THE ROLE OF ASSEMBLY-ACTIVATING PROTEIN IN THE CAPSID
ASSEMBLY OF DIFFERENT ADENO-ASSOCIATED VIRUS SEROTYPES

By Lauriel F Earley

A Dissertation

Presented to the Department of Molecular Microbiology and Immunology and the
Oregon Health & Science University School of Medicine in partial fulfillment of
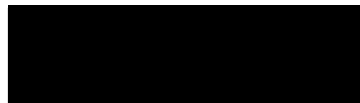the requirements for the degree of

Doctor of Philosophy

April 2016

School of Medicine

Oregon Health & Science University
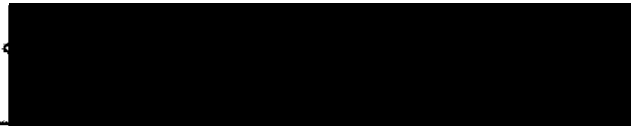
CERTIFICATE OF APPROVAL

This is to certify that the

Ph.D. dissertation of Lauriel Earley
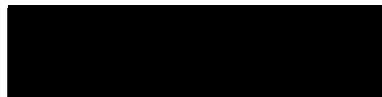
has been approved

Hiroyuki Nakai

Eric Barklis

Georgiana Purdy

Daniel Streblow

Eric Cambronne

TABLE OF CONTENTS

List of Abbreviations

aa	amino acid
Ad	adenovirus
AAP	assembly activating protein
AAV	adeno-associated virus
BR	basic amino acid-rich region
CBB	capsid building block
CDS	condensed DAPI staining
CF	cystic fibrosis
CPV	canine parvovirus
DGC	dense fibrillar centers
DPB	DNA binding protein
FC	fibrillar center
FPV	feline panleukopenia virus
GC	granular component
HEK 293	human embryonic kidney 293 cells
HSPG	heparan sulfate proteoglycan
ITR	inverted terminal repeat
IVIG	intravenous immunoglobulin
NLS	nuclear localization signal
NoLS	nucleolar localization signal
PLA2	phospholipase A2

Pol I	RNA polymerase I
MTLF	major late promoter transcription factor
MVM	minute virus of mice
NLM	nuclear localization motif
NOR	nucleolar organizing region
NS	non-structural
NDF	nucleoli-derived foci
ORF	open reading frame
rAAV	recombinant adeno-associated virus
rDNA	ribosomal DNA
rRNA	ribosomal RNA
RBE	Rep binding element
sc-AAV	self-complementary AAV vectors
snoRNA	small nucleolar RNAs
TRS	terminal resolution site
UBF	upstream binding factor
VA RNA	virus associated RNA
wtAAV	wild-type AAV
YY1	Yin Yang 1

List of Illustrations

Abstract

Adeno-associated virus (AAV) based *in vivo* gene therapy has made great progress in recent years including an approved treatment (Glybera) in Europe and several successful clinical trials in the United States for hemophilia B and Leber congenital amaurosis. One drawback of the vector is the large quantity of virus needed to induce therapeutic levels of transgene expression and thus, production of the vector must be as efficient as possible. Fundamental aspects of AAV's capsid assembly remain poorly characterized, but the recent discovery of assembly-activating protein (AAP) may be the key to finally understanding this crucial part of recombinant AAV (rAAV) vector production and AAV biology.

AAVs are non-enveloped, single-stranded DNA viruses with small icosahedral capsids and simple genomes consisting of two genes, *rep* and *cap*. In 2010, a novel protein was discovered in an overlapping, +1 reading frame within the *cap* gene and termed AAP for its essential role in the capsid assembly of AAV serotype 2 (AAV2). Subsequent work showed that AAP was also required for capsid assembly of AAV8 and AAV9, but its mechanism of action and importance in the life cycle of other AAV serotypes has yet to be explored.

AAP from the canonically studied AAV2 (AAP2) is a nucleolar localizing protein that facilitates the trans-localization of VP proteins from the nucleoplasm to the nucleolus where capsid assembly takes place. In the first part of this dissertation we identified multiple, overlapping nuclear and nucleolar localization signals (NLSs, NoLSs) in the C-terminus of AAP2 that when mutated, resulted in mis-localization of AAP2, decrease in AAV2 titer, and decrease in AAP2 protein

levels. The decrease in AAP2 protein levels was strongly correlated with the decrease in titer, demonstrating that this region is also crucial for sustaining AAP protein levels for high levels of capsid assembly (Chapter 2).

We next compared the amino acids involved in nucleolar localization of AAP2 to the corresponding protein sequence of AAPs from serotypes 1 to 12 and found these regions lacked many of the arginines and lysines seen in AAP2. Characterization of the AAPs 1 to 12 showed that AAP5 and AAP9 were nucleolar excluded. Investigation of capsid assembly in a subset of these serotypes revealed that not all AAVs accumulated capsids in the nucleolus in the same manner as AAV2, that AAP and capsids did not always co-localize, and surprisingly AAV4, 5, and 11 did not require AAP for capsid assembly (Chapter 3). Subsequent work using chimeric AAV11 and AAV12 VP3 proteins identified the region responsible for AAP-independent assembly in AAV11 (Chapter 4).

Taken together, this dissertation work 1) identifies the amino acids in the C-terminus of AAP2 that contribute to nuclear and nucleolar localization, 2) demonstrates that nucleolar-localized capsid assembly found in AAV2 is not generalizable to all AAV serotypes, 3) shows that AAP is not required during capsid assembly for AAV4, 5, and 11, and 4) identifies the region responsible for AAP-independent capsid assembly in AAV11. Current rAAV development uses a variety of AAV serotypes and the implications from this dissertation work will be an important aspect of improving the vector production for rAAVs from non-AAV2 serotypes as well as defining a pathway that may ultimately reveal both the mechanism of AAP and AAV capsid assembly.

Chapter 1: Introduction

Adeno-associated virus

1.1 Classification (*Parvoviridae* family, *Dependoparvovirus* genus)

Adeno-associated viruses belong to the *Dependoparvovirus* genus of the *Parvoviridae* family. Like all *Parvoviridae*, AAVs are non-enveloped, single-stranded DNA viruses with small (~26nm) capsids displaying T=1 icosahedral symmetry and a simple genome consisting of two genes, *rep* and *cap*, flanked by two inverted terminal repeats (ITRs) (**Fig 1.1**). One of the characteristics that distinguishes AAVs from other parvoviruses is the reliance upon a helper virus for productive replication (1). AAV was originally discovered in 1965 as a contaminant in stocks of the eponymous adenovirus (Ad) (1, 2). Further investigation of cell lines (3), adenoviral stocks (2, 4, 5), non-human (6, 7) and human (8, 9) primate tissues led to the discovery of 13 AAV serotypes (AAV1-13) and copious variants, which greatly expanded the AAV family. Recent classification has divided these AAVs into clades A-F, with AAV4 and AAV5 in their own groups (9, 10) (**Fig 1.2**). New viral isolates and AAV genomes are being discovered at a rapid pace in various animal tissues or non-primate adenoviral stocks, including in cow (11), goat (12), bat (13), snake (14), and lizard (15). As the number of sequenced genomes increases, we can expect to learn more about the origin and evolutionary history of AAV.

Intriguingly, based on a genetic analysis, the Anseriform (waterfowl) parvoviruses are also placed within the *Dependoparvovirus* genus even though these do not require a helper virus for productive replication (16-19). This close

relationship may indicate that these parvoviruses serve as an evolutionary link between helper virus-dependent AAVs and a helper-independent ancestor (17, 18).
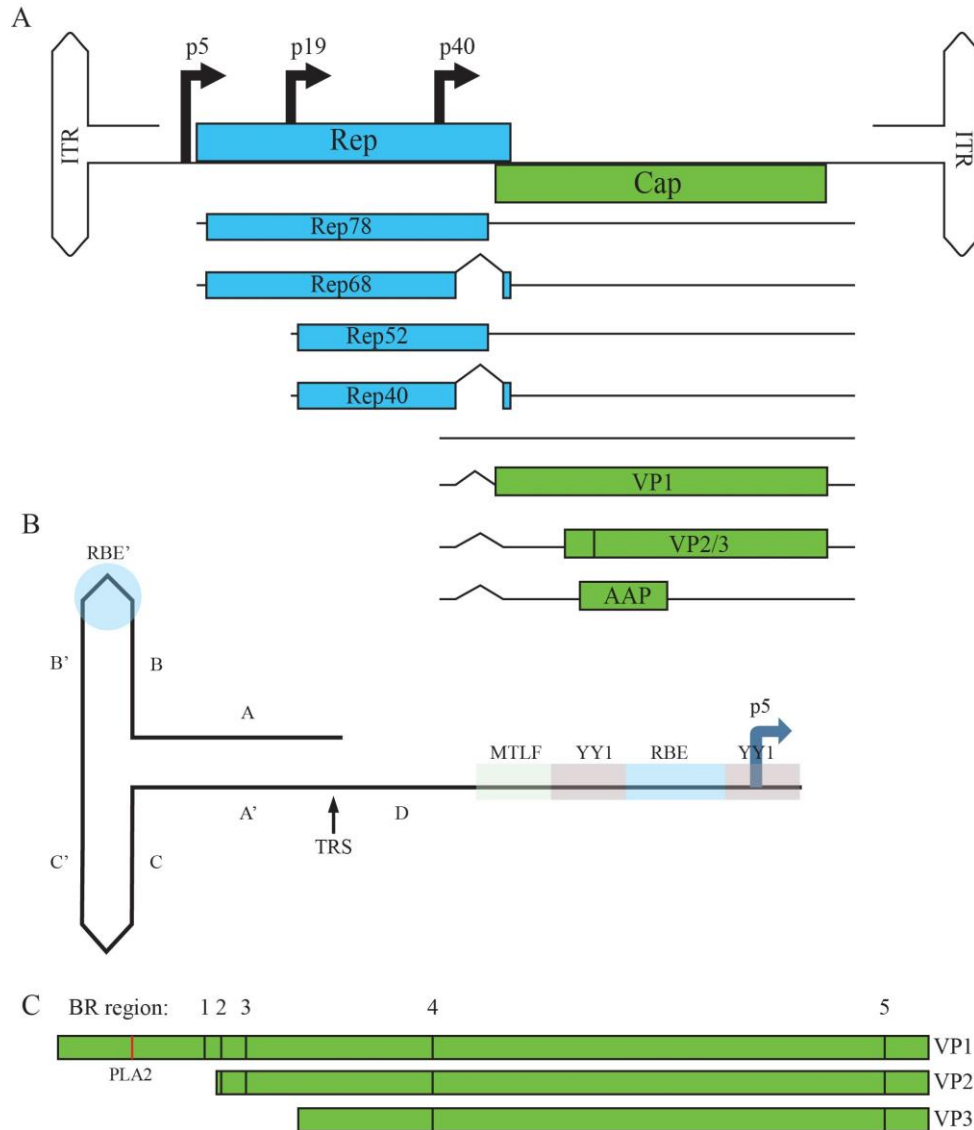
Figure 1.1 Schematic of the wild-type AAV2 genome, ITR, and VP proteins.
 (**A**) The inverted terminal repeats (ITRs) are folded into hairpins at either end of the genome. The p5, p19, and p40 promoters are shown at their map positions. The *rep* gene produces four Rep proteins: Rep78, Rep68, Rep52, and Rep40. The *cap* gene produces the three structural proteins: VP1, VP2, and VP3, and AAP. The 2.3kb mRNA that produces VP1 has an intron at 1906-2201, while the 2.3kb mRNA that produces VP2 and VP3 has an intron at 1906-2228. The mRNA that produces AAP is currently unknown. **AAV genome positions**: **ITR**: 1-145; *rep* **gene**: 321-2252; **Rep78**: 321-2186; **Rep68**: 321-1906, 2228-2252; **Rep52**: 993-2186; **Rep40**: 993-1906, 2228-2252; *cap* **gene**: 2203-4410; **VP1**: 2203-4410; **VP2**: 2614-4410; **VP3**: 2809-4410; **AAP**: 2729-3343; **ITR**: 4534-4679. (**B**) Schematic of the ITR and p5 promoter region. The ITR is designated by complementary regions (A and A', B and B', C and C') and the unpaired D region. It also contains a Rep binding element (RBE) and the terminal resolution site (TRS). The p5 promoter region contains one RBE, two binding sites for the transcription factor YY1, and an MTLF binding site. (**C**) VP1 contains a phospholipase A2 (PLA2) domain and a unique BR in its N-terminus. VP1 and VP2 share two BR regions, and B4 and B5 are found in all the VP proteins.

13
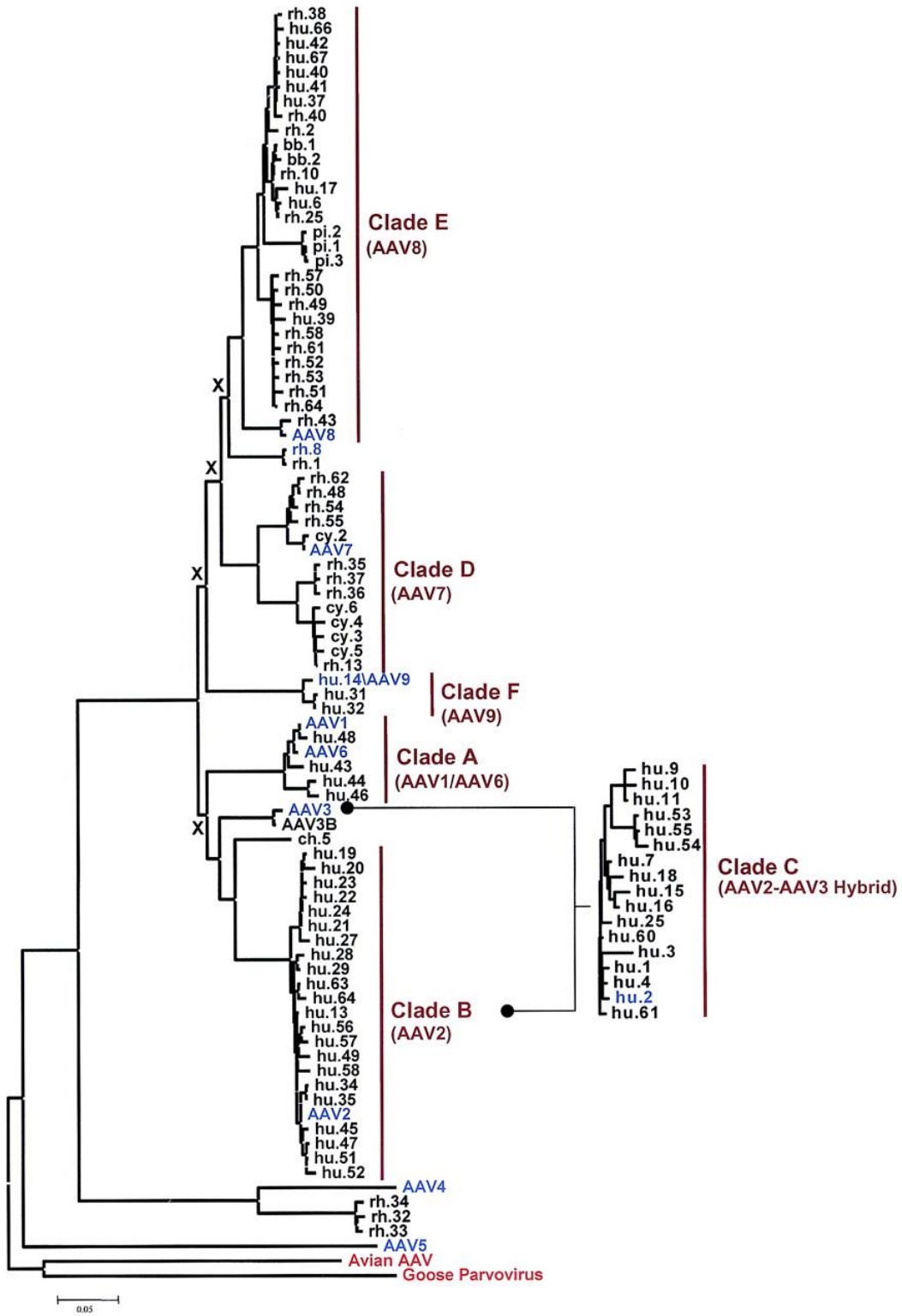
Figure 1.2 Clades of AAV. Reprinted with permission from Gao et al., 2004.

1.2 The AAV genome and protein products

AAV2, the first serotype to be cloned (20) and used to transduced human cells (21), has served as the canonical model for AAV biology. This section is based on the knowledge gained from studying AAV2 unless otherwise specified.

The AAV genome is ~4.7 kb and is composed of two ITRs, two genes (*rep* and *cap*) with three promoters (p5, 19, and p40), and a polyadenylation signal shared by all AAV transcripts, which is cleaved at position 4450 (**Fig 1.1A**) (22). The ITRs contain a Rep binding element (RBE) and are important for genome replication (23-25), genome packaging (26), and insertion of the AAV genome into the host genome (27). Aside from the RBE, they are composed of palindromic repeats designated A, A', B, B', C, C' and an unpaired section, D and a terminal resolution site (TRS) (**Fig 1.1B**). The RBEs are essential for control of both transcription and replication of the genome as well as integration, while the TRS is critical for productive genome replication, as described in section 1.3.

Through mRNA splicing and different start codons, the *rep* gene produces four proteins named for their weight in kilodaltons: Rep78, Rep68, Rep52, and Rep40. The large Reps, Rep78 and Rep68, produced from transcription of the p5 promoter, are multifunctional proteins involved in replication of the AAV genome (28), control of transcription from all three AAV promoters (29-34), and interaction with host proteins (35-40). These proteins have helicase and endonuclease activity and since they are largely functionally redundant, the term Rep68/78 will be used when referring to either of the large rep proteins. The p19 promoter produces the small Rep proteins Rep52 and Rep40, which also have

multiple functions including helicase activity and genome packaging (41). Activation of p40 produces the VP proteins, VP1, VP2, and VP3 all from the same reading frame, with VP1 and VP2 having N-terminal extensions not found in VP3 (**Fig 1.1A and C**). The unique N-terminus of VP1 contains a phospholipase A2 domain (PLA2) and multiple regions rich in basic amino acids that are important for correct intercellular trafficking. In 2010 a second, overlapping +1 reading frame within the *cap* gene was identified (42) and found to encode a previously unknown AAV-specific protein, now termed assembly-activating protein (AAP). AAP is essential for the capsid assembly of AAV serotypes 2, 8, and 9 (43), but its mechanism of action and importance in the life cycle of other AAV serotypes has yet to be explored. AAP from the canonically studied AAV2 is a nucleolar localizing protein that facilitates the trans-localization of VP proteins from the nucleoplasm to the nucleolus (42), presumably through interactions with the C-terminal region common to all the VPs (44).

1.3 Genome replication and integration

Initial insight into AAV genome replication came from the increasing knowledge about its sequence and potential secondary structure (45). Early on, it was clear that the genome was composed of single-stranded DNA, but reports of a double-stranded portion (46) were resolved by a series of experiments showing that the ends of the genome could self-anneal (46, 47). After determining the ITR sequences, Lusby et al. proposed a mechanism of replication initiation in which

the ends of the genome folded into the characteristic AAV hairpins and were thus able to prime replication (23). With our current understanding, the proposed model largely holds true today and is referred to as 'rolling hairpin' replication (45, 48). This mechanism is common to all parvoviruses, but unlike most parvoviruses, that package only minus stranded genomes due the differences in their 3' and 5' ITR sequences, AAV can package either the plus or minus strand of its genome with equal efficiency (45).

Initially, replication begins at the 3' OH provided by the ITR hairpin and continues through the second hairpin. This creates an intermediate form with a closed hairpin on one side and duplexed DNA on the other (**Fig 1.3**). Rep68/78 has ATP-dependent endonuclease activity and is able to nick the replicating genome at the TRS and thus provide a new 3'-OH to initiate transcription through the remaining hairpin (28). This step completes genome replication and results in two complete duplexed AAV genomes. At this point, either end can refold into a hairpin and prime replication. The replication complex forces the other genome out of the duplex via strand displacement. During the nicking process, the Rep68/78 protein is covalently attached to the 5' phosphate at the TRS (49) and this attachment is thought to serve as a packaging signal by allowing interaction between the Rep68/78-AAV genome complex and Rep40/52. Rep40/52 act as motor proteins to package nascent genomes into preformed capsids (41, 50).

The essential proteins and *cis*-acting elements involved in AAV genome replication have been identified and the replication of genomes has been fully reconstituted *in vitro*. Surprisingly, only one Ad gene product, the DNA binding

protein (DBP), is associated with AAV genome replication. This leaves AAV heavily dependent on cellular replication machinery. Currently, both PCNA and polymerase δ have been implicated in the replication process (37), along with replication protein A (RPA), replication factor C (RFC) (39), and the mini chromosome maintenance complex (MCM) (38). These proteins, excluding DBP, make up the minimal elements needed to reconstitute AAV genome replication *in vitro* (38). Polymerase δ is a replicative DNA polymerase with 3′ to 5′ exonuclease proofreading activity and high fidelity. It is commonly found in complex with PCNA, RPA, and RCF. Typically, this complex is responsible for lagging strand synthesis during host genome replication (51). Since AAV only replicates through leading strand synthesis, the function of polymerase δ in replicating AAV genomes is likely tied to its ability to perform gap synthesis after nucleotide removal during DNA repair (52-54), especially since the AAV genome could resemble a damaged DNA structure (gapped DNA, hairpins). Like other parvoviruses, replication of AAV genomes takes place during S-phase when the required host protein complexes are active and thus, AAV is dependent upon the host cell cycle for its own replication. Rep78 acts to maintain the cell in S-phase to prolong AAV replication (36, 55), which also explains, in part, the observations that AAV is anti-oncogenic (56, 57). This leads to a model in which, during S-phase, the ITR hairpin structure and RBE can recruit Rep, PCNA, polymerase δ, RPA, RFC, and MCM to form a replication complex that is able to faithfully copy the AAV genome.
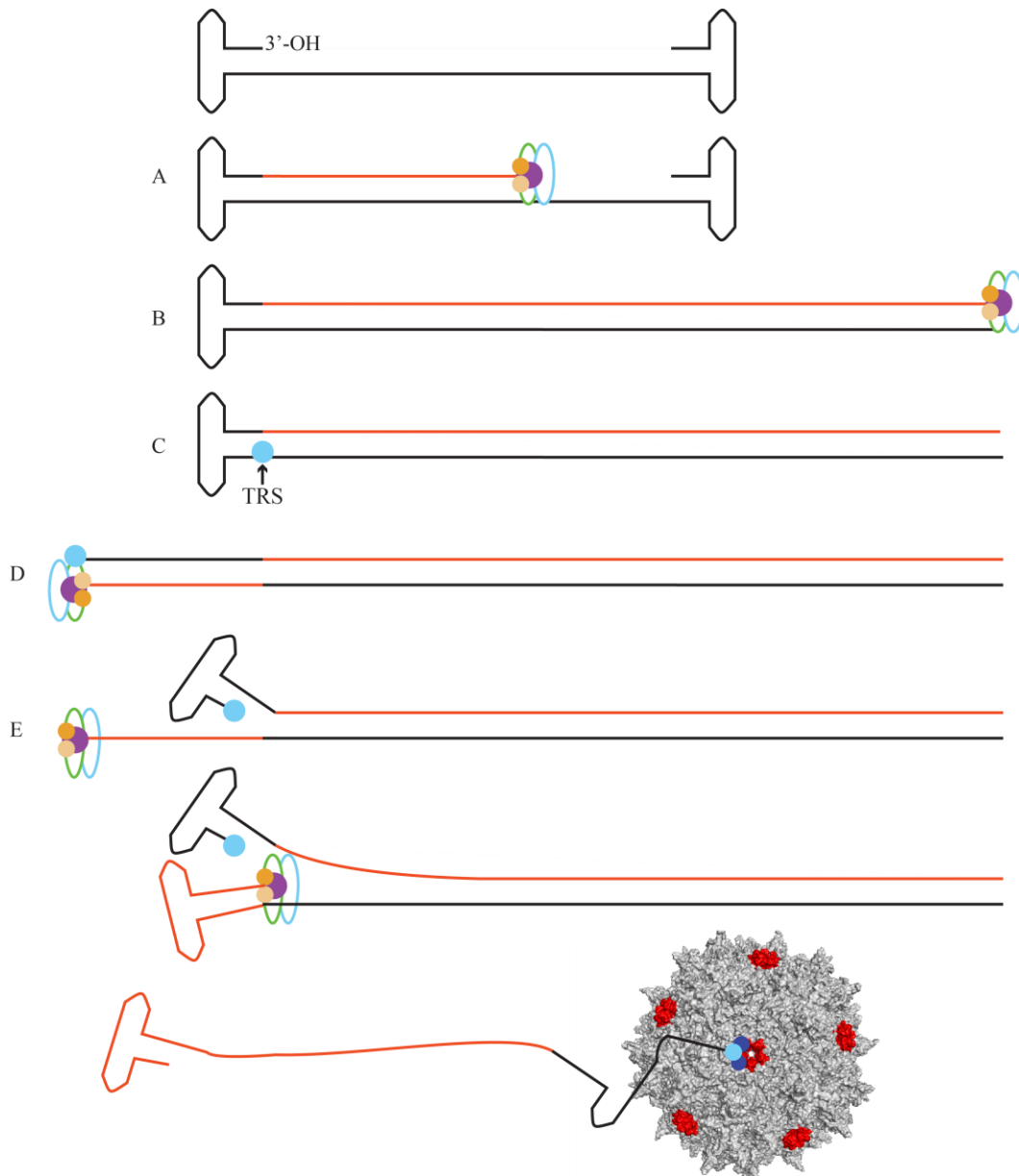
Figure 1.3 Replication of the AAV genome.
(**A**) Replication is primed by the 3'OH provided by the hairpin. Rep68/78 (● and O) binds at the RBE and recruits the replication complex PCNA (O), polymerase δ (●), RPA (●), and RFC (●). (**B**) Replication proceeds, unwinding the duplexed hairpin as needed. (**C**) Rep68/78 nicks the TRS and becomes covalently bound to the 5' phosphate, (**D**) leaving a new 3'OH to initiate replication through the remaining hairpin. (**E**) The hairpins refold and replication can start again. The replication complex releases the new genome through strand displacement. The covalently bound Rep and the ITRs act as signals for Rep52/40 (●) to package the genome into the AAV capsid via the 5-fold pore (marked in red on the AAV2 capsid cartoon). The MCM in is also involved, but not pictured here.

In cell culture systems, integration of the AAV genome (58) predominantly takes place at the AAVS1 site in the human chromosome 19 at position q13.4 (59, 60), although random integrations can take place at a lower frequency (59, 61). This site-specific integration is dependent upon Rep68/78 (62) and the RBE in the AAV genome (63, 64) (**Fig 1.1B**), as well as specific DNA regions at the AAVS1 site that resemble the RBE and TRS in the AAV ITRs (62-64). A proposed mechanism for this integration is that the helicase domain in Rep68/78 mediates non-specific binding and slides along the region until the N-terminal origin-binding domain (OBD) recognizes the RBE-like '-GCTC-' repeats in the AAVS1 locus. The binding of one Rep molecule facilitates the binding of subsequent molecules, eventually forming a heptameric complex which can unwind the integration site (65). Originally, the OBD in the N-terminus of Rep68/78 in complex with the RBE-like '-GCTC-' repeats was determined for AAV5 (66) and recently this interaction was found to be similar for AAV2 (65), indicating it is a likely mechanism for all AAV serotypes. After unwinding the AAVS1 site, the Rep endonuclease domain nicks the homologous TRS site, creating a 3' OH that can act as a primer for replication at AAVS1. This Rep-induced AAVS1 replication can lead to rearrangements at the site and integration of AAV genomes (67, 68). Frequently these integrations are characterized by breakpoints at the ITR or p5 region and if multiple genomes are integrated, they commonly do so in a head-to-tail fashion (69). Weitzman et al. demonstrated that Rep68/78 can bind both the host DNA and the AAV genome (62), placing the AAV genome in close

proximity to the integration site and available for insertion by cellular recombination machinery (69).

Since the site-specific integration of wild-type AAV (wtAAV) is Rep-dependent, rAAV vectors that lack Rep are unable to undergo this type of integration. Numerous studies have been conducted on whether rAAV integration takes place and the general consensus is that it happens very infrequently, at rates reported as <0.5% (70, 71) to <0.0001% (72). The integration sites are described to be either random (72) or in active genes (71), but the sites of these integrations is generally considered safe (71, 73). The more common fate of rAAV DNA is to form concatemeric, double-stranded episomal rings that are maintained long term in non-dividing cells (73-76).

Very few studies have looked at wtAAV integration at an organismal level, but the few that have call into question the specificity of the virus for the AAVS1 site (77, 78). Schnepp et al. were only able to find a single AAV integration event in 175 AAV-positive human tissue samples and the integration took place on chromosome 1 (79). Instead, the majority of detected AAV in their study existed as episomes (79). During a study on endogenous DNA viral elements, Beyli et al. found AAV integration events in various species on chromosomes 1, 2, 3, 8, 10, 13, 19, and the X chromosome (80). Although the homology of these areas to AAVS1 is unspecified, rats and mice both had integration events on separate chromosomes (mice: chromosomes 1 and 3; rats: chromosomes 2, 13, and 19). This area of research is understudied because the current area of focus has been on rAAV integration events, but the potential use of endogenous AAV elements

as a molecular marker in studies on evolution may lead to increased knowledge about the more common sites of AAV integration (80-82).

1.4 mRNA and protein production in the absence of a helper virus

In the absence of a helper virus, AAV enters into a latent phase characterized by reduced transcriptional activity and site-specific integration into the host genome (59, 60). Transcriptional activity is kept to a minimum from interactions between *cis* elements in the AAV genome (29, 31, 32, 40, 83), all four Rep proteins (29, 84), and host factors (40). Both the p5 and p19 promoters are active in HeLa cells, but low levels of Rep production result in auto-suppression of both the p5 and p19 promoters (32). This is mediated through Rep binding to the RBE inside the p5 promoter (85) in combination with the cellular transcription factor YY1 (40), which has two binding sites in p5 at -60 and +1 (**Fig 1.1B**). Interestingly, the suppression of p19 is p5-independent, as demonstrated by a Rep78 mutant (Rep78/K340H) that is able to suppress p5, but not p19 (84). Transcripts from the AAV genome are difficult to detect in the absence of a helper virus, but experiments using plasmid constructs and new deep-sequencing technologies (22) have shed insight into which transcripts might be expressed. The bulk of transcription occurs at the p5 promoter with very little activity from p19 or p40 and the majority of these transcripts from all the promoters are unspliced (22, 29). From the p40 promoter, there are three mRNAs: an unspliced 2.6kb transcript and two spliced transcripts approximately 2.3kb in size that are generated by the same donor site (1906) but two different acceptor sites (2201 and

2228) (86, 87) (**Fig 1.1A**). In the absence of a helper virus, the 2201 and 2228 transcripts are expressed in approximately equal amounts. The larger 2201 transcript produces VP1 and the smaller 2228 transcript produces both VP2 and VP3 (87, 88). The 2.6kb transcript is at low abundance in the cytoplasm and does not appear to produce any protein products (88, 89). While all three transcripts contain intact reading frames for AAP, the transcript that produces the AAP protein has yet to be identified.

1.5 mRNA and protein production in the presence of a helper virus

Multiple viruses can supply helper function for AAV, including several herpes viruses (cytomegalovirus (90), herpes simplex virus 1 (91), and varicella zoster virus (92)), papillomavirus (93), vaccinia virus (94), and adenovirus (1). The helper function of Ad is the most extensively characterized and will be covered in this section, as it also is the most relevant to current vector production methods, which are discussed later in section 4.1 of this chpater.

Adenoviruses are members of the *Adenoviridae* family. Like AAV, Ad has a non-enveloped, icosahedral capsid, but with pseudo T=25 symmetry and characteristic knobbed proteins called fibers that stick out from the capsid at the pentons. The Ad genome is linear, double-stranded DNA, which also contains ITRs that are important for genome replication. Only five Ad genes are required for productive AAV replication. These encode the E1A, E1B, E2A, E4orf6 proteins, and the virus-associated (VA) RNA. The helper function of E1A primarily occurs though gene regulation by relieving p5 suppression via

interactions with YY1 (40) and MLTF (95), consequently displacing Rep and resulting in a dramatic increase of transcription from the p5 promoter. E2A encodes the Ad DPB, which helps to increase AAV titers by several fold, perhaps by promoting efficient splicing of AAV mRNAs (96, 97), however, the specific mechanism for this is still unclear (98-100). The E1B-55k protein can form a complex with the E4orf6 protein which has multifunctional roles in AAV and Ad lifecycles by promoting AAV mRNA export (101, 102) and preventing p53-inducded apoptosis by promoting degradation of p53 via the ubiquitin proteasome pathway (103, 104). Conversely, this complex also leads to degradation of Rep40/52 and AAV capsid proteins (105, 106). Additionally, E4orf6 assists in AAV genome replication by promoting the conversion of single-stranded DNA to double-stranded (107), possibly by indirectly activating DNA repair proteins (108). The role of the VA RNAs is unclear, but its helper function may be more generalized by acting through its ability to circumvent cellular anti-viral defenses (109, 110).

A recent report was able to detect transcription on the minus strand of the p5 promoter upon co-infection with Ad and these divergent transcripts are proposed to play a role in DNA unwinding at p5 (22), but here, only plus strand transcription will be discussed in detail. Once p5 repression is relieved by E1A, there is both an increase in transcription from the p19 and p40 promoters and an increase in the spliced mRNA products, especially for the p40 VP mRNAs. Aside from a general increase in transcripts, there is also a temporal order to the appearance of the mRNAs that matches their positions in the AAV genome, with

p5 expression followed by p19 and then p40 (111). The dramatic shift towards spliced p40 mRNA products and the unconventional ACG start codon for VP2 ultimately leads these structural proteins to be produced in approximately a 1:5:50 stoichiometry (89). This protein ratio is close to the widely reported 1:1:10 stoichiometry in viral particles, but a recent closer examination of individual particles has shown that this ratio can fluctuate from 0-2:8-11:48-51, yet still always equating to 60 VPs in total (112). The protein products from the AAV5 genome are slightly divergent from that of the above described AAV2. The mRNAs from the *rep* gene of AAV5 are inefficiently spliced due to polyadenylation at the internal *rep* intron and thus, AAV5 lacks both Rep68 and Rep40. Instead a Rep40-like protein is produced from an alternative start codon 50 amino acids downstream from the Rep52 start codon (113).

It should be noted that even though AAVs are said to be dependent upon a helper virus while other parvoviruses are autonomous, these distinctions are not completely binary. In fact, the replication of autonomous parvoviruses can be enhanced upon co-infection with a helper virus (114, 115) and AAV can be induced to replicate without a helper virus if the cellular environment is permissive, i.e. during cellular stress. UV radiation, hydroxyurea, and other genotoxic agents have all been used to stimulate AAV replication (116, 117). The reasons for this helper-independent replication are still unclear, but one explanation may be that protein complexes involved in DNA repair that are normally inhibitory to AAV replication are sequestered away from AAV genomes (108). This phenomenon also supports the idea that a major role of helper viruses

in AAV replication is to induce a cellular environment that is permissive to AAV replication.

Capsid assembly

2.1 Icosahedral capsid assembly

All spherical viruses utilize icosahedral symmetry and are composed of 60 structural units or capsid building blocks (CBB), which can be as simple as an individual capsid protein or a complex multi-protein subunit. These CBBs fit together with equivalent (or "quasi-equivalent" for larger viruses) protein interactions between the subunits and are related to each other by 2-fold, 3-fold, and 5-fold axes (**Fig 1.4E**). During the 1950s, the limited crystallographic data available at the time allowed Watson and Crick to deduce that spherical virus capsids must be composed of repeating subunits with cubic symmetry in multiples of 12 (118). As more viruses were studied, the data indicated that all the spherical viruses exhibit icosahedral symmetry. This finding led Capser and Klug to famously lay out their rules for how capsids could be made from repeating copies of CBBs, noting that icosahedral symmetry would be the most economical way possible to build a capsid using limited genetic information. (119). Capsids consisting of only 60 proteins, like AAV, will have 12 pentagons with identical interactions between the proteins, while any capsid composed of more than 60 proteins will result in having non-identical interactions between the proteins due to the requirements for both pentagons and hexagons to be present in the capsid structure. Capser and Klug originally proposed the use of triangulation numbers

(T numbers) to describe the relationships between these hexagons and pentagons. The T numbering system is still in use today (119), although, there are exceptions to these rules: for instance, the capsid of SV40 is considered T=7, but it is composed only of pentamers (120). Assembling an icosahedral capsid, indeed any viral capsid, by necessity must be carefully controlled and orchestrated to produce infectious virions. The timing of capsid protein production is a late-stage event during infection after early genes have been expressed to create a permissive environment for viral replication. Viral genomes are either packaged into preformed capsids or packaged coincident with capsid assembly. Strategies for the capsid assembly process can be simplified into three pathways that are not always mutually exclusive (121): 1) unassisted self assembly in which, under the appropriate conditions, capsid proteins can spontaneously form intact capsids, 2) scaffolding-assisted assembly, such that assembly requires the use of scaffolding proteins that are not found in the intact capsid, and 3) nucleic-acid assisted assembly in which the capsid is formed around the viral genome. Nucleation models of assembly have predicted sigmoidal kinetics (122) during assembly reactions, in which intermediate structures composed of several CBBs are formed during a lag phase, followed by rapid capsid assembly. While experimental data has matched the theoretical models, these intermediate structures have often been difficult to isolate experimentally as these states are transient (121) or not common during *in vivo* assembly (123). Even with those caveats, some intermediate structures during *in vitro* assembly have been identified for bacteriophage MS2 (124), hepatitis B virus (125), and norovirus (125). The forces

driving the association of capsid proteins are thought to be mainly weak protein-protein interactions mediated by hydrophobic contacts (122, 126). Interestingly, the tertiary structure of icosahedral capsid proteins tend to adopt one of two types of folds: the HK97 fold or the β-barrel fold (also known as the jelly roll). The HK97 fold is primarily found bacterial and archaea viruses and the β-barrel roll is predominantly found in eukaryotic viruses. The structure of the β-barrel roll is an antiparallel 8-stranded β-barrel (**Fig 1.4A, B**) (127, 128). Since these folds are found in a wide number of viral families with little amino acid homology, these structures may represent the only way to form capsids with icosahedral symmetry (121).
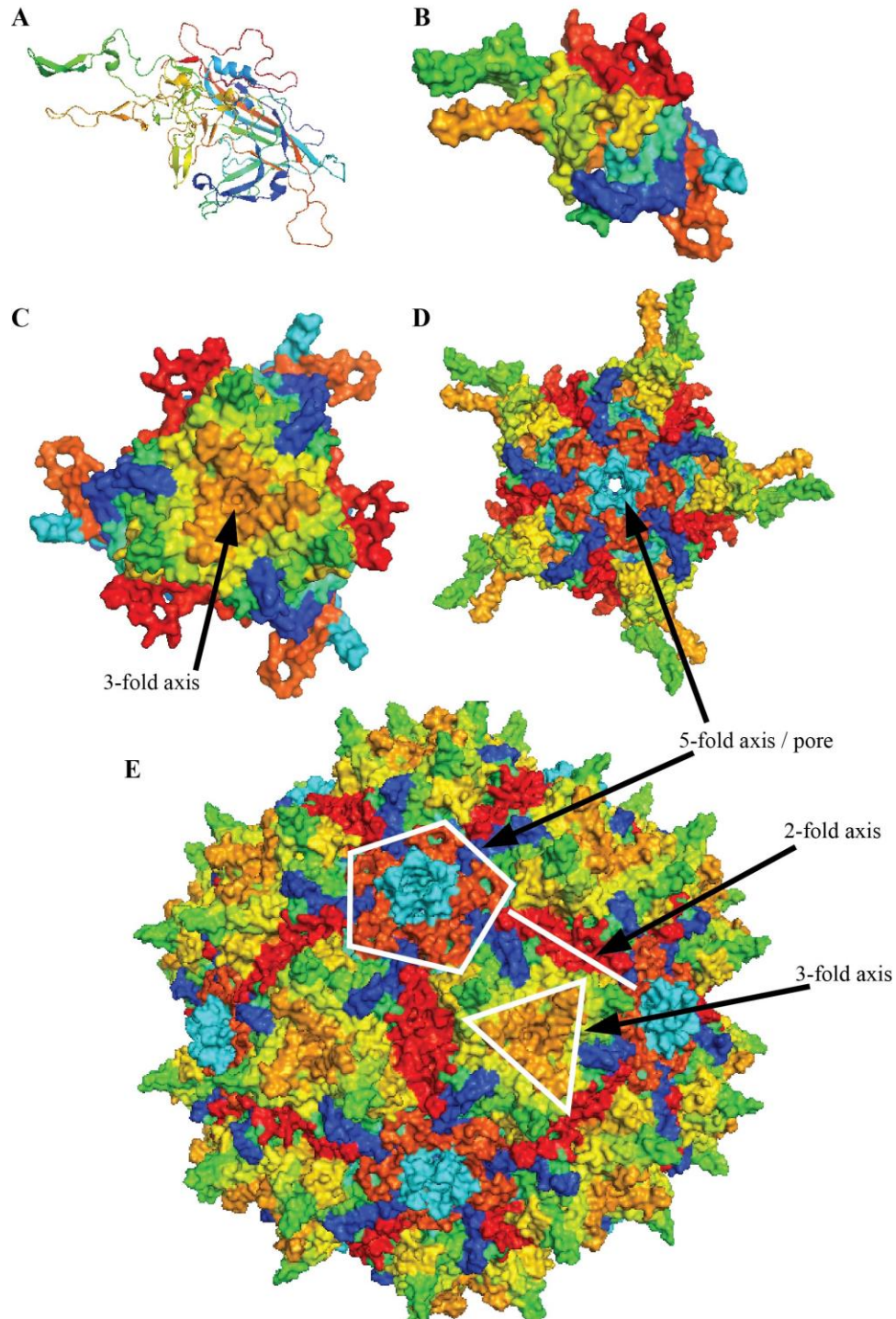
Figure 1.4 PyMOL models of AAV2 VP monomers, oligomers, and full capsid.
(**A**) Ribbon model and (**B**) cartoon of an AAV2 VP monomer displaying β-barrel structure.
(**C**) Cartoon of an AAV2 VP trimer. (**D**) Cartoon of an AAV2 VP pentamer. (**E**) Cartoon of
an AAV2 capsid. 5-fold axis and pore, 2-fold axis, and 3-fold axis are labeled. Images were
created in PyMOL using oligomers from the VIPER database at Scripps from the 1LP3 entry
in the protein database.

(129)

2.2 Parvovirus capsid assembly

Parvovirus capsid assembly takes place in the nucleus of infected cells as empty particles into which viral genomes are packaged (50, 130). Sub-nuclear replication centers or "factories" have been found during infection for numerous different virus types, including parvoviruses (131-133), but these reflect the site of genome packaging, which may not be the site of capsid assembly (134). The capsids are composed of repeat copies of two to four structural proteins termed VPs, which adopt a β-barrel structure. The major capsid protein is capable of forming intact particles without the minor proteins present, but the minor proteins are required for proper intracellular trafficking and infection. These minor proteins contain N-terminal extensions that are usually not observed in structural studies and are thought to be tucked inside the capsid shell. The minor proteins are not capable of forming capsids on their own, likely because the interior of the capsid structure cannot accommodate 60 copies of their extended region (130).

Minute virus of mice (MVM) serves as a model virus for the *Parvoviridae* family and has the most extensively characterized capsid assembly pathway. MVM has two structural proteins, the minor capsid protein VP1 and the major capsid protein VP2. In the cytoplasm, these capsid proteins form two types of trimers, one containing a VP1 protein with two VP2 proteins and the other consisting of three VP2 proteins. These trimers represent the main intermediates in MVM capsid formation (135). VP1 and VP2, which only differ by a short N-

terminal sequence of 142 amino acids, contain a nuclear localization motif in their C-terminal end that transports the proteins into the nucleus (136). VP1 also has a unique N-terminal nuclear localization signal that can transport the VP1 protein and the MVM particle into the nucleus (137). The transport of the trimers into the nucleus is kept in strict stoichiometry by the differing nuclear localization signals and motifs. Under normal conditions, the trimers appear incapable of forming capsids outside of the nucleus, indicating that host proteins or a nuclear environment is required to change these intermediates from an assembly incompetent state to a competent state. The balance of nuclear translocation of the trimers is critical for correct capsid assembly and if altered, results in protein aggregates (135). Hence, nuclear transport of CBBs is hypothesized to be a mechanism parvoviruses utilize to ensure capsid assembly occurs in the correct location and with the correct ratio of CBBs. In conjunction with this, cell cycle also controls nuclear import of the trimers, which can only happen during S-phase (138), demonstrating again the reliance of parvoviruses on S-phase for replication.

2.3 AAV capsid assembly

Even though AAV is also considered a model parvovirus, much less is known about its assembly pathway and no intermediate structures have been identified. AAV VPs first migrate from the cytoplasm into the nucleus and then further on to the nucleolus where capsid assembly is thought to take place. Like MVM, AAV VPs contain regions rich in basic amino acids (BRs) (**Fig 1.1C**), but the evidence for their role in nuclear transport is conflicting. Mutation of BR1 ([122]-KKR-[124])

has little effect on transport of VP1 protein (44) or nuclear entry of virions (139).

