nature microbiology

ARTICLES

https://doi.org/10.1038/s41564-019-0382-0

Target-dependent nickase activities of the CRISPR–Cas nucleases Cpf1 and Cas9

Becky Xu Hua Fu[1,2*], Justin D. Smith[3,4], Ryan T. Fuchs[5], Megumu Mabuchi[5], Jennifer Curcuru[5], G. Brett Robb[5*] and Andrew Z. Fire[1,2*]

Clustered regularly interspaced short palindromic repeats (CRISPR) machineries are prokaryotic immune systems that have been adapted as versatile gene editing and manipulation tools. We found that CRISPR nucleases from two families, Cpf1 (also known as Cas12a) and Cas9, exhibit differential guide RNA (gRNA) sequence requirements for cleavage of the two strands of target DNA in vitro. As a consequence of the differential gRNA requirements, both Cas9 and Cpf1 enzymes can exhibit potent nickase activities on an extensive class of mismatched double-stranded DNA (dsDNA) targets. These properties allow the production of efficient nickases for a chosen dsDNA target sequence, without modification of the nuclease protein, using gRNAs with a variety of patterns of mismatch to the intended DNA target. In parallel to the nicking activities observed with purified Cas9 in vitro, we observed sequence-dependent nicking for both perfectly matched and partially mismatched target sequences in a *Saccharomyces cerevisiae* system. Our findings have implications for CRISPR spacer acquisition, off-target potential of CRISPR gene editing/manipulation, and tool development using homology-directed nicking.

Bacteria and archaea are constantly challenged by invasive genetic elements (for example, bacteriophages, transposons, and plasmids). To combat these threats, prokaryotes evolved an adaptive immune system known as CRISPR–Cas (CRISPR and CRISPR-associated proteins)[1–4]. This immune system is able to capture foreign genetic elements into repeats in the CRISPR loci as short DNA segments. These captured 'spacer' sequences are expressed in the context of precursor CRISPR RNAs (pre-crRNAs), which are processed into small CRISPR RNAs (referred to as gRNAs)[1–4]. CRISPR–Cas proteins use the gRNA complementarity and often a protospacer adjacent motif (PAM) to recognize and cleave the target, thus conferring immunity to the invading elements[1–4]. The CRISPR–Cas systems have been adapted as versatile genome editing tools that are ubiquitously used in many disciplines[5].

Streptococcus pyogenes Cas9 is a type II-A CRISPR–Cas nuclease that is widely studied and used in genome editing and epigenome manipulation[6,7]. Cas9 is a blunt cutting nuclease that can target specific DNA sequences using an NGG PAM and a gRNA[8]. Many variants of Cas9 have been developed (for example, transcription activating/repressing, nicking, nuclease-dead, base editing)[9–11]. One set of observations of particular note in reference to such variants are experiments where nicking of targets can be used to guide homology-dependent gene editing, with apparent advantages in specificity[12] and efficiency[13] over double-stranded break strategies. These assessments were carried out with mutant variants of Cas9 where one of the nuclease domains has been inactivated. Sternberg et al.[14] describe a conformational shift mechanism that would arguably coordinate the two nuclease domains to provide concerted double-strand cleavage without substantial accumulation of nicked intermediates, while Szczelkun et al.[15] (Supplementary Fig. 6b) describe one case of a truncated gRNA:target match that leads to slower coordination and accumulation in vitro of a nicked substrate. In addition, Strohkendl et al.[16] reported temporary nicking with Cpf1 due to mismatches influencing the timescale of cleavage of each DNA strand.

Cpf1 is a minimal type V CRISPR–Cas nuclease that uses a single Cas endonuclease paired with a gRNA to cleave complementary DNA targets[17–20]. Cpf1 homologues from three species have been successfully adapted for genome editing: *Acidaminococcus* sp. (AsCpf1), *Francisella novicida* (FnCpf1), and *Lachnospiraceae bacterium* (LbCpf1)[18]. Cpf1 homologues recognize T-rich PAMs (reported as TTTN/TTTV for As/LbCpf1 and TTV for FnCpf1), theoretically opening a wide array of sites not available for Cas9 editing[17,21]. Cpf1 processes its own pre-crRNA, allowing this to be a potential platform for multi-gene functional analysis[20,21]. Despite numerous possible advantages, Cpf1 homologues have yet to be used as extensively in genome editing applications as Cas9.

Examining the consequences of Cpf1 and Cas9 activities on DNA topology for a broad set of potential substrates, we observed efficient nicking activities with both nucleases on specific classes of mismatched DNA targets. Given that either nicking or double-stranded cleavage of DNA is sufficient to induce a variety of repair and replacement mechanisms in vivo[13,22–26], the resulting profiles illuminate a dual capability of CRISPR–Cas nucleases to initiate genetic change through both types of interaction. These observations challenge binary models in which CRISPR nucleases either cleave or fail to cleave individual targets. Instead, our data indicate a scenario where cleaving and nicking targets can and will coexist in a single experimental or natural condition.

Results

Sequence dependence for CRISPR–Cas nuclease activity was first addressed by building variant libraries of plasmids with a diversity of perfectly matched ('wild-type' (WT)) and mismatched ('mutant') target sequences. Each sequence (WT or mutant) is represented by several barcoded species. We challenged the variant library with CRISPR–Cas nuclease protein programmed with single-guide RNA, and used sequencing to determine which templates remained following nuclease treatment. Additional information on the nature of

[1]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. [2]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [3]Stanford Genome Technology Center, Stanford University, Palo Alto, CA, USA. [4]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [5]New England BioLabs Inc., Ipswich, MA, USA. *e-mail: xhbeckyfu@gmail.com; Robb@neb.com; afire@stanford.edu

the cleaved substrates comes from assays with and without a whole-library (backbone) linearization step after CRISPR–Cas nuclease cleavage of substrates and before PCR (Fig. 1a). In particular, we note that backbone-intact assays provide a potential to distinguish nicked from closed circles due to the higher PCR yields observed from molecules where a nicking event has allowed the two plasmid strands to become topologically unlinked[27].

Cpf1 nicking and cleavage specificity. Sequence dependence for LbCpf1 activity was investigated using a variant library of plasmids for four different canonical guide sequences: enhanced green fluorescent protein 1 (EGFP1), EGFP2, *unc-22A*, and *rol-6* (Fig. 1b). Initial assays were carried out with the backbone linearization step (Method 1, Fig. 1a). For each target, the libraries contained unmodified sequences, single-base variants, double-adjacent transversions, and deletions. Using loss of amplification across the cleavage junction as an assay, LbCpf1 showed a similar specificity profile to previous characterizations of Cas9, less tolerance of mutations in the seed region (positions 1–10, with position 1 being PAM-proximal), while sequence requirements in the distal region (positions 11–20) were more lax (left panel of Figs. 2 and 3; left panel of Supplementary Fig. 1–6). This trend is consistent for all three classes of variants, for all Cpf1s, and on all tested targets (left panel of Supplementary Fig. 1–6).

An unexpected feature of Cpf1 cleavage was observed in assays with the circular variant library where no linearization was carried out before assessing retention (Method 2). Certain target site variants had highly positive retention scores, which suggest these variants increase (rather than decrease) their representation during a short-time cleavage reaction. Log retention scores were observed as high as ~3 (that is, 2^3 or eightfold enrichment) for double mutants from time points 1, 3, 10, and 30 min (right panels of LbCpf1: Figs. 2 and 3; AsCpf1 and FnCpf1: Supplementary Figs. 7–27). For some targets, these effects disappeared at longer reaction times, as expected if there was eventual cleavage of the circular targets (Figs. 2 and 3). The enhanced recovery led to the hypothesis of a rapid nicking of the initial substrate followed by a relatively slow cleavage on the opposite strand (nicked plasmids enhance PCR amplification compared to linear and supercoiled substrates[27]). This would explain the observed over-representation after the addition of Cpf1 and the highly positive retention score followed by much slower loss of the observed retention.

