

Combination of engineered FnCas9 and extended gRNAs for PAM-flexible, robust and nucleobase specific editing and diagnostics

Sundaram Acharya^{1,2,#}, Asgar Hussain Ansari^{1,2}, Seiichi Hirano³, Sajal Sarkar^{1,2}, Riya Rauthan^{1,2}, Manoj Kumar^{1,2}, Rhythm Phutela^{1,2}, Sneha Gulati¹, C. Afzal¹, Deepanjan Paul¹, Abdul Rahman¹, Sudipta Mahato^{4,5}, Savitri Maddileti⁴, Vinay Kumar Pulimamidi^{4,6}, Subhadra Jalali⁷, Hiroshi Nishimasu⁸, Indumathi Mariappan⁴, 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

⁶Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA 02114, USA.

⁷Smt. Kanuri Santhamma Centre for Vitreo Retinal Diseases, LV Prasad Eye Institute, Hyderabad-500034, Telangana, India

⁸Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

#Correspondence:

acharyasundaram.ac@gmail.com (S.A.); 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 enhanced FnCas9 (enFnCas9) variants 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 increasing their accessibility across human genomic sites by ~3.5-fold. The enFnCas9 proteins harbor single mismatch specificity both *in vitro* and *in cellulo* leading to broadened target range of FnCas9-based CRISPR diagnostics for detection of point mutations and pathogenic DNA signatures. 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). Remarkably, enFnCas9 variants can be combined with extended length gRNAs for robust base editing at sites which are inaccessible to PAM-constrained canonical base editors. Finally, we show the complete correction of a disease-specific Retinitis Pigmentosa mutation in patient derived iPSCs using enFnCas9 Adenine Base Editor highlighting its broad application in therapeutics and diagnostics.

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 substrates¹⁻⁴. 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 enzyme⁵⁻⁷. To circumvent these issues, in recent years, alternate high-efficiency Cas systems from other microorganisms have been demonstrated for genome

editing^{8–12}. Notably, none show editing efficiencies higher than SpCas9^{5,6,7}, 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^{13–17} (Supplementary Table 1). For base editing applications, constraints introduced by the targeting window (4-9 bp PAM distal for Cas9 variants) with respect to the nearest PAM has necessitated engineering Cas9 proteins with altered PAM specificities to access nucleobase targets on a case-to-case basis¹⁸. Thus there is an unmet need of Cas effectors that can combine high activity, specificity and a flexible base editing window, that is not constrained by its distance to the nearest available PAM as seen in canonical base editors.

Results

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*^{4,19}. In contrast, SpCas9 and its high-fidelity variants remain bound to off-target sites in a cleavage incompetent fashion, a property that might cause non-specific off-targeting outcomes from such regions^{20,21,22}. To investigate if FnCas9's high DNA binding specificity is reflected on a genome wide level, we constructed catalytically inactive (dead, d) dSpCas9 and dFnCas9 and targeted the *c-Myc* locus where comparable cellular editing efficiencies between SpCas9 and FnCas9 were observed previously⁴. Using chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-Seq)^{23–25}, we found that although both dSpCas9 and dFnCas9 were tightly bound to the on-target sites, 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 (Supplementary Figure 1A). Whereas, all the 27 dSpCas9 off-target sites showed greater enrichment than the on-target, dFnCas9 was bound to 6 off-target sites (0.01 FDR) all of which showed at least 1.2-fold lower enrichment than the on-target (Supplementary Figure 1B-C; Supplementary Table 2). This high specificity of binding *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.

94 Towards engineering the protein, we used an approach of stabilizing FnCas9:DNA
95 binding by introducing non-specific FnCas9:PAM interactions, based on recent
96 mechanistic studies on SpCas9 highlighting the directional PAM-duplex DNA unwinding
97 as the rate-limiting checkpoint of Cas9 action for R-loop expansion (Supplementary Note
98 1)^{21, 26-31}. Additionally, we also investigated if the FnCas9 sgRNA length might also be a
99 factor determining its DNA cleavage activity.

100 To discover the optimal length of gRNA for FnCas9, we performed an *in vitro* cleavage
101 assay using a previously reported target DNA harboring a stretch of guanines with
102 FnCas9 RNP containing variable length of gRNAs ranging from 20 to 24 nucleotides³²
103 (nt, hereafter referred g20-g24). Interestingly, we observed the lowest activity with the
104 canonical g20 while all other extended length gRNAs exhibited enhanced DNA cleavage
105 rate with g21 inducing the fastest rate of cleavage (Supplementary Figure 2A). We used
106 g21 in all our subsequent assays unless stated otherwise.

107 Next, we engineered 49 different FnCas9 variants guided by its crystal structure bearing
108 mostly single amino acid substitutions in the WED-PI domain to introduce novel PAM
109 duplex DNA contacts (Figure 1A, Supplementary Table 3). We then measured *in vitro*
110 DNA cleavage activities of the FnCas9 variants against a DNA target containing GGG
111 PAM (where FnCas9 was shown to be least active)³ and performed DNA cleavage
112 experiments with the engineered variants. Recent reports have suggested that high-
113 fidelity SpCas9 variants have slower enzyme kinetics and concomitant lower editing
114 efficiencies^{5,6}. FnCas9 being an enzyme with high fidelity, we therefore focused on
115 engineered (en) FnCas9 variants showing a combination of faster cleavage rate and
116 minimum structural alterations to ensure that its intrinsic specificity remains unchanged.
117 A subset of nine enFnCas9 variants (containing single/combinatorial mutations) were
118 selected for downstream experiments satisfying this criteria (Figure 1B, Supplementary
119 Note1). Among them, three variants (en1, en15 and en31) had at least 2-fold higher
120 cleavage rates than the wild type protein (Figure 1B). Intrigued by this observation, we
121 tested cleavage efficiency of two of the enFnCas9 variants (en15 and en31) using super-
122 extended gRNAs with 26 to 28-nt protospacer (g26-g28, hereafter referred as sx-gRNA)
123 and confirmed similar cleavage efficiencies as g21 which suggests the compatibility of
124 enFnCas9 variants with sx-gRNAs (Supplementary Figure 2B). To our knowledge, similar

observations have not been made so far for other Cas systems and this might offer further enhancement of specificity and nucleobase accessibility away from PAM as shown later. Earlier reports have shown that engineered SpCas9 variants often create additional phosphate backbone interactions and facilitate these proteins to recognize non-canonical PAMs^{33,34}. 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) and performed an *in vitro* PAM discovery assay. Deep sequencing of the PAM depleted library containing randomized 8 bp sequence ($4^8 = 65,536$ combinations in total) revealed that enFnCas9 variants showed more flexible recognition in second and third nucleotide positions as compared to FnCas9 (Supplementary Figures 3,4). Importantly, for all the enFnCas9 variants tested, NGG PAM was relaxed to NGR/NRG thereby expanding (~3.5 fold over wild type FnCas9) the scope of accessibility across the human genome to just below SpCas9-RY³⁵ and SpCas9-NG³¹ (Figure 1D, Supplementary Figure 6A, Supplementary Table 1).

The remarkable intrinsic specificity of FnCas9 to single-nucleotide mismatches in the target has proven effective both in disease diagnostics and disease correction⁴. At the level of diagnostics, FnCas9 has been utilized for paper strip-based robust detection of nucleic acid targets through the FnCas9 Editor Linked Uniform Detection Assay (FELUDA) and Rapid Variant Assay (RAY) platforms^{36,37}. In contrast to collateral cleavage based platforms employed by Type V effectors (such as Cas12a³⁸ or Cas12f³⁹) or Type VI effectors (such as Cas13⁴⁰), FELUDA and RAY uses the specificity of direct FnCas9:DNA binding as a lateral-flow readout through a combination of FAM-labeled FnCas9:sgRNA complex and paper strip chemistry (Figure 1D)^{36,37}. We anticipated that in comparison to FnCas9-based FELUDA, enFnCas9 (with NRG/NGR PAM)-based FELUDA can now cover ~2-fold higher number of reported Mendelian SNVs across the human genome thereby increasing the scope of detection to more disease-causing variants (Figure 1E). Expectedly, on a lateral flow strip, all enFnCas9 variants tested

Figure 1. Engineering and characterization of FnCas9 variants for enhanced activity on canonical PAM, its altered PAM activity and CRISPR diagnostics

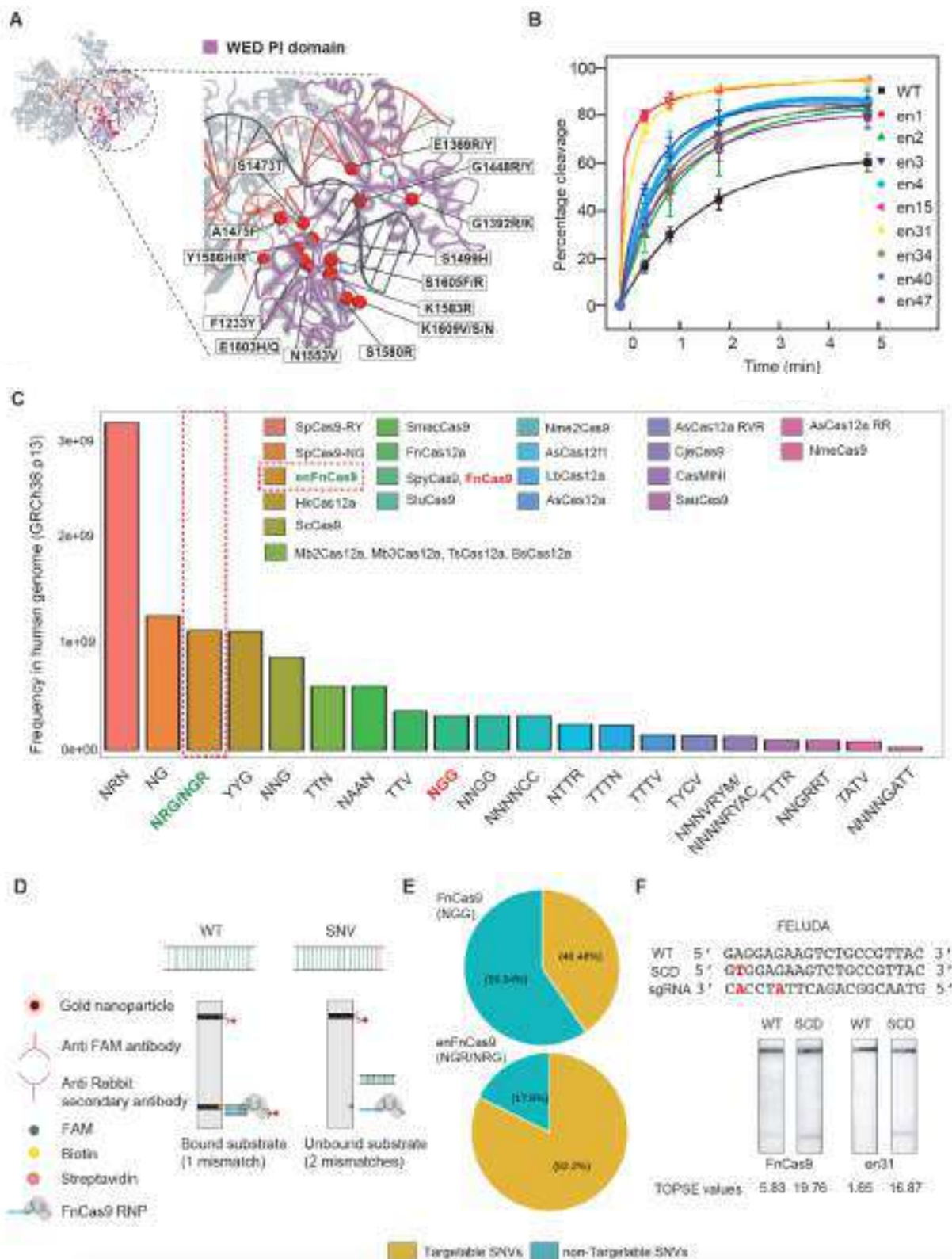


Figure 1:

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. G1243T mutation on phosphate lock loop (PLL) was not shown.

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 mean \pm SD (three independent experiments).

C. Bar plot showing the availability of PAMs of respective Cas effectors 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 enFnCas9 variants.

D. Schematic showing the mode of SNV detection by FELUDA and RAY CRISPRDx platforms.

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

F. 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.

(complexed with 20-nt gRNA, g20) showed robust activity on a substrate carrying the non-canonical NGA PAM whereas FnCas9 did not show any signal (Supplementary Figure 5A). Importantly, enFnCas9 variants showed similar resolution of single nucleotide variant (SNV) diagnosis (4.4-fold) as compared to AaCas12b (4.6-fold) and Cas14a1 (5.1-fold) both of which belong to type V DNA targeting Cas systems and have been reported to have higher intrinsic specificity than SpCas9^{10,11,15,16,41,42} further establishing its utility as a diagnostic platform (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 g20, 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 part 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 positions⁴. To determine if these changes in enFnCas9 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 2F, 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 (Supplementary Figure 5F). 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.

We next investigated if engineering FnCas9 by altering residues that interact with PAM in the substrate had altered its binding affinity to DNA. Using catalytically inactive versions of two of the variants (en1 and en15) we performed microscale thermophoresis (MST) to determine their DNA binding affinities on a substrate (*VEGFA*) with a 20-nt gRNA as reported earlier⁴. 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 (Supplementary Figure 6 B,C). Interestingly, in our previous study⁴, we showed 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 determined 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^{38,43}. Interestingly, Cas12a has been reported to nick off-target DNA substrates with up to four mismatches depending upon the crRNA sequences employed⁴⁴. 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 construction of high-fidelity SpCas9 proteins have improved its overall specificity, this is also accompanied by lower editing efficiencies^{5,45,46}. We selected two such proteins (SpCas9-HF1 and eSpCas9) due to their balanced activity and specificity as reported in literature^{5,45,46} and compared their cellular editing rates (insertion/deletions) with one of the enFnCas9 variants, en1. We used 20-nt protospacer containing gRNAs for which bona-fide off-targets were identified either through *in silico* prediction or GUIDE-Seq^{4,47}. Encouragingly, en1 showed higher editing rates than the wild-type protein or the SpCas9-HF1 and eSpCas9 variants at all the loci tested without any detectable editing at the corresponding off-targets (Figure 2A, Supplementary Figure 6D-E). Similarly, we confirmed successful genome editing by enFnCas9 variants (en1 and en15) in retinal pigmented epithelial cells (ARPE-19) and induced pluripotent stem cells (iPSCs) (Supplementary Figure 6F-G). Notably, in iPSCs, en1 (18.6% indels) and

en15 (23.0% indels) showed superior editing rates at the *PAX6* locus when compared to even SpCas9 (13.8%) in unsorted cell populations (Supplementary Figure 6F).

As seen in our *in vitro* studies, editing rate with enFnCas9 variants went up dramatically reaching ~90% at therapeutically relevant sickle cell locus *HBB* in HEK293T cells when combined with g21 (Figure 2B). Similarly, g21 gave robust genome editing outcomes (up to 90%) with all the enFnCas9 variants at other loci too (*EMX1* and *FASN*) (Supplementary figure 7A-B).

Next, we investigated if the high editing efficiency and DNA binding affinity compromised the single mismatch specificity of the enFnCas9 variants. To this end, we interrogated the *FANCF* site2 in HEK293T cells for which GUIDE-Seq validated off-target with a single PAM proximal mismatch was reported even by high-fidelity SpCas9 variants from independent studies^{21,45,48}. Expectedly, we found comparable off-target editing (25% and 27%) as the on-target site (30% and 29%) by SpCas9-HF1 and eSpCas9 respectively (Figure 2C). In sharp contrast, negligible (~1%) editing at the single mismatch off-target was observed for all the enFnCas9 variants when g20 was used, albeit with lower on-target editing (15-20% across the enFnCas9 variants) while FnCas9 did not induce substantial editing (~2%) (Figure 2C). Interestingly, using g21 or g22 increased the on-target editing efficiency up to 45% with en15 but no increase in off-targeting was seen (Figure 2C). A similar trend was seen for both en1 and en31 although en1 showed a small increase in off-target editing with a g21/22 which was still around three-fold lower than the high fidelity SpCas9 variants tested (Figure 2C). Taken together, this underscores the combinatorial action of enFnCas9 variants and extended length gRNAs for highly precise and robust editing.

The activity of enFnCas9 variants on non-canonical PAMs (NGR/NRG) observed *in vitro* prompted us to evaluate the genome editing efficiencies of these variants on such altered PAM targets in human cells. Given its highest *in vitro* rate of DNA cleavage both at canonical NGG and non-canonical NGA PAM, en31 was additionally examined for cellular genome editing on the targets with NGA/NAG PAM. Two GUIDE-Seq validated gRNAs targeting an NGA PAM at *RUNX1* and *ZNF629* that had previously been reported⁴⁹ with highly promiscuous off-targets were investigated alongside an additional NGA containing *FANCF1* site2 gRNA. We confirmed robust editing at all the three loci (~80% at *FANCF1*,

Figure 2. enFnCas9 variants enable robust and specific nuclease and base editing in human cells

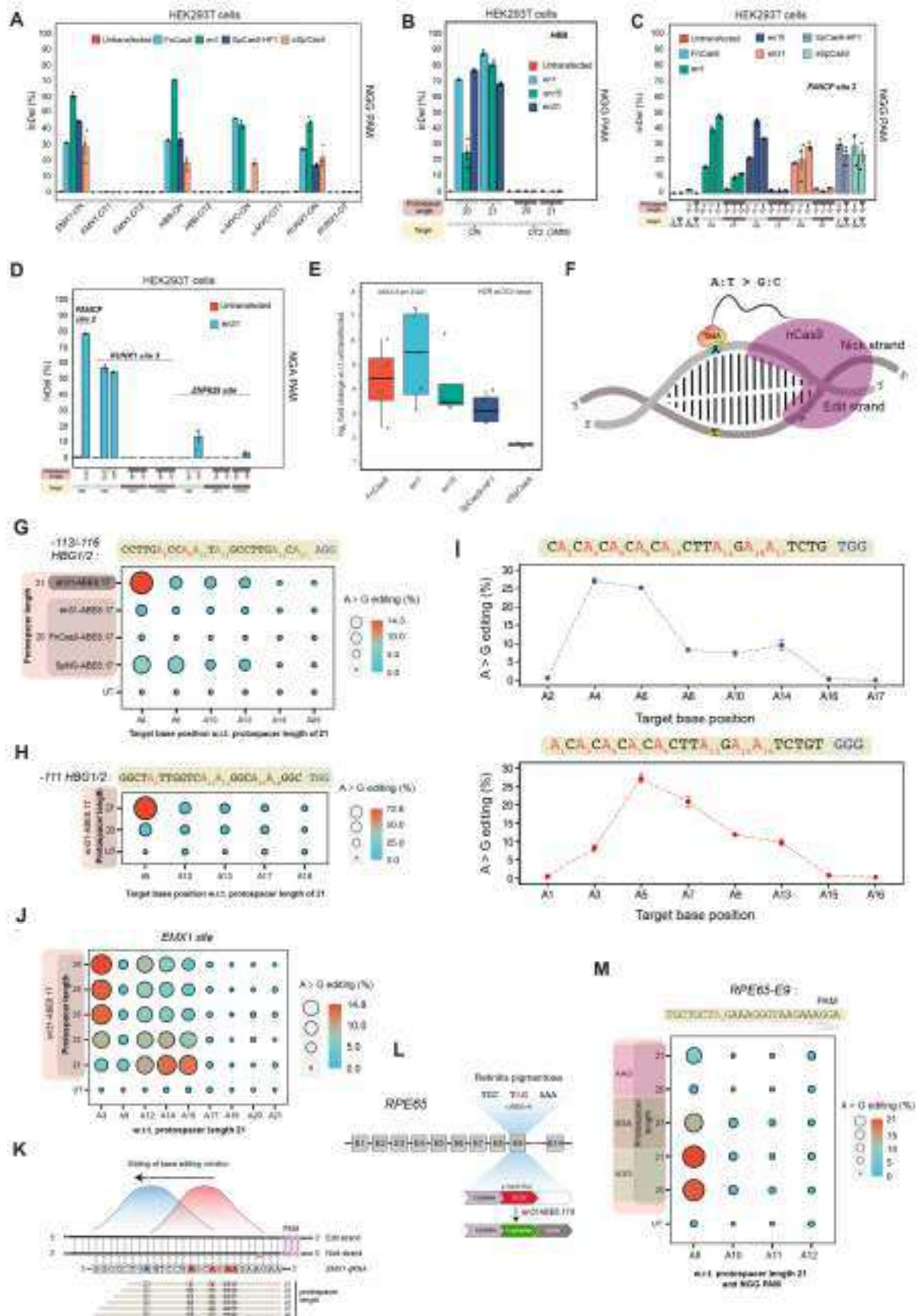


Figure 2:

A. Bar plot showing the InDel events (expressed in percentage) plotted on the Y-axis as obtained from amplicon sequencing upon targeting NGG PAM containing *EMX1*, *HBB*, *c-MYC*, *RUNX1* loci and its respective off-targets (OTs) by FnCas9, en1, SpCas9-HF1 in HEK293T cells. Untransfected cells were used as control. Error bars represent SEM of n=3 independent biological replicates with individual values shown as dots.

B. Bar plot showing the InDel events (expressed in percentage) plotted on the Y-axis as obtained from amplicon sequencing upon targeting NGG PAM containing *HBB* locus and its off-target site (OT2) by en1, en15 and en31 with either sgRNA containing 20-nt protospacer (g20) or sgRNA containing 21-nt protospacer (g21) in HEK293T cells. Untransfected cells were used as control. Error bars represent SEM of n=3 independent biological replicates with individual values shown as dots.

C. Bar plot showing the InDel events (expressed in percentage) plotted on the Y-axis as obtained from amplicon sequencing upon targeting NGG PAM containing *FANCF1* site2 locus and its off-target site (OT) by FnCas9, en1, en15, en31, SpCas9-HF1 and eSpCas9 with sgRNA containing 20/21/22-nt protospacers (g20, g21 and g22 respectively) in HEK293T cells. Untransfected cells were used as control. Error bars represent SEM of n=3 independent biological replicates with individual values shown as dots.

D. Bar plot showing the InDel events (expressed in percentage) plotted on the Y-axis as obtained from amplicon sequencing upon targeting NGA PAM containing *FANCF1* site2, *RUNX1* site3 and *ZNF629* site loci and its respective off-targets by en31 with either sgRNA containing 20-nt protospacer (g20) or sgRNA containing 21-nt protospacer (g21) in HEK293T cells. Untransfected cells were used as control. Error bars represent SEM of n=3 independent biological replicates with individual values shown as dots.

E. 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 w.r.t. untransfected samples and analysed using one-way Anova, p-value is shown (4 independent experiments).

F. Schematic showing the mode of the action of an adenine base editor (ABE) where the nickase version of Cas9 has been fused to a mutant TadA, a deaminase domain capable

of installing A to G substitution on the adenine target base. The location of the nick by ABE is indicated by a scissor on the nick strand. PAM containing strand hosts the target bases for base editor and represented as edit strand.

