bioRχiv

ARTICLE PREPRINT

Versatile genome editing with the prototypical *Streptococcus thermophilus* CRISPR1-Cas9 system

Sophie Carter[1*], Daniel Agudelo[1*], Alexis Duringer[1], Jeremy Loehr[1], Mathilde Mouchiroud[2], Denis Cyr[3], Paula J Waters[3], Mathieu Laplante[2,4], and Yannick Doyon[1,4,5].

The broad spectrum of activities displayed by CRISPR-Cas systems has led to biotechnological innovations that are poised to transform human therapeutics. Therefore, the comprehensive characterization of distinct Cas proteins is highly desirable. Here we expand the repertoire of nucleases for mammalian genome editing using the archetypal *Streptococcus thermophilus* CRISPR1-Cas9 (St1Cas9) and define variables required for robust and efficient editing *in vitro* and *in vivo*. We show that diverse bacterial strains encode divergent St1Cas9s with altered protospacer adjacent motif (PAM) binding specificity offering a simple solution to expand targeting density. Expression of holo-St1Cas9 from a single adeno-associated viral (rAAV) vector in the neonatal liver efficiently rewires metabolic pathways due to robust DNA cleavage activity that is comparable to *Staphylococcus aureus* Cas9 (SaCas9) *in vivo*. Lastly, we identified potent anti-CRISPR proteins that differentially inhibit the activity of St1Cas9 and SaCas9 in cells. This work expands the targeting range and versatility of CRISPR-associated enzymes and should encourage studies to determine St1Cas9 structure, genome-wide specificity profile and sgRNA design rules.

KEYWORDS Genome editing, CRISPR-Cas9, *Streptococcus thermophilus* Cas9 (St1Cas9), *Staphylococcus aureus* Cas9 (SaCas9), Protospacer adjacent motif (PAM), Anti-CRISPR proteins (Acr), Adeno-associated virus (AAV), *In vivo* genome editing, Metabolic rewiring.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) proteins form a prokaryotic adaptive immune system and some of its components have been harnessed for robust genome editing (1). Type II-based editing tools rely on a large multidomain endonuclease, Cas9, guided to its DNA target by an engineered single-guide RNA (sgRNA) chimera (2) (See (3, 4) for a classification of CRISPR-Cas systems). The Cas9-sgRNA binary complex finds its target through recognition of a short sequence called the protospacer adjacent motif (PAM) and the subsequent base pairing between the guide RNA and DNA leads to a double-strand break (DSB) (1, 5). While *Streptococcus pyogenes* (SpCas9) remains the most widely used Cas9 ortholog for genome engineering, the diversity of naturally occurring RNA-guided nucleases is astonishing (4). Hence, Cas9 enzymes from different microbial species can contribute to the expansion of the CRISPR toolset by increasing targeting density, improving activity and specificity as well as easing delivery (1, 6).

In principle, engineering complementary CRISPR-Cas systems from distinct bacterial species should be relatively straightforward, as they have been minimized to only two components. However, many such enzymes were found inactive in human cells despite being accurately reprogrammed for DNA binding and cleavage *in vitro* (7-9). Nevertheless, the full potential of selected enzymes can be unleashed using machine learning to establish sgRNA design rules (10, 11).

Perhaps the most striking example of the value of alternative Cas9 enzymes is the implementation of the type II-A Cas9 from *Staphylococcus aureus* (SaCas9) for *in vivo* editing using a single recombinant adeno-associated virus (rAAV) vector (8, 12). More recently, *Campylobacter jejuni* and *Neisseria meningitidis* Cas9s from the type II-C (13) CRISPR-Cas systems have been added to this repertoire (14, 15). *In vivo* editing offers the possibility to generate phenotypes in animal models in order to better recapitulate the interactions between cell types and organs. In addition, it can be envisioned as a novel class of human therapeutics that enables precise molecular correction of genetic defects underlying diseases. Therefore, further development of robust and wide-ranging CRISPR-based technologies for *in vivo* editing may help to decipher disease mechanisms and offer novel therapeutic options (16, 17).

Here we revisited the properties of *Streptococcus thermophilus* type II-A CRISPR1-Cas9 system (18) and engineered a potent RNA-guided nuclease for both *in vitro* and *in vivo* applications. The distinctive functional PAM

[1]Centre Hospitalier Universitaire de Québec Research Center - Université Laval, Québec, QC, Canada. [2]Centre de Recherche de l'Institut Universitaire de Cardiologie et de Pneumologie de Québec (CRIUCPQ) - Université Laval, QC, Canada. [3]Service de Génétique médicale, Département de Pédiatrie, Centre Hospitalier Universitaire de Sherbrooke (CHUS), Sherbrooke, QC, Canada. [4]Université Laval Cancer Research Centre, Québec, QC, Canada. *These authors contributed equally to this work. [5]Correspondence to: Yannick.Doyon@crchudequebec.ulaval.ca.

sequences (NNAGAA and NNGGAA) of St1Cas9 LMD-9 increase the targeting flexibility and combinatorial potential of CRISPR-based genome editing tools. In addition, we discovered that natural diversity occurring at the strain level in *S. thermophilus* species can be harnessed to easily reprogram the PAM of St1Cas9s. *In vivo* delivery using an all-in-one rAAV led to efficient inactivation of *Hpd* and *Pck1*, perturbing metabolic pathways in the liver of neonatal mice. Lastly, we identified a family of anti-CRISPR proteins acting as off-switches for both SaCas9 and St1Cas9.

RESULTS

Identification of an sgRNA architecture directing robust DNA cleavage by St1Cas9 in human cells

We recently studied the interplay between Cas9 orthologs and anti-CRISPR proteins isolated from virulent phages infecting *S. thermophilus* (19) and were surprised by the substantial levels of editing achieved with St1Cas9 since early reports indicated that it was mildly active in human cells (8, 20). This observation led us to investigate whether we could further boost its activity. First, we added an N-terminal nuclear localization signal (NLS) to a previously described human codon-optimized expression construct (21) and established a K562 cell line constitutively expressing St1Cas9 (*S. thermophilus* strain LMD-9) from the *AAVS1* safe harbor locus (22) (**Fig. 1a** and **Supplementary Fig. 1**). In this setting, maximal nuclease activity can be achieved (23) and the impact of various sgRNA architectures can be determined without concerns related to variable levels of St1Cas9 expression. Then, we compared several sgRNA designs to the wild-type full length crRNA:tracrRNA duplex connected via a tetraloop (v0) in order to optimize the performance of the system (**Fig. 1b** and **Supplementary Fig. 1**). Cells were transfected with increasing amounts of sgRNA constructs targeting *FANCF*, *EMX1*, and *RUNX1* (21) and the Surveyor nuclease assay was used to determine the frequency of indels characteristic of imprecise DSB repair by NHEJ (23, 24) (**Fig. 1c-d** and **Supplementary Fig. 1**). The spectrum and frequency of targeted mutations was also analyzed using the complementary TIDE (Tracking of Indels by DEcomposition) method (25). The best performing sgRNA architecture (v1) was engineered by truncating the repeat:anti-repeat region (26) and substituting a wobble base pair present in the lower stem for a canonical Watson-Crick base pair (**Fig. 1b**). The improvement in activity was also observed when co-expressing St1Cas9 and its sgRNA transiently, a setting more typical of a genome editing experiment (**Supplementary Fig. 1**). This analysis revealed that high gene disruption rates could be obtained under standard conditions using St1Cas9 and a designed sgRNA in human cells.

Functional PAM sequences for St1Cas9 from strain LMD-9 in human cells

Cas9 orthologs rely on different PAMs as the first step in target recognition. The consensus PAM for St1Cas9 (LMD-9 and DGCC7710 *S. thermophilus* strains that differ by only 2 aa) was originally defined as NNAGAAW (27). However, sequences closely related to the consensus, in particular NNAGAA and NNGGAA can be functional in test tubes and in bacterial cells (6, 21, 28). Based solely on PAM occurrence in the DNA sequence, we assembled single vector constructs expressing both St1Cas9 LMD-9 and its sgRNA to target *FANCF*, *EMX1*, *VEGFA*, and *RUNX1* (21) at NNAGAA and NNGGAA PAMs. Transient transfection of K562 cells revealed that St1Cas9 LMD-9 could be targeted to both PAM sequences with equivalent efficacy as gaged by the Surveyor nuclease assay (**Fig. 2a** and **Supplementary Fig. 2**). The length of the nonconserved PAM linker (NN) has also been shown to be flexible and an extension from 2 to 3 bases can be tolerated in bacterial cells (26, 29). However, we failed to reproduce this observation in human cells suggesting a higher stringency of the system in this context (**Fig. 2b**). Hence, while recognition of an A-rich PAM may facilitate targeting A/T-rich regions of genomes, the flexibility of PAM recognition by St1Cas9 LMD-9 enhances its targeting capabilities.

Engineering an St1Cas9 variant with a distinct PAM preference based on natural diversity found within *S. thermophilus* strains

The targeting range of shorter Cas9 orthologs is constrained by their longer and more complex PAM (8, 14, 15). We speculated that tapping into the natural diversity found within *S. thermophilus* strains might allow us to identify St1Cas9 enzymes with distinct PAM specificities. For example, based on protospacer identification *in silico*, inferred consensus PAM sequences for St1Cas9 from strains CNRZ1066 and LMG13811 are NNACAA(W) and NNGYAA(A) (Y = C or T), respectively (30). In addition, the CRISPR1-Cas system of *S. thermophilus* strain LMG13811 transplanted in *E. coli* or reconstituted from purified components can target DNA using the NNGCAAA PAM (29). At the protein level, the sequence of those three St1Cas9 variants diverges mostly within the C-terminal wedge (WED) and PAM-interacting (PI) domains, implying that they have evolved to recognize slightly distinct PAM sequences (**Supplementary Fig. 3**). Using the structure of SaCas9 (31) as a guide (St1Cas9 shares 37% identity with SaCas9, see **Supplementary Fig. 4**), we tested whether swapping the C-terminus of St1Cas9 LMD-9 with the one from St1Cas9 LMG13811 could reprogram the PAM specificity (**Fig. 3a**). While St1Cas9 LMD-9 could only target NNAGAA and NNGGAA PAMs, the hybrid construct targeted only NNGCAA PAMs with high efficacy (**Fig. 3b-c**, **Supplementary Figs. 2,5**). Minimal cross reactivity was observed in this setting. In the few guides tested, the St1Cas9 hybrid was not active at NNGTAA PAMs (**Supplementary Fig. 5**). Note that the same sgRNA architecture was used with all St1Cas9 variants (see **Fig.**

1b). These data highlight the modularity inherent to Cas9 enzymes. As sequence database mining predicts that even more diversity exists within CRISPR1-StCas9 systems (32), this simple strategy may be used to further expand the targeting range of St1Cas9.

An anti-CRISPR family of proteins inhibits both SaCas9 and St1Cas9

Various families of anti-CRISPR proteins (Acrs) have been recently found in viruses and they block the activities of CRISPR-Cas systems in bacteria and archaea. Of interest here, these proteins could potentially be harnessed to fine-tune the activity of CRISPR-Cas9-based genome editing tools (33). In theory, Acrs could be used as safety off-switches to block Cas9 activity *in vivo* and prevent overt off-target activity. They could also be used during manufacturing of recombinant viral vectors to prevent cleavage of the packaging cell line's genome, which is detrimental to vector yield. The same concept would apply to protect the integrity of self-inactivating viral vectors that bear Cas9 cleavage sites. To begin to explore this avenue, we screened type II-A CRISPR-Cas9 inhibitor proteins (AcrIIAs) against SaCas9 and St1Cas9 (34, 35). Interestingly, the AcrIIA5 family could inhibit both Cas9s in cultured cells while the AcrIIA6 family was specific to St1Cas9 (19) (**Fig. 4**). Note that under similar conditions, we previously showed that AcrIIA5 could also block the activity of SpCas9 (19). AcrIIA4, an SpCas9 inhibitor, did not act on SaCas9 or St1Cas9 (**Fig. 4**). The broad range of activities displayed by these natural variants indicates that they could be engineered as highly potent Cas9 inhibitor for *in vivo* applications.

Robust editing of target genes involved in liver metabolism by St1Cas9 in mouse cells

As the small size of St1Cas9 would be permissive to packaging into rAAV vectors for *in vivo* genome editing (see below), we used CRISPOR (36) to design several sgRNAs against *Hpd*, *Pck1*, and *Pcsk9,* three genes affecting liver function when disrupted. We reasoned that these targets would enable us to test the cleavage efficacy of St1Cas9 *in vivo*. When possible, we selected guides targeting essential protein domains and predicted to have few potential off-targets. Transient transfection of single vector constructs expressing both St1Cas9 and its sgRNA in mouse Neuro-2a cells revealed strong cleavage activity (18% to > 50% indels) at 14 out of 15 target sites highlighting the robustness of the system (**Fig. 5** and **Supplementary Fig. 2**). Taken together, these data suggest that St1Cas9 might enable *in vivo* genome editing if it could be packaged into a single rAAV particle alongside its sgRNA and the regulatory elements needed to drive its expression.

Potent *in vivo* genome editing using an all-in-one rAAV vector in newborn mice

Recombinant AAV vectors are prime *in vivo* gene delivery vectors for non-proliferative tissues. However, a limitation in the therapeutic use of rAAV is the loss of episomal vector genomes from actively dividing cells resulting in transient expression of therapeutic transgenes. Hence, the combination of genome editing technology with rAAV-mediated delivery could lead to permanent genome modification and positive therapeutic outcome in young patients when tissues, such as the liver and retina, are still growing (37, 38). As a side benefit, the elimination of vector genomes leads to transient nuclease expression in proliferating tissues that likely prevents accumulation of mutations at off-target sites (37, 38). *In vivo* editing could complement the use of classical mouse models to study inborn errors of metabolism and potentially alleviates concerns related to developmental compensation.

In vivo inactivation of *Hpd* rescues lethality and metabolic defects in a mouse model of hereditary tyrosinemia type I

To deliver holo-St1Cas9 (St1Cas9 + sgRNA) to the liver, we generated a rAAV plasmid based on the original SaCas9 expression construct (8) and produced hepatotropic rAAV serotype 8 (hereafter referred to as AAV8-St1Cas9) vectors (38, 39) (**Supplementary Fig. 6**). To test the cleavage activity of AAV8-St1Cas9 and its capacity to achieve phenotypic correction *in vivo*, we used a mouse model of hereditary tyrosinemia type I (HT-I) (OMIM 276700) (Orphanet ORPHA:882), an autosomal recessive disease caused by a deficiency of fumarylacetoacetate hydrolase (FAH), the last enzyme of the tyrosine catabolic pathway (**Fig. 6a**). *Fah*$^{-/-}$ mutant mice die as neonates with severe hepatic dysfunction and kidney damage due to the accumulation of toxic metabolites unless treated with nitisone (NTBC), a drug that inhibits 4-hydroxyphenylpyruvate dioxygenase (Hpd) upstream in the pathway (**Fig. 6a**) (40). Likewise, genetic ablation of *Hpd* in mice prevents liver damage and lethality leading to a much milder HT-III phenotype (41, 42) (**Fig. 6a**).

We injected NTBC-treated *Fah*$^{-/-}$ mice at day 2 of life into the retro-orbital sinus with increasing amounts of vector targeting exon 13 of *Hpd* (aka AAV8-St1Cas9 *Hpd* G5) and isolated total liver DNA at day 28 post injection in a subset of treated mice. The titration showed that the degree of target editing was substantial and dependent on the dose of AAV8-St1Cas9 (**Supplementary Fig. 6**). It is worth noting that as the genomic DNA was extracted from pieces of total livers, the effective activity is likely to be underestimated since hepatocytes make up 70-85% of the liver's mass. To assess the extent of metabolic correction, the drug was withdrawn shortly after weaning in the remaining groups of treated mice (**Fig. 6b**). Systemic delivery via a single neonatal injection rescued lethality in all mice while saline-treated animals had to be killed after ~3 weeks as they met the weight loss criterion (**Supplementary Fig. 6**). Likewise,

glycemia was normalized in the treatment groups (**Supplementary Fig. 6**). At low vector doses, we observed delayed but complete recovery, which is likely due to the potent selective growth advantage of targeted hepatocytes that can extensively repopulate the diseased organ (40) (**Supplementary Fig. 6**). Notably, the excretion of succinylacetone, a toxic metabolite and a diagnostic marker for HT-I, was inversely correlated with the dose of rAAV demonstrating metabolic correction (**Supplementary Fig. 6**). These observations were recapitulated when targeting exon 8 of *Hpd* (**Supplementary Fig. 7**). Only three missense mutations in *HPD* are known to occur in patients and we could target the corresponding site for two of them with high efficacy *in vivo* (OMIM 276710) (**Supplementary Figs. 6-7**). Targeting the third mutation was not attempted due to the low specificity score of the guide.

We then evaluated if alternative rAAV-St1Cas9 expression cassettes could further improve cleavage efficacy *in vivo* while minimizing vector size (**Fig. 6c** and **Supplementary Fig. 8**). In our best performing design (AAV-St1Cas9 v3), we engineered a liver-specific promoter (LP1b) by combining elements from the human apolipoprotein E/C-I gene locus control region (*ApoE*-HCR), a modified human α1 antitrypsin promoter (h*AAT*), an SV40 intron and used a synthetic polyadenylation signal element (**Fig. 6c**) (43, 44). These modifications increased cleavage efficacy markedly, especially at low rAAV dose, and led to the creation of a vector of ~4.7 kb in size which is optimal for viral particle packaging (39) (**Fig. 6c-d**). Under the same experimental conditions, the levels of *in vivo* editing achieved with rAAV-St1Cas9 v3 were comparable to the ones obtained using the gold standard rAAV-SaCas9 system (8), as shown for two target genes in the tyrosine degradation pathway (**Fig. 6e-f**). As observed with our initial vector systems (v1, v2), lethality and metabolic defects were rescued upon treatment in *Fah*$^{-/-}$ mice (**Fig. 6g-i**). Therefore, rAAV-mediated delivery of St1Cas9 in neonatal mice can result in efficient DNA cleavage and phenotypic correction by rewiring a metabolic pathway.

Impaired response to fasting upon *in vivo* inactivation of *Pck1*

We then assessed whether *in vivo* genome editing with AAV8-St1Cas9 could be used as a tool for better understanding the control of metabolism, as is the case for conventional genetically modified animal models. To do so, we targeted the gene encoding phosphoenolpyruvate carboxykinase (Pck1 aka PEPCK), which plays a broad role in integrating hepatic energy metabolism and gluconeogenesis (45). Mice with a liver-specific deletion of the gene are viable but display phenotypes associated with maladaptation to fasting including elevated plasma triglycerides and free fatty acids, decreased hepatic glycogen, and increased liver triglycerides content (46). Neonatal 2 days old C57BL/6N (wild-type) pups were injected with AAV8-St1Cas9 *Pck1* G1, and at 6 weeks of age, they were fasted for 24 hours and killed for metabolic profiling and evaluating gene disruption efficacy (**Fig. 7a**). Systemic delivery via a single neonatal injection resulted in substantial hepatic *Pck1* gene disruption (**Fig. 7b**). Plasma triglyceride levels were elevated and hepatic triglyceride content was markedly increased (**Fig. 7c-d**). We found no change in circulating free fatty acid levels and hepatic glycogen stores were not depleted (**Fig. 7e-f**), suggesting that the observed phenotype may be intermediary to the one described in a prenatal hepatic knock-out model (46). We speculate that normal Pck1 in any non-targeted hepatocytes can partially compensate for the loss-of-function resulting from *in vivo* editing. Nevertheless, rAAV-mediated delivery of St1Cas9 in neonatal mice can efficiently disrupt the function of a key metabolic enzyme leading to clear and substantial phenotype *in vivo*. Collectively, these data support the notion that St1Cas9 can be engineered as a powerful tool for *in vivo* genome editing.