The nicking hypothesis was tested by reacting Cpf1::gRNA complexes with individual plasmid substrates and examining the topology using ethidium-containing native agarose DNA gels. Sequence variants with highly positive retentions were identified in the high-throughput data and either cloned into a DNA plasmid or synthesized as a gRNA. Using a DNA target with double consecutive transversion mutations in positions 12–13 and 14–15 in WT EGFP1 gRNA, we found LbCpf1 was rapidly capable of converting closed circular to nicked circular substrates in these assays (Fig. 4a,b). In addition, the nicking ability was found to be specific, with no observed nicking of substrates lacking homology to the gRNA (Fig. 4c). The nicking ability was confirmed in AsCpf1 and FnCpf1 with their respective EGFP1 gRNAs (Supplementary Figs. 28 and 29).

To evaluate the determinants for nicking and cleavage, we carried out a number of reciprocal experiments where mutated gRNAs were used in assays with WT targets. These assays showed nicking and some linearization, as expected if the ability to form a nicking enzyme is a general feature of certain classes of gRNA::target mismatch. We note that there was some non-equivalence in the target-mutated versus gRNA-mutated assays, depending on the individual gRNAs (Supplementary Fig. 30). These data indicate an interaction where a specific sequence along with the pattern of mismatches determine the balance between nicking and cleavage activities.

These observations raise the possibility that an appropriately designed gRNA might produce effective single-strand nicking activity on an arbitrary substrate. Some possible guidelines for such design are suggested from the patterns of mismatch effects on initial sequence recovery in Figs. 2 and 3 and Supplementary Figs. 1–6. In particular, we see a strong tendency to nick for templates with a combination of transversion point mutations in the distal region (positions 9–15). Taking consecutive double mutations at positions 12 and 13 as a provisional lead for such assays, we first tested the ability to produce nicking activities towards additional targets where no high-throughput analysis of target specificity had been carried out (targets dm22085 (DNMT1), fc596 (FANCF), and wp1058 (WTAP)). While these three sequences show different degrees of cleavage for matched gRNA::target combinations, all three bias towards nicking with the indicated double mutant target (Supplementary Figs. 31–33).

Cas9 nicking and cleavage specificity. The type II CRISPR effector enzyme Cas9 has been a workhorse tool for genome editing and for a wide variety of experimental applications in vitro and in vivo. We tested *S. pyogenes* Cas9 for nicking activity in assays and libraries that (as with the tested Cpf1 libraries) contained WT, single variants, single deletions, and double consecutive mutations in four targets: EGFP1, EGFP2, *unc-22A*, and *rol-6* (Fig. 1b). While yielding some evidence of preferential nicking of specific substrates, these assays yielded a less dramatic distinction between nicking and double-strand cleavage for individual template sequences than had been seen with Cpf1 (Supplementary Figs. 34–45). Encouraged by the differences (but with the knowledge that an optimal nicking activity might require more extensive mutational analysis), we repeated the Cas9 assays using an *unc-22A* library obtained through random oligonucleotide synthesis with a broader set of multiple mutations (Fu et al.[28,29]; Supplementary Fig. 46). These assays showed a strong preferential nicking for a variety of double mutants (Supplementary Fig. 47). Examining an extended list of the randomly mutagenized targets from this library for which positive retention scores indicated a strong nicking activity (Supplementary Table 1), we chose a design with two mutations, a single deletion in the seed region at position 5 and a mismatch (A–G) in the distal region at position 18 for the *unc-22A* target. DNA templates with this double variant reacted with Cas9 complexed to WT *unc-22A* gRNA confirmed robust target-specific nicking (Fig. 4d). EGFP2 variant targets that had a highly positive retention score in the high-throughput assays were likewise confirmed as nicking substrates in agarose gel assays (Fig. 4e). Conversely, WT *unc-22A* and EGFP2 targets were assayed for nicking with the equivalently mismatched (mutant) gRNA. Similar to Cpf1, the equivalent mutations in the gRNA with WT targets showed nicking and some linearization that varied depending on the gRNAs (Supplementary Figs. 48–49). Finally, we observed that some non-intended targets were nicked (for example, a variant target sequence of *unc-22A* nicking with WT EGFP2 gRNA). We hypothesized that targets with certain minimal and/or broken homologies can induce nicking. We confirmed this by testing non-targeting DNA with gRNAs that showed positive retention scores and confirmed the nicking observed in the high-throughput results. For example, we observed in the high-throughput assays a variant of the *unc-22A* target that evidently nicks with EGFP2 gRNA (Supplementary Fig. 50). This variant brings the *unc-22A* sequence somewhat closer to the EGFP2 sequence, although still providing rather limited homology (contiguous complementarity confined to six consecutive matches in the seed region). While demonstrating a less stringent sequence requirement, nicking remained a specific process, with only residual nicking for unrelated targets observed in the gel assay (for example, >16-h time point, Supplementary Fig. 51).