BR2 (^{140}-PGKKRP-^{145}) and BR3 (^{168}-RKR-^{170}) have a role in nuclear transport of both VP1 and VP2 since deletion (140) or mutation (141) of BR3 results in partial retention of the VPs in the cytoplasm. Even with those mutations, a substantial amount of VP1 and VP2 can still enter the nucleus and the significance of this cytoplasmic retention is questioned since the resulting VP ratios in assembled particles is the same (142) or very similar (141) to wild type. These regions may be more important for nuclear entry of incoming virions during infection through interactions with importin-β (139, 143, 144). BR4 (^{307}-RPKRLN-^{312}) and BR5 (^{688}-KENSKR-^{693}) do not appear to play a role in nuclear localization of the VP proteins (141) and there are conflicting reports about whether BR4 is required for capsid formation (141, 142). Mutation or deletion of an RGN motif (aa position 584-587) also inhibits VP nuclear localization, but it is unclear if this is due to direct disruption of a NLS or cytoplasmic retention of misfolded proteins (145). The mechanism of nuclear transport and capsid assembly of AAV VPs is complicated by the overlapping VP and AAP proteins, the possible existence of conformational NLMs like those seen in MVM (137), oligomerization-assisted assembly in which NLSs in VP1 or VP2 could assist in transporting VP3 into the nucleus (140), or effects of cell-cycle regulation on nuclear import (138). Further work is needed in this area to tease out the method of VP nuclear import.

There are conflicting reports about capsid assembly based solely on efficient nuclear transport of the VP proteins. Hoque et al. added the NLS from SV40 to the N-terminus of VP3 and were able to observe capsid formation, but Sonntag et

al. were not able to achieve capsid assembly when VP3 was tagged with the same NLS or an NoLS from the HIV Rev protein (42, 140). This distinction may be due to the different methods used for identifying capsids (EM vs immunofluorescence), but the generally accepted model is that nuclear localization of the VP proteins is insufficient to induce capsid formation. Even *in vitro*, the VPs are incapable of self-assembly and are only marginally stimulated to do so when HeLa cell cytoplasmic lysate is added (146). To date, no one has successfully created AAV capsids *in vitro* in large quantities, but *in vitro* assembly of B19 has been achieved (147).

Beyond nuclear transport, nucleolar targeting of VPs is also likely important to the capsid assembly of AAV. Nucleolar involvement in the AAV life cycle is based on several observations: 1) the nucleolus serves as the initial site of capsid assembly (148), 2) several nucleolar proteins co-localize with AAV capsids (134, 148-150), 3) AAP is a nucleolar localizing protein that facilities transport of VPs to the nucleolus (42) (**Fig 2.8A, B; Fig 1.5**), and 4) the nucleolus is where AAV capsids localize to at early stages of infection (151). Bevington et al. hypothesized that initial capsid assembly takes place in the nucleolus with the aid of nucleolar proteins and, once assembled, the capsids would be moved out of the nucleolus and into replication centers (152) in a rep-dependent manner (148) for genome packaging (134).
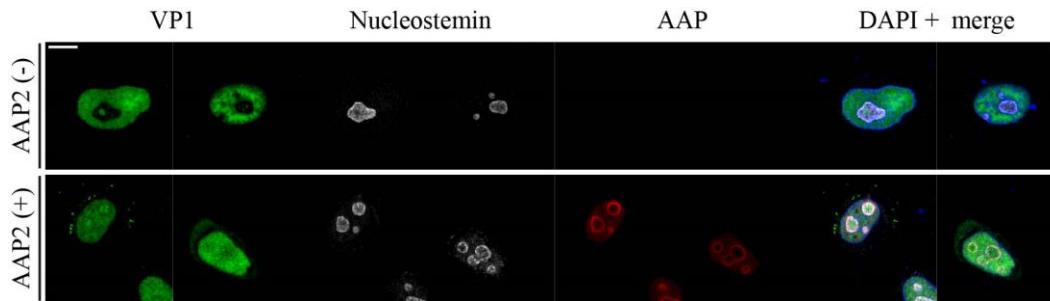
Figure 1.5 VP1 localization without or with AAP2
The VP1 reading frame from pAAV2-ΔAAP2 was cloned into the pCMV$_1$ plasmid to create pCMV$_1$-VP1ΔAAP2. HeLa cells were transfected with pCMV$_1$-VP1ΔAAP2 without or with pCMV$_3$-FLAGcmAAP2. The cells were fixed at 47 hours post-transfection and immunostained with anti-VP1 (green), anti- nucleostemin (white), and anti-FLAG (red) antibodies. Scale bar, 10 μm. Is pAAV2-ΔAAP2 a plasmid with 7 stop codons introduced into the AAV ORF that do not effect the VP3 ORF.

As mentioned previously, stable capsid intermediates have not been isolated, but some work towards this goal has been done both in the context of co-infection with AAV2 and Ad or with AAV VP expression alone. During infection, the VPs exist predominately in monomeric and oligomeric forms in the cytoplasm and upon entering the nucleus, are able to form capsids (153). VP3 can form large, non-capsid oligomers in the nucleus; the nature of which are still unknown (153). However, this oligomerization is not dependent upon the C-terminus, which is involved interactions at the 2-fold axes (44).

To date, a likely model of AAV assembly is that VPs (either as monomers or oligomers (**Fig 1.4**)) are able to enter the nucleus via NLS sequences or NLM motifs. Diffusion is also a possible route in the case of VP3 monomers (154). Once in the nucleus, they interact with the N-terminus of AAP by a common domain in their C-termini (44). AAP transports them to the nucleolus and induces capsid assembly through a scaffolding or chaperone mechanism that pushes the

VPs toward an assembly-competent state. Once formed, the capsid is transported to the sites of genome replication where Rep proteins can insert the genome into the capsid via the 5-fold pore (41).

Pathogenesis

3.1 Parvovirus pathogenesis

Many members of the *Parvoviridae* family can cause disease, ranging from mild to fatal. During acute viremia, viral shedding has been detected in feces, urine, saliva, and nasal mucus, and in some cases shedding can be detected for weeks after the initial infection. Transmission likely involves direct contact with infected secretions, inhalation of particles from respiratory droplets, or through sexual contact. Since replication is dependent upon S-phase associated proteins, the most common pathologies are in cells with high division rates including fetal tissue, intestinal epithelium, hepatocytes, and the bone marrow. Apoptosis is characteristic of parvovirus infections and is mediated by the non-structural (NS) proteins, either through direct manipulation of cellular proteins or by trans-activating host genes involved in inflammation. The pro-inflammatory pathways evoked by some parvovirus can lead to the production of auto-antibodies. Parvovirus infections can result in persistence, but latency is rare except in the case of AAV (155). Due to their requirement for cycling cells, the morbidity of parviral infections tends to correlate with the age of the host, with the most severe consequences occurring in fetuses and newborns. Even AAV, which is regarded as non-pathogenic, can result in fetal death when given to mice during

pregnancy (156). The following section will cover the pathogenesis of several well-known parvoviruses.

B19: Parvovirus B19 (erythrovirus B19) is a member of the *Erythroparvovirus* genus of *Parvoviridae* and the only known parvovirus that can cause severe disease in humans. The viral capsid is composed of two structural proteins, VP1 and VP2 (157). Unlike most parvoviruses, the B19 VP1 unique N-terminus is partially exposed on the capsid surface and susceptible to antibody binding (158). B19 also lacks the prominent 3-fold spikes that are typical of parvovirus capsids (159). Aside from the VPs, there is only one major NS protein, NS1, and two small proteins of unknown function. Transmission commonly arises though respiratory droplets, but can also occur following transfusion of blood products from infected donors (160). The tropism of B19 for the hematopoietic lineage comes from its use of P antigen as a receptor (161). P antigen is produced in many cells types, including immune and endothelial cells, but is highly prominent in erythroid progenitor cells where B19 is known to replicate. Initially, B19 produces a high viremia and the characteristic giant pronormoblasts, or 'lantern cells' (162). Cytopathic effects from the NS1 protein leads to apoptosis, which can result in various cytopenias (bone marrow failure). This is especially a concern for developing fetuses, which can become infected with B19 during the second trimester of pregnancy when P-antigen is present on trophoblasts. Because erythropoiesis is highly active in the liver at this stage of development and due to the short erythrocyte life span during this time, disruptions in erythrocyte

development caused by B19 replication can have severe outcomes including anemia, hydrops fetalis (accumulation of fluid), hepatitis, and myocarditis. These symptoms can resolve without complications, but may also result in fetal death (163). In addition to inducing apoptosis, NS1 can also transactivate TNFα (164) and IL-6 (165), leading to an inflammatory state that contributes to the development of auto-antibodies, arthralgia, and long-lasting autoimmune disorders (166). In children, B19 commonly causes a symmetrical rash on the cheeks known as 'fifth disease', which later spreads to the rest of the body and is associated with general malaise and flu-like symptoms. In healthy patients, infection is controlled by IgG antibodies, but persistence of the virus is common and some patients test positive for B19 DNA-positive blood samples years post-infection (167). The mechanism of persistence is unknown and latency has never been conclusively demonstrated. There is no vaccine or treatment that targets B19, but IVIG can be given if medically necessary.

Carnivore parvoviruses: The carnivore parvoviruses belong to the *carnivore protoparvovirus* genus, which contains several highly related species that infect a wide variety of carnivores including dogs, cats, minks, foxes, and raccoons. Feline panleukopenia virus (FPV) and canine parvovirus (CPV) are a significant concern for domesticated animals, especially in shelters. Vaccination is available, but interference of maternally derived antibodies can leave young animals vulnerable after early vaccination (168). FPV and CPV exhibit the infectious and pathogenic patterns typical of parvoviruses, are easily transmissible, and have

high fatality rates. Like B19, the clinical manifestations of infection vary with the age of the host and the tissue types currently undergoing mitosis.

Feline panleukopenia virus (FPV): The FPV genome encodes two NS proteins and two capsid proteins. It can infect cats and related species, but does not infect dogs due to species specific differences in the transferrin receptor utilized by FPV for entry into host cells. *In utero* transmission can result in fetal death or damage to the developing neuronal tissue. In adult cats, the disease manifests as loss of myeloid cells resulting in panleukopenia and enteritis from infection of rapidly dividing cells in the gut. Infected cats usually die within 5 days (169).

Canine parvovirus: CPV is derived from FPV (or another closely related carnivore parvovirus) and the two viruses differ in a structural change in the VP2 protein that allows the virus to bind the canine transferrin receptor, effectively changing its host range from cat to dogs. The appearance of this new virus in the late 1970's resulted in worldwide dissemination and within a few years a new variant, CPV-2a, emerged that was able to infect both dogs and cats, replacing CPV2 as the most common CPV (170, 171). Unlike FPV, CPV is not characterized by panleukopenia even though it can infect bone marrow cells (172), but it does cause severe enteritis in adult animals and loss of osmotic regulation (173). CPV can also cause neuronal damage in developing fetuses, but myocarditis is more common and has a mortality rate of 20-100% in infected litters (174, 175).

Minute virus of mice: Minute virus of mice (MVM) was originally isolated from a stock of mouse adenovirus (176), but it is an autonomously replicating parvovirus that does not require Ad co-infection for propagation. Transmission is likely through a fecal-oral route. There are currently two common strains of MVM used in the laboratory: MVMi and MVMp. MVMi, but not MVMp, infection can be fatal in several inbred mouse strains when the host pup is still a neonate (177, 178). The difference in lethality between the two stains is due to differences in cellular tropism (179), with MVMi infecting T-cells and MVMp infecting fibroblasts. MVM infection is acute and without persistence. MVM has served as a laboratory model for parvoviruses and small icosahedral viruses, in general. Of note, there is also a mouse parvovirus that can cause persistent, non-pathogenic infections in mice (180) and is able to suppress the immune system (181), which MVMi is not able to do *in vivo*.

3.2 AAV pathogenesis

AAV is ubiquitous in the human population worldwide and seroprevalence is high for multiple serotypes (182, 183). Routes of transmission are not known, but its association with Ad suggests inhalation while tissue distribution of the virus suggests fecal-oral (9) and sexual transmission (184).

AAV has long been considered non-pathogenic and there is no epidemiological evidence to connect AAV to any disease. On the contrary, AAV infection may be beneficial to the host and AAV has been suggested as a commensal virus of mammals (185). There are several reasons for this unusual

proposal. First, since AAV requires a helper virus for substantial protein production and replication, it has little effect to the host cell on its own and its preferred integration site is considered a 'safe harbor' that does not result in cytopathic consequences for the cell (186). The gene interrupted by AAV integration at the AAVS1 site is protein phosphatase 1 regulatory subunit 12C (PPP1R12C), which may be involved in the actomyosin contractility pathway (187), but the function of this subunit is not well described. Expression levels of PPP1R12C remain normal after integration (188) and many studies of site-specific integration at AAVS1 have not shown any overt deleterious phenotype in cell-culture systems (77, 188). Non-integrated AAV remains episomal (79) with low levels of transcription and is thus unlikely to have negative consequences for the host. Second, several reports have shown that AAV has anti-oncogenic properties (189-193) and is considered a negative risk-factor for cervical carcinoma (56, 194-196), a phenomenon likely mediated by Rep (55, 197). Third, AAV is also known to lower titers of pathogenic viruses in cell culture in a Rep dependent-manner (198-200).

While there is no obvious link of AAV to a disease, there have been reports that AAV may be a factor in male infertility (201), spontaneous abortion (202), or other reproductive complications (203), but no studies to date have been able to demonstrate a clear or direct role for AAV in these adverse events. A report of rAAV-induced hepatocellular carcinoma (HCC) in specifically neonatal mice (204) by integration of the vector at the highly active RIAN locus (205) has caused concern. Another recent report by Nault et al. of wtAAV integrations

associated with HCC in human tissue samples led the authors to conclude that AAV had a pathogenic role in the development of these cancers (78, 206). In the Nault study, the wtAAV was not integrated at the orthologous RIAN locus (DLK1-DIO3 in humans) and the strong interpretation of the data has been questioned (207, 208). The controversy over AVV-induced HCC is still on going (209) and further studies are currently being conducted (73).

Gene Therapy

4.1 Development of recombinant AAV vectors for *in vivo* gene therapy

In the early 1980's, the methods for human gene therapy were focused on *ex vivo* manipulation of bone marrow cells that could be transduced with either plasmids (via chemical transfection or microinjection), DNA viruses, or retroviruses, with the latter being the most feasible and popular (210). An early concern with retroviruses was the random nature of the genome integration and the chance for recombination with endogenous retroviral elements to produce infectious virions (210). AAV represented an attractive alternative to retroviruses because of its ability to integrate safely and site-specifically into AAVS1 (21, 211), its lack of association with any known disease, and the production capacity to both grow and concentrate the virus to high titer (212). Despite subsequent work demonstrating that rAAV does not integrate site-specifically, AAV is still considered one of the most promising vectors for gene therapy because it is capable of long-term transgene expression from an episomal state.

In some ways, the path for developing AAV into the vector it is today began with a decision by the NIH RAC in the early 1980s, following suggestion by ASM virologists, to allow the molecular cloning of human viruses into plasmid backbones without the need for high level containment (211). This opened the door for the creation of AAV-containing plasmids and recombinant vectors for study in model systems. Samulski et al. and Laughlin et al. developed the first AAV plasmid constructs in 1982 and 1983, respectively. Transfection of these plasmids into human cell lines along with Ad infection resulted in wtAAV virions (20, 213), demonstrating that Ad could "rescue" AAV genomes that had been integrated into plasmids. Soon after, several groups successfully infected and transduced mammalian cell lines with rAAV using either the SV40 promoter (21) or the native AAV p40 promoter (83). The discovery that AAV was able to integrate (58) site-specifically (59, 60) generated interest in using AAV as a human gene therapy vector (60, 212, 214), but there was uncertainty at the time over how faithfully the rAAV vectors would mimic the site-specific integration seen in their natural counterparts. The vector production methods at the time used a "two-plasmid" approach. The only *cis*-element needed for packaging wild-type or recombinant AAV genomes are the ITRs (215), hence cells were transfected with one plasmid (a vector plasmid) containing a transgene flanked by the ITRs and one plasmid (a helper plasmid) containing the *rep* and *cap* genes. Infection with Ad enabled replication and packaging of the ITR-containing plasmid only. In some systems, only the *cap* gene was replaced in the vector plasmid, so the resulting vectors were Rep(+). Unfortunately, these methods of vector production

often resulted in Rep mediated integration of the rAAV genomes from Rep(+) vectors or contamination with wtAAV though recombination of the plasmids during viral production. Thus, there was an early and incorrect notion that rAAV could mediate long-term transduction through site-specific integration at AAVS1.

The first clinical trial using rAAV to treat cystic fibrosis (CF) began in 1996 (216) and this prompted the researchers involved to determine the fate of rAAV vectors by infecting an immortalized CF bronchial epithelial cell line. Using FISH and Southern blotting, Kearns et al. found that 6% of the rAAV genomes were integrated and that they were integrated randomly (217). This study, and several since then, showed that the predominant mechanism of rAAV transgene persistence is from head-to-tail concatemerized episomes (75, 76, 218, 219). At the same time, second-strand synthesis of the single-stranded vector was identified as a significant rate-limiting step in cellular transduction (220). This finding eventually lead to the creation of self-complementary vectors (sc-AAV), which had a 5-140 fold increase in transduction, but reduced by half the already constrained AAV capacity (221).

Another important step towards the use of rAAV vectors *in vivo* was the development of a triple transfection method that forego the use of helper viruses, thus greatly reducing the concern over wild-type virus contamination (222) (**Fig 1.6**). In this system, the Ad genes necessary for AAV replication are provided in a plasmid, often called pHelper. When plasmids containing the *rep* and *cap* genes from AAV (pRepCap) and a transgene flanked by AAV ITRs (pVector) are

transfected into HEK 293 cells along with pHelper, rAAV can be isolated with a

significantly lower chance for contamination by replication competent viruses.
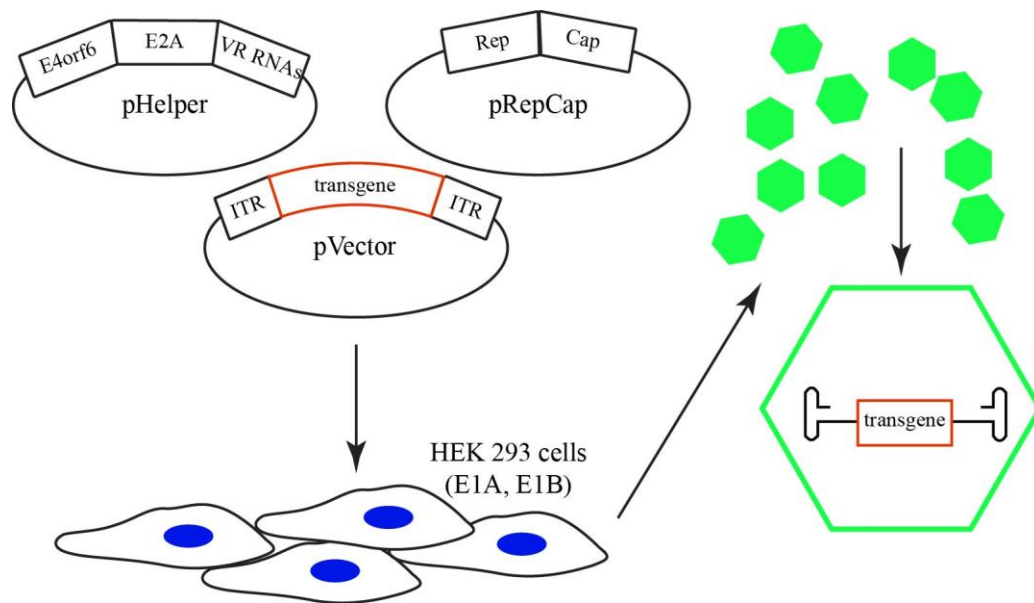
Figure 1.6 Triple transfection method of rAAV production
pVector (a plasmid containing a transgene flanked by the AAV ITRs), pRepCap (a plasmid containing AAV *rep* and *cap*, with their native promoters), and pHelper (a plasmid containing E2A, E4orf6, and the VA-RNAs) are transfected into HEK 293 cells. 2-5 days later the cells and media are harvested and rAAVs are isolated and purified by ultracentrifugation.

This method is the current standard for the creation of rAAV and was used in the first hemophilia B trials for *in vivo* gene therapy in human patients (223). In this initial study rAAV containing a Factor IX (F.IX) transgene was injected intramuscularly into hemophilia B patients and resulted in F.IX levels at ~1% of what would be seen in healthy individuals. While this F.IX production was low, this study demonstrated the safety of using rAAVs in humans and revealed an inconsistency in pre-clinical studies using non-human animals, namely that these studies underestimated the amount of vector that would be needed to achieve therapeutic levels of transgene expression in humans (223). A follow-up dose-escalation study using the hepatic artery for delivery was able to achieve therapeutic levels of F.IX using doses of $4x10^{11}$ - $2x10^{12}$ viral genome / kg, but the induction of a T-cell mediated cytotoxicity against capsid epitopes caused the levels to decrease at ~4 weeks post infusion, eliminating the therapeutic effect (224, 225). This T-cell response was unanticipated since none of the pre-clinical studies had demonstrated any toxic immune responses to the vectors. While these early trials were underway, the discovery of new AAV serotypes and their specific cellular tropism expanded the clinical use of new rAAVs that could target different tissue types *in vivo* (226). Using this knowledge, a third trial for hemophilia B was conducted using an improved sc-AAV F.IX vector genome and a liver-tropic AAV8 capsid. To circumvent the destruction of transduced cells, patients were given corticosteroids if they started to develop an anti-AAV immune response (227). This treatment led to the successful production of F.IX levels (3-11% of normal) in each patient and those levels remained stable in

follow-up years later. Other successful clinical trials using rAAV have been for the treatment of Leber congenital amaurosis (228-230), Duchenne muscular dystrophy (231, 232), and lipoprotein lipase deficiency (233), the latter has led to the first approved gene therapy treatment in Europe and uses an AAV1-based vector. There are several recent reviews that list some of the numerous ongoing trials using rAAV-based vectors (234-236).

4.2 Engineering of recombinant AAV vectors

Because AAV2 was the first AAV to be cloned, the first rAAVs were also AAV2 based, but the discovery of numerous naturally occurring serotypes and variants, each with their own tropism and biological properties, has allowed for a greatly expanded repertoire of AAV capsids with which to work. In conjunction, the discovery that AAV chimeric capsids could easily be assembled from different serotypes has led to the idea of an "AAV vector toolkit" composed of lab-created novel capsids (237, 238). Currently, the creation of unique capsids is a growing and competitive field within the AAV gene therapy community. Naturally occurring capsids are being altered, combined between serotypes, and modified in creative ways to design the most suitable capsid type for specific disease treatments. The next section will briefly cover some of the more common methods, which are predominately focused around the generation of a library of different AAV mutants and then using a selection method for identifying capsids with desirable properties. This selection method is mostly commonly called directed evolution or bio-panning.

Capsid Shuffling: Grimm et al. described the first method for generating chimeric AAV capsids using a modified DNA shuffling (239) procedure in 2008 (240). Previously, chimeric and mosaic capsids had been made through a limited combination of serotype helper plasmids (241, 242) and rational domain swapping (243, 244). The method used by Grimm et al. employed digestion of the *cap* gene from eight different AAV serotypes with DNaseI and then reassembly in a random fashion with primerless PCR. These newly created chimeric *cap* genes were cloned into an AAV helper plasmid and transfected into HEK 293 cells for AAV production using the triple transfection method previously described. This library of chimeric virions was used to infect human hepatocytes in the presence of IVIG and Ad (to allow for replication of successfully infecting AAV particles). The nascent AAV virions were harvested and the process was repeated, eventually resulting in the isolation of AAV chimeras that were the most successful in infecting hepatocytes with the lowest rate of neutralization by anti-AAV antibodies in the IVIG. This resulted in the AAV-DJ capsid, a chimera of AAV2, 8, and 9, which had an increased ability to infect numerous cell types *in vitro* and a higher transduction rate *in vivo*. This technique has now been refined and expanded upon by several labs (245-247).

Peptide Display: AAV capsids can be engineered to give novel tropism by inserting targeting peptides into the VPs, especially if these peptides are receptor ligands. This strategy was made feasible by several early studies that identified regions of the AAV capsid tolerant to peptide insertion (248-250) and are summarized in several reviews (251, 252). This strategy has been successfully

used to generate AAVs that target heart vascular endothelial cells (253), brain vascular endothelial cells (254), and the CNS (255). Peptide insertion has also been used to add fluorescent proteins into the capsid proteins (256), which can be used to track the virus during infection (257).

Immune evasion: The high prevalence of anti-AAV antibodies in the human population (182, 183) is a barrier to AAV-based gene therapy and current trials must exclude patients with high antibody titers. In addition, if the gene therapy requires more than one treatment, the antibodies developed from the first infusion of rAAV vector will diminish the effectiveness of the second treatment. Hence, several groups are working on the creation of novel AAVs that can evade naturally occurring neutralizing antibodies. Aside from the aforementioned use of IVIG during mutant capsid library screening (240), other approaches are to rationally-design vectors in which identified antibody epitopes have been altered (258) and to reconstruct ancestral AAV capsids that are not currently in circulation (259, 260).

Other alternative techniques can include, site-specific point mutations, error-prone PCR, and hexapeptide scanning. These methods, among others, are common ways to create AAV mutant libraries and also identify capsid amino acids that play an important biological role in the AAV life cycle. As technology progresses, so have the ways to generate and identify new AAV mutants. The Nakai lab pioneered a Barcode-Seq approach in which the encapsulated genome contains a unique DNA barcode, allowing for the characterization of hundreds of different AAV variants in a high-throughput manner (261). In another example,

the Zolotukhin lab combined the virtual design of vectors from knowledge of the capsid structure with high-throughout sequencing and *in vivo* selection to identify a liver-tropic AAV2 mutant (262). Together, these techniques paired with increasing knowledge of AAV biology and technological progress should result in the AAV toolkit for which many investigators are hoping.

The nucleolus in cell and viral life cycles

5.1 Structure and biogenesis of the nucleolus

The nucleolus is the largest structure within the nucleus and predominantly known as the site of ribosome biogenesis, but this membraneless organelle is also involved in multiple aspects of cellular function including cell division and stress responses. The nucleolus is formed from the act of creating ribosomes and centers around the ribosomal genes (rDNA) within chromosomes (263). These nucleolar organizing regions (NORs) (264) contain the genes for the 18S, 5.8S, and 28S ribosomal subunits and are located in the short arms of the acrocentric chromosomes (263, 265). The size of the nucleoli is dependent upon the level of ribosome production and is adjustable given the demands of the cell (263). The nucleolus is composed of three distinct regions: the granular component (GC) is the outer layer which surrounds the dense fibrillar centers (DFC), which in turn contain the innermost fibrillar centers (FCs), although this organization can change depending on the organism and cellular state (**Fig 1.7**) (263, 266, 267). Active transcription of ribosomal genes at the NORs by RNA polymerase I (Pol I) takes place at the border between the FC and the DFC zones (**Fig 1.7E**) (266).

The pre-ribosomal RNA (rRNA) transcript is processed by nucleolar proteins and small nucleolar RNAs (snoRNAs) in the DFC and GC, with pre-40S and pre-60S particles migrating to the GC where subunit assembly takes place (266, 268). Even though transcription takes place at the edge of the FC region, the FC itself contains many of the factors required for this process including rDNA, Pol I, DNA topoisomerase I, and upstream binding factor (UBF)
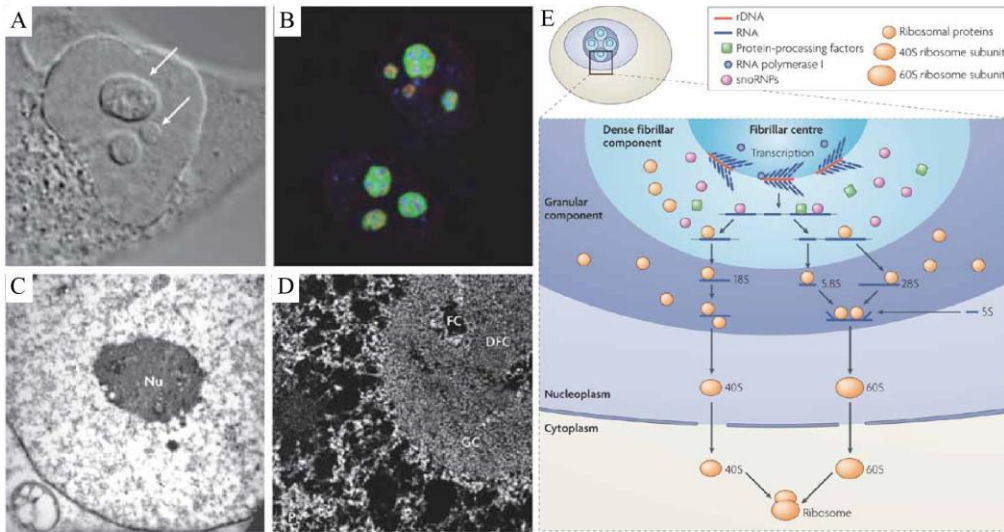
Figure 1.7 The nucleolus and ribosome biogenesis
(**A**) Differential interference contrast (DIC) image of a HeLa cell showing prominent nucleoli within the nucleus (indicated by arrows). (**B**) Immunofluorescence labelling of a HeLa cell with antibodies that are specific for proteins enriched in the GC (B23; shown in green), the DFC (fibrillarin; shown in red) or the FC (RNA polymerase I subunit RPA39; shown in blue). (**C**) DIC image of nucleoli purified from HeLa cells. The inset shows a scanning EM image of a purified HeLa nucleolus. (**D**) Uranyl-acetate-stained cell section showing a characteristic image of a nucleus with a nucleolus imaged by transmission EM. (**E**) Transcription of ribosomal DNA (rDNA) by RNA polymerase I occurs either in the fibrillar centers (FCs) or at the boundary between the FC and the dense fibrillar component (DFC) region. The pre-ribosomal RNA transcripts are spliced and modified by small nucleolar ribonucleoproteins (snoRNPs) in the DFC. Final maturation of the pre-ribosomal ribonucleoprotein and assembly with ribosomal proteins occurs mostly in the granular component (GC) region. In the GC, the 5.8S and 28S ribosomal RNAs (rRNAs) assemble with the 5S rRNA transcript to form the 60S subunit, whereas the 18S rRNA alone assembles into the 40S ribosome subunit. The 40S and 60S ribosome subunits are both exported to the cytoplasm, where they bind to mRNA to form functional ribosomes.

Images and text are from Boisvert et al., 2007, reprinted with permission.

(263). The DFC, as the name implies, contains densely packed fibrils as well as proteins and snoRNAs involved in the initial steps of rRNA processing. The nucleolar marker protein fibrillarin is found primarily in this region of the nucleolus. The GC region is also named because of its contents, in this case densely packed riboprotein granules. The nucleolar marker proteins nucleostemin and nucleophosmin primarily reside here, along with other components required for the assembly of the large ribosome subunits.

Given that the nucleolus is tightly tied to ribosome production, cell cycle phase has considerable effect on the nucleolus. As the cell progresses into M phase and transcription stops, the nucleolus breaks down in early prophase and many of the proteins and snoRNAs normally concentrated in the nucleolus relocalize and become associated with chromosomes throughout mitosis. Some, like nucleophosmin and fibrillarin, become distributed over the chromosomes, while others involved in rDNA transcription, like Pol I and UBF, remain associated with the NORs (263, 266, 269). During anaphase many nucleolar components remain associated with chromosomes, but can also be found in nucleoli-derived foci (NDF) in the cytoplasm along with other nucleolar-associated proteins (266, 270, 271). The exact function of these NFDs is still unclear, but their disappearance correlates with the appearance of prenucleolar bodies during telophase, indicating that they might be a way to sequester and then deliver these components back to reforming nucleoli at the proper time (266,

270). By late telophase, the nucleolus is completely reformed around the NORs (266).

5.2 Cellular stress and the nucleolus

The nucleolus is able to sense and control cellular stress responses though several mechanisms, notably via p53-MDM2 regulation. p53 is an important cellular stress-response transcription factor that regulates DNA damage and cell cycle pathways. There are multiple pathways that control p53 activation, but a primary mechanism is through the regulation of its stability by MDM2. MDM2 is an E3 ubiquitin ligase that can target p53 for proteasome degradation and these two proteins exist in a regulatory feedback loop in which p53 can cause transcription of MDM2 and in turn, MDM2 ubiquitinates p53 to target it to the proteasome (272, 273). The nucleolus contains many proteins that interact with MDM2 or other p53 regulators to enable p53-induced cellular arrest during times of ribosomal stress. Primarily, Arf has been implicated as playing a central role in the stabilization of p53 by binding a central acidic region of MDM2, which prevents proper folding of MDM2 and activation of its ligase activity (273). Additionally, this interaction leads to sequestration of MDM2 in the nucleolus (272, 273). Nucleostemin and several other ribosomal proteins can also bind to MDM2 in this same region and prevent E3 ligase activity (273).

Ribosomal processing destabilization can itself lead to cellular stress. Upon exposure to anisomycin, oxidative stress, or other JNK pathway activators, the Pol 1 transcription factor TIF-IA is phosphorylated by JNK2, which causes relocation

of TIF-IA to the cytoplasm and also prevents interactions with Pol I and TIF-IB (266, 273), thus preventing further rDNA transcription.

5.3 Viral interactions with the nucleolus

Given the importance of the nucleolus in ribosome production, RNA processing, and cell cycling, it's not surprising that many viral families encode proteins that localize to the nucleolus or interact with nucleolar proteins (274). Both positive-strand and negative-strand RNA viruses have interactions with the nucleolus, even though many of these viruses replicate in the cytoplasm (275). The multifunctional role of the nucleolus is also reflected in the myriad of different ways viruses interact with the nucleolus. Because of its role in p53 regulation, viruses such as West Nile Virus can translocate MDM2 to the nucleolus, thus activating p53-induced apoptosis (276) and alphaviruses can translocate p21 to the nucleolus which also leads to apoptosis (277). This is in contrast to viruses that target the nucleolus to prevent apoptosis, such as influenza A which contain an NS1 protein that can bind to p53 (275, 278). Aside from cell cycle control and apoptosis, the nucleolus is also targeted because of its role in cellular transcription. Several picornaviruses can shut down host cell Pol I transcription by inhibiting UBF and nucleophosmin, while leaving viral RNA transcription and translation intact (275, 279, 280).

Since many DNA viruses replicate in the nucleus, its not unexpected that these viruses also have interactions with the nucleolus. A well described mechanism for HSV intronless viral mRNA export occurs via the HSV ORF57

protein and hTREX protein (281, 282). Indeed, many HSV proteins localize to the nucleolus and interact with nucleolar proteins (275, 282). Both nucleophosphim and nucleolin are common targets for viral proteins and nucleolin has been described as a co-receptor for multiple viruses during cellular entry (275) as well as trafficking HSV capsid proteins from the nucleus to the cytoplasm (283).

5.4 AAV2 interactions with the nucleolus

When Wistuba et al. made the first AAV2 capsid antibodies, they were able to visualize AAV2 intact capsids initially appearing in the nucleolus, before spreading to the rest of the infected cell (148, 153). Aside from AAV capsid nucleolar localization, several nucleolar proteins are able to interact with, or at least colocalize with, AAV2 capsids during assembly, including nucleolin (284, 285), nucleostemin (150), and nucleophosmin (134, 285) and incoming AAV2 capsids localize to the nucleolus after infection (151). There is also evidence that the AAV2 Rep78 protein can interact with nucleolin and nucleophosmin (286). Together, these findings indicate a role for the nucleolus in the life cycle of AAV2. A caveat to this is that there is currently no evidence for interaction of VP monomers with the nucleolus or nucleolar proteins, so the functional relevance of these interactions during capsid assembly has yet to be determined. Interestingly, and possibly due to the predilection for AAV2 to localize to the nucleolus, there may be a preference for rAAV genome integration into the rDNA repeats (71, 287). This phenomenon has become the basis for a class of rAAV vectors which

contain homology regions to the rDNA repeats and thus promote rAAV genome integration by homologous recombination (288-290).

AAV antibodies used in research

The AAV field utilizes several common, commercially available antibodies for research purposes. The following section will discuss those antibodies and their binding sites.

6.1 AAV2 antibodies

A20: A20 is a neutralizing mouse monoclonal antibody that recognizes a conformational epitope in intact AAV2 and AAV3 capsids that is not present in individual capsid protein monomers (**Fig 1.8A**) (153, 291, 292). The antibody was created by Wistuba et al. by injecting BALB/c mice subcutaneously with purified AAV2 capsid proteins and complete Freund's adjuvant (153). 4 weeks later, the mice were given a booster of UV-inactivated AAV2 virus. After sacrificing the mice, hybridoma cells were created by fusing harvested spleen cells with X63/Ag8 cells. Specificity of A20 for intact AAV2 capsids was demonstrated by western blot, in which cell lysates from HeLa cells co-infected with AAV2 and Ad were separated on SDS-PAGE and probed with different anti-AAV antibodies. Polyclonal anti-serum was able to recognize the individual capsid proteins, but A20 was not. (148). Later work using the Fab' fragment of A20 identified the binding footprint as extending from the plateau next to the spike at the 3-fold axis, along the 2-fold axis, and into the canyon near the 5-fold pore (293). Of note, this work was done at OHSU by Dustin McCraw as part of his

dissertation project and a full detail of the methods and results can be found in his dissertation: Towards a structural understanding of adeno-associated virus serotype 2 and its recognition by antibodies

 (http://digitalcommons.ohsu.edu/etd/766/).

B1: The B1 antibody is a mouse monoclonal antibody that recognizes the amino acid sequence IGTRYLTR in the very C-terminus of the AAV2 VP proteins (**Fig 1.8B**) (291). The antibody was created in the same manner as the above mentioned A20 antibody (153). This region is common to all the VPs and unavailable for binding when the VPs are assembled into capsids. It resides at the 2-fold axis, and interestingly, overlaps with the proposed AAP-binding region (44). While this antibody was raised against AAV2 capsid proteins, it is capable of recognizing the VPs from AAV serotypes 1-11, but has only weak reactivity with AAV serotypes 4 and 11. AAV4 and AAV 11 are distantly related to AAV2 and their corresponding regions are IGSRYLTH and IGSRYLTN, respectively. The remaining AAV serotypes are completely conserved in this region.

A69: A69 is a mouse monoclonal antibody created in the same manner as the above antibodies and recognizes a linear epitope in the VP1 and VP2 N-terminal regions at amino acids LNFGQTGDADSV (**Fig 1.8B**) (291). The sequence is not found in VP3 and is not available for binding in the intact capsid because it is likely that the VP1/VP2 N-terminal extensions are tucked within the AAV capsid.

A1: A1 is a mouse monoclonal antibody created in the same manner as the above antibodies and recognizes a linear epitope in the VP1 unique N-terminal region at amino acids KRVLEPLGL (**Fig 1.8B**) (291). As with A69, this epitope is unavailable for binding with intact AAV capsids because it is likely that the VP1 N-terminal extension is tucked within the AAV capsid.

All the above antibodies are commercially available and the A20 antibody is also the basis for an ELISA kit specific to intact AAV2 capsids (294).

6.2 Non-AAV2 capsid antibodies

ADK4 and ADK5: ADK4 and ADK5 are a mouse monoclonal antibodies that recognize intact AAV4 and AAV5 capsids, respectively. The creation of these antibodies is described in Kuck et al. 2007 and was done by exposing C57Bl/6 mice to 1–5×10^{10} genome containing particles intranasally, followed by boosting the mice 7 days, 14 days, and 4 months after the first exposure (295). Five days after the last boost, hybridomas were created by fusing harvested spleen cells with X63/Ag8 cells. Specificity of these antibodies to intact capsids was determined by transfecting 293T cells with AAV4 or AAV5 plasmids and running the cell lysate through sucrose gradients, blotting each fraction onto a membrane, and exposing the membrane to the ADK antibodies. The strongest signal for antibodies was at the fraction expected to contain AAV capsids. Denatured lysate did not show any positive signal from the ADK antibodies.

ADK8/9 and ADK9: ADK8/9 and ADK9 were generated in the same manner as ADK4 and ADK5, but first described in Sonntag et al. 2011 (43). ADK8/9 (called ADK8 in the Sonntag paper) recognizes AAV8 intact capsids and has with weak recognition of AAV9 intact capsids, whereas ADK9 is specific to AAV9 capsids. The specificity of the ADK antibodies for intact capsids was tested in the same sucrose density procedure as described previously and against denatured capsid proteins. The strongest recognition was on blotted sucrose density fractions containing intact capsids (43).

C24 and C37: C24 and C37 are mouse monoclonal antibodies generated in a similar manner to the previously described A1, A69, B1, and A20 antibodies, except for the use of synthetic peptide instead of AAV2 capsid proteins (291). These initial exposures were followed with two boosters of intact AAV2 capsid (291). These antibodies are not commercially available but are described here because they are used in Chapter 4, Fig 4.1, which is from Naumer et al. 2012.
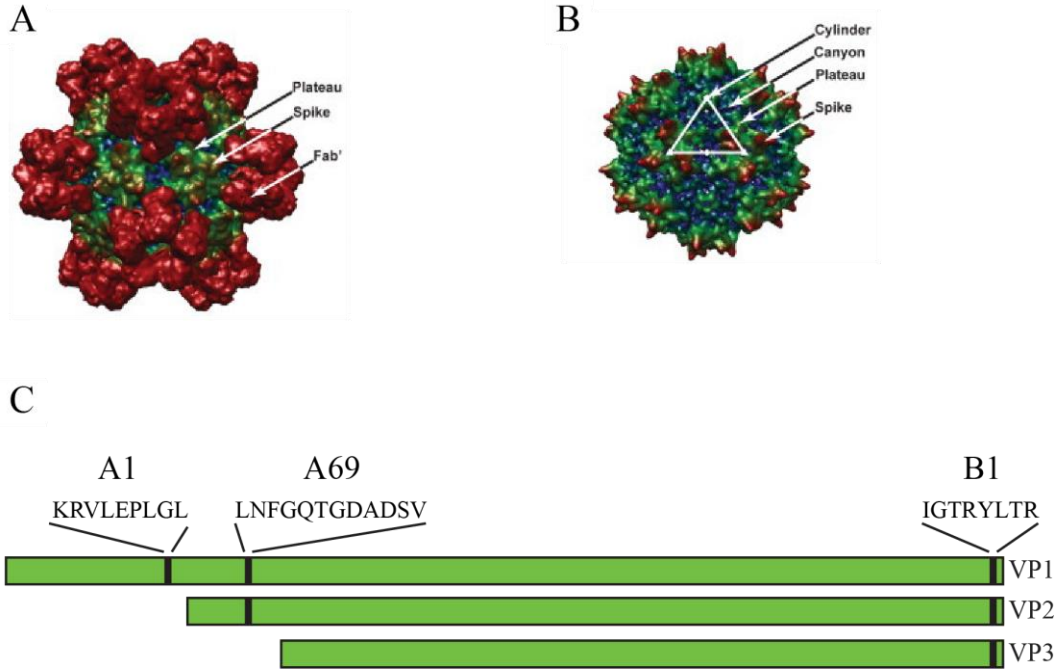
Figure 1.8 Location of AAV2 antibodies
(A) Reconstruction of AAV2 complexed with A20 Fab' compared to a prior *cryo*-EM reconstruction of native AAV2 (B) (O'Donnell et al., 2009) Reprinted with permission from McCraw et al 2009. (C) Location of AAV2 linear antibody epitopes.

Chapter 2: Identification and Characterization of Nuclear and Nucleolar Localization Signals in the Adeno-associated Virus Serotype 2 Assembly-activating Protein

Lauriel F Earley[1], Yasuhiro Kawano[2,3], Kei Adachi[2], Xiao-Xin Sun[2], Mu-Shui Dai[2] & Hiroyuki Nakai[2]

Departments of [1]Molecular Microbiology & Immunology and [2]Molecular & Medical Genetics, Oregon Health and Science University School of Medicine, Portland, Oregon 97239, USA; [3]Takara Bio Inc. Otsu Shiga 520-2134, Japan.

For this work, YK, KA, and HN created the AAPmt plasmids, HN and LE created the LacZ and GFP-AAP fusion plasmids, HN performed the LacZ staining and statistical analyses, LE performed the immunofluorescence microscopy, western blots, ELISAs, XXS assisted with the immunofluorescence microscopy, MSD provided advice about the nucleolus, LE and NH designed the experiments and wrote the manuscript.

Abstract

   Assembly-activating protein (AAP) of adeno-associated virus serotype 2 (AAV2) is a nucleolar localizing protein that plays a critical role in transporting viral capsid VP3 protein to the nucleolus for assembly. Here we identify and characterize AAV2 AAP (AAP2) nuclear and nucleolar localization signals (NLS

and NoLS, respectively) near the carboxy-terminal region of AAP2 (amino acid positions 144-184 or AAP2 144-184). This region contains five basic amino acid-rich (BR) clusters, KSKRSRR (AAP2BR1), RRR (AAP2BR2), RFR (AAP2BR3), RSTSSR (AAP2BR4) and RRIK (AAP2BR5) from the amino terminus to the carboxy terminus. We created 30 AAP2BR mutants by arginine/lysine-to-alanine mutagenesis or deletion of AAP2BRs and 8 and 1 GFP-AAP2BR and β-galactosidase-AAP2BR fusion proteins, respectively, and analyzed their intracellular localization in HeLa cells by immunofluorescence microscopy. The results showed that: AAP2 144-184 has redundant multipartite NLSs and any combinations of 4 AAP2BRs, but not 3 or less, can constitute a functional NLS/NoLS; AAP2BR1 and AAP2BR2 play the most influential role for nuclear localization but either one of these two AAP2BRs is dispensable if all the other 4 AAP2BRs are present, resulting in 3 different, overlapping NLS motifs; and the NoLS is shared redundantly among the five AAP2BRs and functions in a context dependent manner. AAP2BR mutations not only resulted in aberrant intracellular localization but also attenuated AAP2 protein expression to various degrees, and both of these abnormalities have a significant negative impact on capsid production. Thus, this study reveals the organization of the intermingling NLSs and NoLSs in AAP2 and provides insights into their functional role in capsid assembly.

