Comprehensive Analysis of the Adeno-Associated Virus (AAV) Transcriptome Based on Next Generation Sequencing Reveals New Insights into AAV Biology

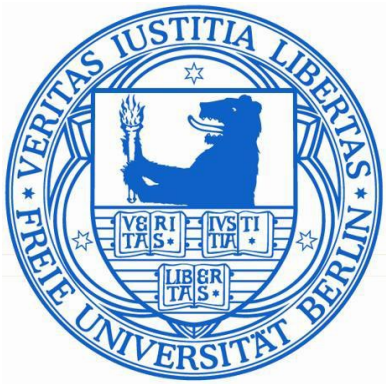
Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

Submitted to the Department of Biology, Chemistry and Pharmacy

of the Freie Universität Berlin



by

Catrin Stutika

from Freiberg, Germany

September 2016

The present doctoral thesis was conducted in the period of October 2012 to September 2016 at the Institute of Virology of the Charité Universitätsmedizin Berlin under the supervision of Prof. Dr. Regine Heilbronn.

1st Reviewer:	Prof. Dr. Regine Heilbronn Charité - Universitätsmedizin Berlin Institut für Virologie Campus Benjamin Franklin Hindenburgdamm 27 12203 Berlin, Germany
2nd Reviewer:	Prof. Dr. Rupert Mutzel FB Biologie, Chemie, Pharmazie Institut für Biologie - Mikrobiologie Freie Universität Berlin Königin-Luise-Straße 12-16 14195 Berlin, Germany
Date of thesis defense:	18.11.2016

Acknowledgement

First and foremost, I want to thank my mentor Regine Heilbronn for giving me the opportunity to work on my doctoral thesis at the Institute of Virology of the Charité Universitätsmedizin Berlin. I am grateful that she gave me an exciting topic on AAV biology with extensive research potential. I am thankful for her supervision and support on my thesis with countless discussions leading to the design and progress of the study. Furthermore, I want to thank her for helping me with applications for scholarships and for giving me the chance to participate at the international parvovirus workshop in 2014. In this regard, I also want to thank the Charité Universitätsmedizin Berlin for providing me a 12-month Ph.D. fellowship to finish my doctoral research study.

Secondly, I especially thank Andreas Gogol-Döring for the several bioinformatic analyses of the sequencing data. I further thank him for helping me on interpreting the data and discussions that considerably promoted this study. I further thank Wei Chen for his offer to use the Illumina NGS platform and all members of his lab who were involved in preparing the sequencing process.

I especially want to thank Stefan Weger and Daniela Hüser of the Heilbronn lab for their scientific input and advice on my study progress. Also, I deeply thank Mario Mietzsch for the endless scientific discussions and all his support. I really enjoyed working with you. Furthermore, I want to thank Melanie Heßler for experimentally supporting me and all my current and former colleagues, namely Kristina von Kietzell, Natalja Rutz, Dina Khalid, Eva-Maria Hammer, Kerstin Winter, Eva Guhl and Ruth Joncker. It was a pleasure to work with all of you.

Last but not least I want to thank my family and friends for always supporting me during the entire period of my doctoral thesis with all ups and downs.

Foreword

This cumulative doctoral thesis is based on the following four peer-reviewed original publications:

Catrin Stutika, Andreas Gogol-Döring, Laura Botschen, Mario Mietzsch, Stefan Weger, Mirjam Feldkamp, Wei Chen and Regine Heilbronn. 2016. A Comprehensive RNA Sequencing Analysis of the Adeno-Associated Virus (AAV) Type 2 Transcriptome Reveals Novel AAV Transcripts, Splice Variants, and Derived Proteins.
Journal of Virology 90 (3): 1278-1289.

Catrin Stutika, Mario Mietzsch, Andreas Gogol-Döring, Stefan Weger, Madlen Sohn, Wei Chen and Regine Heilbronn. 2016. Comprehensive Small RNA-Seq of Adeno-Associated Virus (AAV)-Infected Human Cells Detects Patterns of Novel, Non-Coding AAV RNAs in the Absence of Cellular miRNA Regulation.
PLoS One 11 (9): e0161454.

Stefan Weger, Eva-Maria Hammer, Melanie Gonsior, **Catrin Stutika**, Regine Heilbronn. 2016. A Regulatory Element Near the 3' End of the Adeno-Associated Virus rep Gene Inhibits Adenovirus Replication in cis by Means of p40 Promoter-Associated Short Transcripts.
Journal of Virology 90 (8): 3981-3993.

Catrin Stutika, Daniela Hüser, Stefan Weger, Natalja Rutz, Melanie Heßler and Regine Heilbronn. 2015. Definition of Herpes Simplex Virus Helper Functions for the Replication of Adeno-Associated Virus Type 5.
Journal of General Virology 96 (Pt4): 840-850.

Directory

1 Introduction

1.1 Adeno-Associated Virus (AAV)

Adeno-associated viruses (AAVs) are members of the parvovirus family with small non-enveloped icosahedral capsids of about 20 to 25 nm in diameter. They were first described in 1965 as contaminants of an adenovirus preparation and henceforward named adeno-associated viruses (1). Among the parvoviruses the AAVs belong to the genus *Dependovirus*. As the name implies they depend on the presence of a suitable helper virus, like adenovirus (2) or herpesviruses (3, 4) for productive replication. In the absence of a helper virus AAV establish a latent state and persist in the nucleus as episome (5) or integrated within the host cell genome (6). Upon co-infection or superinfection with a helper virus AAV will be reactivated and switch to productive replication (7).

For the AAVs a large number of serotypes and variants are described which especially differ in their viral capsid structures. Since the capsids mediate host cell entry they determine the cellular tropism of the different AAV serotypes. Among these, AAV2 is the best-characterized serotype and serves as the prototype for the AAV family. AAV5 represents the evolutionary most distant member of the AAV serotypes with some unique features. The AAVs are widely distributed among the human population. But despite the high seroprevalence of approximately 80%, AAV infections have not been associated with any diseases (8). Based on wild-type AAV recombinant vectors have been developed for gene therapy.

1.1.1 AAV Genome Organization and Protein Expression

Adeno-associated viruses possess a linear single-stranded DNA genome with a length of approximately 4.7 kilobases (kb) (9). Both ends of the genome contain identical inverted terminal repeats (ITRs) of about 145 nucleotides (10). Due to embedded complementary repeat sequences, the ITRs can fold back to form a T-shaped secondary structure (Fig. 1). This structure is composed of three palindromic regions, A/A', B/B' and C/C', and a single-stranded unique region of about 20 nucleotides, designated as D-region. The AAV-ITRs contain *cis*-regulatory elements required for AAV gene regulation, DNA replication and genome packaging (11, 12), as well as for the integration of the virus into the host genome (5). The Rep binding element (RBE) within the ITRs enables binding of the large

regulatory proteins Rep78 and Rep68 to the AAV genome (13). Furthermore, the terminal resolution site (*trs*) allows strand-specific nicking of the AAV genome by Rep78/68 during DNA replication (14).

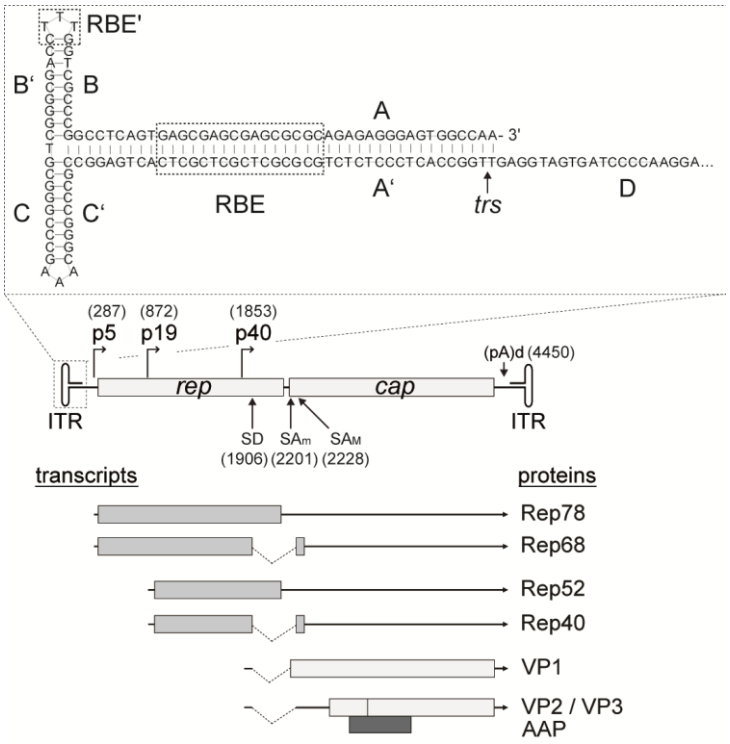


Fig. 1: AAV2 genome organization. Displayed in the center is the AAV2 genome with the two open reading frames *rep* and *cap* flanked by the inverted terminal repeats (ITRs). Indicated are the three promoters, the splice donor (SD) and the minor and major splice acceptor sites ($SA_{m/M}$), and the polyadenylation site (pA), each with corresponding nucleotide positions. Above the genome the ITR secondary structure is enlarged showing the different repeats, the Rep binding element (RBE) and the terminal resolution site (*trs*). Below spliced and unspliced transcripts initiated from the p5, p19 and p40 promoter and resulting proteins are depicted. Different shading patterns imply different reading frames.

The ITRs flank two large open reading frames (ORFs), *rep* and *cap*, which encode overlapping regulatory proteins (Rep) and capsid proteins (Cap), respectively. Furthermore, within the *cap* gene the assembly activating protein (AAP) is encoded (Fig. 1). All described AAV mRNAs are transcribed from one DNA strand, designated as AAV (+) strand. For AAV2 these mRNAs are initiated from the three promoters p5, p19 and p40, named after their relative positions within the genome. Only a single polyadenylation signal at map position p96 is used for transcription termination (15, 16). Furthermore, AAV contains a single intron located in the center of the genome (17). Splicing of AAV2 transcripts occur at the splice donor site at nucleotide position 1906 and either the minor or major splice acceptor site at nucleotide position 2201 or 2228, respectively. Transcripts of the *rep* ORF are initiated at the p5 and the p19 promoter leading to four non-structural Rep proteins (Fig. 1) (18). The large Rep proteins Rep78 and Rep68 are derived from p5-initiated transcripts. The small Rep proteins Rep52 and Rep40 represent N-terminally truncated variants of the large Rep proteins, which are derived from p19-initiated

transcripts. While Rep78 and Rep52 are translated from unspliced mRNAs, Rep68 and Rep40 are translated from spliced mRNAs leading to C-terminally truncated versions of Rep78 and Rep52, respectively. All four Rep proteins possess ATPase activity and a 3'-to-5' helicase activity (19). Only the large Rep proteins exhibit site-specific DNA-binding- and endonuclease activity (20, 21) since the respective protein domains are located N-terminally. The large Rep proteins are involved in various stages of gene expression, genome replication and genomic integration, whereas the small Rep proteins are required for genome packaging (12). The *cap* ORF encodes the structural proteins VP1, VP2 and VP3 that form the viral capsid (Fig. 1). The corresponding transcripts are initiated at the p40 promoter and are differentially spliced (22). The utilization of the AAV2 minor or major splice acceptor sites generate two alternative transcripts (23, 24). In the splice process the major splice acceptor site is used more frequently than the minor splice acceptor site. VP1 can only be translated from the transcript spliced at the minor splice acceptor site and is therefore expressed at a lower level. The major transcript is responsible for the expression of VP2 and VP3. However, since VP2 translation is initiated at a non-canonical and weak ACG start codon its expression level is as low as that of VP1. In contrast, VP3 translation initiation starts at a conventional ATG start codon, resulting in high expression levels. As a consequence VP1, VP2 and VP3 are expressed at ratios of approximately 1:1:10 (23). VP1 and VP2 represent N-terminal extended variants of VP3. The VP1 unique region harbors a phospholipase A2 domain essential for viral infectivity (25). Within the *cap* ORF an alternative reading frame encodes the AAP protein, which is translated from an unconventional CTG start codon (Fig. 1). AAP is required for assembly and maturation of the viral capsids in the nucleolus (26).

1.1.2 AAV Serotypes: Comparison of the AAV2 and AAV5 Genomes

The AAV family comprises a growing number of different variants and virus strains. Up to now 13 AAV serotypes have been described isolated from human and nonhuman primates. AAV2 was the first serotype to be cloned into a plasmid construct (27) and was henceforth used for most molecular studies. Therefore, AAV2 represents the best-characterized serotype and serves as the prototype for most AAV serotypes.

AAV5, the evolutionary most distant member compared to AAV2 shows a sequence homology of only 50 to 60%, at the nucleotide and amino acid level (28). The overall genome organization of AAV5 is comparable to that of AAV2 (Fig. 2). The main

difference between the AAV2 and AAV5 transcription profiles is that RNAs generated from the AAV5 promoters p7 and p19 predominantly use an internal polyadenylation site (pA)p (nt 2193) located within the central intron of the genome (Fig. 2) (29). The AAV5 intron bears two consensus polyadenylation signals (AAUAAA) (17). RNA cleavage and polyadenylation occurs 11 to 14 nt downstream of the first poly(A) motif, upstream of the minor splice acceptor site SA_m (nt 2204). Utilization of the internal polyadenylation signal prevents splicing of p7- and p19-generated transcripts. As a result no spliced Rep isoforms are generated (29). The AAV2 genome also exhibits the consensus polyadenylation signal within the intron, but in contrast to AAV5 it is not utilized (17). In addition to Rep78 and Rep52, high levels of a Rep40-like protein have been observed during AAV5 infection (Fig. 2) (30). Unlike Rep40 of AAV2, the Rep40-like protein of AAV5 is generated from an alternative start codon 50 amino acids downstream of the Rep52 start codon. As a consequence, AAV5 Rep52 and Rep40-like have the same C-terminus, but differ in their N-termini. Similar to the AAV2 Rep40 protein, the Rep40-like protein exhibits helicase activity and was shown to be functional for viral encapsidation (31). AAV5 transcripts initiated at the p41 promoter use the distal polyadenylation site (pA)d (29). Furthermore, the analysis of the AAV5 transcription profile revealed an abundant transcript generated from the ITRs in HEK 293 cells achieving expression levels comparable to the p7 promoter (29). This transcript initiates at nt 142 (Inr) which maps to the terminal resolution site (*trs*) within the AAV5-ITR (Fig. 2). The ITR transcript is polyadenylated at the distal polyadenylation site (pA)d and appears not to be spliced. Its potential role for the AAV5 infection cycle is currently unknown. A similar ITR-initiated transcript has been shown for AAV2 in AAV2-derived vectors (32, 33). But in contrast to AAV5, transcripts initiated at the AAV2-ITRs are only expressed at very low levels.

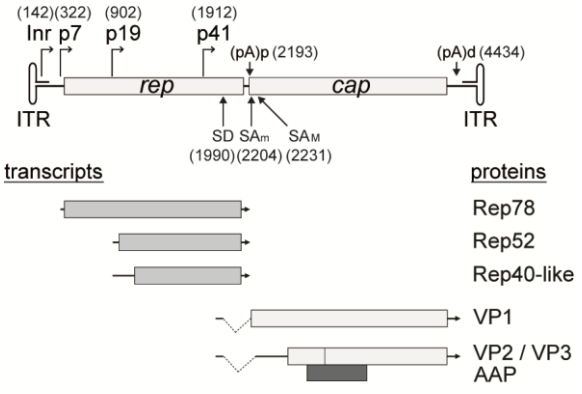


Fig. 2: AAV5 genome organization. The AAV5 genome is depicted indicating the three promoters, the splice donor (SD) and the minor and major splice acceptor sites ($SA_{m/M}$), and the proximal and distal polyadenylation sites (pA)p/d, each with corresponding nucleotide positions. The ITR transcript with its starting nucleotide is indicated. Below the genome transcripts initiated from the p7, p19 and p41 promoter and resulting proteins are displayed. Different shading patterns imply different reading frames.

1.1.3 Biphasic Life Cycle of AAV

Despite the wide distribution of AAVs in the human population, the *in vivo* mechanisms of AAV replication in humans and primates are largely unknown. Since efficient AAV replication requires the presence of a helper virus and no *in vivo* model for wild-type AAV infection exists, most of the knowledge regarding AAV biology is based on experiments performed in cell culture.

1.1.3.1 Host Cell Entry

In the first step of the replication cycle AAV enters the host cell by receptor-mediated endocytosis. This is achieved by binding of the AAV capsids to a specific primary receptor and to co-receptors on the cellular surface. The AAV serotypes use diverse receptors, which are differentially expressed on the various host cells. For AAV2 heparan sulfate proteoglycan (HSPG) is described as primary receptor (34), while a couple of co-receptors, like integrins, the laminin receptor, the fibroblast growth factor receptor 1 (FGFR-1), the hepatocyte growth factor receptor (HGFR) or the tetraspanin CD9 (35-39) are utilized for viral internalization. The evolutionary most divergent member of the AAV family, AAV5, binds to N-linked α2-3 or α2-6 sialic acids on the cellular surface (40). Furthermore, the platelet-derived growth factor receptor (PDGFR) is described to be required for AAV5 infection (41). Recently, a transmembrane protein has been identified that serves as a universal receptor for all members of the AAV family, designated as AAVR (42). However, its exact role in the transduction process is still debated (43, 44).

The AAV capsids can be internalized via the dynamin-dependent pathway by clathrin-coated pits (45) or by a clathrin-independent carrier (CLIC) through a GPI-anchored-protein-enriched endosomal compartment (GEEC) (46). After endocytosis the AAV virions are situated in endosomes. Acidification causes structural rearrangements of the capsids leading to exposure of the phospholipase A2 domain of the viral VP1 proteins which enables the escape from the endosomes (47). Rapid unidirectional cytoplasmic trafficking towards the nucleus is facilitated by the use of the microtubule network (48). The released capsids are assumed to pass the nuclear pores intact, due to their small size. Within the nucleus the capsids are uncoated and the genomes are set free. The complementary strand of the single-stranded AAV genomes are synthesized by cellular DNA polymerases allowing transcription of the viral mRNAs. Depending on the absence

or presence of a helper virus, AAV enters the latent or productive replication cycle, respectively.

1.1.3.2 Latent Replication Cycle

In the absence of a suitable helper virus only low levels of Rep78/68 are expressed. These are sufficient to suppress any further AAV gene transcription by a negative feedback mechanism on the AAV promoters (49, 50), thereby inhibiting viral replication. In the following, the AAV genomes either integrate into the cellular host genome (6) or persist in the nucleus as episomes (5). Traditionally it was shown that AAV integrates at a site called AAVS1 located on human chromosome 19 (51). However, recent studies revealed additional AAV integration sites throughout the human genome near GAGY/C repeats (52). The Rep proteins are believed to specifically bind to GAGY/C repeat sequences of the viral and the host cell genome (53), thereby mediating AAV strand invasion. Usually, viruses that exhibit a latent infection cycle persist in specific cell types until reactivation of productive virus replication. The *in vivo* sites of AAV persistence have not been identified. But recent studies on healthy and immunocompromised human blood donors revealed various AAV serotypes located within human T-lymphocytes, indicating that these cells might represent the *in vivo* sites of AAV persistence (Khalid, Hüser and Heilbronn, pers. communication).

1.1.3.3 Productive Replication Cycle

In the presence of a suitable helper virus AAV switch from latent to productive replication. Initially, specific viral helper genes transactivate the AAV p5 promoter (54, 55). Furthermore, the large Rep proteins transactivate all AAV promoters to support efficient gene expression (56, 57). In addition, the Rep proteins are essential for the replication of the AAV genome.

Starting from the single-stranded AAV genome the free 3'-OH end at the ITR serves as a primer for DNA synthesis by cellular DNA polymerases generating the AAV monomer turnaround (mT) form (Fig. 3) (11). The large Rep proteins are able to specifically bind to the RBE within the AAV-ITRs. After DNA binding the endonuclease domain of Rep78/68 can introduce a site-specific single-stranded nick at the *trs* creating a new free 3'-OH end for DNA synthesis (58). Due to the helicase activity of the Rep proteins DNA unwinding is achieved by the generation of relaxed DNA strands. These are accessible for the

recruitment of cellular or viral DNA polymerases and further helper functions required for efficient genome replication. The complementary sequence of the ITR is elongated by strand displacement leading to the AAV monomer extended (mE) replication form (Fig. 3). At this stage AAV strands of both polarities, positive and negative, are synthesized. Self-annealing of the ITRs at either end of the genome results in new free 3'-OH ends, that can be used for reinitiation of DNA synthesis or alternatively these replication products can be used for genome packaging. Self-annealing at the elongated ITR of the AAV (mT) replication form also leads to a new free 3'-OH end on which DNA polymerases can reinitiate for DNA synthesis. If this occurs before strand-specific nicking at the *trs*, dimeric (dT) or higher AAV replication forms can be generated (Fig. 3).

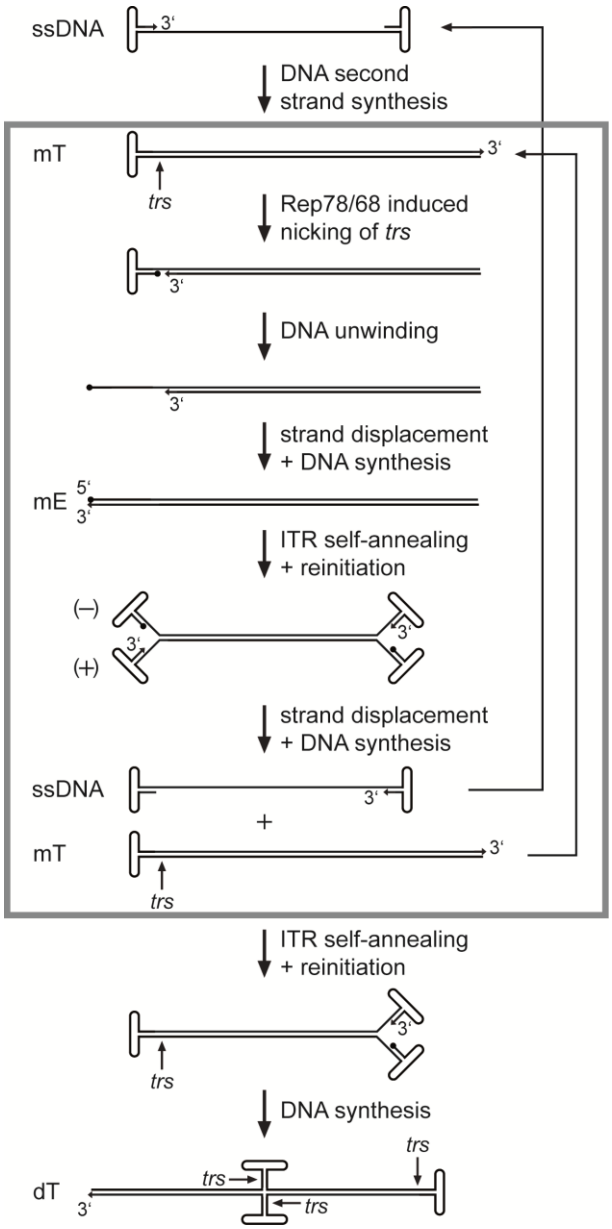


Fig. 3: AAV genome replication. Model of the AAV genome replication, starting with the single-stranded DNA (ssDNA) genome, which results in the generation of multiple AAV genomic copies of positive and negative polarity. mT = monomer turnaround, mE = monomer extended, dT = dimer turnaround, *trs* = terminal resolution site.

Late in the AAV replication cycle, the structural proteins VP1, VP2 and VP3 assemble to preformed viral capsids, supported by AAP (26). AAV ssDNA genomes of positive and negative polarity are equally encapsidated, which is mediated by the helicase activity of the small Rep proteins, Rep52 and Rep40 (12). Newly generated AAV progenies are released from the host cell as a consequence of helper virus induced cell lysis.

1.1.4 Viral Helper Functions

The best-characterized helper viruses enabling productive AAV replication represent adenovirus (Ad2 or Ad5) or herpes simplex virus type 1 (HSV1). These helper viruses differ in their molecular mechanisms by which they support productive AAV replication. While adenovirus mainly promotes AAV gene expression and leads the cell into the S phase, the HSV1 helper genes support AAV genome replication.

It has been shown that five early adenovirus gene products, namely E1A, E1B, E2A, E4orf6 and VA RNA, are required for productive AAV2 replication (59). The early gene products E1A and E1B play an essential role during AAV gene expression (60). E1A has been shown to act as a transcription activator required for initial activation of the AAV promoters (54, 61). E2A represents an ssDNA-binding protein that interacts with Rep and enhances AAV genome replication (62). E4orf6 has been shown to stimulate secondary strand synthesis of the AAV genome, which is the rate-limiting step during early AAV replication (63). Furthermore, the E1B55K/E4orf6 complex acts as an ubiquitin ligase, which is described to promote viral replication by degrading cellular proteins important for apoptosis or DNA repair (64, 65). The VA RNAs stimulate AAV protein synthesis by blocking the activation of the cellular interferon-induced antiviral defense mechanism (66, 67). Similar to AAV2, the same adenoviral gene products are sufficient for productive AAV5 replication (68). But in contrast to AAV2, AAV5 is less dependent on adenoviral helper genes for RNA expression and RNA splicing (17).

In contrast to the identified adenoviral gene products, the helper functions provided by herpes simplex virus type 1 (HSV1) are directly involved in AAV genome replication. A minimal set comprising the ternary HSV1 helicase-primase complex (UL5/UL8/UL52) and the ssDNA-binding protein ICP8 (UL29) is essential and sufficient for productive AAV2 replication (69). In this process, ICP8 has been shown to co-localize with Rep78 on AAV ssDNA genomes *in vitro* and within nuclear replication centers *in vivo* resulting in ternary complex formation required for AAV genome replication initiation (70). Furthermore, it

has been shown that UL5 helicase activity, but not UL52 primase catalytic activity was required for efficient AAV genome replication (71). In addition, the HSV1 DNA polymerase (UL30/UL42) enhances AAV2 genome replication. Furthermore, the HSV1 gene product ICP0 has been shown to transactivate AAV2 *rep* gene expression (55). This effect is further increased by the HSV1 gene functions ICP4 and ICP22, which act synergistically with ICP0 (72). More recently, another gene product, the HSV1 exonuclease (UL12), has been shown to facilitate resolution of high molecular AAV replication forms (73).

Many other viruses are described to support productive AAV replication. These include members of the herpesvirus family, e.g. herpes simplex virus type 2 (HSV2) (3), human cytomegalovirus (HCMV) (4), varicella zoster virus (VZV) (74), Epstein-Barr virus (EBV), human herpesvirus 6 (HHV6) (75) and pseudorabies virus (PrV) (76), as well as members of other virus families such as human papillomavirus (HPV) (77) or baculovirus, an insect virus (78). However, for these viruses the individual helper genes that promote productive AAV replication are less well characterized. In contrast to previous studies, the vaccinia virus (VV) was recently shown to be insufficient to completely promote productive AAV replication and was renominated as subhelper of AAV due to its lack to transactivate AAV promoters (79).

1.1.5 AAV-Derived Vectors and Their Applications

In recent years, AAV-derived vectors have become one of the most favorable candidates for virus-based gene therapeutic applications. Vectors based on AAV have been successfully used in clinical trials for the treatment of monogenetic disorders such as hemophilia B and show stable long-term persistence and transgene expression in post-mitotic cells (80, 81). Moreover, Glybera, an AAV vector for the treatment of the rare genetic disease, lipoprotein lipase deficiency was approved in 2012 by the European Medicines Agency (EMA) as the first virus-based gene therapeutic pharmaceutical (82). In these and many other clinical trials recombinant AAV (rAAV) vectors have been shown to be safe, not exceeding mild immune responses of the host (83).

1.1.5.1 Features of AAV-Derived Vectors

In order to achieve a specific tissue tropism, AAV vectors can be packaged into capsids of the AAV serotype of choice (84). In contrast to the AAV wild-type genome, rAAV vectors

do not contain any viral coding sequences except for the *cis*-active AAV-ITRs which flank the transgene. In absence of the AAV Rep proteins, Rep-mediated integration into the host genome does not occur. However, very rare random integration events have been described (52, 85). The majority of AAV vector genomes predominantly persist as circular monomeric or concatameric episomes (86). In comparison to retroviral or lentiviral vectors, complications regarding insertional mutagenesis have not been reported in any clinical trial. One limitation of AAV vectors is their small genome capacity of approximately 4.5 kb. This however, is sufficient for many human genes, which are generally expressed as cDNAs. For the expression of larger transgenes, methods exist, that separate the therapeutic gene on two or more vector genomes. After successful transduction the full-length transgene is recovered by head-to-tail concatamerization of the vector genomes and subsequent trans-splicing (87, 88).

1.1.5.2 Generation of Recombinant AAV (rAAV) Vectors

In general, for AAV vector generation the transgene of choice is cloned between the AAV-ITRs under the control of a heterologous promoter. The AAV-ITRs contain *cis*-acting elements, required for efficient replication and encapsidation of the vector genome and are the only elements retained for recombinant AAV (rAAV) vectors (89). AAV *rep* and *cap*, as well as the helper functions provided by adenovirus or herpesvirus are expressed *in trans* either from a helper plasmid or a recombinant virus. Alternatively, AAV helper functions can be constitutively expressed in a producer cell line.

The most widely used method for AAV vector production represents transient co-transfection of adherent HEK 293 cells with two or three plasmids encoding the rAAV transgene, the *rep/cap* expression cassette, and the adenovirus helper genes for E2A, E4orf6 and VA RNA (Fig. 4) (90, 91). The helper functions E1A and E1B are constitutively expressed by the 293 cell line (92). The co-transfection approach is efficient for laboratory applications but limited in scalability. Alternative approaches, such as the infection of suspension BHK mammalian cells or *Sf9* insect cells with recombinant herpesvirus (93) or baculovirus strains (94), respectively, have been developed for large scale production as needed for bioreactor scale-up production for clinical applications.

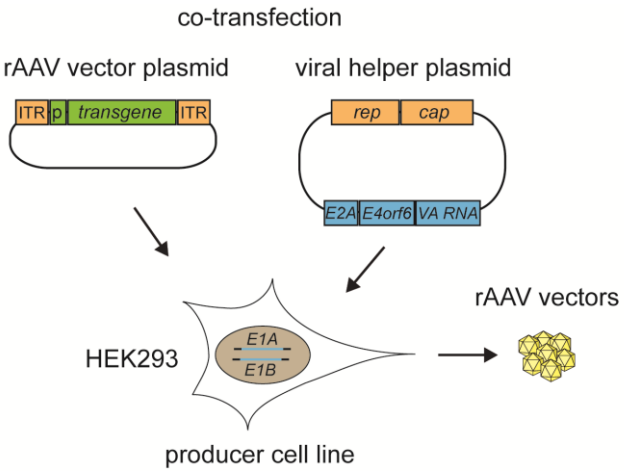


Fig. 4: Generation of rAAV vectors. Schematic depiction of the production of recombinant AAV (rAAV) vectors by transient transfection of HEK 293 cells.

1.2 Non-Coding RNAs

The human genome contains approximately 25,000 protein-coding genes, which correlates to only about 2% of the entire genome. These are transcribed to messenger RNAs (mRNAs) which are eventually translated to proteins at the ribosomes. However, mRNAs or coding RNAs only make up a small portion of total RNAs in mammalian cells. In fact, more than 70% of the genome is transcribed into RNAs (95). The majority is designated as non-coding RNAs (ncRNAs) with a wide range of regulatory functions. The best-characterized ncRNAs include the ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs), as well as small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs), respectively. These ncRNAs are involved in central processes of the cell. The rRNAs as major component of the ribosomes and the tRNAs carrying the amino acids are required for protein biosynthesis. While rRNAs and tRNAs are mostly found in the cytoplasm, snRNAs and snoRNAs are located in the nucleus. There they are involved in posttranscriptional mRNA processing, e.g. splicing or RNA editing. In all these processes the ncRNAs have been found to interact with RNA-binding proteins.

In recent years, increasing numbers of new classes of long non-coding RNAs (lncRNAs) and small non-coding RNAs have been identified. Many of them can regulate gene expression through a diversity of mechanisms. For example, transcripts derived from flanking gene regions can modulate the expression of an adjacent gene, e.g. by altering the chromatin structure (96). Many ncRNAs have been associated with cancers or other diseases (97).

1.2.1 Small Non-Coding RNAs

RNAs with a size smaller than 200 nucleotides that do not code for a protein product are termed small non-coding RNAs (small ncRNAs). The first small ncRNAs with regulatory functions were discovered in the nematode *C. elegans*, where small temporal RNAs, known as lin-4 and let-7, have been shown to regulate the timing of developmental larval stages (98, 99). Soon thereafter, let-7 homologs and further small ncRNAs involved in various cellular regulatory processes were identified in diverse metazoan species, some of which are evolutionary conserved (100, 101). Today, this class of endogenous RNAs, known as microRNAs (miRNAs), makes up one of the largest gene families in higher eukaryotes. The human genome encodes approximately 2000 miRNAs and it is estimated that at least half of the transcriptome is regulated by one or multiple miRNAs (102). Some of the cellular processes modulated by miRNAs include cell cycle progression (103), the innate and adaptive immune response (104), apoptosis (105), hematopoiesis (106) and tumorigenesis (107).

1.2.1.1 Small Non-Coding RNA Species

In the last decade, several classes of small ncRNAs with regulatory functions have been described (Tab. 1). The best-studied small ncRNAs represent microRNAs (miRNAs), small interfering RNAs (siRNAs) and piwi-interacting RNAs (piRNAs). MiRNAs and siRNAs have similar biological roles and are characterized by their double-stranded RNA precursors. Both are typically seen in posttranscriptional gene silencing, but also transcriptional gene silencing achieved by DNA or histone modification has been described (108-110). Nonetheless, clear distinctions can be made particularly regarding their origin and evolutionary conservation (111). While miRNAs are derived from individual genes, siRNAs often derive from mRNAs or transposons. High sequence conservation can be observed for miRNAs of related species, whereas endogenous siRNAs are rarely conserved. Endogenous siRNAs typically silence the same or very similar loci from which they originate, whereas miRNAs silence a broad range of various genes. In contrast to miRNAs and siRNAs, piRNAs are primarily found in animals and appear to be derived from single-stranded RNA precursors (112). They are clearly germline-specific and are usually between 26 - 31 nt in length, which is different from mature si/miRNAs that have a size of about 22 nt. PiRNAs are involved in silencing of transposable elements and the regulation of the chromatin state in germline cells (113, 114).

For a long time, the transfer RNAs (tRNAs) and small nucleolar RNAs (snoRNAs) were believed to be restricted to their generic cellular functions in mRNA translation and rRNA modification, respectively. Surprisingly, studies identified tRNA-derived fragments (tRFs) which are involved in posttranscriptional gene silencing, translational repression and the modulation of cell proliferation (115-117). This new class of small ncRNAs can be divided in different subtypes according to their position of origin within the pre-tRNA or mature tRNA (118). Mammalian snoRNAs are evolutionary conserved, generally derive from introns and range from 60 to 300 nt in length (119). They are located in the nucleolus where they are involved in rRNA processing and site-specific nucleotide modification of rRNAs and snRNAs (120). Furthermore, snoRNA-derived RNAs (sdRNAs) have been reported that are involved in posttranscriptional gene silencing similar to miRNAs (121, 122). The presence of miRNAs located in the nucleolus points to a potential role of snoRNA-derived miRNAs in this subcellular compartment (123). Another RNA species similar to miRNAs is represented by mirtrons (124). These intron-encoded miRNAs account for 5 to 10% of the total miRNA genes in vertebrates, but are less frequently expressed compared to miRNAs generated by the canonical processing pathway. Further classes of small ncRNAs involved in posttranscriptional gene silencing include short hairpin RNAs (shRNAs) and microRNA-offset RNAs (moRs). These RNA species are less well characterized and are classified as separate species based on their divergent biogenesis in comparison to the miRNA processing pathway. MoRs are highly conserved among older families of animal miRNAs, indicative of a distinct class of small ncRNAs (125).

Tab. 1: Summary of small non-coding RNA species described here.

Small non-coding RNAs	*Function*	*Size [nt]*
microRNA (miRNA)	PTGS, DNA/histone modification (TGS)	21 - 25
small interfering RNA (siRNA)	PTGS, DNA/histone modification (TGS)	21 - 23
piwi-interacting RNA (piRNA)	epigenetic gene silencing in germline cells	26 - 31
tRNA-derived fragment (tRF)	PTGS, translational repression and others	18 - 32
snoRNA-derived RNA (sdRNA)	PTGS	17 - 30
mirtron	PTGS	21 - 25
short hairpin RNA (shRNA)	PTGS	21 - 25
microRNA-offset RNA (moR)	(PTGS)	19 - 20
Promoter-associated RNAs		
long paRNA (PALR)	DNA modification, transcription regulation, siRNA precursors	> 200
small paRNA (PASR)	maintenance of active transcription state, TGS	< 200
tiny paRNA (tiRNA)	marking of highly transcribed genes	18

PTGS = posttranscriptional gene silencing, TGS = transcriptional gene silencing

1.2.1.2 Promoter-Associated Small Non-Coding RNAs

Some classes of small ncRNAs are specifically associated with promoter activities. These promoter-associated RNAs (paRNAs) are transcribed close to the transcription start or within the promoter region. In the literature alternative names for this class of RNAs exist, e.g. transcription start site-associated RNAs (TSSaRNAs) (126, 127). PaRNAs can be divided into long, small and tiny RNA subspecies (Tab. 1). Human small paRNAs (PASRs) transcribe in both, the sense and the antisense direction of active promoters with peak fractions around +50 bp downstream and -250 bp upstream of the TSS (126). This bidirectional pattern of transcription initiation is termed divergent transcription and can be recognized for many mammalian promoters, especially for those that lack a strong TATA element (127, 128). The human small paRNAs are 5' capped and arise either as independent transcripts or represent processing products from longer RNAs, likely long paRNAs (PALRs) (129). Their rather low expression level is positively correlated to the promoter activity and mRNA level (126). This small ncRNA class is presumably required to maintain an open chromatin structure and relax DNA-negative supercoils for active transcription from the corresponding promoter (130). Additionally, small paRNAs have been suggested to play an important role in transcriptional gene silencing (TGS) by targeting the promoter sequence (129). Similar functional roles are suggested for the tiny paRNAs or transcription initiation RNAs (tiRNAs), since they share several common features with the small paRNA subspecies. They are transcribed in both directions from the promoter with a preference for the sense direction forming a peak fraction between +10 and +30 bp downstream the TSS (131). The tiny paRNAs are low abundant transcripts of 18 nt in size and are associated with RNA polymerase II binding sites. However, in contrast to the small paRNAs, the abundance of tiny paRNAs does not correlate with the gene expression level (131). They are suggested to be generated by RNApol II pausing, backtracking and cleavage by the transcription factor IIS (TFIIS) and are postulated to mark highly transcribed genes (132). Many more emerging classes of small ncRNAs are described, most of them however, with unknown function and some being controversially discussed.

1.2.2 Biogenesis of Small Non-Coding RNAs

Since research in terms of biogenesis and biological function is mostly advanced in the field of miRNAs, other classes of small ncRNAs are often compared to the canonical

miRNA pathway. In animals, miRNAs are generally transcribed by RNApol II, followed by 5' capping and 3' polyadenylation (133) (Fig. 5A). This leads to the primary miRNA transcript (pri-miRNA) that is usually several kb long and possesses one or multiple internal hairpin structures. The primary miRNA transcript is recognized and processed by the nuclear RNase III endonuclease Drosha and its cofactor DGCR8, which is also known as the microprocessor complex (134). This processing step liberates a stem loop intermediate of approximately 60 - 70 nt that exhibit a 5' phosphate and a 2 nt 3' overhang, known as precursor miRNA (pre-miRNA). The pre-miRNA is transported to the cytoplasm by the nuclear export receptor exportin-5 where a second RNase III endonuclease, called Dicer and its cofactor TRBP cleaves off the terminal loop to release a miRNA duplex of ~22 nt in length (111, 133). Usually one strand of the miRNA duplex, the mature miRNA strand, is chosen to be incorporated into the miRNA ribonucleoprotein complex (miRNP), while the other strand, the star strand, is degraded. Strand selection of the miRNA duplex depends on the relative thermodynamic stability of the duplexes' ends (133). Within the miRNP complex the mature miRNA is directly associated with the argonaute protein, which possesses endonucleolytic activity (108).

Typically, miRNAs act on a posttranscriptional level by base-pairing to their target mRNA. This leads to mRNA cleavage and degradation in the case of perfect complementarity between miRNA and target mRNA, the usual mechanism carried out in plants (111). Binding of the miRNA to the target site with partial sequence complementarity leads to varying degrees of translational repression, which is more common in animals. The so-called 'seed region' which comprises nucleotides 2 - 8 at the 5' end of the mature miRNA is especially important in mediating target specificity (135). In animals the miRNA seed region typically binds the 3' UTR of its target mRNA with perfect complementarity thereby inhibiting mRNA translation.

Endogenous siRNAs are generated from long double-stranded RNA precursors and also require Dicer processing, but are independent from cleavage by Drosha (Fig. 5B). Similar to the miRNA pathway, a short RNA duplex is generated during the canonical RNAi pathway of which one strand, the guide strand, is loaded into the RNA-induced silencing complex (RISC) (108). Binding of the siRNA to the mRNA target site with perfect sequence complementarity leads to endonucleolytic cleavage and degradation of the target mRNA.

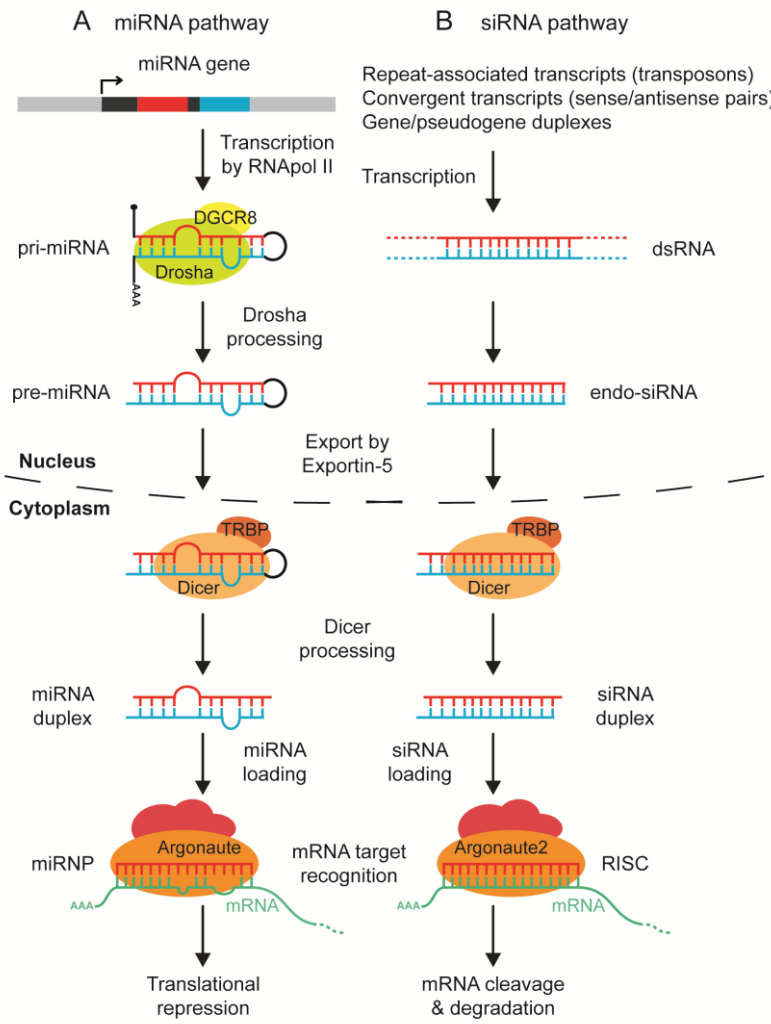


Fig. 5: Biogenesis of small ncRNAs. (A) Schematic depiction of the canonical miRNA processing pathway leading to translational repression. **(B)** Schematic depiction of the siRNA processing pathway resulting in mRNA cleavage and degradation. pri-miRNA = primary miRNA, pre-miRNA = precursor miRNA, miRNP = miRNA ribonucleoprotein complex, dsRNA = double-stranded RNA, endo-siRNA = endogenous siRNA, RISC = RNA-induced silencing complex.

With the exception of piwi-interacting RNAs, which are generated by a different processing pathway, most of the other small ncRNA species described here merge into the miRNA pathway. Mirtrons derive from short introns that are spliced to pre-miRNA like hairpin structures thereby circumventing Drosha processing (136). ShRNAs are derived from intergenic regions and are directly transcribed as short hairpins, which are further processed by Dicer, but are independent of Drosha processing (137). MoRNAs generally derive from sequences adjacent to the mature miRNA and are most likely generated during processing of the primary miRNA hairpin by Drosha (138). Similarly, some of the subtypes of tRFs and sdRNAs are processed by Dicer, but their initial processing differs from the canonical miRNA Drosha processing (139, 140).

1.2.3 Virus-Encoded Small Non-Coding RNAs

Viruses infecting host cells can utilize the host RNA processing pathways to express their own small RNAs. The first virus-encoded microRNAs were described in 2004 in Epstein-Barr virus (EBV), a member of the herpesvirus family (141). Currently, about 1300 experimentally validated viral miRNA sequences (including isoforms thereof) encoded by 44 mostly mammalian virus species are documented in the database of viral microRNAs (VIRmiRNA) (142) (Tab. 2).

Tab. 2: Summary of DNA and RNA viruses that express small ncRNAs.

Virus family or subfamily	*Virus species*	*No. of miRNAs * (pre-miRNAs)*	*Host*
DNA viruses			
Alpha-herpesvirinae	Herpes simplex virus 1 (HSV1)	>26 (>17)	Human
	Herpes simplex virus 2 (HSV2)	24 (17)	Human
	Herpes B virus (HBV)	16 (13)	Simian
	Marek's disease virus 1 (MDV1)	26 (13)	Avian
	Marek's disease virus 2 (MDV2)	36 (18)	Avian
	Herpesvirus of turkeys (HVT)	28 (17)	Avian
	Infectious laryngotracheitis virus (ILTV)	10 (7)	Avian
	Bovine herpesvirus 1 (BHV1)	12 (10)	Bovine
	Pseudorabies virus (PRV)	18 (11)	Porcine
Beta-herpesvirinae	Human cytomegalovirus (HCMV)	17 (11)	Human
	Mouse cytomegalovirus (MCMV)	28 (18)	Murine
	Human herpesvirus 6B (HHV6B)	8 (4)	Human
Gamma-herpesvirinae	Epstein-Barr virus (EBV)	44 (25)	Human
	Kaposi's sarcoma-associated herpesvirus (KSHV)	25 (12)	Human
	Rhesus lymphocryptovirus (RLCV)	68 (36)	Simian
	Rhesus monkey rhadinovirus (RRV)	27 (15)	Simian
	Herpesvirus saimiri strain A11 (HVS)	6 (3)	Simian
	Murine gammaherpesvirus 68 (MHV68)	29 (15)	Murine
Polyomaviridae	Merkel cell polyomavirus (MCV)	2 (1)	Human
	BK polyomavirus (BKV)	2 (1)	Human
	JC polyomavirus (JCV)	2 (1)	Human
	Simian virus 40 (SV40)	2 (1)	Simian
	Simian agent 12 (SA12)	2 (1)	Simian
	Murine polyomavirus (PyV)	2 (1)	Murine
Papillomaviridae	Human papillomavirus (HPV) (different types)	9	Human
Polyoma/ Papillomaviridae	Bandicoot papillomatosis carcinomatosis virus type 1 (BPCV 1)	1 (1)	Bandicoot
	Bandicoot papillomatosis carcinomatosis virus type 2 (BPCV 2)	1 (1)	Bandicoot
Adenoviridae	Human adenovirus type 2 & 5 (HAdV)	3 (2)	Human
Ascoviridae	Heliothis virescens ascovirus (HvAV)	1 (1)	Insect
Baculoviridae	Bombyx mori nucleopolyhedrosis virus (BmNPV)	5 (?)	Insect
RNA viruses			
Retroviridae	Human immunodeficiency virus 1 (HIV1)	4 (3)	Human
	Bovine leukemia virus (BLV)	8 (5)	Bovine
	Bovine foamy virus (BFV)	3 (2)	Bovine
	Simian foamy virus (SFV)	>2	Simian
	Avian leukemia virus subgroup J (ALV-J)	1 (1)	Avian
Flaviviridae	West Nile virus (WNV)	1 (1)	Insect
	Dengue virus 2 (DENV-2)	? (6)	Insect

*Numbers of miRNAs (pre-miRNA) according to the VIRmiRNA database (142).

Most of the virus-encoded miRNAs stem from various members of the herpesvirus family (Tab. 2). In fact, most herpesviruses encode clusters of miRNAs, typically more than 10 per genome (143). In contrast, only a few miRNAs are encoded by members of the polyomavirus and adenovirus group (144, 145) and presumably by the human papillomaviruses (146) (Tab. 2). In addition, miRNAs have also been described in the insect virus families of ascoviruses and baculoviruses (147, 148). All these virus families comprise viruses with DNA genomes that replicate in the nucleus where they have access to the initial part of the miRNA processing machinery.

Whether RNA viruses express miRNAs had been long controversial due to the negative effects on fitness that might result by unintended cleavage of the genome or mRNA transcripts. For the human immunodeficiency virus (HIV1) some virus-encoded miRNAs have been reported (149), but they are not widely accepted. However, it has been established that at least four retroviruses generate miRNAs (150-153) (Tab. 2). Furthermore, some flaviviruses with single-stranded (+) RNA genome have been shown to encode miRNA-like viral small RNAs (154, 155) (Tab. 2).

1.2.4 Functions of Virus-Encoded Small Non-Coding RNAs

Typically, virus-encoded miRNAs are expressed to support virus propagation by generating beneficial host cell conditions or to evade the cellular immune response (156). On the other hand, some of the best-characterized viral miRNA functions include the regulation of viral gene expression for the establishment of long-term persistence (157). Interestingly, some viruses, especially tumor-inducing viruses, express homologs that mimic cellular miRNAs that regulate the same set of cellular target genes (158). However, for most of the identified virus-encoded miRNAs the exact roles have not yet been defined.

1.2.4.1 HSV1 miRNAs

As a member of the herpesvirus family, herpes simplex virus type 1 (HSV1) exhibits a biphasic life cycle and persists throughout the lifetime of its host after the initial infection. During HSV1 latency the only abundantly expressed viral transcript represents the latency-associated transcript (*LAT*) (159). This non-coding RNA is spliced generating a ~6.3 kb exon product and a stable intron of ~2.0 kb. The *LAT* intron is described to maintain the HSV latent state by inhibiting cellular apoptosis and silencing viral gene expression through alteration of the heterochromatin structure at the viral promoters (160, 161). The

exonic product of *LAT* serves as a precursor for several miRNAs that are likely to synergistically function in the maintenance of the latent state (159). HSV1 encodes at least 18 well-described miRNAs, most of them expressed during latency (162), but further miRNAs have been identified, which are expressed in early lytic infection (163). At least for three miRNAs the molecular function during the herpesvirus life cycle has been determined: The miR-H2 and miR-H6 repress the major HSV1 transcriptional activators ICP0 and ICP4, respectively, and thus support the maintenance of the latent state (159), whereas the miR-H4 has been shown to target the HSV1 pathogenicity factor ICP34.5 (164).

1.2.4.2 Adenovirus miRNAs

Adenovirus (Ad), the other major helper virus of AAV, generates miRNAs processed from the longer structured virus-associated RNAs, VA RNA I and II. The VA RNAs represent RNA polymerase III transcribed non-coding RNAs of approximately 160 nt in size, which can adopt secondary structures that resemble miRNA precursors (165). The VA RNAs are required to block the cellular interferon-induced antiviral defense mechanism by binding to protein kinase R (PKR), thereby stimulating viral protein synthesis (166). Furthermore, they are described to suppress the cellular RNA interference (RNAi) pathway by interfering with the activity of Dicer (167). The role of the VA RNA derived miRNAs (mivaRNAs) within the adenovirus life cycle has not been fully defined yet. Several potential cellular target genes have been identified, some of these which are involved in cell growth, gene expression and DNA repair (168).

1.2.4.3 Unconventional Virus-Encoded Small Non-Coding RNAs

Although most virus-encoded small regulatory RNAs represent miRNAs, a few other small RNA species as well as non-canonical miRNA-like structures that contain miRNAs have been described to be expressed from viral genomes. For instance, HSV1, HSV2 and the human cytomegalovirus (HCMV) express microRNA-offset RNAs (moRs) that arise from viral miRNA precursors (162, 169). EBV expresses a viral snoRNA-derived RNA of 24 nt in size, v-snoRNA1(24pp), that target the viral DNA polymerase for mRNA degradation (170). The murine gammaherpesvirus 68 (MHV68) encodes small hybrid ncRNAs that contain tRNA-like fragments and functional miRNAs that are alternatively processed from the tRNA-like structures by RNaseZ (171).

1.3 DNA / RNA Sequencing

The conventional method for determining DNA sequences, so-called Sanger sequencing, is based on the complementary strand synthesis of a DNA template by a DNA polymerase initiated from a synthetic sequence-specific primer molecule. During DNA synthesis either natural deoxynucleotides (dNTPs) or dideoxynucleotides (ddNTPs) can be incorporated into the newly generated strand. However, utilization of a ddNTP leads to strand termination (172). Since nucleotide incorporation occurs at random, strands of various lengths are generated during this process. The resulting DNA sequence is determined by capillary electrophoresis and fluorescence detection that provide four-color plots. This automated Sanger method is considered as the first generation sequencing technology and led to remarkable progress in research including the completion of the first human genome sequence (173). However, this approach has its limits since this conventional method can only sequence one DNA molecule at a time. For a broader application, this technique is too time-consuming and expensive.

1.3.1 Next Generation Sequencing (NGS)

In the last decade, new sequencing technologies were developed referred to as next generation sequencing (NGS). Compared to conventional sequencing methods, these new sequencing platforms produce an enormous volume of data by massively parallel sequencing. Currently, multiple NGS platforms exist, such as the Genome Analyzer (Illumina/Solexa) or the 454-FLX (Roche), all of them having its own specifics, e.g. in terms of the synthesized read length (174). Due to that a broad range of applications are available. For instance, ChIP-seq and methyl-seq are utilized to screen for genome-wide DNA-protein interactions and epigenetic marks within chromatin structures, respectively (175, 176). Another common application using NGS is the detection of sequence variations between genomes, which range from single nucleotide variations to larger targeted regions, such as genomic deletions or insertions (177). In case of human genomes, this can enhance our understanding in terms of how genetic differences can affect health and genetic diseases. Furthermore, RNA-Seq can be used to analyze the expression level of entire transcriptomes, as well as to identify novel transcripts of low abundance and alternative splice variants (178, 179). The rise of next generation sequencing technologies has largely

replaced hybridization-based microarray analyses, which are limited in detecting a comprehensive picture of transcripts expressed from the genome.

1.3.2 Illumina-Based Next Generation Sequencing

In general, NGS technologies rely on a combination of methods comprising library preparation, sequencing and imaging followed by bioinformatic data analysis, such as genome alignment and assembly methods. For Illumina-based sequencing various protocols for template preparation are provided for different applications, such as total RNA sequencing or small RNA sequencing. After library preparation, which includes the ligation of specific adapters, the individual nucleic acid molecules are immobilized on a flow cell of the sequencing device. In the first step, the immobilized templates are clonally amplified to high density clusters via bridge amplification (Fig. 6A) (174). In that process, millions of separate template clusters are generated in parallel. The utilization of multi-channel flow cells allows the sequencing of multiple independent samples simultaneously. For the actual sequencing reaction modified nucleotides are used, designated as 3'-blocked reversible terminators, each containing a different fluorescent dye to achieve four-color imaging (180). After initial primer annealing, the first nucleotide is incorporated by a DNA polymerase (Fig. 6B). The unincorporated nucleotides are washed away and imaging is performed. The detection of fluorescent signals is executed throughout the entire flow cell allowing the simultaneous detection of incorporated nucleotides of millions of clusters. A subsequent cleavage step removes the terminating group, as well as the fluorescent dye. After an additional washing step the sequencing reaction proceeds with a second round of added nucleotides, imaging, cleavage and washing. This cycle is repeated for several rounds. The number of cycles determines the length of the sequence reads. In the end, the generated sequence reads are aligned to known reference genomes or assembled *de novo*, in case of unknown sequenced genomes.

The Illumina NGS platform is especially favorable to produce a massive data volume of sequence reads with rather short read lengths (50 to 150 nucleotides depending on the utilized mode).

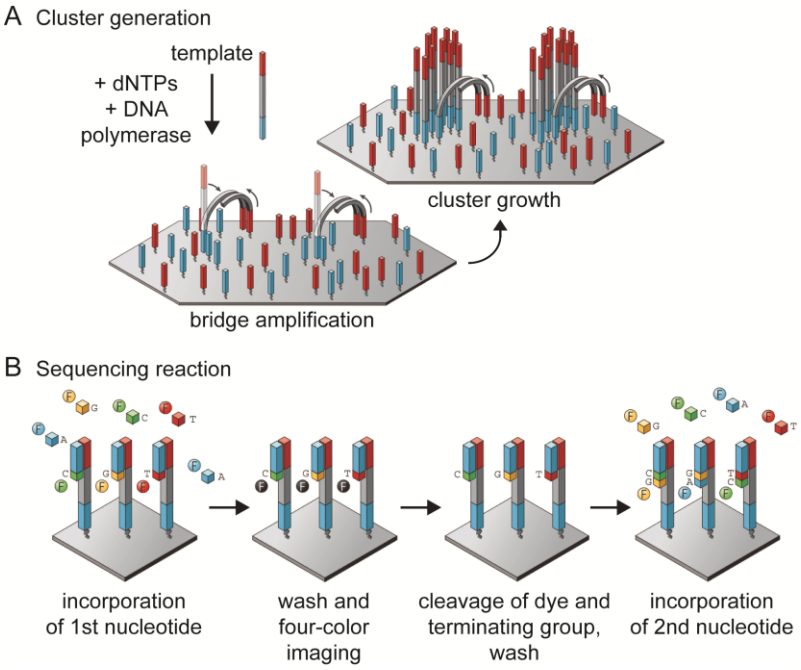


Fig. 6: Illumina-based NGS platform. (A) After template preparation millions of clonal clusters are generated via bridge amplification on a flow cell. **(B)** Next generation sequencing is achieved by the stepwise incorporation of reversible terminators and four-color imaging. Figure adapted from Metzker 2010 (174).

1.4 Aim of the Thesis

In recent years, vectors based on adeno-associated viruses (AAVs) have received significant attention in the field of human gene therapy. Long-term successes for monogenetic disorders in clinical trials and the approval of the first gene therapy medication in the western world further enhanced the development of AAV-derived vectors. Despite the widespread application of AAV vectors, there is an obvious lack of knowledge regarding the *in vivo* replication cycle of AAV wild-type. This is of specific importance since the majority of the human population had contact to wild-type AAVs at some point of their life and might still be carrier of the virus that could potentially interact with AAV vectors.

