Supplemental information for

Versatile and robust genome editing with *Streptococcus thermophilus* CRISPR1-Cas9

Daniel Agudelo[1], Sophie Carter[1], Minja Velimirovic[1], Alexis Duringer[1], Jean-François Rivest[1], Sébastien Levesque[1], Jeremy Loehr[1], Mathilde Mouchiroud[2], Denis Cyr[3], Paula J Waters[3], Mathieu Laplante[2,4], Sylvain Moineau[5,6,7], Adeline Goulet[8,9], and Yannick Doyon[1,4].

Supplemental items:

Supplemental Fig. S1. Engineered CRISPR1-StCas9 system drives robust gene editing in mammalian cells.

Supplemental Fig. S2. Engineered CRISPR1-StCas9 sgRNA improves dSt1Cas9-VPR transcriptional activation in human cells.

Supplemental Fig. S3. St1Cas9 strain variants display unique PAM specificities.

Supplemental Fig. S4. St1Cas9 strain variants are incompatible with SaCas9 PAMs and differential inhibition of St1Cas9 and SaCas9 by anti-CRISPR proteins.

Supplemental Fig. S5. St1Cas9BE4max variants display unique and orthogonal PAM specificities.

Supplemental Fig. S6. *In vivo* genome editing using St1Cas9.

Supplemental Fig. S7. Examples of pathogenic alleles found in patients with inherited monogenic diseases potentially targetable by St1Cas9s in a mutation-specific manner and targeting the mouse albumin gene.

Supplemental Fig. S8. Examples of pathogenic alleles found in patients with inherited monogenic diseases potentially targetable by St1Cas9-derived base editors.

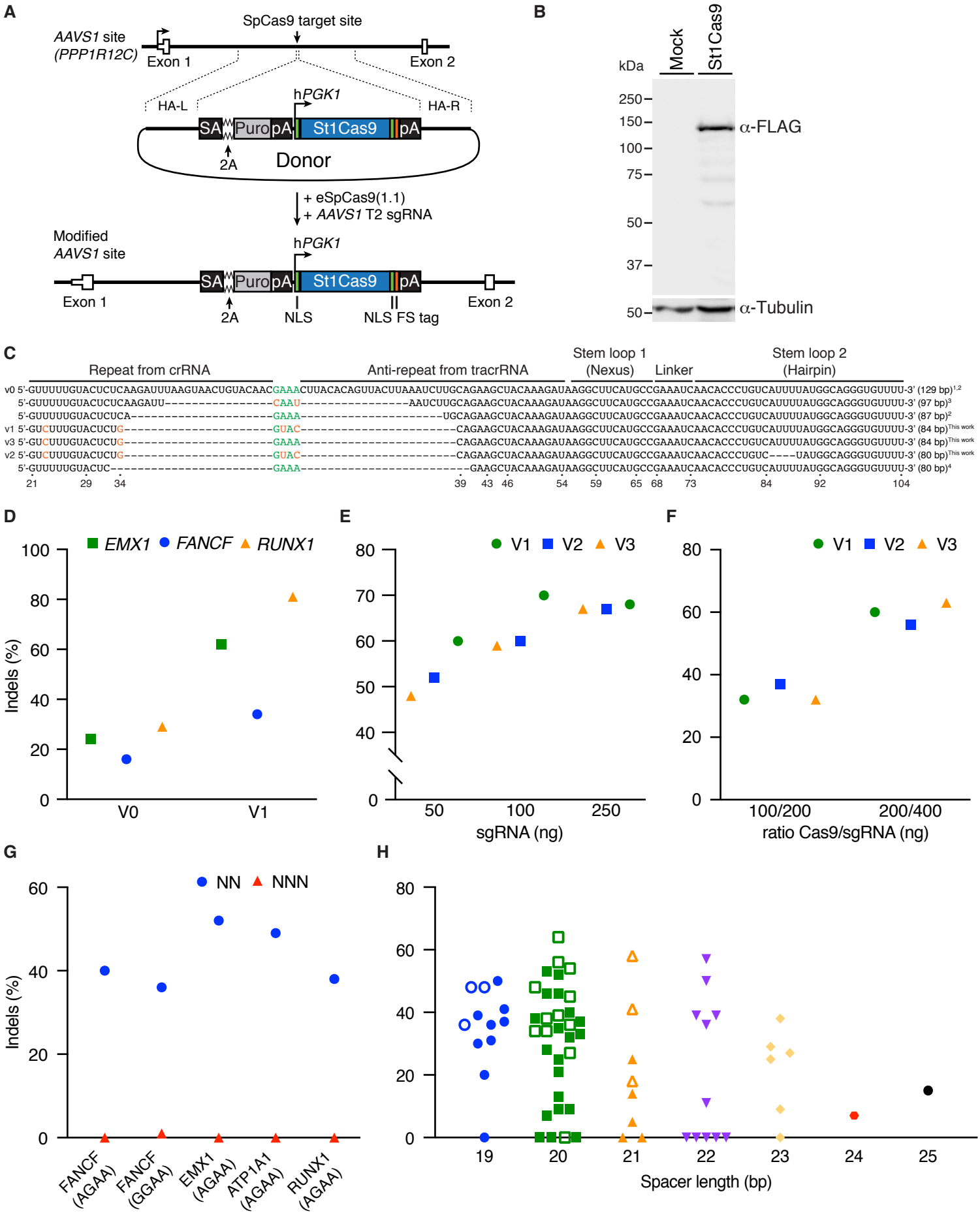
Supplemental Fig. S9. St1Cas9 vectors available from Addgene.

Supplemental Fig. S10. SDS-PAGE gel analysis of purified rAAV8 vectors used in this study.

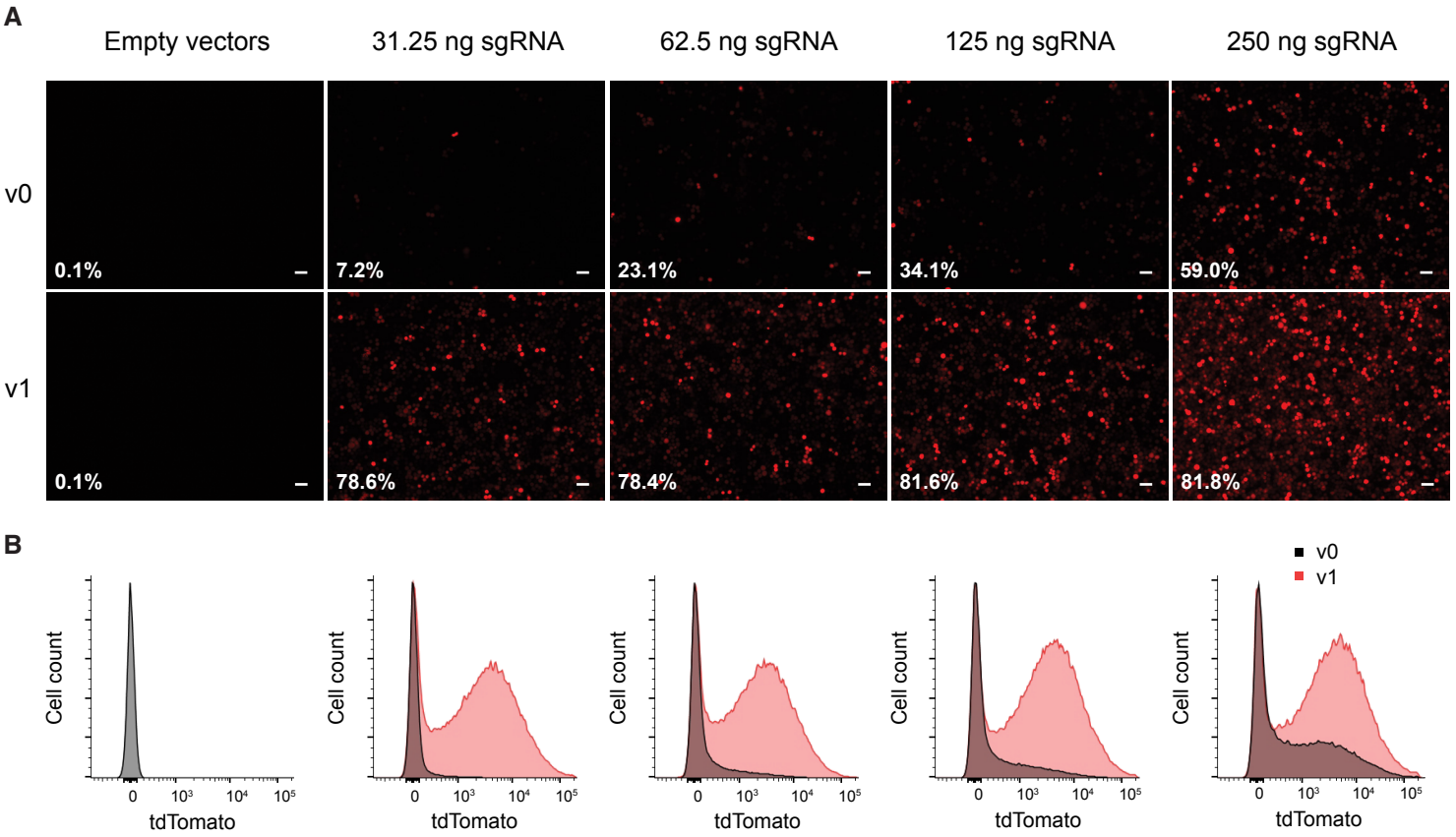
Supplemental Tables S1-19. Detailed list of guides tested with St1Cas9 orthologs and SaCas9, PCR primers, complete sequence of nucleases, base editors, promoters, sgRNAs. Screenshots of base editing chromatograms.

