PROCESS DESIGN FOR SCALABLE RECOMBINANT ADENO-ASSOCIATED VIRUS

Tejash Vijay Patel

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Pharmaceutical Sciences in the Eshelman School of Pharmacy.

Chapel Hill
2017

Approved by:

Xiao Xiao

Leaf Huang

Shawn Hingtgen

Aravind Asokan

Rihe Liu

ABSTRACT

Tejash Vijay Patel: Process Design for Scalable Recombinant Adeno-Associated Virus
(Under the direction of Xiao Xiao)

The versatility of recombinant adeno-associated viruses has garnered significant attention from investors, pharmaceutical companies, and regulatory agencies as more therapies using this vector are finding success in preclinical and late-stage clinical trials, even some reaching the market approval stage in the US. A number of vector production methods have been developed to generate the necessary clinical grade vectors that have high potency, and have high-titers to reduce general manufacturing and therapeutic costs. However, each of these production methods inherently have their own drawbacks, whether it is the concomitantly produced helper vectors used to make the AAV or the quality of the AAV itself. We previously developed a HEK293-cell based producer cell line method for high titer and high potency rAAV vectors. However, this method requires significant time-intensive generation of novel producer cell lines for every rAAV vector of interest and furthermore the E1-, E3-deleted adenovirus is only occupied at the E1-region, leaving the space available in the E3-region unused. In this body of work, we discuss a novel adenovirus construct that utilizes a packaging cell line to generate a high titer and high quality rAAV, and attempts to remove the accompanying adenovirus in the final product batch. The improvements made in this system are: 1) the development of a one-step cloning of a rAAV vector cassette into the E1-region of the adenovirus, 2) use of an efficiently generated packaging cell line to universally package the rAAV vector provided in trans by the adenovirus, 3) high vector yields on different rAAV inverted terminal repeat designs

provided by the adenovirus, and 4) high quality, low. empty-particle containing rAAV product. We attempt to remove contaminating entities in the rAAV production method, using a selective precipitating agent called domiphen bromide. Although this detergent is efficient in removing contaminating materials such as DNA and adenovirus, there are major difficulties to reduce interactions with certain serotypes of AAV. Therefore, an alternative method for removing adenovirus is necessary, possibly by high hydrostatic pressure or nanofilters. The simple adenovirus construct coupled with the packaging cell line can be a pivotal method for large scale AAV vector production.

To my father, Vijay
To my mother, Shabda
To my brother, Hetav.
JSK, JSCA, and thank you.

ACKNOWLEDGEMENTS

There is a saying in Gujarati that describes my journey in this dissertation: "*Tipe tipe sarovar bandhay, kakre kakre paad bandhay*." It means, "Even the biggest oceans can be formed by bringing drops of water one a time. And the biggest of mountains can be made if we gather pebbles one at a time." The path to this conclusion of my dissertation work could not be possible without the support of so many individuals in my personal and professional life. Each of these individuals have brought me the support I needed to achieve this goal of mine. And I thank a higher power, i.e. God or Bhagwan, for bringing these people into my life and shaping me into the scientist and person I am today.

First, I would like to thank Dr. Xiao. It must have been a risky choice to mentor a student without any genetics, molecular biology, or biochemistry background, but I am very glad he took that risk on me. He believed in my ability, supported me in my time of need, and he challenged me with new ideas and ways of thinking for the lab, the AAV field, and my thesis work. I have learned so much from him and I am proud to have worked in his lab these past 5 or so years. Thank you.

I would like to send huge thanks for my parents, Vijay and Shabda, and my brother, Hetav (aka Motu) who I have dedicated this thesis to. There are probably a hundred million words I could type to describe how much I love and am thankful for, for their support, advice, love, and laughter while I finished this degree. It wasn't always easy and it wasn't always the greatest experience. But you are the team I would always want to be a part of. We are what Bo Schembechler wanted in 1983: "The Team. The Team. The Team."

To my lab mates: Dr. Chunping Qiao, Jianbin Li, Dr. Juan Li, Dr. Tao Bian, and Dr. Bin Xiao. You were some of the greatest teachers and individuals to talk about my thesis work with. It was always great to hear your perspectives on experiment design and theory of my thesis aims, but it was also great to have you as people to work with and share holidays, lunches, and stories with. Thank you all so much. To my colleagues who have been in the lab, old and new, I thank you for sharing a cup of tea and your company: Dr. Yi Dai, Dr. Hui Zheng, Dr. Zhenhua Yuan, Dr. Carrie Martin, Dr. Karen Bulaklak, Randolph (Randerson) Qian, and Quan Jin.

Thank you to my committee members. To my committee chair, Dr. Leaf Huang, for helping me get to this stage of my career. To Dr. Shawn Hingtgen, thank you for setting me up with your friend Haytham to help me in Aim 2, and thank you for discussing football with me. One day I will buy an Iowa shirt to remember you, but the recent losses are too much to bear to think about that. To Dr. Aravind Asokan, thank you for providing such great insight into the AAV field, whether it is through theoretical talks in our committee meetings or the outlook of the field currently. To Dr. Rihe Liu, thank you for providing a prospective to my science I hadn't considered before we had our committee meetings. Together you four were a great team to share my work with and constructively move my projects to where it is presented in this document.

I would like to thank a few of my funding sources that assisted me in my dissertation work. A huge thank you to American Foundation for Pharmaceutical Education (AFPE), where I was awarded the Lynn Van Campen Pre-Doctoral Fellowship in Manufacturing Science and Engineering in 2014, and was renewed in 2015. Thank you to the UNC Eshelman Institute for Innovation for partially funding Aim 2 of my dissertation work. Furthermore, I thank Dr. Xiao's NIH grants, particularly Project Number: 5R01NS082536-02, for assisting my graduate work.

Professionally, there are a few behind-the-scenes coworkers, or faculty members in the University of North Carolina that have assisted me quite a bit and deserve my thanks. To Chris Hill and Aaron Todd, thank you for helping me throughout the past 5 years of my dissertation work and lending an ear when I needed it. Thank you to Dr. Nancie M. Archin for providing an approachable faculty to discuss coping skills for all of science's and life's failures and experiences. You were also a great positive presence when I was working during weekends or on holiday breaks! Thank you also to the students of the Department of Pharmaceutical Sciences, especially Alison Mercer-Smith, Dr. Dylan Glatt-Dowd, Carla Coste Sanchez, Dr. Junghyun Kim, Holly Schroeder, Sam Fix, Dr. Kim Barnash, Dr. Paul Himes, Dr. Jarod Waybright, and Dr. Rachel Bleich. Thank you so much!

I would like to take a portion of my acknowledgements to thank the friends and family that have made my life experience these past 5 years more enjoyable. There are so many people to thank. To the unsung heroes in my life - Shaye Hagler, Dr. Kevin Sheets and his wife Cissy and Baby Tyler and Roxie, Dr. Erin Wilson, Phil Marek, Dr. Juli Bago and his wife Alicia and Sinatra, Mike Kawano and his (soon-to-be) wife Mariesol, Nathan Mills and his wife Bekkah and their son Baby Titus, Tiago Ruas, Deb and Dave Mills, the Krue Fam, The Bard Army (aka beyondtheguitar's fanbase), Dr. Maria Abad, Greg Ho, Dr. Félix Olivares, Dr. Chloe Ciarrochi, Glenn Wilkinson, Erin Stuelke, Bryan Walenta and his wife Jen, Rachel and Ken Kirchoff with Baby Bradley and Baby Eloise, Steve Su, Dr. Stephen Liu, Ammy Yuan, Emily Seidel, Eric Biesbrock, Stephen Morrsion and his (soon-to-be) wife Carly, Danielle Forsyth, Dr. Laura Marsh, Dr. Brandon Sheets, Adam Sheets, Dr. Brad Falk, Julia DiFiore, Dr. Monica Mann-Wrobel, and Steven Clemens– I probably would not be where I am today if it was not for your love and support. To my family – Nishith and Krutika and Baby Aashka, Bhavika Patel from California, Palaash and Roshni, Niraja, Jay and Priyanka, Ilamami and Dinumama, Megha,

Roshni, and Neel – thank you so much for helping me when my family and I needed help. Thank you to my extended family for their support along the way. It would be remiss of me to not mention some of the people I am so proud of in my life: my nieces. Anjali, Bhakti, and Dhara I am so thankful for the relationship we developed over the last several months and I am so excited to have this relationship grow as time goes on. I love you all so much.

PREFACE

Around the year 2006 or 2007, I was asked by my parents and counselors the common question of "What do you want to be when you grow up?". I had grown an affinity toward math and sciences and excelled in these subjects, and growing up about 30 minutes away from the motor vehicle hub of Detroit and Dearborn, Michigan, I had a feeling I would be an engineer. I just didn't know if I was cut out for the heavy-physics of mechanical engineering. And even though I was good at patterns, I didn't think I was that good at computers to become a programmer. While this was internal struggle was going on, by serendipity, I was watching a NOVA scienceNOW special on PBS and it was talking about RNA interference, this revolutionary molecule that led scientists to shut off specific genes and provide potential treatments for a wide range of diseases. And immediately I was hooked, asking myself "What degree would get me to a place in my life where I can make more of *that* drug to help more people?"

I obtained my chemical engineering degree from the University of Michigan in 2011, with hopes I could use the degree to get me into a position of scaling up pharmaceuticals. At the time I was in the program, it felt the degree was more tailored towards oil, food, and polymers rather than pharmaceuticals and biotechnologies. Even in the research I performed in some of the labs in the Department of Biomedical Engineering, I would find myself asking the graduate student I worked with "How far away is this to translatable applications?" It was a very theoretical project with inevitable real-world applications, but the inevitability was too far for me. I obtained an internship at Cayman Chemical the summer before I graduated and some of

the PhD's in the company noticed my meticulous work ethic and suggested that I pursue a PhD. I told them my goals of designing scalable methods of a therapeutic, and they had recommended I pursue a higher degree to achieve those goals and look at degrees in biomedical engineering, chemical engineering, or pharmaceutical sciences.

In 2012, when UNC flew me in for an interview at the Division of Molecular Pharmaceutics (now DPMP), I remember explaining my Nova scienceNOW story to Dr. Xiao at the time he interviewed me, and how I wanted to pursue scaling of therapeutics as a career choice. We talked for an engaging 60-minutes, 30 minutes longer than my allocated interview time with Dr. Xiao, and thereby completely missing my following interview with Dr. Mumper. In those 60-minutes, Dr. Xiao and I were discussing scalability of therapeutics, especially in his field, including a tour of his lab where we spent several minutes at the bioreactors and ultracentrifuges. After receiving an acceptance letter from the school, I had already been convinced on which lab I wanted to work for and I worked as hard as I could to have a good foundation for the work I would be doing in Dr. Xiao's lab.

At the time the Nova scienceNOW special, my acceptance into UNC and Dr. Xiao's lab, and the rough idea of my thesis work was very exciting and inspiring. But one other factor played a major role into why I am pursuing this career path with gusto: meeting someone with a rare genetic disorder and the implications it had on himself and his family members. It was at a wedding I attended in California in 2014, where I met someone who had what I guessed to be some slowly progressing muscular dystrophy, perhaps Becker's. In the few days I interacted with this person and his family I learned the impact of the disease: the quality of life for the individual with the disease and his caregivers, his elderly parents. The extended family, after learning of my research, requested of me to look for gene therapies or clinical trials for his form of dystrophy and if their family member could seek treatment. I felt helpless since I barely knew the type of

dystrophy he had, feeling all the more powerless by telling them that I knew of nothing to help cure or improve this individual, especially since the ClinicalTrials.gov showed nothing related to gene therapy treatments for his dystrophy. They had disclosed to me his symptoms were fatal and he had less than a decade to live.

I came back from that wedding and understood the humanity behind these therapeutics, and it gave me more reason to pursue this career. I felt very proud that my advisor, Dr. Xiao, has spent majority of his career studying and treating muscular dystrophies. Coming back from that experience also connected me to the greater picture Dr. Xiao discussed with me early in my dissertation work: how can we translate treatments for major dystrophies to clinical settings more efficiently? Or in other words, like what I have always wanted to do, how can we make more of *that* drug to help more people?

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAV	Adeno-Associated Virus
AAVS1	Adeno-Associated Virus Integration Site
AC	Affinity chromatography
Ad	Adenovirus
AEX	Anion Exchange Chromatography
BHK	Baby Hamster Kidney
bp	base pair
Cap	Capsid
Cas	CRISPR Associated Nuclease
CEX	Cation Exchange Chromatography
CFR	Code of Federal Regulations
cGMP	current Good Manufacturing Practices
CGT	Cell and Gene Therapy
CMC	Critical Micelle Concentration
CMC	Chemistry, Manufacturing and Control
CMV	Cytomegalovirus
CPE	Cytopathic Effect
Cre	Cre Recombinase
CRISPR	Clustered, Regularly Interspaced, Short Palindromic Repeat
crRNA	CRISPR RNAs
DB	Domiphen Bromide
DNA	Deoxyribonucleic acid
ds	Double-stranded

DSB	Double-Strand Break
EGFP	Enhanced Green Fluorescent Protein
FDA	Food and Drug Administration
GFP	Green Fluorescent Protein
GMP	Good Manufacturing Practice
gRNA	Guide RNA
HEK293	Human Embryonic Kidney 293
IEX	Ion Exchange Chromatography
ITR	Inverted Terminal Repeat
IU	Infectious Unit
kDa	kilodalton
MCB	Master Cell Bank
MOI	Multiplicity of Infection
mRNA	messenger Ribonucleic acid
MVB	Master Viral Bank
NIH	National Institutes of Health
nt	nucleotide
ORF	Open Reading Frame
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethyleneimine
qPCR	Quantitative polymerase chain reaction
rAAV	recombinant Adeno-Associated Virus
RBS	Rep Binding Site

rcAAV	replication competent Adeno-Associated Virus
Rep	Replication
rHSV	recombinant Herpes Simplex Virus
RNA	Ribonucleic acid
SEC	Size Exclusion Chromatography
SPA	Selective Precipitating Agent
TEM	Transmission Electron Microscopy
vg	vector genome
VP1	Viral Protein 1
VP2	Viral Protein 2
VP3	Viral Protein 3
VV	Vaccinia Virus
WCB	Working Cell Bank
wtAAV	wild-type Adeno-Associated Virus
wtAd	wild-type Adenovirus
WVB	Working Viral Bank

CHAPTER 1: INTRODUCTION

1.1 An Interest to Genetic Transfer Therapeutics

From the understanding of inheritance and nucleic acid structure occurring between late 19th and mid-20th century, to the sequencing of the 3 billion base pairs in the human genome in the late 20th century, we have seen the progress of genetics research influence a variety of fields, especially the medical and biotechnology fields. Nucleic acid research and delivery of nucleic acid has molded modern medicine into a captivating science where therapies are created to improve the quality of life for individuals whose lives have been compromised by their genetic disorder. The introduction of transduction of genetic material and use of viral mediated gene transfer in the 1950s and 1960s and initial discoveries of monogenic diseases during the latter half of the 20th century was just enough motivation for a few scientists, such as Dr. William French Anderson, to postulate how to deliver and correct genetic diseases using delivery vehicles, or also known as vectors, to dependably transport the nucleic acid cargo *in humans*, calling this new form of medicine gene therapy[1]. The first landmark trial of gene therapy in two human patients, Ashanthi DeSilva (age 4 years old) and Cynthia Cutshall (age 9 years old), occurred between September 1990 and January 1991[2,3]. The result of this trial showed researchers and clinicians the safety and efficiency of the gene transfer, albeit not to the most optimal therapeutic levels desired for longer term correction. The brevity at which this trial came to fruition and its resulting therapeutic effect put the spotlight on human gene transfer as many high-impact journals and press began reporting statements such as, "Once considered a fantasy that would not become reality for generations, human gene transfer moved from feasibility and

safety studies in animals to clinical applications more rapidly than expected by even its most ardent supporters" by Ron Crystal in Science[4]. The breakthrough may have led to an unsubstantiated assurance to the use of similarly engineered vectors, other engineered vectors from different virus families, and conducting clinical trials without the full understanding and adjustment to the patient's medical history especially with those taking medications like immunosuppressants. These oversights led to the significant setbacks and tragic deaths of individuals, most notably: a 1999 SCID-X1 trial using an *ex vivo* infection of engineered retroviruses; the death of Jolee Mohr in 2007 from a growth of a fungal infection uncontained due to her immunosuppressants that eventually poked a hole in her gut intestine during her gene therapy clinical trial; and Jesse Gelsinger in 1999 from a gene therapy clinical trial at the University of Pennsylvania in Philadelphia due to a high administered dose of adenovirus to deliver the therapeutic gene for his partial deficiency of ornithine transcarbamylase, that resulted in his immune system to react immediately and Jesse dying several days later due to multiorgan failure[5-7]. These deaths, especially Gelsinger's in 1999, left a black mark on gene therapy to investors that lasted nearly a decade. For researchers in the field, the deaths were wake up calls to the practice of gene transfer experiments, leading to better targeting vectors, more careful selection of patients and better understanding their medical history, and more calculated designs of the genetic material to be transferred. However, this cannot explain the unquestionable resurgence in the last decade with more investments made to gene transfer therapeutics by major corporations like Amgen with Kite Pharma, Pfizer with Spark Therapeutics, Glaxo Smith Kline with TIGET/San Raffaele, and Celgene with Juno Therapeutics[8,9]. The spotlight to gene therapies can be attributed to the success stories in the field, namely Corey Haas (AAV2 for Leber Congenital Amaurosis Type 2, received in 2008), as well as successes in other therapies such as adrenoleukodystrophy and severe immune deficiencies, and now with more pre-clinical

work showing therapeutic efficacy to previously untreatable diseases, mostly coming from academic investigators[9,10]. This history of gene transfer experiments and therapies and an educated projection at the future of the field has been brilliantly captured by Dr. Ricki Lewis' book, written in 2012, "The Forever Fix: Gene Therapy and The Boy Saved It" and is highly recommended for understanding the field, the humanity of the patients involved, and the history of these therapies[11].

Although the success stories have certainly built the foundation of evidence for the field, the progression of gene therapy are attributed to many scientific contributions that have occurred, not only from the improvements to safety and efficacy of the vector delivering the nucleic acid, but also from the design of the therapeutic nucleic acid for the disease of interest. The design of the therapeutic nucleic acid came from a better understanding of the defective genetic information. Significantly contributing to the understanding of the genetic defects in diseased genomes was the Human Genome Project, initiated in 1990 and completed on April 14th, 2003. The results of the project provided and continues to provide researchers with insight in health and pathology of human disease that would advance modern medicine, biotechnology, and new methods for treating previously untreatable diseases. Although the analysis and understanding of the 3.2 billion nucleotide base genome is certainly in its initial stages – especially with epigenetics complicating expression and study of genetic regions - previous discoveries of disease genes of the genome and those nearly 1,800 disease genes found in the Project has inspired scientists to generate therapies for these genetic disorders[12]. Such pursuits are investigating how to correct – i.e. repair, replace, or delete - the defective gene/genes or impart a new function – i.e. regulatory or additive - to the cell that can induce an assumed therapeutic function for known poly- or monogenic diseases at the molecular level using recombinant nucleic acids. According to analysis of current gene therapy clinical trials performed

by Edelstein et. al., almost 80% of the gene therapy clinical trials address cancer, infectious diseases, and cardiovascular diseases because of its prevalence, impact to society, and potentially fatal outcomes[13]. The remainder of the diseases that use gene therapy are not as prevalent, but still have significant impact to individuals who suffer those diseases. A very small subset of gene therapy clinical trials is for the testing of the delivery vector or as a gene locator. The summary of Edelstein data is summarized in Figure 1.

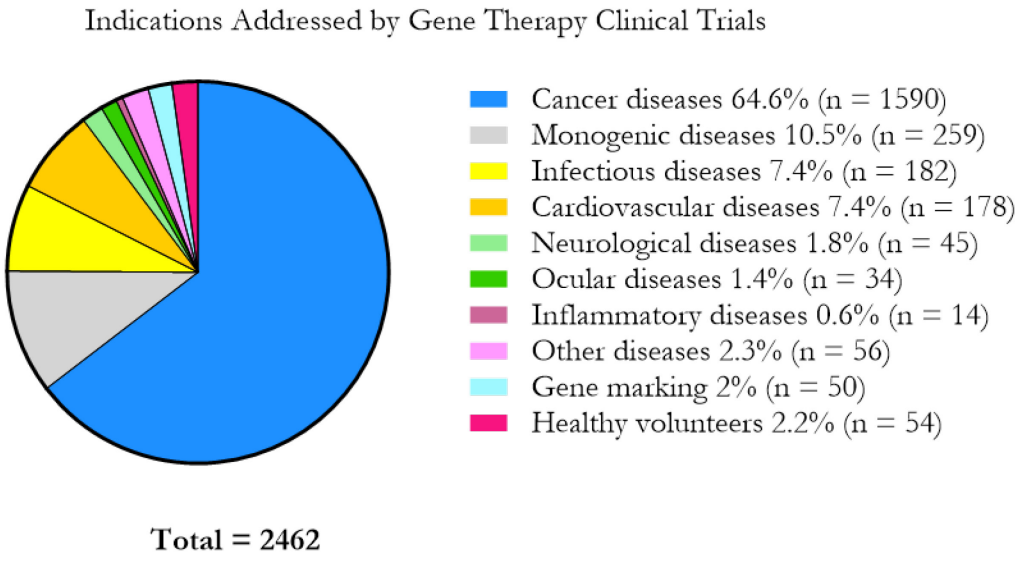


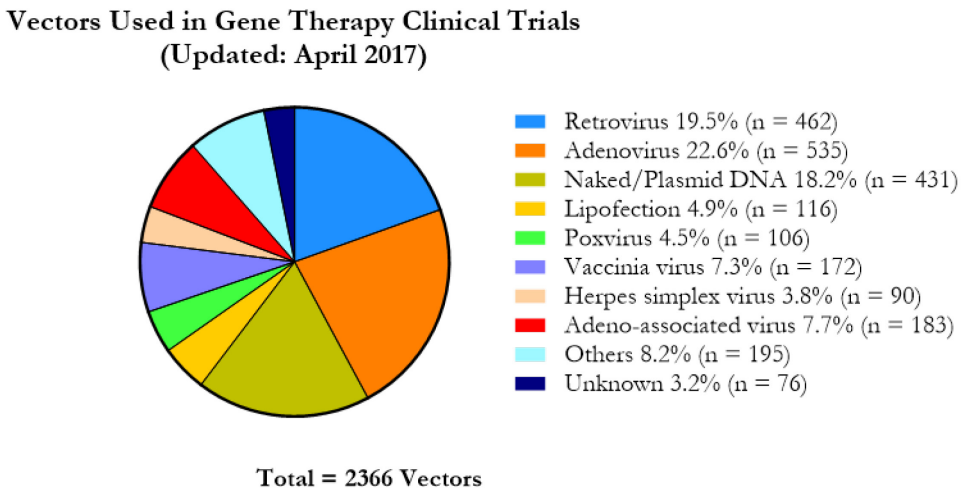
Figure 1: Indications Addressed by Gene Therapy Clinical Trials. Data from The Journal of Gene Medicine.

While the intrigue and rationale of gene therapy are quite apparent, the therapeutic is faced with two major questions: how to dependably deliver the therapeutic gene and how well does the therapeutic gene function in the targeted region? The latter of these two questions is a disease specific assessment that necessitates the knowledge of the mechanism for the natural and defective genetic information as well as proper design of the therapeutic gene for adequate function. The former question is a more global question, independent of the disease mechanism, and is rather focused on how to directly deliver an appropriate amount of the genetic material to the desired cell with limited or without off-target delivery. The vehicle of choice for the delivery

needs to be able to stably hold the genetic material during its transport to the cell of interest, endocytose into the cell, deliver the genetic material into the cytoplasm or in nucleus, and have the cell express the gene contained in the vector. These vehicles can be subdivided into viral vectors and non-viral vectors. Viruses are natural vectors that have evolved many thousands of years to evade the host response and deliver their encapsulated nucleic acid to a variety of cells. Certain viruses can be modified to minimize the likelihood of toxicity and harbor tissue specific epitopes that have been developed to deliver specifically to a target cell. Alternatively, non-viral vectors are vectors designed for delivery of nucleic acid without encapsulating the nucleic acid in a viral protein, be it through synthetic particles that encapsulate the nucleic acid or plasmid DNA that is injected directly *in vivo*. To date, the most effective delivery vehicle has been vectors derived from viruses. Among the design concerns for any vectors used for delivery are limiting off-target effects, limiting interactions with the host immune system, the mechanism of cellular uptake and trafficking of the vector, regulation of delivered gene (hereafter called transgene) expression, and modification of the vector genome to prevent insertional mutagenesis[14].

From the start of the data collection in 2004, Edelstein et. al. has tabulated which vectors have been used in gene therapy clinical trials[15]. As depicted in Figure 2, they have presented the list of vectors used as of April 2017 (Figure 2A) and compiling data from their 2007 update, a new representation of the Edelstein data shows the percentage change of use of vectors in the past decade (Figure 2B)[13,16]. This figure illustrates the trend of the field on the vector of choice for therapy of genetic disorders. It is apparent that the field is favoring the use of viral vectors (n = 1,548) over non-viral counterparts (n = 547), but the percentage change of unique vectors used shows the greatest increase in the use of adeno-associated virus (AAV) in the past decade at +289%, with more than 180 clinical trials being performed worldwide with this vector.

A



B

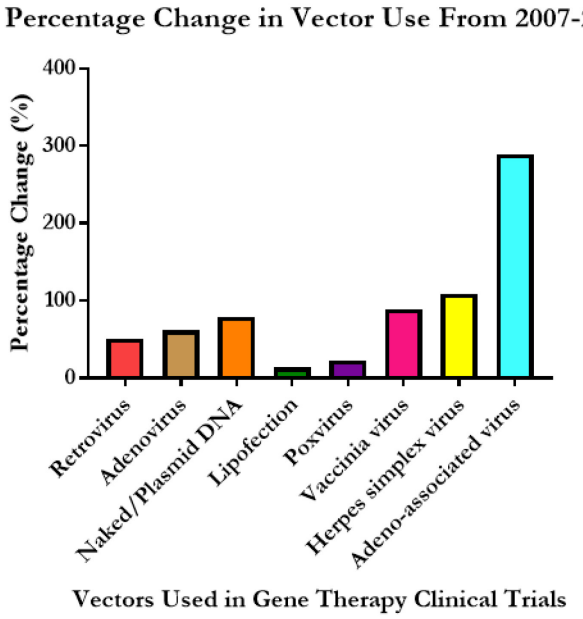


Figure 2: Summary of Journal of Gene Medicine Data. A) Compiled data as of April 2017 of vectors used in gene therapy clinical trials worldwide. B) Percentage change of the use of a specific vector in the past decade. Adeno-associated virus, shown in light blue, has shown the greatest increase in use at 289%.

1.2 Adeno-Associated Virus: An Understanding to its Popularity

First discovered as an associated virus during preparations of the much larger adenovirus in 1965, and later introduced as a gene delivery vector in the 1980s, adeno-associated virus, or AAV, has now become a very useful tool for research labs for *in vitro* and *in vivo* studies[17]. Belonging to the family of viruses known as parvovirus, AAVs are small 20-25 nm in diameter, non-enveloped, and icosahedral viruses that can hold about 4.8 kilobases of single-stranded DNA. Mechanistically, the virus is replication-defective and requires helper virus functions for productive replication, which will be discussed further in later sections. Despite its small carrying capacity, there is a clear interest in using this vector for genetic transfer and so far, the interest has garnered evidence to the vectors utility in delivering its envisioned need: a safe and effective tool for genetic transfer for a multitude of diseases.

Unlike other parvoviruses, AAV is naturally non-pathogenic, and has never been identified as a causative agent to human disease[18,19]. Although it does not directly cause disease, much like other gene therapy vectors namely adenovirus or lentivirus, there has been concern about indirect methods for rAAV to cause disease. As appealing as the 'replication defective' tag has been to AAV, wild-type AAV (wtAAV) is known to preferentially integrate its genome into a specific locus of human chromosome 19 (chr.19q13.3q-ter). This integration is a natural evolution of the virus: to ensure persistence in its primate host, especially since its usual episomal DNA cargo can be diluted out of the system among mitotically active cells[20]. Logically, this has raised concerns for the use of a recombinant AAV (rAAV) virus for long-term genetic therapy because aberrant integrations of genetic material may lead to more health concerns (e.g. cancers). Recombinant forms of AAV, as discussed in more detail in a later section, are stripped of replication components necessary for the virus to replicate its own genome. These same replication proteins, particularly Rep68 and Rep78, were discovered by multiple researchers to be

present in cells for integration into the Chr19 region because of a strong Rep:DNA-binding domain that specifically binds the Rep binding sites (RBS) and displays nuclease activity. This RBS consists of two or more 5'-GCTC-3' repeats and are found at the viral origin of replication, in several promoters, and at the AAVS1 integration site[21,22]. Although, there are possibilities of aberrant integrations to occur in rAAV, this possibility is inefficient and is usually not targeted to Chr19[23]. According to a study performed by researchers at the Stanford University School of Medicine, it has been reported that the frequency of rAAV2 integration events at common integration sites – which was defined as cancer-related genes located by high-throughput, retrovirus-based insertional mutagenesis screens of oncogenesis in mice - was found to be statistically similar to random integration into the tested murine genome. However, these researchers have made conclusions that although insertional mutagenesis has been believed to be negligible due to low integration frequency *in vivo* and no direct evidence of tumorigenesis in a number of preclinical studies, "perhaps this is true in most gene therapy settings that target post mitotic quiescent tissues […] it may become an issue of concern when rAAV2 vectors target proliferating cells such as hematopoietic cells or tissues undergoing regeneration, such as chronic liver inflammation" and integration studies not be limited to only rAAV2 vectors[24].

Although the researchers of this Stanford study provide some evidence for safety of the gene transfer vector being used, the rarity and inefficiency of integration in the host genome by rAAV will likely not perturb the use of the vector in clinical studies and various disease models. In theory, with the sudden interest in using CRISPR/Cas9 technology for gene editing of cells *in vivo*, and with the versatility of rAAV to infect specific cells or tissues, the concern of tumorigenesis occurring in proliferating cells would instead be with higher occurrence from off-target double-strand breaks in gene editing techniques using rAAV as the gene transfer vehicle

rather than from the innate property, albeit extremely low probability, of integration of the rAAV vector genome[25].

One aspect of the field that has garnered great attention recently has been designing a vector that can avoid host immunological reaction and specific targeting of the vector to the desired tissue or cell type. Although no disease is associated with AAV, the vector is composed of biomolecules, i.e. proteins and nucleic acids, that can trigger components of the immune system to develop an immune response to serotypes. The search for a capsid that can generally avoid an immune response, is complicated further because most humans have been exposed to a variety of AAV serotypes and have a pre-existing adaptive response. The development of antibodies can result from natural exposure of AAV serotypes, as shown in the second column of Table 1 which shows the seroprevalence of IgG antibodies produced in serum collected from a population in the Ile de France community in France. This production of antibodies can then result in neutralizing factors that can inhibit vector transduction and therefore have a dramatic effect on AAV clinical efficacy of the rAAV being used, as described in the third column of Table 1. This can certainly be a major shift in momentum of a therapeutic from its preclinical findings to its journey to post-market approval. As shown in Table 1, based on this 2010 study, rAAVs using serotypes 5, 8, and 9 may have an advantage for gene therapy in humans since the anti-AAV IgG produced for serotypes 5, 8, and 9 are 40%, 38% and 47%, respectively, and the neutralizing antibodies are 3.2%, 19%, and 33.5%, respectively.

If the human body is as responsive as it is with natural exposure to AAV serotypes unbeknownst to us, for individuals purposefully receiving the therapeutic of a certain serotype vector, their subsequent introduction of the vector, if applicable to the disease being treated, needs to be of a capsid that does not have preexisting antibodies in the patient. These particular conditions for the therapeutic design entails an investigation to new serotype designs. Generally,

the goal is a one-time administration of the therapeutic and the genetic material is stably inserted into the target cell nucleus. However, for those disease models that involve frequently dividing cells, the genetic material delivered by the rAAV would be diluted from the cells of interest. The challenge for an adequate therapeutic effect in the diseased individual would require multiple doses of the rAAV therapeutic, likely from rAAV vectors with similar specificity to the target tissue and cell, but do not respond similarly from the patient's immune system. This would mean a novel rAAV capsid design able to directly deliver the therapy and avoid the immune system.

Table 1: Analysis of AAV Serotypes: Immune response, receptors, and tropisms

Compilation of Analyses on Human Immunity to Capsid, Receptors for Entry, and Known Tropisms Via *in vivo* Transduction						
AAV Serotype	**Seroprevalence of total IgG Antibodies (Relative to Number of Serum Samples Tested)**	**Seropositive for Neutralizing Factors (Relative to Number of Serum Samples Tested)**	**Receptors**	**Co-Receptors**	**Known Tropism, Via *in vivo* Transduction**	**References**
1	67% (n=210)	50.5% (n=152)	N-linked α-2,3 sialic acid	None Reported	Brain, Muscle, Kidney, Heart, Retina, Liver, Pancreas, Cochlear Inner Ear, Hematopoietic Stem Cells	26-39
2	72% (n=202)	59% (n=89)	Heparin sulfate proteoglycan	α5β1-integrin, hFGFR1, αVβ5-integrin, hHGFP, LamR	Brain, Glioma Cells, Kidney, Photoreceptor Cells, Retina, Solid Tumors/Melanoma, Cochlear Inner Ear	31,33,35,40-48
3	Not Analyzed	Not Analyzed	Heparin sulfate proteoglycan	hFGFR, hHGFR, LamR	Some Retina and Some Solid Tumor/Melanoma	31,43,49-51
4	Not Analyzed	Not Analyzed	O-linked α-2,3 sialic acid	Unknown	Retina and some Brain	31,47,52
5	40% (n=101)	3.2% (n=49)	N-linked α-2,3 sialic acid	PDGFR	Brain, Retina, Photoreceptors, Lung, Liver, Kidney, Some Muscle	31,35 42,53-59
6	46% (n=91)	37% (n=56)	N-linked α-2,6 sialic acid	EGFR	Muscle, Lung, Dendritic Cells, Heart	31,37,57,60-66
7	Not Analyzed	Not Analyzed	Unknown	Unknown	Muscle, Liver, Glioblastoma, Brain	31,67,68
8	38% (n=45)	19% (n=50)	Unknown	LamR	Liver, Muscle, Heart, Pancreas, Glioblastoma	31,40,67-75
9	47% (n=134)	33.5% (n=62)	Galactose	LamR	Heart, Lung, Liver, Muscle	40,69,76-80

For increased specificity, there are a few choices researchers can make: (1) locally introduce the vector rather than systemic distribution, 2) manipulate the capsid to more specifically target desired tissue or cell, or (3) alternatively use a tissue-specific promoter such that the transgene only expresses when it is present in the target tissue or cell. If the area of interest is difficult to administer to, then systemic distribution and tissue-specific promoter strategies are necessary. However, the turnaround time for finding novel capsid designs to the application in clinical work is very lengthy, and so far, many of the clinical trials being performed are more likely to use natural serotypes or rationally designed hybrids of natural serotypes. This will soon change, and in fact a few of the industry sponsored clinical trials for rAAV candidates for various disease models compiled in Table 2 show novel proprietary capsids used to transport gene therapies. These proprietary designs are most likely based on natural serotype receptor interactions or natural *in vivo* transduction, compiled in Table 1, or can be rationally designed to incorporate high-affinity ligands into the AAV capsid to redirect or restrict viral tropism. Other capsids that are less likely to be currently present in industry sponsored clinical trials are capsids generated using direct evolution. Various novel serotypes can be generated using a variety of techniques: error-prone PCR to introduce random point mutations into the Cap open reading frame, chimeric capsids using *in vivo* recombination or DNA shuffling, random peptide insertions, or randomization of surface loops. Although the ease at which capsid manipulation can happen, the resulting novel vectors may be incapable of packaging, not be significantly different to the predecessor, or the vector transduction efficiency does not translate well from murine models to human subjects. All told, the tunability of the capsid with the ease found in rAAV is rare among other viral vectors used for gene therapy, making this viral vector a first-round pick amongst other vectors to consider in the design of a gene transfer vehicle.

Table 2: List of Industry Sponsored rAAV Clinical Trials in progress or recently in progress.

Selected examples of interventional clinical candidates using rAAV, sponsored by industry, Part 1 of 2							
Industry Sponsor	**Candidate Name**	**AAV Serotype Used**	**Current Status (US)**	**Molecular Target**	**Major indication**	**NCT Number**	**Reference**
Amsterdam Molecular Therapeutics, now UniQure	AMT-011	1	Phase 2|Phase 3	Human LPL (S447X)	Familial Lipoprotein Lipase Deficiency	NCT00891306	79
Applied Genetic Technologies Corp	AGTC-402	2	Phase 1|Phase 2	CNGA3	Achromatopsia	NCT02935517	80
	rAAV2tYF-CB-hRS1	2	Phase 1|Phase 2	Retinoschisin	X-linked Retinoschisis	NCT02416622	81
	rAAV2tYF-PR1.7-hCNGB3	2	Phase 1|Phase 2	CNGB3	Achromatopsia	NCT02599922	82
	rAAV1-CB-hAAT	1	Phase 2	Alpha-1 Antitrypsin	Alpha-1 Antitrypsin Deficiency	NCT01054339	83
	rAAV2-CB-hRPE65	2	Phase 1|Phase 2	Human RPE65	Leber Congenital Amaurosis Type 2	NCT00749957	84
Arthrogen|Centre for Human Drug Research (CHDR)	AAV5.NF-kB.IFN-β)	5	Phase 1	hIFN-β	Rheumatoid Arthritis	NCT02727764	85,86
Audentes Therapeutics	AT342	8	Phase 1|Phase 2	UGT1A1	Crigler-Najjar Syndrome	NCT03223194	87
	AT132	8	Phase 1|Phase 2	hMTM1	X-Linked Myotubular Myopathy	NCT03199469	88
AveXis, Inc.	AVXS-101	9	Phase 3	SMN	Spinal Muscular Atrophy 1	NCT03306277	89
Baxalta, now part of Shire	AskBio009	8	Phase 1|Phase 2	Human Factor IX	Hemophilia B	NCT01687608	90
Celladon Corporation	AAV1/SERCA2a	1	Phase 2	SERCA2a	Heart Failure	NCT01643330	91
Ceregene now part of Sangamo Therapeutics	CERE-120 (AAV2-Neurturin)	2	Phase 2	Neurturin	Idiopathic Parkinson's Disease	NCT00400634	92
	CERE-110	2	Phase 1	Beta-Nerve Growth Factor	Alzheimer's Disease	NCT00087789	93,94
Digna Biotech S.L.|UniQure N.V.	rAAV2/5-PBGD	2/5	Phase 1	Porphobilinogen Deaminase	Acute Intermittent Porphyria	NCT02082860	95
Dimension Therapeutics	AAVrh10FIX	rh10	Phase 1|Phase 2	Human Factor IX	Hemophilia B	NCT02618915	96
	scAAV8OTC	8	Phase 1|Phase 2	Human Ornithine Transcarbamylase	Ornithine Transcarbamylase (OTC) Deficiency	NCT02991144	97
Genethon	AAV1-gamma-sarcoglycan	1	Phase 1	Gamma Sarcoglycan	Limb Girdle Muscular Dystrophy Type 2C	NCT01344798	98
GenSight Biologics	GS030-DP	AAV2.7m8	Phase 1|Phase 2	ChrimsonR-tdTomato	Non-syndromic Retinitis Pigmentosa	NCT03326336	99
	GS010	2	Phase 3	ND4	Leber Hereditary Optic Neuropathy	NCT02652767	100
Genzyme, a Sanofi Company|Sanofi	AAV2-sFLT01	2	Phase 1	Flt1 Receptor (anti-VEGF)	Neovascular Age-Related Macular Degeneration	NCT01024998	101
	AAV-hAADC-2	2	Phase 1	Human Aromatic L-Amino Acid Decarboxylase	Parkinson's Disease	NCT00229736	102

Selected examples of interventional clinical candidates using rAAV, sponsored by industry, Part 2 of 2							
Industry Sponsor	**Candidate Name**	**AAV Serotype Used**	**Current Status (US)**	**Molecular Target**	**Major indication**	**NCT Number**	**Reference**
Horama S.A.	AAV2/5-hPDE6B	2/5	Phase 1|Phase 2	HORA-PDE6B	Retinitis Pigmentosa	NCT03328130	103
Lysogene	SAF-301	10	Phase 1|Phase 2	Human SGSH and SUMF1	Sanfilippo Type A Syndrome	NCT02053064	104
MeiraGTx UK II Ltd	AAV RPE65	2/5	Phase 1|Phase 2	Human RPE65	Inherited Retinal Dystrophy Due to RPE65 Mutations	NCT02781480	105
	AAV2/8-hCARp.hCNGB3	2/8	Phase 1|Phase 2	CNGB3	Achromatopsia	NCT03001310	106
	AAV2hAQP1	2	Phase 1|Phase 2	Human Aquaporin-1	Irradiation Induced Parotid Salivary Hypofunction	NCT02446249	107
	AAV2/5-hRKp.RPGR	2/5	Phase 1|Phase 2	Retinitis Pigmentosa GTPase Regulator	X-Linked Retinitis Pigmentosa	NCT03252847	108
Neurologix, Inc.	AAV-GAD	2	Phase 1	Glutamic Acid Decarboxylase	Parkinson's Disease	NCT00195143	109,110
NightstaRx Limited	AAV-RPGR	8	Phase 1|Phase 2	Retinitis Pigmentosa GTPase Regulator	X-Linked Retinitis Pigmentosa	NCT03116113	111,112
Regenxbio Inc.	RGX-314	8	Phase 1	Anti-VEGF	Neovascular Age-related Macular Degeneration|Wet Age-related Macular Degeneration	NCT03066258	113
Sangamo Therapeutics	SB-FIX	2/6	Phase 1	Human Factor IX	Hemophilia B	NCT02695160	114
	SB-318	2/6	Phase 1	IDUA	Mucopolysaccharidosis I	NCT02702115	115
	SB-913	2/6	Phase 1	IDS	Mucopolysaccharidosis II	NCT03041324	116
	SB-525	2/6	Phase 1|Phase 2	Human Factor VIII	Hemophilia A	NCT03061201	117
Spark Therapeutics	SPK-8011	AAV-Spark200	Phase 1|Phase 2	Human Factor VIII	Hemophilia A	NCT03003533	118
	AAV2-hRPE65v2	2	Phase 3	Human RPE65	Inherited Retinal Dystrophy Due to RPE65 Mutations|Leber Congenital Amaurosis	NCT00999609	119
	AAV2-hCHM	2	Phase 1|Phase 2	Choroideremia Gene	Choroideremia|CHM (Choroideremia) Gene Mutations	NCT02341807	120
	AAV8-hFIX19	8	Phase 1	Human Factor IX	Hemophilia B	NCT01620801	121
Spark Therapeutics|Pfizer	SPK-9001	AAV-Spark100	Phase 1|Phase 2	Human Factor IX	Hemophilia B	NCT02484092	122
Tacere Therapeutics, Inc.	TT-034	8	Phase 1|Phase 2	Anti-Hepatitis C Virus	Chronic Hepatitis C Infection	NCT01899092	123
Targeted Genetics Corporation	tgAAC94	2	Phase 1|Phase 2	TNFR-IgG1-Fc	Arthritis, Rheumatoid|Arthritis, Psoriatic|Ankylosing Spondylitis	NCT00126724	124
UniQure Biopharma B.V.|Chiesi Farmaceutici S.p.A.	AAV5-hFIX	5	Phase 1|Phase 2	Human Factor IX	Hemophilia B	NCT02396342	125
	rAAV2/5-hNAGLU	2/5	Phase 1|Phase 2	Human Alpha-N-acetylglucosaminidase cDNA	Sanfilippo Type B Syndrome	NCT03300453	126
Voyager Therapeutics	VY-AADC01	2	Phase 1	hAADC	Idiopathic Parkinson's Disease|Parkinson's Disease|Basal Ganglia Disease	NCT03065192	127

For those in the field of developing a gene therapeutic for a disease model, the tunability of the viral capsid to more specifically target the therapeutic, the low-risk for onset of vector causing diseases, and the replication defective properties of the biologic are what make rAAV a popular vector to use, despite the small carrying capacity of genetic material (4.8 kb of total single-stranded DNA, ~4.4 kb excluding necessary viral components). Arguably, what is probably the most desirable aspect of this avenue of therapy, for researchers and more so the individuals afflicted by the genetic disorder, is the capability of long-term persistence, which has been exploited for gene transfer in a variety of applications, especially for slow proliferating cells. While adenovirus- and retrovirus-based vectors are known for their stable gene transfer and high-level expression, and hence explaining their abundant use in gene therapy clinical trials, both vectors come with considerable safety concerns, which center around the inflammatory potential of adenoviral vectors and the possibility of unwanted chromosomal integration of retroviral vectors. Current efforts are being made to reduce the immunogenicity of adenoviral vectors through deletion of viral early genes, whereas retroviruses are being reengineered to self-inactivate and prevent potential for integration, called SIN retroviral vectors; however, testing of these modifications has been limited to local and *ex vivo* applications [130,131]. The properties for the natural biology of rAAV in conjunction with the achievements in genetic engineering can attest to rAAV being the method of choice for *in vivo* gene transfer.

1.3 Involvement of Biopharma: Reason to Invest in Curative Therapy

For a therapy that is geared towards treating diseases that is colloquially termed 'Orphan Diseases', it logically should not gain the attention of profit-oriented big pharmaceutical companies to invest in a single-dose curative medicine and altruistically help cure diseases that affect a very small population of the world. In 1983, Congress passed the Orphan Drug Act to deal with the unique commercial and regulatory challenges by 'orphan' diseases that afflict fewer than 200,000 Americans[132]. For industrial superpowers, such as Pfizer, Sanofi, Bristol-Myers Squibb, this Act was incentive enough to invest in a therapeutic that will be required by only a small number of patients. However, big pharmaceutical companies are usually not forefronts of the clinical studies for novel gene therapies. Smaller institutes, like hospitals or academic laboratories, would use their preclinical work for their novel therapeutic as evidence to drive their therapeutic into clinical trials. If these smaller groups can capture the attention of big pharmaceutical companies, the big companies are more inclined to back smaller companies or buy out the start-ups and continue the clinical trial work. This is because these start-up companies not only benefit from government support and general money-saving techniques that caught the attention of larger pharmaceutical companies, but they can also act like scapegoats to take the brunt of the impact for the heightened possibility of failure in early-to-middle stage clinical trials due to a small cohort of this already small population being tested for safety, short-term efficacy, and long-term efficacy. On the other hand, with more success they gain from clinical trials, the more support they would receive from government and venture capitalists. For an example of ebb and flow of these start-ups, and also a marque example of a catastrophic failure that can occur for a biologic in clinical trials, we can look at the publicized clinical trial called Mydicar developed by Celladon. Mydicar was designed to work via infusion of the gene for the SERCA2a enzyme, that is deficient in heart failure, into coronary arteries and restoring

enzyme production in cardiac cells and improving heart contractions. In 2012, Celladon had obtained substantial interest from investors to advance development of the drug. When the therapeutic failed to meet primary and secondary endpoints in Phase IIb, Celladon had to suspend all research tied to Mydicar. Even though Celladon had other gene therapy programs and other investigational product candidates to advertise to shareholders, the failure of its lead product consequently led to Celladon stocks to plummet[133,134]. The heart failure therapy failed to show statistically significant reduction in hospitalizations, improvement in cardiovascular survival rates, and all-told unable to free patients from the need for ventricular-assist devices and heart transplants[134]. The shell of a company was bought out by Eiger BioPharmaceuticals with hopes to reenter Wall Street with a handful of rare disease treatments[135]. The pursuit of finding the so called 'niche busting' drug as termed by a 2010 Nature Medicine News snippet, has led to big pharmaceutical companies creating their own rare disease research units and we are seeing more financial backing or direct buying of Biotech companies such as Pfizer/Spark, Bristol-Myers Squibb/UniQure and Sanofi/Genzyme[136]. Not only are these larger companies investing into smaller companies, especially with preclinical successes of rAAV therapies capturing their attention, many venture capitalists are seeing the investments of these larger companies and are also trying to profit themselves[8,137].

All the while small molecule generic medicines are siphoning profits away from these big pharmaceutical companies, and the evermore difficult discovery of new blockbuster drugs becoming more expensive for pharmaceutical companies to invest in, the 'orphan' drug sector is relatively untapped. Perhaps the goal of the big pharmaceutical company in investing in 'orphan' drugs is to reap the benefits of the tax credits on clinical trial expenses, grant funding from FDA, and the seven-year period of market exclusivity for their prized rare disease therapy.

Perhaps also, due to the small cohort size in clinical trials, the evidence for effectiveness of 'orphan' drugs are lower in quality than required in regular drugs and more side effects are tolerated meaning the approval process is quicker compared to small-molecule medications[138]. Even still, because of the rarity of the disease and the exclusivity of their therapeutic, the companies can drive up prices to help recover costs and profit in this small market since their expenditures on the journey to approval are usually not subjected to price ceilings or maximum budgets. The price tag associated with the drug can be unrelated to the effectiveness or prevalence, questioning their cost-effectiveness.

As merciless as it is to charge individuals with a poor quality of life exuberant prices for a life-altering therapeutic, the more than 6,000 classified 'orphan' diseases that affect as many as 25 million Americans according to the National Institutes of Health (NIH) are finally receiving the attention to find a curative therapy[139]. Whether or not patients and their families would invest in the expensive medication is at the discretion and situation of the individuals in need of the therapy. Nevertheless, the field is constantly restructuring: from the regulatory aspect of the therapeutics on rare diseases which would require approval from a smaller population for safety and efficacy, to the pricing cost of the therapeutic from the pharmaceutical companies to break even from manufacturing and regulatory costs and consider a curative single-dose pricing model[9,140-142]. To reduce as much cost for the development of the Cell and Gene Therapy (CGT) therapeutic, big pharma has developed strategies that one Forbes writer describes as an "outside-in" approach: have the clinical translation work be performed outside the company, usually by academic investigators at top tier medical centers. Only after the academics deliver important data would major biotech and venture capital funding provide financial backing to these findings to launch small biotech startups based on their developing therapeutic and enter early phase clinical trials. This Forbes description of the reduction of cost for the pharmaceutical

companies goes into a discussion of third-party vector production facilities to catalyze academic research as well as natural evolution of technologies such as gene editing, basic science research behind the vectors used, and also improvements in administration of the therapeutic[9]. But as described in the next section, these cost reduction strategies are clearly not enough to reduce the seven-figure price tag. So, the question remains: why is the price so high? There are other factors at play for overall cost of the product - regulatory, facilities, quality control, research and development, employee, and manufacturing – all of which are being restructured especially with the nascence and constant developments of rAAV therapies in the current market.

There are only two rAAV therapies that have been approved (or soon-to-be approved) in the US or in the European Union: Glybera™ (alipogene tiparvovec, UniQure, AAV1 containing an intact copy of lipoprotein lipase, afflicted population estimates for lipoprotein lipase deficiency in the general population: 1 in 250,000) and Luxturna™ (voretigene neparvovec, Spark Therapeutics, AAV2 containing human RPE65 cDNA for inherited retinal diseases, afflicted population estimates for inherited retinal dystrophy in US and Europe: 3,500). The pricing of these rAAV therapy is amongst the highest, if not the highest reported for any approved therapeutic in US or the European Union: Glybera™ at $1 million and Luxturna™ estimated at about $1 million[143]. The exorbitant price of $1 million is difficult to justify, but if the manufacturing process behind making the rAAV is elucidated, coupled with the fact of a small market, a majority of the price tag can make sense. The analysis of the price can start with the therapeutic that was approved by the European Medicines Agency for treatment of lipoprotein lipase deficiency (LPLD) in 2012, Glybera™ (alipogene tiparvovec)[144]. Since its approval, its publicity from the "first gene therapy" in the western world quickly soured to "the most expensive drug in history" at an unprecedented $1 million price tag and eventually this drug became a bust. It did ignite the explosion of investment and excitement around treatments

to correct genetic defects, but unfortunately the therapeutic Glybera™ has been used just once. Performed in Berlin, the 43-year-old approximately 60-kg female patient received 40 injections, of 1.5 x 10^{12} genome containing particles per injection, to the muscles to help the patient process fats, with the therapeutic showing clinical efficacy to the patient and allowing her to live a better quality of life. The rarity of the disease, in addition to the price tag, and concerns about the effectivity of the therapeutic all play a variable in the success equation of this therapy. UniQure, who is the company to back the journey of this therapeutic, were trailblazers in the European Union, and the information they presented was too advanced for the regulatory agency involved to approve Glybera™. From the small cohort study, the data presented sub-par results in the patients receiving the therapy. After pushing several times to approve the drug, the drug was approved in 2012, UniQure decided to go public on Nasdaq, and they vouched to commercialize Glybera™ in the US. Instead, the FDA required more stringent and expensive clinical trials that dissuaded UniQure from pursuing the US commercialization goal[145]. Furthermore, several patients in Europe wanted to seek treatment with Glybera™ but could not afford the drug and hardly any insurance companies would front the cost of this gene therapy. After about 5 years since approval, UniQure announced in April 2017 that it will not renew the drug's marketing authorization in Europe stating "The drug's usage has been extremely limited and we do not envision patient demand increasing materially in the years ahead."[146] The price tag can be partially justified by the cost it took to get the drug to the European Medical Agency and through clinical trials, which some reports have said it totaled about $100 million. In their $1 million charge, executives were possibly attempting to profit and recover expenditures lost in the following cost variables: regulatory costs, an attempt at single-dose pricing, the manufacturing of the rAAV, the small market, and cost of keeping the drug in the market. As economists, industry

executives, and regulatory agencies work on ways to improve incentives and pricing schemes, researchers can focus on reducing costs from the manufacturing process perspective.

Likely an unfortunate byproduct to the substantial increase in price for these approved rAAV-based therapies is the bottleneck that arises to supply sufficient rAAV vectors to be used in further clinical studies and market settings, especially for systemic diseases that require a large supply of virus per dose. For rAAV therapeutics that are well into Phase II and Phase III clinical studies, a serious holdup that faces their approval is the paucity of large scale manufacturing methods in accordance with current Good Manufacturing Practices (cGMP) that can produce sufficient high-quality and high-potency rAAV to meet the demand of these trials and beyond. Consequently, a "domino-effect" ensues that interferes with making rAAV the marque therapeutic of that genetic disorder. Due to insufficient production capacity from current methods, the middle- to late-stage clinical trials are subjected to limited number and duration of trials since these trials require multiple human subjects and high amounts of high-quality and high-potency rAAV. This can make further development of a rAAV therapeutic difficult because the investment for the drug to be approved may severely outweigh the gains in market, as seen with Glybera™ in the European market. This hesitation to supply resources from investors or otherwise, coupled with the current tedious and lengthy production process of rAAV, will without a doubt cause rAAV medication to continue to be very expensive. Therefore, therapeutics needing larger clinical studies to receive approval by regulatory agencies more importantly depends upon improvements to production methods to propel them to these clinical studies and beyond.

The methods used to produce Glybera™ and Luxturna™ are not an uncommon approach in rAAV manufacturing, the recombinant baculovirus-rAAV system and HEK293 transfection methods, respectively[147-149]. However, the methods used in each therapeutic has

its own drawbacks and advantages, especially if these methods were to be used for more systemic disorders, such as Duchenne's Muscular Dystrophy. Triple plasmid transfection is the most common method for production of rAAV in laboratory and early Phase I clinical trials. However, as rAAV demand increases, this method becomes laborious, costly, and time-consuming. Alternatively, viruses can be used to carry the AAV genes, thus reducing the number of vectors used to enter a single cell. This method involves infecting a cell line harboring all AAV components, called producer cell line, or a cell line harboring only AAV Rep/Cap genes with the rAAV vector provided by the virus, called packaging cell line. Additionally, suspension cell cultures are preferred over adherent cultures to avoid batch to batch cell confluency inconsistencies, avoid resource consuming equipment such as custom-made roller bottle incubators, and approach commercial scalability via bioreactors holding 1,000 L or more[150]. The six studied methods for large-scale production of rAAV are triple plasmid transfection, dual recombinant Herpes Simplex Virus Type 1 (rHSV) infection, recombinant baculovirus-rAAV hybrid infection (OneBac), recombinant vaccinia virus and adenovirus-rAAV hybrid infection, wild-type adenovirus and adenovirus-rAAV hybrid infection, and single recombinant adenovirus infection (summarized in Table 3). The following sections will first describe the necessary background information of rAAV viral replication biology, segueing into established large-scale production methods that are developed given the rAAV biology, their improvements to meet FDA and International Conference on Harmonisation (ICH) guidelines or Code of Federal Regulations, and scalable separation techniques for contaminants from the final rAAV product.

Table 3: rAAV Vector Manufacturing Methods and Yields.

rAAV Vector Yields from Large-Scale Manufacturing Method						
Vector used	**Number of Vectors**	**Method**	**Cell Line**	**Cell Type**	**Vector Genome (vg/cell)**	**References**
Polyethyleneimine (PEI)	3	Transfection	HEK293SF	Suspension	2×10^5	151
Vaccinia Virus and Recombinant Adenovirus	2-3	Infection	HeLa (S3) or HEK293	Suspension	$3x10^3$-$4x10^4$	152
Recombinant Herpes Simplex Virus Type 1	2	Infection	Baby Hamster Kidney (BHK)	Suspension	1×10^5	153
Recombinant and Wild-type Adenovirus	2	Infection	HeLa (B50)	Suspension	2.2×10^4	154
Recombinant Baculovirus	1	Infection	*Spodoptera frugiperda* (Sf9)	Suspension	5×10^5	155
Recombinant Adenovirus	1	Infection	HEK293	Adherent	1.3×10^5	156

1.4 AAV Biology: How the Virus Gets Made

To understand the bottleneck in manufacturing methods, we must first understand the wtAAV biology. The wild-type AAV genome undertakes a T-shaped double hairpin structure and consists of two open reading frames (ORFs). These ORFs are flanked by two palindromic inverted terminal repeats (ITRs) that are 145 bases in length, of which the distal 125 bases form the palindrome and is capable of folding into the characteristic T-shaped hairpin shape[19]. These ITRs serve as important sites for initiating DNA replication, packaging into the capsid and integration into the host genome. The left ORF contains the Rep gene and encodes for four nonstructural proteins required for AAV DNA replication and transactivation of AAV transcription[157]. The p5 promoter directs the expression of a single transcript encoding the larger Rep 78 and Rep 68 proteins, which have similar functions. Rep78/68 are required for nearly every step of the AAV life cycle. Rep78/68 homodimers autoregulate the AAV gene expression by binding to the double-stranded stem region of the ITRs. These homodimers create nicks at the terminal resolution site (trs) to form a linear structure that permits replication of the AAV genome[19,158]. Rep78/68, along with the ITRs, are also necessary for AAV DNA integration into the host AAVS1 locus, as briefly described earlier in Section 1.2[159]. The p19 promoter drives the expression of a second transcript that contains sequences for the smaller Rep52 and Rep 40 proteins. Rep52/40 possess helicase functions for full-length genome packaging into the preformed capsid[160]. An illustration of the replication and packaging pathway is described in Figure 3 and Figure 4, depending on the ITR design used for the rAAV (described later).

The right ORF contains the Cap gene, which codes for the three structural proteins of the viral capsid. These capsid proteins, VP1, VP2 and VP3, are derived from a single transcript

initiated at the p40 promoter. A total of 60 copies of VP1, VP2 and VP3 are present in a 1:1:10 ratio, respectively, and combine to form an ordered icosahedral structure, at about 20-25 nm in diameter[72]. X-ray crystallography and pseudo-typing studies have elucidated the organization and interaction between these subunits that influence receptor binding, cell entry and endosomal escape[72,161,162]. Different serotypes of AAV are defined by their distinct receptor binding, tissue tropism and antigenicity, as described earlier in Section 1.2 and in Table 1. For example, the AAV2 and AAV3 serotypes recognize heparin sulfate on the cell surface, while AAV1, AAV4 and AAV5 bind sialic acid. While naturally occurring, serotypes have at least 45% homology in their capsid structure, residues within the hypervariable region of the VP3 protein confer unique surface topologies[163-165]. Assembly of the AAV occurs in two distinct steps: VP1, VP2, and VP3 are synthesized and assembled into an empty virion in the nucleus. Secondly, in a rate-limiting step, single-stranded AAV DNA is inserted into the pre-formed capsids.

The temporal expression of AAV ORFs is orchestrated by Rep proteins as well as helper virus products [166]. Adenovirus (Ad) clearly has the best-studied relationship with AAV, although herpesvirus, human papilloma virus (HPV) and vaccinia virus can also provide helper or sub helper functions[167]. Essential Ad genes that aid in AAV replication include E1A, E1B, E2A, E4 and VA RNA. E1A is required for transcription of Ad early genes and initiates Rep and Cap expression through binding of the AAV p5 promoter. The E1A proteins activates the other early genes of adenovirus and induce the cell to enter S phase to create an environment optimal for virus replication. E1B and E4 help to promote AAV replication and second strand synthesis. These two proteins also function as a ubiquitin ligase to degrade certain targets that limit rAAV transduction. E2A is a single-strand DNA binding protein used in AAV replication and generally helps with replication processing. VA RNA influence viral mRNA stability and translation,

particularly for the Cap gene [168]. Altogether, Rep and helper virus functions are critical for different stages of the AAV life cycle and promoting productive AAV infection. The helper functions for HSV and vaccinia virus will be discussed further in Section 1.5.3 and Section 1.5.4, respectively.

The current model for productive AAV replication can be divided into three steps: (1) single-strand genome of AAV being extended into a double-strand template for transcription of the Rep gene, (2) the Rep genes are then transcribed, and (3) extensive DNA replication occurs. The ITRs that flank the AAV genome serve as the viral origin of replication and signals for packaging into the final vector. The self-annealing property of the ITR is instrumental to AAV DNA replication. The result of the self-complimentary sequence lends the secondary structure and a base-paired 3' hydroxyl group for unidirectional DNA synthesis that is mediated by the host replication machinery under adenovirus helper replication. In herpes virus helper replication, the cellular replication machinery is replaced with those provided from the herpes virus, discussed further in Section 1.5.3. When the AAV template has been copied, terminal resolution occurs to replicate the ITR that had initially primed the replication process. To replicate this region, Rep proteins specifically bind to the RBS motif to perform certain actions: (1) render the terminal resolution site (trs) single-stranded using its innate helicase activity to then (2) recreate a 3' hydroxyl end by exacting a site- and strand-specific nick at the trs, allowing the ITRs to be replicated by the cellular replication machinery. The ITR renatures, into a terminal hairpin putting the 3' hydroxyl group in position for single-strand displacement synthesis (otherwise known as 2nd strand synthesis) and elongate the AAV genome into a double-stranded full length AAV genome (duplex monomer) and a single-stranded full-length AAV product that may serve as a template for further replication. The double stranded replication product would be cleaved by AAV Rep proteins at the junction ITR to yield

monomers that are then packaged into an AAV particle as single-stranded DNA[19,169-171]. Modifications were made in a 2003 Gene Therapy paper by Xiao Xiao's group where the single-stranded AAV ITR design was truncated by deleting the D-sequence (the packaging signal) with the adjacent trs on one ITR, choosing the left ITR in their study. This would result in the dimers failing to be resolved in monomers via Rep/trs nicking and the dimer still being packaged because of the remaining intact wtAAV ITR on the right side that still had the intact packaging signal. Because of the resulting generated dimer, the total size of the AAV vector must be truncated to less than 2.5 kb so that the dimer molecule does not exceed the viral packaging limit of 5 kb. Their method showed that the deletion of D-sequence on one ITR did not affect the efficiency of viral replication and packaging, nor did it effect production of the virus in larger scales[172]. The theorized description of the single-stranded replication model and double-stranded replication model are provided in Figure 3 and Figure 4, respectively. Other modifications were made using the properties of AAV ITRs and replication to meet a need for faster expression of the transgene or for delivering >5 kb transgenes. For faster expression, efforts have been made much like Xiao's group, to generate self-complementary AAV products to limit the lag-time in second-strand synthesis, which is a rate-limiting step for transduction since converting single-stranded DNA to double-stranded DNA is a necessary step before the gene is expressed[173,174]. For expanding size of cargo from ~4.4 kb to about ~8 kb, researchers are using two vectors to "expand" AAV packaging capacity, which involves two designed rAAV to infect one cell and using different strength enhancers, recombination, or splicing techniques to reconnect a transgene split into the two AAV vectors[175].

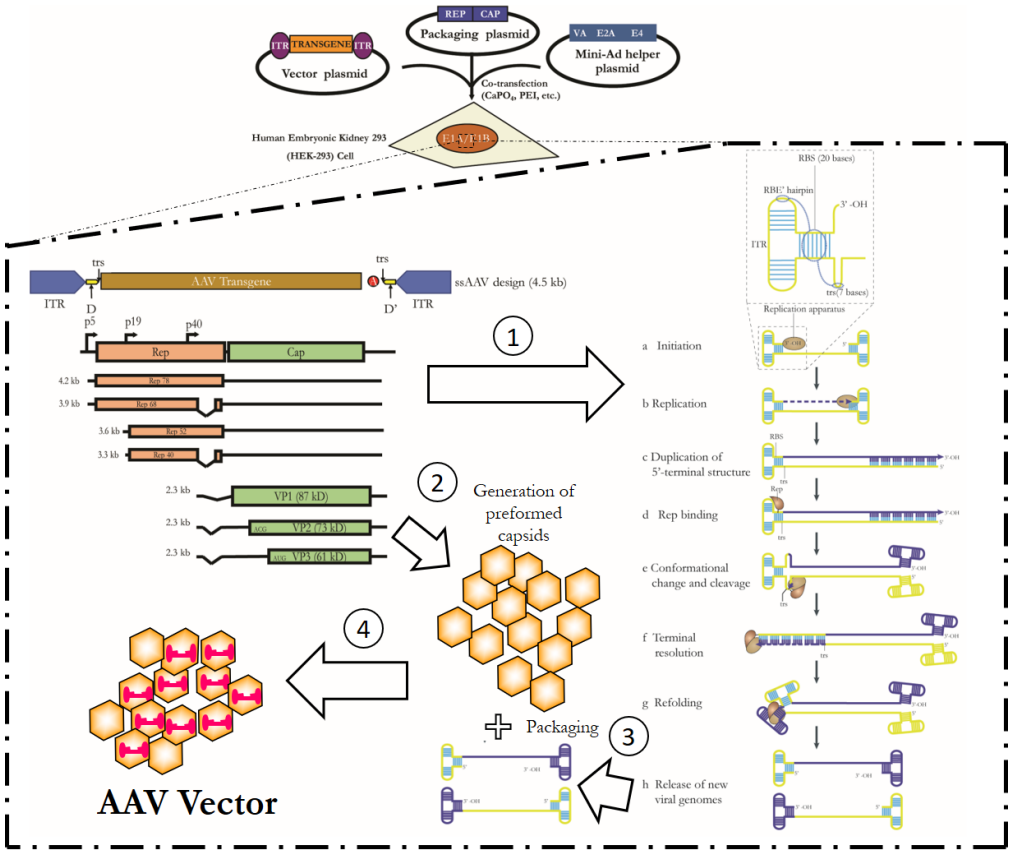


Figure 3: AAV Replication Schematic for Single-Stranded DNA Design, wtAAV ITR genome. (1) In a concerted effort, the Rep proteins (78, 68, 52, and 40) assist in the replication of the genomic DNA. Briefly, the ITR provides a primer within its secondary structure to initiate the DNA replication using host cell replication (or helper virus) machinery. The template is copied, as well as the distal 5' ITR. Next, the 3' ITR that initiated the replication is replicated by a precise nicking at the terminal resolution site. The result after replication is a renatured ITR from both sites of the AAV genome that are then cleaved into monomers via Rep/trs nicking. (2) All the while that the replication process is occurring, the generation of preformed capsids is promoted by p40, and an assembly of 60 total copies VP1, VP2, and VP3 arranged in a 1:1:10 ratio. (3) The resulting monomers are guided to the preformed capsids with the assistance of the Rep52/40 proteins. (4) The final rAAV product is generated, which includes a mixture of full, genome-containing particles and empty-particles. (Image adapted from Chandler et. al[176])

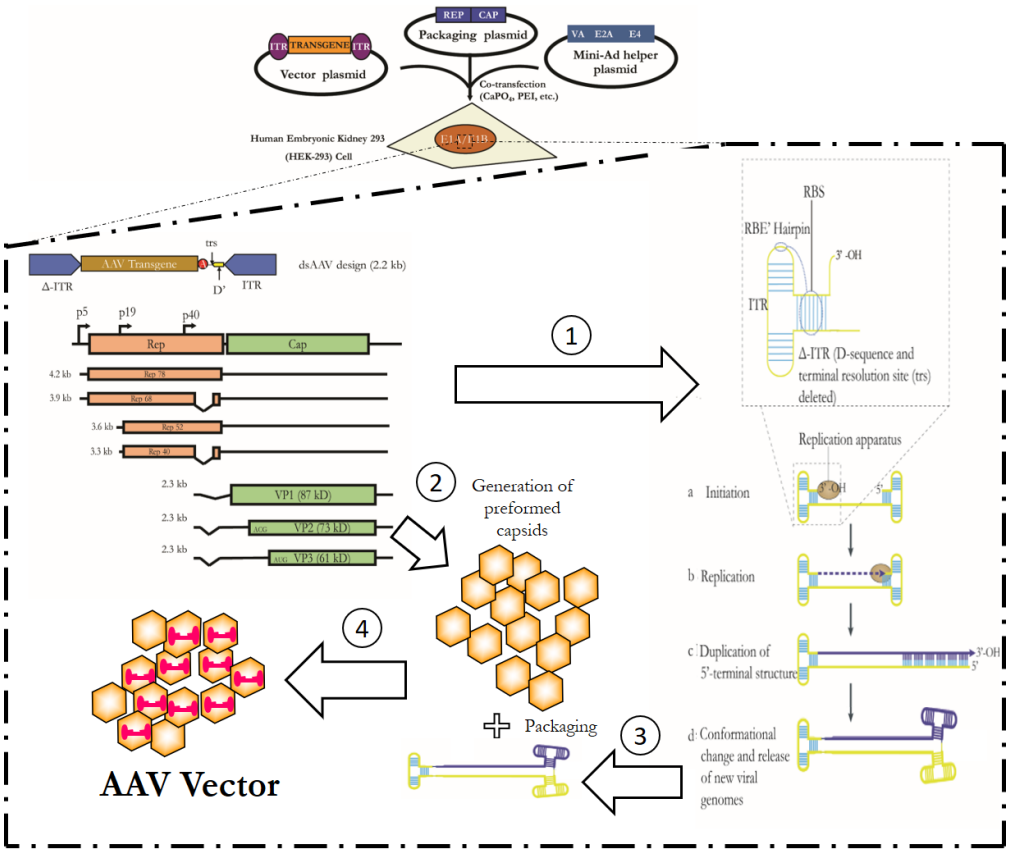


Figure 4: AAV replication Schematic for Double-Stranded DNA Design, Truncated ITR Genome. (1) Again, like the single-strand DNA design, in a concerted effort, the Rep proteins (78, 68, 52, and 40) assist in the replication of the genomic DNA. The DNA to be replicated is nearly half the size of the single-strand DNA design because this self-complementary viral genome was designed by truncating an ITR by removing the trs and D-sequence. The deletion of trs results in the lack of Rep/trs nicking of the dimers, and therefore the genome remains in the dimer form. (2) All the while that the replication process is occurring, the generation of preformed capsids is promoted by p40, and an assembly of 60 total copies VP1, VP2, and VP3 arranged in a 1:1:10 ratio. (3) The resulting dimer, that contains a still intact D-sequence, the packaging signal, on the distal wtAAV ITR are guided to the preformed capsids with the assistance of the Rep52/40 proteins. (4) The final rAAV product is generated, which includes a mixture of full, genome-containing particles and empty-particles and a genome of nearly half the size of the natural single-stranded DNA AAV. (Image adapted from Chandler et. al[176])

In the absence of helper virus, AAV assumes a latent state of infection and viral transcripts are virtually undetectable. Latency establishment and maintenance are mediated by the interaction of Rep proteins and the host YY1 protein with the AAV genome. Without helper virus, all four Rep proteins can act as transcriptional repressors from the p5 and p19 promoters[177]. The complex coordination of these opposing roles of Rep is dependent on the availability of helper virus and influences the infection state (latent or active) of AAV[178].

The first AAV vectors used to transduce mammalian cells were created over 30 years ago through seminal work from the laboratories of Carter and Muzyczska[179,180]. Since then, many groups have exploited the vectors' seemingly infinite possibilities for biological study and therapeutic use, spurring what has been termed a "vector revolution"[181]. Recombinant AAVs, like their natural counterparts, are replication-defective and non-pathogenic, and the genome assumes an episomal conformation after infection, which can persist for years[182]. The success of rAAVs has benefitted from major achievements in genetic engineering (novel expression cassette and capsid design), production, the understanding of AAV biology and discovery of new serotypes. While years of work have contributed to the field as it stands today, the design of AAV vectors remains rather simple. Essentially, the wild-type Rep and Cap genes (~4.4 kb) are replaced by a gene of interest. Traditional small-scale production of rAAVs requires three vital components, consisting of (1) an AAV vector construct (containing the transgene of interest and the AAV ITRs), (2) a packaging construct (Rep and Cap genes) and (3) helper virus functions (Ad E1A, E1B, E2, E4, VA RNA). These components are transiently transfected into E1A/E1B-containing HEK293 cells to mimic the productive wtAAV infection/replication and thereby creating high-titer stocks of infectious rAAVs. Resultant viral stocks are completely free of wild-type AAV virus, DNA, and contaminating helper virus. Large-scale production methods based on these essential process parameters will be described later in Section 1.5. The following

sections will explain the importance and possible design variations of each aspect of traditional rAAV production.

1.4.1 Recombinant Vector Design: AAV Expression Cassette (pAAV)

The basic AAV vector plasmid (pAAV) consists of a promoter, desired transgene, and a polyadenylation (PolyA) element. This sequence is flanked on either side by the AAV ITRs, which are the only viral sequences needed for proper orientation and packaging into the capsid. Any gene $\leq$ 5 kb can be cloned into this expression cassette, and there have been countless variations on this template. Although 80% of all cDNAs range from 3 to 6 kb in length, numerous groups have utilized innovative strategies to overcome the packaging limitation for larger therapeutic genes, such as dystrophin and CFTR. These approaches include mini-expression cassettes, trans-splicing or homologous recombination [183-188].

