Check for updates

Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches

Majda Bratovič[1,2,3,4,10], Ines Fonfara[2,4,5,10], Krzysztof Chylinski[5,6,7], Eric J. C. Gálvez[1,4], Timothy J. Sullivan[1], Stefan Boerno[8], Bernd Timmermann[8], Michael Boettcher [1,9] and Emmanuelle Charpentier [1,2,3,4,5] ✉

The RNA-programmable DNA-endonuclease Cas9 is widely used for genome engineering, where a high degree of specificity is required. To investigate which features of Cas9 determine the sensitivity to mismatches along the target DNA, we performed in vitro biochemical assays and bacterial survival assays in *Escherichia coli*. We demonstrate that arginines in the Cas9 bridge helix influence guide RNA, and target DNA binding and cleavage. They cluster in two groups that either increase or decrease the Cas9 sensitivity to mismatches. We show that the bridge helix is essential for R-loop formation and that R63 and R66 reduce Cas9 specificity by stabilizing the R-loop in the presence of mismatches. Additionally, we identify Q768 that reduces sensitivity of Cas9 to protospacer adjacent motif-distal mismatches. The Cas9_R63A/Q768A variant showed increased specificity in human cells. Our results provide a firm basis for function- and structure-guided mutagenesis to increase Cas9 specificity for genome engineering.

C as9 is an RNA-guided endonuclease that specifically binds and cleaves DNA with complementarity to the spacer sequence of CRISPR RNA (crRNA) located upstream of a protospacer adjacent motif (PAM)[1,2]. Sufficient base pairing of the crRNA spacer of either dual-RNA (trans-activating crRNA (tracrRNA) in complex with crRNA) or single-guide RNA (sgRNA, tracrRNA fused to crRNA) with the target DNA strand leads to displacement of the nontarget strand and R-loop formation[1–5]. On completion of this process, Cas9 endonuclease domains, HNH and RuvC, cleave the target and nontarget strand, respectively, resulting in a double-strand DNA break[1,2]. Due to its programmability, CRISPR-Cas9 was adapted for a wide variety of genome engineering approaches[6–9]. However, a main concern for the application of the CRISPR-Cas9 technology is off-target cleavage that occurs when Cas9 cleaves mismatched targets; that is, targets with imperfect complementarity to the dual-RNA/sgRNA spacer. It is therefore crucial to identify regions and amino acids that determine Cas9 specificity.

Streptococcus pyogenes Cas9 (SpCas9) requires a seed sequence of 10–12 base pairs (bp) adjacent to the PAM for successful cleavage[2]. The REC3 domain senses PAM-distal mismatches and controls the transition of the HNH domain into an active conformation, which serves as a mechanism to block the cleavage in the presence of mismatches[10,11]. Alongside other strategies to reduce off-target cleavage[12], both structure-guided and random mutagenesis have been used to generate enhanced specificity Cas9 variants[10,13–19]. Three of the improved variants[10,15,17] contain mutations that increase the energetic barrier for transition of the HNH domain to the cleavage-competent state[10]. eSpCas9 (ref. [17]) and SpCas9-HF1 (ref. [15]) require one additional base pair between crRNA and target DNA for stable binding, which makes target unwinding more sensitive to

mismatches and slows cleavage in the presence of mismatches[20]. It was shown that mutation of two residues in the loop region of the bridge helix impairs target cleavage and improves specificity[21]. The conformational checkpoint through the HNH domain before cleavage and the seed sequence are the only known specificity determinants for Cas9. In addition, stable R-loop formation is an essential prerequisite for cleavage[5], but residues involved in this process have not been identified. Therefore, we investigated whether other Cas9 domains are involved in mismatch sensing and R-loop stabilization focusing on the PAM-adjacent region of the target.

We show that arginines of the Cas9 bridge helix influence guide RNA binding, and target DNA binding and cleavage. We identify two groups of arginines with opposite effects on Cas9 mismatch sensitivity. We demonstrate that R63, R66 and R70 stabilize the R-loop in the presence of PAM-adjacent mismatches thus reducing the Cas9 specificity. Additionally, we observe that Q768 mediates Cas9 insensitivity to a mismatch in target position 15 and is involved in sensitivity to PAM-distal mismatches. Combination of R63 and Q768 exchanges to alanine increased Cas9 sensitivity to mismatches in vitro, in bacteria and human cells. Our results elucidate the basis for Cas9-mediated sensing of PAM-adjacent mismatches and establish a framework to further improve the specificity of Cas9 for genome engineering applications.

Results

Influence of mismatches on binding and cleavage by Cas9. We wanted to determine how mismatches between crRNA and target DNA influence the binding and cleavage by Cas9. Therefore, we performed in vitro binding and kinetic cleavage assays to determine the binding constants and cleavage rates of SpCas9 on the protospacer 2 (*speM* superantigen) target sequence[3] (Fig. 1a). We also

[1]Max Planck Unit for the Science of Pathogens, Berlin, Germany. [2]Department of Regulation in Infection Biology, Max Planck Institute for Infection Biology, Berlin, Germany. [3]Institute for Biology, Humboldt University, Berlin, Germany. [4]Department of Regulation in Infection Biology, Helmholtz Centre for Infection Research, Braunschweig, Germany. [5]The Laboratory for Molecular Infection Medicine Sweden, Umeå Centre for Microbial Research, Department of Molecular Biology, Umeå University, Umeå, Sweden. [6]Max F. Perutz Laboratories, University of Vienna, Vienna, Austria. [7]Protein Technologies Facility, The Vienna Biocenter Core Facilities GmbH (VBCF), Vienna, Austria. [8]Sequencing Core Facility, Max Planck Institute for Molecular Genetics, Berlin, Germany. [9]Medical Faculty, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [10]These authors contributed equally: Majda Bratovič, Ines Fonfara. ✉e-mail: research-charpentier@mpiib-berlin.mpg.de
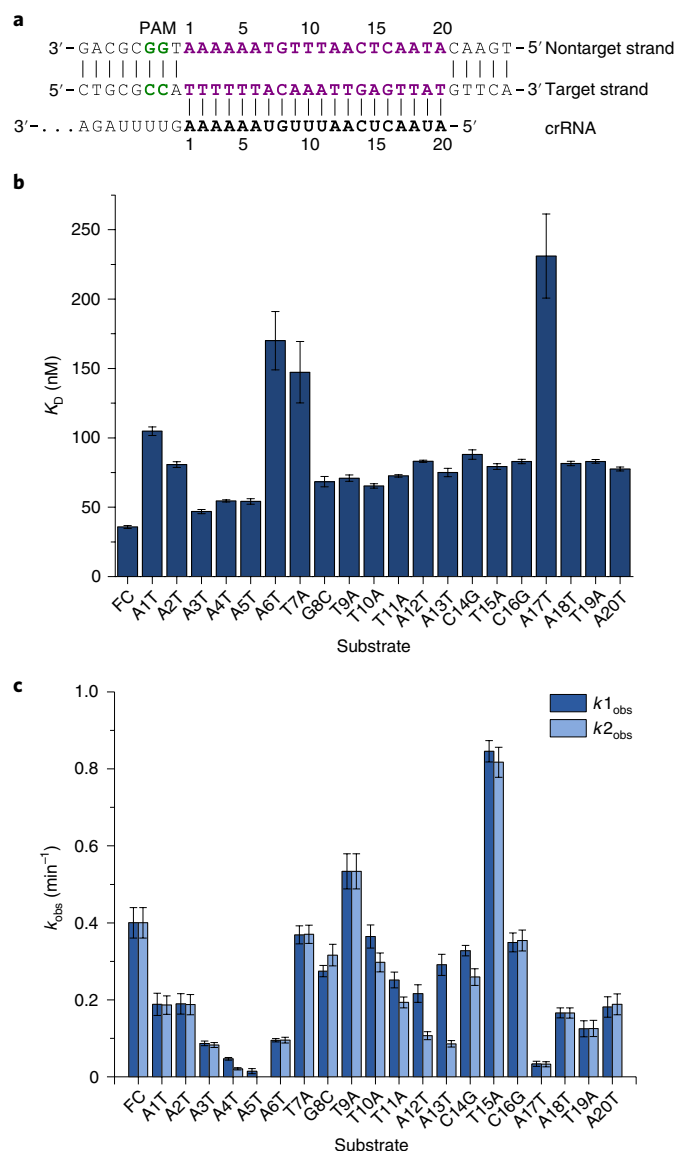
Fig. 1 | The influence of mismatches between the crRNA and target DNA on binding and cleavage by Cas9. a, Scheme of *speM* target site (purple) used in the in vitro assays. This protospacer corresponds to spacer 2, present in the type II-A CRISPR array of *S. pyogenes* SF370. Base pairing of the crRNA to the target strand is indicated (for reasons of clarity, crRNA is shown truncated and tracrRNA is not shown). Mutations were introduced in the DNA, leading to a mismatch between the crRNA spacer and the target DNA. The numbering of mismatches is indicated above the nontarget strand. The PAM is shown in green. **b**, Binding constants (K_D) derived from EMSAs with increasing concentrations of Cas9_wt and two molar excess of dual-RNA on 90 bp 5′-^{32}P labeled PCR products of the target sites used for kinetic cleavage assays. Binding constants were calculated as described in Methods. **c**, In vitro cleavage rates (k_{obs}) obtained by kinetic cleavage assays with 5 nM plasmid DNA (containing wild-type or mutated target site) and 10 nM Cas9 in complex with 20 nM dual-RNA. Mutations along the target site are indicated on the x axis, starting from the first PAM-adjacent nucleotide. The cleavage rate $k1_{obs}$ (dark blue) represents disappearance of the supercoiled plasmid DNA, whereas the cleavage rate $k2_{obs}$ (light blue) represents appearance of the linear DNA. Cleavage rates were calculated as described in Methods. **b,c**, Binding constants and cleavage rates were calculated by concatenated fitting, represented are mean values. Error bars represent the standard deviation (s.d.) ($n \geq 3$ independent experiments, Supplementary Table 1). FC, fully complementary target substrate.

investigated the influence of mismatches on cleavage by Cas9 in vivo in *E. coli*, where the target sequence is located within *rfp* on a plasmid (Methods and Supplementary Figs. 1 and 2a). Single mutations were introduced in each position of the target sequence (in vitro) (Fig. 1a) or the sgRNA (in vivo) (Supplementary Fig. 2a) to prevent base pairing between the crRNA and DNA at the corresponding position. In the in vivo bacterial toxin–antitoxin survival assay, growth is a measure of Cas9 activity; that is, when Cas9 is active, bacterial cells are unable to propagate because Cas9 cleaved the reporter plasmid encoding the antitoxin, whereas bacteria can grow when Cas9 is unable to cleave (Methods and Supplementary Fig. 1).