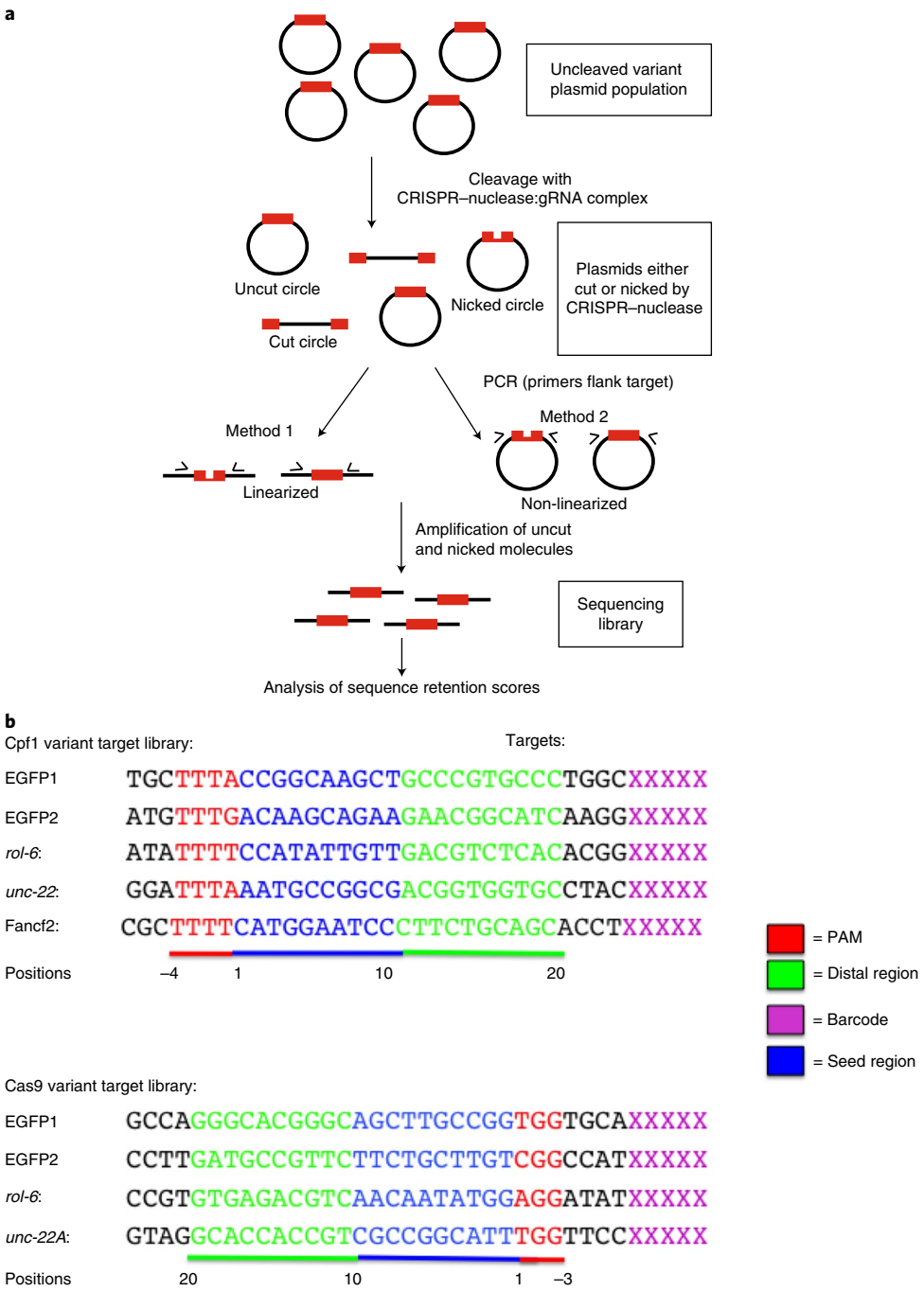


Fig. 1 | High-throughput assay for nicking and cleavage by CRISPR–Cas nucleases. a, Schematic of in vitro high-throughput plasmid libraries and subsequent steps used to assess representation or each sequence before and after CRISPR–Cas interaction. Each library consists of a uniform backbone into which a variety of potential target sites are inserted at a single location. Potential targets match one of several gRNA or control sequences (**b**), with each library including a diversity of both matched and mismatched sequences. To assess representation for each sequence before and after CRISPR–Cas interaction, samples were amplified with primers flanking the potential cleavage site and subjected to high-throughput sequencing. Amplification and sequencing of an unreacted sample (no CRISPR–Cas interaction) provides a baseline fraction of normalization of each variant; this reference amplification and sequencing is carried out for every experiment and used for subsequent normalization. To assess topological state before and after CRISPR–Cas interaction, samples were split into two pools that were treated identically with the exception that one of the two subpools is cleaved at a site outside of the region to be amplified ('Method 1'), while the other pool is not subjected to an outside cleavage ('Method 2'). The amplification of uncleaved circular templates (Method 2) is known to enrich for nicked over closed circular templates[27]. **b**, Depiction of target sequences synthesized for the Cpf1 and Cas9 variant libraries. Regions in red indicate the PAM, blue indicates the seed region (positions 1–10), green indicates the distal region (positions 11–20), and magenta is the barcode region. For each target, the following variants were synthesized: WT; single variants; single deletions; and double consecutive transversion variants.

Tuning of relative nicking activity. Analysis of complex libraries at multiple time points or with differing enzyme:substrate ratios provides a combined view that allows selection of substrates with a high preference for nicking over cleavage, with little or no cleavage even on extended incubation. Such substrates are exemplified for Cas9 by *unc-22A* double transversion (positions 5 and 12)

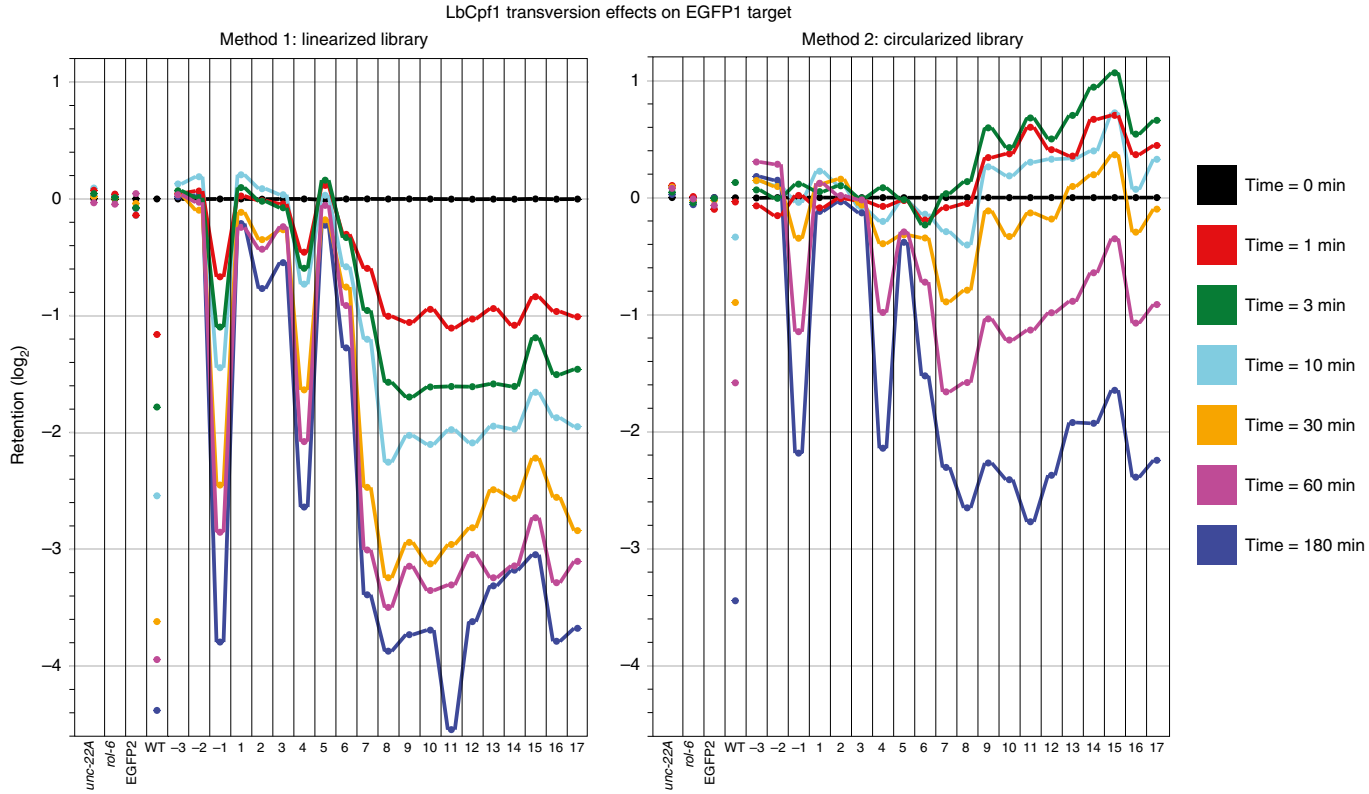


Fig. 2 | Library-based assessment of nicking and cleavage activities of LbCpf1 on single mutant target variants. The line graph on the left shows the retention profiles for the whole-library assays with a post-reaction backbone cleavage step that avoids preferential recovery of nicked substrates. The line graph on the right shows the retention profiles without such a cleavage step (that is, with preferential recovery of nicked sequences). For the assays shown in this figure, nuclease LbCpf1 was programmed with an EGFP1 gRNA and interacted with a mixed target library as described in Fig. 1a. Retention scores are shown for single-base transversions at each indicated position in the EGFP1 target, and have been normalized using median retention for a set of unrelated target sequences included in the library (*unc-22A*, *rol-6*, and EGFP2; sequencing data: AF_SOL_820). For details on retention score calculation, see Methods.

(Supplementary Fig. 47). For LbCpf1, maximal nicking is seen at certain time points with double consecutive transversion mutations in positions 12/13 and 14/15 (Fig. 3 and Supplementary Figs. 1–6). These examples illustrate the value of an optimization round in identifying the most specific nicking reagents for a given target, and of the value of a broad variant survey in characterizing potential off-target nicking consequences for a given gRNA.

Nicking in vivo in *S. cerevisiae*. Assessments of Cas9 activity in vivo were carried out in *S. cerevisiae*, using the methodologies of Cas9 and gRNA expression and of target library production described in Fu et al.[29]. These high-throughput assays allow tracking of topology as a function of sequence for a target library, allowing the determination of which templates in a complex pool are cut, which are nicked, and which are not cut. Assays in yeast allow the assessment of nuclease activities in a system with rapid cell division and where the DNA targets are present in a chromatinized context (in contrast to the supercoiled plasmid context of the assays, which may be much more representative of bacterial than of eukaryotic chromosomes).

While sequencing provides a valuable assay for determining the population of molecules present before and after exposure in vivo to a Cas9 + gRNA pair, it remains important to approach such data with considerable care. One concern is that the in vivo situation in yeast represents a dynamic equilibrium between any cleavage or nicking by the Cas9 enzyme and repair processes (that might fix a nick or break) or other cellular processes (for example, replication across a nick, generating a break) that might interconvert the various states. To investigate in vivo activities, we evaluated the incidence of nicked substrates in vivo rather than the much more complicated kinetic question of determining the rate of new nicks in the presence of repair and other activities.

