Gene Therapy (2008) 15, 817–822
© 2008 Nature Publishing Group All rights reserved 0969-7128/08 $30.00
www.nature.com/gt

REVIEW

Adeno-associated virus integration: virus versus vector

RH Smith
Laboratory of Biochemical Genetics, National Heart, Lung, and Blood Institute, Bethesda, MD, USA

Although a large percentage of the world population is seropositive for exposure to various strains of adeno-associated virus (AAV), a human parvovirus, AAV has never been identified as an etiologic agent of human disease. Most likely contributing to the pronounced lack of pathogenicity is the fact that AAV is a naturally 'defective' virus that requires a helper virus for productive replication of its genome. Another unusual aspect of wild-type AAV biology is the ability of the virus to establish latent infection by preferential integration of its genome into a specific locus of human chromosome 19. Site-specific integration was a major impetus for the development of recombinant AAV vectors, which typically lack all AAV coding sequences. It was soon realized, however, that expression of at least one species of the virally encoded initiator proteins, Rep78 or Rep68, is necessary for targeted integration of AAV-derived DNA constructs to occur. This article will present a chronological outline of studies characterizing site-specific integration of wild-type AAV sequences and the quasi-random target site selection observed with recombinant AAV vectors.
Gene Therapy (2008) **15,** 817–822; doi:10.1038/gt.2008.55; published online 10 April 2008

Keywords: *AAV; adeno-associated virus; integration*

Introduction

In 1965, Atchison *et al*.[1] published a paper reporting the characterization of a newly discovered viral agent, which they named adeno-associated virus (AAV). AAV particles were first observed by these researchers in electron micrographs prepared from a stock of simian adenovirus that had been serially passaged in primary cultures of rhesus monkey kidney cells. Atchison *et al*. were able to separate the approximately 24-nm AAV particles from the larger (approximately 80-nm) adenovirus virions by ultrafiltration. Upon isolation, it was noted that the partially purified AAV virions failed to replicate autonomously when inoculated onto primary cells; however, AAV propagated readily when the cell cultures were coinfected with both AAV and its adenovirus partner. Thus, AAV appeared to be a defective satellite virus that was dependent upon a coinfecting helper virus for efficient replication. In an important extension of these findings, Hoggan and colleagues[2] described a colloquium report in which they observed that a significant number of human and monkey primary kidney cell lots harbored AAV in a quiescent (that is, latent) state and that infectious AAV particles could be recovered from the latently infected cells after superinfection with adenovirus. These observations prompted investigations into the nature of the AAV genome in latently infected cells.

Correspondence: Dr RH Smith, Laboratory of Biochemical Genetics, National, Heart, Lung, and Blood Institute, Building 10, Room 7N264, Bethesda, MD 20892, USA.
E-mail: smithr@nhlbi.nih.gov
Received 25 February 2008; accepted 27 February 2008; published online 10 April 2008

Integration of wild-type AAV

Early work concerning the characterization of AAV latency was greatly facilitated by the establishment of continuous cell lines harboring AAV in a quiescent state. Berns *et al*.[2] reported that a human bone marrow-derived, fibroblast-like cell line (Ruddle's Detroit 6 cells) that had been infected with 250 infectious units per cell of AAV serotype 2 (AAV-2) maintained viral sequences in a latent state for at least 47 passages. Infection of Detroit 6 cells with significantly lower multiplicities of AAV resulted in progressive loss of AAV-positive cells from the total cell population during serial passage. Upon cloning the infected cells, Berns *et al*. observed that 29% of the clones were able to produce AAV after superinfection with adenovirus, suggesting that the establishment of AAV latency was fairly efficient at high multiplicities of infection. Solution-based hybridization analyses (that is, reassociation kinetics analysis) of AAV-positive Detroit 6 cell subclones indicated a copy number of three to five AAV genome equivalents per diploid cellular genome. To determine whether latently infected Detroit 6 cells harbored AAV sequences episomally or in an integrated state, Cheung *et al*.[3] prepared Hirt extracts of cellular DNA from a latently infected Detroit 6 cell subclone and characterized the DNA fractions by restriction enzyme digestion and Southern blot analysis. It was determined that wild-type AAV sequences were integrated into the host cell genome as

tandem, head-to-tail repeats linked to genomic DNA sequences by the viral inverted terminal repeat elements (or ITRs).

