Molecular Therapy
Nucleic Acids
Original Article

Orthogonal characterization of rAAV reveals vector attributes that drive ITR repair, self-complementary genome formation, and transgene expression

Carla Llinares,[1] Suk Namkung,[3,4] Mitchell Yip,[3] África Vales,[1] Cristina Olagüe,[1] Itsaso Mauleón,[1] Sergio Isola,[1] Pedram Moeini,[1] Gloria González-Aseguinolaza,[1,2] Phillip W.L. Tai,[3,4,5] and Carmen Unzu[1]

[1]DNA & RNA Medicine Division, Centro de Investigación Médica Aplicada, CIMA, Pamplona, Navarre 31008, Spain; [2]IdiSNA, Navarra Institute for Health Research, Pamplona, Navarre 31008, Spain; [3]Horae Gene Therapy Center, UMass Chan Medical School, Worcester, MA 01605, USA; [4]Department of Genetic and Cellular Medicine, UMass Chan Medical School, Worcester, MA 01605, USA; [5]Department of Microbiology, UMass Chan Medical School, Worcester, MA 01605, USA

Recombinant adeno-associated virus (rAAVs) vectors are the flagship vehicles for delivering DNA payloads for human gene therapy. However, only a few outstanding therapies have reached the market in the past decade. One reason for the slow development of rAAV-based gene therapies has been our limited knowledge of basic AAV biology. Therefore, the goal of this work was to investigate the influence of rAAV inverted terminal repeat (ITR) design, promoter use, and serotype selection on vector genome heterogeneity and transduction efficiency. Our analyses revealed additional, larger and smaller, vector genome species that were identified as unresolved and bound by the ITRs. Alkaline gel analysis and long-read single-molecule sequencing confirmed the presence of double-stranded genomes with self-complementary (sc) configurations for most of the serotypes and ITR designs tested. sc genome formation and ITR repair were favored under specific capsid-genome feature combinations, resulting in genomes larger than 5 kb in some cases. Furthermore, we found that vector genome length had an impact on ITR repair. Finally, transduction studies confirmed that the AAV capsid plays a role in functional ITR repair and potency. In conclusion, sc genome packaging is favored under specific ITR designs, and their stability is influenced by the AAV capsid.

INTRODUCTION

The development of recombinant adeno-associated virus (rAAV)-based gene therapies has heralded a new era in the treatment of genetic disorders, offering the promise of long-term and potentially curative interventions.[1] Despite their therapeutic potential, the translation of rAAV medicines from bench to bedside is still significantly delayed by our incomplete understanding of AAV biology. First discovered in 1965, AAVs are small (~25 nm), nonpathogenic *dependoparvoviruses* that have a single-stranded (ss)DNA genome protected by an icosahedral protein capsid.[2] They have been proven to be highly suitable vehicles for delivering DNA payloads, because they are easy to engineer and can efficiently deliver genetic material to a wide range of tissues.[3]

To generate a rAAV for gene therapy, the wild-type (wt)AAV genome is modified to remove all viral coding sequences, namely *rep* and *cap*. The only viral elements that are retained are the two T-shaped inverted terminal repeats (ITRs) that flank the therapeutic expression cassette and are needed for correct packaging of the AAV genomes.[4,5] During rAAV production, *rep* and *cap* are provided in *trans* to ensure correct replication and encapsidation of the transgene cassette, which is generally flanked by ITRs from AAV serotype 2 (AAV2).[6] Traditional ssAAV vectors require host cell machinery to synthesize the complementary DNA strand, a process that limits the speed and efficiency of transgene expression.[7,8] In contrast, self-complementary (sc)AAVs,[9–11] lack the terminal resolution site (*trs*) from one ITR (Figure S1A),[12] which generates a double-stranded DNA genome that results in higher vector potency.[13] It has also been described that when the replicating genome is half the length of the wtAAV genome (~4.7 kb), both single- and double-stranded genomes can become packaged during production in the presence of a helper virus.[13,14] Notably, the efficiency of this process is variable and inversely proportional to the genome size, which also affects vector yields during AAV production.[14] Several groups have generated novel ITR designs that would result in better double-stranded genome packaging and titers.[15–17] However, how novel ITR designs interact with the rest of the vector genome elements or the capsid, or how they are repaired, is still poorly understood.

Another bottleneck is the production of empty and partially loaded capsids,[18,19] which is a natural part of wtAAV's life cycle. Therefore, unintentionally encapsulated truncated AAV genomes or DNA impurities are also generated during the rAAV manufacturing process.[20–22] rAAV genome truncations result in shortened vector

Received 1 October 2025; accepted 10 March 2026;
https://doi.org/10.1016/j.omtn.2026.102899.

Correspondence: Phillip W.L. Tai, Horae Gene Therapy Center, UMass Chan Medical School, Worcester, MA 01605, USA.
E-mail: phillip.tai2@umassmed.edu
Correspondence: Carmen Unzu, DNA & RNA Medicine Division, Centro de Investigación Médica Aplicada, CIMA, Pamplona, Navarre 31008, Spain.
E-mail: cunzu@unav.es

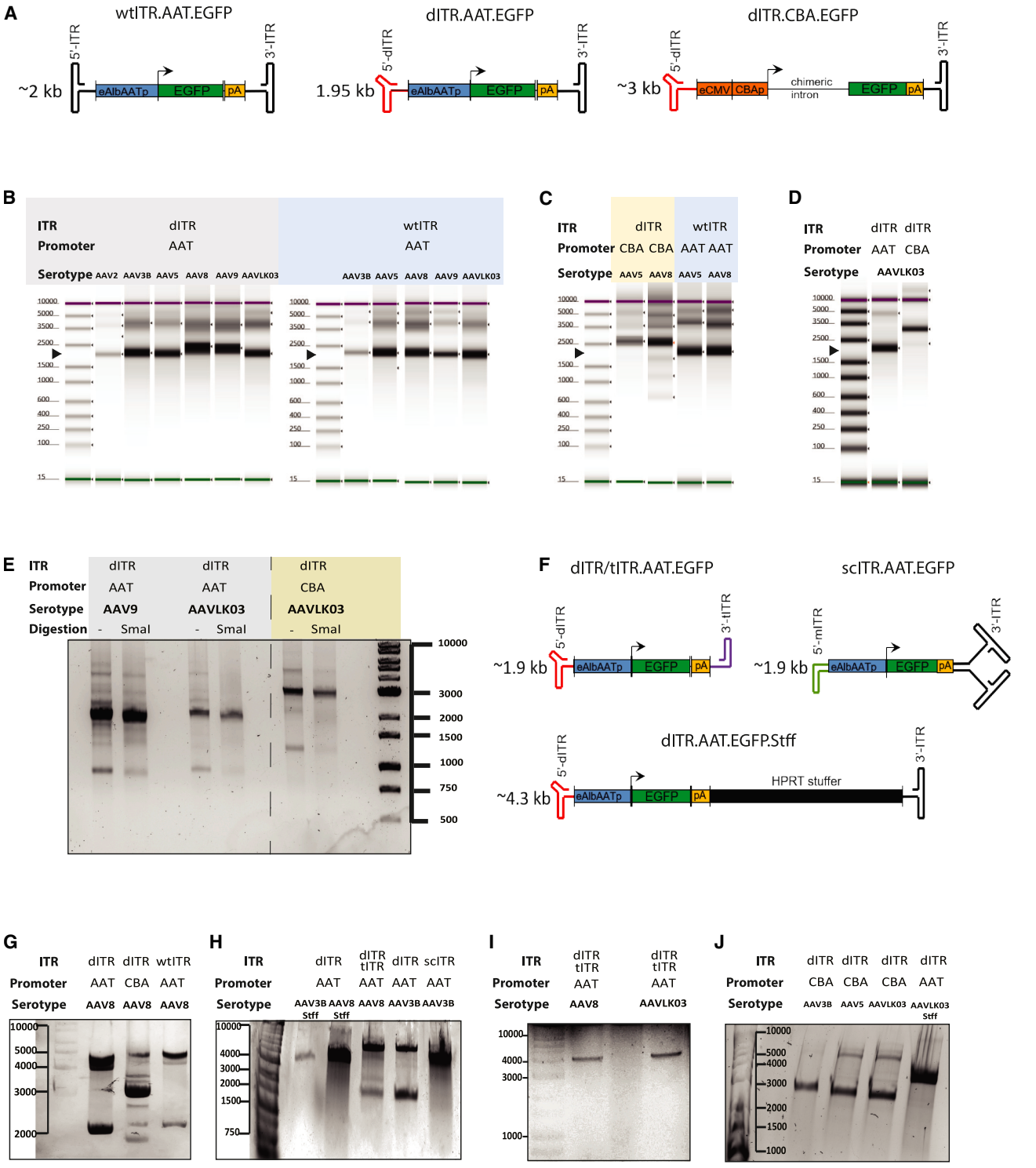


Figure 1. Recombinant AAV genome element and heterogeneity analysis by native and alkaline electrophoresis

(A) Graphic representation of the main recombinant AAV vector genome elements evaluated: wtITR (143 nt, black) versus dITR/wtITR (11-nt deletion in the 5′ C arm, 130 nt, red), *EAlbAAT* (albumin enhancer-alpha1-antitrypsin promoter, blue), *Egfp* (enhanced green fluorescent protein, green), pA (bovine/human growth hormone polyadenylation signal, yellow), and *CBA* (CMV (cytomegalovirus) enhancer-chicken beta actin promoter, orange). Vector genome sizes are shown next to each representation. (B) Native

(legend continued on next page)

genomes that do not have a therapeutic effect, decreasing the proportion of functional vectors and the potency of the therapy. This process is not well understood, and the factors involved in genome truncation and heterogeneity remain veiled.

These gaps in knowledge pose substantial challenges to the rational optimization of rAAV vectors. Therefore, this study aims to shed light on the elements that cause heterogeneity by using complementary analytical techniques to probe specific vector genome designs, their interaction with the rAAV capsid, and the final impact on transgene expression.

RESULTS

Combination of vector elements that affect heterogeneity

First, we analyzed the integrity of rAAV-e*nhanced green fluorescent protein* (*Egfp*) vectors containing either an alpha-1 antitrypsin (*AAT*) promoter (2,011-nt vector genome) or a chicken beta-actin (*CBA*) promoter (2,917-nt vector genome). These constructs were flanked by two different ITR designs: two wtITRs, or a commercially available 11-nt deletion in the C arm of the 5′ ITR (dITR/wtITR) and a 3′ wtITR (Figures 1A and S1A). To factor in the influence of the capsid in rAAV heterogeneity, the constructs were packaged within six different serotypes: AAV2, AAV3B, AAVLK03, AAV5, AAV8, and AAV9. AAV preparations were treated with DNaseI for 1 h at 37°C, and vector genomes were extracted using a commercial kit. Heterogeneity was first analyzed by native automated electrophoresis. Most of the genomes with the *AAT* promoter migrated at their expected size (∼2 kb, black arrows), but additional larger bands were also identified (Figure 1B). The majority of the bands were consistent across serotypes, with the second most abundant one being twice the size of the 2 kb genomes. When AAV5, AAV8, and AAVLK03 serotypes harboring the *CBA* promoter were compared, we also observed that the banding pattern was more heterogeneous with the AAV8.*CBA*.*Egfp* construct (Figures 1C and 1D). These results were confirmed by quantitative PCR (qPCR) vector genome quantification, with primers aligning to different regions in the genome (Figures S1B and S1C). Vector genome quantification at the polyA site was 1 log lower in the constructs harboring the *CBA* promoter compared to genome quantification in the ITR region, something that we did not observe in the constructs with the *AAT* promoter. Based on the detection of larger genomes, we hypothesized that the additional bands could be DNA species that were DNaseI-resistant, un-nicked by Rep (unresolved), and bound at the ITRs. To test this hypothesis, the vector genomes were digested with the SmaI restriction enzyme, which cuts in the C arm. After digestion, most of the additional bands were eliminated (Figures 1E and S1D), suggesting that these species were indeed unresolved genomes bound by ITRs.

