Adeno-Associated Viral Vector Integration: Implications for Long-Term Efficacy and Safety

Author: Paul Batty [1,2], David Lillicrap [2]

Institution Addresses:

[1]Department of Haematology, Cancer Institute, University College London, London, UK

[2]Department of Pathology and Molecular Medicine, Queen's University, Kingston, Ontario, Canada

Corresponding Author: Dr Paul Batty, Department of Haematology, Cancer Institute, University College London, Royal Free Campus, Pond Street, London, NW3 2QG

Telephone: 0207 794 0500 Ext. 35920

Email: paul.batty@ucl.ac.uk

Word Count (Abstract): 250 words

Word Count (Body of Text): 5,099 words

Running Short Title: AAV Integration

Keywords: Adeno-Associated Virus**,** Episome, Gene Therapy, Hemophilia A, Virus Integration

Abstract:

Adeno-associated viral vector (AAV) gene therapy provides a promising platform for treatment of monogenic inherited disorders. Clinical studies have demonstrated long-term expression with reduction in bleeding using this approach for the treatment of hemophilia. Despite these advances, there are unknowns surrounding the natural history of recombinant AAV (rAAV) vectors and the cellular mechanisms mediating vector persistence. These unknowns underpin questions regarding long-term efficacy and safety. The predominant mechanism via which AAV is proposed to persist is in circular double-stranded extrachromosomal DNA structures (episomes) within the nucleus. Studies of wild-type (WT-AAV) and rAAV have demonstrated that AAV also persists via integration into a host cell's DNA. It is important to determine whether these integration events can mediate expression or could result in any long-term safety concerns. WT-AAV infection affects a large proportion of the general population, which is thought to have no long-term sequelae. Recent studies have highlighted that this WT-AAV has been detected in cases of acute hepatitis in children and in a minority of cases of hepatocellular carcinoma. Integration following treatment using rAAV has also been reported in preclinical and clinical studies. There have been variable reports on the potential implications of integration for rAAV vectors with data in some murine studies demonstrating recurrent integration with development of hepatocellular carcinoma. These findings have not been seen in other pre-clinical or clinical studies. In this review, we will summarize current understanding of the natural history of AAV (wild-type and recombinant) with a focus on genomic integration and the cellular implications.

Introduction

Adeno-associated viral (AAV) vector gene therapy offers the potential to provide long-term transformative treatment for individuals with monogenic disorders. This approach has been successfully trialed for the treatment of hemophilia, with pre-clinical and clinical studies demonstrating persistent expression of Factor VIII (hemophilia A) or Factor IX (hemophilia B) and reduction in bleeding events (1, 2). This has resulted in the approval and commercial availability of vectors for both hemophilia B: Etranacogene dezaparvovec (FDA, EMA, and Health Canada) and Fidanacogene elaparvovec (FDA, EMA & Health Canada) and hemophilia A: Valoctocogene roxaparvovec (FDA & EMA). AAV vectors have also received approvals for treatment of inherited retinal dystrophy, spinal muscular atrophy, Duchenne muscular dystrophy and aromatic L-amino acid decarboxylase deficiency and there are multiple clinical trials ongoing in other therapy areas (reviewed in (3)). Despite these advances, there are unknowns surrounding the natural history of recombinant AAV (rAAV) vectors, which underpin questions regarding long-term efficacy and safety. To answer these questions, there is a need to understand how AAV persists long-term. The main mechanism via which AAV is proposed to persist in the nucleus of transduced cells is in circular double-stranded DNA structures (episomes). This mechanism would not be anticipated to result in damage to the native genome or dysregulation of gene expression (genotoxicity). In an episomal form, loss of therapeutic vectors and expression with time would be anticipated as cells divide and lose their episomal content, unless episomal attachment to host chromosomes occurs. A second mechanism by which AAV vectors could persist is by direct inclusion (integration) of viral sequences into a host cell's genome. Although, integration could provide a more stable mechanism for long-term expression, this could theoretically provide one step in the process of tumorigenesis if this occurred in proximity to a protooncogene or resulted in disrupted expression of a tumor suppressor gene (insertional oncogenesis). In this review, we will summarize current understanding of the natural history of AAV (wild-type and recombinant), with a focus on genomic integration and the cellular implications.

Wild-Type Adeno-Associated Virus: Mechanism of Persistence and Integration

Adeno-associated viruses (AAV) are small replication-deficient non-enveloped DNA viruses of the Parvoviridae family (Dependoparvovirus genus). AAV was first described as a contaminant in simian adenovirus preparations in the mid-1960s (4, 5). AAV possesses a single-stranded DNA genome (4.7kb) within an icosahedral protein capsid. The AAV genome consists of *Rep* and *Cap* viral coding sequences, between two inverted tandem repeat structures (each ~200 bps) (Figure 1). Within the open reading frame (ORF) of AAV2, there are three promoters: p5, p19 & p40. *Rep* encodes four proteins: Rep78, Rep68, Rep52 & Rep40. Rep68 and Rep78 are involved in viral DNA replication, site-specific integration

and regulation of gene expression. Rep52 and Rep40 are involved in capsid loading. The *cap* genes encode three structural AAV capsid proteins: VP1, VP2 & VP3, with VP3 accounting for 90% of the capsid protein content. Thirteen natural capsids have been identified, with different primary and secondary receptors for cellular uptake which results in different species and organ-specific transduction profiles (6-8). The AAV genome also encodes three other proteins: assembly accessory protein (AAP), membrane associated accessory protein (MAAP) and Protein X (Figure 1). The AAP is transcribed from a frame shift in the VP2/3 reading frame and is involved in targeting the VPs to the nucleolus and capsid assembly. The MAAP is encoded in a different reading frame of VP1/2 and facilitates AAV replication (9). Protein X is encoded in the 3' region and has been proposed to play a role in DNA replication (10, 11). Finally, a sequence within the 3' untranslated region (3'UTR) has been identified to have enhancer activity (12).

Natural AAV infection is endemic in the population seen from high rates of seropositivity (30-80%) (13-21). Viral transmission occurs via respiratory and gastrointestinal routes, with active replication requiring the presence of a helper virus. Helper viruses include adenoviruses and members of the human herpesvirus family. A detailed review of the interplay of these helper viruses in the AAV lifecycle has recently been presented (22). In the absence of a helper virus AAV persists in a latent form in circular double-stranded episomal structures or via integration into the host genome (23, 24). Studies characterizing integration have predominantly been performed *in vitro* for wild-type (WT)-AAV2. These have demonstrated that WT-AAV integration is a site-specific process, directed by the Rep78/68 complex. This binds to the host genome at consensus Rep-binding sites (RBS) and within the viral DNA at cis-acting elements containing RBS, in the ITRs or p5 promoter (25, 26). This results in a single-stranded cleavage of the host and viral genomes (27).

A recurrent integration site for WT-AAV2, was first identified in HeLa cells on chromosome 19q13.42 in the first exon of the Protein Phosphatase 1 Regulatory Subunit 12C (PPP1R12C) gene; denoted the AAV integration site 1 (AAVS1) locus (28). A genome-wide analysis has identified additional recurrent integration sites in vicinity to consensus RBS sequences containing repeated GAGY/C motifs (29). These integration sites occurred with increased frequency in regions containing genes, with high GC content and areas of open chromatin. This included two further recurrent integration sites: AAVS2 (chr 5p13.3) and AAVS3 (chr 3p24.3) (29). Further studies of WT-AAV2 have demonstrated that this predominantly integrates as single full genome copies (30). Integration profiles have subsequently been evaluated for different cell-lines and AAV serotypes. In diploid human fibroblasts divergent integration hot spots are seen compared to HeLa cells, resulting from differences in chromatin accessibility (31). Studies of WT-AAV5 have also demonstrated different integration patterns (32). This heterogeneity of integration preferences is an area where further study is required. Recent integration

studies evaluating liver samples from non-human primates (n=168) and humans (n=85), support these findings (33). In human samples, 60% had ≥1 WT-AAV integration site comprising 1836 unique integration loci (UILs). A random distribution pattern was seen when the number of UILs was normalized to chromosomal size in human samples. Similar distribution patterns were seen in NHPs except for relative increased frequency in chromosome 19 and decreased frequency in chromosome 3. Integration loci were enriched in genes that are highly expressed in the liver, possess binding regions for RNA Pol II & transcription factors and regions of open chromatin. Site specific integration (AAVS1, AAVS2 and AAV3) was seen in the human and macaque genomes (7.2% of UILs).

