Molecular Therapy
Nucleic Acids
Original Article

Precision and efficacy of RNA-guided DNA integration in high-expressing muscle loci

Made Harumi Padmaswari,[1,2] Gabrielle Bulliard,[1] Shilpi Agrawal,[1] Mary S. Jia,[1] Sabin Khadgi,[3] Kevin A. Murach,[2,3] and Christopher E. Nelson[1,2]

[1]Biomedical Engineering, University of Arkansas, Fayetteville, AR, USA; [2]Cellular and Molecular Biology, University of Arkansas, Fayetteville, AR, USA; [3]Exercise Science Research Center, Molecular Muscle Mass Regulation Laboratory, Department of Health, Human Performance, and Recreation, University of Arkansas, Fayetteville, AR, USA

Gene replacement therapies primarily rely on adeno-associated virus (AAV) vectors for transgene expression. However, episomal expression can decline over time due to vector loss or epigenetic silencing. CRISPR-based integration methods offer promise for long-term transgene insertion. While the development of transgene integration methods has made substantial progress, identifying optimal insertion loci remains challenging. Skeletal muscle is a promising tissue for gene replacement owing to low invasiveness of intramuscular injections, relative proportion of body mass, the multinucleated nature of muscle, and the potential for reduced adverse effects. Leveraging endogenous promoters in skeletal muscle, we evaluated two highly expressing loci using homology-independent targeted integration (HITI) to integrate reporter or therapeutic genes in mouse myoblasts and skeletal muscle tissue. We hijacked the muscle creatine kinase (*Ckm*) and myoglobin (*Mb*) promoters by co-delivering CRISPR-Cas9 and a donor plasmid with promoterless constructs encoding green fluorescent protein (GFP) or human Factor IX (hFIX). Additionally, we deeply profiled our genome and transcriptome outcomes from targeted integration and evaluated the safety of the proposed sites. This study introduces a proof-of-concept technology for achieving high-level therapeutic gene expression in skeletal muscle, with potential applications in targeted integration-based medicine and synthetic biology.

INTRODUCTION

The current wave of therapies transforming genetic medicines relies on gene replacement.[1] Efficient transgene expression is predominantly achieved through the use of adeno-associated virus (AAV) vectors to express a therapeutic gene under an exogenous promoter.[2] While multiple studies have reported the sustainability of AAV gene therapy for years after a single administration, others have noted a decline in transgene expression, with some even observing a return to baseline levels.[3–6] Although the majority of transgene expression is sustained episomally, research indicates that epigenetic silencing of episomal DNA and interactions between the AAV capsid and host factors significantly contribute to the reduction in expression.[7] In line with these findings, recent studies in dogs and non-human primates revealed initially high but short-lived expression from episomal genomes, while long-term expression of the transgene was attributed to clonal expansion of cells containing integrated vectors.[8,9] However, transgene integration in AAV gene replacement happens in a semi-random pattern throughout the genome.[10] Vector integration in unpredictable positions in the genome might result in unforeseeable expression and interactions with the host genome, though no adverse events related to vector integration have been detected in humans.[11,12] Thus, these limitations highlight the need to transition from random integration with viral vectors to targeted, site-specific methods like CRISPR-Cas technology, which can overcome the problem of insertion of correct genes into random genomic sites.

CRISPR-based integration methods have evolved significantly, from using the endogenous repair pathway, non-homologous end joining (NHEJ), or homology-directed repair to engineered transposable elements.[13–15] These gene editing technologies emerged as a potential avenue for long-term therapeutic transgene insertion. Despite considerable advancements in targeted gene integration methods, identifying the locus to insert replacement genes for optimal safety and efficiency remains challenging.[16] While many gene editing approaches target the disease locus itself, some studies indicate that the genomic structure of mutated genes may not favor entire gene replacement. Further, some exogenous promoters, such as cytomegalovirus (CMV) and human elongation factor-1 alpha, are susceptible to promoter silencing.[17,18] Therefore, exploring loci with high transcriptional activity emerges as an intriguing alternative.

The current strategy of "hijacking" the endogenous promoter of high-transcriptional-activity genes or transgene overexpression remains largely organ specific.[19–21] While the results are promising, the safety and versatility of this approach have yet to undergo thorough investigation. Ensuring the safety and versatility of these sites is essential for achieving stable expression of integrated transgenes without adversely affecting the host cell. Empirical studies have identified

Received 22 March 2024; accepted 30 August 2024;
https://doi.org/10.1016/j.omtn.2024.102320.

Correspondence: Christopher E. Nelson, PhD, Department of Biomedical Engineering, University of Arkansas, 120 John A. White Jr. Engineering Hall, Fayetteville, AR 72701, USA.
E-mail: nelsonc@uark.edu

Table 1. Comparison of highly expressed genes in skeletal muscle tissues between mice and humans

		Human		Mouse	
	Gene	RPKM	SE	RPKM	SE
1	*ACTA1*	7.03	0.27	7.14	0.14
2	*CKM*	6.91	0.24	5.65	0.60
3	*MB*	6.63	0.46	6.14	0.49
4	*MYH7*	6.72	0.33	2.26	2.80
5	*MYL2*	6.87	0.26	3.35	2.52
6	*DES*	6.11	0.44	3.35	1.68
7	*TNNC2*	6.28	0.54	6.93	0.46
8	*TNNI*	6.31	0.43	1.93	2.73
9	*MYL3*	5.79	0.24	3.46	2.23
10	*TTN*	6.11	0.70	0.63	4.52

The data were extracted from Abdelmoez et al.,[31] who compiled skeletal muscle tissue gene expression from the Gene Expression Omnibus public database. RPKM, reads per kilobase per million mapped reads; SE, standard error.

safe-harbor sites that support long-term transgene expression. There are three well-established sites—*AAVS1*, *CCR5*, and *hROSA26*—and recently explored sites such as *Rogi1*, *Rogi2*, and *SHS231*.[22–24] However, the expression from these sites is limited or tissue specific and requires the inserted vector to include an exogenous promoter.

In this study, we identified new potential safe-harbor sites in skeletal muscle that offer secure and stable integration, facilitating gene replacement. Choosing skeletal muscle as a gene therapy integration site can reduce procedure invasiveness and complexity. Its unique syncytial nature is less likely to create negative effects *in vivo*, with minimal local or systemic adverse effects associated with intramuscular injection reported in a systematic review of AAV gene therapies.[2] Owing to its local delivery, the exposure to circulating neutralizing antibodies, such as AAV pre-existing antibodies, against the viral capsid is minimized, ensuring efficient tissue transduction.[25,26] The well-vascularized nature of skeletal muscle allows secreted proteins to enter the bloodstream. Previous studies have demonstrated that intramuscular administration of an AAV vector can lead to the secretion of functional proteins.[27–29]

In this study, we devised a strategy to leverage endogenous promoters in muscle cells for expressing either the reporter gene or a therapeutic gene. We adapted the currently established safe-harbor characterization to evaluate two highly expressing skeletal muscle loci. We applied homology-independent targeted integration (HITI) to mediate transgene integration,[30] introducing a reporter gene or therapeutic gene (human Factor 9 [*hF9*]) to express human Factor IX protein (hFIX) from the identified loci. By delivering a promoterless transgene, we were able to detect both protein expression in the reporter gene and hFIX. We deeply profiled genome and transcriptome outcomes after gene editing using high-throughput sequencing, unidirectional sequencing, and nanopore long-read sequencing. Overall, the data support RNA-guided DNA integration strategies as effective therapies for restoring desired gene expression in muscle and can be extended to synthetic biology applications.

RESULTS

Selected muscle-specific integration sites have high expression and a predicted favorable safety profile

The selection of integration sites involves combining established criteria for potential safe-harbor sites with insights from prior *in vivo* studies that employed the highly expressing locus albumin to express therapeutic genes in the liver.[20,22] We tailored this strategy to target integration sites within skeletal muscle. To pinpoint suitable sites, we referred to a comprehensive analysis of public transcriptomics studies on skeletal muscle in humans and mice by Abdelmoez et al.[31] Through their thorough examination and the application of rigorous quality control, normalization, and annotation using official human gene names, we identified the top 10 genes specifically expressed in skeletal muscle and demonstrated high expression levels in both humans and mice (Table 1). The selection of these genes is rooted in the goal of ensuring robust and consistent muscle-specific promoter expression in different models of organisms. Next, we evaluated the essentiality of each gene and the safety implications associated with hijacking its endogenous promoter.

Skeletal actin (*ACTA1*) is at the top of the list as the highest expressed gene in skeletal muscle. However, skeletal actin is the main actin isoform in skeletal muscle, and it plays an essential role, alongside myosin, in facilitating muscle contraction.[32] Further, *ACTA1* disruption in newborn mice causes early demise.[33] In light of this, we propose creatine kinase (*CKM*) and myoglobin (*MB*) as genomic integration sites for skeletal muscle. Mice lacking *Ckm* are viable and exhibit no changes in absolute muscle force, but they show an inability to engage in burst activity.[34,35] Similarly, mice without myoglobin are also viable, with no distinct phenotype apart from depigmentation of the cardiac muscle and cardiac adaptations regarding oxygen delivery.[36,37] Since our approach targets intronic regions and does not involve knocking out the *Ckm* or *Mb* gene, submaximal levels of these genes are adequate to maintain functionality.

To identify whether these sites could serve as potential genomic safe-harbor sites, we assessed the suitability of creatine kinase and myoglobin as safe-harbor sites in comparison to currently known safe-harbor sites. Scores were assigned based on previously established and widely accepted criteria (Figure 1; Table 2).[22] Table 2 indicates that our proposed sites did not meet all of the criteria. However, it is worth noting that even the most commonly used safe-harbor sites fail to achieve 100% clearance based on these criteria. Interestingly, *MB* shows better results in a free oncogene region criterion compared to *CKM* and other sites.

Targeted integration into *Ckm* and *Mb* leads to the expression of a promoterless reporter

To experimentally assess transgene expression from proposed sites (intron 1 of *CKM* or *MB*), we performed a targeted integration approach to knock in a gene construct encoding a promoterless green

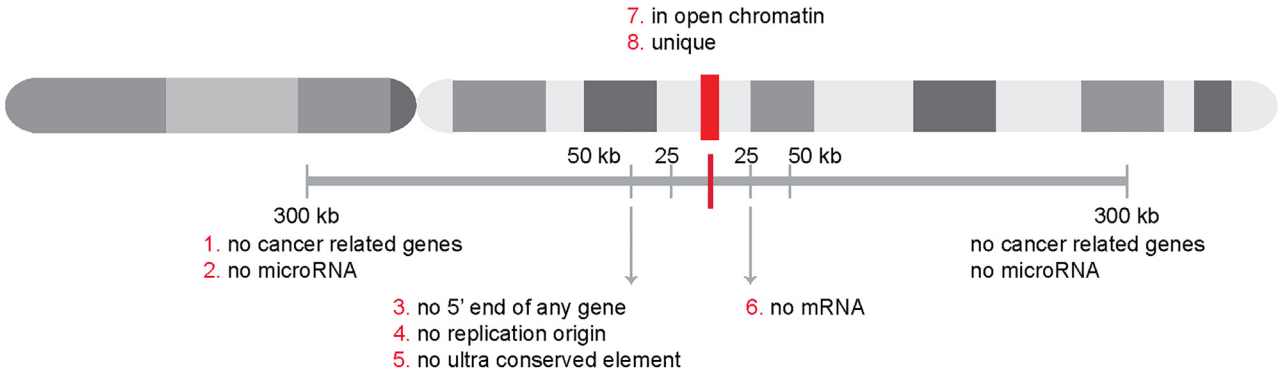


Figure 1. Schematic of ideal safe-harbor sites

The red bar in the middle represents the optimal location of safe-harbor sites. The site should be unique and located in open chromatin.[38,39] Within 300 kb, the site should be free from any cancer-related genes, microRNAs, or other functional small RNAs.[40] Within 50 kb, the site should be free from the 5′ end of a gene, replication origin, and ultra-conserved element.[41–43] Additionally, the site should be in a region of low transcriptional activity or have no mRNA within 25 kb.

fluorescent protein (GFP) into mouse myoblasts (*Ckm* or *Mb*). We identified the knockin site within the intron preceding the coding sequence of the gene, ensuring that the non-edited or inaccurately edited copy maintains functional gene expression via mRNA splicing. The rationale behind selecting the proximal upstream region of exon 2 for designing guide RNAs, despite potential splicing disruptions, is multifaceted. Neural network promoter prediction predictions for eukaryotes indicated that the promoter region for *Mb* extends throughout intron 1, whereas for *Ckm*, the promoter is more centrally located in intron 1, with an enhancer element situated in distal intron 1.[44,45]

Targeting a region close to the gene's coding sequence was essential to effectively capture the entire promoter region and potential enhancer elements. Additionally, this site selection allows for flexibility in transgene incorporation, accommodating features like signaling peptides at the N terminus and skipping peptides at the C terminus to preserve endogenous gene expression.