Wild-type AAV preferentially integrates within a specific region of human chromosome 19

In an important contribution to our understanding of integration site selection by wild-type AAV, Kotin and Berns[4] used a bacteriophage lambda-based genomic DNA library derived from latently infected Detroit 6 cells to isolate cellular sequences flanking an AAV provirus. The cellular flanking sequences were used as probes in Southern blot analyses of DNA extracted from 22 independently derived cell lines latently infected with AAV,[5] including clonal cell lines derived from Detroit 6, HeLa and KB cells. Sixty-eight percent of the independently derived clones displayed AAV-induced alterations in the electrophoretic mobility of DNA restriction fragments derived from the same genomic locus (as identified by the flanking sequence probes). In addition, more than half of the affected DNA bands were also positive for AAV genomic sequences by Southern blot analysis. These findings suggested that in latently infected cells, wild-type AAV utilizes a common integration site to establish latent infection. Hybridization analysis of a panel of rodent–human somatic cell hybrids mapped the common AAV integration locus to human chromosome 19. Using an alternative approach to identify AAV integration sites, Samulski *et al*.[6] established latently infected cell lines containing integrated chimeric AAV genomes bearing DNA-binding sites recognized by the bacteriophage lambda repressor protein. Cellular sequences flanking AAV integrants were obtained by restriction digestion of genomic DNA extracted from latently infected cells and enrichment of AAV-positive sequences by a filter-binding technique. Fragments containing cellular flanking sequences were eluted from the filters, cloned and used to probe Southern blots of DNA extracted from at least eight independently derived cell lines harboring latent AAV genomes. In each case, the cellular flanking sequences hybridized to at least one restriction fragment that was also detected by an AAV-specific probe, indicating that in each cell line AAV had utilized a common genomic locus for integration. Further hybridization analysis indicated that the common AAV integration site was conserved in monkey cells, but not in those of canine, bovine, rodent, chicken or yeast origin. In agreement with Kotin *et al*.,[5] analysis of rodent–human somatic cell hybrids indicated that the common AAV integration site occurred on the long arm of human chromosome 19.

To obtain a genomic clone of the intact preintegration locus, Kotin *et al*.[7] screened bacteriophage lambda libraries derived from diploid human WI38 cells with cellular flanking sequence probes representing both left and right integration junctions of an AAV provirus. An 8.2-kbp *Eco*RI fragment of human chromosome 19 bearing the preintegration locus (termed AAVS1) was isolated from the lambda library. DNA sequence analysis of the first 4000 basepairs of AAVS1 revealed that the region was GC-rich, with an overall G+C content of 65%. The first 900 basepairs displayed an even higher G+C content of 82% and met the criteria of a CpG island. CpG islands are commonly associated with actively transcribed chromatin. A putative cyclic AMP response element and several recognition sequences for upstream binding factor 1 were also observed within the CpG island. The last 400 basepairs of the sequenced region of AAVS1 contained a chromosome 19-abundant minisatellite array. PCR analysis of reverse-transcribed RNA (RT-PCR) indicated that the region of AAVS1 downstream of the CpG island may be transcribed, although at apparently low (or possibly tissue-specific) levels as northern blot analysis did not identify mRNA transcripts in the cells examined. A DNase I hypersensitive site, designated DHS-S1, has been mapped to the CpG island within AAVS1.[8] DHS-S1 demonstrates enhancer-like properties when linked to a reporter gene in transient transfection assays and has been shown to function as a chromosomal insulator sequence.[8,9]

