Signed details of the excellence in research work for which the Sun Pharma Research Award is claimed, including references & illustra- tions (Max. 2.5 MB). The candidate should duly sign on the details *

Broad Overview:

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.

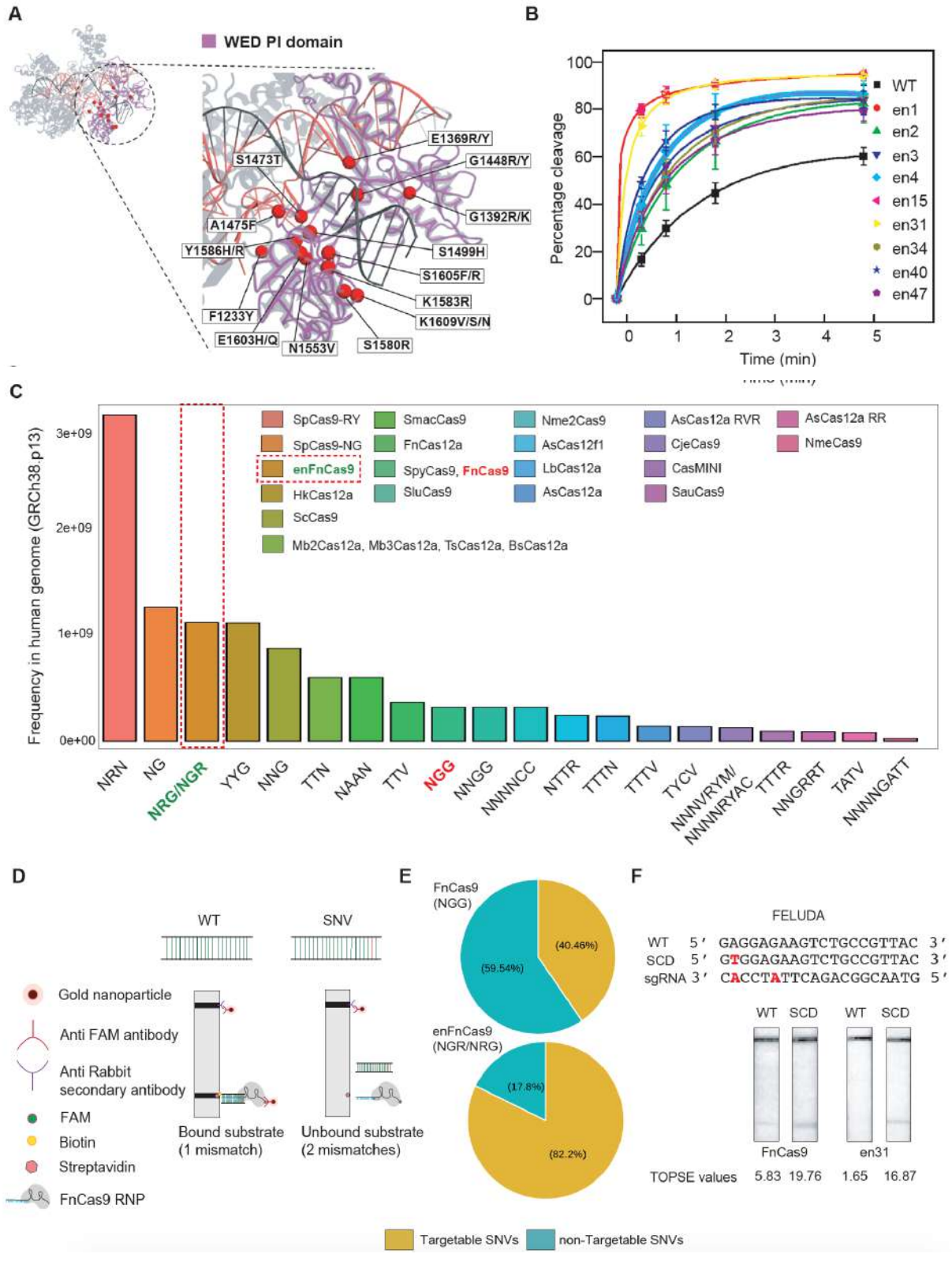

A
WED PI domain
E1369R/Y
G1448R/Y
S1473T
G1392R/K
A1475F
Y1586H/R
S1499H
S1605F/R
K1583R
F1233Y
K1609V/S/N
E1603H/Q
N1553V
S1580R
B
Percentage cleavage
Time (min)
WT
en1
en2
en3
en4
en15
en31
en34
en40
en47
C
Frequency in human genome (GRCh38.p13)
SpCas9-RY
SpCas9-NG
enFnCas9
HkCas12a
ScCas9
Mb2Cas12a, Mb3Cas12a, TsCas12a, BsCas12a
SmacCas9
FnCas12a
SpyCas9, FnCas9
SluCas9
Nme2Cas9
AsCas12f1
LbCas12a
AsCas12a
AsCas12a RVR
CjeCas9
CasMINI
SauCas9
AsCas12a RR
NmeCas9
NRN
NG
NRG/NGR
YYG
NNG
TTN
NAAN
TTV
NGG
NNGG
NNNNCC
NTTR
TTTN
TTTV
TYCV
NNNVRYM/
NNNNRYAC
TTTR
NNGRRT
TATV
NNNNGATT
D
WT
SNV
Gold nanoparticle
Anti FAM antibody
Anti Rabbit
secondary antibody
FAM
Biotin
Streptavidin
FnCas9 RNP
Bound substrate
(1 mismatch)
Unbound substrate
(2 mismatches)
E
FnCas9
(NGG)
(40.46%)
(59.54%)
enFnCas9
(NGR/NRG)
(17.8%)
(82.2%)
F
FELUDA
WT 5′ GAGGAGAAGTCTGCCGTTAC 3′
SCD 5′ GTGGAGAAGTCTGCCGTTAC 3′
sgRNA 3′ CACCTATTCAGACGGCAATG 5′
WT SCD WT SCD
FnCas9
en31
TOPSE values 5.83 19.76 1.65 16.87
Targetable SNVs
non-Targetable SNVs

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 ± 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.

