Comprehensive analysis of off-target and on-target effects resulting from liver-directed CRISPR-Cas9-mediated gene targeting with AAV vectors

Kshitiz Singh,[1] Raffaele Fronza,[3] Hanneke Evens,[1] Marinee K. Chuah,[1,2] and Thierry VandenDriessche[1]

[1]Department of Gene Therapy & Regenerative Medicine, Vrije Universiteit Brussel (VUB), Brussels, Belgium; [2]Center for Molecular & Vascular Biology, Department of Cardiovascular Sciences, University of Leuven, Leuven, Belgium; [3]ProtaGene CGT GmbH, Heidelberg, Germany

Comprehensive genome-wide studies are needed to assess the consequences of adeno-associated virus (AAV) vector-mediated gene editing. We evaluated CRISPR-Cas-mediated on-target and off-target effects and examined the integration of the AAV vectors employed to deliver the CRISPR-Cas components to neonatal mice livers. The guide RNA (gRNA) was specifically designed to target the factor IX gene (F9). On-target and off-target insertions/deletions were examined by whole-genome sequencing (WGS). Efficient F9-targeting (36.45% ± 18.29%) was apparent, whereas off-target events were rare or below the WGS detection limit since only one single putative insertion was detected out of 118 reads, based on >100 computationally predicted off-target sites. AAV integrations were identified by WGS and shearing extension primer tag selection ligation-mediated PCR (S-EPTS/LM-PCR) and occurred preferentially in CRISPR-Cas9-induced double-strand DNA breaks in the F9 locus. In contrast, AAV integrations outside F9 were not in proximity to any of ~5,000 putative computationally predicted off-target sites (median distance of 70 kb). Moreover, without relying on such off-target prediction algorithms, analysis of DNA sequences close to AAV integrations outside the F9 locus revealed no homology to the F9-specific gRNA. This study supports the use of S-EPTS/LM-PCR for direct *in vivo* comprehensive, sensitive, and unbiased off-target analysis.

INTRODUCTION

CRISPR-Cas9-mediated gene editing has been shown to be therapeutically beneficial in several preclinical disease models.[1–3] Moreover, the first gene-therapy trials yielded encouraging results.[4,5] Typically, CRISPR-Cas9-based gene editing results in a double-strand DNA break (DSB) at a specific chromosomal locus, catalyzed by the Cas9 nuclease. Alternatively, Cas9 nickases have been employed to cause single-strand DNA nicks. This DNA cleavage specificity is determined by the binding of the specific guide RNA (gRNA) to the corresponding DNA target sequences and the gRNA-mediated recruitment of Cas9 to the target locus.[6–9] Typically, DSBs are repaired by non-homologous end joining (NHEJ) requiring canonical cellular factors. However, this NHEJ repair process is not error-proof and results in nucleotide insertions and deletions (indels) at the DSB.[10] Alternatively, DNA breaks can be repaired by microhomology-mediated end joining (MMEJ) or alternative NHEJ (alt-NHEJ). MMEJ uses microhomologies at the ends of DNA breaks for DNA repair and does not require NHEJ-associated factors.[11] The addition of "donor" DNA fragments that are homologous to the CRISPR-Cas9-targeted locus results in targeted integration of this donor DNA into that locus by homology-directed repair (HDR).[12] HDR constitutes the basis of targeted gene integration into safe-harbor loci or the substitution of a defective gene with a functional copy as a means to reduce the risk of insertional oncogenesis due to random integration of the gene of interest.[13,14]

Despite the remarkable targeting specificity of CRISPR-Cas9-mediated gene editing, one of the main challenges in the field pertains to the risk of off-target effects. This stems from the observation that gRNA molecules could potentially bind to alternative homologous sites in the genome, resulting in non-specific off-target gene editing. Off-target effects could be reduced by identifying unique target sites with minimal homology to alternative genomic loci or by the use of truncated gRNAs.[15,16] Alternatively, next-generation high-fidelity Cas9 nucleases have been developed that are characterized by reduced off-target effects.[17–20] Temporal and spatial control of Cas9 expression could also potentially limit off-target effects through the use of tissue-specific or regulatable promoters or by using self-inactivating Cas9 expression systems.[21–23]

Different viral vectors and non-viral formulations have been explored and are being further optimized for *in vivo* CRISPR-Cas-based gene editing.[24,25] Adeno-associated virus (AAV)-based vectors constitute

some of the most promising vector platforms for *in vivo* delivery of CRISPR-Cas components to the brain, muscle, liver, heart, or retina.[24,26–29] In particular, we and others have demonstrated effective *in vivo* targeting following liver-directed delivery of CRISPR-Cas components with AAV8 or AAV9 vectors.[16,30–33] Liver-specific gene editing is accomplished by driving Cas9 expression from potent hepatocyte-specific promoters, whereas the gRNAs are typically expressed from a polymerase III U6 promoter.[16,34] More specifically, in one of our previous studies, the gRNAs were designed to target the mouse coagulation factor IX (FIX) gene (*F9*) (and designated as AAV-mF9-Exon1-gRNA).[16] We demonstrated that a relatively robust indel frequency in the range of 40%–50% could be attained in the *F9* gene, but only in the liver, with no evidence of targeting in other organs. This resulted in a substantial loss of FIX activity and the emergence of a bleeding phenotype, consistent with hemophilia B. No off-target indels were detected on the basis of only three different computationally predicted off-target sites. Since this type of analysis was noncomprehensive and based on only limited datasets, it may have failed to capture the most likely off-target genomic sites.[35–39] AAV vectors typically persist episomally in the transduced tissue as extra-chromosomal circular elements containing concatenated AAV genomes that serve as the predominant templates for expression of the transgene of interest in the transduced cells.[40,41] In contrast, in neonatal mice, expression of CRISPR-Cas9 components is transient after liver-directed AAV-mediated transduction due to the loss of AAV episomes in rapidly dividing hepatocytes.[16] However, AAV vectors can also stably integrate randomly into the target cell chromosomes, although this is believed to be relatively inefficient, typically resulting in a low frequency of AAV integrations (0.05%–0.1%).[42–45]

We have now conducted a more comprehensive analysis of on-target and off-target effects after liver-directed AAV-based CRISPR-Cas9-mediated gene editing and vector integrations using whole-genome sequencing (WGS) and shearing extension primer tag selection ligation-mediated PCR (S-EPTS/LM-PCR), respectively. In parallel, we also examined the vector-genome and vector-vector breakpoints. We demonstrate that on-target effects were relatively robust, yielding about 40% indels in the *F9* target locus, whereas off-target effects were very rare and/or below the WGS detection limit. Notably, the majority of AAV integrations occurred most frequently in close proximity to the CRISPR-Cas9-targeted locus in the *F9* gene when the AAV-mF9-Exon1-gRNA was used. This supports the hypothesis that AAV genomes are preferentially integrated at CRISPR-Cas9-induced DSBs at the on-target site in the *F9* gene.

The advantage of using DNA for the >100× WGS and S-EPTS/LM-PCR analysis from the same mouse cohorts as in our previous study[16] is that it allow us to *directly* link the current data to all the previously reported functional data and (limited) on-target and off-target analysis. The current study hereby overcomes some of the limitations of our previous report by conducting a much more comprehensive on-target and off-target analysis revealing the propensity of AAV vectors to stably integrate into the CRISPR-Cas9-induced DSBs of the FIX gene and the very low risk of off-target effects even when sensitive

S-EPTS/LM-PCR analysis was used based on ∼5,000 putative computationally extracted off-target sites or by gRNA homology analysis of the sequences proximal to the AAV integration sites (ISs).

RESULTS

WGS analysis

To comprehensively assess on-target and off-target effects of CRISPR-Cas9 for *in vivo* editing in the liver, WGS was performed on genomic liver DNA obtained from the same cohorts of CRISPR-Cas9-treated mice, which were previously described.[16] No new mouse injections were performed. To achieve specific targeting of the endogenous mouse *F9* gene (Figure 1A), wild-type mice (*n* = 3) (C57BL/6) were co-injected with (1) an AAV vector that expressed an *F9*-specific gRNA (designated as AAV-mF9-Exon1-gRNA) from a U6 promoter, and (2) a second AAV vector that expressed the Cas9 gene (AAV-Cas9) from a potent hepatocyte-targeted promoter.[16,46] Controls were based on WGS of genomic liver DNA of wild-type mice injected with an AAV vector expressing a scrambled gRNA (*n* = 3) (designated as AAV-scrambled-gRNA). Mean coverage of all samples (combining the coverage of AAV-scrambled and AAV-mF9-Exon1 samples) is 133 ± 34 (mean ± SD). Mean WGS coverages for AAV-scrambled and AAV-mF9-Exon1 were 118 ± 10 and 148 ± 46, respectively (Figure 1B). Indel calls at the *F9* on-target site were determined using the CRISPRessoWGS pipeline[47] and compared with the results of manual counting (Figures 1C, 1D, and S1). The indel counts at the on-target site were determined based on the CRISPRessoWGS pipeline (percentage modified: 36% ± 1%, *n* = 3) were consistent with the number of indels determined by manual counting (percentage modified: 35% ± 1%, *n* = 3) (*p* = 0.94; not significant) (Figure 1C). The percentage of indels determined by WGS to be at the *F9* target site was also in accord with the percentage of indels determined by deep sequencing of an *F9* target site-specific amplicon, as determined previously.[16] This further confirms that CRISPRessoWGS is a reliable pipeline with which to detect indels based on a WGS dataset. It could therefore be explored for automated analysis of predicted off-target sites. In contrast, other bioinformatics pipelines that were developed for automated identification of genetic variants (i.e., Lofreq, Bcftools, varscan) failed to reliably detect the expected gene editing on the basis of the current WGS datasets and were therefore not further explored (data not shown). Consequently, only the validated CRISPRessoWGS pipeline was employed for the subsequent analysis of potential off-target effects based on WGS data.

Typically, off-target effects following CRISPR-Cas9-mediated gene editing depend greatly on the extent of the homology of the target site with other sites in the genome. Using the Cas-OFFinder algorithm as part of the "offTargetPredict" pipeline (https://github.com/penn-hui/OfftargetPredict),[48] a total of 128 predicted off-target sites (Table S1) were identified as a subset of the degenerate locations of the mF9-Exon1-gRNA.

Subsequently, the 128 sites were analyzed for the presence of indels (Figure 2; Table S3) on the basis of the WGS dataset. Notably, there was no significant difference at the potential predicted
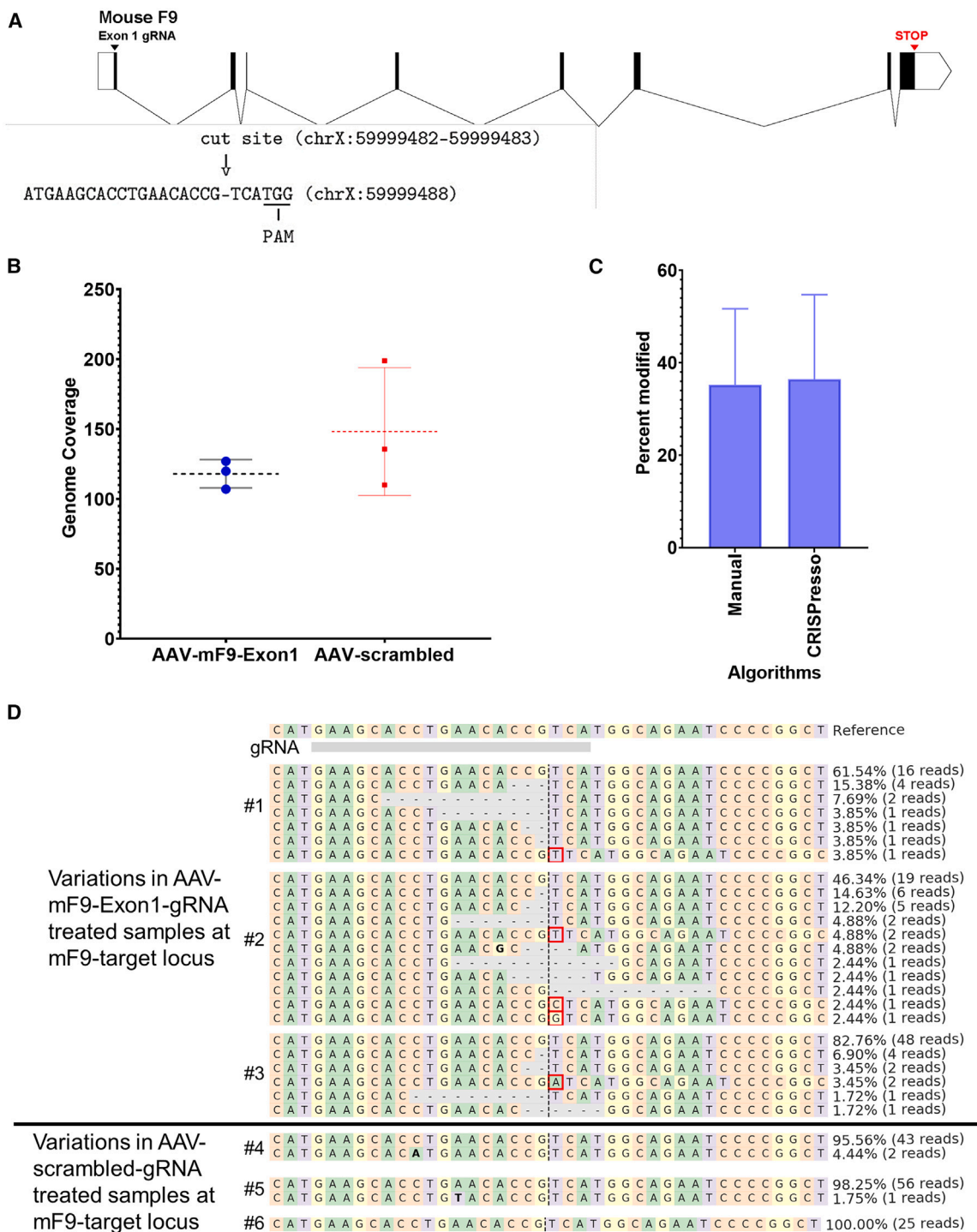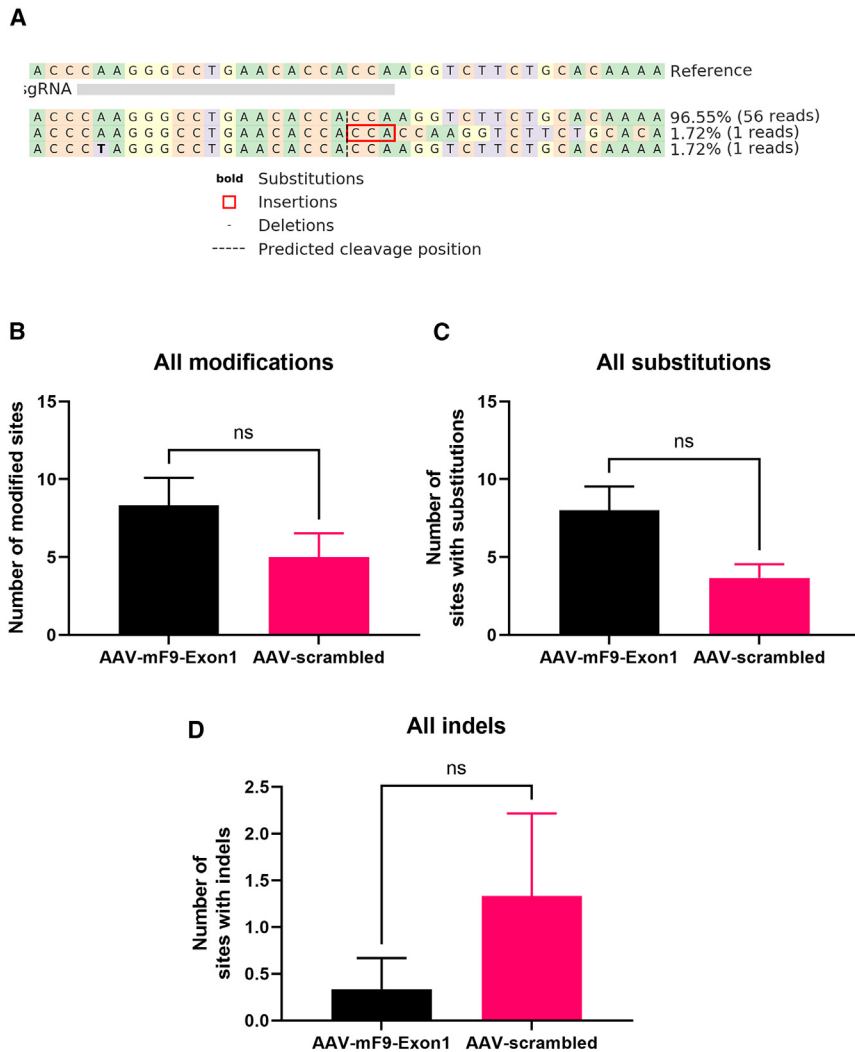
Figure 1. WGS coverage and benchmarking of on-target genetic variant calls

