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

Debojyoti Chakraborty (debojyoti.chakraborty@igib.in)
Institute of Genomics and Integrative Biology

Sundaram Acharya
Institute of Genomics and Integrative Biology https://orcid.org/0000-0001-5164-6328

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

Seiichi Hirano
University of Tokyo

Sajal Sarkar
Institute of Genomics and Integrative Biology

Riya Rauthan
Institute of Genomics and Integrative Biology

Manoj Kumar
Institute of Genomics and Integrative Biology

Rhythm Phutela
Institute of Genomics and Integrative Biology

Sneha Gulati
Institute of Genomics and Integrative Biology

C Afzal
Institute of Genomics and Integrative Biology

Deepanjan Paul
Institute of Genomics and Integrative Biology

Abdul Rahman
Institute of Genomics and Integrative Biology

Sudipta Mahato
LV Prasad Eye Institute

Savitri Maddileti
LV Prasad Eye Institute

Vinay Pulimamidi
LV Prasad Eye Institute

Subhadra Jalali

LV Prasad Eye Institute
Hiroshi Nishimasu
University of Tokyo
Indumathi Mariappan
LV Prasad Eye Institute https://orcid.org/0000-0001-7059-3030
Osamu Nureki
University of Tokyo https://orcid.org/0000-0003-1813-7008
Souvik Maiti
Institute of Genomics and Integrative Biology

Brief Communication

Keywords:

Posted Date: July 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3104171/v1

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 substrates[1–4]. Although high-fidelity versions of SpCas9 have been designed and validated in multiple systems, their editing efficiencies have generally dropped significantly as compared to the wild-type enzyme[5–7]. 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 basis[18]. 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 previously[4]. 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 Fig. 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 Fig. 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.

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

To discover the optimal length of gRNA for FnCas9, we performed an in vitro cleavage assay using a previously reported target DNA harboring a stretch of guanines with FnCas9 RNP containing variable length of gRNAs ranging from 20 to 24 nucleotides[32] (nt, hereafter referred g20-g24). Interestingly, we observed the lowest activity with the canonical g20 while all other extended length gRNAs exhibited enhanced DNA cleavage rate with g21 inducing the fastest rate of cleavage (Supplementary Fig. 2A). We used g21 in all our subsequent assays unless stated otherwise.

Next, we engineered 49 different FnCas9 variants guided by its crystal structure bearing mostly single amino acid substitutions in the WED-PI domain to introduce novel PAM duplex DNA contacts (Fig. 1A, Supplementary Table 3). We then measured *in vitro* DNA cleavage activities of the FnCas9 variants against a DNA target containing GGG PAM (where FnCas9 was shown to be least active)[3] and performed DNA cleavage experiments with the engineered variants. Recent reports have suggested that high-fidelity SpCas9 variants have slower enzyme kinetics and concomitant lower editing efficiencies[5,6]. FnCas9 being an enzyme with high fidelity, we therefore focused on engineered (en) FnCas9 variants showing a combination of faster cleavage rate and minimum structural alterations to ensure that its intrinsic specificity remains unchanged. A subset of nine enFnCas9 variants (containing single/combinatorial mutations) were selected for downstream experiments satisfying this criteria (Fig. 1B, Supplementary Note1). Among them, three variants (en1, en15 and en31) had at least 2-fold higher cleavage rates than the wild type protein (Fig. 1B). Intrigued by this observation, we tested cleavage efficiency of two of the enFnCas9 variants (en15 and en31) using super-extended gRNAs with 26 to 28-nt protospacer (g26-g28,