DISCUSSION

Here we report that St1Cas9 can be harnessed for robust and efficient genome editing *in vitro* and *in vivo*, thereby expanding the CRISPR-Cas toolbox. We optimized this previously overlooked system and validated its use in mice by demonstrating efficient rewiring, rescue, and creation of metabolic defects using all-in-one rAAV vectors. There is considerable interest in harnessing the diversity of Cas enzymes, but their implementation as genome editing tools is not a straightforward process (7-9). Some enzymes simply fail to work and some choose their substrates promiscuously, necessitating thorough biochemical characterization (47-49). Nevertheless, orthologous Cas9 enzymes with different PAM requirements increase targeting flexibility and allow multiplexing for combinatorial genetic screens, as demonstrated for SpCas9 and SaCas9 (6, 10, 50). In that regard, St1Cas9 and SpCas9 also function orthogonally (51). Moreover, sgRNAs for St1Cas9 and SaCas9 are not functionally interchangeable, which is likely due to their unique PAM specificity (**Supplementary Fig. 9**). This growing set of complementary and orthologous Cas9 systems could be adapted and combined for targeted knockout, transcriptional control, base editing, as well as “epigenome” editing (1).

While shorter Cas9 orthologs require a longer and more complex PAM than the relatively simple NGG of SpCas9, this intrinsic feature may reduce the occurrence of off-target mutagenesis (52). High-throughput and genome-wide characterization of St1Cas9 specificity will be required to validate this observation. Structure-guided rational design and selection-based approaches could also enable St1Cas9 to recognize a broad range of PAM sequences (21). Interestingly, making use of the natural diversity within S. thermophilus strains may achieve similar outcomes. Since the crRNA and tracrRNA are completely conserved (26), we envision that a suite of tools based on naturally occurring St1Cas9 variants could be used in order to expand the targeting range of this, and potentially other, CRISPR-based genome editing systems.

ACKNOWLEDGMENTS

This study was supported by grants from the Canadian Institutes of Health Research (CIHR) to M.L. and Y.D. and by the Banting Research Foundation to Y.D. Salary support was provided by the Fonds de la recherche du Québec-Santé (FRQS) to M.L. and Y.D. D.A. holds a Vanier Canada graduate scholarship. Partial salary support to A.D. was provided by a Desjardins scholarship from the Fondation du CHU de Québec. We thank Marie-Ève Paquet and the skilled vector core facility staff at the Canadian neurophotonics platform (molecular tools) for rAAV8 production. Robert Tanguay provided the mouse model of HT-I, nitisone, expertise and support. Sylvain Moineau ignited our interest in *S. thermophilus* CRISPR systems and graciously shared his insights on CRISPR-Cas biology and Acrs.

AUTHOR CONTRIBUTIONS

Conceptualization, S.C., D.A., A.D., and Y.D.; Methodology, S.C., D.A., A.D., J.L., M.M., D.C., P.J.W., M.L., and Y.D.; Investigation, S.C., D.A., A.D., J.L., M.M., D.C.; Writing – Original Draft, Y.D.; Writing – Review and Editing, S.C., D.A., A.D., P.J.W., M.L., and Y.D.; Supervision, P.J.W., M.L., Y.D.; Funding Acquisition, M.L., Y.D. All authors read and approved the final manuscript.

METHODS

Cell culture and transfection

K562 were obtained from the ATCC (CCL-243) and maintained at 37 °C under 5% CO_2 in RPMI medium supplemented with 10% FBS, penicillin-streptomycin and GlutaMAX. Neuro-2a were obtained from the ATCC and maintained at 37 °C under 5% CO2 in DMEM medium supplemented with 10% FBS, penicillin-streptomycin and GlutaMAX. All cell lines are tested for absence of mycoplasma contamination. Cells (2E5 per transfection) were transfected using the Amaxa 4D-Nucleofector (Lonza) per manufacturer's recommendations. K562 cell lines expressing SaCas9 and St1Cas9 from the *AAVS1* safe harbor locus were generated as described (22, 23). Briefly, simultaneous selection and cloning was performed for 10 days in methylcellulose-based semi-solid RPMI medium supplemented with 0.5 mg/ml puromycin starting 3 days post-transfection. Clones were picked and expanded in 96 wells for 3 days and transferred to 12-well plates for another 3 days before cells were harvested for western blot analysis.

Genome editing vectors

Vectors for *in vitro* and *in vivo* genome editing with the CRISPR1-Cas9 (St1Cas9) system of *S. thermophilus* LMD-9 generated in this study are available from Addgene (**Supplementary Fig. 10**). The CRISPOR (36) web tool was used to design guide (spacer) sequences against mouse and human targets (**Supplementary Tables 1-5)**. A fraction of the guides targeting *EMX1*, *FANCF*, and *VEGFA* have been described previously (21). DNA sequence for the spacers were modified at position 1 to encode a “G” due to the transcription initiation requirement of the human U6 promoter when required. Alternatively, the spacer length was increased to capture a naturally occurring “G”. The mammalian expression vector for St1Cas9 (LMD-9) fused to SV40 NLS sequences at the N- and C-terminus (MSP1594_2x_NLS; Addgene plasmid #110625) was constructed from MSP1594 (21) (Addgene plasmid #65775, a gift from Keith Joung). The U6-driven sgRNA expression plasmids for St1Cas9 (LMD-9) (v1) (St1Cas9_LMD-9_sgRNA_pUC19; Addgene plasmid #110627) and SaCas9 (8) (**Supplementary Table 6**) were synthesized as a gBlock gene fragments (Integrated DNA Technologies) and cloned into pUC19. BPK2301 (21) (v0) (Addgene plasmid # 65778, a gift from Keith Joung) was used to compare St1Cas9 sgRNA architectures. The single vector mammalian expression system containing a CAG promoter-driven St1Cas9 LMD-9 and its U6-driven sgRNA (U6_sgRNA_CAG_hSt1Cas9_LMD9; Addgene plasmid #110626) was built from the above-described plasmids. The single vector rAAV-St1Cas9 LMD-9 systems containing liver-specific promoters (**Supplementary Table 7**) were assembled from the above-described components into a derivative of pX602 (8) (Addgene plasmid #61593, a gift from Feng Zhang) containing a deletion within the backbone to eliminate BsmBI restriction sites. The LP1b promoter was engineered by combining elements from previously described AAV expression cassettes (43, 44). We deposited the most active version of this vector (v3) (pAAV_LP1B_St1Cas9_LMD-9_SpA_U6_sgRNA; Addgene plasmid # 110624). To establish clonal K562 cell lines constitutively expressing C-terminally tagged SaCas9 and St1Cas9 under the control of an h*PGK1* promoter, the Cas9 ORFs from pX602 and MSP1594_2x_NLS were subcloned into AAVS1_Puro_PGK1_3xFLAG_Twin_Strep(22) (Addgene plasmid # 68375). Untagged AcrIIA proteins were expressed transiently from a modified pVAX1 vector (Thermo Fisher Scientific) containing a beta-globin intron. The AcrIIA ORFs were codon-optimized for expression in human cells and synthesized as gBlock gene fragments (19) (Integrated DNA Technologies) (**Supplementary Table 8**).

Surveyor nuclease and TIDE assays

Genomic DNA from 2.5E5 cells was extracted with 250 ml of QuickExtract DNA extraction solution (Epicentre) per manufacturer's recommendations. The various loci were amplified by PCR using the primers described in **Supplementary Table 9**. Assays were performed with the Surveyor mutation detection kit (Transgenomics) as described (23, 24). Samples were separated on 10% PAGE gels in TBE buffer. Gels were imaged using a ChemiDoc MP (Bio-Rad) system and quantifications were performed using the Image lab software (Bio-Rad). TIDE analysis was

performed using a significance cut-off value for decomposition of $p<0.001$ (25).

Recombinant adeno-associated virus production

Production of recombinant adeno-associated viral vectors was performed by the triple plasmid transfection method essentially as described (53). Briefly, HEK293T17 cells were transfected using polyethylenimine (PEI, Polysciences) with helper plasmid pxx-680 (A gift from R.J. Samulski), the rep/cap hybrid plasmid pAAV2/8 (A gift from James Wilson) and the rAAV vector plasmid. Twenty-four hours post-transfection, media was replaced with growth media without FBS, and cells were harvested 24 hours later. rAAV particles were extracted from cell extracts by freeze/thaw cycles and purified on a discontinuous iodixanol gradient. Virus were resuspended in PBS 320 mM NaCl + 5% D-sorbitol + 0.001% pluronic acid (F-68), aliquoted and stored at $^{-}80^{o}$C. rAAV were titrated by qPCR (Roche) using SYBR green and ITR primers as described (54). The yields for all vectors varied between 1E13 and 2E13 vg/ml. The purity of the viral preparations was determined by SDS-PAGE analysis on a 10% stain free gel (Biorad) in Tris-Glycine-SDS buffer (**Supplementary Figure 11**). ITR integrity was assessed following a BssHII digestion of the AAV plasmid. The vector core facility at the Canadian neurophotonics platform (molecular tools) produced the rAAV8s.

Animal experiments (*Fah*$^{-/-}$ mouse model)

Fah$^{-/-}$ mice (55) on a C57BL/6 genetic background were group-housed and fed a standard chow diet (Harlan #2018SX) with free access to food and water. *Fah*$^{-/-}$ mice drinking water was supplemented with 7.5 mg (2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione) (NTBC)/L and pH was adjusted to 7.0. Mice were exposed to a 12:12-h dark-light cycle and kept under an ambient temperature of 23 ± 1 °C. Animals were cared for and handled according to the *Canadian Guide for the Care and Use of Laboratory Animals.* The Université Laval Animal Care and Use Committee approved the procedures.

Two days old neonatal mice were injected intravenously in the retro-orbital sinus (56) with different doses of rAAV8 or saline in a total volume of 20 µL. Mice were weaned at 21 days of age and NTBC was removed 7 days later. Body weight and glycemia were monitored daily following NTBC removal. Mice were not fasted for measurement of glycemia, data collection occurred between 9-10 am. Animals were killed by cardiac puncture under anesthesia at predetermined time points or when weight loss reached 20% of body weight. Livers were snap frozen for downstream applications.

Urine collection and succinylacetone quantification

Urine from groups of 3-4 mice was collected overnight in metabolic cages (Tecniplast) 15 days after NTBC removal. Urine was centrifuged at 2000 rpm for 5 minutes, aliquoted and frozen at $^{-}80^{o}$C. Succinylacetone was quantified in urine samples by a sensitive method using gas chromatography–mass spectrometry (GC-MS) as previously described (57). The biochemical genetics laboratory at the centre hospitalier universitaire de Sherbrooke performed the analyses.

Animal experiments (C57BL/6)

Two days old neonatal *C57BL/6N* mice were injected intravenously in the retro-orbital sinus with 1E11 vector genome of rAAV8 targeting *Pck1* or saline in a total volume of 20 µL. At 6 weeks of age, mice were fasted for 24 hours and sacrificed by cardiac puncture under anaesthesia using syringes conditioned with EDTA. Liver sections were snap frozen and fixed in 4% PFA for downstream applications. Plasma was stored at -80°C for further biochemical analyses.

Plasma triglycerides and NEFAs quantification

Plasma metabolites were assayed in duplicate using the following commercial kits; triglycerides (Thermo Fisher Scientific, TR22421), NEFAs (Wako, 999-34691, 995-34791, 991-34891, 993- 35191, 276-76491).

Hepatic triglyceride content extraction and measurement

Hepatic triglycerides extraction was previously described in detail (58). Briefly, liver sections (25-50 mg) were homogenized in a 2:1 chloroform:methanol mixture, combined with methanol, and centrifuged for 15 min at 3,000 rpm. 825 ml of supernatant was transferred to a new glass tube, chloroform and 0.73% NaCl was added, and the resulting mixture was centrifuged at 5,000 rpm for 3 min. The upper phase was discarded and the lower phase was washed 3 times with a 3/48/47 mixture of chloroform:methanol:NaCl (0.58%). The lower phase was then evaporated and resuspended in 1 ml of fresh isopropanol. Triglyceride levels were determined with a standard assay kit (Thermo Fisher Scientific, TR22421) according to the manufacturer's instructions.

Hepatic glycogen content extraction and measurement

Glycogen was measured in liver samples as described (59). Briefly, glycogen was extracted in 30% KOH saturated with Na_2SO_4, precipitated in 95% ethanol, and re-suspended in distilled H_2O. Absorbance at 490 nm was measured in triplicates, after addition of phenol and H_2SO_4.

REFERENCES

1. Komor AC, Badran AH, Liu DR. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell. 2017;169(3):559.

2. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816-21.

3. Koonin EV, Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. Curr Opin Microbiol. 2017;37:67-78.

4. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, et al. Diversity and evolution of class 2 CRISPR-Cas systems. Nat Rev Microbiol. 2017;15(3):169-82.

5. Hille F, Richter H, Wong SP, Bratovic M, Ressel S, Charpentier E. The Biology of CRISPR-Cas: Backward and Forward. Cell. 2018;172(6):1239-59.

6. Esvelt KM, Mali P, Braff JL, Moosburner M, Yaung SJ, Church GM. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods. 2013;10(11):1116-21.

7. Chen F, Ding X, Feng Y, Seebeck T, Jiang Y, Davis GD. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. Nat Commun. 2017;8:14958.

8. Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015;520(7546):186-91.

9. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015;163(3):759-71.

10. Najm FJ, Strand C, Donovan KF, Hegde M, Sanson KR, Vaimberg EW, et al. Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol. 2018;36(2):179-89.

11. Kim HK, Min S, Song M, Jung S, Choi JW, Kim Y, et al. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. Nat Biotechnol. 2018;36(3):239-41.

12. Friedland AE, Baral R, Singhal P, Loveluck K, Shen S, Sanchez M, et al. Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. Genome Biol. 2015;16:257.

13. Mir A, Edraki A, Lee J, Sontheimer EJ. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. ACS Chem Biol. 2018;13(2):357-65.

14. Ibraheim R, C-Q. S, A. M, N. A, W X, J. SE. All-in-One Adeno-associated Virus Delivery and Genome Editing by Neisseria meningitidis Cas9 in vivo. bioRxiv. 2018.

15. Kim E, Koo T, Park SW, Kim D, Kim K, Cho HY, et al. In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. Nat Commun. 2017;8:14500.

16. Schneller JL, Lee CM, Bao G, Venditti CP. Genome editing for inborn errors of metabolism: advancing towards the clinic. BMC Med. 2017;15(1):43.

17. Lau CH, Suh Y. In vivo genome editing in animals using AAV-CRISPR system: applications to translational research of human disease. F1000Res. 2017;6:2153.

18. Barrangou R, Horvath P. A decade of discovery: CRISPR functions and applications. Nat Microbiol. 2017;2:17092.

19. Hynes AP, Rousseau GM, Agudelo D, Goulet A, Amigues B, Loehr J, et al. Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. Nat Commun. 2018;9(1):2919.

20. Chari R, Mali P, Moosburner M, Church GM. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. Nat Methods. 2015;12(9):823-6.

21. Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015;523(7561):481-5.

22. Dalvai M, Loehr J, Jacquet K, Huard CC, Roques C, Herst P, et al. A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. Cell Rep. 2015;13(3):621-33.

23. Agudelo D, Duringer A, Bozoyan L, Huard CC, Carter S, Loehr J, et al. Marker-free coselection for CRISPR-driven genome editing in human cells. Nat Methods. 2017;14(6):615-20.

24. Guschin DY, Waite AJ, Katibah GE, Miller JC, Holmes MC, Rebar EJ. A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol. 2010;649:247-56.

25. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. 2014;42(22):e168.

26. Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, et al. Guide RNA functional modules direct Cas9 activity and orthogonality. Mol Cell. 2014;56(2):333-9.

27. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol. 2008;190(4):1401-12.

28. Leenay RT, Maksimchuk KR, Slotkowski RA, Agrawal RN, Gomaa AA, Briner AE, et al. Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol Cell. 2016;62(1):137-47.

29. Chen H, Choi J, Bailey S. Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. J Biol Chem. 2014;289(19):13284-94.

30. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology. 2005;151(Pt 8):2551-61.

31. Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, et al. Crystal Structure of Staphylococcus aureus Cas9. Cell. 2015;162(5):1113-26.

32. Chatterjee P, Jakimo N, Jacobson JM. Divergent PAM Specificity of a Highly-Similar SpCas9 Ortholog. bioRxiv. 2018.

33. Pawluk A, Davidson AR, Maxwell KL. Anti-CRISPR: discovery, mechanism and function. Nat Rev Microbiol. 2018;16(1):12-7.

34. Hynes AP, Rousseau GM, Lemay ML, Horvath P, Romero DA, Fremaux C, et al. An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. Nat Microbiol. 2017;2(10):1374-80.

35. Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ, Krogan NJ, et al. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell. 2017;168(1-2):150-8 e10.

36. Haeussler M, Schonig K, Eckert H, Eschstruth A, Mianne J, Renaud JB, et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 2016;17(1):148.

37. Yang Y, Wang L, Bell P, McMenamin D, He Z, White J, et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. Nat Biotechnol. 2016;34(3):334-8.

38. Li H, Haurigot V, Doyon Y, Li T, Wong SY, Bhagwat AS, et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature. 2011;475(7355):217-21.

39. Colella P, Ronzitti G, Mingozzi F. Emerging Issues in AAV-Mediated In Vivo Gene Therapy. Mol Ther Methods Clin Dev. 2018;8:87-104.

40. Grompe M. Fah Knockout Animals as Models for Therapeutic Liver Repopulation. Adv Exp Med Biol. 2017;959:215-30.

41. Endo F, Kubo S, Awata H, Kiwaki K, Katoh H, Kanegae Y, et al. Complete rescue of lethal albino c14CoS mice by null mutation of 4-hydroxyphenylpyruvate dioxygenase and induction of apoptosis of hepatocytes in these mice by in vivo retrieval of the tyrosine catabolic pathway. J Biol Chem. 1997;272(39):24426-32.

42. Pankowicz FP, Barzi M, Legras X, Hubert L, Mi T, Tomolonis JA, et al. Reprogramming metabolic pathways in vivo with CRISPR/Cas9 genome editing to treat hereditary tyrosinaemia. Nat Commun. 2016;7:12642.

43. Nathwani AC, Gray JT, Ng CY, Zhou J, Spence Y, Waddington SN, et al. Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver. Blood. 2006;107(7):2653-61.

44. McIntosh J, Lenting PJ, Rosales C, Lee D, Rabbanian S, Raj D, et al. Therapeutic levels of FVIII following a single peripheral vein administration of rAAV vector encoding a novel human factor VIII variant. Blood. 2013;121(17):3335-44.

45. Yang J, Kalhan SC, Hanson RW. What is the metabolic role of phosphoenolpyruvate carboxykinase? J Biol Chem. 2009;284(40):27025-9.

46. She P, Shiota M, Shelton KD, Chalkley R, Postic C, Magnuson MA. Phosphoenolpyruvate carboxykinase is necessary for the integration of hepatic energy metabolism. Mol Cell Biol. 2000;20(17):6508-17.

47. Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science. 2018.

48. Dugar G, Leenay RT, Eisenbart SK, Bischler T, Aul BU, Beisel CL, et al. CRISPR RNA-Dependent Binding and Cleavage of Endogenous RNAs by the Campylobacter jejuni Cas9. Mol Cell. 2018;69(5):893-905 e7.

49. Strutt SC, Torrez RM, Kaya E, Negrete OA, Doudna JA. RNA-dependent RNA targeting by CRISPR-Cas9. Elife. 2018;7.

50. Boettcher M, Tian R, Blau JA, Markegard E, Wagner RT, Wu D, et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. Nat Biotechnol. 2018;36(2):170-8.

51. Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lecrivain AL, Bzdrenga J, et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. 2014;42(4):2577-90.

52. Muller M, Lee CM, Gasiunas G, Davis TH, Cradick TJ, Siksnys V, et al. Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. Mol Ther. 2016;24(3):636-44.

53. Gray SJ, Choi VW, Asokan A, Haberman RA, McCown TJ, Samulski RJ. Production of recombinant adeno-associated viral vectors and use in in vitro and in vivo administration. Curr Protoc Neurosci. 2011;Chapter 4:Unit 4 17.

54. Aurnhammer C, Haase M, Muether N, Hausl M, Rauschhuber C, Huber I, et al. Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences. Hum Gene Ther Methods. 2012;23(1):18-28.