In developing assays to test nicking of DNA in vivo, the simple detection of relaxed circles among extracted DNAs is not sufficient to infer whether nicking had occurred in vivo, or whether nicking occurred during extraction and/or analysis of DNA. To ensure definitive resolution of this issue, we asked a number of additional questions. Is the nicking signal (enhanced PCR yield observed in DNA pools containing close matches to the guide) reproducible in parallel samples and in different biological backgrounds? Is the nicking signal specific to molecules in the target pool with homology to the Cas9 gRNA and PAM site? For partially matched targets, does the nicking signal decrease for target sequences with substantial numbers of gRNA mismatches? Does the nicking signal increase with induction of Cas9 and with longer exposure times in vivo? Does the in vivo nicking signal for different target variants correlate with measured in vitro nicking?

Our analysis found all these criteria fulfilled. The nicking signal (preferential retention of gRNA-matched targets in samples processed from uncut yeast DNA) is consistent between replicates of induced yeast (Fig. 5a), depends on both PAM sequence and gRNA homology (Fig. 5a–c), is lost with multiple mutations in the target (Fig. 5b,c, 4–7 mismatch lane), increases with longer exposure times (Fig. 5b,c, comparing single-generation and 2.5 generation samples), and is correlated ($R = 0.47$, $P = 0.00015$; $n = 66$) with nicking observed in vitro (Fig. 5d–f).Thus, these data provide support for

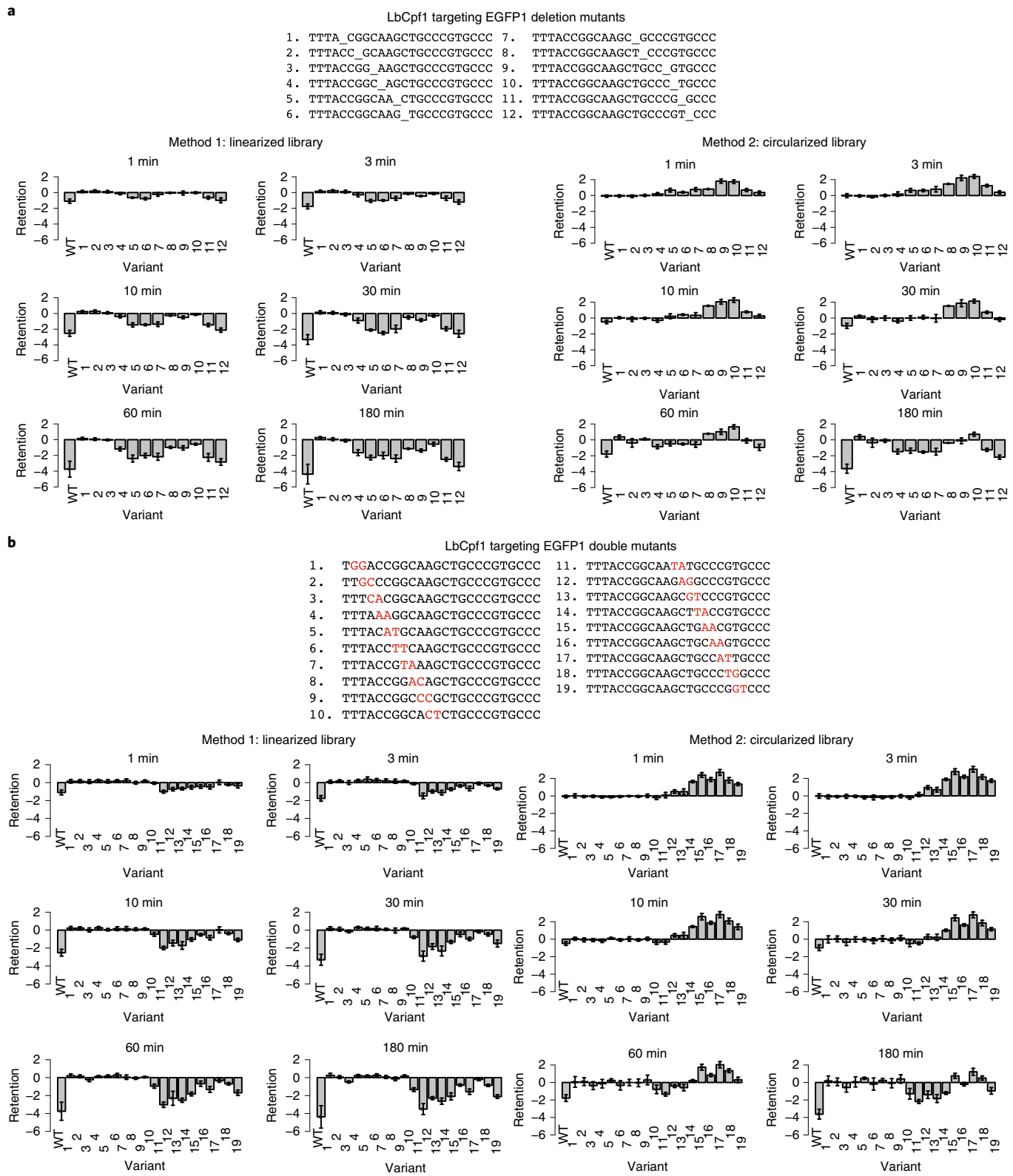


Fig. 3 | Library-based assessment of nicking and cleavage activities of LbCpf1 on double consecutive transversion and deletion target variants. The bar graphs on left show the retention profiles for whole-library assays with a post-reaction backbone cleavage step that avoids preferential recovery of nicked substrates. The bar graphs on the right show the retention profiles without such a cleavage step (that is, with preferential recovery of nicked sequences). **a**, Single deletion variants. **b**, Double consecutive transversion variants. The bar graphs show the median retention of indicated variants at each base, with the error bars representing the s.d. of observed retention among distinct barcoded instances of each variant target sequence (left: linearized library; right: circular library; sequence data: AF_SOL_820). For details on retention score calculation, see Methods. The number of distinct barcoded instances for each variant assessed (*n* for s.d. calculation) were as follows. Left: WT: 26, deletions 1–12: 7, 9, 8, 9, 6, 3, 10, 2, 6, 6, 3, 6. Consecutive transversions 1–19: 19, 24, 11, 4, 5, 8, 7, 6, 6, 9, 9, 5, 8, 11, 7, 8, 8, 5, 5. Right: WT: 24, deletions 1–12: 6, 8, 8, 9, 6, 4, 6,1, 5, 6, 2, 5. Consecutive transversions 1–19: 15, 20, 9, 5, 3, 8, 5, 5, 8, 9, 9, 6, 9, 11, 6, 6, 5, 5, 4.