AAVS1 and integration

To map *cis*-acting elements of AAVS1 essential for site-specific AAV integration, Giraud *et al*.[10] utilized an Epstein–Barr virus (EBV)-derived shuttle plasmid that can be stably maintained as an episome in eukaryotic cells and subsequently recovered as a bacterial plasmid. Various portions of the AAVS1 locus were cloned into the EBV shuttle plasmid and individual constructs were used to establish shuttle plasmid-bearing cell lines. The various cell lines were infected with AAV-2 at a multiplicity of 20 infectious units per cell. At 48 h post-infection, extrachromosomal DNA was isolated and used to transform bacteria. Integration events were characterized by filter-based hybridization analysis using AAV-specific probes. DNA sequences within the first 510 bp of AAVS1 were found to be both necessary and sufficient for AAV site-specific integration. In addition to containing a large portion of the AAVS1 CpG island, this region was found to contain a tandem GCTC repeat element that can serve as a binding site for the large Rep proteins of AAV[11] (Rep78 and Rep68), as well as a Rep-specific nicking site, known as a terminal resolution site or *trs*. The Rep78/68 proteins are AAV-encoded non-structural proteins, which possess DNA-binding, helicase and site-specific endonuclease activities that are essential for AAV DNA replication.[12,13] Using the EBV shuttle plasmid system, Linden *et al*.[14] finely mapped the sequences of AAVS1 necessary for site-specific integration to an approximately 100 bp AAVS1 region containing the Rep-binding and nicking sites. Mutation of the Rep-binding site within the 510-bp parental DNA fragment abolished site-specific integration events detected by the EBV-based system. Moreover, a 33-bp oligonucleotide bearing the Rep-binding and nicking sites successfully mediated site-specific integration when cloned into the EBV shuttle plasmid, whereas a similar oligonucleotide in which the Rep-nicking site was mutated did not. Importantly, Weitzman *et al*.[11] have demonstrated in biochemical assays that the AAV Rep78/68 proteins can stably bind the GCTC repeat motif within AAVS1 and mediate complex formation between an AAVS1-derived substrate and an AAV ITR element *in vitro*. These findings support the notion that the AAV Rep proteins facilitate site-specific recombination by mediating a synapse between viral and cellular (AAVS1) sequences in AAV infected cells. Furthermore, Urcelay *et al*.[15] demonstrated that a discrete region of AAVS1 containing the Rep-binding and -nicking sites

can serve as a unidirectional origin of replication in the presence of Rep78/68 proteins *in vitro*.

The AAV Rep proteins are essential for targeted integration

Most mobile DNA elements and integrating viruses encode *trans*-acting protein factors (or integrases) that mediate recombination of their cognate genetic element into the host cells genomic sequences. An essential role for the AAV Rep78/68 proteins in site-specific integration has been demonstrated by several research groups. Balagué *et al*.[16] showed that under nonselective conditions, the integration frequency of an ITR-flanked green fluorescent protein (GFP) reporter gene increased from 5% to 47% when the AAV *rep* gene was included in the plasmid-based integration construct. Surosky *et al*.[17] used a PCR-based AAVS1 integration assay to demonstrate that mutation of the AAV *rep* gene within a cloned copy of the wild-type AAV genome abolished detectable site-specific integration of the cloned provirus; however, integration at the AAVS1 locus could be restored by providing the Rep proteins in *trans*. In addition, it was reported that site-specific integration of a recombinant AAV vector encoding a beta-galactosidase reporter gene was undetectable in nonselected, transduced 293 cells but was readily detected if the cells were first transfected with a *rep*-expressing plasmid. In a comparison of wild-type and recombinant AAV integration, Kearns *et al*.[18] used fluorescent *in situ* hybridization analysis to examine a human bronchial epithelial cell line, IB3-1 cells, that had been infected with either wild-type AAV or transduced with recombinant AAV encoding a CFTR cDNA. It was observed that in cells infected with wild-type virus, 94% of AAV-positive metaphase spreads demonstrated integration within chromosome 19. In contrast, no chromosome 19-specific integration events were observed with the recombinant vector.

Integration of recombinant AAV

Integration of recombinant AAV genomic sequences in the absence of the AAV Rep proteins is inefficient and is not targeted to chromosome 19

Kay and colleagues[19] performed partial hepatectomy in mice that had undergone portal vein infusion with recombinant AAV (rAAV) vectors to quantify the extent of vector integration under conditions in which nonintegrated vector genomes would be lost due to cellular division *in vivo*. It was estimated that only about 10% of rAAV genomes integrated into the host cell genome, thus indicating that the majority of vector genomes persisted in an extrachromosomal form *in vivo*.

