

SHORT REPORT

Open Access

Adeno-associated virus type 2 preferentially integrates single genome copies with defined breakpoints

Tyler Janovitz^{1,2,3}, Michel Sadelain³ and Erik Falck-Pedersen^{2*}

Abstract

Background: Adeno-associated virus (AAV) serotype 2 prevalently infects humans and is the only described eukaryotic virus that integrates site-preferentially. In a recent high throughput study, the genome wide distribution of AAV-2 integrants was determined using Integrant Capture Sequencing (IC-Seq). Additional insight regarding the integration of AAV-2 into human genomic DNA could be gleaned by low-throughput sequencing of complete viral-chromosomal junctions.

Findings: In this study, 140 clones derived from Integrant-Capture Sequencing were sequenced. 100 met sequence inclusion criteria, and of these 39 contained validated junction sequences. These unique sequences were analyzed to investigate the structure and location of viral-chromosomal junctions.

Conclusions: Overall the low-throughput analysis confirmed the genome wide distribution profile gathered through the IC-Seq analysis. We found no unidentifiable sequence inserted at AAV-2 chromosomal junctions. Assessing both left and right ends of the AAV genome, viral breakpoints predominantly occurred in one hairpin of the inverted terminal repeat and AAV genomes were preferentially integrated as single copies.

Keywords: AAV-2, Integration, Virus chromosomal junctions, Viral breakpoints

Findings

Adeno-associated virus, a human Parvovirus in the genus *Dependovirus*, possesses a linear single-strand 4.7 Kb genome [1]. AAV serotype 2 infects up to eighty percent of the human population [2,3] and is the only described eukaryotic virus that integrates site-preferentially [4-6]. The dominant integration hotspot, AAVS1, is located in the first exon of protein phosphatase 1 regulatory subunit 12C (PPP1R12C) [1,7]. Site-preferential integration requires two cell-extrinsic factors: the large AAV replication proteins, Rep68 or Rep78 [8-11], and DNA integration substrates containing Rep binding sites, which are GAGC repeats [12-14].

The genome-wide integration profile of AAV-2 has recently been revealed by a high-throughput sequencing approach coupled with bioinformatics [15]. That study was the first high-throughput analysis of AAV integration

and led to a number of discoveries, including the presence of several thousand novel genomic hotspots. However, paired-end sequencing generates short reads that do not sequence the entirety of viral-chromosomal junctions.

We reasoned that additional insight regarding the integration of AAV-2 into human genomic DNA could be gleaned by low-throughput sequencing of complete viral-chromosomal junctions. In this study, junctions were assayed from wild-type AAV-2 infected HeLa cells processed through the Integrant-Capture Sequencing (IC-Seq) protocol (Figure 1A and B). AAV-2 generated from helper-free plasmid transfection (Applied Viromics) and applied at 1E4 viral genomes per cell. These conditions provide maximal integration efficiency with minimal residual episomal virus, as previously described [15,16]. Since this protocol generates random chromosomal breaks using sonication and does not rely on locus-specific primers, it should be less biased than previous junction studies [15,17,18]. Primer sets to both the left and right portions of the AAV-2 genome were used to assess each biological replicate. The L1/L2 primer set was