Author Summary

Adeno-associated virus (AAV) has become a popular and successful vector for in vivo gene therapy; however, its biology has yet to be fully understood. In this regard, the recent discovery of the assembly-activating protein (AAP), a non-structural, nucleolar localizing AAV protein essential for viral capsid assembly, has provided us a new opportunity to better understand the fundamental processes required for the virion formation. Here we identify clusters of basic amino acids in the carboxy-terminus of AAP from AAV serotype 2 (AAV2) that act as nuclear and nucleolar localization signals. We also demonstrate their importance in maintaining AAP expression levels and efficient production of viral capsids. Insights into the functions of AAP can elucidate the requirements and process for AAV capsid assembly, which may lead to improved vector production for use in gene therapy. This study also contributes to the growing body of work on nuclear and nucleolar localization signals.

Introduction

Adeno-associated virus (AAV) is a small, single-stranded DNA virus from the parvovirus family, which has become a successful vector for gene delivery. The recent achievements in the field of AAV vector research have called attention to the incompletely understood life cycle of this virus. The AAV genome comprises two genes, *rep* and *cap*, which encode for the non-structural Rep proteins and the structural VP proteins, respectively. The AAV virion is composed of sixty subunits comprising the three VP proteins; VP1, VP2 and VP3, encoded by a

single open reading frame (ORF) in the *cap* gene. Recently, a non-structural viral protein encoded by an alternative ORF within the *cap* gene was identified and termed assembly-activating protein (AAP) for its indispensable role in capsid formation (42-44). AAP from AAV serotype 2 (AAP2) is a nucleolar localizing protein that binds to VP proteins through interacting domains in the amino (N)-terminus of AAP2 (44), transports the VP proteins to the nucleolus and promotes capsid assembly (42). Therefore, AAP2 is expected to have both a nuclear localization signal (NLS) and a nucleolar localization signal (NoLS) within its protein sequence. However, such organelle-targeting sequences in AAP2 remain to be identified and characterized.

The most common mechanism for targeting a protein to the nucleus is by an NLS, which is recognized by one of the nuclear import proteins, termed importins, which are part of the large family of transport proteins known as karyopherins (296). Classical NLSs can be either monopartite, such as the PKKKRKV sequence in SV40 large T-antigen (297), or bipartite, such as the KRPAATKKAGQAKKKK sequence in nucleophosmin (298). These classical NLS signals are bound by the adaptor protein importin-α☐ which is then bound by importin-β, forming a heterotrimeric complex consisting of the two importin proteins and the cargo protein. Importin-β mediates nuclear entry of the heterotrimer through the nuclear pores by its increasing affinity to nucleoporins along the inside of the nuclear pore complex (299). If the cargo protein also contains an NoLS, it can then be targeted to the nucleolus through charge-based

interactions (300) or interactions with nucleolar proteins (301, 302) although the specific requirements defining nucleolar localization are not as well understood as those for nuclear import.

As AAP2 is able to localize to the nucleolus (42), we hypothesized that it would contain both an NLS and an NoLS responsible for this intracellular localization and that these signals would be critical to its function in capsid assembly. Because a protein region rich in basic amino acid residues is a hallmark of NLS and NoLS, we tested our hypothesis on the carboxy (C)-terminal region of AAP2, amino acid positions 144-184 (AAP2$^{144-184}$), where there are five basic amino acid-rich (BR) clusters. By fusing GFP or β-galactosidase protein with an AAP2 protein segment of interest and by creating a series of arginine/lysine-to-alanine mutations or deletions in AAP2$^{144-184}$, we were able to identify NLSs and NoLSs and elucidate their redundant and overlapping nature. Mutations in this NLS/NoLS-containing region resulted in not only aberrant intracellular localization but also substantial reduction in AAP2 expression and capsid production, showing the multifaceted functional importance of the NLS/NoLS in AAP2.

Materials and Methods

Plasmid construction. pCMV$_3$-FLAG-cmAAP2 is a plasmid expressing a codon-modified (cm) version of the wild-type AAP2 with an N-terminal FLAG tag under the control of the human cytomegalovirus (CMV) immediately early gene enhancer/promoter (303). This FLAG-tagged AAP2 is translated from the ATG start codon. The codon-modified AAP2 ORF was utilized to maximize expression in human cells and prevent recombination between AAV2-RepVP3 viral genome (303) and AAP2 plasmid DNA through the homologous sequence during AAV virus production in human kidney embryonic (HEK) 293 cells. pCMV$_3$-FLAG-cmAAP2 mutant plasmids were created by site-directed mutagenesis. pCMV$_3$-GFP is an enhanced GFP (eGFP)-expressing plasmid under the control of the same CMV enhancer/promoter as that in pCMV$_3$-FLAG-cmAAP2 and was used to construct plasmids expressing GFP fused with the AAP2$^{144-184}$ peptide at its C-terminus. pAAV2-RepVP3 is a plasmid that expresses all the Rep proteins and the VP3 protein from AAV2, but does not express VP1, VP2 or AAP2 (303). An Ad helper plasmid, pHelper, was purchased from Agilent. pCMV$_1$-AAV2VP3 is a plasmid expressing the AAV2 VP3 protein under the same CMV enhance/promoter (303). pAAV-CMV-lacZ is a plasmid containing an alcohol dehydrogenase (*Adh*)-*lacZ* fusion transgene under the control of the CMV enhance/promoter (71). This plasmid expresses cytosolic β-galactosidase in cells transfected with the plasmid. To construct pAAV-CMV-lacZ-AAP2$^{144-184}$ plasmid expressing β-galactosidase fused with the AAP2$^{144-184}$ peptide, the peptide-coding nucleotide sequence was introduced at the amino (N)-terminus between the start

and second codons of the *Adh-lacZ* gene ORF in the plasmid pAAV-CMV-lacZ. pCMV$_3$-FLAG-cmAAP2-GFP is a plasmid expressing the wild type full-length AAP2 fused with a FLAG tag and GFP at the N-terminus and C-terminus, respectively. This protein configuration allows for simultaneous detection of the wild-type AAP2 in cells using three different approaches; anti-FLAG antibody immunostaining, direct detection of the GFP fluorescence, and anti-GFP antibody immunostaining.

Cells. HEK 293 cells (AAV293) were purchased from Stratagene. Human cervical cancer cell line HeLa was obtained from American Type Culture Collection (ATCC). HEK 293 cells and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM, Lonza, Basel, Switzerland) supplemented with 10% fetal bovine serum (FBS), L-glutamine and penicillin-streptomycin.

Immunofluorescence microscopy and data analysis. HeLa cells were seeded on cover slips in 12-well plates and transfected with plasmid DNA using polyethyleneimine (PEI). Forty-eight hours after transfection, the cells were fixed with 4% paraformaldehyde at room temperature, permeabilized with 0.2% Tween 20, blocked with 8% bovine serum albumin (BSA), and stained with mouse monoclonal anti-FLAG M2 antibody (F1804, Sigma-Aldrich, St. Louis, MO), rabbit polyclonal anti-nucleostemin antibody (AB5689, Millipore, Billerica, MA or sc-67012, Santa Cruz Biotechnology, Dallas, TX) and then followed by 4',6-diamidino-2-phenylindole (DAPI), Alexa Fluor 488-AffiniPure goat anti-mouse

IgG antibody (115-545-166, Jackson ImmunoResearch, West Grove, PA), and Cy3-AffiniPure goat anti-rabbit IgG antibody (111-165-144 Jackson ImmunoResearch). For imaging of AAP2 and VP proteins together, the cells were stained with rat monoclonal anti-DYKDDDDK (FLAG) antibody (NBP1-06712, Novus Biological Littleton, CO), mouse monoclonal anti-AAV VP1/VP2/VP3 antibody (B1) (03-61058, American Research Products, Inc., Waltham, MA), rabbit polyclonal anti-nucleostemin antibody (sc-67012, Santa Cruz Biotechnology), and then followed by DAPI, Alexa Fluor 488-AffiniPure goat anti-mouse IgG antibody (115-545-166, Jackson ImmunoResearch), Cy3-AffiniPure donkey anti-rat IgG antibody (712-165-153, Jackson ImmunoResearch), and Alexa Fluor 647-AffiniPure goat anti-rabbit IgG antibody (111-605-144, Jackson ImmunoResearch). For a subcellular localization analysis of β-galactosidase, the cells were treated in the same manner as that for the double immunostaining of AAP2 and nucleostemin described above except that mouse monoclonal anti-β-galactosidase antibody (B0271, Sigma-Aldrich) was used in place of mouse monoclonal anti-FLAG M2 antibody. For antigen retrieval by protease treatment, transfected HeLa cells were fixed with 4% paraformaldehyde at 48 hours post-transfection, permeabilized with 0.2% Tween 20, incubated with trypsin (2.5 μg/ml) at 37 °C for 10 minutes, treated with 1mM of PMSF at room temperature for 1 min, and immunostained using mouse monoclonal anti-eGFP antibody (F56-6A1.2.3, Thermo Scientific, Waltham, MA) and rat monoclonal anti-DYKDDDDK (FLAG) antibody (NBP1-06712, Novus Biological Littleton, CO), followed by Alexa Fluor 647-AffiniPure donkey anti-mouse IgG antibody

(A-31571, Invitrogen, Grand Island, NY) and Cy3-AffiniPure donkey anti-rat IgG antibody (712-165-153, Jackson ImmunoResearch). The cells were imaged on a Zeiss LMS 710 laser scanning confocal microscope. In this study, we defined NLS (+) as those that showed exclusive nuclear accumulation, NoLS (+) as those that showed obvious nucleolar enrichment, and nucleolar exclusion (Ex) as those that showed clear exclusion from the nucleolus.

The functional role of each AAP2BR in nuclear and nucleolar localization was evaluated in the following manner. The following mutants harboring a pair of AAP2BR mutations showed impaired nuclear trafficking: AAP2mt12, mt17, mt21, mt22 and mt27. In these 5 mutants, when correction of one AAP2BR mutation back to the wild-type amino acid sequence resulted in an NLS (+) phenotype, we interpreted that the AAP2BR that was corrected has an NLS role. Likewise, we focused on the following nucleolar-excluded mutants harboring a pair of AAP2BR mutations; AAP2mt17, mt22 and mt27, to assess the NoLS role of each AAP2BR. In these 3 mutants, when correction of one AAP2BR mutation back to the wild-type amino acid sequence resulted in an NoLS (+) phenotype, we interpreted that the AAP2BR that was corrected has an NoLS role. The NLS role of a given pair of AAP2BRs were also assessed statistically as described below.

X-Gal staining. HEK 293 cells were seeded on a 6-well plate and transfected with 2 μg of either pAAV-CMV-lacZ or pAAV-CMV-lacZ-AAP2$^{144-184}$ plasmid DNA. Twenty-four hours after transfection, the cells were fixed with 2%

formaldehyde / 0.2% glutaraldehyde in phosphate-buffered saline (PBS) for 5 min, washed with PBS, and stained with 5mM $FeK_4(CN)_6$ / 5mM $FeK_3(CN)_6$ / 2mM $MgCl_2$ / 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (1mg/ml) in PBS at 37°C for 1 h.

AAV production and an ELISA specific for intact AAV2 particles. AAV2 VP3 only viral particles were produced by transcomplementation where VP3 and AAP2 (the wild type or mutant) were expressed in HEK 293 cells from two separate plasmids (303). In brief, HEK 293 cells were transfected with pAAV2-RepVP3, pCMV$_3$-FLAG-cmAAP2 (the wild type or mutant), and pHelper using PEI in 6 cm dishes. Forty-eight hours after transfection, the cells were washed in PBS and resuspended in 100 μl of-HEPES buffered-saline (pH 7.4). The cells were lysed by three cycles of freeze-thaw using a dry ice/ethanol bath and the supernatants were collected. The resulting cell lysates containing viral particles were then subjected to an A20 antibody-based intact AAV2 capsid-specific ELISA using the AAV2 Titration ELISA Kit (Progen, Heidelberg, Germany) per manufacturer's instructions.

Western blot analysis. HEK 293 cells were transfected with either the wild-type or mutant pCMV$_3$-FLAG-cmAAP2 plasmid DNA. Forty-eight hours post-transfection, the HEK 293 cells were lysed in radioimmunoprecipitation assay (RIPA) buffer containing protease inhibitors (Complete Mini, Roche, Indianapolis, IN). Protein concentrations in the cell lysates were determined by a

DC Protein Assay Kit (Bio-Rad, Hercules, CA). The same amount of total cell lysates (60 or 80 μg per lane) was separated on an 8% SDS-PAGE gel, transferred onto a PVDF membrane, reacted with mouse monoclonal anti-FLAG M2 antibody and monoclonal mouse anti-α tubulin antibody (sc-32293, Santa Cruz Biotechnology), followed by goat polyclonal anti-mouse IgG antibody conjugated to horseradish peroxidase (sc-2055, Santa Cruz Biotechnology). The signals on the blots were visualized and quantified using the FluorChem M system (ProteinSimple, Santa Clara, CA). The data were collected from a biologically duplicated set of experiments.

Statistical analyses. Differences in AAV viral particle production yields were statistically assessed by the two-tailed Welch's *t*-test. An unconditional exact test was used to statistically evaluate the association between the presence or absence of a given combination of 2 intact AAP2BRs and the presence or absence of an NLS using a 2 x 2 contingency table. For the purpose of this statistical analysis, the NLS (+) and NLS (-) mutants among AAP2mt10 to 27 were defined as those that showed exclusive nuclear accumulation and those that did not belong to the NLS (+) category, respectively. There were 9 NLS (+) and 9 NLS (-) AAP2BR mutants. The null hypothesis is that there is no association between AAV2BRs and NLS. Since we analyzed only 18 combinations out of all 25 possible combinations of AAP2BR mutations that leave at least 2 intact AAP2BRs (10 combinations of 2 intact BRs with 3 mutated BRs, 10 combinations of 3 intact BRs with 2 mutated BRs, and 5 combinations of 4 intact BRs with 1 mutated BR),

a range of P values are given in which all the possible outcomes from the 7 AAP2BR mutants that were not assessed are taken into account. Such P values that would be obtained if all the 25 AAP2BR mutants were analyzed are expressed as P_{25}. The P values obtained from the 18 AAP2BR mutants are expressed as P_{18}.

Results

The C-terminus of AAP2 contains both an NLS and an NoLS.

AAP2 has an amino acid stretch rich in basic amino acid residues near the C-terminus, AAP2$^{144\text{-}184}$ (**Fig. 2.1**). This region harbors a KRSR sequence that matches the classical monopartite NLS motif, K(K/R)X(K/R) (304), a putative NoLS, RRIK (305, 306), and shares features characteristic of NoLSs including a high proportion of basic amino acids and proximity to the C-terminus (307). It is also be possible that these two basic amino acid clusters might serve as a bipartite NLS with a long linker sequence (304, 308, 309). Therefore, we hypothesized that AAP2$^{144\text{-}184}$ contains both an NLS and an NoLS. To test this hypothesis, we transfected HeLa cells with the pCMV$_3$-GFP-AAP2$^{144\text{-}184}$ plasmid expressing the GFP fused with the AAP2$^{144\text{-}184}$ peptide at its C-terminus or pCMV$_3$-GFP, the control parental GFP plasmid devoid of the AAP2$^{144\text{-}184}$ peptide. Forty-eight hours post-transfection, the control GFP was observed diffusely throughout the cytoplasm and nucleoplasm, while the GFP-AAP2$^{144\text{-}184}$ fusion protein was localized predominantly to the nucleolus (**Fig. 2.2A and B**). Since GFP can diffuse into the nucleus without an NLS due to its small size (27 kDa), we also investigated whether the AAP2$^{144\text{-}184}$ peptide could target a large cytoplasmic protein, β-galactosidase (~120 kDa), to the nucleolus. Both immunofluorescence microscopic and cytochemical analyses demonstrated strong nucleolar enrichment of the otherwise predominantly cytosolic β-galactosidase when this enzyme was fused with the AAP2$^{144\text{-}184}$ peptide (**Fig. 2.2C and D**). Addition of the canonical NLS derived from SV40 large T-antigen (PKKKRKV) at the N-terminus of β-

galactosidase was not sufficient for nucleolar targeting showing a nucleolar exclusion pattern, although the protein accumulated in the nucleus (data now shown). These observations strongly supported our hypothesis that the AAP2$^{144-184}$ region contains both an NLS and an NoLS.

```
           BR1        BR2        BR3       BR4        BR5          N   No
        144                                             184
wt   ..KSKRSRRMTVRRRLPITLPARFRCLLTRSTSSRTSSARRIK..       +    +
mt 1 ..ASKRSRR.................................         +    +
mt 2 ..KSARSRR.................................         +    +
mt 3 ..KSKASRR.................................         +    +
mt 4 ..KSKRSAR.................................         +    +
mt 5 ..KSKRSRA.................................         +    +
mt 6 ..ASARSRR.................................         +    +
mt 7 ..KSAASRR.................................         +    +
mt 8 ..KSKASAR.................................         +    +
mt 9 ..KSKRSAA.................................         +    +
mt10 ..KSAASAR.................................         +    +
mt11 ...............................AAIA..               +    +
mt12 ..KSAASAR......................AAIA..                    Ex
mt13 ...........AAA..............................        +    +
mt14 ....................AFA.....................        +    +
mt15 ............................ASTSSA.........         +    +
mt16 ...........AAA......AFA....ASTSSA..........
mt17 ...........AAA.................AAIA..                    Ex
mt18 ....................AFA.......AAIA..                 +
mt19 ............................ASTSSA....AAIA..         +
mt20 ...........AAA......AFA....ASTSSA....AAIA..               Ex
mt21 ...........AAA......AFA.....................
mt22 ...........AAA............ASTSSA..........               Ex
mt23 ....................AFA....ASTSSA..........          +
mt24 ...........AAA......AFA........AAIA..                    Ex
mt25 ...........AAA............ASTSSA....AAIA..                Ex
mt26 ....................AFA....ASTSSA....AAIA..          +    Ex
mt27 ..KSAASAR...AAA............................               Ex
mt28 ....................___.....................        +    +
mt29 ...............................____..               +    +
mt30 ...........___................____..                     Ex
```

Figure 2.1 Amino acid sequences of the wild type (wt) and mutants (mt) of AAP2 near the C terminus and their roles as an NLS and/or NoLS. The sequence of amino acid positions from 144 to 184 in the wild-type AAP2 is shown at the top, followed by the sequences of a total of 30 AAP2 mutants with arginine/lysine-to-alanine substitutions (mt1 to mt27) or deletions (mt28 to mt30). The 5 BR clusters are indicated by lines above the wild-type sequence. The red and light-blue letters show basic amino acids and alanine mutations, respectively. The blue underlines indicate deletions. The dots in the sequences indicate residues that are the same as those of the wild type. The presence of a fully functional NLS (N) and/or NoLS (No), determined by immunofluorescence microscopy (Fig. 2.3), is indicated on the right. A plus in the N column indicates that the protein is observed exclusively in the nucleus regardless of its subnuclear localization. A plus in the No column indicates that the protein is strongly associated with the nucleolus. The mutants that showed a nucleolar-exclusion pattern are indicated by Ex.
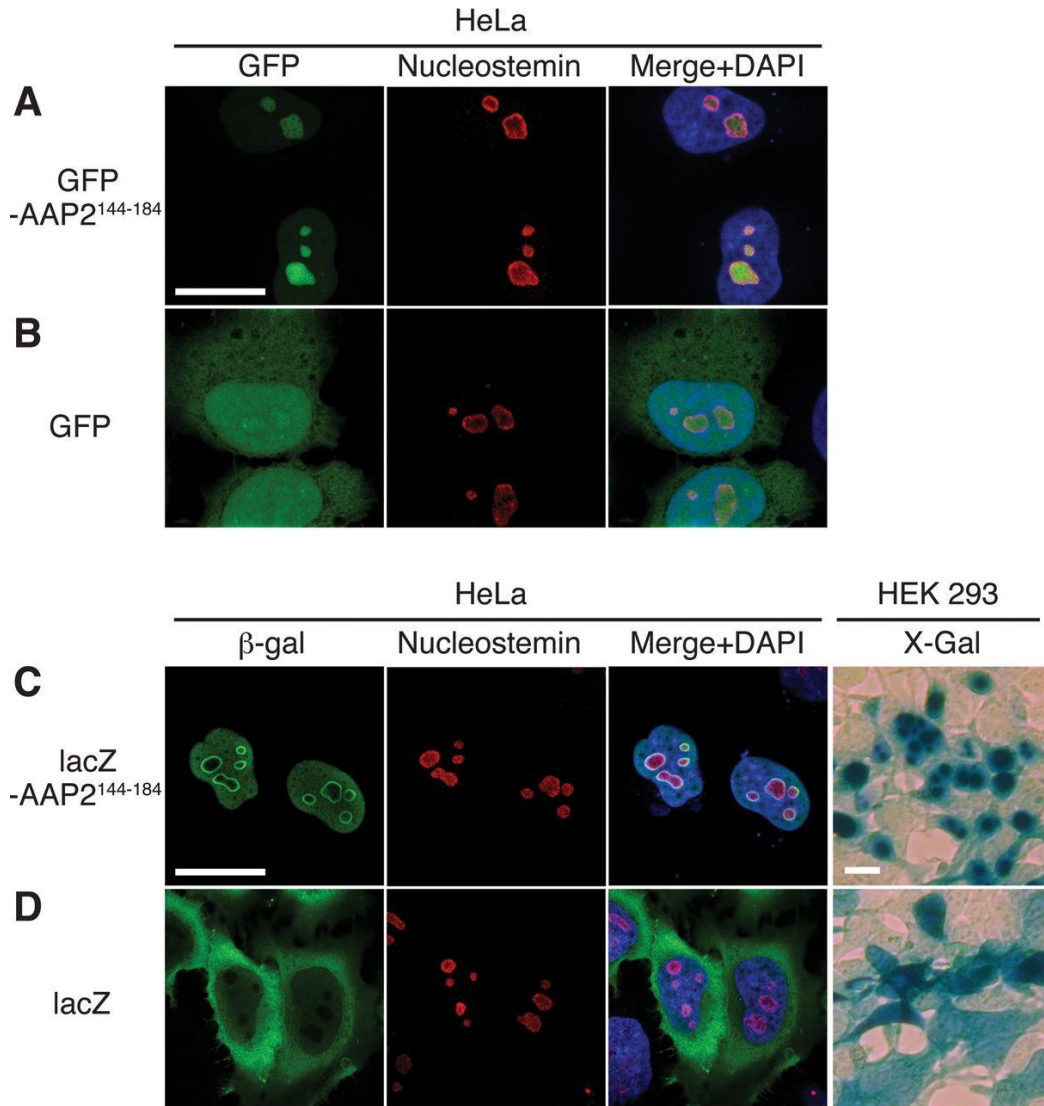
Figure 2.2 Intracellular localization of GFP or bacterial β-galactosidase fused with the AAP2$^{144-184}$ region at the C or N terminus, respectively.

HeLa or HEK 293 cells were transiently transfected with plasmid pCMV$_3$-GFP-AAP2$^{144-184}$ (**A**), pCMV$_3$-GFP (**B**), pAAV-CMV-lacZ-AAP2$^{144-184}$(**C**), or pAAV-CMV-lacZ (**D**), using PEI. pCMV$_3$-GFP-AAP2$^{144-184}$ and pAAV-CMV-lacZ-AAP2$^{144-184}$ are plasmids expressing the GFP-AAP2$^{144-184}$ fusion protein and the β-galactosidase–AAP2$^{144-184}$fusion protein, respectively, under the control of the CMV promoter. For HeLa cells, the signals were detected by immunofluorescence microscopy using corresponding antibodies, except for GFP, for which direct fluorescence was imaged. For HEK 293 cells, X-Gal staining was performed. Scale bars, 20 μm.

The FLAG-tagged wild-type AAP2 and nucleostemin co-localize in the nucleus and nuclear bodies.

It has been established that AU1-tagged AAP2 and fibrillarin, an endogenous nucleolar protein, co-localize in the nucleolus in HeLa cells (42). We first investigated whether the N-terminal FLAG-tagged AAP2 we used for this study also localizes in the nucleolus as expected. To this end, HeLa cells were transfected with pCMV$_3$-FLAG-AAP2 and immunostained with anti-FLAG and anti-nucleostemin antibodies. The FLAG-tagged wild-type AAP2 was found to localize exclusively to the nucleus with significant nucleolar enrichment (**Fig. 2.3A**). The AAP2 signals were also found tightly associated with nucleostemin-positive nuclear bodies (**Fig. 2.3A**). The AAP2 signals in the nucleolus detected by immunostaining were stronger in the periphery than in the center exhibiting a ring-shaped pattern with a central hollow. As discussed below, this nucleolar staining pattern is most likely an artifact caused by overexpression of a nucleolar-localizing protein (310). After trypsin treatment of the fixed cells, FLAG staining was visible throughout the nucleolus (**Fig. 2.4**). This observation confirmed that the FLAG peptide fused with AAP2 at the N-terminus does not interfere with its intracellular localization. The functional integrity of the FLAG-tagged AAP2 had been confirmed previously (303).
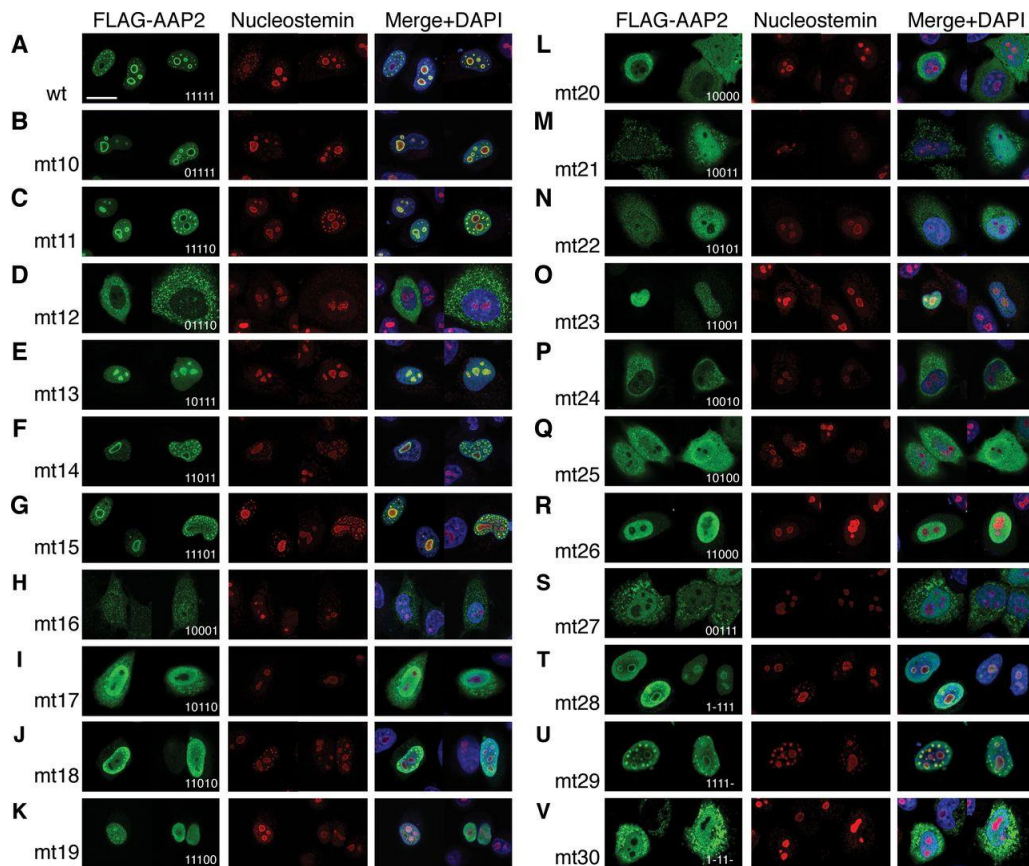
Figure 2.3 Intracellular localization of various AAP2 mutants with arginine/lysine-to-alanine substitutions or deletions within the BR clusters, AAP2BR1 to AAP2BR5. HeLa cells were transiently transfected with a plasmid expressing the wild-type or a mutant AAP2 with an N-terminal FLAG tag. The cells were fixed 48 hours post-transfection, immunostained with anti-FLAG antibody (green) and anti-nucleostemin antibody (red), and counterstained with DAPI (blue), before being imaged on a Zeiss LSM 710 confocal microscope with a 100× objective. Representative cell images are shown for the wild type and each mutant. (**A**) Wild-type AAP. (**B**) AAP2mt10, showing a staining pattern representing all the AAP2BR1 mutants, AAP2mt1 to -mt10 (the data for AAP2mt1 to -mt9 are not shown). (**C** to **V**) AAP2mt11 to -mt30. Formation of multiple nuclear bodies containing both AAP2 and nucleostemin was evident in many cells expressing the wild-type AAP2 (**A**) and the AAP2 mutants showing an intracellular localization pattern similar to that of the wild type (**B**, **C**, **F**, **G**, **T**, and **U**). The identity and the role of these speckled structures have yet to be determined. The 5-digit numbers in the bottom right corner of each FLAG-AAP2 panel indicate mutations introduced in each AAV2BR in each mutant (1, wild type; 0, alanine mutation; -, deletion). For example, 01110 (mt12) indicates BR1−/BR2+/BR3+/BR4+/BR5−. Scale bar, 20 μm.
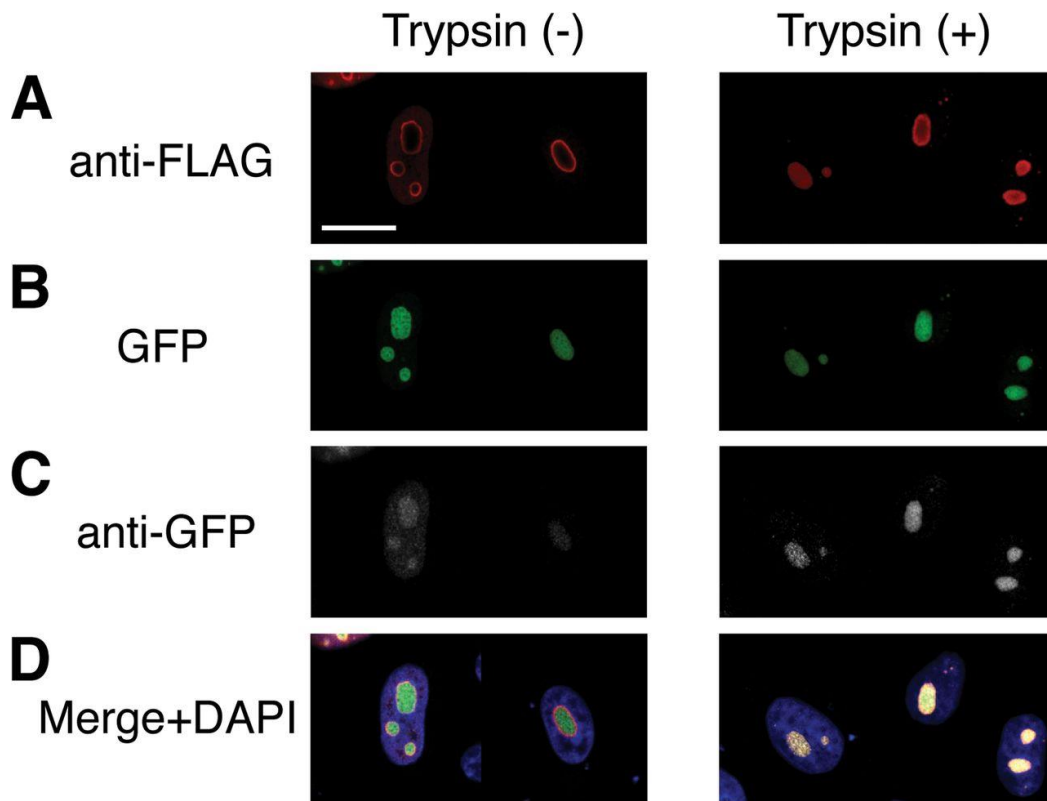
Figure 2.4 Immunofluorescence microscopic analysis of a FLAG-tagged AAP2 fused with GFP in HeLa cells with or without proteinase treatment. (**A** to **C**) HeLa cells were transiently transfected with pCMV$_3$-FLAG-cmAAP2-GFP expressing the wild-type full-length AAP2 fused with a FLAG tag and GFP. Intracellular localization of the fusion protein was analyzed with or without trypsin treatment by anti-FLAG antibody immunostaining (**A**), direct detection of GFP fluorescence (**B**), and anti-GFP antibody immunostaining (**C**) under a fluorescence microscope. (**D**) Merged images. Scale bar, 20 μm.

Basic amino acid clusters in the C-terminus of AAP2 constitute redundant multipartite NLSs and NoLSs.

The C-terminal region AAP2[144-184], where we found that the AAP2 NLS and NoLS reside, contains five basic amino acid-rich (BR) clusters separated by 3 to 7 amino acids: KSKRSRR (AAP2BR1), RRR (AAP2BR2), RFR (AAP2BR3), RSTSSR (AAP2BR4) and RRIK (AAP2BR5) from the N-terminus to the C-terminus (**Fig. 2.1**). To dissect the functional roles of these basic amino acids in the nuclear and nucleolar localization of AAP2, we constructed plasmids expressing a series of mutant AAP2 proteins by site-directed mutagenesis using the pCMV$_3$-FLAG-cmAAP2 plasmid (**Fig. 2.1**). Each mutant AAP2 was expressed in HeLa cells by plasmid DNA transfection and analyzed for its intracellular localization by immunofluorescence microscopy (**Fig. 2.3**).

We first created AAP2 mutants 1 to 12 (mt1 to 12 in **Fig. 2.1**) to investigate the role of AAP2BR1 and AAP2BR5, which carry the classical monopartite NLS motif, <u>KRSR</u>, and the putative NoLS motif, <u>RRIK</u>, respectively. Removal of positive charges from AAP2BR1 (AAP2mt1 to 10) or from AAP2BR5 (AAP2mt11) did not abolish the ability of mutant AAP2s to translocate to the nucleus with strong nucleolar enrichment and robust accumulation in the nucleostemin-positive nuclear bodies (**Fig. 2.3B and C**). Removal of positive charges from AAP2BR1 and AAP2BR5 in conjunction (AAP2mt12), however, resulted in substantial, if not complete, nuclear exclusion and nucleolar exclusion (**Fig. 2.3D**). This first set of experiments indicated that there are at least two NLSs that involve either AAP2BR1 or AAP2BR5 within AAP2[144-184]. We then created

an additional set of 18 AAP2 mutants 13 to 30 (mt13 to 30 in **Fig. 2.1**) to further investigate the role of each AAP2BR in nuclear and nucleolar localization. In these mutants, the positive charge in one or more AAP2BRs was diminished by alanine substitutions or deletions in various combinations. The 7 AAP2 mutants that had only 1 mutant AAP2BR and 4 intact AAP2BRs (mt10, 11, 13, 14, 15, 28 and 29) were all found exclusively in the nucleus with obviously enhanced association with the nucleolus (**Fig. 2.3B, C, E-G, T and U**). In keeping with this observation, GFP-AAP2$^{144-184}$ fusion mutants that had four intact AAP2BRs were observed exclusively in the nucleus with strong enrichment in the nucleolus (**Fig. 2.5A-C**). All the 13 mutants that had 3 or less intact AAP2BRs exhibited attenuated or complete loss of nucleolar association, and among them, 9 (mt16, 17, 20, 21, 22, 24, 25, 27 and 30) showed various degrees of impaired nuclear transport and cytoplasmic retention (**Fig. 2.3D, H-V**). AAP2 is a small protein that can passively cross the nuclear envelope through the nuclear pore complexes. The stronger cytoplasmic signals than the nuclear signals in AAP2mt12, 20 and 24 therefore might indicate the presence of a nuclear export signal (NES) in AAP2. However, we have not observed any nuclear-cytoplasmic shuttling of the AAP2 protein during the course of AAV2 particle production in HEK 293 cells by plasmid transfection, where AAP2 consistently remains located in the nucleolus or nucleostemin-positive nuclear bodies (Earley et al., unpublished observation). In addition, three different NES prediction programs (311-313) fail to identify any potential CRM1-binding NES in the AAP2 amino acid sequence. Thus, the above observations are most likely due primarily to various degrees of

impairment of nuclear import caused by different mutations although involvement

of an unidentified NES cannot be totally ruled out. Taken together, these

observations indicate that the NLS and NoLS in AAP2 are redundant, and various

combinations of 4 AAP2BRs out of 5 AAP2BRs within the AAP2$^{144\text{-}184}$ region

are able to constitute a functional multipartite NLS/NoLS.
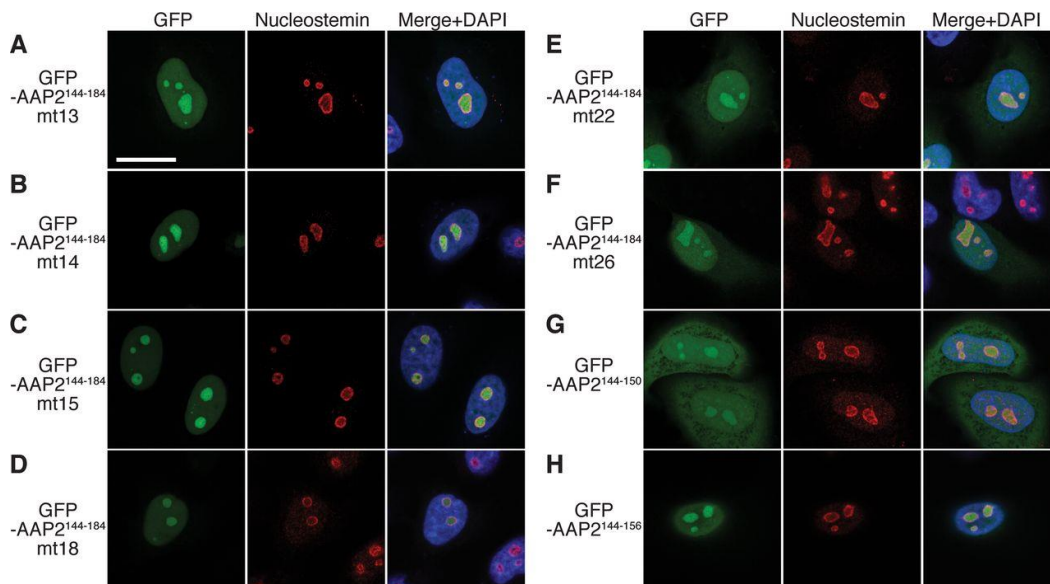
Figure 2.5 Microscopic assessment of the abilities of the AAP2BRs to target GFP to the nucleus and the nucleolus. The method used was the same as that for Fig. 2. (**A** to **F**) GFP-AAP2144–184mt13 to -mt26 are GFPs fused with the 41-amino-acid-long AAP144–184 region containing their corresponding AAP2BR mutations. (**G** and **H**) GFP-AAP2144-150 and GFP-AAP2144-156 are GFPs fused with AAP2BR1 only (7 amino acids) and AAP2BR1-AAP2BR2 only (13 amino acids), respectively. Scale bar, 20 µm.

Co-presence of AAP2BR1 and AAP2BR2 is an important but not absolute requirement for the AAP2 NLS.

The immunofluorescence microscopy analysis described above revealed that all the 10 AAP2BR mutants that exhibited impaired nuclear translocation and various degrees of cytoplasmic retention had either or both AAP2BR1 and AAP2BR2 mutations, and all the 10 combinatorial mutations involving AAP2BR1 or AAP2BR2 impaired nuclear transport of AAP2 to various degrees. A statistical comparison of various combinations of two intact AAP2BRs and the presence of a fully functional NLS in the AAP2 mutants revealed that the combination of AAP2BR1 and AAP2BR2 is strongly associated with nuclear localization (**Table 1**). The GFP-fusion experiment also demonstrated that AAP2BR1 by itself is not sufficient for active nuclear transport and co-presence of AAP2BR1 and AAP2BR2 is required for nuclear accumulation (**Fig. 2.5D-H**). When we assessed the NLS role of each AAP2BR based on its ability to restore a fully functional NLS to AAP2 mutants (see Materials and Methods), we found that all the 5 AAP2BRs can contribute to the facilitation of nuclear localization of AAP2. These observations indicate that, although all the 5 AAP2BRs have an NLS role, the presence of AAP2BR1 and AAP2BR2 in conjunction plays the most influential role in nuclear targeting. Due to the redundant nature of the AAP2 NLS, either one of these two AAP2BRs is dispensable if all the other 4 AAP2BRs are present, as evidenced by the nuclear localization of AAP2mt10 and mt13.

TABLE 1

Associations between various combinations of two AAP2BRs and the presence of an NLS in AAP2[144–184]

AAP2 BR	Association with [no./total (%)][a]:			
	AAP2BR1	AAP2BR2	AAP2BR3	AAP2BR4
AAP2 BR2	7/7 (100) (P_{18} = 0.002; P_{25} =0.001–0.017)			
AAP2 BR3	4/7 (57) (P_{18} = 1.000; P_{25} = 0.186–1.000)	4/5 (80)[b] (P_{18} = 0.211; P_{25} = 0.014–1.000)		
AAP2 BR4	4/7 (57) (P_{18} = 1.000; P_{25} = 0.186–1.000)	4/5 (80)[b] (P_{18} = 0.211; P_{25} = 0.014–1.000)	3/6 (50) (P_{18} = 1.000; P_{25} =0.317–1.000)	
AAP2 BR5	4/7 (57) (P_{18} = 1.000; P_{25} = 0.186–1.000)	4/4(100)[b] (P_{18} = 0.031; P_{25} =0.002–1.000)	3/5 (60) (P_{18} = 0.696; P_{25} =0.113–1.000)	3/5 (60) (P_{18} = 0.696; P_{25} =0.113–1.000)

[a]Eighteen AAP2 mutants (AAP2mt10 to -mt27) were categorized into two groups according to the presence or absence of an NLS as defined in the legend to Fig 2.1. There were 9 NLS (+) and 9 NLS (−) AAP2 mutants. An unconditional exact test was performed to statistically evaluate the association between the presence or absence of a given combination of 2 AAP2BRs and the presence or absence of an NLS. The null hypothesis is that there is no association between AAV2BRs and NLSs. Two types of P values (P_{18} and P_{25}) are provided, as described in Materials and Methods. Shown are the frequencies of NLS (+) AAP2 mutants among all the mutants that carry a given combination of two different AAP2BRs.

[b]The statistical analysis for the combination does not give a conclusive result due to the wide range of P_{25} values across the statistically significant cutoff value of 0.05.

All the AAP2BRs can contribute to nucleolar localization and function in a context-dependent manner. The NoLS role of each AAP2BR was investigated based on the functional restoration of nucleolar-excluded AAP2 mutants, as detailed in the Materials and Methods. This analysis revealed an obvious NoLS role in AAP2BR1, 2, 4, and 5. For example, the NoLS role of AAP2BR2 could be demonstrated by the observation that, when the mutated AAP2BR2 was corrected back to the wild-type sequence in the nucleolar-excluded AAP2mt17 (**Fig. 2.3I**),

nucleolar retention was restored as evidenced with AAP2mt11 (**Fig. 2.3C**). A comparison between AAP2mt19 showing modest nucleolar accumulation and AAP2mt26 showing nucleolar exclusion, which differ only by the presence or absence of AAP2BR3, revealed that AAP2BR3 also functions as a weak NoLS (**Fig. 2.3K and R**). These results indicate that the effective nucleolar accumulation is a consequence of the cooperation of the redundant NoLSs residing in the five AAP2BRs in a given combination of AAP2BRs. Interestingly, AAP2mt26 (AAP2BR1+/BR2+/BR3-/BR4-/BR5-) exhibited a nearly complete nuclear localization with nucleolar exclusion (**Fig. 2.3R**), while the GFP-AAP2$^{144-184}$mt26 carrying an AAP2$^{144-184}$ with the same AAP2BR mutations and the GFP-AAP2$^{144-156}$, in which GFP was fused with a 13-mer peptide containing only AAP2BR1 and AAP2BR2, showed nuclear and nucleolar accumulation (**Fig. 2.5F and H**). The same discrepancy in the pattern of nucleolar accumulation was observed in AAP2mt22 (**Fig. 2.3N**) and GFP-AAP2$^{144-184}$mt22 (**Fig. 2.5E**). It has previously been shown that adding a nonapeptide containing 9 basic amino acids to the C-terminus of GFP allows for nucleolar accumulation of this protein not through a specific NoLS motif, but presumably through non-specific, electrostatic interactions with negatively charged nucleolar components (300). There are 8 basic amino acids within the AAP2BR1/AAP2BR2 segment and there are 10 basic residues in AAP2mt22 (**Fig. 2.1**). Therefore, the extra positive electric charges added to GFP likely explains why the GFP-AAP2$^{144-156}$, GFP-AAP2$^{144-184}$mt22 and GFP-AAP2$^{144-184}$mt26 fusion proteins were still capable of nucleolar retention while AAP2mt26 was not. Taken together, although there is no

inconsistency in the NLS role of AAP2BR1 and AAP2BR2 between the AAP2 mutants and the GFP fusion proteins, their NoLS role is context dependent.

Capsid assembly-promoting functions of AAP2BR mutants.

