Factors affecting AAV production: Mutagenesis of AAV2 rep and AAV transgenes

HARVARD
Kenneth C. Griffin

GRADUATE SCHOOL
OF ARTS AND SCIENCES

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Medical Sciences

in the subject of Biological and Biomedical Sciences

have examined a dissertation entitled

Factors affecting AAV production: Mutagenesis of AAV2 rep and AAV transgenes

presented by Nina K. Jain

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: Aaron G. Schmidt (Nov 21, 2023 15:40 EST)

Typed Name: Dr. Aaron Schmidt

Signature: Amy Wagers (Nov 21, 2023 15:38 EST)

Typed Name: Dr. Amy Wagers

Signature: Casey Maguire (Nov 21, 2023 15:48 EST)

Typed Name: Dr. Casey Maguire

Signature: David Schaffer (Nov 21, 2023 12:37 PST)

Typed Name: Dr. David Schaffer

Date: November 21, 2023

Factors affecting AAV production: Mutagenesis of AAV2 *rep* and AAV transgenes

A DISSERTATION PRESENTED
BY
NINA K. JAIN
TO
THE DIVISION OF MEDICAL SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
BIOLOGICAL AND BIOMEDICAL SCIENCES

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
NOVEMBER 2023

Thesis advisor: Professor George M. Church Nina K. Jain

Factors affecting AAV production: Mutagenesis of AAV2 *rep* and AAV transgenes

Abstract

Recombinant adeno-associated viruses (rAAVs) are the predominant gene therapy vector. Previous work in the AAV field focused on capsid engineering to enable targeting of vectors to specific tissues and has resulted in five approved therapies. Here, I turn my attention to the effect of *rep* and transgene sequence variants on rAAV production. As more rAAV therapies enter the clinic, manufacturing sufficient rAAV quantities to meet patient needs is increasingly difficult. By investigating two relatively understudied components of AAV production, I seek to improve our understanding of the AAV production process and, potentially, identify methods of enhancing it. In Chapter 1, I focus on AAV2 *rep*, which codes for four proteins that are critical for both AAV genome replication and packaging. I measure the effect of all single codon mutations to AAV2 *rep* on AAV production and identify numerous novel functional *rep* variants. This work highlights the importance of the DNA-interacting regions of Rep for AAV production enhancement and provides a useful dataset to enable further Rep engineering. In Chapter 2, I examine the effect of transgene sequence on AAV production. Here, I use synonymous recoding to study the effect of non-coding nucleotide changes to both wild-type and rAAV genomes on production. I demonstrate that nucleotide sequence - absent protein-coding changes - can affect rAAV production. Future work may identify generalizable principles for rAAV transgene design. Through an investigation of *rep* and transgene variants, this thesis informs our understanding of AAV production and lays the groundwork for improvements in large-scale gene therapy production.

Contents

Listing of figures

To my grandfather, Michael Doyle Hughes (1933-2020), the greatest scientist I've ever met.

Acknowledgments

First and foremost, I would like to thank my adviser, George. I am nothing if not a details person, but his willingness to take on big questions has encouraged me to do the same. I'm a better scientist for it. This dissertation would also not have been possible without the mentorship and collaboration of Pierce Ogden. He is a rare kind of scientist that has both the vision to dream up ambitious projects and the technical skill and perseverance to bring them to fruition. He has very generously continued to support both me and my projects even years after leaving the lab, and I'm incredibly grateful to have worked with and learned from him. I would also like to acknowledge the myriad members of the Church lab, who are too numerous to name, but provided essential advice, support, and ideas at different stages of this work.

Thank you to the many mentors I worked with prior to grad school. It is your advice and axioms I hear every time I plan an experiment, give a presentation, or run a Western blot. Thank you to Sharmistha Sarkar, who took me on as a bumbling undergrad and taught me to pipette. To Karen Hasty, who showed me what it looks like to be a woman in science and proved to me that there was a place for me in the research community. To Cait Costello and John March for their continued

support and encouragement. Finally, to Walt Voegtli and Rich Altman. Walt, thank you for sharing your technical expertise, excitement for science, and sarcasm. Rich, you are the kind of scientist and colleague I aspire to be. Thank you for your knowledge, empathy, and generosity.

I would also like to thank the many friends that have supported me throughout grad school. Spending time with you gave me a sense of joy and community when the negative data seemed never ending. To Evi Van Itallie, thank you for being my pandemic buddy and showing me that there is a light at the end of the tunnel. To Aleks Prochera, I am so grateful to have gone through grad school along side you. Thank you for your excellent revision skills, discerning palette, and unending thoughtfulness. To Chara Koutsioumpa, thank you for the constant support and commiseration. Lastly, to Kristen Ewing, thank you for your humor and for always being in my corner. And thank you for taking the time to understand my research well enough to explain it to your family.

Finally, I would like to thank my family. This dissertation may have my name on it, but it has been the work of generations to give me the opportunity to pursue it. To my grandparents, Carol and Michael Hughes and Santosh and Kailash Jain, your confidence in me made me just crazy enough to believe that I could do this. A special acknowledgement to my grandfather, Michael Hughes, who passed away in 2020 and to whom this dissertation is dedicated. He didn't have the opportunity to attend college, never mind grad school, but was without a doubt one of the greatest scientific minds I've had the pleasure of knowing. Thank you for being one of my biggest cheerleaders and for passing along a little bit of your brilliance to me. To my [many] aunts, uncles, and cousins, thank you for the storytelling, raucous family parties, and constant support. To my parents, Colette Hughes and Arun Jain, thank you for everything: your sacrifices, grit, unending empathy, and constant encouragement. Last but not least, to my sister, Maya. It is a true gift to be seen and understood the way that you see me. I couldn't have done it without you!

Nothing is less predictable than the development of an active scientific field.

Charles Francis Richter

0 Introduction

Adeno-associated virus (AAV) was first identified in 1965 as a "defective viral" contaminant of an adenovirus preparation [3]. This initial study demonstrated that AAV required a helper virus, such as adenovirus, to replicate and did not cause disease when administered to mice or hamsters. Over the last six decades, the structure and much of the replication and infection mechanisms of AAV were resolved [5,6,50,60,132]. Notably, Samulski et al. demonstrated that replication-deficient recombinant AAV (rAAV) vectors could be generated by flanking a gene of interest with inverted terminal repeat

sequences (ITRs) derived from the wild-type virus and supplying the rest of the wild-type genome *in trans*[77,97]. This work was critical to the application of AAV as a gene therapy vector and has ultimately resulted in five FDA approved therapies[86,90,95,111,136]. Much of the previous research in the AAV field centered on the capsid, as it is the main structural component of AAV and dictates tissue tropism. Motivated by a need for increased quantities of rAAV vectors, here I turn my attention to rAAV production, the generation of genome-containing and transducing particles. In this thesis, I explore the effects of *rep* and rAAV transgene variants on AAV production, with the goal of improving our understanding of the AAV production process and, potentially, identifying methods of enhancing it.

0.0.1 AAV genes and proteins

Wild-type AAV2, the prototypical AAV serotype, consists of a 4.7 kilobase single-stranded DNA genome, which is packaged into an approximately 26 nanometer icosahedral capsid[108,132]. The AAV genome is flanked on either end by 145 nucleotide long ITRs, which form hairpins, serve as the origins of viral replication, and are the only sequences required *in cis* for packaging of DNA into the capsid[131]. As such, the remainder of the AAV genome can be replaced with a gene of interest to generate rAAV vectors[98]. The wild-type AAV genome consists of two genes, *rep* and *cap* (Figure 1)[108]. During rAAV production, these genes are supplied *in trans* to the ITR plasmid. The *cap* gene encodes three structural proteins, VP1, VP2, and VP3, which assemble in an approximately 1:1:10 ratio to form the 60-mer capsid[17,129]. The *rep* gene encodes four proteins, Rep78, Rep68, Rep52, and Rep40, which are involved in genome replication and packaging and are generated through the use of two promoters, p5 and p19, and alternative splicing[30].

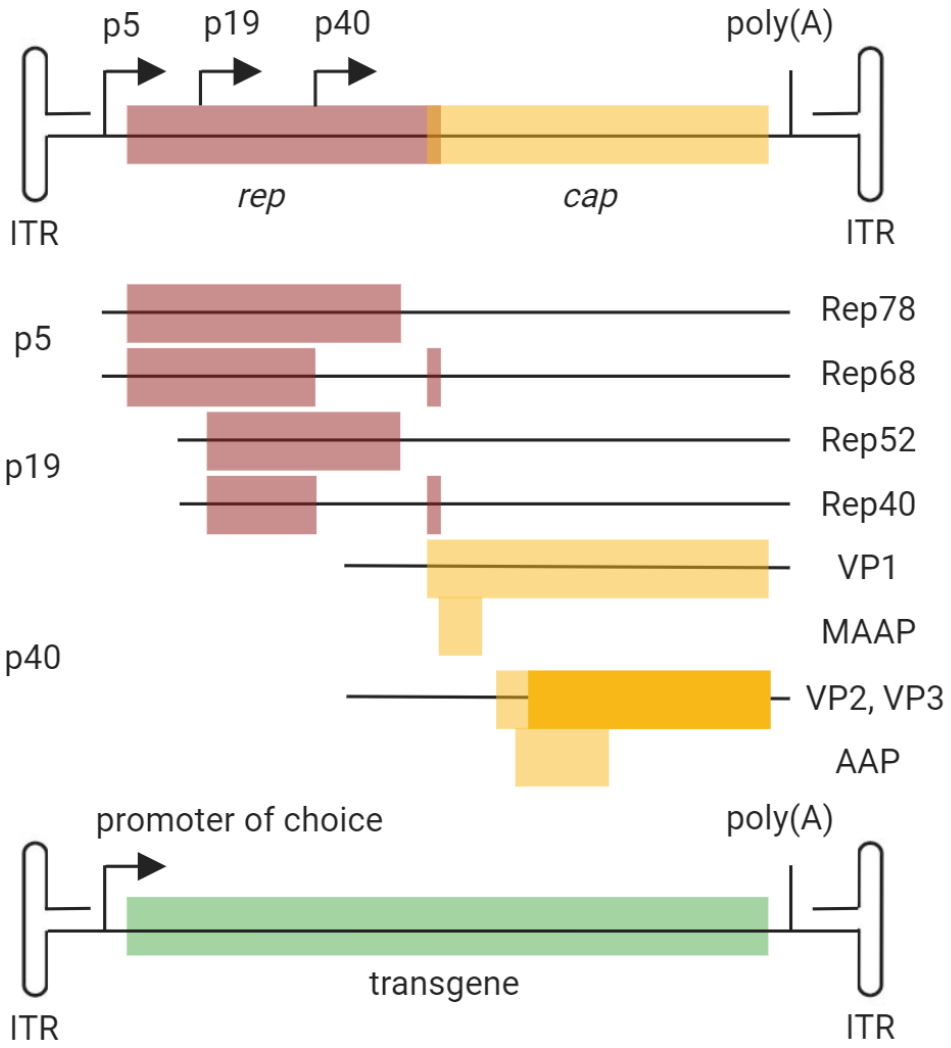


Figure 1: Organization of the wild-type AAV2 genome. Top: schematic of the single-stranded DNA genome, boxes indicate position of genes, middle: representation of AAV2 mRNA transcripts, boxes indicate protein coding open reading frames, bottom: depiction of a rAAV genome.

0.0.2 Capsid engineering for enhanced transduction

Previous work in the AAV field has centered on the capsid. The capsid is the main structural component of AAV and determines multiple parameters that are essential for use of AAV as a gene therapy vector, such as tissue tropism and neutralizing antibody evasion [107,120]. Capsid engineering efforts - as well as evaluation of diverse naturally occurring serotypes - have resulted in rAAV vectors capable of efficiently transducing multiple tissues, including motor neurons, skeletal muscle, and retinal epithelium [111,117,95]. However, work to improve rAAV transduction *via* capsid engineering is ongoing [46,66,1,85]. More recent efforts have focused on engineering capsids to target specific cell types, evade antibody recognition, and detarget the liver, where rAAV are predominantly sequestered after systemic injection [23,100,120].

The success of capsid engineering efforts and the subsequent increased clinical use of rAAV vectors has resulted in another problem: insufficient rAAV production. Current manufacturing processes yield 10^4-10^5 viral genomes per cell, or about 10^{14} viral genomes per liter of suspension cell culture [26,93]. However, doses greater than 10^{14} viral genomes per kilogram are required for therapies, such as delandistrogene moxeparvovec-rokl (Elevidys), which targets skeletal muscle for the treatment of Duchenne muscular dystrophy [111]. Improvements in rAAV production are required to support the continued application of rAAV as a gene therapy vector.

While previous studies of the AAV capsid have identified variants that enhance titer, here I turn my attention to other components of rAAV vectors and the AAV production process [85]. The capsid sequence must be optimized for numerous parameters outside of production, including stability, antibody evasion, and tissue targeting. By focusing on other components of rAAV production, I aim to enable capsid engineering efforts to concentrate on these downstream functions and/or to identify capsid-sequence-independent methods of enhancing rAAV production. In this dissertation, I investigate the effect of AAV2 *rep* and AAV transgene sequence variants on rAAV production.

The AAV2 *rep* gene and its products, the four Rep proteins, are essential for AAV production but understudied relative to the AAV capsid. Here, I seek to enhance our understanding of the Rep proteins and their mutational landscape. Similarly, previous studies demonstrated that the AAV genome or transgene sequence may affect AAV production, but the nucleotide sequence determinants of AAV production have not been explored[45,138]. Here, I aim to better understand the transgene-sequence-AAV production relationship.

0.1 Wild-type AAV replication and rAAV production

The rAAV production process is analogous to the later steps of wild-type AAV replication. Figure 2A summarizes the wild-type AAV2 replication process, and Figure 2B diagrams the process of rAAV2 production. In the case of wild-type AAV2, genome-containing capsids first bind to heparin sulfate proteoglycan receptors on the cell surface are are then taken up into the endosome[115,48]. They must escape the endosome and traffic to the nucleus where genome release primarily occurs[83,106]. Once exposed, the single-stranded genome is converted into a double-stranded, transcriptionally active form through the action of host cell polymerases[53]. In the case of rAAV production, the AAV *rep* and *cap* genes are introduced into the cell on plasmids that can be directly transcribed. In both cases, in the presence of helper virus genes - introduced either by co-infection of the host cell or on an accessory helper plasmid - Rep78/68 activate transcription from the AAV promoters[89]. Both adenovirus and herpes simplex virus can act as helper viruses to support wild-type AAV replication or rAAV production, but plasmid-based rAAV production methods most often make use of adenoviral helper genes[29,3]. Following Rep78/68 expression, VP expression is activated. The VPs assemble into small oligomers in the cytoplasm, which are stabilized by the assembly-activating protein (AAP)[72,126,125]. In the case of AAV2, AAP also facilitates transport of these oligomers into the nucleolus where they assemble to form empty capsids[33,34]. Previous work demonstrated that some

AAV serotypes, such as AAV2, are AAP-dependent, while other serotypes, such as AAVrh32.33, do not require AAP expression for capsid assembly[72].

Meanwhile, the larger Rep proteins, Rep78/68, facilitate AAV genome replication through a process known as rolling hairpin replication. Rep78/68 become covalently attached to the AAV genome during this process[103]. This attachment may act as a signal, directing genomes to pre-formed empty capsids into the nucleolus. Here, AAV2 genomes are inserted into capsids, likely through a channel at the five-fold axis of symmetry[8]. Genome packaging proceeds in a 3' to 5' direction and is facilitated by the Rep proteins[62]. Single-stranded genomes are almost universally associated with capsids, indicating that the process of genome replication may be linked to the process of genome packaging[5]. The smaller Rep proteins, Rep52/40, are not required for AAV production but have been shown to increase viral titers by facilitating genome packaging[56]. Depending on the capsid serotype used, genome-containing particles may be secreted from or retained in the cells[122]. The recently discovered membrane-associated accessory protein (MAAP) has been shown to facilitate secretion of AAV, possibly by facilitating AAV interaction with extracellular vesicles[85,36].

rAAV2 transduce cells in a process analogous to the early steps of wild-type AAV2 infection: rAAV2 vectors are endocytosed and trafficked to the nucleus, where the genome is exposed. If the rAAV2 genomes are single-stranded they must be converted into double-stranded forms before the transgenes can be expressed. Days to months post-transduction, rAAV2 genomes form large concatemers with adjacent genomes separated by ITRs[32,134]. These concatemers can enable long-term transgene expression or, in some instances, undergo repression[75,76].

0.2 rAAV production systems

rAAV production in the lab involves cell culture to generate genome-containing particles and purification to remove host cell proteins and empty capsids, which are generated at high levels in all

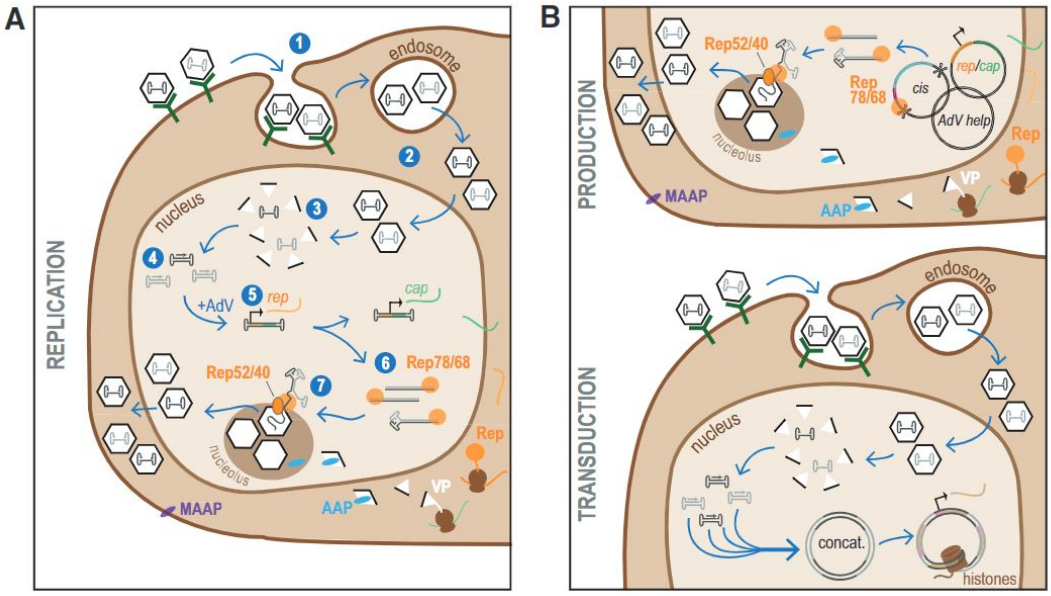


Figure 2: Comparison of wild-type AAV2 replication and rAAV2 production and transduction. *cis*: rAAV genome plasmid containing transgene and ITRs. Reproduced with permission from Maurer and Weitzman [73]

rAAV production systems. The majority of production systems utilize mammalian cells, either transient transfection of HEK293 cells, generation of stable cell lines, helper-virus dependent systems, or some combination thereof[93]. Baculovirus infection of Spodoptera frugiperda (Sf9) insect cells has also been used to produce rAAV[121]. Alipogene tiparvovec (Glybera), the first rAAV therapy to gain approval in Europe, was produced in Sf9 cells but has been withdrawn from the market[39]. Work from the Paulk group reported differences in post-translational modifications, rAAV genome methylation, and potency between HEK293- and baculovirus/Sf9-derived rAAV vectors[94]. As such, the field, and this review, are focused on mammalian production systems.

The most common system for rAAV vector production in both research and manufacturing contexts is transient transfection of HEK293 cells[130,93]. Generally, rAAV are produced *via* transfection of three different plasmids: a plasmid containing the transgene of interest flanked by two ITRs, a plasmid containing the adenovirus genes E4, E2A, and VA RNAs, and a plasmid carrying the AAV2

rep and *cap* genes (Figure 2B). HEK293 cells have copies of the adenoviral E1A and E1B genes integrated into their genomes; although these genes are required for rAAV production they do not need to be supplied on plasmids[69]. Additionally, while ITR and *rep* sequences from other AAV serotypes have been identified, AAV2 ITRs and AAV2 *rep* are used nearly universally for rAAV production across capsid serotypes. The broad use of AAV2 is likely due to its early identification and characterization, which have cemented it as the prototypical AAV serotype. Additionally, work from the Nakai lab demonstrated that the AAV2 *rep*-AAV2 ITR combination resulted in the highest titers for a panel of AAV capsids, even when compared to the cognate *rep* and ITRs for each capsid[96]. Transient transfection methods are advantageous as they allow for easy modification of the transgene, *rep*, or *cap* sequence. However, plasmid manufacturing to support large-scale transfection is costly[93]. Despite this, transient transfection remains the most popular method of rAAV production for clinical applications.

Stable cell lines are also used for rAAV production and are generated by integration of the AAV *rep* and *cap* genes, adenoviral helper genes, and rAAV genome. Importantly, the Rep78/68 proteins are cytostatic and, if under the control of the endogenous p5 and p19 promoters, their expression is stimulated by the adenoviral E1A proteins[20]. E1A expression also stimulates expression of the adenoviral E2A and E4 genes, which are important for rAAV production but toxic to cells[42]. Given this, stable cell lines must be generated to enable conditional expression of Rep78/68 and the adenoviral helper genes as constitutive or leaky expression can lead to cell cycle arrest and impede cell growth[7,59]. Stable cell lines do not require generation of large amounts of plasmid, as transient transfection does. However, generation of stable cell lines is time intensive and does not allow for easy modification of the transgene, *rep*, or *cap* sequence.

Finally, some mammalian rAAV production systems make use of helper viruses, namely adenovirus and herpes simplex virus. Wild-type adenovirus can be used along with adenoviral vectors to produce rAAV. In this instance, adenoviral vectors packaging both the *rep* and *cap* genes as well

as the rAAV genome are generated [114]. Adenoviral-based systems yield similar amounts of rAAV to transient transfection systems, and there is some evidence that rAAVs produced in adenoviral systems have greater potency than rAAVs produced by helper-virus free triple plasmid transfection [114]. However, great care must be taken to remove contaminating adenovirus from rAAV preparations, and the generation of recombinant adenoviral vectors carrying the rAAV genome and *rep* and *cap* genes can be costly. Similar herpes simplex virus-based systems have been reported [29]. Improvements in rAAV potency were also demonstrated with herpes simplex virus-based systems [25]. But, as with adenoviral systems, great care must be taken to remove contaminating herpes virus vectors from rAAV preparations, and the generation of recombinant herpes simplex virus can be complex and costly [26]. As such triple plasmid transfection remains the most common method of rAAV production for clinical use and it will be the focus of the experiments in this thesis. [26].

Whatever cell culture system is used, rAAV must be purified to remove host cell proteins, empty capsids, and possibly helper viruses. In research settings, ultracentrifugation is commonly used to enrich or purify completely genome-containing particles. Iodixanol ultracentrifugation makes use of a step gradient and allows for enrichment of genome-containing particles, while cesium-chloride centrifugation uses a continuous gradient to enrich genome-containing capsids to near 100% purity. However, cesium-chloride methods require multiple rounds of centrifugation and cesium-chloride purified vectors can induce adverse effects in animals [140,113]. While these ultracentrifugation methods work well in academic settings they do not scale effectively. As such, rAAV manufacturing processes generally make use of sequential column chromatography [35]. First, an affinity column is used to bind all capsids and remove host cell proteins. A second ion exchange step is then used to separate empty and genome-containing capsids based on charge. Notably, these ion exchange methods must be optimized for each capsid-rAAV genome pairing and complete removal of empty capsids is difficult. Therefore, there is interest in identifying methods of enhancing the proportion of genome-containing capsids in the crude cell harvest. Given the role of the Rep proteins in genome

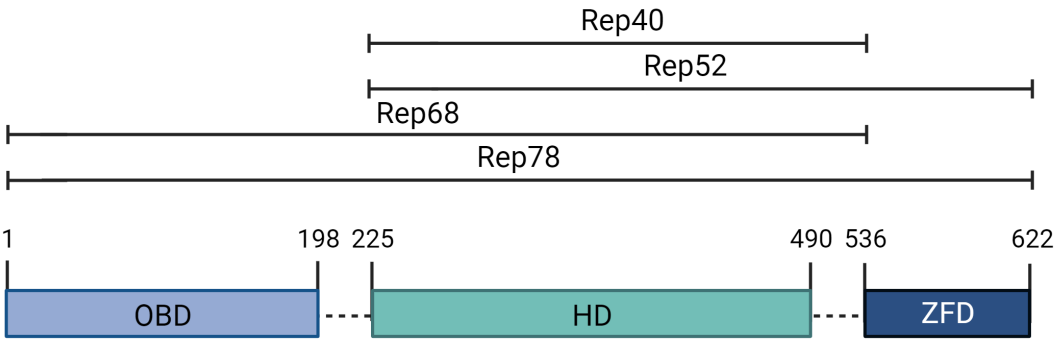


Figure 3: The *rep* gene codes for three protein domains. Each of the four Rep proteins, Rep78, Rep68, Rep52, and Rep40 contain a different combination of the three protein domains. OBD: origin-binding domain, HD: helicase domain, ZFD: zinc-finger domain. Numbers indicate Rep78 amino acid position.

replication and packaging, *rep* variants may enable increases in the proportion of full capsids.

0.3 Rep protein domains: Structure and function

The *rep* gene encodes three protein domains, an origin-binding domain, a helicase domain, and a zinc-finger domain (Figure 3)[31,58,102]. The origin-binding domain plays a critical role in genome replication, while the helicase domain is involved in both genome replication and packaging[58,103,11,62]. The zinc-finger domain interacts with various host cell proteins, but has it been shown to bind DNA not is it required for rAAV production[31]. All four Rep proteins contain the helicase domain as well as a nuclear localization signal[16]. Only Rep78 and Rep68 contain the origin-binding domain and only Rep78 and Rep52 contain the zinc-finger domain. Additionally, residues in the linker domain between the origin-binding and helicase domains cooperate with residues at the N-terminus of the helicase domain to facilitate oligomerization of the larger Rep proteins, which is required for their function and subsequently for AAV production[137].

The origin-binding domain contains three separate DNA binding motifs that are important for genome replication[50,51]. Firstly, the origin-binding domain recognizes the Rep binding site, which

consists of GCTC repeats present in the double stranded region of the inverted terminal repeats (Figure 4A)[82]. Unwinding of this double stranded DNA by the helicase domain enables the origin-binding domain to interact with single stranded DNA at its second motif, the active site pocket (Figure 4B). Here, the origin-binding domain nicks the single stranded DNA at the terminal resolution site, an essential step in viral genome replication[58,103]. Finally, the origin-binding domain contains a single stranded DNA hairpin binding site, which interacts with one of the inverted terminal repeat hairpins (Figure 4C). Hairpin binding, however, is not required for terminal resolution site nicking or genome replication[127]. The origin-binding domain can also recognize GCTC repeats present in the AAV p5 promoter; in the presence of a helper virus, such as adenovirus, the Rep proteins activate transcription from the endogenous promoters and in the absence of helper virus the Rep proteins repress transcription[64,81].

The helicase domain plays a definitive role in all steps of AAV production while the zinc-finger domain is seemingly dispensable for production. The helicase domain unwinds the double stranded inverted terminal repeats during genome replication and its activity has been shown to facilitate genome packaging (Figure 4D)[11,62]. There is also evidence that residues within the helicase domain mediate capsid interactions. Mutations to the helicase domain of Rep52/40, as well as mutations near the five-fold axis of symmetry in the capsid, reduced the interaction between the Rep proteins and capsid and resulted in lower rAAV titers[8,62]. The third Rep domain, the zinc-finger domain, has not been shown to bind DNA but is involved in binding to various host cell proteins[31]. Previous work indicated that premature stop codons introduced into the zinc-finger domain-encoding region of *rep* do not affect rAAV titers, suggesting that this domain is not required for rAAV production[78].

Engineering the Rep proteins is a promising avenue for improvement of rAAV production. The Rep proteins are involved in all steps of viral production, including genome replication, regulation of VP expression, and genome packaging. However, the Rep proteins are not structural components of the final vector. As such, the Rep proteins can be designed to optimize vector produc-

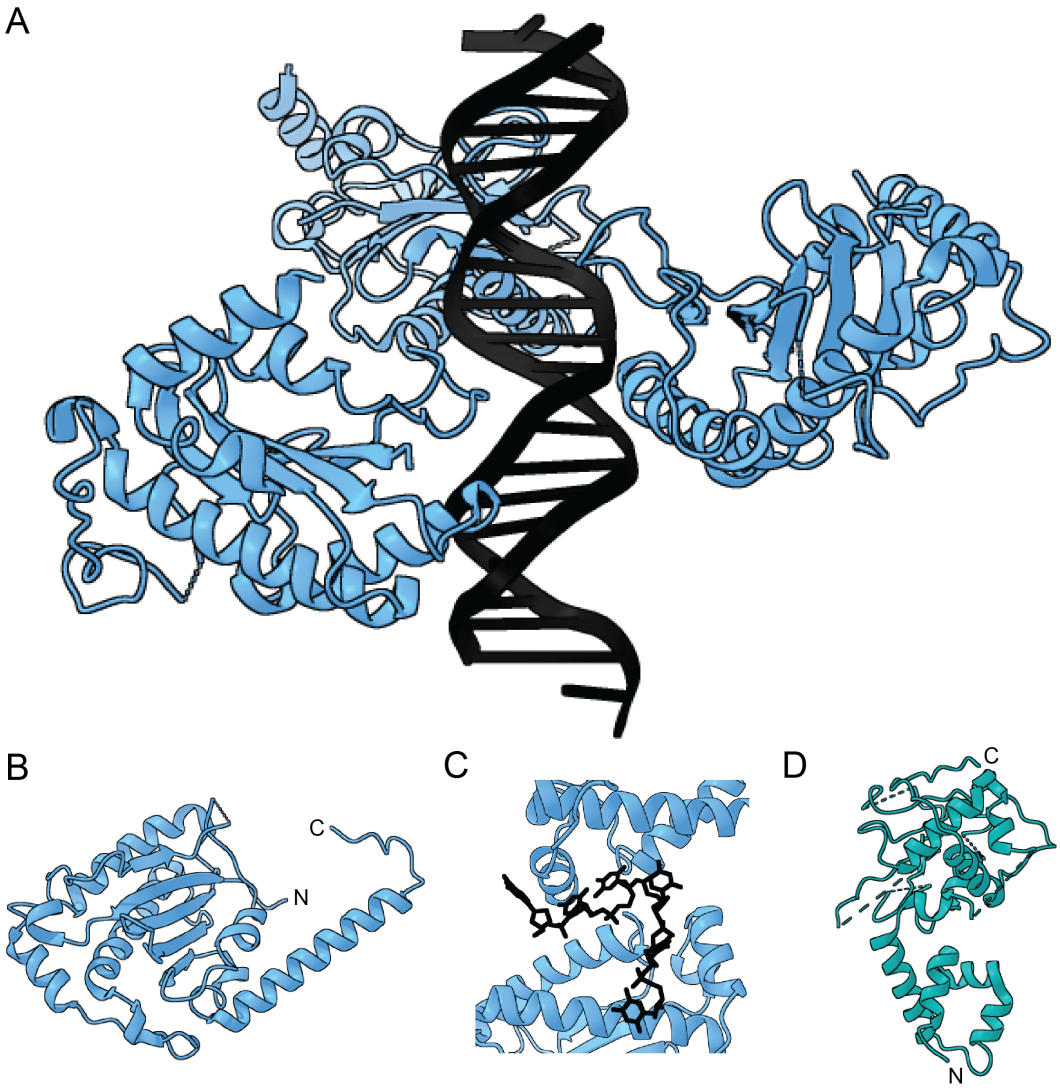


Figure 4: Structure of the origin-binding and helicase domains. (A) Three monomers of the origin-binding domain in complex with the Rep binding site in the double-stranded region of the ITR (PDB 4ZQ9). (B) Origin-binding domain active site (PDB 5DCX). (C) Two monomers of the origin-binding domain in complex with the single-stranded ITR hairpin (PDB 6XB8). (D) Helicase domain (PDB 1S9H).

tion without affecting downstream processes, such as cell targeting, which are driven by the capsid. While previous mutational studies of the Rep proteins have been conducted, these studies have focused primarily on identifying non-functional Rep variants to identify the location of key motifs and have not yielded *rep* variants that enhance production [30,40,57,135]. As such, Rep engineering efforts may benefit from the exploration of additional sequence diversity.