The exogenous cytomegalovirus (CMV) and hybrid CMV/chicken β-actin (CBA) promoters are commonly employed to provide robust, constitutive gene expression. However, these promoters can be susceptible to silencing over time [189]. Tissue-specific promoters, enhancer elements and hybrid promoters are utilized to customize transgene expression and provide an added level regulatory control [162]. For example, the neuron-specific platelet-derived growth factor-β (PDGF-β) promoter displayed higher transduction levels compared to exogenous promoters in the rat brain [190]. Similarly, increased muscle-specific expression has been attained with a muscle creatine kinase (MCK) promoter and α-myosin heavy chain enhancer [191].

The possibility of regulating host genes via AAV-mediated delivery of non-coding elements has also been explored. AAV expression cassettes containing sequences for short hairpin RNAs (shRNAs) and microRNAs (~21 to 25 bps in length) utilize host RNA interference (RNAi) mechanisms. AAV-encoded precursors are processed into mature single

stranded molecules, which are subsequently incorporated into the RISC complex to silence target mRNAs [162]. This strategy has been utilized to treat animal models of spinal cerebellar ataxia, Huntington's disease and ALS [192-194]. Although this strategy may be more amenable to vector capacity constraints, a balance between potency and toxicity must be achieved due to possible oversaturation of the cellular pathways [195]. Dose optimization and the use microRNAs over shRNAs appear to reduce the likelihood of toxicity [196,197].

1.4.2 Recombinant Vector Design: Packaging Construct (AAV Rep and Cap)

The AAV Rep and Cap genes are supplied by a packaging plasmid to enable replication of vector construct DNA and provide the necessary building blocks for the viral capsid. As mentioned briefly before in a previous section, careful coordination of the Rep gene is required to induce an active AAV infection and is thus of equal importance for optimal recombinant vector production. Significant work by Li et. al. revealed that unregulated Rep expression, particularly of Rep78/68, can greatly reduce rAAV yields by decreasing rAAV DNA replication and Cap gene expression[198]. In addition, Rep78 has a demonstrated cytostatic effect on host cells, inducing S phase arrest [199,200]. To solve this problem, Li and colleagues introduced an inefficient translation initiation codon (ACG) at the p5 promoter to attenuate Rep 78/68 expression, which resulted in higher titers of rAAV[198]. Commonly, the AAV2 Rep coding sequence is used for generation of most rAAV vectors, but other Rep sequences from other serotypes are being investigated to use.

In addition to regulatory elements contained within the vector construct, the choice of capsid provides yet another opportunity to customize transgene expression. To date, twelve naturally-occurring AAV serotypes, AAV1 through AAV12, have been isolated and over 100 variants have been identified[164]. Different serotypes display unique tissue tropism and

distribution, which has piqued the interest of those in the gene therapy field[201]. While AAV2 is the most commonly used serotype, alternative serotypes have been successful in preclinical studies to treat diseases such as alpha-1 antitrypsin deficiency, muscular dystrophy and heart disease, encouraging translation into clinical studies, some of which are shown in Table 2[202-205].

Moreover, the possibility of preventing off-target effects and increasing the genetic payload has fueled study into two lines of vector engineering to fit clinical applications. Rational design involves alterations to the capsid through addition of chemical moieties or site-directed mutagenesis. In doing so, antibody neutralization of rAAVs and cytotoxic T-lymphocyte detection of transduced cells within the host can potentially be avoided, increasing therapeutic efficacy, as described in detail in Section 1.2[206,207]. For example, Li et. al. demonstrated that a single amino acid modification of the AAV2 VP1 subunit increased muscle transduction and changed the neutralizing antibody (Nab) profile, which may prove useful for repeated administration[208]. In directed evolution, diverse capsid libraries are created through alteration of wild-type Cap genes. These chimeric and mosaic capsids are exposed to selective pressures, such as tissue transport barriers and neutralizing antibodies. Common techniques are error-prone PCR, DNA shuffling and degenerate oligonucleotide insertion (for reviews, see References 207 and 209). Vectors created with these methods also exhibit unique transduction capabilities and immune responses.

1.4.3 Recombinant Vector Design: Helper Virus Construct

Helper virus products are usually provided *in trans* through an adenovirus-based construct encoding genes essential for inducing rAAV replication. Other helper viruses have been explored; however, their exact functions in rAAV production are not well-characterized. For adenovirus, the known helper functions are the E1, E2a, and E4 genes and VA RNA. Previously, rAAVs were generated in cells infected with wild-type adenovirus. While critical Ad

helper functions were presented, this approach posed issues with safety and purification due to contaminating adenovirus that are replication competent, and ultimately affected the clinical quality of AAV vectors. Xiao and colleagues first described a completely virus-free technique of rAAV production, which increased titers over 40-fold compared to conventional methods at the time [168]. They constructed a plasmid containing a mini-Ad genome with the E2, E4 and VA RNA regions. The vector plasmid, packaging plasmid and mini Ad plasmid were transfected into HEK293 cells stably expressing Ad E1A and E1B, bringing about the traditional triple transfection method used in many laboratories today.

1.5 Manufacturing a Clinical-Grade rAAV

After extensive preclinical study and improvements upon the AAV vector, we have reached a crossroads between our basic understanding of rAAVs and success in clinical trials. AAV vectors are powerful tools for gene replacement and have been utilized in a multitude of clinical indications, such as retinal, neurological, muscle and metabolic disease, cancer and hemophilia. However, current large-scale production and purification limitations pose a considerable obstacle for obtaining the necessary amount of virus for many systemic diseases that require a considerable amount of therapeutic dose or for later phases in clinical trials, especially for those requiring a larger quantity of high-quality rAAV. As mentioned earlier, there are six large scale production methods currently being investigated for use in the clinical setting, as described in Table 3 and each method design depicted in Figure 5. Unfortunately, most of these novel production methods generate a mixed population of product virus particles and contaminating helper virus or virus-based expression systems per batch of rAAV product. These contaminating entities can affect the safety, potency, and purity of the final rAAV product and are to be subjected to further purification. According to 21 CFR 610.13, of the Electronic Code of Federal Regulations General Biological Products Standards: "Products shall be free of extraneous material except that which is unavoidable in the manufacturing process described in the approved biologics license application."[210] One concern is of replication competent AAV (rcAAV) that occurs via rare Rep protein mediated or nonhomologous recombination events between the rAAV vector DNA and the Rep/Cap DNA in the nucleus. This is shared by most methods since these methods utilize vectors that deliver Rep/Cap and vector genes to the nucleus directly. Another concern is some methods can generate ineffective product, such as empty and/or non-infectious rAAV that can induce a capsid-specific T-cell response[211]. Therefore, post-production purification methods are critical for the final product to be a highly-

infective, therapy carrying, safe-to-administer product. The FDA Cellular & Gene Therapy Guidance document on Chemistry, Manufacturing, and Control (CMC) offers guidelines for human gene therapy investigational new drugs that should be adopted by all production methods[212]. This information is summarized in Table 4. If viruses are used in the production system, with or without a therapeutic gene, a Master Viral Bank (MVB) and Working Viral Bank (WVB) should be created and characterized[213]. Furthermore, when a cell line is to be used over many manufacturing cycles, two banks of cells should be established: A Master Cell Bank (MCB) created from a single clone and characterized and extensively tested for contaminants such as bacteria, fungi, and mycoplasmas; and the Working Cell Bank (WCB), where cells are expanded from the MCB to use for the manufacturing process and is tested for cell viability[213]. According to the FDA, the more information that is provided to them will allow for a more thorough assessment of the identity, quality, purity, and potency of the product. The information they have detailed in regard to the manufacturing process are detailed descriptions of where and how the gene therapy is manufactured, by including all the components and materials used during the manufacturing of the gene therapy product (e.g. vector, cells, cell bank, reagents, excipients), as well as the procedures (e.g. purification, preparation of cell banks, vector derivation)[212]. If FDA guidelines are held to consistently yield clinical-grade vectors, the approval process for a rAAV based therapy can be accelerated and find its way to sooner treat afflicted patients nationally.

Table 4: Recommendations Offered by FDA Guidance for Industry

Nonbinding Recommendations for Product Manufacturing: Material Information to be Submitted by Sponsors		
Materials Used in Production Method	**Information to be Provided**	**Details**
Vector	Construct	Gene map (with relevant restriction sites), gene insert, regulatory elements, selection markers
	Sequence	Fully sequenced if <40 kb, including for those used in MVB and MCB
		Summary of sequence: analysis of gene insert, flanking regions, any regions of vector that are modified
Cell Bank System	Master Cell Bank (MCB) and Working Cell Bank (WCB)	History, source, derivation, characterization, frequency at which testing is performed, processes critical to product safety (microbiologic characteristic, presence of pathogens, culture conditions use, method of introduction of vector, analysis and selection of cell clone, cryopreservation, and genetic and phenotypic stability)
Viral Bank System	Master Viral Bank (MVB) and Working Viral Bank (WVB)	History, derivation, culture conditions used during tissue culture scale up, processes critical to product safety (microbiologic characteristic, presence of pathogens) tests to identify presence of replication competent virus, identity testing to establish the presence of gene therapy vector and therapeutic transgene, and cryopreservation
Reagents	Concentration of Reagent	Concentration reported by sponsor or supplier
	Vendor/Supplier	Manufacturer or supplier of reagent used in method
	Source	Whether the reagent is derived from human, porcine, or bovine (e.g. albumin, sera)
	Reagent Quality	Quality assurance of reagent (e.g. percentage purity, sterility, etc.)
	Certificates of Analysis (COA) or Cross-reference letters	If using research grade, and not FDA-approved or cleared, reagent then it is recommended to provide information regarding source, safety, and performance of reaction

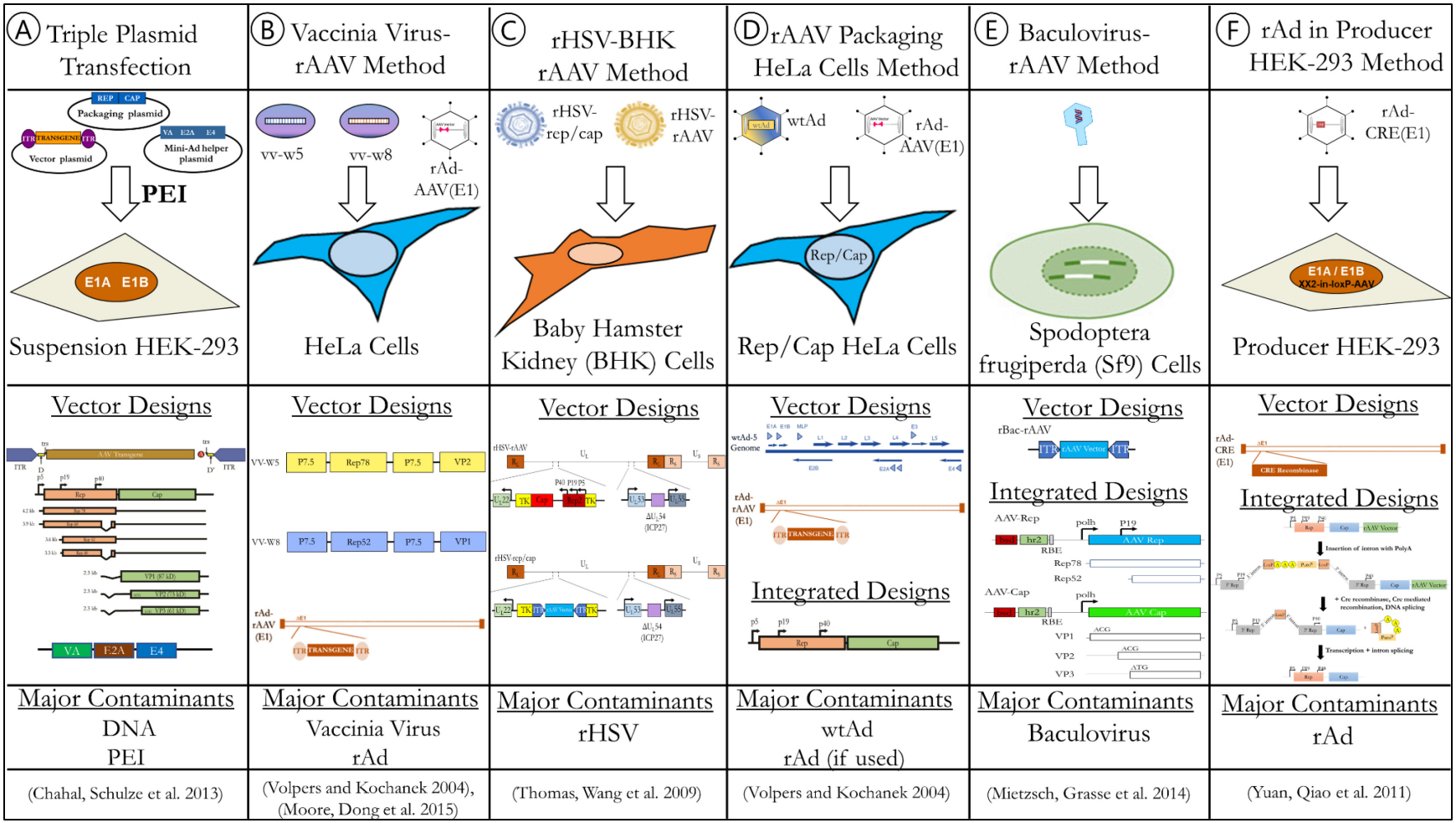


Figure 5: Large-Scale Manufacturing Methods for rAAV Production. This figure shows the most popular large-scale manufacturing methods used for rAAV production, including a visual display of the number of vectors, the cell type, the vector designs, the integrated designs, and the major contaminants of each system.

1.5.1 Triple Plasmid Transfection in HEK293 Cells

Plasmid DNA can be coprecipitated with calcium phosphate and is a well-established and simple method to achieve exogenous DNA transfer and expression in mammalian cells [214]. As discussed in earlier sections, the vector, packaging, and mini Ad plasmids can be co-transfected using calcium phosphate into adherent E1A/E1B containing HEK293 cells to generate rAAV. Although avoiding helper virus, production on a per cell basis is inefficient since three unique plasmids must hit a single cell. In order to achieve reliably high productivity, particular attention to serum conditions, cell confluency, and time of harvests post transfection must be controlled [168]. If followed correctly, this method can reliably yield rAAV upwards of $1x10^5$ vg/cell and is the method of choice for production in preclinical studies and early-phase clinical studies. With its success in creating adequate viral vectors in laboratory scales, there have been numerous attempts to scale this method. However, this is very difficult and laborious to do for adherent HEK293 cultures and there are only a few technologies that can scale adherent cultures to about 3 L, with roller bottles or Nunc cell factories [215,216].

According to some estimates, for typical GMP manufacturing efforts, more than one hundred CellSTACKs can yield about $1x10^{15}$ of clinical rAAV product. If the therapeutic effect requires $1x10^{11}$ to $2x10^{12}$ vector genomes per kilogram of the patient, the $1x10^{15}$ vector genome product would yield enough for up to 100 100-kg patients. If the virus generated is low potency, say about 10-fold less than vectors used in the previous example, the total product generated could treat as little as 5 100-kg patients[217]. These CellSTACKs, sold by Corning Life Sciences, are 636 cm^2 growth area and are available in a variety of stacking configuration used for adherent cell culture[218]. The team responsible for a rAAV clinical trial in Hemophilia B used 432 10-stack CellSTACKs, equating to $2.7x10^6$ cm^2 of surface area for the adherent cultures of

HEK293 with an overall yield of about $2x10^{15}$ vector genomes. The group used calcium phosphate-mediated transfection and adherent HEK293 cells[211]. To equate this quantity of virus necessary for a patient, if a typical rAAV lab generates $1x10^{13}$ total vector genomes for their small-animal studies, it would require 20 to 40 15-cm culture plates (surface area: 177 cm^2). If the same plate system was used for generation of rAAV for the Hemophilia B clinical trial, a total of 15,550 15-cm plates would be necessary. With regards to both the rAAV product quality - in terms of infectivity and full to empty particles - and the overall yield – in terms of vector genome copy numbers – there are several factors that are difficult to control from system-to-system (plates vs. CellSTACKs vs. roller bottles) or person-to-person. Such things are: batch-to-batch inconsistencies in transfection efficiencies, batch-to-batch inconsistencies in cell-density at time of infection, space requirements, and man-power. An estimated manufacturing cost is provided in Table 5. It should be noted that this estimation only includes cost using the materials necessary to produce the virus in a triple plasmid calcium phosphate transfection, and not materials nor resources to purify the product to clinical grade. The assumptions made in this analysis revolve around making about $2x10^{15}$ vectors for every batch no matter the clinical indication. With this in mind, the analysis was done to illustrate the cost of systemic diseases, especially at the scale of vector being produced and not necessarily to show cost for a localized disease because the amount of vector assumed to produce per batch is in great excess to what is actually required for a locally occurring indication.

Table 5: Estimated Manufacturing Cost Per Batch Using Triple Plasmid Transfection Adherent Cultures

Assumptions	**Popular Diseases Investigated in Gene Therapy Clinical Trials**				
(1) Estimated Employee Cost Based on Roller Bottle Method (ref: 219): Salaries $84,240.00 Overhead $46,980.00 Consumables $74,520.00 (2) Each production method yields $2.00x10^{15}$ vg rAAV (ref: 211) from a surface area of 2,750,000 cm^2 (3) Population estimates from references: 147,220-223 (4) Vector estimates from references: 224 (5) Method parameter references: 211, 218, 225, 226	**Estimated Disease Parameters**	**Gene Therapy Clinical Indication**			
		Duchenne's Muscular Dystrophy	Hemophilia B	Parkinson's	Leber's congenital amaurosis
	Average Dose Per Patient (vg)	$1x10^{15}$	$5.65x10^{13}$	$6.46x10^{11}$	$5.82x10^{10}$
	Patients in US Market	16,000	4,000	1,000,000	3,500

Triple Plasmid Transfection Method Used for Manufacturing rAAV			**Production Cost for $2x10^{15}$ vg per Batch per Clinical Indication** *(Total Cost of Production Using Manufacturing Method for All Estimated Patients of Disease Indication)*			
Culture Surface Method	**Method Parameters**					
Corning CellSTACKs	*Surface Area per CellSTACK (cm^2)*	636	$47,131,891,200	$665,737,963	$1,902,950,107	$600,048
	Number Used	432				
	Cost Per CellStack	$151.53				
	Total Cost of CellSTACKs	$654,610				
	Total Cost	$5,891,487				
Roller Bottle	*Surface Area per Roller Bottle (cm^2)*	1700	$1,854,576,000	$26,195,886	$74,878,506	$23,611
	Number Used	1620				
	Cost Per Roller Bottle	$16.10				
	Total Cost of Roller Bottles	$26,082				
	Total Cost	$231,822				
Traditional 15-cm Culture Plates	*Surface Area per 15-cm Plate (cm^2)*	156.4	$4,510,127,923	$63,705,557	$182,096,415	$57,420
	Number Used	17572				
	Cost Per 15-cm Plate	$3.56				
	Total Cost of 15-cm Plates	$62,641				
	Total Cost	$563,766				

The traditional method of transfection in adherent HEK293 cultures is the basis of Spark Therapeutics' Luxturna™ (voretigene neparvovec) gene therapy product, mentioned earlier. From their briefing document with the FDA on October 12th, 2017, Spark disclosed a manufacturing flow chart to the FDA Advisory Committee in hopes to obtain a recommendation for approval. Figure 6 shows a snapshot of the document, showing the typical process flow of manufacturing processes. It should be noted the various steps required to make a final FDA grade product, starting with culturing of the cells, the physical generation of rAAV with transfection methods described in this section and shown in Figure 5, collection of all cell products including the media in which the cells were cultured in, a crude lysis of the collected material, a concentration step to reduce the overall volume of product being worked on, another lysis step to break any remaining debris that can damage the chromatography technique, chromatography to concentrate and collect the sample, CsCl gradient centrifugation to remove empty particle contaminants from the desired full particle rAAV, and finally exchanging of buffers to remove the CsCl from the gradient step[147]. The purification methods will be discussed in Section 1.6, but it should be noted that this industrial scale is to generate product for the eye, therefore requiring much less viral product than compared to more systemic diseases like Duchenne's Muscular Dystrophy or Hemophilia. If this production method were to be used, it would certainly drive up the price of the therapeutic for systemic diseases. BioMarin Pharmaceutical, who is developing their own rAAV therapeutic for Hemophilia A, is aware of this concern and are relying on the baculovirus technique to scale up their product after approval. Amit Nathwani, the investigator who led the University College London team to develop the treatment, posed a concern shared by all researchers behind systemic therapies such as BioMarin Pharmaceutical's product: "Will they be able to manufacture enough to support the world's hemophilia A market?"[227]

Figure 16: Voretigene Neparvovec Manufacturing Process Flow Diagram

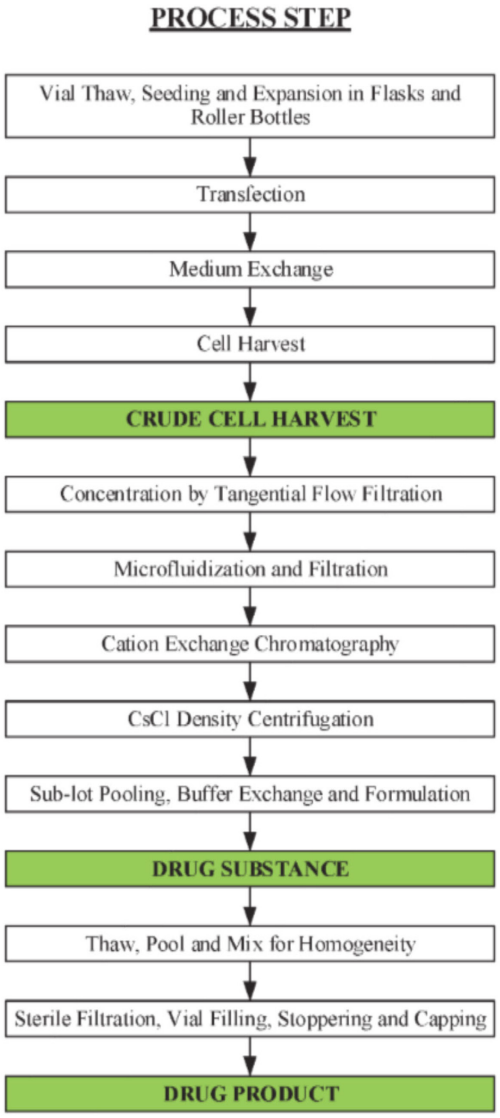


Figure 6: Process Flow Diagram of Gene Therapy Product, Luxturna™, for FDA Approval

Recently, a current Good Manufacturing Practice (cGMP) approved serum-free suspension HEK293 cell line has been developed for rAAV production by triple plasmid transfection using polyethyleneimine (PEI, 25-kDa), a cost-effective cationic polymer DNA transfection agent that can effectively deliver genes both *in vivo* and *in vitro* [151,228]. Figure 5A shows the diagram of the system, which is like tradition methods of triple plasmid transfection, only with PEI used instead of calcium phosphate ($CaPO_4$). In the context of the production process, PEI transfection methods are also useful because it is well-tolerated by HEK293 cells, does not require a media change after transfection that are required for calcium phosphate transfection, and is inexpensive to commercially available lipid transfection reagents like Lipofectamine™ ($330 for 1L of 25-kDa PEI vs $4,835 for 15 mL of Lipofectamine™ 2000) [229,230]. This scalable method can generate rAAV much like its adherent cell culture predecessor without the use of sera, antibiotics, or medium exchange post-transfection. However, the optimal density for transfection is low and selection of a cell that shows high transfection capability and high rAAV vector production per cell production efficiency can be challenging [231]. Consequently, batch-to-batch inconsistencies in transfection efficiencies and viral titers can occur. This is of concern regarding FDA guidelines, since Master Cell Banks (MCBs) and Working Cell Banks (WCBs) should be adequately established as a safe, pure, and stable cell line [212]. Furthermore, the implementation of purification is critical in PEI systems, since PEI itself is nonbiodegradable and can be moderately cytotoxic. In 2016, work by Grieger et. al. has led to a qualified clinical master cell bank derived from a clone selected for high transfection efficiency and rAAV production. They've improved on the previous system by exploring other preparations of PEI, namely a PEI preparation called PEI Max that contains more than 11% additional free nitrogen and was able to increase transfection efficiency. Their system involves a simplistic method to generate clinical rAAV vectors, covering all facets of the production

process including cell culture, transfection methodology, and purification strategies[232]. Furthermore, the recently innovated, low-cost, scaled production of plasmid preparations has made this system even more advantageous. According to Genetic Engineering & Biotechnology News, Nature Technology's HyperGro™ is an inducible fed-batch fermentation process that can enable high yield production of optimized plasmids and even certain unstable or toxic vectors[233]. One disadvantage to other systems is the need to generate two- to three-unique high-grade plasmids for transfection. Despite these pitfalls, this method is comparable to other methods in a small-scale bioreactor and has all prerequisites to enable rapid and scalable rAAV production for large-scale manufacturing campaigns[217,228]. Advantages to this system is there is no need to generate and characterize several cell clones for Master and Working Cell Banks, and there is no use of viruses, so there is no need to document viral banks (Master Viral Bank and Working Viral Banks). Unfortunately, there remains the issue exactly found in traditional transfection methods: a single cell must be hit by three unique, GMP grade DNA plasmids for rAAV production, affecting the per-cell efficiency and overall material cost.

1.5.2 Recombinant Baculovirus-rAAV into Insect Packaging Cells

The baculovirus/insect cell-based technology is a popular avenue for rAAV vector production, and has been rooted into the manufacturing process for the rAAV companies UniQure and BioMarin Pharmaceutical Inc. for their rAAV products[148,227]. The OneBac system, the latest update in the technology, is attractive to generate large-scale rAAV due to its easy implementation, scalability (up to 200 L in recent reports), suspension and serum-free media cultures, growth at 27°C, and its current use in protein manufacturing [234,235]. At a low multiplicity of infection (MOI), a single baculovirus, derived from the *Autographa californica* nuclear polyhedrosis virus (AcmNPV), containing the rAAV vector is infected into a stable transcriptionally silent inducible Rep/Cap packaging *Spodoptera frugiperda* (*Sf9*) cell line. A variety

of these cell lines with serotypes AAV1 through AAV12 Cap genes have been made. The *Sf*9 packaging cell lines are genetically stable and lead to undiminished rAAV yield, albeit questionable quality [236]. The baculovirus itself, a rod-shaped 260 nm length by 20 nm diameter 134 kb virion, has significantly less safety concerns than other viral vector counterparts. If there were baculovirus contaminants in the end-product, the virus does not replicate inside the transduced cells, nor integrate its DNA into host chromosome. Humans do not possess pre-existing antibody and T-cells against baculovirus, and the virus is nonpathogenic to humans. In fact, the baculovirus can be readily constructed and propagated to high titers in biosafety level 1 facilities[237,238]. However, in compliance with industry guidance documents – specifically the viral safety evaluation of biotechnology products derived from cell lines of human or animal origin (ICH Q5A) – the process for purification of a biological pharmaceutical can remove any non-product virus, to avoid the potential pathogenicity or immunogenicity of the non-product virus. Removal of these viral contaminants are performed by "viral clearance" or "viral removal" process steps usually by chromatography or virus filtration. Viral inactivation can be done to decrease potential pathogenic effects caused by non-product viruses, but these usually require extreme treatments – e.g. pH, temperature - or chemical conditions – e.g. detergents, solvents[239]. For the case of baculovirus, removal of the contaminating rBac-rAAV is a challenge that faces companies like UniQure or BioMarin Pharmaceutical who use this scalability technique as their platform. In fact, UniQure filed for a patent on such a technology to comply with the European Medical Agency (EMA), and meet the needs for industrial setting. Their technology attempts to clarify the product via the physical shape of baculovirus, rod-shaped, and the rAAV, essentially spherical. Their invention uses two steps, the first is a pre-purification step by the following, individually, or in combination: density gradient separation, prefiltration, chromatography. Next, a virus filter/ultrafilter would be used to only allow virions of an

essentially spherical shape to pass in a pre-designed filter with a nominal pore size of 35 nm, such as the Asahi-Kasai Planova 35N membrane, or molecular weight cut-off of 10,000-100,000. Despite the efforts to reduce overall baculovirus quantities in the final preparation, these separation and filtration techniques does not guarantee complete removal of the baculovirus and thus leaves the possibility of its existence in the final product. The invention indicates that a removal of 5.5-10 orders of magnitude is preferred with their technique, to which they can detect using an infection assay. In their patent, they indicate a reduction of at least 6 orders of magnitude[240].

Expression of the Rep gene must be silent because of the cytotoxic property of Rep78. It is strictly controlled by the Bac-derived polyhedron promoter polh and the cis-acting enhancer element hr2-0.9 placed upstream. The Cap genes were separated, and expression was also strictly controlled by this same device [236,241]. Figure 5E shows the vector design and the integrated design. polh expression is only induced upon baculovirus infection by the immediate-early gene-mediated transactivation of the hr2-0.9 enhancer. An AAV Rep-binding element (RBE) was inserted adjacent to the hr2-0.9 enhancer because it led to enhanced viral protein expression. This is due to a feed-forward loop that is initiated when Rep78 is expressed that up-regulates the integrated Rep (Rep78 and Rep52) and Cap (VP1, VP2, and VP3) genes by interaction with RBE, and therefore leading to larger titers compared to the previous three- to four-baculovirus infection system. Since the Rep52 sequence is in the Rep78 gene, its expression remains under control of p19, and enhanced by RBE. This allows production of both Rep78 and Rep52 from a common open reading frame (ORF) only after activation of polh from the hr2-0.9 enhancer. There was little to no expression of Rep68/Rep40 through western blot likely from *Sf*9 cells having poor splicing recognition of mammalian signals in the AAV Rep gene [235,242]. Cap proteins VP1, VP2, and VP3 are normally synthesized from two differently spliced mRNA in

wild-type AAV. These splicing signals were eliminated by expressing all three VP polypeptides from a single transcript and in the correct ratio by mutating the start codon for VP1 from ATG to ACG, and incorporating downstream start codons ACG for VP2 and ATG for VP3 [235,236]. Attention was made for VP1 - critical for AAV particle infectivity – which notoriously shows lower levels in baculovirus-rAAV production systems and correlates to an increase in non-infectious rAAV product [236].

Although currently used for commercial protein manufacturing and by UniQure and BioMarin Pharmaceutical, the use of this method for clinical-grade rAAV production is questionable. First, the infectivity/transduction efficiency of resulting OneBac made rAAV is not comparable to mammalian made counterparts, despite advances in VP1:VP2: VP3 expression levels, especially for serotypes AAV3, AAV5, AAV8 and AAV9[235]. This can again be attributed to the issues with VP1 expression from an inefficient VPs expression design in the capsid transcript[217]. Second, the ineffective splicing signals in *Sf*9 cells are crippling to the final product which would contain a noticeable amount of contaminating empty AAV particles. The over expression of capsid proteins or inefficient AAV genome packaging is another problem with this system, also leading to a large amount of contaminating empty rAAV particles. This can be from the noticeable loss of Rep68 (associated with replication of the AAV transgene) and Rep40 (associated with efficient packaging), despite efforts of Rep78 and Rep52 to substitute their loss[243]. Third, the possibility of different posttranslational modifications of the rAAV capsid made in insect cells, as compared to mammalian cell counterparts, can impact the final product[18,231,243-245]. Indeed, we can assume these attributes are culminating to the product that BioMarin Pharmaceutical has generated for its FVIII rAAV, where they require an astounding $6x10^{13}$ vg/kg per patient for their Hemophilia A (recall that triple plasmid

transfection methods required $2x10^{11}$ vg/kg albeit for their Hemophilia B indication)[246]. Finally, no matter if OneBac is used or the alternative three-baculovirus system is used as the method for rAAV production, there remains the biggest drawback: instability of AAV cassette in baculoviruses during the expansion phase, with notable loss in ITR-transgene cassette containing baculovirus over five consecutive passages[247]. These drawbacks related to the molecular/cellular biological learning curve over the past decade has certainly made this technology limited in use in the clinical setting.

1.5.3 Recombinant Herpes Simplex Viruses Infection in Suspension BHK Cells

Recombinant herpes simplex virus type 1 (rHSV) has become another popular method of rAAV production in clinical trials [248]. The rHSV virus down-regulates cellular functions to force AAV to use the HSV replication complex to propagate instead of the cellular machinery alone[169]. 50% of the HSV genome encodes for nonessential gene products and can be deleted and replaced with large foreign DNA sequences without jeopardizing viral amplification[169,249]. It was initially proposed that the replication proteins of HSV provided the helper functions of rAAV. These were the helicase/primase complex UL5, UL8, and UL52 and the DNA binding protein ICP8 encoded by the UL29 gene. However, as more is elucidated about HSV, other proteins can enhance helper functions in combination with the replication proteins. ICP0 transactivator activates Rep gene expression, ICP4 and ICP22 enhances AAV replication in the presence of six HSV replication genes, and the polymerase UL30 and co-factor UL42 can contribute to efficient AAV replication[169,250].

The most-advanced version of rHSV infection systems are based on infection of two rHSVs into baby hamster kidney (BHK) suspension-adapted cells, HeLa cells, or HEK293 adherent cells [169,249]. These two vectors contain all cis- and trans- acting elements necessary for

rAAV production and include helper functions naturally found in HSV, but uses the d27.1 variant of HSV that lacks ICP27 expression: one HSV vector containing rAAV Rep/Cap genes, the other HSV vector harboring the rAAV transgene with its respective rAAV inverted terminal repeats (ITRs), and helper functions supporting replication of the rAAV genome including activation of Rep gene expression. The rHSV has stability to carry Rep expression without severely impacting replication of rHSV stocks especially during rHSV serial expansion in V27 cells, the cells responsible for replication of rHSV [153,231]. The V27 cells, also known as Vero cells, are used to propagate rHSV because these cells are stably transformed with the UL54, specifically containing ICP27, the gene deleted in the rHSV vector responsible for replication of HSV[217,251]. The scalable version requires lower MOI of rHSV, all the while can be used in higher cell densities during scale up of BHK cells, decreased time of harvest, a high yield (reports of $2x10^4$ to $1x10^5$ vg/cell which is greater than transfection methods of approximately $1x10^4$ vg/cell), scalable to 100 L bioreactors, and can accommodate multiple AAV capsid serotypes without loss of infectivity[215,217,231,235,252,253]. The resulting rAAV product has increased potency, as seen by vector genome unit over infectious units and increased cell transduction expression, and overall reduction in empty capsids.

Several issues are associated with the rHSV technology. First, the rHSV vectors have a large genome size (152 kb) and thus difficult to fully characterize, a particular concern in FDA guidelines[212]. This is much like the baculovirus system, where a significant amount of time and effort is needed to characterize rAAV generation at all scales. It is also unknown if HSV DNA is packaged into AAV or if there is generation of replication competent AAV (rcAAV). Next, rHSVs have potential to be pathogenic and/or immunogenic. Efforts have been made for replication defective rHSVs, for instance by using the d27.1 variant, but generating a high-titer infectious stock of these rHSV variants is difficult and the overall yield is greatly diminished.

Nevertheless, the assurance that the rAAV product is free of adventitious agents, in this case the rHSV, is critical to meet the 21 CFR 610.13 requirement for purity, quoted earlier. To the concern of FDA, HSV has tropism to the central nervous system and has a potential for latency and reactivation[254]. Methods to do this are suggested by certain FDA guidance documents, and these tests involve *in vitro* cell culture tests in various cell cultures, like Human Diploid cell lines or Monkey Kidney Cell Lines to test for human viruses, and could also include transmission electron microscopy, PCR, or other specific *in vitro* tests[255]. Removal of these viruses could be from chromatography, or using nanofilters since rHSV are 120-200 nm in size, about 6 to 8 times bigger than rAAV. Third, rHSV particles can be easily inactivated to production and processing conditions due to capsid fragility of the enveloped virus[231]. This is a significant challenge when establishing a FDA standard Master Viral Bank (MVB) and Working Viral Bank (WVB) for both rHSVs used[243]. For a Master Viral Banks, the seed vector should be tested for genetic integrity and stability, bioactivity of the vector should be determined, or expression of the gene to be assessed by the sponsor. Finally, the requirement of two viruses to infect a single cell is a drawback to this technology on a per-cell basis.

1.5.4 Recombinant Vaccinia Viruses and Adenovirus-rAAV Infection into HeLa Cells

The newest development of rAAV production uses vaccinia virus (VV) to carry helper DNA of AAV into the cytoplasm, and not the nucleus of the host cell, in hopes to eliminate any concerns of replication competent AAV (rcAAV). Vaccinia virus is a member of the poxvirus family, has dimensions of 360 × 270 × 250 nm, undergoes its entire life cycle in the cytoplasm of the host cell, and is most commonly known for its use in eradicating smallpox[256,257]. With the knowledge of VV using solely the cytoplasm for its viral life cycle, Moore et. al. and Dong et. al. developed a novel approach utilizing two recombinant VVs for rAAV production in HeLa cells, but not limited to this cell line. In one recombinant VV are the Rep78 gene and both VP2

and VP3 genes expressed from a single promoter, the vaccinia p7.5 promoter. In the other recombinant VV are the Rep52 and VP1 genes, also driven by the p7.5 promoter. It is possible to accommodate all genes into one VV, since VV can tolerate >20-25 kb inserts into its final 200 kb genome. These Rep and Cap DNA are transcribed and translated in the cytoplasm and the resulting proteins are migrated into the nucleus where it mediates rAAV replication and packaging. The rAAV vector DNA can be incorporated into the host chromosome or delivered to the nucleus by an adenovirus harboring the AAV vector, overall resulting in a high yielding rAAV preparation, devoid of rcAAV. Furthermore, the contaminating VVs can be eliminated by simple 0.22 μm filtration since the VVs diameter is approximately 0.4 μm [152]. Furthermore, if VVs are present still after filtration, the density of the VV is 1.24-1.27 g/cm^3, which is significantly lighter than AAV (1.41 g/cm^3). This means that in a gradient centrifugation for polishing steps of rAAV purification, VV can be removed without major qualms. After removal of the VV contaminant, the remainder of the process can follow typical purification methods to obtain the clinical-grade vector product[152].

Reports have stated that vaccinia virus can provide helper functions to rAAV, but it should really be categorized as a sub-helper. Vaccinia fails to produce rAAV and activate AAV promoters, but it could initiate replication and packaging when AAV promoter activation is not necessary. This is because of the p7.5 vaccinia promoter that drives expression of Rep78 and Rep 52 in the cytoplasm and translates in the nucleus. The DNA that is delivered via VV will not have an opportunity to recombine with the rAAV vector to generate rcAAV. Estimates have suggested that 1x10^{16} rAAV vectors can be made from just 100 L of suspension HeLa cells, but this technique remains in a proof-of-concept stage and these estimates have not been verified.

Since this is relatively new, there remain some concerns to this method. First, the removal of Rep68 and Rep40 may provide issues in contaminating empty AAV particles, like in

OneBac discussed earlier. Removal of Rep68 and Rep40 appears to be a common strategy for systems that require a more stable recombinant delivery vector of the Rep genes, however it is unclear if this would explicitly affect the rAAV quality and the potency of the eventual clinical grade product. Second, this method requires at the minimum two different vectors for adequate rAAV production, reducing per cell production efficiency. The vectors in this system are the optional helper adenovirus-rAAV hybrid and the mandatory two VVs containing rAAV Rep78, Rep52, and Cap genes. Adenovirus can be avoided by alternatively having an integrated cell line with the rAAV vector gene integrated into the cell and subsequent infection with the two VVs. In larger scales, removal of the VVs is simple, but the removal of contaminating adenovirus is far more challenging. In this situation, manufacturers would more likely integrate the rAAV into the genome of the cell, rather than use recombinant adenovirus to deliver this vector into the nucleus of the cell. Third, the use of vaccinia virus may pose some safety concerns for development of this manufacturing process, and not necessarily for the final rAAV product since removal of VV is relatively simple. Although this virus is of the MVA strain, which is an attenuated strain of wtVV unable to replicate and form infectious virus in mammalian cells, contamination of stocks of avirulent viruses with replication-competent poxvirus is of moderate concern and proper measures should be enforced to reduce possibility of accidental infections[167]. According to FDA guidelines this poxvirus, albeit replication-deficient, may have the: "(1) ability to infect and replicate in many types of human tissues and cells, (2) potential for toxicity in immune-compromised populations such as cancer patients, and (3) renal/cardiac concerns."[254] Safety measures that should be followed include Biosafety Level 1 (BSL-1) for attenuated strains (or BSL-2 if virulent strains), use of a biosafety cabinet, and proper personal protective equipment (including lab coat, gloves, and eye protection)[258]. Finally, the large VV genome size (~200 kb) would follow similar issues as rHSV technology, and may be of concern

in FDA guidelines, especially for developing a MVB and WVB. All told, this method provides a unique alternative to generate a scalable, cost-effective method relative to triple plasmid transfection. Further investigation is necessary to verify that the rAAV product is of clinical-grade quality.

1.5.5 Wild-Type and Recombinant Adenovirus Infection into HeLa Producer Cells

First used as a vector to assist in rAAV production in 1995, adenovirus continues to be of interest for large scale production of rAAV. Adenovirus (~30-40 kb, ~70-90 nm diameter) promotes AAV replication directly by trans-activators that stimulate AAV Rep gene expression, and indirectly by driving the cell cycle to enter S phase, which provides AAV with active cellular replication machinery. Furthermore, adenovirus can be used at a low MOI, is suitable for suspension culture, it can be produced in animal-free conditions, easy to produce at high yield, well-established purification process in small-scale settings using ultracentrifugation, the capsid is stable (unlike rHSV or baculovirus), the size is much smaller than rHSV and baculovirus, the genome is stable, has fully characterized helper functions, there is no or little rcAAV generated, and adenovirus DNA is not packaged into AAV[243].

Rep/Cap and rAAV vector genes were stably integrated into HeLa cells, generating a producer cell line. A high titer of rAAV was achieved from these rAAV producer cells following an infection of wild-type adenovirus or an E1-intact adenovirus to activate the p5, p19, and p40 transcription units of Rep/Cap gene [259]. Further improvements were made by replacing the E1 region of an adenovirus with the rAAV vector and infecting both this recombinant adenovirus vector with the wild-type adenovirus into the stable Rep/Cap HeLa packaging cell line, also generating a high yield of rAAV [154,260]. HeLa-based cell lines can be adapted for suspension culture and grown in serum free media. Since the HeLa cells do not harbor E1 genes, the p5 and p19 promoters were not able to be activated and potentiate cytostatic or cytotoxic effects of

Rep78 to the cell, and thus HeLa cells are a prime cell line to generate a stable packaging or producer cell line[261]. According to recent estimates, this method can supply highly pure vector in the order of $1x10^{16}$ DNase resistant particles for clinical trials at a volume of 250 L. Furthermore, this report projected no impediments to scale up to commercial manufacturing in 2000 L bioreactors [262].

However, the contamination of wild-type adenovirus in the final vector preparations is highly undesirable for vector safety [263,264]. According to the FDA, adenovirus has the potential for a significant immune response and inflammatory response to the vector and possible adverse effects from any contaminating replication-competent adenovirus, such as wild-type adenovirus[254]. The FDA guidance document recommends an appropriate maximum level of replication competent adenovirus would be <1 in $3x10^{10}$ viral particles, with an adenoviral particle to infectious unit ratio (vg:IU) of ≤30:1[212]. At a large scale this can be a problem, since there may be difficulty in removing contaminating adenovirus from vector preparations while keeping the high yields of rAAV. The methods to remove the contaminating adenovirus would be done using density gradient centrifugation (adenovirus density 1.35 g/cm^3 and AAV density 1.41 g/cm^3), but this is only economically feasible on small volumes, not industrial scaled volumes and the minor difference in densities would mean a narrow proximity to one another in the density gradient. Heating a mixed population of adenovirus and rAAV to 56 °C – 65 °C (e.g. a 30-60-minute incubation of the rAAV vector preparations to inactivate adenovirus effectively) can lead to more selective inactivation of the helper virus without effecting the rAAV infectivity. However, the denatured helper virus proteins from the adenoviral capsid would still be present in the product and upon use the gene therapy would evoke a cellular immune response in the patient[265,266]. Removal of the helper virus can be performed using ion exchange and affinity

chromatography, or using a filter membrane with a pore size of 35-50 nm, since AAV is 20-25 nm in diameter and adenovirus is 65-90 nm[267,268]. Another safety concern regarding the use of HeLa cell cultures is the presence of human papilloma sequences, the HPV 18 sequence [243,244]. Finally, much like the rHSV system, the system is also inefficient regarding number of viruses to infect a single cell.

1.5.6 Recombinant Adenovirus Infection into HEK293 Producer Cell Line

The existence of the adenovirus 5 E1 region integrated into the HEK293 genome has made generation of a stable Rep/Cap packaging and rAAV producer cell line difficult due to the E1 activation of the Rep gene, that leads to generation of cytostatic or cytotoxic properties from Rep78[261,264]. In the Xiao lab, they have established an innovative inducible Rep expression using a dual splice or intron splice switch. Briefly, the Rep/Cap gene had been inactivated by an intron insert in between the Rep gene. Between the intron signals is an insert of a LoxP cassette containing PolyA termination signals to inactivate splicing of the intron. A rAAV vector was included downstream of this switch design, carrying the rAAV transgene flanked by the appropriate ITRs, and together this plasmid was integrated into the HEK293 cell genome, resulting in a producer cell line. An infection of an E1-, E3-deleted adenovirus harboring a Cre recombinase gene in the E1 region (hereafter, Ad-Cre (E1)) into this cell line results in a Cre-LoxP site-specific recombination with the LoxP sites, and removal of the PolyA termination signals. After mRNA splicing to remove the inserted intron, the Rep/Cap genes are restored to its normal function. [156,269]. E1A/E1B of the HEK293 cell activates the expression of Rep, Rep and Cap proteins together identify the ITRs, package the vector into the capsid, and the resulting rAAV can be collected and purified at a high titer and high potency [261]. This method has an improved safety profile by removing the need of a wild-type adenovirus because of the E1 genes

present in HEK293. Recombinant adenoviruses are used often in clinical trials, but in this method, removal of the adenovirus is necessary for increased safety in the final rAAV product.

This approach follows similar issues to those in the previous adenovirus technique as well as innate issues found in the system developed by Qiao et. al. and Yuan et. al. First, it requires a time-consuming selection and characterization process to identify the best cell clone, a critical process for MCBs and WCBs. Furthermore, for every novel rAAV product to be generated, a unique HEK293 producer cell line has to be generated. Second, currently this method uses adherent HEK293 cells that rely on sera for cell growth, and therefore only scalable to about 3 L, e.g. roller bottles or Nunc cell factories, and has not yet been optimized for serum-free, suspension HEK293 cultures. There should not be any major complications to transition from adherent to suspension cultures, as evident by work in the suspension transfection methods, but this has not been performed for these particular cell lines. Next, the recombinant adenovirus is harvested alongside the resulting rAAV. Although it is an E1-deleted adenovirus, a widely used type of vector in gene therapy for humans, the virus can recombine with the E1 region integrated in HEK293 cells and become replication-competent[270]. If it is not adequately removed the recombinant adenovirus can cause induction of innate immune responses characteristic of the adenovirus capsid interaction with cells [265,266].

1.6 rAAV Large Scale Purification Methods

During harvest in smaller scale productions, the cells are collected separately from the 1-3L of media which is also collected since rAAV is released from cells into media [271,272]. The cells are lysed by freeze-thaw or sonication releasing rAAV in or attached to the cells into the viral resuspension buffer and the rAAV in media are concentrated and collected. Both preparations are combined and subjected to two to three ultracentrifugation spins. The first ultracentrifugation spin removes most of the contaminating materials via density differences [273]. The second or third spins are specifically required to purify and concentrate full transgene containing rAAV from the empty rAAV, by density differences in empty (~1.32 g/cm^3) and full rAAV (~1.40 g/cm^3), regardless of serotypes [211].

As the volumes of collected post-harvest material increases, using ultracentrifugation for all purification steps would be resource intensive and time-consuming. Therefore, several upstream purification processes must be used ending with a polishing step or ultracentrifugation. Many academic and industrial researchers are investigating purification strategies to streamline purification, especially when certain manufacturing methods approach industrial scales[232,274]. The use of the nanofilter is a promising developing technology for purifying mixed populations of larger non-product viruses and the desired product virus. However, the material must be pre-filtered with a larger pore-sized filter (e.g. 0.1 μm) before it encounters the nanofiltration system such that of all other contaminating entities are removed, elsewise the nanofilter will be obstructed by insoluble cell constituents or other non-product entities, and thus reducing yield of desired rAAV product. Given the advancements of chromatography for purification of a variety of AAV serotypes that are being used by companies like Pfizer, chromatography may be the most effective way of purifying a variety of AAV serotypes. Companies like Pfizer are using an affinity ligand made from a proprietary camelid-derived single-domain antibody fragment for

AAV affinity purification. Sold as a commercially available resin, called POROS CaptureSelect AAVX resin, the CaptureSelect ligand is a 13-kDa fragment that comprises the three complementary determining regions that form the antigen-binding domain. The resin is able to be used for large-scale downstream purification for a broad range adeno-associated virus used for gene therapy applications. Reports from retailers of this resin have said this resin features one-step AAV purification from crude material, high specificity, and can handle high flow rates. However, if the solution is clarified of other crude lysate materials, the resulting viral products can be sent through this high-precision affinity chromatography to more precisely purify AAVs of serotypes AAV1 through AAV8, AAVrh10, and other serotypes without interference occurring from materials in the crude lysate. The following sections will discuss upstream steps to remove crude materials after harvesting from earlier production and it will end with a brief overview on efforts to remove empty vectors, in more realistic industrial systems.

1.6.1 Removal of Cellular, Nucleic Acid, and Viral Contaminants

After the cell fraction and media fraction have been collected, the rAAV in the media is concentrated. One choice method is tangential flow filtration (TFF), a method that can quickly reduce solution volume at least 50-fold by operating at pressures of 10-12 psi to remove salt, water, and small molecular weight protein at a desired molecular weight cutoff (usually 100 kDa) without significant loss of vectors [275]. The membrane used in the process is easily scaled up and handled for cGMP manufacturing. Furthermore, TFF is a desirable tool if the media is to be exchanged for an appropriate buffer used in the following chromatography steps. Removal of contaminating virus is a challenging avenue of research and several inventions have been made to remove these contaminating entities using size exclusion filters, as briefly described earlier for VV, baculovirus, and adenovirus[152,240,268]. It may be possible to use these filters for rHSV as well since this virus is larger than rAAV.

Recovery of the rAAV from the cell lysate fraction and the media fraction is difficult to do in one purification step and is usually divided in two column chromatography steps. The first column is used to capture rAAV from other contaminating materials. As with any chromatography, the appropriate binding resin, binding buffer and elution buffer is essential so the rAAV product is not lost or damaged while it passes in the column. The options of chromatography used in the first steps are: ion-exchange chromatography (IEX) or affinity chromatography (AC). In IEX, it is further divided into anion exchange chromatography (AEX) and cation exchange chromatography (CEX). Compared with most contaminating impurities, rAAV contains multiple sites to bind to anion exchangers, making AEX the more commonly used chromatography for rAAV purification. In AEX, transgene containing AAV vector is predominantly negatively charged when the pH is above the isoelectric point (pI). In the case of purifying empty particles, the electronegativity of the empty particle is weaker than the packaged rAAV and the empty vector will be eluted first followed by the packaged vector. With IEX, the purity of rAAV can reach up to 98% or more, the purification can be handled by automated equipment, resins can be regenerated, rAAV interaction with resins is gentle to rAAV, and scale is readily adjustable according to quantity. However, some resins are difficult to optimize for certain serotypes, the resins have a finite lifespan, and resins can be contaminated with organic materials if poorly handled. Nevertheless, AEX is being used in industry, and POROS HQ is a common industrial resin used for AEX with a variety of rAAV serotypes [211,276,277].

The common forms of AC are associated with capsid interaction with antibodies, native proteins, or with inorganic salt. Use of antibodies, called immunoaffinity chromatography, is not necessarily economical, but it can recognize several serotypes of rAAV making this type of chromatography very useful for a system that generates different serotypes of rAAV. For instance, the use of the native protein heparin, in heparin affinity chromatography, is usually

reserved for rAAV2 in large scale. Use of inorganic salts in chromatography, called apatite chromatography, uses affinity of the rAAV vector surface protein and the rAAV genome to capture rAAV of various serotypes, sometimes with assistance of PEG [278]. It is the choice method to separate rAAV and adenovirus, using a common industrial AC resin called ceramic hydroxyapatite (CHT) [277].

1.6.2 Removal of Empty Vectors

Most contaminants are removed after the first chromatography step. The second chromatography step is to get a purer rAAV, vector containing, product. This is done primarily with either IEX or using size exclusion chromatography (SEC). Size exclusion chromatography, known as gel-filtration chromatography, is a method to separate large biomolecules, such as helper viruses or any contaminating cellular material, by their size or weight in solution. It is considered a low-resolution chromatography and a 10% difference in molecular mass is enough for good separation [279]. SEC would not be able to remove empty particles since the size and shape of full and empty particles are identical. In this case, IEX would be used with either CEX or AEX to remove empty particles. However, in some cases, there are lower resolution signals between full and empty particles, making removal difficult. Although achieving a low contaminant product can be done with several runs, with every purification step that is undertaken, there will be an accumulation of product loss. It is because of this the purification ends with a final purification step, called the polishing step, wherein any other remaining contaminants and empty vectors are removed from the preparation. The polishing step usually ends in the use of ultracentrifugation to concentrate the eluted pure final product from previous steps and effectively separate it away from the contaminating materials and contaminating empty vectors. Ultracentrifugation may be necessary since it effectively separates empty virions better than current separation techniques, without as much risk of accumulated product loss.

1.7 Hypothesis and Specific Aims

The purpose of this study is to design a scalable process for production of high-titer clinical-grade recombinant adeno-associated virus (rAAV) with an adenovirus (Ad)-rAAV hybrid infected into a packaging cell line. To comply with the increased demand of rAAV, a robust, reproducible, and scalable production system that can also be adjustable for a variety of rAAV serotypes is necessary. Unfortunately, current methods of scalable production fail to do so either from uneconomical equipment/procedures, safety concerns of production vectors to assist in rAAV production, genetic instability of production vector, and/or decreased potency of rAAV.

Previously, a rAAV producer cell line was generated by our lab using human embryonic kidney 293 (HEK293) cells to increase titer of rAAV and avoid use of wild-type Ad/helper Ad since HEK293 cells contain integrated copies of the adenovirus E1A and E1B genes. We introduced and integrated an inducible AAV replication (Rep) gene, capsid (Cap) gene, and a rAAV vector into the HEK293 genome. Using an intron splice switching method previously developed in our lab, the Rep function is restored and activated by the HEK293 E1A/E1B genes leading to the production of rAAV that can be collected and purified at a high titer and high potency. However, this approach has multiple issues: (1) the Ad harboring the Cre gene in the E1 region is also deleted in an unused E3 region; (2) using the previous method requires generation of a new producer cell line for each novel rAAV vector desired to be made into the final viral form; (3) identifying HEK293 cell clones with stably integrated inducible Rep, Cap, and rAAV vector is a time-intensive process, and generally yield few ideal candidates with enough copy numbers of the integrated plasmid due to random integration of the plasmid in the genome; and (4) the Ad-Cre has potential to recombine with HEK293 E1A/E1B during infection to become replication-competent, and can only be removed in an unscalable ultracentrifugation or partially inactivated due to its thermal instability.

To circumvent these issues, three major adjustments will be proposed. First, the technology will be converted to a rAAV packaging cell line, since both regions of the E1-, E3-deleted Ad will be replaced with the rAAV vector and Cre-recombinase gene, respectively (hereafter described as Ad-Cre(E3)-rAAV(E1)). Second, site-specific nucleases that can introduce a double strand break (DSB) in specific areas of the cellular genome can allow for an efficient and precise stable cell line generation. CRISPR/Cas9, a recently developed site-specific nuclease, will be used due to its facile implementation and design to introduce a DSB at the AAVS1 integration site on Chromosome 19 (chr.19q13.3q-ter). Potentially, integration of the packaging plasmid DNA at a greater reproducibility and efficiency in the HEK293 cells will occur, and therefore characterization of cell lines can be streamlined. Third, a 2013 Merck patent has shown that adenovirus can be precipitated using a selective precipitating agent (SPA), domiphen bromide, along with cellular DNA, even at high cell densities. With these ideas in mind, we are positioned to test our general hypothesis: Ad-Cre(E3)-rAAV(E1) can produce high potent and high titer rAAV in an inducible packaging cell line and the resulting rAAV vector preparation can be purified from contaminating cellular DNA and adenovirus by precipitation using domiphen bromide. The hypothesized process is depicted in Figure 7, which primarily focuses on upstream production methods and briefly delves into upstream crude lysate clarification. The realm of this dissertation will not focus on downstream purification processes.

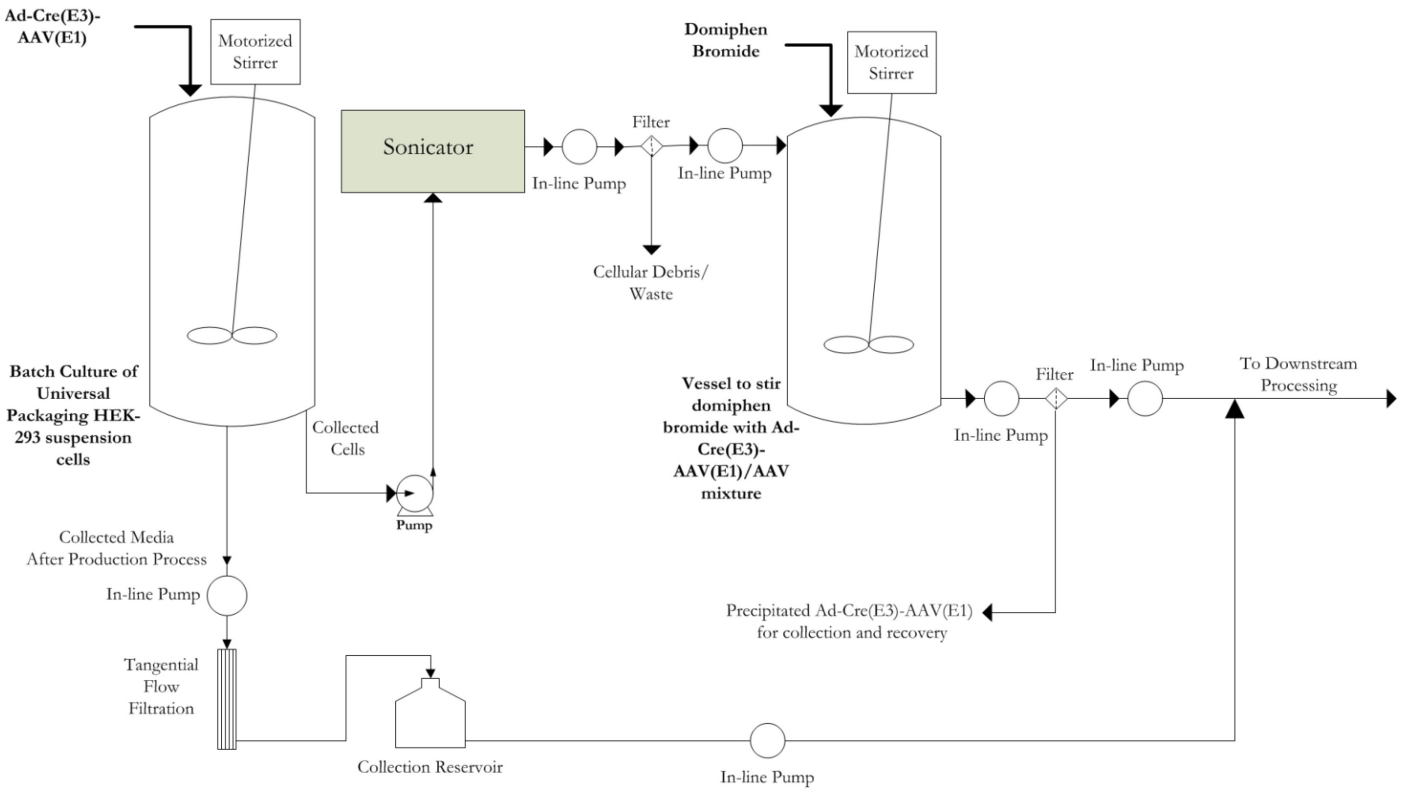


Figure 7: Hypothesized upstream process design for large scale rAAV using Ad-Cre(E3)-rAAV(E1) in universal packaging cell line. First the novel adenovirus construct is infected into a suspension inducible packaging cell line, then the cells and media are harvested. Next, the cells are disintegrated, releasing viral products and contaminants into a crude lysate. This crude lysate is initially clarified by the selective precipitating agent, hypothetically yielding a clarified rAAV product that is sent to downstream purification schemes

Aim 1. Generation and Characterization of Ad-Cre (E3)-rAAV (E1)

The unused E3 region of the E1-, E3- deleted adenovirus can be used to insert the Cre recombinase gene, and the E1-region for insertion of the rAAV vector with flanking inverted terminal repeats (ITRs). This would still be a single helper recombinant adenovirus infection, but instead of repeatedly performing cell line generation for every novel rAAV vector intended for a viral product, like in the producer cell line method developed in our lab, an easier universal inducible packaging cell line can be developed for a particular AAV serotype or capsid variants with the rAAV vector provided in trans by the adenovirus. Also provided in trans is the Cre recombinase gene, but instead of being placed at the E1 region, this Cre is placed in the unnecessary to in vitro use E3 region of the adenovirus. The method of placing the Cre gene is done by standard molecular cloning, since the Cre is unchanged between each Ad construct hypothesized in this aim. However, the method to place the vector is too cumbersome to perform using restriction enzyme digests and the recombination steps of the AdEasy method for each rAAV vector intended to be placed in the E1 region of the adenovirus. Our reconstruction of the Ad to place the rAAV vector specifically to the E1 region is intended to simplify the incorporation of the rAAV vector in E1 of Ad. We made this simpler by incorporating a Gateway Recombination cassette into the E1 region of the adenovirus to specifically recombine any new rAAV vector into the E1 region of the adenovirus and avoid the more time-consuming methods for adenovirus design. Hypothesis: Using Gateway recombination to place an rAAV vector into the E1 region and Cre recombinase into the E3 region, resulting in Ad-Cre(E3)-rAAV(E1), can produce high potent, high quality, and high titer rAAV, both single- and double-stranded ITRs of a variety of transgenes, when infected into a universal rAAV packaging HEK293 cell line.

Aim 2. CRISPR/Cas9 Modification to HEK293 Cell

Our previous methods involved the random integration of a linearized inducible packaging plasmid in the HEK293 genome, which caused a particularly time-consuming characterization process. This integration rate was rather low because the integration was dependent on the available randomly formed double-strand breaks in the genome. We will utilize the state of the art CRISPR/Cas9 genome (the clustered, regularly interspaced, short palindromic repeat (CRISPR)-associated protein 9 nuclease (Cas9)) editing technology to create an inducible Rep/Cap packaging cell line via targeted integration, hypothetically resulting in a cell line with sufficient copy numbers of Rep/Cap to generate rAAV product in tandem with Aim 1. Recently, this emergent commercially available technology is able to efficiently create site-specific double strand breaks and has been gaining popularity for its ease-of-use and adaptability. In this aim, we ask a simple question: can we utilize this powerful technology to create a superior AAV packaging cell line with better performance to our previous method in concert with findings from Aim 1? We intend to target the AAVS1 site on chromosome 19 (chr.19q13.3q-ter) because AAV has a natural affinity to this region and wild type forms of AAV can integrate site-specifically into this area during its latent infection. Here we hypothesize that if we target the inducible AAV packaging plasmid into the AAVS1 site of HEK293 cells, the yield of AAV using the methods in Aim 1 will be increased when compared to cell lines generated from no assistance of the targeted integration.

Aim 3. Purification of Vector Preparation

Adenovirus removal from our preparations is a difficult challenge. With the prospects of the aforementioned technology entering industrial scale, the design for the removal of the adenovirus from the rAAV product is paramount. We will make use of information from a 2013 Merck patent that has shown adenovirus precipitation using a selective precipitating agent (SPA), domiphen bromide, along with cellular DNA, even at high cell densities. We have replicated the data presented by Merck and extended concentration range of domiphen bromide to show that as concentration increases, both adenovirus and DNA are able to precipitate out of solution with about 99% clarity. However, we were interested to see if this detergent can be selective enough to disregard rAAV, allowing us to remove majority of the adenoviral contaminants and have rAAV remain in solution for downstream purification methods. Hypothesis: The addition of the selective precipitating agent, domiphen bromide, will allow for contaminating adenovirus to be cleared from rAAV vector preparations without the need for heat inactivation or chromatography and remove host cellular nucleic acid contamination without need for nucleases like Benzonase.

REFERENCES

1 Anderson, W. F. Human Gene Therapy. *Science (New York, N.Y.)* **256**, 808-813, doi:10.1126 (1992).

2 *IDF Advocate* (Immune Deficiency Foundation, 2013).

3 DeWitt Stetten, J. *A Revolution in Progress: Human Genetics and Medical Research*, <https://history.nih.gov/exhibits/genetics/index.htm> (2016).

4 Crystal, R. G. Transfer of Genes to Humans: Early Lessons and Obstacles to Success. *Science (New York, N.Y.)* **270**, 404-410 (1995).

5 Kohn, D. B., Sadelain, M. & Glorioso, J. C. Occurrence of Leukaemia Following Gene Therapy of X-Linked SCID. *Nature Reviews Cancer* **3**, 477-488, doi:10.1038/nrc1122 (2003).

6 Frank, K. M. *et al.* Investigation of the Cause of Death in a Gene-Therapy Trial. *New England Journal of Medicine* **361**, 161-169, doi:10.1056/NEJMoa0801066 (2009).

7 Raper, S. E. *et al.* Fatal Systemic Inflammatory Response Sydrome in a Ornithine Transcarbamylase Deficient Patient Following Adenoviral Gene Transfer. *Molecular Genetics and Metabolism* **80**, 148-158, doi:https://doi.org/10.1016/j.ymgme.2003.08.016 (2003).

8 Herper, M. Gene Therapy's Big Comeback. *Forbes* (2014). <http://www.forbes.com/sites/matthewherper/2014/03/26/once-seen-as-too-scary-editing-peoples-genes-with-viruses-makes-a-618-million-comeback>.

9 Booth, B. *Cell And Gene Therapy: An 'Outside-In' Technology Evolution*, <https://www.forbes.com/sites/brucebooth/2015/11/12/cell-and-gene-therapy-an-outside-in-technology-evolution/#1968aff42ff0> (2015).

10 Cideciyan, A. V. *et al.* Vision 1 Year after Gene Therapy for Leber's Congenital Amaurosis. *New England Journal of Medicine* **361**, 725-727, doi:10.1056/NEJMc0903652 (2009).

11 Lewis, R. *The Forever Fix: Gene Therapy and The Boy Who Saved It.* (St. Martin's Press, 2012).

12 National Human Genome Research Institute, (NHGRI) *Human Genome Project*, <https://report.nih.gov/NIHfactsheets/ViewFactSheet.aspx?csid=45> (2013).

13 *Gene Therapy Clinical Trials Worldwide*, <http://www.abedia.com/wiley/> (2017).

14 Nayerossadat, N., Maedeh, T. & Ali, P. A. Viral and Nonviral Delivery Systems for Gene Delivery. *Advanced Biomedical Research* **1**, doi:10.4103/2277-9175.98152. (2012).

15 Edelstein, M. L., Abedi, M. R., Wixon, J. & Edelstein, R. M. Gene therapy clinical trials worldwide 1989–2004 – an overview. *The Journal of Gene Medicine* **6**, 597-602 (2004).

16 Edelstein, M. L., Abedi, M. R. & Wixon, J. Gene therapy clinical trials worldwide to 2007 – an update. *The Journal of Gene Medicine* **9**, 833-842 (2007).

17 Atchison, R. W., Casto, B. C. & Hammon, W. M. Adenovirus-Associated Defective Virus Particles. *Science (New York, N.Y.)* **149**, 754-755, doi:10.1126/science.149.3685.754 (1965).

18 Grieger, J. C. & Samulski, R. J. Adeno-associated virus as a gene therapy vector: vector development, production and clinical applications. *Advances in biochemical engineering/biotechnology* **99**, 119-145 (2005).

19 Cotmore, S. F. & Tattersall, P. 28 Parvovirus DNA Replication. *Cold Spring Harbor Monograph Archive* **31**, 799-813 (1996).

20 Huser, D., Weger, S. & Heilbronn, R. Packaging of human chromosome 19-specific adeno-associated virus (AAV) integration sites in AAV virions during AAV wild-type and recombinant AAV vector production. *Journal of Virology* **77**, 4881-4887, doi:10.1128/JVI.77.8.4881-4887.2003 (2003).

21 Weitzman, M. D., Kyöstiö, S. R., Kotin, R. M. & Owens, R. A. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 5808-5812 (1994).

22 Zarate-Perez, F. *et al.* Determination of Adeno-associated Virus Rep DNA Binding Using Fluorescence Anisotropy. *Bio-protocol* **7**, e2194 (2017).

23 Smith, R. Adeno-associated virus integration: virus versus vector. *Gene therapy* **15**, 817-822, doi:doi:10.1038/gt.2008.55 (2008).

24 Nakai, H. *et al.* Large-Scale Molecular Characterization of Adeno-Associated Virus Vector Integration in Mouse Liver. *Journal of Virology* **79**, 3606-3614, doi:doi: 10.1128/JVI.79.6.3606-3614.2005 (2005).

25 Ran, F. A. *et al. In vivo* genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191, doi:10.1038/nature14299 (2015).

26 Boutin, S. *et al.* Prevalence of Serum IgG and Neutralizing Factors Against Adeno-Associated Virus (AAV) Types 1, 2, 5, 6, 8, and 9 in the Healthy Population: Implications for Gene Therapy Using AAV Vectors. *Human Gene Therapy* **21**, 704-712, doi:DOI: 10.1089/hum.2009.182 (2010).

27 Burger, C. *et al.* Recombinant AAV Viral Vectors Pseudotyped with Viral Capsids from Serotypes 1, 2, and 5 Display Differential Efficiency and Cell Tropism after Delivery to Different Regions of the Central Nervous System. *Molecular Therapy* **10**, 302-317, doi:10.1016/j.ymthe.2004.05.024 (2004).

28 Du, L. *et al.* Differential Myocardial Gene Delivery by Recombinant Serotype-Specific Adeno-associated Viral Vectors. *Molecular Therapy* **10**, 604-608, doi:10.1016/j.ymthe.2004.06.110 (2004).

29 Hauck, B. *et al.* Efficient AAV1-AAV2 Hybrid Vector for Gene Therapy of Hemophilia. *Human gene therapy* **17**, 46-54, doi:10.1089/hum.2006.17.46 (2006).

30 Loiler, S. A. *et al.* Targeting recombinant adeno-associated virus vectors to enhance gene transfer to pancreatic islets and liver. *Gene therapy* **10**, 1551, doi:10.1038/sj.gt.3302046 (2003).

31 Rabinowitz, J. E. *et al.* Cross-Packaging of a Single Adeno-Associated Virus (AAV) Type 2 Vector Genome into Multiple AAV Serotypes Enables Transduction with Broad Specificity. *Journal of Virology* **76**, 791-801, doi:doi: 10.1128/JVI.76.2.791-801.2002 (2002).

32 Seppen, J. *et al.* Adeno-associated Virus Vector Serotypes Mediate Sustained Correction of Bilirubin UDP Glucuronosyltransferase Deficiency in Rats. *Molecular Therapy* **13**, 1085-1092, doi:10.1016/j.ymthe.2006.01.014 (2006).

33 Stone, I. M., Lurie, D. I., Kelley, M. W. & Poulsen, D. J. Adeno-associated virus-mediated gene transfer to hair cells and support cells of the murine cochlea. *Molecular Therapy* **11**, 843-848, doi:10.1016/j.ymthe.2005.02.005 (2005).

34 Su, H. *et al.* AAV Serotype-1 mediates early onset of gene expression in mouse hearts and results in better therapeutic effect. *Gene therapy* **13**, 1495, doi:10.1038/sj.gt.3302787 (2006).

35 Takeda, S. i. *et al.* Successful Gene Transfer Using Adeno-Associated Virus Vectors into the Kidney: Comparison among Adeno-Associated Virus Serotype 1–5 Vectors *in vitro* and *in vivo*. *Nephron Experimental Nephrology* **96**, e119-e126 (2004).

36 Wang, C., Wang, C. M., Clark, K. R. & Sferra, T. J. Recombinant AAV serotype 1 transduction efficiency and tropism in the murine brain. *Gene therapy* **10**, 1528, doi:10.1038/sj.gt.3302011 (2003).

37 Wu, Z., Miller, E., Agbandje-McKenna, M. & Samulski, R. J. α2,3 and α2,6 N-Linked Sialic Acids Facilitate Efficient Binding and Transduction by Adeno-Associated Virus Types 1 and 6. *Journal of Virology* **80**, 9093-9103, doi:doi: 10.1128/JVI.00895-06 (2006).

38 Xiao, W. *et al.* Gene Therapy Vectors Based on Adeno-Associated Virus Type 1. *Journal of Virology* **73**, 3994-4003 (1999).

39 Zhong, L. *et al.* Evaluation of Primitive Murine Hematopoietic Stem and Progenitor Cell Transduction In Vitro and In Vivo by Recombinant Adeno-Associated Virus Vector Serotypes 1 Through 5. *Human Gene Therapy* **17**, 321-333, doi:10.1089/hum.2006.17.321 (2006).

40 Akache, B. *et al.* The 37/67-Kilodalton Laminin Receptor Is a Receptor for Adeno-Associated Virus Serotypes 8, 2, 3, and 9. *Journal of Virology* **80**, 9831-9836, doi:10.1128/JVI.00878-06 (2006).

