

Engineered PAM-flexible FnCas9 variants for robust and specific genome editing and diagnostics

Sundaram Acharya

Institute of Genomics and Integrative Biology

Asgar Ansari

Institute of Genomics and Integrative Biology <https://orcid.org/0000-0002-1172-9521>

Seiichi Hirano

University of Tokyo

Deepanjan Paul

Institute of Genomics and Integrative Biology

Riya Rauthan

Institute of Genomics and Integrative Biology

Manoj Kumar

Institute of Genomics and Integrative Biology

Rhythm Phutela

Institute of Genomics and Integrative Biology

Sajal Sarkar

Institute of Genomics and Integrative Biology

Sneha Gulati

Institute of Genomics and Integrative Biology

Sudipta Mahato

LV Prasad Eye Institute

Savitri Maddileti

LV Prasad Eye Institute

Vinay Pulimamidi

LV Prasad Eye Institute

Indumathi Mariappan

LV Prasad Eye Institute

Hiroshi Nishimasu

University of Tokyo

Osamu Nureki

University of Tokyo <https://orcid.org/0000-0003-1813-7008>

Souvik Maiti

Institute of Genomics and Integrative Biology

Debojyoti Chakraborty (✉ debojyoti.chakraborty@igib.in)

Brief Communication

Keywords: Cas9 from *Francisella novicida* (FnCas9), enhanced variants (enFnCas9), genome editing and diagnostics

Posted Date: November 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-990232/v2>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Engineered PAM-flexible FnCas9 variants with enhanced activity for robust and specific genome editing and diagnostics

Sundaram Acharya^{1,2}, Asgar Hussain Ansari^{1,2}, Seiichi Hirano³, Deepanjan Paul¹, Riya Rauthan^{1,2}, Manoj Kumar^{1,2}, Rhythm Phutela^{1,2}, Sajal Sarkar¹, Sneha Gulati¹, Sudipta Mahato^{4,5}, Savitri Maddileti⁴, Vinay Kumar Pulimamidi^{4,6}, Indumathi Mariappan⁴, Hiroshi Nishimasu⁷, Osamu Nureki³, Souvik Maiti^{1,2}, Debojyoti Chakraborty^{1,2}

¹CSIR-Institute of Genomics & Integrative Biology, Mathura Road, New Delhi- 110025, India

²Academy of Scientific & Innovative Research (AcSIR), Ghaziabad, 201002, India

³Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

⁴Center for Ocular Regeneration, Prof. Brien Holden Eye Research Centre, Hyderabad Eye Research Foundation, LV Prasad Eye Institute, Hyderabad-500034, Telangana, India

⁵Manipal Academy of Higher Education, Manipal University, India

⁶School of Life Sciences, University of Hyderabad, India

⁷Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

*Correspondence: debojyoti.chakraborty@igib.in (D.C.)

Abstract

The clinical success of CRISPR therapies is dependent on the safety and efficacy of Cas proteins. The Cas9 from *Francisella novicida* (FnCas9) has negligible affinity for mismatched substrates enabling it to discriminate off-targets in DNA with very high precision even at the level of binding. However, its cellular targeting efficiency is low, limiting its use in therapeutic applications. Here, we rationally engineer the protein to develop engineered(enFnCas9) variants with enhanced activity and expand its cellular editing activity to genomic loci previously inaccessible. Notably, some of the variants release the protospacer adjacent motif (PAM) constraint from NGG to NGR/NRG making them rank just below SpCas9-RY and SpCas9-NG in their accessibility across

human genomic sites. The enFnCas9 proteins, similar to Cas12a and Cas12f, harbor high intrinsic specificity and can diagnose single nucleotide variants accurately. Importantly, they provide superior outcomes in terms of editing efficiency, knock-in rates, and off-target specificity over other engineered high-fidelity versions of SpCas9 (SpCas9-HF1 and eSpCas9). Broad targeting range coupled with remarkable specificity of DNA interrogation underscores the utility of these variants for safe and efficient therapeutic gene correction across multiple cell lines and target loci.

Main

Like the orthogonal *Streptococcus pyogenes* Cas9 (SpCas9) protein, FnCas9 too interacts with the minimal NGG protospacer adjacent motif (PAM) yet shows a much higher sgRNA sequence-dependent specificity when interrogated with DNA substrates(1–4). Although high-fidelity versions of SpCas9 have been designed and validated in multiple systems, their editing efficiencies have generally dropped significantly as compared to the wild-type enzyme(5),(6). To circumvent these issues, in recent years, alternate high-efficiency Cas systems from other microorganisms have been demonstrated for genome editing(7–11). Notably, very few show editing efficiencies either similar or higher than SpCas9, and the majority of these enzymes have a PAM requirement that is more complex and less available in the human genome than SpCas9, limiting the number of possible sites accessible for therapeutic correction(12–16) (Supplementary Table 1).

In earlier studies we and others had reported that FnCas9 has a very high intrinsic specificity, resulting in dissociation from off-targets presented *in vitro*(4, 17). In contrast, SpCas9 and its high-fidelity variants remain bound to off-target sites in a cleavage incompetent fashion(18, 19). Thus, although off-target cleavage has been reduced in these variants, they are still able to negotiate these loci at the level of binding, a feature that might cause non-specific off-targeting outcomes from such regions(20). We speculated that if the high specificity of off-target discrimination at the level of DNA binding is also retained *in vivo*, FnCas9 might present a more specific editing scope, particularly relevant in therapeutic applications. To compare their genome-wide binding propensities, we constructed Hemagglutinin (HA)-tagged catalytically inactive (dead, d) dSpCas9 and dFnCas9 and targeted a locus (*c-Myc*)

where comparable editing efficiencies between SpCas9 and FnCas9 were observed previously(4). Using chromatin immunoprecipitation followed by massively parallel sequencing, we found that both SpCas9 and FnCas9 were tightly bound to the on-target sites explaining their high editing rates (Supplementary Figure 1A). Similar to earlier reports for other loci(21–23), dSpCas9 showed promiscuous binding at multiple off-targets (27 sites, 0.01 FDR) across the genome, even at sites with up to 6 mismatches in the sgRNA. Interestingly, all the 27 sites showed greater enrichment than the on-target (Supplementary Table 2). In contrast, dFnCas9 showed only 6 off-target bound sites (0.01 FDR) with the on-target having 1.2-fold more enrichment over the next off-target site (Supplementary Figure 1B, Supplementary Table 2). This high binding specificity *in vivo* thus presented an attractive scenario for structure-guided engineering to enhance the activity of the FnCas9 enzyme at sites where editing was minimal.

FnCas9 is evolutionarily divergent to SpCas9 and harbors some structural dissimilarities such as unique interactions between the RuvC and REC3 domains, and the PI and WED domains with the latter sharing contacts with the REC1 and REC2 domains(1, 3). However, PAM recognition is conserved among Cas9 orthologs which trigger directional target DNA unwinding, R-loop formation and expansion. This eventually leads to reorientation of the HNH endonuclease domain to DNA cutting and concomitant RuvC activation leading to concerted DNA cleavage(24). Recent mechanistic studies showed that the directional PAM-duplex DNA unwinding serves as the rate-limiting checkpoint of Cas9 action and a conformational switch discriminates Cas9 DNA binding and cleavage events(18, 25–29). Moreover, the loss of nucleobase-specific interaction between the target DNA and Cas9 was reported to be rescued by base non-specific Cas9 interactions(3, 30). Thus, we reasoned that stabilizing FnCas9:DNA duplex binding by introducing base non-specific interactions between PAM duplex and the protein might improve FnCas9 nuclease activity without compromising its intrinsic specificity (Supplementary Note 1).

We engineered 49 different FnCas9 variants guided by its crystal structure bearing mostly single amino acid substitutions in the WED-PI domain to introduce novel PAM duplex DNA contacts (Figure 1A, Supplementary Table 3). We then measured *in vitro* DNA cleavage activities of the FnCas9 variants against a DNA target containing GGG PAM (where FnCas9 was shown to be least active)(3) and performed target DNA cleavage experiments with the FnCas9 variants. Recent reports suggested that high-

fidelity SpCas9 variants result in lower cellular editing efficiency than SpCas9, a property attributed to their slower enzyme kinetics and in particular their ability to sense PAM distal mismatches(5, 6). FnCas9 being an enzyme with high fidelity, we focussed on engineered FnCas9 (enFnCas9) variants with faster cleavage activity on a DNA substrate and selected a subset of nine enFnCas9 variants (containing single/combinatorial mutations) ranked based on the number of amino acid substitutions and its position on the protein (Figure 1B, Supplementary Note 1, and Supplementary Figure 2A). All these variants showed faster cleavage rates than FnCas9 with three variants (en1, en15 and en31) having the highest activity (Figure 1B).

Two independent groups had earlier reported that SpCas9 and its engineered altered-PAM variants with multiple amino acid substitutions create additional phosphate backbone interactions which synergistically induce unexpected displacement in the DNA backbone of the PAM duplex and facilitate these variants to recognize the non-canonical PAMs(31, 32). To test if enFnCas9 variants show a similar relaxation in PAM recognition, we selected a subset of five enFnCas9 variants based on their enhanced activity at the non-canonical NGA PAM containing DNA substrates (Supplementary Figure 2B-C). Next, we performed an *in vitro* PAM discovery assay where a pooled library with a target DNA sequence (protospacer region) adjacent to a randomized 8 bp sequence ($4^8 = 65,536$ combinations in total) was cleaved with respective FnCas9 or enFnCas9-sgRNA complexes (Supplementary Figure 3). This was followed by deep sequencing of the PAM depleted library to comprehensively determine the PAM specificities of the variants. We observed that the enFnCas9 variants showed more flexible recognition in second and third nucleotide positions as compared to FnCas9 (Supplementary Figure 4). Importantly, for all the variants tested, NGG was relaxed to NGR/NRG thereby expanding the scope of enFnCas9 accessibility across the human genome (Supplementary Figure 4). With this broadened PAM accessibility, enFnCas9 variants rank just below SpCas9-RY(33) and SpCas9-NG(30) in their ability to target sites on the human genome when compared to other CRISPR systems reported to date (Figure 1D, Supplementary Table 1).

Figure 1. Engineering and characterization of enFnCas9 variants for enhanced activity and altered PAM activity

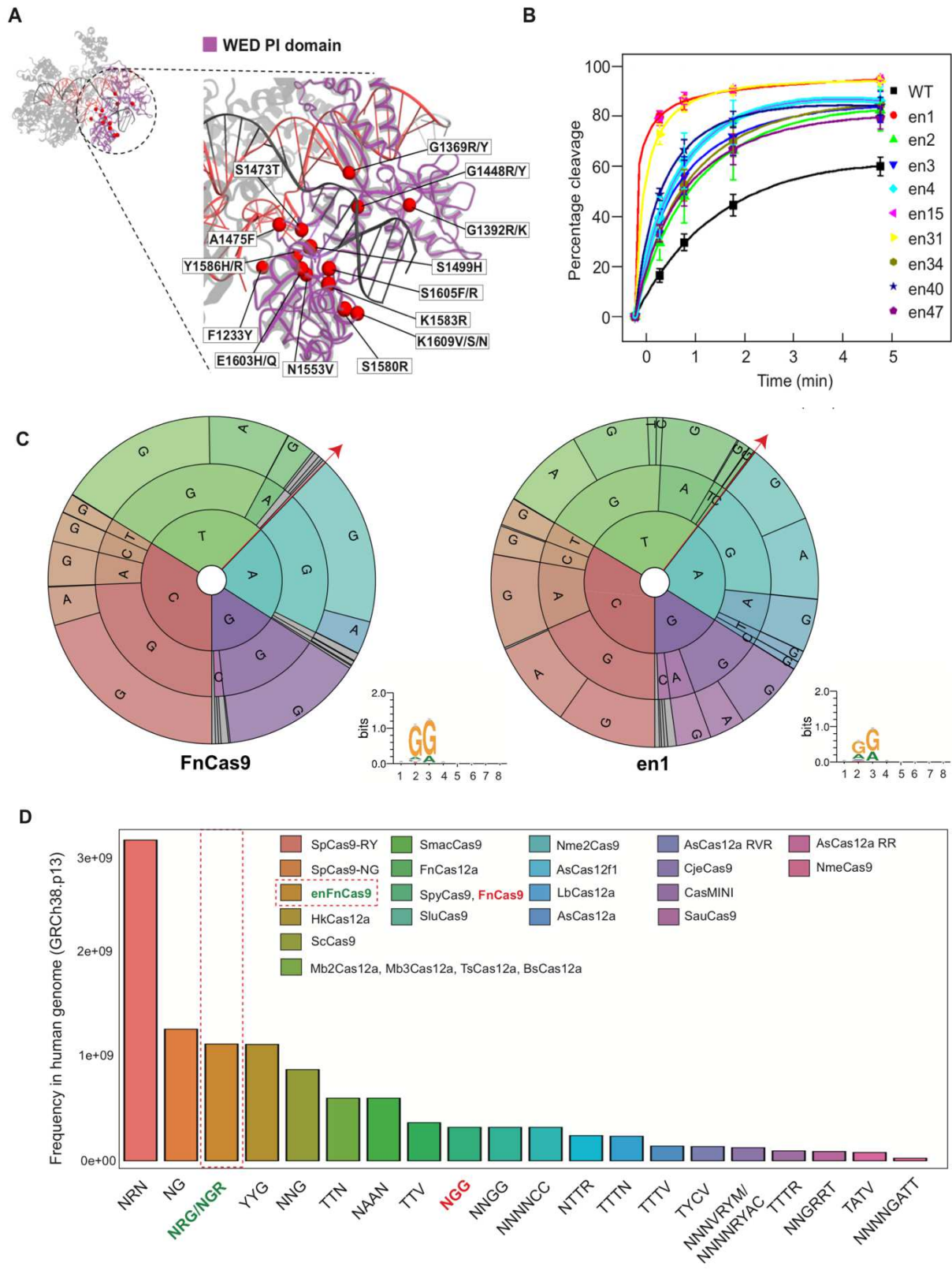


Figure 1.