It is thought that infection with WT-AAV does not lead to any significant long-term pathology. This concept has been challenged recently and is an area where further study is required. Recent studies identified the presence of WT-AAV2 in liver samples obtained from children diagnosed with acute liver failure (34-36). Studies have also evaluated whether AAV may play a role in the development of hepatocellular carcinoma (HCC), with conflicting results. One case series demonstrated WT-AAV2 integration in 11/193 of HCC cases. Integration was seen in known cancer genes (CCNA2, TERT, CCNE1 and KMT2B), with upregulation of gene expression suggesting a pathogenic role (37). Data presented from whole genome evaluation of liver cancer cases (n=300) in a Japanese population demonstrated WT-AAV integration in three cases with hepatitis B or C co-infection (38). Integration was seen in KMT2B, CCNE1 and an intergenic region on chromosome 5. Finally, a larger study of 1461 patients, demonstrated clonal WT-AAV integration in 30 HCC cases, with recurrent integration sites in CCNA2, CCNE1, TERT, TNFSF10, KMT2B and GLI1/INHBE (24). In contrast, other groups have demonstrated no evidence of AAV-mediated oncogene insertion (39, 40). With the high seroprevalence of WT-AAV and low incidence seen in these studies, the pathogenic involvement of WT-AAV is likely a minor contributing factor for most cases of HCC, although this is an area requiring further study.

Recombinant Adeno-Associated Viral Vectors: Mechanisms of Persistence and Integration

With the relatively low immunogenicity of AAV and potential for persistence via episomal forms, this virus has become an attractive vehicle for therapeutic gene delivery. Engineered recombinant AAVs (rAAV) demonstrate significant differences from the WT-AAV (Figure 1). Generation of rAAV vectors involves removal of viral sequences, with the single-stranded rAAV genome retaining only the two ITRs (AAV2 ITRs). The *Rep* and *Cap* sequences are replaced by a transgene of interest and a promoter/enhancer sequence, that regulates transgene expression. Natural or bioengineered capsids provide a degree of tissue tropism with the aim of reducing off-target effects. Within current therapeutic vectors the 3' UTR enhancer has been removed, although this was present in some earlier preclinical studies. With a lack of viral sequences in rAAV vectors, studies performed in WT-AAV are

limited in their ability to inform the natural history of rAAV. With rAAV vectors lacking the Rep sequence one might anticipate that if integration were to occur this would be a more random or passive process in genomic regions of accessible chromatin (31). In the following sections we will summarize current understanding of the natural history of rAAV and address questions arising from preclinical and clinical studies.

Laboratory Characterization of Integration

Prior to considering these studies it is important to provide context on the laboratory approaches used to characterize integration. These assays detect junctions between viral sequences and the host genome, which are then aligned to a reference genome to characterize where integration has occurred. Approaches are either untargeted (e.g., whole genome sequencing) or targeted to enrich integration sites. Targeted approaches use primers to pre-specified sequences within the vector (PCR-based: e.g., linear amplification mediated PCR [LAM-PCR], ligation mediated-PCR [LM-PCR] or non-PCR based: e.g., target enrichment sequencing). It is likely that there will be differences in the integration sites detected depending on the methodology used, a critical factor in the interpretation of study outcomes (41).

Small Animal Studies: Evidence of Genotoxicity in Murine Studies

The presence of integration *in vivo* following delivery of rAAV, was first proposed in studies conducted in the late 1990's in mice treated with intramuscular (42-45) or liver-directed rAAV (46). Integration was confirmed for a liver-directed rAAV (AAV-EF1α-FIX), which identified integrated forms containing ITR deletions, vector rearrangements, and specific integration sites in the mouse rRNA and α1 collagen genes (47). Subsequent studies by this group using a partial hepatectomy model proposed that non-integrated episomal forms were the main rAAV structures responsible for stable transduction (48). Detailed studies following intramuscular gene delivery supported these findings, describing that rAAV predominantly (>99.5%) persisted in transcriptionally active monomeric or concatemeric episomes (49).

Following on from these studies, there was a shift in focus of the relevance of integration following a report of tumor formation following neonatal gene therapy in mucopolysaccharidosis type VII (MPSVII) mice (50). In this study, angiosarcoma and/or hepatocellular carcinoma was seen in mice with up to 18 months follow-up. This study, which was not designed to assess toxicity, concluded that tumor formation had probably not resulted from insertional mutagenesis with very low vector copy numbers detected by qPCR, although this finding required further investigation. A larger follow-up study provided clearer insights into mechanisms underlying these findings (51). Treatment of neonatal

wild-type or MPSVII mice using the same AAV-GUSB construct recapitulated these findings, with increased incidence of HCC. A single unique amplification product was retrieved from 4 tumors, which mapped to a 6kb region of the RNA imprinted and accumulated in nucleus (*Rian*) locus on chromosome 12, a region enriched in murine miRNA and snoRNA genes. Overexpression of adjacent genes at this locus suggested a role for insertional mutagenesis contributing to the pathogenesis of HCC. These findings are supported by targeted activation of this locus inducing HCC in another murine study (52). Following on from these important findings, there have been variable reports with some studies showing similar findings (53-56), and others not reporting rAAV integration in the *Rian* locus with associated tumorigenesis (57-62). In contrast to the murine data, a study in a neonatal rat model performed specifically to evaluate tumorigenicity demonstrated no evidence of rAAV triggering pre-neoplastic lesions (63). Within murine studies where insertional mutagenesis was reported, this appears to be associated with higher vector dose, young age at treatment, and the rAAV promoter/enhancer elements utilized (64). For more detailed discussion of murine studies evaluating genotoxicity, we refer the reader to a recently published review (64).

Recent murine studies have evaluated the impact of host factors on the safety of rAAV gene therapy. Metabolic-associated fatty liver disease (MAFLD) in humans, is an important risk factor for HCC formation. The impact of experimentally induced non-alcoholic fatty liver disease has recently been reported in mice treated with a *Rian* locus targeting-rAAV construct (65). Within this study, adult mice did not develop HCC unless they were fed a high fat diet. Interestingly, a lower HCC incidence was seen in female mice, which appears to be an estrogen-related effect. Consistent with this theory, male mice treated with estrogen exhibited less inflammation and immune exhaustion associated with oncogenesis compared to those without estrogen. A recent murine study has suggested the impact of metabolic dysfunction may have additional complexity. This study used a plasmid containing B-domain deleted FVIII (i.e., without an rAAV vector) delivered by hydrodynamic tail vein injection (66). Different plasmids were studied to evaluate the effect of FVIII misfolding: a well-folded control (cytosolic dihydrofolate reductase, DHFR), BDD-FVIII prone to misfolding, and a N6-FVIII which folds more efficiently than BDD-FVIII. One week after plasmid delivery, when FVIII expression was undetectable, mice were fed a high-fat diet for 65 weeks. All mice treated with the BDD-FVIII construct developed liver tumors, in contrast to 56% treated with N6-FVIII, and no mice treated with the DHFR plasmid. This suggests that even a transient effect of protein misfolding within the ER may predispose to HCC formation with a high-fat diet, and that viral vector integration events are not required for the oncogenic outcome. Whether the process of plasmid delivery via hydrodynamic delivery, that results in significant hepatocyte injury, acts an additional factor in the multi-stage process of tumorigenesis requires further study. This study may have important translational implications, firstly with the rising

prevalence of MAFLD and secondly, providing additional theoretical factors that could contribute to tumor formation aside from insertional mutagenesis.

In summary, integration following treatment with rAAV vectors was first reported more than two decades ago in studies conducted in mouse models. Variable findings from these murine studies have been reported on implications of integration, with some studies demonstrating recurrent sites of integration with associated tumorigenesis. Recent studies evaluating the impact of liver disease and diet/obesity are of significant translational interest and require further study.

Evidence of Integration from Long-Term Canine Studies

To address questions posed from these murine studies, long-term follow-up studies have been performed in large animal models (Table 1). These have allowed evaluation of whether there is potential for longer-term genotoxicity and provided more information on relevance of integration. Long-term follow-up studies have now been reported in both hemophilia A and B dog models. These animals represent an excellent model of severe hemophilia, with animals requiring treatment for spontaneous bleeding episodes. In the case of the hemophilia A dog model, the underlying natural *F8* mutation is similar to the *F8* intron 22 inversion seen in 45% of persons with severe hemophilia A (67, 68). Studies conducted in two hemophilia B dogs treated with an AAV-(ApoE)4/hAAT-cFIX followed up for 5.5-6 years demonstrated no conclusive evidence of integration using an earlier LAM-PCR based approach (69). Studies have recently been presented from two long-term follow-up studies in the hemophilia A dog model (70, 71). In both studies, sustained FVIII expression was seen with reduction in the annualized bleed rate. The first study, performed in the University of North Carolina colony, followed up 9 dogs for between 2.2 and 10.1 years, using either single chain or two chain AAV-FVIII constructs. Integration was seen throughout the canine genome, with an increase in transcriptional units, near CpG islands, and modestly increased close to cancer-associated genes. Recurrent integration was seen in association with EGR2, DUSP1, ALB, CCND1 and EGR3. Two dogs demonstrated a rise in FVIII activity 4 years after dosing, with evidence of clonal expansion of hepatocytes containing integrated vectors, although with no evidence of malignant transformation. Data from the Queen's University dog colony has provided further insights into the long-term fate of rAAV after more than a decade of follow-up. In this study, eight severe hemophilia A dogs were treated with a portal vein infusion of an AAV-canine FVIII-SQ construct (71-73). Episomal AAV was detected in the liver using orthogonal methods, with targeted sequencing demonstrating that vector-vector forms comprised the majority of sequencing reads (average=95%). Integration was seen in all liver samples, at an average frequency of 9.3e-4 IS/cell (74). Integration was seen throughout the canine genome, with most events falling in intergenic regions and with no significant enrichment in proximity to cancer

