bioRχiv

ARTICLE PREPRINT

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

Sophie Carter[1]*, Daniel Agudelo[1]*, Alexis Duringer[1], Jeremy Loehr[1], and Yannick Doyon[1,2,3].

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). We define functional protospacer adjacent motif (PAM) sequences and variables required for robust and efficient editing *in vitro*. Expression of holo-St1Cas9 from a single adeno-associated viral (rAAV) vector in the neonatal liver rescued lethality and metabolic defects in a mouse model of hereditary tyrosinemia type I demonstrating effective cleavage activity *in vivo*. Furthermore, we identified potent anti-CRISPR proteins to regulate the activity of both St1Cas9 and the related type II-A *Staphylococcus aureus* Cas9 (SaCas9). This work expands the targeting range and versatility of CRISPR-associated enzymes and should encourage studies to determine its structure, genome-wide specificity profile and sgRNA design rules.

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 subsequent base pairing of the guide RNA with the DNA to generate a specific double-strand break (DSB)[1, 5]. While *Streptococcus pyogenes* (SpCas9) remains the most widely used Cas9 variant 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*[7-10]. Nevertheless, the full potential of selected enzymes can be unleashed using machine learning to establish sgRNA design rules[11, 12].

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 recombinant adeno-associated virus (rAAV) vectors[7, 13, 14]. More recently, *Campylobacter jejuni* and *Neisseria meningitidis* Cas9s from the type II-C[15] CRISPR-Cas systems have been added to this repertoire[16, 17].

In vivo genome 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. As such, we have previously shown that rAAV- and zinc-finger nuclease (ZFN)-mediated liver targeting can correct disease phenotypes in neonatal and adult mouse models, a process currently under clinical investigation[18-21]. 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[22, 23].

Here we revisited the properties of *Streptococcus thermophilus* type II-A CRISPR1-Cas9 system, a model nuclease of paramount importance to the entire CRISPR field, and engineered a potent RNA-guided nuclease for both *in vitro* and *in vivo* applications. *S. thermophilus* encodes up to two active type II-A systems (CRISPR1 and CRISPR3) that epitomize the instrumental role played by researchers focusing on the biology of CRISPR-Cas systems in the development of genome-engineering tools[5, 24]. In particular, study of the CRISPR1-Cas system has led to the following seminal discoveries; (i) spacer sequences are derived from phage genomes and plasmids[25], (ii)

[1]Centre Hospitalier Universitaire de Québec Research Center - Université Laval, Québec, QC, Canada. [2]Université Laval Cancer Research Centre, Québec, QC, Canada. [3]Address correspondence to: Yannick.Doyon@crchudequebec.ulaval.ca

CRISPR-Cas constitutes a bacterial adaptive immune system[26], (iii) Cas9 is an endonuclease that cleaves DNA precisely 3 nucleotides upstream of the protospacer adjacent motif (PAM)[27-29], (iv) sgRNA structural motifs govern function and orthogonality[30], and (v) anti-CRISPR proteins (Acrs) block Cas9 activity[31]. The distinctive functional PAM sequences (NNAGAA and NNGGAA) of St1Cas9 increase the targeting flexibility and combinatorial potential of CRISPR-based genome editing tools. As shown for SaCas9, St1Cas9 can be efficiently packaged into an all-in-one rAAV for *in vivo* delivery to the liver. Moreover, we identified a family of anti-CRISPR proteins acting as off-switches for these two Cas9s. Currently, it is the only Cas9 functional in genome editing applications to be isolated from a nonpathogenic bacterium[32].

RESULTS

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

While characterizing the interplay between St1Cas9 and diverse Acr families isolated from virulent phages infecting *S. thermophilus*[31], we were surprised by the substantial levels of editing achieved in human cells[33]. This observation contrasts with early reports indicating that this ortholog was mildly active[7, 34]. This 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[35] and established a K562 cell line stably expressing St1Cas9 (*S. thermophilus* strain LMD-9) from the *AAVS1* safe harbor locus[36, 37] (**Fig. 1a** and **Supplementary Fig. 1**). St1Cas9 (1,121 aa) shares 17% and 37% identity with SpCas9 (1,368 aa) and SaCas9 (1,053 aa), respectively. Second, we adapted an sgRNA sequence used to monitor St1Cas9 activity in the heterologous host *Escherichia coli*[30]. We substituted a wobble base pair present in the lower stem of the repeat:anti-repeat region for a canonical Watson-Crick base pair in order to interrupt the RNA polymerase III termination signal (**Fig. 1b**). Then, we compared this sgRNA architecture (v1) to its counterpart containing a wild-type full length crRNA:tracrRNA duplex connected via a tetraloop (v0) by targeting *FANCF*, *EMX1*, and *RUNX1* genes[35] (**Fig. 1c** and **Supplementary Fig. 1**). St1Cas9-expressing human cells were transfected with increasing amounts of each sgRNA construct and the Surveyor nuclease assay was used to determine the frequency of indels characteristic of imprecise DSB repair by NHEJ[36, 38] (**Fig. 1d**). The spectrum and frequency of targeted mutations was also analyzed using the complementary TIDE (Tracking of Indels by DEcomposition) method[39] (**Fig. 1d** and **Supplementary data set 1**). Irrespective of the quantification method, the potency of sgRNA v1 was markedly superior. The increased 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.