G. Ballon plot showing the A to G editing events (expressed in percentage) as obtained from amplicon sequencing upon targeting -113/-116 sites of *HBG1/2* promoter by FnCas9-ABE8.17d, en31ABE8.17d and SpNG-ABE8.17d with sgRNA containing 20/21-nt protospacer (g20 and g21 respectively) in HEK293T cells. Target bases (As) for adenine base editing are on the X-axis and numbered w.r.t. g21. -113 /-116 sites are highlighted in red. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates.

H. Ballon plot showing the A to G editing events (expressed in percentage as obtained from amplicon sequencing upon targeting -111 site of *HBG1/2* promoter by en31ABE8.17d with sgRNA containing 20/21-nt protospacer (g20 and g21 respectively) in HEK293T cells. Target bases (As) for adenine base editing are numbered on the X-axis w.r.t. g21 with the target site (A5) indicated in red. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates.

I. Line plot showing A to G editing events (expressed in percentage) plotted on the Y-axis obtained from amplicon sequencing upon targeting endogenous loci having alternately present target 'A' bases in HEK293T cells. Target bases (As) for adenine base editing are numbered on the X-axis and marked red. Error bars represent SEM of n=3 independent biological replicates.

J. Ballon plot showing the modulation of base editing window by en31ABE8.17d expressed in percentage of A to G editing using gRNAs with extended protospacers (g21 to g26) as obtained from amplicon sequencing upon targeting *EMX1* locus against GGG PAM. Target bases (As) for adenine base editing are on the X-axis and numbered w.r.t. g21. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates.

K. Schematic showing the sliding of the base editing window from primary window (shown in red) to secondary window (shown in blue) by extended gRNAs and indicated by higher efficiency of A to G editing on *EMX1* locus with GGG PAM. A3 (highlighted in blue), the

inaccessible target base by en31ABE8.17d due to PAM restriction with g21 becomes amenable to edit with g23-g26 due to sliding of the optimal editing window to A5-A8 respectively.

L. Schematic showing the patient-specific mutation on exon 9 of *RPE65* led to the generation of premature stop codon and correction of the mutation by en31ABE8.17d treatment leading to restoration of the reading frame of *RPE65*.

M. Ballon plot showing the A to G editing events (expressed in percentage) as obtained from amplicon sequencing upon targeting *RPE65-E9* by en31ABE8.17d with sgRNA containing 20/21-nt protospacer (g20 and g21 respectively) in patient-specific iPSC line. The pathogenic mutation is numbered on the X-axis (A8) and is indicated in red. Both the target base (A8) and the bystander bases are counted w.r.t g21 and AGG PAM. AGG, GGA and AAG PAMs are underlined in the sequence downstream of the respective protospacers of the sgRNAs. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates.

~60% at *RUNX1* and ~20% at *ZNF629*) (Figure 2D). Expectedly, g21 was able to induce editing outcomes wherever g20 failed to do (*ZNF629*) (Figure 2D). Remarkably, while previous reports had shown greater off-target editing than on-target activity with SpCas9 variants with special emphasis on OT12 of *ZNF629* site which is an identical stretch of the on-target site⁴⁹, we were unable to detect any off-target editing with the en31 variant at any of these loci except at OT12 of *ZNF629* site (Figure 2D). Despite being identical to the on-target of *ZNF629* site, off-targeting at OT12 was marginally detected on contrary to SpCas9 variant⁴⁹. Furthermore, we also confirmed robust editing by en31 in one (~70% at *FANCF*) out of three sites having NAG PAM with g21 (Supplementary Figure 7C). Our results suggest that the PAM preference of en31 nuclease ranges from NGG>NGA>NAG while retaining superior specificity of DNA interrogation even in the sites showing preponderance of off-targeting by high-fidelity SpCas9 variants. Finally, we speculated that higher editing outcomes by enFnCas9 variants might reflect in both higher NHEJ mediated indels or HDR mediated knock-in rates. Expectedly, 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 2E). Collectively, en1 nuclease showed a higher rate of gene editing (NHEJ/HDR) at all the target loci tested highlighting its suitability as a highly potent genome-editing protein. Despite the promises in Cas9 nuclease-based gene editing approaches, on-target genotoxicity combined with complex gene rearrangements has raised concerns about its use in therapeutic settings^{50–54}. In contrast, the development of double-strand break (DSB)-free editing approaches such as base editing and prime editing has shown tremendous promise as safer alternatives⁵⁵. Nevertheless, both the approaches suffer from guide-dependent off-targeting due to its reliance on enzymatically defective or inactive Cas9 for binding, an imperative feature for DSB free editing^{56,57}. We sought to develop FnCas9/enFnCas9 base editors owing to its remarkable specificity of binding to cognate nucleobases both *in vitro* and in human cells (Figure 2F). Among the enFnCas9 variants, en31 showed the broadest PAM flexibility and coupled with its robust indel activity in human cells appeared to be an ideal candidate for evaluation as a base editor. To this end, we generated adenine base editor variants for FnCas9/en31 following previously reported ABEmax (ABE8.17dV106W) configurations which are shown to be

highly efficient with improved gRNA-independent editing profiles - a feature important to maintain transcriptome fidelity during base editing^{58,59}. Given the larger share of ABE for pathogenic SNP correction⁵⁵, we characterized FnCas9/en31-ABE for editing in human cells and compared it with SpNG-ABEmax8.17d, another PAM flexible ABE variant that has been widely reported in literature⁵⁹. Given the larger share of ABE for pathogenic SNP correction⁶⁰, we characterized FnCas9/en31-ABE for editing in human cells and compared it with SpNG-ABEmax8.17d, another PAM flexible ABE variant that has been widely reported in literature⁵⁹. We chose the therapeutically relevant -113/-116 sites of *HBG1/2* promoter responsible for hereditary persistence of fetal hemoglobin (HPFH)⁶⁰, a rare genetic condition known to ameliorate Sickle cell disease phenotype and the commonly used *EMX1* site in HEK293T. For both loci, we observed low A>G substitution (1.7%/0.0% A6/A9 of -113/-116 and 3.7% A9 of *EMX1*) with en31ABEmax8.17d with sg20 but drastically improved A>G substitutions (14%/2.5% A6/A9 of -113/-116 and 3.7%/10.7%/13.33%/12.7% of A9, A12, A15 of *EMX1*) with sg21 (Figure 2G, Supplementary Figure 7D). Notably, SpNG-ABEmax8.17d showed reduced editing at both loci (6.7%/5.7% A6/A9 of -113/-116 and 0% of A12, A15 of *EMX1*) while the off-targeting profile in the validated off-target of *EMX1* (*EMX1-OT1*) was poor albeit the off-target editing was very low owing to natural specificity of ABEs (Figure 2G, Supplementary Figure 7D). However, wild type FnCas9ABEmax8.17d did not induce any appreciable A>G substitution over the baseline. We confirmed robust A>G substitution efficiency up to 72% with en31ABEmax8.17d at different sites of the therapeutically relevant *HBG1/2* promoter (-111, -123/124, -175, -198) with g21 outperforming g20 in all the sites tested (Supplementary Figure 7E, F, G). Thus, en31ABEmax8.17d with a g21 showed robust base editing in human cells with higher target base substitutions than SpNG-ABEmax8.17d in the tested loci.

We speculated that widened PAM accessibility coupled with extended length sgRNAs might offer en31ABE distinct possibilities of base editing where conventional base editors might not be able to target the desired base. Because of protospacer length restrictions to 19/20-nt, SpCas9 base editors (ABE8s) can only target bases which are within the targeting window of the deaminase (PAM-distal 3rd to 9th bases counting PAM at positions 21-23)⁵⁹. For editing other sites far away from the nearest available PAM, protein

engineering to recognize a new PAM has been reported¹⁸. The en31ABE protein showed a wider editing window with respect to the PAM (PAM-distal 3rd to 14th bases counting PAM at positions 22-24) when interrogated at a loci with alternate adenine bases (Figure 2I, Supplementary Figure 7H-I). Moreover, since enFnCas9 can tolerate sx-gRNA such as sg26 or sg28 (Supplementary Figure 2B), we hypothesized that combining the two properties could facilitate the shifting of adenine base editing window to target inaccessible bases away from PAM. To validate this, we chose two loci in the human genome *EMX1* (NGG PAM) and *SERPINI1* (NGA PAM) with a target base situated inaccessible PAM-distal positions (3rd position for *EMX1* and 1st position for *SERPINI1*) (Figure 2J, Supplementary Figure 8B). Remarkably, by systematic modulation of gRNA lengths (g22-g26), we were successful in gradually shifting the editing window to the desired target while the target base editing in the primary window got serially diluted (Figure 2J, K; Supplementary Figure 8A, B). Thus, combining PAM flexibility (NGR/NRG) and extended length sgRNAs (up to 26) theoretically improves the target range of en31ABE to 99.39% of all human G>A pathogenic SNVs identified in ClinVar^{62,63} (Supplementary Figure 7J).

Finally, towards the proof-of-concept validation of en31ABE_{max8.17}-based disease correction, we tested this protein for an ophthalmic condition where genetically corrected iPSC-derived Retinal Pigment Epithelium (RPE) sheet transplantation can be a viable therapeutic modality⁶⁴. To this end, we isolated human dermal fibroblasts (HDFs) of a patient with retinitis pigmentosa from skin biopsy sample and reprogrammed it to generate hiPSCs that were further characterized for genetic identity, stemness markers and pluripotency (Supplementary Figure 8A-D). This patient was diagnosed with significant retinal thinning and attenuated photoreceptor cell layer in Optical Coherence Tomography (OCT) due to generation of premature stop codon (p.Trp331Ter) at c.992 stemming from single base substitution from G to A on exon9 of *RPE65* (TGG>TAG) (Figure 2L). The hiPSC line was treated with en31ABE_{8.17d} and the mutation specific sgRNA. We confirmed the successful installation of A>G substitution of the disease mutation in an unsorted population (21% with sg21, NGG PAM). Since, there were adjacent alternate PAMs present at this locus, we also validated successful A>G editing using these PAM sites (13% GGA and 8% AAG) conforming to the earlier observations

of en31 nuclease activity NGG>NGA>NAG. Importantly, two of the clonally expanded iPSC lines derived from the edited cells showed 100% correction of the mutated base (A8) with very high base purity (undetectable bystander edits at A10-12) (Supplementary Figure 8E). Thus, en31ABEmax8.17d can be successfully utilized for robust, precise, nucleobase correction with undetectable bystander edits in therapeutic conditions.

Discussion

In the present study, we have shown the 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)⁶⁵, lipid nanoparticle encapsulated (LNP) mRNA⁶⁶, or split Adeno associated viruses (AAVs)⁶⁷.

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, enFnCas9 variants showed minimal editing at a GUIDE-Seq validated single mismatch off-target (Figure 2C). This is in sharp contrast to both eSpCas9 and SpCas9-HF1 or a naturally occurring specific ortholog ScCas9 reported in literature for the same locus^{12,21,48}. These results show that enFnCas9 variants possibly negotiate off-targets through a different mechanism than high-fidelity SpCas9 proteins⁶⁸.

Furthermore, we compared the activity and specificity of one of the enFnCas9 variants (en31) with the recently developed Superfi-Cas9, a next-generation high-fidelity engineered Cas9 variant which delicately balances the trade-off between DNA interrogation efficiency and specificity of Cas9 enzyme⁶⁹. Using in vitro assays, we observed that Superfi-Cas9 exhibited slower rate of DNA cleavage compared to en31 which might explain the markedly decreased cellular editing by Superfi-Cas9 as reported recently⁷⁰ (Supplementary Figure 9A, B). Next, we evaluated the single mismatch

specificity of Superfi-Cas9 at the PAM-distal stretch of gRNA (serving as a checkpoint of DNA interrogation fidelity which guides nuclease activation) using mismatch walking assay. Interestingly, we found lower PAM-distal single mismatch specificity of Superfi-Cas9 compared with en31, an attribute of superior specificity of enFnCas9 variants observed in our study (Supplementary Figure 9C). Notably, a careful scrutiny of editing efficiency on *FANCF* site2 by one of the enFnCas9 variants, en15, suggests drastically improved on-target to off-target ratio (~45% on-target editing and ~1.4% off-target editing, ~32-fold) of enFnCas9 variant than Superfi-Cas9 (2852 on-target reads and 500 off-target reads, ~6-fold)⁷⁰, a feature indicating superior efficiency and specificity compared to SpCas9 and its engineered variants described in literature so far (Figure 2C).

The ability to extend sgRNA lengths allow wide targeting range and base editing scope across the genome with enFnCas9 variants. It should be noted that Cas9 variants such as SpRY, that are not limited by PAM constraints, have been reported in literature and have been applied in genome editing³⁵. However a recent biophysical study highlights ~1000 fold slower kinetics of SpRY as compared to SpCas9 and the possibility of off-target binding⁷¹. In therapeutic situations where rapid editing and clearance of the editor from cells is desirable, enFnCas9 variants might prove beneficial.

Contrary to the general belief of universal suitability of ABE8 versions of highly processive deaminases with orthogonal Cas effectors^{59,72}, our data suggests otherwise as we did not observe any significant base editing using FnCas9ABE8.17d (Figure 2G, Supplementary Figure 7D). This clearly necessitates Cas9 engineering for its compatibility as base editors. Our results indicate that engineering residues that regulate PAM duplex contacts in the Cas9 backbone can significantly improve editing efficiency without affecting specificity. This strategy can be potentially extended to other orthogonal Cas systems that possess higher intrinsic specificity but have low cellular activity. Furthermore, it is worth noting that our engineering strategy breaks the notion of the trade-off involved between activity and specificity where specificity has been achieved by sacrificing the efficiency of the enzyme even on the canonical PAM upon engineering.

We recommend a combination of en1 for general cellular editing assays when encountering NGG PAM, en15 when super fidelity is required using NGG PAM and en31 for nuclease and base editing on altered PAMs (NRG/NGR). The en31 variant coupled

with sx-gRNA (more than 24-nt long protospacer) is recommended when editing window modulation for base editing is warranted. Notably, compatibility of enFnCas9 variants with sx-gRNAs provides alternatives for enhancing specificity when targeting highly promiscuous repetitive regions in the genome, given the natural rarity of occurrence of perfectly or near-perfectly matched off-targets with longer lengths of matched spacer. Thus, we propose coupling of super specific enFnCas9 variants with sx-gRNAs to provide efficient yet safe editing outcomes in the difficult regions of the genome. Altogether, the enFnCas9 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) when recently reported prime editing by FnCas9 severely suffered from general low activity⁷³.

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 PX458-3xHA-FnCas9 and PX458-3xHA-SpCas9 vectors were done to generate catalytically inactive (dead) double mutants for ChIP-Seq assay. Point mutations on the pET-His6-dFnCas9GFP and PX458-3xHA-FnCas9 (Addgene130969) were done to generate respective en1, en15 and en31 constructs by essentially following the method described earlier⁴. Base editing constructs (PX458-3xHA-FnCas9ABEmax8.17d, PX458-3xHA-en31FnCas9ABEmax8.17d and PX458-3xHA-SpCas9-NG-ABEmax8.17d) were synthesized as gene blocks (GenScript) and subcloned in modified PX458-3xHA-FnCas9 and modified PX458-3xHA-SpCas9 (with a unique EcoRI site generated by site-directed mutagenesis) in AgeI and EcoRI sites. gRNAs were cloned in the BbsI sites of PX458-3xHA-FnCas9, PX458-3xHA-en1FnCas9, PX458-3xHA-en15FnCas9, PX458-3xHA-en31FnCas9, PX458-3xHA-SpCas9-HF1, eSpCas9(1.1) (Addgene 71814), PX458-

3xHA-FnCas9ABEmax8.17d, PX458-3xHA-en31FnCas9ABEmax8.17d and PX458-3xHA-SpCas9-NG-ABEmax8.17d constructs for cellular genome editing and ChIP-Seq assays by essentially following the method described earlier⁷⁴ (Supplementary Figure 1A, 6D). All of the constructs were sequenced before being used. The plasmids reported in the work have been deposited in addgene (Deposit ID 82698). The Superfi-Cas9 plasmid was a kind gift from the lab of David Taylor (University of Texas).

Protein Purification and sgRNA purification

The proteins used in this study were purified as reported previously^{4,31}. 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, Cas14a1 and Superfi-Cas9 were purified essentially by following the purification methods described earlier with some modifications^{66,75}. 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

633 Fisher Scientific) purification as described earlier⁴. IVT sgRNAs were stored at -20°C until
634 further use.

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

636 For the DNA cleavage study, PCR linearized pUC119 plasmid containing the target
637 sequence and the respective PAM (mentioned in respective legends) was used as the
638 substrate for in vitro cleavage experiments. The linearized pUC119 plasmid (50 ng or
639 ~5nM) was incubated at 37°C for 0.5–5 min with the Cas9–sgRNA complex (50 nM) in
640 10 µL of reaction buffer, containing 20 mM HEPES, pH 7.5, 150 mM KCl, 10 mM MgCl₂,
641 1 mM DTT, and 5% glycerol. The reaction was stopped by the addition of a quenching
642 buffer, containing EDTA (20 mM final) and Proteinase K (40 ng). The reaction products
643 were resolved, visualized, and quantified with a MultiNA microchip electrophoresis device
644 (SHIMADZU)³¹.

645 The rest of the IVC assays were done as described earlier⁴. Details of substrates,
646 concentrations, and incubation time are mentioned in the respective figure legends.

647 Quantification of IVC assays from agarose electrophoresis was done using ImageJ and
648 the kinetics data were fitted with a one-phase exponential association curve using
649 Graphpad Prism.

650 **PAM discovery assay**

651 The PAM discovery assays were performed, as previously described³¹. Briefly, a library
652 of pUC119 plasmids containing eight randomized nucleotides downstream of the target
653 sequence was incubated at 37°C for 5 min with the FnCas9–sgRNA complex (50 nM), in
654 50 µL of the reaction buffer. The reactions were quenched by the addition of Proteinase
655 K and then purified using a Wizard DNA Clean-Up System (Promega). The purified DNA
656 samples were amplified for 25 cycles, using primers containing common adapter
657 sequences. After column purification, each PCR product (~5 ng) was subjected to the
658 second round of PCR for 15 cycles, to add custom Illumina TruSeq adapters and sample
659 indices. The sequencing libraries were quantified by qPCR (KAPA Biosystems) and then
660 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 log₂-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 MgCl₂) at 37°C for 10 min. The reaction was inactivated by using 1μl 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 labeled-Chimeric gRNA and en/FnCas9 in a buffer containing 20 mM HEPES, pH7.5, 150 mM KCl, 1 mM DTT, 10% glycerol, 10 mM MgCl₂ and rested for 10 min at RT) for 10 min at 37°C. Wherever active en/FnCas9 was used, MgCl₂ was omitted from the buffer for making it catalytically inactive. After incubation, an 80μl 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.

688 Further background-corrected band intensity values were calculated through a
689 smartphone application (TOPSE)^{36,37}.

690 **Fluorescence assay (dFnCas9)**

691 250nM biotin-labelled PCR amplicons carrying 580 bp long SARS-CoV-2 region with
692 N501Y mutation were used for attaching DNA substrate to the wells of streptavidin coated
693 plate by 10 mins incubation at room temperature. Wells were rinsed thrice with the wash
694 buffer to get rid of the unbound amplicons (25mM Tris-Cl, pH 7.2; 300mM NaCl; 0.1%
695 BSA, 0.05% Tween®-20 Detergent) before using for the binding assay. dFnCas9-GFP
696 RNP complex was pre-assembled in the binding buffer (20 mM HEPES, pH7.5, 150mM
697 KCl, 1mM DTT, 10mM MgCl₂) by incubating 200nM dFnCas9-GFP with 200nM sgRNA
698 for 10 min at room temperature. Reaction was initiated by adding pre-assembled RNP to
699 the wells of 96-well streptavidin coated plate (Thermo Fisher Scientific; Cat 15119) pre-
700 attached with biotin labeled amplicons and incubated at 37°C for 10 mins. Fluorescence
701 was measured using a fluorescence plate reader (λ_{ex} : 485 nm; λ_{em} : 528 nm, transmission
702 gain: optimal) (Tecan Infinite Pro F200).

703 **Fluorescence assay (AaCas12b and Cas14a1)**

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

715

DNA Binding assay

MST was performed as described previously⁴. 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 37⁰ 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 earlier⁷⁶. 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 (ChIP-Seq)

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 1500*g for 10 min 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 1 ml 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 1500*g for 10 min at 4°C. The pellet was resuspended in 1 ml 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 were amplified using forward and reverse primers along with overhang adapter sequences using Phusion High-Fidelity DNA polymerase (ThermoFisher). 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 50 ng 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 sec. Log₂ fold change values were calculated by the $2^{-\Delta\Delta Ct}$

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 bowtie2⁷⁷. Peaks were called over input samples using MACS2⁷⁸ 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' option⁷⁹. Overlap between the dSpCas9 and dFnCas9 ChIP peaks were calculated using bedtools⁸⁰ and plotted as weighted Venn diagrams with the help of Intervene⁸¹.

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 or base editing analysis for each target to remove sequencing read depth bias across the samples. Raw amplicon sequencing reads were subjected to indel frequency estimation for nucleases and A to G nucleotide conversion for adenine base editors using CRISPResso2 v2.0.45⁸² with parameters such as ignoring substitutions for indel analysis and keeping minimum overlap between the forward and reverse read to the 10bp for the both the cases. The base editing data were visualized as balloon plot using a customized R script and the ggpubr package.

Analysis of targetable pathogenic SNPs for Adenine base editors (ABEs)

A variant summary file was downloaded from ClinVar FTP sites in March 2023 to check for targetable G>A variations in ClinVar. Several pre-processing steps are performed to filter the data before evaluating the editable adenine alternate base in ClinVar. Initially, only mutations reported in the 'GRCh38' assembly were kept, and then only single nucleotide mutations with clinical significance for pathogenicity were chosen. Only mutations with the alternate allele 'A' and the reference allele 'G' were then allowed to proceed. The positions of mutations were used to map them onto the human reference genome after pre-processing (GRCh38.p13). Furthermore, the PAM sequences (which differed between base editors) were located near the mutations, as was the base editing window, which differed between base editors. For the calculation of the targetable adenine mutations, pathogenic (G>A) mutations that met the above criteria were considered. Finally, using Bioconductor's circlize package, the targetability of various base editors was plotted.

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 with 1X PBS for 5 minutes each. 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.