genes. Common integration events were seen in proximity to the canine *KCNIP2*, *CLIC2*, and *ABCB1* genes and the native *F8* locus, with no evidence of dysregulation of gene expression. Integration was seen in areas of increased chromatin accessibility. Most liver samples demonstrated integration profiles with variable integration site abundance across different dogs and samples, and no obvious clonal dominance. Two biopsies; however, exhibited integration sites with frequencies >30% of the total IS sequences, a threshold set for clonality in retroviral integration studies. Long-read sequencing detected only rearranged and truncated integrated AAV forms, with the investigators concluding that expression was derived predominantly from episomal forms. A recent case report in a privately owned severe hemophilic dog treated with an AAV8-codon optimized canine FVIII-BDD(ΔF/V3) vector has added to discussions on the risk of insertional mutagenesis (75). This describes studies investigating multicentric B-cell lymphoma detected 3.5 years after treatment. Low vector copies were detected in the liver (0.05 copies/diploid genome) and spleen (0.02 copies/diploid genome), with levels below the limit of detection in the lymph nodes. Six integration sites were determined to be potentially clonally expanded (i.e., present in ≥5 cells) in the liver (n=2) or spleen (n=4). The most abundant of these upstream of *TNFRSF19,* was present in only 11 cells. One integration site without clonal expansion was detected in a single cell in the lymph nodes studied. These findings suggest that integration was unlikely to have contributed to tumor formation.

In summary, integration following rAAV was seen in all canine studies at relatively low frequencies. Some overlap in common integration sites has been reported (EGR2/MIR1296, ALB, CCND1 and EGR3/MIR320), which differ from sites seen in murine studies. Some evidence of expansion of clones containing rAAV has been reported, albeit with no evidence of malignant transformation. These findings suggest that some of these clonal integration events may represent benign age-related clonal expansions, although this hypothesis requires further study.

Evidence of Integration from Non-Human Primate Studies

Studies evaluating rAAV integration have also been conducted in non-human primates (NHP) treated in adult life and in utero (Table 1). Intra-uterine gene transfer (IUGT) studies have been reported with both late and early IUGT (76-78). The first study evaluated late IUGT (0.9G) in 12 NHP fetuses using a scAAV-LP1-hFIXco construct (76). Liver samples were evaluated early post-vector delivery in 3 NHP and sequential sampling was also performed in 1 NHP at 2, 6 & 11 months. Most sequencing reads were concatemeric (vector-vector) and sequential biopsies demonstrated no identical integration sites. Follow up studies, have been reported at 18-49 months of age from 5 different NHPs in this cohort, with two having postnatal vector re-administration due to low or falling expression (77). No preference was seen for integration within or near to coding regions, in cancer-related genes, and no

hotspots were seen. In contrast to other studies, most sequencing reads corresponded to integration sites (87.4-100%). Interestingly, in two animals that were retreated postnatally, reappearance of concatemers was seen. Studies have also reported using earlier IUGT (0.35-0.4G) with scAAV-LP1-hFIXco or scAAV-LP1-hFXco vectors (78). In this study, 43.9-73.7% of sequencing reads represented integration events at 10 months, with this increasing to 87.5% at 42 months. No targeting of integration to specific chromosomes or gene regions was seen. No enrichment of IS in proximity to cancer-related genes was seen. Within the most frequently occurring IS, only two corresponded to events near cancer-related genes (RAD23B & RAB1A), with these not seen at consecutive timepoints in the same or different animals. In summary, higher proportions of integrated compared to concatemeric sequencing reads were seen in two of these studies. Although these results are based on a small number of samples and integration loci, one interpretation of these IUGT studies is that treatment during a time of active cellular division facilitates rAAV integration. Importantly, despite the findings of stable integration there was no evidence of malignancy at 4 or 6 years follow up after early or later IUGT (77, 78).

Vector persistence has been evaluated in adult NHPs after intramuscular or liver-directed rAAV (Table 1). Early studies demonstrated persistence of rAAV in episomal forms following intramuscular delivery of an AAV-RSV-LEA29Y-WPRE-pA construct encoding LEA29Y (belatacept), a mutant form of the human immunosuppressive molecule CTLA4-Ig. Vector persistence evaluated after 34 months demonstrated integration frequencies ranging from 1e-5 to 1e-4 IS/diploid genome (79). There were no common integration sites and no increase in integration in CpG islands or within genes. Detailed studies have since been reported in NHPs treated with an rAAV2/5-hcoPBGD construct containing the porphobilinogen deaminase (PBGD) transgene, used in a phase 1 clinical trial for treatment of acute intermittent porphyria. Vector-vector (concatemeric) sequences were the predominant AAV form detected, supporting predominant episomal persistence. Integration was seen in all tissues studied, with frequencies ranging from 7.4e-5 to 1.0e-4~~5~~ IS/cell. Integration was seen throughout the NHP genome, with no common integration sites. Three hemophilia gene therapy NHP studies have also presented data in abstract format. The first of these used an AAV5-hFVIII-SQ construct (n=12), with analysis of liver samples 13-26 weeks post-vector administration (80). Most sequencing reads represented episomal copies (>99.9%) with integration occurring in <1 in 600 cells. Integration was distributed across the NHP genome, with no evidence of clonal expansion or gene-specific targeting. Recurrent integration was seen near active genes, such as the liver-restricted albumin locus. The second abstract, reported a study of NHPs treated with an AAV5-hFIX construct, demonstrated that rAAV persisted almost exclusively in non-integrated episomal forms (62). Integration events were distributed throughout the NHP genome with no IS clusters. The third study, reported long-term

follow-up of 7 NHPs treated with the scAAV-LP1-hFIXco construct used in the UCL/St Jude clinical study (81). Sequential liver biopsies performed at weeks 1-8 and 1-4.6 years demonstrated sequencing reads predominantly resulted from concatemeric forms. Integration was distributed throughout the macaque genome, with no hotspots or association with known oncogenes.

Two recent studies have provided further detail of integration following rAAV in NHPs (33, 82). The first compared integration in the liver of WT-AAV in naïve NHPs (n=168) to rAAV-treated NHPs (n=86). Included rAAV studies used liver-directed vectors with different constructs/doses and follow up ranging from 7 days to 15 years. Integration was seen in all rAAV treated NHPs, with more unique integration loci (UIL) when compared to WT-AAV. Only a small proportion of rAAV integration sites demonstrated clonal expansion, with the largest consisting of 29 copies of a single UIL. No difference in the number of UILs was seen based on treatment age, transgene, promoter, or timing post-treatment. Random distribution of UILs was seen relating to chromosomal size. Integration was increased in highly expressed genes, open chromatin regions, highly transcriptionally active rRNA repetitive regions, and/or regions susceptible to DNA damage. The second study presented longitudinal data from NHPs treated using different rAAV constructs: macaque derived β-choriogonadotropin hormone (rhβCG), human and rhesus low density lipoprotein receptor (hLDLR and rhLDLR) or GFP (82). Integration frequencies of 0.1–1.6 per 100 genomes were reported at day 182 (rhβCG construct). Time course studies (day 14, 77 and 760) for the LDLR and GFP constructs demonstrated similar integration frequencies, with an initial decline and stabilization (0.1–0.7 per 100 genomes). Integration was seen across the genome, with increased frequencies in and around genes that are highly expressed in the liver. Integration was not seen in regions associated with HCC. Long-read sequencing detected complete integrated vectors, with complex concatemeric structures containing a mixture of rearranged and truncated genomes. These findings challenge assumptions that AAV predominantly integrates as single copies and suggest that integrated vectors may play a role in mediating long-term transgene expression. Most importantly, these variable findings may well result from differences in the analytical methodologies utilized, and indicate that further studies are required to resolve these discrepancies (30).