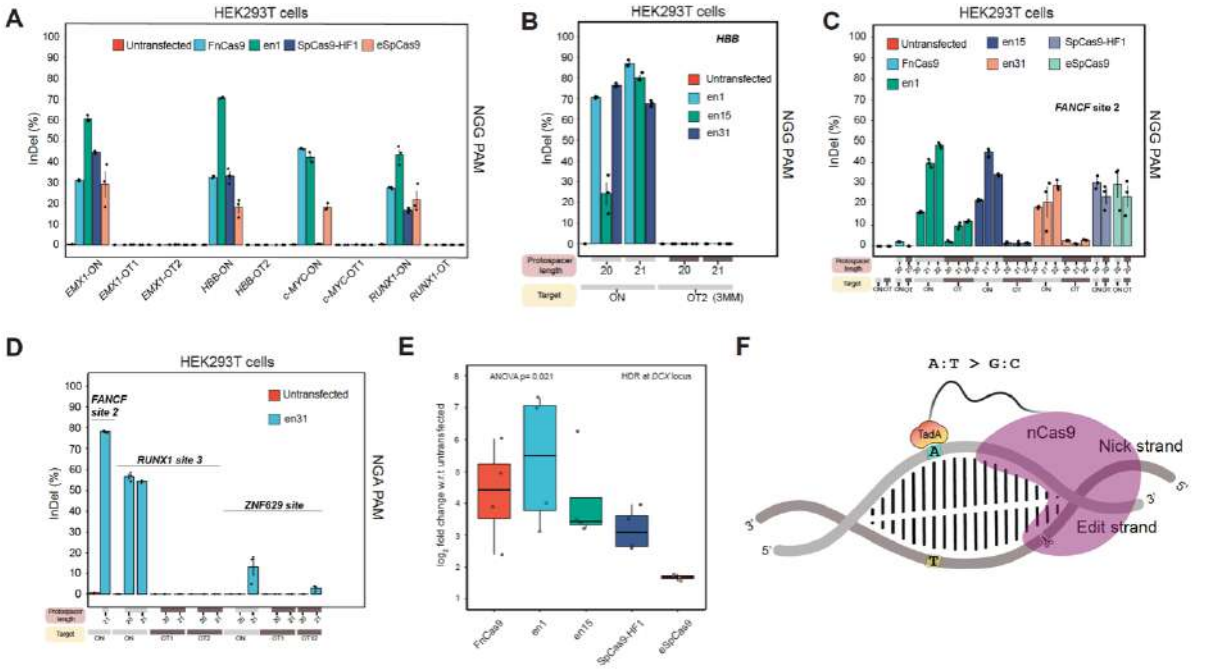

A
HEK293T cells
Untransfected
FnCas9
en1
SpCas9-HF1
eSpCas9
InDel (%)
NGG PAM
EMX1-ON
EMX1-OT1
EMX1-OT2
HBB-ON
HBB-OT2
c-MYC-ON
c-MYC-OT1
RUNX1-ON
RUNX1-OT
B
HEK293T cells
HBB
Untransfected
en1
en15
en31
InDel (%)
NGG PAM
Protospacer length
Target
ON
OT2 (3MM)
C
HEK293T cells
Untransfected
FnCas9
en1
en15
en31
SpCas9-HF1
eSpCas9
FANCF site 2
InDel (%)
NGG PAM
D
HEK293T cells
FANCF site 2
RUNX1 site 3
ZNF629 site
Untransfected
en31
InDel (%)
NGA PAM
E
ANOVA p= 0.021
HDR at DCK locus
log2 fold change w.r.t. untransfected
FnCas9
en1
en15
SpCas9-HF1
eSpCas9
F
A:T > G:C
TadA
nCas9
Nick strand
Edit strand
G
-113/-116 HBG1/2 :
CCTTGACCAAATATGCCTTGACACAGG
Protospacer length
en31-ABE8.17
en31-ABE8.17
FnCas9-ABE8.17
SpNG-ABE8.17
UT
A > G editing (%)
Target base position w.r.t. protospacer length of 21
H
-111 HBG1/2 :
GGCTATTGGTCAAAGGCAAAGGC TGG
en31-ABE8.17 Protospacer length
A > G editing (%)
Target base position w.r.t. protospacer length of 21
I
CACACACACACACTTAGAAATCTG TGG
A > G editing (%)
Target base position
ACACACACACTTAGAATCTGT GGG
A > G editing (%)
Target base position
J
EMX1 site
en31-ABE8.17 Protospacer length
UT
A > G editing (%)
w.r.t. protospacer length 21
K
Sliding of base editing window
PAM
Edit strand
Nick strand
EMX1 gRNA
protospacer length
L
Retinitis pigmentosa
TGC TAG AAA
RPE65
E1 E2 E3 E4 E5 E6 E7 E8 E9 E10
Cysteine
STOP
en31ABE8.17d
Cysteine
Tryptophan
Cysteine
M
RPE65-E9 :
PAM
TGCTGCTAGAAAGGGTAAGAAAGGA
AAG
GGA
AGG
Protospacer length
UT
A > G editing (%)
w.r.t. protospacer length 21 and NGG PAM

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.

enFnCas9 can achieve only exclusive HDR (Homology directed repair) mediated insertion at the therapeutically relevant sickle cell anemia mutation locus in human CD34+ cells without any associated NHEJ (Non Homologous End Joining) based insertion/deletion mutations. To our knowledge this has not been achieved in any reported study (Dewitt et al. 2016, Pattabhi et al. 2019, Wilkinson et al. 2021 etc.). Precise correction of the sickle cell locus (with minimum indels) is therapeutically critical because indels caused due to Cas9 breaks lead to the risk of beta thalassemia mutations being created at the HBB site. Although such a strategy has been attempted before using ssODN or AAV (Magis et al. 2022) and recently in a clinical trial (CEDAR Trial, NCT04819841) every report has show far shown higher level of indels over precise gene correction limiting its clinical application. Importantly, the CEDAR trial (GRAPHITE Bio) has also been recently halted due to complications in the recipient post therapy.
We took cues from the fact that FnCas9 produces staggered DNA ends (Acharya et al., 2019; Chen et al. 2017) and that staggered DNA ends might lead to higher

HDR:NHEJ (Wang et al. 2018), and now show that in human CD34+ cells, enFnCas9 can reach > 80% precise insertion of a donor DNA template at the sickle cell locus with no detectable indels. We will be happy to share this human cell data in the revised manuscript to further emphasize the clinical application of enFnCas9 variants to two disease specific loci (sickle cell anemia and RPE65) and two different patient specific cells (CD34+ hematopoietic stem cells and iPSCs). In our opinion, the exclusive HDR outcome at the sickle cell disease loci further underscores the novelty and uniqueness of enFnCas9 among other engineered Cas9 proteins reported so far.