Genetic screening and identification of mutation in LCA patients:

As a part of an ongoing genetic screen on a large case series, the proband (LCA27) was clinically diagnosed as a candidate for Leber congenital amaurosis (LCA), based on the clinical characteristics described earlier⁸³. The genomic DNA was extracted from the blood sample of all patients and were genotyped for a select set of LCA candidate genes, by targeted amplification of all exons, along with the flanking intronic regions by PCR, followed by Sanger sequencing. The proband, LCA27 was found to carry a pathogenic, nonsense mutation within the exon 9 of the human *RPE65* gene, c.992 G>A, p.Trp331Ter (TGG>TAG), which resulted in a stop codon, premature translational termination and formation of a C-terminally truncated and non-functional protein (personal communication). This patient volunteer was counselled and recruited for skin biopsy collection, to generate human dermal fibroblast cultures, which were further reprogrammed to generate the patient-specific iPSC line, LVIP-EPI-LCA2-2.

HDF culture, reprogramming and generation of hiPSC lines:

The study was approved by the Institutional Ethics Committee (IEC), Institutional Bio-Safety Committee (IBSC) and Institutional Committee for Stem Cell Research (IC-SCR) at the LV Prasad Eye Institute, Hyderabad. Full-thickness skin biopsy of about 2X2 mm was taken from the retro-auricular surface of the patient volunteer with the informed consent. The tissue was cut into small pieces and digested with Collagenase (1 mg/mL) for 3 hrs, diluted with DPBS and then centrifuged at 200 g for 3 minutes. The pellet containing the released cells, along with the partially digested tissues are suspended in 5 mL of HDF culture medium (Gibco™ Medium 106, with the addition of Low Serum Growth Supplement- LSGS, Thermo Fisher Scientific Inc.) and cultured in T25 flasks at 37°C with 5% CO₂ for 3-5 days or until the culture reaches 80-90% confluence. The human dermal fibroblast (HDFs) cultures are further passaged using 1X TrypLE at 1:3 split ratio. For HDF reprogramming, a cocktail of three episomal plasmids encoding the human OCT3/4, SOX2, KLF4, L-MYC and LIN28 (Addgene Plasmid # 27077, 27078,

27080)⁸⁴ were nucleofected into passage 3 HDFs, using the P2 nucleofection kit and EO-114 program of the 4D-Nucleofector, X Unit System, as per the manufacturer's instructions (Lonza, Basel, Switzerland). The cells were then plated onto vitronectin coated cell culture dishes and cultured in HDF medium for the first 2 days. On the third day, the cultures were shifted to hiPSC growth medium (Gibco™ Essential 8™ Medium, Thermo Fisher Scientific Inc.) and maintained till day 30, with regular media changes on alternate days. Well reprogrammed hiPSC colonies with distinct margins emerged at around D20-D25. Individual colonies are manually picked by gentle nudging using a P2 tip and are mildly triturated to form 5-10 cell clusters and further cultured on vitronectin coated 12-well plates. Well expanding hiPSC clones are further passaged using ReLeSR (STEMCELL Technologies Inc.) at 1:6 split ratio till passage 10, for further molecular characterizations to confirm stemness, pluripotency, genetic identity, genomic stability and for the loss of episomal plasmids. The cells of stable patient-specific iPSC lines at passage 12 were used in base editing experiments. The stable patient-specific iPSC line (LVIP-EPI-LCA2-2) was characterized for the genetic identity (patient-specific mutation, STR profiling), stemness marker expression (by RT-PCR, Immunocytochemistry) and pluripotency (by EB formation and three lineage marker expression by RT-PCR).

Base editing and mutation correction in patient-specific iPSCs:

The patient-specific iPSC culture was harvested using ReLeSR (STEMCELL Technologies Inc.) to prepare single cell suspensions and the cell count was taken using a haemocytometer. About 1 million cells were prepared for the nucleofection of en31-ABE plasmids with cloned mutation specific gRNA using the P3 nucleofection kit and CA-137 program of the 4D-Nucleofector, X Unit System, as per the manufacturer's instructions (Lonza, Basel, Switzerland). After nucleofection, the cells were suspended in hiPSC growth medium (Gibco™ Essential 8™ Medium, Thermo Fisher Scientific Inc.), with the addition of Gibco™ RevitaCell™ Supplement to support single cell culture and were plated onto two wells of a vitronectin coated 6-well plate. After 18-24 hours, the cultures are shifted to the complete Essential 8 medium, with regular media changes on alternate days and are maintained in a controlled condition incubator at 37°C, with 5% CO₂ supply. At 72 hrs post nucleofection or when the cultures reach 70-80% confluence, one well of

the edited cell pool was harvested using ReLeSR and cryopreserved using Gibco™ Synth-a-Freeze™ Cryopreservation Medium. The second well was used for DNA isolation to evaluate edit efficiencies, either by T7 endonuclease cleavage assay or by deep sequencing of PCR amplicons of the targeted genomic region. Once the presence of edits is confirmed, a vial of cryopreserved cell pool was revived and plated at clonal densities (<1000 cells per 100 mm dish) on vitronectin coated plates and cultured in Essential 8 Medium with the addition of RevitaCell Supplement during the initial 18-24 hrs of single cell passaging. Well grown single cell clones (N=13), with clear margins are manually picked and individually plated onto vitronectin-coated 12-well plates and further cultured for clonal expansion. Subsequent clonal passaging was done using ReLeSR for harvesting the cells and replica plates were maintained, to harvest the cells for expansion, cryopreservation, and genomic DNA isolation for editing analysis.

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. S.A. designed and performed the protein engineering experiments with inputs from S.H., H.N. and O.N. A.H.A designed and implemented bioinformatics tools for design, validation and execution of enFnCas9-based diagnostics and editing experiments with inputs from S.A. and D.C. R.R. performed HDR experiments with enFnCas9 variants. S.S. performed comparison of single mismatch specificity with Cas12/14 proteins, cloning of guides in base editing constructs and assisted in sequencing experiments. M.K., R.P., and S.G. contributed to validating the FELUDA/RAY platform with enFnCas9 proteins. C.A. performed *in vitro* assays for comparison of Superfi-Cas9 and en31. D.P. contributed to

a part of cellular editing experiments using enFnCas9. A.R. purified the Superfi-Cas9 protein. S.A. performed ChIP assay and all the sequencing experiments. S.A., Su. M, Sa. M, V.K.P and I.M. performed and validated enFnCas9 activity experiments in HEK293T, ARPE-19 and iPSC lines. S.J. and I.M. contributed to the LCA patient resource. S.A. and D.C. drafted the manuscript with inputs from all other authors.

Acknowledgments

We thank all members of Chakraborty, Maiti, Nureki, Mariappan 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 Lady Tata Young Investigator award (GAP0198) to D.C., Department of Biotechnology (I.M.), (BT/PR32404/MED/30/2136/2019) to I.M. and Senior Research Fellowships from ICMR (Su.M.) and CSIR (V.K.P.), Government of India.

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.

1013 **Supplementary Figures**

1014

1015

1016

1017

1018

1019

1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

1035

1036

1037

1038

1039

1040

1041

1042

1043

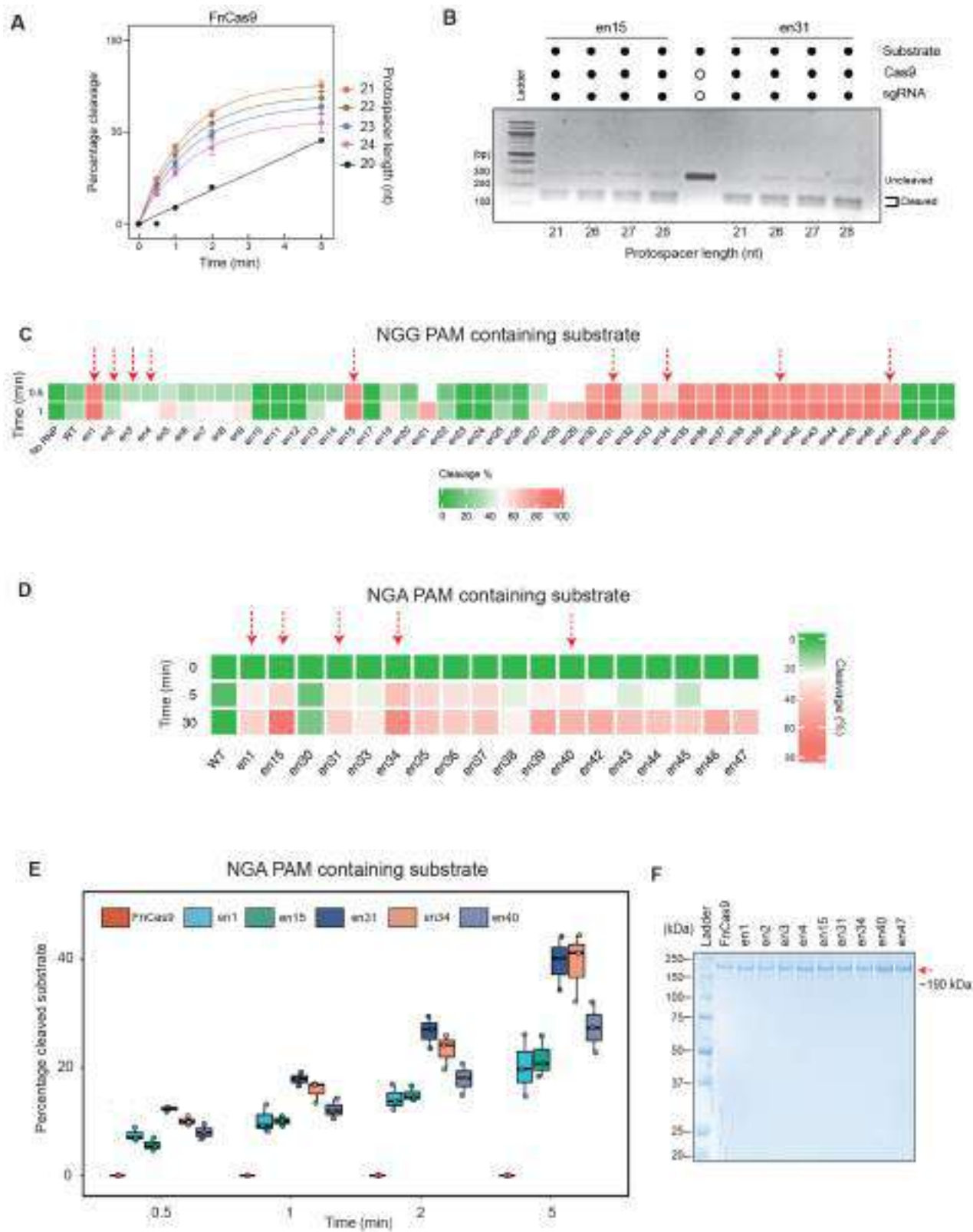
Supplementary Figure 1:

A. Schematic showing the ChIP-Seq assay 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 assay on *c-MYC* locus.

C. Circos plot showing the sgRNA homology dependent off-target peaks (calculated upto 6 mismatches, MM) with NGG PAM for dSpCas9 and dFnCas9. Chromosomes are arranged radially and shown in megabase (Mb) for which sequence reads were captured for both the dCas9s. On-target peaks for both the dCas9s are shown and marked in red.

Supplementary Figure 2. *In vitro* assays for characterization of cleavage activity of enFnCas9 variants



Supplementary Figure 2:

A. Fitted line plot showing the rates of *in vitro* DNA cleavage of FnCas9 (expressed in percentage, Y-axis) as a function of time (expressed in minutes, on X-axis) upon programmed with variable length of sgRNAs having protospacer length ranging from 20 to 24 nucleotides (nt).

B. Agarose gel image of the *in vitro* cleavage assay showing the DNA cleavage efficiency of en15 and en31 upon programmed with variable length of sgRNAs having protospacer length ranging from 21, 26 to 28 nucleotides (nt). 25nM DNA substrate was incubated with 250nM RNP for 15 minutes. Reaction was quenched and treated with Proteinase K before running on 2% agarose gel. Solid dot and open dot represent the presence and absence of components in the assay respectively.

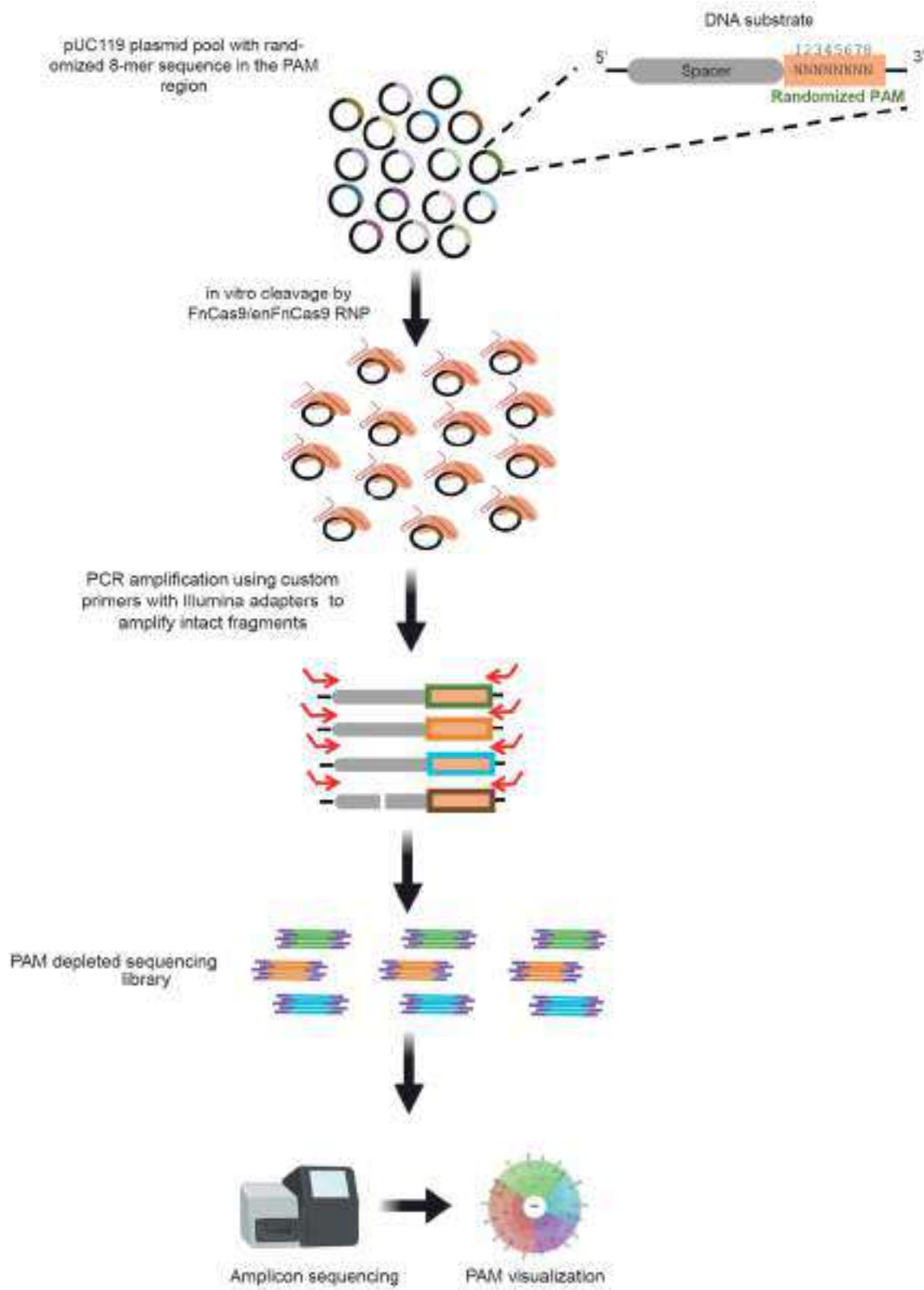
C. Heat map showing the *in vitro* cleavage screening assay of FnCas9 (WT) and enFnCas9 variants using GGG PAM containing DNA substrate expressed as cleavage percentage for 0.5 and 1 minutes. Red dotted arrows indicate the subset of nine enFnCas9 variants selected for studying enzymatic activity as a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

D. Heat map showing the *in vitro* cleavage screening assay of a subset of FnCas9 (WT) and enFnCas9 variants using GGA PAM containing DNA substrate expressed as cleavage percentage for 5 and 30 minutes. Red dotted arrows indicate the subset of five enFnCas9 variants selected for studying enzymatic activity as a function of time.

E. 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).

F. Coomassie stained SDS-PAGE gel showing purified FnCas9 and enFnCas9 protein variants used in the study.

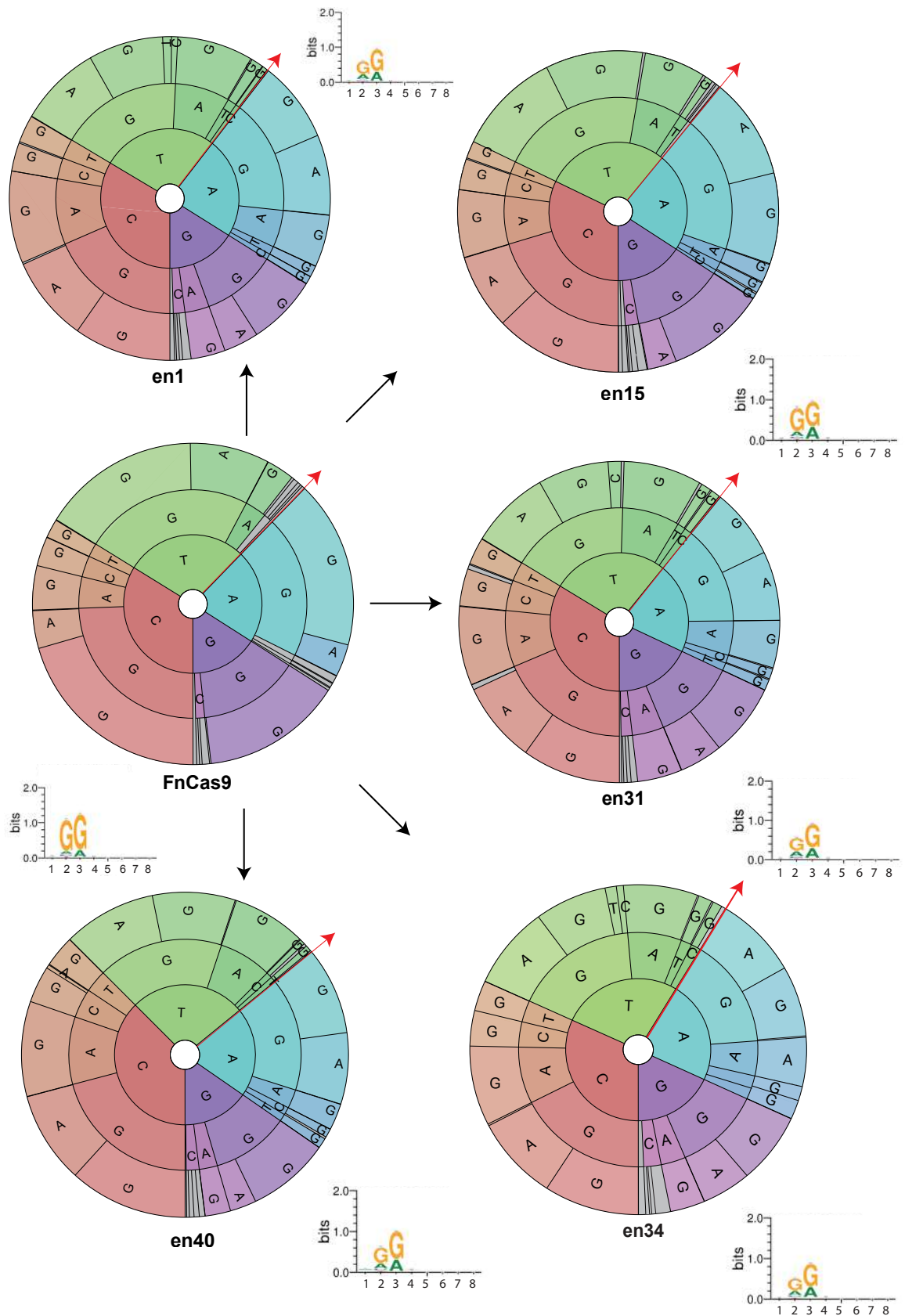
Supplementary Figure 3. Pipeline of PAM discovery assay to comprehensively characterize PAM preferences of enFnCas9 variants



Supplementary Figure 3:

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

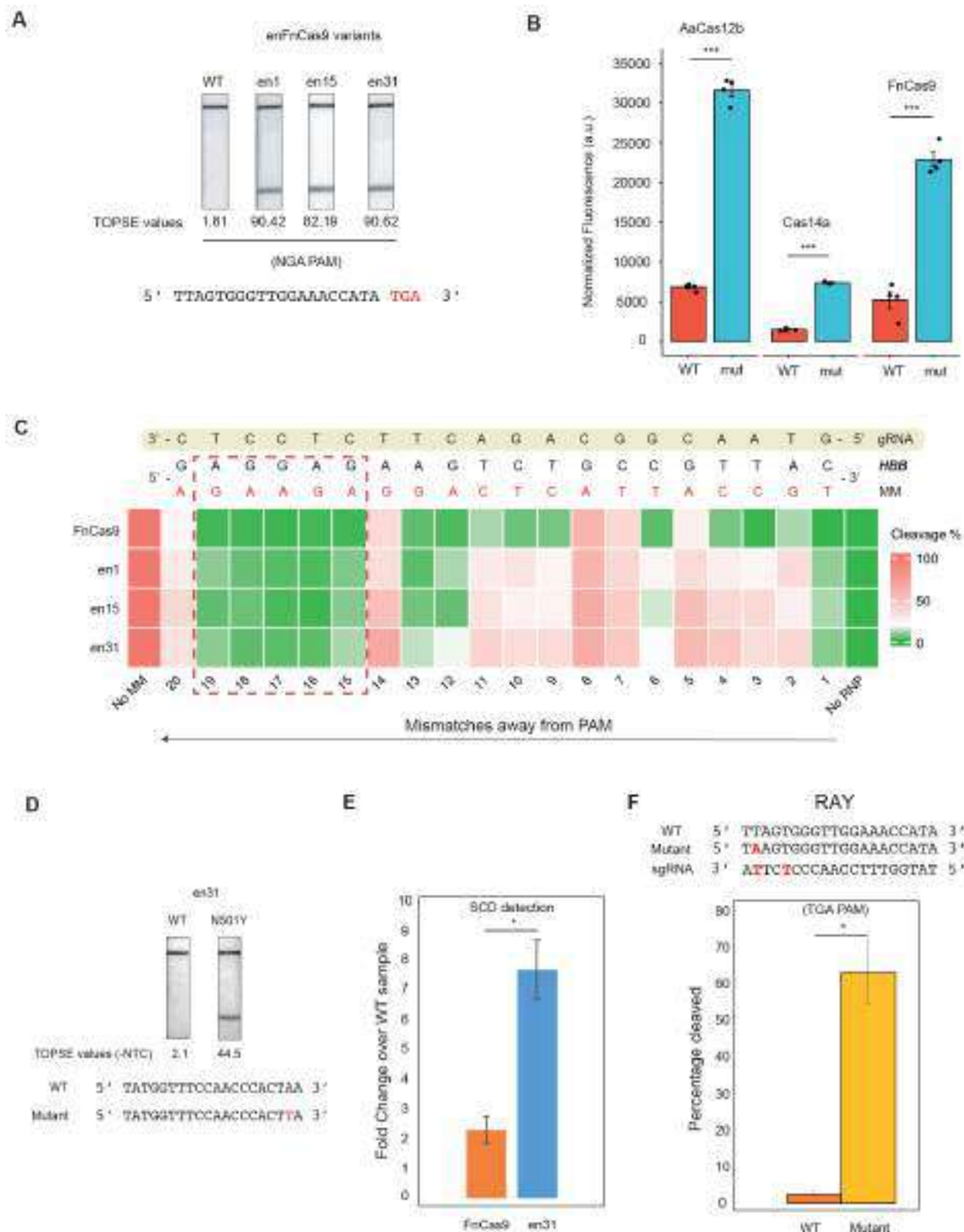
Supplementary Figure 1. Characterization of PAM preferences of enFnCas9 variants using PAM discovery assay



Supplementary Figure 4:

The PAM wheels and sequence logos showing the results obtained from PAM discovery assay for FnCas9, en1, 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. *In vitro* characterization of enFnCas9 variants for mismatch sensitivity and its potential in CRISPRDx platforms



Supplementary Figure 5:

A. Outcome of lateral flow assay (LFA) on a TGA PAM containing substrate 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 SEM (n=3 independent experiments). Student's unpaired *t*-test p-values are represented for *** <0.001.