hereafter referred as sx-gRNA) and confirmed similar cleavage efficiencies as g21 which suggests the compatibility of enFnCas9 variants with sx-gRNAs (Supplementary Fig. 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 Fig. 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 Figs. 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-RY[35] and SpCas9-NG[31] (Fig. 1D, Supplementary Fig. 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 correction[4]. 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 Cas12a[38] or Cas12f[39]) or Type VI effectors (such as Cas13[40]), 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 (Fig. 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 (Fig. 1E). Expectedly, on a lateral flow strip, all enFnCas9 variants tested (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 Fig. 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 Fig. 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 Fig. 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 positions[4]. 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 (Fig. 2F, Supplementary Fig. 5D) with an improved signal discrimination (> 3.5-fold) as compared to FnCas9 (Supplementary Fig. 5E). We confirmed that the same specificity of SNV discrimination was also extended for an NGA PAM-containing substrate as well (Supplementary Fig. 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 earlier[4]. We found that these variants showed stronger DNA binding (K_d = 91.33 ± 29.8 nM for en1, K_d = 49.16 ± 10.96 nM for en15) as compared to FnCas9 (K_d = 170 ± 31.53 nM), with en15 showing ~ 3.5-fold higher DNA binding affinity (Supplementary Fig. 6B,C). Interestingly, in our previous study[4], 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 employed[44]. 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

(Fig. 2A, Supplementary Fig. 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 Fig. 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 Fig. 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 (Fig. 2B). Similarly, g21 gave robust genome editing outcomes (up to 90%) with all the enFnCas9 variants at other loci too (*EMX1* and *FASN*) (Supplementary Fig. 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 (Fig. 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%) (Fig. 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 (Fig. 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 (Fig. 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 reported[49] 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*, ~ 60% at *RUNX1* and ~ 20% at *ZNF629*) (Fig. 2D). Expectedly, g21 was able to induce editing outcomes wherever g20 failed to do (*ZNF629*) (Fig. 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 site[49], we were unable to detect any off-target editing with the en31 variant at any of these loci except at OT12 of *ZNF629* site (Fig. 2D). Despite being identical to the on-target of *ZNF629* site, off-targeting at OT12 was marginally detected on contrary to SpCas9 variant[49]. Furthermore, we also confirmed robust editing by en31 in one (~ 70% at *FANCF*) out of three sites having NAG PAM with g21 (Supplementary Fig. 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 (Fig. 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 alternatives[55]. 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 (Fig. 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 correction[55], 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 literature[59]. Given the larger share of ABE for pathogenic SNP correction[60], 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 literature[59]. 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 (Fig. 2G, Supplementary Fig. 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 (Fig. 2G, Supplementary Fig. 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 Fig. 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 reported[18]. 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 (Fig. 2I, Supplementary Fig. 7H-I). Moreover, since enFnCas9 can tolerate sx-gRNA such as sg26 or sg28 (Supplementary Fig. 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*) (Fig. 2J, Supplementary Fig. 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 (Fig. 2J, K; Supplementary Fig. 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 Fig. 7J).

Finally, towards the proof-of-concept validation of en31ABEmax8.17-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 modality[64]. 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 Fig. 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) (Fig. 2L). The hiPSC line was treated with en31ABE8.17d 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 Fig. 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) mRNA[66], 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 (Fig. 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 proteins[68].

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 enzyme[69]. 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 recently[70] (Supplementary Fig. 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 Fig. 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 (Fig. 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 editing[35]. However a recent biophysical study highlights ~ 1000 fold slower kinetics of SpRY as compared to SpCas9 and the possibility of off-target binding[71]. 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 (Fig. 2G, Supplementary Fig. 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 activity[73].

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 earlier[4]. 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 earlier[74] (Supplementary Fig. 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 OD_{600} 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 Fig. 5F.

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

In vitro cleavage (IVC) assay`

For the DNA cleavage study, PCR linearized pUC119 plasmid containing the target sequence and the respective PAM (mentioned in respective legends) was used as the substrate for in vitro cleavage experiments. The linearized pUC119 plasmid (50 ng or ~ 5nM) was incubated at 37°C for 0.5–5 min with the Cas9–sgRNA complex (50 nM) in 10 µL of reaction buffer, containing 20 mM HEPES, pH 7.5, 150 mM KCl, 10 mM $MgCl_2$, 1 mM DTT, and 5% glycerol. The reaction was stopped by the addition of a quenching buffer, containing EDTA (20 mM final) and Proteinase K (40 ng). The reaction products were resolved, visualized, and quantified with a MultiNA microchip electrophoresis device (SHIMADZU)[31].