A mismatch in the first two PAM-adjacent nucleotides negatively affected the binding of Cas9 to the DNA (Fig. 1b). Consequently, the cleavage rates for these two targets were reduced compared to that of the target fully complementary to the crRNA spacer (fully complementary, FC) (Fig. 1c). This demonstrates that Cas9 is sensitive to mismatches in the first two target positions, in agreement with previous studies[5,22,23]. Cas9 cleaves between the third and fourth base pair of the target[2]. The cleavage of substrates with mismatches at positions 3–5 was decreased, yet these substrates were bound with an affinity comparable to the one on FC substrate, indicating an impairment affecting only catalysis. Reduced cleavage rate in presence of a mismatch at position 6 is due to poorer binding of Cas9 to this substrate (Fig. 1b). In contrast to the first six PAM-adjacent positions in vitro, complementarity at only four PAM-adjacent positions was important for cleavage in *E. coli* (Supplementary Fig. 2b). This discrepancy can be due to the different nucleotide sequences used, as the spacer sequence and the identity of the mismatch can influence the sensitivity to the mismatch[24,25]. Additionally, the differences could be explained by the measured parameters; that is, cleavage rate in vitro and end-point for bacterial survival.

We observed that on substrates with mismatches in positions 10–14, the rates for cleavage were different for each DNA strand (Fig. 1c), implying that one endonuclease domain was cleaving faster than the other. Cleavage assays on linear substrates demonstrate that cleavage by the RuvC domain was slower (Supplementary Fig. 3). Substrates with mismatches in positions 17–20 were cleaved with reduced rates compared to the FC substrate, in line with previous studies demonstrating that a certain degree of PAM-distal complementarity is necessary for cleavage[11,26,27].

Cas9 not only tolerated a mismatch in position 15, but displayed an enhanced cleavage rate (Fig. 1c). Bacteria in which Cas9-sgRNA C15G was expressed also showed reduced survival when compared to bacteria in which Cas9-sgRNA wild type (WT) was expressed, indicating that the effect is not sequence-dependent (Supplementary Fig. 2b). The same was detected when *Streptococcus mutans* Cas9 was tested for in vitro cleavage of the substrate with mismatch 15 (Supplementary Fig. 4).

Q768 is involved in sensitivity to PAM-distal mismatches. In an unrelated project, we analyzed a set of *cas9* mutants that retained activity on the FC substrate. Exchange of Q768 to alanine resulted in a cleavage rate on substrate with a mismatch at position 15 that was comparable to the cleavage rate on FC substrate. Examination of the cocrystal structure of *S. pyogenes* Cas9 (ref. [28]) revealed that the side-chain of Q768 is located at the border between RuvC and HNH domains and is in proximity to the DNA at position 15 (Fig. 2a). A mismatch might perturb the contact between Q768 and the RNA–DNA hybrid, and as a result affect the cleavage rate. To determine whether Q768 is responsible for this, we replaced it with glutamate, asparagine or alanine, and assessed Cas9 activity in *E. coli* (Fig. 2b). Expression of these Cas9 variants resulted in survival comparable to bacteria where Cas9_wt was expressed. However, in the presence of the sgRNA C15G, Cas9_Q768A, Cas9_Q768E and Cas9_Q768N showed increased survival when compared to Cas9_wt, indicating that replacement of Q768 increases the specificity of
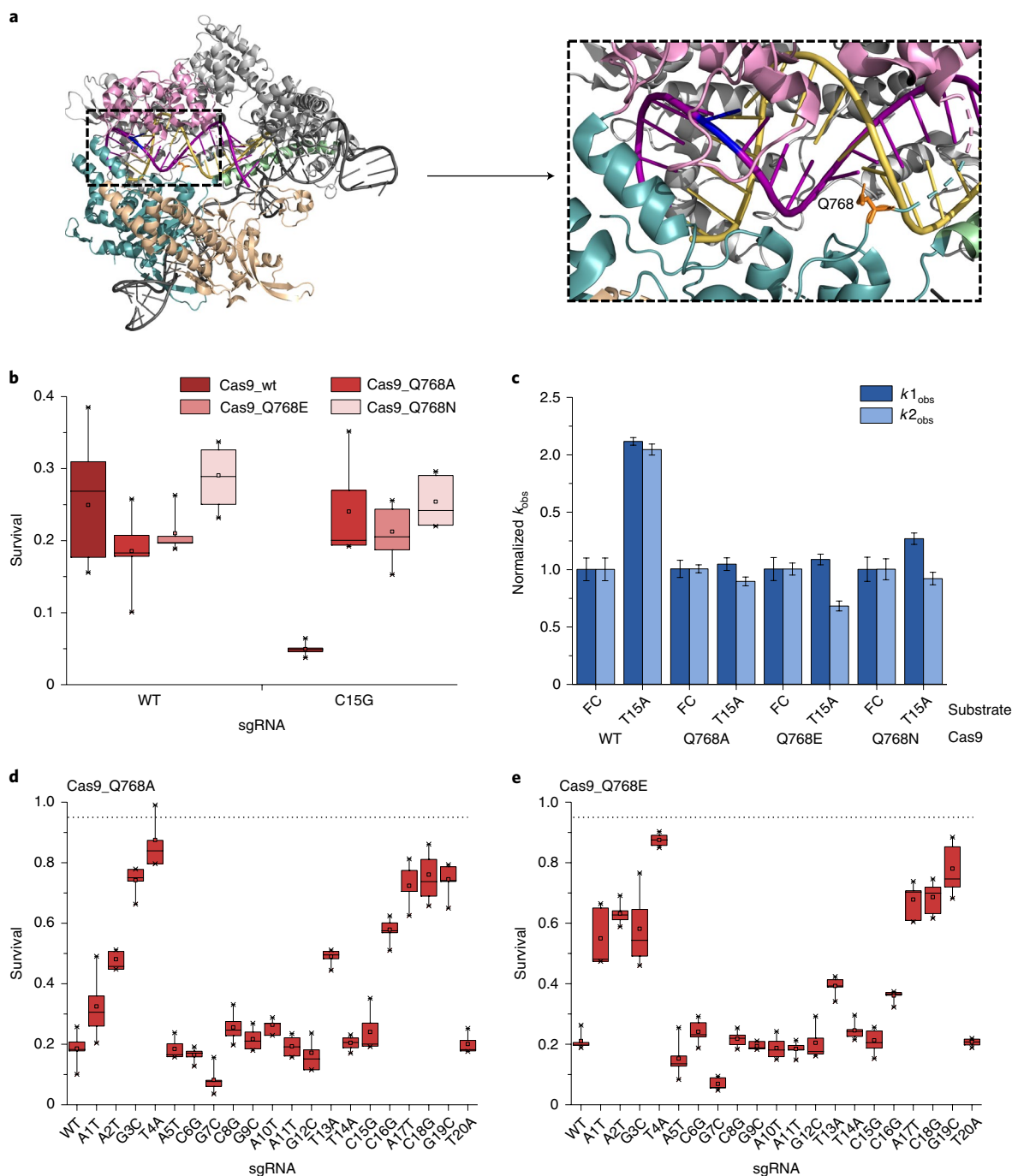
Fig. 2 | Q768 is involved in Cas9 sensitivity to PAM-distal mismatches. a, Cocrystal structure of Cas9 bound to sgRNA (spacer in yellow, sgRNA scaffold in dark gray) and DNA target strand (purple) (Protein Data Bank accession no. 4OO8). HNH motif is shown in pink, RuvC in blue, bridge helix in green, recognition lobe in gray and C-terminal domain in beige. Q768 (orange) and DNA position 15 (dark blue) are zoomed in (dashed box). The distance between Q768 and position 15 is 20.3 Å. **b,** Bacterial survival assay with Cas9_wt, Cas9_Q768N, Cas9_Q768E and Cas9_Q768A, in the presence of WT or C15G sgRNAs. Survival represents the $OD_{600\,nm}$ of induced versus suppressed expression of WT or mutant Cas9. **c,** In vitro cleavage rates of Cas9_wt and Q768 mutants on target DNA having a mismatch at position 15 normalized to the cleavage rate of each protein on the FC substrate. Cleavage rates ($k1_{obs}$ in dark blue and $k2_{obs}$ in light blue) were calculated as described in Methods using a concatenated fit, represented are mean values. Error bars represent normalized s.d. ($n \geq 3$ independent experiments, Supplementary Table 1). **d,e,** Bacterial survival assay with Cas9_Q768A (**d**) and Cas9_Q768E (**e**), in the presence of WT and mismatched sgRNAs (labeled on x axis). The dotted line at 0.95 indicates the value obtained with dCas9 (catalytically inactive Cas9) that represents no cleavage. Statistical analysis of at least four replicates was performed with Origin Software. **b,d,e,** Boxes represent the interquartile range between 25th and 75th percentiles. The median and mean ($n \geq 4$ independent experiments, Supplementary Table 2) are indicated by a horizontal line and the square, respectively. The whiskers show minimum and maximum.

Cas9 in the presence of a mismatch on position 15. In vitro cleavage rates of Cas9_Q768A and Cas9_Q768E on the FC and T15A substrates were either in the same range or slightly slower on T15A,