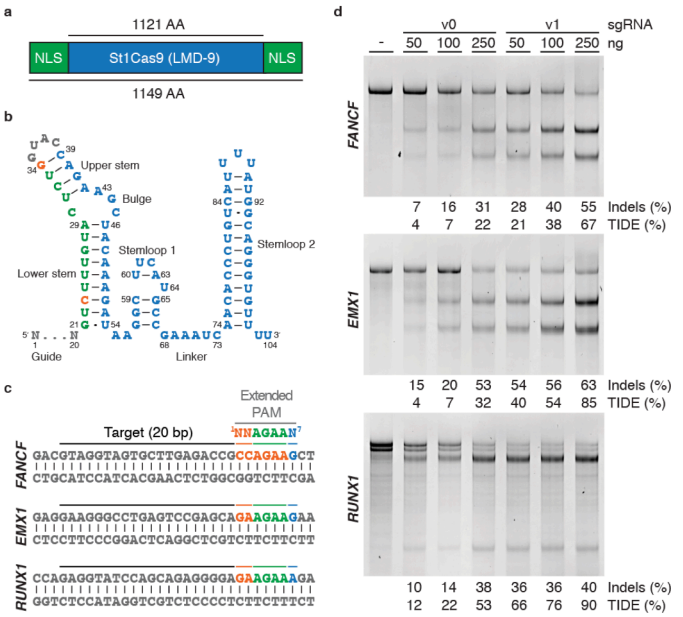

Figure 1 | Engineered CRISPR1-StCas9 system drives robust gene editing in human cells. (**a**) Schematic representation of St1Cas9 from the LMD-9 strain flanked by nuclear localization signals (NLS). (**b**) Nucleotide sequence, predicted secondary structure, and functional modules of St1Cas9 sgRNA (v1); 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.

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

To further support these observations, we used CRISPOR[40] to design several sgRNAs against *Pck1*, *Pcsk9,* and *Hpd*, 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. 2a-c**). Of note, this screen identified highly active sgRNAs targeting in the vicinity of mutations found in human *HPD*[41, 42]. Deficiency of 4-hydroxyphenyl-pyruvate dioxygenase (HPD), the second enzyme in the tyrosine catabolic pathway, causes Tyrosinemia type III (Orphanet ORPHA:69723) (**Fig. 3a**). Only three missense mutations are known to cause this rare disease (Prevalence <1/1,000,000) and we could target two of them with high

efficacy (OMIM 276710) (**Fig. 3c** and **Supplementary Fig. 2**). Targeting the third mutation was not attempted due to the low specificity score of the guide. 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.

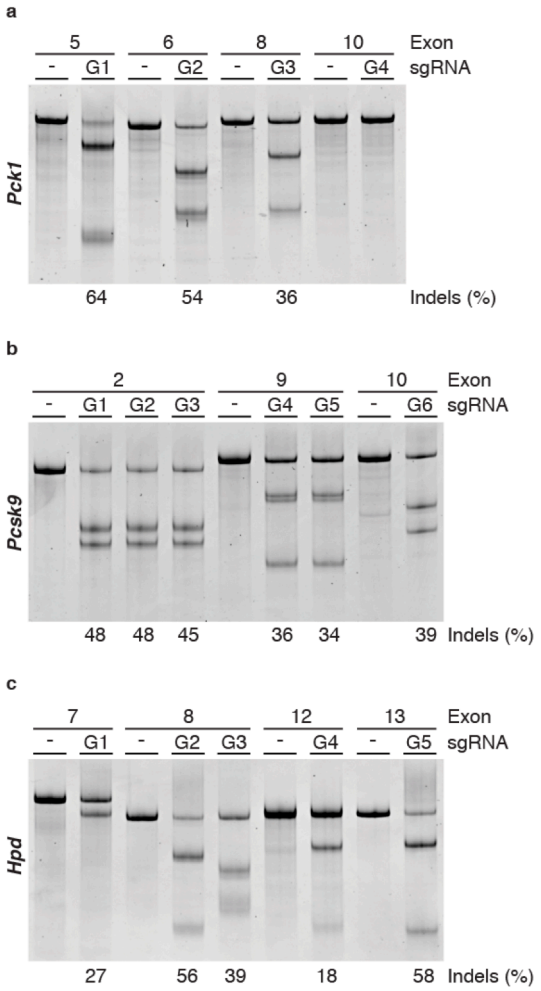

Figure 2 | Screening for active sgRNAs targeting genes affecting liver function in mouse cells. (a) Surveyor assays to determine St1Cas9 activity programmed with various sgRNAs targeting *Pck1*. Neuro-2a cells were transiently transfected with a single vector (0.5µg) driving the expression of St1Cas9 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 *Pcsk9*. (c) Same as in (a) but targeting *Hpd*.

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 as demonstrated in clinical trials for hemophilia B and inherited retinal dystrophy[43-45]. 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[46-48]. 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[19, 49]. 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[19, 49].