41 Asokan, A., Hamra, J. B., Govindasamy, L., Agbandje-McKenna, M. & Samulski, R. J. Adeno-Associated Virus Type 2 Contains an Integrin α5β1 Binding Domain Essential for Viral Cell Entry. *Journal of Virology* **80**, 8961-8969, doi:10.1128/JVI.00843-06 (2006).

42 Glushakova, L. G., Timmers, A. M., Pang, J., Teusner, J. T. & Hauswirth, W. W. Human Blue-Opsin Promoter Preferentially Targets Reporter Gene Expression to Rat S-Cone Photoreceptors. *Investigative Ophthalmology & Visual Science* **47**, 3505-3513, doi:10.1167/iovs.05-1670 (2006).

43 Hacker, U. T. *et al.* Adeno-associated virus serotypes 1 to 5 mediated tumor cell directed gene transfer and improvement of transduction efficiency. *The Journal of Gene Medicine* **7**, 1429-1438, doi:10.1002/jgm.782 (2005).

44 Kashiwakura, Y. *et al.* Hepatocyte Growth Factor Receptor Is a Coreceptor for Adeno-Associated Virus Type 2 Infection. *Journal of Virology* **79**, 609-614, doi:10.1128/JVI.79.1.609-614.2005 (2005).

45 Qing, K. *et al.* Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nature Medicine* **5**, 71, doi:10.1038/4758 (1999).

46 Summerford, C. & Samulski, R. J. Membrane-Associated Heparan Sulfate Proteoglycan Is a Receptor for Adeno-Associated Virus Type 2 Virions. *Journal of Virology* **72**, 1438-1445 (1998).

47 Tenenbaum L., Chtarto A., Lehtonen E., Velu T., Brotchi J., & Leivier M. Recombinant AAV-mediated gene delivery to the central nervous system. *The Journal of Gene Medicine* **6**, S212-222, doi:10.1002/jgm.506 (2004).

48 Thorsen, F. *et al.* Adeno-associated virus (AAV) serotypes 2, 4 and 5 display similar transduction profiles and penetrate solid tumor tissue in models of human glioma. *The Journal of Gene Medicine* **8**, 1131-1140, doi:10.1002/jgm.939 (2006).

49 Handa A. Muramatsu S., Qiu J., Mizukami H., & Brown K.E. Adeno-associated virus (AAV)-3-based vectors transduce haematopoietic cells not susceptible to transduction with AAV-2-based vectors. *Journal of General Virology* **81**, 2077-2084, doi:10.1099/0022-1317-81-8-2077 (2000).

50 Lerch, T. F. & Chapman, M. S. Identification of the heparin binding site on adeno-associated virus serotype 3B (AAV-3B). *Virology* **423**, 6-13, doi:https://doi.org/10.1016/j.virol.2011.10.007 (2012).

51 Ling, C. *et al.* Human Hepatocyte Growth Factor Receptor Is a Cellular Coreceptor for Adeno-Associated Virus Serotype 3. *Human Gene Therapy* **21**, 1741-1747, doi:10.1089/hum.2010.075 (2010).

52 Kaludov, N., Brown, K. E., Walters, R. W., Zabner, J. & Chiorini, J. A. Adeno-Associated Virus Serotype 4 (AAV4) and AAV5 Both Require Sialic Acid Binding for Hemagglutination and Efficient Transduction but Differ in Sialic Acid Linkage Specificity. *Journal of Virology* **75**, 6884-6893, doi:10.1128/JVI.75.15.6884-6893.2001 (2001).

53 Davidson, B. L. *et al.* Recombinant adeno-associated virus type 2, 4, and 5 vectors: Transduction of variant cell types and regions in the mammalian central nervous system. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 3428-3432 (2000).

54 Pasquale, G. D. *et al.* Identification of PDGFR as a receptor for AAV-5 transduction. *Nature Medicine* **9**, 1306, doi:10.1038/nm929 https://www.nature.com/articles/nm929#supplementary-information (2003).

55 Duan, D., Yan, Z., Yue, Y., Ding, W. & Engelhardt, J. F. Enhancement of Muscle Gene Delivery with Pseudotyped Adeno-Associated Virus Type 5 Correlates with Myoblast Differentiation. *Journal of Virology* **75**, 7662-7671, doi:10.1128/JVI.75.16.7662-7671.2001 (2001).

56 Mingozzi, F. *et al.* Improved Hepatic Gene Transfer by Using an Adeno-Associated Virus Serotype 5 Vector. *Journal of Virology* **76**, 10497-10502, doi:10.1128/JVI.76.20.10497-10502.2002 (2002).

57 Seiler, M. P., Miller, A. D., Zabner, J. & Halbert, C. L. Adeno-Associated Virus Types 5 and 6 Use Distinct Receptors for Cell Entry. *Human Gene Therapy* **17**, 10-19, doi:10.1089/hum.2006.17.10 (2006).

58 Walters, R. W. *et al.* Structure of Adeno-Associated Virus Serotype 5. *Journal of Virology* **78**, 3361-3371, doi:10.1128/JVI.78.7.3361-3371.2004 (2004).

59 Walters, R. W. *et al.* Binding of Adeno-associated Virus Type 5 to 2,3-Linked Sialic Acid Is Required for Gene Transfer. *Journal of Biological Chemistry* **276**, 20610-20616, doi:10.1074/jbc.M101559200 (2001).

60 Aldrich, W. A. *et al.* Enhanced transduction of mouse bone marrow-derived dendritic cells by repetitive infection with self-complementary adeno-associated virus 6 combined with immunostimulatory ligands. *Gene therapy* **13**, 29, doi:10.1038/sj.gt.3302601 (2005).

61 Blankinship, M. J. *et al.* Efficient transduction of skeletal muscle using vectors based on adeno-associated virus serotype 6. *Molecular Therapy* **10**, 671-678, doi:10.1016/j.ymthe.2004.07.016 (2004).

62 Ghosh, A., Yue, Y. & Duan, D. Viral serotype and the transgene sequence influence overlapping adeno-associated viral (AAV) vector-mediated gene transfer in skeletal muscle. *The Journal of Gene Medicine* **8**, 298-305, doi:10.1002/jgm.835 (2006).

63 Kawamoto, S., Shi, Q., Nitta, Y., Miyazaki, J.-i. & Allen, M. D. Widespread and early myocardial gene expression by adeno-associated virus vector type 6 with a β-

actin hybrid promoter. *Molecular Therapy* **11**, 980-985, doi:10.1016/j.ymthe.2005.02.009 (2005).

64 Limberis, M. P., Vandenberghe, L. H., Zhang, L., Pickles, R. J. & Wilson, J. M. Transduction Efficiencies of Novel AAV Vectors in Mouse Airway Epithelium In Vivo and Human Ciliated Airway Epithelium In Vitro. *Molecular Therapy: the Journal of the American Society of Gene Therapy* **17**, 294-301, doi:10.1038/mt.2008.261 (2009).

65 Ng, R. *et al.* Structural Characterization of the Dual Glycan Binding Adeno-Associated Virus Serotype 6. *Journal of Virology* **84**, 12945-12957, doi:10.1128/JVI.01235-10 (2010).

66 Weller, M. L. *et al.* Epidermal growth factor receptor is a co-receptor for adeno-associated virus serotype 6. *Nature medicine* **16**, 662-664, doi:10.1038/nm.2145 (2010).

67 Harding, T. C. *et al.* Enhanced Gene Transfer Efficiency in the Murine Striatum and an Orthotopic Glioblastoma Tumor Model, Using AAV-7- and AAV-8-Pseudotyped Vectors. *Human Gene Therapy* **17**, 807-820, doi:10.1089/hum.2006.17.807 (2006).

68 Denby, L., Nicklin, S. A. & Baker, A. H. Adeno-associated virus (AAV)-7 and -8 poorly transduce vascular endothelial cells and are sensitive to proteasomal degradation. *Gene therapy* **12**, 1534, doi:10.1038/sj.gt.3302564 (2005).

69 Gao, G. *et al.* Clades of Adeno-Associated Viruses Are Widely Disseminated in Human Tissues. *Journal of Virology* **78**, 6381-6388, doi:10.1128/JVI.78.12.6381-6388.2004 (2004).

70 Loiler, S. A. *et al.* Localized Gene Expression Following Administration of Adeno-associated Viral Vectors via Pancreatic Ducts. *Molecular Therapy* **12**, 519-527, doi:10.1016/j.ymthe.2005.04.017 (2005).

71 Nakai, H. *et al.* Unrestricted Hepatocyte Transduction with Adeno-Associated Virus Serotype 8 Vectors in Mice. *Journal of Virology* **79**, 214-224, doi:10.1128/JVI.79.1.214-224.2005 (2005).

72 Nam, H.-J. *et al.* Structure of Adeno-Associated Virus Serotype 8, a Gene Therapy Vector. *Journal of Virology* **81**, 12260-12271, doi:10.1128/jvi.01304-07 (2007).

73 Pan, X. *et al.* Long-Term Robust Myocardial Transduction of the Dog Heart from a Peripheral Vein by Adeno-Associated Virus Serotype-8. *Human Gene Therapy* **24**, 584-594, doi:10.1089/hum.2013.044 (2013).

74 Wang, A. Y., Peng, P. D., Ehrhardt, A., Storm, T. A. & Kay, M. A. Comparison of Adenoviral and Adeno-Associated Viral Vectors for Pancreatic Gene Delivery In Vivo. *Human Gene Therapy* **15**, 405-413, doi:10.1089/104303404322959551 (2004).

75 Wang, Z. *et al.* Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. *Nature Biotechnology* **23**, 321, doi:10.1038/nbt1073 https://www.nature.com/articles/nbt1073#supplementary-information (2005).

76 Bell, C. L. *et al.* The AAV9 receptor and its modification to improve *in vivo* lung gene transfer in mice. *The Journal of Clinical Investigation* **121**, 2427-2435, doi:10.1172/JCI57367 (2011).

77 Limberis M.P. & Wilson J.M. Adeno-associated virus serotype 9 vectors transduce murinealveolar and nasal epithelia and can be readministered. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12993-12998, doi:10.1073/pnas.0601433103 (2006).

78 Pacak C.A. *et al.* Recombinant adeno-associated virus serotype 9 leads to preferential cardiac transduction *in vivo*. *Circulation research* **99**, e3-9, doi:10.1161/01.RES.0000237661.18885.f6 (2006).

79 Shen, S., Bryant, K. D., Brown, S. M., Randell, S. H. & Asokan, A. Terminal N-Linked Galactose Is the Primary Receptor for Adeno-associated Virus 9. *The Journal of Biological Chemistry* **286**, 13532-13540, doi:10.1074/jbc.M110.210922 (2011).

80 Vandendriessche, T. *et al.* Efficacy and safety of adeno-associated viral vectors based on serotype 8 and 9 vs. lentiviral vectors for hemophilia B gene therapy. *Journal of Thrombosis and Haemostasis* **5**, 16-24, doi:10.1111/j.1538-7836.2006.02220.x (2007).

81 (https://ClinicalTrials.gov/show/NCT00891306).

82 (https://ClinicalTrials.gov/show/NCT02935517).

83 (https://ClinicalTrials.gov/show/NCT02416622).

84 (https://ClinicalTrials.gov/show/NCT02599922).

85 (https://ClinicalTrials.gov/show/NCT01054339).

86 (https://ClinicalTrials.gov/show/NCT00749957).

87 MacLachlan, T. K. *et al.* Preclinical Safety Evaluation of AAV2-sFLT01— A Gene Therapy for Age-related Macular Degeneration. *Molecular Therapy* **19**, 326-334, doi:10.1038/mt.2010.258 (2011).

88 (https://ClinicalTrials.gov/show/NCT02727764).

89 (https://ClinicalTrials.gov/show/NCT03223194).

90 (https://ClinicalTrials.gov/show/NCT03199469).

91 (https://ClinicalTrials.gov/show/NCT03306277).

92 (https://ClinicalTrials.gov/show/NCT01687608).

93 (https://ClinicalTrials.gov/show/NCT01643330).

94 (https://ClinicalTrials.gov/show/NCT00400634).

95 (https://ClinicalTrials.gov/show/NCT00087789).

96 Mandel R.J. CERE-110, an adeno-associated virus-based gene delivery vector expressing human nerve growth factor for the treatment of Alzheimer's disease. *Current opinion in molecular therapeutics* **12**, 240-247 (2010).

97 (https://ClinicalTrials.gov/show/NCT02082860).

98 (https://ClinicalTrials.gov/show/NCT02618915).

99 (https://ClinicalTrials.gov/show/NCT02991144).

100 (https://ClinicalTrials.gov/show/NCT01344798).

101 (https://ClinicalTrials.gov/show/NCT03326336).

102 (https://ClinicalTrials.gov/show/NCT02652767).

103 (https://ClinicalTrials.gov/show/NCT01024998).

104 (https://ClinicalTrials.gov/show/NCT00229736).

105 (https://ClinicalTrials.gov/show/NCT03328130).

106 (https://ClinicalTrials.gov/show/NCT02053064).

107 (https://ClinicalTrials.gov/show/NCT02781480).

108 (https://ClinicalTrials.gov/show/NCT03001310).

109 (https://ClinicalTrials.gov/show/NCT02446249).

110 (https://ClinicalTrials.gov/show/NCT03252847).

111 (https://ClinicalTrials.gov/show/NCT00195143).

112 Kaplitt, M. G. *et al.* Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial. *Lancet (London, England)* **369**, 2097-2105, doi:10.1016/s0140-6736(07)60982-9 (2007).

113 Fischer, M. D. *et al.* Codon-Optimized RPGR Improves Stability and Efficacy of AAV8 Gene Therapy in Two Mouse Models of X-Linked Retinitis Pigmentosa. *Molecular Therapy* **25**, 1854-1865, doi:10.1016/j.ymthe.2017.05.005 (2017).

114 (https://ClinicalTrials.gov/show/NCT03116113).

115 (https://ClinicalTrials.gov/show/NCT03066258).

116 (https://ClinicalTrials.gov/show/NCT02695160).

117 (https://ClinicalTrials.gov/show/NCT02702115).

118 (https://ClinicalTrials.gov/show/NCT03041324).

119 (https://ClinicalTrials.gov/show/NCT03061201).

120 (https://ClinicalTrials.gov/show/NCT03003533).

121 (https://ClinicalTrials.gov/show/NCT00999609).

122 (https://ClinicalTrials.gov/show/NCT02341807).

123 (https://ClinicalTrials.gov/show/NCT01620801).

124 (https://ClinicalTrials.gov/show/NCT02484092).

125 (https://ClinicalTrials.gov/show/NCT01899092).

126 (https://ClinicalTrials.gov/show/NCT00126724).

127 (https://ClinicalTrials.gov/show/NCT02396342).

128 (https://ClinicalTrials.gov/show/NCT03300453).

129 (https://ClinicalTrials.gov/show/NCT03065192).

130 Thomas, C. E., Ehrhardt, A. & Kay, M. A. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* **4**, 346-358 (2003).

131 Wirth, T., Parker, N. & Ylä-Herttuala, S. History of gene therapy. *Gene* **525**, 162-169, doi:http://dx.doi.org/10.1016/j.gene.2013.03.137 (2013).

132 The needs of the few. *Nature* **466**, 160, doi:10.1038/466160a (2010).

133 Rodgers, P. *Gene Therapy To Replace Heart Transplants*, <https://www.forbes.com/sites/paulrodgers/2014/08/12/injection-to-replace-heart-pumps-and-transplants/#7016bfbd23f9> (2014).

134 Garde, D. *Celladon pulls the plug on R&D after a crushing gene therapy failure*, <https://www.fiercebiotech.com/financials/celladon-pulls-plug-on-r-d-after-a-crushing-gene-therapy-failure> (2015).

135 Garde, D. *Eiger BioPharma eyes the shell of Celladon to back its way onto Wall Street*, <https://www.fiercebiotech.com/financials/eiger-biopharma-eyes-shell-of-celladon-to-back-its-way-onto-wall-street> (2015).

136 Dolgin, E. Big pharma moves from 'blockbusters' to 'niche busters'. *Nature Medicine* **16**, 837, doi:10.1038/nm0810-837a (2010).

137 Timmerman, L. *Following The Biobucks: Cancer Immunotherapy, Gene Therapy And Synthetic Biology Startups Rake In Cash*, <https://www.forbes.com/sites/luketimmerman/2016/07/20/following-the-biobucks-cancer-immunotherapy-gene-therapy-synthetic-biology-startups-rake-in-cash/#7e500d1c76a0> (2016).

138 Gronde, T. v. d., Uyl-de Groot, C. A. & Pieters, T. Addressing the challenge of high-priced prescription drugs in the era of precision medicine: A systematic review of drug life cycles, therapeutic drug markets and regulatory frameworks. *PLOS ONE* **12**, e0182613, doi:10.1371/journal.pone.0182613 (2017).

139 U. S. Food and Drug Administration. *Orphan Products: Hope for People With Rare Diseases*, <https://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm143563.htm> (2016).

140 Burton, T. M. *FDA Modernizing Evaluations as Gene, Cell Therapy Field Grows*, <https://www.wsj.com/articles/fda-modernizing-evaluations-as-gene-cell-therapy-fields-grow-1504802513s> (2017).

141 Weintraub, A. *Big Pharma Doesn't Want Health Outcomes To Sway Pricing -- At Least Not Yet*, <https://www.forbes.com/sites/arleneweintraub/2017/09/29/big-pharma-doesnt-want-health-outcomes-to-sway-pricing-at-least-not-yet/#15a68b0e7033> (2017).

142 Weintraub, A. *Here Come More Gene Therapies -- And More Pricing Debates*, <https://www.forbes.com/sites/arleneweintraub/2017/10/11/here-come-more-gene-therapies-and-more-pricing-debates/#3ae79f1e577e> (2017).

143 Taylor, P. *Spark Therapeutics' Luxturna advisory committee vote sets gene therapy landmark*, <https://www.fiercebiotech.com/biotech/spark-therapeutics-luxturna-adcomm-vote-sets-gene-therapy-landmark> (2017).

144 Salmon, F., Grosios, K. & Petry, H. Safety profile of recombinant adeno-associated viral vectors: focus on alipogene tiparvovec (Glybera®). *Expert Review of Clinical Pharmacology* **7**, 53-65 (2014).

145 Regalado, A. *The World's Most Expensive Medicine Is a Bust*, <https://www.technologyreview.com/s/601165/the-worlds-most-expensive-medicine-is-a-bust/> (2016).

146 Mullin, E. *The World's Most Expensive Medicine Is Being Pulled from the Market*, <https://www.technologyreview.com/s/604252/the-worlds-most-expensive-medicine-is-being-pulled-from-the-market/> (2017).

147 Spark Therapeutics, I. FDA ADVISORY COMMITTEE BRIEFING DOCUMENT: Spark Therapeutics, Inc LUXTURNA TM (voretigene neparvovec). (FDA.gov, 2017).

148 UniQure. *GENE THERAPY: Intellectual Property* <http://uniqure.com/gene-therapy/gene-therapy-intellectual-property.php>

149 UniQure. *Gene Therapy: Manufacturing*, <http://www.uniqure.com/gene-therapy/manufacturing.php>

150 Dietmair, S., Nielsen, L. K., Timmins, N. E. & J. Engineering a mammalian super producer. *Technol Biotechnol* **86 SRC - GoogleScholar**, 905-914 (2011).

151 Park, J. Y., Lim, B.-P., Lee, K., Kim, Y.-G. & Jo, E.-C. Scalable production of adeno-associated virus type 2 vectors via suspension transfection. *Biotechnology and bioengineering* **94**, 416-430 (2006).

152 Dong, B. *et al.* A concept of eliminating nonhomologous recombination for scalable and safe AAV vector generation for human gene therapy. *Nucleic Acids Res* **41**, 6609-6617, doi:10.1093/nar/gkt404 (2013).

153 Johnston, K. M. *et al.* HSV/AAV hybrid amplicon vectors extend transgene expression in human glioma cells. *Human gene therapy* **8**, 359-370 (1997).

154 Fisher, K. J., Kelley, W. M., Burda, J. F. & Wilson, J. M. A novel adenovirus-adeno-associated virus hybrid vector that displays efficient rescue and delivery of the AAV genome. *Human gene therapy* **7**, 2079-2087 (1996).

155 Mietzsch, M. *et al.* OneBac: Platform for Scalable and High-Titer Production of Adeno-Associated Virus Serotype 1–12 Vectors for Gene Therapy. *Human Gene Therapy* **25**, 212-222 (2014).

156 Yuan, Z., Qiao, C., Hu, P., Li, J. & Xiao, X. A versatile adeno-associated virus vector producer cell line method for scalable vector production of different serotypes. *Human gene therapy* **22**, 613-624 (2011).

157 McCarty, D. M., Christensen, M. & Muzyczka, N. Sequences required for coordinate induction of adeno-associated virus p19 and p40 promoters by Rep protein. *Journal of Virology* **65**, 2936-2945 (1991).

158 McCarty, D. M. *et al.* Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. *Journal of Virology* **68**, 4988-4997 (1994).

159 Balagúe, C., Kalla, M. & Zhang, W. W. Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. *Journal of Virology* **71**, 3299-3306 (1997).

160 King, J. A., Dubielzig, R., Grimm, D. & Kleinschmidt, J. A. DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. *The EMBO Journal* **20**, 3282-3291, doi:10.1093/emboj/20.12.3282 (2001).

161 Xie, Q. *et al.* The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proceedings of the National Academy of Sciences* **99**, 10405-10410, doi:10.1073/pnas.162250899 (2002).

162 Agbandje-McKenna, M. & Kleinschmidt, J. in *Adeno-Associated Virus* Vol. 807 *Methods in Molecular Biology* (eds Richard O. Snyder & Philippe Moullier) Ch. 3, 47-92 (Humana Press, 2011).

163 Van Vliet, K. K. Adeno-associated virus capsid serotype identification: Analytical methods development and application. *Journal of virological methods* **159**, 167-177, doi:10.1016/j.jviromet.2009.03.020. (2009)

164 Wu, Z., Asokan, A. & Samulski, R. J. Adeno-associated Virus Serotypes: Vector Toolkit for Human Gene Therapy. *Mol Ther* **14**, 316-327 (2006).

165 Gurda, B. L. *et al.* Capsid Antibodies to Different Adeno-Associated Virus Serotypes Bind Common Regions. *Journal of Virology* **87**, 9111-9124, doi:10.1128/jvi.00622-13 (2013).

166 Mouw, M. B. & Pintel, D. J. Adeno-Associated Virus RNAs Appear in a Temporal Order and Their Splicing Is Stimulated during Coinfection with Adenovirus. *Journal of Virology* **74**, 9878-9888 (2000).

167 Moore, A. R., Dong, B., Chen, L. & Xiao, W. Vaccinia virus as a subhelper for AAV replication and packaging. *Molecular Therapy. Methods & Clinical Development* **2**, 15044, doi:10.1038/mtm.2015.44 (2015).

168 Xiao, X., Li, J. & Samulski, R. J. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J Virol* **72**, 2224-2232 (1998).

169 Weitzman, M. D. & Linden, R. M. in *Adeno-Associated Virus: Methods and Protocols* Vol. 807 *Methods in Molecular Biology* (eds Richard O Snyder & Philippe Moullier) Ch. 1, 1-23 (Humana Press, 2011).

170 Berns, K. I. Parvovirus replication. *Microbiological Reviews* **54**, 316-329 (1990).

171 Gonçalves, M. A. F. V. Adeno-associated virus: from defective virus to effective vector. *Virology Journal* **2**, 43, doi:10.1186/1743-422X-2-43 (2005).

172 Wang, Z. *et al.* Rapid and highly efficient transduction by double-stranded adeno-associated virus vectors *in vitro* and *in vivo*. *Gene therapy* **10**, 2105-2111 (2003).

173 McCarty, D. M. Self-complementary AAV Vectors; Advances and Applications. *Molecular Therapy* **16**, 1648-1656, doi:10.1038/mt.2008.171 (2008).

174 McCarty, D. M., Monahan, P. E. & Samulski, R. J. Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene therapy* **8**, 1248, doi:10.1038/sj.gt.3301514 (2001).

175 Ghosh, A. & Duan, D. Expanding Adeno-associated Viral Vector Capacity: A Tale of Two Vectors. *Biotechnology and Genetic Engineering Reviews* **24**, 165-178, doi:10.1080/02648725.2007.10648098 (2007).

176 Chandler, M. *et al.* Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature Reviews Microbiology* **11**, 525, doi:10.1038/nrmicro3067

https://www.nature.com/articles/nrmicro3067#supplementary-information (2013).

177 Kyöstiö, S. R. *et al.* Analysis of adeno-associated virus (AAV) wild-type and mutant Rep proteins for their abilities to negatively regulate AAV p5 and p19 mRNA levels. *Journal of Virology* **68**, 2947-2957 (1994).

178 Pereira, D. J., McCarty, D. M. & Muzyczka, N. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *Journal of Virology* **71**, 1079-1088 (1997).

179 Hermonat, P. L. & Muzyczka, N. Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. *Proceedings of the National Academy of Sciences* **81**, 6466-6470 (1984).

180 Tratschin, J. D., West, M. H., Sandbank, T. & Carter, B. J. A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase. *Molecular and Cellular Biology* **4**, 2072-2081, doi:10.1128/mcb.4.10.2072 (1984).

181 Grimm, D. & Kay, M. A. From Virus Evolution to Vector Revolution: Use of Naturally Occurring Serotypes of Adeno-associated Virus (AAV) as Novel Vectors for Human Gene Therapy. *Current Gene Therapy* **3**, 281-304, doi:http://dx.doi.org/10.2174/1566523034578285 (2003).

182 Penaud-Budloo, M. *et al.* Adeno-Associated Virus Vector Genomes Persist as Episomal Chromatin in Primate Muscle. *Journal of Virology* **82**, 7875-7885, doi:10.1128/JVI.00649-08 (2008).

183 Halbert, C. L., Allen, J. M. & Miller, A. D. Efficient mouse airway transduction following recombination between AAV vectors carrying parts of a larger gene. *Nat Biotech* **20**, 697-701 (2002).

184 Nakai, H. H. Increasing the size of rAAV-mediated expression cassettes *in vivo* by intermolecular joining of two complementary vectors. *Nature biotechnology* **18**, 527-532, doi:10.1038/75390. (2000).

185 Samulski, R. J. R. Expanding the AAV package. *Nature biotechnology* **18**, 497-498, doi:10.1038/75352. (2000).

186 Sun, L., Li, J. & Xiao, X. Overcoming adeno-associated virus vector size limitation through viral DNA heterodimerization. *Nat Med* **6**, 599-602 (2000).

187 Wang, B., Li, J. & Xiao, X. Adeno-associated virus vector carrying human minidystrophin genes effectively ameliorates muscular dystrophy in mdx mouse model. *Proceedings of the National Academy of Sciences* **97**, 13714-13719, doi:10.1073/pnas.240335297 (2000).

188 Yan, Z., Zhang, Y., Duan, D. & Engelhardt, J. F. Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6716-6721 (2000).

189 Gray, S. J. *et al.* Optimizing Promoters for Recombinant Adeno-Associated Virus-Mediated Gene Expression in the Peripheral and Central Nervous System Using Self-Complementary Vectors. *Human Gene Therapy* **22**, 1143-1153, doi:10.1089/hum.2010.245 (2011).

190 Paterna, J. C., Moccetti, T., Mura, A., Feldon, J. & Bueler, H. Influence of promoter and WHV post-transcriptional regulatory element on AAV-mediated transgene expression in the rat brain. *Gene therapy* **7**, 1304-1311, doi:10.1038/sj.gt.3301221 (2000).

191 Salva, M. Z. *et al.* Design of Tissue-specific Regulatory Cassettes for High-level rAAV-mediated Expression in Skeletal and Cardiac Muscle. *Mol Ther* **15**, 320-329 (2007).

192 Harper, S. Q. *et al.* RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5820-5825, doi:10.1073/pnas.0501507102 (2005).

193 Miller, T. M. *et al.* Virus-delivered small RNA silencing sustains strength in amyotrophic lateral sclerosis. *Annals of Neurology* **57**, 773-776, doi:10.1002/ana.20453 (2005).

194 Xia, H. H. RNAi suppresses polyglutamine-induced neurodegeneration in a model of spinocerebellar ataxia. *Nature medicine* **10**, 816-820, doi:10.1038/nm1076. (2004).

195 Grimm, D. *et al.* Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**, 537-541, doi:10.1038/nature04791 (2006).

196 McBride, J. L. *et al.* Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: Implications for the therapeutic development of RNAi. *Proceedings of the National Academy of Sciences* **105**, 5868-5873, doi:10.1073/pnas.0801775105 (2008).

197 Boudreau, R. L. R. L. Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs *in vitro* and *in vivo*. *Molecular therapy* **17**, 169-175, doi:10.1038/mt.2008.231. (2009).

198 Li, J., Samulski, R. J. & Xiao, X. Role for highly regulated rep gene expression in adeno-associated virus vector production. *Journal of Virology* **71**, 5236-5243 (1997).

199 Yang, Q., Chen, F. & Trempe, J. P. Characterization of cell lines that inducibly express the adeno-associated virus Rep proteins. *Journal of Virology* **68**, 4847-4856 (1994).

200 Berthet, C., Raj, K., Saudan, P. & Beard, P. How adeno-associated virus Rep78 protein arrests cells completely in S phase. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13634-13639, doi:10.1073/pnas.0504583102 (2005).

201 Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Mol Ther* **16**, 1073-1080, doi:10.1038/mt.2008.76 (2008).

202 Bish, L. T. *et al.* Adeno-associated virus (AAV) serotype 9 provides global cardiac gene transfer superior to AAV1, AAV6, AAV7, and AAV8 in the mouse and rat. *Human gene therapy* **19**, 1359-1368 (2008).

203 Kornegay, J. N. *et al.* Widespread muscle expression of an AAV9 human mini-dystrophin vector after intravenous injection in neonatal dystrophin-deficient dogs. *Molecular Therapy* **18**, 1501-1508 (2010).

204 Lu, Y. *et al.* Therapeutic level of functional human alpha 1 antitrypsin (hAAT) secreted from murine muscle transduced by adeno-associated virus (rAAV1) vector. *The journal of gene medicine* **8**, 730-735 (2006).

205 Asokan, A., Schaffer, D. V. & Jude Samulski, R. The AAV Vector Toolkit: Poised at the Clinical Crossroads. *Mol Ther* **20**, 699-708 (2012).

206 Bartel, M., Schaffer, D. & Büning, H. Enhancing the Clinical Potential of AAV Vectors by Capsid Engineering to Evade Pre-Existing Immunity. *Frontiers in Microbiology* **2**, 204, doi:10.3389/fmicb.2011.00204 (2011).

207 Kotterman, M. A. & Schaffer, D. V. Engineering adeno-associated viruses for clinical gene therapy. *Nature Reviews Genetics* **15**, 445-451 (2014).

208 Li, C. *et al.* Single Amino Acid Modification of Adeno-Associated Virus Capsid Changes Transduction and Humoral Immune Profiles. *Journal of Virology* **86**, 7752-7759, doi:10.1128/jvi.00675-12 (2012).

209 Choi, V. W., McCarty, D. M. & Samulski, R. J. AAV Hybrid Serotypes: Improved Vectors for Gene Delivery. *Current gene therapy* **5**, 299-310 (2005).

210 Electronic Code of Federal Regulations (Electronic Code of Federal Regulations, ecfr.gov, 2017).

211 Allay, J. A. *et al.* Good manufacturing practice production of self-complementary serotype 8 adeno-associated viral vector for a hemophilia B clinical trial. *Human gene therapy* **22**, 595-604 (2011).

212 Food and Drug Administration, Center for Biologics Evaluation and Research, in Guidance for and Sponsors: Content and Review of Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs). (2008).

213 U.S. Department of Health and Human Services, Food and Drug Administration & Center for Biologics Evaluation and Research *Guidance for Industry: Guidance for Human Somatic Cell Therapy and Gene Therapy* (ed Center for Biologics Evaluation and Research) (Food and Drug Administration, FDA.gov, 1998).

214 Wright, J. F. Transient Transfection Methods for Clinical Adeno-Associated Viral Vector Production. *Human gene therapy* **20**, 698-706, doi:10.1089/hum.2009.064 (2009).

215 Aucoin, M. G., Perrier, M. & Kamen, A. A. Critical assessment of current adeno-associated viral vector production and quantification methods. *Biotechnology advances* **26**, 73-88 (2008).

216 Mezzina, M. & Merten, O.-W. Adeno-associated viruses. *Methods in molecular biology (Clifton, N.J.)* **737**, 211-234 (2011).

217 Clément, N. & Grieger, J. C. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Molecular Therapy. Methods & Clinical Development* **3**, 16002, doi:10.1038/mtm.2016.2 (2016).

218 Corning Life Sciences. (Corning Life Sciences, corning.com, 2007).

219 Danos, O. in *Parent Project Muscular Dystrophy AAV Manufacturing Workshop.* (International Collaborative Effort, 2006).

220 National Hemophilia Foundation. *Fast Facts I National Hemophilia Foundation*, <https://www.hemophilia.org/About-Us/Fast-Facts> (2017).

221 NORD - National Organization for Rare Disorders. *Duchenne Muscular Dystrophy -NORD (National Organization for Rare Disorders)*, <http://rarediseases.org/rarediseases/duchenne-muscular-dystrophy/> (2017).

222 Fay, C. M. & Caroline, C. *Success of gene therapy to treat blind fades with follow-up*, <https://www.bloomberg.com/news/articles/2015-05-03/success-of-gene-therapy-to-treat-blind-fades-during-follow-up?utm_source=dlvr.it&utm_medium=linkedin> (2015).

223 Parkinson's Foundation. *Statistics on Parkinson's*, <http://parkinson.org/Understanding-Parkinsons/Causes-and-Statistics/Statistics> (2017).

224 Templeton, N. S. *Gene and Cell Therapy : Therapeutic Mechanisms and Strategies, Fourth Edition.* Vol. Fourth edition (CRC Press, 2015).

225 Sigma-Aldrich. (Sigma-Aldrich, sigmaaldrich.com).

226 VWR. (VWR, VWR.com).

227 Herper, M. *BioMarin Gene Therapy Yields Dramatic Results In Hemophilia*, <https://www.forbes.com/sites/matthewherper/2016/07/27/biomarin-gene-therapy-stops-yields-dramatic-results-in-hemophilia/#2ead27e6634d> (2016).

228 Chahal, P. S., Schulze, E., Tran, R., Montes, J. & Kamen, A. A. Production of adeno-associated virus (AAV) serotypes by transient transfection of HEK293 cell suspension cultures for gene delivery. *Journal of virological methods* **196**, 163-173 (2014).

229 ThermoFisher Scientific (ed ThermoFisher Scientific) (ThermoFisher Scientific, thermofisher.com).

230 Sigma-Aldrich. (ed Sigma-Aldrich) (Sigma-Aldrich, sigmaaldrich.com).

231 Grieger, J. C. & Samulski, R. J. Adeno-associated virus vectorology, manufacturing, and clinical applications. *Methods in enzymology* **507**, 229-254 (2012).

232 Grieger, J. C., Soltys, S. M. & Samulski, R. J. Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector. *Molecular Therapy* **24**, 287-297, doi:10.1038/mt.2015.187 (2016).

233 Carnes, A. E. & James A. Williams, P. D. in *Genetic Engineering & Biotechnology News* Vol. 32 (Genetic Engineering & Biotechnology News, www.genengnews.com, 2012).

234 Cecchini, S., Virag, T. & Kotin, R. M. Reproducible high yields of recombinant adeno-associated virus produced using invertebrate cells in 0.02- to 200-liter cultures. *Human gene therapy* **22**, 1021-1030 (2011).

235 Mietzsch, M. *et al.* OneBac: platform for scalable and high-titer production of adeno-associated virus serotype 1-12 vectors for gene therapy. *Human gene therapy* **25**, 212-222 (2014).

236 Galibert, L. & Merten, O.-W. Latest developments in the large-scale production of adeno-associated virus vectors in insect cells toward the treatment of neuromuscular diseases. *Journal of invertebrate pathology* **107 Suppl**, S80-93 (2011).

237 Luo, W.-Y. *et al.* Adaptive Immune Responses Elicited by Baculovirus and Impacts on Subsequent Transgene Expression In Vivo. *Journal of Virology* **87**, 4965-4973, doi:10.1128/JVI.03510-12 (2013).

238 Chen, C.-Y., Lin, C.-Y., Chen, G.-Y. & Hu, Y.-C. Baculovirus as a gene delivery vector: Recent understandings of molecular alterations in transduced cells and latest applications. *Biotechnology Advances* **29**, 618-631, doi:https://doi.org/10.1016/j.biotechadv.2011.04.004 (2011).

239 U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research & Center for Biologics Evaluation and Research. *Guidance for Industry: Q5A Viral Safety Evaluation of Biotechnology Products Derived From Cell Lines of Human or Animal Origin* (eds U.S. Department of Health and Human Services & Food and Drug Administration) (Food and Drug Administration, FDA.gov, 1998).

240 Hermens, W. T. J. M. C. & Smith, J. P. (Google Patents, 2014).

241 Aslanidi, G., Lamb, K. & Zolotukhin, S. An inducible system for highly efficient production of recombinant adeno-associated virus (rAAV) vectors in insect Sf9 cells. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5059-5064 (2009).

242 Urabe, M., Ding, C. & Kotin, R. M. Insect cells as a factory to produce adeno-associated virus type 2 vectors. *Human gene therapy* **13**, 1935-1943 (2002).

243 Wang, L., Blouin, V., Brument, N., Bello-Roufai, M. & Francois, A. Production and purification of recombinant adeno-associated vectors. *Methods in molecular biology (Clifton, N.J.)* **807**, 361-404 (2011).

244 Merten, O. W., Gény-Fiamma, C. & Douar, A. M. Current issues in adeno-associated viral vector production. *Gene therapy* **12 Suppl 1**, S51-61 (2005).

245 Smith, R. H., Levy, J. R. & Kotin, R. M. A simplified baculovirus-AAV expression vector system coupled with one-step affinity purification yields high-titer rAAV stocks from insect cells. *Molecular therapy : the journal of the American Society of Gene Therapy* **17**, 1888-1896 (2009).

246 BioMarin Pharmacuetical. *BioMarin's Investigational Gene Therapy for Hemophilia A at 6e13 vg/kg Dose Maintains Average Factor VIII Levels within Normal Range for over One Year*, <http://investors.biomarin.com/2017-07-11-BioMarins-Investigational-Gene-Therapy-for-Hemophilia-A-at-6e13-vg-kg-Dose-Maintains-Average-Factor-VIII-Levels-within-Normal-Range-for-over-One-Year> (2017).

247 Kohlbrenner, E. *et al.* Successful Production of Pseudotyped rAAV Vectors Using a Modified Baculovirus Expression System. *Molecular therapy : the journal of the American Society of Gene Therapy* **12**, 1217-1225, doi:10.1016/j.ymthe.2005.08.018 (2005).

248 Chulay, J. D. *et al.* 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**, 155-165 (2011).

249 Conway, J. E., Zolotukhin, S., Muzyczka, N., Hayward, G. S. & Byrne, B. J. 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**, 8780-8789 (1997).

250 Stutika, C. *et al.* Definition of herpes simplex virus helper functions for the replication of adeno-associated virus type 5. *Journal of General Virology* **96**, 840-850, doi:doi:10.1099/vir.0.000034 (2015).

251 Park, D., Lalli, J., Sedlackova-Slavikova, L. & Rice, S. A. Functional Comparison of Herpes Simplex Virus 1 (HSV-1) and HSV-2 ICP27 Homologs Reveals a Role for ICP27 in Virion Release. *Journal of Virology* **89**, 2892-2905, doi:10.1128/JVI.02994-14 (2015).

252 Clement, N., Knop, D. R. & Byrne, B. J. Large-Scale Adeno-Associated Viral Vector Production Using a Herpesvirus-Based System Enables Manufacturing for Clinical Studies. *Human Gene Therapy* **20 SRC - GoogleScholar**, 796-806 (2009).

253 Thomas, D. L. *et al.* Scalable recombinant adeno-associated virus production using recombinant herpes simplex virus type 1 coinfection of suspension-adapted mammalian cells. *Human gene therapy* **20**, 861-870 (2009).

254 U.S. Department of Health and Human Services, Food and Drug Administration & Center for Biologics Evaluation and Research. *Guidance for Industry: Preclinical Assessment of Investigational Cellular and Gene Therapy Products* (ed Center for Biologics Evaluation and Research) (Food and Drug Administration, FDA.gov, 2013).

255 U.S. Department of Health and Human Services, Food and Drug Administration & Center for Biologics Evaluation and Research. *Guidance for Industry: Characterization and Qualification of Cell Substrates and Other Biological Materials Used in the Production of Viral Vaccines for Infectious Disease Indications* (ed Center for Biologics Evaluation and Research) (Food and Drug Administration, FDA.gov, 2010).

256 Fenner, F. *et al.* Smallpox and its eradication. (World Health Organization, World Health Organization: Institutional Repository for Information Sharing, 1988).

257 Johnson, L., Gupta, A. K., Ghafoor, A., Akin, D. & Bashir, R. Characterization of vaccinia virus particles using microscale silicon cantilever resonators and atomic force microscopy. *Sensors and Actuators B: Chemical* **115**, 189-197, doi:https://doi.org/10.1016/j.snb.2005.08.047 (2006).

258 Isaacs, S. N. in *Vaccinia Virus and Poxvirology: Methods and Protocols* (ed Stuart N. Isaacs) 1-22 (Humana Press, 2012).

259 Clark, K. R., Voulgaropoulou, F., Fraley, D. M. & Johnson, P. R. Cell lines for the production of recombinant adeno-associated virus. *Human gene therapy* **6**, 1329-1341 (1995).

260 Gao, G. P. *et al.* High-titer adeno-associated viral vectors from a Rep/Cap cell line and hybrid shuttle virus. *Human gene therapy* **9**, 2353-2362 (1998).

261 Qiao, C., Wang, B., Zhu, X., Li, J. & Xiao, X. A novel gene expression control system and its use in stable, high-titer 293 cell-based adeno-associated virus packaging cell lines. *Journal of virology* **76**, 13015-13027 (2002).

262 Thorne, B. A., Takeya, R. K. & Peluso, R. W. Manufacturing recombinant adeno-associated viral vectors from producer cell clones. *Human gene therapy* **20**, 707-714 (2009).

263 Monahan, P. E. *et al.* Direct intramuscular injection with recombinant AAV vectors results in sustained expression in a dog model of hemophilia. *Gene therapy* **5**, 40-49 (1998).

264 Zhang, H., Xie, J., Xie, Q., Wilson, J. M. & Gao, G. Adenovirus-adeno-associated virus hybrid for large-scale recombinant adeno-associated virus production. *Human gene therapy* **20**, 922-929 (2009).

265 Smith, C. A., Woodruff, L. S., Kitchingman, G. R. & Rooney, C. M. Adenovirus-pulsed dendritic cells stimulate human virus-specific T-cell responses *in vitro*. *Journal of Virology* **70**, 6733-6740 (1996).

266 Volpers, C. & Kochanek, S. Adenoviral vectors for gene transfer and therapy. *The journal of gene medicine* **6 Suppl 1**, S164-171 (2004).

267 Kennedy, M. A. & Parks, R. J. Adenovirus Virion Stability and the Viral Genome: Size Matters. *Molecular Therapy: the Journal of the American Society of Gene Therapy* **17**, 1664-1666, doi:10.1038/mt.2009.202 (2009).

268 Bogedain, C., Maass, G. & Hörer, M. (Google Patents, 2002).

269 Suzuki, Y. *et al.* A novel high-throughput (HTP) cloning strategy for site-directed designed chimeragenesis and mutation using the Gateway cloning system. *Nucleic acids research* **33**, e109 (2005).

270 Duigou, G. J. & Young, C. S. H. Replication-competent adenovirus formation in 293 cells: the recombination-based rate is influenced by structure and location of the transgene cassette and not increased by overproduction of HsRad51, Rad51-interacting, or E2F family proteins. *Journal of virology* **79**, 5437-5444 (2005).

271 Lock, M. *et al.* Rapid, simple, and versatile manufacturing of recombinant adeno-associated viral vectors at scale. *Human gene therapy* **21**, 1259-1271 (2010).

272 Vandenberghe, L. H. *et al.* Efficient serotype-dependent release of functional vector into the culture medium during adeno-associated virus manufacturing. *Human gene therapy* **21**, 1251-1257 (2010).

273 Ayuso, E. *et al.* High AAV vector purity results in serotype- and tissue-independent enhancement of transduction efficiency. *Gene therapy* **17**, 503-510 (2010).

274 Potter, M. *et al.* A simplified purification protocol for recombinant adeno-associated virus vectors. *Molecular Therapy. Methods & Clinical Development* **1**, 14034, doi:10.1038/mtm.2014.34 (2014).

275 Doria, M., Ferrara, A. & Auricchio, A. AAV2/8 vectors purified from culture medium with a simple and rapid protocol transduce murine liver, muscle, and retina efficiently. *Human gene therapy methods* **24**, 392-398 (2013).

276 Kaludov, N., Handelman, B. & Chiorini, J. A. Scalable Purification of Adeno-Associated Virus Type 2, 4, or 5 Using Ion-Exchange Chromatography. *Human Gene Therapy* **13**, 1235-1243 (2004).

277 Qu, W., Wang, M., Wu, Y. & Xu, R. Scalable downstream strategies for purification of recombinant adeno- associated virus vectors in light of the properties. *Current pharmaceutical biotechnology* **16**, 684-695 (2015).

278 Zhou, J. *et al.* PEG-modulated column chromatography for purification of recombinant adeno-associated virus serotype 9. *Journal of virological methods* **173**, 99-107 (2011).

279 Wang, Y., Teraoka, I., Hansen, F. Y., Peters, G. H. & Hassager, O. A Theoretical Study of the Separation Principle in Size Exclusion Chromatography. *Macromolecules* **43**, 1651-1659 (2010).

CHAPTER 2: A NOVEL CONSTRUCTION OF RECOMBINANT ADENOVIRUS HARBORING ADENO-ASSOCIATED VIRUS VECTOR FOR USE IN UNIVERSAL PACKAGING CELL LINE IN SCALABLE VECTOR PRODUCTION

2.1 Overview

The versatility of recombinant adeno-associated viruses has garnered significant attention from investors, pharmaceutical companies, and regulatory agencies as more therapies using this vector are finding success in preclinical and late-stage clinical trials, even some reaching the market approval stage in the US. A number of vector production methods have been developed to generate the necessary clinical grade vectors that have high potency, and have high-titers to reduce general manufacturing and therapeutic costs. However, each of these production methods inherently have their own drawbacks. We previously developed a HEK293-cell based producer cell line method for high titer and high potency rAAV vectors. However, this method requires significant time-intensive generation of novel producer cell lines for every rAAV vector of interest and furthermore the E1-, E3-deleted adenovirus is only occupied at the E1-region, leaving the space available in the E3-region unused. Here, we discuss a novel adenovirus construct that utilizes a packaging cell line to generate a high titer and high quality rAAV. The improvements made in this system are: 1) the development of a one-step cloning of a rAAV vector cassette into the E1-region of the adenovirus using precision site-specific recombination, 2) use of a high Rep/Cap copy number packaging cell line to universally package the rAAV vector provided in trans by the adenovirus, 3) high vector yields on different rAAV inverted terminal repeat designs harbored stably in the adenovirus, and 4) high quality, low empty-particle

containing rAAV product. The vector yields were consistent between the two adenovirus constructs, with about 1 to $2x10^{12}$ vector genome particles per 20 15-cm plates. Furthermore, these two adenovirus constructs yielded high infectivity product. In conclusion, the simple and versatile adenovirus construct coupled with the packaging cell line can be useful for large scale AAV vector production as the vector finds its way to late stage clinical trials and market approval.

2.2 Introduction

Recombinant adeno-associated virus (rAAV) vectors are a powerful tool for research labs for *in vitro* and *in vivo* studies. The popularity of this viral vector is escalating, evident in the increase of use in gene therapy clinical trials in the last decade and lately gaining acknowledgement as a potential vector for use in gene editing *in vivo*[1-3]. The success of preclinical and subsequent early-stage clinical trials is now beginning to crescendo, reaching Phase III trials and some rAAV therapeutics are recommended for or have been given approval at US or European markets. The rAAV therapeutics that have been or will soon be approved are Glybera™ (alipogene tiparvovec, UniQure, AAV1 containing an intact copy of lipoprotein lipase) and Luxturna™ (voretigene neparvovec, Spark Therapeutics, AAV2 containing human RPE65 cDNA for inherited retinal diseases), respectively. Glybera™ has been pulled off the market because it has received 1 patient during its 5-year span of market approval, at a cost of $1 million per dose. Luxturna™ is slated to be priced at $1 million per dose, making critics of the therapeutic question where is the cost-sink is occurring and whether this therapeutic can be reduced in price. Arguably, one major cost-sink is the scalable production method to supply the necessary dose per patient. At the scale and method of production for Luxturna™, the manufacturing output is relatively small given the localized region of vector delivery, thereby requiring less resource-intensive equipment, labor, and space. However, as the diseases broaden from localized to systemic diseases, (e.g. muscular dystrophies) the manufacturing output will have to be amplified by 2-3 orders of magnitude, making an overwhelming demand that requires more robust scalable production methods.

For rAAV vector production, four popular strategies differing in process design are being used. The most widely used method, including the disclosed manufacturing method for Spark Therapeutics' Luxturna™ product, is based on the helper-virus-free transient transfection

method of all cis and trans components (vector plasmid, packaging plasmid, and helper genes isolated from adenovirus) into the host cell HEK293. While this method is simple in vector plasmid construction, generating high-titer rAAV, and free of any non-product helper or subhelper viruses (e.g. recombinant herpes simplex virus (rHSV), adenovirus, or vaccinia virus (VV)), this method is simply too resource-intensive for scalability, especially for higher demand clinical trials. A second popular strategy is the rHSV-based AAV production system, using rHSV vectors to bring rAAV vector and AAV replication (Rep) and capsid (Cap) genes into the fast-growing Baby Hamster Kidney (BHK) cells. The third popular strategy is an adapted baculovirus system requiring a single baculovirus vector to deliver the rAAV vector cassette into a stable transcriptionally silent inducible Rep/Cap packaging insect cell line. For both of these systems, it is inconvenient to prepare large quantities of helper and vector viruses and maintain their purity and stability. The fourth system is based on the AAV producer cell lines derived from HeLa or stably integrated AAV Rep/Cap genes in a HeLa cell line, called the A549 cell line. The rAAV vector cassette was either stably integrated in the host genome or introduced by a recombinant adenovirus that contained the cassette. Although this HeLa cell line method is easy to scale up and produces relatively high titers of AAV vector comparable to transient transfection method, these cell lines required wild-type adenovirus as the helper. Contamination of wild-type adenovirus in the final vector preparations is an undesirable by-product.

To eliminate transient transfection step and avoid use of wild-type helper adenovirus, we established rAAV producer cell lines using HEK293 cells as the parental cell line. Integrated with the early genes E1A and E1B from the adenovirus based on work by Frank Graham in the 1970s, the HEK293 cells were further adjusted by Qiao et. al. to include a well-designed inducible AAV Rep and Cap genes giving the ability to use an E1A/E1B-defective adenovirus for helper functions. The safety profile of the E1-deleted adenovirus is better than wild-type

adenovirus and has been widely used as a gene therapy vector in human clinical trials. However, the existence of the adenovirus type 5 E1 region integrated into the HEK293 genome has made generation of a stable packaging and producer cell line difficult due to the E1 activation of p5 and p19 promoters located in the Rep gene, that leads to generation of AAV replication proteins (Rep78, Rep52, Rep68, and Rep40) that are cytostatic or cytotoxic if constantly expressed [4,5]. Tightly regulating the p19 promoter is difficult because of its location within the coding region of Rep78 and Rep68. In our lab, we established an innovative inducible Rep expression, called the dual splicing switch. Briefly, an intron - containing two LoxP sites that flank a drug-resistant gene and three polyadenylation elements - was inserted into the Rep gene-coding region disrupting all Rep transcripts. The Rep gene was restored to its original expression by providing an E1- and E3- deleted adenovirus containing the Cre-recombinase gene in the E1 region, called Ad-Cre(E1). Here, both DNA splicing by Cre-LoxP and RNA splicing to remove the intron, and hence the nomenclature of dual splicing, reconstitute and reactivate Rep gene expression on the AAV packaging or producer cell line[5]. With the Rep gene spliced together to its wild-type form, the E1A/E1B of the HEK293 cell activates the expression of Rep. The Rep and Cap proteins together identify the ITRs, package the transgene into a viral vector, and then can be collected and purified at a high titer and high potency. This method was able to package the 5-kb single-stranded rAAV DNA and can also package the 2.5 kb double-stranded rAAV DNA – 1.2 kb accounting for a common promoter sequence – which is shown to accelerate the transgene expression despite the decrease in the genome size[6]. Further work using this inducible system led to a rAAV producer HEK293 cell line that could increase titer of rAAV and continue to avoid use of wild-type adenovirus. In this work, the rAAV transgene flanked by the ITRs was placed downstream to the inducible Rep/Cap gene using the Gateway system, and this plasmid

construct was linearized and integrated randomly into the HEK293 genome[7,8]. The dual splicing switch is illustrated in Figure 8.

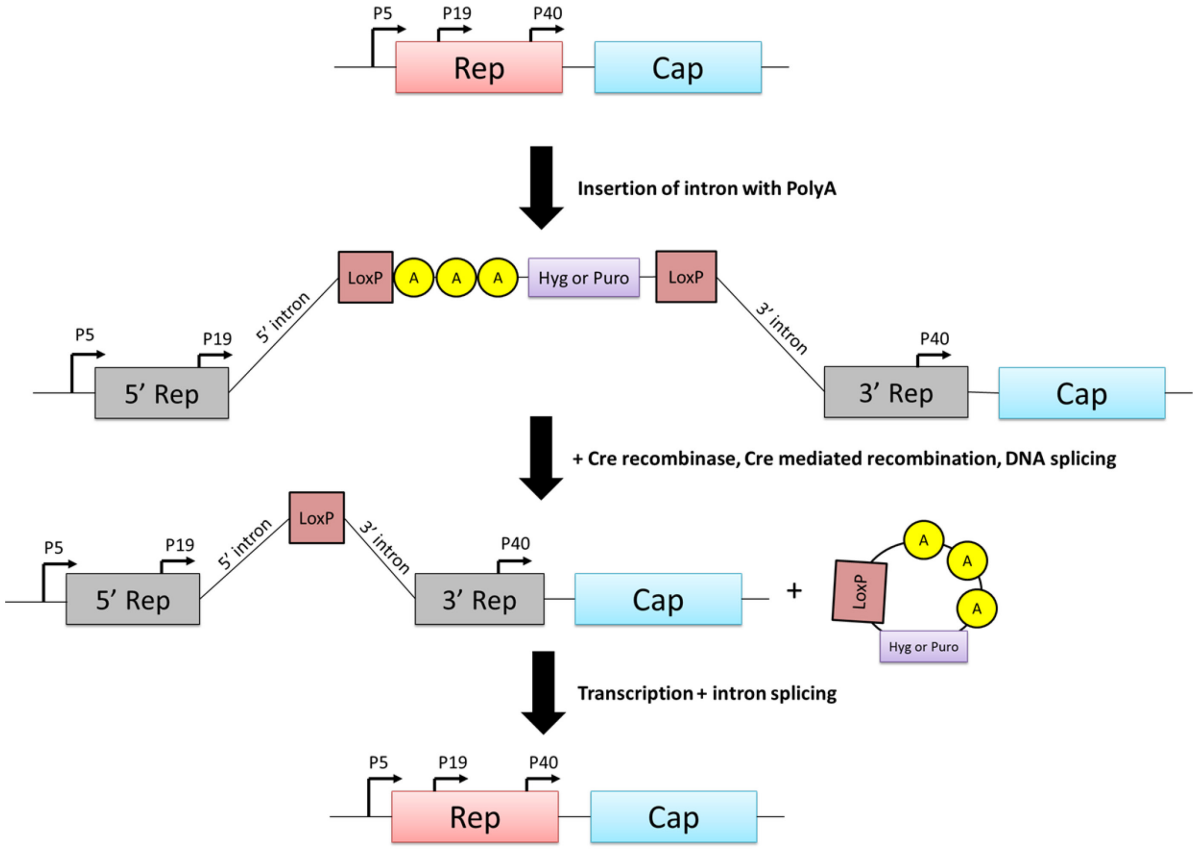


Figure 8: Dual splice switch depiction. The termination cassette Int-3A-Hyg (or Puro) was inserted into the shared Rep coding sequence downstream of promoter p19 to block Rep gene transcriptions. A Cre recombinase is provided in trans by an adenovirus infection. The enzyme recognizes the two LoxP sites and splices out the inserted DNA fragment that contains the PolyA sequences between the LoxP sites. The removal of the PolyA sequences allows transcription to proceed and full-length mRNA to be generated. After RNA splicing, the inserted intron is precisely removed from the full-length mRNA, and the coding sequence is restored.

However, this approach has two significant issues. Firstly, the adenovirus harboring the Cre recombinase gene in the E1 region is also deleted in an unused E3 region. This region is large enough to accommodate the Cre recombinase gene and the larger E1 region can be used to hold another component, such as the rAAV vector. Secondly, it requires a time-consuming selection and characterization process of each HEK293 cell clone to confirm stable expression of the introduced plasmid, albeit shorter than the previous method established by Qiao et. al. in our lab in 2002[5]. This issue is complicated further because for every rAAV vector of interest, a novel cell line of sufficient quantity of integrated plasmid has to be generated, thereby restricting that cell line for generation of the particular rAAV vector. The second issue will be addressed in further detail in Chapter 3, but the results from Chapter 3 will be used to test the hypothesis in the current chapter. A third issue that arises from the development of an E1 containing gene in adenovirus is the need to recombine linearized pShuttle with the pAd-Easy vector in BJ5183 bacteria for every gene of interest to be placed in E1. This complicates the design of every rAAV vector intended to be in the E1 region because specific restrictions need to be identified to insert the rAAV vector without effecting the adenovirus shuttle plasmid or the overall adenovirus plasmid.

It has already been reported that the E3 region is not vital for adenovirus. The function of the E3 region is to subvert the host immune response and allow persistence of infected cells, which is only vital for *in vivo* applications of this virus[9]. Therefore, its deletion is not important for this *in vitro* method and together with an E1-deletion, the adenovirus can offer a total of 8 kb of transgene capacity[10]. Furthermore, since there are a multitude of rAAV vectors in mind, the need to quickly insert the rAAV vector into the E1 region of this novel adenovirus is necessary. Therefore, we have developed an adenovirus that takes advantage of the unused E3 region of the E1-, E3- deleted adenovirus to insert the Cre recombinase gene, and the E1-region for

insertion of the Gateway destination cassette that can allow for efficient, reproducible rAAV vector insertion by Gateway specific recombination into the E1 region. This would still be a single virus infection, but instead of integrating a large plasmid for every rAAV gene of interest, like in the producer cell line method developed in our lab, an easier universal inducible packaging cell line can be developed for a particular AAV serotype or capsid variants. The inducible system is activated with the Cre recombinase gene provided in trans by an adenovirus that also holds the rAAV vector. This prompts the following general hypothesis for this aim: Ad-Cre(E3)-rAAV(E1) can produce high potent, high quality, and high titer rAAV, both single- and double-stranded of a variety of transgenes, when infected into a universal rAAV packaging HEK293 cell line.

2.3 Material and Methods

Plasmid Construction for Adenoviral Vectors Using Molecular Cloning and Gateway System

To redesign the E1-, E3- deleted adenovirus to contain a Cre-recombinase gene in the E3 and the rAAV transgene in the E1, the large adenovirus plasmid has to either be subdivided in a separate plasmid to not affect other regions of the adenovirus plasmid or a very unique restriction site must be utilized during the molecular cloning, especially when considering the AdEasy system in this aim. We used pAd-Easy-1, an E1 and E3 deleted first generation adenovirus plasmid, to first clone the Cre gene into the E3 region. Several unique plasmids had to be generated to accomplish this task. First, a PacI/SpeI was inserted into pBSKS by PCR to form pBSKS-PacI-SpeI. This was digested at these two sites to allow for ligation of the E3 region of the pAd-Easy-1 plasmid removed also using PacI/SpeI, hereby the ligated product is called pBSKS-Ad-E3. Next, a BamHI site was introduced into E3 region 5' to a PolyA signal present in the E3, resulting in pBSKS-Ad-E3-BamHI. In parallel to this cloning work, a CMV-Cre-sPa was prepared by cloning the Cre gene between a CMV promoter and a SV40 PolyA signal (sPa), using a cloned 5'-SalI-Cre-XhoI-3' and ligating it into pAAV-CMV-sPa at the XhoI site (XhoI and SalI have compatible cohesive ends). Here, the entire CMV-Cre-sPa cassette, which was temporarily cloned into pAAV-CMV-Cre-sPa, was digested out using BglII digestion. This cassette was then ligated into the BamHI site of pBSKS-Ad-E3-BamHI, since BglII and BamHI have compatible cohesive ends. Finally, the fragment containing the CMV-Cre-sPa gene in the E3 region is cleaved using PacI and SpeI again and ligated into the original pAd-Easy-1 vector, hereby known as pAd-Easy-1-Cre(E3). Functionality was verified using a co-transfection with LacZ-2LoxP-in in HEK293 cells, as shown in the results section. The LacZ-2LoxP-in is an inducible LacZ reporter gene that is silenced with the LoxP intron cassette found in our inducible Rep/Cap dual splice switch system. As a control, a Cre carried by an AAV plasmid was

co-transfected with the inducible LacZ cassette. As a negative control, the inducible LacZ cassette was transfected to show that LacZ expression is controlled by the inserted intron.

The AdEasy system was developed as a simple and efficient method for generation of recombinant adenoviruses[11,12]. The recombination step, in particular, allows for performance in E. coli rather than in mammalian cells and takes advantage of the highly efficient homologous recombination that occurs in bacteria. The system can be divided into three steps. 1) Subcloning the genes of interest into the shuttle vector containing two "arms" of viral sequence for homologous recombination with the adenoviral backbone vectors. 2) Generate a recombinant adenoviral plasmid between the PmeI linearized shuttle vector and the supercoiled backbone vector in specific bacterial cells called, BJ5183 cells using kanamycin selection. 3) Generate a recombinant adenovirus by transfecting the recombinant adenoviral DNA into HEK293 cells. For the first step, we decided to insert the Gateway destination cassette into the E1 region to reduce time to produce novel adenoviruses. Our goal was to use the pAd-Cre-E3-DEST product generated from the recombination system in AdEasy, and deliver the AAV vector using the Gateway entry vector methods. However, with fears of the attR1 and attR2 recombining out from the BJ5183 recombination step, we decided to introduce a unique restriction site not found anywhere else on the original pAd-Easy-Cre-E3 plasmid. Unfortunately, no such single cut enzyme existed in the E1 that did not cut anywhere else in the adenovirus plasmid, and therefore we needed to insert a unique restriction site by PCR. To design the shuttle vector that carries the restriction site into the E1 region of the adenovirus, we used similar PCR techniques to introduce the restriction site SwaI in the multiple cloning site of the pShuttle vector. The pShuttle-SwaI vector was linearized using PmeI and was transformed with supercoiled pAd-Easy-Cre-E3 into BJ5183 cells. The SwaI in the E1 region is now part of the pAd, hereby pAd-Cre-E3-SwaI-E1. The Gateway destination vector was cloned out of pSPG8-DEST, a

destination vector that was used by Yuan in his 2011 producer cell line system[8]. The cloned area, which includes all components of the Gateway destination cassette – attR1, CmR gene, ccdB gene, attR2 – was cloned with SwaI sites flanking the ends of the cassette. Next the pAd-Cre-E3-SwaI-E1 was digested with SwaI and the destination cassette – 5' SwaI-attR1-CmR-ccdB-attR2-SwaI 3' – was ligated specifically into the E1 region, resulting in the pAd-Cre(E3)-ccdB vector.

With the Gateway destination vector stably located in the E1 region of the adenoviral plasmid, we can now use Gateway recombination to incorporate the rAAV transgene in the E1 region by the Gateway entry cassette. Gateway technology was used to introduce two adenovirus constructs: (1) a rAAV with the transgene CMV-EGFP flanked by double-stranded ITRs, and (2) a rAAV with the transgene CMV-LacZ-nLs flanked by sub201 ITRs. The entry vector contains the AAV transgene flanked by the attL1 and attL2. Using an LR clonase reaction in the Gateway system, a system that uses lambda recombination to facilitate recombination of an attL substrate with an attR substrate, the entry vector containing the AAV transgene recombines into the E1 region of pAd-Cre(E3)-ccdB, yielding a pAd-Cre(E3)-rAAV(E1) plasmid[7]. Two designs were made using this method: pAd-CMV-Cre (E3)-AAV-ds-GFP(E1) and pAd-CMV-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1). A figure showing the differences in the method is displayed on Figure 9. Although it appears to be more complex in Figure 9B, the outcome is easier to perform than it is for those described in Figure 9A.

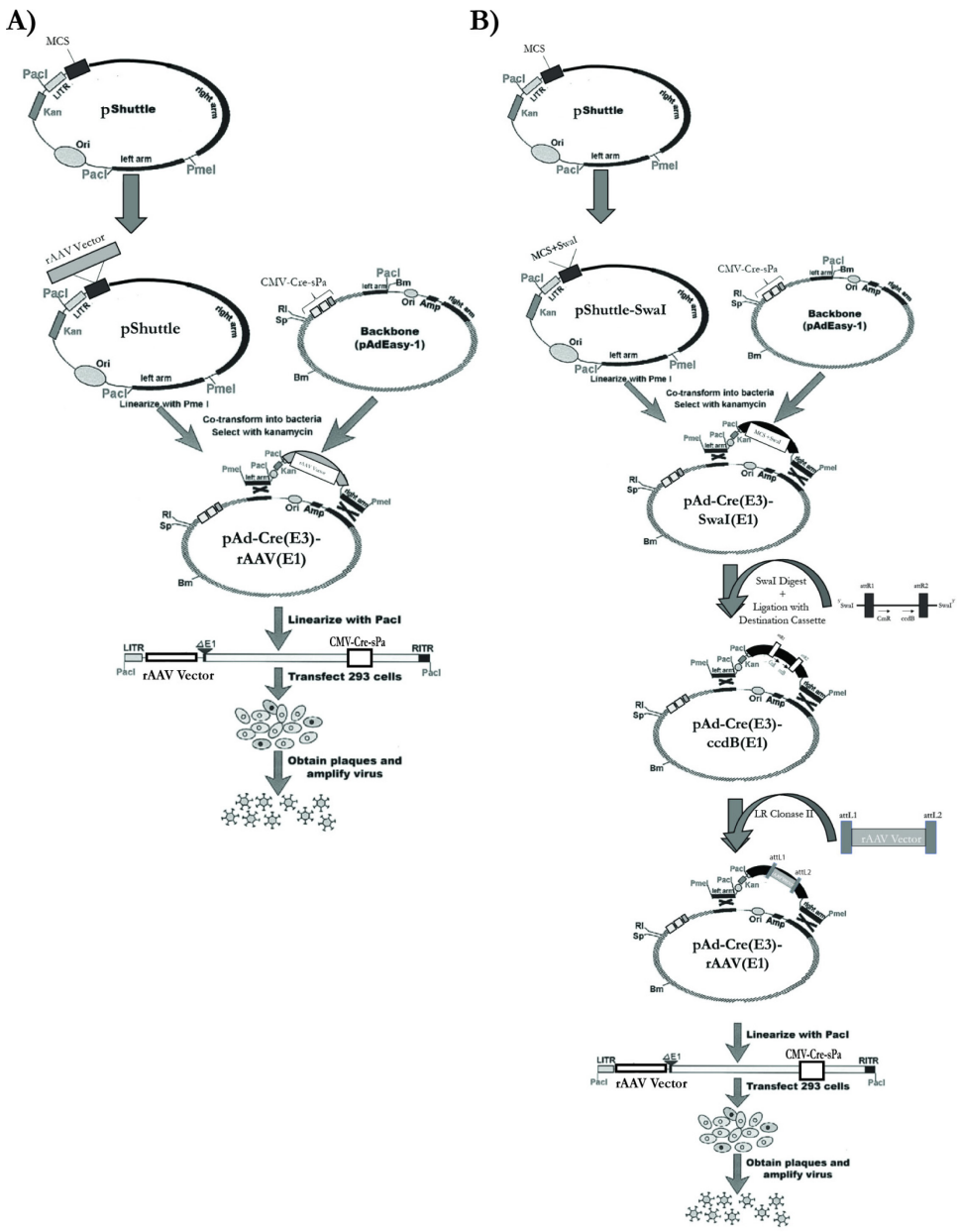


Figure 9: Use of Gateway technology for adenovirus construct. A) Traditional methods to generate Ad-rAAV(E1) using AdEasy protocol. B) Novel method to generate Ad-rAAV(E1) by reducing overall steps of novel rAAV vector harboring adenoviruses by inserting Gateway destination cassette into a unique sequence of the E1 region, introduced into the adenovirus plasmid by a modified pShuttle vector in the AdEasy system. Figure adapted from He et al. (1998). See methods for details.

Cells and Viruses

This plasmid, now with the necessary insert, follows the latter portions of the AdEasy method which are a PacI digestion to linearize the plasmid before transfecting the plasmid into HEK293 cells[11]. The recombinant adenoviruses begin to form plaques after about 14 days, depending on the transfection efficiency (i.e. plaques can form as early as 10 days). The plaques are collected by scarping, centrifuged, and resuspended in PBS. The collected cells undergo three cycles of freeze and thaw and vortex, being mindful of avoiding warming virus supernatants longer than necessary. The cellular debris is spun out of solution and the remaining lysate is used to infect HEK293 cells for expansion. It may take two to four rounds of amplification to arrive at a large-scale preparation of high titer adenoviruses. The number of amplification rounds is dependent on the initial titers of the primary transfection lysates. When scaled to 20 15-cm plates, we purified and concentrated adenovirus in CsCl density gradients[11]. The viruses, Ad-Cre(E3)-AAV-ds-CMV-GFP(E1) and Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1), were stored in an equal volume of adenovirus storage buffer (Storage Buffer (2x): 10 mM Tris Cl, pH 8.0, 100 mM NaCl, 0.1% bovine serum albumin (BSA), 50% (v/v) glycerol, filtered to sterilize).

We used two cell lines developed by Qiao et. al. in 2002 called XX2-GFP-145, an AAV2-GFP producer cell line, and XX2-in-19-sub, an AAV2 parental packaging cell line[5]. These were used primarily to assist in initial characterization studies of the Ad-Cre(E3)-rAAV(E1) virus, particularly for Ad-Cre(E3) and the Ad-Cre(E3)-AAV-ds-GFP(E1) constructs. We also use cell lines developed in Chapter 3 for a more robust study of the Ad constructs, particular of the Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs construct. These cell lines are XX2-in-Puro-AAVS1 Clone 152.69, and XX2-in-Puro-AAVS1 Clone 152.74. XX2-in-Puro-AAVS1 Clone 152.74 was used to provide MOI characteristics for Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1). At the time of performing MOI studies for Ad-Cre(E3)-AAV-ds-GFP(E1), we used XX2-GFP-145, despite the

generation of two undifferentiable vectors: AAV-ds-GFP and AAV-ss-GFP. When the cell lines XX2-in-Puro-AAVS1 and XX2-in-Puro were generated, supply of the Ad-Cre(E3)-AAV-ds-GFP(E1) and timing of experiments was only sufficient for large scale studies and Rep gene amplification. Therefore, the MOI was determined from results obtained from characterizations in XX2-GFP-145. Ad-Cre(E3) and Ad-Cre(E1) are used to provide controls for the function of the Cre gene in our new Ad constructs. The other virus used in this aim is the negative control virus Ad-GFP.

Infection assays are used to assess production capabilities of AAV-GFP and AAV-LacZ in small scale. Products that contain both AAV and adenovirus are subjected to 56 °C for 60 minutes to inactivate the adenovirus so that expression is coming from AAV and not influenced by adenovirus. AAV-LacZ expression was evaluated using X-gal staining. If AAV expression were to be accelerated in HEK-293 cells, 4 μM of Hoechst 33342 (ThermoFisher, H3570) dye was used with expression occurring 48 hours post infection.

All other cell culture chemicals or buffers used to perform the tasks in this aim, such as transfection agents, CsCl for gradient, trypsin, or media, are used commonly in the lab.

Plasmid Equimolar Rescue and Viral Rescue Experiment Designs

During characterization, we use a variety of plasmids to see if the rAAV vector is able to be rescued in plasmid and in viral form. First, to test if the rAAV vector is able to rescue in the plasmid form, especially for the pAd-Cre(E3)-rAAV(E1) constructs we supplied $1x10^7$ HEK293 cells an equimolar quantity of rAAV production components via calcium phosphate transfection. In this experiment, 4 unique sets of experiments were performed: 0.6 pmoles of pAd-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) + 0.8 pmoles of pXX2, 0.6 pmoles of pAAV2.1-CMV-LacZ-nLs(E1) + 0.8 pmoles of pXX2 + 0.6 pmoles of pHelper, 0.6 pmoles of pAd-Cre(E3)-AAV-ds-GFP(E1) + 0.8 pmoles of pXX2, and finally 0.6 pmoles of pEMBL-AAV-ds-

GFP(E1) + 0.8 pmoles of pXX2 + 0.6 pmoles of pHelper. The cells were harvested 72 hours post transfection, subjected to sonication to break apart the cells, spun down to remove cellular debris, and then an aliquot of the samples was clarified using methods to limit interference in qPCR quantification. These methods include an initial DNase reaction to remove preexisting DNA, followed by addition of EDTA to disrupt DNase activity, followed by Proteinase K to break apart the viral particles, and ending with a Phenol/chloroform extraction for cleaning of the crude lysate. The remaining DNA after phenol/chloroform extraction was precipitated using standard ethanol precipitation. For the quantification of the viral particles, we used CMV promoter FAM probes and primers for real-time quantitative assay (ABI PRISM 7700 Sequence Detector, Applied Biosystems). The qPCR materials are: CMV Probe 5'-FAM-TCAATGGGTGGAGTATTTA-3', CMV Forward: 5'-GTATGTTCCCATAGTAACGCCAATAG-3', and CMV Reverse: 5'-GGCGTACTTGGCATATGATACACT-3'.