while for Cas9_Q768N the first cleavage rate was faster on T15A substrate compared to the FC substrate, although not to the same extent as for Cas9_wt (Fig. 2c).
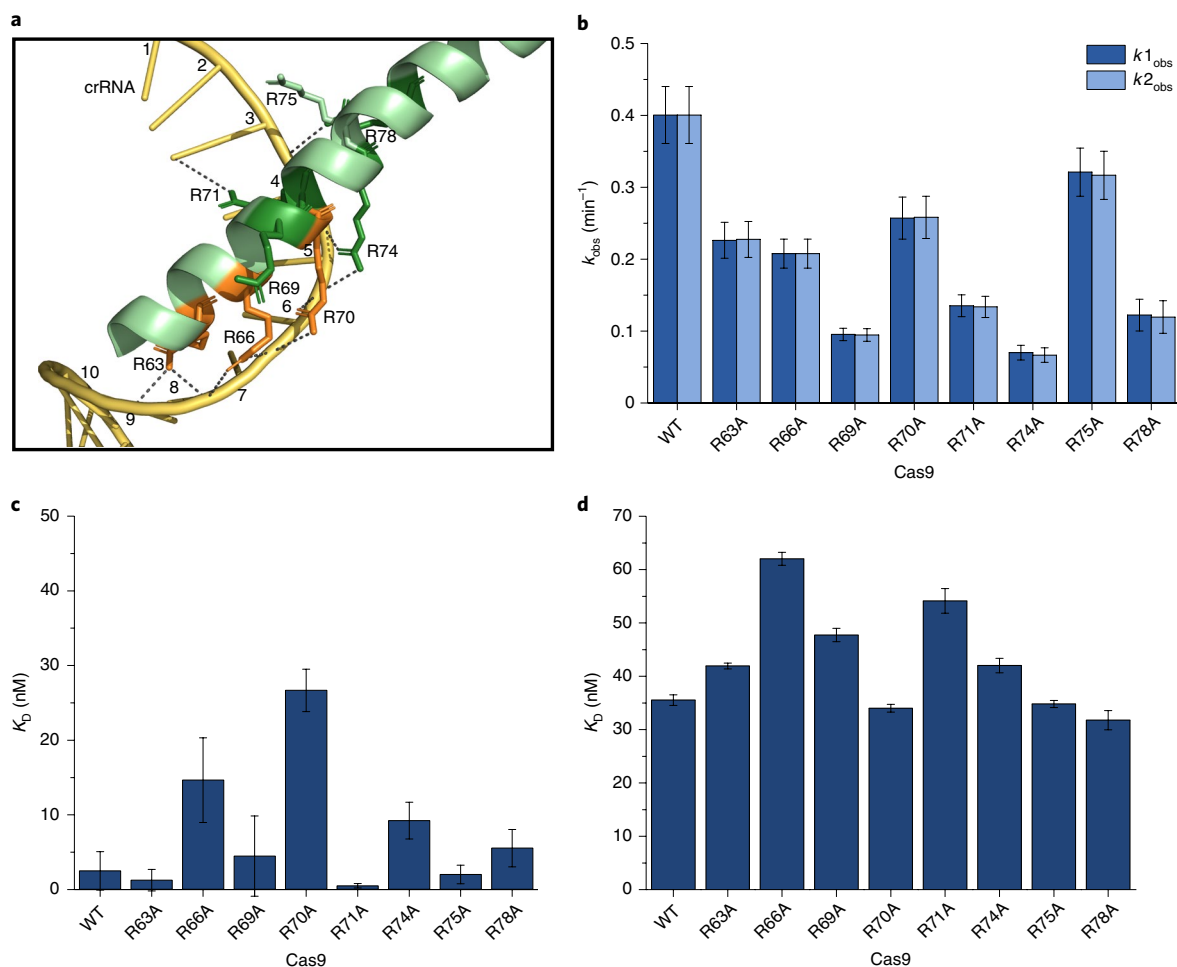
Fig. 3 | Arginine residues from the bridge helix affect DNA cleavage, RNA binding and target DNA binding by Cas9. a, Zoom in on the bridge helix (light green) and crRNA spacer (yellow) from Fig. 2a. The conserved arginines are shown in stick representation with dashed lines indicating hydrogen bonds or salt bridges to the crRNA (according to Nishimasu et al.[28]). R63, R66 and R70 are shown in orange; R69, R71, R74 and R78 in dark green and R75 in light green. **b**, In vitro cleavage rates of Cas9_wt and arginine mutants obtained by kinetic cleavage assays using 5 nM plasmid DNA (containing the WT target site) and 10 nM Cas9 in complex with 20 nM dual-RNA. $k1_{obs}$ is shown in dark blue and $k2_{obs}$ in light blue. **c**, Binding constants (K_D) of Cas9_wt and arginine mutants (Cas9_R63A, Cas9_R66A, Cas9_R69A, Cas9_R70A, Cas9_R71A, Cas9_R74A, Cas9_R75A and Cas9_R78A) derived from RNA-binding assays on dual-RNA, performed and calculated as described in Methods. Error bars represent the s.d. of at least three independent experiments. **d**, Binding constants resulting from EMSAs of Cas9_wt and arginine mutants on 90 bp 5′-³²P labeled PCR product of the WT target site. **b–d**, Cleavage rates and binding constants were calculated by concatenated fitting, represented are mean values. Error bars represent the s.d. ($n \geq 3$ independent experiments, Supplementary Tables 1 and 3).

Since Q768 affects Cas9 sensitivity to a mismatch on position 15, its replacement with alanine or glutamate might also influence the sensitivity to other PAM-distal mismatches. Hence, we tested the activity of these Cas9 variants in *E. coli* with mismatched sgRNAs (Fig. 2d,e). The respective bacterial strains showed increased survival in presence of sgRNAs with mismatches in positions 13 and 15–19 when compared to the WT sgRNA, and when compared to survival obtained with Cas9_wt. This indicates that these variants cleaved the target to a lesser extent than Cas9_wt in the presence of PAM-distal mismatches. We also noted that the sensitivity to PAM-adjacent mismatches was decreased when compared to Cas9_wt (Supplementary Fig. 2b).

Bridge helix arginines affect Cas9 binding and cleavage. To explore whether other amino acids and/or domains influence Cas9 sensitivity to mismatches in the PAM-adjacent part of the target, we investigated the conserved arginine-rich bridge helix (Supplementary Fig. 5). The bridge helix of SpCas9 comprises eight arginine residues (R63, R66, R69, R70, R71, R74, R75, R78), which

make contacts to the sgRNA phosphate backbone in the specificity-determining seed region[28–31] (Fig. 3a). We tested Cas9 variants of these arginine residues for binding to dual-RNA and target DNA, and cleavage of the FC target. All variants showed reduced in vitro cleavage rates when compared to Cas9_wt (Fig. 3b). Combination of these results with the data from binding assays (Fig. 3c,d) reveals three groups of mutants. First, Cas9_R66A, Cas9_R70A and Cas9_R74A have higher RNA binding constants than Cas9_wt (Fig. 3c), indicating that these residues are important for RNA binding, in agreement with the multiple contacts of these arginines to the sgRNA backbone[28–31]. Second, Cas9_R66A and Cas9_R71A are deficient in DNA binding (Fig. 3d), which in the case of Cas9_R66A might be due to its reduced RNA binding. For Cas9_R71A, however, the RNA binding affinity is not decreased indicating that R71 is directly involved in DNA binding. Possibly as a consequence of these binding deficiencies, the observed cleavage rates of Cas9_R66A, Cas9_R70A, Cas9_R71A and Cas9_R74A were slower when compared to Cas9_wt. Slower cleavage rates were also observed for Cas9_R63A, Cas9_R69A and Cas9_R78A. However, these variants
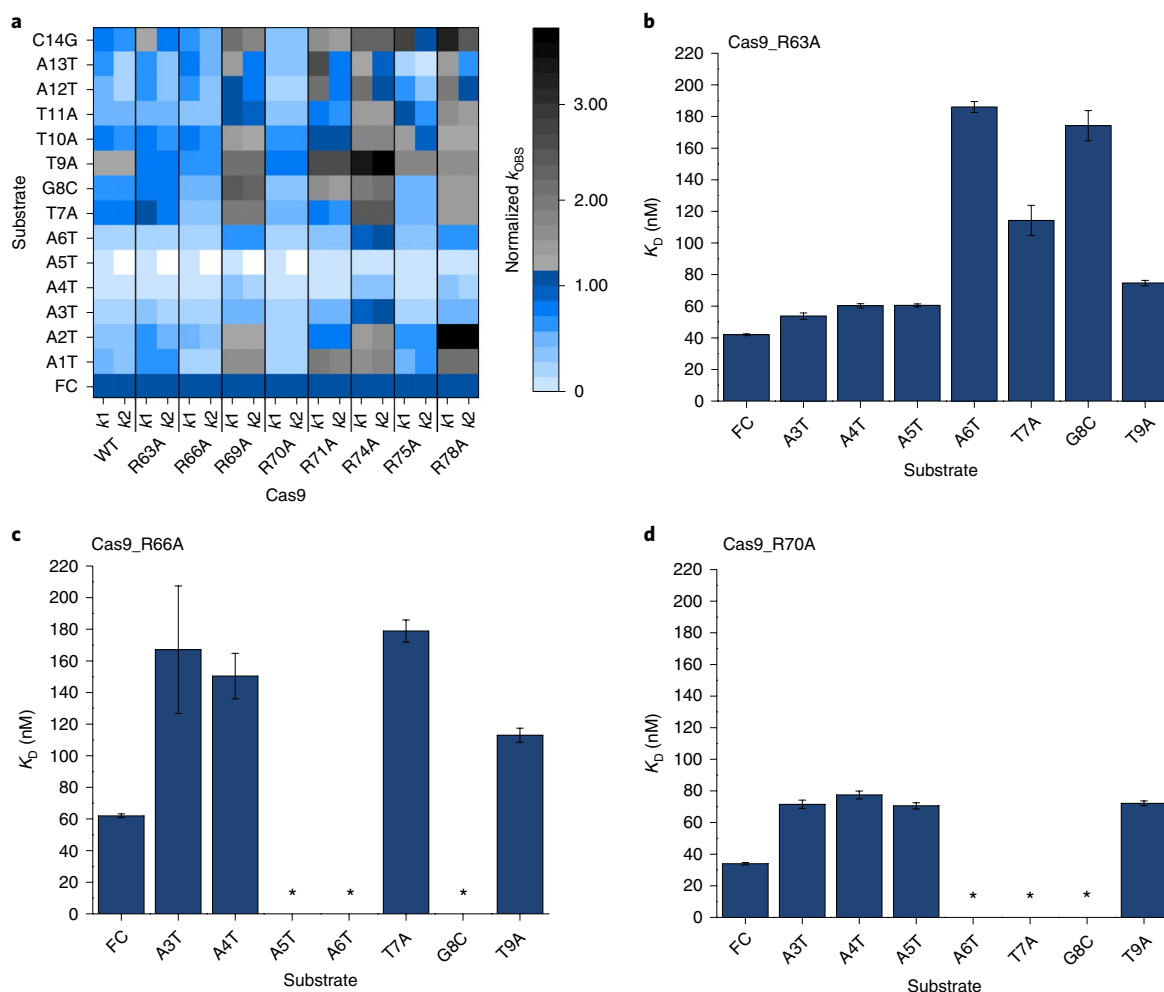
Fig. 4 | Two groups of arginine residues with opposite effects on Cas9 sensitivity to mismatches. a, Heat map of in vitro cleavage rates of Cas9_wt and arginine mutants normalized to each protein's rate on the FC substrate. Cleavage rates were determined as described in Methods from at least three independent experiments. Shades of blue (values less than 1) indicate that the cleavage rate is slower or similar to the cleavage rate on the WT substrate, whereas shades of gray (values more than 1) indicate that the cleavage rate is similar to or faster than the cleavage rate on the wt substrate. **b–d**, Binding constants of Cas9_R63A (**b**), Cas9_R66A (**c**) and Cas9_R70A (**d**) on the WT and mismatched substrates, determined by concatenated fitting, represented are mean values. Error bars represent the s.d. ($n \geq 3$ independent experiments, Supplementary Table 1). Asterisks indicate that due to poor binding of the protein to the substrate, binding constants could not be determined with confidence under the tested conditions.

displayed RNA and DNA binding constants comparable to Cas9_wt, implying that these residues are important for catalysis. In *E. coli*, increased survival was observed only in the presence of Cas_R66A, Cas9_R70A and Cas9_R74A (Supplementary Fig. 6). Last, exchange of R75 slightly reduced the in vitro cleavage rate, but had no effect on RNA and DNA binding or survival in vivo.