55. Grompe M, al-Dhalimy M, Finegold M, Ou CN, Burlingame T, Kennaway NG, et al. Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. Genes Dev. 1993;7(12A):2298-307.

56. Yardeni T, Eckhaus M, Morris HD, Huizing M, Hoogstraten-Miller S. Retro-orbital injections in mice. Lab Anim (NY). 2011;40(5):155-60.

57. Cyr D, Giguere R, Villain G, Lemieux B, Drouin R. A GC/MS validated method for the nanomolar range determination of succinylacetone in amniotic fluid and plasma: an analytical tool for tyrosinemia type I. J Chromatogr B Analyt Technol Biomed Life Sci. 2006;832(1):24-9.

58. Caron A, Mouchiroud M, Gautier N, Labbe SM, Villot R, Turcotte L, et al. Loss of hepatic DEPTOR alters the metabolic transition to fasting. Mol Metab. 2017;6(5):447-58.

59. Lo S, Russell JC, Taylor AW. Determination of glycogen in small tissue samples. J Appl Physiol. 1970;28(2):234-6.

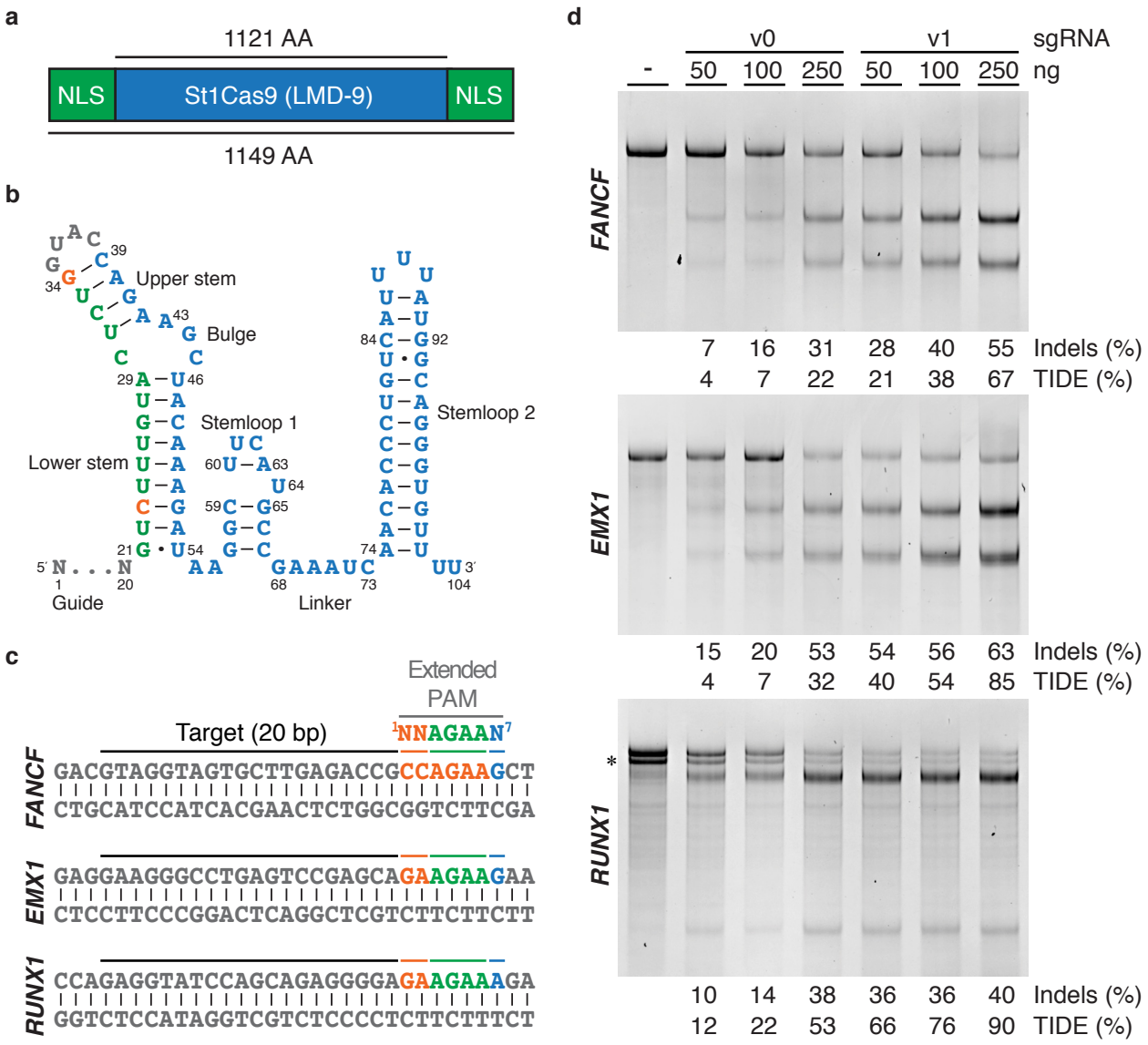


Figure 1 | Engineered CRISPR1-StCas9 system drives robust gene editing in human cells. (**a**) Schematic representation of St1Cas9 from the LMD-9 strain flanked by nuclear localization signals (NLS). (**b**) Nucleotide sequence, predicted secondary structure, and functional modules of St1Cas9 sgRNA (v1); crRNA (green), loop (grey), tracrRNA (blue), mutated nucleotides (orange). (**c**) St1Cas9 target sites and PAM sequences in *FANCF*, *EMX1*, and *RUNX1*. (**d**) K562 cells stably expressing St1Cas9 were transfected with indicated sgRNA expression vectors at increasing doses and the Surveyor and TIDE assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. *A stretch of 15 consecutive "A" within the *RUNX1* amplicon in proximity to the cleavage site prevents accurate quantification of indels by the Surveyor assay. An expression vector encoding EGFP (-) was used as a negative control.

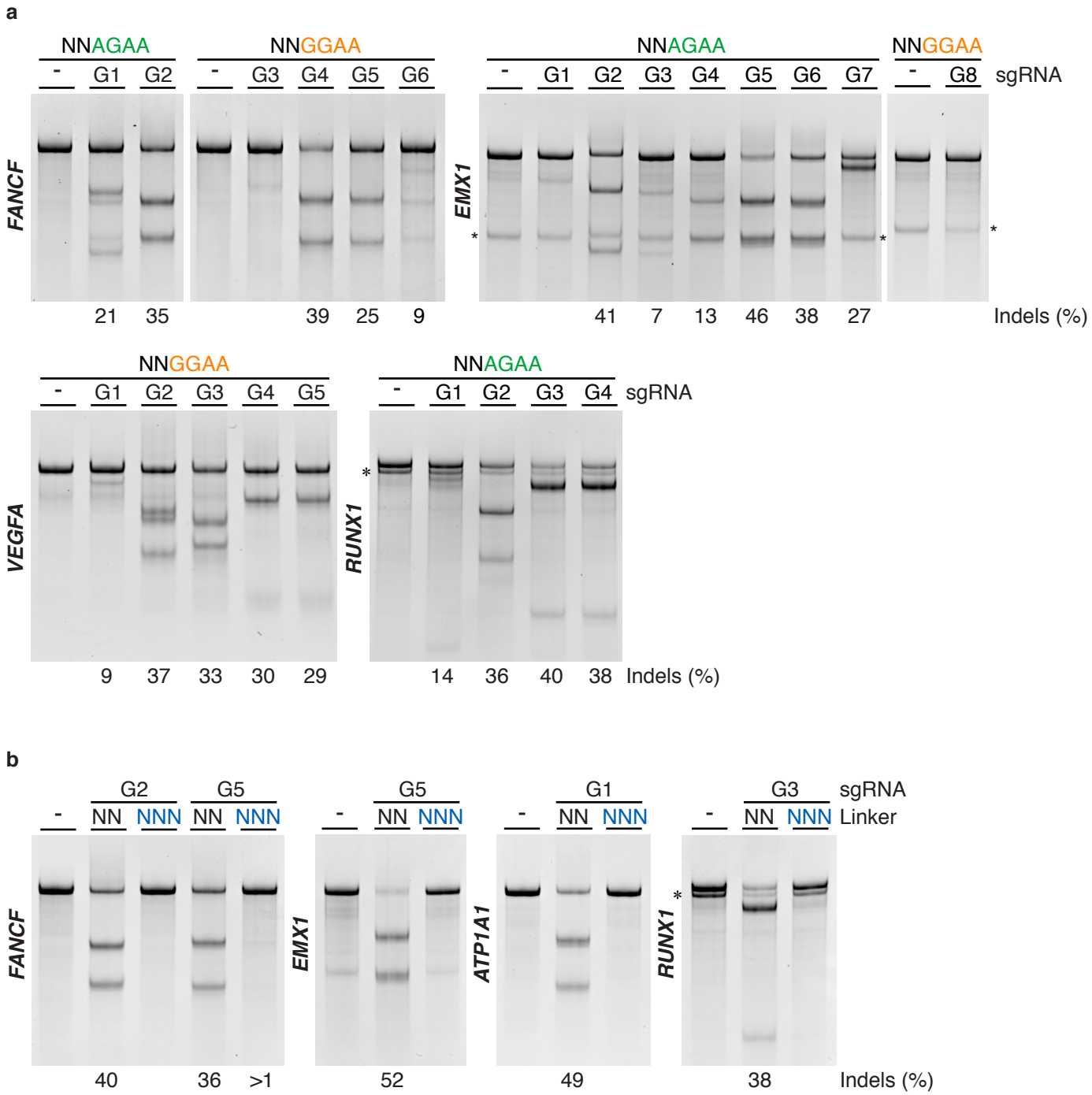


Figure 2 | St1Cas9 LMD-9 is functional at both NNAGAA and NNGGAA PAM sequences in human cells. (**a**) Surveyor assays to determine St1Cas9 activity programmed with various sgRNAs targeting *FANCF*, *EMX1*, *VEGFA*, and *RUNX1*. K562 cells were transiently transfected with a single vector (1µg) driving the expression of St1Cas9 and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control. *Indicates a non-specific PCR amplification product that generates a signal in the *EMX1* Surveyor assay. This signal was subtracted from all quantifications. *A stretch of 15 consecutive "A" within the *RUNX1* amplicon in proximity to the cleavage site prevents accurate quantification of indels. (**b**) sgRNAs specifying cleavage by St1Cas9 at PAMs with an NN linker were modified to test their functionality with an NNN linker. Cleavage activity was determined as in (**a**).

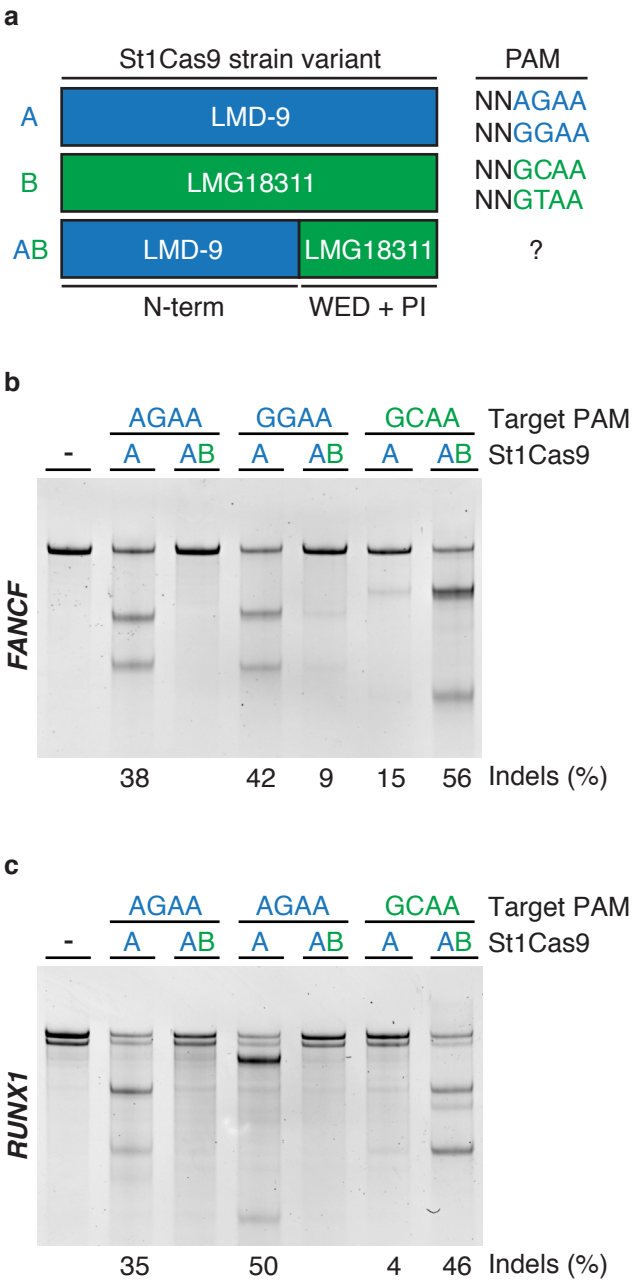


Figure 3 | Rewiring St1Cas9 LMD-9 to target a distinct PAM sequence using a naturally occurring variant. (**a**) Schematic representation of St1Cas9 from the LMD-9 (A) and LMG18311 (B) strains along with their predicted and experimentally determined PAM. The hybrid protein (AB) containing the N-terminal of St1Cas9 LMD-9 and the C-terminal domain (WED + PI) of St1Cas9 LMG18311 is also represented. (**b**) Surveyor assays to determine the activity of St1Cas9 variants programmed with sgRNAs targeting different PAM within *FANCF*. K562 cells were transiently transfected with a single vector (1µg) driving the expression of St1Cas9's and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control. (**c**) Same as in (**b**) but targeting *RUNX1*. *A stretch of 15 consecutive "A" within the *RUNX1* amplicon in proximity to the cleavage site prevents accurate quantification of indels.

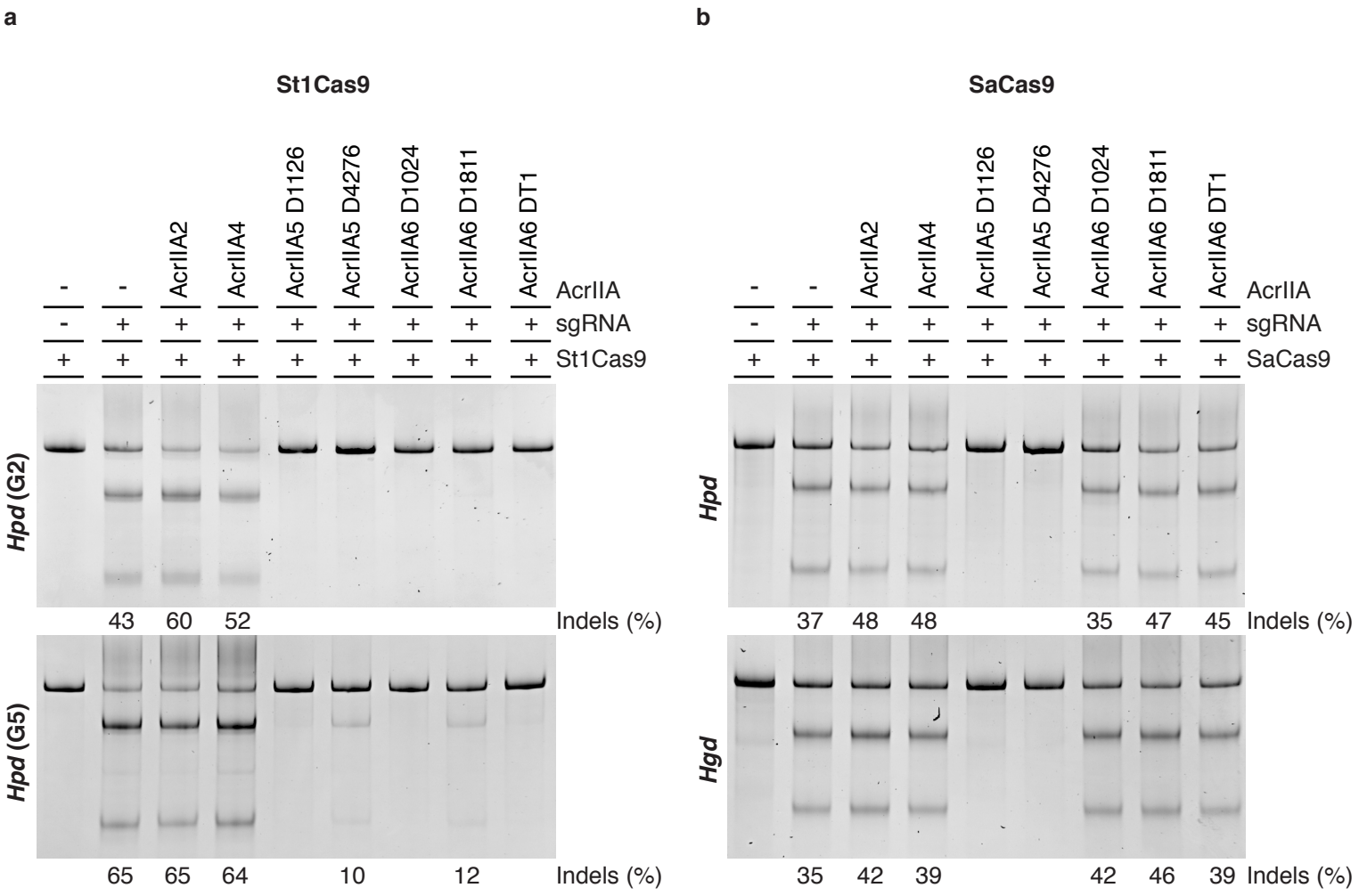


Figure 4 | Anti-CRISPR activity of AcrIIAs against St1Cas9 and SaCas9. (**a**) Neuro-2a cells were transfected with a dual St1Cas9 and sgRNA expression plasmid (0.5µg) targeting *Hpd* (G2) and *Hpd* (G5). Where indicated, a second vector (0.5µg) was used to co-express human codon-optimized AcrIIAs. The Surveyor assay was performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. (**b**) Same as in (**a**) but using SaCas9 to target *Hpd* and *Hgd*.

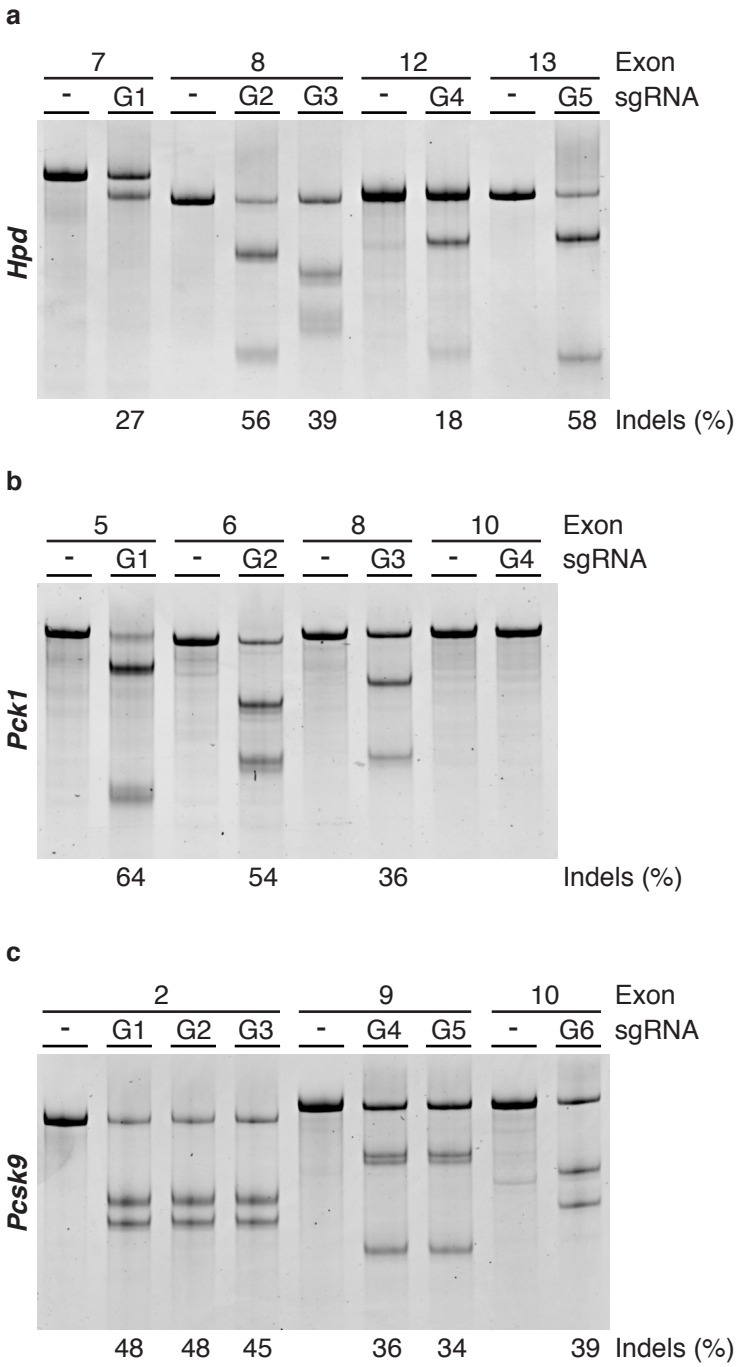


Figure 5 | Screening for active sgRNAs targeting genes affecting liver function in mouse cells. (**a**) Surveyor assays to determine St1Cas9 LMD-9 activity programmed with various sgRNAs targeting *Hpd*. Neuro-2a cells were transiently transfected with a single vector (0.5µg) driving the expression of St1Cas9 and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control. (**b**) Same as in (**a**) but targeting *Pck1*. (**c**) Same as in (**a**) but targeting *Pcsk9*.