To study the kinetics of rAAV integration *in vivo*, Miao *et al*.[20] infused mice with rAAV particles encoding a human blood-clotting factor, factor IX, and isolated liver DNA from the animals at various times post-transduction for Southern blot analysis of vector sequences. It was observed that AAV vector DNA persisted within the livers of the transduced animals in a nonintegrated, single-stranded form. The abundance of single-stranded AAV genomes gradually decreased over a period of at least 5 weeks post-transduction. Pulsed-field gel electrophoresis of extracted DNA demonstrated that during this time, the recombinant AAV genomes became incorporated into high-molecular-weight DNA with little indication of long-lived rAAV duplex intermediates. Similarly, Fisher *et al*.[21] found that skeletal muscle of C57Bl/6 mice that had been injected with rAAV vector harbored slowly decreasing amounts of single-stranded rAAV genomes for at least 4 weeks. By 9 weeks post-transduction, rAAV genomes were no longer detectable as single-stranded DNA but were, instead, found to be integrated into host cell chromatin. In both studies, the integrated vector genomes appeared to be in the form of head-to-tail concatemers. Thomas *et al*.[22] found that rAAV virion DNA (particularly that of AAV serotype 2) persisted within the livers of vector-transduced mice in a DNase-resistant form that could be recovered from cell extracts by immunoprecipitation with a monoclonal antibody specific to intact AAV virions. This suggested that vector uncoating was a rate-limiting step in vector-mediated gene expression and integration. In a corroborative study, Hauck *et al*.[23] labeled recombinant AAV genomes with bromodeoxyuridine (BrdU) and monitored the uncoating of the vector DNA within transduced cells using a BrdU-specific antibody. It was observed that recombinant AAV vectors did not immediately uncoat upon entering the nucleus, but rather persisted within transduced cells in an encapsidated form.

To molecularly characterize rAAV integration events *in vivo*, Nakai *et al*.[24] injected C57Bl/6 mice with recombinant AAV vectors expressing either a human factor IX transgene or a GFP-expressing shuttle vector bearing a bacterial origin of replication and an antibiotic resistance gene, thus allowing rescue of vector–cellular DNA junction sequences in bacteria. At 3 months post-injection, liver DNA was extracted from animals infused with the factor IX-encoding vector and analyzed for rAAV integration events by restriction enzyme digestion and Southern hybridization. It was observed that AAV-specific sequences co-migrated with high-molecular weight genomic DNA. Digestion of the extracted DNA with various restriction enzymes whose recognition sites occur only once within the vector genome revealed that the majority of integrated rAAV genomes occurred as concatemers of tandem vector sequences in a head-to-tail orientation. Although less abundant, head-to-head and tail-to-tail concatemers were also detected. Southern blot analysis indicated that the ITR sequences of the vector genome often sustained deletions upon integration. Eighteen *bona fide* rAAV vector–host chromosomal DNA junctions were recovered from liver DNA extracted from mice that had received the GFP-encoding shuttle vector, thus confirming that some percentage of rAAV sequences were indeed integrated in host cell chromatin. The majority of vector-cellular recombination events occurred within the ITR sequences of the vector and demonstrated no large-scale homology between vector and chromosomal sequences.

To further characterize recombinant AAV chromosomal integration events in live animals, Nakai *et al*.[25] used an *in vivo* hepatocyte selection strategy, which allowed for clonal expansion of rAAV-positive hepatocytes within the livers of hereditary tyrosinemia type 1 (HT1) mice, which are deficient in the activity of the metabolic enzyme fumarylacetoacetate hydrolase (FAH).

In HT1 mice, hepatocytes transduced *in vivo* with an FAH-encoding rAAV vector demonstrate a selective growth advantage over nontransduced cells and can repopulate the liver of affected animals, thereby, diluting episomal rAAV genomes (which have no mechanism for segregation among mitotically active cells) and enriching for integrated rAAV vector sequences within the total liver DNA. The FAH-expressing rAAV vector also contained an antibiotic resistance gene and bacterial origin of replication for recovery of rAAV-cellular junction sequences in bacteria. Fourteen recombinant AAV proviruses with cellular flanking sequences were recovered from the HT1 mice following *in vivo* selection. All 14 of the preintegration sites within the mouse genome displayed deletion of cellular sequences at the site of integration. The majority of the preintegration loci examined (13 of 14) bore minor deletions of less than 300 bp, whereas one clone displayed a deletion of approximately 2-kbp. Six of the integrants demonstrated microhomologies of one-to-four nucleotides occurring at both ends of the integrated vector DNA, whereas the remaining eight integrants demonstrated microhomologies at one end of the integrated DNA. Two of the integrants had small heterologous insertions of two to four basepairs at one end of the structure. In five of the clones, both ITR elements were totally deleted. The remainder of the clones retained one or both ITR elements in a partially deleted form. It was not determined whether any of the integrants were competent for rescue in the presence of the appropriate *trans*-acting factors.