Generation of sc genomes depends on vector size, sequence, and ITR design

To explore the impact of the ITRs on the generation of larger genome species, we produced additional constructs flanked by different ITR designs: a 22-nt deletion in the B arm of the 3′-ITR (dITR/tITR), a sc ITR lacking the *trs* (scITR), and a dITR.*EalbAAT*.*Egfp*.*Stff* construct that includes a stuffer sequence to expand the genome length to 4.3 kb (Figure 1F). As previously described, alkaline gel analysis revealed ss and double-stranded genomes for compact vectors carrying the wtITRs or the dITR design (<2 kb) packaged with AAV8, AAV3B, or AAV5 (Figures 1G and 1H). However, vector genomes including the dITR/tITR showed mostly double-stranded genomes, such as in AAV8 and AAVLK03 serotypes (Figures 1H and 1I), although we observed variations in the proportion of double-stranded forms depending on the AAV8 production batch (Figure 1H versus Figure 1I). Of note, productions harboring the dITR/tITR design always presented lower titers compared to the other ITR designs (Table S1). Again, genome heterogeneity was observed for the AAV8.*CBA*.*Egfp* vector (Figure 1G) but not for the AAV3B, AAVLK03, or AAV5.*CBA* constructs (Figure 1J). Interestingly, the latter two showed a faint band larger than 5 kb that would correspond to a 3 kb double-stranded genome. As expected, the 4.3 kb stuffer constructs with AAV3B, AAVLK03, and AAV8 capsids showed mainly ss forms (Figures 1H–1J).

To interrogate the effect of the transgene sequence, we swapped the reporter *Egfp* transgene with the clinically relevant human gene survival motor neuron 1 (*SMN1*) to keep a similar cargo size to *Egfp*. Newly produced vectors did not show differences in yields compared to *Egfp* constructs (Table S1). Alkaline gel analysis showed similar formation of ss and double-stranded genomes when the *SMN1* gene was included in constructs packaged within an AAV8 capsid and flanked by the different ITR designs (Figure S2A), although we identified an extra 2.8 kb band in the dITR/tITR vector preparation. Given its therapeutic relevance, we also analyzed the promoter-transgene combination used in the approved gene therapy onasemnogene abeparvovec,

automated electrophoresis of extracted rAAV genomes from dITR/wtITR and wtITR.*EalbAAT*.*Egfp* vectors by TapeStation. From left to right: AAV2, AAV3B, AAV5, AAV8, AAV9, and AAVLK03 serotypes. As a run reference, the upper marker (10,000 bp) and lower marker (15 bp) of the ladder are highlighted in purple and green, respectively. (C) Native automated electrophoresis of extracted rAAV genomes from AAV5 and AAV8 serotypes with dITR vs. wtITR and *CBA* versus *EalbAAT* promoters. As a run reference, the upper marker (10,000 bp) and lower marker (15 bp) of the ladder are highlighted in purple and green, respectively. (D) Native automated electrophoresis of extracted rAAV genomes from the AAVLK03 serotype with dITR and *CBA* vs. *EalbAAT* promoters. As a run reference, the upper marker (10,000 bp) and lower marker (15 bp) of the ladder are highlighted in purple and green, respectively. (E) Native electrophoresis of extracted rAAV genomes from AAV9 and AAVLK03 serotypes with dITR and *CBA* versus *EalbAAT* promoters, before and after digestion at 37°C for 1 h with SmaI. (F) Graphic representation of the dITR/tITR design (11-nt deletion in the 5′ C arm and 22-nt deletion of the 3′ B arm, 121 nt, purple), scITR (C arm deletion, 106 nt, green), and the dITR.*AAT*.*Egfp* expression cassette including an hypoxanthine phosphoribosyltransferase 1 (*HPRT1*) DNA stuffer sequence (GenBank: NG_012329.2) (bottom). Vector genome sizes are shown next to each design. (G) Alkaline electrophoresis of AAV8 vectors with different ITR configurations and *CBA* versus *EalbAAT* promoters. (H) Alkaline electrophoresis of AAV8 and AAV3B vectors with dITR/wtITR, dITR/tITR, and scITR configurations. (I) Alkaline electrophoresis of AAV8 and AAVLK03 vectors with the dITR/tITR configuration. (J) Alkaline electrophoresis of AAV3B, AAV5, and AAVLK03 vectors with dITR/wtITR configuration and *CBA* versus *EalbAAT* promoters. Combinations are displayed on top of each lane. Agarose gel ladder: FastGene 1 kb DNA marker (10,000–100 bp).

an AAV9.*CBA.SMN1*, and compared its vector genome integrity to AAV9.*CBA.Egfp* or AAV9.*AAT.Egfp* (Figure S2B), although our construct was ss (wtITRs), and the approved therapeutic vector is sc. Again, our alkaline gel results showed no differences in genome heterogeneity between *Egfp* or *SMN1* constructs.

To validate these findings, long-read, single-molecule, real-time (SMRT) sequencing was performed on vector genomes with different ITR and promoter designs extracted from AAV3B, AAVLK03, and AAV8 serotypes (Figure 2). Whereas the 2 kb AAV3B vector preparation showed a similar proportion of ss and double-stranded genomes, AAVLK03 appeared to be mostly double-stranded (Figure 2A). Long-read sequencing also confirmed the presence of genome dimers in vectors with the *CBA* promoter that were over 6 kb in size (Figures 2A–2C). Genome heterogeneity was also analyzed by examining the most frequent start and end positions within the whole genome (Figures 2B–2D). Aligning with the alkaline gels, SMRT sequencing of the AAVLK03 and AAV8 dITR/tITR and scITR designs showed a unique peak at 4 kb (Figures 2A–2C). Both the AAV3B and the AAVLK03.*CBA.Egfp* genomes showed hotspots for truncations within the *CBA* promoter, the chimeric intron, and the *HGH* polyA (Figure 2B). Similar truncated genomes and start/end hotspots were found within the AAV8.*CBA.Egfp* genome (Figure 2D). Interestingly, both the scITR.AAVLK03 and AAV8 genomes showed two dominant starting positions located at the left (5′) and right (3′) ITRs but only one stop position at the right ITR (Figures 2B–2D).

Noteworthy, we were able to reproduce these results with the AAV8.*AAT.SMN1* vectors flanked by the different ITR designs, and the AAV9.*CBA.SMN1* vector, using Oxford Nanopore long-read sequencing (Figures S2C–S2F). A similar proportion of ss and double-stranded genomes was found with wtITRs or dITR/wtITR, a higher proportion of double-stranded genomes was quantified with the dITR/tITR design, and the scITR design resulted in 100% double-stranded genomes (Figures S2C and S2D). As previously shown with the *Egfp* constructs, a high proportion of truncated genomes originating from the *CBA* promoter was found in the AAV9.*CBA.SMN1* vector (Figures S2E and S2F). Furthermore, long-read sequencing allowed us to identify the extra band observed in the dITR/tITR alkaline gel as contaminating backbone reverse packaging between the ITRs (Figure S2D), something that was also detected to a lower extent (<2%) in the AAV9.*CBA.SMN1* preparation (Figure S2F).

These results highlight the impact of the ITR design and the *CBA* promoter on recombinant AAV genome heterogeneity and the sensitivity of long-read sequencing to detect truncated forms and contaminants.

Intermolecular species form before DNA extraction and long-read sequencing library preparation steps

To rule out that the detected dimers, especially the oversized ones, were DNaseI-resistant unpackaged genomes undergoing replication during AAV production, we quantified DNA levels for different AAV8 and AAVLK03 preparations by fluorimetry (Qubit) before (Table 1) and after (Table 2) DNaseI treatment. Interestingly, before DNaseI digestion, DNA levels were always below the limit of detection for the AAVLK03 preparations but not for the AAV8 preparations, which varied between 0.14 and 0.9 ng/μL (Table 1). This result may have been associated with the higher AAV8 production yields versus AAVLK03 (Table S1). Furthermore, we tested DNaseI ability to digest increasing concentrations of a pAAV plasmid DNA spike-in and compared it to levels after Benzonase digestion (Table 2). As previously described, Benzonase proved to be more effective than DNaseI for digesting increasing concentrations of spiked-in plasmid DNA under their respective conditions for both preparations (Table 2). However, DNA was still detectable in the AAV8 preparation after digestion, whereas the levels in the AAVLK03 preparation were below the limit of detection. The heterogeneous banding pattern observed in the AAV8.*CBA.Egfp* productions was not affected by DNaseI treatment (Figures S3A and S3B), but preparations were cleaner after the digestion. Moreover, we did not observe any differences in the banding pattern after DNaseI or Benzonase treatment (Figure S3C), confirming that the identified rAAV genomes were packaged or protected by the capsid.

To determine whether the detection of oversized genomes was due to artifacts from target ligation or recombination during SMRT library preparation steps, the 2 and 3 kb vector preparations were mixed, and vector genomes were extracted before library preparation (Figure 3A). Fragment analysis was performed right after vector genome extraction, and two independent peaks corresponding to the 2 and the 3 kb genomes were clearly identified (Figure S3D). After SMRT sequencing, clear individual peaks corresponding to the ss genomes and their respective double-stranded forms were detected, with no sign of recombination between the two genomes (Figure 3B). Further alignment of the dimers to the reference sequence (Figures 3C and 3D) or to the 2 kb plus 3 kb combination reference sequences (Figures 3E and 3F) confirmed these results. Therefore, oversized intermolecular dimers are formed during AAV production, although our experiments do not have the resolution to prove whether they are fully or partly packaged.

ITR repair is affected by rAAV design

Because AAVLK03 exhibited the most consistent genome formation independently of the features analyzed, and AAV8 showed substantial variability depending on genome characteristics, we performed a bioinformatic analysis of the different ITR species from AAV8 and AAVLK03 extracted genomes and compared them to their original sequence by design (Figures 4A, 4B, and S1A). From the total number of ITR counts, we observed a high percentage of unresolved wtITR (u-wtITRs) in most AAV8 preparations (165-nt dark blue peaks in Figure 4C). Unresolved ITRs are generated when Rep-dependent terminal resolution fails or is incomplete during AAV DNA replication, resulting in complementary D arms and leading to the packaging of replication intermediates.[23] This species was particularly abundant in the *CBA* construct, associating u-wtITRs with higher heterogeneity.

