

METHOD

Open Access



# Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides

Sojung Kim<sup>1</sup>, Taegeun Bae<sup>1,2</sup>, Jaewoong Hwang<sup>3</sup> and Jin-Soo Kim<sup>1,2,3\*</sup>

## Abstract

We report that engineered Cas9 variants with improved specificity—eCas9-1.1 and Cas9-HF1—are often poorly active in human cells, when complexed with single guide RNAs (sgRNAs) with a mismatch at the 5' terminus, relative to target DNA sequences. Because the nucleotide at the 5' end of sgRNAs, expressed under the control of the commonly-used U6 promoter, is fixed to a guanine, these attenuated Cas9 variants are not useful at many target sites. By using sgRNAs with matched 5' nucleotides, produced by linking them to a self-cleaving ribozyme, the editing activity of Cas9 variants can be rescued without sacrificing high specificity.

**Keywords:** CRISPR-Cas, Off-target effect, Engineered Cas9 variants, Hammerhead ribozyme-linked sgRNA

## Background

Clustered, regularly interspaced, short palindromic repeats (CRISPR) – CRISPR-associated (Cas) RNA-guided endonucleases, derived from adaptive immune systems in bacteria and archaea, have been repurposed for targeted genome editing in various cells and organisms [1–4]. These nucleases cleave chromosomal DNA in a targeted manner, producing site-specific DNA double-strand breaks (DSBs), the repair of which via non-homologous end-joining (NHEJ) induces small insertions or deletions (indels) at target sites. Unfortunately, off-target DNA cleavage at sites that are highly homologous to on-target sites can lead to mutations at undesired genomic loci [5, 6] and to chromosomal rearrangements such as translocations and inversions [6–8]. Both *S. pyogenes* Cas9 [9, 10] and single guide RNAs (sgRNAs) [6, 11, 12] have been modified to minimize or eliminate these off-target effects. In particular, two groups have independently presented engineered Cas9 variants, termed enhanced Cas9-1.1 (eCas9-1.1) [9] and Cas9 high-fidelity variant 1 (Cas9-HF1) [10], with minimal or no detectable off-target effects in human cells. These high-specificity Cas9 variants contain alanine substitutions to weaken non-specific ionic interactions between the Cas9 protein and the non-target or target DNA strand.

Here, we show that these attenuated Cas9 variants are poorly active at sites with a mismatched 5' nucleotide relative to their sgRNA sequences in human cells. By using sgRNAs with matched 5' nucleotides relative to their target DNA sequences, generated by self-cleaving ribozyme fusion, the cleavage activity of the Cas9 variants was rescued in human cells without sacrificing their high specificities.

## Results

We hypothesized that the attenuated Cas9 variants might be poorly active at sites with a mismatch at the 5' terminus. Because the U6 promoter, which is commonly used to express sgRNAs in eukaryotic cells, requires a guanosine (G) nucleotide to initiate transcription, sgRNAs typically contain a G nucleotide at the 5' terminus. Three out of four DNA target sites will contain a mismatch at this position and thus might be poorly edited in cells by attenuated Cas9 variants in complex with gX<sub>19</sub> sgRNAs (Fig. 1a), where “g” or “G” is a mismatched or matched guanosine, respectively. Note that high-specificity Cas9 variants have been previously tested at target sites with a G nucleotide at the 5' end of the target DNA strand using GX<sub>19</sub> sgRNAs [9, 10].