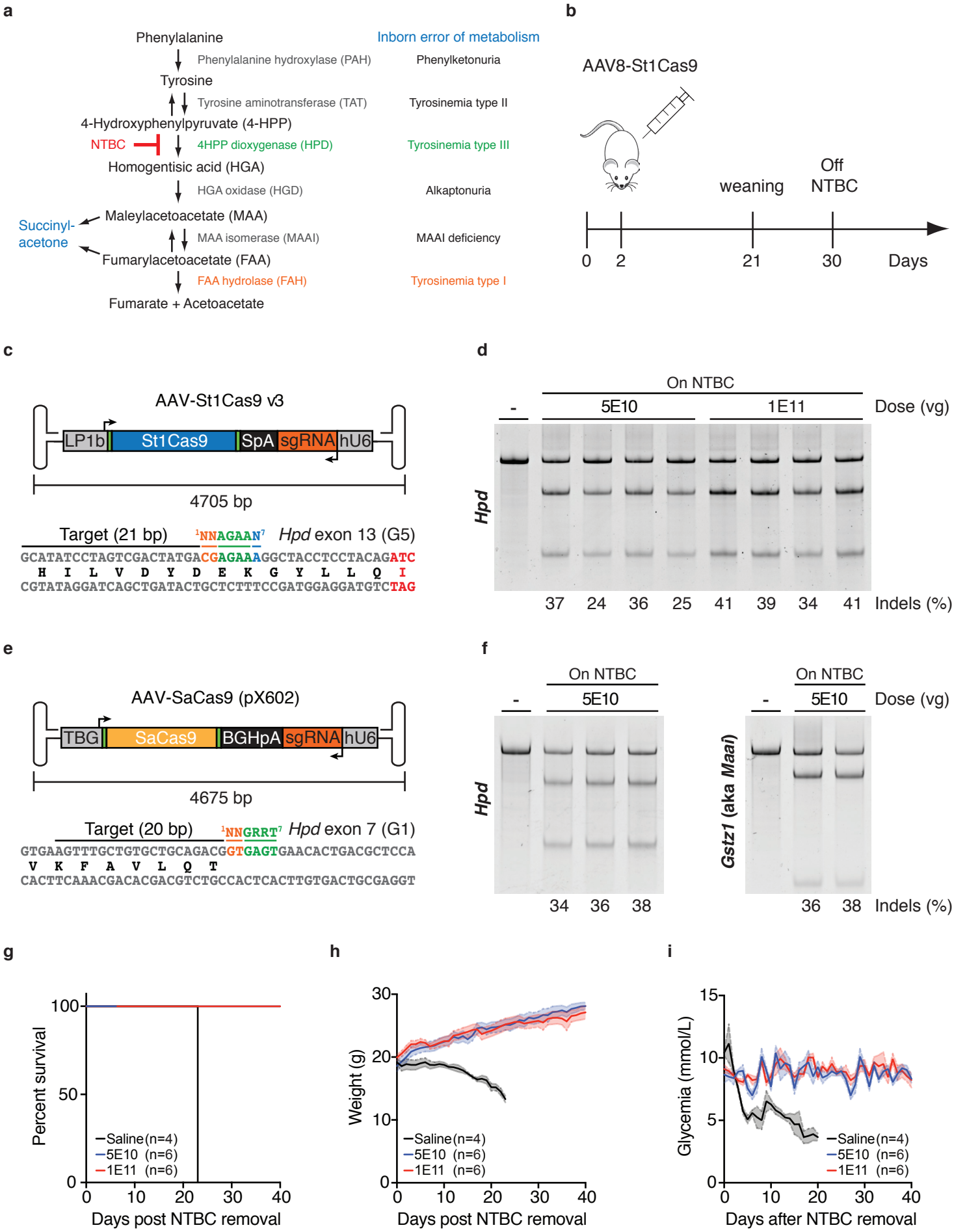

a
Phenylalanine
Inborn error of metabolism
Phenylalanine hydroxylase (PAH)
Phenylketonuria
Tyrosine
Tyrosine aminotransferase (TAT)
Tyrosinemia type II
4-Hydroxyphenylpyruvate (4-HPP)
NTBC
4HPP dioxygenase (HPD)
Tyrosinemia type III
Homogentisic acid (HGA)
HGA oxidase (HGD)
Alkaptonuria
Succinyl-acetone
Maleylacetoacetate (MAA)
MAA isomerase (MAAI)
MAAI deficiency
Fumarylacetoacetate (FAA)
FAA hydrolase (FAH)
Tyrosinemia type I
Fumarate + Acetoacetate
b
AAV8-St1Cas9
weaning
Off NTBC
0
2
21
30
Days
c
AAV-St1Cas9 v3
LP1b
St1Cas9
SpA
sgRNA
hU6
4705 bp
Target (21 bp)
1NNAGAAN7
Hpd exon 13 (G5)
GCATATCCTAGTCGACTATGACGAGAAAGGCTACCTCCTACAGATC
H I L V D Y D E K G Y L L Q I
CGTATAGGATCAGCTGATACTGCTCTTTCCGATGGAGGATGTCTAG
d
On NTBC
-
5E10
1E11
Dose (vg)
Hpd
37 24 36 25 41 39 34 41
Indels (%)
e
AAV-SaCas9 (pX602)
TBG
SaCas9
BGHpA
sgRNA
hU6
4675 bp
Target (20 bp)
1NNGRRT7
Hpd exon 7 (G1)
GTGAAGTTTGCTGTGCTGCAGACGGTGAGTGAACACTGACGCTCCA
V K F A V L Q T
CACTTCAAACGACACGACGTCTGCCACTCACTTGTGACTGCGAGGT
f
On NTBC
-
5E10
Dose (vg)
Hpd
Gstz1 (aka Maai)
34 36 38
36 38
Indels (%)
g
Percent survival
100
50
0
Saline (n=4)
5E10 (n=6)
1E11 (n=6)
0 10 20 30 40
Days post NTBC removal
h
Weight (g)
30
20
10
0
Saline (n=4)
5E10 (n=6)
1E11 (n=6)
0 10 20 30 40
Days post NTBC removal
i
Glycemia (mmol/L)
15
10
5
0
Saline (n=4)
5E10 (n=6)
1E11 (n=6)
0 10 20 30 40
Days after NTBC removal

Figure 6 | *In vivo* metabolic pathway rewiring via rAAV8-mediated delivery of holo-St1Cas9. (**a**) The tyrosine degradation pathway and associated genetic disorders. (**b**) Experimental design for *in vivo* editing. Neonatal (2 days old) *Fah*$^{-/-}$ mice were injected with rAAV8-St1Cas9 or saline into the retro-orbital sinus, weaned at 21 days, and NTBC was removed at 30 days of age. Mice were assayed for phenotypic and metabolic correction and for gene disruption efficacy. Mice off NTBC were killed when they lost 20% of their body weight. (**c**) Schematic representations of the rAAV vector and St1Cas9 target site (G5) within exon 13 of *Hpd*. Target sequence, PAM and position of the I335M mutation (red) causing type III tyrosinemia in humans are shown. Also annotated is the liver-specific promoter (LP1b) promoter, synthetic polyadenylation sequence (SpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. (**d**) Neonatal *Fah*$^{-/-}$ mice were injected into the retro-orbital sinus with either 5E10 or 1E11 vector genomes (vg) of rAAV8-St1Cas9 v3 targeting *Hpd* exon 13 (G5) and killed 28 days following injection. Genomic DNA was extracted from whole liver samples and the Surveyor assay was used to determine the frequency of St1Cas9-induced gene disruption as the % Indels indicated at the base of each lane. Each lane represents a different mouse. A mouse injected with saline (-) was used as a negative control. (**e**) Same as in (**c**) but for SaCas9 with its target site (G1) in exon 7 of *Hpd*. Annotated is the human thyroxine binding globulin (TBG) promoter and bovine growth hormone polyadenylation sequence (BGHpA). (**f**) Same as in (**d**) but for SaCas9 at *Hpd* and *Gstz1* (*Maai*). (**g**) Survival analysis following NTBC removal in mice treated with rAAV8-St1Cas9 as described in (**b**). Number of mice per group (n) and rAAV doses (vg) is indicated. (**h**) Same as in (**g**) but body weight was measured daily. (**i**) Same as in (**g**) but glycemia was monitored in non-fasted mice. Solid lines designate the mean and error bars are represented by shaded areas and denote s.e.m.

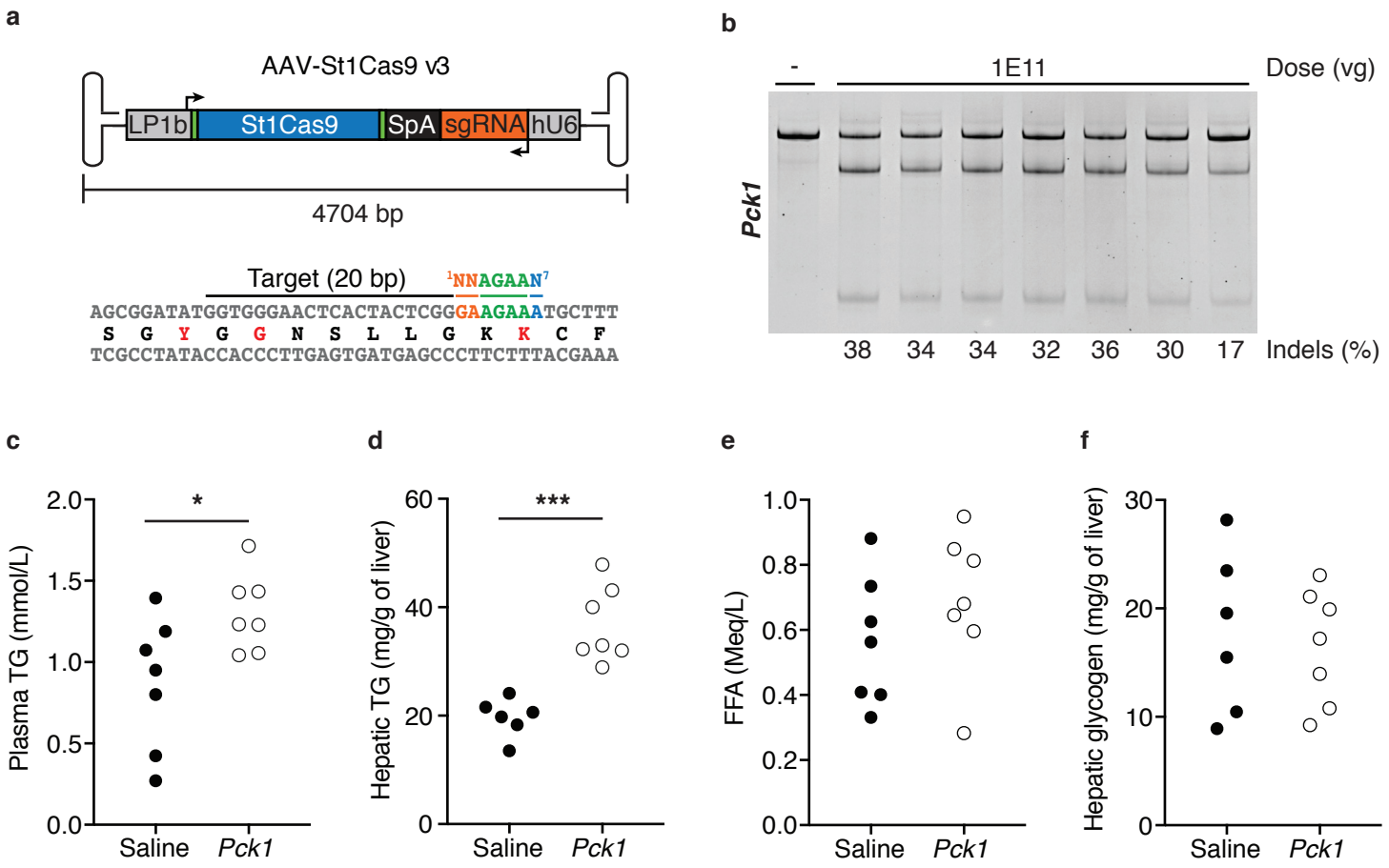
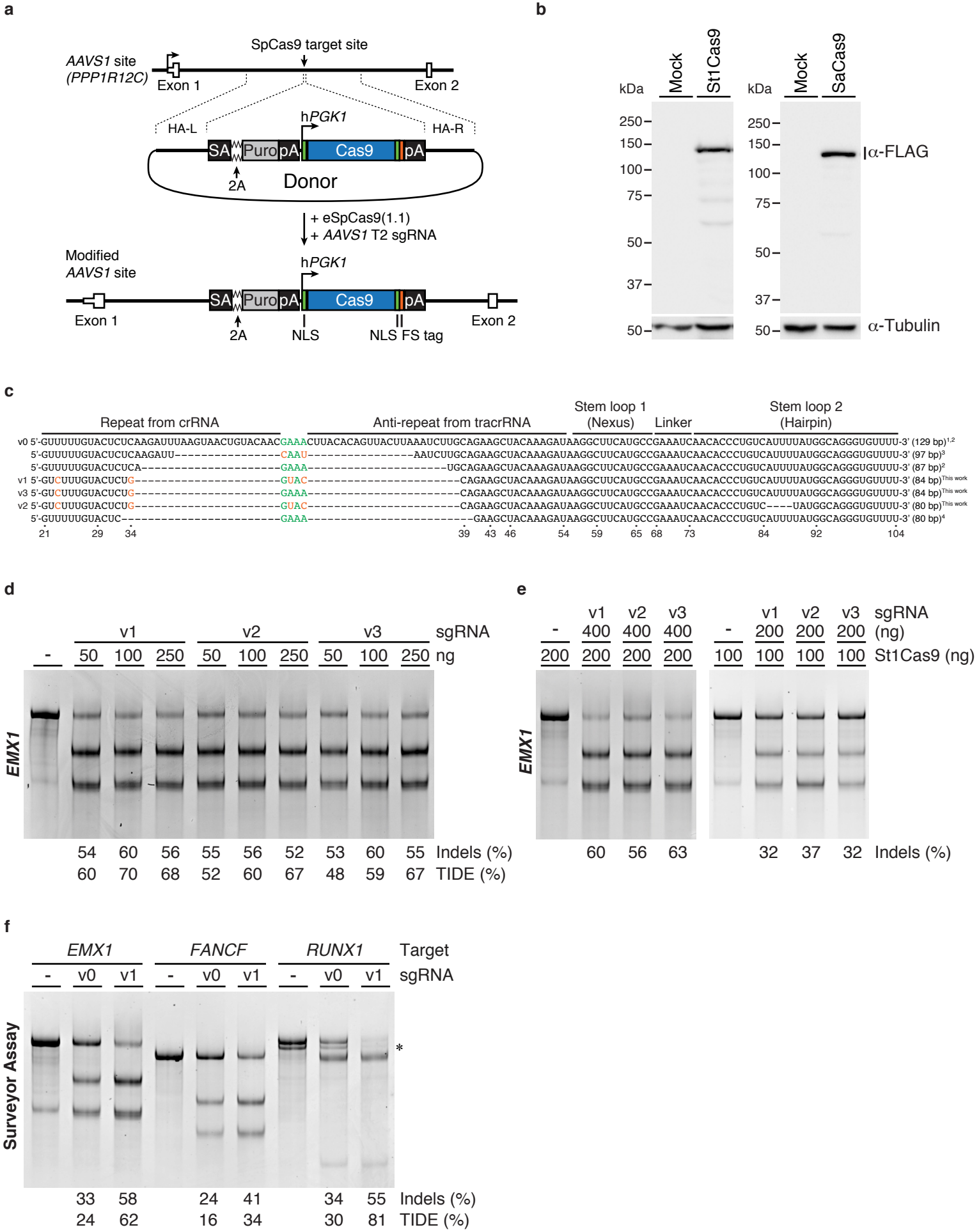


Figure 7 | *In vivo* disruption of *Pck1* via rAAV8-St1Cas9 raises hepatic and plasma triglyceride levels in C57BL/6N mice. (**a**) Schematic representations of the rAAV vector and St1Cas9 target site (G1) within exon 5 of *Pck1*. Target sequence, PAM and position of amino acids involved in important product interactions (Tyr 235, Gly 237) and Mn^{2+} coordination (Lys 244) are shown. Also annotated are the liver-specific promoter (LP1b) promoter, synthetic polyadenylation sequence (SpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. (**b**) Neonatal C57BL/6N mice were injected into the retro-orbital sinus with 1E11 vector genomes (vg) of rAAV8-St1Cas9, 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 of St1Cas9-induced gene disruption as the % Indels indicated at the base of each lane. Each lane represents a different mouse. A mouse injected with saline (-) was used as a negative control. (**c**) Plasma triglyceride levels were measured after 24 hours of fasting in mice injected with saline and rAAV8-St1Cas9. Each dot represents a different animal. (**d**) Same as (**c**) but hepatic triglycerides were extracted and quantified. (**e**) Same as (**c**) but for free fatty acids. (**f**) Same as (**c**) but hepatic glycogen was extracted and quantified. * $p < 0.05$, *** $p < 0.0005$ according to Student's t-test.


a
AAVS1 site
(PPP1R12C)
SpCas9 target site
Exon 1
Exon 2
HA-L
HA-R
hPGK1
SA
Puro
pA
Cas9
pA
2A
Donor
+ eSpCas9(1.1)
+ AAVS1 T2 sgRNA
Modified
AAVS1 site
hPGK1
SA
Puro
pA
Cas9
pA
Exon 1
2A
NLS
NLS FS tag
Exon 2
b
kDa
Mock
St1Cas9
250
150
100
75
50
37
50
α-FLAG
α-Tubulin
kDa
Mock
SaCas9
250
150
100
75
50
37
50
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
v1
v2
v3
sgRNA
- 50 100 250 50 100 250 50 100 250 ng
EMX1
54 60 56 55 56 52 53 60 55 Indels (%)
60 70 68 52 60 67 48 59 67 TIDE (%)
e
v1 v2 v3 sgRNA
- 400 400 400 (ng)
200 200 200 200 St1Cas9 (ng)
EMX1
60 56 63 Indels (%)
v1 v2 v3 sgRNA
- 200 200 200 (ng)
100 100 100 100 St1Cas9 (ng)
32 37 32 Indels (%)
f
EMX1
FANCF
RUNX1
Target
- v0 v1 - v0 v1 - v0 v1 sgRNA
Surveyor Assay
*
33 58 24 41 34 55 Indels (%)
24 62 16 34 30 81 TIDE (%)