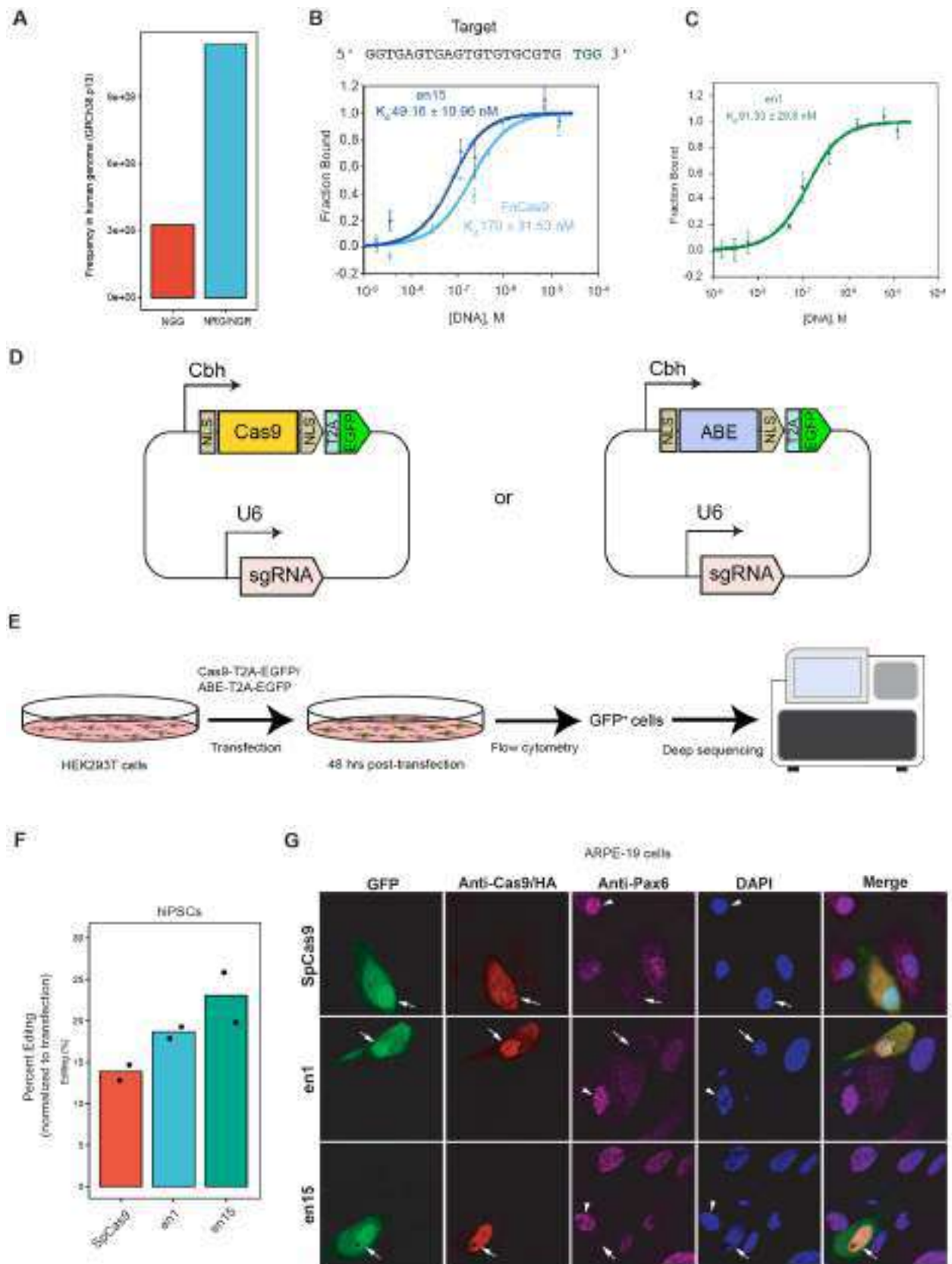
C. Heat map showing the *in vitro* cleavage assay outcome (expressed in percentage of cleavage) of FnCas9, en1, en15 and en31 on HBB and its mutant substrates. Each substrate is harboring a single mutation (position is indicated by counting away from PAM) across the target length. Mismatch containing base is indicated in red while other unaltered bases are shown in black. 25nM of DNA substrates was incubated with respective 100nM RNPs for 15 min at 37°C. Reaction was quenched and treated with Proteinase K before running on 2% agarose gel. Dotted box shown in red represents PAM-distal sites showing minimum mismatch tolerance.

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. Bar plot showing the SCD mutation discrimination by en31 and FnCas9 using FELUDA. Data is shown as a fold enrichment over WT sample (n=3 independent experiments). Student's unpaired *t*-test p-value is represented for * <0.05.

F. Bar plot showing outcome of *in vitro* cleavage based detection of N501Y (with an NGA PAM) by en31 (n=3 independent experiments). WT and mutant substrate sequences are shown along with sequence of gRNA. Mismatches are indicated in red on the respective sequences. Student's unpaired *t*-test p-value is represented for * <0.05.

Supplementary Figure 6. Broad targeting range coupled with improved DNA binding underscore its utility in efficient targeting of human genome



Supplementary Figure 8:

A. Bar plot showing the calculated frequency of NGG PAM and NRG/NGR PAM in the human genome (GRCh38.p13).

B. 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).

C. 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).

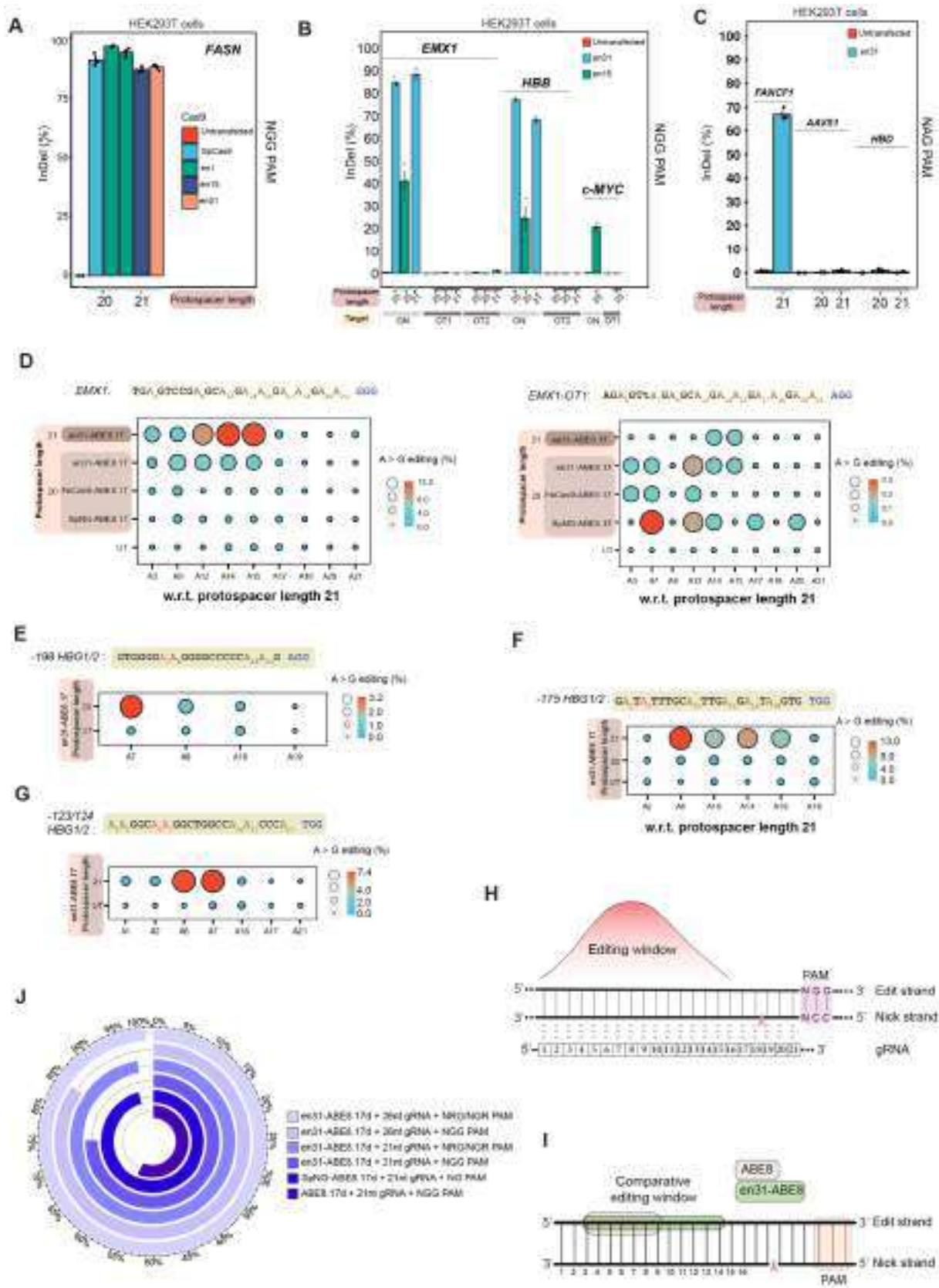
D. Schematic showing the architecture of all-in-one plasmids expressing nucleases or base editors as T2A-EGFP fusion with gRNA cloning sites.

E. Schematic showing the experimental pipeline used for genome editing efficiency evaluation.

F. Bar plot showing the InDel percentage (editing) normalized to the transfection efficiency as obtained from T7E1 assay upon targeting *PAX6* locus by SpCas9, en1 and en15 in hiPSC cells. Values represent the mean of n=2 independent biological replicates.

G. Immunofluorescence image showing the 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 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 cells without Cas9 expression are marked by arrowheads.

Supplementary Figure 7. Characterization of enFnCas9 variants as nucleases and base editors in human cells



Supplementary Figure 7:

A. Bar plot showing the InDel events (expressed in percentage) plotted on the Y-axis as obtained from amplicon sequencing upon targeting NGG PAM containing FASN locus by SpCas9, en1, en15 and en31 with either sgRNA containing 20-nt protospacer or sgRNA containing 21-nt protospacer in HEK293T cells. Untransfected cells were used as control. Error bars represent SEM of n=3 independent biological replicates with individual values shown as dots.

B. Bar plot showing the InDel events (expressed in percentage) plotted on the Y-axis as obtained from amplicon sequencing upon targeting NGG PAM containing *EMX1*, *HBB* and *c-MYC* loci by en31 and en15 with either sgRNA containing 20-nt protospacer (g20) or sgRNA containing 21-nt protospacer (g21) in HEK293T cells. Untransfected cells were used as control. Error bars represent SEM of n=3 independent biological replicates with individual values shown as dots.

C. Bar plot showing the InDel events (expressed in percentage) plotted on the Y-axis as obtained from amplicon sequencing upon targeting NAG PAM containing *FANCF1* site2, *AAVS1* and *HBD* loci by en31 with either sgRNA containing 20-nt protospacer (g20) or sgRNA containing 21-nt protospacer (g21) in HEK293T cells. Untransfected cells were used as control. Error bars represent SEM of n=3 independent biological replicates with individual values shown as dots.

D. (Left) Ballon plot showing the A to G editing events (expressed in percentage) as obtained from amplicon sequencing upon targeting *EMX1* locus by FnCas9-ABE8.17d, en31ABE8.17d and SpNG-ABE8.17d with either sgRNA containing 20-nt protospacer (g20) or sgRNA containing 21-nt protospacer (g21) in HEK293T cells. Target bases (As) for adenine base editing are on the X-axis and numbered w.r.t. g21. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates. (Right) Ballon plot showing the A to G editing events (expressed in percentage) as obtained from amplicon sequencing for the validated off-target, *EMX1*-OT1 upon targeting *EMX1* locus by FnCas9-ABE8.17d, en31ABE8.17d and SpNG-ABE8.17d with either sgRNA containing 20-nt protospacer (g20) or sgRNA containing 21-nt protospacer (g21) in HEK293T cells. Mismatches on the protospacer are

shown in small caps. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates.

E. Ballon plot showing the A to G editing events (expressed in percentage) as obtained from amplicon sequencing upon targeting -198 site of *HBG1/2* promoter by en31ABE8.17d in HEK293T cells. Target bases (As) for adenine base editing are numbered on the X-axis with target site (A7) indicated in red. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates.

F. Ballon plot showing the A to G editing events (expressed in percentage) as obtained from amplicon sequencing upon targeting -175 site of *HBG1/2* promoter by en31ABE8.17d with either sgRNA containing 20-nt protospacer (g20) or sgRNA containing 21-nt protospacer (g21) in HEK293T cells. Target bases (As) for adenine base editing are numbered on the X-axis w.r.t. g21 with target site (A4) indicated in red. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates.

G. Ballon plot showing the A to G editing events (expressed in percentage as obtained from amplicon sequencing upon targeting -123/124 sites of *HBG1/2* promoter by en31ABE8.17d in HEK293T cells. Target bases (As) for adenine base editing are numbered on the X-axis with target sites (A6-A7) indicated in red. The area of the dots is proportional to the magnitude of editing (numerical values). Values represent the mean of n=3 independent biological replicates.

H. Schematic showing the adenine base editing window of en31ABE8.17d. Apex of the bell curve indicates the optimal editing space. Numbers indicate the positions across the protospacer of the gRNA. The location of the nick by en31ABE8.17d (nD11A) is indicated by a scissor on the nick strand. PAM containing strand hosts the target bases for base editor and represented as edit strand.

I. Schematic showing the comparative base editing window between ABE8.17d (shaded in pink) and en31ABE8.17d (shaded in green). Numbers indicate the positions across the protospacer of the gRNA. The location of the nick by ABE8.17d (D10A) and en31ABE8.17d (nD11A) is indicated by a scissor on the nick strand. PAM containing strand hosts the target bases for base editor and represented as edit strand.

1504 **J.** Circular bar plot showing the theoretically targetable pathogenic SNPs (expressed in
1505 percentage) as identified from the ClinVar for SpCas9ABE8.17d (NGG PAM), SpNG-
1506 ABE8.17d (NG PAM) and en31ABE8.17d (NGG/NRG/NGR PAMs) with respective
1507 gRNAs containing specified protospacer length.

1508

1509

Su



1539

1540 **A.** Ballon plot showing the modulation of base editing window by en31ABE8.17d
1541 expressed in percentage of A to G editing plotted on Y-axis using gRNAs with extended
1542 protospacers (g20 to g25) as obtained from amplicon sequencing upon targeting
1543 *SERPINI1* locus against TGA PAM. Target bases (As) for adenine base editing are on
1544 the X-axis and numbered w.r.t. g20. Values represent the mean of n=3 independent
1545 biological replicates.

1546 **B.** Schematic showing the sliding of the base editing window from primary window (shown
1547 in red) to secondary window (shown in blue) by extended gRNAs and indicated by higher
1548 efficiency of A to G editing on *SERPINI1* locus with TGA PAM.

1549 **C.** Images showing the skin explant culture and human dermal fibroblast (HDF) expansion
1550 (right) isolated from patient diagnosed with Leber congenital amaurosis (LCA) followed
1551 by reprogramming of HDFs into iPSC clones by episomal plasmids (left).

1552 **D.** Image iPSC colony stained for pluripotency markers (Oct4, Nanog, Sox2 and Klf4) and
1553 merged with DAPI (in blue).

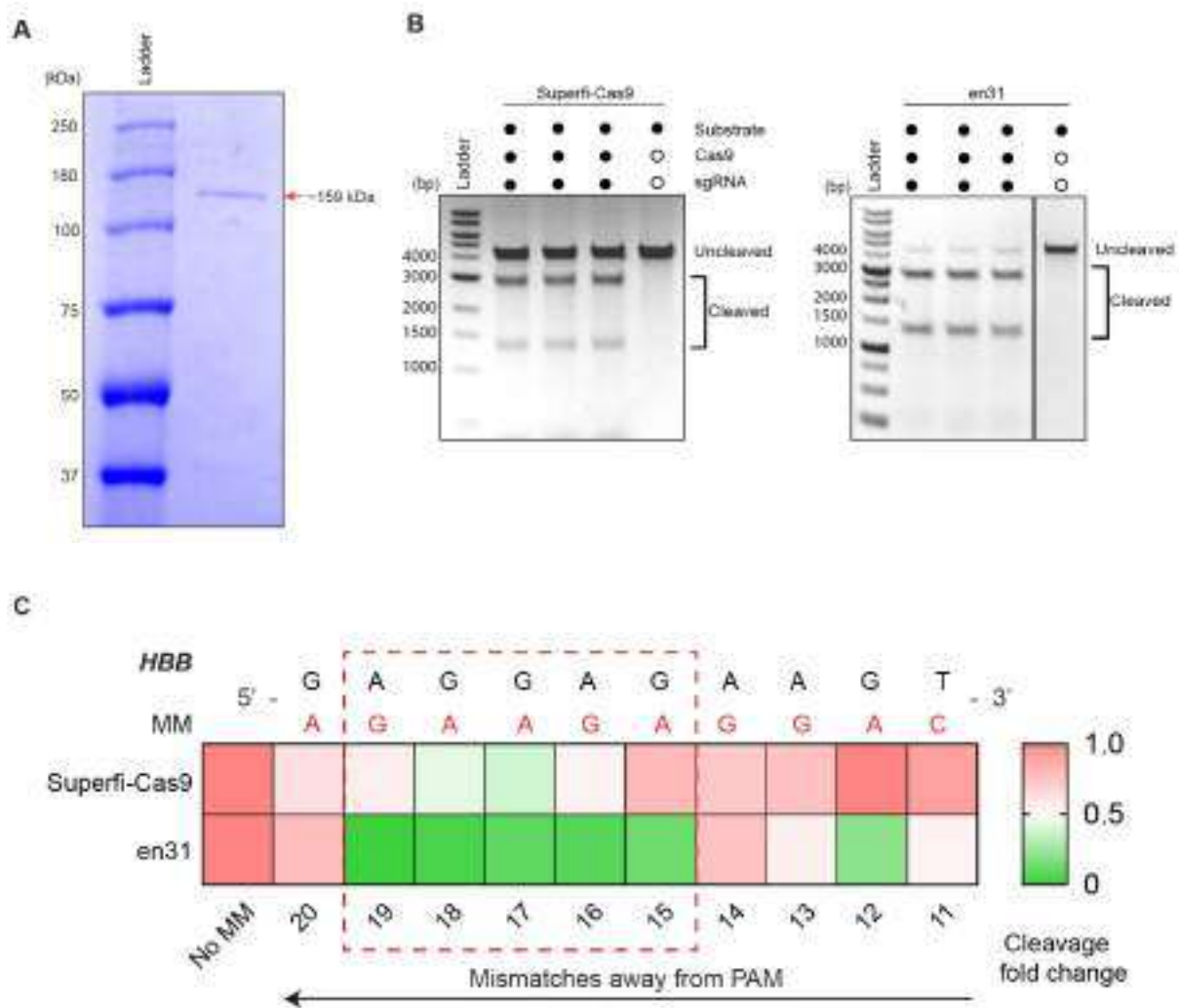
1554 **E.** Agarose gel image showing the RT-PCR results for expression of stemness markers
1555 (hOCT4, hSOX2, hNANOG and hTERT) of patient-specific iPSC line. GAPDH was used
1556 as internal control.

1557 **F.** Agarose gel image showing the RT-PCR results for expression of trilineage markers
1558 upon embryoid body (EB) formation from patient-specific iPSC line. Undifferentiated
1559 iPSCs (UD) were used as a control for trilineage markers. eEF1 α was used as an internal
1560 control.

1561 **G.** Bar plot showing the efficiency of A to G editing of en31ABE8.17d-treated patient-
1562 specific iPSC clones for correction of the disease mutation (in red). Target bases (As)
1563 were shown on the X-axis and base editing efficiency expressed as A to G editing
1564 percentage was plotted on the Y-axis as obtained from amplicon sequencing. A8 is the
1565 patient mutation and A10-A12 are bystander bases for the base editor across the
1566 protospacer. Error bars represent SEM of n=3 independent sequencing replicates with
1567 individual values shown as dots.

1568

Supplementary Figure 9: Comparison between Superfi-Cas9 and enFnCas9 variant for evaluating cleavage activity and specificity of DNA interrogation



Supplementary Figure 9:

A. Coomassie stained SDS-PAGE gel image of purified Superfi-Cas9.

B. Agarose gel image of *in vitro* cleavage assay showing the efficiency of DNA cleavage as a function of time of Superfi-Cas9 and en31 on *HBB* substrate cloned in TOPO vector. 5nM linearized substrate (4200 bp) was incubated with respective 50nM ribonucleoproteins (RNPs) for 10 minutes. Reaction was quenched and treated with Proteinase K before running on 0.8% gel. Experiments were done in triplicates. Solid dot and open dot represent the presence and absence of components in the assay respectively.

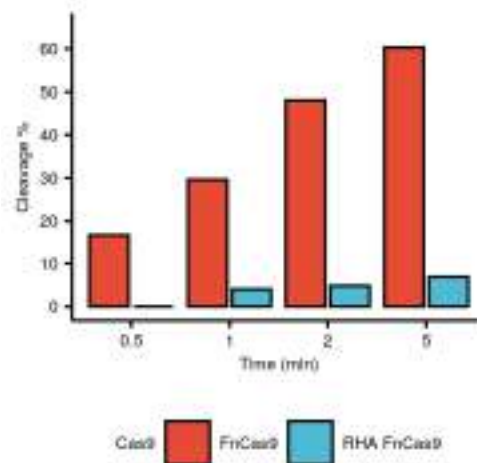
C. Heat map showing the *in vitro* cleavage-based mismatch (MM) walking assay outcome between Superfi-Cas9 and en31 on *HBB* and its mutant substrates, each carrying a single mismatch. Reaction was quenched and treated with Proteinase K before running on 2% agarose gel. Mismatches are indicated in red and counted away from PAM as indicated on the x-axis. Dotted box shown in red represents PAM-distal sites showing minimum mismatch tolerance. Data is shown as cleavage fold change normalized to *HBB* substrate (No MM).