0.4 Engineering rAAV genomes

While there have been relatively few efforts to engineering Rep, there has been some work to engineer rAAV genomes. These efforts have centered on modifying rAAV genomes to facilitate transduction. Early work by McCarty, Monahan, and Samulski demonstrated that rAAV containing self-complementary genomes resulted in higher transduction efficiencies than rAAV containing single-stranded genomes [74]. By eliminating the need for second-strand synthesis, self-complementary rAAV vectors result in higher transgene expression. Other groups have demonstrated that inclusion of tissue- or cell-specific promoters can be used to restrict transgene expression to target tissues, thereby reducing immune responses generated when transgenes are expressed in antigen-presenting cells [2,19,27,55]. Efforts have also been directed at designing shorter promoter sequences to facilitate tissue-specific expression while remaining within the packaging limit of AAV [24,43]. Codon optimization algorithms are routinely used to design transgene sequences with improved expression levels. Some algorithms take into account tissue- and cell-specific codon bias to generate transgene sequences with enhanced expression specifically in the target cells [12]. Notably, codon optimization algorithms look at codon usage and tRNA abundance in different organisms and tissues. No algorithms have been developed to optimize codon usage specifically for transgenes delivered in rAAV vectors. Finally, the discovery that rAAV genomes can stimulate toll-like receptor 9 (TLR9) and activate anti-transgene and anti-capsid CD8+ T-cell responses has led groups to generate rAAV

genomes that are depleted for CpG sequences, the ligand for TLR9[139]. There is evidence that CpG depletion enhances transgene expression in animal models[37]. Taken together, these data demonstrate that rAAV genome engineering can be used to enhance vector transduction.

There have also been observations that different AAV genomes result in different genome-containing particle and transducing particle titers[45,138]. Zeltner et al. reported that wild-type AAV2 production resulted in a near 1:1 ratio of physical: infectious particles while rAAV2 production resulted in a 100:1 physical: infectious particle ratio, raising the possibility that the wild-type AAV2 genome may contain uncharacterized nucleotide sequences that facilitate production[138]. A similar study from Grimm et al. characterized the physical, genome-containing, and transducing particle titers of both wild-type AAV2 and a panel of different rAAV2 vectors[45]. Wild-type AAV2 was also reported to have a much lower physical: transducing particle ratio than rAAV2 here. Additionally, differences of an order of magnitude or more in the physical: transducing particle ratio of different rAAV vectors were observed, indicating that alterations to rAAV genome sequence - absent identification of production-enhancing sequences in the wild-type AAV2 genome - may be able to improve rAAV2 titers. Differences in the physical: genome-containing particle ratio, or the proportion of full capsids, were also observed. These reports, together with the success of efforts to modify the rAAV genome to enhance transduction, motivate a more in-depth interrogation of the effect of changes to rAAV genome sequence on production.

0.5 Comprehensive mutagenesis and synonymous recoding to understand AAV biology

Recent advances in DNA synthesis have enabled the generation of comprehensive mutagenesis libraries. Previously, our lab generated a library of all single codon variants of the AAV2 *cap* gene[85]. By barcoding individual variants and using Illumina sequencing to track library members through

screens for properties such as genome-containing particle production, tropism, and neutralizing antibody evasion, general principles of AAV assembly, tropism, and stability were inferred. This library also led to the identification of the MAAP open reading frame, highlighting the utility of unbiased approaches for understanding AAV biology. Here I apply the cloning and screening approaches developed for the *cap* gene to the *rep* gene to assay the effect of *rep* variants on AAV production.

Advances in DNA synthesis have also enabled facile synonymous recoding of entire genes, including viral genes. As discussed above, synonymous recoding in the form of codon optimization is routinely used to increase transgene expression, both in the context of rAAV vectors and for protein production generally. Conversely, codon deoptimization has been employed to generate vaccine candidates for a number of viruses, including poliovirus, vesicular stomatitis virus, and influenza virus [14,28,80,123]. Finally, synonymous recoding has also been used as a tool to study viral biology. By controlling for parameters such as codon bias, dinucleotide content, and mean free energy of RNA folding, synonymous recoding can be used to generate constructs with wild-type protein expression levels but vastly different nucleotide sequences [70]. These constructs can then be used to identify functional nucleotide motifs. For example, synonymous recoding of the poliovirus genome was used to identify two functional RNA elements in the capsid coding region [105]. Similarly, synonymous recoding of the AAV *rep* gene identified a DNA sequence that inhibited adenovirus replication when the *rep* gene was present in an adenoviral vector [101].

Here, I use comprehensive mutagenesis libraries and synonymous recoding to interrogate the effect of *rep* and transgene variants on AAV production. In Chapter 1, I report a detailed sequence-to-function map of the AAV2 *rep* gene that enhances our understanding of Rep protein function and provides a useful data set for further Rep engineering. In Chapter 2, I demonstrate that non-coding nucleotide changes to transgene sequences can affect rAAV production. This thesis lays the groundwork for enhancement of large-scale gene therapy production.

They might say, "it's not science," or "gentlemen don't do random mutagenesis." But I'm not a scientist, and I'm not a gentleman...

Frances Arnold

1

Comprehensive mutagenesis maps the effect of all single codon mutations in the AAV2 *rep* gene on AAV production

The contents of Chapter 1 were adapted from:

Jain, N.K., Ogden, P.J., Church, G.M. (2023). *Comprehensive mutagenesis maps the effect of all single codon mutations in the AAV2 rep gene on AAV production*. *eLife*, 12:RP87730.

Author contributions: N.K.J., P.J.O., and G.M.C. conceived the study and designed experiments. N.K.J. performed experiments. N.K.J. performed data analysis with input from P.J.O. N.K.J. wrote the paper with input from all authors.

rAAVs are a popular tool for the delivery of gene therapies as wild-type AAV is not pathogenic. The wild-type AAV genome consists of two genes, *rep* and *cap*, which are supplied *in trans* to the ITR-flanked transgene to enable rAAV production [108]. The *cap* gene encodes three structural pro-

teins, VP1, VP2, and VP3, which assemble in an approximately 1:1:10 ratio to form the 60-mer capsid[17,129]. Engineering of the *cap* gene has enabled targeting of AAV vectors to specific tissues and cell populations. The *rep* gene encodes four proteins, Rep78, Rep68, Rep52, and Rep40, which are generated through the use of two promoters, p5 and p19, and alternative splicing[30].

The larger Rep proteins, Rep78 and Rep68, are required for genome replication while the smaller Rep proteins, Rep52 and Rep40, facilitate genome packaging. Notably, Rep78 and Rep68 alone are each sufficient for rAAV vector production[56]. However, the presence of Rep52/40 enhances genome packaging and therefore rAAV titer[21,62]. The *rep* gene encodes three protein domains, an origin-binding domain, a helicase domain, and a zinc-finger domain (Figure 1A)[31,58,102]. All four Rep proteins contain the helicase domain as well as a nuclear localization signal[16]. Only Rep78 and Rep68 contain the origin-binding domain and only Rep78 and Rep52 contain the zinc-finger domain.

The origin-binding domain is responsible for sequence-specific recognition of the ITRs, ITR hairpin binding, and nicking of the AAV genome during replication[82,58,103,127]. The origin-binding domain also mediates recognition and regulation of the p5 promoter[64,81]. The helicase domain unwinds double-stranded DNA to facilitate nicking of the ITRs by the origin-binding domain and its activity is also important for genome packaging[11,62]. Previous studies demonstrated that premature stop codons introduced into the zinc-finger domain-encoding region of *rep* do not affect rAAV titers, suggesting that this domain is dispensable for rAAV production[78].

As more rAAV vectored therapies are approved, manufacturing sufficient rAAV quantities to meet patient need is an increasingly important issue. Engineering the Rep proteins is a promising avenue for improving AAV production. The Rep proteins are involved in all steps of viral production, but are not structural components of the final vector. As such, the Rep proteins can be designed to optimize viral production without affecting downstream processes, such as cell targeting, which are driven by the capsid. The sequence of the *rep* genes in naturally occurring serotypes is relatively

well conserved (average of 78% amino acid sequence identity between AAV2 *rep* and AAV1-13 *rep*). While previous mutational studies of the Rep proteins have been conducted, these studies have focused primarily on identifying non-functional Rep variants to identify the location of key motifs [30,40,57,135]. As such, Rep engineering efforts may benefit from the exploration of additional sequence diversity.

Towards this end, we generated a library of all possible single codon variants of the AAV2 *rep* gene, the *rep* gene most commonly used for AAV production. We assayed the effect of these mutations on AAV2 production and generated a detailed sequence-to-function map of the AAV2 *rep* gene.

1.1 Results

1.1.1 Comprehensive mutagenesis assays the effect of all single codon mutations in the AAV2 *rep* gene on AAV2 production

To better understand how mutations to *rep* affect AAV production, we generated two *rep* libraries. The first library, referred to as pCMV-Rep78/68, consisted of a cytomegalovirus (CMV) promoter followed by the *rep* open reading frame with a M225G mutation introduced to prevent expression of the smaller Rep proteins. This library allowed us to assay the effect of mutating only Rep78 and Rep68. The second library, referred to as WT AAV2, was analogous to the sequence of the wild-type virus and contained the p5 promoter and *rep* and *cap* open reading frames. Use of the endogenous promoters in the WT AAV2 library allowed us to capture the effect of *rep* mutations on Rep and VP expression. This library also enabled simultaneous mutation of all four Rep proteins. For each library, we generated all 39,297 possible single codon substitution and deletion variants spanning from the Rep78/68 start codon to the Rep78/52 stop codon (Figure 1.1A). All variants contained unique 20 base pair barcodes at the 3' end of the *rep* gene (pCMV-Rep78/68 library) or *cap* gene (WT AAV2 library), and each codon variant was represented by a minimum of two barcode

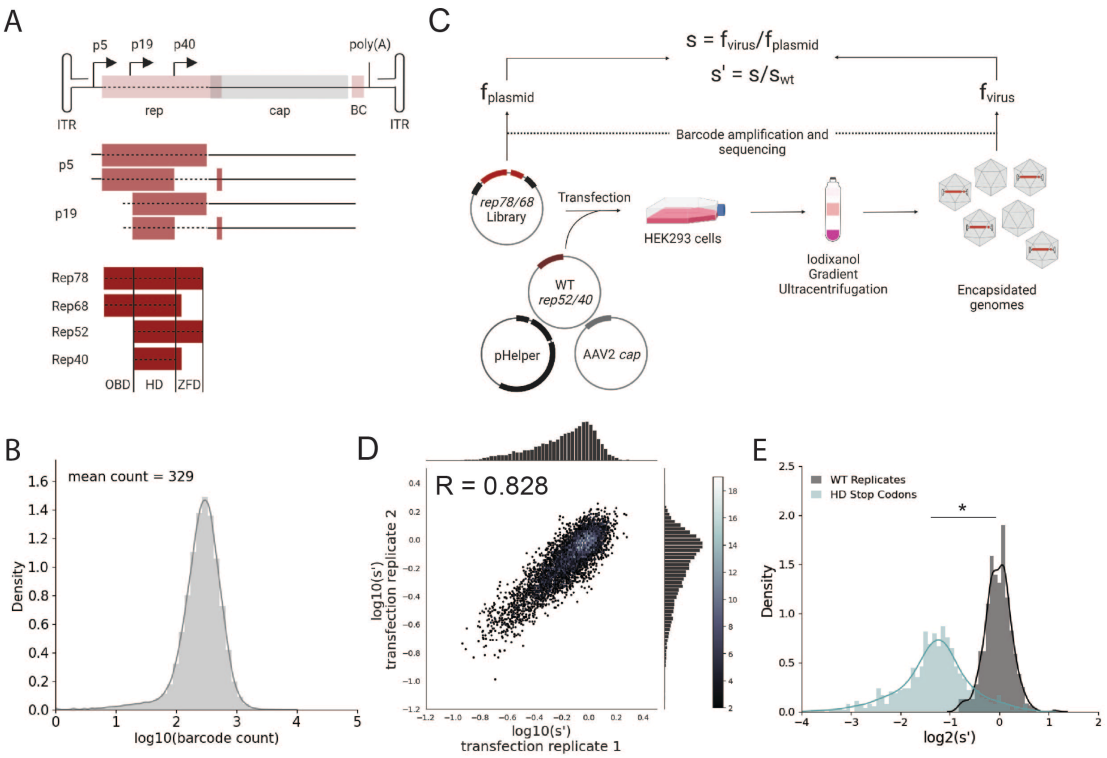


Figure 1.1: Comprehensive mutagenesis library design and production assay. (A) Organization of the AAV2 genome and Rep protein domains. Top: single-stranded DNA genome, middle: RNA transcripts, bottom: Rep proteins. Dotted lines indicate mutated regions. (B) Density plot of barcode counts in the pCMV-Rep78/68 plasmid library. (C) Overview of production assay for the pCMV-Rep78/68 library and calculation of wild-type normalized production fitness values (s'). (D) Amino acid level production fitness values from replicate transfections of the pCMV-Rep78/68 library. Pearson R correlation coefficient calculated after log transformation. (E) Density plot of production fitness values for wild-type (black) and premature stop codon (blue) controls for the pCMV-Rep78/68 library. *p < 10-20 (Mann Whitney U test)

sequences. To assess the diversity of the plasmid libraries, we amplified and sequenced the barcodes from each of the plasmid pools (Figure 1.1B, Figure A.1A, and Table A.1). We observed 100 and 99.9% of the expected amino acid variants in the pCMV-Rep78/68 and WT AAV2 libraries, respectively. Next, we transfected the plasmid libraries into HEK293T cells to assay the effect of all single codon mutations on production of genome-containing viral particles (Figure 1.1C). Production fitness values (s) for each variant were calculated and normalized to internal wild-type controls as shown in Figure 1C. For each library, we performed transfections in duplicate and compared

the production fitness values calculated from each replicate (Figure 1.1D and Figure A.1B). For both libraries, production fitness values were well-correlated across replicates, indicating that the genotype-phenotype linkage was maintained during viral production. After examining the correlation between biological replicates, we next looked at the distribution of fitness values for the wild-type and premature stop codon controls (Figure 1.1E and Figure A.1C). As expected, premature stop codons had a deleterious effect on AAV2 production and fitness values for replicate wild-type controls clustered together.

1.1.2 Annotation of the AAV2 *rep* sequence-to-function map

To better understand the sequence-to-function map of the AAV2 Rep proteins, we visualized the data from our production assay in multiple ways. First, we generated heatmaps containing the wild-type normalized production fitness values for all single amino acid substitutions and deletions (Figure 1.2 and Figure A.2). Additionally, we calculated the "mutability" of each residue by averaging the normalized production fitness values at each position across all substitutions. We mapped the mutability of each residue onto structures of the origin-binding domain in complex with the Rep binding site (PDB: 4ZQ9, Figures 1.3A-B), the origin-binding domain in complex with single-stranded DNA from the inverted terminal repeat hairpin (PDB: 6XB8, Figure 1.3C), the origin-binding domain alone (PDB: 5D6X, Figure A.5A), and the helicase domain (PDB: 1S9H, Figure A.5B)[60,82,99] Finally, we used the production fitness values calculated for individual barcodes to determine which amino acid changes resulted in production fitness values significantly different from wild-type (Figure A.4A-B). We observed that the origin-binding and zinc-finger domains were more tolerant of mutation than the helicase domain. Interestingly, a stark boundary between amino acids D212 and A213 was observed; at residue A213 the larger Rep proteins become much less tolerant of mutation. This boundary is adjacent to linker domain residues P214-Y224, which were previously reported to be important for Rep78/68 oligomerization[82,137]. When looking across

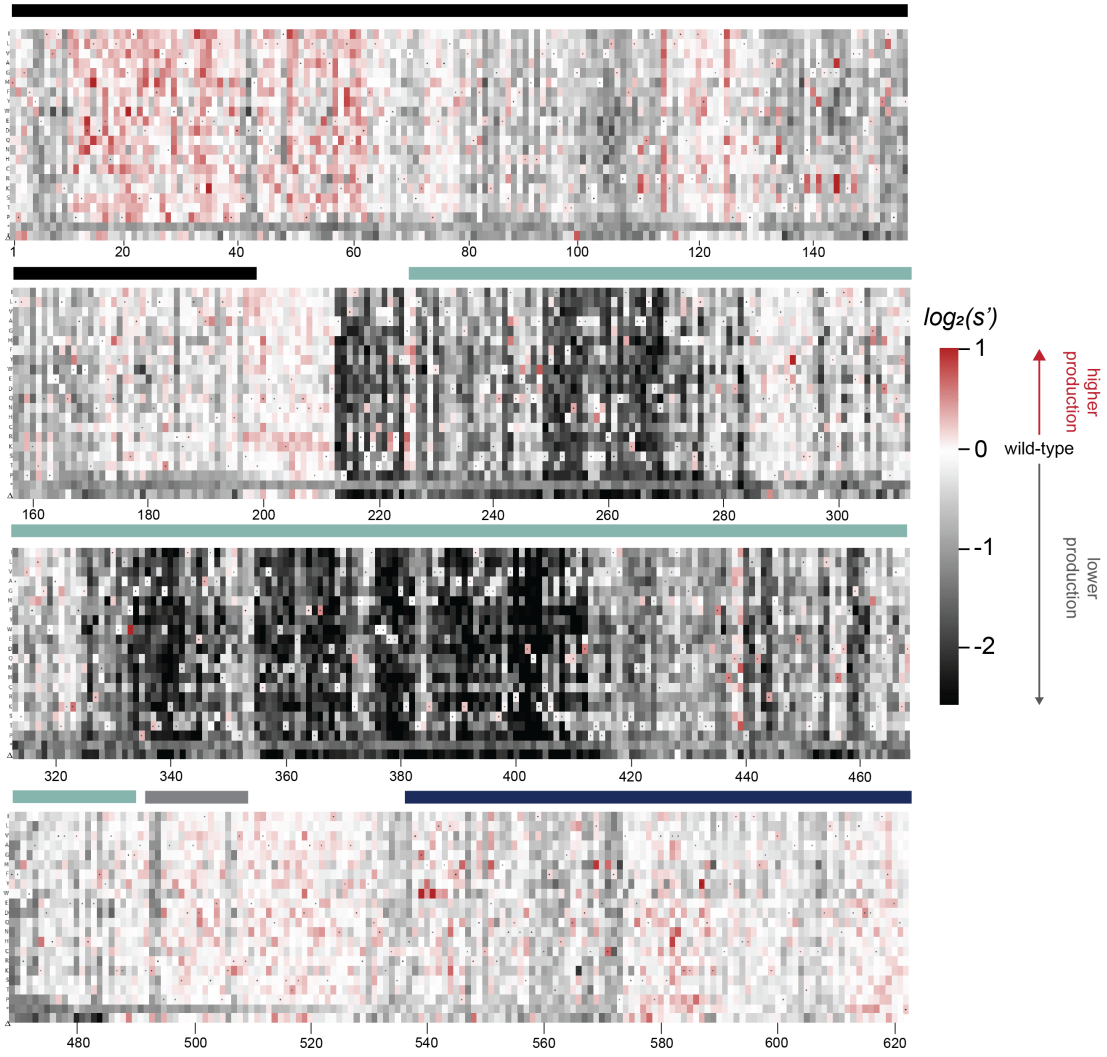


Figure 1.2: Effects of all single amino acid substitutions and deletions in the Rep78/68 proteins on AAV2 production. Amino acid level production fitness values from the pCMV-Rep78/68 production assay were calculated as in Figure 1.1C by summing barcode counts for synonymous mutations. Rectangles are colored by mutational effect on the production of genome-containing particles, with black indicating deleterious mutations and red indicating beneficial mutations. Colored bars above the heatmaps indicate protein domains. Black: origin-binding domain, light blue: helicase domain, gray: nuclear localization signal, and navy blue: zinc-finger domain. Black dots indicate wild-type amino acid identity.

all library members, the production fitness values determined for the pCMV-Rep78/68 and WT AAV2 libraries were well-correlated (Figure A.3A). However, as expected, introduction of a methionine upstream of the Rep52/40 start codon was deleterious in the WT AAV2 format but not in the pCMV-Rep78/68 format, where Rep52/40 were expressed *in trans*.

The majority of beneficial substitutions clustered in the origin-binding domain. In particular, substitutions between residues V11 and D62 of the origin-binding domain enhanced AAV2 production relative to wild-type. These residues are involved in recognition of the inverted terminal repeat hairpin (Figure 1.3C)[82]. Origin-binding domain-hairpin interactions have been shown to enhance terminal resolution site nicking but are not required for nicking to occur[127]. The majority of Rep binding site-interacting residues, on the other hand, were intolerant of mutation (Figure 1.3B). The origin-binding domain-Rep binding site interaction is important in mediating Rep-genome interactions during both genome replication and promoter regulation[64,81,82]. Interestingly, there are a handful of Rep binding site-interacting residues that were somewhat mutable, including S110 and N139, which form contacts with the phosphate backbone, and A141, which forms contacts with bases in the Rep binding site. Mutation of S110, N139, and A141, as well as G144 and V147, to positively charged residues had a beneficial effect on AAV2 production. Residues N139, A141, G144, and V147 are part of the loop that connects "β-strand 4 to "α-helix E[82]. The sequence of this loop is highly variable across serotypes[82]. Notably, the asparagine at position 139 is positively charged in AAV5 (K139). However, no serotypes contain positively charged amino acids at positions 141, 144, or 147. Our data indicate that substitution of DNA-interacting residues in the origin-binding domain can enhance AAV production and identifies beneficial substitutions not observed in nature. A clear pattern in the location of mutable residues in the origin-binding domain active site can also be observed (Figure A.5A). This active site is responsible for cleaving single stranded DNA and is formed by the origin-binding domain beta sheet and Y156, the active site nucleophile[51]. Residues with side chains directed towards the active site are less mutable than adja-

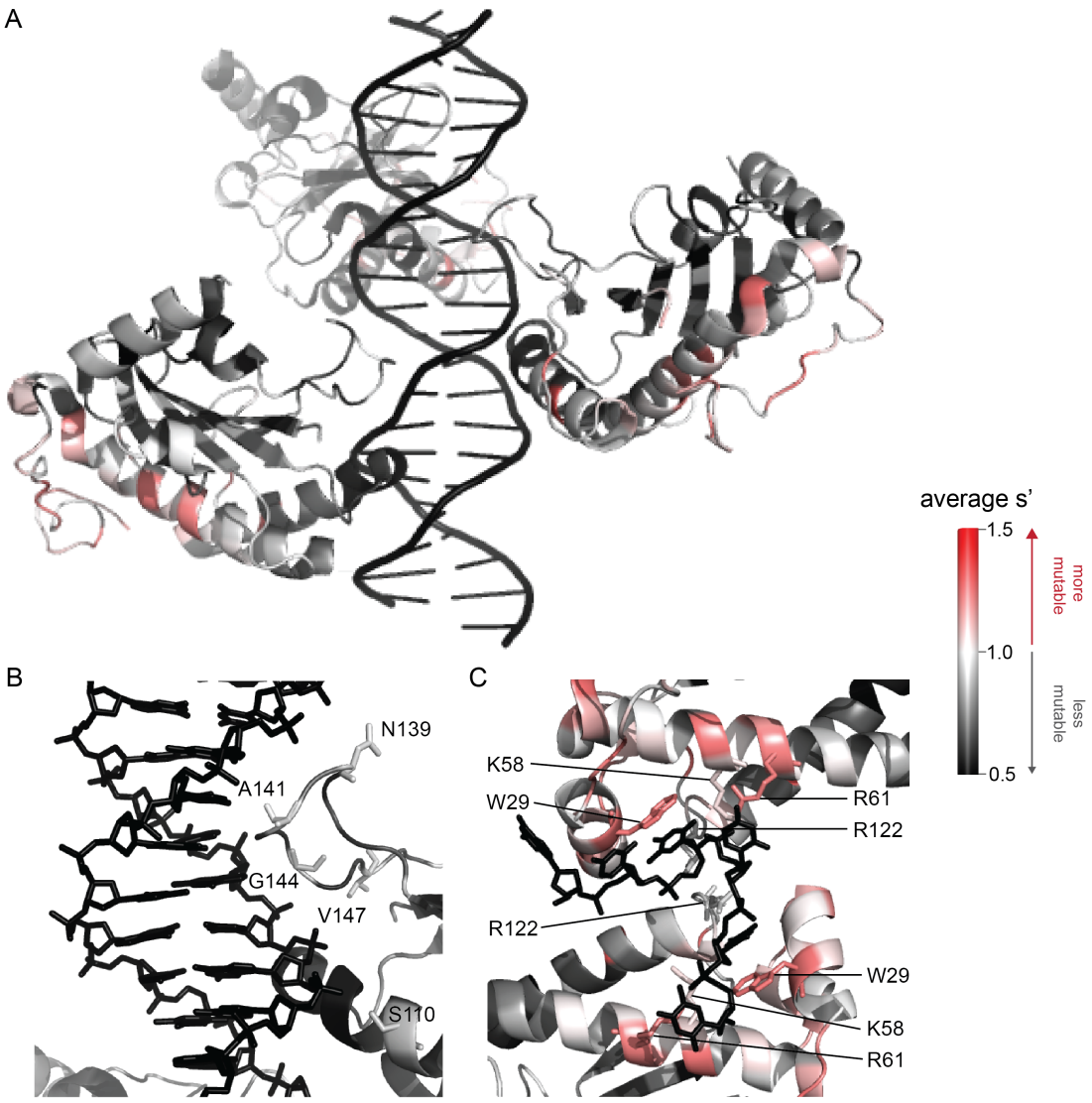


Figure 1.3: Beneficial substitutions cluster in DNA-interacting residues. (A) Average production fitness values for all substitutions at each position mapped onto the structure of the origin-binding domain in complex with the AAVS1 Rep binding site (PDB 4ZQ9). (B) Close up view of origin-binding domain-Rep binding site interactions. Residues where substitutions to positively charged residues are beneficial are shown as sticks. (C) Average production fitness values mapped onto the structure of the origin-binding domain in complex with the single stranded inverted terminal repeat hairpin (PDB 6XB8). DNA interacting residues are shown as sticks. Residues are colored by mutability, with red indicating higher mutability and black indicating lower mutability.

cent residues with their side chains directed away from the active site. It is more difficult to discern positional trends in mutability within the helicase domain as it is less tolerant of mutation than the origin-binding domain (Figure A.5B). No structures of the ZFD are available. However, the majority of variants in this domain do not have production fitness values that are significantly different from wild-type (Figures A.4A-B).

1.1.3 Generation of all single codon variants enables interrogation of nucleotide-level effects

As our libraries contain all possible single codon mutations, we were able to search for variants with nucleotide-level effects on production by comparing differences in production fitness values between synonymous codon variants (Figures A.6-7). In general, there was good agreement in the fitness values for synonymous variants. The production fitness values for synonymous codon variants were more tightly distributed than the fitness values for nonsynonymous variants (Figures A.8A-B). Comparison of fitness values between synonymous codons that do and do not introduce premature stop codons into alternate reading frames allowed us to search for possible frameshifted open reading frames. However, no evidence of frameshifted open reading frames was observed (Figures A.9A-B).

We did observe an interesting pattern at amino acid Y283 (*rep* nucleotides 847-849). Variants with a c.849C>G mutation had lower production fitness values than synonymous variants without a c.849C>G ($p < 10^{-8}$, Mann-Whitney U-test). The deleterious effect of this mutation can also be observed by plotting the average fitness value at each nucleotide position for each of the four bases (Figures A.9C-D). The negative effect of a c.849C>G mutation was consistent across the pCMV-Rep78/68 and WT AAV2 library formats and was also observed in experiments with alternate *cap* genes. In each library, there were fifteen codon variants with a c.849C>G mutation, thirteen of which had synonymous codons without a c.849C>G mutation for comparison. Our results indi-

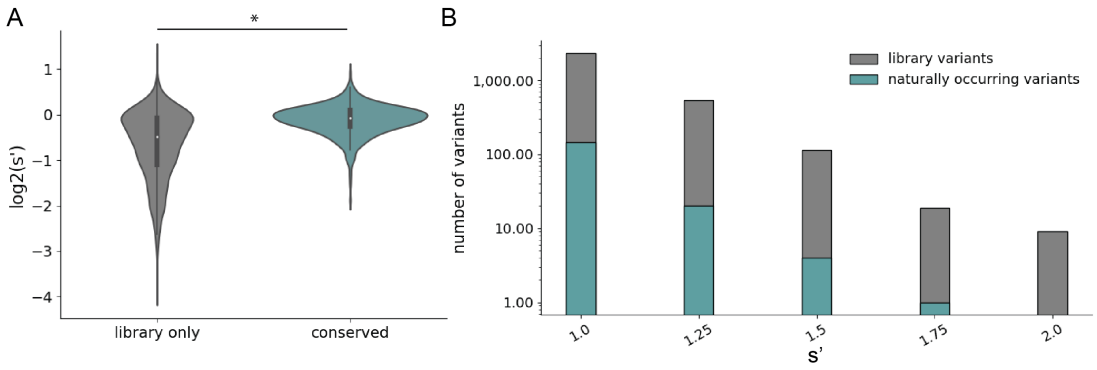


Figure 1.4: Comparison of comprehensive mutagenesis measurements to variation in nature. (A) Distribution of wild-type normalized production fitness values for conserved variants (blue) and variants found only in the library (gray). (B) Total number of variants and number of conserved variants with s' > 1). *$p < 10^{-20}$ (Mann Whitney U test)

cate that the nucleotide sequence of the *rep* gene is optimized. However, we were not able to detect broad nucleotide-level effects in our data.

1.1.4 Comprehensive mutagenesis identifies beneficial variants not observed in nature

We observed that mutations to amino acids found in other naturally occurring serotypes were better tolerated than mutations to amino acids that are not found in nature (Figure 1.4A). However, a majority of the variants with production fitness values greater than that of wild-type are not observed in nature (Figure 1.4B). Only 245/2351 (6.17%) of variants with s' > 1 and only 4/115 (3.48%) of variants with s' > 1.5 are observed in AAV serotypes 1-13. These data emphasize the power of our comprehensive approach to identify novel functional sequence diversity.

1.1.5 Validation of multiplexed production assay results

To validate the results of our multiplexed production assay we selected fourteen variants, cloned them individually into the pCMV-Rep78/68 format, and determined their effect on production

by measuring DNase-resistant particle titers (Figure 1.5A). We included H92A and K340H, mutations to the origin-binding domain and helicase domain active sites, respectively, as controls[82,102]. Twelve variants with mutations in the origin-binding domain, linker domain, and helicase domain and fitness values greater than wild-type were also assayed. Finally, we included two deleterious variants identified in our multiplexed assay, N139G and A213V. The DNase-resistant particle titers determined from individual transfections are well-correlated with the fitness values determined from the multiplexed production assay (Figure 1.5B). We performed similar validation for a panel of WT AAV2 format variants (Figure A.3B).