Supplemental Methods. Methods related to Supplemental Figs. S2,6.

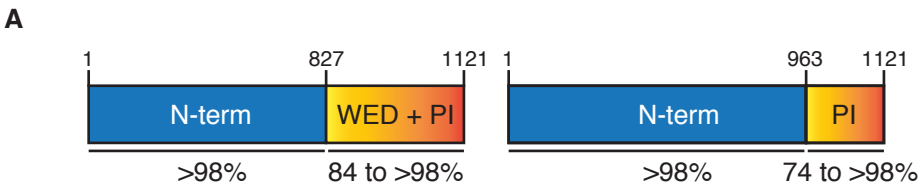
Supplemental Data. Raw Sanger sequencing data generated in this study.

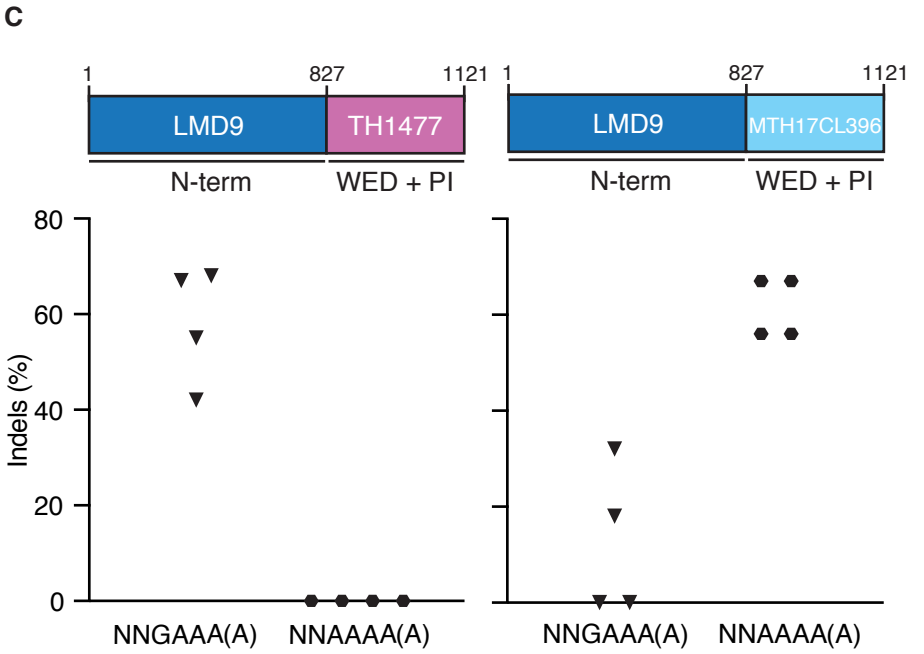

A
AAVS1 site
(PPP1R12C)
SpCas9 target site
Exon 1
Exon 2
HA-L
HA-R
hPGK1
SA
Puro
pA
St1Cas9
pA
2A
Donor
+ eSpCas9(1.1)
+ AAVS1 T2 sgRNA
Modified
AAVS1 site
Exon 1
2A
NLS
NLS FS tag
Exon 2
B
kDa
Mock
St1Cas9
250
150
100
75
50
37
50
α-FLAG
α-Tubulin
C
Repeat from crRNA
Anti-repeat from tracrRNA
Stem loop 1
(Nexus)
Linker
Stem loop 2
(Hairpin)
v0 5'-GUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAACGAAACUUACACAGUUACUUAAAUCUUGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU-3' (129 bp)[1,2]
5'-GUUUUUGUACUCUCAAGAUU----------CAAU----------AAUCUUGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU-3' (97 bp)[3]
5'-GUUUUUGUACUCUCA----------GAAA----------UGCAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU-3' (87 bp)[2]
v1 5'-GUCUUUGUACUCUG----------GUAC----------CAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU-3' (84 bp)[This work]
v3 5'-GUCUUUGUACUCUG----------GAAA----------CAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU-3' (84 bp)[This work]
v2 5'-GUCUUUGUACUCUG----------GUAC----------CAGAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUC----UAUGGCAGGGUGUUUU-3' (80 bp)[This work]
5'-GUUUUUGUACUC----------GAAA----------GAAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUAUGGCAGGGUGUUUU-3' (80 bp)[4]
21 29 34 39 43 46 54 59 65 68 73 84 92 104
D
EMX1
FANCF
RUNX1
Indels (%)
V0
V1
E
V1
V2
V3
sgRNA (ng)
F
V1
V2
V3
100/200
200/400
ratio Cas9/sgRNA (ng)
G
NN
NNN
Indels (%)
FANCF (AGAA)
FANCF (GGAA)
EMX1 (AGAA)
ATP1A1 (AGAA)
RUNX1 (AGAA)
H
Spacer length (bp)