* Correspondence: efalckp@med.cornell.edu

²Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10065, USA

Full list of author information is available at the end of the article

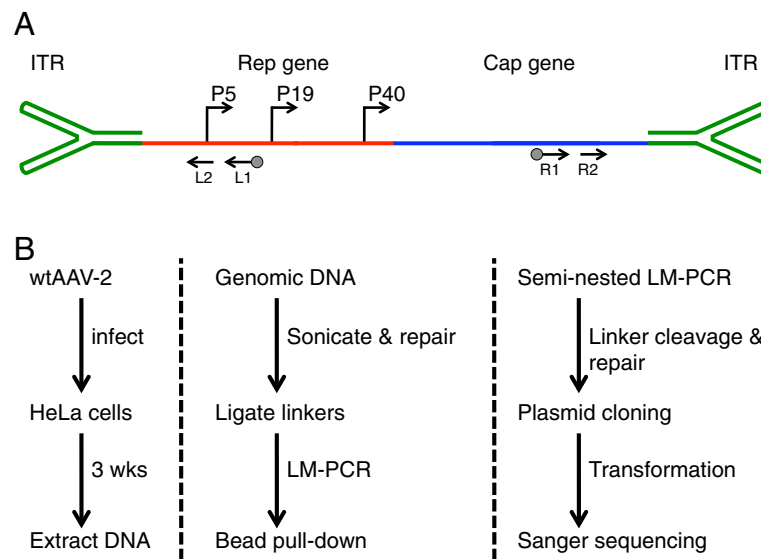


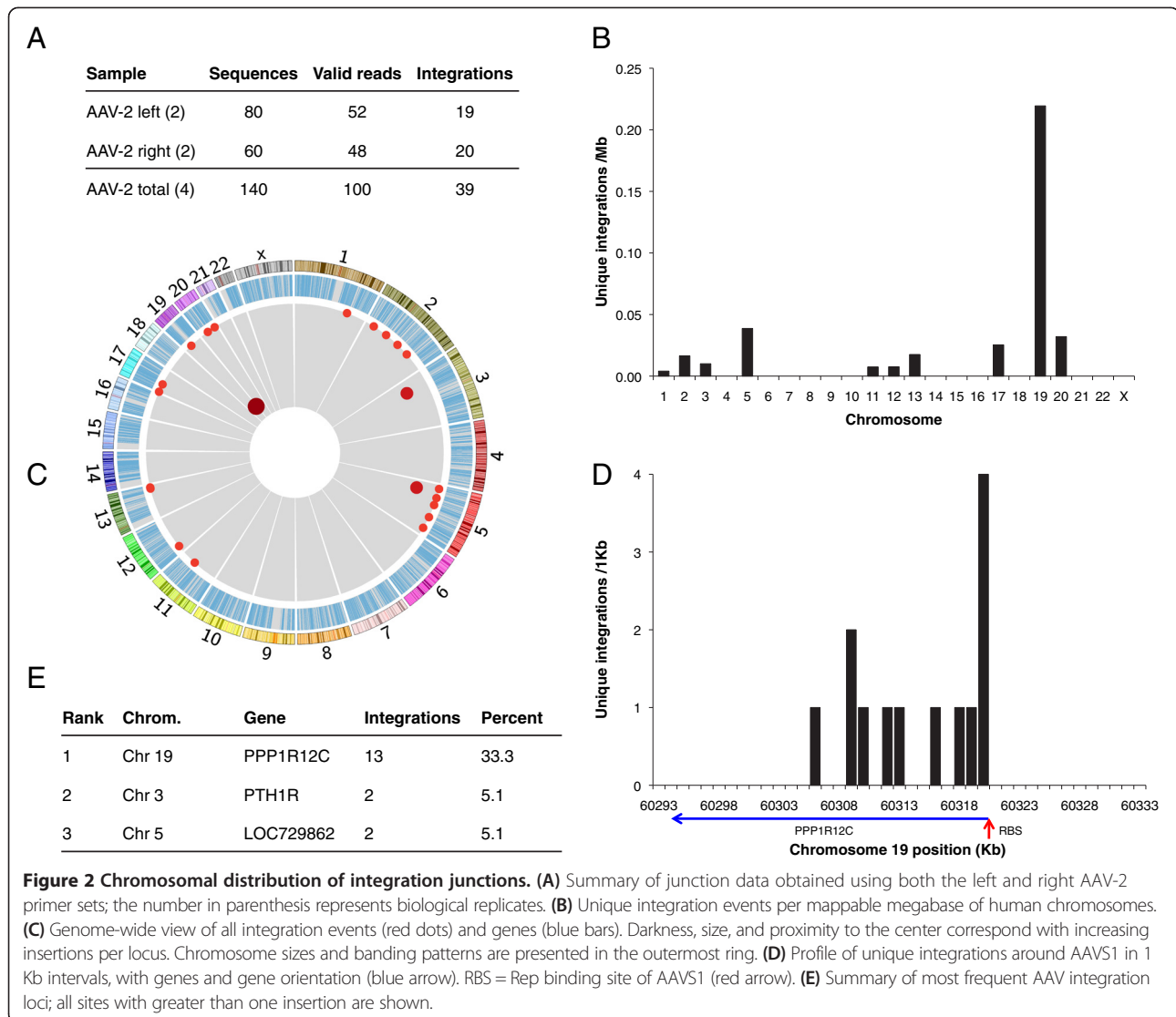
Figure 1 Schematic of AAV genome and experimental design. (A) Overview of AAV genome features (elements of this diagram are not to scale). Inverted terminal repeats (green) cap the ends of the single-strand 4.7 Kb viral genome. Expression of two viral genes, Rep (red) and Cap (blue), is driven by three promoters, P5, P19, P40. L1/L2 and R1/R2 (black arrows) are locations for left and right primer pair binding sites, with biotinylation indicated (grey circle). **(B)** Modified IC-Seq outline for junction analysis. AAV-2 infected HeLa cells were grown for three weeks prior to DNA extraction. Human genomic DNA was sonicated, blunted, A-tailed, and ligated to T-tailed asymmetric linkers. Integration junctions were amplified by semi-nested ligation-mediated PCR, incorporating bead pull-down target enrichment. Libraries were then cloned into bacterial plasmids, transformed, and sequenced.