To deliver holo-St1Cas9 (St1Cas9 + sgRNA) to the liver, we generated a hepatotropic rAAV serotype 8[13, 18-20] vector targeting *Hpd* exon 13 (aka AAV8-St1Cas9 *Hpd* G5) (**Fig. 3c**). To test the cleavage activity of St1Cas9 *in vivo*, we injected mice at day 2 of life into the retro-orbital sinus with increasing amounts of vector and isolated total liver DNA at day 28 post injection (**Fig. 3b**). The titration showed that the degree of target editing was substantial and dependent on the dose of AAV8-St1Cas9 (**Fig. 3d**).

To test if AAV8-St1Cas9 can lead to 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. 3a**). Of particular relevance to us, the incidence of HT-I reaches 1/1846 in a region of the province of Québec (Canada) while it is around 1/100,000 births worldwide[50]. *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 *Hpd* upstream in the pathway (**Fig. 3a**)[51]. Likewise, genetic ablation of *Hpd* in mice prevents liver damages and lethality leading to a much milder HT-III phenotype[52, 53]. *Fah*$^{-/-}$ mutant pups maintained on NTBC were injected at day 2 of life with AAV8-St1Cas9 *Hpd* G5 and then the drug was withdrawn shortly after weaning (**Fig. 3b**). 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 criteria (**Fig. 3e,f**). Likewise, glycemia was normalized in the treatment groups (**Fig. 3g**). 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[51] (**Fig. 3g**). 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 (**Fig. 3h**). These observations were recapitulated when targeting *Hpd* exon 8 at a site corresponding to a mutation also found in human patients (**Supplementary Fig. 2**). Therefore, rAAV-mediated delivery of St1Cas9 *in vivo* can correct a phenotype in neonatal mice by rewiring a metabolic pathway.

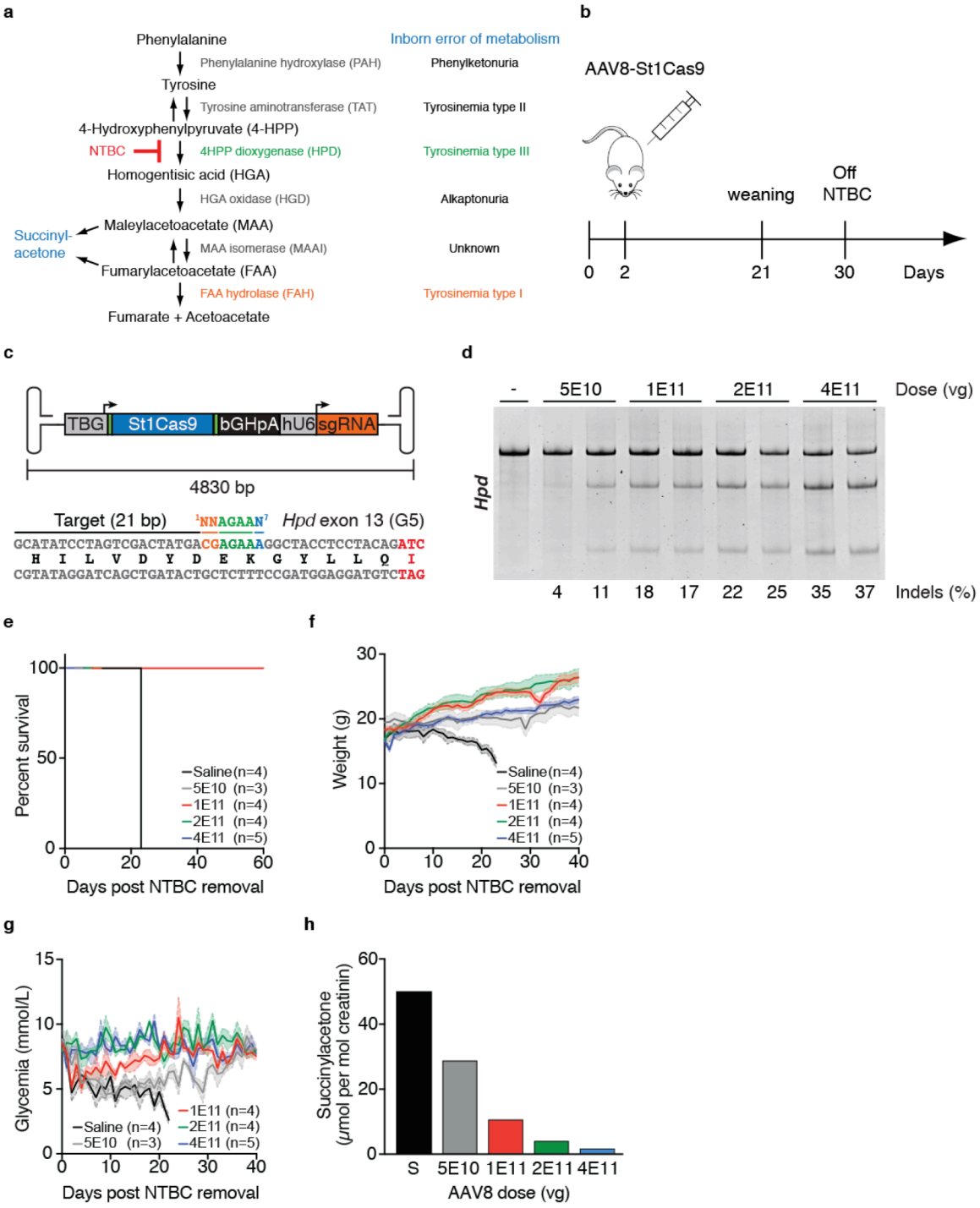

Figure 3 | *In vivo* metabolic pathway rewiring via rAAV8-mediated delivery of holo-St1Cas9. (**a**) The tyrosine degradation pathway and associated genetic disorders. (**b**) Experimental design for *in vivo* editing. Neonatal (2 days old) *Fah*$^{-/-}$ mice were injected with rAAV8-St1Cas9 or saline 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.