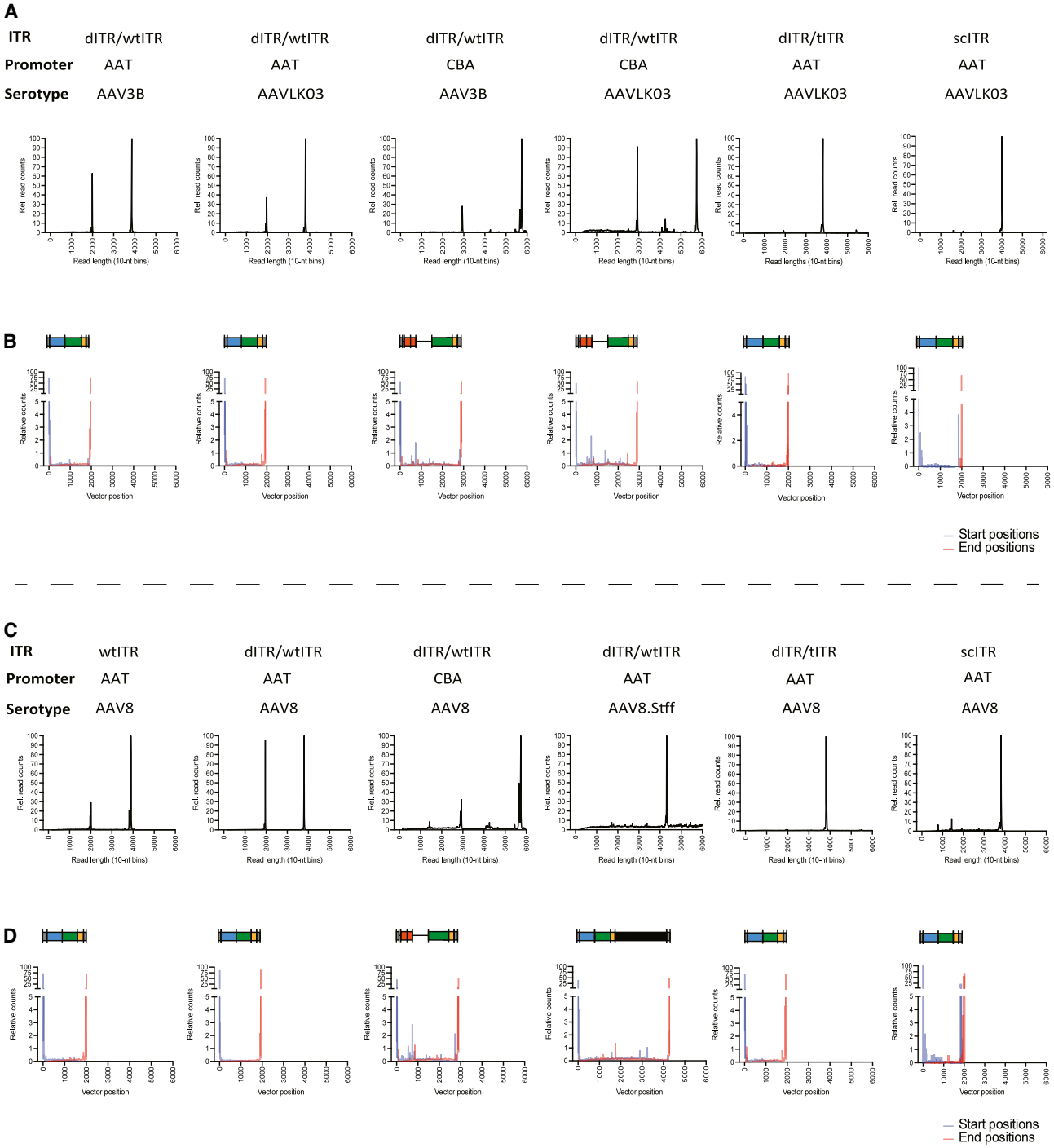


Figure 2. Recombinant AAV genome element effect and heterogeneity analysis by long-read single molecule, real-time SMRT sequencing

(A) Read counts and length of rAAV3B and rAAVLK03 vector genomes encoding the *Egfp* transgene with dITR, dITR/tITR, and scITR designs, and *AAT* versus *CBA* promoters. (B) Identification of vector genome start and end sequencing positions from the AAV3B and AAVLK03 constructs (displayed on top). Start positions are labeled in blue, and end positions are labeled in red. Quantification of read counts is displayed in the *Y* axis. Vector genome elements are displayed on top as a reference. (C) Read counts and length of rAAV8 vector genomes encoding the *Egfp* reporter transgene with different ITR designs (from left to right: wtITR, dITR/wtITR, dITR/tITR, and scITR) under the control of the *EAlbAAT* or *CBA* promoters. The dITR.AAT.*Egfp*.*Stuffer* vector is also included. (D) Identification of vector genome start and end sequencing positions from the aforementioned AAV8 constructs. Start positions are labeled in blue, and end positions are labeled in red. Quantification of read counts is displayed on the *Y* axis. Vector genome elements are displayed on top as a reference.

Table 1. DNA concentrations in AAV8 and AAVLK03 vector productions before DNaseI treatment

Vector production	DNA (ng/μL)
dITR.AAV8.*CBA.Egfp*	0.978
wtITR.AAV8.*AAT.Egfp*	0.267
dITR.AAV8.*AAT.Egfp*	0.146
scITR.AAV8.*AAT.Egfp*	0.142
tITR.AAV8.*AAT.Egfp*	0.162
dITR.LK03.*CBA.Egfp*	<0.1
wtITR.AAVLK03.*AAT.Egfp*	<0.1
dITR.AAVLK03.*AAT.Egfp*	<0.1
scITR.AAVLK03.*AAT.Egfp*	<0.1
tITR.AAVLK03.*AAT.Egfp*	<0.1

Table 2. DNA concentrations in AAV8 and AAVLK03 vector productions with spiked-in DNA before and after DNaseI and Benzonase treatment

	Pre-treatment		Post-DNaseI		Post-Benzonase	
Spiked-in DNA	Average	*SD*	Average	*SD*	Average	*SD*
AAV8						
0 ng/μL	0.045	*0.057*	0.082	*0.085*	<0.05	*ND*
0.05 ng/μL	0.053	*0.074*	<0.05	*ND*	<0.05	*ND*
0.1 ng/μL	0.064	*0.090*	0.241	*0.089*	0.028	*0.176*
0.2 ng/μL	0.656	*0.473*	0.252	*0.094*	0.092	*0.001*
0.5 ng/μL	0.528	*0.138*	0.291	*0.121*	0.186	*0.117*
AAVLK03						
0 ng/μL	<0.05	*ND*	<0.05	*ND*	<0.05	*ND*
0.05 ng/μL	<0.05	*ND*	<0.05	*ND*	<0.05	*ND*
0.1 ng/μL	0.06	*ND*	0.153	*0.055*	<0.05	*ND*
0.2 ng/μL	0.08	*ND*	0.167	*0.064*	<0.05	*ND*
0.5 ng/μL	0.5	*ND*	0.159	*0.048*	<0.05	*ND*

SD, standard deviation; *ND*, not detected.

Interestingly, this proportion was consistently lower in genomes packaged within the AAVLK03 serotype (Figure 4D). Furthermore, we observed more than 70% 5′-dITR repair to the wtITR sequence in the case of the compact genomes with the *AAT* promoter (Figure 4B) and nearly 100% repair in constructs with the *CBA* promoter or constructs with stuffer sequences. This finding suggests that vector genome length plays a role in ITR repair. In the dITR/tITR design, the 3′-tITR was 70% converted to the 5′-dITR, and *de novo* wtITR forms were also detected in a small percentage of genomes (Figures 4B–4D). Most strikingly, we did not observe any kind of repair of the scITR form despite having a 3′-wtITR as a template (Figure 4B). The same ITR repair pattern was reproduced in the *SMN1* constructs (data not shown). These results demonstrate that the ability of rAAVs to efficiently repair truncated ITRs depends on the genome length and that this repair can be influenced by the capsid.

rAAV potency is dependent on ITR and serotype combination

Next, we studied the effect of ITR repair and sc genome formation on AAV potency. Huh7 cells were transduced with increasing doses of AAV8 or AAVLK03 vector genomes (vg) per cell (10^3, 10^4, and 10^5 vg/cell) harboring the wtITR, dITR/wtITR, scITR, and dITR/tITR designs expressing the *Egfp* transgene under the control of the *AAT* promoter (Figure 5). AAV8 flow cytometry results showed the highest EGFP expression with the scITR design, compared to the other designs, at all vg/cell tested, with no differences in fluorescence intensity (Figures 5A–5C). No significant differences were found between the wtITR and dITR/wtITR expression levels, consistent with the ITR repair results (Figure 4C) and their similarities in the proportion of single- and double-stranded genomes (Figure 2). Finally, the dITR/tITR vector design showed a 2–2.5× increase in EGFP expression in cells transduced with AAV8 that was stronger at the highest dose (Figures 5A–5C and 5I), aligning with the fact that these vector genomes were mostly sc (Figures 1I and 2C). To confirm that there were no differences in the transduction pattern between different transgenes *in vitro*, Huh7 cells were also transduced with the AAV8.*AAT.SMN1* vectors at 10^5 vg/cell (Figure S4), showing similar results.

Interestingly, the AAVLK03 vector designs displayed different transduction dynamics (Figures 5E–5G and 5I). First, more vg entered the cells compared to the AAV8 vector (Figures 5D–5H), with comparable amounts of vg/ng of DNA between AAV8 at a dose of 10^5 vg/cell and AAVLK03 at a dose of 10^3 vg/cell (Figures 5D and 5H). These results were expected based on the well-known strong transduction of human hepatic cells *in vitro* by AAVLK03.[24] Nevertheless, as opposed to AAV8, the scITR did not outperform the wtITR design, showing similar potency with this serotype at all the tested doses despite its double-stranded genome (Figure 2). Noteworthy, the dITR/tITR design was the most potent at the lowest dose and the brightest at the other two doses (Figures 5E–5G). AAVLK03 reached a transduction plateau at the highest dose (Figure 5G), although EGFP-positive cells were at least 2× brighter with the dITR/tITR construct compared to the other design. In this case, the AAVLK03 dITR/wtITR combination negatively affected transgene expression compared to the use of wtITRs (Figures 5E and 5F). Altogether, these results confirm that functional ITR repair has an impact on AAV potency, but it is also dependent on the AAV serotype.

To validate the translational application of our findings, we dosed 3×10^{10} AAV8.*AAT.SMN1* vg with the different ITR designs (wtITR, dITR/wtITR, scITR, and dITR/tITR) intravenously into mice, and expression was analyzed 14 days after dosing (Figure 6). Only AAV8 vectors were used, since AAVLK03 transduces mouse cells but does not express the transgene due to epigenetic silencing.[25] In alignment with the results obtained *in vitro*, we observed the highest *SMN1* mRNA expression in the mouse livers from the scITR construct, followed by the dITR/tITR (Figure 6A). The lowest expression was obtained from the wtITR and dITR/wtITR vectors, with no significant differences between the latter two (Figure 6A).