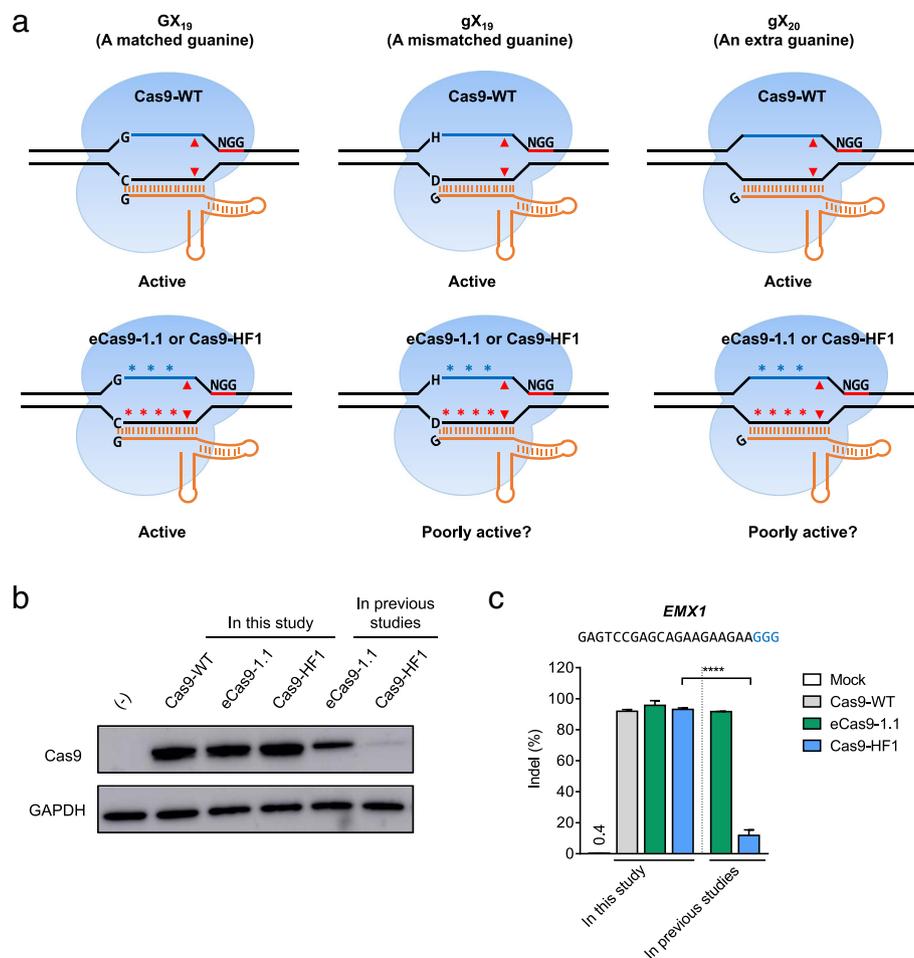
Before testing this hypothesis, we compared expression levels of Cas9 variants in human cells and found that the two variants, especially Cas9-HF1, were poorly expressed in HeLa cells (Fig. 1b). We noted that our plasmid encoding wild-type Cas9 (termed Cas9-WT hereinafter) and the two plasmids encoding Cas9 variants [9, 10] contained different promoters, tags, and

\* Correspondence: jskim01@snu.ac.kr

<sup>1</sup>Center for Genome Engineering, Institute for Basic Science, Seoul, South Korea

<sup>2</sup>IBS School, Korea University of Science and Technology, Seoul, South Korea  
Full list of author information is available at the end of the article