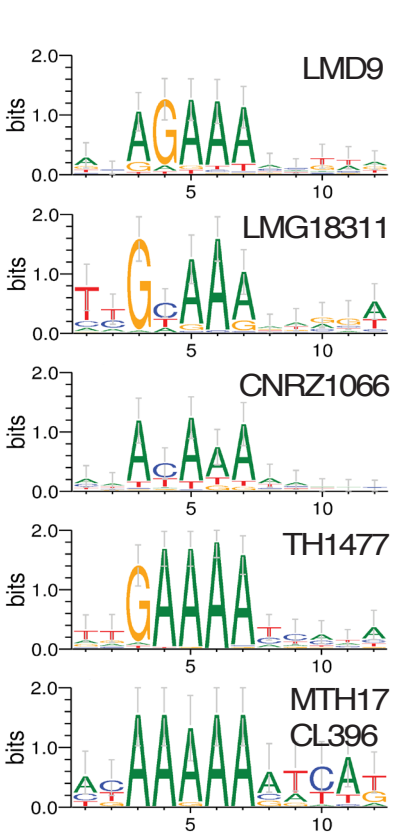
Supplemental Fig. S1. Related to Fig. 1. Engineered CRISPR1-StCas9 system drives robust gene editing in mammalian cells. (A) Schematic representation of the targeted integration of tagged St1Cas9 to the *AAVS1* safe harbor locus. The donor construct and the locus following cDNA addition are displayed. The first two exons of the *PPP1R12C* gene are shown as open boxes. Also annotated are the locations of the splice acceptor site (SA), 2A self-cleaving peptide sequence (2A), puromycin resistance gene (Puro), polyadenylation sequence (pA), human phosphoglycerate kinase 1 promoter (h*PGK1*), nuclear localization signals (NLS), and 3xFLAG-2xSTREP tandem affinity tag (Tag), homology arms left and right (HA-L, HA-R) are respectively 800 and 840 bp.
(B) Western blots showing St1Cas9-tag protein expression in a K562 clone and in cells expressing only the tag (Mock). The FLAG M2 antibody was used to detect Cas9 and the tubulin antibody was used as a loading control. (C) Alignment of previously described sgRNA sequences for St1Cas9 as well as novel designs from the present study (v1, v2, v3). [1]Kleinstiver, Nature 2015b, [2]Esvelt, Nat Methods 2013, [3]Muller, Mol Ther 2016, [4]Ran, Nature 2015. (D) K562 cells were transiently transfected with a St1Cas9 LMD-9 expression vector (0.5µg) in addition to the indicated sgRNA expression plasmids (0.8µg) and Surveyor assays were performed 3 days later to determine the frequency of indels at the 3 specified targets. (E) K562 cells stably expressing St1Cas9 LMD-9 were transfected with indicated sgRNA expression vectors targeting *EMX1* at increasing doses and TIDE assays were performed 3 days later to determine the frequency of indels. (F) K562 cells were transiently transfected with a St1Cas9 LMD-9 expression vector (100 or 200 ng) in addition to the indicated sgRNA expression plasmids (200 and 400 ng) targeting *EMX1*. Indel frequency was determined as in (D). (G) sgRNAs specifying cleavage by St1Cas9 LMD-9 at indicated PAMs with an NN linker were modified to test their functionality with an NNN linker at the specified targets. Cleavage activity was determined as in (D). (H) Cleavage activity of sgRNAs shown in Figure 1C ordered by guide length. Solid shapes (K562), open shapes (Neuro-2A).



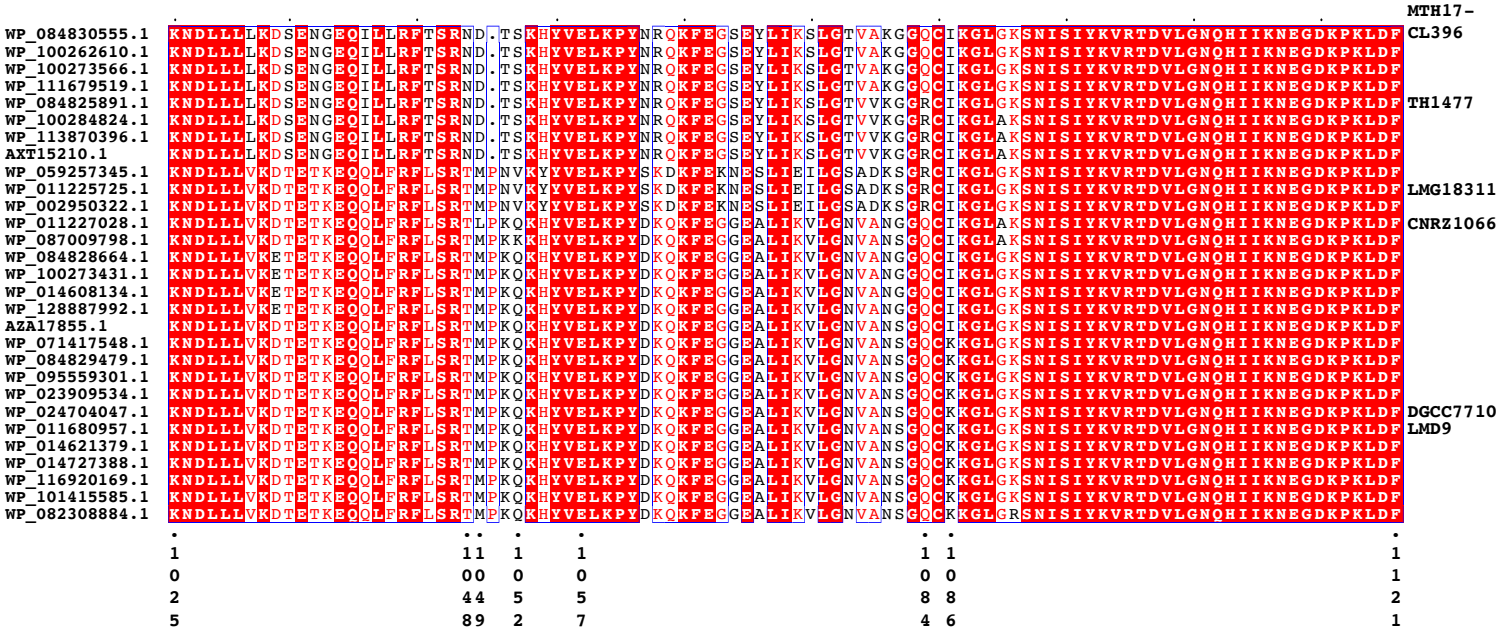
Supplemental Fig. S2. Related to Fig. 1. Engineered CRISPR1-StCas9 sgRNA improves dSt1Cas9-VPR transcriptional activation in human cells. (A) K562 cells were transfected with the tdTomato transcriptional reporter along with dSt1Cas9-VPR and the targeting sgRNAs (v0 or v1) at increasing doses. Fluorescence microscopy images were taken 3 days post-transfection. Scale bars represent 50 µm. (B) Transcriptional activation was quantified by FACS 3 days post-transfection. % of tdTomato-positive cells are indicated in panel (A).


A
1
827
1121
N-term
WED + PI
>98%
84 to >98%
1
963
1121
N-term
PI
>98%
74 to >98%