A series of studies have demonstrated that nucleolar localization of viral components is important for the AAV life cycle (134, 148, 151, 284). Indispensable roles of the nucleolus in the virus life cycle have also been demonstrated in other viruses (282, 314, 315). Having created a panel of AAP2 mutants whose intracellular localizations are characterized, we addressed how aberrant localization of AAP2 might affect its assembly-promoting function. To this end, we performed a transcomplementation assay in which AAV2 VP3 only particles were produced in HEK 293 cells in the presence of the wild-type AAP2 or a series of AAP2 mutants, and quantified AAV2 viral particle yields using an AAV2 intact particle-specific ELISA. There was a strong correlation between aberrant intracellular localization of AAP2 and attenuated AAV2 capsid production (**Fig. 2.6A**). In particular, AAP2 mutants with nucleolar exclusion consistently showed a 3-log or more fold reduction of capsid production compared to the wild type. Interestingly, AAP2mt8, 9, 10, and 13, which showed exclusive nuclear localization with enhanced nucleolar association similar to the wild type, exhibited a more than 20 to 100-fold decrease in capsid production (**Fig. 2.6A**). Thus, normal intracellular localization of AAP2 is prerequisite but is not sufficient for effective capsid production. The impaired function observed in AAP2 mutants showing wild type-like intracellular localization implies that

AAP2BRs are not only the redundant organelle-targeting signals, but also play a direct or indirect role in the capsid assembly process, which we address in the following sections.
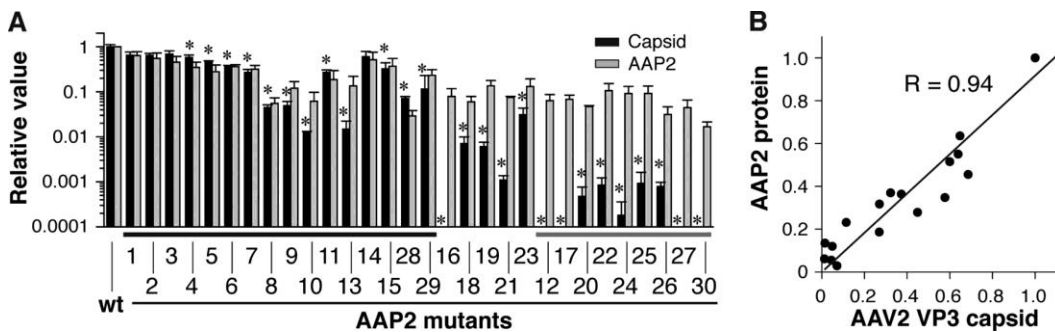
Figure 2.6 AAV2 VP3 capsid production titers and wild-type and mutant AAP2 protein expression levels in HEK 293 cells. HEK 293 cells seeded on 6-cm dishes were transfected with pAAV2-RepVP3, pCMV$_3$-FLAG-cmAAP2 expressing either the full-length wild-type AAP2 or a full-length AAP2 mutant with an AAP2BR mutation(s), and pHelper, using PEI (12). The cells were harvested at 48 hours post-transfection and made into a 100-μl crude cell lysate in HEPES-buffered saline after three cycles of freezing and thawing. The AAV2 particle concentration in each crude lysate was quantified by an A20 antibody-based ELISA. (**A**) Viral titers relative to the titer obtained with the wild-type AAP2 (black bars). A relative titer of 1.0 corresponds to 7.2×1012 particles/ml. The thick horizontal black line between the graph and the x axis labels indicates the wild type and AAP2 mutants that showed exclusive nuclear localization with enhanced nucleolar association. The thick horizontal gray line indicates the mutants showing nucleolar exclusion. The data were collected from biologically triplicate experiments. *, $P < 0.05$ (two-tailed Welch's t test) compared to the wild-type values. In a separate experiment, HEK 293 cells seeded on 6-cm dishes were transfected with pCMV$_3$-FLAG-cmAAP2 (the wild type or mutant) using PEI. The cells were harvested 48 hours post-transfection, and mutant AAP2 protein expression levels relative to the level of the wild type were determined by a quantitative western blot analysis using biologically duplicate samples (gray bars). The error bars represent standard errors of the mean (SEM) (black bars) or the difference between each value and the mean value (gray bars). (**B**) Correlations between AAP2 protein levels of the wild type and mutants indicated by the thick horizontal black line in panel A and AAV2 VP3 capsid production titers are shown in a scatter plot. All values are relative to the levels of the wild type.

AAP2BRs maintain AAP2 expression levels for effective capsid assembly.

During the microscopic analysis, we had noticed that the number of transfected HeLa cells and signal intensity in many AAP2 mutants were lower than those from the wild type to various degrees. Since AAP2BR mutations may have affected not only intracellular localization but also protein expression levels, we performed a quantitative western blot assay on lysates from HEK 293 cells transfected with either the wild-type or mutant $pCMV_3$-FLAG-cmAAP2 plasmids. The molecular weight (MW) of the FLAG-tagged AAP2 protein was determined as 27 kDa by SDS-PAGE (26 kDa for AAP2), which is slightly higher than a calculated molecular weight of 24 kDa. Mutation or deletion of the AAP2BR regions decreased AAP2 protein expression levels to various degrees between 0.02 and 0.64 compared to that of 1.0 of the wild type (**Figs. 2.6A and 2.7**). Expression levels of all the mutants showing aberrant localization were reduced substantially with the relative levels ranging from 0.02 to 0.14. The reduction of protein levels was particularly pronounced in the nucleolar-excluded mutants. The AAP2BR1 or AAP2BR2 mutants showing localization similar to that of the wild type had greater quantities of protein on average than the mutants showing aberrant localization; however, they also showed attenuated AAP2 expression ranging from 0.03 to 0.64. Importantly, the results showed a very nice correlation between intracellular AAP2 levels, nucleolar accumulation of AAP2 and AAV capsid production. When AAP2 mutants retained the ability to localize normally (such mutants are indicated with a horizontal black line in **Fig. 2.6A**), AAV capsid yields were positively correlated to the AAP2 mutant protein levels

with Pearson's correlation coefficient of 0.94 (**Fig. 2.6B**). When AAP2 mutants were excluded from the nucleolus, AAV capsid yields were disproportionately lower than the AAP2 mutant protein levels (such mutants are indicated with a horizontal gray line in **Fig. 2.6A**). Although the underlying mechanism has yet to be elucidated, these results indicate that AAP2BRs play a pivotal role in maintaining the AAP2 levels for effective capsid assembly. The AAP2BRs' role in this aspect well explains why some of the AAP2BR1 or AAP2BR2 mutants showing the wild-type localization failed to produce capsids effectively.
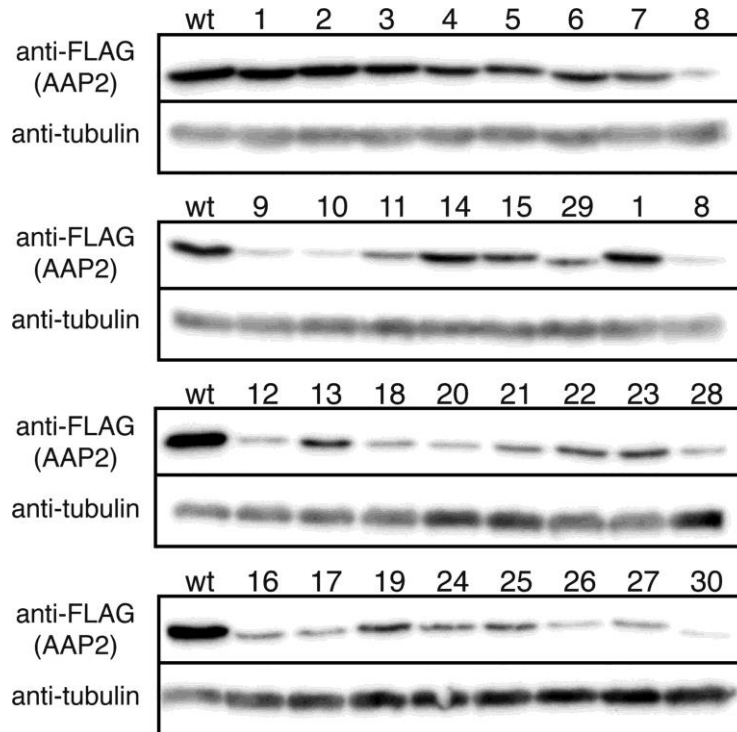
Figure 2.7 Western blot analysis of the wild-type and mutant AAP2 proteins expressed in HEK 293 cells.
Cells seeded on 6-cm dishes were transfected with pCMV3-FLAG-cmAAP2 (wild type or mutant) using PEI. The same quantity of crude lysates prepared at 48 hours post-transfection were separated on an 8% SDS-PAGE gel and blotted onto a membrane, which was then probed with both anti-FLAG and anti-α-tubulin antibodies. The AAP2 mutant numbers are shown above the lanes.

The AAP2 NoLS is important for nucleolar accumulation of AAV2 VP3.

The data presented above provided insights into the AAP2BRs' functional roles in intracellular localization of AAP2 and capsid assembly; however the data did not address the AAP2 role as a VP3 transporter. Since the nucleolus has been thought to be important for the AAV2 capsid assembly, and nucleolar-excluded AAP2BR mutants showed a substantially reduced ability to produce viral capsids, it is possible that AAP2BR mutants would not associate with VP3 leading to impaired nuclear transport and/or failure to accumulate VP3 in the nucleolus even if VP3 is actively localized to the nucleus. To address this, HeLa cells were transfected with $pCMV_1$-AAV2VP3 and $pCMV_3$-FLAG-cmAAP2 (either the wild type or AAP2mt26) and intracellular localization of AAP2 and VP3 was analyzed by immunofluorescence microcopy at 48 hours post-transfection. As Sonntag et al. have previously shown, VP3 diffusely localizes to the cytoplasm and nucleoplasm, but remains outside the nucleus unless AAP2 is present (42) (**Fig. 2.8A, B**). In the presence of AAP2mt26, which had the ability to exclusively translocate to the nucleus but was not able to accumulate in the nucleolus, VP3 was seen predominantly in the nucleoplasm, co-localizing with AAP2mt26, and leaving only a small quantity of VP3 proteins in the cytoplasm. Importantly VP3 failed to accumulate in the nucleolus (**Fig. 2.8C**). We also have found that, using a plasmid carrying an AAV2 cap gene that does not express AAP2, expression of the AAV2 VP1, VP2, and VP3 proteins does not result in nucleolar localization of any of the VP proteins in HeLa cells when AAP2 is not co-expressed (Earley et al., an unpublished observation). Taken together, these observations indicate that

the NoLS in AAP2 plays a primary role in targeting capsid VP proteins to the nucleolus.
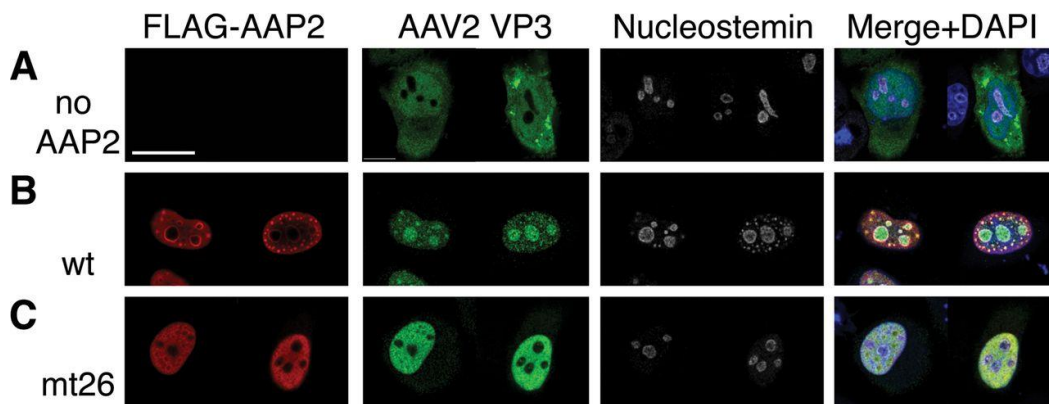
Figure 2.8 Intracellular localization of AAV2 VP3 in the presence or absence of wild-type AAP2 or in the presence of the nuclear-targeted, nucleolar-excluded AAP2 mutant, AAP2mt26. HeLa cells were transfected with the following plasmids: $pCMV_1$-AAV2VP3 only (**A**), $pCMV_1$-AAV2VP3 and $pCMV_3$-FLAG-cmAAP2 (wild type) (**B**), and $pCMV_1$-AAV2VP3 and $pCMV_3$-FLAG-cmAAP2mt26 (**C**). The cells were fixed at 48 hours post-transfection and immunostained with anti-FLAG, anti-AAV VP1/VP2/VP3, and anti-nucleostemin antibodies. Scale bar, 20 μm.

Discussion

The NLS in AAP2 had remained unidentified in a series of studies reported from Kleinschmidt's group (42-44), which prompted us to seek to determine amino acid residues responsible for nuclear translocation of AAP2. The consensus sequences of the classical, importin-α-binding NLSs have been well defined. The monopartite NLSs have the consensus sequence, K(K/R)X(K/R), while the bipartite NLSs show a consensus sequence composed of two short basic amino acid clusters separated by a linker sequence such as $(K/R)(K/R)X_{10-12}(K/R)_{3/5}$ where the C-terminal side cluster contains at least three basic amino acids within 5 consecutive residues (304, 308, 309). Many non-canonical, importin-α-binding NLSs have also been identified including bipartite NLSs with a long linker of up to 29 residues (304, 308, 309). In light of this knowledge about various types of NLSs, one could assume that AAP2 has one classical monopartite NLS in AAP2BR1 (KS<u>KRSR</u>R) and/or one non-canonical bipartite NLS with a long linker in AAP2BR1-X_{30}-AAP2BR5 (<u>RRIK</u>). Our experimental data, however, do not support this assumption. None of the AAP2BRs were capable of serving as a monopartite NLS. Substantial nuclear enrichment required the presence of AAP2BR1/AAP2BR2 (KSKRSRR-X_3-RRR) or at least four AAP2BRs when either AAP2BR1 or AAP2BR2 is absent (i.e., AAP2BP1/AAP2BP3/AAP2BP4/AAP2BP5 and AAP2BP2/AAP2BP3/AAP2BP4/AAP2BP5). Thus, we identified at least three different motifs that can function as an NLS and overlap within the AAP2[144-184] region. This redundant nature of the AAP2 NLS explains why a pervious

investigation concluded that the AAP2BR1 was not involved in nuclear localization even though the region contained a classical NLS (44).

In nucleolar localizing proteins, an NLS and an NoLS can reside in separate locations, but they often co-reside in a region highly rich in lysine and arginine residues. The NLS and NoLS of AAP2 that we identified is a joint NLS/NoLS because all the AAP2BR regions play either, or both, of the NLS and NoLS roles. Many such joint NLS/NoLSs have been reported as NLSs in literature mainly due to incomplete investigation for nucleolar accumulation; and therefore joint NLS/NoLSs have been overlooked in many instances (307). A recent attempt to extract commonalities of 48 human NoLSs including joint NLS/NoLSs has revealed that NoLSs are highly basic, predominantly located near the N- or C-terminus of proteins, in α-helices or coils and rarely in β-strands, and are solvent accessible (307). In this regard, the AAP2$^{144-184}$ region containing the joint NLS/NoLSs is consistent with these general trends. That is, in addition to the high proportion of basic amino acids, the NLS/NoLS resides near the C-terminus; the AAP2$^{144-184}$ residues are exclusively in α-helices (49%) or coils (51%) according to the Jpred3 prediction (316); and proportions of buried residues in the AAP2$^{144-184}$ region predicted by Jpred3 are 37%, 10%, and 2% at prediction thresholds of 25%, 5%, and 0%, respectively, which are lower than the full-length protein control in the study reported by Scott et al. (307) The AAP2$^{144-184}$ region also contains the previously identified (K/R)(K/R)X(K/R) NoLS motif (305). Through literature search, we could identify a joint NLS/NoLSs that resembles the AAP2

NLS/NoLS. Liu et al. reported that a stretch of 18 amino acid residues, $_{2744}$KKKMKKHKNKSEAKKRKI$_{2761}$, is a part of a joint NLS/NoLS found in 1A6/downregulated in metastasis (1A6/DRIM), a large nucleolar targeting protein (2785 amino acids) involved in ribosomal RNA processing (317). This amino acid sequence is reminiscent of $_{154}$RRRLPITLPARFRCLLTRSTSSRTSSARRIK$_{184}$, one of the complete joint NLS/NoLSs that we identified in AAP2. Although the observations obtained from the arginine/lysine-to-alanine mutagenesis experiments were not exactly the same between Liu et al.'s and our studies, both identified basic amino acid clusters that play a dual NLS/NoLS role. In addition, both studies showed that a decrease in electric charge in the signals resulted in nucleolar exclusion. As demonstrated with AAP2 and many other nucleolar localizing proteins (307), NLSs and NoLSs are inherently difficult to clearly define particularly when they reside within the same region with a portion of the signal playing a dual role.

How NoLSs exert their role in nucleolar targeting has not been fully understood. Unlike NLSs, NoLSs do not have well defined consensus sequences, although they share common characteristics to some degree as described above. The nucleolus does not have any compartmentalizing membranous structures and therefore nucleolar accumulation presumably does not require sophisticated machinery that actively targets this organelle like the nuclear transport machinery. Rather, nucleolar accumulation of proteins has been thought to be mediated by a retention mechanism through interactions with nucleolar components, such as

RNA and other nucleolar proteins (318). RNA binding motifs have been identified in many nucleolar proteins, including nucleolin, fibrillarin and viral RNA-binding proteins (319-321). Nucleophosmin, a putative shuttle protein that transports cargos between the cytoplasm and the nucleolus (322-324), has been shown to mediate nucleolar retention of other proteins through the interaction between two highly acidic regions ($_{120}$EDAESEDEEEED$_{132}$ and $_{161}$DEDDDDDDEEDDDEDDDDDDFDDEEAEE$_{188}$) in nucleophosmin and positively charged amino acid regions in the other (325). Thus, nucleolar accumulation is most likely mediated by electrostatic interaction rather than by a defined set of particular amino acid sequences (300). Concordant with this notion, we could only vaguely define the AAP2 NoLS, as opposed to its NLSs and found that the abundance of positively charged AAP2BR regions, rather than a specific amino acid sequence or a specific combination of AAP2BRs, confers AAP2 with the ability to translocate from the nucleoplasm to the nucleolus. This is in line with the notion that the nucleolar accumulation of AAP2 results from non-specific interaction with negatively charged nucleolar components and also raises a possibility that the degree of accumulation is determined by the net electric charge of the AAP2$^{144-184}$ region. The electric charge of AAP2$^{144-184}$ is 15, and at least 12 are found to be required for efficient nucleolar accumulation of the AAP2BR mutants (**Fig. 2.1**). The net electric charges of the AAP2$^{144-184}$-corresponding regions of various serotype AAPs (AAP1-13) (44) vary from 8 to 15, with AAP5 being the lowest and AAP2 and AAP3B being the highest. Assuming that nucleolar accumulation is determined by the abundance of the net positive

charges within this region, AAP5 would likely be devoid of an NoLS and excluded from the nucleolus. If this is true, AAP5 would not be able to transport AAV5 capsid VP proteins to the nucleolus, the organelle that is believed to be important for AAV capsid assembly, at least for AAV2 (134, 148). How would AAV5 viral capsid proteins assemble if they are not transported to the nucleolus? Although our study demonstrates that the nucleolar localization of AAP2 strongly correlates with the ability for AAV2 to form capsids, it is intriguing to speculate that the nucleolar accumulation of AAP2 might merely be a coincidental, non-specific consequence associated with a function(s) other than the nucleolar targeting and this targeting might not be functionally important. Such non-specific nucleolar accumulation with no functional role has been demonstrated in the human histone H2B protein (300). Future studies on the subcellular localizations of AAPs derived from various serotypes will address this potential paradox and elucidate the biological significance of the presence of an NoLS in AAPs. Nonetheless, our observations corroborate the driving role of the electric charge in nucleolar targeting and support an idea that NoLSs pay a lower cost for the acquisition of their role by intermingling their signals with NLSs that already have high electric charges. This is particularly advantageous to small viruses that have a limited capacity to store genetic information.

Surprisingly, we could identify at least three complete NLSs that overlap within the same AAP2$^{144-184}$ region. They are AAP2BR1+BR2, AAP2BR1+BR3+BR4+BR5, and AAP2BR2+ BR3+BR4+BR5. Redundant

NLSs, either as separate NLSs or overlapping NLSs, have been identified in many proteins including those derived from viruses. For example, the polyomavirus large T antigen has two separate NLSs, VSRKRPR and PKKARED, each of which mediates nuclear localization in the absence of the other (326, 327); the nonstructural protein 1 of influenza A virus (NS1A) has a functionally competent monopartite NLS, DRLRRDQKSLR, and a classical bipartite NLS-like sequence, KRKMARTARSKVRRDKMAD (328), which also conveys nucleolar localization; and bovine adenovirus-3 33K protein has a 40-residue long region where two karyopherin-binding motifs, an importin-α5-binding motif and a transportin 3-binding motif, overlap (329). Utilization of different nuclear transport pathways through the interactions between an NLS and multiple karyopherins have also been demonstrated in human immunodeficiency virus (HIV) Rev protein (330). The straightforward inference of the NLS redundancy is that it ensures and promotes efficient nuclear entry by increasing the avidity to the nuclear transport system and/or by using more than one nuclear entry pathways. However, there are also many instances in which the amino acid residues composing an NLS exert functions distinct from those in nuclear translocation. For example, the unconventional NLS in the nucleoprotein (NP) of the influenza A virus plays a crucial role in nuclear import of the viral genomic RNA (331) and the second NLS in the transcription factor Fli-1 also functions as a DNA binding domain (332). In fact, DNA-binding domains frequently overlap with NLSs (332) and this has led to speculation about the possible co-evolutionary origins of these two motifs (333, 334). In the case of AAP2, our present and previous studies

provide clues to understanding the necessity of redundancy. In our previous study, we used a directed evolution approach to experimentally and computationally evolve functionally competent mutant AAP2s from a randomly mutagenized library (303). We demonstrated that the most functionally optimal AAP2 mutants harbor a strongly basic and hydrophilic heptapeptide, $(R/S)_7$, in the AAP2BR1 region, which remarkably resembles the native sequence, KSKRSRR (303). Since all the AAP2 mutants used in this previous study had a completely functional NLS/NoLS composed of AAP2BR2, 3, 4 and 5, AAP2BR1 is most likely not a simply redundant sequence composing an NLS/NoLS but has an important non-NLS/NoLS role. In keeping with this notion, mutations in this region could substantially reduce AAP2 protein levels without affecting its intracellular localization, and result in significantly impaired capsid assembly (AAP2mt8, 9 and 10). The reduction of capsid production by an AAP2BR1 mutation has also been reported by Naumer et al. (44) Although the mechanism for the decreased protein expression has yet to be determined, it could be either due to intrinsic instability caused by protein misfolding or due to decreased protein-protein or protein-nucleic acid interactions that stabilize AAP2. Taken together, the redundancy of the NLS/NoLS in AAP2 is likely a necessity for optimal capsid production, and presumably results from their alternative functional roles.

In summary, this study identifies a joint NLS/NoLS near the C-terminus of AAP2. The identified basic amino acid region is not only important for nuclear and nucleolar localization but also for maintaining AAP2 levels, indicating a non-

NLS/NoLS role in this region. The system described here can readily be applied to AAPs derived from other serotypes. Further studies on interactions between AAP, capsid, and cellular proteins will help understand the mechanism of AAV capsid assembly and may lead to improved AAV vector production for gene therapy.

Acknowledgments

Chapter 3: Adeno-Associated Virus Assembly-Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5 and 11

Lauriel F. Earley[1], John M. Powers[2], Kei Adachi[2], Joshua Baumgart[2], Nancy L. Meyer[3], Qing Xie[3], Michael S. Chapman[3], and Hiroyuki Nakai[2,4]

Departments of [1]Molecular Microbiology & Immunology, [2]Molecular & Medical Genetics, and [3]Biochemistry & Molecular Biology, Oregon Health & Science University School of Medicine, Portland, Oregon 97239; [4]Division of Neuroscience, Oregon National Primate Research Center, Beaverton, Oregon 97006, United States of America.

For this manuscript, LE, JMP, JB, and HN made the plasmids. LE performed the immunofluorescence microscopy, western blots, ELISAs, produced and purified the viruses for the infection assays. JB assisted with the immunofluorescence microscopy. JMP performed the VP3 – AAP cross complementation study. KA produced and purified the viruses for electron microscopy. QX and NLM performed the electron microscopy. MSC provided expertise and advise. LE, JMP, and HN designed the experiments. LE, HN, JMP, and KA wrote the manuscript.

Abstract

Adeno-associated virus (AAV) not only offers a promising gene therapy vector but also serves as a model that helps understand functional and structural biology of small icosahedral viruses. In this regard, the recent discovery of the assembly-activating proteins (AAPs), a novel class of AAV non-structural proteins, shed new light on the mechanism of AAV capsid assembly. Previous studies have shown AAP to be essential for assembly of AAV1, 2, 5, 8, and 9 capsids; however, its mechanistic role in the viral life cycle remains uncharacterized. Here, we present comprehensive functional studies on AAPs from AAV serotypes 1 to 12 (AAP1 to 12, respectively). Subcellular localizations of AAP1 to 12 and assembly of AAV2, 4, 5, 8 and 9 capsids were examined by immunofluorescence microscopy. AAP's ability for cross-complementation among the 12 serotypes was investigated by the Illumina sequencing-based Barcode-Seq and quantitative dot blot analysis. Transmission electron microscopy was performed to confirm capsid assembly. These analyses revealed that AAP1 to 12 are all localized in the nucleus with serotype-specific contrastive patterns of nucleolar association; AAPs and assembled capsids do not necessarily co-localize; with the exception of AAP4, 5, 11 and 12, AAPs are promiscuous in promoting capsid assembly of other serotypes; assembled AAV5, 8 and 9 capsids are excluded from the nucleolus in contrast to nucleolar enrichment of assembled AAV2 capsids; and surprisingly, AAV4, 5 and 11 capsids are not dependent on AAP for assembly. Thus, both nucleolar-localized capsid assembly and the strict requirement for AAP found in AAV2, does not extend to all AAV serotypes.

Author Summary

Adeno-associated virus (AAV) is a small, non-enveloped, icosahedral DNA virus that belongs to the parvovirus family. Even though AAV is a popular and successful vector for gene therapy, many fundamental aspects of AAV biology remain unanswered, including the process and proteins involved in capsid assembly. The recent discovery of a new class of AAV proteins, the assembly-activating protein (AAP), has provided new opportunities for research in this area. Previous studies on AAV serotype 2, the canonically studied AAV, show that assembly takes place in the nucleolus and is dependent upon AAP. Here, we significantly expand the current knowledge about AAP in capsid assembly by characterizing the AAPs from AAV serotypes 1 to 12 for their sub-cellular localization, association with assembled capsids, and ability to promote assembly of heterologous AAV serotypes. We find that AAV serotypes 4, 5, and 11 are capable of AAP-independent assembly, that nucleolar-localized capsid assembly is not a general characteristic of AAVs, and that many AAPs are capable of promoting assembly in a wide range of heterologous AAV serotypes. These findings challenge widely held beliefs about the AAP-dependent and nucleolus-associated AAV capsid assembly and show the heterogeneous nature of the assembly process within the AAV family.

Introduction

Capsid assembly of icosahedral viruses has been an important area of research with an impact on multiple fields. Foremost is the basic biology behind how

pathogenic viral proteins hijack the host cell to aid in their assembly and how viral capsid proteins fit together with the ultimate end of devising anti-viral therapies (335). While there are a multitude of morphologies viruses can take, icosahedral symmetry is found frequently in a wide number of viral families and may constitute roughly a half of the known forms of viral capsids (336). A model icosahedral virus with T=1 symmetry (119) is adeno-associated virus (AAV), a parvovirus that belongs to the genus Dependoparvovirus of the family Parvoviridae. AAV has recently become a well-regarded vector for in vivo gene therapy with successful clinical trials for hemophilia B (227), lipoprotein lipase deficiency (233), Leber congenital amaurosis (228, 229), among others (reviewed in Mingozzi and High (235), thus making the study of its capsid assembly an attractive pursuit for both gene therapy applications and for furthering our knowledge of parvovirus biology.

AAV is a small non-enveloped virus with a single-stranded DNA genome of 4.7 kb containing two genes, *rep* and *cap*, between two inverted terminal repeats. The rep gene produces non-structural Rep proteins essential for viral genome replication and packaging. The cap gene produces the three structural proteins, VP1, VP2, and VP3, translated from different start codons in a single open reading frame (ORF). Alternative mRNA splicing and a combined use of ATG and an alternative start codon for initiation of VP protein translation leads to the appropriate capsid stoichiometry, a VP1:VP2:VP3 ratio of approximately 1:1:10 (89, 337). It had long been believed that the AAV viral genome encodes only the

Rep and VP proteins until 2010, when a second +1 frame-shifted ORF that encodes a 204 amino acid-long non-structural protein was identified within the cap gene of AAV serotype 2 (AAV2) (42). This new AAV protein has been named assembly-activating protein (AAP) after the role it plays in capsid assembly (42).

The AAP ORFs have been found in all the parvoviruses that belong to the genus Dependoparvovirus (42, 43), and among them AAP derived from AAV2 (i.e., AAP2) has been the main focus of the studies to date. AAP2 is a nucleolar localizing protein essential for AAV2 capsid assembly (42, 150). When the AAV2 VP proteins are expressed in cultured cells in the absence of AAP, the VP proteins can be found in both the cytoplasm and the nucleus but are excluded from the nucleolus, and there is no detectable capsid assembly (42, 150). When the AAV2 VP proteins and AAP2 are co-expressed in cells, the VP proteins translocate to and accumulate in the nucleolus together with AAP2 and assemble into capsids (42, 150). It has been shown that AAP2 can form high-molecular-weight oligomers and change the conformation of a wide range of VP protein oligomer intermediates, leading to the formation of capsid-specific antibody-positive oligomers before the capsids are fully assembled (44). This, together with the demonstration of AAP2-AAV2 VP3 interactions through hydrophobic regions, may suggest that AAP functions as a scaffolding protein in the capsid assembly reaction as well as a transporter, targeting VP proteins to the nucleolus for assembly (44, 148). As for the role of AAPs derived from other AAV

serotypes, Sonntag et al. have demonstrated that AAP is essential for AAV1, AAV8 and AAV9 assembly by showing that expression of VP3 alone does not yield capsids, but that assembly can be restored by co-expressing heterologous AAP2 (43). They have also shown that such cross-complementation of capsid assembly with heterologous AAP2 does not easily extend to AAV5, one of the most divergent serotypes (43). In addition, they have found that AAP1, AAP2 and AAP5 protein expression levels could be significantly affected by the nature of co-expressed homologous and heterologous VP proteins (43). While these findings are intriguing, a number of questions remain about the roles of AAP in capsid assembly and elsewhere in the life cycle.

In this study, to further understand the roles of AAPs in AAV capsid assembly, we comprehensively characterized AAPs derived from AAV1 to 12 (AAP1 to 12, respectively). To this end, we investigated subcellular localizations of AAP1 to 12 and their assembly-promoting abilities for homologous and heterologous AAV VP3 proteins derived from all the 12 serotypes by immunofluorescence microscopy, a comprehensive AAP-VP3 cross-complementation assay, and transmission electron microscopy. These analyses revealed AAP's varied capability of nucleolar enrichment and heterologous capsid assembly among the serotypes and serotype-dependent differences in the sites of capsid assembly inside the nucleus. The most striking finding was that VP proteins derived from AAV4, 5 and 11 could assemble without the requirement of AAP in contrast to the VP proteins from the other nine serotypes that required AAP for assembly. Recombinant AAV5 vector produced in the absence of AAP

was found to be infectious and capable of transducing cells, indicating that AAP is not a necessary component for the production of infectious virions for some AAV serotypes.

Results

Successful expression of AAP1 to 12 in cultured cells

Initial work by Sonntag et al. found that the expression level of AAP in transfected cells was variable by serotype and was especially low for AAP5. In effort to try and increase AAP production in our transfections, we used the 750-bp CMV-IE enhancer-promoter fused with the 132-bp intervening sequence (IVS) known to enhance the stability of mRNA (338) to drive expression of a FLAG-tagged codon-optimized AAP2 from the strong ATG start codon in place of the native non-ATG start codon. Previously, we had found that this expression plasmid construction led to strong steady-state levels of AAP2 (150). Being prompted by this observation, we constructed a panel of FLAG-tagged AAP expression plasmids for the AAPs from AAV1 to 12, pCMV$_3$-FLAG-cmAAPx (x=1 to 12), using the same plasmid backbone. All the AAPs could be readily detected by western blot analysis in plasmid-transfected HEK 293 cells in the absence of co-expressed VP proteins, with each primarily showing a discrete single band, except for AAP8, 9, and 10 which had a faint secondary band (**Fig. 3.1**). Interestingly, the molecular weights of FLAG-tagged AAPs that were experimentally determined by western blot analysis were all found to be heavier than the theoretical molecular weights by 13% to 63% with AAP4 showing the largest discrepancy (**Table S1**). Such a discrepancy has previously been observed for AAP5 (43). The slower-than-expected migration might simply be due to differences in amino acid sequences, although the possibility of post-translational modification cannot be totally ruled out.
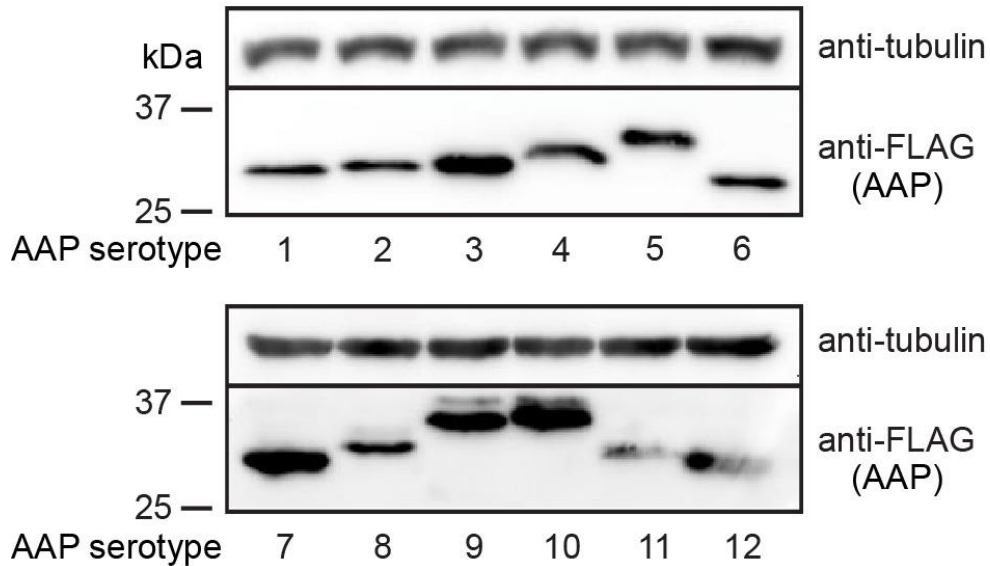
Fig. 3.1 Expression of AAP1 to 12 in HEK 293 cells.
HEK 293 cells were transiently transfected with a plasmid expressing the respective FLAG-tagged AAP indicated below each panel and AAP expression in transfected cells was analyzed by western blot using anti-FLAG antibody. α-tubulin was used as a loading control. Molecular weight markers (kDa) are shown to the left.

AAPs show varied sub-nuclear localization

Our previous work on identifying amino acids involved in AAP2 nuclear and nucleolar localization identified five clusters of basic amino acids (AAP2BR1 to AAP2BR5) near its C-terminus that contained overlapping nuclear and nucleolar localization signals (NLS-NoLS) (150). While these regions are mostly well-conserved amongst the AAPs from different AAV serotypes (AAP BR1 to AAP BR5 in **Fig. 3.2**), certain AAPs lack some or nearly all of the lysine and arginine

107

residues found in their corresponding protein sequence (**Fig. 3.2**). For instance, AAP5 has more proline residues and fewer basic residues than AAP2 in these regions, leading us to hypothesize that AAP5 might be nucleolar excluded, as discussed in our previous publication (150). To test this hypothesis and determine the subcellular localization of AAP1 to 12, HeLa cells were transiently transfected with each FLAG-tagged AAP-expressing plasmid, pCMV$_3$-FLAG-cmAAPx (x=1 to 12), individually, and localization was determined 48 hours later using an anti-FLAG antibody. All of the AAPs were found in the nucleus, but their nucleolar localization varied by serotype. AAP5 and AAP9, which show the two lowest theoretical isoelectric point (pI) values within the amino acid stretch spanning AAP BR1 to AAP BR5 (**Fig. 3.2 and Table S2**), were nucleolar excluded and all the other ten AAPs showed nucleolar localization to various degrees (**Fig. 3.3**). In a random model in which only two AAPs are nucleolar excluded among a total of twelve AAPs, the probability that the two AAPs showing the lowest pI values are the two AAPs that exhibit a nucleolar exclusion pattern is 0.015. This indicates strong correlation between pI values and nucleolar exclusion. There is another basic amino acid-rich region further down toward the C-terminus (AAP BR6); however, such a strong relationship between low pI and nucleolar exclusion was not observed in the larger segment containing AAP BR6 (**Table S2**). Therefore, the amino acid stretch containing the AAP BR1 to AAP BR5, and not the larger segment, is the most likely determinant of nucleolar localization. Among those that showed nucleolar localization, AAP1, AAP3B, AAP6, AAP7, AAP11 and AAP12 exhibited strong nucleolar enrichment similar to AAP2, while AAP4,

AAP8 and AAP10 showed more even distribution throughout the nucleus. These results demonstrate that, while all of the 12 AAPs are nuclear localized, the strict nucleolar localization observed in AAP2 is not a general characteristic of AAPs, and that AAPs with low pI values in the well-conserved basic amino acid-rich region near the C-terminus are devoid of the ability to translocate to the nucleolus.


```
       144  BR1___      BR2        BR3    __BR4___    BR5     _BR6_              204
AAP2   KSKRSRRMTVRRRLPITLPARFRCLLTRSTSSRTSSARRIKDASRRSQQTSSWCHSMDTSP
AAP1   KSRRSRRMMASQPSLITLPARFKSSRTRSTSFRTSSALRTRAASLRSRRTCS---------
AAP3B  KLKRSRRTMARRLLPITLPARFKCLRTRSISSRTCSGRRTKAVSRRFQRTSSWSLSMDTSP
AAP4   RSRRSRRRTARQRWLITLPARFRSLRTRRTNCRT---------------------------
AAP5   KSKRSRCRTPPPPSPTTSPPPSKCLRTTTTSCPTSSATGPRDACRPSLRRSLRCRSTVTRR
AAP6   KSRRSRRMMASRPSLITLPARFKSSRTRSTSCRTSSALRTRAASLRSRRTCS---------
AAP7   RSRRSRRMTALRPSLITLPARFRYSRTRNTSCRTSSALRTRAACLRSRRTSS---------
AAP8   RSRRSRRMKAPRPSPITSPAPSRCLRTRSTSCRTFSALPTRAACLRSRRTCS---------
AAP9   RSKRLRTTMESRPSPITLPARSRSSRTQTISSRTCSGRLTRAASRRSQRTFS---------
AAP10  RSRRSRRMKAPRPSPITLPARFRYLRTRNTSCRTSSAPRTRAACLRSRRMSS---------
AAP11  KLRRSQRRTARLRSLITLPARFRYLRTRRMSSRT---------------------------
AAP12  RSRRSRRQTARLRSLITLPARFRSLRIRRMNSHT---------------------------
```

Fig. 3.2 Sequence alignment of the C-termini of AAP1 to 12. Sequences of the basic amino acid-rich C-termini of AAP1 to 12 are aligned. The basic amino acid-rich regions (BRs), BR1 to BR6, are indicated with overlines. These 6 BRs are defined arbitrarily based on the sequence alignment data. The BR1 to BR5 have previously been identified as NLS-NoLS in the context of AAP2. Please note that the definition of BR4 is different from the definition described in our previous publication in that it extends toward the N-terminus by 2 amino acid residues to include the arginine residue that is conserved among all the serotypes except for AAV2.
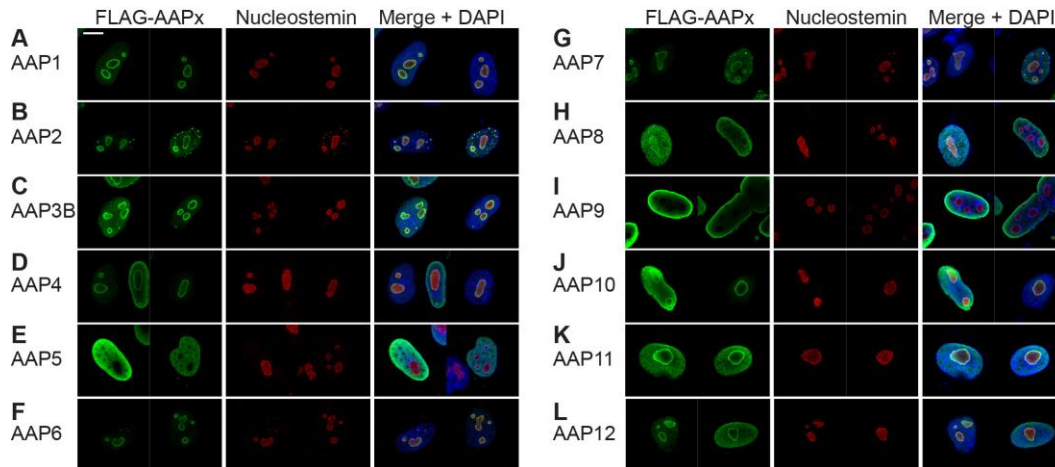
Fig. 3.3 Intracellular localization of the AAP1 to 12.
HeLa cells were transiently transfected with a plasmid expressing the respective FLAG-tagged AAP indicated in each panel. The cells were fixed 48 hours post-transfection and immunostained with anti-FLAG antibody (green) and anti-nucleostemin antibody (red), and counterstained with DAPI (blue), before being imagined on a Zeiss LSM 710 confocal microscope with a 100x/1.46 NA objective. Representative cell images are shown. (**A**) AAP1, (**B**) AAP2, (**C**) AAP3B, (**D**) AAP4, (**E**) AAP5, (**F**) AAP6, (**G**) AAP7, (**H**) AAP8, (**I**) AAP9, (**J**) AAP10, (**K**) AAP11, and (**L**) AAP12. All AAPs are enriched in the nucleus. AAP5 and 9 show a nucleolar exclusion pattern, while nucleolar signals are observed with other AAPs to various degrees. Please note that a ring-shaped anti-FLAG antibody staining pattern with a central hollow (Panels **A**, **B**, **C**, **D**, **F**, **G**, **J**, **K** and **L**) is most likely an artifact caused by overexpression of a nucleolar-localizing protein and does not necessarily indicate that nucleolar expression is enriched in the nucleolar periphery.

Investigation of assembly of various serotype capsids without AAP

It has been well established that AAV2 capsid assembly requires AAP (42, 150). As for other serotypes, Sonntag et al. have previously reported that AAV1, AAV5, AAV8 and AAV9 capsid assembly also requires AAP (42). To reproduce the observations by Sonntag et al. and investigate AAP-independent assembly for other serotypes, we expressed AAV1 to 12 VP3 proteins from $pCMV_1$-AAVxVP3 plasmids (x=1 to 12) in HEK 293 cells in the presence or absence of cognate AAP expressed from a separate plasmid. In this transfection system, we also provided the components necessary for packaging of recombinant AAV viral genomes encoding enhanced green fluorescence protein (eGFP) into assembled capsids. We then quantified packaged viral genomes as surrogates of assembly by quantitative dot blot in a biologically duplicated experiment. In the presence of AAP, all the serotypes showed assembled capsids at readily detectable levels as expected (**Fig. 3.4**). Surprisingly, in the absence of AAP, the signals from AAV4, 5 and 11 were clearly beyond the background in contrast to other serotypes (**Fig. 3.4**). This result strongly indicated that AAV4, 5 and 11, but not the other nine serotypes, could assemble without AAP.
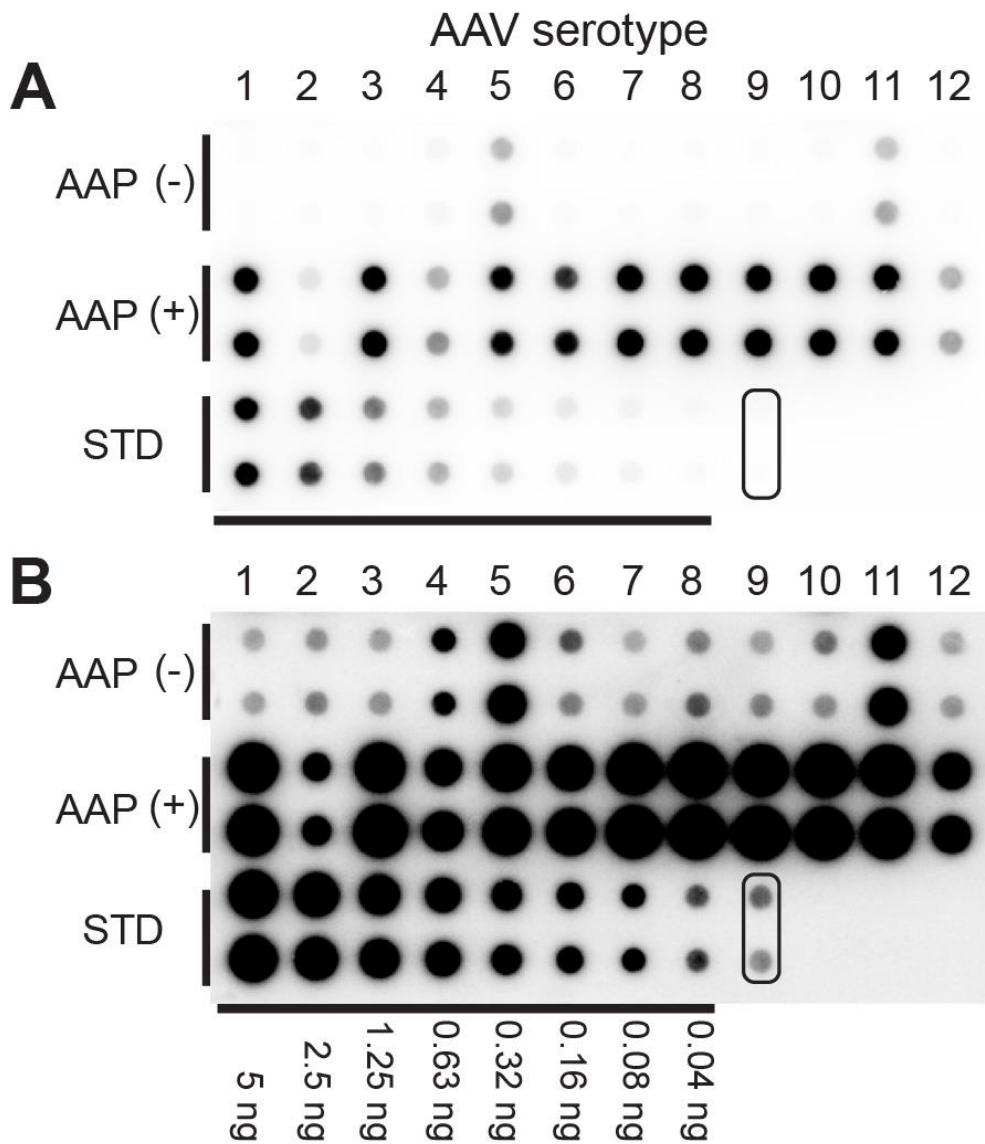
Fig. 3.4 Dot blot titering of VP3 capsids produced in the absence of AAP.
(**A**) AAP-independent assembly of VP3 proteins from AAV1 to 12 was assessed by a quantitative dot blot assay in a biologically duplicated experiment. VP3 only particles derived from AAV1 to 12 that contained a double-stranded AAV-CMV-GFP genome were produced in 6-well plates in the presence or absence of their cognate AAPs. Benzonase-resistant DNA was recovered from 7% of the samples obtained from each well, blotted on a nylon membrane together with standards (i.e., linearized pEMBL-CMV-GFP plasmid), and probed with a ^{32}P-labeled GFP probe. Pairs of dots in each combination represent the results obtained from two separate transfections. The pair of dots indicated with a rounded rectangle are negative controls. (**B**) The same blot as that shown in Panel A. The signals are intensified.