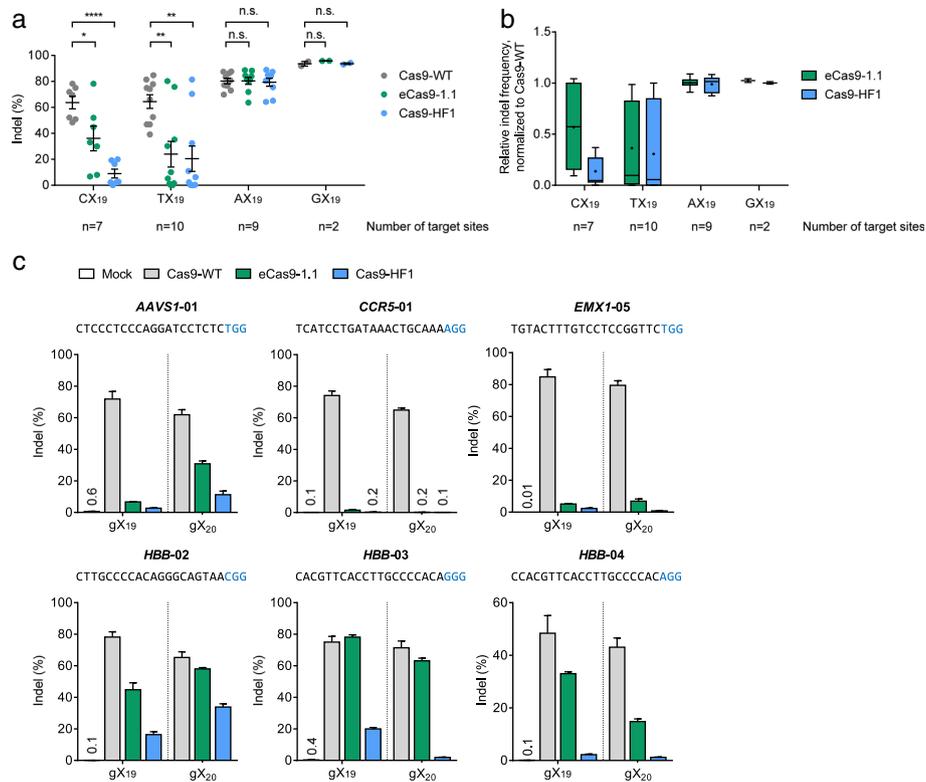
**Fig. 1** High-specificity Cas9 variants with attenuated cleavage activity. **a** Schematics of Cas9-WT, engineered Cas9 variants, and sgRNA variants. Locations of the alanine substitutions introduced in Cas9-WT to create eCas9-1.1 or Cas9-HF1 are shown with blue or red asterisks, respectively. Red triangles indicate cleavage positions. GX<sub>19</sub> sgRNA starts with a G matched to its protospacer (blue line). gX<sub>19</sub> sgRNA has a mismatched G at its 5'-end. gX<sub>20</sub> sgRNA includes an extra guanine at its 5' terminus. The protospacer-adjacent motif (PAM) is shown by a red line labeled NGG. H, not G (A or C or T); D, not C (A or G or T). **b**, **c** Western blot (**b**) and indel frequencies at the *EMX1* site (**c**) of HeLa cells co-transfected with plasmids encoding Cas9-WT or variants and *EMX1*-specific sgRNA. Cas9 variants were expressed by new plasmids used in this study or by old plasmids used in previous studies. Error bars, s.e.m of two or three biological replicates

codon sequences [4, 6]. We performed site-directed mutagenesis in our Cas9-WT plasmid to obtain constructs encoding the two high-specificity Cas9 variants. Western blot analysis showed that all three proteins were highly expressed in HeLa cells (Fig. 1b). Consistent with this result, eCas9-1.1 and Cas9-HF1 expressed using the new constructs induced indels at the *EMX1* site with efficiencies comparable to Cas9-WT (Fig. 1c). Based on these results, we used the newly cloned constructs to express eCas9-1.1 and Cas9-HF1 throughout this study.

#### Reduced editing activity of high-fidelity Cas9 variants at target sites with a mismatched 5' nucleotide

To test whether the attenuated Cas9 variants are poorly active at sites with a mismatch at the 5' terminus, we compared editing activities of eCas9-1.1 and Cas9-HF1

with those of Cas9-WT at 26 sites whose 5' terminal nucleotides are not guanosine in HeLa cells using gX<sub>19</sub> sgRNAs: the sites with a 5' cytosine (C) were termed CX<sub>19</sub> (seven sites); those with a 5' thymine (T), TX<sub>19</sub> (ten sites); and those with a 5' adenosine (A), AX<sub>19</sub> (nine sites) (Fig. 2a and b, Additional file 1: Table S1 and Table S2). As expected, Cas9-WT was not sensitive to the mismatch at the 5' end, inducing indels at high frequencies (64 ± 5% at CX<sub>19</sub> sites; 65 ± 5% at TX<sub>19</sub> sites; 80 ± 2% at AX<sub>19</sub> sites, on average). eCas9-1.1 showed much lower indel frequencies at CX<sub>19</sub> sites (36 ± 10%) and TX<sub>19</sub> sites (24 ± 10%), a 1.8-fold or 2.7-fold reduction in average indel frequencies at CX<sub>19</sub> or TX<sub>19</sub> sites, respectively. Cas9-HF1 was least active among the three Cas9 nucleases, with average indel frequencies of 9.0 ± 3% at CX<sub>19</sub> sites and 20 ± 10% at TX<sub>19</sub> sites, which corresponds to 7.1-fold and 3.2-fold reductions,