(A) Representative schema of AAV-mF9-Exon1-gRNA on-target site. Position chrX:59999482-59999483 represents the AAV-mF9-Exon1-gRNA cut site in the exon 1 of *F9* gene. Vertical lines, *F9* exons; horizontal lines, *F9* introns; black downward arrow, AAV-mF9-Exon1-gRNA on-target site; red downward arrow, stop codon of *F9* gene; PAM, protospacer-adjacent motif. (B) Plot showing the distributions of the coverage of the AAV-mF9-Exon1-gRNA and AAV-scrambled-gRNA WGS datasets separately. (C) Detection of editing by manual counting and CRISPResso in the WGS dataset at AAV-mF9-Exon1-gRNA target site in liver. (D) Indels at AAV-mF9-Exon1 target site by CRISPResso pipeline. Red squares, insertions; dots, deletions; vertical dashed line, expected cut site; bold, substitution. Manual counting of indels at this site can be found in Figure S1.

Off-target analysis at predicted sites

A

B

C

D

particular, this CCA insertion was detected in only one out of 118 total reads and only in one of the three DNA samples obtained from the three different CRISPR-Cas9-treated mice. The presence of such an indel at this putative off-target site is likely the consequence of an NHEJ-mediated repair event. One other sequence read showed the presence of an A-T substitution at position chr2:25583602 in the mm10 reference genome assembly. Whether this A-T substitution is the consequence of a *bona fide* CRISPR-Cas9-mediated off-target event or a possible technical artifact such as a sequencing error is currently uncertain.

Most importantly, WGS analysis revealed the presence of AAV inverted terminal repeats (ITRs) across the genome, consistent with AAV vector integration. The AAV integrations discovered in the WGS datasets are described in Table S2. Notably, in AAV-mF9-Exon1-gRNA-treated mice, there was a common AAV IS on chromosome X, in proximity to the AAV-mF9-Exon1-gRNA target site in the

AAV-mF9-Exon1-gRNA off-target sites in terms of the number of indels, substitutions, or all modifications when comparing the AAV-mF9-Exon1-gRNA and AAV-scrambled-gRNA control samples (Figures 2 and S2). Collectively, the off-target analysis based on the WGS datasets suggests that there is only a limited probability of a potential off-target effect associated with the use of the AAV-mF9-Exon1-gRNA.

Nevertheless, as compared with the reference sequence, only one potential off-target site (designated as OT77, Figures 2A and S2) out of the 128 computationally predicted sites was detected in only one out of the three AAV-mF9-Exon1-gRNA-treated samples. This single indel corresponded to a 3-nucleotide (nt) insertion (CCA), 3 nt upstream of the putative NRG PAM at the predicted CRISPR-Cas off-target site (chr2:25,583,600–25,583,619 in the mm10 reference genome). In

F9 gene (Table S2). Although AAV integrations were also apparent in the AAV-scrambled-gRNA-treated samples, AAV integrations at the targeted *F9* locus were only detected in the AAV-mF9-Exon1-gRNA and not in the AAV-scrambled-gRNA cohorts (Table S2).

S-EPTS/LM-PCR analysis

For a more comprehensive independent analysis of the AAV integration events, S-EPTS/LM-PCR was employed, given its superior sensitivity compared with WGS. AAV integration occurred across the entire genome, with no common ISs that were shared between the AAV-mF9-Exon1-gRNA and AAV-scrambled-gRNA conditions (Figure 3). Moreover, our analysis revealed the distributions of AAV integrations were similar between the AAV-mF9-Exon1-gRNA versus AAV-scrambled-gRNA cohorts (Figures S3 and S4). These values are in line with similar observation in other studies.[49]
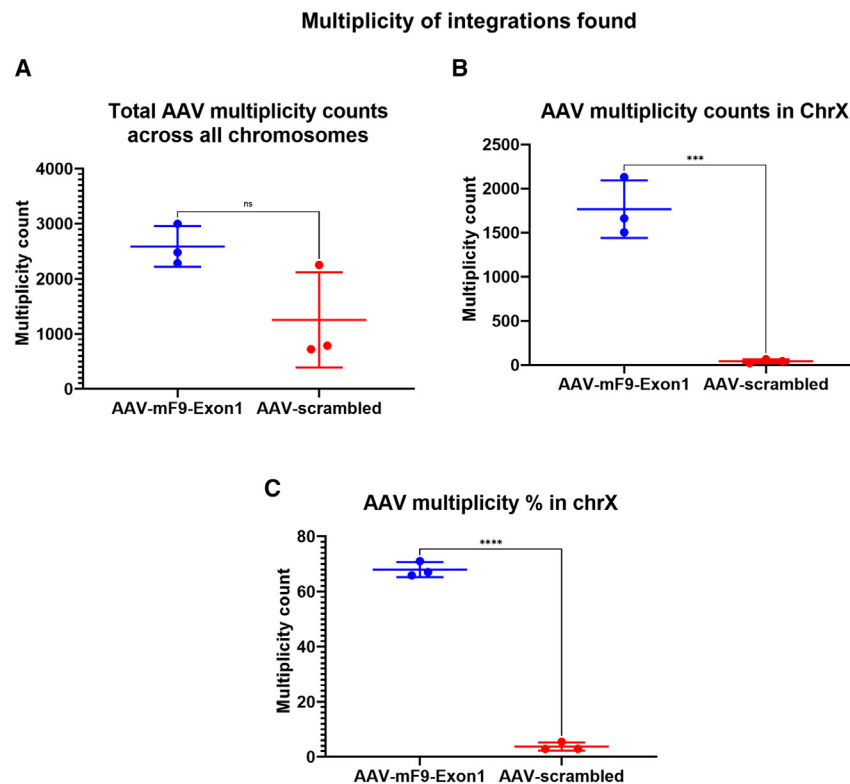
Multiplicity of integrations found

A

Total AAV multiplicity counts across all chromosomes

B

AAV multiplicity counts in ChrX

C

AAV multiplicity % in chrX

Figure 3. AAV integration multiplicity counts based on the S-EPTS/LM-PCR dataset

(A) Multiplicity of AAV integration over the entire genome.
(B) Multiplicity count of AAV integration events in chromosome X.
(C) Multiplicity percentage of AAV integration events in chromosome X.
Statistical analysis was based on unpaired t test (GraphPad). Mean ± SD is shown (ns, $p > 0.05$; *$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$; ****$p \leq 0.0001$).

This indicates that these recombinant AAV vectors do not have a propensity to integrate into specific common ISs. This in contrast to the preferential AAV Rep-mediated integration of wild-type AAV into the AAVS1 locus located on human chromosome 19 in human cells.[50]

Total multiplicity counts across the entire genome did not differ significantly between DNA from the mice treated with AAV-mF9-Exon1-gRNA and those treated with AAV-scrambled-gRNA (Figure 3A). Although AAV integration events were distributed all over the genome, AAV integration multiplicity counts were significantly higher in chromosome X when the AAV-mF9-Exon1-gRNA was employed as compared with the AAV-scrambled-gRNA control condition (Figures 3B and 3C). A comparative analysis of AAV integration multiplicity counts in all chromosomes is shown in Figure S4. Notably, when the AAV-mF9-Exon1-gRNA was employed most of these AAV integration events were at or in close proximity to the AAV-mF9-Exon1-gRNA target site (Figures 3 and 4). Hence, these S-EPTS/LM-PCR-based AAV integration data are consistent with the WGS analysis (Table S2). The AAV integration frequency decreased with increasing distance from the CRISPR-Cas cut site in the *F9* gene (Figure 4).

Corresponding Circos plots of AAV integration events over the entire genome demonstrated that for AAV-mF9-Exon1-gRNA-treated samples, the maximum frequency of AAV integration events was found in chromosome X at the AAV-mF9-Exon1-gRNA-targeted locus (Figure 5). In c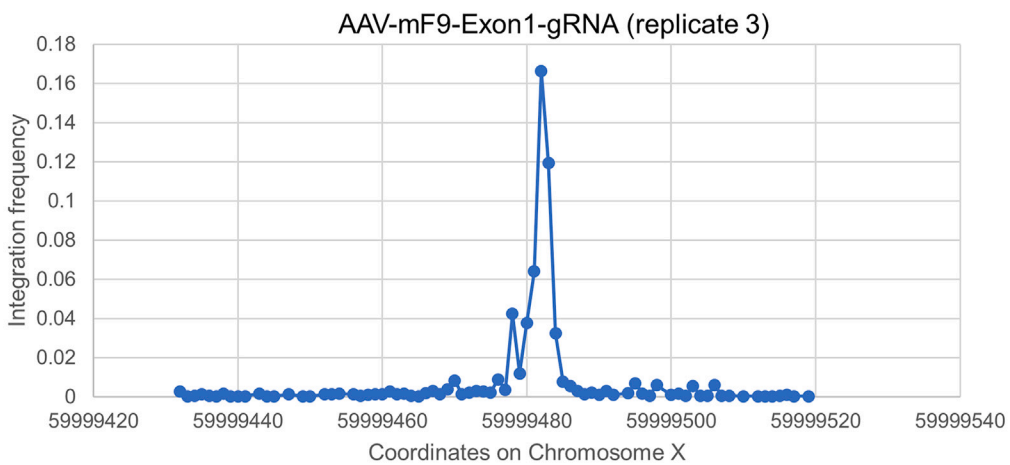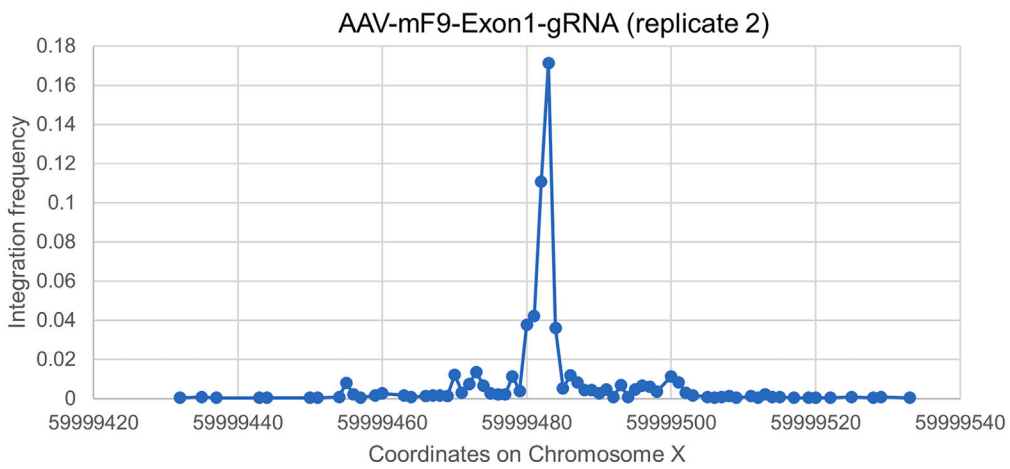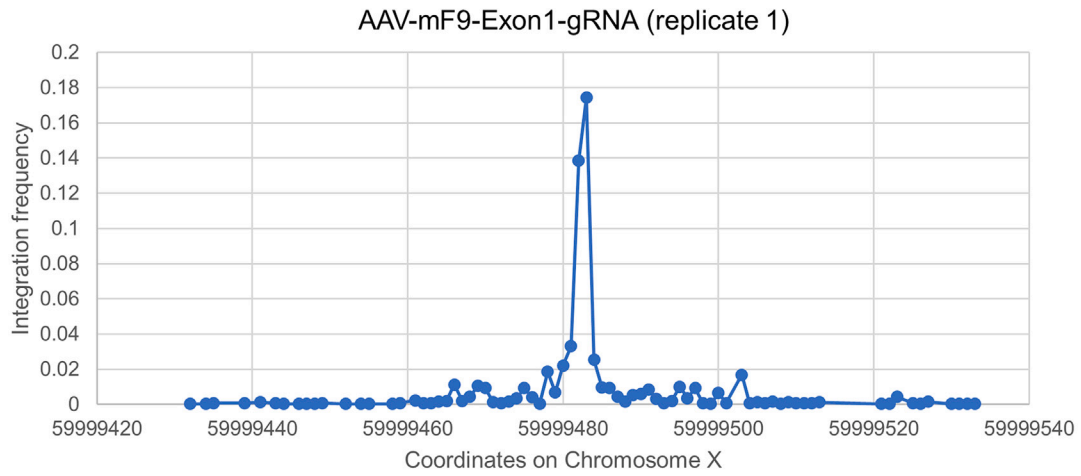ontrast, when the AAV-scrambled-gRNA was employed, AAV integrations at the AAV-mF9-Exon1-gRNA target site could not be detected (Figures 5 and S4). Taken together, these data are consistent with the hypothesis that AAV integrates into CRISPR-Cas-induced DSBs.