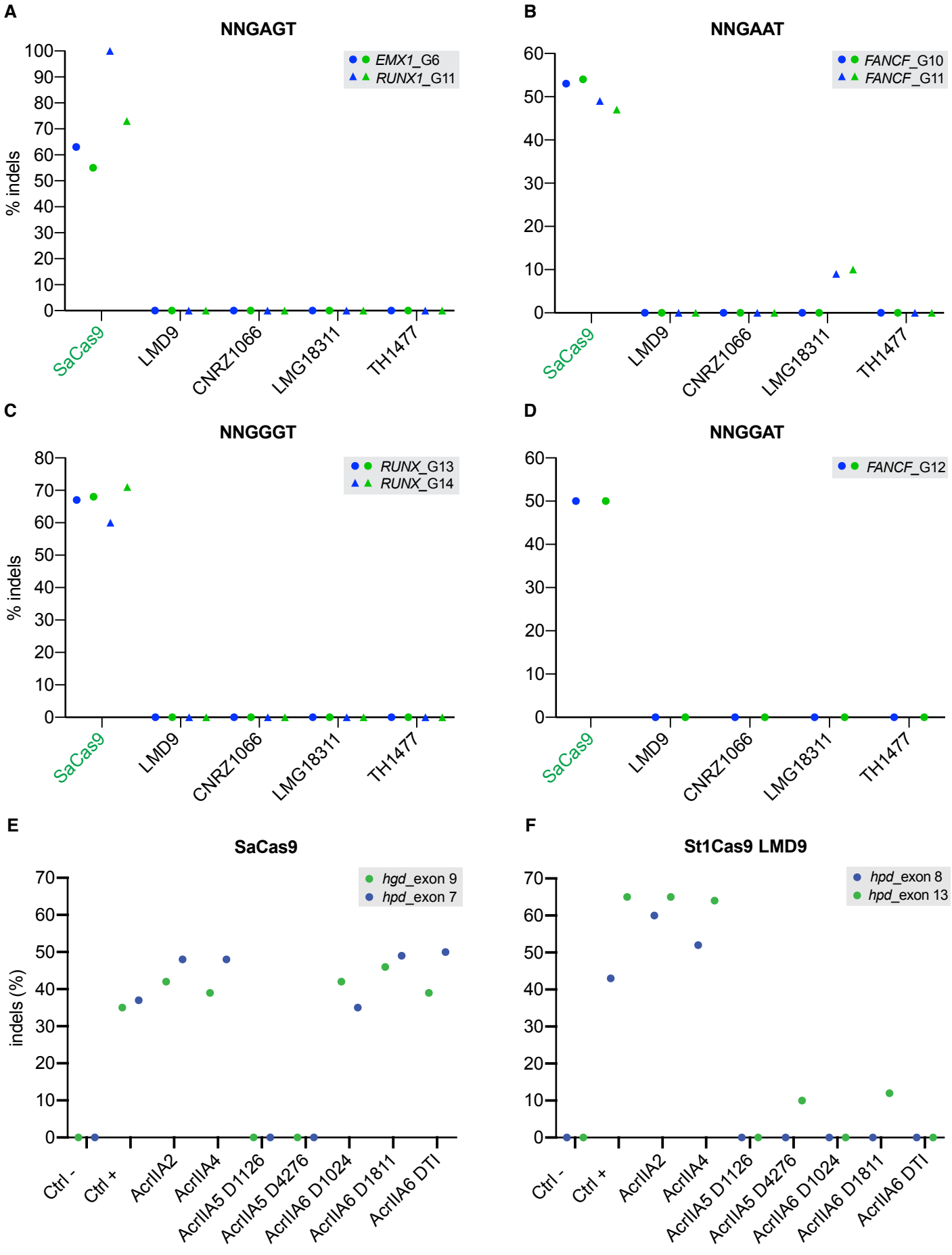

C
1
827
1121
LMD9
TH1477
N-term
WED + PI
1
827
1121
LMD9
MTH17CL396
N-term
WED + PI
Indels (%)
80
60
40
20
0
NNGAAA(A)
NNAAAA(A)
NNGAAA(A)
NNAAAA(A)