AAV5 VP3 can assemble capsids without AAP

To examine potential AAP-independent assembly further, we chose to study AAV5 in-depth because there are commercially available reagents for this serotype that allow for verification of capsid assembly on multiple levels and because, contrary to our experience, AAV5 assembly was previously reported to require cognate AAP5 (43). In all the experiments described above, the VP3 plasmids used still retained the C-terminal 80% of the AAP ORFs. Without its start codon, AAP should not be expressed, but it was important to rule out any possibility that an unidentified cryptic start codon could have provided a functional phenotype through expression of an N-terminally truncated AAP5. To exclude this possibility, we created plasmids carrying codon-modified AAV2 and AAV5 VP3 ORFs (pCMV$_1$-AAV2cmVP3 and pCMV$_1$-AAV5cmVP3) and used them in the subsequent experiments. In these new plasmids, the amino acids encoded by the codon-modified AAP ORFs were less than 50% identical when compared to the corresponding wild-type AAPs and there were 4 to 6 stop codons within the AAP ORFs. It would therefore be very unlikely for the plasmids to produce functional AAPs. To verify that the AAV5cmVP3 proteins could also assemble capsids without AAP, we transiently transfected HeLa cells with pCMV$_1$-AAV2cmVP3 or pCMV$_1$-AAV5cmVP3 together with either the cognate AAP-expressing plasmid (pCMV$_3$-FLAG-cmAAP2 or pCMV$_3$-FLAG-cmAAP5) or pCMV$_3$-GFP, a GFP-expressing control plasmid devoid of AAP expression. Forty-eight hours after transfection, the cells were stained with anti-FLAG and anti-AAV capsid antibodies (A20 to detect assembled AAV2 VP3 capsids and

ADK5a to detect assembled AAV5 VP3 capsids). As previously shown (42), AAV2 VP3 was only able to assemble capsids when AAP2 was supplied in trans (**Fig. 3.5A**). In contrast, AAV5 VP3 was able to produce capsid antibody-positive staining regardless of the presence or absence of AAP5 (**Fig. 3.5A**). The ADK5a antibody is an anti-AAV5 antibody that should be specific only to intact AAV5 capsids and not monomers or oligomers of AAV5 VP proteins (295). Thus, the positive staining with ADK5a strongly indicates that the AAV5 VP3 can assemble into capsids without a need for AAP.
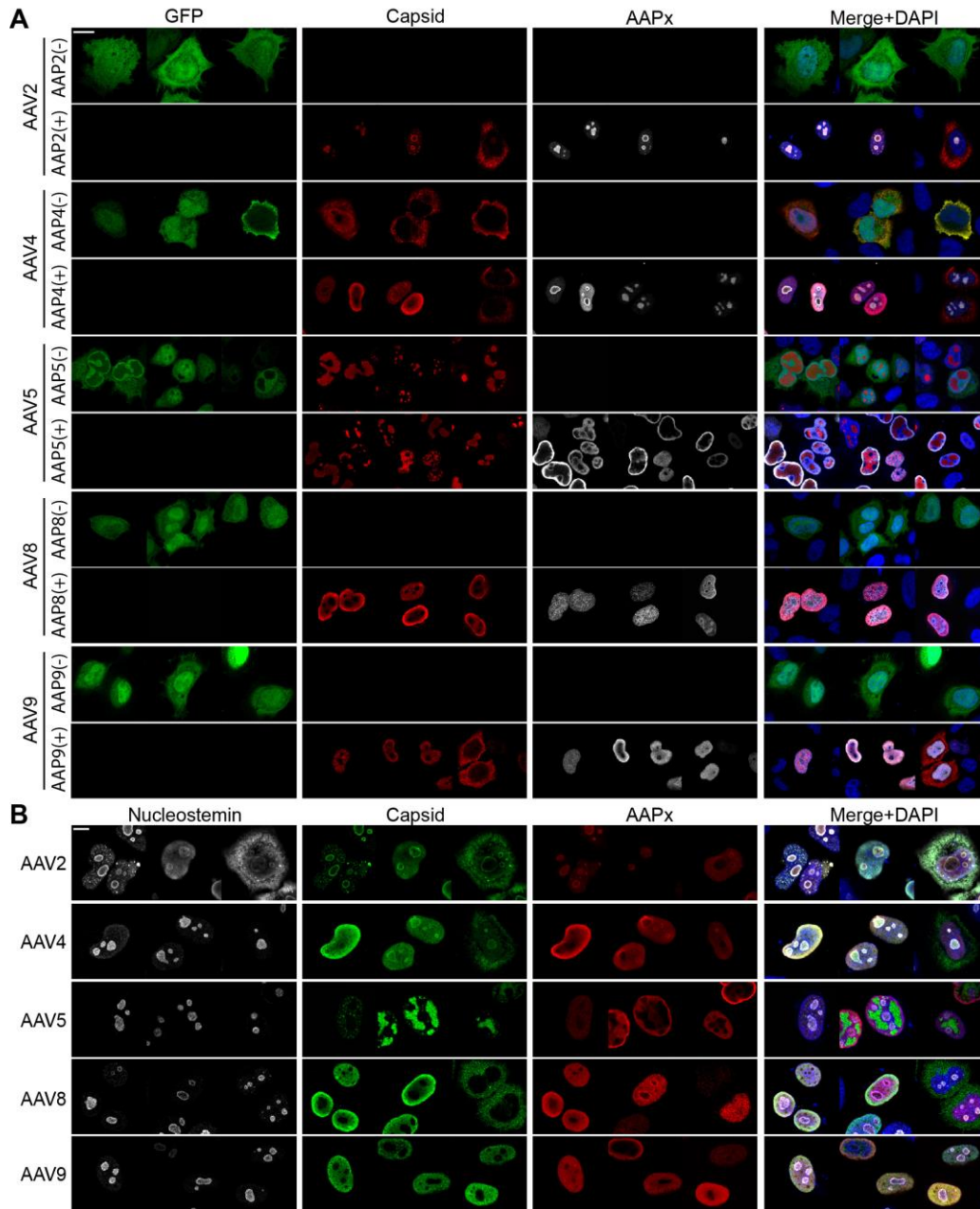
Fig. 3.5 Capsid assembly of AAV2, 4, 5, 8 and 9 in HeLa cells.
HeLa cells were transiently transfected with plasmids expressing VP3 proteins derived from AAV2, 4, 5, 8 or 9 in the presence (+) or absence (-) of a co-transfected plasmid expressing their cognate AAP. The groups that received no AAP-expressing plasmid were transfected with pCMV$_3$-GFP instead to ensure successful transfection and maintain the total quantity of transfected DNA constant across the groups. The cells were fixed 48 hours post-transfection, immunostained with the antibodies indicated in each panel, and counterstained with DAPI. The images were obtained using Zeiss LSM 710 confocal microscope with a 63x/1.4 NA objective. (**A**) The cells were stained with anti-AAV capsid antibody (A20 for AAV2, ADK4 for AAV4, ADK5a for AAV5, ADK8 for AAV8 and ADK8/9 for AAV9; red) and anti-FLAG antibody (white). (**B**) The cells were stained with anti-nucleostemin antibody (white), anti-AAV capsid antibodies (green) and anti-FLAG antibody (red). Scale bar, 10 μm.

Finally, transmission electron microscopy (TEM) was used to visualize directly the assembly (or not) of AAV5 "VP3 only" capsid in the presence or absence of AAP5. HEK 293 cells were transfected using either $pCMV_1$-AAV2cmVP3, with or without $pCMV_3$-FLAG-cmAAP2, or $pCMV_1$-AAV5cmVP3, with or without $pCMV_3$-FLAG-cmAAP5 plasmids. Five days post-transfection, the media and cells were harvested and subjected to three rounds of purification by cesium chloride (CsCl) density-gradient ultracentrifugation to prepare samples for TEM. On the final round, a discrete band was found at refractive indices (RIs) of 1.3640 to 1.3645, except for AAV2 VP3 without AAP2 (**Fig. 3.6A and D**). After fractions were collected, an enzyme-linked immunosorbent assay (ELISA) specific for intact capsids, was performed on each fraction to confirm the presence of assembled AAV2 or AAV5. Assembled AAV2 capsids were not detected in the absence of AAP2, but were present when AAP2 was co-expressed (**Fig. 3.6B**), consistent with the observed CsCl-banding and immunofluorescence microscopy. AAV5 capsids were detected by ELISA in both the AAV5 VP3 without AAP5 and the AAV5 VP3 with AAP5 samples (**Fig. 3.6E**). The analysis of TEM images of the fractions showing high ELISA optical density (OD) values confirmed the presence of assembled capsids while the AAV2 VP3 without AAP2 sample of the corresponding RI (1.3640) did not contain any capsids (**Fig. 3.6C and F**). AAV5 VP3 capsids produced in the absence of co-expressed AAP5 were morphologically indistinguishable from those produced in the presence of AAP5 (**Fig. 3.6F**). Thus, we concluded that the AAV5 VP3 proteins are capable of forming capsids in the absence of AAP.
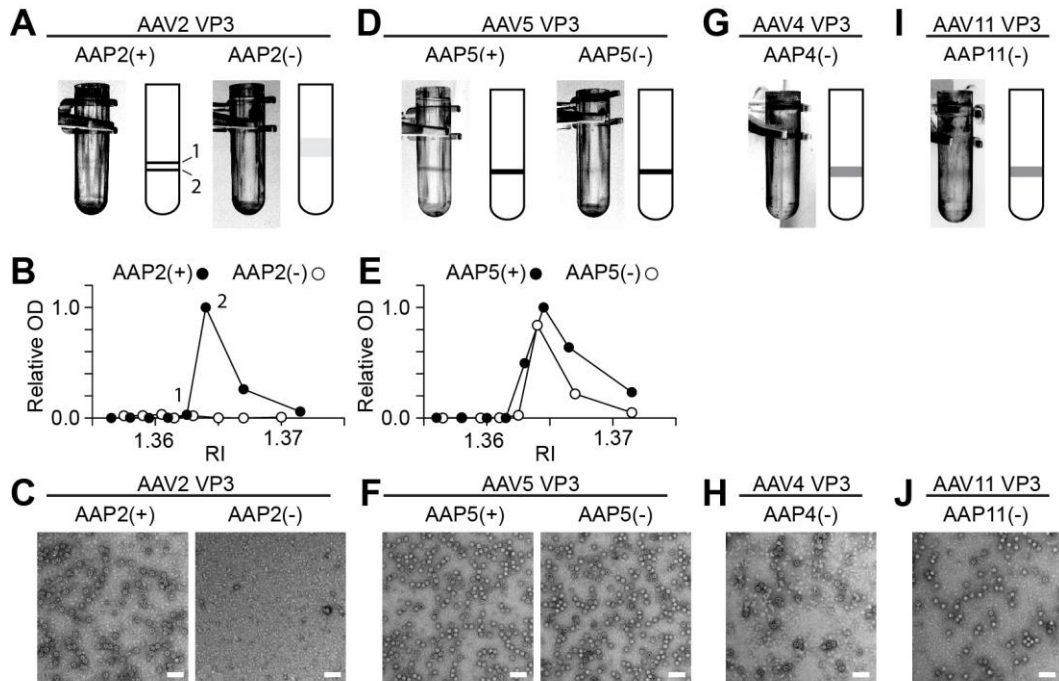
Fig. 3.6 Transmission electron microscopy of AAV "VP3 only" capsids produced with or without AAP. HEK 293 cells were transfected with plasmids expressing the VP3 proteins and AAP proteins as indicated. Five days post-transfection, capsids were purified by three rounds of CsCl density-gradient ultracentrifugation. (**A, D, G,** and **I**) Negative image photos and accompanying cartoons show bands formed after the third CsCl ultracentrifugation. AAV2 VP3 with AAP2 [AAP2(+)] contained two bands at RIs of 1.3630 and 1.3645. The following ELISA indicated that the heavier of the two bands contained more capsids and this was the band used for TEM imaging. (**B** and **E**) Assembled AAV capsids positive for anti-AAV capsid antibody in each CsCl fraction were assessed by an AAV capsid-specific ELISA. Relative optical density (OD_{450}) values are plotted against RIs of CsCl fractions. The OD_{450} value obtained in the peak fraction is set as to 1.0. Closed and open circles represent the samples prepared with or without AAP. (**C, F, H** and **J**) Representative TEM images of the samples prepared with or without AAP. The CsCl fractions showing RIs of 1.3645, 1.3640, 1.3645, 1.3640, 1.3640 and 1.3640 were used for TEM imaging of the AAV2 VP3 with and without AAP2 [AAP2(+) and AAP2(-)], AAV5 VP3 with and without AAP5 [AAP5(+) and AAP5(-)], AAV4 VP3 without AAP4, and AAV11 VP3 without AAP11 samples, respectively. Scale bar, 100 nm. Panels **A**, **B** and **C** are for AAV2; Panels **D**, **E** and **F** are for AAV5; Panels **G** and **H** are for AAV4; and Panels **I** and **J** are for AAV11.

Infectious AAV5 virions can be produced without AAP

While AAV capsids can be composed entirely of VP3, VP1 is required for viral infectivity and cell transduction (139, 339, 340). Having established that AAV5 VP3 could assemble capsids in the absence of co-expressed AAP, we sought to determine if infectious AAV5 virions could be produced by expressing all of the AAV5 VP proteins, VP1, VP2 and VP3, without supplying the AAP5 protein. To this end, we utilized two types of AAV5 helper plasmids, pHLP-AAV5(AAP5+) and pHLP-AAV5(AAP5-). The pHLP-AAV5(AAP5+) plasmid is our standard AAV5 helper plasmid for vector production that expresses AAV2 Rep, AAV5 VP1, VP2 and VP3, and AAP5 proteins. The pHLP-AAV5(AAP5-) expresses all the AAV proteins except AAP5 due to extensive codon modification of the AAP-VP overlapping ORFs that abolishes only AAP5 expression (**Fig. 3.7A**). Western blot analysis of the cell lysates obtained from HEK 293 cells transfected with pHLP-AAV5(AAP5-) or pHLP-AAV5(AAP5+) with or without pCMV$_3$-FLAG-cmAAP5 showed that the AAV5 VP3 protein was detectable in small amounts when AAP5 was not expressed, and that the steady-state level of expression of AAV5 VP proteins could be increased substantially when AAP5 was co-expressed (**Fig. 3.7B**). Using the adenovirus-free plasmid transfection method (222) with pHLP-AAV5(AAP5-), we could successfully produce AAV5 vectors containing a recombinant AAV vector genome in the absence of AAP5 expression, although the vector yield was a magnitude lower than the yields that could be obtained when AAP5 was expressed (**Fig. 3.7C**).
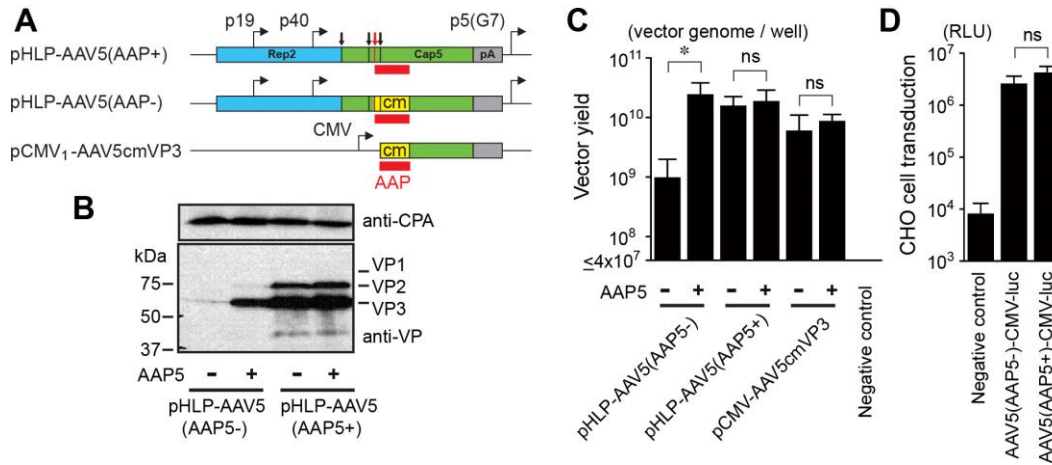
Fig. 3.7 Production and characterization of AAV5 VP1/VP2/VP3 particles produced with and without AAP5.

(**A**) Schematic representation of the plasmids expressing AAV5 VP proteins. These plasmids were used for the experiments shown in Panels **B**, **C** and **D**. The VP1, VP2, VP3 translation start sites are indicated with black vertical lines and arrows from the left to the right, respectively. The AAP translation start site is indicated with red vertical lines and a red arrow. The AAP ORF is shown with red boxes. The codon-modified region is indicated with yellow boxes. pA, polyadenylation signal; p5(G7), the AAV2 p5 promoter with the TATA box sequence TATTTAA replaced with GGGGGGG. (**B**) HEK 293 cells were transiently transfected with an AAV5 helper plasmid, pHLP-AAV5(AAP5+) or pHLP-AAV5(AAP5-), with or without a plasmid expressing AAP5. All the groups were also transfected with an adenovirus helper plasmid, pHelper, to induce protein expression from the AAV5 helper plasmids. Forty-eight hours post-transfection, AAV5 VP1, VP2 and VP3 proteins were probed with anti-AAV VP antibody (B1) by western blot. Cyclophilin A (CPA) was used as a loading control. (**C**) HEK 293 cells were transfected with plasmids indicated in the figure to produce AAV5 VP1/VP2/VP3 particles or "VP3 only" particles containing a double-stranded AAV-CMV-GFP vector genome in the presence or absence of AAP5 expressed from a separate plasmid, pCMV$_3$-FLAG-cmAAP5. Five days post-transfection, the media and cells were harvested, and Benzonase-resistant viral genomes in each sample were quantified by a dot blot assay in a biologically triplicated experiment. Y-axis shows AAV vector titer (vector genomes) per well in a 6-well plate format. (**D**) CHO-K1 cells were infected with either AAV5(AAP5+)-CMV-luc or AAV5(AAP5-)-CMV-luc vector, which were produced with or without AAP5, respectively, at an MOI of 10^6. Luciferase activity was measured 46 hours post-infection in a biologically triplicated experiment. The negative control group received the luciferase-containing samples prepared in the same manner except for the absence of AAV5 VP protein expression, which provides a measure of pseudo-transduction. For the pseudo-transduction control, the same sample volume as that for the AAV5(AAP5-)-CMV-luc vector preparation was used. Y-axis shows relative light units (RLUs). Error bars represent standard deviations. An asterisk indicates statistical significance with $P < 0.05$ (two-tailed Welch's t-test). ns, not significant

We then produced two types of recombinant AAV5 vectors expressing firefly luciferase, AAV5(AAP5-)-CMV-luc and AAV5(AAP5+)-CMV-luc, using pHLP-AAV5(AAP5-) and pHLP-AAV5(AAP5+), respectively, without supplying AAP5 in trans from the pCMV$_3$-FLAG-cmAAP5 plasmid. The two vector preparations were added on Chinese hamster ovary (CHO)-K1 cells at a multiplicity of infection (MOI) of 10^6, and luciferase activity was measured 46 hours post-infection in a biologically triplicated experiment. The results showed that the AAV5(AAP5-)-CMV-luc vector was able to transduce CHO-K1 cells and there was no statistically significant difference in transduction efficiency between AAV5(AAP5-)-CMV-luc and AAV5(AAP5+)-CMV-luc vectors (two-tailed Welch's *t*-test, P=0.18, **Fig. 3.7D**). Taken together, our data demonstrate that AAV5 does not require AAP5 for assembly of infectious virions containing a recombinant viral genome.

AAV4 and 11 can assemble capsids without AAP

Having determined that AAV5 was capable of assembling infectious virions in the absence of AAP, we next sought to determine conclusively if AAV4 and AAV11 could also assemble capsids without AAP. AAV4 has a commercially available mouse monoclonal antibody against assembled AAV4 capsid, ADK4, and thus we were able to use immunofluorescence microscopy to determine if capsids could be made from AAV4 VP3 without AAP4. However, this approach was not applicable to AAV11 because an antibody specific to assembled AAV11 capsid is currently not available.

In the immunofluorescence microscopy approach, we transfected HeLa cells with pCMV$_1$-AAV4VP3 with either pCMV$_3$-FLAG-cmAAP4 or pCMV$_3$-GFP. Forty-eight hours post-transfection, the cells were fixed and stained with anti-FLAG antibody or anti-AAV4 capsid antibody (ADK4). This analysis confirmed that AAV4 VP3 was also able to assemble antibody-positive capsids regardless of whether AAP4 was co-expressed (**Fig. 3.5A**). In the TEM approach, we followed the same procedure that we took for the AAV5 TEM experiment described earlier. In brief, to completely abolish any possible functional AAP expression, we constructed pCMV$_1$-AAV4cmVP3 and pCMV$_1$-AAV11cmVP3 in which the AAP-VP overlapping ORFs were extensively codon-modified compared to the native nucleotide sequences. These two plasmids were used to produce viral capsids in HEK 293 cells in the absence of co-expressed AAP, and the TEM samples were then prepared by the three rounds of CsCl density-gradient ultracentrifugation. On the final round of the ultracentrifugation, a discrete faint band could be seen in each tube at a position showing RIs of 1.3640 for both AAV4 and AAV11, respectively (**Fig. 6G and I**). The TEM analysis revealed that these bands contained capsid structures of the expected size and shape for AAV (**Fig. 6H and J**). These observations provide direct evidence that AAV4 and AAV11 VP3's are capable of assembling capsids without AAP.

Heterologous AAPs efficiently promote capsid assembly except for AAP4, 5, 11 and 12.

The promiscuity of AAPs in assembling heterologous AAV serotype capsids has previously been shown with combinations of AAPs and AAV VP3 proteins derived from AAV1, 2, 5, 8 and 9 (43). In this study by Sonntag et al., they reported that AAP2 could stimulate assembly for heterologous AAV1, 8 and 9 capsids, but not AAV5 capsid; AAP1 could stimulate AAV2 capsid assembly; and AAP5 could weakly support AAV1 capsid assembly. While the observation obtained from these limited combinations has given us a glimpse of the nature of the AAP-VP compatibility in capsid assembly, more expansive investigation of numerous different AAP and VP3 combinations for heterologous capsid assembly is imperative for deeper understanding of the role of AAP. In order to fill this knowledge gap, we performed an AAP-VP3 cross-complementation study in which we investigated all the possible 121 combinations of AAP1 to 11 and the VP3 proteins from AAV1 to 11 in a biologically triplicated experiment. Including the no AAP controls that determined the background levels of the assay, a total of 396 assessments of capsid formation was required for undertaking this comprehensive experiment. For this reason, we applied a massively parallel capsid assembly assay using the Illumina sequencing-based AAV Barcode-Seq approach reported earlier (261). In this approach, packaged viral genomes were quantified as surrogates of assembly (see Materials and Methods). A subset of 7 combinations were validated by quantitative dot blot in a biologically replicated experiment (n=2 to 4). This validation experiment revealed that the results

obtained by these two different methods were primarily concordant (**Fig. S1**). As for AAP12, which was not included in the massively parallel analysis, we determined the yield by quantitative dot blot. The results showed that AAPs derived from AAV1, 2, 3B, 6, 7, 8, 9, and 10 could promote heterologous serotype capsid assembly at least 30% as efficiently as the native combinations (**Fig. 3.8** and **Fig. S2**). As observed in the previous study (43), AAP5 also weakly supported assembly of many heterologous VP3 proteins. AAP12 supported assembly efficiently only for AAV11 and 12. AAP4, 11 and 12 were found to be least efficient for heterologous capsid assembly (**Fig. 3.8**). AAV12 VP3, which could not assemble without AAP, could form capsids exclusively by AAP4 or AAP11 besides its cognate AAP12 (**Fig. 3.8**). The capsid assembly promoting role of heterologous AAPs were difficult to assess for the AAV4, 5 and 11 VP3 proteins because of the high background signals due to the AAP-independent assembly (**Fig. S2**). Nonetheless, the decreased ability for AAP4, 5, 11 and 12 to promote assembly of VP3 capsids of heterologous serotypes is congruent with the previously published data showing that AAP4, 5, 11 and 12 were the most phylogenetically dissimilar AAPs when compared to the other AAPs derived from AAV1 to 13 (44). These observations establish that the AAPs promiscuously promote efficient assembly of capsids among wide groups of closely related serotypes, but that promiscuity does not extend to more distantly related serotypes such as AAV4, 5, 11 and 12.
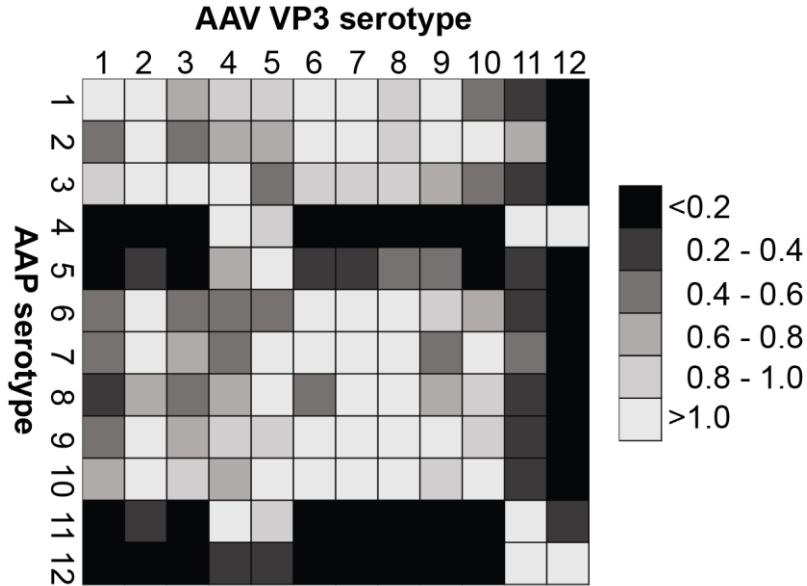
Fig. 3.8 The ability for each AAP to assemble homologous and heterologous VP3 proteins derived from AAV1 to 12 is shown in a matrix heatmap.
The data were obtained by AAV Barcode-Seq for 121 combinations between AAP1 to 11 and AAV1 to 11 VP3's in a biologically triplicated experiment. As for AAV12, the data were obtained by a quantitative dot blot assay in a biologically duplicated experiment.

Concordant and discordant subcellular localization between AAP and capsids

During our examination of the subcellular localization of various AAPs and capsid assembly, we observed both concordant and discordant subcellular localization patterns between particular AAV capsids and their cognate AAPs. In the case of AAV2, both AAP2 and assembled AAV2 VP3 capsids were localized

to the nucleoli and nucleostemin-positive sub-nuclear bodies (**Fig. 3.5B**). Co-localization of assembled capsids and AAP was also observed for AAV4 and AAV9 (**Fig. 3.5A and B**). In the case of AAV5, although assembled AAV5 VP3 capsids and AAP5 were both nucleolar excluded, they exhibited contrastive patterns of nucleoplasmic distribution. The AAV5 VP3 capsid signals formed discrete spheres and globules while the AAP5 signals were found diffusely in the nucleoplasm excluding the nucleoli and the areas where the AAV5 VP3 capsids were localized (**Fig. 5A and B**). In contrast to AAV2, which displayed tight association in the nucleus between assembled capsid, AAP and nucleostemin, neither of the AAV5 capsids or AAP5 was co-localized with nucleostemin (**Fig. 3.5B**). Likewise, such discordance was also observed with AAV8. AAP8 was expressed diffusely in the nucleus with occasional nucleolar enrichment or exclusion while assembled AAV8 VP3 capsids were always completely nucleolar excluded, and there was no association between the assembled AAV8 VP3 capsids and nucleostemin (**Fig. 3.5A and B**). AAV4 capsids were enriched in the nucleolus, but did not show tight association with nucleostemin in the same manner as AAV2 (**Fig. 3.5B**). Both AAP9 and AAV9 capsids were nucleolar excluded and, like AAV4, 5 and 8, did not show any association with nucleostemin (**Fig. 3.5A and B**). These observations demonstrate that the sites where AAP proteins are enriched do not necessarily correspond with the sites where capsids assemble and accumulate, although the mechanisms underlying these concordances and discordances in the sub-nuclear localizations of AAPs and assembled AAV capsids have yet to be elucidated. In addition, our data presented

here challenge a long-believed notion that tight association with the nucleolus is a hallmark of AAV capsid assembly. Because assembled capsids for certain serotypes such as AAV5, 8 and 9 are only found outside the nucleolus and are not associated with the nucleolar protein nucleostemin, the involvement of the nucleolus and nucleolar proteins seen with AAV2 capsid assembly is not generalizable to all AAV serotypes.

Discussion

In this study, we comprehensively analyzed AAPs derived from twelve AAV serotypes, AAV1 to 12, for their subcellular localization and their ability to promote assembly of capsids derived from homologous and heterologous serotypes. For nearly a half century, AAV2 has been the prototype for the studies on the AAV viral life cycle including infection, replication and assembly, and much of what is currently known about AAV comes from experiments with this serotype. Although all the currently known AAV serotypes and variants display significant similarity to each other in their viral protein amino acid sequences and virion structures, the studies of AAV as gene delivery vectors for the last two decades have delineated that there are substantial differences in their biological properties including cell and tissue tropism (238). This has raised a possibility that certain aspects of the fundamental AAV biology are also diverse among different serotypes and what we had learned from the prototype AAV2 might not be applicable to the basic biology of the entire AAV family, including the process of capsid assembly. In this regard, the study reported here has clearly demonstrated that: 1) AAP, which had been believed to be essential for capsid assembly based on the observations obtained with AAV2, 8 and 9 (42, 43), is not necessarily so for other serotypes; 2) tight association between AAP and assembled capsids in the nucleolus during or post-assembly, which has been shown with AAV2, is not always the case in other serotypes; and 3) targeting to the nucleolus, which serves as an indispensable organelle for many viruses including AAV2 (134, 148, 151, 274, 282, 284, 315), may not be required for

127

certain AAV serotypes. To be more concrete, we demonstrated here the dispensability of AAP in assembly of AAV4, 5 and 11 capsids and in production of infectious virions at least for AAV5; serotype-specific concordant (in the case of AAV2, AAV4 and AAV9) and discordant (in the case of AAV5 and AAV8) association between AAP and assembled capsids; and nucleolar exclusion of assembled AAV5, 8 and 9 capsids as opposed to the tight nucleolar association observed in AAV2 capsids.

Structural organization of AAP has been partially revealed in previous studies (44, 150). AAP has two hydrophobic domains of approximately 15 amino acids near the N-terminus. The first domain closer to the N-terminus consists of four short hydrophobic motifs that are evolutionarily highly conserved among AAV1, 2, 3B, 6, 7, 8, 9, 10 and 13 and partially conserved among AAV4, 5, 11 and 12. The second domain, which is closer to the C-terminus, is highly conserved across AAV1 to 13 and termed the "conserved core" of AAP (44). A co-immunoprecipitation analysis using a panel of AAP2 and AAV2 VP3 mutants has shown that these two hydrophobic domains in the AAP proteins play a crucial role in interacting with the capsid VP proteins through a highly conserved hydrophobic patch near the VP C-terminus that includes I682 (44) (note: the amino acid position is based on the AAV2 capsid protein). This high conservation of amino acids that constitutes the interface of the AAP-VP interaction might allow an AAP derived from one serotype to assist in the assembly of AAV capsids from other serotypes. Therefore, the promiscuity of AAPs in cross-

complementing assembly of heterologous capsids that we observed in the study conforms well to the inference drawn from the evolutionarily conserved interfaces in the AAP-VP complex. Although the mechanism underlying why AAP4, 5, 11 and 12 show no, only weak or limited promiscuity in heterologous capsid assembly has yet to be elucidated, the decreased hydrophobicity in the first hydrophobic region that is found only in these four most-distantly related AAPs might correspondingly decrease the ability for the AAPs to interact with the hydrophobic C-terminal regions in the VP proteins. A note of interest in the context of AAV capsid engineering by DNA shuffling (238), which is being actively pursued by many laboratories aiming to create the next generation AAV vectors, is that, while overall AAPs are generally able to cross-complement many different serotypes, capsid libraries which include AAV4, 5, 11 and 12 might benefit from providing their cognate AAPs during capsid assembly reactions.

To our surprise, our study convincingly demonstrated that AAV4, 5, and 11 can assemble capsids without a need of AAP. This was not anticipated because a previously published study had shown that AAV5 capsid assembly required AAP (43). Although the reasons for this discrepancy are still unknown, a few possible explanations exist. First, there was a difference in the AAV5 VP3 expression plasmid constructs used in the studies. The plasmid used in the previous study contained extra 154 base pairs after the stop codon derived from the AAV5 viral genome. Second, in our experiment, the AAV5 VP3 might have been expressed in cells at a level higher than that in the previous study. As detailed in the Materials

and Methods section, we used a strong CMV-IE enhancer-promoter with an enhancing element to express VP proteins. Taking this into account, a higher concentration of AAV5 VP3 proteins might be required for AAP-independent capsid assembly. Interestingly, AAV12 VP3 was found to require AAP for assembly. This was unexpected because AAV12 VP3 is evolutionarily closely related to VP3 proteins from AAV4 and 11 (43) exhibiting AAP-independent assembly and because the subcellular localization and the specificity in the cross-complementation of heterologous serotype assembly are conserved among these serotypes. Such similarity and dissimilarity in these closely related serotypes warrants furthers studies to delineate the AAP-independent assembly properties seen in AAV4 and AAV11 VP3, which will lead to a better understanding of AAV assembly mechanisms.

It is worth noting that, despite the dispensable nature of AAP5 for infectious AAV5 virion formation, the capsid assembly-promoting role of AAP5 is still obvious as the abolishment of AAP5 expression from our standard AAV5 helper plasmid expressing AAV2 Rep, AAV5 VP1, VP2 and VP3 resulted in titers that were one order of magnitude lower than the titers obtainable with the helper plasmid proficient in AAP5 expression (**Fig. 3.7C**). AAP4 and AAP11 also shared this role (**Fig. 3.8**) although they were not essential for the capsid formation. These observations lead us to speculate that AAP might be contributing to the promotion of capsid assembly through mechanisms other than its already described essential role. In this respect, it is intriguing that, when we

used the pCMV$_1$-AAV5cmVP3 plasmid to express AAV5 VP3 in HEK 293 cells (**Fig. 3.7A**), we could produce assembled AAV5 VP3 capsids at equivalent levels irrespective of the presence or absence of co-expressed AAP5 (**Fig. 3.7C**) while this was not the case when we used the pCMV$_1$-AAV5VP3 plasmid (**Fig. 3.4**). The pCMV$_1$-AAV5cmVP3 plasmid differs from the pCMV$_1$-AAV5VP3 plasmid in that the AAV5 VP3 ORF had been codon-optimized for human cell expression. Hence, the concentration of VP proteins in cells may be a rate-limiting step in the AAP-independent assembly process of the AAV5 capsid. At a lower concentration, the AAV5 VP3 proteins might require the presence of AAP5 to become stabilized and accumulate to the sites where capsids assemble through the interaction with AAP5. Therefore, it is plausible to propose a stabilization-accumulation mechanism via AAP-VP interactions through which capsid assembly could be further enhanced. Such a mechanism might be exerted not only for AAV5 but also other serotypes, although AAP-dependent serotypes still require the AAP's essential role for capsids to assemble.

For many years, the nucleolus has been viewed as an organelle important for the AAV2 life cycle. Early work showed that AAV2 capsids are first seen in the nucleolus and sub-nuclear bodies during replication (134, 148). Several studies have found a close association between nucleolar proteins and AAV2 capsids (134, 284), and it was found that AAV2 capsids are trafficked to the nucleolus following infection (151). These observations led the field to speculate a role for the nucleolus or nucleolar proteins in AAV2 replication. A proposed hypothesis is

that AAV capsid assembly occurs in the nucleolus and capsids are subsequently moved to the nucleoplasm for genome packaging in a Rep-dependent manner (134). The nucleolus as the site of AAV2 capsid assembly was further supported by the identification of the nucleolar-localizing AAP2 protein (42). However, the role of the nucleolus had not been examined in other AAV serotypes. This study addressed this question and revealed that AAP5 and 9 display a subcellular localization markedly different from that of AAP2, (i.e., complete exclusion from the nucleolus, and that AAV5, 8 and 9 capsids do not accumulate in the nucleolus). Thus, the question arises as to whether the nucleolus plays an important role in capsid assembly only for a subset of AAV serotypes including AAV2. It could be possible that certain AAV serotypes require only transient interaction with the nucleolus for capsid assembly and the assembled capsids leave the nucleolus for packaging viral genomes replicating in the nucleoplasm. Alternatively, for certain serotypes such as AAV5, assembly of the capsid may not have to rely upon factors in the nucleolus. Thus, our observations challenge the generalized view on the significance of the nucleolar roles in the AAV life cycle, and highlight potential heterogeneity of the mechanisms of viral capsid assembly and replication among different AAV serotypes.

In summary, we show that capsid assembly in the nucleolus and its strict dependence on AAP are not universal phenomena applicable to all AAV serotypes. A potential caveat in our study is that we used an artificial plasmid transient transfection system to study AAPs, which might not necessarily

recapitulate the AAV infection and replication that takes place in nature. Nonetheless, this study could at least reveal that the processes and mechanisms involved in the AAV life cycle among different serotypes are more heterogeneous than previously thought. Further study into the roles and functions of AAPs in the AAV life cycle will advance our foundational knowledge of icosahedral capsid assembly mechanisms and lead to improved methods for production of AAV vectors for gene therapy.

Materials and Methods

Plasmid construction. Plasmids pCMV$_3$-FLAG-cmAAPx (x=1 to 12) are plasmids expressing codon-modified (cm) versions of AAP with an N-terminal FLAG tag under the control of the human cytomegalovirus (CMV) immediate-early (IE) gene enhancer-promoter and an ATG start-codon. Each AAP ORF was codon-modified to optimize expression in human cells and cloned into the pCMV$_3$-FLAG-cmAAP2 parental plasmid used in our previous study (150) by replacing the cmAAP2 ORF with a new cmAAPx ORF. Plasmids pCMV$_3$-FLAG-AAP2 carries the native AAP2 ORF sequence in place of the cmAAP2 ORF. Plasmid pCMV$_3$-GFP is a plasmid expressing enhanced green fluorescence protein (eGFP) under the control of the same CMV-IE enhancer-promoter. Plasmids pCMV$_1$-AAVxVP3 (x=1 to 12) are plasmids expressing each VP3 protein from the native ORF initiating at the ATG start codon and ending at the TAA stop codon. pCMV$_1$-AAVxcmVP3 (x=2, 4, 5 and 11) are plasmids expressing each VP3 protein from a codon-modified ORF in which the AAP-VP overlapping ORFs were codon-modified to optimize VP3 expression in human cells. The modification resulted in extensive changes in amino acids coded by each AAP ORF with identities and the number of new stop codons being 37% and 6 for AAP2, 48% and 5 for AAP4, 47% and 4 for AAP5, and 44% and 8 for AAP11, respectively. This modification most likely abolishes functional AAP expression completely. Each VP3 ORF was cloned in the pCMV$_1$-AAV2VP3 parental plasmid used in our previous study (150) by replacing the AAV2VP3 ORF with a new VP3 ORF. The CMV-IE enhancer-promoters we used in the

134

study contained an intervening sequence (IVS) consisting of a splice donor, an intron and a splice acceptor from pIRES (Clontech, Mountain View, CA) for the AAP-expressing plasmids or the IVS from pAAV-MCS (Agilent, Santa Clara, CA) for the VP3-expressing plasmids. An Ad helper plasmid, pHelper, was purchased from Agilent. pHLP-AAV5(AAP5+) and pHLP-AAV5(AAP5-) are AAV5 helper plasmids carrying the AAV2 rep gene and the AAV5 cap gene. pHLP-AAV5(AAP5+) is the same as pHLP19-5 (341) that has been used for recombinant AAV5 vector production in our laboratory. In the pHLP19-5 helper plasmid, the AAV2 p5 promoter is moved from the native location to the downstream of the AAV2 polyadenylation signal and the TATA box sequence in the p5 promoter TATTTAA is replaced with the sequence GGGGGGG to reduce the expression of the large Rep proteins. pHLP-AAV5(AAP5-) is a derivative of pHLP-AAV5(AAP5+) that carries the codon-modified AAP-VP ORFs that abolishes functional AAP5 expression while preserving expression of the wild-type AAV5 VP1, VP2 and VP3 proteins. The identity of AAP5 amino acids coded by the native and codon-modified AAP-VP ORFs is 44% with 4 new stop codons being introduced within the AAP5 ORF. pHLP-Rep is a plasmid that expresses all the AAV2 Rep proteins in human embryonic kidney (HEK) 293 cells in the presence of co-transfected pHelper. pHLP-Rep was constructed by removing a 1.8-kilobase (kb) *Xho* I-*Xcm* I fragment from the wild-type AAV2 genome contained in our standard AAV2 helper plasmid pHLP19-2 (341), and expressed only AAV2 Rep proteins. pEMBL-CMV-GFP is an AAV vector plasmid for production of double-stranded AAV vector expressing eGFP under

the control of the CMV-IE enhancer-promoter, and is a gift from X. Xiao. pAAV-CMV-luc is an AAV vector plasmid for production of single-stranded AAV vector expressing firefly luciferase under the control of the CMV-IE enhancer-promoter and created from pAAV-MCS. Plasmids pdsAAV-U6-VBCx (x is an integer identification number indicating each different DNA barcode contained in each pdsAAV-U6-VBCx plasmid) are all double-stranded AAV vector plasmids carrying a 135-bp DNA fragment (nucleotide positions from 4445 to 4579 of pAAV9-SBBANN-VBCLib, GenBank accession code KF032296) that harbors a pair of 12-nucleotide-long DNA barcodes (Virus Bar Codes (VBCs)). Besides the DNA barcodes and flaking PCR primer binding sites, the pdsAAV-U6-VBCx vector plasmids contain a human U6 snRNA promoter-driven nonfunctional non-coding RNA expression cassette of 0.6 kb in length and a 1.0-kb stuffer DNA derived from the bacterial *lacZ* gene between the two AAV2 inverted terminal repeats. The pdsAAV-U6-VBCx vector plasmids were designed and created for the purpose of a separate study, and their feature of noncoding RNA expression was not utilized in this study. We have confirmed that expression of non-coding RNA from the pdsAAV-U6-VBCx plasmids does not affect AAV vector production in HEK 293 cells. $pCMV_1$ and $pCMV_3$ are control empty plasmids carrying no transgene in the $pCMV_1$ and $pCMV_3$ backbones, respectively. The native capsid ORFs used for the plasmid construction were from pHLP19-1 to 6 (341), AAV7, 8 and 9 helper plasmids provided by J. M. Wilson and G. Gao, AAV10 and 11 helper plasmids provided by S. Mori, and a plasmid containing a

de novo-synthesized AAV12 cap ORF provided by Voyager Therapeutics. The codon-modified AAP and VP3 ORFs were synthesized at GenScript.

Cells. HEK 293 cells (AAV293) were purchased from Stratagene. The HeLa human cervical cancer cell line and the Chinese hamster ovary (CHO)-K1 cell line were obtained from the American Type Culture Collection (ATCC). HEK 293 cells and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Lonza, Basel, Switzerland) supplemented with 10% fetal bovine serum (FBS), L-glutamine, and penicillin-streptomycin. The CHO-K1 cells were grown in F-12K medium supplemented with 10% FBS.