In an expanded study employing the HT1 mouse model and *in vivo* selection of hepatocytes bearing rAAV integrants, Nakai *et al.*[26] injected HT1 mice with 3×10^{11} rAAV serotype 2 particles encoding the FAH gene. Libraries of integrated proviral sequences were constructed from liver DNA isolated from mice that had begun the *in vivo* selection process at either 3 or 6 weeks post-infusion. Of 733 molecular clones analyzed, the researchers were able to identify and unambiguously map 347 unique integration events within the liver DNA of HT1 mice transduced with the FAH-rAAV vector. Sixty percent of the vector-cellular junctions examined occurred within the vector ITR elements, resulting in partial loss of ITR sequences. The observation that the majority of vector breakpoints occurred within the vector ITR elements may reflect the molecular biology of recombinant AAV integration, but may also be biased by the requirement for active expression of the FAH gene during the *in vivo* selection process and the need for an active origin of replication and antibiotic resistance gene during rescue of the integrated vector DNA, as vector sequences bearing breakpoints that disrupt genes essential to the selection strategy would not be recovered. The most common crossover point within the vector ITR occurred at nt 76 of the AAV ITR sequence. A paucity of crossover points occurred within the first seventy nucleotides of the extended ITR sequence. This is consistent with the ITRs being in the hairpin conformation during the process of vector-chromosomal DNA strand exchange. In terms of chromosomal target site selection, fully one-half of the integration events analyzed (53%) occurred within known genes (including both coding and regulatory sequences). Approximately, one third of the integration events identified were within 1-kb of either a transcriptional start site or a CpG island. The authors concluded that rAAV2 sequences demonstrate a preference for insertion within, or in close proximity to, gene regulatory elements. Interestingly, five genes hosted a number of integration events greater than expected by chance. These hotspots of rAAV integration included the genes for upstream binding transcription factor, ubiquitin C, serologically defined colon cancer antigen 33, Riken cDNA 1300002F13 gene and (representing 3% of total criteria-matched integrants) the 45S pre-rRNA gene repeats. The 5′ regulatory elements of the 45S pre-rRNA gene hosted the majority rAAV integration events within the rRNA gene repeats. Approximately, 92% of the preintegration loci displayed evidence of deletion of chromosomal DNA sequences upon rAAV integration. The majority of the deletion events were under 300 bp, although 17% of the deletions were over 1 kbp in size. Two percent of analyzed integration events appeared to be associated with translocation of chromosomal sequences. It was observed that slightly more than 3% of rAAV integration events occurred in or near cancer-related genes as identified in two tumor gene databases, the Retroviral Tagged Cancer Gene Database and the Tumor Gene Database from the Baylor College of Medicine. A majority of the cancer-related genes identified in the study were previously recognized as common integration sites defined by high-throughput, retrovirus-based insertional mutagenesis screens of oncogenesis in mice. The frequency of rAAV integration events at common integration sites , however, was found to be statistically similar to that expected for random integration into the murine genome.

A large-scale analysis of rAAV vector integration sites in primary human fibroblasts was conducted by Miller *et al.*[27] In this study, cultured human lung fibroblasts were transduced with 5×10^4 genome-containing particles per cell of a rAAV shuttle vector capable of recovery in the form of a bacterial plasmid after restriction enzyme digestion and ligation of isolated cellular genomic DNA. Transduced cells were amplified by serial passage without biochemical selection for rAAV sequences. The authors obtained 977 unique integration junctions between rAAV and host cell chromosomal DNA sequences, including both flanking junctions for 323 of the proviral integrants. Recombinant AAV integration events were found to occur in each of the 23 pairs of diploid human chromosomes. Chromosomes 7 and 19 exhibited a statistically significant increase in the number of vector integration events compared to a computer-generated random integration site control set. Chromosome 5, in contrast, demonstrated significantly fewer integration events than expected for random targeting. Based on stringent criteria, rAAV sequences appeared at the site of a chromosomal translocation in at least three instances. Hotspots of AAV vector integration included CpG islands and, in agreement with the results obtained in mice by Nakai *et al.*,[26] ribosomal DNA repeats. In fact, integration within ribosomal DNA repeats accounted for 8% of total vector integration sites identified. In the subset of defined integration events in which chromosomal flanking sequences were identified for a given proviral structure, 70% of the preintegration loci had undergone deletion of host DNA sequences. Most deletions were under 100 bp in length.