To integrate GFP, we used a CRISPR-Cas9-based genome editing strategy that uses the HITI method capitalizing on the NHEJ pathway that is accessible in muscle cells (Figure 2A). This approach designs Cas9 target sites in the donor DNA as reverse complements of the genomic target site to facilitate the re-cleavage of reverse integration products, providing a means to dictate the specific directionality of the knockin. We applied this strategy in C2C12 immortalized mouse myoblast cell lines, as it is the most commonly used cellular model to study murine skeletal muscle *in vitro*.[31,46]

As a proof of concept, we identified five guide RNA (gRNA) sites within intron 1 of each gene, positioned distal to the enhancer and proximal to exon 2, specifically within the last 250 bp leading up to the noncoding portion of exon 2 (Figure 2B; Table S1) based on the web tool for gRNA screening (CRISPOR) and screened their cutting efficiency in the NIH 3T3 cell line[47] (Figure S1). Additionally, we conducted PCR-enriched next-generation sequencing (NGS) using Iseq short-read sequencing to quantify the efficiency of gRNAs in the C2C12 cell line. Within the *Ckm* gene, g1 exhibited the highest efficiency, making it the selected gRNA for HITI-mediated GFP integration experiments. In the *Mb* gene, although g3 demonstrated slightly higher cutting efficiency, it targeted the non-coding exon sequence. Therefore, we opted for g1 as the gRNA for the downstream integration experiment (Figures 2B and 2C).

We used Lipofectamine 3000 to co-transfect C2C12 muscle myoblasts with a ratio of 2:2:1 for SpCas9, gRNA, and site-specific promoterless insert plasmids, respectively. Three days after transfection, we performed a genotyping PCR to confirm the integration. PCR primers were used spanning the intron of *Ckm* or *Mb* to the 5′ junction of the GFP insert and from the 3′ junction of the GFP insert to the intron of *Ckm* or *Mb*. We detected bands at the expected length of chimeric *Ckm-GFP* or *Mb*-GFP integration in both 5′ and 3′ end junctions at the genomic DNA and cDNA levels, but no bands were observed in the control groups (Figures 2D and 2E). In myotubes, after differentiation, we observed GFP expression in the CRISPR-treated group at both loci. In the Scr-GFP group, Cas9 is paired with a scrambled (scr) gRNA and a HITI insert. Under these conditions, GFP expression should be entirely suppressed unless the GFP gene is integrated into the genome in the correct orientation. However, we observed low-level or leaky expression of the promoterless donor GFP, as seen in Figure 2F. This leaky expression of promoterless GFP has also been observed previously *in vivo*.[48] This phenomenon could be attributed to the intrinsic properties of the donor DNA, leading to minimal GFP expression even without integration.

To determine whether the GFP expression we observed was a result of integration or leakage, we conducted flow cytometry on C2C12 cells 48 h after transfection. The analysis from flow cytometry showed that the group treated with scrambled gRNA had about 0.5% of cells expressing GFP. In contrast, the groups treated with *Ckm* and *Mb* had 6% and 10% GFP-positive cells, respectively (the sample gating process is shown in Figure S2). The significant increase in

Table 2. Assessment of currently widely used safe-harbor sites (*AAVS1*, *CCR5*, and *hROSA26*) and proposed safe-harbor sites

	Criteria							
Location	1	2	3	4	5	6	7	8
AAVS1	no	yes	no	yes	yes	no	yes	yes
CCR5	no	yes	no	yes	yes	no	yes	yes
hROSA26	no	yes	no	yes	yes	no	yes	yes
CKM intron	no	yes	no	yes	yes	no	yes	yes
MB intron	yes	yes	no	yes	yes	no	yes	yes

The assessment was performed based on the ideal criteria of safe-harbor sites (Figure 1). Each criterion was evaluated in the UCSC genome browser track following recommendations from Pellenz et al.[22]

GFP-positive cells in the *Ckm*- and *Mb*-treated groups suggests that the GFP expression is mainly due to the integration of the GFP gene into the host gene. These findings support the notion that the observed GFP expression in the cells is mainly a consequence of transgene expression driven by the *Ckm* or *Mb* promoter rather than leakage expression (Figure 2G).

Targeted integration of a promoterless human F9 gene at *Ckm* and *Mb* leads to sustained expression of *hF9*

Next, we sought to investigate whether this strategy applies to therapeutic genes. The GFP transgene in the HITI insert plasmid was substituted with a transgene encoding *hF9* (Figure 3A). The selection of *hF9* as a proof of concept for therapeutic gene integration was motivated by previous reports indicating that even with <1% of targeted integration events of *hF9* under the albumin promoter, it proved adequate to attain 5%–20% of FIX levels, effectively correcting bleeding in hemophilia B mice.[21] In this approach, we utilized the complete coding sequence of *hF9*, given the absence of signaling peptides in the *Ckm* or *Mb* gene. Following a method similar to GFP transfection, we co-delivered a site-specific *hF9* HITI insert plasmid with SpCas9 and *Ckm* or *Mb* gRNA into C2C12 myoblasts using Lipofectamine 3000. Post transfection, as observed in GFP integration, we verified *hF9* integration exclusively within the CRISPR-mediated HITI group at both genomic DNA and cDNA levels in both sites (Figures 3B and 3C).

To evaluate changes in gene expression between groups, we conducted RT-qPCR using Fluorescein amidite (FAM) probe primers spanning from exon 6 to exon 7 of *hF9*. The *Ppia* gene served as a housekeeping gene for normalizing expression levels. This selection was based on our data that showed that the *Ppia* gene has the most consistent expression in both myoblasts and myotubes. In contrast, *Gapdh* has variable expression levels between the myoblast and myotube stages. Furthermore, we evaluated four different reference genes from a previous study to identify the most consistent expression in both myoblasts and myotubes[49] (Table S2).

During the myotube stage, the relative expression of *hF9* was found to be elevated in the CRISPR-mediated HITI group compared to the scrambled group, reaching up to a 3-fold increase in *Ckm*-mediated expression and up to a 10-fold increase in *Mb*-mediated expression (Figure 3D). The scrambled group with promoter-less *hF9* demonstrated some level of *hF9* expression without integration. While this phenomenon might be due to the leaky expression of the promoterless insert, similar to what we observed with GFP expression, another reason could be residual donor plasmid lingering in the cells and being extracted in RNA. To address potential DNA contamination, we performed a rigorous two-step DNA removal process using a genomic DNA (gDNA) eliminator column followed by DNase treatment. Despite these efforts, we still detected background expression of *hF9* in qPCR. To ensure the accuracy of our analysis, we normalized the qPCR results against the scrambled group to account for this background expression. This normalization allows us to distinguish between actual integration events and residual plasmid expression, providing a more accurate representation of targeted *hF9* expression levels.

Furthermore, to determine whether identified sites supported long-term stable *hF9* expression, we maintained the myoblast culture for 30 days. A 30-day culture time point was chosen to balance the need for repassaging the cells every other day to prevent spontaneous myotube differentiation and the limited lifespan of C2C12 cells, which can only be passaged 10–50 times before senescence.[50] Despite our efforts to incorporate a reporter gene marker in Cas9 for cell sorting, we encountered difficulties in maintaining the sorted cultured cells. This obstacle has also been noted in previously published literature, which indicates that myoblasts tend to lose their differentiation potential after single-cell cloning, and those that are successfully edited often fail to survive the stress of sorting.[51] Since generating clonal populations was not feasible, we opted to gather informative data from bulk populations by routinely passaging the cells when they reached 60%–70% confluency before spontaneously differentiating into myotubes.

Following our 30-day culture, we detected integration in genomic DNA (Figure S3). In the time course analysis, *hF9* gene expression was assessed at various intervals starting from 3 days post transfection, with cell passaging every 2–3 days, followed by differentiation from days 23–30. Despite the continuous passaging of cells, this longitudinal study demonstrated generally consistent expression of *hF9* throughout the observation period. In Figure 3D, cells were treated in the myoblast stage and then differentiated into myotubes, whereas in Figure 3E, cells were treated in the myoblast stage and repassaged every other day to maintain myoblast characteristics. Previous qPCR data have indicated varying expression levels of *Ckm* and *Mb* in myoblasts and myotubes of C2C12 cells, with *Mb* expression rising post differentiation (Figure S4). This condition might contribute to the observed differences in the strength of the promoter to express the transgene. Therefore, the differences seen in Figure 3D may not be transient but, rather, a consequence of the different experimental conditions used in Figures 3D and 3E.

To provide a better understanding of the extent of *hF9* expression over time, we also performed absolute quantification of hF9 expression through an external standard (linearized plasmid DNA).

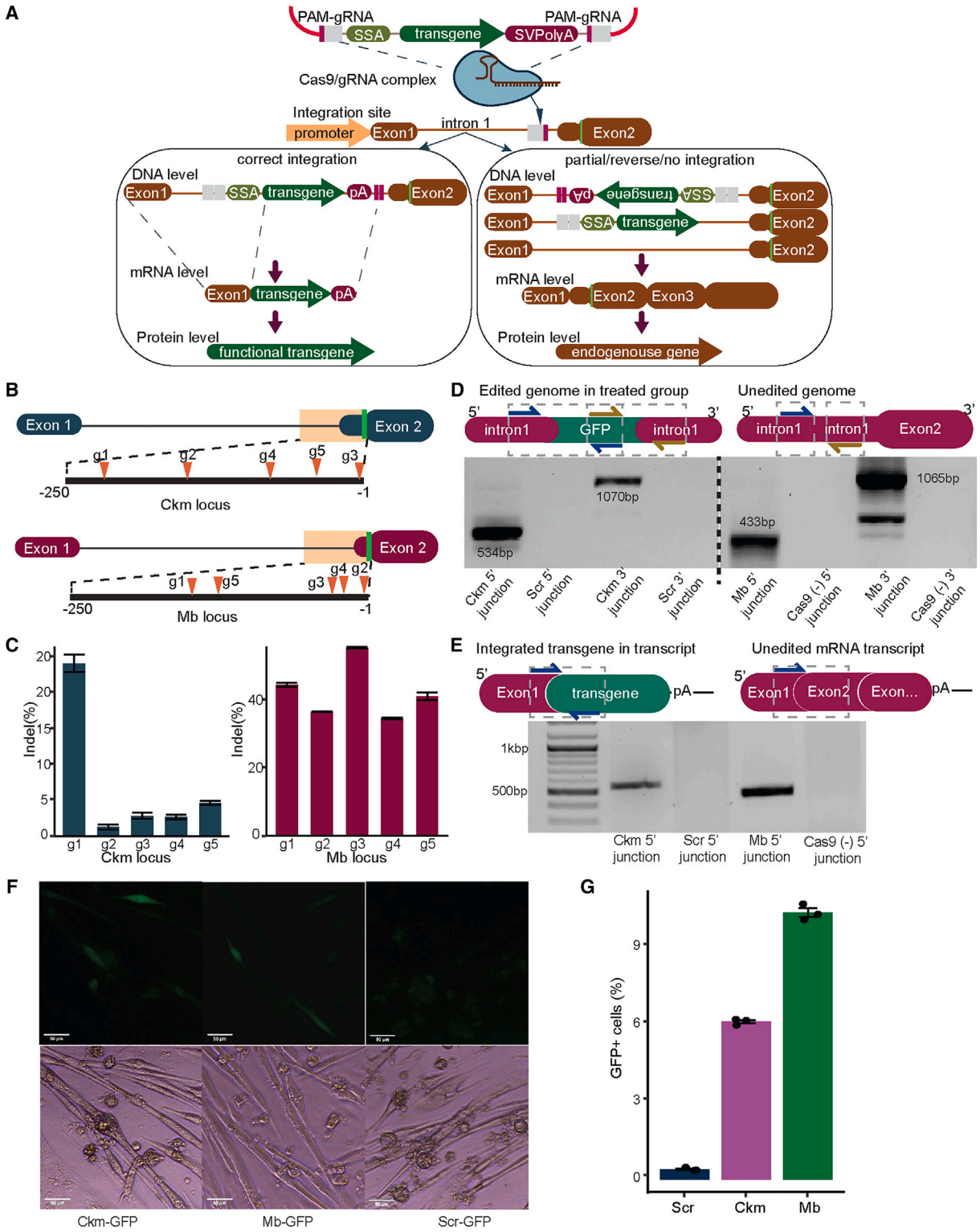


Figure 2. Targeted integration of a promoterless GFP at Mb and Ckm leads to GFP expression in myotubes

(A) Top: overview of transgene integration to hijack endogenous *Ckm* or *Mb* promoters. The HITI-based gene editing system is delivered via plasmid transfection, which encodes the CMV-driven *S. pyogenes* Cas9 protein, U6 promoter-gRNA expression cassette, and a donor plasmid containing a transgene fragment flanked by Cas9 target

(legend continued on next page)

Following the PREXCEL-Q method (previously referred to as "FocusField2-6Gallup-qPCRSet-upTool-001"),[52] we calculated the copy number estimate based on the cycle cumber (Cq) values of the samples. Since our standard was made from double-stranded DNA, we divided the copy number results by 2 to adjust for the single-stranded condition of cDNA (Table S4). We then normalized the copy number per nanogram of mRNA. In a typical mammalian cell, mRNA constitutes 4% of the total RNA mass.[53] Given that each reaction used 50 ng of total RNA converted into cDNA, 4% of 50 ng is 2 ng. Therefore, the final copy number estimate after Cq value quantification was divided by 2 ng to obtain the copy number per nanogram of mRNA.

Based on the time course copy number assessment, we observed that, at the initial time points, the copy number of the transgene was higher, even in the scrambled samples, and gradually decreased over time. While the magnitude of the fold change in copy number differs from the magnitude of fold change in relative expression due to some variance in the reference gene, in general, the copy number of treated samples remained higher compared to scrambled samples until day 30 (Figure S5).