The aim of this thesis was to analyze novel and basic aspects of the AAV transcriptome, in order to achieve a profound understanding of transcription regulation during AAV latent and lytic infection. For this purpose an in-depth sequencing approach based on Illumina RNA next generation sequencing was chosen. To this end, RNA libraries generated from total RNAs on the one hand and size-selected small RNAs on the other hand were required that represent different stages of the AAV latent and productive replication cycle, respectively.

The total RNA sequencing analysis was conducted to receive a comprehensive picture of the AAV transcriptome during latent and lytic infection. This approach allowed for the detection of low expressed transcripts, which may not be detectable with other methods used in previous studies. Furthermore, AAV splicing can be analyzed which may reveal alternative splice variants that can lead to novel transcripts and/or proteins. The combination of Illumina-based total RNA sequencing and small RNA sequencing enables a variety of further analyses, such as the effect of AAV infection on the host cell transcriptome.

In particular, one major objective of this thesis was to screen for the presence of small non-coding RNAs encoded by AAV, which have not been described for any of the parvoviruses yet. Many other DNA viruses have been described to express miRNAs that are involved in the regulation of the viral life cycle. A striking parallel to AAV is that most of these viruses exhibit a biphasic life cycle and establish long-term persistence. These arguments speak in favor for the presence of AAV-encoded small RNAs. In fact, early AAV transcription studies showed small transcripts that were described as degradation products. The parallel analysis of adenovirus and herpes simplex virus infected cells, which both express miRNAs, allowed an internal validation of this approach by the detection of known small RNAs.

Furthermore, a comparative study of AAV2 and the most distantly related serotype AAV5 on the dependence of HSV1 helper functions during AAV productive replication was performed to further evaluate the principles of AAV biology.

2 Publications

2.1 A Comprehensive RNA Sequencing Analysis of the Adeno-Associated Virus (AAV) Type 2 Transcriptome Reveals Novel AAV Transcripts, Splice Variants, and Derived Proteins

Authors: Catrin Stutika, Andreas Gogol-Döring, Laura Botschen, Mario Mietzsch, Stefan Weger, Mirjam Feldkamp, Wei Chen, Regine Heilbronn

Year: 2016

Journal: Journal of Virology 90 (3): 1278-1289

NCBI Sequence Read Archive (SRA) accession: PRJNA322856

2.1.1 Contribution to the Publication

The aim of this study was an unbiased and comprehensive analysis of the AAV transcriptome during latent and lytic infection based on an Illumina in-depth total RNA genomic sequencing approach.

For this publication I initially produced a highly purified AAV2 virus stock and performed preliminary AAV expression studies. I further generated high-quality DNase-digested RNA samples as a prerequisite for Illumina RNA next generation sequencing (NGS). The preparation of total RNA libraries and the sequencing on the Illumina HiSeq system was carried out in cooperation with Wei Chen and Mirjam Feldkamp at the Berlin Institute for Medical Systems Biology at the Max Delbrück Center in Berlin-Buch, Germany. The bioinformatic analysis of the sequencing raw data was performed by Andreas Gogol-Döring at the Technische Hochschule Mittelhessen in Gießen, Germany. The resulting sequencing data were interpreted by me, supported by Andreas Gogol-Döring and my colleague Mario Mietzsch and led to the concept for further experiments unraveling the significance of the findings for AAV biology. Detection analyses of the novel splice variants and expression analyses of potentially generated proteins thereof were performed by me. Furthermore, I analyzed how the novel splice variants effect AAV replication. Mario Mietzsch helped by generating the required AAV splice mutant AAV-A1. Additional AAV splice mutants (AAV-D1 and AAV-D2) generated by Stefan Weger were

used to complement the analysis of alternative AAV splice variants. I created and compiled the tables and illustrations with their corresponding descriptions. Laura Botschen added Fig. 1D, E and F. The manuscript was written by me, Mario Mietzsch and my mentor Regine Heilbronn.

2.1.2 Article

DOI: http://dx.doi.org/10.1128/JVI.02750-15

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

A Comprehensive RNA Sequencing Analysis of the Adeno-Associated Virus (AAV) Type 2 Transcriptome Reveals Novel AAV Transcripts, Splice Variants, and Derived Proteins

Catrin Stutika,[a] Andreas Gogol-Döring,[b]* Laura Botschen,[a] Mario Mietzsch,[a] Stefan Weger,[a] Mirjam Feldkamp,[c] Wei Chen,[c] Regine Heilbronn[a]

Institute of Virology, Campus Benjamin Franklin, Charité Medical School, Berlin, Germany[a]; German Centre for Integrative Biodiversity Research, Halle-Jena-Leipzig, Germany[b]; Laboratory for Functional Genomics and Systems Biology, Berlin Institute for Medical Systems Biology, Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany[c]

ABSTRACT

Adeno-associated virus (AAV) is recognized for its bipartite life cycle with productive replication dependent on coinfection with adenovirus (Ad) and AAV latency being established in the absence of a helper virus. The shift from latent to Ad-dependent AAV replication is mostly regulated at the transcriptional level. The current AAV transcription map displays highly expressed transcripts as found upon coinfection with Ad. So far, AAV transcripts have only been characterized on the plus strand of the AAV single-stranded DNA genome. The AAV minus strand is assumed not to be transcribed. Here, we apply Illumina-based RNA sequencing (RNA-Seq) to characterize the entire AAV2 transcriptome in the absence or presence of Ad. We find known and identify novel AAV transcripts, including additional splice variants, the most abundant of which leads to expression of a novel 18-kDa Rep/VP fusion protein. Furthermore, we identify for the first time transcription on the AAV minus strand with clustered reads upstream of the p5 promoter, confirmed by 5' rapid amplification of cDNA ends and RNase protection assays. The p5 promoter displays considerable activity in both directions, a finding indicative of divergent transcription. Upon infection with AAV alone, low-level transcription of both AAV strands is detectable and is strongly stimulated upon coinfection with Ad.

IMPORTANCE

Next-generation sequencing (NGS) allows unbiased genome-wide analyses of transcription profiles, used here for an in depth analysis of the AAV2 transcriptome during latency and productive infection. RNA-Seq analysis led to the discovery of novel AAV transcripts and splice variants, including a derived, novel 18-kDa Rep/VP fusion protein. Unexpectedly, transcription from the AAV minus strand was discovered, indicative of divergent transcription from the p5 promoter. This finding opens the door for novel concepts of the switch between AAV latency and productive replication. In the absence of a suitable animal model to study AAV *in vivo*, combined *in cellulae* and *in silico* studies will help to forward the understanding of the unique, bipartite AAV life cycle.

Adeno-associated viruses (AAV) are helper-dependent members of the parvovirus group that require coinfection with an unrelated helper virus, particularly adenovirus (Ad) for productive replication (1). Despite of several identified AAV serotypes most research has been done with prototype AAV type 2. The AAV2 genome consists of a linear, single-stranded DNA of 4.7 kb. Both ends carry identical inverted terminal repeats (ITR) of 145 bp (2), which flank the two major open reading frames (ORF), called *rep* and *cap*. The *rep* ORF codes for four nonstructural proteins, Rep78 and a C-terminally spliced variant, Rep68. In addition, N-terminally truncated versions thereof are expressed, called Rep52 and Rep40, respectively. The Rep proteins are required as regulators for various steps of the AAV life cycle. AAV *cap* harbors the ORF for the capsid proteins (VP1 to -3) and a separate ORF for the assembly-activating protein (AAP) (3).

The current knowledge of AAV transcription dates back to early research in the 1980s when the three AAV2 promoters, p5, p19, or p40 and major transcripts derived thereof were characterized (4, 5). All AAV transcripts identified since then map to only one DNA plus strand (6), which led to the assumption that the complementary AAV minus strand was not transcribed. Furthermore, spliced AAV transcripts were identified displaying a single, shared intron located in the center of the AAV2 genome (4, 7). Initially, a single splice donor site at AAV nucleotide 1906 and a single splice acceptor site located at nucleotide (nt) 2228 were described. Later, an alternative splice acceptor site at nt 2201 was detected (8), allowing the expression of VP1 protein from the same promoter (p40) as VP2 and VP3. Compared to the major splice acceptor site (nt 2228), the so-called minor splice acceptor site was more divergent from the ideal consensus sequence, likely explaining its less-frequent use

Received 28 October 2015 **Accepted** 6 November 2015

Accepted manuscript posted online 11 November 2015

Citation Stutika C, Gogol-Döring A, Botschen L, Mietzsch M, Weger S, Feldkamp M, Chen W, Heilbronn R. 2016. A comprehensive RNA sequencing analysis of the adeno-associated virus (AAV) type 2 transcriptome reveals novel AAV transcripts, splice variants, and derived proteins. J Virol 90:1278–1289. doi:10.1128/JVI.02750-15.

Editor: L. Banks

Address correspondence to Regine Heilbronn, regine.heilbronn@charite.de.

* Present address: Andreas Gogol-Döring, Technische Hochschule Mittelhessen, Giessen, Germany.

C.S. and A.G.-D. contributed equally to this article.

TABLE 1 Oligonucleotides for mutagenesis PCR, qPCR, RT-PCR, and PCRs for detection of AAV2 splice events, Northern blot hybridization, and 5′RACE

Oligonucleotide	Sequence (5′-3′)
AAV-D1-Fwd	CTGACGGAATGGCGCCGTCTTTCGAAGGCCCCGGAGGCCCTT
AAV-D1-Rev	AAGGGCCTCCGGGGCCTTCGAAAGACGGCGCCATTCCGTCAG
AAV-D2-Fwd	CTGACGGAATGGCGCCGTGTTTCGAAGGCCCCGGAGGCCCTT
AAV-D2-Rev	AAGGGCCTCCGGGGCCTTCGAAACACGGCGCCATTCCGTCAG
AAV-A1-Fwd	CTTAAACACCCTCCTCCACAAATTCTCATCAAGAACACC
AAV-A1-Rev	GGTGTTCTTGATGAGAATTTGTGGAGGAGGGTGTTTAAG
Flag-Fwd	CCTGACTCGTAATCTGTAATGACTACAAGGATGACGATGACAAGTGATGCTTGTTAATCAATAAACCG
Flag-Rev	CGGTTTATTGATTAACAAGCATCACTTGTCATCGTCATCCTTGTAGTCATTACAGATTACGAGTCAGG
Stop-Mut-Fwd	CCATTGGCACCAGATACCTTACTCGTAATCTGTAATGAC
Stop-Mut-Rev	GTCATTACAGATTACGAGTAAGGTATCTGGTGCCAATGG
Rep2-Fwd	AGAAGGAATGGGAGTTGCCG
Rep2-Rev	TCTGACTCAGGAAACGTCCC
RT-Primer	CAGATTACGAGTCAGGTAT-CTGGTGCCAATGGGGCG
SD527-Fwd	AGAAGGAATGGGAGTTGCCG
SD988-Fwd	GTAAACGGTTGGTGGCGCAG
SD1906-Fwd	GCCATCGACGTCAGACGCGGAAGCTTC
SD3184-Fwd	CTTCAACAGATTCCACTGCC
SA807-Rev	GCTCCGTGAGATTCAAACAG
SA985-Rev	AGGCCGCATTGAAGGAGATG
SA1437-Rev	TTGCTTCCTCCGAGAATGGC
SA1635-Rev	TTGGTGACCTTCCCAAAGTC
SA2228-Rev	CGTAGGCTTTGTCGTGCTCG
SA4138-Rev	TGAAGGTGGTCGAAGGATTCGCAGGT
SDA1	CCTCGAGCCAATACGGCGCCATTC
p5-RACE-GSP1	GGTGGAGTCGTGACGTGAATT
p5-RACE-GSP2	TCATAGGGTTAGGGAGGTCCT

for splicing. During AAV replication the majority of p5- or p19-initiated transcripts are not spliced, whereas transcripts initiated from p40 are almost exclusively spliced (9, 10). All AAV transcripts share a 3' end (7), coinciding with the polyadenylation signal identified between AAV2 nt 4424 and 4429 (2). The only other AAV serotype whose transcription pattern has been analyzed is AAV5, the genetically most distant member of the AAV family. Several variations in comparison to the AAV2 transcription pattern were described. Most importantly, AAV5 *rep* transcripts terminate at a second, internal polyadenylation signal located in the AAV5 intron (11).

Early analysis of AAV transcription relied on S1 nuclease mapping and reverse transcriptase-mediated primer extension (4, 7, 12, 13). These methods allowed the mapping and sequence analysis of defined and highly expressed transcripts. Subsequently, quantitative RNase protection was used to complement the currently accepted AAV transcription map (10). Using the described methods, only AAV RNAs in regions explicitly probed for could be detected. Recent NGS-based technologies allow high-throughput analysis of entire cellular transcriptomes, leading to unbiased identification and parallel quantification of all existing cellular and viral transcripts. With RNA sequencing (RNA-Seq), novel transcripts and splicing variants in human cells and of viral transcripts have been discovered (14, 15).

In this report we used Illumina-based RNA-Seq to provide a comprehensive picture of AAV2 transcription in the presence or absence of helper adenovirus. We were able to identify known and novel AAV2 transcripts, splicing variants, and a new protein derived therefrom. In addition, previously unrecognized RNAs transcribed from the AAV minus strand were detected.

MATERIALS AND METHODS

Cell culture and viruses. HEK 293- and HeLa cells were cultivated as described previously (16). Human adenovirus type 2 (Ad2) was propagated and assayed in 293 cells.

Cloning and mutagenesis. For the generation of different luciferase constructs, the corresponding AAV2 wild-type promoter sequences were cloned upstream of the firefly luciferase gene in a sense orientation, AAV(+)190-320-*luc*, or in a reverse complementary orientation, AAV(–)586-155-*luc*, AAV(–)320-180-*luc*, and AAV(–)320-256-*luc*, respectively. The constructs were verified by restriction analysis. For the generation of AAV mutants AAV-D1, AAV-D2, and AAV-A1, the corresponding mutations were introduced by the QuikChange site-directed mutagenesis protocol (Stratagene) into the AAV2 *rep* or *cap* genes of pTAV2-0 (17). The mutations were verified by DNA sequencing and restriction enzyme analysis. Similarly, the nucleotide exchange G4395T and the 9 amino acid C-terminal Flag-tag insertion (DYKDDDDK*) between AAV2 nucleotide positions 4411 and 4412 led to the constructs pTAV2-0 G4395T Flag (pAAVwt-Flag), and pTAV2-0 G4137A G4395T Flag (pAAV-A1-Flag). For the primers, see Table 1.

AAV2 production, purification, and quantification. For AAV2 production, HEK 293 cells seeded at 30% confluence were transfected 24 h later with AAV plasmids and pHelper as described previously (18). AAV2 was purified from benzonase-treated, cleared freeze-thaw supernatants by one-step AVB Sepharose affinity chromatography and quantified by qPCR as described previously (19) with primers for AAV2 *rep* (Table 1). AAV2 infectious titers were determined by endpoint dilutions on Ad2-infected HeLa cells (20).

Virus infection and DNA plasmid transfection. HeLa cells seeded at a density of 30% were infected 20 h later with AAV2 wild-type or AAV-A1 mutant (multiplicity of infection [MOI] 250) and/or Ad2 (MOI 25). Transfection of HeLa cells with AAV plasmids was performed as described previously (21). Infection with Ad2 (MOI 25) was performed 16 h post-

transfection and cells were harvested at the indicated time points postinfection.

RNA extraction. Total RNA was isolated by TRIzol reagent (Ambion) according to the manufacturer's protocol. RNA samples were treated with RNase-free Turbo DNase (Ambion) to remove residual DNA, followed by phenol-chloroform extraction and precipitation. RNA quality and integrity was verified on 0.8% agarose gels and on bioanalyzer (Agilent). Total RNA samples of proven high quality were used for reverse transcription-PCR (RT-PCR) and for RNA library generation. Poly(A)$^+$ selected RNA was purified from total RNA with the Oligotex mRNA Midi purification kit (Qiagen).

RT-PCR. Newly identified AAV2 splice sites were validated using 1 μg of DNase-digested total RNA, reverse transcribed by SuperScript II reverse transcriptase (Invitrogen) according to the manufacturer's protocol with a primer that binds at the rear 3' end of the AAV2 genome (Table 1). The cDNAs were PCR amplified with AAV2-specific primer pairs (Table 1) located upstream of the predicted splice donor and downstream of the predicted splice acceptor sites.

Library preparation and Illumina sequencing. 1 μg of each of the four different total RNA samples (cells, AAV2, AAV2+Ad2, and Ad2) were subjected to total RNA library generation using the TruSeq stranded total RNA sample preparation kit according to the manufacturer's protocol (Illumina). The libraries were pooled and sequenced on a HiSeq 2000 platform (multiplexed SR 1×101+7 high-output mode).

Read mapping. Demultiplexed sequencing reads (length, 101 bp; single end) were mapped after trimming Illumina adapters from the 3' ends on the human genome (hg19, GRCh37) and the human transcriptome by using the Bowtie 2 read mapping tool (22). The transcriptome was derived from the RefSeq gene annotation downloaded from the UCSC table browser (https://genome.ucsc.edu/cgi-bin/hgTables). All reads that could not be mapped to the human genome/transcriptome were mapped to the reference sequences of Ad2 (RefSeq AC_000007.1) and of both the flip and the flop variants of AAV2 (RefSeq NC_001401.2). In all cases, only read mappings with an alignment score (as given in the XS field of the SAM output) of at least −20 were accepted. A detailed analysis of the reads mapped on AAV2 revealed that the genome of the sequenced AAV2 only differs from the reference at a single nucleotide (a C instead of a G at nt 3284).

Detection of splice junctions. We generated a splice junction reference, i.e., a set of sequences containing fusions of all possible combination of canonical splicing donor sites (GT dinucleotides) and downstream splicing acceptor sites (AG dinucleotides) in AAV2. A split alignment of the sequencing read onto the AAV2 genome using the Segemehl mapping tool (23) revealed no evidence of splicing at noncanonical splicing sites (data not shown). The sequencing reads were mapped to the splice junction reference using Bowtie 2, where insertions and deletions were suppressed by setting the gap penalties to very high values. We then determined the number of uniquely mapped reads for each possible junction of donor and acceptor sites. Only reads that covered at least 20 bp upstream of the splice donor site and 20 bp downstream of the splice acceptor site were counted.

Detection of polyadenylation sites. We searched the reverse complement sequencing reads for polyadenine repeats of a length of 8 (i.e., AAAAAAAA) and mapped the parts of the reads upstream of the polyadenine repeats to the AAV2 genome. Note, that AAV2 contains no polyadenine repeat longer than 6. Mappings with quality below 20 were discarded.

Promoter prediction and ORF finder. Promoters on the AAV2 wild-type genome were predicted by the Neural Network Promoter Prediction platform (NNPP) of the BDGP (http://www.fruitfly.org/seq_tools/promoter.html) (24). For the translation of alternative splice variants into potential new proteins an open reading frame (ORF) finder (NCBI) was used (http://www.ncbi.nlm.nih.gov/gorf/gorf.html).

Luciferase assay and Western blot analysis. Luciferase assays and Western blot analysis were performed as described previously (16). Mouse monoclonal antibodies (MAbs) and rabbit polyclonal antibodies used for immunoblot analysis were anti-Rep MAb 303.9 (1:10; Progen), anti-VP MAb B1 (1:10; Progen), anti-VP1 MAb A1 (1:10; Progen), anti-VP1/2 MAb A69 (1:10; Progen) or rabbit and mouse anti-Flag (1:1,000; Sigma), respectively, followed by reaction with horseradish peroxidase-conjugated secondary antibodies and ECL detection.

Extraction of viral DNA and Southern hybridization. Extraction of viral DNA by a modified Hirt procedure and Southern blot analysis were performed as described previously (16). Samples of DpnI-digested Hirt DNA (6 μg) were analyzed on Southern blots for AAV replication with a 1.6-kb fragment from the *cap* region of pTAV2-0 as probe labeled either with [^{32}P]dCTP for radioactive detection or with Biotin-11-dUTP for nonradioactive detection.

Northern hybridization with oligonucleotide probes. Northern blot hybridization of total or poly(A)$^+$ selected RNA samples was performed with a 24-nt oligonucleotide spanning 12 nt upstream and downstream of the AAV2 527:2228 splice junction (SDA1, see Table 1) labeled with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP. Hybridization was performed for 4 h at 68°C in hybridization solution (0.5% [wt/vol] sodium dodecyl sulfate [SDS], 6× SSC, 50 mM HEPES [pH 7.8], 150 μg/ml salmon sperm DNA, and 5× Denhardt solution). The filters were washed four times in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS at room temperature for 5 min and subsequently twice with 6× SSC–0.1% SDS at 68°C for 30 min.

RNase protection assay. RNase protection assays were performed with the RPA III kit (Ambion) essentially as described by the supplier. Two AAV2 sense RNA probes (nt 170 to 250 and nt 170 to 340) were radiolabeled with [α-^{32}P]UTP by *in vitro* transcription with T7 polymerase (Riboprobe system; Promega) using XbaI-linearized pBluescript vector containing the corresponding AAV nucleotides as the templates. The *in vitro*-transcribed RPA probes were gel purified on an 8% PAA–8 M urea denaturing gel. A total of 3×10^4 cpm of purified radiolabeled probe were hybridized with 20 μg of total RNA, followed by RNase A/RNase T_1 treatment according to the manufacturer's protocol with an additional final ethanol precipitation step. The protected RNA fragments were electrophoresed on an 8% PAA–8 M urea gel and subsequently fixed in a solution containing 20% ethanol and 5% acetic acid for 30 min before vacuum drying for 2 h at 80°C. Gels were exposed to a storage phosphor screen and scanned by a Fujifilm FLA-3000 imager.

RACE. Rapid amplification of cDNA ends (RACE) was conducted using the 5'RACE system (version 2.0; Invitrogen) according to the manufacturer's protocol. First-strand cDNA synthesis was performed on total RNA from AAV2/Ad2-coinfected HeLa cells with a primer binding to AAV2 nt 151 to 171 (p5-RACE-GSP1), followed by treatment with RNase H. The cDNA was purified and poly(C)-tailed using terminal deoxynucleotidyltransferase (TdT). After TdT removal, PCR amplification was performed using the TrueStart Hot Start *Taq* DNA polymerase (Thermo Scientific) and a nested primer binding to AAV2 nt 175 to 195 (p5-RACE-GSP2) and the anchor primer supplied by the manufacturer. The PCR products were analyzed and excised from agarose gels and subsequently cloned into pBluescript II SK(+). Randomly picked clones were subjected to DNA sequencing (Eurofins Genomics). For the primer sequences, see Table 1.

RESULTS

RNA-Seq library preparation. To identify the ideal time point for RNA-Seq analysis, time course experiments with total RNAs and proteins extracted from HeLa cells infected with AAV2 in presence or absence of adenovirus (Ad2) were performed. At 27 h postinfection (hpi), AAV2 transcripts were clearly detectable on Northern blots, as were Rep and VP on Western blots. This time point was chosen for total RNA extraction from cells infected with AAV2 in presence or absence of Ad2 and from uninfected and Ad2-infected cells that served as controls. All RNA samples were analyzed by bioanalyzer tests displaying RNA integrity numbers higher than 8.90, which are indicative of highly intact RNA. Total

TABLE 2 RNA-Seq analysis: assignment of the reads to the species from which they originate

Data set	Total no. of reads	No. (%) of mappable reads	No. (%) of human reads	No. (%) of AAV2 reads			No. (%) of Ad2 reads
				Plus strand	Minus strand	Ambiguous	
Cells	23,380,236	22,665,799 (96.94)	22,662,280 (96.93)	1,404 (0.01)	14 (0.00)	0 (0.00)	2,101 (0.01)
AAV2	21,200,968	20,601,318 (97.17)	20,591,869 (97.13)	7,616 (0.04)	241 (0.00)	1 (0.00)	1,591 (0.01)
AAV2 + Ad2	23,047,796	22,167,728 (96.18)	16,327,433 (70.84)	3,061,388 (13.28)	68,099 (0.30)	184 (0.00)	2,710,624 (11.76)
Ad2	26,564,586	25,351,109 (95.43)	10,920,188 (41.11)	1,614 (0.01)	16 (0.00)	0 (0.00)	14,429,291 (54.32)

RNA was rRNA depleted and digested to 100 nucleotide fragments. RNA libraries were sequenced on an Illumina HiSeq 2000 platform.

Illumina RNA-Seq next-generation sequencing. For each set of data, between 21.2 and 26.6 million reads were obtained by Illumina RNA-Seq (Table 2). The vast majority of reads (>95%) could be successfully mapped either to the human genome or to the viral genomes of AAV2 or Ad2. In AAV2-infected cells, in the absence of Ad2 ca. 97% of all RNA reads were assigned to the human genome, which is comparable to uninfected control cells. Only few AAV-specific transcripts (0.04%) could be detected at 27 hpi. In the presence of Ad, AAV-specific transcripts rose to 13.6% of the total read count (Table 2). The majority of AAV reads mapped to the coding plus strand of the AAV2 genome. However, a lower, but significant proportion of reads mapped to the AAV2 minus strand. Parallel to the Ad-induced increase of AAV-specific transcripts, a reduction of the percentage of RNAs mappable to the human genome was found. This was also seen for cells infected with Ad2 alone, suggestive of an Ad-induced effect (Table 2). AAV2/Ad2-coinfected cells showed a 5-fold reduction of Ad2-specific reads compared to cells infected with Ad2 alone. This finding likely reflects the established effect that AAV inhibits adenoviral replication (25–28).

RNA read mapping on the AAV2 plus strand. The population of AAV2-specific RNA reads for each data set was mapped to the wild-type AAV2 genome. Based on this alignment, a coverage map of the AAV2 genome was created that displays the number of reads that map to a specific AAV2 genome position. The genome coverage of the data set from AAV2/Ad2-coinfected cells is based on approximately 3.0 million AAV2-specific reads (Fig. 1A). Steep increases of RNA-Seq read counts are indicative of either transcription initiation or of splicing. The reads assigned to the AAV2 plus strand largely confirmed the well-established transcription initiation at the p5, p19, and p40 promoters. Also in line with previous studies, transcripts initiated at the p40 promoter were more frequent compared to p5- or p19-initiated transcripts in AAV2/Ad2-coinfected cells, whereas in cells infected solely with AAV2 the p5-initiated transcripts were the most prevalent (Fig. 1B). In the absence of Ad, few AAV-specific transcripts were detectable. With the exception of the prevalent p5-initiated transcripts, the overall AAV transcription profile was very similar to that in the presence of Ad, although at a >200-fold reduced level. Additionally, occasional reads were detected upstream of the p5 promoter on the AAV plus strand in the absence or presence of Ad2 (Fig. 1A and B), a finding indicative of transcription initiated within the left ITR of AAV2, as described previously (11, 29, 30). A single polyadenylation site near the right end of the AAV2 genome was defined before (31). RNA-Seq shows that indeed the vast majority of Ad2-induced AAV2 transcripts terminate by polyadenylation at nt 4450 (Fig. 1C). No other hotspots for polyadenylation were detected.

Analysis of transcription on the AAV minus strand. As stated above, a significant proportion of reads mapped to the AAV2 minus strand in the absence, as well as in the presence of Ad coinfection (Fig. 1A and B). These new transcripts, detected here for the first time, were largely restricted to the AAV region between the left ITR and the p5 promoter. Their abundance was about one-third compared to p5-initiated transcripts on the AAV plus strand (Fig. 1A). To obtain hints, where transcription on the AAV2 minus strand was initiated we first searched for candidate upstream promoters. *In silico* promoter prediction (NNPP) (24) under stringent conditions with a cutoff score above 0.9 was applied to both AAV strands. While the well-characterized AAV2 p5 and p40 promoters could be readily detected with cutoff scores of 0.95 and 1.0, respectively, the p19 promoter was only detected when lowering the cutoff to a score of 0.53. On the AAV2 minus strand three candidate promoters were detected with cutoff scores above 0.9. Two of these are located near the left ITR, with putative transcription start sites at nt 230 (score 0.93) and nt 305 (score 0.94), respectively (Fig. 1D). These sites are in agreement with the assumed start sites for AAV2 minus strand transcription (Fig. 1A). The activity of the candidate promoters was tested using reporter gene constructs. Various AAV2 fragments comprising the p5 promoter region were cloned in reverse orientation upstream of the luciferase gene. Ad2-infected HeLa cells were transfected with the constructs, as depicted in Fig. 1D. The luciferase activity measured at 48 h posttransfection showed strong promoter activity of AAV(–)320-180-*luc*, the reverse complement of the established AAV p5 promoter AAV(+)190-320-*luc*, which was 7-fold more active. Deletion of nt 180 to 255 in the AAV(–)320-180-*luc* sequence led to an >30-fold drop in promoter activity, whereas an extended version of the p5 reverse complement, AAV(–)586-155-*luc*, was marginally stronger compared to AAV(–)320-180-*luc*. Relative promoter activities in sense and antisense orientation correlated well with the relative proportions of transcripts detected on the AAV plus and minus strands (Fig. 1A).

To more closely map transcription start sites on the AAV minus strand, we used 5'RACE with the gene-specific primer used for first-strand cDNA synthesis binding to AAV nt 151 to 171. The PCR products were cloned and individual representative clones sequenced to identify the respective transcription start site (TSS). The identified TSSs mapped to the region between AAV2 nt 227 to 320 (Fig. 1E). AAV minus strand transcription was further confirmed by RNase protection analysis (RPA) with probes covering AAV2 nt 170 to 250 or nt 170 to 340, respectively (Fig. 1F). With both probes, a rather heterogeneous population of AAV-specific protected RNA fragments in the size range of about 40 to 80 nt was

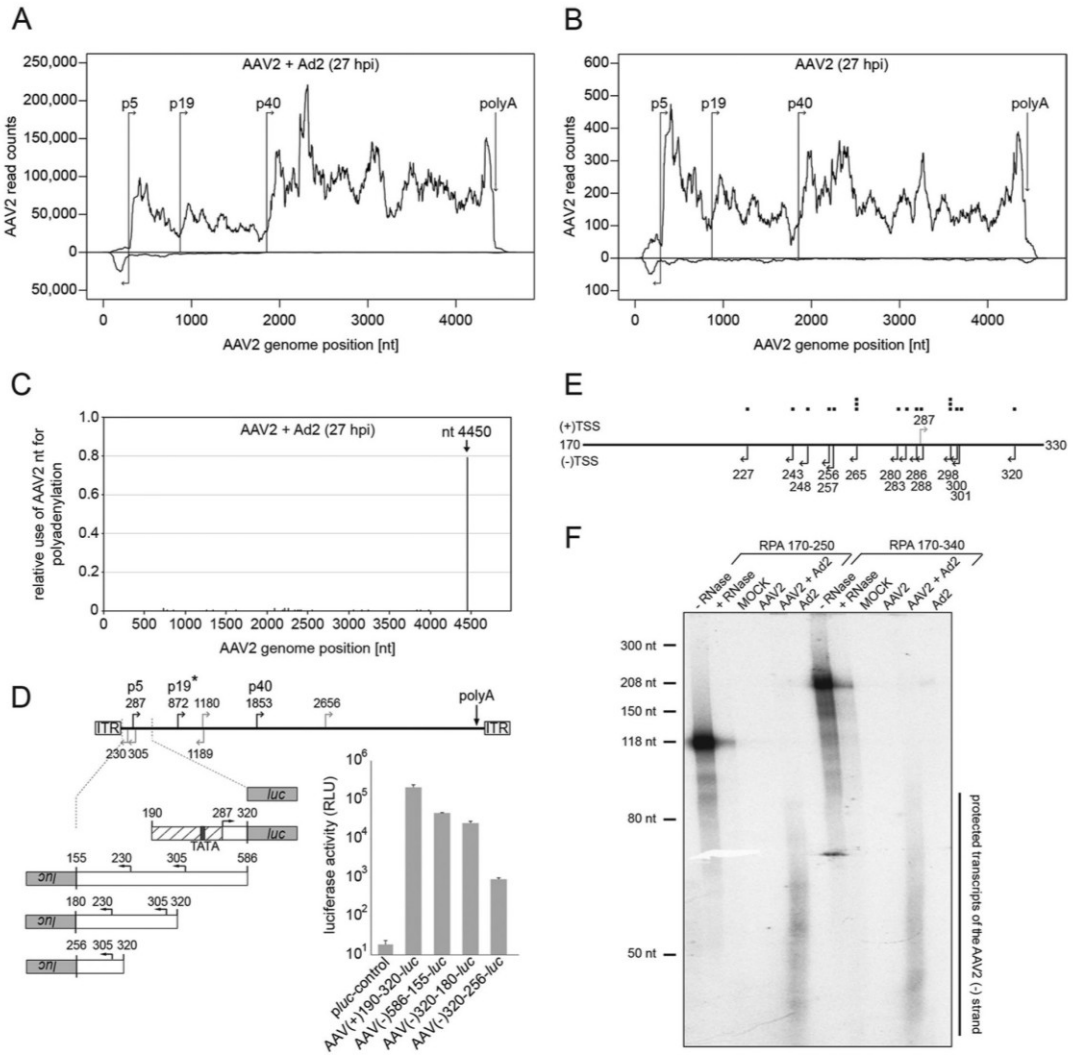


FIG 1 RNA-Seq: mapping and transcription analysis on both strands of the AAV2 genome. (A) Coverage map of AAV2-specific RNA-Seq reads to the AAV2 genome of AAV2/Ad2-coinfected cells at 27 hpi. The AAV2 genome scale is displayed in the center. Reads mapped to the plus strand are presented above, and reads mapped to the minus strand are presented below. (B) Display as in panel A. RNA-Seq reads of AAV2-infected cells at 27 hpi are presented. Note the different scales of the read counts in panels A and B. (C) Mapping of RNA-Seq reads ending with a poly(A) tail. The genome position of poly(A) addition is displayed relative to the totality of polyadenylated AAV reads in the data set of AAV2/Ad2-coinfected cells. (D) Promoter activity of the AAV2 p5 region in reverse direction. The upper part shows the AAV2 genome with the nucleotide positions of the transcription start sites of the p5, p19, and p40 promoters (black arrows). In addition, putative promoters (gray arrows) originating from a promoter prediction program with a cutoff score above 0.9 are indicated on the plus or minus strand of the AAV genome with the respective nucleotide positions of the TSS. An asterisk (*) highlights the established p19 promoter, detected with a cutoff score below 0.9. The lower part shows the relative luciferase activities of the indicated candidate reverse p5 promoter sequences on the AAV2 minus strand. AAV(+)190-320-*luc* represents the AAV2 p5 promoter in the forward orientation serving as a positive control and empty p*luc* serving as a negative control. The relative luciferase activities of three different experiments were determined in the presence of Ad2 and are depicted as means ± standard deviations (SD). (E) 5'RACE analysis of AAV minus strand transcription. Displayed are the TSSs (arrows) derived from DNA sequences of cloned 5'RACE products from the AAV minus strand (nt 330 to 170). Single TSSs are displayed as black squares above the genome. The p5 promoter TSS at nt 287 (gray arrow) on the AAV plus strand is shown for orientation. (F) An RNase protection assay was performed with *in vitro*-transcribed hybridization probes covering AAV2 nt 170 to 250 (left part) or AAV2 nt 170 to 340 (right part) in a sense orientation for the detection of AAV minus strand transcripts in total RNA from HeLa cells either mock infected, AAV2 infected, Ad2 infected, or AAV2/Ad2 coinfected for 27 h. The undigested probes (–RNase) and the probes digested in the presence of yeast t-RNA (+RNase) are shown as controls. Both probes contain additional non-AAV2 sequences from the pBluescript vector used for T7 polymerase-dependent *in vitro* transcription. The maximal protected AAV2 regions are 81 nt for probe RPA 170-250, and 171 nt for probe RPA 170-340, respectively.

TABLE 3 Summary of AAV2 splice events by RNA-Seq analysis: AAV2 + Ad2

Splice donor	Read count	% not spliced	% spliced with acceptor[a]							
			807	985	1437	1635	2201	2228	4138	Others
527	34,472	93.64	**0.53**	**0.33**	0.07	0.11	0.09	**0.50**	**4.21**	0.52
988	26,404	95.05			**0.17**	**0.56**	0.23	**0.89**	**2.45**	0.65
1906	35,746	72.25					**3.57**	**23.42**	**0.36**	0.40
3184	50,575	99.30							**0.51**	0.19

[a] The displayed values give the percentages for the splice donors joined to the various splice acceptors. Values in boldface are splice variants that were validated by RT-PCR.

detected specifically in AAV2/Ad2-coinfected cells (Fig. 1F) but not in mock- or Ad-infected cells.

Analysis of AAV2 splicing. Spliced AAV transcripts were identified by RNA-Seq analysis as AAV reads joined to distant, defined parts of the AAV genome. Bioinformatic tools with stringent filters were utilized to exclude background noise. Only splice events above a critical abundance were investigated further. Reads covering the known AAV2 splice donor site at position 1906 were highly abundant, with >35,000 reads in the data set for AAV2/Ad2-coinfected cells. In ca. 72% of those reads, no splice event was detectable (Table 3). In about 23% of the reads, the splice donor site was joined to the described major splice acceptor site (nt 2228), and 3.5% of the reads were fused to the minor splice acceptor site (nt 2201), as expected from previous studies (8). Less frequently used splice donor sites were identified at nt 527, 988, and 3184. In addition, occasional, alternative AAV splice acceptor sites were detected mostly at nt 4138 (Table 3).

The existence of alternative splice products was verified by RT-PCR with sets of primers whose binding sites were located upstream of the respective splice donor and downstream of the corresponding splice acceptor sites (Fig. 2). Additionally, AAV mutants were generated with an inactivated splice donor site at nt 527 (Fig. 2B, D1 and D2 mutants) or splice acceptor site at nt 4138 (Fig. 2D, A1 mutant), respectively. The described splice products 1906:2201/2228 were detected by RT-PCR, as shown for AAV wild-type and mutant A1 after coinfection with Ad2 (Fig. 2A). The 527:2228 splice variant was clearly detected with the AAV wild type but, as expected, not with the D1 and D2 mutants (Fig. 2B). Northern blots with total and with poly(A)$^+$-selected RNA from AAV2/Ad2-coinfected HeLa cells were probed with an oligonucleotide spanning the 527:2228 splice junction. A single band of about 2.4 kb was detected (Fig. 2C) corresponding to a p5-initiated spliced mRNA that extends to the common poly(A)$^+$ site (theoretical size, 2,436 nt). The splice events involving the splice acceptor site at nt 4138 leading to the splice products 527:4138, 988:4138, and 3184:4138 were detected by RT-PCR for the AAV2 wild type but, as expected, not for AAV mutant A1 (Fig. 2D to F). Notably, splice product 527:4138 had already been faintly detected in the absence of Ad2 infection, but none of the other splice products had (Fig. 2D). Further splice variants with lower frequencies in the RNA-Seq screen were also confirmed by RT-PCR (Table 3, values in boldface).

Involvement of the novel splice variants in the AAV replication cycle. To address the functional significance of the newly identified splice variants for the AAV2 replication cycle, the AAV mutants for the splice donor site at nt 527 (D1 and D2) and the splice acceptor site at nt 4138 (A1) were assayed for Rep and Cap expression and for AAV DNA replication. In the D1 mutant the absolutely conserved G nucleotide at the 5' end of the intron had been abolished, at the expense of a conservative exchange of valine to isoleucine in Rep78/68, whereas the D2 mutant maintains the amino acid sequence of Rep78/68 (Fig. 2B). On Western blots, mutant D1 showed strongly reduced Rep78 and capsid protein expression and some reduction of Rep52, whereas mutant D2 showed a less-pronounced effect on Rep, and VP expression was indistinguishable from that of AAV wild type (Fig. 3A). Also, AAV DNA replication was reduced more extensively for mutant D1 than D2 (Fig. 3B). In line with these findings, the titers of infectious AAV were reduced by 3 to 4 logs for mutant D1 and by 1 log for mutant D2 compared to AAV wild type (data not shown).

The AAV-A1 mutant, with inactivated splice acceptor site at nt 4138 and maintained VP amino acid sequence (Fig. 2D) showed markedly reduced Rep and Cap expression early (24 hpi), but not later (48 hpi) after infection (Fig. 3C). Furthermore, in contrast to AAV2 wild type, Rep52 was undetectable for mutant A1 in the absence of Ad2 (Fig. 3C). In line with these findings, AAV DNA replication was reduced, as evidenced by diminished monomeric and dimeric AAV2 replicative forms (Fig. 3D). In summary, the newly described, alternative splice products appear to play some role in the regulation of Ad-induced AAV replication, especially in the early stages of AAV replication.

Prediction and identification of novel AAV2 proteins. Some of the alternatively spliced AAV transcripts could possibly lead to the translation of novel proteins, as predicted *in silico*. Candidate proteins of at least 50 amino acids, initiated by an ATG start codon or alternatively by ACG or CTG, were considered for further analysis. The most abundant splice variants and the predicted proteins are depicted in Fig. 4A. The splice variant 527:2228 could lead to a small protein of 9.1 kDa consisting of the N terminus of Rep78 fused to a short C terminus translated in a different reading frame as the Rep and VP proteins (Fig. 4A). Unfortunately for this predicted protein, no antibody was available, nor was an antibody available for the splice variant 988:2228-derived 26-kDa protein that lacks the central portion of Rep68. However, the presumed translation product of splice variant 527:4138 could be assayed by MAb B1, which is directed against AAV2 VP1-3. Indeed, an 18-kDa protein was detected in the low-molecular-weight range in cells infected with AAV wild type but not with mutant A1 (Fig. 4B). The 15-kDa bands visible in all extracts infected with AAV most likely represent autolytic capsid cleavage products, as described previously (32). The newly identified 18-kDa protein consists of 159 amino acids (Fig. 4C), of which the first 69 amino acids correspond to the N-terminal part of Rep78 and the C-terminal 90 amino acids correspond to VP. As described above, the novel 18-kDa Rep/VP fusion protein might have a function on the regulation of early AAV gene expression (see above).

The predicted protein of 43-kDa translated from splice variant 3184:4138, although containing the epitopes of MAb A1 and of

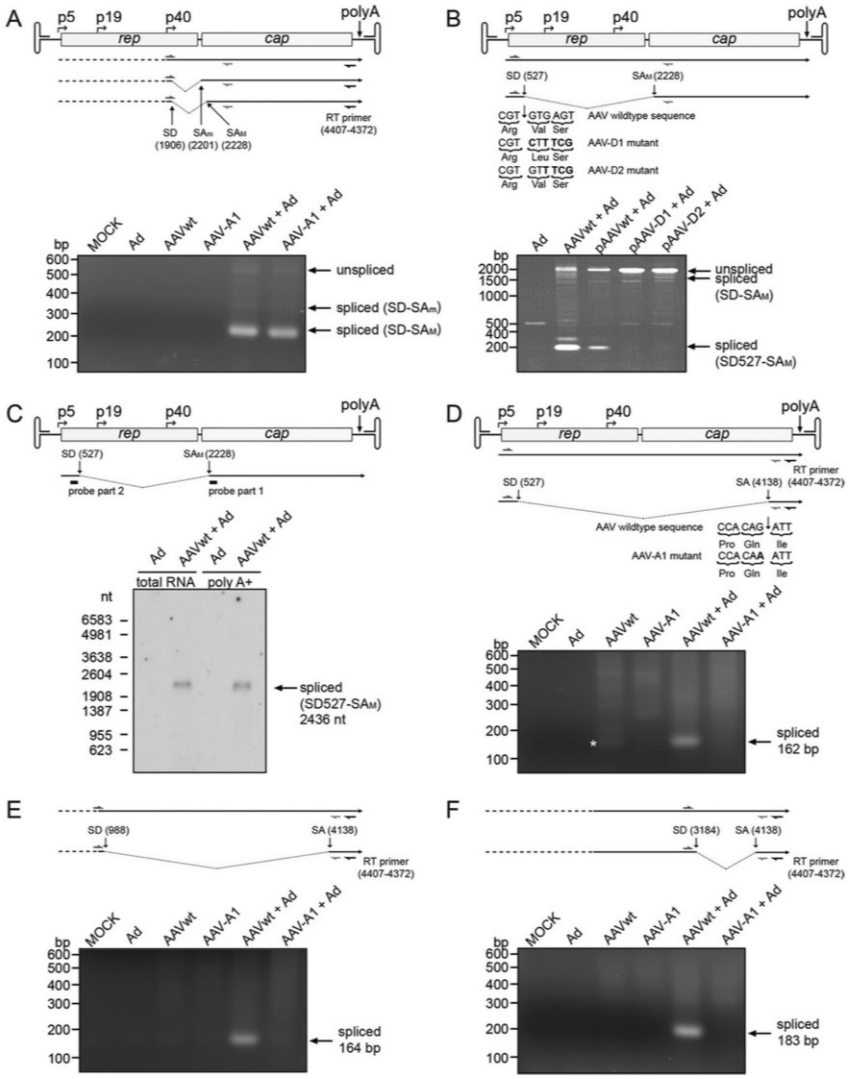


FIG 2 Validation of RNA-Seq-derived AAV2 splice events by RT-PCR and Northern blot analysis. (A) The AAV2 genome with unspliced/spliced transcripts is displayed at the top. The primer for reverse transcription is shown with its binding position. The forward primer (nt 1862 to 1889) and reverse primer (nt 2440 to 2421) are displayed for PCR analysis of 1906:2201/2228 splicing. Below, the results of RT-PCR for AAV wild-type- and mutant A1-infected cells are displayed, separated on an agarose gel. PCR products are indicated by arrows. (B) RT-PCR of splice event 527:2228, performed as in panel A with cells transfected with plasmids for AAV wild-type or mutants D1 or D2. Recoded nucleotide sequences of AAV-D1 and AAV-D2 are displayed below the AAV genome. (C) Northern blot analysis of total and poly(A)$^+$ selected RNA of cells coinfected with AAV2 and Ad2 using a 24-nt oligomer as a probe that spans the AAV2 splice junction 527:2228, as depicted below the AAV genome. The arrow points to the spliced transcripts. (D to F) RT-PCR analysis of additional AAV RNA splice events, performed as in panel A. (D) Forward primer (nt 415 to 434) and reverse primer (nt 4186 to 4161) used for the analysis of 527:4138 splicing. The corresponding mutation of AAV-A1 is displayed below the AAV genome. The asterisk (*) marks a faint band of the splice product in the absence of Ad. (E) Forward primer (nt 874 to 893) and reverse primer (nt 4186 to 4161) for the analysis of 988:4138 splicing. (F) Forward primer (nt 3051 to 3070) and reverse primer (nt 4186 to 4161) for the analysis of 3184:4138 splicing.

MAb A69 could not be identified with either antibody. To verify expression of other predicted proteins, especially from splice variant 988:4138, an in-frame Flag tag was introduced in the AAV wild-type and mutant A1 DNA sequences (Fig. 4E). Neither of the predicted 35- or 13-kDa proteins could be detected on Western blots with two alternative anti-Flag antibodies, which successfully detected the positive control (data not shown). Furthermore, the Flag tag allowed to search for the proposed X protein (18 kDa) (33) derived from the same ORF. Indeed, on very long exposures faint bands of approximately 18 kDa were detected in AAV wild-type and in AAV-A1 transfected cells, a finding consistent with the existence of the X protein (Fig. 4D).

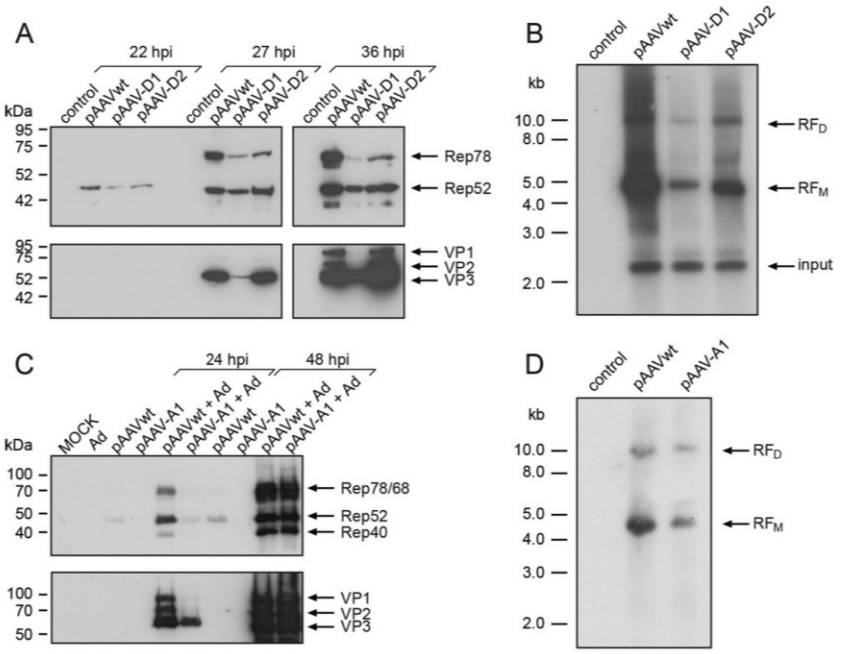


FIG 3 Influence of alternative splicing on AAV replication. (A) Analysis of AAV2 Rep and VP expression in AAV wild type-, splice mutant AAV-D1-, or AAV-D2 (see Fig. 2B)-transfected HeLa cells at the indicated time points after Ad infection. AAV Rep and VP proteins are detected by MAb 303.9 or MAb B1, respectively. (B) Southern blot analysis of Hirt extracts of AAV wild type-, AAV-D1-, or AAV-D2-transfected and Ad-infected HeLa cells. AAV-specific monomeric (RF_M) and dimeric (RF_D) replicative forms, detected by probing with a labeled 1.6-kb *cap*-derived DNA sequence, are indicated by arrows. Input refers to undigested input DNA. (C) Western blot analysis as described in panel A with mutant AAV-A1 (see Fig. 2D). Mock- and Ad-infected cells served as controls. (D) Southern blot analysis as performed for panel B with mutant AAV-A1.

Comparison of the splice sites to other AAV serotypes. The biological relevance of the new splice variants was further evaluated by comparing their conservation in the genomes of alternative AAV serotypes. The genomes of AAV serotypes 1 to 9 were aligned, and the splice sites were compared to the consensus sequence of mammalian splice donor and acceptor sites (34). The well-documented splice donor site at AAV2 nt 1906 is highly conserved among all AAV serotypes but shows minor variations compared to the consensus sequence (Fig. 5A). The splice donor site at nt 527 is similarly conserved among AAV serotypes and is therefore assumed to be functional in these serotypes as well. The splice donor sites at nt 988 and 3184 are more divergent from the consensus sequence. Although the splice donor site nt 988 may be also functional in AAV3 and AAV4, the splice donor site at nt 3184 exists only in AAV2.

The described major splice acceptor site at nt 2228 of AAV2 is highly conserved among all AAV serotypes (Fig. 5B), as is the minor splice acceptor site at nt 2201, although with reduced pyrimidine levels indicative of lower utilization rates. The newly identified splice acceptor site at nt 4138 is also conserved, with the exception of AAV3, AAV4, and AAV5. Another splice acceptor site at nt 807 validated by PCR for splice variant 527:807 is conserved among the AAV serotypes (Fig. 5B). Unfortunately, the predicted protein of 60 kDa is very similar to the molecular mass of Rep68, and we were therefore unable to differentiate it on Western blots.

DISCUSSION

We present here the first NGS analysis of the AAV2 transcriptome during latency and productive AAV replication. Each of the four data sets analyzed by Illumina RNA-Seq gave rise to more than 21 million reads, more than 95% of which could be successfully mapped to the human or viral genomes. The high quality of the raw data allowed a comprehensive and comparative analysis of the transcription profiles of AAV2 with or without Ad2 infection. The high sensitivity of RNA-Seq combined with unbiased RNA detection uncovered previously unrecognized transcripts, alternative splice variants, and a new 18-kDa protein that appears to play a role in early AAV replication. Most importantly, we are the first to document transcription from the AAV minus strand, leading to an update of the current concept of AAV gene expression.

Analysis of the AAV2 transcription profile. As a single-stranded DNA virus AAV is dependent on the synthesis of the complementary strand to generate a double-stranded template for transcription initiation. In the absence of a helper virus, only limited Rep78/68 expression occurs from the p5 promoter (35). This is reflected in the RNA-Seq analysis of AAV2-infected cells in the absence of Ad2, where few AAV transcripts (0.04%) are found, only ~4-fold above background level. Of these, p5 promoter-initiated transcripts are the most abundant RNAs, but transcripts originating from p19 and p40 are also detected. The findings are in line with AAV2 biology, where Rep78/68 was shown to repress

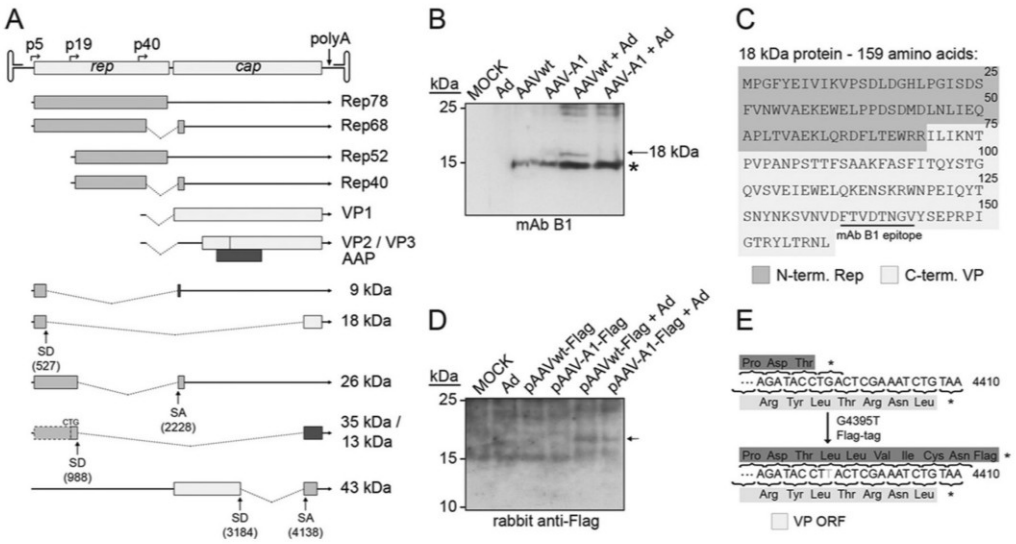


FIG 4 Detection of novel AAV proteins derived from alternative splicing. (A) The AAV2 genome and current transcription/translation map is represented at the top. The different shading patterns indicate alternative reading frames. Below, the newly identified spliced transcripts are displayed with the deduced novel fusion ORFs. The molecular mass of the derived protein candidates is indicated to the right in kilodaltons. (B) Analysis of the predicted protein (18 kDa) expressed from splice variant 527:4138 on a Western blot reacted with MAb B1. AAV-A1 represents the AAV2 mutant at nt 4138 (see Fig. 2D). An asterisk (*) marks the previously described 15- to 17-kDa band, which is indicative of degraded input capsids. (C) Amino acid sequence of the novel 18-kDa protein. The first 69 amino acids correspond to the N-terminal part of Rep78; the last 90 amino acids correspond to the C terminus of the VP proteins, including the epitope for MAb B1. (D) Detection of AAVwt-Flag and AAV-A1-Flag constructs on a Western blot reacted with rabbit anti-Flag upon extensive exposure. The arrow marks a faint band of ~18 kDa consistent with the previously proposed X protein (38). (E) Displayed is the proposed X-gene-derived C terminus of the resulting protein (dark gray) before and after insertion of the Flag tag in comparison to the confirmed capsid (VP) protein sequence (light gray). The asterisk (*) denotes the stop codon in the wild-type sequence.

AAV2 promoter activity, which is viewed as a mechanism to establish and maintain the latent state (35, 36).

In the presence of Ad2, AAV2-specific reads are increased by >200-fold, with a shift of transcription toward p40 expressed genes, which is consistent with previous reports (10). RNA-Seq also detects more-abundant transcription from the p5 promoter than from the p19 promoter, a finding that is at variance with the described 1:3 p5/p19 ratio found by RNase protection-based quantification late in AAV infection (10) and is likely explained by experimental differences. Furthermore, RNA-Seq identified reads mapping to the AAV plus strand upstream of the p5 promoter. These transcripts appear to be derived from the described transcription initiation site (Inr) located within the ITR, observed before for AAV5 and for recombinant AAV2-based vectors (11, 29, 30). In addition, RNA-Seq precisely mapped the previously defined, unique AAV2 polyadenylation site (31). Mammalian mRNAs typically cleave the RNA molecule 10 to 30 nt behind the polyadenylation signal (37). In the case of AAV2, we show that cleavage occurs exactly 21 nt behind the AAUAAA polyadenylation signal at position 4450 (2).

Previously, a candidate promoter (p81) was described on the AAV plus strand near genome position 3800 (38). We were unable to detect this promoter *in silico* under the stringent conditions applied. Furthermore, RNA-Seq did not detect an increase of AAV-specific reads around nt 3800, as would have been expected downstream of an active promoter. However, on Western blots with Flag-tagged constructs a faint 18-kDa protein became visible consistent with the proposed X protein (33). Its very weak expression likely explains the difficulties in confirming its existence by alternative methods but also poses questions as to its *in vivo* significance.

Transcription on the AAV2 minus strand. All AAV2 transcripts described in previous reports were derived from a single DNA strand, referred to as the AAV plus strand. Therefore, the complementary AAV2 minus strand has not been considered transcribed until now (6, 7, 12). RNA-Seq analysis, however, showed that a small proportion of AAV-specific reads from AAV2/Ad2-coinfected cells (0.30%) map to the AAV minus strand. AAV minus strand reads map to a region between the left ITR and the p5 promoter and have an abundance of approximately one-third of the transcripts initiated at the p5 promoter of the AAV2 plus strand. *In silico*, candidate promoters are predicted on the AAV2 minus strand with transcription start sites at nt 230 and 305, respectively. Reporter gene assays revealed that the AAV2 sequence from nt 320 to 180 was sufficient to initiate transcription in minus strand direction. An extended sequence from nt 586 to 155 only marginally enhanced transcriptional activity. Obviously, the p5 promoter region is also active in opposite direction, which likely explains transcription of the AAV2 minus strand. This finding is unique for p5; transcripts in the opposite direction were not detected upstream of p19 or p40.