Supplementary Note 1

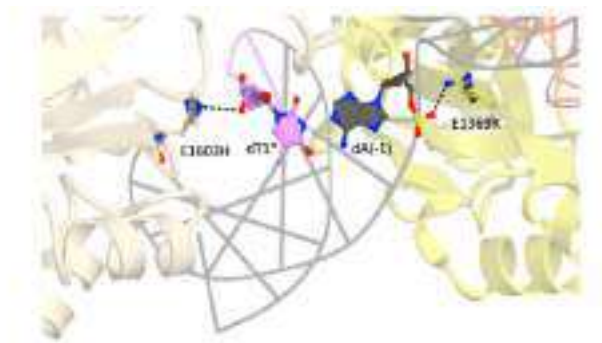
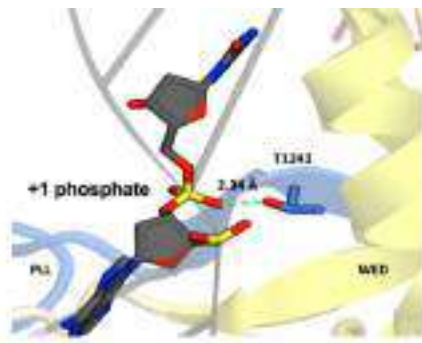
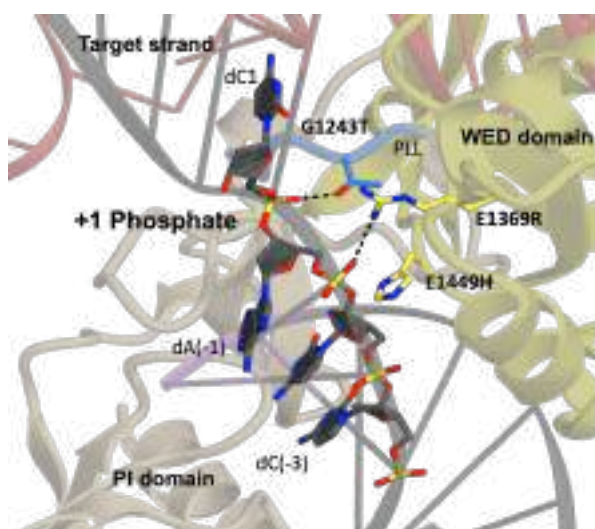
Details of structure-guided engineering

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 nontarget 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'-NGA3' PAM by FnCas9. Towards developing a PAM-relaxed FnCas9-NG 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 PAM³. 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^{2,85}. 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: Comparative evaluation of *in vitro* cleavage activity between FnCas9 and RHA-FnCas9 on TGG 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 G1243T substitution on phosphate lock loop (PLL) (en4) with +1 Phosphate group. **C)** Interactions of en31 (E1369R/E1603H/G1243T) with PAM duplex and PLL loop.

1678 **References**

- 1679 1. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target
1680 DNA. *Cell* 156, 935–949 (2014).
- 1681 2. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-
1682 dependent target DNA recognition by the Cas9 endonuclease. *Nature* 513, 569–573
1683 (2014).
- 1684 3. Hirano, H. et al. Structure and Engineering of *Francisella novicida* Cas9. *Cell* 164,
1685 950–961 (2016).
- 1686 4. Acharya, S. et al. Cas9 interrogates genomic DNA with very high specificity and
1687 can be used for mammalian genome editing. *Proc. Natl. Acad. Sci. U. S. A.* 116, 20959–
1688 20968 (2019).
- 1689 5. Kim, N. et al. Prediction of the sequence-specific cleavage activity of Cas9
1690 variants. *Nat. Biotechnol.* 38, 1328–1336 (2020).
- 1691 6. Liu, M.-S. et al. Engineered CRISPR/Cas9 enzymes improve discrimination by
1692 slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* 11, 3576 (2020).
- 1693 7. Kim, Y.-H. et al. Sniper2L is a high-fidelity Cas9 variant with high activity. *Nat.*
1694 *Chem. Biol.* (2023) doi:10.1038/s41589-023-01279-5.
- 1695 8. Liu, J.-J. et al. CasX enzymes comprise a distinct family of RNA-guided genome
1696 editors. *Nature* 566, 218–223 (2019).
- 1697 9. Pausch, P. et al. CRISPR-CasΦ from huge phages is a hypercompact genome
1698 editor. *Science* 369, 333–337 (2020).
- 1699 10. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-
1700 Cas system. *Cell* 163, 759–771 (2015).
- 1701 11. Karvelis, T. et al. PAM recognition by miniature CRISPR–Cas12f nucleases

1702 triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Research*
1703 vol. 48 5016–5023 Preprint at <https://doi.org/10.1093/nar/gkaa208> (2020).

1704 12. Chatterjee, P., Jakimo, N. & Jacobson, J. M. Minimal PAM specificity of a highly
1705 similar SpCas9 ortholog. *Sci Adv* 4, eaau0766 (2018).

1706 13. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free
1707 nuclease. *Nat. Commun.* 12, 555 (2021).

1708 14. Schmidt, M. J. et al. Improved CRISPR genome editing using small highly active
1709 and specific engineered RNA-guided nucleases. *Nat. Commun.* 12, 4219 (2021).

1710 15. Kim, D. Y. et al. Efficient CRISPR editing with a hypercompact Cas12f1 and
1711 engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* 40, 94–
1712 102 (2022).

1713 16. Wu, Z. et al. Programmed genome editing by a miniature CRISPR-Cas12f
1714 nuclease. *Nat. Chem. Biol.* 17, 1132–1138 (2021).

1715 17. Xu, X. et al. Engineered miniature CRISPR-Cas system for mammalian genome
1716 regulation and editing. *Mol. Cell* 81, 4333–4345.e4 (2021).

1717 18. Newby, G. A. et al. Base editing of haematopoietic stem cells rescues sickle cell
1718 disease in mice. *Nature* 595, 295–302 (2021).

1719 19. Chen, F. et al. Targeted activation of diverse CRISPR-Cas systems for mammalian
1720 genome editing via proximal CRISPR targeting. *Nat. Commun.* 8, 14958 (2017).

1721 20. Tycko, J. et al. Mitigation of off-target toxicity in CRISPR-Cas9 screens for
1722 essential non-coding elements. *Nat. Commun.* 10, 4063 (2019).

1723 21. Chen, J. S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting
1724 accuracy. *Nature* 550, 407–410 (2017).

1725 22. Jones, S. K., Jr et al. Massively parallel kinetic profiling of natural and engineered
1726 CRISPR nucleases. *Nat. Biotechnol.* 39, 84–93 (2021).

- 1727 23. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in
1728 mammalian cells. *Nat. Biotechnol.* 32, 670–676 (2014).
- 1729 24. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis
1730 reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat.*
1731 *Biotechnol.* 32, 677–683 (2014).
- 1732 25. O’Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F. & Segal, D. J. A genome-
1733 wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture.
1734 *Nucleic Acids Res.* 43, 3389–3404 (2015).
- 1735 26. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA
1736 interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507, 62–67
1737 (2014).
- 1738 27. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control
1739 of DNA target cleavage by CRISPR-Cas9. *Nature* 527, 110–113 (2015).
- 1740 28. Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A
1741 conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci*
1742 *Adv* 3, eaao0027 (2017).
- 1743 29. Okafor, I. C. et al. Single molecule analysis of effects of non-canonical guide RNAs
1744 and specificity-enhancing mutations on Cas9-induced DNA unwinding. *Nucleic Acids*
1745 *Res.* 47, 11880–11888 (2019).
- 1746 30. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA Unwinding Is the Primary
1747 Determinant of CRISPR-Cas9 Activity. *Cell Rep.* 22, 359–371 (2018).
- 1748 31. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting
1749 space. *Science* 361, 1259–1262 (2018).
- 1750 32. Hirano, S. et al. Structural basis for the promiscuous PAM recognition by
1751 *Corynebacterium diphtheriae* Cas9. *Nat. Commun.* 10, 1968 (2019).

- 1752 33. Hirano, S., Nishimasu, H., Ishitani, R. & Nureki, O. Structural Basis for the Altered
1753 PAM Specificities of Engineered CRISPR-Cas9. *Mol. Cell* 61, 886–894 (2016).
- 1754 34. Anders, C., Bargsten, K. & Jinek, M. Structural Plasticity of PAM Recognition by
1755 Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol. Cell* 61, 895–902
1756 (2016).
- 1757 35. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained
1758 genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* 368,
1759 290–296 (2020).
- 1760 36. Azhar, M. et al. Rapid and accurate nucleobase detection using FnCas9 and its
1761 application in COVID-19 diagnosis. *Biosensors and Bioelectronics* 183, 113207 (2021).
- 1762 37. Kumar, M. et al. FnCas9-based CRISPR diagnostic for rapid and accurate
1763 detection of major SARS-CoV-2 variants on a paper strip. *Elife* 10, (2021).
- 1764 38. Chen, J. S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-
1765 stranded DNase activity. *Science* 360, 436–439 (2018).
- 1766 39. Harrington, L. B. et al. Programmed DNA destruction by miniature CRISPR-Cas14
1767 enzymes. *Science* 362, 839–842 (2018).
- 1768 40. Gootenberg, J. S. et al. Nucleic acid detection with CRISPR-Cas13a/C2c2.
1769 *Science* 356, 438–442 (2017).
- 1770 41. Kleinstiver, B. P. et al. Publisher Correction: Engineered CRISPR-Cas12a variants
1771 with increased activities and improved targeting ranges for gene, epigenetic and base
1772 editing. *Nat. Biotechnol.* 38, 901 (2020).
- 1773 42. Kleinstiver, B. P. et al. Engineered CRISPR-Cas12a variants with increased
1774 activities and improved targeting ranges for gene, epigenetic and base editing. *Nat.*
1775 *Biotechnol.* 37, 276–282 (2019).
- 1776 43. Wei, Y. et al. Indiscriminate ssDNA cleavage activity of CRISPR-Cas12a induces

1777 no detectable off-target effects in mouse embryos. *Protein Cell* 12, 741–745 (2021).

1778 44. Murugan, K., Seetharam, A. S., Severin, A. J. & Sashital, D. G. CRISPR-Cas12a
1779 has widespread off-target and dsDNA-nicking effects. *J. Biol. Chem.* 295, 5538–5553
1780 (2020).

1781 45. Kleinstiver, B. P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable
1782 genome-wide off-target effects. *Nature* 529, 490–495 (2016).

1783 46. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved
1784 specificity. *Science* 351, 84–88 (2016).

1785 47. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage
1786 by CRISPR-Cas nucleases. *Nat. Biotechnol.* 33, 187–197 (2015).

1787 48. Casini, A. et al. A highly specific SpCas9 variant is identified by in vivo screening
1788 in yeast. *Nat. Biotechnol.* 36, 265–271 (2018).

1789 49. Kleinstiver, B. P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM
1790 specificities. *Nature* 523, 481–485 (2015).

1791 50. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9
1792 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 24, 927–930
1793 (2018).

1794 51. Ihry, R. J. et al. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem
1795 cells. *Nat. Med.* 24, 939–946 (2018).

1796 52. Leibowitz, M. L. et al. Chromothripsis as an on-target consequence of CRISPR-
1797 Cas9 genome editing. *Nat. Genet.* 53, 895–905 (2021).

1798 53. Cullot, G. et al. CRISPR-Cas9 genome editing induces megabase-scale
1799 chromosomal truncations. *Nat. Commun.* 10, 1136 (2019).

1800 54. Zuccaro, M. V. et al. Allele-Specific Chromosome Removal after Cas9 Cleavage
1801 in Human Embryos. *Cell* 183, 1650–1664.e15 (2020).

1802 55. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas
1803 nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* 38, 824–844
1804 (2020).

1805 56. Kim, D., Kang, B.-C. & Kim, J.-S. Identifying genome-wide off-target sites of
1806 CRISPR RNA-guided nucleases and deaminases with Digenome-seq. *Nat. Protoc.* 16,
1807 1170–1192 (2021).

1808 57. Zhao, Z., Shang, P., Mohanraju, P. & Geijsen, N. Prime editing: advances and
1809 therapeutic applications. *Trends Biotechnol.* (2023) doi:10.1016/j.tibtech.2023.03.004.

1810 58. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression
1811 optimization and ancestral reconstruction. *Nat. Biotechnol.* 36, 843–846 (2018).

1812 59. Gaudelli, N. M. et al. Directed evolution of adenine base editors with increased
1813 activity and therapeutic application. *Nat. Biotechnol.* 38, 892–900 (2020).

1814 60. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and
1815 transcriptome of living cells. *Nat. Rev. Genet.* 19, 770–788 (2018).

1816 61. Li, C. et al. In vivo HSPC gene therapy with base editors allows for efficient
1817 reactivation of fetal γ -globin in β -YAC mice. *Blood Adv* 5, 1122–1135 (2021).

1818 62. Landrum, M. J. et al. ClinVar: public archive of relationships among sequence
1819 variation and human phenotype. *Nucleic Acids Res.* 42, D980–5 (2014).

1820 63. Landrum, M. J. et al. ClinVar: improvements to accessing data. *Nucleic Acids Res.*
1821 48, D835–D844 (2020).

1822 64. Kamao, H. et al. Characterization of human induced pluripotent stem cell-derived
1823 retinal pigment epithelium cell sheets aiming for clinical application. *Stem Cell Reports* 2,
1824 205–218 (2014).

1825 65. Ehrke-Schulz, E. et al. CRISPR/Cas9 delivery with one single adenoviral vector
1826 devoid of all viral genes. *Sci. Rep.* 7, 17113 (2017).

- 1827 66. Han, J. P. et al. In vivo delivery of CRISPR-Cas9 using lipid nanoparticles enables
1828 antithrombin gene editing for sustainable hemophilia A and B therapy. *Sci Adv* 8,
1829 eabj6901 (2022).
- 1830 67. Chew, W. L. et al. A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat.*
1831 *Methods* 13, 868–874 (2016).
- 1832 68. Mekler, V., Kuznedelov, K. & Severinov, K. Quantification of the affinities of
1833 CRISPR–Cas9 nucleases for cognate protospacer adjacent motif (PAM) sequences.
1834 *Journal of Biological Chemistry* vol. 295 6509–6517 Preprint at
1835 <https://doi.org/10.1074/jbc.ra119.012239> (2020).
- 1836 69. Bravo, J. P. K. et al. Publisher Correction: Structural basis for mismatch
1837 surveillance by CRISPR-Cas9. *Nature* 604, E10 (2022).
- 1838 70. Kulcsár, P. I., Tálas, A., Ligeti, Z., Krausz, S. L. & Welker, E. SuperFi-Cas9 exhibits
1839 remarkable fidelity but severely reduced activity yet works effectively with ABE8e. *Nat.*
1840 *Commun.* 13, 6858 (2022).
- 1841 71. Hibshman, G. N. et al. Unraveling the mechanisms of PAMless DNA interrogation
1842 by SpRY Cas9. *bioRxiv* 2023.06.22.546082 (2023) doi:10.1101/2023.06.22.546082.
- 1843 72. Richter, M. F. et al. Author Correction: Phage-assisted evolution of an adenine
1844 base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* 38, 901
1845 (2020).
- 1846 73. Oh, Y. et al. Expansion of the prime editing modality with Cas9 from *Francisella*
1847 *novicida*. *Genome Biol.* 23, 92 (2022).
- 1848 74. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat.*
1849 *Protoc.* 8, 2281–2308 (2013).
- 1850 75. Ansari, A. H., Kumar, M., Sarkar, S., Maiti, S. & Chakraborty, D. CriSNPr, a single
1851 interface for the curated and de novo design of gRNAs for CRISPR diagnostics using
1852 diverse Cas systems. *Elife* 12, (2023).

1853 76. Susaimanickam, P. J. et al. Generating minicorneal organoids from human
1854 induced pluripotent stem cells. *Development* 144, 2338–2351 (2017).

1855 77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat.*
1856 *Methods* 9, 357–359 (2012).

1857 78. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137
1858 (2008).

1859 79. Georgiou, G. & van Heeringen, S. J. fluff: exploratory analysis and visualization of
1860 high-throughput sequencing data. *PeerJ* 4, e2209 (2016).

1861 80. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
1862 genomic features. *Bioinformatics* 26, 841–842 (2010).

1863 81. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of
1864 multiple gene or genomic region sets. *BMC Bioinformatics* 18, 287 (2017).

1865 82. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing
1866 sequence analysis. *Nat. Biotechnol.* 37, 224–226 (2019).

1867 83. Shukla, R., Kannabiran, C. & Jalali, S. Genetics of Leber congenital amaurosis: an
1868 update. *Expert Rev. Ophthalmol.* (2014) doi:10.1586/eop.12.14.

1869 84. Okita, K. et al. A more efficient method to generate integration-free human iPS
1870 cells. *Nat. Methods* 8, 409–412 (2011).

1871 85. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid-base
1872 interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level.
1873 *Nucleic Acids Res.* 29, 2860–2874 (2001).

Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing

Sundaram Acharya^{a,b,1}, Arpit Mishra^{a,1,2}, Deepanjan Paul^{a,1}, Asgar Hussain Ansari^{a,b}, Mohd. Azhar^{a,b}, Manoj Kumar^{a,b}, Riya Rauthan^{a,b}, Namrata Sharma^a, Meghali Aich^{a,b}, Dipanjali Sinha^{a,b}, Saumya Sharma^{a,b}, Shivani Jain^a, Arjun Ray^{a,3}, Suman Jain^c, Sivaprakash Ramalingam^{a,b}, Souvik Maiti^{a,b,d}, and Debojyoti Chakraborty^{a,b,4}

^aGenomics and Molecular Medicine Unit, Council of Scientific and Industrial Research—Institute of Genomics & Integrative Biology, New Delhi, 110025, India; ^bAcademy of Scientific & Innovative Research, Ghaziabad, 201002, India; ^cKamala Hospital and Research Centre, Thalassemia and Sickle Cell Society, Rajendra Nagar, Hyderabad, 500052, India; and ^dInstitute of Genomics and Integrative Biology (IGIB)-National Chemical Laboratory (NCL) Joint Center, Council of Scientific and Industrial Research—National Chemical Laboratory, Pune, 411008, India

Edited by K. VijayRaghavan, Tata Institute of Fundamental Research, Bangalore, India, and approved September 6, 2019 (received for review October 27, 2018)

Genome editing using the CRISPR/Cas9 system has been used to make precise heritable changes in the DNA of organisms. Although the widely used *Streptococcus pyogenes* Cas9 (SpCas9) and its engineered variants have been efficiently harnessed for numerous gene-editing applications across different platforms, concerns remain regarding their putative off-targeting at multiple loci across the genome. Here we report that *Francisella novicida* Cas9 (FnCas9) shows a very high specificity of binding to its intended targets and negligible binding to off-target loci. The specificity is determined by its minimal binding affinity with DNA when mismatches to the target single-guide RNA (sgRNA) are present in the sgRNA:DNA heteroduplex. FnCas9 produces staggered cleavage, higher homology-directed repair rates, and very low nonspecific genome editing compared to SpCas9. We demonstrate FnCas9-mediated correction of the sickle cell mutation in patient-derived induced pluripotent stem cells and propose that it can be used for precise therapeutic genome editing for a wide variety of genetic disorders.

CRISPR Cas9 | genome editing | gene therapy | iPSCs | sickle cell anemia

The process of introducing changes in the DNA of cells using CRISPR/Cas proteins has emerged as a powerful technique in molecular biology with potentially far-reaching applications in gene therapy (1–6). The method involves harnessing the prokaryotic type II CRISPR/Cas protein Cas9, which in complex with a single-guide RNA (sgRNA), can be directed to any region in the DNA upstream of a protospacer adjacent motif (PAM) with which the sgRNA sequence finds a match (7–9). Upon double-stranded cleavage at the target site, the endogenous repair machinery of the cell can be utilized to make nucleotide changes in the DNA (6, 10–12). Although several Cas9 proteins recognizing different PAM sequences have so far been reported in the literature, only a subset of these have been characterized and have demonstrated genome-editing ability in eukaryotic cells (5, 13–16).

Cas9 from *Francisella novicida* (FnCas9) is one of the largest Cas9 orthologs and has been shown to predominantly interact with the 5'-NGG-3' PAM motif in DNA (17). The crystal structure of FnCas9 ribonucleoprotein (RNP) in complex with target DNA has revealed both conserved and divergent features of interaction that is unique among the Cas9 enzymes studied. Unlike SpCas9, FnCas9 does not form a bilobed structure, has a different sgRNA scaffold, and has been implicated in RNA targeting (17, 18). By structure-guided protein engineering, FnCas9 can be made to recognize a 5'-YGG-3' PAM (17). Although the protein can efficiently cleave DNA substrates in vitro, its in vivo activity at several genomic loci is significantly diminished as compared to SpCas9 (17, 19).

FnCas9 has been reported to have high intrinsic specificity for its targets, most notably by tolerating only a single mismatch in the sgRNA at the 5' position of the PAM distal end (17). This is in stark contrast to *Streptococcus pyogenes* Cas9 (SpCas9), which

has shown variable levels of off targeting due to tolerance of mismatches predominantly in the “nonseed” region in the sgRNA, wherever these are encountered in the genome (20). To what extent FnCas9 mediates this high specificity of target interrogation is not known and whether these properties can be harnessed for highly specific genome editing at a given DNA loci has not been investigated. The distinct structural attributes of FnCas9 and its low tolerance for mismatches led us to investigate its DNA interaction properties and role in genome editing.

Results

FnCas9 Shows Sequence-Dependent DNA Binding Affinity and Cleavage Kinetics. The interaction of Cas9 to its substrate occurs in a multistep process involving PAM binding, target unwinding, and DNA:RNA

Significance

Therapeutic genome editing has been significantly accentuated by the advent of CRISPR-based gene correction. However, genomic off-targeting has been a major setback for clinical translation. Although high-fidelity versions of Cas9 have been designed, these recognize and bind to off-targets. In this study, we characterize a naturally occurring Cas9 from *Francisella novicida* (FnCas9) that shows negligible binding affinity to off-targets differing by 1 or more mismatches, rendering it highly specific in target recognition. We show that FnCas9 can direct both homology-directed repair (HDR) and nonhomologous end-joining-mediated DNA repair, generates a higher rate of HDR and negligible off-targeting. Finally, we show its potential in therapeutic genome editing by correcting the sickle cell anemia mutation in patient-derived induced pluripotent stem cells.