Additionally, to assess whether this increased mRNA expression translates to hFIX protein expression, we performed an hFIX sandwich ELISA on cell media after myotube differentiation. We observed an increased level of protein expression in both *Ckm*- and *Mb*-treated cell serum relative to the scrambled control (Figure S6).

Multipronged targeted NGS and long-read sequencing show relative precision of HITI-based integration

While we have confirmed that the *Ckm* and *Mb* endogenous promoter can express the promoterless protein, the precision of editing remains unknown. Imprecise integration could negatively impact protein expression; thus, we need to assess several factors affecting editing efficiency. To begin with, we examined the off-target gRNA editing at the selected sites. Using Cas-OFFinder, we identified 11 predicted off-target sites for the *Ckm* gRNA and 10 predicted off-target sites for the *Mb* gRNA *in vitro* (Tables S5 and S6).[54] These sites were then analyzed through deep amplicon sequencing. The results showed no insertion or deletion (indel) formation beyond background levels, indicating minimal or no off-target editing (Figure S7). Next, we evaluated factors affecting protein expression, such as the precision and the efficiency of integration. We performed PCR-enriched short-read sequencing for fragments spanning the genomic insertion site or chimeric transcript to evaluate whether the integrated chimeric sequence is as precise based on the mechanism of HITI or whether unintentional on-target events could be detected. Genomic DNA-level analysis revealed levels of precise integration ranging from 15% to 35% in 5′ and 3′ integration junctions in both loci (Figures 4A and 4B). The modified sequence predominantly resulted from NHEJ repair at gRNA cut sites, with a preference for removing portions of the inserted sequence (Figures 4A and 4B). Although genomic DNA-level precision was not optimal, we saw a higher average of precision integration at the mRNA level, with around 80% of precise integration (Figure 4C) with some modifications on the junction site of exon 1 *Ckm* and start codon GFP (Figure 4D).

Precision integration data provide a selective perspective on how effectively this method incorporates the transgene insert. However, the selection of primers and PCR will bias the result. Moreover, this method does not quantify the proportion of integrated sequences due to the PCR enrichment method that selectively amplifies the integrated sequences. To quantify the integration efficiency, we used a uni-directional targeted sequencing (UDiTaS) methodology that is based on Tn5 transposase-assisted tagmentation short-read sequencing. This method incorporates a unique molecular identifier (UMI) to remove PCR amplification bias. We quantified the number of sequences integrated with our transgene by tagging primers specific to the *Ckm* intron before the gRNA cut site (Figure 4E). Reading from this site will capture unedited, precisely edited, and unintentionally modified alleles. In Figure 4E, we deduplicated the UMI and aligned the resulting BAM files to the expected fusion of *Ckm* intron-GFP integration. After implementing filters (see details in STAR Methods) to verify that the plotted reads were authentic and not sequencing artifacts, the analysis revealed that correct GFP integration represented 2.8% of the total aligned reads. This percentage was calculated by dividing the proportion of edited reads by the proportion of unedited reads that aligned with the reference amplicon from the total reads (4%). Separately, a primer specific to the GFP insert genome was used in conjunction with the same transposon-specific primer to map genome-wide GFP episome integration into the mouse genome. Following a similar pipeline as with UdiTaS analysis, we did not detect GFP integration in other locations in the genome (Figure S8).

The cDNA-selective PCR-enriched sequencing method revealed a high percentage of precise integration. However, it may overlook large structural variants. Therefore, to get a more comprehensive

sites in the reverse orientation relative to genomic DNA. Bottom: the mechanism at the molecular level shows that, by integrating the transgene in the intronic region, only correct integration is expected to produce the protein, while partial or reverse integration is expected to not affect endogenous genes. The figure is not drawn to scale. SSA, strong splice acceptor; PAM, protospacer adjacent motif; gRNA, guide RNA; SVpoly(A), simian virus polyadenylation signal; pA, polyadenylation signal. (B) The gRNA target site map in *Ckm* and *Mb* locus. Five gRNA target sites were selected within 250 bp in the splice acceptor and 5′ UTR of the gene. (C) The quantification of gRNA efficiency in the C2C12 cell line with short-read next-generation sequencing (NGS) based on indel percentage. The error bars were computed as the standard deviation obtained through the bootstrap resampling technique. (D) Validation of correct GFP integration in genomic DNA in both 5′ and 3′ integration regions in the treated group. The schematic shows that the forward primer in the GFP transgene overlaps with the reverse primer (primer list). (E) Validation of correct mRNA splicing from GFP integration by cDNA PCR in both loci. At the transcriptome level, only 5′ integration was assessed. (F) Fluorescence microscopy images assessing GFP expression in edited myotubes 10 days after editing; the scale bar represents 50 μm. (G) Flow cytometry analysis of GFP-positive cells 48 h after treatment in the C2C12 cell line. Three independent biological replicates; means ± SEM.

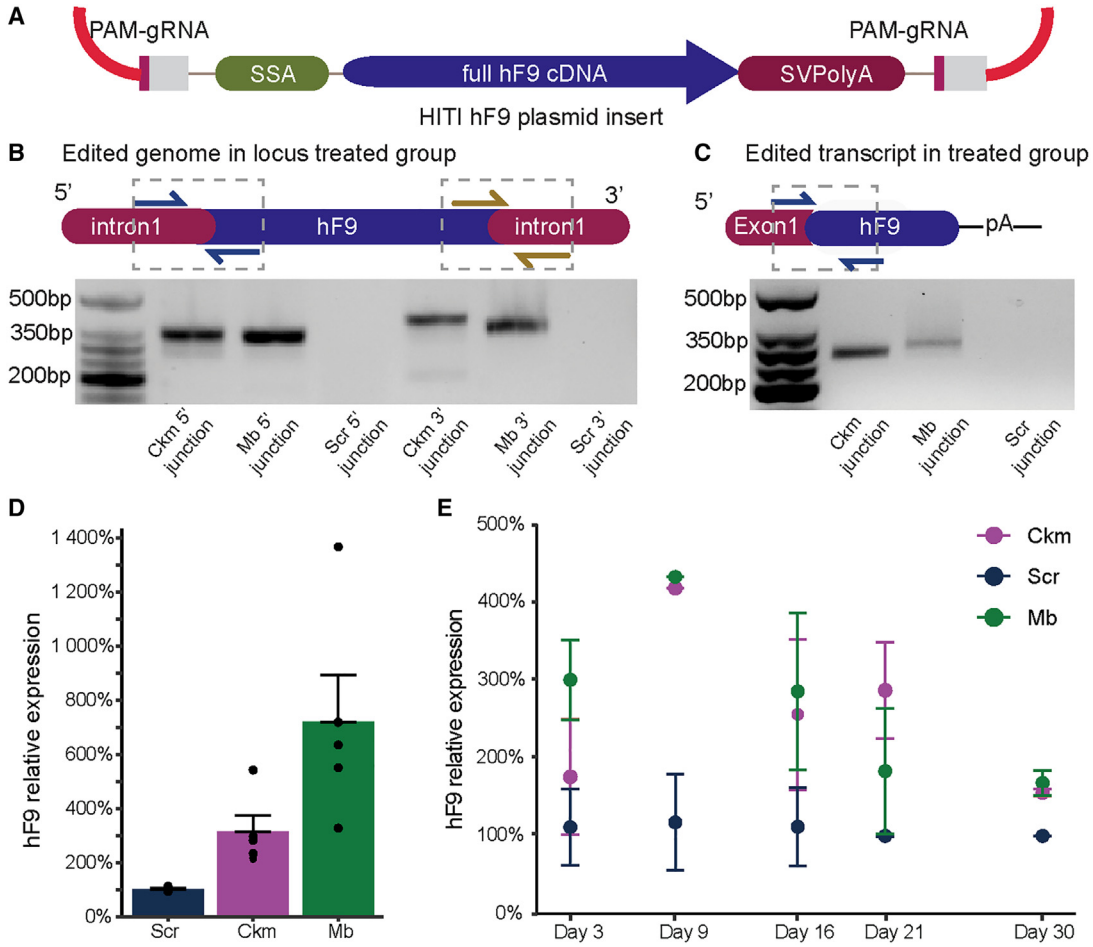


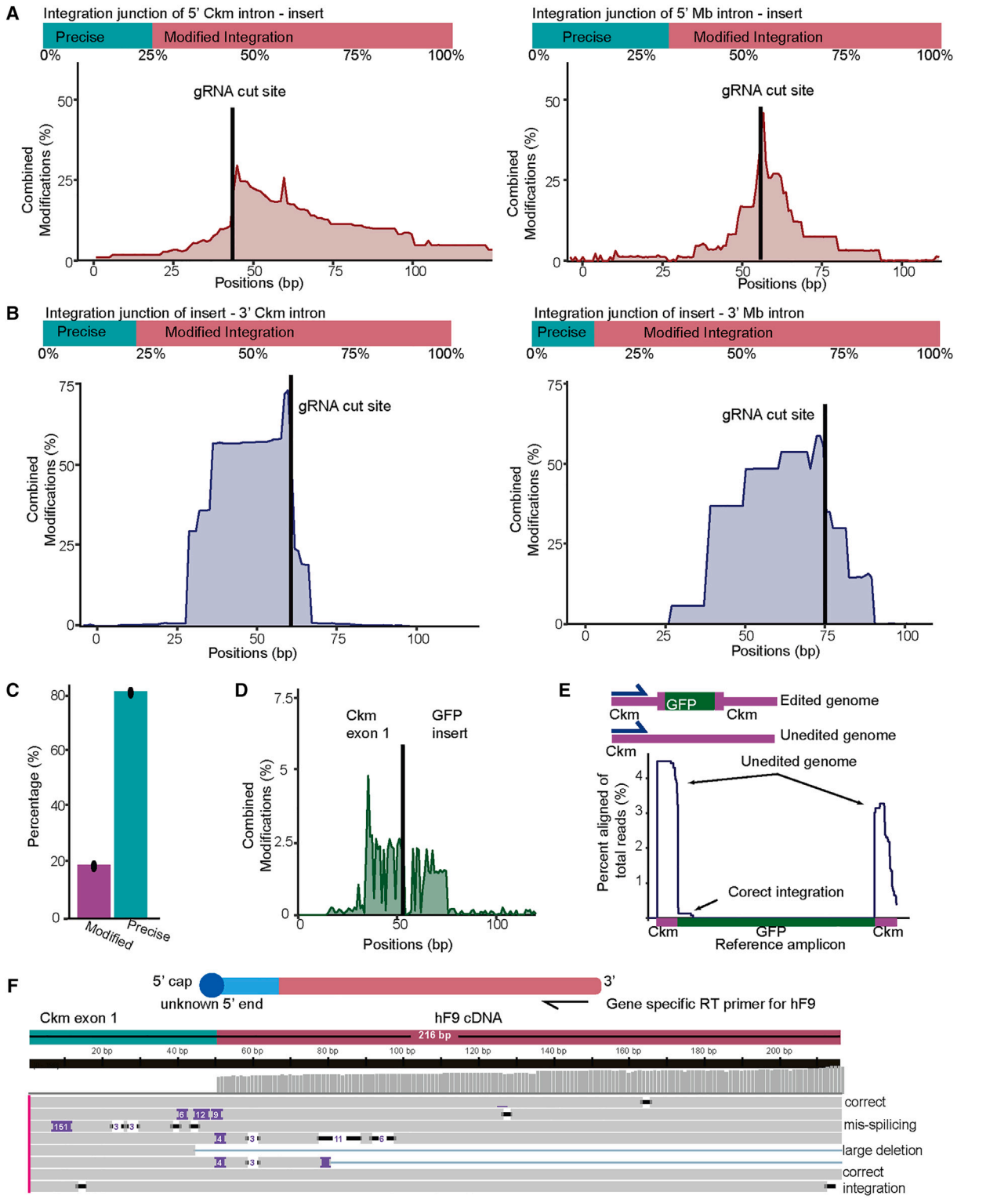
Figure 3. Targeted integration of a promoterless *hF9* gene leads to sustained expression

(A) Schematic of the structure of the HITI insert plasmid for human F9 expression. The full-length cDNA region from exon 1 to exon 8 is cloned to replace the GFP transgene in the previous construct. (B and C) Genotyping of the F9 integration in the C2C12 cell line using primers spanning the junction between the integration site and the transgene in genomic DNA and cDNA. (D) Relative *hF9* expression following plasmid transfection. Five biological independent samples, mean ± SEM. All samples were processed 10 days post transfection, including 7 days of myotube differentiation. (E) Time course cell culture to assess the expression of *hF9* in treated C2C12 cells for 30 days without selection compared to the scrambled group. Two independents biological replicates; means ± SEM.

picture of the precision, we aimed to characterize the 5′ integration junction of the fusion between *Ckm* and the transgene at the transcriptome level. This approach is motivated by the expectation that intron rearrangements would be spliced out in mRNA processing. To achieve this, we conducted 5′ rapid amplification of cDNA ends (RACE) using cDNA from Lipofectamine-transfected cells and amplified *Ckm-hF9* fusion transcripts using a gene-specific primer (GSP) positioned at the end of the *hF9* region.[55,56] Long-read nanopore sequencing confirmed the addition of exon 1 of *Ckm* in the 5′ region of *hF9*. However, we also observed several structural rearrangement events, including significant deletions and insertions (Figure 4F). Each line in Figure 4F represents a single alignment. Upon investigation of the 151-bp insertion, we determined that it originated from the splice acceptor region incorporated from the HITI insert plasmid. This suggests the occurrence of mis-splicing events at the transcriptome level.