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

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

PAM discovery assay

The PAM discovery assays were performed, as previously described[31]. Briefly, a library of pUC119 plasmids containing eight randomized nucleotides downstream of the target sequence was incubated at 37°C for 5 min with the FnCas9–sgRNA complex (50 nM), in 50 µL of the reaction buffer. The reactions were quenched by the addition of Proteinase K and then purified using a Wizard DNA Clean-Up System (Promega). The purified DNA samples were amplified for 25 cycles, using primers containing common adapter sequences. After column purification, each PCR product (~ 5 ng) was subjected to the second round of PCR for 15 cycles, to add custom Illumina TruSeq adapters and sample indices. The sequencing libraries were quantified by qPCR (KAPA Biosystems) and then subjected to paired-end sequencing on a MiSeq sequencer (Illumina) with 20% PhiX spike-in (Illumina). The sequencing reads were demultiplexed by primer sequences and sample indices, using NCBI Blast+ (version 2.8.1) with the blastn-short option. For each sequencing sample, the number of reads for every possible 8-nt PAM sequence pattern (4^8 = 65,536 patterns in total) was counted and normalized by the total number of reads in each sample. For a given PAM sequence, the enrichment score was calculated as 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 $MgCl_2$) 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 $MgCl_2$ and rested for 10 min at RT) for 10 min at 37°C. Wherever active en/FnCas9 was used, $MgCl_2$ was omitted from the buffer for making it catalytically inactive. After incubation, an 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.

Further background-corrected band intensity values were calculated through a smartphone application (TOPSE)[36,37].

Fluorescence assay (dFnCas9)

250nM biotin-labelled PCR amplicons carrying 580 bp long SARS-CoV-2 region with N501Y mutation were used for attaching DNA substrate to the wells of streptavidin coated plate by 10 mins incubation at room temperature. Wells were rinsed thrice with the wash buffer to get rid of the unbound amplicons (25mMTris-Cl, pH 7.2; 300mM NaCl; 0.1% BSA, 0.05% Tween®-20 Detergent) before using for the binding assay. dFnCas9-GFP RNP complex was pre-assembled in the binding buffer (20 mM HEPES, pH7.5, 150mM KCl, 1mM DTT, 10mM $MgCl_2$) by incubating 200nM dFnCas9-GFP with 200nM sgRNA for 10 min at room temperature. Reaction was initiated by adding pre-assembled RNP to the wells of 96-well streptavidin coated plate (Thermo Fisher Scientific; Cat 15119) pre-attached with biotin labeled amplicons and incubated at 37 C for 10 mins. Fluorescence was measured using a fluorescence plate reader (λ_{ex}: 485 nm; λ_{em}: 528 nm, transmission gain: optimal) (Tecan Infinite Pro F200).

Fluorescence assay (AaCas12b and Cas14a1)

AaCas12b and Cas14a1 RNP complexes were pre-assembled by incubating 200nM AaCas12b and Cas14a1 with 200 nM respective sgRNA for 10 min at room temperature. Reaction was initiated by adding pre-assembled RNP, 20nM ssDNA activator, 100ng background genomic DNA and 200 nM custom synthesized homopolymer ssDNA FQ reporter as described earlier (37, 44) (GenScript) in a cleavage buffer (40mM Tris-HCl, pH 7.5, 60mM NaCl, 6mM $MgCl_2$). The reaction was incubated in a 96-well flat bottom clear, black polystyrene microplate (Corning, cat no. CLS3603) at 37 C up to 180 mins with fluorescent measurements taken every 10 min (λ_{ex}: 485 nm; λ_{em}: 528 nm, transmission gain: optimal) using fluorescence plate reader (Tecan Infinite Pro F200). The resulting data were background-subtracted using the readings taken in the absence of ssDNA activator.

DNA Binding assay

MST was performed as described previously[4]. Briefly, dFnCas9-GFP and variant proteins were complexed with PAGE purified respective IVT sgRNAs (purified by 12% Urea-PAGE). The binding affinities of the Cas9 proteins and sgRNA RNP complexes were calculated using Monolith NT. 115 (NanoTemper Technologies GmbH, Munich, Germany). RNP complex (Protein:sgRNA molar ratio,1:1*)* was reconstituted at 25 for 10 mins in reaction buffer (20 mM HEPES, pH7.5, 150mM KCl, 1mM DTT, 10mM $MgCl_2$) 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 37^0 C temperature for 30 min in reaction buffer. The sample was loaded into NanoTemper standard treated capillaries and measurements were performed at 25°C using 20% LED power and 40% MST power. Data analyses were done using NanoTemper analysis software and the data were plotted by OriginLab.