A) FnCas9 crystal structure in complex with sgRNA-DNA (PDB: 5B2O) in ribbon model with highlighted WED-PI domain marked in dotted circle. Zoomed inset shows amino acid residues substituted for protein engineering.

B) *In vitro* cleavage assay of FnCas9 and a subset of nine enFnCas9 variants on GGG PAM containing PCR linearized DNA substrate expressed as percentage cleavage (y-axis) as a function of time (x-axis). Error bars represent SD (three independent experiments).

C) The PAM wheels and sequence logos showing the results obtained after PAM discovery assay for FnCas9 and en1. Individual sections of the pie charts in the PAM wheels with $\leq 2\%$ depletion enrichment are shown in gray. Bases from the inner to the outer circle in the PAM wheels map the PAM reads away from the target region in the 5' to 3' direction as shown by red arrows.

D) Bar plot showing the availability of PAMs of respective CRISPR systems in the human genome expressed as frequency in human genome on y-axis and PAM sequence on x-axis. Respective NGG and NRG/NGR PAMs of FnCas9 and enFnCas9 are highlighted in red and green accordingly. Red dotted box highlights PAM preference for a subset of enFn variants.

The remarkable intrinsic specificity of FnCas9 to single-nucleotide mismatches in the target has applications both in disease diagnostics and disease correction⁴. At the level of diagnostics, FnCas9 has recently been utilized for paper strip-based robust diagnostics of nucleic acid targets through the FnCas9 Editor Linked Uniform Detection Assay (FELUDA) and Rapid Variant Assay (RAY) platforms(34, 35). These platforms utilize a direct FnCas9:DNA binding-based readout as opposed to collateral cleavage of DNA or RNA reporters employed by Type V effectors (such as Cas12a(36) or Cas12f(37)) or Type VI effectors (such as Cas13(38)) respectively. We had earlier shown that placing an additional second mismatch in defined positions in the sgRNA causes the enzyme to dissociate from its target and can be used as a readout for discriminating point mismatches such as disease-causing single nucleotide variants (SNVs) or Variants of Concern (VOCs) in SARS-CoV-2 infections(34, 35). At a visual level, this can be used for diagnosing such mutants on a lateral flow device using FAM-labeled sgRNA (Figure 2A)(34, 35). We predicted that due to their broadened PAM accessibility, enFnCas9 variants (NRG/NGR) can now cover 82.2 % of the reported Mendelian SNVs across the human genome (compared to 40.46 % by FnCas9) thereby increasing the scope of detection to more disease-causing variants (Figure 2B). Expectedly, on a lateral flow strip, all enFnCas9 variants tested showed robust activity on a substrate carrying the non-canonical NGA PAM whereas FnCas9 did not show any signal (Supplementary Figure 5A).

In recent studies, several type V DNA targeting Cas systems (such as Cas12a or Cas12f) have been characterized and engineered for genome editing(9, 10, 14–16, 39, 40). These Cas effectors have a naturally occurring high intrinsic specificity to mismatches(41, 42)(43)(15) and induce non-specific single-stranded DNA (ssDNA) trans-cleavage upon target activation(36, 37, 44). Indiscriminate cleavage of ssDNA reporters has been utilized as a read-out for detecting point mismatches in targets. To compare the inherent specificities across these naturally occurring Cas proteins, we purified FnCas9(34), AaCas12b(44), and Cas14a1(37) and compared their single mismatch specificities using sgRNA design principles reported for each of the CRISPR proteins with their respective reporter system (trans-cleavage for Cas12/14 and affinity for FnCas9). All the three Cas effectors were able to discriminate SNVs from the WT sequence with a signal resolution suitable for diagnosis (4.4-fold for FnCas9, 4.6-fold for AaCas12b, and 5.1-fold for Cas14a1) suggesting that they are all useful for *in vitro*

Figure 2. enFnCas9 variants show high sensitivity for SNV detection on a lateral flow device

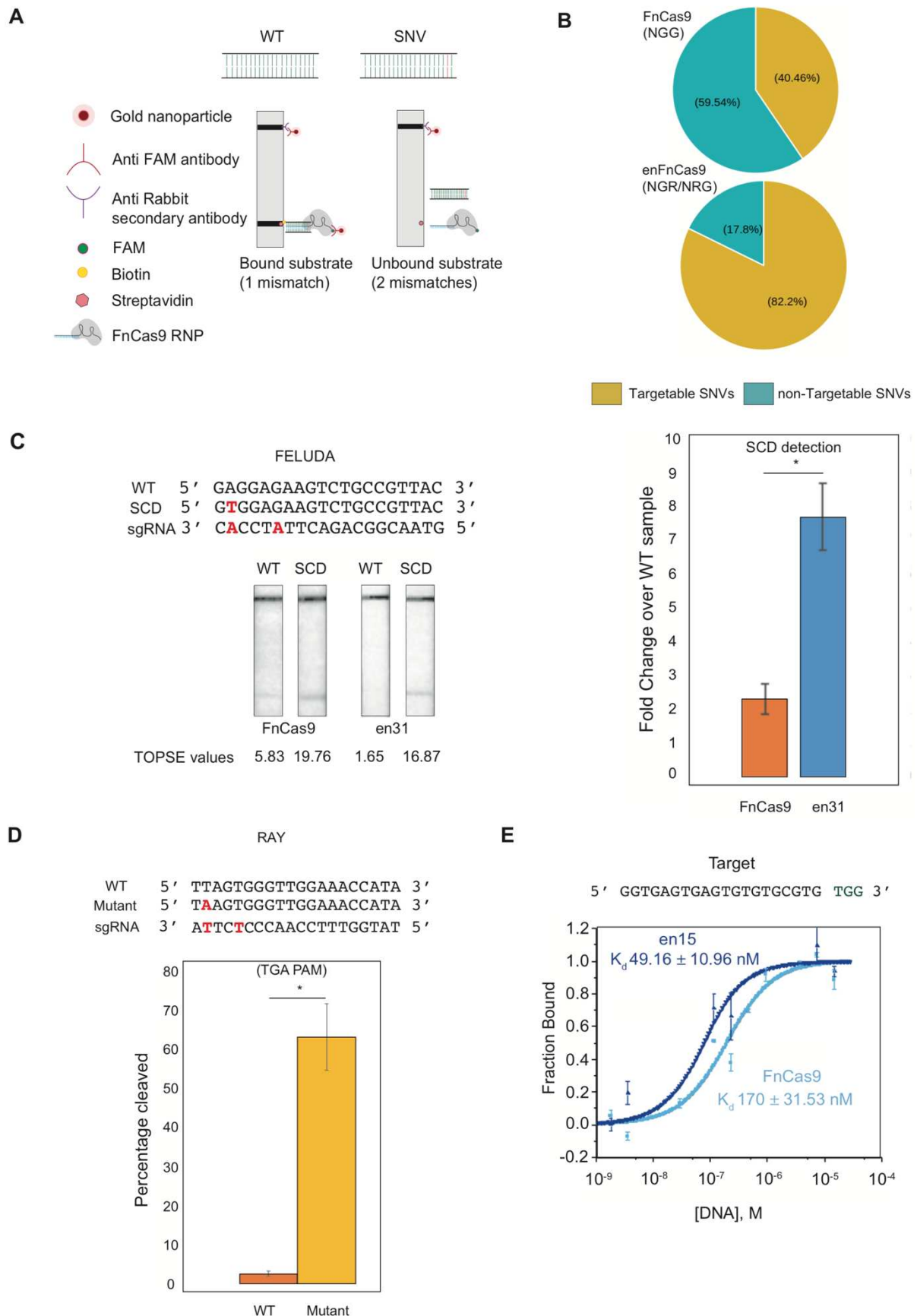


Figure 2.

A) Schematic representation showing the mode of SNV detection by FELUDA and RAY CRISPRDx platforms.

B) Pie chart showing the percentage of targetable and non-targetable SNVs by FnCas9 and enFnCas9 variants.

C) (Right) Outcome of lateral flow assay (LFA) for SCD detection by FELUDA using FnCas9 and en31. WT and SCD target sequences are shown. The sickle cell mutation and FELUDA specific sgRNAs with mismatch positions are represented in red. Corresponding TOPSE values are given at the bottom. (Left) Bar plot showing SCD mutation discrimination by en31 and FnCas9 using FELUDA. Data is shown as fold enrichment over WT sample (n=3 independent experiments). Student's unpaired *t-test* p-value is represented for * <0.05.

D) Bar plot showing outcome of *in vitro* cleavage based detection of N501Y (with an NGA PAM) by en31(n=3 independent experiments). Student's unpaired *t-test* p-value is represented for * <0.05.

E) MicroScale Thermophoresis (MST) result showing the comparative binding affinity between FnCas9 and en15 on *VEGFA3* substrate DNA. Data is represented as a fraction bound RNP (y-axis) with respect to purified DNA substrate (Molar units M, x-axis). Error bars represent SD (3 independent experiments).

discrimination of mismatched substrates and have very high intrinsic specificity (Supplementary Figure 5B).

Since enFnCas9 variants were constructed by altering residues that stabilize the PAM duplex binding keeping the DNA interacting domains (responsible for PAM distal mismatch sensitivity) untouched, we speculated that they should still retain the high specificity as WT FnCas9. Indeed, upon performing a mismatch walking assay along the full sequence of the sgRNA, the three highest activity enFnCas9 variants en1, en15, and en31 all showed grossly similar specificity for mismatch tolerance as FnCas9 (Supplementary Figure 5C). For all the enzymes, tolerance to mismatches was lowest at the most PAM proximal (1st and 2nd) and distal (15th-19th) bases. However, unlike FnCas9, the stringency for mismatch tolerance for all the variants was lower towards the middle of the sgRNA (PAM distal 9-11 bases). This can be attributed to faster cleavage rates of enFnCas9 variants since even for FnCas9, longer incubation times can lead to substrate cleavage with mismatches in these positions(4). To determine if these changes in enFn variants might affect their diagnostic potential, we selected the enFnCas9 variant with the broadest activity at altered PAM sites (en31) and investigated if it was able to distinguish single mismatches in two targets with pathogenic mutations related to Sickle Cell Anemia and the SARS-CoV-2 Alpha VOC signature (N501Y). Remarkably en31 accurately distinguished both the target SNVs on a lateral flow device (Figure 2C, Supplementary Figure 5D) with an improved signal discrimination (>3.5-fold) as compared to FnCas9 (Supplementary Figure 5E). We confirmed that the same specificity of SNV discrimination was also extended for an NGA PAM-containing substrate as well (Figure 2D). Taken together, enFnCas9 variants have a very high specificity of mismatch discrimination similar to Cas12a or Cas12f but due to their wider PAM accessibility, these can potentially target more genomic sites and pathogenic SNVs for detection (Figure 1D).

We next investigated if engineering FnCas9 by altering residues that interact with PAM in the substrate had altered its binding affinity to DNA. To test this, we constructed recombinant Green Fluorescent Protein (GFP)-tagged catalytically inactive versions of two of the variants (en1 and en15) and performed microscale thermophoresis (MST) to determine their DNA binding affinities on a substrate (*VEGFA*) as reported earlier(4). We found that these variants showed stronger DNA binding ($K_d = 91.33 \pm 29.8$ nM for en1, $K_d = 49.16 \pm 10.96$ nM for en15) as compared to FnCas9 ($K_d = 170 \pm 31.53$ nM), with en15 showing ~3.5-fold higher DNA binding affinity (Figure 2E and

Supplementary Figure 5E). Interestingly, in our previous study(4), we had seen that FnCas9 showed weaker binding to the same substrate as SpCas9 (3.02-fold). Thus, engineering improved enFnCas9:DNA binding affinity, reaching similar levels as SpCas9 but with superior specificity.

The safety of therapeutic genome editing is guided by off-target interrogation of CRISPR effectors. Although Cas12a and Cas12f have higher specificity than SpCas9, their therapeutic success relies on minimum ssDNA cleavage inside the cell such as those formed during replication, homology-directed repair, or transcription(36, 45). Interestingly, Cas12a has been recently reported to nick off-target DNA substrates with up to four mismatches depending upon the crRNA sequences employed(46). On the contrary, enFnCas9 does not produce trans-cleavage products, and its high specificity both at the level of DNA interrogation and cleavage might be beneficial for safe nuclease-mediated genome editing. Although SpCas9 has shown robust gene editing capabilities across different genomic loci, the intrinsic non-specific nature of the protein has warranted the development of high-fidelity versions for potential therapeutic editing(47, 48). Interestingly, high-fidelity SpCas9 proteins generally show lower editing efficiencies as compared to the wild-type protein⁵. We selected two such proteins (SpCas9-HF1 and eSpCas9) due to their balanced activity and specificity as reported in literature(5)(49)(50) and compared their cellular editing rates (insertion/deletions) with enFnCas9 variants en1 and en15 in HEK293T cells. Among the two, enFn1 showed higher editing rates as compared to high-fidelity SpCas9 proteins at all four loci investigated and no detectable off-targets at any of the off-target sites identified either through GUIDE-Seq or *in silico* prediction(4, 47) (Figure 3A). Expectedly, at all four loci, en1 had equal or higher editing efficiencies than the wildtype FnCas9 protein (Figure 3A). These results showed that en1 achieves higher genome editing efficiency than high-fidelity SpCas9 variants but retains similarly high on-target specificity. Similarly, we confirmed that both en1 and en15 variants showed successful genome editing in other human cell lines as well such as the induced pluripotent stem cells (iPSCs) and retinal pigmented epithelial cells (ARPE-19) (Figure 3B, C). Notably, in the iPSC line, en1 (18.6% indels) and en15 (23.0% indels) showed superior editing rates at the *PAX6* locus when compared to even SpCas9 (13.8%) (Figure 3B). Taken together, these results established that enFnCas9 variants perform genome editing with superior efficiency and specificity than high-fidelity SpCas9 variants.