previously described [15,19] and the R1/R2 oligonucleotides were five-prime modified from a previous study [20] to include [Bio-TEG]C and CGTTT respectively. Methods for IC-Seq are presented in references [15,17,18], and we hope to publish a step-by-step methods guide for future reference. Subsequent to the main phase of IC-Seq, sample pools were cloned into bacterial plasmids, and individual clones were sequenced.

In total, 140 clones were sequenced encompassing two biological replicates assayed using both primer sets (Figure 2A). One hundred of the clones met the inclusion criteria for valid sequences, as previously described [15]. Briefly, these criteria included the presence of correct AAV sequence following the AAV-specific primer and correct linker-tag sequence on the opposing end. These sequence constraints mitigate the possibility of artifactual products [15,17,18]. Of the 100 validated sequences, 39 contained AAV-chromosomal junctions. The remaining sequences were relatively short, representing either uninterrupted viral genome, or viral sequence with a DNA fragment too small to unambiguously assign to the human genome. The 39 confirmed cellular sequences captured averaged 103 base pairs, allowing high-confidence placement in the human genome. Since both the location of the integration junction in the viral inverted terminal repeats and the nature of cellular sequences recovered for the left and right primer sets were extremely similar, they were pooled for further analysis.

Integration junction sequences mapped to ten chromosomes, with chromosome 19 receiving 36% of all events (Figure 2B). Three genomic loci were represented by greater than one unique integrant (Figure 2C and E). AAVS1 was the most frequent site of viral genome insertion, accounting for one-third of all events, while the other two sites, PTH1R (chromosome 3) and LOC729862 (chromosome 5), each represented five percent of detected integrations. These were also the three largest hotspots identified via IC-Seq [15], and two of these hotspots were detected in a previous low-throughput analysis [19]. The thirteen unique integrants identified in AAVS1 begin proximal to the AAV Rep binding site and span the first 15 Kb of PPP1R12C (Figure 2D). This distribution mirrors, on a diminutive scale, the peak-and-tail integration phenotype described in the high-throughput analysis [15].

Examining the viral portion of the recombination junctions revealed additional insights into AAV-2 integration biology. Of the 39 integrations, 92.3% involved contiguous, identifiable viral sequence ligated directly to human chromosomal sequence. The three instances that displayed non-contiguous viral sequence involved viral-viral recombination events in addition to the viral-chromosomal recombination. For the sequences in which contiguous viral regions recombined with chromosomal DNA, 91.6% of viral junctions occurred in the external 120 bp of the inverted terminal repeats (Figure 3A). Within this region, a 19 bp span (position 65-83) accounts