AAV viral particle production. AAV VP1/VP2/VP3 particles and "VP3 only" particles were produced in HEK 293 cells by an adenovirus-free plasmid transfection method (222) with modification. In brief, we changed the complete culture media to serum free media immediately before transfection, transfected cells with a mixture of a required amount of each plasmid DNA by polyethyleneimine (PEI) at a DNA:PEI weight ratio of 1:2, and harvested both media and cells for viral particle recovery five days post-transfection. The plasmids used for the production of each viral particle preparation are described in each relevant section. AAV5(AAP5-)-CMV-luc and AAV5(AAP5+)-CMV-luc vectors were produced using one 225-cm2 flask, and concentrated from an initial volume of 25 ml to a final volume of 250 μl using Amicon Ultra Centrifugal Filter Units with a molecular weight cut off of 100 kDa. For transmission electron

microscopy (TEM), we produced "VP3 only" virus-like particles using fifteen 225-cm2 flasks. The harvested media and cells underwent one cycle of freezing and thawing and the cell debris was removed by centrifugation at $10,000 \times g$ for 15 min. The culture medium supernatants were made to 8% polyethylene glycol (PEG) 8000 and 0.5 M NaCl, incubated on ice for 3 h, and spun at $10,000 \times g$ for 30 min to precipitate viral particles. The pellets were resuspended in a buffer containing 50 mM Tris-HCl (pH 8.5) and 2 mM MgCl2, treated with Benzonase (EMD Millipore, Darmstadt, Germany) at a concentration of 200 units per mL for 1 h, and subjected to purification by three rounds of cesium chloride (CsCl) density-gradient ultracentrifugation (341) followed by dialysis with a buffer (25 mM HEPES, 150 mM NaCl, pH 7.4) for TEM.

Cell infection. CHO-K1 cells seeded on a 96-well plate were infected with AAV5(AAP5-)-CMV-luc or AAV5(AAP5+)-CMV-luc at a multiplicity of infection (MOI) of 1x106 in the absence of a helper virus. Forty-eight hours post-infection, luciferase activity was quantified using the Bright-GloTM Luciferase Assay System (Promega, Madison, WI) and the CentroXS LB960 plate reader (Berthold, Bad Wildbad Germany). The data were collected from a biologically triplicated experiment.

Immunofluorescence microscopy. HeLa cells were seeded on coverslips in 12-well plates and transfected with plasmid DNA using PEI. Forty-eight hours after transfection, the cells were fixed with 4% paraformaldehyde at room temperature,

permeabilized with 0.2% Tween 20, and blocked with 8% bovine serum albumin (BSA). For AAP localization, the cells were stained with mouse monoclonal anti-FLAG M2 antibody (F1804; Sigma-Aldrich, St. Louis, MO) and rabbit polyclonal anti-nucleostemin antibody (sc-67012; Santa Cruz Biotechnology, Dallas, TX), followed by DAPI (4',6-diamidino-2- phenylindole), Alexa Fluor 488-AffiniPure goat anti-mouse IgG antibody (115-545-166; Jackson ImmunoResearch, West Grove, PA), and Cy3-AffiniPure goat anti-rabbit IgG antibody (111-165-144; Jackson ImmunoResearch). For imaging of AAV capsids, cells were stained with mouse monoclonal anti-AAV2 capsid antibody (A20 clone, 03-61055; American Research Products Inc., Waltham, MA), mouse monoclonal anti-AAV4 capsid antibody (ADK4 clone, 03-610147; American Research Products Inc.), mouse monoclonal anti-AAV5 capsid antibody (ADK5a clone, 03-61048; American Research Products Inc.), mouse monoclonal anti-AAV8 capsid antibody (ADK8 clone, 03-651160; American Research Products Inc.), or mouse monoclonal anti-AAV8/9 capsid antibody (ADK8/9 clone, 03-651161; American Research Products Inc.) and rat monoclonal anti-DYKDDDDK (FLAG) antibody (NBP1-06712; Novus Biological, Littleton, CO), followed by DAPI, Alexa Fluor 488-AffiniPure goat anti-mouse IgG antibody (115-545-166; Jackson ImmunoResearch) and Cy3-AffiniPure donkey anti-rat IgG antibody (712-165-153; Jackson ImmunoResearch). The nucleolus was visualized by rabbit polyclonal anti-nucleostemin antibody (sc-67012; Santa Cruz Biotechnology) and Alexa Fluor 647-AffiniPure goat anti-rabbit IgG antibody (111-605-144; Jackson ImmunoResearch). GFP was directly visualized with fluorescence microscopy.

The cells were imaged on a Zeiss LMS 710 laser scanning confocal microscope using either a 63x/1.4 NA or a 100x/1.46 NA objective.

Quantitative dot blot. HEK 293 cells were seeded on 6-well plates one day before transfection. We changed complete culture media to serum free media before transfection and transfected cells with 0.4 μg each of the following 5 plasmids, pCMV$_3$-FLAG-cmAAPx (x=1 to 12) or pCMV$_3$ (an empty plasmid as no AAP control), pCMV$_1$-AAVxVP3 (x=1 to 12), pEMBL-CMV-GFP, pHLP-Rep and pHelper using PEI. Five days post-transfection, we harvested both media and cells, disrupted cells by one cycle of freezing and thawing, and released viral particles into the culture media. After cell debris was removed by centrifugation at 21,100 × g for 5 min, 200 μl of the culture medium supernatant was subjected to nuclease treatment with 200 units per mL of Benzonase at 37°C for 4 h, followed by proteinase K treatment at 55°C for 1 h. Viral genome DNA was purified by phenol-chloroform extraction, ethanol-precipitated, and dissolved in 1 x Tris-HCl EDTA (TE) buffer (pH 8.0). The viral DNA and linearized standard plasmid DNA were then denatured with 0.4 N NaOH, blotted on a Zeta Probe nylon membrane (Bio-Rad, Hercules, CA), and hybridized with a ^{32}P-labeled GFP probe. The hybridized signals were imaged and quantified using a Typhoon FLA7000 scanner (GE Healthcare Bio-Science, Uppsala, Sweden). The negative control contained double-stranded AAV-CMV-GFP genomes that had undergone AAV2 Rep-mediated replication but were not protected by viral capsids. The negative control ensured efficient nuclease digestion in the assay.

AAP-VP cross-complementation analysis. For the AAP-VP cross-complementation analysis, we used two methods; the Illumina sequencing-based AAV Barcode-Seq (261) and the conventional quantitative dot blot assay as described above. In the AAV Barcode-Seq, HEK 293 cells were seeded on 12-well plates one day before transfection and "VP3 only" particles containing a DNA-barcoded viral genome were produced essentially in the same manner as described earlier except that we used 0.24 μg of each plasmid, pdsAAV-U6-VBCx in place of pEMBL-CMV-GFP, and pCMV$_3$-FLAG-AAP2 in place of pCMV$_3$-FLAG-cmAAP2. There were a total of 132 AAP and VP combinations and each combination received different DNA-barcoded AAV vector plasmid, pdsAAV-U6-VBCx (x=1 to 132). Five days post-transfection, we harvested both media and cells and pooled them in a bottle. We performed this procedure in triplicate and produced three pooled samples. Viral genome DNA was extracted from 200 μl of each pooled sample and purified by phenol-chloroform extraction, ethanol-precipitated, and dissolved in 20 μl of 1 × TE buffer (pH 8.0). We then PCR-amplified both the left-VBCs (lt-VBCs) and the right-VBCs (rt-VBCs) separately using two different sets of PCR primers and 2 μl each of the resulting DNA solution. The PCR primer sequences are as follows: lt-VBC-For (FSN-SBC-ACCTACGTACTTCCGCTCAT), lt-VBC-Rev (FSN-SBC-TCCCGACATCGTATTTCCGT), rt-VBC-For (FSN-SBC-ACGGAAATACGATGTCGGGA), and rt-VBC-Rev (FSN-SBC-CTTCTCGTTGGGGTCTTTGC). Each primer contained 7-nucleotide-long

Sample-specific Bar Code (SBC) and a 1 to 5 Frame-Shifting Nucleotides (FSN) at the 5' end. The primer combinations, lt-VBC-For and lt-VBC-Rev, and rt-VBC-For and rt-VBC-Rev, were used to amplify the lt-VBCs and rt-VBCs, respectively, in each of the biologically triplicated sets of the experiment. The resulting six PCR products were mixed at an equimolar ratio and subjected to multiplexed Illumina sequencing as previously described (261) together with other PCR products prepared in the same manner in separate AAV Barcode-Seq studies. One to five μg of PCR products attached to Illumina sequencing adaptors were sent to Elim Biopharmaceuticals Inc. (Hayward, CA) and sequenced with a 50-cycle single-end run on an Illumina HiSeq 2500. The quality measures of Illumina raw sequence reads determined by FastQC (i.e., per base sequence quality, per sequence quality scores, per base N content, and sequence length distribution) were all met in the dataset used in the study. We analyzed the Illumina sequencing data at the Pittsburgh Supercomputing Center using an algorithm we developed. In this experimental scheme, a pair of the DNA barcodes carried by each AAV vector plasmid could provide a measure of AAV vector yield from each individual transfection by means of the AAV Barcode-Seq.

Barcode-Seq data analysis. We determined relative viral particle yields in each of the triplicated sets of the experiment using the same principle as that used for our previous study (261). First, we globally normalized Illumina raw sequence read numbers of all the 132 VBCs to obtain a relative read number data for each of the lt-VBCs and rt-VBCs. We then adjusted the relative read number data for

each of the lt-VBCs and rt-VBCs by each VBC-specific PCR amplification efficiency factor. The VBC-specific PCR amplification efficiency factor was determined in the following manner. We created two sets of an equimolar mixture of 379 pdsAAV-U6-VBCx plasmids (x=1 to 379) independently. The 132 pdsAAV-U6-VBCx plasmids (x=1 to 132) used in the study were included in each equimolar plasmid mixture. We then PCR-amplified 379 lt-VBCs together using the primers lt-VBC-For and lt-VBC-Rev, and 379 rt-VBCs together using the primers rt-VBC-For and rt-VBC-Rev, in each of the two equimolar plasmid mixtures. The resulting four PCR products were mixed at an equimolar ratio and subjected to multiplexed Illumina sequencing together with other PCR products prepared in the same manner in separate AAV Barcode-Seq studies as described earlier. This gave us raw sequence read numbers of all the 132 lt-VBCs and 132 rt-VBCs in each set. We then globally normalized the sequence read numbers and determined a relative quantity value of each lt-VBC and each rt-VBC in each set. The relative quantity values of each VBC obtained from the duplicated sets of the experiment were averaged and used as the PCR amplification efficiency factor. Since the experiment was done in triplicate and two DNA barcodes, lt-VBC and rt-VBC, were used, we obtained a total of 6 relative quantity values that could quantify AAV vector yield in each AAP-VP3 combination. Among the 6 values in each AAP-VP3 combination, we excluded outliers showing values more than three times the interquartile range beyond the upper and lower quartiles. AAV vector yield of each AAP-VP3 combination relative to the native combination was determined in each serotype.

Transmission electron microscopy. The viral particle preparations purified by three rounds of CsCl density-gradient ultracentrifugation followed by dialysis were spun at $6,100 \times$ g. for 10 min to remove any viral precipitate, and carbon-coated copper grids (Pacific Grid-Tech Cu-300CN, San Francisco, CA) were glow-discharged for 25 s at 15 mA using the PELCO easiGlowTM Glow Discharge Cleaning System (Ted Pella Inc., Redding, CA) immediately prior to use. Four μl of each sample was placed onto the grids for 3 min and manually blotted with a Whatman filter paper (1001-125; GE Healthcare, Pittsburg, PA). The grids were then placed face down onto 45 μl drops of a buffer (50mM HEPES, 25mM MgCl2, 50mM NaCl, pH 7.4) for 30 s, then washed three times similarly with distilled water, then blotted. The samples on the grids were stained with 5.5 μl of 0.75% uranyl formate (pH 4.5) for 30 s, washed with distilled water, followed by blotting. The grids were allowed to dry, then stored at room temperature in a petri dish sealed with Parafilm until imaging.

Western blot. We seeded HEK 293 cells on 6-well plates and transfected them with a total of 2 μg of a plasmid or a mixture of plasmids using PEI. The plasmids used in each experiment are described in each relevant section. Forty-eight hours post-transfection, we lysed HEK 293 cells in radioimmunoprecipitation assay (RIPA) buffer containing protease inhibitors (Complete Mini; Roche, Indianapolis, IN), sonicated, and determined protein concentrations in the cell lysates with a DC Protein Assay Kit (Bio-Rad, Hercules, CA). The same amount

of total cell lysates (40 µg per lane) was separated along with a molecular weight marker on a 10% SDS-PAGE gel, transferred onto a polyvinylidenedifluoride (PVDF) membrane, and reacted with mouse monoclonal anti-FLAG M2 antibody and mouse anti-α-tubulin antibody (sc-32293; Santa Cruz Biotechnology) or mouse monoclonal anti-AAV VP1/VP2/VP3 antibody (B1) (03-61058; American Research Products) and rabbit polyclonal anti-cyclophilin A antibody (Cell Signaling Technology, Danvers, MA) followed by goat polyclonal anti-mouse IgG antibody (sc-2055; Santa Cruz Biotechnology) or goat polyclonal anti-rabbit IgG antibody (sc-2004; Santa Cruz Biotechnology) conjugated to horseradish peroxidase. The signals on the blots were visualized with Immobilon Western Chemiluminescent HRP Substrate (EMD Millipore) and imaged on X-ray films or using the FluorChem M system (ProteinSimple, Santa Clara, CA). Molecular weights of FLAG-tagged AAP1 to 12 were determined in triplicated western blots.

Statistics. In the cross-complementation study using AAV Barcode-Seq, the null hypothesis that there was no enhancement of assembly by AAP was examined by one-tailed Mann-Whitney U-test in each serotype. In the AAV5 vector production and infection assays, we used the two-tailed Welch's t-test to assess differences in the mean values between two groups. P values of < 0.05 were considered statistically significant.
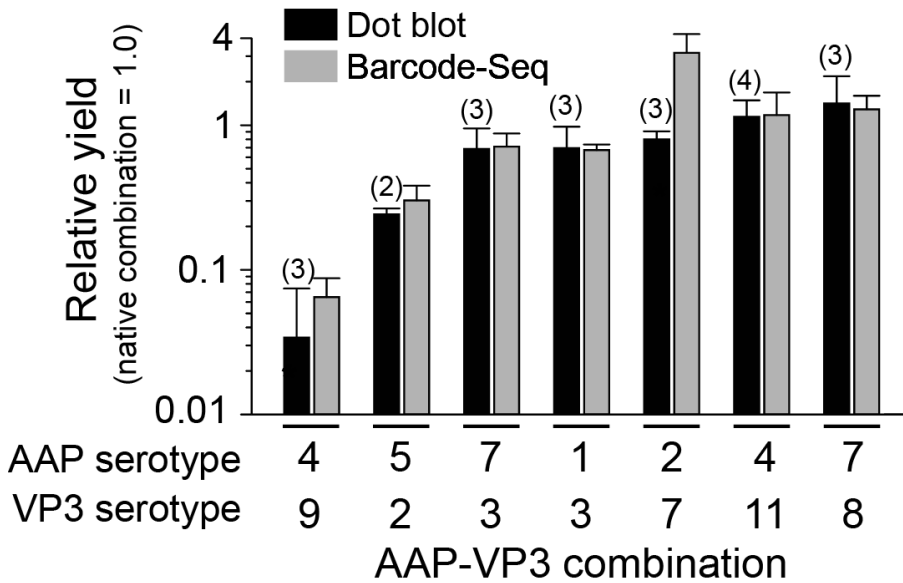
Fig. S1 AAP-VP3 cross-complementation analysis by a quantitative dot blot assay.
A total of 7 AAP-VP3 combinations were also analyzed for cross-complementation of capsid assembly by a quantitative dot blot assay in a biologically replicated experiment. The number of replicates in each AAP-VP combination is indicated in parentheses. The results were compared to those obtained by AAV Barcode-Seq shown in Fig. 3.8 and Fig. S2. Values represent relative AAV "VP3 only" particle yields in each AAP-VP combination relative to those obtained by the native AAP-VP3 combination. Black and gray bars represent the results obtained from dot blot and AAV Barcode-Seq, respectively. Error bars represent standard deviations except for the AAP5-AAV2 VP3 combination. For the AAP5-AAV2 VP3 combination, the error bar represents the difference between each value and the mean value because the data were collected from biologically duplicated samples.
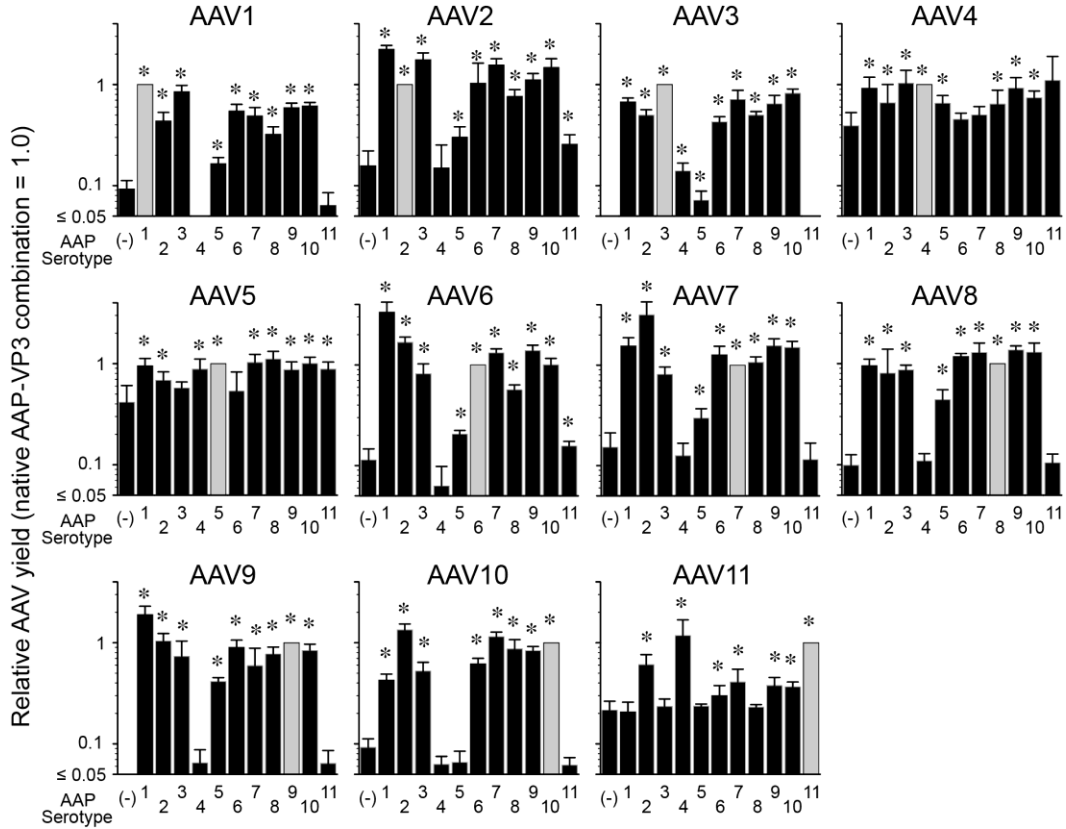
Fig. S2 AAV Barcode-Seq data on AAP-VP3 cross-complementation between AAV1 to 11.
The ability for each AAP to assemble homologous and heterologous VP3 proteins derived from
AAV1 to 11 is shown. All possible 12 combinations including no AAP controls were assessed by
AAV Barcode-Seq. No AAP controls are shown as AAP (-) to the leftmost in each panel. Y-axis
shows AAV "VP3 only" particle yields in each AAP-VP combination relative to the yields
obtained with the native AAP-VP3 combination. Error bars represent standard deviations.
Asterisks indicate that values are higher than those of the no AAP controls with P < 0.05 (one-
tailed Mann-Whitney U-test).

Table S1. Theoretically and experimentally determined molecular weights of FLAG-tagged AAPs

Molecular weight (Mw)	AAP											
	1	2	3B	4	5	6	7	8	9	10	11	12
Theoretical (kDa)	22	24	23	19	22	22	22	22	22	22	20	20
Experimentally determined (kDa)[a]	27 ± 0.6	27 ± 0.6	28 ± 0.7	31 ± 0.5	34 ± 0.7	27 ± 0.4	27 ± 0.6	30 ± 0.4	33 ± 0.4	34 ± 0.8	30 ± 0.8	30 ± 1.3
Difference[b] (%)	+27	+17	+24	+63	+55	+26	+29	+36	+54	+59	+58	+51

[a]Molecular weight (Mw) of each FLAG-tagged AAP was determined by western blot analysis of three biologically replicated experiments.
[b]Percentage increase of Mw determined by western blot analysis compared to the theoretical Mw.

Table S2. Profiles of isoelectric point (pI) values in the basic amino acid-rich region in the C-terminus of AAPs.

AAP BR1 - 5				AAP BR1-6 toward the C-term			
	No. of residues				No. of residues		
AAP[a]	H, K or R	D or E	pI	AAP[a]	H, K or R	D or E	pI
AAP5 (NoE[b])	8	0	11.45	AAP5 (NoE)	15	1	11.86
AAP9 (NoE)	11	1	12.37	AAP2	18	2	12.26
AAP3B	15	0	12.43	AAP8	15	0	12.26
AAP8	12	0	12.43	AAP3B	18	1	12.40
AAP7	13	0	12.52	AAP7	16	0	12.48
AAP10	14	0	12.52	AAP10	17	0	12.48
AAP6	13	0	12.65	AAP9 (NoE)	14	1	12.52
AAP2	15	0	12.70	AAP6	16	0	12.57
AAP11	13	0	12.70	AAP11	13	0	12.70
AAP4	14	0	12.78	AAP1	15	0	12.74
AAP1	12	0	12.96	AAP4	14	0	12.78
AAP12	13	0	13.04	AAP12	13	0	13.04

[a]The order of AAPs is sorted by pI (the smallest to the largest). pI values were determined using BioPerl pICalculator.
[b]NoE, nucleolar excluded.

Chapter 4: Discussion, Conclusions, and Future Directions

4.1 Discussion and Future Directions

Nuclear and nucleolar localization signals in AAP

As discussed at the end of Chapter 2, the C-terminus of AAP2 contains multiple overlapping NLS and NoLSs that contribute to the robust accumulation of this protein within the nucleolus. BR1 is highly conserved across the AAPs in serotypes 1-12 (**Fig 3.2**) as the motif (K/R)-(S/L)-(K/R)-R and probably accounts for the nuclear localization seen in all of these AAPs (**Fig 3.3**). The conservation of this strong classical nuclear localization signal denotes the importance of targeting AAP to the nucleus where capsid assembly must occur to result in proper packaging of the AAV genome. In this light, the nuclear targeting of an important assembly factor likely reflects the evolutionary pressure to maintain capsid assembly in the nucleus and prevent non-productive assembly in the cytoplasm. Once assembled, there is no known mechanism for AAV capsids to enter the nucleus without undergoing conformational changes associated with the endosomal environment that lead to VP1 externalization, hence capsids assembled in the cytoplasm would be unable to reach the sites of AAV genome replication.

The NoLS in AAP2 and other AAPs is especially interesting in context of the wider knowledge and discussion about NoLSs. Since the nucleolus is not a membrane-bound organelle, it is doubtful that there are specific proteins that regulate traffic as the importins do for the nucleus. Instead, it is likely that proteins capable of diffusion into the nucleolus are retained there by resident nucleolar proteins. Given that the main consensus for NoLSs is the presence of

positively charged amino acids and that several proteins can associate with the nucleolus in a non-functional manner though electrostatic interactions (300), the localization of AAP to the nucleolus may simply be a consequence of its multiple NLSs. Regardless, the correlation between nucleolar targeting and high pI values in the C-terminus of AAPs 1 to 12 support the idea of net positive electric charge being the main determinant in nucleolar localization or exclusion of AAP (**Fig. 3.2, 3.3, Table S2**). In addition to the effect of point mutations on AAP2 nucleolar localization, the swapping of this C-terminal region can retarget AAP. When amino acids 144-204 in AAP2 are replaced from the corresponding amino acids in AAP8 or 9, the protein localization changes from tightly nucleolar to diffusion throughout the nucleus (**Appendix Fig i**), again supporting the previous results showing that this region is the main determinant in AAP localization.

Localization of AAP capsid assembly

The first study of sub-cellular localization of AAV2 capsids during assembly found that capsids initially appear in nucleoli and that both Rep and AAV-DNA are primarily nucleolar-excluded (148). AAV2 capsids are associated with several nucleolar proteins including nucleophosmin (134, 342), nucleolin (284, 342), and nucleostemin (150). At least in the case of nucleostemin, this association remains after the capsids have left the nucleolus and migrate into the cytoplasm (**Appendix Fig ii**). Nucleolin and nucleophosmin co-purify with viral particles, indicating that this association is also not restricted to the inside of the nucleus (284, 342). These observations have led to speculation on a possible role for the nucleolus in AAV2 capsid assembly. Intriguingly, AAV2 capsids do not require

nucleolar targeting for assembly per se. When AAPmt26 was used in place of AAP2, AAV2 capsids were still associated with nucleostemin and a portion were localized to the nucleolus, even though VP3 and AAP were nucleolar-excluded (**Appendix Fig 2.8 and iii**). Since nucleostemin can be found in low amounts in the nucleoplasm, these interactions could have taken place outside the nucleolus. After binding the capsid surface, nucleostemin could have transported the capsids into the nucleolus and retained them there. Hence, if there is a role of the nucleolus in AAV2 assembly, it may be the requirement for specific nucleolar proteins and not the need to take place in the nucleolus. This notion is supported by the sub-nuclear bodies seen in HeLa cells during capsid assembly that stain positive for nucleostemin, AAP2, and AAV2 capsids. These sub-nuclear bodies also stain positive for nucleolin (**Appendix Fig iv**), but not positive for fibrillarin (Anna Maurer, personal correspondence), therefore the recruitment of nucleolar proteins is likely specific, but this needs to be investigated in further detail before definitive conclusions can be made. It is noteworthy that while AAP2 is associated with nucleostemin, unincorporated VP3 is not. The main question to address is whether these AAV2-associated nucleolar proteins are required during assembly or only bind to capsids after assembly. There are a few ways to approach this subject. First, the nuclear excluded AAP2mt12 has a ~10-fold reduction in protein levels and no discernable capsid assembly (**Fig 2.6**). If the protein levels of AAP2mt12 could be maintained to near wtAAP2 protein levels, perhaps with different amino acid substitutions or with MG132, it would be informative to know if this mutant was still capable of promoting capsid assembly

outside the nucleus. If this were true, it would indicate that nuclear and nucleolar host factors are not required for capsid assembly. A more direct method would be to measure capsid production after RNAi knockdown of specific nucleolar proteins.

When we characterized capsid assembly localization and nucleostemin association in AAV serotypes 4, 5, 8, and 9, we found that none of these capsids had association with nucleostemin and only AAV4 capsids were enriched in the nucleolus. In addition, the AAPs from these serotypes were also not associated with nucleostemin and in the case of AAP5 and AAP9, were specifically nucleolar excluded (**Fig 3.3, 3.5**). Based on these findings it may be that only AAV2 (or possibly other Clade B AAVs) have an association with nucleolar proteins. Clearly, more investigation is required to determine if this is the case since in the studies presented here, only one nucleolar protein was probed for and other proteins such as nucleolin or nucleophosmin may still be associated with these AAPs or capsids. Unfortunately, there are a limited number of available anti-AAV antibodies, which constrains the ability to do more co-localization studies. Recently, an anti-AAV6 antibody became commercially available. AAV6 is a Clade A AAV closely related to AAV2 and it also has a strongly nucleolar localized AAP (**Fig 3.3**) much like AAV2. Immunofluorescence studies like the ones conducted in Chapter 3 would reveal if this AAV also has nucleolar-localized capsid assembly and association with nucleolar proteins.

It is important for the AAV field to know if the association of AAV2 with the nucleolus is a significant factor in assembly for other AAVs, if it is significant

only for AAV2, or if it is merely a consequence of having a positively charged region on the capsid surface. AAV2 binds its primary receptor, heparan sulfate proteoglycan (HSPG), though a patch of arginines and lysines near the spikes at the 3-fold axis (343). This region of highly positively charged amino acids interact with the negatively charged HSPG molecule to mediate weak binding of the virus to the cell surface. Nucleophosmin has two acid rich regions of amino acids that are capable of mediating interactions with other proteins by electrostatic interactions. In conjunction, nucleolar localization is likely mediated though interactions between its negatively changed nucleolar components and positively charged proteins (325), thus the association of AAV with nucleolar proteins may be a consequence of the positively charged amino acids required for initial receptor binding. AAV2 and 3 use HSPG as a primary receptor, AAV1, 4, and 5 all use sialic acid, while AAV6 can use both (226). Sialic acids are also negatively charged molecules, but there is clear a difference in binding ability between AAV serotypes and this may also reflect in the ability of the virus to bind nucleolar proteins. Aside from the already proposed RNAi and AAV6 staining experiments, another way to examine this question is to take advantage of the commonly used AAV2 R585E mutant, which is unable to bind HSPG. An absence of co-localization between AAV2 R585E and nucleostemin or nucleophosmin would strongly indicate that this hypothesis is true. Footprint mapping of nucleolar proteins in complex with AAV2 would also be informative to see if these interaction regions overlap with the receptor binding area. While none of these experiments on its own will conclusively prove the role of nucleolar proteins in

the assembly of AAV, the preponderance of evidence generated by the results may inform the field if the nucleolus is worth targeting as tactic to improve vector production yields.

AAP dependent vs independent capsid assembly

Previous to this work, the majority of information on capsid assembly came from studying the nucleolar-localizing, AAP-dependent AAV2. These studies have uncovered that in the context of infection with AAV2 and Ad, VPs exist predominately in monomeric and oligomeric forms in the cytoplasm and upon entering the nucleus, they are able to form capsids if AAP is present. VP3 is capable of forming large non-capsid oligomers, the nature of which are still unknown (153), but are not dependent upon the C-terminal 2-fold axes interactions (44). A recent study looking only at sucrose density gradients of VP3, with or without AAP, found that although the presence of AAP did not change the amount of VP monomers, it did increase the amount of intermediately sized oligomers and importantly, it induced A20 binding in these pools of VP3s (44). This binding was specific for A20, as the other AAV2 capsid antibodies C24 and C37 did not bind in the presence of VP3 alone (**Fig 4.1B**). The epitope for the A20 antibody has previously only been found in fully assembled capsids (291), but in the immunoblotting done by Naumer et al., AAP is inducing A20 binding in VP complexes before the capsids are fully assembled, as observed by EM (**Fig 4.1C**). Interestingly, both the A20 conformational epitope and AAP-VP binding site are located at the 2-fold axis (44, 293), so it may be that AAP is able to induce dimers of trimers that can lead to further capsid assembly, although the

A20 binding domain also covers areas near the 5-fold axis (**Fig 1.8A**). A model in which VP trimers are formed and then undergo a conformational change induced by AAP that allows for rapid capsid assembly in the nucleus would be in line with the capsid assembly pathway for the closely related autonomous parvovirus MVM, in which the CBBs are trimers that form in the cytoplasm, but are incapable of assembly until entering the nucleus and undergoing a conformational change (135, 344). Naumer et al. suggest that the role of AAP might be to assist VP interactions at the 3-fold or 5-fold interfaces since these axes consist of complex folds between the VP monomers (44), as opposed to the relatively simple interactions at the 2-fold axis. Under this hypothesis, it may be that the VPs of AAV4, 5, and 11 are able to create these complex structures at the 3-fold or 5-fold on their own, but AAP can act to help over come the required energy barriers and lead higher capsid formation during assembly.
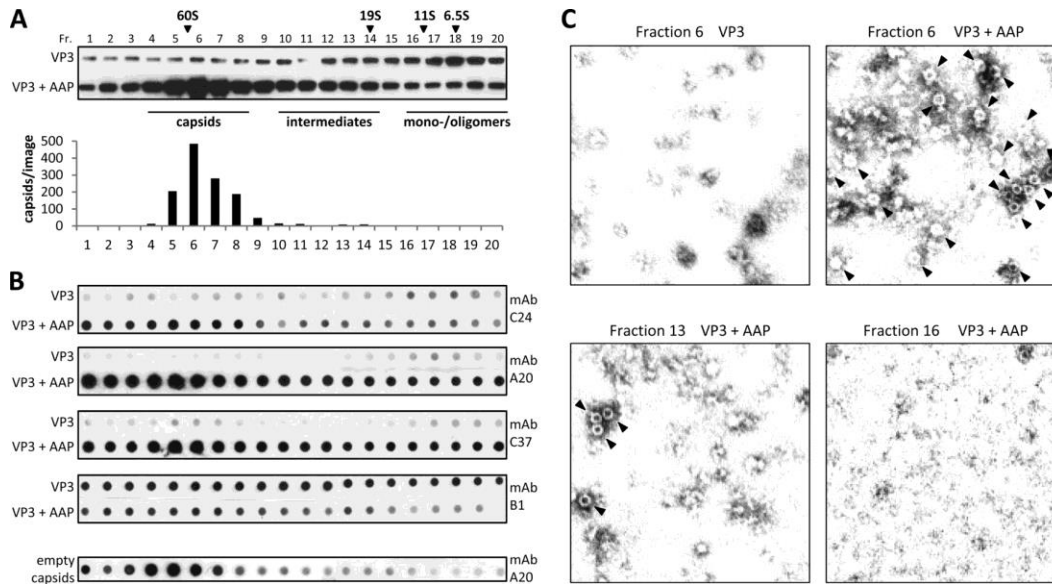
Figure 4.1 Analysis of VP3 conformations in the presence or absence of AAP.
Figure is reprinted with permission from Naumer et al., 2012.
(**A**) Western blot analysis of sucrose density gradient fractions (Fr. 1 to 20) of lysates from 293T cells expressing VP3 alone or in combination with AAP. VP3 was detected using MoAb B1. The same amount of each fraction was analyzed under the same conditions. A reference gradient with purified proteins with known S values was used for calibration. The distribution of assembled AAV2 capsids (number of capsids per image) was analyzed by quantification of at least three representative electron microscopy images from each gradient fraction.
(**B**) Native dot blot analysis of sucrose density gradient fractions from cell lysates containing VP3 or VP3 and AAP, respectively. The same amount of each fraction was dotted onto nitrocellulose membranes under the same nondenaturing conditions and detected by the capsid-specific antibodies C24, A20, and C37 or the B1 antibody, which is specific for a linear epitope at the VP3 C terminus, in order to draw conclusions about AAP-induced conformational changes. The distribution of purified empty AAV2 capsids served as a control.
(**C**) Electron microscopy images of sucrose density gradient fractions containing VP3 alone (fraction 6) or in combination with AAP (fractions 6, 13, and 16). Capsids are indicated by arrowheads.

The surprising result of finding AAP-independent capsid assembly in AAVs 4, 5, and 11 is an exciting opportunity to gain further insight into AAV capsid assembly. These serotypes come from two distinct AAV groups, with AAV5 in its own group and AAV4, 11, and 12 in the AAV4 group (9) (**Fig 1.2**), showing that AAP-independence is not restricted to only one linage of AAVs and that other AAVs may also be capable of AAP-independent assembly. The variance between AAV11 and AAV12 is a unique chance to determine the differences in AAP-dependent and independent capsid assembly in these VP proteins and delineate the place of AAP in capsid assembly. As mentioned above, there are several axes that AAP could be functioning on and by identifying the regions of difference between 11 and 12 VPs that are responsible for AAP-independent assembly, we might be able to determine which axis these differences are placed at. The amino acids sequences between 11 and 12 VP3 are 87% identical and, more importantly, of the same length. Thus, swapping regions between the two proteins to create chimeric 11_12 VP3s is a good initial strategy for identifying the region that allows for AAP-independent assembly. A previous attempt at this with the VP3 between differently sized AAV2 and AAV5 VP3s was unsuccessful as none of the chimeras could form capsids, regardless of the presence of AAP (**Appendix Fig v**). Digestion of the pCMV$_1$-AAV11cmVP3 and pCMV$_1$-AAV12VP3 plasmids used in Chapter 3 with *Pvu*II allowed for a swap between the two plasmids to easily create 11 and 12 chimeric VP3s (**Appendix Fig vi**). A preliminary assessment of the ability of these chimeric VP3s to form capsids without AAP shows that the C-terminus of the AAV11 VP3 is the region that is

responsible for, or at least involved in, AAP-independent assembly (**Appendix Fig vi**). When the different amino acids are mapped on the crystal structure of AAV4 VP3 (structures for AAV11 and AAV12 VP3 are not currently available), the resulting model shows several areas of divergence involved in VP-VP interactions (**Appendix Fig vi**). Especially interesting are the amino acids K718, T720, P722, and Y729 which are involved in dimer interactions along the 2-fold axis and reside inside a pocket that is adjacent to and over-laps the AAP binding region identified by Naumer et al. (44). Another interesting cluster are amino acids A657, V665, and I669 which are at the 5-fold axis (**Fig. 1.4D, E**). Future work to precisely determine the amino acids that are critical for AAP-independent assembly, in accompaniment with protein models and biochemical studies, could reveal the VP-VP interaction sites that require AAP assistance.

Other functions of AAP

Aside from its role in capsid assembly, it is possible that AAP has other roles in the AAV life cycle. There is circumstantial evidence supporting the notion that AAP is a DNA binding protein that can remodel host chromatin. When transfected into HeLa cells, AAPs 1, 5, 6, 7, 8, 9, can induce condensation of DAPI staining (CDS) in correlation with strong over-expression, whereas AAPs 2, 4, 12 do not (**Appendix Fig vii** (AAP2 not pictured)). This change in chromatin structure is unlikely to be a sign of apoptosis since these cells lack activated caspase-3 staining and have intact nuclear membranes by gross examination (**Appendix Fig vii**). When amino acids 144-204 in the C-terminal of AAP2 are replaced with the corresponding amino acids from AAP8 or AAP9, the chimeric

AAP2.8 and 2.9 can cause CDS, indicating that this is the region responsible for inducing the chromatin remodeling (**Appendix Fig vii**). Direct evidence that AAP can bind DNA, histones, or induce other mechanisms of chromatin remodeling is still lacking, but could be directly tested via an EMSA or ChIP-seq techniques. Interestingly, AAP8 and Snake AAV AAP can co-localize with DNA during mitosis, but AAPs 2, 4, 5, and 10 do not (**Appendix Fig viii**).

If AAP is not directly binding to DNA, it may be interacting with host proteins and indirectly inducing the CDS. The FLAG-tagged AAP constructs in the Nakai lab are a valuable resource that can be utilized to identify AAP protein-protein interactions. Commercially available FLAG immunoprecipitation kits can be used to pull down potential AAP interacting partners en masse for identification by mass spectrometry. The elution by excess FLAG peptide should reduce the non-specific interactions seen when potential interaction partners are eluted from protein A/G beads via boiling or pH (Earley, data not show). Isolating AAP interaction partners may also help identify host proteins involved in AAV capsid assembly. Attempts at *in vitro* AAV capsid assembly using only AAV2 VP proteins were largely unsuccessful, but HeLa cell cytoplasmic lysate did moderately stimulate capsid formation indicating that there may be some host factors that can assist in this process. The addition of AAP or nuclear lysate would likely result in more efficient capsid assembly, but this experiment has not been attempted yet.

4.2 Conclusions

The work presented here significantly expands the limited knowledge about the role of AAP in capsid assembly for AAV serotypes 1 to 12. We initially began this work by delineating the NLS and NoLSs in the C-terminus of AAP2 and we identified three distinct NLSs that could target AAP to the nucleus. In conjunction, the amino acids that constituted these NLSs were also involved in nucleolar localization, likely based on the number of positively charged amino acids instead of a specific sequence. The corresponding regions in AAPs 1 to 12 contain the conserved NLS in BR1, but vary in conservation of the remaining BR regions, which influences their nucleolar localization or exclusion. We also showed that mutations or deletions in this region in AAP2 result in decreased AAP protein and correspondingly decreased titer, indicating that this region is important for maintaining optimal levels of AAP protein during capsid assembly.

When we expanded our work to the other AAPs in serotypes 1 to 12, we found AAV4, 5, and 11 were capable of AAP-independent assembly and that the region responsible for this independence in AAV11 lies within the last 197 amino acids of the C-terminus. Additionally, nucleolar localization of AAP and AAV capsids are not generalizable to all AAV serotypes. Specifically, AAV5 and AAV8 had nucleolar excluded AAP and AAV5, 8, and 9 had nucleolar excluded capsids. Only AAV2 capsids were associated with nucleostemin. Taken together, these findings challenge current viewpoints about the importance of the nucleolus in the life cycle of non-AAV2 serotypes and demonstrate that AAV biology is more heterogeneous than previously appreciated.

Appendix

Contained in this section is a collection of data that are unpublished, but important to support the conclusions made in this dissertation and also to act as preliminary work for future AAP studies. The materials and methods for this section are included at the end, starting at page 170.