Mechanism of Persistence and Integration of rAAV Vectors in Humans

There are limited studies evaluating rAAV vector integration in clinical studies (Table 2). The first study reported on integration after intramuscular delivery of an AAV1-LPL^{S447X} vector for treatment of lipoprotein lipase deficiency (83). In this study, muscle biopsies (n=5) were obtained 14-52 weeks after treatment (84). Most rAAV copies persisted in episomal forms with rearrangements and largely deleted vector ITRs. Integration frequencies were reported at 1e-4 to 1e-5 IS/cell, with no preferential

integration sites within genes, CpG islands, palindromic sequences or AAVS1. Common integration sites were seen near OR4F29 and within PCBD2. Interestingly alignment of IS to the mitochondrial genome, revealed greater homology than to the nuclear genome. Transcriptionally active episomal rAAV has also been demonstrated in biopsies (n=8), 3-12 months post intramuscular delivery for the treatment of Alpha-1-Antitrypsin deficiency (85).

Integration studies have been reported after liver-directed gene therapy for the treatment of acute intermittent porphyria. In a phase 1 study of a rAAV2/5-cohPBGD vector liver transduction was seen at 1 year in six participants, with a trend towards reduced hospitalization and heme treatment, but with no metabolic correction (86). Liver biopsies (n=3) demonstrated integration frequency of 1.17e-3 IS/cell, with no common integration sites, or integration in regions associated with HCC (87). Recent data reported from liver biopsies obtained from patients with severe hemophilia A (n=5) treated with an AAV5-HLP-cohFVIII-SQ construct, has provided further evidence of episomal persistence after 2.6-4.1 years follow-up (88). Integration studies have been presented in abstract form from 5 participants treated with this AAV5-HLP-cohFVIII-SQ construct, after 0.5-4.1 years follow up (89). Integration was seen in all samples at an average frequency of 3.97e-3 IS/cell (range 1.33-5.71e-3 IS/cell). Integration was seen throughout the human genome, with mild enrichment in proximity to liver-expressed genes. There was no evidence of enrichment in proximity to cancer genes or clonal expansion. Further data is awaited from this and another study evaluating liver biopsy samples after long-term follow-up in patients treated in the UCL/St. Jude's & Freeline studies (ClinicalTrials.gov Identifier: NCT04817462).

Investigation of Malignancy in Clinical Gene Therapy Studies

The key clinical concerns regarding integration is whether this can drive the formation of malignancy, as has been reported in murine studies and for retroviral vectors (90). Data in the previous sections, describes predominantly random rAAV integration sites, with no common integration sites across species. With theoretical concerns that integration could drive clonal expansion, robust protocols have been initiated to investigate malignancies occurring in participants in rAAV studies (Figure 2). Seven different malignancies have been reported in individuals in rAAV gene therapy studies, with full reports published on 4 of these cases (Table 3) (91-94). Molecular studies have suggested no evidence of vector integration contributing to these tumors. A detailed case report has been published from a man with severe hemophilia B who was found to have developed hepatocellular carcinoma on routine imaging, 1 year after treatment with an AAV5-hFIXco-Padua construct (92). Analyses from a resected secondary liver lesion were compared to healthy surrounding liver tissue. Similar proportions of total AAV copies were seen in the tumor versus surrounding normal tissue (3.2 v 4.11 copies/cell). Many of these copies represented vector-vector sequencing reads, in keeping with episomal forms. Integration

frequencies were reported at 1 in 10,000 IS/cell, with no dominant IS and no evidence of integration-derived clonal expansion. Profiling using WGS demonstrated mutations seen in HCC, unrelated to rAAV, including chromosome 7 deletions and mutations in HCC genes (TP53, NFE2L2 & PTPRK). Transcriptomics demonstrated differential expression independent of rAAV of COL1A1, LCN2, AEBP1 and CRP. The conclusion of these studies was that molecular/integration profiles demonstrated no relationship between this tumor and rAAV. The second report describes a tonsillar tumor in a man with severe hemophilia B treated with an AAV8.sc-TTR-FIXR338Lopt vector. No evidence of integration was found in this tumor (91). There has been a recent report of an epithelioid neoplasm of the spinal cord occurring in a child 14 months after treatment for spinal muscular atrophy (93). Vector presence was demonstrated by in-situ hybridization with broad distribution in tumor cells, which was absent in non-neoplastic stroma and infiltrating immune cells. This analysis is unable to differentiate episomal from integrated rAAV. Integration studies failed to detect high confidence integration sites (5 IS, across 3 samples), with analyses being inconclusive due to limited tissue. Data has been published on a parotid acinic cell carcinoma, diagnosed in year 6 after treatment with an AAV5-HLP-hFVIII-SQ construct for hemophilia A. No increase in integration frequency was seen within the tumor (healthy: 1.00e-4 per diploid cell v tumor: ~~8.25e-5~~ 6.49e-5 per diploid cell) and there was no evidence of clonal expansion. Data has also been presented in abstract format on a case of acute lymphoblastic leukemia after treatment with an AAV5-HLP-hFVIII-SQ construct, with no evidence of rAAV-associated insertional mutagenesis (89). Finally, data presented in abstract form from the UCL/St Jude FIX gene therapy study, reported cases of lung adenocarcinoma and prostate carcinoma in two study participants after long-term follow-up with no evidence of causality (95). Further data on these three cases are awaited. Collectively, these studies have shown no evidence of vector integration contributing to tumor formation.

Formatted: Not Highlight

Discussion

Studies over the last two decades have demonstrated that despite the majority of AAV particles persisting in episomal forms, that both wild-type and recombinant AAV integrate into the host genome. There have been variable reports on the potential implications of integration for rAAV vectors. Although some murine studies demonstrated recurrent integration with development of hepatocellular carcinoma, these findings have not been recapitulated in other small or large animal models. Data from a small numbers of human liver biopsies post rAAV have demonstrated integration at similar frequencies to large animal studies, without recurrent sites of integration or insertional mutagenesis. Nevertheless, despite these findings, insertional oncogenesis remains a theoretical concern. There is a need for greater understanding of the natural history of rAAV to assess potential short and long-term side effects. Further studies are required to characterize the source(s) of

transgene expression, whether from episomal or integrated forms. This will provide additional data on long-term efficacy and help inform shared decision making for individuals considering gene therapy. Finally, with the commercial availability of rAAV vectors, robust national and international infrastructures are required for post-marketing monitoring for adverse events (96).

Acknowledgements

Figures created in biorender.com

Author Contribution

PB wrote the first draft. PB and DL revised the manuscript.

Conflicts of Interest

PB has received research support from BioMarin, and consulting fees or honoraria from BioMarin, Octapharma, Novo Nordisk, CSL Behring, Pfizer, and Institute for Nursing and Medication Education (IMNE). DL has received research support from BioMarin, CSL-Behring, and Sanofi and has received consulting fees or honoraria from BioMarin, CSL-Behring, Novo Nordisk, Pfizer, and Sanofi.

References

1. Pipe SW, Leebeek FWG, Recht M, Key NS, Castaman G, Miesbach W, et al. Gene Therapy with Etranacogene Dezaparvovec for Hemophilia B. The New England journal of medicine. 2023;388(8):706-18.
2. Mahlangu J, Kaczmarek R, von Drygalski A, Shapiro S, Chou SC, Ozelo MC, et al. Two-Year Outcomes of Valoctocogene Roxaparvovec Therapy for Hemophilia A. The New England journal of medicine. 2023;388(8):694-705.
3. Wang JH, Gessler DJ, Zhan W, Gallagher TL, Gao G. Adeno-associated virus as a delivery vector for gene therapy of human diseases. Signal Transduct Target Ther. 2024;9(1):78.
4. Atchison RW, Casto BC, Hammon WM. Adenovirus-Associated Defective Virus Particles. Science. 1965;149(3685):754-6.
5. Hoggan MD, Blacklow NR, Rowe WP. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. Proceedings of the National Academy of Sciences of the United States of America. 1966;55(6):1467-74.
6. Li C, Samulski RJ. Engineering adeno-associated virus vectors for gene therapy. Nat Rev Genet. 2020;21(4):255-72.
7. Issa SS, Shaimardanova AA, Solovyeva VV, Rizvanov AA. Various AAV Serotypes and Their Applications in Gene Therapy: An Overview. Cells. 2023;12(5).
8. Meyer NL, Chapman MS. Adeno-associated virus (AAV) cell entry: structural insights. Trends Microbiol. 2022;30(5):432-51.
9. Galibert L, Hyvonen A, Eriksson RAE, Mattola S, Aho V, Salminen S, et al. Functional roles of the membrane-associated AAV protein MAAP. Sci Rep. 2021;11(1):21698.
10. Cao M, You H, Hermonat PL. The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication. PLoS One. 2014;9(8):e104596.
11. Cao M, Chiriva-Internati M, Hermonat PL. AAV2 X increases AAV6 rep/cap-driven rAAV production. Virology. 2015;482:84-8.