The genomic DNA adjacent to the AAV IS was subsequently analyzed. The characterization of both sides of the target locus in a subset of clusters in one of the samples is shown in Table S5. The Cas9 protein is expected to cleave the genomic DNA at position 59999482, provided the F9-Exon1-gRNA is supplied. In Table S4, the two most represented clusters with the highest multiplicities (432 and 343 unique reads) are composed of genomic sequences of integrations in the reverse directions that show minor deviations from the expected *F9* target sequence. Notably, this implies that most AAV vectors integrated precisely at the CRISPR-Cas9 target site (more than 90% of the reads found).

We have now extended our IS analysis beyond the analysis of the multiplicity counts by comparing common ISs (CISs) and unique AAV ISs in *individual* mice. For clarity, we define CISs as clusters of ISs situated within a proximity of less than 50 kilobases to one another, adhering to the criteria set forth by Fronza et al.[51] The top 15 unique ISs for each individual mouse were meticulously delineated, presenting these findings in visually informative rainbow plots that include detailed gene annotations (Figures 6 and 7). An alternative representation based on a compilation of CISs in all of the mice injected with AAV-Cas9 and either AAV-mF9-Exon1-gRNA or AAV-scrambled-gRNA is shown in Figure 8, based on our previously described graph-based framework.[51] As estimated 55%–60% of AAV integrations occurred in the CRISPR-induced DSB in the on-target F9 locus. Through this refined analysis, we have successfully identified CISs (other than the F9 on-target locus), some of which overlap among individual mice as shown below in Figures 6, 7, and 8.

Some AAV CISs mapped to several highly expressed genes in the liver, including Alb (reads per kilobase per million mapped reads

AAV integrations at AAV-mF9-Exon1-gRNA target site ChrX:59999466-59999488

AAV-mF9-Exon1-gRNA (replicate 1)

AAV-mF9-Exon1-gRNA (replicate 2)

AAV-mF9-Exon1-gRNA (replicate 3)

(legend on next page)

[RPKM], 14891; https://www.ncbi.nlm.nih.gov/gene/11657) and Trf (RPKM, 3915; https://www.ncbi.nlm.nih.gov/gene/22041). In addition, CISs were also detected in the miR101c gene. These CISs were found in at least two different mice belonging to different cohorts (i.e., injected with AAV-Cas9 and either AAV-mF9-Exon1-gRNA or AAV-scrambled gRNA). This suggests that the frequent occurrence of these CIS in the Alb, Trf, and miR101c gene may be related to their high levels of expression and/or other structural features pertaining to chromatin accessibility. This is consistent with previous reports indicating that actively transcribed genes are preferred targets for AAV integration.[52] Functional annotations of the top AAV ISs outside the *F9* locus revealed that some of these genes are involved in apoptosis and immune system related functions, which may raise some potential safety concerns (Table S6).

Detection of unbiased off-target AAV integrations

The propensity of AAV vectors to integrate into CRISPR-Cas-induced DSBs provides a unique opportunity to identify potential off-target effects in an unbiased manner. The integrated AAV vectors could serve as distinctive "molecular anchor-points" that serve as templates for subsequent S-EPTS/LM-PCR analysis, which is more sensitive than WGS. The most likely putative off-target sites were identified, taking into account sequence degeneracy compared with the AAV-mF9-Exon1-gRNA or AAV-scrambled-gRNA by allowing a maximum variation of 6 nt, including 4 nt with respect to the AAV-mF9-Exon1-gRNA or AAV-scrambled-gRNA and 2 nt in the SpCas9 PAM sequence itself (i.e., NRG), given its known redundancy. Taking this degeneracy into account, a total of 5,215 putative off-target sites (composed of 2,625 on the plus and 2,590 on the minus AAV strand; Table S5) that aligned with the AAV-mF9-Exon1-gRNA were identified. Based on the same strategy, 10,207 putative off-target sites (composed of 5,105 on the plus and 5,102 on the minus AAV strand; Table S5) that aligned with the scrambled gRNA were identified. The median distance between the AAV ISs and the putative off-target sites was more than 100 kb for the gRNA and more than 70 kb for the scrambled gRNA (Table S5). Notably, not a single genomic AAV IS was identified in close proximity (<100 nt) to any putative AAV-mF9-Exon1-gRNA off-target site. The closest AAV genomic integration was 157 nt away from a putative AAV-mF9-Exon1-gRNA off-target site, and it is therefore unlikely that this represents a *bona fide* off-target site. Similarly, only a single genomic AAV IS was identified that was in close proximity (14 nt) to a putative scrambled gRNA off-target site. In conclusion, CRISPR-Cas9-mediated off-target events are undetectable or relatively rare, based on SEPT/LM-PCR analysis, consistent with the results obtained by the high-coverage WGS analysis.

gRNA homology detection at AAV ISs and Monte Carlo analysis

We subsequently aimed to determine whether there was an enrichment of ISs homologous to the mF9-Exon1-gRNA sequence in the mF9-Exon1-gRNA-treated group compared to the scrambled

RNA-treated group. In the mF9-Exon1-gRNA-treated group, the maximum number of observed matches between the mF9-Exon1-gRNA sequence and AAV ISs was 12 for the plus strand and 14 for the minus strand (Figures S5A and S5C). In the scrambled RNA-treated control group, the maximum observed matches for the gRNA sequence were 17 for the plus strand and 14 for the minus strand (Figures S5B and S5D). For the scrambled RNA sequence, no significant enrichment of homology was observed in either the mF9-Exon1-gRNA-treated or scrambled RNA-treated groups, as all observed matches were below or equal to the 99th percentile threshold from the Monte Carlo (MC) simulations. These findings indicate no preferential homology between the mF9-Exon1-gRNA sequence and the AAV ISs in the mF9-Exon1-gRNA-treated mice, suggesting an absence of specific off-target nuclease activity. The same lack of enrichment was observed for the scrambled RNA sequence in both groups, further supporting the conclusion that neither guide RNA (mF9-Exon1-gRNA or scRNA) led to off-target cleavage events that would have revealed AAV integrations at these locations.

It is worth noting that the maximum of 17 matches observed for the mF9-Exon1-gRNA sequence in the scrambled RNA-treated group suggest a very rare random integration event, as it exceeded the 99th percentile of the MC simulation ($p < 0.01$). However, this event is likely unrelated to CRISPR-Cas9 activity, given that the scrambled RNA group did not receive the FIX-specific gRNA.

Consequence of genomic integration on AAV-ITR configuration

AAV can exist either as non-integrated concatemeric episomal forms or integrate into the target cell chromosomes. This genomic AAV integration can occur either randomly or at the CRISPR-Cas9 on-target or an off-target site. The genomic DNA sequence in proximity to the CRISPR-Cas9 on-target site that captures the AAV genome was previously shown (Figure 1D). However, the consequences of this genomic AAV integration into CRISPR-Cas-induced DSBs is not well understood.

We therefore examined the ITR sequence of the integrated AAV vector genomes at this CRISPR-Cas on-target site (designated as the "R1-on" configuration). For comparison, we also determined the AAV-ITR sequence following random genomic integration (designated as the "R1" configuration) and that of unintegrated concatemeric AAV episomes (designated as the "R2" configuration). The R1-on and R1 configurations encompass vector-chromosome junctions, whereas the R2 configuration corresponds to vector-vector junctions. The theoretical normal ITR sequence present in an AAV vector is shown in Figure S6, whereas the actual fraction of sequence reads at each position in the 3′ ITR is shown in Figures S7A and S7B. The B-arm is defined as 5′-CGGGCGACCTTTGGTCGCCCG-3′ or its reverse complement, and the C-arm is defined as 5′-CGC CCGGGCTTTGCCCGGGCG-3′ or its reverse complement.[53] The

---

Figure 4. AAV integration frequency at the AAV-mF9-Exon1-gRNA target site chrX:59999466-59999488

AAV integration frequency was determined by S-EPTS/LM-PCR and was based on *n* = 3 independent liver samples after cotransduction with AAV-Cas9 and AAV-mF9-Exon1-gRNA.
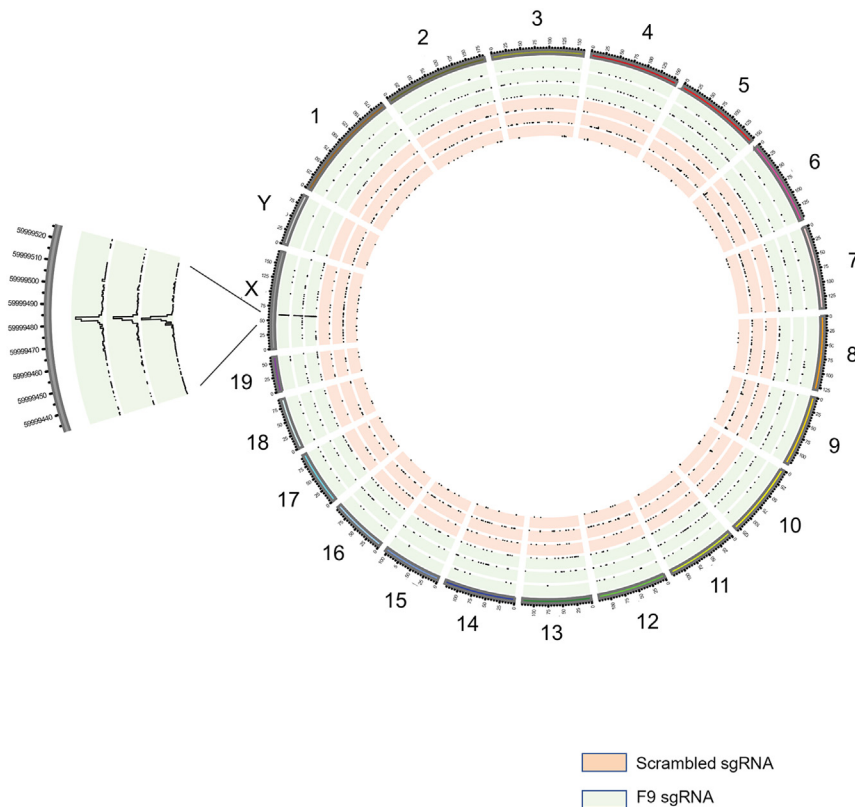
Figure 5. Circos plot of AAV integration events in all mouse chromosomes

The zoomed region shows the integration events at the target site of AAV-mF9-Exon1-gRNA, where AAV integration events were found to be most frequent.

In the current study, we therefore conducted a more comprehensive analysis that relied on a high-coverage (>100×) Illumina WGS platform. A relatively high percentage of indels was detected in the WGS dataset on the *F9* on-target site (about 40%) following automated analysis with the CRISPRessoWGS pipeline,[47] consistent with our previous analysis.[16] In contrast, based on this WGS analysis in conjunction with the CRISPRessoWGS pipeline, no indels could be detected in two out of three liver DNA samples, taking into consideration the most likely 128 computationally predicted sites. Only a single indel was detected in only one out of three liver DNA samples at only a single potential off-target site out of the 128 predicted sites based on high-coverage WGS. This indicates that the occurrence of off-target indels following liver-directed gene editing is relatively rare and/or below the detection limit of the WGS. This underscores the potential and limitations of off-target analysis using high-coverage WGS and biocomputational analysis using the CRISPRessoWGS pipeline. Other studies that relied on WGS for analyzing CRISPR-Cas off-target effects found that off-target effects were rare, whereas in some cases a higher prevalence of off-target effects was apparent.[56–58] However, these studies were all limited to *in vitro* editing applications and/or were limited by a substantially lower coverage. Although the WGS coverage in the current study was relatively high, uncovering rare off-target events (<1%) with this technology remains challenging.[59–61] Moreover, pre-existing genetic variation, DNA replication errors, or other non-editing sources of mutation may confound interpretation of the data.[62–65] However, in the current study, scrambled gRNAs were used as controls in littermates, and that should have minimized the effects of some of these confounding variables.