RNA sequencing reveals differentially expressed genes after targeted integration

To investigate whether the *hF9* targeted integration into identified sites resulted in alterations in the overall transcriptome profiles, bulk RNA sequencing (RNA-seq) analysis was performed. After 7 days of myotube differentiation, samples exhibiting high *hF9* expression levels in *Ckm* and *Mb* based on qPCR analysis were compared with scramble-treated cells from the same experimental batch (Figure 5A). Paired-end sequencing on the DNBSEQTM Sequencing System from BGI with an average read length of 100 bp was used for sequencing two biological replicates from each treatment. Principal-component analysis (PCA) was performed, followed by visualization of each sample in two dimensions using the first two principal components. PCA revealed biological variation between samples in the same treatment groups. However, we observed transcriptional similarity within the *Ckm*-integrated and scrambled group



(legend on next page)

and transcriptional variations in *Mb*-treated samples compared to the scrambled group (Figure 5B).

Consistent with the PCA results, more genes were differentially expressed in *Mb*-treated samples compared to *Ckm*-treated samples (Figure 5B). We also mapped the chromosomal distribution of differentially expressed genes (Figures 5C and S9). No specific chromosome had a higher number of differentially expressed genes compared to others. Although the distribution is not even, there is no pattern indicating a specific off-target effect in any particular chromosome.

To evaluate changes in the transcriptome post treatment, we performed differential gene expression analysis. This involved combining read alignments to the mouse genome with those from the *hF9* alignment. The analysis revealed that *hF9* was upregulated in both *Ckm*- and *Mb*-treated groups. Additionally, several genes showed differential expressions in both treatments, such as *C3* and *Inmt*. Differentially expressed genes were observed to be dispersed across various chromosomes (e.g., *C3* on chromosome 17, *Inmt* on chromosome 6, and *Mmp12* on chromosome 12) rather than being clustered within the targeted chromosome (*Ckm* on chromosome 7 and *Mb* on chromosome 15), where more local contacts are expected to occur (Figures 5D and 5E).

To evaluate the biological phenomena of differentially expressed genes in *Ckm* and *Mb* compared to the scramble group, we performed Gene Ontology (GO) analysis.[57] In *Ckm*, the GO analysis indicates significant involvement in immune response and cellular mechanisms responding to environmental changes, with the most notable contributions from the genes *Mmp12* and *Oas2* (Figures 5D and S10). In the *Mb*-treated group, a broader range of genes was differentially expressed, and the *Mb* gene itself was downregulated (Figures 5E and S11). However, we did not observe GO results indicating a significant contribution of the *Mb* gene to any biological phenomena in the ontology list.

We then focused our ontology analysis on cancer-related GO terms in the *Mb* sample, as cancer development is our primary safety concern. Using cancer terminology derived from the Hallmark of Cancer and adapted to GO terms, we filtered GO terms with potential relevance to cancer development.[58,59] We found that multiple genes were involved in various biological processes: for the regulation of angiogenesis (GO:0045765), the top three genes are *Itgb8*, *Ccl11*, and *Id1*; in the process of epithelial-to-mesenchymal transition (GO:0001837), the top three genes are *Epha4*, *Epb41l5*, and *Sox9*, and for the regulation of inflammatory response (GO:0050727), the top three genes are *Sucnr1*, *Tslp*, and *Zfp36*. While these processes can be found in normal physiological systems, they are also critical in cancer development and progression, highlighting the potential impact of the observed gene expression changes. Importantly, from our list of genes, based on the Cancer Gene Census, no known oncogene was dysregulated. Similarly, we conducted a cancer ontology analysis on the *Ckm* sample and found that none of the GO terms were associated with cancer development.

In vivo validation of the AAV-CRISPR HITI strategy for hF9 integration in the *Ckm* genome

To evaluate whether this method can be applied to other types of muscle cells or different model systems, we performed the targeted integration experiment using mouse primary myoblasts and wild-type mice. Based on our initial safety assessment, which included off-target editing assessment and RNA-seq analysis, we found that the *Ckm* site did not show any off-target increased variability in the genome or general transcriptome after editing while still maintaining a high level of *hF9* integration. Therefore, we proceeded with *Ckm*-targeted site integration.

We used a dual AAV vector strategy with enhanced muscle specificity serotypes, myoAAV1C; one myoAAV vector encoded SpCas9 under a muscle-specific promoter (CK8), and the other myoAAV vector contained a *Ckm* gRNA expression cassette along with the full *hF9* coding sequence flanked by the same gRNA target site, with a viral genome ratio of 1:1[60,61] (Figure 6A). As an episomal control, we delivered the second AAV vector without the Cas9 AAV vector to the control group.

We tested this vector in mouse primary skeletal muscle cells and performed 60 h of differentiation due to the limited lifespan of differentiated primary myotubes. At this stage, nascent multinuclear myotubes were observed. Transcript analysis showed a pattern similar to the C2C12-lipofection treatment, with the *Ckm*-treated group showing higher expression compared to the scrambled control. This relative expression was approximately 10-fold higher than that observed in lipofection-treated C2C12 cells (Figure 6B). This increased expression is likely due to the higher efficiency of the viral vector compared to lipofection and the fact that primary muscle cells share a more similar characteristics to actual skeletal muscle, where *Ckm* gene expression is naturally high.[31]

For the *in vivo* experiment, we performed intramuscular injections in the first week after acclimatization and then evaluated the mice 3 and 8 weeks post injection. Transcript analysis of tibialis anterior (TA) muscle of the *Ckm*-treated group using RT-qPCR revealed that *hF9* expression was 89-fold and 60-fold higher compared to the respective

Figure 4. Deep sequencing reveals precise and imprecise outcomes of targeted integration

(A and B) Deep sequencing shows a varied range of precise integration in 5′ and 3′ junctions. A similar pattern was observed in 5′ and 3′ junctions in both loci—vector chewback that predominantly happened on the insert side. (C) Deep sequencing at the mRNA level showed a higher percentage of precise integration in three biological replicates in the treated *Ckm* locus. (D) Modification frequency trace from the fusion of *Ckm* and the GFP transgene. (E) Top: Schematic Tn5 tagmentation-based sequencing to quantify integration efficiency. Bottom: The graph shows the percentage of reads aligned to the *Ckm* gene and reference amplicon. (F) Characterization of the fusion of *Ckm* and *hF9* cDNA using 5′ RACE with the GSP reverse transcriptase primer.

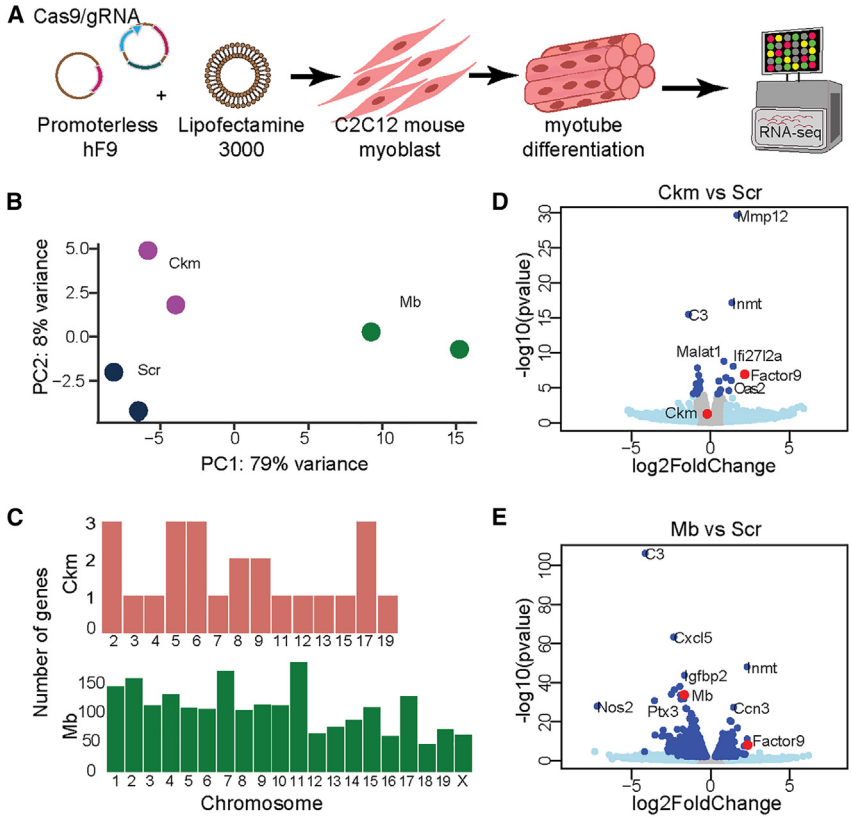


Figure 5. RNA-seq reveals significant increases in transgene expression and relative precision of integration at *Ckm* over *Mb*

(A) Pipeline for the bulk RNA-seq experiment on *Ckm*- and *Mb*-integrated and scrambled non-integrated C2C12 myotube cells. (B) PCA of two biological replicates of C2C12 myotubes with *Ckm*, *Mb*, or scrambled treatment. (C) Chromosome distribution of differentially expressed genes. (D) Differential expression of genes following *hF9* integration in the *Ckm* locus; highlighted in red are *Ckm* and *hF9*. (E) Differential expression of genes following *hF9* integration in the *Mb* locus; highlighted in red are *Mb* and *hF9*.

scrambled controls 3 weeks and 8 weeks post treatment. This expression was significantly higher than the levels observed in cells treated with lipofection. Although *hF9* expression in the 8-week *Ckm*-treated group, while higher than control, was lower than at the 3-week time point, this decline was anticipated. As we did not expect 100% integration efficiency, some AAVs might act as episomal vectors contributing to the high transcript expression levels observed in the early weeks. However, this spike in episomal expression is typically short lived, while integrated expression persists long term. In a study evaluating the sustainability of AAV episomal vectors, Greig et al. demonstrated a significant reduction in total vector RNA over time, with RNA levels initially declining after day 14 and stabilizing between days 98 and 182.[8]

We then evaluated the editing activity in the treated group using an unbiased approach with Tn5 transposon-based library preparation methods. In addition to measuring correctly oriented donor integrations, we also analyzed genome editing events, such as donor inversion integrations and AAV-Inverted Terminal Repeat (ITR) integrations at the gDNA level. We located the GSP in the 5′ and 3′ regions of the *Ckm* site for the expected integration. This GSP unidirectionally amplified sequences following the *Ckm* site. We measured correct integration rates ranging from 0.9% to 2.9%, depending on the GSP region, with 8 weeks of treatment showing higher amounts of integration. The difference between 5′ and 3′ region integration efficiency could be due to varying primer amplification efficiency or transposase tagging efficiency. While we expected the transgene to integrate only in the correct direction, we also detected reverse transgene integration in the *Ckm* target sites, with rates ranging from 0.4% to 1.1%. Additionally, as reported previously regarding the potential integration of AAV fragments, we also detected ITR integration in the *Ckm* target sites, with rates ranging from 0.2% to 0.7%.

We also assessed two types of off-target effects: off-target gRNA editing and off-target integration. For off-target gRNA editing, we performed deep amplicon sequencing at 11 predicted off-target sites, similar to our *in vitro Ckm* off-target analysis. We detected an indel rate of around 25% at the on-target gRNA site at both time points (Table S7). However, we did not identify any indel formation above background levels at the off-target sites at either time point, suggesting that no or minimal off-target editing occurred (Table S8; Figure S7). To evaluate off-target integration events in the treated group at the transcriptome level, similar to the method used for measuring editing activity, we employed Tn5 transposon-based library preparation to map integration events across the whole genome. Using 3′ end GSP flanking the *hF9* insert, we detected and mapped the unknown 5′ regions of the sequences through pattern analysis.[62] Deep sequencing revealed some reads corresponding to the AAV episomal vector, while the majority were integrated into the *Ckm* transcript. No off-target integration events into other genes were detected, except for a notable number of chimeric sequences involving 18S rRNA (Figure S12). While this could potentially represent true off-target integration events, a previous report suggests that, during PCR amplification of rRNA genes, the formation of chimeric sequences between different copies is likely to occur.[63] Additionally, we did not perform rRNA depletion during both first-strand and second-strand cDNA synthesis, which could have contributed to the observed chimeric sequences.