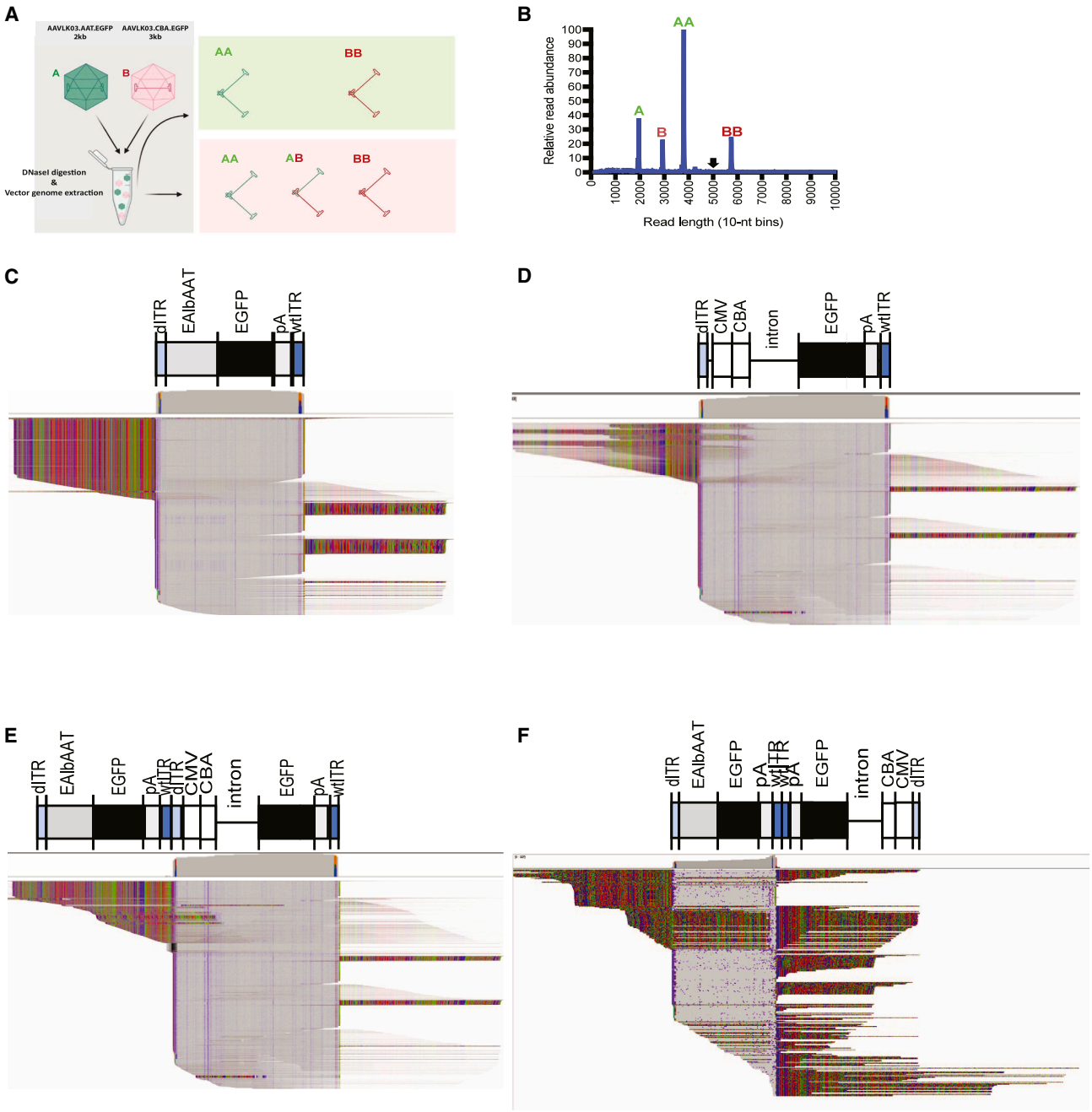
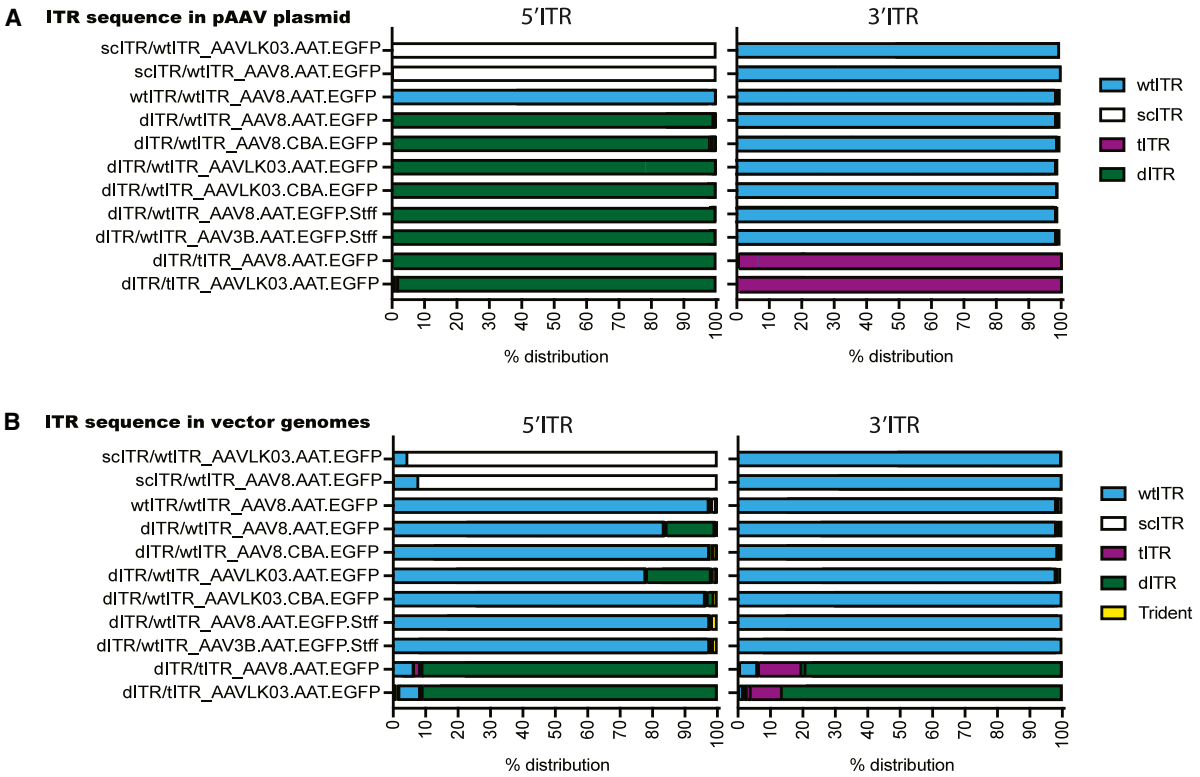
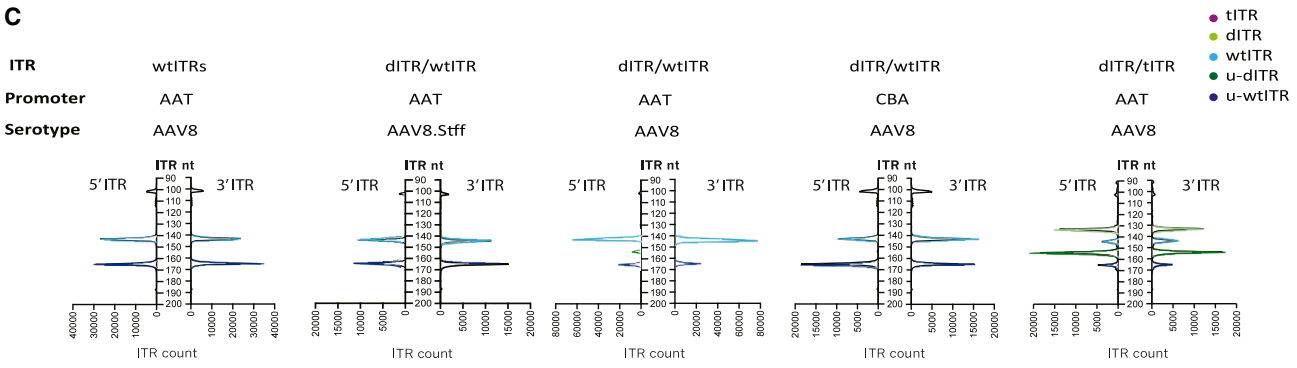


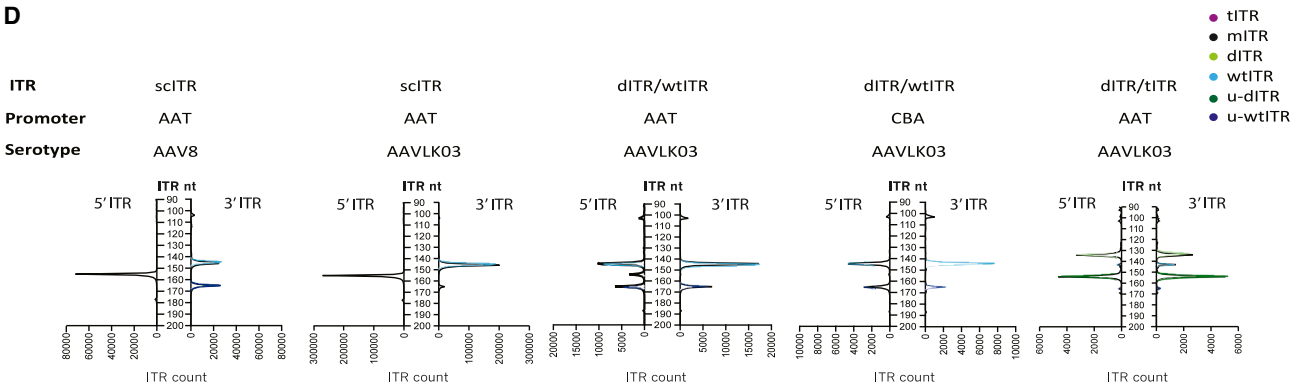
Figure 3. Impact of vector genome extraction and SMRT sequencing library preparation on dimer formation

(A) Graphical summary of the experiment and possible outcomes depending on the timing of dimer generation. Briefly, two rAAVLK03 vectors with different genomes (*AAT.Egfp*, 2 kb, and *CBA.Egfp*, 3 kb) were mixed in a 1:1 proportion, treated with DNaseI, and vector genomes were extracted. (B) Read counts and lengths of the different species identified in the sample after extraction. (C) Integrative Genomics Viewer (IGV) alignment to the dITR.*EAlbAAT.Egfp* reference genome. (D) IGV alignment to the dITR.*CBA.Egfp* genome. (E) IGV alignment to the dITR.*EAlbAAT.Egfp*-dITR.*CBA.Egfp* same 5′-3′ orientation dimer genome. (F) IGV alignment to the 5′ dITR.*EAlbAAT.Egfp* 3′-3′dITR.*CBA.Egfp*-5′ complementary dimer genome.