**Fig. 2** Reduced editing activity of high-fidelity Cas9 variants at target sites with a non-G 5' nucleotide. **a** Indel frequencies of Cas9-WT and Cas9 variants at 26 endogenous target sites with an HX<sub>19</sub> sequence obtained using gX<sub>19</sub> sgRNAs in HeLa cells (H: not G [C or T or A]). The sequences and indel frequencies (%) of the 26 target sites are described in Additional file 1: Table S1. \*  $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*\*  $P < 0.0001$ . **b** Distributions of relative indel frequencies of eCas9-1.1 and Cas9-HF1 normalized to that of Cas9-WT. Means and medians of relative indel frequencies are represented in Additional file 1: Table S2. *Box and whisker plots*: center lines show the medians; crosses show the means. **c** Indel frequencies (measured by targeted deep sequencing) at six on-target sites with a 5' C or 5' T nucleotide obtained using gX<sub>19</sub> and gX<sub>20</sub> sgRNAs in HeLa cells. PAM sequences are shown in blue. Error bars, s.e.m. of three biological replicates

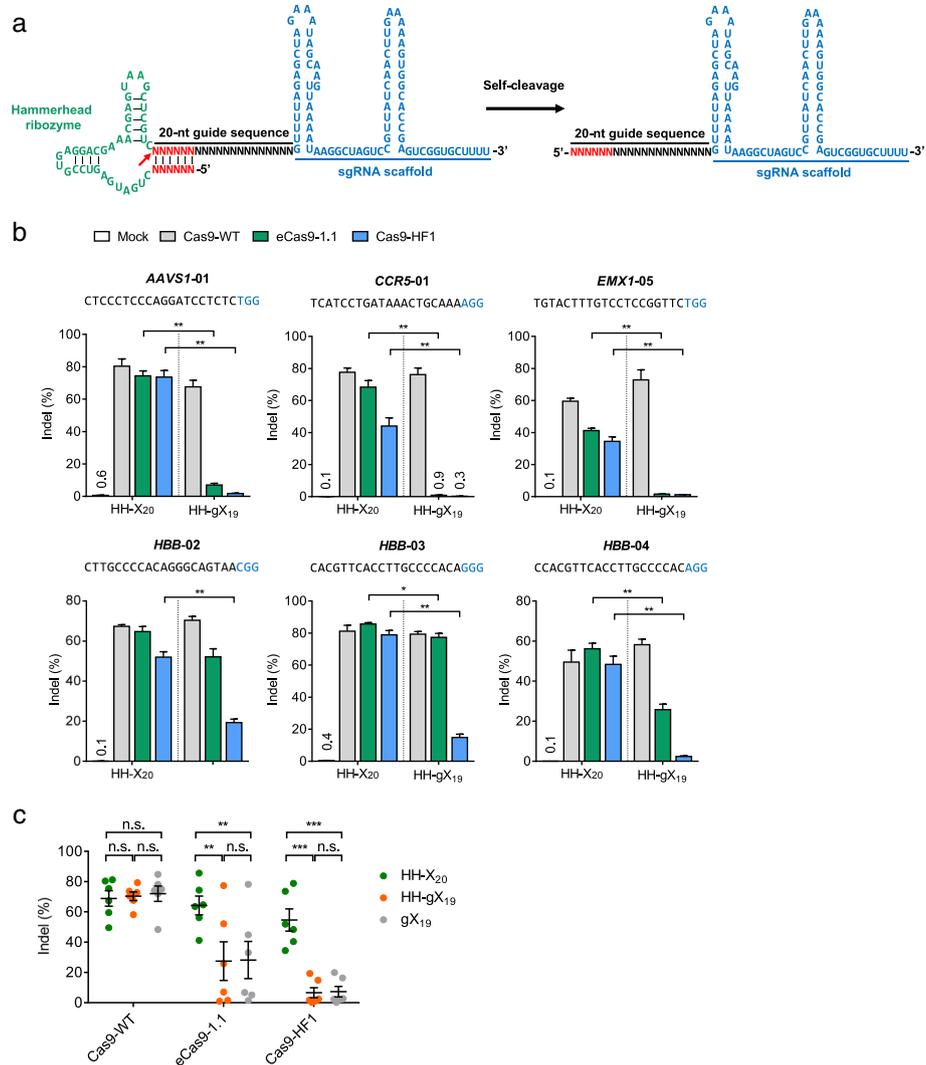
respectively. At AX<sub>19</sub> target sites, however, both eCas9-1.1 and Cas9-HF1 showed indel efficiencies ( $81 \pm 3\%$  and  $79 \pm 3\%$ ) comparable to that of Cas9-WT ( $80 \pm 2\%$ ), suggesting that a G:T mismatch at the 5' terminus may still form a wobble base pair. These results are in line with a previous report showing Cas9-HF1 activities with three and one gX<sub>19</sub> sgRNAs at CX<sub>19</sub> and AX<sub>19</sub> sites, respectively [10].