To assess whether there were any effects at the transcriptome level, we also performed RNA-seq on muscle tissue. The PCA plot distribution showed a more uniform sample distribution, even in comparison with

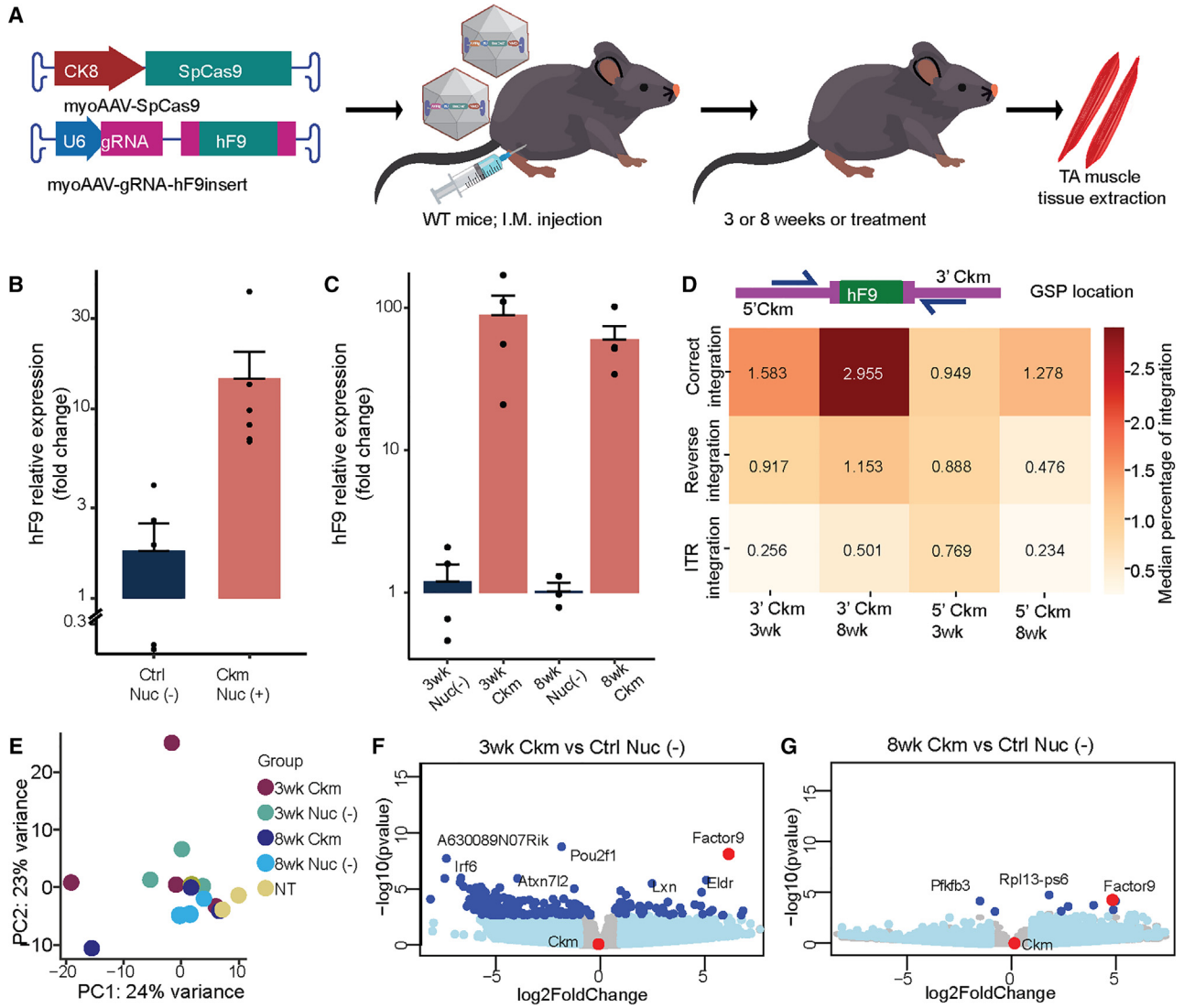


Figure 6. *In vivo* validation of *hF9*-targeted integration at the *Ckm* locus

(A) Schematic of the AAV-CRISPR mediated integration *in vivo* experiment. (B) Relative *hF9* expression in primary skeletal muscle cells following myoAAV-CRISPR transduction; mean ± SEM. Ctrl Nuc (−), control without the nuclease group; *Ckm* Nuc (+), *Ckm* target with the nuclease group. (C) Relative *hF9* expression in mice following myoAAV CRISPR injections at two time points with respective control groups. Mean ± SEM. 3wk Nuc(−) and 8wk Nuc(−), no nuclease group at the 3- and 8-week time point, respectively; 3wk *Ckm*, *Ckm* target with the nuclease group at the 3- and 8-week time point, respectively. (D) Integration results from 5′ and 3′ gene-specific primer direction in gDNA. (E) PCA plot of treated, control without nuclease, and untreated mice (NT). (F) Differential expression of genes following *hF9* integration in the *Ckm* locus at the 3-week time point; highlighted in red are *Ckm* and *hF9*. (G) Differential expression of genes following *hF9* integration in *Ckm* locus at the 8-week time point; highlighted in red are *Ckm* and *hF9*.

non-treated TA muscle tissue. *In vivo* differential expression analysis at the 3-week time point showed more differentially expressed genes compared to the 8-week time point relative to their respective control groups (Figures 6F and 6G). The GO analysis for the 3-week mice highlights processes related to the immune response, with a particular focus on adaptive immunity, antigen processing and presentation, and T cell activation. At the 8-week time point, the GO analysis continues to reflect broader aspects of immune response. The *Ckm* gene, similar to what we observed *in vitro*, did not change in terms of expression. We also performed transcripts per million (TPM) analysis for the *Ckm* gene, comparing expression levels between the treated group, scrambled, and non-treated samples. No differences in TPM expression were observed among these three categories, suggesting that the integration did not affect the expression level of *Ckm* (Figure S13). Importantly, GO analysis did not reveal any cancer-related ontologies, which we had seen in the *Mb* analysis.

Finally, we tested hFIX protein expression in both cell serum and mouse plasma at the 3-week and 8-week time points. We observed higher hFIX expression in the primary cell serum treated with AAV compared to the C2C12 cell serum treated with lipofection. In treated *Ckm*-group mice, there was a trend toward higher hFIX expression compared to the scrambled group (Figure S14).

DISCUSSION

Our current work aims to address several significant challenges associated with AAV vector-based *in vivo* gene therapy, including long-term efficacy, promoter or transgene silencing, and the safety of transgene integration.[10,64,65] To enhance the sustainability of the treatment, targeted integration into the host genome has emerged as a potential solution. However, the integration site must be chosen strategically, as the integrating vector with the constitutive promoter might pose risks. Despite their ability to facilitate robust and consistent transgene expression, such promoters are associated with increased risks, including an elevated likelihood of inactivation, amplified toxicity resulting from transgene overexpression, off-target transgene expression, and the potential for severe immune responses due to inadvertent transgene expression in antigen-presenting cells.[65]

Utilizing endogenous promoters for transgene expression can circumvent potential silencing mechanisms associated with exogenous promoters. In this study, we identified two highly expressed genes, *Ckm* and *Mb*, in skeletal muscle and exploited the transcriptional output of their endogenous promoters. Our findings demonstrate that these loci can drive the expression of promoterless GFP and enhance RNA expression of promoterless *hF9* in cells and muscle tissue. While in the beginning we suspected that the limited level of integration was primarily due to the efficiency of lipofection in a skeletal muscle cell line, we also observed limited integration efficiency *in vivo* despite using a muscle-specific AAV serotype and achieving a high indel rate at the targeted site *in vivo*. Similar findings have been reported by Pickar-Oliver et al. and Stephenson et al., who observed up to 1% on-target HITI-mediated integration in skeletal muscle or heart muscle in mice through AAV-mediated delivery.[66,67] Several factors could contribute to these outcomes. First, the unique nature of skeletal muscle tissue poses obstacles for efficient editing. Second, the immune response may also play a role. While intramuscular injection has been proven to evoke a reduced immune response, local immune responses might still hinder editing efficiency. This issue is not unique to our method; growing evidence indicates that immune responses are a common challenge in gene editing.[68] Another potential factor is the ratio of donor DNA to Cas9. Stephenson et al. examined various Cas9:donor ratios and found that a 1:5 ratio achieved maximal efficiency, resulting in 1.4% genome integration in heart muscle.[67] Optimizing this ratio could potentially enhance integration efficiency in skeletal muscle as well.

Regarding the protein expression that is shown in *in vitro* levels, the expression levels of *Ckm* and *Mb* in C2C12 cells were not as high as those reported in skeletal muscle tissue[31] (Table S3). This difference may result in the lower expression of GFP and hFIX protein observed. However, *in vivo*, we also noticed low protein expression (around 2 ng/mL), which is lower than the required 50 ng/mL to reach therapeutic levels of FIX.[27] One key factor could be the rich presence of collagen IV in skeletal muscle that may facilitate the local attachment of FIX, limiting its release into circulation.[69] Previous studies have suggested the potential use of hFIX variants harboring mutations such as lysine to alanine at residue 5 (K5A) or valine to lysine at residue 10 (V10K), which show reduced binding to endothelial cells, thus enabling synthesis of hFIX in skeletal muscle while maintaining normal clotting activity.[70,71] Additionally, the injection was administered into the TA muscle, a relatively small muscle compartment compared to the entire body musculature. This likely limited the number of nuclei that were edited and subsequently expressed the edited transcript and therapeutic protein. Future attempts could benefit from targeting larger muscles, such as the quadriceps, to potentially increase the number of edited nuclei. Finally, the precision of integration, as detailed in our sequencing analysis, may also play a role in the observed expression levels. Addressing these factors could significantly improve future outcomes in achieving therapeutic levels of FIX.

The general safety evaluation of the proposed sites indicates that the group treated with *Mb* exhibits alterations in the global transcriptome. The downregulation of the *Mb* gene may be attributed to the higher indel activity of the CRISPR/gRNA complex at that target site compared to the indel activity at the *Ckm* target sites. Additionally, while the target site location might present challenges for splice site disruption owing to indels created in splice motifs, the downregulation of *Mb* could be due to the disruption of the promoter region in the distal intron 1 rather than splicing disruptions.[41] Conversely, the *Ckm* gene, which lacks promoter features around the selected region, did not show any downregulation in *in vitro* or *in vivo* studies. Despite observing minimal changes in the expression of upregulated and downregulated genes at the *Ckm* locus, suggesting initial safety validation, a more comprehensive safety analysis is still warranted.

Based on our ontology analysis and PCA plot, the observed changes in gene expression levels were not expected from *hF9* integration. This is supported by the *Ckm* group differential expression and GO analysis both *in vitro* and *in vivo*, which also has *hF9* integration but did not show diverse differential gene expression like the *Mb* group. The PCA plot further supports this, as the scrambled group and *Ckm*-treated group are plotted in proximity. Therefore, these changes in gene expression can be mainly attributed to the selected *Mb* site. While the transcriptional changes observed in the *Mb* site may not directly result in cancer, they could potentially lead to toxic effects. A longer study period is necessary to thoroughly evaluate the safety and efficacy of this integration.

Our sequencing studies also show patterns of deletions following integration induced by double-strand breaks (DSBs). These sequences demonstrate that indel-induced damage tends to occur toward the insertion site, as evidenced by the selected PCR-enriched integration sequencing. This trend is consistently observed across loci and in both

5′ and 3′ integrations. This insight is valuable for future template design efforts, emphasizing the importance of incorporating padding sequences before the coding sequence of the transgene to mitigate "chewback" into the desired gene before integration can occur.

Our approach of integrating the transgene into intronic regions offers the advantage of splicing, as indels occurring at the intronic level are typically spliced out, resulting in more precise mRNA. Another study has also reported the advantages of intron targeting, which results in the production of nearly error-free mRNA.[72] However, identifying the strongest or most optimal splice acceptor poses a significant challenge. Despite selecting a robust splice acceptor from a highly conserved gene in skeletal muscle, based on our 5′ RACE results, we observed instances of mis-splicing due to the failure of the splice acceptor to be properly excised. To date, there is a lack of practical guidelines available to determine the ideal length of splice acceptors or splice donors for a given gene. Thus, design of the splice acceptor allows additional optimization of gene expression.

The two sequencing approaches that we performed, both short and long read, provided us with a general overview of the effects of DSBs on integration sites. As targeted integration gains prominence as a viable gene editing therapy approach, we should not overlook the analysis of post-DSB-induced integration events. A more thorough analysis regarding structural rearrangements and vector integration is needed to get a full understanding of these consequences. Another potential avenue involves exploring recent DSB-less or non-DSB approaches, such as ShCAST (*Scytonema hofmanni CRISPR-associated transposase)*, TwinPE (Twin prime editing), PASTE (Programmable Addition via Site-specific Targeting Elements), CAST (CRISPR-associated transposase), HELIX, (a nicking homing endonuclease fusion to TnsB) and PRINT (precise RNA-mediated insertion of transgenes).[73–78]

In summary, this work introduces a proof-of-concept technology for achieving high-level therapeutic gene expression in skeletal muscle. This innovative method shows potential for advancing RNA-guided DNA integration, which transforms skeletal muscle into a biofactory capable of producing desired therapeutic agents. While further refinement is needed to enhance various aspects of the concept, our findings provide a promising pathway toward realizing the potential of gene editing therapy, broadening its reach and impact on muscle gene therapy.

