HEK293 cells using targeted deep sequencing. At most of the *in vitro* cleavage sites, indels were not detected above the basal 'noise' of sequencing errors. Indel frequencies at most of the validated off-target sites were below 1% (**Fig. 3d** and **Supplementary Fig. 8**), much lower than those at corresponding on-target sites. To quantitate genome-wide off-target effects, we calculated an off-target effect index (OTI)[11], the ratio of the sum of indel rates at validated off-target sites to the on-target indel rate. The OTIs for LbCpf1 were 0.005 and 0.012 at the two *DNMT1* sites, whereas those for AsCpf1 were 0.267 and 0.024, respectively. These results suggest that off-target effects are site dependent and that LbCpf1 is more specific than AsCpf1. OTIs for SpCas9 nucleases targeting two different sites have been shown to be >2.0 (ref. 11). Engineered SpCas9 variants are more specific than wild-type SpCas9 (refs. 15,16), but their genome-wide specificities have not been tested using Digneome-seq.

To test whether the low indel frequencies at these validated off-target sites was caused by local chromatin inaccessibility, we transfected crRNAs with sequences that matched off-target sites (**Fig. 3e**). Each 'off-target-specific' crRNA induced indels at its respective site, but not at the on-target site, indicating that off-target sites were still accessible and cut by Cpf1. It remains plausible that RGEN activity is affected by chromatin accessibility[17,18]. Further studies using Cas9 or Cpf1 are warranted to address this issue.

We noted that the OT6 off-target site harbored a non-canonical, 5′-TCTN-3′ PAM sequence. OT6 and OT12 (which differ by one nucleotide from each other at the 3′ end) site-specific crRNAs were able to induce indels at the OT6 site with a frequency of 3.8% and 8.5%, respectively. This result suggests that Cpf1 can cleave

© 2016 Nature America, Inc. All rights reserved.