We chose six CX<sub>19</sub> or TX<sub>19</sub> target sites at which the two Cas9 variants were poorly active and tested gX<sub>20</sub> sgRNAs with an extra guanosine at the 5' terminus rather than gX<sub>19</sub> sgRNAs (Fig. 2c). Note that gX<sub>20</sub> sgRNAs, unlike gX<sub>19</sub> sgRNAs, have matched nucleotides at the 5' end. Use of gX<sub>20</sub> sgRNAs enhanced the activity of Cas9 variants at AAVS1-01 and HBB-02 sites but reduced the activity at the other four sites, compared to gX<sub>19</sub> sgRNAs. We also noted that Cas9-WT was more efficient with gX<sub>19</sub> sgRNAs than with gX<sub>20</sub> sgRNAs at all six sites. These results show that gX<sub>20</sub> sgRNAs cannot rescue the genome editing activities of high-specificity Cas9 variants. These Cas9 variants in combination with gX<sub>19</sub> and gX<sub>20</sub> sgRNAs also showed lower indel

frequencies than Cas9-WT in HEK293T, another human cell line (Additional file 1: Figure S1).

### Rescue of high-specificity Cas9 variants using Hammerhead ribozyme-linked sgRNAs

To expand the utility of high-fidelity Cas9 variants, we produced sgRNAs with matched 5' nucleotides by using a self-cleaving ribozyme. Thus, each sgRNA was fused to a Hammerhead (HH) ribozyme at its 5'-end [13], which generates mature 20-nucleotide (X<sub>20</sub>) sgRNAs after self-cleavage (Fig. 3a). HH ribozyme-fused sgRNAs with matched 5' nucleotides (termed HH-X<sub>20</sub>) or the mismatched 5' guanosine nucleotide (termed HH-gX<sub>19</sub>) were tested in combination with Cas9-WT and high-fidelity Cas9 variants in HeLa cells (Fig. 3b). Use of HH-X<sub>20</sub> sgRNAs rescued the activity of the two Cas9 variants at all six target sites. Thus, indel frequencies obtained with eCas9-1.1 ( $64 \pm 6\%$ ) and Cas9-HF1 ( $55 \pm 7\%$ ) using HH-X<sub>20</sub> sgRNAs were comparable to those obtained with Cas9-WT ( $69 \pm 5\%$  or  $70 \pm 3\%$ ) using HH-X<sub>20</sub> sgRNAs or HH-gX<sub>19</sub> sgRNAs, respectively (Fig. 3c and



**Fig. 3** Recovery of editing efficiency of high-fidelity Cas9 variants using HH ribozyme-linked sgRNAs. **a** A schematic of a self-processing ribozyme fused sgRNA. The pre-sgRNA contains a HH ribozyme at its 5'-end. The pre-sgRNA undergoes self-cleavage to release a mature sgRNA. The red arrow indicates the self-cleavage position. **b** HH ribozyme-fused sgRNAs with a matched 5' nucleotide (HH-X<sub>20</sub>) or a mismatched guanosine (HH-gX<sub>19</sub>) were tested in combination with Cas9-WT and high-fidelity Cas9 variants at six target sites in HeLa cells. Indel frequencies were measured using targeted deep sequencing. The PAM is shown in blue. Error bars, s.e.m. of three biological replicates. Statistical significances were calculated by unpaired t-test. \*  $P < 0.05$ , \*\*  $P < 0.01$ . **c** Mean indel frequencies  $\pm$  s.e.m. at the six target sites in HeLa cells. Statistical significances were calculated by paired t-test. \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$