MATERIALS AND METHODS

In vitro gRNA screening

gRNA plasmids were constructed by cloning the targeted sequences into the pX330 vector (Addgene 42230) via the BbsI sites under the human U6 promoter. A panel of gRNAs was designed based on high efficiency and predicted minimal off-target effects using the CRISPOR tool from University of California Santa Cruz (UCSC) to target intron 1 in each site.[47] The SpCas9 protein is expressed in a separate plasmid under the CMV promoter. gRNA activity was compared by surveyor assay in NIH 3T3 cells and by NGS in the C2C12 cell line. NIH 3T3cells were obtained from the ATCC (CRL-1658) and maintained in Dulbecco's modified Eagle's medium (DMEM) (Gibco) supplemented with 10% fetal bovine serum (FBS) (Sigma) and 1% penicillin-streptomycin (P/S) (Gibco). C2C12 cell lines were obtained from the ATCC (CRL-1772) and cultured in DMEM supplemented with 20% FBS and 1% P/S. Both cell lines were grown in a tissue culture incubator at 37°C and 5% CO_2. NIH 3T3 cells were transfected with plasmids encoding SpyCas9 or SauCas9. After a 72-h incubation period, genomic DNA was isolated using a DNeasy kit (QIAGEN). Indels were identified by PCR of the region of interest (Tables S3), followed by incubation with the Surveyor nuclease and electrophoresis on Tris/Borate/EDTA agarose gels as described previously. C2C12 transfection was performed using Lipofectamine 3000 (Thermo Fisher Scientific). In 24-well plates, C2C12 cells were transfected using 250 ng pSpCas9 and 250 ng pgRNA plasmid. Three days after transfection, cells were harvested, and gDNA was extracted as described above. In the second PCR round, full Illumina flowcell adapter sequences along with experiment-specific barcodes were appended to the 5′ and 3′ ends of the PCR product, respectively (Table S9). After pooling the resultant PCR products, sequencing was performed with 150-bp paired-end reads on an Illumina ISeq instrument (Iseq 100 i1 Reagent v.2, 300-cycle). Samples were demultiplexed based on assigned barcode sequences, and the Illumina adapter sequences were trimmed from the reads. The indel analysis was carried out using CRISPResso analysis, following the default parameters.[79]

HITI insert plasmid preparation and C2C12 transfection

Plasmid constructs were produced via either gene synthesis or conventional subcloning methodologies. To promote transparency and accessibility, all constructs not previously disclosed, along with their associated sequences, will be submitted to Addgene. Depending on the particular downstream experiments, the donor constructs may include either promoterless GFP or *hF9* coding sequences, both of which are surrounded by the same gene-targeting cutting site, along with the splice acceptor and simian virus polyadenylation signal (SVpoly(A)). Exon 1 of *hF9* was ordered from Integrated DNA Technologies (IDT) as a gene block, and exons 2–8 were cloned from Addgene 182141. HITI insert plasmids were designed as reported previously.[30] Cloned plasmids were confirmed with Sanger sequencing, and confirmed plasmids were isolated and purified using the QIAGEN Plasmid Plus Maxi Kit, following the manufacturer's protocol.

Cell culture transfection was carried out using Lipofectamine 3000 (Thermo Fisher Scientific) following the standard protocol. Briefly, 24 h before transfection, 25,000 C2C12 cells were seeded into 24-well plates to achieve 50%–60% confluency the next day. On the day of transfection, 2 h before transfection, the growth medium was replaced with fresh growth medium. Three plasmids—Cas9, gRNA, and the HITI insert—were first mixed at a ratio of 2:2:1 with the reagent P3000 (ratio of DNA:P3000 of 1:2) in Opti-MEM Reduced Serum Medium (Gibco) and then added to Opti-MEM Reduced Serum Medium (Gibco) containing Lipofectamine 3000. DNA/RNA:reagent complexes were briefly vortexed, kept at room

temperature for 15 min according to the supplier's instructions, and added dropwise onto the cell monolayer. Three days post transfection, when confluency exceeded 90%, myoblasts were differentiated with DMEM supplemented with 2% donor equine serum, 1% P/S, and insulin. Cells were differentiated for 7 days and then harvested. For long-term culture, after transfection, cells were passaged and harvested every 2 days for 23 days before reaching 80% confluency to prevent spontaneous myotube formation. On day 23, cells were differentiated into myotubes as described above.

Transfer plasmid design, AAV production, and primary cell and animal experiments

The ITR-containing muscle-specific promoter CK8-*Staphylococcus pyogenes* Cas9 with the mini poly(A) (pAAV-SpCas9) transfer plasmid was generated by replacing the elongation factor 1alpha binding sequence with the CK8 gene block promoter. A separate transfer plasmid with the hU6-driven Ckm gRNA cassette and donor insert was created using Gibson cloning to insert the donor into the ITR-containing plasmid. The transfer plasmids were verified by whole plasmid sequencing by Eurofins, and intact ITRs were confirmed by SmaI digestion before AAV production. The myoAAV1C capsid plasmid was generated using PCR stitching to mutate the AAV9 capsid plasmid (Addgene 112865), following a report published previously.[60] The capsid plasmid was also verified by whole-plasmid sequencing. Two myoAAV1C vectors were produced by the University of North Carolina at Chapel Hill Vector Core. Titers were measured by qPCR with a plasmid standard curve.

Primary myogenic progenitor cells were harvested from >2-month-old male C57BL/6J mice following a published protocol,[80] as we have performed previously.[81,82] The cells were cultured on extracellular matrix-coated Primaria plates (Corning 353801) in Ham's F-10 Nutrient Mix (Thermo Fisher Scientific 11550043) supplemented with fibroblast growth factors (5 ng/mL, Millipore). Transduction was performed using a 1:1 AAV Cas9:donor ratio at an MOI of 10,000 without the addition of Polybrene. The growth medium was replaced every other day until cell confluency reached over 70%, at which point it was switched to differentiation medium (high-glucose DMEM + 10% normal horse serum + 1% P/S). The differentiation process was carried out for 60 h, with medium changes every day, as described previously; this time point is sufficient to induce large multinuclear myotubes.[82]

All animal experiments were performed in full compliance with the NIH guidelines for laboratory animal care and use. The Institutional Animal Care and Use Committee at the University of Arkansas approved all experimental procedures. The C57BL/6J mice were supplied by The Jackson Laboratory. Eight-week-old mice received injections into the TA muscle with a 1:1 ratio of pAAV-CK8-SpCas9 and pAAV-U6-gRNA-donor insert hF9, with a total AAV concentration of 7e12 vector genomes/kg of body weight. Tissues were harvested at two time points: 3 weeks and 8 weeks post injection. At each time point, mice were sacrificed, and TA muscle and plasma were harvested.

Genomic DNA, RNA, and protein analyses

gDNA extraction from cell pellets was carried out using the QIAGEN DNeasy kit (69504) according to the manufacturer's protocol. For muscle tissue, gDNA extraction was performed using the High Molecular Weight DNA Extraction Kit (T3060), following the manufacturer's protocol for low-input samples. RNA extraction was performed using New England Biolab's Monarch Total RNA Miniprep Kit (T2010), following the manufacturer's instructions, including DNAseI treatment. mRNA was then reverse transcribed using LunaScript RT-Supermix (New England Biolabs [NEB] E3010). Genotyping PCR was then conducted using Q5 polymerase (NEB) for genomic DNA and cDNA. qPCR analysis for *hF9* was executed using Luna Universal Probe qPCR Master Mix (M3004) with FAM probe primers (IDT), while the *Ppia* reference gene was assessed using Luna Universal qPCR Master Mix (M3003). For primer location, sequences, and combinations, please refer to Table S9.

To analyze off-target gRNA editing, we performed PCR amplification of 124- to 245-bp fragments encompassing both the target sites and predicted off-target sites (Tables S5 and S6). This was done using gDNA extracted from two edited cells per site and two scrambled cells per site as well as gDNA from three Ckm-treated muscle tissues and three control muscle tissues. All PCR reactions used Q5 High-Fidelity DNA Polymerase (M0491). The primers used for amplification included adapter sequences to which custom-designed barcodes and Illumina-compatible sequencing adapters were added in a subsequent PCR reaction, as described previously. Analysis was performed using the CRISPResso pool feature.[79]

For precision deep sequencing analysis, PCR-based genotyping of *Ckm* and *Mb* sites using primers spanning the 5′ and 3′ junctions of the targeted site and the inserted GFP transgene was performed for 15 cycles. The second round of PCR and the analysis were performed as described previously. In CRISPResso2 analysis, the amplicon sequence was the expected precise integration based on HITI integration.[79] The percentage of unmodified sequences was tallied as precise integration, while the percentage of modified sequences was categorized as non-precise integration and then mapped to identify the regions where modifications predominantly occurred.

For protein analysis, cell supernatant was collected on days 3 and 7 of differentiation or from plasma and stored at −80 before analysis. Analysis of hFIX protein was performed using the Human Coagulation Factor IX Total Antigen ELISA Kit (Innovative Research) following the supplier's protocol, with 1 M HCl used as a stop solution.

Tn5-mediated tagmentation sequencing

Unloaded Tn5 transposase proteins were acquired from Diagenode (C01070010-10). Tn5 transposase was then loaded with oligos containing mosaic ends incorporating 13-bp UMIs and the i5 sequencing adapter according to the manufacturer's protocol. Tagmentation of 50 ng gDNA with 1:8 dilutions of loaded Tn5 was carried out as described previously.[83] Double-stranded cDNA was synthesized using DNA Polymerase I, Large (Klenow Fragment) (NEB M0210),

following a protocol published previously.[84] Briefly, the cDNA was heated to 95°C for 2 min and immediately placed on ice for 2 min. Subsequently, 1 μL of Klenow enzyme was added for every 20 μL of the cDNA reaction. The mixture was then incubated at 37°C for 60 min, followed by 75°C for 20 min. Nested PCRs were then performed using the full i5 adapter and a GSP positioned in the intron of *Ckm* to enrich the target site and incorporate sample barcodes/the i7 adapter. All PCRs were performed with Q5 DNA polymerase, with an annealing temperature of 60°C for 15 s and an extension temperature of 68°C for 1 min. Sequencing was conducted on the Illumina iSeq platform or Nova-seq using the 150-bp paired-end reagent cartridge following the standard protocol.

Sequence data from the clusters that pass filter on a flow cell (FASTQ files) were demultiplexed using the list of barcodes assigned to each sample. Trimmomatic (v.0.33) was then used to trim sequencing adapters and remove low-quality reads.[85] Paired-end reads were merged using Paired-End reAd mergeR, and then UMI sequences were annotated using UMI-tools.[86,87] Afterward, the annotated FASTQ files were aligned to a reference amplicon using bwa-mem2.[88] The reference amplicons were constructed to align with the targeted locus and anticipated edits. Sequence Alignment Map (SAM) files resulting from the alignment were converted to Binary Alignment Map (BAM) files and then indexed. Deduplication of the indexed BAM reads was performed based on the UMI annotation using UMI-tools. Following deduplication, reads were filtered using seqkit tools to eliminate reads attributed to false priming (reads lacking the 20 bases directly adjacent to the GSP expected sequence were filtered out). Additionally, reads failing to extend adequately beyond the edit site and those falling short of the minimum required length were filtered out. Finally, the deduplicated and filtered reads were analyzed and mapped to quantify the frequency of integration.

For hF9 integration quantification sequencing, FASTQ files were aligned to the mouse whole genome, including the hF9 AAV insert (from ITR to ITR). After deduplication, BAM files were categorized into chimeric reads from the target site (*Ckm*) to the hF9 insert and reads from the target site (*Ckm*) only were used as the denominator. Reads were then identified for direction based on expected integration and were checked for origin from AAV fragments. Browser Extensible Data (BED) tool extraction was performed to quantify Ckm-only amplification. For mapping off-target integration, BAM files were categorized for chimeric reads from the hF9 insert to any detected chromosome. These reads were extracted to FASTQ format and filtered to ensure that the sequences contained the on-target primer and did not contain sequencing artifacts. They were then aligned to the hF9 insert only using the mappy aligner. The unaligned FASTQ reads were then parsed and analyzed using the motif-based sequence pattern analyzer.[62] Identified patterns were confirmed using BLAST-Like Alignment Tool or BLAST to determine the source of integration.

5′ RACE and long-read sequencing

The 5′ RACE protocol using the template-switching reverse transcriptase enzyme mix from NEB was executed following the manufacturer's guidelines. Briefly, 1 μg of RNA was annealed with a gene-specific RT primer containing deoxynucleotide triphosphates at 70°C for 5 min. Next, the mixture was supplemented with template-switching buffer, template-switching oligo, and RT enzyme mix (NEB M0466), followed by an incubation step at 42°C for 90 min and then at 85°C for 5 min. Touchdown PCR amplification was carried out using gene-specific reverse primer and Template Switching Oligo-specific primer on 1 μL of the template-switching sample. To assess the enrichment of the amplified fragment in the targeted gene, qPCR was performed on the bead-purified PCR product. Qubit measurement was used to determine the concentration of the purified PCR product. A comparative analysis of qPCR results for cDNA with similar amounts served as a control. A Cq value of <10 was considered indicative of good enrichment for RACE amplification. Enriched samples were then prepared for Oxford Nanopore long-read sequencing library preparation following the manufacturer's protocol (SQK-NBD112.24), and sequencing was conducted using a MinION flow cell (R10.4.1). FASTQ reads were aligned to reference amplicon that has the fusion of *Ckm* exon 1 with the *hF9* coding sequence using minimap2 alignment[89] and then visualized using Integrative Genome Viewer.[90]

RNA-seq analysis

After qPCR analysis, the best-expressing samples in *Ckm*-treated and MB-treated cells, as well as the scrambled samples of C2C12 cells, were further assessed for RNA integrity using Tapestation with High Sensitivity RNA ScreenTape (5067-5579). RNA samples with (RNA Integrity Number) RIN > 7 and a 260/280 ratio of around 2 were selected for sequencing. The RNA-seq assay was conducted by BGI Genomics (Shenzhen, China) or by the Oklahoma Medical Research Foundation Clinical Genomics Center. Following RNA-seq, the FASTQ data underwent quality control using fastqc, and poor-quality reads were trimmed using Trimmomatic (v.0.33).[85] High-quality reads were then aligned to mouse and human genomes using HISAT2.[91] The number of raw reads associated with each gene was quantified using featureCounts. Next, the raw read data were subjected to PCA and differential expression analysis using DESeq2 in R. GO analysis was performed using gprofiler2, and chromosomal distribution analysis was visualized using biomaRt.[57,92]

DATA AND CODE AVAILABILITY

The custom code developed for this study is publicly accessible at https://github.com/harumiswari/sequencing_analysis. All raw sequencing data were submitted to the National Center for Biotechnology Information Sequence Read Archive under BioProject PRJNA1139805. Any additional raw data relevant to this study can be obtained from the corresponding author upon request.