An expanding set of tools for gene therapy

An exciting contribution to the arsenal of gene therapy reagents is the development of hybrid viral vectors that utilize the *cis*- and *trans*-acting components of AAV to specifically target transgenes to the AAVS1 integration locus on human chromosome 19. Hybrid vectors based on engineered derivatives of adenovirus and herpes simplex virus can potentially accommodate genetic constructs reaching tens-of-thousands of basepairs in size, and, in addition, can expand the repertoire of cell types accessible to gene therapy approaches seeking targeted genomic integration. In a study of hybrid vector-mediated site-specific integration, Wang and Lieber[28] coinfected Mo7e cells (a human erythroleukemic cell line) with (i) a helper-dependent adenovirus vector encoding the AAV *rep78* gene under the transcriptional control of the globin locus control region, and (ii) a second helper-dependent hybrid adenovirus-AAV vector containing an AAV ITR-flanked, 27-kbp GFP reporter gene construct. AAVS1-specific integration of the reporter cassette was detected in up to 30% of the integration sites analyzed, although with the majority of vector junctions occurring within the Ad, rather than AAV, ITR elements. In another example of hybrid vector technology, Gonçalves *et al*.[29] demonstrated that two copies of a 14-kbp dystrophin cDNA borne by a hybrid adenovirus-AAV vector could be targeted to the AAVS1 locus within human HeLa cells when the AAV Rep proteins were provided in *trans* from a transfected expression plasmid. A demonstration of hybrid vector-mediated site-specific integration *in vivo* was provided by Breakefield and colleagues[30] who used a herpes simplex virus type 1-AAV hybrid amplicon vector to successfully target the integration of an ataxia-telangiectasia mutated-encoding cDNA into the AAVS1 locus of AAVS1-bearing transgenic mice.

Another exciting prospect is the use of AAV components to genetically modify cultured human embryonic stem cells by integration of DNA sequences at a defined locus. Smith *et al*.[31] have used lipid-mediated co-transfection of a *rep*-expressing plasmid and an ITR-flanked GFP reporter gene to introduce reporter gene sequences into the AAVS1 locus of human embryonic stem cells. Site-specific integration within AAVS1 was observed in 4% of the stem cell clones examined. The AAVS1-targeted clones demonstrated a lower incidence of variegated reporter gene expression and decreased levels of gene silencing during prolonged culture when compared to nontargeted controls.

Taken together, the proven utility and exceptional safety profile of AAV-based gene therapy vectors, the potential for targeted integration of very large segments of DNA by AAV-hybrid vectors and the potential union of gene- and cell-based therapies hold great promise for the molecular treatment of human disease.

Acknowledgements

This work was supported by the Intramural Research Program of the National Heart, Lung, and Blood Institute, National Institutes of Health.