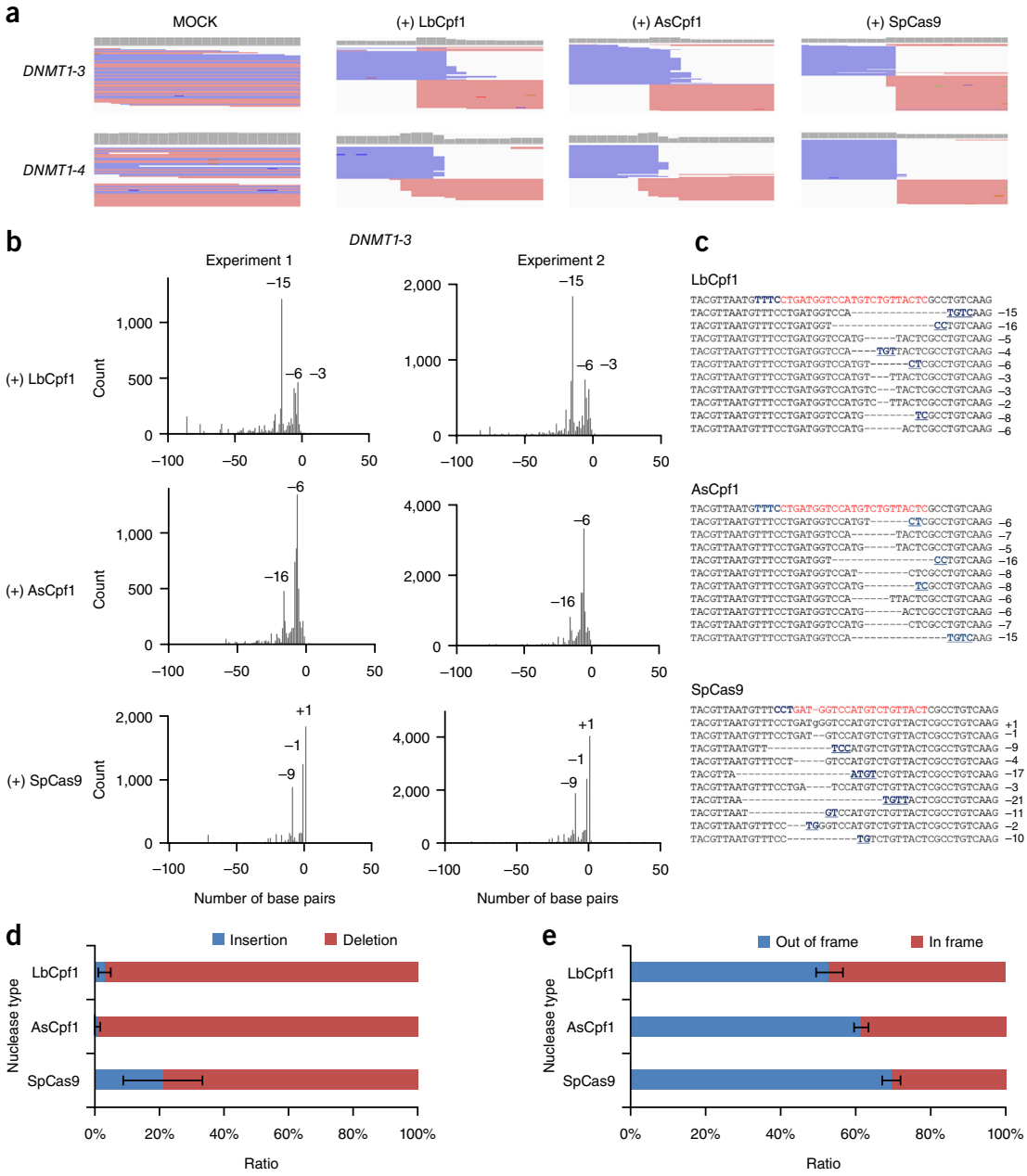


Figure 4 Overhang patterns and mutation signatures of Cpf1 and SpCas9 RGENs. (**a**) Representative IGV images showing overhang patterns at on-target cleavage sites. (**b**) The number of mutant sequence reads binned by the deletion or insertion size in base pairs. (**c**) Mutant sequences induced at the Cpf1 or Cas9 target site. PAM sequences are shown in blue. crRNA and sgRNA target sites are shown in red. Microhomology sequences are underlined and in blue. The number of deleted or inserted bases is shown on the right. (**d**,**e**) Mutation signatures of LbCpf1, AsCpf1 and SpCas9. Mutant sequences are divided into two fractions, deletions versus insertions (**d**) or in-frame versus out-of-frame indels (**e**). Data represent mean ± s.e.m. (n = 10 target sites).

chromosomal target sites that contain non-canonical PAMs, which increases the range of Cpf1-mediated genome editing.

To reduce off-target effects, we transfected preassembled Cpf1 RNPs into HEK293 cells. We assumed that Cpf1 RNPs, similar to Cas9 RNPs, would cleave target sites immediately after transfection and would be degraded rapidly by endogenous proteases and RNases, thereby reducing off-target effects without sacrificing on-target effects[19]. Indeed, Cpf1 RNPs did not induce any indels above basal sequencing error levels at four off-target sites that were validated with plasmids (**Fig. 3f**). OTIs for AsCpf1 RNPs and LbCpf1 RNPs were <0.0004, suggesting that these RNPs had virtually no off-target effects. We also found that crRNAs truncated at the 3′ end (termed tru-crRNAs)[20] enabled an up to ninefold reduction of indel frequencies at certain off-target sites (**Supplementary Fig. 9**). Unfortunately, this approach is limited, as most off-target sites harbor mismatches at the PAM-distal 3′ end.

One important advantage of Digenome-seq over other (cell-based) methods[12,21–23] is that overhang patterns at cleavage sites can be observed using Integrative Genomics Viewer (IGV)[24] (**Fig. 4a**). LbCpf1 produced mainly 5′ 3-nt overhangs, but no 2-nt overhangs at cleavage

© 2016 Nature America, Inc. All rights reserved.

sites. AsCpf1, however, produced 5′ 2–4-nt overhangs. SpCas9 yielded blunt ends and 5′ 1-nt overhangs. We examined whether these different overhang patterns produced different mutation signatures[25] (**Fig. 4b,c** and **Supplementary Fig. 10**). Notably, LbCpf1, AsCpf1 and Cas9 induced different edited sequences, although microhomologies were often found at deletion junctions. Single-base insertions or deletions were very rare with Cpf1 nucleases, but were more common than other mutation patterns with SpCas9. These results indicate that differences in cleavage positions of Cpf1 and SpCas9 and overhang patterns result in different mutation signatures. We also note that, unlike SpCas9, Cpf1 rarely induced insertions (**Fig. 4d**). Consequently, in-frame mutations caused by 3-nt, 6-nt, 9-nt or other triple nucleotide combination deletions were more likely to be induced by Cpf1 than by Cas9 (**Fig. 4e**). This suggests that a microhomology-based choice of target sites is more important with Cpf1 than with Cas9 if one wishes to inactivate a protein-coding gene[26].