B
LMD9
LMG18311
CNRZ1066
TH1477
MTH17
CL396
bits
2.0
1.0
0.0
5
10

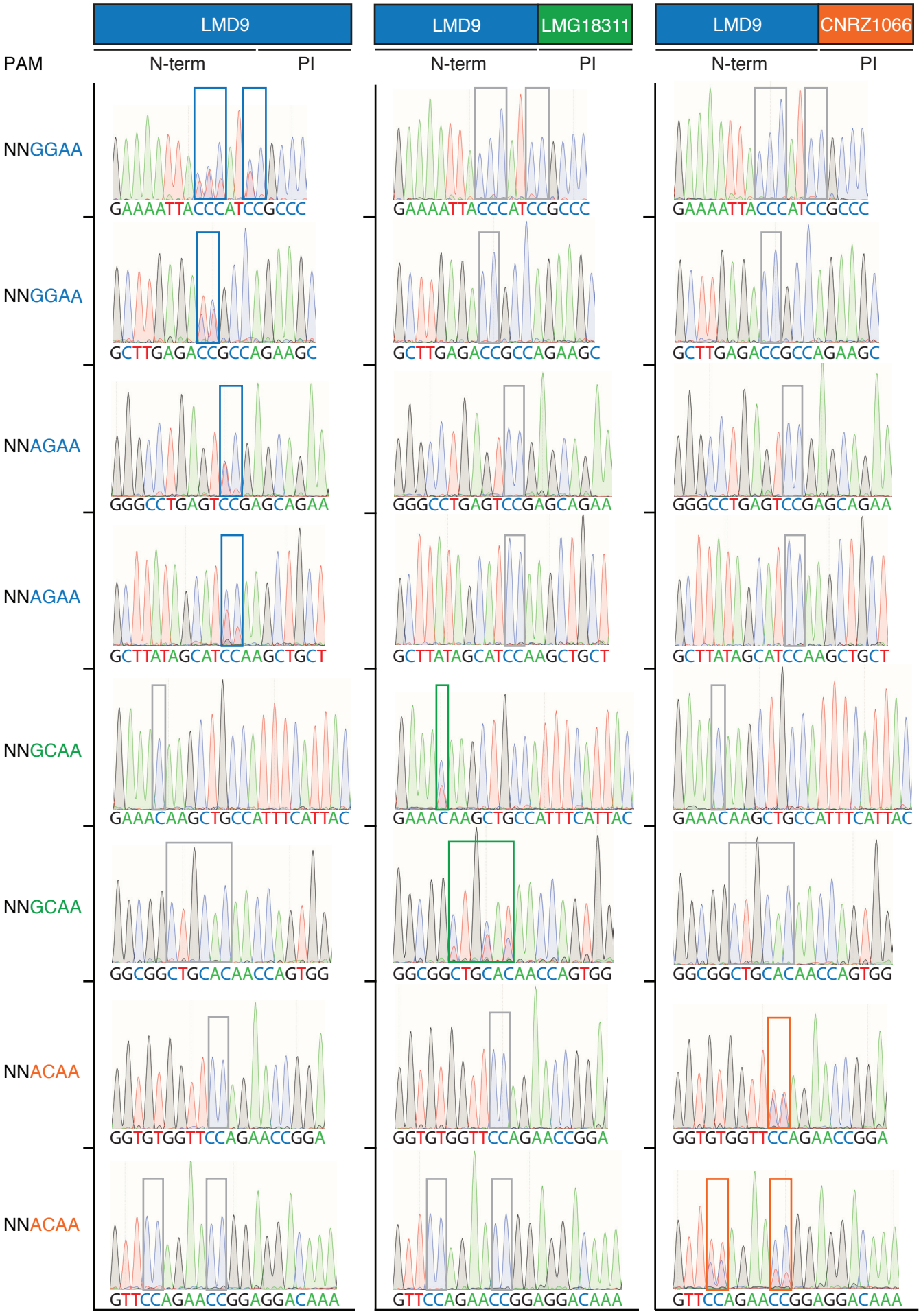
D


MTH17-
CL396
TH1477
LMG18311
CNRZ1066
DGCC7710
LMD9

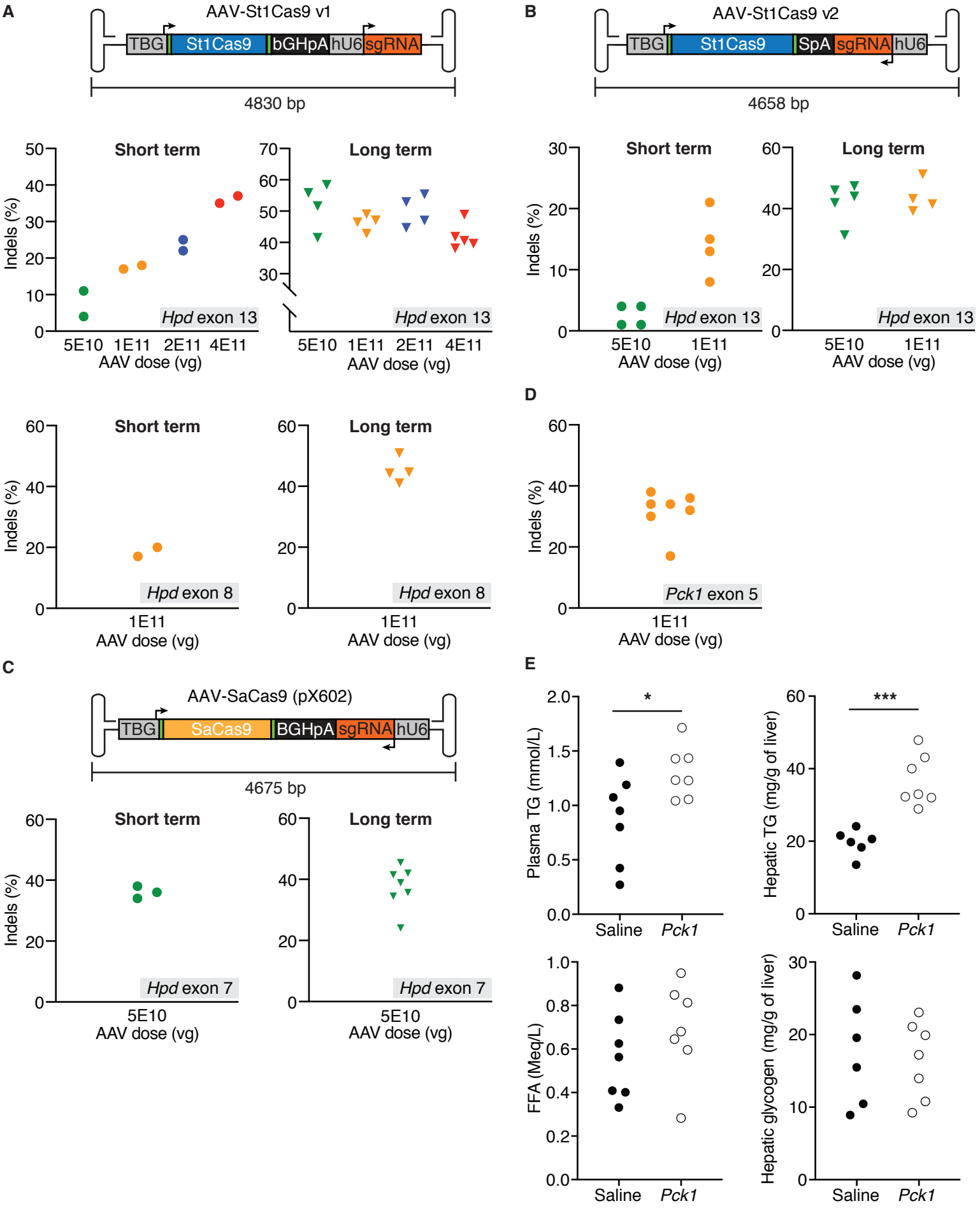
Supplemental Fig. S3. Related to Fig. 2. St1Cas9 strain variants display unique PAM specificities. (A) Percentage of identity within the N-terminal versus the WED and PI domains of St1Cas9s. Amino acid sequence alignment of St1Cas9 respective domains from all 29 variants available in NCBI's Identical Protein Groups (IPG) resource were performed with Clustal Omega. (B) Sequence logos of predicted PAMs for the indicated St1Cas9 "families" using SPAMALOT. (C) Schematic representation of St1Cas9 hybrid proteins containing the N-terminal of LMD9 and the C-terminal domains (WED + PI) of TH1477 and MTH17CL396. To determine the activity of St1Cas9 variants programmed with sgRNAs compatible with indicated PAMs, K562 cells were transiently transfected with single vector constructs driving expression of St1Cas9 and its sgRNA. For each PAM and nuclease combination, four different sgRNAs (targets) were tested. Surveyor assays were performed 3 days later to determine the frequency of indels. An expression vector encoding EGFP (-) was used as a negative control. The experiment was performed twice and yielded equivalent results, only one is shown. (D) Amino acid sequence alignment of St1Cas9 CTD domains from all 29 variants available in NCBI's Identical Protein Groups (IPG) resource. Alignment was performed with Clustal Omega and rendered using ESPript 3.


A
NNGAGT
% indels
EMX1_G6
RUNX1_G11
SaCas9
LMD9
CNRZ1066
LMG18311
TH1477
B
NNGAAT
FANCF_G10
FANCF_G11
C
NNGGGT
RUNX_G13
RUNX_G14
D
NNGGAT
FANCF_G12
E
SaCas9
indels (%)
hgd_exon 9
hpd_exon 7
Ctrl -
Ctrl +
AcrIIA2
AcrIIA4
AcrIIA5 D1126
AcrIIA5 D4276
AcrIIA6 D1024
AcrIIA6 D1811
AcrIIA6 DTI
F
St1Cas9 LMD9
hpd_exon 8
hpd_exon 13

Supplemental Fig. S4. Related to Fig. 2. St1Cas9 strain variants are incompatible with SaCas9 PAMs and differential inhibition of St1Cas9 and SaCas9 by anti-CRISPR proteins. (A) K562 cells were transfected with expression vectors for SaCas9 (0,4 µg) and its sgRNA (0,6 µg) or single vector constructs driving expression of St1Cas9 and its sgRNA. sgRNAs where programmed to target the nucleases to NNGAGT PAMs. Surveyor assays were performed 3 days later to determine the frequency of indels. An expression vector encoding EGFP (-) was used as a negative control. The experiments were performed twice and are represented by two colors (blue for n=1 and green for n=2). (B-D) Same as in (A) but for NNGAAT, NNGGGT, NNGGAT PAMs, respectively. (E) Anti-CRISPR activity of AcrIIAs against SaCas9. Neuro-2a cells were transfected with a dual SaCas9 and sgRNA expression plasmid (0.5µg) targeting *Hpd* and *Hgd*. Where indicated, a second vector (0.5µg) was used to co-express human codon-optimized AcrIIAs. The Surveyor assay was performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. (F) Same as in (E) but using St1Cas9 to target *Hpd* (G2) and *Hpd* (G5).


LMD9
LMD9
LMG18311
LMD9
CNRZ1066
PAM
N-term
PI
N-term
PI
N-term
PI
NNGGAA
GAAAATTACCCATCCGCCC
GAAAATTACCCATCCGCCC
GAAAATTACCCATCCGCCC
NNGGAA
GCTTGAGACCGCCAGAAGC
GCTTGAGACCGCCAGAAGC
GCTTGAGACCGCCAGAAGC
NNAGAA
GGGCCTGAGTCCGAGCAGAA
GGGCCTGAGTCCGAGCAGAA
GGGCCTGAGTCCGAGCAGAA
NNAGAA
GCTTATAGCATCCAAGCTGCT
GCTTATAGCATCCAAGCTGCT
GCTTATAGCATCCAAGCTGCT
NNGCAA
GAAACAAGCTGCCATTTCATTAC
GAAACAAGCTGCCATTTCATTAC
GAAACAAGCTGCCATTTCATTAC
NNGCAA
GGCGGCTGCACAACCAGTGG
GGCGGCTGCACAACCAGTGG
GGCGGCTGCACAACCAGTGG
NNACAA
GGTGTGGTTCCAGAACCGGA
GGTGTGGTTCCAGAACCGGA
GGTGTGGTTCCAGAACCGGA
NNACAA
GTTCCAGAACCGGAGGACAAA
GTTCCAGAACCGGAGGACAAA
GTTCCAGAACCGGAGGACAAA

Supplemental Fig. S5. Related to Fig. 3. St1Cas9BE4max variants display unique and orthogonal PAM specificities. Schematic representation of St1Cas9BE4max hybrid proteins containing the N-terminal of LMD9 and the C-terminal PI domains of LMD9, LMG18311, and CNRZ1066. To determine the activity of St1Cas9BE4max variants programmed with sgRNAs compatible with indicated PAMs, K562 cells were transiently transfected with single vector constructs driving expression of St1Cas9 and its sgRNA. For each PAM and base editor combination, two different sgRNAs (targets) were tested. Genomic DNA was harvested 3 days later, and quantification of base editing was performed on PCR amplified target sites using EditR. The experiment was performed twice and yielded equivalent results, only one is shown.