Author contributions: S.A., A.M., D.P., M. Azhar, S.M., and D.C. designed research; S.A., A.M., D.P., M. Azhar, M.K., R.R., N.S., M. Aich, D.S., S.S., Shivani Jain, and D.C. performed research; Suman Jain and S.R. contributed new reagents/analytic tools; D.P., A.H.A., A.R., S.M., and D.C. analyzed data; and S.A. and D.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: Sequencing data reported in this paper have been deposited in NCBI Sequence Read Archive (accession no. SRP119217), and plasmid sequences have been deposited in Addgene (<https://www.addgene.org/browse/article/28203957/>).

¹S.A., A.M., and D.P. contributed equally to this work.

²Present address: Division of Medical Genetics, University of Washington School of Medicine, Seattle, WA 98195.

³Present address: Center for Computational Biology, Indraprastha Institute of Information Technology, New Delhi, 110020, India.

⁴To whom correspondence may be addressed. Email: debojyoti.chakraborty@igib.in.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1818461116/-DCSupplemental.

First published September 30, 2019.

heteroduplex formation (21, 22). Because several components in this process are dependent on enzyme conformation, we speculated that FnCas9 might possess DNA interrogation parameters that are intrinsically linked to its structure. To dissect the binding affinity of the FnCas9 RNP complex with target DNA, we used microscale thermophoresis (MST) to determine the dissociation constant (K_d) of the FnCas9 RNP complex:target interaction. To perform binding-affinity measurements, we generated recombinant catalytically dead (d, inactive) FnCas9 (dFnCas9) tagged with green fluorescent protein (GFP) and confirmed its inability to perform in vitro cleavage of a DNA substrate (SI Appendix, Fig. S1 A and B). Electrophoretic gel-mobility shift experiments (EMSA) concluded that FnCas9 and its sgRNA sequence interaction reaches saturation at a molar ratio of $\sim 1:1.5$, similar to that reported for SpCas9 binding with its sgRNA (9, 23) (SI Appendix, Fig. S1C).

Next, we investigated the binding of the dFnCas9:sgRNA RNP complex with 2 different DNA substrates (*VEGFA site3* [*VEGFA3*] and *EMX1*) and observed a K_d of 150.7 ± 36.6 nM and 78.8 ± 18.5 nM, respectively (Fig. 1). To compare the binding of dSpCas9 to the same targets, we purified recombinant dSpCas9-GFP and interrogated the substrates using MST. We observed a lower K_d of substrate binding for dSpCas9 for each of the targets (49.6 nM \pm 9.78 nM and 10.9 ± 6.2 nM, respectively),

suggesting that in comparison with dFnCas9, dSpCas9 might have a generally higher affinity for the same substrate sequence (Fig. 1). Notably, the K_d of SpCas9 binding to its substrates as reported in the literature using other techniques (such as EMSA, Beacon assays, or active site titration assays) also fall in a similar range (≤ 100 nM) as observed under our experimental conditions, suggesting concordance between the different methods used in determining binding affinity (9, 24–27). We tested 2 more substrate sequences (*EMX1_2*, *c-MYC*) with dFnCas9 and once again observed a wide distribution of dissociation constants (K_d 462 ± 52.4 nM and 30.6 ± 6.7 nM, respectively) (SI Appendix, Fig. S1D). Collectively, these data suggest that FnCas9 shows a diverse range of binding affinities with different DNA substrates and its affinity for targets is generally lower than that of SpCas9.

We then proceeded to investigate if these differences in binding affinity might lead to variability in cleavage efficiencies across different substrates. To investigate substrate cleavage, we purified wild-type FnCas9 and SpCas9 and performed in vitro cleavage (IVC) assays. When different DNA substrates were incubated with equal amounts of FnCas9 RNP, we observed that cleavage reached completion at different time points (30 min to 2 h) for these substrates, suggesting that the sequence of the target might affect the rate of cleavage; however, SpCas9 did not

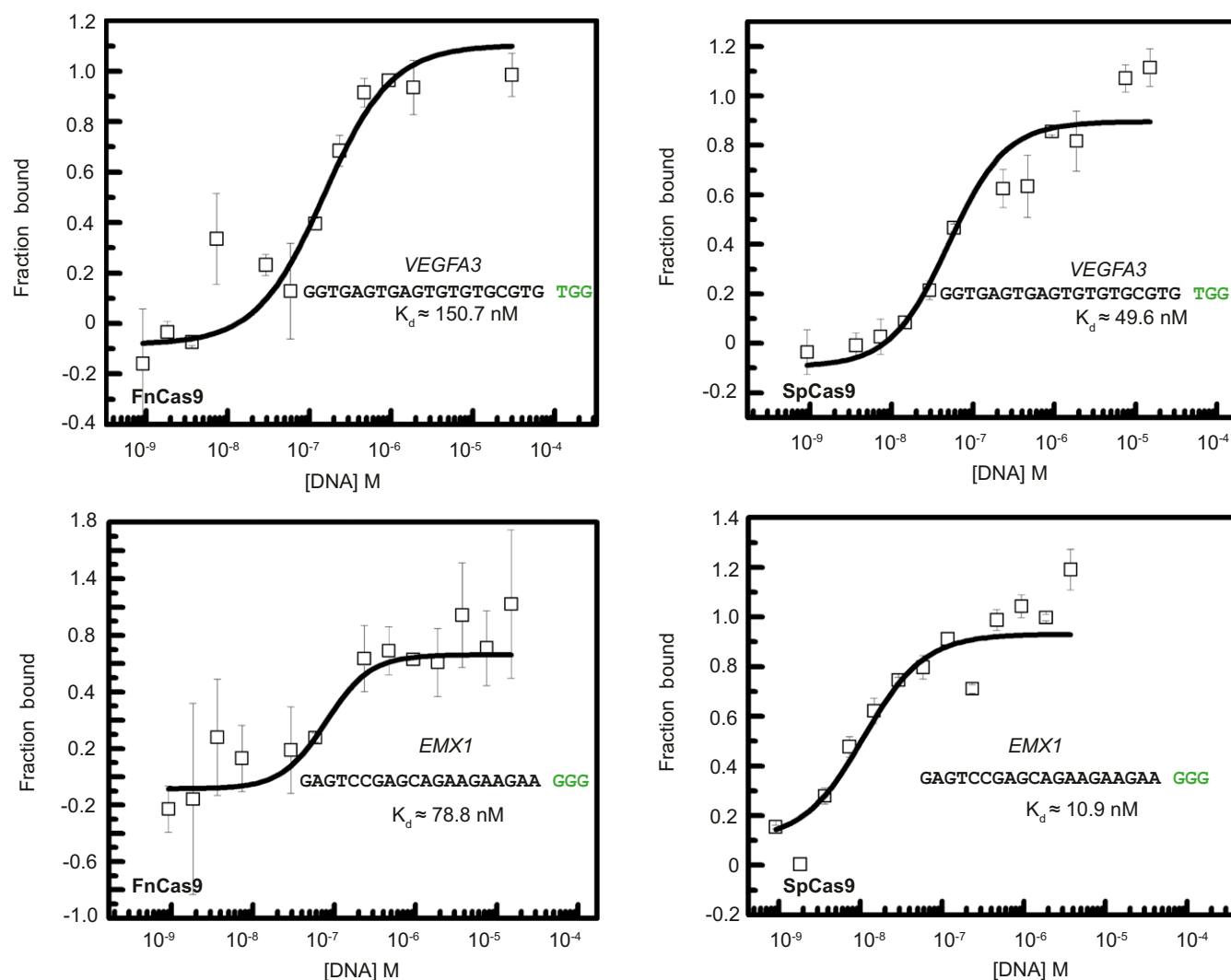


Fig. 1. FnCas9 shows different affinity and cleavage kinetics for substrates as compared to SpCas9. MST results showing binding of dFnCas9-GFP (Left) and dSpCas9-GFP (Right) to 2 substrates, *VEGFA3* (Upper) and *EMX1* (Lower) expressed as fraction bound (y axis) with respect to varying concentrations of purified DNA substrate (x axis). The substrate sequences are indicated in the box with the PAM shown in green. Error bars represent SEM (2 independent experiments).

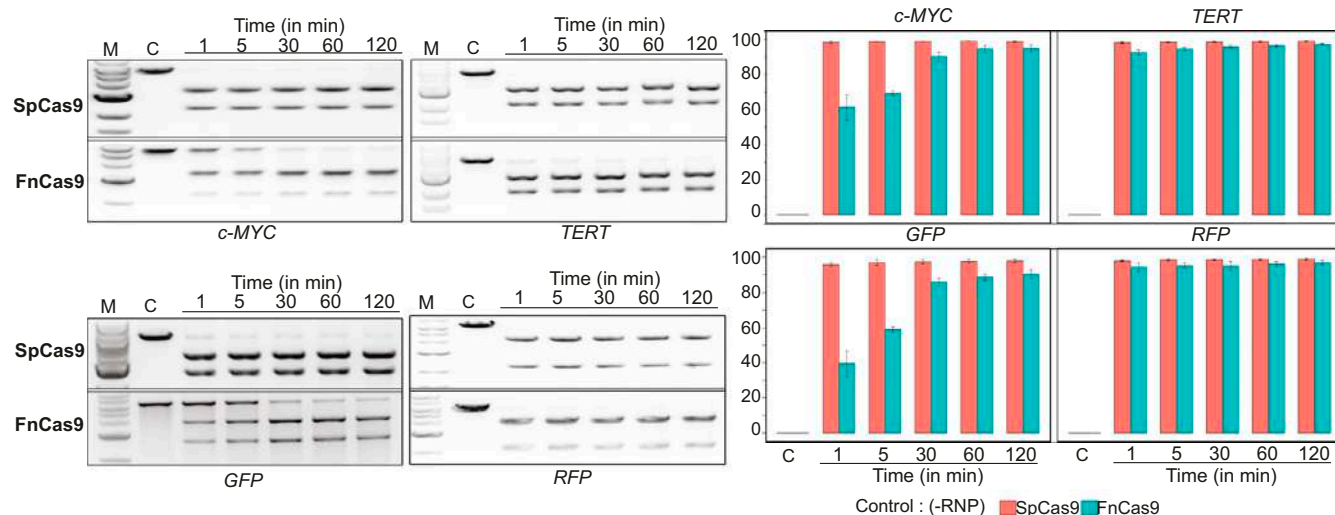


Fig. 2. FnCas9 shows different cleavage kinetics for substrates as compared to SpCas9. In vitro cleavage assay showing activity of wild-type SpCas9 and wild-type FnCas9 (50 nM Cas9:150 nM sgRNA) on indicated substrates (linearized plasmids, 250 ng) at different time points (1 min to 2 h). The top band represents uncleaved target while the 2 bottom bands represent cleaved products. Quantification is shown, *Right*. The y axis represents percentage cleavage. Error bars represent SD (3 independent experiments, C, control).

show such variability with any of the targets (Fig. 2). Interestingly, when we took a substrate that showed gradual completion of cleavage (*c-MYC*) and performed IVC with increasing molar concentrations of RNP, we observed that substrate cleavage completion could be shifted to an earlier time point upon incubation with higher concentration of FnCas9 RNP (*SI Appendix, Fig. S24*). Thus, FnCas9 cleavage efficiency varies proportionally to RNP concentration, indicating that it might act as a single turnover enzyme, a feature reported earlier for SpCas9 (9). Collectively, these results suggest that FnCas9 has a generally high threshold for substrate recognition and even in the case of com-

plete match of CRISPR RNA with its target, exhibits different cleavage kinetics when encountering different target sequences.

More than 1 Mismatch in sgRNA:DNA Heteroduplex Abolishes FnCas9 Cleavage. The specificity of a genome-editing protein is guided by a balance between its affinity for target and ability to discriminate off-targets. A recent report has suggested that FnCas9 shows higher intrinsic specificity than SpCas9 to its target by showing less tolerance to single mismatches at certain sgRNA positions (19). We investigated the in vitro cleavage efficiency of both SpCas9 and FnCas9 by systematically changing every base

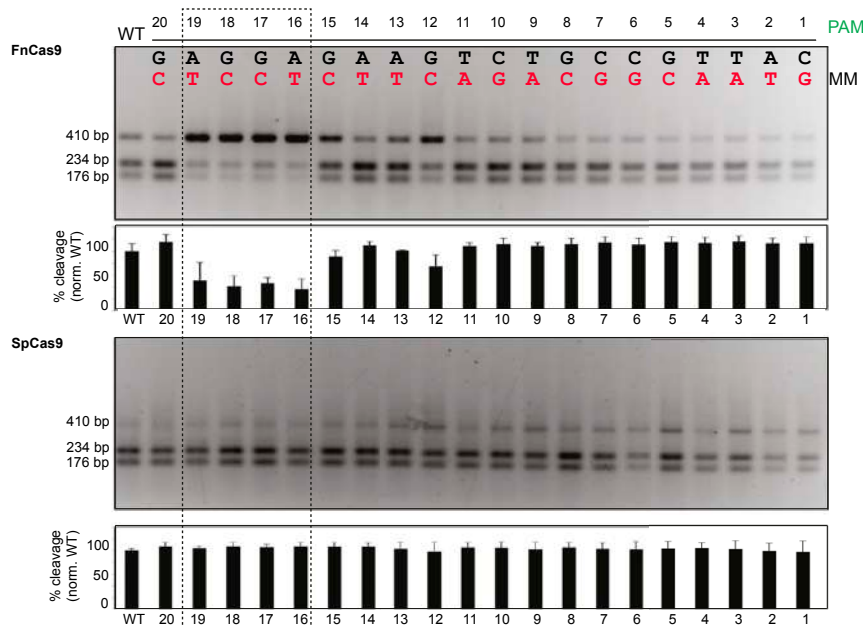


Fig. 3. Mismatches (MM) in target abrogates FnCas9 cleavage activity. Mismatches at different positions along the substrate alter FnCas9 cleavage outcomes while retaining SpCas9 activity. Representative in vitro cleavage outcome of HBB substrate (purified PCR product, 100 ng) showing mismatches at a single position in every lane (indicated in red, other bases remaining unchanged) interrogated with FnCas9 (*Upper*) or SpCas9 (*Lower*) in the form of RNP complexes (250 nM) is shown, PAM is indicated in green. The top band represents uncleaved target while the 2 bottom bands represent cleaved products. Dotted box represents positions showing minimum tolerance to mismatches for FnCas9. Quantification for each reaction is shown below the gel images. Error bars represent SD (3 independent experiments).

in a given substrate and observed that, whereas SpCas9 cleaved nearly equally at all mismatched positions, FnCas9 was less tolerant to single mismatches particularly at base positions 19, 18, 17, and 16 at the PAM distal end (Fig. 3), where considerable abolishment of cleavage was observed ($\sim 56 \pm 7\%$). Interestingly, engineered highly specific Cas9 variants that are able to prevent cleavage at off targets containing PAM distal mismatches adopt a conformational structure that renders the HNH DNA cleavage domain inactive (24). Recent biophysical studies using single-molecule FRET have revealed a highly dynamic conformation of SpCas9 where allosteric interactions between the PAM distal end and the HNH domain of the Cas9 enzyme renders a cleavage-impaired conformationally "closed" configuration upon encountering mismatches close to the 5' end of the sgRNA (28, 29).

To dissect if a greater stringency of target recognition could be achieved in case of FnCas9 by increasing the number of mismatches at the 5' end of the substrate, we selected 2 well-studied loci *EMX1* and *VEGFA3*, amplified their genomic off-targets with 2 and 3 mismatches at the nonseed region (PAM distal end), and interrogated the in vitro cleavage efficiency of FnCas9.

Remarkably, FnCas9 was unable to cleave the substrate in the presence of 2 or 3 mismatches for both loci, suggesting that it is extremely specific in target recognition, particularly when the mismatches occur together in the PAM distal region (Fig. 4A). However, SpCas9 was able to cleave both of the mismatched substrates (Fig. 4A). Previous studies have highlighted the importance of defined mutations in the REC3 domain of highly specific engineered versions of Cas9 in determining target specificity by allosterically regulating the HNH domain from adopting a cleavage-competent form (24, 30). However, the different engineered variants showed similar binding affinity for their off-targets as wild-type SpCas9, even though they did not cleave these targets, suggesting that they probably remain bound to off-targets in a cleavage incompetent state (24). We asked if FnCas9 too shows similar properties and investigated the binding affinity for off-targets using MST. Strikingly, FnCas9 showed negligible to no binding affinity for substrates having 2 mismatches where no cleavage was observed, suggesting that it either interacts extremely weakly or is evicted from the substrate following off-target interrogation (Fig. 4B and *SI Appendix, Fig. S2A*). However, SpCas9 showed a strong binding affinity for both on- and

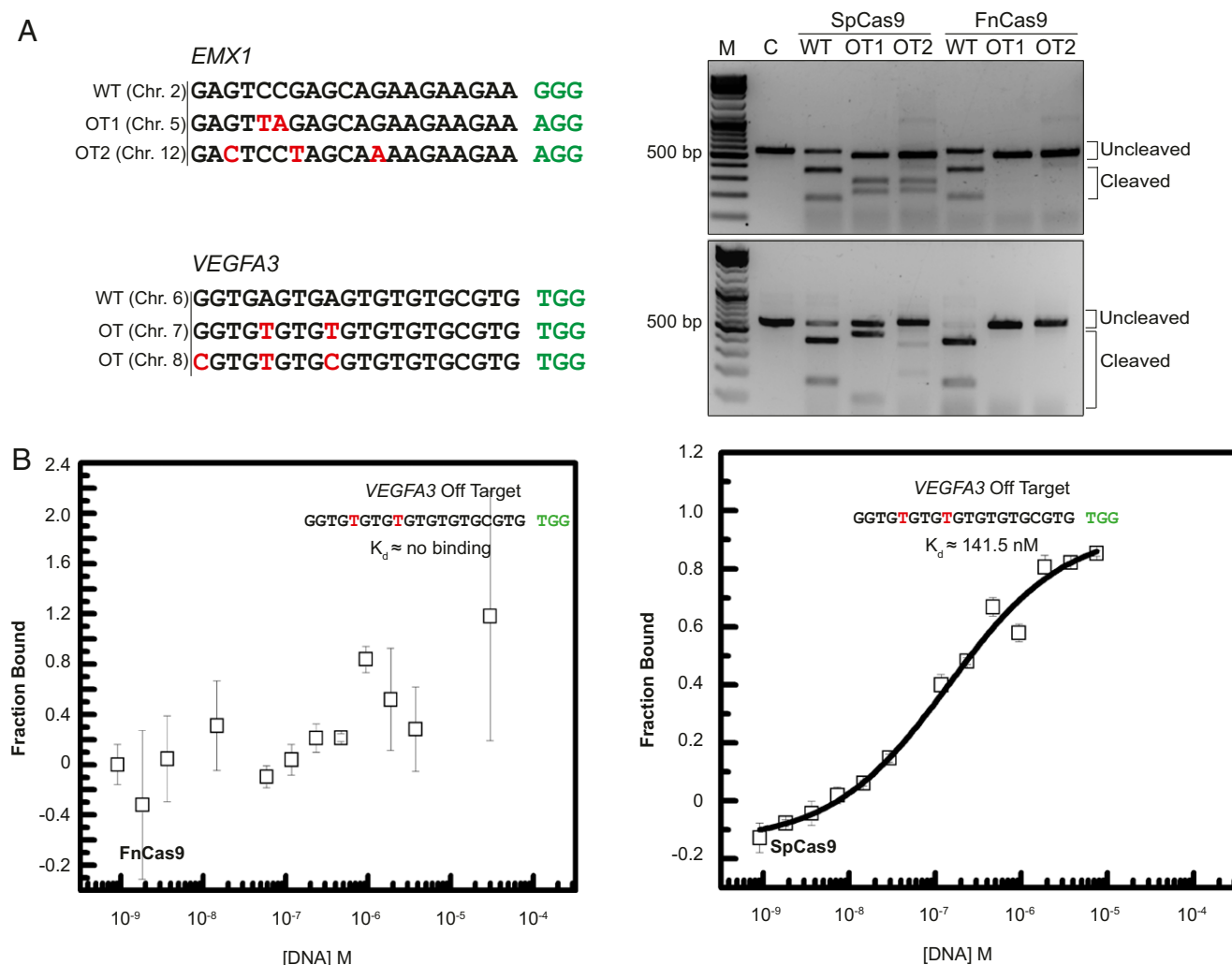


Fig. 4. FnCas9 does not bind to mismatched substrates. (A) FnCas9 does not tolerate 2 or more mismatches in the substrate. In vitro cleavage outcomes of *EMX1* (Upper) and *VEGFA3* (Lower) targets interrogated by wild-type FnCas9 and SpCas9 are shown (Right). (Left) The sequences with 2 or more mismatches (indicated as red bases) are presented. PAM is indicated in green. Data representative of 3 independent experiments. (B) MST results showing binding outcomes of dFnCas9-GFP (Left) and dSpCas9-GFP (Right) to *VEGFA3* off-target substrate expressed as fraction bound (y axis) with respect to varying concentrations of purified DNA substrate (x axis). The substrate sequences are indicated in the box with the PAM shown in green. Error bars represent SEM (2 independent experiments).

off-targets, as reported previously (24) (Fig. 4*B* and *SI Appendix, Fig. S2B*). Taken together, these results indicate that FnCas9 has a fundamentally distinct outcome of off-target recognition and binding as compared to SpCas9 and its engineered derivatives.

Intrigued by the ability of FnCas9 to discern mismatches in the target with high accuracy, we looked at the protein structure to understand the basis for this specificity. The interaction of FnCas9 with its substrate is predominantly mediated by an expanded REC3 and the associated REC2 domains that are structurally distinct from those found in SpCas9 or *Staphylococcus aureus* Cas9 (SaCas9) (17). In the crystal structure of FnCas9, the REC3 domain was reported to contain a structural zinc ion in a coordination sphere consisting of 4 cysteine residues (17). Notably, complete abolishment of FnCas9 activity was observed when in vitro cleavage reaction was supplanted with Zn^{2+} (1 mM), suggesting that the binding of the metal ion inhibits FnCas9-mediated cleavage even in the presence of Mg^{2+} (*SI Appendix, Fig. S2C*). Upon analyzing the crystal structure of FnCas9 RNP in complex with target DNA, we observed that the FnCas9 has a much higher electrostatic potential than SpCas9 and interacts with several bases, both on the PAM distal and proximal ends of the substrate (*SI Appendix, Fig. S2D*). We speculate that these interactions might be necessary for FnCas9 to adopt a cleavage-competent state and in the presence of at least 2 mismatches, the complex dissociates. In fact, the electrostatic contributions of FnCas9 on each of the 20 bases in the substrate are higher than that of SpCas9, possibly leading to stringent interrogation and subsequent dissociation upon mismatch recognition (*SI Appendix, Fig. S2E*).