Figure 3. Superior cellular genome editing outcomes by enFnCas9 variants

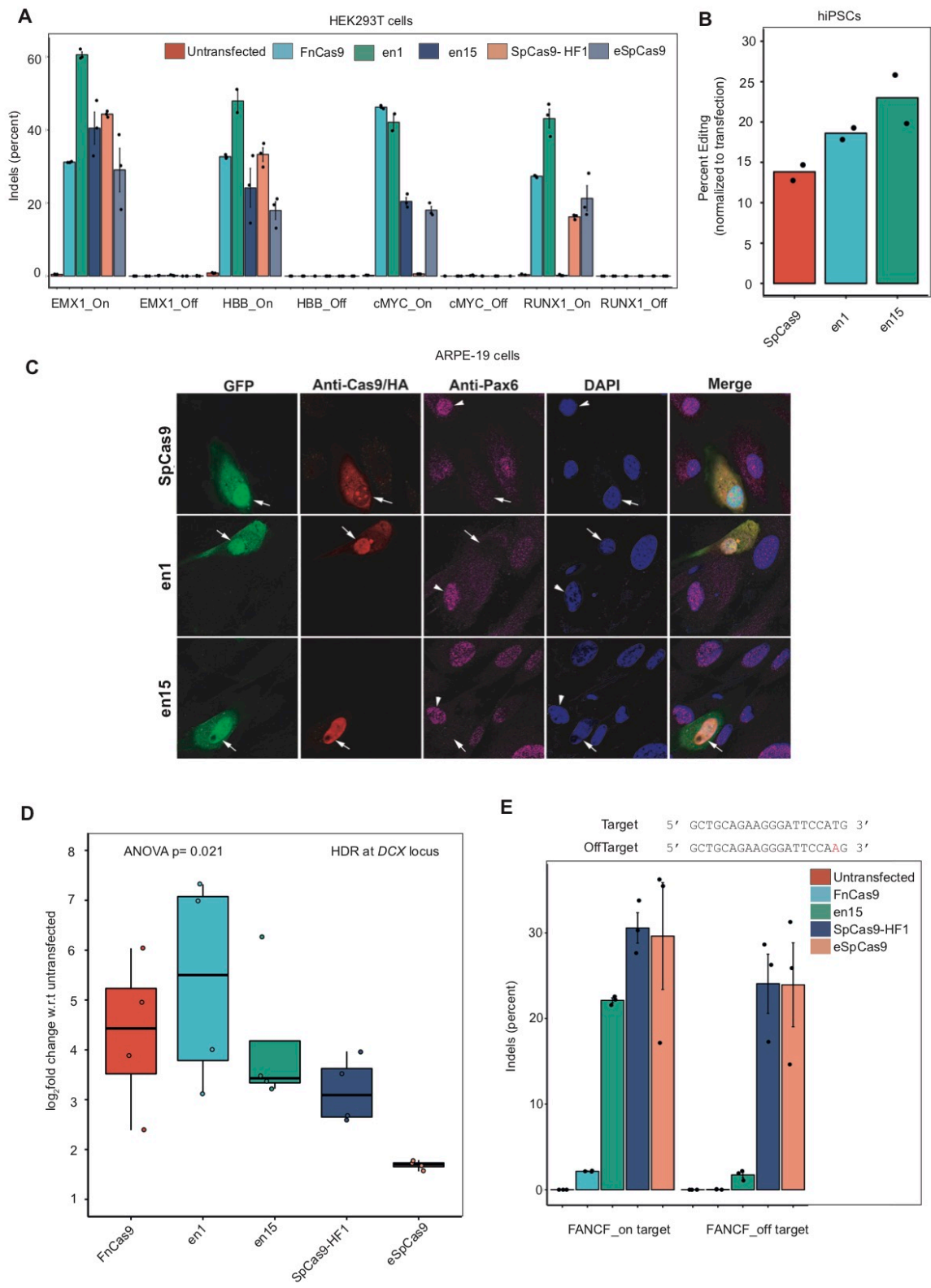


Figure 3

A) Indel events (expressed in percentage and shown as black dots) as obtained from amplicon sequencing upon targeting *EMX1*, *HBB*, *c-MYC*, *RUNX1* by FnCas9, en1, en15, SpCas9-HF1 and eSpCas9 in HEK293T cells. Untransfected cells shown as control. Error bars represent SEM (3 independent experiments except for *HBB* and *c-MYC* ontarget of en1).

B) Bar plot showing indel percentage normalized to transfection efficiency as obtained from T7E1 assay upon targeting *PAX6* locus by SpCas9, en1 and en15 in hiPSC cells (2 independent experiments).

C) Immunofluorescence imaging showing fixed cells stained with anti-HA/anti-Cas9 (red), anti-PAX6 (pink) and DAPI (blue) upon targeting *PAX6* locus by SpCas9, en1 and en15 in ARPE-19 cells. The GFP panel is showing Cas9 transfected cells. Merged panel depicts all the cells present in the field. Cas9 expressing cells are marked by arrows and non-expressing cells are marked by arrowheads.

D) Box plot showing knock-in of a dsDNA template at *DCX* locus by FnCas9, en1, en15, SpCas9-HF1 and eSpCas9 in HEK293T cells. Data is represented as log₂ fold change with respect to untransfected samples and analyzed using one-way Anova, p-value is shown (4 independent experiments).

E) Indel events (expressed in percentage) as obtained from amplicon sequencing upon targeting *FANCF* locus in HEK293T cells. Untransfected cells serve as control. Error bars represent SEM (3 independent experiments).

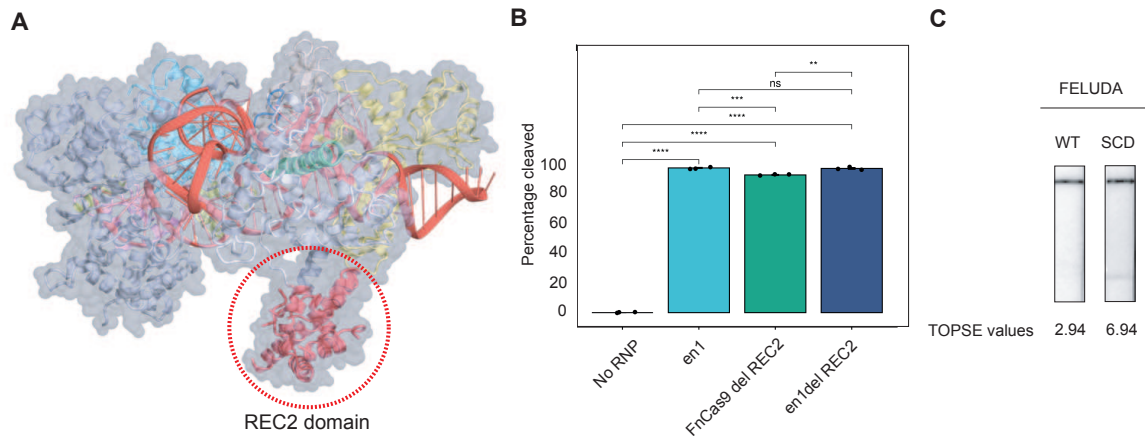
Finally, we investigated if the higher editing efficiency of the enFnCas9 variants can also lead to greater homology-directed repair (HDR) when presented with a double-stranded DNA (dsDNA) template. Here too, we observed higher HDR mediated knock-in of a long donor template (4.1 kb) at the *DCX* locus in HEK293T cells for both en1 and en15 as compared to SpCas9-HF1 and eSpCas9 (Figure 3D). Collectively enFn1 showed a higher rate of gene editing at all the target loci tested both for insertions/deletions as well HDR mediated knock-in highlighting its suitability as a highly potent genome-editing protein.

In the present study, we have shown the remarkable efficacy and specificity of enFnCas9 variants in targeted genome editing and diagnostics. Two aspects of these variants would require further investigation. Firstly, the genome-wide editing specificity of enFnCas9 variants has not been explored and is a subject of ongoing experiments. Secondly, enFnCas9 variants are larger (~190 kDa) than SpCas9 (~159 kDa) and other smaller-sized Cas9 variants, thus limiting their delivery only through larger or combinatorial packaging vehicles such as Adenoviruses (AVs), mRNA, or split Adeno associated viruses (AAVs).

Interestingly, the specificity of these variants appears to stem from the DNA interaction properties of FnCas9 independent of the engineered residues in the enzyme. Thus, we observed that even after substantially improving DNA binding affinity and activity, en15 showed minimal editing at a GUIDE-Seq validated off-target with a single mismatch at the PAM proximal end (Figure 3E). This is in sharp contrast to both eSpCas9 and SpCas9-HF1 which showed editing efficiencies comparable to the target site(51, 52) (Figure 3E). These results show that enFnCas9 variants possibly negotiate off-targets through a different mechanism than high-fidelity SpCas9 proteins. Similarly, due to its inherent modularity, FnCas9 has options for further engineering to reduce its size. To this end, we constructed a truncated enFn1 by deleting its REC2 domain (which does not show any tertiary contacts with the rest of the protein) and reduced the size of en1 to ~170 kDa, closer to that of SpCas9 (~159 kDa) (Figure 4A). Remarkably, en1- Δ REC2 can retain both its activity and specificity to similar levels as en1 (Figure 4B,C). This is in sharp contrast to SpCas9- Δ REC2 where the substantial loss of activity was seen upon deletion(1).

Our results indicate that engineering residues that regulate PAM duplex contacts in the Cas9 backbone can significantly improve editing efficiency without affecting

Figure 4. REC2 truncation retains activity and specificity of enFnCas9 variants



A) Crystal structure of FnCas9 in complex with DNA and RNA (PDB: 52BO) is shown in ribbon model with space fill overlay. Truncated REC2 domain (Δ S112-A297) is highlighted in red and marked by a red-dotted circle.

B) Bar plot showing in vitro cleavage efficiency of en1, FnCas9 Δ REC2 and en1 Δ REC2 on GGG PAM containing DNA substrate. The 5nM DNA substrate was incubated with 100nM RNPs for 1hr at 37°C. Error bars represent SD (3 independent experiments). Student's unpaired t-test p-values are represented for ** <0.01, *** <0.001.

C) Outcome of lateral flow assay (LFA) for SCD detection by FELUDA using en1 Δ REC2. Corresponding TOPSE values are given at the bottom.

specificity. This strategy can be potentially extended for other orthogonal Cas systems that possess higher intrinsic specificity but have low cellular activity. The enFn variants hold a lot of promise for safe and efficient nuclease mediated genome editing and also present potentially attractive avenues for double-strand break-free editing (such as base and prime editors) where the extent of off-target interrogation and concomitant nucleobase editing has not been understood to the fullest.

Materials and Methods

Plasmid construction

Point mutations and deletions were done by inverse PCR method on FnCas9 cloned in pE-SUMO vector backbone (LifeSensors) where intended changes were made on the forward primer and the entire plasmid was amplified by inverse PCR. Point mutations on the pET-His6-dFnCas9GFP and PX458-3xHA-FnCas9 (Addgene 130969) were done by essentially following the method described earlier(4). gRNAs were cloned in the BbsI sites of PX458-3xHA-FnCas9, PX458-3xHA-en1FnCas9, PX458-3xHA-en15FnCas9, PX458-3xHA-SpCas9HF1 and eSpCas9(1.1) (Addgene 71814) for cellular genome editing assays by essentially following the method described earlier(53). All of the constructs were sequenced before being used.

Protein Purification and sgRNA purification

The proteins used in this study were purified as reported previously(4, 30) Briefly, plasmids for Cas9 from *Francisella novicida* were expressed in Escherichia coli Rosetta2 (DE3) (Novagen). The protein-expressing Rosetta2 (DE3) cells were cultured at 37°C in LB medium (supplemented with 50mg/ml kanamycin) until OD₆₀₀ reached 0.6 and protein expression was induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG). The Rosetta2 (DE3) cells were further cultured at 18°C overnight and harvested by centrifugation. The *E.coli* cells were resuspended in buffer A (20 mM Tris-HCl, pH 8.0, 20 mM imidazole, and 1 M NaCl), and lysed by sonication, and centrifuged. The lysate was mixed with Ni-NTA beads (Roche), the mixture was loaded into a Poly-Prep Column (BioRad) and the protein was eluted by buffer B (20 mM Tris-HCl, pH 8.0, 0.3 M imidazole, and 0.3 M NaCl). The affinity eluted protein was mixed with ion-exchange beads (SP Sepharose Fast Flow, GE Healthcare) equilibrated with buffer C (20 mM Tris-HCl, pH 8.0, and 0.15 M NaCl) and the protein was eluted by buffer D (20 mM Tris-HCl, pH 8.0, and 1 M NaCl). AaCas12b and Cas14a1 were purified essentially by following the purification methods described earlier with some modifications(37, 44). The concentration of purified proteins was measured by the Pierce BCA protein assay kit (Thermo Fisher Scientific). The purified proteins were stored at -80 °C until further use. The proteins used in the study are depicted in Supplementary Figure 5F.