Because fewer vg were quantified in the mice dosed with the dITR/wtITR vector, normalization of mRNA expression to vg in the liver was performed, but it did not change the expression pattern (Figure 6B). SMN1 immunohistochemistry staining in the livers confirmed the results at the protein level (Figure 6C) and the transduction pattern observed *in vitro* (Figure S4).







(legend on next page)

DISCUSSION

In the last decade, rAAV-mediated gene therapy has made significant advances, transforming the landscape of genetic medicine. However, harnessing the full potential of rAAVs remains an ongoing challenge as a result of our limited understanding of how vector design features can impact the potency of these biological drugs. Therefore, a systematic and complementary characterization of different design elements used in rAAV gene therapy is critical for identifying the key components that permit proper vector packaging. By following this strategy, we were able to unravel specific interactions between the ITRs, the vector genome, and the capsid that directly affect rAAV potency.

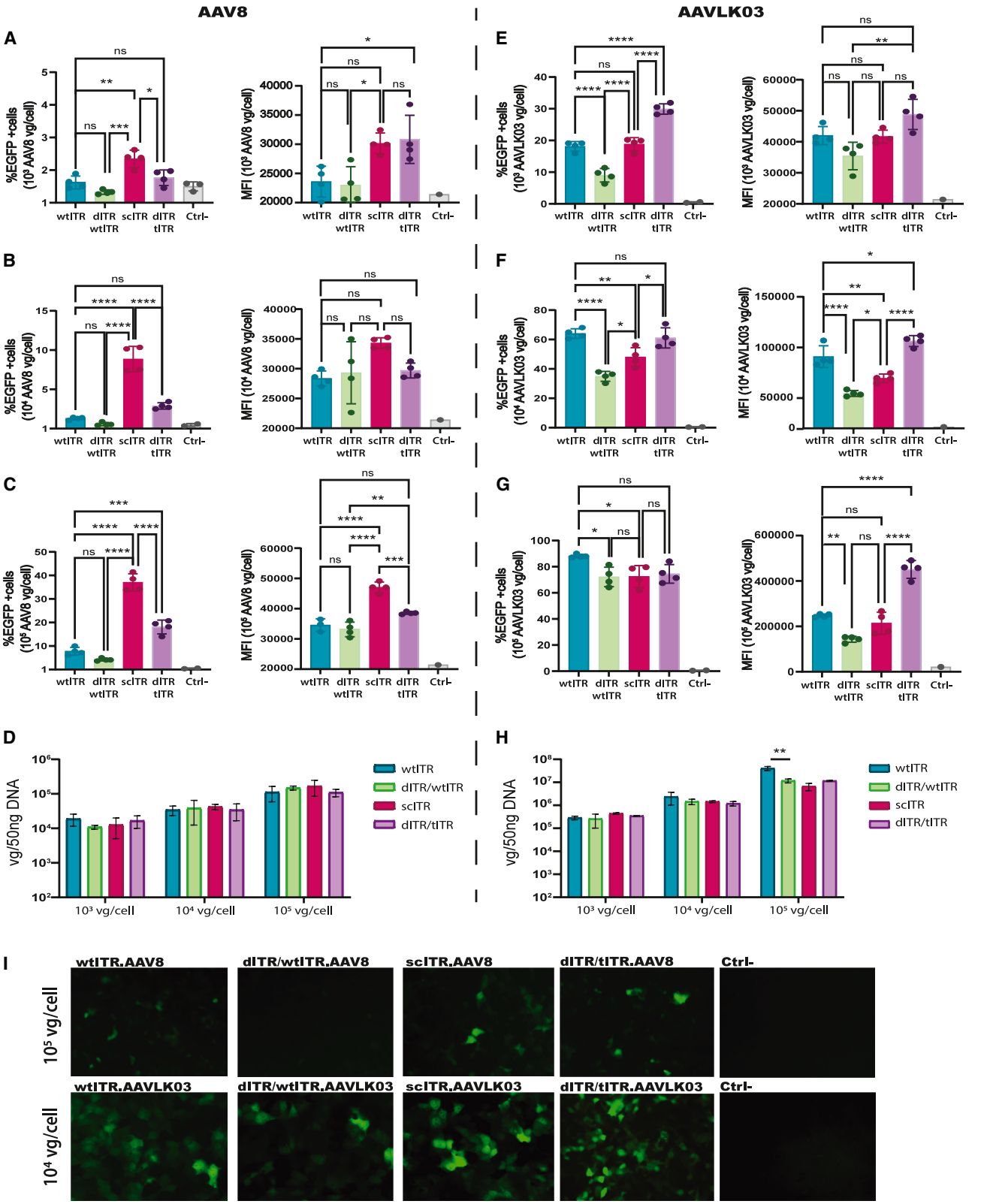
First, we looked into vector genome heterogeneity across samples, where we found a high abundance of unresolved dimer species bound by the ITRs, including oversized genomes. These sc AAV forms can be produced if one of the ITRs lack the *trs* sequence in the D arm. However, a compact-sized genome can also trigger spontaneous generation of sc species.[9,14,26] These occurrences are also limited by the size of the genome (<3.3 kb). However, we were able to identify, by alkaline gels and long-read sequencing, sc genomes larger than previously described. Our rAAV production protocol is based on adherent HEK293T cells, where vector particles are collected from both the supernatant and the cells and purified by iodixanol gradient ultracentrifugation. In this protocol, we have two DNaseI treatment steps: one right after particle collection and one before vector genome extraction. The DNaseI concentration and duration of the treatment are based on standard protocols.[27,28] Although we cannot rule out insufficient DNase or Benzonase digestion as the reason for detecting oversized genomes, our results with AAVLK03, where no DNA can be detected after digestion, but oversized genomes are present, suggest that these vector genomes escaped detection or digestion because they are at least partly protected by the protein capsid. Given this theory, there may be serotypes that are *stickier* or more efficient at packaging their genome compared to others, such as AAVLK03 or AAV5; therefore, the presence of oversized sc genomes and their potential functionality should not be overlooked. Second, we have confirmed that the use of the ubiquitous *CBA* promoter-intron sequence generated more truncated genomes compared to the tissue-specific *AAT* promoter. In a complementary study, Xie et al. demonstrated that hairpin-like sequences within an rAAV8 generated intramolecular, double-stranded genomes, and that both the *CMV* enhancer and the *CBA* promoter are rich in secondary structures, which trigger vector genome truncations.[29] We have been able to reproduce this effect with two independent rAAV8 productions (Table S1; Figure 1). Noteworthy, when the same genomes were packaged within other serotypes, such as AAV9, AAV3B, AAVLK03, or AAV5, the truncation events were strongly reduced and were only detected by the sensitivity of PacBio SMRT and Oxford Nanopore long-sequencing.

Besides vector genome heterogeneity, long-read sequencing allowed for the detection of reverse-packaged genomes in the dITR/tITR *SMN1* and *CBA.SMN1* productions (Figure S2). Noteworthy, this is a type of contaminant we have only detected in unstable cassette designs: on one hand, the dITR/tITR design affects rAAV replication, resulting in lower yields (Table S1) and variable genome formation (Figures 1H and 1I). On the other hand, 57% of the packaged *CBA.SMN1* genomes were truncated (Figure S2), something not observed with the *AAT* promoter constructs. Our results align with the recent findings reported by Buddle et al., describing a case of severe acute hepatitis in a 7-year-old patient treated with onasemnogene abeparvovec.[30] Using long-read nanopore sequencing of liver tissue, the authors identified plasmid-derived sequences originating from manufacturing components, as well as heterogeneous vector genomes, including recombined fragments and vector-host or vector-plasmid chimeric DNA molecules. The authors proposed that these findings might represent potential contributing factors to hepatic toxicity, although the original rAAV production was not sequenced and causality could not be established. Taken together, theirs and our datasets indicate that expression cassette design influences vector genome heterogeneity and impurity packaging during production, especially the use of the *CBA* promoter, and underscore the importance of routinely including long-read sequencing as a quality control measure in rAAV production.

The other key elements in vector genome design are the ITRs, which are the only remaining viral sequences in rAAVs. They bear intrinsic promoter activity and are known to be essential for genome replication, packaging, and vector persistence, among others.[29,31] Despite the fact that ITRs are involved at each stage of the wt viral genome replication process,[29,32] their interaction with the AAV capsid has been poorly studied. By combining several ITR modifications, expression cassette designs, and serotypes, we have been able to identify how ITRs from serotype 2 are repaired during rAAV production and reveal an intriguing ITR-capsid relationship. So far, most of the studies performed on this subject have shown that ITR truncations compromise packaging yields.[33,34] In our experiments, long-read SMRT sequencing has helped to demonstrate that, during rAAV genome replication, small ITR truncations are repaired based on the most functional sequence, leading to the generation of intact ITRs in the process. Moreover, we also found a high percentage of u-wtITRs in most AAV8 preparations compared to the AAVLK03 productions. Interestingly, the scITR sequence, which lacks the full

Figure 4. ITR repair analysis from representative rAAV vectors

Summary table of the percentage distribution of the different ITR forms before (A) and after (B) vector production of representative rAAV vectors, including scITR.AAV3B and dITR.AAV3B.*AAT.Egfp* with the *HPRT1* DNA stuffer sequence. (C) Quantification of the different 5′ ITR and 3′ ITR configurations from rAAV8 vectors with wtITR, dITR/wtITR, and dITR/tITR designs. (D) rAAV8 vectors with the scITR design, and rAAVLK03 vector genomes with scITR, dITR/wtITR, and dITR/tITR designs (ITR nucleotides, nt). wtITR – 143 nt size, light blue; dITR – 130 nt size, light green; tITR – 121 nt size, purple; scITR – 124 nt, black; unresolved wtITR (u-wtITR – 165nt, dark blue); unresolved dITR (u-dITR – 154 nt, dark green).