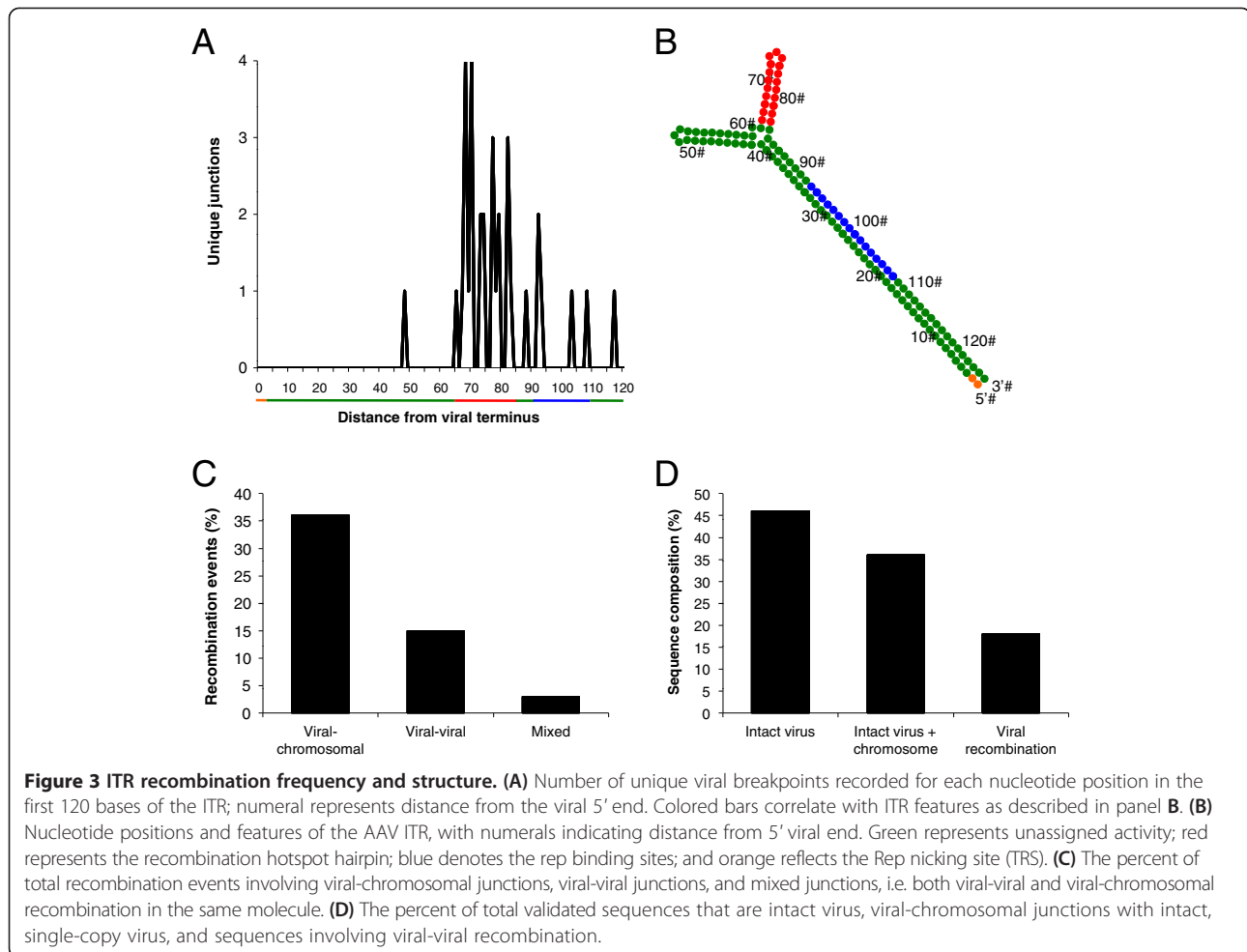


for 69.4% of all viral junctions. This small sequence forms one of the two hairpin loops of the viral ITR and is part of the region that occurs in either a flip or flop orientation [21]. Its position relative to the Rep binding and nicking sites provides insight that may explain the targeting of this region (Figure 3B). After a Rep complex engages the Rep binding sequence (RBS), the terminal resolution site (TRS) is nicked, and the complex proceeds with 3'-5' helicase activity [22-24]. Therefore, the hairpin loop identified as a viral recombination hotspot is the first strong secondary structure encountered by the amplification polymerase complex and may serve to halt progression long enough to facilitate recombination with host DNA via cellular pathways. Additionally, previous work has identified that this internal hairpin loop is specifically bound by AAV Rep during ITR nicking [25]. Thus, the positioning of the

Rep nicking complex may contribute to the creation of the observed recombination hotspot.

Several previous studies, mostly involving AAV vectors, have identified the ITRs as frequent viral recombination points in the absence of Rep [20,21,26,27]. Since the AAV genome is linear and flanked by ITRs, viral-cellular recombination would be expected to occur in this region. Additionally, the complex secondary structure of the ITRs is sufficient to induce a host DNA damage response [28-30]. Based on the data presented in this study, and considering the accumulated insight from previous work [20,21,26-30], the identification of the extreme targeting of one specific ITR hairpin as the primary recombination hotspot is an important observation.

Interestingly, the data provided in this study offer insight into the question of whether wild-type AAV



genomes integrate as single copies or concatamers. Previous work using Southern blotting to characterize integrations from several cell lines suggested that AAV integrates as head-to-tail concatamers [31]. The data analyzed in this study are one hundred unique sequences from a diverse cell population. Of the one hundred sequences that met our inclusion criteria, forty-six were intact viral sequence, thirty-six were direct viral-chromosomal events, fifteen were viral-viral recombinations and three sequences possessed both viral-viral and viral-chromosomal recombination. Therefore, 66.7% of all recombination events captured were between single viral genomes and human chromosomal DNA (Figure 3C). Additionally, we noted that 82% of all sequences were free of viral-viral recombinations (Figure 3D). Thus, analyzing both ends of integrated AAV-2 sequences, the data indicate viral genomes predominantly integrate into host DNA as single copies.