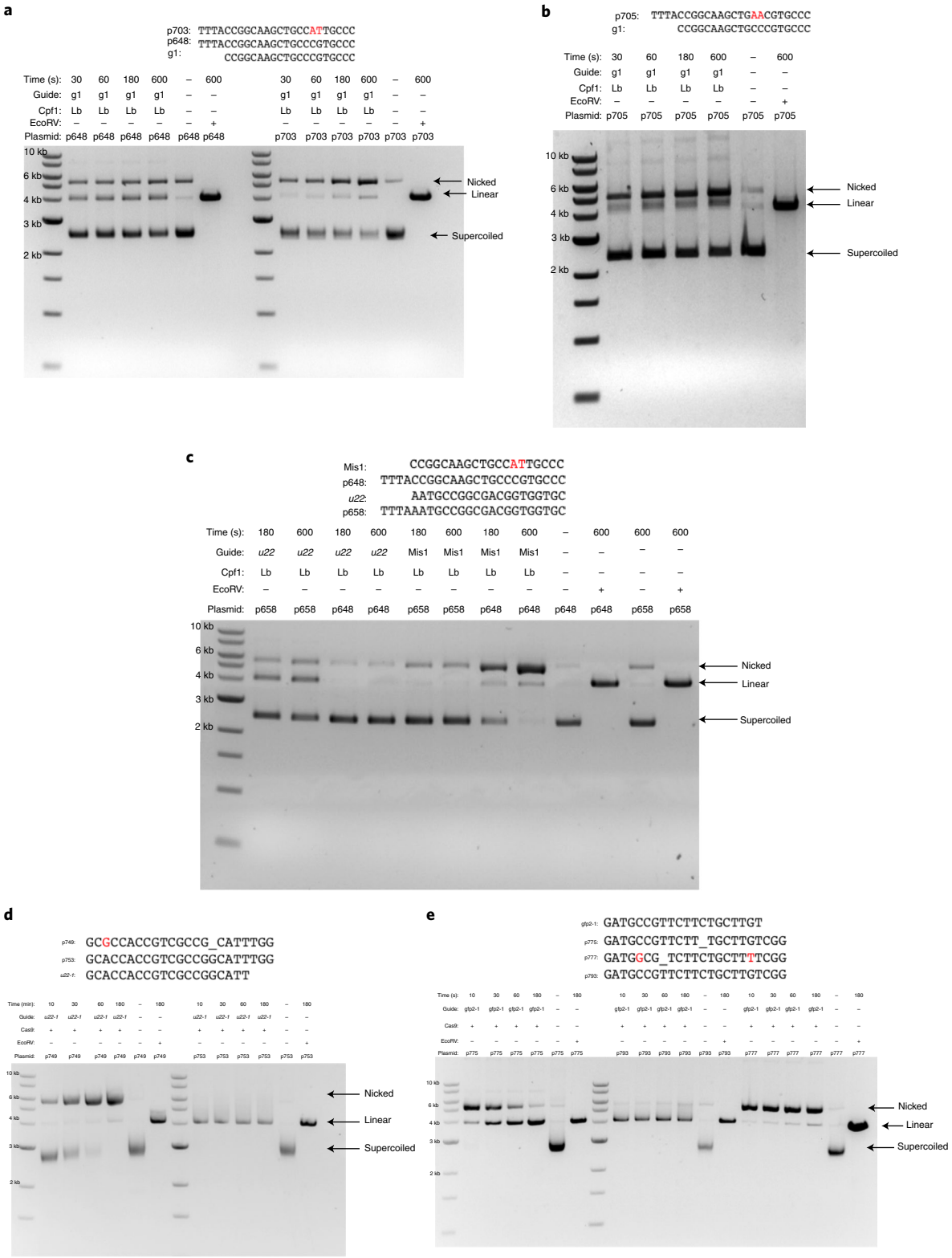


Fig. 4 | Gel-based assessments of nicking and cleavage by LbCpf1 and Cas9. a–**c**, Gel-based assays of nicking by LbCpf1. **a**, Left: WT EGFP1 target (p648) reacted with WT EGFP1 gRNA. Gels show a mixture of linear, nicked, and uncut target plasmids. Right: a mutated EGFP1 target (p703) with WT EGFP1 gRNA showed preferential accumulation of nicked plasmid. **b**, Gel assay with mutated EGFP1 target (p705) and WT EGFP1 gRNA. Preferential nicking with some linearization is observed. **c**, Specificity assessment using nicking gel assay. Tested gRNAs are WT *unc-22A* (*u22*) and mismatched EGFP1 (Mis1). Tested targets are WT *unc-22A* (p658) and EGFP1 (p648). The *u22* gRNA linearizes the WT *unc-22A* (p658) target while having no effect on the EGFP1 (p648) target. The mismatched EGFP1 gRNA (Mis1) promotes nicking when paired with the EGFP1 target but not with *unc-22A*. Nicking is thus shown to be RNA-guided and specific. **d**,**e**, Specificity and mismatch effects on Cas9 nicking and cleavage activities. **d**, Cas9 activity WT *unc-22A* gRNA and mutated *unc-22A* target (left), and with WT *unc-22A* gRNA and WT *unc-22A* target (right). Cas9 can be seen to efficiently nick at the mismatched target. **e**, WT EGFP2 gRNA with mutated EGFP2 DNA targets (p775 and p777) and WT EGFP2 with WT EGFP2 DNA (middle). Both mutated EGFP2 targets are nicked efficiently. The p775 mutation, if given enough time, eventually linearizes the plasmid (left). The p777 mutation remains nicked through the time course (right).