(legend on next page)

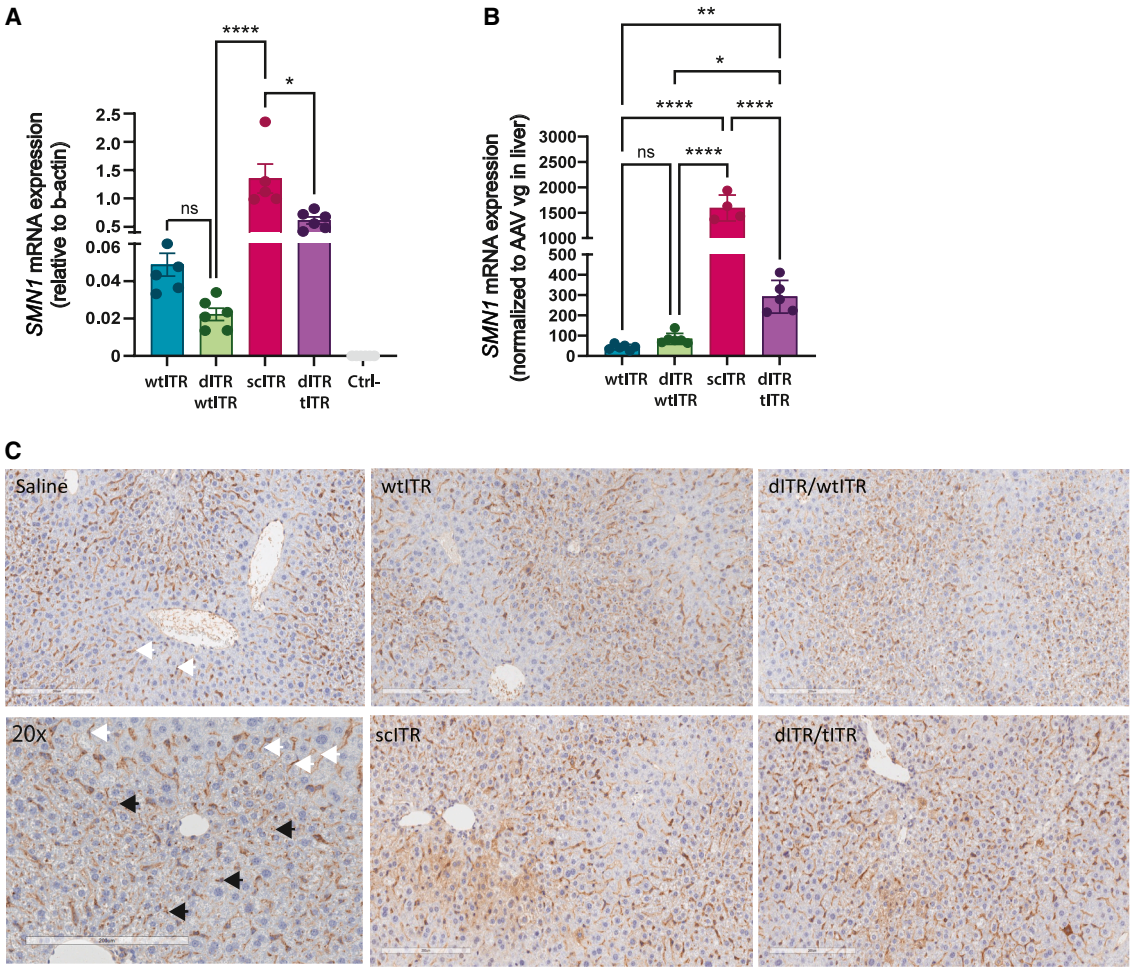


Figure 6. *In vivo* characterization of rAAV8-*SMN1* constructs harboring different ITR configurations

(A) Relative human *SMN1* mRNA levels in liver determined by RT-qPCR and normalized to β-actin. Mice ($n = 6$ per group) were injected with 3×10^{10} vg/mouse of rAAV8-*SMN1* vectors containing wtITR, dITR/wtITR, scITR, or dITR/tITR configurations. (B) *SMN1* transcript levels normalized to vector genome (vg) copy number in liver. (C) Representative immunohistochemistry images from mouse liver sections stained for SMN1 expression. SMN1 staining is identified by specific puncti in the hepatocyte cytoplasm (representative bottom-left image, black arrows). Non-specific mouse IgG staining due to anti-mouse secondary antibody binding in the liver sinusoids is detected (bottom top and left images, white arrows). Scale bars, 200 μm. Statistical significance was determined using one-way ANOVA followed by Dunnett's multiple comparison test. Error bars represent the standard deviation within the group. Significant differences are shown as $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$.

C arm, remained intact after rAAV replication, despite having a 3′-wtITR as a template. This outcome was probably due to the selective advantage of this ITR to generate sc functional genomes, a more stable nucleic acid form with higher potency.

Finally, ITR repair and the generation of double-stranded genomes are also influenced by the serotype. Our ITR repair results showed that rAAV8 genomes harboring a tITR were repaired based on the dITR; therefore, the latter became the most abundant configuration,

Figure 5. rAAV transduction with different ITR configurations *in vitro*

(A) EGFP expression by fluorescence microscopy in Huh7 cells transduced with 1×10^5 vector genomes (vg)/cell of rAAV8 (top) or 1×10^4 vg/cell of rAAVLK03 (bottom) harboring wtITR, dITR/wtITR, scITR, and dITR/tITR designs. (B, D, and F) Quantification of EGFP-positive Huh7 cells and mean fluorescence intensity of positive cells by flow cytometry 48 h post-AAV transduction with increasing doses of rAAV8 (10^3, 10^4, and 10^5 vg/cell). (C, E, and G) Quantification of EGFP-positive Huh7 cells and mean fluorescence intensity of positive cells by flow cytometry 48 h post-AAV transduction with increasing doses of rAAVLK03 (10^3, 10^4, and 10^5 vg/cell). Quantification of rAAV vg in Huh7 cells 48 h post-AAV8 (H) or post-AAVLK03 (I) transduction in each condition by qPCR. Statistical significance was determined using one-way ANOVA followed by Dunnett's multiple comparison test. Error bars represent the standard deviation within the group. Significant differences are shown as $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, and $^{****}p < 0.0001$.

which led mostly to sc species and inadvertently increased vector potency by 2.5×. Interestingly, the same ITR configuration within an AAVLK03 capsid not only improved expression versus wtITRs or dITR/wtITR vectors but also outperformed the scITR design under non-saturating conditions (Figure 5C). When a transduction plateau was reached, with 80%–100% of EGFP-positive cells in all conditions, cells with the dITR/wtITR design were 2× brighter (Figure 5G), despite vectors with this ITR design having a negative impact on AAV replication during production and yielding lower titers (Table S1). A potential explanation is that AAVLK03 could be very efficient at packaging and unpackaging resolved, double-stranded functional vectors, a feature that would require post-entry capsid-ITR interaction. This aligns with the fact that we have consistently found lower proportions of unresolved ITR forms with AAVLK03 compared to AAV8, independently of ss or sc genomes (Figures 4C and 4D), as well as less DNA bound to the capsid (Tables 1 and 2). It is important to note that most of our rAAV biology knowledge is based on wtITR serotype 2 sequences in combination with different capsids. Further research on ITR-capsid interactions and their impact on vector transgene expression will shed light on specific ITR designs, including other serotypes' ITRs, that will result in more potent rAAV combinations.

In conclusion, our results underscore that the combination of multiple techniques to advance our understanding of AAV biology is crucial for overcoming the current limitations in rAAV design and unlocking the full potential of rAAV-mediated gene therapy.

MATERIAL AND METHODS

Cell lines

HEK293T and HUH7 cell lines were purchased from ATCC and cultured at 37°C, 5% CO_2 in DMEM (Thermo Fisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS, Thermo Fisher Scientific), penicillin (100 μg/mL), streptomycin (100 μg/mL), and pyruvate (100 μg/mL). Cell lines were periodically tested for mycoplasma (MycoAlert Mycoplasma Detection Kit).

rAAV vector design

EGFP reporter cassettes were driven either by the human albumin enhancer/α-1 antitrypsin promoter (*EAlbAAT*; 2-kb vector genome; pAAV-*EAlbAAT-Egfp*) or by the cytomegalovirus enhancer/chicken β-actin promoter with a chimeric intron (*CBA*; 3-kb; pAAV-*CBA-Egfp*). Expression cassettes were flanked by wt AAV2 ITRs (wtITRs); commercially available AAV2 ITRs containing an 11-nt deletion in the C arm of the 5′ ITR combined with a 3′ wtITR (dITR/wtITR; VPK-410, Cell Biolabs); truncated ITRs comprising a 22-nt deletion in the B arm of the 3′ ITR combined with a 5′ dITR (dITR/tITR); or modified sc ITRs lacking the entire C arm (scITRs), kindly provided by Dr. Dirk Grimm and first published elsewhere.[17] Human SMN1 coding sequence (GenBank: NM_000344.4)[35] were driven by the EAlbAAT promoter (2 kb vector genome; pAAV-AAT-*SMN1*) and flanked by wtITRs, dITR/wtITRs, dITR/tITRs, or scITRs. The *CBA* promoter construct was flanked by wtITRs (3.2 kb; pAAV-CBA-*SMN1*). All constructs were generated by traditional cloning strategies using restriction enzymes and were used for AAV production.

Vector production and quality analysis

rAAVs were produced by double- or triple-transfection systems (Table S1), depending on the serotype. Six serotypes were tested. To control for variability among productions, each vector was produced at least twice. HEK293T cells (2% FBS DMEM; 65% cell confluency) were co-transfected using linear polyethyleneimine 25 kDa (Polysciences, Warrington, PA, USA), with two plasmids for the double-transfection system (transgene plasmid and pDP plasmid (*rep/cap* and Ad helper genes); ratio 2.75:1) and three plasmids for the triple-transfection system (transgene plasmid, the *rep* and *cap* genes, and pAd Delta F6; at a ratio of 2:1:1). Vector particles were obtained from cells and supernatants following different steps. To isolate vector particles, after 72 h, supernatants were treated with polyethylene glycol solution (PEG8000, 8% final concentration) (48 h; 4°C), then centrifuged (3000 rpm; 15 min), and pellets were resuspended in lysis buffer (50 mM Tris-Cl, 150 mM NaCl, 2 mM $MgCl_2$, 0.1% Triton X-100) and kept at −80°C. Then, both supernatants and cells were frozen and thawed (3×), centrifuged, and treated with DNaseI and RNase solutions (Invitrogen; 0.1 mg/10 μL). This lysate was purified in an iodixanol gradient (15%, 25%, 40%, and 54% iodixanol) by ultracentrifugation (69,000 rpm, 16°C, 2.5 h; in a Beckman type 70 Ti rotor) according to the method of Zolotukhin et al.[36] and concentrated using Ultra-15 mL Amicon columns (Amicon; Millipore, Bedford, MA). Vectors were stored at −80°C until use. For rAAV titration, vector genomes were extracted using the High Pure Viral Nucleic Acid Kit (Roche) according to the manufacturer's specifications and quantified by real-time qPCR using the qPCR assay (CFX Connect TM Real-Time System, Bio-Rad; iQ SYBR Green Supermix (Bio-Rad, 1708882) with specific primers for the ITRs (Table S2). Vector titers ranged between 1.1×10^{11} and 6.8×10^{13} vg/mL for all productions. Capsid viral protein (VP) ratios were analyzed by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) combined with SYPRO Ruby staining.