In vitro transcribed sgRNAs were synthesized using MegaScript T7 Transcription kit (Thermo Fisher Scientific) using T7 promoter containing template as substrates. IVT reactions were incubated overnight at 37°C followed by NucAway spin column (Thermo Fisher Scientific) purification as described earlier(4). IVT sgRNAs were stored at -20°C until further use.

***In vitro* cleavage (IVC) assay**

For the DNA cleavage study, PCR linearized pUC119 plasmid containing the target sequence and the respective PAM (mentioned in respective legends) was used as the substrate for *in vitro* cleavage experiments. The linearized pUC119 plasmid (50 ng or ~5nM) was incubated at 37°C for 0.5–5 min with the Cas9–sgRNA complex (50 nM) in 10 µL of reaction buffer, containing 20 mM HEPES, pH 7.5, 150 mM KCl, 10 mM MgCl₂, 1 mM DTT, and 5% glycerol. The reaction was stopped by the addition of a quenching buffer, containing EDTA (20 mM final) and Proteinase K (40 ng). The reaction products were resolved, visualized, and quantified with a MultiNA microchip electrophoresis device (SHIMADZU)(30).

The rest of the IVC assays were done as described earlier(4). Details of substrates, concentrations, and incubation time are mentioned in respective figure legends.

PAM discovery assay

The PAM discovery assays were performed, as previously described(30). Briefly, a library of pUC119 plasmids containing eight randomized nucleotides downstream of the target sequence was incubated at 37°C for 5 min with the FnCas9–sgRNA complex (50 nM), in 50 µL of the reaction buffer. The reactions were quenched by the addition of Proteinase K and then purified using a Wizard DNA Clean-Up System (Promega). The purified DNA samples were amplified for 25 cycles, using primers containing common adapter sequences. After column purification, each PCR product (~5 ng) was subjected to the second round of PCR for 15 cycles, to add custom Illumina TruSeq adapters and sample indices. The sequencing libraries were quantified by qPCR (KAPA Biosystems) and then subjected to paired-end sequencing on a MiSeq sequencer (Illumina) with 20% PhiX spike-in (Illumina). The sequencing reads were demultiplexed by primer sequences and sample indices, using NCBI Blast+ (version 2.8.1) with the blastn-short option. For each sequencing sample, the

number of reads for every possible 8-nt PAM sequence pattern ($4^8 = 65,536$ patterns in total) was counted and normalized by the total number of reads in each sample. For a given PAM sequence, the enrichment score was calculated as \log_2 -fold enrichment as compared to the untreated sample. PAM sequences with enrichment scores of ≥ 2.0 were used to generate the PAM wheel using KronaTools (v2.7) (<https://hpc.nih.gov/apps/kronatools.html>) and the sequence logo representation using WebLogo 3 (<http://weblogo.threeplusone.com/create.cgi>).

en/FnCas9 based SNP detection:

(i) in vitro cleavage (IVC) assay

The RNA substrates were reverse transcribed into cDNA (Qiagen), followed by PCR amplification or the DNA substrates were only PCR amplified (Invitrogen) and further purified. The substrates were treated with a pre-assembled 500 nM en/FnCas9-sgRNA (1:1) RNP complex in a tube containing reaction buffer (20 mM HEPES, pH7.5, 150mM KCl, 1 mM DTT, 10% glycerol, 10 mM $MgCl_2$) at 37°C for 10 min. The reaction was inactivated by using 1 μ l of Proteinase K (Ambion) at 55°C for 10 min, followed by the removal of residual gRNA by RNase A (Purelink) at 37°C for 10 min. The cleaved products were visualized on a 2% agarose gel and quantified.

(ii) via lateral flow assay

5' biotin-labeled amplicons were treated with reconstituted en/FnCas9 RNP complex (prepared by equimolar mixing 3' FAM labelled-Chimeric gRNA and en/FnCas9 in a buffer containing 20 mM HEPES, pH7.5, 150 mM KCl, 1 mM DTT, 10% glycerol, 10 mM $MgCl_2$ and rested for 10 min at RT) for 10 min at 37°C. Wherever active en/FnCas9 was used, $MgCl_2$ was omitted from the buffer for making it catalytically inactive. After incubation, an 80 μ l Dipstick buffer was added to the reaction tube followed by the addition of one Milenia HybriDetect lateral flow strip and kept for 2–5 min at RT to observe test and control bands. Further background-corrected band intensity values were calculated through a smartphone application (TOPSE) (35-36).

Fluorescence assay (dFnCas9)

250nM biotin labelled PCR amplicons carrying 580bp long SARS-CoV-2 region with N501Y mutation were used for attaching DNA substrate to the wells of streptavidin coated plate by 10 mins incubation at room temperature. Wells were rinsed thrice with the wash buffer to get rid of the unbound amplicons (25mM Tris-Cl, pH 7.2; 300mM NaCl; 0.1% BSA, 0.05% Tween®-20 Detergent) before using for the binding assay. dFnCas9-GFP RNP complex was pre-assembled in the binding buffer (20 mM HEPES, pH7.5, 150mM KCl, 1mM DTT, 10mM MgCl₂) by incubating 200nM dFnCas9-GFP with 200nM sgRNA for 10 min at room temperature. Reaction was initiated by adding pre-assembled RNP to the wells of 96-well streptavidin coated plate (Thermo Fisher Scientific; Cat 15119) pre-attached with biotin labelled amplicons and incubated at 37°C for 10 mins. Fluorescence was measured using a fluorescence plate reader (λ_{ex} : 485 nm; λ_{em} : 528 nm, transmission gain: optimal) (Tecan Infinite Pro F200).

Fluorescence assay (AaCas12b and Cas14a1)

AaCas12b and Cas14a1 RNP complexes were pre-assembled by incubating 200nM AaCas12b and Cas14a1 with 200 nM respective sgRNA for 10 min at room temperature. Reaction was initiated by adding pre-assembled RNP, 20nM ssDNA activator, 100ng background genomic DNA and 200 nM custom synthesized homopolymer ssDNA FQ reporter as described earlier(37, 44) (GenScript) in a cleavage buffer (40mM Tris-HCl, pH 7.5, 60mM NaCl, 6mM MgCl₂). The reaction was incubated in a 96-well flat bottom clear, black polystyrene microplate (Corning, cat no.CLS3603) at 37°C up to 180 mins with fluorescent measurements taken every 10 min (λ_{ex} : 485 nm; λ_{em} : 528 nm, transmission gain: optimal) using fluorescence plate reader (Tecan Infinite Pro F200). The resulting data were background-subtracted using the readings taken in the absence of ssDNA activator.

DNA Binding assay

MST was performed as described previously(4). Briefly, dFnCas9-GFP and variant proteins were complexed with PAGE purified respective IVT sgRNAs (purified by 12% Urea-PAGE). The binding affinities of the Cas9 proteins and sgRNA RNP complexes were calculated using Monolith NT. 115 (NanoTemper Technologies GmbH, Munich, Germany). RNP complex (Protein:sgRNA molar ratio,1:1) was reconstituted at 25 for

10 mins in reaction buffer (20 mM HEPES, pH7.5, 150mM KCl, 1mM DTT, 10mM MgCl₂) HPLC purified 30 bp dsDNA (IDT) of different genomic loci with varying concentrations (ranging from 0.09 nM to 30 μM) were incubated with RNP complex at 37° C temperature for 30 min in reaction buffer. The sample was loaded into NanoTemper standard treated capillaries and measurements were performed at 25°C using 20% LED power and 40% MST power. Data analyses were done using NanoTemper analysis software and the data were plotted by OriginLab.

Cell culture

HEK293T cells were grown in DMEM media supplemented with high glucose (Invitrogen), 2 mM GlutaMax, 10% FBS (Invitrogen), 1X antibiotic and antimycotic (Invitrogen) at 37°C in 5% CO₂. Human iPS cells (LVP-F2-3F) were derived and maintained as described earlier(54). Briefly, the cells were cultured using Essential 8™ complete media kit (Gibco, Cat No. A1517001), along with the addition of 1X Penicillin-Streptomycin antibiotics solution (Gibco, Cat No. 15140122) and cultured on Vitronectin coated (Gibco, Cat No. A14700) cell culture plates at 37°C in 5% CO₂. The human RPE cell line, ARPE19 (ATCC, Cat No. CRL2302) was cultured in DMEM/F-12 medium (Gibco, Cat No. 10565018) supplemented with 10% FBS (Gibco, Cat No. 26140079) and 1X Penicillin-Streptomycin antibiotics solution (Gibco, Cat No. 15140122) at 37°C in 5% CO₂.

Transfection of HEK293T and ARPE19 cells were performed using Lipofectamine 3000 Reagent (Invitrogen) following the manufacturer's protocol. Transfections of hiPSCs were performed on the next day of seeding 5X10⁴ cells onto Vitronectin-coated 24-well plates using Lipofectamine™ Stem Transfection Reagent (Invitrogen, Cat No. STEM00003).

T7 endonuclease assay

48 hrs post-transfection, the cells were lysed with 250 μL of extraction buffer (100mM Tris pH 8.0, 1% SDS, 5mM EDTA, 200 μg/mL Proteinase K) and incubated at 56°C for 2 hrs and the genomic DNA was precipitated with the addition of isopropanol. The DNA pellet was washed with 70% ethanol, air-dried, and dissolved in the TE buffer. The human *PAX6* exon6 target region was amplified by PCR using screening primer sets and DreamTaq DNA polymerase (ThermoFisher, Cat No EP0702), as per

manufacturer's protocol. The PCR amplicons were gel-purified (Qiagen, Cat No. 28104) and about 1 µg of DNA was subjected to denaturation at 95°C for 5 mins and renaturation by slow cooling in a dry thermostat. The annealed DNA amplicons with heteroduplexes were incubated with 1 µL of T7 endonuclease 1 (NEB, Cat No. M0302S) and incubated at 37°C for 1 hour. The cleavage products in the reaction mix were separated by 8% agarose gel electrophoresis. Densitometry analysis was done using BioRad Image Lab software. The NHEJ event was calculated using the following formula:

$$\% \text{ NHEJ events} = 100 \times [1 - (1 - \text{fraction cleaved})^{1/2}]$$

where, fraction cleaved = (density of digested product)/(density of digested product + density of undigested product).

The cleavage fraction was normalized for transfection efficiency (% GFP⁺ cells) and the average values were plotted with standard deviations.

ChIP sequencing

HEK293T cells on 10 cm dishes were transfected with 30µg of plasmids. 48 hrs post-transfection GFP-positive cells were FACS sorted (BD FACSMelody Cell Sorter). ChIP was done by essentially following the earlier reported protocol with modifications as per requirements of our experiments(23). Sorted cells were cross-linked with 1% formaldehyde (Sigma) with gentle rotation at room temperature for 15 mins followed by quenching by adding 125mM glycine. Cells were rinsed twice chilled PBS. Cells were centrifuged at 1500g for 10min at 4°C and the cell pellet was snap-frozen in liquid nitrogen before storing at -80°C. The cell pellet was resuspended in pre-chilled 1ml lysis buffer 1 (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1x Roche protease inhibitor cocktail), rotated for 15 min at 4°C and centrifuged at 1500g for 10min at 4°C. The pellet was resuspended in 1ml pre-chilled lysis buffer 2 (10 mM Tris-Cl pH 8, 200 mM NaCl, 1 mM EDTA, 1x Roche protease inhibitor cocktail) and treated similarly as previous. Now, the nuclear pellet was resuspended in 500ul pre-chilled sonication buffer (20 mM Tris-Cl pH 8, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100, 1x Roche protease inhibitor cocktail) and sonicated for 10 min using Covaris S220 focused ultrasonicator (duty factor 20%, duty cycle 5, PIP 140, CPB 200, water temperature 4°C). The lysates were

centrifuged by placing it in DNA LoBind microfuge tubes (Eppendorf) at maximum speed for 15 min at 4°C and the supernatant was collected. 25 ul of lysate was saved as input (5%). Precleared diluted lysates were incubated with 5 ug anti-HA ChIP grade antibody (abcam #9110) overnight at 4°C. The antibody-protein complexes were incubated with 15 ul of protein G magnetic beads (Dynabeads, Life Technologies) for 2 hrs at 4°C. Beads were repeatedly washed using three of the buffers by adding pre-chilled ChIP dilution buffer, high salt buffer and LiCl buffer. Washed beads were next washed two times by TE buffer. The chromatin was recovered from the beads by incubating with the ChIP elution buffer for 15 min at room temperature with rotation. The eluted chromatin was reverse crosslinked, digested with ProteinaseK treatment and contaminating RNA was removed by RNase followed by purification of DNA using ethanol precipitation. Purified DNA was tested for fold enrichment at sgRNA target region before library preparation for massively parallel sequencing. Sequencing libraries were prepared using NEBNext® Ultra™ II DNA Library Prep Kit by essentially following the manufacturer's protocol and sequenced on HiSeq X platform at MedGenome Labs Pvt. Ltd. (Bangalore, India).