Lastly, we evaluated two additional vector architectures in order to minimize the size of rAAV and test the impact of the promoter on overall activity (**Fig. 4**). An rAAV vector (v3) containing an engineered liver-specific promoter (LP1b) combining the human apolipoprotein E/C-I gene locus control region (*ApoE*-HCR) and a modified human α1 antitrypsin promoter (h*AAT*) coupled to an SV40 intron and a synthetic polyadenylation element improved (>2 fold) cleavage efficacy at low rAAV dose as compared to the TBG promoter (**Fig. 4a,b**)[54, 55]. These modifications also led to the creation of a vector of ~4.7 kb in size which is optimal for viral particle packaging[13]. Collectively, these data support the notion that St1Cas9 is a powerful tool for *in vivo* genome editing.

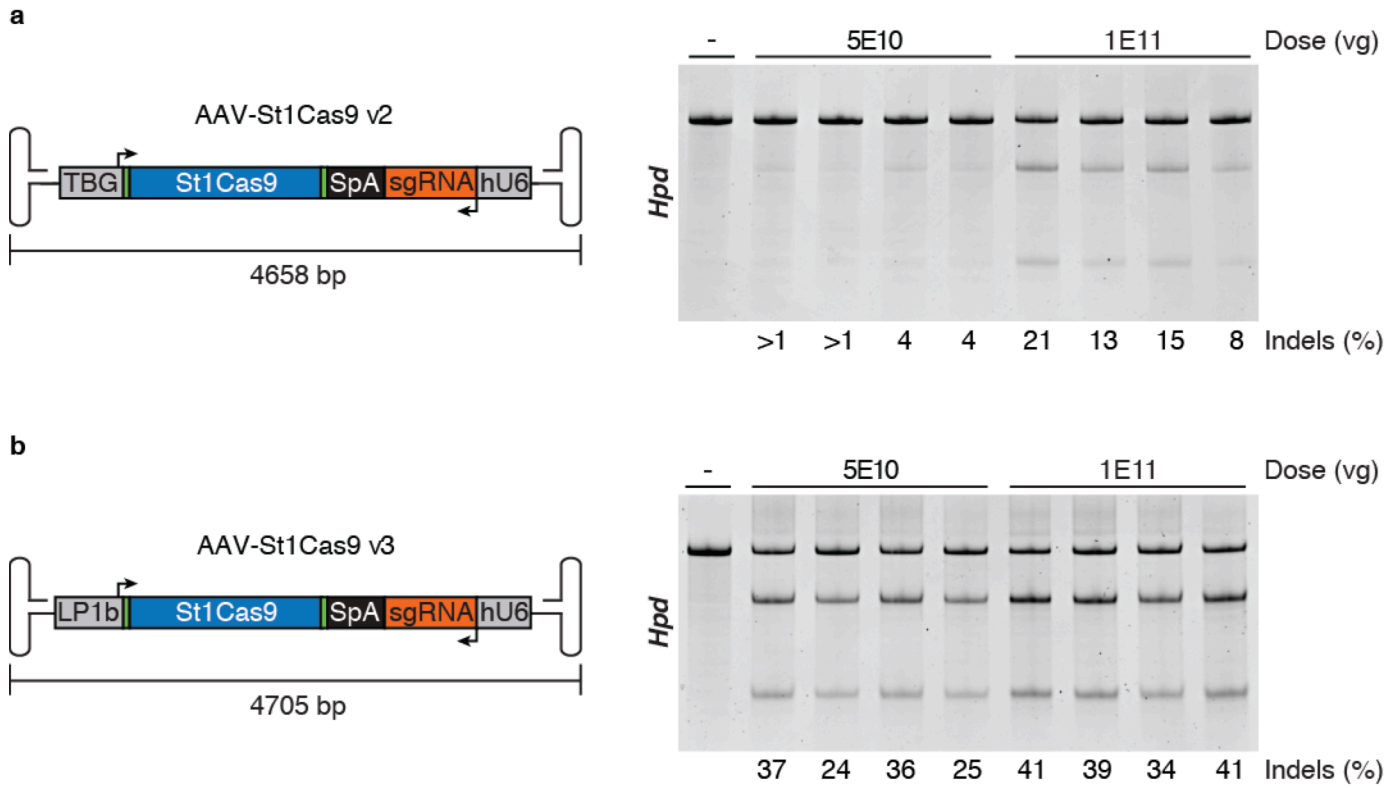

Figure 4 | Alternative rAAV-St1Cas9 vector architectures can further improve potency. (a) Schematic representations of the second-generation rAAV-St1Cas9 (v2) vector of similar size to the parent AAV genome (~4.7kb). Annotated is the human thyroxine binding globulin (TBG) promoter, synthetic polyadenylation sequence (SpA) and hU6 promoter. Arrows indicate the direction of transcriptional unit. 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. (b) Same as in (a) but the TBG promoter was swapped for the composite liver-specific LP1b promoter to generate rAAV8-St1Cas9 v3.