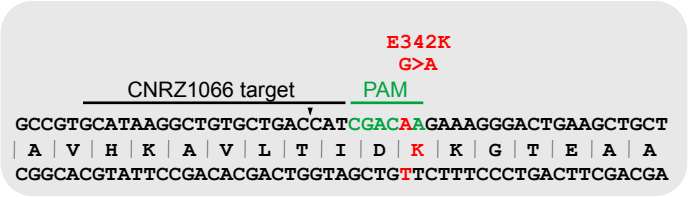

A
AAV-St1Cas9 v1
TBG
St1Cas9
bGHpA
hU6
sgRNA
4830 bp
Short term
Long term
Indels (%)
Hpd exon 13
5E10
1E11
2E11
4E11
AAV dose (vg)
B
AAV-St1Cas9 v2
TBG
St1Cas9
SpA
sgRNA
hU6
4658 bp
Short term
Long term
Indels (%)
Hpd exon 13
5E10
1E11
AAV dose (vg)
C
AAV-SaCas9 (pX602)
TBG
SaCas9
BGHpA
sgRNA
hU6
4675 bp
Short term
Long term
Indels (%)
Hpd exon 7
5E10
AAV dose (vg)
Hpd exon 8
1E11
AAV dose (vg)
D
Indels (%)
Pck1 exon 5
1E11
AAV dose (vg)
E
Plasma TG (mmol/L)
*
Saline
Pck1
Hepatic TG (mg/g of liver)
***
FFA (Meq/L)
Hepatic glycogen (mg/g of liver)

Supplemental Fig. S6. Related to Fig. 4. *In vivo* genome editing using St1Cas9. (A) Schematic representations of the AAV St1Cas9 v1 vector. Human thyroxine binding globulin (TBG) promoter, bovine growth hormone polyadenylation sequence (BGHpA) and hU6 promoter are shown. Arrows indicate the direction of transcriptional unit. Neonatal *Fah*$^{-/-}$ mice were injected into the retro-orbital sinus with either 5E10, 1E11, 2E11 or 4E11 vector genomes (vg) of AAV8-St1Cas9 v1 targeting *Hpd* exon 13 (G5) (Top) or 1E11 vector genomes (vg) of AAV8-St1Cas9 v1 targeting *Hpd* exon 8 (G2) (Bot). Mice were killed 28 days following injection (short term) or kept alive for 8 months (long term) post NTBC removal. Genomic DNA was extracted from whole liver samples and the Surveyor assay (short term) or TIDE assay (long term) were used to determine the frequency of indels. Each dot represents a different mouse. A mouse injected with saline (-) was used as a negative control. (B) Same as in (A) but mice were injected with either 5E10 or 1E11 vector genomes (vg) of AAV8-St1Cas9 v2 targeting *Hpd* exon 13 (G5). (C) Same as in (A) but using 5E10 vector genomes (vg) of AAV8-SaCas9 targeting exon 7 of *Hpd*. Mice were killed after 12 months for the long-term condition. (D) Neonatal C57BL/6N mice were injected into the retro-orbital sinus with 1E11 vector genomes (vg) of AAV8-St1Cas9 v3 targeting exon 5 of *Pck1*, weaned at 21 days and fasted for 24 hours at 6 weeks of age before sacrifice. Genomic DNA was extracted from whole liver samples and the Surveyor assay was used to determine the frequency. Each dot represents a different mouse. A mouse injected with saline (-) was used as a negative control. (E) Plasma triglyceride (Top left) and free fatty acids (Bot left) levels were measured after 24 hours of fasting in mice injected with saline or AAV8-St1Cas9 v3. Hepatic triglycerides (Top right) and Hepatic glycogen (Bot right) were extracted and quantified. Each dot represents a different animal. $* p < 0.05$, $*** p < 0.0005$ according to Student's test.

A

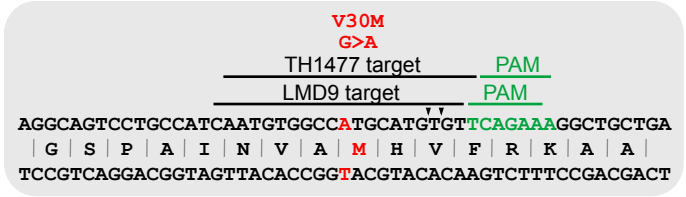
SERPINA1 (Alpha-1 antitrypsin deficiency)

Mutation creates a novel PAM



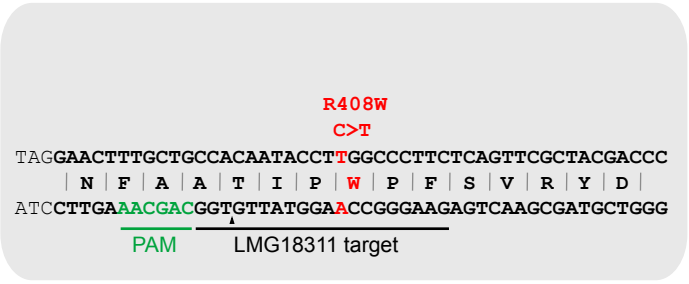
TTR (Transthyretin amyloidosis)

Mutation occurs in seed region; 10bp or 11bp upstream of PAM



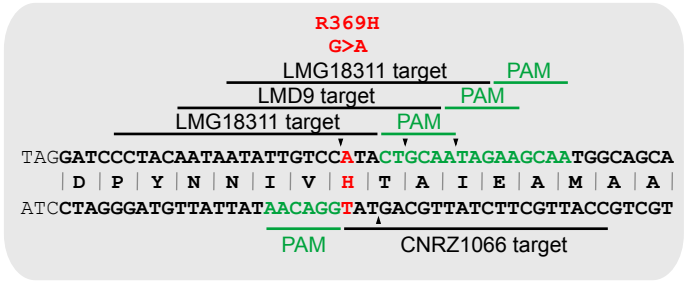
PAH (Phenylketonuria)

Mutation occurs in seed region; 12bp upstream of PAM



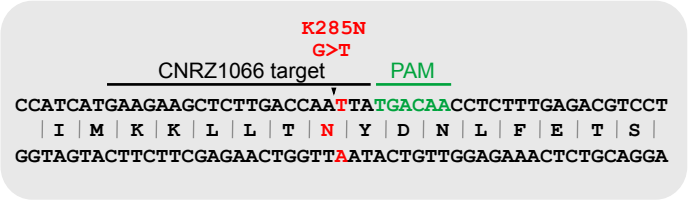
MMUT (Methylmalonic acidemia)

Mutation occurs in seed region; 1bp, 3bp, 8bp or 12bp upstream of PAM



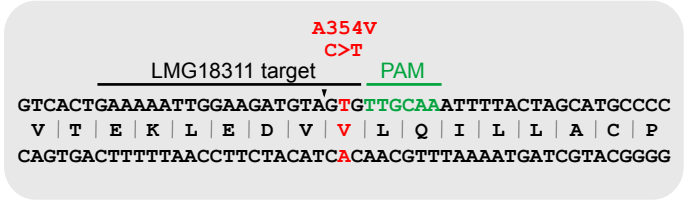
GALT (Classical galactosemia)

Mutation in seed region, 3bp upstream of PAM



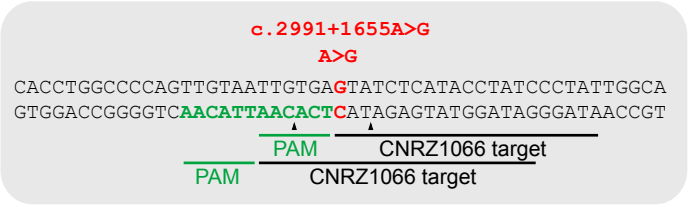
LRPPRC (French Canadian Leigh syndrome)

Mutation occurs in seed region; 2bp upstream of PAM



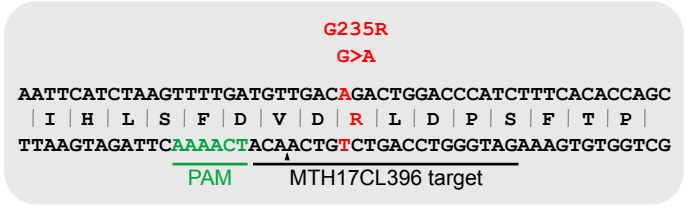
CEP290 (Leber congenital amaurosis)