In summary, our data indicate that Cpf1 is a highly specific programmable nuclease that is suitable for precision genome editing, although it is less efficient than SpCas9 in certain human cell lines. Cpf1 orthologs, or variants, with improved efficiency or altered PAM specificity may be identified or engineered in the near future. Finally, we note that the marked specificity of Cpf1 is paradoxical; it is an open question how a highly specific Cpf1 can function as an adaptive immune system in bacteria cells against hypervariable foreign genetic elements, such as phages.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. NCBI Sequence Read Archive: SRP070794.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by IBS-R021-D1. We thank F. Zhang (MIT) for providing human codon-optimized Cpf1 expression plasmids.

AUTHOR CONTRIBUTIONS

J.-S.K. supervised the research. D.K., J.K., J.K.H., K.W.B. and S.Y. carried out the experiments. D.K. performed bioinformatics analyses.

COMPETING FINANCIAL INTERESTS

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

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Cho, S.W., Lee, J., Carroll, D., Kim, J.S. & Lee, J. Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics* **195**, 1177–1180 (2013).
2. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
3. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
4. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
5. 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).
6. Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
7. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
8. Kim, H. & Kim, J.S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15**, 321–334 (2014).
9. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
10. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
11. Kim, D., Kim, S., Kim, S., Park, J. & Kim, J.S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Res.* **26**, 406–415 (2016).
12. Ran, F.A. *et al. In vivo* genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).
13. Bae, S., Park, J. & Kim, J.S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
14. Kim, D. *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat. Methods* **12**, 237–243 (2015).
15. Slaymaker, I.M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
16. Kleinstiver, B.P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
17. Horlbeck, M.A. *et al.* Nucleosomes impede Cas9 access to DNA *in vivo* and *in vitro*. *eLife* **5**, e12677 (2016).
18. Isaac, R.S. *et al.* Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* **5**, e13450 (2016).
19. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24**, 1012–1019 (2014).
20. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).
21. Wang, X. *et al.* Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* **33**, 175–178 (2015).
22. Tsai, S.Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
23. Frock, R.L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* **33**, 179–186 (2015).
24. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
25. Kim, Y., Kweon, J. & Kim, J.S. TALENs and ZFNs are associated with different mutation signatures. *Nat. Methods* **10**, 185 (2013).
26. Bae, S., Kweon, J., Kim, H.S. & Kim, J.S. Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* **11**, 705–706 (2014).

© 2016 Nature America, Inc. All rights reserved.

ONLINE METHODS

Cas9 and Cpf1 ribonucleoproteins. Recombinant Cas9 protein was purchased from ToolGen. SpCas9 sgRNAs and Cpf1 crRNAs were synthesized by *in vitro* transcription using T7 RNA polymerase as described previously[19]. For preparation of recombinant Cpf1 proteins, the protein sequences were codon optimized for *E.coli* expression. Synthetic oligonucleotide fragments were cloned into bacterial protein expression vectors that encode 6xHis and maltose-binding protein tags (pMAL-c5x, New England BioLabs; and pDEST-hisMBP)[27]. The resulting expression vectors were used to transform Rosetta expression cells (EMD Millipore). 2 l of Luria broth (LB) growth medium with 50 mg/ml carbenicilin were inoculated with 10 ml overnight culture of Rosetta cells containing Cpf1 plasmids. Cells were cultured at 37 °C until OD600 reached 0.6, and then cooled to 16 °C before induction with 0.5 mM IPTG for 14-18 h. Cells were then harvested and frozen at −80 °C until protein purification.