Additional file 1: Table S3a). The ratios of indel frequencies of Cas9 variants in complex with HH-X<sub>20</sub> sgRNAs relative to that of Cas9-WT had medians of 0.9 for eCas9-1.1 and 0.8 for Cas9-HF1 (Additional file 1: Figure S2, Table S3b). The two Cas9 variants were poorly active when combined with HH-gX<sub>19</sub> sgRNAs, demonstrating that the rescue of high-fidelity variants was due to matched nucleotides at the 5' end rather than the ribozyme fusion itself. As expected, Cas9-WT was equally efficient with HH-X<sub>20</sub>, HH-X<sub>19</sub>, and gX<sub>19</sub> sgRNAs ( $P = 0.36$ , HH-X<sub>20</sub> vs HH-X<sub>19</sub>;  $P = 0.28$ , HH-X<sub>20</sub> vs gX<sub>19</sub>;  $P = 0.31$ , HH-X<sub>19</sub> vs gX<sub>19</sub>) (Fig. 3c and

Additional file 1: Table S3a). Editing efficiencies of eCas9-1.1 and Cas9-HF1 were also increased with HH-X<sub>20</sub> sgRNAs in HEK293T cells (Additional file 1: Figure S3 and Table S4).

#### Specificities of Cas9-WT and high-fidelity Cas9 variants in combination with HH-X<sub>20</sub> sgRNAs

Next, we compared the specificities of the two Cas9 variants in complex with HH-X<sub>20</sub> sgRNAs by measuring mutation frequencies at known off-target sites in HeLa cells. The CCR5-01- and EMX1-05-specific sgRNAs have no known off-target sites and were

excluded from this analysis. At most of the off-target sites that differed from their respective on-target sites by one to three nucleotides, the two Cas9 variants showed much lower indel frequencies than Cas9-WT (Fig. 4). Of note, Cas9-HF1 was able to discriminate against three off-target sites (one *HBB-03* off-target site and two *HBB-04* off-target sites), each with a single nucleotide mismatch. These results show that attenuated Cas9 variants retain their high specificities when combined with HH- $X_{20}$  sgRNAs. eCas9-1.1 and Cas9-HF1 also showed high specificities with HH- $X_{20}$  sgRNAs in HEK293T cells (Additional file 1: Figure S4).

## Discussion and conclusions

In summary, we showed here that newly developed, high-specificity Cas9 variants, unlike the wild-type (WT) protein, are often inefficient at target sites with a mismatch at the 5' terminus, unequivocally demonstrating the contribution of the 5' nucleotide to the high specificity of CRISPR-Cas9 in human cells for the first time. Of note, a single 5'-end mismatch between the sgRNA and target DNA is largely tolerated by Cas9-WT. The two attenuated Cas9 variants, however, contain several alanine substitutions to weaken ionic interactions between the protein and target DNA, which can make them sensitive to a single mismatch even at the 5' terminus. By matching the first nucleotide of sgRNAs to target DNA via the self-cleaving activity of a HH-ribozyme fusion, highly

specific genome editing was achieved without sacrificing on-target editing efficiency. As an alternative to using a HH-ribozyme fusion, sgRNAs with matched 5' non-G nucleotides could be created via tRNA fusion [14] or chemical synthesis [15] and combined with the two high-fidelity Cas9 variants. Delivery of pre-assembled Cas9 variant ribonucleoproteins [16] rather than Cas9- and sgRNA-encoding plasmids may further improve genome-wide target specificities of CRISPR genome editing. Our method expands targetable sites for high-specificity Cas9 variants, allowing broad applications in research and medicine.