Cell culture

HEK293T cells were grown in DMEM media supplemented with high glucose (Invitrogen), 2 mM GlutaMax, 10% FBS (Invitrogen), 1X antibiotic and antimycotic (Invitrogen) at 37°C in 5% CO2. Human iPS cells (LVP-F2-3F) were derived and maintained as described earlier[76]. 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 $5X10^4$ 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:

% NHEJ events = 100 X [1-(1-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^{+ve} 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 (Thermo Fisher). AMPure XP beads (A63881, Beckman Coulter) were used to separate out amplicons from free primers and primer dimers. Dual indexing was done using Nextera XT V2 index kit followed by a second round of bead-based purification. The libraries were quantified using a Qubit dsDNA HS Assay kit (Invitrogen, Q32853) and were also loaded on agarose gel for the qualitative check. Libraries were normalized, pooled and were loaded onto the Illumina MiniSeq platform.

HDR assay at DCX locus in HEK293T

HEK293T cells were cultured in DMEM with GlutaMAX supplement (ThermoFisher Scientific Cat. No. 10566016) with 10% FBS serum. 70%-80% confluent HEK293T cells were harvested from a 6 well plate using Trypsin-EDTA (0.05%) (ThermoFisher Scientific Cat. No.: 25300062) and pipetted to make a single-cell suspension. For each electroporation reaction, a total 15ug plasmid was mixed in Resuspension buffer R, in which linearized donor plasmid DNA and Cas9-gRNA vector were taken in a 1:2 ratio. 6 x 10^5 cells were resuspended in 100µL of Resuspension Buffer R containing plasmids and electroporation was performed using Neon Transfection System 100µL 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℃ for 5 min followed by 40 amplification cycles of 95℃ for 10 sec; 63℃ for 30 sec; 72℃ for 30 sec. Log_2 fold change values were calculated by the 2^ΔΔCt method for each sample with respect to untransfected control. A non-targeting region in genomic DNA was used for normalization.

ChIP Seq analysis

Raw sequencing reads were mapped to the human reference genome GRCh38 using bowtie2[77]. Peaks were called over input samples using MACS2[78] with default parameters. Finally, scrambled sample peaks were used to remove background and false positive peaks from the dSpCas9 and dFnCas9 test samples. These filtered peaks were searched for off-targets based on sgRNA sequence homology with a maximum of 6 mismatches. On-target peak coverage plots were generated by the fluff profiles command with ‘remove duplicates’ option[79]. Overlap between the dSpCas9 and dFnCas9 ChIP peaks were calculated using bedtools[80] and plotted as weighted Venn diagrams with the help of Intervene[81].

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 v2.0.45[82] 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 ballon 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 earlier[83]. 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% CO_2 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.

Declarations

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.

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.

References

1. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156, 935–949 (2014).
2. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513, 569–573 (2014).
3. Hirano, H. et al. Structure and Engineering of Francisella novicida Cas9. Cell 164, 950–961 (2016).
4. Acharya, S. et al. Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. Proc. Natl. Acad. Sci. U. S. A. 116, 20959–20968 (2019).
5. Kim, N. et al. Prediction of the sequence-specific cleavage activity of Cas9 variants. Nat. Biotechnol. 38, 1328–1336 (2020).
6. Liu, M.-S. et al. Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. Nat. Commun. 11, 3576 (2020).
7. Kim, Y.-H. et al. Sniper2L is a high-fidelity Cas9 variant with high activity. Nat. Chem. Biol. (2023) doi:10.1038/s41589-023-01279-5.
8. Liu, J.-J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. Nature 566, 218–223 (2019).
9. Pausch, P. et al. CRISPR-CasΦ from huge phages is a hypercompact genome editor. Science 369, 333–337 (2020).
10. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015).
11. Karvelis, T. et al. PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. Nucleic Acids Research vol. 48 5016–5023 Preprint at https://doi.org/10.1093/nar/gkaa208 (2020).