Mutation occurs in seed region; 1bp or 7bp upstream of PAM



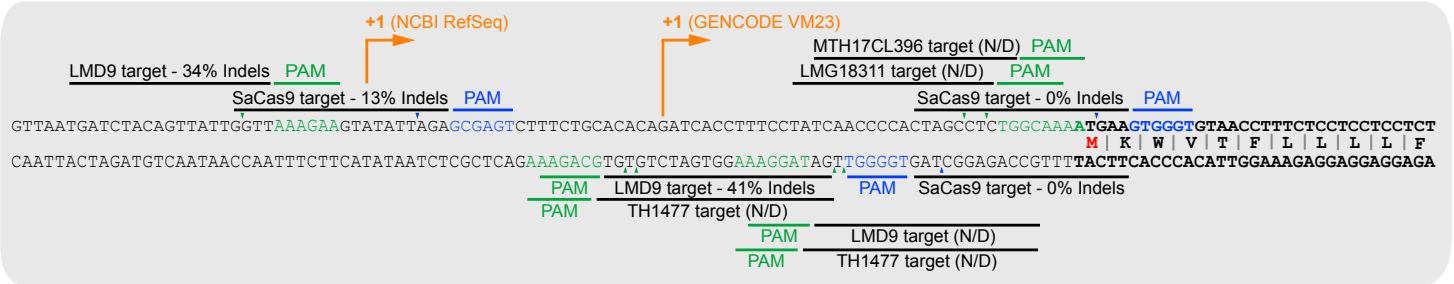
ARG1 (Argininemia)

Mutation occurs in seed region; 8bp upstream of PAM



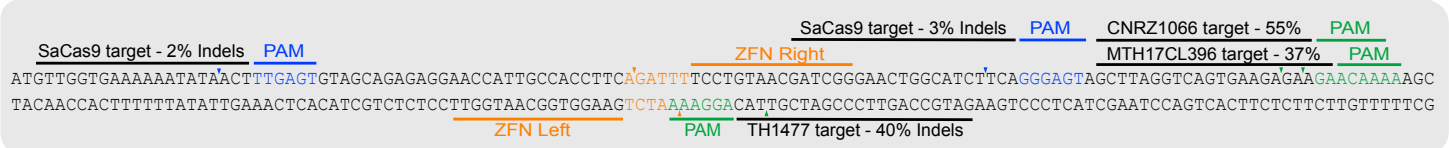
B

Alb **(5'UTR)**



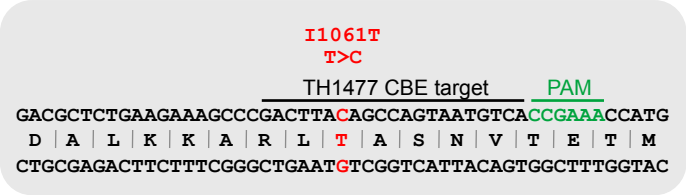
C

Alb **(Intron 1)**

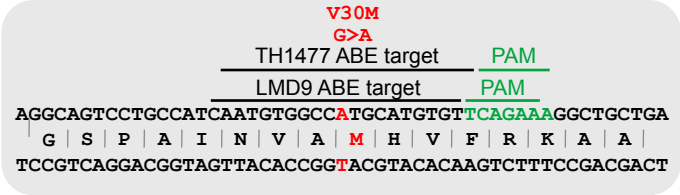


Supplemental Fig. S7. Related to Discussion. (A) Examples of pathogenic alleles found in patients with inherited monogenic diseases potentially targetable by St1Cas9s in a mutation-specific manner. No SaCas9 PAMs occur at proximity of most of these mutations. (B) Activity of St1Cas9 versus SaCas9 at the 5'UTR of the albumin gene. Neuro-2A cells were transfected with single vector constructs (0.5µg) driving expression of St1Cas9 or SaCas9 and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels indicated next to the target sequence. An expression vector encoding EGFP (-) was used as a negative control. The experiment was performed twice and yielded equivalent results, only one is shown. Potential targets for St1Cas9 variants are also shown but have not been tested (N/D). (C) Same as in (B) but at intron 1 of albumin.

NPC1 (Niemann-Pick disease type C)

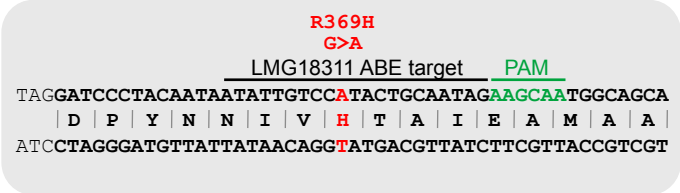


TTR (Transthyretin amyloidosis)



PAH (Phenylketonuria)

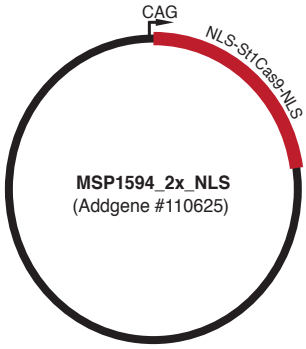
MMUT (Methylmalonic acidemia)



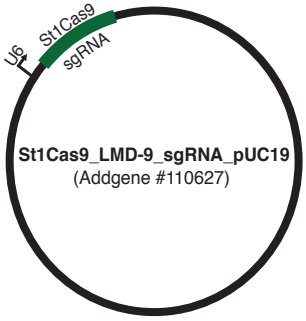
Supplemental Fig. S8. Related to Discussion. (A) Examples of pathogenic alleles found in patients with inherited monogenic diseases potentially targetable by St1Cas9-derived base editors. No SaCas9 PAMs are appropriately positioned to target these mutations using current base editors.

Plasmids for *in vitro* genome editing

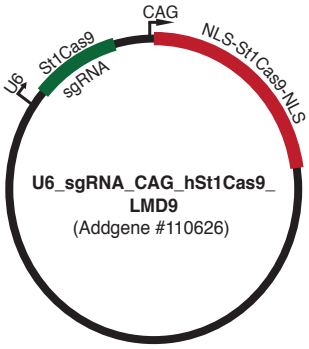
Mammalian expression vector for St1Cas9 LMD-9



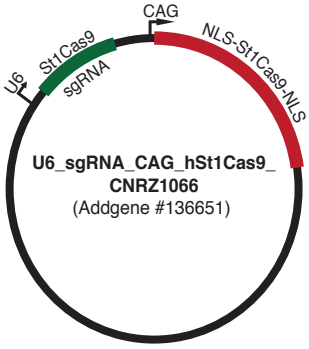
hU6-driven sgRNA expression plasmid for St1Cas9 LMD-9



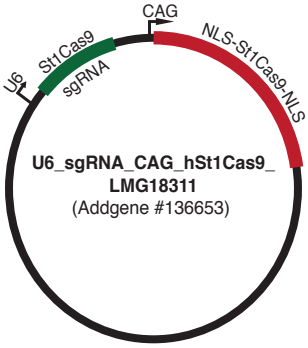
Dual expression vector for St1Cas9 LMD-9 and its sgRNA



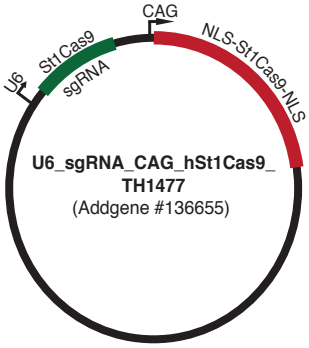
Dual expression vector for St1Cas9 CNRZ1066 and its sgRNA



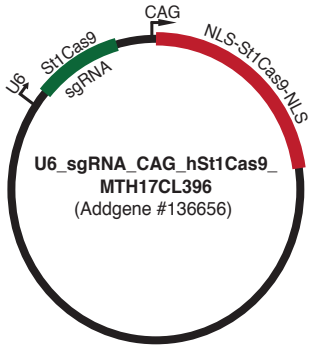
Dual expression vector for St1Cas9 LMG18311 and its sgRNA