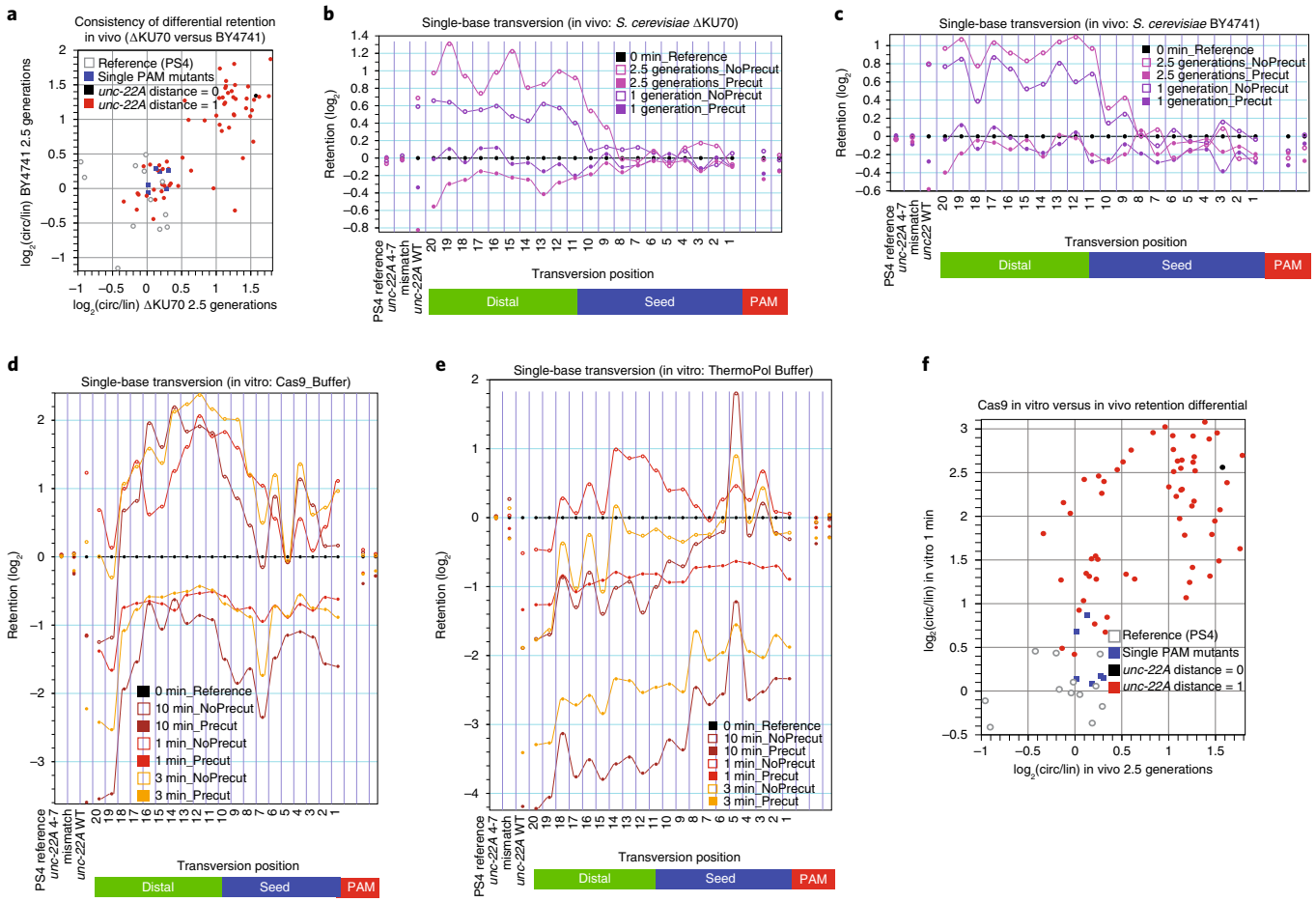


Fig. 5 | Assays for target-match-dependent nicking of precise and imprecise targets by Cas9 in vivo. a, Reproducible nicking signal for matched targets following Cas9 induction in yeast. This plot compares differential retention ($(\log_2$(circular_assay_retention) – $\log_2$(linear_assay_retention)) for induced libraries from two different but functionally equivalent yeast strains, BY4741 and ΔKU70. The individual dots represent different target sequences, each assessed as a median over multiple barcoded instances in each library. Targets unrelated to *unc-22A* (grey hollow circles; 'protospacer 4' variants[28]) and with perfect *unc-22A* match in the gRNA homology but mismatches in the GG PAM (blue squares) show no substantial differential retention in either sample. Perfectly matched targets with canonical 'NGG' PAM sites (black circle) show substantial differential retention in the two yeast populations. Targets with single mismatch (red dots) show a spectrum of different retention differentials, ranging from no difference to differences comparable to the perfect target match. The two yeast populations give highly similar results with calculated sample-to-sample correlations between the two yeast populations of 0.81 (Pearson's correlation; two-tailed $P = 9.6 \times 10^{-17}$, $n = 66$) and 0.76 (Spearman's correlation; two-tailed $P = 1.6 \times 10^{-13}$, $n = 66$). For details on retention score calculation, see Methods. **b**–**e**, Similar sequence requirements for nicked substrate accumulation under diverse in vitro and in vivo conditions. The top graphs show circular and linear assay retention scores for yeast (in vivo) experiments in the BY4741 and ΔKU70 genetic backgrounds. Each plot shows the median retention for multiple (barcoded) transversion variants at each position (averaged where duplicate samples for the relevant conditions are available). Retentions are calculated using the unrelated PS4 sequence as an internal standard, with initial library abundance obtained from the measurement of target species incidence in parallel libraries with yeast that have not been induced. Below are the equivalent retention score plots for in vitro analysis where an equivalent library of targets was interacted with purified Cas9 in one of two buffer conditions (a relatively active 'ThermoPol' buffer condition and a less active 'Cas9 buffer' condition; see Methods for buffer details). **f**, Differential retention comparison between in vitro and in vivo samples. The in vivo sample shown here is from BY4741 (at 2.5 generations) compared to the initial rate (1 min) of nicking in vitro (Cas9 buffer). Correlations are Pearson's (0.54, two-tailed $P = 3.6 \times 10^{-6}$) and Spearman's (0.49, two-tailed $P = 3.2 \times 10^{-5}$); $n = 66$. Each point on the plot represents a single gRNA homology + PAM sequence class, showing a median of differential retentions derived from independent flanking-barcoded instances of each sequence.

the hypothesis that the state of extracted DNA reflects the configuration in vivo and that a fraction of this DNA is indeed in a nicked form. Based on the observed differential retention (a maximum of approximately threefold in vivo) and the maximum enhancement for fully nicked DNA (similarly found in Lin et al.[27] and our observations), we estimate that between one-sixth and one-third of susceptible targets are nicked at any time in vivo. We note a modest difference among in vitro conditions (two different buffers) and the in vivo yeast observations. All show nicking of mismatched (and to some extent, fully matched) targets, while differences are seen in the relative proportions for different sequences. Of interest, a substantial proportion of nicked targets is observed to persist even for fully matched Cas9 targets in the yeast assays (Fig. 5b,c, black dots).

Discussion

We show that two families of RNA-programmed CRISPR–Cas nucleases, As/Fn/LbCpf1 and Cas9 can efficiently nick DNA targets with perfect and/or mismatched homology to the programming gRNA. The ability to programme native CRISPR–Cas nucleases to nick has uses in technology and implications in the biology of CRISPR immunity.

The ability to tailor some CRISPR–Cas nucleases to nick at specific sites 'on demand' provides a potential alternative to mutated CRISPR nuclease variants for gene editing/manipulation. While mutation of defined cleavage domains[8,30,31] provides a capability for nicking activities for some CRISPR nucleases, the approach is not available for all CRISPR–Cas systems (particularly Cpf1, where a single domain may execute both cleavage reactions[20,32,33]). In addition, such approaches limit the multifunctional applications of CRISPR in a single system, since co-expression of nickase and WT enzymes will generate cleavage at intended nickase sites as well as nicking at intended cleavage sites. Use of a single WT enzyme with gRNAs for nicking and cleavage would surmount this challenge, with tuning of gRNAs a likely requirement in making such an approach effective. While we have not investigated such applications in microbial or other systems (for example, mammalian cells), the use of mismatched gRNAs with WT CRISPR enzymes to direct nicking could be valuable for gene editing or replace nickases in gene manipulation. As an example, base editors composed of a catalytically dead Cas9 fused to a deaminase could potentially be redesigned to involve a fully functional Cas9 fused to a deaminase that can be guided to specific areas via mismatched gRNAs.

This study also highlights the fact that a fraction of Cas9's observed activities in vivo, either in native systems or in engineering applications, may reflect nicking rather than cleavage activities, with the balance presumably dependent on in vivo conditions as well as the sequences of gRNAs and targets. Of particular importance in this regard, observations that nicking activities can provide advantages in genome editing[13,22,23,25,26] indicate that such conditions could prove advantageous. We note that the consequences of nicking depend in each system on the kinetic balances between nick ligation, single-stranded exonuclease and endonuclease activities that might extend or convert nicks[34], other modes of DNA repair, and DNA replication/division rates. These kinetic parameters vary substantially based on intrinsic cellular properties, on specific genomic positions, and on stochastic ordering of events. Of particular interest, recent work by Davis and Maizels[13] provides assays for nick-induced repair and editing in mammalian systems.

An understanding of sequence requirements for nicking as well as full cleavage of targets is critical for the identification and assessment of potential off-target effects of CRISPR–Cas nucleases in vitro and in vivo as these are applied in experimental, biotechnological and (eventually) clinical settings. In particular, the existence of a guide-specific nicking repertoire impacts the selection of gRNA targets to avoid off-target consequences during genome editing. Although nicked double-stranded DNA in vivo is less detrimental than a double-stranded break, nicks lead to downstream repair events that can cause unexpected mutations (for example, see Kuzminov[24]). Of the various algorithms that score and/or pick gRNAs based on potential off-target effects[35,36], many entail a user-selected threshold for candidate gRNA mismatch and/or leave the user to select candidate gRNAs based on a mismatch count-based estimate of off-target cleavage potential. Understanding the nicking abilities of CRISPR–Cas nucleases may thus offer considerable value in gRNA design and selection.

Imperfect homology-dependent nicking of CRISPR–Cas nucleases has implications on the mechanism of spacer acquisition in CRISPR immunity. Spacer acquisition is the process where nucleic acids from foreign genetic elements (for example, plasmids, bacteriophages) are integrated in the CRISPR loci. The integrated spacers are later transcribed and processed and used for host defence[1,37]. Cas1 and Cas2 catalyse spacer acquisition[38,39]. It has been shown that the effector CRISPR–Cas nucleases (active and inactive nuclease) in type II CRISPR systems are necessary for spacer acquisition[40,41]. In addition, there is evidence that type I CRISPR–Cas systems in the presence of non-targeting gRNAs can increase spacer acquisition with evident strand bias, a phenomenon that is called 'primed adaptation'[42–44]. Although primed adaptation has not been reported for type II systems, it is possible that spacer acquisition is conserved among CRISPR–Cas systems. These observations in conjunction with knowledge that nicking can be induced with mismatched gRNAs could be relevant in the increased ability to acquire spacers in the presence of non-targeting gRNAs. In particular, the nicked product could provide a strand-specific advantage for the integration of nucleic acids in the CRISPR loci. An additional possibility is that the minimal homology between gRNA and DNA target produces a nick that serves as an anchor for the CRISPR–Cas nuclease to recruit Cas1 and Cas2.

Methods

Variant plasmid library. The Cpf1 library was created using pooled oligonucleotide synthesis and was inserted into a plasmid vector population as described Fu et al.[28]. A Cas9 *unc-22A* variant library created using degenerate oligonucleotide synthesis and a library created using pooled oligonucleotide synthesis were previously characterized in Fu et al.[29] and were used in this study. The library created using pooled oligonucleotide synthesis was retransformed and regrown from Fu et al.[29].