12. Chatterjee, P., Jakimo, N. & Jacobson, J. M. Minimal PAM specificity of a highly similar SpCas9 ortholog. Sci Adv 4, eaau0766 (2018).
13. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. Nat. Commun. 12, 555 (2021).
14. Schmidt, M. J. et al. Improved CRISPR genome editing using small highly active and specific engineered RNA-guided nucleases. Nat. Commun. 12, 4219 (2021).
15. Kim, D. Y. et al. Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. Nat. Biotechnol. 40, 94–102 (2022).
16. Wu, Z. et al. Programmed genome editing by a miniature CRISPR-Cas12f nuclease. Nat. Chem. Biol. 17, 1132–1138 (2021).
17. Xu, X. et al. Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. Mol. Cell 81, 4333–4345.e4 (2021).
18. Newby, G. A. et al. Base editing of haematopoietic stem cells rescues sickle cell disease in mice. Nature 595, 295–302 (2021).
19. Chen, F. et al. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. Nat. Commun. 8, 14958 (2017).
20. Tycko, J. et al. Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. Nat. Commun. 10, 4063 (2019).
21. Chen, J. S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature 550, 407–410 (2017).
22. Jones, S. K., Jr et al. Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. Nat. Biotechnol. 39, 84–93 (2021).
23. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nat. Biotechnol. 32, 670–676 (2014).
24. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat. Biotechnol. 32, 677–683 (2014).
25. O'Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F. & Segal, D. J. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. Nucleic Acids Res. 43, 3389–3404 (2015).
26. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507, 62–67 (2014).
27. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature 527, 110–113 (2015).
28. Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. Sci Adv 3, eaao0027 (2017).
29. Okafor, I. C. et al. Single molecule analysis of effects of non-canonical guide RNAs and specificity-enhancing mutations on Cas9-induced DNA unwinding. Nucleic Acids Res. 47, 11880–11888 (2019).

30. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. Cell Rep. 22, 359–371 (2018).
31. Nishimasu, H. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. Science 361, 1259–1262 (2018).
32. Hirano, S. et al. Structural basis for the promiscuous PAM recognition by Corynebacterium diphtheriae Cas9. Nat. Commun. 10, 1968 (2019).
33. Hirano, S., Nishimasu, H., Ishitani, R. & Nureki, O. Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. Mol. Cell 61, 886–894 (2016).
34. Anders, C., Bargsten, K. & Jinek, M. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. Mol. Cell 61, 895–902 (2016).
35. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. Science 368, 290–296 (2020).
36. Azhar, M. et al. Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis. Biosensors and Bioelectronics 183, 113207 (2021).
37. Kumar, M. et al. FnCas9-based CRISPR diagnostic for rapid and accurate detection of major SARS-CoV-2 variants on a paper strip. Elife 10, (2021).
38. Chen, J. S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science 360, 436–439 (2018).
39. Harrington, L. B. et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science 362, 839–842 (2018).
40. Gootenberg, J. S. et al. Nucleic acid detection with CRISPR-Cas13a/C2c2. Science 356, 438–442 (2017).
41. Kleinstiver, B. P. et al. Publisher Correction: Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat. Biotechnol. 38, 901 (2020).
42. Kleinstiver, B. P. et al. Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. Nat. Biotechnol. 37, 276–282 (2019).
43. Wei, Y. et al. Indiscriminate ssDNA cleavage activity of CRISPR-Cas12a induces no detectable off-target effects in mouse embryos. Protein Cell 12, 741–745 (2021).
44. Murugan, K., Seetharam, A. S., Severin, A. J. & Sashital, D. G. CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. J. Biol. Chem. 295, 5538–5553 (2020).
45. Kleinstiver, B. P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529, 490–495 (2016).
46. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. Science 351, 84–88 (2016).
47. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187–197 (2015).