The second set of experiments we were interested in the viral rescue of the rAAV, especially for the Ad-Cre(E3)-AAV(E1) constructs. Since we do not have an adequate packaging cell line (see Chapter 3), with the only functional cell line also harboring its own rAAV vector, generating a mixed population of resulting vector, we had to find a new strategy to validate the rAAV rescue in the viral form. We used a variety of plasmids including Rep/Cap genes such as pXXHH67.2m which contains replication gene from AAV2 and the capsid of our unique capsid design HH67-double mutant, as well as a Rep/Cap gene where the Rep is disrupted by an intron and can be restored with Cre-LoxP recombination. We also used a plasmid to showcase the mixed population phenomena we observed in our producer cell line. This plasmid was pSPHH67.2m-AAV-ds-GFP, which is the disrupted Rep with an AAV-ds-GFP vector located downstream to the Rep/Cap. Finally, to validate the system we used in Chapter 3, we performed

viral rescue using pXX2-SseI-2LoxP-Puro to generate AAV2-CMV-LacZ-nLs from our Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) virus.

Histochemical Staining of Monolayer Tissue Culture Cells for LacZ Activity (X-gal Staining)

For cells that express the LacZ cassette, we used X-gal staining. Cultured cells were rinsed with 1x Phosphate Buffered Saline (PBS, pH 7.3) and then fixed for >5 min at 4 °C in 2% formaldehyde and 0.2% glutaraldehyde, mixed in PBS. The cells were then overlaid with a histochemical reaction mixture containing 1 mg/mL 4-Cl-5-Br-3-indolyl-β-galactosidase (X-gal), 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, and 2 mM $MgCl_2$. The X-gal was dissolved in dimethoxysulfoxide (DMSO) at 40 mg/mL, and then diluted into the reaction mixture. Incubation was for 8-14 hours at 37 °C.[13] Cells were then counted for LacZ expression, i.e. blue cells were counted.

Infectious Titer Calculation

HEK293 cells were used for infectious titer assays. For LacZ expression, cells were quantified for infectious titer, by counting number of blue cells in the microscope or camera field under bright-field microscopy. For GFP, cells were quantified by counting number of green cells in the microscope or camera field under fluorescent microscope (excitation of GFP is 488 nm and emission is 509 nm). The yield of infectious particles was presented as infectious unit. The units of infectious units (IU) were either IU/mL or IU/10-cm plate. IU/10-cm plate was calculated by extrapolating from the surface area of the microscope field to the surface area of the well being used to for the infection assay. IU/mL was calculated by dividing the extrapolated well infectious unit by the total volume infected into the well. Larger scaled products were typically presented in IU/mL, whereas smaller scale characterization studies were presented in IU/10-cm plate. Both measurements were consistently used to compare infectivity at a specific scale between various production methods.

AAV Vector Production and Purification

For the production of AAV2, three systems were used: Ad-Cre(E1) infection into XX2-GFP-145 producer cell lines, Ad-Cre(E3)-rAAV(E1) into XX2-in-Puro-AAVS1 cell lines, and triple plasmid calcium phosphate transfection (pXX2, pAAV2.1-CMV-LacZ-nLs or pEMBL-AAV-CMV-ds-GFP, pHelper). These were performed on 20 15-cm plates.

Briefly, the HEK293 cells or cell lines were propagated in Dulbecco's modified Eagle's medium (DMEM) in 15 cm plates supplemented with 10% fetal bovine serum (FBS) at 37°C with the appropriate drug selection, if applicable: puromycin (1 μg/mL) for cell lines with the puromycin resistant gene and hygromycin (100 μg/mL) for XX2-GFP-145 that has a hygromycin resistant gene. For triple plasmid transfection, cells were transfected at 80-90% confluency. Each plate was transfected with 16.7 μg vector plasmid, 16.7 μg adenovirus helper plasmid, and 16.7 μg AAV2 packaging plasmid dissolved in 40 mL of 0.25M $CaCl_2$. For every 2 mL of dissolved DNA in 0.25M $CaCl_2$, 2 mL of 2x HEPES-buffered saline (HBS buffer: 50 mM HEPES, 280 mM NaCl and 1.5 mM Na_2HPO_4; pH 7.1). Eight to twelve hours later, the medium was replaced with fresh DMEM supplemented with 2% FBS. Cells were harvested at 72 hours post transfection, and media was also collected for AAV purification. Infected cells were harvested at 48 hours post infection after achieving cytopathic effect (CPE).

Cells were resuspended in suspension buffer I (50mM HEPES, 150 mM NaCl, 50 mM NaH_2PO_4, 2 mM $MgCl_2$, 2.5 mM KCl, pH 8.0) and sonicated. The cell lysate was treated with DNase (100 units/ml) and RNase A (4 units/ml) and incubated at 37°C for 1 hour. Debris was removed by centrifugation at 2,500 rpm at 4°C for 15 minutes. For preparations using adenovirus, a heat inactivation of 56 °C was performed for both the collected media and the cell lysate before moving onto the next steps.

PEG-8000 and NaCl solutions were added to the clarified lysate, to a final concentration of 8% PEG-8000 and 0.5 M NaCl, and incubated at 4°C overnight. For the culture medium, powdered forms of PEG-8000 and NaCl were added to also reach 8% PEG-8000 and 0.5 M NaCl and incubated at 4°C overnight. The cell lysate and medium were centrifuged at 2,500 rpm for 30 minutes and the resulting pellets from both cell lysate and medium were combined and thoroughly resuspended in about 20 mL of resuspension buffer #2 (50 mM HEPES, 150 mM NaCl, 1% Sarkosyl, 20 mM EDTA, pH 8.0). The solution was place in an ultracentrifuge at 31,000 rpm for 16 hours in a CsCl density gradient. The AAV band was collected and subjected to a second round of CsCl density gradient ultracentrifugation at 38,000 rpm for 48 hours. The AAV band was collected in drop-wise fractions and stored at -80°C. Vector titers were determined by the DNA dot-blot method.

AAV titer was quantified by dot blot hybridization. Five microliters of AAV stock was added to 200 μL DMEM and treated with 50 μg/ml DNase I at 37°C for 1 hour to degrade unencapsidated DNA. Then, 200 μL proteinase K buffer (20 mM Tris Cl pH 8.0, 20 mM EDTA pH 8.0, 1% SDS) was added to inhibit DNase activity, followed by addition of 40 μg of proteinase K, and the sample was incubated at 55°C for 1 hour to degrade the capsid. The vector DNA was precipitated using ethanol precipitation with addition of glycogen (40 μg) to visually observe the pelleted DNA. The resultant pellet was resuspended in alkaline buffer (0.4 M NaOH and 10 mM EDTA pH 8.0) and bound to a hybridization transfer membrane (PerkinElmer). A standard of the original vector plasmid was applied to the same membrane. A biotin-labeled probe against CMV, LacZ, or GFP was hybridized to the membrane at 55°C overnight. Bound probe was detected using the North2South Chemiluminescent Nucleic Acid Hybridization and Detection Kit (Pierce). AAV titers were in the range of 10^{11} to 10^{13} vector genomes (vg)/ml.

Heparin Sulfate Column Chromatography

Buffers used in chromatography: Buffer A: 50 mM NaH_2PO_4, 2 mM $MgCl_2$, 2.5 mM KCl. pH 7.5-7.6, 0.22 μm filtered; Buffer B: 1x Buffer A, 1 M NaCl, 25 mM HEPES. pH 7.4, 0.22 μm filtered; Virus Resuspension Buffer: 1 x Buffer A; 50 mM HEPES; 150 mM NaCl. pH 8.0, 0.22 μm filtered

AAV serotype 2 was generated using two unique methods at a scale of 20 15-cm plates each: triple plasmid transfection (pXX2, pAAV2.1-CMV-LacZ-nLs, pHelper) and adenovirus in packaging cell line (Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) infected at approximately 10 MOI into XX2-in-Puro-AAVS1 Clone 152.69 and Clone 152.74). Triple plasmid transfected cells were harvested 72 hours post transfection, whereas packaging cell line methods were harvested when cytopathic effect (CPE) showed at 48 hours post infection. Cells were spun and media was aspirated. 30 mL of virus resuspension buffer was used to resuspend the collected cells, cells were subjected to three cycles of freeze-thaw. DNase was added to a final concentration of 50 U/mL and digested for 1 hour at 37 °C. Sarkosyl was added to a final concentration of 1% and this was mixed for 30 minutes at room temperature. Next, the colysate was clarified in two spins. The first spin the lysate was centrifuged at 8,000 rpm for 10 minutes at 16 °C to remove debris. The supernatant was collected from this first spin. The remaining debris from the first spin was resuspended with about 10 mL of virus resuspension buffer and was centrifuged according to the first spin, and again the supernatant was collected. The clarified lysates were filtered through a 0.8 μm filter.

The column was 1 mL HiTrap® Heparin High Performance (obtained from Sigma-Aldrich: GE Healthcare, 17-0406-01). Due to the materials and equipment at hand, the columns were operated by hand. Concentrations of wash and elution buffers were rationalized according to past experience with gradient-driven heparin sulfate chromatography and based on salt

content of the buffer used during these steps. The column was pretreated according to the following steps to remove stored 20% ethanol: wash with 10 volumes (i.e. 10 mL) of water followed by 10 volumes of 1x Buffer A. The viruses were loaded onto the column. Next, 5 volumes of 15% Buffer B were used to rinse the loaded viruses from other contaminating materials. The viruses were eluted in 5 volumes of 60% Buffer B. The viruses that still remain on the column were then fully eluted from the column using 5 volumes of 100% Buffer B. All eluants after the virus was loaded were collected to measure vector titers.

The columns were cleaned with 1x Buffer A, water, 70% ethanol, and were stored in 20% ethanol for reuse.

Negative Staining for Transmission Electron Microscopy

The Microscopy Services Laboratory (MSL) in the Department of Pathology & Laboratory Medicine at UNC-Chapel Hill was used to perform negative staining and transmission electron microscopy. Carbon-coated formvar grids (01754-F Formvar/Carbon 400 mesh, Copper) were glow discharged using Pelco easi-Glow unit to render the surface hydrophilic. The viruses adsorbed onto the grid and subsequently stained using the "Grid-on-Stain or 'Float' Method). Briefly, 25 μL of sample was placed onto a hydrophobic surface and the grid was placed film-side down onto the surface of the droplet for approximately 5 minutes to allow for viral adsorption onto the grid. Next, the grid was quickly transferred to two subsequent droplets of filtered deionized water to remove salts/fixative before staining and the grid was then placed film-side down onto a drop of 2% uranyl acetate in water (pH 4.5, 1-gram uranyl acetate dissolved in 50 mL deionized water for 20-30 minutes, spun for 10-15 minutes to remove debris that may interfere with microscopy) for one minute. Excess stain was wicked off by touch the edge of the grid to filter paper and the grid was then air dried. Grids were then

loaded onto a JEOL JEM 1230 Transmission Electron Microscope and images were taken between 100,000X and 150,000X. Data was analyzed for images in 100,000X or 120,000X.

Exclusion criteria and observational criteria to measure full and empty rAAV particles as a result of the 2% uranyl acetate stain are given in detail in Appendix A. Furthermore, all TEM images taken for each production method have been included in Appendix A for further analysis of the reader.

2.4 Results

Functionality of the Cre-recombinase gene was tested using a LacZ assay, where activation of LacZ gene is controlled by the dual-splicing switch. Cre recombinase provided in trans by an adenovirus infection would recognize the two LoxP sites and splice out the inserted DNA fragment that contains the PolyA sequences between the LoxP sites. The removal of PolyA sequences then allows transcription to proceed and full-length mRNA is generated. After RNA splicing, the inserted intron is precisely removed from the full-length mRNA, and the coding sequence is restored. This phenomenon is shown brilliantly for the Cre in the E3 region of the adenovirus in the plasmid form, as shown in Figure 10A. Here as a positive control, an AAV construct harboring a Cre recombinase gene was transfected into the HEK293 cells with the plasmid carrying the LacZ interrupted by the dual splice switch. As a negative control, no Cre recombinase was provided into the cells. The blue cells that appear in the third panel of Figure 10A are leaky expressing LacZ proteins that can arise because the expression of the LacZ gene cannot be fully turned off by the inserted intron.

Although functionality can be great in plasmid form, it may not translate well in the viral form. Therefore, we applied the latter half of the AdEasy protocol and generated Ad-Cre(E3) in the viral form. We used this virus to test side-by-side to the previously used adenovirus, Ad-Cre(E1) in Qiao's 2002 paper where she and her colleagues made the XX2-GFP-145 producer cell line. We infected the virus at approximately 5 multiplicity of infection per cell (MOI/cell) into the XX2-GFP-145 cell line and waited for cytopathic effect (CPE) to occur at about 48 hours post infection. The resulting colysate was heat treated to inactivate infectivity of lingering adenovirus, and confirmation of adenovirus inactivation was done with parallel negative controls of Ad-CMV-mVenus that showed no expression of mVenus after the heat treatment (data not shown). The now heat treated colysate was infected into HEK293 where green fluorescent cells

were counted to measure infectious titer resulting from these two viruses. The result of this infection assay is shown in Figure 10B. Whether the Cre recombinase gene is in the Cre(E3) or Cre(E1) region of the adenovirus makes no significant difference (p-value>0.95 by t-test).

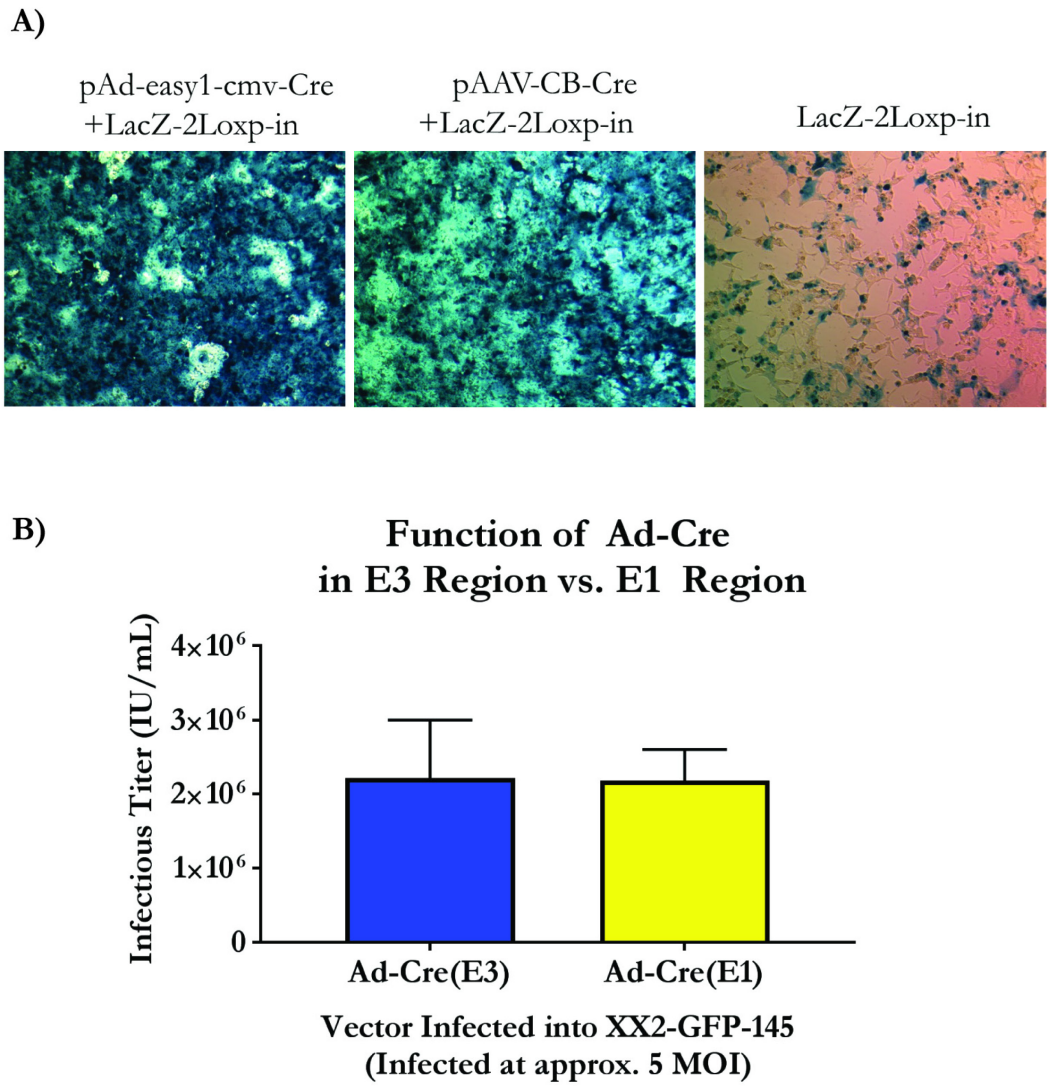
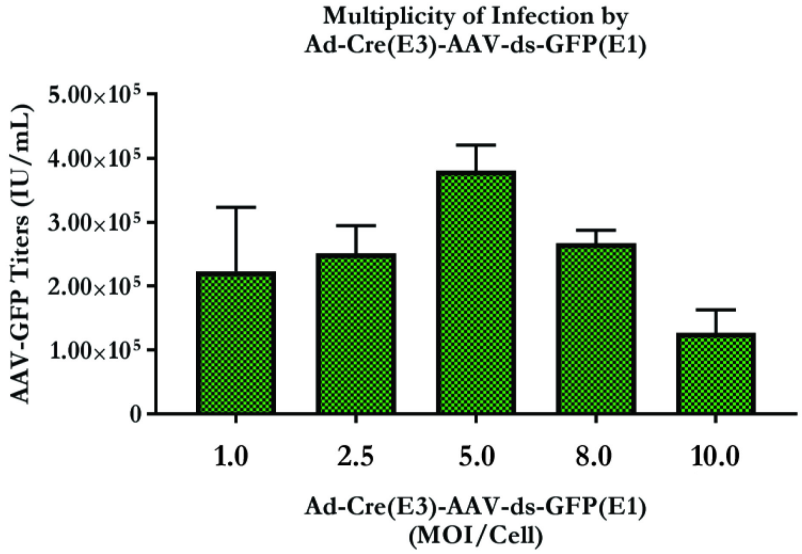


Figure 10: Cre(E3) function in plasmid and viral form. A) Plasmid forms of Cre in the E3 region of an AdEasy plasmid were transfected to observe the ability to recognize LoxP sites in a LacZ cassette whose gene expression has been controlled by the dual splice switch. B) Viral forms of Cre in the E3 and Cre in the E1 region of the adenovirus were infected into the XX2-GFP-145 producer cell line and resulting GFP expression was quantified.

After creating the adenoviral vector harboring the rAAV vector in the E1 region, via Gateway recombination as described in the methods section, characterization of the adenovirus construct was to be performed to observe if the rAAV is recognized by Rep/Cap proteins, and if this system can generate high titers of rAAV in an adequate packaging cell line. First, we needed to identify the appropriate multiplicity of infection per cell in a packaging cell line. At the time of generation of Ad-Cre(E3)-AAV-ds-GFP(E1), we did not have a sufficient packaging cell line to use and therefore relied on XX2-GFP-145 to generate data for the MOI/cell that would give the best yield AAV-GFP. Furthermore, when the packaging cell line was developed, the supply of Ad-Cre(E3)-AAV-ds-GFP(E1) had diminished substantially and a decision was made to use the supply and the time for more important large-scale experiments instead of replicating MOI/cell assays. In Figure 11A, we see the results of the MOI experiment and from this data we started using an MOI/cell of 5. At the time of generation of Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1), we had generated a sufficient packaging cell line (details described in more detail in Chapter 3), and used the cell line XX2-in-Puro-AAVS1 Clone 152.74 to decide the optimal MOI/cell to use. In Figure 11B, Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) showed an increase in infectious titers as MOI/cell increased. However, to not expend all supply of the Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) virus in stock, we chose to use an MOI of 10 for all experiments.

A)



B)

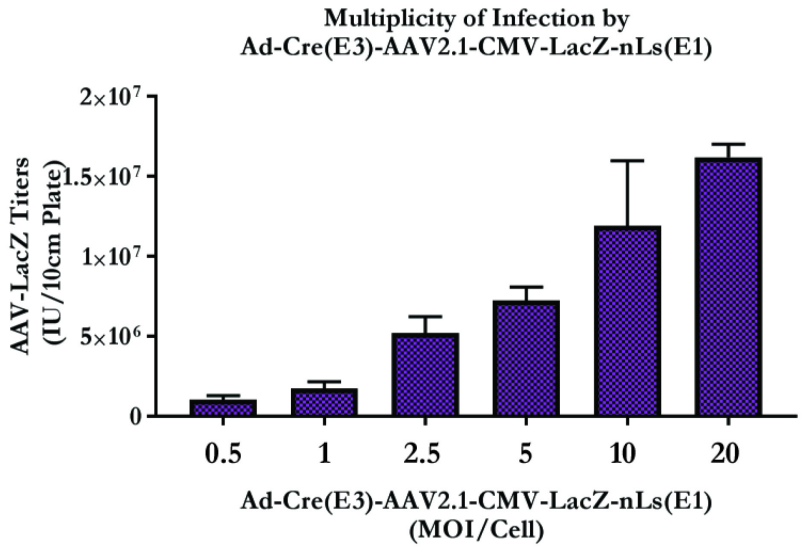


Figure 11: Determining optimal multiplicity of infection per cell for each Ad construct. A) This is the test of Ad-Cre(E3)-AAV-ds-GFP in XX2-GFP-145, since there was no available packaging cell lines at the time of this assay. B) MOI/cell assay for Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) in XX2-in-Puro-AAVS1 Clone 152.74.

To further characterize our virus, we needed to know whether or not the rAAV vector, more specifically the inverted terminal repeats (ITRs), was being recognized by the restored Rep and Cap proteins for adequate rAAV replication and encapsidation. To address this inquiry, we had to use another experimental design because of the lack of an adequate packaging cell line at the time this experiment was performed. The method that was divided into two sets of experiments: analysis of rAAV rescue from the plasmid adenoviral construct and analysis of the rAAV rescue from the adenovirus.

An equimolar quantity of necessary rAAV components in DNA plasmid form was transfected into HEK293 cells, with the controls of triple plasmid transfection generating the same rAAV vector found in the adenovirus designs. The purpose of this experiment was to provide the cells an equal amount of rAAV components and observe if titers were lower than triple plasmid transfection controls. If the titers were lower than the controls, this suggests there is an inherent issue with rAAV in the adenovirus plasmid. The results of this experiment have been placed on Figure 12A. Although the titer from qPCR for adenoviral constructs are higher than the triple plasmid control, this does not mean production titers of the novel adenovirus system supersedes the traditional transfection methods. Instead, this is more likely a property of transfection efficiency per cell between the double plasmid transfection for the adenoviral sets and the triple plasmid transfection sets. Nonetheless, the results of the experiment suggest that in the plasmid construct of pAd-Cre(E3)-rAAV(E1), the ITRs of the rAAV are being sufficiently recognized, processed by Rep/Cap, and is able to package into the final rAAV virus.

After production of the Ad-Cre(E3)-rAAV(E1), we needed to test if the vector was unperturbed during the production of the adenovirus and still maintains the function found in the plasmid analysis. Figures 12B and Figure 13 show these in various arrangements based on available plasmids we had at cell culture grade. First, in Figure 12B we observed the titers after

cells were infected with the Ad-Cre(E3)-AAV2.1-CMV-LacZ(E1) with the remaining rAAV component to generate the AAV2-CMV-LacZ product. The control used was triple plasmid control. The pXX2-SseI-2LoxP-Puro plasmid that is used in this experiment is the controlled Rep gene by the dual splicing switch and also provides AAV2 capsid. This is the pre-cell line method to see if Cre is able to function to remove the LoxP sites and initiate the dual splice switch system. As controls, we provided undisrupted AAV2 Rep/Cap which would identify the AAV-CMV-LacZ in the adenovirus and package into AAV2 capsid particles. As can be shown in 12B, triple plasmid transfection still functioned far greater than the virus system, but still this does not give a picture of the full capabilities of the adenovirus construct. This experiment also shows that the Cre-recombinase and ITR recognition of the rAAV are functional in the adenovirus.

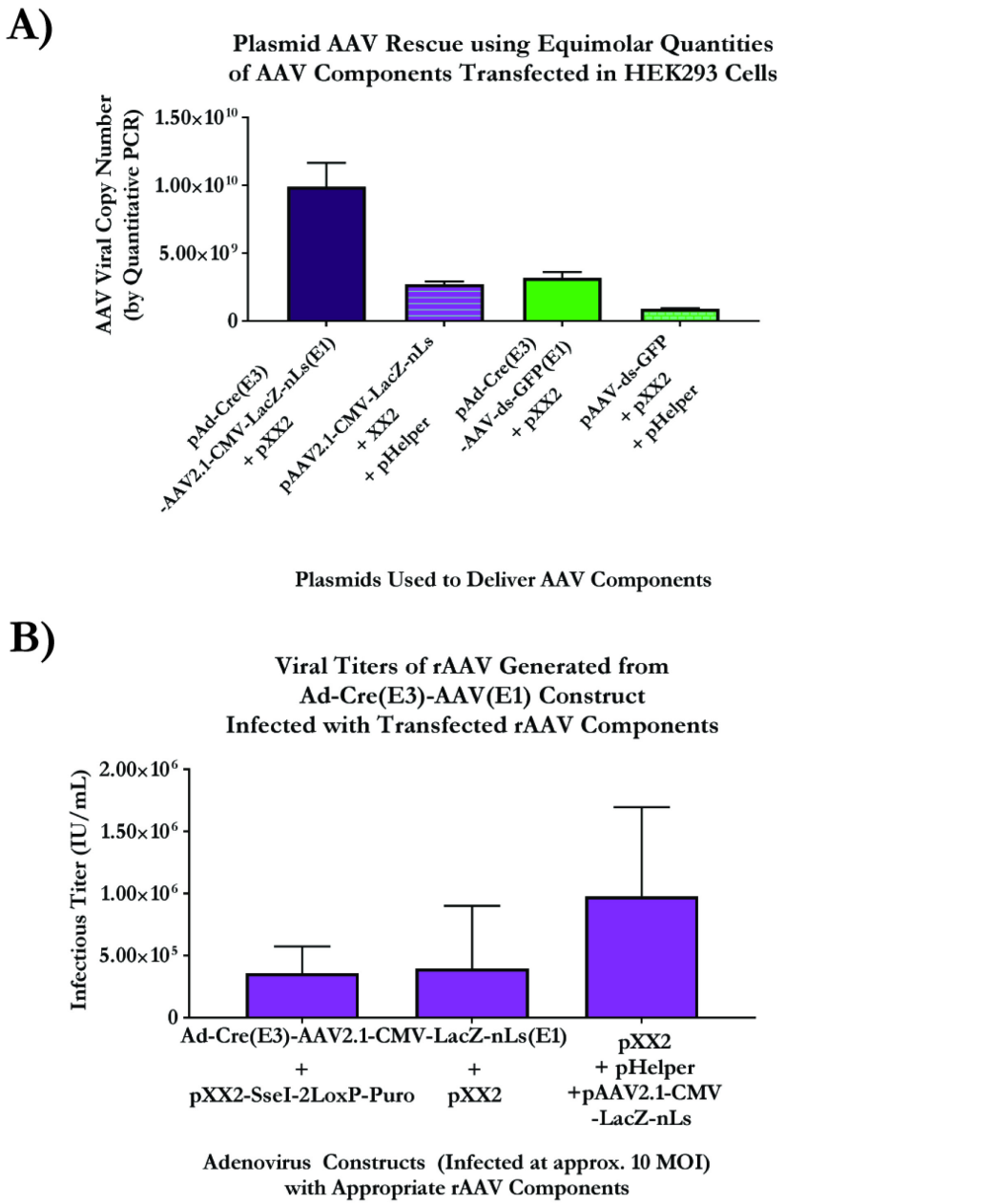


Figure 12: Rescue ability of rAAV vector in plasmid and viral forms. A) Use of equimolar rAAV components supplied to HEK293 for ability to recognize and package rAAV vector in various constructs of adenovirus, with triple plasmid transfection as control. B) Viral function of both Cre recombinase on Rep controlled plasmid, pXX2-SseI-2LoxP-Puro and rescue of rAAV vector to generate AAV2-CMV-LacZ-nLs product.

Figures 13 shows the ability to rescue the rAAV vector in both the Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1), used earlier in Figure 12, and Ad-Cre(E3)-AAV-ds-GFP(E1). In 13A, we observe the AAV-ds-GFP production from Ad-Cre(E3)-AAV-ds-GFP when using pSPGHH67.2m, a packaging plasmid with the Rep controlled dual splicing switch produces the HH67.2m serotype AAV. We see that the ability to rescue is comparable to triple plasmid transfection method. In another set of experiments, we used Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) transfected with the Rep dual splice switch that also includes the rAAV vector AAV-ds-GFP. As a control to this, we used Ad-Cre(E3) and this same Rep controlled producer plasmid. Of course, the two products that should arise from the Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) with the transfected Rep controlled producer plasmid are AAVHH67.2m-CMV-LacZ-nLs(E1) and AAVHH67.2m-ds-GFP. Interestingly, if we compare the titers in Figure 13A and Figure 13B for the products as a result of this virus with the producer plasmid, there is more AAV-ds-GFP compared to AAV-CMV-LacZ-nLs. This is possibly because of the copy numbers of the AAV-ds-GFP vector provided by the producer plasmid are significantly higher than the copy numbers of the AAV-CMV-LacZ-nLs vector provided by the adenovirus. Furthermore, after Rep splicing of the ITRs from the rAAV vector, it is possible that the adenovirus that has been spliced can no longer replicate, thereby perturbing the influx of supply of the vector AAV-CMV-LacZ-nLs. In Figure 13B, we compared the delayed Rep response from the Rep controlled dual splice switch to Rep that is not delayed in expression upon transfection. These results show that the delayed response of Rep increases infectious titers to nearly comparable levels of triple plasmid transfection controls. This data fits with early rAAV production research performed by Juan Li, et. al. in 1997[14]. Using their work as an explanation for the observations seen in Figure 13B, the overexpression of Rep in pXXHH67.2m is leading to substantially lower rAAV yields in the presence of adenovirus, possibly because this overexpression is also reducing

DNA replication and inhibiting capsid gene expression. The lower levels of Rep proteins occurring in the delayed Rep gene expression could support normal DNA replication and enhance Cap gene expression, and therefore increasing the yield significantly. By the time the Rep gene expression increases, cells would have already had an optimal frame of time for rAAV capsid production and generation of vectors, and at this stage the cells infected with the adenovirus would already undergo the cytopathic effect, CPE, thereby terminating cell viability. The balance between the right amount of Rep before the cells start to enter CPE is a necessary investigation for ideal rAAV vector production.

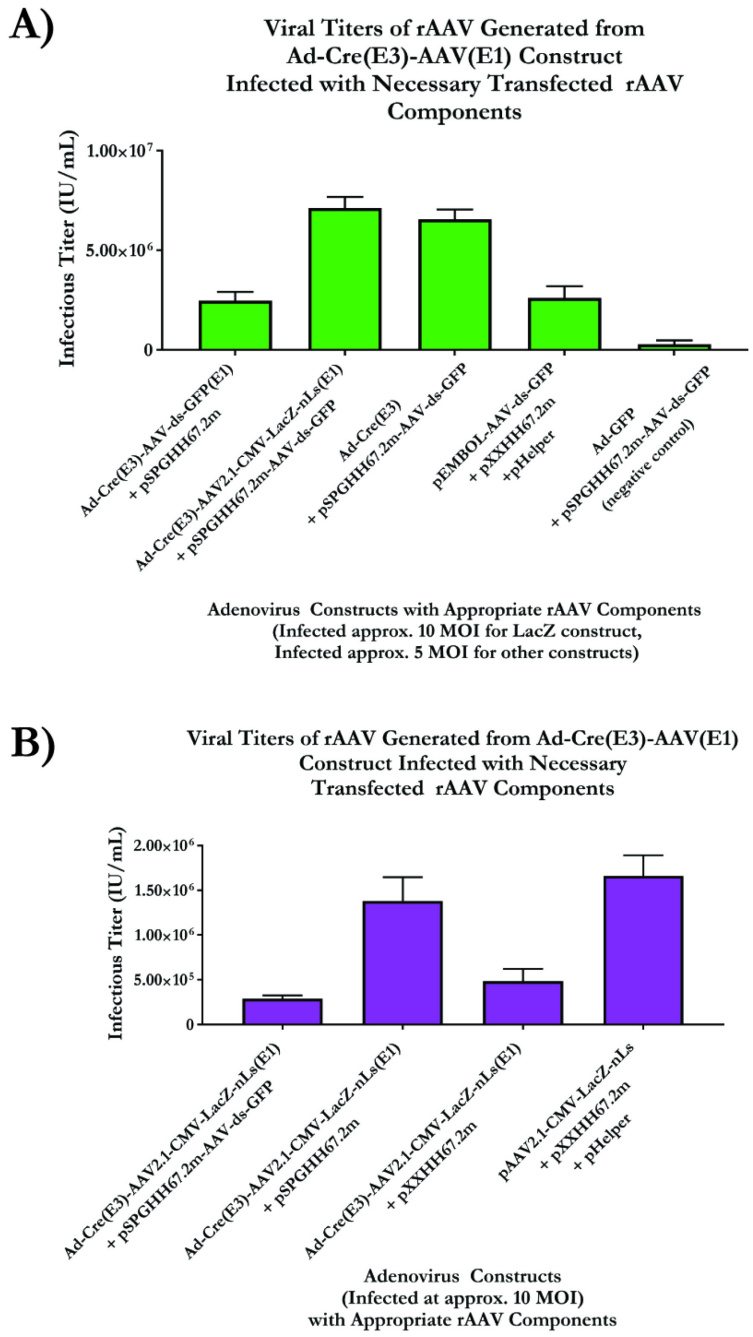


Figure 13: Rescue ability of virus using intron disrupted Rep/Cap. A) Infectious assay of AAV-ds-GFP using Rep disrupted HH67.2m packaging plasmid, with controls of Ad-Cre(E3) in Rep disrupted HH67.2m AAV-ds-GFP and triple plasmid transfection for AAV-ds-GFP. B) Infectious assay of AAV-CMV-LacZ-nLs using Rep disrupted HH67.2m packaging plasmid, with controls of undisrupted HH67.2m packaging plasmid and triple plasmid transfect for AAV-CMV-LacZ-nLs.

The outcome of a competitive production of rAAV vector was an interesting observation that was seen in Figure 13. However, we wanted to characterize the adenovirus construct in a more realistic situation before an actual packaging cell line was generated and attempted to do so using an established producer cell line, XX2-GFP-145. The purpose of this experiment was to see if an already integrated form of Rep/Cap can recognize and package adenoviral provided rAAV. Figure 14 shows these results. As a control Ad-Cre(E3) and Ad-Cre(E1) were used, to generate AAV2-ss-GFP exclusively. The competitive production of rAAV resulted in a nearly 50-50 distribution of AAV-CMV-LacZ-nLs and AAV-ss-GFP from the Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) infection into the XX2-GFP-145 cell line. All together these characterization studies have shown that the rAAV vector in the adenovirus is recognizable and able to generate a rAAV product. However, the full potential of the adenoviral vector could not be verified due to the lack of a packaging cell line prior to the work done in Aim 3. The use of a producer cell line for generation of a rAAV would have been unacceptable for determining the fullest potential of the viral vector because of the competitive production between the AAV vector within the producer cell line and the AAV vector within the adenovirus.

It was only after the methods discussed in Chapter 3, that we could have observed the fullest potential of the adenovirus construct. We have generated packaging cell lines called XX2-in-Puro-AAVS1 Clone 152.69 and XX2-in-Puro-AAVS1 Clone 152.74 that contain a high copy number of inducible Rep/Cap integrated into the genome of the HEK293 cell. Both of these cell lines are subclones of the XX2-in-Puro-AAVS1 Clone 152 cell line generated from integration of the pXX2-SseI-in-2LoxP-Puro integrated with intended site-specificity for the AAVS1 integration site using AAVS1 guided CRISPR/Cas9 genome editing. As a control, we infected these adenovirus constructs in cell lines that were generated using previous methods, i.e.

in HEK293 that had randomly integrated the packaging plasmid. The best of these type of cell lines was XX2-in-Puro Clone 253. As another control for the AAV-GFP products generated from the Ad-Cre(E3)-AAV-ds-GFP(E1) construct, we infected Ad-Cre(E1) in XX2-GFP-145 to compare yields of an AAV-GFP from the new strategy with the strategy developed by Qiao et. al. in 2002. After scaling these cell lines to 20 15-cm plates (equivalent cell number of about 600 million cells), we infected with the adenovirus constructs and compared them to triple plasmid transfection controls. The results for AAV2-CMV-GFP production from various methods is presented in Table 6 and the results for AAV2-CMV-LacZ-nLs production from various methods is presented in Table 7.

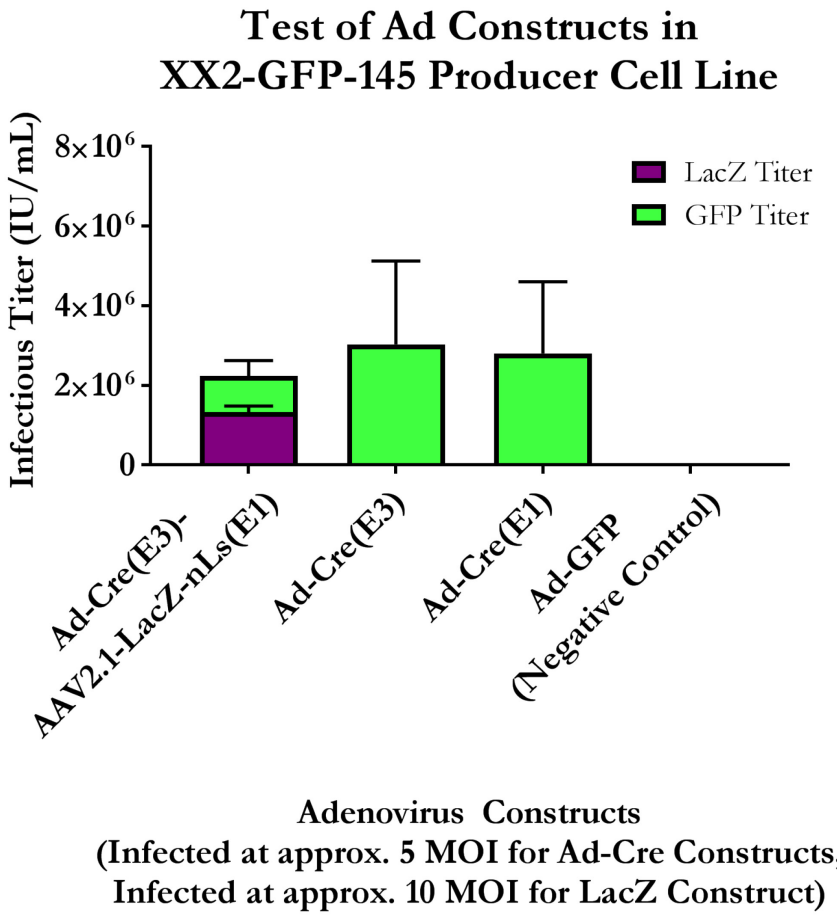


Figure 14: Function of Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) in XX2-GFP-145 Producer Cell Line

Table 6: Yield of AAV2-CMV-GFP from Various Production Methods

Introduced Vector	Cell Line	Clone	Serotype	rAAV Vector Produced	Purified vector (vg/mL) from 20x15-cm plates	Infectious Unit Titer (IU/mL)	DNA Containing Particles to Infectious Units Ratio
Ad-Cre(E3)-AAV-ds-GFP(E1)	XX2-in-Puro-AAVS1	152.69	AAV2	AAV-CMV-ds-GFP	1.89E+12	1.54E+09	1.23E+03
Ad-Cre(E3)-AAV-ds-GFP(E1)	XX2-in-Puro-AAVS1	152.74	AAV2	AAV-CMV-ds-GFP	1.75E+12	1.35E+09	1.29E+03
Ad-Cre(E3)-AAV-ds-GFP(E1)	XX2-in-Puro	253	AAV2	AAV-CMV-ds-GFP	1.54E+11	8.26E+07	1.87E+03
Ad-Cre-(E1)	XX2-GFP	145	AAV2	AAV-CMV-ss-GFP	8.20E+11	4.36E+08	1.88E+03
Triple Plasmid Transfection (control)	HEK293	N/A	AAV2	AAV-CMV-ds-GFP	3.71E+12	3.83E+09	9.70E+02

Table 7: Yield of AAV2-CMV-LacZ-nLs from Various Production Methods

Introduced Vector	Cell Line	Clone	Serotype	rAAV Vector Produced	Purified vector (vg/mL) from 20x15-cm plates	Infectious Unit Titer (IU/mL)	DNA Containing Particles to Infectious Units Ratio
Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1)	XX2-in-Puro-AAVS1	152.69	AAV2	AAV-CMV-LacZ-nLs	1.33E+12	4.46E+08	2.99E+03
Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1)	XX2-in-Puro-AAVS1	152.74	AAV2	AAV-CMV-LacZ-nLs	1.87E+12	4.24E+08	4.42E+03
Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1)	XX2-in-Puro	253	AAV2	AAV-CMV-LacZ-nLs	2.80E+11	1.34E+08	2.09E+03
Triple Plasmid Transfection (control)	HEK293	N/A	AAV2	AAV-CMV-LacZ-nLs	2.78E+12	7.59E+08	3.66E+03

The product from the larger scale can give us a multitude of information. First, as seen in vector genome values obtained from DNA dot blot methods, we see that the packaging cell lines developed in Chapter 3 can provide sufficient Rep/Cap copy numbers to generate a high quantity of virus when infected with the novel adenovirus construct. In the GFP products, the Ad-Cre(E3)-AAV-ds-GFP(E1) into the XX2-in-Puro-AAVS1 packaging cell line produced $1.8x10^{12}$ vg/mL, whereas the previous method of Ad-Cre(E1) into the AAV-CMV-ss-GFP producer cell line XX2-GFP-145 produced nearly half, at $8.2x10^{11}$ vg/mL. Triple plasmid transfection, producing AAV-CMV-ds-GFP, yielded $3.7x10^{12}$ vg/mL, about twice the yield of our new method. Next, if we infect the product into HEK293 cells to measure the infectious titer, we can then formulate whether or not this product is infectious. The infectious titers for Ad-Cre(E3)-AAV-ds-GFP(E1) in XX2-in-Puro-AAVS1 was $1.5x10^{9}$ IU/mL, $8.3x10^{7}$ IU/mL for the same adenovirus construct in XX2-in-Puro, $4.4x10^{8}$ IU/mL for Ad-Cre(E1) into XX2-GFP-145 producer cell line, and $3.83x10^{9}$ IU/mL for triple plasmid transfection product. We can take this information and generate a ratio of the vector genomes determined from dot blot over the infectious titer from the infection assay, to give a sense of the quality of virus that is produced. Basically, the smaller the ratio, the higher quality the vector is, or in other words how many vector genomes are required to produce one infectious unit. The conclusion from these ratios is that the AAV2-ds-GFP product formed from the Ad-Cre(E3)-AAV-ds-GFP(E1) into XX2-in-Puro-AAVS1 packaging cell line are highly infectious.

If we look at the capability of the Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) construct in a XX2-in-Puro-AAVS1 packaging cell line, with the XX2-in-Puro packaging cell line and triple plasmid transfection as controls, we can see that the resulting AAV2-CMV-LacZ-nLs product follows similar trends to the AAV2-ds-GFP product discussed previously: highly infectious, high

yielding product. Here, we were able to generate $1.8x10^{12}$ vg/mL vectors in the packaging cell line system compared to the $2.8x10^{12}$ vg/mL vectors in triple plasmid transfection. The DNA containing vectors to infectious ratios suggest that the products generated from the packaging cell line system are high quality.

The quality of the vectors can also be evidenced by the percentage of empty particles versus full particles present in a preparation. One of the major issues in vector production is the presence of empty particles in the final preparation that can induce a capsid specific T-cell response[15]. Ultracentrifugation can be a sufficient method to remove empty particles from preparations, but a good indicator of an efficient production system is the percentage of empty particles in a production batch. For this, we produced three batches of AAV virus purified through heparin sulfate columns. Two of these batches are Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) into XX2-in-Puro-AAVS1 packaging cell lines and the third batch was from triple plasmid transfection. The yield of the vector preparations through the heparin sulfate column are indicated in Figure 15.

The heparin sulfate chromatography method is used because of the affinity of AAV2 capsid to heparin sulfate. In theory, both empty and full particles would have affinity towards the heparin sulfate in these columns and both would be eluted out in moderate salt concentrations in the chromatography process. This is supported in Figure 15, where most of the viral particles were eluted at 60% Buffer B, which was 600 mM of NaCl.

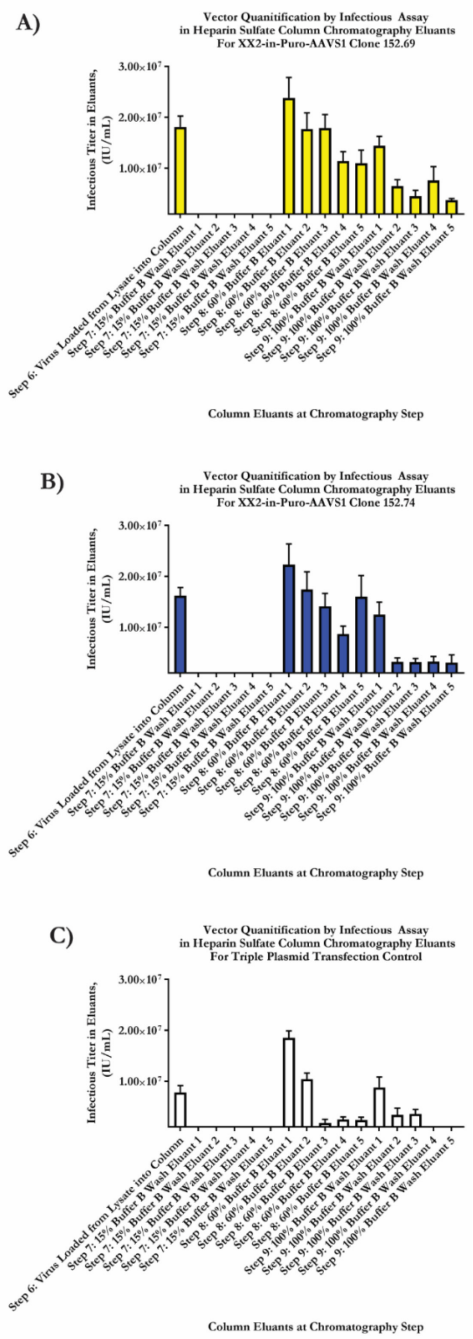


Figure 15: Vector quantification by infection assay of AAV2-CMV-LacZ-nLs in heparin sulfate chromatography eluants for three batches: A) from Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) infection into XX2-in-Puro-AAVS1 Clone 152.69, B) from Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) infection into XX2-in-Puro-AAVS1 Clone 152.74, and C) from triple plasmid transfection to generate AAV2-CMV-LacZ-nLs.

Taking aliquots of a few eluants obtained from Step 8 of the column chromatography process, we then adsorbed the viruses onto copper grids and the viruses were stained with 2% uranyl acetate, a negative staining solution. These copper grids, with viruses from a particular batch adsorbed onto it, were then taken to a transmission electron microscopy (TEM), where electron images of the viruses are generated. These images can be qualitatively used to analyze all particles in a copper grid, specifically quantifying full and empty particles from these batches to qualitatively analyze empty particle content.

It should be noted that a high quantity of virus or viral particles in the aliquot would mean more viral capsids that are adsorbed onto the copper grids. However, this does not mean that the electron images should be used to determine titer, since adsorption on the copper grid is dependent on a number of variables: the wettability of the copper grid, time of adsorption, quality of wash steps, salt concentration of initial aliquot, quality of the negative staining solution, and scanning under the transmission electron microscope for the ideal image frame in the copper mesh that contains an adequate gradient of particles. More often than intended, even a high-titer virus preparation can mean more than a single unique grid preparation to find a mesh in the copper grid with sufficient viral particles for imaging, especially since these images are taken from 100,000X or 125,000X magnification levels.

Here is a brief description on why this technique is used to analyze empty particles in AAV preparations. A water-soluble heavy metal-containing negative staining salt is used to surround and permeate within any aqueous compartment of a biological particle. After air-drying, a thin amorphous film of stain supports and embeds the biological material, all the while generating differential electron scattering between the relatively electron-transparent biological material and the electron-opaque negative stain[16]. In the case of AAV, if stain is able to enter the AAV capsid, that particular AAV particle has to be considered an empty particle. This is

displayed on an electron image as a dark grey to black colored dot (see Appendix B for exclusion criteria). In our datasets, we consistently identified viral particles with prominent dark gray or black dots as empty particles to reduce as much selection bias as possible. There is no automatic method available for quantifying, and therefore this a manual visual inspection of staining that is highly subjective and therefore varied between individuals. The criteria for exclusion was defined by knowledgeable staff members of the TEM facility and were consistently implemented during the analysis of these images. If the reader finds the exclusion criteria to be too strict or finds the data presented to be inconsistent with current literature, all high resolution TEM images for each batch have been provided in Appendix B for further investigation. A few examples of the TEM images that were taken are shown on Figure 16.

Analysis of the empty and full particle content, including total particles counted per batch is provided in Table 8.

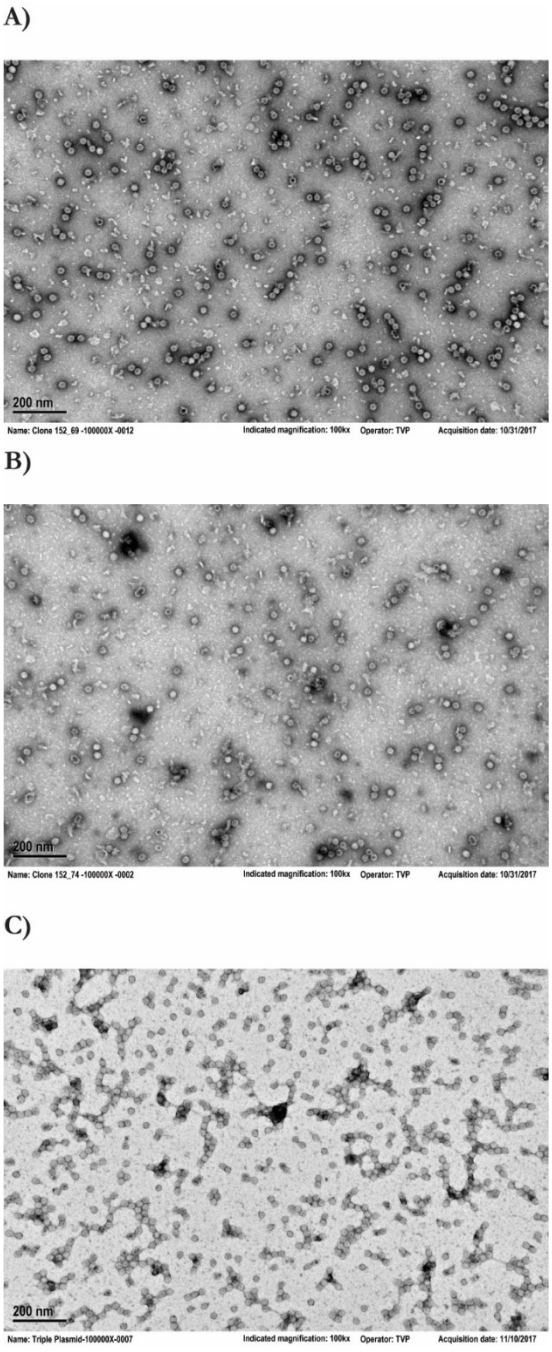


Figure 16: Sample transmission electron microscopy images from AAV2-CMV-LacZ-nLs preparations purified in heparin sulfate columns. A) from Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) infection into XX2-in-Puro-AAVS1 Clone 152.69, B) from Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) infection into XX2-in-Puro-AAVS1 Clone 152.74, and C) from triple plasmid transfection to generate AAV2-CMV-LacZ-nLs.

Table 8: Analysis of Full and Empty Particles from AAV2-CMV-LacZ-nLs Preparations

Introduced Vector	Cell Line	Clone	Serotype	Vector Produced	Counted Particles			Percentages		Full:Empty Ratio
					Empty	Full	Total	Empty	Full	
Triple Plasmid Transfection (Control)	HEK293	N/A	AAV2	AAV-CMV-LacZ-nLs	2244	12800	15044	14.92%	85.08%	5.7:1
Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs (E1)	XX2-in-Puro-AAVS1	152.69	AAV2	AAV-CMV-LacZ-nLs	814	4624	5438	14.97%	85.03%	5.7:1
Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs (E1)	XX2-in-Puro-AAVS1	152.74	AAV2	AAV-CMV-LacZ-nLs	600	3246	3846	15.60%	84.40%	5.4:1

As can be seen in Table 8, there does not seem to be a distinct difference between empty particle content in the TEM electron images analyzed. We luckily were able to find a high concentration of virus on a particular mesh of the copper grid, leading to the astounding 15,000+ total particle count for triple plasmid transfection preparations. Although this may not fit current understanding of empty particle generation from current production methods, our exclusion criteria were extremely consistent. The rigidity at which we performed these exclusions gave us less selection bias, and all particles that were analyzed as empty contained distinct dark-grey or black dots in the center of the viral capsid.

This analysis of empty particles shows that the product generated from these batches are high quality, where only 15% of the total particles analyzed are considered empty. With all characterization studies of the Ad-Cre(E3)-rAAV(E1) infected into a sufficient packaging cell line XX2-in-Puro-AAVS1, we show that the products that are generated are of a high-titer, high-quality, and highly infectious product.

2.5 Discussion

AAV vectors have been gaining a significant amount of popularity in gene therapy studies in the past decade for a variety of genetic diseases, and soon the gene editing revolution will find its need to deliver in vivo, with AAV vectors being the frontrunner as the vector of choice. Due to the high demand currently for AAV vectors in basic science and clinical applications, continued efforts are being made to improve the vector production method.

A depiction of the two systems, the previous system generated by our lab in 2011 and the updated system discussed in this chapter, is shown in Figure 17. We have previously made strides in developing a HEK293 producer cell line for high yielding AAV products, that avoids the use of triple plasmid transfection and also avoids use of wild-type adenovirus as a helper. However, a significant drawback to our previous methods is the need to create a novel HEK293 producer cell line for every rAAV vector of interest, as seen in Figure 17A. We also postulated a pitfall that occurs when generating the adenovirus. Our previous system used an engineered adenovirus that had potential for an additional 4-5 kb of any insertion. This extra space was enough to fit an rAAV vector. Therefore, we transitioned the Cre-recombinase to an alternate unused area, the E3 region, and placed a Gateway destination cassette in the E1 region. The Cre function was not lost when transitioning the Cre recombinase from the E1 region to the E3 region, and the titers of the adenovirus when scaling from plasmid form to the final viral form were not different to adenoviruses constructed only on the E1 region.

With the Cre unchanged between adenoviral constructs, the ability to change the E1 region in an efficient manner was necessary. Our novel design can create new adenoviral vectors with a rapid site-specific recombination of the transgene of interest without the need for endonuclease and ligase to insert the gene into the complex adenovirus shuttle plasmid or final adenovirus plasmid. We have created a method to swap for different rAAV vector cassette as

evidenced by the different ITR designs placed into the adenovirus. In this new design we can avoid the need for bacteria to do the recombination of AAV into the E1 region of Ad, and instead use a recombination reaction occurring on the bench. The reduction of time during generation of the adenoviral plasmid using the Gateway recombination into E1 can in turn save several days for a single construct. When compared to the previous method of generating a novel AAV producer cell line, the time and effort saved in shifting the rAAV from the producer cell line to the adenovirus construction is immensely reduced, and our use of an on-the-bench recombination can reproducibly create novel adenovirus products, further streamlining our generation of a novel adenovirus.

With the adenovirus now harboring the rAAV vector in the E1 region, through Gateway technology, we needed to make a packaging cell line that was sufficient to recognize the rAAV in the adenovirus and package rAAV product to a high quantity and quality, as shown in Figure 17B. This will be discussed further in Chapter 3. Our data before the generation of the packaging cell line shows that the rAAV vector is being recognized and packaged by Rep/Cap proteins of various serotypes of AAV. Our data also shows that if the Rep is interrupted with the inducible intron construction, this Rep can be restored with our adenovirus and the restored Rep and Cap can identify the rAAV vector and create a rAAV product. However, our designed systems at the time of characterization, namely the producer cell line XX2-GFP-145 and the packaging cell line XX2-in-19-sub, were unsatisfactory for the purposes of our adenovirus construct, either from a competitive production of the adenoviral rAAV with the already integrated rAAV in the XX2-GFP-145 producer cell line, or insufficient integrated copy numbers of inducible Rep/Cap in XX2-in-19-sub. We have shown in this method that we have the capability of producing a variety of rAAV vectors at high quality and quantity when infected into a predesigned packaging cell line capable of packaging into an AAV serotype of choice. We

show that with a cell line with sufficient inducible Rep/Cap copy numbers stably integrated into the HEK293 genome, we can yield higher quantities of the rAAV product when compared to a cell line that contains a smaller copy number of integrated inducible Rep/Cap. This indicates that the copy number of Rep/Cap integrated into the cell line is critical for a packaging cell line system described here. According to results described in greater detail in Chapter 3, the possible reason for this in our design is due to the reduction in adenovirus titers when Rep/Cap are activated and nick the AAV vector from the adenovirus, thereby disrupting the ability of adenovirus to replicate further in the HEK293 cell. It is possible that more MOI/cell of adenovirus is to be provided in a packaging cell line that contains less inducible Rep/Cap copy numbers per cell. However, the excessive adenovirus would sicken the cells at an accelerated rate and the time to produce rAAV vector would be stunted due to adenovirus-mediated apoptosis occurring earlier than the adenovirus-mediated S phase cell cycle transitions for rAAV to be replicated using functional cellular replication machinery. Therefore, an interesting exploration can be performed to supply the cells with sufficient, possibly excessive, copy numbers of the inducible packaging plasmid such that a high yielding product can be generated with less MOI/cell of the adenovirus.

According to the studies performed in this Chapter, there does not appear to be a difference in quality of vector generated from the packaging cell line and triple plasmid transfection control. Both systems generate a vector that is infectious and show relatively low amounts of empty particles. Of course, the analysis of the empty particles was subject to a manual counting of the electron images obtained in TEM, and are highly subjective. Our exclusion criteria were strictly followed, and although the values obtained were not necessarily those that are generally reported anecdotally or in literature, the results that we presented had little selection bias and had exact exclusion parameters. All told, the full potential of the novel

adenoviral construct was shown in our best packaging cell line generated in Chapter 3, and the quality we desired was met in our system.

A chief advantage of the HEK293-based AAV packaging cell line over other cell line is the use of replication defective adenovirus rather than the wild-type adenovirus. As mentioned in earlier chapters, this system also has the advantage of producing highly infectious product using a single vector entering a cell and not two to three vectors like those found in transfection or certain helper/subhelper production methods. When compared to baculovirus system, our system has no issue with recognizing splicing signals nor does it truncate the Rep or Cap genes in the dual splice switch, and generates all Rep and Cap proteins that would give more assurance to a better rAAV vector in a molecular scale. However, these replication deficient adenoviruses should be thoroughly inactivated and eliminated from the rAAV preparations in view of safety and immunity. Thermally inactivating the adenovirus is a common method used to inactivate the adenovirus. However, this does not completely inactivate the adenovirus. As discussed in Chapter 4, there are numerous methods that try to address this problem. In Chapter 4, we discuss a method using a selective precipitating agent to remove the adenovirus, but other systems are available to remove the virus such as nanofilters, high hydrostatic pressures, or scalable chromatography[17-19]. The combination of the Gateway-mediated generation of the adenovirus, the packaging cell line, and downstream purification systems to remove contaminating adenovirus should facilitate vector production for preclinical and clinical applications.

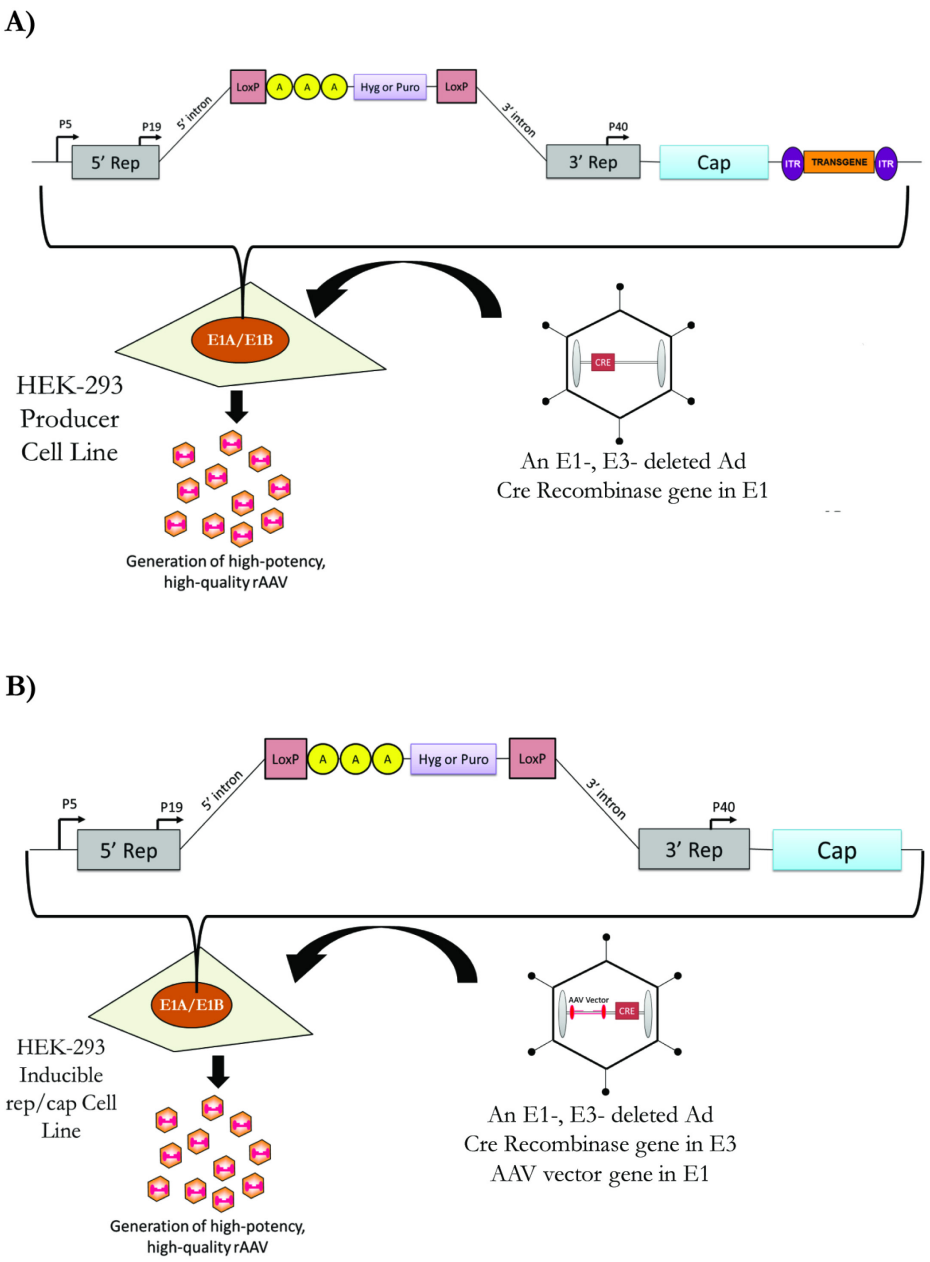


Figure 17: Depiction of the previous Ad-Cre/dual intron splice switch system (in A) compared to the new Ad-Cre(E3)-rAAV(E1)/dual intron splice switch system (in B).

2.6 Conclusions

With the close of this project came a technology that can help address the pressing concern of efficient manufacturing methods. The novel adenovirus design can allow for a variety of E1 deleted recombinant adenoviruses to be generated, using a very simple recombination technique. The packaging cell lines generated for the purposes of this chapter will be discussed further in the next chapter. They have played an instrumental role in the characterization of the adenovirus design presented in this chapter, but it is unique in its own right by using more frequent double-strand breaks, even with site-specificity to introduce the packaging plasmid. Hand-in-hand these two systems can generate the highly desirable clinical grade product that is necessary for rAAV gene therapy clinical trials.

REFERENCES

1 Gene Therapy Clinical Trials Worldwide, <http://www.abedia.com/wiley/> (2017).

2 Gearing, M. A Match Made in Heaven: CRISPR/Cas9 and AAV, <http://blog.addgene.org/a-match-made-in-heaven-crispr/cas9-and-aav> (2015).

3 Senís, E. et al. CRISPR/Cas9-mediated genome engineering: An adeno-associated viral (AAV) vector toolbox. Biotechnology Journal 9, 1402-1412, doi:10.1002/biot.201400046 (2014).

4 Zhang, H., Xie, J., Xie, Q., Wilson, J. M. & Gao, G. Adenovirus–Adeno-Associated Virus Hybrid for Large-Scale Recombinant Adeno-Associated Virus Production. Human Gene Therapy 20, 922-929 (2009).

5 Qiao, C., Wang, B., Zhu, X., Li, J. & Xiao, X. A novel gene expression control system and its use in stable, high-titer 293 cell-based adeno-associated virus packaging cell lines. Journal of virology 76, 13015-13027 (2002).

6 Wang, Z. et al. Rapid and highly efficient transduction by double-stranded adeno-associated virus vectors in vitro and in vivo. Gene therapy 10, 2105-2111 (2003).

7 Suzuki, Y. et al. A novel high-throughput (HTP) cloning strategy for site-directed designed chimeragenesis and mutation using the Gateway cloning system. Nucleic acids research 33, e109 (2005).

8 Yuan, Z., Qiao, C., Hu, P., Li, J. & Xiao, X. A versatile adeno-associated virus vector producer cell line method for scalable vector production of different serotypes. Human gene therapy 22, 613-624 (2011).

9 McConnell, M. J. & Imperiale, M. J. Biology of Adenovirus and Its Use as a Vector for Gene Therapy. Human Gene Therapy 15, 1022-1033 (2004).

10 Volpers, C. & Kochanek, S. Adenoviral vectors for gene transfer and therapy. The journal of gene medicine 6 Suppl 1, S164-171 (2004).

11 He, T.-C. in Current Protocols in Human Genetics (John Wiley & Sons, Inc., 2001).

12 Luo, J. et al. A protocol for rapid generation of recombinant adenoviruses using the AdEasy system. Nature Protocols 2, 1236-1247 (2007).

13 Sanes, J. R., Rubenstein, J. L. & Nicolas, J. F. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. The EMBO Journal 5, 3133-3142 (1986).

14 Li, J., Samulski, R. J. & Xiao, X. Role for highly regulated rep gene expression in adeno-associated virus vector production. Journal of Virology 71, 5236-5243 (1997).

15 Qu, W., Wang, M., Wu, Y. & Xu, R. Scalable downstream strategies for purification of recombinant adeno- associated virus vectors in light of the properties. Current pharmaceutical biotechnology 16, 684-695 (2015).

16 Harris, J. R. in Electron Microscopy: Methods and Protocols (ed John Kuo) 107-142 (Humana Press, 2007).

17 Leonard, J. N., Ferstl, P., Delgado, A. & Schaffer, D. V. Enhanced preparation of adeno-associated viral vectors by using high hydrostatic pressure to selectively inactivate helper adenovirus. Biotechnology and Bioengineering 97, 1170-1179, doi:10.1002/bit.21355 (2007).

18 Bogedain, C., Maass, G. & Hörer, M. (Google Patents, 2002).

19 Thorne, B. A., Takeya, R. K. & Peluso, R. W. Manufacturing recombinant adeno-associated viral vectors from producer cell clones. Human gene therapy 20, 707-714 (2009).

CHAPTER 3: GENERATION OF AN AAV UNIVERSAL PACKAGING CELL LINE WITH SITE-SPECIFIC INTEGRATION USING CRISPR/CAS9 AT THE AAVS1 INTEGRATION SITE

3.1 Overview

Recombinant adeno-associated virus (rAAV) producer or packaging cell lines using adenovirus(es) are created by linearized plasmids transfected into a particular cell line, such as HEK293 or HeLa cells. For production systems created previously by our group, we use a single recombinant adenovirus infected into a producer cell line for generation of high-yielding rAAV product. Although the adenovirus system has been modified, the generation of the cell line has not changed much, relying on the inherently low-frequency and in turn low efficiency random integration of the linearized plasmid into the host cell genome. Here, we present work that has been done to increase the frequency and efficiency integration of a linearized packaging plasmid using a site-specific endonuclease, CRISPR/Cas9, designed to target the AAVS1 locus, a natural integration locus for wild type AAV. The integration frequency of the CRISPR/Cas9 is apparent by the increased number in single cell colonies formed compared to randomly integrated methods, but is more pronounced when comparing the average infectious unit yields at a small scale, as evident by comparing Rep/Cap copy numbers of the best yielding packaging cell lines for both cell line generation methods. The resulting packaging cell lines using the AAVS1 guide RNA (gRNA), are capable of stably harboring the inducible designed dual-splicing switch, grow at the same viability and rate as parental HEK293 cells, and show greater than 9 times more integrated Rep/Cap copy numbers per cell than a randomly integrated control.

3.2 Introduction

Targeted gene disruption techniques have advanced of late, particularly with homing meganucleases, zinc finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs), and most recently the clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated gene 9 (Cas9) systems[1]. ZFNs, TALENs, and meganucleases can be designed to cleave genomes at specific locations[1-3]. Unfortunately, these technologies remain difficult and time-consuming to design, develop, and empirically test in a cellular context due to engineering DNA binding domains for each target gene, and the fusion of a Fok1 restriction endonuclease to DNA binding motifs for ZFNs and TALENs[2]. Recent advancements in CRISPR/Cas9 technology have accelerated introduction of double strand breaks (DSBs) into the cellular genome, followed by insertion of genes of interest.

CRISPR/CRISPR-associated (Cas) nuclease system is a RNA-guided endonuclease system identified in bacteria and archaea as part of an adaptive immune system[4]. It employs CRISPR RNAs (called crRNAs or guide RNA, gRNA, or single guide RNA, sgRNA) and Cas nucleases to induce double strand breaks (DSBs) to complimentary sequences of exogenous viral and plasmid DNA in accordance with Watson-Crick complementary rules, briefly illustrated in Figure 18. Modifications of the Streptococcus pyogenes (S. pyogenes) Type II CRISPR/Cas9 system were made to allow genome-editing in mammalian cells. Among them, crRNAs were fused to a normally trans-encoded tracrRNA (called a chimeric guide RNA or gRNA, expressed by a U6 promoter) to direct a human codon-optimized Cas9 (hSpCas9) nuclease to cleave target DNA sequences matching the crRNA[3]. Plasmid constructs are now available to make the molecular cloning of this CRISPR/hSpCas9 simplified with only the need of ligating the gRNA into a guide sequence insertion site. Additionally, inclusion of another gRNA sequence in the arrayed spacer architecture of CRISPR loci, imitating the multiplex nature of bacteria and

archaea CRISPR loci, can allow simultaneous editing of several sites within the mammalian genome, including the human genome, with efficient cleavage [5]. The only prerequisite for the target DNA to be a CRISPR target site is a 3' NGG sequence (protospacer adjacent motif, PAM). A single mismatch within 13-14 bp of the 3'-terminal gRNA sequences (5' to PAM site) has the ability to abolish the nuclease activity, but reports have been made that even three to five base pair mismatches in the PAM-distal part of the gRNA, suggesting a high degree of specificity yet the ability to make off-target nuclease activity [6-8].

The gene sequence for the human genome is readily available and the particular target can be screened to identify for the required 3' NGG protospacer adjacent motif (PAM) used by hSpCas9 for target sequence recognition. To assist in this endeavor, several software and tools are available for design of the gRNA. The Zinc Finger Consortium ZiFiT Targeter software, for instance can be used to design the gRNA with the U6 promoter included in the design for quick insertion into a hSpCas9 plasmid donated by the Zhang lab[9,10].

Our previous methods for generation of an rAAV packaging cell line involved random integration of the linearized plasmid in the HEK293 genome, resulting in a particularly time-consuming process especially if the linearized plasmid is large (>10kb) like those used in the rAAV producer cell line by Yuan et. al.[11]. This system can take up to 4 months to develop a suitable cell line, and even then, the Rep/Cap copy numbers may not be adequate for use for the methods indicated in Chapter 2. There are several ideas that can improve on this low integration rate: increase the frequency of double strand breaks and target an area that AAV has a natural affinity towards. If this can become a reproducible site-specific region of integration of the packaging plasmid, then the generation of an adequate packaging cell line can be accelerated and better characterized, transitioned to suspension cultures used for scalable reactors, and be used in large-scale production process designs of rAAV. Engineered nucleases allow random or

designed genomic modification at precise loci through cellular double strand break (DSB) repair, by either homologous recombination (HR) or non-homologous end-joining (NHEJ). However, insertion of an exogenous linearized plasmid can be streamlined using host cellular repair machinery to ligate the linear plasmid into the newly formed DSB. Therefore, our approach is to utilize the natural NHEJ DSB repair to integrate our linearized plasmid at more frequently occurring DSBs courtesy of the engineered nuclease CRISPR/Cas9.

The RNA-guided endonuclease system, CRISPR/Cas9, will be used to induce a site-specific double strand break (DSB) in the HEK293 genome to insert the packaging plasmid. This technology is easy-to-use and can generate DSBs with great precision, and more frequency especially when considering any off-target effects that may arise using the guide RNA of choice. The decision as to where to generate the DSB was not a critical factor in the purpose of this adjustment because our main objectives were the reproducibility of plasmid insertion and an increased number of integrated packaging plasmid within the HEK293 genome. However, for an optimal location of an insertion we looked at the AAVS1 integration site, with a guide RNA designed by Mali et.al. who have donated their construct to Addgene and have shown in their previous work that this guide RNA design is able to create distinct double strand breaks at the AAVS1 locus[3]. AAVS1 is considered a safe locus to integrate transgenes because previous attempts at integration in this region show cell function or viability to be uncompromised. Furthermore, AAV has a natural tendency towards AAVS1, since the AAV Rep proteins target AAVS1 due to specific binding of the Rep binding sequence (RBS) present within AAVS1. Luo et. al. has identified yields of rAAV producer cell lines and have noticed that many of their medium and high yielding producer cell lines have integrated plasmid at the AAVS1 locus[12].