We asked if these features could make FnCas9 extend its specificity for off-target discrimination at both the PAM proximal and distal ends. To test this, we designed substrates carrying the therapeutically relevant human *HBB* sequence and introduced 2 mutations at different positions throughout the substrate. We observed that cleavage activity went down drastically, suggesting that FnCas9 has very low tolerance for double mismatches along most of the sgRNA:DNA duplex, while SpCas9 activity remained largely unaffected (*SI Appendix, Fig. S3A*). This discrimination of off-targets is not correlated with distance between the 2 mismatches (*SI Appendix, Fig. S3A*). We also inquired if off-targets can be distinguished by introducing double mismatches in the sgRNA sequence. To this end, we targeted this substrate using sgRNAs having double mismatches at various positions in the PAM proximal and distal ends, and observed that complete loss of cleavage activity could be effected with mismatch combinations both at PAM proximal and distal ends (*SI Appendix, Fig. S3B*). To rule out that this specificity is dependent on the sequence being interrogated, we looked at 3 more targets (*c-MYC*, *VEGFA3*, and *EMX1*) with sgRNAs containing 2 mismatches and found that in these targets too, cleavage was completely abrogated in the presence of these mismatches (*SI Appendix, Fig. S3C*). Taken together these results suggest that the trigger for mismatch discrimination is embedded in the FnCas9 structure and contributes to the intrinsic specificity of the enzyme, as reported previously (19).

FnCas9 Mediates Cellular Genome Editing with Very High Precision.

Intrigued by the high specificity of substrate recognition under in vitro conditions, we next investigated if FnCas9 can function in vivo as a genome-editing agent and how its genome-editing properties compared with that of SpCas9. We first investigated if FnCas9 can access nuclear DNA by complexing a dFnCas9-GFP protein with sgRNAs that target the telomeric TTAGGA repeats and performing in situ labeling of genomic loci in mouse embryonic stem (ES) cells that are known to have longer telomeres amenable to imaging (31). We observed identical nuclear punctate dots for both dSpCas9-GFP and dFnCas9-GFP using deconvolution microscopy establishing that FnCas9 can localize to genomic DNA efficiently (Fig. 5*A*). We then asked if DNA localization also translates to actual DNA binding events. To this end, we used a mouse cell line containing a GFP sequence (Neuro2A-GFP),

transfected it with a plasmid expressing FnCas9, and in vitro-transcribed sgRNAs targeting the GFP sequence, and measured chromatin binding of FnCas9 using chromatin immunoprecipitation. We observed up to 100-fold enrichment of the protein at the target 24 h postsgRNA transfection (*SI Appendix, Fig. S3D*), thus confirming that both active and dead FnCas9 were able to access their target sites on DNA in 2 different mouse cell lines.

Having established that FnCas9 localizes and binds to DNA in mouse cells, we next asked if it can perform genome editing in human cells and how this compares with SpCas9. For a bona-fide comparison between the 2 different Cas9 proteins with identical expression parameters, we generated constructs with either SpCas9 or FnCas9 and their corresponding sgRNAs from the same backbone. To select cells that received this plasmid by FACS sorting, we supplemented this construct with a T2A-eGFP sequence. To compare on-target and off-target editing, we selected 2 loci, *c-MYC* and *EMX1*, where genomic off-targets with 2 mismatches could be identified and 1 locus *HBB*, where genomic off-targets with 3 mismatches could be identified. In the case of *EMX1*, these off-targets were earlier validated by an unbiased genome-wide GUIDE-Seq study (32). We performed genome editing with SpCas9 and FnCas9 in HEK293T cells and a construct containing a scrambled sgRNA sequence was used as control. After sorting GFP⁺ cells from both SpCas9 and FnCas9 transfected cells, we isolated genomic DNA from these cells, amplified the target and off-target loci for each of the genes, and performed deep sequencing to quantify the number of insertions/deletions (indels). We observed that FnCas9 was able to cleave all of the 3 loci efficiently although with different efficiencies as compared to SpCas9. In the case of *c-MYC*, ~60% indels were seen, which dropped to ~22% for *EMX1* and ~19% for *HBB*. In contrast, SpCas9 showed greater than 50% indels for each of the 3 loci tested (Fig. 5*B* and *C* and *SI Appendix, Fig. S3E*). Analysis of the percentage of either insertions or deletions at the different loci did not reveal a consistent trend for any of these events being favored by both SpCas9 or FnCas9 (*SI Appendix, Fig. S4A*).

Although SpCas9 cleaved more efficiently at each of the 3 loci, it showed modest to high cleavage at the off-targets for each locus. Among those with 2 mismatches, these included 1 off-target for *c-MYC* (~8%) and 2 off-targets for *EMX1* (~16% each) (Fig. 5*B* and *C*). Even with 3 mismatches to the sgRNA, it showed ~2% cleavage for at least 1 off-target. Strikingly, FnCas9 did not show any cleavage at any of the off-targets tested except *EMX1* off-target 2, where a very low ~1.15% of indels were detected (Fig. 5*C*). Of note, the off-targets previously validated for cleavage by SpCas9 through GUIDE-Seq (*EMX1*) also showed the highest SpCas9 off-target activity in our analysis, suggesting concordance between the 2 studies in terms of bona-fide off-targeting. However, FnCas9 showed almost no cleavage at these sites either, suggesting a highly stringent mismatch detection mechanism. Although it might be argued that FnCas9 shows a generally lower efficiency of DNA cleavage due to which off-target cleavage events are diluted, the case of *c-MYC* where it exhibits comparable cleavage efficiency as SpCas9 yet maintains no cleavage at off-target loci rules out such an explanation (Fig. 5*B*). Taking these data together, in our studies we find that FnCas9 did not show any off-target activity, a prerogative for therapeutic gene targeting.

FnCas9-Mediated Genome Editing Shows a Higher Efficiency of Genetic Insertions as Compared to SpCas9.

In a previous study (19), it was reported that FnCas9 produces a staggered pattern of DNA cleavage with a predominantly 4-nt 5' overhang in the target strand. Interestingly, similar properties of staggered DNA cleavage have also been reported for Cpf1 protein, which recognizes a T-rich PAM, generates sticky ends, and is less tolerant to single or double mismatches in the CRISPR RNA sequence (33). In addition, Cpf1 has also been associated with very low off-targeting in mammalian cells (34). Observing certain parallels in the mode of cleavage between Cpf1 and FnCas9, we first examined the cleavage

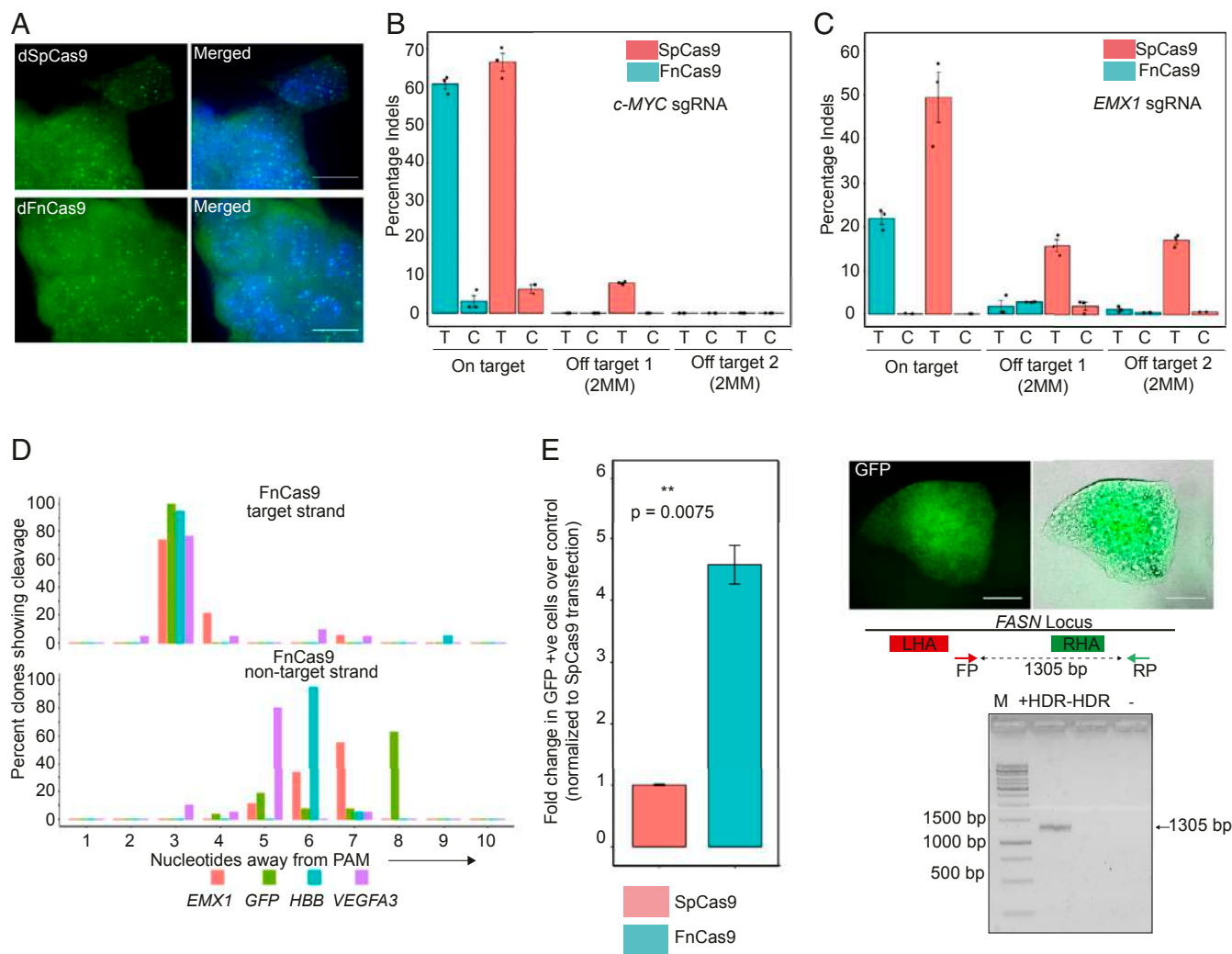


Fig. 5. Fncas9 performs genome editing with very high specificity. (A) Representative immunofluorescence image showing the nuclear localization of dSpCas9-NLS-GFP or dFncas9-NLS-GFP protein (shown in green) with sgRNA (as RNP complex) against telomeric repeats in mouse ES cells. DNA is counterstained with DAPI and represented in blue. (Scale bar, 10 μ m.) (B) Indel events (expressed as percentage) obtained by amplicon sequencing upon SpCas9 or Fncas9 targeting *c-MYC* in HEK293T cells. Individual data points from independent experiments are shown. Error bars represent SEM (3 independent experiments except for 1 target control for SpCas9). C, control; T, test. (C) Indel events (expressed as percentage) obtained by amplicon sequencing upon SpCas9 or Fncas9 targeting *EMX1* in HEK293T cells. Individual data points from independent experiments are shown. Error bars represent SEM (3 independent experiments). (D) Cleavage positions on target and nontarget strand by Fncas9 determined using Sanger sequencing for different targets indicated. The y axis represents percentage of sequenced clones showing cleavage at a given nucleotide. The x axis represents the position of the base away from PAM in the sgRNA. (E) HDR represented as fold-change in GFP⁺ cells over a scrambled sgRNA control (normalized to SpCas9 transfection). Error bars represent SEM (3 independent experiments). Student's *t* test *P* values are shown. A representative image of GFP-tagged HEK293T cells after Fncas9 mediated HDR is shown. (Scale bar, 50 μ m.) (F) Genotyping PCR confirming targeted GFP integration at the *FASN* locus in sorted GFP⁺ (+HDR) or GFP⁻ (-HDR) and no template control (-) for Fncas9-mediated HDR. Integration at the desired locus generates a 1,305-bp PCR product.

site of Fncas9 for multiple targets (*GFP*, *EMX1*, *VEGFA3*, and *HBB*). We cloned sticky-ended products of Fncas9-mediated in vitro cleavage reaction into a destination vector and performed bidirectional Sanger sequencing of the clones. We observed that in nearly all clones, the target strand was cleaved at 3 bp upstream of the PAM. However, the nontarget strand showed different positions of cleavage (3 to 8 bp away from the PAM), depending upon the sequence being cleaved (Fig. 5D). In contrast, SpCas9 showed cleavage 3 bp upstream of the PAM for the nontarget strand, as reported previously (8) (*SI Appendix, Fig. S4B*). Taken together, these experiments reveal a flexible nontarget strand cleavage activity by Fncas9, which has not been seen so far in other naturally occurring Cas9 proteins. Single-molecule studies would be required to dissect how Fncas9 maintains this flexibility in nontarget strand cleavage and the conformational changes that it encounters upon binding to a target.

We speculated that the sticky ends generated by Fncas9 might impact the rate of homology-directed repair (HDR) -mediated genomic insertions. Since HDR is an overarching component of therapeutic genome editing, we interrogated if Fncas9 can integrate foreign DNA in the form of a GFP reporter in a targeted fashion. We transfected HEK293T cells with a DNA template containing an out-of-frame GFP sequence flanked by homology sites to the mammalian *FASN* locus, along with SpCas9 or Fncas9 with corresponding sgRNAs. A scrambled sgRNA sequence was used as control for both. We sorted cells on the basis of transient GFP expression from Cas9-containing plasmids after 72 h and allowed them to grow for 10 d to generate in-frame GFP from the endogenous *FASN* locus targeted by HDR. After 10 d, we analyzed the levels of the endogenous GFP by FACS. We consistently observed higher (~4.6 fold) GFP⁺ cells in Fncas9-transfected cells compared to SpCas9-transfected cells,

suggesting that FnCas9 produces higher efficiency of genetic insertions possibly through HDR (Fig. 5E). Using genotyping primers, we successfully validated these HDR events at the targeted locus (Fig. 5E). Collectively, these results suggest that FnCas9 shows higher levels of site-specific genomic insertions in HEK293T cells.

FnCas9 Generates Detectable Levels of Genome Editing in Patient-Derived Induced Pluripotent Stem Cells. Therapeutic genome-editing trials for hemoglobinopathies have so far relied on ex vivo gene-editing strategies in patient-derived hematopoietic or induced pluripotent stem cells (iPSCs). Intrigued by the specificity and higher HDR efficiency of FnCas9-mediated genome editing and to examine its efficacy in a clinical set-up, we next tested its activity in a disease model (Fig. 6A). To this end, we collected peripheral blood mononuclear cells (PBMCs) from a patient carrying the homozygous sickle cell anemia mutation and generated iPSCs using Sendai virus-mediated reprogramming (*SI Appendix, Fig. S4C*). We characterized the resultant iPSC colonies with respect to their characteristic dome-shaped morphology and expression of pluripotency markers OCT4 and SOX2 (Fig. 6B). To further establish their pluripotent character, we differentiated the iPSCs using defined factors into the 3 lineages and validated them by successful expression of ecto-, meso-, and endodermal marker proteins SOX1, α -SMA, and GATA4, re-

spectively (*SI Appendix, Fig. S4D*). Digital karyotyping revealed no abnormal chromosomal aberrations in these iPSCs at different passages (*SI Appendix, Fig. S4E*).

We then proceeded to electroporate these cells with FnCas9 or SpCas9. Since CRISPR clinical trials are currently focused on RNP-mediated gene editing due to shorter persistence of the CRISPR complex and subsequent lesser off-targeting inside the cell (35), we generated RNP complexes targeting the mutant sickle cell anemia locus. For targeting the HBB locus harboring the sickle cell mutation, we attempted 2 different strategies: A single-stranded donor nucleotide (ssODN) with 50-bp homology-arms containing the wild-type β -globin sequence or a double-stranded DNA (dsDNA) fragment carrying the same sequence. We electroporated patient iPSCs with RNP complexes carrying FnCas9/SpCas9-GFP along with donor templates, sorted GFP⁺ cells, and analyzed the targeting locus using next-generation sequencing. Consistent with reports in literature of very low targeting efficiency in iPSCs without the use of adeno-associated virus (AAV) donor vectors (36), we observed very few HDR events in both SpCas9 and FnCas9 over the scrambled control. Strikingly, the number of significant HDR events was higher for FnCas9 as compared to SpCas9 when an ssODN was used (Fig. 6C). This is similar to what is reported for Cpf1 where the staggered cleavage on DNA has been attributed to a higher incorporation rate of ssODNs (37) and a lower rate of incorporation

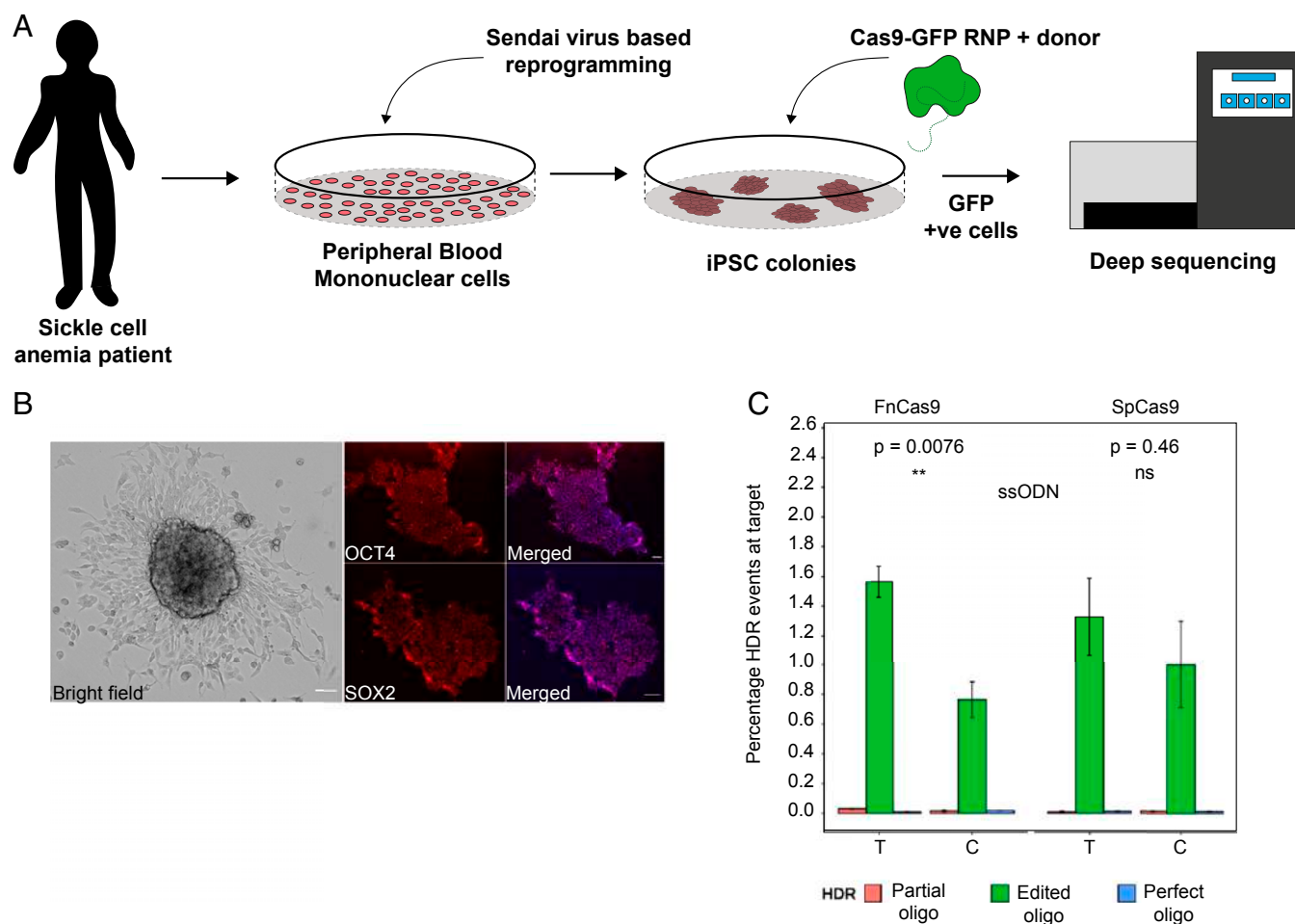


Fig. 6. FnCas9-based genome editing in iPSCs. (A) Schematic showing genome editing using FnCas9 in patient derived iPSCs. (B) Micrograph showing a sickle cell disease patient-derived iPSC colony (Left). iPSC colony stained for pluripotency markers (Oct4 and Sox2) and merged image with DAPI (in blue) is shown (Right). (Scale bars, 25 μ m.) (C) HDR-based correction of the sickle cell mutation in patient iPSCs using an ssODN donor shown as percentage of HDR events at the target. Error bars represent SEM (3 independent experiments). Student's *t* test *P* values are shown; ns, not significant. Types of HDR detected are represented (for description, see *Materials and Methods*).

seen when blunt-ended oligos are used. Indeed, we observed almost no HDR events with FnCas9 when a blunt double-stranded oligo was used, even though SpCas9 showed very low but significant detectable HDR events at the target (*SI Appendix, Fig. S4F*). Taken together, these results suggest that HDR events can be generated by FnCas9-based editing, particularly with the use of ssODNs, and this can be utilized for correcting disease-causing mutations in patient-derived cells.

Discussion

Precise genome engineering depends on the ability of the genome-editing agent to interrogate a given DNA locus with high specificity (38). The bacterial type II CRISPR/Cas system has been efficiently exploited to edit a variety of genomes and has recently gained prominence for gene-therapy applications, predominantly for monogenic disorders like β -thalassemia and sickle cell anemia (39, 40). As one of the first Cas9 proteins described, SpCas9 has been highly successful in gene editing in diverse biological systems, but reports of off-targeting remain a concern, particularly those of large deletions and complex rearrangements (41). In this study, we show that FnCas9, a type II orthogonal Cas9 protein, shows very high specificity to its target and negligible affinity with off-targets, thereby increasing the repertoire of naturally occurring Cas9 proteins that can be used for therapy.

Binding studies of FnCas9 with its intended targets and off-targets reveal that FnCas9 has a fundamentally different mode of delineating off-targets. Whereas highly specific engineered domains of SpCas9 remain in an associated “cleavage incompetent” state with its off-targets (24), FnCas9 shows greatly reduced affinity to such substrates, possibly by eviction from the target upon encountering mismatches. Since the REC3 domain, which determines this specificity in the engineered variants, is structurally unique in FnCas9, further biophysical studies aimed at visualizing the mode of action of this domain might elucidate the exact nature of target discrimination.

Systematic dissection of FnCas9 cleavage across different substrates reveals that its activity is largely abolished in the presence of 2 mismatches, most prominently at the PAM distal end. This understanding sets up a framework for FnCas9 sgRNA design algorithms with a high degree of specificity by rationally introducing mismatches in sgRNA sequences at defined positions that can distinguish between the target and a very similar off-target.

Although FnCas9 has a large size that can potentially deter its genomic interrogation to various sites, we observed successful in vivo genome editing at each of the genomic loci interrogated, suggesting that it can access endogenous targets once it is inside the nucleus. Smaller sized Cas9 proteins offer the advantage of ease of delivery but can be more promiscuous in binding to unintended sites in the genome, highlighting the importance of a balance between size and specificity.