12. Logan GJ, Dane AP, Hallwirth CV, Smyth CM, Wilkie EE, Amaya AK, et al. Identification of liver-specific enhancer-promoter activity in the 3' untranslated region of the wild-type AAV2 genome. Nat Genet. 2017;49(8):1267-73.
13. Calcedo R, Vandenberghe LH, Gao G, Lin J, Wilson JM. Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. J Infect Dis. 2009;199(3):381-90.
14. Boutin S, Monteilhet V, Veron P, Leborgne C, Benveniste O, Montus MF, Masurier C. Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. Hum Gene Ther. 2010;21(6):704-12.
15. Li C, Narkbunnam N, Samulski RJ, Asokan A, Hu G, Jacobson LJ, et al. Neutralizing antibodies against adeno-associated virus examined prospectively in pediatric patients with hemophilia. Gene Ther. 2012;19(3):288-94.
16. Mimuro J, Mizukami H, Shima M, Matsushita T, Taki M, Muto S, et al. The prevalence of neutralizing antibodies against adeno-associated virus capsids is reduced in young Japanese individuals. J Med Virol. 2014;86(11):1990-7.
17. Stanford S, Pink R, Creagh D, Clark A, Lowe G, Curry N, et al. Adenovirus-associated antibodies in UK cohort of hemophilia patients: A seroprevalence study of the presence of adenovirus-associated virus vector-serotypes AAV5 and AAV8 neutralizing activity and antibodies in patients with hemophilia A. Res Pract Thromb Haemost. 2019;3(2):261-7.
18. Boyce S, James I, Rangarajan S, Curry N, Bagot C, Austin S, et al. Seroprevalence to adeno-associated virus type 6 in people with hemophilia B from a UK adult cohort. Res Pract Thromb Haemost. 2022;6(4):e12705.
19. Kashiwakura Y, Baatartsogt N, Yamazaki S, Nagao A, Amano K, Suzuki N, et al. The seroprevalence of neutralizing antibodies against the adeno-associated virus capsids in Japanese hemophiliacs. Mol Ther Methods Clin Dev. 2022;27:404-14.
20. Klamroth R, Hayes G, Andreeva T, Gregg K, Suzuki T, Mitha IH, et al. Global Seroprevalence of Pre-existing Immunity Against AAV5 and Other AAV Serotypes in People with Hemophilia A. Hum Gene Ther. 2022;33(7-8):432-41.
21. Schulz M, Levy DI, Petropoulos CJ, Bashirians G, Winburn I, Mahn M, et al. Binding and neutralizing anti-AAV antibodies: Detection and implications for rAAV-mediated gene therapy. Mol Ther. 2023;31(3):616-30.
22. Meier AF, Fraefel C, Seyffert M. The Interplay between Adeno-Associated Virus and its Helper Viruses. Viruses. 2020;12(6).
23. Schnepp BC, Jensen RL, Chen CL, Johnson PR, Clark KR. Characterization of adeno-associated virus genomes isolated from human tissues. J Virol. 2005;79(23):14793-803.
24. La Bella T, Imbeaud S, Peneau C, Mami I, Datta S, Bayard Q, et al. Adeno-associated virus in the liver: natural history and consequences in tumour development. Gut. 2020;69(4):737-47.
25. Snyder RO, Im DS, Ni T, Xiao X, Samulski RJ, Muzyczka N. Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. J Virol. 1993;67(10):6096-104.
26. Philpott NJ, Giraud-Wali C, Dupuis C, Gomos J, Hamilton H, Berns KI, Falck-Pedersen E. Efficient integration of recombinant adeno-associated virus DNA vectors requires a p5-rep sequence in cis. J Virol. 2002;76(11):5411-21.
27. Im DS, Muzyczka N. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell. 1990;61(3):447-57.
28. Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, et al. Site-specific integration by adeno-associated virus. Proceedings of the National Academy of Sciences of the United States of America. 1990;87(6):2211-5.
29. Huser D, Gogol-Doring A, Lutter T, Weger S, Winter K, Hammer EM, et al. Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome. PLoS Pathog. 2010;6(7):e1000985.

30. Janovitz T, Sadelain M, Falck-Pedersen E. Adeno-associated virus type 2 preferentially integrates single genome copies with defined breakpoints. Virol J. 2014;11:15.
31. Huser D, Gogol-Doring A, Chen W, Heilbronn R. Adeno-associated virus type 2 wild-type and vector-mediated genomic integration profiles of human diploid fibroblasts analyzed by third-generation PacBio DNA sequencing. J Virol. 2014;88(19):11253-63.
32. Janovitz T, Oliveira T, Sadelain M, Falck-Pedersen E. Highly divergent integration profile of adeno-associated virus serotype 5 revealed by high-throughput sequencing. J Virol. 2014;88(5):2481-8.
33. Martins KM, Breton C, Zheng Q, Zhang Z, Latshaw C, Greig JA, Wilson JM. Prevalent and Disseminated Recombinant and Wild-Type Adeno-Associated Virus Integration in Macaques and Humans. Hum Gene Ther. 2023;34(21-22):1081-94.
34. Morfopoulou S, Buddle S, Torres Montaguth OE, Atkinson L, Guerra-Assuncao JA, Moradi Marjaneh M, et al. Genomic investigations of unexplained acute hepatitis in children. Nature. 2023;617(7961):564-73.
35. Ho A, Orton R, Tayler R, Asamaphan P, Herder V, Davis C, et al. Adeno-associated virus 2 infection in children with non-A-E hepatitis. Nature. 2023;617(7961):555-63.
36. Servellita V, Sotomayor Gonzalez A, Lamson DM, Foresythe A, Huh HJ, Bazinet AL, et al. Adeno-associated virus type 2 in US children with acute severe hepatitis. Nature. 2023;617(7961):574-80.
37. Nault JC, Datta S, Imbeaud S, Franconi A, Mallet M, Couchy G, et al. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. Nat Genet. 2015;47(10):1187-93.
38. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat Genet. 2016;48(5):500-9.
39. Schaffer AA, Dominguez DA, Chapman LM, Gertz EM, Budhu A, Forgues M, et al. Integration of adeno-associated virus (AAV) into the genomes of most Thai and Mongolian liver cancer patients does not induce oncogenesis. BMC Genomics. 2021;22(1):814.
40. Qin W, Xu G, Tai PWL, Wang C, Luo L, Li C, et al. Large-scale molecular epidemiological analysis of AAV in a cancer patient population. Oncogene. 2021;40(17):3060-71.
41. Oziolor EM, Kumpf SW, Qian J, Gosink M, Sheehan M, Rubitski DM, et al. Comparing molecular and computational approaches for detecting viral integration of AAV gene therapy constructs. Mol Ther Methods Clin Dev. 2023;29:395-405.
42. Clark KR, Sferra TJ, Johnson PR. Recombinant adeno-associated viral vectors mediate long-term transgene expression in muscle. Hum Gene Ther. 1997;8(6):659-69.
43. Fisher KJ, Jooss K, Alston J, Yang Y, Haecker SE, High K, et al. Recombinant adeno-associated virus for muscle directed gene therapy. Nat Med. 1997;3(3):306-12.
44. Herzog RW, Hagstrom JN, Kung SH, Tai SJ, Wilson JM, Fisher KJ, High KA. Stable gene transfer and expression of human blood coagulation factor IX after intramuscular injection of recombinant adeno-associated virus. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(11):5804-9.
45. Snyder RO, Spratt SK, Lagarde C, Bohl D, Kaspar B, Sloan B, et al. Efficient and stable adeno-associated virus-mediated transduction in the skeletal muscle of adult immunocompetent mice. Hum Gene Ther. 1997;8(16):1891-900.
46. Miao CH, Snyder RO, Schowalter DB, Patijn GA, Donahue B, Winther B, Kay MA. The kinetics of rAAV integration in the liver. Nat Genet. 1998;19(1):13-5.
47. Nakai H, Iwaki Y, Kay MA, Couto LB. Isolation of recombinant adeno-associated virus vector-cellular DNA junctions from mouse liver. J Virol. 1999;73(7):5438-47.
48. Nakai H, Yant SR, Storm TA, Fuess S, Meuse L, Kay MA. Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. J Virol. 2001;75(15):6969-76.