Supplementary Figure 1 | Engineered CRISPR1-StCas9 system drives robust gene editing in human cells. (**a**) Schematic representation of the targeted integration of tagged St1Cas9 and SaCas9 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 Cas9-tag protein expression in K562 clones and 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). (**d**) K562 cells stably expressing St1Cas9 were transfected with indicated sgRNA expression vectors at increasing doses and the Surveyor and TIDE assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. (**e**) K562 cells were transiently transfected with an St1Cas9 expression vector (100 or 200 ng) in addition to the indicated sgRNA expression plasmids (200 and 400 ng) and Surveyor assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. (**f**) Same as (**e**) but K562 cells were transiently transfected with an St1Cas9 expression vector (0.5μg) in addition to the indicated sgRNA expression plasmids (0.8μg). *A stretch of 15 consecutive "A" within the *RUNX1* amplicon in proximity to the cleavage site prevents accurate quantification of indels by the Surveyor assay. An expression vector encoding EGFP (-) was used as a negative control.

a

	Target	Extended PAM ¹NNAGAAN⁷
***EMX1* G2**	AAACCGGAGGACAAAGTACAAACG	GCAGAAGCT
***EMX1* G3**	GAACCGGAGGACAAAGTACAAACG	GCAGAAGCT
***EMX1* G4**	GGAGGAGGAAGGGCCTGAGTCCGA	GCAGAAGAA
***EMX1* G5**	GGAGGAAGGGCCTGAGTCCGAGCA	GAAGAAGAA
***EMX1* G6**	GGAAGGGCCTGAGTCCGAGCAGAA	GAAGAAGGG
***EMX1* G7**	CGCCCAGGCAGGCAGGCTCTCCGA	GGAGAAGGC
***FANCF* G1**	GCTCGGAAAAGCGATCCAGGTGCT	GCAGAAGGG
***FANCF* G2**	TGACGTAGGTAGTGCTTGAGACCG	CCAGAAGCT
***RUNX1* G1**	CTGTACTAATCAGATGGAAGCTCT	TCAGAAATG
***RUNX1* G2**	AAGAAATCATTGAGTCCCCCGCCT	TCAGAAGTG
***RUNX1* G3**	CCCTGAGGTATCCAGCAGAGGGGA	GAAGAAAGA
***RUNX1* G4**	AGTCCCAGAGGTATCCAGCAGAGG	GGAGAAGAA
***ATP1A1* G1**	TGTAGCAGCTTGGATGCTATAAGC	CAAGAAACA
***Hpd* G1**	ACTTGGTACCCCACGCAGAAAGCT	CGAGAACGG
***Hpd* G2**	GTATGTTGGGGCCTCGAATCCAGG	TAAGAAACG
***Hpd* G3**	TGGAGATACCACACACACCCTGGT	GGAGAAGAT
***Hpd* G4**	TTTGTAGTAAGAAGATGGGGCGGC	CAAGAACTC
***Hpd* G5**	GCTGCATATCCTAGTCGACTATGA	CGAGAAAGG
***Pcsk9* G1**	CAACAGGTCACTGCTCATCTTCAC	CAAGAAGCC
***Pcsk9* G2**	AACAGGGTCACTGCTCATCTTCAC	CAAGAAGCC
***Pcsk9* G3**	AACGAGGTCACTGCTCATCTTCAC	CAAGAAGCC
***Pcsk9* G4**	TCACGCACGACGCCTCCCGCTCCT	GGAGAAGCT
***Pcsk9* G5**	TCAGCCACGACGCCTCCCGCTCCT	GGAGAAGCT
***Pcsk9* G6**	CCTGGAGACCCATGTCCACTGCCA	CCAGAAGGA
***Pck1* G1**	ATATGGTGGGAACTCACTACTCGG	GAAGAAATG
***Pck1* G2**	CCTGGGCATAACTAACCCCGAAGG	CAAGAAGAA
***Pck1* G3**	AATGGGGCACTGGCTGGCAGGGGT	GCAGAATCT

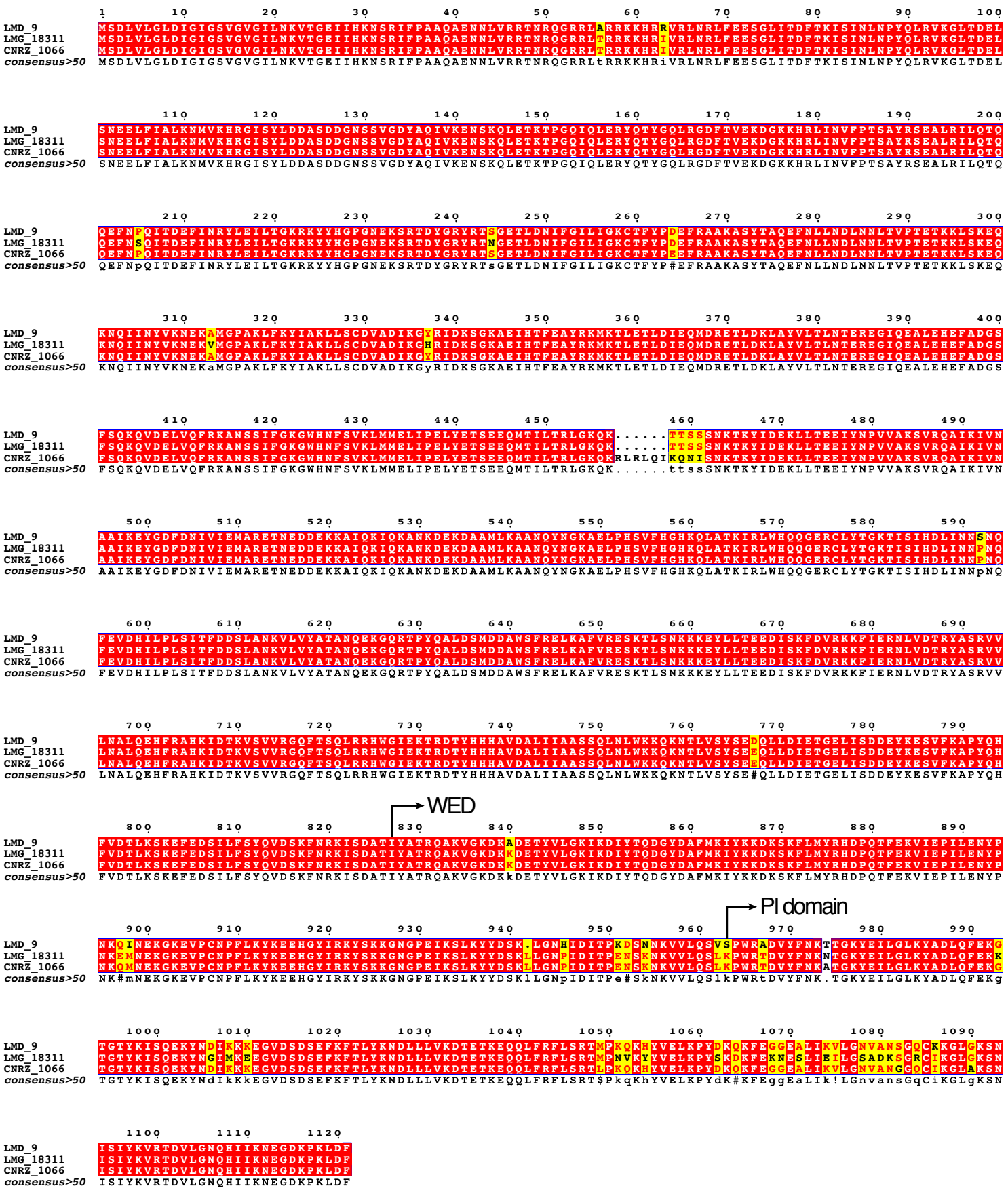
b

	Target	Extended PAM ¹NNGGAAN⁷
***FANCF* G4**	GTAGTGCTTGAGACCGCCAGAAGC	TCGGAAAAG
***FANCF* G5**	GTAGTGCTTGAGACCGCCAGAAGC	TCGGAAAAG
***FANCF* G6**	CGAGGGCCTGGAAGTTCGCTAATC	CCGGAACTG
***VEGFA* G1**	TGGGGAGAGGGACACACAGATCTA	TTGGAATCC
***VEGFA* G2**	GCCTGAGAGCCGTTCCCTCTTTGC	TAGGAATAT
***VEGFA* G3**	CGCGGGGCATTGGCGAGGAGGGAG	CAGGAAAGT
***VEGFA* G4**	AGCCTGAAAATTACCCATCCGCCC	CCGGAAACT
***VEGFA* G5**	AGCCTGAAAATTACCCATCCGCCC	CCGGAAACT

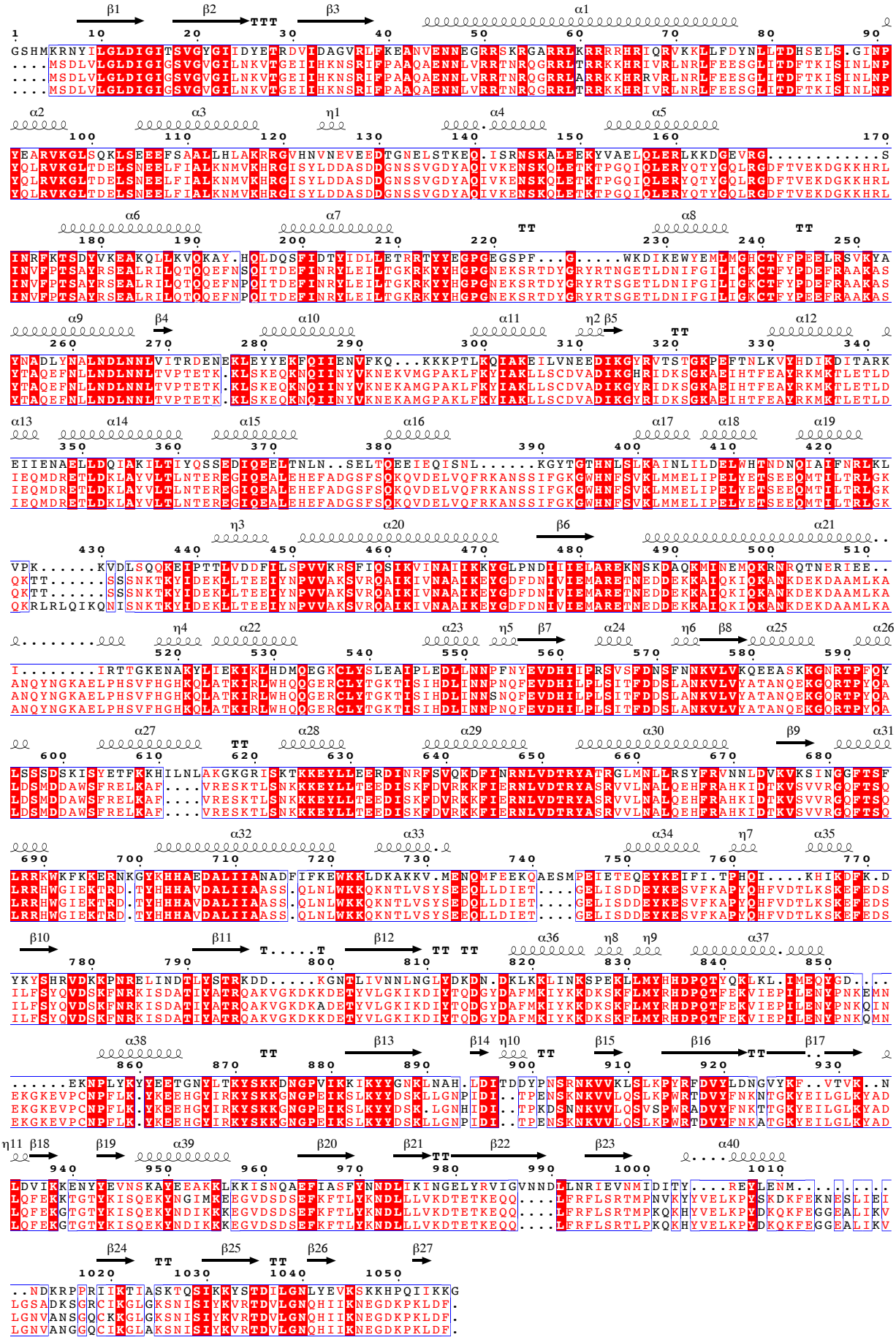
c

	Target	Extended PAM ¹NNGCAAN⁷
***FANCF* G202**	GAGCGGCGGCTGCACAACCAGTGG	AGGCAAGAG
***RUNX1* G202**	TGCCATTTCATTACAGGCAAAGCT	GAGCAAAAG
***AAVS1* G201**	TCGGGGACACAGGATCCCTGGAGG	CAGCAAACA

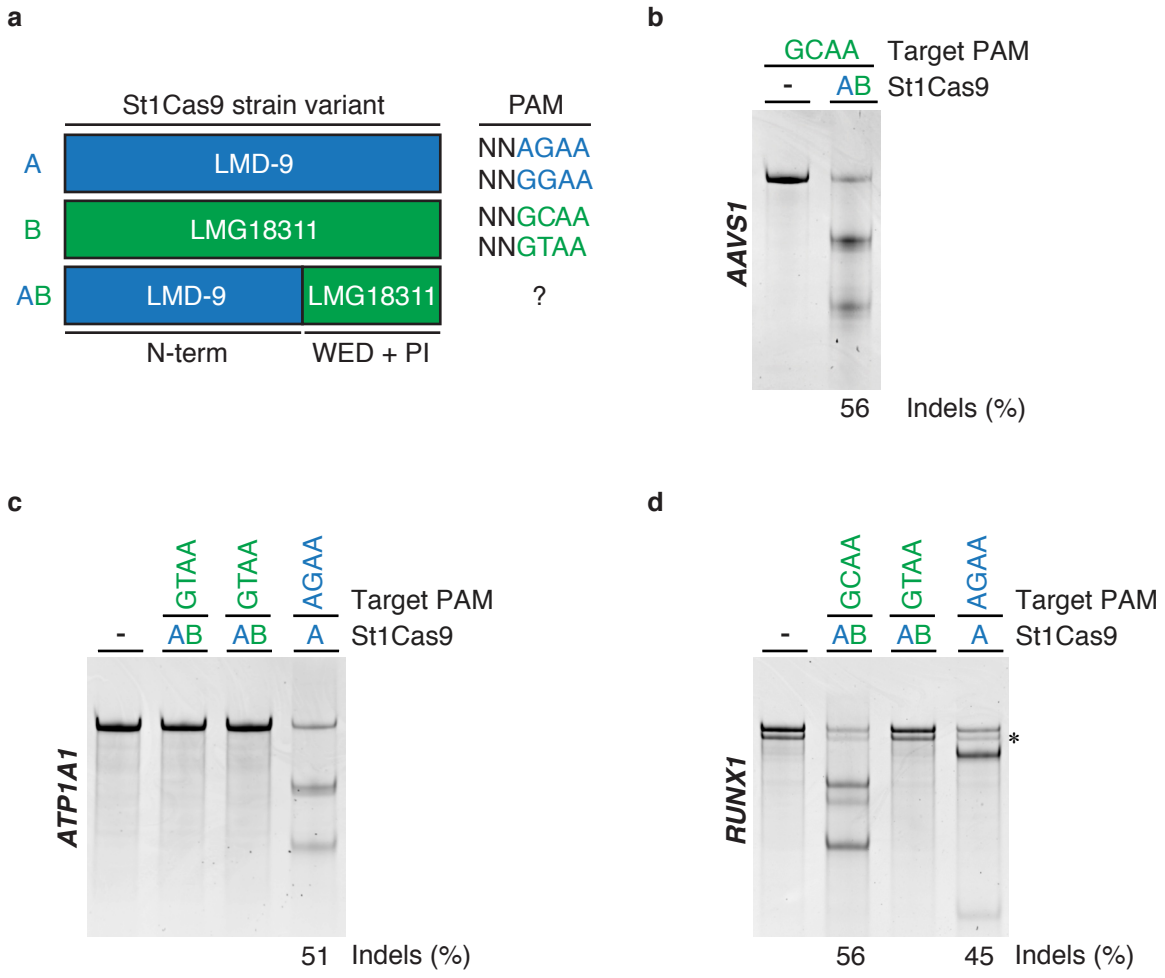
Supplementary Figure 2 | Alignment of all active sgRNAs as a function of the PAM sequence. (**a**) NNAGAA target PAM for St1Cas9 LMD-9. (**b**) NNGGAA target PAM for St1Cas9 LMD-9. (**c**) NNGCAA target PAM for St1Cas9 LMD-9/LMG18311 hybrid.



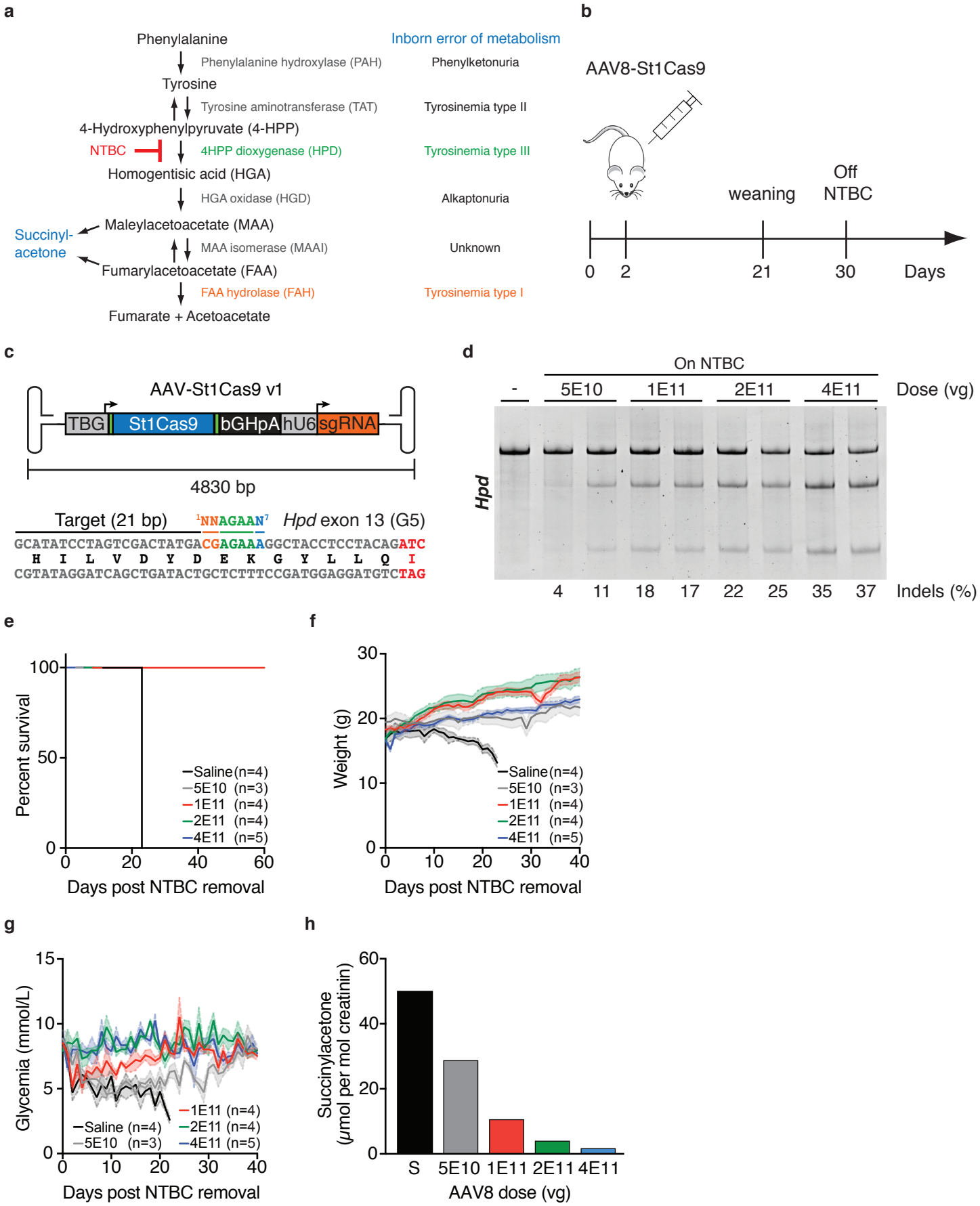
Supplementary Figure 3 | Amino acid sequence alignment of St1Cas9 from different strains. Identical residues are highlighted in red. The position of the WED and PAM-interating domain (PI) are indicated by arrows. This region of the protein has diverged the most as compared to the N-terminal segment. In SaCas9, the PAM duplex is sandwiched between the WED and PI domains[5]. Alignment was performed with Clustal Omega[6] and ESPript[7].