We hypothesize that if the inducible AAV packaging plasmid is integrated into the AAVS1 site of the HEK293, more Rep/Cap copy numbers can be integrated and that would

increase the yield of rAAV using an Ad-Cre(E3)-rAAV(E1) construct. We have shown these results in Chapter 2. This chapter is focused on why the yield of rAAV is increased in the novel adenovirus construct infected into the packaging cell line that is made using assisted double strand break creation versus randomly generated double strand breaks. To address these focus points, we will show the methods of producing the packaging cell line, reasoning behind using AAV2 serotype instead of more clinically relevant serotypes, measuring the Rep gene amplification over time, measuring the Rep gene amplification with different adenovirus constructs, and mention the growth characteristics of the cell line.

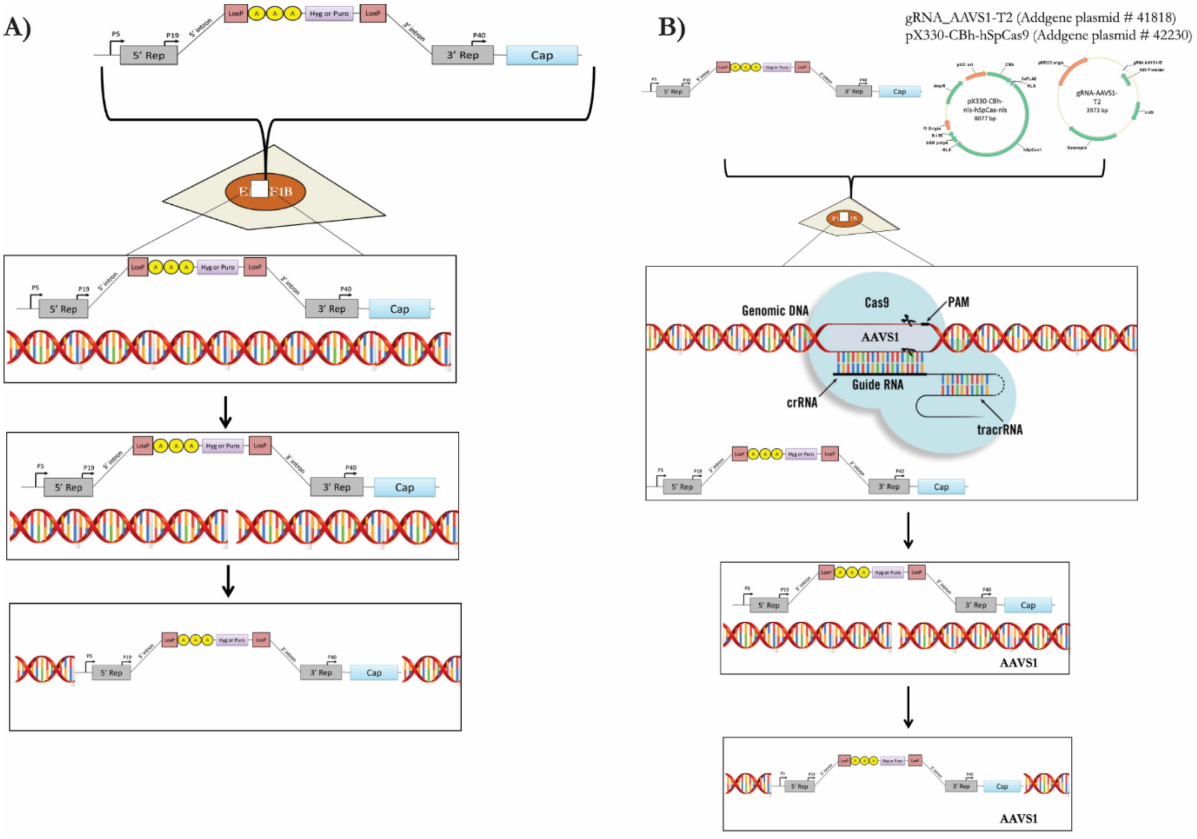


Figure 18: Depiction of two cell line production systems. A) This panel shows the randomly integrated packaging plasmid that has been used in previous methods of cell line development. B) This panel shows the hypothesized application of CRISPR/Cas9 at a specific site to introduce the linearized packaging plasmid.

3.3 Material and Methods

Program Design For gRNA/Integrated Plasmid Interaction

A concern that arose during analysis of appropriate CRISPR guide RNAs to use in this aim was the potential for an interaction to occur between Cas9 and the integrated plasmid. The species of Cas9 used in this aim was Streptococcus pyogenes (S. pyogenes), and specifically the Type II CRISPR/Cas9 that must recognize a 3' NGG protospacer adjacent motif, PAM, for the target DNA to be cleaved by Cas9. However, Cas9 can still be activated in the following scenarios, albeit at lower frequency or lower efficiency: a 3' NAG PAM (one-fifth of the binding efficiency compared to NGG) and three to five base pair mismatches in the PAM-distal part of the guide RNA sequence. If the mismatches are more frequent closer to the PAM sequence, Cas9 will likely not cause a double strand break[7,8]. To search through all plasmid of interest with these conditions (e.g. NGG, NAG, up to 5 mismatches, location of the mismatches, direct strand, reverse complementary strand) took far too much time, especially considering the number of integrated plasmids we were interested in using to generate cell lines (e.g. pXX2-SseI-2LoxP-in-Puro, pSPG9-AAV-ds-GFP, pSPGHH67.2m-AAV-ds-GFP, etc.).

With the knowledge in the field in mind for SpCas9, a Visual Basic (VBA) script was written in a program titled "FindcrRNAinIntroducedPlasmid," modified nine times to its latest version "FindcrRNAinIntroducedPlasmid_ver1.9." After inputting the guide RNA intended for use for genome editing using SpCas9, and entering the sequence desired to be integrated, the program is ready to analyze both the direct strand, the reverse complementary strand, for 3' NGG PAM, and for 3' NAG PAM, and output the location, sequence, number of mismatches, and location of mismatches of the guide RNA on the integrated sequence onto the Excel spreadsheet. The output can be used for sequences as big as the size limit of an Excel cell, 32767 characters and the guide RNA can be placed for any length of gRNA, though this gRNA

sequence is usually about 20-22 bp. Typically our constructs range from 10-14kb, and therefore there are no character size restrictions foreseen in our designs when using this program. Only A, T, C, G characters can be inputted into the cell and those that do not fit these inputs will be ignored in the analysis.

The program will prompt the following error messages: "Enter crRNA and Sequence of Interest in Cell B1 and Cell B2, respectively" when no inputs have been placed; "Enter Sequence of Interest in Cell B2" when no intended integrated plasmid sequence has been placed; and "Enter crRNA sequence in Cell B1" when no intended crRNA (i.e. guide RNA or gRNA) sequence has been placed. After inputs have been placed, the user would click on the "Find for Direct Sequence" to analyze in the direct strand or the "Find for Reverse Comp Sequence" to analyze in the reverse complementary strand. The following are prompts if no errors have been reported, and no matches have occurred: "No Exact or Partial Matches in the Direct Strand Were Found" and "No Exact or Partial Matches in the Reverse Complementary Strand Were Found." If a match were to occur, the program would jump to the section of the strand being analyzed (under "DIRECT STRAND RESULTS" or "REVERSE COMPLEMENTARY STRAND RESULTS") and then go to the closest 3' NRG the match occurred on underneath the subsection indicating whether the match occurred in an exact match or partial match. The program will then print all matches that have occurred for that gRNA and sequence of interest. The user can return to the top of the spreadsheet by clicking on "Return to Top." The user is also given instructions to help decipher the program output if the printed materials are unclear. After completion, the user can clear contents and continue to search for partial or exact matches in other sequence of interest for a particular gRNA.

The entirety of the code for this VBA script can be found in Appendix B.

Plasmid Construction For CRISPR/Cas9 System

There are currently three unique plasmids that are necessary in this Aim, two of which are for the CRISPR/Cas9 system and the other plasmid is an AAV2 packaging plasmid we would like to integrate into the genome. The CRISPR/Cas9 genes are called gRNA-AAVS1-T2 (Addgene plasmid # 41818) which is an AAVS1 guide sequence from Mali et al (2013) and pX330-CBh-hSpCas9 (Addgene plasmid # 42230), which is the human codon optimized S. pyogenes Cas9 modified to remove its guide sequence. The removal of the guide sequence found in the original pX330-U6-Chimeric_BB-CBh-hSpCas9 was done by AflIII/XbaI digestion and blunt end ligation, creating the pXX30-CBh-hSpCas9 used in the aim. The AAV packaging plasmid is pXX2-SseI-in-2LoxP-Puro, where it contains an AAV2 capsid gene and AAV2 Rep that is disrupted by an intron placed specifically to turn off the cytostatic and cytotoxic function of the Rep gene.

Initial attempts at the CRISPR/Cas9 system were made by generating a novel guide RNA sequence using the ZiFiT consortium gRNA design tool. It was initially postulated to target for a cell cycle inhibitor: p16 exon 2. Whether or not this guide RNA was functional was not investigated further due to the progressively lowering titers of the generated producer cell line, called ZY9-AAV-ds-GFP Clone 1.24. This was more pronounced in cell lines generated from co-transfection of AseI linearized pSPGHH67.2m-DEST using the AAVS1 guide sequence from Mali et. al. and modified pX330-CBh-hSpCas9 plasmids mentioned above. In the ZYHH67.2m-DEST cell lines, the cells would either grow for a short amount of time, or not at all, suggesting some leaky expression of the Rep gene. It was puzzling to us at the time why this was occurring, but we later sequenced the intron region of the plasmid using the following primers to elucidate the problem: 5' Rep 5’-GGGATTACCTCGGAGAAGCAGTGG-3’, HCG New 5’-GTAAGAAGATCGAGGTC-3’, 3' Rep 5’-ACCAGATCACCATCTTGTCG-3’, 3'

LoxP 5'-TCTTCCTAGTGGATCTGCGAC-3', and SV40Promoter 5'-CCATAGTCCCGCCCCTAACTCC-3'.

Selection of Cell Colonies and Viruses Used During Characterization

pXX2-SseI-2LoxP-in-Puro was digested using AseI. This plasmid was cleaned using ethanol precipitation. Two sets of cell lines were generated using the linearized pXX2-SseI-2LoxP-in-Puro: the AAVS1 CRISPR/Cas9 cell lines, labeled as XX2-in-Puro-AAVS1, or the randomly integrated cell lines, labeled as XX2-in-Puro. Briefly, the methods to generate the former cell lines, XX2-in-Puro-AAVS1, is half of the total mass of transfected plasmid to be the linearized pXX2-SseI-2LoxP-in-Puro, a quarter of the total mass to be the gRNA-AAVS1-T2, and the last quarter to be the pXX30-CBh-hSpCas9. The latter cell line, XX2-in-Puro, is the equivalent of the total mass of plasmid used for XX2-in-Puro-AAVS1 cell lines, but solely using linearized pXX2-SseI-2LoxP-in-Puro.

After these plasmids were cleaned and resuspended in about 20 μL of TE (pH 7.0) they were transfected into 3 6-wells (approx. 750,000 to 1,000,000 cells per well) using calcium phosphate methods. Media was changed 6-12 hours after transfection. Cells were scaled about 48 hours post transfection to up to 20 15-cm plates to thoroughly dilute the cells per plate. After about four days, the media in the plates were changed from DMEM, 10% FBS, to DMEM, 10% FBS, 2 μg/mL puromycin to leave only cells that can resist the puromycin drug. Media was changed every four days. Between days 12 and 16, when colonies were about 1-3 mm in size, single clone colonies were counted to tabulate total single clone colonies generated. A small subset of these colonies was selected according to the following protocol: 0.5 cm x 0.5 cm Whatman filter papers were cut and autoclaved along with tweezers. TrypLE Select (Gibco, 12563011) or TrypLE Express (Gibco, 12605010) was used to evenly dissociate cells without damaging the integrity of the cells. This recombinant form of trypsin was critical for proper

healthy growing of cells from single clone colonies captured by the Whatman filter paper to scaled cell lines used for characterization studies. Porcine trypsin was previously used in early trials of cell line generation, but the cells would clump and result in very poor growth characteristics that was difficult to troubleshoot. The Whatman filter papers were soaked in this TrypLE recombinant trypsin, and the soaked paper was placed on top of the cell colony for approximately 5 minutes before removing the paper and shaking the cells off into a 24 well plate prefilled with media not containing the puromycin drug. From there, cells were allowed to grow for 2-3 days, and the media was changed to puromycin containing media at the selection concentration of 2 μg/mL.

Cells were scaled from a 24 well plate, eventually to a 6 well size. When cells were scaled, passage entries were noted to determine if titer of rAAV reduces with increasing passages. At the 6-well scale, cells were tested for rAAV production capabilities using the novel adenovirus design in Chapter 2, Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1). They were infected at approximately 10 MOI to generate cytopathic effect within 48 hours. The cells were then harvested, subjected to three freeze/thaw cycles, spun to collect crude lysate, heat inactivated at 56 °C for 1 hour, and infected into HEK293 cells with the addition of 4 μM of Hoechst 33342 dye to accelerate AAV expression. Once a novel cell line was chosen, it was repeated several times at various passages to see if titers decreased. The cell colonies narrowed in choices from the hundreds selected, to a few dozen prominent candidates, to a single candidate.

Subcloning of the best candidate was an inevitable scenario because selection methods of the single clone colony often, but not always, resulted in two or more single cells growing very close to each other during colony growth that the Whatman filter paper caught these populations to some degree. The population selected had to be assumed to have a mixed population and the cells were seeded at a very dilute concentration of 0.5 cells/mL in a 96 well

plate. These subclones were documented and subjected to the same characterization assay and the best candidate(s) was selected from these subcloned colonies.

Cells and Viruses Used in Characterization

We used two cell lines developed by Qiao et. al. in 2002 called XX2-GFP-145, an AAV2-GFP producer cell line, and XX2-in-19-sub, an AAV2 parental packaging cell line. These were used primarily to assist in initial characterization studies of the Ad-Cre(E3)-rAAV(E1) virus, particularly for Ad-Cre(E3) and the Ad-Cre(E3)-AAV-ds-GFP(E1) constructs, that justified the need to generate a packaging cell line in this aim. XX2-GFP-145, in particular, was used to see the differences in yield at larger production scales from our previous producer cell line method to this modified cell line method. As mentioned before Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) was used extensively in this aim. Other viruses that were used were: Ad-Cre(E3)-AAV-ds-GFP(E1), Ad-GFP, wtAd, Ad-Cre(E3), and Ad-Cre(E1).

Infectious Titer Calculation

HEK293 cells were used for infectious titer assays. For LacZ expression, cells were quantified for infectious titer, by counting number of blue cells in the microscope or camera field under bright-field microscopy. For GFP, cells were quantified by counting number of green cells in the microscope or camera field under fluorescent microscope (excitation of GFP is 488 nm and emission is 509 nm). The yield of infectious particles was presented as infectious unit. The units of infectious units (IU) were either IU/mL or IU/10-cm plate. IU/10-cm plate was calculated by extrapolating from the surface area of the microscope field to the surface area of the well being used to for the infection assay. IU/mL was calculated by dividing the extrapolated well infectious unit by the total volume infected into the well. Larger scaled products were typically presented in IU/mL, whereas smaller scale characterization studies were presented in

IU/10-cm plate. Both measurements were consistently used to compare infectivity at a specific scale between various production methods.

AAV Vector Production and Large-Scale Purification

For the production of AAV2, three systems were used: Ad-Cre(E1) infection into XX2-GFP-145 producer cell lines, Ad-Cre(E3)-rAAV(E1) into XX2-in-Puro-AAVS1 cell lines, and triple plasmid calcium phosphate transfection (pXX2, pAAV2.1-CMV-LacZ-nLs or pEMBL-AAV-CMV-ds-GFP, pHelper). For large scale, these were performed on 20 15-cm plates, and on 6-well plates for smaller scale. The viruses, Ad-Cre(E3)-AAV-ds-CMV-GFP(E1) and Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1), were used for titers in both scales of packaging cell lines.

Briefly, the HEK293 cells or cell lines were propagated in Dulbecco's modified Eagle's medium (DMEM) in 15 cm plates supplemented with 10% fetal bovine serum (FBS) at 37°C with the appropriate drug selection, if applicable: puromycin (1 μg/mL) for cell lines with the puromycin resistant gene and hygromycin (100 μg/mL) for XX2-GFP-145 that has a hygromycin resistant gene. For triple plasmid transfection, cells were transfected at 80-90% confluency. Each plate was transfected with 16.7 μg vector plasmid, 16.7 μg adenovirus helper plasmid, and 16.7 μg AAV2 packaging plasmid dissolved in 40 mL of 0.25M $CaCl_2$. For every 2 mL of dissolved DNA in 0.25M $CaCl_2$, 2 mL of 2x HEPES-buffered saline (HBS buffer: 50 mM HEPES, 280 mM NaCl and 1.5 mM Na_2HPO_4; pH 7.1). In a smaller scale (i.e. 6-well), 0.67 μg of each plasmid was dissolved in 125 μL if $CaCl_2$ and quickly mixed with 125 μL of 2x HBS. Eight to twelve hours later, the medium was replaced with fresh DMEM supplemented with 2% FBS. Cells were harvested at 72 hours post transfection, and media was also collected. Infected cells were harvested at 48 hours post infection after achieving cytopathic effect (CPE).

In purification of large scale rAAV, cells were resuspended in suspension buffer I (50mM HEPES, 150 mM NaCl, 50 mM NaH_2PO_4, 2 mM $MgCl_2$, 2.5 mM KCl, pH 8.0) and sonicated. The cell lysate was treated with DNase (100 units/ml) and RNase A (4 units/ml) and incubated at 37°C for 1 hour. Debris was removed by centrifugation at 2,500 rpm at 4°C for 15 minutes. For preparations using adenovirus, a heat inactivation of 56 °C was performed for both the collected media and the cell lysate before moving onto the next steps.

PEG-8000 and NaCl solutions were added to the clarified lysate, to a final concentration of 8% PEG-8000 and 0.5 M NaCl, and incubated at 4°C overnight. For the culture medium, powdered forms of PEG-8000 and NaCl were added to also reach 8% PEG-8000 and 0.5 M NaCl and incubated at 4°C overnight. The cell lysate and medium were centrifuged at 2,500 rpm for 30 minutes and the resulting pellets from both cell lysate and medium were combined and thoroughly resuspended in about 20 mL of resuspension buffer #2 (50 mM HEPES, 150 mM NaCl, 1% Sarkosyl, 20 mM EDTA, pH 8.0). The solution was placed in an ultracentrifuge at 31,000 rpm for 16 hours in a CsCl density gradient. The AAV band was collected and subjected to a second round of CsCl density gradient ultracentrifugation at 38,000 rpm for 48 hours. The AAV band was collected in drop-wise fractions and stored at -80°C. Vector titers were determined by the DNA dot-blot method.

AAV titer was quantified by dot blot hybridization. Five microliters of AAV stock was added to 200 μL DMEM and treated with 50 μg/ml DNase I at 37°C for 1 hour to degrade unencapsidated DNA. Then, 200 μL proteinase K buffer (20 mM Tris Cl pH 8.0, 20 mM EDTA pH 8.0, 1% SDS) was added to inhibit DNase activity, followed by addition of 40 μg of proteinase K, and the sample was incubated at 55 °C for 1 hour to degrade the capsid. The vector DNA was precipitated using ethanol precipitation with addition of glycogen (40 μg) to visually observe the pelleted DNA. The resultant pellet was resuspended in alkaline buffer (0.4

M NaOH and 10 mM EDTA pH 8.0) and bound to a hybridization transfer membrane (PerkinElmer). A standard of the original vector plasmid was applied to the same membrane. A biotin-labeled probe against CMV, LacZ, or GFP was hybridized to the membrane at 55°C overnight. Bound probe was detected using the North2South Chemiluminescent Nucleic Acid Hybridization and Detection Kit (Pierce). AAV titers were in the range of 10^{11} to 10^{13} vector genomes (vg)/ml.

Histochemical Staining of Monolayer Tissue Culture Cells for LacZ Activity (X-gal Staining)

Cultured cells were rinsed with 1x Phosphate Buffered Saline (PBS, pH 7.3) and then fixed for >5 min at 4 °C in 2% formaldehyde and 0.2% glutaraldehyde, mixed in PBS. The cells were then overlaid with a histochemical reaction mixture containing 1 mg/mL 4-Cl-5-Br-3-indolyl-β-galactosidase (X-gal), 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, and 2 mM $MgCl_2$. The X-gal was dissolved in dimethoxysulfoxide (DMSO) at 40 mg/mL, and then diluted into the reaction mixture. Incubation was for 8-14 hours at 37 °C[13]. Cells were then counted for LacZ expression, i.e. blue cells were counted.

Growth Rate of Cell Lines Using CellTiter-Glo®

The CellTiter-Glo® 2.0 Luminescent Cell Viability Assay (Promega, Catalog Number: G9241[14]) was used to determine the growth rate of the packaging cell lines that were developed in this aim. This assay provides a homogeneous method to determine the number of viable cells in culture based on quantitation of ATP present, which indicates the presence of metabolically active cells. This assay results in cell lysis and generation of a luminescent signal proportional to the amount of ATP present. The amount of ATP is directly proportional to the number of cells present in culture.

Cells were seeded into CELLSTAR® 96-well white plates (Supplier Catalog Number 675083). To generate the standard curve, the following cell counts were placed into the 96-well

plate, with the final volume reaching 100 μL: 0 cells (only media), 50, 100, 200, 400, 1000, 2500, 5000, 7500, 12500, 25000, and 50000 cells in triplicates. An ATP standard, just as a control for function of CellTiter-Glo® luminescence, was provided with each plate that was tested using CellTiter-Glo® for concentrations of ATP ranging from $1x10^{-12}$ moles to $1x10^{-10}$ moles. Once this plate was seeded, luminescence versus cell count can be measured and a standard curve could be generated for later luminescent recordings. The standard curve was tolerable for the purposes of this growth experiment (R^2=0.989). In addition to this standard plate, 5 plates were used to measure cell growth: Plate 1 for Day 0 (0 hr.), Plate 2 for Day 1 (24 hr.), Plate 3 for Day 2 (48 hr.), Plate 4 for Day 3 (72 hr.), and Plate 5 for Day 4 (96 hr.). For each cell line, 2500 cells were plated in triplicate and ATP standards were performed on the same day as luminescence measurements.

First, on the day of luminescent testing the plate was equilibrated to room temperature for approximately 30 minutes. Next, a volume of CellTiter-Glo® Reagent equal to the volume of cell culture medium was added to each well. Contents were mixed at room temperature for 2 minutes using an orbital shaker to induce cell lysis. The plate was then incubated at room temperature for 10 minutes to stabilize luminescent signal and then the luminescence of the plate was recorded using a luminometer. Doubling time of the cells were determined using the following equation: $Doubling\ Time = ln(2)/[ln(N(t)/N(0))/t]$, where $N(t)$ is number of cells at time t, $N(0)$ is the number of cells at time 0, and time is t in hours.

Quantification of Rep Gene Copy Number by Using Real-Time Polymerase Chain Reaction

For the quantification of Rep gene copy number in stable 293-based cell lines, we used SYBR green–based real-time quantitative assay (ABI PRISM 7700 Sequence Detector, Applied Biosystems). We designed the primers to amplify a 317-bp fragment of the Rep gene. The sequences of the forward and reverse primer are: Rep-5': 5'-GGG ATT ACC TCG GAG AAG

CAG TGG-3'; and Rep-3': 5'-CTT CCC GGT AGT TGC AGG-3'. We also designed the primer to amplify a 300-bp fragment of the human glucagon gene as the internal cell copy number control. Sequences of the forward and reverse primers are as follows: human-glucagon-F: 5'-TGA GAG ACA TGC TGA AGG GAC-3'; human-glucagon-R: 5'-CTT TCA CCA GCC AAG CAA TG-3'.

Total cellular DNA was extracted from cells by using the DNeasy Tissue Kit (Qiagen). Copy numbers of the Rep gene detected by real-time polymerase chain reaction (PCR) were reported as Rep copies per cell by the following equation: (vector copy numbers/human glucagon gene internal control) x 2.

Two designs were made to determine the Rep gene amplification properties in each cell line. One design is to determine the time course of Rep gene amplification depending on two Ad constructs, Ad-Cre(E3) (MOI 5) and Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) (MOI 10), measured from 0 hrs. post-infection to 96 hrs. post-infection in the cell lines XX2-in-Puro Clone 253, XX2-in-Puro-AAVS1 Clone 152.69, and XX2-in-Puro-AAVS1 Clone 152.74. This was done in the scale of a 6-well plate, where cells were infected at about 80% confluency and these conditions were done in triplicates for statistical purposes. The second design is to determine the Rep gene amplification depending on a variety of Ad constructs, using the following Ad designs infected to achieve CPE at 48 hrs. post infection: Ad-GFP, wild-type Ad (wtAd), Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1), Ad-Cre(E3)-AAV-CMV-ds-GFP(E1), Ad-Cre(E3), and Ad-Cre(E1). The cell lines that were used were HEK293 (negative control), XX2-in-Puro Clone 253, XX2-in-Puro-AAVS1 Clone 152.69, and XX2-in-Puro-AAVS1 Clone 152.74. Cells were infected at about 80% confluency in a 6-well plate and these conditions were done in triplicates for statistical purposes.

3.4 Results

To show the need for a packaging cell line, Figure 19 shows what was available to us when we initially generated our adenovirus constructs from Chapter 2. The abysmal yield of rAAV vector, especially when considering the yields obtained through normal triple plasmid transfection gave us misdirection on the utility of this novel adenovirus design. It was later identified through the intense characterization studies in Chapter 2, that indeed the adenovirus construct was functional and that our existing packaging cell lines failed to provide enough Rep/Cap copy numbers for our application.

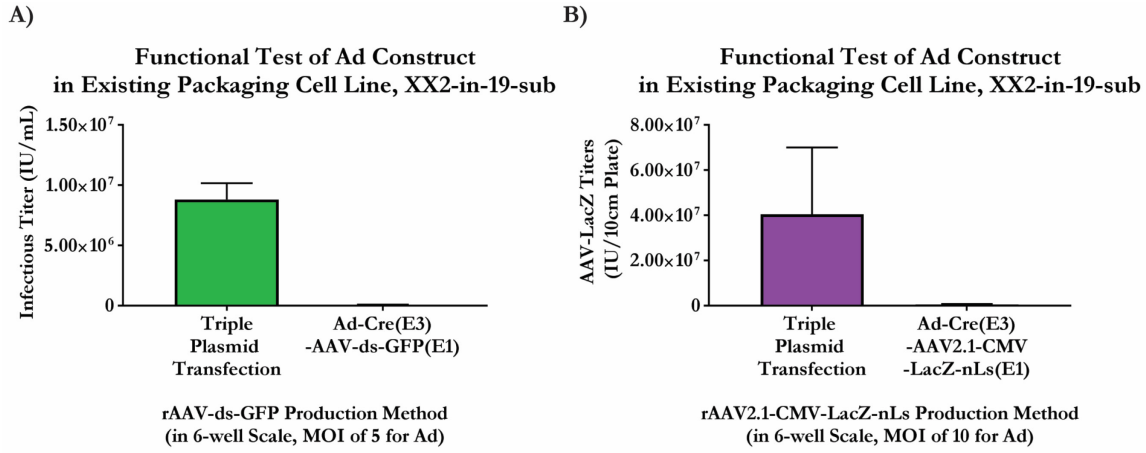


Figure 19: The need for a packaging cell line. A) Infection of Ad-Cre(E3)-AAV-ds-GFP(E1) into XX2-in-19-sub with a control of triple plasmid transfection generating AAV-ds-GFP. B) Infection of Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1)

According to Qiao's studies of this cell line in 2002, XX2-in-19-sub has 50 copies of Rep/Cap per 15 μg of total cellular DNA that was analyzed. Furthermore, there was hardly an increase in Rep/Cap gene amplification after the addition of Ad-Cre, suggesting that the initial copy number of the inducible Rep/Cap was low, and minimally increased after the Rep gene was restored. In comparison, XX2-GFP-145 had about 75 or so copies of integrated inducible Rep/Cap plasmid, but with the addition of Cre to restore the Rep gene, the copy numbers in

this cell line reached about 500 copies within the 15-µg used in the experiment. The system developed in Qiao's work in 2002, was improved by Yuan et. al. in 2011. He had shown that the Rep/Cap copy numbers were not critical for the producer cell line system. However, in the transition of producer cell line to packaging cell line, the Rep/Cap copy number may be of great importance, especially when seeing results from Figure 19.

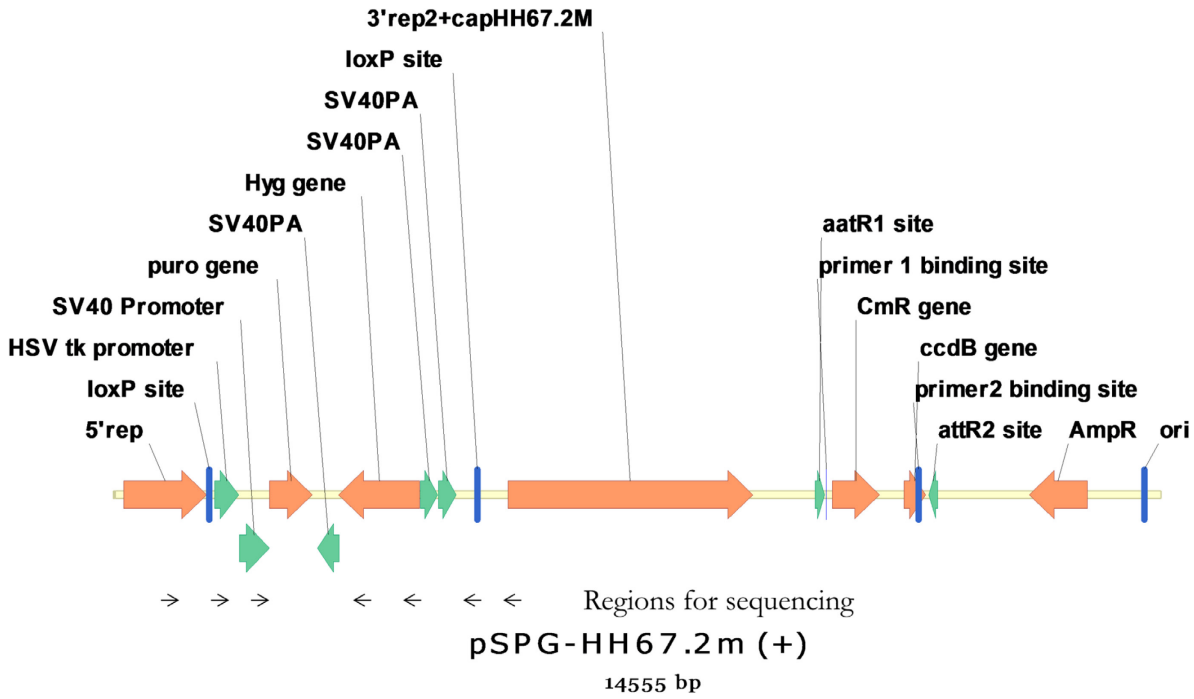


Figure 20: Sequencing of past plasmid

It was evident there was a need for a packaging cell line from Figure 19, but we wanted to use a more clinically or soon to be clinically relevant capsid rather than using AAV2, like in Qiao's earlier cell line designs. A capsid that we were using in our lab is the HH67.2m capsid. Placing this capsid according to the methods described in Yuan's 2011 producer cell line method, we created pSPG-HH67.2m, an inducible Rep plasmid that contains the HH67.2m capsid sequence. Unfortunately, when we generated a packaging cell line, we saw several things that in hindsight indicated something was wrong with the system before it ever entered the cell.

These were: 1) low colony numbers, or none at all, 2) yields of resulting rAAV decreased substantially over time, 3) growth of the cultures were becoming slower over time, and 3) health of the cells were progressively deteriorating. Our troubleshooting events were to first repeat it with different researchers performing the transfection and then to try different amounts of transfected plasmid. Finally, we realized there may have been an issue with the plasmid itself and therefore we set out to investigate the regions of the plasmid that specifically regarding the expression of Rep. To do so, we used multiple sequencing plasmids to identify areas between the two LoxP sites to see if the intron is still intact, as shown in Figure 20 with the arrows drawn below the plasmid that indicate the primers and direction of primers used for sequencing. Unfortunately, the intron had somehow lost its PolyA signals, as seen in the results of the sequencing displayed as updated maps of the packaging or producer plasmid in Figure 21.

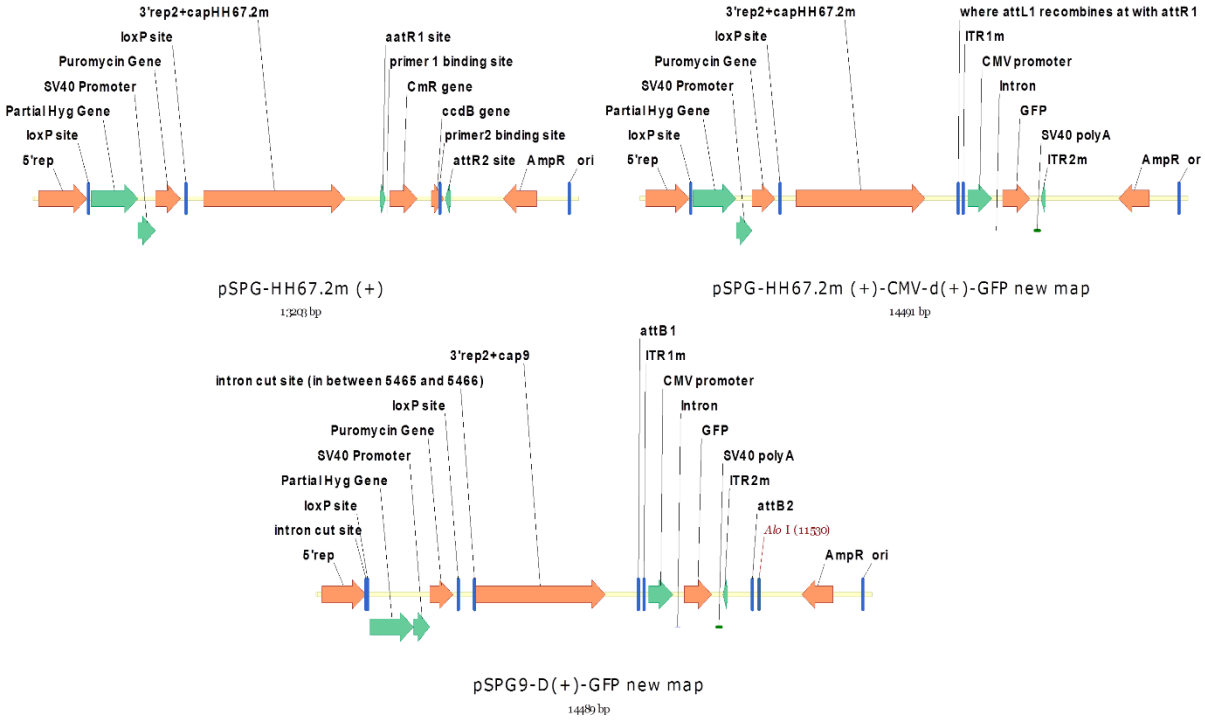


Figure 21: Updated plasmid maps after sequencing of the dual-splice switch

With these results, we had to resort to using the packaging plasmid used in Qiao's 2002 AAV2 cell line generation called pXX2-SseI-2LoxP-in-Puro. At the time of this report, we are still struggling to make higher plasmid yields of the pSPGHH67.2m, pSPGHH67.2m-ds-GFP, and pSPG9-ds-GFP. We hypothesize the reason for this issue is the puromycin resistant gene between the two LoxP sites is somehow toxic in bacterial cells, the backbone of the pSPG plasmids is not the right backbone to use for adequate plasmid production, or we are using an incorrect bacteria strain to generate more plasmid product. Yields of pXX2-SseI-2LoxP-in-Puro did not suffer scalability issues during large scale plasmid production, the PolyA signals were still present in the intron, and the puromycin resistant gene in this plasmid was not detrimental to the plasmid. Since the issues were minimal to generate this plasmid we used it for the remainder of the study to generate AAV2 packaging cell lines. In turn, in lieu of the inherent issues of the plasmids we initially intended to use that were more clinical relevant to the field, this project has become a proof-of-concept study to help converge designs from Aim 1 and Aim 2.

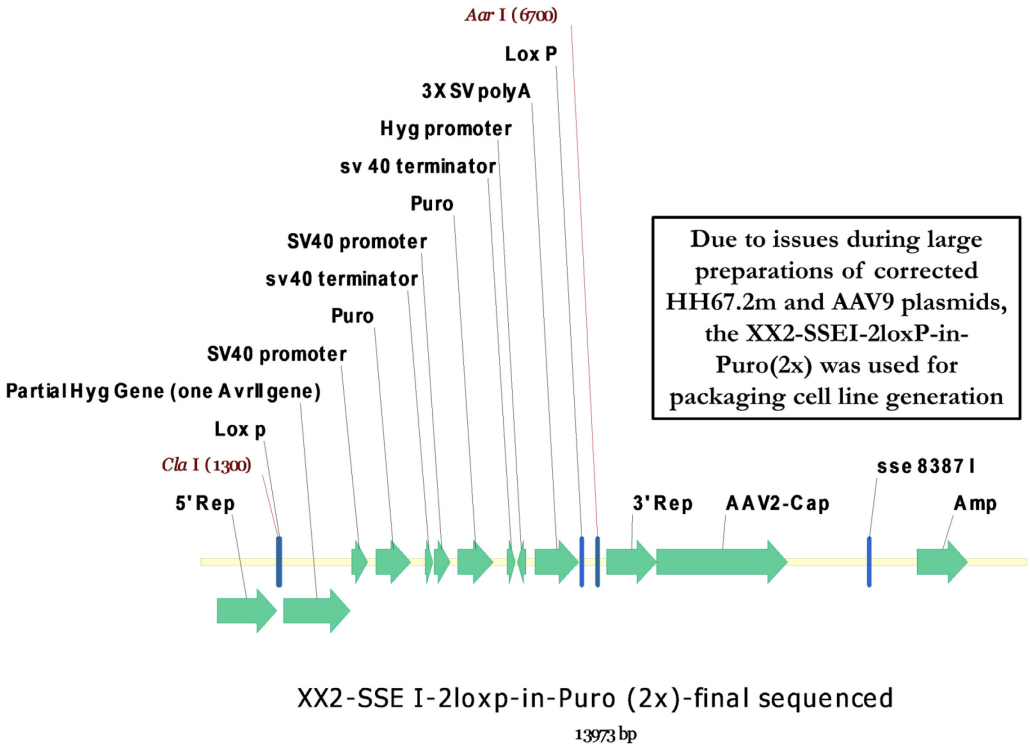


Figure 22: Construct of inducible AAV packaging plasmid used in packaging cell lines

Table 9 is compiled data of the cell colony generation for cell lines that had linearized plasmids randomly integrated and the cell lines generated using the AAVS1 CRISPR/Cas9 system. As can be seen from this data, there were more total colonies using AAVS1 CRISPR/Cas9 than for randomly integrated, suggesting that the frequency and integration rate of the linearized plasmid is increased. Of the 3,792 colonies that were generated in the AAVS1 CRISPR/Cas9 set of cell lines, 223 were selected for further characterization. For the 2,969 colonies generated in the randomly integrated set, 92 were selected for further characterization. Of these subset of total colonies, the several hundred candidates were narrowed down, eventually leading to a single candidate colony from each generation method.

Table 9: Cell Colony Generation Statistics

	Using AAVS1 CRISPR/Cas9	Randomly Integrated
Total Colonies Generated	3792	2969
Number of Colonies Selected	223	92
Selected Colonies for Further Characterization	26 (11.7%)	4 (4.3%)

Figure 23 shows the characterization results of the cell lines that were selected from these two cell line generation methods. We used the adenovirus construct Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) to measure LacZ infectious titers from the resultant AAV2-CMV-LacZ-nLs. From the results gathered from this characterization study, there are several observations that can be made for both generation systems. First, on average, colonies selected from the AAVS1 CRISPR/Cas9 cell line generation systems had about a 2-fold increase in infectious unit titers when compared to the average cell colony in the randomly integrated generation system. Second, the number of colonies increased, suggesting that there was an increase of double strand breaks with effective integration to yield more candidate colonies. Third, the overall number of colonies that generated a moderate-to-high infectious titer, determined by candidates producing $>1.5 \times 10^7$ IU/10-cm plate at the small scale, was nearly 3-times higher in AAVS1 CRISRP/Cas9 systems. As can be seen in Figure 23, Clone 152 was the best for AAVS1 CRISPR/Cas9 cell lines, and Clone 253 was considered the best overall – i.e. best cell viability, growth, etc. - for randomly integrated cell lines. Clone 152 was subcloned to generate a more uniform population, since the original population may have had low-yielding contaminating populations that would impact overall vector yields. The cell lines to show best yields after subcloning were Clone 152.69 and Clone 152.74.

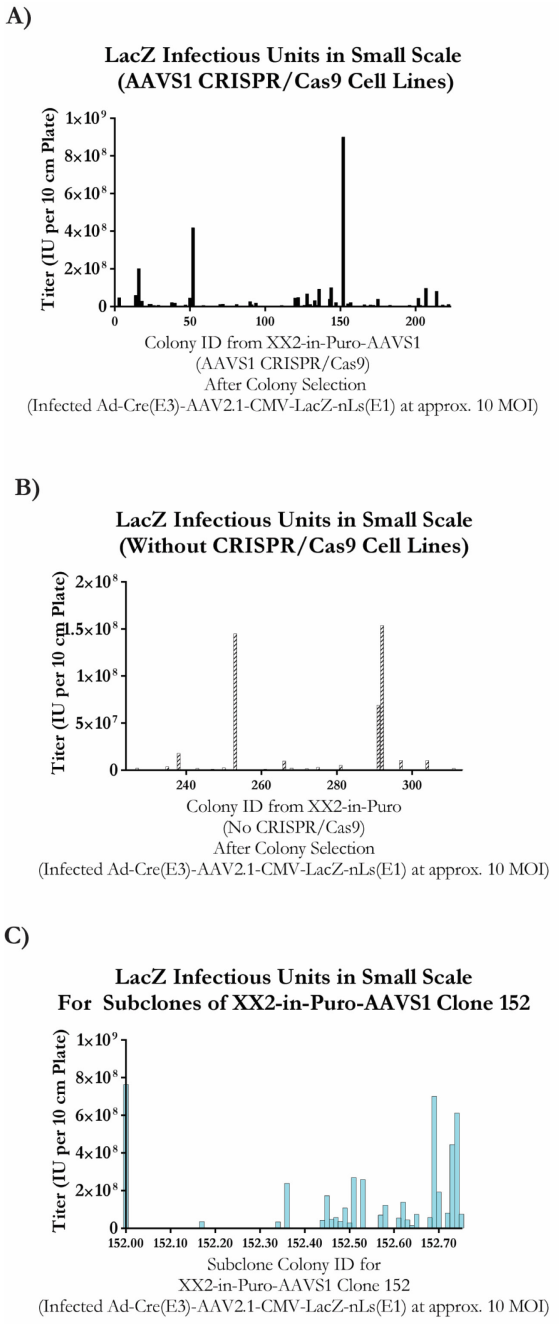


Figure 23: LacZ infectious units for selection of best packaging cell candidate. A) Cell lines for AAVS1 CRISPR/Cas9 integration of packaging plasmid. Clone 152 was considered the best yielding cell line. B) Cell lines for random integration of packaging plasmid. Clone 253 and 292 were considered the best yielding cell lines. C) Subclone of clone 152 to have a purified population of packaging cell line. Subclones 152.69 and 152.74 were the best.

These three cell lines were used extensively in studies for Chapter 2. We saw earlier the yields of the novel adenovirus construct, but the cell line also plays an equal role to a high yielding system. Therefore, we analyzed the relationship between cell line productivity with the Rep/Cap gene copy numbers integrated in cells. These results are counter to what was reported by previous studies by Yuan et. al. in 2011 because the systems used to make rAAV are different. In the previous method, the rAAV vector was produced in cis while the adenovirus was only used to activate Rep/Cap. This would in turn continually produce Ad, increasing helper functions, and generate just enough Rep/Cap for high yielding vector. In our modified system, however, we are provided rAAV and the "on-switch" by the same adenovirus vector. Here, the Ad can no longer continually replicate as freely since the adenovirus is cleaved when translated Rep proteins identify the harbored ITRs in the Ad virus. Therefore, it is critical to have a substantial amount of Rep/Cap present and an appropriate amount of Ad present for adequate vector production. This method difference is enough to see a relationship between the integrated Rep/Cap gene copy number and vector productivity, that was not seen in this investigation from our previous studies. It is more pronounced depending on the adenovirus construct used to activate the dual-splice switch, discussed below. We found that in our best producing cell lines, XX2-in-Puro-AAVS1 Clones 152.69 and 152.74, the Rep/Cap copy number was at least 9 times more than XX2-in-Puro Clone 253. XX2-in-Puro-AAVS1 Clone 152.69 had an average Rep/Cap copy number of 87 copy numbers/cell, 152.74 with 64 copy numbers/cell, and XX2-in-Puro Clone 253 had an average of 7 copy numbers/cell. As shown in Chapter 2, this difference in Rep/Cap copy number between the AAVS1 CRISPR/Cas9 assisted integration and random integration can lead to up to an order of magnitude of titer difference, independent of the novel Ad-Cre(E3)-rAAV(E1) construct. By this comparison alone, there is a

clear importance of total integrated packaging plasmid present for our packaging cell line and its influence on vector yield.

Figure 24 shows the relationship between adenovirus construct used and Rep/Cap copy numbers generated per cell in the cell lines XX2-in-Puro-AAVS1 Clone 152.69, XX2-in-Puro-AAVS1 Clone 152.74, and XX2-in-Puro Clone 253. Ad-Cre constructs, whether in E3 or E1, showed a significant increase ($p=0.0001$) in copy number for XX2-in-Puro-AAVS1 cell lines, going from 60-90 copy numbers/cell to 630-770 copy numbers/cell for Ad-Cre(E3) and 1,000-1,370 copy numbers/cell in Ad-Cre(E1). The novel Ad-Cre(E3)-AAV(E1) constructs showed significant increases in clone 152.69 for both adenovirus constructs ($p<0.001$), but only Ad-Cre(E3)-AAV-ds-GFP(E1) showed significant increase in clone 152.74 ($p=0.0001$). XX2-in-Puro cell lines showed no statistically significant increases in Rep/Cap gene copy numbers no matter which construct was used, but the registered copy numbers went from 7 copy numbers/cell to >125 copy number/cell after Ad-Cre infection. It is interesting to see the substantial increase in Rep/Cap for Ad-Cre constructs that do not contain an AAV vector. It is theorized that the Ad-Cre constructs will continue to amplify, unimpeded by any restored Rep/Cap from the dual splice switch. However, the Ad-Cre(E3)-AAV(E1) constructs cannot amplify to the same extent as Ad-Cre constructs because the restored Rep gene will identify the AAV vector in the adenovirus and remove it from the adenovirus genome, thereby nullifying the adenovirus from replicating in the cell line. This is evidenced multiple times in our studies, particularly in Chapter 2 (in Figures 12 and 13), and in Figures 25 and 26.

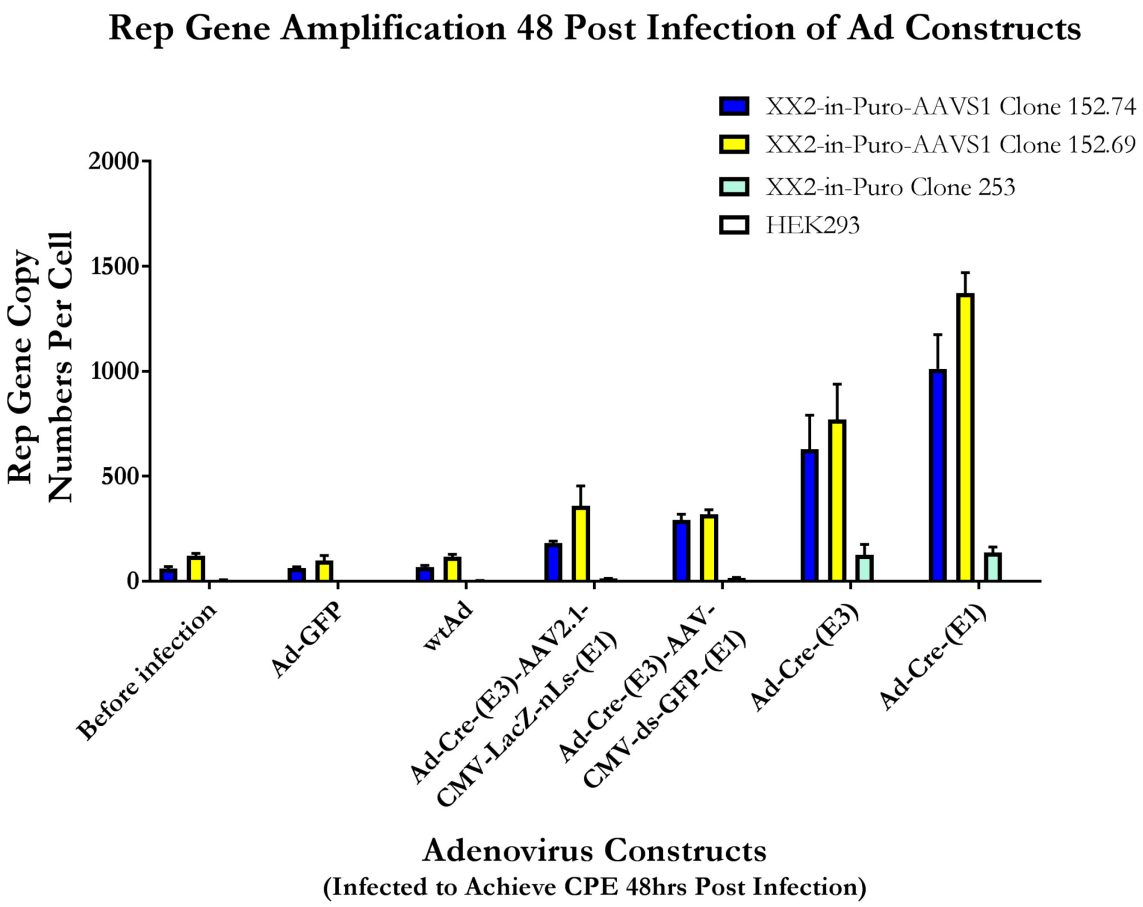


Figure 24: Rep gene amplification on different Ad constructs. Negative controls were wtAd and Ad-GFP.

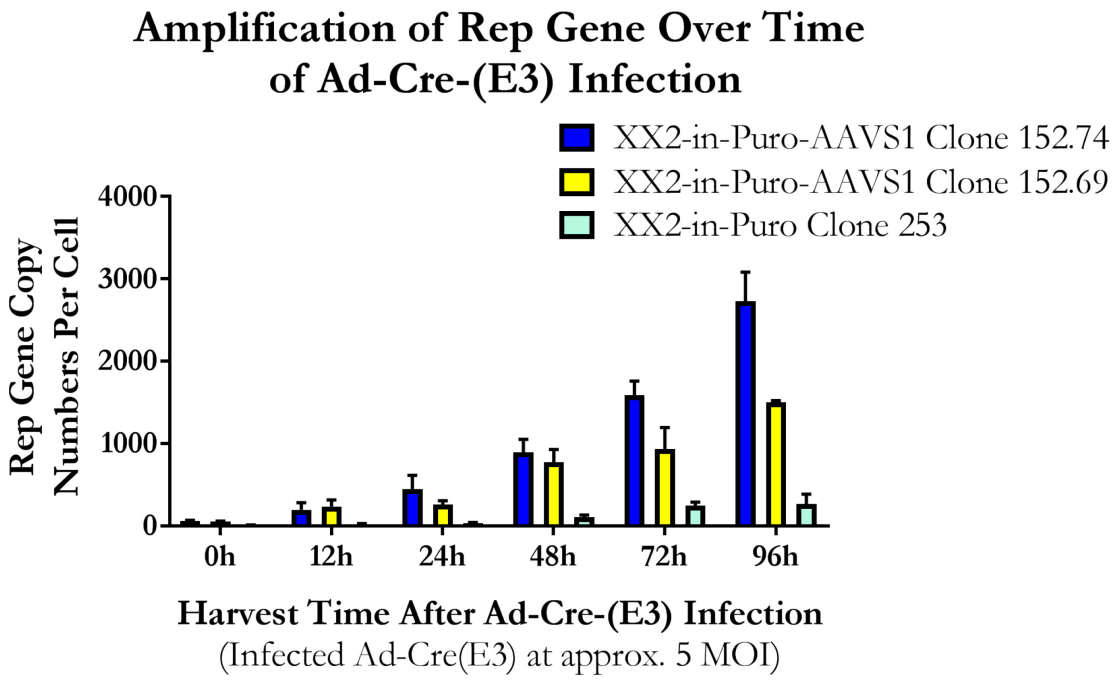


Figure 25: Rep gene amplification over time using Ad-Cre(E3)

To see if when the Rep and Cap genes were amplified in the packaging cell lines after adenovirus infection, we analyzed a time course Rep and Cap copy numbers/cell over 96 hours after infection of either Ad-Cre(E3) or Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1). For Ad-Cre(E3) infection, as shown in Figure 25, we see a statistically significant increase in Rep/Cap gene copy numbers after 48 hours for the XX2-in-Puro-AAVS1 cell lines, and again we see no statistically significant increase in XX2-in-Puro Clone 253. The Rep/Cap gene copy numbers increase to greater than 1,500 copy numbers/cell after 96 hours. In this gene amplification study, we see similar results to what was shown in Figure 24, where the Ad-Cre constructs are amplifying Rep to high copy numbers. This is most likely attributed to the unperturbed replication of Ad-Cre in the packaging cell line, and nearly continual production of the Cre recombinase protein.

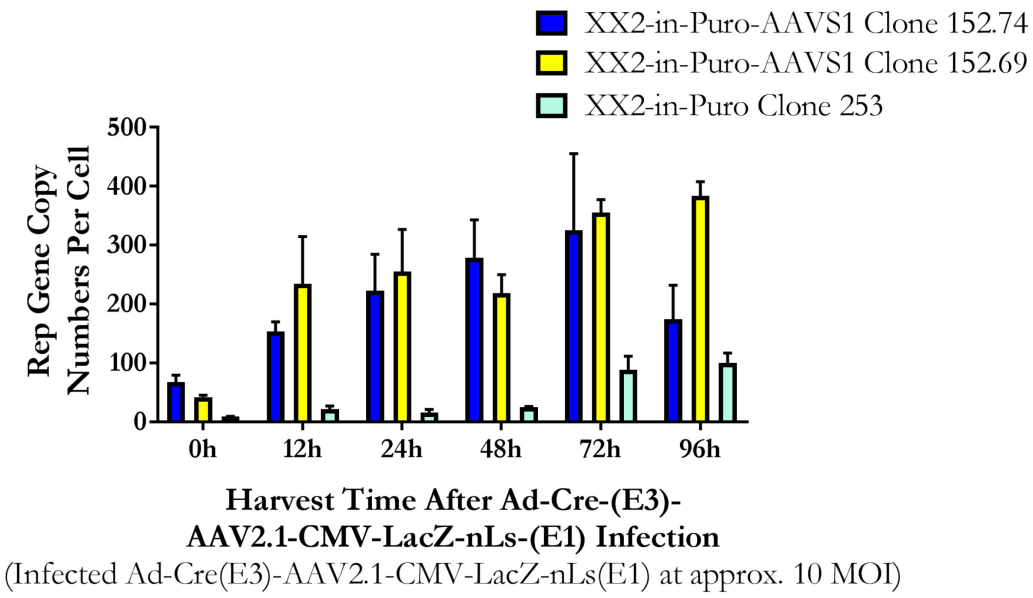


Figure 26: Rep gene amplification over time using Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1)

For Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1) infection, as shown in Figure 26, we see a statistically significant increase in Rep/Cap gene copy numbers after 24 hours for the XX2-in-Puro-AAVS1 cell lines, and again we see no statistically significant increase in XX2-in-Puro Clone 253. Based on Figure 24, we unsurprisingly have the Rep/Cap gene copy numbers increase to greater than 300 copy numbers/cell after 72 hours, which is 5 times less than Ad-Cre(E3) Rep/Cap copy numbers/cell. As described before, this is likely from the adenovirus no longer being to replicate as freely as its Ad-Cre counterparts. Here the Rep proteins can identify the AAV vector and excise it from the adenovirus, leaving the adenovirus to be replication defective. The total Rep copy number for the XX2-in-Puro-AAVS1 cell lines likely comes from the total Cre protein that was able to be generated before the adenovirus was disrupted by the activated Rep protein. Since several copies of adenovirus containing Cre are presented into a cell initially, there are enough Cre proteins to reactivate the Rep integrated into the genome. This is

seen in the increase in Rep copy number from 0 to 12 hours. When Rep proteins are generated, they would excise the rAAV from the adenovirus, thereby turning off new Cre generation. The rate at which new restored Rep copy numbers are present is consequently slowed, leading to Rep copy numbers to start to stabilize after about 48 hours. The decrease seen in XX2-in-Puro-AAVS1 Clone 152.74 at 96 hours can be attributed to adenovirus-induced apoptotic effects on the DNA present in the cells. The explanation observed in the control randomly integrated XX2-in-Puro Clone 253 is harder to realize in the graphical form since the increase is observed as marginal when compared to the XX2-in-Puro-AAVS1 cells. However, this control condition follows similar observations to what is seen in the XX2-in-Puro-AAVS1 cells, but do not have statistically significant changes.

For statistical analysis of the statistical significance during the time course or adenovirus construct experiments, please refer to Appendix C.

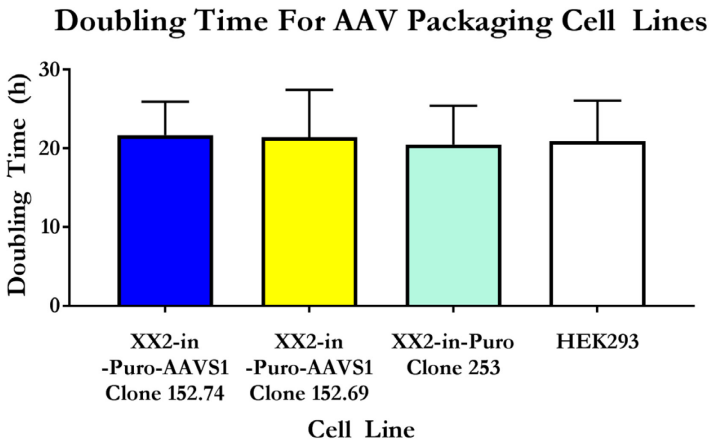


Figure 27: Growth rate characteristics of developed packaging cell lines

The AAV packaging cell lines were stable in its titers and had normal growth rates and morphology indistinguishable from the parental HEK293 cells, as seen in Figure 27. There was no statistical difference in the growth rates (statistics provided in Appendix C).

3.5 Discussion

Our lack of a packaging cell line was a clear bottleneck to our proposed adenovirus/packaging cell line method design. With a sufficient packaging cell line, it is possible the adenovirus design could become a technology that would be effective in suspension cultures to make clinical grade rAAV at larger late-stage clinical scale or during market manufacturing. In our initial attempts at generating a packaging cell line, we wanted to use a more clinical relevant AAV serotype to package our rAAV vector provided in trans by the adenovirus. However, we came across issues to make a more clinically relevant serotype, especially for a serotype for systemic muscular diseases. This was because of an either unstable backbone plasmid for our packaging construct, a toxicity associated with the puromycin resistant gene, or an incorrect bacterium used for large scale plasmid production. While we try to troubleshoot the issue with these constructs, we went ahead to use an AAV2 inducible packaging plasmid. Because of the use of this less clinically relevant serotype, the generation of an AAV2 inducible packaging cell line was a proof-of-concept.

In this study, we have increased the efficiency of double-strand breaks by using AAVS1 CRISPR/Cas9. The system provides several advantages. First, the increased efficiency of double strand breaks means an increase in Rep/Cap copy number integrations per cell. We can see this indirectly with the generation of more colonies when compared to the randomly integrated cell line control. We can more definitively when quantifying Rep copy numbers per cell in our best generated cell lines XX2-in-Puro-AAVS1 Clones 152.69 and Clone 152.74 when compared to the randomly integrated cell line XX2-in-Puro Clone 253. We found that in our best producing cell lines, XX2-in-Puro-AAVS1 Clones 152.69 and 152.74, the Rep/Cap copy number was at least 9 times more than XX2-in-Puro Clone 253. XX2-in-Puro-AAVS1 Clone 152.69 had an average Rep/Cap copy number of 87 copy numbers/cell, 152.74 with 64 copy numbers/cell,

and XX2-in-Puro Clone 253 had an average of 7 copy numbers/cell. Second, we see that with the increase of colonies generated using CRISPR/Cas9 we can see a direct relationship of increased number of characterized colonies showing moderate to high titers compared to the randomly integrated cell lines. About 35% of the colonies characterized from XX2-in-Puro-AAVS1 generated no noticeable product during the LacZ quantification assay, and 44% showed no product for XX2-in-Puro cell colonies. On average, the XX2-in-Puro-AAVS1 cell lines produced about $1.35x10^7$ IU/10-cm plate out of the remaining 145 colonies capable of generating noticeable AAV2-LacZ product and on average XX2-in-Puro cell lines produced about $5.77x10^6$ IU/10-cm plate out of the remaining 53 colonies capable of generating noticeable AAV2-LacZ product. Finally, this new cell line generation, in tandem with the adenovirus construct described in Aim 1, eliminates the need to create a novel cell line for every rAAV vector gene desired to be in a final AAV product. Integration of only the inducible packaging plasmid, with the assistance of a double strand break generator in CRISPR/Cas9, can create a packaging cell line for any rAAV vector delivered in trans with Cre recombinase and thereby lead to a system where the creation of a cell line can be performed once for a specific serotype, and the resource intensive process would transition to the much-improved construction of the novel adenovirus. Therefore, the upgraded method of introducing several double strand breaks, as opposed to randomly formed double strand breaks, can save significant time to generate a high copy number of the inducible Rep/Cap gene in the HEK293 cell.

Ideally, the cell line that is generated in this aim can be transitioned to a suspension culture by gradually modifying the media the cells are growing in, much like what was done for the suspension cultures used for PEI-mediated triple plasmid transfection[16]. This would allow for the cells to be grown in a larger container, such as a 100L or greater sized bioreactor, and with the single vector infection into the cells, the confluency of the cells can be increased

without compromising the yield of the product. In this theorized scenario, the demand of rAAV vector supply for systemic diseases can be met with less intensive production protocols to generate the rAAV product. The packaging cell line reduces upfront labor by removing the need to generate a novel cell line for each desirable rAAV vector in need of large scale production. Furthermore, this process is expedited by introducing more double strand breaks using CRISPR/Cas9 as the generator of multiple double strand breaks, ideally meant for site-specificity, but prone to off-target effects especially for the Streptococcus pyogenes (S. pyogenes) Type II CRISPR/Cas9 system. The downstream generation and infection of the novel adenovirus eliminates the necessity for a more difficult recombination of the adenovirus using AdEasy methods and also contributes to the design of the cell line by moving the rAAV vector onto the adenovirus.

3.6 Conclusions

As mentioned in the conclusion of the previous chapter, hand in hand the generation of this packaging cell line with the novel design of the adenovirus would give a new methodology for rAAV production, requiring an easy-to-construct adenovirus infected into an established stable packaging cell line. Increased copy numbers of an inducible Rep/Cap in the HEK293 genome plays a critical role in this packaging cell line/Ad-Cre(E3)-rAAV(E1) system. With the assistance of a double strand break generator, such as the easy-to-use CRISPR/Cas9, the copy number of integrated plasmid would increase per cell and give better conditions for rAAV yields in this novel production method. The pitfall of this system is the adenovirus contaminant which will be discussed further in the next chapter.

REFERENCES

1 Gaj, T., Gersbach, C. A. & Barbas III, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Cell Press: Trends in Biotechnology 31, 397-405 (2013).

2 Klug, A. The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. Annu. Rev. Biochem. 79, 213-231 (2010).

3 Mali, P. et al. RNA-Guided Human Genome Engineering via Cas9. Science (New York, N.Y.) 339, 823-826 (2013).

4 Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. Nature 482, 331-338 (2012).

5 Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (New York, N.Y.) 339, 819-823 (2013).

6 Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature Biotechnology 32, 279-284 (2014).

7 Cho, S. W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Research 24, 132-141, doi:10.1101/gr.162339.113 (2014).

8 Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S. & Yang, S.-H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Molecular Therapy - Nucleic Acids 4, doi:10.1038/mtna.2015.37 (2015).

9 Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31, 227-229 (2013).

10 Sander, J. D. et al. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Research 38, W462-W468, doi:10.1093/nar/gkq319 (2010).

11 Yuan, Z., Qiao, C., Hu, P., Li, J. & Xiao, X. A versatile adeno-associated virus vector producer cell line method for scalable vector production of different serotypes. Human gene therapy 22, 613-624 (2011).

12 Luo, Y. et al. AAVS1-Targeted Plasmid Integration in AAV Producer Cell Lines. Human Gene Therapy Methods 28, 124-138, doi:10.1089/hgtb.2016.158 (2017).

13 Sanes, J. R., Rubenstein, J. L. & Nicolas, J. F. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. The EMBO Journal 5, 3133-3142 (1986).

14 Corporation, P. CellTiter-Glo® 2.0 Assay. Instructions for Use of Products: G9241, G9242 and G9243. (2017). <https://www.promega.com/-/media/files/resources/protocols/technical-manuals/101/celltiterglo-2-0-assay-protocol.pdf?la=en>.

15 Grieger, J. C., Soltys, S. M. & Samulski, R. J. Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector. Molecular Therapy 24, 287-297, doi:10.1038/mt.2015.187 (2016).

CHAPTER 4: ATTEMPTS AT REMOVAL OF ADENOVIRUS, DNA, AND CELLULAR CONTAMINANTS FROM AAV PREPARATIONS USING THE SELECTIVE PRECIPITATING AGENT DOMIPHEN BROMIDE

4.1 Overview

Recombinant adeno-associated viruses have significant therapeutic potential, but current production methods are inherently filled with unneeded components to the final rAAV product. In particular, for methods requiring the adenovirus as a helper vector for large scale production, the rAAV product post-harvest would have contaminating materials concomitantly mixed into solution. These adenoviruses must be inactivated and removed from AAV preparations for safety. Thermally inactivating these viruses is problematic since it can still lead to immune responses if left over virus or viral components are present. Here, we attempt to remove contaminating entities in the rAAV production method, using a selective precipitating agent called domiphen bromide. Although this detergent is great in removing contaminating materials such as DNA and adenovirus, there are major difficulties to make the detergent ignore certain serotypes of AAV, particularly AAV8 and AAV9, but is able to leave AAV2 titers unperturbed with increasing concentrations of the detergent. In general, for AAV8 and AAV9 the change in the surrounding environment of the AAV after the addition of this detergent possibly makes the viral particles sticky to one another, other contaminating entities, or the partially formed micelle structures of domiphen bromide in solution. From this project, an alternative method for removing adenovirus is necessary, possibly by high hydrostatic pressure or by nanofilters.

4.2 Introduction

The current process of rAAV manufacturing using the method involving adenovirus infection into an inducible rAAV packaging or producing HEK293 cell line yields concomitantly produced adenovirus, albeit E1-, E3- deleted. Fears of using adenovirus in gene therapy medications are still present since the 1999 death of Jesse Gelsinger, an 18-year-old ornithine transcarbamylase (OTC) deficient patient, who died after a very high dose of an E1-, E4-deleted adenovirus vector harboring a functional OTC enzyme. The very high dose ($3.8x10^{13}$ total particles) most likely caused acute liver injury that triggered an adenovirus vector-induced shock, due to a cytokine cascade that led to disseminated intravascular coagulation, acute pulmonary complications, and multiorgan failure. Since this incident there has been a reevaluation of Ad vector dosage, safety and toxicity in clinical trials, and guidelines have been issued for studies to be continued[1]. The method of our rAAV production is to use a recombinant adenovirus to help produce rAAV, and since the concomitantly produced adenovirus is no longer of use to the remainder of the process or therapy, it should be removed in compliance with. 21 CFR 610.13, of the Electronic Code of Federal Regulations General Biological Products Standards, stating "Products shall be free of extraneous material except that which is unavoidable in the manufacturing process described in the approved biologics license application."[2]

One main safety issue of importance is to remove the replication capabilities of the adenovirus since this can assist replication of rAAV generated, a potential issue if replication-competent adenovirus is present in rAAV vector stock preparations. It is known that the advent of E1-deleted adenovirus has provided an important safety feature to adenoviral gene therapy, as well as viruses that may use adenovirus during production like our rAAV production method. This deleted E1 region renders the recombinant virus replication-defective, and it was predicted that this would prevent production of unwanted viral proteins by infected cells and limiting both

direct adenovirus toxicity and possible harmful consequences of anti-adenovirus immune responses[3]. However, viral genes are expressed at low levels in cells transduced with E1-deleted vectors, causing direct toxicity and immunogenicity of the viral gene products. Fortunately, a cytotoxic T lymphocyte (CTL)-mediated immune response often leads to the clearance of vector-transduced cells[1]. This has led to the creation of E1-(+/-E3-), E2/4-deleted adenoviruses, and "gutless" adenovirus devoid of all coding viral genes. We have decided to use the E1-, E3-deleted adenovirus because it provides sufficient space to insert our genes of interest, and deletion of other regions is unnecessary and may affect rAAV production.

As this technology has potential for industrialization, we must realize the possibility for generation of replication competent adenovirus (RCA) by recombination of the E1 region in the adenoviral genome with the HEK293 cell line[4]. Although we can use alternative adenovirus constructs, vector production under GMP standards for drug development will likely require use of more recent cell lines which have been transformed with well defined, non-overlapping E1 DNA fragments where homologous recombination is not possible (like in PER.C6 or N52.E6 cells). Since these two cell lines are relatively expensive, the proof-of-concept of rAAV production will remain in HEK293 cultures before investing in more sophisticated cells that are more inclined to meet GMP standards. The appearance of RCAs in vector preparations involving recombinant adenoviruses is a rare and unpredictable event, and therefore it is difficult to control[5]. In manufacturing GMP, this is a major issue since a largescale batch of the material is prepared and a number of reports have already published RCA formation in their large-scale vector preparations. If not removed, RCA may replicate in an uncontrolled manner in the patient and induce significant pathological side-effects. Although this phenomenon is especially frustrating if the E1-deleted adenovirus is meant for use as a gene therapy, our rAAV production method uses adenovirus as a helper function and should be removed immediately after

generation of the rAAV. Nevertheless, the Food and Drug Administration (FDA) have recognized the seriousness of RCA and have required a series of labor-intensive and expensive RCA screening tests.

The separation of adenovirus and rAAV vector preparations is possible with significant differences in the estimated densities of hydrated Ad and AAV particles, 1.35 g/cm^3 and 1.41 g/cm^3, respectively. Unfortunately, this requires an unscalable ultracentrifugation in a CsCl gradient, which may not completely remove all adenovirus or yield strictly rAAV particles (e.g. presence of proteins of similar densities). Resistance of AAV to high temperatures can heat inactivate the helper adenovirus, although this is only a partial heat inactivation of the entire adenovirus present, coupled with the fact that even a minor contamination with heat-inactivated adenovirus can result in a local inflammatory response at the virus injection site[6,7]. Forms of chromatography can be used to separate adenovirus and rAAV, like size-exclusion and hydrophobic interaction chromatography, but these suffer from loss of product and difficulty to separate virus from high molecular weight contaminants. Furthermore, expensive ion-exchange chromatography techniques that have high binding capacity and selectivity for adenovirus can be used, but can be impacted with contaminating cellular DNA that interferes with the binding of adenovirus to the resin.

Given the advancements of chromatography for purification of a variety of AAV serotypes that are being used by companies like Pfizer, chromatography may be the most effective way of purifying a variety of AAV serotypes. Companies like Pfizer are using an affinity ligand made from a proprietary camelid-derived single-domain antibody fragment for AAV affinity purification. Sold as a commercially available resin, called POROS CaptureSelect AAVX resin, the CaptureSelect ligand is a 13-kDa fragment that comprises the three complementary determining regions that form the antigen-binding domain. The resin is able to be used for large-

scale downstream purification for a broad range adeno-associated virus used for gene therapy applications. Reports from retailers of this resin have said this resin features one-step AAV purification from crude material, high specificity, and can handle high flow rates. However, if the solution is clarified of other crude lysate materials, the resulting viral products can be sent through this high-precision affinity chromatography to more precisely purify AAVs of serotypes AAV1 through AAV8, AAVrh10, and other serotypes without interference occurring from materials in the crude lysate.

Consequently, removal of cellular DNA from the vector preparation is an important component to vector purity, especially when considering large scale preparations without the capabilities of an ultracentrifuge. After harvesting and collecting virus containing cells, the cells must be lysed to release virions within the cell. Cell lysis by either freeze/thaw or sonication results in the release of large amounts of host cell DNA and RNA, as well as unencapsulated viral DNA. The DNA and RNA must be digested with nuclease in order to remove unwanted nucleic acids, reduce the viscosity of the cell lysate, and avoid aggregation complicating further purification steps. Benzonase is the most commonly used nuclease because it is available in quantities sufficient for commercial virus purification. Unfortunately, removing residual Benzonase from cell lysates can be difficult and problematic especially at the scale and pace of production needed for industry[8].

Recently, a group of researchers at Merck research laboratories were identifying methods to remove cellular DNA from their high-density cell cultures when generating adenovirus vector preparations since current methods were too expensive and cumbersome for their production scale. They developed a purification process that incorporates selective precipitation of host cell DNA, using selective precipitating agents, enabling a reduction in use of costly nucleases and chromatographic resins while improving DNA and protein clearance capabilities[9].