49. Schnepp BC, Clark KR, Klemanski DL, Pacak CA, Johnson PR. Genetic fate of recombinant adeno-associated virus vector genomes in muscle. J Virol. 2003;77(6):3495-504.
50. Donsante A, Vogler C, Muzyczka N, Crawford JM, Barker J, Flotte T, et al. Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. Gene Ther. 2001;8(17):1343-6.
51. Donsante A, Miller DG, Li Y, Vogler C, Brunt EM, Russell DW, Sands MS. AAV vector integration sites in mouse hepatocellular carcinoma. Science. 2007;317(5837):477.
52. Wang PR, Xu M, Toffanin S, Li Y, Llovet JM, Russell DW. Induction of hepatocellular carcinoma by in vivo gene targeting. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(28):11264-9.
53. Zhong L, Malani N, Li M, Brady T, Xie J, Bell P, et al. Recombinant adeno-associated virus integration sites in murine liver after ornithine transcarbamylase gene correction. Hum Gene Ther. 2013;24(5):520-5.
54. Walia JS, Altaleb N, Bello A, Kruck C, LaFave MC, Varshney GK, et al. Long-term correction of Sandhoff disease following intravenous delivery of rAAV9 to mouse neonates. Mol Ther. 2015;23(3):414-22.
55. Kang HR, Gjorgjieva M, Smith SN, Brooks ED, Chen Z, Burgess SM, et al. Pathogenesis of Hepatic Tumors following Gene Therapy in Murine and Canine Models of Glycogen Storage Disease. Mol Ther Methods Clin Dev. 2019;15:383-91.
56. Li Y, Miller CA, Shea LK, Jiang X, Guzman MA, Chandler RJ, et al. Enhanced Efficacy and Increased Long-Term Toxicity of CNS-Directed, AAV-Based Combination Therapy for Krabbe Disease. Mol Ther. 2021;29(2):691-701.
57. Bell P, Wang L, Lebherz C, Flieder DB, Bove MS, Wu D, et al. No evidence for tumorigenesis of AAV vectors in a large-scale study in mice. Mol Ther. 2005;12(2):299-306.
58. Embury JE, Frost S, Charron CE, Madrigal E, Perera O, Poirier AE, et al. Hepatitis Virus Protein X-Phenylalanine Hydroxylase fusion proteins identified in PKU mice treated with AAV-WPRE vectors. Gene Therapy and Molecular Biology. 2008;12(1):69-76.
59. Reiss J, Hahnewald R. Molybdenum cofactor deficiency: Mutations in GPHN, MOCS1, and MOCS2. Hum Mutat. 2011;32(1):10-8.
60. Li H, Malani N, Hamilton SR, Schlachterman A, Bussadori G, Edmonson SE, et al. Assessing the potential for AAV vector genotoxicity in a murine model. Blood. 2011;117(12):3311-9.
61. Kao CY, Yang SJ, Tao MH, Jeng YM, Yu IS, Lin SW. Incorporation of the factor IX Padua mutation into FIX-Triple improves clotting activity in vitro and in vivo. Thrombosis and haemostasis. 2013;110(2):244-56.
62. Spronck L, de Haan M, Heijink L, Twisk J, Ferreira V, van Deventer S. Assessment of Vector Integration of AAV5-hFIX in Mice and Non-human Primates Indicates No Association with Tumorigenic Risk. Res Pract Thromb Haemost. 2020;4:1.
63. Gauttier V, Pichard V, Aubert D, Kaeppel C, Schmidt M, Ferry N, Conchon S. No tumour-initiating risk associated with scAAV transduction in newborn rat liver. Gene Ther. 2013;20(7):779-84.
64. Sabatino DE, Bushman FD, Chandler RJ, Crystal RG, Davidson BL, Dolmetsch R, et al. Evaluating the state of the science for adeno-associated virus integration: An integrated perspective. Mol Ther. 30 (8):2646-2663. 2022.
65. Dalwadi DA, Torrens L, Abril-Fornaguera J, Pinyol R, Willoughby C, Posey J, et al. Liver Injury Increases the Incidence of HCC following AAV Gene Therapy in Mice. Mol Ther. 2021;29(2):680-90.
66. Kapelanski-Lamoureux A, Chen Z, Gao ZH, Deng R, Lazaris A, Lebeaupin C, et al. Ectopic clotting factor VIII expression and misfolding in hepatocytes as a cause for hepatocellular carcinoma. Mol Ther. 2022;30(12):3542-51.
67. Lozier JN, Dutra A, Pak E, Zhou N, Zheng Z, Nichols TC, et al. The Chapel Hill hemophilia A dog colony exhibits a factor VIII gene inversion. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(20):12991-6.

68. Hough C, Kamisue S, Cameron C, Notley C, Tinlin S, Giles A, Lillicrap D. Aberrant splicing and premature termination of transcription of the FVIII gene as a cause of severe canine hemophilia A: similarities with the intron 22 inversion mutation in human hemophilia. Thrombosis and haemostasis. 2002;87(4):659-65.
69. Niemeyer GP, Herzog RW, Mount J, Arruda VR, Tillson DM, Hathcock J, et al. Long-term correction of inhibitor-prone hemophilia B dogs treated with liver-directed AAV2-mediated factor IX gene therapy. Blood. 2009;113(4):797-806.
70. Nguyen GN, Everett JK, Kafle S, Roche AM, Raymond HE, Leiby J, et al. A long-term study of AAV gene therapy in dogs with hemophilia A identifies clonal expansions of transduced liver cells. Nat Biotechnol. 2021;39(1):47-55.
71. Batty P, Mo AM, Hurlbut D, Ishida J, Yates B, Brown C, et al. Long-term follow-up of liver-directed, adeno-associated vector-mediated gene therapy in the canine model of hemophilia A. Blood. 2022;140(25):2672-83.
72. Jiang H, Lillicrap D, Patarroyo-White S, Liu T, Qian X, Scallan CD, et al. Multiyear therapeutic benefit of AAV serotypes 2, 6, and 8 delivering factor VIII to hemophilia A mice and dogs. Blood. 2006;108(1):107-15.
73. Scallan CD, Lillicrap D, Jiang H, Qian X, Patarroyo-White SL, Parker AE, et al. Sustained phenotypic correction of canine hemophilia A using an adeno-associated viral vector. Blood. 2003;102(6):2031-7.
74. Batty P, Fong S, Franco M, Sihn CR, Swystun LL, Afzal S, et al. Vector integration and fate in the hemophilia dog liver multi-years following AAV-FVIII gene transfer. Blood. 2024.
75. Van Gorder L, Doshi BS, Willis E, Nichols TC, Cook E, Everett JK, et al. Analysis of vector genome integrations in multicentric lymphoma after AAV gene therapy in a severe hemophilia A dog. Mol Ther Methods Clin Dev. 2023;31:101159.
76. Mattar CN, Nathwani AC, Waddington SN, Dighe N, Kaeppel C, Nowrouzi A, et al. Stable human FIX expression after 0.9G intrauterine gene transfer of self-complementary adeno-associated viral vector 5 and 8 in macaques. Mol Ther. 2011;19(11):1950-60.
77. Mattar CNZ, Gil-Farina I, Rosales C, Johana N, Tan YYW, McIntosh J, et al. In Utero Transfer of Adeno-Associated Viral Vectors Produces Long-Term Factor IX Levels in a Cynomolgus Macaque Model. Mol Ther. 2017;25(8):1843-53.
78. Chan JKY, Gil-Farina I, Johana N, Rosales C, Tan YW, Ceiler J, et al. Therapeutic expression of human clotting factors IX and X following adeno-associated viral vector-mediated intrauterine gene transfer in early-gestation fetal macaques. FASEB J. 2019;33(3):3954-67.
79. Nowrouzi A, Penaud-Budloo M, Kaeppel C, Appelt U, Le Guiner C, Moullier P, et al. Integration frequency and intermolecular recombination of rAAV vectors in non-human primate skeletal muscle and liver. Mol Ther. 2012;20(6):1177-86.
80. Sullivan L, Zahn M, Gil-Farina I, Kasprzyk T, O'Neill CA, Eggan K, et al. Rare genomic integrations of AAV5-hFVIII-SQ occur without evidence of clonal activation or gene-specific targeting. Mol Ther 2021;29(4):425.
81. McIntosh JH, Kaeppel C, Davidoff A, Morton CL, Schmidt M, Nathwani AC. Assessment of the integration profile of self-complementary AAV vector over time in non-human primates. Hum Gene Ther. 2017;28(12):A1 - A125.
82. Greig JA, Martins KM, Breton C, Lamontagne RJ, Zhu Y, He Z, et al. Integrated vector genomes may contribute to long-term expression in primate liver after AAV administration. Nat Biotechnol. 2023.
83. Kaeppel C, Beattie SG, Fronza R, van Logtenstein R, Salmon F, Schmidt S, et al. A largely random AAV integration profile after LPLD gene therapy. Nat Med. 2013;19(7):889-91.
84. Carpentier AC, Frisch F, Labbe SM, Gagnon R, de Wal J, Greentree S, et al. Effect of alipogene tiparvovec (AAV1-LPL(S447X)) on postprandial chylomicron metabolism in lipoprotein lipase-deficient patients. J Clin Endocrinol Metab. 2012;97(5):1635-44.