To enable multiplexed analysis of *rep* variants in our library assay, the mutant *rep* genes themselves are packaged into the AAV capsids. Relatively small amounts of *rep* library plasmids are also used in transfection to reduce the frequency at which multiple *rep* variant plasmids enter the same cell. These procedures maintain a genotype-phenotype linkage and allow us to identify the *rep* sequences that enable production of genome-containing particles. We sought to confirm that the effects observed in the library production assay would be conserved when the *rep* variants were supplied *in trans* to the ITR plasmid, as in the traditional triple plasmid transfection method used for rAAV production. To this end, we selected a small subset of *rep* variants and cloned them into an AAV2 pRepCap plasmid without inverted terminal repeats. Here, we assayed variants with single codon mutations in each of the Rep domains.

We used these mutant pRepCap plasmids to produce rAAV vectors containing four different genomes, three single-stranded genomes and one self-complementary genome. The DNase-resistant particle titers for these pRepCap variants were relatively consistent across the different genomes (Figure 1.5C). Importantly, the DNase-resistant particle titers determined when the *rep* variants were supplied *in trans* correlated well with the normalized fitness values determined in our production assay (Figure 1.5D). Several single codon variants, including S110R, N139R, and K566L, showed 50-100% improvements in DNase-resistant particle titer over wild-type. Western blot analy-

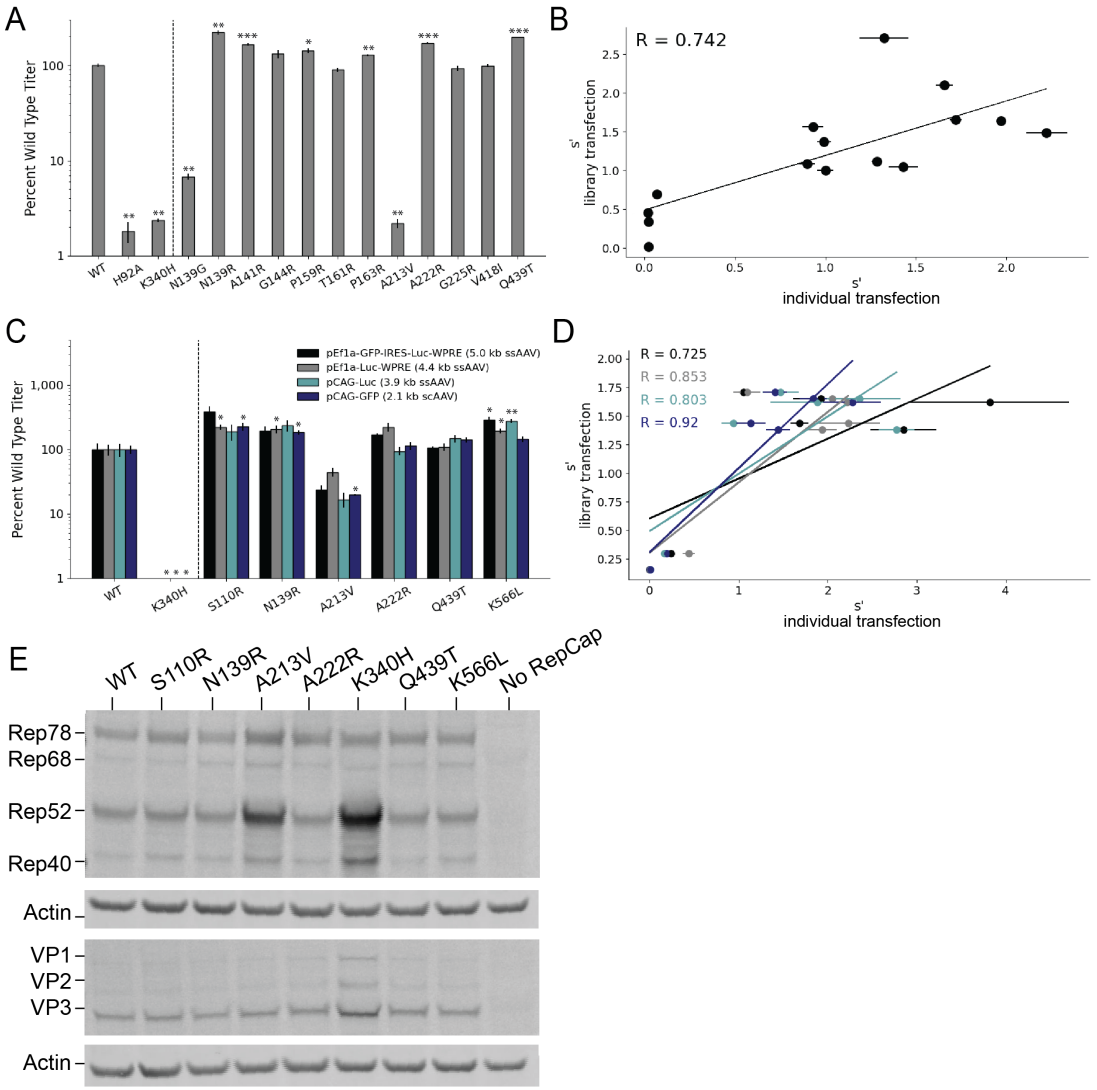


Figure 1.5: Validation of AAV2 library production assay results. (A) DNase-resistant particle titers for fourteen single amino acid pCMV-Rep78/68-inverted terminal repeat variants produced individually. Titers for previously characterized variants are plotted to the left of the dotted line. (B) Relationship between normalized production fitness values from library experiments and DNase-resistant particle titers from individual transfections. (C) DNase-resistant particle titers for four rAAV genomes produced with the indicated Rep variants. Titers for previously characterized variants are plotted to the left of the dotted line. (D) Relationship between rAAV DNase-resistant particle titers and normalized production fitness values from library experiments. (E) Expression of Rep and VP proteins from variant pRepCap plasmids by Western blot. For panels A and C, asterisks indicate significance of titer differences between the Rep variant and the relevant wild-type control. *p < 0.05, **p < 0.01, ***p < 0.001 (Welch's t-test)

sis indicated that most of these variants had little to no effect on Rep protein or VP expression (Figure 1.5E). Interestingly, in the case of the deleterious variants, A213V and K340H, Rep52/40 protein levels were increased. Codon GCG213 is located just upstream of the Rep52/40 start codon. As such, mutations to this position likely affect the strength of the p19 promoter and therefore Rep52/40 expression levels. Notably, the effects of this mutation on the p19 promoter are separate from the deleterious effects of the A213V amino acid change; the negative effects of the A213V change are observed even when Rep52/40 are expressed *in trans* (Figure 1.5A).

1.1.6 The S110R variant increases transducing particle titer

We determined the relative transduction efficiency of rAAV2 produced with wild-type Rep and two different Rep variants, S110R and Q439T. We set up individual transfections with these Rep variants and affinity purified the resulting rAAV2 vectors. Use of the S110R Rep variant resulted in approximately 2-fold higher genome-containing particle titers (Figure A.10A). We also performed capsid ELISAs and determined that the S110R variant resulted in a similar increase in total particle titers; while genome-containing particle titers were increased relative to wild-type Rep, the ratio of viral genomes: capsids was unchanged. We then used a constant volume of each rAAV2 prep to transduce HEK293T cells and quantified relative transduction efficiency by measuring transgene expression (luciferase activity, Figure A.10B). Again, an approximately 2-fold increase in relative transduction was observed with rAAV2 produced with S110R Rep as compared to wild-type Rep.

1.1.7 Mutations in the AAV2 *rep* gene have similar effects on production of AAV2, AAV5, and AAV9 capsids

Many different AAV capsid serotypes are of clinical interest for their unique tissue targeting properties. Given the putative physical interaction between Rep proteins and the capsid during genome

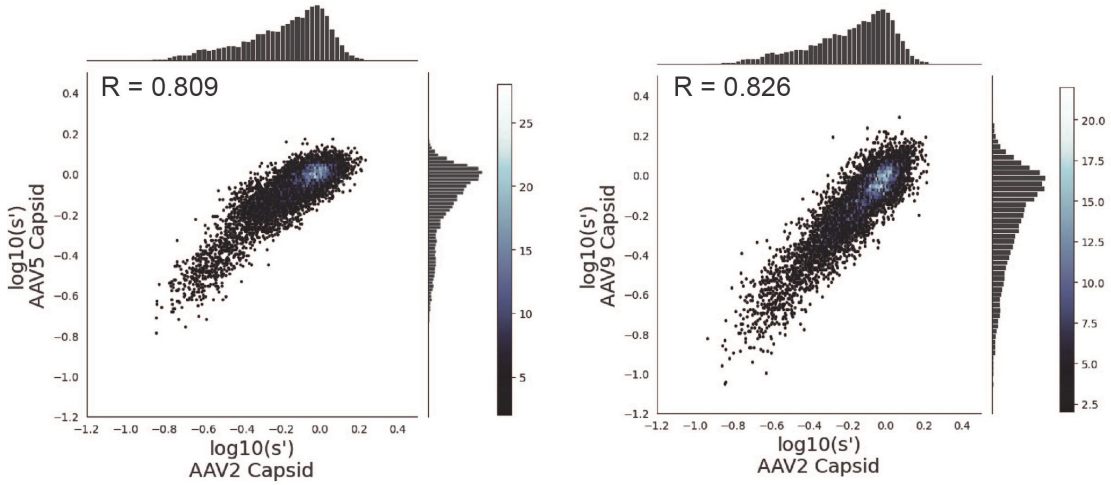


Figure 1.6: Mutations to AAV2 *rep* have similar effects on AAV2, AAV5, and AAV9 capsid production. Wild-type normalized production fitness values from the library production assay with (A) AAV5 and AAV2 capsids and (B) AAV9 and AAV2 capsids. Pearson R correlation coefficient calculated after log transformation.

packaging, we hypothesized that some *rep* mutations would have unique effects on the production of different capsid serotypes. To identify *rep* variants with serotype-specific effects on production, we repeated our pCMV-Rep78/68 library production assay using AAV5 and AAV9 *cap* genes in place of AAV2 (Figures A.11-14). The AAV5 capsid was chosen as it has low overall sequence identity with the AAV2 capsid (57.7% amino acid identity). The AAV9 capsid is more similar to the AAV2 capsid (81.7% amino acid identity) but is of particular clinical interest and differs from AAV2 at amino acid positions 329 and 330, which were previously identified as important for formation of Rep/capsid complexes (AAV2: T329/T330, AAV9: V331/K332)[8]. Surprisingly, the correlation in fitness values across capsid serotypes was comparable to the correlation between biological replicates, indicating that the effect of AAV2 *rep* variants on production is largely consistent across capsid serotypes (Figures 1.6A-B). We individually produced rAAV5 and rAAV9 vectors using a subset of our *rep* variants and confirmed that the titers with each *rep* variant are well correlated across capsid serotype (Figure A.15).

1.2 Discussion

We have generated a comprehensive mutagenesis library of the AAV2 *rep* gene, containing all possible single codon substitutions and deletion variants. Multiplexed assay of this library allowed us to generate a sequence-to-function map, linking all variants to their effect on AAV production.

Many of the previous AAV mutagenesis studies have focused on the AAV *cap* gene. Previous work in our lab generated a library containing all possible single codon substitutions, deletions, and insertions of the AAV2 *cap* gene and assayed their effect on AAV2 production[85]. Notably, we observed a smaller range of fitness values in our production assay with the *rep* library than with the previously reported *cap* library. This observation aligns with our knowledge of Rep and capsid biology. The Rep proteins' primary function is to facilitate AAV production, while the capsid has evolved to not only enable assembly and genome packaging, but also to facilitate cell targeting, entry, and nuclear trafficking. Additionally, the Rep proteins possess multiple enzymatic activities that require the proteins to adopt specific and dynamic conformations. The capsid, in contrast, is not known to possess enzymatic activity outside of the phospholipase domain located at the N-terminus of VP1 [109]. It follows that there are greater mutational constraints on *rep* than on *cap* in the context of AAV2 production.

Despite the smaller range of fitness values observed here, we identified numerous Rep variants with production fitness values greater than or equal to that of wild-type AAV2 Rep that are not observed in naturally occurring AAV serotypes. We attribute this discrepancy to the fact that optimal rAAV production may not equate to optimal fitness of AAV in the endogenous host. Wild-type AAV has both a lytic and a latent cycle[67]. The latent cycle occurs when no helper virus is present. Rep and VP expression are repressed and the AAV genome is integrated into the host cell genome in a process that also involves Rep[89,116]. In the presence of a helper virus, such as adenovirus, the lytic cycle proceeds. In this instance, Rep78/68 activate transcription from the AAV promoters

and genome replication and packaging occur[89]. rAAV production is somewhat analogous to the lytic cycle. However, naturally-occuring AAV must balance the effect of any mutations on fitness in both the lytic and latent contexts. An additional explanation for this finding is the relatively small number of AAV serotypes, for which *rep* sequences are available. In our analysis, we compared the beneficial variants identified in our library to the *rep* sequences from twelve different AAV serotypes (AAV1-13, no sequence for AAV9 *rep*). This is a small number of naturally occurring *rep* sequences when compared to the size of our library. While there have been efforts to identify additional naturally-occurring *cap* sequences, there has been relatively little effort to expand the number of naturally-occurring *rep* sequences[124]. We expect that such work may identify *rep* sequences containing many of the beneficial variants identified in our library.

We observed that substitutions in the origin-binding and zinc-finger domains were better tolerated than substitutions in the linker and helicase domains. A sharp drop in mutability was observed between residues D212 and A213. Previous work has identified residues V215 through Y224 as the minimal linker sequence required to facilitate Rep78/68 oligomerization[137]. In a separate study, it was reported that mutation of P214 reduced the ability of Rep to oligomerize[82]. Our data indicate that residue A213 is intolerant of mutation. A213 likely plays an important role in AAV production and, given its position, may be involved in Rep78/68 oligomerization. Additional work is needed to confirm that mutations to A213 affect AAV production by interfering with Rep78/68 oligomerization. In both library formats, the zinc-finger domain was relatively tolerant of mutation and few substitutions in this domain resulted in production fitness values that were significantly different from wild-type. At some positions, even introduction of a premature stop codon was not deleterious. It was recently reported that premature stop codons introduced at positions 522 and 553 of Rep did not affect AAV2 production when supplied in a pRepCap plasmid[78]. (The zinc-finger domain begins at amino acid position 536 of Rep78.) Our results provide further evidence that the zinc-finger domain is dispensable for AAV production and identify A213 as a potential linker

domain residue.

The majority of beneficial substitutions clustered in the origin-binding domain. In a recent investigation of *rep* hybrids, Mietzsch et al. reported that replacement of the entire AAV2 origin-binding domain with that of AAV1 or AAV8 improved the proportion of full capsids[78]. Our data supports the importance of the origin-binding domain for AAV production and identifies specific origin-binding domain regions where beneficial mutations cluster. Most substitutions of inverted terminal repeat-hairpin interacting residues and substitutions of residues in the Rep binding site-interacting loop to positively charged amino acids were beneficial. Previous studies have demonstrated that Rep binding to the inverted terminal repeat-hairpin improves the efficiency of terminal resolution site nicking but is not required for the nicking reaction to occur[127]. The Rep binding site-interacting loop, on the other hand, is important for recognition of double-stranded DNA during genome replication and promoter binding. Interestingly, further characterization of the S110R variant, which falls in the Rep binding site-interacting region, indicated that it increases genome-containing particle and transducing particle titer relative to wild-type Rep but did not increase the ratio of viral genome: capsids, a proxy for the proportion of full capsids. Additional work is needed to understand which, if any, of the beneficial OBD variants identified here also increase the proportion of full capsids. Taken together, our results indicate that enhancement of Rep-DNA interactions may be a fruitful avenue for further improvement of AAV production.

Inclusion of all single codon variants in our libraries allowed us to investigate the effect of synonymous mutations on AAV production. Interestingly, we observed that introduction of a c.849C>G mutation resulted in lower production fitness values compared to synonymous variants. Interestingly, this position is located downstream of the p5 and p19 promoters. Additionally, this mutation had a negative effect on production even when the *cap* gene was supplied *in trans*, making the activity of the p40 promoter irrelevant for AAV production. The c.849C>G mutation also does not fall near any known AAV2 splice sites[112]. While the mechanism by which this mutation affects produc-

tion remains unclear, our results emphasize that the *rep* gene is optimized for AAV production at both the amino acid and nucleotide levels.

While we were able to discern a trend in the effect of synonymous variants at nucleotide position 849 of *rep*, the selection values for synonymous codon variants in general were relatively closely distributed. This precluded further investigation of the effect of synonymous codons on AAV production. Synonymous variants likely have an effect on aspects of AAV production, such as genome replication, transcriptional regulation, mRNA stability, and protein expression. However, our assay measures the aggregate effect of *rep* variants on all steps in the AAV production process and is likely unable to detect the effects of synonymous variants on specific steps in this process if those steps are not rate-limiting.

In an attempt to identify Rep substitutions with a beneficial effect on the production of specific capsid serotypes, we repeated the pCMV-Rep78/68 library assay using AAV5 and AAV9 *cap* genes in place of AAV2 *cap*. Surprisingly, the correlation in fitness values across experiments was similar to the correlation between biological replicates, suggesting that the majority of AAV2 *rep* mutations have a similar effect on the production of AAV2, AAV5, and AAV9 capsids. These results suggest that either: *rep* mutations have a similar effect on Rep78/68-capsid interaction across serotypes, that the effect of perturbing Rep78/68-capsid interactions is obscured by the deleterious effects of the same mutations on other Rep activities, or that Rep78/68-capsid interactions are not a limiting factor in AAV production.

We have generated a comprehensive sequence-to-function map of the effect of all single codon AAV2 *rep* mutations on AAV2, AAV5, and AAV9 production. Our experiments identified thousands of functional Rep variants, laying the groundwork for further engineering of these proteins and enhancement of large-scale gene therapy production.

1.3 Materials and Methods

1.3.1 Library cloning

We generated *rep* variant libraries through pooled oligonucleotide (oligo) synthesis (Twist Biosciences) and subsequent Golden Gate Assembly using methods previously developed in our lab[85]. To begin, we generated two wild-type backbone plasmids, containing either the Rep78/68 open reading frame or the *rep* and *cap* open reading frames, and removed all BsaI, BbsI, EcoRV, SphI, and XbaI sites. Within the *rep* open reading frame we introduced a synonymous mutation at G339 (c.1017G>C). Within the VP open reading frame we introduced a synonymous mutation at V118 (c.354C>G) and a coding mutation F370Y (c.1109T>A). The *rep* gene was divided into eleven tiles and cloning for each tile was carried out separately. We designed 300-mer oligos to include 207 nucleotides of *rep*-coding sequence immediately followed by a BbsI site, an EcoRV site, another BbsI site, and a unique 20 nucleotide barcode sequence. All of these elements were flanked by BsaI and primer binding sites on either end of the synthesized sequence. Unique primer binding sites were used for each tile. To enable cloning of both the Rep78/68 and WT AAV2 format libraries, oligos containing the Rep52/40 start codon (M225) were synthesized with and without the M225G mutation. All possible single codon substitutions and deletions, including synonymous variants, were included in the library. All positions from the Rep78/68 start codon to the Rep78/52 stop codon (622 codons) were mutated. We did not include the eight codons at the end of the Rep68/40 open reading frame in our mutagenesis as these positions overlap with the VP1 open reading frame. Each codon variant was represented by at least two unique barcodes. For each tile, a minimum of ten uniquely barcoded wild-type controls were included. The Rep78/68 and WT AAV2 libraries each had a total of 81,116 uniquely barcoded variants.

Following synthesis, we amplified the oligos for each tile (Q5 Hot Start High-Fidelity 2X Master Mix, NEB). In parallel, we amplified the backbone vector using primers that introduced BsaI

sites. Vector PCR products were digested with BsaI-HF v2, DpnI, and recombinant shrimp alkaline phosphatase (rSAP, NEB) overnight and PCR purified (Qiagen QIAQuick PCR Purification Kit) the following day. We then performed Golden Gate Assembly with the amplified oligos and vector PCR digest products (NEBridge Golden Gate Assembly Kit, BsaI-HF v2). Golden Gate Assembly reactions were cycled 100X (16°C for 5 minutes, 37°C for minutes) and then heat inactivated. Golden Gate Assembly products were PCR purified and eluted into 25 uL of Buffer EB (Qiagen). Eluates were drop-dialyzed against 30 mL of water for 1 hour and transformed (Lucigen E. cloni 10G ELITE electrocompetent cells). Transformed cells were recovered in 1 mL Lucigen recovery media at 37°C for 1 hour and the entire volume of outgrowth was used to inoculate 50 mL LB + Kanamycin cultures, which we grew at 30°C overnight. The following day, we midi-prepped these cultures via alkaline lysis (Qiagen Plasmid Plus Midi-Prep Kit). This first cloning step enabled generation of intermediate products, which contained any wild-type *rep* sequence upstream of the mutated oligo followed by the mutant oligo. In this step, all wild-type sequences downstream of the mutant oligo were removed.

To re-introduce these missing wild-type sequences, we used a second round of Golden Gate Assembly and the internal BbsI sites present in each oligo. Before performing Golden Gate Assembly, we ran rolling circle amplification on our intermediate cloning products. 10 ng of the intermediate plasmid products were incubated with 10 uM random hexamer primers and 1X phi29 DNA polymerase buffer (NEB) at 95°C for 3 minutes and then cooled to room temperature. We then mixed these samples with rolling circle amplification solution (10 uM random hexamers, 1X phi29 DNA polymerase buffer, 5 U phi29 DNA polymerase, 1 mM dNTPs, 2 mg/mL recombinant albumin, 0.02 U inorganic pyrophosphatase) at a 1:1 ratio and incubated them at 30°C overnight. The resulting rolling circle amplification products were directly digested with BbsI-HF, DpnI, and recombinant shrimp alkaline phosphatase (all NEB) at 37°C overnight. The following day, we ran the digest products on 1% Tris-acetate EDTA gels and extracted and purified the correctly sized

products (Qiagen QIAQuick Gel Extraction Kit). In parallel, we PCR amplified the missing downstream wild-type sequences for each tile from the backbone vectors; the primers used here also added BbsI sites. We ran Golden Gate Assembly with the rolling circle amplification digest products and vector PCR products, BbsI-HF (NEB), and T4 DNA Ligase (NEB), cycling as described above. Golden Gate Assembly products were transformed and midi-prepped as above. This second cloning step resulted in plasmids containing a full-length *rep* gene with a single codon mutation followed by a twenty nucleotide barcode. WT AAV2 library plasmids also contained a wild-type copy of the AAV2 *cap* gene between the variant *rep* genes and barcodes.

To enable packaging of variant *rep* sequences into AAV capsids, the step two cloning products were moved into inverted terminal repeat-containing vectors. Step two cloning products were subject to rolling circle amplification as above and digested with XbaI, SphI-HF, and EcoRV-HF. The *rep* and/or *cap* open reading frames in the backbone vectors were flanked by XbaI and SphI sites. EcoRV sites were part of the synthesized oligos and should only be present in step one cloning products. Digest products were gel extracted and ligated into inverted terminal repeat-containing vectors. In the case of the WT AAV2 format library, the inverted terminal repeat vector contained the p5 promoter and endogenous AAV2 polyA sequence. In the case of the Rep78/68 and Rep52/40 libraries, the inverted terminal repeat vector contained a CMV promoter, WPRE, and bGH polyA sequence. The integrity of the inverted terminal repeat destination plasmids was confirmed by SmaI digest and complete plasmid sequencing (MGH DNA Core).

1.3.2 Viral library production assay

HEK293T cells were cultured in DMEM supplemented with 10% FBS and seeded in five-layer cell culture multi-flasks (Corning 353144) at 8 x 107 cells/flask two days prior to transfection. We transfected HEK293T cells using polyethylenimine. For each library, replicate transfections were performed in separate cell stacks. For the pCMV-Rep78/68 library, transfections were performed with

1 ug of pCMV-Rep78/68 library plasmids, 50 ug of pHelper, 25 ug of pCMV-AAV2cap, and 1.5 ug of p19-Rep52/40 per cell stack. For the WT AAV2 library, transfections were performed with 1 ug of WT AAV2 library plasmids and 50 ug of pHelper. Additional plasmid ratios were tested for each library; conditions that resulted in the strongest correlation in production fitness values between replicate transfections were used; these are the conditions listed above. AAV5 and AAV9 capsid production assays were performed using the pCMV-Rep78/68 library as described above.

Three days post-transfection NaCl was added to a final concentration of 0.5 M and samples were incubated at 37°C for 3 hours to detach and lyse cells. Samples were transferred to fresh 500 mL bottles and incubated at 4°C overnight. The following day any precipitate was removed by pipetting and the remaining volume was passed through a sterile 0.22 um PES filter. PEG-8000 was added to a final concentration of 8% and samples were again incubated at 4°C overnight. The following day, we centrifuged samples at 3000 x g for 20 minutes to pellet the PEG-precipitated virus. Supernatants were discarded and pellets resuspended in 8 mL of DPBS. We then digested samples with a 1:10,000 dilution benzonase (Millipore Sigma 1.101695.0001) at 37°C for 45 minutes and subjected them to iodixanol gradient ultracentrifugation as previously described[85,140]. Briefly, benzonase-treated samples were underlaid with an iodixanol gradient (Millipore Sigma D1556) in polypropylene tubes (Beckman Coulter, 362183) and centrifuged at 242,000 x g for 1 hour at 16°C in a VTi50 rotor. Following ultracentrifugation, 500 uL fractions were collected from the 40% iodixanol layer and buffer-exchanged into DPBS using 100 kDa molecular weight cutoff centrifugal filter units (Millipore Sigma UFC910024).

Rep sequences in the purified pool represent variants capable of producing genome-containing viral particles. Barcodes from this viral pool, along with barcodes from the plasmid libraries, were amplified using flanking primer sequences. Illumina sequencing adapters and indices were added in a subsequent PCR. The resulting PCR amplicons were pooled and sequenced using the Illumina NextSeq platform (Biopolymers Facility, HMS). Barcode sequences were extracted from the result-

ing sequencing reads and barcode counts (c_v) from replicate transfections were summed. We calculated the frequency of each variant (f_v) in the viral library as: $f_v = c/\sum f_v$ and then determined the production fitness for each variant as: $s = f_{viral}/f_{plasmid}$. Finally, production fitness was normalized to that of the wild-type controls: $s' = s/s_{WT}$.

1.3.3 Determination of DNase-resistant particle titers by qPCR

To validate the results of our multiplexed library production assay, fourteen single codon *rep* variants were selected and individually transfected into HEK293T cells. Transfections were performed in triplicate. We seeded cells at 4 x 10^5 cells/well in 6-well plates 24 hours prior to transfection. Small amounts of pCMV-Rep78/68 format variants, 50 pg per transfection, were used to recapitulate the low plasmid levels used in the library production assay. pCMV-Rep78/68 variants were transfected via PEI together with pHelper (2 ug), pCMV-AAV2cap (1 ug), and p19-Rep52/40 (2 ng) plasmids as in the library assay. Three days post-transfection, we lysed cells by 3X freeze-thaw. Samples were then clarified by centrifugation at 15,000 x *g* for 5 minutes. 5 uL of supernatant were subjected to DNase digest (ThermoFisher, EN0521) at 37°C for 30 minutes followed by heat inactivation at 65°C for 10 minutes. We then incubated samples at 98°C for 10 minutes to denature the capsids and measured DNase-resistant particle titers with qPCR (IDT PrimeTime Gene Expression Master Mix). Primers and a probe binding to the 5' region of the *rep* gene were used. The sequence of the forward primer was: 5'-GAGCATCTGCCCGGCATTTC-3', the sequence of the reverse primer was: 5'-ATCTGGCGGCAACTCCCATT-3', and the sequence of the probe was 5'-HEX-ACAGCTTTG-ZEN-TGAACTGGGTGGCCGA-3IABkFQ-3' (IDT).

Transfections with *rep* variants cloned into pRepCap plasmids were performed as described above, using the following amounts of plasmid for each transfection: 1 ug of pRepCap, 1 ug of inverted terminal repeat plasmid, and 2 ug of pHelper. The pEf1a-Luc-WPRE (ssAAV, Addgene plasmid #87951; http://n2t.net/addgene:87951; RRID:Addgene_87951), pCAG-Luc (ssAAV,

Addgene plasmid #83281; http://n2t.net/addgene:83281; RRID:Addgene_83281), and pCAG-GFP (ssAAV, Addgene plasmid #83279; http://n2t.net/addgene:83279; RRID: Addgene_83279) Inverted terminal repeat plasmids were a gift from Mark Kay[87,88]. We used qPCR to determine plasmid concentration for transfection. The *rep* primer and probe sequences mentioned above were used to quantify pRepCap plasmids. To determine the concentration of GFP-expressing plasmids we used a forward primer with the sequence 5'-GAACCGCATCGAGCTGAA-3', a reverse primer with the sequence 5'-TGCTTGTCGGCCATGATATAG-3', and a probe with the sequence 5'-FAM-ATCGACTTC-ZEN-AAGGAGGACGGCAAC-3IABKFQ-3' (IDT). To determine the concentration of luciferase-expressing plasmids we used a forward primer with the sequence 5'-CATGTACCGCTTCGAGGAG-3', a reverse primer with the sequence 5'-GAAGCTAAATAGTGTGGGCACC3', and a probe with the sequence 5'-FAM-CTTGCGCAG-ZEN-CTTGCAAGACTATAAGATTC-3IABKFQ-3' (IDT). The same GFP- and luciferase-specific primers and probes were used to quantify the resulting rAAV vectors.

1.3.4 Western blot analysis of VP and Rep protein levels

Following transfection, media was aspirated from cell culture plates and cells were washed with DPBS. We then lysed cells with 50 uL of lysis buffer per well (6-well plates). The lysis buffer contained: 10 mM Tris-HCl (pH 8), 150 mM NaCl, 1% Triton X-100, and cOmplete mini protease inhibitor (1 tablet/10 mL lysis buffer). Lysates were transferred to fresh 1.5 mL tubes and centrifuged at 15,000 x g for 5 minutes to clarify. Total protein concentrations were determined by Bradford assay (Thermo Scientific Pierce Coomassie Bradford Protein Assay Kit). We loaded 75 ug of total protein from each sample on 4-12% Bis-Tris gels (ThermoFisher) and ran the gels for 45 minutes at 140V. We then transferred the proteins to PVDF membranes (ThermoFisher). Membranes were blocked with 5% milk in PBST for 1 hour at 4°C. Blots were then incubated with a 1:250 dilution

of B1 anti-VP antibody or a 1:100 dilution of anti-Rep 303.9 antibody (both American Research Products) in blocking buffer overnight. Both primary antibody mixtures also included a 1:20,000 dilution of anti-□-actin antibody (Abcam, ab8227). The following day, blots were washed 3X with PBST and incubated with secondary antibodies for 1 hour in blocking buffer with 0.01% SDS. Goat anti-mouse IRDye 800CW (LiCor 925-32210) and donkey anti-rabbit IRDye 680RD (LiCor 925-68073) secondary antibodies were used. Blots were washed 3X with PBST and 1X with PBS, and the near-infrared fluorescence of the secondary antibodies was visualized using the Sapphire imager.

Code and data availability

Code and data for this paper can be accessed here: https://github.com/churchlab/aav_rep_scan.git. Illumina sequencing reads were uploaded to NCBI GEO (series accession number GSE226265).

Science, my lad, is made up of mistakes, but they are mistakes which it is useful to make, because they lead little by little to the truth.