The finding is reminiscent of a mechanism called divergent transcription (39–42) that is related to transcription initiation in opposite directions from the same promoter. The rather short transcripts initiated at nonuniform start sites that have been described for the upstream direction in divergent transcription are

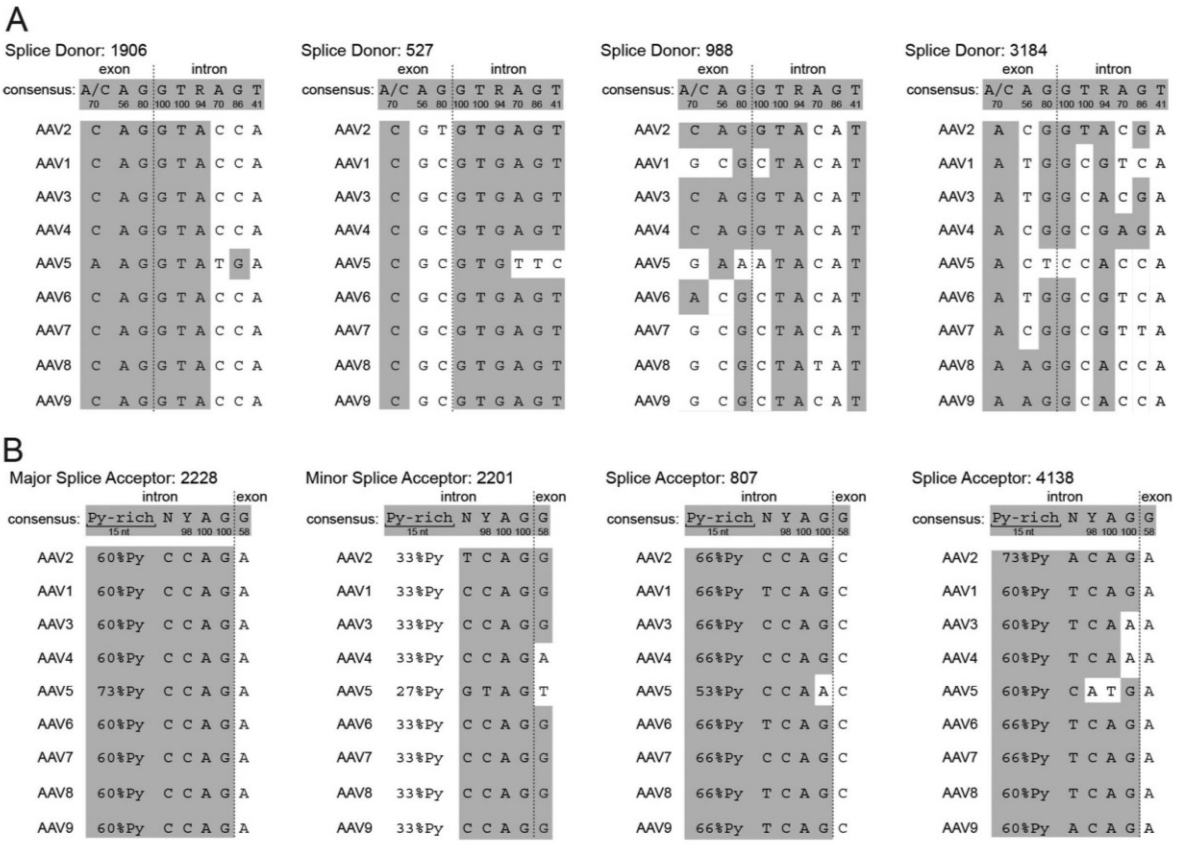


FIG 5 Alignment of splice donor and acceptor sites for AAV serotypes 1 to 9. (A) Splice donor sites at nt positions 1906, −527, −988, and −3184 of AAV2 were aligned to the corresponding sites of AAV serotypes 1 to 9 and compared to the mammalian consensus sequence of splice donor sites (34). Exon-intron boundaries are indicated by dashed lines. Below the consensus sequence the frequency (%) of a particular nucleotide in average splice donor sites is displayed. White areas indicate aberrations to the consensus sequence in the given AAV serotype. (B) Splice acceptor sites at nt positions 2228, −2201, −807, and −4138 of AAV2 were aligned to the corresponding sites of AAV serotypes 1 to 9 and compared to the mammalian consensus sequence of splice acceptor sites. The display is as described for panel A. Py, pyrimidine.

in line with the heterogeneous population of AAV2 minus strand transcripts that we identified both by 5'RACE and by RNase protection. One assumed role for divergent transcription is to facilitate DNA unwinding, so that transcription initiation rates and expression of the downstream gene are enhanced (43). Furthermore, the short antisense transcripts are believed to regulate the transcription of downstream genes by affecting the chromatin and thereby modulating promoter activity (44). An analogous mechanism may be advantageous for the regulation of AAV p5-initiated transcripts, since these transcripts are the first AAV2 genes to be transcribed both in the presence and in the absence of adenovirus. It will be interesting to see whether divergent transcription at p5 serves as a key for the switch between the latent and the lytic AAV replication cycles.

AAV splicing. Splicing of AAV2 transcripts is essential for the generation of Rep68 and Rep40 and for the correct stoichiometry of VP1, VP2, and VP3. The corresponding splice sites have been known for nearly 2 decades and are located at AAV2 nt 1906 (the splice donor) and nt 2201 and 2228 (the splice acceptor) (7, 8). The majority of splice products detected by RNA-Seq correspond to these sites. The ratio of usage for the established minor or major splice acceptor site, respectively, joined to the splice donor site at nt 1906, was roughly 1:7, a finding consistent with previous reports (31). The proportion of unspliced versus spliced transcripts at these sites is higher than previously reported (10) and is likely explained by the use of total cellular RNA for RNA-Seq compared to cytoplasmic RNA in previous studies. In addition, splicing enhances the export of mRNAs (45). Therefore, the total cellular RNA used here may be enriched for unspliced transcripts still located in the nucleus.

The first indication of additional splice products came from the genome coverage of total AAV2-specific reads. The read count at the far right end of the AAV genome, downstream of nt 4000, was significantly increased, despite the absence of known or predicted promoters in this region. Of several newly detected splice donor and acceptor sites utilized at variable efficiencies, the splice products involving a splice donor site at nt 527 and a splice acceptor site at nt 4138 were particularly noteworthy. Reads covering the 527:4138 splice variant were even more abundant than the known splice event 1906:2201, allowing expression of the newly

described 18-kDa Rep/VP protein. Mutants involving these particular splice donor (AAV2 nt 527) or splice acceptor (AAV2 nt 4138) sites showed reduced expression levels of Rep and Cap and also some decrease in DNA replication.

Unlike the AAVs, most parvovirus genera exhibit more than one intronic sequence to generate alternative transcripts. The autonomous parvovirus minute virus of mice (MVM) possesses a large intron on the left-hand side of the genome, in addition to a short intron analogous to the single intron of the AAVs (46–48). For MVM, splicing of the large intron generates a transcript, which is translated into the nonstructural protein NS2. NS2 is required for viral growth in its natural murine host, while it is dispensable in many nonmurine cells (49). Furthermore, the human parvovirus B19 contains two introns and the simian parvovirus (SPV) contains three introns, which leads to various spliced transcripts with partially unknown protein function (50, 51). The newly identified AAV2 splice donor site at nt 527 appears to be analogous to the described splice donor site of MVM (nt 514) and the erythroviruses B19 (nt 531) and SPV (nt 279). All of these are located close to the left ends of their genomes. Similarly, the newly identified splice acceptor site at AAV2 nt 4138 may have counterparts in B19 (nt 4883) and SPV (nt 4638) at the right ends of the respective genomes. For parvovirus B19, a multiply spliced transcript involving the above-described splice sites is translated to a small protein of 11 kDa, which appears to increase viral infectivity *in vivo* (52) and may play a role in apoptosis and cell signaling pathways (53, 54). In cell culture the novel 18-kDa Rep/VP protein only had a minor effect on AAV replication. However, analogous to NS2 of MVM, the 18-kDa protein may have a more pronounced role for the AAV2 life cycle *in vivo*. Unfortunately, *in vivo* models of wild-type AAV replication do not exist, and in humans and primates the site(s) and mechanisms of AAV latency and productive replication are largely unknown. It will be challenging to further unravel possible functions of the novel RNAs and derived proteins during the AAV life cycle in its natural host.

ACKNOWLEDGMENTS

We thank Eva-Maria Hammer of the Heilbronn lab for help with cells and virus propagation and Madlen Sohn of the Chen lab for assistance during RNA library preparation.

C.S. acknowledges the receipt of a Charité-funded Ph.D. fellowship.

FUNDING INFORMATION

Charité Universitätsmedizin Berlin provided funding to Catrin Stutika.

REFERENCES

1. **Blacklow NR, Hoggan MD, Rowe WP.** 1967. Isolation of adenovirus-associated viruses from man. Proc Natl Acad Sci U S A **58:**1410–1415. http://dx.doi.org/10.1073/pnas.58.4.1410.
2. **Srivastava A, Lusby EW, Berns KI.** 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol **45:**555–564.
3. **Sonntag F, Schmidt K, Kleinschmidt JA.** 2010. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc Natl Acad Sci U S A **107:**10220–10225. http://dx.doi.org/10.1073/pnas.1001673107.
4. **Green MR, Roeder RG.** 1980. Definition of a novel promoter for the major adenovirus-associated virus mRNA. Cell **22:**231–242. http://dx.doi.org/10.1016/0092-8674(80)90171-3.
5. **Lusby EW, Berns KI.** 1982. Mapping of the 5' termini of two adeno-associated virus 2 RNAs in the left half of the genome. J Virol **41:**518–526.
6. **Jay FT, de la Maza LM, Carter BJ.** 1979. Parvovirus RNA transcripts containing sequences not present in mature mRNA: a method for isolation of putative mRNA precursor sequences. Proc Natl Acad Sci U S A **76:**625–629. http://dx.doi.org/10.1073/pnas.76.2.625.
7. **Laughlin CA, Westphal H, Carter BJ.** 1979. Spliced adenovirus-associated virus RNA. Proc Natl Acad Sci U S A **76:**5567–5571. http://dx.doi.org/10.1073/pnas.76.11.5567.
8. **Trempe JP, Carter BJ.** 1988. Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. J Virol **62:**3356–3363.
9. **Marcus CJ, Laughlin CA, Carter BJ.** 1981. Adeno-associated virus RNA transcription in vivo. Eur J Biochem **121:**147–154. http://dx.doi.org/10.1111/j.1432-1033.1981.tb06443.x.
10. **Mouw MB, Pintel DJ.** 2000. Adeno-associated virus RNAs appear in a temporal order and their splicing is stimulated during coinfection with adenovirus. J Virol **74:**9878–9888. http://dx.doi.org/10.1128/JVI.74.21.9878-9888.2000.
11. **Qiu J, Nayak R, Tullis GE, Pintel DJ.** 2002. Characterization of the transcription profile of adeno-associated virus type 5 reveals a number of unique features compared to previously characterized adeno-associated viruses. J Virol **76:**12435–12447. http://dx.doi.org/10.1128/JVI.76.24.12435-12447.2002.
12. **Green MR, Straus SE, Roeder RG.** 1980. Transcripts of the adenovirus-associated virus genome: multiple polyadenylated RNAs including a potential primary transcript. J Virol **35:**560–565.
13. **Green MR, Roeder RG.** 1980. Transcripts of the adeno-associated virus genome: mapping of the major RNAs. J Virol **36:**79–92.
14. **Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B.** 2012. Landscape of transcription in human cells. Nature **489:**101–108. http://dx.doi.org/10.1038/nature11233.
15. **Zhao H, Chen M, Pettersson U.** 2014. A new look at adenovirus splicing. Virology **456-457:** 329–341. http://dx.doi.org/10.1016/j.virol.2014.04.006.
16. **Winter K, von Kietzell K, Heilbronn R, Pozzuto T, Fechner H, Weger S.** 2012. Roles of E4orf6 and VA I RNA in adenovirus-mediated stimulation of human parvovirus B19 DNA replication and structural gene expression. J Virol **86:**5099–5109. http://dx.doi.org/10.1128/JVI.06991-11.
17. **Heilbronn R, Bürkle A, Stephan S, zur Hausen H.** 1990. The adeno-associated virus *rep* gene suppresses herpes simplex virus-induced DNA-amplification. J Virol **64:**3012–3018.
18. **Hüser D, Gogol-Doring A, Chen W, Heilbronn R.** 2014. 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 **88:**11253–11263. http://dx.doi.org/10.1128/JVI.01356-14.
19. **Mietzsch M, Broecker F, Reinhardt A, Seeberger PH, Heilbronn R.** 2014. Differential adeno-associated virus serotype-specific interaction patterns with synthetic heparins and other glycans. J Virol **88:**2991–3003. http://dx.doi.org/10.1128/JVI.03371-13.
20. **Weindler FW, Heilbronn R.** 1991. A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. J Virol **65:**2476–2483.
21. **Stutika C, Hüser D, Weger S, Rutz N, Hessler M, Heilbronn R.** 2015. Definition of herpes simplex virus helper functions for the replication of adeno-associated virus type 5. J Gen Virol **96:**840–850. http://dx.doi.org/10.1099/vir.0.000034.
22. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods **9:**357–359. http://dx.doi.org/10.1038/nmeth.1923.
23. **Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermuller J.** 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput Biol **5:**e1000502. http://dx.doi.org/10.1371/journal.pcbi.1000502.
24. **Reese MG.** 2001. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. Comput Chem **26:**51–56. http://dx.doi.org/10.1016/S0097-8485(01)00099-7.
25. **Casto BC, Atchison RW, Hammon WM.** 1967. Studies on the relationship between adeno-associated virus type I (AAV-1) and adenoviruses. I. Replication of AAV-1 in certain cell cultures and its effect on helper adenovirus. Virology **32:**52–59.
26. **Casto BC, Armstrong JA, Atchison RW, Hammon WM.** 1967. Studies on the relationship between adeno-associated virus type 1 (AAV-1) and

adenoviruses. II. Inhibition of adenovirus plaques by AAV; its nature and specificity. Virology **33:**452–458.
27. **Weitzman MD, Fisher KJ, Wilson JM.** 1996. Recruitment of wild-type and recombinant adeno-associated virus into adenovirus replication centers. J Virol **70:**1845–1854.
28. **Timpe JM, Verrill KC, Trempe JP.** 2006. Effects of adeno-associated virus on adenovirus replication and gene expression during coinfection. J Virol **80:**7807–7815. http://dx.doi.org/10.1128/JVI.00198-06.
29. **Flotte TR, Afione SA, Solow R, Drumm ML, Markakis D, Guggino WB, Zeitlin PL, Carter BJ.** 1993. Expression of the cystic fibrosis transmembrane conductance regulator from a novel adeno-associated virus promoter. J Biol Chem **268:**3781–3790.
30. **Haberman RP, McCown TJ, Samulski RJ.** 2000. Novel transcriptional regulatory signals in the adeno-associated virus terminal repeat A/D junction element. J Virol **74:**8732–8739. http://dx.doi.org/10.1128/JVI.74.18.8732-8739.2000.
31. **Qiu J, Pintel D.** 2008. Processing of adeno-associated virus RNA. Front Biosci **13:**3101–3115. http://dx.doi.org/10.2741/2912.
32. **Salganik M, Venkatakrishnan B, Bennett A, Lins B, Yarbrough J, Muzyczka N, Agbandje-McKenna M, McKenna R.** 2012. Evidence for pH-dependent protease activity in the adeno-associated virus capsid. J Virol **86:**11877–11885. http://dx.doi.org/10.1128/JVI.01717-12.
33. **Cao M, You H, Hermonat PL.** 2014. The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication. PLoS One **9:**e104596. http://dx.doi.org/10.1371/journal.pone.0104596.
34. **Zhang MQ.** 1998. Statistical features of human exons and their flanking regions. Hum Mol Genet **7:**919–932. http://dx.doi.org/10.1093/hmg/7.5.919.
35. **Pereira DJ, McCarty DM, Muzyczka N.** 1997. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. J Virol **71:**1079–1088.
36. **Beaton A, Palumbo P, Berns KI.** 1989. Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein. J Virol **63:**4450–4454.
37. **Proudfoot NJ.** 2011. Ending the message: poly(A) signals then and now. Genes Dev **25:**1770–1782. http://dx.doi.org/10.1101/gad.17268411.
38. **Hermonat PL, Santin AD, De Greve J, De Rijcke M, Bishop BM, Han L, Mane M, Kokorina N.** 1999. Chromosomal latency and expression at map unit 96 of a wild-type plus adeno-associated virus (AAV)/Neo vector and identification of p81, a new AAV transcriptional promoter. J Hum Virol **2:**359–368.
39. **Seila AC, Core LJ, Lis JT, Sharp PA.** 2009. Divergent transcription: a new feature of active promoters. Cell Cycle **8:**2557–2564. http://dx.doi.org/10.4161/cc.8.16.9305.
40. **Flynn RA, Chang HY.** 2012. Active chromatin and noncoding RNAs: an intimate relationship. Curr Opin Genet Dev **22:**172–178. http://dx.doi.org/10.1016/j.gde.2011.11.002.
41. **Duttke SH, Lacadie SA, Ibrahim MM, Glass CK, Corcoran DL, Benner C, Heinz S, Kadonaga JT, Ohler U.** 2015. Human promoters are intrinsically directional. Mol Cell **57:**674–684. http://dx.doi.org/10.1016/j.molcel.2014.12.029.
42. **Scruggs BS, Gilchrist DA, Nechaev S, Muse GW, Burkholder A, Fargo DC, Adelman K.** 2015. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. Mol Cell **58:**1101–1112. http://dx.doi.org/10.1016/j.molcel.2015.04.006.
43. **Kouzine F, Sanford S, Elisha-Feil Z, Levens D.** 2008. The functional response of upstream DNA to dynamic supercoiling in vivo. Nat Struct Mol Biol **15:**146–154. http://dx.doi.org/10.1038/nsmb.1372.
44. **Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA.** 2008. Divergent transcription from active promoters. Science **322:**1849–1851. http://dx.doi.org/10.1126/science.1162253.
45. **Valencia P, Dias AP, Reed R.** 2008. Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc Natl Acad Sci U S A **105:**3386–3391. http://dx.doi.org/10.1073/pnas.0800250105.
46. **Pintel D, Dadachanji D, Astell CR, Ward DC.** 1983. The genome of minute virus of mice, an autonomous parvovirus, encodes two overlapping transcription units. Nucleic Acids Res **11:**1019–1038. http://dx.doi.org/10.1093/nar/11.4.1019.
47. **Jongeneel CV, Sahli R, McMaster GK, Hirt B.** 1986. A precise map of splice junctions in the mRNAs of minute virus of mice, an autonomous parvovirus. J Virol **59:**564–573.
48. **Haut DD, Pintel DJ.** 1998. Intron definition is required for excision of the minute virus of mice small intron and definition of the upstream exon. J Virol **72:**1834–1843.
49. **Ruiz Z, D'Abramo A, Jr, Tattersall P.** 2006. Differential roles for the C-terminal hexapeptide domains of NS2 splice variants during MVM infection of murine cells. Virology **349:**382–395. http://dx.doi.org/10.1016/j.virol.2006.01.039.
50. **Ozawa K, Ayub J, Hao YS, Kurtzman G, Shimada T, Young N.** 1987. Novel transcription map for the B19 (human) pathogenic parvovirus. J Virol **61:**2395–2406.
51. **Liu Z, Qiu J, Cheng F, Chu Y, Yoto Y, O'Sullivan MG, Brown KE, Pintel DJ.** 2004. Comparison of the transcription profile of simian parvovirus with that of the human erythrovirus B19 reveals a number of unique features. J Virol **78:**12929–12939. http://dx.doi.org/10.1128/JVI.78.23.12929-12939.2004.
52. **Zhi N, Mills IP, Lu J, Wong S, Filippone C, Brown KE.** 2006. Molecular and functional analyses of a human parvovirus B19 infectious clone demonstrates essential roles for NS1, VP1, and the 11-kilodalton protein in virus replication and infectivity. J Virol **80:**5941–5950. http://dx.doi.org/10.1128/JVI.02430-05.
53. **Chen AY, Zhang EY, Guan W, Cheng F, Kleiboeker S, Yankee TM, Qiu J.** 2010. The small 11 kDa nonstructural protein of human parvovirus B19 plays a key role in inducing apoptosis during B19 virus infection of primary erythroid progenitor cells. Blood **115:**1070–1080. http://dx.doi.org/10.1182/blood-2009-04-215756.
54. **Fan MM, Tamburic L, Shippam-Brett C, Zagrodney DB, Astell CR.** 2001. The small 11-kDa protein from B19 parvovirus binds growth factor receptor-binding protein 2 in vitro in a Src homology 3 domain/ligand-dependent manner. Virology **291:**285–291. http://dx.doi.org/10.1006/viro.2001.1217.

Supplementary Material, in proof

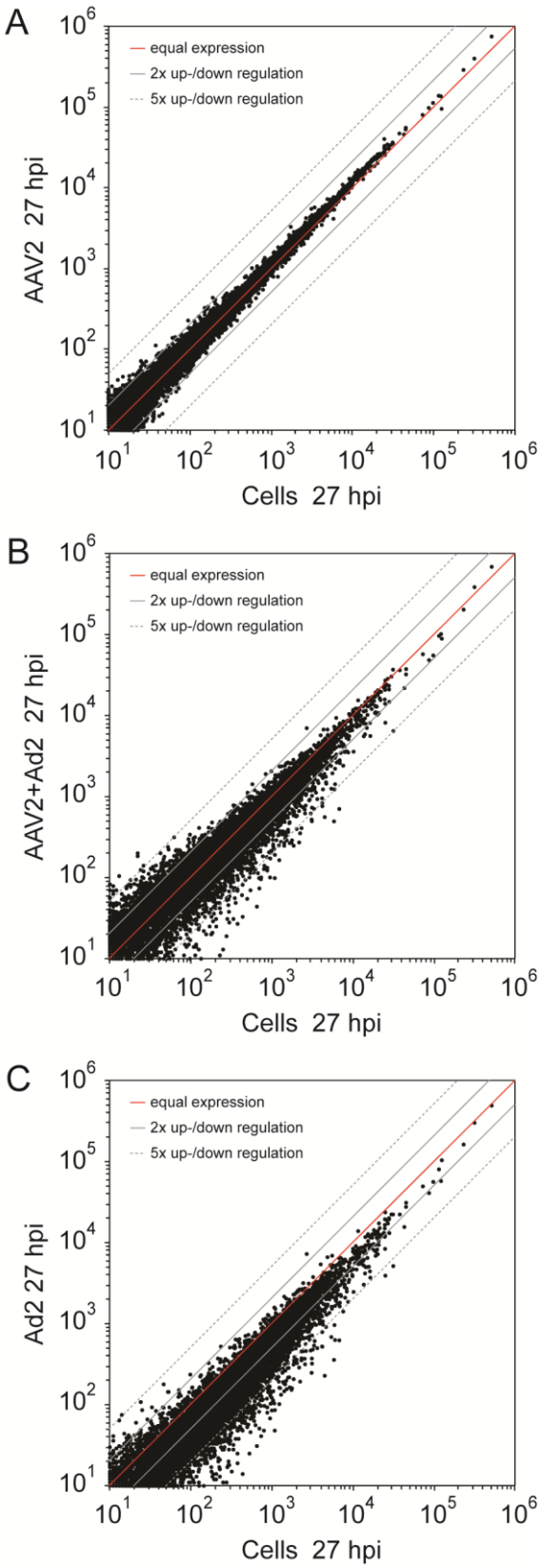


Fig. S1: Analysis of the effect of AAV2 infection on host cell transcription.

(A) Comparison of the expression levels of 12,526 RefSeq genes in AAV2 infected versus uninfected cells 27 hpi diagrammed in read counts. The figure shows only genes onto which at least 10 reads per data set were uniquely mapped. Dots near the red line represent equal gene expression levels in both data sets. Fold changes of 2 and 5 are marked by solid and dashed lines, respectively. **(B-C)** Comparison of the expression levels of 12,526 RefSeq genes in (B) AAV2/Ad2 co-infected versus uninfected cells 27 hpi or (C) Ad2 infected versus uninfected cells 27 hpi, displayed as in (A).

2.2 Comprehensive Small RNA-Seq of Adeno-Associated Virus (AAV)-Infected Human Cells Detects Patterns of Novel, Non-Coding AAV RNAs in the Absence of Cellular miRNA Regulation

Authors: Catrin Stutika, Mario Mietzsch, Andreas Gogol-Döring, Stefan Weger, Madlen Sohn, Wei Chen, Regine Heilbronn

Year: 2016

Journal: PLoS One 11 (9): e0161454

NCBI Sequence Read Archive (SRA) accession: PRJNA338126

2.2.1 Contribution to the Publication

This study is the first to search and describe the existence and potential function of small non-coding RNAs encoded by AAV.

For this publication I initially generated a set of high-quality DNase-digested RNA samples from AAV2 infected cells and control cells, respectively. From these I prepared multiple sets of small RNA libraries. The small RNA next generation sequencing was carried out by Wei Chen's group, conducted by Madlen Sohn at the Berlin Institute for Medical Systems Biology at the Max Delbrück Center in Berlin-Buch, Germany. The bioinformatic analysis of the sequencing raw data was performed by Andreas Gogol-Döring at the Technische Hochschule Mittelhessen in Gießen, Germany, in the course of an ongoing collaboration between my mentor Regine Heilbronn and him. The analysis of the resulting sequencing data including the calculation of read frequencies, alignment of small RNA reads to the different genomes and the calculation of read length distributions was performed by me. I further analyzed all AAV small RNA hotspots on Northern blots and performed co-immunoprecipitation (Co-IP) of the most promising candidates. I generated the tables and figures of this study and their corresponding descriptions. Mario Mietzsch developed the model illustrated in Fig. 7A. Furthermore, Andreas Gogol-Döring added Table S3 for supplementary material. The initial draft of the manuscript was written by me and Mario Mietzsch. It was revised and finalized by Regine Heilbronn, Stefan Weger and myself.

2.2.2 Article

DOI: http://dx.doi.org/10.1371/journal.pone.0161454

PLOS | ONE

RESEARCH ARTICLE

Comprehensive Small RNA-Seq of Adeno-Associated Virus (AAV)-Infected Human Cells Detects Patterns of Novel, Non-Coding AAV RNAs in the Absence of Cellular miRNA Regulation

Catrin Stutika[1], Mario Mietzsch[1¤], Andreas Gogol-Döring[2], Stefan Weger[1], Madlen Sohn[3], Wei Chen[3], Regine Heilbronn[1]*

1 Charité Medical School, Campus Benjamin Franklin, Institute of Virology, Berlin, Germany, **2** Technische Hochschule Mittelhessen, Gießen, Germany, **3** Max-Delbrück-Centrum für Molekulare Medizin, Berlin Institute for Medical Systems Biology, Laboratory for Functional Genomics and Systems Biology, Berlin, Germany

¤ Current address: Department of Biochemistry and Molecular Biology, Center for Structural Biology, McKnight Brain Institute, University of Florida, Gainesville, Florida, United States of America
* regine.heilbronn@charite.de

OPEN ACCESS

Citation: Stutika C, Mietzsch M, Gogol-Döring A, Weger S, Sohn M, Chen W, et al. (2016) Comprehensive Small RNA-Seq of Adeno-Associated Virus (AAV)-Infected Human Cells Detects Patterns of Novel, Non-Coding AAV RNAs in the Absence of Cellular miRNA Regulation. PLoS ONE 11(9): e0161454. doi:10.1371/journal.pone.0161454

Editor: Jianming Qiu, University of Kansas Medical Center, UNITED STATES

Received: June 27, 2016

Accepted: August 5, 2016

Published: September 9, 2016

Data Availability Statement: All sequences were deposited at: NCBI Sequence Read Archive (SRA), Acc.No.: PRJNA338126.

Funding: RH received funding by the Sonnenfeld-Stiftung, Berlin; http://www.sonnenfeld-stiftung.de/. CS received a Ph.D. scholarship from Charité Universitätsmedizin Berlin.

Competing Interests: I have read the journal's policy and the authors of this manuscript have the following

Abstract

Most DNA viruses express small regulatory RNAs, which interfere with viral or cellular gene expression. For adeno-associated virus (AAV), a small ssDNA virus with a complex biphasic life cycle miRNAs or other small regulatory RNAs have not yet been described. This is the first comprehensive Illumina-based RNA-Seq analysis of small RNAs expressed by AAV alone or upon co-infection with helper adenovirus or HSV. Several hotspots of AAV-specific small RNAs were detected mostly close to or within the AAV-ITR and apparently transcribed from the newly identified anti-p5 promoter. An additional small RNA hotspot was located downstream of the p40 promoter, from where transcription of non-coding RNAs associated with the inhibition of adenovirus replication were recently described. Parallel detection of known Ad and HSV miRNAs indirectly validated the newly identified small AAV RNA species. The predominant small RNAs were analyzed on Northern blots and by human argonaute protein-mediated co-immunoprecipitation. None of the small AAV RNAs showed characteristics of bona fide miRNAs, but characteristics of alternative RNA processing indicative of differentially regulated AAV promoter-associated small RNAs. Furthermore, the AAV-induced regulation of cellular miRNA levels was analyzed at different time points post infection. In contrast to other virus groups AAV infection had virtually no effect on the expression of cellular miRNA, which underscores the long-established concept that wild-type AAV infection is apathogenic.

competing interests: RH is an inventor of patents related to AAV vector technology and owns equity in a company that is commercializing AAV for gene therapy. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Introduction

Adeno-associated viruses (AAV) belong to the family of parvoviruses and possess a single-stranded DNA genome of approximately 4.7 kb. A characteristic feature of AAV is its biphasic life cycle. In the absence of a helper virus AAV establishes latent infection and integrates into the host genome or persist as nuclear episome [1–3]. Co-infection with a helper virus, e.g. adenovirus or herpesvirus results in AAV replication and progeny formation [4–8]. AAV type 2 represents the best-studied serotype and is commonly accepted as AAV prototype. The AAV2 genome contains two major open reading frames, *rep* and *cap*, which are flanked by hairpin-structured, 145 bp inverted terminal repeats (ITRs) at either end [9]. The *rep* gene encodes the four regulatory proteins Rep78 and Rep68, and N-terminally truncated versions thereof, called Rep52 and Rep40, respectively. The AAV capsid proteins VP1, VP2 and VP3 are encoded by the *cap* gene. Furthermore, *cap* encodes the assembly activating protein (AAP) by use of an alternative open reading frame [10]. Early AAV2 transcription mapping only defined transcripts derived from the coding AAV positive (+) strand. These mRNAs initiate at the p5, p19 or p40 promoters, named according to their relative positions on the AAV2 genome. In a total RNA-Seq analysis, we have recently discovered transcription on the AAV negative (-) strand opposite to the p5 promoter, indicative of non-coding RNA species [11]. In addition, we have identified p40 promoter-associated short non-coding transcripts on the (+) strand relevant for the inhibition of adenovirus replication [12]. Apparently, non-coding RNA species are involved in the regulation of the AAV life cycle.

Small non-coding RNAs represent a growing class of diverse, regulatory RNAs. Of these, microRNAs (miRNAs) and short interfering RNAs (siRNAs) represent the best-characterized species. These RNAs are approximately 22 nucleotides in length and are processed by the cellular enzyme Dicer from longer double-stranded RNA precursors that form a distinctive, secondary RNA structure [13, 14]. One strand of the processed RNA duplex is loaded into the RNA-induced silencing complex (RISC) allowing recognition of the mRNA target sequence. Mammalian miRNAs and siRNAs typically represent posttranscriptional inhibitors by specifically binding to a target RNA leading to translational repression or mRNA degradation [15, 16]. Less well characterized classes of small regulatory RNAs have since been described, whose functions are largely unknown. Of these, tRNA-derived fragments (tRFs) or microRNA-offset RNAs (moRs) have been suggested to play a miRNA-like role in posttranscriptional gene silencing [17, 18]. Others, such as promoter-associated RNAs (paRNAs) appear to be specifically involved in regulating promoter activity [19].

Most DNA virus genera and also certain RNA viruses express small non-coding RNAs [20, 21], but often the molecular function has not been fully defined. Adenovirus (Ad) generates miRNAs processed from the longer structured virus-associated RNAs, VA-RNA I and II. The VA-RNAs themselves are described to suppress the cellular RNA interference (RNAi) pathway by interfering with the activity of Dicer [22]. For the VA-RNA derived miRNAs (mivaRNAs) cellular target genes were identified, some of which are involved in the regulation of cellular gene expression [23, 24]. The exact role of the Ad mivaRNAs for the adenovirus life cycle has not yet been defined. Most members of the herpesvirus family encode clusters of miRNAs, which are differentially expressed during latent or lytic infection. During HSV1 latency only a single abundant viral transcript is expressed. This latency-associated transcript (*LAT*) represents a non-coding RNA that serves as a precursor for several miRNAs [25], assumed regulators of HSV1 latency. Of the 18 well described HSV1 miRNAs, a role during the herpesviral life cycle has only been determined for three: miR-H2 and miR-H6 repress the major HSV1 transcriptional activators ICP0 and ICP4, respectively, and thus maintain the latent state [25], and miR-H4 has been shown to target the HSV1 pathogenicity factor ICP34.5 [26]. For certain

herpesviruses microRNA-offset RNAs have been identified, whose functions are presently unknown [27, 28].

AAV represents one of the very few DNA viruses for which small regulatory RNAs have not yet been described. Its bipartite life cycle combined with the hairpin-structured AAV-ITRs reminiscent of the structure of miRNA precursors led to the hypothesis that small RNA species with regulatory function might be expressed by AAV as well. In the first comprehensive Illumina-based small RNA-Seq analysis of AAV-infected cells we discovered and validated several classes of previously unknown small non-coding RNAs.

Materials and Methods

Cell culture and viruses

HEK 293 and HeLa cells were obtained over 30 years ago from ATCC and passaged since as described [29]. Human adenovirus type 2 (Ad2) was propagated and quantified in 293 cells. The HSV1 KOS strain was produced in Vero cells and titrated by plaque assay.

AAV2 production, purification and quantification

For AAV2 production HEK 293 cells were seeded at 30% confluency and transfected 24 hours later with pTAV2-0 [30] and pHelper as described [2]. AAV2 was purified from benzonase-treated, cleared freeze-thaw supernatants by one-step AVB sepharose affinity chromatography and quantified by Light-Cycler PCR as described [31] with primers specific for AAV2 *rep* (Rep2-Fwd: `AGAAGGAATGGGAGTTGCCG` and Rep2-Rev: `TCTGACTCAGGAAACGTCCC`). AAV2 infectious titers were determined by end point dilutions on Ad2 infected HeLa cells [32].

Virus infection

HeLa cells were seeded at a density of 30% and infected 20 hours later with AAV2 wild-type (MOI 250), Ad2 (MOI 25) or HSV1 (MOI 10), or indicated combinations thereof.

RNA extraction

Total RNA was extracted 8 hpi and 27 hpi from AAV2 infected cells in presence or absence of HSV1 or Ad2, respectively. Total RNA was isolated using TRIzol Reagent (Ambion) according to the manufacturer's protocol. Subsequently, RNA samples were treated with RNase-free Turbo DNase (Ambion), followed by phenol-chloroform extraction and precipitation. RNA quality and integrity was verified on 0.8% agarose gels and on bioanalyzer (Agilent). Total RNA samples of proven high quality were used for small RNA library generation and Northern blot analysis.

Northern blot analysis

For electrophoresis 15% polyacrylamide gels containing 8 M urea and 0.5x MOPS running buffer (10x MOPS is 200 mM MOPS, 50 mM sodium acetate and 10 mM EDTA, pH 7.0) were used. Prior to loading of the samples the gel was pre-run at 100 V for 30 min. RNA samples were treated using the FDF-PAGE method as described [33]. In short, equal amounts of total RNA (5 μg to 25 μg per lane) in a volume of 4 μl were mixed with 11 μl FDF buffer (2.75 μl formaldehyde (37%), 7.5 μl formamide (>99.5%) and 0.75 μl 10x MOPS buffer), incubated at 55°C for 15 min and subsequently mixed with 2 μl of 10x dyes (0.05% [w/v] xylene cyanol and 0.05% [w/v] bromphenol blue in nuclease-free water) before loading. In addition 1 μl of microRNA ladder (NEB) and 5 μl of low range ssRNA ladder (NEB) were loaded onto the gel. Electrophoresis was run at 150–200 V until the bromphenol blue dye reached about 2 cm at the bottom

of the gel. After electrophoresis, gels were stained in ethidium bromide (10 μg/ml) to visualize the RNA size markers. RNAs were transferred to a positively charged nylon membrane (GeneScreen Plus, PerkinElmer) at 1 mA/cm^2 for 1.5 h or overnight by electroblotting, and cross-linked by UV (Stratagene) or EDC for 2 h at 60°C as described [34]. The membranes were pre-hybridized for 2 h in pre-warmed hybridization buffer (5x SSC [10x SSC is 1.5 M NaCl plus 0.15 M sodium citrate], 1% SDS and 1x Denhardt´s Solution [100x Denhardt´s Solution is 2% [w/v] BSA, 2% [w/v] Ficoll 400 and 2% [w/v] Polyvinylpyrrolidone]). DNA-oligonucleotides (MWG Eurofins) or DNA/LNA-oligonucleotides (Exiqon) [35] with sequences complementary to sRNAs (see S1 Table) were 5' end labeled with [γ-^{32}P]-ATP (Hartmann Analytic), and purified on a Bio-Gel P-10 column (Bio-Rad). Membranes were hybridized for 16 hours at 37°C or at 55–65°C with the labeled DNA- or DNA/LNA-probes, respectively. Subsequently, membranes were washed three times for 10 min with low stringency buffer (2x SSC / 0.1% SDS), followed by two washing steps at higher stringency (1x SSC / 0.1% SDS) for 5 min at 37°C or at 55–65°C for membranes hybridized with DNA- or DNA/LNA-probes, respectively. Membranes were exposed to X-ray films between intensifying screens at -80°C overnight or longer.

Co-immunoprecipitation (Co-IP) of human argonaute protein complexes

Co-immunoprecipitation was carried out as described [23]. In brief, uninfected or infected HeLa cells were washed twice with ice-cold PBS 27 hpi and lysed in cell lysis buffer containing 25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM $MgCl_2$, 0.5% Nonidet P40, 5 mM dithiothreitol (DTT), protease inhibitor (cOmplete Tablets, Mini Easypack, Roche) and 40 U/ml RNase inhibitor (Recombinant RNasin Ribonuclease Inhibitor, Promega) for 30 min at 4°C on an overhead tumbler. Cell debris was removed by centrifugation at 20,000 g for 30 min at 4°C. For co-immunoprecipitation, 100 μl Protein G-Agarose beads (Roche) per preparation were washed three times in ice-cold nuclease-free PBS and subsequently resuspended in 1 ml nuclease-free PBS supplemented with protease inhibitor cocktail. 15 μg of anti-pan Ago mAb clone 2A8 (Merck Millipore) or anti-Rep mAb 76–3 (Progen) were added to the pre-washed beads, respectively, and incubated for 6 hours at 4°C under constant rotation. Antibody-coupled beads were blocked at 4°C over night in blocking solution consisting of nuclease-free PBS containing 1.2 mg/ml BSA, 0.6 mg/ml yeast tRNA (Roche) and protease inhibitor cocktail. The beads were washed three times in ice-cold blocking solution and once with cell lysis buffer. For immunoprecipitation, the cell lysates were incubated with the prepared antibody-coupled beads under constant rotation at 4°C over night. Following, the beads were washed twice with cell lysis buffer supplemented with BSA and tRNA, three times with high salt washing buffer (25 mM Tris-HCl (pH 8.0), 900 mM NaCl, 2 mM $MgCl_2$, 1% Nonidet P40, 5 mM DTT, protease inhibitor cocktail, 40 U/ml RNase inhibitor, BSA and tRNA) and twice with low salt washing buffer (25 mM Tris-HCl (pH 8.0), 150 mM NaCl, 2 mM $MgCl_2$, 0.05% Nonidet P40, 5 mM DTT, protease inhibitor cocktail, 40 U/ml RNase inhibitor, BSA and tRNA). Co-immunoprecipitated RNAs were isolated from the beads using the TRIzol Reagent (Ambion) followed by RNA precipitation and purification.

Small RNA library preparation and Illumina Sequencing

1 μg of each of the eight different total RNA samples (cells, AAV2, AAV2 + HSV1, HSV1 extracted 8 hpi and cells, AAV2, AAV2 + Ad2, Ad2 extracted 27 hpi) were subjected to small RNA library generation using the TruSeq Small RNA Sample Prep Kit according to the manufacturer's protocol (Illumina). Briefly, after ligation of the RNA 3'- and the RNA 5' adapter, total RNA samples were reverse transcribed to cDNA libraries and subsequently PCR-amplified. The PCR products were run on 6% PAGE gels and bands of adapter-ligated DNA libraries

between 145 and 160 basepairs (bp) were excised from the gel and purified. These bands correspond to RNAs of a length of approximately 20 to 35 nucleotides. Subsequently, the libraries were validated by bioanalyzer, pooled and sequenced on a HiSeq 2000 platform (multiplexed 1x51+7).

Processing of sequencing reads

The sequencing reads were demultiplexed and the sequencing adapters were removed. Reads shorter than 16 bp were discarded; the remaining reads were mapped without mismatch to the human genome (hg19), transcriptome (generated from RefSeq annotations) and pre-miRNA sequences (downloaded from the miRBase [36], http://www.mirbase.org) using Bowtie [37]. The unmapped reads were further aligned to HSV1 (accession number NC_001806.1), Ad2 (accession number AC_000007.1) and AAV2 (accession number NC_001401.2) reference genomes. Additional sequence mapping was performed gradually allowing zero, one and two mismatches (see S2 Table).

Database of microRNAs, secondary structure prediction and miRNA target prediction

Known human, HSV1 and Ad microRNAs were obtained from miRBase, the database of microRNAs (http://www.mirbase.org/). For secondary structure prediction of possible precursor RNAs the Vienna RNAfold WebServer was used (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). To search for putative target sites of described human miRNAs the miRNA target prediction tool "Target Scan" was applied (http://www.targetscan.org).

Results

Small RNA-Seq library preparation

To identify AAV2-derived small RNA species we performed Illumina small RNA next generation sequencing. HeLa cells were infected with AAV2 alone or co-infected with AAV2 and the helper viruses Ad2 or HSV1, respectively. Uninfected, Ad2 and HSV1 infected cells served as controls. Total RNAs were extracted at 8 hpi after HSV1 infection or 27 hpi after Ad2 infection. Together, two sets of four total RNA samples were generated, each representing the latent and lytic AAV2 infection state including controls. All samples contained highly intact RNA as validated by bioanalyzer displaying RNA integrity numbers (RIN) higher than 8.90. Using the Illumina TruSeq small RNA sample preparation kit, libraries of RNAs with sizes between 20 to 35 nucleotides were generated. The integrity of the libraries was re-checked immediately prior to sequencing on the Illumina HiSeq 2000 platform.

Illumina small RNA-Seq next generation sequencing

For each data set between 18.0 and 24.9 million RNA reads were obtained by Illumina small RNA-Seq analysis (Table 1). The majority of these reads (>88%) corresponded to RNAs with a size of at least 16 nucleotides (nt). Based on these, a total of 50 to 85% could be assigned to sequences of the human genome without allowing any mismatch. Due to the inherent error rate of Illumina sequencing, higher numbers of reads could be assigned to the human and viral genomes when allowing one and two mismatches (S2 Table). While the numbers per mapped read increased, this did not lead to the detection of further small RNAs. In all of the 8 data sets a significant portion of reads could be mapped to known human miRNAs (Table 1). In the absence of a helper virus very low numbers of small AAV2-specific RNA reads were detectable in AAV2 infected cells. Their numbers were increased by more than 200-fold in presence of

Table 1. Small RNA-Seq—Assignment of the reads to the species from which they originate.

Data set	Total no. of reads	No. (%) of reads ≥ 16 nt	No. (%) of human reads[a]	No. (%) of small reads assigned to:				No. (%) of unknown reads[b]
				Human miRNAs	AAV2	Ad2	HSV1	
Cells (27 hpi)	19,395,459	17,429,392 (89.9)	14,653,485 (84.1)	4,793,871 (27.5)	17 (<0.1)	348 (<0.1)	6 (<0.1)	809,469 (4.6)
AAV2 (27 hpi)	18,085,432	16,228,519 (89.7)	13,331,538 (82.1)	5,840,779 (36.0)	1,193 (<0.1)	1,209 (<0.1)	9 (<0.1)	1,037,657 (6.4)
AAV2 + Ad2 (27 hpi)	21,333,021	18,814,505 (88.2)	11,238,151 (59.7)	3,307,157 (17.6)	270,577 (1.4)	2,644,519 (14.1)	12 (<0.1)	2,142,730 (11.4)
Ad2 (27 hpi)	24,841,074	24,202,341 (97.4)	12,151,399 (50.2)	5,129,522 (21.2)	24 (<0.1)	5,332,713 (22.0)	14 (<0.1)	6,079,458 (25.1)
Cells (8 hpi)	21,851,591	20,985,998 (96.0)	17,475,849 (83.3)	6,899,774 (32.9)	15 (<0.1)	286 (<0.1)	13 (<0.1)	2,644,242 (12.6)
AAV2 (8 hpi)	20,375,107	20,022,269 (98.3)	16,464,292 (82.2)	7,551,710 (37.7)	160 (<0.1)	449 (<0.1)	10 (<0.1)	3,204,520 (16.0)
AAV2 + HSV1 (8 hpi)	23,031,470	20,904,023 (90.8)	16,445,457 (78.7)	7,202,911 (34.5)	146,807 (0.7)	3,933 (<0.1)	108,860 (0.5)	2,071,519 (9.9)
HSV1 (8 hpi)	22,391,766	20,895,686 (93.3)	16,671,702 (79.8)	9,745,884 (46.6)	14 (<0.1)	326 (<0.1)	124,055 (0.6)	2,603,509 (12.5)

[a]Included are reads mapped to the human genome, transcriptome and miRNAs.

[b]Included are reads ≥ 16 nt that were unmappable during sequence mapping allowing zero mismatches.

doi:10.1371/journal.pone.0161454.t001

HSV1 or Ad2 accounting for 0.7 to 1.4% of total small RNA reads displaying at least 16 nt in size (Table 1). The consistently lower read numbers in the HSV data set may be due to the short HSV replication cycle and consequently earlier time point of cell harvest (8 hpi versus 27 hpi). Similar to AAV2, HSV1 showed a low number of specific small RNAs (0.5 to 0.6% of total reads), whereas the number of small RNAs mappable to Ad2 was significantly higher (14.1 to 22.0% of total reads). On the other hand, Ad2-specific read numbers were significantly reduced in presence of AAV2, presumably reflecting the known inhibitory effect of AAV on adenovirus replication [38, 39]. Furthermore, in the presence of Ad2 human miRNAs were slightly less abundant compared to those in the other data sets (Table 1) confirming the recent finding that certain human miRNAs are dramatically deregulated during adenovirus infection [40]. In addition, adenovirus was shown to block the RNAi processing machinery in order to generate its own miRNAs late during infection [22]. In the following all further analyses were performed using zero mismatches per mapped read (Table 1).

Small RNA read mapping to the AAV2 (+) and (-) strand

The population of AAV2-specific small RNA reads for each data set was mapped to the wild-type AAV2 genome. For a better overview RNA reads were initially counted in closed intervals of 10 nucleotides. In the presence of Ad2, AAV2 infected cells showed a high number of small RNAs located within the ITRs (Fig 1A). Since small RNAs located in the ITRs cannot be mapped unambiguously to one ITR, they were assigned to either end. A significant amount of reads mapped to the region opposite of the p5 promoter (anti-p5) on the AAV (-) strand and to a region close to the p40 promoter on the AAV (+) strand (Fig 1A). The transcription profile of small AAV2 RNAs (27 hpi) was comparable in cells with and without Ad2 infection, but the absolute read numbers differ by a factor of around 200. However, no accumulation of p40 promoter-associated reads could be observed in the absence of Ad2 (Fig 1B). In the presence of HSV1 the AAV transcription profile of small RNAs was different to that observed in the

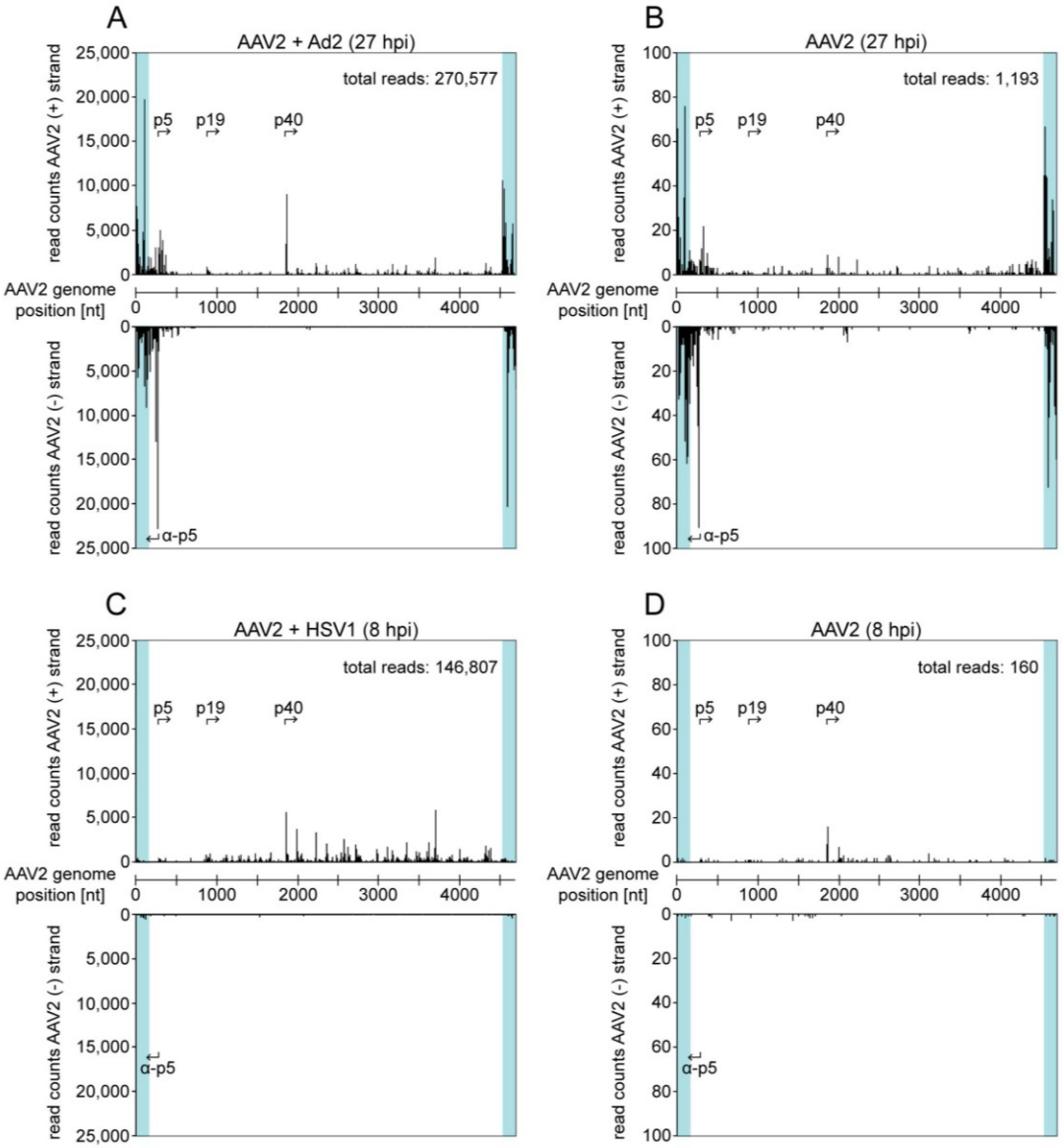


Fig 1. Small RNA-Seq: Read mapping of small AAV2-specific RNAs to either strand of the genome. (A) Mapping of AAV2-specific reads identified by the small RNA-Seq analysis in AAV2/Ad2 co-infected HeLa cells 27 hours p.i. (hpi). The AAV2 genome scale is displayed in the center. Reads mapped to the AAV2 (+) strand are displayed above and those mapped to the AAV2 (-) strand below the genome, read counts are shown in intervals of 10 nucleotides. The ITR regions are highlighted in turquoise. The positions of the known AAV promoters on either strand are indicated by arrows. The total read count is indicated. Note the ambiguous assignment of reads mapping to the ITRs. (B) Mapping of small AAV2-specific reads in AAV2 infected HeLa cells 27 hpi, displayed as in (A). (C) Mapping of small AAV2-specific reads in AAV2/HSV1 co-infected HeLa cells 8 hpi, displayed as in (A). (D) Mapping of small AAV2-specific reads in AAV2 infected HeLa cells 8 hpi, displayed as in (A). Note the different scales of read counts in (A to B) and (C to D).

doi:10.1371/journal.pone.0161454.g001

presence of Ad2 (Fig 1C). Neither the ITR nor the AAV (-) strand showed an accumulation of small RNA reads, but similarly to Ad2 co-infected cells, there was indication of a RNA hotspot close to the p40 promoter (compare Fig 1A and 1C). Additional small RNA reads were located downstream of the p40 promoter (Fig 1C). RNA harvested from AAV infected cells (8 hpi) showed hardly any AAV2-specific reads (Fig 1D). Obviously, AAV transcription in the absence of a helpervirus starts delayed (compare Fig 1A to 1B and 1C to 1D).

Mapping of small RNA reads to the Ad2 and HSV1 genomes

Comparative analysis of Ad2 and HSV1 as helper viruses for productive AAV2 replication allowed the validation of small RNA-Seq data by analysis of known Ad or HSV encoded miRNAs, respectively. Indeed, the majority of Ad2-specific small RNA reads (>95%) could be assigned to the adenoviral mivaRNAs in the absence, as well as in the presence of AAV2 (Fig 2A and 2B). These small RNAs, processed from the longer structured VA RNA I and II, are expressed in lytic, as well as in persistently infected cells [41, 42]. Co-infection with AAV2 reduced the numbers of mivaRNA I and II by half, likely explained by the repressive activity of co-replicating AAV2. Similarly, many of the HSV1-specific small RNA reads could be assigned to known HSV1 encoded miRNAs in the data sets of HSV1 infected cells in the presence or absence of AAV2 (Fig 2C and 2D). Since small RNA reads within repeat regions cannot be unambiguously mapped, they were assigned to either HSV repeat as applied here for the small AAV RNAs. Prominent hotspots could be attributed to the miRNAs miR-H4-3p and miR-H6-3p located within the HSV1 repeat region IR_L (TR_L) (Fig 2C and 2D). Furthermore, HSV1-miR-1 and -2 were moderately expressed with 300 and 700 reads, respectively, whereas HSV1-miR-3, -5, -11, -12 and -16 were hardly detectable (<100 reads). HSV miRNAs are differently expressed during latent and lytic infection in highly susceptible cell types [27]. In HeLa cells used here, a prominent small RNA hotspot was detected at HSV1 genome position 132,142 [+] or 146,091 [−] (Fig 2C and 2D, named IR_S-HS / TR_S-HS). This hotspot has also been reported in another study during early lytic HSV1 infection just recently [43].