References

1 Atchison RW, Casto BC, Hammon WM. Adenovirus-associated defective virus particles. *Science* 1965; **149**: 754–756.
2 Berns KI, Pinkerton TC, Thomas GF, Hoggan MD. Detection of adeno-associated virus (AAV)-specific nucleotide sequences in DNA isolated from latently infected Detroit 6 cells. *Virology* 1975; **68**: 556–560.
3 Cheung AKM, Hoggan MD, Hauswirth WW, Berns KI. Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. *J Virol* 1980; **33**: 739–748.
4 Kotin RM, Berns KI. Organization of adeno-associated virus DNA in latently infected Detroit 6 cells. *Virology* 1989; **170**: 460–467.
5 Kotin RM, Siniscalo M, Samulski RJ, Zhu X, Hunter L, Laughlin CA *et al*. Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci USA* 1990; **87**: 2211–2215.
6 Samulski RJ, Zhu X, Xiao X, Brook JD, Housman DE, Epstein N *et al*. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J* 1991; **10**: 3941–3950.
7 Kotin RM, Linden RM, Berns KI. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *EMBO J* 1992; **11**: 5071–5078.
8 Lamartina S, Sporeno E, Fattori E, Toniatti C. Characteristics of the adeno-associated virus preintegration site in human chromosome 19: open chromatin conformation and transcription-competent environment. *J Virol* 2000; **74**: 7671–7677.
9 Ogata T, Kozuka T, Kanda T. Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. *J Virol* 2003; **77**: 9000–9007.
10 Giraud C, Winocour E, Berns KI. Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. *Proc Natl Acad Sci USA* 1994; **91**: 10039–10043.
11 Weitzman MD, Kyostio SRM, Kotin RM, Owens RA. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc Natl Acad Sci USA* 1994; **91**: 5808–5812.
12 Im D-S, Muzyczka N. The AAV origin binding protein Rep68 ia an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell* 1990; **61**: 447–457.
13 Owens RA, Weitzman MD, Kyostio SR, Carter BJ. Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. *J Virol* 1993; **67**: 997–1005.
14 Linden RM, Winocour E, Berns KI. The recombination signals for adeno-associated virus site-specific integration. *Proc Natl Acad Sci USA* 1996; **93**: 7966–7972.
15 Urcelay E, Ward P, Wiener SM, Safer B, Kotin RM. Asymmetric replication *in vitro* from a human sequence element is dependent on adeno-associated virus Rep protein. *J Virol* 1995; **69**: 2038–2046.
16 Balagué C, Kalla M, Zhang W-W. Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. *J Virol* 1997; **71**: 3299–3306.
17 Surosky RT, Urabe M, Godwin SG, McQuiston SA, Kurtzman GJ, Ozawa K *et al*. Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J Virol* 1997; **71**: 7951–7959.
18 Kearns WG, Afione SA, Fulmer SB, Pang MG, Erikson D, Egan M *et al*. Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. *Gene Ther* 1996; **3**: 748–755.
19 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**: 6969–6976.
20 Miao CH, Snyder RO, Schowalter DB, Patijn GA, Donahue B, Winther B *et al*. The kinetics of rAAV integration in the liver. *Nat Genet* 1998; **19**: 13–15.

21 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**: 306–312.
22 Thomas CE, Storm TA, Huang Z, Kay MA. Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors. *J Virol* 2004; **78**: 3110–3122.
23 Hauck B, Zhao W, High K, Xiao W. Intracellular viral processing, not single-stranded DNA accumulation, is crucial for recombinant adeno-associated virus transduction. *J Virol* 2004; **78**: 13678–13686.
24 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**: 5438–5447.
25 Nakai H, Montini E, Fuess S, Storm TA, Grompe M, Kay MA. AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat Genet* 2003; **34**: 297–302.
26 Nakai H, Wu X, Fuess S, Storm TA, Munroe D, Montini E *et al*. Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. *J Virol* 2005; **79**: 3606–3614.
27 Miller DG, Trobridge GD, Petek LM, Jacobs MA, Kaul R, Russell DW. Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. *J Virol* 2005; **79**: 11434–11442.
28 Wang H, Lieber A. A helper-dependent capsid-modified adenovirus vector expressing adeno-associated virus Rep78 mediates site-specific integration of a 27-kilobase transgene cassette. *J Virol* 2006; **80**: 11699–11709.
29 Gonçalves MAFV, van Nierop GP, Tijssen MR, Lefesvre P, Knaän-Shanzer S, van der Velde I *et al*. Transfer of the full-length dystrophin-coding sequence into muscle cells by a dual high-capacity hybrid viral vector with site-specific integration ability. *J Virol* 2005; **79**: 3146–3162.
30 Cortés ML, Oehmig A, Saydam O, Sanford JD, Perry KF, Fraefel C *et al*. Targeted integration of functional human ATM cDNA into genome mediated by HSV/AAV hybrid amplicon vector. *Mol Ther* 2008; **16**: 81–88.
31 Smith JR, Maguire S, Davis LA, Alexander M, Yang F, Chandran S *et al*. Robust, persistent transgene expression in human ES cells is achieved with AAVS1-targeted integration. *Stem Cells* 2008; **26**: 496–504.