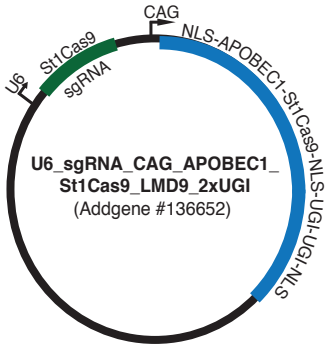
Dual expression vector for St1Cas9 TH1477 and its sgRNA



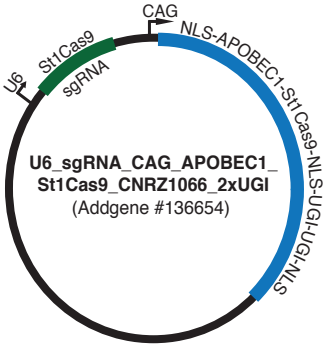
Dual expression vector for St1Cas9 MTH17CL396 and its sgRNA



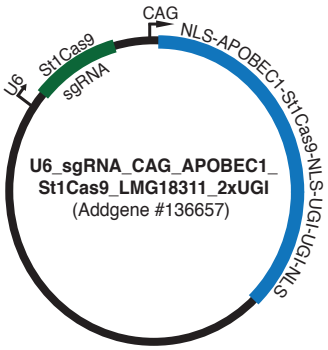
Dual expression vector for St1Cas9 LMD9 cytidine Base editor and its sgRNA



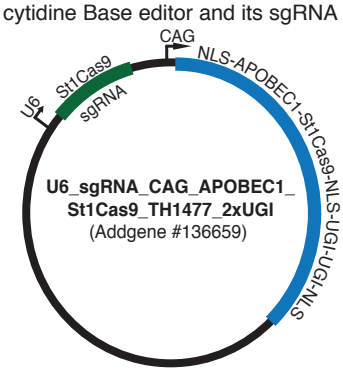
Dual expression vector for St1Cas9 CNRZ1066 cytidine Base editor and its sgRNA



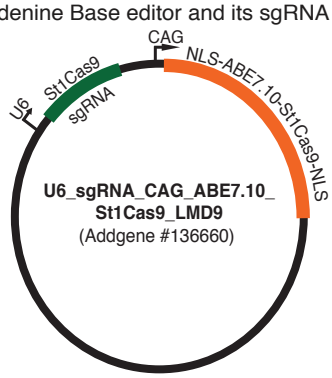
Dual expression vector for St1Cas9 LMG18311 cytidine Base editor and its sgRNA



Dual expression vector for St1Cas9 TH1477 cytidine Base editor and its sgRNA

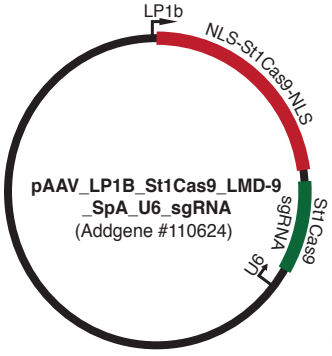


Dual expression vector for St1Cas9 LMD9 Adenine Base editor and its sgRNA



Plasmid for *in vivo* genome editing

Single AAV vector with a liver-specific promoter to express St1Cas9 LMD-9 and its hU6-driven sgRNA



Cloning

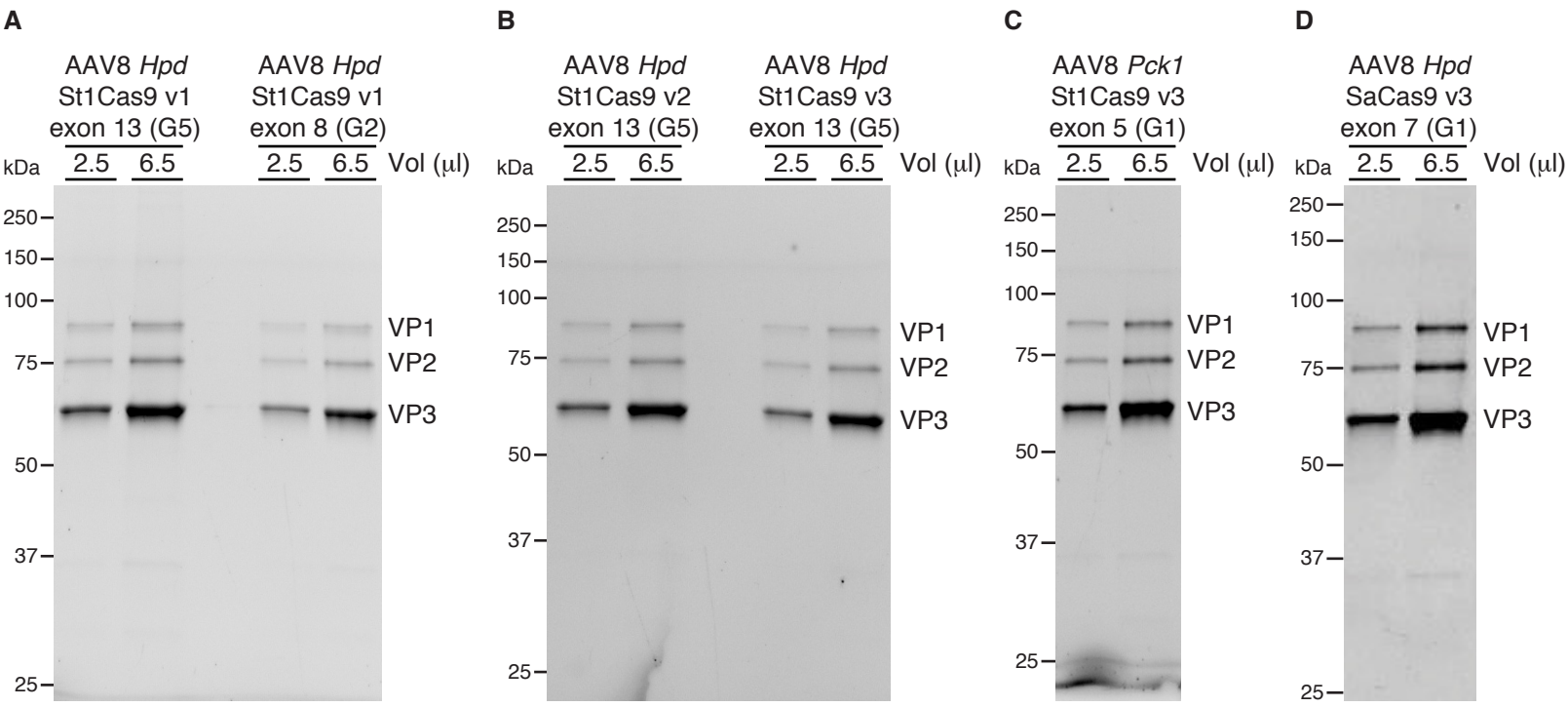
- The guides $N_{(x)}$ can vary in size between 18bp and 22bp (We favor 20bp guides)
- Annealed oligonucleotides must have the indicated overhangs.
- sgRNA expression cassettes are linearized by digestion with BsmBI.

$$\overset{N_{(18-22)}}{\overline{}}$$

5'-**CACCG**NNNNNNNNNNNNNNNNNNN-3'

CNNNNNNNNNNNNNNNNNNN**CAGA**-5'

Supplemental Fig. S9. Related to Methods. St1Cas9 vectors available from Addgene.



Supplemental Fig. S10. Related to Methods. (A-D) Purified rAAV8 vectors used in this study. SDS-PAGE gel analysis of rAAV8-St1Cas9 and rAAV8-SaCas9 vectors. Fixed volumes of purified viruses were run on mini-PROTEAN TGX Stain-Free Gels. Molecular weight (kDa) of protein standards are indicated.