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[56]. In theory, Acrs could be used as safety off-switches to block Cas9 activity *in vivo* and prevent overt off-target activity[57]. To begin to explore this avenue, we screened type II-A CRISPR-Cas9 inhibitor proteins (AcrIIAs) against SaCas9 and St1Cas9[31, 33, 58]. Interestingly, the AcrIIA5 -family could inhibit both Cas9s in cultured cells (**Fig. 5**). AcrIIA4, an SpCas9 inhibitor[57, 58], did not act on SaCas9 and St1Cas9 (**Fig. 5**). 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.

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 and rescue 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[7-10]. Some enzymes simply fail to work and some choose their substrates promiscuously, necessitating thorough biochemical characterization[59-65]. 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, 11, 66]. In that regard, St1Cas9 and SpCas9 also function orthogonally[67]. Moreover, sgRNAs for St1Cas9 and SaCas9 are not functionally interchangeable, which is likely due to their unique PAM specificity (**Supplementary Fig. 3**). These orthologous Cas9 systems could be adapted and combined for targeted knockout, transcriptional control, base editing, as well as "epigenome" editing[1].

Further improvements in rAAV vector design; such as the use of alternative promoters, introns, and codon optimization schemes may yield to even higher activity *in vivo*[13]. The current size of the rAAV vector (v3) (**Fig. 4b**) is ~4.7kb and, if needed, there is sufficient room to incorporate target sequences for the hematopoietic-specific microRNA miR-142 into the St1Cas9 expression cassette to prevent its expression in antigen-presenting cells (APCs) and induce robust tolerance to Cas9[13, 68-72].

Cas9 orthologs used for rAAV-mediated *in vivo* genome editing require a more complex PAM than the relatively simple NGG of SpCas9[1]. This restricts the range of accessible targets but may reduce the occurrence of off-target mutagenesis[73]. The consensus PAM for St1Cas9 (LMD-9 and DGCC7710 *S. thermophilus* strains that differ by only 2 aa) has been defined as $N^1N^2A^3G^4A^5A^6(W^7)$, however sequences closely related to the consensus can be functional in test tubes and in bacterial cells[27, 35, 74-77]. While recognition of an A-rich PAM may ease targeting A/T-rich regions of genomes, we found that St1Cas9 can be targeted to both NNAGAA and NNGGAA PAMs in mammalian cells (**Supplementary Fig. 4**). Of note, in our limited dataset, the presence of an A at position 7 of an extended PAM correlates with high activity but is neither necessary nor sufficient (**Supplementary Fig. 4** and **Supplementary Tables 1-2**). The length of the nonconserved linker (N^1N^2) has also been shown to be flexible and an extension from 2 to 3 bases is tolerated[30, 78]. However, we failed to reproduce this observation in human cells suggesting a higher stringency of the system in this context (**Supplementary Fig. 4** and **Supplementary Table 3**). Structure-guided rational design and selection-based approaches could also enable St1Cas9 to recognize a broad range of PAM sequences[35, 79, 80]. Interestingly, making use of the natural diversity within *S. thermophilus*