Detailed mapping of AAV2-specific small RNAs

For more precise mapping of the small AAV2-specific RNAs the regions with high read counts were enlarged to single nucleotide level in the data set of AAV2/Ad2 co-infected cells (Fig 3A). The 5' end of any small RNA was defined as starting nucleotide used henceforth as nomenclature for these RNAs. Even on a single nucleotide level hotspots of small RNAs were detected. The threshold was arbitrarily set to 3,000 reads, which is approximately 100-fold above background. Only small RNAs above this threshold were defined as hotspots and analyzed further. Four hotspots were located within the AAV-ITRs, sR-1 and sR-108 on the AAV (+) strand, and sR-125 and sR-141 on the AAV (-) strand (Fig 3A). Due to the ambiguous character of the ITRs sR-1 on the AAV (+) strand is equivalent to sR-125 on the AAV (-) strand with 3,170 reads (Fig 3B). Furthermore, sR-1, sR-108 and sR-141 can be mapped to either ITR on the AAV (+) or (-) strand (see S3 Table). In fact, sR-108 represented one of the two most abundant small RNAs with 12,002 read counts (Fig 3B) with a predominant mean read length of 21 nt (Fig 3C). The other prominent hotspot, displaying sR-271, was located downstream of the newly identified anti-p5 promoter on the AAV (-) strand [11] with 12,737 reads and showed varying read lengths between 16 and 23 nt (Fig 3C). Furthermore, additional minor hotspots clustered in that region (Fig 3A and 3B). Another hotspot, displaying sR-1862, was located just downstream the transcription start site (TSS) of the p40 promoter with 3,780 read counts. This particular small RNA showed a read length between 18 and 19 nt (Fig 3C). All hotspots displayed an average read length between 19 and 22 nt, typically seen for small non-coding regulatory RNAs

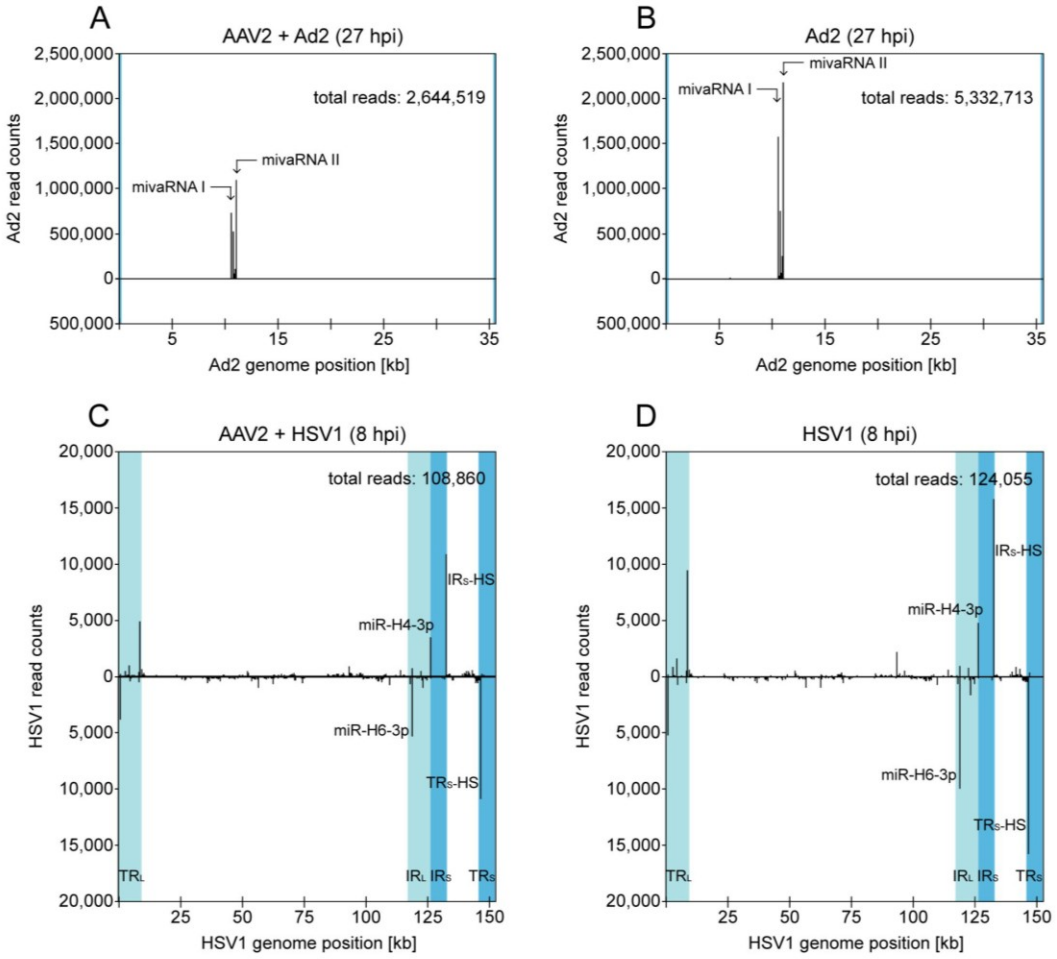
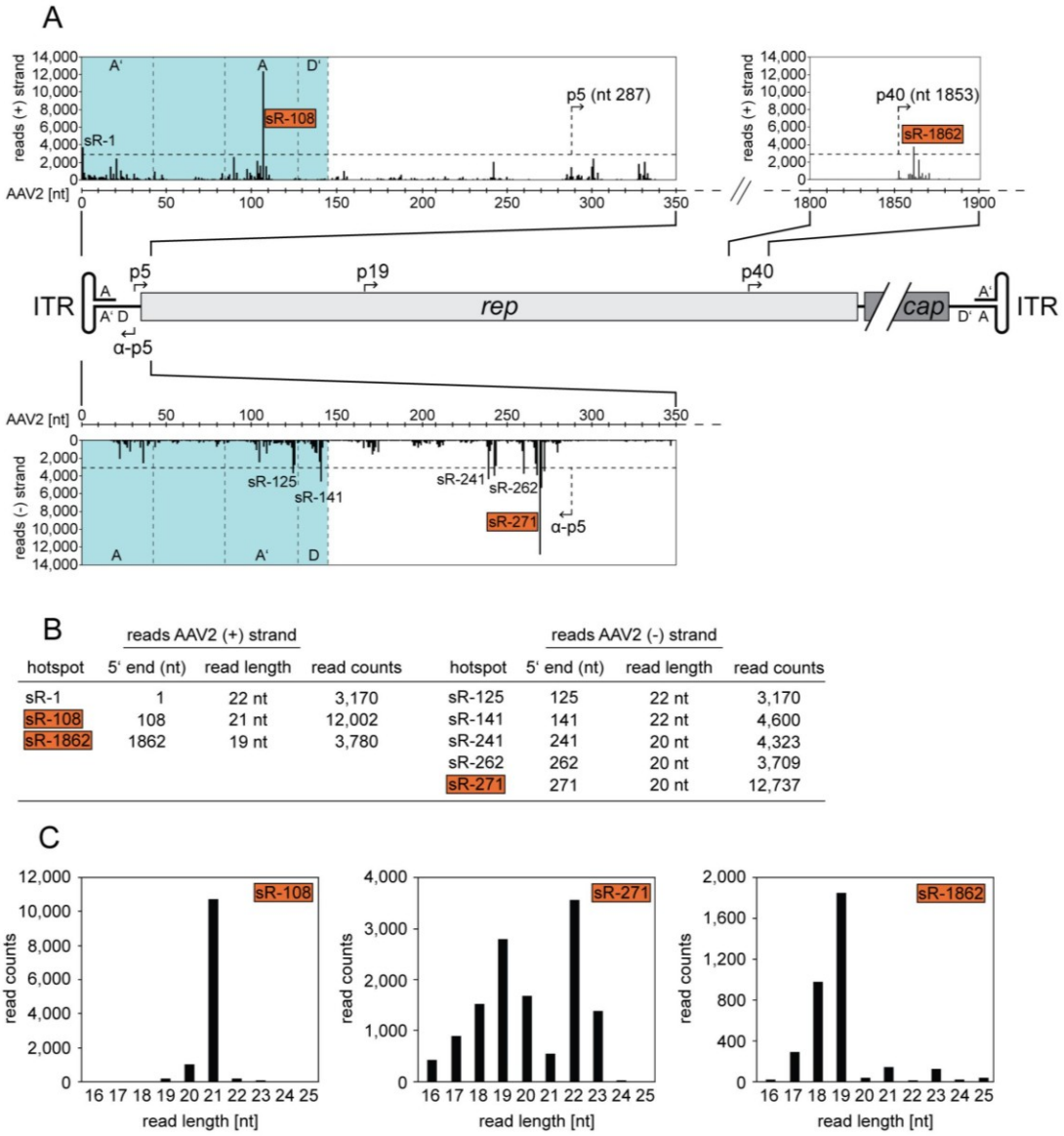


Fig 2. Small RNA-Seq: Mapping of small RNA reads to the genomes of Ad2, or HSV1. (A) Mapping of Ad2 reads identified by the small RNA-Seq analysis to the Ad2 genome in AAV2/Ad2 co-infected HeLa cells 27 hpi. The Ad2 genome scale is displayed in the center. Reads mapped to the Ad2 (+) strand are displayed above and to the Ad2 (-) strand below, read counts are shown in intervals of 10 nucleotides. Terminal repeats flanking the adenoviral genome are highlighted in turquoise. Reads assigned to the described adenoviral mivaRNAs I and II are indicated by arrows. The total read count of small RNAs is indicated. (B) Mapping of small Ad2 reads in Ad2 infected HeLa cells 27 hpi, displayed as in (A). (C) Mapping of HSV1 reads identified by the small RNA-Seq analysis to the HSV1 genome (KOS strain) in AAV2/HSV1 co-infected HeLa cells 8 hpi, displayed as in (A), but in intervals of 100 nucleotides. Terminal (TR) and internal repeat (IR) regions within the HSV1 genome are highlighted in shades of turquoise. Small RNAs of high read frequencies were assigned to known HSV1 miRNAs according to miRBase. A small RNA with very high read counts was designated as IR_S-HS / TR_S-HS due to its location within these repeat regions. Note the ambiguous assignment of reads mapping to the repeat regions. (D) Mapping of small HSV1 reads in HSV1 infected HeLa cells 8 hpi, displayed as in (C). Note the different scales of read counts in Ad2 infected cells (A, B) and HSV1 infected cells (C, D).

doi:10.1371/journal.pone.0161454.g002

(Fig 3B). The overall expression of the AAV2-specific small RNAs was comparable to that of HSV1 small RNAs but lower than that of Ad mivaRNAs. For AAV2/HSV1 infected cells the small AAV2-specific reads were of low abundance and mostly below the threshold of 3,000 reads. A list of the top hundred AAV-specific small RNAs in the presence of the helper viruses



B

reads AAV2 (+) strand				reads AAV2 (-) strand			
hotspot	5' end (nt)	read length	read counts	hotspot	5' end (nt)	read length	read counts
sR-1	1	22 nt	3,170	sR-125	125	22 nt	3,170
sR-108	108	21 nt	12,002	sR-141	141	22 nt	4,600
sR-1862	1862	19 nt	3,780	sR-241	241	20 nt	4,323
				sR-262	262	20 nt	3,709
				sR-271	271	20 nt	12,737

Fig 3. Small RNA-Seq: Detailed mapping of small AAV2-specific reads to both strands of the genome. (A) Detailed mapping of small AAV2-specific reads in regions of high read frequencies in AAV2/Ad2 co-infected cells from Fig 1A. A schematic representation of the AAV2 genome is displayed in the center. Read assignments to the (+) strand are presented above and to the (-) strand below the genome, with read counts displayed at single nucleotide level. A horizontal dashed line marks the threshold value set to 3,000 reads. Any small RNA (sR) above 3,000 read counts was defined as hotspot candidate, designated according to the 5' starting nucleotide. Different genomic regions of the AAV2-ITRs (highlighted in turquoise) are separated by vertical dashed lines. Complementary regions of the hairpin-structured ITR are indicated by letters, e.g. A, A'; D, D'. Promoters on the AAV2 (+) or (-) strand are indicated by arrows. (B) Summary of small RNA hotspot candidates on the AAV2 genome, displayed in (A), designated according to their starting nucleotide (5' end). Average read lengths are shown for the small RNAs. (C) Depicted are the detailed small RNA read length distributions for sR-108, sR-271 and sR-1862. Small RNA hotspots highlighted in red were validated further.

doi:10.1371/journal.pone.0161454.g003

Ad and HSV, respectively, is given in S3 Table, which includes data of genomic location, read counts, median read length and sequence of the respective small AAV RNAs.

Validation of the AAV2-specific small RNA hotspots

To confirm the presence of AAV2-specific small RNAs by an alternative method, total RNA was harvested from AAV2 infected and AAV2/Ad2 co-infected HeLa cells and respective controls. RNA samples treated with RNase-free DNase were analyzed on Northern blots with specific probes directed against the most prominent AAV2-specific small RNAs. As internal controls, the human miRNA hsa-let-7a-1 and the splicosomal U6-snRNA were detected in all RNA extracts (Fig 4A) showing the high quality of the RNA samples. The highly abundant sR-108 located in the ITR of the AAV (+) strand was detected with a specific DNA probe (Fig 4A) and also with a specific DNA/LNA probe (Fig 4B) only in extracts harvested from AAV2/Ad2 co-infected cells, but not in AAV2/HSV1 co-infected cells (Fig 4A), which is in agreement with the data of the RNA-Seq analysis described above. The size of the detected bands was in the range of approximately 20–30 nt. Furthermore, larger RNA fragments with sizes of approximately 100 nt, 200 nt, 250 nt and above were visible (Fig 4B). Similar bands were also detected with a DNA probe, specific to sR-271, located on the AAV (-) strand (compare Fig 4C and 4B). Furthermore, a very faint band at a size around 25 nt was visible, corresponding to sR-271 (Fig 4C). In agreement with the varying read length distribution of sR-271 described above (see Fig 3C), it was not surprising that no sharp band could be detected in the gel for this small RNA. In contrast to sR-108 and sR-271, no AAV2-specific small RNA indicative of sR-1862 could be detected using a DNA/LNA probe (Fig 4D). The failure to validate sR-1862 by Northern blot analysis might be due to its 3- to 4-fold lower abundance compared to sR-108 and sR-271 (see Fig 3), probably lower than the detection limit of the used method. Unexpectedly, either probe (DNA and DNA/LNA) detected a strong Ad2-specific band of $>$ 300 nt (Fig 4D). Even with highly stringent hybridization and washing conditions the bands persisted. An alignment to the Ad2 genome did not show a target sequence of the designed AAV2 probe. None of the AAV2-specific small RNAs was detected upon HSV co-infection, which was anticipated from their low abundance in RNA-Seq.

Detection of human argonaute protein (Ago) co-immunoprecipitated RNAs

To determine whether sR-108, sR-271 and sR-1862 represent functional miRNAs, their loading into the RNA-induced silencing complex (RISC) was studied. Ago2 is a component of RISC, which binds the mature miRNA and thus enables mRNA target recognition. We performed co-immunoprecipitations (Co-IP) of human Ago2 complexes with total RNAs from infected cells, using an antibody that detects the human argonaute proteins 1 to 4. Subsequently, the RNAs isolated from the precipitates were analyzed on Northern blots using probes directed against the AAV2-specific small RNAs sR-108, sR-271 or sR-1862, respectively. Probes directed against the human miRNA hsa-let-7a and the cellular U6-snRNA, a component of the splicosome, were used as controls. The human miRNA hsa-let-7a-1 could be detected in all total RNA and Co-IP RNA extracts, with the exception of co-immunoprecipitated RNAs in lane 7 for which an unrelated control antibody was used (Fig 5). In the AAV/Ad2 co-infected and Ad2 infected Co-IPs, hsa-let-7a-1 was reduced (Fig 5, lanes 6 and 8), in agreement with data that adenovirus down-regulate human hsa-let-7a [40] and blocks the RNAi pathway [22]. Since the U6-snRNA is not incorporated into RISC, it is only detectable in total RNA extracts (Fig 5, lanes 1 to 4). In contrast to hsa-let-7a, AAV-specific small RNAs could not be detected after immunoprecipitation of Ago2 complexes with any of the probes for sR-108, sR-271, or sR-1862 (Fig 5). Obviously, the identified small RNAs do not represent functional miRNAs.

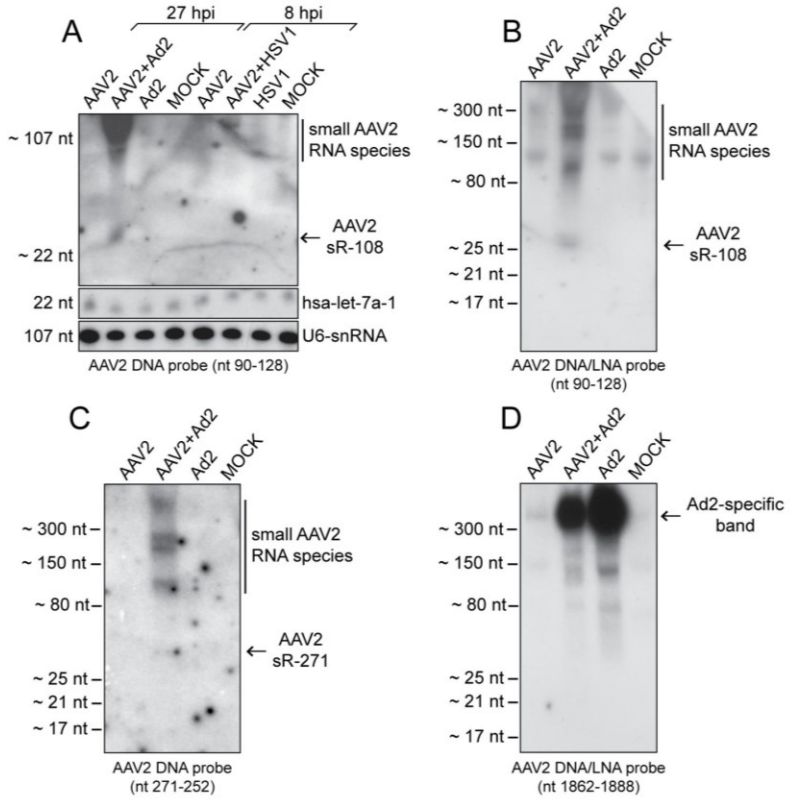


Fig 4. Detection of AAV2-specific small RNAs by Northern blot analysis. (A) Northern blot analysis of DNase treated total RNA extracts from HeLa cells, either none infected (Mock), or infected with AAV2, AAV2 and Ad2, AAV2 and HSV1, Ad2, or HSV1, harvested at the indicated time points. Total RNA extracts separated on 15% PAA-Urea gels, transferred to a positively charged membrane were hybridized with radiolabeled DNA probes directed against the depicted AAV2 small RNA sR-108, the human miRNA hsa-let-7a, or the cellular U6-snRNA, respectively. To the right detected small RNAs are indicated. (B) Experimental procedure as in (A) but RNAs were detected by a radiolabeled DNA/LNA probe directed against the AAV2 small RNA sR-108. (C) Experimental procedure as in (A) but RNAs were detected by a radiolabeled DNA probe directed against the AAV2 small RNA sR-271. (D) Experimental procedure as in (A) but RNAs were detected by a radiolabeled DNA/LNA probe targeting the AAV2 small RNA sR-1862. To the left, the molecular sizes of the bands of the miRNA ladder and the ssRNA ladder, respectively, are depicted.

doi:10.1371/journal.pone.0161454.g004

Instead, these RNAs may belong to another class of small non-coding regulatory RNAs, or are alternatively processed.

Alignment of the small AAV2 RNAs to other AAV serotypes

To evaluate the relevance of the newly identified small RNAs the genomes of AAV serotypes 1 to 7 were aligned and compared to the AAV2 sequences of sR-108, sR-271 and sR-1862 (S1 Fig). The small AAV2 RNAs sR-108 and sR-1862 showed a high degree of conservation among the AAV serotypes, with the exception of AAV5, the most distantly related member. AAV2 sR-271 on the AAV (-) strand displayed some variations from other AAV serotypes in the central part of the small RNA.

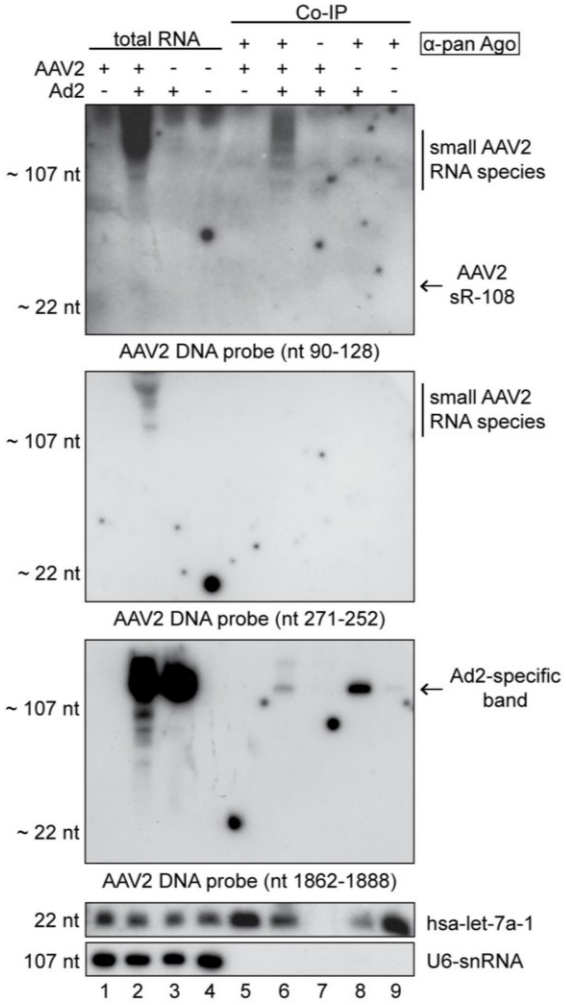


Fig 5. Human argonaute protein (Ago) co-immunoprecipitated RNAs of AAV2 infected cells. Northern blot analysis of DNase treated total RNA extracts or co-immunoprecipitated Ago-RNA complexes as indicated from HeLa cells, either mock infected, or infected with AAV2, AAV2 and Ad2, or Ad2, harvested 27 hpi. For co-immunoprecipitation, cell lysates were incubated over night with anti-pan Ago mAb or anti-Rep mAb 76–3 antibodies coupled to protein G agarose beads, respectively, as outlined in the methods. Northern blot analysis of purified RNAs was performed as described in Fig 4, with radiolabeled DNA probes directed against the AAV2-specific small RNAs sR-108, sR-271 or sR-1862, the human miRNA hsa-let-7a or the cellular U6-snRNA, respectively, as indicated. Note that instead of the human anti-pan Ago antibody an unrelated antibody (anti-Rep76-3) was used in lane 7 for co-immunoprecipitation.

doi:10.1371/journal.pone.0161454.g005

Effect of AAV2 on the expression levels of cellular miRNAs

To assess whether AAV2 infection has an effect on the cellular miRNA expression profile, the read counts of 979 cellular miRNAs were determined, that are listed in the miRNA database (miRBase) [44]. Their expression levels determined by small RNA-Seq in AAV2 infected cells

8 hpi, 27 hpi were compared to those in uninfected cells (Fig 6). In addition, cellular miRNA regulation by lytic AAV replication in the presence of either helper virus was analyzed (S2 Fig). In the light of known extensive effects of either Ad or HSV on host cell transcription, miRNA effects cannot be readily attributed to lytic AAV infection. In the case of AAV infection alone the majority of cellular miRNAs were unaltered within a 2- to 5-fold range of regulation. One single human miRNA, hsa-mir-3687, was down-regulated slightly above 5-fold upon AAV infection (Fig 6A and 6B). To evaluate the significance of hsa-mir-3687 regulation the program "TargetScan" for miRNA target prediction was used and putative target genes were searched [45]. Eight human genes were predicted as potential targets of hsa-mir-3687, with MTA2 (metastasis associated 1 family, member 2) as the most likely candidate. However, the RNA expression levels of the predicted cellular target genes retrieved from our previously total RNA-Seq analysis performed under the same conditions were unaltered. Thus, AAV infection even at high MOIs does not significantly affect cellular miRNAs. This result extends earlier findings from microarray-based RNA screens demonstrating that AAV infection has no major effect on host mRNA expression [46]. The absence of miRNA induced transcriptional effects upon AAV infection further underlines the long-held notion that AAV infection is largely apathogenic.

Discussion

In this report we present the first comprehensive analysis of AAV-encoded small RNAs from infected cells using Illumina small RNA-Seq. Whereas the cellular miRNA profile was largely unchanged by AAV infection, previously unrecognized small AAV RNAs were detected. Their abundance, AAV serotype conservation and predominant localization in close proximity to the established AAV promoters, p40 and p5, calls for roles as transcriptional regulators during latency and active replication.

Putative generation of the small AAV2 RNAs

Based on their location we postulate that sR-108 and sR-271 are generated from the recently identified anti-p5 promoter, whereas sR-1862 is expressed from the p40 promoter. We recently detected the novel anti-p5 promoter on the AAV (-) strand [11]. The previously performed total RNA-Seq designed to detect long AAV transcripts led to its discovery.

In order to initiate transcription, second-strand synthesis of the AAV ssDNA genome is required [47]. Thereby a monomer turnaround structure (mT) is generated with the ITR connected to both complementary strands (Fig 7A). RNA polymerase can proceed through the ITR as shown recently for AAV mutants lacking the authentic polyA signal [48]. By continuing transcription on the opposite AAV strand double-stranded RNA can be formed (Fig 7A) as shown recently for the related parvovirus, minute virus of mice (MVM) [49]. The postulated AAV-ITR precursor transcript would comprise the two small RNAs, sR-108 and sR-271. Alternatively, RNA transcription could also be initiated independently at the ITR to yield the RNAs on the AAV (+) strand (sR-108). Transcripts mapping to the AAV2-ITRs have been reported previously [11, 50, 51], but neither 5' ends nor function have been further investigated. For AAV5, the 5' end of the analogous transcripts could be mapped to nucleotide position 142 [52]. Small RNAs are often processed from longer precursor RNAs by various enzymes [53]. Due to the structure of the ITR the RNA product will automatically adopt a T-hairpin loop secondary structure (Fig 7A and 7B). An additional hairpin loop is generated at the terminal resolution site (*trs*) during Rep-dependent, strand-specific nicking of the AAV genome required to resolve the ssDNA genomes during AAV DNA replication [54, 55]. Recently, AAV2 Rep was shown to also bind to ITRs composed of RNA [48]. Nicking of the *trs* RNA-bound Rep, would

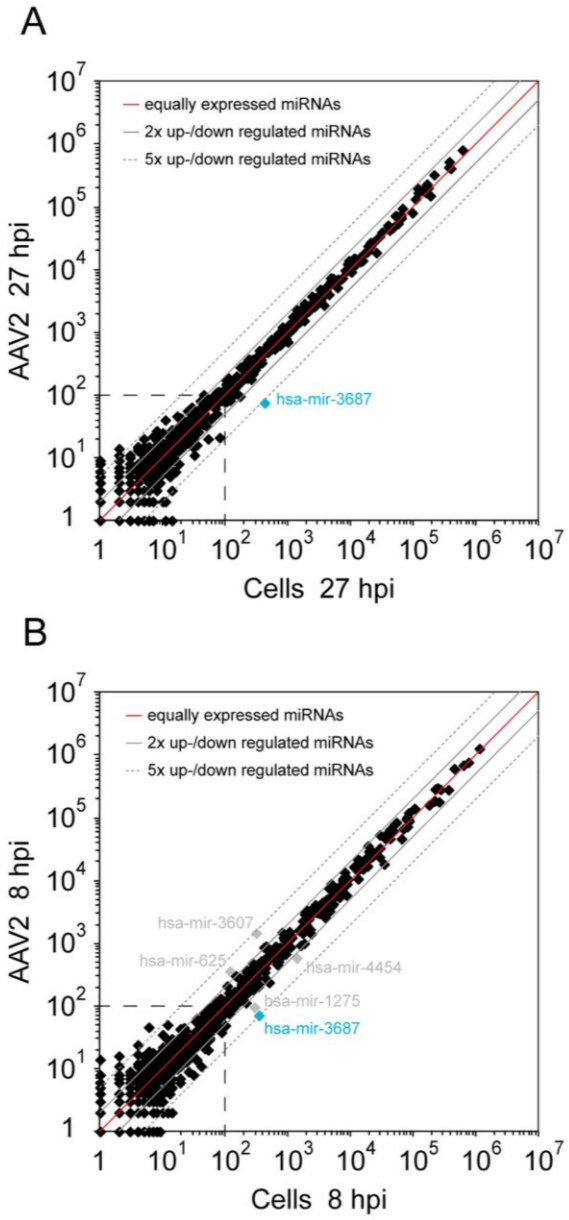


Fig 6. Effect of AAV2 infection on the expression of cellular miRNAs. (A) The expression levels of 979 described cellular miRNAs were analyzed by small RNA-Seq analysis in AAV2 infected versus uninfected HeLa cells at 27 hpi. Dots near the red line represent equally expressed miRNAs in both data sets, the solid gray lines mark 2-fold and the dashed gray lines 5-fold up- or down-regulation of miRNAs. The most highly up- or down-regulated miRNAs were highlighted and specified. Cellular miRNAs below a level of 100 read counts, within the dashed square, were not considered as representative. (B) Expression level of 979 described cellular miRNAs in AAV2 infected versus uninfected HeLa cells 8 hpi, displayed as in (A).

doi:10.1371/journal.pone.0161454.g006

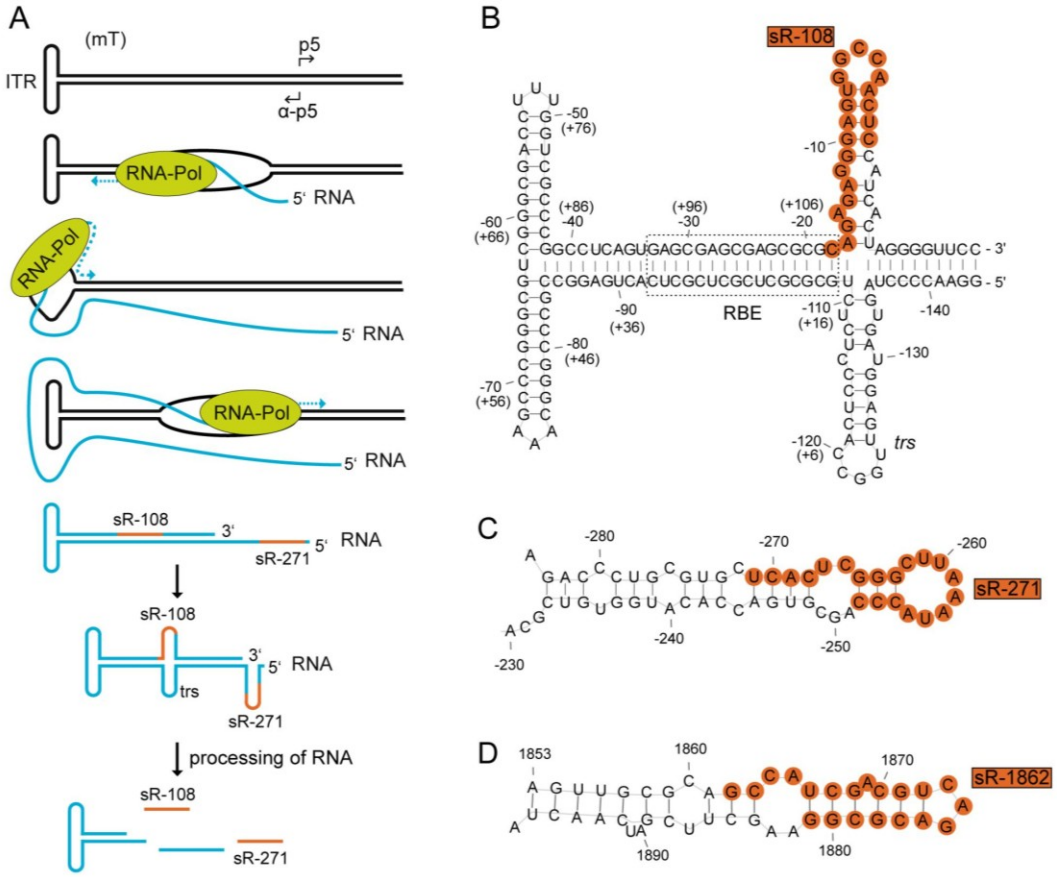


Fig 7. Model for the putative generation of the AAV2 small RNAs derived from the ITR and secondary structure formation. (A) A model for the putative generation of AAV2-specific small RNAs located within and close to the ITR is given. RNA transcription is initiated from the previously identified anti-p5 promoter on the AAV2 (-) strand. The RNA polymerase reads through the AAV2-ITR generating a transcript (highlighted in blue) that reforms the T-shape structure of the ITR. Furthermore, this transcript is able to form additional hairpin loops, which might be recognized by specific processing enzymes that generate the small RNAs sR-271 and sR-108 (highlighted in red). (B) Secondary structure formation of the AAV2-ITR RNA. Depicted are the nucleotide positions of the (-) and in brackets of the (+) strand. Indicated are the Rep binding element (RBE) framed in a dashed rectangle and the terminal resolution site (*trs*) at nucleotide position -125 (+1). The newly identified small RNA sR-108 ranging from nucleotide position 108 to 128 on the AAV (+) strand is highlighted in red. (C) Secondary structure formation of AAV2 RNA nucleotides 284 to 230 on the (-) strand. The newly identified small RNA sR-271 ranging from nucleotide position 271 to 252 on the AAV (-) strand is highlighted in red. (D) Secondary structure formation of AAV2 RNA nucleotides 1853 to 1897 on the (+) strand. The newly identified small RNA sR-1862 ranging from nucleotide position 1862 to 1881 on the AAV (+) strand is highlighted in red. Secondary structure prediction was generated using RNAfold.

doi:10.1371/journal.pone.0161454.g007

result in the here-described small RNAs starting at nucleotide position 1 or 125, respectively (Fig 3, sR-1 and sR-125). Furthermore, genome-wide AAV integration analysis implied that Rep-induced nicking within the human genome does not necessarily require a *trs* adjacent to the Rep binding element (RBE) [2], leaving the possibility of Rep endonucleolytic attack at other sites. Similar to the ITR, the region near the anti-p5 promoter potentially allows formation of a RNA secondary structure (Fig 7C). The sequence of sR-271 partially overlaps with the p5 RBE and with the p5 TATA-box on the complementary strand. Since binding of Rep was

also shown on the RBE within the p5 promoter [56], Rep could also play a role in RNA processing at this site.

Transcription of the small RNA sR-1862, located a few nucleotides downstream of the p40 TSS (nt 1853) could also lead to a secondary structure (Fig 7D). Recently, we have identified this region as *cis*-regulatory element for the inhibition of recombinant adenovirus replication [12]. Here we show by small RNA-Seq that p40-initiated transcripts although highly heterogeneous (Fig 3A), mostly represent sR-1862.

Interestingly, in all stem loop models the small RNAs, sR-108, sR-271 and sR-1862, span sequences, which cover both sides of the stem loop including the terminal loop (Fig 7B–7D). It is apparent that a processing mechanism other than the canonical miRNA processing pathway is required to generate these small RNAs, since Dicer regularly removes the terminal loop for miRNA generation.

Potential role of small AAV2 RNAs

Typically, virus encoded small non-coding RNAs are expressed either to support virus propagation by generating beneficial conditions or to evade the cellular immune response [21]. The majority of identified small RNAs represent microRNAs, but also other types of small RNAs have been described [57–59]. For the parvovirus family including AAV, small RNAs have not yet been reported. The here identified AAV-specific small RNAs are candidate regulators to adapt to the environment of the host cell and/or co-replicating helper viruses.

The functional role of the small RNAs located within the ITRs are difficult to analyze by genetic means, due to an overlap with elements required for DNA replication and/or virus packaging [60, 61]. A series of previously described mutants in these elements showed a replication- and/or packaging-deficient phenotype [62, 63]. In addition to affecting the functions of the AAV RBE and *trs*, the described mutations might have inadvertently abolished the generation of the newly detected small RNAs.

In a recent study we have shown that short transcripts initiated at the anti-p5 promoter are generated [11]. Due to the heterogeneous character of these anti-p5 transcripts, we concluded a mechanism reminiscent of divergent transcription from the AAV p5 promoter. In agreement with that, small RNA-Seq showed several hotspots of small AAV2 RNAs initiated from the anti-p5 promoter with various 5' ends (Fig 3). The short antisense transcripts generated during divergent transcription are believed to play a role in gene regulation by modulating promoter activity [64]. Similarly, the here identified small RNAs sR-271, sR-262 and sR-241 might play a role in AAV p5 promoter activity regulation.

In addition, a heterogeneous group of small AAV p40 promoter-associated transcripts has been described [12]. A mechanism of RNA polymerase II pausing accompanied with the inhibition of recombinant adenoviral replication was postulated. The here identified small RNA sR-1862 starts next to the p40 TSS (nt 1853) and might be a likely candidate for adenoviral inhibition. On the other hand, sR-1862 could also represent a transcription initiation RNA (tiRNA). These small RNAs of 18 nt in size are typically located just downstream the TSS of promoters and are generated by RNA polymerase II backtracking and cleavage by TFIIS. For eukaryotes, tiRNAs have been described to be associated with highly expressed transcripts [65]. Driving the expression of AAV capsid proteins the AAV p40 promoter is extremely active late during AAV replication.

The complex regulation of the AAV life cycle has not yet been fully unraveled, particularly not *in vivo*. The discovery of the small AAV RNA species some derived from transcripts from within the AAV-ITRs and associated with the major AAV promoters adds a new piece of information to the unsolved puzzle of the AAV switch from latency to productive replication. In the

absence of a suitable animal model to study the AAV biology *in vivo*, systems biology approaches to study viral interaction with the host cell transcriptome offer detailed insight onto potentially pathogenic host cell effects. Surprisingly and in contrast to most other virus infections studied, AAV infection neither induces own miRNAs nor regulates host cell miRNA levels to any significant extend. These findings are in line with earlier microarray-based cellular mRNA analysis, where AAV infection showed virtually no regulation [46]. Together the data underline the notion of AAV's apathogenicity. Further studies particularly *in vivo* will be required to unravel the role of the newly identified, obviously regulatory small AAV RNAs for the AAV life cycle.

Supporting Information

S1 Fig. Alignment of AAV2-specific small RNAs for AAV serotypes 1 to 7. The small AAV2-specific RNAs sR-108, sR-271 and sR-1862 were aligned to the corresponding sites of AAV serotypes 1 to 7. Gray areas represent conserved nucleotide regions; white areas indicate aberrations to the AAV2 small RNAs in the given AAV serotype.
(TIF)

S2 Fig. Effect of lytic AAV2 infection on expression levels of cellular miRNAs. (A) The expression levels of 979 described cellular miRNAs were analyzed by small RNA-Seq analysis in AAV2/Ad2 co-infected versus uninfected HeLa cells at 27 hpi. Dots near the red line represent equally expressed miRNAs in both data sets; the solid gray lines mark 2-fold and the dashed gray lines 5-fold up- or down-regulation of miRNAs. The most highly up- or down-regulated miRNAs were highlighted and specified. Cellular miRNAs below a level of 100 read counts, within the dashed square, were considered as not representative. (B-D) Expression level of 979 described cellular miRNAs in (B) AAV2/HSV1 co-infected versus uninfected HeLa cells 8 hpi; (C) Ad2 infected versus uninfected HeLa cells 27 hpi; (D) HSV1 infected versus uninfected HeLa cells 8 hpi, displayed as in (A).
(TIF)

S1 Table. Oligonucleotide sequences used for Northern blot analysis.
(DOC)

S2 Table. Small RNA-Seq analysis—Assignment of the reads to the species from which they originate (gradually allowing zero, one and two mismatches per read).
(DOCX)

S3 Table. Small AAV2 RNA list. Listed are the top hundred AAV2-specific small RNA candidates detected by small RNA-Seq analysis in the presence of the helper viruses Ad2 and HSV1, respectively. Given are genomic location, read counts, median read length and sequence of the respective small AAV RNAs.
(XLS)

Acknowledgments

The authors thank Eva-Maria Hammer of the Heilbronn lab for help with cell culture and virus propagation.

Author Contributions

Conceptualization: RH CS.

Data curation: CS AGD.

Formal analysis: AGD.

Funding acquisition: RH CS.

Investigation: CS MM MS AGD.

Methodology: RH CS WC AGD.

Project administration: RH CS AGD.

Resources: WC AGD.

Software: AGD.

Supervision: RH.

Validation: CS RH SW AGD WC.

Visualization: CS MM AGD.

Writing – original draft: CS MM RH.

Writing – review & editing: CS RH SW.

References

1. Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, et al. Site-specific integration by adeno-associated virus. Proc Natl Acad Sci U S A. 1990; 87(6):2211–5. PMID: 2156265
2. Hüser 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. doi: 10.1128/JVI.01356-14 PMID: 25031342; PubMed Central PMCID: PMC4178796.
3. Yang J, Zhou W, Zhang Y, Zidon T, Ritchie T, Engelhardt JF. Concatamerization of adeno-associated virus circular genomes occurs through intermolecular recombination. J Virol. 1999; 73(11):9468–77. PMID: 10516055; PubMed Central PMCID: PMC112981.
4. Blacklow NR, Hoggan MD, Rowe WP. Isolation of adenovirus-associated viruses from man. Proc Natl Acad Sci U S A. 1967; 58(4):1410–5. Epub 1967/10/01. PMID: 4295829; PubMed Central PMCID: PMCPMC223939.
5. Buller RM, Janik JE, Sebring ED, Rose JA. Herpes simplex virus types 1 and 2 completely help adeno-virus-associated virus replication. J Virol. 1981; 40(1):241–7. PMID: 6270377; PubMed Central PMCID: PMCPMC256613.
6. Georg-Fries B, Biederlack S, Wolf J, zur Hausen H. Analysis of proteins, helper dependence, and ser-oepidemiology of a new human parvovirus. Virology. 1984; 134(1):64–71. PMID: 6200995.
7. McPherson RA, Rosenthal LJ, Rose JA. Human cytomegalovirus completely helps adeno-associated virus replication. Virology. 1985; 147(1):217–22. PMID: 2998066
8. Thomson BJ, Weindler FW, Gray D, Schwaab V, Heilbronn R. Human herpesvirus 6 (HHV6) is a helper virus for adeno-associated virus type 2 (AAV2) and the rep gene homologue in HHV6 can mediate AAV-2 DNA replication and regulate gene expression. Virology. 1994; 204(1):304–11. PMID: 8091661
9. Srivastava A, Lusby EW, Berns KI. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol. 1983; 45(2):555–64. PMID: 6300419; PubMed Central PMCID: PMC256449.
10. Sonntag F, Schmidt K, Kleinschmidt JA. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc Natl Acad Sci U S A. 2010; 107(22):10220–5. Epub 2010/05/19. doi: 10.1073/pnas.1001673107 PMID: 20479244; PubMed Central PMCID: PMC2890453.
11. Stutika C, Gogol-Doring A, Botschen L, Mietzsch M, Weger S, Feldkamp M, et al. A Comprehensive RNA-Seq Analysis of the Adeno-Associated Virus type 2 Transcriptome Reveals Novel AAV Tran-scripts, Splice Variants, and Derived Proteins. J Virol. 2015. doi: 10.1128/JVI.02750-15 PMID: 26559843.
12. Weger S, Hammer E, Gonsior M, Stutika C, Heilbronn R. A regulatory element near the 3' -end of the AAV rep gene inhibits adenovirus replication in cis by means of p40 promoter-associated short tran-scripts. J Virol. 2016. doi: 10.1128/JVI.03120-15 PMID: 26842470.

13. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116(2):281–97. PMID: 14744438.

14. Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol. 2005; 6 (5):376–85. doi: 10.1038/nrm1644 PMID: 15852042.

15. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009; 136 (4):642–55. doi: 10.1016/j.cell.2009.01.035 PMID: 19239886; PubMed Central PMCID: PMC2675692.

16. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136(2):215–33. doi: 10.1016/j.cell.2009.01.002 PMID: 19167326; PubMed Central PMCID: PMC3794896.

17. Keam SP, Hutvagner G. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. Life. 2015; 5(4):1638–51. doi: 10.3390/life5041638 PMID: 26703738; PubMed Central PMCID: PMCPMC4695841.

18. Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF. Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics. 2009; 25(18):2298–301. doi: 10.1093/bioinformatics/btp419 PMID: 19584066.

19. Yan BX, Ma JX. Promoter-associated RNAs and promoter-targeted RNAs. Cell Mol Life Sci. 2012; 69 (17):2833–42. doi: 10.1007/s00018-012-0953-1 PMID: 22415323.

20. Kincaid RP, Sullivan CS. Virus-encoded microRNAs: an overview and a look to the future. PLoS Pathog. 2012; 8(12):e1003018. doi: 10.1371/journal.ppat.1003018 PMID: 23308061; PubMed Central PMCID: PMC3534370.

21. Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, Xie M, et al. Viral noncoding RNAs: more surprises. Genes Dev. 2015; 29(6):567–84. doi: 10.1101/gad.259077.115 PMID: 25792595; PubMed Central PMCID: PMC4378190.

22. Andersson MG, Haasnoot PC, Xu N, Berenjian S, Berkhout B, Akusjarvi G. Suppression of RNA interference by adenovirus virus-associated RNA. J Virol. 2005; 79(15):9556–65. doi: 10.1128/JVI.79.15.9556-9565.2005 PMID: 16014917; PubMed Central PMCID: PMC1181602.

23. Bellutti F, Kauer M, Kneidinger D, Lion T, Klein R. Identification of RISC-associated adenoviral microRNAs, a subset of their direct targets, and global changes in the targetome upon lytic adenovirus 5 infection. J Virol. 2015; 89(3):1608–27. doi: 10.1128/JVI.02336-14 PMID: 25410853; PubMed Central PMCID: PMC4300742.

24. Aparicio O, Carnero E, Abad X, Razquin N, Guruceaga E, Segura V, et al. Adenovirus VA RNA-derived miRNAs target cellular genes involved in cell growth, gene expression and DNA repair. Nucleic Acids Res. 2010; 38(3):750–63. doi: 10.1093/nar/gkp1028 PMID: 19933264; PubMed Central PMCID: PMCPMC2817457.

25. Umbach JL, Kramer MF, Jurak I, Karnowski HW, Coen DM, Cullen BR. MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. Nature. 2008; 454(7205):780–3. doi: 10.1038/nature07103 PMID: 18596690; PubMed Central PMCID: PMC2666538.

26. Flores O, Nakayama S, Whisnant AW, Javanbakht H, Cullen BR, Bloom DC. Mutational inactivation of herpes simplex virus 1 microRNAs identifies viral mRNA targets and reveals phenotypic effects in culture. J Virol. 2013; 87(12):6589–603. doi: 10.1128/JVI.00504-13 PMID: 23536669; PubMed Central PMCID: PMC3676078.

27. Jurak I, Kramer MF, Mellor JC, van Lint AL, Roth FP, Knipe DM, et al. Numerous conserved and divergent microRNAs expressed by herpes simplex viruses 1 and 2. J Virol. 2010; 84(9):4659–72. doi: 10.1128/JVI.02725-09 PMID: 20181707; PubMed Central PMCID: PMC2863732.

28. Meshesha MK, Veksler-Lublinsky I, Isakov O, Reichenstein I, Shomron N, Kedem K, et al. The microRNA Transcriptome of Human Cytomegalovirus (HCMV). Open Virol J. 2012; 6:38–48. doi: 10.2174/1874357901206010038 PMID: 22715351; PubMed Central PMCID: PMC3377890.

29. Winter K, von Kietzell K, Heilbronn R, Pozzuto T, Fechner H, Weger S. Roles of E4orf6 and VA I RNA in adenovirus-mediated stimulation of human parvovirus B19 DNA replication and structural gene expression. J Virol. 2012; 86(9):5099–109. doi: 10.1128/JVI.06991-11 PMID: 22357277; PubMed Central PMCID: PMC3347340.

30. Heilbronn R, Bürkle A, Stephan S, zur Hausen H. The adeno-associated virus *rep* gene suppresses herpes simplex virus-induced DNA-amplification. J Virol. 1990; 64(6):3012–8. PMID: 2159559

31. Mietzsch M, Broecker F, Reinhardt A, Seeberger PH, Heilbronn R. Differential adeno-associated virus serotype-specific interaction patterns with synthetic heparins and other glycans. J Virol. 2014; 88 (5):2991–3003. doi: 10.1128/JVI.03371-13 PMID: 24371066; PubMed Central PMCID: PMC3958061.

32. Weindler FW, Heilbronn R. A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. J Virol. 1991; 65(5):2476–83. PMID: 1850024; PubMed Central PMCID: PMC240602.

33. Harris CJ, Molnar A, Muller SY, Baulcombe DC. FDF-PAGE: a powerful technique revealing previously undetected small RNAs sequestered by complementary transcripts. Nucleic Acids Res. 2015; 43 (15):7590–9. doi: 10.1093/nar/gkv604 PMID: 26071954; PubMed Central PMCID: PMCPMC4551911.

34. Damm K, Bach S, Muller KM, Klug G, Burenina OY, Kubareva EA, et al. Improved Northern Blot Detection of Small RNAs Using EDC Crosslinking and DNA/LNA Probes. Methods Mol Biol. 2015; 1296:41–51. doi: 10.1007/978-1-4939-2547-6_5 PMID: 25791589.

35. Kim SW, Li Z, Moore PS, Monaghan AP, Chang Y, Nichols M, et al. A sensitive non-radioactive northern blot method to detect small RNAs. Nucleic Acids Res. 2010; 38(7):e98. doi: 10.1093/nar/gkp1235 PMID: 20081203; PubMed Central PMCID: PMC2853138.

36. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014; 42(Database issue):D68–73. doi: 10.1093/nar/gkt1181 PMID: 24275495; PubMed Central PMCID: PMCPMC3965103.

37. Langmead B. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics. 2010;Chapter 11:Unit 11 7. doi: 10.1002/0471250953.bi1107s32 PMID: 21154709; PubMed Central PMCID: PMCPMC3010897.

38. Jing XJ, Kalman-Maltese V, Cao X, Yang Q, Trempe JP. Inhibition of adenovirus cytotoxicity, replication, and E2a gene expression by adeno-associated virus. Virology. 2001; 291(1):140–51. doi: 10.1006/viro.2001.1192 PMID: 11878883.

39. Timpe JM, Verrill KC, Trempe JP. Effects of adeno-associated virus on adenovirus replication and gene expression during coinfection. J Virol. 2006; 80(16):7807–15. doi: 10.1128/JVI.00198-06 PMID: 16873238; PubMed Central PMCID: PMC1563798.

40. Zhao H, Chen M, Tellgren-Roth C, Pettersson U. Fluctuating expression of microRNAs in adenovirus infected cells. Virology. 2015; 478:99–111. doi: 10.1016/j.virol.2015.01.033 PMID: 25744056.

41. Aparicio O, Razquin N, Zaratiegui M, Narvaiza I, Fortes P. Adenovirus virus-associated RNA is processed to functional interfering RNAs involved in virus production. J Virol. 2006; 80(3):1376–84. doi: 10.1128/JVI.80.3.1376-1384.2006 PMID: 16415015; PubMed Central PMCID: PMCPMC1346933.

42. Furuse Y, Ornelles DA, Cullen BR. Persistently adenovirus-infected lymphoid cells express microRNAs derived from the viral VAI and especially VAII RNA. Virology. 2013; 447(1–2):140–5. doi: 10.1016/j.virol.2013.08.024 PMID: 24210108; PubMed Central PMCID: PMCPMC3825519.

43. Munson DJ, Burch AD. A novel miRNA produced during lytic HSV-1 infection is important for efficient replication in tissue culture. Arch Virol. 2012; 157(9):1677–88. doi: 10.1007/s00705-012-1345-4 PMID: 22661375.

44. Griffiths-Jones S. miRBase: the microRNA sequence database. Methods Mol Biol. 2006; 342:129–38. doi: 10.1385/1-59745-123-1:129 PMID: 16957372.

45. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015; 4. doi: 10.7554/eLife.05005 PMID: 26267216; PubMed Central PMCID: PMCPMC4532895.

46. Stilwell JL, Samulski RJ. Role of viral vectors and virion shells in cellular gene expression. Mol Ther. 2004; 9(3):337–46. doi: 10.1016/j.ymthe.2003.11.007 PMID: 15006600.

47. Ferrari FK, Samulski T, Shenk T, Samulski RJ. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J Virol. 1996; 70(5):3227–34. PMID: 8627803; PubMed Central PMCID: PMCPMC190186.

48. Wang L, Yin Z, Wang Y, Lu Y, Zhang D, Srivastava A, et al. Productive life cycle of adeno-associated virus serotype 2 in the complete absence of a conventional polyadenylation signal. J Gen Virol. 2015; 96(9):2780–7. doi: 10.1099/jgv.0.000229 PMID: 26297494.

49. Son KN, Liang Z, Lipton HL. Double-Stranded RNA Is Detected by Immunofluorescence Analysis in RNA and DNA Virus Infections, Including Those by Negative-Stranded RNA Viruses. J Virol. 2015; 89 (18):9383–92. doi: 10.1128/JVI.01299-15 PMID: 26136565; PubMed Central PMCID: PMCPMC4542381.

50. Flotte TR, Afione SA, Solow R, Drumm ML, Markakis D, Guggino WB, et al. Expression of the cystic fibrosis transmembrane conductance regulator from a novel adeno-associated virus promoter. J Biol Chem. 1993; 268(5):3781–90. PMID: 7679117.

51. Haberman RP, McCown TJ, Samulski RJ. Novel transcriptional regulatory signals in the adeno-associated virus terminal repeat A/D junction element. J Virol. 2000; 74(18):8732–9. PMID: 10954575; PubMed Central PMCID: PMC116385.

52. Qiu J, Nayak R, Tullis GE, Pintel DJ. Characterization of the transcription profile of adeno-associated virus type 5 reveals a number of unique features compared to previously characterized adeno-associated viruses. J Virol. 2002; 76(24):12435–47. Epub 2002/11/20. PMID: 12438569; PubMed Central PMCID: PMCPMC136721.

53. Rother S, Meister G. Small RNAs derived from longer non-coding RNAs. Biochimie. 2011; 93 (11):1905–15. doi: 10.1016/j.biochi.2011.07.032 PMID: 21843590.

54. Brister JR, Muzyczka N. Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. J Virol. 1999; 73(11):9325–36. PMID: 10516041; PubMed Central PMCID: PMCPMC112967.

55. Brister JR, Muzyczka N. Mechanism of Rep-mediated adeno-associated virus origin nicking. J Virol. 2000; 74(17):7762–71. PMID: 10933682; PubMed Central PMCID: PMCPMC112305.

56. Pereira DJ, McCarty DM, Muzyczka N. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. J Virol. 1997; 71 (2):1079–88. PMID: 8995628; PubMed Central PMCID: PMC191159.

57. He S, Yang Z, Skogerbo G, Ren F, Cui H, Zhao H, et al. The properties and functions of virus encoded microRNA, siRNA, and other small noncoding RNAs. Crit Rev Microbiol. 2008; 34(3–4):175–88. doi: 10.1080/10408410802482008 PMID: 18972284.

58. Lin YT, Kincaid RP, Arasappan D, Dowd SE, Hunicke-Smith SP, Sullivan CS. Small RNA profiling reveals antisense transcription throughout the KSHV genome and novel small RNAs. RNA. 2010; 16 (8):1540–58. doi: 10.1261/rna.1967910 PMID: 20566670; PubMed Central PMCID: PMC2905754.

59. Tuddenham L, Jung JS, Chane-Woon-Ming B, Dolken L, Pfeffer S. Small RNA deep sequencing identifies microRNAs and other small noncoding RNAs from human herpesvirus 6B. J Virol. 2012; 86 (3):1638–49. doi: 10.1128/JVI.05911-11 PMID: 22114334; PubMed Central PMCID: PMC3264354.

60. Hauswirth WW, Berns KI. Origin and termination of adeno-associated virus DNA replication. Virology. 1977; 78(2):488–99. PMID: 867815.

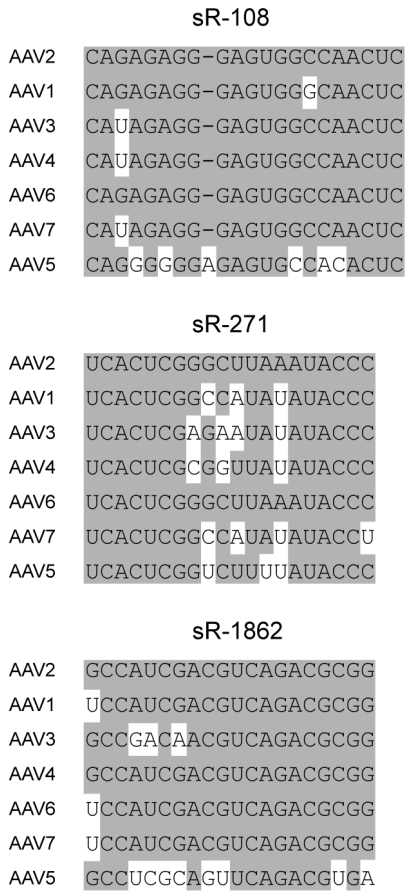
61. King JA, Dubielzig R, Grimm D, Kleinschmidt JA. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. Embo J. 2001; 20(12):3282–91. PMID: 11406604

62. Wang XS, Ponnazhagan S, Srivastava A. Rescue and replication of adeno-associated virus type 2 as well as vector DNA sequences from recombinant plasmids containing deletions in the viral inverted terminal repeats: selective encapsidation of viral genomes in progeny virions. J Virol. 1996; 70(3):1668–77. PMID: 8627687; PubMed Central PMCID: PMCPMC189990.

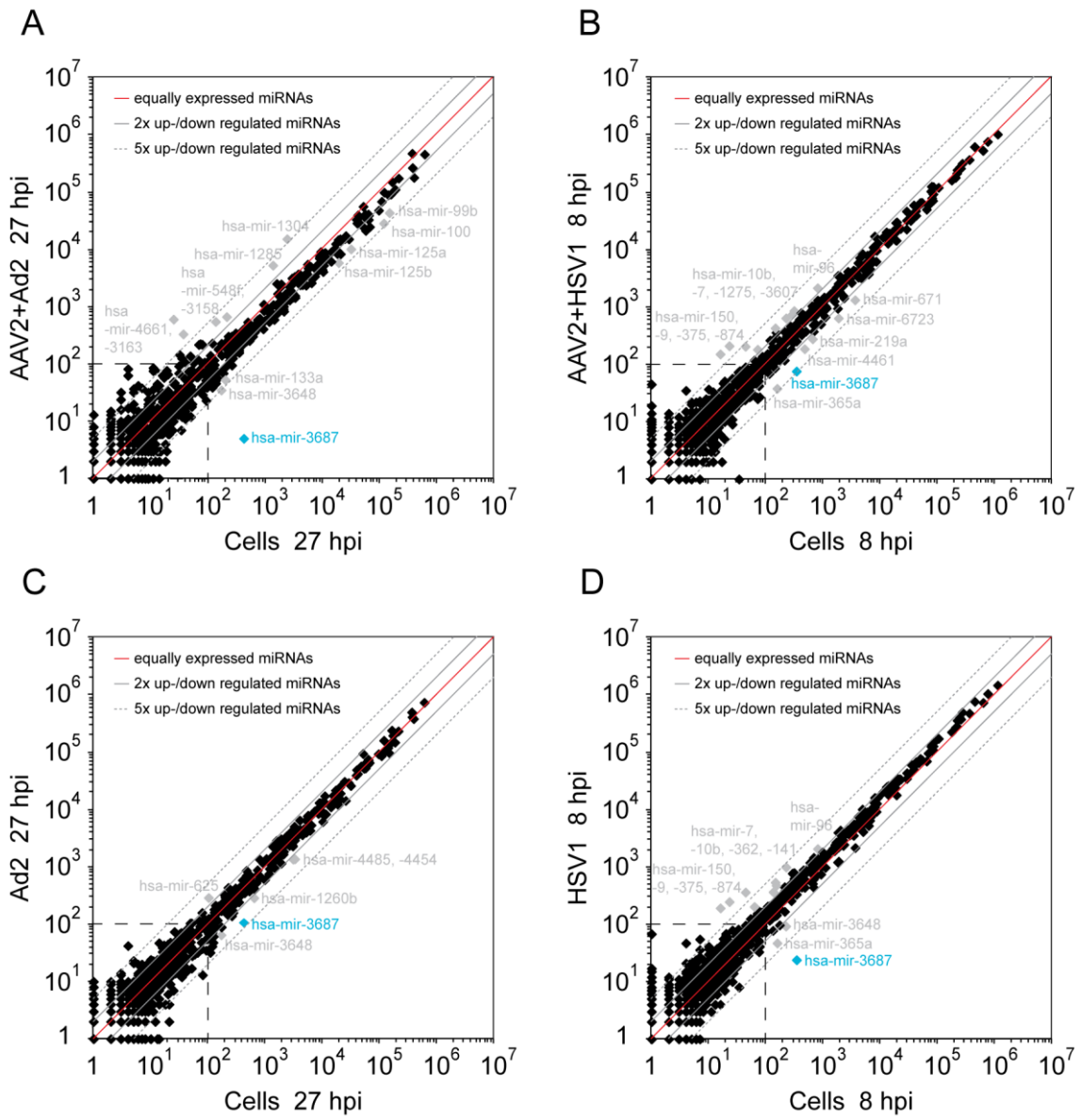
63. Wang XS, Qing K, Ponnazhagan S, Srivastava A. Adeno-associated virus type 2 DNA replication in vivo: mutation analyses of the D sequence in viral inverted terminal repeats. J Virol. 1997; 71(4):3077–82. PMID: 9060669; PubMed Central PMCID: PMCPMC191438.

64. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, et al. Divergent transcription from active promoters. Science. 2008; 322(5909):1849–51. doi: 10.1126/science.1162253 PMID: 19056940; PubMed Central PMCID: PMC2692996.

65. Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, et al. Tiny RNAs associated with transcription start sites in animals. Nat Genet. 2009; 41(5):572–8. doi: 10.1038/ng.312 PMID: 19377478.


sR-108
AAV2 CAGAGAGG-GAGUGGCCAACUC
AAV1 CAGAGAGG-GAGUGGGCAACUC
AAV3 CAUAGAGG-GAGUGGCCAACUC
AAV4 CAUAGAGG-GAGUGGCCAACUC
AAV6 CAGAGAGG-GAGUGGCCAACUC
AAV7 CAUAGAGG-GAGUGGCCAACUC
AAV5 CAGGGGGGAGAGUGCCACACUC
sR-271
AAV2 UCACUCGGGCUUAAAUACCC
AAV1 UCACUCGGCCAUAUAUACCC
AAV3 UCACUCGAGAAUAUAUACCC
AAV4 UCACUCGCGGUUAUAUACCC
AAV6 UCACUCGGGCUUAAAUACCC
AAV7 UCACUCGGCCAUAUAUACCU
AAV5 UCACUCGGUCUUUUAUACCC
sR-1862
AAV2 GCCAUCGACGUCAGACGCGG
AAV1 UCCAUCGACGUCAGACGCGG
AAV3 GCCGACAACGUCAGACGCGG
AAV4 GCCAUCGACGUCAGACGCGG
AAV6 UCCAUCGACGUCAGACGCGG
AAV7 UCCAUCGACGUCAGACGCGG
AAV5 GCCUCGCAGUUCAGACGUGA

S1 Fig


A
B
C
D
equally expressed miRNAs
2x up-/down regulated miRNAs
5x up-/down regulated miRNAs
AAV2+Ad2 27 hpi
Cells 27 hpi
AAV2+HSV1 8 hpi
Cells 8 hpi
Ad2 27 hpi
HSV1 8 hpi
hsa-mir-1304
hsa-mir-1285
hsa-mir-548f, -3158
hsa-mir-4661, -3163
hsa-mir-99b
hsa-mir-100
hsa-mir-125a
hsa-mir-125b
hsa-mir-133a
hsa-mir-3648
hsa-mir-3687
hsa-mir-10b, -7, -1275, -3607
hsa-mir-96
hsa-mir-671
hsa-mir-6723
hsa-mir-219a
hsa-mir-4461
hsa-mir-365a
hsa-mir-150, -9, -375, -874
hsa-mir-625
hsa-mir-4485, -4454
hsa-mir-1260b
hsa-mir-7, -10b, -362, -141

S2 Fig

S1 Table. Oligonucleotide sequences used for Northern blot analysis.

sRNA name	Probe sequence 5' → 3' (source)
sR-108	5' GAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACT 3' (DNA)
sR-108	5' GAGTTGGCC+ACTCCC+T+CTCT+GCGCG+C+TCGC+TCGC+TCACT 3' (DNA/LNA)
sR-1862	5' GAAGCTTCCGCGTCTGACGTCGATGGC 3' (DNA)
sR-1862	5' GAAGCTTC+C+GCGTC+TGACG+T+CGA+TGGC 3' (DNA/LNA)
sR-271	5' GGGTATTTAAGCCCGAGTGA 3' (DNA)
hsa-let-7a-1	5' AACTATACAACCTACTACCTCA 3' (DNA)
U6-snRNA	5' CACGAATTTGCGTGTCATCCTT 3' (DNA)

Underlined is the exact complementary sequence of the corresponding sRNA identified in the small RNA-Seq analysis.