ACKNOWLEDGMENTS

The authors would like to thank all members of the Nelson lab and Dr. Nicholas Greene for helpful conversations about muscle biology. This work was supported by NIH/NIBIB R00 #EB023979, an ASGCT Career Development Award, a University of Arkansas Chancellor's Innovation Grant, and The Arkansas Bioscience Institute. C.E.N. was supported by the 21st Century Chair in Biomedical Engineering. M.H.P. was supported by the Fulbright Indonesia Research Science and Technology (First) Master's Degree Program and PhRMA Foundation 2023 Predoctoral Fellowship in Drug Discovery. G.B. was supported by the University of Arkansas Honors College Undergraduate Research Grant. M.S.J. was supported by a State Undergraduate Research Fellowship (SURF), the Bodenhamer

Fellowship, and the Goldwater Scholarship. S.K. and K.A.M. were supported by NIH/NIA R01 AG080047.

AUTHOR CONTRIBUTIONS

M.H.P. and C.E.N. designed experiments and wrote the manuscript. M.H.P. performed experiments, analyzed data, and created figures. M.H.P., G.B., and S.A. performed molecular cloning and molecular analysis. M.H.P. and S.K. performed primary cells experiments. K.A.M. provided resources, intellectual input, and edits to the manuscript. M.H.P. and M.S.J. performed short-read sequencing.

DECLARATION OF INTERESTS

M.H.P. and C.E.N. are named inventors on patents and patent applications related to genome editing.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2024.102320.

REFERENCES

1. Dunbar, C.E., High, K.A., Joung, J.K., Kohn, D.B., Ozawa, K., and Sadelain, M. (2018). Gene therapy comes of age. Science *359*, eaan4672. https://doi.org/10.1126/science.aan4672.
2. Kuzmin, D.A., Shutova, M.V., Johnston, N.R., Smith, O.P., Fedorin, V.V., Kukushkin, Y.S., van der Loo, J.C.M., and Johnstone, E.C. (2021). The clinical landscape for AAV gene therapies. Nat. Rev. Drug Discov. *20*, 173–174. https://doi.org/10.1038/d41573-021-00017-7.
3. Niemeyer, G.P., Herzog, R.W., Mount, J., Arruda, V.R., Tillson, D.M., Hathcock, J., van Ginkel, F.W., High, K.A., and Lothrop, C.D., Jr. (2009). Long-term correction of inhibitor-prone hemophilia B dogs treated with liver-directed AAV2-mediated factor IX gene therapy. Blood *113*, 797–806. https://doi.org/10.1182/blood-2008-10-181479.
4. Nathwani, A.C., Reiss, U., Tuddenham, E., Chowdary, P., McIntosh, J., Riddell, A., Pie, J., Mahlangu, J.N., Recht, M., Shen, Y.-M., et al. (2018). Adeno-Associated Mediated Gene Transfer for Hemophilia B:8 Year Follow up and Impact of Removing "Empty Viral Particles" on Safety and Efficacy of Gene Transfer. Blood *132*, 491. https://doi.org/10.1182/blood-2018-99-118334.
5. Rangarajan, S., Walsh, L., Lester, W., Perry, D., Madan, B., Laffan, M., Yu, H., Vettermann, C., Pierce, G.F., Wong, W.Y., and Pasi, K.J. (2017). AAV5–Factor VIII Gene Transfer in Severe Hemophilia A. N. Engl. J. Med. *377*, 2519–2530. https://doi.org/10.1056/NEJMoa1708483.
6. Jacobson, S.G., Cideciyan, A.V., Roman, A.J., Sumaroka, A., Schwartz, S.B., Heon, E., and Hauswirth, W.W. (2015). Improvement and Decline in Vision with Gene Therapy in Childhood Blindness. N. Engl. J. Med. *372*, 1920–1926. https://doi.org/10.1056/NEJMoa1412965.
7. Das, A., Vijayan, M., Walton, E.M., Stafford, V.G., Fiflis, D.N., and Asokan, A. (2022). Epigenetic Silencing of Recombinant Adeno-associated Virus Genomes by NP220 and the HUSH Complex. J. Virol. *96*, e02039-21. https://doi.org/10.1128/jvi.02039-21.
8. Greig, J.A., Martins, K.M., Breton, C., Lamontagne, R.J., Zhu, Y., He, Z., White, J., Zhu, J.-X., Chichester, J.A., Zheng, Q., et al. (2024). Integrated vector genomes may contribute to long-term expression in primate liver after AAV administration. Nat. Biotechnol. *42*, 1232–1242. https://doi.org/10.1038/s41587-023-01974-7.
9. 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 dogs with hemophilia A identifies clonal expansions of transduced liver cells. Nat. Biotechnol. *39*, 47–55. https://doi.org/10.1038/s41587-020-0741-7.
10. Colella, P., Ronzitti, G., and Mingozzi, F. (2018). Emerging Issues in AAV-Mediated In Vivo Gene Therapy. Mol. Ther. Methods Clin. Dev. *8*, 87–104. https://doi.org/10.1016/j.omtm.2017.11.007.
11. Sabatino, D.E., Bushman, F.D., Chandler, R.J., Crystal, R.G., Davidson, B.L., Dolmetsch, R., Eggan, K.C., Gao, G., Gil-Farina, I., Kay, M.A., et al. (2022). Evaluating the state of the science for adeno-associated virus integration: An integrated perspective. Mol. Ther. *30*, 2646–2663. https://doi.org/10.1016/j.ymthe.2022.06.004.
12. Kaeppel, C., Beattie, S.G., Fronza, R., van Logtenstein, R., Salmon, F., Schmidt, S., Wolf, S., Nowrouzi, A., Glimm, H., von Kalle, C., et al. (2013). A largely random AAV integration profile after LPLD gene therapy. Nat. Med. *19*, 889–891. https://doi.org/10.1038/nm.3230.
13. Yamamoto, Y., and Gerbi, S.A. (2018). Making ends meet: Targeted integration of DNA fragments by genome editing. Chromosoma *127*, 405–420. https://doi.org/10.1007/s00412-018-0677-6.
14. Lampe, G.D., and Sternberg, S.H. (2023). Novel recombinases for large DNA insertions. Nat. Biotechnol. *41*, 471–472. https://doi.org/10.1038/s41587-022-01600-y.
15. Jasin, M., and Rothstein, R. (2013). Repair of Strand Breaks by Homologous Recombination. Cold Spring Harbor Perspect. Biol. *5*, a012740. https://doi.org/10.1101/cshperspect.a012740.
16. Sadelain, M., Papapetrou, E.P., and Bushman, F.D. (2011). Safe harbours for the integration of new DNA in the human genome. Nat. Rev. Cancer *12*, 51–58. https://doi.org/10.1038/nrc3179.
17. Battulin, N., Korablev, A., Ryzhkova, A., Smirnov, A., Kabirova, E., Khabarova, A., Lagunov, T., Serova, I., and Serov, O. (2022). The human EF1a promoter does not provide expression of the transgene in mice. Transgenic Res. *31*, 525–535. https://doi.org/10.1007/s11248-022-00319-5.
18. Brooks, A.R., Harkins, R.N., Wang, P., Qian, H.S., Liu, P., and Rubanyi, G.M. (2004). Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. J. Gene Med. *6*, 395–404. https://doi.org/10.1002/jgm.516.
19. Pavani, G., Laurent, M., Fabiano, A., Cantelli, E., Sakkal, A., Corre, G., Lenting, P.J., Concordet, J.-P., Toueille, M., Miccio, A., and Amendola, M. (2020). Ex vivo editing of human hematopoietic stem cells for erythroid expression of therapeutic proteins. Nat. Commun. *11*, 3778. https://doi.org/10.1038/s41467-020-17552-3.
20. Sharma, R., Anguela, X.M., Doyon, Y., Wechsler, T., DeKelver, R.C., Sproul, S., Paschon, D.E., Miller, J.C., Davidson, R.J., Shivak, D., et al. (2015). In vivo genome editing of the albumin locus as a platform for protein replacement therapy. Blood *126*, 1777–1784. https://doi.org/10.1182/blood-2014-12-615492.
21. Barzel, A., Paulk, N.K., Shi, Y., Huang, Y., Chu, K., Zhang, F., Valdmanis, P.N., Spector, L.P., Porteus, M.H., Gaensler, K.M., and Kay, M.A. (2015). Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. Nature *517*, 360–364. https://doi.org/10.1038/nature13864.
22. Pellenz, S., Phelps, M., Tang, W., Hovde, B.T., Sinit, R.B., Fu, W., Li, H., Chen, E., and Monnat, R.J. (2019). New Human Chromosomal Sites with "Safe Harbor" Potential for Targeted Transgene Insertion. Hum. Gene Ther. *30*, 814–828. https://doi.org/10.1089/hum.2018.169.
23. Aznauryan, E., Yermanos, A., Kinzina, E., Devaux, A., Kapetanovic, E., Milanova, D., Church, G.M., and Reddy, S.T. (2022). Discovery and validation of human genomic safe harbor sites for gene and cell therapies. Cell Rep. Methods *2*, 100154. https://doi.org/10.1016/j.crmeth.2021.100154.
24. Papapetrou, E.P., and Schambach, A. (2016). Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. Mol. Ther. *24*, 678–684. https://doi.org/10.1038/mt.2016.38.
25. Büning, H., Morgan, M., and Schambach, A. (2023). Skeletal muscle-directed gene therapy: hijacking the fusogenic properties of muscle cells. Sig Transduct. Targeted Ther. *8*, 1–3. https://doi.org/10.1038/s41392-023-01584-4.
26. Wang, D., Zhong, L., Nahid, M.A., and Gao, G. (2014). The potential of adeno-associated viral vectors for gene delivery to muscle tissue. Expert Opin. Drug Deliv. *11*, 345–364. https://doi.org/10.1517/17425247.2014.871258.
27. Herzog, R.W., Yang, E.Y., Couto, L.B., Hagstrom, J.N., Elwell, D., Fields, P.A., Burton, M., Bellinger, D.A., Read, M.S., Brinkhous, K.M., et al. (1999). Long-term correction of canine hemophilia B by gene transfer of blood coagulation factor IX mediated by adeno-associated viral vector. Nat. Med. *5*, 56–63. https://doi.org/10.1038/4743.
28. Herzog, R.W., Hagstrom, J.N., Kung, S.-H., Tai, S.J., Wilson, J.M., Fisher, K.J., and High, K.A. (1997). Stable gene transfer and expression of human blood coagulation factor IX after intramuscular injection of recombinant adeno-associated virus. Proc. Natl. Acad. Sci. USA *94*, 5804–5809.

29. Kessler, P.D., Podsakoff, G.M., Chen, X., McQuiston, S.A., Colosi, P.C., Matelis, L.A., Kurtzman, G.J., and Byrne, B.J. (1996). Gene delivery to skeletal muscle results in sustained expression and systemic delivery of a therapeutic protein. Proc. Natl. Acad. Sci. USA *93*, 14082–14087.

30. Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z., et al. (2016). In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. Nature *540*, 144–149. https://doi.org/10.1038/nature20565.

31. Abdelmoez, A.M., Sardón Puig, L., Smith, J.A.B., Gabriel, B.M., Savikj, M., Dollet, L., Chibalin, A.V., Krook, A., Zierath, J.R., and Pillon, N.J. (2020). Comparative profiling of skeletal muscle models reveals heterogeneity of transcriptome and metabolism. Am. J. Physiol. Cell Physiol. *318*, C615–C626. https://doi.org/10.1152/ajpcell.00540.2019.

32. Laing, N.G., Dye, D.E., Wallgren-Pettersson, C., Richard, G., Monnier, N., Lillis, S., Winder, T.L., Lochmüller, H., Graziano, C., Mitrani-Rosenbaum, S., et al. (2009). Mutations and Polymorphisms of the Skeletal Muscle α-Actin Gene (ACTA1). Hum. Mutat. *30*, 1267–1277. https://doi.org/10.1002/humu.21059.

33. Crawford, K., Flick, R., Close, L., Shelly, D., Paul, R., Bove, K., Kumar, A., and Lessard, J. (2002). Mice lacking skeletal muscle actin show reduced muscle strength and growth deficits and die during the neonatal period. Mol. Cell Biol. *22*, 5887–5896. https://doi.org/10.1128/MCB.22.16.5887-5896.2002.

34. van Deursen, J., Heerschap, A., Oerlemans, F., Ruitenbeek, W., Jap, P., ter Laak, H., and Wieringa, B. (1993). Skeletal muscles of mice deficient in muscle creatine kinase lack burst activity. Cell *74*, 621–631. https://doi.org/10.1016/0092-8674(93)90510-W.

35. Momken, I., Lechêne, P., Koulmann, N., Fortin, D., Mateo, P., Doan, B.T., Hoerter, J., Bigard, X., Veksler, V., and Ventura-Clapier, R. (2005). Impaired voluntary running capacity of creatine kinase-deficient mice. J. Physiol. *565*, 951–964. https://doi.org/10.1113/jphysiol.2005.086397.