High-throughput in vitro target specificity assays. Cpf1 in vitro target specificity assays were performed as reported by Fu et al.[28], in 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 μg ml^{-1} bovine serum albumin, pH 7.9. Cpf1 gRNAs were synthesized by Integrated DNA Technologies. Each reaction contained 50 ng μl^{-1} of Cpf1 protein and gRNA at a 1:1 ratio and was incubated at 37 °C.

For Cas9, we used the modified single (chimeric) gRNA structure detailed in Jinek et al.[8]. Target specificity in vitro assays for Cas9 were performed as detailed in Fu et al.[28] (ThermoPol buffer: 20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.8). In addition to the buffer used for these studies, we performed a set of parallel assays using an alternative buffer similar to that of Jinek et al.[8] (Cas9 buffer: reaction buffer: 20 mM HEPES, 100 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, pH 6.5). Similar results were obtained in the two buffers, albeit (as previously observed by Fu et al.[29]) with differences in cleavage kinetics and completion.

Our initial assessment of nicking versus cleavage relies on the known stimulation of PCR yields from nicked plasmid compared to circular (supercoiled) plasmids, while linearized plasmids decrease skew in PCR yields[27]. As an initial assay to assess target nicking and cleavage after Cas9/Cpf1 interaction, we started with circular plasmid libraries that carry a wide variety of target variants between defined PCR primers. Aliquots of the library were interacted with Cas9/Cpf1 and DNA was extracted, along with control aliquots of the circular (unreacted) library. Each extracted DNA sample was then split into two pools, one of which was amplified and sequenced after digestion outside of the target region with a restriction enzyme that linearizes the population of molecules outside of the amplified region, so there is no longer a preference for nicked circles in the overall PCR yield. The second pool is amplified without linearization, resulting in counts that would retain a preference for nicked over closed circular or double-stranded cleaved molecules. We note that nicking by Cpf1/Cas9 in the amplified (target homology) region resulted in an increase of PCR yields and that both gel and PCR enhancement assays would see nicked molecules with a single cleaved phosphodiester bond in additional to molecules where other processes had extended an initial phosphodiester break in vitro[34] or in vivo. In practice, the observed net stimulations of PCR yields from nicking are on the order of 2^3.

Supplementary Table 3 lists the gRNA and target plasmid sequences.

Calculation of retention scores. For each sequence '*X*': Retention (*X*) = $\log_2$ (representation of *X* in library with addition of CRISPR–Cas nuclease)/(representation of *X* in uncleaved library).

First, we normalized retentions to the total reads from each experimental condition that had the expected length (35–36 base pairs (bp) and barcode (scaling to the total library counts)). As an alternative normalization for comparison, we scaled different libraries by counting the tags matching a non-targeted gRNA sequence or all non-targeted gRNA sequences. For example, experiments with gRNAs targeting EGFP1, EGFP2, and *unc-22A* were normalized to the sum of all variants of the *rol-6* target sequence or normalized to the sum of all non-targeted variants. For experiments where *rol-6* was the intended target, the library was normalized to the total sum of *unc-22A* variants. As expected, the different normalization approaches yielded highly comparable results. Normalizations used for display are noted for each figure.

Cases of positive retention (as noted earlier and in reference to increased yield for nicked circles in PCR) are indicative of potential target nicking rather than cleavage. We note that a maximum positive retention in our experiments of approximately 3 $\log_2$ units, yielded a comparable estimate. A set of assays using $Cas9^{D10A}$ nickase also allows us a more quantitative conversion between degree of enhancement in PCR yield and quantitative fraction of nicked templates (100% nicking yields a retention ~2.5–3, again in the same range; Supplementary Figs. 52–54).

Nicking agarose gel assays. Target DNA was cloned into a parent vector (pHRL-TK; Promega) using NotI (New England Biolabs) and Acc65I (New England Biolabs) restriction sites. Cpf1 or Cas9 reactions were set up as stated earlier for the library assays. For time points, samples were stopped with 100 mM EDTA, 2% SDS, and 80 U μl^{-1} Proteinase K and/or flash-frozen with dry ice. For visualization, frozen samples were thawed and immediately loaded on ~1–1.5% agarose gel (Tris-acetate-EDTA buffer, ethidium bromide concentration (0.3 mg l^{-1})). Supplementary Table 3 lists the gRNA and target plasmid sequences.

Protein components. As/Fn/LbCpf1 constructs were generated by assembly of synthetic gene fragments into an *Escherichia coli* expression plasmid using the NEBuilder HiFi DNA Assembly Cloning Kit (catalogue no. E5520S; New England Biolabs). The As/Fn/LbCpf1 expression vectors contained N-terminal 6×His-tag and SV40 NLS, and C-terminal SV40 NLS. Recombinant proteins were expressed in modified *E. coli* NiCo21 (DE3) cells (catalogue no. C2925H; New England Biolabs) harbouring the Cpf1 expression plasmid by growing in lysogeny broth medium at 23 °C for 16 h in the presence of isopropyl β-D-1-thiogalactopyranoside at 0.4 mM. Cells were disrupted by sonication before chromatographic purification.

As/Fn/LbCpf1 was purified using HiTrap DEAE Sepharose FF (GE Healthcare Life Sciences), HisTrap HP (Ni-NTA resin; GE Healthcare Life Sciences), and HiTrap SP HP (GE Healthcare Life Sciences) columns. Recombinant proteins were dialysed and concentrated into 20 mM Tris-HCl (pH 7.4), 500 mM NaCl, 1 mM dithiothreitol, 0.1 mM EDTA, and 50% glycerol.

The Cas9 and LbCpf1 protein preparations were sourced from New England Biolabs (catalogue nos. M0386S and M0653S, respectively).

Cas9 in vivo assays. Two yeast strains were transfected with a pooled library cloned into a plasmid carrying a Cas9 expression cassette (driven by the GalL promoter), a gRNA expression construct (driven by a tetracycline-inducible RNase P ribonucleoprotein promoter), and a target or non-target sequence adjacent to a potential 'PAM' site ('NGG') (plasmid map available from https://benchling.com/s/O5VobNjd). Target sequences were inserted into the plasmid at a defined location, with the majority deriving from a library of variants of the *unc-22A*, while variants of a sequence of unrelated origin (PS4) provide a number of internal controls[28]. The two yeast populations analysed were from strains BY4741 (BY) and ΔKU70 deletion strain (KU) from *MATa* collection. BY is an auxotrophic WT S288C laboratory strain, while KU was chosen for analysis based on its intrinsic defects in non-homologous double-strand break repair. We found no major differences between the two strains. Each strain was analysed without specific induction of Cas9 ('baseline levels'). Additionally, each strain was analysed after one division and at 2.5 generations as assessed by optical density (these were ~90 and ~230 min after an initial 4-h metabolic adaptation to galactose media from dextrose) following induction with galactose and 250 ng ml^{-1} anhydrotetracycline. Control, single-generation, and 2.5 generation time points in yeast were analysed in duplicate for ΔKU70, with control and single-generation time points analysed in duplicate for BY4741.

In our analysis (Fig. 5 and Supplementary Table 1), we used the difference in log retention score for any given target (circular-linear) as the provisional metric for evaluation, validating this measurement as described. Retention scores for these comparisons were calculated from individual DNA pools that had been subdivided following extraction. One aliquot was analysed following an experimental template linearization cleavage step, which we call a 'linear' assay (Method 1), while a second aliquot was analysed without such a step (designated a 'circular' assay or Method 2). As already noted, previous biochemical analysis[27] and our further analysis in vitro confirm that this assessment indeed provides a metric that assesses DNA nicking.

Statistics and reproducibility. Several features of experimental design with complex libraries are key to consistent assessment of nicking activities. First, we ensured that each potential target sequence (WT or mutant) was present in several barcoded versions in each library, allowing independent measurements for each sequence based on the individual barcodes (the number of such sequences used for the analysis are indicated in each figure with a value *n*). Second, each library contains multiple internal reference standards—plasmids with a parallel structure but with target sequences unrelated to the gRNA being assayed. For the Cas9 libraries described in this study, the unrelated reference sequences were from several distinct unrelated targets. As shown on the left of each line plot figure, the relative representation of these unrelated targets was constant following interaction with a specific gRNA, while representation of targeted sequences was, as noted, consistently modulated. Additional internal control support comes from a subset of gRNA-homologous targets with central PAM variants, which also show little or no change relative to reference sequences (shown for Cpf1 targets as positions −2 and −3). Third, we carried out assessments of retention at multiple time points and with more than one target for each enzyme, with comparable results. Cpf1 assays on each target were carried out with several different cognate enzymes and time points with similar results. Cas9 in vitro assays were each performed in several independent time series with comparable results, and Cas9 in vivo assays were performed in two different yeast strains with comparable results.