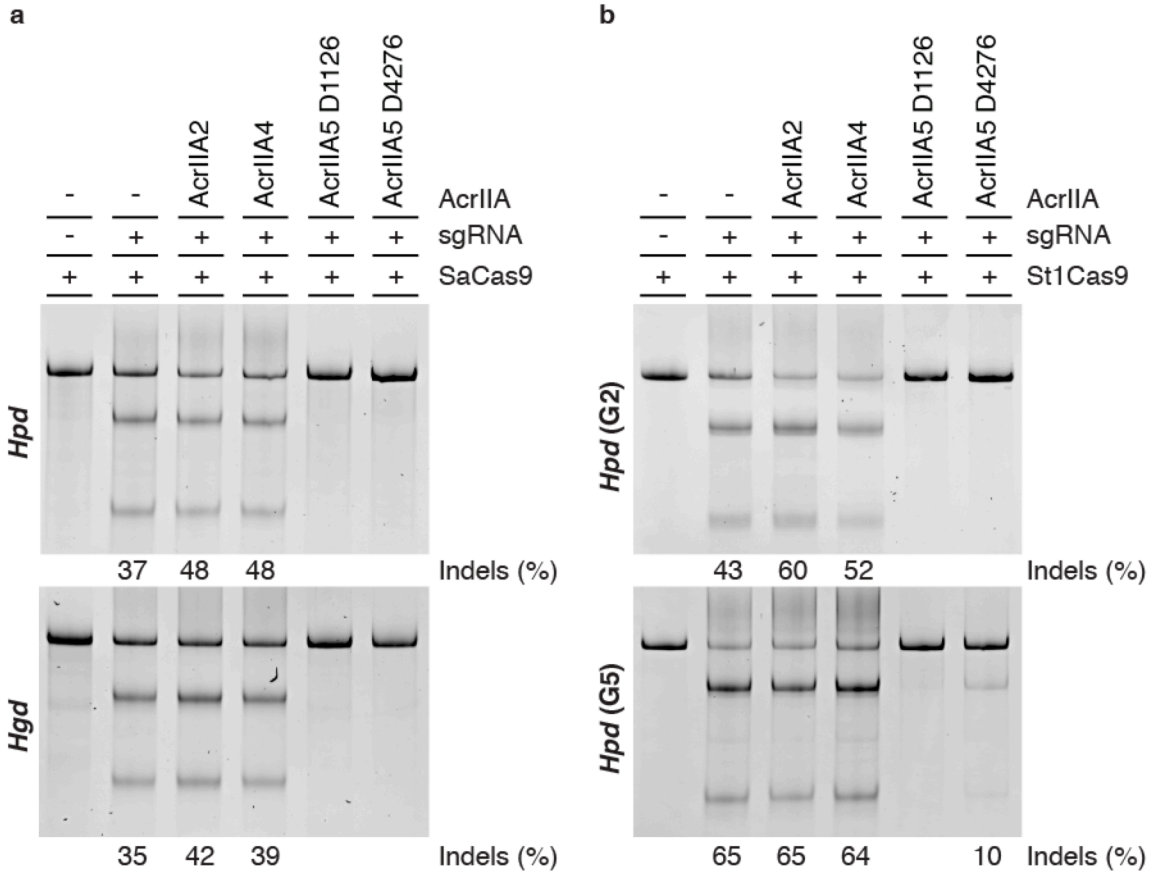

Figure 5 | Anti-CRISPR activity of AcrIIAs against SaCas9 and St1Cas9. (a) Neuro-2a cells were transfected with a dual SaCas9 and sgRNA expression plasmid (0.5µg) targeting *Hpd* or *Hgd*. Where indicated, a second vector (0.5µg) was used to co-express human codon-optimized AcrIIAs. The Surveyor assay was performed 3 days later to determine the frequency of indels, as indicated at the base of each lane. (b) Same as in (a) but using St1Cas9 to target *Hpd* (G2) and *Hpd* (G5).

strains may achieve similar outcomes. For example, inferred consensus PAM sequences for St1Cas9 from strains CNRZ1066 and LMG13811 are NNACAA(W) and NNGCAA(A), respectively[25, 32]. Notably, 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[78]. At the protein level, the sequence of those three St1Cas9 variants diverges mostly within the C-terminal PAM-interacting (PI) domain, implying that they have evolved to recognize slightly distinct PAM sequences (**Supplementary Figs. 5-6**). Database mining predicts that even more diversity exists within CRISPR1-StCas9 systems[81]. Since the crRNA and tracrRNA are completely conserved[30], we envisioned that a suite of tools based on naturally occurring St1Cas9 variants could be used in order to expand the targeting range of this genome editing system.

ACKNOWLEDGMENTS

This study was supported by grants from the Canadian Institutes of Health Research (CIHR) and the Banting Research Foundation to Y.D. Salary support was provided by the Fonds de la recherche du Québec-Santé (FRQS) to 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, as well as Paula Waters and Denis Cyr for their high level analytical expertise on the quantitation of metabolites. 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., and Y.D.; Investigation, S.C., D.A., A.D., J.L.; Writing – Original Draft, Y.D.; Writing – Review and Editing, S.C., D.A., A.D., and Y.D.; Supervision, Y.D.; Funding Acquisition, Y.D.

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[36, 37]. Briefly, simultaneous selection and cloning was performed for 10 days in methylcellulose-based semi-solid RPMI medium supplemented with 0.5 μg/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.

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. 7**). The CRISPOR[40] web tool was used to design guide (spacer) sequences against mouse and human targets (**Supplementary Tables 1-4)**. A fraction of the guides targeting *EMX1*, *FANCF*, and *VEGFA* have been described previously[35] (**Supplementary Tables 1,4)**. 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[35] (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[7] (**Supplementary Table 5**) were synthesized as a gBlock gene fragments (Integrated DNA Technologies) and cloned into pUC19. BPK2301[35] (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 6**) were assembled from the above-described components into a derivative of pX602[7] (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[54, 55]. 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[37] (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 (Integrated DNA Technologies) (**Supplementary Table 7**).

Surveyor nuclease and TIDE assays

Genomic DNA from 2.5E5 cells was extracted with 250 μl of QuickExtract DNA extraction solution (Epicentre) per manufacturer's recommendations. The various loci were amplified by PCR using the primers described in **Supplementary Table 8**. Assays were performed with the Surveyor mutation detection kit (Transgenomics) as described[36, 38]. 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 $p<0.001$[39].

Recombinant adeno-associated virus production

Production of recombinant adeno-associated viral vectors was performed by the triple plasmid transfection method essentially as described[82]. 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[83]. 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 8**). 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$^{-/-}$ mice[84] 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[85] 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[86]. The biochemical genetics laboratory at the centre hospitalier universitaire de Sherbrooke performed the analyses.

REFERENCES

1. Komor, A.C., Badran, A.H. & Liu, D.R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **169**, 559 (2017).

2. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).