As the name suggests, selective precipitating agents are certain chemicals that can be added to the cell lysate preparation, at a certain concentration, to precipitate a high percentage of contaminating materials such as nucleic acid molecules, both DNA and RNA, in a selective manner from the macromolecule of interest. It is an inexpensive alternative to remove DNA, RNA, and even proteins for various applications. Selective nucleic acid precipitating agents already exist in PEG, PEI, spermine, spermidine, and inorganic salts. Selective precipitation of nucleic acids from adenovirus is a difficult challenge since these two are highly negatively charged and relatively hydrophobic. Furthermore, precipitating agents are generally cationic detergents that have use in sanitizing agents due to their harsh antimicrobial properties, and have potential to inactivate the adenovirus/rAAV particles during exposure. The Merck research group has identified a cationic detergent, domiphen bromide, which has superior selectivity for host cell DNA and can eliminate the need for nuclease treatment and/or anion exchange chromatography immediately after harvesting.

In their study, the researchers compared cationic detergents that are specifically quaternary ammonium compounds. The detergents tested were monoalkyltrimethyl ammonium salts (cetyltrimethyl ammonium bromide, CTAB), dialkyldimethyl ammonium salts (domiphen bromide, DB), and heteroaromatic ammonium salts (cetylpyridinium chloride, CPC). These detergents contain certain properties: CTAB has a quaternary ammonium group separated by two carbons from quaternary ammonium and is used in purification of plasmid DNA, CPC has a pyridinium group separated by two carbons from quaternary ammonium and is an active ingredient in mouthwash, and DB which has a phenyl ether group separated by two carbons from quaternary ammonium and is also an active ingredient in mouthwash. Therefore, none of them posed a special regulatory concern, if clearance was demonstrated. It is theorized that the "addition of SPA following cell lysis results in the positively charged groups (and aromatic rings,

if present) on the compounds binding to the negatively charged phosphate groups (and base pairs) on the DNA molecules. The hydrophobic tails on the detergents then interact with each other resulting in precipitation," likely due to micellular formation[10].

It was shown, that as concentration of the detergent increased, the DNA in solution approaches 0%. However, the adenovirus concentration in solution is dramatically reduced for all detergents as the detergent concentration increases, but that DB sustained a large decrease in DNA in solution while maintaining adenovirus in high concentrations in solution. Less DB is necessary to remove more contaminants and the minimal amount of product, a feature that is magnified in high cell density work in the Merck patents. Finally, this group published the effects of DB on various adenovirus serotypes, concluding that there is not much serotype preference of DB.

Consequently, in our investigation we will challenge our adenovirus, rAAV, and adenovirus-rAAV solutions against increasing concentrations of domiphen bromide to identify if domiphen bromide can effectively remove host cell nucleic acids and contaminating adenovirus without removing our rAAV. The patents alluded to the use of domiphen bromide for the precipitation of cellular DNA and the contaminating helper adenovirus in large scale production of AAV[10]. It is to this reference that inspired our group to investigate further.

From their preliminary studies, these researchers have identified that domiphen bromide – an extensively used active ingredient in oral hygiene products and topical antibiotic creams produced at cGMP quality – can form micellular like structures around nucleic acid molecules without effecting the adenovirus and its infectivity. This group further stated that nucleic acids will have a higher affinity to the DB than adenovirus, only if the nucleic acid concentration is higher than the adenovirus, elsewise both can be subjected to precipitation. To test the limits of domiphen bromide, the group set out to test the detergent in high-cell density environments

containing adenovirus to simulate their industry production scale. In the high-cell density experiments, the domiphen bromide concentrations can extend further to allow for more nucleic acids to be precipitated without compromising adenovirus recovery, since the host cell DNA is much more abundant than adenovirus[11]. This leads to the following conclusion of these experiments: as the cell density increases, the appropriate domiphen bromide concentration to precipitate most host cell DNA without compromising adenovirus recovery increase because of the greater DNA than adenovirus content in solution. The patents suggest that Benzonase or any other nuclease is no longer necessary, but can be used for a more robust purification. The hypothetical use of this selective precipitating agent is diagramed in Figure 29, where we theorize that the additions of the detergent can selectively remove DNA and adenovirus contaminants and keep rAAV in solution.

This prompts the general hypothesis for this aim: The addition of the selective precipitating agent, domiphen bromide, will allow for contaminating adenovirus to be cleared from rAAV vector preparations and also clear host cellular nucleic acid contamination without need for nucleases like Benzonase.

A)

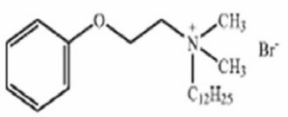
Acronym	DB
Molecular Weight (g/mol)	414
CMC (mM)	1.2-1.8
Structure	

Figure 28: Use of domiphen bromide to selectively remove adenovirus and DNA contaminants. A) the physical properties of domiphen bromide including the structure and critical micelle concentration (CMC). B) Theoretical application of the detergent in rAAV applications.

4.3 Material and Methods

Cell Culture and Virus Propagation

We use three unique producer cell lines in this aim that can produce AAV upon Ad-Cre infection: AAV8-ds-GFP Clone 23 producing AAV8-ds-GFP, AAV9-ds-GFP Clone 1.24 producing AAV9-ds-GFP, and XX2-GFP-145 producing AAV2-ss-GFP. The viruses used in this aim are Ad-Cre(E1) or Ad-Cre(E3). In an initial analysis, cells were scaled to $7x10^8$ cells and infected with Ad-Cre(E1) because these cell lines had the dual splicing switch integrated in its genome. The resulting product was resuspended in DMEM only and underwent three freeze-thaw cycles, spun down at 2500 rpm for 15 minutes, and the supernatant was collected and aliquoted.

Ad-GFP was infected into HEK293 cells to replicate data from Merck. Cells were scaled to about $6x10^8$ total cells and infected at 70-80% confluency, achieving CPE 48 hours post infection. The cells were spun and resuspended with DMEM to match the cell-density reported in the Merck patents. For instance, if using 20 15-cm plates of HEK-293 cells, there is an estimated $8x10^6$ cells per plate and a total of $1.6x10^8$ cells for 20 plates. If the cells were centrifuged and resuspended in 8 mL of DMEM, it would result in an equivalent of $2x10^7$ cells/mL.

Domiphen Bromide Addition to Samples

A pre-calculated amount of 60 mM domiphen bromide dissolved in water was added to the aliquots, mixed for 5-10 seconds, and spun down for 30 minutes at 4 °C. The resulting supernatant of these aliquots were collected and analyzed. Next, domiphen bromide was added to these aliquots at various concentrations, ranging from 0 mM to 6 mM. These were then mixed, and spun down at 4°C for 30 minutes at 13,500 rpm on a table top microfuge. 180 μL of the supernatant was collected to analyze the residual DNA content. The remainder of the

supernatant was subjected to quantification of residual adenovirus, by infecting HEK293 cells with supernatant and counting GFP-positive cells, and also used for qPCR analysis.

Measurement of Adenovirus and GFP in Preparations

qPCR was done to quantify both adenovirus at a conserved sequence in the Ad contaminant and GFP for this aim (ABI PRISM 7700 Sequence Detector, Applied Biosystems). Alternative methods of quantification were dot blot for GFP in the AAV product and Cre in the Ad contaminant. The following primers were used for qPCR. EGFP-F: GTCCGCCCTGAGCAAAGA, EGFP-R: TCCAGCAGGACCATGTGATC, EGFP-FAM: FAM-CCCAACGAGAAGCG-MGB, Ad-F: CAGCGTAGCCCCGATGTAA, and Ad-R: TTTTTGAGCAGCACCTTGCA.

Determining Total DNA Concentrations

The procedure of the DNA extraction follows accordingly: Addition of extraction buffer (10 mM Tris•Cl pH 8.0, 20 mM EDTA pH 8.0, 0.5% SDS) plus 4-5 μL of RNAseA, incubated at 37°C for 1 hr. Then proteinase K was added at a concentration of100 μg/mL, incubated at 50°C for 3 hrs. Next, the solution underwent a phenol-chloroform extraction, followed by ethanol precipitation procedures. The precipitated DNA was diluted with 50 μL of TE pH 8.0 (or 100 μL depending on if DNA pellet was difficult to dilute with 50 μL). The DNA is quantified using OD or Nanodrop readings of the product.

Measurement of Sample Infectivity by Infectious Titer Assay

HEK293 cells were used for infectious titer assays. For LacZ expression, cells were quantified for infectious titer, by counting number of blue cells in the microscope or camera field under bright-field microscopy. For GFP, cells were quantified by counting number of green cells in the microscope or camera field under fluorescent microscope (excitation of GFP is 488 nm and emission is 509 nm). The yield of infectious particles was presented as infectious unit.

The units of infectious units (IU) were either IU/mL or IU/10-cm plate. IU/10-cm plate was calculated by extrapolating from the surface area of the microscope field to the surface area of the well being used to for the infection assay. IU/mL was calculated by dividing the extrapolated well infectious unit by the total volume infected into the well. Larger scaled products were typically presented in IU/mL, whereas smaller scale characterization studies were presented in IU/10-cm plate. Both measurements were consistently used to compare infectivity at a specific scale between various production methods.

Histochemical Staining of Monolayer Tissue Culture Cells for LacZ Activity (X-gal Staining)

Cultured cells were rinsed with 1x Phosphate Buffered Saline (PBS, pH 7.3) and then fixed for >5 min at 4 °C in 2% formaldehyde and 0.2% glutaraldehyde, mixed in PBS. The cells were then overlaid with a histochemical reaction mixture containing 1 mg/mL 4-Cl-5-Br-3-indolyl-β-galactosidase (X-gal), 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide, and 2 mM $MgCl_2$. The X-gal was dissolved in dimethoxysulfoxide (DMSO) at 40 mg/mL, and then diluted into the reaction mixture. Incubation was for 8-14 hours at 37 °C. Cells were then counted for LacZ expression, i.e. blue cells were counted.

Buffers, Salts, and Disruptive Agents

Domiphen Bromide (247480 Sigma Aldrich), a cationic detergent with a CMC of about 1.2 mM, is used as the selective precipitating agent in this aim. Its stock concentration of 60 mM is prepared from powdered form and dissolved in various buffers, depending on the desired pH and/or salt concentration. Buffers that are used are: citrate at pH 3-5, MES at pH 6, HEPES at pH 7, TrisCl at pH 7.5-8, and water. Desired salt concentrations were met using NaCl and $MgCl_2$. In some situations, a buffer of a particular pH and salt concentration were added to the cell pellet to assess recovery of final product after addition of domiphen bromide that was

dissolved in the same buffer. Guanidine HCl was made at 6M dissolved in water to use as a disruptive agent between domiphen bromide and the viral particles.

Negative Staining for Transmission Electron Microscopy

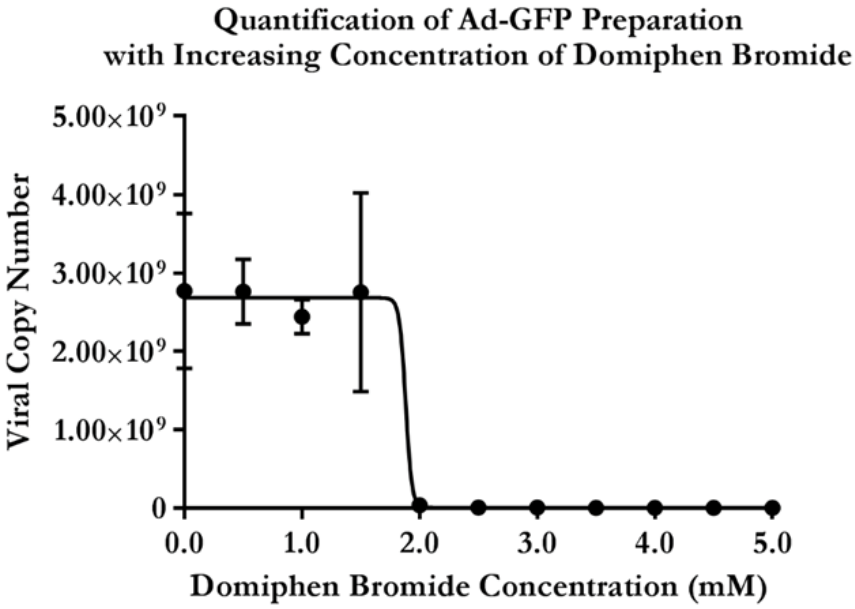
The Microscopy Services Laboratory (MSL) in the Department of Pathology & Laboratory Medicine at UNC-Chapel Hill was used to perform negative staining and transmission electron microscopy. Carbon-coated formvar grids (01754-F Formvar/Carbon 400 mesh, Copper) were glow discharged using Pelco easi-Glow unit to render the surface hydrophilic. The viruses adsorbed onto the grid and subsequently stained using the "Grid-on-Stain or 'Float' Method). Briefly, 25 μL of sample was placed onto a hydrophobic surface and the grid was placed film-side down onto the surface of the droplet for approximately 5 minutes to allow for viral adsorption onto the grid. Next, the grid was quickly transferred to two subsequent droplets of filtered deionized water to remove salts/fixative before staining and the grid was then placed film-side down onto a drop of 2% uranyl acetate in water (pH 4.5, 1-gram uranyl acetate dissolved in 50 mL deionized water for 20-30 minutes, spun for 10-15 minutes to remove debris that may interfere with microscopy) for one minute. Excess stain was wicked off by touch the edge of the grid to filter paper and the grid was then air dried. Grids were then loaded onto a JEOL JEM 1230 Transmission Electron Microscope and images were taken between 100,000X and 150,000X. Data was analyzed for images in 100,000X or 120,000X.

Exclusion criteria and observational criteria to measure full and empty rAAV particles as a result of the 2% uranyl acetate stain are given in detail in Appendix A.

4.4 Results

Firstly, as validation for the work done by the Merck group in their 2013 patent, we decided to use an adenovirus with the reporter function GFP for ease of quantification, by counting in infectious units and by qPCR methods. We intended to verify whether domiphen bromide could clarify the solution of cellular DNA and adenovirus. We imitated cell-density conditions, infected Ad-GFP, and harvested the cells after showing signs of the adenovirus cytopathic effect and detachment. As seen in Figure 30, nearly 99% reduction in adenovirus and DNA is observed, thereby validating the data submitted by Merck in their patent.

A)



B)

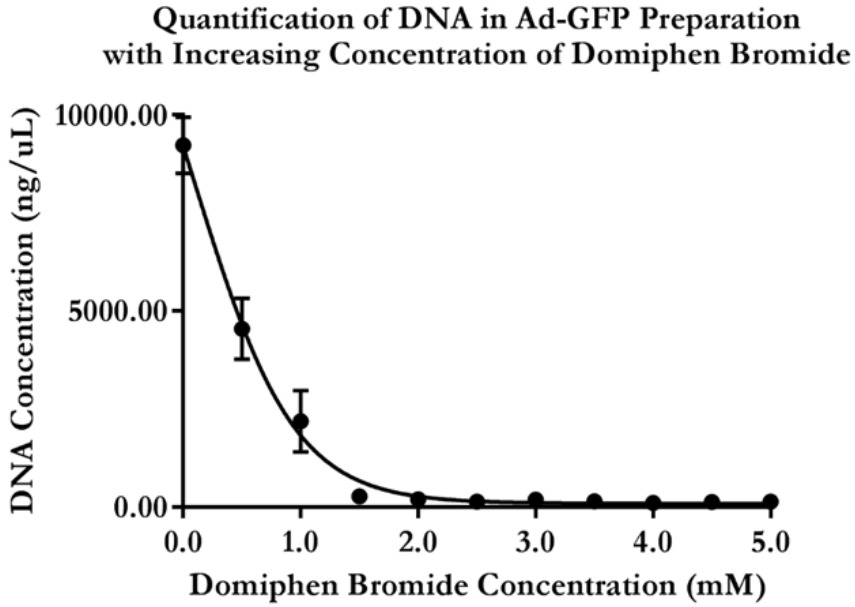
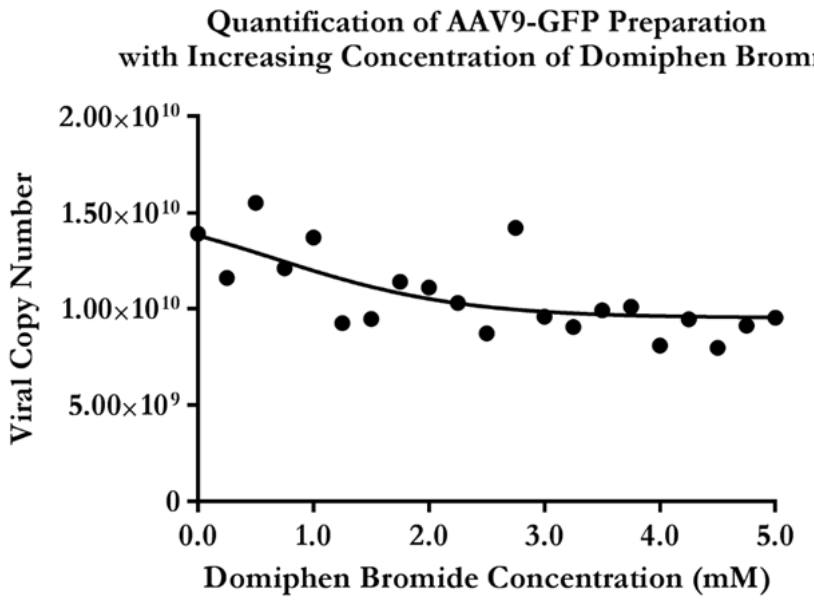


Figure 29: Replication of Merck data for Ad-GFP in HEK293 cells

A)



B)

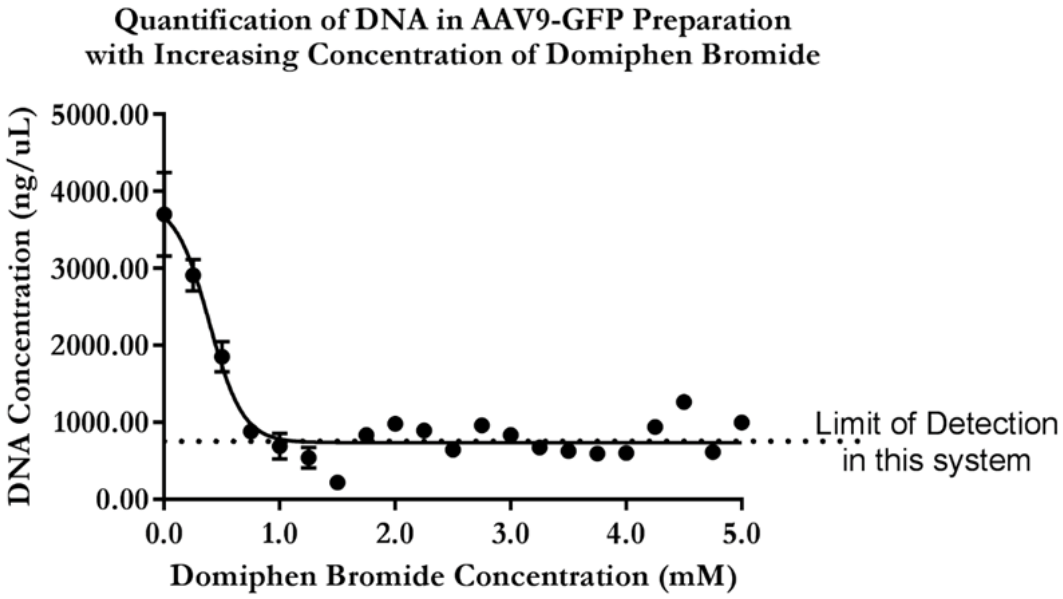
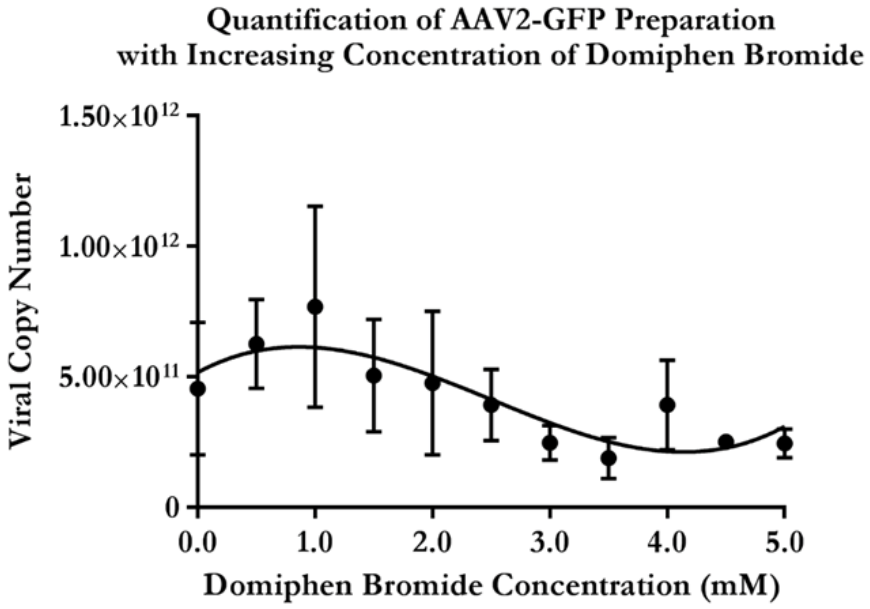


Figure 30: Domiphen bromide interaction with AAV9 generated from triple plasmid transfection

Next, based on the statement in the Merck patent that indicated potential use in AAV processes, we were interested if the detergent interacts with AAV in a production scheme. We used two serotypes for this study: AAV2 and AAV9. In Figure 31, we see that AAV9

precipitates as the detergent concentration increases, losing about 25% of the product, while removing most of the DNA contaminants. This suggests that AAV9 has some interaction with the detergent.

A)



B)

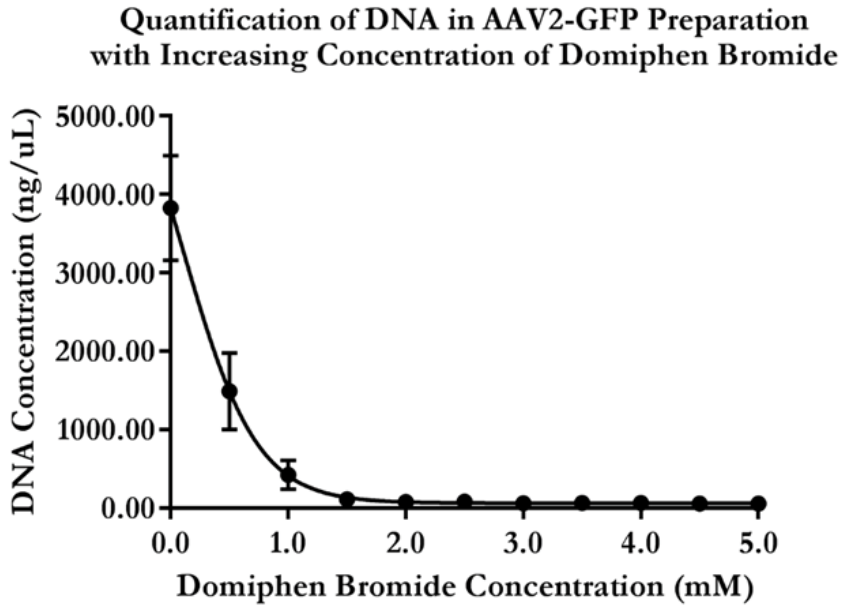
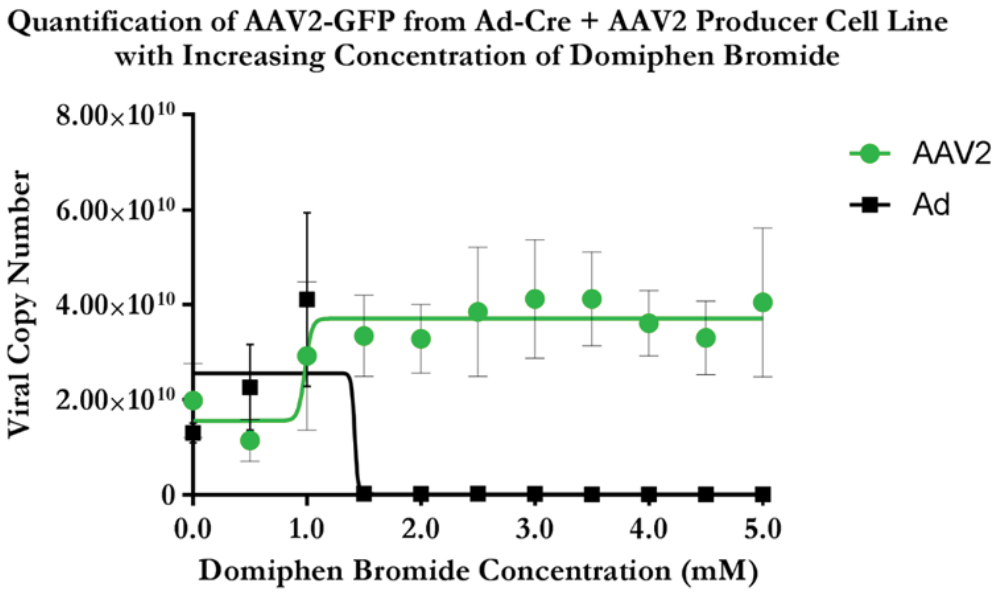


Figure 31: Domiphen bromide interaction with AAV2 generated from triple plasmid transfection

In Figure 32, we see that AAV2 interactions with domiphen bromide are about 50% in reduction at the highest concentration of the detergent. However, when DNA concentration starts to approach roughly 99% removal, at about 2 mM, the AAV concentration remains about the same. This is unlike the results seen in AAV9, where AAV9 precipitated nearly 25% at 2 mM. This difference between the two serotypes reveals an interaction between domiphen bromide and specific serotypes that varies depending on the unique capsid structure of the AAV product. However, we next explored if this interaction can be influenced by more contaminating entities added to the production scheme.

A)



B)

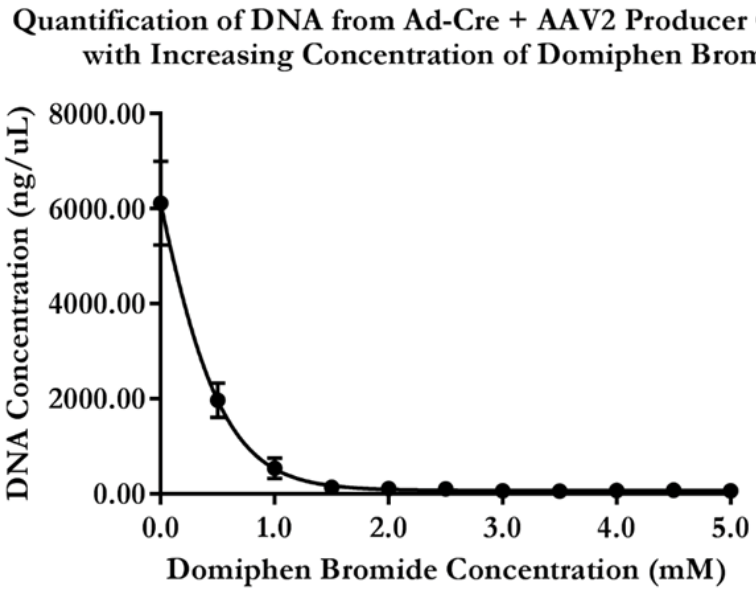


Figure 32: Domiphen bromide interactions with Ad-Cre/AAV2 preparations

To add these additional contaminants, we used the AAV producer cell lines that utilize Ad-Cre to initiate the dual-splicing switch, as described earlier, to generate the AAV product. With these production systems, we would identify if a different profile exists between domiphen bromide and AAV when the additional contaminant of adenovirus is added to the crude lysate.

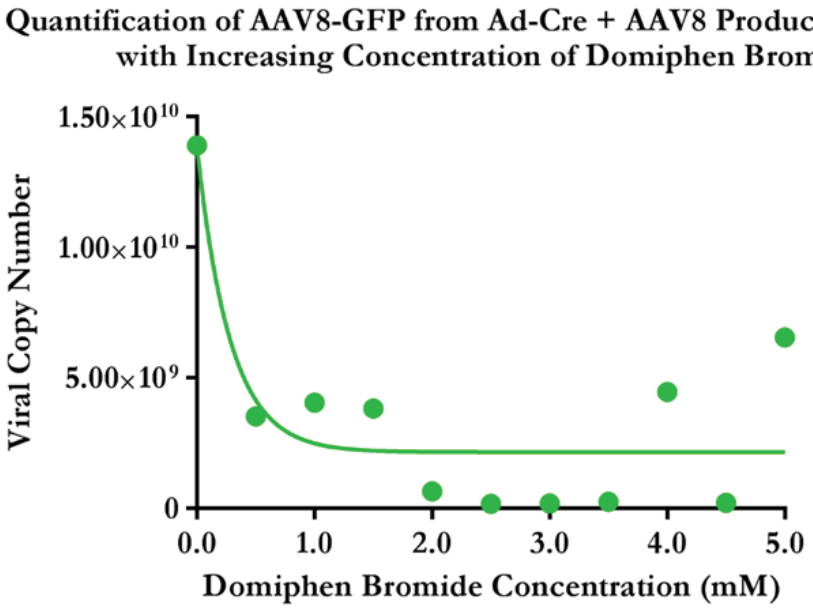
The reasoning behind this was if more contaminants were introduced to the system, the more the AAV can be disguised and not be brought down by the detergent and instead the detergent would have more affinity to the DNA and adenovirus contaminants. We initially tested our hypothesis in a moderate cell density of about $8x10^6$ cells/mL XX2-GFP-145 that generates AAV2-ss-GFP when infected with Ad-Cre. In Figure 33, we see as detergent concentration increases, the adenovirus and DNA concentrations both precipitation out of solution to a final reduction of >95%, all the while AAV2 product remains in solution.

We theorized the use of domiphen bromide in our current purification schemes in laboratory-scale production. Unfortunately, when we used domiphen bromide into our current system of purification involving PEG precipitation and CsCl gradient purification, we seemed to have lost product either through the PEG precipitating step, meaning the detergent interferes with the PEG precipitation, or through the gradient itself. The loss in gradient product could be attributed to Sarkosyl-domiphen bromide interactions, but this was not studied in the timeframe of this aim.

Our next step was to observe interactions between domiphen bromide and more clinical relevant serotypes in our AAV8 and AAV9 producer cell line production methods. Figures 34 and 35 shows the response of more clinically relevant serotypes in these production methods. In Figure 34, we see that we lose a significant amount of AAV8 product as detergent concentration increases and it appears to drop earlier than adenovirus, much like the profile of DNA precipitation. We attempted to lower the loss of product using a multitude of adjustments that could potentially interfere with the interactions between AAV8 capsid and the domiphen bromide. These adjustments were high salt concentrations, acidic pH, temperature of the spin, and the zwitterion L-arginine. Yet, all of these resulted in the same trend, where the loss of AAV8 occurred at around the CMC of domiphen bromide of 1.2 mM. The only modification

that seemed to have worked was using a chaotropic agent called guanidine HCl which seemed to have brought concentrations of AAV8 back up to original levels after domiphen bromide was used, all the while still precipitating the adenovirus. However, removal of guanidine HCl was a bottleneck to the progression of this investigation and it was unknown what the infectivity of the AAV8 was after addition of guanidine HCl.

A)



B)

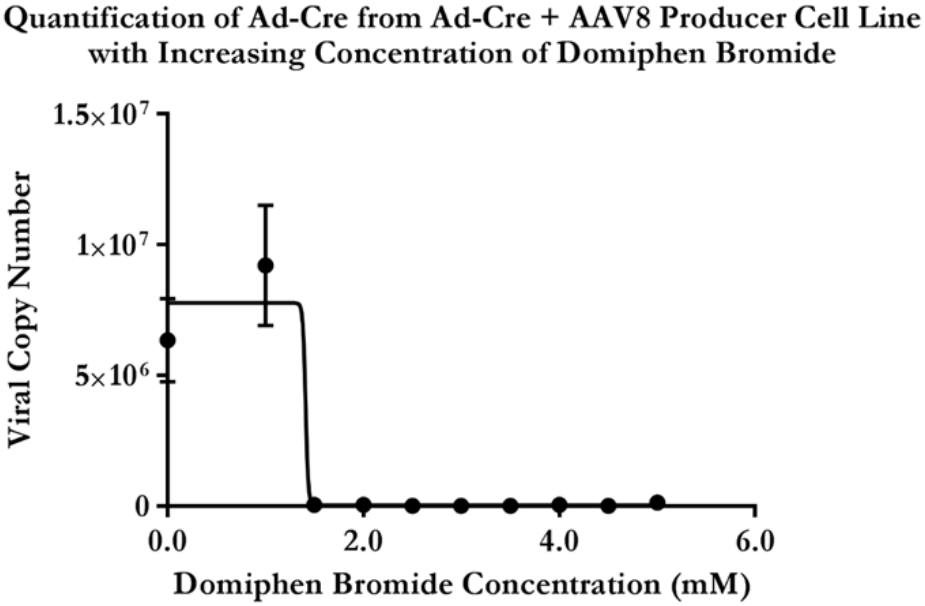
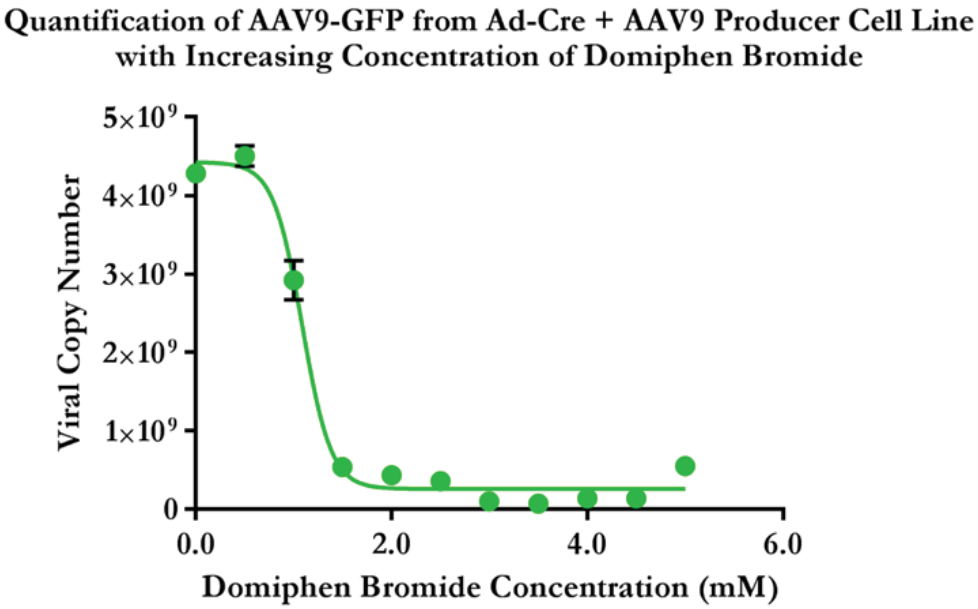


Figure 33:Domiphen bromide interactions with Ad-Cre/AAV8 preparations

Figure 35 shows the profile for AAV9 in panel A and adenovirus in panel B in the AAV9/Ad-Cre production system. Here, the adenovirus is lost out of solution around the CMC of domiphen bromide, 1.2 mM. The AAV9 product is also lost at around the CMC of 1.2 mM. We also performed the same interference techniques used in AAV8/Ad-Cre preparations to reduce the net loss in product, and again it appeared that only guanidine HCl led to the best results. The data for these adjustments are not shown mostly because we do not know the effects of guanidine HCl on the AAV8 or AAV9 viruses, particularly its infectivity. Guanidine HCl is a chaotropic agent that is one of the strongest denaturing agents for proteins. It has the ability to increase the solubility of hydrophobic proteins, but the proteins have the ability to renature upon removal of the guanidine HCl. We did not explore removal of this guanidine HCl due to time constraints. This data was furthermore not shown because we are still unsure how to implement this detergent to our CsCl gradient applications. The use of Sarkosyl (N-lauroylsarcosine) in our purification methods is meant for solubilization and separation of membrane proteins in case sonication, DNase, or RNase could not remove the nuclear membrane of the cells. Furthermore, any cells unable to be lysed by sonication or nucleases could be lysed using this detergent. The investigation on the molecular interactions between domiphen bromide and Sarkosyl for the purposes of implementing domiphen bromide to our purification methods was outside the intent of this Aim. We intended to understand if domiphen bromide can universally precipitate adenoviral and DNA contaminants from crude lysates and leave AAV of varying serotypes in solution for downstream purification. Despite our intent, the difficulty with two of the more clinically relevant serotypes resulted in adjustments that may impede the potency of the AAV product, by rearranging critical proteins on the capsid structure. Therefore, if this system were to be used, other adjustments that have not been already explored

in this aim are necessary. These subtler adjustments would ideally be serotype independent, yet still be able to reliably remove adenovirus and DNA contaminants.

A)



B)

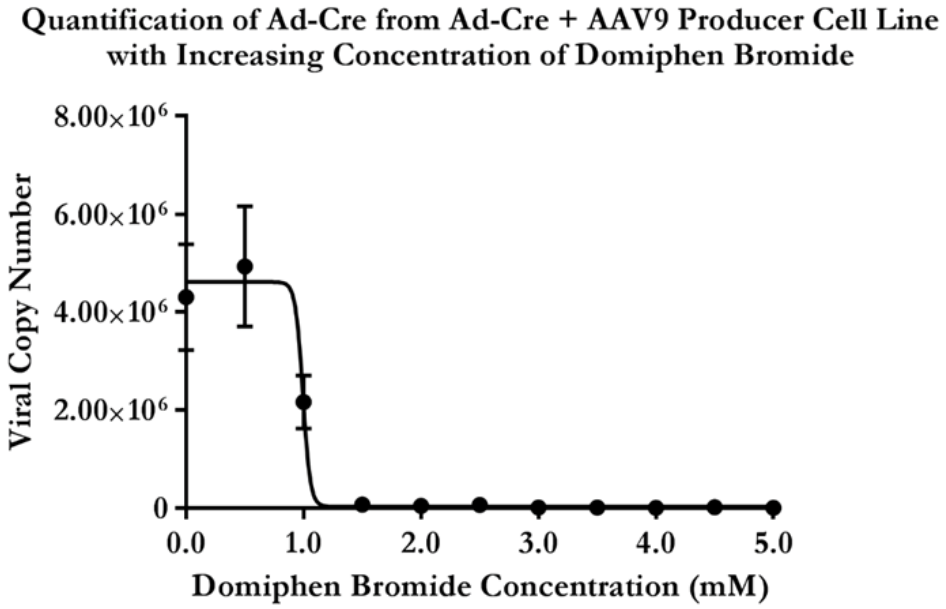


Figure 34:Domiphen bromide interactions with Ad-Cre/AAV9 preparations

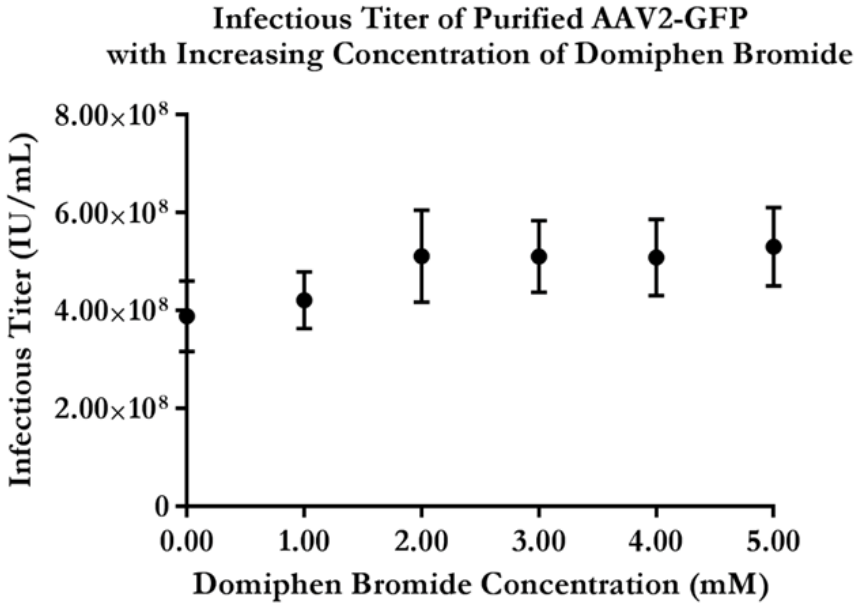
Table 10: Quantification of Purified Viral Particles After Addition of Domiphen Bromide

Domiphen Bromide Concentration (mM)	Purified AAV Virus Titer (vg/mL)			
	AAV2-CMV-ds-GFP	AAV8-CMV-GFP	AAV9-CMV-LacZ-Op	AAVHH67.2m-ds-CMV-GFP
0	6.03E+11	8.16E+10	1.17E+12	2.50E+12
1	7.43E+11	7.78E+10	1.51E+12	2.63E+12
2	9.53E+11	1.83E+11	1.33E+12	2.69E+12
3	1.05E+12	1.66E+11	1.17E+12	1.96E+12
4	2.03E+12	1.91E+11	1.55E+12	2.33E+12
5	7.26E+11	1.95E+11	1.47E+12	2.45E+12

Curious with the outcome of these results, especially for the more difficult to purify serotypes in AAV8 and AAV9, we then investigated if the detergent interacts with AAV to any extent. This was performed to see if precipitation occurs between the detergent and certain serotypes of AAV and if this can be observed from purified stocks. This would allow us to see if there is an innate affinity of the domiphen bromide with the specific serotype and can help rationalize the phenomena occurring in production schemes with more contaminants present. To do this investigation, we purified AAV of various serotypes that had been purified using CsCl gradient and dialyzed to remove CsCl from the final AAV preparation. The serotypes that were investigated were AAV2, AAV8, AAV9, and AAVHH67.2m. Because the AAV solution is placed into a 1x dialysis buffer, we prepared domiphen bromide dissolved in the 1x dialysis buffer before adding to these AAV virus preparations. According to Table 10, the vector genomes remain relatively stable, particularly for AAV2, AAV9 and AAVHH67.2m. According to these dot blot results for these three serotypes, there does not appear to be an interaction between the detergent and the AAV capsid. This generally supports data obtained for AAV2 in the triple plasmid transfection production scheme, but it does not fully elucidate what is occurring for serotype AAV9. In AAV9 production methods, we see the loss of AAV9 product when other contaminants are involved whereas no losses are observed in vector genome

quantification after domiphen bromide concentrations have increased. It is curious to note that AAV8 had a nearly doubled increase in titer, despite the starting aliquots starting at approximately the same titer. Hypothetically, what we think could occur from this observation is that the AAV8 virus is being aggregated with the help of domiphen bromide, and what was aliquoted in our dot blot analysis were aggregates of the AAV8 viruses, rather than a homogenous solution whose aliquots would yield results more apparent in other serotypes like AAV2 or AAV9 in the dot blot assessment. There is a possibility of vector genome results varying depending on the quality of the dot blot assay performed. Because of this unusual result that we observed for AAV8, we investigated with an infectious assay to see if the trends seen in the dot blot are verifiable. These results are presented in Figures 36 for AAV2 and AAV9 and Figure 37 for AAVHH67.2m and AAV8.

A)



B)

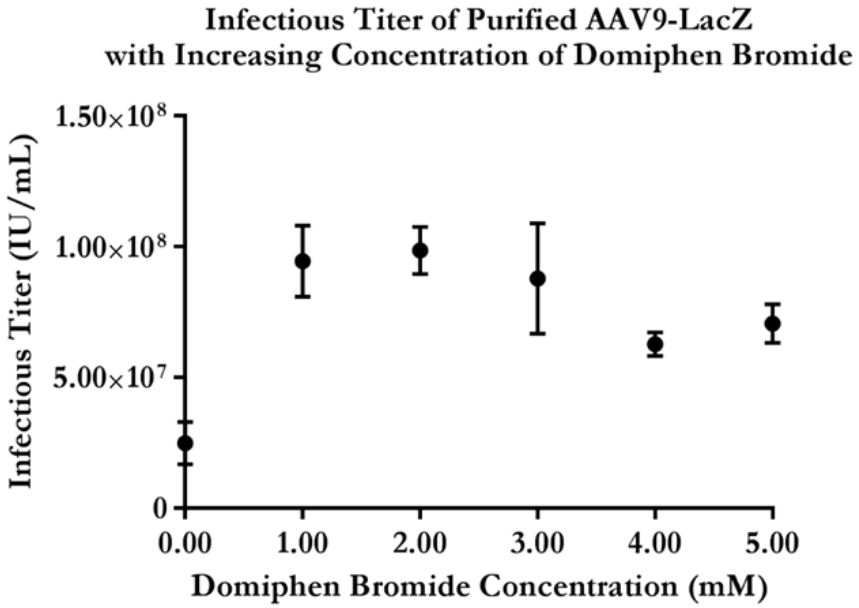
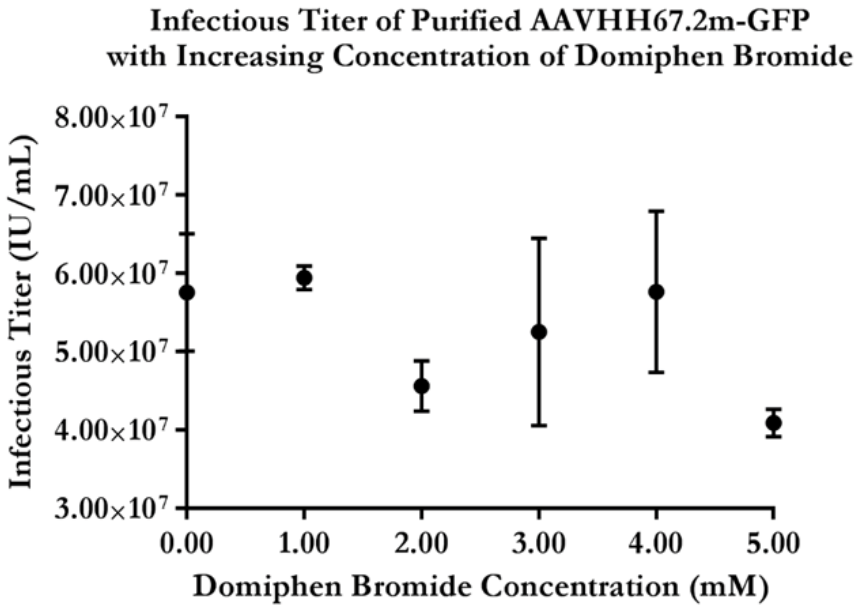


Figure 35: Infectious titer of from interactions between purified AAV with increasing concentrations of domiphen bromide

A)



B)

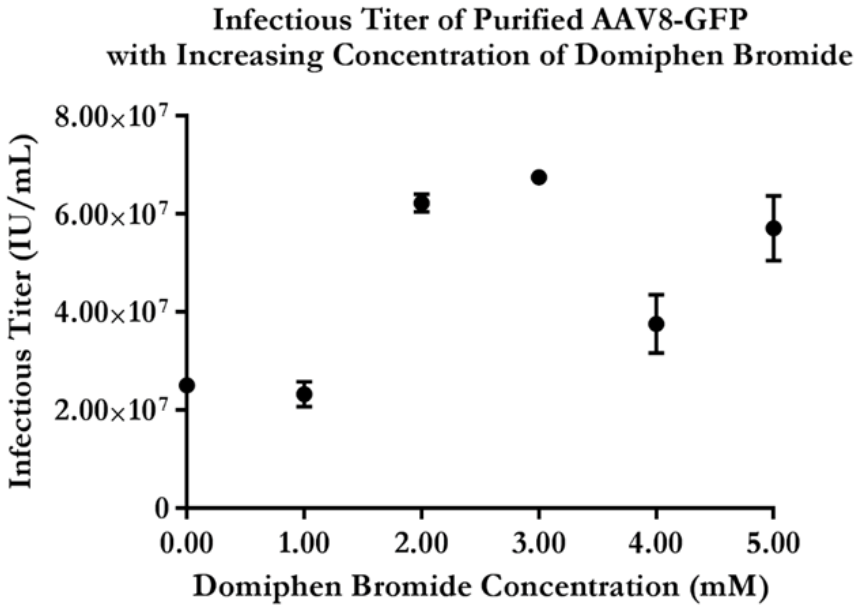


Figure 36: Infectious titer of from interactions between purified AAV with increasing concentrations of domiphen bromide

According to Figure 36, the infectious titers for both AAV2 and AAV9 remain about the same, supporting dot blot results. These affirms that there are no distinct interactions between a purified AAV9 or purified AAV2 with the detergent. However, as seen in previous results, when

there is a mixture of other contaminants in the preparation of the AAV, there seems to be an interaction dependent on the type of contaminant present. For AAV2, domiphen bromide interacts with the virus at higher concentrations, but if the additional adenovirus contaminant is concomitantly introduced, AAV2 does not interact with domiphen bromide. Alternatively, for AAV9, there is a slight interaction between the detergent with the additional DNA contaminant, but the interaction is exacerbated when the additional adenovirus is concomitantly present in the AAV9 preparation. Figure 37 shows that HH67.2m titers decrease, which somewhat supports dot blot results. It is unclear the exact interactions of HH67.2m and detergent when there is a concomitant presence of DNA, since this was not explored in the time of this project using triple plasmid transfection production schemes, nor in the presence of adenovirus, since no such producer cell line exists. Interestingly, we see that AAV8 shows a similar trend to those that appeared in dot blot. We see that there is an increase in AAV8 infectious titers as domiphen bromide is increasing in concentrations. Since these were coming from the same source, and no helper functions were provided during this purification analysis, we would not have an innate increase in vector. Therefore, we suggest that there is some sort of interaction going on with domiphen bromide and the virus as detergent increases. One idea is that the surface of the viral particle is interacting with a micellular or close to micellular domiphen bromide structure, but the virus is not necessarily within the resulting domiphen bromide micelle itself. Multiple viral particles can attach to this micelle much like fibers onto Velcro and when the aliquot is taken from the preparation of detergent and the virus, a non-homogenous distribution of virus is aliquoted, yielding a seemingly higher titer of virus. This would explain the drastic fluctuation in titers in both the dot blot assay and also the infection assay.

Since the amount of virus to analyze was very limited, we chose AAV8 as the only virus to see under transmission electron microscopy to see anything that is occurring to the structure

of the virus or if we are seeing removal of empty particles. Since we were limited to one replicate the results that appear in this set of images are not meant to be conclusive. These images are provided in Figure 38. As the detergent increased from 0 mM to 3 mM to 5 mM, there were curious physical features that showed in the images, most notably in 5 mM where these physical features are interacting with the viral particles. Without staining for these physical features, it is difficult to say what their composition is and what they are, but it is possible they are micellular structures that formed from domiphen bromide.

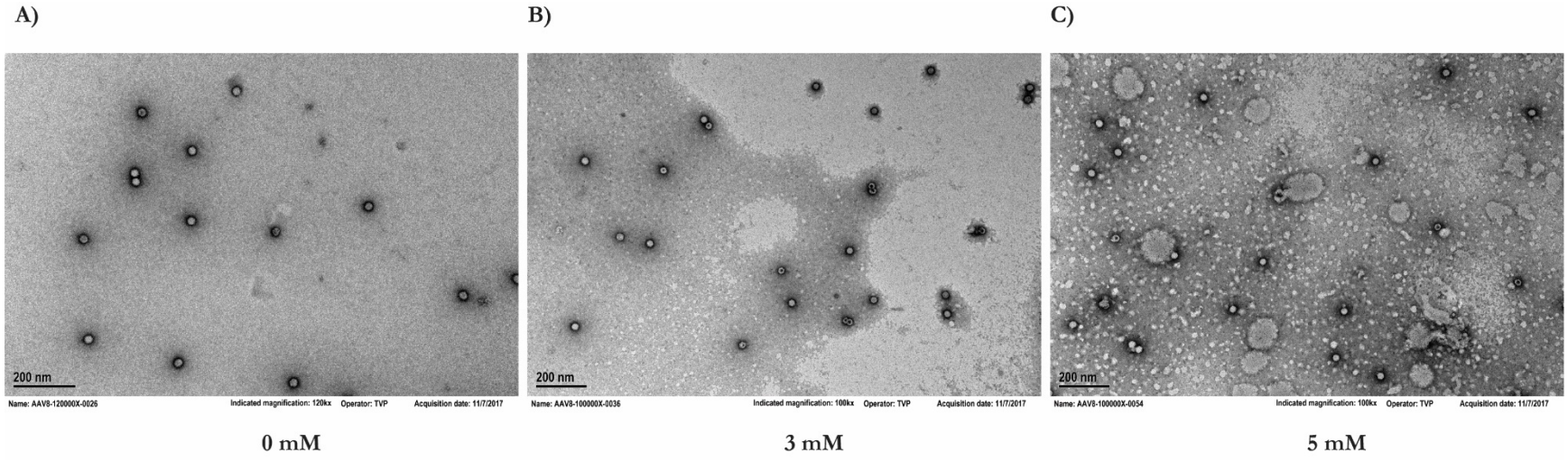


Figure 37: TEM Images of Various Concentrations of Domiphen Bromide in Purified AAV8 Samples

Table 11: TEM Analysis of AAV8 at Specific Detergent Concentrations

Domiphen Bromide Concentration (mM)	Full Particle Count	Empty Particle Count	Total Particles	Percentage of Full Particles	Percentage of Empty Particles
0	185	6	191	96.86%	3.14%
3	105	26	131	80.15%	19.85%
5	66	15	81	81.48%	18.52%

4.5 Discussion

The difficulty in removing contaminating materials is quite evident in the experiences of this aim. Heat inactivation can only prevent adequate removal for in vitro experiments, and not necessarily useful for larger scale in vivo work. Furthermore, use of an ultracentrifuge to remove adenovirus may seem like a safe choice, but even in this scenario adenoviral contaminants can sneak into the viral preparations, whether it be a component of the adenovirus, or the entire virus itself.

Our investigation for a novel method for removal of the contaminating adenovirus entities led us to the materials found in the Merck patent where they had spent significant efforts to investigate the ideal selective precipitating agent to use for their adenoviral production process. There work identified domiphen bromide to be used, and suggested briefly that cationic agents be used for selective precipitation of DNA based on charge or hydrophobicity. They theorized that "any cationic product (or anion with lower charge density than DNA) should remain free in solution when precipitating DNA with cationic detergents or polymers." The patent continued to say "since purification using cation exchange chromatography has been demonstrated […] cationic detergents could be used to precipitate both host cell DNA and the contaminating helper adenovirus," with the recommendation of domiphen bromide as the selective precipitating agent of choice. The rationalization of domiphen bromide was its versatility in removal of "any number of cellular components, especially nucleic acids, from any number of different type of biological products" and its "availability as a GMP grade raw material and current use in other products intended for human use."[12] The statements pointed out by this patent led us to first validate their adenovirus results and use the detergent for our production schemes that utilize the adenovirus for production of rAAV.

Our findings supported adenovirus and DNA removal, as shown in the Merck patents. When this was tested in AAV product generated from triple plasmid transfection, yielding mainly DNA contaminants and no adenoviral contaminants, domiphen bromide appeared to have some interaction with the AAV leading to a loss of product as domiphen bromide concentrations were increased. The loss was more apparent in an AAV9 production scheme, whereas the loss was less substantial in the AAV2 production scheme. When we transitioned to a production scheme more relevant to the realm of this dissertation work, namely an adenoviral product concomitantly generated with the rAAV product, we see more dramatic changes. For AAV2/Ad-Cre systems, we see that the AAV2 remains in solution and the adenovirus and DNA contaminants are precipitated and removed to up to 99% clarity. Conversely, in AAV9/Ad-Cre systems or AAV8/Ad-Cre systems, we unfortunately see the AAV products are lost at around the CMC of domiphen bromide of 1.2 mM, which coincides with the loss of DNA contaminants and the start of reduction in adenoviral contaminants. Attempts were made to try to reduce the extent of loss of these two products, but the only improvement that we investigated could in turn effect the potency of the desired viral product. This is because of the chaotropic properties of the guanidine hydrochloride that was used to potentially interfere with domiphen bromide and the AAV of a particular serotype.

To address if any natural interactions occurred between this detergent and the AAV, a simplified experiment was performed to observe any reduction of titer or infectious unit between domiphen bromide and the following AAV serotypes or variants in AAV preparations that had minimal contaminants: AAV2, AAV8, AAV9, and AAVHH67.2m. AAV2 and AAV9 show no decrease in titers, suggesting domiphen bromide does not interact with these viruses in the absence of contaminating entities. In the presence of contaminating entities AAV9 seems to tag-along with the precipitating materials and lost in solution. This could be due to an affinity of

AAV9 with the domiphen bromide-contaminant complex. For AAVHH67.2m, the results from a viral genome assessment and an infectious unit assessment were in conclusive. Infectious assays are generally more reliable, and in the case of the infectious assay performed in this experiment there may be an interaction occurring between the detergent and AAVHH67.2m due to the reduced infectious titers as detergent concentration increases. A striking interaction is observed in AAV8 serotypes. Here we see that we seemingly increase in titers as detergent concentrations increase, despite no helper functions present to initiate replication of the AAV8 in solution. We hypothesize what is occurring to describe the phenomena reproduced between vector genome assessment and infectious assays, is that the AAV8 has an affinity to domiphen bromide, but the affinity is not enough to precipitate the material from solution. Instead, we think a non-homogenous solution is present, where the AAV8-domiphen bromide complex are adhering to others of the same complex resulting in an apparent increased amount of vector in the same volume. The results from this interaction study for AAV8 supports results we see in the AAV8/Ad-Cre production system, in that with these aggregates that may form between AAV8 and domiphen bromide coupled with complexes forming between DNA-domiphen bromide and Ad-domiphen bromide, it would likely be more thermodynamically stable to come out of solution with other detergent complexes than remain in solution.

Based on the results obtained from this chapter, the use of domiphen bromide is likely restricted to use in AAV2 serotypes that are produced using an adenovirus based production system, while selectively purifying more clinically relevant serotypes made from an adenovirus based system remains a significant challenge. Therefore, for serotypes that are not AAV2, the best currently investigated methods to use for removal of adenovirus are chromatography, nanofilters, or high pressure. The use of a selective precipitating agent would be a great avenue to explore, but complications of its own removal before entering the final clinical grade product

are also of concern. In fact, in some studies of domiphen bromide toxicity in vitro on HEK293 cells, we found that the concentration of domiphen bromide that was tolerated by the cells was about 30 μM before cells started to die. This may be of importance for in vivo use and therefore it is important to consider removal of this agent to the final AAV product.

4.6 Conclusions

Although the systems presented in Chapters 2 and 3 are great for large scale production, the removal of adenovirus will have to be performed using another system such as nanofilters, high pressure environments, or chromatography if the serotype of AAV being generated is not AAV2. Given the advancements of chromatography for purification of a variety of AAV serotypes that are being used by companies like Pfizer, chromatography may be the most effective way of purifying a variety of AAV serotypes. Pfizer is using an affinity ligand made from a proprietary camelid-derived single-domain antibody fragment for AAV affinity purification and can purify a broad range adeno-associated virus used for gene therapy applications. However, if the solution is clarified of other crude lysate materials, the resulting viral products can be sent through this high-precision affinity chromatography to more precisely purify AAVs. The investigation in this chapter was intended to use the selective precipitating agent with hopes it can work on a multitude of serotypes, but seeing the great fluctuation that occurred in the 3 of the 4 serotypes tested, it appears this reagent is too troublesome to use in our current purification scheme, despite its ability to remove DNA and adenovirus with great efficiency in AAV2 producer cell line preparations. The adjusted process flow diagram for the thesis aims described in this dissertation are presented in Figure 39. Here, a black box was placed over the intended domiphen bromide purification scheme presented in Chapter 1 because the domiphen bromide system failed to be universally for all serotypes of AAV.

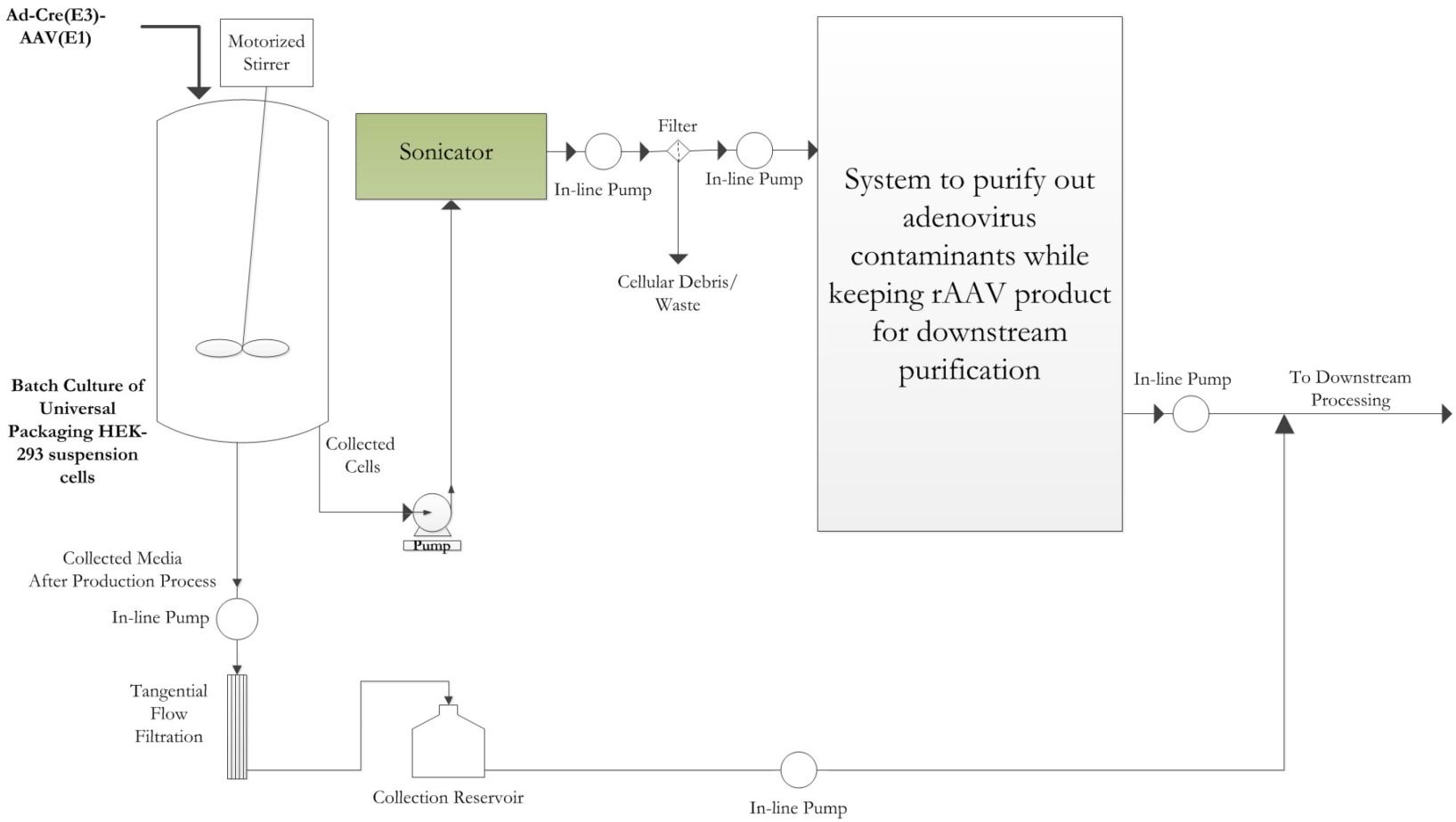


Figure 38: New process flow diagram. The process flow diagram removed the domiphen bromide part of the process since this selective precipitating agent may not be an effective tool for high-yielding rAAV processes.

REFERENCES

1 Volpers, C. & Kochanek, S. Adenoviral vectors for gene transfer and therapy. The journal of gene medicine 6 Suppl 1, S164-171 (2004).

2 Regulations, E. C. o. F. (Electronic Code of Federal Regulations, ecfr.gov, 2017).

3 Young, L. S. & Mautner, V. The promise and potential hazards of adenovirus gene therapy. Gut 48, 733-736 (2001).

4 Duigou, G. J. & Young, C. S. H. Replication-competent adenovirus formation in 293 cells: the recombination-based rate is influenced by structure and location of the transgene cassette and not increased by overproduction of HsRad51, Rad51-interacting, or E2F family proteins. Journal of virology 79, 5437-5444 (2005).

5 Fallaux, F. J., van der Eb, A. J. & Hoeben, R. C. Who's afraid of replication-competent adenoviruses? Gene therapy 6, 709-712 (1999).

6 Burova, E. & Ioffe, E. Chromatographic purification of recombinant adenoviral and adeno-associated viral vectors: methods and implications. Gene therapy 12, S5-S17 (2005).

7 Maheshwari, G., Jannat, R., McCormick, L. & Hsu, D. Thermal inactivation of adenovirus type 5. Journal of Virological Methods 118, 141-145 (2004).

8 Drittanti, L. et al. Optimised helper virus-free production of high-quality adeno-associated virus vectors. The Journal of Gene Medicine 3, 59-71 (2001).

9 Goerke, A. R., To, B. C. S., Lee, A. L., Sagar, S. L. & Konz, J. O. Development of a Novel Adenovirus Purification Process Utilizing Selective Precipitation of Cellular DNA. Biotechnology and Bioengineering 91, 12-21 (2005).

10 Konz, J., John O. & Rahway, N. Methods of Adenovirus Purification. EP1506287B1 (2007).

11 De Vocht, M. L. & Veenstra, M. Process for Adenovirus Purification from High Cell Density Cultures. US8470585B2 (2013).

12 Konz, J. O., Lee A. L, To, B. C. S., & Goerke, A. R. Methods of Adenovirus Purification. US7326555B2 (2008).

CHAPTER 5: FUTURE DIRECTIONS – CELL LINE ADJUSTMENTS FOR INCREASED PROLIFERATION

As we are investigating new production methods to meet the demand of rAAV necessary for clinical trials and market, rAAV products are gaining success with great rapidity. It is imperative that our technologies should move towards industrialization, elsewise the cost of the therapeutic will continue to rise as the indication veers away from localized administration. In this chapter, I will discuss the path our technology should take based on the aims detailed earlier to increase vector yields in less time.

One of the major issues in large-scale production in industry is the proliferation rate of cells as they grow from small bioreactors to the eventual bioreactors reaching a 10,000 L capacity. Modifications to generate a "super cell" to grow as fast as possible without undergoing apoptosis are constantly being investigated by modulating cell growth conditions, improving cellular metabolism, induction of oncogenes for cell proliferation, induction of anti-apoptotic signals, as well as arrest of the cell cycle with intentions to increase cell-specific productivity. Unfortunately, outcomes from these strategies have been mixed, with few instances where improvements in product yield have been achieved[1]. Furthermore, cell-proliferation enhancement has previously focused on overexpression of exogenously introduced oncogenes, transcription enhancers, and cyclin-dependent kinases with success to increase proliferation and viable cell densities. For the purposes of large-scale rAAV production using adenovirus infection into a packaging HEK-293 cell line, we brainstormed to manipulate the cell cycle such that it is additive to the cell cycle deregulation mechanisms of both adenoviral E1A and E1B genes already integrated in the HEK-293 genome. The manipulation would involve disruption or

knock-out of signaling pathways governing cell proliferation, instead of overexpression of an exogenously introduced cell proliferation enhancer.

The mammalian cell cycle can be divided into five phases: G0-G1-S-G2-M. G0, G1, and G2 are considered gap phases: G0 is the resting phase where cells have zero growth but are metabolically active, G1 is the first period of growth where cells are stimulated by growth factors to enter the cell cycle, and G2 is the second period of growth after the cell has duplicated its chromosomes and is prepared to divide into two daughter cells during the period of mitosis. The S phase is the DNA synthesis phase where DNA replication occurs and M phase is the mitosis phase where the cell segregates its duplicated chromosomes and prepares for cytokinesis which is cell division. The G1 phase is where control of cell proliferation generally occurs, since it is here that a decision to enter S phase from G1 constitutes a "point of no return," and the cells are committed to complete the cell cycle and divide barring any DNA damage[2]. This decision point, called the restriction point or "R" point, is tightly controlled primarily by the retinoblastoma pathway (RB) and its family of retinoblastoma proteins (pRB). The tumor suppressive properties of the pRB family inhibit transcription factors, such as E2F target-genes, that activate necessary genes for G1/S transition[3].

To appreciate the role of adenovirus in the cell-cycle, it is important to understand the biology of the wild-type adenovirus. Only a portion of the adenovirus genome directly impacts cell cycle progression and apoptosis prevention, namely E1A, E1B, and E4.

The early gene E1A encodes for the E1A proteins that activates the other early genes of adenovirus and induce the cell to enter S phase in order to create an environment optimal for virus replication. E1A proteins target multiple areas of the cell cycle to allow for G1/S transition, including sequestering of RB and the modulation of other pRB family proteins p107 and p130. E1A can directly bind and inhibit components of the CKI p21, as well as several host

factors involved in mediating chromatin structure that is thought to increase the accessibility of DNA to the transcriptional machinery[4].

The increased deregulation of the cell cycle by E1A results in the accumulation of the tumor suppressor p53, which if activated can result in the cell to undergo apoptosis. In fact, under conditions of mitogen deficiency in mouse fibroblasts, E1A can actually promote p53-mediated apoptosis by stabilizing p53 through the involvement of p19ARF (known as p14ARF in humans) which inhibits the activity of a negative regulator of p53 (discussed later)[5]. The presence of adenovirus E1B-55K protein acts to block p53-dependent apoptosis by directly binding to p53, by this means inhibiting its ability to induce expression of proapoptotic genes[4]. The E4orf6, from the E4 region of adenovirus, cooperates with E1B-55K to inhibit p53. The second product of E1B, E1B-19K protein, blocks downstream mediators of tumor necrosis factor α- (TNF-α) and TRAIL (TNF-related apoptosis-inducing ligand)-mediated death receptor pathways. This is done by directly binding to the proapoptotic proteins Bak and Bax to prevent mitochondria-mediated apoptosis[4].

Adenovirus E4 genes, and its resulting proteins, play an important supporting role to E1A and E1B proteins, as well as a p53-independent apoptosis that is speculated to facilitate the release of progeny virions during the late stages of infection[4]. Indeed, when the anti-apoptotic gene Bcl2 was overexpressed in HEK-293 cultures, in hopes to increase productivity by prolonging cell survival, the titer of a GFP expressing adenovirus decreased. Interestingly, when the apoptotic gene for pro-caspase-3 was overexpressed, the cell survival was expectedly reduced, but there was an increase in adenoviral titer[6]. This shows the importance of apoptosis to adenovirus production, and justification to avoid apoptotic signaling pathways for disruption or knock-out as this could greatly affect adenovirus and adeno-associated virus maturation. Furthermore, E4orf6 has been shown to inhibit the binding of p53 to cellular transcription

factors, and, together with E1B-55K, can target p53 for degradation by ubiquitin ligases, lending insight that apoptosis signaling from E4 genes is most likely not p53-mediated[4]. Finally, the E4 product, E4orf6/7 protein, binds directly to the transcription factor E2F and promotes its activation of viral and cellular promoters that are utilized in adenoviral gene expression and viral assembly[4].

The cell can progress to S phase upon inhibitory phosphorylation of the pRB proteins via three principle cyclin-dependent kinases (CDKs): cyclin-D-dependent CDK4 and CDK6 and cyclin-E-dependent CDK2. These kinases form complexes with cyclins to promote phosphorylation in as many as 16 sites in pRB proteins - CDK4 and CDK6 form complexes with cyclin D1, D2, and D3 and CDK2 forms complexes with cyclins E1 and E2[3,7].

CDK inhibitors (CKIs) can prevent phosphorylation of the pRB proteins, causing cell cycle arrest in G1 phase in response to a variety of stimuli such as DNA damage, senescence, and cellular stress[3]. These CKIs can be divided into two major families: the CIP/KIP and INK4 families. The CIP/KIP family includes p21 (also known as CDKN1A, CIP1, or WAF1), p27 (also known as CDKN1B or KIP1), and p57 (also known as CDKN1C or KIP2). These CKIs inhibits the activity of multiple cyclin-CDK complexes by contacting both subunits and blocking kinase activity and substrate binding. The INK4 family is divided into four CKIs: p16 (also known as p16INK4A or INK4A), p15 (also known as p15INK4B or INK4B), p18 (also known as p18INK4C), and p19 (also known as p19INK4D, and not to be confused with the rep gene p19 promoter). These CKIs specifically bind to the catalytic subunits of two CDKs, CDK4 and CDK6, thereby preventing cyclin-CDK complex formation[3].

The protein levels of CDKs remain relatively constant during the cell cycle and quiescence, so the induction of cell cycle arrest occurs in regulation of CDK activity by controlling abundance of their cyclin partners and also CKIs[3]. Furthermore, regulation of

cyclins and CKIs is mainly through transcriptional regulation and ubiquitin-dependent proteolysis, even though they can be also regulated at the level of mRNA, translation, and subcellular localization. The reason for this complexity of the pRB pathway is because the cell needs to adjust quickly to positive and negative signals before it can commit to replicating an exact stable genomic copy of itself in S phase[3]. Although disruption of these mechanisms is an important explanation for the understanding of diseases like cancer, the redundancy of G1 cell proliferation control is superfluous for an industrial scale cell culture where large quantities of cells and product are needed to meet intense demand. Since the HEK-293 cell is of human origin and harbors the E1A/E1B genes of the adenovirus, mammalian cell proliferation inhibitors can be disrupted in an additive manner to the E1A/E1B genes.

With the substantial impact of adenoviral early genes to the cell cycle, only a few CKIs remain for disruption in hopes to promote cell proliferation. These are: p16INK4A, p15INK4B, p18INK4C, p19INK4D, p27, and p57. It has been reported that there is an intrinsic cooperation between INK4 families and CIP/KIP families to inhibit proliferation and induce cellular senescence. Such cooperation exists for p16INK4A/p21, p19INK4D/p27, p18INK4C/ p21, and also p18INK4C/ p27[8-10]. Individual knock-outs of p16INK4A, p18INK4C, and p19INK4D may not result in major proliferation rate changes, at least in murine models, unless the CIP/KIP family protein it cooperates with is also simultaneously disrupted or knocked out[10]. Genetic alterations of p18INK4C are rare in human tumors, although gene silencing by promoter methylation has been observed in Hodgkin's lymphoma and medulloblastoma[11]. Therefore, relying on p18INK4C knock-out studies in animal models - particularly murine models – data has shown p18INK4C to be a haploinsufficient tumor suppressor, i.e. can induce tumor suppression if both alleles are functional and carcinogen-induced tumorigenesis can occur if mutated in one or both alleles of the p18INK4C gene[12]. Furthermore, tumorigenesis can be

induced if there was a simultaneous disruption or knock-out of p27 and p18INK4C, suggesting the cooperation between the two proteins for tumor suppression[9]. In murine models, p19INK4D do not develop tumors or other types of proliferative disorders[11].