48. Casini, A. et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nat. Biotechnol. 36, 265–271 (2018).
49. Kleinstiver, B. P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523, 481–485 (2015).
50. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. Nat. Med. 24, 927–930 (2018).
51. Ihry, R. J. et al. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. Nat. Med. 24, 939–946 (2018).
52. Leibowitz, M. L. et al. Chromothripsis as an on-target consequence of CRISPR-Cas9 genome editing. Nat. Genet. 53, 895–905 (2021).
53. Cullot, G. et al. CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. Nat. Commun. 10, 1136 (2019).
54. Zuccaro, M. V. et al. Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. Cell 183, 1650–1664.e15 (2020).
55. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. Nat. Biotechnol. 38, 824–844 (2020).
56. Kim, D., Kang, B.-C. & Kim, J.-S. Identifying genome-wide off-target sites of CRISPR RNA-guided nucleases and deaminases with Digenome-seq. Nat. Protoc. 16, 1170–1192 (2021).
57. Zhao, Z., Shang, P., Mohanraju, P. & Geijsen, N. Prime editing: advances and therapeutic applications. Trends Biotechnol. (2023) doi:10.1016/j.tibtech.2023.03.004.
58. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol. 36, 843–846 (2018).
59. Gaudelli, N. M. et al. Directed evolution of adenine base editors with increased activity and therapeutic application. Nat. Biotechnol. 38, 892–900 (2020).
60. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. Nat. Rev. Genet. 19, 770–788 (2018).
61. Li, C. et al. In vivo HSPC gene therapy with base editors allows for efficient reactivation of fetal γ-globin in β-YAC mice. Blood Adv 5, 1122–1135 (2021).
62. Landrum, M. J. et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 42, D980–5 (2014).
63. Landrum, M. J. et al. ClinVar: improvements to accessing data. Nucleic Acids Res. 48, D835–D844 (2020).
64. Kamao, H. et al. Characterization of human induced pluripotent stem cell-derived retinal pigment epithelium cell sheets aiming for clinical application. Stem Cell Reports 2, 205–218 (2014).
65. Ehrke-Schulz, E. et al. CRISPR/Cas9 delivery with one single adenoviral vector devoid of all viral genes. Sci. Rep. 7, 17113 (2017).

66. Han, J. P. et al. In vivo delivery of CRISPR-Cas9 using lipid nanoparticles enables antithrombin gene editing for sustainable hemophilia A and B therapy. Sci Adv 8, eabj6901 (2022).
67. Chew, W. L. et al. A multifunctional AAV-CRISPR-Cas9 and its host response. Nat. Methods 13, 868–874 (2016).
68. Mekler, V., Kuznedelov, K. & Severinov, K. Quantification of the affinities of CRISPR–Cas9 nucleases for cognate protospacer adjacent motif (PAM) sequences. Journal of Biological Chemistry vol. 295 6509–6517 Preprint at https://doi.org/10.1074/jbc.ra119.012239 (2020).
69. Bravo, J. P. K. et al. Publisher Correction: Structural basis for mismatch surveillance by CRISPR-Cas9. Nature 604, E10 (2022).
70. Kulcsár, P. I., Tálas, A., Ligeti, Z., Krausz, S. L. & Welker, E. SuperFi-Cas9 exhibits remarkable fidelity but severely reduced activity yet works effectively with ABE8e. Nat. Commun. 13, 6858 (2022).
71. Hibshman, G. N. et al. Unraveling the mechanisms of PAMless DNA interrogation by SpRY Cas9. bioRxiv 2023.06.22.546082 (2023) doi:10.1101/2023.06.22.546082.
72. Richter, M. F. et al. Author Correction: Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. Nat. Biotechnol. 38, 901 (2020).
73. Oh, Y. et al. Expansion of the prime editing modality with Cas9 from Francisella novicida. Genome Biol. 23, 92 (2022).
74. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).
75. Ansari, A. H., Kumar, M., Sarkar, S., Maiti, S. & Chakraborty, D. CriSNPr, a single interface for the curated and de novo design of gRNAs for CRISPR diagnostics using diverse Cas systems. Elife 12, (2023).
76. Susaimanickam, P. J. et al. Generating minicorneal organoids from human induced pluripotent stem cells. Development 144, 2338–2351 (2017).
77. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
78. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137 (2008).
79. Georgiou, G. & van Heeringen, S. J. fluff: exploratory analysis and visualization of high-throughput sequencing data. PeerJ 4, e2209 (2016).
80. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
81. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics 18, 287 (2017).
82. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. 37, 224–226 (2019).
83. Shukla, R., Kannabiran, C. & Jalali, S. Genetics of Leber congenital amaurosis: an update. Expert Rev. Ophthalmol. (2014) doi:10.1586/eop.12.14.