Jules Verne

2

The effect of transgene sequence on AAV2 production

Pierce J. Ogden and George M. Church also contributed to this work. N.K.J., P.J.O., and G.M.C. conceived the study and designed experiments. N.K.J. performed experiments and data analysis with input from all authors.

rAAV vectors consist of two main components, a capsid shell and a DNA genome. Much effort has been dedicated to engineering both rAAV capsids and genomes to enhance transduction of specific tissues and cell types [1,46,85,117,74,12,37,19,27]. There have also been efforts to understand the relationship between capsid sequence and production of genome-containing and transducing particles [9,91,128,44]. However, there has been comparably little study of the effect of rAAV transgene sequence on production. rAAV production involves genome replication and packaging, both processes that could theoretically be affected by rAAV genome sequence. Likewise, transduction necessitates release of encapsidated genomes, synthesis of double-stranded DNA, and transcrip-

tion of transgenes. By affecting these processes, rAAV genome identity could impact not only genome-containing particle titer but also transducing particle titer. Interestingly, multiple groups have reported a marked difference in both genome-containing and transducing particle titer when comparing wild-type AAV2 and rAAV2 vectors produced with the same capsid[45,138]. Additionally, the numbers of genome-containing and transducing particles have been reported to vary between rAAV2 vectors containing different transgenes by an order of magnitude or more[45]. Taken together, these data suggest that encapsidated sequence affects rAAV2 production. Here, we further probe the relationship between encapsidated, or transgene, sequence and rAAV production, building towards the identification of generalizable rules for rAAV transgene design.

2.1 Results

2.1.1 Wild-type AAV2 titers are 2-to-10-fold higher than rAAV2 titers

We performed transfections with a wild-type-AAV2-ITR plasmid as well as three rAAV2-ITR plasmids, which each contained a different transgene. All rAAV2-ITR plasmids also included non-coding regions to bring the length of each rAAV2 genome to that of wild-type AAV2 (4.7 kilobases). Three days post-transfection, viral particles were purified and both genome-containing particle and capsid titers were determined. Wild-type AAV2 genome-containing particle and capsid titers were 2- to 10-fold higher than rAAV2 titers (Figure 2.1). These data confirm previous reports that wild-type AAV2 has a production advantage over rAAV2, although the difference is smaller than previously reported.

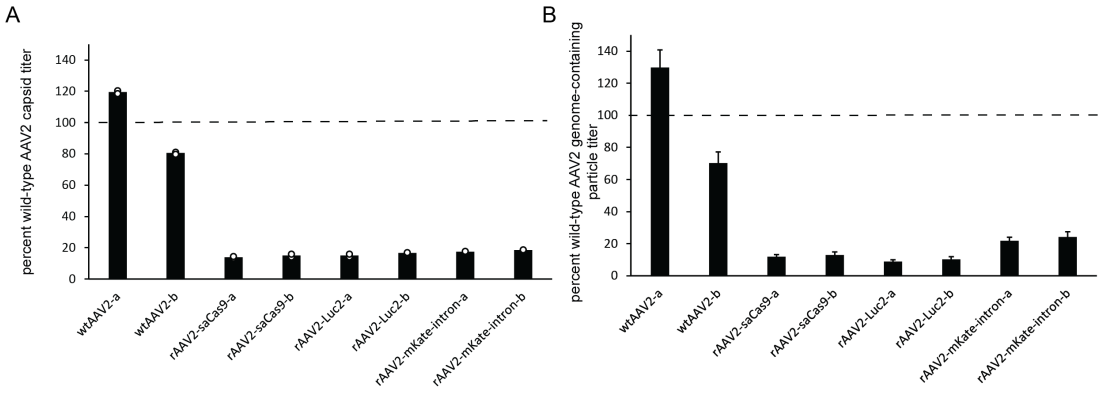


Figure 2.1: Comparison of wild-type AAV2 and rAAV2 titers following iodixanol purification. (A) Capsid titers and (B) genome-containing titers. For each construct two replicate transfections and purifications were performed.

2.1.2 Wild-type AAV2 transduction is not universally higher than rAAV2 transduction

Previous *in vitro* experiments demonstrated that capsids containing genomes shorter than the wild-type genome uncoated less efficiently [54]. Additionally, differences in the ratio of genome-containing particles: transducing particles between wild-type and rAAV2 have been reported [45,138]. These data suggest that wild-type AAV may have not only a production advantage but also a transduction advantage over rAAV. As such, we sought to directly compare wild-type and rAAV2 transduction in cell culture. We generated three rAAV2 vectors that each contained a different transgene (*Staphylococcus aureus* Cas9 [SaCas9], luciferase [Luc2], or mKate) under the control of the AAV2 p5 promoter. This enabled us to compare transgene expression to wild-type AAV2 *rep* expression without having to account for differences in promoter strength. We transduced HeLaRC32 cells with adenovirus and either wild-type AAV2 or rAAV2 and measured p5 transcript levels 24 hours post-transduction (Figure 2.2A). We used HeLaRC32 cells here as they have two copies of the AAV2 *rep* and *cap* genes stably integrated into their genomes [119,18]. *Rep* expression is required for activation

of the AAV p5 promoter but only the wild-type AAV2 particles contain the *rep* gene, so we supplied *rep in trans* to the rAAV2 particles *via* the HeLaRC32 genome. Expression of the *rep* and *cap* sequences within the HeLaRC32 genome is stimulated by adenovirus infection, so we performed AAV transduction together with adenovirus infection. We spiked empty capsids into each AAV preparation to normalize both genome-containing and capsid particle titers across preps. Importantly, we also synonymously recoded a small segment of the *rep* gene in the wild-type AAV2 particles used here to enable discrimination of transduced genomes from HeLaRC32-originating AAV genomes. As indicated in Figure 2.2B, p5 transcripts from wild-type AAV2 were present at higher levels than p5-SaCas9 and p5-Luc2 transcripts but at comparable levels to p5-mKate transcripts. Wild-type AAV2 did not have a universal transduction advantage over rAAV2.

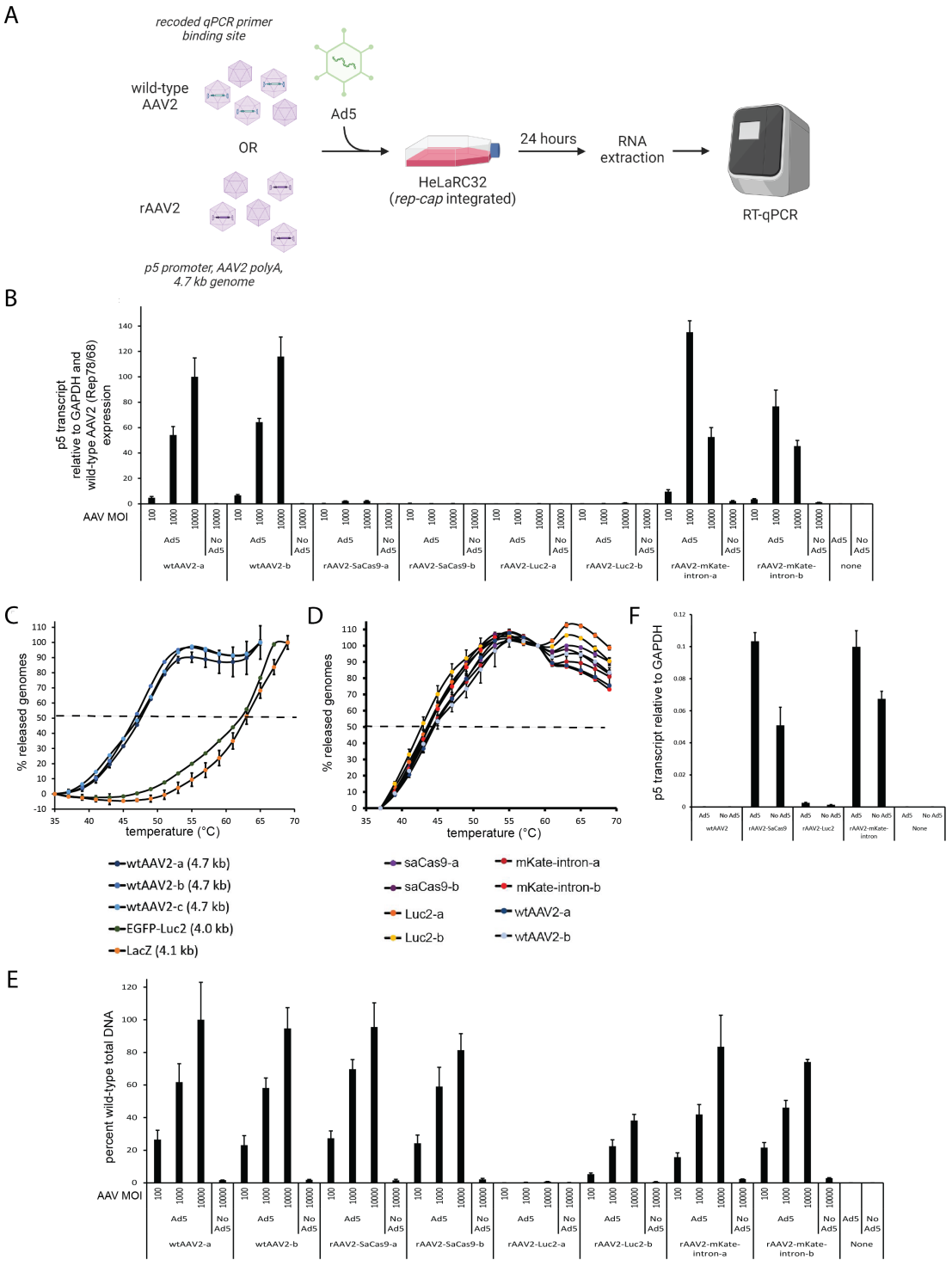
Differences in p5 transcript abundance between rAAV2 carrying different transgenes were also observed. This could be attributed to differences in genome release, genome replication, or transgene transcription between rAAV vectors. Since sequence length was previously reported to affect *in vitro* genome release, we first wanted to determine if rAAV2 genome identity - absent differences in length - also affected *in vitro* genome release. While we observed differences in *in vitro* genome release between rAAV2 vectors containing genomes of different length, no differences in release were observed between rAAV2 vectors containing different genomes of the same length (Figures 2.2C-D). These results suggest that differences in rAAV2 transduction are not driven by differences in genome release. However, the mechanism of *in vivo* genome release is not fully understood. As such, differences in *in vitro* genome release may or may not correspond to differences in *in vivo* release. While this assay represents a useful method for characterizing the biophysical properties of rAAV2 vectors, direct measurement of genome release *in vivo* is needed to determine definitely whether the uncoating process is affected by genome identity.

Next, we wanted to determine if the differences in rAAV2 p5 transcript abundance following transduction were driven by differences in genome replication. Notably, genome replication is not part of the usual rAAV transduction process, but occurs here because of adenovirus co-infection and subsequent *rep* and *cap* expression. To assess the effect of rAAV2 genome sequence on genome replication, we transduced HeLaRC32 cells as above, extracted total intracellular DNA 24 hours post-transduction, and quantified AAV genome abundance by qPCR (Figure 2.2E). For the majority of constructs tested, total DNA levels were consistent. Notably, little to no rAAV2-Luc2 DNA was detected from one of the two replicate preps, which likely reflects technical variability. These data indicate that differences in p5 transcript abundance post-transduction are likely not driven by differences in rAAV2 genome replication. Given that genome replication occurs subsequent to genome release, or uncoating, these results also provide further evidence that rAAV2 genome sequence does not affect genome release.

The differences in p5 transcript levels following rAAV2 transduction could be inherent to the transgene sequences themselves and independent of AAV transduction. To test this, we directly transfected HeLaRC32 cells with the ITR plasmids used for AAV production and measured p5 transcript abundance 24 hours post-transfection (Figure 2.2F). As in the transduction experiment, mKate expression was high and luciferase expression was low relative to the other constructs. However, in contrast to the transduction experiment, Rep78/68 (wild-type AAV2) expression was low and SaCas9 expression was high. These results indicate that the differences in p5 transcript levels following AAV transduction are likely not driven by inherent differences in the mRNA stability of the transgene sequences and that the differences in mRNA abundance that we observed are specific to AAV transduction. Work from the Kay lab recently demonstrated that the capsid plays a role in regulating the epigenetic status and subsequent transcription of transgenes, which may - in part, explain the differences in relative expression from plasmids and rAAV2 vectors observed here[41].

Figure 2.2 *(following page)*: Wild-type versus rAAV2 transduction. (A) Experimental workflow for transduction assay. (B) Expression of p5 transcripts from wild-type AAV2 is not universally higher than expression of p5 transcripts from rAAV2 vectors. (C) Comparison of *in vitro* genome release from wild-type AAV2 and two rAAV2 vectors, containing genomes shorter than wild-type AAV2. (D) Comparison of genome release from wild-type AAV2 and three rAAV2 vectors. All AAV tested in this experiment contain 4.7 kb single-stranded genomes. Letters a, b, and c denote separate AAV preparations. (E) p5 transcript levels following transduction are not correlated with total intracellular DNA levels. Transductions were performed as in (A). Total intracellular DNA was extracted 24 hours post-transduction and AAV genome abundance was quantified by qPCR using primers against the p5 promoter. (F) p5 transcript levels following transduction are not correlated with p5 transcript levels following plasmid transfection. The ITR plasmids used to produce the wild-type and rAAV2 used in (B) were directly transfected into HeLaRC32 cells. p5 transcript levels were quantified as in (B).

Figure 2.2: (continued)



2.1.3 Co-transfection of wild-type AAV2 and rAAV2 enhances rAAV2 genome-containing particle titer

While wild-type AAV2 did not result in universally higher transduction than rAAV2 vectors, wild-type AAV2 yielded higher genome-containing particle titers than rAAV2. To further probe the cause of wild-type and rAAV2 genome-containing particle titer differences, we wanted to determine if co-production of wild-type and rAAV2 influenced the titer of either construct, i.e. if wild-type AAV2 maintained a production advantage when produced in the same cells as rAAV2. To this end we set up co-transfections of wild-type AAV2-ITR and rAAV2-ITR plasmids and measured titers using orthogonal qPCR probes (Figure 2.3A). Here, we tested four different rAAV2 genomes. Previous work from our lab demonstrated that production differences between AAV genomes are most easily observed when small amounts of ITR plasmid are used for transfection[85]. As such, we used relatively small amounts of ITR plasmid (5×10^9 plasmid molecules or approximately 30 ng of plasmid/well of 12-well plate). In the case of rAAV2 production, *rep* and *cap* genes must be supplied *in trans* to the ITR plasmid to enable particle production. Wild-type AAV2 production was also performed both with and without *rep* and *cap* supplied *in trans via* a pRepCap plasmid.

Wild-type AAV2 was produced as expected, and addition of pRepCap plasmid to wild-type transfections had little effect on wild-type AAV2 titer (Figure 2.3B). Transfection of rAAV2-ITR and pRepCap plasmids, however, resulted in little to no virus production, likely because of the relatively small amounts of plasmid used for transfection. Interestingly, co-production of wild-type and rAAV2 enhanced rAAV2 titer. In the co-transfection conditions, wild-type and rAAV titers were comparable. These data indicate that the wild-type AAV2-ITR plasmids facilitated virus production when pRepCap plasmids did not and are similar to results reported by Song et al.[104]. These results can be explained in multiple ways. Firstly, it is possible that wild-type AAV2-ITR and pRepCap plasmids result in different Rep and VP expression levels and/or kinetics. The amount or timing

of AAV protein expression off of the wild-type AAV2-ITR plasmid may be more optimal for AAV production. Secondly, AAV2 genome replication is localized to specific foci within the nucleus [125]. While the pRepCap plasmid provides all the components necessary for rAAV2 production, they likely do not co-localize with replicating, ITR-containing genomes. *Cis*-acting sequences within the replicating wild-type AAV2 genome may interact with host cell factors and recruit them to centers of AAV genome replication in ways that the pRepCap plasmid cannot.

2.1.4 Changes in wild-type AAV2 titer on synonymous recoding are correlated with changes in protein expression

To probe for potential *cis*-acting sequences in the wild-type AAV2 genome that affect genome-containing particle production, we synonymously recoded segments of the genome. We started by dividing the genome into three segments, the *rep* gene, the 5' region of the *cap* gene containing the overlapping assembly-activating protein (AAP) and membrane-associated accessory protein (MAAP) reading frames, and the 3' end of the *cap* gene containing only the VP3 coding region. For each of these regions we generated at least four recoded sequences, two that were codon shuffled and two that were "maximally" recoded. Briefly, the codon shuffling method preserved codon abundance across the recoded segment while maximal recoding forced a codon change wherever possible. In both cases, we avoided introducing changes into annotated promoters and splice sites. Our aim was to introduce as many nucleotide changes as possible while preserving protein expression so that we could look specifically at the effect of DNA sequence changes on wild-type AAV2 production. For the 5' region of the *cap* gene containing the overlapping AAP and MAAP reading frames, we did not preserve the amino acid sequence of AAP or MAAP and supplied these coding sequences *in trans*.

We transfected synonymously recoded variants into HEK293T cells and determined DNase-resistant particle titers three days post-transfection. We ran Western blots against both the Rep pro-

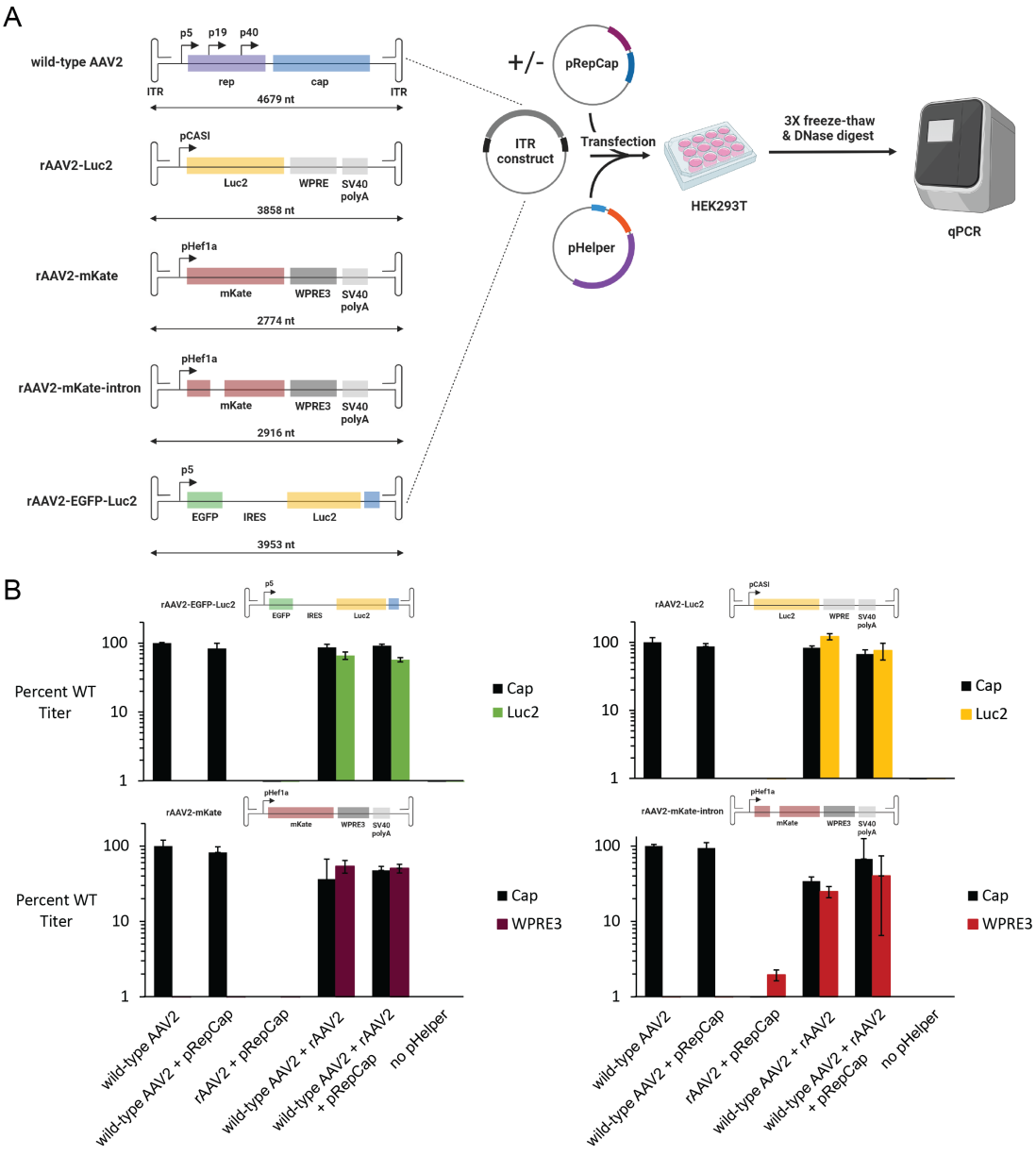


Figure 2.3: Co-production of wild-type and rAAV2. Co-transfection of wild-type and rAAV2-encoding plasmids enhances rAAV2 titer. (A) Schematic of experimental workflow. (B) DNase-resistant particle titers determined by qPCR. Data is the average of three separate transfections and is reported as percentage of the wild-type AAV2 titer from individual transfections without pRepCap.

teins and VPs to determine if protein expression was perturbed. As shown in Figure 2.4A, codon shuffling of the *rep* gene and all recodings of the VP3 coding region preserved Rep and VP expression. For all of these constructs, DNase-resistant particle titer was equivalent to, or even slightly higher than, wild-type, indicating that no *cis*-acting sequnces affecting production are present in these regions. Reductions in titer were only observed when protein expression was also perturbed, as in the case of the *rep* maximally recoded variants and all 5' *cap* recoded variants. Complementation of 5'-*cap* recoded variants with AAP and MAAP *in trans* did not improve VP expression or titer to wild-type levels. Since reductions in titer were only observed when protein expression was also perturbed, we were unable to identify any *cis*-acting signals within the wild-type AAV2 genome that facilitate genome-containing particle production. However, given that the *rep* gene and 3' end of the *cap* gene could be recoded without affecting production, it is unlikely that these regions contain any sort of *cis*-acting nucleotide sequences that affect production.

2.1.5 Synonymous recoding of SaCas9 transgene affects both AAV2 titer and protein expression

As an orthogonal approach to understand the AAV genome sequence-production relationship and motivated by the observation that different rAAV2 transgenes yielded different titers, we synonymously recoded a transgene and measured the effect of these changes on rAAV2 production. We generated two maximally recoded SaCas9 variants and first transfected them into HEK293T cells using small amounts of rAAV2-SaCas9-ITR plasmids. Under these conditions, approximately 10-fold increases in genome-containing particle titer were observed with the maximally recoded rAAV2-SaCas9-ITR constructs (Figure 2.5A). Notably, however, maximal synonymous recoding reduced SaCas9 expression in the producer cells (Figure 2.5B). These results suggest that synonymous recoding can be used to improve rAAV2 production, but there may be a tradeoff between enhanced rAAV2 production and protein expression.

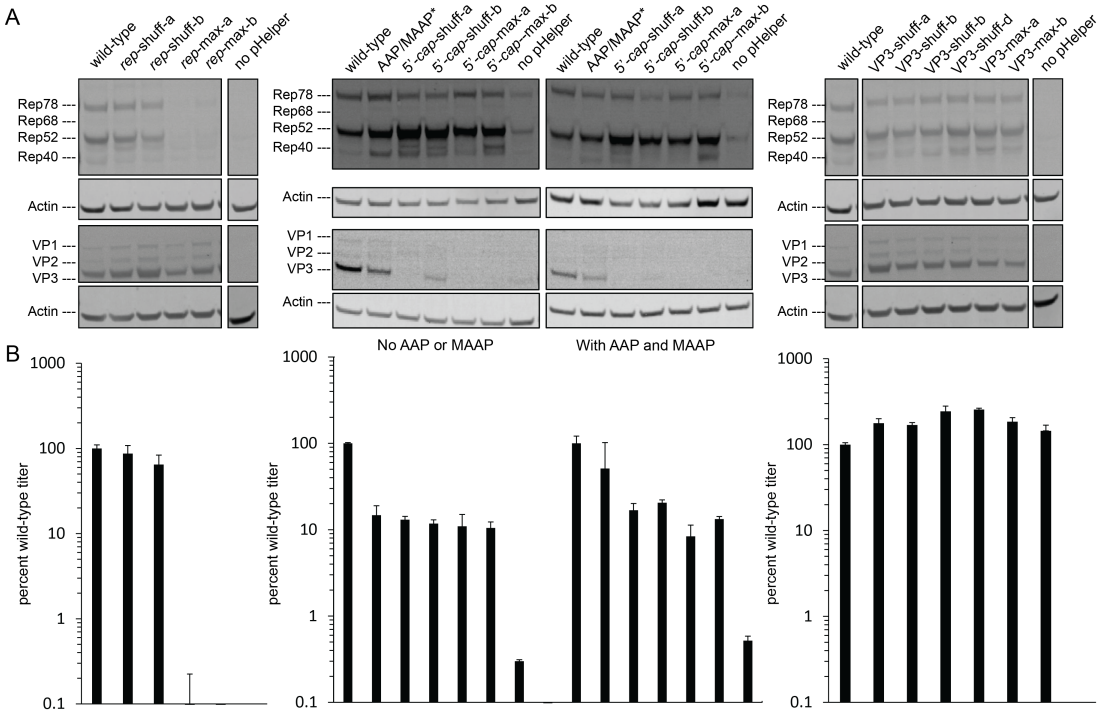


Figure 2.4: Synonymous recoding affects DNase-resistant particle titer through changes in protein expression. (A) Effect of codon shuffling and maximal synonymous recoding of *rep* (left), 5' region of *cap* (middle), and VP3-coding region of *cap* (right) on Rep and VP expression. (B) Effect of codon shuffling and maximal recoding of *rep* (left), 5' region of *cap* (middle), and VP3-coding region of *cap* (right) on DNase-resistant particle titers. AAP/MAAP*: control construct with wild-type nucleotide sequence but with premature stop codon in AAP reading frame and mutations of MAAP start codon. a, b, c, and d refer to unique recoded sequences.

Previous work has shown that protein overexpression can reduce rAAV2 production[47]. To exclude the possibility that the increase in rAAV2 genome-containing particle titer we observed was due solely to reductions in SaCas9 expression, we measured the effect of SaCas9 overexpression on the production of rAAV2 with unrelated genomes (Figure 2.5C). We produced rAAV2 containing either a luciferase or mKate transgene in the presence of: a pCMV-SaCas9 plasmid, a pCMV-LacZ plasmid, or a SaCas9 plasmid lacking a promoter. The titers of both rAAV2 vectors were similar under all three conditions, indicating that overexpression of this SaCas9 construct does not reduce rAAV2 titer and was not responsible for the increases in rAAV2 titer observed with the maximally recoded rAAV2-SaCas9-ITR constructs. When comparing wild-type to rAAV2 titers, we observe smaller differences when more ITR plasmid is used for transfection. As such, we repeated the above rAAV2-SaCas9 production experiment using larger amounts of rAAV2-SaCas9-ITR plasmid and purifying the resulting viral particles. Under these more typical triple plasmid transfection conditions, we observed approximately 2-fold differences in genome-containing particle titer between wild-type and maximally recoded SaCas9 sequences.

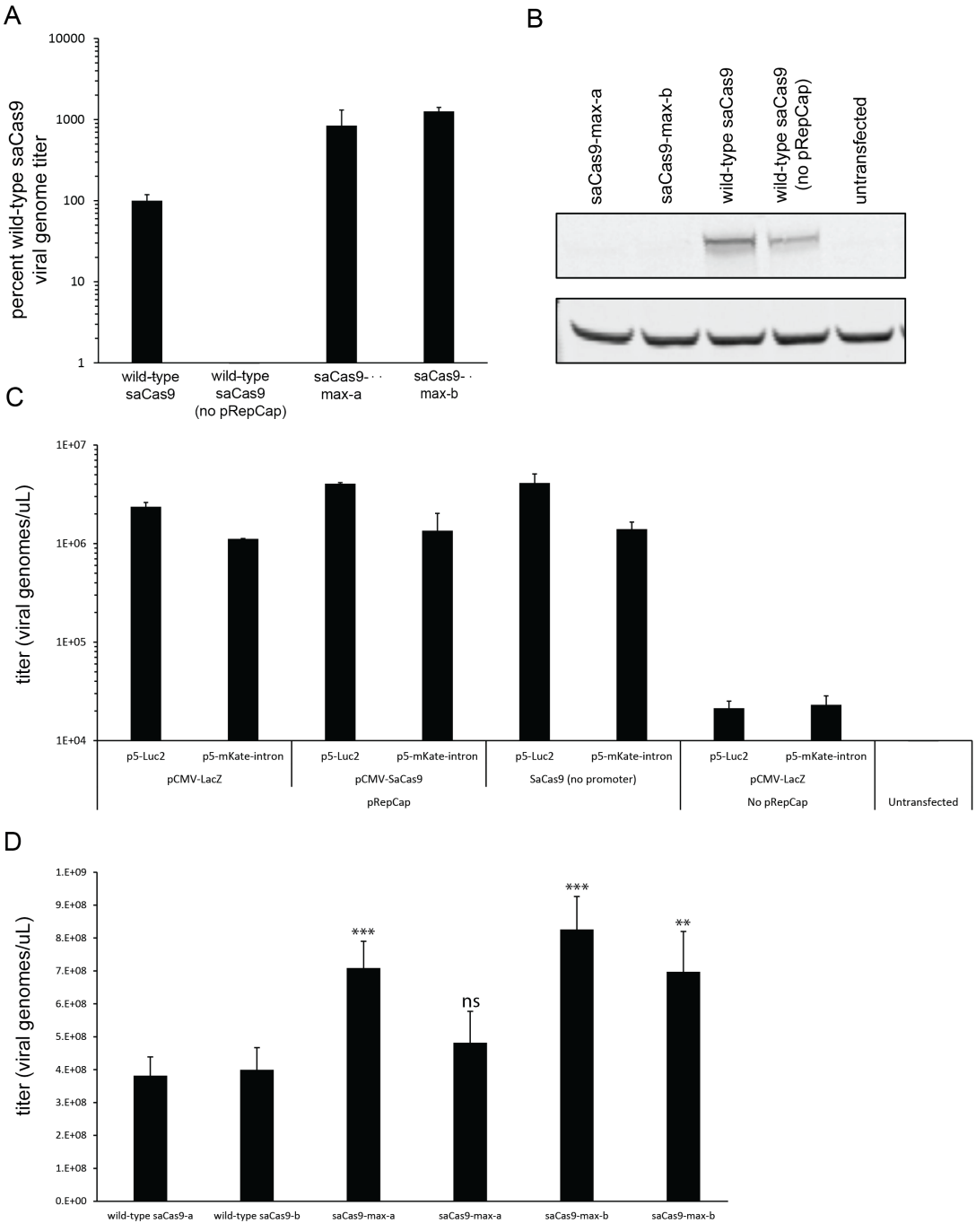
2.2 Discussion

Since the discovery of AAV over fifty years ago, much effort has been dedicated to engineering AAV capsids to enhance tissue-specific transduction. As this work comes to fruition and more rAAV therapies gain approval, there is an increasing demand to improve rAAV manufacturing processes to meet patient needs. Here, we focus on the effect of encapsidated sequence on AAV production. Our goal is to identify methods of enhancing rAAV2 production without requiring modification to the capsid, enabling use of the capsid best-suited for transduction of the target tissue. Our focus on rAAV transgene sequences, or encapsidated sequence, is motivated by previous reports that there are approximately 100-fold differences in titer between wild-type and rAAV2 produced with the same capsid[45,138]. First, we confirmed that wild-type AAV2 indeed has a production advantage over rAAV2, ranging from 2-to-10-fold. This is in line with a recent report from Song et al. that observed wild-type AAV2 titers that were 2.5-to-10.7-fold higher than rAAV2 titers[104]. Notably, while we used rAAV2 constructs containing the p5 promoter and the same length as wild-type AAV2, Song et al. investigated rAAV2 constructs with heterologous promoters and of different lengths. Taken together, these results indicate that wild-type AAV2 does have a production advantage over rAAV2, but the difference is smaller than initially reported. Additionally, titer differences are seemingly independent of genome length or promoter identity.