Viral genome integrity evaluation

Before vector genome extraction (High Pure Viral Nucleic Acid Kit, Roche), 50–100 μL of rAAV were digested with DNaseI (3 μL of DNase for 1 h at 37°C; Invitrogen 2 μg/μL). qPCR (CFX Connect Real-Time System, Bio-Rad; SYBR Green Supermix, #1708882 Bio-Rad) was used to determine AAVs genome integrity by titration of different regions of the recombinant AAV genome (Table S2). Native Tape Station (Agilent 2200 D500-BR automated electrophoresis platform, Agilent Technologies, Waldbronn, Germany) and agarose gels were used to analyze AAV genome integrity. Tape Station electrophoresis was performed by CIMALabs Diagnostics following Agilent recommendations (5 μL of sample/lane). For agarose electrophoresis, samples were run for 50 min at 85 V. The ladder used was FastGene 1-kb DNA marker, and gels were visualized using the ChemiDoc MP System. SmaI restriction enzyme (20,000 units/mL; New England Biolabs) was used to digest rAAV genomes for 1 h at 37°C. To prepare the alkaline agarose gels, 5 mL of 10× alkaline

buffer (prepared by mixing 50 mL of 10 N NaOH and 20 mL of 0.5 M EDTA, then bringing to 800 mL dH2O in a final volume of 1 L) were added to a 1% agarose gel. For sample loading, 9 μL of extracted vector genomes were mixed with 1 μL of 10× alkaline solution and 2 μL of loading dye/buffer (6 mg/mL of xylene cyanol, 300 mM of NaOH, 6 mM of EDTA, 0.18 g/mL of Ficoll). The ladder used was FastGene 1-kb DNA marker, and the gel was run at 75 V for 2.5 h in an ice bath, inside a cold room, and then rinsed with tap water. Subsequently, the gel was immersed in neutralization buffer (1 M Tris–HCl, pH 7.6; 1.5 M NaCl) for 30 min, followed by incubation in 1× SYBR Gold (SYBR Gold 10,000×, Thermo Fisher Scientific Inc, S11494) for 45 min, and then in distilled water for 20 min prior to visualization using a UV transilluminator.

Plasmid spike-in digestion

Increasing concentrations of spiked-in pAAV plasmid (0, 0.05, 0.1, 0.5, 1, and 5 ng/μL) were added to 10 μL of vector preparation and digested for 1 h at 37°C with either DNase (3 μL of 2 μg/μL Dnase, Invitrogen) or Benzonase (50 U/mL, Nuclease Benzonase, Merk, ≥250 units/μL)). Samples were collected for DNA quantification before spike-in and after DNase/Benzonase incubation. Three independent extractions were performed and quantified. DNA was quantified using Qubit with the dsDNA HS Assay Kit, following the manufacturer's instructions (Qubit Flex Fluorometer, Invitrogen, Q32851). Samples were also analyzed by running on a native agarose gel.

AAV genome sequencing and analysis

PacBio library preparation, long-read SMRT sequencing, and data analysis

Purified vector DNAs were processed for SMRT sequencing by the UMass Chan Medical School Deep Sequencing Core. Libraries for vector DNAs were constructed using the Express Template Prep Kit 2.0 (End Repair/A-tailing, PN 100-938-900) and ligated to indexed SMRT bell adapters with the Barcoded Overhang Adapter Kit (PN 101-628-400/500). Libraries were pooled and purified using 1.8× AMPure beads. Sequencing was performed on a Sequel II instrument. Subreads from each library underwent preprocessing steps prior to analysis as described previously.[37] Briefly, subreads were first processed through recalladapters version 9.0.0 with the following parameters: –minSnr = 2.0 and –disableAdapterCorrection. This step recovered long palindromic sequences that were artificially cleaved from subreads generated by SMRT link (v12). Subsequent subreads were then processed using the circular consensus sequencing (CCS) tool in SMRT link version 12.0.0 with the following parameters: –minSnr = 3.75, –min = passes = 2.0, and –by strand. The following bioinformatic workflows were conducted using the Galaxy platform.[38,39] Resulting consensus FASTQ files were mapped to the reference sequences using the Burrows-Wheeler Aligner-Maximal Exact Match (BWA-MEM)[40,41] in PacBio mode (-x pacbio). Aligned reads were visualized with the Integrative Genomics Viewer (IGV) tool version 2.14.0 with soft clipping enabled.[42]

The procedure for categorizing and quantifying ITR configurations was carried out using custom workflows in USEARCH and publicly available tools on Galaxy, as described previously.[39] Reads from all libraries were mapped to the corresponding vector reference from ITR to ITR using BWA-MEM. Mapped reads were selected, and the ITR sequences were extracted from these mapped reads by selecting sequences flanking the D region. ITRs were analyzed for their length distribution, and major peaks were identified. For each identified peak, the corresponding population was extracted, and unique ITR lengths were selected to cluster reads at 95% similarity using USEARCH v.10 with the command -cluster_fast and the parameter, -id 0.95. Clusters containing five or more members were selected, and the consensus sequences were defined as the ITR configurations. Using this method, ITR configurations, including the parental wtITR, tITR, dITR, and scITR sequences, were identified. The left and right (5′ and 3′, respectively) ITRs were separated and mapped to the defined ITR configurations, as reference genomes, to categorize them accordingly.

Oxford Nanopore sequencing

For the *SMN1* vectors, AAV genome integrity was analyzed using the AAV sequencing service provided by Plasmidsaurus Inc. (Eugene, OR, USA). AAV genomes were sequenced using Oxford Nanopore long-read technology, followed by custom analysis and annotation.

Transduction efficiency assays

A total of 1×10^5 Huh7 cells/well were seeded in 48-well tissue-culture-treated plates (50% confluency). The following day, cells were transduced with AAV vectors in DMEM supplemented with 2% FBS. After 2 h of incubation, 275 μL of DMEM containing 10% FBS was added. For the *Egfp* transgene, six wells per condition were transduced with 10^3, 10^4, and 10^5 vector genomes (vg)/cell of AAV8 and AAVLK03 in two independent experiments. EGFP expression was analyzed at 24 and 48 h post-transduction by fluorescence microscopy (Zeiss Axio Observer Z1, Zeiss Germany), and cells were collected at 48 h for flow cytometry analysis and vg quantification. For the *SMN1* transgene, two wells per condition were transduced with 10^5 vg/cell of AAV8. SMN1 expression was analyzed 48 h post-transduction by immunofluorescence.

Immunofluorescence

Cells were fixed with 4% paraformaldehyde (PFA; ITW Reagents, 252931), permeabilized with 0.1% Triton X-100 (Sigma-Aldrich, Product No. 282103; CAS 92046-34-9), and blocked with 3% BSA (Sigma-Aldrich, A3059; CAS 9048-46-8). Samples were incubated overnight at 4°C with anti-SMN1 primary antibody (mouse monoclonal, Abcam, #ab5831), followed by incubation with donkey anti-mouse IgG conjugated to Alexa Fluor 647 (Invitrogen, A31571). Nuclei were counterstained with Hoechst (Thermo Fisher Scientific, H21486), and samples were mounted prior to imaging using a Zeiss LSM 800 laser scanning confocal microscope. For quantitative image analysis (QuPath version 0.5.1.[43]), cells were segmented based on the nuclear channel (Hoechst), and single-cell mean fluorescence intensity (MFI) was quantified for SMN1 staining. Cells were classified as positive or negative based on the basal fluorescence of non-transduced control cells. Two random fields of view (FOV) were analyzed per experimental group.

Flow cytometry

Cells were harvested with trypsin (Gibco, trypsin-EDTA [0.05%]), washed with 1 mL of fluorescence-activated cell sorting (FACS) buffer (DPBS, Gibco; 5% FCS, 0.1% (w/v) NaN_3), and resuspended in 200 μL of FACS buffer. Cells were acquired in a CytoFLEX LX (Beckman Coulter). For the GFP-positive gating strategy, dead cells were excluded using 7-amino-actinomycin D (7-AAD) staining solution (Thermo Fisher Scientific, V35123). Negative controls were used for 7-AAD marking and AAV transduction. Flow cytometry analysis was performed using BD CytExpert 2.1 software (Beckman Coulter) at the Flow Cytometry Core, CIMA (Pamplona, Spain). Data were analyzed using FlowJo software (v.10.7.1; BD Life Sciences). MFI was calculated considering only EGFP-positive gated cells.

Animal in vivo *study*

Thirty 4-week-old male C57BL/6 mice (n = 6 per group) were intravenously administered either saline or 3×10^{10} vg/mouse of rAAV-EAlbAAT-*SMN1*, flanked by wtITRs, dITR/wtITRs, dITR/tITRs, or scITRs. Fourteen days post-injection, animals were sacrificed, and liver samples were collected. The median lobe was fixed in 4% paraformaldehyde for 48 h for SMN1 immunohistochemistry, while the remaining lobes were snap-frozen in liquid nitrogen for molecular analyses. Animals that were not properly dosed were excluded from the study. All experimental procedures were performed under general anesthesia (Isoflurane Isovet, Piramal; 1000 mg/g) and were approved by the Ethics Committee for Animal Experimentation of the University of Navarra (Protocol No.058-24).

Quantification of vector genomes and SMN1 *gene expression*

Vector genomes were quantified by qPCR using ITR-specific primers, as described in the Vector production and quality analysis section, starting from 50 ng of total DNA. For gene expression analysis, 0.5 μg of total RNA was treated with DNaseI at 37°C for 30 min, followed by EDTA inactivation at 75°C for 10 min. Reverse transcription was performed using RNaseOUT recombinant ribonuclease inhibitor (Invitrogen, 10777019), M-MLV reverse transcriptase, 0.1 M DTT, 5× first-strand buffer (Invitrogen, 28025013), 10 mM deoxyribonucleotide triphosphate (dNTP) mix (Invitrogen, 362275), and random primers (Invitrogen, 48190011) at 37°C for 1 h, followed by 95°C for 1 min. Human *SMN1* mRNA expression levels were quantified by qPCR using iQ SYBR Green Supermix (Bio-Rad, 1708882) and normalized to mouse β-actin using the $2^{-}\Delta\Delta$Ct method (Table S2).

Immunohistochemical analysis

Previously fixed and paraffin-embedded liver sections were incubated overnight at 4°C with a mouse monoclonal anti-SMN/Gemin 1 antibody (Abcam, ab5831) to detect *SMN1* expression.

Statistical analyses

Statistical analyses were performed using GraphPad Prism version 10.6.1 (San Diego, California, USA). Data normality was assessed using the Shapiro-Wilk test, with a significance level of $\alpha = 0.05$. A one-way ordinary ANOVA was performed to evaluate the effect of group on transgene expression. Post hoc multiple comparisons between groups were conducted using Dunnett's multiple comparison test. In all analyses, statistical significance was set at $p < 0.05$.

DATA AND CODE AVAILABILITY

All data are provided in the main text, figures, or supplemental information. Raw data files and vectors used in this study are available upon request from the corresponding authors.

ACKNOWLEDGMENTS

This work has been funded by MCIN/AEI/10.13039/501100011033 (Ministerio de Ciencia e Innovación, Spain) and by the European Union Next GenerationEU/PRTR (MCI PID2022-137327OA-I00) to C.U., and P.W.L.T. is supported by the National Institutes of Health (1R21AI183080-01A1), an award from The Bassick Family Foundation, and a BRIDGE Fund Award (a UMass Chan Medical School internal grant).

The authors deeply thank Ngoc Tam Tran for her support with sequencing analysis; Gorka Alkorta (Cimalabs Diagnostics) for assistance with Tapestation; Mirja Hommel, Diego Alignani, and the staff from the CIMA flow cytometry core for their support with flow cytometry. The authors also thank Dirk Grimm and his laboratory for kindly sharing their self-complementary pAAV plasmid, as well as the UMass Chan Deep Sequencing core (Maria Zapp, Daniella Wilmot, and Ellie Kittler) for their assistance with sample preparation and SMRT sequencing. Finally, the authors thank Guangping Gao for his support and guidance on the project. The graphical abstract was created using Biorender.com.