The current study cautions that the sensitivity of the WGS analysis for off-target analysis is limited even at a high sequence coverage. Nevertheless, WGS allowed us to uncover the propensity of AAV vectors to integrate into the CRISPR-Cas-induced DSB in the *F9* locus, which we had not detected in our previous study that relied on Sanger sequencing of cloned PCR-amplified products or deep sequencing corresponding to the on-target FIX loci.[22] Typically, PCR products were generated of size 200–230 bp that were sequenced using the MiSeq PE150 platform. Consequently, PCR amplification introduced an intrinsic bias preventing efficient amplification of target regions that contained large integrated (concatemeric) AAV genomes. This

integrity of the AAV 3′ ITR is maintained in all configurations (R1-on, R1, and R2) until nucleotide 5,950 (Figure S7). However, beyond this nucleotide, the 3′ ITR is progressively deleted to the extent that only less than 10% of the sequence reads contain the 3′ ITR up to position 5,005. Notably, at position 4,998, a discontinuity in the breakpoint profile was observed. In the R2 configuration (episomal AAV), more than 25% of the sequences have a breakpoint at that position, whereas this breakpoint is less abundant in the case of integrated AAV vector genomes, with less than 10% of the R1-on and 5% of the R1 configurations showing this breakpoint at 4,998 (red arrow, Figure S6).
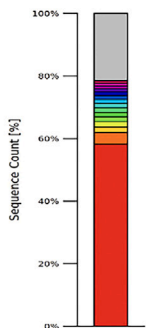
DISCUSSION

Robust delivery of CRISPR-Cas9 components is required to achieve efficient gene editing *in vivo*. Typically, this can be accomplished using AAV vectors that express the gRNA and the Cas9 nuclease.[54,55] In particular, in our previous study we demonstrated that AAV vectors expressing SpCas9 from a potent hepatocyte-specific promoter in conjunction with a U6-driven gRNA vector targeting the endogenous mouse *F9* locus provide relatively efficient gene targeting (up to 40%–50%) at the *F9* locus.[16] In our previous study, no apparent off-target effects could be detected on the basis of a very limited set of computationally predicted potential off-target sites (*n* = 3). However, genome-wide analyses are required to expand off-target analysis that goes beyond this initial limited set of off-target sites.

Total Number of IS: 163

Rank	Gene	Freq. [%]	Location
Top1	F9	58.225	chrX+59999483
Top2	Gnao1	3.663	chr8+93810837
Top3	Lef1	1.761	chr3+131110296
Top4	Sec23b	1.691	chr2+144556228
Top5	Mir6238	1.585	chr7-53891760
Top6	Fhit	1.444	chr14-9550093
Top7	Ptk2	1.444	chr15-73205104
Top8	Ptk2	1.444	chr15-73215556
Top9	Unc5cl	1.303	chr17+48454900
Top10	Ppargc1a	1.198	chr5-51454248
Top11	Tango6	1.198	chr8+106683067
Top12	Ankrd23	0.951	chr1-36530533
Top13	9330117O12Rik	0.881	chr18+54611743
Top14	Lef1	0.881	chr3+131112790
Top15	Sars	0.845	chr3-108424863
#All Other mapp. IS	148	21.486	

Seq Count 15 strongest	2229
Seq Count all other mapp. IS	610
Total Seq Counts Used	2839

Total Number of IS: 293

Rank	Gene	Freq. [%]	Location
Top1	F9	53.904	chrX+59999483
Top2	Tmem196	2.465	chr12+119945961
Top3	Dis3l2	2.167	chr1+86703803
Top4	Asmt	1.868	chrX+170672643
Top5	Wdr33	1.644	chr18+31804056
Top6	Cnih3	1.569	chr1+181352627
Top7	Nlgn1	1.419	chr3-25431840
Top8	Ahsg	1.195	chr16+22892014
Top9	Gm21708|Gm3376|Rb	1.195	chrY+3771672
Top10	Gm22109|Gm22291|G	0.897	chr8-123543870
Top11	D830032E09Rik	0.822	chr1+107917258
Top12	Tmbim6	0.784	chr15+99392946
Top13	Dapk1	0.747	chr13+60601946
Top14	Msra	0.747	chr14-64122620
Top15	Akap7	0.672	chr10-25169089
#All Other mapp. IS	278	27.904	

Seq Count 15 strongest	1930
Seq Count all other mapp. IS	747
Total Seq Counts Used	2677

Total Number of IS: 325
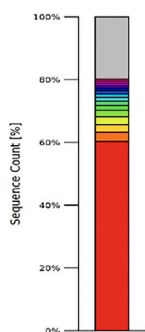
Rank	Gene	Freq. [%]	Location
Top1	F9	60.358	chrX+59999483
Top2	Asmt	2.876	chrX+170672643
Top3	Tpcn2	2.465	chr7-145253922
Top4	Gm21708|Gm3376|Rb	2.465	chrY+3771672
Top5	E330011O21Rik	2.054	chr16-78250862
Top6	Ido2	1.555	chr8-24531891
Top7	Alb	1.379	chr5+90460888
Top8	Gm22109|Gm22291|G	1.232	chr8-123538783
Top9	Ttc39b	1.056	chr4-83220300
Top10	Kcnma1	0.998	chr14-23298693
Top11	Tnfaip8l1	0.910	chr17+56162490
Top12	1700065L07Rik	0.792	chr6-73436964
Top13	Mtmr3	0.763	chr11-4480867
Top14	Gm22109|Gm22291|G	0.704	chr8-123543870
Top15	Wls	0.646	chr3+159839694
#All Other mapp. IS	310	19.748	

Seq Count 15 strongest	2735
Seq Count all other mapp. IS	673
Total Seq Counts Used	3408

Figure 6. Rainbow plot of ISs in mice injected with AAV-mF9-Exon1-gRNA and AAV-Cas9

Each rainbow plot refers to an individual mouse. The corresponding table shows the IS hotspots and respective frequency.
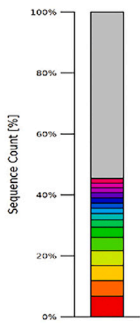
Total Number of IS: 192

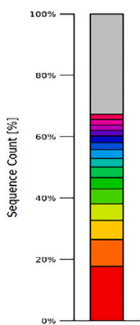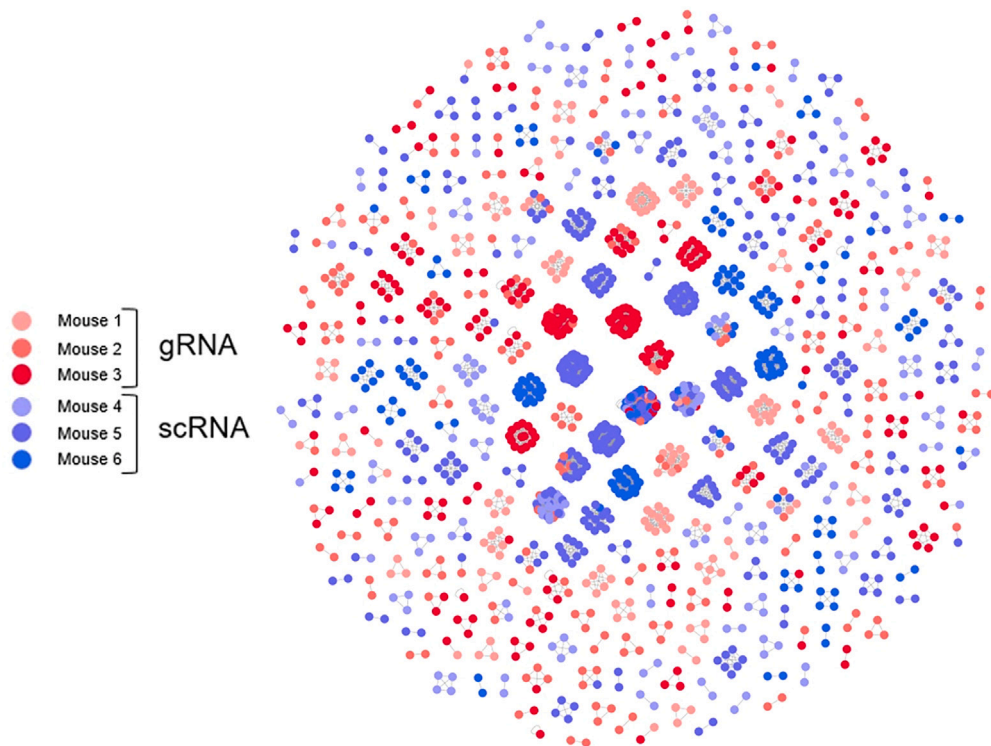| Rank		Gene	Freq. [%]	Location
Top1		Sox6	6.742	chr7-115470871
Top2		Lrrc4c	5.056	chr2+96318168
Top3		Lrrc4c	5.056	chr2+97467656
Top4		Mir101c	4.775	chr9-3038668
Top5		Egfr	4.494	chr11+16752202
Top6		Pex2	3.277	chr3-5560187
Top7		Gm5134	2.434	chr10+75954513
Top8		Ctnnbl1	2.060	chr2+157737400
Top9		Esp31	1.779	chr17+38639446
Top10		Mgll	1.685	chr6+88724411
Top11		Mgll	1.685	chr6+88724703
Top12		Tacc1	1.685	chr8-25154551
Top13		Slc15a1	1.592	chr14-121459620
Top14		Birc3	1.592	chr9-7848700
Top15		Ralgapa1	1.498	chr12-55602889
#All Other mapp. IS		177	54.588	

Seq Count 15 strongest	485
Seq Count all other mapp. IS	583
Total Seq Counts Used	1068

Total Number of IS: 381

Rank		Gene	Freq. [%]	Location
Top1		Slc25a13	12.275	chr6-6041217
Top2		Atf2	5.623	chr2-73816508
Top3		Fam214b	5.591	chr4-43032413
Top4		Ralyl	3.856	chr3+13471654
Top5		C130030K03Rik	3.695	chr10-49093680
Top6		Lars2	3.631	chr9+123366939
Top7		4933422A05Rik	3.567	chr9+35281104
Top8		Ncam2	3.535	chr16+81200696
Top9		A530053G22Rik	3.117	chr6-60279042
Top10		Cntnap5b	2.763	chr1+99772764
Top11		Ifrd1	2.731	chr12-40203128
Top12		Bhmt	2.474	chr13-93616890
Top13		Nlgn1	1.864	chr3-25431840
Top14		Alb	1.864	chr5+90460888
Top15		Trf	1.864	chr9-103208875
#All Other mapp. IS		366	41.549	

Seq Count 15 strongest	1819
Seq Count all other mapp. IS	1293
Total Seq Counts Used	3112

Total Number of IS: 129

Rank		Gene	Freq. [%]	Location
Top1		Rap1gds1	17.758	chr3-138925896
Top2		Mms22l	8.700	chr4+24496461
Top3		Alb	6.278	chr5+90460888
Top4		Usp15	5.381	chr10-123105005
Top5		Celf4	4.933	chr18-25477619
Top6		4933411G06Rik	3.677	chr10-51756173
Top7		Mfsd7c	3.408	chr12+85746538
Top8		Zmynd8	2.870	chr2-165784151
Top9		Wisp1	2.780	chr15+66891392
Top10		Prkca	2.242	chr11-107933386
Top11		Dis3	2.242	chr14-99076633
Top12		Adk	1.794	chr14+21052573
Top13		Adk	1.794	chr14+21076151
Top14		Dusp11	1.794	chr6-85942268
Top15		Cypt14	1.704	chrX-39862918
#All Other mapp. IS		114	32.646	

Seq Count 15 strongest	751
Seq Count all other mapp. IS	364
Total Seq Counts Used	1115

Figure 7. Rainbow plot of ISs in mice injected with AAV-scrambled-gRNA and AAV-Cas9

Each rainbow plot refers to an individual mouse. The corresponding table shows the IS hotspots and respective frequencies.
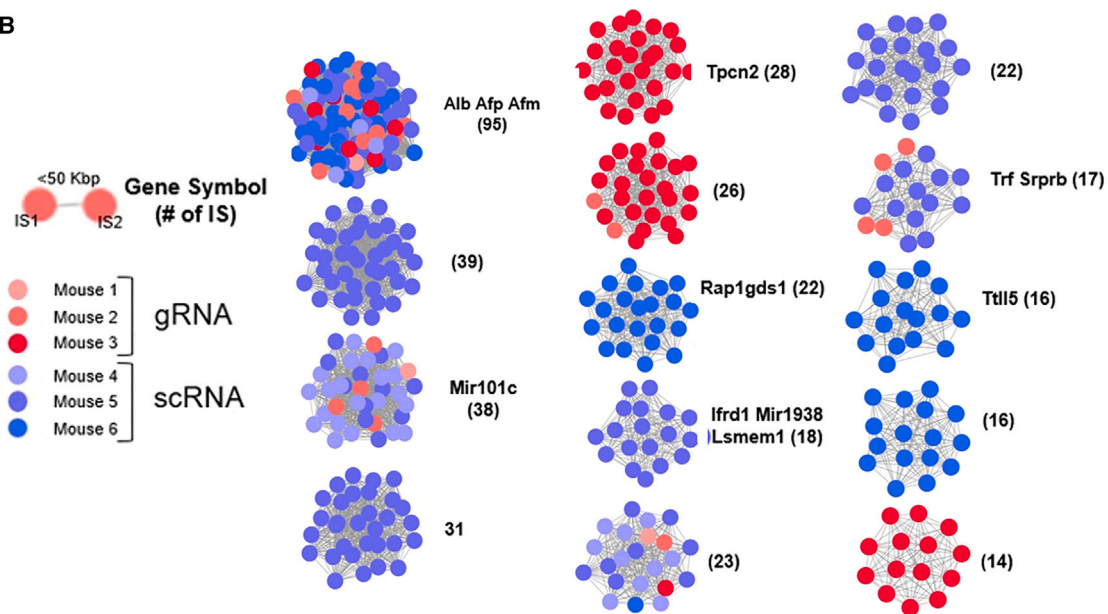
Figure 8. CISs in mice injected with AAV-Cas9 and either AAV-mF9-Exon1-gRNA (mouse 1, 2, and 3, abbreviated as gRNA) or AAV-scrambled-gRNA (mouse 4, 5, and 6, abbreviated as scRNA)

CISs identified in different individual mice have a different color code, as indicated. A CIS is represented as a network of nodes that symbolize a single integration event. An edge between two nodes indicates that the two corresponding ISs are within a range of 50 kbp. The number in parentheses corresponds to the number of ISs that compose the CIS (order). The bottom panel (B) is an enlargement of the CIS in (A) that contain 14 or more integrations (order ≥ 14).

allowed only for the shorter amplicons devoid of AAV integrations to be characterized by subsequent sequencing analysis. In addition, the PCR amplifications were based on primers that mapped to the FIX target site and were not specifically designed to map to the AAV vector allowing identification of AAV ISs in the genome.