Previous work indicated that wild-type AAV2 production not only resulted in higher genome-containing particle titer but also in higher transducing particle titers[138]. Additionally, it has been shown that genome length affects genome release *in vitro*[54]. We measured the relative transduction efficiency of wild-type and rAAV2 cells, accounting for promoter strength and *rep* and *cap* expression. While differences in mRNA levels between rAAV2 constructs were observed, wild-type AAV2 transduction resulted in comparable expression to a rAAV2-mKate construct. These results indicate that wild-type AAV2 does not have universally higher transduction efficiency than rAAV2. We did,

Figure 2.5 ***(following page)*****:** Maximal synonymous recoding of a SaCas9 transgene affects rAAV titer and SaCas9 expression. (A) DNase-resistant particle titers for wild-type and two maximally recoded rAAV2-SaCas9 constructs. Viral production was performed using small amounts of ITR plasmids. (B) SaCas9 expression in HEK293T rAAV2 producer cells. (C) SaCas9 overexpression does not reduce rAAV2 titer. rAAV2 containing either a p5-Luc2 or a p5-mKate-intron genome were produced in the presence of a pCMV-LacZ plasmid, a pCMV-SaCas9 plasmid, or an SaCas9 plasmid lacking a promoter. (D) Genome-containing particle titers for wild-type and two maximally recoded rAAV2-SaCas9 constructs. Each rAAV2 construct was transfected and purified in duplicate, using larger amounts of ITR plasmid than in (A). Asterisks indicate statistics for comparison of wild-type SaCas9-b titer and the titer of the relevant recoded variant (Welch's t-test). **p < 0.01, ***p<0.001

Figure 2.5: (continued)



however, observe differences in mRNA levels across different rAAV2 vectors, which all used the p5 promoter. Further experiments demonstrated that the differences in p5 transcript abundance were likely not due to differences in genome release or genome replication. These differences may be caused by differences in transcription of rAAV2 transgenes, but more work is needed to pinpoint the exact cause of differences in rAAV2 transcript abundance.

Interestingly, relative p5 transcript levels differed when transgenes were supplied on plasmids and when transgenes were supplied in rAAV2 vectors. Other groups have recently identified a role for the AAV capsid in regulating the epigenetic status and subsequent transcription of transgenes[41]. It has also been demonstrated that the effect of capsid mutations on transgene expression can be promoter-specific[10]. Our data indicate that the transgene sequences, not just the promoter sequence, may also play a role in capsid-mediated transcription regulation. Future work may uncover the mechanisms of transgene-specific differences in mRNA abundance and, hopefully, identify generalizable rules for rAAV transgene sequence design.

While wild-type AAV2 does not have enhanced transduction relative to rAAV2 vectors, wild-type AAV2 does have a consistent production advantage over rAAV2 vectors. Interestingly, we and others also report that co-production of wild-type and rAAV2 results in an equalization of wild-type and rAAV2 titers[104]. While additional work is needed to determine the exact mechanism of this titer enhancement, the observation that wild-type and rAAV2 co-production enhanced rAAV2 titer indicates that sequence-based enhancement of rAAV2 production is possible. There are a few possible explanations for the observed rAAV2 titer improvements. Firstly, it is possible that expression of *rep* and *cap* from the ITR plasmid enables optimal timing of protein expression. Secondly, *cis*-acting signals within the wild-type AAV2 genome may recruit host cell factors to the site of genome replication and packaging. ITR mediated recombination of wild-type and rAAV2 genomes could also enable these sequences to act *in cis*.

Previous studies suggested that the ITRs, in enabling genome replication and through their

enhancer activity, regulate *rep* and *cap* expression to facilitate particle formation [15,38]. Wild-type AAV2-ITR plasmids may enhance rAAV2 titer by enabling optimal timing and levels of Rep and VP expression. Notably, co-transfection of wild-type AAV2-ITR and rAAV2-ITR plasmids enhanced rAAV titer both with and without pRepCap plasmid present. Even with a proportion of Rep and VP protein being expressed off a plasmid lacking ITRs, wild-type AAV2-ITR co-transfection improved rAAV2 titer. This suggests that the timing of Rep and VP expression may not be the sole or primary driver of rAAV2 titer enhancements, but direct measurement of Rep and VP levels under different conditions is needed to confirm this.

A second explanation for rAAV2 titer increases on wild-type AAV2 co-production is that the wild-type AAV2 genome contains *cis*-acting sequences that facilitate genome replication and/or packaging. We attempted to identify *cis*-acting sequences important for AAV production by synonymously recoding the wild-type AAV2 genome. Our experiments do not provide evidence for *cis*-acting sequences in the wild-type AAV2 genome. However, we cannot rule out the existence of *cis*-acting signals in the 3' region of the *cap* gene, which we were not able to recode without perturbing protein expression, or in the promoter sequences within the *rep* gene, which we did not recode. The observation that wild-type AAV2-ITR plasmids increase rAAV2 titer *in trans* raises the possibility that rAAV2 titer enhancement may be driven by wild-type AAV2-mediated co-localization of host cell proteins to sites of particle production. It is possible that host cell proteins interact with sequences within the wild-type AAV2 genome and facilitate some aspect of genome-containing particle production. AAV2 genome replication is localized to the nucleoplasm while genome packaging appears to be localized to the nucleolus [125]. The recruitment of host cell factors to the site of AAV replication and packaging could explain the rAAV2 titer improvements observed on co-transfection. Conversely, the ITRs are known to facilitate intermolecular recombination of AAV genomes [84,133]. Another possible explanation for the enhancement of rAAV2 titer by co-production with wild-type AAV2 is that ITR-mediated recombination results in the inclusion of portions of the wild-type

AAV2 genome in rAAV2 constructs and these wild-type AAV2-derived sequences enhance the replication and/or encapsidation of rAAV2 sequences *in cis*. Long-read sequencing of AAV genomes isolated from co-transfected cells and the resulting viral particles could be used to determine if recombination is occurring between wild-type AAV2 and rAAV2 plasmids and if certain regions of the genome are over represented in viral particles as compared to the co-transfected cells. Other groups are working to identify host cell proteins that interact with the ITRs during production[71]. It will be interesting to expand these techniques to probe host cell protein interactions along the full length of the wild-type AAV2 genome.

As an orthogonal approach to understanding the genome sequence-AAV production relationship, we also synonymously recoded a rAAV2 transgene. We observed that synonymous recoding of a SaCas9-rAAV2 genome can increase genome-containing particle titers. Here, we explored only a small number of recoded sequences. Exploration of additional sequences or inclusion of specific functional nucleotide motifs may enable more robust improvements in rAAV2 titer. Recently, groups have reported deep-learning-enabled codon optimization methods to enhance tissue-specific rAAV transduction[92]. It will be interesting to determine if these models can be expanded to incorporate rAAV production data and enable increases in rAAV titer.

In conclusion, we have confirmed the importance of encapsidated sequence for rAAV production and transduction. We showed that wild-type AAV2 has a production but not a transduction advantage over rAAV2. We also report that differences in transduction between rAAV2 vectors may be driven by differences in transgene transcription, possibly mediated by interactions with the capsids. Finally, we demonstrate that synonymous recoding can be used to alter rAAV2 production; future work may expand on these results, providing further insight into the mechanisms at play and identifying generalizable rules for rAAV transgene design to optimize production.

2.3 Materials and Methods

2.3.1 Iodixanol purification of AAV particles

Two days prior to transfection, HEK293T cells were seeded in five-layer cell culture multi-flasks (Corning 353144) at 8 x 10^7 cells/flask. We transfected HEK293T cells using PEI. Replicate transfections were performed in separate cell stacks. All transfections included pHelper (25 ug), pRepCap (12.5 ug) and the indicated ITR plasmid (12.5 ug). We changed the media at the time of transfection and added 62.5 mL fresh complete media three days post-transfection. Six days post-transfection, NaCl was added to a final concentration of 0.5 M and samples were incubated at 37°C for 3 hours to detach and lyse cells. Samples were transferred to fresh 500 mL bottles and incubated at 4°C overnight. The following day any precipitate was removed by pipetting and the remaining volume was passed through a sterile 0.22 um PES filter. PEG-8000 was added to a final concentration of 8% and samples were again incubated at 4°C overnight. The following day, we centrifuged samples at 3000 x *g* for 20 minutes to pellet the PEG-precipitated virus. Supernatants were discarded and pellets resuspended in 8 mL of DPBS. We then digested samples with a 1:10,000 dilution benzonase (Millipore Sigma 1.101695.0001) at 37°C for 45 minutes and subjected them to iodixanol gradient ultracentrifugation as previously described[85,140]. Briefly, benzonase-treated samples were underlaid with an iodixanol gradient (Millipore Sigma D1556) in polypropylene tubes (Beckman Coulter, 362183) and centrifuged at 242,000 x *g* for 1 hour at 16°C in a VTi50 rotor. Following ultracentrifugation, 500 uL fractions were collected from the 40% iodixanol layer and buffer-exchanged into DPBS using 100 kDa molecular weight cutoff centrifugal filter units (Millipore Sigma UFC910024). Genome-containing particle titers were determined by qPCR as described above. Capsid titers were determined *via* capsid ELISA (Progen PRATV) according to the manufacturer's protocol. Transfections, affinity purifications, and titer determinations were performed in duplicate.

2.3.2 *In vitro* genome release assay to measure AAV thermostability

We measured *in vitro* genome release using a fluorometric assay as previously described[54]. For each reaction, 1 x 10^{10} viral genomes were mixed with 1X SYBR Gold (Thermo Fisher, S11494) in DPBS (no calcium, no magnesium). Samples were heated from 25°C to 95°C in 2°C increments and fluorescent signal was measured after a 5 minutes incubation at each temperature. Pre-heated controls were also included. All AAV preps were assayed in triplicate.

2.3.3 Transduction of HeLaRC32 cells, RNA extraction and RT-qPCR

Prior to transduction, the genome-containing particle and capsid titers of iodixanol-purified rAAV2 and wild-type AAV2 preps were normalized by spiking in empty capsids. The wild-type AAV2 preps used here contain a synonymously recoded region of the *rep* gene to enable discrimination between transcripts expressed from these constructs and transcripts expressed from *rep* genes integrated into the HeLaRC32 genome. Twenty-four hours prior to transduction, HeLaRC32 cells were seeded in 6-well plates at 1 x 10^5 cells/well. Media was aspirated from the cells and replaced with serum-free media containing adenovirus 5 (MOI = 10) and the indicated AAV preparation (MOI = 100-10,000). Four hours later, the virus-containing media was replaced with fresh complete media. Twenty-four hours post-transduction, RNA was extracted using the RNeasy kit (Qiagen 74104) according to the manufacturer's protocol. On column DNase digests were performed during RNA extraction. RNA quality was confirmed *via* agarose gel and reverse transcription was performed with the Maxima First Strand Synthesis cDNA Kit (Thermo Fisher, K1641). We performed qPCR as described above to quantify p5 transcript levels. P5 transcript levels in wild-type AAV2 transductions were determined using primers and probes that specifically recognize the recoded region of rep. The forward primer had the sequence: 5'-GAACATCTGCCAGGCATTAG-3', the reverse primer had the sequence: 5'-ATTCTTTCGTGAATTGGGTGGCAGA-3', and the probe

had the sequence: 5'-HEX-GTCCGGGGG-ZEN-AAGCTCCCACT-3IABkFQ-3'. p5 transcript levels were normalized to GAPDH abundance. To quantify GAPDH abundance we used a forward primer with the sequence: 5'-ACCACAGTCCATGCCATCAC-3' and a reverse primer with the sequence : 5'-TCCACCACCCTGTTGCTGTA-3'. The same primers were used to amplify GAPDH from the cDNA pool for use as a standard curve.

2.3.4 Transfection of HeLaRC32 cells

Twenty-four hours prior to transfection, HeLaRC32 cells were seeded in 6-well plates at 1 x 10^5 cells/well. Cells were transfected with 500 ng of the indicated ITR plasmids *via* PEI. Six hours post-transfection, serum-containing media was aspirated, cells were washed with DPBS, and serum-free media containing adenovirus 5 (MOI = 10) was added. Four hours post-transduction, Adenovirus 5-containing media was replaced with fresh complete media. Cells were harvested and RNA was extracted and quantified 24 hours post-transfection as described above.

2.3.5 Isolation and quantification of total intracellular DNA

Twenty-four hours prior to transduction, HeLaRC32 cells were seeded in 12-well plates at 3 x 10^4 cells/well. AAV and Ad5 co-transductions were performed as described above using the same MOI. Twenty-four hours post-transduction, we removed the media from the cells, washed with DPBS, and added 100 uL of DNA extraction buffer [1 mM Tris-HCl (pH 8), 1 mM EDTA, 0.001% SDS, 0.062% Deoxycholate, 0.45% Tween, 0.2 mg/mL Proteinase K] per well. Samples were mixed thoroughly by pipetting, transferred to 96-well PCR plates, and incubated at 37°C for 1 hour, 55°C for 2 hours, and 95°C for 30 minutes. 2 uL of each sample were used directly in qPCR. To quantify total intracellular DNA abundance, we used primers and a probe that bound to the p5 promoter. The sequence of the forward primer was 5'-CAGAGAGGGAGTGGCCAACT-3', the sequence of

the reverse primer was 5'-TACCCAGCGTGACCACATGG-3', and the sequence of the probe was 5'-HEX-TGGAGGGGT-ZEN-GGAGTCGTGACGTGAA-3IABkFQ-3'.

2.3.6 Determination of DNase-resistant particle titers by qPCR

All transfections were performed in triplicate. HEK293T cells were cultured in DMEM supplemented with 10% FBS. 24 hours prior to transfection, we seeded HEK293T cells at 2.5 x 10^5 cells/well in 12-well plates. ITR plasmids (5 x 10^9 plasmid molecules/well, 30 ng/well) were transfected along with pRepCap (5 x 10^9 plasmid molecules/well, 40 ng/well) and pHelper (2 x 10^{10} plasmid molecules/well, 240 ng/well) plasmids *via* polyethylenimine (PEI). Three days post-transfection, we lysed cells by 3X freeze-thaw. Samples were then clarified by centrifugation at 15,000 x *g* for 5 minutes. 5 uL of supernatant were subjected to DNase digest (ThermoFisher, EN0521) at 37°C for 30 minutes followed by heat inactivation at 65°C for 10 minutes. We then incubated samples at 98°C for 10 minutes to denature the capsids and measured DNase-resistant particle titers with qPCR (IDT PrimeTime Gene Expression Master Mix). To titer wild-type AAV2, primers and a probe binding to either the 5' region of the *rep* gene or to the *cap* gene were used. The sequence of the forward *rep* primer was: 5'-GAGCATCTGCCCGGCATTTC-3', the sequence of the reverse *rep* primer was: 5'-ATCTGGCGGCAACTCCCATT-3', and the sequence of the *rep* probe was 5'-HEX-ACAGCTTTG-ZEN-TGAACTGGGTGGCCGA-3IABkFQ-3' (IDT). The sequence of the forward *cap* primer was: 5'-CAGACCATGCCTGGAAGAACG-3', the sequence of the reverse *cap* primer was: 5'-GCTCAGAGAAAACAAATGTGGAC-3', and the sequence of the *cap* probe was 5'-HEX-CAGGACAAC-ZEN-CAATCCCGTGGCTA-3IABkFQ-3' (IDT). To titer rAAV2, primers and probes binding to either EGFP, luciferase, or the WPRE3 element were used.The sequence of the forward EGFP primer was: 5'-GAACCGCATCGAGCTGAA-3', the sequence of the reverse EGFP primer was: 5'-TGCTTGTCGGCCATGATATAG-3', and the sequence of the EGFP probe was: 5'-FAM-ATCGACTTC-ZEN-AAGGAGGACGGCAAC-3IABKFQ-3' (IDT).

The sequence of the luciferase forward primer was: 5'-CATGTACCGCTTCGAGGAG-3', the sequence of the luciferase reverse primer was: 5'-GAAGCTAAATAGTGTGGGCACC-3', and the sequence of the luciferase probe was: 5'-FAM-CTTGCGCAG-ZEN-CTTGCAAGACTATAAGATTC-3IABKFQ-3' (IDT). The sequence of the WPRE3 forward primer was:

5'-GCTTTAATGCCTTTGTATCATGCTAT-3', the sequence of the WPRE3 reverse primer was: 5'-GCCGTGGCAAGAACTAACCA-3', and the sequence of the WPRE3 probe was:

5'-FAM-CTTCCCGTA-ZEN-TGGCTTTCATTTTCTCCTCC-3IABkFQ-3' (IDT).

2.3.7 Synonymous recoding of wild-type AAV2 and SaCas9

The wild-type AAV2 genome was divided into three segments, the *rep* gene, the 3' region of the *cap* gene, and the VP3-coding region of the *cap* gene. Each of these three segments was synonymously recoded separately. The goal of synonymous recoding was to alter the nucleotide sequence as much as possible without altering protein expression. To accomplish this, synonymous recoding was performed in two ways: *via* codon shuffling to preserve codon abundance across the recoded region and *via* maximal recoding, in which synonymous codon changes were made wherever possible. Synonymous recoding was implemented in Python. Briefly, for each region, a table was generated containing each amino acid and a list of all the codons used to code for that amino acid throughout the region to be recoded. Codon shuffling was performed by iterating over the amino acid sequence of a given region and randomly selecting a codon from that list. Maximal recoding was performed by iterating over the amino acid sequence and forcing a change to a synonymous codon. Codon abundance was not preserved in maximally recoded sequences. This resulted in changes to all codons except for methionine and tryptophan, which are coded for by only a single codon.

Following generation of a recoded sequence, we calculated the CpG and TpA dinucleotide abundances, the mean free energy (MFE) of the predicted RNA secondary structure (ViennaRNA), the Hamming distance of that sequence from the input wild-type sequence, and codon pair bias

(CPB)[52,68].We aimed to design recoded sequences that had minimal changes to these parameters as they have all been shown to affect protein expression. CpG-specific methyltransferase and deamainases can result in the generation of C → T mutations within CpG dinucleotides[118]. Additionally, unmethylated CpGs are recognized by toll-like receptor 9[49]. TpA dinucleotides are the most energetically unstable dinucleotides and are often found in regions where DNA is bent and unwound; depletion of TpAs may alter regulatory factor binding[63]. Additionally, UpA dinucleotides are preferential targets for ribonucleases and so may reduce mRNA stability[63]. The MFE of the predicted RNA secondary structure is also a predictor of mRNA stability. Finally, CPB is a measure of the proportion of overrepresented codon pairs within a given sequence and is calculated from codon pair score (CPS)[28].

$$CPS = ln(\frac{F(AB)}{\frac{F(A)xF(B)}{F(X)xF(Y)}xF(XY)})$$

Where F(A) is the frequency of codon A, F(B) is the frequency of codon B, F(AB) is the frequency of codon pair AB, F(X) is the frequency of the amino acid coded for by codon A, F(Y) is the frequency of the amino acid coded for by codon B, and F(XY) is the frequency of the amino acid pair XY in the human genome. CPB is the mean of the CPS for all codon pairs in a given open reading frame.

$$CPB = \sum_{i=1}^{k} \frac{CPS_{\mathrm{i}}}{k-1}$$

Given the importance of these parameters for regulatory factor binding, mRNA stability, and protein expression, we sought to limit changes to them in the synonymously recoded sequences. Multiple codon shuffled and maximally recoded sequences were randomly generated as described above. Sequences were selected for synthesis and assay based on the similarity of their dinucleotide

abundances, RNA MFE, and codon pair bias to the input wild-type sequence.

For synonymous recoding of the 5' region of the *cap* gene we also had to consider the effect of VP recoding on AAP and MAAP, which are expressed from frameshifted open reading frames within the *cap* gene. If we maintained the amino acid sequences of VP3, AAP, and MAAP, very few nucleotide changes could be introduced. Instead we chose to prevent expression of AAP and MAAP from the recoded *cap* gene by mutating the MAAP start codon and introducing a premature stop codon into the AAP frame. Both of these mutations were previously shown to ablate MAAP and AAP expression and correspond to synonymous changes in the VP frame[72,85]. When using these recoded variants to produce AAV, we carried out transfections with and without wild-type AAP and MAAP supplied *in trans*.

The above approach was also used for SaCas9 recoding. Here, the entire SaCas9 open reading frame was treated as a single region.

2.3.8 Western blot analysis of VP and Rep protein levels

Following transfection, media was aspirated from cell culture plates and cells were washed with DPBS. We then lysed cells with 50 uL of lysis buffer per well (6-well plates). The lysis buffer contained: 10 mM Tris-HCl (pH 8), 150 mM NaCl, 1% Triton X-100, and cOmplete mini protease inhibitor (1 tablet/10 mL lysis buffer). Lysates were transferred to fresh 1.5 mL tubes and centrifuged at 15,000 x *g* for 5 minutes to clarify. Total protein concentrations were determined by Bradford assay (Thermo Scientific Pierce Coomassie Bradford Protein Assay Kit). We loaded 75 ug of total protein from each sample on 4-12% Bis-Tris gels (ThermoFisher) and ran the gels for 45 minutes at 140V. We then transferred the proteins to PVDF membranes (ThermoFisher). Membranes were blocked with 5% milk in PBST for 1 hour at 4°C. Blots were then incubated with a 1:250 dilution of B1 anti-VP antibody or a 1:100 dilution of anti-Rep 303.9 antibody (both American Research Products) in blocking buffer overnight. Both primary antibody mixtures also included a 1:20,000

dilution of anti-□-actin antibody (Abcam, ab8227). The following day, blots were washed 3X with PBST and incubated with secondary antibodies for 1 hour in blocking buffer with 0.01% SDS. Goat anti-mouse IRDye 800CW (LiCor 925-32210) and donkey anti-rabbit IRDye 680RD (LiCor 925-68073) secondary antibodies were used. Blots were washed 3X with PBST and 1X with PBS, and the near-infrared fluorescence of the secondary antibodies was visualized using the Sapphire imager.

Science is about knowing; engineering is about doing.

Henry Petroski

3

Conclusion

In this thesis, I examined the effect of both *rep* and transgene sequence variants on AAV production. I began by assaying the effect of all single codon variants of AAV2 *rep* on AAV2 capsid production. This work provides a useful data set, mapping the relationship between *rep* sequence and function. I identified many functional *rep* sequences not observed in nature, highlighting the utility of this comprehensive mutagenesis approach for the discovery of novel functional sequence diversity. I also observed that beneficial substitutions clustered in the DNA-interacting regions of the Rep

origin-binding domain. Further characterization of a subset of variants indicated that the effect of *rep* variants on production is generally consistent across rAAV genomes, and additional library experiments demonstrated that the effect of *rep* variants on production is also consistent across capsid serotypes.

Notably, I observed that, generally, the effect of single codon *rep* variants on production was smaller than the effect of previously characterized single codon *cap* variants[85]. The top-performing *rep* variants identified here resulted in two-fold increases in viral genome titer. These titer improvements were reproduced in small and large scale triple plasmid transfections with individual *rep* variants. For at least one variant, I demonstrated that improvements in viral genome titer were correlated with improvements in transducing particle titer. However, the utility of this work could be expanded by identifying *rep* variants that result in larger titer improvements. Separate from the work discussed here, I contributed to efforts to develop algorithms to predict the fitness of capsid variants containing tens of amino acid substitutions[13]. These algorithms used data from the *cap* comprehensive mutagenesis library to predict the effect of multiple substitutions on AAV production. Future work to apply these same algorithms to the *rep* comprehensive mutagenesis data may identify *rep* variants with larger effects on rAAV titer. Additionally, further characterization of rAAV produced with individual *rep* variants by analytical ultracentrifugation and/or cryo-electron microscopy could provide more insight into the effect of *rep* variants on the proportion of full capsids, a key parameter in rAAV production. Recent work from the Agbandje-McKenna lab identified *rep* hybrids that increased the proportion of full capsids, mainly via changes to the origin-binding domain[78]. It remains to be seen if any of the novel functional variants identified here affect the proportion of full capsids.

In addition to generating a data set on the effect of *rep* variants on rAAV production, I have also generated libraries that I hope will be useful tools for further investigation of *rep* biology. The Rep proteins are involved not only in production of AAV particles but also in integration of AAV

genomes into the host cell genome[116]. Purification and sequencing of genomic DNA could provide insight into how *rep* variants affect genome integration. Additionally, alternative AAV production systems, using stable cell lines and/or helper virus co-infection, are employed, and it will be interesting to determine if *rep* variants affect production differently in these contexts[114,59].

Concurrent with this work, there has been an explosion of protein structure prediction algorithms, such as AlphaFold and, subsequently, amazing progress in both *de novo* protein design and variant effect prediction[61,4,22]. These advances are already streamlining protein engineering efforts and it will be interesting to see them applied to both the AAV capsid and Rep proteins. Advances in protein design algorithms, however, do not completely eliminate the utility of comprehensive mutagenesis libraries, like the one described here. Comprehensive mutagenesis is a useful tool not only for protein engineering but also for investigation of viral biology. This work, other efforts from our lab, and deep mutational scanning work from the Bloom lab have leveraged large, unbiased libraries to uncover novel biology across a range of viruses, including AAV, SARS-CoV-2, and influenza[110,65]. Furthermore, the function, or fitness, of many proteins represents an aggregate of multiple characteristics and activities that may be hard to capture without empirical data. For example, the Rep proteins effects on rAAV production are determined by a combination of expression level, binding affinity to the ITRs, helicase activity, endonuclease activity, and binding affinity to the capsid. Empirical measurement of complex protein activities may continue to be useful to complement protein design algorithms.

After investigating the effect of *rep* variants on AAV production, I shifted my attention to the effect of transgene sequence variants on production. First, I confirmed previous reports that wild-type AAV2 production results in higher genome-containing particle titers than rAAV2 production. Wild-type AAV2 transduction, however, did not result in uniformly higher mRNA levels than rAAV2 transduction. Interestingly, I observed that transgenes delivered in rAAV2 vectors had different relative expression levels than the same transgenes expressed off of plasmids. Other groups

have identified a role for the capsid and capsid-promoter pairings in regulating the transcriptional activity of rAAV transgenes[41,10]. My data suggest that transgene sequence, independent of promoter identity, may also affect capsid-mediated transcriptional regulation. However, more work is needed to better understand the sequence determinants of capsid-mediated transcriptional regulation and to identify the specific mechanisms driving this regulation. Finally, I turned my attention back to AAV production and demonstrated that synonymous recoding of a rAAV2 transgene can be used to improve genome-containing particle titers, although there was a tradeoff between rAAV2 titer and transgene expression. Future work could assay additional recoded sequences and, potentially, identify generalizable rules for rAAV transgene design.

The work presented here investigates the effect of two relatively understudied components of rAAV vectors, the *rep* gene and transgene sequence, on rAAV production. I generated a comprehensive data set mapping the effect of *rep* variants on rAAV production, and I demonstrated that non-coding nucleotide changes to transgene sequences can affect rAAV production. Taken together, these results provide insight into the AAV production process and lay the groundwork for enhancement of large-scale gene therapy production.

A

Supplemental Material for Chapter 1

A.1 Materials and Methods

A.1.1 Analysis of stop codons in alternate reading frames

At each codon position, the average fitness value for all variants that introduced a premature stop codon into the +1 reading frame was determined. The average fitness value for all variants that did not introduce a stop codon into the +1 frame but were synonymous in the Rep frame to those that

did was also calculated. Ten amino acid rolling averages for the +1 stops and +1 non-stops were calculated. These average fitness values were plotted and compared. The same procedure was used to determine the effect of premature stop codons in the +2 reading frame.

A.1.2 Comparison of effects on synonymous and non-synonymous codons on production fitness

For this analysis we used production fitness values (s') calculated at the codon level. For each fitness value, we calculated two mean centered fitness values. Positional mean centered fitness values were calculated as:

$$fitness_{\text{positional mean centered}} = \left(\frac{s'_{\text{codon}}}{averages'_{\text{amino acid position}}}\right)$$

Synonymous codon mean centered fitness values were calculated as:

$$fitness_{\text{synonymous codon mean centered}} = \left(\frac{s'_{\text{codon}}}{averages'_{\text{synonymous codon variants}}}\right)$$

Methionine and tryptophan codons were excluded as these amino acids are coded for by only a single codon. Mean centered fitness values for each codon variant were plotted versus amino acid position.

A.1.3 Affinity purification of AAV2 capsids

We purified rAAV2 vectors produced with variant pRepCap plasmids with AVB Sepharose (Cytiva) using a previously reported protocol, which we modified for use with gravity columns[79]. A 4.4 kb inverted terminal repeat genome (pAAV-EF1a-FLuc-WPRE-HGHpA) was used. pRepCap plasmids and the pAAV-EF1a-FLuc-WPRE-HGHpA inverted terminal repeat plasmid were quantified

by qPCR as discussed above. HEK293T cells were harvested three days post-transfection and lysed by 3X freeze-thaw. We then subject samples to benzonase digest as described above and centrifuged samples at 6,000 x g for 30 minutes to pellet cell debris. Samples were then passed through 0.2 um PES filters, diluted 1:1 with TD Buffer (DPBS, 1 mM $MgCl_2$, 2.5 mL KCl), and incubated with 4 mL of AVB Sepharose slurry at 4°C overnight with shaking. The following day, we loaded the samples onto gravity columns, washed with 20 mL of 1X TD Buffer, and eluted samples with 5 mL 0.1 M glycine-HCl (pH 2.6). Eluates were immediately neutralized with 500 uL of 1 M Tris-HCl (pH 10). All elution fractions were pooled and buffer-exchanged as above and samples volumes were normalized to 400 uL. rAAV viral genome titers were determined via qPCR using the same luciferase primer and probe sequences discussed above. Total particle titers were determined via capsid ELISA (Progen PRATV) according to the manufacturer's protocol. Transfections, affinity purifications, and titer determinations were performed in duplicate.

A.1.4 Determination of relative transduction efficiency

Twenty-four hours prior to transduction, we seeded HEK293T cells at 2 x 10^4 cells/well in opaque white-bottom 96-well plates. Each rAAV prep was diluted 1:10,000 in serum-free media. 100 uL of this dilution was added to each well of a 96-well plate. We performed eight transductions with each rAAV prep. Forty-eight hours after transduction, we lysed the cells and measured luciferase activity (Promega One-Glo EX Luciferase Assay System, E8110).