Amplicon sequencing

HEK293T cells on six well dishes were transfected with 2µg of respective Cas9 containing sgRNAs. 48 hrs post-transfection GFP-positive cells were FACS sorted (BD FACSMelody Cell Sorter) and gDNA was isolated (Lucigen QuickExtract Extraction solution). PCR primers were designed flanking the predicted double stranded break site and amplified with Phusion High-Fidelity DNA polymerase (Thermo Fisher Scientific). The 16S Metagenomic sequencing library preparation protocol was adapted for library preparation. Briefly, the respective loci was amplified using forward and reverse primers along with overhang adapter sequences using Phusion High-Fidelity DNA polymerase (Thermo Fisher). AMPure XP beads (A63881, Beckman Coulter) were used to separate out amplicons from free primers and primer dimers. Dual indexing was done using Nextera XT V2 index kit followed by a second round of bead-based purification. The libraries were quantified using a Qubit dsDNA HS Assay kit (Invitrogen, Q32853) and were also loaded on agarose gel for the qualitative check. Libraries were normalized, pooled and were loaded onto the Illumina MiniSeq platform.

HDR assay at DCX locus in HEK293T

HEK293T cells were cultured in DMEM with GlutaMAX supplement (ThermoFisher Scientific Cat. No. 10566016) with 10% FBS serum. 70%-80% confluent HEK293T cells were harvested from a 6 well plate using Trypsin-EDTA (0.05%) (ThermoFisher Scientific Cat. No.: 25300062) and pipetted to make a single-cell suspension. For each electroporation reaction, a total 15ug plasmid was mixed in Resuspension buffer R, in which linearized donor plasmid DNA and Cas9-gRNA vector were taken in a 1:2 ratio. 6×10^5 cells were resuspended in 100 μ L of Resuspension Buffer R containing plasmids and electroporation was performed using Neon Transfection System 100 μ L Kit (ThermoFisher Scientific Cat. No. MPK10096) with double pulses at 950 V, 30 milliseconds pulse width. The electroporated cells were transferred immediately to a 6 well plate containing 2 ml of pre-warmed culture medium and incubated at 37°C and 5% CO₂. After 24 hours cells were washed and re-incubated with a fresh culture medium. 72 hours post electroporation GFP positive cells per sample were sorted using BD FACSMelody Cell Sorter (BD Biosciences-US) and gDNA was isolated from the sorted cells using Wizard Genomic DNA Purification Kit (Promega) for qPCR genotyping.

qPCR reactions were performed using LightCycler 480 SYBR Green I Master (Roche) added to 50ng DNA for each sample. The cycling conditions on the instrument were as follows: Initial denaturation 95°C for 5 min followed by 40 amplification cycles of 95°C for 10 sec; 63°C for 30 sec; 72°C for 30. Log₂ fold change values were calculated by the $2^{-\Delta\Delta C_t}$ method for each sample with respect to untransfected control. A non-targeting region in genomic DNA was used for normalization.

ChIP Seq analysis

Raw sequencing reads were mapped to the human reference genome GRCh38 using bowtie2(55). Peaks were called over input samples using MACS2(56) with default parameters. Finally, scrambled sample peaks were used to remove background and false positive peaks from the dSpCas9 and dFnCas9 test samples. These filtered peaks were searched for off-targets based on sgRNA sequence homology with a maximum of 6 mismatches. On-target peak coverage plots were generated by the fluff profiles command with 'remove duplicates' option(57). Overlap between the dSpCas9

and dFnCas9 ChIP peaks were calculated using bedtools(58) and plotted as weighted Venn diagrams with the help of Intervene(59).

PAM frequency analysis

PAM frequencies were calculated for more than 167 Cas systems (146 unique PAM sequences) from the human reference genome (GRCh38.p13) using in-house python script.

Amplicon sequencing analysis

Sequencing reads from different replicates were down-sampled prior to indel analysis for each target to remove sequencing read depth bias across the samples. Raw amplicon sequencing reads were subjected to indel frequency estimation using CRISPResso2 v2.0.45(60) with parameters such as ignoring substitutions and keeping minimum overlap between the forward and reverse read to the 10bp.

Immunostaining and confocal imaging

The ARPE-19 cells on glass coverslips were washed with phosphate-buffered saline (1X PBS) 48 hrs after transfection, fixed with 3.5% formaldehyde in 1X PBS for 10 minutes, followed by three washes. The cells were then permeabilized with 0.5% Triton X-100 in 1X PBS for 10 mins, followed by three washes, and then blocked with 10% FBS in 1X PBS for 1 hour. The cells were then sequentially incubated with anti-PAX6 (Abcam, ab195045, 1:200) and anti-HA (CST, Cat No. 3724S, 1:800) primary antibodies diluted in blocking buffer for 1 hr. The cells were then washed three times with 1X PBS and incubated with species-specific secondary antibodies conjugated to different fluorophores for 45 minutes. The cells were then washed, counterstained with DAPI, and mounted on a glass slide using the Vectashield mountant. The samples were then imaged and analyzed using the Zeiss LSM 880 confocal laser scanning microscope and Zeiss Zen software and the images were assembled into collages using Adobe Photoshop.

Primer Sequences

All primer sequences used in the study are listed in Supplementary Table 4.

Data Availability

Deep sequencing data from ChIP and amplicon sequencing experiments were deposited as a BioProject under Project ID PRJNA766155.

Contributions

S.A., S.M. and D.C. conceived the project and designed the experimental pipeline. A.H.A designed and implemented bioinformatics tools for design, validation and execution of enFnCas9 based diagnostics and editing experiments. S.A. designed and performed the protein engineering experiments with inputs from S.H., H.N. and O.N. D.P. designed target and off-target guides for in cellulo experiments with enFn variants. R.R. performed HDR experiments with enFn variants. M.K., R.P. and S.G. contributed to validating the FELUDA/RAY platform with enFn proteins. S.S. performed comparison of single mismatch specificity with Cas12/14 proteins. S.A. performed sequencing and protein truncation experiments. S.A., Su. M, Sa. M, V.K.M and I.M performed and validated enFn activity experiments in HEK293T, ARPE-19 and iPSC lines. S.A. and D.C. drafted the manuscript with inputs from all other authors.

Acknowledgments

We thank all members of Chakraborty, Maiti, Nureki and Nishimasu labs for helpful discussions and valuable insights pertaining to this work. This study was funded by CSIR Sickle Cell Anemia Mission (HCP0008 and HCP0023), and a Lady Tata Young Investigator award (GAP0198) to D.C.

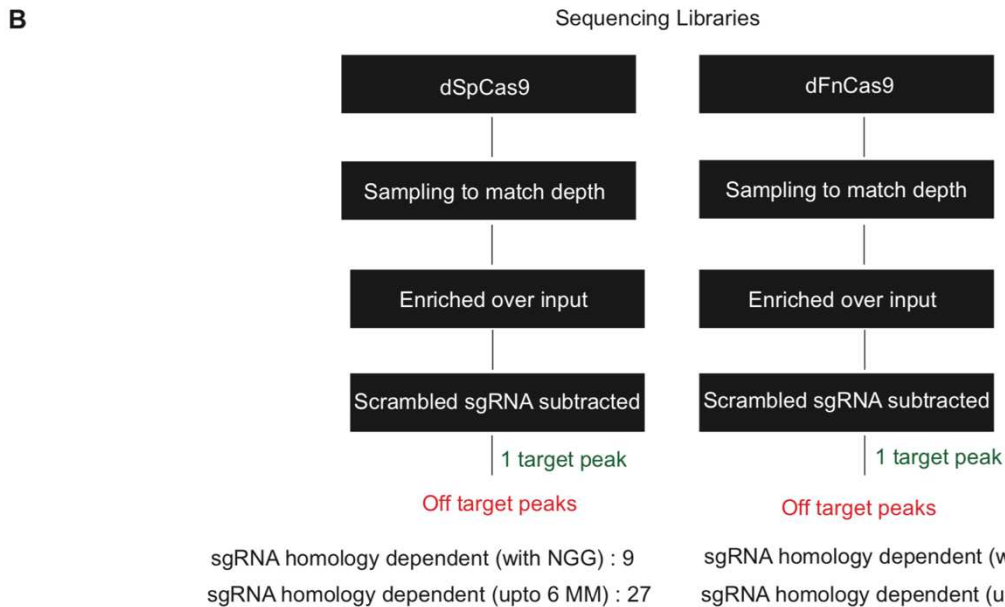
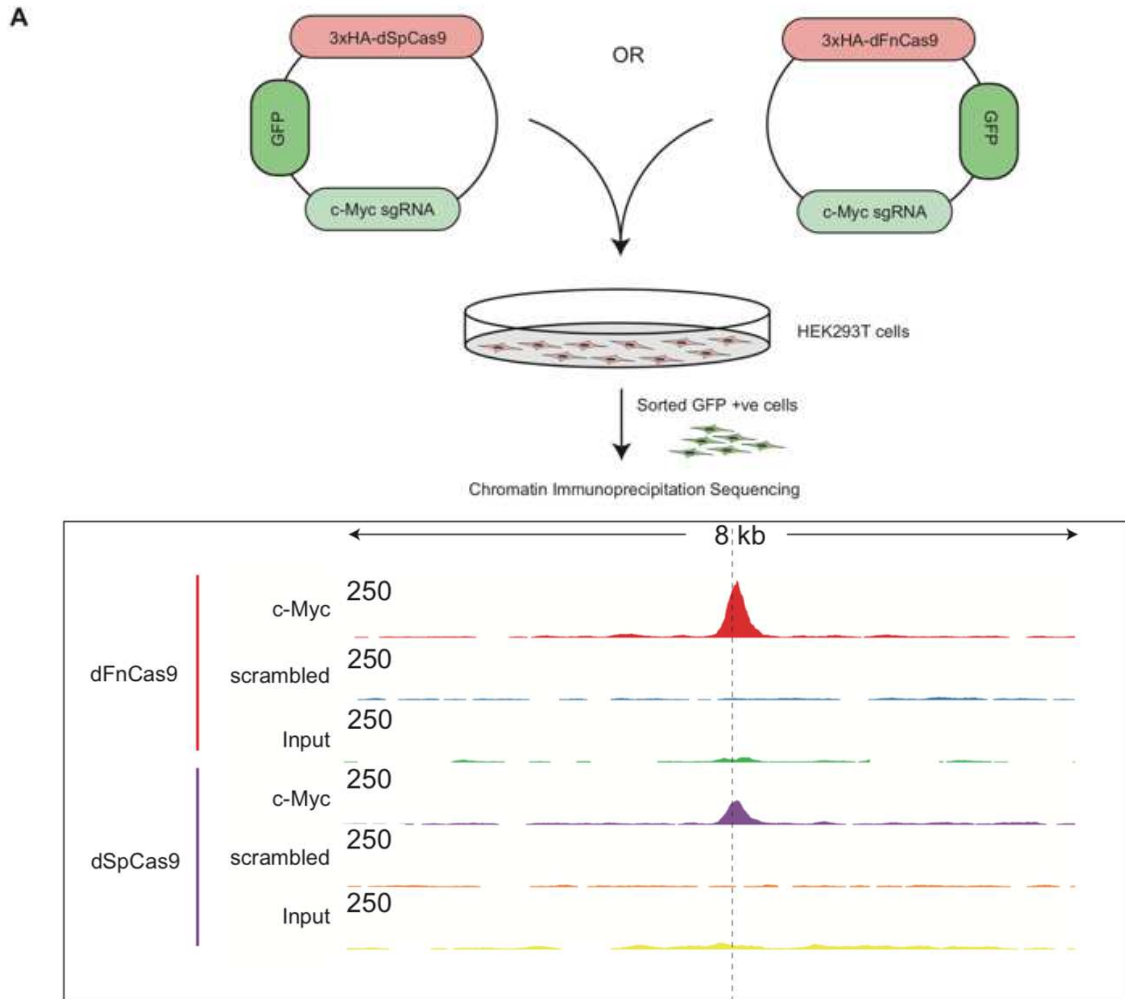
Ethics Declarations

Competing Interests

S.A., S.H., H.N., S.M., O.N., and D.C. are co-inventors on Indian provisional patent application 0023NF2021, relating to the Cas proteins described in this manuscript. The other authors declare no competing interests.