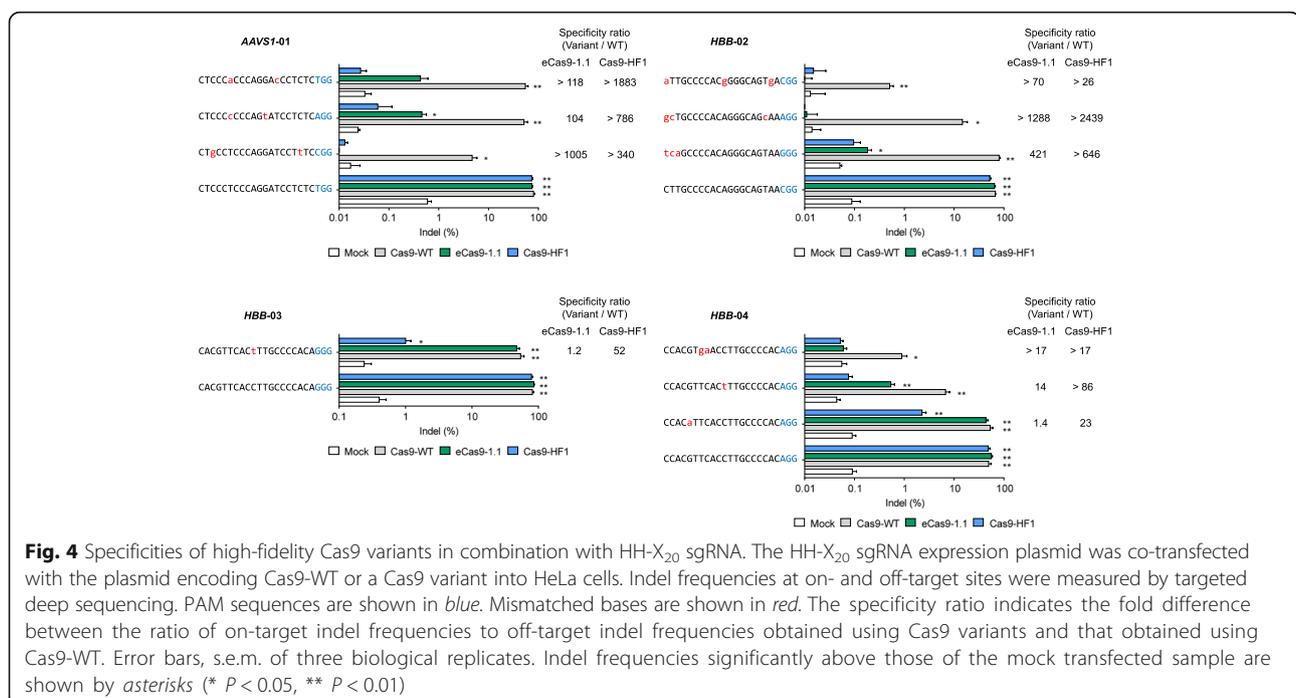
## Methods

### Construction of high-fidelity Cas9 variant-encoding plasmids and the HH-ribozyme-fused sgRNA-encoding plasmid

eCas9-1.1- and Cas9-HF1-encoding plasmids (p3s-eCas9-1.1, Addgene #104172; p3s-Cas9-HF1, Addgene #104173) were created via site-directed mutagenesis of a WT Cas9 construct (p3s-Cas9-HN, Addgene #104171). HH-ribozyme sgRNA constructs were cloned via ligation of annealed oligonucleotides that included a HH-ribozyme sequence and a protospacer sequence into a plasmid (pRG2, Addgene #104174) in which sgRNA expression is under the control of the U6 promoter.

### Cell culture and transfection

HeLa cells (ATCC, CCL-2) and HEK 293 T/17 cells (ATCC, CRL-11268) were maintained in Dulbecco's



modified Eagle's medium (DMEM) supplemented with 100 units/mL penicillin, 100 µg/mL streptomycin, 0.1 mM non-essential amino acids, and 10% fetal bovine serum (FBS).  $0.8 \times 10^5$  HeLa cells and  $2 \times 10^5$  HEK293T/17 cells were transfected with the Cas9-encoding plasmid (0.5 µg) and sgRNA expression plasmid (0.5 µg) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol.

### Western blotting

The Cas9-WT and Cas9 variant proteins expressed in HeLa cells after transfection were detected using western blotting. Cas9 and GAPDH were detected using anti-Cas9 (Abcam, ab191468) and anti-GAPDH (Santa Cruz Biotechnology, sc-32233) primary antibodies. Goat anti-mouse IgG-HRP antibody (Santa Cruz Biotechnology, sc-2005) was used for signal detection. ImageQuant LAS4000 (GE healthcare) was used for digital imaging.

### Targeted deep sequencing

The on-target and off-target regions were polymerase chain reaction (PCR) amplified for NGS library construction. Pooled PCR amplicons were sequenced using MiniSeq with a TruSeq HT Dual Index system (Illumina). Indel frequencies were obtained using Cas-Analyzer [17].