A.2 Supplemental Figures

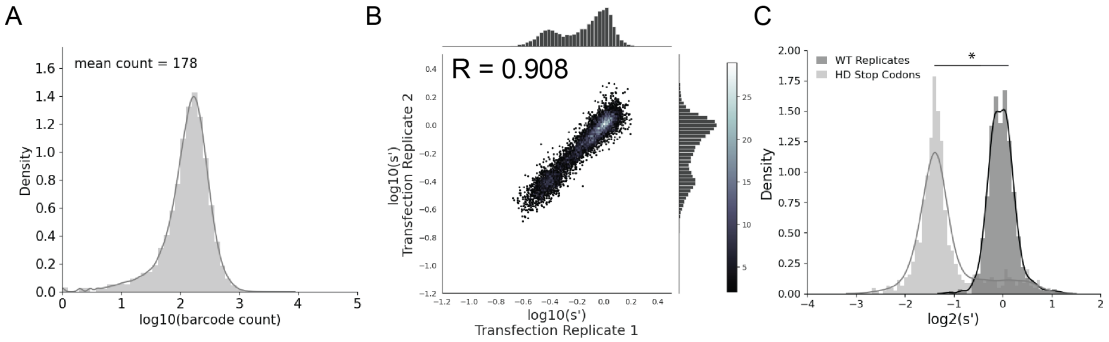


Figure A.1: Comprehensive mutagenesis library design and production assay results for WT AAV2 format library. (A) Density plot of barcode counts in the WT AAV2 plasmid library. (B) Amino acid level production fitness values from replicate transfections for the WT AAV2 library. Pearson R correlation coefficient calculated after log transformation. (C) Density plot of wild-type (black) and premature stop codon (gray) controls for the WT AAV2 library. *p < 10-20 (Mann Whitney U test)

Table A.1: Percent of expected variants sequenced in plasmid and viral libraries

	pCMV-Rep78/68		WT AAV2	
	Plasmid	Viral	Plasmid	Viral
Barcodes	99.8%	98.0%	98.3%	97.6%
Codon Variants	100%	99.9%	99.7%	99.7%
Amino Acid Variants	100%	100%	99.9%	99.9%

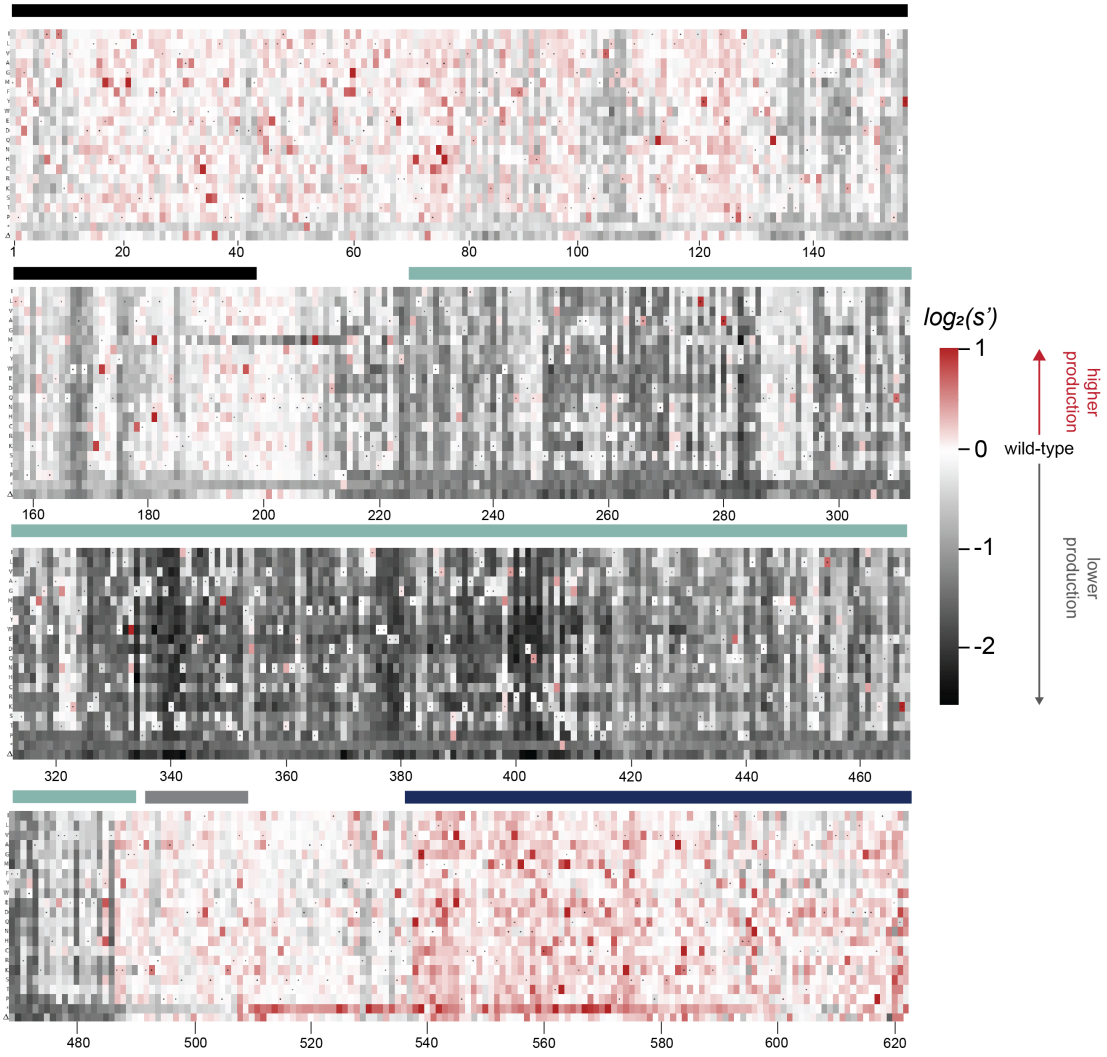


Figure A.2: Effects of all single amino acid substitutions and deletions in Rep78, Rep68, Rep52, and Rep40 on AAV2 production. Amino-acid level production fitness values from the WT AAV2 library production assay were calculated as in Figure 1C by summing barcode counts for synonymous variants. Rectangles are colored by mutational effect on the production of genome-containing particles, with black indicating deleterious mutations and red indicating beneficial mutations. Colored bars above the heatmaps indicate protein domains. Black: origin-binding domain, light blue: helicase domain, gray: nuclear localization signal, and navy blue: zinc-finger domain. Black dots indicate wild-type amino acid identity.

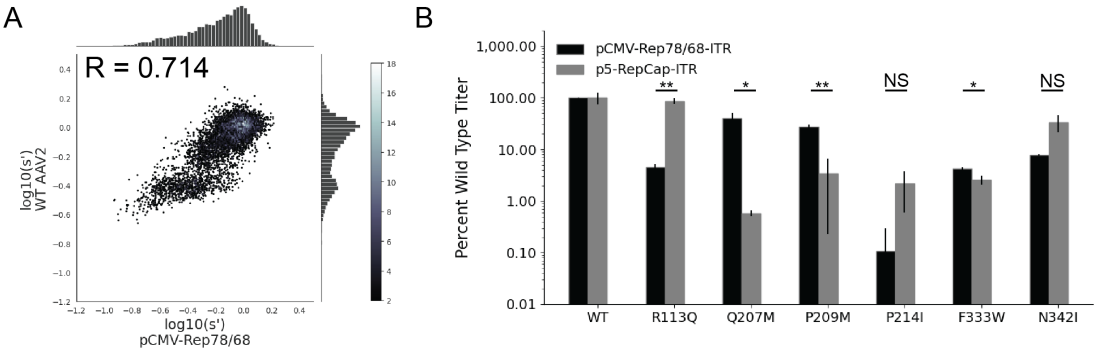


Figure A.3: Mutations to AAV2 rep have similar effects on AAV2 production in pCMV-Rep78/68 and WT AAV2 format libraries. (A) Production fitness values for library production assay with pCMV-Rep78/68 and WT AAV2 format libraries, Pearson R correlation coefficient calculated after log transformation, (B) DNase-resistant particle titers for AAV2 capsids produced with pCMV-Rep78/68-inverted terminal repeat or p5-RepCap-inverted terminal repeat plasmids containing the indicated Rep substitutions. *p < 0.05, **p < 0.01 (Welch's t test)

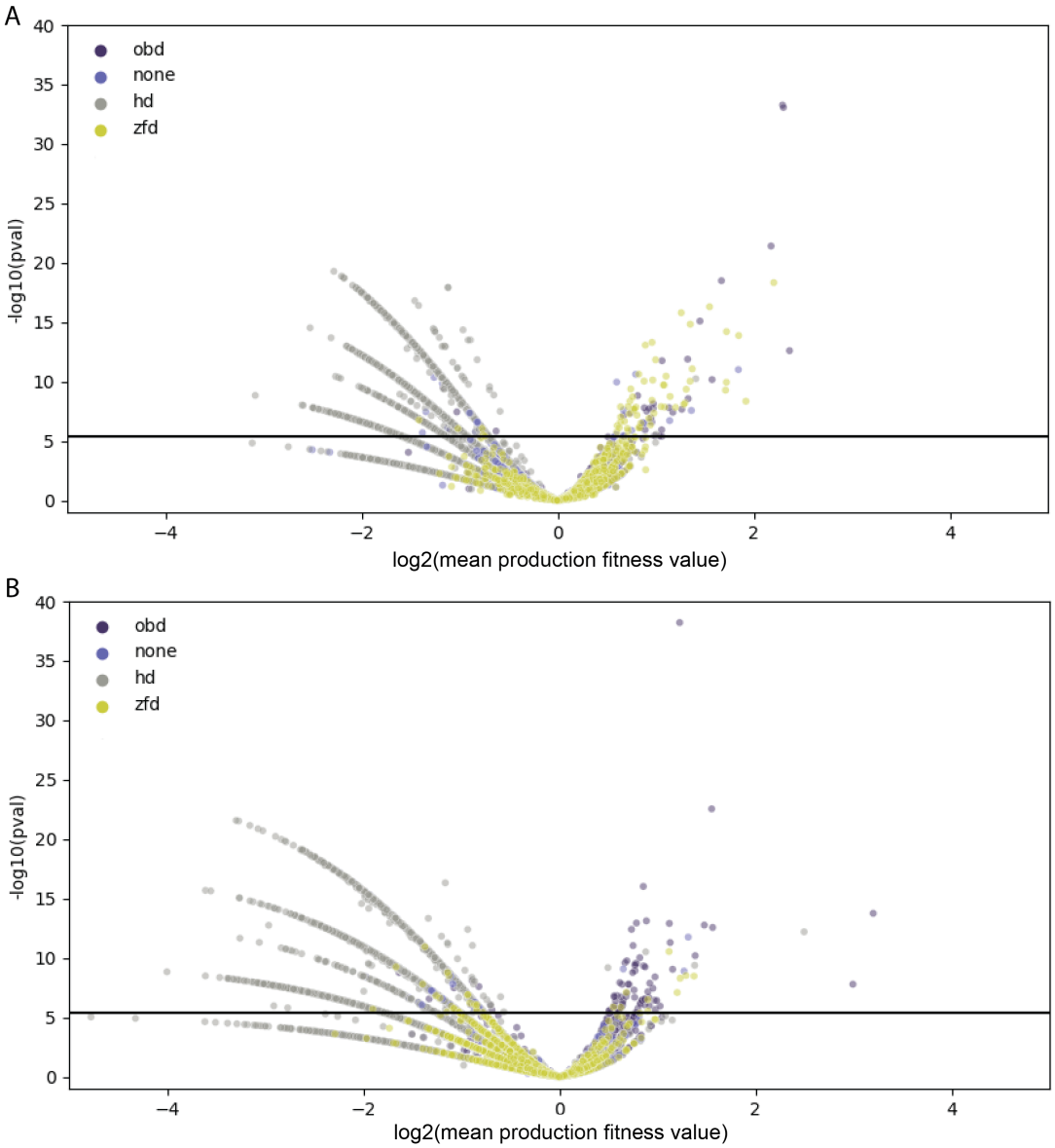


Figure A.4: Comparison of Rep variant and wild-type production fitness values. For each amino acid variant, mean production fitness values were calculated by averaging the production fitness values for all uniquely barcoded variants coding for the relevant amino acid change. The mean production fitness value for each amino acid variant was compared to the mean production fitness value for the wild-type controls using a t-test. (A) Data for the WT AAV2 library. (B) Data for the pCMV-Rep78/68 library. The solid black line indicates the significance threshold after Bonferroni correction. Data points are colored according to their position in Rep where "none" refers to variants that fall outside of an annotated protein domain.

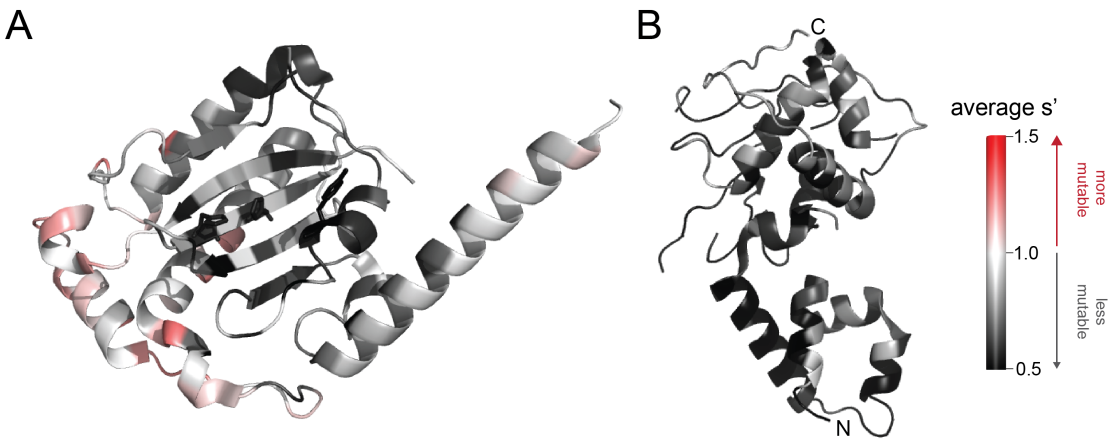


Figure A.5: Average production fitness values from the pCMV-Rep78/68 library production assay mapped onto (A) the structure of the origin-binding domain active site (PDB 5DCX) and (B) the structure of the helicase domain (PDB 1S9H). H90, H92, and the active site nucleophile, Y156, are shown as sticks in (A). Residues are colored by mutability, with red indicating higher mutability and black indicating lower mutability.

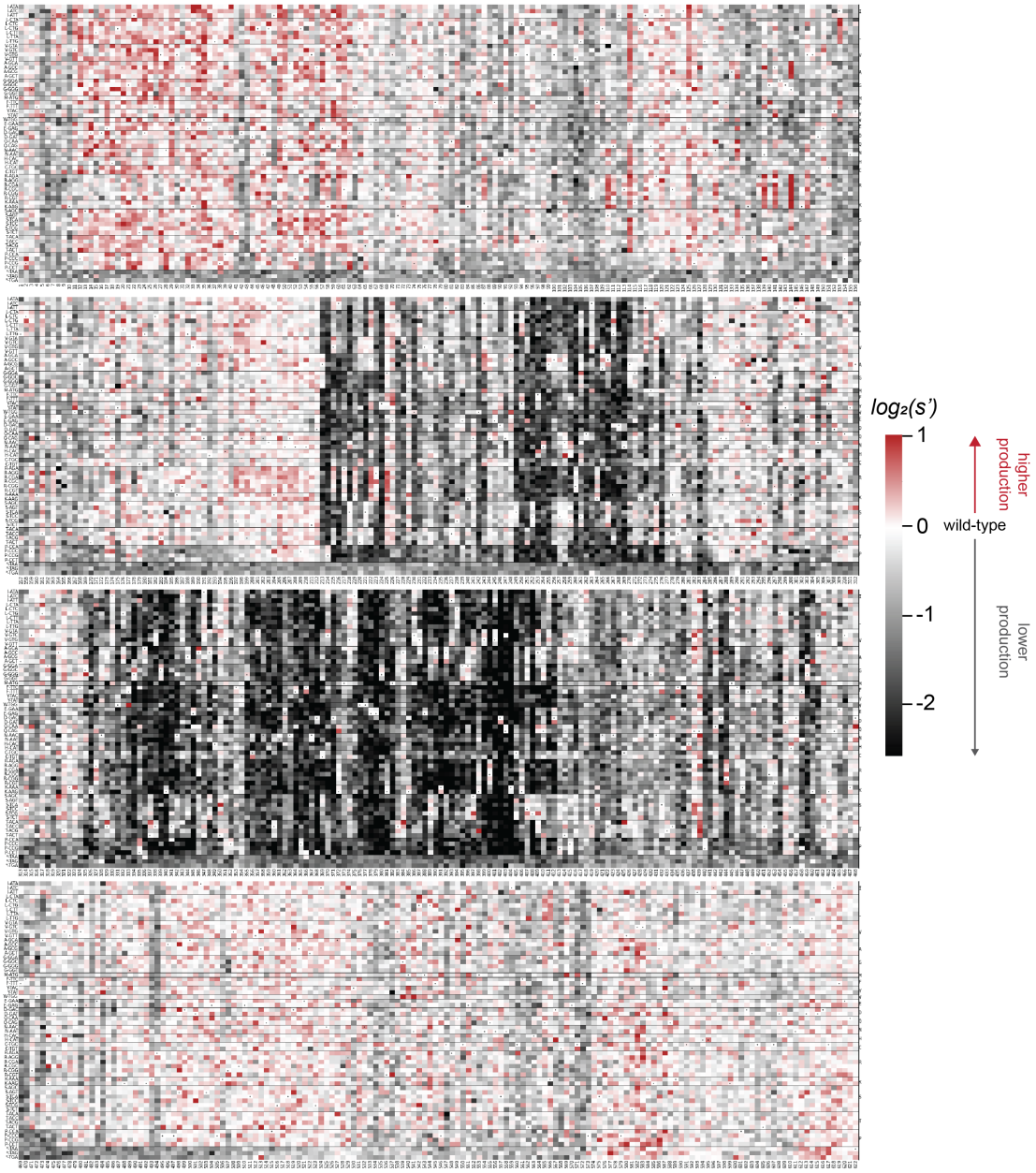


Figure A.6: Codon level production fitness values for the pCMV-Rep78/68 format library. Codon level production fitness values were calculated as in Figure 1C by summing counts for all barcodes corresponding to a given codon variant. Rectangles are colored by mutational effect on the production of genome-containing particles, with black indicating deleterious mutations and red indicating beneficial mutations. Black dots indicate wild-type codon identity.

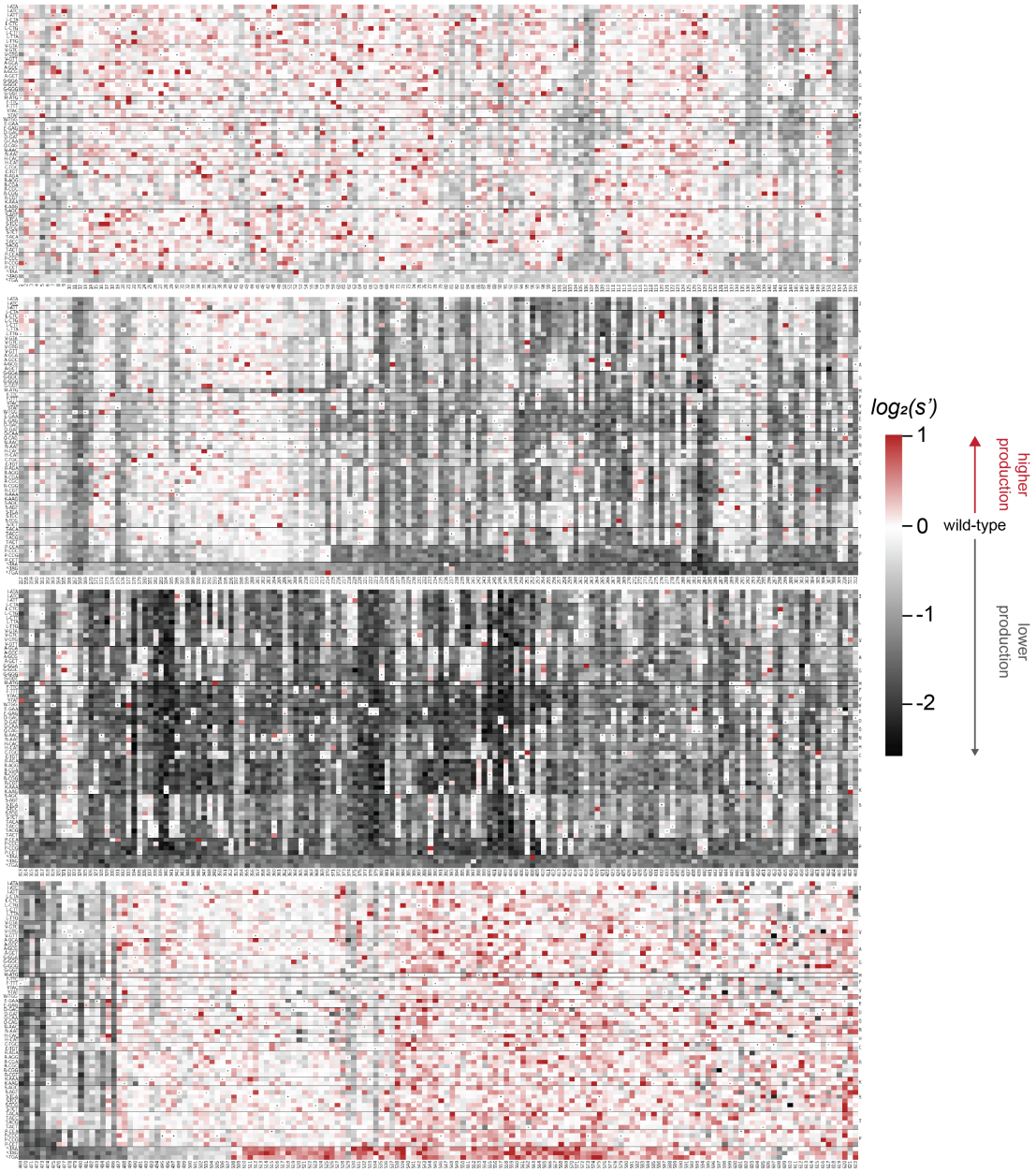


Figure A.7: Codon level production fitness values for the WT AAV2 format library. Codon level production fitness values were calculated as in Figure 1C by summing counts for all barcodes corresponding to a given codon variant. Rectangles are colored by mutational effect on the production of genome-containing particles, with black indicating deleterious mutations and red indicating beneficial mutations. Black dots indicate wild-type codon identity.

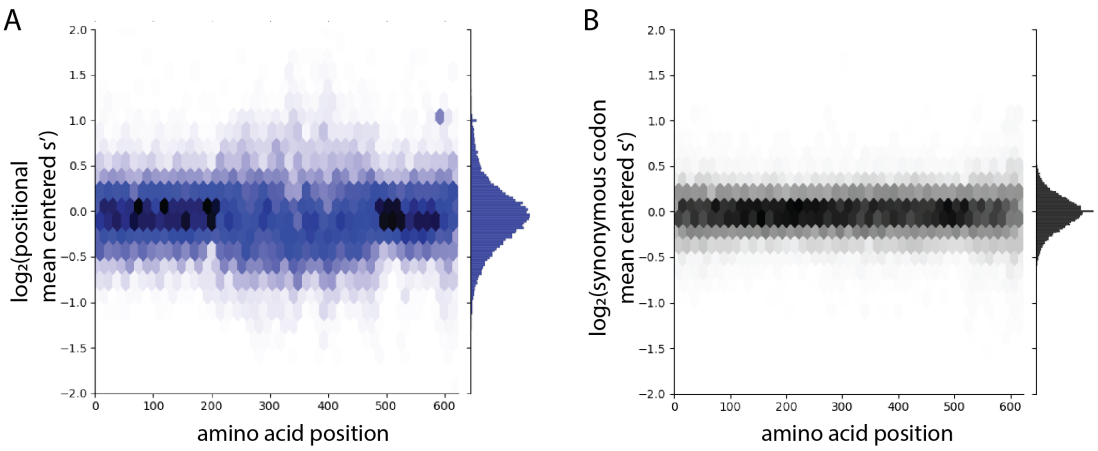


Figure A.8: The distribution of production fitness values is narrower for synonymous variants than for nonsynonymous variants. (A) Positional mean centered fitness values for all codon variants in the WT AAV2 library and (B) Synonymous codon mean centered fitness values for all codon variants in the WT AAV2 library. For each codon variant, positional mean centered fitness values were calculated by normalizing to the average selection value of all codon variants at that position. Synonymous mean centered fitness values were calculated by normalizing the fitness values of each codon variant to the average fitness value of all its synonymous codons.

Figure A.9 *(following page)*: Inclusion of synonymous codon variants enables interrogation of nucleotide-level effects. Analysis of pCMV-Rep78/68 variants with premature stop codons in +1 (A) and +2 (B) reading frames does not identify any frameshifted open reading frames. Pink dots: average production fitness for mutations that introduce stop codons into +1 or +2 frame, black dots: average production fitness for mutations synonymous to +1 or +2 stop codon mutations in Rep open reading frame, lines indicate 10 bp rolling averages. (C) Average production fitness at each nucleotide position for variants that introduce the indicated nucleotide change in the pCMV-Rep78/68 library, orange: T, blue, G, green: C, red: A. (D) Average production fitness for variants that introduce the indicated nucleotide change in the WT AAV2 library.

Figure A.9: (continued)

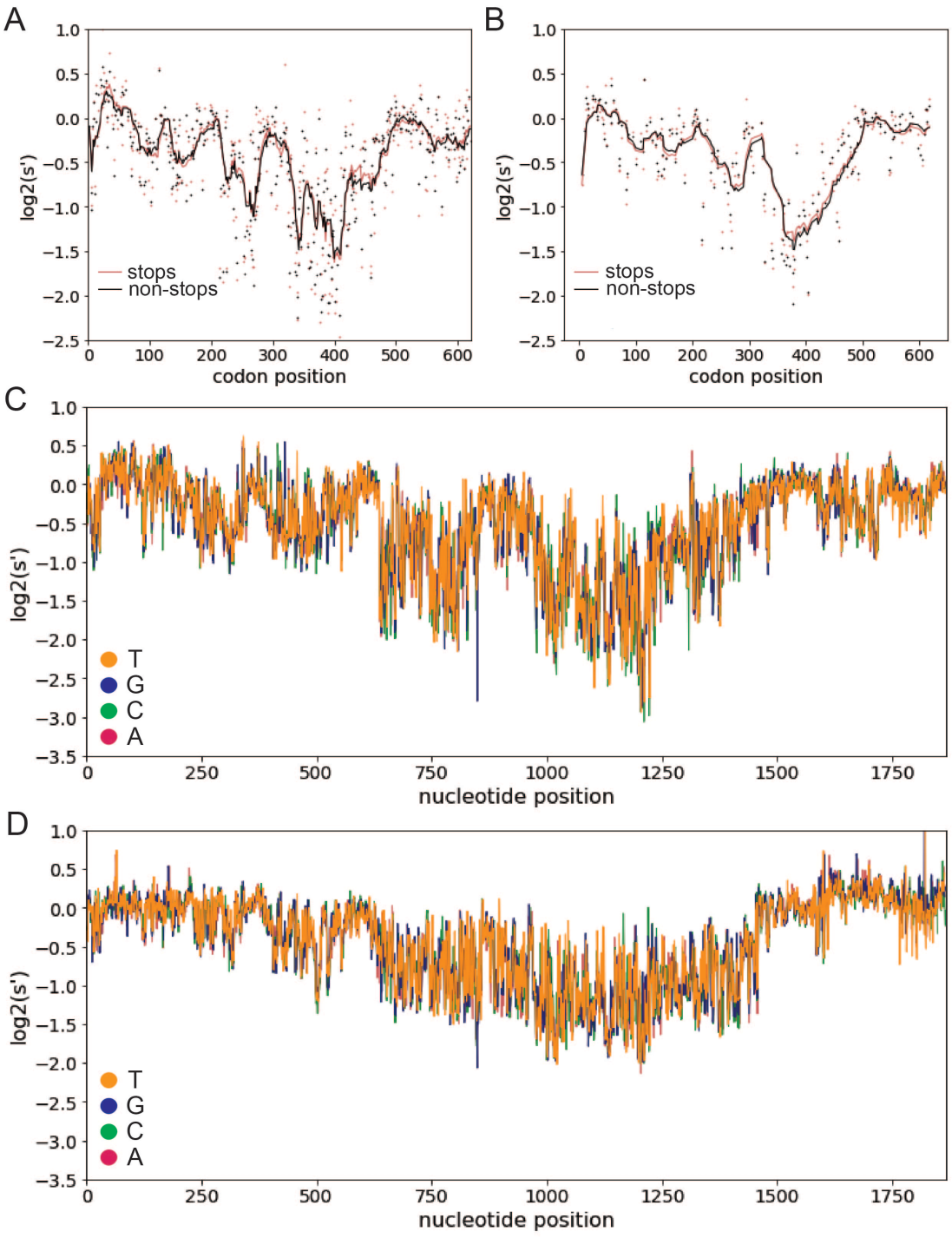
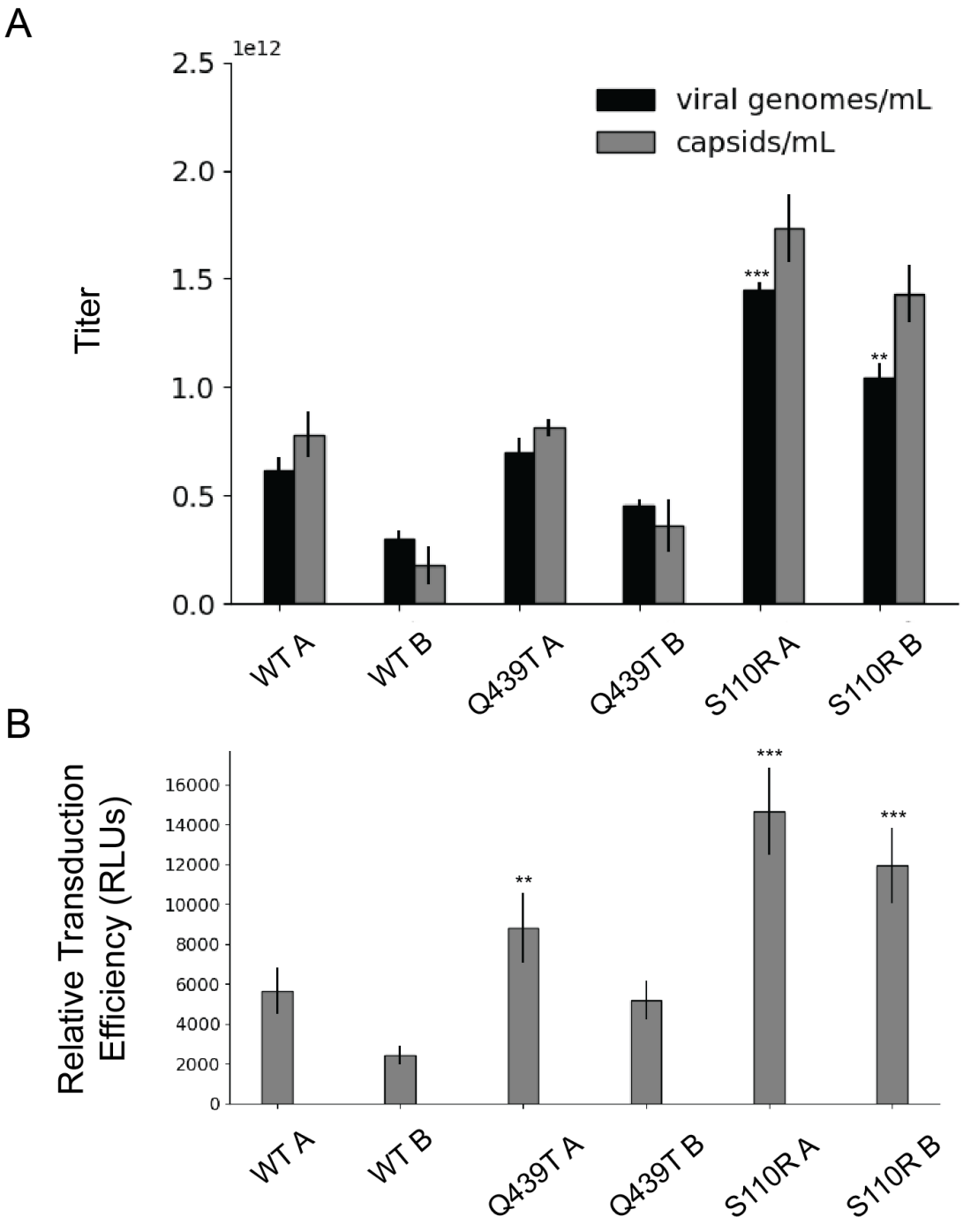