Supplementary Figures

Supplementary Figure 1. Mapping genome-wide binding sites of dSpCas9 and dFnCas9 in *c-MYC* locus in HEK293T cells

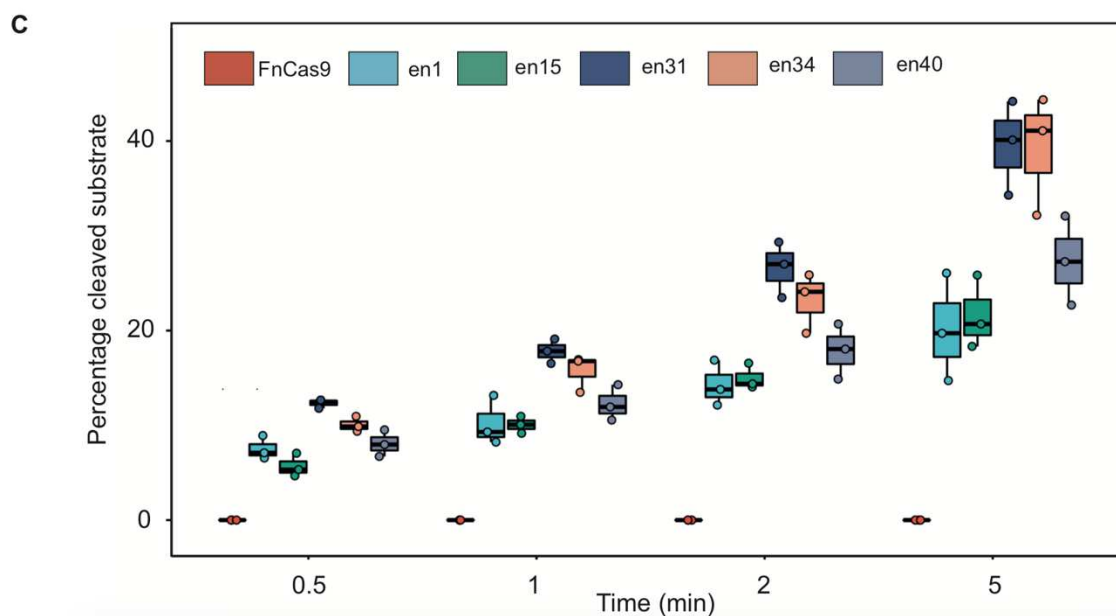
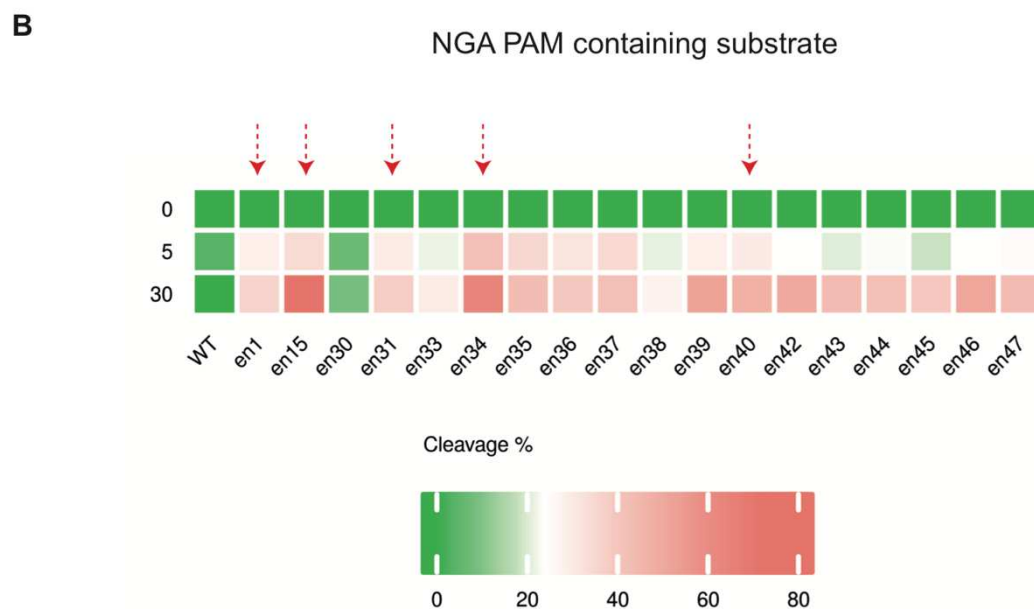
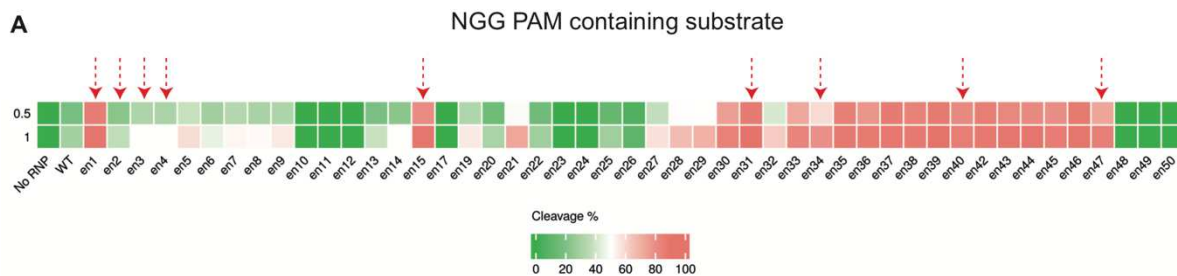


Supplementary Figure 1.

A) Schematic representation showing ChIP-Seq for targeting *c-MYC* locus by 3xHA-dSpCas9 and 3xHA-dFnCas9. ChIP-Seq signal intensity of dSpCas9 and dFnCas9 with respect to input sample and scrambled sgRNA transfection at the intended on-target region is shown.

B) Peak calling pipeline showing the off-target sites of 3xHA-dSpCas9 and 3xHA-dFnCas9 ChIP-Seq on *c-MYC* locus.

Supplementary Figure 2. *In vitro* cleavage assay for testing cleavage activity of enFnCas9 variants



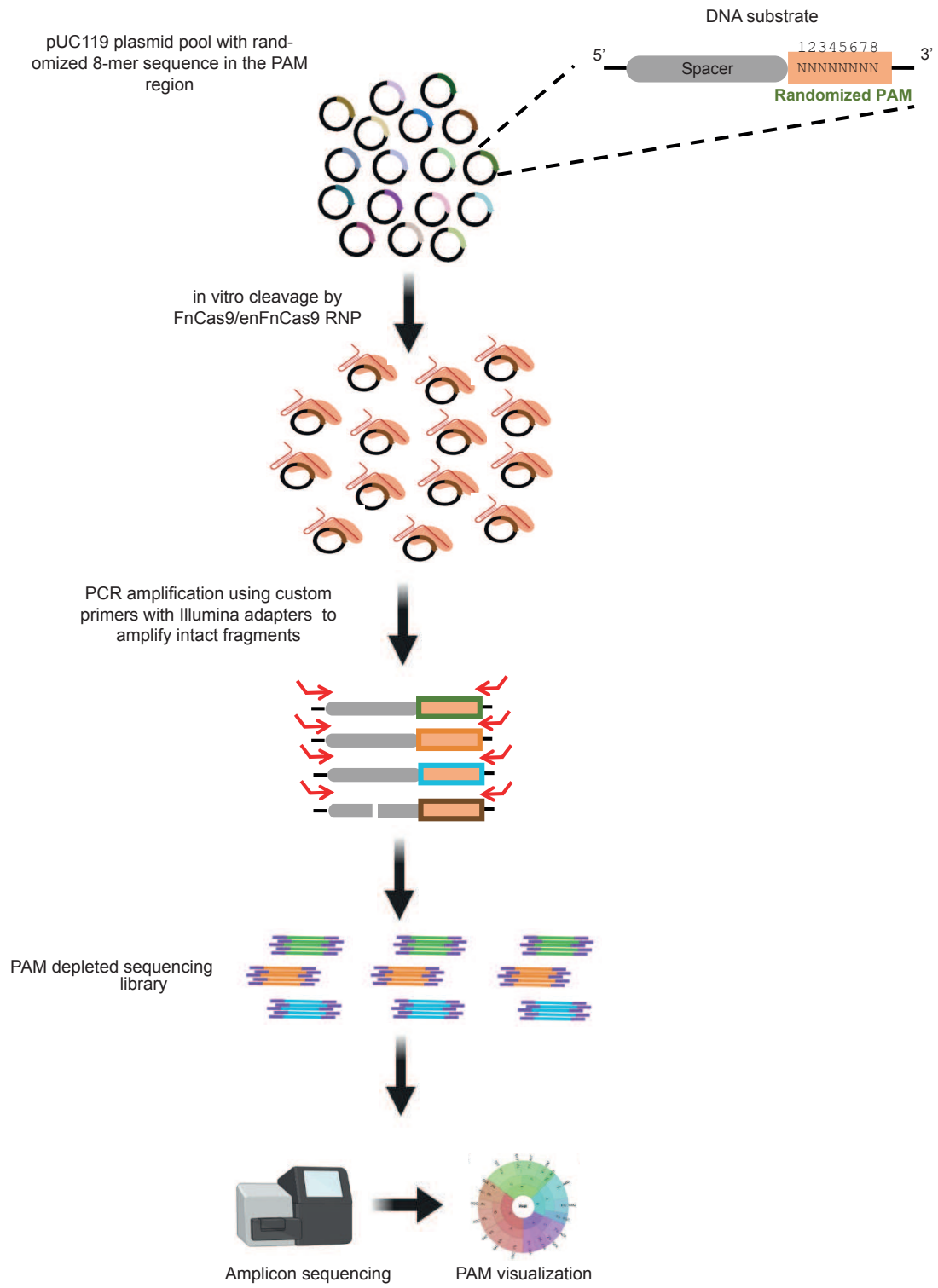
Supplementary Figure 2.

A) Heat map showing the *in vitro* cleavage screening of FnCas9 (WT) and enFnCas9 variants using GGG PAM containing DNA substrate expressed as cleavage percentage for 0.5 and 1 min. Red dotted arrows indicate the subset of nine enFnCas9 variants selected for studying enzymatic activity.

B) Heat map showing the *in vitro* cleavage screening of a subset of FnCas9 (WT) and enFnCas9 variants using GGA PAM containing DNA substrate expressed as cleavage percentage for 0.5 and 1 min. Red dotted arrows indicate the subset of five enFnCas9 variants selected for further evaluation.

C) Box plot showing *in vitro* cleavage assay using GGA PAM containing DNA substrate expressed as cleavage percentage (y-axis) as a function of time (x-axis). Error bars represent SD (three independent experiments).

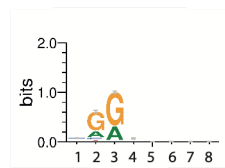
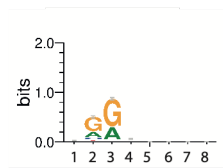
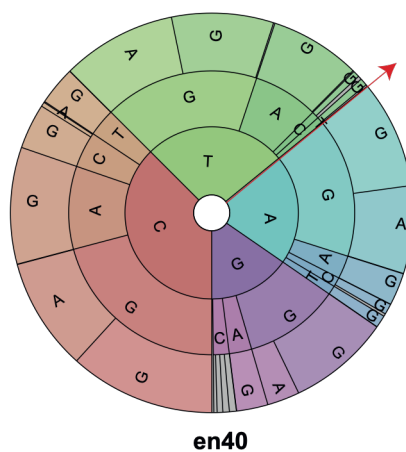
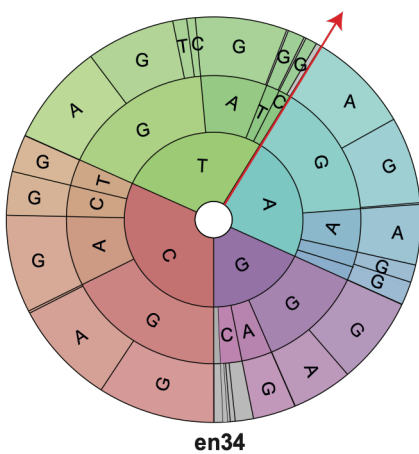
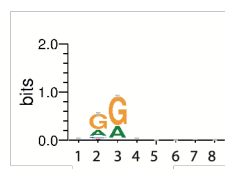
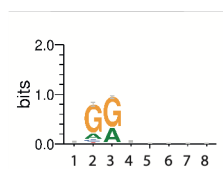
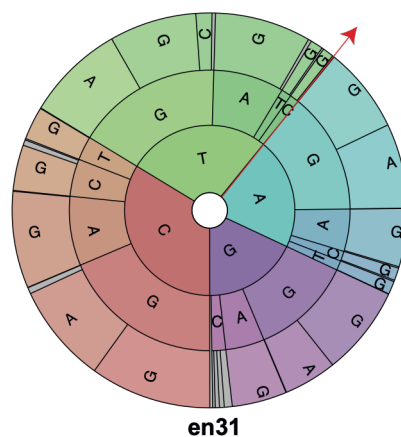
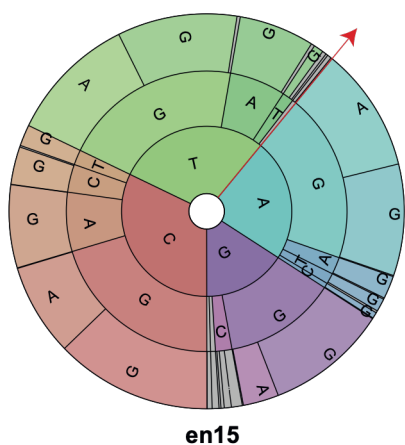
Supplementary Figure 3. Pipeline of PAM discovery assay to comprehensively characterize PAM preferences of enFnCas9 variants



Supplementary Figure 3

Schematic representation showing the pipeline of PAM discovery assay. Main steps in the assay are highlighted.