3. Koonin, E.V., Makarova, K.S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67-78 (2017).

4. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182 (2017).

5. Hille, F. et al. The Biology of CRISPR-Cas: Backward and Forward. *Cell* **172**, 1239-1259 (2018).

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

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

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

9. Anderson, E.M. et al. Lactobacillus gasseri CRISPR-Cas9 characterization In Vitro reveals a flexible mode of protospacer-adjacent motif recognition. *PLoS One* **13**, e0192181 (2018).

10. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771 (2015).

11. Najm, F.J. et al. Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. *Nat Biotechnol* **36**, 179-189 (2018).

12. Kim, H.K. et al. Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat Biotechnol* **36**, 239-241 (2018).

13. Colella, P., Ronzitti, G. & Mingozzi, F. Emerging Issues in AAV-Mediated In Vivo Gene Therapy. *Mol Ther Methods Clin Dev* **8**, 87-104 (2018).

14. Friedland, A.E. et al. Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol* **16**, 257 (2015).

15. Mir, A., Edraki, A., Lee, J. & Sontheimer, E.J. Type II-C CRISPR-Cas9 Biology, Mechanism, and Application. *ACS Chem Biol* **13**, 357-365 (2018).

16. Ibraheim, R. et al. All-in-One Adeno-associated Virus Delivery and Genome Editing by Neisseria meningitidis Cas9 in vivo. *bioRxiv* (2018).

17. Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**, 14500 (2017).

18. Anguela, X.M. et al. Robust ZFN-mediated genome editing in adult hemophilic mice. *Blood* **122**, 3283-3287 (2013).

19. Li, H. et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217-221 (2011).

20. Sharma, R. et al. In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777-1784 (2015).

21. Nami, F. et al. Strategies for In Vivo Genome Editing in Nondividing Cells. *Trends Biotechnol* (2018).

22. Schneller, J.L., Lee, C.M., Bao, G. & Venditti, C.P. Genome editing for inborn errors of metabolism: advancing towards the clinic. *BMC Med* **15**, 43 (2017).

23. Lau, C.H. & Suh, Y. In vivo genome editing in animals using AAV-CRISPR system: applications to translational research of human disease. *F1000Res* **6**, 2153 (2017).

24. Barrangou, R. & Horvath, P. A decade of discovery: CRISPR functions and applications. *Nat Microbiol* **2**, 17092 (2017).

25. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S.D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561 (2005).

26. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).

27. Deveau, H. et al. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* **190**, 1390-1400 (2008).

28. Garneau, J.E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71 (2010).

29. Horvath, P. et al. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *J Bacteriol* **190**, 1401-1412 (2008).

30. Briner, A.E. et al. Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol Cell* **56**, 333-339 (2014).

31. Hynes, A.P. et al. An anti-CRISPR from a virulent streptococcal phage inhibits Streptococcus pyogenes Cas9. *Nat Microbiol* **2**, 1374-1380 (2017).

32. Bolotin, A. et al. Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus. *Nat Biotechnol* **22**, 1554-1558 (2004).

33. Hynes, A.P. et al. Widespread anti-CRISPRs in virulent bacteriophages inhibit a range of Cas9 proteins. *Under review* (2018).

34. Chari, R., Mali, P., Moosburner, M. & Church, G.M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* **12**, 823-826 (2015).

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

36. Agudelo, D. et al. Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat Methods* **14**, 615-620 (2017).

37. Dalvai, M. et al. A Scalable Genome-Editing-Based Approach for Mapping Multiprotein Complexes in Human Cells. *Cell Rep* **13**, 621-633 (2015).

38. Guschin, D.Y. et al. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* **649**, 247-256 (2010).

39. Brinkman, E.K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* **42**, e168 (2014).

40. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol* **17**, 148 (2016).

41. Tomoeda, K. et al. Mutations in the 4-hydroxyphenylpyruvic acid dioxygenase gene are responsible for tyrosinemia type III and hawkinsinuria. *Mol Genet Metab* **71**, 506-510 (2000).

42. Ruetschi, U. et al. Mutations in the 4-hydroxyphenylpyruvate dioxygenase gene (HPD) in patients with tyrosinemia type III. *Hum Genet* **106**, 654-662 (2000).

43. Russell, S. et al. Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* **390**, 849-860 (2017).

44. George, L.A. et al. Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. *N Engl J Med* **377**, 2215-2227 (2017).

45. Nathwani, A.C. et al. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N Engl J Med* **371**, 1994-2004 (2014).

46. McKay, T.R. et al. Perinatal gene transfer to the liver. *Curr Pharm Des* **17**, 2528-2541 (2011).

47. Wang, L. et al. AAV8-mediated hepatic gene transfer in infant rhesus monkeys (Macaca mulatta). *Mol Ther* **19**, 2012-2020 (2011).

48. Wang, L., Wang, H., Bell, P., McMenamin, D. & Wilson, J.M. Hepatic gene transfer in neonatal mice by adeno-associated virus serotype 8 vector. *Hum Gene Ther* **23**, 533-539 (2012).