Two sets of arginines with opposite effects on specificity. To study whether the arginines have an effect on Cas9's sensitivity to mismatches, we exchanged them to alanine and tested these Cas9 variants for binding and cleavage of mismatched substrates in vitro, and in the bacterial survival assay with mismatched sgRNAs. Cleavage rates of Cas9_R75A on mismatched substrates (Fig. 4a and Supplementary Fig. 7g) and bacterial survival (Supplementary Fig. 8g) were like those of Cas9_wt. We identified two groups of mutants with opposite effects on Cas9 mismatch sensitivity (Fig. 4a and Supplementary Figs. 7 and 8). First, Cas9_R69A, Cas9_R71A, Cas9_R74A and Cas9_R78A cleaved most mismatched substrates with a similar cleavage rate or faster than the FC substrate (Fig. 4a, dark blue and shades of gray). This indicates that R69, R71, R74 and R78 confer specificity to Cas9 and changing them renders the protein less sensitive to

mismatches; that is, the specificity decreases. Cas9_R78A tolerates a mismatch in position 4 in vivo (Supplementary Fig. 8h), whereas Cas9_wt and the remaining arginine variants are all highly sensitive to this mismatch (Supplementary Fig. 8a–g). R78 could be acting as a mismatch-sensing residue on target position 4. This is consistent with the Cas9 structure where R78 contacts the sgRNA nucleotides that base pair with the target DNA positions 3–4 (refs. [28–31]).

The second group comprises Cas9_R63A, Cas9_R66A and Cas9_R70A. Cleavage rates of these variants on mismatched substrates were similar to or slower than on FC substrate (Fig. 4a, shades of light blue and Supplementary Fig. 7a,b,d). This is partially due to impaired binding of these mutants to several substrates containing a PAM-adjacent mismatch (Fig. 4b,c,d), indicating that replacing these residues renders Cas9 more sensitive to mismatches. Cas9_R70A was inactive in *E. coli* (Supplementary Fig. 6) consistent with a previous report in eukaryotic cells[28]. A possible explanation could be its reduced RNA binding (Fig. 3c). Hence, we focused on the remaining two variants, which displayed increased specificity.

Next, we wanted to determine whether changing both R63 and R66 would further increase the sensitivity to mismatches.
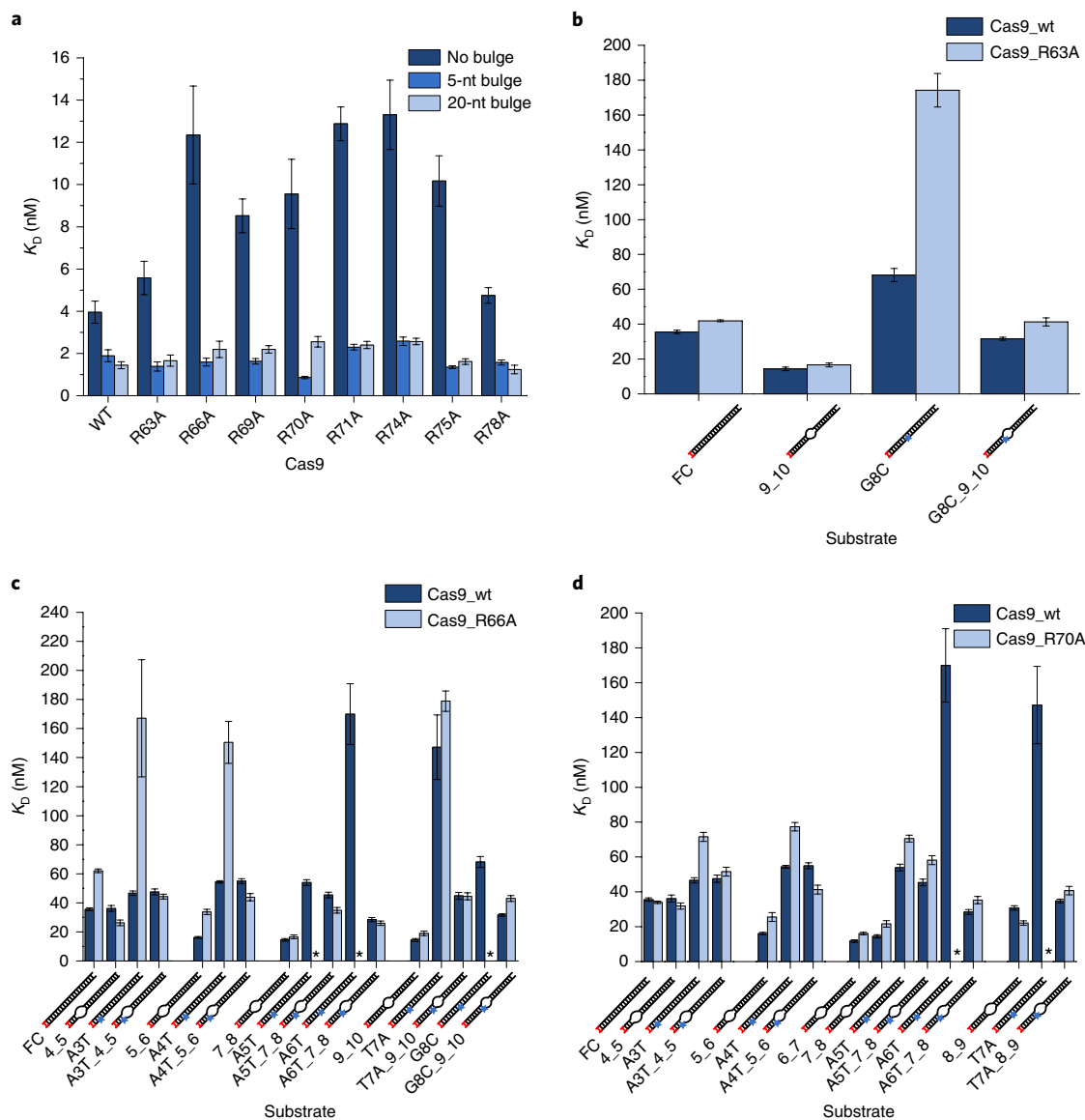
Fig. 5 | Arginines 63, 66 and 70 stabilize the R-loop in the presence of mismatches. a, Binding constants derived from EMSAs of Cas9_wt and arginine mutants on 140 bp 5′-³²P labeled oligonucleotides containing the NGG PAM and the target site without a bulge (target and nontarget strand fully base paired), a 5-nt bulge starting from the PAM-adjacent side of the target or with a complete 20-nt bulge (the scheme of the oligonucleotides can be found in Supplementary Fig. 10, whereas the full sequences of the oligonucleotides can be found in Supplementary Table 8). EMSAs were done as previously described, but in KG buffer instead of the Binding buffer. **b**–**d**, Binding constants of Cas9_wt (dark blue) and Cas9_R63A (light blue) (**b**), Cas9_wt (dark blue) and Cas9_R66A (light blue) (**c**), and Cas9_wt (dark blue) and Cas9_R70A (light blue) (**d**) on the different oligonucleotide substrates (with the target site indicated on the x axis). PAM is indicated in red, positions where the substrate is opened is shown as a bulge and a mismatch is marked by the blue asterisk. The black asterisks indicate that due to poor binding of the protein to the substrate, it was not possible to determine the K_D with confidence. **a**–**d**, Binding constants were determined by concatenated fitting, represented are mean values. Error bars represent the s.d. ($n \geq 3$ independent experiments, Supplementary Table 3).

Cas9_RR63,66AA was able to fully cleave the WT substrate in vitro, although slower than Cas9_wt (Supplementary Fig. 9a). Analysis of the cleaved supercoiled plasmid DNA (Supplementary Fig. 9b) revealed that Cas9_wt fully cleaved most of the substrate plasmids (indicated by 100% cleavage). In contrast, Cas9_RR63,66AA cleaved most of the plasmids with PAM-adjacent mismatches only up to 50%. This decrease in the amount of cleavage product is possibly due to poorer binding in the presence of a PAM-adjacent mismatch, since we could only measure the binding constant for the FC substrate (Supplementary Fig. 9c). Bacterial survival in presence of Cas9_RR63,66AA and WT sgRNA was comparable to survival in presence of dCas9 (Supplementary Fig. 9d), indicating that this variant is inactive in vivo.

R63 and R66 stabilize the R-loop in presence of mismatches. A previous study indicated that the bridge helix might play a role in the initiation of R-loop formation[32]. Therefore, we examined arginine variants for binding to DNA substrates, where five or twenty mismatches were introduced between the target and nontarget strand of the substrate resulting in unpaired DNA, mimicking a partially or fully formed R-loop (Supplementary Fig. 10). Binding of the arginine variants on opened substrates was improved compared to the WT substrate, indicating that these Cas9 variants are involved in R-loop formation and/or stability (Fig. 5a).

This led us to hypothesize that R63, R66 and R70 render Cas9 less specific by positively affecting the R-loop stability in the presence of a mismatch. Therefore, we tested binding of Cas9_R63A,

Cas9_R66A and Cas9_R70A on two sets of opened DNA substrates. The first set allowed full base pairing between the crRNA and DNA, but mismatches were introduced between the two DNA strands, resulting in a bulge. The second set contained a mismatch preventing base pairing of the crRNA and DNA at a specific position and were opened in the two subsequent positions to facilitate R-loop formation (Fig. 5b–d).

Cas9_R63A binds the FC substrate comparable to Cas9_wt (Fig. 3c), but has a binding defect on the substrate with mismatched position 8 (G8C) (Fig. 4b, note that the binding defect on A6T is also observed with Cas9_wt and therefore not related to the R63A exchange). This suggests that R63 stabilizes the R-loop in the presence of a mismatch on target position 8. On the other hand, if the substrate containing the G8C mismatch is opened at positions 9 and 10 (G8C 9_10), Cas9_R63A binds with the same affinity as to the FC substrate (Fig. 5b), indicating that the negative effect of replacing R63 is neutralized by opening the substrate and facilitating R-loop formation. This indicates that R63 stabilizes the R-loop in the presence of a mismatch on position 8 and thus could lower the sensitivity of Cas9 to this mismatch.