This study of complete viral-chromosomal junctions derived from cloning and sequencing IC-Seq DNA pools provides valuable insight into AAV integration.

The structurally complex, repetitive, and GC-rich nature of these sequences may hinder capture of the entire junction-population. We have taken many steps to mitigate these effects. These steps included using: short sequences from random breaks, two primer sets, stringent sequence validation, robust polymerases, and high melting temperatures. Therefore, we believe that the junctions captured and analyzed in this study are not unduly influenced by sequence constraints, and present a valuable representation of the AAV-2 junction population. The insertion profile of AAV-2 maintained the same top three hotspots found using high-throughput technology and the distribution around AAVS1, the largest hotspot, was also quite similar. In the absence of Rep, the unique AAV-2 ITR structure is a target for cellular DNA repair and recombination pathways which can vary in a cell dependent manner [21,30,32,33]. In the case of wild-type AAV-2, Rep binding to the RBE as well as the hairpin stem influences helicase activity [25]. Therefore, Rep, in concert with cellular DNA repair complexes, may contribute to formation of the internal stem-loop ITR

recombination hotspot identified in this study. We anticipate that cell-specific differences in DNA repair proteins and Rep interacting proteins may also influence the integration profile to some extent. However, direct Rep-DNA interactions appear to play the dominant role in defining the genome-wide targets for AAV-2 integration [15,19]. Finally, based on the population of junctions captured, AAV-2 genomes were found to predominately integrate as single genome copies, and viral-viral recombination was modest. This study may impact Rep-mediated gene therapy approaches and highlights how long read length, even on a modest scale, may serve to significantly augment the understanding of high-throughput data sets.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TJ designed and performed experiments and analysis and wrote the manuscript. MS provided material assistance and made suggestions on the manuscript. EFP designed experiments and analysis and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

T.J. was supported by a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the National Institutes of Health under award number T32GM07739 to the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD-PhD Program. E.F.P. received support from the WR Hearst Foundation and PHS grant RO1 AI094050. The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author details

¹Tri-Institutional MD-PhD Program, Weill Medical College of Cornell University, New York, NY 10065, USA. ²Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10065, USA. ³Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA.