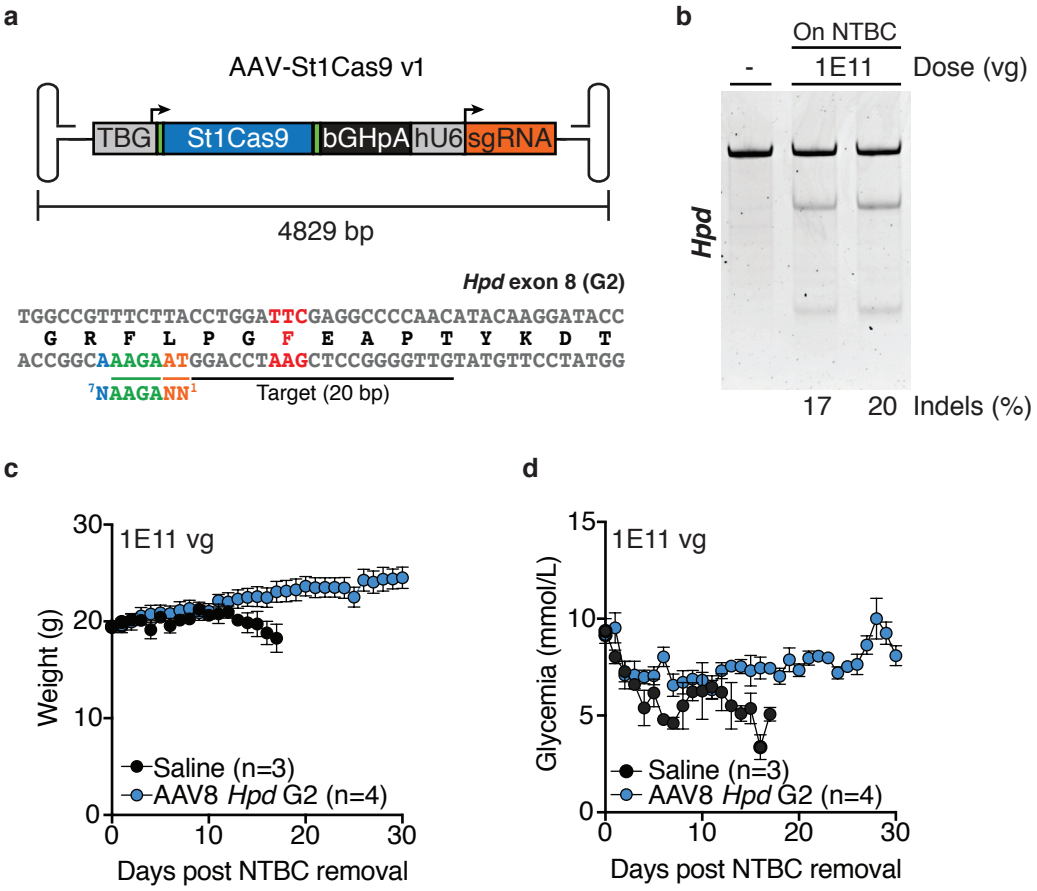
Supplementary Figure 4 | Amino acid sequence alignment of SaCas9 with St1Cas9 from different strains. The secondary structure of SaCas9[5] (5CZZ) is displayed above the sequences, which are numbered according to the residues of SaCas9. Identical residues are highlighted in red. Alignment was performed with Clustal Omega[6] and ESPript[7].



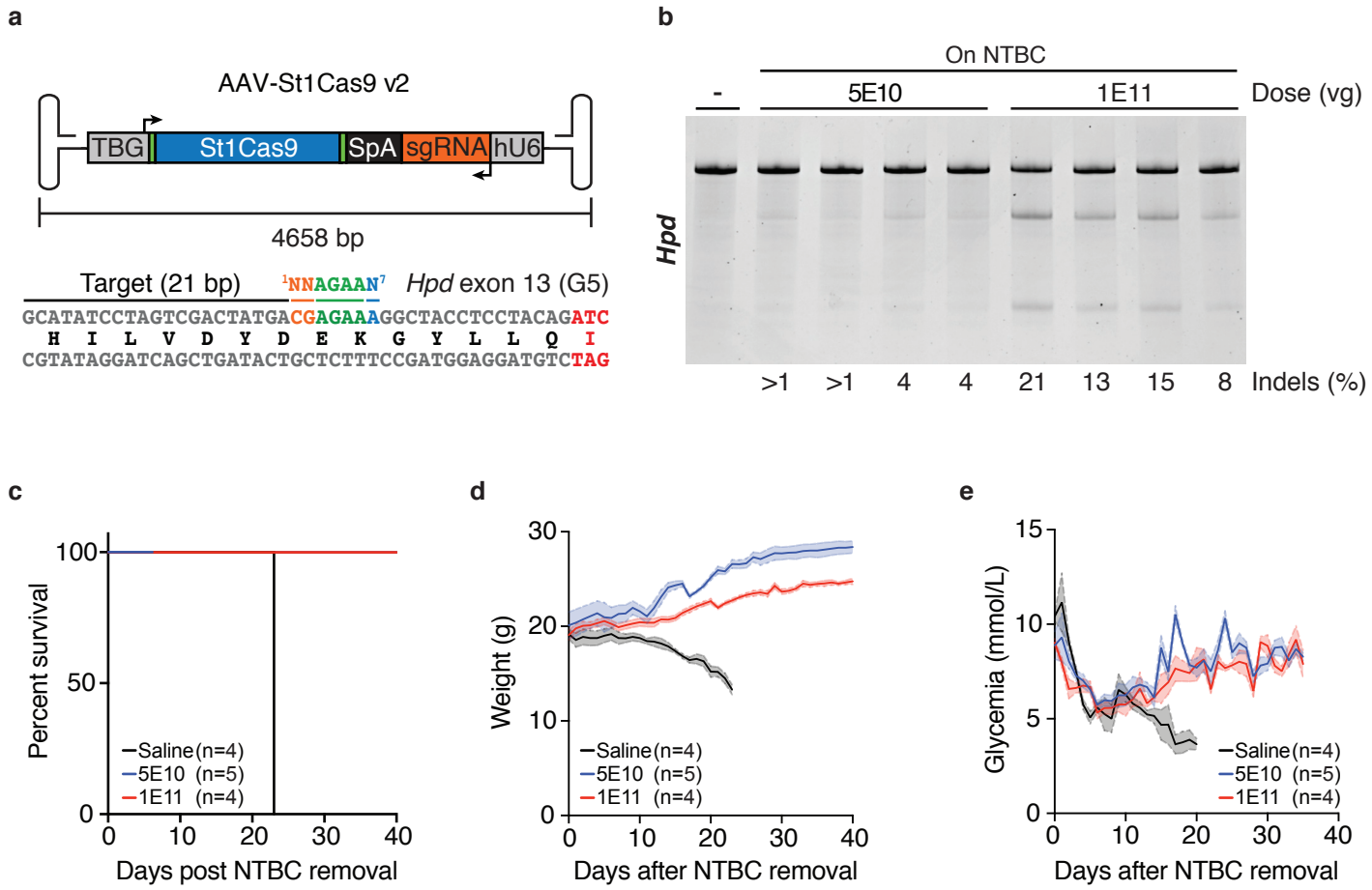
Supplementary Figure 5 | Rewiring St1Cas9 LMD-9 to target a distinct PAM sequence using a naturally occurring variant. (**a**) Schematic representation of St1Cas9 from the LMD-9 (A) and LMG18311 (B) strains along with their predicted and experimentally determined PAM. The hybrid protein (AB) containing the N-terminal of St1Cas9 LMD-9 and the C-terminal domain (WED + PI) of St1Cas9 LMG18311 is also represented. (**b**) Surveyor assays to determine the activity of St1Cas9 variant AB programmed with a sgRNA targeting a NNGCAA PAM within *AAVS1*. K562 cells were transiently transfected with a single vector (1µg) driving the expression of St1Cas9 and its sgRNA. Surveyor assays were performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control. (**c**) Same as in (**b**) but attempting to target *ATP1A1* at NNGTAA PAM. (**d**) Same as (**c**) but targeting *RUNX1*. *A stretch of 15 consecutive "A" within the *RUNX1* amplicon in proximity to the cleavage site prevents accurate quantification of indels.


a
Phenylalanine
Phenylalanine hydroxylase (PAH)
Tyrosine
Tyrosine aminotransferase (TAT)
4-Hydroxyphenylpyruvate (4-HPP)
NTBC
4HPP dioxygenase (HPD)
Homogentisic acid (HGA)
HGA oxidase (HGD)
Maleylacetoacetate (MAA)
Succinyl-acetone
MAA isomerase (MAAI)
Fumarylacetoacetate (FAA)
FAA hydrolase (FAH)
Fumarate + Acetoacetate
Inborn error of metabolism
Phenylketonuria
Tyrosinemia type II
Tyrosinemia type III
Alkaptonuria
Unknown
Tyrosinemia type I
b
AAV8-St1Cas9
weaning
Off NTBC
0
2
21
30
Days
c
AAV-St1Cas9 v1
TBG
St1Cas9
bGHpA
hU6
sgRNA
4830 bp
Target (21 bp)
1NNAGAAN7
Hpd exon 13 (G5)
GCATATCCTAGTCGACTATGACGAGAAAGGCTACCTCCTACAGATC
H I L V D Y D E K G Y L L Q I
CGTATAGGATCAGCTGATACTGCTCTTTCCGATGGAGGATGTCTAG
d
On NTBC
-
5E10
1E11
2E11
4E11
Dose (vg)
Hpd
4
11
18
17
22
25
35
37
Indels (%)
e
Percent survival
Days post NTBC removal
Saline (n=4)
5E10 (n=3)
1E11 (n=4)
2E11 (n=4)
4E11 (n=5)
f
Weight (g)
Days post NTBC removal
Saline (n=4)
5E10 (n=3)
1E11 (n=4)
2E11 (n=4)
4E11 (n=5)
g
Glycemia (mmol/L)
Days post NTBC removal
Saline (n=4)
5E10 (n=3)
1E11 (n=4)
2E11 (n=4)
4E11 (n=5)
h
Succinylacetone (μmol per mol creatinin)
S
5E10
1E11
2E11
4E11
AAV8 dose (vg)

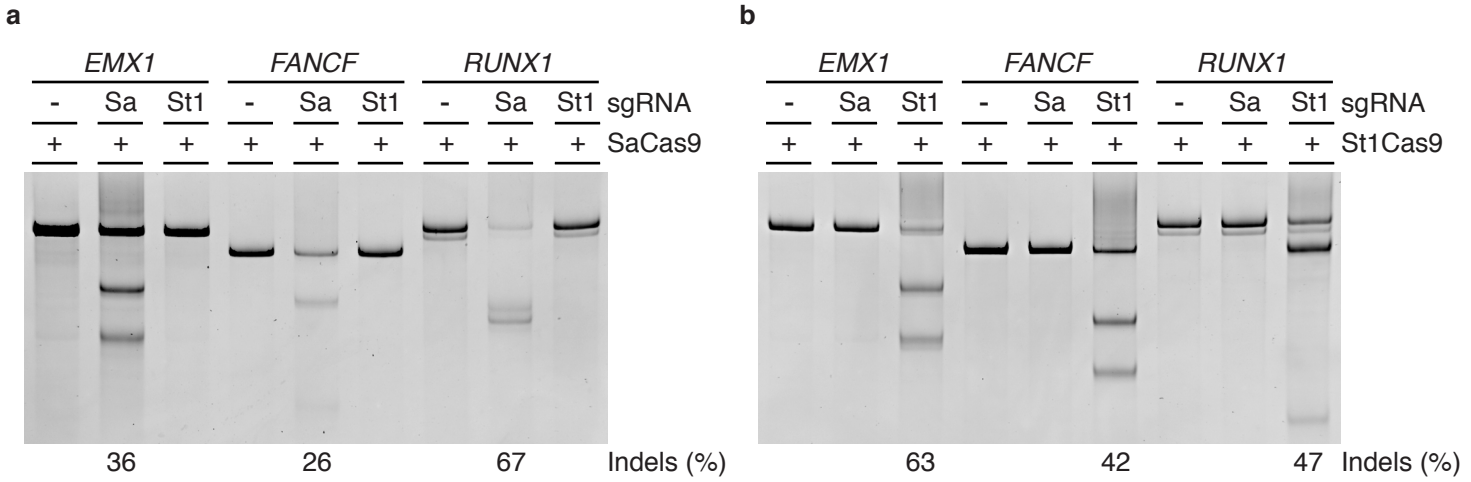
Supplementary Figure 6 | *In vivo* metabolic pathway rewiring via rAAV8-mediated delivery of holo-St1Cas9. (**a**) The tyrosine degradation pathway and associated genetic disorders. (**b**) Experimental design for *in vivo* editing. Neonatal (2 days old) *Fah*$^{-/-}$ mice were injected with rAAV8-St1Cas9 or saline into the retro-orbital sinus, weaned at 21 days, and NTBC was removed at 30 days of age. Mice were assayed for phenotypic and metabolic correction and for gene disruption efficacy. Mice off NTBC were killed when they lost 20% of their body weight. (**c**) Schematic representations of the rAAV vector and St1Cas9 target site (G5) within exon 13 of *Hpd*. Target sequence, PAM and position of the I335M mutation (red) causing type III tyrosinemia in humans are shown. Also annotated is the human thyroxine binding globulin (TBG) promoter, bovine growth hormone polyadenylation sequence (BGHpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. (**d**) Neonatal *Fah*$^{-/-}$ mice were injected into the retro-orbital sinus with either 5E10, 1E11, 2E11 or 4E11 vector genomes (vg) of rAAV8-St1Cas9 targeting *Hpd* exon 13 (G5) and killed 28 days following injection. Genomic DNA was extracted from whole liver samples and the Surveyor assay was used to determine the frequency of St1Cas9-induced gene disruption as the % Indels indicated at the base of each lane. Each lane represents a different mouse. A mouse injected with saline (-) was used as a negative control. (**e**) Survival analysis following NTBC removal in mice treated as described in (**b**). Number of mice per group (n) and rAAV doses (vg) is indicated. (**f**) Same as in (**e**) but body weight was measured daily. Solid lines designate the mean and error bars are represented by shaded areas and denote s.e.m. (**g**) Same as in (**f**) but glycemia was monitored in non-fasted mice. (**h**) Same as in (**e**) but succinylacetone levels in urine were determined 15 days following NTBC removal. Samples were collected from the indicated treatment groups over a 24 hours period using metabolic cages.



Supplementary Figure 7 | *In vivo* metabolic pathway rewiring via rAAV8-mediated delivery of holo-St1Cas9. (**a**) Schematic representations of the rAAV vector and St1Cas9 target site (G2) within exon 8 of *Hpd*. Target sequence, PAM and position of the Y160C mutation (red) causing type III tyrosinemia in humans are shown. Note that while this region of the protein is well conserved between human and mouse, a phenylalanine is found at this position in mouse, rat, pig, and *C. elegans*. Also annotated is the human thyroxine binding globulin (TBG) promoter, bovine growth hormone polyadenylation sequence (BGHpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. (**b**) Neonatal $Fah^{-/-}$ mice were injected into the retro-orbital sinus with 1E11 vector genomes (vg) of rAAV8-St1Cas9 targeting *Hpd* exon 8 (G2) and killed 28 days following injection. Genomic DNA was extracted from whole liver samples and the Surveyor assay was used to determine the frequency of St1Cas9-induced gene disruption as the % Indels indicated at the base of each lane. Each lane represents a different mouse. A mouse injected with saline (-) was used as a negative control. (**c**) Body weight was measured daily following NTBC removal in mice treated as in **Supplementary Fig. 6**. Number of mice per group (n) is indicated. Dots designate the mean and error bars denote s.e.m. (**d**) Same as in (**c**) but glycemia was monitored in non-fasted mice.



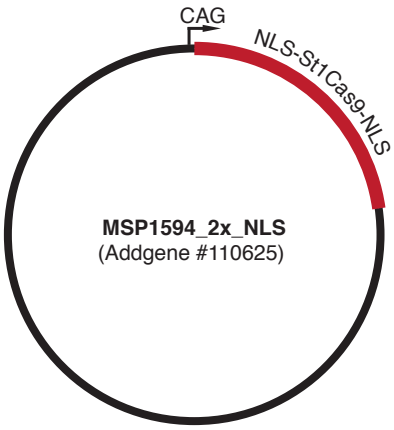
Supplementary Figure 8 | Testing alternative rAAV-St1Cas9 vector architectures. (**a**) Schematic representations of the second-generation rAAV-St1Cas9 (v2) vector of similar size to the parent AAV genome (~4.7kb). Annotated is the human thyroxine binding globulin (TBG) promoter, synthetic polyadenylation sequence (SpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. (**b**) Neonatal (2 days old) *Fah*$^{-/-}$ mice were injected with either 5E10 or 1E11 vg rAAV8-St1Cas9 v2 targeting *Hpd* exon 13 (G5) or saline into the retro-orbital sinus and killed 28 days post injection. Genomic DNA was extracted from whole liver samples and the Surveyor assay was used to determine the frequency of St1Cas9-induced gene disruption as the % Indels indicated at the base of each lane. Each lane represents a different mouse. A mouse injected with saline (-) was used as a negative control. (**c**) Survival analysis following NTBC removal in mice treated as described **Supplementary Fig. 6**. Number of mice per group (n) and rAAV doses (vg) is indicated. (**d**) Same as in (**c**) but body weight was measured daily. Solid lines designate the mean and error bars are represented by shaded areas and denote s.e.m. (**e**) Same as in (**d**) but glycemia was monitored in non-fasted mice.



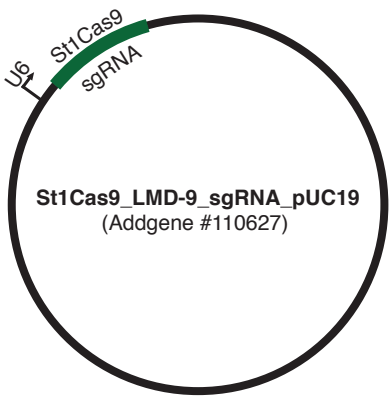
Supplementary Figure 9 | sgRNAs for SaCas9 and St1Cas9 are not functionally interchangeable. (**a**) A stable K562 cell line constitutively expressing SaCas9 was transfected with expression vectors (0.25μg) for its cognate sgRNA or the St1Cas9 sgRNA. The Surveyor assay was performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. An expression vector encoding EGFP (-) was used as a negative control. (**b**) Same as in (**a**) but in a K562 cell line constitutively expressing St1Cas9. These data indicate that an sgRNA programmed to specify cleavage by St1Cas9 at one site cannot recruit and induce cutting by SaCas9 at this same site and vice-versa.