Cas9_R66A displayed binding defects on substrates containing mismatches on positions 3–9 (Fig. 4c). We tested this Cas9 variant for binding of substrates containing single mismatches on these positions that are opened on the two consecutive positions. The binding defects were reversed and the binding constant of Cas9_R66A became similar to or even smaller than for Cas9_wt on the same mismatched substrate without the bulge (Fig. 5c). The gain in binding of Cas9_R66A on opened substrates versus FC substrate is less when compared to the gain in binding between substrates having a mismatch and opened substrates with the same mismatch. Therefore, the decreased binding of Cas9_R66A on mismatched substrates containing the bulge is not only due to the opening of the DNA. In addition to being important for RNA binding, reflected in its interaction with the sgRNA backbone in positions that base-pair to target DNA positions 7 and 8 (refs. [28–31]), these results indicate that R66 has a stabilizing effect on R-loop formation in the presence of a mismatch over almost the entire seed sequence.

Cas9_R70A binds the target DNA comparably to Cas9_wt, but displays a binding defect on substrates with mismatches from positions 3–9 (Fig. 4d). However, if the mismatched substrate is followed by a bulge, Cas9_R70A can bind these substrates equally or better than Cas9_wt (Fig. 5d). This indicates that R70, which interacts with the sgRNA backbone in positions that base pair to target DNA positions 6 and 7 (refs. [28–31]), has a stabilizing effect on the R-loop in the presence of mismatches from positions 3–7, similar to Cas9_R66A (Fig. 5c).

Cas9_R63A/Q768A is an enhanced specificity Cas9 variant. Replacing R63 or R66 with alanine increases Cas9 sensitivity to mismatches in the PAM-adjacent part of the target, whereas mutation Q768A increases sensitivity to PAM-distal mismatches. To assess whether Cas9 variants with a combination of these changed residues have an additive effect on specificity, we determined the in vitro cleavage rates of Cas9_R63A/Q768A and Cas9_R66A/Q768A (Supplementary Fig. 11a,b), and tested them also in the bacterial survival assays (Supplementary Fig. 11c,d). Both Cas9 variants were sensitive to mismatches in most positions of the target in vitro and in bacterial cells, as both showed reduced cleavage rates and increased survival compared to the WT targets (Supplementary Fig. 11).

Cas9_R63A/Q768A shows enhanced gene editing specificity. In addition, we tested Cas9_R63A/Q768A and Cas9_R66A/Q768A for gene editing activity in the human cell line MCF-7 (Supplementary Table 5) with four sgRNAs targeting *EpCAM*. We used flow cytometry

to determine the fraction of EpCAM[positive] versus EpCAM[negative] cells (Supplementary Fig. 12) as read-out for editing efficiency. Cas9_R63A/Q768A showed activity, although lower than Cas9_wt (Fig. 6a), whereas Cas9_R66A/Q768A did not exhibit any editing activity (Supplementary Fig. 13). To examine whether Cas9_R63A/Q768A is more sensitive to mismatches than Cas9_wt, we repeated the experiment with sgRNAs EpCAM-1 and EpCAM-4, containing a mismatch in different PAM-distal positions (Supplementary Fig. 14). Cas9_R63A/Q768A showed increased specificity in a sgRNA dependent manner. With sgRNA EpCAM-4, Cas9_R63A/Q768A was more sensitive to most PAM-distal mismatches when compared to Cas9_wt (Fig. 6b). Cas9_R63A/Q768A shows no editing in the presence of a mismatch in position 15, in agreement with the results obtained in vitro and in bacteria (Supplementary Fig. 11). With sgRNA EpCAM-1 we observed minimal to no editing for positions 14 and 16–19, probably due to a difference in the on-target activity between Cas9_wt and the mutant. Considering this, Cas9_R63A/Q768A was significantly more specific than Cas9_wt only on position 13 (Fig. 6c).

Finally, we performed gene editing experiments in human embryonic kidney 293 (HEK293) cells with two sgRNAs targeting *VEGFA* and one sgRNA targeting *EMX1* with previously characterized off-target sites[33]. On- and off-target sites (Supplementary Table 6) from Cas9-treated cells were analyzed by amplicon sequencing (Fig. 6d–f). Cas9_R63A/Q768A cleaved the on-target sites with all three guides with similar (VEGFA3 and EMX1.4) or lower (VEGFA1) efficiency than Cas9_wt. Cas9_R63A/Q768A showed significantly higher specificity at off-target sites compared to Cas9_wt, except with sgRNA targeting *EMX1* (Fig. 6d–f).

Discussion

Designing highly specific Cas9 variants that do not cleave off-targets is crucial for genome editing. Considering their natural function as defense systems against invading nucleic acids, Cas9 enzymes evolved to tolerate certain mismatches to be able to cleave viral escape mutants[34,35]. Cas9 is sensitive to mismatches in the PAM-adjacent and PAM-distal part of the target, but shows flexibility toward mismatches in the middle of the target[2]. Here, we show that a mismatch at position 15 causes hyperactivity of Cas9 in vitro and in bacteria, but not in human cells. This hyperactivation is dependent on Q768 given that it is no longer observed when Q768 was exchanged. How Q768 is involved in the hyperactivation of Cas9 is not yet understood. Given its distance from the RNA–DNA hybrid, the effect is most probably indirect (Fig. 2a). Therefore, other residues in closer proximity to target DNA position 15 might be of interest for future studies. Enhanced cleavage in presence of this mismatch was also detected with Cas9 of *S. mutans* (Supplementary Fig. 4), indicating that it may be a common feature of some type II-A Cas9 proteins.

Structural and biochemical characterization of Cas9 led to the identification of several domains responsible for sgRNA and DNA binding (REC lobe), PAM recognition (PI-domain) or DNA cleavage (HNH and RuvC domains)[2,28–31,36]. Here, we analyzed biochemical properties of Cas9_wt and specific variants in vitro. It has to be noted that due to differences in the experimental setup, kinetic parameters for Cas9_wt differ from previously published ones for dCas9 (ref. [4]), namely 0.49 nM versus 35.5 nM. These differences include the nature of the protein (Cas9_wt versus dCas9), the substrate length for DNA binding assays (55 bp versus 90 bp), preparation of the substrates (annealed oligos versus gel purified PCR products), concentration of dual-RNA used (either 1 or 10 M excess of dual-RNA over Cas9 versus 2 M excess), presence and nature of a competitor, differences in Cas9 purification, reaction time and so on.

We determined all values using the same setup for all Cas9 variants, which only serve as a measure of comparison to analyze the
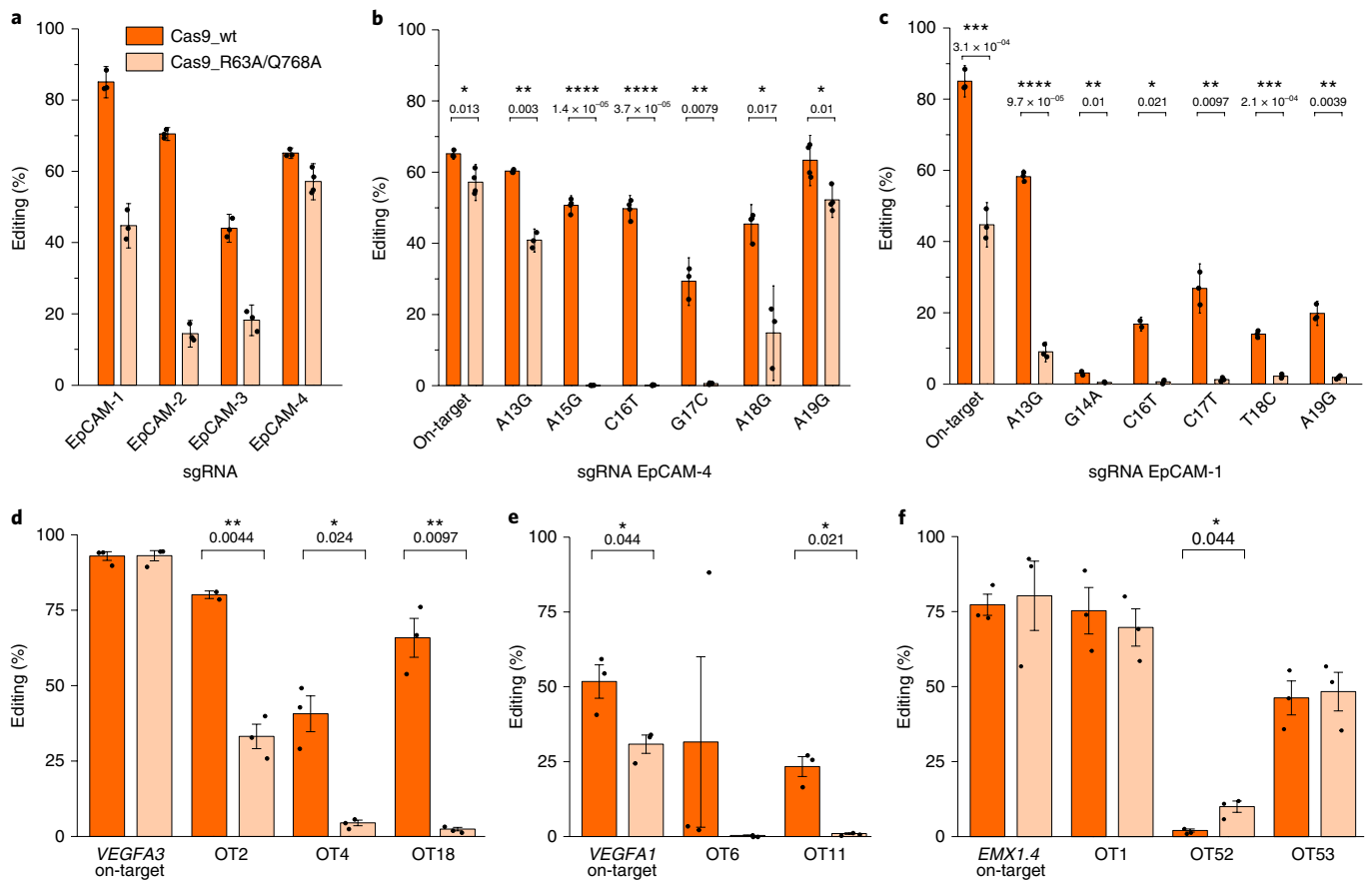
Fig. 6 | Cas9_R63A/Q768A displays enhanced editing specificity in human cells. a, Percentage of *EpCAM* on-target editing in MCF-7 cells by Cas9_wt (dark orange) and Cas9_R63A/Q768A (light orange) in the presence of four different sgRNAs, determined by flow cytometry as described in Methods. **b,c**, Percentage of *EpCAM* editing by Cas9_wt (dark orange) and Cas9_R63A/Q768A (light orange) in the presence of EpCAM-4 (**b**) and EpCAM-1 (**c**) sgRNAs, that were either fully complementary to the target site (on-target), or contained the indicated mismatches to the PAM-distal part of the target site. Editing percentage was determined by flow cytometry as described in Methods. **d–f**, Percentage of on- and off-target editing in HEK293 cells by Cas9_wt (dark orange) and Cas9_R63A/Q768A (light orange) with sgRNAs targeting *VEGFA3* (**d**), *VEGFA1* (**e**) and *EMX1.4* (**f**) sites, determined by amplicon sequencing with a mean coverage of 64,000 paired-end reads per library (also see Methods). **a–f**, OT stands for off-target. Error bars represent s.d. ($n \geq 3$ independent experiments, Supplementary Table 4). Values from independent replicates are shown as black dots and the center value represents the mean. Statistical significance was determined by a two-sided standard *t*-test (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$, ****$P \leq 0.0001$, exact values are shown below the asterisks).

influence of investigated Cas9 residues on RNA or DNA binding and cleavage. We were also able to confirm the in vitro results in bacteria and human cells. The in vitro analysis showed that the bridge helix arginines are involved in RNA and target DNA binding or catalysis and most importantly in mismatch sensing and R-loop stabilization. These arginines can be separated into two groups. First, R69, R71, R74 and R78 were found to confer sensitivity toward mismatches. Changing these residues rendered Cas9 more unspecific. Second, R63, R66 and R70 confer tolerance to Cas9 in the presence of a mismatch. Replacement of these residues rendered Cas9 more specific. These contrasting observations indicate that these Cas9 residues are involved in the balance between specificity and efficiency in the natural CRISPR system.