49. Yang, Y. et al. A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat Biotechnol* **34**, 334-338 (2016).

50. Morrow, G. & Tanguay, R.M. Biochemical and Clinical Aspects of Hereditary Tyrosinemia Type 1. *Adv Exp Med Biol* **959**, 9-21 (2017).

51. Grompe, M. Fah Knockout Animals as Models for Therapeutic Liver Repopulation. *Adv Exp Med Biol* **959**, 215-230 (2017).

52. Endo, F. 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* **272**, 24426-24432 (1997).

53. Pankowicz, F.P. et al. Reprogramming metabolic pathways in vivo with CRISPR/Cas9 genome editing to treat hereditary tyrosinaemia. *Nat Commun* **7**, 12642 (2016).

54. Nathwani, A.C. 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* **107**, 2653-2661 (2006).

55. McIntosh, J. et al. Therapeutic levels of FVIII following a single peripheral vein administration of rAAV vector encoding a novel human factor VIII variant. *Blood* **121**, 3335-3344 (2013).

56. Pawluk, A., Davidson, A.R. & Maxwell, K.L. Anti-CRISPR: discovery, mechanism and function. *Nat Rev Microbiol* **16**, 12-17 (2018).

57. Shin, J. et al. Disabling Cas9 by an anti-CRISPR DNA mimic. *Sci Adv* **3**, e1701620 (2017).

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

59. Rousseau, B.A., Hou, Z., Gramelspacher, M.J. & Zhang, Y. Programmable RNA Cleavage and Recognition by a Natural CRISPR-Cas9 System from Neisseria meningitidis. *Mol Cell* **69**, 906-914 e904 (2018).

60. Strutt, S.C., Torrez, R.M., Kaya, E., Negrete, O.A. & Doudna, J.A. RNA-dependent RNA targeting by CRISPR-Cas9. *Elife* **7** (2018).

61. Zhang, Y., Rajan, R., Seifert, H.S., Mondragon, A. & Sontheimer, E.J. DNase H Activity of Neisseria meningitidis Cas9. *Mol Cell* **60**, 242-255 (2015).

62. Stephenson, A.A., Raper, A.T. & Suo, Z. Bidirectional Degradation of DNA Cleavage Products Catalyzed by CRISPR/Cas9. *J Am Chem Soc* **140**, 3743-3750 (2018).

63. Ma, E., Harrington, L.B., O'Connell, M.R., Zhou, K. & Doudna, J.A. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol Cell* **60**, 398-407 (2015).

64. Dugar, G. et al. CRISPR RNA-Dependent Binding and Cleavage of Endogenous RNAs by the Campylobacter jejuni Cas9. *Mol Cell* **69**, 893-905 e897 (2018).

65. Chen, J.S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* (2018).

66. Boettcher, M. et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. *Nat Biotechnol* **36**, 170-178 (2018).

67. Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res* **42**, 2577-2590 (2014).

68. Goudy, K.S., Annoni, A., Naldini, L. & Roncarolo, M.G. Manipulating Immune Tolerance with Micro-RNA Regulated Gene Therapy. *Frontiers in microbiology* **2**, 221 (2011).

69. Charlesworth, C.T. et al. Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans. *bioRxiv* (2018).

70. Moreno, A.M. et al. Exploring protein orthogonality in immune space: a case study with AAV and Cas9 orthologs. *bioRxiv* (2018).

71. Wagner, D.L. et al. High prevalence of S. pyogenes Cas9-specific T cell sensitization within the adult human population - A balanced effector/regulatory T cell response. *bioRxiv* (2018).

72. Chew, W.L. et al. A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods* **13**, 868-874 (2016).

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

74. Karvelis, T., Gasiunas, G. & Siksnys, V. Methods for decoding Cas9 protospacer adjacent motif (PAM) sequences: A brief overview. *Methods* **121-122**, 3-8 (2017).

75. Leenay, R.T. & Beisel, C.L. Deciphering, Communicating, and Engineering the CRISPR PAM. *J Mol Biol* **429**, 177-191 (2017).

76. Rock, J.M. et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol* **2**, 16274 (2017).

77. Makarova, K. et al. Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci U S A* **103**, 15611-15616 (2006).

78. Chen, H., Choi, J. & Bailey, S. Cut site selection by the two nuclease domains of the Cas9 RNA-guided endonuclease. *J Biol Chem* **289**, 13284-13294 (2014).

79. Hu, J.H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63 (2018).

80. Kleinstiver, B.P. et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**, 1293-1298 (2015).

81. Chatterjee, P., Jakimo, N. & Jacobson, J.M. Divergent PAM Specificity of a Highly-Similar SpCas9 Ortholog. *bioRxiv* (2018).

82. Gray, S.J. et al. Production of recombinant adeno-associated viral vectors and use in in vitro and in vivo administration. *Curr Protoc Neurosci* **Chapter 4**, Unit 4 17 (2011).

83. Aurnhammer, C. 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* **23**, 18-28 (2012).

84. Grompe, M. et al. Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. *Genes Dev* **7**, 2298-2307 (1993).

85. Yardeni, T., Eckhaus, M., Morris, H.D., Huizing, M. & Hoogstraten-Miller, S. Retro-orbital injections in mice. *Lab Anim (NY)* **40**, 155-160 (2011).

86. 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* **832**, 24-29 (2006).