The limitations of WGS prompted us to validate a more sensitive method based on S-EPTS/LM-PCR to comprehensively assess CRISPR-Cas-induced on-target (i.e., F9) and off-target effects in the liver. Using both WGS and S-EPTS/LM-PCR technology, we have now demonstrated that AAV vector genomes frequently integrate into CRISPR-Cas-induced DSBs precisely at the on-target site in the *F9* locus following liver-directed delivery of the Cas9 and gRNA vectors. The on-target AAV integration often occurred precisely at the CRISPR-Cas cut site itself, while minor nucleotide changes at the AAV vector-genome junctions also occurred, consistent with possible indels. Although AAV genomes typically persist as non-integrated concatemeric episomes in transduced cells, it is known that a fraction of AAV genomes integrate randomly into the target cell chromosomes.[40,66] Miller and colleagues had previously demonstrated that these AAV integrations can occur at pre-existing DSBs, which is consistent with our results.[67] Furthermore, they also showed that AAV-mediated gene targeting is enhanced by creating DSBs.[68] Similarly, intrathymic AAV delivery resulted primarily in AAV integrations clustered within the T cell receptor α, β, and γ genes that coincide with DSBs created by the enzymatic activity of recombination activating genes (RAGs) during VDJ recombination, in accordance with our current data.[69] We and others have previously shown that exogenous DNA can be trapped and integrate into pre-existing DSBs through an NHEJ-mediated mechanism.[14] The integration of AAV into DSBs is not limited to CRISPR-induced DSBs but is also consistent with previous reports demonstrating AAV integration into DSBs induced by zinc-finger nucleases (ZFNs).[14]

The propensity of AAV genomes to integrate into CRISPR-Cas9-induced DSBs could therefore be used to identify potential CRISPR-Cas9 off-target effects by exploiting these integrated AAV genomes to tag off-target sites by S-EPTS/LM-PCR, which is one of the most sensitive methods available to date to identify AAV ISs. Notably, no such AAV integrations could be identified in proximity (<100 bp) to any of the more than 5,000 computationally predicted potential CRISPR-Cas9 off-target sites. Typically, the large majority of these AAV integrations were distantly located (median distance of 100 kb) from more than 5,000 putative mF9-Exon1-gRNA off-target sites. It is therefore extremely unlikely that these particular AAV integrations would have been associated with a CRISPR-Cas9-induced off-target DSB. Instead, apart from the AAV integrations specifically into the CRISPR-Cas9 on-target site in the *F9* gene, AAV vectors integrated randomly throughout the mouse genome, with no bias for any of the more than 5,000 potential CRISPR-Cas off-target sites. This is consistent with the outcome of the gRNA homology analysis near the AAV ISs. The analysis aimed to determine whether there was an enrichment of ISs homologous to the mF9-Exon1-gRNA sequence in the mF9-Exon1-gRNA-treated group compared to the scrambled RNA-treated group. The results showed an absence of mF9-Exon1-gRNA homology enrichment within a window of 150 nt near the AAV ISs, relative to the random background established by the MC analysis. This comprehensive direct *in vivo*, unbiased genome-wide analysis reinforces the on-target specificity of the CRISPR-Cas9 system and the absence of any detectable off-target effects *in vivo*. Hence, the MC simulation effectively shaped the random homology background, reinforcing the conclusion that observed AAV integration patterns are not driven by CRISPR-Cas9-induced off-target effects. Hence, the absence of any genomic AAV integrations near any of these potential off-target sites strongly suggests that the occurrence of any CRISPR-Cas9-induced off-target effects after liver-directed AAV delivery of the CRISPR components is very rare and/or below the detection limit.

The low probability of off-target effects in our current study is likely due to the design of the gRNA. In addition, we previously demonstrated that the episomal AAV genomes encoding the CRISPR-Cas9 components decline in the transduced dividing neonatal hepatocytes, resulting in a decline in expression of the CRISPR-Cas9 components.[16] This short-term CRISPR-Cas9 expression likely further diminished the risk of off-target effects. Since some CRISPR-Cas9-induced off-target effects, such as large chromosomal translocations,[70] may have escaped detection by WGS and S-EPTS/LM-PCR, the use of complementary detection methods based on long-read sequencing may be required to detect such events.[71–73]

Our current results provide a comprehensive genome-wide analysis of AAV integration in the liver. The demonstration that AAV vectors integrate into CRISPR-induced DSBs in hepatocytes following *in vivo* delivery is consistent with a recent study demonstrating AAV integration in DSBs in brain, muscle, and cochlea and may thus represent a general, tissue type-independent phenomenon.[27,45]

Conceptionally, mapping off-target effects using the S-EPTS/LM-PCR approach resembles the VIVO technique since both strategies exploit the integration of exogenous DNA sequences into CRISPR-induced DSBs.[74] Similarly, our strategy is complementary to the capture of protected double-stranded oligonucleotides (GUIDE-seq) to identify CRISPR-induced DSBs at off-target sites in cells transfected *in vitro*.[75,76] These integrated DNA sequences could then serve as a "molecular beacon" for subsequent IS mapping using PCR-based DNA amplification. The main limitation of VIVO and GUIDE-seq techniques is that they rely on *in vitro* transfection of (hepatic) cell lines *in vitro* with oligonucleotides, which is not efficient for *in vivo* applications. Similarly, based on the same principle of DNA capture, integration-deficient lentiviral vectors (IDLVs) have been employed to identify CRISPR-induced DSBs in transfected cells *in vitro*.[77] Given the limited transduction efficacy of hepatocytes following *in vivo* IDLV delivery,[78] this strategy may not be readily amenable to *in vivo* applications. Consequently, the pattern of CRISPR-Cas-induced DSBs at putative off-target sites may be different in primary hepatocytes *in vivo* compared to transfected (hepatic) cell lines *in vitro* due, for instance, to epigenetic differences at the level of chromatin

accessibility. This can then, in turn, impact on the integration pattern of the exogenous DNA (AAV or double-strand oligo) and give a different readout about the putative off-target effects. It is likely that the *in vivo* AAV-based approach in combination with S-EPTS/LM-PCR may more closely mimic what can be expected in a clinical context, which may therefore be complementary and perhaps more accurate compared to alternative, *in vitro* cell-based or cell-free methods[79] for off-target analysis such as Digenome-seq, CIRCLE-seq, DIG-seq,[80–82] BLISS,[76,83] or DISCOVER-seq.[84] Another advantage of the S-EPTS/LM-PCR-based integration analysis described in the current study is that it allows for whole-genome screening for any AAV integration events in a comprehensive manner, providing a more complete picture of off-target activity. This is in contrast to hybrid capture and Amp-seq-based methods that are limited to capturing sequences that are predefined or predicted, potentially missing off targets that occur at unexpected locations. By directly measuring AAV ISs, our approach offers a clear and unbiased method for detecting off-target effects of CRISPR-Cas9. Hence, this strategy is more reflective of actual genomic events than *in silico* predictions as it directly captures *in vivo* outcomes.

Our current study is complementary to the recent work of Ferrari and colleagues.[85] While Ferrari et al.'s research parallels ours in demonstrating AAV vector integration at CRISPR-Cas9-induced DSBs, our study diverges in its lack of a homologous AAV donor, relying instead on alternative DNA repair mechanisms such as NHEJ. Additionally, the edited target cells are different (hematopoietic/stem progenitor cells versus hepatocytes) Moreover, our use of AAV8 for CRISPR-Cas9 delivery contrasts with their employment of ribonucleoproteins. Methodologically, while both studies leverage sonication-based PCR techniques for identifying AAV ISs, the specific protocols (S-EPTS/LM-PCR in our study versus SLiM in Ferrari et al.'s work) share foundational principles but differ in some technical aspects (e.g., the usage of unique molecular identifier to quantify the clonality), contributing to the diverse yet complementary insights generated by our respective investigations.[85]

To further address the consequences of AAV-mediated gene editing, we analyzed the ITR breakpoint profile based on the S-EPTS/LM-PCR sequencing data at genomic ISs, the gRNA-Cas9 on-target IS, and vector-vector junctions. The analysis revealed that the ITR integrity was maintained up to nucleotide 5950. The ITRs were progressively deleted beyond this position, with evidence of a discontinuous breakpoint. Though ITR breakpoints were more common in the episomal AAV vectors than the integrated AAV genomes, even AAV vectors integrated into the target site contained rearranged forms of the ITRs. It is likely that the ITR sequences facilitated intermolecular recombination of monomeric viral genomes.[40,41] The presence of integrated AAV-ITR sequence can be exploited to map potential off-target sites, similar to the LAM-PCR method described herein and constitutes the basis of the ITR-seq method.[86] In future, studying ITR-genome and ITR-vector fusions by newer sequencing methods, including those that generate longer sequences, will reveal further insights into AAV integration and its underlying mechanisms.

It is well established that the majority of recombinant AAV genomes are predominately non-integrated in AAV-transduced hepatocytes. AAV integration frequencies typically fall in the low range (0.1%–0.5%), although chromosomal integrations were somewhat higher in a humanized liver mouse model (1%–3%).[87] This contrasts starkly with the high number of integrations that are typically found upon lentiviral liver transduction.[88–90] In our study, the identification of approximately 800 ISs after hepatic liver transduction and was sufficient to uncover AAV integration hotspots among different recipient mice (such as Alb, Asmt, Nlgn1, Gm21708|Gm3376| Rbmy, Gm22109|Gm22291). While increasing the number of identified ISs could potentially reveal additional, less prominent integration hotspots, our study's primary aim was not to exhaustively catalog AAV integration but to elucidate the influence of CRISPR-Cas9 on AAV integration patterns, particularly at the F9 locus and at the computed predicted CRISPR-Cas9 off-target sites. This underscores the sensitivity of our S-EPTS/LM-PCR-based methodology and corroborates the paper's primary conclusion that AAV predominantly integrates at CRISPR-Cas9-induced breaks in the F9 locus. The number of AAV ISs that we uncovered falls within the ballpark of what has been reported in other studies that focus on AAV-based hepatic transduction and integration. For instance, in one recent groundbreaking study by the group of Sabatino et al., 1,741 AAV ISs were analyzed, which was sufficient to uncover clonal expansions of AAV-transduced liver cells.[42] Similarly, in another impactful study, Chandler and colleagues identified and mapped 2,834 unique AAV ISs in the liver.[91] The number of AAV ISs in these studies was sufficient to uncover clonal expansions and/or hepatocellular carcinoma of AAV-transduced liver cells. (e.g., *Rian* locus). More recently, using a xenogeneic humanized-liver mouse, Dalwadi et al. identified >1,200 sequencing reads containing the AAV genome, 370 of which contained the rAAV/cellular genomic junction at the IS.[87]

In conclusion, our current study reinforces the notion that AAV integration can occur at high frequency at nuclease-induced DSBs and shows that it is possible to target the genome specifically with very low prevalence of off-target events near the detection limit of available assays. A combinatorial application of several different strategies such as better bioinformatics tools and sensitive detection systems such as S-EPTS/LM-PCR, transient tissue-targeted expression of CRISPR-Cas9 expression, and high-fidelity Cas enzymes can help us to achieve efficient on-target gene editing, with (near-)zero levels of off-target gene editing.

MATERIALS AND METHODS
Animals
The liver DNA samples used in the current study were based on the animal experiments that we previously described and no new mouse injections were performed.[16] Briefly, the animal experiments were approved by the University's Animal Ethics Committee. C57BL/6 mice (Taconic, Denmark; Janvier Labs, France) were used in this study. One- to 2-day-old neonatal mice were given, by injection into the facial vein, two different doses of the AAV-Cas9 vector (i.e., 6.25×10^{10} vg/mouse intravenously [i.v.]) and AAV-mF9-Exon1-gRNA, or control

AAV-scrambled-gRNA (i.e., 1.25×10^{11} vg/mouse i.v.). Genomic DNA was extracted from different tissues using the DNeasy Blood and Tissue Kit (Qiagen, Chatsworth, CA, USA).

Library preparation and WGS

The NEBNext Ultra II FS DNA module (New England Biolabs catalog # E7810 S/L) and the NEBNext Ultra II Ligation module (catalog # NEB #E7595 S/L) were used to process the samples. Fragmentation, A-tailing, and ligation of sequencing adapters to the resulting product were performed according to the procedure described in the NEBNext UltraII FS DNA module and NEBNext Ultra II Ligation module instruction manual. The quality and yield after sample preparation were measured with the fragment analyzer. The size of the resulting product was consistent with the expected size of approximately 500–700 bp. Clustering and DNA sequencing using the NovaSeq6000 was performed according to manufacturer's protocols. A DNA concentration of 1.1 nM was used. NovaSeq control software NCS v1.6 was used. WGS ($100\times$ coverage) was performed by using the Illumina NovaSeq 6000 platform (GenomeScan, the Netherlands). Alignments were done with the GRCm38 (mm10) genome. All the samples used in this study were processed under identical conditions to minimize the possibility of any batch effects.