Additionally, R63, R66 and R70 reduce mismatch sensitivity by stabilizing the R-loop. Mutation of any one of these residues resulted in increased mismatch sensitivity at position 8 for Cas9_R63A or along the seed sequence for Cas9_R66A and Cas9_R70A. A bulge in the substrate around these positions neutralized this sensitivity and led to binding constants comparable to Cas9_wt on the same substrate.

According to the kinetic model for the specificity of RNA-guided nucleases proposed by Bisaria et al.[37], it was suggested that

by increasing the dissociation constant and/or decreasing the cleavage rate the rate-limiting step of the reaction can be shifted from binding to cleavage. This would increase the time Cas9 has to equilibrate its binding and could result in preference of on- over off-target cleavage[37]. In accordance with this model, a recent study[38] showed that R-loop formation is the rate-limiting step, which reduces the DNA dissociation rate of Cas9. Replacing R63 and R66 increased the dissociation constant and/or decreased the cleavage rate of Cas9 by possibly destabilizing the R-loop. Therefore, this model for specificity enhancement[37] could be a possible explanation for the observed increased mismatch sensitivity of our mutants.

The conformational transition of the HNH domain from inactive to the catalytically active state in the other enhanced specificity Cas9 variants such as eSpCas9 (ref. [17]) (mutations in HNH and PAM-interacting domains), SpCas9-HF1 (ref. [15]) (mutations in HNH and REC3 domains) and HypaCas9 (ref. [10]) (mutations in REC3 domain) was blocked in the presence of PAM-distal mismatches. This mechanism is mediated by the REC3 domain[10]. Considering that R63, R66 and Q768 are located in the bridge helix and RuvC domain, respectively, a mechanism similar to the one mediated by the REC3 domain is unlikely, but cannot be excluded without further investigation.

The trade-off between on-target activity and improved specificity might sometimes be disadvantageous as demonstrated by the double mutants Cas9_R66A/Q768A and Cas9_RR63,66AA. The first is more specific in vitro and bacteria, but was not active in human cells. The second showed an increased specificity in vitro, but is inactive in vivo.

Combination of Q768A, which increases Cas9 specificity at the PAM-distal part, with a replacement of R63 led to an overall improved specificity of Cas9 in vitro and in bacteria, and also displayed increased specificity in human cells. This demonstrates that a combination of variants altering the kinetic properties and variants derived from structure-guided analysis is a viable approach to optimize Cas9 specificity and provides a basis for further development of highly specific Cas9 enzymes.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-020-0490-4.

References

1. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci. USA* **109**, E2579–E2586 (2012).
2. Jinek, M. et al. A programmable dual-RNA–Guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
3. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602 (2011).
4. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62 (2014).
5. Szczelkun, M. D. et al. Direct observation of R-loop formation by single RNA-guided Cas9 and cascade effector complexes. *Proc. Natl Acad. Sci. USA* **111**, 9798–9803 (2014).
6. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).
7. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
8. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
9. Wu, W. Y., Lebbink, J. H. G., Kanaar, R., Geijsen, N. & van der Oost, J. Genome editing by natural and engineered CRISPR-associated nucleases. *Nat. Chem. Biol.* **14**, 642–651 (2018).
10. Chen, J. S. et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407 (2017).
11. Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci. Adv.* **3**, eaao0027 (2017).
12. Tycko, J., Myer, VicE. & Hsu, PatrickD. Methods for optimizing CRISPR-Cas9 genome editing specificity. *Mol. Cell* **63**, 355–370 (2016).
13. Casini, A. et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotech.* **36**, 265 (2018).
14. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57 (2018).
15. Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490 (2016).
16. Kulcsár, P. I. et al. Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. *Genome Biol.* **18**, 190 (2017).
17. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
18. Lee, J. K. et al. Directed evolution of CRISPR-Cas9 to increase its specificity. *Nat. Commun.* **9**, 3048 (2018).
19. Vakulskas, C. A. et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* **24**, 1216–1224 (2018).
20. Singh, D. et al. Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.* **25**, 347–354 (2018).
21. Babu, K. et al. Bridge Helix of Cas9 modulates target DNA cleavage and mismatch tolerance. *Biochemistry* **58**, 1905–1917 (2019).
22. Mekler, V., Minakhin, L. & Severinov, K. Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proc. Natl Acad. Sci. USA* **114**, 5443–5448 (2017).
23. Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* **7**, 12778 (2016).
24. Kunne, T. et al. Role of nucleotide identity in effective CRISPR target escape mutations. *Nucleic Acids Res.* **46**, 10395–10404 (2018).
25. Xue, C. et al. CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* **43**, 10831–10847 (2015).
26. Cencic, R. et al. Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One* **9**, e109213 (2014).
27. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110 (2015).
28. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
29. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569 (2014).
30. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
31. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
32. Zeng, Y. et al. The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic Acids Res.* **46**, 350–361 (2018).
33. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotech.* **32**, 279 (2014).
34. Datsenko, K. A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
35. Fineran, P. C. et al. Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl Acad. Sci. USA* **111**, E1629–E1638 (2014).
36. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
37. Bisaria, N., Jarmoskaite, I. & Herschlag, D. Lessons from enzyme kinetics reveal specificity principles for RNA-guided nucleases in RNA interference and CRISPR-based genome editing. *Cell Syst.* **4**, 21–29 (2017).
38. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* **22**, 359–371 (2018).

Methods

DNA handling. Plasmid DNA preparation (QIAprep Spin MiniPrep Kit, Qiagen), PCR (Phusion High-Fidelity DNA Polymerase, Thermo Scientific; Taq DNA Polymerase, Fermentas), DNA digestion with restriction enzymes (Thermo Scientific or New England Biolabs), DNA ligation (T4 DNA ligase, Fermentas), purification of PCR products (QIAquick PCR Purification Kit, Qiagen), gel extraction (QIAquick Gel Extraction Kit) and gel electrophoresis were done according to the manufacturer's instructions or using standard protocols[39]. Site-directed mutagenesis was done following the PCR-based protocol[40]. All constructed plasmids used in this study were validated by sequencing (LGC Genomics and SeqLab) and are listed in Supplementary Table 7. All oligonucleotides (Sigma-Aldrich and Biomers) used in this study are listed in Supplementary Table 8.

Construction of plasmids for kinetic cleavage assays. Oligonucleotides containing the NGG PAM and wt or mutant *speM* target site (OLEC2286-OLEC2341) were annealed to generate the double-stranded insert. This was cloned using T4 DNA ligase (ThermoFisher) into the beta-galactosidase-encoding gene of pEC286, which was digested with Eco91I. The ligation products were transformed in *E. coli* and positive clones were selected by blue-white screening[41,42].

Construction of plasmids for the bacterial survival assay. The *rfp* gene was PCR-amplified from pEC1217 using primers containing XbaI and EcoRI restriction sites, and cloned into the corresponding sites of the plasmid pStaby1.2 (Delphi Genetics). Plasmids encoding nuclease-inactive Cas9 (pEC1215) or the *rfp* targeting sgRNA (pEC1218) were obtained from a previously published study[43]. The gene encoding Cas9_wt was PCR-amplified from pEC626 (ref.[44]) using primers containing SpeI and EagI restriction sites, and cloned into the corresponding sites of pEC1215. All Cas9 mutants were generated by site-directed mutagenesis of the Cas9_wt-encoding plasmid. To generate plasmids encoding sgRNA with single mismatches in each position of the spacer sequence, site-directed mutagenesis was performed using the appropriate mutation-introducing forward primer and OLEC6032 as the reverse primer. Resulting plasmids (pEC1537-1556) express sgRNA under control of the J23119 promoter[45]. To reduce the amount of sgRNA expressed, the promoter was mutagenized with OLEC3908 and OLEC3909 to a weaker version called J23100 (http://parts.igem.org).

Construction of plasmids for in vivo gene editing in eukaryotic cells. Oligonucleotides containing sgRNA spacers (OLEC10121-10132, OLEC10341-10479) targeting the *EpCAM* gene (fully complementary or with single point mutations that caused a mismatch to the target DNA sequence) were phosphorylated with T4 polynucleotide kinase (Fermentas) and annealed to generate the double-stranded inserts. To obtain the variants Cas9_R63A/Q768A and Cas9_R66A/Q768A, site-directed mutagenesis was performed on the plasmid pCROPseq_Cas9_wt, which is based on CROPseq-Guide-Puro (Addgene no. 86708)[46] with an added human codon optimized SpCas9 containing a C-terminal NLS tag. Cas9-encoding plasmids were digested with Esp3I (Thermo Scientific) and dephosphorylated with alkaline phosphatase (Thermo Scientific).

For cloning of sgRNAs containing spacers targeting the *VEGFA* sites 1 and 3, and the *EMX1* target site 4, Cas9-encoding plasmids were digested with BpiI FD (ThermoFisher) and purified (GeneJET Gel Extraction Kit, ThermoFisher) following the manufacturer's instructions. Oligonucleotides containing sgRNAs (CR3373-3378) were mixed and annealed. The inserts were cloned into the digested vectors to generate full sgRNAs expressed under the control of the U6 promoter.