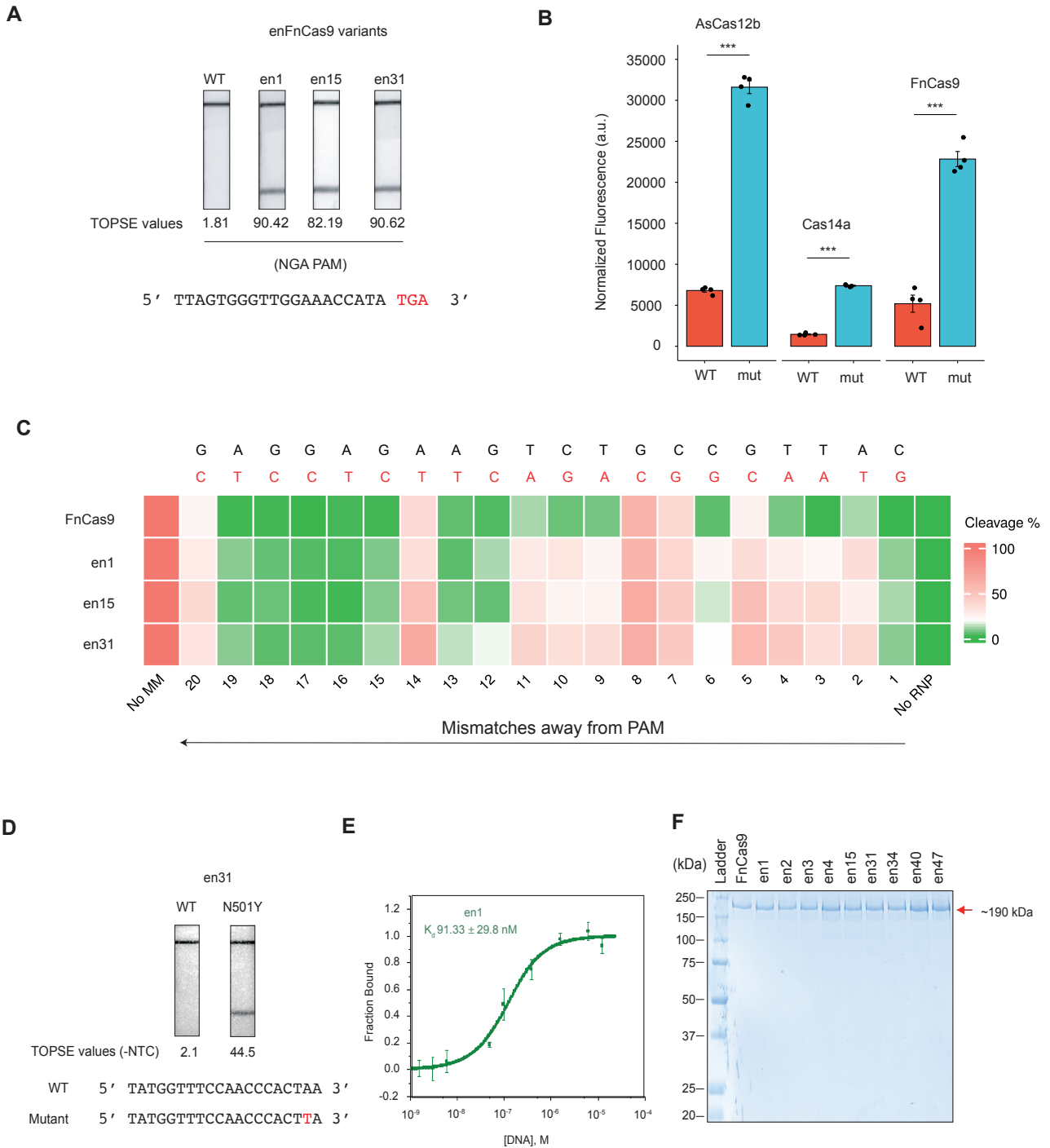
Supplementary Figure 4. PAM wheels and sequence logos to represent PAM preferences of enFnCas9 variants



Supplementary Figure 4.

The PAM wheels and sequence logos showing the results obtained after PAM discovery assay for en15, en31, en34 and en40. Individual sections of the pie charts in the PAM wheels with $\leq 2\%$ depletion enrichment are shown in gray. Bases from the inner to the outer circle in the PAM wheels map the PAM reads away from the target region in the 5' to 3' direction as shown by red arrows.

Supplementary Figure 5. enFnCas9 variants retains specificity of target DNA interrogation



Supplementary Figure 5.

A) Outcome of lateral flow assay (LFA) on TGA PAM for WT, en1, en15 and en31. Corresponding TOPSE values are given at the bottom.

B) Bar plot showing the discrimination of a single mismatched substrate by dFnCas9, AaCas12b and Cas14a1 using fluorescence based assays (affinity based for FnCas9 and trans-cleavage for AaCas12b and Cas14a1). Error bars represent s.e.m (n=3 independent experiments). Student's unpaired *t-test* p-values are represented for *** <0.001.

C) Heat map showing the *in vitro* cleavage outcome of FnCas9, en1, en15 and en31 on HBB and its mutant substrates. Each substrate is harboring a single mutation (position is indicated by counting away from PAM) across the target length. Mismatch containing base is indicated in red while other unaltered bases are shown in black. 25nM of DNA substrates was incubated with respective 100 nM RNPs for 15 min at 37°C.

D) Outcome of lateral flow assay (LFA) for SARS-Cov2 N501Y (with a NGG PAM) mutation detection by en31 using RAY. WT and N501Y target sequences are shown. The N501Y mutation is represented in red. Corresponding TOPSE values are given at the bottom.

E) MST result showing the binding affinity of en1 on *VEGFA3* substrate DNA. Data is represented as a fraction bound RNP (y-axis) with respect to purified DNA substrate (Molar units M, x-axis). Error bars represent SD (3 independent experiments).

F) Coomassie gel showing purified FnCas9 and enFnCas9 protein variants used in the study.

Supplementary Note 1

Screening of enFnCas9 variants on the basis of substrate cleavage

A subset of nine enFnCas9 variants were ranked based on the number of amino acid substitutions and its position on the protein. Out of nine variants, five enFnCas9 variants (en1, en2, en3, en4 and en15) harbor single amino acid substitution and four variants (en31, en34, en40 and en47) carry a combination of single amino acid substitutions (Supplementary Table 3). The variants en1, en15, en31 and en40 showed >80% substrate DNA cleavage within 0.5 min (Supplementary Figure 2A). Furthermore, en1, en15 and en31 showed around two fold higher cleavage activity w.r.t wild-type protein in the *in vitro* time kinetics assay (Figure 2B).

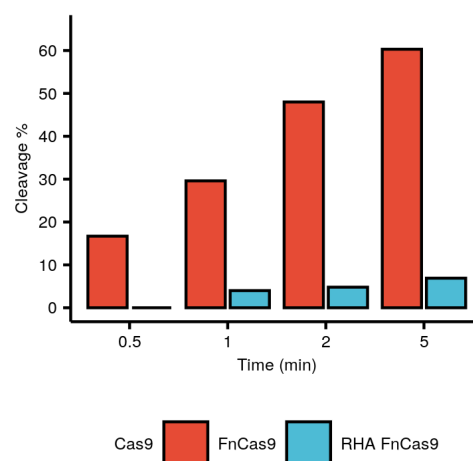
Details of structure-guided engineering

The crystal structure of FnCas9 showed that the PAM duplex is nestled in the FnCas9 WED-PI domain, and the major groove dG(2) and dG(3) of 5'-NGG-3' PAM on the non-target strand are recognized by R1585 and R1556 respectively through bidentate hydrogen bonds while dN(1) is free from any protein contacts. R1556 also recognizes dA(3) by a single hydrogen bond which signifies the very weak interaction with 5'-NGA-3' PAM by FnCas9. Towards developing a PAM-relaxed FnCas9NG variant, earlier work reported that substituting R1556A abrogates the protein function and thus relaxing the PAM constraint to NG PAM cannot be done by this approach. Partial rescue of functional activity was achieved by incorporating base non-specific interactions E1369R and E1449H. This led to the creation of the RHA-FnCas9 variant with activity on YG PAM(3). However, we observed that the RHA variant has very poor activity even on the canonical NGG PAM necessitating the development of alternate variants with improved activity (current manuscript) (Supplementary Note 1, Figure 1). In previous studies, the major groove adenine:glutamine contact had been reported(2, 61). Substituting R1556Q (en49) completely destroyed the protein function *in vitro*, recapitulating the earlier observation seen with R1556A. Expectedly, R1556T(en17) substitution designed for 5'-NGT-3' PAM recognition considering the preference of threonine for thymine base failed to show any protein activity. This confirms the interaction between R1556 with dG(3) of non-target strand of the PAM duplex is

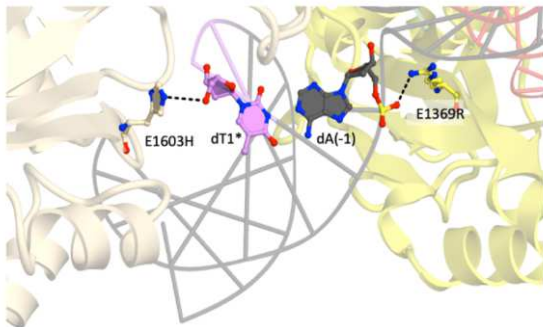
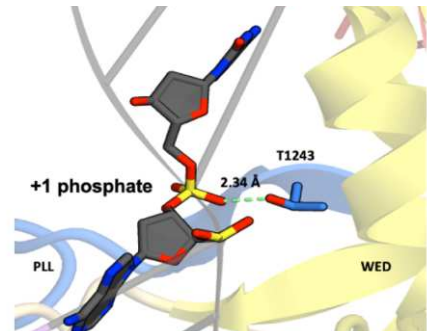
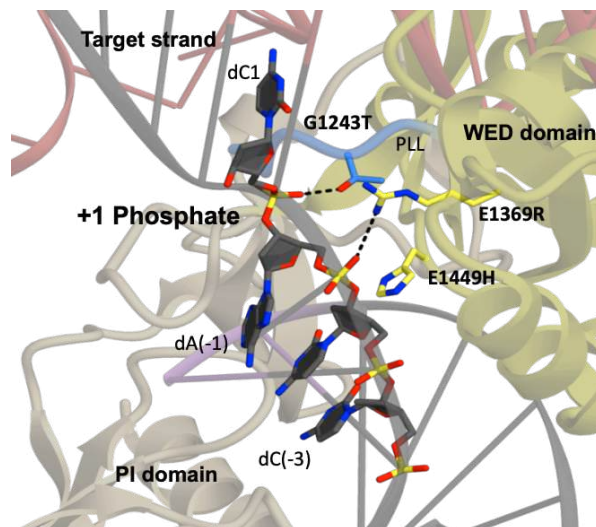
indispensable for FnCas9 functional activity. Interestingly, the incorporation of novel base non-specific interactions with the WED-PI domain of the protein created a subset of three variants (en1, en15 and en31) which showed enhanced enzymatic activity on both canonical NGG PAM and non-canonical NGA PAM. This subset also showed NGR/NRG PAM recognition which broadens the targeting scope in the genome. The crystal structure entails that the en1(E1369R) and en15(E1603H) variants create additional interactions with the backbone phosphate group between dC(-2) and dA(-1) in the target DNA strand and between dG(2) and dT(1) in the non-target strand respectively. In the triple mutant (E1369R/E1449H/G1243T, en31), E1369R/E1449H makes hydrogen bonding interaction with phosphate backbone between dC(-2) and dA(-1) in the target DNA strand and the phosphate-lock loop (PLL) G1243T makes additional hydrogen bonding interaction with +1 phosphate in the target strand from where DNA unwinding ensues (Supplementary Note 1 Figure 2).

Notably, amino acid substitutions of the protein with PAM-duplex base non-specific interactions which results in enhanced kinetic activity (en48, en50) failed to rescue the activity of R1556Q FnCas9.

It is important to note that G1243T alone cannot enhance functional activity to a greater extent. However, G1243T in combination with E1369R and E1603H (en31) exhibited robust *in vitro* kinetic enhancement on both NGG and NGA PAM.



Supplementary Note 1 Figure 1: In vitro activity of RHA FnCas9 on GGG PAM containing substrates represented by percentage cleavage (y-axis) at different time points (x-axis).

A.**B****C**

Supplementary Note 1 Figure 2: Structural models showing interaction between substituted amino acids and PAM duplex. **A)** Interactions of en1 (E1369R) and en15 (E1603H) are shown. **B)** Interaction of en4 (G1243T) with +1 Phosphate group. **C)** Interactions of en31 (E1369R/E1603H/G1243T) with PAM duplex and PLL loop.

References

1. H. Nishimasu, F. A. Ran, P. D. Hsu, S. Konermann, S. I. Shehata, N. Dohmae, R. Ishitani, F. Zhang, O. Nureki, Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*. **156**, 935–949 (2014).
2. C. Anders, O. Niewoehner, A. Duerst, M. Jinek, Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. **513**, 569–573 (2014).
3. H. Hirano, J. S. Gootenberg, T. Horii, O. O. Abudayyeh, M. Kimura, P. D. Hsu, T. Nakane, R. Ishitani, I. Hatada, F. Zhang, H. Nishimasu, O. Nureki, Structure and Engineering of *Francisella novicida* Cas9. *Cell*. **164**, 950–961 (2016).
4. S. Acharya, A. Mishra, D. Paul, A. H. Ansari, M. Azhar, M. Kumar, R. Rauthan, N. Sharma, M. Aich, D. Sinha, S. Sharma, S. Jain, A. Ray, S. Jain, S. Ramalingam, S. Maiti, D. Chakraborty, *Francisella novicida* Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20959–20968 (2019).
5. N. Kim, H. K. Kim, S. Lee, J. H. Seo, J. W. Choi, J. Park, S. Min, S. Yoon, S.-R. Cho, H. H. Kim, Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat. Biotechnol.* **38**, 1328–1336 (2020).
6. M.-S. Liu, S. Gong, H.-H. Yu, K. Jung, K. A. Johnson, D. W. Taylor, Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nature Communications*. **11** (2020), , doi:10.1038/s41467-020-17411-1.
7. J.-J. Liu, N. Orlova, B. L. Oakes, E. Ma, H. B. Spinner, K. L. M. Baney, J. Chuck, D. Tan, G. J. Knott, L. B. Harrington, B. Al-Shayeb, A. Wagner, J. Brötzmann, B. T. Staahl, K. L. Taylor, J. Desmarais, E. Nogales, J. A. Doudna, CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*. **566**, 218–223 (2019).