Analysis of predicted off-target sites in WGS dataset

For predicting the potential off-target sites for the sequence GAAG CACCTGAACACCGTCATGG based on the WGS data, we used the ensemble-learning-based offTargetPredict pipeline (https://github.com/penn-hui/OfftargetPredict).[48] In this pipeline, the Cas-OFFinder module was first used to predict the pool of candidate off-target sites, using the following parameters: PAM type "SpCas9 from *Streptococcus pyogenes*: 5′-NRG-3′ (R = A or G)" for the *Mus musculus* (mm10) genome with less than or equal to six mismatches. Then, offTargetPredict was used to further refine the predicted off-target sites. The results can be found in predict_results.csv. Somatic variant analysis in the predicted off-target sites was performed using the CRISPRessoWGS platform[47] (https://github.com/pinellolab/CRISPResso2) and thus quantified insertions, deletions, and substitutions.

Analysis of AAV ISs in WGS dataset

WGS samples underwent processing for the detection of AAV ISs. Initially, raw sequencing reads were subjected to quality filtering and trimming, ensuring a Q score of 30 or higher. Next, the BWA-MEM aligner was employed to align the reads with the reference vector sequence. Following alignment, reads exhibiting vector signatures were extracted, and this specific subset of data underwent further processing for IS detection. To achieve this, we utilized an improved version of the GENE-IS tool suite in conjunction with a combined human and vector reference genome (mm10).[92]

S-EPTS LM-PCR

S-EPTS/LM-PCR was performed using a Cas9-specific megaprimer to detect Cas9-ITR-mouse genome fusions at the 5′ end of the Cas9 sequence (primer extension, CATTTTATGTTTCAGGTTCAGG;

first-strand primers, GTTCAGGGGGAGGTGTGG + GACCCGGG AGATCTGAATTC; second-strand primers, GAGATCCACTAGGG CCGC + AGTGGCACAGCAGTTAGG). Briefly, samples were analyzed in triplicate, each with 500 ng of input genomic DNA. DNA was sheared to a median length of 500 nt using a Covaris M220 instrument. Sheared DNA was purified and primer extension was performed using a specific biotinylated primer proximal to the 3′ ITR. The extension product was again purified, followed by magnetic capture of the biotinylated DNA and washing with H_2O. The captured DNA was ligated to linker cassettes including a molecular barcode. The ligation product was amplified in a first exponential PCR using biotinylated vector and sequencing adaptor-specific primers. Biotinylated PCR products were magnetically captured, washed, and used as template for a second exponential PCR step with primers allowing deep sequencing by MiSeq technology (Illumina).

Analysis of potential off-target sites in S-EPTS/LM-PCR dataset

To determine the proximity of the ISs to potential off-target sites, the criteria used in this search were slightly different from those described above. The sequences of AAV-mF9-Exon1-gRNA (GCACCTGAAC ACCGTCA) and AAV-scrambled-gRNA (GGGTCTTCGAGAAGA CCT) with an adjacent PAM sequence for SpCas9 (NRG) were used in an exhaustive search of the mouse genome (mm10), allowing up to four mismatches in the guide sequence plus the redundancy in the PAM. We did not apply any selection criteria to the potential off-target set, and all potential off-target sites were treated as equiprobable. This strategy was designed to ensure a comprehensive and unbiased analysis of potential off-target activity of the nuclease by assessing the insertion pattern of the integration events. The distances from each IS to the nearest potential off-target site, excluding the on-target integrations, were calculated for each sample, and the two sets of distances were compared. This method provided a quantitative approach to assess the potential off-target effects of the gRNA and scrambled RNA.

S-EPTS/LM-PCR reads preprocessing

An on-target location was defined as a single region in the mouse genome that matches the gRNA sequence. An off-target region is any location in the mouse genome that matches a degenerate version of the gRNA. This degenerate version may contain some variations, but the overall sequence pattern is conserved. An AAV integration in a location of the mouse genome that does not match the degenerate version of the gRNA is considered a random AAV integration event.

To explore the two sides of the on-target locus, we exploit the randomness of the AAV integration process with respect to the strand orientation. In our experimental setup, we capture the 3′ end of the vector. The integration event, depending on its orientation, will produce genomic sequences that align either to the plus strand (forward orientation) or the minus strand (reverse orientation) of the host DNA. The reads are demultiplexed and trimmed for adapters and barcode sequences. We only process forward reads with a well-formed structure (starting with the correct primer position), as described in Gil-Farina et al.[49] Reads that start within 5 bp of the primer's end are considered for further analysis. We then analyzed the reads to detect the number

and location of vector fragments. Reads containing a single vector fragment (ITR) followed by an unknown sequence are labeled as R1 and used for IS detection and breakpoint analysis. The ITR portion is trimmed, and the remaining unknown sequences with a length greater than 20 nt are utilized for an extended IS analysis.

S-EPTS/LM-PCR ISs detection

To characterize the structure of expected multiple integration events in the on-target site, we used the dnaclust[93] algorithm to cluster all the R1 reads in each sample. Briefly, sequences with identity equal to or greater than 95% were clustered together, and the longest sequence in each cluster was designated as the seed of the cluster. We quantified the number of distinct sequences in each cluster, and the seeds were aligned to the mouse genome assembly (mm10) using BLAT.[94] The same integration locus was assigned to label all the sequences in the cluster. We then counted the number of unique reads that composed the cluster as the numerosity of independent integration events, under the simplification that each sequence in a cluster originated from a different targeted cell. We refer to this number as the multiplicity count. The IS position and the corresponding target location quantification were combined into the final integration-site table. In this table, we estimated the fraction of independent events that contributed to the integration in the on-target site.

For the distribution of the ISs in the whole mouse genome, each one was categorized according to its location relative to the gene: (1) upstream (located within 10 kb from the transcription start site [TSS]), (2) exonic (within an exon), (3) intronic (within an intron), (4) downstream (within 10 kb from the end of the gene), and (5) other (locations that do not fall under the aforementioned categories). The gene annotation used in this study corresponds to the RefSeq annotation of the mm10 mouse genome, as sourced from the UCSC database (https://genome.ucsc.edu/index.html).

On-target integration events characterization

For one of the samples (sample 1), we identified four major clusters observed upstream and downstream of the target locus. The seed sequences of each cluster found in the same location were aligned with each other using the MUSCLE algorithm.[95] From these alignments, we extracted the consensus sequence and visualized it as a sequence logo.[96] We then manually verified the alignments, and any unaligned sequences with more than 5 nt (which could indicate insertions in the target site) were compared with the vector genome. This comparison helped identify any signals of small vector-vector rearrangements.

gRNA homology detection at AAV ISs and MC analysis

To evaluate potential off-target activity of the CRISPR-Cas9 system, the homology between the regions surrounding AAV ISs and the mF9-Exon1-gRNA or scrambled RNA sequences was first assessed by comparing the number of homologous sequences in the mice treated with the mF9-Exon1-gRNA and those treated with the scRNA. The analysis was performed on both the forward and reverse complement sequences of the mF9-Exon1-gRNA and scRNA, examining a ±150 nucleotide (nt) window around each IS. The primary goal of this analysis was to determine if there was an enrichment of mF9-Exon1-gRNA homologs in the mF9-Exon1-gRNA-treated mice, which would indicate specific off-target nuclease activity. In contrast, the scRNA-treated group served as a control, where no mF9-Exon1-gRNA homology was expected.

Following this initial comparison, an MC simulation was conducted to further evaluate the signal-to-noise ratio of the observed homologies. The MC simulation aimed to establish the random background level of homology between the target sequences and the ISs. Each simulation involved scrambling the target sequences (mF9-Exon1-gRNA and scRNA) 100 times, aligning the scrambled sequences to the ISs and recording the maximum number of matches in each iteration. The 99th percentile of the simulated match distribution was used as a threshold to determine whether the observed number of homologs exceeded what would be expected by random chance, providing a measure of the statistical significance of the observed homologies.

AAV breakpoint analysis

As many vector genomes persist as non-integrated forms, we investigated rearranged vector sequences as described by Gil-Farina et al.[49] The reads that contain one or more vector fragments (ITR plus any vector fragment) were called Rx, where x = 1, 2, …, 7. The ITR portion was trimmed from the reads. The nature of the remaining portion was used to classify the reads into three classes:

(1) R1: the sequence without the ITR had a length greater than 20 nt and did not map with the AAV-Cas vector sequence; this class was used for the ISs and for the breakpoints analysis.
(2) R2: the remaining sequence mapped uninterruptedly on the vector and was used for breakpoint analysis.
(3) R*: the remaining sequence reads that map uninterruptedly on the vector or were composed of three or more vector fragments; we discarded these reads as no breakpoint or more than one breakpoint is present in the sequence.

All the seed sequences that specifically aligned with the on-target location are part of the R1-on subset within the R1 group.

The reads in the three sets (R1, R1-on, and R2) were analyzed to determine the breakpoint locations. The coordinate of the end of the first identified vector fragment was used as the breakpoint location in each read. For each position of the 3′ ITR, starting from the end of the megaprimer and moving in the 3′ direction, we calculated the fraction of reads containing that position. These fractions were then passed to an R script for visualization and plotting.

Data wrangling and analysis

Custom scripts written in R Studio (version 1.4.1) and Python 3.9 in combination with bash scripts were used for data wrangling and analysis. Further, for data import, preprocessing, and filtering, the R packages readxl, tidyverse, stringr, and janitor were used. Data plotting and statistical analysis were done using Circos[97] and

GraphPad Prism version 9.3 for windows (GraphPad Software, San Diego, CA, USA).

Rainbow plots

For each sample, the following metrics were computed and tabulated: (1) Seq Count 10 Strongest, highlighting the dominant clones within the sample; (2) Seq Count All Other Mappable IS, to capture the diversity of the less dominant clones; (3) Total Seq Count Used, providing a comprehensive view of the total clonal representation; (4) Frequency [%], calculated as the proportion of the total sequence count, to assess the distribution of clones within the sample. Additionally, for each IS, the RefSeq names of the genes located in closest proximity were identified and included in the table (gene name) to provide a genomic context. Using the data compiled in these tables, stacked histograms (referred to as rainbow plots) were constructed in Excel for each sample. These plots visually represent the clonality by stacking the sequence counts of the 10 most prominent ISs at bottom of the aggregated count of all other mappable ISs, with distinct colors used for each IS.

Graph-based analysis of CISs

The objective of our CIS analysis was to identify clusters of ISs within the mouse genome where the frequency of occurrence was higher than would be expected by random chance, as outlined by Shen et al. and Abel et al.[98,99] Such clusters may indicate regions that confer a selective advantage to the cells or are preferentially targeted by the vector used for delivery. To this end, we adopted a graph-based approach, as described by Fronza et al.,[51] to elucidate biologically significant CISs. In our analysis, each IS identified in any of the samples was represented as a unique node within a graph G, with each sample assigned a distinct color for differentiation. A pair of nodes within this graph was connected by an edge if the linear genomic distance between their corresponding IS was less than a predefined threshold (50 kilobase pairs [kbp]). This approach facilitated the construction of a network of nodes, where each connected sub-graph within G was indicative of a CIS.

DATA AND CODE AVAILABILITY

The data that support the findings of this study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

K.S. and R.F. designed and performed experiments, analyzed data, and wrote the paper. H.E. designed and performed experiments. T.V. and M.K.C. designed the experiments, analyzed data, wrote and approved the paper, and led the study. T.V. and M.K.C. applied for all grant support for this study. All authors approved the final manuscript. T.V., M.C. and R.F. are joint corresponding authors.

DECLARATION OF INTERESTS

R.F. is an employee of ProtaGene CGT GmbH. K.S. is the founder of Nanoformatics Inc., incorporated on October 8, 2024, which provides custom biomedical data analysis and bioinformatics solutions. Additionally, K.S. is a professor at St. Lawrence College, Kingston, Ontario, Canada.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtm.2024.101365.

REFERENCES

1. Pickar-Oliver, A., and Gersbach, C.A. (2019). The next generation of CRISPR–Cas technologies and applications. Nat. Rev. Mol. Cell Biol. 20, 490–507. https://doi.org/10.1038/s41580-019-0131-5.

2. Sharma, G., Sharma, A.R., Bhattacharya, M., Lee, S.-S., and Chakraborty, C. (2021). CRISPR-Cas9: A Preclinical and Clinical Perspective for the Treatment of Human Diseases. Mol. Ther. 29, 571–586. https://doi.org/10.1016/j.ymthe.2020.09.028.

3. Raguram, A., Banskota, S., and Liu, D.R. (2022). Therapeutic in vivo delivery of gene editing agents. Cell 185, 2806–2827. https://doi.org/10.1016/j.cell.2022.03.045.

4. Gillmore, J.D., Gane, E., Taubel, J., Kao, J., Fontana, M., Maitland, M.L., Seitzer, J., O'Connell, D., Walsh, K.R., Wood, K., et al. (2021). CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis. N. Engl. J. Med. 385, 493–502. https://doi.org/10.1056/NEJMoa2107454.

5. Zarghamian, P., Klermund, J., and Cathomen, T. (2022). Clinical genome editing to treat sickle cell disease-A brief update. Front. Med. 9, 1065377. https://doi.org/10.3389/fmed.2022.1065377.

6. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821. https://doi.org/10.1126/science.1225829.

7. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-Guided Human Genome Engineering via Cas9. Science 339, 823–826. https://doi.org/10.1126/science.1232033.

8. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science 339, 819–823. https://doi.org/10.1126/science.1231143.

9. Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell 157, 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010.

10. Jasin, M., and Haber, J.E. (2016). The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. DNA Repair 44, 6–16. https://doi.org/10.1016/j.dnarep.2016.05.001.

11. Fu, Y.-W., Dai, X.-Y., Wang, W.-T., Yang, Z.-X., Zhao, J.-J., Zhang, J.-P., Wen, W., Zhang, F., Oberg, K.C., Zhang, L., et al. (2021). Dynamics and competition of CRISPR–Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. Nucleic Acids Res. 49, 969–985. https://doi.org/10.1093/nar/gkaa1251.