84. Okita, K. et al. A more efficient method to generate integration-free human iPS cells. Nat. Methods 8, 409–412 (2011).
85. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. 29, 2860–2874 (2001).

Figures

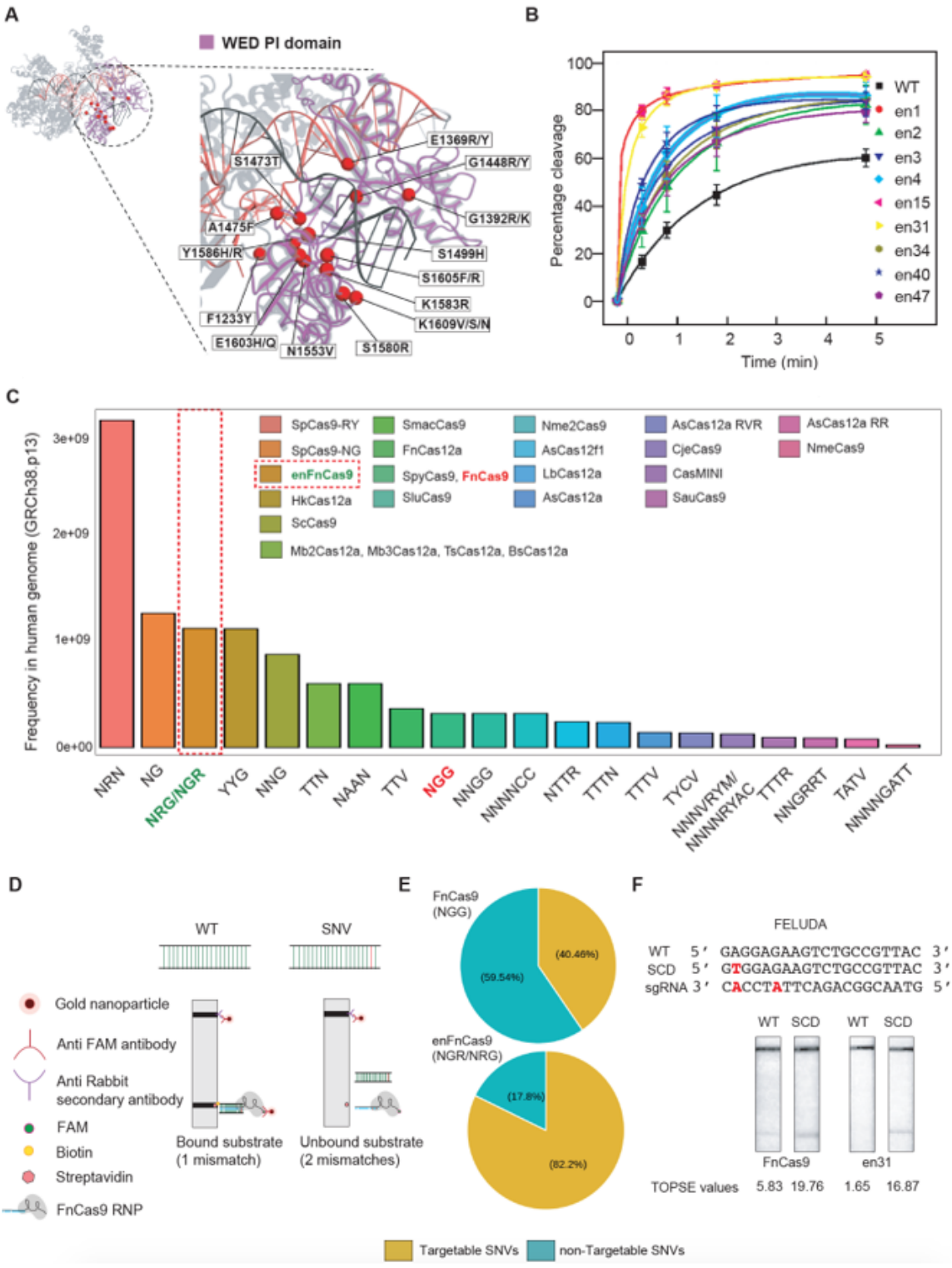

Figure 1

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

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 ± 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.