S2 Table. Small RNA-Seq - Assignment of the reads to the species from which they originate (gradually allowing zero, one and two mismatches per read).

Data set	Total no. of reads	No. (%) of reads ≥ 16 nt	No. (%) of human reads[a]	No. (%) of small reads assigned to:			No. (%) of unknown reads[b]
				AAV2	Ad2	HSV1	
Cells (27 hpi)	19,395,459	17,429,392 (89.9)	16,672,836 (95.7)	22 (<0.1)	621 (<0.1)	10 (<0.1)	755,903 (4.3)
AAV2 (27 hpi)	18,085,432	16,228,519 (89.7)	15,441,982 (95.2)	1,400 (<0.1)	1,703 (<0.1)	14 (<0.1)	783,420 (4.8)
AAV2 + Ad2 (27 hpi)	21,333,021	18,814,505 (88.2)	13,008,523 (69.1)	328,917 (1.7)	4,663,151 (24.8)	20 (<0.1)	813,894 (4.3)
Ad2 (27 hpi)	24,841,074	24,202,341 (97.4)	14,184,796 (58.6)	29 (<0.1)	9,214,482 (38.1)	23 (<0.1)	803,011 (3.3)
Cells (8 hpi)	21,851,591	20,985,998 (96.0)	19,964,098 (95.1)	24 (<0.1)	544 (<0.1)	18 (<0.1)	1,021,314 (4.9)
AAV2 (8 hpi)	20,375,107	20,022,269 (98.3)	18,992,333 (94.9)	176 (<0.1)	791 (<0.1)	16 (<0.1)	1,028,953 (5.1)
AAV2 + HSV1 (8 hpi)	23,031,470	20,904,023 (90.8)	19,108,356 (91.4)	158,561 (0.8)	7,867 (<0.1)	127,303 (0.6)	1,501,936 (7.2)
HSV1 (8 hpi)	22,391,766	20,895,686 (93.3)	19,600,347 (93.8)	15 (<0.1)	616 (<0.1)	145,417 (0.7)	1,149,291 (5.5)

[a]Included are reads mapped to the human genome, transcriptome and miRNAs.

[b]Included are reads ≥ 16 nt that were unmappable during sequence mapping allowing two mismatches.

S3 Table. Top 100 AAV2-specific small RNA candidates detected by small RNA-Seq analysis in the presence of the helper viruses Ad2 and HSV1, respectively.

small RNA	*alternative locations*	*AAV2/Ad read counts*	*AAV2/HSV read counts*	*AAV2/Ad median length*	*AAV2/HSV median length*	*sequence*
sR-271-		12.737	21	20	19	TCACTCGGGCTTAAATACCC
sR-108+	sR-4572-	12.004	6	21	21	CAGAGAGGGAGTGGCCAACTC
sR-272-		5.305	15	20	19	CTCACTCGGGCTTAAATACC
sR-141-	sR-4539+	4.600	91	22	22	ACCCCTAGTGATGGAGTTGGCC
sR-1862+		3.780	860	19	19	GCCATCGACGTCAGACGCG
sR-241-		4.323	6	20	21	CATGGTGTCGCAAAATGTCGC
sR-244-		3.926	13	20	20	CCACATGGTGTCGCAAAATG
sR-270-		3.900	23	21	17	CACTCGGGCTTAAATACCCAG
sR-3703+		649	3.169	28	28	TCGTGGACTGGAGCTACCAAGTACCAC C
sR-262-		3.709	0	20		CTTAAATACCCAGCGTGACC
sR-274-		3.467	3	22	22	TGCTCACTCGGGCTTAAATACC
sR-1+	sR-125- sR-4555+ sR-4679-	3.170	38	22	22	TTGGCCACTCCCTCTCTGCGCG
sR-245-		2.859	1	21	21	ACCACATGGTGTCGCAAAATG
sR-1865+		2.271	558	22	24	ATCGACGTCAGACGCGGAAGCTTC
sR-3704+		949	1.879	25	26	CGTGGACTGGAGCTACCAAGTACCAC
sR-126-	sR-4554+	2.667	15	21	25	GTTGGCCACTCCCTCTCTGCGCGCT
sR-269-		2.619	6	18	17	ACTCGGGCTTAAATACCC
sR-303+		2.388	150	18	18	TTTGAACGCGCAGCCGCC
sR-36-	sR-90+ sR-4590- sR-4644+	2.488	11	22	22	AGTGAGCGAGCGAGCGCGCAGA
sR-140-	sR-4540+	2.374	107	21	21	CCCCTAGTGATGGAGTTGGCC
sR-105-	sR-105- sR-4575+ sR-4659-	2.426	21	21	22	CGCTCGCTCGCTCACTGAGGCC
sR-2001+		268	1.924	25	26	TTCACTCACGGACAGAAAGACTGTTT
sR-104+	sR-104+ sR-4576- sR-4658+	2.172	2	21	21	CGCGCAGAGAGGGAGTGGCCA
sR-1866+		357	1.712	23	23	TCGACGTCAGACGCGGAAGCTTC
sR-333+		2.037	18	22	21	TACGAGATTGTGATTAAGGTCC
sR-244+		2.007	0	21		GTCACGCTGGGTATTTAAGCC
sR-330+		1.792	15	22	27	TTTTACGAGATTGTGATTAAGGTCCCC
sR-1853+		999	737	34	34	AGTTGCGCAGCCATCGACGTCAGACGC GGAAGCT
sR-172-		1.542	26	17	17	TAATTCACGTCACGACT
sR-4329+		623	945	21	21	TGTGGACTTTACTGTGGACAC
sR-106+	sR-106+ sR-4574- sR-4660+	1.565	0	22		CGCAGAGAGGGAGTGGCCAACT
sR-2229+		528	1.005	25	28	TTGGCTCGAGGACACTCTCTCTGAAGGA
sR-290+		1.438	91	22	29	TTGAAGCGGGAGGTTTGAACGCGCAGC

						CG
sR-302+		1.449	63	18	19	GTTTGAACGCGCAGCCGCC
sR-139-	sR-4541+	1.404	104	20	20	CCCTAGTGATGGAGTTGGCC
sR-109+	sR-4571-	1.494	0	20		AGAGAGGGAGTGGCCAACTC
sR-109-	sR-109- sR-4571+ sR-4663-	1.460	10	22	22	TGCGCGCTCGCTCGCTCACTGA
sR-2000+		175	1.275	27	27	CTTCACTCACGGACAGAAAGACTGTTT
sR-2354+		500	921	25	25	AGTACCTCGGACCCTTCAACGGACT
sR-138-	sR-4542+	1.342	59	18	19	CCTAGTGATGGAGTTGGCC
sR-1860+		649	728	21	26	CAGCCATCGACGTCAGACGCGGAAGC
sR-282-		1.328	4	22	22	ACCCTGCGTGCTCACTCGGGCT
sR-2719+		501	830	17	17	TTTGGTCAGACTGGAGA
sR-175-		1.295	10	20	20	ACGTAATTCACGTCACGACT
sR-1869+		480	775	18	18	ACGTCAGACGCGGAAGCT
sR-1861+		539	685	22	24	AGCCATCGACGTCAGACGCGGAAG
sR-173-		1.206	14	19	19	GTAATTCACGTCACGACTC
sR-242-		1.210	10	20	16	ACATGGTGTCGCAAAATGTC
sR-1863+		352	863	19	24	CCATCGACGTCAGACGCGGAAGCT
sR-28-	sR-98+ sR-4582- sR-4652+	1.199	13	22	22	AGCGAGCGCGCAGAGAGGGAGT
sR-3625+		296	901	21	21	AGGAACTGGCTTCCTGGACCC
sR-261-		1.188	5	22	19	TTAAATACCCAGCGTGACCACA
sR-103-	sR-103- sR-4577+ sR-4657-	1.096	87	19	19	CTCGCTCGCTCACTGAGGC
sR-2228+		175	944	30	31	ATTGGCTCGAGGACACTCTCTCTGAAGGAAT
sR-4385+		280	749	23	23	CCAGATACCTGACTCGTAATCTG
sR-1867+		727	300	21	22	CGACGTCAGACGCGGAAGCTTC
sR-155+		1.006	1	21	21	GAGTCGTGACGTGAATTACGT
sR-374+		798	200	22	22	TCTGCCCGGCATTTCTGACAGC
sR-1474+		121	861	16	17	AGGTCGTGGAGTCGGCC
sR-2579+		160	817	27	27	TTGAACCTCTGGGCCTGGTTGAGGAAC
sR-3611+		260	707	24	22	TTCGGGACCAGTCTAGGAACTGGC
sR-1871+		760	204	18	18	GTCAGACGCGGAAGCTTC
sR-2234+		161	800	23	23	TCGAGGACACTCTCTCTGAAGGA
sR-874+		652	301	20	29	GTAAACGGTTGGTGGCGCAGCATCTGACG
sR-2629+		286	657	23	23	AAGAGGCCGGTAGAGCACTCTCC
sR-43+	sR-4597+	911	20	32	32	GGGCGACCAAAGGTCGCCCGACGCCCGGGCTT
sR-1859+		610	306	21	26	GCAGCCATCGACGTCAGACGCGGAAG
sR-167-		879	34	21	22	CACGTCACGACTCCACCCCTCC
sR-2235+		334	540	22	21	CGAGGACACTCTCTCTGAAGGA
sR-142-	sR-4538+	818	47	22	23	AACCCCTAGTGATGGAGTTGGCC
sR-3344+		249	616	26	26	ACAACGGGAGTCAGGCAGTAGGACGC

sR-3920+		197	636	18	18	CGGAGCAGTATGGTTCTG
sR-305+		784	44	21	18	TGAACGCGCAGCCGCCATGCC
sR-34-	sR-92+ sR-4588- sR-4646+	809	14	22	22	TGAGCGAGCGAGCGCGCAGAGA
sR-100+	sR-100+ sR-4580- sR-4654+	783	37	21	19	CGAGCGCGCAGAGAGGGAGTG
sR-136-	sR-4544+	754	62	21	17	TAGTGATGGAGTTGGCCACTC
sR-268-		800	12	17	26	CTCGGGCTTAAATACCCAGCGTGACC
sR-169-		779	32	21	22	TTCACGTCACGACTCCACCCCT
sR-199-		792	15	21	21	ATACAGGACCTCCCTAACCCT
sR-445-		788	6	21	22	TCAGAATCTGGCGGCAACTCCC
sR-124-	sR-124- sR-4556+ sR-4678-	765	28	23	29	TGGCCACTCCCTCTCTGCGCGCTCGCTCG
sR-1280+		213	567	28	28	CACGAAAAAGTTCGGCAAGAGGAACACC
sR-43-	sR-83+ sR-4597- sR-4637+	759	20	22	24	CGGCCTCAGTGAGCGAGCGAGCGC
sR-3119+		141	635	21	21	TCCGACCCAAGAGACTCAACT
sR-211-		769	6	21	20	ACGTGACCTCTAATACAGGAC
sR-2991+		246	527	18	18	ATCAGGAGCCTCGAACGA
sR-4633-	sR-4633-	702	71	18	18	GGGCAAAGCCCGGGCGTC
sR-174-		751	18	19	24	CGTAATTCACGTCACGACTCCACC
sR-335+		726	42	22	22	CGAGATTGTGATTAAGGTCCCC
sR-4534+		758	2	22	22	AAGGAACCCCTAGTGATGGAGT
sR-170-		735	11	21	25	ATTCACGTCACGACTCCACCCCTCC
sR-287+		615	128	31	32	ATTTTGAAGCGGGAGGTTTGAACGCGCAGCCG
sR-107-	sR-107- sR-4573+ sR-4661-	693	26	21	23	CGCGCTCGCTCGCTCACTGAGGC
sR-196-		618	96	18	18	CAGGACCTCCCTAACCCT
sR-212-		698	14	21	18	CACGTGACCTCTAATACAGGA
sR-137-	sR-4543+	666	44	18	18	CTAGTGATGGAGTTGGCC
sR-331+		702	6	22	29	TTTACGAGATTGTGATTAAGGTCCCCAGC
sR-240-		697	0	21		ATGGTGTCGCAAAATGTCGCA
sR-4363+		165	522	27	26	TCAGAGCCTCGCCCCATTGGCACCAGA
sR-4328+		305	361	22	22	ATGTGGACTTTACTGTGGACAC

2.3 A Regulatory Element Near the 3' End of the Adeno-Associated Virus rep Gene Inhibits Adenovirus Replication in cis by Means of p40 Promoter-Associated Short Transcripts

Authors: Stefan Weger, Eva-Maria Hammer, Melanie Gonsior, Catrin Stutika, Regine Heilbronn

Year: 2016

Journal: Journal of Virology 90 (8): 3981-3993

2.3.1 Contribution to the Publication

This study was conducted to further evaluate the long known inhibitory effect of AAV on the replication of helper adenovirus. Based on a preceding study by Weger *et al*. (181), in which a *cis* inhibitory sequence in the 3' part of the *rep* gene was shown to be involved in this effect, mutational studies were performed to analyze the contribution of individual sequence elements located at the 3' end of *rep*. This study is of specific interest for AAV vector production systems based on recombinant adenovirus/AAV hybrid vectors for the generation of rAAVs.

For this publication my contribution was to help evaluate the transcriptional profile of the AAV *rep*-expressing adenovirus. My small RNA-Seq data showed a unique hotspot of small AAV-specific transcripts located exactly in the genomic region of interest. By further analyzing the hotspot *in silico*, size and read length distributions showed features of small promoter-associated RNAs rather than bona-fide miRNAs. By Ago2 co-immunoprecipitation I further confirmed that these small transcripts do indeed not represent canonical miRNAs.

2.3.2 Article

DOI: http://dx.doi.org/10.1128/JVI.03120-15

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

A Regulatory Element Near the 3′ End of the Adeno-Associated Virus *rep* Gene Inhibits Adenovirus Replication in *cis* by Means of p40 Promoter-Associated Short Transcripts

Stefan Weger, Eva Hammer, Melanie Gonsior, Catrin Stutika, Regine Heilbronn

Institute of Virology, Charité-University Medicine Berlin, Berlin, Germany

ABSTRACT

Adeno-associated virus (AAV) has long been known to inhibit helper adenovirus (Ad) replication independently of AAV Rep protein expression. More recently, replication of Ad serotype 5 (Ad5)/AAV serotype 2 (AAV-2) hybrid vectors was shown to be inhibited in *cis* by a sequence near the 3′ end of AAV *rep*, termed the Rep inhibition sequence for adenoviral replication (RIS-Ad). RIS-Ad functions independently of Rep protein expression. Here we demonstrate that inhibition of adenoviral replication by RIS-Ad requires an active AAV p40 promoter and the 5′ half of the intron. In addition, Ad inhibition is critically dependent on the integrity of the p40 transcription start site (TSS) leading to short p40-associated transcripts. These do not give rise to effector molecules capable of inhibiting adenoviral replication in *trans*, like small polypeptides or microRNAs. Our data point to an inhibitory mechanism in which RNA polymerase II (Pol II) pauses directly downstream of the p40 promoter, leading to interference of the stalled Pol II transcription complex with the adenoviral replication machinery. Whereas inhibition by RIS-Ad is mediated exclusively in *cis*, it can be overcome by providing a replication-competent adenoviral genome in *trans*. Moreover, the inhibitory effect of RIS-Ad is not limited to AAV-2 but could also be shown for the corresponding regions of other AAV serotypes, including AAV-5. These findings have important implications for the future generation of Ad5/AAV hybrid vectors.

IMPORTANCE

Insertion of sequences from the 3′ part of the *rep* gene of adeno-associated virus (AAV) into the genome of its helper adenovirus strongly reduces adenoviral genome replication. We could show that this inhibition is mediated exclusively in *cis* without the involvement of *trans*-acting regulatory RNAs or polypeptides but nevertheless requires an active AAV-2 p40 promoter and p40-associated short transcripts. Our results suggest a novel inhibitory mechanism that has so far not been described for AAV and that involves stalled RNA polymerase II complexes and their interference with adenoviral DNA replication. Such a mechanism would have important implications both for the generation of adenoviral vectors expressing the AAV *rep* and *cap* genes and for the regulation of AAV gene expression in the absence and presence of helper virus.

Adeno-associated virus (AAV) is characterized by a bipartite replication cycle, with productive infection being critically dependent upon the presence of a helper virus, such as adenovirus (Ad) (1) or herpes simplex virus (HSV) (2). In the absence of helper virus, AAV can establish latent infection by integration into the host genome (3, 4). During the productive replication cycle, AAV is able to inhibit the propagation of the coinfecting helper virus, especially evident in the case of adenovirus (5, 6). This inhibition has been attributed to the expression of the large regulatory (Rep) proteins Rep78 and Rep68. These promote the replication of the AAV genome and induce the expression of the AAV serotype 2 (AAV-2) small Rep proteins Rep52 and Rep40 and the three structural proteins VP1, VP2, and VP3 (3). Rep78/Rep68 strongly suppress early adenoviral gene expression and genome replication (7) and to a lesser extent suppress HSV genome replication (8, 9). In addition to the large Rep proteins, AAV-mediated inhibition of adenoviral replication in *trans* critically depends on the presence of the AAV inverted terminal repeats (ITRs) (10). These represent the only AAV sequences required in *cis* for AAV genome replication and packaging.

The AAV ITRs also represent the key element of recombinant AAV (rAAV) vectors, which have been developed into very effective tools for the long-term expression of therapeutic genes in the clinical treatment of monogenic diseases, such as congenital blindness and hemophilia (11, 12). rAAVs are constructed as AAVs with gene deletions created by flanking the therapeutic transgene with the AAV ITRs. For production of the corresponding rAAV, the AAV *rep* and *cap* genes are provided in *trans* together with the required helper virus genes. This is mostly achieved by plasmid cotransfection (13, 14). Since transfection-based production systems are limited in scalability, a variety of systems that combine the AAV *rep* and *cap* genes and the required viral helper functions on recombinant helper viruses have been developed. Whereas such infection-based systems have been successfully implemented for HSV in mammalian cells (15–17) and baculovirus in insect cells (18–20), recombinant adenoviruses (rAds) carrying the AAV *rep* and *cap* genes either could not be

Received 10 December 2015 **Accepted** 26 January 2016

Accepted manuscript posted online 3 February 2016

Citation Weger S, Hammer E, Gonsior M, Stutika C, Heilbronn R. 2016. A regulatory element near the 3′ end of the adeno-associated virus *rep* gene inhibits adenovirus replication in *cis* by means of p40 promoter-associated short transcripts. J Virol 90:3981–3993. doi:10.1128/JVI.03120-15.

Editor: G. McFadden

Address correspondence to Stefan Weger, stefan.weger@charite.de.

propagated at all or were genetically unstable (21). Since inhibition of adenoviral replication by AAV sequences present in *cis* was also observed with rAd/AAV hybrid vectors exclusively harboring the *rep* gene (22–25), the effect was attributed to Rep protein expression. However, even tightly regulated expression of the Rep proteins could not relieve the *rep*-mediated inhibition of adenoviral replication (22, 24, 26). In line with these findings, Rep recoding experiments in the context of rAd vectors first suggested a *cis* inhibitory sequence in the 3′ part of the *rep* gene (26). In our recent study (10), we could confirm that this 3′ *rep* sequence, which we termed the Rep inhibition sequence for adenoviral replication (RIS-Ad), and not Rep protein expression is the major hurdle to the propagation of rAd/AAV hybrid vectors. RIS-Ad, which comprises the AAV-2 p40 promoter and the intron, seems to function at least partly in *cis* by a yet unknown mechanism (10).

To elucidate the molecular mechanisms involved in RIS-Ad function, the present study addressed the contribution of individual sequence elements contained within this small regulatory region. We demonstrate a transcription-associated inhibitory mechanism apparently mediated by promoter-proximal paused RNA polymerase II (Pol II) complexes. A role of small inhibitory RNA molecules or peptides could largely be ruled out.

MATERIALS AND METHODS

Cell culture and transfection. HEK-293 (human embryonal kidney) and HeLa (human cervix carcinoma) cells were propagated in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum and 100 μg/ml of penicillin and streptomycin at 37°C with 5% CO_2. Transfections were performed by the calcium phosphate precipitate technique as described previously (27) with 5×10^5 cells that had been seeded in 25-cm^2 flasks or 6-cm dishes on the day before transfection for HEK-293 cells or 4 to 5 h before transfection for HeLa cells.

Generation of rAds. Generation of rAds was performed with an AdEasy system (28, 29) largely as described previously. Briefly, the sequences of interest were cloned into the pShuttle vector, and PmeI-linearized pShuttle constructs were electroporated into *Escherichia coli* BJ5183-AD-1 cells for homologous recombination with the Ad serotype 5 (Ad5) genome with an E1/E3 deletion contained on the stably transformed pAdEasy (pAdE) plasmid. Recombinant pAdEasy plasmid DNA was amplified in *E. coli* XL1-Blue cells, linearized with PacI, and transfected into HEK-293 cells. The primary rAd preparation was harvested at 12 to 14 days posttransfection by four freeze-thaw cycles in liquid nitrogen and in a 37°C water bath in 2 ml of medium. An aliquot of the supernatant (0.5 ml) was used for the first round of amplification with 1×10^6 cells that had been seeded in 6-cm dishes on the day before infection. Further rounds of amplification and rAd purification were performed as described previously (29). Unless otherwise indicated, the generation of rAds was conducted at least three times for each AdEasy construct. The data are presented as the mean ± standard deviation.

Plasmids. The pShuttle-CMV-GFP transfer construct and the corresponding pAdEasy-CMV-GFP plasmid, used as a positive control for the production of rAd, have been described previously, as have the pShuttle-Rep1701/2186 and pShuttle-sRep1701/2186 constructs (10). pShuttle-Rep1701/2186 and pShuttle-sRep1701/2186 contain the AAV-2 wild-type sequence from nucleotides (nt) 1701 to 2186 or the corresponding sequence after recoding of nt 1782 to 1916 (indicated by sRep for scrambled Rep), respectively, as described by Sitaraman et al. (26). The constructs containing deletions of the 5′ and/or 3′ sequence of Rep1701/2186 were generated by PCR amplification of the corresponding AAV-2 sequences from the pTAV2-0 template (30) with oligonucleotides which additionally introduced an upstream XbaI site and a downstream SalI site for cloning into the pShuttle vector as described previously (10). All constructs are denoted according to the first and last AAV-2 nucleotides present (e.g., pShuttle-Rep1781/2060 for AAV-2 nucleotides 1781 to 2060). Domain-swapping constructs pShuttle-sRep-p40 and pShuttle-sRep-Int were generated by PCR amplification of partially recoded AAV nucleotides 1701 to 1882 or 1882 to 2186, respectively, from pShuttle-sRep1701/2186 with the concomitant reintroduction of the HindIII restriction site at nucleotide 1882 of the recoded AAV-2 sequence by the PCR primers, followed by cloning into XbaI/HindIII- or HindIII/SalI-restricted pShuttle-Rep1701/2186, respectively. For pShuttle-p40-Luci, the NotI/XhoI fragment from p1701/p40-Luci (see below) containing the p40 promoter, the luciferase (Luci) open reading frame (ORF), and the simian virus 40 poly(A) sequence was subcloned into the NotI/XhoI-digested pShuttle vector. pShuttle-CMV/Int was generated by replacement of the complete p40 promoter up to the HindIII site in pShuttle-Rep1701/2186 with a NotI/HindIII fragment from pHCMV-Luci (see below) containing the human cytomegalovirus (HCMV) immediate early enhancer and promoter (CMV-IE). For pShuttle-CMV/TATA-Int, two PCR fragments containing CMV-IE up to the TATA box and the 3′ Rep sequence from the p40 TATA box to nt 2186 were amplified from pShuttle-CMV/Int and pShuttle-Rep1701/2186, respectively. For fusion at the two TATA boxes, BbsI sites were introduced by the PCR primers, while terminal NotI and SalI sites were introduced for assembly of the two digested PCR fragments in NotI/SalI-digested pShuttle-CMV/Int after removal of the original CMV-Int insert. The internal deletions from AAV-2 nt 1840, 1850, 1860, or 1870 to the HindIII site at nt 1882 in the pShuttle-Rep1701/2186 plasmid were introduced through PCR amplification of AAV-2 nt 1701 to the respective 5′ boundary of the deletion from pShuttle-Rep1701/2186, with additional generation of a HindIII site and recloning into XbaI/HindIII-restricted pShuttle-Rep1701/2186. pShuttle-dl1882/1962, which carries AAV-2 from which nt 1882 to 1962 was deleted, was generated by Pst/HindIII restriction of pShuttle-Rep1701/2186, a fill-in reaction, and religation. All point mutants of the pShuttle-Rep1701/2186 plasmid were generated by site-directed mutagenesis, and their sequences were confirmed by sequencing. All the corresponding pAdEasy plasmids were generated as described above by homologous recombination of PmeI-linearized pShuttle DNA with pAdEasy in *E. coli* BJ5183-AD-1 cells.

The pHCMV-Luci vector containing CMV-IE in front of the luciferase reporter gene has been described previously (31). The AAV-2 p40 luciferase reporter constructs, in which the promoter part generally terminated at AAV-2 nt 1882, were generated by subcloning of the XbaI/HindIII fragments from the corresponding pShuttle-Rep constructs into the SpeI/HindIII-digested pBL vector (32) and were denoted according to either the starting AAV-2 nucleotide (e.g., p1757/p40-Luci for the construct with the p40 promoter from AAV-2 nt 1757 to 1882) or the region deleted within the p40 sequence from nt 1701 to 1882 (e.g., dl1860/1882 for deletion of nt 1860 to 1882). The pBluescript-RPA1843/1980 plasmid for *in vitro* transcription of the probe used in the RNase protection assay was generated by PCR amplification of the corresponding AAV-2 sequences from template pTAV2-0 by introduction of XbaI and SalI restriction sites for cloning.

Extraction of viral DNA for quantitation of newly replicated adenoviral genomes. Extraction of viral DNA by a modified Hirt procedure and quantification of newly replicated adenoviral genomes were performed essentially as described previously (10).

Quantification of rAd genomic particles. rAd genomic particles in a small aliquot of freeze-thaw supernatants were quantified by use of a Light-Cycler-based real-time PCR with a QuantiTect SYBR green PCR kit (Qiagen) as described recently (10). Primers E4-Q1 and E4-Q2 for amplification of a 202-bp region from the Ad5 E4 region have also been described previously (10). Primers GFP-Q1 (5′-GAG CTG AAG GGC ATT GAC TT-3′) and GFP-Q2 (5′-GCC GAT TGG AGT GTT CTG TT-3′) were used for amplification of a 198-bp region from the green fluorescent protein (GFP) ORF (nt 370 to 567), while primers RIS-Q1 (5′-AAG GGT GGA GCC AAG AAA AG-3′) and RIS-Q2 (5′-GAT GCA GTC ATC CAA ATC CA-3′) amplified a 393-bp sequence of the 3′ AAV-2 *rep* region.

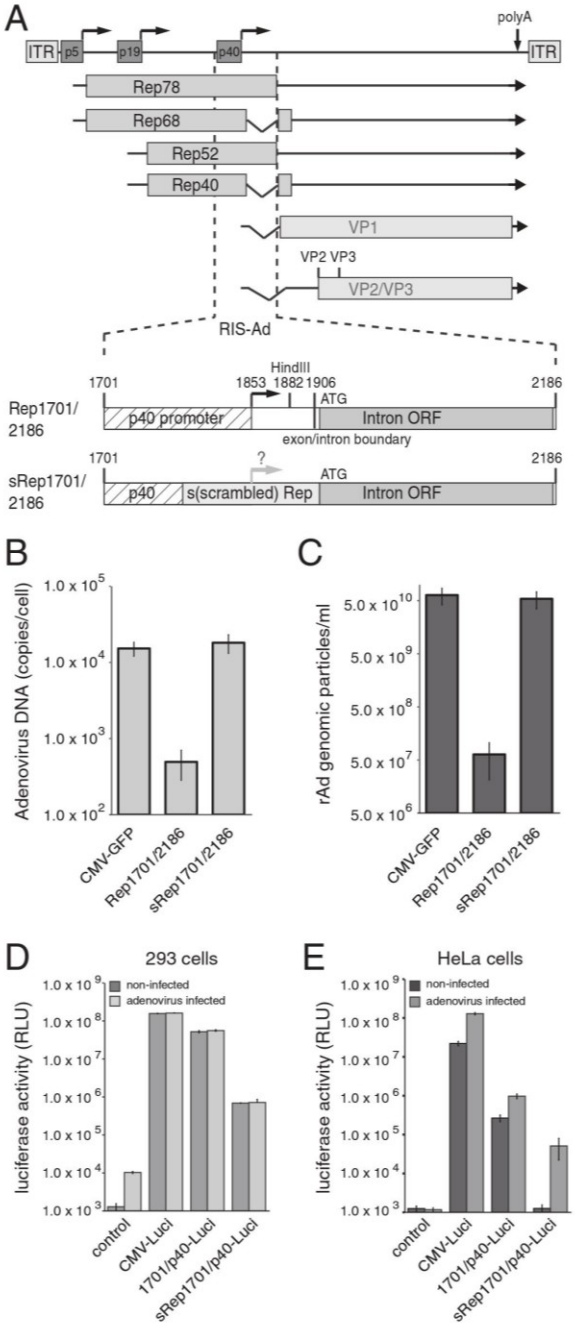


FIG 1 Ablation of the inhibitory effect of the 3′ *rep* region (RIS-Ad) on adenoviral replication through recoding is accompanied by strongly reduced p40 promoter activity. (A) Genome organization of AAV-2 and schematic representation of the 3′ *rep* sequence from AAV-2 (RIS-Ad) involved in inhibition of rAd replication. (Top) The viral genome with the inverted terminal repeats (ITRs), the four Rep proteins Rep78, Rep68, Rep52, and Rep40, and the capsid proteins VP1 to VP3, indicated by boxes with different shading. Right-angled arrows, the three promoters at map units 5, 19, and 40; vertical arrow, the common polyadenylation [poly(A)] site for all transcripts at map position 96. (Bottom) A closeup of AAV-2 nucleotides 1701 to 2186 (the RIS-Ad region) cloned into the pShuttle transfer vector of the AdEasy system. RIS-Ad comprises the p40 promoter, including the p40 transcription start site at nt 1853 and the HindIII site at nt 1882; the exon/intron boundary at nt 1906; and the internal open reading frame in the intron region (designated Intron ORF). s(scrambled) Rep, the exchange of AAV-2 nucleotides 1782 to 1916 for a recoded sequence as described previously (26). Characteristic nucleotide positions are given above the boxes. (B) Amount of newly amplified, DpnI-resistant adenoviral DNA extracted by a modified Hirt procedure, harvested at 12 to 13 days posttransfection of HEK-293 cells with the indicated PacI linearized, pAdEasy construct, presented as the number of copies per cell. (C) Amounts of recombinant adenoviral particles obtained in freeze-thaw cell supernatants after transfection of PacI-linearized pAdEasy plasmids into HEK-293 cells for 12 to 13 days and amplification of the primary supernatants in HEK-293 cells for 6 days determined by real-time PCR as the number of genomic particles per milliliter of the final cell supernatant. (D) HEK-293 (293) cells were transfected with the indicated luciferase reporter constructs, and relative luciferase activities in the cytoplasmic extracts were determined at 40 h posttransfection. Where indicated, cells were infected directly prior to transfection with Ad2 at an MOI of 20. (E) Transfection of HeLa cells with luciferase reporter constructs and determination of relative luciferase activities as described in the legend to panel D for HEK-293 cells. (D and E) The data from three independent experiments are presented as means ± standard deviations. RLU, relative light units.

Luciferase assays. For determination of firefly luciferase activities, cells in 6-cm dishes were washed twice in phosphate-buffered saline and scraped into 400 μl Triton lysis buffer (1% [vol/vol] Triton X-100, 25 mM glycylglycine [pH 7.8], 15 mM $MgSO_4$, 4 mM EGTA, 1 mM dithiothreitol, 10% [vol/vol] glycerin), and the cell debris was removed by centrifugation. Quantification was performed with 10 to 20 μl of supernatant and 10 μl of Bright-Glo luciferase substrate (Promega GmbH, Mannheim, Germany) using a Centro LB960 luminometer (Berthold Technologies, Bad Wildbad, Germany).

Western analysis. Immunoblot analysis for AAV-2 Cap protein expression was performed with monoclonal antibody B1 diluted 1:20 (Progen, Heidelberg, Germany) followed by incubation with horseradish peroxidase-conjugated secondary antibodies and enhanced chemiluminescence (ECL) detection as described previously (27).

Nucleic acid secondary structure prediction. For the prediction of the RNA secondary structure, the CentroidFold tool (33, 34), provided by the Computational Biology Research Center (CBRC) at the National Institute of Advanced Industrial Science and Technology (AIST) (http://www.ncrna.org/), in the basic mode and the software tools provided by the lab of David H. Mathews (Department of Biochemistry & Biophysics at the University of Rochester Medical Center; http://rna.urmc.rochester.edu/RNAstructureWeb/) were used. The prediction tool of the Mathews lab was used with the following default values: temperature, 310.15 K; maximum loop size, 30 bp; maximum percent energy difference (maximum free energy [MFE], MaxExpect algorithm [MEA]), 10%; maximum number of structures (MFE, MEA), 20; window size (MFE, MEA), 3; gamma (MEA), 1; number of iterations (pseudoknot prediction), 1; and minimum helix length (pseudoknot prediction), 3.

RPAs. RNase protection assay (RPAs) were performed with an RPA III kit (Ambion) essentially as described by the supplier. The AAV-2 antisense RNA probe (nt 1980 to 1843) was radiolabeled with [α-^{32}P]UTP by *in vitro* transcription with T7 polymerase (Riboprobe system; Promega) using an XbaI-linearized pBluescript-RPA1843/1980 plasmid as the template. The *in vitro*-transcribed RPA probes were gel purified on an 8% polyacrylamide (PAA)–8 M urea denaturing gel. Purified radiolabeled probe (3×10^4 cpm) was hybridized with 20 μg of total RNA followed by RNase A-RNase T1 treatment according to the manufacturer's protocol with the addition of ethanol at the final precipitation step. The protected RNA fragments were electrophoresed on an 8% PAA–8 M urea gel and subsequently fixed in a solution containing 20% ethanol and 5% acetic acid for 30 min before vacuum drying for 2 h at 80°C. The gels were exposed to a storage phosphor screen and scanned with a Fujifilm FLA-3000 imager.

RESULTS

Recoding of the 3′ part of *rep* ablates both inhibition of adenoviral replication and p40 promoter activity. Based on the findings of prior studies with recombinant adenovirus (rAd) vectors expressing a recoded AAV-2 Rep78 protein (26), we have recently demonstrated that an isolated sequence element from the 3′ part of the AAV-2 *rep* gene spanning nucleotides (nt) 1701 to 2186 (Rep1701/2186; Fig. 1A) completely blocks rAd replication when it is present in *cis* but not when it is provided in *trans* in the form of a nonamplifiable plasmid DNA (10). This *rep* sequence element, which we termed the Rep inhibition sequence for adenoviral replication (RIS-Ad), harbors the p40 promoter and most of the major intron (Fig. 1A). In agreement with these results, replacement of AAV-2 nucleotides 1782 to 1916 by the recoded *rep* sequence (termed sRep1701/2186, where sRep indicates scrambled Rep; Fig. 1A) as described by Sitaraman et al. (26) completely abolished the RIS-Ad-mediated inhibition of adenoviral genomic DNA replication after transfection of the corresponding pAdEasy constructs and by use of an AdEasy construct harboring a GFP-tagged cytomegalovirus (CMV-GFP) cassette as a positive control (Fig. 1B). Adenoviral genomic particle formation, as determined upon first-round amplification, was likewise suppressed to background levels by wild-type RIS-Ad but not by the recoded sequence (Fig. 1C). Since the changes introduced by the sRep sequence mainly affected the p40 promoter part of RIS-Ad (Fig. 1A), we assayed the transcriptional activities of both the wild-type and the sRep p40 sequences from AAV-2 nt 1701 to 1882 by luciferase reporter assays. RIS-Ad includes the transcription start site (TSS) located at nt 1853, which is mutated in the recoded sequence (Fig. 1A). The wild-type p40 promoter was highly active in HEK-293 cells, with its activity being only about 3-fold lower than that of the HCMV promoter and enhancer (CMV-IE), used as a control (Fig. 1D; compare the activity of 1701/p40-Luci with that of CMV-Luci). The promoter activity of the sRep construct was strongly reduced by 2 orders of magnitude (80-fold) compared to that of the wild type (Fig. 1D, sRep1701/p40-Luci). Interestingly, neither construct was further activated upon infection with Ad2 at a multiplicity of infection (MOI) of 20 (Fig. 1D). Apparently, Ad E1A and E1B constitutively expressed in HEK-293 cells are sufficient for adenovirus-mediated p40 promoter activation. In line with this notion, analogous experiments in HeLa cells showed a strongly reduced basal activity of both promoters and a clear enhancement of activity in the presence of adenovirus (Fig. 1E).

The properties of RIS-Ad closely correlate with p40 promoter activity. The strongly reduced transcriptional activity of the sRep-modified p40 promoter region pointed to a possible contribution of p40 promoter strength to RIS-Ad activity. We therefore investigated a series of constructs with 5′ deletions which successively removed candidate transcription factor-binding sites (35) located upstream of the p40 transcription start site in the wild-type sequence (Fig. 2A). Deletion of p40 promoter sequences between AAV-2 nt 1701 and 1781, including consensus sites for binding of transcription factors EF1A, MLTF, and ATF and one of two so-called GGT motifs, did not lead to an increase in the formation of recombinant adenoviral particles (Fig. 2A and B, mutants Rep1757/2186 and Rep 1781/2186). Further removal of the second GGT motif, however, restored adenoviral particle formation to a level only 2- to 3-fold lower than that for the CMV-GFP control after first-round amplification (Fig. 2A and B, Rep1797/2186). When an additional 24 nucleotides harboring the potential Sp1 and AP1 binding sites were deleted, the inhibitory potential of RIS-Ad was completely lost (Fig. 2A and B, Rep1821/2186). Reporter gene assays performed with the corresponding luciferase constructs in noninfected HEK-293 cells (Fig. 2C) demonstrated that the formation of rAd particles inversely correlated with p40 transcriptional activity. Whereas deletion of the upstream elements up to the ATF site had no major impact on promoter activity (Fig. 2C, 1757/p40-Luci and 1781/p40-Luci), a clear decrease in p40 activity was observed in constructs with deletions beyond the downstream GGT motif (Fig. 2C, 1797/p40-Luci), and activity was largely lost after deletion of the potential Sp1 and AP1 tran-

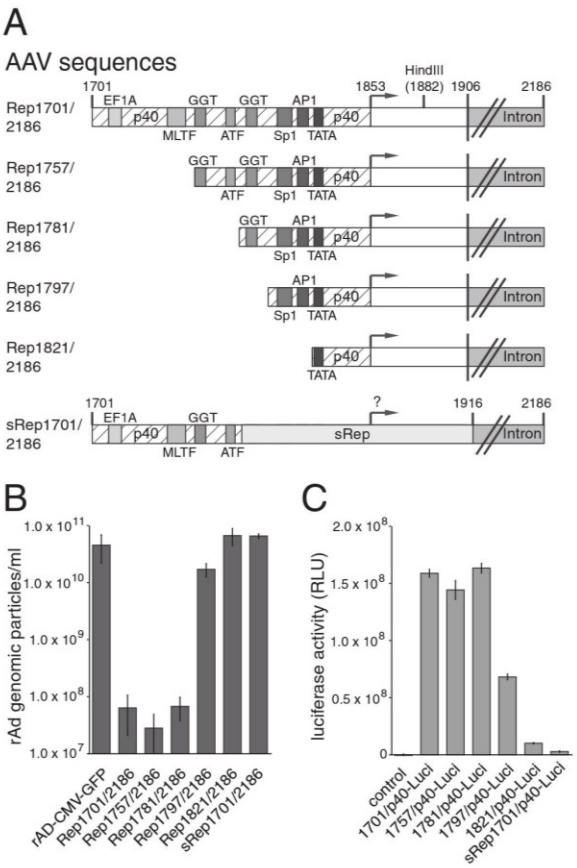


FIG 2 RIS-Ad function closely correlates with the p40 promoter activity defining the 5′ boundary of the inhibitory region. (A) Schematic representation of the 5′ deletion variants of the Rep1701/2186 sequence (RIS-Ad) assayed for both inhibition of adenoviral replication and p40 promoter activity. Boxes with different shading represent the binding elements for characteristic transcription factors, as indicated. The p40 TSS at nt 1853 is represented by a right-angled vector. Additionally, the HindIII site at nt 1882 constituting the 3′ boundary of the p40 promoter in the luciferase reporter constructs, the exon/intron boundary, and intron region are shown, and the characteristic nucleotide positions are indicated. (B) Amounts of recombinant adenoviral particles determined in freeze-thaw cell supernatants after transfection and first-round amplification of the indicated AdEasy-Rep1701/2186 deletion variants as described in the legend to Fig. 1C. (C) Relative luciferase activities of the indicated p40 reporter constructs determined as described in the legend to Fig. 1D.

scription factor-binding sites (1821/p40-Luci). Thus, both sequence deletions and recoding, as in the sRep construct (Fig. 1; also included in Fig. 2 for better comparison), led to a reduction in p40 promoter activity which closely correlated with the loss of RIS-Ad activity. These findings strongly implicate a functional p40 promoter as an essential component of the inhibitory mechanism.

Characterization of AAV-2 intron sequences required for the inhibitory effects of RIS-Ad. In a previous study (10), we already demonstrated that at least part of the AAV-2 intron sequence is required for the inhibitory effect of RIS-Ad. The 3′ boundary of RIS-Ad was delineated more closely in the present study through the use of a series of deletion mutants, in all of which the deletion started at nt 1701 and terminated at different sites within the intron (Fig. 3A). Sequences beyond nt 2060 were largely dispensable for inhibition of adenoviral replication (Fig. 3A and B, Rep1701/2122 and Rep1701/2060). A small increase in the level of formation of genomic particles of about 5- to 10-fold above the background level for Rep1701/2186 after first-round amplification was observed for Rep1701/2020 (Fig. 3A and B). Rep1701/2020 also gave rise to infectious particles in plaque assays, and these could be further propagated in successive amplification rounds (data not shown). Deletion of further nucleotides from the 3′ end led to a continuous loss of inhibitory activity (Fig. 3A and B, Rep1701/1980 and Rep1701/1938), with the mutant terminating at nt 1938 showing genomic titers similar to those of the CMV-GFP control vector. Thus, only the first half of the intron sequence is required for RIS-Ad-mediated inhibition. In summary, the minimal *cis* inhibitory sequence could be assigned to AAV-2 nt 1781 to 2060 (Fig. 3C and D).

By introduction of different subfragments of the recoded sRep sequence ranging from nt 1782 to 1882 or from nt 1882 to 1916 (Fig. 3E), we could demonstrate that the abrogating effect of the recoded sequence is limited to the p40 promoter region and does not involve the region around the splice donor site (Fig. 3E and F, sRep-p40 and sRep-Int). The successive loss of inhibitory activity observed for the series of 3′ deletion mutants (Fig. 3B) suggested that the sheer length of potentially transcribed sequences downstream from the p40 promoter played a role. To address this question, we replaced the sequence downstream from nt 1882 with the luciferase ORF (total fragment length, 2.7 kb, including the 3′ nontranslated region) (Fig. 3E, p40-Luc). The respective AdEasy constructs replicated and were packaged at an efficiency similar to that for the control pAdEasy-CMV-GFP construct (Fig. 3F), and high-titer rAd preparations expressing large amounts of luciferase could be produced by further amplification (not shown). Thus, specific sequences located within the AAV-2 intron are necessary for inhibition.

No evidence for involvement of RIS-Ad intron-encoded polypeptides or miRNAs in RIS-Ad activity. Based on the findings that both a functional p40 promoter and sequences extending into the AAV-2 intron were required for inhibition of adenoviral replication, we hypothesized that p40-driven expression of an intron-encoded effector molecule, either a small polypeptide or a regulatory RNA, could be involved. Whereas RIS-Ad has been shown not to inhibit adenoviral replication in *trans* (10), initial replication of the corresponding AdEasy plasmids could possibly generate large amounts of the hypothetical effector molecule. The AAV-2 intron not only contains the C-terminal part of the Rep78/52 reading frame but also harbors an additional ORF of 82 amino acids (the corresponding hypothetical polypeptide is denoted "intron protein" in the following) with a canonical ATG translation initiation codon. This ATG codon (AAV-2 nt 1919 to 1921) is located just a few bases downstream from the splice donor site (Fig. 4A). A second ATG initiation codon located in the region from nt 1938 to 1940 could theoretically initiate a polypeptide

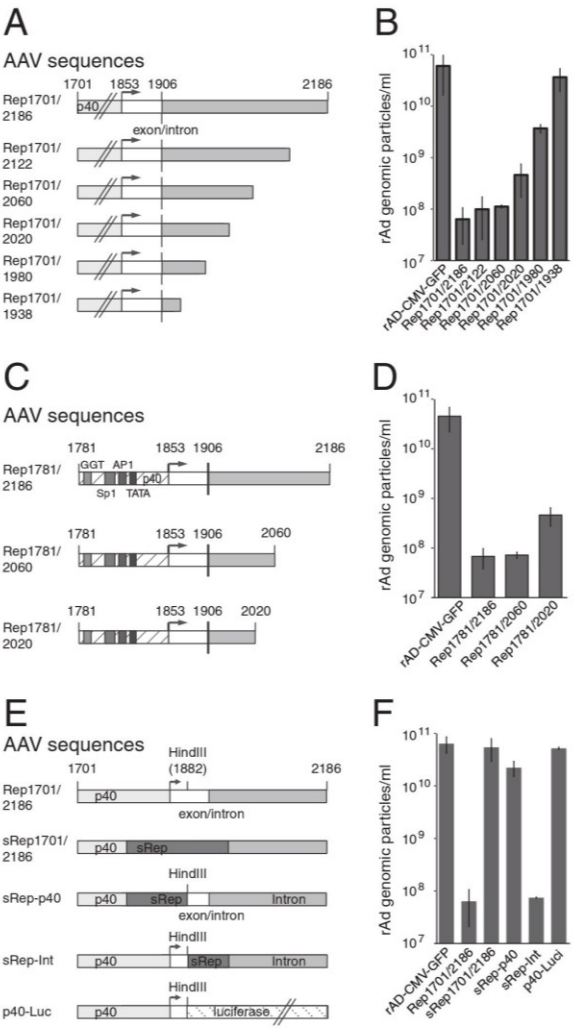


FIG 3 Only the 5′ part of the intron is required for RIS-Ad-mediated inhibition of adenoviral replication. (A, C, and E) Schematic representation of the variants with 3′ deletions of the Rep1701/2186 sequence (RIS-Ad) (A), variants with both 5′ and 3′ deletions (C), and domain swap constructs (E), all of which were assayed for inhibition of adenoviral replication. (A and C) Boxes with different shading and open boxes represent the p40 promoter with defined transcription factor-binding sites, the 3′ nontranslated region of the cap transcripts (nt 1853 to 1906), and the intron sequence. (E) Different shaded and open boxes represent the wild-type and recoded (sRep) p40 promoter, the 3′ nontranslated cap region, and the AAV-2 intron sequences. The luciferase ORF is also indicated. (B, D, and F) Amounts of recombinant adenoviral particles determined in freeze-thaw cell supernatants after transfection and first-round amplification of the indicated AdEasy constructs as described in the legend to Fig. 1C.

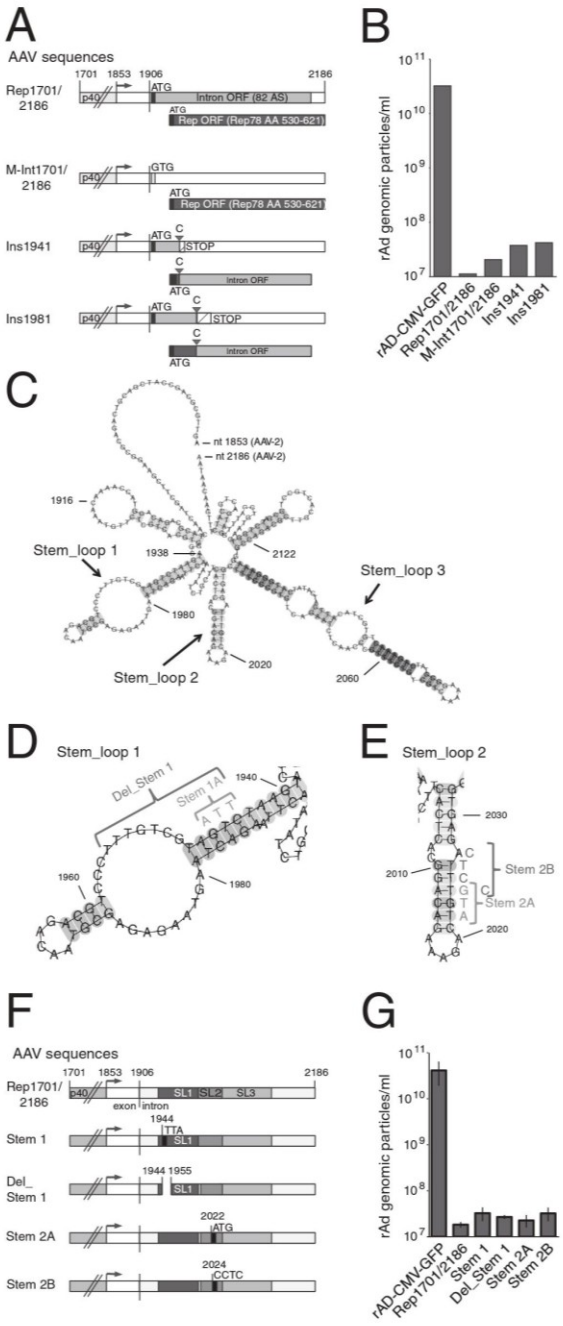


FIG 4 RIS-Ad point mutations targeting potential small polypeptides and RNA stem-loop structures do not abrogate inhibition. (A) Schematic representation of potential polypeptides encoded by the intron and point mutations disrupting these ORFs. Boxes with different shading and open boxes represent the p40 promoter, the 3′ nontranslated region, and the two larger ORFs found in the intron. Putative translation initiation codons (ATG) are shown. Filled triangles, insertions of cytidine residues; STOP, translation termination codons; AA, amino acid. (B) Amounts of recombinant adenoviral particles determined in freeze-thaw cell supernatants after primary transfection and first-round amplification of the indicated pAdEasy constructs as described in the legend to Fig. 1C. (C) RNA secondary structure of a transcript initiating at the described p40 TSS and terminating at AAV-2 nt 2186, as predicted by the CentroidFold software tool as described in the Materials and Methods section. The numbering of the nucleotides corresponds to the complete AAV-2 genome. (D and E) Close-ups of predicted stem-loop structures 1 and 2 from panel C with the indicated nucleotide changes introduced to disrupt these potential secondary structures. (F) Schematic representation of the point mutations in predicted stem-loop structures 1 and 2 (SL1 and SL2, respectively) of Rep1701/2186 with the indicated nucleotide changes. (G) Amounts of recombinant adenoviral particles determined in freeze-thaw cell supernatants after primary transfection and first-round amplification of the pAdEasy constructs shown in panel F as described in the legend to Fig. 1C.

with the 82 C-terminal Rep78/52 amino acids (Fig. 4A). No further potential polypeptides longer than 20 amino acids could be initiated by canonical ATG start codons within the intron. The possibility of the involvement of the hypothetical intron protein in RIS-Ad function could be excluded by point mutation of its potential ATG start codon to GTG (Fig. 4A and B, M-Int1701/2186). This finding could be further confirmed by the placement of 1-nt insertions at nucleotide positions 1941 and 1981 (Ins1941 and Ins1981, respectively; Fig. 4A and B), which disrupt the corresponding ORF. These insertions also affect the ORF encoding the C-terminal Rep78/Rep52 amino acids, which shift into the intron protein ORF. Thus, the marginal increases in rAd particle formation observed with these frameshift mutants largely exclude the possibility of a role of intron-encoded polypeptides in the inhibition. We next focused on regulatory RNAs as potential effectors. Since cellular processing of the precursors of regulatory RNAs, like microRNAs (miRNAs), requires characteristic stem-loop structures, the potential secondary structure of an unspliced p40-driven transcript initiating at the described AAV-2 p40 transcription start site (TSS; nt 1853) and terminating at nt 2186 was investigated *in silico* with several complementary software tools, as described in Materials and Methods. All tools predicted two prominent stem-loop structures (Fig. 4C; arbitrarily denoted Stem_loop 1 [nt 1939 to 1991] and Stem_loop 2 [nt 2003 to 2032]) in the AAV-2 intron region, and these were found to be required for RIS-Ad function (compare Fig. 3). A third stem-loop structure (Stem_loop 3; Fig. 4C) was only partially located in this region and was therefore not investigated further. Of note, AAV-2 nt 1939 to 1991 were also scored with the miRAbela miRNA prediction tool, run for the complete AAV-2 genome (threshold level, −2; 3 additional sequences were scored). Point mutations and deletions, which should disrupt the secondary RNA structure (structure prediction data are not shown), were introduced into both Stem_loop 1 (Fig. 4D) and Stem_loop 2 (Fig. 4E) in the context of AAV-2 nt 1701 to 2186. However, none of the point mutations (Fig. 4D to F) abrogated the inhibitory effect of RIS-Ad on the formation of adenoviral particles (Fig. 4G), strongly arguing against the involvement of AAV-encoded miRNAs in the inhibition.

The p40 transcription start site is required for RIS-Ad function. As RIS-Ad activity strongly correlated with p40 promoter activity (compare Fig. 2), we next examined whether CMV-IE, which has activity even stronger than that of p40 in HEK-293 cells (compare Fig. 1D), could functionally replace p40 in the RIS-Ad sequence. Rather surprisingly, however, replacement of AAV-2 p40 sequences upstream of nt 1882 by the CMV-IE sequence, including the TATA box and transcription start site (Fig. 5A; the

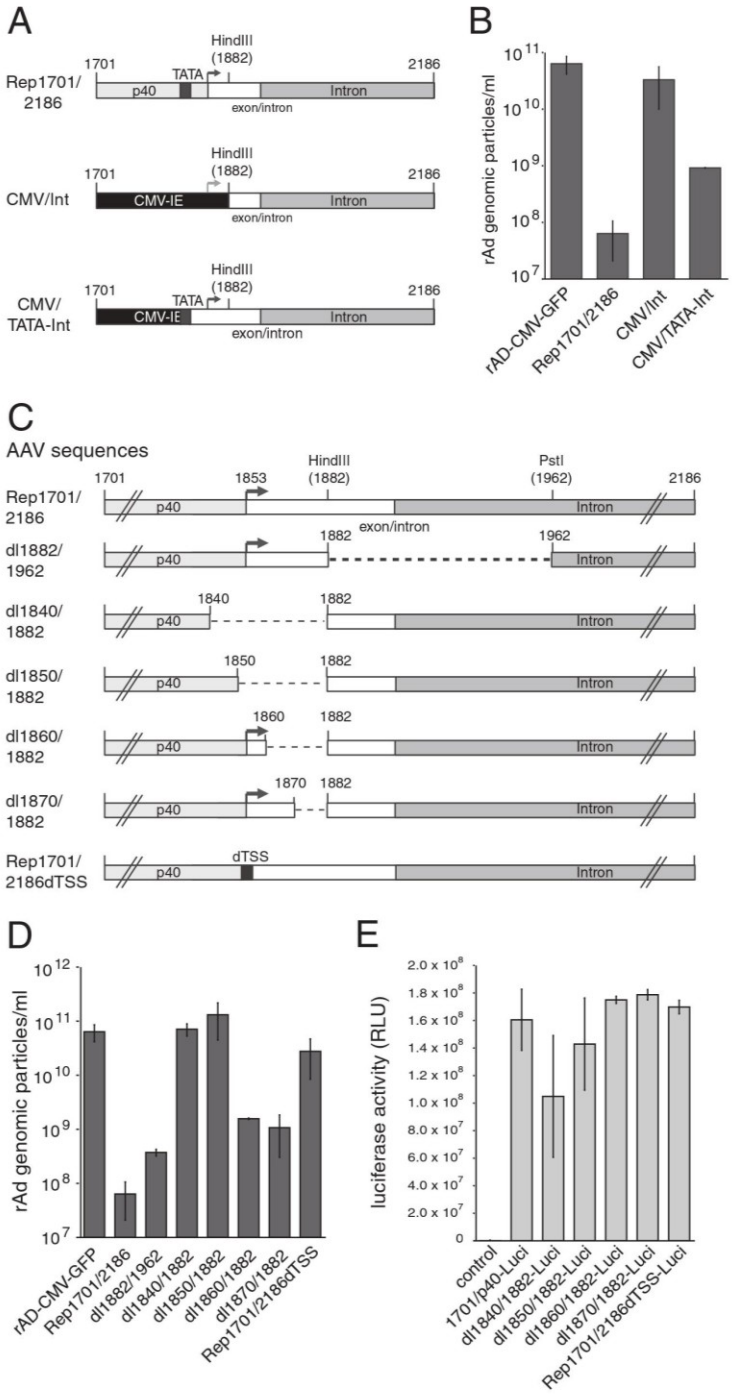


FIG 5 RIS-Ad-mediated inhibition requires the p40 TSS. (A and C) Schematic representation of the chimeric constructs containing various parts of CMV-IE and the RIS-Ad (A) and the internal Rep1701/2186 deletion constructs (C). (B and D) Amounts of recombinant adenoviral particles determined in freeze-thaw cell supernatants after transfection and first-round amplification of the indicated AdEasy constructs as described in the legend to Fig. 1C. (E) Relative luciferase activities of the indicated p40 reporter constructs harboring internal deletions in the AAV sequence as described in the legend to Fig. 1D.

identical sequences were assayed in the luciferase assay, the results of which are presented in Fig. 1D), led to a complete loss of inhibition (Fig. 5B, CMV/Int). This result pointed to a requirement for the authentic AAV p40 TSS in RIS-Ad function. A second construct with CMV-IE fused to p40 directly at the TATA boxes (located at nt 1821 for AAV-2; compare Fig. 2A) (Fig. 5A, CMV/TATA-Int) led to a partial restoration of the inhibition (Fig. 5B). The requirement for the AAV-2 region surrounding the p40 TSS was further confirmed with a series of internal deletions. In agreement with the previous results (Fig. 3E and F, sRep-Int), nt 1882 to 1962 downstream from the TSS contributed only weakly to the inhibition (Fig. 5C and D, dl1882/1962). In contrast, deletions which affected the TSS completely abrogated inhibition (Fig. 5C and D, dl1840/1882 and dl1850/1882) with only a minor loss of promoter activity, as determined by luciferase reporter gene assays (Fig. 5E; note the linear scale). The requirement for the TSS was then directly confirmed by a short 6-bp deletion (Fig. 5C, Rep1701/2186dTSS), which largely abolished inhibition, while overall transcriptional activity was preserved (Fig. 5D and E).