Complete knock-out of the CIP/KIP family may provide mixed results for cell proliferation even though these are bona fide inhibitors of cyclin-CDK complexes. p27 is a critical component to the cell cycle, especially in G1 phase. p27 knock-out would result in potential issues in assembly and nuclear import of D-type cyclin-CDK complexes[13]. The least studied CIP/KIP protein, p57, may not be as critical to matured, fully differentiated cell proliferation as its fellow family members, p21 and p27. Although this protein does have capabilities to suppress several G1 cyclin-CDK complexes, the importance of the protein lies in development of organisms such as the control of organogenesis, embryogenesis, and cell differentiation. p57 is also reported to be necessary for endoreduplication as it can suppress CDK1 activity, a CDK critical for cytokinesis[14]. Several reports have stated a paradoxical role of p21 and p27 in addition to their role as a CKI, where this family of proteins may even promote cell cycle progression by facilitating cyclin-D association to CDK4/6 and also activation of cyclin-E-CDK2 complexes[15]. Furthermore, if p27 or p21 are generally in the cytoplasm of the cell, unable to interact with its cyclin-CDK complex targets in the nucleus, it may be an indicator of tumor generation[16-18]. The E1A gene may play a role to inhibit the binding of the other CIP/KIP proteins, p27 and p57, to bind to its cyclin-CDK complexes and therefore progress into S phase and prevent growth arrest, but this is not fully elucidated in all cell types[19]. These multifaceted and more importantly, paradoxical roles of the CIP/KIP family of proteins dissuaded our use for CRISPR/Cas9 targeting in our initial investigates, especially since the purpose of our genome editing was not for mechanistic understanding, but rather for

reproducible integration of a packaging genome with secondary ambitions of increased proliferation of the HEK-293 cells. However, to meet the need for rAAV supply demands, perhaps it is a necessary avenue to investigate since faster growth rates can reduce production times and material costs.

Other avenues that can be done to achieve this fast-growing producer or packaging cell line goal is to generate an E1 harboring cell line for a cell line that is known to be fast growing and commonly being used in industrial settings for biologics, namely Chinese Hamster Ovary (CHO) cells or Baby Hamster Kidney (BHK) cells. Once a cell line stably integrates the E1 gene, the methods described in Chapter 3 can be performed to create a fast-growing packaging cell line that can use methods described in Chapter 2 to generate a rAAV product in suspension culture.

REFERENCES

1 Dietmair, S., Nielsen, L. K. & Timmins, N. E. Engineering a mammalian super producer. J Chem Technol Biotechnol 86, 905-914 (2011).

2 Rhind, N. & Russell, P. Signaling Pathways that Regulate Cell Division. Cold Spring Harbor Perspectives in Biology 4 (2012).

3 Duronio, R. J. & Xiong, Y. Signaling Pathways that Control Cell Proliferation. Cold Spring Harbor Perspectives in Biology 5 (2013).

4 McConnell, M. J. & Imperiale, M. J. Biology of Adenovirus and Its Use as a Vector for Gene Therapy. Human Gene Therapy 15, 1022-1033 (2004).

5 Chattopadhyay, D., Ghosh, M. K., Mal, A. & Harter, M. L. Inactivation of p21 by E1A Leads to the Induction of Apoptosis in DNA-Damaged Cells. Journal of Virology 75, 9844-9856 (2001).

6 Sandhu, K. S. & Al-Rubeai, M. The Effect of Bcl-2, YAMA, and XIAP Over-Expression on Apoptosis and Adenovirus Production in HEK293 Cell Line. Biotechnology and Bioengineering 104, 752-765 (2009).

7 Berridge, M. J. Cell Cycle and Proliferation. Cell Signalling Biology (2012). <http://www.biochemj.org/csb/>.

8 Takeuchi, S., Takahashi, A. & Motoi, N. Intrinsic cooperation between p16INK4a and p21Waf1/Cip1 in the onset of cellular senescence and tumor suppression in vivo. Cancer Research 70, 9381-9390 (2010).

9 Franklin, D. S. et al. CDK inhibitors p18INK4c and p27Kip1 mediate two separate pathways to collaboratively suppress pituitary tumorigenesis. Genes & Development 12, 2899-2911 (1998).

10 Ortega, S., Malumbres, M. & Barbacid, M. Cyclin D-dependent kinases, INK4 inhibitors and cancer. Biochimica et Biophysica Acta 1602, 73-87 (2002).

11 Canepa, E. T. et al. INK4 Proteins, a Family of Mammalian CDK Inhibitors with Novel Biological Functions. IUBMB Life 59, 419-426 (2007).

12 Bai, F., Pei, X.-H., Godfrey, V. L. & Xiong, Y. Haploinsufficiency of p18INK4c Sensitizes Mice to Carcinogen-Induced Tumorigenesis. Molecular and Cellular Biology 23, 1269-1277 (2003).

13 Wander, S. A., Zhao, D. & Slingerland, J. M. p27: a barometer of signaling deregulation and potential predictor of response to targeted therapies. Clinical Cancer Research 17, 12-18 (2011).

14 Borriello, A. et al. p57Kip2 and cancer: time for a critical appraisal. Molecular Cancer Research 9 (2011).

15 Abukhdeir, A. M. & Park, B. H. p21 and p27: roles in carcinogenesis and drug resistance. Expert Reviews in Molecular Medicine 10 (2008).

16 Xia, W. et al. Phosphorylation/Cytoplasmic Localization of p21Cip1/WAF1 Is Associated with HER2/neu Overexpression and Provides a Novel Combination Predictor for Poor Prognosis in Breast Cancer Patients. Clinical Cancer Research 10, 3815-3824 (2004).

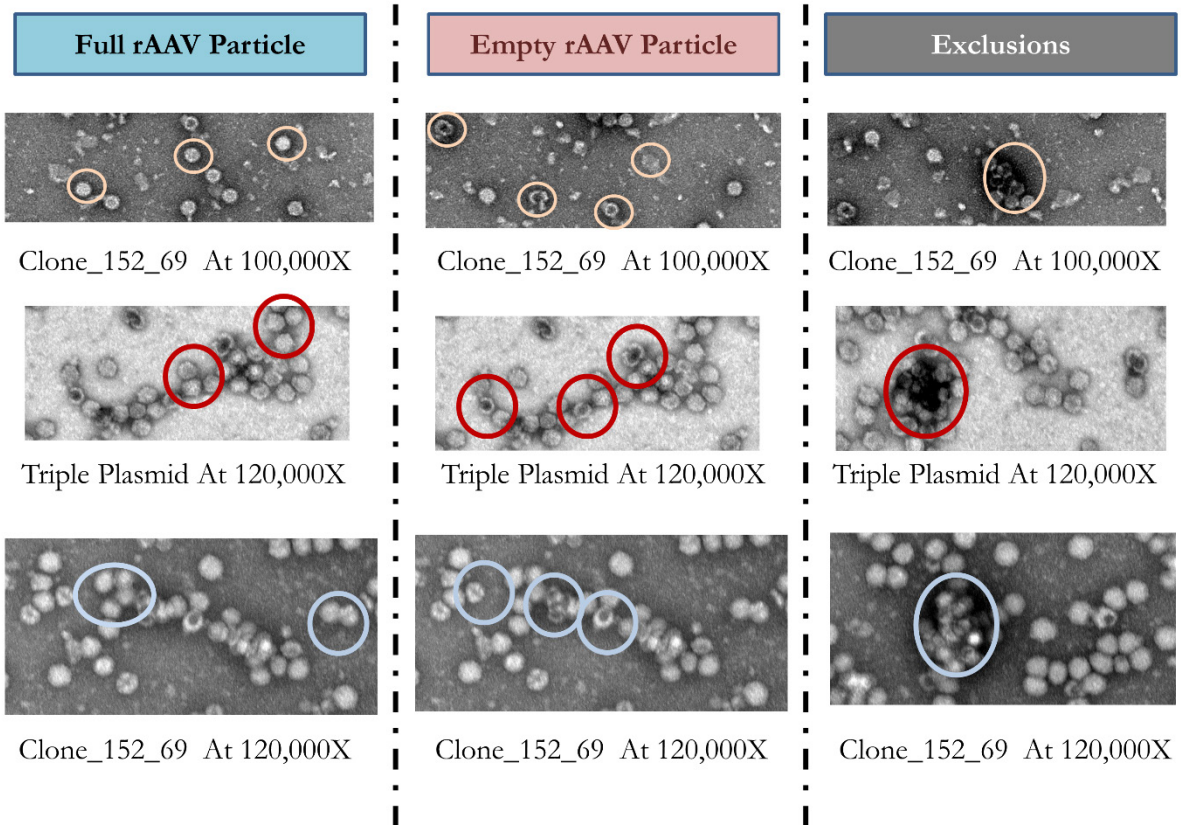
17 Porter, P. L. et al. Expression of cell-cycle regulators p27Kip1 and cyclin E, alone and in combination, correlate with survival in young breast cancer patients. Nature Medicine 3, 222-225 (1997).

18 Bachman, K. E. et al. p21(WAF1/CIP1) mediates the growth response to TGF-beta in human epithelial cells. Cancer Biology & Therapy 3, 221-225 (2004).

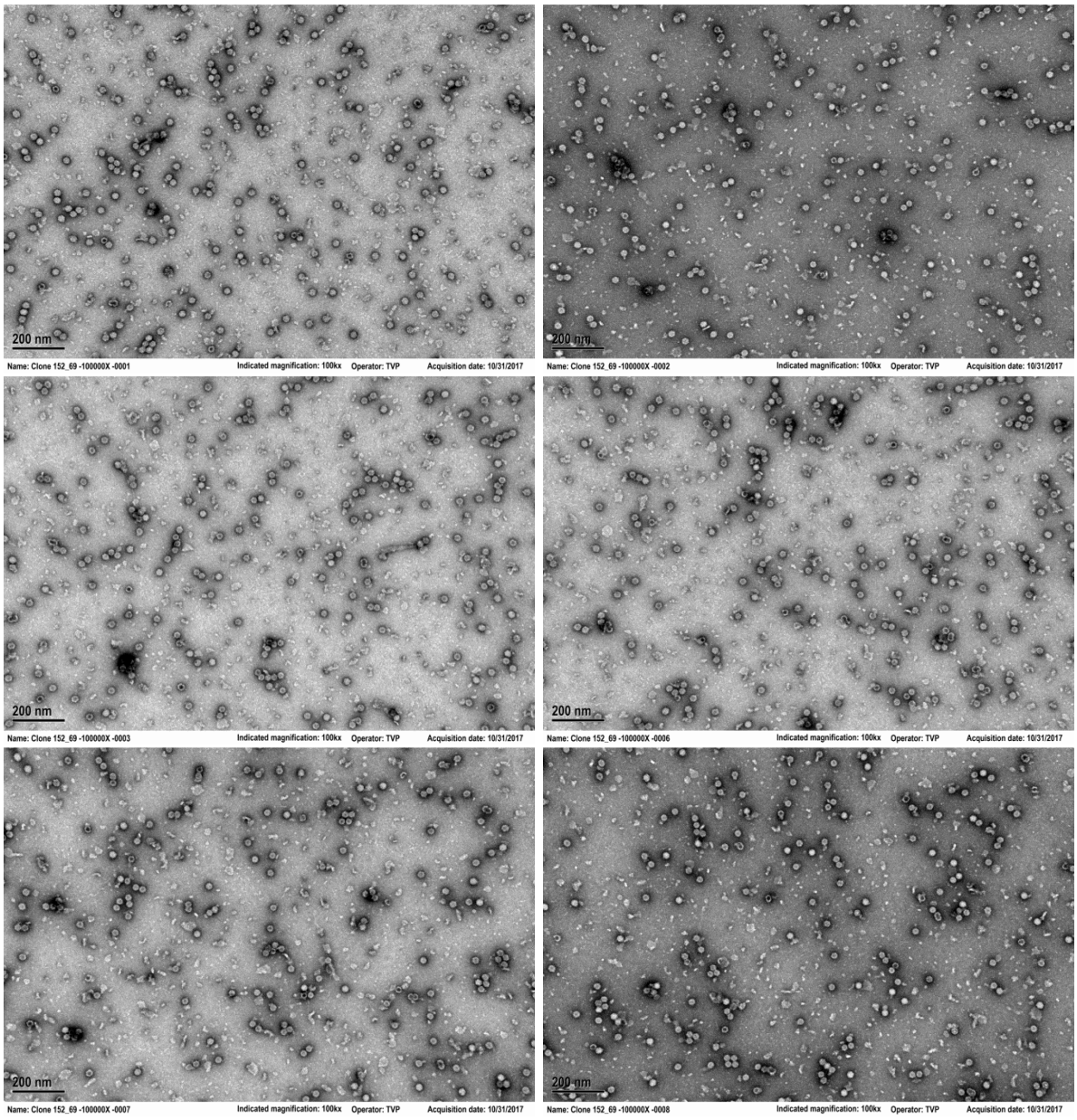
19 Frisch, S. M. & Mymryk, J. S. Adenovirus-5 E1A: Paradox and Paradigm. Nature Reviews 3, 441-452 (2002).

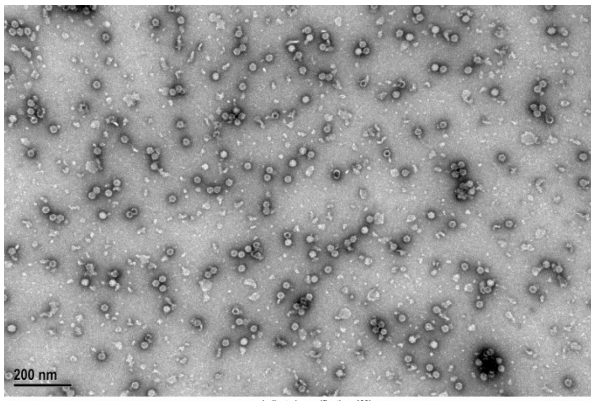
APPENDIX A – EXCLUSION CRITERIA AND RAW TRANSMISSION ELECTRON IMAGES

OBSERVATION CRITERIA

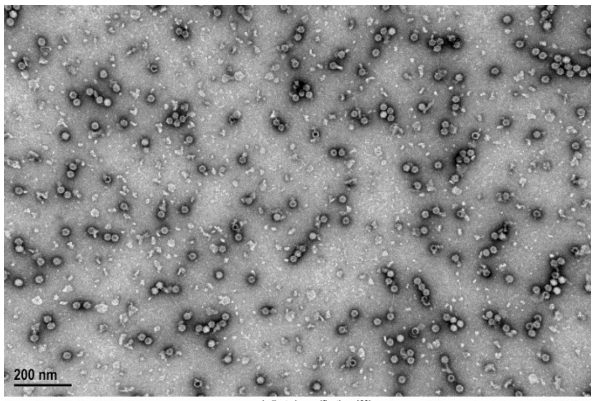


XX2-in-Puro-AAVS1 Clone 152.69 infected with Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1)

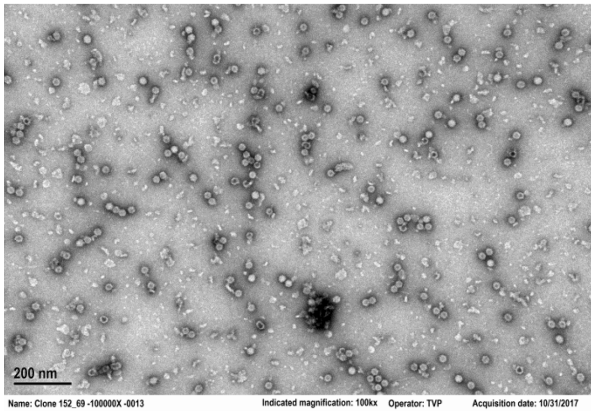




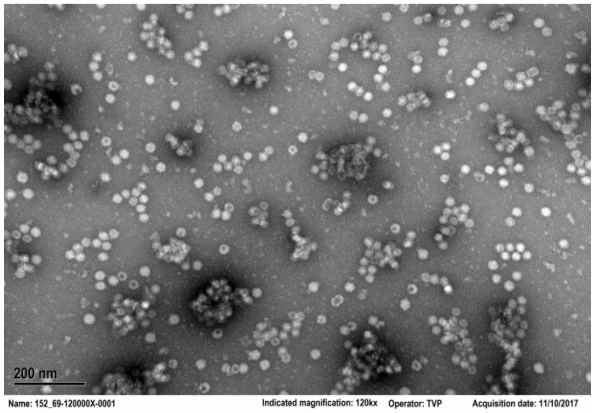
Name: Clone 152_69 -100000X -0010 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017



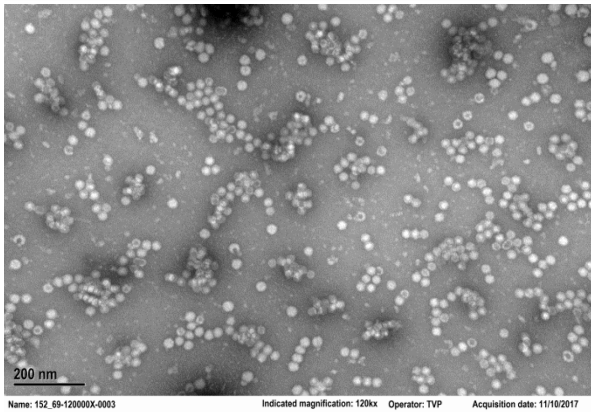
Name: Clone 152_69 -100000X -0012 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017



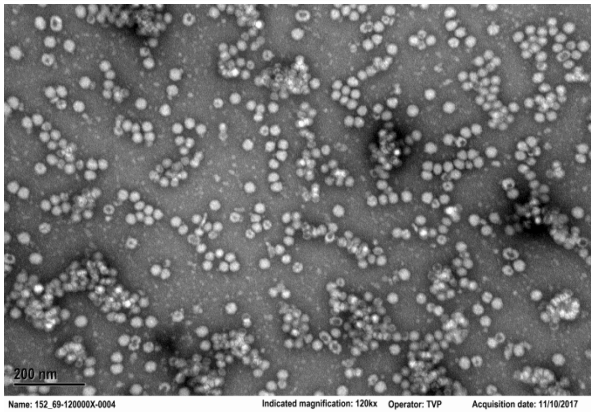
Name: Clone 152_69 -100000X -0013 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017



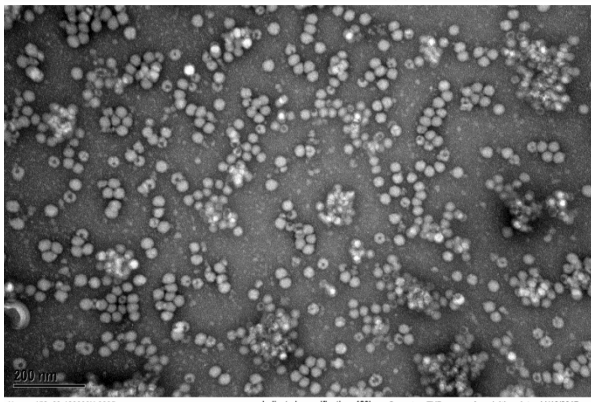
Name: 152_69-120000X-0001 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



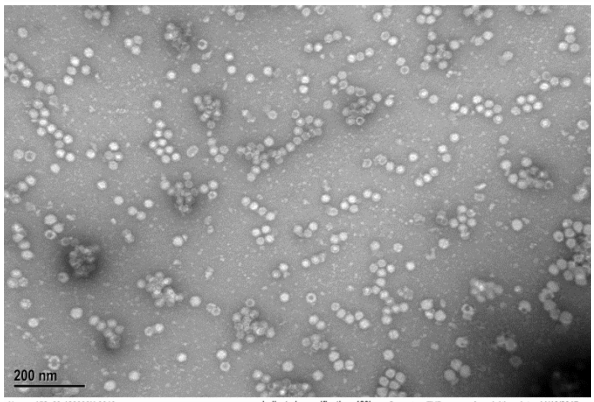
Name: 152_69-120000X-0003 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



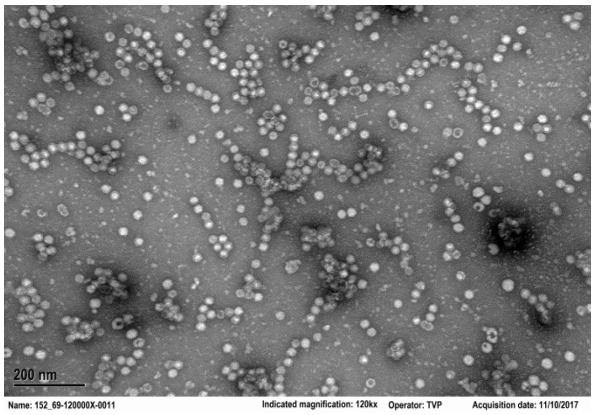
Name: 152_69-120000X-0004 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

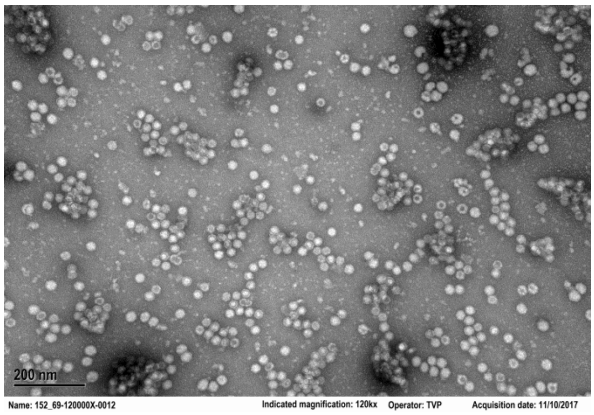


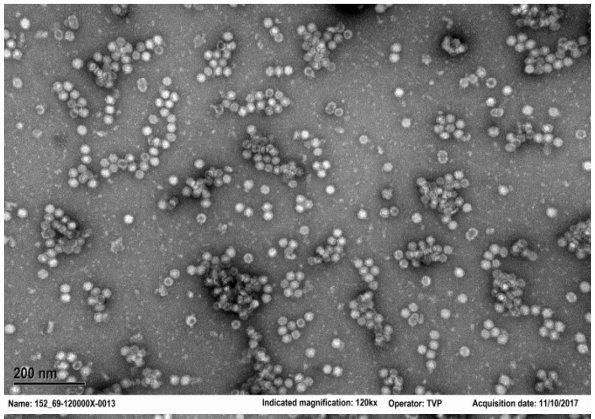
Name: 152_69-120000X-0005 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

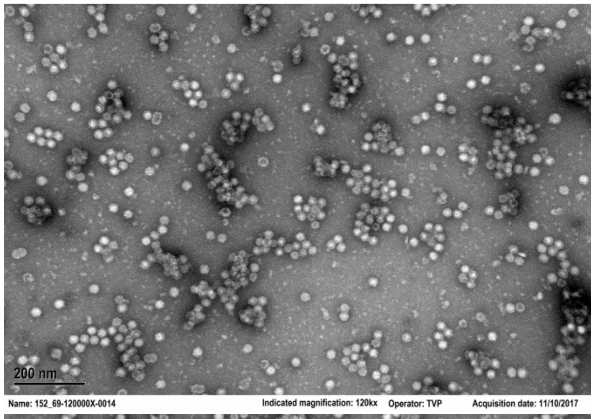


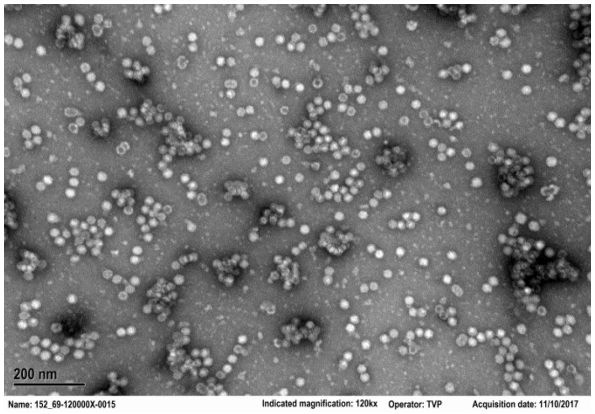
Name: 152_69-120000X-0010 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

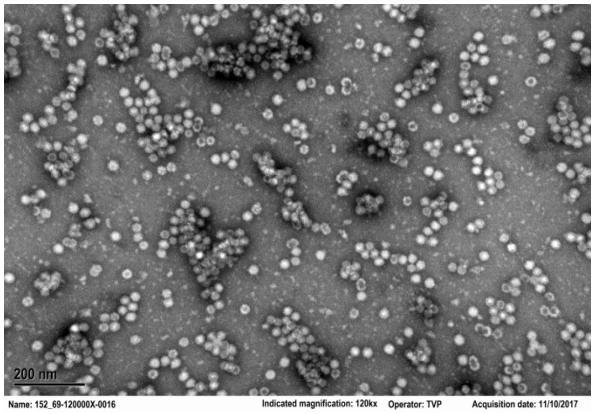

200 nm
Name: 152_69-120000X-0011 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017


200 nm
Name: 152_69-120000X-0012 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

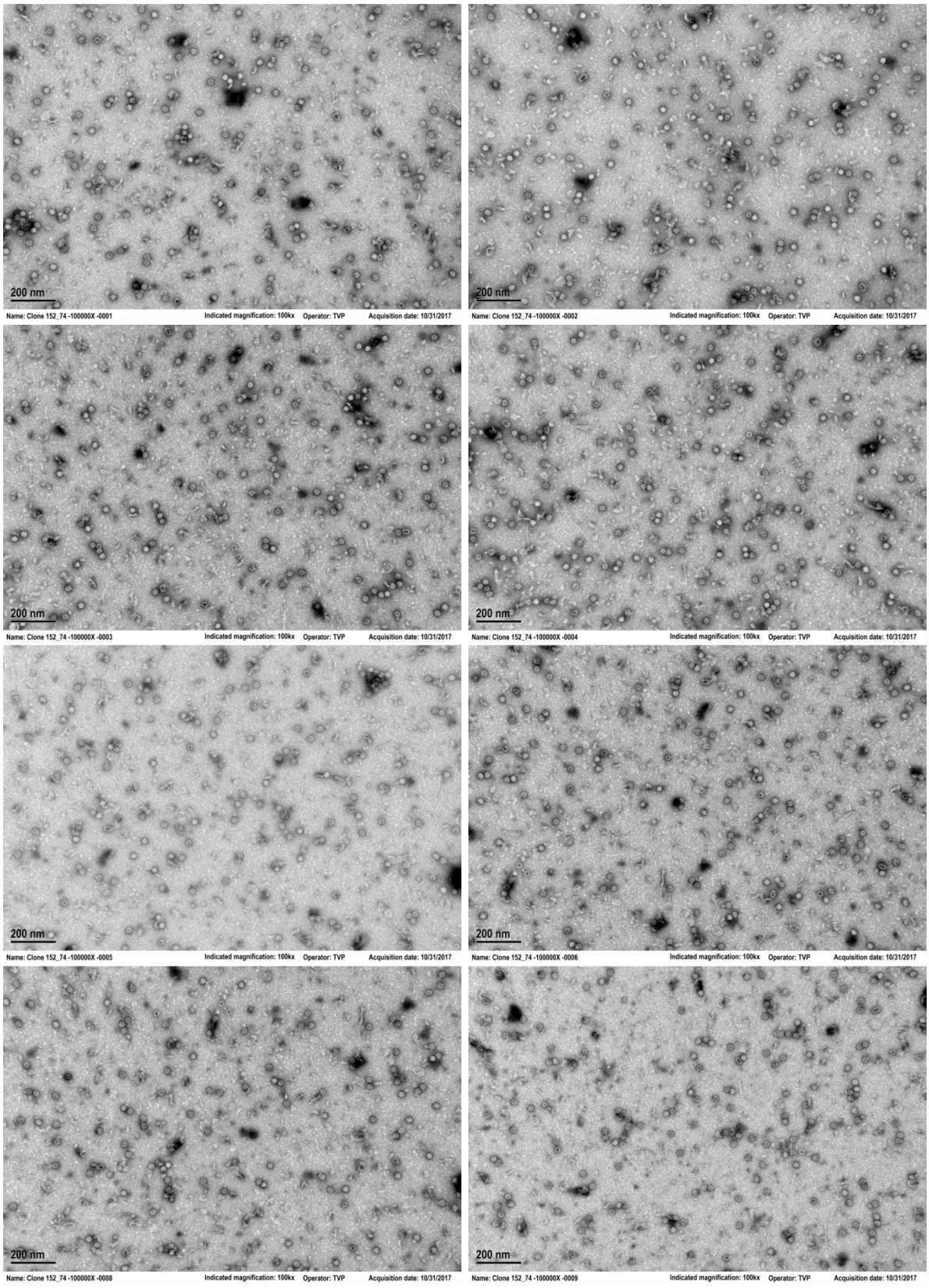

200 nm
Name: 152_69-120000X-0013 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

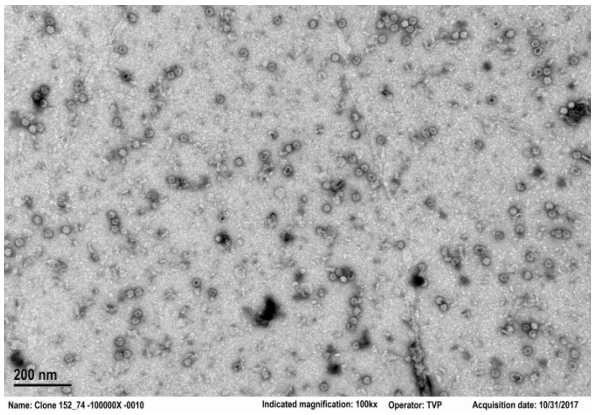

200 nm
Name: 152_69-120000X-0014 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

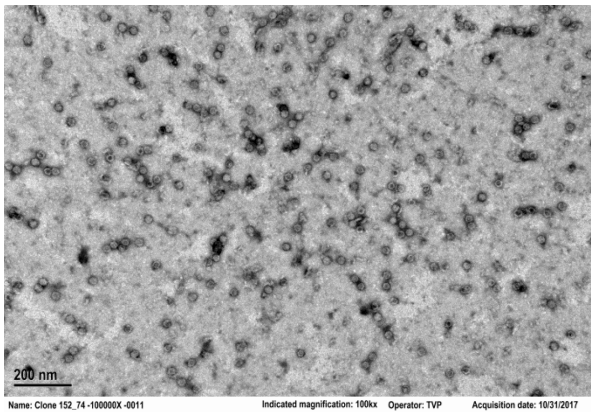

200 nm
Name: 152_69-120000X-0015 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

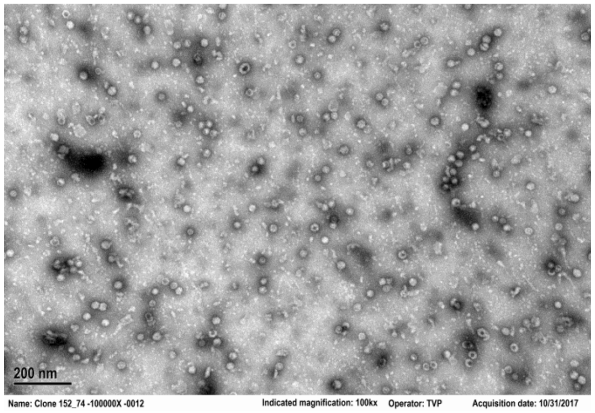

200 nm
Name: 152_69-120000X-0016 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

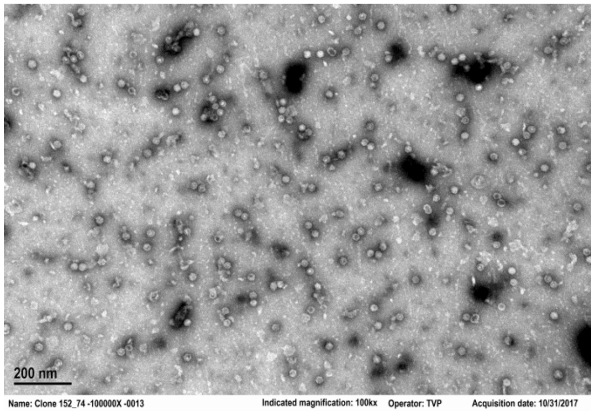
XX2-in-Puro-AAVS1 Clone 152.74 infected with Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1)

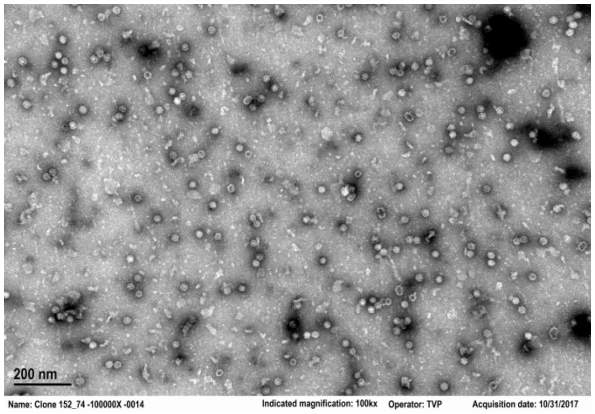


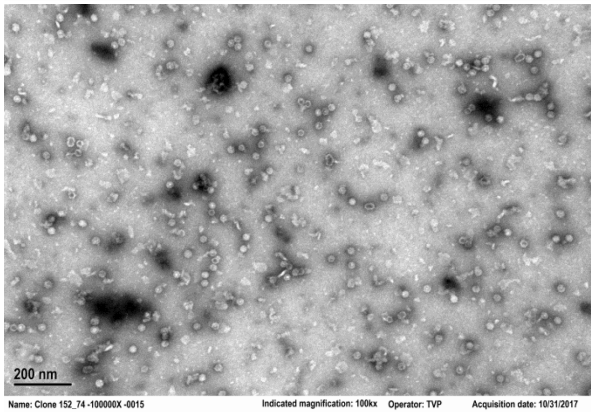

200 nm
Name: Clone 152_74 -100000X -0010 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017

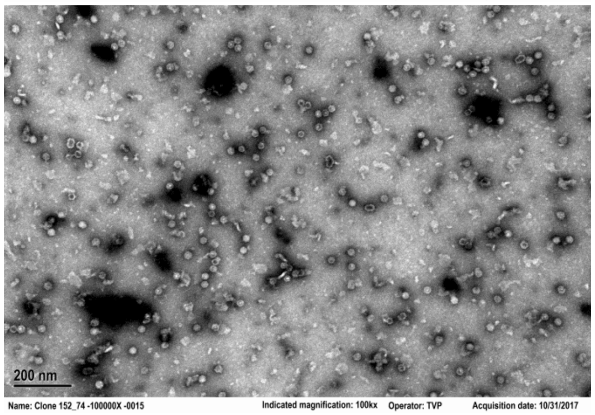

200 nm
Name: Clone 152_74 -100000X -0011 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017

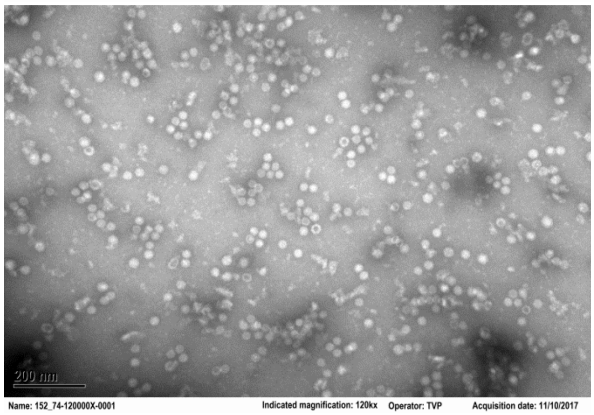

200 nm
Name: Clone 152_74 -100000X -0012 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017

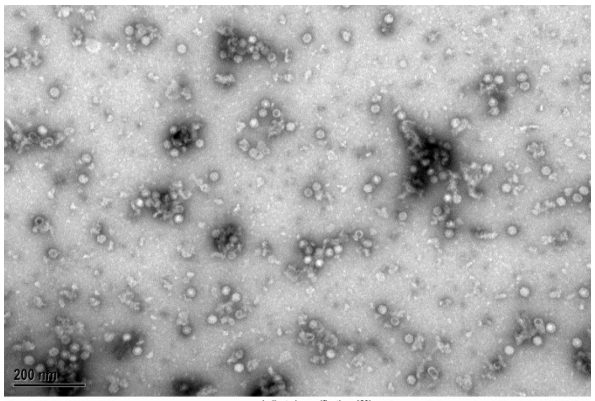

200 nm
Name: Clone 152_74 -100000X -0013 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017


200 nm
Name: Clone 152_74 -100000X -0014 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017

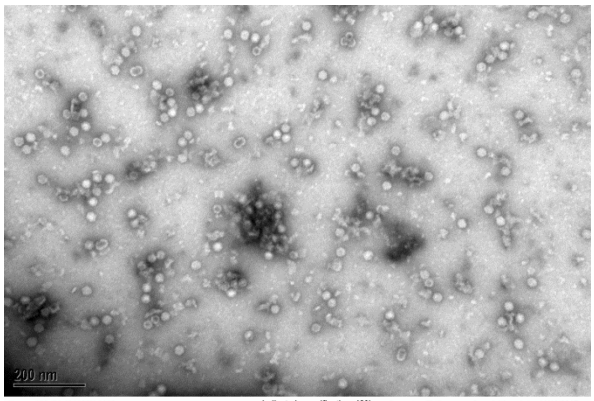

200 nm
Name: Clone 152_74 -100000X -0015 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017


200 nm
Name: Clone 152_74 -100000X -0015 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017

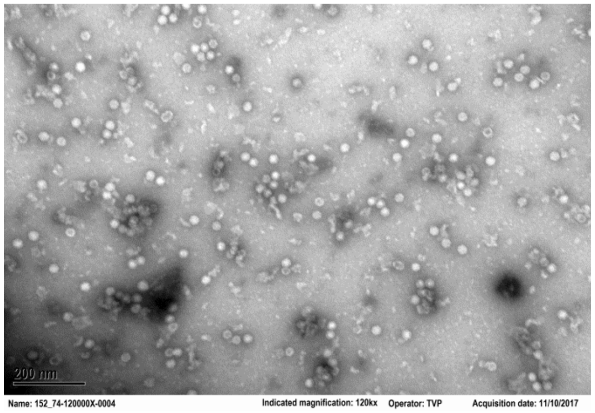

200 nm
Name: 152_74-120000X-0001 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



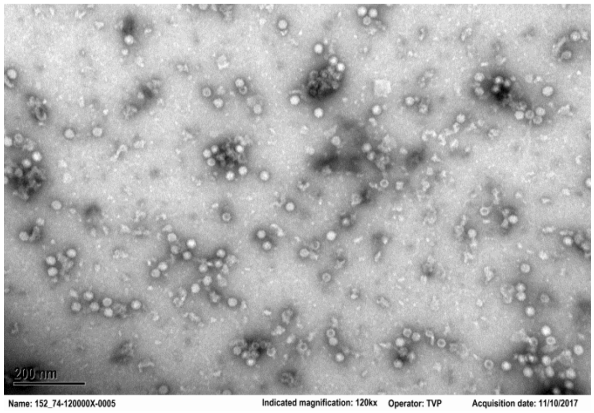
Name: 152_74-120000X-0002 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



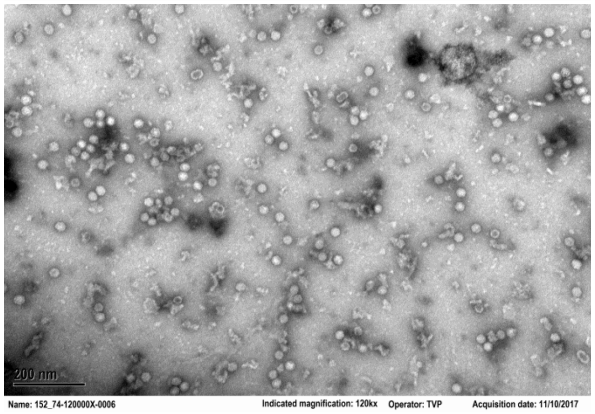
Name: 152_74-120000X-0003 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



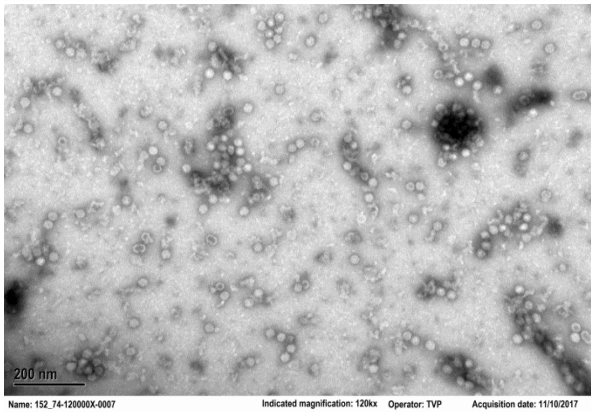
Name: 152_74-120000X-0004 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



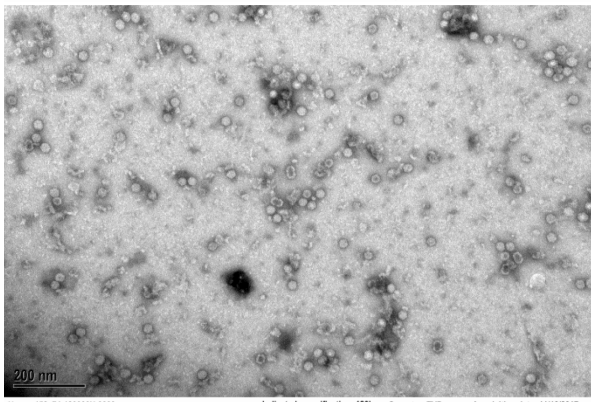
Name: 152_74-120000X-0005 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



Name: 152_74-120000X-0006 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

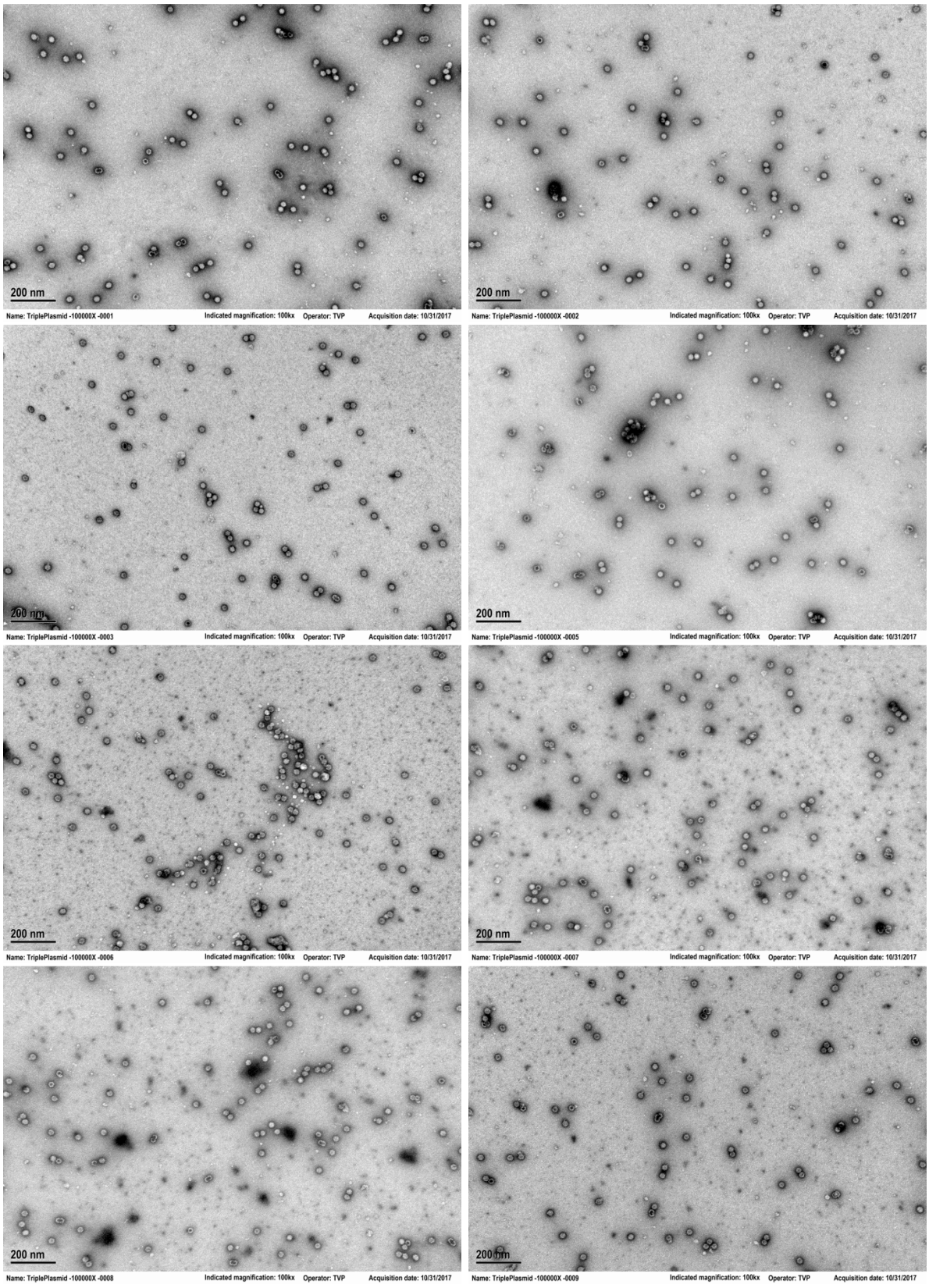


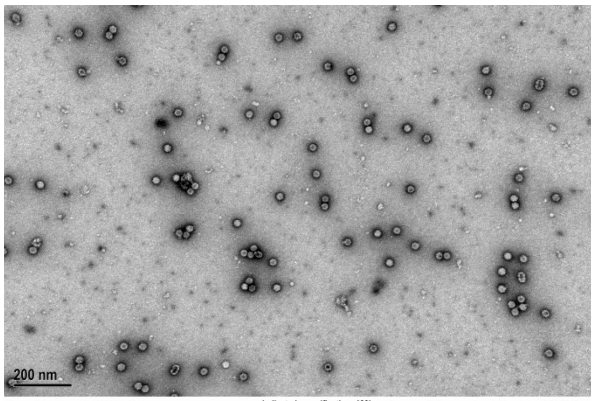
Name: 152_74-120000X-0007 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

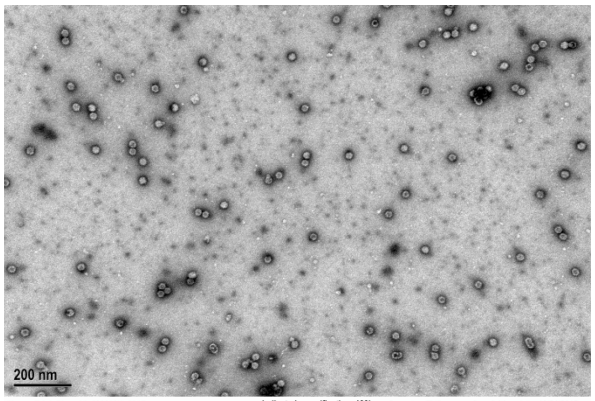


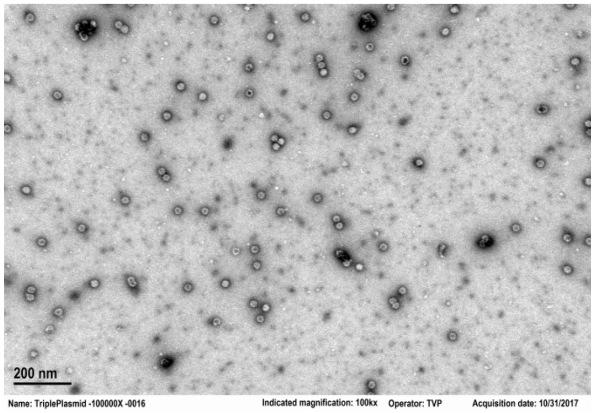
Name: 152_74-120000X-0009 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

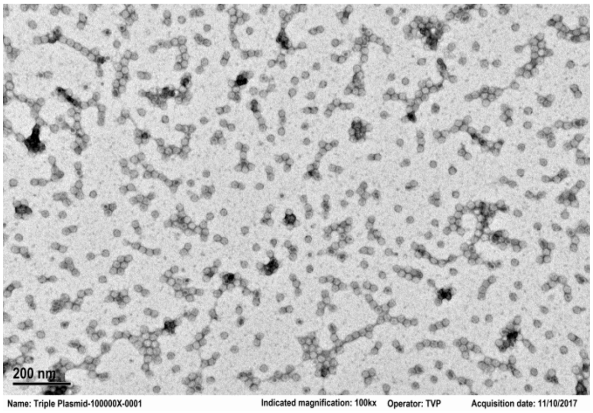
Triple plasmid transfection for AAV2-CMV-LacZ-nLs

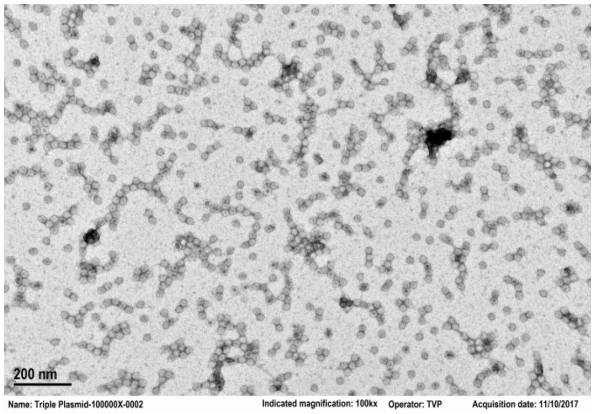


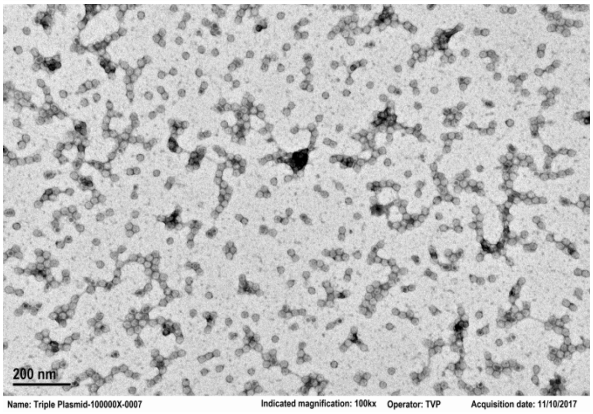

200 nm
Name: TriplePlasmid -100000X -0011 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017

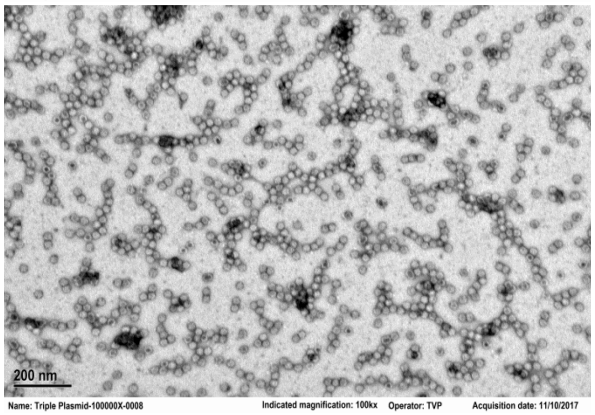

200 nm
Name: TriplePlasmid -100000X -0015 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017

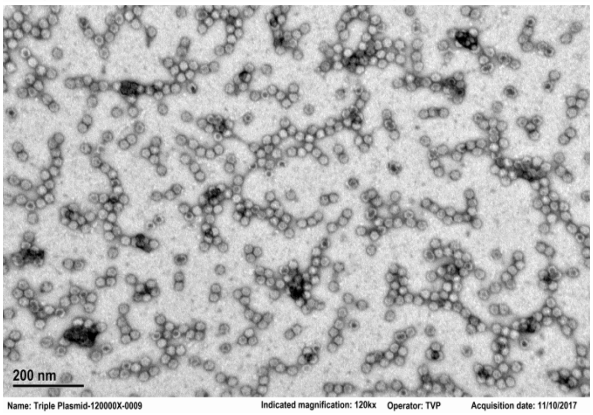

200 nm
Name: TriplePlasmid -100000X -0016 Indicated magnification: 100kx Operator: TVP Acquisition date: 10/31/2017

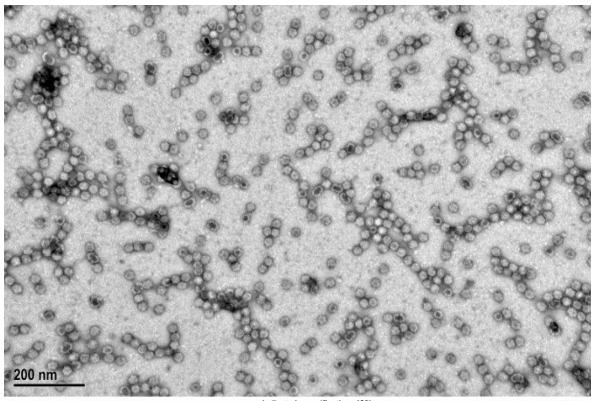

200 nm
Name: Triple Plasmid-100000X-0001 Indicated magnification: 100kx Operator: TVP Acquisition date: 11/10/2017


200 nm
Name: Triple Plasmid-100000X-0002 Indicated magnification: 100kx Operator: TVP Acquisition date: 11/10/2017

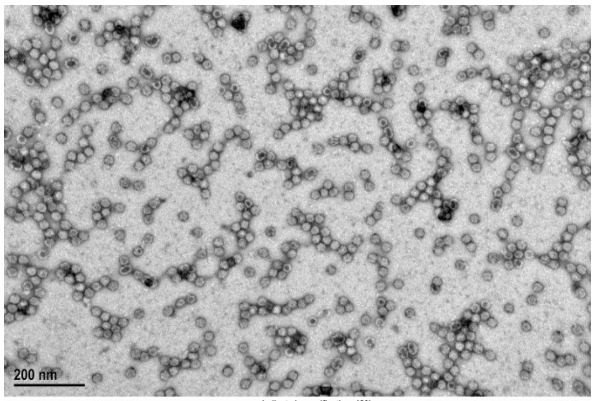

200 nm
Name: Triple Plasmid-100000X-0007 Indicated magnification: 100kx Operator: TVP Acquisition date: 11/10/2017


200 nm
Name: Triple Plasmid-100000X-0008 Indicated magnification: 100kx Operator: TVP Acquisition date: 11/10/2017

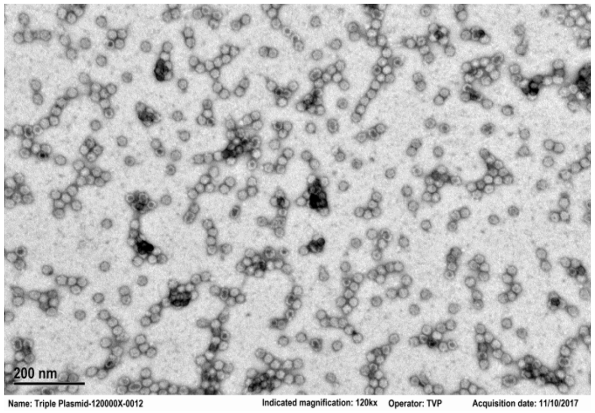

200 nm
Name: Triple Plasmid-120000X-0009 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



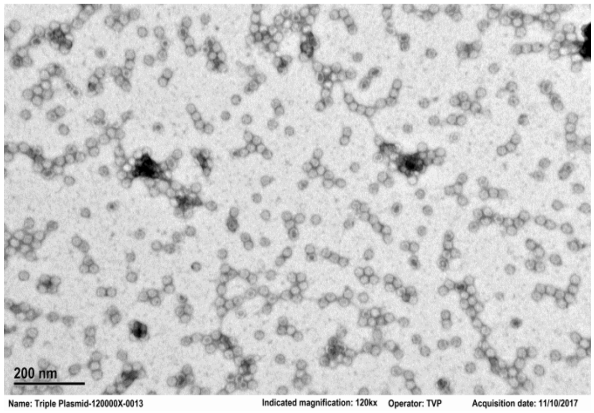
Name: Triple Plasmid-120000X-0010 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



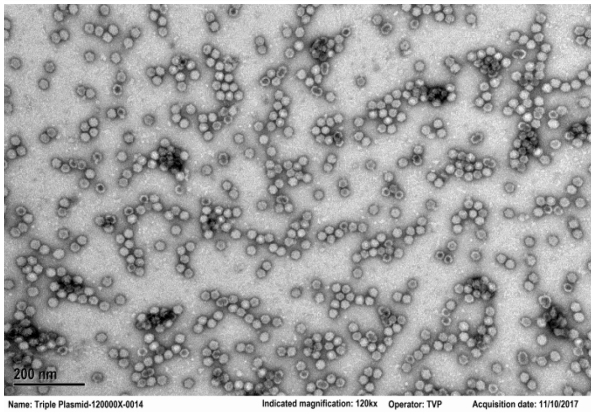
Name: Triple Plasmid-120000X-0011 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



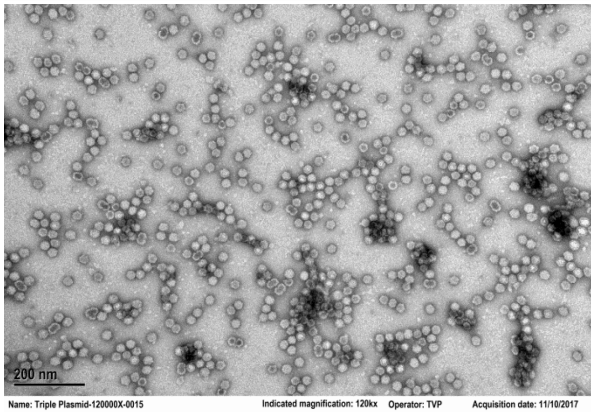
Name: Triple Plasmid-120000X-0012 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



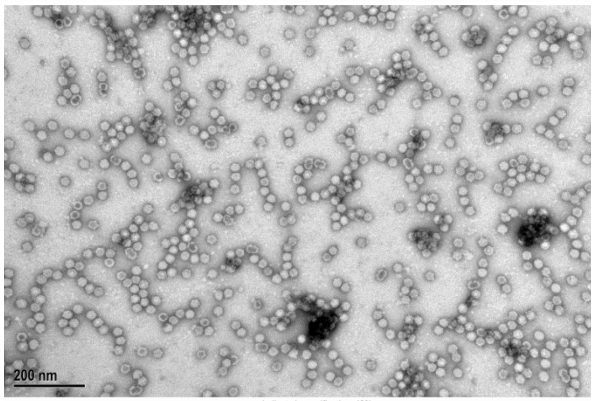
Name: Triple Plasmid-120000X-0013 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



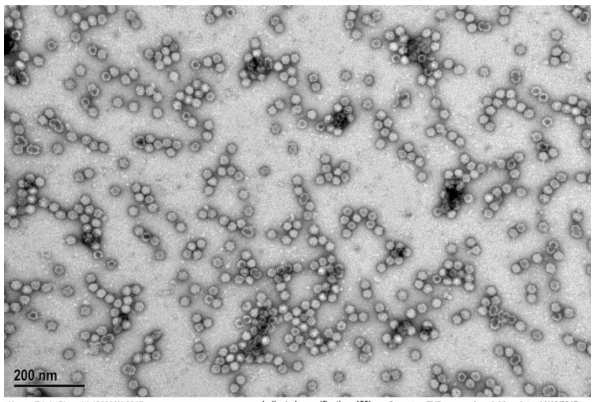
Name: Triple Plasmid-120000X-0014 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



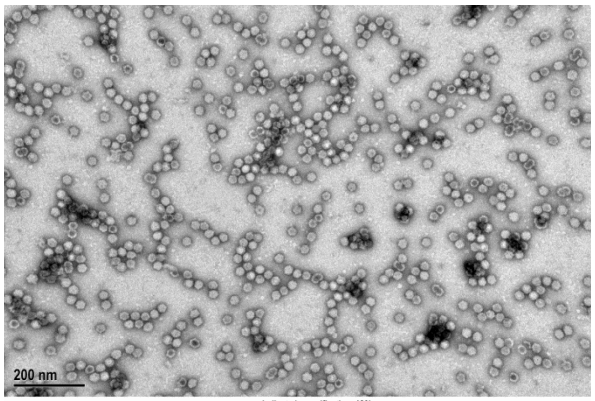
Name: Triple Plasmid-120000X-0015 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



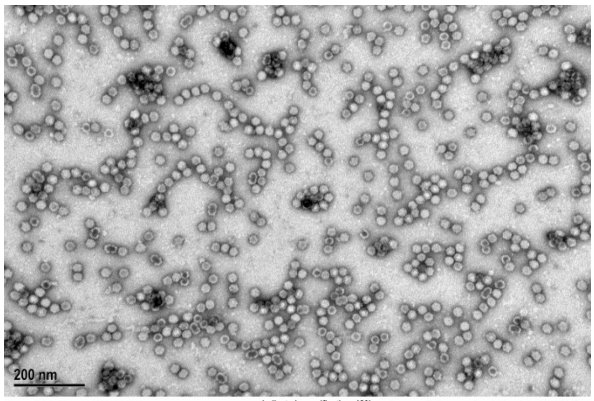
Name: Triple Plasmid-120000X-0016 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



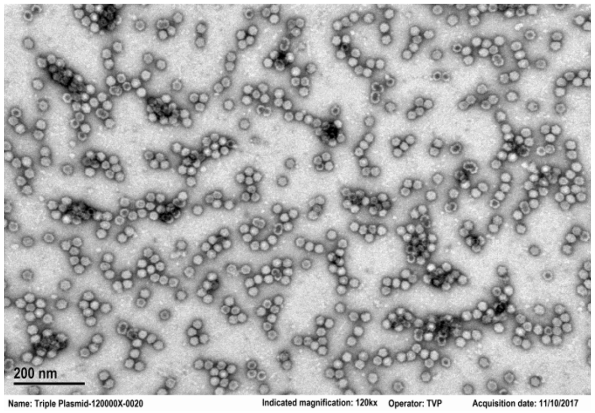
Name: Triple Plasmid-120000X-0017 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



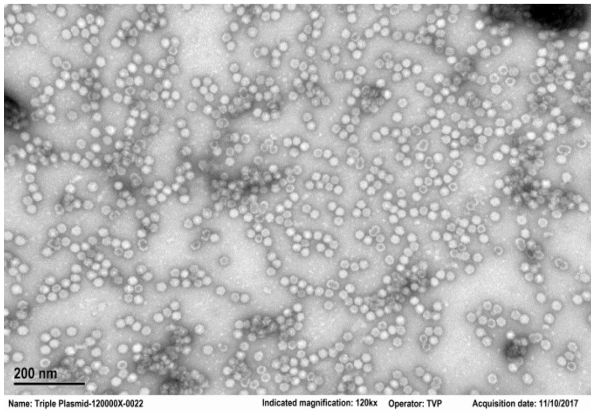
Name: Triple Plasmid-120000X-0018 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



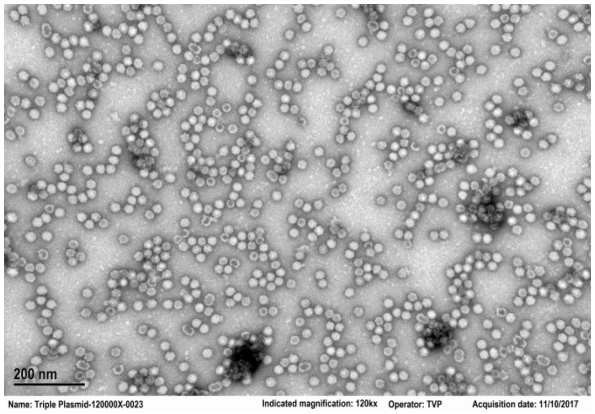
Name: Triple Plasmid-120000X-0019 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



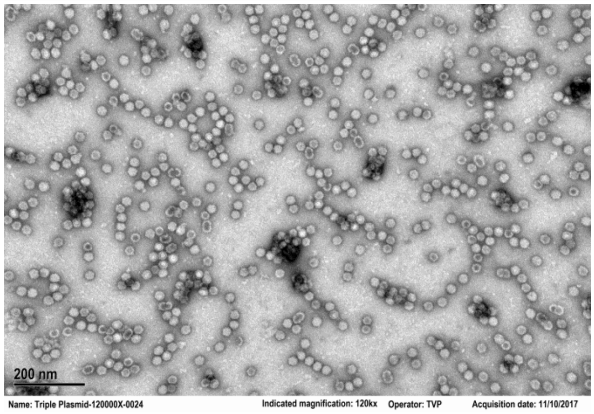
Name: Triple Plasmid-120000X-0020 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



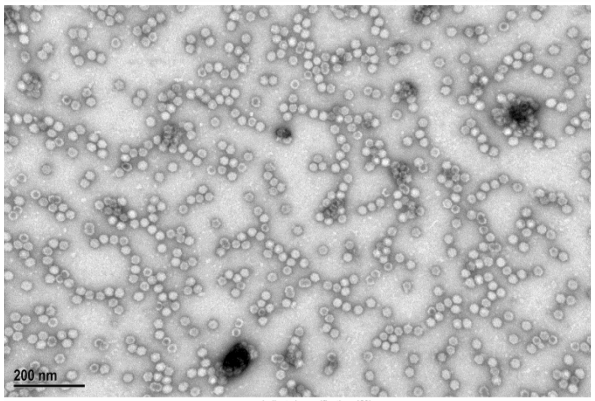
Name: Triple Plasmid-120000X-0022 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



Name: Triple Plasmid-120000X-0023 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

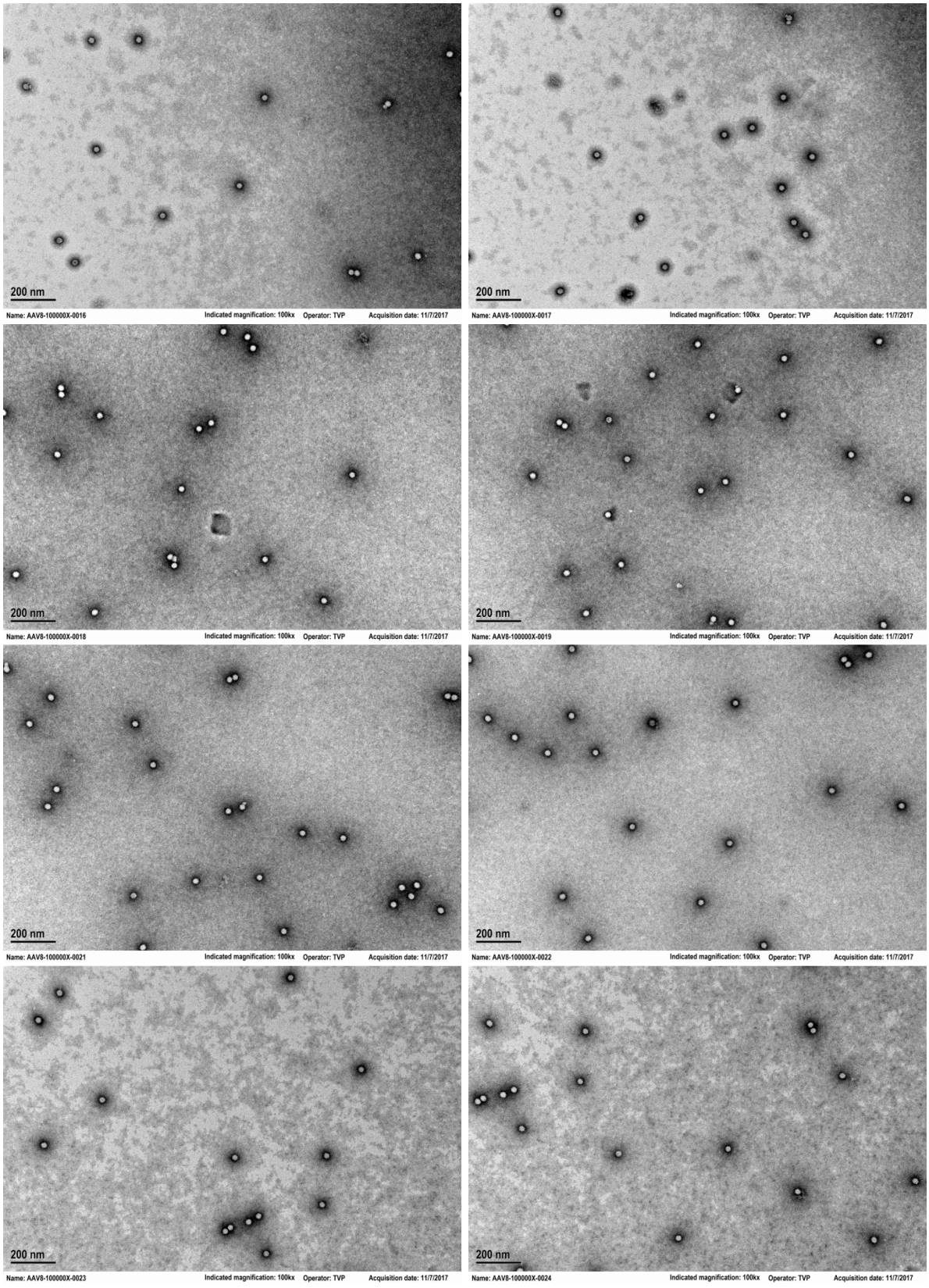


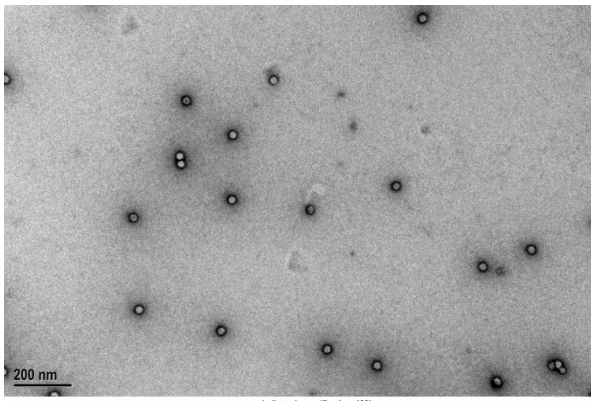
Name: Triple Plasmid-120000X-0024 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017



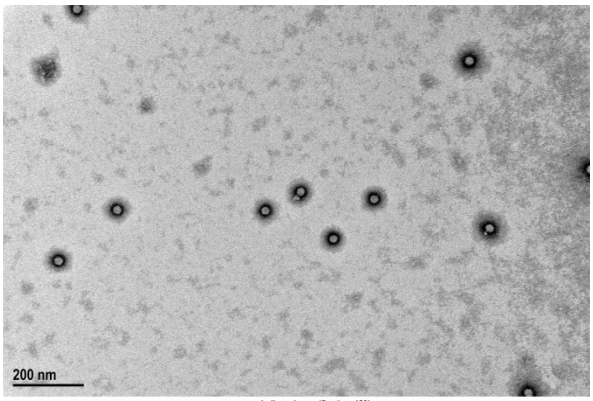
Name: Triple Plasmid-120000X-0026 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/10/2017

Domiphen bromide 0 mM for AAV8

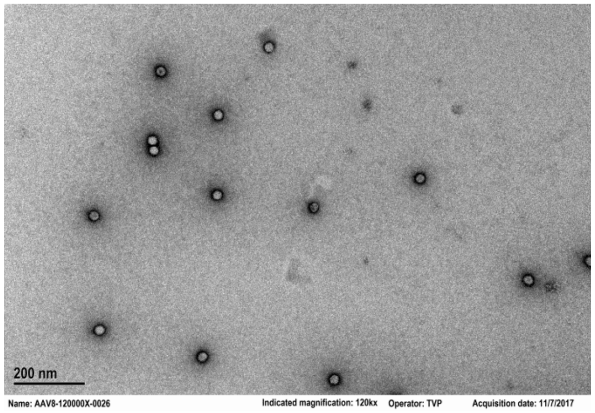




Name: AAV8-100000X-0025 Indicated magnification: 100kx Operator: TVP Acquisition date: 11/7/2017

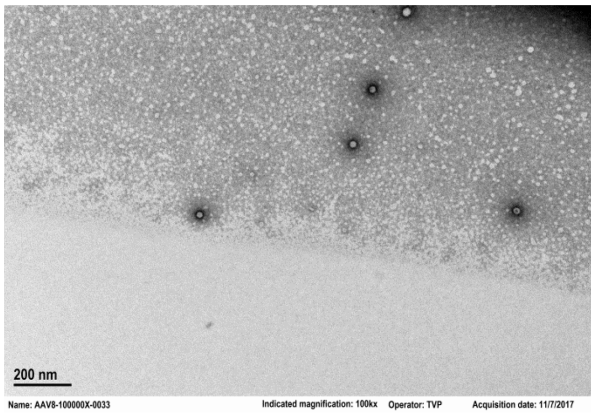


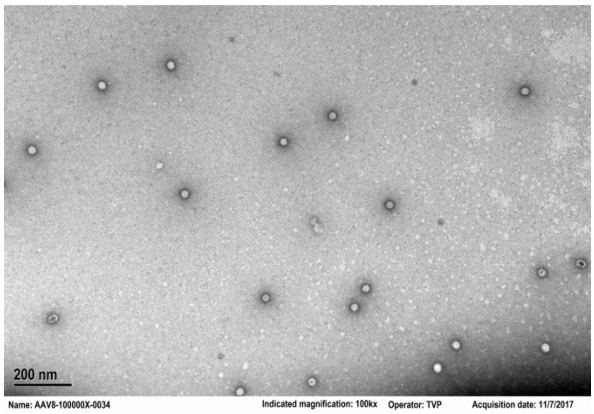
Name: AAV8-120000X-0014 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/7/2017