12. Schubert, M.S., Thommandru, B., Woodley, J., Turk, R., Yan, S., Kurgan, G., McNeill, M.S., and Rettig, G.R. (2021). Optimized design parameters for CRISPR Cas9 and Cas12a homology-directed repair. Sci. Rep. 11, 19482. https://doi.org/10.1038/s41598-021-98965-y.

13. Park, C.-Y., Lee, D.R., Sung, J.J., and Kim, D.-W. (2016). Genome-editing technologies for gene correction of hemophilia. Hum. Genet. 135, 977–981. https://doi.org/10.1007/s00439-016-1699-x.

14. Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S.Y., Bhagwat, A.S., Malani, N., Anguela, X.M., Sharma, R., Ivanciu, L., et al. (2011). In vivo genome editing restores hemostasis in a mouse model of hemophilia. Nature 475, 217–221. https://doi.org/10.1038/nature10177.

15. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. 32, 279–284. https://doi.org/10.1038/nbt.2808.

16. Singh, K., Evens, H., Nair, N., Rincón, M.Y., Sarcar, S., Samara-Kuko, E., Chuah, M.K., and VandenDriessche, T. (2018). Efficient In Vivo Liver-Directed Gene Editing Using CRISPR/Cas9. Mol. Ther. 26, 1241–1254. https://doi.org/10.1016/j.ymthe.2018.02.023.

17. Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature 529, 490–495. https://doi.org/10.1038/nature16526.

18. Wang, D., Zhang, C., Wang, B., Li, B., Wang, Q., Liu, D., Wang, H., Zhou, Y., Shi, L., Lan, F., and Wang, Y. (2019). Optimized CRISPR guide RNA design for two high-fidelity Cas9 variants by deep learning. Nat. Commun. 10, 4284. https://doi.org/10.1038/s41467-019-12281-8.

19. Vakulskas, C.A., Dever, D.P., Rettig, G.R., Turk, R., Jacobi, A.M., Collingwood, M.A., Bode, N.M., McNeill, M.S., Yan, S., Camarena, J., et al. (2018). A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. Nat. Med. 24, 1216–1224. https://doi.org/10.1038/s41591-018-0137-0.

20. Tang, H., Wang, D., and Shu, Y. (2022). Structural insights into Cas9 mismatch: promising for development of high-fidelity Cas9 variants. Signal Transduct. Targeted Ther. 7, 271–273. https://doi.org/10.1038/s41392-022-01139-z.

21. Knapp, D.J.H.F., Michaels, Y.S., Jamilly, M., Ferry, Q.R.V., Barbosa, H., Milne, T.A., and Fulga, T.A. (2019). Decoupling tRNA promoter and processing activities enables specific Pol-II Cas9 guide RNA expression. Nat. Commun. 10, 1490. https://doi.org/10.1038/s41467-019-09148-3.

22. Gao, Z., Herrera-Carrillo, E., and Berkhout, B. (2019). A Single H1 Promoter Can Drive Both Guide RNA and Endonuclease Expression in the CRISPR-Cas9 System. Mol. Ther. Nucleic Acids 14, 32–40. https://doi.org/10.1016/j.omtn.2018.10.016.

23. Merienne, N., Vachey, G., de Longprez, L., Meunier, C., Zimmer, V., Perriard, G., Canales, M., Mathias, A., Herrgott, L., Beltraminelli, T., et al. (2017). The Self-Inactivating KamiCas9 System for the Editing of CNS Disease Genes. Cell Rep. 20, 2980–2991. https://doi.org/10.1016/j.celrep.2017.08.075.

24. Hu, X., Zhang, B., Li, X., Li, M., Wang, Y., Dan, H., Zhou, J., Wei, Y., Ge, K., Li, P., and Song, Z. (2023). The application and progression of CRISPR/Cas9 technology in ophthalmological diseases. Eye 37, 607–617. https://doi.org/10.1038/s41433-022-02169-1.

25. Finn, J.D., Smith, A.R., Patel, M.C., Shaw, L., Youniss, M.R., van Heteren, J., Dirstine, T., Ciullo, C., Lescarbeau, R., Seitzer, J., et al. (2018). A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing. Cell Rep. 22, 2227–2235. https://doi.org/10.1016/j.celrep.2018.02.014.

26. Torregrosa, T., Lehman, S., Hana, S., Marsh, G., Xu, S., Koszka, K., Mastrangelo, N., McCampbell, A., Henderson, C.E., and Lo, S.-C. (2021). Use of CRISPR/Cas9-mediated disruption of CNS cell type genes to profile transduction of AAV by neonatal intracerebroventricular delivery in mice. Gene Ther. 28, 456–468. https://doi.org/10.1038/s41434-021-00223-3.

27. Nelson, C.E., Wu, Y., Gemberling, M.P., Oliver, M.L., Waller, M.A., Bohning, J.D., Robinson-Hamm, J.N., Bulaklak, K., Castellanos Rivera, R.M., Collier, J.H., et al. (2019). Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. Nat. Med. 25, 427–432. https://doi.org/10.1038/s41591-019-0344-3.

28. Zhang, Y., Nishiyama, T., Li, H., Huang, J., Atmanli, A., Sanchez-Ortiz, E., Wang, Z., Mireault, A.A., Mammen, P.P.A., Bassel-Duby, R., and Olson, E.N. (2021). A consolidated AAV system for single-cut CRISPR correction of a common Duchenne muscular dystrophy mutation. Mol. Ther. Methods Clin. Dev. 22, 122–132. https://doi.org/10.1016/j.omtm.2021.05.014.

29. Hakim, C.H., Wasala, N.B., Nelson, C.E., Wasala, L.P., Yue, Y., Louderman, J.A., Lessa, T.B., Dai, A., Zhang, K., Jenkins, G.J., et al. (2018). AAV CRISPR editing rescues cardiac and muscle function for 18 months in dystrophic mice. JCI Insight 3, e124297. https://doi.org/10.1172/jci.insight.124297.

30. Li, Q., Su, J., Liu, Y., Jin, X., Zhong, X., Mo, L., Wang, Q., Deng, H., and Yang, Y. (2021). In vivo PCSK9 gene editing using an all-in-one self-cleavage AAV-CRISPR system. Mol. Ther. Methods Clin. Dev. 20, 652–659. https://doi.org/10.1016/j.omtm.2021.02.005.

31. Martinez-Turrillas, R., Martin-Mallo, A., Rodriguez-Diaz, S., Zapata-Linares, N., Rodriguez-Marquez, P., San Martin-Uriz, P., Vilas-Zornoza, A., Calleja-Cervantes, M.E., Salido, E., Prosper, F., and Rodriguez-Madoz, J.R. (2022). In vivo CRISPR-Cas9 inhibition of hepatic LDH as treatment of primary hyperoxaluria. Mol. Ther. Methods Clin. Dev. 25, 137–146. https://doi.org/10.1016/j.omtm.2022.03.006.

32. Li, N., Gou, S., Wang, J., Zhang, Q., Huang, X., Xie, J., Li, L., Jin, Q., Ouyang, Z., Chen, F., et al. (2021). CRISPR/Cas9-Mediated Gene Correction in Newborn Rabbits with Hereditary Tyrosinemia Type I. Mol. Ther. 29, 1001–1015. https://doi.org/10.1016/j.ymthe.2020.11.023.

33. Yang, Y., Wang, L., Bell, P., McMenamin, D., He, Z., White, J., Yu, H., Xu, C., Morizono, H., Musunuru, K., et al. (2016). A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. Nat. Biotechnol. 34, 334–338. https://doi.org/10.1038/nbt.3469.

34. De Giorgi, M., Li, A., Hurley, A., Barzi, M., Doerfler, A.M., Cherayil, N.A., Smith, H.E., Brown, J.D., Lin, C.Y., Bissig, K.-D., et al. (2021). Targeting the Apoa1 locus for liver-directed gene therapy. Mol. Ther. Methods Clin. Dev. 21, 656–669. https://doi.org/10.1016/j.omtm.2021.04.011.

35. Naeem, M., Majeed, S., Hoque, M.Z., and Ahmad, I. (2020). Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. Cells 9, 1608. https://doi.org/10.3390/cells9071608.

36. Rose, J.C., Popp, N.A., Richardson, C.D., Stephany, J.J., Mathieu, J., Wei, C.T., Corn, J.E., Maly, D.J., and Fowler, D.M. (2020). Suppression of unwanted CRISPR-Cas9 editing by co-administration of catalytically inactivating truncated guide RNAs. Nat. Commun. 11, 2697. https://doi.org/10.1038/s41467-020-16542-9.

37. Li, C., Chu, W., Gill, R.A., Sang, S., Shi, Y., Hu, X., Yang, Y., Zaman, Q.U., and Zhang, B. (2023). Computational tools and resources for CRISPR/Cas genome editing. Dev. Reprod. Biol. 21, 108–126. https://doi.org/10.1016/j.gpb.2022.02.006.

38. Yan, J., Xue, D., Chuai, G., Gao, Y., Zhang, G., and Liu, Q. (2020). Benchmarking and integrating genome-wide CRISPR off-target detection and prediction. Nucleic Acids Res. 48, 11370–11379. https://doi.org/10.1093/nar/gkaa930.

39. Yaish, O., Asif, M., and Orenstein, Y. (2022). A systematic evaluation of data processing and problem formulation of CRISPR off-target site prediction. Briefings Bioinf. 23, bbac157. https://doi.org/10.1093/bib/bbac157.

40. Nowrouzi, A., Penaud-Budloo, M., Kaeppel, C., Appelt, U., Le Guiner, C., Moullier, P., von Kalle, C., Snyder, R.O., and Schmidt, M. (2012). Integration Frequency and Intermolecular Recombination of rAAV Vectors in Non-human Primate Skeletal Muscle and Liver. Mol. Ther. 20, 1177–1186. https://doi.org/10.1038/mt.2012.47.

41. Yang, J., Zhou, W., Zhang, Y., Zidon, T., Ritchie, T., and Engelhardt, J.F. (1999). Concatamerization of Adeno-Associated Virus Circular Genomes Occurs through Intermolecular Recombination. J. Virol. 73, 9468–9477.

42. Nguyen, G.N., Everett, J.K., Kafle, S., Roche, A.M., Raymond, H.E., Leiby, J., Wood, C., Assenmacher, C.-A., Merricks, E.P., Long, C.T., et al. (2021). A long-term study of AAV gene therapy in hemophilia A dogs identifies clonal expansions of transduced liver cells. Nat. Biotechnol. 39, 47–55. https://doi.org/10.1038/s41587-020-0741-7.

43. Deyle, D.R., and Russell, D.W. (2009). Adeno-associated virus vector integration. Curr. Opin. Mol. Therapeut. 11, 442–447.

44. Inagaki, K., Piao, C., Kotchey, N.M., Wu, X., and Nakai, H. (2008). Frequency and spectrum of genomic integration of recombinant adeno-associated virus serotype 8 vector in neonatal mouse liver. J. Virol. 82, 9513–9524. https://doi.org/10.1128/JVI.01001-08.

45. Hanlon, K.S., Kleinstiver, B.P., Garcia, S.P., Zaborowski, M.P., Volak, A., Spirig, S.E., Muller, A., Sousa, A.A., Tsai, S.Q., Bengtsson, N.E., et al. (2019). High levels of AAV vector integration into CRISPR-induced DNA breaks. Nat. Commun. 10, 4439. https://doi.org/10.1038/s41467-019-12449-2.

46. Nair, N., Rincon, M.Y., Evens, H., Sarcar, S., Dastidar, S., Samara-Kuko, E., Ghandeharian, O., Man Viecelli, H., Thöny, B., De Bleser, P., et al. (2014). Computationally designed liver-specific transcriptional modules and hyperactive factor IX improve hepatic gene therapy. Blood 123, 3195–3199. https://doi.org/10.1182/blood-2013-10-534032.

47. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., and Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. 37, 224–226. https://doi.org/10.1038/s41587-019-0032-3.

48. Peng, H., Zheng, Y., Zhao, Z., Liu, T., and Li, J. (2018). Recognition of CRISPR/Cas9 off-target sites through ensemble learning of uneven mismatch distributions. Bioinformatics 34, i757–i765. https://doi.org/10.1093/bioinformatics/bty558.

49. Gil-Farina, I., Fronza, R., Kaeppel, C., Lopez-Franco, E., Ferreira, V., D'Avola, D., Benito, A., Prieto, J., Petry, H., Gonzalez-Aseguinolaza, G., and Schmidt, M.

(2016). Recombinant AAV Integration Is Not Associated With Hepatic Genotoxicity in Nonhuman Primates and Patients. Mol. Ther. *24*, 1100–1105. https://doi.org/10.1038/mt.2016.52.

50. Kotin, R.M., Linden, R.M., and Berns, K.I. (1992). Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J. *11*, 5071–5078. https://doi.org/10.1002/j.1460-2075.1992.tb05614.x.

51. Fronza, R., Vasciaveo, A., Benso, A., and Schmidt, M. (2016). A Graph Based Framework to Model Virus Integration Sites. Comput. Struct. Biotechnol. J. *14*, 69–77. https://doi.org/10.1016/j.csbj.2015.10.006.

52. Nakai, H., Montini, E., Fuess, S., Storm, T.A., Grompe, M., and Kay, M.A. (2003). AAV serotype 2 vectors preferentially integrate into active genes in mice. Nat. Genet. *34*, 297–302. https://doi.org/10.1038/ng1179.

53. Tran, N.T., Heiner, C., Weber, K., Weiand, M., Wilmot, D., Xie, J., Wang, D., Brown, A., Manokaran, S., Su, Q., et al. (2020). AAV-Genome Population Sequencing of Vectors Packaging CRISPR Components Reveals Design-Influenced Heterogeneity. Mol. Ther. Methods Clin. Dev. *18*, 639–651. https://doi.org/10.1016/j.omtm.2020.07.007.