Plasmids for *in vitro* genome editing

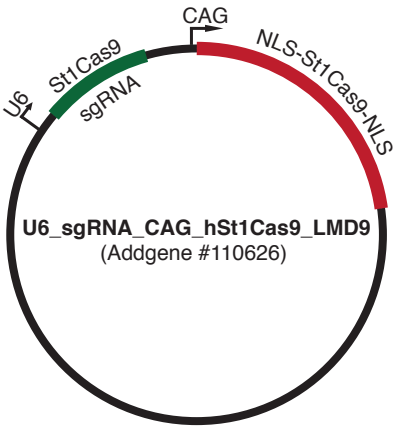
Mammalian expression vector for St1Cas9 LMD-9



hU6-driven sgRNA expression plasmid for St1Cas9 LMD-9

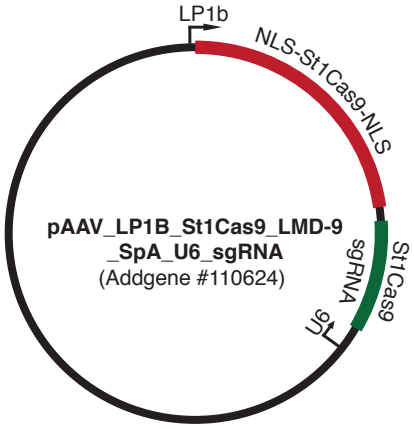


Dual expression vector for St1Cas9 LMD-9 and its sgRNA



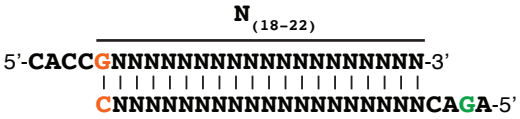
Plasmid for *in vivo* genome editing

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

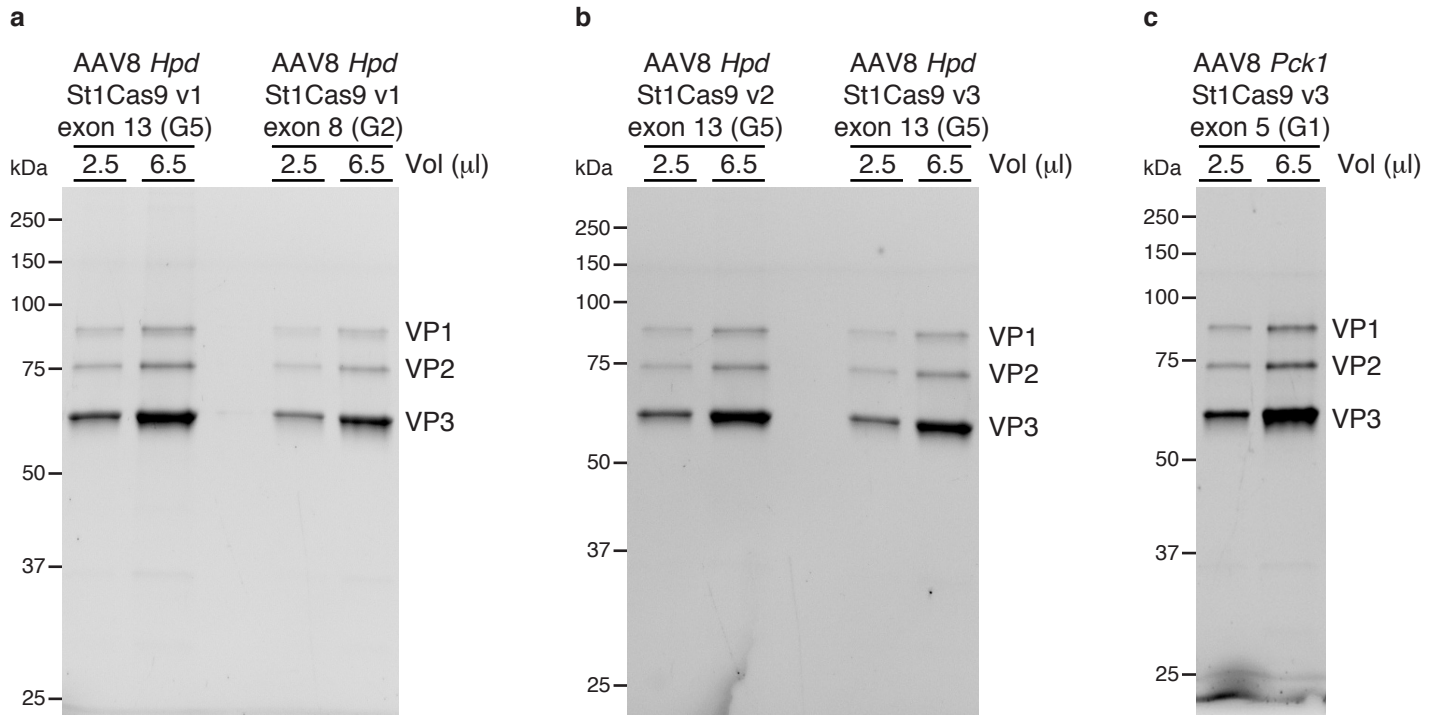


Cloning

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



Supplementary Figure 10 | St1Cas9 vectors available from Addgene.



Supplementary Figure 11 | Purified rAAV8 vectors used in this study. (**a**) SDS-PAGE gel analysis of first-generation rAAV8-St1Cas9 vectors targeting *Hpd*. Fixed volumes of purified viruses were run on mini-PROTEAN TGX Stain-Free Gels. Molecular weight (kDa) of protein standards are indicated. (**b**) Same as in (**a**) but for second and third generation rAAV8-St1Cas9 vectors. (**c**) Same as in (**a**) but analyzing the third generation rAAV8-St1Cas9 vector targeting *Pck1*.

Supplementary Table 1 | Sequences for Sth1Cas9 LMD-9 spacers with NNAGAA PAMs

	ID	bp	5′	Target	PAM	3′
EMX1	**G1**	22	ACCTGG	GCCAGGGAGGGAGGGGCACAGA	TGAGAA	ACTCAG
	G2	19	AGAACC	GGAGGACAAAGTACAAACG	GCAGAA	GCTGGA
	G3	24	GTTCCA	GAACCGGAGGACAAAGTACAAACG	GCAGAA	GCTGGA
	G4	20	CTGGAG	GAGGAAGGGCCTGAGTCCGA	GCAGAA	GAAGAA
	G5	20	GAGGAG	GAAGGGCCTGAGTCCGAGCA	GAAGAA	GAAGGG
	G6	20	GAGGAA	GGGCCTGAGTCCGAGCAGAA	GAAGAA	GGGCTC
	G7	23	GGGCCC	GCCCAGGCAGGCAGGCTCTCCGA	GGAGAA	GGCCAA
FANCF	**G1**	20	AAGCTC	GGAAAAGCGATCCAGGTGCT	GCAGAA	GGGATT
	G2	20	GCTGAC	GTAGGTAGTGCTTGAGACCG	CCAGAA	GCTCGG
RUNX1	**G1**	21	ATTACT	GTACTAATCAGATGGAAGCTCT	TCAGAA	ATGTTT
	G2	22	GTAAAA	GAAATCATTGAGTCCCCCGCCT	TCAGAA	GTGGGT
	G3	20	GTCCCT	GAGGTATCCAGCAGAGGGGA	GAAGAA	AGAGAG
	G4	23	TGGGGA	GTCCCAGAGGTATCCAGCAGAGG	GGAGAA	GAAAGA
ATP1A1	**G1**	20	TCTGTA	GCAGCTTGGATGCTATAAGC	CAAGAA	ACAAAG
Hpd	**G1**	20	CTACTT	GGTACCCCACGCAGAAAGCT	CGAGAA	CGGGGC
	G2	20	TTGTAT	GTTGGGGCCTCGAATCCAGG	TAAGAA	ACGGCC
	G3	20	TATGGA	GATACCACACACACCCTGGT	GGAGAA	GATCAA
	G4	21	CAGTTT	GTAGTAAGAAGATGGGGCGGC	CAAGAA	CTCCGT
	G5	21	GGAGCT	GCATATCCTAGTCGACTATGA	CGAGAA	AGGCTA
Pcsk9	**G1**	19	CCAACA	GGTCACTGCTCATCTTCAC	CAAGAA	GCCAGG
	G2	19	CCAACA	GGGTCACTGCTCATCTTCAC	CAAGAA	GCCAGG
	G3	20	CCCAAC	GAGGTCACTGCTCATCTTCAC	CAAGAA	GCCAGG
	G4	19	AATCAC	GCACGACGCCTCCCGCTCCT	GGAGAA	GCTGGA
	G5	20	CAATCA	GCCACGACGCCTCCCGCTCCT	GGAGAA	GCTGGA
	G6	20	GGCCTG	GAGACCCATGTCCACTGCCA	CCAGAA	GGACCA
Pck1	**G1**	20	GGATAT	GGTGGGAACTCACTACTCGG	GAAGAA	ATGCTT
	G2	20	ATCCTG	GGCATAACTAACCCCGAAGG	CAAGAA	GAAATA
	G3	20	ATAATG	GGGCACTGGCTGGCAGGGGT	GCAGAA	TCTCGA
	G4	20	GCCAGG	GTATTTGCCGAAGTTGTAGCC	GAAGAA	GGGTCG

"G" in position 1 of the guide indicates a mismatch to the genome and it is not counted in the size (bp) of the guide

Supplementary Table 2 | Sequences for Sth1Cas9 LMD-9 spacers with NNGGAA PAMs

	ID	bp	5′	Target	PAM	3′
EMX1	**G8**	22	GACAAA	GTACAAACGGCAGAAGCTGGAG	GAGGAA	GGGCCT
FANCF	**G3**	20	GCGGAA	GTAGGGCCTTCGCGCACCTC	ATGGAA	TCCCTT
	G4	19	GGTAGT	GCTTGAGACCGCCAGAAGC	TCGGAA	AAGCGA
	G5	21	TAGGTA	GTGCTTGAGACCGCCAGAAGC	TCGGAA	AAGCGA
	G6	20	ACCGAG	GGCCTGGAAGTTCGCTAATC	CCGGAA	CTGGAC
VEGFA	**G1**	20	GGTGGG	GAGAGGGACACACAGATCTA	TTGGAA	TCCTGG
	G2	20	GGGCCT	GAGAGCCGTTCCCTCTTTGC	TAGGAA	TATTGA
	G3	20	CCCGCG	GGGCATTGGCGAGGAGGGAG	CAGGAA	AGTGAG
	G4	19	CAGCCT	GAAAATTACCCATCCGCCC	CCGGAA	ACTCTG
	G5	23	TTCACA	GCCTGAAAATTACCCATCCGCCC	CCGGAA	ACTCTG

Supplementary Table 3 | Sequences for Sth1Cas9 LMD-9 spacers with NNN linker

	ID	bp	5′	Target	PAM	3′
EMX1	**G5**	20	GGAGGA	GGAAGGGCCTGAGTCCGAGC	AGAAGAA	GAAGGG
FANCF	**G2**	19	GCTGAC	GTAGGTAGTGCTTGAGACC	GCCAGAA	GCTCGG
	G5	20	TAGGTA	GTGCTTGAGACCGCCAGAAG	CTCGGAA	AAGCGA
ATP1A1	**G1**	19	TCTGTA	GCAGCTTGGATGCTATAAG	CCAAGAA	ACAAAG
RUNX1	**G3**	19	GTCCCT	GAGGTATCCAGCAGAGGGG	AGAAGAA	AGAGAG

Supplementary Table 4 | Sequences for Sth1Cas9 LMD-9/LMG-18311 hybrid spacers with NNGCAA PAMs

	ID	bp	5′	Target	PAM	3′
FANCF	**G202**	20	TGGAGC	GGCGGCTGCACAACCAGTGG	AGGCAA	GAGGGC
RUNX1	**G202**	23	CAAGCT	GCCATTTCATTACAGGCAAAGCT	GAGCAA	AAGTAG
AAVS1	**G201**	20	GCTCGG	GGACACAGGATCCCTGGAGG	CAGCAA	ACATGC

Supplementary Table 5 | Sequences for SaCas9 spacers with NNGRRT PAMs

	ID	bp	5′	Target	PAM	3′
EMX1	**G1**	21	GGGTGG	GCAACCACAAACCCACGAGGG	CAGAGT	GCTGCT
FANCF	**G1**	21	GCGGAA	GTAGGGCCTTCGCGCACCTCA	TGGAAT	CCCTTC
RUNX1	**G1**	23	CAGCAT	GTACTCACCTCTCATGAAGCACT	GTGGGT	ACGAAG
Hgd	**G1**	20	CATCCT	GGAGGTCTATGGTGTCCACT	TTGAGT	TACCTG
Hpd	**G1**	20	AGGTGA	GAGTTTGCTGTGCTGCAGACG	GTGAGT	GAACAC
Gstz1	**G3**	20	TTGATG	GAATCACCATTGTCCAGTCA	GTGAGT	AGAGGA

"G" in position 1 of the guide indicates a mismatch to the genome and it is not counted in the size (bp) of the guide

Supplementary Table 6 | SaCas9 and Sth1Cas9 sgRNAs

tracrRNA	Sequence
Sth1Cas9_v0	GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAACGAAACTTACACAGTTACTTAAATCTTGCA GAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTTT TT
Sth1Cas9_v1	GTCTTTGTACTCTGGTACCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCA TTTTATGGCAGGGTGTTTT
Sth1Cas9_v2	GTCTTTGTACTCTGGTACCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCT ATGGCAGGGTGTTTT
Sth1Cas9_v2	GTCTTTGTACTCTGGAAACAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCA TTTTATGGCAGGGTG TTTT
SaCas9	GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACT TGTTGGCGAGATTTTT

Supplementary Table 7 | Sequences for TBG and LP1B promoters

Promoter	Sequence
TBG	GGGCTGGAAGCTACCTTTGACATCATTTCCTCTGCGAATGCATGTATAATTTCTACAGAACCTAT TAGAAAGGATCACCCAGCCTCTGCTTTTGTACAACTTTCCCTTAAAAAACTGCCAATTCCACTGC TGTTTGGCCCAATAGTGAGAACTTTTTCCTGCTGCCTCTTGGTGCTTTTGCCTATGGCCCCTATT CTGCCTGCTGAAGACACTCTTGCCAGCATGGACTTAAACCCCTCCAGCTCTGACAATCCTCTTTC TCTTTTGTTTTACATGAAGGGTCTGGCAGCCAAAGCAATCACTCAAAGTTCAAACCTTATCATTT TTTGCTTTGTTCCTCTTGGCCTTGGTTTTGTACATCAGCTTTGAAAATACCATCCCAGGGTTAAT GCTGGGGTTAATTTATAACTAAGAGTGCTCTAGTTTTGCAATACAGGACATGCTATAAAAATGGA AAGAT
LP1B	CCCTAAAATGGGCAAACATTGCAAGCAGCAAACAGCAAACACACAGCCCTCCCTGCCTGCTGACC TTGGAGCTGGGGCAGAGGTCAGAGACCTCTCTGGGCCCATGCCACCTCCAACATCCACTCGACCC CTTGGAATTTCGGTGGAGAGGAGCAGAGGTTGTCCTGGCGTGGTTTAGGTAGTGTGAGAGGGTGG ACACAGGACGCTGTGGTTTCTGAGCCAGGGGGCGACTCAGATCCCAGCCAGTGGACTTAGCCCCT GTTTGCTCCTCCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCCTCCCCCGTTGCCCC TCTGGATCCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTCAGGCACCACCA CTGACCTGGGACAGTGAATCCGGACTCTAAGGTAAATATAAAATTTTTAAGTGTATAATGTGTTA AACTACTGATTCTAATTGTTTCTCTCTTTTAGATTCCAACCTTTGGAACTGA

Supplementary Table 8 | AcrIIA Gblocks

Gblock	Sequence
AcrIIA2[8]	ATGACGCTGACCCGCGCTCAGAAAAAATACGCCGAGGCGATGCATGAGTTTATCAATATGGTTGATGACTTT GAAGAATCAACGCCTGACTTTGCAAAAGAGGTTCTGCACGACTCCGACTATGTGGTCATTACAAAAAACGAG AAATATGCCGTGGCACTCTGTAGTCTCTCCACAGATGAATGTGAGTACGATACTAACTTGTATTTGGATGAA AAGCTCGTCGATTACAGCACAGTTGATGTCAACGGAGTGACATATTACATCAATATAGTGGAAACAAATGAC ATAGATGATCTTGAAATTGCGACCGACGAGGACGAGATGAAGTCTGGAAACCAAGAGATTATTCTTAAGTCC GAACTGAAGTGA*TAA*
AcrIIA4[8]	ATGAACATTAACGACCTCATACGAGAGATTAAGAACAAAGATTACACCGTCAAACTGTCAGGAACTGATAGT AACTCAATCACCCAGCTTATTATCAGGGTAAACAATGATGGGAATGAATATGTGATATCTGAGAGCGAAAAC GAGTCTATCGTCGAGAAATTCATTTCCGCTTTTAAGAACGGGTGGAATCAGGAATATGAGGATGAAGAAGAA TTTTACAATGACATGCAGACGATCACGTTGAAAAGTGAACTGAACTGA*TAA*
AcrIIA5 D1126	ATGGCCTATGGTAAAAGTAGGTATAACTCCTATCGCAAACGCAACTTCAGCATCTCTGACAATCAAAGGAGA GAATACGCAAAGAAAATGAAGGAACTTGAACAGGCATTTGAGAACCTCGACGGATGGTACTTGAGTAGCATG AAAGATTCTGCGTATAAGGATTTCGGTAAGTACGAAATCCGACTTTCAAATCACTCAGCCGACAATCGCTAT CATGACCTGGAGAACGGCCGCTTGATCGTGAATGTGAAAGCAAGCAAACTTAACTTTGTCGATATTATCGAG AATAAACTCGGCAAGATCATCGAAAAAATTGATACCCTGGACCTTGATAAATATCGCTTCATCAATGCGACC AAGCTGGAGAGGGACATTAAGTGC TACTATAAAGGGTATAAGACTAAGAAAGACGTTATCTGA*TAA*
AcrIIA5 D4276	ATGGCCTATGGTAAAAGTAGGTATAACTCCTATCGCAAACGCTCTTTCAACAGGTCTAACAAACAAAGGAGA GAATACGCACAGGAAATGGACCGCCTTGAAAAGGCATTTGAGAACCTCGACGGATGGTACTTGAGTAGCATG AAAGATTCTGCGTATAAGGATTTCGGTAAGTACGAAATCCGACTTTCAAATCACTCAGCCGACAATAAATAT CATGACCTGGAGAACGGCCGCTTGATCGTGAATATCAAAGCAAGCAAACTTAACTTTGTCGATATTATCGAG AATAAACTCGACAAGATCATCGAAAAAATTGATAAGCTGGACCTTGATAAATATCGCTTCATCAATGCGACC AACCTGGAGCACGACATTAAGTGC TACTATAAAGGGTTTAAGACTAAGAAAGAGGTTATCTGA*TAA*
AcrIIA6 D1024	ATGAAGATCAACGACGACATAAAGGAATTGATCCTCGAATACATGTCCCGATATTTTAAATTCGAGAATGAC TTCTATAAACTGCCGGGAATCAAATTTACCGATGCTAACTGGCAGAAGTTCAAAAATGGGGGGACTGATATC GAGAAGATGGGGGCGGCCAGGGTGAATGCGATGCTTTCATGTCTCTTCGAAGACTTCGAGTTGGCAATGATT GGAAAAGCCCAAACTAACTATTATATTGACAATTCTCTCAAGCTTAATATGCCATTTTACGCTTATTACGAC ATGTTTAAAAAGCAGTTGCTTATTAACTGGCTGAAAAATAATCGCGACGACGTTATTTGCGGCACGGGTAGG ATGTATACTGCGTCAGGGAACTATATTGCAAACGCGTACCTGGAGGTGGCTCTGGAGTCTAGCCGACTGGGC GGTGGAGAGTATATGCTTCAGATGCGCTTCAAAAATTATAGTCGCAGTCAAGAGCCTATCCCATCAGGACGA CAGAACCGACTTGAATGGATCGAAAACAATCTTGAGAACATCCGGTGA*TAA*
AcrIIA6 D1811	ATGAAGATCAACGACGACATAAAGGAATTGATCCTCGAATACATGTCCCGATATTTTAAATTCGAGAATGAC TTCTATAAACTGCCGGGAATCAAATTTACCGATGCTAACTGGCAGAAGTTCAAAAATGGGGGGACTGATATC GAGAAGATGGGGGCGGCCAGGGTGAATGCGATGCTTGATTGTCTCTTCGATGACTTCGAGTTGGCAATGATT GGAAAAGCCCAAACTAACTATTATAATGACAATTCTCTCAAGATGAATATGCCATTTTACACTTATTACGAC ATGTTTAAAAAGCAGCAACTTCTCAAGTGGCTGAAAAATAATCGCGACGACGTTATTGGCGGCACGGGTAGG ATGTATACTGCGTCAGGGAACTATATTGCAAACGCGTACCTGGAGGTGGCTCTGGAGTCTAGCTCACTGGGC TCAGGAAGCTATATGCTTCAGATGCGCTTCAAAGATTATAGTAAAGGACAAGAGCCTATCCCATCAGGACGA CAGAACCGACTTGAATGGATCGAAAACAATCTTGAGAACATCCGGTGA*TAA*
AcrIIA6 DT1	ATGAAGATCAACAACGACATAAAGGAATTGATCCTCGAATACGTGTCCCGATATTTTAAATTCGAGAATGAC TTCTATAAACTGCCGGGAATCAAATTTACCGATGCTAACTGGCAGAAGTTCAAAAATGGGGAAACTTCCATC GAGAAGATGGGGGCGGCCAGGGTGAATGCGATGCTTTCATGTCTCTTCGAAGACTTCGAGTTGGCAATGATT GGAAAAGCCCAAACTAACTATTATATTGACAATTCTCTCAAGCTTAATATGCCATTTTACGCTTATTACGAC ATGTTTAAAAAGCAGTTGCTTATTAACTGGCTGAAAAATAATCGCGACGACGTTATTGGCGGCACGGGTAGG ATGTATACTGCGTCAGGGAACTATATTGCAAACGCGTACCTGGAGGTGGCTCTGGAGTCTAGCTCACTGGGC TCAGGAAGCTATATGATACAGATGCGCTTCAAAGATTATAGTAAAGGACAAGAGCCTATCCCATCAGGACGA AAGAACCGACTTGAATGGATCGAAAACAATCTTGAGAACATCCGGTGA*TAA*

Supplementary Table 9 | PCR primers used for Surveyor assays and size of expected products.