85. Schnepp BC, Chulay JD, Ye GJ, Flotte TR, Trapnell BC, Johnson PR. Recombinant Adeno-Associated Virus Vector Genomes Take the Form of Long-Lived, Transcriptionally Competent Episomes in Human Muscle. Hum Gene Ther. 2016;27(1):32-42.
86. D'Avola D, Lopez-Franco E, Sangro B, Paneda A, Grossios N, Gil-Farina I, et al. Phase I open label liver-directed gene therapy clinical trial for acute intermittent porphyria. J Hepatol. 2016;65(4):776-83.
87. Gil-Farina I, Fronza R, Kaeppel C, Lopez-Franco E, Ferreira V, D'Avola D, et al. Recombinant AAV Integration Is Not Associated With Hepatic Genotoxicity in Nonhuman Primates and Patients. Mol Ther. 2016;24(6):1100-5.
88. Fong S, Yates B, Sihn CR, Mattis AN, Mitchell N, Liu S, et al. Interindividual variability in transgene mRNA and protein production following adeno-associated virus gene therapy for hemophilia A. Nat Med. 2022;28(4):789-97.
89. Eggan K. Molecular Characterization of Recombinant AAV5 Encoding FVIII After Human Administration. Hum Gene Ther. 2022;33.
90. Bushman FD. Retroviral Insertional Mutagenesis in Humans: Evidence for Four Genetic Mechanisms Promoting Expansion of Cell Clones. Mol Ther. 2020;28(2):352-6.
91. Konkle BA, Walsh CE, Escobar MA, Josephson NC, Young G, von Drygalski A, et al. BAX 335 hemophilia B gene therapy clinical trial results: potential impact of CpG sequences on gene expression. Blood. 2021;137(6):763-74.
92. Schmidt M, Foster GR, Coppens M, Thomsen H, Dolmetsch R, Heijink L, et al. Molecular Evaluation and Vector Integration Analysis of HCC Complicating AAV Gene Therapy for Hemophilia B. Blood Adv. 2023.
93. Retson L, Tiwari N, Vaughn J, Bernes S, Adelson PD, Mansfield K, et al. Epithelioid neoplasm of the spinal cord in a child with spinal muscular atrophy treated with onasemnogene abeparvovec. Mol Ther. 2023;31(10):2991-8.
94. Symington E, Rangarajan S, Lester W, Madan B, Pierce GF, Raheja P, et al. Long-term safety and efficacy outcomes of valoctocogene roxaparvovec gene transfer up to 6 years post-treatment. Haemophilia. 2024;30(2):320-30.
95. Reiss U, Davidoff AM, Tuddenham EGD, Chowdary P, McIntosh J, Riddell A, et al. Stable Therapeutic Transgenic FIX Levels for More Than 10 Years in Subjects with Severe Hemophilia B Who Received scAAV2/8-LP1-Hfixco Adeno-Associated Virus Gene Therapy. Blood. 2023;142:1056.
96. Konkle B, Pierce G, Coffin D, Naccache M, Clark RC, George L, et al. Core data set on safety, efficacy, and durability of hemophilia gene therapy for a global registry: Communication from the SSC of the ISTH. Journal of thrombosis and haemostasis : JTH. 2020;18(11):3074-7.

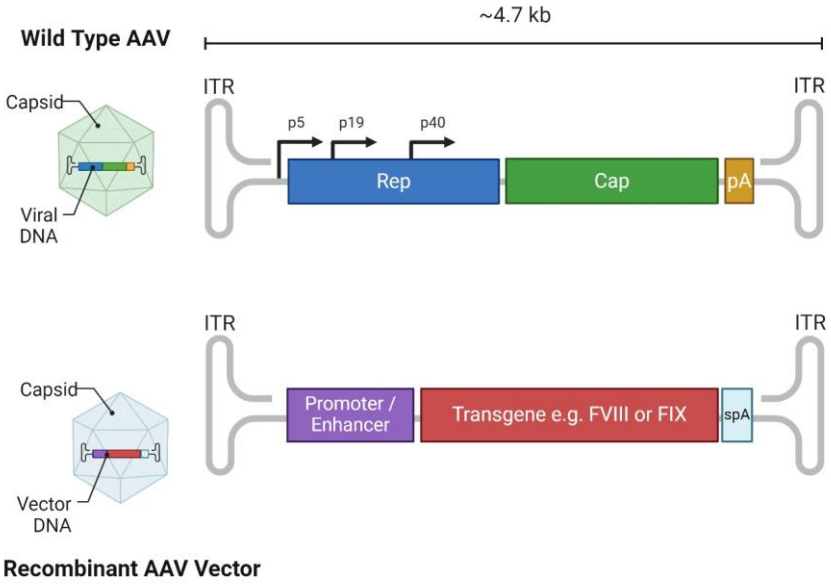

Figure 1: Comparison of the Structure of Wild Type Adeno-associated Virus and Recombinant Adeno-associated Viral (AAV) vectors. ITR = inverted terminal repeat. pA = poly-adenylation sequence. kb = kilobase.

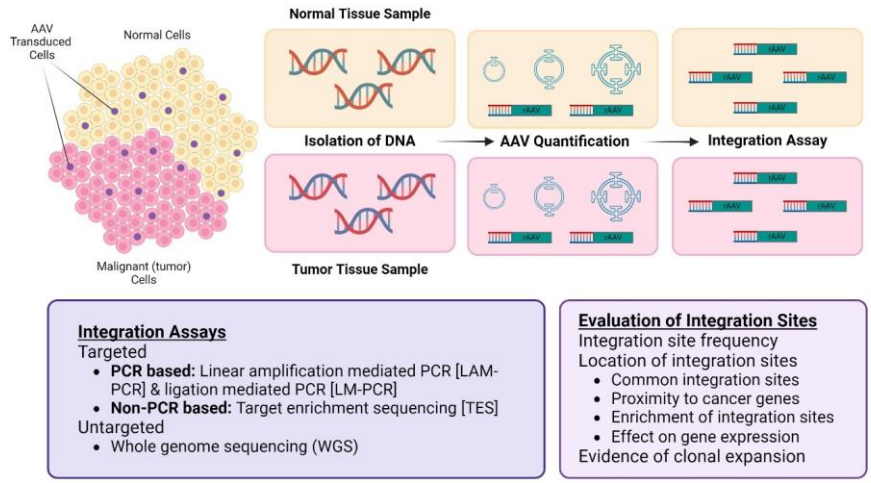

Figure 2: Summary of Current Strategies to Investigate Malignancies in Gene Therapy Trials. Investigations involve vector quantification (integration & episomal) and specific integration studies within a tumor biopsy (pink) and surrounding healthy tissue (yellow). In the example shown similar numbers of cells have been transduced with AAV with no clear evidence of clonal expansion.

Canine Studies												
Vector	State	Capsid	Dose (vg/kg)	n (bx)	Route	FU	Assay	Unique IS	Integration Frequency[+]	CIS	Integration Site Features	Reference
AAV2-ApoEhAAT-cFIX	HB	2	$1.1x10^{12}$ - $1.2x10^{12}$	2	IV	5.5 - 6 years	LAMPCR	Nil	Nil	No	No IS sites detected	Niemeyer et al (59)
ssAAV-TBG-cFVIII ssAAV-hAAT-cFVIII	HA	8 & 9	$1.2 x10^{13}$ - $4.0 x10^{13}$	6 (20)	PV IV	2.3 – 10.1 years	LMPCR	1741	nr	Yes	IS throughout the canine genome Integration favored in transcription units, near CpG islands and areas of active transcription Modest increase in IS in association with cancer genes Integrated AAV partially deleted or rearranged Evidence of clonal expansion with no malignant transformation CIS: EGR2. DUSP1, ALB, CCND1 & EGR3	Nguyen et al. (60)
ssAAV-TTR-cFVIII-SQ	HA	2, 6 & 8	$6.0x10^{12}$ - $2.7x10^{13}$	8 (16)	PV	8.2 – 12.0 years	LAMPCR TES	5746	$1.2x10^{-4}$ – $2.7x10^{-3}$	Yes	IS throughout the canine genome Integration predominantly No increase in IS in proximity to cancer genes Increased IS in chromatin accessible regions No full-length integrated vectors retrieved CIS: KCNIP2, CLIC2, ABCB1 & F8 No evidence of dysregulation at CIS	Batty et al. (64)
Non-Human Primate Studies												
Vector	State	Capsid	Dose (vg/kg)	n (bx)	Route	FU	Assay	Unique IS	Integration Frequency[+]	CIS	Integration Site Features	Reference
scAAV-LP1-hFIXco	WT	8	$4x10^{12}$ – $1.5x10^{13}$	3 (5)	IUGT 0.9G	IU – 11 months	LAMPCR	38	nr	No	Random IS	Mattar et al. (66)
scAAV-LP1-hFIXco	WT	5 & 8	$1.4x10^{13}$ – $1.9x10^{13}$	5	IUGT 0.9G	18 – 49 months	LAMPCR	119	1.3 - 74 IS/µg	No	No increase in IS in-gene +/- 10kb No increase in IS in cancer genes	Mattar et al. (67)
scAAV-LP1-hFIXco	WT	5 & 8	$4.1x10^{12}$ – $3.8x10^{13}$	4	IUGT 0.36-0.48G	10 – 42 months	LAMPCR	111	1.3 – 132 IS/µg	No	No targeting of chromosome or gene coding regions +/- 10kb No difference in IS between AAV5 & AAV8 No increase in IS in cancer genes	Chan et al. (68)
AAV-RSV-LEA29Y-WPRE-pA	WT	1 & 8	$5x10^{12}$	5 4	IM IV	14 – 34 months	LAMPCR	L 178 M 153	L: $1.9x10^{-5}$ - $1.5x10^{-4}$ IS/dg	No	No preferential IS hotspots No preference for IS in-gene, CpG islands or cancer genes	Nowrouzi et al. (69)