Figure i: Intracellular localization of AAP2.8 and AAP2.9 chimeric AAPs
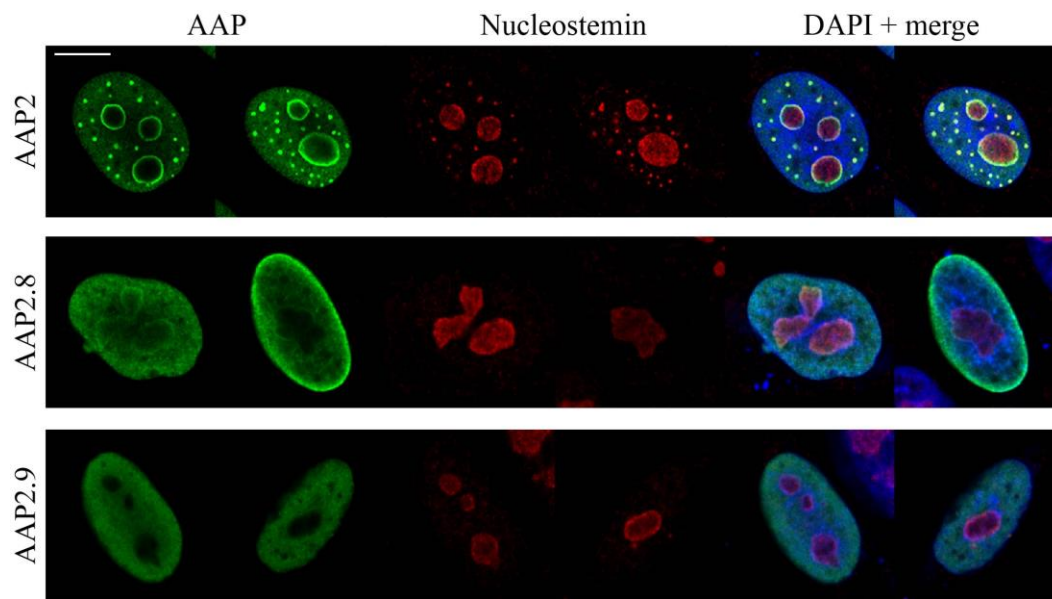
Figure i: Intracellular localization of AAP2.8 and AAP2.9 chimeric AAPs
HeLa cells were transfected with the following plasmids: (**A**) pCMV$_3$-FLAG-AAP2, (**B**) pCMV$_3$-FLAGcmAAP2.8, or (**C**) pCMV$_3$-FLAGcmAAP2.9. The cells were fixed at 48 hours post-transfection and immunostained with anti-FLAG and anti-nucleostemin antibodies. Scale bar, 10 μm

Figure ii: Intracellular localization of AAV2 and nucleostemin capsids with

Rep (A) and during co-infection with Ad (B)

A

B
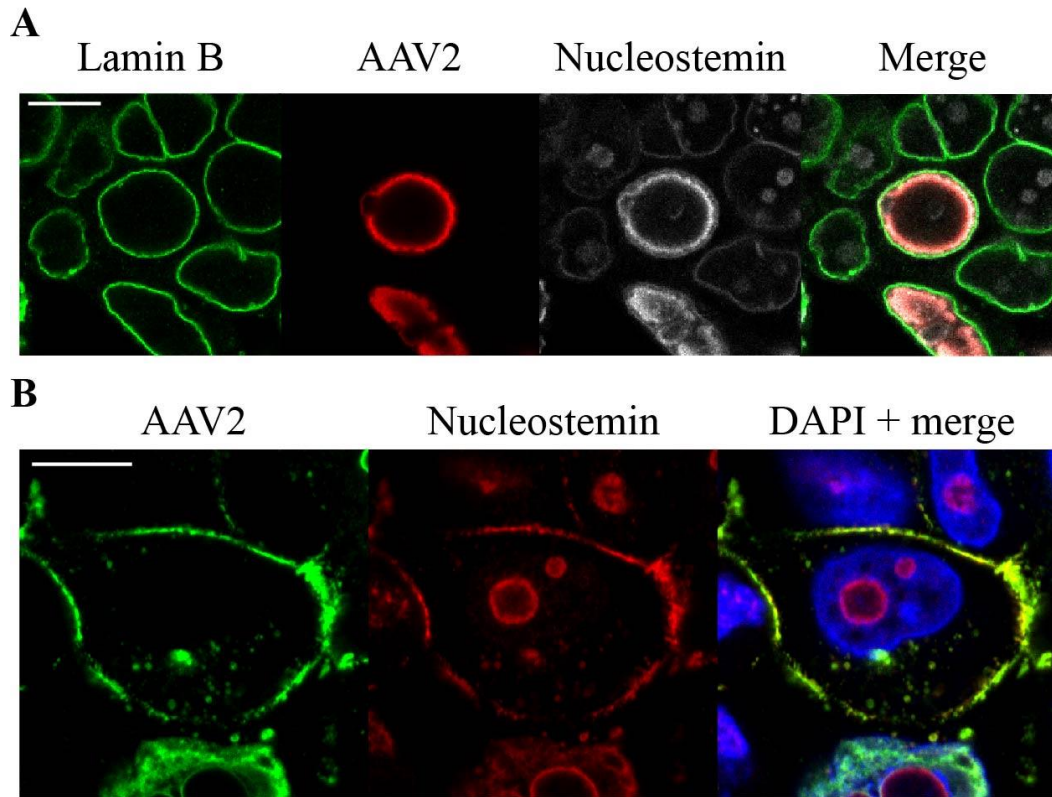
Figure ii: Intracellular localization of AAV2 and nucleostemin capsids with Rep (A) and during co-infection with Ad (B)
(**A**) HEK 293 cells were transfected with pHelper, pAAV-RepVP3, and pCMV$_3$-FLAGcmAAP2. Forty-nine hours post-transfection the cells were fixed and immunostained with anti-lamin B (green), anti-AAV2 (red), and anti-nucleostemin (white) antibodies. Lamin B is a marker for the nuclear membrane. (**B**) HeLa cells were co-infected with adenovirus (MOI 100) and wild-type AAV2 (MOI 1000). Forty-eight hours post infection the cells were fixed and immunostained with anti-AAV2 (green) and anti-nucleostemin (red) antibodies. Scale bar, 10

Figure iii Intracellular localization of AAV2 capsids in the presence of wild-type AAP2 or in the presence of the nuclear-targeted, nucleolar-excluded AAP2 mutant, AAP2mt26
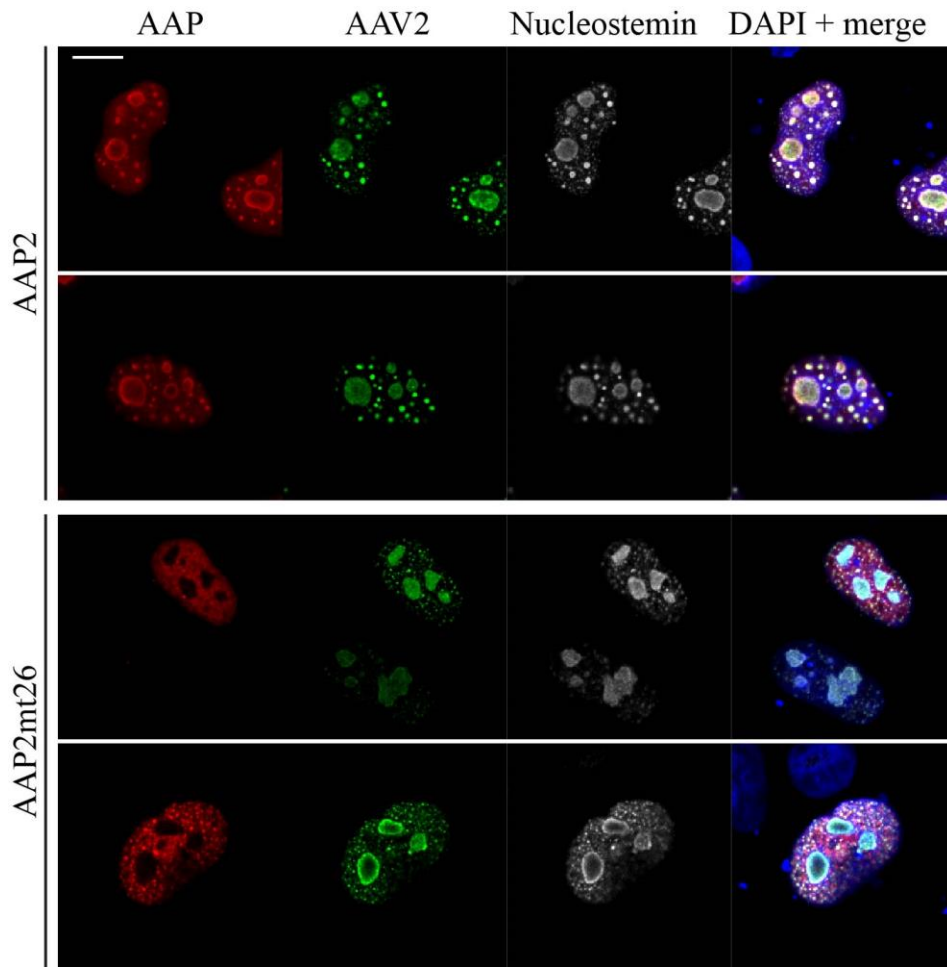
Figure iii Intracellular localization of AAV2 capsids in the presence of wild-type AAP2 or in the presence of the nuclear-targeted, nucleolar-excluded AAP2 mutant, AAP2mt26
HeLa cells were transfected with the following plasmids: pCMV$_1$-AAV2VP3 and pCMV$_3$-FLAG-cmAAP2 (wild type) (**A**), or pCMV$_1$-AAV2VP3 and pCMV$_3$-FLAG-cmAAP2mt26 (**B**). The cells were fixed at 48 hours post-transfection and immunostained with anti-FLAG, anti-AAV2, and anti-nucleostemin antibodies. Scale bar, 10 μm

Figure iv: Intracellular localization of AAV2, nucleolin, and AAP2
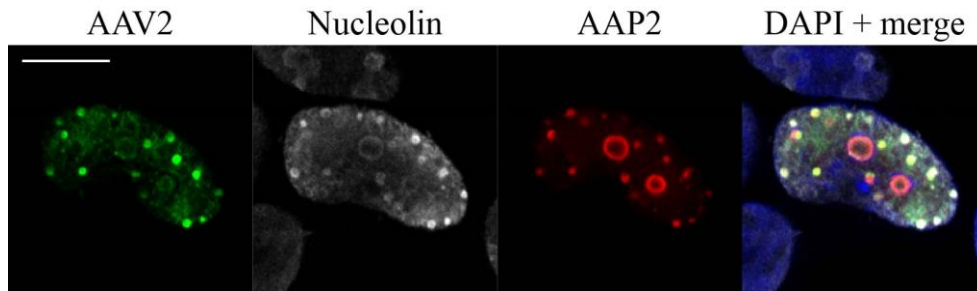
Figure iv: Intracellular localization of AAV2, nucleolin, and AAP2
HEK 293 cells were transfected with pCMV$_1$-AAV2VP3 and pCMV$_3$-FLAG-cmAAP2. Forty-eight hours post-transfection the cells were fixed stained and immunostained with anti-AAV, anti, nucleolin, and anti-FLAG antibodies. Scale bar, 10 μm

Figure v: Dot blot of encapsulated AAV genomes from particles made with AAV2 and AAV5 chimeric VP3s.
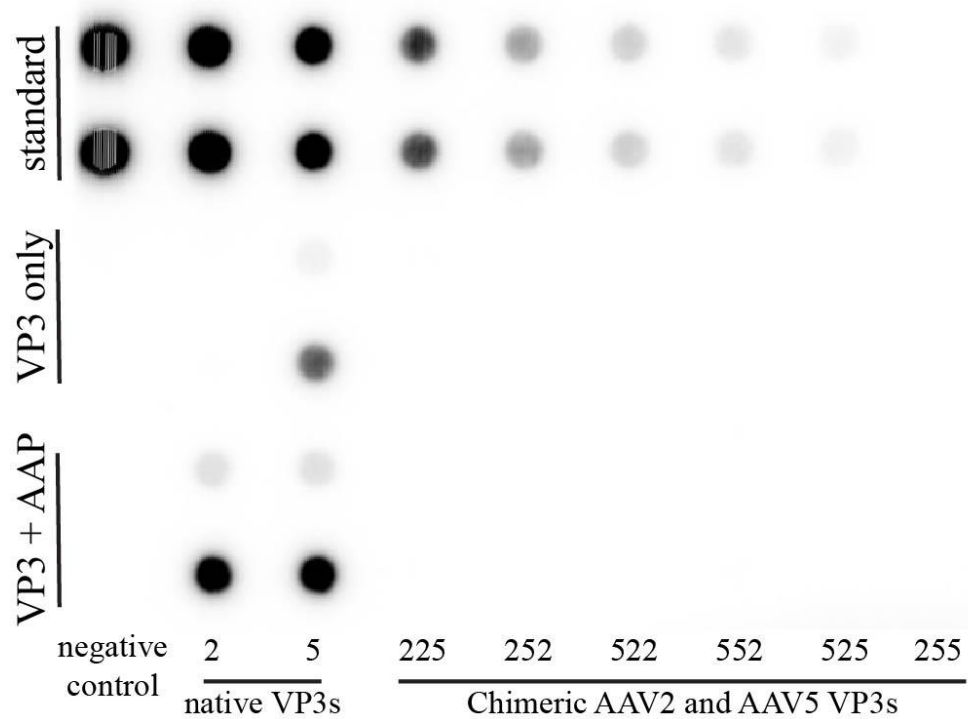
Figure v: Dot blot of encapsulated AAV genomes from particles made with AAV2 and AAV5 chimeric VP3s.
HEK 293 cells were transfected with pAAV-CMV-GFP, pHLP-Rep, pHelper, and a p4.1-CMV-VP3 plasmid in which the VP3 was the native sequence from AAV2, the native sequence of AAV5, or a chimeric VP3 made from swapping domains between the AAV2 and AAV5 VP3 reading frames. AAP function was supplied as indicated by inclusion of both p4.3c-FLAGcmAAP2 and p4.3c-FLAGcmAAP5 plasmids. The negative control lacked a VP3 plasmid.

Figure vi: The C-terminus of AAV11 contains a region that supports AAP-independent capsid assembly
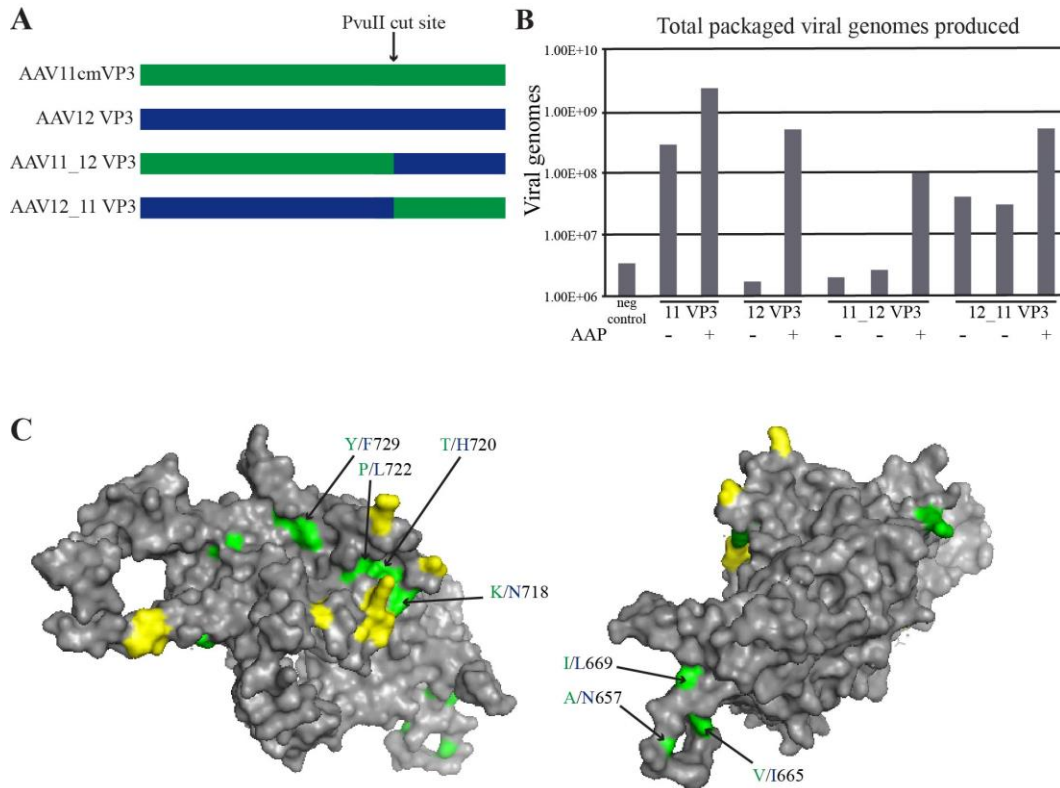
Figure vi: The C-terminus of AAV11 contains a region that supports AAP-independent capsid assembly.
(**A**) Schematic of AAV VP3 proteins: 11 and 12, and the chimeric VP3 proteins: 11_12 and 12_11. (**B**) Titer of encapsulated genomes from particles made with AAV VP3 11_12 and 12_11 VP3 chimeric proteins produced without or with both pCMV$_3$-FLAGcmAAP11 and pCMV$_3$-FLAGcmAAP12. (**C**) 3D model of AAV4 VP3 with amino acid differences between AAV11 VP3 and AAV12 VP3 C-terminal regions highlighted. Green denotes amino acids that are the same in AAV4 and AAV11, but different in AAV12. Yellow denotes amino acids different in all three serotypes. Arrows indicate amino acids that may be important to AAP-independent assembly based on their position within the structure of VP3. Green letters indicate the amino acid identity in AAV11 and blue letters indicated the amino acid identity in AAV12.

Figure vii: DAPI staining in HeLa cells highly over-expressing various AAPs
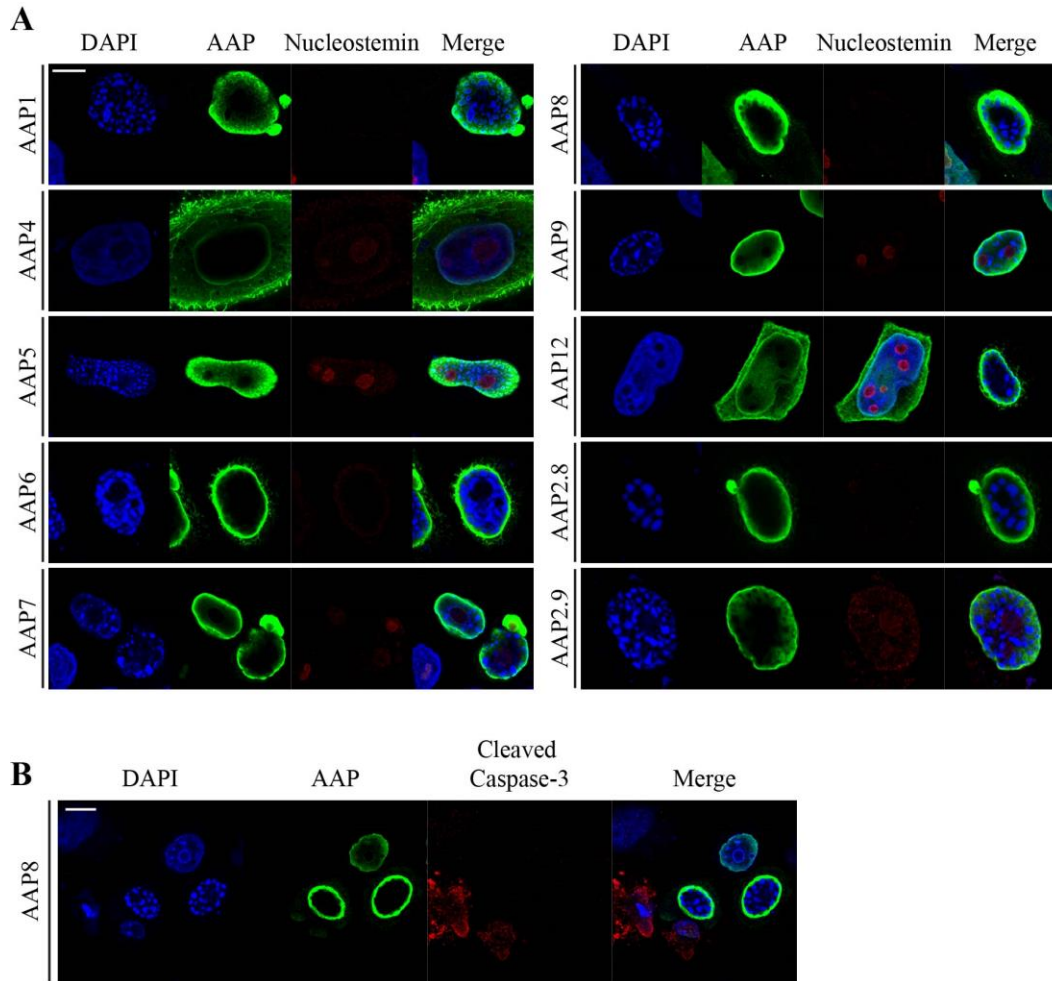
Figure vii: DAPI staining in HeLa cells highly over-expressing various AAPs.
(**A**) HeLa cells were transfected with p4.3c-FLAGcmAAP (AAP serotype is indicated). The cells were fixed 24-48 hours post-transfection and immunostained with anti-FLAG (green) and anti-nucleostemin (red) antibodies. DAPI (blue) was used to visualize the cellular DNA. (**B**) HeLa cells were transfected with p4.3c-FLAGcmAAP8. The cells were fixed 48 hours post-transfection and immunostained with anti-FLAG (green) and anti-cleaved caspase-3 (red) antibodies. DAPI (blue) was used to visualize the cellular DNA. Scale bar, 10 μm

Figure viii: AAP localization during mitosis.
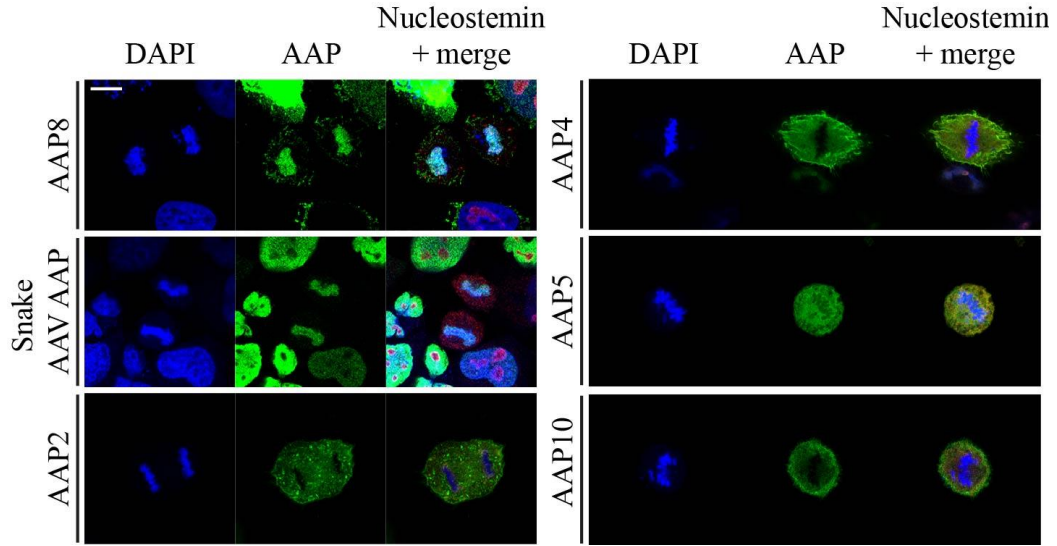
Figure viii: AAP localization during mitosis.
HeLa cells were transfected with p4.3c-FLAGcmAAP (AAP serotype is indicated). The cells were fixed 24-48 hours post-transfection and immunostained with anti-FLAG (green) and anti-nucleostemin (red) antibodies. DAPI (blue) was used to visualize the cellular DNA. Scale bar, 10 μm.

Materials and Methods

This section covers the materials and methods for the preceding Appendix.

Plasmid construction. Plasmids pCMV$_3$-FLAG-cmAAPx (x= AAV serotype, indicated in the figure legend) are plasmids expressing codon-modified (cm) versions of AAP with an N-terminal FLAG tag under the control of the human cytomegalovirus (CMV) immediate-early (IE) gene enhancer-promoter and an ATG start-codon. Each AAP ORF was codon-modified to optimize expression in human cells. Plasmids pCMV$_1$-AAVxVP3 (x= AAV serotype indicated in the figure legend) are plasmids expressing each VP3 protein from the native ORF initiating at the ATG start codon and ending at the TAA stop codon. pCMV$_3$-FLAG-cmAAP2.8 and pCMV$_3$-FLAG-cmAAP2.9 are plasmids expressing chimeric AAP proteins in which the amino acids 144-204 in AAP2 were replaced with the corresponding amino acids from either AAP8 or AAP9. pHelper, an Ad helper plasmid, was purchased from Agilent. pAAV-RepVP3 is a plasmid that expresses all the AAV2 Rep proteins in HEK 293 cells in the presence of co-transfected pHelper, but only the VP3 protein for the cap gene and not VP1, VP2, or AAP2. pCMV$_1$-AAV11_12-VP3 and pCMV$_1$-AAV12_11-VP3 were created by digesting pCMV$_1$-AAV11cmVP3 and pCMV$_1$-AAV12VP3 with *Pvu*II and *Nde*I, gel purifying the digestion products and ligating the chimeric plasmids together with T4 DNA ligase (M0202, New England Biolabs, Ipswich, MA). pCMV$_1$-AAV11_12-VP3 contained an ORF for amino acids 1- 340 of AAV11 VP3 fused to amino acids 341-538 of AAV12 VP3 and pCMV$_1$-AAV12_11- VP3 contained an ORF for amino acids 1- 340 of AAV12 VP3 fused to amino acids

341-538 of AAV11 VP3. pCMV$_3$-FLAGcmAAP2mt26 is the same plasmid used in Chapter 2 and is AAP2BR1+/BR2+/BR3-/BR4-/BR5-.

Cells. HEK 293 cells (AAV293) were purchased from Stratagene. The HeLa human cervical cancer cell line was obtained from the American Type Culture Collection (ATCC). HEK 293 cells and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Lonza, Basel, Switzerland) supplemented with 10% fetal bovine serum (FBS), L-glutamine, and penicillin-streptomycin.

Infections. HeLa cells were co-infected with wild-type AAV2 (MOI 100) and Ad (MOI 100) in serum free media. One hour post-infection the media was changed to DMEM (10% FBS) and cells were fixed 48 hours later.

Immunofluorescence microscopy. HeLa cells were seeded on coverslips in 12-well plates and transfected with plasmid DNA using PEI. Forty-eight hours after transfection, the cells were fixed with 4% paraformaldehyde at room temperature, permeabilized with 0.2% Tween 20, and blocked with 8% bovine serum albumin (BSA). The cells were stained with DAPI (4',6-diamidino-2- phenylindole) and the following antibodies as indicated: mouse monoclonal anti-FLAG M2 antibody (F1804; Sigma-Aldrich, St. Louis, MO), rabbit polyclonal anti-nucleostemin antibody (sc-67012; Santa Cruz Biotechnology, Dallas, TX), mouse monoclonal anti-AAV2 capsid antibody (A20 clone, 03-61055; American Research Products Inc., Waltham, MA), rat monoclonal anti-DYKDDDDK (FLAG) antibody

(NBP1-06712; Novus Biological, Littleton, CO), goat polyclonal anti-Lamin B antibody (sc-6217; Santa Cruz Biotechnology) and is a gift from Dr. Thayer, mouse monoclonal anti-nucleolin (C23) antibody (sc-8031; Santa Cruz Biotechnology) and is a gift from Dr. Dai, or rabbit polyclonal anti-Cleaved Caspase-3 (#9664, Cell Signaling Technology, Danvers, MA) and was a gift from Dr. Mandel. Secondary antibodies used were: Alexa Fluor 488-AffiniPure goat anti-mouse IgG antibody (115-545-166; Jackson ImmunoResearch), Cy3-AffiniPure donkey anti-rat IgG antibody (712-165-153; Jackson ImmunoResearch), Cy3-AffiniPure goat anti-rabbit IgG antibody (111-165-144 Jackson ImmunoResearch), Cy3-AffiniPure goat anti-mouse IgG antibody (115-165-166 Jackson ImmunoResearch), AF568 donkey anti-goat IgG and was a gift from Dr. Kurre, or Alexa Fluor 647-AffiniPure goat anti-rabbit IgG antibody (111-605-144; Jackson ImmunoResearch). The cells were imaged on a Zeiss LMS 710 laser scanning confocal microscope using either a 63x/1.4 NA or a 100x/1.46 NA objective.

Quantitative dot blot. HEK 293 cells were seeded on 6-well plates one day before transfection. We changed complete culture media to serum free media before transfection and transfected cells with 0.2 μg each of the following 3 plasmids, pAAV-CMV-GFP, pHLP-Rep and pHelper and the indicated plasmids: $pCMV_1$-AAV11cmVP3, $pCMV_1$-AAV12VP3, $pCMV_1$-AAV11_12-VP3, $pCMV_1$-AAV12_11-VP3, $pCMV_3$-FLAGcmAAP11, $pCMV_3$-FLAGcmAAP12, and $pCMV_1$ using PEI at a 1:2 weight ratio of DNA:PEI. Five days post-

transfection, we harvested both media and cells, disrupted cells by one cycle of freezing and thawing, and released viral particles into the culture media. After cell debris was removed by centrifugation at 21,100 × g for 5 min, 150 μl of the culture medium supernatant was subjected to nuclease treatment with 200 units per mL of Benzonase at 37°C for 4 h, followed by proteinase K treatment at 55°C for 1 h. Viral genome DNA was purified by phenol-chloroform extraction, ethanol-precipitated, and dissolved in 1 x Tris-HCl EDTA (TE) buffer (pH 8.0). The viral DNA and linearized standard plasmid DNA were then denatured with 0.4 N NaOH, blotted on a Zeta Probe nylon membrane (Bio-Rad, Hercules, CA), and hybridized with a ^{32}P-labeled GFP probe. The hybridized signals were imaged and quantified using a Typhoon FLA7000 scanner (GE Healthcare Bio-Science, Uppsala, Sweden). The negative control contained double-stranded AAV-CMV-GFP genomes that had undergone AAV2 Rep-mediated replication but were not protected by viral capsids. The negative control ensured efficient nuclease digestion in the assay.

References

1. **Atchison RW, Casto BC, Hammon WM.** 1965. Adeno-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. **Schmidt M, Grot E, Cervenka P, Wainer S, Buck C, Chiorini JA.** 2006. Identification and characterization of novel adeno-associated virus isolates in ATCC virus stocks. J Virol **80:**5082-5085.
4. **Schmidt M, Voutetakis A, Afione S, Zheng C, Mandikian D, Chiorini JA.** 2008. Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity. J Virol **82:**1399-1406.
5. **Rutledge EA, Halbert CL, Russell DW.** 1998. Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. J Virol **72:**309-319.
6. **Gao GP, Alvira MR, Wang L, Calcedo R, Johnston J, Wilson JM.** 2002. Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. Proc Natl Acad Sci U S A **99:**11854-11859.
7. **Mori S, Wang L, Takeuchi T, Kanda T.** 2004. Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein. Virology **330:**375-383.
8. **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.
9. **Gao G, Vandenberghe LH, Alvira MR, Lu Y, Calcedo R, Zhou X, Wilson JM.** 2004. Clades of Adeno-associated viruses are widely disseminated in human tissues. J Virol **78:**6381-6388.
10. **Vandenberghe LH, Breous E, Nam HJ, Gao G, Xiao R, Sandhu A, Johnston J, Debyser Z, Agbandje-McKenna M, Wilson JM.** 2009. Naturally occurring singleton residues in AAV capsid impact vector performance and illustrate structural constraints. Gene Ther **16:**1416-1428.
11. **Schmidt M, Katano H, Bossis I, Chiorini JA.** 2004. Cloning and characterization of a bovine adeno-associated virus. J Virol **78:**6509-6516.
12. **Arbetman AE, Lochrie M, Zhou S, Wellman J, Scallan C, Doroudchi MM, Randlev B, Patarroyo-White S, Liu T, Smith P, Lehmkuhl H, Hobbs LA, Pierce GF, Colosi P.** 2005. Novel caprine adeno-associated virus (AAV) capsid (AAV-Go.1) is closely related to the primate AAV-5 and has unique tropism and neutralization properties. J Virol **79:**15238-15245.
13. **Li Y, Ge X, Hon CC, Zhang H, Zhou P, Zhang Y, Wu Y, Wang LF, Shi Z.** 2010. Prevalence and genetic diversity of adeno-associated viruses in bats from China. J Gen Virol **91:**2601-2609.

14. **Farkas SL, Zadori Z, Benko M, Essbauer S, Harrach B, Tijssen P.** 2004. A parvovirus isolated from royal python (Python regius) is a member of the genus Dependovirus. J Gen Virol **85:**555-561.

15. **Penzes JJ, Pham HT, Benko M, Tijssen P.** 2015. Novel parvoviruses in reptiles and genome sequence of a lizard parvovirus shed light on Dependoparvovirus genus evolution. J Gen Virol **96:**2769-2779.

16. **Brown KE, Green SW, Young NS.** 1995. Goose parvovirus--an autonomous member of the dependovirus genus? Virology **210:**283-291.

17. **Zadori Z, Erdei J, Nagy J, Kisary J.** 1994. Characteristics of the genome of goose parvovirus. Avian Pathol **23:**359-364.

18. **Zadori Z, Stefancsik R, Rauch T, Kisary J.** 1995. Analysis of the complete nucleotide sequences of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-associated virus 2. Virology **212:**562-573.

19. **Lukashov VV, Goudsmit J.** 2001. Evolutionary relationships among parvoviruses: virus-host coevolution among autonomous primate parvoviruses and links between adeno-associated and avian parvoviruses. J Virol **75:**2729-2740.

20. **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.

21. **Hermonat PL, Muzyczka N.** 1984. Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. Proc Natl Acad Sci U S A **81:**6466-6470.

22. **Stutika C, Gogol-Doring A, Botschen L, Mietzsch M, Weger S, Feldkamp M, Chen W, Heilbronn R.** 2015. 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.

23. **Lusby E, Fife KH, Berns KI.** 1980. Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. J Virol **34:**402-409.

24. **Wang XS, Ponnazhagan S, Srivastava A.** 1995. Rescue and replication signals of the adeno-associated virus 2 genome. J Mol Biol **250:**573-580.

25. **Xiao X, Xiao W, Li J, Samulski RJ.** 1997. A novel 165-base-pair terminal repeat sequence is the sole cis requirement for the adeno-associated virus life cycle. J Virol **71:**941-948.

26. **Wang XS, Ponnazhagan S, Srivastava A.** 1996. 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 **70:**1668-1677.

27. **Feng D, Chen J, Yue Y, Zhu H, Xue J, Jia WW.** 2006. A 16bp Rep binding element is sufficient for mediating Rep-dependent integration into AAVS1. J Mol Biol **358:**38-45.

28. **Im DS, Muzyczka N.** 1990. The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell **61:**447-457.

29. **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.

30. **Labow MA, Hermonat PL, Berns KI.** 1986. Positive and negative autoregulation of the adeno-associated virus type 2 genome. J Virol **60:**251-258.

31. **Horer 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.

32. **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.

33. **Weger S, Wistuba A, Grimm D, Kleinschmidt JA.** 1997. Control of adeno-associated virus type 2 cap gene expression: relative influence of helper virus, terminal repeats, and Rep proteins. J Virol **71:**8437-8447.

34. **Qiu J, Pintel DJ.** 2002. The Adeno-Associated Virus Type 2 Rep Protein Regulates RNA Processing via Interaction with the Transcription Template. Molecular and Cellular Biology **22:**3639-3652.

35. **Pereira DJ, Muzyczka N.** 1997. The cellular transcription factor SP1 and an unknown cellular protein are required to mediate Rep protein activation of the adeno-associated virus p19 promoter. J Virol **71:**1747-1756.

36. **Berthet C, Raj K, Saudan P, Beard P.** 2005. How adeno-associated virus Rep78 protein arrests cells completely in S phase. Proc Natl Acad Sci U S A **102:**13634-13639.

37. **Nash K, Chen W, McDonald WF, Zhou X, Muzyczka N.** 2007. Purification of host cell enzymes involved in adeno-associated virus DNA replication. J Virol **81:**5777-5787.

38. **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.

39. **Ni TH, McDonald WF, Zolotukhin I, Melendy T, Waga S, Stillman B, Muzyczka N.** 1998. Cellular proteins required for adeno-associated virus DNA replication in the absence of adenovirus coinfection. J Virol **72:**2777-2787.

40. **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.

41. **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.

42. **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.

43. **Sonntag F, Kother K, Schmidt K, Weghofer M, Raupp C, Nieto K, Kuck A, Gerlach B, Bottcher B, Muller OJ, Lux K, Horer M, Kleinschmidt JA.** 2011. The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. J Virol **85:**12686-12697.

44. **Naumer M, Sonntag F, Schmidt K, Nieto K, Panke C, Davey NE, Popa-Wagner R, Kleinschmidt JA.** 2012. Properties of the adeno-associated virus assembly-activating protein. J Virol **86:**13038-13048.

45. **Ward P.** 2005. Replication of adeno-associated virus DNA, p 189-212. In Kerr J (ed), Parvoviruses. Hodder Arnold, Great Britain.

46. **Carter BJ, Khoury G, Rose JA.** 1972. Adenovirus-associated virus multiplication. IX. Extent of transcription of the viral genome in vivo. J Virol **10:**1118-1125.

47. **Koczot FJ, Carter BJ, Garon CF, Rose JA.** 1973. Self-complementarity of terminal sequences within plus or minus strands of adenovirus-associated virus DNA. Proc Natl Acad Sci U S A **70:**215-219.

48. **Tattersall P, Ward DC.** 1976. Rolling hairpin model for replication of parvovirus and linear chromosomal DNA. Nature **263:**106-109.

49. **Snyder RO, Im DS, Muzyczka N.** 1990. Evidence for covalent attachment of the adeno-associated virus (AAV) rep protein to the ends of the AAV genome. J Virol **64:**6204-6213.

50. **Myers MW, Carter BJ.** 1980. Assembly of adeno-associated virus. Virology **102:**71-82.

51. **Nick McElhinny SA, Gordenin DA, Stith CM, Burgers PM, Kunkel TA.** 2008. Division of labor at the eukaryotic replication fork. Mol Cell **30:**137-144.

52. **Nishida C, Reinhard P, Linn S.** 1988. DNA repair synthesis in human fibroblasts requires DNA polymerase delta. J Biol Chem **263:**501-510.

53. **Ogi T, Limsirichaikul S, Overmeer RM, Volker M, Takenaka K, Cloney R, Nakazawa Y, Niimi A, Miki Y, Jaspers NG, Mullenders LH, Yamashita S, Fousteri MI, Lehmann AR.** 2010. Three DNA polymerases, recruited by different mechanisms, carry out NER repair synthesis in human cells. Mol Cell **37:**714-727.

54. **Prindle MJ, Loeb LA.** 2012. DNA polymerase delta in DNA replication and genome maintenance. Environ Mol Mutagen **53:**666-682.

55. **Saudan P, Vlach J, Beard P.** 2000. Inhibition of S-phase progression by adeno-associated virus Rep78 protein is mediated by hypophosphorylated pRb. EMBO J **19:**4351-4361.

56. **Schlehofer JR.** 1994. The tumor suppressive properties of adeno-associated viruses. Mutat Res **305:**303-313.

57. **Hermonat PL.** 1994. Adeno-associated virus inhibits human papillomavirus type 16: a viral interaction implicated in cervical cancer. Cancer Res **54:**2278-2281.

58. **Cheung AK, Hoggan MD, Hauswirth WW, Berns KI.** 1980. Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. J Virol **33:**739-748.

59. **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.

60. **Samulski RJ, Zhu X, Xiao X, Brook JD, Housman DE, Epstein N, Hunter LA.** 1991. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. Embo j **10:**3941-3950.

61. **Ward P, Walsh CE.** 2012. Targeted integration of a rAAV vector into the AAVS1 region. Virology **433:**356-366.

62. **Weitzman MD, Kyostio SR, 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.

63. **Linden RM, Winocour E, Berns KI.** 1996. The recombination signals for adeno-associated virus site-specific integration. Proc Natl Acad Sci U S A **93:**7966-7972.

64. **Young SM, Jr., Samulski RJ.** 2001. Adeno-associated virus (AAV) site-specific recombination does not require a Rep-dependent origin of replication within the AAV terminal repeat. Proc Natl Acad Sci U S A **98:**13525-13530.

65. **Musayev FN, Zarate-Perez F, Bishop C, Burgner JW, 2nd, Escalante CR.** 2015. Structural Insights into the Assembly of the Adeno-associated Virus Type 2 Rep68 Protein on the Integration Site AAVS1. J Biol Chem **290:**27487-27499.

66. **Hickman AB, Ronning DR, Perez ZN, Kotin RM, 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.

67. **Giraud C, Winocour E, Berns KI.** 1995. Recombinant junctions formed by site-specific integration of adeno-associated virus into an episome. J Virol **69:**6917-6924.

68. **Hamilton H, Gomos J, Berns KI, Falck-Pedersen E.** 2004. Adeno-associated virus site-specific integration and AAVS1 disruption. J Virol **78:**7874-7882.

69. **Yang CC, Xiao X, Zhu X, Ansardi DC, Epstein ND, Frey MR, Matera AG, Samulski RJ.** 1997. Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro. J Virol **71:**9231-9247.

70. **McCarty DM, Young SM, Jr., Samulski RJ.** 2004. Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu Rev Genet **38:**819-845.

71. **Nakai H, Wu X, Fuess S, Storm TA, Munroe D, Montini E, Burgess SM, Grompe M, Kay MA.** 2005. Large-scale molecular characterization

of adeno-associated virus vector integration in mouse liver. J Virol **79:**3606-3614.

72. **Kaeppel C, Beattie SG, Fronza R, van Logtenstein R, Salmon F, Schmidt S, Wolf S, Nowrouzi A, Glimm H, von Kalle C, Petry H, Gaudet D, Schmidt M.** 2013. A largely random AAV integration profile after LPLD gene therapy. Nat Med **19:**889-891.

73. **Gil-Farina I, Fronza R, Kaeppel C, Lopez-Franco E, Ferreira V, D DA, Benito A, Prieto J, Petry H, Gonzalez-Aseguinolaza G, Schmidt M.** 2016. Recombinant AAV integration is not associated with hepatic genotoxicity in non-human primates and patients. Mol Ther doi:10.1038/mt.2016.52.

74. **Xiao X, Li J, Samulski RJ.** 1996. Efficient long-term gene transfer into muscle tissue of immunocompetent mice by adeno-associated virus vector. J Virol **70:**8098-8108.

75. **Schnepp BC, Clark KR, Klemanski DL, Pacak CA, Johnson PR.** 2003. Genetic fate of recombinant adeno-associated virus vector genomes in muscle. J Virol **77:**3495-3504.

76. **Nakai H, Yant SR, Storm TA, Fuess S, Meuse L, Kay MA.** 2001. Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. J Virol **75:**6969-6976.

77. **Sadelain M, Papapetrou EP, Bushman FD.** 2012. Safe harbours for the integration of new DNA in the human genome. Nat Rev Cancer **12:**51-58.

78. **Nault JC, Datta S, Imbeaud S, Franconi A, Mallet M, Couchy G, Letouze E, Pilati C, Verret B, Blanc JF, Balabaud C, Calderaro J, Laurent A, Letexier M, Bioulac-Sage P, Calvo F, Zucman-Rossi J.** 2015. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. Nat Genet **47:**1187-1193.

79. **Schnepp BC, Jensen RL, Chen CL, Johnson PR, Clark KR.** 2005. Characterization of adeno-associated virus genomes isolated from human tissues. J Virol **79:**14793-14803.

80. **Belyi VA, Levine AJ, Skalka AM.** 2010. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the parvoviridae and circoviridae are more than 40 to 50 million years old. J Virol **84:**12458-12462.

81. **Katzourakis A, Gifford RJ.** 2010. Endogenous viral elements in animal genomes. PLoS Genet **6:**e1001191.

82. **Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, Li G, Peng Y, Yi X, Jiang D.** 2011. Widespread endogenization of densoviruses and parvoviruses in animal and human genomes. J Virol **85:**9863-9876.

83. **Tratschin JD, West MH, Sandbank T, Carter BJ.** 1984. A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase. Mol Cell Biol **4:**2072-2081.

84. **Kyostio SR, Owens RA, Weitzman MD, Antoni BA, Chejanovsky N, Carter BJ.** 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.

85. **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.

86. **Laughlin CA, Westphal H, Carter BJ.** 1979. Spliced adenovirus-associated virus RNA. Proc Natl Acad Sci U S A **76:**5567-5571.

87. **Trempe JP, Carter BJ.** 1988. Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. J Virol **62:**3356-3363.

88. **Jay FT, Laughlin CA, Carter BJ.** 1981. Eukaryotic translational control: adeno-associated virus protein synthesis is affected by a mutation in the adenovirus DNA-binding protein. Proc Natl Acad Sci U S A **78:**2927-2931.

89. **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.

90. **McPherson RA, Rosenthal LJ, Rose JA.** 1985. Human cytomegalovirus completely helps adeno-associated virus replication. Virology **147:**217-222.

91. **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.

92. **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.

93. **Ogston P, Raj K, Beard P.** 2000. Productive replication of adeno-associated virus can occur in human papillomavirus type 16 (HPV-16) episome-containing keratinocytes and is augmented by the HPV-16 E2 protein. J Virol **74:**3494-3504.

94. **Schlehofer JR, Ehrbar M, zur Hausen H.** 1986. Vaccinia virus, herpes simplex virus, and carcinogens induce DNA amplification in a human cell line and support replication of a helpervirus dependent parvovirus. Virology **152:**110-117.

95. **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.

96. **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.

97. **Janik JE, Huston MM, Cho K, Rose JA.** 1989. Efficient synthesis of adeno-associated virus structural proteins requires both adenovirus DNA binding protein and VA I RNA. Virology **168:**320-329.

98. **Chang LS, Shenk T.** 1990. The adenovirus DNA-binding protein stimulates the rate of transcription directed by adenovirus and adeno-associated virus promoters. J Virol **64:**2103-2109.

99. **Carter BJ, Antoni BA, Klessig DF.** 1992. Adenovirus containing a deletion of the early region 2A gene allows growth of adeno-associated virus with decreased efficiency. Virology **191:**473-476.

100. **Ward P, Dean FB, O'Donnell ME, Berns KI.** 1998. Role of the adenovirus DNA-binding protein in in vitro adeno-associated virus DNA replication. J Virol **72:**420-427.

101. **Pilder S, Moore M, Logan J, Shenk T.** 1986. The adenovirus E1B-55K transforming polypeptide modulates transport or cytoplasmic stabilization of viral and host cell mRNAs. Mol Cell Biol **6:**470-476.

102. **Samulski RJ, Shenk T.** 1988. Adenovirus E1B 55-Mr polypeptide facilitates timely cytoplasmic accumulation of adeno-associated virus mRNAs. J Virol **62:**206-210.

103. **Querido E, Marcellus RC, Lai A, Charbonneau R, Teodoro JG, Ketner G, Branton PE.** 1997. Regulation of p53 levels by the E1B 55-kilodalton protein and E4orf6 in adenovirus-infected cells. J Virol **71:**3788-3798.

104. **Steegenga WT, Riteco N, Jochemsen AG, Fallaux FJ, Bos JL.** 1998. The large E1B protein together with the E4orf6 protein target p53 for active degradation in adenovirus infected cells. Oncogene **16:**349-357.

105. **Nayak R, Farris KD, Pintel DJ.** 2008. E4Orf6-E1B-55k-dependent degradation of de novo-generated adeno-associated virus type 5 Rep52 and capsid proteins employs a cullin 5-containing E3 ligase complex. J Virol **82:**3803-3808.

106. **Farris KD, Fasina O, Sukhu L, Li L, Pintel DJ.** 2010. Adeno-associated virus small rep proteins are modified with at least two types of polyubiquitination. J Virol **84:**1206-1211.

107. **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.

108. **Adachi K, Nakai H. 2012.** The role of DNA repair pathways in adeno-associated virus infection and viral genome replication/recombination/ integration. In: DNA Repair and Human Health. Vengrova S (Ed.). InTech, NY, USA.

109. **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.

110. **Kitajewski J, Schneider RJ, Safer B, Munemitsu SM, Samuel CE, Thimmappaya B, Shenk T.** 1986. Adenovirus VAI RNA antagonizes the antiviral action of interferon by preventing activation of the interferon-induced eIF-2 alpha kinase. Cell **45:**195-200.

111. **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.

112. **Snijder J, van de Waterbeemd M, Damoc E, Denisov E, Grinfeld D, Bennett A, Agbandje-McKenna M, Makarov A, Heck AJ.** 2014. Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. J Am Chem Soc **136:**7295-7299.

113. **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.

114. **Fox E, Moen PT, Jr., Bodnar JW.** 1990. Replication of minute virus of mice DNA in adenovirus-infected or adenovirus-transformed cells. Virology **176:**403-412.

115. **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.

116. **Yakobson B, Koch T, Winocour E.** 1987. Replication of adeno-associated virus in synchronized cells without the addition of a helper virus. J Virol **61:**972-981.

117. **Yalkinoglu AO, Heilbronn R, Burkle A, Schlehofer JR, zur Hausen H.** 1988. DNA amplification of adeno-associated virus as a response to cellular genotoxic stress. Cancer Res **48:**3123-3129.

118. **Crick FH, Watson JD.** 1956. Structure of small viruses. Nature **177:**473-475.

119. **Caspar DL, Klug A.** 1962. Physical principles in the construction of regular viruses. Cold Spring Harb Symp Quant Biol **27:**1-24.

120. **Liddington RC, Yan Y, Moulai J, Sahli R, Benjamin TL, Harrison SC.** 1991. Structure of simian virus 40 at 3.8-A resolution. Nature **354:**278-284.

121. **Mateu MG.** 2013. Assembly, stability and dynamics of virus capsids. Arch Biochem Biophys **531:**65-79.

122. **Perlmutter JD, Hagan MF.** 2015. Mechanisms of virus assembly. Annu Rev Phys Chem **66:**217-239.

123. **Dokland T.** 2000. Freedom and restraint: themes in virus capsid assembly. Structure **8:**R157-162.

124. **Knapman TW, Morton VL, Stonehouse NJ, Stockley PG, Ashcroft AE.** 2010. Determining the topology of virus assembly intermediates using ion mobility spectrometry-mass spectrometry. Rapid Commun Mass Spectrom **24:**3033-3042.

125. **Uetrecht C, Heck AJ.** 2011. Modern biomolecular mass spectrometry and its role in studying virus structure, dynamics, and assembly. Angew Chem Int Ed Engl **50:**8248-8262.