### Additional file

**Additional file 1: Figure S1.** Comparison of editing efficiencies of Cas9-WT and high-fidelity Cas9 variants using gX19 and gX20 sgRNAs in HEK293T cells. **Figure S2.** Comparison of relative indel frequencies of Cas9 variants normalized to that of Cas9-WT in HeLa cells. **Figure S3.** Comparison of indel frequencies of HH-X<sub>20</sub>, HH-gX<sub>19</sub>, and gX<sub>19</sub> sgRNAs in combination with Cas9-WT and Cas9 variants in HEK293T cells. **Figure S4.** Specificities of high-fidelity Cas9 variants in combination with HH-X<sub>20</sub> sgRNA in HEK293T cells. **Table S1.** Indel frequencies of Cas9-WT, eCas9-1.1, and Cas9-HF1 combined with gX<sub>19</sub> sgRNAs at 26 target sites with an HX<sub>19</sub> sequence. **Table S2.** Comparison of Cas9-WT and Cas9 variants using gX<sub>19</sub> sgRNAs at 26 target sites with an HX<sub>19</sub> sequence. **Table S3.** Comparison of indel frequencies of HH-X<sub>20</sub>, HH-gX<sub>19</sub>, and gX<sub>19</sub> sgRNAs in combination with Cas9-WT and Cas9 variants in HeLa cells. **Table S4.** Comparison of indel frequencies of HH-X<sub>20</sub>, HH-gX<sub>19</sub>, and gX<sub>19</sub> sgRNAs in combination with Cas9-WT and Cas9 variants in HEK293T cells. (PDF 497 kb)

### Abbreviations

Cas9-HF1: Cas9 high-fidelity variant 1; Cas9-WT: Wild-type Cas9; DSBs: DNA double strand breaks; eCas9-1.1: Enhanced Cas9-1.1; HH: Hammerhead; Indels: Insertions or deletions; NHEJ: Non-homologous end-joining; sgRNAs: Single guide RNAs

### Funding

This work was supported by the Institute for Basic Science (IBS-R021-D1).

### Availability of data and materials

The deep sequencing data are available at the NCBI Sequence Read Archive (SRA) under accession number SRP118952 [18].

### Authors' contributions

SK, TB, and JH performed the experiments. SK and J-SK wrote the manuscript. All authors read and approved the final manuscript.

### Ethics approval and consent to participate

No ethics approval was required for this study.

### Competing interests

J-SK is a co-founder of and holds stocks in ToolGen, Inc.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details

<sup>1</sup>Center for Genome Engineering, Institute for Basic Science, Seoul, South Korea. <sup>2</sup>IBS School, Korea University of Science and Technology, Seoul, South Korea. <sup>3</sup>Department of Chemistry, Seoul National University, Seoul, South Korea.

Received: 27 June 2017 Accepted: 2 November 2017

Published online: 15 November 2017

### References

- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816–21.
- Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, et al. RNA-guided human genome engineering via Cas9. *Science*. 2013;339:823–6.
- Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339:819–23.
- Cho SW, Kim S, Kim JM, Kim JS. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol*. 2013;31:230–2.
- Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*. 2013;31:827–32.
- Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res*. 2014;24:132–41.
- Lee HJ, Kim E, Kim JS. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res*. 2010;20:81–9.
- Lee HJ, Kweon J, Kim E, Kim S, Kim JS. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res*. 2012;22:539–48.
- Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F. Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016;351:84–8.
- Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016;529:490–5.
- Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*. 2014;32:279–84.
- Kim D, Bae S, Park J, Kim E, Kim S, Yu HR, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods*. 2015;12:237–43. 231 p following 243.
- Gao Y, Zhao Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J Integr Plant Biol*. 2014;56:343–9.
- Port F, Bullock SL. Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nat Methods*. 2016;13:852–4.
- Hendel A, Bak RO, Clark JT, Kennedy AB, Ryan DE, Roy S, et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol*. 2015;33:985–9.
- Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res*. 2014;24:1012–9.
- Park J, Lim K, Kim JS, Bae S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics*. 2017;33:286–8.
- Kim S, Bae T, Hwang J, Kim JS. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5' nucleotides. NCBI Sequence Read Archive (SRA). <https://www.ncbi.nlm.nih.gov/sra/?term=SRP118952>.