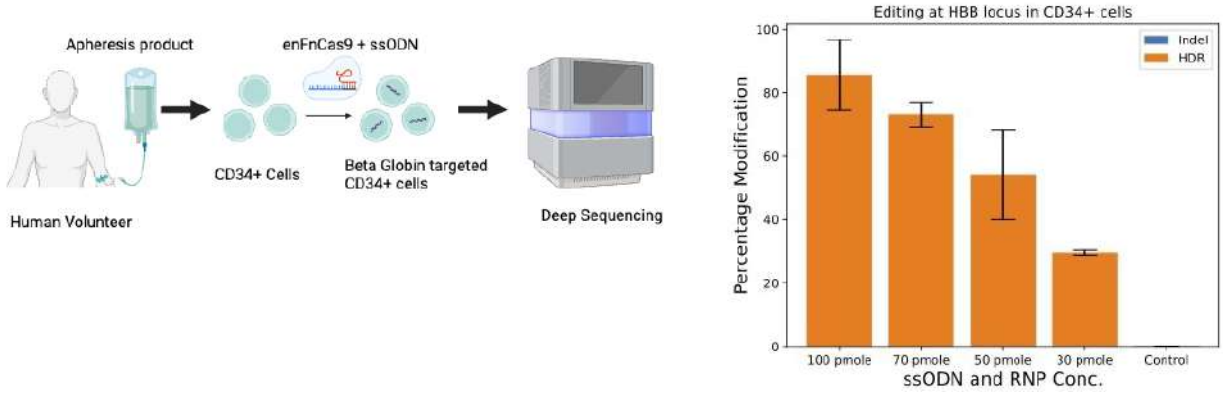


Figure 3: ssODN mediated HDR in Human CD34 cells at the Beta Globin Locus showing percentage HDR conversion with different concentrations of en1FnCas9 RNP used. Left, schematic of the experiment. Right, HDR percentages at the target locus.

enFnCas9ABE mediated correction of a retinal disease: A immunofluorescence and complete restoration of RPE65 upon iPSC differentiation as a result of the gene correction to conclusively establish the therapeutic value of enFnCas9-ABE mediated gene correction.