To assay the possible transcripts initiated from the p40 TSS by the different plasmids harboring the Ad5/AAV-2 hybrid investigated in this study, these plasmids were transfected into HEK-293 cells and RNase protection assays (RPAs) were performed on total RNA isolated at 2 days posttransfection (Fig. 6). Transcripts initiated at the p40 TSS (AAV-2 nt 1853) and extending at least to the right boundary of the RPA probe at AAV-2 nt 1980 (Fig. 6, top) would have been expected to lead to a protected RNA fragment of about 128 nt (Fig. 6, arrow). However, all RNAs which led to detectable signals in the RPAs showed clusters of protected RNA fragments over a size range of 30 to 70 nt (Fig. 6). This pattern is highly reminiscent of RNA polymerase II pausing, which has been shown to occur in the promoter-proximal region of a variety of promoters (36). Importantly, the intensity of these short transcripts initiated from the different constructs correlated closely with the inhibition of adenoviral replication. A continuous decrease in intensity was observed after deletion of the GGT, Sp1, and AP1 transcription factor-binding sites (Fig. 6, bottom, lanes 3 to 6, constructs Rep1701/2186 to 1821/2186; compare Fig. 2A), and no transcripts could be detected for the recoded sRep1701/2186 construct, the dl1840/1882 construct with the TSS deletion, or the Rep1701/1938 construct terminating at AAV-2 nt 1938 (Fig. 6, lanes 8, 9, and 12, respectively). The last three constructs were all negative for inhibition (compare Fig. 1B and C, 3A and B, and 5C and D). These results strongly suggest a transcription-dependent mechanism for RIS-Ad-mediated inhibition of adenoviral replication involving initiation at the well-characterized p40 TSS and subsequent RNA polymerase II pausing.

The inhibitory function of the AAV-2 RIS-Ad region can be overcome by an excess of replication-competent adenoviral genomes in *trans*. We have previously shown that RIS-Ad is not able to inhibit adenoviral replication when provided from a second construct in *trans* (10). In line with these former results, we were not able to identify a potential *trans*-acting effector molecule involved in the inhibition detected in the present study (see above). To further exclude the possibility that such a hypothetical *trans*-acting effector molecule might be produced in large amounts by initial replication of the recombinant adenoviral construct, we modified the *trans* inhibition assay from the assay used in the earlier study. Instead of the nonamplifiable pShuttle-Rep1701/2186 plasmid harboring RIS-Ad, the corresponding pAdEasy

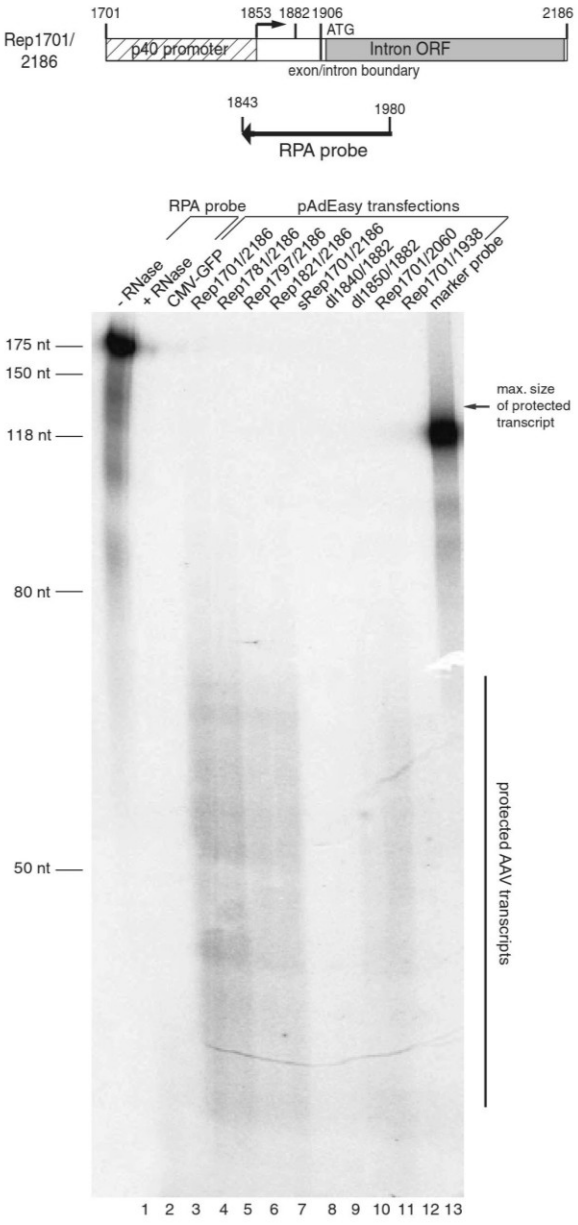


FIG 6 Short transcripts are expressed from the RIS-Ad sequence. (Top) RNase protection assay with an *in vitro*-transcribed hybridization probe covering AAV-2 nucleotides 1843 to 1980. (Bottom) This AAV antisense probe was used to detect p40-initiated transcripts in total RNA isolated 2 days posttransfection of HEK-293 cells with the indicated pAdEasy constructs. The results for the nondigested RPA probe (−RNase) and the probe digested in the presence of *Torula utilis* yeast tRNA (+RNase) are shown as controls in lanes 1 and 2, respectively. The probe contains additional non-AAV-2 sequences from the pBluescript vector used for T7 polymerase-dependent *in vitro* transcription. The maximum length of the protected AAV sequence is 128 nt, and this size is indicated by the arrow on the right. As an additional size marker, in addition to an unlabeled RNA length standard, a labeled probe with a size of 118 nt was loaded in the last lane. For better orientation, the lanes are additionally numbered at the bottom of the gel.

(pAdE)-Rep1701/2186 construct was cotransfected with the indicator construct pAdE-CMV-GFP. However, even at a 2:1 ratio of pAdE-Rep1701/2186 to pAdE-CMV-GFP, we did not observe a decrease in the number of rAd-CMV-GFP genomic particles produced after first-round reinfection compared to the number observed in the absence of RIS-Ad in *trans* (Fig. 7A; compare the 3 columns on the right). We then asked whether the presence of the replication-competent pAdE-CMV-GFP genome might, conversely, actually rescue the *cis* inhibitory effect of RIS-Ad. After cotransfection of the AdE-p40-Cap construct with increasing amounts of pAdE-CMV-GFP as a helper plasmid, the rAd particles harboring either the GFP or the AAV genome sequence were scored separately. The pAdE-p40-Cap construct (Fig. 7D) contains the complete RIS-Ad and had been shown before to be defective for the formation of adenoviral particles (10). This construct rather than the pAdE-Rep1701/2186 construct was chosen for use in the rescue assay, since it allowed an easy determination of the integrity of the AAV sequences by using the capsid protein expression as a readout (see below). In line with the results presented above, only background amounts of rAd-p40-Cap particles were obtained in the absence of cotransfected pAdE-CMV-GFP (Fig. 7B). The apparent presence of rAd-CMV-GFP in these samples is due to a relatively high background level of this specific real-time PCR in combination with a high level of predilution of the probes. However, with increasing amounts of the pAdE-CMV-GFP plasmid used for rescue, the number of rAd particles containing the AAV-2 p40-Cap sequence increased strongly and reached levels of about 1×10^{10} genomic particles/ml (Fig. 7B). The titers of these rAd-p40-Cap viral particles that copropagated with rAd-CMV-GFP remained at high levels over three additional amplification rounds (Fig. 7C), as did the rAd-CMV-GFP titers. Thus, replication-competent rAd-CMV-GFP did not outgrow rAd-p40-Cap but could fully rescue this normally replication-deficient recombinant virus. The mixed rAd-CMV-GFP/rAd-p40-Cap preparations were also capable of producing large amounts of the AAV-2 Cap proteins (shown in Fig. 7D for the fourth amplification round) with a VP3/VP2/VP1 stoichiometry very similar to that in a wild-type AAV-2/Ad2 coinfection (data not shown), confirming the integrity of the p40 promoter and the Cap open reading frame. The successful expression of the Cap proteins was also not due to recombination events between the CMV promoter of rAd-CMV-GFP and the 5′ end of the *cap* gene, since the region PCR amplified for determination of the rAd genomic titers (see the Materials and Methods section) contained both the p40 promoter and large parts of the AAV-2 intron, and no clear sequence variations were observed within this region (date not shown).

The inhibitory function of the AAV-2 RIS-Ad region is conserved among other AAV serotypes. A sequence alignment of the AAV-2 p40 region up to the splice donor (SD) site at nt 1906 with the corresponding sequences of other serotypes showed that, in particular, the potential transcription factor-binding sites found to be important for RIS-Ad activity (compare Fig. 2) and the region surrounding the AAV-2 TSS were highly conserved not only in AAV-4 and AAV-6 but also in AAV of a more distantly related serotype, AAV-5 (Fig. 8A). Within the intron region, the sequences of AAV-2, AAV-4, and AAV-6 were highly conserved (Fig. 8B), while the sequence of AAV-5 showed a variety of deletions, insertions, and substitutions compared to the sequences of the three other aligned serotypes. To analyze the inhibitory potential of the 3′ part of the *rep* gene of the other AAV serotypes, the corresponding regions starting with the second GGT motif and terminating close to the end of the intron were introduced into the

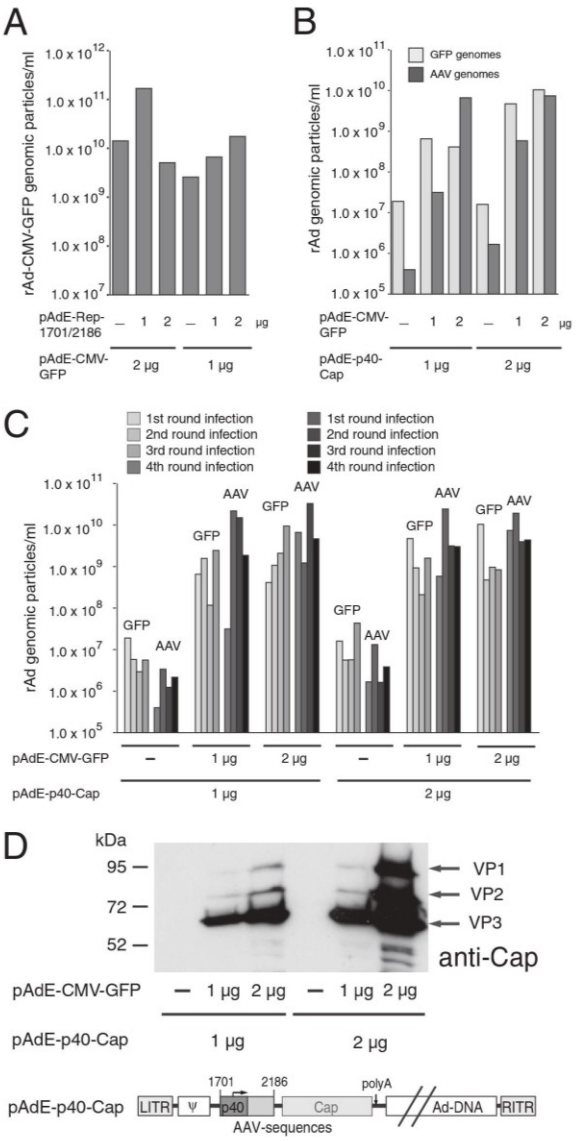


FIG 7 Rescue of replication-deficient rAd/AAV hybrid vectors by a replication-competent rAd in *trans*. (A) *trans* inhibition assay. PacI-linearized pAdEasy plasmid constructs pAdE-CMV-GFP as indicator DNA and pAdE-Rep1701/2186 as effector DNA were cotransfected into HEK-293 cells at different ratios, as indicated, and rAd particles containing the GFP genome were scored after first-round amplification of the primary freeze-thaw supernatants (obtained at 12 days post-transfection) of HEK-293 cells for 6 days. (B) Rescue assay. PacI-linearized pAdEasy plasmid constructs pAdE-CMV-GFP and pAdE-p40-Cap, which are shown schematically in panel D, were cotransfected into HEK-293 cells at different ratios, as indicated, and rAd particles containing either the GFP or the 3′ Rep sequences (AAV-2 nt 1782 to 2174, denoted AAV) were separately scored after transfection and amplification of the primary freeze-thaw supernatants as described in the legend to panel A. The apparent presence of rAd-CMV-GFP particles in the transfections devoid of the pAde-CMV-GFP plasmid is due to a relatively high background level of this specific real-time PCR product in combination with a high level of predilution of the probes. (C) For the rescue assay described in the legend to panel B, three further amplification rounds for 3, 5, and 4 days, respectively, were conducted, and the amounts of rAd-CMV-GFP and rAd-p40-Cap genomic particles were determined as described above. (D) Western analysis for AAV-2 Cap proteins in total cell extracts obtained after the fourth amplification round of the rescue assay described in the legends to panels B and C. LITR and RITR, left and right ITRs, respectively. Ψ, adenoviral packaging signal.

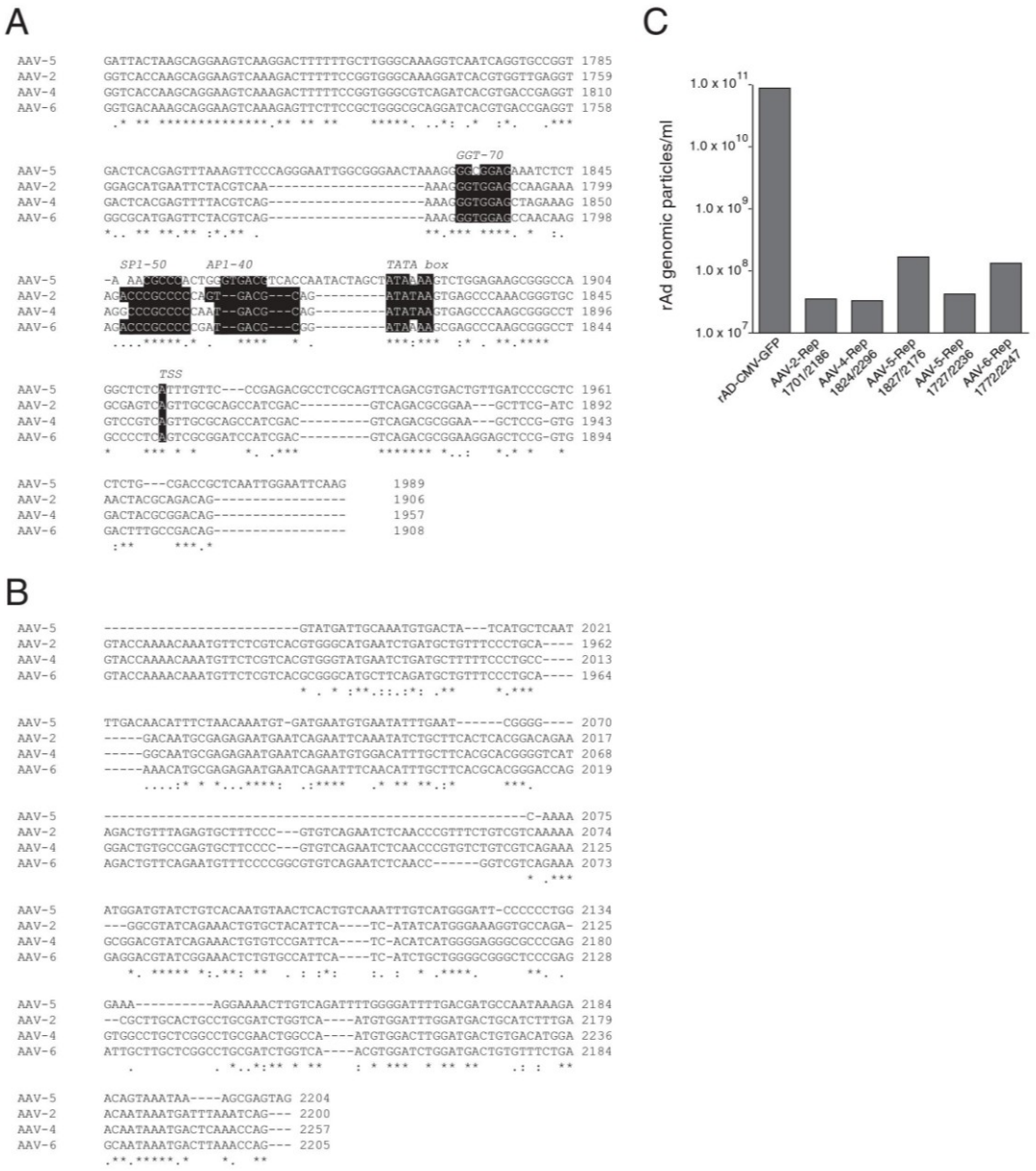


FIG 8 The inhibitory effect of the 3′ region of the AAV *rep* gene is conserved in other AAV serotypes. (A) Sequence alignment of the AAV-2 sequence from nt 1701 to the splice donor (SD) site at nt 1906 with the corresponding regions from AAV serotypes 4, 5, and 6. Transcription factor-binding sites are indicated by white type on a black background. (B) Sequence alignment of the intron region of AAV-2 (nt 1907 to 2200, minor splice acceptor site) to the intron regions of AAV serotypes 4, 5, and 6. (C) Amounts of recombinant adenoviral particles determined in freeze-thaw cell supernatants after transfection and first-round amplification of the indicated AdEasy constructs harboring 3′ *rep* sequences from different AAV serotypes.

AdEasy system and assayed for the formation of adenoviral particles in HEK-293 cells. As expected, the closely related AAV-4 and AAV-6 sequences inhibited adenoviral replication in a fashion similar to that of the AAV-2 RIS-Ad (Fig. 8C). Despite the variations in the intron region, the C-terminal AAV-5 Rep sequence also strongly inhibited adenoviral replication (Fig. 8C, AAV-5 Rep1827/2176). The minor amounts of particles observed were further reduced when the complete AAV-5 p41 promoter was included in the pAdEasy-AAV-5-Rep construct (Fig. 8C, AAV-5-Rep1727/2236). Thus, the inhibitory function of RIS-Ad probably represents a general feature of AAV and is not unique to the AAV-2 serotype.

DISCUSSION

AAV-mediated inhibition of the replication of its helper adenovirus was already described shortly after its initial identification as a contaminant of adenoviral preparations (5, 6). In subsequent studies, this inhibition could conclusively be assigned to the expression of the Rep proteins (7, 37, 38). Therefore, the limited replication capability of adenoviral vectors harboring the AAV *rep* gene (22, 23, 39) was also attributed to Rep protein expression. Recently, however, it became evident that a constrained sequence element in the 3′ part of the *rep* gene contributes to the inhibition of rAd/AAV hybrid vector propagation independently of Rep protein expression (10, 26).

In the present study, we have further pinpointed the functional elements and possible mechanisms involved in the action of this 3′ *rep* inhibitory sequence, which we termed RIS-Ad. Importantly, we could further corroborate the validity of the hypothesis raised by the earlier studies (10, 26) that RIS-Ad exclusively functions as a *cis* regulatory element. This view is supported by two major findings. (i) In our modified *trans* inhibition assay, we provided RIS-Ad in the same adenviral vector backbone as the cotransfected recombinant adenoviral reporter vector, which harbored the CMV-GFP cassette. Even under these conditions, where the initial replication of rAd containing the RIS-Ad cassette may produce large amounts of template for the expression of a potentially *trans*-acting effector molecule, replication of the recombinant adenoviral CMV-GFP reporter vector was not impaired at all but, rather, was stimulated by increasing amounts of cotransfected RIS-Ad. (ii) Furthermore, despite extensive mutational analysis affecting the potential open reading frames or RNA secondary structures characteristic of candidate miRNA precursors within the AAV-2 intron, we were not able to obtain any evidence for the involvement of small regulatory RNAs or of polypeptides in the inhibition.

Although *trans*-acting molecules are obviously not involved in RIS-Ad function, inhibition still requires an active p40 promoter. With the deletion of the GGT motif proximal to the TATA box, p40 promoter activity dropped markedly and the RIS-Ad-mediated inhibition of adenoviral replication was almost completely abolished. GGT motifs have been identified in the AAV-2 p19 and p40 promoters and have been shown to be important for Rep-mediated activation of transcription from these promoters in HeLa cells in the presence of adenovirus (35). However, it has not been determined whether the large Rep proteins Rep78 and Rep68 or adenoviral factors directly bind to the GGT motif. Additional deletion of the Sp1 and AP1 transcription factor-binding sites proximal to the TATA box further reduced p40 activity and completely restored replication of the corresponding Ad5/AAV-2 hybrid vector. One possible explanation for the inverse correlation between p40 activity and the inhibition of adenoviral replication may be competition between the corresponding transcription factor sites in the RIS-Ad and those in the adenoviral genome for the limited amount of transcription factors (40). However, such a mechanism does not explain either the exclusive function of RIS-Ad as a *cis*-acting sequence or the requirement for additional downstream sequences, unless these may also compete for transcription factors. Furthermore, replacement of the p40 sequences up to and beyond the transcription start site by the strong CMV-IE completely abolished the inhibition of adenoviral replication, although CMV-IE also contains a number of Sp1 and AP1 transcription factor-binding sites and was demonstrated to have a higher transcriptional activity than the p40 promoter in reporter gene assays. By the use of additional chimeras of CMV enhancer/promoter sequences with AAV sequences and RIS-Ad deletion mutants specifically targeting the p40 transcription start site, we could then demonstrate that inhibition of adenoviral replication by the RIS-Ad sequence involves a transcription-dependent mechanism obviously closely related to the exact nature of the 5′ region of the nascent transcripts.

At first glance it seems hard to envision how high levels of transcription from the p40 promoter may lead to inhibition of genome replication without producing an effector molecule also capable of inhibiting adenoviral replication in *trans*. Some clues about possible mechanisms were obtained when we measured p40-driven transcription from the 3′ Rep sequences capable of inhibiting adenoviral replication in an RNase protection assay. We could not detect major amounts of the 128-nt protected fragment expected after transcription initiation at the p40 TSS and further elongation of the nascent transcripts. Rather, we observed a heterogeneous population of transcripts in the size range of 30 to 70 nt. This transcript pattern was highly reminiscent of a phenomenon called RNA Pol II pausing, which in recent years has been realized to be widespread in metazoans (41). For as many as an estimated 30% of all promoters (41), RNA Pol II can pause and accumulate reversibly at very high levels 30 to 60 nt downstream of the TSS (36). RNA Pol II pausing and the subsequent release to the phase of elongation have been realized to be key steps in transcriptional regulation which also affect cotranscriptional processes, such as splicing and termination (36). In addition to the AAV transcription pattern observed in the RPAs, several indirect lines of evidence indicate that the isolated AAV-2 p40 promoter may be subject to RNA Pol II pausing. A detailed analysis of a paused versus a constitutive promoter has revealed that some well-characterized sequence motifs are overrepresented in the former (42, 43). Whereas a TATA box is not significantly associated with paused promoters, initiator elements (Inr) like the one found directly around the p40 TSS (Fig. 9A) are enriched in paused promoters, as is the so-called pause button (PB) (42), which may also be regarded as a core promoter element. A sequence matching the PB consensus sequence is found downstream from the p40 TSS at nt positions +26 to +32 (Fig. 9). In addition, a GAGA motif is also found downstream from the TSS at nt 1968 to 1975 and is thereby located in the part of the intron identified by mutational analysis to be of major importance for inhibition. The combined presence of GAGA, Inr, and PB elements has been found to be the most predictive for paused promoters with a 7.5-fold enrichment over constitutive promoters. As noted before, Pol II pausing, especially in exon regions (44), is proposed to have a role in promoting splicing, and it may therefore be no mere coincidence that the major AAV-2 splice site is located 53 bp downstream from the p40 TSS.

These direct and indirect lines of evidence point to a high level of paused Pol II proximal to the p40 promoter and, together with the other data presented in this study, suggest a possible mechanism for AAV RIS-Ad-mediated inhibition of adenoviral replication in *cis*. Although the proposed mechanism remains speculative at present and will need further investigation, its implications are nevertheless intriguing. We envision that a stalled RNA Pol II transcription complex downstream from the p40 promoter might keep chromatin-associated adenoviral DNA (45) in an open con-

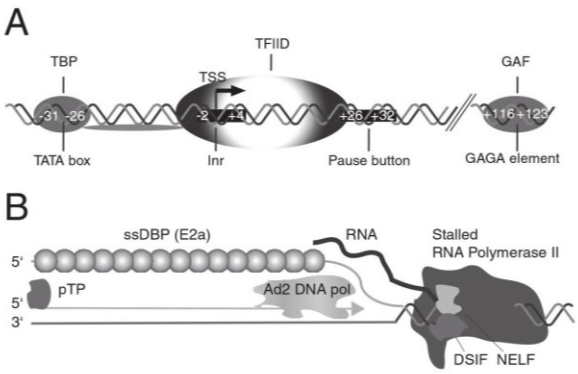


FIG 9 Elements associated with RNA Pol II pausing at the AAV-2 p40 promoter and a model for *cis* inhibition of adenoviral replication by RNA Pol II pausing. (A) Elements in the AAV-2 p40 promoter possibly involved in RNA Pol II pausing. The combination of the Inr element, the pause button (PB), and the GAGA element has been shown to be highly predictive for promoters with disproportionately high levels of Pol II near the TSS (42). Nucleotide positions are indicated relative to the TSS. Abbreviations: TBP, TATA box binding protein; GAF, GAGA factor. (B) Proposed model of inhibition of Ad5/AAV hybrid vector genome replication through steric hindrance by a stalled RNA Pol II complex at the p40 promoter. Abbreviations: DSIF, DRB sensitivity-inducing factor; NELF, negative elongation factor; pTP, preterminal protein; ssDBP, single-stranded DNA binding protein; TFIID, general transcription factor IID.

formation and sterically interfere with adenoviral DNA replication, which involves the adenovirus genome-encoded DNA polymerase, the terminal protein, and the single-stranded DNA binding protein (Fig. 9B). During rescue, adenoviral factors expressed in excess amounts after a replication-competent adenoviral genome is provided in *trans* may directly or indirectly lead to a release of the well-characterized cellular factors that stabilize the paused Pol II complex (41, 46), like negative elongation factor (NELF) and dichloro-1-ß-D-ribofuranosylbenzimidazole (DRB) sensitivity-inducing factor (DSIF).

Due to the resulting mixture of two or (through homologous recombination events) even more different vectors, the rescue of replication-deficient Ad5/AAV hybrid vectors by a replication-competent vector very obviously is not the method of choice to produce large quantities of adenoviral vectors producing the AAV Rep and/or Cap protein. Future studies will address the possibility of the use of direct interference with paused transcription at the p40 promoter to relieve the inhibition of adenoviral replication, for example, by artificial tethering of the positive transcription elongation factor (p-TEFb), as has been described for the Hsp70 promoter (47).

Our findings that the 3′ part of the AAV *rep* gene, dependent on p40-mediated transcription, exhibits strong inhibitory *cis* effects on DNA replication also have important implications for the basic regulation of the AAV replication cycle in the absence and the presence of helper virus. Both AAV capsid protein expression and DNA replication are completely dependent on the combined action of helper virus functions and the AAV large Rep proteins. We therefore propose that helper virus functions and/or the large Rep proteins may interfere with paused RNA Pol II at the p40 promoter to coordinately stimulate transcript elongation and splicing (48) as well as AAV DNA replication. Interestingly, when RIS-Ad is extended at the 5′ end by the remaining part of the *rep* gene, including the p5 promoter, the inhibition of adenoviral replication in *cis* is at least partially relieved (10). Whereas we had not explicitly determined whether this effect was due to Rep protein expression or the additional *cis*-acting sequences, a possible function of RNA Pol II pausing in the regulation of AAV gene expression seems worthy of being addressed in further studies.

FUNDING INFORMATION

C.S. was supported by a Charité-funded Ph.D. fellowship. R.H. acknowledges support from the Sonnenfeld-Stiftung, Berlin.

REFERENCES

1. **Atchison RW, Casto BC, Hammon WM.** 1965. Adenovirus-associated defective virus particles. Science **149:**754–756. http://dx.doi.org/10.1126/science.149.3685.754.
2. **Buller RM, Janik JE, Sebring ED, Rose JA.** 1981. Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. J Virol **40:**241–247.
3. **Berns KI, Giraud C.** 1996. Biology of adeno-associated virus. Curr Top Microbiol Immunol **218:**1–23.
4. **Huser D, Gogol-Doring A, Lutter T, Weger S, Winter K, Hammer EM, Cathomen T, Reinert K, Heilbronn R.** 2010. Integration preferences of wildtype AAV-2 for consensus rep-binding sites at numerous loci in the human genome. PLoS Pathog **6:**e1000985. http://dx.doi.org/10.1371/journal.ppat.1000985.
5. **Casto BC, Armstrong JA, Atchison RW, Hammon WM.** 1967. Studies on the relationship between adeno-associated virus type 1 (AAV-1) and adenoviruses. II. Inhibition of adenovirus plaques by AAV; its nature and specificity. Virology **33:**452–458.
6. **Casto BC, Atchison RW, Hammon WM.** 1967. Studies on the relationship between adeno-associated virus type I (AAV-1) and adenoviruses. I. Replication of AAV-1 in certain cell cultures and its effect on helper adenovirus. Virology **32:**52–59.
7. **Timpe JM, Verrill KC, Trempe JP.** 2006. Effects of adeno-associated virus on adenovirus replication and gene expression during coinfection. J Virol **80:**7807–7815. http://dx.doi.org/10.1128/JVI.00198-06.
8. **Glauser DL, Seyffert M, Strasser R, Franchini M, Laimbacher AS, Dresch C, de Oliveira AP, Vogel R, Buning H, Salvetti A, Ackermann M, Fraefel C.** 2010. Inhibition of herpes simplex virus type 1 replication by adeno-associated virus Rep proteins depends on their combined DNA-binding and ATPase/helicase activities. J Virol **84:**3808–3824. http://dx.doi.org/10.1128/JVI.01503-09.
9. **Glauser DL, Strasser R, Laimbacher AS, Saydam O, Clement N, Linden RM, Ackermann M, Fraefel C.** 2007. Live covisualization of competing adeno-associated virus and herpes simplex virus type 1 DNA replication: molecular mechanisms of interaction. J Virol **81:**4732–4743. http://dx.doi.org/10.1128/JVI.02476-06.
10. **Weger S, Hammer E, Heilbronn R.** 2014. Differential contribution of adeno-associated virus type 2 Rep protein expression and nucleic acid elements to inhibition of adenoviral replication in *cis* and in *trans*. J Virol **88:**14126–14137. http://dx.doi.org/10.1128/JVI.02350-14.
11. **Maguire AM, High KA, Auricchio A, Wright JF, Pierce EA, Testa F, Mingozzi F, Bennicelli JL, Ying GS, Rossi S, Fulton A, Marshall KA, Banfi S, Chung DC, Morgan JI, Hauck B, Zelenaia O, Zhu X, Raffini L, Coppieters F, De Baere E, Shindler KS, Volpe NJ, Surace EM, Acerra C, Lyubarsky A, Redmond TM, Stone E, Sun J, McDonnell JW, Leroy BP, Simonelli F, Bennett J.** 2009. Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. Lancet **374:**1597–1605. http://dx.doi.org/10.1016/S0140-6736(09)61836-5.
12. **Mingozzi F, High KA.** 2011. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet **12:**341–355. http://dx.doi.org/10.1038/nrg2988.
13. **Grimm D, Kern A, Rittner K, Kleinschmidt JA.** 1998. Novel tools for production and purification of recombinant adenoassociated virus vectors. Hum Gene Ther **9:**2745–2760. http://dx.doi.org/10.1089/hum.1998.9.18-2745.
14. **Xiao X, Li J, Samulski RJ.** 1998. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. J Virol **72:**2224–2232.
15. **Clement N, Knop DR, Byrne BJ.** 2009. Large-scale adeno-associated viral vector production using a herpesvirus-based system enables manufactur-

ing for clinical studies. Hum Gene Ther **20:**796–806. http://dx.doi.org/10.1089/hum.2009.094.
16. **Conway JE, Rhys CM, Zolotukhin I, Zolotukhin S, Muzyczka N, Hayward GS, Byrne BJ.** 1999. High-titer recombinant adeno-associated virus production utilizing a recombinant herpes simplex virus type I vector expressing AAV-2 Rep and Cap. Gene Ther **6:**986–993. http://dx.doi.org/10.1038/sj.gt.3300937.
17. **Thomas DL, Wang L, Niamke J, Liu J, Kang W, Scotti MM, Ye GJ, Veres G, Knop DR.** 2009. Scalable recombinant adeno-associated virus production using recombinant herpes simplex virus type 1 coinfection of suspension-adapted mammalian cells. Hum Gene Ther **20:**861–870. http://dx.doi.org/10.1089/hum.2009.004.
18. **Cecchini S, Virag T, Kotin RM.** 2011. Reproducible high yields of recombinant adeno-associated virus produced using invertebrate cells in 0.02- to 200-liter cultures. Hum Gene Ther **22:**1021–1030. http://dx.doi.org/10.1089/hum.2010.250.
19. **Urabe M, Ding C, Kotin RM.** 2002. Insect cells as a factory to produce adeno-associated virus type 2 vectors. Hum Gene Ther **13:**1935–1943. http://dx.doi.org/10.1089/10430340260355347.
20. **Mietzsch M, Grasse S, Zurawski C, Weger S, Bennett A, Agbandje-McKenna M, Muzyczka N, Zolotukhin S, Heilbronn R.** 2014. OneBac: platform for scalable and high-titer production of adeno-associated virus serotype 1-12 vectors for gene therapy. Hum Gene Ther **25:**212–222. http://dx.doi.org/10.1089/hum.2013.184.
21. **Zhang HG, Wang YM, Xie JF, Liang X, Hsu HC, Zhang X, Douglas J, Curiel DT, Mountz JD.** 2001. Recombinant adenovirus expressing adeno-associated virus Cap and Rep proteins supports production of high-titer recombinant adeno-associated virus. Gene Ther **8:**704–712. http://dx.doi.org/10.1038/sj.gt.3301454.
22. **Carlson CA, Shayakhmetov DM, Lieber A.** 2002. An adenoviral expression system for AAV rep78 using homologous recombination. Mol Ther **6:**91–98. http://dx.doi.org/10.1006/mthe.2002.0634.
23. **Recchia A, Parks RJ, Lamartina S, Toniatti C, Pieroni L, Palombo F, Ciliberto G, Graham FL, Cortese R, La Monica N, Colloca S.** 1999. Site-specific integration mediated by a hybrid adenovirus/adeno-associated virus vector. Proc Natl Acad Sci U S A **96:**2615–2620. http://dx.doi.org/10.1073/pnas.96.6.2615.
24. **Recchia A, Perani L, Sartori D, Olgiati C, Mavilio F.** 2004. Site-specific integration of functional transgenes into the human genome by adeno/AAV hybrid vectors. Mol Ther **10:**660–670. http://dx.doi.org/10.1016/j.ymthe.2004.07.003.
25. **Wang H, Lieber A.** 2006. A helper-dependent capsid-modified adenovirus vector expressing adeno-associated virus Rep78 mediates site-specific integration of a 27-kilobase transgene cassette. J Virol **80:**11699–11709. http://dx.doi.org/10.1128/JVI.00779-06.
26. **Sitaraman V, Hearing P, Ward CB, Gnatenko DV, Wimmer E, Mueller S, Skiena S, Bahou WF.** 2011. Computationally designed adeno-associated virus (AAV) Rep 78 is efficiently maintained within an adenovirus vector. Proc Natl Acad Sci U S A **108:**14294–14299. http://dx.doi.org/10.1073/pnas.1102883108.
27. **Winter K, von Kietzell K, Heilbronn R, Pozzuto T, Fechner H, Weger S.** 2012. Roles of E4orf6 and VA I RNA in adenovirus-mediated stimulation of human parvovirus B19 DNA replication and structural gene expression. J Virol **86:**5099–5109. http://dx.doi.org/10.1128/JVI.06991-11.
28. **He TC, Zhou S, da Costa LT, Yu J, Kinzler KW, Vogelstein B.** 1998. A simplified system for generating recombinant adenoviruses. Proc Natl Acad Sci U S A **95:**2509–2514. http://dx.doi.org/10.1073/pnas.95.5.2509.
29. **Luo J, Deng ZL, Luo X, Tang N, Song WX, Chen J, Sharff KA, Luu HH, Haydon RC, Kinzler KW, Vogelstein B, He TC.** 2007. A protocol for rapid generation of recombinant adenoviruses using the AdEasy system. Nat Protoc **2:**1236–1247. http://dx.doi.org/10.1038/nprot.2007.135.
30. **Heilbronn R, Burkle A, Stephan S, zur Hausen H.** 1990. The adeno-associated virus rep gene suppresses herpes simplex virus-induced DNA amplification. J Virol **64:**3012–3018.
31. **Wistuba A, Kern A, Weger S, Grimm D, Kleinschmidt JA.** 1997. Subcellular compartmentalization of adeno-associated virus type 2 assembly. J Virol **71:**1341–1352.
32. **de Wet JR, Wood KV, DeLuca M, Helinski DR, Subramani S.** 1987. Firefly luciferase gene: structure and expression in mammalian cells. Mol Cell Biol **7:**725–737. http://dx.doi.org/10.1128/MCB.7.2.725.
33. **Sato K, Hamada M, Asai K, Mituyama T.** 2009. CENTROIDFOLD: a web server for RNA secondary structure prediction. Nucleic Acids Res **37:**W277–W280. http://dx.doi.org/10.1093/nar/gkp367.
34. **Hamada M, Kiryu H, Sato K, Mituyama T, Asai K.** 2009. Prediction of RNA secondary structure using generalized centroid estimators. Bioinformatics **25:**465–473. http://dx.doi.org/10.1093/bioinformatics/btn601.
35. **McCarty DM, Christensen M, Muzyczka N.** 1991. Sequences required for coordinate induction of adeno-associated virus p19 and p40 promoters by Rep protein. J Virol **65:**2936–2945.
36. **Jonkers I, Lis JT.** 2015. Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol **16:**167–177. http://dx.doi.org/10.1038/nrm3953.
37. **Di Pasquale G, Chiorini JA.** 2003. PKA/PrKX activity is a modulator of AAV/adenovirus interaction. EMBO J **22:**1716–1724. http://dx.doi.org/10.1093/emboj/cdg153.
38. **Weitzman MD, Fisher KJ, Wilson JM.** 1996. Recruitment of wild-type and recombinant adeno-associated virus into adenovirus replication centers. J Virol **70:**1845–1854.
39. **Fisher KJ, Kelley WM, Burda JF, Wilson JM.** 1996. A novel adenovirus-adeno-associated virus hybrid vector that displays efficient rescue and delivery of the AAV genome. Hum Gene Ther **7:**2079–2087. http://dx.doi.org/10.1089/hum.1996.7.17-2079.
40. **Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R.** 2014. The transcription factor titration effect dictates level of gene expression. Cell **156:**1312–1323. http://dx.doi.org/10.1016/j.cell.2014.02.022.
41. **Adelman K, Lis JT.** 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nat Rev Genet **13:**720–731. http://dx.doi.org/10.1038/nrg3293.
42. **Hendrix DA, Hong JW, Zeitlinger J, Rokhsar DS, Levine MS.** 2008. Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. Proc Natl Acad Sci U S A **105:**7762–7767. http://dx.doi.org/10.1073/pnas.0802406105.
43. **Kwak H, Fuda NJ, Core LJ, Lis JT.** 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science **339:**950–953. http://dx.doi.org/10.1126/science.1229386.
44. **Carrillo Oesterreich F, Preibisch S, Neugebauer KM.** 2010. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol Cell **40:**571–581. http://dx.doi.org/10.1016/j.molcel.2010.11.004.
45. **Giberson AN, Davidson AR, Parks RJ.** 2012. Chromatin structure of adenovirus DNA throughout infection. Nucleic Acids Res **40:**2369–2376. http://dx.doi.org/10.1093/nar/gkr1076.
46. **Kwak H, Lis JT.** 2013. Control of transcriptional elongation. Annu Rev Genet **47:**483–508. http://dx.doi.org/10.1146/annurev-genet-110711-155440.
47. **Lis JT, Mason P, Peng J, Price DH, Werner J.** 2000. P-TEFb kinase recruitment and function at heat shock loci. Genes Dev **14:**792–803.
48. **Qiu J, Pintel DJ.** 2002. The adeno-associated virus type 2 Rep protein regulates RNA processing via interaction with the transcription template. Mol Cell Biol **22:**3639–3652. http://dx.doi.org/10.1128/MCB.22.11.3639-3652.2002.

2.4 Definition of Herpes Simplex Virus Helper Functions for the Replication of Adeno-Associated Virus Type 5

Authors: Catrin Stutika, Daniela Hüser, Stefan Weger, Natalja Rutz, Melanie Heßler, Regine Heilbronn

Year: 2015

Journal: Journal of General Virology 96 (Pt4): 840-850

2.4.1 Contribution to the Publication

The aim of this study was to define the individual HSV1 helper genes required for productive replication of the distantly related serotype AAV5.

The initial experimental design for this publication was developed by my mentor Regine Heilbronn. All experiments were performed by me with assistance by a technician, Melanie Heßler, whom I supervised during this study. I analyzed and evaluated the resulting data. All illustrations and their descriptions, with the exception of Fig. 4A were generated by me. The co-localization of AAV5 Rep and HSV1 ICP8 displayed in Fig. 4A was performed and analyzed by Natalja Rutz and Stefan Weger. The manuscript was written by my mentor Regine Heilbronn and me. Furthermore, I presented the data ahead of publication on the XV. International Parvovirus Workshop in 2014.

2.4.2 Article

DOI: http://dx.doi.org/10.1099/vir.0.000034

Definition of Herpes Simplex Virus Helper Functions for the Replication of Adeno-Associated Virus Type 5

Catrin Stutika, Daniela Hüser, Stefan Weger, Natalja Rutz, Melanie Heßler, and Regine Heilbronn*

Institute of Virology, Campus Benjamin Franklin, Charité Medical School, Germany

*To whom correspondence should be addressed:

Regine Heilbronn, Institute of Virology, Campus Benjamin Franklin, Charité Universitätsmedizin, Berlin, Hindenburgdamm 27, 12203 Berlin, Germany; phone +49 30 84453696; fax +49 30 84454485; e-mail regine.heilbronn@charite.de

Running title: HSV helper functions for AAV5 replication

Abstract

Adeno-associated virus (AAV) type 5 represents the genetically most distant AAV serotype and the only isolated directly from human tissue. Seroepidemiological evidence suggests HSV as helper virus for human AAV5 infections underlining the *in vivo* relevance of the AAV-herpesvirus relationship. In this study we analyzed for the first time herpes simplex virus (HSV) helper functions for productive AAV5 replication and compared these to AAV2. Using a combination of HSV strains and plasmids for individual genes, the previously defined HSV helper functions for AAV2 replication were shown to induce AAV5 gene expression, DNA replication and production of infectious progeny. The helper functions comprise the replication genes for ICP8 (UL29), helicase-primase (UL5/8/52), and DNA polymerase (UL30/42). HSV immediate-early genes for ICP0 and ICP4 further enhanced AAV5 replication, mainly by induction of *rep* gene expression. In the presence of HSV helper functions AAV5 Rep colocalized with ICP8 in nuclear replication compartments and HSV alkaline exonuclease (UL12) enhanced AAV5 replication, similar to AAV2. UL12 in combination with ICP8 was shown to induce DNA strand exchange on partially double- stranded templates to resolve and repair concatemeric HSV replication intermediates. Similarly, concatemeric AAV replication intermediates appeared to be processed to yield AAV unit-length molecules, ready for AAV packaging. Taken together, our findings show that productive AAV5 replication is promoted by the same combination of HSV helper functions as AAV2.

Introduction

Adeno-associated virus type 5 (AAV5) represents the genetically most distant member of the AAV family with AAV2 being considered as prototype strain. In contrast to the AAV serotypes 1 to 4 isolated as contaminants of human or simian adenovirus stocks (Atchison *et al.*, 1965; Hoggan *et al.*, 1966; Parks *et al.*, 1967), AAV5 was directly isolated from human tissue (Bantel-Schaal & zur Hausen, 1984). Although it was obvious from the beginning that AAV5 infection was prevalent in the human population (Georg-Fries *et al.*, 1984; Samulski *et al.*, 1982) research focused on AAV2, the first AAV strain to be cloned and sequenced (Samulski *et al.*, 1982; Srivastava *et al.*, 1983) as prerequisites for further molecular analysis and the development of AAV2-based gene therapy vectors.

The AAVs are small replication-deficient, and apathogenic viruses that require coinfection with an unrelated helper virus for productive infection. Besides adenovirus (Ad) all herpesvirus genera tested so far, including HSV, CMV, and HHV6 were shown to support productive AAV2 replication (Blacklow *et al.*, 1967; Buller *et al.*, 1981; Georg-Fries *et al.*, 1984; McPherson *et al.*, 1985; Thomson *et al.*, 1994). In the absence of a helper virus AAV2 was shown to establish latency. Both, chromosomal integration and episomal modes of persistence have been described (Hüser *et al.*, 2014; Hüser *et al.*, 2010; Schnepp *et al.*, 2005).

AAV5 has a 4.7 kb ssDNA genome that comprises two open reading frames, one coding for the capsid proteins (*cap*) and another (*rep*) coding for a family of non-structural regulatory proteins (Chiorini *et al.*, 1999b). Whereas AAV2 expresses Rep78, Rep52 and C-terminally spliced versions thereof, called Rep68 and Rep40, AAV5 only transcribes unspliced mRNAs leading to Rep78 and Rep52 (Qiu *et al.*, 2002). In addition, a Rep40-like protein generated by alternative translation initiation was recently identified (Farris & Pintel, 2010; Fasina & Pintel, 2013). In analogy to AAV2, AAV5 Rep78 is assumed to regulate AAV gene expression and DNA replication. AAV5 Rep was shown to bind to the AAV5 inverted terminal repeats (ITR) that flank the genome and serve as origins of AAV5 DNA replication. Whereas the ITRs of most AAV serotypes are relatively conserved and in part interchangeable, the AAV5-ITR is more divergent and genetically unstable (Chiorini *et al.*, 1999a; Chiorini *et al.*, 1999b). This has long hampered the cloning of a fully infectious wild- type AAV5 genome as a prerequisite for further analysis (Chiorini *et al.*, 1999b). On the other hand, Rep78 derived from AAV5 was the first Rep protein to be

crystallized together with its cognate ITR visualizing Rep-binding to the conserved Rep-binding site (RBS) within the AAV5-ITR (Hickman *et al.*, 2002; Hickman *et al.*, 2004). A conformational change of the ITR upon binding of Rep leads to nicking of the adjacent terminal resolution site (trs) required for processive AAV DNA replication (Chiorini *et al.*, 1999a).

In cell culture AAV5 can be propagated by coinfection with either adenovirus or HSV. In contrast to what was assumed from AAV1-4, seroepidemiology for AAV5 showed a parallel increase of HSV and AAV5 antibody titers from early childhood until adulthood suggestive of HSV as a preferred *in vivo* helper virus (Georg-Fries *et al.*, 1984). HSV functions involved in AAV5 replication have not yet been analyzed. For productive AAV2 replication the responsible HSV functions consist of a subset of six HSV replication genes comprising those for the HSV DNA-binding protein, also called ICP8 (UL29), the three-component helicase- primase complex (HP) encoded by UL5, UL8 and UL52 and the two-component HSV DNA polymerase (UL30/UL42). The minimal HSV four protein complex consisting of ICP8 and HP (UL5/8/52/29) leads to low-level, productive AAV2 replication (Weindler & Heilbronn, 1991). *In vivo*, AAV2 Rep colocalizes to ICP8 in a ssDNA-dependent manner for which complex formation with UL5/8/52 is required (Alex *et al.*, 2012; Heilbronn *et al.*, 2003; Slanina *et al.*, 2006). The HSV immediate-early gene for ICP0 was shown to transactivate *rep* gene expression from an integrated AAV2 genome (Geoffroy *et al.*, 2004) and together with the HSV immediate-early functions ICP4 and ICP22 further enhances AAV replication in the presence of the six identified HSV replication genes (Alazard-Dany *et al.*, 2009; Nicolas *et al.*, 2010). More recently, HSV alkaline exonuclease encoded by UL12 was shown to add to AAV replication, likely by helping to resolve concatemeric AAV replication intermediates (Nicolas *et al.*, 2010). Specific HSV mutants defective in these genes (Alazard-Dany *et al.*, 2009; Nicolas *et al.*, 2010; Slanina *et al.*, 2006; Weindler & Heilbronn, 1991) corroborated the transfection data. The HSV *ori*-binding protein (UL9) is dispensable (Weindler & Heilbronn, 1991). This is not surprising since AAV Rep serves as *ori*-binding protein attracting HSV-ICP8 to the AAV-ITR, comparable to the action of UL9 on the HSV *ori* (Alex *et al.*, 2012; Heilbronn *et al.*, 2003). Upon infection with HSVΔUL9 HSV DNA replication was blocked whereas AAVwt replication was undiminished (Alazard-Dany *et al.*, 2009; Weindler & Heilbronn, 1991). These findings clearly document that all HSV helper functions required for AAV2 replication are expressed before HSV DNA replication.

In this study we tested the HSV helper functions previously identified for AAV2 for their role in AAV5 replication. By a combination of infection and transfection experiments we analyzed AAV5 gene expression, DNA replication and production of infectious AAV5 particles. The dependence of AAV5 replication on various combinations of transfected HSV functions were analyzed and compared side by side to the results of HSV-induced AAV2 replication.

Results

Comparison of AAV5 and -2 replication with Ad-2 or HSV-1 as helper virus

Initial experiments were performed to compare AAV5 replication in HeLa cells upon coinfection with Ad-5, Ad-2, or HSV-1. All helper virus strains led to very comparable levels of *rep* and *cap* gene expression, AAV DNA replication and AAV particle production (data not shown). In time course experiments AAV2, as well as AAV5 DNA replication with either helper virus reached a plateau at 24 hours that remained stable until 48 hours post infection (h.p.i.). To evaluate AAV5 replication upon plasmid transfection, pAAV5, containing the AAV5 wild-type genome, was transfected into HeLa cells, which were subsequently infected with HSV-1 (MOI = 5) or Ad-2 (MOI = 5), respectively. At 48 h.p.i. cells were harvested for parallel analysis of AAV *rep* and *cap* gene expression, AAV DNA replication, and AAV virus production. As shown in Fig. 1A, AAV5 Rep78, Rep52, and a Rep40-like protein are induced with either HSV-1 or Ad-2 infection. The Rep expression profile is indistinguishable from the one described for AAV5 induced by Ad-5 helper genes (Nayak & Pintel, 2007). Expression of VP proteins is slightly higher in the presence of HSV-1 (Fig. 1A). A parallel experiment was conducted transfecting an AAV2 wild-type plasmid (pAAV2), leading to the AAV2 typical expression pattern for Rep and VP proteins (Fig. 1B). AAV DNA replication was analyzed on Southern blots, where the AAV replication forms (RF) indicative of AAV DNA replication are visible throughout. RF1 represent 4.7 kb double-stranded AAV monomers, and RF2, 9.4 kb double-stranded AAV dimers (Fig. 1C). Production of AAV5 virus particles was quantified by qPCR as genomic particles (gp/ml) determined in cell extracts, as outlined in the methods. High titers of AAV5 were produced with either helper virus (Fig. 1D).

HSV helper functions for AAV5 replication are expressed before HSV DNA replication

To identify the HSV functions responsible for AAV5 replication, pAAV5-transfected cells were infected with HSVΔUL9, a replication-deficient host range mutant for the HSV *ori*-binding protein (Malik *et al.*, 1992) previously shown to support AAV2 replication similarly well as wild-type HSV-1 (Weindler & Heilbronn, 1991). In the time course experiments displayed in Fig. 2, the AAV5 Rep and VP protein profiles are identical in cells infected with HSV-1, or HSVΔUL9, respectively. HSV UL5 helicase, expressed as "early" protein, shows stable expression levels for either HSV strain, whereas UL9 is only expressed in HSV-1 infected cells. As expected, UL9 is not expressed in cells infected with HSVΔUL9 (Fig. 2B, D). Similar results were achieved in a parallel experiment with pAAV2 (Fig. 2C, D). AAV virus production was analyzed by qPCR in benzonase-treated freeze-thaw cell extracts, as outlined in the methods. AAV5 titers up to $5x10^{10}$ gp/ml were reached at 48 h.p.i. with HSV-1, or HSVΔUL9 as helper virus. AAV2 reach higher titers up to $8x10^{11}$ gp/ml (Fig. 2E). In summary, the HSV functions required for AAV5 replication are fully expressed before and in the absence of HSV DNA replication, as shown previously for AAV2 (Weindler & Heilbronn, 1991).

HSV immediate-early genes induce AAV5 rep gene expression

The results shown so far suggested that the previously identified set of HSV helper genes described for AAV2 replication are similarly involved in productive AAV5 replication. The set of HSV helper genes consists of the immediate-early (IE) genes for ICP0 and ICP4, which transactivate the six identified "early" HSV replication genes (UL5, -8, 29, -30, -42, -52). Together these are required for productive AAV2 replication (Weindler & Heilbronn, 1991). Subsequently, ICP0 was shown to induce *rep* gene expression from chromosomally integrated AAV2 genomes (Alazard-Dany *et al.*, 2009; Geoffroy *et al.*, 2004). ICP4 alone did not induce Rep but enhanced its expression induced by ICP0 in the absence of detectable AAV DNA replication (Alazard-Dany *et al.*, 2009). Here, the impact of HSV-IE genes on AAV5 gene expression was analyzed, as displayed in Fig. 2F. The combination of ICP0 and ICP4 led to the highest Rep protein levels in the absence of VP proteins. Marginal VP3 levels were detected upon overexposure of the blot (data not shown). ICP22 did not further enhance AAV5 expression (Fig. 2F). Therefore, the combination of ICP0 with ICP4 was selected for the subsequent analysis of HSV helper functions for AAV5 replication.

Identification of HSV replication genes as helper functions for AAV5 replication

The finding of undiminished productive AAV5 replication in the presence of HSVΔUL9 was reminiscent of our previous study identifying the HSV helicase-primase complex (UL5/UL8/UL52) and the single-strand DNA-binding protein ICP8 (UL29) as minimal helper functions for AAV2 and the two-component HSV DNA polymerase (UL30/UL42) further enhancing AAV2 replication (Weindler & Heilbronn, 1991). Recently, the HSV-1 alkaline exonuclease (UL12) was identified as additional enhancing factor when combined with the identified set of HSV-IE-, and DNA replication genes (Nicolas *et al.*, 2010). To test the assumption that these genes are also involved in AAV5 replication induced by HSV, various combinations thereof were co-transfected with pAAV5 and analyzed for expression of *rep* and *cap*, for AAV5 DNA replication and the production of infectious AAV5 particles (Fig. 3). Rep expression levels are highest with the combination of ICP0, ICP4 and the six HSV replication genes for ICP8 (UL29), helicase-primase (UL5/8/52) and DNA polymerase (UL30/UL42) (Fig. 3A, upper panel). This set of genes also led to the highest levels of VP proteins (Fig. 3A, lower panel). Expression of HSV exonuclease (UL12) did not further enhance AAV5 expression levels. The identical experimental set-up performed with AAV2 led to comparable results (Fig. 3B). The data suggest that VP expression is most pronounced when HSV-IE genes are combined with HSV replication genes that lead to AAV template amplification as outlined below (Fig. 3C and D).

DNA replication of AAV5, or AAV2 was analyzed on Southern blots of size-separated, genomic DNAs from Hirt extracts digested with *Dpn*I to degrade input plasmid DNA. Upon infection with HSV, the typical RF1 and RF2 bands are visible for AAV2 and for AAV5 (Fig. 1C; Fig. 3C and D lane 10). Upon cotransfection of HSV helper genes AAV2 replication is apparent as described before (Alazard-Dany *et al.*, 2009; Weindler & Heilbronn, 1991) whereas AAV5 DNA replication is barely detectable (Fig. 3C). Upon cotransfection of UL12, RF1 and RF2 become clearly visible (Fig. 3C lane 7). The specificity of the UL12 effect is underlined by the disappearance of RF bands upon addition of an exonuclease-deficient version of UL12 (Fig. 3C lane 8), as initially described for AAV2 (Nicolas *et al.*, 2010). In summary, AAV5 DNA replication increases with increasing numbers of HSV helper genes, which is consistent with a parallel increase of VP protein levels.

Production of infectious AAV particles was analyzed in second round infection with cleared supernatants of freeze-thaw extracts. Since AAV titers were initially higher with

AAV2 (Fig. 2E), the AAV2 extracts were prediluted as outlined in the methods. AAV titers were quantified in triplicate cultures of Ad-2-infected HeLa cells as gp/ml by qPCR as outlined in the methods. All combinations of HSV helper functions led to production of infectious AAV and titers increased with increasing numbers of HSV helper genes (Fig. 3E and F). In summary, HSV helper functions for AAV5 replication are identical to the ones described before for AAV2 (Alazard-Dany *et al.*, 2009; Nicolas *et al.*, 2010; Weindler & Heilbronn, 1991). The highest AAV5 replication rates are achieved by the combined expression of the HSV-IE genes, ICP0 and ICP4, six HSV replication genes, UL5/8/29/30/42/52, and HSV alkaline exonuclease UL12.

Nuclear colocalization of AAV5 Rep and HSV ICP8

We have shown before that Rep of AAV2 colocalizes to HSV ICP8 in nuclear replication compartments in an AAV-ssDNA-dependent manner (Alex *et al.*, 2012; Heilbronn *et al.*, 2003; Slanina *et al.*, 2006). ICP8 and the helicase-primase complex (UL5/8/52) are required to initiate the formation of the replication compartments that further mature when the two- component HSV polymerase (UL30/42) is coexpressed (Wilkinson & Weller, 2004). Here we studied colocalization of AAV5 Rep (tagged with HA at the carboxy terminus) with HSV ICP8 in the presence of different combinations of HSV helper functions. Although colocalization of AAV5 Rep with ICP8 was not as pronounced as previously described for AAV2 (Alex *et al.*, 2012; Heilbronn *et al.*, 2003; Slanina *et al.*, 2006) it was clearly apparent when UL30/42 and ICP0/4 were added (Fig. 4A). Thus, AAV5 Rep can be recruited by HSV to nuclear replication compartments similar to Rep of AAV2 (Alex *et al.*, 2012; Heilbronn *et al.*, 2003; Slanina *et al.*, 2006). The relatively reduced frequency of colocalization may reflect less efficient AAV5 plasmid rescue and/or AAV replication as described above (Fig. 2).