									M: $1.9x10^{-5}$ - $2.2x10^{-4}$ IS/dg			
AAV2/5-cohPBGD	WT	5	$1x10^{13}$ $5x10^{13}$	3 3	IV	4 weeks	LAMPCR^	353 - 365	$7.4x10^{-5}$ $1.0x10^{-4}$ IS/cell	No	No targeting of IS to specific regions No increase in IS in coding regions +/- 10kb No IS in regions associated with HCC	Gil Farina et al (77)
AAV5-hFVIII-SQ	WT	5	$2 – 6x10^{13}$	12	IV	13-26 weeks	TES	nr	< 1 in 600 liver cells	Yes	Integration across the NHP genome No evidence of clonal expansion <10% IS within 100kb of TSS of cancer genes (read frequency <0.6%) Recurrent IS near active genes including Albumin	Sullivan et al* (70)
AAV5-hFIX	WT	5	$5x10^{11}$ - $9x10^{13}$	4	IV	6 months	LAMPCR	1392	nr	No	Nearly random distribution of low-level IS across the NHP genome	Spronck et al.* (52)
scAAV-LP1-hFIXco	WT	5 & 8	$2x10^{12}$ $2x10^{13}$	6 1	IV (6) PV (1)	1-8 weeks 1-4.6 years	LAMPCR	32	$0 – 4x10^{-4}$ IS/dg	No	Largely random IS profile No IS in proximity to oncogenes or tumor suppressor genes	McIntosh et al* (71)
Multiple	WT	8, rh10, hu37 & 3B	$3x10^{12}$ - $5x10^{13}$	86	IV	7 days – 15 years	ITRSeq	61,685	nr	No	Random distribution of IS relating to chromosomal size No difference in the number of IS based on treatment age, transgene, promoter, or timing post-treatment. IS increased in highly expressed genes, open chromatin regions, rRNA repetitive regions and regions susceptible to DNA damage	Martins et al (24)
AAV8.TBG.rh-β-CG AAVrh10.TBG.rh-β-CG AAV8.TBG.rhLDLR AAV8.TBG.hLDLR AAV8.TBG.GFP	WT	8 & rh10	$1x10^{13}$	6 6 2 2 2	IV	14 – 760 days multiple samples	ITRSeq ONP	nr	1.6 per 100 – 1 per 1000 genomes (rh-β-CG) 0.1 – 0.7 per 100 genomes (GFP/LDLR)	No	IS across the genome. Increased IS frequencies in/and around genes that are highly expressed in the liver. No IS in regions associated with HCC. Integrated full-length genomes detected, with complex concatemeric structures	Greig et al (72)

Table 1: Summary of AAV Integration Findings in Large Animal Models. AAV = adeno-associated virus. vg = vector genomes. n = number of animals in which integration studies were performed. bx = number of biopsies/samples analyzed. FU = follow up. IS = integration site. CIS = common integration sites. HA = hemophilia A. HB = hemophilia B. WT = wild-type. G = gestation. LAMPCR = linear amplification mediated PCR. LMPCR = ligation mediated PCR. TES = target enrichment sequencing. ONP = Oxford nanopore sequencing. HCC = hepatocellular carcinoma. NHP = non-human primate. IV = intravenous. IM = intramuscular. PV = portal vein. IUGT = Intrauterine gene therapy. nr = not reported. kb = kilobase. L = liver. M = muscle. ^Conventional and multiplex LAM-PCR. ⁺Integration frequency as reported, note units used vary in different studies. *published in abstract format only.

Clinical (Human) Studies												
Vector	State	Capsid	Dose (vg/kg)	n	Route	FU	Assay	Unique IS	Integration Frequency[+]	CIS	Integration Site Features	Reference
AAV1-LPL-S447X	LPLD	1	$1x10^{12}$	5	IM	14-52 weeks	LAMPCR	1969	$5.2x10^{-5}$ – $8.3x10^{-4}$ IS/cell	Yes	No increase in IS within genes, CpG islands, palindromic sequences or ribosomal DNA CIS: OR4F29 & PCBD2. Greater homology to mitochondrial genome	Kaeppel et al (73)
AAV2/5-cohPBGD	AIP	5	$5x10^{11}$ - $6x10^{12}$	3	IV	52 weeks	LAMPCR	134	$2.3x10^{-4}$ - $4.2x10^{-3}$ IS/cell	No	No targeting of IS to specific genomic regions No increase in IS in coding regions +/- 10kb No IS in regions associated with HCC	Gil-Farina et al (77)
AAV5-hFVIII-SQ	HA	5	$4x10^{13}$ - $6x10^{13}$	5	IV	0.5 – 4.1 years	TES	541	$1.3\ x10^{-3}$– $5.71x10^{-3}$ IS/cell	nr	Integration across human genome Mild enrichment in proximity to liver-expressed genes No evidence of enrichment in proximity to cancer genes or clonal expansion	Eggan et al* (79)

Table 2: Summary of AAV Integration Findings in Clinical Studies. AAV = adeno-associated virus. vg = vector genomes. n = number of participants in which integration studies were performed. FU = follow up at biopsy sample. IS = integration site. CIS = common integration sites. HA = hemophilia A. LPLD = lipoprotein lipase deficiency. AIP = acute intermittent porphyria. LAMPCR = linear amplification mediated PCR. TES = target enrichment sequencing. HCC = hepatocellular carcinoma. IV = intravenous. IM = intramuscular. nr = not reported. kb = kilobase. [+]Integration frequency as reported, note units used vary in different studies. *published in abstract format only.

Vector	State	Dose (vg/kg)	Tumor	Time Post AAV	Age at diagnosis	Assay	Key Findings (published)	AAV Related	Reference
Published Studies									
scAAV2/8-TTR-FIXR338L	HB	1 x 10^{12}	Tonsillar carcinoma	nr	69 years #	LAMPCR	No vector genomes or integration detected in the tumor	No	Konkle et al. (81)
AAV2/5-LP1-hFIXco-R338L	HB	2 x 10^{13}	Hepatocellular carcinoma	1 year	69 years	SEPTS/LMPCR	**Total VCN (copies/cell):** 3.21 (T) v 4.11 (N) **Integration Frequency:** 1 in 10,000 cells **Unique IS:** 56 (T) v 39 (N) No evidence of clonal expansion	No	Schmidt et al. (82)
scAAV2/9-CMVen/CB-SMN1	SMA	nr	Spinal cord epithelioid neoplasm	1.2 years	16 months	SEPTS/LMPCR	**In Situ Hybridization:** Vector (total) copies detected in tumor cells **IS:** 5 unique in 3 samples. Limited tissue for full analysis	Inconclusive	Retson et al. (83)
ssAAV2/5-HLP-hFVIII-SQ	HA	6 x 10^{13}	Parotid acinic cell carcinoma	c. 6 years	47 years	TES	**Integration Frequency:** 6.49 x 10^{-5} (T) v 1 x 10^{-4} (H) copies/diploid cell **IS** 26 (T1) & 35 (T2) v 37 (H1) & 54 (H2) No evidence of clonal expansion	No	Symington et al. (84)
Conference Abstracts									
ssAAV2/5-HLP-hFVIII-SQ	HA	6 x 10^{13}	B-Acute Lymphoblastic Leukemia (ALL)	nr	c. 20 years	TES	n/a*	No	Eggan et al. (79)
scAAV2/8-LP1-hFIXco	HB	n/a	Non-mucinous lung adenocarcinoma	5 years	48 years	nr	n/a*	No	Reiss et al. (85)
scAAV2/8-LP1-hFIXco	HB	n/a	Prostate adenocarcinoma	12 years	72 years	nr	n/a*	Ongoing studies	Reiss et al. (85)

Table 3: Summary of reported malignancies in clinical studies and the results of integration & molecular studies. HB =hemophilia B. HA = hemophilia A. SMA = spinal muscular atrophy. nr = not reported. LAMPCR = linear amplification mediated PCR. SEPTS/LMPCR = Shearing Extension Primer Tag Selection Ligation-Mediated PCR. TES = target enrichment sequencing. IS = Integration Sites. VCN = vector copy number (integrated & episomal). T = tumor. H = healthy. # = age at consent. * = published in abstract format.