54. Vertex Pharmaceuticals Incorporated (2023). A Phase 3b Study to Evaluate Efficacy and Safety of a Single Dose of Autologous CRISPR Cas9 Modified CD34+ Human Hematopoietic Stem and Progenitor Cells (CTX001) in Subjects With Transfusion-Dependent β-Thalassemia or Severe Sickle Cell Disease (clinicaltrials.gov).

55. Musunuru, K., Chadwick, A.C., Mizoguchi, T., Garcia, S.P., DeNizio, J.E., Reiss, C.W., Wang, K., Iyer, S., Dutta, C., Clendaniel, V., et al. (2021). In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. Nature *593*, 429–434. https://doi.org/10.1038/s41586-021-03534-y.

56. Dong, Y., Li, H., Zhao, L., Koopman, P., Zhang, F., and Huang, J.X. (2019). Genome-Wide Off-Target Analysis in CRISPR-Cas9 Modified Mice and Their Offspring. G3 (Bethesda) *9*, 3645–3651. https://doi.org/10.1534/g3.119.400503.

57. Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. *31*, 839–843. https://doi.org/10.1038/nbt.2673.

58. Atkins, A., Chung, C.-H., Allen, A.G., Dampier, W., Gurrola, T.E., Sariyer, I.K., Nonnemacher, M.R., and Wigdahl, B. (2021). Off-Target Analysis in Gene Editing and Applications for Clinical Translation of CRISPR/Cas9 in HIV-1 Therapy. Front. Genome Ed. *3*, 673022.

59. Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., Wang, Y., Brodsky, R.A., Zhang, K., Cheng, L., and Ye, Z. (2014). Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. Cell Stem Cell *15*, 12–13. https://doi.org/10.1016/j.stem.2014.06.011.

60. Veres, A., Gosis, B.S., Ding, Q., Collins, R., Ragavendran, A., Brand, H., Erdin, S., Cowan, C.A., Talkowski, M.E., and Musunuru, K. (2014). Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell *15*, 27–30. https://doi.org/10.1016/j.stem.2014.04.020.

61. Iyer, V., Shen, B., Zhang, W., Hodgkins, A., Keane, T., Huang, X., and Skarnes, W.C. (2015). Off-target mutations are rare in Cas9-modified mice. Nat. Methods *12*, 479. https://doi.org/10.1038/nmeth.3408.

62. Nutter, L.M.J., Heaney, J.D., Lloyd, K.C.K., Murray, S.A., Seavitt, J.R., Skarnes, W.C., Teboul, L., Brown, S.D.M., and Moore, M. (2018). Response to "Unexpected mutations after CRISPR-Cas9 editing *in vivo*". Nat. Methods *15*, 235–236. https://doi.org/10.1038/nmeth.4559.

63. Lareau, C.A., Clement, K., Hsu, J.Y., Pattanayak, V., Joung, J.K., Aryee, M.J., and Pinello, L. (2018). Response to "Unexpected mutations after CRISPR-Cas9 editing *in vivo*". Nat. Methods *15*, 238–239. https://doi.org/10.1038/nmeth.4541.

64. Lescarbeau, R.M., Murray, B., Barnes, T.M., and Bermingham, N. (2018). Response to "Unexpected mutations after CRISPR–Cas9 editing *in vivo*". Nat. Methods *15*, 237. https://doi.org/10.1038/nmeth.4553.

65. Wilson, C.J., Fennell, T., Bothmer, A., Maeder, M.L., Reyon, D., Cotta-Ramusino, C., Fernandez, C.A., Marco, E., Barrera, L.A., Jayaram, H., et al. (2018). Response to "Unexpected mutations after CRISPR-Cas9 editing *in vivo*". Nat. Methods *15*, 236–237. https://doi.org/10.1038/nmeth.4552.

66. Penaud-Budloo, M., Le Guiner, C., Nowrouzi, A., Toromanoff, A., Chérel, Y., Chenuaud, P., Schmidt, M., von Kalle, C., Rolling, F., Moullier, P., and Snyder, R.O. (2008). Adeno-Associated Virus Vector Genomes Persist as Episomal Chromatin in Primate Muscle. J. Virol. *82*, 7875–7885. https://doi.org/10.1128/JVI.00649-08.

67. Miller, D.G., Petek, L.M., and Russell, D.W. (2004). Adeno-associated virus vectors integrate at chromosome breakage sites. Nat. Genet. *36*, 767–773. https://doi.org/10.1038/ng1380.

68. Miller, D.G., Petek, L.M., and Russell, D.W. (2003). Human Gene Targeting by Adeno-Associated Virus Vectors Is Enhanced by DNA Double-Strand Breaks. Mol. Cell Biol. *23*, 3550–3557. https://doi.org/10.1128/MCB.23.10.3550-3557.2003.

69. Calabria, A., Cipriani, C., Spinozzi, G., Rudilosso, L., Esposito, S., Benedicenti, F., Albertini, A., Pouzolles, M., Luoni, M., Giannelli, S., et al. (2023). Intrathymic AAV delivery results in therapeutic site-specific integration at TCR loci in mice. Blood *141*, 2316–2329. https://doi.org/10.1182/blood.2022017378.

70. Hunt, J.M.T., Samson, C.A., Rand, A.D., and Sheppard, H.M. (2023). Unintended CRISPR-Cas9 editing outcomes: a review of the detection and prevalence of structural variants generated by gene-editing in human cells. Hum. Genet. *142*, 705–720. https://doi.org/10.1007/s00439-023-02561-1.

71. Zeng, X., Lin, D., Liang, D., Huang, J., Yi, J., Lin, D., and Zhang, Z. (2023). Gene sequencing and result analysis of balanced translocation carriers by third-generation gene sequencing technology. Sci. Rep. *13*, 7004. https://doi.org/10.1038/s41598-022-20356-8.

72. Au, C.H., Ho, D.N., Ip, B.B.K., Wan, T.S.K., Ng, M.H.L., Chiu, E.K.W., Chan, T.L., and Ma, E.S.K. (2019). Rapid detection of chromosomal translocation and precise breakpoint characterization in acute myeloid leukemia by nanopore long-read sequencing. Cancer Genet. *239*, 22–25. https://doi.org/10.1016/j.cancergen.2019.08.005.

73. Hu, L., Liang, F., Cheng, D., Zhang, Z., Yu, G., Zha, J., Wang, Y., Xia, Q., Yuan, D., Tan, Y., et al. (2019). Location of Balanced Chromosome-Translocation Breakpoints by Long-Read Sequencing on the Oxford Nanopore Platform. Front. Genet. *10*, 1313.

74. Akcakaya, P., Bobbin, M.L., Guo, J.A., Malagon-Lopez, J., Clement, K., Garcia, S.P., Fellows, M.D., Porritt, M.J., Firth, M.A., Carreras, A., et al. (2018). In vivo CRISPR editing with no detectable genome-wide off-target mutations. Nature *561*, 416–419. https://doi.org/10.1038/s41586-018-0500-9.

75. Nobles, C.L., Reddy, S., Salas-McKee, J., Liu, X., June, C.H., Melenhorst, J.J., Davis, M.M., Zhao, Y., and Bushman, F.D. (2019). iGUIDE: an improved pipeline for analyzing CRISPR cleavage specificity. Genome Biol. *20*, 14. https://doi.org/10.1186/s13059-019-1625-3.

76. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V.V., Thapar, V., Wyvekens, N., Khayter, C., Iafrate, A.J., Le, L.P., et al. (2015). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–197. https://doi.org/10.1038/nbt.3117.

77. Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.-J., and Yee, J.-K. (2015). Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nat. Biotechnol. *33*, 175–178. https://doi.org/10.1038/nbt.3127.

78. Mátrai, J., Cantore, A., Bartholomae, C.C., Annoni, A., Wang, W., Acosta-Sanchez, A., Samara-Kuko, E., De Waele, L., Ma, L., Genovese, P., et al. (2011). Hepatocyte-targeted expression by integrase-defective lentiviral vectors induces antigen-specific tolerance in mice with low genotoxic risk. Hepatology *53*, 1696–1707. https://doi.org/10.1002/hep.24230.

79. Clement, K., Hsu, J.Y., Canver, M.C., Joung, J.K., and Pinello, L. (2020). Technologies and Computational Analysis Strategies for CRISPR Applications. Mol. Cell *79*, 11–29. https://doi.org/10.1016/j.molcel.2020.06.012.

80. Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.-I., and Kim, J.-S. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat. Methods *12*, 237–243. https://doi.org/10.1038/nmeth.3284.

81. Tsai, S.Q., Nguyen, N.T., Malagon-Lopez, J., Topkar, V.V., Aryee, M.J., and Joung, J.K. (2017). CIRCLE-seq: a highly sensitive *in vitro* screen for genome-wide CRISPR–Cas9 nuclease off-targets. Nat. Methods *14*, 607–614. https://doi.org/10.1038/nmeth.4278.

82. Kim, D., and Kim, J.-S. (2018). DIG-seq: a genome-wide CRISPR off-target profiling method using chromatin DNA. Genome Res. *28*, 1894–1900. https://doi.org/10.1101/gr.236620.118.

83. Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., Wernersson, E., Li, Y., Gao, L., et al. (2017). BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. Nat. Commun. *8*, 15058. https://doi.org/10.1038/ncomms15058.

84. Wienert, B., Wyman, S.K., Yeh, C.D., Conklin, B.R., and Corn, J.E. (2020). CRISPR off-target detection with DISCOVER-Seq. Nat. Protoc. *15*, 1775–1799. https://doi.org/10.1038/s41596-020-0309-5.

85. Ferrari, S., Jacob, A., Cesana, D., Laugel, M., Beretta, S., Varesi, A., Unali, G., Conti, A., Canarutto, D., Albano, L., et al. (2022). Choice of template delivery mitigates the genotoxic risk and adverse impact of editing in human hematopoietic stem cells. Cell Stem Cell *29*, 1428–1444.e9. https://doi.org/10.1016/j.stem.2022.09.001.

86. Breton, C., Clark, P.M., Wang, L., Greig, J.A., and Wilson, J.M. (2020). ITR-Seq, a next-generation sequencing assay, identifies genome-wide DNA editing sites *in vivo* following adeno-associated viral vector-mediated genome editing. BMC Genom. *21*, 239. https://doi.org/10.1186/s12864-020-6655-4.

87. Dalwadi, D.A., Calabria, A., Tiyaboonchai, A., Posey, J., Naugler, W.E., Montini, E., and Grompe, M. (2021). AAV integration in human hepatocytes. Mol. Ther. *29*, 2898–2909. https://doi.org/10.1016/j.ymthe.2021.08.031.

88. Cantore, A., Nair, N., Della Valle, P., Di Matteo, M., Màtrai, J., Sanvito, F., Brombin, C., Di Serio, C., D'Angelo, A., Chuah, M., et al. (2012). Hyperfunctional coagulation factor IX improves the efficacy of gene therapy in hemophilic mice. Blood *120*, 4517–4520. https://doi.org/10.1182/blood-2012-05-432591.

89. Milani, M., Annoni, A., Moalli, F., Liu, T., Cesana, D., Calabria, A., Bartolaccini, S., Biffi, M., Russo, F., Visigalli, I., et al. (2019). Phagocytosis-Shielded Lentiviral Vectors Improve Liver Gene Therapy in Non Human Primates. Sci. Transl. Med. *11*, eaav7325. https://doi.org/10.1126/scitranslmed.aav7325.

90. Ranzani, M., Cesana, D., Bartholomae, C.C., Sanvito, F., Pala, M., Benedicenti, F., Gallina, P., Sergi, L.S., Merella, S., Bulfone, A., et al. (2013). Lentiviral vector-based insertional mutagenesis identifies genes associated with liver cancer. Nat. Methods *10*, 155–161. https://doi.org/10.1038/nmeth.2331.

91. Chandler, R.J., LaFave, M.C., Varshney, G.K., Trivedi, N.S., Carrillo-Carrasco, N., Senac, J.S., Wu, W., Hoffmann, V., Elkahloun, A.G., Burgess, S.M., and Venditti, C.P. (2015). Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. J. Clin. Invest. *125*, 870–880. https://doi.org/10.1172/JCI79213.

92. Afzal, S., Wilkening, S., von Kalle, C., Schmidt, M., and Fronza, R. (2017). GENE-IS: Time-Efficient and Accurate Analysis of Viral Integration Events in Large-Scale Gene Therapy Data. Mol. Ther. Nucleic Acids *6*, 133–139. https://doi.org/10.1016/j.omtn.2016.12.001.

93. Ghodsi, M., Liu, B., and Pop, M. (2011). DNACLUST: accurate and efficient clustering of phylogenetic marker genes. BMC Bioinf. *12*, 271. https://doi.org/10.1186/1471-2105-12-271.

94. Kent, W.J. (2002). BLAT–the BLAST-like alignment tool. Genome Res. *12*, 656–664. https://doi.org/10.1101/gr.229202.

95. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340.

96. Schneider, T.D., and Stephens, R.M. (1990). Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. *18*, 6097–6100. https://doi.org/10.1093/nar/18.20.6097.

97. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109.

98. Shen, H., Suzuki, T., Munroe, D.J., Stewart, C., Rasmussen, L., Gilbert, D.J., Jenkins, N.A., and Copeland, N.G. (2003). Common Sites of Retroviral Integration in Mouse Hematopoietic Tumors Identified by High-Throughput, Single Nucleotide Polymorphism-Based Mapping and Bacterial Artificial Chromosome Hybridization. J. Virol. *77*, 1584–1588. https://doi.org/10.1128/jvi.77.2.1584-1588.2003.

99. Abel, U., Deichmann, A., Nowrouzi, A., Gabriel, R., Bartholomae, C.C., Glimm, H., von Kalle, C., and Schmidt, M. (2011). Analyzing the Number of Common Integration Sites of Viral Vectors – New Methods and Computer Programs. PLoS One *6*, e24247. https://doi.org/10.1371/journal.pone.0024247.