126. **Zlotnick A.** 2003. Are weak protein–protein interactions the general rule in capsid assembly? Virology **315:**269-274.

127. **Prasad BV, Schmid MF.** 2012. Principles of virus structural organization. Adv Exp Med Biol **726:**17-47.

128. **Almendral JM.** 2013. Assembly of simple icosahedral viruses. Subcell Biochem **68:**307-328.

129. **Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, Natarajan P, Johnson JE, Brooks CL, 3rd, Reddy VS.** 2009. VIPERdb2: an enhanced and web API enabled relational database for structural virology. Nucleic Acids Res **37:**D436-442.

130. **Kleinschmidt J, King J.** 2005. Molecular interactions involved in assembling the viral particle and packagin the genome, p 305-319. In Kerr J (ed), Parvoviruses. Hodder Arnold, Great Britain.

131. **Bashir T, Rommelaere J, Cziepluch C.** 2001. In vivo accumulation of cyclin A and cellular replication factors in autonomous parvovirus minute virus of mice-associated replication bodies. J Virol **75:**4394-4398.

132. **Oleksiewicz MB, Costello F, Huhtanen M, Wolfinbarger JB, Alexandersen S, Bloom ME.** 1996. Subcellular localization of Aleutian mink disease parvovirus proteins and DNA during permissive infection of Crandell feline kidney cells. J Virol **70:**3242-3247.

133. **Cziepluch C, Lampel S, Grewenig A, Grund C, Lichter P, Rommelaere J.** 2000. H-1 parvovirus-associated replication bodies: a distinct virus-induced nuclear structure. J Virol **74:**4807-4815.

134. **Bevington JM, Needham PG, Verrill KC, Collaco RF, Basrur V, Trempe JP.** 2007. Adeno-associated virus interactions with B23/Nucleophosmin: identification of sub-nucleolar virion regions. Virology **357:**102-113.

135. **Riolobos L, Reguera J, Mateu MG, Almendral JM.** 2006. Nuclear transport of trimeric assembly intermediates exerts a morphogenetic control on the icosahedral parvovirus capsid. J Mol Biol **357:**1026-1038.

136. **Lombardo E, Ramirez JC, Agbandje-McKenna M, Almendral JM.** 2000. A beta-stranded motif drives capsid protein oligomers of the parvovirus minute virus of mice into the nucleus for viral assembly. J Virol **74:**3804-3814.

137. **Lombardo E, Ramirez JC, Garcia J, Almendral JM.** 2002. Complementary roles of multiple nuclear targeting signals in the capsid proteins of the parvovirus minute virus of mice during assembly and onset of infection. J Virol **76:**7049-7059.

138. **Gil-Ranedo J, Hernando E, Riolobos L, Dominguez C, Kann M, Almendral JM.** 2015. The Mammalian Cell Cycle Regulates Parvovirus Nuclear Capsid Assembly. PLoS Pathog **11:**e1004920.

139. **Nicolson SC, Samulski RJ.** 2014. Recombinant adeno-associated virus utilizes host cell nuclear import machinery to enter the nucleus. J Virol **88:**4132-4144.

140. **Hoque M, Ishizu K, Matsumoto A, Han SI, Arisaka F, Takayama M, Suzuki K, Kato K, Kanda T, Watanabe H, Handa H.** 1999. Nuclear transport of the major capsid protein is essential for adeno-associated virus capsid formation. J Virol **73:**7912-7915.

141. **Popa-Wagner R, Sonntag F, Schmidt K, King J, Kleinschmidt JA.** 2012. Nuclear translocation of adeno-associated virus type 2 capsid proteins for virion assembly. J Gen Virol **93:**1887-1898.
142. **Grieger JC, Snowdy S, Samulski RJ.** 2006. Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly. J Virol **80:**5199-5210.
143. **Johnson JS, Li C, DiPrimio N, Weinberg MS, McCown TJ, Samulski RJ.** 2010. Mutagenesis of adeno-associated virus type 2 capsid protein VP1 uncovers new roles for basic amino acids in trafficking and cell-specific transduction. J Virol **84:**8888-8902.
144. **Popa-Wagner R, Porwal M, Kann M, Reuss M, Weimer M, Florin L, Kleinschmidt JA.** 2012. Impact of VP1-specific protein sequence motifs on adeno-associated virus type 2 intracellular trafficking and nuclear entry. J Virol **86:**9163-9174.
145. **Ruffing M, Heid H, Kleinschmidt JA.** 1994. Mutations in the carboxy terminus of adeno-associated virus 2 capsid proteins affect viral infectivity: lack of an RGD integrin-binding motif. J Gen Virol **75 (Pt 12):**3385-3392.
146. **Steinbach S, Wistuba A, Bock T, Kleinschmidt JA.** 1997. Assembly of adeno-associated virus type 2 capsids in vitro. J Gen Virol **78 (Pt 6):**1453-1462.
147. **Sanchez-Rodriguez SP, Munch-Anguiano L, Echeverria O, Vazquez-Nin G, Mora-Pale M, Dordick JS, Bustos-Jaimes I.** 2012. Human parvovirus B19 virus-like particles: In vitro assembly and stability. Biochimie **94:**870-878.
148. **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.
149. **Hunter LA, Samulski RJ.** 1992. Colocalization of adeno-associated virus Rep and capsid proteins in the nuclei of infected cells. J Virol **66:**317-324.
150. **Earley LF, Kawano Y, Adachi K, Sun XX, Dai MS, Nakai H.** 2015. Identification and characterization of nuclear and nucleolar localization signals in the adeno-associated virus serotype 2 assembly-activating protein. J Virol **89:**3038-3048.
151. **Johnson JS, Samulski RJ.** 2009. Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus. J Virol **83:**2632-2644.
152. **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.
153. **Wistuba A, Weger S, Kern A, Kleinschmidt JA.** 1995. Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins. J Virol **69:**5311-5319.
154. **Mohr D, Frey S, Fischer T, Guttler T, Gorlich D.** 2009. Characterisation of the passive permeability barrier of nuclear pore complexes. EMBO J **28:**2541-2553.

155. **Marshall E, Kerr J.** 2005. Pathogenesis of parvovirus infections, p 323-341. *In* Kerr J (ed), Parvoviruses. **Hodder Arnold**, Great Britain.
156. **Botquin V, Cid-Arregui A, Schlehofer JR.** 1994. Adeno-associated virus type 2 interferes with early development of mouse embryos. J Gen Virol **75 (Pt 10):**2655-2662.
157. **Servant-Delmas A, Lefrere JJ, Morinet F, Pillet S.** 2010. Advances in human B19 erythrovirus biology. J Virol **84:**9658-9665.
158. **Rosenfeld SJ, Yoshimoto K, Kajigaya S, Anderson S, Young NS, Field A, Warrener P, Bansal G, Collett MS.** 1992. Unique region of the minor capsid protein of human parvovirus B19 is exposed on the virion surface. J Clin Invest **89:**2023-2029.
159. **Agbandje M, Kajigaya S, McKenna R, Young NS, Rossmann MG.** 1994. The structure of human parvovirus B19 at 8 A resolution. Virology **203:**106-115.
160. **Yu MY, Alter HJ, Virata-Theimer ML, Geng Y, Ma L, Schechterly CA, Colvin CA, Luban NL.** 2010. Parvovirus B19 infection transmitted by transfusion of red blood cells confirmed by molecular analysis of linked donor and recipient samples. Transfusion **50:**1712-1721.
161. **Brown KE, Anderson SM, Young NS.** 1993. Erythrocyte P antigen: cellular receptor for B19 parvovirus. Science **262:**114-117.
162. **Takahashi T, Ozawa K, Takahashi K, Asano S, Takaku F.** 1990. Susceptibility of human erythropoietic cells to B19 parvovirus in vitro increases with differentiation. Blood **75:**603-610.
163. **Tolfvenstam T, Broliden K.** 2009. Parvovirus B19 infection. Semin Fetal Neonatal Med **14:**218-221.
164. **Fu Y, Ishii KK, Munakata Y, Saitoh T, Kaku M, Sasaki T.** 2002. Regulation of tumor necrosis factor alpha promoter by human parvovirus B19 NS1 through activation of AP-1 and AP-2. J Virol **76:**5395-5403.
165. **Moffatt S, Tanaka N, Tada K, Nose M, Nakamura M, Muraoka O, Hirano T, Sugamura K.** 1996. A cytotoxic nonstructural protein, NS1, of human parvovirus B19 induces activation of interleukin-6 gene expression. J Virol **70:**8485-8491.
166. **Kerr J, Modrow S.** 2005. Human and primate erythrovirus infections and associated disease, p 385-416. *In* Kerr J (ed), Parvovirues. Hooder Arnold, Great Britian.
167. **Kerr JR, Curran MD, Moore JE, Coyle PV, Ferguson WP.** 1995. Persistent parvovirus B19 infection. Lancet **345:**1118.
168. **Jakel V, Cussler K, Hanschmann KM, Truyen U, Konig M, Kamphuis E, Duchow K.** 2012. Vaccination against Feline Panleukopenia: implications from a field study in kittens. BMC Vet Res **8:**62.
169. **Stuetzer B, Hartmann K.** 2014. Feline parvovirus infection and associated diseases. Vet J **201:**150-155.
170. **Hoelzer K, Parrish CR.** 2010. The emergence of parvoviruses of carnivores. Vet Res **41:**39.

171. **Hueffer K, Parker JS, Weichert WS, Geisel RE, Sgro JY, Parrish CR.** 2003. The natural host range shift and subsequent evolution of canine parvovirus resulted from virus-specific binding to the canine transferrin receptor. J Virol **77:**1718-1726.

172. **Parrish C, Hueffer K.** 2005. Parvovirus host range, cell tropism and evolution – studies of canine and feline parvoviruses, minute virus of mice, porcine parvovirus, and Aleutian mink disease virus, p 343-350. *In* Kerr J (ed), Parvoviruses. Hodder Arnold Great Britain.

173. **Meunier PC, Cooper BJ, Appel MJ, Slauson DO.** 1984. Experimental viral myocarditis: parvoviral infection of neonatal pups. Vet Pathol **21:**509-515.

174. **Robinson WF, Huxtable CR, Pass DA.** 1980. Canine parvoviral myocarditis: a morphologic description of the natural disease. Vet Pathol **17:**282-293.

175. **Parrish CR, Oliver RE, McNiven R.** 1982. Canine parvovirus infections in a colony of dogs. Vet Microbiol **7:**317-324.

176. **Crawford LV.** 1966. A minute virus of mice. Virology **29:**605-612.

177. **Kimsey PB, Engers HD, Hirt B, Jongeneel CV.** 1986. Pathogenicity of fibroblast- and lymphocyte-specific variants of minute virus of mice. J Virol **59:**8-13.

178. **Brownstein DG, Smith AL, Jacoby RO, Johnson EA, Hansen G, Tattersall P.** 1991. Pathogenesis of infection with a virulent allotropic variant of minute virus of mice and regulation by host genotype. Lab Invest **65:**357-364.

179. **Brownstein DG, Smith AL, Johnson EA, Pintel DJ, Naeger LK, Tattersall P.** 1992. The pathogenesis of infection with minute virus of mice depends on expression of the small nonstructural protein NS2 and on the genotype of the allotropic determinants VP1 and VP2. J Virol **66:**3118-3124.

180. **Smith AL, Jacoby RO, Johnson EA, Paturzo F, Bhatt PN.** 1993. In vivo studies with an "orphan" parvovirus of mice. Lab Anim Sci **43:**175-182.

181. **McMaster GK, Beard P, Engers HD, Hirt B.** 1981. Characterization of an immunosuppressive parvovirus related to the minute virus of mice. J Virol **38:**317-326.

182. **Erles K, Sebokova P, Schlehofer JR.** 1999. Update on the prevalence of serum antibodies (IgG and IgM) to adeno-associated virus (AAV). J Med Virol **59:**406-411.

183. **Calcedo R, Vandenberghe LH, Gao G, Lin J, Wilson JM.** 2009. Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. J Infect Dis **199:**381-390.

184. **Han L, Parmley TH, Keith S, Kozlowski KJ, Smith LJ, Hermonat PL.** 1996. High prevalence of adeno-associated virus (AAV) type 2 rep DNA in cervical materials: AAV may be sexually transmitted. Virus Genes **12:**47-52.

185. **Flotte TR, Berns KI.** 2005. Adeno-associated virus: a ubiquitous commensal of mammals. Hum Gene Ther **16:**401-407.

186. **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.

187. **Mulder J, Ariaens A, van den Boomen D, Moolenaar WH.** 2004. p116Rip targets myosin phosphatase to the actin cytoskeleton and is essential for RhoA/ROCK-regulated neuritogenesis. Mol Biol Cell **15:**5516-5527.

188. **Henckaerts E, Dutheil N, Zeltner N, Kattman S, Kohlbrenner E, Ward P, Clement N, Rebollo P, Kennedy M, Keller GM, Linden RM.** 2009. Site-specific integration of adeno-associated virus involves partial duplication of the target locus. Proc Natl Acad Sci U S A **106:**7571-7576.

189. **Mayor HD, Houlditch GS, Mumford DM.** 1973. Influence of adeno-associated satellite virus on adenovirus-induced tumours in hamsters. Nat New Biol **241:**44-46.

190. **Cukor G, Blacklow NR, Kibrick S, Swan IC.** 1975. Effect of adeno-associated virus on cancer expression by herpesvirus-transformed hamster cells. J Natl Cancer Inst **55:**957-959.

191. **Ostrove JM, Duckworth DH, Berns KI.** 1981. Inhibition of adenovirus-transformed cell oncogenicity by adeno-associated virus. Virology **113:**521-533.

192. **Hermonat PL.** 1991. Inhibition of H-ras expression by the adeno-associated virus Rep78 transformation suppressor gene product. Cancer Res **51:**3373-3377.

193. **Hermonat PL.** 1992. Inhibition of bovine papillomavirus plasmid DNA replication by adeno-associated virus. Virology **189:**329-333.

194. **Freitas LB, Tonani de Mattos A, Lima BM, Miranda AE, Spano LC.** 2012. Adeno-associated virus may play a protective role against human papillomavirus-induced cervical lesions independent of HIV serostatus. Int J STD AIDS **23:**258-261.

195. **Smith J, Herrero R, Erles K, Grimm D, Munoz N, Bosch FX, Tafur L, Shah KV, Schlehofer JR.** 2001. Adeno-associated virus seropositivity and HPV-induced cervical cancer in Spain and Colombia. Int J Cancer **94:**520-526.

196. **Coker AL, Russell RB, Bond SM, Pirisi L, Liu Y, Mane M, Kokorina N, Gerasimova T, Hermonat PL.** 2001. Adeno-associated virus is associated with a lower risk of high-grade cervical neoplasia. Exp Mol Pathol **70:**83-89.

197. **Batchu RB, Shammas MA, Wang JY, Shmookler Reis RJ, Munshi NC.** 2002. Interaction of adeno-associated virus Rep78 with SV40 T antigen: implications in Rep protein expression leading to the inhibition of SV40-mediated cell proliferation. Intervirology **45:**115-118.

198. **Liu T, Cong M, Wang P, Jia J, Liu Y, Hermonat PL, You H.** 2009. Adeno-associated virus Rep78 protein inhibits Hepatitis B virus

replication through regulation of the HBV core promoter. Biochem Biophys Res Commun **385:**106-111.

199. **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.

200. **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.

201. **Rohde V, Erles K, Sattler HP, Derouet H, Wullich B, Schlehofer JR.** 1999. Detection of adeno-associated virus in human semen: does viral infection play a role in the pathogenesis of male infertility? Fertil Steril **72:**814-816.

202. **Tobiasch E, Rabreau M, Geletneky K, Larue-Charlus S, Severin F, Becker N, Schlehofer JR.** 1994. Detection of adeno-associated virus DNA in human genital tissue and in material from spontaneous abortion. J Med Virol **44:**215-222.

203. **Schlehofer JR, Boeke C, Reuland M, Eggert-Kruse W.** 2012. Presence of DNA of adeno-associated virus in subfertile couples, but no association with fertility factors. Hum Reprod **27:**770-778.

204. **Donsante A, Vogler C, Muzyczka N, Crawford JM, Barker J, Flotte T, Campbell-Thompson M, Daly T, Sands MS.** 2001. Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. Gene Ther **8:**1343-1346.

205. **Donsante A, Miller DG, Li Y, Vogler C, Brunt EM, Russell DW, Sands MS.** 2007. AAV vector integration sites in mouse hepatocellular carcinoma. Science **317:**477.

206. **Russell DW, Grompe M.** 2015. Adeno-associated virus finds its disease. Nat Genet **47:**1104-1105.

207. **Berns KI, Byrne BJ, Flotte TR, Gao G, Hauswirth WW, Herzog RW, Muzyczka N, VandenDriessche T, Xiao X, Zolotukhin S, Srivastava A.** 2015. Adeno-Associated Virus Type 2 and Hepatocellular Carcinoma? Hum Gene Ther **26:**779-781.

208. **Buning H, Schmidt M.** 2015. Adeno-associated Vector Toxicity-To Be or Not to Be? Mol Ther **23:**1673-1675.

209. **Nault JC, Datta S, Imbeaud S, Franconi A, Mallet M, Couchy G, Letouze E, Pilati C, Verret B, Blanc JF, Balabaud C, Calderaro J, Laurent A, Letexier M, Bioulac-Sage P, Calvo F, Zucman-Rossi J.** 2016. AAV2 and hepatocellular carcinoma. Hum Gene Ther doi:10.1089/hum.2016.002.

210. **Anderson WF.** 1984. Prospects for human gene therapy. Science **226:**401-409.

211. **Carter BJ.** 2004. Adeno-associated virus and the development of adeno-associated virus vectors: a historical perspective. Mol Ther **10:**981-989.

212. **Carter BJ.** 1992. Adeno-associated virus vectors. Curr Opin Biotechnol **3:**533-539.

213. **Laughlin CA, Tratschin JD, Coon H, Carter BJ.** 1983. Cloning of infectious adeno-associated virus genomes in bacterial plasmids. Gene **23:**65-73.

214. **Kotin RM.** 1994. Prospects for the use of adeno-associated virus as a vector for human gene therapy. Hum Gene Ther **5:**793-801.

215. **McLaughlin SK, Collis P, Hermonat PL, Muzyczka N.** 1988. Adeno-associated virus general transduction vectors: analysis of proviral structures. J Virol **62:**1963-1973.

216. **Flotte T, Carter B, Conrad C, Guggino W, Reynolds T, Rosenstein B, Taylor G, Walden S, Wetzel R.** 1996. A phase I study of an adeno-associated virus-CFTR gene vector in adult CF patients with mild lung disease. Hum Gene Ther **7:**1145-1159.

217. **Kearns WG, Afione SA, Fulmer SB, Pang MC, Erikson D, Egan M, Landrum MJ, Flotte TR, Cutting GR.** 1996. Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. Gene Ther **3:**748-755.

218. **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.

219. **Duan D, Sharma P, Yang J, Yue Y, Dudus L, Zhang Y, Fisher KJ, Engelhardt JF.** 1998. Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. J Virol **72:**8568-8577.

220. **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.

221. **McCarty DM, Monahan PE, Samulski RJ.** 2001. Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. Gene Ther **8:**1248-1254.

222. **Matsushita T, Elliger S, Elliger C, Podsakoff G, Villarreal L, Kurtzman GJ, Iwaki Y, Colosi P.** 1998. Adeno-associated virus vectors can be efficiently produced without helper virus. Gene Ther **5:**938-945.

223. **Kay MA, Manno CS, Ragni MV, Larson PJ, Couto LB, McClelland A, Glader B, Chew AJ, Tai SJ, Herzog RW, Arruda V, Johnson F, Scallan C, Skarsgard E, Flake AW, High KA.** 2000. Evidence for gene transfer and expression of factor IX in haemophilia B patients treated with an AAV vector. Nat Genet **24:**257-261.

224. **Manno CS, Pierce GF, Arruda VR, Glader B, Ragni M, Rasko JJ, Ozelo MC, Hoots K, Blatt P, Konkle B, Dake M, Kaye R, Razavi M, Zajko A, Zehnder J, Rustagi PK, Nakai H, Chew A, Leonard D, Wright JF, Lessard RR, Sommer JM, Tigges M, Sabatino D, Luk A, Jiang H, Mingozzi F, Couto L, Ertl HC, High KA, Kay MA.** 2006. Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat Med **12:**342-347.

225. **Mingozzi F, Maus MV, Hui DJ, Sabatino DE, Murphy SL, Rasko JE, Ragni MV, Manno CS, Sommer J, Jiang H, Pierce GF, Ertl HC, High KA.** 2007. CD8(+) T-cell responses to adeno-associated virus capsid in humans. Nat Med **13:**419-422.

226. **Balakrishnan B, Jayandharan GR.** 2014. Basic biology of adeno-associated virus (AAV) vectors used in gene therapy. Curr Gene Ther **14:**86-100.

227. **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.

228. **Jacobson SG, Cideciyan AV, Ratnakaram R, Heon E, Schwartz SB, Roman AJ, Peden MC, Aleman TS, Boye SL, Sumaroka A, Conlon TJ, Calcedo R, Pang JJ, Erger KE, Olivares MB, Mullins CL, Swider M, Kaushal S, Feuer WJ, Iannaccone A, Fishman GA, Stone EM, Byrne BJ, Hauswirth WW.** 2012. Gene therapy for leber congenital amaurosis caused by RPE65 mutations: safety and efficacy in 15 children and adults followed up to 3 years. Arch Ophthalmol **130:**9-24.

229. **Pierce EA, Bennett J.** 2015. The Status of RPE65 Gene Therapy Trials: Safety and Efficacy. Cold Spring Harb Perspect Med **5:**a017285.

230. **Simonelli F, Maguire AM, Testa F, Pierce EA, Mingozzi F, Bennicelli JL, Rossi S, Marshall K, Banfi S, Surace EM, Sun J, Redmond TM, Zhu X, Shindler KS, Ying GS, Ziviello C, Acerra C, Wright JF, McDonnell JW, High KA, Bennett J, Auricchio A.** 2010. Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. Mol Ther **18:**643-650.

231. **Bowles DE, McPhee SW, Li C, Gray SJ, Samulski JJ, Camp AS, Li J, Wang B, Monahan PE, Rabinowitz JE, Grieger JC, Govindasamy L, Agbandje-McKenna M, Xiao X, Samulski RJ.** 2012. Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector. Mol Ther **20:**443-455.

232. **Mendell JR, Sahenk Z, Malik V, Gomez AM, Flanigan KM, Lowes LP, Alfano LN, Berry K, Meadows E, Lewis S, Braun L, Shontz K, Rouhana M, Clark KR, Rosales XQ, Al-Zaidy S, Govoni A, Rodino-Klapac LR, Hogan MJ, Kaspar BK.** 2015. A phase 1/2a follistatin gene therapy trial for becker muscular dystrophy. Mol Ther **23:**192-201.

233. **Carpentier AC, Frisch F, Labbe SM, Gagnon R, de Wal J, Greentree S, Petry H, Twisk J, Brisson D, Gaudet D.** 2012. Effect of alipogene tiparvovec (AAV1-LPL(S447X)) on postprandial chylomicron metabolism in lipoprotein lipase-deficient patients. J Clin Endocrinol Metab **97:**1635-1644.

234. **Mueller C, Flotte TR.** 2008. Clinical gene therapy using recombinant adeno-associated virus vectors. Gene Ther **15:**858-863.
235. **Mingozzi F, High KA.** 2011. Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet **12:**341-355.
236. **Lisowski L, Tay SS, Alexander IE.** 2015. Adeno-associated virus serotypes for gene therapeutics. Curr Opin Pharmacol **24:**59-67.
237. **Wu Z, Asokan A, Samulski RJ.** 2006. Adeno-associated virus serotypes: vector toolkit for human gene therapy. Mol Ther **14:**316-327.
238. **Asokan A, Schaffer DV, Samulski RJ.** 2012. The AAV vector toolkit: poised at the clinical crossroads. Mol Ther **20:**699-708.
239. **Stemmer WP.** 1994. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proc Natl Acad Sci U S A **91:**10747-10751.
    See also: https://www.youtube.com/watch?v=KQ6zr6kCPj8
240. **Grimm D, Lee JS, Wang L, Desai T, Akache B, Storm TA, Kay MA.** 2008. In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J Virol **82:**5887-5911.
241. **Hauck B, Chen L, Xiao W.** 2003. Generation and characterization of chimeric recombinant AAV vectors. Mol Ther **7:**419-425.
242. **Bowles DE, Rabinowitz JE, Samulski RJ.** 2003. Marker rescue of adeno-associated virus (AAV) capsid mutants: a novel approach for chimeric AAV production. J Virol **77:**423-432.
243. **Hauck B, Xiao W.** 2003. Characterization of tissue tropism determinants of adeno-associated virus type 1. J Virol **77:**2768-2774.
244. **Shen X, Storm T, Kay MA.** 2007. Characterization of the relationship of AAV capsid domain swapping to liver transduction efficiency. Mol Ther **15:**1955-1962.
245. **Grimm D, Zolotukhin S.** 2015. E Pluribus Unum: 50 Years of Research, Millions of Viruses, and One Goal-Tailored Acceleration of AAV Evolution. Mol Ther **23:**1819-1831.
246. **Li W, Asokan A, Wu Z, Van Dyke T, DiPrimio N, Johnson JS, Govindaswamy L, Agbandje-McKenna M, Leichtle S, Redmond DE, Jr., McCown TJ, Petermann KB, Sharpless NE, Samulski RJ.** 2008. Engineering and selection of shuffled AAV genomes: a new strategy for producing targeted biological nanoparticles. Mol Ther **16:**1252-1260.
247. **Dalkara D, Byrne LC, Klimczak RR, Visel M, Yin L, Merigan WH, Flannery JG, Schaffer DV.** 2013. In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. Sci Transl Med **5:**189ra176.
248. **Rabinowitz JE, Xiao W, Samulski RJ.** 1999. Insertional mutagenesis of AAV2 capsid and the production of recombinant virus. Virology **265:**274-285.
249. **Wu P, Xiao W, Conlon T, Hughes J, Agbandje-McKenna M, Ferkol T, Flotte T, Muzyczka N.** 2000. Mutational analysis of the adeno-

associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. J Virol **74:**8635-8647.

250. **Shi W, Arnold GS, Bartlett JS.** 2001. Insertional mutagenesis of the adeno-associated virus type 2 (AAV2) capsid gene and generation of AAV2 vectors targeted to alternative cell-surface receptors. Hum Gene Ther **12:**1697-1711.

251. **Muzyczka N, Warrington KH, Jr.** 2005. Custom adeno-associated virus capsids: the next generation of recombinant vectors with novel tropism. Hum Gene Ther **16:**408-416.

252. **Buning H, Bolyard CM, Hallek M, Bartlett JS.** 2011. Modification and labeling of AAV vector particles. Methods Mol Biol **807:**273-300.

253. **Nicklin SA, Buening H, Dishart KL, de Alwis M, Girod A, Hacker U, Thrasher AJ, Ali RR, Hallek M, Baker AH.** 2001. Efficient and selective AAV2-mediated gene transfer directed to human vascular endothelial cells. Mol Ther **4:**174-181.

254. **Chen YH, Chang M, Davidson BL.** 2009. Molecular signatures of disease brain endothelia provide new sites for CNS-directed enzyme therapy. Nat Med **15:**1215-1218.

255. **Choudhury SR, Harris AF, Cabral DJ, Keeler AM, Sapp E, Ferreira JS, Gray-Edwards HL, Johnson JA, Johnson AK, Su Q, Stoica L, DiFiglia M, Aronin N, Martin DR, Gao G, Sena-Esteves M.** 2015. Widespread Central Nervous System Gene Transfer and Silencing After Systemic Delivery of Novel AAV-AS Vector. Mol Ther **24**(4):726-35.

256. **Warrington KH, Jr., Gorbatyuk OS, Harrison JK, Opie SR, Zolotukhin S, Muzyczka N.** 2004. Adeno-associated virus type 2 VP2 capsid protein is nonessential and can tolerate large peptide insertions at its N terminus. J Virol **78:**6595-6609.

257. **Lux K, Goerlitz N, Schlemminger S, Perabo L, Goldnau D, Endell J, Leike K, Kofler DM, Finke S, Hallek M, Buning H.** 2005. Green fluorescent protein-tagged adeno-associated virus particles allow the study of cytosolic and nuclear trafficking. J Virol **79:**11776-11787.

258. **Huttner NA, Girod A, Perabo L, Edbauer D, Kleinschmidt JA, Buning H, Hallek M.** 2003. Genetic modifications of the adeno-associated virus type 2 capsid reduce the affinity and the neutralizing effects of human serum antibodies. Gene Ther **10:**2139-2147.

259. **Zinn E, Pacouret S, Khaychuk V, Turunen HT, Carvalho LS, Andres-Mateos E, Shah S, Shelke R, Maurer AC, Plovie E, Xiao R, Vandenberghe LH.** 2015. In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. Cell Rep **12:**1056-1068.

260. **Santiago-Ortiz J, Ojala DS, Westesson O, Weinstein JR, Wong SY, Steinsapir A, Kumar S, Holmes I, Schaffer DV.** 2015. AAV ancestral reconstruction library enables selection of broadly infectious viral variants. Gene Ther **22:**934-946.

261. **Adachi K, Enoki T, Kawano Y, Veraz M, Nakai H.** 2014. Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing. Nat Commun **5:**3075.

262. **Marsic D, Govindasamy L, Currlin S, Markusic DM, Tseng YS, Herzog RW, Agbandje-McKenna M, Zolotukhin S.** 2014. Vector design Tour de Force: integrating combinatorial and rational approaches to derive novel adeno-associated virus variants. Mol Ther **22:**1900-1909.

263. **Hernandez-Verdun D.** 2011. Structural Organization of the Nucleolus as a Consequence of the Dynamics of Ribosome Biogenesis, p 3-28. *In* Olson MOJ (ed), The Nucleolus. Springer, New York, NY.

264. **McClintock B.** 1934. The relation of a particular chromosomal element to the development of the nucleoli in Zea mays. Zeitschrift für Zellforschung und Mikroskopische Anatomie **21:**294-328.

265. **Henderson AS, Warburton D, Atwood KC.** 1972. Location of ribosomal DNA in the human chromosome complement. Proc Natl Acad Sci U S A **69:**3394-3398.

266. **Boisvert FM, van Koningsbruggen S, Navascues J, Lamond AI.** 2007. The multifunctional nucleolus. Nat Rev Mol Cell Biol **8:**574-585.

267. **Lam YW, Trinkle-Mulcahy L.** 2015. New insights into nucleolar structure and function. F1000Prime Rep **7:**48.

268. **Henras AK, Plisson-Chastang C, O'Donohue MF, Chakraborty A, Gleizes PE.** 2015. An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip Rev RNA **6:**225-242.

269. **Roussel P, Andre C, Comai L, Hernandez-Verdun D.** 1996. The rDNA transcription machinery is assembled during mitosis in active NORs and absent in inactive NORs. J Cell Biol **133:**235-246.

270. **Dundr M, Meier UT, Lewis N, Rekosh D, Hammarskjold ML, Olson MO.** 1997. A class of nonribosomal nucleolar components is located in chromosome periphery and in nucleolus-derived foci during anaphase and telophase. Chromosoma **105:**407-417.

271. **Olson MO, Dundr M, Szebeni A.** 2000. The nucleolus: an old factory with unexpected capabilities. Trends Cell Biol **10:**189-196.

272. **Tollini L, RA F, Y Z.** 2011. The Role of the Nucleolus in the Stress Response, p 281-299. *In* Olson MO (ed), The nucleolus. Springer, New York.

273. **James A, Wang Y, Raje H, Rosby R, DiMario P.** 2014. Nucleolar stress with and without p53. Nucleus **5:**402-426.

274. **Salvetti A, Greco A.** 2014. Viruses and the nucleolus: the fatal attraction. Biochim Biophys Acta **1842:**840-847.

275. **Rawlinson SM, Moseley GW.** 2015. The nucleolar interface of RNA viruses. Cell Microbiol **17:**1108-1120.

276. **Yang MR, Lee SR, Oh W, Lee EW, Yeh JY, Nah JJ, Joo YS, Shin J, Lee HW, Pyo S, Song J.** 2008. West Nile virus capsid protein induces p53-mediated apoptosis via the sequestration of HDM2 to the nucleolus. Cell Microbiol **10:**165-176.

277. **Hu J, Cai XF, Yan G.** 2009. Alphavirus M1 induces apoptosis of malignant glioma cells via downregulation and nucleolar translocation of p21WAF1/CIP1 protein. Cell Cycle **8:**3328-3339.

278. **Wang X, Shen Y, Qiu Y, Shi Z, Shao D, Chen P, Tong G, Ma Z.** 2010. The non-structural (NS1) protein of influenza A virus associates with p53 and inhibits p53-mediated transcriptional activity and apoptosis. Biochem Biophys Res Commun **395:**141-145.

279. **Aminev AG, Amineva SP, Palmenberg AC.** 2003. Encephalomyocarditis virus (EMCV) proteins 2A and 3BCD localize to nuclei and inhibit cellular mRNA transcription but not rRNA transcription. Virus Res **95:**59-73.

280. **Weidman MK, Sharma R, Raychaudhuri S, Kundu P, Tsai W, Dasgupta A.** 2003. The interaction of cytoplasmic RNA viruses with the nucleus. Virus Res **95:**75-85.

281. **Boyne JR, Whitehouse A.** 2006. Nucleolar trafficking is essential for nuclear export of intronless herpesvirus mRNA. Proc Natl Acad Sci U S A **103:**15190-15195.

282. **Matthews D EE, Hiscox J** 2011. Viruses and the nucleolus, p 321-345. *In* Olson MO (ed), The nucleolus. Springer, New York.

283. **Sagou K, Uema M, Kawaguchi Y.** 2010. Nucleolin is required for efficient nuclear egress of herpes simplex virus type 1 nucleocapsids. J Virol **84:**2110-2121.

284. **Qiu J, Brown KE.** 1999. A 110-kDa nuclear shuttle protein, nucleolin, specifically binds to adeno-associated virus type 2 (AAV-2) capsid. Virology **257:**373-382.

285. **Dong B, Duan X, Chow HY, Chen L, Lu H, Wu W, Hauck B, Wright F, Kapranov P, Xiao W.** 2014. Proteomics analysis of co-purifying cellular proteins associated with rAAV vectors. PLoS One **9:**e86453.

286. **Nash K, Chen W, Salganik M, Muzyczka N.** 2009. Identification of cellular proteins that interact with the adeno-associated virus rep protein. J Virol **83:**454-469.

287. **Miller DG, Trobridge GD, Petek LM, Jacobs MA, Kaul R, Russell DW.** 2005. Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J Virol **79:**11434-11442.

288. **Wang Z, Lisowski L, Finegold MJ, Nakai H, Kay MA, Grompe M.** 2012. AAV vectors containing rDNA homology display increased chromosomal integration and transgene persistence. Mol Ther **20:**1902-1911.

289. **Lisowski L, Lau A, Wang Z, Zhang Y, Zhang F, Grompe M, Kay MA.** 2012. Ribosomal DNA integrating rAAV-rDNA vectors allow for stable transgene expression. Mol Ther **20:**1912-1923.

290. **Russell DW.** 2012. AAV vectors for the nucleolus. Mol Ther **20:**1842-1843.

291. **Wobus CE, Hugle-Dorr B, Girod A, Petersen G, Hallek M, Kleinschmidt JA.** 2000. Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and

identification of capsid domains involved in AAV-2-cell interaction and neutralization of AAV-2 infection. J Virol **74:**9281-9293.

292. **Moskalenko M, Chen L, van Roey M, Donahue BA, Snyder RO, McArthur JG, Patel SD.** 2000. Epitope mapping of human anti-adeno-associated virus type 2 neutralizing antibodies: implications for gene therapy and virus structure. J Virol **74:**1761-1766.

293. **McCraw DM, O'Donnell JK, Taylor KA, Stagg SM, Chapman MS.** 2012. Structure of adeno-associated virus-2 in complex with neutralizing monoclonal antibody A20. Virology **431:**40-49.

294. **Grimm D, Kern A, Pawlita M, Ferrari F, Samulski R, Kleinschmidt J.** 1999. Titration of AAV-2 particles via a novel capsid ELISA: packaging of genomes can limit production of recombinant AAV-2. Gene Ther **6:**1322-1330.

295. **Kuck D, Kern A, Kleinschmidt JA.** 2007. Development of AAV serotype-specific ELISAs using novel monoclonal antibodies. J Virol Methods **140:**17-24.

296. **Chook YM, Suel KE.** 2011. Nuclear import by karyopherin-betas: recognition and inhibition. Biochim Biophys Acta **1813:**1593-1606.

297. **Kalderon D, Roberts BL, Richardson WD, Smith AE.** 1984. A short amino acid sequence able to specify nuclear location. Cell **39:**499-509.

298. **Robbins J, Dilworth SM, Laskey RA, Dingwall C.** 1991. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell **64:**615-623.

299. **Ben-Efraim I, Gerace L.** 2001. Gradient of increasing affinity of importin beta for nucleoporins along the pathway of nuclear import. J Cell Biol **152:**411-417.

300. **Musinova YR, Lisitsyna OM, Golyshev SA, Tuzhikov AI, Polyakov VY, Sheval EV.** 2011. Nucleolar localization/retention signal is responsible for transient accumulation of histone H2B in the nucleolus through electrostatic interactions. Biochim Biophys Acta **1813:**27-38.

301. **Wang Y, Chen B, Li Y, Zhou D, Chen S.** 2011. PNRC accumulates in the nucleolus by interaction with B23/nucleophosmin via its nucleolar localization sequence. Biochim Biophys Acta **1813:**109-119.

302. **Szebeni A, Herrera JE, Olson MO.** 1995. Interaction of nucleolar protein B23 with peptides related to nuclear localization signals. Biochemistry **34:**8037-8042.

303. **Kawano Y, Neeley S, Adachi K, Nakai H.** 2013. An experimental and computational evolution-based method to study a mode of co-evolution of overlapping open reading frames in the AAV2 viral genome. PLoS One **8:**e66211.

304. **Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, Corbett AH.** 2007. Classical nuclear localization signals: definition, function, and interaction with importin alpha. J Biol Chem **282:**5101-5105.

305. **Weber JD, Kuo ML, Bothner B, DiGiammarino EL, Kriwacki RW, Roussel MF, Sherr CJ.** 2000. Cooperative signals governing ARF-mdm2

interaction and nucleolar localization of the complex. Mol Cell Biol **20:**2517-2528.

306. **Viranaicken W, Gasmi L, Chaumet A, Durieux C, Georget V, Denoulet P, Larcher JC.** 2011. L-Ilf3 and L-NF90 traffic to the nucleolus granular component: alternatively-spliced exon 3 encodes a nucleolar localization motif. PLoS One **6:**e22296.

307. **Scott MS, Boisvert FM, McDowall MD, Lamond AI, Barton GJ.** 2010. Characterization and prediction of protein nucleolar localization sequences. Nucleic Acids Res **38:**7388-7399.

308. **Lange A, McLane LM, Mills RE, Devine SE, Corbett AH.** 2010. Expanding the definition of the classical bipartite nuclear localization signal. Traffic **11:**311-323.

309. **Kosugi S, Hasebe M, Matsumura N, Takashima H, Miyamoto-Sato E, Tomita M, Yanagawa H.** 2009. Six classes of nuclear localization signals specific to different binding grooves of importin alpha. J Biol Chem **284:**478-485.

310. **Svistunova DM, Musinova YR, Polyakov VY, Sheval EV.** 2012. A simple method for the immunocytochemical detection of proteins inside nuclear structures that are inaccessible to specific antibodies. J Histochem Cytochem **60:**152-158.

311. **Kosugi S, Yanagawa H, Terauchi R, Tabata S.** 2014. NESmapper: accurate prediction of leucine-rich nuclear export signals using activity-based profiles. PLoS Comput Biol **10:**e1003841.

312. **Fu SC, Imai K, Horton P.** 2011. Prediction of leucine-rich nuclear export signal containing proteins with NESsential. Nucleic Acids Res **39:**e111.

313. **Prieto G, Fullaondo A, Rodriguez JA.** 2014. Prediction of nuclear export signals using weighted regular expressions (Wregex). Bioinformatics **30:**1220-1227.

314. **Hiscox JA.** 2002. The nucleolus--a gateway to viral infection? Arch Virol **147:**1077-1089.

315. **Pederson T.** 2011. The nucleolus. Cold Spring Harb Perspect Biol **3**.

316. **Cole C, Barber JD, Barton GJ.** 2008. The Jpred 3 secondary structure prediction server. Nucleic Acids Res **36:**W197-201.

317. **Liu J, Du X, Ke Y.** 2006. Mapping nucleolar localization sequences of 1A6/DRIM. FEBS Lett **580:**1405-1410.

318. **Olson MO, Dundr M.** 2005. The moving parts of the nucleolus. Histochem Cell Biol **123:**203-216.

319. **Hiscox JA.** 2007. RNA viruses: hijacking the dynamic nucleolus. Nat Rev Microbiol **5:**119-127.

320. **Schmidt-Zachmann MS, Nigg EA.** 1993. Protein localization to the nucleolus: a search for targeting domains in nucleolin. J Cell Sci **105 (Pt 3):**799-806.

321. **Aris JP, Blobel G.** 1991. cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera. Proc Natl Acad Sci U S A **88:**931-935.

322. **Fankhauser C, Izaurralde E, Adachi Y, Wingfield P, Laemmli UK.** 1991. Specific complex of human immunodeficiency virus type 1 rev and nucleolar B23 proteins: dissociation by the Rev response element. Mol Cell Biol **11:**2567-2575.

323. **Borer RA, Lehner CF, Eppenberger HM, Nigg EA.** 1989. Major nucleolar proteins shuttle between nucleus and cytoplasm. Cell **56:**379-390.

324. **Valdez BC, Perlaky L, Henning D, Saijo Y, Chan PK, Busch H.** 1994. Identification of the nuclear and nucleolar localization signals of the protein p120. Interaction with translocation protein B23. J Biol Chem **269:**23776-23783.

325. **Adachi Y, Copeland TD, Hatanaka M, Oroszlan S.** 1993. Nucleolar targeting signal of Rex protein of human T-cell leukemia virus type I specifically binds to nucleolar shuttle protein B-23. J Biol Chem **268:**13930-13934.

326. **Richardson WD, Roberts BL, Smith AE.** 1986. Nuclear location signals in polyoma virus large-T. Cell **44:**77-85.

327. **Howes SH, Bockus BJ, Schaffhausen BS.** 1996. Genetic analysis of polyomavirus large T nuclear localization: nuclear localization is required for productive association with pRb family members. J Virol **70:**3581-3588.

328. **Melen K, Kinnunen L, Fagerlund R, Ikonen N, Twu KY, Krug RM, Julkunen I.** 2007. Nuclear and nucleolar targeting of influenza A virus NS1 protein: striking differences between different virus subtypes. J Virol **81:**5995-6006.

329. **Kulshreshtha V, Ayalew LE, Islam A, Tikoo SK.** 2014. Conserved arginines of bovine adenovirus-3 33K protein are important for transportin-3 mediated transport and virus replication. PLoS One **9:**e101216.

330. **Arnold M, Nath A, Hauber J, Kehlenbach RH.** 2006. Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1. J Biol Chem **281:**20883-20890.

331. **Cros JF, Garcia-Sastre A, Palese P.** 2005. An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. Traffic **6:**205-213.

332. **Cokol M, Nair R, Rost B.** 2000. Finding nuclear localization signals. EMBO Rep **1:**411-415.

333. **Hu W, Philips AS, Kwok JC, Eisbacher M, Chong BH.** 2005. Identification of nuclear import and export signals within Fli-1: roles of the nuclear import signals in Fli-1-dependent activation of megakaryocyte-specific promoters. Mol Cell Biol **25:**3087-3108.

334. **LaCasse EC, Lefebvre YA.** 1995. Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. Nucleic Acids Res **23:**1647-1656.

335. **Lingappa VR, Hurt CR, Garvey E.** 2013. Capsid assembly as a point of intervention for novel anti-viral therapeutics. Curr Pharm Biotechnol **14:**513-523.

336. **Mateu MG.** 2013. The Structural Basis of Virus Function, p 3-51. *In* Mateu MG (ed), Structure and Physics of Viruses: An Integrated Textbook. Springer, New York.

337. **Trempe JP, Carter BJ.** 1988. Regulation of adeno-associated virus gene expression in 293 cells: control of mRNA abundance and translation. J Virol **62:**68-74.

338. **Huang MT, Gorman CM.** 1990. Intervening sequences increase efficiency of RNA 3' processing and accumulation of cytoplasmic RNA. Nucleic Acids Res **18:**937-947.

339. **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.

340. **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.

341. **Grimm D, Zhou S, Nakai H, Thomas CE, Storm TA, Fuess S, Matsushita T, Allen J, Surosky R, Lochrie M, Meuse L, McClelland A, Colosi P, Kay MA.** 2003. Preclinical in vivo evaluation of pseudotyped adeno-associated virus vectors for liver gene therapy. Blood **102:**2412-2419.

342. **Strobel B, Miller FD, Rist W, Lamla T.** 2015. Comparative Analysis of Cesium Chloride- and Iodixanol-Based Purification of Recombinant Adeno-Associated Viral Vectors for Preclinical Applications. Hum Gene Ther Methods **26:**147-157.

343. **O'Donnell J, Taylor KA, Chapman MS.** 2009. Adeno-associated virus-2 and its primary cellular receptor--Cryo-EM structure of a heparin complex. Virology **385:**434-443.

344. **Perez R, Castellanos M, Rodriguez-Huete A, Mateu MG.** 2011. Molecular determinants of self-association and rearrangement of a trimeric intermediate during the assembly of a parvovirus capsid. J Mol Biol **413:**32-40.