Figure A.10 *(following page)*: Effect of single amino acid Rep substitutions on the genome-containing particle titer, physical particle titer, and relative transduction efficiency of affinity purified rAAV2. (A) Capsid (gray) and viral genome (black) titers for pAAV-EF1a-FLuc-WPRE-HGHpA rAAV2 produced with the indicated Rep variant and affinity purified with AVB Sepharose. For each variant, samples A and B represent replicate transfections and affinity purifications. (B) Relative transduction efficiency of purified rAAV preps as measured by luciferase signal. Asterisks indicate significance of difference between relevant prep and WT A prep. *p < 0.05, **p < 0.01, ***p < 0.001 (Welch's t-test)

Figure A.10: (continued)



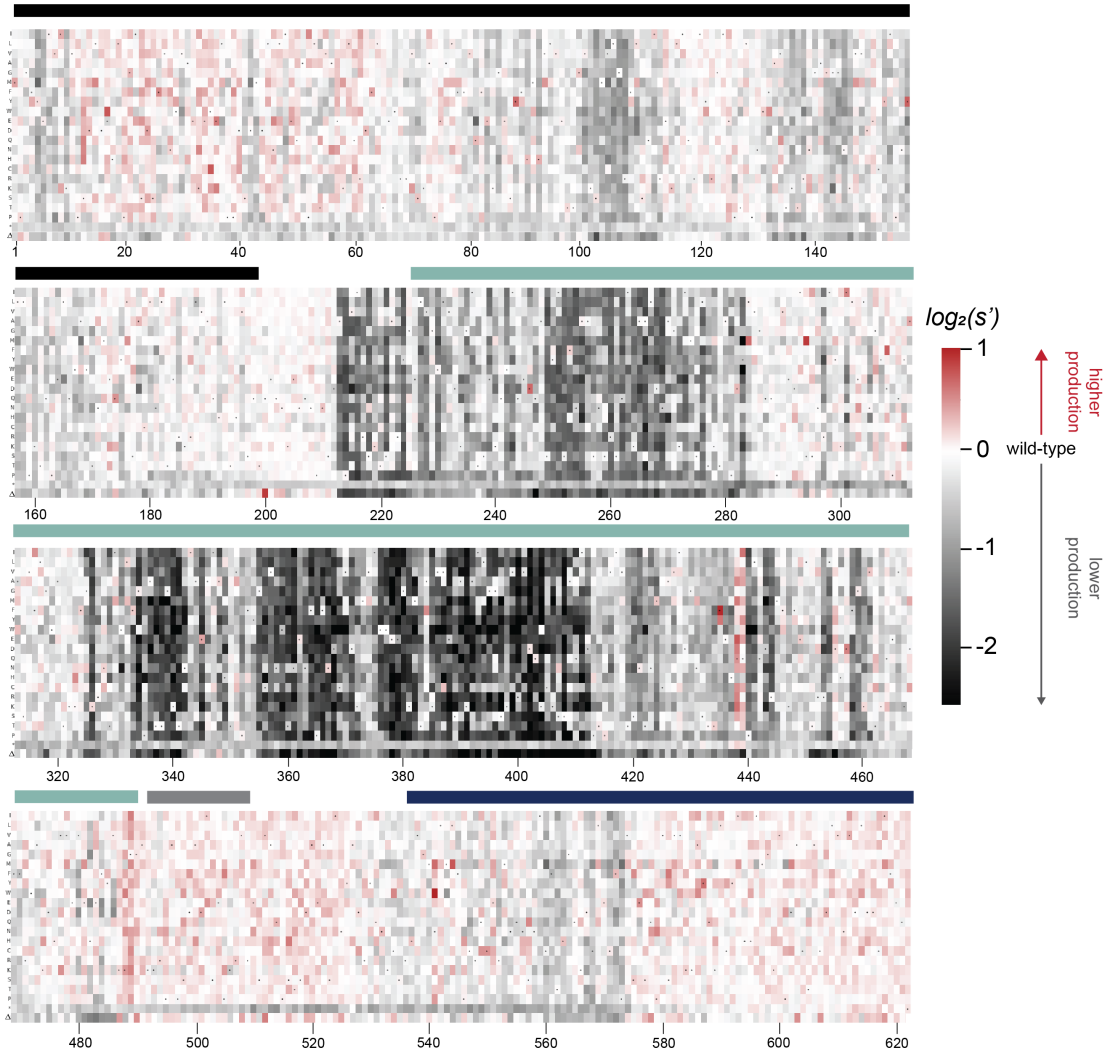


Figure A.11: Effects of all single amino acid substitutions and deletions in the Rep78/68 proteins on AAV5 capsid production. Amino acid level production fitness values from the pCMV-Rep78/68 production assay were calculated as in Figure 1C by summing barcode counts for synonymous mutations. Rectangles are colored by mutational effect on the production of genome-containing particles, with black indicating deleterious mutations and red indicating beneficial mutations. Colored bars above the heatmaps indicate protein domains. Black: origin-binding domain, light blue: helicase domain, gray: nuclear localization signal, and navy blue: zinc-finger domain. Black dots indicate wild-type amino acid identity.

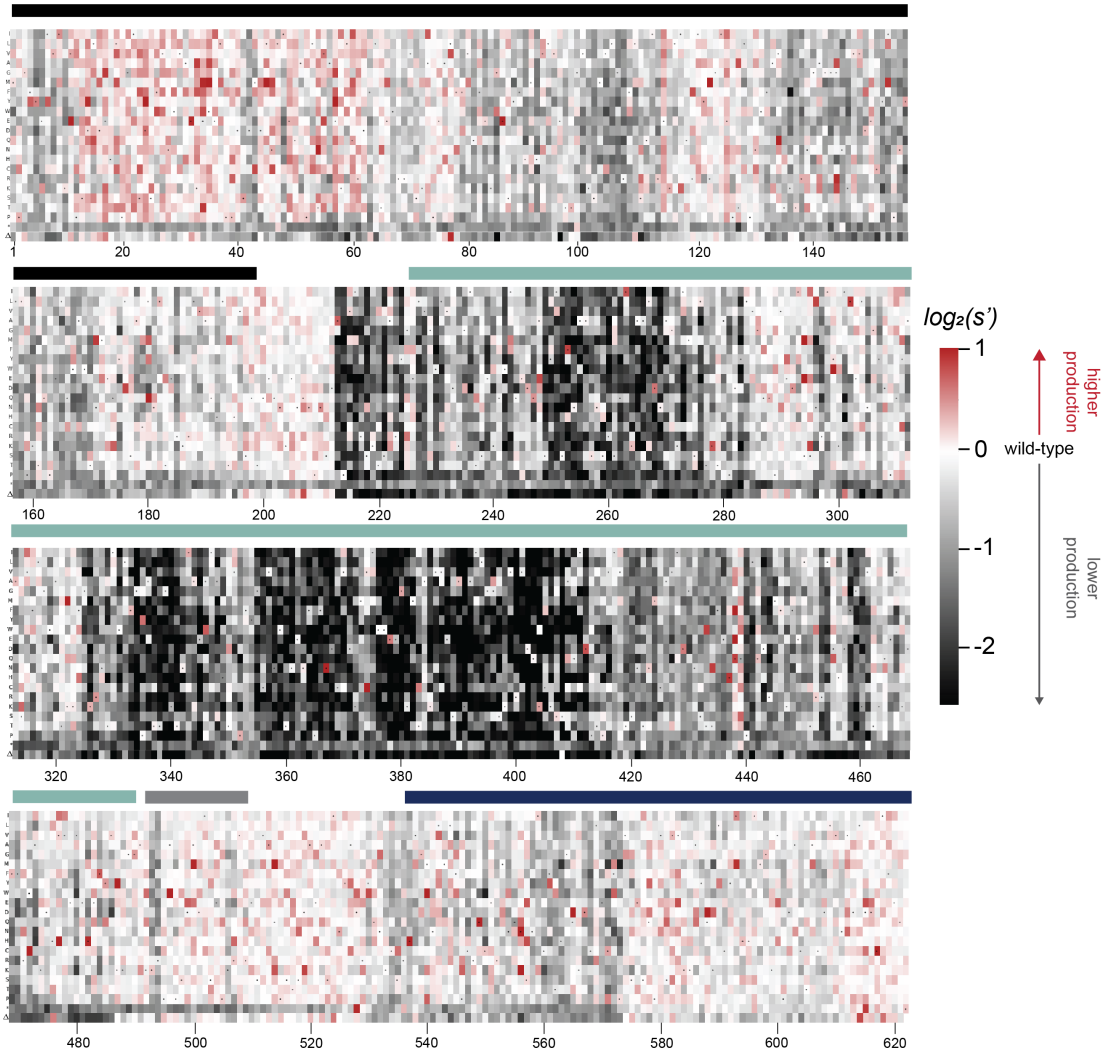


Figure A.12: Effects of all single amino acid substitutions and deletions in the Rep78/68 proteins on AAV9 capsid production. Amino acid level production fitness values from the pCMV-Rep78/68 production assay were calculated as in Figure 1C by summing barcode counts for synonymous mutations. Rectangles are colored by mutational effect on the production of genome-containing particles, with black indicating deleterious mutations and red indicating beneficial mutations. Colored bars above the heatmaps indicate protein domains. Black: origin-binding domain, light blue: helicase domain, gray: nuclear localization signal, and navy blue: zinc-finger domain. Black dots indicate wild-type amino acid identity.

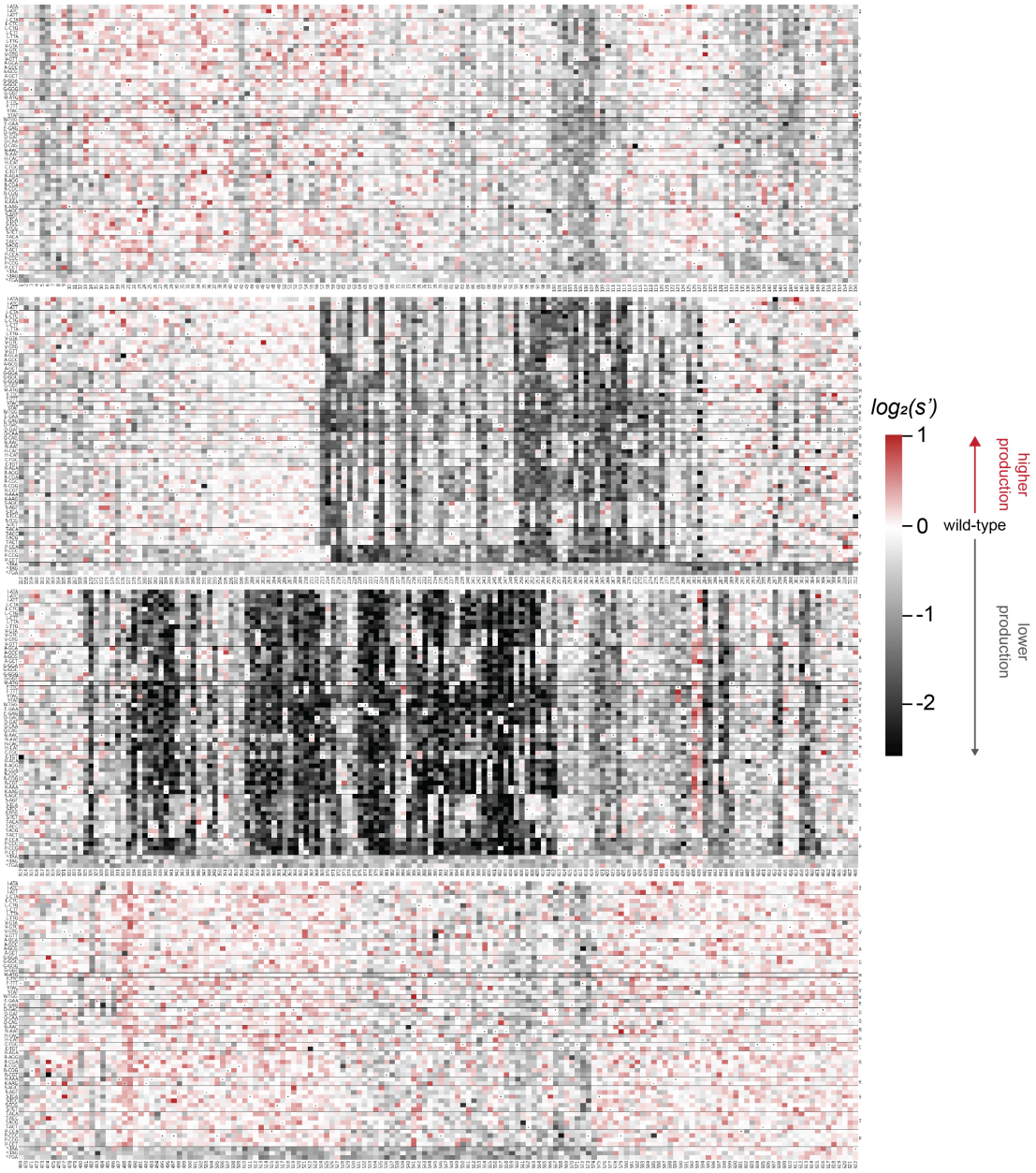


Figure A.13: Codon level production fitness values for the AAV5 capsid production assay. Codon level production fitness values were calculated as in Figure 1C by summing counts for all barcodes corresponding to a given codon variant. Rectangles are colored by mutational effect on the production of genome-containing particles, with black indicating deleterious mutations and red indicating beneficial mutations. Black dots indicate wild-type codon identity.

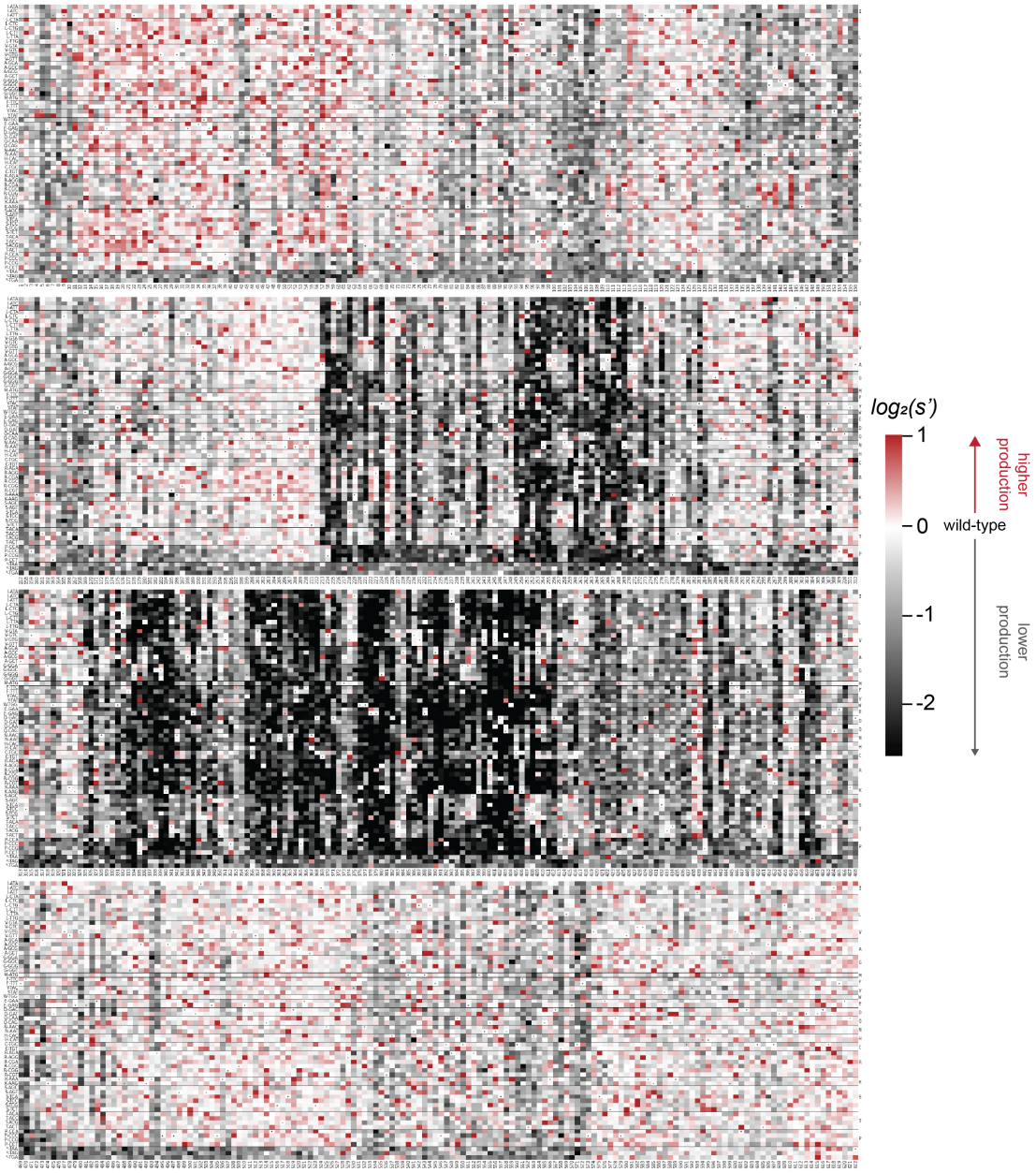


Figure A.14: Codon level production fitness values for the AAV9 capsid production assay. Codon level production fitness values were calculated as in Figure 1C by summing counts for all barcodes corresponding to a given codon variant. Rectangles are colored by mutational effect on the production of genome-containing particles, with black indicating deleterious mutations and red indicating beneficial mutations. Black dots indicate wild-type codon identity.

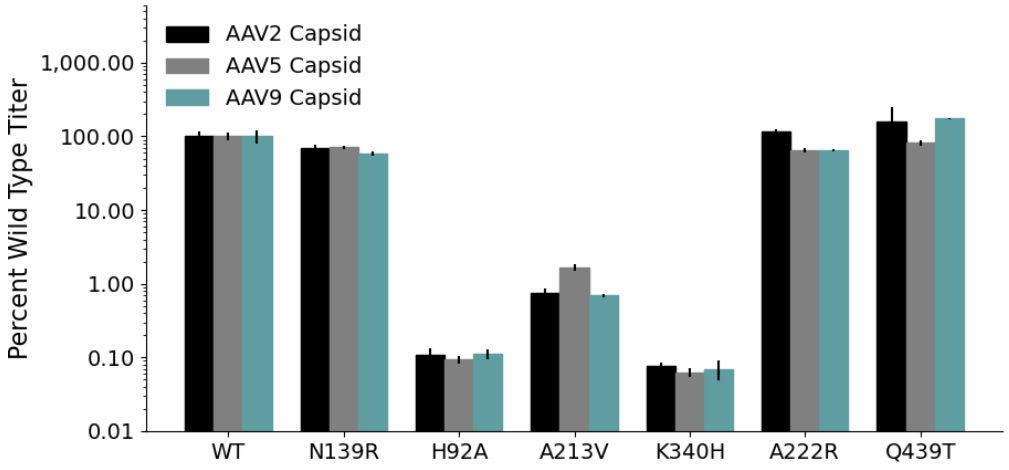


Figure A.15: DNase-resistant particle titers for AAV2, AAV5, and AAV9 capsids produced individually with the indicated Rep variants.

References

[1] 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. *Nature Communications*, 5(1), 3075. Number: 1 Publisher: Nature Publishing Group.

[2] Amoasii, L., Long, C., Li, H., Mireault, A. A., Shelton, J. M., Sanchez-Ortiz, E., McAnally, J. R., Bhattacharyya, S., Schmidt, F., Grimm, D., Hauschka, S. D., Bassel-Duby, R., & Olson, E. N. (2017). Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. *Science Translational Medicine*, 9(418), eaan8081. Publisher: American Association for the Advancement of Science.

[3] Atchison, R. W., Casto, B. C., & Hammon, W. M. (1965). Adenovirus-Associated Defective Virus Particles. *Science*, 149(3685), 754–756. Publisher: American Association for the Advancement of Science.

[4] Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., Opperman, D. J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M. K., Dalwadi, U., Yip, C. K., Burke, J. E., Garcia, K. C., Grishin, N. V., Adams, P. D., Read, R. J., & Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557), 871–876. Publisher: American Association for the Advancement of Science.

[5] Bennett, A., Mietzsch, M., & Agbandje-McKenna, M. (2017). Understanding capsid assembly and genome packaging for adeno-associated viruses. *Future virology*, 12(6), 283–297.

[6] Berry, G. E. & Asokan, A. (2016). Cellular transduction mechanisms of adeno-associated viral vectors. *Current Opinion in Virology*, 21, 54–60.

[7] Berthet, C., Raj, K., Saudan, P., & Beard, P. (2005). How adeno-associated virus Rep78 protein arrests cells completely in S phase. *Proceedings of the National Academy of Sciences*, 102(38), 13634–13639. Publisher: Proceedings of the National Academy of Sciences.

[8] Bleker, S., Pawlita, M., & Kleinschmidt, J. A. (2006). Impact of capsid conformation and Rep-capsid interactions on adeno-associated virus type 2 genome packaging. *Journal of Virology*, 80(2), 810–820.

[9] Bleker, S., Sonntag, F., & Kleinschmidt, J. A. (2005). Mutational analysis of narrow pores at the fivefold symmetry axes of adeno-associated virus type 2 capsids reveals a dual role in genome packaging and activation of phospholipase A2 activity. *Journal of Virology*, 79(4), 2528–2540.

[10] Bohlen, M. O., McCown, T. J., Powell, S. K., El-Nahal, H. G., Daw, T., Basso, M. A., Sommer, M. A., & Samulski, R. J. (2020). Adeno-Associated Virus Capsid-Promoter Interactions in the Brain Translate from Rat to the Nonhuman Primate. *Human Gene Therapy*, 31(21-22), 1155–1168. Publisher: Mary Ann Liebert, Inc., publishers.

[11] Brister, J. R. & Muzyczka, N. (1999). Rep-Mediated Nicking of the Adeno-Associated Virus Origin Requires Two Biochemical Activities, DNA Helicase Activity and Transesterification. *Journal of Virology*, 73(11), 9325–9336. Publisher: American Society for Microbiology.

[12] Brown, H. C., Zakas, P. M., George, S. N., Parker, E. T., Spencer, H. T., & Doering, C. B. (2018). Target-Cell-Directed Bioengineering Approaches for Gene Therapy of Hemophilia A. *Molecular Therapy - Methods & Clinical Development*, 9, 57–69.

[13] Bryant, D. H., Bashir, A., Sinai, S., Jain, N. K., Ogden, P. J., Riley, P. F., Church, G. M., Colwell, L. J., & Kelsic, E. D. (2021). Deep diversification of an AAV capsid protein by machine learning. *Nature Biotechnology*, 39(6), 691–696. Number: 6 Publisher: Nature Publishing Group.

[14] Burns, C. C., Shaw, J., Campagnoli, R., Jorba, J., Vincent, A., Quay, J., & Kew, O. (2006). Modulation of Poliovirus Replicative Fitness in HeLa Cells by Deoptimization of Synonymous Codon Usage in the Capsid Region. *Journal of Virology*, 80(7), 3259–3272. Publisher: American Society for Microbiology.

[15] Cao, L., During, M., & Xiao, W. (2002). Replication competent helper functions for recombinant AAV vector generation. *Gene Therapy*, 9(18), 1199–1206. Number: 18 Publisher: Nature Publishing Group.

[16] Cassell, G. D. & Weitzman, M. D. (2004). Characterization of a nuclear localization signal in the C-terminus of the adeno-associated virus Rep68/78 proteins. *Virology*, 327(2), 206–214.

[17] Cassinotti, P., Weitzand, M., & Tratschin, J.-D. (1988). Organization of the adeno-associated virus (AAV) capsid gene: Mapping of a minor spliced mRNA coding for virus capsid protein. *Virology*, 167(1), 176–184.

[18] Chadeuf, G., Favre, D., Tessier, J., Provost, N., Nony, P., Kleinschmidt, J., Moullier, P., & Salvetti, A. (2000). Efficient recombinant adeno-associated virus production by a stable rep-cap HeLa cell line correlates with adenovirus-induced amplification of the integrated rep-cap genome. *The Journal of Gene Medicine*, 2(4), 260–268. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1521-2254%28200007/08%292%3A4%3C260%3A%3AAID-JGM111%3E3.0.CO%3B2-8.

[19] Chaffiol, A., Caplette, R., Jaillard, C., Brazhnikova, E., Desrosiers, M., Dubus, E., Duhamel, L., Macé, E., Marre, O., Benoit, P., Hantraye, P., Bemelmans, A.-P., Bamberg, E., Duebel, J., Sahel, J.-A., Picaud, S., & Dalkara, D. (2017). A New Promoter Allows Optogenetic Vision Restoration with Enhanced Sensitivity in Macaque Retina. *Molecular Therapy*, 25(11), 2546–2560. Publisher: Elsevier.

[20] Chang, L. S., Shi, Y., & Shenk, T. (1989). Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. *Journal of Virology*, 63(8), 3479–3488. Publisher: American Society for Microbiology.

[21] Chejanovsky, N. & Carter, B. J. (1989). Mutagenesis of an AUG codon in the adeno-associated virus rep gene: Effects on viral DNA replication. *Virology*, 173(1), 120–128.

[22] Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., Schneider, R. G., Senior, A. W., Jumper, J., Hassabis, D., Kohli, P., & Avsec, ♦. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*, 381(6664), eadg7492. Publisher: American Association for the Advancement of Science.

[23] Cheng, M., Dietz, L., Gong, Y., Eichler, F., Nammour, J., Ng, C., Grimm, D., & Maguire, C. A. (2021). Neutralizing Antibody Evasion and Transduction with Purified Extracellular Vesicle-Enveloped Adeno-Associated Virus Vectors. *Human Gene Therapy*, 32(23-24), 1457–1470. Publisher: Mary Ann Liebert, Inc., publishers.

[24] Chuah, M. K., Petrus, I., De Bleser, P., Le Guiner, C., Gernoux, G., Adjali, O., Nair, N., Willems, J., Evens, H., Rincon, M. Y., Matrai, J., Di Matteo, M., Samara-Kuko, E., Yan, B., Acosta-Sanchez, A., Meliani, A., Cherel, G., Blouin, V., Christophe, O., Moullier, P., Mingozzi, F., & VandenDriessche, T. (2014). Liver-Specific Transcriptional Modules Identified by Genome-Wide In Silico Analysis Enable Efficient Gene Therapy in Mice and Non-Human Primates. *Molecular Therapy*, 22(9), 1605–1613.

[25] Chulay, J. D., Ye, G.-J., Thomas, D. L., Knop, D. R., Benson, J. M., Hutt, J. A., Wang, G., Humphries, M., & Flotte, T. R. (2011). Preclinical Evaluation of a Recombinant Adeno-Associated Virus Vector Expressing Human Alpha-1 Antitrypsin Made Using a Recombinant Herpes Simplex Virus Production Method. *Human Gene Therapy*, 22(2), 155–165. Publisher: Mary Ann Liebert, Inc., publishers.

[26] Clément, N. & Grieger, J. C. (2016). Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Molecular Therapy - Methods & Clinical Development*, 3, 16002.

[27] Colella, P., Sellier, P., Verdera, H. C., Puzzo, F., Wittenberghe, L. v., Guerchet, N., Daniele, N., Gjata, B., Marmier, S., Charles, S., Sola, M. S., Ragone, I., Leborgne, C., Collaud, F., & Mingozzi, F. (2019). AAV Gene Transfer with Tandem Promoter Design Prevents Anti-transgene Immunity and Provides Persistent Efficacy in Neonate Pompe Mice. *Molecular Therapy - Methods & Clinical Development*, 12, 85–101. Publisher: Elsevier.

[28] Coleman, J. R., Papamichail, D., Skiena, S., Futcher, B., Wimmer, E., & Mueller, S. (2008). Virus Attenuation by Genome-Scale Changes in Codon Pair Bias. *Science*, 320(5884), 1784–1787. Publisher: American Association for the Advancement of Science.

[29] Conway, J. E., Zolotukhin, S., Muzyczka, N., Hayward, G. S., & Byrne, B. J. (1997). Recombinant adeno-associated virus type 2 replication and packaging is entirely supported by a herpes simplex virus type 1 amplicon expressing Rep and Cap. *Journal of Virology*, 71(11), 8780–8789. Publisher: American Society for Microbiology.

[30] Davis, M. D., Wu, J., & Owens, R. A. (2000). Mutational Analysis of Adeno-Associated Virus Type 2 Rep68 Protein Endonuclease Activity on Partially Single-Stranded Substrates. *Journal of Virology*, 74(6), 2936–2942.

[31] Di Pasquale, G. & Stacey, S. N. (1998). Adeno-Associated Virus Rep78 Protein Interacts with Protein Kinase A and Its Homolog PRKX and Inhibits CREB-Dependent Transcriptional Activation. *Journal of Virology*, 72(10), 7916–7925. Publisher: American Society for Microbiology.

[32] Duan, D., Sharma, P., Yang, J., Yue, Y., Dudus, L., Zhang, Y., Fisher, K. J., & Engelhardt, J. F. (1998). Circular Intermediates of Recombinant Adeno-Associated Virus Have Defined Structural Characteristics Responsible for Long-Term Episomal Persistence in Muscle Tissue. *Journal of Virology*, 72(11), 8568–8577. Publisher: American Society for Microbiology.

[33] Earley, L. F., Kawano, Y., Adachi, K., Sun, X.-X., Dai, M.-S., & Nakai, H. (2015). Identification and Characterization of Nuclear and Nucleolar Localization Signals in the Adeno-Associated Virus Serotype 2 Assembly-Activating Protein. *Journal of Virology*, 89(6), 3038–3048. Publisher: American Society for Microbiology.

[34] Earley, L. F., Powers, J. M., Adachi, K., Baumgart, J. T., Meyer, N. L., Xie, Q., Chapman, M. S., & Nakai, H. (2017). Adeno-associated Virus (AAV) Assembly-Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5, and 11. *Journal of Virology*, 91(3), 10.1128/jvi.01980–16. Publisher: American Society for Microbiology.

[35] El Andari, J. & Grimm, D. (2021). Production, Processing, and Characterization of Synthetic AAV Gene Therapy Vectors. *Biotechnology Journal*, 16(1), 2000025. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/biot.202000025.

[36] Elmore, Z. C., Patrick Havlik, L., Oh, D. K., Anderson, L., Daaboul, G., Devlin, G. W., Vincent, H. A., & Asokan, A. (2021). The membrane associated accessory protein is an adeno-associated viral egress factor. *Nature Communications*, 12(1), 6239. Number: 1 Publisher: Nature Publishing Group.

[37] Faust, S. M., Bell, P., Cutler, B. J., Ashley, S. N., Zhu, Y., Rabinowitz, J. E., & Wilson, J. M. (2013). CpG-depleted adeno-associated virus vectors evade immune detection. *The Journal of Clinical Investigation*, 123(7), 2994–3001.

[38] Flotte, T. R., Solow, R., Owens, R. A., Afione, S., Zeitlin, P. L., & Carter, B. J. (1992). Gene Expression from Adeno-associated Virus Vectors in Airway Epithelial Cells. *American Journal of Respiratory Cell and Molecular Biology*, 7(3), 349–356. Publisher: American Thoracic Society - AJRCMB.

[39] Gaudet, D., Méthot, J., Déry, S., Brisson, D., Essiembre, C., Tremblay, G., Tremblay, K., de Wal, J., Twisk, J., van den Bulk, N., Sier-Ferreira, V., & van Deventer, S. (2013). Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPLS447X) gene therapy for lipoprotein lipase deficiency: an open-label trial. *Gene Therapy*, 20(4), 361–369. Number: 4 Publisher: Nature Publishing Group.