Discussion

In this study the HSV helper functions for productive AAV5 replication were identified to consist of the HSV-IE genes for ICP0 and ICP4, and the HSV replication functions for ssDNA-binding protein ICP8 (UL29), the helicase-primase complex (UL5/8/52), the two-subunit DNA polymerase (UL30/42). In addition, alkaline exonuclease (UL12) adds to AAV5 replication. The results are corroborated by side-by-side analysis with AAV2,

confirming the findings described previously (Alazard-Dany *et al.*, 2009; Nicolas *et al.*, 2010; Weindler & Heilbronn, 1991).

Comparison of HSV and Ad helper functions for AAV5

The major viral helpers for productive AAV2 replication, adenovirus and HSV, were also the most intensely studied in the past. For AAV2 replication five Ad-5 functions, namely E1A, E1B, E2A, E4Orf6, and VA RNA were identified to support DNA replication and production of progeny virus (Muzyczka & Berns, 2001). These induce AAV2 replication by a combination of effects on AAV transcription, translation, induction of DNA second-strand synthesis and diverse modulatory effects on host cell proteins. AAV5 displays a divergent transcription profile, only unspliced mRNAs are transcribed for expression of Rep78 and Rep52 (Qiu *et al.*, 2002). An additional Rep40-like protein is initiated downstream of Rep52 from the same mRNA (Farris & Pintel, 2010; Fasina & Pintel, 2013). In spite of these variations the same combinations of Ad-5 genes that replicate AAV2 were identified to replicate AAV5 (Nayak & Pintel, 2007). In view of the similarity of the HSV- with the Ad- induced AAV5 expression profile described here, it was not surprising to find that the same HSV helper functions identified for AAV2 were active with AAV5 as well. Other than adenovirus, HSV exerts its helper effects, by a combination of transcriptional activation, DNA replication and genome maturation for both AAV2 and AAV5 (Fig. 4B).

Mechanisms of AAV induction by HSV immediate-early genes

The major HSV immediate-early proteins ICP0 and ICP4 serve as transcriptional activators and are constituents of HSV replication compartments in the nucleus. Their role as transactivators of silent HSV "early" replication genes is needed in the context of HSV infection (Slanina *et al.*, 2006; Weindler & Heilbronn, 1991). In the absence of a helper virus, AAV2 gene expression is repressed by Rep78/68, which binds to AAV2 promoters (Beaton *et al.*, 1989; Kyöstiö *et al.*, 1994). ICP0 was shown to induce Rep in a cell line harboring latent AAV2 (Geoffroy *et al.*, 2004). It is long established that Rep78/68 is required but not sufficient for AAV replication. Ad E1A, a potent transcriptional activator relieves the repression and strongly stimulates *rep* gene expression (Chang *et al.*, 1989). HSV ICP0 appears to exert a comparable effect when inducing *rep* from an integrated silent AAV2 genome (Geoffroy *et al.*, 2004) or from transfected plasmids for wild-type

AAV2 or AAV5, as shown here. Similar to E1A, ICP0 does not bind to DNA directly but exerts its transcriptional effects indirectly. HSV mutants with specific amino acid exchanges in ICP0 showed that ICP0 exploits the ubiquitin-proteasome pathway, presumably to degrade factors that repress the AAV2 p5 promoter. Neither the localization of ICP0 to nuclear ND10 domains, nor their destruction was required for *rep* gene expression (Geoffroy *et al.*, 2004). HSV ICP4, the major HSV-IE transcriptional activator had hardly an effect on Rep induction but enhanced the stimulatory effect of ICP0 (Alazard-Dany *et al.*, 2009). The additive effect of ICP0 and ICP4 on Rep of AAV5 shown here, is consistent with previous findings for AAV2 (Alazard-Dany *et al.*, 2009).

AAV ITR-dependent initiation of DNA replication by HSV replication functions

For the initial identification of HSV helper functions for AAV replication, HSV "early" replication genes were expressed from constitutive heterologous promoters to make them independent of transactivation by HSV-IE genes. This led to the identification of the minimal set of four helper genes for AAV replication, encoding ICP8 and the helicase-primase complex. These were sufficient for low-level AAV2 production including *rep* gene expression (Alex *et al.*, 2012; Slanina *et al.*, 2006; Weindler & Heilbronn, 1991). Here we show that ICP8 and helicase-primase also constitute the minimal helper genes for AAV5 replication, although the efficiency is very low. Addition of ICP0/4 stimulates Rep and VP levels and boosts AAV production. In the absence of HSV DNA polymerase a cellular polymerase is assumed to induce AAV5 DNA replication. Leading-strand DNA pol∂ was shown to replicate AAV2 *in vitro* (Nash *et al.*, 2008) and to colocalize *in vivo* with Rep in nuclear HSV/AAV replication compartments (Nicolas *et al.*, 2010). The full replication complex is assembled when the two-component HSV DNA polymerase joins in (Fig. 4B). AAV5 DNA replication and particle production are enhanced by more than 10-fold, as shown for AAV2 replication before (Alazard-Dany *et al.*, 2009; Weindler & Heilbronn, 1991).

AAV DNA replication is initiated by binding of Rep78/68 to the RBS of the ITR. Although the primary DNA sequences are divergent, both AAV2 and AAV5 display comparable hairpin-structured ITRs and the RBS sequences are conserved (Chiorini *et al.*, 1999a; Chiorini *et al.*, 1999b; Hickman *et al.*, 2004). AAV2- and AAV5-Rep78 display DNA- binding, ATPase, helicase and endonuclease activities (Chiorini *et al.*, 1999a; Maggin *et al.*, 2012) required for the initiation of AAV DNA replication. Co-

crystallization of the endonuclease domain of AAV5 Rep with the AAV5-ITR identified conserved amino acids that contact the RBS and the adjacent *trs* on the opposite DNA strand (Hickman *et al.*, 2004). Although the sequence of the *trs* cleavage site differs between AAV2 and AAV5 the mechanisms appear to be similar and nicking is required for ongoing AAV DNA replication (Chiorini *et al.*, 1999a).

HSV-induced AAV2 DNA replication is initiated by AAV ssDNA-dependent ternary complex formation of Rep78 and ICP8 on the hairpin-structured AAV ITR as shown *in vitro*. *In vivo*, Rep78 was shown to colocalize to ICP8 in nuclear replication compartments in an AAV ssDNA-dependent manner (Alex *et al.*, 2012; Heilbronn *et al.*, 2003). Ternary complex formation of Rep and ICP8 on the AAV-ITR is reminiscent of HSV *ori*-binding by the HSV ori-binding protein, UL9 that first recruits ICP8 to cover and stabilize ssDNA regions of the unwinding HSV *ori* (Weller & Coen, 2012). The growing replication fork then recruits the helicase-primase (HP) complex, as shown *in vitro* (Biswas & Weller, 2001; Chen *et al.*, 2011; Weller & Coen, 2012). *In vivo,* ICP8 and HP initiate the formation of nuclear HSV replication compartments (Livingston *et al.*, 2008; Weller & Coen, 2012). With AAV ssDNA genomes as templates, Rep colocalizes to ICP8 in these nuclear foci and the four helper proteins initiate AAV DNA replication (Slanina *et al.*, 2006). We show here that this is the case also for AAV5. In contrast to the HSV *ori*, the AAV-ITR represents a partially double-stranded genome that can adopt a fork-shaped DNA configuration. It contains a self-primed free 3' end, ready for elongation by DNA polymerase even in the absence of UL52 primase activity, as shown by us before (Slanina *et al.*, 2006). The three protein components of the helicase-primase complex were needed, apparently as structural backbone of the replication complex. Here we show that the components of the HSV replication complex for AAV5 replication are the same, leading to the assumption that the details of their interactions can be extrapolated from earlier findings with AAV2 (Fig. 4B).

Role of UL12 for AAV genome maturation

The most recent addition to the list of HSV helper genes was UL12, which codes for alkaline exonuclease. UL12 was identified as Rep-interacting protein and shown to enhance the accumulation of unit-length AAV replication forms (Nicolas *et al.*, 2010). UL12 is not directly involved in HSV DNA replication. In cooperation with ICP8 UL12 mediates DNA strand exchange of newly replicated HSV DNA concatemers, apparently

for repair of DNA damage emerging during rolling circle replication (Reuven *et al.*, 2003). HSV infection was shown to induce DNA repair by homologous recombination and UL12 was identified to interact with the components of the ATM pathway (Balasubramanian *et al.*, 2010). Recently, isolated UL12 exonuclease proved to be sufficient to induce DNA repair by the single-strand annealing (SSA) pathway that exploits small DNA sequence homologies to anneal double- strand breaks at the expense of small deletions of varying sizes. (Schumacher *et al.*, 2012; Weller & Coen, 2012). Here we used catalytically active or inactive UL12 variants to demonstrate that the AAV5 replication pattern with predominant AAV intermediates of undefined lengths is relieved by UL12. Unit-length AAV replication forms accumulate leading to a moderate enhancement of AAV5 production, as shown before for AAV2 (Nicolas *et al.*, 2010). It will be interesting to see whether UL12 by binding to both, Rep and ICP8 induces AAV DNA strand exchange of concatemeric AAV replication intermediates (Fig. 4B). The AAV-ITR contains multiple short DNA repeats which are known to invert and recombine during AAV DNA replication, resulting in variable alternative ITR conformations.

It is conceivable that UL12 exploits these DNA repeats to mediate DNA strand exchange of AAV genomes in conjunction with ICP8. Thereby unit length AAV ssDNA strands will accumulate, ready for packaging. The titer of infectious AAVs directly depends on the proportion of particles with packaged full-length genomes, likely explaining the enhancement of infectivity by addition of UL12 seen here for AAV5 and reported before for AAV2 (Nicolas *et al.*, 2010).

Materials and Methods

Cells and Viruses

HSV-1 strain KOS and the HSVΔUL9 mutant (hr94) (Malik *et al.*, 1992) were grown in Vero cells, as described (Heilbronn *et al.*, 1990). Adenovirus type 2 (Ad-2) was propagated in HeLa cells.

Plasmids

The AAV5 plasmids p5TRV-RepCap and p5TRV-Rep380HACap were kind gifts of D. Pintel. p5TRV-RepCap spans the AAV5 wild-type genome with two 4 nt insertions between the ITRs and the coding region and is abbreviated as pAAV5. Plasmid p5TRV-

380HA contains a hemagglutinin tag (HA) at AAV5 nucleotide position 2158 (Farris & Pintel, 2010). The AAV2 plasmid (pTAV2-0) comprises the AAV2 wild-type genome (abbreviated as pAAV2). Expression plasmids for the HSV-1 replication genes UL5, UL8, UL52, UL29 (ICP8), UL30 (pol), UL42, and plasmids for HSV-1 ICP0 and ICP4 under their cognate promoters were described before (Heilbronn & zur Hausen, 1989). Plasmid pTF3pol for ICP4/0/22, and UL30/UL42 (Alazard-Dany *et al.*, 2009), pSAKUL12 and pUL12exo$^-$ for HSV-1 exonuclease (Nicolas *et al.*, 2010; Reuven *et al.*, 2003) were under heterologous promoter control.

DNA Transfection and Cell Infection

DNA transfection of HeLa cells was performed as described (Winter *et al.*, 2012) and infected with HSV-1 wild-type or HSVΔUL9 at MOI 2 to 5, or Ad-2 at a MOI of 5. For AAV infections equal proportions (1/30) of AAV-containing freeze-thaw supernatants were mixed with Ad-2 (MOI = 20) to co-infect HeLa cells in a total volume of 200 µl per well for 1.5 hours. AAV2-containing cleared supernatants were serially diluted up to 10^{-4} before infection. The inoculum was sucked off, fresh medium was added and the cells were incubated for another 48 hours.

Southern Blot Analysis

DNA samples were generated by Hirt extraction as described (Winter *et al.*, 2012). For the detection of AAV5 and AAV2 DNA replication intermediates, DNA probes were labeled with the DecaLabel DNA Labeling Kit (Thermo Scientific) and α-^{32}P-dCTP for radioactive labeling, or biotin-11-dUTP labeling, according to the manufacturer`s protocol. The AAV5 probe spans a 1.1 kb *Eco*RI fragment within the *rep* ORF (AAV5 nt 532 to 1637). For AAV2 a 1.1 kb *Nco*I / *Eco*RI fragment located in the *rep* ORF (AAV2 nt 626 to 1768) was used. Membranes hybridized with ^{32}P-labeled DNA probes were detected by autoradiography on X-ray films. Biotinylated probes were detected with Streptavidin horseradish peroxidase (High-Sensitivity HRP, Thermo Scientific) by enhanced chemiluminescence (ECL-kit, Perkin Elmer).

Western Blot Analysis

Western blots were prepared as described (Winter *et al.*, 2012). Protein extracts were separated on 10% SDS polyacrylamide gels and transferred to a nitrocellulose membrane (Protran BA85, GE Healthcare). The membrane was reacted with mAb 303.9 (anti-Rep) or mAb B1 (anti-VP) (Progen) at a 1:10 dilution, mAb 152 (anti-UL5) diluted 1:15 (Slanina *et al.*, 2006), and mAb 13924 (anti-UL9) diluted 1:2500 (Stow *et al.*, 1998). The anti-β-actin rabbit polyclonal antibody was diluted 1:5000. A horseradish peroxidase-conjugated anti- mouse or anti-rabbit secondary antibody (1:2500) was added to the membrane for 1.5 hours at RT, followed by ECL detection. For internal loading control membranes were stripped with boiling 0.1% SDS and exposed to anti-β-actin primary antibody.

Quantification of infectious AAV particles

HeLa cells were infected with 100 µl of supernatants, which were heat-inactivated at 56°C for 30 min in the case of HSV infection, and coinfected with Ad-2 for productive AAV replication. At 48 h.p.i. cells were lysed by three freeze-thaw cycles and harvested as crude extracts. Crude extracts were treated with 250 U/ml benzonase (Merck) at 37°C for at least 1 hour to degrade residual plasmid DNA. After centrifugation at 8000xg for 20 min a 500 µl aliquot of each supernatant was lysed in buffer (1% [w/v] N-Lauroylsarcosine, 25 mM Tris pH 8.5, 10 mM EDTA pH 8.0) containing proteinase K (Roth, Germany) at 56°C for 2 hours, to degrade AAV capsids. AAV DNA was purified by phenol-chloroform extraction and precipitation with ethanol. Copy numbers of AAV genomes were quantified on a Light- Cycler using the QuantiTect SYBR Green PCR Kit (Qiagen). PCR primers used for AAV5 genome quantification were AAV5 Rep for (5'-GTC CCA TTT GAC GTG GAG GAA CA- 3') and AAV5 Rep rev (5'-GGG GTT CAA TTC CCT GGA AGA CC-3'). PCR primers used for AAV2 genome analysis were pAAV-F1 (5'-GCC AAC TCC ATC ACT AGG GG-3') and pAAV-W1 (5'-CCC GCT TCA AAA TGG AGA CC-3').

Immunofluorescence analysis

HeLa cells were grown on coverslips and transfected as described above. After fixation in 3.7% formaldehyde in PBS at pH 7.4 for 30 min, cells were permeabilized for 10 min in 1% Triton X-100 in PBS and washed twice with PBS before reaction with primary antibodies for 60 min. After three washes in PBS, cells were reacted for 45 min with secondary antibodies. All incubations were performed at room temperature. Primary

antibody mixture contained monoclonal mouse-anti-ICP8 antibody HB8180 at a final dilution of 1:5 and rabbit anti-HA (hemagglutinin) antibody Y-11 (Santa Cruz) for detection of HA-tagged Rep proteins at a final dilution of 1:200 (antibodies were diluted in 2% FCS/PBS). Secondary antibody mixture consisted of rhodamine-labeled goat anti-rabbit and fluorescein-labeled goat anti-mouse IgG, both diluted 1:1000. Coverslips were washed three times in PBS and mounted in polyvinyl alcohol (Elvanol) containing 1% DABCO as anti-fading agent. Image acquisition was performed with a Zeiss LSM 510 laser-scanning microscope.

Acknowledgements

We thank David J. Pintel (University of Missouri, USA), Sandra K. Weller (University of Connecticut, USA), Anna Salvetti (Université de Lyon, France), and Nigel Stow (University of Glasgow, U.K.) for plasmids and antibodies, and all members of the Heilbronn lab for advice, helpful discussions and critical reading of the manuscript.

Literature

Alazard-Dany, N., Nicolas, A., Ploquin, A., Strasser, R., Greco, A., Epstein, A. L., Fraefel, C. & Salvetti, A. (2009). Definition of herpes simplex virus type 1 helper activities for adeno- associated virus early replication events. *PLoS Pathog* **5**, e1000340.

Alex, M., Weger, S., Mietzsch, M., Slanina, H., Cathomen, T. & Heilbronn, R. (2012). DNA-binding activity of adeno-associated virus Rep is required for inverted terminal repeat-dependent complex formation with herpes simplex virus ICP8. *J Virol* **86**, 2859-2863.

Atchison, R. W., Casto, B. C. & Hammon, W. M. (1965). Adenovirus-Associated Defective Virus Particles. *Science* **149**, 754-756.

Balasubramanian, N., Bai, P., Buchek, G., Korza, G. & Weller, S. K. (2010). Physical interaction between the herpes simplex virus type 1 exonuclease, UL12, and the DNA double-strand break-sensing MRN complex. *J Virol* **84**, 12504-12514.

Bantel-Schaal, U. & zur Hausen, H. (1984). Characterization of the DNA of a defective human parvovirus isolated from a genital site. *Virology* **134**, 52-63.

Beaton, A., Palumbo, P. & Berns, K. I. (1989). Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein. *J Virol* **63**, 4450-4454.

Biswas, N. & Weller, S. K. (2001). The UL5 and UL52 subunits of the herpes simplex virus type 1 helicase-primase subcomplex exhibit a complex interdependence for DNA binding. *J Biol Chem* **276**, 17610-17619.

Blacklow, N. R., Hoggan, M. D. & Rowe, W. P. (1967). Isolation of adenovirus-associated viruses from man. *Proc Natl Acad Sci U S A* **58**, 1410-1415.

Buller, R. M. L., Janik, J. E., Sebring, E. D. & Rose, J. A. (1981). Herpes simplex virus type 1 and 2 completely help adenovirus-associated virus replication. *J Virol* **10**, 241-247.

Chang, L. S., Shi, Y. & Shenk, T. (1989). Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. *J Virol* **63**, 3479-3488.

Chen, Y., Bai, P., Mackay, S., Korza, G., Carson, J. H., Kuchta, R. D. & Weller, S. K. (2011). Herpes simplex virus type 1 helicase-primase: DNA binding and consequent protein oligomerization and primase activation. *J Virol* **85**, 968-978.

Chiorini, J. A., Afione, S. & Kotin, R. M. (1999a). Adeno-associated virus (AAV) type 5 Rep protein cleaves a unique terminal resolution site compared with other AAV serotypes. *J Virol* **73**, 4293-4298.

Chiorini, J. A., Kim, F., Yang, L. & Kotin, R. M. (1999b). Cloning and characterization of adeno-associated virus type 5. *J Virol* **73**, 1309-1319.

Farris, K. D. & Pintel, D. J. (2010). Adeno-associated virus type 5 utilizes alternative translation initiation to encode a small Rep40-like protein. *J Virol* **84**, 1193-1197.

Fasina, O. & Pintel, D. J. (2013). The adeno-associated virus type 5 small rep proteins expressed via internal translation initiation are functional. *J Virol* **87**, 296-303.

Geoffroy, M. C., Epstein, A. L., Toublanc, E., Moullier, P. & Salvetti, A. (2004). Herpes simplex virus type 1 ICP0 protein mediates activation of adeno-associated virus type 2 rep gene expression from a latent integrated form. *J Virol* **78**, 10977-10986.

Georg-Fries, B., Biederlack, S., Wolf, J. & zur Hausen, H. (1984). Analysis of proteins, helper dependence, and seroepidemiology of a new human parvovirus. *Virology* **134**, 64-71.

Heilbronn, R., Engstler, M., Weger, S., Krahn, A., Schetter, C. & Boshart, M. (2003). ssDNA-dependent colocalization of adeno-associated virus Rep and herpes simplex virus ICP8 in nuclear replication domains. *Nucleic Acids Res* **31**, 6206-6213.

Heilbronn, R., Weller, S. K. & zur Hausen, H. (1990). Herpes simplex virus type 1 mutants for the origin-binding protein induce DNA amplification in the absence of viral replication. *Virology* **179**, 478-481.

Heilbronn, R. & zur Hausen, H. (1989). A subset of herpes simplex replication genes induces DNA amplification within the host cell genome. *J Virol* **63**, 3683-3692.

Hickman, A. B., Ronning, D. R., Kotin, R. M. & Dyda, F. (2002). Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol Cell* **10**, 327-337.

Hickman, A. B., Ronning, D. R., Perez, Z. N., Kotin, R. M. & Dyda, F. (2004). The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. *Mol Cell* **13**, 403-414.

Hoggan, M. D., Blacklow, N. R. & Rowe, W. P. (1966). Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. *Proc Natl Acad Sci U S A* **55**, 1467-1474.

Hüser, D., Gogol-Döring, A., Chen, W. & Heilbronn, R. (2014). 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* **88**, 11253-11263.

Hüser, D., Gogol-Döring, A., Lutter, T., Weger, S., Winter, K., Hammer, E. M., Cathomen, T., Reinert, K. & Heilbronn, R. (2010). Integration preferences of wildtype AAV-2 for consensus rep- binding sites at numerous loci in the human genome. *PLoS Pathog* **6**, e1000985.

Kyöstiö, S. R. M., Owens, R. A., Weitzman, M. D., Antoni, B. A., Chejanovsky, N. & Carter, B. J. (1994). Analysis of adeno-associated virus (AAV) wild-type and mutant Rep proteins for their abilities to negatively regulate AAV p5 and p19 mRNA levels. *J Virol* **68**, 2947-2957.

Livingston, C. M., DeLuca, N. A., Wilkinson, D. E. & Weller, S. K. (2008). Oligomerization of ICP4 and rearrangement of heat shock proteins may be important for herpes simplex virus type 1 prereplicative site formation. *J Virol* **82**, 6324-6336.

Maggin, J. E., James, J. A., Chappie, J. S., Dyda, F. & Hickman, A. B. (2012). The amino acid linker between the endonuclease and helicase domains of adeno-associated virus type 5 Rep plays a critical role in DNA-dependent oligomerization. *J Virol* **86**, 3337-3346.

Malik, A. K., Martinez, R., Muncy, L., Carmichael, E. P. & Weller, S. K. (1992). Genetic analysis of the herpes simplex virus type 1 UL9 gene: isolation of a LacZ insertion mutant and expression in eukaryotic cells. *Virology* **190**, 702-715.

McPherson, R. A., Rosenthal, L. J. & Rose, J. A. (1985). Human cytomegalovirus completely helps adeno-associated virus replication. *Virology* **147**, 217-222.

Muzyczka, N. & Berns, K. I. (2001). Parvoviridae: The viruses and their replication. In *Fields Virology*, pp. 2327-2359. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott.

Nash, K., Chen, W. & Muzyczka, N. (2008). Complete in vitro reconstitution of adeno-associated virus DNA replication requires the minichromosome maintenance complex proteins. *J Virol* **82**, 1458-1464.

Nayak, R. & Pintel, D. J. (2007). Positive and negative effects of adenovirus type 5 helper functions on adeno-associated virus type 5 (AAV5) protein accumulation govern AAV5 virus production. *J Virol* **81**, 2205-2212.

Nicolas, A., Alazard-Dany, N., Biollay, C., Arata, L., Jolinon, N., Kuhn, L., Ferro, M., Weller, S. K., Epstein, A. L., Salvetti, A. & Greco, A. (2010). Identification of rep-associated factors in herpes simplex virus type 1-induced adeno-associated virus type 2 replication compartments. *J Virol* **84**, 8871-8887.

Parks, W. P., Melnick, J. L., Rongey, R. & Mayor, H. D. (1967). Physical assay and growth cycle studies of a defective adeno-satellite virus. *J Virol* **1**, 171-180.

Qiu, J., Nayak, R., Tullis, G. E. & Pintel, D. J. (2002). Characterization of the transcription profile of adeno-associated virus type 5 reveals a number of unique features compared to previously characterized adeno-associated viruses. *J Virol* **76**, 12435-12447.

Reuven, N. B., Staire, A. E., Myers, R. S. & Weller, S. K. (2003). The herpes simplex virus type 1 alkaline nuclease and single-stranded DNA binding protein mediate strand exchange in vitro. *J Virol* **77**, 7425-7433.

Samulski, R. J., Berns, K. I., Tan, M. & Muzyczka, N. (1982). Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. *Proc Natl Acad Sci U S A* **79**, 2077-2081.

Schnepp, B. C., Jensen, R. L., Chen, C. L., Johnson, P. R. & Clark, K. R. (2005). Characterization of adeno-associated virus genomes isolated from human tissues. *J Virol* **79**, 14793-14803.

Schumacher, A. J., Mohni, K. N., Kan, Y., Hendrickson, E. A., Stark, J. M. & Weller, S. K. (2012). The HSV-1 exonuclease, UL12, stimulates recombination by a single strand annealing mechanism. *PLoS Pathog* **8**, e1002862.

Slanina, H., Weger, S., Stow, N. D., Kuhrs, A. & Heilbronn, R. (2006). Role of the herpes simplex virus helicase-primase complex during adeno-associated virus DNA replication. *J Virol* **80**, 5241- 5250.

Srivastava, A., Lusby, E. W. & Berns, K. I. (1983). Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J Virol* **45**, 555-564.

Stow, N. D., Brown, G., Cross, A. M. & Abbotts, A. P. (1998). Identification of residues within the herpes simplex virus type 1 origin-binding protein that contribute to sequence-specific DNA binding. *Virology* **240**, 183-192.

Thomson, B. J., Weindler, F. W., Gray, D., Schwaab, V. & Heilbronn, R. (1994). Human herpesvirus 6 (HHV6) is a helper virus for adeno-associated virus type 2 (AAV2) and the rep gene homologue in HHV6 can mediate AAV-2 DNA replication and regulate gene expression.
Virology **204**, 304-311.

Weindler, F. W. & Heilbronn, R. (1991). A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. *J Virol* **65**, 2476-2483.

Weller, S. K. & Coen, D. M. (2012). Herpes simplex viruses: mechanisms of DNA replication. *Cold Spring Harbor perspectives in biology* **4**, a013011.

Wilkinson, D. E. & Weller, S. K. (2004). Recruitment of cellular recombination and repair proteins to sites of herpes simplex virus type 1 DNA replication is dependent on the composition of viral proteins within prereplicative sites and correlates with the induction of the DNA damage response. *J Virol* **78**, 4783-4796.

Winter, K., von Kietzell, K., Heilbronn, R., Pozzuto, T., Fechner, H. & Weger, S. (2012). Roles of E4orf6 and VA I RNA in adenovirus-mediated stimulation of human parvovirus B19 DNA replication and structural gene expression. *J Virol* **86**, 5099-5109.

Figure legends:

Fig. 1: Comparison of AAV5 and AAV2 replication with Ad-2 or HSV-1 as helper viruses. Plasmids pAAV5 or pAAV2 were transfected into HeLa cells and infected with Ad-2 or HSV-1 at an MOI of 5. At 48 h.p.i. cells were harvested. **(A)** AAV5 or **(B)** AAV2 protein expression was analyzed on Western blots using mAb 303.9 against AAV Rep, or mAb B1 against AAV VP. A polyclonal anti-β-actin antibody detects actin as internal control. Secondary antibodies were detected by ECL, as outlined in the methods. **(C)** Hirt DNA extracts were digested with *Dpn*I and analyzed on Southern blots detected with biotin-labeled *rep* gene-derived probes of similar lengths specific for AAV5 or AAV2. AAV replicative intermediates for monomeric (RF1) or dimeric (RF2) replication forms are indicated by arrows. Input plasmid DNA fragments digested by *Dpn*I are indicated by a vertical bar. Membranes were exposed for 10 min (AAV5) or 2 min (AAV2), respectively. **(D)** AAV5 genomic particle (gp/ml) yields with Ad-2 or HSV-1 as helper virus. AAV gp/ml were analyzed in purified cell extracts by qPCR. AAV gp/ml of three experiments are depicted as mean ± standard deviation.

Fig. 2: AAV5 and AAV2 replication with HSV-1 or HSVΔUL9 or induced by HSV-IE genes. A time course of pAAV5, or pAAV2 transfected cells after infection with wild-type HSV-1 or the replication-deficient strain HSVΔUL9 was performed as described in Fig. 1. **(A, B)** Western blot of pAAV5 transfected cells infected with HSV-1 or HSVΔUL9 as indicated. AAV Rep was detected with mAb 303.9, AAV VP with mAb B1, HSV-1 early proteins UL5 and UL9 with the mAb 152 (anti-UL5) or mAb 13924 (anti-UL9), respectively. Actin detected as outlined in Fig. 1 served as control. **(C, D)** Western blot with pAAV2 transfected cells performed in parallel to (A) and (B). **(E)** AAV5 and AAV2 genomic particle yields with HSV-1wt or HSVΔUL9 were analyzed as outlined in Fig. 1D. **(F)** HeLa cells were cotransfected with pAAV5 and equal amounts of plasmids for the HSV-IE genes, ICP0, ICP4, and ICP22 in the combinations depicted. Cotransfection with pAAV2 and ICP0 (lane 7) served as control. Cells were harvested 48 hours post transfection and analyzed for Rep expression on Western blots with mAb 303.9 as described in Fig. 1.

Fig. 3: AAV5 and AAV2 replication induced by combinations of HSV-1 helper genes. Plasmids pAAV5 or pAAV2 were cotransfected in HeLa cells with combinations of HSV-1 genes ICP0, ICP4, UL5, UL8, UL52, UL29 (ICP8), UL30, UL42, UL12, or UL12exo$^-$, as indicated. AAV plasmid transfected cells infected with HSV-1 (MOI = 5) served as references. **(A, B)** Cells were analyzed on Western blots for AAV Rep, AAV VP and actin expression as outlined in Fig. 1. **(C, D)** AAV DNA replication was analyzed on Southern blots with ^{32}P-labeled *rep* gene-derived probes specific for either AAV5 or AAV2, as outlined in Fig. 1. Asterisks depict low intensity bands of the 4.7 kb AAV5 replication form (RF1). **(E, F)** Production of infectious AAV particles was analyzed after second round infection of HeLa cells in triplicate cultures, infected with cleared, transfected cell extracts as described in the methods. AAV titers were analyzed by qPCR as gp/ml, as described in Fig. 1.

Fig. 4: Colocalization of AAV5 Rep and HSV-1 ICP8 and model of HSV-dependent AAV replication. **(A)** HeLa cells grown on coverslips were cotransfected with pAAV5-Rep-HA and different combinations of HSV-1 genes for ICP0, ICP4, UL5, UL8, UL52, ICP8 (UL29), UL30 and UL42, as indicated. Cells were stained with antibodies against HSV ICP8 and the HA-tag of AAV5 Rep, and were analyzed by confocal microscopy as described in the methods. Displayed images represent cross sections of 0.8 µm. Anti-ICP8 reactivity is displayed in green, while Rep is represented in red. Merged foci appearing in yellow indicate colocalization. To determine the extent of colocalization, the correlation coefficient R of green (ICP8) and red (Rep) fluorescence above a fixed background value was determined for a total of 4 to 10 cells for each construct and displayed as mean ± S.E.M. *NA,* not applicable. **(B)** A model of HSV-induced AAV DNA replication is depicted. The AAV5 genome with essential genetic elements is depicted on top. In the middle, the initiation of Rep78-dependent AAV DNA replication is depicted, proposing a stepwise recruitment of HSV replication proteins, single-strand DNA binding protein (ICP8), helicase-primase complex (UL5/8/52) and two-component HSV DNA polymerase (UL30/42). The model for alkaline exonuclease (UL12) and ICP8-dependent resolution of AAV concatemers in the presence of Rep is depicted below.

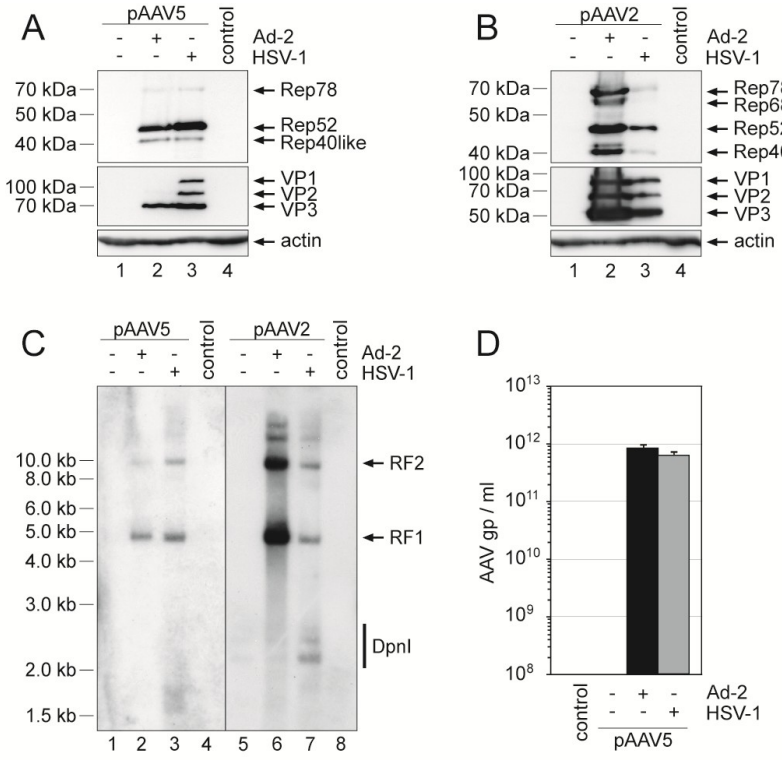

A
pAAV5
control
Ad-2
HSV-1
70 kDa
50 kDa
40 kDa
Rep78
Rep52
Rep40like
100 kDa
70 kDa
VP1
VP2
VP3
actin
1 2 3 4
B
pAAV2
control
Ad-2
HSV-1
70 kDa
50 kDa
40 kDa
Rep78
Rep68
Rep52
Rep40
100 kDa
70 kDa
50 kDa
VP1
VP2
VP3
actin
1 2 3 4
C
pAAV5
pAAV2
control
control
Ad-2
HSV-1
10.0 kb
8.0 kb
6.0 kb
5.0 kb
4.0 kb
3.0 kb
2.0 kb
1.5 kb
RF2
RF1
DpnI
1 2 3 4 5 6 7 8
D
AAV gp / ml
10^{13}
10^{12}
10^{11}
10^{10}
10^{9}
10^{8}
control
pAAV5
Ad-2
HSV-1

Fig. 1

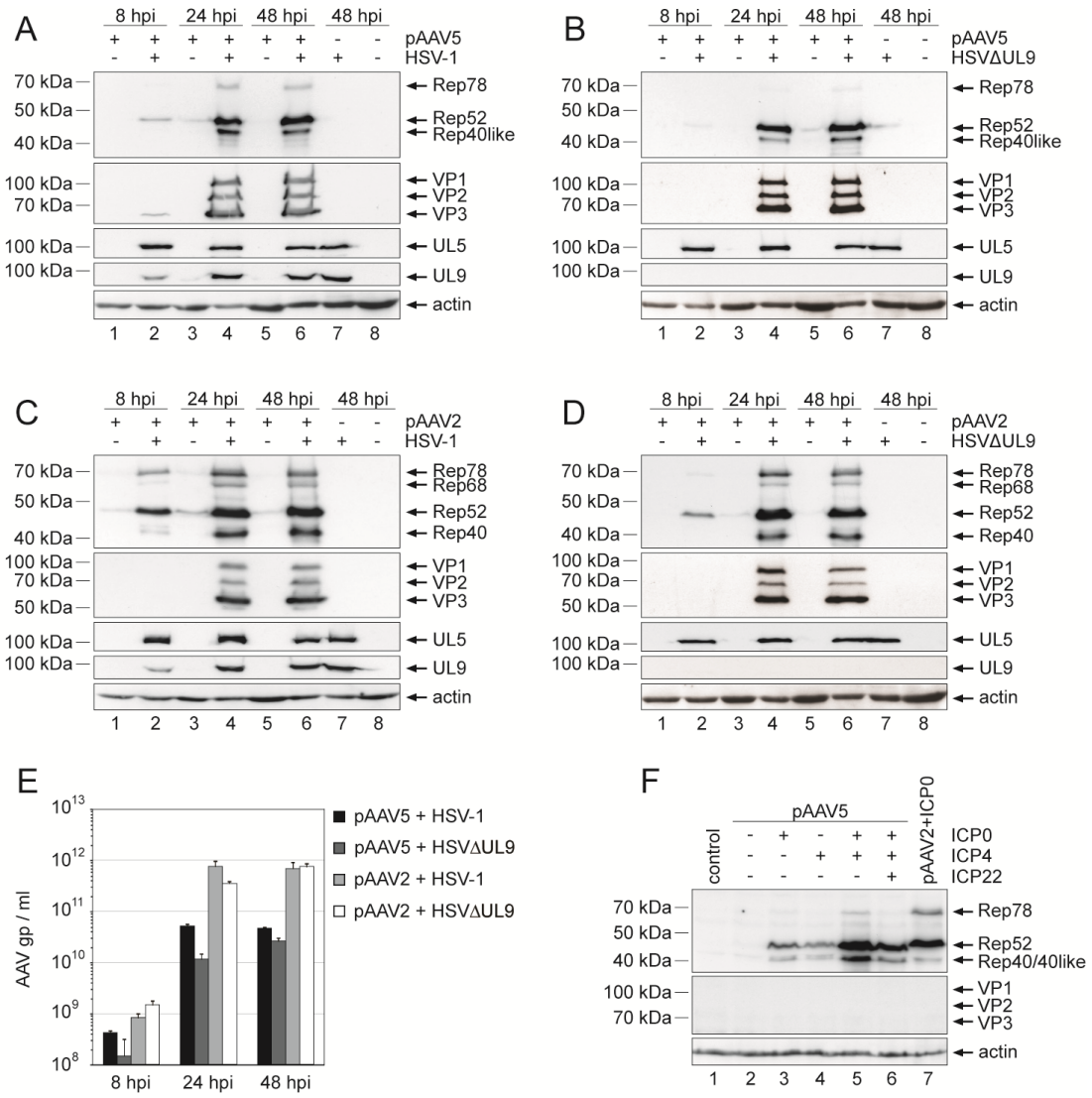

A
8 hpi
24 hpi
48 hpi
48 hpi
pAAV5
HSV-1
70 kDa
50 kDa
40 kDa
Rep78
Rep52
Rep40like
100 kDa
70 kDa
VP1
VP2
VP3
100 kDa
UL5
100 kDa
UL9
actin
1 2 3 4 5 6 7 8
B
pAAV5
HSVΔUL9
C
pAAV2
HSV-1
Rep68
Rep40
D
pAAV2
HSVΔUL9
E
AAV gp / ml
pAAV5 + HSV-1
pAAV5 + HSVΔUL9
pAAV2 + HSV-1
pAAV2 + HSVΔUL9
F
control
pAAV5
pAAV2+ICP0
ICP0
ICP4
ICP22
Rep40/40like

Fig. 2

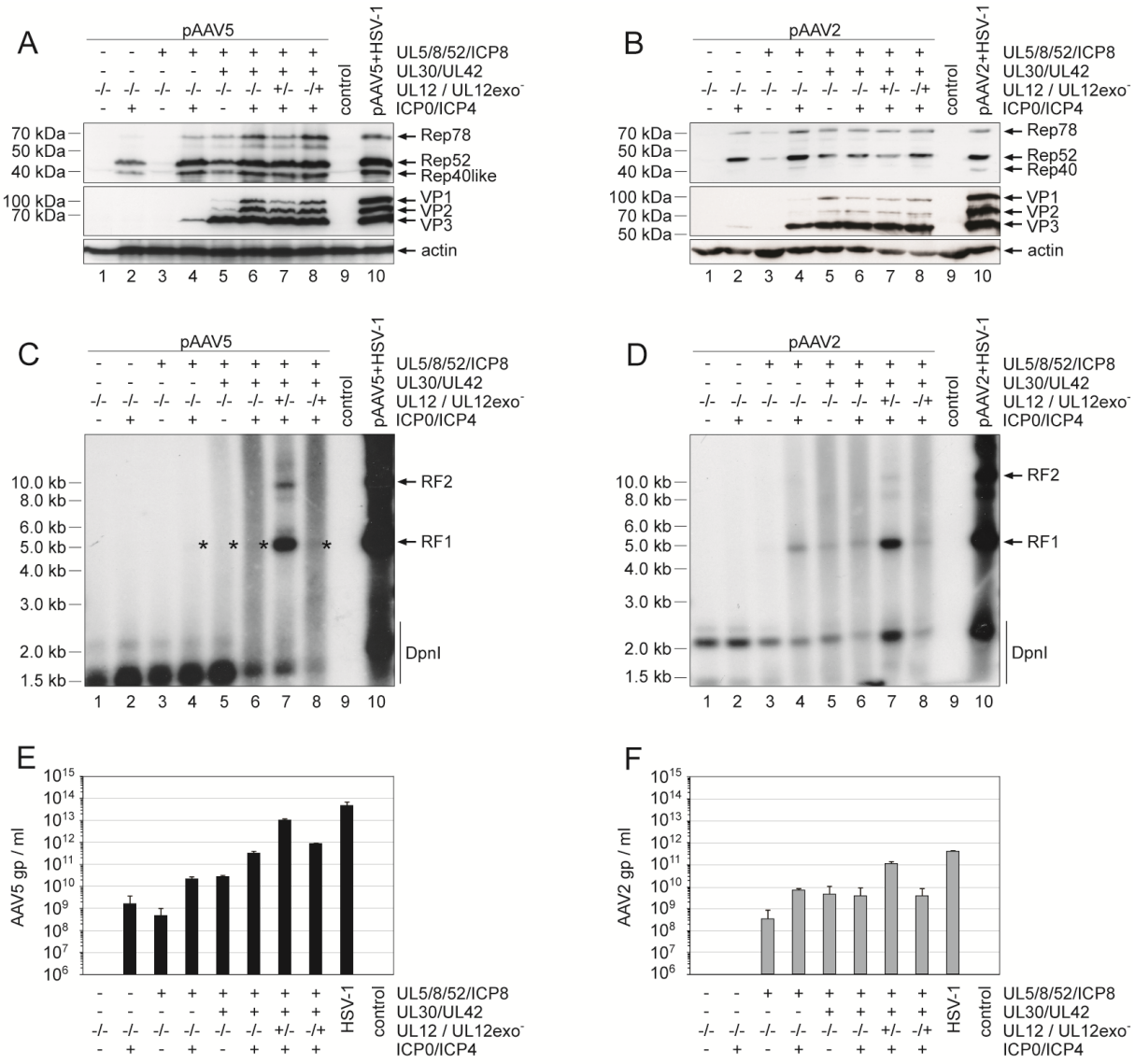

A
pAAV5
pAAV5+HSV-1
control
UL5/8/52/ICP8
UL30/UL42
UL12 / UL12exo-
ICP0/ICP4
70 kDa
50 kDa
40 kDa
100 kDa
70 kDa
Rep78
Rep52
Rep40like
VP1
VP2
VP3
actin
1 2 3 4 5 6 7 8 9 10
B
pAAV2
pAAV2+HSV-1
70 kDa
50 kDa
40 kDa
100 kDa
70 kDa
50 kDa
Rep78
Rep52
Rep40
VP1
VP2
VP3
actin
C
pAAV5
10.0 kb
8.0 kb
6.0 kb
5.0 kb
4.0 kb
3.0 kb
2.0 kb
1.5 kb
RF2
RF1
DpnI
D
pAAV2
E
AAV5 gp / ml
HSV-1
F
AAV2 gp / ml

Fig. 3

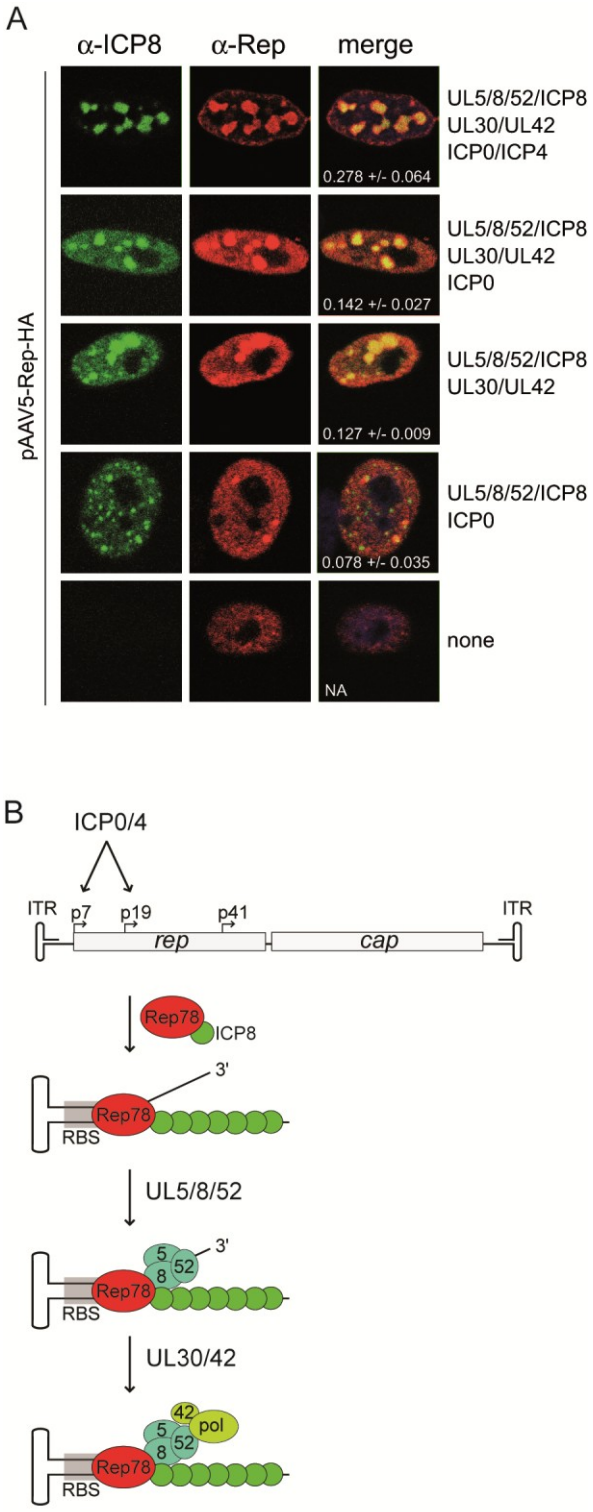

A
α-ICP8
α-Rep
merge
pAAV5-Rep-HA
UL5/8/52/ICP8
UL30/UL42
ICP0/ICP4
0.278 +/- 0.064
UL5/8/52/ICP8
UL30/UL42
ICP0
0.142 +/- 0.027
UL5/8/52/ICP8
UL30/UL42
0.127 +/- 0.009
UL5/8/52/ICP8
ICP0
0.078 +/- 0.035
none
NA
B
ICP0/4
ITR
p7
p19
p41
ITR
rep
cap
Rep78
ICP8
3'
Rep78
RBS
UL5/8/52
5
8
52
3'
Rep78
RBS
UL30/42
42
pol
5
8
52
Rep78
RBS

Fig. 4

3 Discussion

Adeno-associated viruses (AAVs) represent a group of replication-deficient, small single-stranded DNA viruses that exhibit a complex biphasic life cycle. In the absence of a helper virus AAV establish latency and can integrate into the host cell genome. AAV productive infection requires the presence of an unrelated helper virus. Due to the lack of a suitable animal model to study AAV biology *in vivo*, the mechanisms of AAV latency and productive infection in its natural host are largely unknown. In human gene therapy, vectors that are derived from adeno-associated viruses are increasingly applied in the clinic with a focus on the treatment of a variety of monogenetic disorders. Therefore, a more profound understanding of the wild-type AAV replication cycle and its potential interaction with transduced AAV vectors *in vivo* is required. In fact, a recent study in our lab revealed human T-lymphocytes as putative sites of AAV persistence (Khalid, Hüser and Heilbronn, pers. communication). In this study 34% of the healthy and 76% of the immunocompromised blood donors were tested AAV-positive. In addition, due to the high AAV seroprevalence of approximately 80% in the adult population (8) and the increasing application of AAV vectors in the clinic, safety considerations require to understand the biology of latent AAV wild-type infection and the impact on AAV vector transduction.

3.1 AAV Transcriptome Analysis

One major requirement to complete the AAV replication cycle is the ability to transcribe the genetic information of its viral genes. The currently accepted AAV transcription profile mostly dates back to research done in the 1980s (15, 16, 182). At that time, the three AAV2 promoters and the derived transcripts were first characterized. These studies were based on S1 nuclease mapping and primer extension analyses that only allowed for the detection of highly expressed transcripts. Later on, RNase protection assays were used to quantify AAV transcription (22). However, only those transcripts could be detected that were explicitly probed for. All AAV transcripts characterized so far are derived from the AAV positive strand, while the complementary AAV negative strand is considered not to be transcribed. The recent development of next generation sequencing (NGS) technologies allows to perform unbiased gene expression studies. Illumina-based RNA sequencing (RNA-Seq) enables the high-throughput analysis of entire cellular transcriptomes. Using this approach a comprehensive transcriptome analysis of herpes simplex virus (HSV1)

infected cells has revealed new insights into the viral gene expression profile (183). Similar studies with adenovirus has revealed novel splice variants of previously described adenovirus transcripts (184).

The AAV transcription profile has not been analyzed with an unbiased approach. In the present thesis Illumina-based in-depth RNA sequencing was chosen for a comprehensive analysis of the AAV2 transcriptome during latent and lytic infection. This allowed re-evaluating the notion whether AAV is indeed only transcribed on the so-called (+) strand. The overall AAV2 transcription profile largely confirmed the well-established mRNA expression kinetics during AAV2 latency and productive infection. In addition, the AAV negative strand was found to be transcribed as well. These new transcripts initiate opposite of the AAV p5 promoter. Furthermore, novel AAV splice variants were detected by RNA-Seq, the most abundant of which allowed the expression of a new 18 kDa Rep/VP fusion protein with yet unknown function.

3.1.1 AAV Negative Strand Transcription

The here newly detected transcripts on the AAV (-) strand initiate opposite of the p5 promoter on the AAV (+) strand (see Fig. 1A/B, Stutika *et al*. (2016), JVI). Their characterization revealed a heterogeneous group of transcripts of rather short length and varying 5' ends (Fig. 1E/F, Stutika *et al*. (2016), JVI). Since an ORF could not be identified in that region these transcripts initiating from the anti-p5 promoter are proposed to represent non-coding RNAs. A bidirectional pattern of transcription initiation has been described for many mammalian promoters (128), but not for mammalian viruses so far. This mechanism called divergent transcription is described to be associated with the regulation of promoter activity by non-coding promoter-associated RNAs (130). Similarly, transcripts from the AAV p5 promoter running in opposite direction might also contribute to transcription regulation in the AAV replication cycle. It is known that early during latency small amounts of the p5 promoter-derived proteins Rep78/68 are expressed that repress further *rep* gene expression by shutdown of the p5 promoter *in trans* (50). Yet, the maintenance of AAV latency is not fully understood. The newly identified anti-p5 transcripts are candidates for a role in the regulation of the AAV p5 promoter and hence might be involved in the regulatory switch between AAV latency and lytic infection. It is conceivable that the anti-p5 transcripts modulate p5 promoter activity thereby affecting

Rep78/68 protein levels in order to maintain the AAV latent or productive replication state, respectively.

To further study the correlation between anti-p5 transcription and p5 promoter activity levels, mutational analyses of the AAV2 p5 promoter upstream regions have to be performed in order to inhibit the bidirectional transcription. However, it might be difficult to separate the transcription initiation in forward and reverse orientation. Any mutation might affect specific *cis*-regulatory elements important for either orientation of transcription. Previous mutation analyses have shown that the *cis*-regulatory p5 Rep binding element (RBE), located directly upstream of the p5 transcription start site, mediates transcriptional repression of the p5 promoter during latent and lytic infection by Rep78/68 binding (50, 57). Mutating or deleting the p5 RBE (nucleotide position 262 to 277), directly affects transcripts from the anti-p5 promoter since it overlaps with the small RNAs (discussed in 3.2) found in this region. The fact that Rep also inhibits diverse heterologous promoters without *cis*-regulatory RBE elements (185) indicates that the p5 RBE binding by Rep might not be the crucial step for p5 promoter repression. For example, Rep78-mediated inhibition of the HPV18 promoter/enhancer has shown to involve a combination of multiple *cis*-regulatory elements in the absence of an RBE element suggesting that Rep indirectly exerts its effect by interacting with cellular transcription factors (185).

In addition to the TATA box and the RBE element, the p5 promoter exhibits *cis*-regulatory elements that bind the transcription factor YY1 and the major late transcription factor (MLTF) (57). YY1 is described to repress p5 promoter activity in the absence of a helper virus (186). The adenovirus transactivator protein E1A is described to induce p5 promoter activity by interacting with the cellular proteins MLTF and YY1 (54, 186). However, the exact mechanisms on how Rep is involved in the activation of p5 and how transcription initiation from AAV p5 is activated in the presence of other helper viruses is not fully understood. Therefore, further regulatory factors are potentially involved in the p5 promoter control, which might include unknown transcripts generated by divergent transcription from the AAV p5 promoter.

3.1.2 Novel AAV Splice Products

The series of newly identified splice sites potentially allow the generation of yet unknown proteins (see Fig. 4A, Stutika *et al*. (2016), JVI). Due to the lack of available antibodies

only one of these proteins has been confirmed. This 18 kDa protein is hybrid of Rep and VP since it contains the N-terminus of the Rep proteins and the C-terminus of the VP proteins. AAV mutants lacking the necessary splice sites required for the generation of the 18 kDa protein showed minor effects on Rep and VP expression levels (Fig 3A/C, Stutika *et al.* (2016), JVI). This observation indicates a potential role as transcription factor.

The existence of multiple smaller proteins derived by alternative splicing would be analogous to other members of the parvovirus family, such as the parvovirus B19 or the autonomous parvovirus minute virus of mice (MVM), which express similar proteins that seem to be relevant for viral growth in its natural host (187, 188). Since no animal model exists to study AAV biology *in vivo*, it will be challenging to further evaluate the roles of the AAV alternative splice variants *in vivo*.

3.2 AAV Small RNA Analysis

The presence of an anti-p5 promoter close to the end of the left hand side of the AAV genome was a surprising discovery. Since no ORF could be identified in that region, it is an intriguing hypothesis that these anti-p5 transcripts might represent non-coding RNAs instead. In recent years small non-coding RNAs have been identified for many virus families, including herpesviruses and adenoviruses, the major helpers of AAV. Unlike most DNA viruses that establish a latent state, no small non-coding RNAs have been described for the parvovirus group yet.

In order, to search for small RNAs encoded by AAV, libraries from AAV2 infected cells in the presence or absence of Ad2 or HSV1 harvested at two different stages of the AAV replication cycle were generated and subjected to Illumina-based small RNA sequencing (small RNA-Seq). The utilization of this technology allowed the identification of several small AAV2-specific RNAs located in close proximity to viral promoters and within the ITRs (see Fig. 3A, Stutika *et al*. (2016), PONE).

3.2.1 Small AAV2-specific RNAs

In agreement with the results of the total RNA-Seq, a group of small AAV2-specific RNAs near the anti-p5 promoter on the AAV negative strand could be detected by small RNA-Seq with sR-271 being one of the two most abundantly expressed small RNAs in this analysis (Fig. 3, Stutika *et al*. (2016), PONE). Based on its location divergent transcription

from the AAV p5 promoter is the most likely origin of this small RNA. The small RNA-Seq data adds to this hypothesis, since the small AAV2-specific RNAs detected in that region exhibit various 5' ends and a diverse read length distribution of the transcripts, typical features for transcripts derived by divergent transcription (189). While these small AAV2 RNAs are of low abundance at a late stage in latently infected cells, expression levels are highly increased in Ad infected cells. For small promoter-associated RNAs it is described that their expression level correlates with the downstream gene expression level (126). Similarly, the anti-p5 derived small AAV2 RNA sR-271 might affect downstream gene expression of the AAV p5 promoter thereby regulating Rep78/68 protein expression levels during latent and lytic AAV infection, respectively.

The other highly abundant small AAV2-specific RNA sR-108 was located within the ITR next to the Rep binding element (RBE) (nucleotide position 93 to 108) (Fig. 3, Stutika *et al*. (2016), PONE). Even though this small AAV2 RNA possesses a distinct 5' end and a uniform read length of 21 nt, typically seen for miRNAs, argonaute co-immunoprecipitation indicated an alternative processing mechanism different from the canonical miRNA pathway is utilized. For the generation of this small AAV2 RNA two mechanisms are imaginable. (I) Transcription can be initiated within the ITR or (II) from the newly defined anti-p5 promoter on the AAV negative strand assuming that transcription proceeds through the entire AAV-ITR (see Fig. 7A, Stutika *et al*. (2016), PONE). AAV transcription from the ITR has been shown before for AAV5 and certain AAV-derived vectors (29, 33) and is being confirmed here for wild-type AAV2 by the total RNA-Seq analysis. However, the initiator element that drives the AAV5 ITR transcript maps further downstream (nt 142) relatively to the small AAV2 RNA sR-108 (29). In addition, the total RNA-Seq showed that longer transcripts can be generated from the anti-p5 promoter, since this analysis was designed to detect transcripts with a minimal size of 100 nt. Furthermore, RNA transcription across the ITR has been shown just recently (190). Since the abundance of the small AAV2 sR-108 correlated with the anti-p5 derived small AAV2 RNAs on the negative strand during AAV latent and lytic infection, these small RNAs are likely to be generated from the anti-p5 promoter simultaneously.

Small non-coding RNAs are usually processed from longer non-coding RNAs that form a distinct secondary structure. Transcription from the anti-p5 promoter across the ITR would generate a transcript that can form the well-established T-shaped secondary structure of the AAV-ITR, which might facilitate accessibility for cellular or viral processing enzymes.

Since small RNA processing by the common miRNA pathway could be virtually excluded experimentally, a processing mechanism involving Rep might be possible. Rep binding to ITR transcripts has been shown to occur just recently (190).

AAV p5 promoter activity is described to be mediated by the Rep binding elements (RBE) located within the ITR and upstream of the p5 TSS *in cis* during latent and lytic AAV replication (57). In addition, AAV p5 promoter regulation might involve the here detected small AAV2 RNAs sR-108 and sR-271, since these are closely located to the ITR RBE and p5 RBE, respectively, and are differentially expressed during latent and lytic infection.