Received: 1 November 2013 Accepted: 22 January 2014

Published: 27 January 2014

References

- Linden RM, Ward P, Giraud C, Winocour E, Berns KI: **Site-specific integration by adeno-associated virus.** *Proc Natl Acad Sci U S A* 1996, **93**:11288–11294.
- Halbert CL, Miller AD, McNamara S, Emerson J, Gibson RL, Ramsey B, Aitken ML: **Prevalence of neutralizing antibodies against adeno-associated virus (AAV) types 2, 5, and 6 in cystic fibrosis and normal populations: implications for gene therapy using AAV vectors.** *Hum Gene Ther* 2006, **17**:440–447.
- Calcedo R, Vandenbergh LH, Gao G, Lin J, Wilson JM: **Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses.** *J Infect Dis* 2009, **199**:381–390.
- Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, McLaughlin S, Muzyczka N, Rocchi M, Berns KI: **Site-specific integration by adeno-associated virus.** *Proc Natl Acad Sci U S A* 1990, **87**:2211–2215.
- 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.
- Samulski RJ, Zhu X, Xiao X, Brook JD, Housman DE, Epstein N, Hunter LA: **Targeted integration of adeno-associated virus (AAV) into human chromosome 19.** *EMBO J* 1991, **10**:3941–3950.
- Linden RM, Winocour E, Berns KI: **The recombination signals for adeno-associated virus site-specific integration.** *Proc Natl Acad Sci U S A* 1996, **93**:7966–7972.
- Urceley 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.
- Surosky RT, Urabe M, Godwin SG, McQuiston SA, Kurtzman GJ, Ozawa K, Natsoulis G: **Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome.** *J Virol* 1997, **71**:7951–7959.
- Urabe M, Kogure K, Kume A, Sato Y, Tobita K, Ozawa K: **Positive and negative effects of adeno-associated virus Rep on AAVS1-targeted integration.** *J Gen Virol* 2003, **84**:2127–2132.
- Young SM, McCarty DM, Degtyareva N, Samulski RJ: **Roles of adeno-associated virus Rep protein and human chromosome 19 in site-specific recombination.** *J Virol* 2000, **74**:3953–3966.
- Young SM, Samulski RJ: **Adeno-associated virus (AAV) site-specific recombination does not require a Rep-dependent origin of replication within the AAV terminal repeat.** *Proc Natl Acad Sci U S A* 2001, **98**:13525.
- Pieron L, Fipaldini C, Monciotti A, Cimini D, Sgura A, Fattori E, Epifano O, Cortese R, Palombo F, La Monica N: **Targeted integration of adeno-associated virus-derived plasmids in transfected human cells.** *Virology* 1998, **249**:249–259.
- 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**:5411–5421.
- Janovitz T, Klein IA, Oliveira T, Mukherjee P, Nussenzweig MC, Sadelain M, Falck-Pedersen E: **High-throughput sequencing reveals principles of adeno-associated virus serotype 2 integration.** *J Virol* 2013, **87**:8559–8568.
- Hamilton H, Gomos J, Berns KI, Falck-Pedersen E: **Adeno-associated virus site-specific integration and AAVS1 disruption.** *J Virol* 2004, **78**:7874–7882.
- Klein IA, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, Di Virgilio M, Bothmer A, Nussenzweig A, Robbiani DF, Casellas R, Nussenzweig MC: **Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes.** *Cell* 2011, **147**:95–106.
- Oliveira TY, Resch W, Jankovic M, Casellas R, Nussenzweig MC, Klein IA: **Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements.** *J Immunol Methods* 2012, **375**:176–181.
- Hüser D, Gogol-Döring A, Lutter T, Weger S, Winter K, Hammer E-M, Cathomen T, Reinert K, Heilbronn R: **Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome.** *PLoS Pathog* 2010, **6**:e1000985.
- Drew HR, Lockett LJ, Both GW: **Increased complexity of wild-type adeno-associated virus-chromosomal junctions as determined by analysis of unselected cellular genomes.** *J Gen Virol* 2007, **88**:1722–1732.
- Yang CC, Xiao X, Zhu X, Ansardi DC, Epstein ND, Frey MR, Matera AG, Samulski RJ: **Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro.** *J Virol* 1997, **71**:9231–9247.
- Im DS, Muzyczka N: **The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity.** *Cell* 1990, **61**:447–457.
- Wu JJ, Davis MDM, Owens RAR: **Factors affecting the terminal resolution site endonuclease, helicase, and ATPase activities of adeno-associated virus type 2 Rep proteins.** *J Virol* 1999, **73**:8235–8244.
- Zhou X, Zolotukhin I, Im DS, Muzyczka N: **Biochemical characterization of adeno-associated virus rep68 DNA helicase and ATPase activities.** *J Virol* 1999, **73**:1580–1590.
- Brister JR, Muzyczka N: **Mechanism of rep-mediated adeno-associated virus origin nicking.** *J Virol* 2000, **74**:7762–7771.
- Miller DG, Petek LM, Russell DW: **Adeno-associated virus vectors integrate at chromosome breakage sites.** *Nat Genet* 2004, **36**:767–773.
- 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.
- Schwartz RA, Carson CT, Schubert C, Weitzman MD: **Adeno-associated virus replication induces a DNA damage response coordinated by DNA-dependent protein kinase.** *J Virol* 2009, **83**:6269–6278.
- Raj K, Ogston P, Beard P: **Virus-mediated killing of cells that lack p53 activity.** *Nature* 2001, **412**:914–917.
- Cataldi MP, McCarty DM: **Hairpin-end conformation of adeno-associated virus genome determines interactions with DNA-repair pathways.** *Gene Ther* 2013, **20**:686–693.
- Cheung A, Hoggan M, Hauswirth W, Berns KI: **Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells.** *J Virol* 1980, **33**:739–4832.

32. Daya S, Cortez N, Berns KI: **Adeno-associated virus site-specific integration is mediated by proteins of the nonhomologous end-joining pathway.** *J Virol* 2009, **83**:11655–11664.
33. Inagaki K, Ma C, Storm TA, Kay MA, Nakai H: **The role of DNA-PKcs and artemis in opening viral DNA hairpin termini in various tissues in mice.** *J Virol* 2007, **81**:11304–11321.

doi:10.1186/1743-422X-11-15

Cite this article as: Janovitz *et al.*: Adeno-associated virus type 2 preferentially integrates single genome copies with defined breakpoints. *Virology Journal* 2014 11:15.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