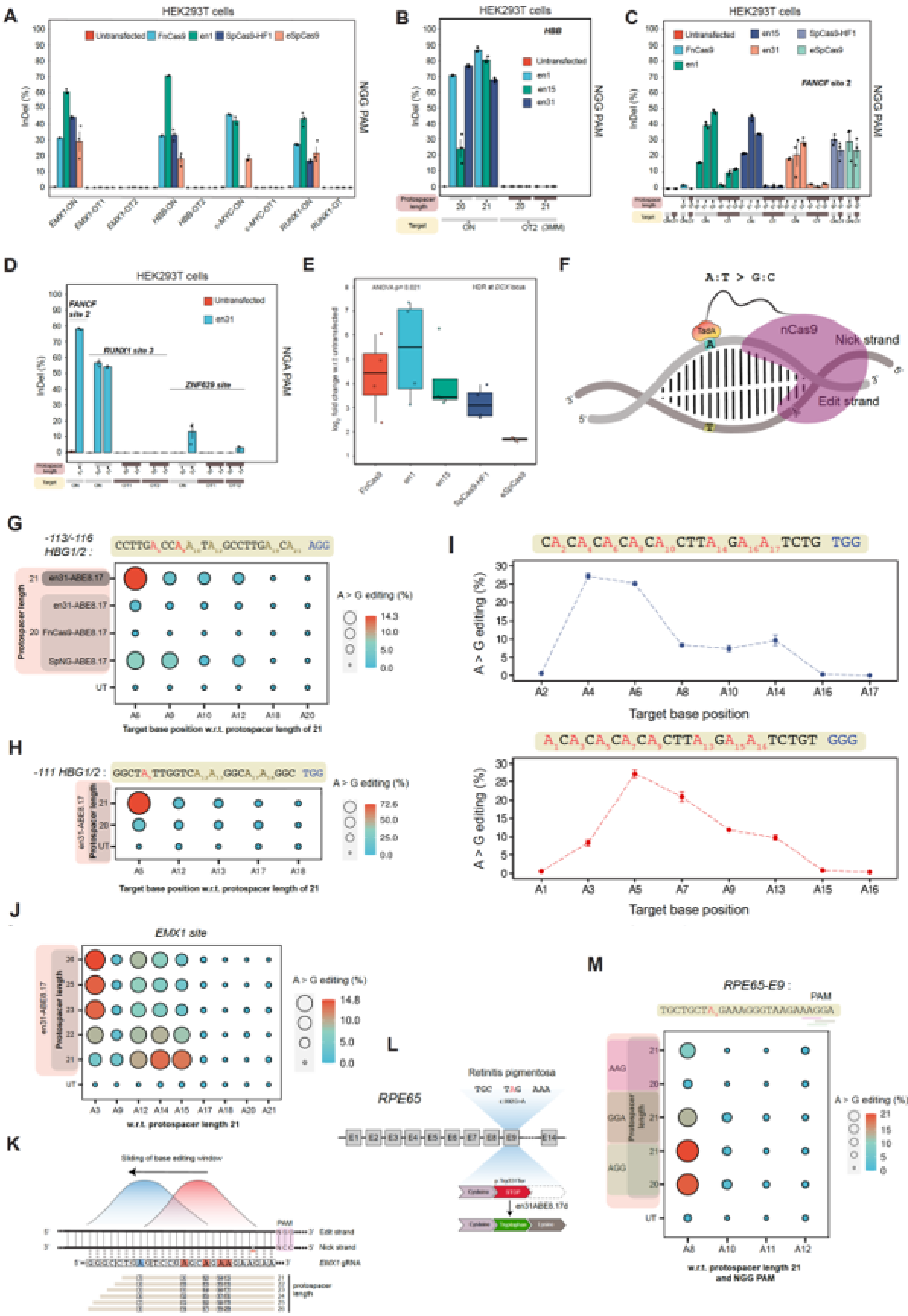

Figure 2

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

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.

Supplementary Files

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

- reportingsummary.pdf
- EditorialPolicy.pdf
- SupplementaryFigures.docx