The small RNA-Seq analysis also detected a hotspot of small AAV2-specific RNAs near the p40 promoter, termed sR-1862 (Fig. 3, Stutika *et al*. (2016), PONE). Due to the very short read length of 18 to 19 nt and its location next to the TSS of p40 a role similar to that of transcription initiation RNAs (tiRNAs) is proposed for this small AAV2 RNA. The tiRNAs are also described to be associated with highly expressed transcripts (132). In accordance with that, sR-1862 can only be detected by small RNA-Seq during AAV productive infection, in which p40 promoter-derived transcripts are highly expressed for the generation of the viral capsid proteins. Besides the described transactivation of p40 by Rep78/68 during AAV productive infection (57), the AAV2 sR-1862 might exert an internal p40 promoter regulatory effect. Furthermore, AAV p40 promoter-derived small RNAs were identified to be associated with the inhibition of adenovirus replication *in cis* during AAV vector production (Fig. 6, Weger *et al*. (2016), JVI). This study adds to the idea that these small AAV2 RNAs likely represent small promoter-associated RNAs that are involved in p40 promoter regulation during AAV productive infection (see Fig. 1A, Stutika *et al*. (2016), PONE). Further evaluation of the role of these small AAV RNAs is therefore of specific interest for AAV vector production systems involving adenovirus vectors.

3.2.2 Impact of AAV on Cellular Transcription Profiles

In order to study viral interactions with the host cell transcriptome, the cellular mRNA and miRNA expression profiles have been analyzed upon the infection of wild-type AAV.

The total RNA-Seq data showed that AAV infection alone did not alter the HeLa cell transcriptome and no differentially expressed genes beyond a maximal twofold range could be detected (see Fig. S1 (A), Stutika *et al.* (2016), JVI, Supplementary Material in proof). This is in line with previous studies on cellular mRNA expression levels upon AAV

infection based on microarray analyses that also showed virtually no effect (191). In contrast, the presence of adenovirus significantly changed cellular transcription levels in the data sets of AAV2/Ad2 co-infected and Ad2 infected cells (Fig. S1 (B and C), Stutika *et al.* (2016), JVI, Supplementary Material in proof). Adenovirus-induced regulation of the host cell transcriptome has been shown before (192).

The results for the miRNA-Seq screen were very similar. The small RNA-Seq showed that the expression level of cellular miRNAs was basically unchanged in AAV infected cells compared to uninfected cells at two different time points (see Fig. 6, Stutika *et al*. (2016), PONE). Only a single cellular miRNA (hsa-mir-3687) was found to be fivefold down-regulated by AAV infection. For this miRNA no target genes have been identified so far. Using a prediction tool designed to find cellular genes that are likely be targeted by a specific miRNA, few putative target sites could be identified here *in silico* for this cellular miRNA. However, the mRNA expression levels of the identified putative target genes received from the total RNA-Seq analysis remained largely unchanged in AAV infected and uninfected cells. Thus, this specific miRNA (hsa-mir-3687) likely targets other than the here predicted cellular mRNAs, e.g. viral mRNAs.

Hence, neither cellular mRNA expression levels nor cellular miRNA expression levels are significantly affected by AAV infection. This is in contrast to the major AAV helpers, adenovirus and HSV, both of which are described to alter the cellular mRNA and miRNA profiles in specific cell types (192-195). This divergence to AAV underlines the long-established idea that AAV wild-type infection is largely apathogenic.

3.3 Advantages and Limitations of Illumina-Based NGS Analysis

To date, next generation sequencing (NGS) approaches are the methods of choice for comprehensive DNA or RNA genome analyses. The primary advantage of Illumina NGS is the high-throughput analysis generating a gigantic volume of sequencing data per sequencing run (196). Therefore the expression profile of entire cellular transcriptomes can be analyzed in parallel. In comparison to conventional methods used for gene expression studies, such as microarray analysis, Illumina-based in-depth sequencing not only detects a comprehensive picture of the expression profile but can also identify novel transcripts, transcripts of low abundance or alternative splice variants.

Similar to other PCR-based methods, amplification biases also exist for various NGS approaches. Therefore RNA-Seq-based methods do not necessarily provide absolute

reliable quantification of the detected transcripts. Illumina genome analysis studies are known to lead to an underrepresentation of AT- and GC-rich regions likely resulting from an amplification bias during library preparation (197). This preselection can only be avoided by choosing an alternative method without prior amplification rounds required for library preparation, e.g. direct RNA sequencing (DRS) as supplied by the Helicos BioSciences platform (198). With this approach polyadenylated RNAs are directly sequenced in a quantitative manner. However, the major challenge of this technique is the generation of high read frequencies as required for a comprehensive RNA analysis. A recent study, profiling the sequencing error types of three different Illumina platforms has revealed substitutions, but not insertions or deletions, as the most prevalent error type (199). Substitutions most frequently occurred after the incorporation of multiple ddGTPs. The here applied HiSeq system showed lowest substitution rates among the tested Illumina platforms in this study. Potential error rates can be corrected by trimming and bioinformatic error correction after the sequencing process. Furthermore, template switching during bridge amplification or sequencing errors within the sample-specific barcode can lead to the assignment of reads to the wrong data set during bioinformatic analysis. This can be avoided by using separate channels of the Illumina flow cell for each of the prepared RNA libraries. Also, major variations of replicative samples might be a first indication of sequencing errors.

On the Illumina platform rather short sequencing reads of 50 to 150 nucleotides in length are generated (174), but according to the supplier also longer reads can be sequenced by using the MiSeq system (up to 300 nucleotides). For this reason, it might be difficult to assign reads to the correct genome position that lie within repetitive regions in the reference genome. To circumvent this problem one can choose the paired-end sequencing approach in which a specific fragment is sequenced from both ends leaving a distance of equal length between each paired read. According to the manufacturer, this allows a more precise gene mapping provided that one read in the pair is unique to the genome. Alternatively, longer read lengths can be generated on the Pacific Biosciences platform (200). While this approach can potentially sequence up to tens of thousands nucleotides of a single molecule, it also has a higher error rate of read accuracy consisting of deletions, insertions and mismatches compared to other NGS platforms and is limited in generating high read frequencies. During the sequencing process lagging or leading strands within a cluster of identical templates might be generated by various NGS platforms. This results

from incomplete extension or addition of multiple nucleotides or probes during the stepwise sequencing cycles, referred to as phasing or pre-phasing, respectively (201). As a consequence, sequence reads that originate from identical templates will be assigned to different genome positions (e.g. -1 or +1) within the reference genome.

Among NGS platforms, Illumina is currently the most widely used system, which stands out for producing the highest data volume within relatively short run times at low costs (196). On the other hand, the Illumina platform provides a lower multiplexing sample capability in comparison to other NGS platforms and generates rather short read length.

3.4 HSV Helper Functions for Productive AAV5 Replication

Various helper viruses have been described to support productive AAV replication. However, unlike the major helpers, adenovirus (Ad) or herpes simplex virus (HSV), the individual helper genes that promote productive AAV replication have been poorly studied for most of these viruses. While the adenovirus helper functions required for efficient AAV productive replication are well characterized for the distantly related serotypes AAV2 and AAV5 (59, 68), helper functions supplied by HSV have been only described for the prototype AAV2 (69, 72). Early studies on AAV seroepidemiology showed a parallel increase of AAV5 and HSV antibody titers (74). In contrast, AAV2 seroconversion has been shown to closely follow that of Ad, which appears to occur earlier during childhood (202) compared to that of HSV. This suggests that HSV, rather than Ad, may represent the natural helper virus for AAV5.

To further study the HSV-AAV5 relationship, a comparative analysis of AAV2 and AAV5 productive replication upon the dependence on the individual HSV helper genes was performed. This analysis revealed that the same HSV helper functions previously identified for AAV2 were required for AAV5 productive replication. In contrast to AAV2, AAV5 was shown to depend on the full set of HSV helper genes for efficient replication that are the six replication genes, UL5/8/29/30/42/52; the immediate-early genes, ICP0 and ICP4; and the alkaline exonuclease, UL12. For AAV2, a minimal set composed of the HSV helicase-primase complex (UL5/8/52) and the single-stranded DNA-binding protein ICP8 (UL29) has been described to support productive replication, which is enhanced by the additional function of HSV DNA polymerase (UL30/42) (69). For AAV5, the sole presence of the four HSV replication genes (UL5/8/29/52) did not promote productive replication, but this can be overcome by the additional helper function for HSV DNA

polymerase (UL30/42). While the HSV replication genes are directly involved in AAV DNA replication, the synergistic effect of the HSV transactivators ICP0 and ICP4 is necessary for efficient AAV5 Rep expression. This is consistent with previous studies on AAV2 (72), with the only difference that HSV ICP22 did not further enhance AAV5 Rep expression. The HSV alkaline exonuclease (UL12) promoted the resolution of AAV5 replicative forms, as described before for AAV2 (73).

Similarly to AAV2 Rep, nuclear co-localization of AAV5 Rep with ICP8 in HSV replication compartments could be detected (70), albeit at reduced levels that were increased in the presence of the full set of HSV helper genes. The relatively reduced frequency of co-localization may reflect less efficient AAV5 plasmid rescue due to the more complex AAV5-ITR (203). This might also explain the generally reduced AAV5 replication levels observed in this analysis, compared to that of AAV2. On the other hand, since it is assumed that HSV represents the natural helper of AAV5, it might be conceivable that AAV5 is actually more dependent on HSV helper functions than AAV2 due to the co-evolving viruses. This assumption is supported by the finding that AAV5 is less dependent on adenovirus than AAV2 (68), for which Ad is assumed to serve as the preferred helper *in vivo*.

3.5 Considerations for AAV Vector Biology

In human gene therapy vectors based on adeno-associated viruses are one of the most successful tools for the treatment of a variety of monogenetic diseases. For their construction the only elements that need to be retained from the wild-type virus are the AAV-ITRs. The AAV Rep and Cap proteins are separated from the ITRs and expressed *in trans* during vector production from a different plasmid, another vector, or by the producer cell line itself.

Since it has been shown that high levels of Rep78/68 negatively influence AAV vector yield, the AAV p5 promoter is mostly replaced by a weak heterologous promoter, such as the MMTV promoter, in order to limit Rep78/68 expression during vector production (91). Consequently, the newly defined anti-p5 promoter opposite of p5 is absent in AAV vectors and therefore plays no role during AAV vector generation. However, the here detected small AAV2 RNA encoded within the ITRs is present in all generated AAV vectors (see Fig. 3A and 7B, Stutika *et al*. (2016), PONE). Since this small RNA is presumed to initiate from the AAV anti-p5 promoter it probably cannot be expressed by the vector due to the

replacement of the viral p5 promoter by a promoter specific for the expression of the transgene. However, choosing a promoter that exhibits bidirectional activity might trigger the expression of that small RNA. In the case that transcription initiation occurs within the ITR, expression of the small ITR-encoded RNA cannot be prevented in any AAV vector. The exact nucleotide position from which the AAV2 ITR transcripts arise has not yet been discovered, but is believed to involve the A/D region of the ITR (see Fig. 1) (33) that would not be suitable of activating the upstream-located small AAV2 RNA sR-108. Still, the putative correlation of the transcriptional activity from the ITR and the small AAV ITR-encoded RNAs has to be further analyzed for AAV-derived vectors.

In addition, studying different helper viruses upon the individual helper genes that promote AAV replication is of high significance, since viral helper factors are required for the generation of AAV-derived vectors irrespective of the vector production system and AAV serotype used. The here identified HSV helper functions for AAV5 replication adds to our knowledge of AAV biology. This is a further prerequisite for advanced generation of AAV vectors, so that only viral helper factors are used that are indispensable for efficient AAV vector production.

3.6 Outlook

Further analyses are required to evaluate the relevance of AAV p5 antisense transcription and the newly detected small AAV2 RNAs for AAV biology. Therefore, it would be interesting to analyze the AAV transcription profile of related serotypes by RNA-Seq and in other cell lines, e.g. in a primary cell line. The expression pattern of specific transcription factors differ between cell lines, which might result in a transcription profile divergent from that seen here for AAV2 in HeLa cells. Furthermore, it would be interesting to know whether other parvoviruses also exhibit transcription from their negative strands. A mechanism resembling divergent transcription from promoters might be a relevant regulatory factor for the switch between latent and lytic infection, especially for the replication cycles of autonomous parvoviruses that do not rely on additional viral helper genes.

4 Summary

Adeno-associated viruses (AAVs) exhibit a complex biphasic life cycle in cell culture, with productive replication dependent upon co-infection with a helper virus, whereas in the absence of a helper virus AAV latency is established. While vectors based on adeno-associated viruses are already widely used in human gene therapy, open questions remain concerning the regulation of AAV productive and latent infection, especially in the *in vivo* situation, for which no suitable animal model exists so far. To fill this gap, a combination of *in vitro* and *in silico* analyses were chosen in the present thesis to perform a comprehensive analysis of the AAV transcriptome based on state of the art, high-throughput total genomic RNA sequencing.

The analysis of the transcription profiles and kinetics of the AAV (+) strand largely confirmed a variety of previous studies performed with classical RNA methods. However, several novel AAV splice variants were identified, the most abundant of which led to the expression of an 18 kDa Rep/VP fusion protein. Genetic analyses with corresponding virus mutants unable to express the newly identified splice variants showed that these have a major impact on early AAV replication steps, highly suggestive of an important role for *in vivo* AAV propagation. Maybe even more interesting, the AAV transcriptome analysis revealed for the first time the expression of transcripts derived from the AAV (-) strand. These transcripts were encoded in the p5 promoter region, which drives the expression of the large AAV2 regulatory proteins Rep78/68. The pattern of p5 sense and antisense transcription was highly reminiscent of a phenomena described in recent years as bidirectional transcription. This adds a further possible mechanistic level to the regulation of p5 driven gene expression, which represents a major switch point between productive and latent AAV infection.

To identify small RNAs with regulatory roles in the AAV life cycle, the first AAV small RNA-Seq analysis was performed in AAV infected cells. A variety of novel AAV2-specific small RNAs located close to or within the AAV-ITRs was identified, which are apparently transcribed from the newly defined anti-p5 promoter. Validation on Northern blots and by argonaute Co-IP indicated alternative small RNA processing, divergent from the canonical miRNA pathway. In addition, small AAV2-specific RNAs in close proximity to viral promoters were suggestive of small promoter-associated RNAs that might act as transcriptional regulators during AAV latent and productive infection. Furthermore, the

cellular miRNA expression profile showed no alteration upon AAV infection reassuring the long-established apathogenic nature of wild-type AAV.

As a further prerequisite for a better understanding of AAV vector biology, the individual HSV helper genes for productive AAV5 replication were analyzed in a comparative study of the AAV2 and AAV5 replication cycle. This analysis revealed that the same combination of HSV1 helper genes identified for AAV2, promote AAV5 productive replication.

5 Zusammenfassung

Adeno-Assoziierte Viren (AAV) besitzen einen komplexen biphasischen *in vitro* Lebenszyklus. Für die produktive Replikation ist AAV auf die Co-Infektion eines Helfervirus angewiesen, während in Abwesenheit eines Helfervirus der latente Zyklus eingeleitet wird. Während von Adeno-Assoziierten Viren abgeleitete Vektoren bereits weitreichend in der humanen Gentherapie eingesetzt werden, bestehen noch offene Fragen bezüglich der Regulation des produktiven und latenten AAV Infektionszyklus, insbesondere zur Regulation *in vivo*, wofür bisher noch kein geeignetes Tiermodel existiert. Um diese Lücke zu schließen, wurde in der vorliegenden Doktorarbeit eine Kombination aus *in vitro* und *in silico* Experimenten angewendet um eine umfassende Analyse des AAV-Transkriptoms durchzuführen, basierend auf einer der aktuellsten RNA-Analysetechniken, der Hochdurchsatz-Sequenzierung (NGS).

Die Analyse des Transkriptionsprofils und der Transkriptionskinetik des AAV (+)-Strangs bestätigte weitgehend frühere AAV Transkriptionsstudien, welche mit klassischen RNA-Methoden durchgeführt wurden. Darüber hinaus wurden aber auch verschiedene neue AAV-Spleißvarianten identifiziert, wobei für die am häufigsten vertretende Spleißvariante die Expression eines neuen 18 kDa Rep/VP Fusionsproteins nachgewiesen werden konnte. Genetische Analysen mit Virusmutanten, die diese neue Spleißvariante nicht exprimierten, zeigten, dass diese einen großen Einfluss auf frühe Schritte im AAV Infektionszyklus ausüben, was auf eine wichtige Rolle für die AAV Propagation *in vivo* hindeutet. Zusätzlich wurden in der AAV Transkriptomanalyse interessanterweise erstmalig Transkripte vom AAV (-)-Strang identifiziert. Diese Transkripte codierten in der p5 Promoterregion, welcher die Expression der großen regulatorischen AAV2 Proteine, Rep78/68 treibt. Das Transkriptionsprofil vom p5 Promoter, in Sense- und Antisense-Richtung, erinnerte an ein in den letzten Jahren beschriebenes Phänomen, bekannt als "bidirektionale Transkription". Dieser Mechanismus könnte eine weitere Ebene für die Regulation der p5 Promoter-Genexpression darstellen, um vom latenten zum produktiven AAV Infektionszyklus umzuschalten.

Für die Identifizierung kleinerer RNAs mit regulatorischer Funktion für den AAV Infektionszyklus wurde hier erstmalig eine Illumina Sequenzanalyse kleiner RNAs (small RNA-Seq) von AAV-infizierten Zellen durchgeführt. Hierbei konnten viele kleine AAV2-spezifische RNAs in unmittelbarer Nähe bzw. innerhalb der AAV-ITRs identifiziert

werden, welche vermutlich am neu definierten anti-p5 Promoter initiieren. Durch die Validierung mittels Northern-Blot Analyse und Argonaute Co-IP konnte der allgemeine miRNA Prozessierungsmechanismus für die Generierung dieser kleinen AAV2-spezifischen RNAs ausgeschlossen werden. Die unmittelbare Nähe einiger neu detektierter kleiner RNAs zu viralen Promotern lässt allerdings vermuten, dass es sich hierbei um kleine Promoter-assoziierte RNAs handelt, die während des latenten und produktiven Infektionszyklus als Transkriptionsregulatoren fungieren. Weiterhin konnte durch die Analyse des zellulären miRNA-Profils gezeigt werden, dass AAV keinen Einfluss auf die Expression zellulärer miRNAs nimmt, was den lang bekannten apathogenen Charakter von AAV unterstreicht.

Weiterhin wurde hier in einer vergleichenden Studie des AAV2 und AAV5 Replikationszyklus die Abhängigkeit der produktiven AAV5 Replikation von den individuellen HSV1 Helfergenen analysiert, als eine Voraussetzung für die Weiterentwicklung der AAV Vektor-Biologie. Dabei konnte gezeigt werden, dass für AAV5 die gleichen HSV Helferfunktionen benötigt werden, wie für AAV2.

6 References

1. **Atchison RW, Casto BC, Hammon WM.** 1965. Adenovirus-Associated Defective Virus Particles. Science **149:**754-756.
2. **Hoggan MD, Blacklow NR, Rowe WP.** 1966. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. Proc Natl Acad Sci U S A **55:**1467-1474.
3. **Buller RM, Janik JE, Sebring ED, Rose JA.** 1981. Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. J Virol **40:**241-247.
4. **McPherson RA, Rosenthal LJ, Rose JA.** 1985. Human cytomegalovirus completely helps adeno-associated virus replication. Virology **147:**217-222.
5. **McLaughlin SK, Collis P, Hermonat PL, Muzyczka N.** 1988. Adeno-associated virus general transduction vectors: analysis of proviral structures. J Virol **62:**1963-1973.
6. **Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, McLaughlin S, Muzyczka N, Rocchi M, Berns KI.** 1990. Site-specific integration by adeno-associated virus. Proc Natl Acad Sci U S A **87:**2211-2215.
7. **Weitzman MD, Linden RM.** 2011. Adeno-associated virus biology. Methods Mol Biol **807:**1-23.
8. **Boutin S, Monteilhet V, Veron P, Leborgne C, Benveniste O, Montus MF, Masurier C.** 2010. 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 **21:**704-712.
9. **Srivastava A, Lusby EW, Berns KI.** 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol **45:**555-564.
10. **Lusby E, Fife KH, Berns KI.** 1980. Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. J Virol **34:**402-409.
11. **Hauswirth WW, Berns KI.** 1977. Origin and termination of adeno-associated virus DNA replication. Virology **78:**488-499.
12. **King JA, Dubielzig R, Grimm D, Kleinschmidt JA.** 2001. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. Embo J **20:**3282-3291.
13. **McCarty DM, Pereira DJ, Zolotukhin I, Zhou X, Ryan JH, Muzyczka N.** 1994. Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J Virol **68:**4988-4997.
14. **Snyder RO, Samulski RJ, Muzyczka N.** 1990. In vitro resolution of covalently joined AAV chromosome ends. Cell **60:**105-113.
15. **Green MR, Roeder RG.** 1980. Definition of a novel promoter for the major adenovirus-associated virus mRNA. Cell **22:**231-242.
16. **Laughlin CA, Westphal H, Carter BJ.** 1979. Spliced adenovirus-associated virus RNA. Proc Natl Acad Sci U S A **76:**5567-5571.
17. **Qiu J, Pintel D.** 2008. Processing of adeno-associated virus RNA. Front Biosci **13:**3101-3115.
18. **Mendelson E, Trempe JP, Carter BJ.** 1986. Identification of the trans-acting Rep proteins of adeno-associated virus by antibodies to a synthetic oligopeptide. J Virol **60:**823-832.
19. **Smith RH, Kotin RM.** 1998. The Rep52 gene product of adeno-associated virus is a DNA helicase with 3'-to-5' polarity. J Virol **72:**4874-4881.
20. **Im D-S, Muzyczka N.** 1990. The AAV origin-binding protein Rep68 is an ATP-dependent site-specific endonuclease with helicase activity. Cell **61:**447-457.

21. **Brister JR, Muzyczka N.** 1999. Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. J Virol **73:**9325-9336.
22. **Mouw MB, Pintel DJ.** 2000. Adeno-associated virus RNAs appear in a temporal order and their splicing is stimulated during coinfection with adenovirus. J Virol **74:**9878-9888.
23. **Becerra SP, Koczot F, Fabisch P, Rose JA.** 1988. Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript. J Virol **62:**2745-2754.
24. **Trempe JP, Carter BJ.** 1988. Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. J Virol **62:**3356-3363.
25. **Girod A, Wobus CE, Zadori Z, Ried M, Leike K, Tijssen P, Kleinschmidt JA, Hallek M.** 2002. The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. J Gen Virol **83:**973-978.
26. **Sonntag F, Schmidt K, Kleinschmidt JA.** 2010. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc Natl Acad Sci U S A **107:**10220-10225.
27. **Samulski RJ, Berns KI, Tan M, Muzyczka N.** 1982. Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. Proc Natl Acad Sci U S A **79:**2077-2081.
28. **Bantel-Schaal U, Delius H, Schmidt R, zur Hausen H.** 1999. Human adeno-associated virus type 5 is only distantly related to other known primate helper-dependent parvoviruses. J Virol **73:**939-947.
29. **Qiu J, Nayak R, Tullis GE, Pintel DJ.** 2002. Characterization of the transcription profile of adeno-associated virus type 5 reveals a number of unique features compared to previously characterized adeno-associated viruses. J Virol **76:**12435-12447.
30. **Farris KD, Pintel DJ.** 2010. Adeno-associated virus type 5 utilizes alternative translation initiation to encode a small Rep40-like protein. J Virol **84:**1193-1197.
31. **Fasina O, Pintel DJ.** 2013. The adeno-associated virus type 5 small rep proteins expressed via internal translation initiation are functional. Journal of virology **87:**296-303.
32. **Flotte TR, Afione SA, Solow R, Drumm ML, Markakis D, Guggino WB, Zeitlin PL, Carter BJ.** 1993. Expression of the cystic fibrosis transmembrane conductance regulator from a novel adeno-associated virus promoter. J Biol Chem **268:**3781-3790.
33. **Haberman RP, McCown TJ, Samulski RJ.** 2000. Novel transcriptional regulatory signals in the adeno-associated virus terminal repeat A/D junction element. J Virol **74:**8732-8739.
34. **Summerford C, Samulski RJ.** 1998. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. J Virol **72:**1438-1445.
35. **Summerford C, Bartlett JS, Samulski RJ.** 1999. AlphaVbeta5 integrin: a co-receptor for adeno-associated virus type 2 infection. Nat Med **5:**78-82.
36. **Akache B, Grimm D, Pandey K, Yant SR, Xu H, Kay MA.** 2006. The 37/67-kilodalton laminin receptor is a receptor for adeno-associated virus serotypes 8, 2, 3, and 9. J Virol **80:**9831-9836.
37. **Qing K, Mah C, Hansen J, Zhou S, Dwarki V, Srivastava A.** 1999. Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. Nat Med **5:**71-77.
38. **Kashiwakura Y, Tamayose K, Iwabuchi K, Hirai Y, Shimada T, Matsumoto K, Nakamura T, Watanabe M, Oshimi K, Daida H.** 2005. Hepatocyte growth factor receptor is a coreceptor for adeno-associated virus type 2 infection. J Virol **79:**609-614.
39. **Kurzeder C, Koppold B, Sauer G, Pabst S, Kreienberg R, Deissler H.** 2007. CD9 promotes adeno-associated virus type 2 infection of mammary carcinoma cells with low cell surface expression of heparan sulphate proteoglycans. Int J Mol Med **19:**325-333.

40. **Kaludov N, Brown KE, Walters RW, Zabner J, Chiorini JA.** 2001. Adeno-associated virus serotype 4 (AAV4) and AAV5 both require sialic acid binding for hemagglutination and efficient transduction but differ in sialic acid linkage specificity. J Virol **75:**6884-6893.
41. **Di Pasquale G, Davidson BL, Stein CS, Martins I, Scudiero D, Monks A, Chiorini JA.** 2003. Identification of PDGFR as a receptor for AAV-5 transduction. Nat Med **9:**1306-1312.
42. **Pillay S, Meyer NL, Puschnik AS, Davulcu O, Diep J, Ishikawa Y, Jae LT, Wosen JE, Nagamine CM, Chapman MS, Carette JE.** 2016. An essential receptor for adeno-associated virus infection. Nature **530:**108-112.
43. **Chiorini JA.** 2016. And one to bind them all. Oral Dis doi:10.1111/odi.12495.
44. **Summerford C, Samulski RJ.** 2016. AAVR: A Multi-Serotype Receptor for AAV. Mol Ther **24:**663-666.
45. **Bartlett JS, Wilcher R, Samulski RJ.** 2000. Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors. J Virol **74:**2777-2785.
46. **Nonnenmacher M, Weber T.** 2011. Adeno-associated virus 2 infection requires endocytosis through the CLIC/GEEC pathway. Cell Host Microbe **10:**563-576.
47. **Stahnke S, Lux K, Uhrig S, Kreppel F, Hosel M, Coutelle O, Ogris M, Hallek M, Buning H.** 2011. Intrinsic phospholipase A2 activity of adeno-associated virus is involved in endosomal escape of incoming particles. Virology **409:**77-83.
48. **Xiao PJ, Samulski RJ.** 2012. Cytoplasmic trafficking, endosomal escape, and perinuclear accumulation of adeno-associated virus type 2 particles are facilitated by microtubule network. J Virol **86:**10462-10473.
49. **Beaton A, Palumbo P, Berns KI.** 1989. Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein. J Virol **63:**4450-4454.
50. **Kyostio SR, Wonderling RS, Owens RA.** 1995. Negative regulation of the adeno-associated virus (AAV) P5 promoter involves both the P5 rep binding site and the consensus ATP-binding motif of the AAV Rep68 protein. J Virol **69:**6787-6796.
51. **Kotin RM, Linden RM, Berns KI.** 1992. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J **11:**5071-5078.
52. **Hüser D, Gogol-Doring A, Chen W, Heilbronn R.** 2014. 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 **88:**11253-11263.
53. **Weitzman MD, Kyöstiö SRM, Kotin RM, Owens RA.** 1994. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. Proc Natl Acad Sci U S A **91:**5808-5812.
54. **Chang LS, Shi Y, Shenk T.** 1989. Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. J Virol **63:**3479-3488.
55. **Geoffroy MC, Epstein AL, Toublanc E, Moullier P, Salvetti A.** 2004. Herpes simplex virus type 1 ICP0 protein mediates activation of adeno-associated virus type 2 rep gene expression from a latent integrated form. J Virol **78:**10977-10986.
56. **Labow MA, Hermonat PL, Berns KI.** 1986. Positive and negative autoregulation of the adeno-associated virus type 2 genome. J Virol **60:**251-258.
57. **Pereira DJ, McCarty DM, Muzyczka N.** 1997. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. J Virol **71:**1079-1088.
58. **Brister JR, Muzyczka N.** 2000. Mechanism of Rep-mediated adeno-associated virus origin nicking. J Virol **74:**7762-7771.
59. **Janik JE, Huston MM, Rose JA.** 1981. Locations of adenovirus genes required for the replication of adenovirus-associated virus. Proc Natl Acad Sci U S A **78:**1925-1929.

60. **Laughlin CA, Jones N, Carter BJ.** 1982. Effect of deletions in adenovirus early region 1 genes upon replication of adeno-associated virus. J Virol **41:**868-876.
61. **Ye C, Qiu J, Pintel DJ.** 2006. Efficient expression of the adeno-associated virus type 5 p41 capsid gene promoter in 293 cells does not require Rep. J Virol **80:**6559-6567.
62. **Stracker TH, Cassell GD, Ward P, Loo YM, van Breukelen B, Carrington-Lawrence SD, Hamatake RK, van der Vliet PC, Weller SK, Melendy T, Weitzman MD.** 2004. The Rep protein of adeno-associated virus type 2 interacts with single-stranded DNA-binding proteins that enhance viral replication. J Virol **78:**441-453.
63. **Ferrari FK, Samulski T, Shenk T, Samulski RJ.** 1996. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J Virol **70:**3227-3234.
64. **Querido E, Blanchette P, Yan Q, Kamura T, Morrison M, Boivin D, Kaelin WG, Conaway RC, Conaway JW, Branton PE.** 2001. Degradation of p53 by adenovirus E4orf6 and E1B55K proteins occurs via a novel mechanism involving a Cullin-containing complex. Genes Dev **15:**3104-3117.
65. **Schwartz RA, Palacios JA, Cassell GD, Adam S, Giacca M, Weitzman MD.** 2007. The Mre11/Rad50/Nbs1 complex limits adeno-associated virus transduction and replication. J Virol **81:**12936-12945.
66. **West MH, Trempe JP, Tratschin JD, Carter BJ.** 1987. Gene expression in adeno-associated virus vectors: the effects of chimeric mRNA structure, helper virus, and adenovirus VA1 RNA. Virology **160:**38-47.
67. **Nayak R, Pintel DJ.** 2007. Adeno-associated viruses can induce phosphorylation of eIF2alpha via PKR activation, which can be overcome by helper adenovirus type 5 virus-associated RNA. J Virol **81:**11908-11916.
68. **Nayak R, Pintel DJ.** 2007. Positive and negative effects of adenovirus type 5 helper functions on adeno-associated virus type 5 (AAV5) protein accumulation govern AAV5 virus production. J Virol **81:**2205-2212.
69. **Weindler FW, Heilbronn R.** 1991. A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. J Virol **65:**2476-2483.
70. **Alex M, Weger S, Mietzsch M, Slanina H, Cathomen T, Heilbronn R.** 2012. DNA-binding activity of adeno-associated virus Rep is required for inverted terminal repeat-dependent complex formation with herpes simplex virus ICP8. J Virol **86:**2859-2863.
71. **Slanina H, Weger S, Stow ND, Kuhrs A, Heilbronn R.** 2006. Role of the herpes simplex virus helicase-primase complex during adeno-associated virus DNA replication. J Virol **80:**5241-5250.
72. **Alazard-Dany N, Nicolas A, Ploquin A, Strasser R, Greco A, Epstein AL, Fraefel C, Salvetti A.** 2009. Definition of herpes simplex virus type 1 helper activities for adeno-associated virus early replication events. PLoS Pathog **5:**e1000340.
73. **Nicolas A, Alazard-Dany N, Biollay C, Arata L, Jolinon N, Kuhn L, Ferro M, Weller SK, Epstein AL, Salvetti A, Greco A.** 2010. Identification of rep-associated factors in herpes simplex virus type 1-induced adeno-associated virus type 2 replication compartments. J Virol **84:**8871-8887.
74. **Georg-Fries B, Biederlack S, Wolf J, zur Hausen H.** 1984. Analysis of proteins, helper dependence, and seroepidemiology of a new human parvovirus. Virology **134:**64-71.
75. **Thomson BJ, Weindler FW, Gray D, Schwaab V, Heilbronn R.** 1994. Human herpesvirus 6 (HHV6) is a helper virus for adeno-associated virus type 2 (AAV2) and the rep gene homologue in HHV6 can mediate AAV-2 DNA replication and regulate gene expression. Virology **204:**304-311.

76. **Shiau AL, Liu PS, Wu CL.** 2005. Novel strategy for generation and titration of recombinant adeno-associated virus vectors. J Virol **79:**193-201.
77. **You H, Liu Y, Prasad CK, Agrawal N, Zhang D, Bandyopadhyay S, Liu H, Kay HH, Mehta JL, Hermonat PL.** 2006. Multiple human papillomavirus genes affect the adeno-associated virus life cycle. Virology **344:**532-540.
78. **Mietzsch M, Grasse S, Zurawski C, Weger S, Bennett A, Agbandje-McKenna M, Muzyczka N, Zolotukhin S, Heilbronn R.** 2014. OneBac: platform for scalable and high-titer production of adeno-associated virus serotype 1-12 vectors for gene therapy. Hum Gene Ther **25:**212-222.
79. **Moore AR, Dong B, Chen L, Xiao W.** 2015. Vaccinia virus as a subhelper for AAV replication and packaging. Mol Ther Methods Clin Dev **2:**15044.
80. **Nathwani AC, Tuddenham EG, Rangarajan S, Rosales C, McIntosh J, Linch DC, Chowdary P, Riddell A, Pie AJ, Harrington C, O'Beirne J, Smith K, Pasi J, Glader B, Rustagi P, Ng CY, Kay MA, Zhou J, Spence Y, Morton CL, Allay J, Coleman J, Sleep S, Cunningham JM, Srivastava D, Basner-Tschakarjan E, Mingozzi F, High KA, Gray JT, Reiss UM, Nienhuis AW, Davidoff AM.** 2011. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N Engl J Med **365:**2357-2365.
81. **Nathwani AC, Reiss UM, Tuddenham EG, Rosales C, Chowdary P, McIntosh J, Della Peruta M, Lheriteau E, Patel N, Raj D, Riddell A, Pie J, Rangarajan S, Bevan D, Recht M, Shen YM, Halka KG, Basner-Tschakarjan E, Mingozzi F, High KA, Allay J, Kay MA, Ng CY, Zhou J, Cancio M, Morton CL, Gray JT, Srivastava D, Nienhuis AW, Davidoff AM.** 2014. Long-term safety and efficacy of factor IX gene therapy in hemophilia B. N Engl J Med **371:**1994-2004.
82. **Kastelein JJ, Ross CJ, Hayden MR.** 2013. From mutation identification to therapy: discovery and origins of the first approved gene therapy in the Western world. Hum Gene Ther **24:**472-478.
83. **Zinn E, Vandenberghe LH.** 2014. Adeno-associated virus: fit to serve. Curr Opin Virol **8:**90-97.
84. **Grimm D, Kay MA.** 2003. From virus evolution to vector revolution: use of naturally occurring serotypes of adeno-associated virus (AAV) as novel vectors for human gene therapy. Curr Gene Ther **3:**281-304.
85. **Smith RH.** 2008. Adeno-associated virus integration: virus versus vector. Gene Ther **15:**817-822.
86. **Penaud-Budloo M, Le Guiner C, Nowrouzi A, Toromanoff A, Cherel Y, Chenuaud P, Schmidt M, von Kalle C, Rolling F, Moullier P, Snyder RO.** 2008. Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. J Virol **82:**7875-7885.
87. **Yan Z, Zhang Y, Duan D, Engelhardt JF.** 2000. Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. Proc Natl Acad Sci U S A **97:**6716-6721.
88. **Lai Y, Yue Y, Liu M, Ghosh A, Engelhardt JF, Chamberlain JS, Duan D.** 2005. Efficient in vivo gene expression by trans-splicing adeno-associated viral vectors. Nat Biotechnol **23:**1435-1439.
89. **Samulski RJ, Chang LS, Shenk T.** 1989. Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. J Virol **63:**3822-3828.
90. **Xiao X, Li J, Samulski RJ.** 1998. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. J Virol **72:**2224-2232.
91. **Grimm D, Kern A, Rittner K, Kleinschmidt JA.** 1998. Novel tools for production and purification of recombinant adenoassociated virus vectors. Hum Gene Ther **9:**2745-2760.
92. **Graham FL, Smiley J, Russell WC, Nairn R.** 1977. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol **36:**59-74.

93. **Thomas DL, Wang L, Niamke J, Liu J, Kang W, Scotti MM, Ye GJ, Veres G, Knop DR.** 2009. Scalable recombinant adeno-associated virus production using recombinant herpes simplex virus type 1 coinfection of suspension-adapted mammalian cells. Hum Gene Ther **20:**861-870.
94. **Urabe M, Ding C, Kotin RM.** 2002. Insect cells as a factory to produce adeno-associated virus type 2 vectors. Hum Gene Ther **13:**1935-1943.
95. **Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakrabortty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, et al.** 2012. Landscape of transcription in human cells. Nature **489:**101-108.
96. **Imamura T, Yamamoto S, Ohgane J, Hattori N, Tanaka S, Shiota K.** 2004. Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem Biophys Res Commun **322:**593-600.
97. **Pang KC, Stephen S, Engstrom PG, Tajul-Arifin K, Chen W, Wahlestedt C, Lenhard B, Hayashizaki Y, Mattick JS.** 2005. RNAdb--a comprehensive mammalian noncoding RNA database. Nucleic Acids Res **33:**D125-130.
98. **Lee RC, Feinbaum RL, Ambros V.** 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75:**843-854.
99. **Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G.** 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature **403:**901-906.
100. **Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degnan B, Muller P, Spring J, Srinivasan A, Fishman M, Finnerty J, Corbo J, Levine M, Leahy P, Davidson E, Ruvkun G.** 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature **408:**86-89.
101. **Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ.** 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res **36:**D154-158.
102. **Friedman RC, Farh KK, Burge CB, Bartel DP.** 2009. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res **19:**92-105.
103. **Cui Y, Su WY, Xing J, Wang YC, Wang P, Chen XY, Shen ZY, Cao H, Lu YY, Fang JY.** 2011. MiR-29a inhibits cell proliferation and induces cell cycle arrest through the downregulation of p42.3 in human gastric cancer. PLoS One **6:**e25872.
104. **Pedersen I, David M.** 2008. MicroRNAs in the immune response. Cytokine **43:**391-394.
105. **Subramanian S, Steer CJ.** 2010. MicroRNAs as gatekeepers of apoptosis. J Cell Physiol **223:**289-298.
106. **Azzouzi I, Moest H, Winkler J, Fauchere JC, Gerber AP, Wollscheid B, Stoffel M, Schmugge M, Speer O.** 2011. MicroRNA-96 directly inhibits gamma-globin expression in human erythropoiesis. PLoS One **6:**e22838.
107. **Lovat F, Valeri N, Croce CM.** 2011. MicroRNAs in the pathogenesis of cancer. Semin Oncol **38:**724-733.
108. **Carthew RW, Sontheimer EJ.** 2009. Origins and Mechanisms of miRNAs and siRNAs. Cell **136:**642-655.
109. **Kim DH, Saetrom P, Snove O, Jr., Rossi JJ.** 2008. MicroRNA-directed transcriptional gene silencing in mammalian cells. Proc Natl Acad Sci U S A **105:**16230-16235.
110. **Younger ST, Corey DR.** 2011. Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. Nucleic Acids Res **39:**5682-5691.

111. **Bartel DP.** 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell **116:**281-297.
112. **Klattenhoff C, Theurkauf W.** 2008. Biogenesis and germline functions of piRNAs. Development **135:**3-9.
113. **Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ.** 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science **322:**1387-1392.
114. **Pillai RS, Chuma S.** 2012. piRNAs and their involvement in male germline development in mice. Dev Growth Differ **54:**78-92.
115. **Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, Dalla-Favera R.** 2013. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proc Natl Acad Sci U S A **110:**1404-1409.
116. **Sobala A, Hutvagner G.** 2013. Small RNAs derived from the 5' end of tRNA can inhibit protein translation in human cells. RNA Biol **10:**553-563.
117. **Lee YS, Shibata Y, Malhotra A, Dutta A.** 2009. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev **23:**2639-2649.
118. **Keam SP, Hutvagner G.** 2015. tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. Life (Basel) **5:**1638-1651.
119. **Falaleeva M, Stamm S.** 2013. Processing of snoRNAs as a new source of regulatory non-coding RNAs: snoRNA fragments form a new class of functional RNAs. Bioessays **35:**46-54.
120. **Kiss T.** 2002. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell **109:**145-148.
121. **Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J.** 2011. Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. Nucleic Acids Res **39:**675-686.
122. **Scott MS, Ono M.** 2011. From snoRNA to miRNA: Dual function regulatory non-coding RNAs. Biochimie **93:**1987-1992.
123. **Politz JC, Hogan EM, Pederson T.** 2009. MicroRNAs with a nucleolar location. RNA **15:**1705-1715.
124. **Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC.** 2007. Mammalian mirtron genes. Mol Cell **28:**328-336.
125. **Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF.** 2009. Evidence for human microRNA-offset RNAs in small RNA sequencing data. Bioinformatics **25:**2298-2301.
126. **Core LJ, Waterfall JJ, Lis JT.** 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science **322:**1845-1848.
127. **Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA.** 2008. Divergent transcription from active promoters. Science **322:**1849-1851.
128. **Wu X, Sharp PA.** 2013. Divergent transcription: a driving force for new gene origination? Cell **155:**990-996.
129. **Affymetrix ETP, Cold Spring Harbor Laboratory ETP.** 2009. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. Nature **457:**1028-1032.
130. **Seila AC, Core LJ, Lis JT, Sharp PA.** 2009. Divergent transcription: a new feature of active promoters. Cell Cycle **8:**2557-2564.
131. **Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J, Daub CO, Hume DA, Suzuki H, Orlando V, Carninci P, Hayashizaki Y, Mattick JS.** 2009. Tiny RNAs associated with transcription start sites in animals. Nat Genet **41:**572-578.

132. **Taft RJ, Kaplan CD, Simons C, Mattick JS.** 2009. Evolution, biogenesis and function of promoter-associated RNAs. Cell Cycle **8:**2332-2338.
133. **Kim VN.** 2005. MicroRNA biogenesis: coordinated cropping and dicing. Nat Rev Mol Cell Biol **6:**376-385.
134. **Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ.** 2004. Processing of primary microRNAs by the Microprocessor complex. Nature **432:**231-235.
135. **Bartel DP.** 2009. MicroRNAs: target recognition and regulatory functions. Cell **136:**215-233.
136. **Ruby JG, Jan CH, Bartel DP.** 2007. Intronic microRNA precursors that bypass Drosha processing. Nature **448:**83-86.
137. **Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R.** 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev **22:**2773-2785.
138. **Shi W, Hendrix D, Levine M, Haley B.** 2009. A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. Nat Struct Mol Biol **16:**183-189.
139. **Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA.** 2010. Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA **16:**673-695.
140. **Taft RJ, Glazov EA, Lassmann T, Hayashizaki Y, Carninci P, Mattick JS.** 2009. Small RNAs derived from snoRNAs. RNA **15:**1233-1240.
141. **Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, Ju J, John B, Enright AJ, Marks D, Sander C, Tuschl T.** 2004. Identification of virus-encoded microRNAs. Science **304:**734-736.
142. **Qureshi A, Thakur N, Monga I, Thakur A, Kumar M.** 2014. VIRmiRNA: a comprehensive resource for experimentally validated viral miRNAs and their targets. Database (Oxford) **2014**.
143. **Kincaid RP, Sullivan CS.** 2012. Virus-encoded microRNAs: an overview and a look to the future. PLoS Pathog **8:**e1003018.
144. **Chen CJ, Cox JE, Kincaid RP, Martinez A, Sullivan CS.** 2013. Divergent MicroRNA targetomes of closely related circulating strains of a polyomavirus. J Virol **87:**11135-11147.
145. **Bellutti F, Kauer M, Kneidinger D, Lion T, Klein R.** 2015. Identification of RISC-associated adenoviral microRNAs, a subset of their direct targets, and global changes in the targetome upon lytic adenovirus 5 infection. J Virol **89:**1608-1627.
146. **Qian K, Pietila T, Ronty M, Michon F, Frilander MJ, Ritari J, Tarkkanen J, Paulin L, Auvinen P, Auvinen E.** 2013. Identification and validation of human papillomavirus encoded microRNAs. PLoS One **8:**e70202.
147. **Hussain M, Taft RJ, Asgari S.** 2008. An insect virus-encoded microRNA regulates viral replication. J Virol **82:**9164-9170.
148. **Wang X, Tang SM, Shen XJ.** 2014. Overview of research on Bombyx mori microRNA. J Insect Sci **14:**133.
149. **Li L, Feng H, Da Q, Jiang H, Chen L, Xie L, Huang Q, Xiong H, Luo F, Kang L, Zeng Y, Hu H, Hou W, Feng Y.** 2016. Expression of HIV-encoded microRNA-TAR and its inhibitory effect on viral replication in human primary macrophages. Arch Virol doi:10.1007/s00705-016-2755-5.
150. **Kincaid RP, Burke JM, Sullivan CS.** 2012. RNA virus microRNA that mimics a B-cell oncomiR. Proc Natl Acad Sci U S A **109:**3077-3082.
151. **Yao Y, Smith LP, Nair V, Watson M.** 2014. An avian retrovirus uses canonical expression and processing mechanisms to generate viral microRNA. J Virol **88:**2-9.

152. **Whisnant AW, Kehl T, Bao Q, Materniak M, Kuzmak J, Lochelt M, Cullen BR.** 2014. Identification of novel, highly expressed retroviral microRNAs in cells infected by bovine foamy virus. J Virol **88:**4679-4686.
153. **Kincaid RP, Chen Y, Cox JE, Rethwilm A, Sullivan CS.** 2014. Noncanonical microRNA (miRNA) biogenesis gives rise to retroviral mimics of lymphoproliferative and immunosuppressive host miRNAs. MBio **5:**e00074.
154. **Hussain M, Torres S, Schnettler E, Funk A, Grundhoff A, Pijlman GP, Khromykh AA, Asgari S.** 2012. West Nile virus encodes a microRNA-like small RNA in the 3' untranslated region which up-regulates GATA4 mRNA and facilitates virus replication in mosquito cells. Nucleic Acids Res **40:**2210-2223.
155. **Hussain M, Asgari S.** 2014. MicroRNA-like viral small RNA from Dengue virus 2 autoregulates its replication in mosquito cells. Proc Natl Acad Sci U S A **111:**2746-2751.
156. **Grundhoff A, Sullivan CS.** 2011. Virus-encoded microRNAs. Virology **411:**325-343.
157. **Murphy E, Vanicek J, Robins H, Shenk T, Levine AJ.** 2008. Suppression of immediate-early viral gene expression by herpesvirus-coded microRNAs: implications for latency. Proc Natl Acad Sci U S A **105:**5453-5458.
158. **Skalsky RL, Samols MA, Plaisance KB, Boss IW, Riva A, Lopez MC, Baker HV, Renne R.** 2007. Kaposi's sarcoma-associated herpesvirus encodes an ortholog of miR-155. J Virol **81:**12836-12845.
159. **Umbach JL, Kramer MF, Jurak I, Karnowski HW, Coen DM, Cullen BR.** 2008. MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. Nature **454:**780-783.
160. **Inman M, Perng GC, Henderson G, Ghiasi H, Nesburn AB, Wechsler SL, Jones C.** 2001. Region of herpes simplex virus type 1 latency-associated transcript sufficient for wild-type spontaneous reactivation promotes cell survival in tissue culture. J Virol **75:**3636-3646.
161. **Cliffe AR, Garber DA, Knipe DM.** 2009. Transcription of the herpes simplex virus latency-associated transcript promotes the formation of facultative heterochromatin on lytic promoters. J Virol **83:**8182-8190.
162. **Jurak I, Kramer MF, Mellor JC, van Lint AL, Roth FP, Knipe DM, Coen DM.** 2010. Numerous conserved and divergent microRNAs expressed by herpes simplex viruses 1 and 2. J Virol **84:**4659-4672.
163. **Munson DJ, Burch AD.** 2012. A novel miRNA produced during lytic HSV-1 infection is important for efficient replication in tissue culture. Arch Virol **157:**1677-1688.
164. **Flores O, Nakayama S, Whisnant AW, Javanbakht H, Cullen BR, Bloom DC.** 2013. Mutational inactivation of herpes simplex virus 1 microRNAs identifies viral mRNA targets and reveals phenotypic effects in culture. J Virol **87:**6589-6603.
165. **Ma Y, Mathews MB.** 1996. Structure, function, and evolution of adenovirus-associated RNA: a phylogenetic approach. J Virol **70:**5083-5099.
166. **O'Malley RP, Mariano TM, Siekierka J, Mathews MB.** 1986. A mechanism for the control of protein synthesis by adenovirus VA RNAI. Cell **44:**391-400.
167. **Andersson MG, Haasnoot PC, Xu N, Berenjian S, Berkhout B, Akusjarvi G.** 2005. Suppression of RNA interference by adenovirus virus-associated RNA. J Virol **79:**9556-9565.
168. **Aparicio O, Carnero E, Abad X, Razquin N, Guruceaga E, Segura V, Fortes P.** 2010. Adenovirus VA RNA-derived miRNAs target cellular genes involved in cell growth, gene expression and DNA repair. Nucleic Acids Res **38:**750-763.
169. **Meshesha MK, Veksler-Lublinsky I, Isakov O, Reichenstein I, Shomron N, Kedem K, Ziv-Ukelson M, Bentwich Z, Avni YS.** 2012. The microRNA Transcriptome of Human Cytomegalovirus (HCMV). Open Virol J **6:**38-48.

170. **Hutzinger R, Feederle R, Mrazek J, Schiefermeier N, Balwierz PJ, Zavolan M, Polacek N, Delecluse HJ, Huttenhofer A.** 2009. Expression and processing of a small nucleolar RNA from the Epstein-Barr virus genome. PLoS Pathog **5:**e1000547.
171. **Diebel KW, Oko LM, Medina EM, Niemeyer BF, Warren CJ, Claypool DJ, Tibbetts SA, Cool CD, Clambey ET, van Dyk LF.** 2015. Gammaherpesvirus small noncoding RNAs are bifunctional elements that regulate infection and contribute to virulence in vivo. MBio **6:**e01670-01614.
172. **Sanger F, Nicklen S, Coulson AR.** 1977. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A **74:**5463-5467.
173. **International Human Genome Sequencing C.** 2004. Finishing the euchromatic sequence of the human genome. Nature **431:**931-945.
174. **Metzker ML.** 2010. Sequencing technologies - the next generation. Nat Rev Genet **11:**31-46.
175. **Nakato R, Shirahige K.** 2016. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. Brief Bioinform doi:10.1093/bib/bbw023.
176. **Sun Z, Cunningham J, Slager S, Kocher JP.** 2015. Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. Epigenomics **7:**813-828.
177. **Medvedev P, Stanciu M, Brudno M.** 2009. Computational methods for discovering structural variation with next-generation sequencing. Nat Methods **6:**S13-20.
178. **Wang Z, Gerstein M, Snyder M.** 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10:**57-63.
179. **Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B.** 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods **5:**621-628.
180. **Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, et al.** 2008. Accurate whole human genome sequencing using reversible terminator chemistry. Nature **456:**53-59.
181. **Weger S, Hammer E, Heilbronn R.** 2014. Differential contribution of adeno-associated virus type 2 Rep protein expression and nucleic acid elements to inhibition of adenoviral replication in cis and in trans. J Virol **88:**14126-14137.
182. **Green MR, Roeder RG.** 1980. Transcripts of the adeno-associated virus genome: mapping of the major RNAs. J Virol **36:**79-92.
183. **Harkness JM, Kader M, DeLuca NA.** 2014. Transcription of the herpes simplex virus 1 genome during productive and quiescent infection of neuronal and nonneuronal cells. J Virol **88:**6847-6861.
184. **Zhao H, Chen M, Pettersson U.** 2014. A new look at adenovirus splicing. Virology **456-457:**329-341.
185. **Hörer M, Weger S, Butz K, Hoppe-Seyler F, Geisen C, Kleinschmidt JA.** 1995. Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters. J Virol **69:**5485-5496.
186. **Shi Y, Seto E, Chang LS, Shenk T.** 1991. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell **67:**377-388.

187. **Ruiz Z, D'Abramo A, Jr., Tattersall P.** 2006. Differential roles for the C-terminal hexapeptide domains of NS2 splice variants during MVM infection of murine cells. Virology **349:**382-395.
188. **Zhi N, Mills IP, Lu J, Wong S, Filippone C, Brown KE.** 2006. Molecular and functional analyses of a human parvovirus B19 infectious clone demonstrates essential roles for NS1, VP1, and the 11-kilodalton protein in virus replication and infectivity. J Virol **80:**5941-5950.
189. **Scruggs BS, Gilchrist DA, Nechaev S, Muse GW, Burkholder A, Fargo DC, Adelman K.** 2015. Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. Mol Cell **58:**1101-1112.
190. **Wang L, Yin Z, Wang Y, Lu Y, Zhang D, Srivastava A, Ling C, Aslanidi GV, Ling C.** 2015. Productive life cycle of adeno-associated virus serotype 2 in the complete absence of a conventional polyadenylation signal. J Gen Virol **96:**2780-2787.
191. **Stilwell JL, Samulski RJ.** 2004. Role of viral vectors and virion shells in cellular gene expression. Mol Ther **9:**337-346.
192. **Zhao H, Granberg F, Elfineh L, Pettersson U, Svensson C.** 2003. Strategic attack on host cell gene expression during adenovirus infection. J Virol **77:**11006-11015.
193. **Zhao H, Chen M, Tellgren-Roth C, Pettersson U.** 2015. Fluctuating expression of microRNAs in adenovirus infected cells. Virology **478:**99-111.
194. **Hu B, Li X, Huo Y, Yu Y, Zhang Q, Chen G, Zhang Y, Fraser NW, Wu D, Zhou J.** 2016. Cellular responses to HSV-1 infection are linked to specific types of alterations in the host transcriptome. Sci Rep **6:**28075.
195. **Hill JM, Zhao Y, Clement C, Neumann DM, Lukiw WJ.** 2009. HSV-1 infection of human brain cells induces miRNA-146a and Alzheimer-type inflammatory signaling. Neuroreport **20:**1500-1505.
196. **Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M.** 2012. Comparison of next-generation sequencing systems. J Biomed Biotechnol **2012:**251364.
197. **Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA.** 2009. Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biol **10:**R32.
198. **Ozsolak F, Milos PM.** 2011. Single-molecule direct RNA sequencing without cDNA synthesis. Wiley Interdiscip Rev RNA **2:**565-570.
199. **Schirmer M, D'Amore R, Ijaz UZ, Hall N, Quince C.** 2016. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. BMC Bioinformatics **17:**125.
200. **Rhoads A, Au KF.** 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics **13:**278-289.
201. **Ledergerber C, Dessimoz C.** 2011. Base-calling for next-generation sequencing platforms. Brief Bioinform **12:**489-497.
202. **Blacklow NR, Hoggan MD, Rowe WP.** 1968. Serologic evidence for human infection with adenovirus-associated viruses. J Natl Cancer Inst **40:**319-327.
203. **Chiorini JA, Kim F, Yang L, Kotin RM.** 1999. Cloning and characterization of adeno-associated virus type 5. J Virol **73:**1309-1319.

7 List of Abbreviations

AAP	assembly activating protein
AAV	adeno-associated virus
Ad	adenovirus
BHK	baby hamster kidney cell line
bp	base pairs
C. elegans	*Caenorhabditis elegans*
cDNA	complementary DNA
ChIP-seq	chromatin immunoprecipitation sequencing
Co-IP	co-immunoprecipitation
ddNTP	dideoxynucleotide
DGCR8	DiGeorge syndrome chromosomal region 8
DNase	deoxyribonuclease
dNTP	deoxynucleotide
ds	double-stranded
dT	dimer turnaround
E	early
EBV	Epstein-Barr virus
Fig.	figure
gp	genomic particle
h.p.i.	hours post infection
HCMV	human cytomegalovirus
HEK	human embryonic kidney cell line
HeLa	human cervical cancer cell line
HHV6	human herpesvirus 6
HIV1	human immunodeficiency virus 1
HPV	human papillomavirus
HSV	herpes simplex virus
ICP	infected cell protein
Inr	initiator element
ITR	inverted terminal repeat
kb	kilobase

kDa	kilodalton
LAT	latency-associated transcript
lncRNA	long non-coding RNA
mAb	monoclonal antibody
mE	monomer extended
Methyl-seq	methylation sequencing
MHV68	murine gammaherpesvirus 68
miRNA	microRNA
miRNP	microRNA ribonucleoprotein complex
mivaRNA	virus-associated RNA derived microRNA
MLTF	major late transcription factor
MMTV	mouse mammary tumor virus
MOI	multiplicity of infection
moR	microRNA-offset RNA
mRNA	messenger RNA
mT	monomer turnaround
MVM	minute virus of mice
ncRNA	non-coding RNA
NGS	next generation sequencing
nt	nucleotide
ORF	open reading frame
pA	polyadenylation site
paRNA	promoter-associated RNA
PCR	polymerase chain reaction
piRNA	piwi-interacting RNA
PKR	protein kinase R
pre-miRNA	precursor microRNA
pri-miRNA	primary microRNA
PrV	pseudorabies virus
rAAV	recombinant adeno-associated virus
RBE / RBS	Rep binding element / Rep binding site
RFU	relative fluorescent unit
RISC	RNA-induced silencing complex

RNAi	RNA interference
RNApol II	RNA polymerase II
RNA-Seq	RNA sequencing
rRNA	ribosomal RNA
RT-PCR	reverse transcriptase PCR
SA	splice acceptor site
SD	splice donor site
sdRNA	sno-derived RNA
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Sf	*Spodoptera frugiperda*
shRNA	short hairpin RNA
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
SRA	Sequence Read Archive (NCBI)
ss	single-stranded
Tab.	table
tiRNA	transcription initiation RNA, tiny RNA
TRBP	(HIV1) TAR RNA-binding protein
tRF	tRNA-derived fragment
tRNA	transfer RNA
trs	terminal resolution site
TSS	transcription start site
UL	unique long
VA RNA	virus-associated RNA
VIRmiRNA	database of viral microRNAs
VV	vaccinia virus
VZV	varicella zoster virus
WT	wild-type
YY1	Ying Yang 1 transcription factor

8 Curriculum Vitae

The resume is not included in the online version for privacy reasons.

9 Professional Expertise

The resume is not included in the online version for privacy reasons.