Target	Primer	Size (bp)
EMX1 Forward	CCATCCCCTTCTGTGAATGT	639
EMX1 Reverse	GGAGATTGGAGACACGGAGA	
FANCF Forward	AGGAACACGGATAAAGACGCTGG	492
FANCF Reverse	AGTTGCTGCACCAGGTGGTAACG	
RUNX1 Forward	CCAGCACAACTTACTCGCACTTGAC	601
RUNX1 Reverse	CATCACCAACCCACAGCCAAGG	
ATP1A1 Forward	GGATTAACATCTGCTCGTGCAGC	419
ATP1A1 Reverse	CACTTGTAAGAGCATCTACAACG	
VEGFA Forward	GAGAAGGCCAGGGGTCACTCCAG	278
VEGFA Reverse	AGCCCGCCGCAATGAAGG	
AAVS1 Forward	TTCGGGTCACCTCTCACTCC	469
AAVS1 Reverse	GGCTCCATCGTAAGCAAACC	
Hpd exon 7 Forward	GCAGGCGCAGTGCCCAAGACAC	498
Hpd exon 7 Reverse	CAGCACATGCCCAGGTCACATGG	
Hpd exon 8 Forward	GCCATGAGGACAGAAAGAGCATC	409
Hpd exon 8 Reverse	GATATTCCAGTCTCCCAGAGAAG	
Hpd exon 12 Forward	CTCGCATACTTGAAGGCTGTGCC	429
Hpd exon 12 Reverse	GATAGGGACTCTGCTACCTCCTG	
Hpd exon 13 Forward	GGCTTTGGTGGTGCAGTAGCCTT	458
Hpd exon 13 Reverse	GACTGATGTGTCCAGCACTTTGC	
Pck1 exon 5 Forward	GATGTGCACCACAAGCTCACTGT	567
Pck1 exon 5 Reverse	GAGTTCGGTGCTGTTGTCTAAGA	
Pck1 exon 6 Forward	GCTAGTTTGGAAGACAGTCCTAG	494
Pck1 exon 6 Reverse	GTCCCTCTCTATCCAGATGATCC	

Pck1 exon 8 Forward	GCAACTTAAGGGCTATCAACCCA	556
Pck1 exon 8 Reverse	GTCTGGATATAGGAGGGAGATCT	
Pck1 exon 10 Forward	GAACACAAGGGTGAGTCACAGTC	565
Pck1 exon 10 Reverse	CATCTGGCTGATTCTCTGTTTCA	
Pcsk9 exon 2 Forward	GATACGCATGCTACACTGAGATG	534
Pcsk9 exon 2 Reverse	CACAGGCAGTAGACAAACCAG	
Pcsk9 exon 9 Forward	GCACCCATGAGACAGGTGAGCAG	559
Pcsk9 exon 9 Reverse	GCAGAGACAATGGGTGGCTAATA	
Pcsk9 exon 10 Forward	GCTGGAAGCTTTATGATGGAGAT	599
Pcsk9 exon 10 Reverse	GACACACCTCAGAGCCTTCCCTT	

NLS-St1Cas9 LMD-9-NLS ORF DNA sequence

ATGGGCGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAGCGAC
CTGGTGCTGGGCCTGGACATCGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTG
ACCGGCGAGATCATCCACAAGAACAGTCGCATCTTCCCTGCTGCTCAGGCTGAGAACAAC
CTGGTGCGCCGCACCAACCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGC
GTGCGCCTGAACCGCCTGTTCGAGGAGAGCGGCCTGATCACCGACTTCACCAAGATCAGC
ATCAACCTGAACCCCTACCAGCTGCGCGTGAAGGGCCTGACCGACGAGCTGAGCAACGAG
GAGCTGTTCATCGCCCTGAAGAACATGGTGAAGCACCGCGGCATCAGCTACCTGGACGAC
GCCAGCGACGACGGCAACAGCAGCGTGGGCGACTACGCCCAGATCGTGAAGGAGAACAGC
AAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGGAGCGCTACCAGACCTACGGC
CAGCTGCGCGGCGACTTCACCGTGGAGAAGGACGGCAAGAAGCACCGCCTGATCAACGTG
TTCCCCACCAGCGCCTACCGCAGCGAGGCCCTGCGCATCCTGCAGACCCAGCAGGAGTTC
AACCCCCAGATCACCGACGAGTTCATCAACCGCTACCTGGAGATCCTGACCGGCAAGCGC
AAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCACCGACTACGGCCGCTACCGCACC
AGCGGCGAGACCCTGGACAACATCTTCGGCATCCTGATCGGCAAGTGCACCTTCTACCCC
GACGAGTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGGAGTTCAATCTGCTGAACGAC
CTGAACAACCTGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAG
ATCATCAACTACGTGAAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATC
GCCAAGCTGCTGAGCTGCGACGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGC
AAGGCCGAGATCCACACCTTCGAGGCCTACCGCAAGATGAAGACCCTGGAGACCCTGGAC
ATCGAGCAGATGGACCGAGAGACCCTGGACAAGCTGGCCTACGTGCTGACCCTGAACACC
GAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGAGTTCGCCGACGGCAGCTTCAGCCAG
AAACAGGTGGACGAGCTGGTGCAGTTCCGCAAGGCCAACAGCAGCATCTTCGGCAAGGGC
TGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCCCCGAGCTGTACGAGACCAGC
GAGGAGCAGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAGCAGCAGCAAC
AAGACCAAGTACATCGACGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGGTGGCC
AAGAGCGTGCGCCAGGCCATCAAGATCGTGAACGCCGCCATCAAGGAGTACGGCGACTTC
GACAACATCGTGATCGAGATGGCCCGCGAGACCAACGAGGACGACGAGAAGAAGGCCATC
CAGAAGATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCTGAAGGCCGCCAAC
CAGTACAACGGCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTGGCC
ACCAAGATCCGCCTGTGGCACCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATC
AGCATCCACGACCTGATCAACAACAGCAACCAGTTCGAGGTGGACCACATCCTGCCCCTG
AGCATCACCTTCGACGACAGCCTGGCCAACAAGGTGCTGGTGTACGCCACCGCCAACCAG
GAGAAGGGCCAGCGCACCCCCTACCAGGCCCTGGACAGCATGGACGACGCCTGGAGCTTC
CGCGAGCTGAAGGCCTTCGTGCGCGAGAGCAAGACCCTGAGCAACAAGAAGAAGGAGTAT
CTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCAAGAAGTTCATCGAGCGCAAC
CTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTGCAGGAGCACTTCCGC
GCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAGCCAGCTGCGC
CGCCACTGGGGCATCGAGAAGACCCGCGACACCTACCACCACCACGCCGTGGACGCCCTG
ATCATTGCGGCTTCTAGCCAGCTGAACCTGTGGAAGAAGCAGAAGAACACCCTGGTGAGC
TACAGCGAGGACCAGCTGCTGGACATCGAGACCGGCGAGCTGATCAGCGACGACGAGTAC
AAGGAGAGCGTGTTCAAGGCCCCCTACCAGCACTTCGTGGACACCCTGAAGAGCAAGGAG
TTCGAGGACAGCATCCTGTTCAGCTACCAGGTGGACAGCAAGTTCAACCGCAAGATCAGC
GACGCCACCATCTACGCCACCCGCCAGGCCAAGGTGGGCAAGGACAAGGCCGACGAGACC
TACGTGCTGGGCAAGATCAAGGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAG
ATCTACAAGAAGGACAAGAGCAAGTTCCTGATGTACCGCCACGACCCCCAGACCTTCGAG
AAGGTGATCGAGCCCATCCTGGAGAACTACCCCAACAAGCAGATCAACGAGAAAGGCAAG
GAGGTGCCCTGCAACCCCTTCCTGAAGTACAAGGAGGAGCACGGCTACATCCGCAAGTAC
AGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGAAGTACTACGACAGCAAGCTGGGC
AACCACATCGACATCACCCCCAAGGACAGCAACAACAAGGTGGTGCTGCAGAGCGTGAGC
CCCTGGCGCGCCGACGTGTACTTCAACAAGACCACCGGCAAGTACGAGATCCTGGGGCTG
AAGTACGCCGATCTGCAGTTTGAGAAAGGCACAGGCACCTACAAGATCAGCCAGGAGAAG

TACAACGACATCAAGAAGAAGGAGGGCGTGGACAGCGACAGCGAGTTCAAGTTCACCCTG
TACAAGAACGACCTTCTGCTGGTGAAGGACACCGAGACCAAGGAGCAACAGCTGTTCCGC
TTCCTGAGCCGCACCATGCCCAAGCAGAAGCACTACGTGGAGCTGAAGCCCTACGACAAG
CAGAAGTTCGAGGGCGGCGAGGCCCTGATCAAGGTGCTGGGCAACGTGGCCAACAGCGGC
CAGTGCAAGAAGGGCCTGGGCAAGAGCAACATCAGCATCTACAAGGTGCGCACCGACGTG
CTGGGCAACCAGCACATCATCAAGAACGAGGGCGACAAGCCCAAGTTGGACTTCAGCAGG
GCTGACCCCAAGAAGAAGAGGAAGGTGTGA

NLS-St1Cas9 LMD-9-NLS AA sequence

MGAPKKKRKVGIHGVPAASDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRT
NRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMV
KHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKH
RLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRT
SGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYV
KNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLD
KLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPE
LYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDF
DNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRL
WHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQA
LDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNAL
QEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVS
YSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIY
ATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENY
PNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVV
LQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTL
YKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKG
LGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDFSRADPKKKRKV*

NLS-St1Cas9 LMD-9/LMG18311 Hybrid-NLS ORF DNA sequence

ATGGGCGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAGCGAC
CTGGTGCTGGGCCTGGACATCGGCATCGGCAGCGTGGGCGTGGGCATCCTGAACAAGGTG
ACCGGCGAGATCATCCACAAGAACAGTCGCATCTTCCCTGCTGCTCAGGCTGAGAACAAC
CTGGTGCGCCGCACCAACCGCCAGGGTCGCCGGCTTGCTCGCCGCAAGAAGCACCGGCGC
GTGCGCCTGAACCGCCTGTTCGAGGAGAGCGGCCTGATCACCGACTTCACCAAGATCAGC
ATCAACCTGAACCCCTACCAGCTGCGCGTGAAGGGCCTGACCGACGAGCTGAGCAACGAG
GAGCTGTTCATCGCCCTGAAGAACATGGTGAAGCACCGCGGCATCAGCTACCTGGACGAC
GCCAGCGACGACGGCAACAGCAGCGTGGGCGACTACGCCCAGATCGTGAAGGAGAACAGC
AAGCAGCTGGAGACCAAGACCCCCGGCCAGATCCAGCTGGAGCGCTACCAGACCTACGGC
CAGCTGCGCGGCGACTTCACCGTGGAGAAGGACGGCAAGAAGCACCGCCTGATCAACGTG
TTCCCCACCAGCGCCTACCGCAGCGAGGCCCTGCGCATCCTGCAGACCCAGCAGGAGTTC
AACCCCCAGATCACCGACGAGTTCATCAACCGCTACCTGGAGATCCTGACCGGCAAGCGC
AAGTACTACCACGGCCCCGGCAACGAGAAGAGCCGCACCGACTACGGCCGCTACCGCACC
AGCGGCGAGACCCTGGACAACATCTTCGGCATCCTGATCGGCAAGTGCACCTTCTACCCC
GACGAGTTCCGCGCCGCCAAGGCCAGCTACACCGCCCAGGAGTTCAATCTGCTGAACGAC
CTGAACAACCTGACCGTGCCCACCGAGACCAAGAAGCTGAGCAAGGAGCAGAAGAACCAG
ATCATCAACTACGTGAAGAACGAGAAGGCTATGGGCCCCGCCAAGCTGTTCAAGTACATC
GCCAAGCTGCTGAGCTGCGACGTGGCCGACATCAAGGGCTACCGCATCGACAAGAGCGGC
AAGGCCGAGATCCACACCTTCGAGGCCTACCGCAAGATGAAGACCCTGGAGACCCTGGAC
ATCGAGCAGATGGACCGAGAGACCCTGGACAAGCTGGCCTACGTGCTGACCCTGAACACC
GAGCGCGAGGGCATCCAGGAGGCCCTGGAGCACGAGTTCGCCGACGGCAGCTTCAGCCAG

AAACAGGTGGACGAGCTGGTGCAGTTCCGCAAGGCCAACAGCAGCATCTTCGGCAAGGGC
TGGCACAACTTCAGCGTGAAGCTGATGATGGAGCTGATCCCCGAGCTGTACGAGACCAGC
GAGGAGCAGATGACCATCCTGACCCGCCTGGGCAAGCAGAAGACCACCAGCAGCAGCAAC
AAGACCAAGTACATCGACGAGAAGCTGCTGACCGAGGAGATCTACAACCCCGTGGTGGCC
AAGAGCGTGCGCCAGGCCATCAAGATCGTGAACGCCGCCATCAAGGAGTACGGCGACTTC
GACAACATCGTGATCGAGATGGCCCGCGAGACCAACGAGGACGACGAGAAGAAGGCCATC
CAGAAGATCCAGAAGGCCAACAAGGACGAGAAGGACGCCGCCATGCTGAAGGCCGCCAAC
CAGTACAACGGCAAGGCCGAGCTGCCCCACAGCGTGTTCCACGGCCACAAGCAGCTGGCC
ACCAAGATCCGCCTGTGGCACCAGCAGGGCGAGCGCTGCCTGTACACCGGCAAGACCATC
AGCATCCACGACCTGATCAACAACAGCAACCAGTTCGAGGTGGACCACATCCTGCCCCTG
AGCATCACCTTCGACGACAGCCTGGCCAACAAGGTGCTGGTGTACGCCACCGCCAACCAG
GAGAAGGGCCAGCGCACCCCCTACCAGGCCCTGGACAGCATGGACGACGCCTGGAGCTTC
CGCGAGCTGAAGGCCTTCGTGCGCGAGAGCAAGACCCTGAGCAACAAGAAGAAGGAGTAT
CTGCTGACCGAGGAGGACATCAGCAAGTTCGACGTGCGCAAGAAGTTCATCGAGCGCAAC
CTGGTGGACACCCGCTACGCCAGCCGCGTGGTGCTGAACGCCCTGCAGGAGCACTTCCGC
GCCCACAAGATCGACACCAAGGTGAGCGTGGTGCGCGGCCAGTTCACCAGCCAGCTGCGC
CGCCACTGGGGCATCGAGAAGACCCGCGACACCTACCACCACCACGCCGTGGACGCCCTG
ATCATTGCGGCTTCTAGCCAGCTGAACCTGTGGAAGAAGCAGAAGAACACCCTGGTGAGC
TACAGCGAGGAgCAGCTGCTGGACATCGAGACCGGCGAGCTGATCAGCGACGACGAGTAC
AAGGAGAGCGTGTTCAAGGCCCCCTACCAGCACTTCGTGGACACCCTGAAGAGCAAGGAG
TTCGAGGACAGCATCCTGTTCAGCTACCAGGTGGACAGCAAGTTCAACCGCAAGATCAGC
GACGCCACCATCTACGCCACCCGCCAGGCCAAGGTGGGCAAGGACAAGaagGACGAGACC
TACGTGCTGGGCAAGATCAAGGACATCTACACCCAGGACGGCTACGACGCCTTCATGAAG
ATCTACAAGAAGGACAAGAGCAAGTTCCTGATGTACCGCCACGACCCCCAGACCTTCGAG
AAGGTGATCGAGCCCATCCTGGAGAACTACCCCAACAAGCAGATgAACGAGAAAGGCAAG
GAGGTGCCCTGCAACCCCTTCCTGAAGTACAAGGAGGAGCACGGCTACATCCGCAAGTAC
AGCAAGAAGGGCAACGGCCCCGAGATCAAGAGCCTGAAGTACTACGACAGCAAGCTGctg
GGCAACCcCATCGACATCACCCCCgAaaACAGCAAgAACAAGGTGGTGCTGCAGAGCcTt
AagCCCTGGCGCaCCGACGTGTACTTCAACAAGaaCACCGGCAAGTACGAGATCCTGGGG
CTGAAGTACGCCGATCTGCAGTTTGAGAAAaagACAGGCACCTACAAGATCAGCCAGGAG
AAGTACAACGgCATCAtGAAGgAGGAGGGCGTGGACAGCGACAGCGAGTTCAAGTTCACC
CTGTACAAGAACGACCTTCTGCTGGTGAAGGACACCGAGACCAAGGAGCAACAGCTGTTC
CGCTTCCTGAGCCGCACCATGCCCAAcgtGAAGtACTACGTGGAGCTGAAGCCCTACagC
AAGgAcAAGTTCGAGaagaaCGAGagCCTGATCgAGaTcCTGGGCAgCGccGaCAAgAGC
GGCagGTGCAtcAAGGGCCTGGGCAAGAGCAACATCAGCATCTACAAGGTGCGCACCGAC
GTGCTGGGCAACCAGCACATCATCAAGAACGAGGGCGACAAGCCCAAGTTGGACTTCAGC
AGGGCTGACCCCAAGAAGAAGAGGAAGGTGGGATCCTGA

NLS-St1Cas9 LMD-9/LMG18311 Hybrid-NLS AA sequence

MGAPKKKRKVGIHGVPAASDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRT
NRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMV
KHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKH
RLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRT
SGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYV
KNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLD
KLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPE
LYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDF
DNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRL
WHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQA
LDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNAL
QEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVS

YSEEQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIY
ATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENY
PNKQMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKV
VLQSLKPWRTDVYFNKNTGKYEILGLKYADLQFEKKTGTYKISQEKYNGIMKEEGVDSDSEFKFT
LYKNDLLLVKDTETKEQQLFRFLSRTMPNVKYYVELKPYSKDKFEKNESLIEILGSADKSGRCIK
GLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDFSRADPKKKRKVGS*

SUPPLEMENTARY REFERENCES

1. Esvelt, K.M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10**, 1116-1121 (2013).

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

3. Muller, M. et al. Streptococcus thermophilus CRISPR-Cas9 Systems Enable Specific Editing of the Human Genome. *Mol Ther* **24**, 636-644 (2016).

4. Ran, F.A. et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191 (2015).

5. Nishimasu, H. et al. Crystal Structure of Staphylococcus aureus Cas9. *Cell* **162**, 1113-1126 (2015).

6. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

7. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324 (2014).

8. Rauch, B.J. et al. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell* **168**, 150-158 e110 (2017).