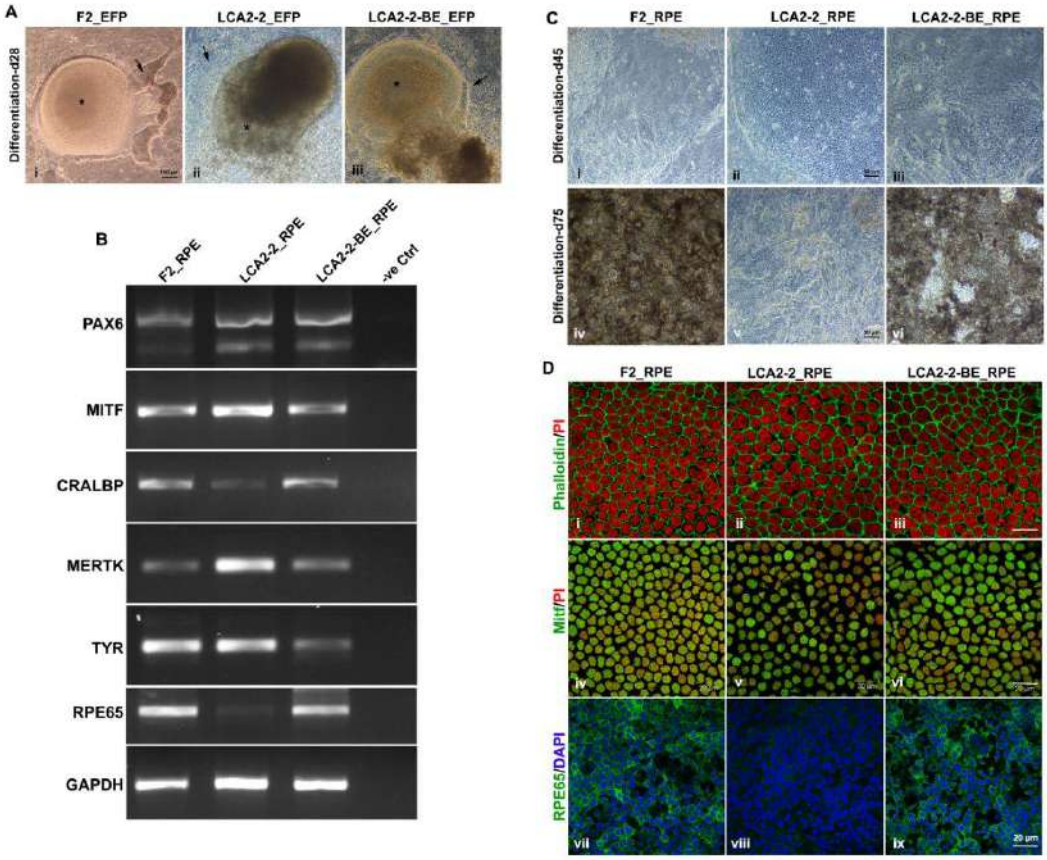
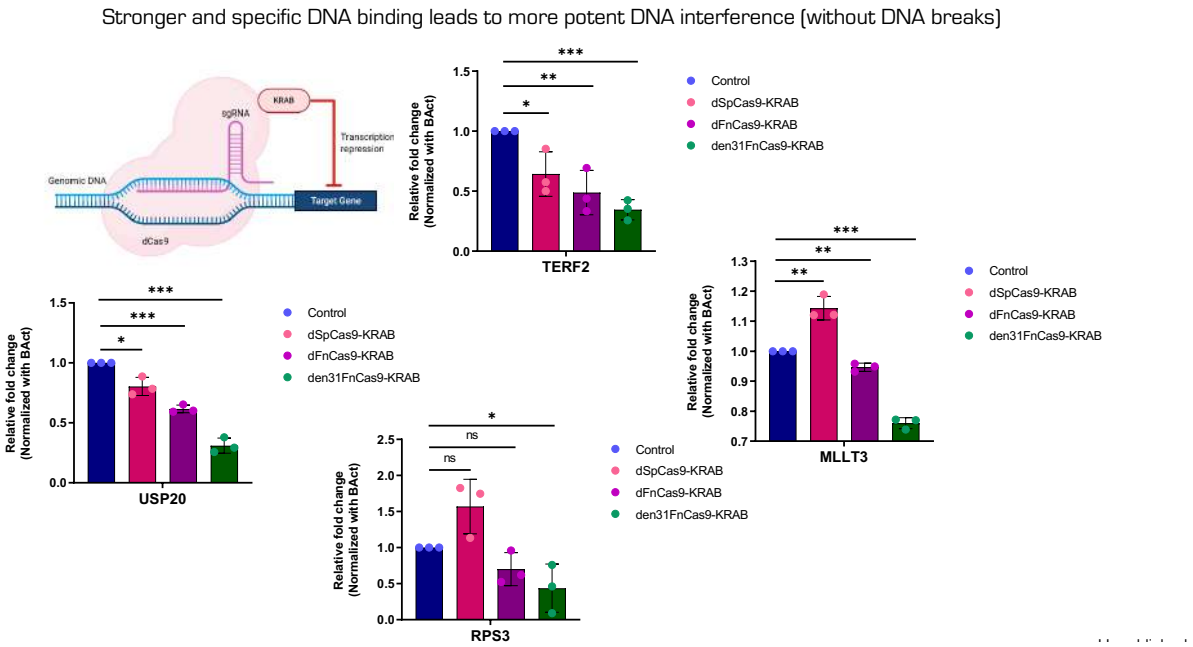


Figure 4 : Retinal differentiation of iPSCs and confirmation of RPE65 expression in iPSC-RPE cells. A. Morphology of eye field primordial (EFP) clusters derived from healthy control iPSC (F2), patient-specific mutant iPSC (LCA-2-2) and the mutation corrected patient-specific iPSC (LCA2-2-BE) at 3-4 weeks of retinal lineage differentiation. Asterisk (*) marks the central neuro-retinal island and black arrows indicate the migrating retinal pigmented epithelium. B. Agarose gel images of gene transcripts of RPE-specific markers, PAX6, MITF, CRALBP, MERTK, TYR and RPE65, as analysed by semi-quantitative RT-PCR in d45 iPSC-RPE cells. C. Phase contrast imaging and lower magnification view of the cellular morphology and levels of pigmentation in d45 and d75 iPSC derived RPE cell cultures. D. Fluorescent immunolabelling and higher magnification view of the RPE cell morphology and the expression of RPE-specific markers, as marked by Phalloidin-FITC (green), PI (red) (i-iii); MITF (green), PI (red) (iv-vi); and RPE65 (green), DAPI (blue) (vii-ix).

Conjugation of enFnCas9 with KRAB leads to highly potent DNA double strand break-free reduction of transcription



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).

Signed

Dr.Debojyoti Chakraborty