In our cellular studies, FnCas9 showed negligible off-targeting at the various loci tested. Although certain engineered Cas9 proteins have shown very high discrimination of off-targets in the genome, to our knowledge this property is not seen among naturally occurring Cas9 proteins. How FnCas9 has evolved into a more precise genome editor than its other bacterial orthologs is a matter of further investigation. A study of structurally similar proteins from other species of bacteria and archaea might reveal genome-editing proteins with beneficial attributes.

Finally, we show that FnCas9 can be used for HDR-based modification of a desired target in the genome. The generation of staggered DNA cleavage, similar to Cpf1, is unique among naturally occurring Cas9 proteins and opens up the possibility of novel mechanisms of nuclease domain engagement with the target and nontarget strands. Sticky ends might also explain higher rates of HDR seen in cells, a highly attractive feature of FnCas9 that can be suitably exploited in genetic engineering of a target loci and lays the foundation for including this protein to currently available genome-editing agents for potential gene-therapy applications. Although we did not achieve high levels of gene editing in patient-derived iPSCs using FnCas9 RNPs, de-

livery of both the protein and donor can be vastly improved using AAV vectors, as reported in some recent studies (36). The bulky nature of the protein will necessitate shorter engineered versions that will be more apt for delivery, potentially by removal of redundant domains yet retaining the on-target specificity of FnCas9. Due to the modularity of the Cas9 enzyme, such constructs would need to exploit the steadily evolving functional link between REC3 and substrate mismatch discrimination.

Our studies indicate that FnCas9 might dissociate from its mismatched substrates rather rapidly, resulting in minimal affinity with off-targets, a property that is strikingly dissimilar from other rationally designed Cas9 proteins that are highly specific (24). Association with off-target loci might result in transcriptional repression in these regions that can confound outcomes of on-target activity and lead to cleavage-independent gene-expression changes. Although it remains to be tested, FnCas9, due to its low binding affinity to off-targets, might present a more attractive scenario for genome editing, where even long-term persistence of the FnCas9 inside the cell might not induce off-targeting because of the high threshold for DNA binding when mismatches in the targets are encountered.

FnCas9, a type II-B Cas system, has evolved separately from the II-A and II-C systems comprising of the majority of the commonly used Cas9 proteins used for genome editing (42). Its distinct structural attributes, leading to high targeting-specificity, hints at the existence of more members of the II-B family that can be similarly explored for mammalian genome editing. Furthermore, FnCas9 can be structurally engineered to render it competent for base editing, an attractive therapeutic application for monogenic disorders where the current generation of base editors have so far shown variability in off-targeting (43).

Materials and Methods

Plasmid Construction. The gene encoding full-length *F. novicida* (Fn) Cas9 nuclease residues 1 to 1,629 bp was PCR-amplified using PX408 (Addgene 68705) as a template and cloned in pET28-His-10-Smt3 vector (a kind gift from Stewart Shuman, Molecular Biology Program Sloan-Kettering Institute, New York, and K. M. Sinha, Amity Institute of Biotechnology, Amity University, Gurgaon, Haryana, India) and pET-His6-GFP-TEV-LIC vector (Addgene 29663), following restriction enzyme-based cloning and ligation-independent cloning (LIC), respectively. Catalytically inactive FnCas9 double mutants were generated on pET-His6-FnGFP-TEV-LIC (this paper) plasmid backbone by a QuikChange II site-directed mutagenesis kit (Agilent) following the manufacturer's protocol, with some modifications.

For the expression of SpCas9 and FnCas9 in mammalian cells, 3xHA-SpCas9, 3xHA-FnCas9, and a FnCas9 BbsI sgRNA cloning site were synthesized as gene blocks (GenScript) and cloned into PX458 (Addgene 48138) backbone by restriction enzyme-based cloning. The FnCas9 BbsI sgRNA cloning site was cloned in PciI and XbaI sites of PX458 to make it suitable for FnCas9 sgRNA expression and in this vector (PX458-3xHA-FnCas9) 3xHA-FnCas9 gene block was cloned in AgeI and FseI sites. Desired guide RNA sequences were cloned as annealed oligonucleotides with appropriate overhangs in BbsI sites following the established protocol (44). A complete list of sgRNA target sites is available in [Dataset S1](#). All of the constructs were sequenced before being used. The plasmids reported in the paper have been deposited in Addgene (<https://www.addgene.org/browse/article/28203957/>).

Protein Purification. The proteins used in this study were purified as reported previously (8). Briefly, plasmids for different Cas9 proteins were expressed in *Escherichia coli* Rosetta2 (DE3) (Novagen). The protein-expressing Rosetta2 (DE3) cells were cultured at 37 °C in LB medium (supplemented with 50 mg/L kanamycin) until OD₆₀₀ reached 0.6 and protein expression was induced by 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 lysis buffer (20 mM HEPES, pH 7.5, 500 mM NaCl, 5% glycerol supplemented with 1× PIC (Roche), 100 μ g/mL lysozyme) and lysed by sonication and centrifuged. The lysate was affinity purified by Ni-NTA beads (Roche) and SUMO tag was removed using Ulp1 protease. The eluted protein was further purified by size-exclusion chromatography on a HiLoad Superdex 200 16/60 column (GE Healthcare) in 20 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, 1 mM DTT. The concentration of purified proteins were measured by a Pierce BCA protein assay

kit (Thermo Fisher Scientific). The purified proteins were stored at -80°C until further use.

In case of 6 \times His-MBP-dSpCas9, the 6 \times His-MBP tag was removed by incubating the affinity-bound protein with PreScission Protease overnight in cleavage buffer (50 mM Tris-Cl, pH 8, 150 mM NaCl, 1 mM EDTA, 1 mM DTT). The cleaved Cas9 protein was separated from fusion tag on HiLoad Superdex 200 16/60 column (GE Healthcare) in 20 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, 1 mM DTT.

MST Assay. MST assay was performed as previously described (45, 46). dSpCas9-GFP and dFnCas9-GFP proteins were used along with PAGE purified respective in vitro transcription (IVT) sgRNAs. Note that IVT sgRNAs were purified by 12% Urea-PAGE. The binding affinities of the Cas9 proteins and sgRNA (RNP) complexes towards different genomic loci were calculated using Monolith NT.115 (NanoTemper Technologies). RNP complex (Protein:sgRNA molar ratio, 1:1) was reconstituted at 25°C for 10 min in reaction buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 1 mM DTT, 10 mM MgCl_2). 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 for 30 min in reaction buffer. The sample was loaded into NanoTemper standard-treated capillaries. Measurements were performed at 25°C using 20% LED power and 40% MST power. All experiments were repeated at least 2 times for each measurement. All data analyses were done using NanoTemper analysis software.

EMSA. EMSA was performed following the protocol reported previously (47). Briefly, 250 nM sgRNA was incubated with different concentrations of catalytically inactive (D11A-H969A FnCas9) Cas9 at 25°C for 10 min in reaction buffer (20 mM HEPES, pH 7.5, 150 mM KCl, 1 mM DTT, 10% glycerol, 10 mM MgCl_2). The samples were resolved by 8% native PAGE (0.5 \times TBE buffer with 2 mM MgCl_2) run at 4°C and the gel was stained by SYBR Gold Nucleic acid Gel stain (ThermoFisher Scientific) for 30 min at room temperature in shaking condition. RNA was visualized by Typhoon FLA 7000 (GE Healthcare) and quantified by ImageJ.

CASFISH. CASFISH was done following standard CASFISH protocol (48). Cells were seeded on 22×22 mm 0.1% gelatin (Sigma)-coated coverslips and 24 h postseeding; cells were fixed at -20°C for 20 min in a prechilled solution of methanol and acetic acid at a 1:1 ratio. Fixed cells were washed 3 \times for 5 min each with 1 \times PBS with gentle shaking followed by incubation for 30 min at 37°C in blocking buffer (20 mM HEPES pH 7.5, 150 mM KCl, 10 mM MgCl_2 , freshly added 1 mM DTT, 5% [vol/vol] glycerol, 2% BSA, 0.1% Triton X). CASFISH probes were assembled by mixing 200 mM dCas9-GFP with 200 mM sgRNA targeting Telomeric region (1:1 molar ratio) in blocking buffer and incubated at room temperature for 10 min. The assembled RNP complex was applied on preblocked cells and incubated for 2 to 4 h in a humid chamber kept at 37°C . The reaction was terminated by removing the RNP complex solution followed by washing 3 \times with blocking buffer. Washed samples were stained with 5 $\mu\text{g}/\text{mL}$ DAPI and were mounted on glass slides before imaging. CASFISH samples were imaged on a DeltaVision Ultra microscope (GE Healthcare, software Acquire Ultra 1.1.1) equipped with 60 \times oil-immersion objective (NA 1.4). Images were processed using ImageJ.

Comparative Structural Study between FnCas9 and SpCas9.

Electrostatic analysis. For the purpose of understanding the underlying electrostatic contributions in the FnCas9 and SpCas9 structures, we used the standalone version of PDB2PQR (49, 50) to assign charge and radius parameters for CHARMM force field to the Fn- (PDB ID code 5B2O) and Sp- (PDB ID code 5F9R) crystal structures.

The electrostatics was calculated using Adaptive Poisson-Boltzmann Solver (APBS) (51). APBS uses iterative solvers to solve the nonlinear algebraic equations resulting from the discretized Poisson-Boltzmann equation with a fixed error tolerance of 10^{-6} . The values are colored in the range of -10 to $35 k_B T e_c^{-1}$, where k_B is Boltzmann's constant $1.3806504 \times 10^{-23} \text{ J K}^{-1}$, T is the temperature of your calculation in K, and e_c is the charge of an electron $1.60217646 \times 10^{-19} \text{ C}$. The images were generated using the University of California, San Francisco Chimera tool.

Contact analysis. To study the interaction between the nucleotide and protein, atomic contacts were calculated using a 4 \AA cutoff. The graph was generated using MATLAB.

Derivation of PBMCs from Sickle Cell Anemia Patient. The present study was approved by the Ethics Committee (Ref no. CSIR/IGIB/IHEC/17-18) and Institutional Committee for Stem Cell Research (Ref no. IGIB/IC-SCR/9), Institute of Genomics and Integrative Biology, New Delhi, India. Written informed

consent was obtained from the sickle cell patient and blood was collected. Peripheral blood was collected in heparin-coated vacutainers from a sickle cell anemia patient (6 y, male). Mononuclear cells were isolated from a 6-mL blood sample using standard Ficoll-Paque gradient centrifugation (52). Briefly, heparinized blood was diluted 1:1 in cell culture grade Dulbecco's PBS (DPBS; Gibco) and was layered over Ficoll-Paque PLUS (density 1.077 g/mL; GE Healthcare) in a 15-mL falcon tube. The tubes were centrifuged for $330 \times g$ for 30 min at room temperature with brake-off. The cell interface layer (buffy coat) was harvested and washed twice with 10% FBS (ES qualified) in DPBS solution for 15 min at $330 \times g$. PBMCs were resuspended in StemPro-34 SFM Medium supplemented with L-glutamine to a final concentration of 2 mM; and cytokines with recommended final concentrations, SCF (C-Kit Ligand) Human Recombinant Protein (Cat. No. PHC2111) 100 ng/mL, FLT-3 Human Recombinant Protein (Cat. No. PHC9414) 100 ng/mL, IL-3 Human Recombinant Protein (Cat. No. PHC 0034) 20 ng/mL, and IL-6 Human Recombinant (Cat. No. PHC0034) 20 ng/mL. Subsequently 5×10^5 cells/mL were plated in 4 wells of a 24-well plate and maintained in standard culture conditions (37°C , 5% CO_2) and fed daily for next 3 d.

Generation of iPSCs and Maintenance. Four days after initial culture, PBMCs were reprogrammed using CytoTune-iPS 2.0 Sendai Reprogramming Kit (Thermo Fisher Scientific) containing 3 Sendai virus vectors encoding *OCT3/4*, *SOX2*, *KLF4*, and *c-MYC* (53). Briefly, the volume of each viral vector for the transduction mix was calculated as per the manufacturer's protocol and added to 3×10^5 cells in 1 mL PBMC medium followed by centrifugation at $1,000 \times g$ for 30 min at room temperature. Cells were resuspended and plated in 2 mL media in 1 well of a 12-well plate. Two days posttransduction, cells were seeded on Matrigel (Corning)-coated 6-well culture plates in 2 mL StemPro-34 SFM Medium without cytokines (54, 55). Furthermore, after 4 d, cells were transitioned to mTeSR1 (Stemcell Technologies) medium. After 2 wk, iPSC colonies were picked and transferred to Matrigel-coated 12-well plates. A sickle cell disease patient-derived iPSC line was maintained in mTeSR1 medium and subcultured using enzyme free passaging reagent, ReLeSR (Stemcell Technologies). Mutations in patient derived-iPSCs were confirmed using Sanger sequencing.

Cas9 Genome Editing in iPSCs. iPSCs were treated with 10 μM ROCK inhibitor (Y-27632) 1 h before electroporation. Next, 70 to 80% confluent iPSC colonies were harvested using StemPro Accutase Cell Dissociation Reagent (Gibco) and pipetted to make a single-cell suspension. A 5- μL volume of RNP complex mix was made in 1:3 molar ratio in Resuspension Buffer R for each electroporation, and incubated for 20 min at room temperature. Next, 3.5×10^5 cells were resuspended in 10 μL of Resuspension Buffer R per electroporation condition and 20 μM ssODN or 1 μg dsDNA with 5 μL of complete RNP complex was added and electroporation performed using Neon Transfection System 10- μL Kit (Thermo Fisher Scientific) with a single pulse at 1,200 V, 30-ms pulse width (56). The electroporated cells were transferred immediately to a Matrigel-coated 24-well plate containing 0.5 mL of mTeSR1 with 10 μM ROCK Inhibitor and incubated at 37°C humidified incubator supplied with 5% CO_2 .

After 12 h, cells were washed and reincubated with fresh mTeSR1 medium with 10 μM ROCK Inhibitor for 1 h and then harvested. Next, 25,000 GFP⁺ cells per sample were sorted using BD FACS Melody Cell Sorter (BD Biosciences) and replated. gDNA was isolated after 7 d using the Wizard Genomic DNA Purification Kit (Promega) for preparation of 16S Metagenomic Sequencing Library.

iPSC Library Preparation and Sequencing. The 16S Metagenomic sequencing library preparation protocol was adapted for library preparation. Briefly, the HBB locus 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 another 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 qualitative check. Libraries were normalized, pooled and was loaded onto illumina MiniSeq platform for a 150-bp paired-end sequencing run.

Data Analysis. Determination of indel frequency from sequencing data were performed using CRISPResso2 v2.0.29 (57) with the following parameters “-ignore_substitution-min_paired_end_reads_overlap 10-max_paired_end_reads_overlap” of 500. We have detected by Sanger sequencing that the cleavage positions for FnCas9 on the nontarget strand varied from 3 to 8 bp upstream of the PAM, so the quantification window for indel detection was set 3 to 8 bp

upstream of PAM. For *c-MYC* samples (ontarget and OT1) amplicon sequences were shorter (165 and 171 respectively) than the sequencing read length of 2×250 bp; so, 50 bp were cropped from the end of reads to achieve the required minimum homology (60% according to CRISPResso2) with the amplicon. CRISPR-DAV has been used with default parameters (58) for HDR analysis of sickle cell iPSCs where HDR has been classified as “partial” if some but not all intended base changes occur along with no indels, “edited” if one or more intended base changes occur along with indels, and “perfect” if all intended base changes occur along with no indels.

ACKNOWLEDGMENTS. This work was supported by Council for Scientific and Industrial Research (CSIR) Mission Mode Project Grant HCP0008 (to D.C. and

S.M.). The *MYC* construct was a kind gift from the laboratory of Shantanu Chaudhury, CSIR Institute of Genomics & Integrative Biology (IGIB) and the *FASN* construct used for HDR experiments was generously shared by Mihail Sarov, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden. The authors thank the laboratory of Frank Buchholz and Beena Pillai for constructs and cell lines used in the study, and Debasis Dash (CSIR IGIB), Sambit Dalui (CSIR Indian Institute of Chemical Biology), and all members of the S.M. and D.C. laboratories for helpful discussions, experimental support, and feedback. The support of Manish Kumar, Imaging Facility CSIR IGIB, Manish Kumar Rai and Nanotemper, Bangalore (MST experiments) is also gratefully acknowledged. D.C. is a young researcher of the Lady Tata Memorial Trust.

1. F. Hille *et al.*, The biology of CRISPR-Cas: Backward and forward. *Cell* **172**, 1239–1259 (2018).
2. P. D. Hsu, E. S. Lander, F. Zhang, Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
3. C. H. Huang, K. C. Lee, J. A. Doudna, Applications of CRISPR-Cas enzymes in cancer therapeutics and detection. *Trends Cancer* **4**, 499–512 (2018).
4. G. J. Knott, J. A. Doudna, CRISPR-Cas guides the future of genetic engineering. *Science* **361**, 866–869 (2018).
5. A. C. Komor, A. H. Badran, D. R. Liu, CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
6. A. Singh, D. Chakraborty, S. Maiti, CRISPR/Cas9: A historical and chemical biology perspective of targeted genome engineering. *Chem. Soc. Rev.* **45**, 6666–6684 (2016).
7. L. Cong *et al.*, Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
8. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
9. 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).
10. F. Hille, E. Charpentier, CRISPR-Cas: Biology, mechanisms and relevance. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150496 (2016).
11. M. L. Maeder, C. A. Gersbach, Genome-editing technologies for gene and cell therapy. *Mol. Ther.* **24**, 430–446 (2016).
12. J. van der Oost, E. R. Westra, R. N. Jackson, B. Wiedenheft, Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **12**, 479–492 (2014).
13. M. Adli, The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).
14. A. Cebrian-Serrano, B. Davies, CRISPR-Cas orthologues and variants: Optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm. Genome* **28**, 247–261 (2017).
15. K. M. Esvelt *et al.*, Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods* **10**, 1116–1121 (2013).
16. F. J. Najm *et al.*, Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat. Biotechnol.* **36**, 179–189 (2018).
17. H. Hirano *et al.*, Structure and engineering of Francisella novicida Cas9. *Cell* **164**, 950–961 (2016).
18. A. A. Price, T. R. Sampson, H. K. Ratner, A. Grakoui, D. S. Weiss, Cas9-mediated targeting of viral RNA in eukaryotic cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6164–6169 (2015).
19. F. Chen *et al.*, Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat. Commun.* **8**, 14958 (2017).
20. P. D. Hsu *et al.*, DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
21. M. Shibata *et al.*, Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat. Commun.* **8**, 1430 (2017).
22. D. Singh, S. H. Sternberg, J. Fei, J. A. Doudna, T. Ha, Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* **7**, 12778 (2016).
23. C. Anders, O. Niewoehner, M. Jinek, In vitro reconstitution and crystallization of Cas9 endonuclease bound to a guide RNA and a DNA target. *Methods Enzymol.* **558**, 515–537 (2015).
24. J. S. Chen *et al.*, Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
25. V. Mekler, L. Minakhin, K. Severinov, Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5443–5448 (2017).
26. A. T. Raper, A. A. Stephenson, Z. Suo, Functional insights revealed by the kinetic mechanism of CRISPR/Cas9. *J. Am. Chem. Soc.* **140**, 2971–2984 (2018).
27. C. D. Richardson, G. J. Ray, M. A. DeWitt, G. L. Curie, J. E. Corn, Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
28. S. Osuka *et al.*, Real-time observation of flexible domain movements in CRISPR-Cas9. *EMBO J.* **37**, e96941 (2018).
29. M. Yang *et al.*, The conformational dynamics of Cas9 governing DNA cleavage are revealed by single-molecule FRET. *Cell Rep.* **22**, 372–382 (2018).
30. A. Casini *et al.*, A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).
31. E. Varela, R. P. Schneider, S. Ortega, M. A. Blasco, Different telomere-length dynamics at the inner cell mass versus established embryonic stem (ES) cells. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15207–15212 (2011).
32. S. Q. Tsai *et al.*, GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
33. D. Kim *et al.*, Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.* **34**, 863–868 (2016).
34. B. P. Kleinstiver *et al.*, Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.* **34**, 869–874 (2016).
35. C. A. Lino, J. C. Harper, J. P. Carney, J. A. Timlin, Delivering CRISPR: A review of the challenges and approaches. *Drug Deliv.* **25**, 1234–1257 (2018).
36. R. M. Martin, *et al.*, Highly efficient and marker-free genome editing of human pluripotent stem cells by CRISPR-cas9 RNP and AAV6 donor-mediated homologous recombination. *Cell Stem Cell* **24**, 821–828.e5 (2019).
37. Y. Wang *et al.*, Systematic evaluation of CRISPR-Cas systems reveals design principles for genome editing in human cells. *Genome Biol.* **19**, 62 (2018).
38. H. Wang, M. La Russa, L. S. Qi, CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
39. C. Lee *et al.*, Gene editing with crispr-cas9 for treating beta-hemoglobinopathies. *Blood* **126**, 3376 (2015).
40. E. A. Traxler *et al.*, A genome-editing strategy to treat β -hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat. Med.* **22**, 987–990 (2016).
41. M. Kosicki, K. Tomberg, A. Bradley, Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, 765–771 (2018).
42. K. Chylinski, K. S. Makarova, E. Charpentier, E. V. Koonin, Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6105 (2014).
43. D. Kim *et al.*, Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).
44. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
45. C. J. Wienken, P. Baaske, U. Rothbauer, D. Braun, S. Duhr, Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* **1**, 100 (2010).
46. D. Dong *et al.*, The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).
47. C. Anders, M. Jinek, In vitro enzymology of Cas9. *Methods Enzymol.* **546**, 1–20 (2014).
48. W. Deng, X. Shi, R. Tjian, T. Lionnet, R. H. Singer, CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11870–11875 (2015).
49. E. Jurrus *et al.*, Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).
50. T. J. Dolinsky, *et al.*, PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).
51. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).
52. S. K. Panda, B. Ravindran, Isolation of human PBMCs. *Bio Protoc.* **3**, e323 (2013).
53. P. T. Lieu, A. Fontes, M. C. Vemuri, C. C. MacArthur, Generation of induced pluripotent stem cells with CytoTune, a non-integrating Sendai virus. *Methods Mol. Biol.* **997**, 45–56 (2013).
54. W. Yang *et al.*, iPSC reprogramming from human peripheral blood using Sendai virus mediated gene transfer. *StemBook*, L. Girard, Ed. (Harvard Stem Cell Institute, Cambridge, MA) (2012).
55. N. Fusaki, H. Ban, A. Nishiyama, K. Saeki, M. Hasegawa, Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **85**, 348–362 (2009).
56. X. Liang *et al.*, Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J. Biotechnol.* **208**, 44–53 (2015).
57. K. Clement *et al.*, CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
58. X. Wang *et al.*, CRISPR-DAV: CRISPR NGS data analysis and visualization pipeline. *Bioinformatics* **33**, 3811–3812 (2017).