AUTHOR CONTRIBUTIONS

Conceptualization, C.L., P.W.L.T., G.G.-A., and C.U.; investigation and methodology, C.L., M.Y., S.N., A.V., C.O., I.M., P.M., S.I., P.W.L.T., and C.U.; analysis, C.L., M.Y., S.N., P.W.L.T., G.G.-A., and C.U.; funding acquisition, P.W.L.T., G.G.-A., and C.U.; writing, review and editing, C.L., G.G.-A., P.W.L.T., and C.U.

DECLARATION OF INTERESTS

C.U. is a co-inventor on patent applications for the use of AAV vectors, licensed to biotechnology companies unrelated to this research. P.W.L.T. is an inventor on patents with royalties, licensed to biopharmaceutical companies unrelated to this research. G.G.-A. is a co-inventor on patent applications for the use of AAV vectors, licensed to biotechnology companies unrelated to this research.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2026.102899.

REFERENCES

1. Wang, D., Tai, P.W.L., and Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. Nat. Rev. Drug Discov. *18*, 358–378.
2. Atchison, R.W., Casto, B.C., and Hammon, W.M. (1965). Adenovirus-associated defective virus particles. Science *149*, 754–756.
3. Pupo, A., Fernández, A., Low, S.H., François, A., Suárez-Amarán, L., and Samulski, R.J. (2022). AAV vectors: The Rubik's cube of human gene therapy. Mol. Ther. *30*, 3515–3541.
4. Hermonat, P.L., and Muzyczka, N. (1984). Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. Proc. Natl. Acad. Sci. USA *81*, 6466–6470.
5. Zhou, S.Z., Broxmeyer, H.E., Cooper, S., Harrington, M.A., and Srivastava, A. (1993). Adeno-associated virus 2-mediated gene transfer in murine hematopoietic progenitor cells. Exp. Hematol. *21*, 928–933.
6. Robert, M.A., Chahal, P.S., Audy, A., Kamen, A., Gilbert, R., and Gaillet, B. (2017). Manufacturing of recombinant adeno-associated viruses using mammalian expression platforms. Biotechnol. J. *12*, 1600193.

7. Ferrari, F.K., Samulski, T., Shenk, T., and Samulski, R.J. (1996). Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J. Virol. *70*, 3227–3234.
8. Fisher, K.J., Gao, G.P., Weitzman, M.D., DeMatteo, R., Burda, J.F., and Wilson, J.M. (1996). Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. J. Virol. *70*, 520–532.
9. McCarty, D.M., Monahan, P.E., and Samulski, R.J. (2001). Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. Gene Ther. *8*, 1248–1254.
10. McCarty, D.M., Fu, H., Monahan, P.E., Toulson, C.E., Naik, P., and Samulski, R.J. (2003). Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo. Gene Ther. *10*, 2112–2118.
11. Xiao, X., Xiao, W., Li, J., and Samulski, R.J. (1997). A novel 165-base-pair terminal repeat sequence is the sole cis requirement for the adeno-associated virus life cycle. J. Virol. *71*, 941–948.
12. Snyder, R.O., Im, D.S., Ni, T., Xiao, X., Samulski, R.J., and Muzyczka, N. (1993). Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. J. Virol. *67*, 6096–6104.
13. McCarty, D.M. (2008). Self-complementary AAV vectors; advances and applications. Mol. Ther. *16*, 1648–1656.
14. Wu, J., Zhao, W., Zhong, L., Han, Z., Li, B., Ma, W., Weigel-Kelley, K.A., Warrington, K.H., and Srivastava, A. (2007). Self-complementary recombinant adeno-associated viral vectors: packaging capacity and the role of rep proteins in vector purity. Hum. Gene Ther. *18*, 171–182.
15. Zhou, Q., Tian, W., Liu, C., Lian, Z., Dong, X., and Wu, X. (2017). Deletion of the B-B' and C-C' regions of inverted terminal repeats reduces rAAV productivity but increases transgene expression. Sci. Rep. *7*, 5432.
16. Zhang, J., Frabutt, D.A., Chrzanowski, M., Li, N., Miller, L.M., Tian, J., Mulcrone, P.L., Lam, A.K., Draper, B.E., Jarrold, M.F., et al. (2024). A novel class of self-complementary AAV vectors with multiple advantages based on cceAAV lacking mutant ITR. Mol. Ther. Methods Clin. Dev. *32*, 101206.
17. Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F., and Kay, M.A. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature *441*, 537–541.
18. McColl-Carboni, A., Dollive, S., Laughlin, S., Lushi, R., MacArthur, M., Zhou, S., Gagnon, J., Smith, C.A., Burnham, B., Horton, R., et al. (2024). Analytical characterization of full, intermediate, and empty AAV capsids. Gene Ther. *31*, 285–294.
19. Heckel, J., Martinez, A., Elger, C., Haindl, M., Leiss, M., Ruppert, R., Williams, C., Hubbuch, J., and Graf, T. (2023). Fast HPLC-based affinity method to determine capsid titer and full/empty ratio of adeno-associated viral vectors. Mol. Ther. Methods Clin. Dev. *31*, 101148.
20. Werle, A.K., Powers, T.W., Zobel, J.F., Wappelhorst, C.N., Jarrold, M.F., Lyktey, N.A., Sloan, C.D.K., Wolf, A.J., Adams-Hall, S., Baldus, P., and Runnels, H.A. (2021). Comparison of analytical techniques to quantitate the capsid content of adeno-associated viral vectors. Mol. Ther. Methods Clin. Dev. *23*, 254–262.
21. 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.
22. Wright, J.F. (2014). Product-Related Impurities in Clinical-Grade Recombinant AAV Vectors: Characterization and Risk Assessment. Biomedicines *2*, 80–97.
23. Zhang, J., Chrzanowski, M., Frabutt, D.A., Lam, A.K., Mulcrone, P.L., Li, L., Konkle, B.A., Miao, C.H., and Xiao, W. (2023). Cryptic resolution sites in the vector plasmid lead to the heterogeneities in the rAAV vectors. J. Med. Virol. *95*, e28433.
24. Lisowski, L., Dane, A.P., Chu, K., Zhang, Y., Cunningham, S.C., Wilson, E.M., Nygaard, S., Grompe, M., Alexander, I.E., and Kay, M.A. (2014). Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. Nature *506*, 382–386.
25. Gonzalez-Sandoval, A., Pekrun, K., Tsuji, S., Zhang, F., Hung, K.L., Chang, H.Y., and Kay, M.A. (2023). The AAV capsid can influence the epigenetic marking of rAAV delivered episomal genomes in a species dependent manner. Nat. Commun. *14*, 2448.
26. de la Maza, L.M., and Carter, B.J. (1980). Molecular structure of adeno-associated virus variant DNA. J. Biol. Chem. *255*, 3194–3203.
27. Torella, L., Klermund, J., Bilbao-Arribas, M., Tamayo, I., Andrieux, G., Chmielewski, K.O., Vales, A., Olagüe, C., Moreno-Luqui, D., Raimondi, I., et al. (2024). Efficient and safe therapeutic use of paired Cas9-nickases for primary hyperoxaluria type 1. EMBO Mol. Med. *16*, 112–131.
28. Maestro, S., Gómez-Echarte, N., Camps, G., Usai, C., Suárez, L., Vales, Á., Olagüe, C., Aldabe, R., and González-Aseguinolaza, G. (2021). AAV-HDV: An Attractive Platform for the In Vivo Study of HDV Biology and the Mechanism of Disease Pathogenesis. Viruses *13*, 788.
29. Xie, J., Mao, Q., Tai, P.W.L., He, R., Ai, J., Su, Q., Zhu, Y., Ma, H., Li, J., Gong, S., et al. (2017). Short DNA Hairpins Compromise Recombinant Adeno-Associated Virus Genome Homogeneity. Mol. Ther. *25*, 1363–1374.
30. Buddle, S., Brown, L.A.K., Morfopoulou, S., Torres Montaguth, O.E., Scoto, M., Herder, V., Dhawan, A., Brown, J.R., Atkinson, L., Kopec, A., et al. (2026). Contaminating plasmid sequences and disrupted vector genomes in the liver following adeno-associated virus gene therapy. Nat. Med. *32*, 472–480.
31. Shitik, E.M., Shalik, I.K., and Yudkin, D.V. (2023). AAV- based vector improvements unrelated to capsid protein modification. Front. Med. *10*, 1106085.
32. Earley, L.F., Conatser, L.M., Lue, V.M., Dobbins, A.L., Li, C., Hirsch, M.L., and Samulski, R.J. (2020). Adeno-Associated Virus Serotype-Specific Inverted Terminal Repeat Sequence Role in Vector Transgene Expression. Hum. Gene Ther. *31*, 151–162.
33. Chen, Y., Hu, S., Lee, W., Walsh, N., Iozza, K., Huang, N., Preston, G., Drouin, L.M., Jia, N., Deng, J., et al. (2024). A Comprehensive Study of the Effects by Sequence Truncation within Inverted Terminal Repeats (ITRs) on the Productivity, Genome Packaging, and Potency of AAV Vectors. Microorganisms *12*, 310.
34. Samulski, R.J., Srivastava, A., Berns, K.I., and Muzyczka, N. (1983). Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. Cell *33*, 135–143.
35. Day, J.W., Finkel, R.S., Chiriboga, C.A., Connolly, A.M., Crawford, T.O., Darras, B.T., Iannaccone, S.T., Kuntz, N.L., Peña, L.D.M., Shieh, P.B., et al. (2021). Onasemnogene abeparvovec gene therapy for symptomatic infantile-onset spinal muscular atrophy in patients with two copies of SMN2 (STR1VE): an open-label, single-arm, multicentre, phase 3 trial. Lancet Neurol. *20*, 284–293. https://doi.org/10.1016/S1474-4422(21)00001-6.
36. Zolotukhin, S., Byrne, B.J., Mason, E., Zolotukhin, I., Potter, M., Chesnut, K., Summerford, C., Samulski, R.J., and Muzyczka, N. (1999). Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. Gene Ther. *6*, 973–985.
37. Tran, N.T., Lecomte, E., Saleun, S., Namkung, S., Robin, C., Weber, K., Devine, E., Blouin, V., Adjali, O., Ayuso, E., et al. (2022). Human and Insect Cell-Produced Recombinant Adeno-Associated Viruses Show Differences in Genome Heterogeneity. Hum. Gene Ther. *33*, 371–388.
38. Afgan, E., Sloggett, C., Goonasekera, N., Makunin, I., Benson, D., Crowe, M., Gladman, S., Kowsar, Y., Pheasant, M., Horst, R., and Lonie, A. (2015). Genomics Virtual Laboratory: A Practical Bioinformatics Workbench for the Cloud. PLoS One *10*, e0140829.
39. Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. *19*, 19.10.1–19.10.21.
40. Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome Res. *15*, 1451–1455.
41. Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.
42. 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.
43. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open-source software for digital pathology image analysis. Sci. Rep. *7*, 16878.