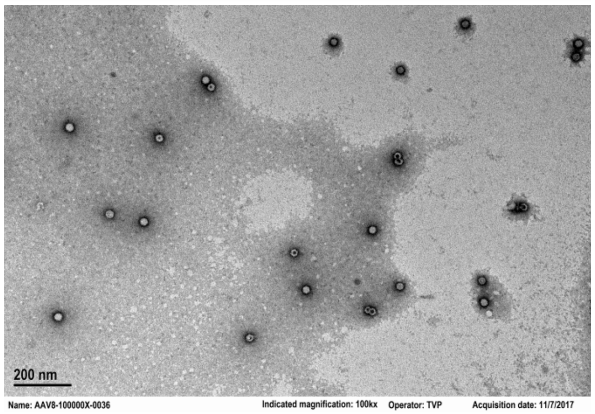


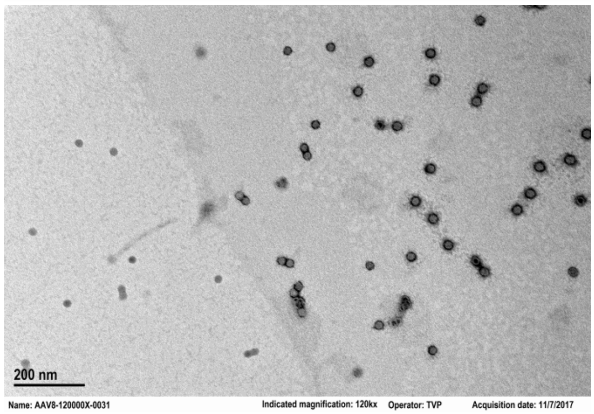
Name: AAV8-120000X-0026 Indicated magnification: 120kx Operator: TVP Acquisition date: 11/7/2017

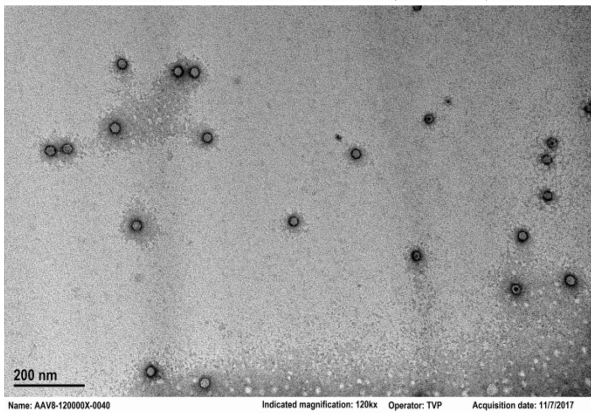
Domiphen bromide 3 mM for AAV8



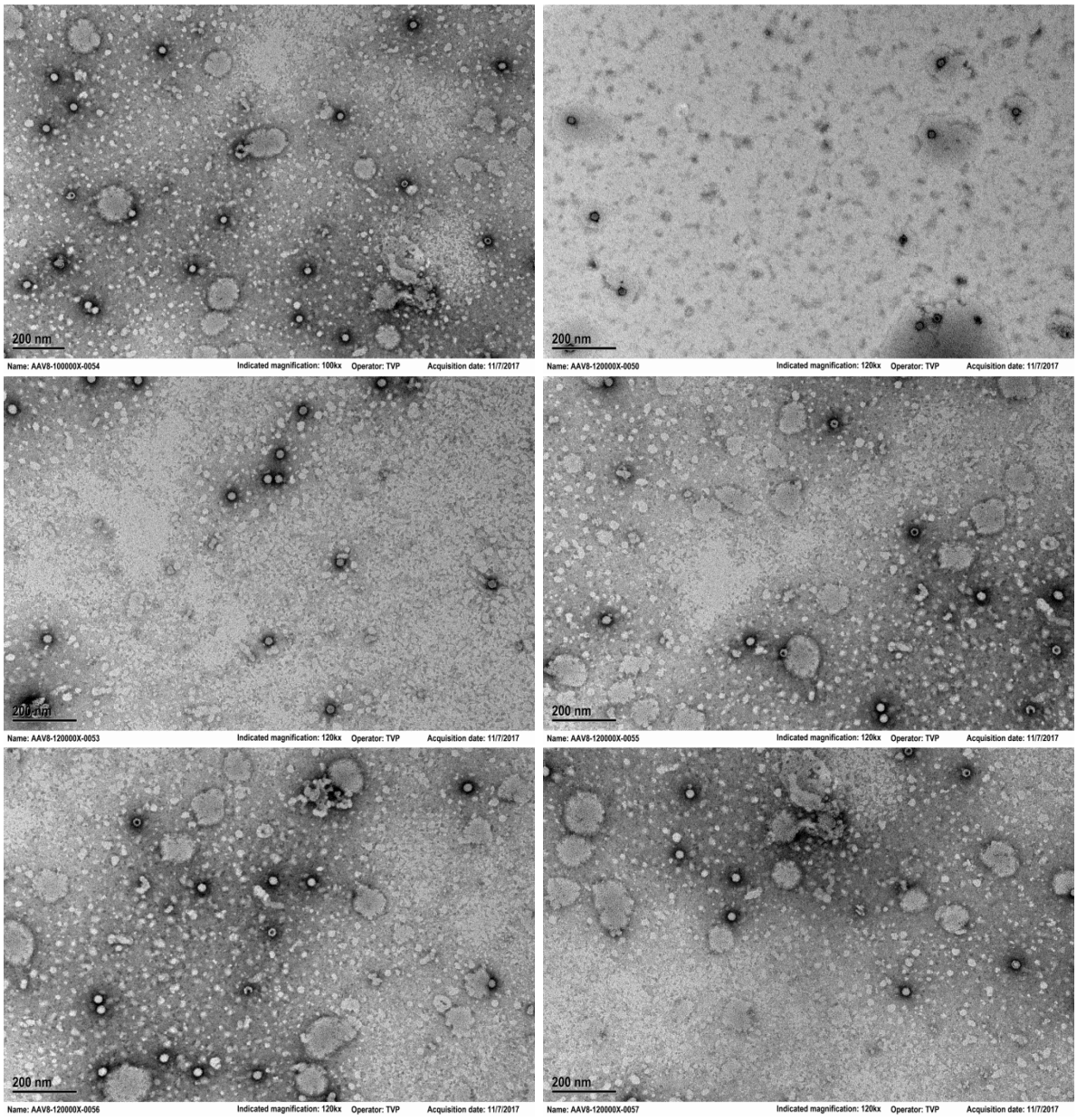








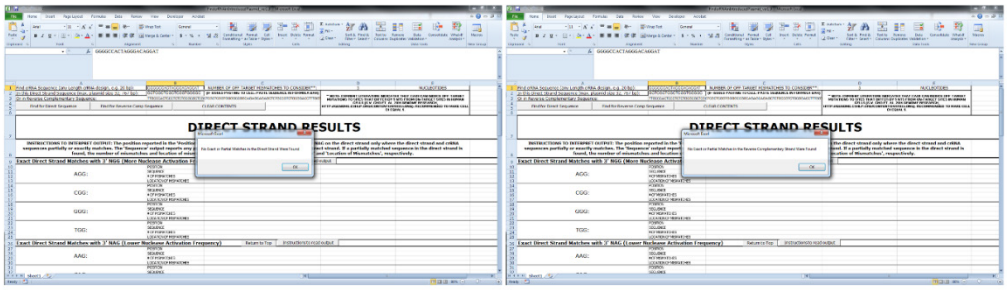
Domiphen bromide 5 mM for AAV8



APPENDIX B – PROGRAM OUTPUTS AND SOURCE CODE FOR PROGRAM IN THE SECOND AIM

a) Prompt indicating 'No Exact or Partial Matches in the Direct Strand Were Found'

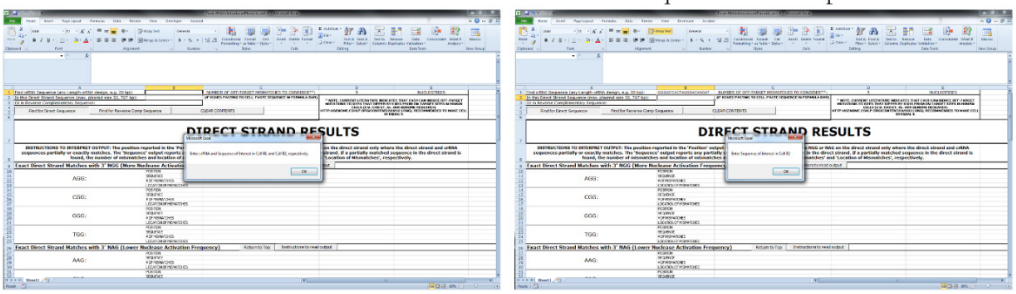
b) Prompt indicating 'No Exact or Partial Matches in the Reverse Complementary Were Found'



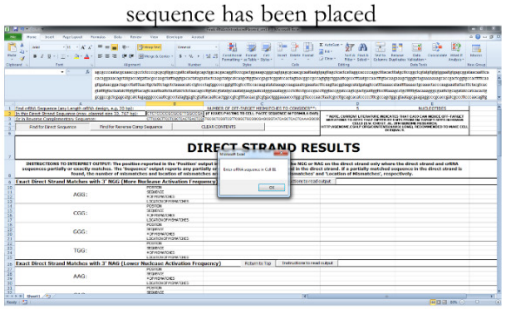
Program for AAVS1 guide-RNA (GGGGCCACTAGGGACAGGAT) screened against AAV1 Genome

a) Error message when no inputs have been placed

b) Error message when no intended integrated plasmid sequence has been placed



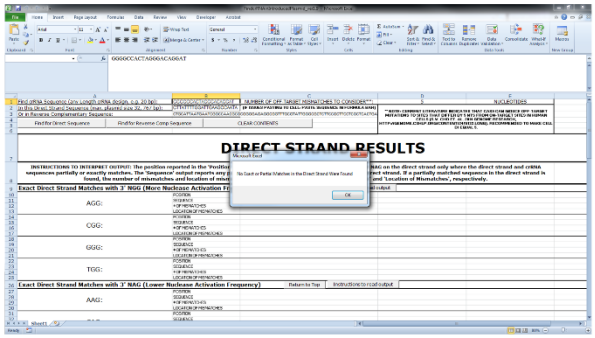
c) Error message when no intended crRNA (i.e. gRNA) sequence has been placed

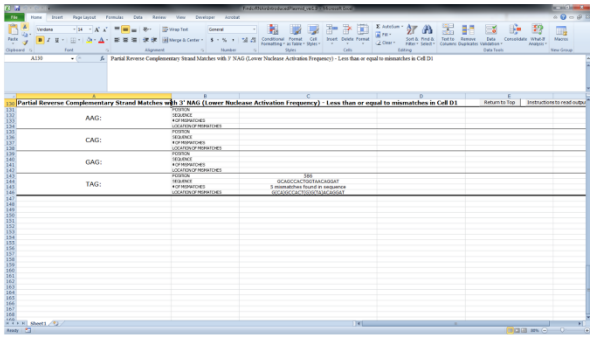


a) Prompt indicating 'No Exact or Partial Matches in the Direct Strand Were Found'

b) Prompt indicating Partial Match on Reverse Complementary Strand:

- PAM – TAG
- POSITION - 586
- SEQUENCE - GCAGCCACTGGTAACAGGAT
- # OF MISMATCHES - 5 mismatches found in sequence
- LOCATION OF MISMATCHES G(CA)GCCACT(G)G(TA)ACAGGAT



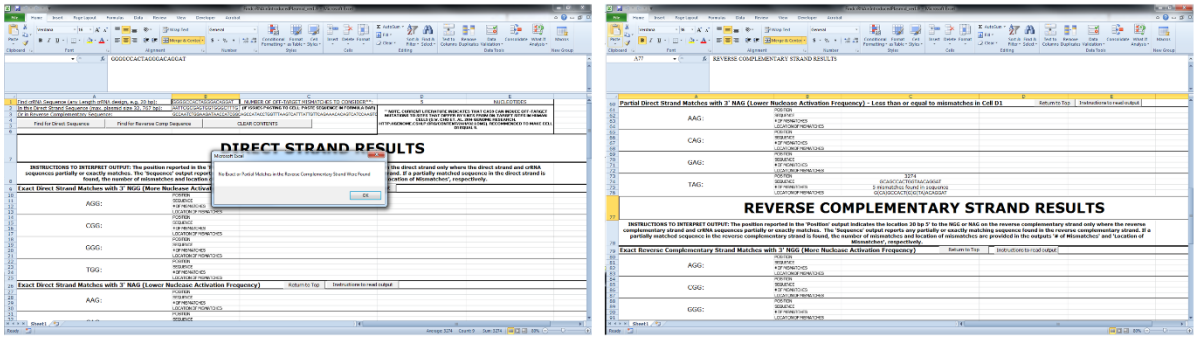


Program for AAVS1 guide-RNA (GGGGCCACTAGGGACAGGAT) screened against pXX2-SseI-2LoxP-Puro

a) Prompt indicating 'No Exact or Partial Matches in the Direct Strand Were Found'

b) Prompt indicating Partial Match on Direct Strand:

- PAM - TAG:
- POSITION - 3274
- SEQUENCE - GCAGCCACTGGTAACAGGAT
- # OF MISMATCHES – 5 mismatches found in sequence
- LOCATION OF MISMATCHES G(CA)GCCACT(G)G(TA)ACAGGAT



Program for AAVS1 guide-RNA (GGGGCCACTAGGGACAGGAT) screened against pBSKS-AAV9

```
'The Exact of Partial Match Function for Direct Strand of DNA sequence


Sub Find()


        Dim str0 As String

        Dim str1 As String

        Dim str2 As String

        Dim a As Integer


        Application.ScreenUpdating = False

        Application.Calculation = xlCalculationManual

        Dim s As String

        For Each c In ActiveSheet.UsedRange

                s = c.Value

                If Trim(Application.Clean(s)) <> s Then

                        s = Trim(Application.Clean(s))

                        c.Value = s

                End If

        Next

        Application.ScreenUpdating = True

        Application.Calculation = xlCalculationAutomatic


'uppercase function

        Application.ScreenUpdating = False


        Dim Cell As Range

        For Each Cell In Range("$B$1:" & _
        Range("$B$1").SpecialCells(xlLastCell).Address)

                If Len(Cell) > 0 Then Cell = UCase(Cell)

        Next Cell
```

```
        Application.ScreenUpdating = True


'clear contents
        Sheet1.Range("C10:AZZ25, C27:AZZ42, C44:AZZ59, C61:AZZ76").CLEARCONTENTS
        str0 = Sheet1.Range("B3").Value 'REV COMP STRAND SEQUENCE
        str1 = Sheet1.Range("B2").Value 'DIRECT STRAND SEQUENCE
        str2 = Sheet1.Range("B1").Value 'crRNA SEQUENCE
        str3 = "AGG"
        str4 = "CGG"
        str5 = "GGG"
        str6 = "TGG"
        str7 = "AAG"
        str8 = "CAG"
        str9 = "GAG"
        str10 = "TAG"
        a = Len(str2)

        If Sheet1.Range("D1").Value = 0 Then
                MsgBox "Enter Number of Off-Target Mismatches to Consider in Cell D1(value _
                        must be greater than 0. Recommended to enter value at 5)."
                GoTo 18
        End If
18
        If Sheet1.Range("B1").Value = 0 Then
                If Sheet1.Range("B2").Value = 0 Then
                        MsgBox "Enter crRNA and Sequence of Interest in Cell B1 and Cell B2, _
                                respectively."
                        GoTo 17
                Else
                        MsgBox "Enter crRNA sequence in Cell B1"
```

```
                GoTo 17
        End If
    Else
        If Sheet1.Range("B2").Value = 0 Then
                MsgBox "Enter Sequence of Interest in Cell B2"
                GoTo 17
        End If
    End If

    If ATCGcheck(str1) = "fail" Then
    MsgBox "Make sure the sequence entered has only nucleotides A, T, C, and G. _
        Be sure to delete any whitespaces (e.g. spaces)."
    GoTo 17
    End If

    If ATCGcheck(str2) = "fail" Then
    MsgBox "Make sure the sequence entered has only nucleotides A, T, C,and G. _
        Be sure to delete any whitespaces (e.g. spaces)."
    GoTo 17
    End If

    i = 1
    col = 3
1
    'InStr(starting index, sequence, subsequence) --> returns the index where
    'the subsequence is found in the sequence (returns 0 if none found)

    'AGG EXACT MATCH
    ind = InStr(i, str1, str3)
    If ind <> 0 Then
```

```
If ind - a < 0 Then
        Sheet1.Cells(11, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "exact")
        If Sheet1.Cells(11, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                        Sheet1.Cells(10, col).Value = Len(str1) - (a - ind)
                Else
                        Sheet1.Cells(10, col).Value = Len(str1) - a + 1
                End If
                Sheet1.Cells(12, col).Value2 = _
                        possibleDSB(ind, str1, str2, "number", "exact")
                Sheet1.Cells(13, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
Else
        Sheet1.Cells(11, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "exact")
        If Sheet1.Cells(11, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(10, col).Value = Len(str1)
                Else
                        Sheet1.Cells(10, col).Value = ind - a
                End If
                Sheet1.Cells(12, col).Value2 = _
                        possibleDSB(ind, str1, str2, "number", "exact")
                Sheet1.Cells(13, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
End If
i = ind + 1
```

```
        GoTo 1
    End If


    i = 1
    col = 3
2
    'CGG EXACT MATCH
    ind = InStr(i, str1, str4)
    If ind <> 0 Then
        If ind - a < 0 Then
            Sheet1.Cells(15, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "exact")
            If Sheet1.Cells(15, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                    Sheet1.Cells(14, col).Value = Len(str1) - (a - ind)
                Else
                    Sheet1.Cells(14, col).Value = Len(str1) - a + 1
                End If
                Sheet1.Cells(16, col).Value2 = _
                    possibleDSB(ind, str1, str2, "number", "exact")
                Sheet1.Cells(17, col).Value = "N/A Exact Match was Found"
                col = col + 1
            End If
        Else
            Sheet1.Cells(15, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "exact")
            If Sheet1.Cells(15, col).Value <> 0 Then
                If ind - a = 0 Then
                    Sheet1.Cells(14, col).Value = Len(str1)
                Else
```

```
                                Sheet1.Cells(14, col).Value = ind - a
                        End If
                        Sheet1.Cells(16, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "exact")
                        Sheet1.Cells(17, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 2
End If
i = 1
col = 3
3
'GGG EXACT MATCH
ind = InStr(i, str1, str5)
If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(19, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "exact")
                If Sheet1.Cells(19, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(18, col).Value = Len(str1) - (a - ind)
                        Else
                                Sheet1.Cells(18, col).Value = Len(str1) - a + 1
                        End If
                        Sheet1.Cells(20, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "exact")
                        Sheet1.Cells(21, col).Value = "N/A Exact Match was Found"
                        col = col + 1
```

```
                End If
        Else
                Sheet1.Cells(19, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "exact")
                If Sheet1.Cells(19, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(18, col).Value = Len(str1)
                        Else
                                Sheet1.Cells(18, col).Value = ind - a
                        End If
                        Sheet1.Cells(20, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "exact")
                        Sheet1.Cells(21, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 3
End If
i = 1
col = 3
4
'TGG EXACT MATCH
ind = InStr(i, str1, str6)
If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(23, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "exact")
                If Sheet1.Cells(23, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
```

```
                        Sheet1.Cells(22, col).Value = Len(str1) - (a - ind)
                    Else
                        Sheet1.Cells(22, col).Value = Len(str1) - a + 1
                    End If
                    Sheet1.Cells(24, col).Value2 = _
                        possibleDSB(ind, str1, str2, "number", "exact")
                    Sheet1.Cells(25, col).Value = _"N/A Exact Match was Found"
                    col = col + 1
                End If
            Else
                Sheet1.Cells(23, col).Value = _
                    possibleDSB(ind, str1, str2, "Nothing", "exact")
                If Sheet1.Cells(23, col).Value <> 0 Then
                    If ind - a = 0 Then
                        Sheet1.Cells(22, col).Value = Len(str1)
                    Else
                        Sheet1.Cells(22, col).Value = ind - a
                    End If
                    Sheet1.Cells(24, col).Value2 = _
                        possibleDSB(ind, str1, str2, "number", "exact")
                    Sheet1.Cells(25, col).Value = "N/A Exact Match was Found"
                    col = col + 1
                End If
            End If
            i = ind + 1
            GoTo 4
    End If

    i = 1
    col = 3
```

```
5
        'AAG EXACT MATCH
        ind = InStr(i, str1, str7)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(28, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "exact")
                        If Sheet1.Cells(28, col).Value <> 0 Then
                                If ind - 1 <> 0 Then
                                        Sheet1.Cells(27, col).Value = Len(str1) - (a - ind)
                                Else
                                        Sheet1.Cells(27, col).Value = Len(str1) - a + 1
                                End If
                                Sheet1.Cells(29, col).Value2 = _
                                        possibleDSB(ind, str1, str2, "number", "exact")
                                Sheet1.Cells(30, col).Value = "N/A Exact Match was Found"
                                col = col + 1
                        End If
                Else
                        Sheet1.Cells(28, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "exact")
                        If Sheet1.Cells(28, col).Value <> 0 Then
                                If ind - a = 0 Then
                                        Sheet1.Cells(27, col).Value = Len(str1)
                                Else
                                        Sheet1.Cells(27, col).Value = ind - a
                                End If
                                Sheet1.Cells(29, col).Value2 = _
                                        possibleDSB(ind, str1, str2, "number", "exact")
```

```
                        Sheet1.Cells(30, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 5
End If


i = 1


col = 3


6
        'CAG EXACT MATCH
        ind = InStr(i, str1, str8)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(32, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "exact")
                        If Sheet1.Cells(32, col).Value <> 0 Then
                                If ind - 1 <> 0 Then
                                        Sheet1.Cells(31, col).Value = Len(str1) - (a - ind)
                                Else
                                        Sheet1.Cells(31, col).Value = Len(str1) - a + 1
                                End If
                                Sheet1.Cells(33, col).Value2 = _
                                        possibleDSB(ind, str1, str2, "number", "exact")
                                Sheet1.Cells(34, col).Value = "N/A Exact Match was Found"
                                col = col + 1
                        End If
```

```
        Else
                Sheet1.Cells(32, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "exact")
                If Sheet1.Cells(32, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(31, col).Value = Len(str1)
                        Else
                                Sheet1.Cells(31, col).Value = ind - a
                        End If
                        Sheet1.Cells(33, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "exact")
                        Sheet1.Cells(34, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 6
    End If


    i = 1
    col = 3


7
    'GAG EXACT MATCH
    ind = InStr(i, str1, str9)
    If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(36, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "exact")
                If Sheet1.Cells(36, col).Value <> 0 Then
```

```
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(35, col).Value = Len(str1) - (a - ind)
                        Else
                                Sheet1.Cells(35, col).Value = _Len(str1) - a + 1
                        End If
                        Sheet1.Cells(37, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "exact")
                        Sheet1.Cells(38, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        Else
                Sheet1.Cells(36, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "exact")
                If Sheet1.Cells(36, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(35, col).Value = Len(str1)
                        Else
                                Sheet1.Cells(35, col).Value = ind - a
                        End If
                        Sheet1.Cells(37, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "exact")
                        Sheet1.Cells(38, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 7
End If


i = 1
```

```
        col = 3


8

        'TAG EXACT MATCH
        ind = InStr(i, str1, str10)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(40, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "exact")
                        If Sheet1.Cells(40, col).Value <> 0 Then
                                If ind - 1 <> 0 Then
                                        Sheet1.Cells(39, col).Value = Len(str1) - (a - ind)
                                Else
                                        Sheet1.Cells(39, col).Value = Len(str1) - a + 1
                                End If
                                Sheet1.Cells(41, col).Value2 = _
                                        possibleDSB(ind, str1, str2, "number", "exact")
                                Sheet1.Cells(42, col).Value = "N/A Exact Match was Found"
                                col = col + 1
                        End If
                Else
                        Sheet1.Cells(40, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "exact")
                        If Sheet1.Cells(40, col).Value <> 0 Then
                                If ind - a = 0 Then
                                        Sheet1.Cells(39, col).Value = Len(str1)
                                Else
                                        Sheet1.Cells(39, col).Value = ind - a
                                End If
                                Sheet1.Cells(41, col).Value2 = _
```

```
                                possibleDSB(ind, str1, str2, "number", "exact")
                        Sheet1.Cells(42, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 8
End If


i = 1
col = 3


9
        'AGG PARTIAL MATCH
        ind = InStr(i, str1, str3)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(45, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "partial")
                        If Sheet1.Cells(45, col).Value <> 0 Then
                                If ind - 1 <> 0 Then
                                        Sheet1.Cells(44, col).Value = Len(str1) - (a - ind)
                                Else
                                        Sheet1.Cells(44, col).Value = Len(str1) - a + 1
                                End If
                                Sheet1.Cells(46, col).Value2 = _
                                        possibleDSB(ind, str1, str2, "number", "partial")
                                Sheet1.Cells(47, col).Value = _
                                        highlightDifference1(str2, Sheet1.Cells(45, col).Value)
                                col = col + 1
```

```
                End If
        Else
                Sheet1.Cells(45, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "partial")
                If Sheet1.Cells(45, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(44, col).Value = Len(str1)
                        Else
                                Sheet1.Cells(44, col).Value = ind - a
                        End If
                        Sheet1.Cells(46, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "partial")
                        Sheet1.Cells(47, col).Value = -
                                highlightDifference1(str2, Sheet1.Cells(45,col).Value)
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 9
End If


i = 1
col = 3


10
        'CGG PARTIAL MATCH
        ind = InStr(i, str1, str4)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(49, col).Value = _
```

```
                possibleDSB(ind, str1, str2, "Nothing", "partial")
        If Sheet1.Cells(49, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                        Sheet1.Cells(48, col).Value = Len(str1) - (a - ind)
                Else
                        Sheet1.Cells(48, col).Value = Len(str1) - a + 1
                End If
                Sheet1.Cells(50, col).Value2 = _
                        possibleDSB(ind, str1, str2, "number", "partial")
                Sheet1.Cells(51, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(49, col).Value)
                col = col + 1
        End If
Else
        Sheet1.Cells(49, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "partial")
        If Sheet1.Cells(49, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(48, col).Value = Len(str1)
                Else
                        Sheet1.Cells(48, col).Value = ind - a
                End If
                Sheet1.Cells(50, col).Value2 = _
                        possibleDSB(ind, str1, str2, "number", "partial")
                Sheet1.Cells(51, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(49, col).Value)
                col = col + 1
        End If
End If
i = ind + 1
```

```
        GoTo 10
    End If


    i = 1
    col = 3


11
    'GGG PARTIAL MATCH
    ind = InStr(i, str1, str5)
    If ind <> 0 Then
        If ind - a < 0 Then
            Sheet1.Cells(53, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "partial")
            If Sheet1.Cells(53, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                    Sheet1.Cells(52, col).Value = Len(str1) - (a - ind)
                Else
                    Sheet1.Cells(52, col).Value = Len(str1) - a + 1
                End If
                Sheet1.Cells(54, col).Value2 = _
                    possibleDSB(ind, str1, str2, "number", "partial")
                Sheet1.Cells(55, col).Value = _
                    highlightDifference1(str2, Sheet1.Cells(53, col).Value)
                col = col + 1
            End If
        Else
            Sheet1.Cells(53, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "partial")
            If Sheet1.Cells(53, col).Value <> 0 Then
                If ind - a = 0 Then
```

```
                                Sheet1.Cells(52, col).Value = Len(str1)
                        Else
                                Sheet1.Cells(52, col).Value = ind - a
                        End If
                        Sheet1.Cells(54, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "partial")
                        Sheet1.Cells(55, col).Value = _
                                highlightDifference1(str2, Sheet1.Cells(53, col).Value)
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 11
End If

i = 1
col = 3

12
        'TGG PARTIAL MATCH
        ind = InStr(i, str1, str6)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(57, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "partial")
                        If Sheet1.Cells(57, col).Value <> 0 Then
                                If ind - 1 <> 0 Then
                                        Sheet1.Cells(56, col).Value = Len(str1) - (a - ind)
                                Else
                                        Sheet1.Cells(56, col).Value = Len(str1) - a + 1
```

```
                        End If
                        Sheet1.Cells(58, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "partial")
                        Sheet1.Cells(59, col).Value = _
                                highlightDifference1(str2, Sheet1.Cells(57, col).Value)
                        col = col + 1
                End If
        Else
                Sheet1.Cells(57, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "partial")
                If Sheet1.Cells(57, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(56, col).Value = Len(str1)
                        Else
                                Sheet1.Cells(56, col).Value = ind - a
                        End If
                        Sheet1.Cells(58, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "partial")
                        Sheet1.Cells(59, col).Value = _
                                highlightDifference1(str2, Sheet1.Cells(57, col).Value)
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 12
End If


i = 1
col = 3
```

```
13

        'AAG PARTIAL MATCH
        ind = InStr(i, str1, str7)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(62, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "partial")
                        If Sheet1.Cells(62, col).Value <> 0 Then
                                If ind - 1 <> 0 Then
                                        Sheet1.Cells(61, col).Value = _Len(str1) - (a - ind)
                                Else
                                        Sheet1.Cells(61, col).Value = Len(str1) - a + 1
                                End If
                                Sheet1.Cells(63, col).Value2 = _
                                        possibleDSB(ind, str1, str2, "number", "partial")
                                Sheet1.Cells(64, col).Value = _
                                        highlightDifference1(str2, Sheet1.Cells(62, col).Value)
                                col = col + 1
                        End If
                Else
                        Sheet1.Cells(62, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "partial")
                        If Sheet1.Cells(62, col).Value <> 0 Then
                                If ind - a = 0 Then
                                        Sheet1.Cells(61, col).Value = Len(str1)
                                Else
                                        Sheet1.Cells(61, col).Value = ind - a
                                End If
                                Sheet1.Cells(63, col).Value2 = _
                                        possibleDSB(ind, str1, str2, "number", "partial")
```

```
                        Sheet1.Cells(64, col).Value = _
                                highlightDifference1(str2, Sheet1.Cells(62, col).Value)
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 13
End If


i = 1
col = 3


14
        'CAG PARTIAL MATCH
        ind = InStr(i, str1, str8)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(66, col).Value = _
                                possibleDSB(ind, str1, str2, "Nothing", "partial")
                        If Sheet1.Cells(66, col).Value <> 0 Then
                                If ind - 1 <> 0 Then
                                        Sheet1.Cells(65, col).Value = Len(str1) - (a - ind)
                                Else
                                        Sheet1.Cells(65, col).Value = Len(str1) - a + 1
                                End If
                                Sheet1.Cells(67, col).Value2 = _
                                        possibleDSB(ind, str1, str2, "number", "partial")
                                Sheet1.Cells(68, col).Value = _
                                        highlightDifference1(str2, Sheet1.Cells(66, col).Value)
                                col = col + 1
```

```
                End If
        Else
                Sheet1.Cells(66, col).Value = _
                        possibleDSB(ind, str1, str2, "Nothing", "partial")
                If Sheet1.Cells(66, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(65, col).Value = Len(str1)
                        Else
                                Sheet1.Cells(65, col).Value = ind - a
                        End If
                        Sheet1.Cells(67, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "partial")
                        Sheet1.Cells(68, col).Value = _
                                highlightDifference1(str2, Sheet1.Cells(66, col).Value)
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 14
End If


i = 1
col = 3


15
        'GAG PARTIAL MATCH
        ind = InStr(i, str1, str9)
        If ind <> 0 Then
                If ind - a < 0 Then
                        Sheet1.Cells(70, col).Value = _
```

```
                possibleDSB(ind, str1, str2, "Nothing", "partial")
        If Sheet1.Cells(70, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                        Sheet1.Cells(69, col).Value = Len(str1) - (a - ind)
                Else
                        Sheet1.Cells(69, col).Value = Len(str1) - a + 1
                End If
                Sheet1.Cells(71, col).Value2 = _
                        possibleDSB(ind, str1, str2, "number", "partial")
                Sheet1.Cells(72, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(70, col).Value)
                col = col + 1
        End If
Else
        Sheet1.Cells(70, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "partial")
        If Sheet1.Cells(70, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(69, col).Value = Len(str1)
                Else
                        Sheet1.Cells(69, col).Value = ind - a
                End If
                Sheet1.Cells(71, col).Value2 = _
                        possibleDSB(ind, str1, str2, "number", "partial")
                Sheet1.Cells(72, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(70, col).Value)
                col = col + 1
        End If
End If
i = ind + 1
```

```
        GoTo 15
    End If


    i = 1
    col = 3


16
    'TAG PARTIAL MATCH
    ind = InStr(i, str1, str10)
    If ind <> 0 Then
        If ind - a < 0 Then
            Sheet1.Cells(74, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "partial")
            If Sheet1.Cells(74, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                    Sheet1.Cells(73, col).Value = Len(str1) - (a - ind)
                Else
                    Sheet1.Cells(73, col).Value = Len(str1) - a + 1
                End If
                Sheet1.Cells(75, col).Value2 = _
                    possibleDSB(ind, str1, str2, "number", "partial")
                Sheet1.Cells(76, col).Value = _
                    highlightDifference1(str2, Sheet1.Cells(74, col).Value)
                col = col + 1
            End If
        Else
            Sheet1.Cells(74, col).Value = _
                possibleDSB(ind, str1, str2, "Nothing", "partial")
            If Sheet1.Cells(74, col).Value <> 0 Then
                If ind - a = 0 Then
```

```
                                Sheet1.Cells(73, col).Value = Len(str1)
                        Else
                                Sheet1.Cells(73, col).Value = ind - a
                        End If
                        Sheet1.Cells(75, col).Value2 = _
                                possibleDSB(ind, str1, str2, "number", "partial")
                        Sheet1.Cells(76, col).Value = _
                                highlightDifference1(str2, Sheet1.Cells(74, col).Value)
                        col = col + 1
                End If
        End If
        i = ind + 1
        GoTo 16
End If

Range("B3").Select
ActiveCell.FormulaR1C1 = "=revstr(R2C2)"

'SELECT FOR AND MOVE TOWARDS THOSE THAT HAVE RESULTS

If WorksheetFunction.CountA(Sheet1.Range("C10:C25")) = 0 Then
If WorksheetFunction.CountA(Sheet1.Range("C27:C42")) = 0 Then
        If WorksheetFunction.CountA(Sheet1.Range("C44:C59")) = 0 Then
                If WorksheetFunction.CountA(Sheet1.Range("C61:C76")) = 0 Then
                        ActiveWindow.ScrollRow = 1
                        Sheet1.Range("B1").Select
                        MsgBox "No Exact or Partial Matches in the Direct _
                        Strand Were Found"
                        GoTo 17
                Else
```

```
                        ActiveWindow.ScrollRow = 60
                        Sheet1.Range("A60").Select
                        GoTo 17
                End If
        Else
                ActiveWindow.ScrollRow = 43
                Sheet1.Range("A43").Select
                GoTo 17
        End If
Else
        ActiveWindow.ScrollRow = 26
        Sheet1.Range("A26").Select
        GoTo 17
End If
Else
ActiveWindow.ScrollRow = 9
Module1 - 11
Sheet1.Range("A9").Select
GoTo 17
End If

17

End Sub

Sub scroll()

        ActiveWindow.ScrollRow = 1
        ActiveWindow.ScrollColumn = 1
        Sheet1.Range("B1").Select
```

```
End Sub

'The Clear contents button macro

Sub CLEARCONTENTS()


        Range("B1:B2, G1:AZZ6, C10:AZZ25, C27:AZZ42, C44:AZZ59, C61:AZZ76,C80:AZZ95, _

        C97:AZZ112, C114:AZZ129, C131:AZZ146").Select

        Range("B1").Activate

        Selection.CLEARCONTENTS

        Range("B3").Select

        ActiveCell.FormulaR1C1 = "=revstr(R2C2)"

        Range("B1").Select


End Sub
```

```
'Converting the direct strand sequence into the reverse complementary sequence


Function revstr(c)

        Dim i As Long

        Dim newstr As String


        i = Len(c)

        For i = i To 1 Step -1

                Select Case UCase(mid(c, i, 1))

                        Case "A"

                        newstr = newstr & "T"

                        Case "C"

                        newstr = newstr & "G"

                        Case "G"

                        newstr = newstr & "C"

                        Case "T"

                        newstr = newstr & "A"

                End Select

        Next


        revstr = newstr

End Function
```

```
'The Exact or Partial Match Function for Reverse Complementary Strand of DNA sequence


Sub Find()


        Dim str0 As String

        Dim str1 As String

        Dim str2 As String

        Dim a As Integer


        Application.ScreenUpdating = False

        Application.Calculation = xlCalculationManual

        Dim s As String

        For Each c In ActiveSheet.UsedRange

                s = c.Value

                If Trim(Application.Clean(s)) <> s Then

                        s = Trim(Application.Clean(s))

                        c.Value = s

                End If

        Next

        Application.ScreenUpdating = True

        Application.Calculation = xlCalculationAutomatic


'uppercase function

        Application.ScreenUpdating = False

        Dim Cell As Range

        For Each Cell In Range("$B$1:" & Range("$B$1").SpecialCells(xlLastCell).Address)

                If Len(Cell) > 0 Then Cell = UCase(Cell)

        Next Cell


        Application.ScreenUpdating = True
```

```
'clear contents
        Sheet1.Range("C80:AZZ95, C97:AZZ112, C114:AZZ129, C131:AZZ146").CLEARCONTENTS


        str0 = Sheet1.Range("B3").Value 'REV COMP STRAND SEQUENCE
        str1 = Sheet1.Range("B2").Value 'DIRECT STRAND SEQUENCE
        str2 = Sheet1.Range("B1").Value 'crRNA SEQUENCE
        str3 = "AGG"
        str4 = "CGG"
        str5 = "GGG"
        str6 = "TGG"
        str7 = "AAG"
        str8 = "CAG"
        str9 = "GAG"
        str10 = "TAG"
        a = Len(str2)


        If Sheet1.Range("D1").Value = 0 Then
        MsgBox "Enter Number of Off-Target Mismatches to Consider in Cell D1 _
                (value must be greater than 0. Recommended to enter value at 5)"
        GoTo 18
        End If


18


        If Sheet1.Range("B1").Value = 0 Then
        If Sheet1.Range("B2").Value = 0 Then
                MsgBox "Enter crRNA and Sequence of Interest in Cell B1 and Cell B2, _
                respectively."
                GoTo 17
        Else
```

```
                MsgBox "Enter crRNA sequence in Cell B1"
                GoTo 17
        End If
        Else
        If Sheet1.Range("B2").Value = 0 Then
                MsgBox "Enter Sequence of Interest in Cell B2"
                GoTo 17
        End If
        End If

If ATCGcheck(str0) = "fail" Then
        MsgBox "Make sure the sequence entered has only nucleotides A, T, C, and G. Be sure _
                to delete any whitespaces (e.g. spaces)"
        GoTo 17
End If

If ATCGcheck(str1) = "fail" Then
        MsgBox "Make sure the sequence entered has only nucleotides A, T, C, and G. Be sure _
                to delete any whitespaces (e.g. spaces)"
        GoTo 17
End If

i = 1
col = 3

1
        'InStr(starting index, sequence, subsequence) --> returns the index where
        'the subsequence is found in the sequence (returns 0 if none found)

        'AGG EXACT MATCH
```

```
ind = InStr(i, str0, str3)
If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(81, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(81, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(80, col).Value = Len(str0) - (a - ind)
                        Else
                                Sheet1.Cells(80, col).Value = Len(str0) - a + 1
                        End If
                        Sheet1.Cells(82, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "exact")
                        Sheet1.Cells(83, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        Else
                Sheet1.Cells(81, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(81, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(80, col).Value = Len(str0)
                        Else
                                Sheet1.Cells(80, col).Value = ind - a
                        End If
                        Sheet1.Cells(82, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "exact")
                        Sheet1.Cells(83, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
```

```
        End If
        i = ind + 1
        GoTo 1
    End If

    i = 1
    col = 3

2

    'CGG EXACT MATCH
    ind = InStr(i, str0, str4)
    If ind <> 0 Then
    If ind - a < 0 Then
        Sheet1.Cells(85, col).Value = _
            possibleDSB(ind, str0, str2, "Nothing", "exact")
        If Sheet1.Cells(85, col).Value <> 0 Then
            If ind - 1 <> 0 Then
                Sheet1.Cells(84, col).Value = Len(str0) - (a - ind)
            Else
                Sheet1.Cells(84, col).Value = Len(str0) - a + 1
            End If
            Sheet1.Cells(86, col).Value2 = _
                possibleDSB(ind, str0, str2, "number", "exact")
            Sheet1.Cells(87, col).Value = "N/A Exact Match was Found"
            col = col + 1
        End If
    Else
        Sheet1.Cells(85, col).Value = _
            possibleDSB(ind, str0, str2, "Nothing", "exact")
```

```
        If Sheet1.Cells(85, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(84, col).Value = Len(str0)
                Else
                        Sheet1.Cells(84, col).Value = ind - a
                End If
                Sheet1.Cells(86, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "exact")
                Sheet1.Cells(87, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
End If
i = ind + 1
GoTo 2
End If

i = 1
col = 3

3
        'GGG EXACT MATCH
        ind = InStr(i, str0, str5)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(89, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(89, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(88, col).Value = Len(str0) - (a - ind)
                        Else
```

```
                        Sheet1.Cells(88, col).Value = Len(str0) - a + 1
                End If
                Sheet1.Cells(90, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "exact")
                Sheet1.Cells(91, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
Else
        Sheet1.Cells(89, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "exact")
        If Sheet1.Cells(89, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(88, col).Value = Len(str0)
                Else
                        Sheet1.Cells(88, col).Value = ind - a
                End If
                Sheet1.Cells(90, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "exact")
                Sheet1.Cells(91, col).Value = _
                        "N/A Exact Match was Found"
                col = col + 1
        End If
End If
i = ind + 1
GoTo 3
End If


i = 1
col = 3
```

```
4

        'TGG EXACT MATCH
        ind = InStr(i, str0, str6)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(93, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(93, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(92, col).Value = Len(str0) - (a - ind)
                        Else
                                Sheet1.Cells(92, col).Value = Len(str0) - a + 1
                        End If
                        Sheet1.Cells(94, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "exact")
                        Sheet1.Cells(95, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
                Else
                Sheet1.Cells(93, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(93, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(92, col).Value = Len(str0)
                        Else
                                Sheet1.Cells(92, col).Value = ind - a
                        End If
                        Sheet1.Cells(94, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "exact")
                        Sheet1.Cells(95, col).Value = "N/A Exact Match was Found"
```

```
                col = col + 1
        End If
End If
i = ind + 1
GoTo 4
End If

i = 1
col = 3

5
'AAG EXACT MATCH
ind = InStr(i, str0, str7)
If ind <> 0 Then
If ind - a < 0 Then
        Sheet1.Cells(98, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "exact")
        If Sheet1.Cells(98, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                        Sheet1.Cells(97, col).Value = Len(str0) - (a - ind)
                Else
                        Sheet1.Cells(97, col).Value = Len(str0) - a + 1
                End If
                Sheet1.Cells(99, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "exact")
                Sheet1.Cells(100, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
Else
        Sheet1.Cells(98, col).Value = _
```

```
                possibleDSB(ind, str0, str2, "Nothing", "exact")
        If Sheet1.Cells(98, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(97, col).Value = Len(str0)
                Else
                        Sheet1.Cells(97, col).Value = ind - a
                End If
                Sheet1.Cells(99, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "exact")
                Sheet1.Cells(100, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
End If
i = ind + 1
GoTo 5
End If


i = 1
col = 3


6
        'CAG EXACT MATCH
        ind = InStr(i, str0, str8)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(102, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(102, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(101, col).Value = Len(str0) - (a - ind)
```

```
                Else
                        Sheet1.Cells(101, col).Value = Len(str0) - a + 1
                End If
                Sheet1.Cells(103, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "exact")
                Sheet1.Cells(104, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
Else
        Sheet1.Cells(102, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "exact")
        If Sheet1.Cells(102, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(101, col).Value = Len(str0)
                Else
                        Sheet1.Cells(101, col).Value = ind - a
                End If
                Sheet1.Cells(103, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "exact")
                Sheet1.Cells(104, col).Value = _
                        "N/A Exact Match was Found"
                col = col + 1
        End If
End If
i = ind + 1
GoTo 6
End If


i = 1
col = 3
```

```
7
        'GAG EXACT MATCH
        ind = InStr(i, str0, str9)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(106, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(106, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(105, col).Value = Len(str0) - (a - ind)
                        Else
                                Sheet1.Cells(105, col).Value = Len(str0) - a + 1
                        End If
                        Sheet1.Cells(107, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "exact")
                        Sheet1.Cells(108, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        Else
                Sheet1.Cells(106, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(106, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(105, col).Value = Len(str0)
                        Else
                                Sheet1.Cells(105, col).Value = ind - a
                        End If
                        Sheet1.Cells(107, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "exact")
```

```
                Sheet1.Cells(108, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
        End If
        i = ind + 1
        GoTo 7
End If

i = 1
col = 3

8
        'TAG EXACT MATCH
        ind = InStr(i, str0, str10)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(110, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "exact")
                If Sheet1.Cells(110, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(109, col).Value = Len(str0) - (a - ind)
                        Else
                                Sheet1.Cells(109, col).Value = Len(str0) - a + 1
                        End If
                        Sheet1.Cells(111, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "exact")
                        Sheet1.Cells(112, col).Value = "N/A Exact Match was Found"
                        col = col + 1
                End If
        Else
```

```
        Sheet1.Cells(110, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "exact")
        If Sheet1.Cells(110, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(109, col).Value = Len(str0)
                Else
                        Sheet1.Cells(109, col).Value = ind - a
                End If
                Sheet1.Cells(111, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "exact")
                Sheet1.Cells(112, col).Value = "N/A Exact Match was Found"
                col = col + 1
        End If
End If
        i = ind + 1
        GoTo 8
End If

i = 1
col = 3

9
        'AGG PARTIAL MATCH
        ind = InStr(i, str0, str3)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(115, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "partial")
                If Sheet1.Cells(115, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
```

```
                Sheet1.Cells(114, col).Value = Len(str0) - (a - ind)
            Else
                Sheet1.Cells(114, col).Value = Len(str0) - a + 1
            End If
            Sheet1.Cells(116, col).Value2 = _
                possibleDSB(ind, str0, str2, "number", "partial")
            Sheet1.Cells(117, col).Value = _
                highlightDifference1(str2, Sheet1.Cells(115, col).Value)
            col = col + 1
        End If
    Else
        Sheet1.Cells(115, col).Value = _
            possibleDSB(ind, str0, str2, "Nothing", "partial")
        If Sheet1.Cells(115, col).Value <> 0 Then
            If ind - a = 0 Then
                Sheet1.Cells(114, col).Value = Len(str0)
            Else
                Sheet1.Cells(114, col).Value = ind - a
            End If
            Sheet1.Cells(116, col).Value2 = _
                possibleDSB(ind, str0, str2, "number", "partial")
            Sheet1.Cells(117, col).Value = _
                highlightDifference1(str2, Sheet1.Cells(115, col).Value)
            col = col + 1
        End If
        End If
        i = ind + 1
        GoTo 9
    End If
```

```
        i = 1
        col = 3

10
        'CGG PARTIAL MATCH
        ind = InStr(i, str0, str4)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(119, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "partial")
                If Sheet1.Cells(119, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(118, col).Value = Len(str0) - (a - ind)
                        Else
                                Sheet1.Cells(118, col).Value = Len(str0) - a + 1
                        End If
                        Sheet1.Cells(120, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "partial")
                        Sheet1.Cells(121, col).Value = _
                                highlightDifference1(str2, Sheet1.Cells(119, col).Value)
                        col = col + 1
                End If
        Else
                Sheet1.Cells(119, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "partial")
                If Sheet1.Cells(119, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(118, col).Value = Len(str0)
                        Else
                                Sheet1.Cells(118, col).Value = ind - a
```

```
                End If
                Sheet1.Cells(120, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(121, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(119, col).Value)
                col = col + 1
        End If
End If
i = ind + 1
GoTo 10
End If


i = 1
col = 3


11
        'GGG PARTIAL MATCH
        ind = InStr(i, str0, str5)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(123, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "partial")
                If Sheet1.Cells(123, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(122, col).Value = Len(str0) - (a - ind)
                        Else
                                Sheet1.Cells(122, col).Value = Len(str0) - a + 1
                        End If
                        Sheet1.Cells(124, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "partial")
```

```
                Sheet1.Cells(125, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(123, col).Value)
                col = col + 1
        End If
        Else
                Sheet1.Cells(123, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "partial")
                If Sheet1.Cells(123, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(122, col).Value = Len(str0)
                        Else
                                Sheet1.Cells(122, col).Value = ind - a
                        End If
                        Sheet1.Cells(124, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(125, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(123, col).Value)
                col = col + 1
        End If
End If
        i = ind + 1
        GoTo 11
End If


i = 1
col = 3


12
        'TGG PARTIAL MATCH
        ind = InStr(i, str0, str6)
```

```
If ind <> 0 Then
If ind - a < 0 Then
        Sheet1.Cells(127, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "partial")
        If Sheet1.Cells(127, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                        Sheet1.Cells(126, col).Value = Len(str0) - (a - ind)
                Else
                        Sheet1.Cells(126, col).Value = Len(str0) - a + 1
                End If
                Sheet1.Cells(128, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(129, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(127, col).Value)
                col = col + 1
        End If
Else
        Sheet1.Cells(127, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "partial")
        If Sheet1.Cells(127, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(126, col).Value = Len(str0)
                Else
                        Sheet1.Cells(126, col).Value = ind - a
                End If
                Sheet1.Cells(128, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(129, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(127, col).Value)
                col = col + 1
```

```
        End If
        End If
        i = ind + 1
        GoTo 12
    End If

    i = 1
    col = 3

13
    'AAG PARTIAL MATCH
    ind = InStr(i, str0, str7)
    If ind <> 0 Then
    If ind - a < 0 Then
        Sheet1.Cells(132, col).Value = _
            possibleDSB(ind, str0, str2, "Nothing", "partial")
        If Sheet1.Cells(132, col).Value <> 0 Then
            If ind - 1 <> 0 Then
                Sheet1.Cells(131, col).Value = Len(str0) - (a - ind)
            Else
                Sheet1.Cells(131, col).Value = Len(str0) - a + 1
            End If
            Sheet1.Cells(133, col).Value2 = _
                possibleDSB(ind, str0, str2, "number", "partial")
            Sheet1.Cells(134, col).Value = _
                highlightDifference1(str2, Sheet1.Cells(132, col).Value)
            col = col + 1
        End If
    Else
        Sheet1.Cells(132, col).Value = _
```

```
                possibleDSB(ind, str0, str2, "Nothing", "partial")
        If Sheet1.Cells(132, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(131, col).Value = Len(str0)
                Else
                        Sheet1.Cells(131, col).Value = ind - a
                End If
                Sheet1.Cells(133, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(134, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(132, col).Value)
                col = col + 1
        End If
        End If
        i = ind + 1
        GoTo 13
End If

i = 1
col = 3

14
'CAG PARTIAL MATCH
ind = InStr(i, str0, str8)
If ind <> 0 Then
If ind - a < 0 Then
        Sheet1.Cells(136, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "partial")
        If Sheet1.Cells(136, col).Value <> 0 Then
                If ind - 1 <> 0 Then
```

```
                        Sheet1.Cells(135, col).Value = Len(str0) - (a - ind)
                Else
                        Sheet1.Cells(135, col).Value = Len(str0) - a + 1
                End If
                Sheet1.Cells(137, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(138, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(136, col).Value)
                col = col + 1
        End If
Else
        Sheet1.Cells(136, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "partial")
        If Sheet1.Cells(136, col).Value <> 0 Then
                If ind - a = 0 Then
                        Sheet1.Cells(135, col).Value = Len(str0)
                Else
                        Sheet1.Cells(135, col).Value = ind - a
                End If
                Sheet1.Cells(137, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(138, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(136, col).Value)
                col = col + 1
        End If
End If
i = ind + 1
GoTo 14
End If
```

```
        i = 1
        col = 3


15
        'GAG PARTIAL MATCH
        ind = InStr(i, str0, str9)
        If ind <> 0 Then
        If ind - a < 0 Then
                Sheet1.Cells(140, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "partial")
                If Sheet1.Cells(140, col).Value <> 0 Then
                        If ind - 1 <> 0 Then
                                Sheet1.Cells(139, col).Value = Len(str0) - (a - ind)
                        Else
                                Sheet1.Cells(139, col).Value = Len(str0) - a + 1
                        End If
                        Sheet1.Cells(141, col).Value2 = _
                                possibleDSB(ind, str0, str2, "number", "partial")
                        Sheet1.Cells(142, col).Value = _
                                highlightDifference1(str2, Sheet1.Cells(140, col).Value)
                        col = col + 1
                End If
        Else
                Sheet1.Cells(140, col).Value = _
                        possibleDSB(ind, str0, str2, "Nothing", "partial")
                If Sheet1.Cells(140, col).Value <> 0 Then
                        If ind - a = 0 Then
                                Sheet1.Cells(139, col).Value = Len(str0)
                        Else
                                Sheet1.Cells(139, col).Value = ind - a
```

```
                End If
                Sheet1.Cells(141, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(142, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(140, col).Value)
                col = col + 1
        End If
End If
i = ind + 1
GoTo 15
End If


i = 1
col = 3


16
'TAG PARTIAL MATCH
ind = InStr(i, str0, str10)
If ind <> 0 Then
If ind - a < 0 Then
        Sheet1.Cells(144, col).Value = _
                possibleDSB(ind, str0, str2, "Nothing", "partial")
        If Sheet1.Cells(144, col).Value <> 0 Then
                If ind - 1 <> 0 Then
                        Sheet1.Cells(143, col).Value = Len(str0) - (a - ind)
                Else
                        Sheet1.Cells(143, col).Value = Len(str0) - a + 1
                End If
                Sheet1.Cells(145, col).Value2 = _
                        possibleDSB(ind, str0, str2, "number", "partial")
```

```
                Sheet1.Cells(146, col).Value = _
                        highlightDifference1(str2, Sheet1.Cells(144, col).Value)
                col = col + 1
            End If
        Else
            Sheet1.Cells(144, col).Value = _
                    possibleDSB(ind, str0, str2, "Nothing", "partial")
            If Sheet1.Cells(144, col).Value <> 0 Then
                If ind - a = 0 Then
                    Sheet1.Cells(143, col).Value = Len(str0)
                Else
                    Sheet1.Cells(143, col).Value = ind - a
                End If
                Sheet1.Cells(145, col).Value2 = _
                    possibleDSB(ind, str0, str2, "number", "partial")
                Sheet1.Cells(146, col).Value = _
                    highlightDifference1(str2, Sheet1.Cells(144, col).Value)
                col = col + 1
            End If
        End If
        i = ind + 1
        GoTo 16
        End If

        Range("B3").Select
        ActiveCell.FormulaR1C1 = "=revstr(R2C2)"

'SELECT FOR AND MOVE TOWARDS THOSE THAT HAVE RESULTS
        If WorksheetFunction.CountA(Sheet1.Range("C80:C95")) = 0 Then
        If WorksheetFunction.CountA(Sheet1.Range("C97:C112")) = 0 Then
```

```
        If WorksheetFunction.CountA(Sheet1.Range("C114:C129")) = 0 Then
                If WorksheetFunction.CountA(Sheet1.Range("C131:C146")) = 0 Then
                        ActiveWindow.ScrollRow = 1
                        Sheet1.Range("B1").Select
                        MsgBox "No Exact or Partial Matches in the Reverse _
                                Complementary Strand Were Found"
                        GoTo 17
                Else
                        ActiveWindow.ScrollRow = 130
                        Sheet1.Range("A130").Select
                        GoTo 17
                End If
        Else
                ActiveWindow.ScrollRow = 113
                Sheet1.Range("A113").Select
                GoTo 17
        End If
Else
                ActiveWindow.ScrollRow = 96
                Sheet1.Range("A96").Select
                GoTo 17
End If
Else
ActiveWindow.ScrollRow = 77
Sheet1.Range("A77").Select
GoTo 17
End If

17
```

```
End Sub

Function highlightDifference1(refcell As String, testCell As String)

        Dim bolSame As Boolean

        Dim refString As String, testString As String, s As String, strResult As String

        Dim i1 As Long, i2 As Long, i As Long


        bolSame = True

        i1 = 1: i2 = 1


        For i = 1 To Len(refcell)

                refString = mid(refcell, i1, 1)

                testString = mid(testCell, i2, 1)

                s = testString


                If refString = testString Then

                If bolSame = False Then

                        bolSame = True

                        s = ")" & s

                        ' s = "</strong></font>" & s

                End If

                i1 = i1 + 1: i2 = i2 + 1

                Else

                If bolSame = True Then

                        bolSame = False

                        s = "(" & s

                        ' s = "<font color=red><strong>" & s

                End If

                i1 = i1 + 1: i2 = i2 + 1

                End If

                strResult = strResult & s
```

```
        Next


        If bolSame = False Then

        strResult = strResult & ")"

        End If


        highlightDifference1 = strResult

End Function
```

```
'This locates the position number in the sequence I enter and compares it with crRNA
'character by character

Function possibleDSB(ByVal position As Integer, ByVal mysequence As String, _

        ByVal crRNA As String, ByVal output As String, ByVal expa As String)


Dim newtempstring As String

Dim i As Integer

Dim g As Integer

Dim b As Integer

Dim anothernewtempstring As String

Dim a As Integer

Dim m As Integer



m = Sheet1.Range("D1").Value


If position > Len(mysequence) Then

        position = position - Len(mysequence)

End If


a = Len(crRNA)

i = 0

g = 0

b = 0

Do

        If position > a Then

                newtempstring = mid(mysequence, position - a, a)

                If mid(newtempstring, 1 + i, 1) = mid(crRNA, 1 + i, 1) Then

                        g = g + 1

                        GoTo 1
```

```
        Else: b = b + 1
                GoTo 1
        End If
Else
        If position = 0 Then
                anothernewtempstring = ""
                possibleDSB = anothernewtempstring
                GoTo 2
        Else
        If position < a Then
                If position - 1 <> 0 Then
                        newtempstring = _
                                Right(mysequence, Abs(position - a - 1)) _
                                & Left(mysequence, position)
                Else
                        newtempstring = Right(mysequence, a)
                End If

                If mid(newtempstring, 1 + i, 1) = mid(crRNA, 1 + i, 1) Then
                        g = g + 1
                Else: b = b + 1
                        GoTo 1
                End If
        Else
                If position = a Then
                        newtempstring = _
                                Right(mysequence, 1) & mid(mysequence, 1, a - 1)
                        If mid(newtempstring, 1 + i, 1) = _
                        mid(crRNA, 1 + i, 1) Then
                                g = g + 1
```

```
                                Else: b = b + 1

                                        GoTo 1

                                End If

                        End If

                End If

        End If

1

        i = i + 1

        End If

Loop Until i = a


If expa = "partial" Then

        If b > 0 And b <= m Then

        If output = "number" Then

                possibleDSB = b & " mismatches found in sequence"

                GoTo 2

        Else

                If output = "Nothing" Then

                        possibleDSB = newtempstring

                        GoTo 2

                End If

        End If

        Else

                possibleDSB = ""

        End If

Else

        If g = a Then

        If output = "number" Then

                possibleDSB = "0. Exact Match was Found"

                GoTo 2
```

```
        Else
                If output = "Nothing" Then
                        possibleDSB = newtempstring
                        GoTo 2
                End If
        End If
        Else
        possibleDSB = ""
        End If
End If


2


End Function
```

```
Function ATCGcheck(refcell As String)
        Dim lngRow As Long
        Dim response As Integer
        Dim a As Integer
        Dim b As Integer
        Dim c As Integer
        Dim d As Integer
        Dim e As Integer
        Dim f As Integer
        a = 0
        b = 0
        c = 0
        d = 0
        strA = "A"
        strC = "C"
        strG = "G"
        strT = "T"
        lngRow = 1

1
        response = InStr(lngRow, refcell, strA)
        If response <> 0 Then
                a = a + 1
                lngRow = response + 1
                GoTo 1
        End If

        lngRow = 1

2
```

```
        response = InStr(lngRow, refcell, strC)
        If response <> 0 Then
                b = b + 1
                lngRow = response + 1
                GoTo 2
        End If

        lngRow = 1

3
        response = InStr(lngRow, refcell, strG)
        If response <> 0 Then
                c = c + 1
                lngRow = response + 1
                GoTo 3
        End If

        lngRow = 1

4
        response = InStr(lngRow, refcell, strT)
        If response <> 0 Then
                d = d + 1
                lngRow = response + 1
                GoTo 4
        End If

e = a + b + c + d
f = Len(refcell)
```

```
If e < f Then
        ATCGcheck = "fail"
        GoTo 5
Else
        ATCGcheck = "success"
        GoTo 5
End If


5


End Function
```

```
Sub interpretdirect()

        ActiveWindow.ScrollRow = 7

        ActiveWindow.ScrollColumn = 1

End Sub




Sub interpretrev()

        ActiveWindow.ScrollRow = 77

        ActiveWindow.ScrollColumn = 1

End Sub
```

APPENDIX C – STATISTICAL DATA FOR THE SECOND AIM

Two-way ANOVA of Rep Gene Amplification 48 Hours Post Infection of Ad Constructs								
Within each column	compare rows							
Number of families	4							
Number of comparisons per family	6							
Alpha	0.05							
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value			
XX2-in-Puro-AAVS1 Clone 152.74								
Before infection vs. Ad-GFP	-2.55	-135 to 129.9	No	ns	0.9999			
Before infection vs. wtAd	-6.251	-138.7 to 126.2	No	ns	0.9998			
Before infection vs. Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1)	-120.1	-252.6 to 12.36	No	ns	0.0891			
Before infection vs. Ad-Cre-(E3)-AAV-ds-GFP(E1)	-230.3	-362.7 to -97.84	Yes	***	0.0001			
Before infection vs. Ad-Cre-(E3)	-568	-700.4 to -435.5	Yes	****	0.0001			
Before infection vs. Ad-Cre-(E1)	-950.9	-1083 to -818.4	Yes	****	0.0001			
XX2-in-Puro-AAVS1 Clone 152.69								
Before infection vs. Ad-GFP	21.46	-111 to 153.9	No	ns	0.9955			
Before infection vs. wtAd	4.329	-128.1 to 136.8	No	ns	0.9999			
Before infection vs. Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1)	-236.8	-369.3 to -104.4	Yes	****	0.0001			
Before infection vs. Ad-Cre-(E3)-AAV-ds-GFP(E1)	-197	-329.4 to -64.5	Yes	**	0.0013			
Before infection vs. Ad-Cre-(E3)	-649.7	-782.2 to -517.3	Yes	****	0.0001			
Before infection vs. Ad-Cre-(E1)	-1251	-1383 to -1118	Yes	****	0.0001			
XX2-in-Puro Clone 253								
Before infection vs. Ad-GFP	2.073	-130.4 to 134.5	No	ns	0.9999			
Before infection vs. wtAd	2.622	-129.8 to 135.1	No	ns	0.9999			
Before infection vs. Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1)	-6.881	-139.3 to 125.6	No	ns	0.9998			
Before infection vs. Ad-Cre-(E3)-AAV-ds-GFP(E1)	-10.55	-143 to 121.9	No	ns	0.9997			
Before infection vs. Ad-Cre-(E3)	-119.2	-251.7 to 13.26	No	ns	0.0928			
Before infection vs. Ad-Cre-(E1)	-130	-262.4 to 2.478	No	ns	0.0564			
HEK293								
Before infection vs. Ad-GFP	0.005264	-132.4 to 132.5	No	ns	0.9999			
Before infection vs. wtAd	0.002771	-132.5 to 132.5	No	ns	0.9999			
Before infection vs. Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1)	0.002417	-132.5 to 132.5	No	ns	0.9999			
Before infection vs. Ad-Cre-(E3)-AAV-ds-GFP(E1)	0.004993	-132.4 to 132.5	No	ns	0.9999			
Before infection vs. Ad-Cre-(E3)	0.004674	-132.4 to 132.5	No	ns	0.9999			
Before infection vs. Ad-Cre-(E1)	0.004788	-132.4 to 132.5	No	ns	0.9999			
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	N1	N2	q	DF
XX2-in-Puro-AAVS1 Clone 152.74								
Before infection vs. Ad-GFP	61.04	63.59	-2.55	50.04	3	3	0.05096	56
Before infection vs. wtAd	61.04	67.29	-6.251	50.04	3	3	0.1249	56
Before infection vs. Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1)	61.04	181.1	-120.1	50.04	3	3	2.4	56
Before infection vs. Ad-Cre-(E3)-AAV-ds-GFP(E1)	61.04	291.3	-230.3	50.04	3	3	4.603	56
Before infection vs. Ad-Cre-(E3)	61.04	629	-568	50.04	3	3	11.35	56
Before infection vs. Ad-Cre-(E1)	61.04	1012	-950.9	50.04	3	3	19	56
XX2-in-Puro-AAVS1 Clone 152.69								
Before infection vs. Ad-GFP	121.3	99.82	21.46	50.04	3	3	0.4289	56
Before infection vs. wtAd	121.3	117	4.329	50.04	3	3	0.08651	56
Before infection vs. Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1)	121.3	358.1	-236.8	50.04	3	3	4.734	56
Before infection vs. Ad-Cre-(E3)-AAV-ds-GFP(E1)	121.3	318.2	-197	50.04	3	3	3.936	56
Before infection vs. Ad-Cre-(E3)	121.3	771	-649.7	50.04	3	3	12.99	56
Before infection vs. Ad-Cre-(E1)	121.3	1372	-1251	50.04	3	3	24.99	56
XX2-in-Puro Clone 253								
Before infection vs. Ad-GFP	6.089	4.016	2.073	50.04	3	3	0.04143	56
Before infection vs. wtAd	6.089	3.468	2.622	50.04	3	3	0.05239	56
Before infection vs. Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1)	6.089	12.97	-6.881	50.04	3	3	0.1375	56
Before infection vs. Ad-Cre-(E3)-AAV-ds-GFP(E1)	6.089	16.64	-10.55	50.04	3	3	0.2108	56
Before infection vs. Ad-Cre-(E3)	6.089	125.3	-119.2	50.04	3	3	2.382	56
Before infection vs. Ad-Cre-(E1)	6.089	136.1	-130	50.04	3	3	2.598	56
HEK293								
Before infection vs. Ad-GFP	0.006526	0.001262	0.005264	50.04	3	3	0.000105	56
Before infection vs. wtAd	0.006526	0.003755	0.002771	50.04	3	3	5.54E-05	56
Before infection vs. Ad-Cre-(E3)-AAV2.1-CMV-LacZ-nLs(E1)	0.006526	0.004109	0.002417	50.04	3	3	4.83E-05	56
Before infection vs. Ad-Cre-(E3)-AAV-ds-GFP(E1)	0.006526	0.001533	0.004993	50.04	3	3	9.98E-05	56
Before infection vs. Ad-Cre-(E3)	0.006526	0.001852	0.004674	50.04	3	3	9.34E-05	56
Before infection vs. Ad-Cre-(E1)	0.006526	0.001738	0.004788	50.04	3	3	9.57E-05	56

Two-way ANOVA of Rep Gene Amplification Time Course for Ad-Cre(E3)-AAV2.1-CMV-LacZ-nLs(E1)								
Within each column	compare rows							
Number of families	3							
Number of comparisons per family	5							
Alpha	0.05							
Dunnett's multiple comparisons test	Mean	95.00% CI of diff.	Significant?	Summary	Adjusted P Value			
XX2-in-Puro-AAVS1 Clone 152.74								
0h vs. 12h	-86.43	-191.5 to 18.63	No	ns	0.1368			
0h vs. 24h	-155.3	-260.4 to -50.24	Yes	**	0.0019			
0h vs. 48h	-211.2	-316.2 to -106.1	Yes	****	0.0001			
0h vs. 72h	-257.9	-362.9 to -152.8	Yes	****	0.0001			
0h vs. 96h	-107.3	-212.3 to -2.219	Yes	*	0.044			
XX2-in-Puro-AAVS1 Clone 152.69								
0h vs. 12h	-192.5	-297.6 to -87.46	Yes	***	0.0001			
0h vs. 24h	-213.3	-318.3 to -108.2	Yes	****	0.0001			
0h vs. 48h	-176.4	-281.4 to -71.33	Yes	***	0.0004			
0h vs. 72h	-313	-418.1 to -208	Yes	****	0.0001			
0h vs. 96h	-341.8	-446.9 to -236.8	Yes	****	0.0001			
XX2-in-Puro Clone 253								
0h vs. 12h	-12.55	-117.6 to 92.5	No	ns	0.9977			
0h vs. 24h	-6.471	-111.5 to 98.59	No	ns	0.9997			
0h vs. 48h	-15.61	-120.7 to 89.44	No	ns	0.9941			
0h vs. 72h	-79.62	-184.7 to 25.44	No	ns	0.1902			
0h vs. 96h	-90.84	-195.9 to 14.22	No	ns	0.1091			
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	N1	N2	q	DF
XX2-in-Puro-AAVS1 Clone 152.74								
0h vs. 12h	67.16	153.6	-86.43	39.92	3	3	2.165	36
0h vs. 24h	67.16	222.5	-155.3	39.92	3	3	3.89	36
0h vs. 48h	67.16	278.3	-211.2	39.92	3	3	5.29	36
0h vs. 72h	67.16	325	-257.9	39.92	3	3	6.46	36
0h vs. 96h	67.16	174.4	-107.3	39.92	3	3	2.687	36
XX2-in-Puro-AAVS1 Clone 152.69								
0h vs. 12h	41.88	234.4	-192.5	39.92	3	3	4.823	36
0h vs. 24h	41.88	255.1	-213.3	39.92	3	3	5.342	36
0h vs. 48h	41.88	218.3	-176.4	39.92	3	3	4.419	36
0h vs. 72h	41.88	354.9	-313	39.92	3	3	7.842	36
0h vs. 96h	41.88	383.7	-341.8	39.92	3	3	8.563	36
XX2-in-Puro Clone 253								
0h vs. 12h	8.996	21.55	-12.55	39.92	3	3	0.3145	36
0h vs. 24h	8.996	15.47	-6.471	39.92	3	3	0.1621	36
0h vs. 48h	8.996	24.61	-15.61	39.92	3	3	0.3911	36
0h vs. 72h	8.996	88.61	-79.62	39.92	3	3	1.995	36
0h vs. 96h	8.996	99.84	-90.84	39.92	3	3	2.276	36

Two-way ANOVA of Rep Gene Amplification Time Course for Ad-Cre(E3)								
Within each column	compare rows							
Number of families	3							
Number of comparisons per family	5							
Alpha	0.05							
Dunnett's multiple comparisons test	Mean	95.00% CI of diff.	Significant?	Summary	Adjusted P Value			
XX2-in-Puro-AAVS1 Clone 152.74								
0h vs. 12h	-135.4	-428 to 157.2	No	ns	0.6246			
0h vs. 24h	-385.8	-678.4 to -93.24	Yes	**	0.0061			
0h vs. 48h	-837.9	-1131 to -545.3	Yes	****	0.0001			
0h vs. 72h	-1530	-1822 to -1237	Yes	****	0.0001			
0h vs. 96h	-2669	-2962 to -2376	Yes	****	0.0001			
XX2-in-Puro-AAVS1 Clone 152.69								
0h vs. 12h	-179.3	-471.9 to 113.3	No	ns	0.3665			
0h vs. 24h	-203.5	-496.1 to 89.1	No	ns	0.2562			
0h vs. 48h	-720.8	-1013 to -428.2	Yes	****	0.0001			
0h vs. 72h	-877.2	-1170 to -584.6	Yes	****	0.0001			
0h vs. 96h	-1447	-1739 to -1154	Yes	****	0.0001			
XX2-in-Puro Clone 253								
0h vs. 12h	-11.95	-304.6 to 280.6	No	ns	0.9999			
0h vs. 24h	-24.72	-317.3 to 267.9	No	ns	0.9997			
0h vs. 48h	-94.43	-387 to 198.2	No	ns	0.8636			
0h vs. 72h	-236.7	-529.3 to 55.89	No	ns	0.1469			
0h vs. 96h	-255.2	-547.8 to 37.4	No	ns	0.1048			
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	N1	N2	q	DF
XX2-in-Puro-AAVS1 Clone 152.74								
0h vs. 12h	58.35	193.8	-135.4	111.2	3	3	1.218	36
0h vs. 24h	58.35	444.2	-385.8	111.2	3	3	3.47	36
0h vs. 48h	58.35	896.3	-837.9	111.2	3	3	7.537	36
0h vs. 72h	58.35	1588	-1530	111.2	3	3	13.76	36
0h vs. 96h	58.35	2727	-2669	111.2	3	3	24.01	36
XX2-in-Puro-AAVS1 Clone 152.69								
0h vs. 12h	54.06	233.4	-179.3	111.2	3	3	1.613	36
0h vs. 24h	54.06	257.6	-203.5	111.2	3	3	1.83	36
0h vs. 48h	54.06	774.8	-720.8	111.2	3	3	6.483	36
0h vs. 72h	54.06	931.3	-877.2	111.2	3	3	7.89	36
0h vs. 96h	54.06	1501	-1447	111.2	3	3	13.01	36
XX2-in-Puro Clone 253								
0h vs. 12h	9.722	21.68	-11.95	111.2	3	3	0.1075	36
0h vs. 24h	9.722	34.44	-24.72	111.2	3	3	0.2223	36
0h vs. 48h	9.722	104.2	-94.43	111.2	3	3	0.8494	36
0h vs. 72h	9.722	246.4	-236.7	111.2	3	3	2.129	36
0h vs. 96h	9.722	264.9	-255.2	111.2	3	3	2.295	36

Ordinary one-way ANOVA for Doubling Time of Packaging Cell Lines								
Number of families	1							
Number of comparisons per family	3							
Alpha	0.05							
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Significant?	Summary	Adjusted P Value	D-?		
HEK293 vs. XX2-in-Puro-AAVS1 Clone 152.74	-0.766	-10.5 to 8.964	No	ns	0.9932	A	XX2-in-Puro-AAVS1 Clone 152.74	
HEK293 vs. XX2-in-Puro-AAVS1 Clone 152.69	-0.5034	-10.23 to 9.226	No	ns	0.998	B	XX2-in-Puro-AAVS1 Clone 152.69	
HEK293 vs. XX2-in-Puro Clone 253	0.4368	-9.293 to 10.17	No	ns	0.9986	C	XX2-in-Puro Clone 253	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
HEK293 vs. XX2-in-Puro-AAVS1 Clone 152.74	20.92	21.68	-0.766	3.626	4	4	0.2112	12
HEK293 vs. XX2-in-Puro-AAVS1 Clone 152.69	20.92	21.42	-0.5034	3.626	4	4	0.1388	12
HEK293 vs. XX2-in-Puro Clone 253	20.92	20.48	0.4368	3.626	4	4	0.1204	12