Preparation of ultra-competent *E. coli*. Ultra-competent *E. coli* cells for bacterial survival assays were prepared according to refs.[47,48]. Bacteria were grown in LB medium at 37 °C with 180 r.p.m. and shaking for 6–8 h. This preculture was used to inoculate SOB medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄), in a ratio of 1:200, and grown at 18 °C with 180 r.p.m. and shaking overnight, until an optical density (OD$_{600nm}$) of 0.55 was reached. The cells were chilled on ice for 10 min and centrifuged at 5,915*g* for 10 min at 4 °C. The supernatant was removed, cells were resuspended in ice-cold transformation buffer (10 mM HEPES pH 6.7, 15 mM CaCl₂, 250 mM KCl, 55 mM MnCl₂) and incubated for 10 min on ice. The procedure was repeated and 7.5% of DMSO was added to the resuspended cells. The cells were incubated for 10 min on ice and the aliquots were frozen in liquid nitrogen.

For transformation, cells mixed with 50 ng plasmid DNA were incubated for 30 min on ice, subjected to heat-shock at 42 °C for 45 s, after which they were shortly put on ice. SOC medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose) was added and cells were recovered for 1 h at 37 °C shaking. After recovery, cells were shortly centrifuged for 1 min at 6,000 r.p.m. and plated on LB agar plates with 50 μg ml⁻¹ streptomycin, 100 μg ml⁻¹ carbenicillin, 30 μg ml⁻¹ chloramphenicol and 1% glucose.

RNA in vitro transcription. All RNAs used in the in vitro assays were in vitro transcribed with the AmpliScribe-T7 Flash Transcription kit (Epicentre) according to manufacturer's instructions. The templates for the reaction were either synthetic

oligonucleotides or generated by PCR (listed in the Supplementary Table 8). In vitro transcribed RNAs were purified as described previously[44]. Equimolar amounts of tracrRNA and crRNA were annealed in 5× RNA annealing buffer (1 M NaCl, 100 mM HEPES, pH 7.5) at 95 °C for 5 min, and slowly cooled to room temperature.

Cas9 protein purification. *E. coli* NiCo21 (DE3) competent cells (New England Biolabs) (Supplementary Table 9) were transformed with overexpression plasmids encoding WT or mutant Cas9. Protein overexpression and purification by Ni-NTA affinity chromatography was performed as previously described[44]. Protein containing fractions were further purified over chitin beads (New England Biolabs), as in ref.[44], and dialyzed against dialysis buffer (20 mM HEPES pH 7.5, 150 mM KCl, 50% glycerol). Protein concentration was determined using the Bradford assay and nucleic acid contamination was assessed by measuring the A$_{260}$/A$_{280}$ ratio.

Generation of substrates for electrophoretic mobility shift assays (EMSAs). Substrates used for EMSAs were generated by PCR or from synthetic oligonucleotides (Sigma-Aldrich). In the first case, the target sequence (wt or mutated protospacer 2, *speM*) was amplified from plasmids (pEC576-pEC599) using primers OLEC4816 and OLEC4817. Products were precipitated with sodium acetate and ethanol, and purified over a 1.5% agarose gel in 1× TBE buffer (0.1 M Tris, 0.1 M boric acid, 0.002 M EDTA). The corresponding bands were excised from the gel and the DNA was isolated using the QIAquick Gel Extraction Kit (Qiagen).

In the second case, HPLC-purified synthetic oligonucleotides containing the target sequence were used. To generate a double-stranded EMSA substrate, oligonucleotides containing the target and nontarget DNA strand were annealed in 5× RNA annealing buffer at 95 °C for 5 min and left at room temperature for slow cooling. The substrates were purified over 6% polyacrylamide gel in 1× TBE buffer and the corresponding bands were excised. Gel pieces containing the samples were incubated overnight at 4 °C in 1× TE buffer (1 M Tris-HCl pH 8, 0.5 M EDTA) and afterward DNA was precipitated with sodium acetate and ethanol. Reactions were centrifuged for 30 min at 15,000*g* and DNA pellets were resuspended in MQ water. The concentration of the DNA was determined by measuring the absorbance at 260 nm with NanoDrop (Thermo Fisher Scientific).

EMSAs. Substrates for EMSAs were radiolabeled with [ɣ-³²P]-ATP (Hartmann Analytics) using T4 polynucleotide kinase (Fermentas) and purified on Illustra Microspin G-25 columns (GE Healthcare). Binding reactions with increasing concentrations of wt or mutant Cas9 and a 2× molar excess of dual-RNA over protein were preincubated in Binding buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM CaCl₂·2H₂O, 5% glycerol, 1 mM DTT) or KG buffer (100 mM potassium glutamate, 25 mM Tris-acetate pH 7.5, 10 mM Mg-acetate, 0.5 mM 2-mercaptoethanol, 10 mg ml⁻¹ bovine serum albumin)[49], for 15 min at 37 °C, before the addition of 1 nM labeled substrates. After addition of DNA, binding reactions took place at 37 °C for 1 h. Protein–DNA complexes were separated by 5% native polyacrylamide gel electrophoresis in 0.5× TBE buffer with 5 mM CaCl₂·2H₂O. The gels were exposed on autoradiography film overnight and visualized by phosphorimaging. Results of at least three independent experiments were quantified with Gel Analyzer (http://www.gelanalyzer.com/index.html) and the binding constants were determined by fitting to the Hill's equation $y_b = V_{max} \times x_C^n / (k^n + x_C^n)$ (y_b, % of bound DNA; x_C, enzyme concentration (nM); V_{max}, maximally bound DNA; k, K_D; and n, Hill coefficient) using Origin Software (OriginLab).

RNA-binding assays. Purified WT or variant Cas9 proteins were labeled with the NT-647-His labeling fluorescent dye (NanoTemper Technologies) in 1× PBST buffer (NanoTemper Technologies) according to the manufacturer's instructions. Labeled Cas9 was incubated with increasing concentrations of dual-RNA in RNA Binding buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM CaCl₂·2H₂O, 5% glycerol, 1 mM DTT, 0.05% Tween-20). To remove potential aggregates, reactions were centrifuged before loading into the Monolith NT.115 MST Premium Coated Capillaries (NanoTemper Technologies). Binding of Cas9 to dual-RNA was measured by microscale thermophoresis using the NanoTemper Monolith NT.115. Binding constants were determined by fitting data from at least three independent experiments to the exponential binding equation using the MO Affinity Analysis Software (NanoTemper Technologies).

Kinetic cleavage assays. Dual-RNA (20 nM) and Wt or mutant Cas9 (10 nM) were preincubated for 15 min at 37 °C in KG buffer. For kinetic cleavage assays with *S. mutans* Cas9, 50 nM dual-RNA and 25 nM Cas9 were used. Plasmids containing WT or mutated target sequence (pEC576-599) were added to a final concentration of 5 nM and incubated for 90 min at 37 °C. At defined time points (1, 2, 5, 10, 20, 40, 60 and 90 min) samples were withdrawn and the reaction was stopped by addition of 5× loading buffer (250 mM EDTA, 30% glycerol, 1.2% SDS, 0.1% bromophenol blue). Cleavage products were resolved by 1% agarose gel electrophoresis in 1× TAE buffer (0.04 M Tris, 0.02 M acetic acid, 1 mM EDTA). DNA was visualized by ethidium bromide staining. Band intensity of open circular, linear and supercoiled DNA was analyzed by densitometry using

Image Lab software (Bio-Rad Laboratories, Inc.). Data obtained from at least three independent experiments were fitted to the equations stated below in Origin Software (OriginLab). The rate of disappearance of the supercoiled plasmid DNA ($k1_{obs}$) was calculated using $y_{sc} = y_{sc}0 + A_{sc} \times \exp(-k1 \times x_b)$ (y_{sc}, amount of supercoiled DNA; $y_{sc}0$, amount of supercoiled DNA at infinite time 0; A_{sc}, maximal amount of supercoiled DNA that was cleaved; $k1$, $k1_{obs}$; x_t, time in min). The rate of appearance of the linear plasmid DNA ($k2_{obs}$) was calculated using: $y_{li} = y_{li}0 + A_{li} \times (1 - \exp(-k2 \times x_t))$ (y_{li}, amount of linear DNA; $y_{li}0$, amount of linear DNA at infinite time 0; A_{li}, maximal amount of linear DNA after cleavage; $k2$, $k2_{obs}$; and x_t, time in min).

Kinetic cleavage assays on linear substrates were performed on oligonucleotides containing the NGG PAM and *speM* target site (WT or with mismatches in positions 10–14), with either the target or nontarget strand radioactively labeled as described above. Equimolar amounts of labeled and nonlabeled substrate were annealed at 95 °C and slowly cooled to room temperature. Double-stranded substrates were purified with DNA Clean & Concentrator −5 (Zymo Research). Kinetic cleavage assays were performed as described using 2 nM radioactively labeled DNA substrates, 7.5 nM Cas9 and 15 nM dual-RNA. Reactions were stopped by the addition of 2× loading buffer (95% formamide, 0.025% SDS, 25 mM EDTA, 0.025% bromophenol blue) and incubated for 5 min at 95 °C, before loading on a 12% PAA-urea gel. The gels were exposed on autoradiography film overnight and visualized by phosphorimaging.

Bacterial survival assay. The bacterial survival assay is based on a three-plasmid system (Supplementary Fig. 1). The three plasmids encoding red fluorescent protein (RFP), Cas9 and sgRNA, respectively, are based on a previously published system[43]. Cas9 is expressed under control of the arabinose promoter, the sgRNA targeting the 5′ region of *rfp* is constitutively expressed by the J23100 promoter and RFP expression is controlled by T7 promoter and *lacO* operator. The bacterial cells used in the assay are *E. coli* SE4 (Delphi genetics), an engineered derivative of BL21DE3, additionally encoding the toxin CcdB (Supplementary Table 9). The corresponding antitoxin CcdA is encoded on the RFP expressing plasmid. *E. coli* SE4 was transformed with these three plasmids in a consecutive manner. In short, plasmid encoding RFP (pEC2132) was transformed using a standard heat-shock transformation protocol[39]. The resulting clones were made ultra-competent as described above. pEC2132-containing ultra-competent cells were transformed with plasmids encoding WT sgRNA or sgRNA with single mismatches in the spacer (pEC2013 and pEC2052-2071), and the resulting cells were made ultra-competent again. Finally, SE4 cells that contained the RFP-encoding plasmid and the sgRNA-encoding plasmid were transformed with 50 ng plasmid encoding Cas9_wt or mutant Cas9 proteins. The complete transformation mix was plated on LB agar plates containing 50 µg ml⁻¹ streptomycin, 100 µg ml⁻¹ carbenicillin, 30 µg ml⁻¹ chloramphenicol and 1% glucose. The concentration of the antibiotics was kept constant throughout the experiment. The next day, individual colonies (2–3) were used to inoculate 2 ml of LB medium with the addition of streptomycin, carbenicillin, chloramphenicol and 1% glucose, and grown at 37 °C for 6.5 h with shaking at 180 r.p.m. The $OD_{600\,nm}$ was measured using a plate reader (BioTek) and the appropriate volume of each culture was taken to normalize the OD_{600nm} to 0.03. Samples were centrifuged for 1 min at 6,000 r.p.m., the supernatant was removed and the pellet was resuspended in 100 µl LB medium with streptomycin, carbenicillin and chloramphenicol. Then, 2 µl of each sample was added to the 96-well plate containing 198 µl of either LB medium with streptomycin, carbenicillin, chloramphenicol and 1% glucose (suppressing conditions), or LB medium with streptomycin, carbenicillin, chloramphenicol, 33 mM arabinose and 0.1 mM IPTG (inducing conditions).