36. Garry, D.J., Ordway, G.A., Lorenz, J.N., Radford, N.B., Chin, E.R., Grange, R.W., Bassel-Duby, R., and Williams, R.S. (1998). Mice without myoglobin. Nature *395*, 905–908. https://doi.org/10.1038/27681.

37. Gödecke, A., Flögel, U., Zanger, K., Ding, Z., Hirchenhain, J., Decking, U.K.M., and Schrader, J. (1999). Disruption of myoglobin in mice induces multiple compensatory mechanisms. Proc. Natl. Acad. Sci. USA *96*, 10495–10500. https://doi.org/10.1073/pnas.96.18.10495.

38. John, S., Sabo, P.J., Thurman, R.E., Sung, M.-H., Biddie, S.C., Johnson, T.A., Hager, G.L., and Stamatoyannopoulos, J.A. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat. Genet. *43*, 264–268. https://doi.org/10.1038/ng.759.

39. Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82. https://doi.org/10.1038/nature11232.

40. Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res. *32*, D109–D111. https://doi.org/10.1093/nar/gkh023.

41. Thurman, R.E., Day, N., Noble, W.S., and Stamatoyannopoulos, J.A. (2007). Identification of higher-order functional domains in the human ENCODE regions. Genome Res. *17*, 917–927. https://doi.org/10.1101/gr.6081407.

42. Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc. Natl. Acad. Sci. USA *107*, 139–144. https://doi.org/10.1073/pnas.0912402107.

43. Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic Acids Res. *35*, D88–D92. https://doi.org/10.1093/nar/gkl822.

44. Reese, M.G. (2001). Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comput. Chem. *26*, 51–56. https://doi.org/10.1016/S0097-8485(01)00099-7.

45. ENCODE Project Consortium, Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710. https://doi.org/10.1038/s41586-020-2493-4.

46. Burattini, S., Ferri, P., Battistelli, M., Curci, R., Luchetti, F., and Falcieri, E. (2004). C2C12 murine myoblasts as a model of skeletal muscle development: morpho-functional characterization. Eur. J. Histochem. *48*, 223–233.

47. Concordet, J.-P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. *46*, W242–W245. https://doi.org/10.1093/nar/gky354.

48. Tsunekawa, Y., Terhune, R.K., Fujita, I., Shitamukai, A., Suetsugu, T., and Matsuzaki, F. (2016). Developing a *de novo* targeted knock-in method based on *in utero* electroporation into the mammalian brain. Development *143*, 3216–3222. https://doi.org/10.1242/dev.136325.

49. Ma, J., Chen, J., Gan, M., Chen, L., Zhao, Y., Niu, L., Zhu, Y., Zhang, S., Li, X., Guo, Z., et al. (2022). Comparison of reference gene expression stability in mouse skeletal muscle via five algorithms. PeerJ *10*, e14221. https://doi.org/10.7717/peerj.14221.

50. Li, L., Huang, C., Pang, J., Huang, Y., Chen, X., and Chen, G. (2023). Advances in research on cell models for skeletal muscle atrophy. Biomed. Pharmacother. *167*, 115517. https://doi.org/10.1016/j.biopha.2023.115517.

51. Goullée, H., Taylor, R.L., Forrest, A.R.R., Laing, N.G., Ravenscroft, G., and Clayton, J.S. (2021). Improved CRISPR/Cas9 gene editing in primary human myoblasts using low confluency cultures on Matrigel. Skelet. Muscle *11*, 23. https://doi.org/10.1186/s13395-021-00278-1.

52. Gallup, J.M., and Ackermann, M.R. (2008). The 'PREXCEL-Q Method' for qPCR. Int. J. Biomed. Sci. *4*, 273–293.

53. Wu, J., Xiao, J., Zhang, Z., Wang, X., Hu, S., and Yu, J. (2014). Ribogenomics: the science and knowledge of RNA. Dev. Reprod. Biol. *12*, 57–63. https://doi.org/10.1016/j.gpb.2014.04.002.

54. Bae, S., Park, J., and Kim, J.-S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. Bioinformatics *30*, 1473–1475. https://doi.org/10.1093/bioinformatics/btu048.

55. Adamopoulos, P.G., Tsiakanikas, P., Stolidi, I., and Scorilas, A. (2022). A versatile 5′ RACE-Seq methodology for the accurate identification of the 5′ termini of mRNAs. BMC Genom. *23*, 163. https://doi.org/10.1186/s12864-022-08386-y.

56. Schramm, G., Bruchhaus, I., and Roeder, T. (2000). A simple and reliable 5′-RACE approach. Nucleic Acids Res. *28*, e96.

57. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. *9*, ELIXIR-709. https://doi.org/10.12688/f1000research.24956.2.

58. Chen, Y., Verbeek, F.J., and Wolstencroft, K. (2021). Establishing a consensus for the hallmarks of cancer based on gene ontology and pathway annotations. BMC Bioinf. *22*, 178. https://doi.org/10.1186/s12859-021-04105-8.

59. Knijnenburg, T.A., Bismeijer, T., Wessels, L.F.A., and Shmulevich, I. (2015). A multilevel pan-cancer map links gene mutations to cancer hallmarks. Chin. J. Cancer *34*, 439–449. https://doi.org/10.1186/s40880-015-0050-6.

60. Tabebordbar, M., Lagerborg, K.A., Stanton, A., King, E.M., Ye, S., Tellez, L., Krunnfusz, A., Tavakoli, S., Widrick, J.J., Messemer, K.A., et al. (2021). Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. Cell *184*, 4919–4938.e22. https://doi.org/10.1016/j.cell.2021.08.028.

61. Martari, M., Sagazio, A., Mohamadi, A., Nguyen, Q., Hauschka, S.D., Kim, E., and Salvatori, R. (2009). Partial Rescue of Growth Failure in Growth Hormone (GH)-Deficient Mice by a Single Injection of a Double-Stranded Adeno-Associated Viral Vector Expressing the GH Gene Driven by a Muscle-Specific Regulatory Cassette. Hum. Gene Ther. *20*, 759–766. https://doi.org/10.1089/hum.2008.197.

62. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res. *43*, W39–W49. https://doi.org/10.1093/nar/gkv416.

63. Wang, G.C., and Wang, Y. (1997). Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Appl. Environ. Microbiol. *63*, 4645–4650. https://doi.org/10.1128/aem.63.12.4645-4650.1997.

64. Herzog, R.W. (2020). Encouraging and Unsettling Findings in Long-Term Follow-up of AAV Gene Transfer. Mol. Ther. *28*, 341–342. https://doi.org/10.1016/j.ymthe.2020.01.007.

65. Burdett, T., and Nuseibeh, S. (2023). Changing trends in the development of AAV-based gene therapies: a meta-analysis of past and present therapies. Gene Ther. *30*, 323–335. https://doi.org/10.1038/s41434-022-00363-0.

66. Pickar-Oliver, A., Gough, V., Bohning, J.D., Liu, S., Robinson-Hamm, J.N., Daniels, H., Majoros, W.H., Devlin, G., Asokan, A., and Gersbach, C.A. (2021). Full-length dystrophin restoration via targeted exon integration by AAV-CRISPR in a humanized mouse model of Duchenne muscular dystrophy. Mol. Ther. *29*, 3243–3257. https://doi.org/10.1016/j.ymthe.2021.09.003.

67. Stephenson, A.A., Nicolau, S., Vetter, T.A., Dufresne, G.P., Frair, E.C., Sarff, J.E., Wheeler, G.L., Kelly, B.J., White, P., and Flanigan, K.M. (2023). CRISPR-Cas9 homology-independent targeted integration of exons 1–19 restores full-length dystrophin in mice. Mol. Ther. Methods Clin. Dev. *30*, 486–499. https://doi.org/10.1016/j.omtm.2023.08.009.

68. Hakim, C.H., Kumar, S.R.P., Pérez-López, D.O., Wasala, N.B., Zhang, D., Yue, Y., Teixeira, J., Pan, X., Zhang, K., Million, E.D., et al. (2021). Cas9-specific immune responses compromise local and systemic AAV CRISPR therapy in multiple dystrophic canine models. Nat. Commun. *12*, 6769. https://doi.org/10.1038/s41467-021-26830-7.

69. Cheung, W.F., van den Born, J., Kühn, K., Kjellén, L., Hudson, B.G., and Stafford, D.W. (1996). Identification of the endothelial cell binding site for factor IX. Proc. Natl. Acad. Sci. USA *93*, 11068–11073.

70. Schuettrumpf, J., Herzog, R.W., Schlachterman, A., Kaufhold, A., Stafford, D.W., and Arruda, V.R. (2005). Factor IX variants improve gene therapy efficacy for hemophilia B. Blood *105*, 2316–2323. https://doi.org/10.1182/blood-2004-08-2990.

71. Sabatino, D.E., and Arruda, V.R. (2012). Muscle Gene Therapy for Hemophilia. J. Genet. Syndr. Gene. Ther. *1*, S1–S010.

72. Zhong, H., Ceballos, C.C., Massengill, C.I., Muniak, M.A., Ma, L., Qin, M., Petrie, S.K., and Mao, T. (2021). High-fidelity, efficient, and reversible labeling of endogenous proteins using CRISPR-based designer exon insertion. Elife *10*, e64911. https://doi.org/10.7554/eLife.64911.

73. Yarnall, M.T.N., Ioannidi, E.I., Schmitt-Ulms, C., Krajeski, R.N., Lim, J., Villiger, L., Zhou, W., Jiang, K., Garushyants, S.K., Roberts, N., et al. (2023). Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. Nat. Biotechnol. *41*, 500–512. https://doi.org/10.1038/s41587-022-01527-4.

74. Anzalone, A.V., Gao, X.D., Podracky, C.J., Nelson, A.T., Koblan, L.W., Raguram, A., Levy, J.M., Mercer, J.A.M., and Liu, D.R. (2022). Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. Nat. Biotechnol. *40*, 731–740. https://doi.org/10.1038/s41587-021-01133-w.

75. Tou, C.J., Orr, B., and Kleinstiver, B.P. (2023). Precise cut-and-paste DNA insertion using engineered type V-K CRISPR-associated transposases. Nat. Biotechnol. *41*, 968–979. https://doi.org/10.1038/s41587-022-01574-x.

76. Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V., and Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. Science *365*, 48–53. https://doi.org/10.1126/science.aax9181.

77. Lampe, G.D., King, R.T., Halpin-Healy, T.S., Klompe, S.E., Hogan, M.I., Vo, P.L.H., Tang, S., Chavez, A., and Sternberg, S.H. (2024). Targeted DNA integration in human cells without double-strand breaks using CRISPR-associated transposases. Nat. Biotechnol. *42*, 87–98. https://doi.org/10.1038/s41587-023-01748-1.

78. Zhang, X., Van Treeck, B., Horton, C.A., McIntyre, J.J.R., Palm, S.M., Shumate, J.L., and Collins, K. (2024). Harnessing eukaryotic retroelement proteins for transgene insertion into human safe-harbor loci. Nat. Biotechnol. 1–10. https://doi.org/10.1038/s41587-024-02137-y.

79. 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.

80. Liu, L., Cheung, T.H., Charville, G.W., and Rando, T.A. (2015). Isolation of skeletal muscle stem cells by fluorescence-activated cell sorting. Nat. Protoc. *10*, 1612–1624. https://doi.org/10.1038/nprot.2015.110.

81. Murach, K.A., Peck, B.D., Policastro, R.A., Vechetti, I.J., Van Pelt, D.W., Dungan, C.M., Denes, L.T., Fu, X., Brightwell, C.R., Zentner, G.E., et al. (2021). Early satellite cell communication creates a permissive environment for long-term muscle growth. iScience *24*, 102372. https://doi.org/10.1016/j.isci.2021.102372.

82. Murach, K.A., Vechetti, I.J., Van Pelt, D.W., Crow, S.E., Dungan, C.M., Figueiredo, V.C., Kosmac, K., Fu, X., Richards, C.I., Fry, C.S., et al. (2020). Fusion-Independent Satellite Cell Communication to Muscle Fibers During Load-Induced Hypertrophy. Function (Oxf) *1*, zqaa009. https://doi.org/10.1093/function/zqaa009.

83. Giannoukos, G., Ciulla, D.M., Marco, E., Abdulkerim, H.S., Barrera, L.A., Bothmer, A., Dhanapal, V., Gloskowski, S.W., Jayaram, H., Maeder, M.L., et al. (2018). UDiTaS™, a genome editing detection method for indels and genome rearrangements. BMC Genom. *19*, 212. https://doi.org/10.1186/s12864-018-4561-9.

84. Olarte-Castillo, X.A., Goodman, L.B., and Whittaker, G.R. (2024). Molecular detection using hybridization capture and next-generation sequencing reveals cross-species transmission of feline coronavirus type-1 between a domestic cat and a captive wild felid. Microbiol. Spectr. e0006124. https://doi.org/10.1101/2024.01.02.573944.

85. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

86. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620. https://doi.org/10.1093/bioinformatics/btt593.

87. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499. https://doi.org/10.1101/gr.209601.116.

88. Vasimuddin, M., Misra, S., Li, H., and Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS), pp. 314–324. https://doi.org/10.1109/IPDPS.2019.00041.

89. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

90. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative Genomics Viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

91. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

92. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics *21*, 3439–3440. https://doi.org/10.1093/bioinformatics/bti525.