For protein purification, the cell pellet was lysed by sonication in 50 ml of lysis buffer (50 mM, HEPES pH 7, 200 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 10 mM imidazole) supplemented with lysozyme (Sigma) and protease inhibitor (Roche complete, EDTA-free). The cell lysate was cleared by centrifugation at 16,000*g* for 30 min, followed by passage through a syringe filter (0.22 μm). The cleared lysate was applied to a nickel column (Ni-NTA agarose, Qiagen), washed with 2 M salt, and eluted with 250 mM imidazole. The eluted protein solution was buffer exchanged and concentrated with lysis buffer without magnesium and imidazole. Purified Cpf1 proteins were examined with SDS-PAGE. Cpf1 protein activity was assessed by an *in vitro* cleavage assay.

Plasmids encoding Cpf1 and crRNAs. Plasmids encoding human codon optimized Cpf1 proteins[10] were obtained from Addgene. U6 promoter-based crRNA expression plasmids were modified from the sgRNA expression plasmid described previously[28]. Briefly, the repeat region of the sgRNA was replaced with crRNA repeat sequences.

Cell culture and transfection conditions. HEK293T cells (ATCC, CRL-11268) were maintained in DMEM medium supplemented with 10% FBS and 1% antibiotics. For Cpf1-mediated genome editing, HEK293T cells were seeded into 24-well plates at 70–80% confluency before transfection and transfected with Cpf1 expression plasmids (500 ng) and crRNA plasmids (500 ng) using lipofectamine 2000 (Invitrogen). Genomic DNA was isolated with the DNeasy Blood & Tissue Kit (Qiagen) 72 h post-transfection. Cells were not tested for mycoplasma contamination.

T7 endonuclease I (T7E1) assay. PCR amplicons containing nuclease target sites were denatured and cooled down to form heteroduplex DNA using a thermal cycler. Reannealed PCR amplicons were incubated with T7 endonuclease I (New England BioLabs) for 30 min at 37 °C and subjected to agarose gel electrophoresis. Gel images were analyzed using a densitometer to quantitate DNA band intensities. Indel frequencies were calculated using the following equation, Indel frequency (%) = $100 \times [1 - \{1 - (\text{cleaved fraction})\}^{1/2}]$.

Targeted deep sequencing. Target sites and potential off-target sites were analyzed by targeted deep sequencing. Deep sequencing libraries were generated by PCR. TruSeq HT Dual Index primers were used to label each sample. Pooled libraries were subjected to paired-end sequencing using MiSeq (Illumina).

Digenome preparation: *In vitro* cleavage of genomic DNA. Genomic DNA was purified from HeLa cells with the DNeasy Tissue kit (Qiagen). The Cpf1 protein (300 nM) and crRNA (900 nM) were pre-incubated at room temperature for 10 min to form RNP complexes. Genomic DNA (8 μg) was incubated with RNP complexes in a reaction buffer (100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 μg/ml BSA, pH 7.9) at 37 °C for 8 h. Digested genomic DNA was treated with RNase A (50 μg/ml) to degrade sgRNAs and purified again with the DNeasy Tissue kit (Qiagen).

Whole genome and digenome sequencing. Cas9- or Cpf1-digested genomic DNA was subjected to WGS, which was performed at a sequencing depth of 30× to 40× using an Illumina HiSeq X Ten Sequencer at Macrogen (South Korea), as described previously[11]. A DNA cleavage score was assigned to each nucleotide position across the entire genome, using WGS data, according to the equation presented in Kim *et al.*[11] (Cas9) and that in **Supplementary Figure 12** (Cpf1). These equations assume that Cas9 produces 5′ and 3′ 1- to 2-nt overhangs in addition to blunt ends and that Cpf1 produces 5′ 1- to 5-nt overhangs. *In vitro* cleavage sites with DNA cleavage scores above the cutoff value of 2.5 were computationally identified.

27. Nallamsetty, S., Austin, B.P., Penrose, K.J. & Waugh, D.S. Gateway vectors for the production of combinatorially-tagged His6-MBP fusion proteins in the cytoplasm and periplasm of *Escherichia coli*. *Protein Sci.* **14**, 2964–2971 (2005).
28. Cho, S.W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).

© 2016 Nature America, Inc. All rights reserved.