Under suppressing conditions, neither RFP nor Cas9 is expressed and CcdB is neutralized by the presence of CcdA. Under inducing conditions three scenarios are possible. (1) Cas9 is cleavage and binding deficient. RFP and CcdA are expressed, the cells grow and fluorescence can be detected. (2) Cas9 is cleavage deficient, but able to bind to its target site. The cells grow since CcdA is still expressed, but *rfp* expression is repressed by Cas9 binding to the 5′ region of the gene. (3) Cas9 is able to bind and cleave the RFP and CcdA encoding plasmid. This leads to death of the cells, since the CcdA antitoxin is no longer expressed and cannot neutralize the toxin CcdB. To distinguish between these scenarios, the OD_{600nm} and red fluorescence units (excitation wavelength 555 nm, emission wavelength 588 nm) were measured with a fluorescence plate reader (Biotek) in 5-min intervals during a 10 h kinetic experiment, at 37 °C with shaking. After subtraction of the blank samples, survival was calculated by dividing the OD_{600nm} at inducing conditions by the OD_{600nm} at suppressing conditions. Statistical analysis of at least five replicates was performed using Origin Software (OriginLab).

Human cell culture and transfections. MCF-7 cells (Supplementary Table 5) were cultured at 37 °C with 5% CO_2 in advanced DMEM (Thermo Scientific) supplemented with 10% heat-inactivated fetal bovine serum (Thermo Scientific), 2 mM GlutaMax (Thermo Scientific) and penicillin-streptomycin (Sigma-Aldrich). HEK293 cells (Supplementary Table 5) were cultured at 37 °C with 5% CO_2 in DMEM (Sigma-Aldrich) supplemented with 10% fetal bovine serum (Gibco), 2 mM L-glutamine (Sigma-Aldrich) and Normocin (Invivogen).

Cells were seeded in six-well or 24-well plates 24 h before transfection at a density of 100,000 cells per ml. Transfections of plasmids with sgRNAs targeting *EpCAM* were performed with the jetPRIME transfection reagent (Polyplus) according to the manufacturer's instructions. MCF-7 cells were transfected with 500 ng of plasmid DNA. Transfected cells were selected by adding 2 µg ml⁻¹ of puromycin (Sigma-Aldrich) 1 d after transfection. Growth medium with puromycin was replaced by standard growth medium after 2 d. Cells were analyzed by flow cytometry 8–10 d post transfection.

HEK293 cells were transfected with 1 µg plasmids p1490-1492, p1498-1500 in 24-well plates using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol. Transfected cells were selected with 1 µg ml⁻¹ puromycin (Invivogen) for 3 d starting at day 1 post transfection. Cells were collected 5 d post transfection and lysed with the DirectPCR Lysis Reagent (Cell) (Viagen), supplemented with Proteinase K (ThermoFisher) according to the manufacturer's protocol. Genomic DNA extracted from these cells was used for PCR amplification of on- and off-target sites and amplicon sequencing.

Flow cytometry. To determine the levels of *EpCAM* editing by Cas9_wt, Cas9_R63A/Q768A and Cas9_R66A/Q768A, cells were stained with human EpCAM (CD326) antibody conjugated to FITC (Miltenyi Biotec) according to the manufacturer's instructions. Dead cells were excluded from the analysis by staining with the 7-AAD viability staining solution (BioLegend). Samples were acquired with the Sony SH800 cell sorter and data were analyzed with the FlowJo software v.10.5.3 (Tree Star).

Amplicon sequencing. On-target and off-target sites were PCR-amplified with Phusion High-Fidelity DNA Polymerase (Thermo Scientific) using primers listed in Supplementary Table 8. The following PCR program was used: (98 °C, 10 s; appropriate annealing temperature for each primer pair, 15 s; 72 °C, 30 s) × 35 cycles (30 cycles for nested PCRs), with the addition of DMSO if necessary. Libraries were prepared with 10 ng DNA for each sample using the KAPA HyperPrep-Kit (Roche), according to the manufacturer's instructions, without fragmentation and size selection. This was followed by eight cycles of PCR to add sequencing adapters. After quality control, libraries with similar size were pooled together, resulting in four pools of 14 libraries, respectively one pool of 20 libraries. These pools were quantified with the KAPA Library Quantification Kit (Roche), normalized to 2 nM and pooled again equimolarly to load them on the MiSeq.

Fastq data was analyzed using the ampliCan[50] pipeline with the following parameters: fastqfiles=0.5, average_quality=30, min_quality=0. Briefly, each read was subjected to quality control, requiring an average base call quality greater than 30, and no ambiguous nucleotides. Filtered reads were aligned via Needleman–Wunsch algorithm to their expected amplicon sequence, given their flanking primers, as extracted from human genome reference GRCh38. The results of the ampliCan pipeline were tallied and reported as the total edited and frameshift indels divided by the number of filtered reads passing quality control.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The datasets generated and analyzed during this study are available from the corresponding author upon request. Raw pictures for all in vitro experiments are available as Supplementary Note 1 online. Amplicon sequencing data are available online (accession code PRJNA577558).

References

39. Maniatis, T., F Fritsch, E. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Wiley, 2018).
40. Kirsch, R. D. & Joly, E. An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes. *Nucleic Acids Res.* **26**, 1848–1850 (1998).
41. Langley, K. E., Villarejo, M. R., Fowler, A. V., Zamenhof, P. J. & Zabin, I. Molecular basis of beta-galactosidase alpha-complementation. *Proc. Natl Acad. Sci. USA* **72**, 1254–1257 (1975).
42. Ullmann, A., Jacob, F. & Monod, J. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*. *J. Mol. Biol.* **24**, 339–343 (1967).
43. Richter, F. et al. Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Res.* **44**, 10003–10014 (2016).
44. Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 2577–2590 (2014).
45. Zuo, Z. et al. Engineering *Pseudomonas putida* KT2440 for simultaneous degradation of organophosphates and pyrethroids and its application in bioremediation of soil. *Biodegradation* **26**, 223–233 (2015).
46. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297 (2017).
47. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).

48. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23–28 (1990).
49. McClelland, M., Hanish, J., Nelson, M. & Patel, Y. KGB: a single buffer for all restriction endonucleases. *Nucleic Acids Res.* **16**, 364 (1988).
50. Labun, K. et al. Accurate analysis of genuine CRISPR editing events with ampliCan. *Genome Res.* **29**, 843–847 (2019).

Author contributions

E.C. oversaw the study. M. Bratovič, I.F., K.C., M.Boettcher and E.C. designed experiments. M. Bratovič and I.F. performed all in vitro experiments and bacterial survival assays, and analyzed all data. M. Bratovič and K.C. performed experiments in eukaryotic cells. S.B. and B.T. performed library preparation and amplicon sequencing. E.J.C.G. and T.J.S. analyzed the sequencing data and performed statistical analysis. M. Bratovič, I.F. and E.C. wrote the manuscript, which all other authors commented on and approved.

Competing interests

E.C. is cofounder of ERS Genomics and CRISPR Therapeutics and is a member of the scientific advisory board of CRISPR Therapeutics. All other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-020-0490-4.

Correspondence and requests for materials should be addressed to E.C.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Emmanuelle Charpentier

Last updated by author(s): Nov 7, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Image Lab TM software (Bio-Rad Laboratories, Inc, version 5.2.1), Typhoon TM FLA9500 (GE Healthcare, version 1.1 Build 1.1.0187 (2015)) and Gen5 TM (Biotek, Version 3.03) were used to collect data from kinetic cleavage assays, DNA binding assays and bacterial survival assays, respectively. MO.Control software (NanoTemper Technologies, version 1.5.3*) was used to collect data from RNA binding assays. SH800S software (Sony, Version 2.1.5) was used to collect data from flow cytometry.
Data analysis	GelAnalyzer (available from http://www.gelanalyzer.com/index.html, version 2010a) and Image Lab TM software (Bio-Rad Laboratories, Inc, version 5.2.1) were used to quantify binding assays and kinetic cleavage assays, respectively. Origin Software (OriginLab, Northampton, MA, Version 9.4) was used to calculate the cleavage rates and binding constants, and to perform statistical analysis of bacterial survival assays. FlowJo TM version 10 software (Tree Star) was used to analyze flow cytometry data. Amplicon sequencing data was analyzed by ampliCan.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated and analyzed during this study are available from the corresponding author upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to predetermine the sample size. Sample sizes were determined based on at least three independent experiments or more. These sample sizes are in accordance to other studies in the field and enable statistical analysis and reproducibility.
Data exclusions	No data were excluded.
Replication	All attempts of replication were successful. All experiments are described in detail to allow replication.
Randomization	Not applicable, not a common procedure for the methods employed.
Blinding	Investigators were not blinded during data acquisition and analysis since it is not a common procedure for the methods employed.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CD326 (EpCAM)-FITC, human (Miltenyi Biotec), clone HEA125, order number: 130-113-263, dilution 1:50.
Validation	The antibody was validated by the vendor and the data is available on their official website (https://www.miltenyibiotec.com/DE-en/products/macs-flow-cytometry/antibodies/primary-antibodies/cd326-epcam-antibodies-human-hea-125-1-50.html#fitc:for-100-tests)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	All cell lines used in the study were obtained from ATCC (MCF-7 cells: ATCC HTB-22TM, HEK293 cells: ATCC CRL-1573TM)
Authentication	None of the cell lines were authenticated.
Mycoplasma contamination	HEK293 cells were tested for mycoplasma contamination and the result was negative. MCF-7 cells were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	MCF-7 cell line was used due to the fact that this cell line expresses EpCAM, which was targeted in the in vivo gene editing experiments. HEK293 cell line is not listed in ICLAC register.

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	MCF-7 cells were stained with the human EpCAM (CD326) antibody conjugated to FITC (Miltenyi Biotec) according to the manufacturer's instructions.
Instrument	Sony SH800 cell sorter.
Software	SH800S software (Sony) was used to collect data, whereas lowJo TM v10 software (Tree Star) was used to analyze data.
Cell population abundance	Not applicable (no cell sorting was performed).
Gating strategy	Dead cells and doublets were excluded from the analysis based on FSC-A/SSC-A scatter plot and by staining with 7-AAD viability staining solution.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.