For nicking assessments using electrophoresis (Fig. 4a–e), the gRNA::target combination was tested at multiple time points, with multiple experimental trials. For Supplementary Figs. 28–33 and 48–51, a single series with multiple time points was carried out. In each case, an expected migration shift on nicking was observed.

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

Data availability

The raw data that support our findings are available in Sequence Read Archive under accession no. PRJNA503740 (see Supplementary Table 2 for information on the data for each corresponding figure).

Received: 28 September 2018; Accepted: 21 January 2019;
Published online: 04 March 2019

References

1. Terns, M. P. & Terns, R. M. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14**, 321–327 (2011).
2. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
3. Barrangou, R. & Marraffini, L. A. CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* **54**, 234–244 (2014).
4. Heler, R., Marraffini, L. A. & Bikard, D. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol. Microbiol.* **93**, 1–9 (2014).
5. Carroll, D. Genome editing: past, present, and future. *Yale J. Biol. Med.* **90**, 653–659 (2017).
6. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–963 (2013).
7. Terns, R. M. & Terns, M. P. CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends Genet.* **30**, 111–118 (2014).
8. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
9. Murovec, J., Pirc, Ž. & Yang, B. New variants of CRISPR RNA-guided genome editing enzymes. *Plant Biotechnol. J.* **15**, 917–926 (2017).
10. Cebrian-Serrano, A. & Davies, B. CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm. Genome* **28**, 247–261 (2017).
11. Eid, A., Alshareef, S. & Mahfouz, M. M. CRISPR base editors: genome editing without double-stranded breaks. *Biochem. J.* **475**, 1955–1964 (2018).
12. Shen, B. et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat. Methods* **11**, 399–402 (2014).
13. Davis, L. & Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl Acad. Sci. USA* **111**, E924–E932 (2014).
14. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110–113 (2015).
15. 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).
16. Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J. & Russell, R. Kinetic basis for DNA target specificity of CRISPR-Cas12a. *Mol. Cell* **71**, 816–824 (2018).
17. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
18. Bayat, H., Modarressi, M. H. & Rahimpour, A. The conspicuity of CRISPR-Cpf1 system as a significant breakthrough in genome editing. *Curr. Microbiol.* **75**, 107–115 (2018).
19. Fernandes, H., Pastor, M. & Bochtler, M. Type II and type V CRISPR effector nucleases from a structural biologist's perspective. *Postepy Biochem.* **62**, 315–326 (2016).
20. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).
21. Kim, H. K. et al. In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat. Methods* **14**, 153–159 (2017).
22. Vriend, L. E. M. & Krawczyk, P. M. Nick-initiated homologous recombination: protecting the genome, one strand at a time. *DNA Repair (Amst.)* **50**, 1–13 (2017).
23. Vriend, L. E. M. et al. Distinct genetic control of homologous recombination repair of Cas9-induced double-strand breaks, nicks and paired nicks. *Nucleic Acids Res.* **44**, 5204–5217 (2016).
24. Kuzminov, A. Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc. Natl Acad. Sci. USA* **98**, 8241–8246 (2001).
25. Gao, Y. et al. Single Cas9 nickase induced generation of NRAMP1 knockin cattle with reduced off-target effects. *Genome Biol.* **18**, 13 (2017).

26. Satomura, A. et al. Precise genome-wide base editing by the CRISPR Nickase system in yeast. *Sci. Rep.* **7**, 2095 (2017).
27. Lin, C.-H., Chen, Y.-C. & Pan, T.-M. Quantification bias caused by plasmid DNA conformation in quantitative real-time PCR assay. *PLoS ONE* **6**, e29101 (2011).
28. Fu, B. X. H., Hansen, L. L., Artiles, K. L., Nonet, M. L. & Fire, A. Z. Landscape of target:guide homology effects on Cas9-mediated cleavage. *Nucleic Acids Res.* **42**, 13778–13787 (2014).
29. Fu, B. X. H., St Onge, R. P., Fire, A. Z. & Smith, J. D. Distinct patterns of Cas9 mismatch tolerance in vitro and in vivo. *Nucleic Acids Res.* **44**, 5365–5377 (2016).
30. 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).
31. Ran, F. A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380–1389 (2013).
32. Dong, D. et al. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).
33. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
34. Chen, J. S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity.*Science* **360**, 436–439 (2018).
35. Koo, T., Lee, J. & Kim, J.-S. Measuring and reducing off-target activities of programmable nucleases including CRISPR-Cas9. *Mol. Cells* **38**, 475–481 (2015).
36. Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol. Ther. Nucleic Acids* **4**, e264 (2015).
37. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **45**, 273–297 (2011).
38. Sternberg, S. H., Richter, H., Charpentier, E. & Qimron, U. Adaptation in CRISPR-Cas systems. *Mol. Cell* **61**, 797–808 (2016).
39. Jackson, S. A. et al. CRISPR-Cas: adapting to change. *Science* **356**, eaal5056 (2017).
40. Wei, Y., Terns, R. M. & Terns, M. P. Cas9 function and host genome sampling in type II-A CRISPR-Cas adaptation. *Genes Dev.* **29**, 356–361 (2015).
41. Heler, R. et al. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202 (2015).
42. Fineran, P. C. et al. Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl Acad. Sci. USA* **111**, E1629–E1638 (2014).
43. Staals, R. H. J. et al. Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* **7**, 12853 (2016).
44. Richter, C. et al. Priming in the type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.* **42**, 8516–8526 (2014).

Acknowledgements

We thank the D. Herschlag, P. Fineran, G. Hess, N. Jain, and colleagues in our laboratories for their input and discussion. We are grateful to J. A. Meacham for advice on the reagents used and C. Lee for lending reagents.

Author contributions

B.X.H.F conceived and designed the study, and carried out the experiments. A.Z.F. and G.B.R. participated in the experimental design and analysis of data. J.D.S. participated in the experimental design, execution, and discussion of the in vivo yeast experiments. R.T.F., M.M., and J.C. participated in enzyme purification and early investigations of nuclease activities. All authors participated in the contextualization of the results and in the preparation of the manuscript.

Competing interests

R.T.F., M.M., J.C., and G.B.R. are employees of New England Biolabs. A provisional patent 'Compositions and Methods for Nicking Target DNA Sequences' related to this work has been filed by Stanford University (inventors: B.X.H.F., A.F., and J.D.S.).

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0382-0.

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

Correspondence and requests for materials should be addressed to B.X.H.F., G.B.R. or A.Z.F.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

nature research

Corresponding author(s): Becky Xu Hua Fu, Andrew Fire, Brett Robb

Last updated by author(s): Dec 9, 2018

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	The data processing and normalization is noted in manuscript. Data is uploaded on SRA (PRJNA503740).
Data analysis	No commercial software was used.

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

Data can be found on SRA (PRJNA503740).

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	As bulk biochemical assays are involved, each experiment reports the behavior of a large number of molecules in ensemble, with molecule numbers calculable from concentrations provided. Individual variants are represented in most cases by multiple barcoded species in each library, with numbers of such species noted in the bar graphs. Each species in turn is reported only in cases where a minimum number of reads (50) is observed for the tag. Minimum reference read counts were chosen to avoid substantial stochastic variation in signal.
Data exclusions	N/A
Replication	To ensure reproducibility of the high throughput assays, parallel assessments for each primary input library were done at least twice and amplified independently, with consistent results in each case. Additional verification comes from the parallel use of two library types (RVL and TypeII), which gave equivalent results (for the in vitro assessments using Cas9). Validating the high throughput assays, top hits from were further verified with native agarose gel migration assays.
Randomization	Randomization was not part of the experimental design for this study.
Blinding	Blinding was not part of the experimental design for this study.

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