[40] Gavin, D. K., Young, S. M., Xiao, W., Temple, B., Abernathy, C. R., Pereira, D. J., Muzyczka, N., & Samulski, R. J. (1999). Charge-to-Alanine Mutagenesis of the Adeno-Associated Virus Type 2 Rep78/68 Proteins Yields Temperature-Sensitive and Magnesium-Dependent Variants. *Journal of Virology*, 73(11), 9433–9445. Publisher: American Society for Microbiology.

[41] Gonzalez-Sandoval, A., Pekrun, K., Tsuji, S., Zhang, F., Hung, K. L., Chang, H. Y., & Kay, M. A. (2023). The AAV capsid can influence the epigenetic marking of rAAV delivered episomal genomes in a species dependent manner. *Nature Communications*, 14(1), 2448. Number: 1 Publisher: Nature Publishing Group.

[42] Gorziglia, M. I., Lapcevich, C., Roy, S., Kang, Q., Kadan, M., Wu, V., Pechan, P., & Kaleko, M. (1999). Generation of an Adenovirus Vector Lacking E1, E2a, E3, and All of E4 except Open Reading Frame 3. *Journal of Virology*, 73(7), 6048–6055. Publisher: American Society for Microbiology.

[43] Greig, J. A., Wang, Q., Reicherter, A. L., Chen, S.-J., Hanlon, A. L., Tipper, C. H., Clark, K. R., Wadsworth, S., Wang, L., & Wilson, J. M. (2017). Characterization of Adeno-Associated Viral Vector-Mediated Human Factor VIII Gene Therapy in Hemophilia A Mice. *Human Gene Therapy*, 28(5), 392–402. Publisher: Mary Ann Liebert, Inc., publishers.

[44] Grieger, J. C., Snowdy, S., & Samulski, R. J. (2006). Separate Basic Region Motifs within the Adeno-Associated Virus Capsid Proteins Are Essential for Infectivity and Assembly. *Journal of Virology*, 80(11), 5199–5210. Publisher: American Society for Microbiology.

[45] Grimm, D., Kern, A., Pawlita, M., Ferrari, F. K., Samulski, R. J., & Kleinschmidt, J. A. (1999). Titration of AAV-2 particles via a novel capsid ELISA: packaging of genomes can limit production of recombinant AAV-2. *Gene Therapy*, 6(7), 1322–1330. Number: 7 Publisher: Nature Publishing Group.

[46] Grimm, D., Lee, J. S., Wang, L., Desai, T., Akache, B., Storm, T. A., & Kay, M. A. (2008). In Vitro and In Vivo Gene Therapy Vector Evolution via Multispecies Interbreeding and Retargeting of Adeno-Associated Viruses. *Journal of Virology*, 82(12), 5887–5911. Publisher: American Society for Microbiology.

[47] Guimaro, M. C., Afione, S. A., Tanaka, T., & Chiorini, J. A. (2020). Rescue of Adeno-Associated Virus Production by shRNA Cotransfection. *Human Gene Therapy*, 31(19-20), 1068–1073. Publisher: Mary Ann Liebert, Inc., publishers.

[48] Hansen, J., Qing, K., Kwon, H.-J., Mah, C., & Srivastava, A. (2000). Impaired Intracellular Trafficking of Adeno-Associated Virus Type 2 Vectors Limits Efficient Transduction of Murine Fibroblasts. *Journal of Virology*, 74(2), 992–996. Publisher: American Society for Microbiology.

[49] Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K., & Akira, S. (2000). A Toll-like receptor recognizes bacterial DNA. *Nature*, 408(6813), 740–745. Number: 6813 Publisher: Nature Publishing Group.

[50] Hickman, A. B., Ronning, D. R., Kotin, R. M., & Dyda, F. (2002). Structural Unity among Viral Origin Binding Proteins: Crystal Structure of the Nuclease Domain of Adeno-Associated Virus Rep. *Molecular Cell*, 10(2), 327–337. Publisher: Elsevier.

[51] Hickman, A. B., Ronning, D. R., Perez, Z. N., Kotin, R. M., & Dyda, F. (2004). The Nuclease Domain of Adeno-Associated Virus Rep Coordinates Replication Initiation Using Two Distinct DNA Recognition Interfaces. *Molecular Cell*, 13(3), 403–414. Publisher: Elsevier.

[52] Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., & Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie / Chemical Monthly*, 125(2), 167–188.

[53] Hong, G., Ward, P., & Berns, K. I. (1992). In vitro replication of adeno-associated virus DNA. *Proceedings of the National Academy of Sciences*, 89(10), 4673–4677. Publisher: Proceedings of the National Academy of Sciences.

[54] Horowitz, E. D., Rahman, K. S., Bower, B. D., Dismuke, D. J., Falvo, M. R., Griffith, J. D., Harvey, S. C., & Asokan, A. (2013). Biophysical and Ultrastructural Characterization of Adeno-Associated Virus Capsid Uncoating and Genome Release. *Journal of Virology*, 87(6), 2994–3002. Publisher: American Society for Microbiology.

[55] Huang, W., Liu, X., Queen, N. J., & Cao, L. (2017). Targeting Visceral Fat by Intraperitoneal Delivery of Novel AAV Serotype Vector Restricting Off-Target Transduction in Liver. *Molecular Therapy - Methods & Clinical Development*, 6, 68–78. Publisher: Elsevier.

[56] Hölscher, C., Kleinschmidt, J. A., & Bürkle, A. (1995). High-level expression of adeno-associated virus (AAV) Rep78 or Rep68 protein is sufficient for infectious-particle formation by a rep-negative AAV mutant. *Journal of Virology*, 69(11), 6880–6885. Publisher: American Society for Microbiology.

[57] Hörer, M., Weger, S., Butz, K., Hoppe-Seyler, F., Geisen, C., & Kleinschmidt, J. A. (1995). Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters. *Journal of Virology*, 69(9), 5485–5496. Publisher: American Society for Microbiology.

[58] Im, D.-S. & Muzyczka, N. (1990). The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. *Cell*, 61(3), 447–457. Publisher: Elsevier.

[59] Jalšić, L., Lytvyn, V., Elahi, S. M., Hrapovic, S., Nassoury, N., Chahal, P. S., Gaillet, B., & Gilbert, R. (2023). Inducible HEK293 AAV packaging cell lines expressing Rep proteins. *Molecular Therapy - Methods & Clinical Development*, 30, 259–275. Publisher: Elsevier.

[60] James, J. A., Escalante, C. R., Yoon-Robarts, M., Edwards, T. A., Linden, R. M., & Aggarwal, A. K. (2003). Crystal Structure of the SF3 Helicase from Adeno-Associated Virus Type 2. *Structure*, 11(8), 1025–1035. Publisher: Elsevier.

[61] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. Number: 7873 Publisher: Nature Publishing Group.

[62] King, J. A., Dubielzig, R., Grimm, D., & Kleinschmidt, J. A. (2001). DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *The EMBO Journal*, 20(12), 3282–3291.

[63] Kunec, D. & Osterrieder, N. (2016). Codon Pair Bias Is a Direct Consequence of Dinucleotide Bias. *Cell Reports*, 14(1), 55–67. Publisher: Elsevier.

[64] Labow, M. A., Hermonat, P. L., & Berns, K. I. (1986). Positive and negative autoregulation of the adeno-associated virus type 2 genome. *Journal of Virology*, 60(1), 251–258.

[65] Lee, J. M., Huddleston, J., Doud, M. B., Hooper, K. A., Wu, N. C., Bedford, T., & Bloom, J. D. (2018). Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proceedings of the National Academy of Sciences*, 115(35), E8276–E8285. Publisher: Proceedings of the National Academy of Sciences.

[66] Li, C., Diprimio, N., Bowles, D. E., Hirsch, M. L., Monahan, P. E., Asokan, A., Rabinowitz, J., Agbandje-McKenna, M., & Samulski, R. J. (2012). Single Amino Acid Modification of Adeno-Associated Virus Capsid Changes Transduction and Humoral Immune Profiles. *Journal of Virology*, 86(15), 7752–7759. Publisher: American Society for Microbiology.

[67] Liu, L. (2014). Fields Virology, 6th Edition. *Clinical Infectious Diseases*, 59(4), 613.

[68] Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., & Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, 6(1), 26.

[69] Louis, N., Evelegh, C., & Graham, F. L. (1997). Cloning and Sequencing of the Cellular–Viral Junctions from the Human Adenovirus Type 5 Transformed 293 Cell Line. *Virology*, 233(2), 423–429.

[70] Martínez, M. A., Jordan-Paiz, A., Franco, S., & Nevot, M. (2016). Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness. *Trends in Microbiology*, 24(2), 134–147. Publisher: Elsevier.

[71] Maurer, A. C., Benyamini, B., Fan, V. B., Whitney, O. N., Dailey, G. M., Darzacq, X., Weitzman, M. D., & Tjian, R. (2023). Double-Strand Break Repair Pathways Differentially Affect Processing and Transduction by Dual AAV Vectors. Pages: 2023.09.19.558438 Section: New Results.

[72] Maurer, A. C., Pacouret, S., Cepeda Diaz, A. K., Blake, J., Andres-Mateos, E., & Vandenberghe, L. H. (2018). The Assembly-Activating Protein Promotes Stability and Interactions between AAV's Viral Proteins to Nucleate Capsid Assembly. *Cell Reports*, 23(6), 1817–1830.

[73] Maurer, A. C. & Weitzman, M. D. (2020). Adeno-Associated Virus Genome Interactions Important for Vector Production and Transduction. *Human Gene Therapy*, 31(9-10), 499–511. Publisher: Mary Ann Liebert, Inc., publishers.

[74] McCarty, D. M., Monahan, P. E., & Samulski, R. J. (2001). Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene Therapy*, 8(16), 1248–1254. Number: 16 Publisher: Nature Publishing Group.

[75] McCarty, D. M., Young, S. M., & Samulski, R. J. (2004). Integration of Adeno-Associated Virus (AAV) and Recombinant AAV Vectors. *Annual Review of Genetics*, 38(1), 819–845. _eprint: https://doi.org/10.1146/annurev.genet.37.110801.143717.

[76] McCown, T. J., Xiao, X., Li, J., Breese, G. R., & Jude Samulski, R. (1996). Differential and persistent expression patterns of CNS gene transfer by an adeno-associated virus (AAV) vector. *Brain Research*, 713(1), 99–107.

[77] McLaughlin, S. K., Collis, P., Hermonat, P. L., & Muzyczka, N. (1988). Adeno-associated virus general transduction vectors: analysis of proviral structures. *Journal of Virology*, 62(6), 1963–1973. Publisher: American Society for Microbiology.

[78] Mietzsch, M., Eddington, C., Jose, A., Hsi, J., Chipman, P., Henley, T., Choudhry, M., McKenna, R., & Agbandje-McKenna, M. (2021). Improved Genome Packaging Efficiency of Adeno-associated Virus Vectors Using Rep Hybrids. *Journal of Virology*, 95(19), e00773–21. Publisher: American Society for Microbiology.

[79] Mietzsch, M., Smith, J. K., Yu, J. C., Banala, V., Emmanuel, S. N., Jose, A., Chipman, P., Bhattacharya, N., McKenna, R., & Agbandje-McKenna, M. (2020). Characterization of AAV-Specific Affinity Ligands: Consequences for Vector Purification and Development Strategies. *Molecular Therapy - Methods & Clinical Development*, 19, 362–373.

[80] Mueller, S., Coleman, J. R., Papamichail, D., Ward, C. B., Nimnual, A., Futcher, B., Skiena, S., & Wimmer, E. (2010). Live attenuated influenza virus vaccines by computer-aided rational design. *Nature Biotechnology*, 28(7), 723–726. Number: 7 Publisher: Nature Publishing Group.

[81] Murphy, M., Gomos-Klein, J., Stankic, M., & Falck-Pedersen, E. (2007). Adeno-Associated Virus Type 2 p5 Promoter: a Rep-Regulated DNA Switch Element Functioning in Transcription, Replication, and Site-Specific Integration. *Journal of Virology*, 81(8), 3721–3730. Publisher: American Society for Microbiology.

[82] Musayev, F. N., Zarate-Perez, F., Bardelli, M., Bishop, C., Saniev, E. F., Linden, R. M., Henckaerts, E., & Escalante, C. R. (2015). Structural Studies of AAV2 Rep68 Reveal a Partially Structured Linker and Compact Domain Conformation. *Biochemistry*, 54(38), 5907–5919. Publisher: American Chemical Society.

[83] Nicolson, S. C. & Samulski, R. J. (2014). Recombinant Adeno-Associated Virus Utilizes Host Cell Nuclear Import Machinery To Enter the Nucleus. *Journal of Virology*, 88(8), 4132–4144. Publisher: American Society for Microbiology.

[84] Nowrouzi, A., Penaud-Budloo, M., Kaeppel, C., Appelt, U., Guiner, C. L., Moullier, P., Kalle, C. v., Snyder, R. O., & Schmidt, M. (2012). Integration Frequency and Intermolecular Recombination of rAAV Vectors in Non-human Primate Skeletal Muscle and Liver. *Molecular Therapy*, 20(6), 1177–1186. Publisher: Elsevier.

[85] Ogden, P. J., Kelsic, E. D., Sinai, S., & Church, G. M. (2019). Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science*, 366(6469), 1139–1143. Publisher: American Association for the Advancement of Science.

[86] Ozelo, M. C., Mahlangu, J., Pasi, K. J., Giermasz, A., Leavitt, A. D., Laffan, M., Symington, E., Quon, D. V., Wang, J.-D., Peerlinck, K., Pipe, S. W., Madan, B., Key, N. S., Pierce, G. F., O'Mahony, B., Kaczmarek, R., Henshaw, J., Lawal, A., Jayaram, K., Huang, M., Yang, X., Wong, W. Y., & Kim, B. (2022). Valoctocogene Roxaparvovec Gene Therapy for Hemophilia A. *New England Journal of Medicine*, 386(11), 1013–1025. Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa2113708.

[87] Paulk, N. K., Pekrun, K., Zhu, E., Nygaard, S., Li, B., Xu, J., Chu, K., Leborgne, C., Dane, A. P., Haft, A., Zhang, Y., Zhang, F., Morton, C., Valentine, M. B., Davidoff, A. M., Nathwani, A. C., Mingozzi, F., Grompe, M., Alexander, I. E., Lisowski, L., & Kay, M. A. (2018). Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. *Molecular Therapy*, 26(1), 289–303.

[88] Pekrun, K., Alencastro, G. D., Luo, Q.-J., Liu, J., Kim, Y., Nygaard, S., Galivo, F., Zhang, F., Song, R., Tiffany, M. R., Xu, J., Hebrok, M., Grompe, M., & Kay, M. A. (2019). Using a barcoded AAV capsid library to select for clinically relevant gene therapy vectors. *JCI Insight*, 4(22). Publisher: American Society for Clinical Investigation.

[89] Pereira, D. J., McCarty, D. M., & 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. *Journal of Virology*, 71(2), 1079–1088. Publisher: American Society for Microbiology.

[90] Pipe, S. W., Leebeek, F. W., Recht, M., Key, N. S., Lattimore, S., Castaman, G., Coppens, M., Cooper, D., Gut, R., Slawka, S., Verweij, S., Dolmetsch, R., Li, Y., Monahan, P. E., & Miesbach, W. A. (2022). Adults with Severe or Moderately Severe Hemophilia B Receiving Etranacogene Dezaparvovec in the HOPE-B Phase 3 Clinical Trial Continue to Experience a Stable Increase in Mean Factor IX Activity Levels and Durable Hemostatic Protection after 24 Months' Follow-up. *Blood*, 140(Supplement 1), 4910–4912.

[91] Popa-Wagner, R., Porwal, M., Kann, M., Reuss, M., Weimer, M., Florin, L., & Kleinschmidt, J. A. (2012). Impact of VP1-Specific Protein Sequence Motifs on Adeno-Associated Virus Type 2 Intracellular Trafficking and Nuclear Entry. *Journal of Virology*, 86(17), 9163–9174. Publisher: American Society for Microbiology.

[92] Rav, S., Sharma, T., Yip, M., Gao, G., & Tail, P. W. (2023). Improving codon optimization for gene therapy vectors using deep learning. Presentation at the American Society of Gene and Cell Therapy Meeting, Washington, DC.

[93] Robert, M.-A., Chahal, P. S., Audy, A., Kamen, A., Gilbert, R., & Gaillet, B. (2017). Manufacturing of recombinant adeno-associated viruses using mammalian expression platforms. *Biotechnology Journal*, 12(3), 1600193. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/biot.201600193.

[94] Rumachik, N. G., Malaker, S. A., Poweleit, N., Maynard, L. H., Adams, C. M., Leib, R. D., Cirolia, G., Thomas, D., Stamnes, S., Holt, K., Sinn, P., May, A. P., & Paulk, N. K. (2020). Methods Matter: Standard Production Platforms for Recombinant AAV Produce Chemically and Functionally Distinct Vectors. *Molecular Therapy - Methods & Clinical Development*, 18, 98–118. Publisher: Elsevier.

[95] Russell, S., Bennett, J., Wellman, J. A., Chung, D. C., Yu, Z.-F., Tillman, A., Wittes, J., Pappas, J., Elci, O., McCague, S., Cross, D., Marshall, K. A., Walshire, J., Kehoe, T. L., Reichert, H., Davis, M., Raffini, L., George, L. A., Hudson, F. P., Dingfield, L., Zhu, X., Haller, J. A., Sohn, E. H., Mahajan, V. B., Pfeifer, W., Weckmann, M., Johnson, C., Gewaily, D., Drack, A., Stone, E., Wachtel, K., Simonelli, F., Leroy, B. P., Wright, J. F., High, K. A., & Maguire, A. M. (2017). Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *The Lancet*, 390(10097), 849–860.

[96] Sairavi, A., Cole, J., Micheletti, C., Gu, B., Adachi, K., & Nakai, H. (2022). Genome packaging efficiencies of Anc80 and AAV9 vectors using non-canonical Rep-ITR combinations. Presentation at the American Society of Gene and Cell Therapy Meeting, Washington, DC.

[97] Samulski, R. J., Chang, L. S., & Shenk, T. (1989). Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. *Journal of Virology*, 63(9), 3822–3828. Publisher: American Society for Microbiology.

[98] Samulski, R. J. & Muzyczka, N. (2014). AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annual Review of Virology*, 1(1), 427–451.

[99] Santosh, V., Musayev, F. N., Jaiswal, R., Zárate-Pérez, F., Vandewinkel, B., Dierckx, C., Endicott, M., Sharifi, K., Dryden, K., Henckaerts, E., & Escalante, C. R. (2020). The Cryo-EM structure of AAV2 Rep68 in complex with ssDNA reveals a malleable AAA+ machine that can switch between oligomeric states. *Nucleic Acids Research*, 48(22), 12983–12999.

[100] Selot, R., Arumugam, S., Mary, B., Cheemadan, S., & Jayandharan, G. R. (2017). Optimized AAV rh.10 Vectors That Partially Evade Neutralizing Antibodies during Hepatic Gene Transfer. *Frontiers in Pharmacology*, 8, 441.

[101] Sitaraman, V., Hearing, P., Ward, C. B., Gnatenko, D. V., Wimmer, E., Mueller, S., Skiena, S., & Bahou, W. F. (2011). Computationally designed adeno-associated virus (AAV) Rep 78 is efficiently maintained within an adenovirus vector. *Proceedings of the National Academy of Sciences*, 108(34), 14294–14299. Publisher: Proceedings of the National Academy of Sciences.

[102] Smith, R. H. & Kotin, R. M. (1998). The Rep52 Gene Product of Adeno-Associated Virus Is a DNA Helicase with 3′-to-5′ Polarity. *Journal of Virology*, 72(6), 4874–4881. Publisher: American Society for Microbiology.

[103] Snyder, R. O., Im, D. S., & Muzyczka, N. (1990). Evidence for covalent attachment of the adeno-associated virus (AAV) rep protein to the ends of the AAV genome. *Journal of Virology*, 64(12), 6204–6213. Publisher: American Society for Microbiology.

[104] Song, L., Samulski, R. J., & Hirsch, M. L. (2020). Adeno-Associated Virus Vector Mobilization, Risk Versus Reality. *Human Gene Therapy*, 31(19-20), 1054–1067. Publisher: Mary Ann Liebert, Inc., publishers.

[105] Song, Y., Liu, Y., Ward, C. B., Mueller, S., Futcher, B., Skiena, S., Paul, A. V., & Wimmer, E. (2012). Identification of two functionally redundant RNA elements in the coding sequence of poliovirus using computer-generated design. *Proceedings of the National Academy of Sciences*, 109(36), 14301–14307. Publisher: Proceedings of the National Academy of Sciences.

[106] Sonntag, F., Bleker, S., Leuchs, B., Fischer, R., & Kleinschmidt, J. A. (2006). Adeno-Associated Virus Type 2 Capsids with Externalized VP1/VP2 Trafficking Domains Are Generated prior to Passage through the Cytoplasm and Are Maintained until Uncoating Occurs in the Nucleus. *Journal of Virology*, 80(22), 11040–11054. Publisher: American Society for Microbiology.

[107] Srivastava, A. (2016). In vivo tissue-tropism of adeno-associated viral vectors. *Current Opinion in Virology*, 21, 75–80.

[108] Srivastava, A., Lusby, E. W., & Berns, K. I. (1983). Nucleotide sequence and organization of the adeno-associated virus 2 genome. *Journal of Virology*, 45(2), 555–564. Publisher: American Society for Microbiology.

[109] Stahnke, S., Lux, K., Uhrig, S., Kreppel, F., Hösel, M., Coutelle, O., Ogris, M., Hallek, M., & Büning, H. (2011). Intrinsic phospholipase A2 activity of adeno-associated virus is involved in endosomal escape of incoming particles. *Virology*, 409(1), 77–83.

[110] Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., Navarro, M. J., Bowen, J. E., Tortorici, M. A., Walls, A. C., King, N. P., Veesler, D., & Bloom, J. D. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell*, 182(5), 1295–1310.e20.

[111] Strauss, K. A., Farrar, M. A., Muntoni, F., Saito, K., Mendell, J. R., Servais, L., McMillan, H. J., Finkel, R. S., Swoboda, K. J., Kwon, J. M., Zaidman, C. M., Chiriboga, C. A., Iannaccone, S. T., Krueger, J. M., Parsons, J. A., Shieh, P. B., Kavanagh, S., Wigderson, M., Tauscher-Wisniewski, S., McGill, B. E., & Macek, T. A. (2022). Onasemnogene abeparvovec for presymptomatic infants with three copies of SMN2 at risk for spinal muscular atrophy: the Phase III SPR1NT trial. *Nature Medicine*, 28(7), 1390–1397. Number: 7 Publisher: Nature Publishing Group.

[112] Stutika, C., Gogol-Döring, A., Botschen, L., Mietzsch, M., Weger, S., Feldkamp, M., Chen, W., & Heilbronn, R. (2016). A Comprehensive RNA Sequencing Analysis of the Adeno-Associated Virus (AAV) Type 2 Transcriptome Reveals Novel AAV Transcripts, Splice Variants, and Derived Proteins. *Journal of Virology*, 90(3), 1278–1289. Publisher: American Society for Microbiology.

[113] Su, Q., Sena-Esteves, M., & Gao, G. (2020). Purification of Recombinant Adeno-Associated Viruses (rAAVs) by Cesium Chloride Gradient Sedimentation. *Cold Spring Harbor Protocols*, 2020(8), pdb.prot095604. Publisher: Cold Spring Harbor Laboratory Press.

[114] Su, W., Patrício, M. I., Duffy, M. R., Krakowiak, J. M., Seymour, L. W., & Cawood, R. (2022). Self-attenuating adenovirus enables production of recombinant adeno-associated virus for high manufacturing yield without contamination. *Nature Communications*, 13(1), 1182. Number: 1 Publisher: Nature Publishing Group.

[115] Summerford, C. & Samulski, R. J. (1998). Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *Journal of Virology*, 72(2), 1438–1445.

[116] Surosky, R. T., Urabe, M., Godwin, S. G., McQuiston, S. A., Kurtzman, G. J., Ozawa, K., & Natsoulis, G. (1997). Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *Journal of Virology*, 71(10), 7951–7959. Publisher: American Society for Microbiology.

[117] Tabebordbar, M., Lagerborg, K. A., Stanton, A., King, E. M., Ye, S., Tellez, L., Krunnfusz, A., Tavakoli, S., Widrick, J. J., Messemer, K. A., Troiano, E. C., Moghadaszadeh, B., Peacker, B. L., Leacock, K. A., Horwitz, N., Beggs, A. H., Wagers, A. J., & Sabeti, P. C. (2021). Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. *Cell*, 184(19), 4919–4938.e22. Publisher: Elsevier.

[118] Takata, M. A., Gonçalves-Carneiro, D., Zang, T. M., Soll, S. J., York, A., Blanco-Melo, D., & Bieniasz, P. D. (2017). CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature*, 550(7674), 124–127. Number: 7674 Publisher: Nature Publishing Group.

[119] Tessier, J., Chadeuf, G., Nony, P., Avet-Loiseau, H., Moullier, P., & Salvetti, A. (2001). Characterization of Adenovirus-Induced Inverted Terminal Repeat-Independent Amplification of Integrated Adeno-Associated Virus rep-cap Sequences. *Journal of Virology*, 75(1), 375–383. Publisher: American Society for Microbiology.

[120] Tse, L. V., Klinc, K. A., Madigan, V. J., Castellanos Rivera, R. M., Wells, L. F., Havlik, L. P., Smith, J. K., Agbandje-McKenna, M., & Asokan, A. (2017). Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. *Proceedings of*

the National Academy of Sciences, 114(24), E4812–E4821. Publisher: Proceedings of the National Academy of Sciences.

[121] Urabe, M., Ding, C., & Kotin, R. M. (2002). Insect Cells as a Factory to Produce Adeno-Associated Virus Type 2 Vectors. *Human Gene Therapy*, 13(16), 1935–1943. Publisher: Mary Ann Liebert, Inc., publishers.

[122] Vandenberghe, L. H., Xiao, R., Lock, M., Lin, J., Korn, M., & Wilson, J. M. (2010). Efficient Serotype-Dependent Release of Functional Vector into the Culture Medium During Adeno-Associated Virus Manufacturing. *Human Gene Therapy*, 21(10), 1251–1257. Publisher: Mary Ann Liebert, Inc., publishers.

[123] Wang, B., Yang, C., Tekes, G., Mueller, S., Paul, A., Whelan, S. P. J., & Wimmer, E. (2015). Recoding of the Vesicular Stomatitis Virus L Gene by Computer-Aided Design Provides a Live, Attenuated Vaccine Candidate. *mBio*, 6(2), 10.1128/mbio.00237–15. Publisher: American Society for Microbiology.

[124] Wang, D., Tai, P. W. L., & Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. *Nature Reviews Drug Discovery*, 18(5), 358–378. Number: 5 Publisher: Nature Publishing Group.

[125] Wistuba, A., Kern, A., Weger, S., Grimm, D., & Kleinschmidt, J. A. (1997). Subcellular compartmentalization of adeno-associated virus type 2 assembly. *Journal of Virology*, 71(2), 1341–1352. Publisher: American Society for Microbiology.

[126] Wistuba, A., Weger, S., Kern, A., & Kleinschmidt, J. A. (1995). Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins. *Journal of Virology*, 69(9), 5311–5319. Publisher: American Society for Microbiology.

[127] Wu, J., Davis, M. D., & Owens, R. A. (1999). Factors Affecting the Terminal Resolution Site Endonuclease, Helicase, and ATPase Activities of Adeno-Associated Virus Type 2 Rep Proteins. *Journal of Virology*, 73(10), 8235–8244. Publisher: American Society for Microbiology.

[128] 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. *Journal of Virology*, 74(18), 8635–8647. Publisher: American Society for Microbiology.

[129] Wörner, T. P., Bennett, A., Habka, S., Snijder, J., Friese, O., Powers, T., Agbandje-McKenna, M., & Heck, A. J. R. (2021). Adeno-associated virus capsid assembly is divergent and stochastic. *Nature Communications*, 12(1), 1642. Number: 1 Publisher: Nature Publishing Group.

[130] Xiao, X., Li, J., & Samulski, R. J. (1998). Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus. *Journal of Virology*, 72(3), 2224–2232. Publisher: American Society for Microbiology.

[131] Xiao, X., Xiao, W., Li, J., & Samulski, R. J. (1997). A novel 165-base-pair terminal repeat sequence is the sole cis requirement for the adeno-associated virus life cycle. *Journal of Virology*, 71(2), 941–948. Publisher: American Society for Microbiology.

[132] Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A., & Chapman, M. S. (2002). The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proceedings of the National Academy of Sciences*, 99(16), 10405–10410. Publisher: Proceedings of the National Academy of Sciences.

[133] Yan, Z., Zak, R., Zhang, Y., & Engelhardt, J. F. (2005). Inverted Terminal Repeat Sequences Are Important for Intermolecular Recombination and Circularization of Adeno-Associated Virus Genomes. *Journal of Virology*, 79(1), 364–379. Publisher: American Society for Microbiology.

[134] Yang, J., Zhou, W., Zhang, Y., Zidon, T., Ritchie, T., & Engelhardt, J. F. (1999). Concatamerization of Adeno-Associated Virus Circular Genomes Occurs through Intermolecular Recombination. *Journal of Virology*, 73(11), 9468–9477. Publisher: American Society for Microbiology.

[135] Yang, Q., Kadam, A., & Trempe, J. P. (1992). Mutational analysis of the adeno-associated virus rep gene. *Journal of Virology*, 66(10), 6058–6069. Publisher: American Society for Microbiology.

[136] Zaidman, C. M., Proud, C. M., McDonald, C. M., Lehman, K. J., Goedeker, N. L., Mason, S., Murphy, A. P., Guridi, M., Wang, S., Reid, C., Darton, E., Wandel, C., Lewis, S., Malhotra, J., Griffin, D. A., Potter, R. A., Rodino-Klapac, L. R., & Mendell, J. R. (2023). Delandistrogene Moxeparvovec Gene Therapy in Ambulatory Patients (Aged ≥4 to <8 Years) with Duchenne Muscular Dystrophy: 1-Year Interim Results from Study SRP-9001-103 (ENDEAVOR). *Annals of Neurology*, 94(5), 955–968. Publisher: John Wiley & Sons, Ltd.

[137] Zarate-Perez, F., Bardelli, M., Burgner, J. W., Villamil-Jarauta, M., Das, K., Kekilli, D., Mansilla-Soto, J., Linden, R. M., & Escalante, C. R. (2012). The interdomain linker of AAV-2 Rep68 is an integral part of its oligomerization domain: role of a conserved SF3 helicase residue in oligomerization. *PLoS pathogens*, 8(6), e1002764.

[138] Zeltner, N., Kohlbrenner, E., Clément, N., Weber, T., & Linden, R. M. (2010). Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. *Gene Therapy*, 17(7), 872–879. Number: 7 Publisher: Nature Publishing Group.

[139] Zhu, J., Huang, X., & Yang, Y. (2009). The TLR9-MyD88 pathway is critical for adaptive immune responses to adeno-associated virus gene therapy vectors in mice. *The Journal of Clinical Investigation*, 119(8), 2388–2398.

[140] Zolotukhin, S., Byrne, B. J., Mason, E., Zolotukhin, I., Potter, M., Chesnut, K., Summerford, C., Samulski, R. J., & Muzyczka, N. (1999). Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Therapy*, 6(6), 973–985. Number: 6 Publisher: Nature Publishing Group.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont's humanist typeface. The above illustration, "Science Experiment 02", was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.