8. P. Pausch, B. Al-Shayeb, E. Bisom-Rapp, C. A. Tsuchida, Z. Li, B. F. Cress, G. J. Knott, S. E. Jacobsen, J. F. Banfield, J. A. Doudna, CRISPR-Cas Φ from huge phages is a hypercompact genome editor. *Science*. **369**, 333–337 (2020).
9. B. Zetsche, J. S. Gootenberg, O. O. Abudayyeh, I. M. Slaymaker, K. S. Makarova, P. Essletzbichler, S. E. Volz, J. Joung, J. van der Oost, A. Regev, E. V. Koonin, F. Zhang, Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. **163**, 759–771 (2015).
10. T. Karvelis, G. Bigelyte, J. K. Young, Z. Hou, R. Zedaveinyte, K. Budre, S. Paulraj, V. Djukanovic, S. Gasior, A. Silanskas, Č. Venclovas, V. Siksnyš, PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Research*. **48** (2020), pp. 5016–5023.
11. P. Chatterjee, N. Jakimo, J. M. Jacobson, Minimal PAM specificity of a highly similar SpCas9 ortholog. *Sci Adv*. **4**, eaau0766 (2018).
12. D. Collias, C. L. Beisel, CRISPR technologies and the search for the PAM-free nuclease. *Nat. Commun*. **12**, 555 (2021).
13. M. J. Schmidt, A. Gupta, C. Bednarski, S. Gehrig-Giannini, F. Richter, C. Pitzler, M. Gamalinda, C. Galonska, R. Takeuchi, K. Wang, C. Reiss, K. Dehne, M. J. Lukason, A. Noma, C. Park-Windhol, M. Allocca, A. Kantardzhieva, S. Sane, K. Kosakowska, B. Cafferty, J. Tebbe, S. J. Spencer, S. Munzer, C. J. Cheng, A. Scaria, A. M. Scharenberg, A. Cohnen, W. M. Coco, Improved CRISPR genome editing using small highly active and specific engineered RNA-guided nucleases. *Nat. Commun*. **12**, 4219 (2021).
14. D. Y. Kim, J. M. Lee, S. B. Moon, H. J. Chin, S. Park, Y. Lim, D. Kim, T. Koo, J.-H. Ko, Y.-S. Kim, Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol*. (2021), doi:10.1038/s41587-021-01009-z.

15. Z. Wu, Y. Zhang, H. Yu, D. Pan, Y. Wang, Y. Wang, F. Li, C. Liu, H. Nan, W. Chen, Q. Ji, Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat. Chem. Biol.* (2021), doi:10.1038/s41589-021-00868-6.
16. X. Xu, A. Chemparathy, L. Zeng, H. R. Kempton, S. Shang, M. Nakamura, L. S. Qi, Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell* (2021), doi:10.1016/j.molcel.2021.08.008.
17. F. Chen, X. Ding, Y. Feng, T. Seebeck, Y. Jiang, G. D. Davis, Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat. Commun.* **8**, 14958 (2017).
18. J. S. Chen, Y. S. Dagdas, B. P. Kleinstiver, M. M. Welch, A. A. Sousa, L. B. Harrington, S. H. Sternberg, J. K. Joung, A. Yildiz, J. A. Doudna, Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*. **550**, 407–410 (2017).
19. S. K. Jones Jr, J. A. Hawkins, N. V. Johnson, C. Jung, K. Hu, J. R. Rybarski, J. S. Chen, J. A. Doudna, W. H. Press, I. J. Finkelstein, Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. *Nat. Biotechnol.* **39**, 84–93 (2021).
20. J. Tycko, M. Wainberg, G. K. Marinov, O. Ursu, G. T. Hess, B. K. Ego, Aradhana, A. Li, A. Truong, A. E. Trevino, K. Spees, D. Yao, I. M. Kaplow, P. G. Greenside, D. W. Morgens, D. H. Phanstiel, M. P. Snyder, L. Bintu, W. J. Greenleaf, A. Kundaje, M. C. Bassik, Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat. Commun.* **10**, 4063 (2019).
21. H. O'Geen, I. M. Henry, M. S. Bhakta, J. F. Meckler, D. J. Segal, A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* **43**, 3389–3404 (2015).
22. X. Wu, D. A. Scott, A. J. Kriz, A. C. Chiu, P. D. Hsu, D. B. Dadon, A. W. Cheng, A. E. Trevino, S. Konermann, S. Chen, R. Jaenisch, F. Zhang, P. A. Sharp, Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676 (2014).

23. C. Kuscu, S. Arslan, R. Singh, J. Thorpe, M. Adli, Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683 (2014).
24. F. Jiang, J. A. Doudna, CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics.* **46** (2017), pp. 505–529.
25. S. H. Sternberg, S. Redding, M. Jinek, E. C. Greene, J. A. Doudna, DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature.* **507**, 62–67 (2014).
26. S. H. Sternberg, B. LaFrance, M. Kaplan, J. A. Doudna, Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature.* **527**, 110–113 (2015).
27. Y. S. Dagdas, J. S. Chen, S. H. Sternberg, J. A. Doudna, A. Yildiz, A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv.* **3**, eaao0027 (2017).
28. I. C. Okafor, D. Singh, Y. Wang, M. Jung, H. Wang, J. Mallon, S. Bailey, J. K. Lee, T. Ha, Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic Acids Res.* **47**, 11880–11888 (2019).
29. S. Gong, H. H. Yu, K. A. Johnson, D. W. Taylor, DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* **22**, 359–371 (2018).
30. H. Nishimasu, X. Shi, S. Ishiguro, L. Gao, S. Hirano, S. Okazaki, T. Noda, O. O. Abudayyeh, J. S. Gootenberg, H. Mori, S. Oura, B. Holmes, M. Tanaka, M. Seki, H. Hirano, H. Aburatani, R. Ishitani, M. Ikawa, N. Yachie, F. Zhang, O. Nureki, Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science.* **361**, 1259–1262 (2018).
31. S. Hirano, H. Nishimasu, R. Ishitani, O. Nureki, Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol. Cell.* **61**, 886–894 (2016).

32. C. Anders, K. Bargsten, M. Jinek, Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Molecular Cell*. **61** (2016), pp. 895–902.
33. R. T. Walton, K. A. Christie, M. N. Whittaker, B. P. Kleinstiver, Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*. **368**, 290–296 (2020).
34. M. Azhar, R. Phutela, M. Kumar, A. H. Ansari, R. Rauthan, S. Gulati, N. Sharma, D. Sinha, S. Sharma, S. Singh, S. Acharya, S. Sarkar, D. Paul, P. Kathpalia, M. Aich, P. Sehgal, G. Ranjan, R. C. Bhoyar, Indian CoV2 Genomics & Genetic Epidemiology (IndiCovGEN) Consortium, K. Singhal, H. Lad, P. K. Patra, G. Makharia, G. R. Chandak, B. Pesala, D. Chakraborty, S. Maiti, Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis. *Biosens. Bioelectron.* **183**, 113207 (2021).
35. M. Kumar, S. Gulati, A. H. Ansari, R. Phutela, S. Acharya, M. Azhar, J. Murthy, P. Kathpalia, A. Kankan, R. Maurya, J. S. Vasudevan, A. S. R. Pandey, S. Maiti, D. Chakraborty, FnCas9-based CRISPR diagnostic for rapid and accurate detection of major SARS-CoV-2 variants on a paper strip. *Elife*. **10** (2021), doi:10.7554/eLife.67130.
36. J. S. Chen, E. Ma, L. B. Harrington, M. Da Costa, X. Tian, J. M. Palefsky, J. A. Doudna, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*. **360**, 436–439 (2018).
37. L. B. Harrington, D. Burstein, J. S. Chen, D. Paez-Espino, E. Ma, I. P. Witte, J. C. Cofsky, N. C. Kyrpides, J. F. Banfield, J. A. Doudna, Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*. **362**, 839–842 (2018).
38. J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins, F. Zhang, Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*. **356**, 438–442 (2017).

39. B. P. Kleinstiver, A. A. Sousa, R. T. Walton, Y. E. Tak, J. Y. Hsu, K. Clement, M. M. Welch, J. E. Horng, J. Malagon-Lopez, I. Scarfò, M. V. Maus, L. Pinello, M. J. Aryee, J. K. Joung, Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).
40. L. Zhang, J. A. Zuris, R. Viswanathan, J. N. Edelstein, R. Turk, B. Thommandru, H. T. Rube, S. E. Glenn, M. A. Collingwood, N. M. Bode, S. F. Beaudoin, S. Lele, S. N. Scott, K. M. Wasko, S. Sexton, C. M. Borges, M. S. Schubert, G. L. Kurgan, M. S. McNeill, C. A. Fernandez, V. E. Myer, R. A. Morgan, M. A. Behlke, C. A. Vakulskas, AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. *Nat. Commun.* **12**, 3908 (2021).
41. B. P. Kleinstiver, S. Q. Tsai, M. S. Prew, N. T. Nguyen, M. M. Welch, J. M. Lopez, Z. R. McCaw, M. J. Aryee, J. K. Joung, Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34**, 869–874 (2016).
42. D. Kim, J. Kim, J. K. Hur, K. W. Been, S.-H. Yoon, J.-S. Kim, Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
43. W. X. Yan, R. Mirzazadeh, S. Garnerone, D. Scott, M. W. Schneider, T. Kallas, J. Custodio, E. Wernersson, Y. Li, L. Gao, Y. Federova, B. Zetsche, F. Zhang, M. Bienko, N. Crosetto, BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nature Communications.* **8** (2017), , doi:10.1038/ncomms15058.
44. F. Teng, L. Guo, T. Cui, X.-G. Wang, K. Xu, Q. Gao, Q. Zhou, W. Li, CDetection: CRISPR-Cas12b-based DNA detection with sub-attomolar sensitivity and single-base specificity. *Genome Biol.* **20**, 132 (2019).
45. Y. Wei, Y. Zhou, Y. Liu, W. Ying, R. Lv, Q. Zhao, H. Zhou, E. Zuo, Y. Sun, H. Yang, C. Zhou, Indiscriminate ssDNA cleavage activity of CRISPR-Cas12a induces no detectable off-target effects in mouse embryos. *Protein Cell.* **12**, 741–745 (2021).

46. K. Murugan, A. S. Seetharam, A. J. Severin, D. G. Sashital, CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. *J. Biol. Chem.* **295**, 5538–5553 (2020).
47. S. Q. Tsai, Z. Zheng, N. T. Nguyen, M. Liebers, V. V. Topkar, V. Thapar, N. Wyvekens, C. Khayter, A. J. Iafrate, L. P. Le, M. J. Aryee, J. K. Joung, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
48. B. Wienert, S. K. Wyman, C. D. Richardson, C. D. Yeh, P. Akcakaya, M. J. Porritt, M. Morlock, J. T. Vu, K. R. Kazane, H. L. Watry, L. M. Judge, B. R. Conklin, M. Maresca, J. E. Corn, Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science.* **364**, 286–289 (2019).
49. B. P. Kleinstiver, V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng, J. K. Joung, High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature.* **529**, 490–495 (2016).
50. I. M. Slaymaker, L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, F. Zhang, Rationally engineered Cas9 nucleases with improved specificity. *Science.* **351**, 84–88 (2016).
51. B. P. Kleinstiver, V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, J. Keith Joung, 731. High-Fidelity CRISPR-Cas9 Nucleases with No Detectable Genome-Wide Off-Target Effects. *Molecular Therapy.* **24** (2016), p. S288.
52. A. Casini, M. Olivieri, G. Petris, C. Montagna, G. Reginato, G. Maule, F. Lorenzin, D. Prandi, A. Romanel, F. Demichelis, A. Inga, A. Cereseto, A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).
53. F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, F. Zhang, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
54. P. J. Susaimanickam, S. Maddileti, V. K. Pulimamidi, S. R. Boyinpally, R. R. Naik, M. N. Naik, G. B. Reddy, V. S. Sangwan, I. Mariappan, Generating minicorneal

- organoids from human induced pluripotent stem cells. *Development*. **144**, 2338–2351 (2017).
55. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**, 357–359 (2012).
 56. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
 57. G. Georgiou, S. J. van Heeringen, fluff: exploratory analysis and visualization of high-throughput sequencing data. *PeerJ*. **4**, e2209 (2016).
 58. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).
 59. A. Khan, A. Mathelier, Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics*. **18** (2017), , doi:10.1186/s12859-017-1708-7.
 60. K. Clement, H. Rees, M. C. Canver, J. M. Gehrke, R. Farouni, J. Y. Hsu, M. A. Cole, D. R. Liu, J. K. Joung, D. E. Bauer, L. Pinello, CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
 61. N. M. Luscombe, R. A. Laskowski, J. M. Thornton, Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.* **29**, 2860–2874 (2001).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)